Antonio Incampo è Professore ordinario di Filosofia del diritto all'Università degli Studi di Bari Aldo Moro. È stato Direttore del "Dipartimento per lo studio del Diritto penale, del Diritto processuale penale e della Filosofia del diritto" e Presidente dei Corsi di laurea della Seconda Facoltà di Giurisprudenza nella stessa Università. Autore di numerosi studi di filosofia della validità giuridica, di semiotica e logica del linguaggio normativo, dirige per i tipi di Cacucci la collana editoriale "Unità del sapere giuridico. Quaderni di scienze penalistiche e filosofico-giuridiche". È stato insignito nel 1998 dalla Presidenza della Repubblica d'Austria della "Gran Decorazione d'Onore" [*Großes Ehrenzeichen*] per meriti culturali

Nicola Triggiani è Professore ordinario di Diritto processuale penale all'Università degli Studi di Bari Aldo Moro e avvocato. È stato Direttore del "Dipartimento per lo studio del Diritto penale, del Diritto processuale penale e della Filosofia del diritto" e Coordinatore del Dottorato di ricerca in "Sistema penale e processo" nella stessa Università, nonché Giudice onorario presso il Tribunale per i minorenni di Bari. Autore di monografie e altri studi soprattutto in tema di investigazioni, prove e rapporti tra *media* e giustizia penale, è componente del comitato di direzione della rivista di classe A "Processo penale e Giustizia" e codirettore per i tipi di Cacucci della collana editoriale "Giustizia penale della post-modernità".

La leggera colomba [...] potrebbe immaginare di volare più agevolmente in uno spazio privo d'aria

Immanuel Kant, *Critica della ragion pura* (1787)

ISBN 979-12-5965-606-3

copia omaggio

24

GIUDIZIO PENALE E INTELLIGENZA ARTIFICIALE a cura di A. Incampo e N. Triggiani

Unità del sapere giuridico

Quaderni di scienze penalistiche e filosofico-giuridiche

Collana diretta da
Antonio Incampo, Vito Mormando, Adolfo Scalfati

Giudizio penale e intelligenza artificiale

Una riflessione sistemica

a cura di Antonio Incampo e Nicola Triggiani

CACUCCI EDITORE
BARI

Il volume esplora l'impatto dell'intelligenza artificiale sul diritto, mettendo in luce le profonde trasformazioni ontologiche, epistemologiche e normative che i sistemi esperti introducono nella razionalità giuridica contemporanea. Attraverso l'originale intreccio interdisciplinare di contributi teorici, tecnici e applicativi, l'opera riflette criticamente sulle nuove sfide poste dall'automazione dei processi decisionali e sui rischi connessi alle forme predittive di governance penale e all'uso della logica algoritmica nei contesti giudiziari e amministrativi. Ne emerge il ritratto di un diritto in transizione, chiamato a confrontarsi con un linguaggio "muto e opaco", fondato sul calcolo più che sul significato, e con l'esigenza di preservare la centralità dell'essere umano in un ecosistema digitale dominato dalla potenza computazionale e dall'interconnessione globale.

The volume examines the impact of artificial intelligence on law, highlighting the profound ontological, epistemological and normative transformations that expert systems introduce into contemporary legal rationality. Through an original interdisciplinary interplay of theoretical, technical, and applied contributions, the work critically addresses the new challenges posed by the automation of decision-making processes, the risks associated with predictive forms of criminal governance, and the use of algorithmic logic in judicial and administrative contexts. What emerges is a portrait of a law in transition, called upon to confront a "mute and opaque" language grounded in calculation rather than meaning, and the need to preserve the centrality of the human being within a digital ecosystem dominated by computational power and global interconnection.

Annalisa la Forgia, *Non nomina nuda tenemus*
olio su tela, 37 x 38 cm (2012)
collezione privata

ISBN 979-12-5965-606-3

copia omaggio

GIUDIZIO PENALE E INTELLIGENZA ARTIFICIALE a cura di A. Incampo e N. Triggiani

24

Unità del sapere giuridico

Quaderni di scienze penalistiche e filosofico-giuridiche

Collana diretta da
Antonio Incampo, Vito Mormando, Adolfo Scalfati

Giudizio penale e intelligenza artificiale

Una riflessione sistemica

a cura di Antonio Incampo e Nicola Triggiani

CACUCCI EDITORE
BARI

Unità del sapere giuridico
Quaderni di scienze penalistiche e filosofico-giuridiche

Collana diretta da
Antonio Incampo, Vito Mormando, Adolfo Scalfati

Comitato scientifico

Paolo Ferrua, Angiola Filipponio, Cristina Hermida del Llano, Heribert Franz Köck, Oliviero Mazza, Mariano Menna, Francesco Palazzo, Carlo Enrico Paliero, Luigi Pannarale, Otto Pfersmann, Aldo Regina, Herbert Schambeck, Giorgio Spangher, Gunther Teubner, Nicola Triggiani, Daniela Vigoni, Wojciech Żełaniec.

Comitato di redazione

Filippo Bottalico, Marilena Colamussi, Enzo Dell'Andro, Lucia Iandolo Pisanelli, Piero Marra, Maria Antonella Pasculli, Porzia Teresa Persio, Guglielmo Siniscalchi, Maurizio Sozio.

I volumi inseriti nella collana, ritenuti preliminarmente ammissibili dai Direttori con l'ausilio del Comitato Scientifico, sono sottoposti alla revisione tra pari [*peer review*] secondo i regolamenti attualmente vigenti per le pubblicazioni scientifiche.

I contributi di Miriana Calvano, Antonio Curci, Wanda D'Avanzo, Andrea Esposito, Antonio Incampo, Rosa Lanzilotti, Francesca Alessandra Lisi, Donato Malerba, Piero Marra, Salvatore Antonello Parente, Antonio Piccinno, Lorenzo Pulito, Nicola Triggiani, sono parzialmente sostenuti dal progetto FAIR – Future AI Research (PE00000013), nell'ambito del programma MUR del PNRR finanziato dal NextGenerationEU e illustrano i risultati della ricerca nell'ambito del medesimo progetto.

Il volume, in accesso aperto, è pubblicato con i fondi dell'Università degli Studi di Bari Aldo Moro afferenti allo Spoke 6 (Symbiotic AI) del progetto FAIR – Future AI Research (PE00000013), nell'ambito del programma MUR del PNRR finanziato dal NextGenerationEU.

Giudizio penale e intelligenza artificiale

Una riflessione sistemica

a cura di
Antonio Incampo e Nicola Triggiani

CACUCCI EDITORE
BARI

© 2026 Cacucci Editore – Bari
Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220
http://www.cacucciditore.it e-mail: info@cacucci.it

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.

Indice

Prefazione 9

I.
Tecnologie

Miriana Calvano, Antonio Curci, Andrea Esposito, Rosa Lanzilotti, Antonio Piccinno, *L'*AI Act *come rampa per un'intelligenza artificiale simbiotica* 15

Valentina Colcelli, *IA e principio di equità: base dati, collaborazione interdisciplinare e nuova generazione di giuristi* 33

Maurizio Goretti, *Geopolitica dell'intelligenza artificiale: reti,* data center *e sovranità digitale* 51

Francesca Alessandra Lisi, *Intelligenza Artificiale Affidabile: dalle applicazioni in ambito forense alla simbiosi fra umani e macchine intelligenti* 87

Donato Malerba, *Dall'IA agente all'IA simbiotica: un paradigma relazionale dell'intelligenza artificiale* 109

II.
Diritto

Giovanni Canzio, AI Act *e processo penale: sfide e opportunità* 139

Santo Davide Ferrara, *Il futuro artificiale delle Scienze Bio-Medico-Legali e Forensi* 157

Luca Lupária Donati, *Giustizia tecnologica* vs. *tecnologia giudiziaria. Il processo penale al crocevia* 171

Vincenzo Muscatiello, *L'euristica del* quasi. *Il diritto penale nell'era dell'intelligenza superficiale* 185

Salvatore Antonello Parente, *Intelligenza artificiale simbiotica tra procedimento di accertamento e processo tributario* 209

Lorenzo Pulito, *Il principio di proporzionalità algoritmica* 245

Serena Quattrocolo, *Nuove sfide per la giustizia penale: la rilevanza processuale di immagini e suoni sintetici* 279

Nicola Triggiani, *L'utilizzo investigativo e probatorio dei sistemi automatici di riconoscimento facciale dopo l'*AI Act *e la l. n. 132/2025: dubbi e perplessità* 307

Cristiana Valentini, *L'IA e la 'scoperta' dei* bias *cognitivi nel processo penale* 339

III.
Filosofia

Salvatore Amato, *Condizioni processuali e condizionamenti digitali. Un algoritmo per il processo penale?* 369

Agata Amato Mangiameli, Maria Novella Campagnoli, *Intelligenza artificiale e sistema penale. Nuove applicazioni e nuovi paradigmi* 391

Wanda D'Avanzo, *Intelligenza artificiale e giustizia predittiva. L'orientamento giuridico europeo* 415

Antonio Incampo, *L'intelligenza artificiale a confine dell'universo del diritto* 433

Giuseppe Lorini, Olimpia Luciana Loddo, Simone Pinna, *Artefatti regolativi digitali, intelligenza artificiale e diritti dell'utente* 451

Piero Marra, *Proceduralità dell'intelligenza artificiale simbiotica* 469

Paolo Moro, *Dialettica processuale e intelligenza artificiale. Il problema del metodo* 487

Paolo Sommaggio, *L'intelligenza artificiale tra tetica, zetetica e maieutica* 505

Autori 539

Indice dei nomi 543

Prefazione

L'intelligenza artificiale è tra i fattori di cambiamento più profondi e pervasivi delle società contemporanee. Essa incide in modo strutturale sui saperi, sulle istituzioni e sulle forme dell'agire umano. Il giudizio penale, come il processo giuridico in generale, si trova al centro di questa metamorfosi, chiamato a confrontarsi con tecnologie che non offrono soltanto strumenti, ma introducono anche logiche inedite di conoscenza, decisione e governo. L'IA, infatti, non si limita a replicare il ragionamento umano: opera attraverso una razionalità algoritmica priva di intenzionalità emotiva, espressa in un linguaggio digitale "muto" e "opaco", basato su calcoli e correlazioni più che su significati e argomentazioni. La sua rivoluzione è, in larga misura, una rivoluzione grafica che sostituisce le parole con sequenze numeriche. Gli effetti sono evidenti. Genera risposte, ma non ha alcuna consapevolezza di ciò che "dice"; interviene nei processi decisionali, ma senza comprenderne appieno il senso. Le implicazioni ontologiche, epistemologiche e normative di questa trasformazione sono molteplici e tali da mettere in discussione categorie giuridiche consolidate, oltre al ruolo stesso del giurista.

Il volume muove innanzitutto da una questione di metodo che assegna all'interdisciplinarietà il valore di requisito essenziale per comprendere adeguatamente le nuove tecnologie nel campo del diritto. In questa prospettiva, grazie al progetto di partenariato esteso FAIR [*Future AI Research*][1], mette insieme, in un'opera davvero unica nel suo genere, le ricerche di giuristi, magistrati,

[1] FAIR è un progetto di partenariato esteso alle università, ai centri di ricerca, alle aziende per il finanziamento di progetti di ricerca di base, nell'ambito del programma MUR del PNRR finanziato dal *NextGenerationEU*. Il volume che qui presentiamo è, in particolare, uno dei prodotti dello *Spoke* 6 (*Symbiotic AI*), che si svolge presso l'Università degli Studi di Bari Aldo

informatici, matematici e imprenditori del settore tecnologico. La forza di ogni analisi e di ogni conclusione risiede proprio nel confronto costante tra prospettive diverse attraverso cui i fenomeni sono osservati e interpretati. È il risvolto propriamente teoretico del metodo.

Quali sono i risultati? Alla base c'è una profonda convinzione. Dalle forme di *governance* anticipatoria e di giudizi predittivi nella giustizia civile e penale, all'uso di sistemi esperti di riconoscimento facciale e di analisi comportamentale, fino all'impiego dell'IA nella sanità, nella medicina legale e nella gestione della prova digitale, emergono opportunità senza precedenti, ma anche rischi significativi per i diritti fondamentali, la libertà personale, la responsabilità penale e le garanzie del giusto processo. In questo scenario, il pericolo di una progressiva riduzione dell'essere umano a oggetto passivo di processi automatizzati impone un'analisi critica del rapporto tra la decisione algoritmica e quella giuridica.

Le direttrici principali sono tre: (i) applicativa, (ii) regolatoria e (iii) riflessiva. Ognuna di queste corrisponde a un capitolo del volume: (i) Tecnologie, (ii) Diritto e (iii) Filosofia.

Sul piano *applicativo*, il volume affronta il tema della fiducia nell'intelligenza artificiale, esplorando ambiti cruciali quali la *digital forensics*, l'analisi delle prove, la medicina legale e i modelli di *human-AI symbiosis*. Attraverso casi di studio e ricerche applicate, emerge l'esigenza di sviluppare soluzioni tecniche conformi ai requisiti di trasparenza, spiegabilità e controllo umano, nella prospettiva di una "giustizia ibrida" in cui l'IA rimanga uno strumento di supporto e non un sostituto del ragionamento dialettico e argomentativo che caratterizza la decisione giudiziaria.

Poi c'è l'aspetto *regolatorio*. Particolare attenzione è dedicata al quadro normativo europeo, *in primis* all'*AI Act*, letto come espressione di una complessa intersezione tra le dimensioni tecnico-giuridiche ed etiche. Le clausole aperte, le deroghe e le previsioni specifiche in materia di *law enforcement* evidenziano la necessità di un'armonizzazione dinamica tra il diritto dell'Unio-

Moro e che vede strutturalmente coinvolte competenze diverse che spaziano da quelle tecnico-scientifiche a quelle giuridiche e filosofiche.

ne e le legislazioni nazionali, nonché di un rinnovato impegno interpretativo da parte dei giuristi. La regolazione dell'IA si presenta come una sfida storica che richiede non solo nuove norme, ma anche nuove competenze, nuovi modelli organizzativi e una formazione giuridica capace di governare l'interazione tra esseri umani e sistemi esperti.

La riflessione si estende, infine, alle implicazioni *filosofiche* dell'intelligenza artificiale. Il volume si interroga sui paradigmi morali più idonei a orientarne lo sviluppo – dalla deontologia al consequenzialismo, fino all'etica delle virtù – e sulla distinzione tra artefatti digitali deontici e non deontici, capaci di incidere sulla libertà individuale anche al di fuori delle tecnologie propriamente "intelligenti". Parallelamente, l'analisi dell'infrastruttura materiale dell'IA – *data center*, reti, *chip* e potenza di calcolo – ne mette in evidenza la dimensione geopolitica e la natura di bene infrastrutturale globale, sollevando interrogativi cruciali in tema di sovranità digitale, sostenibilità e cooperazione internazionale.

In conclusione, l'intelligenza artificiale non è soltanto una questione tecnologica, ma anche una sfida culturale e giuridica di portata epocale. Governarla significa ripensare il diritto come spazio di mediazione tra innovazione e tutela dei valori fondamentali, riaffermando la centralità della persona, della responsabilità e del controllo umano. Solo attraverso un approccio interdisciplinare, critico e consapevole sarà possibile costruire un futuro in cui l'IA possa porsi davvero al servizio dell'uomo e della giustizia, senza comprometterne i principi fondanti. D'altronde, a tradire questo futuro potrebbero essere proprio i suoi stessi vantaggi. Si pensi alla facilità di accesso e alla sorprendente accelerazione delle risposte alle nostre domande. Per comprenderne i pericoli, non vi è forse immagine migliore della celebre colomba di Kant: la colomba, nel suo volo, avvertendo l'aria come resistenza, potrebbe credere di volare meglio nel vuoto assoluto. Nulla, però, sarebbe più sbagliato per sé e per il suo stesso volo.

Bari, nel giorno di S. Silvestro 2025

Antonio Incampo e Nicola Triggiani

I.
TECNOLOGIE

M. Calvano, A. Curci, A. Esposito,
R. Lanzilotti, A. Piccinno

L'*AI Act* come rampa per un'intelligenza artificiale simbiotica

Abstract: **The AI Act as a Road Toward Symbiotic Artificial Intelligence.** Artificial Intelligence (AI) is rapidly spreading, impacting many areas of our lives. In this context, to ensure the protection of the fundamental rights of human beings and of the environment that surrounds them, it is necessary to create systems that comply with the law. In this regard, the European Union has introduced the AI Act, a regulatory framework aimed at governing the development and use of AI through a risk-based approach. At the same time, researchers emphasize the need to design human-centered AI systems capable of fostering a symbiotic relationship between humans and AI. The objective of Symbiotic AI systems is to enhance rather than replace humans through continuous collaboration and learning. Considering the need to create legally compliant Symbiotic AI systems, this chapter describes the FrameSAI framework (*Framework to create Symbiotic Artificial Intelligence*), which aims to guide practitioners throughout the creation process of Symbiotic AI systems compliant with the AI Act.

Keywords: Symbiotic AI, European AI Act, Symbiosis-by-design, Human-Centered.

Sommario: 1. Introduzione – 2. Intelligenza artificiale simbiotica e centrata sulle persone – 3. L'*AI Act* per la simbiosi – 4. Indicazioni pratiche – 5. Conclusioni.

1. Introduzione

L'intelligenza artificiale (generalmente abbreviata in IA o, partendo dall'inglese *artificial intelligence*, AI) è ormai ubiqua. Ci interfacciamo con sistemi di IA giornalmente: basta pensare ai motori di raccomandazione che permettono a servizi di *streaming* digitale (per esempio, Spotify e Netflix) di suggerire il prossimo contenuto da visualizzare. Un altro esempio è fornito dai *chatbot*, a partire da quelli con cui siamo costretti a interfacciarci quando vogliamo contattare l'assistenza clienti di varie aziende, fino a quelli basati su IA generativa e *large language*

models (LLM), come ChatGPT, che utilizziamo per ottenere informazioni o per scrivere testo.

In aggiunta ai diversi domini sopra citati, i sistemi di IA promettono di rivoluzionare settori che impattano in modo significativo la vita degli esseri umani, come quello medico e legale. Per citare un esempio, vari studi mirano a realizzare sistemi di IA in grado di riconoscere e diagnosticare malattie come il morbo di Alzheimer o tumori cerebrali nelle loro fasi iniziali[1]. Un ulteriore esempio è fornito dai casi in cui l'IA è utilizzata in contesti giudiziari per riconoscere sospettati di un crimine e intervenire in maniera tempestiva[2].

Tuttavia, l'uso di tecniche di IA non è privo di rischi, ed è una scelta che va ponderata adeguatamente. Infatti, le tecniche di IA attualmente maggiormente utilizzate (in quanto più performanti) sono tecniche di *machine learning*, o più precisamente di *deep learning*, che sono spesso caratterizzate sia dalla loro impossibilità di rendere note le motivazioni dietro le decisioni prese, sia dall'impossibilità di lasciare il controllo delle azioni all'essere umano. È noto, infatti, quanto sia pericoloso automatizzare un processo senza considerare appieno l'elemento umano con cui, inevitabilmente, il sistema deve interfacciarsi: i suoi utenti. Esistono diversi esempi a supporto di questa necessità, raccolti sul sito web *AI Incident Database*[3], il quale riporta dati riguardo agli incidenti causati dall'intelligenza artificiale.

[1] F. Altay *et al.*, "Preclinical stage Alzheimer's disease detection using magnetic resonance image scans", in *Proceedings of the AAAI conference on artificial intelligence*, 35 (2021), pp. 15088-15097, https://doi.org/10.1609/aaai.v35i17.17772; G. Castellano *et al.*, "Automated Detection of Alzheimer's Disease: A Multi-Modal Approach with 3D MRI and Amyloid PET", in *Scientific Reports* 14 (2024), n. 5210, https://doi.org/10.1038/s41598-024-56001-9; A. Curci, A. Esposito, "Detecting Brain Tumors Through Multimodal Neural Networks", in *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods*, SCITEPRESS - Science and Technology Publications, 26 febbraio 2024, pp. 995-1000, https://doi.org/10.5220/0012608600003654.

[2] S. Raaijmakers, "Artificial Intelligence for Law Enforcement: Challenges and Opportunities", in *IEEE Security & Privacy* 17 (2019), n. 7, pp. 74-77, https://doi.org/10.1109/MSEC.2019.2925649.

[3] https://incidentdatabase.ai

In tale contesto, funge da ammonimento il rapporto della *National Transportation Safety Board* (NTSB, un'agenzia investigativa indipendente del Governo degli USA che indaga ed emette rapporti in merito agli incidenti che coinvolgono aeroplani, navi, treni, oleodotti e gasdotti) su un incidente tra una Tesla e un autocarro con rimorchio. Il passaggio più interessante è il seguente (tradotto dagli autori):

> Poiché oltre il 90% degli incidenti stradali è attribuito a errori umani, l'automazione nei veicoli offre un notevole potenziale per salvare decine di migliaia di vite ogni anno, sostituendo progressivamente il conducente. Tuttavia, introdurre l'automazione in un ambiente così complesso e non strutturato sarà molto impegnativo e dovrà essere affrontato con attenzione e grande cautela.
>
> La storia iniziale dell'automazione in aviazione comprende molti esempi in cui l'automazione è stata introdotta semplicemente perché l'industria possedeva la capacità tecnologica per farlo. L'esperienza ha insegnato che automatizzare "perché possiamo" non significa necessariamente migliorare il funzionamento del sistema uomo-automazione. Questo ha portato a un'evoluzione verso un'automazione "umano-centrica", il cui obiettivo è migliorare le prestazioni complessive del sistema uomo-automazione. Questo incidente è un esempio di ciò che può accadere quando l'automazione viene introdotta "perché possiamo" senza un'adeguata considerazione del fattore umano.[4]

In risposta ai rischi associati all'uso dell'IA, l'Unione Europea (UE) ha introdotto il *framework* legislativo comunemente denominato "AI Act"[5]. L'AI Act è caratterizzato da un insieme di limitazioni e indicazioni sull'uso dell'IA in base al livello di *rischio* associato in contesti specifici. Più precisamente, l'AI Act

[4] National Transportation Safety Board, *Collision Between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016*, Highway Accident Report NTSB/HAR-17/02 (National Transportation Safety Board, 2017), https://www.ntsb.gov/investigations/accidentreports/reports/har1702.pdf.

[5] Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (2024).

individua quattro livelli di rischio: minimale, limitato, alto, inaccettabile; maggiore è il livello di rischio, più restrizioni ci sono sull'uso dell'IA – fino ad arrivare al totale divieto in contesti con rischio inaccettabile (per esempio, per applicazioni di *social scoring*).

Parallelamente all'introduzione nel panorama legislativo dell'AI Act, la comunità accademica ha riconosciuto i rischi dell'uso di IA, portando alla nascita di una nuova corrente di ricerca denominata "intelligenza artificiale centrata sulle persone" (o, in inglese, *human-centered AI*, abbreviato in HCAI)[6]. Differentemente dal classico approccio all'IA, che si focalizza sulle *performance* tecniche del modello intelligente, la HCAI mira a creare sistemi che siano comprensibili, controllabili e allineati con gli obiettivi degli esseri umani – ossia noi, gli utenti finali degli stessi sistemi. La HCAI cambia il modo in cui gli utenti sono concepiti all'interno del ciclo di vita di un sistema di IA: la visione classica li pone a semplici osservatori passivi, che si limitano a ricevere gli *output* del sistema; al contrario, la HCAI li eleva a una posizione attiva, rendendo il processo decisionale un processo collaborativo. Questa visione dell'utente è ulteriormente specializzata nel concetto di IA *simbiotica*[7].

Prendendo in prestito dalla terminologia comunemente usata in biologia, l'IA simbiotica instaura un rapporto molto più profondo con i propri utenti umani. Questo rapporto, collaborativo per natura, è caratterizzato da un mutuale miglioramento delle due forme di intelligenza (umana e artificiale). Infatti, un'IA simbiotica mira a *migliorare* (invece di sostituire) il proprio utente umano, aumentandone le capacità e le prestazioni; al contempo, l'IA viene costantemente migliorata apprendendo continuamente dalle azioni che compie l'utente.

Benché l'approccio basato sul rischio dell'AI Act e le visioni di IA simbiotica e centrata sull'essere umano possano sembrare

[6] B. Shneiderman, *Human-Centered AI*, 1ª ed., Oxford University Press, Oxford, 2022, https://doi.org/10.1093/oso/9780192845290.001.0001.

[7] S.S. Grigsby, "Artificial Intelligence for Advanced Human-Machine Symbiosis", in D.D. Schmorrow, C.M. Fidopiastis (eds.), *Augmented Cognition: Intelligent Technologies*, vol. 10915, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-319-91470-1_22.

distanti fra loro, sono in realtà estremamente vicine. Infatti, uno dei principi fondamentali dell'AI Act è la necessità di sviluppare sistemi "*human-centric*" che lascino il controllo all'essere umano, rispettandone i diritti e le norme etiche della società in cui viviamo. Tuttavia, una delle principali difficoltà dell'implementazione dell'AI Act è proprio nella differenza di linguaggio tra la legislazione e la matematica (linguaggio basilare dell'IA). Questo capitolo si pone quindi l'obiettivo di delineare un "ponte" fra i due mondi, fornendo indicazioni concrete. Saranno descritti linee guida e criteri di successo, estratti dalla letteratura e componenti di *FrameSAI* (dall'inglese: *Framework to create Symbiotic Artificial Intelligence*)[8], per la creazione di sistemi di IA simbiotici nel pieno rispetto dell'AI Act – punto di partenza della definizione delle linee guida e dei criteri di successo stessi.

2. Intelligenza artificiale simbiotica e centrata sulle persone

In questa sezione, saranno forniti il contesto e la terminologia necessari a guidare la lettura di *FrameSAI*. Il fulcro delle linee guida e dei criteri di successo presentati in questo capitolo è il concetto di intelligenza artificiale simbiotica, che a sua volta, si basa sulla più generale disciplina dell'intelligenza artificiale centrata sulle persone (si veda fig. 1).

[8] M. Calvano *et al.*, "FrameSAI: a Three-Layer Framework to Create Symbiotic AI Systems", in *Workshop "AI Responsabile e Affidabile" co-located with the 5th National Conference on Artificial Intelligence, organized by CINI (Ital-IA 2025)*, CEUR Workshop Proceedings, 2025.

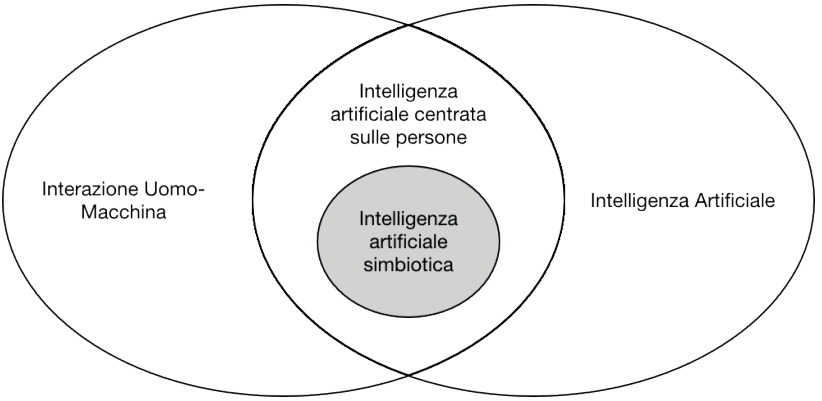

Figura 1. Relazione tra IA simbiotica e IA centrata sulle persone

L'interazione persona-macchina è un campo multidisciplinare (è basato sull'informatica, ma anche su psicologia, ergonomia, ecc.) che si focalizza sulla progettazione e sulla valutazione di sistemi interattivi (tipicamente informatici) enfatizzando l'importanza dell'interazione tra gli utenti e i sistemi[9]. L'intelligenza artificiale, invece, si occupa dello studio di agenti che mostrano caratteristiche assimilabili a un agente "intelligente", essendo quindi in grado di percepire il mondo loro circostante e agire autonomamente[10]. L'intelligenza artificiale centrata sulle persone (dall'inglese *human-centered artificial intelligence*, HCAI), quindi, ha l'obiettivo di progettare e valutare sistemi interattivi *intelligenti*, focalizzandosi sull'interazione tra essi e i loro utenti, affinché siano affidabili, sicuri e degni di fiducia[11]. Più precisamente, i sistemi di intelligenza artificiale centrati sulle persone sono definiti come segue:

> I sistemi di intelligenza artificiale centrati sulle persone (Human-Centered AI) sono sistemi di IA progettati, sviluppati e valutati coinvolgendo gli utenti nel processo, con l'obiettivo di aumentare le prestazioni e la soddisfazione delle persone in specificati contesti. I sistemi HCAI mirano a esse-

[9] A. Dix (ed.), *Human-Computer Interaction*, 3ª ed., Pearson Prentice-Hall, Upper Saddle River, 2010.

[10] S.J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 3ª ed., Pearson, London, 2016.

[11] B. Shneiderman, *Human-Centered AI*, cit.

> re utili e usabili per specificati utenti, i quali possono essere descritti attraverso un modello formale, al fine di raggiungere i propri obiettivi all'interno del loro contesto d'uso, garantendo al contempo affidabilità, sicurezza d'uso e fiducia[12].

Uno dei capisaldi dell'IA centrata sulle persone è, quindi, la necessità di sviluppare sistemi che *potenzino* le capacità umane, preservino l'autonomia dell'utente e includano meccanismi di responsabilità e *governance*[13]. Partendo da queste premesse, è emersa una nuova direzione identificata con il termine "IA *simbiotica*"[14] che immagina sistemi progettati non solo per potenziare le capacità umane in un dato momento, ma anche per apprendere ed evolversi continuamente attraverso l'interazione.

L'IA simbiotica aspira a promuovere una relazione in cui sia il sistema di IA sia l'utente umano crescono e si adattano *insieme*, migliorando reciprocamente le proprie capacità in una collaborazione continua e reciprocamente vantaggiosa tra intelligenza artificiale e intelligenza umana. Così come avviene per la HCAI, quando correttamente implementati, i sistemi simbiotici mirano a supportare i propri utenti senza sostituirli, riconoscendo che gli utenti restano partecipanti essenziali nel ciclo d'interazione: invece di automatizzare completamente decisioni o azioni, i sistemi di IA possono apprendere dai comportamenti, dalle preferenze e dalle correzioni degli utenti, adattandosi nel tempo per allinearsi sempre meglio agli obiettivi umani. Possiamo quindi definire un sistema di IA simbiotica come segue:

> I sistemi di Intelligenza Artificiale Simbiotica (*Symbiotic AI*, SAI) sono sistemi HCAI basati su una collaborazione continua e profonda tra esseri umani e intelligenza artificiale, ovvero una simbiosi tra intelligenza umana e intelligenza artificiale. In questi sistemi, esseri umani e IA si potenziano

[12] G. Desolda *et al.*, "From Human-Centered to Symbiotic Artificial Intelligence: A Focus on Medical Applications", in *Multimedia Tools and Applications* 84 (2024), n. 27, pp. 32109-32150, https://doi.org/10.1007/s11042-024-20414-5.

[13] B. Shneiderman, "Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems", in *ACM Transactions on Interactive Intelligent Systems* 10 (2020), n. 4, pp. 1-31, https://doi.org/10.1145/3419764.

[14] S.S. Grigsby, *op. cit.*

> reciprocamente, bilanciando i rispettivi punti di forza e di debolezza, senza compromettere né l'autonomia degli esseri umani né le prestazioni dell'IA, mantenendo il controllo delle decisioni del sistema nelle mani dell'essere umano. In un sistema SAI, l'IA trae beneficio da un flusso continuo di nuovi dati forniti dagli utenti per perfezionarsi, mentre gli esseri umani traggono vantaggio dal miglioramento delle prestazioni e delle conoscenze dell'IA[15].

La crescente adozione e integrazione di sistemi di IA in vari processi decisionali a cui stiamo assistendo rendono fondamentale una modifica alle modalità di progettazione, sviluppo e valutazione di sistemi IA. L'uso sempre crescente di IA in domini particolarmente delicati (come la medicina) rende infatti necessario un approccio che consenta agli utenti di mantenere il controllo del processo decisionale, lasciando loro la stessa autonomia di giudizio che avrebbero senza l'uso di sistemi intelligenti, ma supportandoli per rendere le loro attività più rapide e ridurre il numero di errori. È questa necessità che sta alimentando la crescita dell'adozione del "paradigma simbiotico" sopra descritto.

Infatti, i sistemi di IA simbiotica, collaborando con i loro utenti umani invece di sostituirli, garantiscono la tutela dei loro diritti fondamentali, oltre che dell'ambiente in cui sono utilizzati. In questo contesto, l'AI Act non è solo una componente dell'ambiente circostante, né tantomeno una semplice legge che impone delle limitazioni o istituisce dei requisiti (come la necessità di registrare sistemi ad alto rischio in un apposito *database* della Commissione Europea), ma diventa un'utile guida che supporti i progettisti di sistemi intelligenti a creare IA che non rappresentino un rischio per gli esseri umani[16].

3. L'*AI Act* per la simbiosi

In questo contesto, è necessario guidare i progettisti verso l'adozione di un approccio per la creazione di sistemi di IA

[15] G. Desolda *et al.*, "From Human-Centered to Symbiotic Artificial Intelligence", cit.

[16] M. Wörsdörfer, "Mitigating the Adverse Effects of AI with the European Union's Artificial Intelligence Act: Hype or Hope?", in *Global Business and Organizational Excellence*, 14 novembre 2023, joe.22238, https://doi.org/10.1002/joe.22238.

chiamato *symbiosis-by-design*, che mira a creare sistemi di IA simbiotica che rispettano gli standard etici delineati nell'AI Act. Tale approccio stabilisce requisiti specifici per i sistemi di IA progettati per lavorare a stretto contatto con gli esseri umani, che si basano su quattro principi fondamentali, che caratterizzano l'IA simbiotica: Livello di Automazione, Trasparenza, Equità e Protezione[17].

Il principio del *livello di automazione* si traduce in meccanismi di interazione specifici che devono essere stabiliti per raggiungere una relazione simbiotica tra le due parti coinvolte. Ad esempio, i sistemi di IA devono essere progettati garantendo un adeguato controllo e supervisione umana (la scelta o il compromesso tra le due opzioni è fortemente influenzata dal dominio specifico in cui l'IA è utilizzata). Un'attenzione particolare deve essere posta anche alla *trasparenza* dell'IA, in modo che possa essere monitorata dagli esseri umani. Affinché ciò sia possibile, gli esseri umani devono poter comprendere il comportamento dell'IA, che deve fornire spiegazioni chiare che consentono di comprendere le motivazioni dell'*output* fornito. Inoltre, questo principio supporta anche quello dell'*equità*, poiché i sistemi di IA devono evitare comportamenti discriminatori e assicurare l'uso di informazioni corrette e legittime. Infine, durante il processo di interazione, è importante assicurare la protezione degli esseri umani da comportamenti malevoli dell'IA salvaguardando la loro *privacy* (restano infatti validi i requisiti del GDPR[18]) e prevenendo qualsiasi forma di danno, sia al singolo utente umano che alla più vasta società e all'ambiente circostante[19].

[17] J. Mökander *et al.*, "Conformity Assessments and Post-Market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation", in *Minds and Machines*, 32 (2022), n. 2, pp. 241-268, https://doi.org/10.1007/s11023-021-09577-4; M. Calvano *et al.*, "Building Symbiotic AI: Reviewing the AI Act for a Human-Centred, Principle-Based Framework", in *Minds and Machines*, (2025), https://doi.org/10.48550/arXiv.2501.08046.

[18] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 32016R0679 (2016), http://data.europa.eu/eli/reg/2016/679/oj.

[19] M. Calvano *et al.*, "Building Symbiotic AI", cit.

Per garantire la creazione di sistemi di IA simbiotica che siano conformi con l'AI Act, è necessario prendere in considerazione, oltre ai principi, anche tre proprietà fondamentali: Sostenibilità, Affidabilità e Robustezza. È importante specificare che queste proprietà non sono una condizione sufficiente per la creazione di sistemi di IA simbiotica, ma emergono come componenti fondamentali da considerare durante il processo di creazione assieme ai principi. Difatti, la fig. 2 mostra la connessione tra i principi e le proprietà.

Le proprietà identificate sono definite nel seguente modo. La *Sostenibilità* riguarda la necessità di sviluppare sistemi di IA conformi con la legge e con i requisiti di sostenibilità, in grado di essere affidabili e resistenti anche di fronte a cambiamenti ambientali[20]. L'*Affidabilità* si riferisce alla capacità dei sistemi di IA di consentire agli esseri umani di supervisionare e/o controllare in modo adeguato le loro prestazioni, mantenendo un livello appropriato di automazione, adottando comportamenti equi e rispettando i diritti umani[21]. La *Robustezza*, infine, indica la capacità dei sistemi di operare in modo affidabile ed efficace in diverse condizioni, comprese quelle inaspettate o particolarmente complesse, senza compromettere le proprie funzioni o la sicurezza[22].

[20] U. Pagallo *et al*., "The Environmental Challenges of AI in EU Law: Lessons Learned from the Artificial Intelligence Act (AIA) with Its Drawbacks", in *Transforming Government: People, Process and Policy*, 16 (2022), n. 3, pp. 359-376, https://doi.org/10.1108/TG-07-2021-0121.

[21] G. Stettinger *et al.*, "Trustworthiness Assurance Assessment for High-Risk AI-Based Systems", *IEEE Access*, 12 (2024), pp. 22718-45, https://doi.org/10.1109/ACCESS.2024.3364387; European Commission: Directorate-General for Communications Networks, Content and Technology, *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self Assessment.* (Publications Office, 2020), https://doi.org/10.2759/002360.

[22] J. Tallberg *et al.*, "AI Regulation in the European Union: Examining Non-State Actor Preferences", *Business and Politics* 26, fasc. 2 (2024): 218–39, https://doi.org/10.1017/bap.2023.36; Johann Laux et al., "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk", *Regulation & Governance*, 18 (2024), n 1, pp. 3-32, https://doi.org/10.1111/rego.12512.

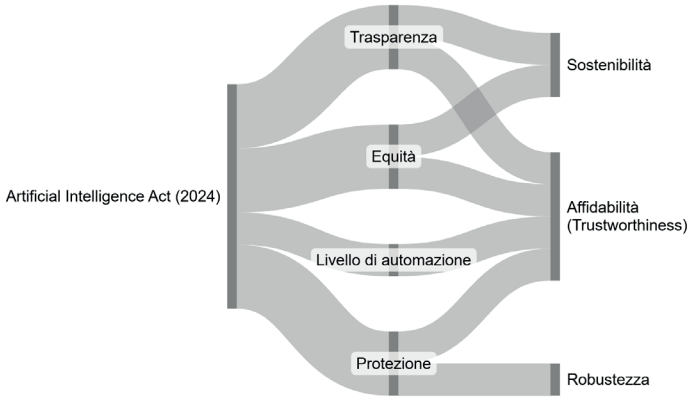

Figura 2. Corrispondenza tra i principi e le proprietà che caratterizzano i sistemi di IA simbiotica conformi con l'AI Act

4. Indicazioni pratiche

L'obbiettivo di creare sistemi di IA che riescano a stabilire una relazione *simbiotica* con gli esseri umani può risultare una grande sfida per progettisti e sviluppatori, i quali devono porre l'attenzione su molteplici aspetti restando fedeli all'approccio centrato sulle persone (dall'inglese, *human-centered*). L'IA simbiotica, infatti, necessita di un approccio multidisciplinare per essere costruita, coinvolgendo l'interazione persona-macchina, l'ingegneria del *software*, l'intelligenza artificiale, la legge e l'etica[23]. Tale approccio ha guidato la costruzione del *framework FrameSAI* (dall'inglese: *Framework to create Symbiotic Artificial Intelligence*)[24] che ha l'obiettivo di stabilire un approccio sistematico e con diversi livelli di granularità per creare sistemi di IA simbiotici. Questo *framework*, illustrato nella fig. 3, pre-

[23] G. Desolda *et al.*, "From Human-Centered to Symbiotic Artificial Intelligence", cit.

[24] M. Calvano *et al.*, "Guiding Designers and Developers in Creating Symbiotic AI Systems", in *Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter*, ACM, 6 ottobre 2025, pp. 1-5, https://doi.org/10.1145/3750069.3750095.

senta una struttura a tre livelli: *Principi*, *Linee Guida* e *Criteri di Successo*. Come suggeriscono questi nomi, i tre livelli sono stati realizzati con l'obiettivo di raggiungere diversi livelli di astrazione, partendo da un'alta generalità dei principi per finire alla pragmaticità e concretezza dei criteri di successo.

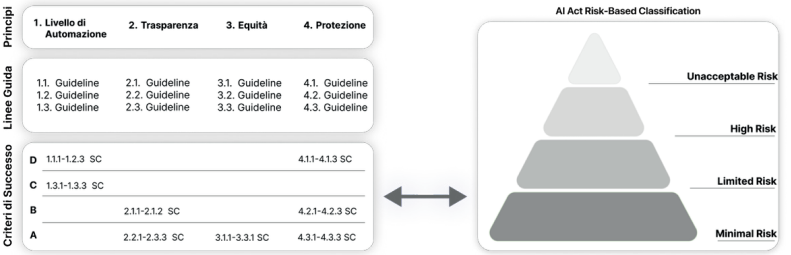

Figura 3. Struttura di FrameSAI con i tre livelli e mappatura con l'AI Act

Nello specifico, questa sezione presenta le linee guida e i criteri di successo che consentono di creare sistemi di IA simbiotica conformi all'AI Act.

Le linee guida definiscono le caratteristiche che il sistema di IA deve possedere e il comportamento che deve manifestare per favorire una collaborazione efficace con gli esseri umani. In particolare, sono state definite 12 linee guida, ciascuna associata a uno o più principi o linee guida e presenta la seguente struttura: *<Soggetto> deve <obbligo> <azione> [<motivazione>] (<Principio/i>), dove*:

- *Soggetto* può essere un modello di IA, un sistema di IA o un umano;
- *Obbligo* consiste nel requisito o la caratteristica da soddisfare;
- *Azione* riguarda il comportamento che il soggetto deve manifestare;
- *Motivazione* è un parametro opzionale che arricchisce la linea guida, fornendo ulteriori dettagli sui motivi per i quali l'azione o l'obbligo devono essere soddisfatti da parte del soggetto;
- *Principio/i* riporta il singolo o più principi da cui la linea guida deriva.

I criteri di successo sono istruzioni verificabili che derivano da una specifica linea guida e consentono di comprendere se il sistema è conforme a essa. Nello specifico, i criteri di successo sono enunciati con un basso livello di generalizzabilità rispetto alle linee guida, dettagliando le istruzioni in base alla tecnologia, al contesto d'uso e al livello di rischio del sistema. Essi sono stati definiti con la seguente struttura: <*Soggetto*> deve <*obbligo*> <*azione*>, dove:

- *Soggetto* può essere un modello di IA, un sistema di IA, un umano o l'*output* generato dall'IA;
- *Obbligo* consiste nel requisito o la caratteristica da soddisfare;
- *Azione* riguarda il comportamento che il soggetto deve manifestare;

La struttura dei criteri di successo è stata definita a partire da quella delle linee guida, ma con un livello di specificità maggiore in cui è importante comprendere anche quali sono le caratteristiche dell'*output* generato dall'IA.

4.1. Esempi e applicazione delle linee guida

Alcuni esempi di linee guida sono:

(G1) "*L'essere umano deve poter verificare, monitorare e supervisionare il comportamento del sistema di IA, tenendo conto di come esso funziona e delle decisioni che prende*" (Livello di Automazione, Trasparenza).

(G4) "*L'essere umano deve essere guidato verso comportamenti etici durante la riconfigurazione del modello di IA, al fine di evitare danni a entrambe le parti*" (Equità, Livello di Automazione).

Nel caso di (G1), la linea guida esplicita che il sistema di IA deve dare la possibilità all'essere umano di poter comprendere il processo decisionale del modello attraverso meccanismi che vedono coinvolti sia i paradigmi dell'Interazione Persona-Macchina che i tecnicismi appartenenti all'Intelligenza Artificiale[25]. Per ri-

[25] A. Esposito *et al.*, "The Fine Line between Automation and Augmentation in Website Usability Evaluation", in *Scientific Reports*, 14 (2024), n. 10129, https://doi.org/10.1038/s41598-024-59616-0; R.A. Calvo *et al.*, "Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry",

trovare l'applicazione concreta di questa indicazione, si riporta un esempio nell'ambito medico: dominio in cui questi sistemi sono spesso progettati con l'obbiettivo di classificare dati per supportare i medici nel processo diagnostico. Si fornisce il seguente scenario: un modello viene utilizzato per classificare delle risonanze magnetiche del cervello per determinare se il paziente è affetto da un tumore generando una spiegazione esaustiva e comprensibile che l'utente possa usare come strumento per valutare la qualità della risposta dell'IA ed effettuare la decisione. Nello specifico, la spiegazione indica quali caratteristiche del cervello hanno influenzato l'*output* dell'IA per poi generare la diagnosi. Tale aspetto è molto importante poiché se la spiegazione è presentata in un formato non familiare al medico o che non può comprendere, nonostante sia matematicamente e concettualmente corretta, il risultato finale potrebbe essere fortemente influenzato negativamente. Al contrario, se il sistema è stato creato coinvolgendo l'utente finale e, dunque, conoscendo le modalità più adatte e consone per la presentazione della spiegazione, il processo di interazione risulterà più efficace[26]. In altre parole, costruire un sistema di IA senza conoscere le caratteristiche dell'utente finale, potrebbe portare ad effettuare decisioni sbagliate con potenziali gravi conseguenze, specialmente in domini ad alto rischio.

Mentre (G1) sottolinea caratteristiche più architetturali, (G4) pone l'attenzione sulla relazione inevitabile tra i meccanismi d'interazione dei sistemi e l'eticità degli stessi. Infatti, con questa linea guida si indica che il sistema deve guidare l'utente nel processo di riconfigurazione (ovverosia richieste di modifiche al funzionamento del sistema) affinché questi possa raggiungere i suoi obbiettivi in maniera corretta e senza causare danni a se stes-

in C. Burr, L. Floridi (eds.), *Ethics of Digital Well-Being*, vol. 140, Philosophical Studies Series, Springer International Publishing, Cham, 2020, https://doi.org/10.1007/978-3-030-50585-1_2; B. Shneiderman, *Human-Centered AI*, cit.; U. Farooq *et al.*, "Human Computer Integration *versus* Powerful Tools", in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, 6 maggio 2017, pp. 1277-1282, https://doi.org/10.1145/3027063.3051137.

[26] G. Desolda *et al.*, "A Human–AI Interaction Paradigm and Its Application to Rhinocytology", in *Artificial Intelligence in Medicine*, 155 (2024), n. 102933, https://doi.org/10.1016/j.artmed.2024.102933.

so o al sistema che sta utilizzando[27]. Per coerenza con l'esempio precedente, si ripropone lo stesso scenario dell'IA applicata alla medicina per illustrare un esempio di corretta applicazione della linea guida. Se un medico non è d'accordo con l'*output* del sistema, nonostante abbia analizzato la spiegazione e correttamente interpretato il processo di ragionamento del modello, deve avere la possibilità di cambiare il risultato fornendo il suo riscontro. In questo caso, il medico dovrà essere guidato dal sistema per poter segnalare l'errore riscontrato al sistema e, eventualmente, accompagnare questo *feedback* con una motivazione. Se il riscontro del medico viene utilizzato per migliorare il modello, questi deve esserne consapevole prima di completarne il processo, in modo tale che sappia che le sue azioni avranno conseguenze sul sistema che sta usando e su eventuali diagnosi future[28].

4.2. Esempi e applicazione dei criteri di successo

In maniera analoga alle linee guida, due esempi di criteri di successo per (G1) sono:

(SC1) Il sistema mostra la spiegazione corrispondente al processo decisionale del modello di IA.

(SC2) Il sistema mostra una spiegazione che può essere sia testuale che visiva, in contesti ad alto rischio in cui l'utente deve avere una visione complessiva e dettagliata del processo decisionale del modello di IA.

Considerando i due criteri di successo, è possibile constatare un livello di granularità maggiore rispetto alle linee guida rappresentato dall'enfasi posta sugli elementi specifici da conside-

[27] G. Pavlidis, "Unlocking the Black Box: Analysing the EU Artificial Intelligence Act's Framework for Explainability in AI", in *Law, Innovation and Technology* 16 (2024), n. 1, pp. 293-308, https://doi.org/10.1080/17579961.2024.2313795; A. Năstasă *et al.*, "Artificial Intelligence: Friend or Foe? Experts' Concerns on European AI Act", in *Economic Computation and Economic Cybernetics Studies and Research*, 57 (2023), n. 3, pp. 5-22, https://doi.org/10.24818/18423264/57.3.23.01.

[28] A. Curci, A. Esposito, "An AI Act-Driven Design for Detecting Brain Tumors through Reconfiguration", in D. Dell'Anna et al. (eds.), *Proceedings of the Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence*, vol. 4074, CEUR Workshop Proceedings, 2025, https://ceur-ws.org/Vol-4074/short9-6.pdf.

rare (ad esempio, la spiegazione) e dalle corrispondenti caratteristiche (ad esempio, visuale o testuale).

Considerando il primo criterio di successo (SC1), si fa riferimento alla possibilità di visualizzare la spiegazione del processo decisionale consentendo agli utenti di comprendere come e perché il modello abbia prodotto un determinato risultato. Inoltre, il secondo criterio di successo (SC2) fornisce un livello di dettaglio maggiore riguardo alle caratteristiche della spiegazione affinché gli utenti possano comprendere a pieno la motivazione per cui è stato fornito loro un determinato *output*.

Ad esempio, nel dominio medico e nello stesso scenario descritto per la linea guida (G1), tali criteri di successo possono considerarsi soddisfatti se il sistema IA fornisce al medico una spiegazione comprensibile e rappresentativa del processo decisionale. In particolare, la spiegazione deve indicare quali caratteristiche del cervello hanno portato alla diagnosi, la quale deve essere supportata da una rappresentazione visiva, come ad esempio una risonanza magnetica del cervello, che evidenzi le aree rilevanti considerate dal modello di IA durante il processo decisionale.

4.3. Processo iterativo di definizione

Come menzionato nei paragrafi precedenti, le linee guida sono state definite in modo iterativo ed empirico. Una prima fase ha riguardato la decomposizione dei principi in istruzioni atomiche con il fine di costruire le radici delle linee guida. A partire da questa versione preliminare, sono stati analizzati i risultati di alcuni esperimenti eseguiti sull'applicazione dell'IA simbiotica in scenari reali. Nello specifico, è stato considerato un sistema applicato nel campo medico per supportare rinocitologi nella classificazione di cellule attraverso l'utilizzo dell'IA. I risultati ottenuti sono stati utilizzati per arricchire le linee guida con il fine di renderle più complete ed esplicative, arrivando ad una prima versione stabile. Successivamente, è stato necessario valutarle in termini di comprensibilità e rilevanza rispetto ai pilastri dell'IA simbiotica. Pertanto, è stato eseguito uno studio pilota con due ricercatori in Interazione Uomo-Macchina ed Intelligenza Artificiale, chiedendo loro di valutare ogni linea

guida assegnando un punteggio nell'intervallo 1-5, dove 1 rappresenta il giudizio più negativo e 5 quello più positivo. Inoltre, per avere una visione più ampia riguardo ai punteggi inseriti, è stato chiesto ai ricercatori di motivare il punteggio assegnato e di esplicitare il cambiamento richiesto. Chiedendo ai ricercatori di commentare e spiegare la loro valutazione con dei commenti, è stato possibile migliorare le linee guida, generando una nuova versione. Questo studio ha generato risultati promettenti, con riscontri positivi rispetto alla comprensibilità e alla struttura delle linee guida. Di seguito, si riporta un esempio di cambiamento per (G5) che originariamente indicava "*L'essere umano non deve essere manipolato dal sistema di IA attraverso comportamenti persuasivi o ingannevoli*"; i ricercatori hanno riscontrato ambiguità e poca chiarezza rispetto ai casi in cui l'umano abbia intenzioni negative e non etiche nel momento in cui interagisce con un sistema di IA. Pertanto, è stata modificata come segue: "*L'essere umano non deve essere manipolato dal sistema di IA attraverso comportamenti persuasivi, a meno che non debba essere dissuaso da intenzioni non etiche*"[29].

Quest'ultima è in fase ulteriore di valutazione, rispetto agli stessi criteri, attraverso uno studio più estensivo e approfondito; al momento della scrittura di questo capitolo, i dati raccolti sono in fase di analisi per la delineazione di un'ultima e definitiva versione delle linee guida con il fine di dare un risvolto più pratico a questa ricerca. Nello specifico, si definiranno indicazioni concrete per la progettazione – attraverso dei *Design Pattern* – e per la valutazione – con i *Criteri di Successo* – sottolineando l'approccio *top-down* e mappando le "foglie" (nello specifico, i criteri di successo) del *framework FrameSAI* alla sua "radice": una relazione simbiotica tra umani e tecnologia.

5. Conclusioni

L'evoluzione dell'IA e la sua pervasività in numerosi ambiti della società richiedono la creazione di sistemi conformi alla legge e in grado di tutelare i diritti fondamentali degli esseri

[29] M. Calvano *et al.*, "Guiding Designers and Developers in Creating Symbiotic AI Systems", cit.

umani e l'ambiente che li circonda. In questo contesto, l'AI Act rappresenta un passo fondamentale per garantire che lo sviluppo e l'utilizzo dei sistemi di IA avvengano in modo etico e responsabile senza recare danno. Per raggiungere questi obiettivi, è necessario impiegare un approccio *centrato sulle persone* per creare sistemi di IA focalizzati sulle caratteristiche e necessità degli utenti finali, considerandoli non solo come utenti passivi ma come partecipanti attivi ai processi decisionali dell'IA.

In questo contesto, l'IA simbiotica, una specializzazione dell'IA *centrata sulle persone*, promuove un rapporto di collaborazione e continuo miglioramento reciproco tra le due parti. *FrameSAI* costituisce un contributo significativo per la progettazione di sistemi di IA simbiotica conformi all'AI Act, fornendo uno strumento capace di integrare principi, linee guida e criteri di successo che, attraverso diversi livelli di astrazione, possono guidare i progettisti nella creazione di sistemi intelligenti simbiotici.

Valentina Colcelli

IA e principio di equità: base dati, collaborazione interdisciplinare e nuova generazione di giuristi

Abstract: **AI and the Principle of Fairness: Databases, Interdisciplinary Collaboration and a New Generation of Jurists.** The fundamental rights and the rule of law have to be incorporated from the design stage onwards into the development of Artificial Intelligence (AI). This requirement calls for lawyers, judges and jurists to conduct an in-depth analysis of the application of traditional legal principles, such as the principle of non-discrimination, to algorithms. Algorithmic discrimination differs from traditional forms of discrimination in its abstract, non-intuitive and often hidden nature, resulting from biases in the data used for training, testing and validation. This complexity compromises the effectiveness of legal remedies historically grounded in intuition. Data governance – including origin, collection, verification and purpose of use – plays a central role, particularly in AI systems intended for sensitive areas such as medicine and the judiciary. To mitigate subjectivity in database construction, a multidisciplinary approach that promotes interaction between the technical community and the social sciences is essential, with lawyers playing a primary role. European Union law must guide the definition of fairness requirements, adopting a risk-based approach and providing statistical metrics consistent with the European Court of Justice standards for the assessment of discrimination. It is necessary to prevent developers from independently determining local regulatory thresholds and to ensure external regulatory and judicial input. Finally, the abstract nature of applying the principles of direct and indirect discrimination requires training a new generation of lawyers through interdisciplinary courses to address the challenges posed by fairness in AI.

Keywords: Fairness, Artificial Intelligence Systems, Database, Interdisciplinary Collaboration, Jurists.

Sommario: 1. Introduzione: la non intuitività del principio di non discriminazione applicato ai sistemi di Intelligenza artificiale – 2. Le sfide poste dagli algoritmi opachi e dalla *governance* dei dati: origine, raccolta, verifica e uso previsto della base dati – 3. Caratterizzazione dell'equità degli algoritmi – 4. La dimensione sostanziale dell'equità secondo le Linee Guida Etiche per un'IA affidabile – 5. Un approccio multidisciplinare per ridurre la soggettività nella formazione delle basi dati – 6. Le dimensioni procedurali dell'equità e la comprensione della componente umana relativa allo sviluppo e applicazione del sistema di IA – 7. Qualche elemento conclusivo.

1. Introduzione: la non intuitività del principio di non discriminazione applicato ai sistemi di Intelligenza artificiale

Nell'aprile 2021, la Commissione Europea (CE) ha iniziato una prima riflessione su come regolare lo sviluppo e uso dell'Intelligenza Artificiale (IA) in Europa, conclusasi momentaneamente con Regolamento (UE) 2024/1689, il così detto AI Act. In realtà, l'Unione Europa (UE) ha pubblicato nel tempo un gran numero di testi, seppur non vincolanti, sull'IA. La questione dell'uso degli algoritmi e dell'intelligenza artificiale, in generale, è oggetto di molti documenti quali la Risoluzione del Parlamento europeo del 16 febbraio 2017, con le Raccomandazioni alla Commissione sulle "norme di diritto civile in materia di robotica"; la Comunicazione della Commissione Europea del 25 aprile 2018 sull' "Intelligenza artificiale per l'Europa" (COM(2018) 237); la Comunicazione della Commissione "Costruire la fiducia nell'intelligenza artificiale incentrata sull'uomo" dell'8 aprile 2019, (COM(2019) 168) e, più recentemente, il Libro bianco sull'intelligenza artificiale del 19 febbraio 2020 (COM(2020) 65). Inoltre, nel quadro del Consiglio d'Europa, il Comitato di esperti sugli intermediari Internet ha pubblicato lo studio "Algoritmi e diritti umani" nel marzo 2018 e, poco dopo, il 3 dicembre 2018, la Commissione europea per l'efficienza dei sistemi giudiziari (CEPEJ) ha approvato la "Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari". Il percorso piuttosto lungo realizzato dall'Unione Europea (UE) attraverso i vari documenti di *soft law* tra cui quelli citati, ha mirato a definire un approccio all'IA basato sui diritti fondamentali.

Questo approccio, oggi, trova la sua concretizzazione nell'AI Act, per quale i sistemi di IA sviluppati ed utilizzati in diverse applicazioni, debbano essere classificati in base al rischio che comportano per gli utenti. I diversi livelli di rischio implicano una regolamentazione più o meno stringente.

L'idea principale alla base dell'approccio UE all'IA è che gli algoritmi di intelligenza artificiale dovrebbero incorporare i principi di democrazia, Stato di diritto e dei diritti fondamentali fin dalla fase di progettazione.

Il Regolamento e l'insieme dei documenti che lo precedono, sottolineano l'idea di creare una nuova cultura dell'IA attraverso una progettazione tecnologica ed informatica che, dovendo incorporare il rispetto dei diritti fondamentali e dello Stato di diritto *by design,* richieda in primo luogo a giuristi, giudici e avvocati, un'analisi approfondita di come applicare principi giuridici tradizionali, come ad esempio il principio di non discriminazione, allo sviluppo di un algoritmo.

Possiamo cercare di condurre questa riflessione usando proprio come paradigma il principio equità.

Sappiamo che i principi alla base dello sviluppo dell'IA nel rispetto dei diritti fondamentali implicano che questa debba essere affidabile ed attendibile. Tre componenti devono, pertanto, essere sempre presenti durante l'intero ciclo di vita del sistema di IA: legalità, eticità e solidità.

La linea guida dell'UE sull'affidabilità dell'IA (Orientamenti etici per un'IA affidabile dell'HLEGAI)[1] realizzata da un gruppo di esperti per la Commissione europea e ora richiamate dal regolamento sulla IA, definisce l'affidabilità attraverso quattro questioni etiche:

1. il rispetto dell'autonomia umana;
2. la prevenzione dei danni;
3. la spiegabilità, e
4. l'equità.

Parlandone in termini generali e per forza di cose anche esemplificati, possiamo sin d'ora dire che l'equità e il principio di non discriminazione sono realizzabili attraverso la corretta selezione dei dati utilizzati per l'addestramento e il collaudo degli algoritmi di IA.

Un esempio molto noto e scolastico, ma che può essere utile a spiegare velocemente il punto, è quello relativo all'algoritmo di IA utilizzato da una banca americana per l'attribuzione di mutui che ha prodotto una discriminazione verso richiedenti prestiti ipotecari quando questi erano donne e di colore. Praticamente l'algoritmo non ne concedeva a quest'ultimi. La violazione del

[1] High-Level Expert Group on AI (2019), "Ethics guidelines for trustworthy AI", in Shaping Europe's digital future – Ethics guidelines for trustworthy AI. Recuperato da https://data.europa.eu/doi/10.2759/346720.

principio di discriminazione si sarebbe concretizzato nel fatto che la base dati usata per il suo sviluppo conteneva uno sbilanciamento a favore di informazioni relative a uomini e di pelle bianca, anche solo perché questi erano i soggetti dei quali si possedevano più informazioni numeriche utili allo sviluppo del sistema in quanto, comunemente, aventi più accesso al credito rispetto a donne e di colore.

Ed invero "l'affidamento a dati storici e a pregiudizi in essi impliciti, difficili da individuare senza strumenti dedicati di interpretabilità, è una delle principali sfide che devono affrontare i ricercatori e gli sviluppatori di sistemi che utilizzano l'IA"[2].

Nella complessa vita reale dello sviluppo di un sistema di IA, però la violazione del principio di non discriminazione nell'applicazione dell'IA non è così intuitiva come quella sopra riportata.

Molti autori[3] ci ricordano che gli esseri umani discriminano a causa di atteggiamenti negativi (ad esempio stereotipi, pregiudizi) e pregiudizi involontari (ad esempio pratiche organizzative o stereotipi interiorizzati), che possono fungere da segnale per le vittime che si è verificata una discriminazione. Questi meccanismi di segnalazione non esistono nei sistemi algoritmici.

Questo significa che il crescente utilizzo di algoritmi sconvolge i tradizionali rimedi giuridici e le procedure di individuazione, indagine, prevenzione e correzione della discriminazione, che si sono basati storicamente anche e prevalentemente sull'intuizione[4].

La discriminazione nei sistemi di IA, rispetto alle forme tradizionali di discriminazione, è più astratta e meno intuitiva, sottile, intangibile e difficile da individuare[5], perché spesso è intrinseca e nascosta nei *bias* della base dati utilizzata per il training, testing e validazione dell'algoritmo.

[2] A. Barucci, V. Colcelli, S. De Masi *et al.*, "Ethical and Regulatory Frameworks for Artificial Intelligence in Clinical Research: A European Perspective Post-AI Act for Ethics Committees and Researchers", in *European Cardiology Review*, (in corso di pubblicazione).

[3] S. Wachter, B. Mittelstadt, C. Russell, "Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI", in *Computer Law & Security Review*, 41 (2021), pp. 105567 ss.

[4] S. Wachter, B. Mittelstadt, C. Russell, *op. cit.*

[5] *Ibidem.*

2. Le sfide poste dagli algoritmi opachi e dalla *governance* dei dati: origine, raccolta, verifica e uso previsto della base dati

Insieme alla complessità tecnica del modello di IA caratterizzato da troppe variabili e parametri per essere interpretato in modo significativo dagli esseri umani, alla mancanza di trasparenza da parte degli sviluppatori riguardo al funzionamento, alle scelte di progettazione degli algoritmi, in tema *governance* dei dati, recente letteratura scientifica[6] riconosce nell'uso dei c.d. dati storici e nei pregiudizi in essi impliciti, uno tra i fattori che maggiormente contribuiscono a rendere i sistemi IA opachi.

Una delle questioni centrali allora è quella, appunto, della *governance* dei dati, in particolare la loro origine, la raccolta, la verifica e l'uso previsto.

Questi aspetti sono rilevanti specialmente per i sistemi di IA sviluppati in ambito medico, ma l'origine, la raccolta, il test e l'uso previsto dei dati riguarda ogni sistema IA, e sicuramente anche quelli che dovranno essere selezionati per addestrare, testare e validare un algoritmo che dovrà essere impiegato ad esempio in alcune fasi del giudizio penale.

Ad esempio, l'origine, la raccolta, la verifica e l'uso previsto è particolarmente importante rispetto ai dati sanitari perché questi possono provenire da una moltitudine di fonti: cartelle cliniche elettroniche (EHR), sistemi di imaging medico (come TC, RM o ecografia), esami di laboratorio, analisi genomiche e proteomiche, patologia digitale, dati provenienti da sensori indossabili e persino risultati riferiti dai pazienti raccolti tramite app mobili o questionari. Questi set di dati sono spesso eterogenei e comprendono campi strutturati (ad esempio, segni vitali, prescrizioni di farmaci), testi non strutturati (ad esempio, note cliniche) e dati ad alta dimensionalità (ad esempio, profili di imaging o multi-omici)[7].

[6] A. Fernández, "Opacity, Machine Learning and Explainable AI", in F. Lara, J. Deckers (eds.), *Ethics of Artificial Intelligence. The International Library of Ethics, Law and Technology*, Springer, Cham, 2023, pp. 39-58.

[7] A. Barucci, V. Colcelli, A. Gottard, "Imaging Biobank: What Are the Areas of the GDPR Bearing on an Image Biobank?", in V. Colcelli, R. Cippitani, C. Brochhausen-Delius, R. Arnold (eds.), *GDPR Requirements for Biobanking Activities Across Europe*, Springer, Cham, 2023, pp. 241-251.

Nei sistemi di IA applicati al processo penale la varietà delle fonti dovrebbe essere inferiore alla precedente descritta, in ogni caso i metodi di raccolta dei dati variano a seconda del contesto: estrazione retrospettiva da database esistenti e/o acquisizione prospettica, ad esempio. Ciascun metodo pone sfide distinte in termini di qualità, completezza, standardizzazione e distorsione dei dati.

Inoltre, il processo di annotazione dei dati, fondamentale per gli approcci di apprendimento automatico, introduce ulteriore complessità, in particolare quando le etichette dipendono da giudizi soggettivi di chi li realizza o sono definite in modo incoerente tra le diverse istituzioni coinvolte nel processo di sviluppo e validazione del sistema.

Altra questione fondamentale che i creatori di sistemi di IA dovrebbero affrontare nella predisposizione dei protocolli di sviluppo è la necessità di essere sempre molto chiari sullo scopo per cui i dati saranno utilizzati.

Per cui, usando sempre il campo medico come riferimento data la sua complessità, se l'obiettivo è quello di addestrare un nuovo modello predittivo, convalidare e testare un algoritmo esistente su una nuova popolazione, identificare sottotipi di malattie precedentemente sconosciuti o supportare il processo decisionale clinico, allora la quantità, le caratteristiche e la qualità dei dati devono essere adeguatamente allineate all'uso previsto dei sistemi di IA, che deve essere sempre esplicitamente indicato. Ad esempio, come in qualsiasi altro approccio statistico, l'uso dei dati delle cartelle cliniche elettroniche per prevedere gli esiti dei pazienti può risentire di variabili confondenti o di informazioni mancanti non raccolte originariamente a fini di ricerca.

Gli sviluppatori, inoltre, devono articolare i dati in base all'obiettivo del sistema, chiarendo su cosa esso si concentra. Questo allineamento non è solo una necessità metodologica, ma anche un imperativo etico: esso garantisce che i dati siano utilizzati in modo responsabile e che i contributi dei titolari dei dati che partecipano ad uno studio siano rispettati e che i rischi associati all'uso secondario dei dati (ad esempio, violazioni della *privacy*, uso improprio di informazioni sensibili) siano ridotti al minimo attraverso adeguati quadri di *governance by design* degli stessi.

3. Caratterizzazione dell'equità degli algoritmi

Equità, universalità, tracciabilità, usabilità, robustezza e spiegabilità sono i principi guida perché le tecnologie di IA siano affidabili. Universalità, tracciabilità, usabilità, robustezza e spiegabilità non saranno trattati in questa sede, sebbene qualsiasi elemento derivante dall'analisi dell'equità abbia un impatto diretto sugli altri principi, in particolare sulla spiegabilità.

Riassumiamo, invece, qui, le principali questioni relative agli aspetti che caratterizzano l'equità degli algoritmi concentrandoci sugli elementi più rilevanti dal punto di vista giuridico.

Prima di tutto, come sopra riportato, i protocolli di raccolta dei dati devono garantire un'adeguata rappresentanza della popolazione target (ad esempio, età, sesso, razza, etnia) nei set di dati. Ciò promuove risultati generalizzabili, mitiga i pregiudizi e aiuta a valutare le prestazioni del modello in condizioni diverse, identificando potenziali limitazioni e garantendo un uso sicuro ed efficace in contesti reali.

Nello stesso tempo, gli algoritmi di IA dovrebbero essere imparziali e mantenere le stesse prestazioni quando applicati a individui in situazioni simili (equità individuale) o a gruppi diversi di individui, compresi i gruppi sottorappresentati (equità di gruppo) (equità sostanziale).

Inoltre, deve essere sempre ammessa la possibilità di contestare e di cercare un rimedio efficace contro le decisioni suggerite dai sistemi di IA e prese dagli esseri umani che li gestiscono e applicano (equità procedurale).

Possiamo considerare gli aspetti sopra menzionati dal punto di vista giuridico.

Anche se la legislazione esistente dovrà essere rivista affinché possa adattarsi alle nuove opportunità e sfide poste dall'IA, compresa la normativa dell'UE in materia di non discriminazione, possiamo dire che l'Unione europea e i quadri giuridici nazionali hanno già un sistema normativo embrionale che consente lo sviluppo, l'implementazione e l'uso dei sistemi di IA basato sui diritti fondamentali: ci riferiamo alla legislazione in materia di protezione dei dati, protezione dei consumatori, le direttive in tema di non discriminazione e parità di genere, la direttiva sull'applicazione della legge (direttiva (UE) 2016/680), la diret-

tiva sulla sicurezza dei prodotti che sono già una base giuridica applicabili alla progettazione, allo sviluppo e all'uso di determinati sistemi di IA.

Questo richiamo serve a ricordare che ogni sistema di IA – anche a riguardo all'applicazione del principio di non discriminazione – deve essere valutato alla luce della sua attuale validità giuridica e secondo il diritto dell'UE.

Tuttavia, proprio la protezione giuridica offerta dalla normativa antidiscriminatoria sopra richiamata viene messa in discussione quando è l'IA – e non gli esseri umani – a discriminare.

Per cercare affrontare la discriminazione astratta e non intuitiva nei sistemi di IA, alcuni studiosi[8] da tempo ci suggeriscono di far riferimento ai criteri e i requisiti probatori utilizzati dalla Corte di giustizia dell'Unione europea (CGUE) nei giudizi di *prima facie* relativi a una presunta discriminazione diretta o indiretta.

Sebbene la giurisprudenza della Corte di giustizia europea offra pochissimi esempi chiari dei requisiti o delle soglie statistiche che possano qualificarsi come standard in grado di definire i concetti chiave per definire i gruppi alla base della discriminazione, si può ricorrere ai criteri adottati dalla Corte stessa per stabilire che cosa si intenda per discriminazione.

Ciò è necessario per evitare la situazione per cui gli sviluppatori del sistema IA stabiliscano da soli, in modo soggettivo e senza un contributo di competenze giuridiche esterne[9] le soglie su cui misurare la discriminazione nella formazione della base dati.

4. La dimensione sostanziale dell'equità secondo le Linee Guida Etiche per un'IA affidabile

Le Linee guida etiche per un'IA affidabile[10] affrontano l'equità in due dimensioni, sostanziale e procedurale.

La dimensione sostanziale dell'equità implica un impegno del sistema a:

[8] S. Wachter, B. Mittelstadt, C. Russell, *op. cit.*

[9] *Ibidem.*

[10] High-Level Expert Group on AI, "Ethics guidelines for trustworthy AI", cit.

1) Garantire una distribuzione equa e giusta sia dei benefici che dei costi e assicurare che gli individui e i gruppi siano liberi da pregiudizi ingiusti, discriminazioni e stigmatizzazioni;

2) Promuovere pari opportunità in termini di accesso all'istruzione, ai beni, ai servizi e alla tecnologia;

3) Usare sistemi di IA in modo da non portare mai le persone a essere ingannate o indebitamente limitate nella loro libertà di scelta;

4) Rispettare da parte dei professionisti dell'IA il principio di proporzionalità tra mezzi e fini.

5) Valutare attentamente come bilanciare interessi e obiettivi contrastanti.

La dimensione sostanziale dell'equità è strettamente correlata alla parità di trattamento ai sensi dell'articolo 19 del Trattato sul Funzionamento dell'Unione europea (TFUE), dell'articolo 21 della Carta dei diritti fondamentali dell'Unione europea (EU-CFR) e dell'articolo 14 della Convenzione europea dei diritti dell'uomo (CEDU), nonché della legislazione dell'UE in materia di non discriminazione.

Da un punto di vista sostanziale, il primo punto da affrontare allora è la corretta rappresentazione dei gruppi di persone i cui dati sono utilizzati nell'addestramento dei sistemi di IA.

Questo significa che, se dal punto di vista dello sviluppo del sistema l'equità può essere influenzata negativamente da pregiudizi sia quantitativi che qualitativi, da un punto di vista giuridico gli errori – che potrebbero derivare dall'addestramento degli algoritmi di IA – diventano distorti nei confronti di già gruppi sottorappresentati nella società e quindi aggravano le disparità esistenti.

La corretta rappresentazione dei gruppi di persone di cui si usano le informazioni, deve essere realizzata alla luce del sesso, dell'etnia, della geografia e della rappresentazione socioeconomica, quando e se tali elementi hanno un impatto sulle disparità magari di accesso a servizi. Le banche dati storiche (per esempio, quelle di immagini medico-diagnostiche per proseguire nella scia degli esempi precedenti) sono spesso sbilanciate in base

al sesso, all'etnia, alla geografia e alla socioeconomia[11] e sono pertanto capaci di esacerbare le disparità di accesso al servizio sanitario[12].

Per questo motivo, gli strumenti di IA possono generare errori rilevati, con conseguenze dannose per i destinatari del risultato prodotto dal sistema IA quando questi vengono testati a condizioni reali, dove queste possono differire o deviare in modo imprevisto, anche leggermente, da quelle utilizzate per l'addestramento[13].

Diversi studi dimostrano che la causa principale della distorsione algoritmica nell'IA è rappresentata dai set di dati di addestramento che spesso mancano della diversità e dell'equilibrio quantitativo e qualitativo necessari per ottenere soluzioni di IA che mantengano le stesse prestazioni tra i gruppi umani e le sottopopolazioni.

In particolare, se un algoritmo di IA viene addestrato con dati sbilanciati in termini di genere[14] socioeconomico o etnico[15] e a causa delle differenze di all'interno e tra questi gruppi, il modello risultante porterà probabilmente a previsioni distorte.

[11] G.H. Rauscher, J.A. Khan, M.L. Berbaum, E.F. Conant, "Potential missed detection with screening mammography: Does the quality of radiologist's interpretation vary by patient socioeconomic advantage/disadvantage?", in *Annals of Epidemiology*, 23 (2013), n. 4, pp. 210-214.

[12] K. Lekadir, R. Osuala, C. Gallin, N. Lazrak, K. Kushibar, G. Tsakou, S. Aussó, L. Cerdá Alberich, K. Marias, M. Tsiknakis, S. Colantonio, N. Papanikolaou, Z. Salahuddin, H.C. Woodruff, P. Lambin, L. Martí-Bonmatí, "FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging", in *arXiv*, (2024), https://doi.org/10.48550/arXiv.2109.09658.

[13] A.J. Larrazabal, N. Nieto, V. Peterson, D.H. Milone, E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis", in *Proceedings of National Academy of Sciences of United Stated of America*, 23 (2020), n. 117, pp. 12592-12594, https://doi.org/10.1073/pnas.191901211.

[14] H. Marzo-Ortega, V. Navarro-Compán, S. Akar, U. Kiltz, Z. Clark, E. Nikiphorou, "The impact of gender and sex on diagnosis, treatment outcomes and health-related quality of life in patients with axial spondyloarthritis", in *Clinical Rheumatology*, 41 (2022), pp. 3573-3581, https://doi.org/10.1007/s10067-022-06228-6.

[15] G.H. Rauscher, J.A. Khan, M.L. Berbaum, E.F. Conant, *op. cit.*

Ma gli attributi personali quali genere, età, etnia e condizioni socioeconomiche, essenziali per rispettare il principio di non discriminazione diretta o indiretta richiesto dalla norma, non sempre vengono conservati durante il processo di preparazione dei dati e di eventuale loro anonimizzazione.

Per evitare queste conseguenze legate all'addestramento dell'algoritmo di IA in relazione al sesso, alla situazione socioeconomica o all'etnia, e date le differenze all'interno e tra gruppi di riferimento, spesso si offre una soluzione che fa appello alla raccolta di dati multicentrici in grado di aumentare la diversità dei set di dati al momento della sua formazione. In realtà questa soluzione da sola potrebbe non essere sufficiente a correggere le potenziali distorsioni sulla base del principio di non discriminazione. Sebbene la raccolta di dati multicentrici sia una buona soluzione se guidata dalla selezione preventiva dei gruppi rappresentati, non è certo che solo con esso si possa raggiungere il perfetto equilibrio tra i diversi gruppi nei set di dati di addestramento e test dell'IA, in modo da garantire il rispetto del principio di equità e non discriminazione stabilito dal diritto dell'UE e dalla sua Corte nei giudizi di *prima facie*[16].

5. Un approccio multidisciplinare per ridurre la soggettività nella formazione delle basi dati

Come detto una semplice raccolta di dati multicentrica potrebbe non essere sufficiente per evitare pregiudizi nella valutazione di potenziali discriminazioni, poiché i soli elementi statistici potrebbero non essere soddisfacenti per correggere la distorsione su base discriminatoria.

Occorre sicuramente ricorrere ad un approccio multidisciplinare per ridurre la soggettività nella formazione delle basi dati. Questo permette di identificare il maggior numero possibile di fonti di pregiudizi e iniquità specifici dell'applicazione attraverso l'interazione tra comunità tecnica e scienziati sociali, in particolare giuristi.

È il diritto dell'UE che dovrebbe guidare la selezione dei requisiti generali di equità – utilizzando l'approccio basato sul

[16] S. Wachter, B. Mittelstadt, C. Russell, *op. cit.*

rischio per selezionare i requisiti – che dovrebbero essere soddisfatti in generale e in obiettivi specifici del sistema che si sta sviluppando.

Il diritto europeo in materia di non discriminazione si applica sia nel diritto primario che in quello derivato. Abbiamo già ricordato il diritto primario (paragrafo 4). Per quanto riguarda il diritto derivato, quattro direttive in materia di non discriminazione si applicano sia al settore privato che a quello pubblico: la direttiva sull'uguaglianza razziale (2000/43/CE), la direttiva sull'uguaglianza di genere (rifusione) (2006/54/CE), la direttiva sull'accesso in base al genere (2004/113/CE) e la direttiva sull'occupazione (2000/78/CE).

Le direttive sulla non discriminazione hanno stabilito uno standard minimo e forniscono un quadro generale che deve essere recepito nei sistemi giuridici degli Stati membri.

Il diritto europeo in materia di non discriminazione affronta due tipi generali di discriminazione: diretta e indiretta[17].

La discriminazione diretta si riferisce a un trattamento sfavorevole basato su qualità della persona protetta dalla norma come l'orientamento sessuale o il genere.

La discriminazione indiretta, invece, descrive una situazione in cui una disposizione, un criterio o una pratica apparentemente neutri svantaggiano in modo sproporzionato un gruppo protetto rispetto ad altre persone[18].

Per intentare una causa innanzi alla CGEU relativa ad una presunta discriminazione diretta o indiretta in materia di non discriminazione, il ricorrente deve soddisfare diversi requisiti probatori che, nel loro insieme, dimostrino la discriminazione *prima facie*.

Per fondare una richiesta di risarcimento sulla discriminazione *prima facie*, i ricorrenti devono dimostrare che:

1) si è verificato o si sta verificando un danno specifico[19];

[17] *Ibidem.*

[18] *Ibidem.*

[19] Causa C-7/12 Nadežda Riežniece contro Zemkopības ministrija e Lauku atbalsta dienests ECLI:EU:C:2013:410 [2013] ECR.

2) il danno si manifesta o si sta manifestando in modo significativo all'interno di un gruppo di persone protette[20];

3) il danno è sproporzionato rispetto ad altri che si trovano in una situazione simile[21].

Il giudice deve decidere se il gruppo sia stato o possa essere significativamente svantaggiato rispetto ad un altro gruppo in una situazione simile[22] e, quindi, se sussista una discriminazione *prima facie*[23]. Ad esempio, ai sensi delle direttive sull'uguaglianza razziale e sull'occupazione, occorre sempre effettuare un confronto tra gli effetti della misura contestata su due gruppi[24].

La definizione di uno o più gruppi svantaggiati, di gruppi di confronto legittimi e di prove di "particolare svantaggio" richiede scelte specifiche per ciascun caso che riflettano le dimensioni politiche, sociali e giuridiche locali del caso, nonché l'attenzione alle argomentazioni proposte dai possibili ricorrenti e dai presunti trasgressori.

Come suggerito dagli stessi studiosi sopra citati[25] l'approccio alla valutazione di un sistema di IA in termini di equità potrebbe essere realizzato attraverso i requisiti probatori utilizzati dalla Corte di giustizia europea, anche in presenza di una riconosciuta incompatibilità critica tra le nozioni di diritto europeo di discri-

[20] Causa C-167/97 Regina contro Segretario di Stato per l'occupazione, Ex Parte Nicole Seymour- Smith e Laura Perez ECLI:EU:C:1999:60 [1999] ECR I-00623.

[21] Causa C-104/09 Pedro Manuel Roca Álvarez contro Sesa Start España ETT SAECLI:EU:C:2010:561 [2010] ECR I-08661.

[22] Cause riunite C-399/92, C-409/92, C-425/92, C-34/93, C-50/93 e C-78/93 Stadt Lengerich contro Angelika Helmig e Waltraud Schmidt contro Deutsche Angestellten-Krankenkasse e Elke Herzog contro Arbeiter-Samariter Bund Landverband Hamburg eV e Dagmar Lange contro Bundesknappschaft Bochum e Angelika Kussfeld contro Firma Detlef Bogdol GmbH e Ursula Ludewig contro Kreis Segeberg ECLI:EU:C:1994:156 [1994] Racc. I-05727.

[23] Causa C-127/92 Dr. Pamela Mary Enderby contro Frenchay Health Authority e Secretary of State for Health ECLI:EU:C:1993:859 [1993] ECR I-05535.

[24] Causa C-167/97 Regina contro Segretario di Stato per l'occupazione, Ex Parte Nicole Seymour- Smith e Laura Perez ECLI:EU:C:1999:60 [1999] ECR I-00623.

[25] S. Wachter, B. Mittelstadt, C. Russell, *op. cit.*

minazione e il lavoro in materia di equità algoritmica e automatizzata.

Una metrica di equità di base statistica, in linea con l'unità di misura dell'equità stabilita dalla Corte di giustizia europea per valutare la potenziale discriminazione, dovrebbe essere definita dai tecnici/sviluppatori per supportare i professionisti del diritto nella collaborazione interdisciplinare.

6. Le dimensioni procedurali dell'equità e la comprensione della componente umana relativa allo sviluppo e applicazione del sistema di IA

Dall'altro lato, l'equità descritta nella sua dimensione procedurale dalle Linee guida etiche per un'IA affidabile implica la possibilità di contestare e cercare un rimedio efficace contro le decisioni prese dai sistemi di IA e dagli esseri umani che li gestiscono e applicano[26]. Questo è il motivo per cui l'entità responsabile della decisione deve essere identificabile e riferibile all'uomo e i processi decisionali devono essere spiegabili.

Solo l'identificazione corretta e completa delle persone coinvolte nel processo di creazione e utilizzo degli algoritmi e dei loro ruoli specifici consente di definire correttamente gli aspetti giuridici che devono essere rispettati in questo campo.

Le persone coinvolte in qualsiasi fase del ciclo di vita di un prodotto di IA devono, prima di tutto, comprenderlo a fondo[27]. Diversi gruppi di "persone/parti interessate" ricoprono ruoli diversi nel processo di verifica dell'applicazione dei requisiti per l'IA: le persone che sviluppano i requisiti del sistema, le persone che li utilizzano e gli utenti finali che, ad esempio nel settore sanitario, possono essere rappresentati dai pazienti che utilizzano trattamenti basati sui risultati prodotti dagli strumenti di intelligenza artificiale, ma che nel sistema giudiziario penale sono rappresentati dagli imputati che sono destinatari del provvedi-

[26] High-Level Expert Group on AI, "Ethics guidelines for trustworthy AI", cit.

[27] S. Wachter, B. Mittelstadt, C. Russell, *op. cit.*

mento realizzato con il supporto del sistema di IA, o gli stessi professionisti che operano nel settore giudiziario[28].

Per quanto riguarda i sistemi di IA, la caratterizzazione delle persone durante la raccolta e la strutturazione dei dati deve essere presa in considerazione per affrontare le questioni sopra menzionate anche in tema di equità. Questo approccio consentirà la verifica dell'intero ciclo dei sistemi di IA alla luce del quadro normativo dell'UE sull'IA, la cui corretta applicazione è requisito fondamentale per l'intero processo che riguarda il sistema[29].

Infatti, i pregiudizi che incidono sull'equità possono consistere in pregiudizi cognitivi degli sviluppatori che generano, interpretano o annotano i dati usati, pertanto comprendere la componente umana è utile per diversi aspetti, in particolare per l'applicazione del principio di equità così come la responsabilità della decisione.

Nel caso del quadro diretto a testare la discriminazione nei sistemi di IA, una recente ricerca in fase iniziale riporta prove di come gli algoritmi di intelligenza artificiale possano identificare la razza di una persona dalle sue scansioni mediche (ad esempio radiografie del torace e delle mani e mammografie), anche quando gli esperti medici umani non sarebbero in grado di comprenderlo dalle immagini senza altri riferimenti ai dati personali[30]. In questo caso particolare, i problemi di discriminazione algoritmica si potrebbero legare con quella umana, sebbene diversi per natura e carattere.

Questa combinazione potrebbe, infatti, generare una decisione che si dice essere basata sull'IA, ma che invece è generata da un trattamento diverso e iniquo – in questo caso dei pazienti –

[28] K. Lekadir, A.F. Frangi, A.R. Porras Perez *et al.*, "FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare", in *arXiv*, (2023), https://doi.org/10.48550/arXiv.2309.12325.

[29] L. Burzagli, "Users. 'Who Are the Users in the Context of Biobanks? What Are the Factors to Consider When Dealing with the Human Component?'", in V. Colcelli, R. Cippitani, C. Brochhausen-Delius, R. Arnold (eds.), *op. cit.*, pp. 403-407.

[30] I. Banerjee, A. Reddy Bhimireddy, J.L. Burns *et al.*, "Reading Race: AI Recognises Patient's Racial Identity in Medical Image", in *arXiv*, (2021), p. 10356, https://doi.org/10.48550/arXiv.2107.10356.

basato sulla loro razza, anche nei casi in cui quest'ultima non è un criterio medico rilevante per il trattamento.

Gli strumenti di IA, inoltre, influenzano le capacità decisionali e interpretative degli user esperti e meno esperti nell'ambito in cui il sistema è applicato.

È ben noto, per esempio, nella letteratura medica che l'integrazione dell'IA nella pratica clinica reale può sollevare questioni in tema di suo utilizzo da parte di medici esperti e meno esperti per l'effetto che ha sulle loro capacità decisionali. Ad esempio, per quanto riguarda l'interpretazione delle immagini mammografiche, uno studio ha riportato come il supporto automatizzato dell'IA abbia influenzato positivamente il processo decisionale dei radiologi con capacità interpretative meno avanzate. Lo stesso studio però ha descritto come il sistema di IA abbia danneggiato il processo decisionale dei radiologi con capacità avanzate di interpretazione delle immagini[31]. Ciò è legato all'eccessiva fiducia nei risultati del supporto automatizzato e alla svalutazione dell'esperienza professionale dei medici accumulata in anni di pratica medica.

Si dovrebbe infatti considerare che ciò può essere dettato da idee errate sui limiti e sui punti di forza dell'IA e non permettere all'utilizzatore del sistema di intervenire correggendolo su di un suggerimento con elementi discriminanti per l'eccessivo affidamento che la persona fa nei risultati prodotti dall'algoritmo.

7. Qualche elemento conclusivo

La discriminazione nei sistemi di IA è più astratta e meno intuitiva, sottile, intangibile e difficile da individuare rispetto alle forme tradizionali di discriminazione[32].

È ampiamente documentato in articoli scientifici internazionali che un algoritmo di IA addestrato con dati sbilanciati rispetto al sesso, alla situazione socioeconomica o all'etnia porterà probabilmente a un modello che darà luogo a previsioni distorte,

[31] G.H. Rauscher, J.A. Khan, M.L. Berbaum, E.F. Conant, *op. cit.*

[32] V. Colcelli, "EU Guideline on AI Trustworthy and 'Not Intuitive' Application of the Principle of Fairness", in R. Arnold, J. Cremades (eds.), *Rule of Law, Technology and Environment*, Springer, Cham, 2025, pp. 117-135.

date le differenze di trattamento all'interno e tra i gruppi dei destinatari dell'elaborazione.

Gli algoritmi di IA sono imparziali e mantengono le stesse prestazioni quando applicati a individui in situazioni simili (equità individuale) o a gruppi diversi di individui, compresi i gruppi sottorappresentati (equità di gruppo).

L'aspetto centrale di un approccio equo nell'uso dell'IA risiede nella scelta di un algoritmo adeguato e nella costruzione di un set di dati di addestramento in grado di rappresentare i gruppi di persone su cui avrà effetto senza nascondere in sé pregiudizi comuni basati su genere, etnia, reddito e origine geografica.

Dal punto di vista giuridico, gli errori che potrebbero derivare dall'addestramento degli algoritmi di IA diventano distorti nei confronti dei gruppi sottorappresentati o sviluppano alcune caratteristiche discriminatorie sui gruppi di persone su cui avranno effetto, esacerbando le disparità esistenti (violazione del principio di equità sostanziale, oppure esasperano pregiudizi cognitivi degli sviluppatori e/o user che generano, interpretano o annotano i dati usati (violazione del principio di equità procedurale).

Per affrontare il problema dell'equità nell'IA, "ciò che deve essere evitato è una situazione in cui gli sviluppatori e i controllori del sistema stabiliscano da soli le soglie normative per la discriminazione a livello locale e soggettivo senza un contributo normativo o giudiziario esterno"[33]; d'altra parte, poiché l'applicazione del principio tradizionale di discriminazione diretta e indiretta è astratta e non intuitiva, è necessario preparare una nuova generazione di giuristi attraverso una collaborazione interdisciplinare[34].

[33] S. Wachter, B. Mittelstadt, C. Russell, *op. cit.*

[34] *Ibidem*.

Maurizio Goretti

Geopolitica dell'intelligenza artificiale: reti, *data center* e sovranità digitale

Abstract: **The Geopolitics of Artificial Intelligence: Networks, Data Centers, and Digital Sovereignty.** Artificial intelligence is not merely a software entity but relies on a set of material infrastructures that include telecommunications networks, data centers, advanced chips, and global interconnection systems. This article examines this ecosystem, highlighting how the availability of computing power has become a crucial geopolitical factor, capable of influencing economic, technological, and social balances. Particular attention is devoted to the European dimension, caught between the challenges of digital sovereignty and dependence on U.S. and Chinese hyperscalers. The interplay between physical infrastructures, regulation, and ethical implications suggests that AI should be regarded as a new global infrastructural good, to be governed according to principles of cooperation, sustainability, and respect for fundamental rights.

Keywords: Infrastructure, Data Centers, Geopolitics, Sovereignty, Ethics.

Sommario: 1. Introduzione – 2. Le reti come sistema circolatorio dell'AI – 2.1. La banda larga come prerequisito dell'AI distribuita – 2.2. Cavi sottomarini e interconnessioni globali – 2.3. *Carrier*, *Access Provider*, *Content Provider* e IXPs: avvicinare l'AI agli utenti – 3. *Data center* e potenza di calcolo – 3.1. Dall'*hyperscale* al GPU-*cluster*: nuove architetture per l'AI – 3.2. Consumi energetici e raffreddamento: il costo fisico dell'intelligenza artificiale – 3.3. Sicurezza e resilienza delle infrastrutture critiche – 4. *Hardware* e piattaforme *cloud* – 4.1. La corsa ai *chip* – 4.2. Il ruolo degli *hyperscaler* – 4.3. Sovranità tecnologica e alternative emergenti – 5. La Geopolitica della potenza di calcolo – 5.1. Stati Uniti e Cina: la competizione per l'AI – 5.2. Europa: tra dipendenza e ambizioni – 5.3. Il resto del mondo: tra opportunità e dipendenze – 6. Impianti e sfide globali – 6.1. Sostenibilità ambientale – 6.2. *Governance* e regolamentazione – 6.3. Accesso diseguale e nuove disuguaglianze – 7. Conclusioni.

1. Introduzione

Quando si parla di intelligenza artificiale (AI), l'attenzione tende a concentrarsi sugli algoritmi, sui modelli linguistici o sulle applicazioni visibili che ormai fanno parte della nostra vita quotidiana: motori di ricerca "intelligenti", traduttori automatici,

chatbot, strumenti di analisi predittiva. Ma queste manifestazioni sono solo la superficie di un fenomeno molto più profondo. Dietro l'AI, infatti, si nasconde un immenso apparato materiale, fatto di fibre ottiche, server, cavi sottomarini, sistemi satellitari, chip avanzati, data center e centrali elettriche. In altre parole, una vera e propria infrastruttura che ne sostiene lo sviluppo e ne determina la portata.

L'AI non è "eterea" né puramente digitale: è ancorata a una fisicità che implica consumi energetici elevati, filiere tecnologiche complesse e scelte geopolitiche decisive. Ogni volta che addestriamo un modello, dietro c'è un cluster di GPU[1] che lavora per settimane, alimentato da grandi quantità di energia elettrica e raffreddato da sistemi sofisticati. Ogni volta che usiamo un'applicazione basata su AI, i nostri dati viaggiano lungo dorsali in fibra ottica, attraversano cavi sottomarini, passano da un *Internet Exchange Point* e finiscono in un *data center*.

Questa prospettiva "materiale" è oggi più necessaria che mai. Le infrastrutture dell'AI non solo abilitano la potenza di calcolo, ma la distribuiscono – o la concentrano – in specifici luoghi del mondo. Alcune regioni diventano poli strategici di innovazione e potere, altre rischiano di restare escluse. L'asimmetria infrastrutturale diventa così una forma nuova di disuguaglianza digitale e, al tempo stesso, un terreno di competizione tra Stati e grandi aziende private.

C'è poi un ulteriore rischio: la concentrazione del controllo di queste risorse in mano a pochi attori globali. I grandi fornitori di *cloud*, i produttori di *chip* avanzati e i costruttori di cavi sottomarini stanno acquisendo un potere senza precedenti, in grado di influenzare non solo i mercati, ma anche la capacità degli Stati di esercitare la propria sovranità digitale. Un futuro in cui l'AI dipenda da pochi operatori dominanti significherebbe nuove forme di monopolio infrastrutturale, con conseguenze economiche e politiche rilevanti.

[1] GPU (*Graphics Processing Unit*): *chip* creato per elaborare grafica in parallelo, oggi usato anche per l'addestramento e l'inferenza dei modelli AI grazie alla sua capacità di gestire calcoli simultanei su ampia scala.

Il dibattito pubblico tende spesso a sottovalutare questa dimensione. Si discute molto di etica dell'AI, di *bias* algoritmici[2], di impatti sul lavoro. Si parla meno, invece, di chi controlla le reti, i *data center* e i *chip* che rendono l'AI possibile. Eppure è lì che si giocano alcune delle partite decisive del nostro tempo: dalla sovranità digitale europea alla guerra dei semiconduttori tra Stati Uniti e Cina, passando per il ruolo crescente delle *big tech* nell'infrastruttura di rete globale.

Questo articolo intende affrontare il tema delle infrastrutture dell'AI in una prospettiva integrata. Verranno analizzati tre grandi ambiti:

i. Le reti di telecomunicazioni, che costituiscono il sistema circolatorio della società digitale e trasportano i dati necessari all'AI.

ii. I *data center* e la potenza di calcolo, dove l'intelligenza artificiale prende forma grazie a *cluster* di GPU e architetture dedicate.

iii. La geopolitica della potenza di calcolo, cioè la distribuzione globale di queste risorse, le tensioni tra Stati e l'impatto delle politiche industriali.

iv. Infine, l'articolo discuterà anche gli impatti e le sfide globali, dal consumo energetico alle disuguaglianze nell'accesso.

In altre parole, se vogliamo capire davvero cosa significa "intelligenza artificiale", dobbiamo guardare sotto la superficie: non solo al codice, ma al ferro, ai cavi e all'energia che la rendono possibile, e anche a chi ne detiene il controllo.

2. Le reti come sistema circolatorio dell'AI

2.1. La banda larga come prerequisito dell'AI distribuita

L'intelligenza artificiale vive di dati. Ogni modello, linguistico o di visione, viene addestrato su miliardi di esempi. Questa mole deve essere raccolta, trasmessa, archiviata ed elaborata in

[2] *Bias* algoritmici: distorsioni nei sistemi *AI* che, a causa di dati o progettazioni sbilanciate, possono produrre risultati discriminatori in contesti come assunzioni o giustizia predittiva.

modo rapido e affidabile. Senza connessioni veloci e stabili, l'AI non potrebbe esistere nella forma attuale.

La banda larga è quindi il prerequisito fondamentale. Non riguarda solo la connettività domestica o aziendale, ma soprattutto infrastrutture ad altissima capacità: fibre ottiche, dorsali a lunga distanza, reti metropolitane, sistemi satellitari e *wireless* avanzati. L'addestramento di un modello di grandi dimensioni distribuisce il lavoro su migliaia di processori collocati in *data center* diversi, persino in continenti diversi. Questo "calcolo distribuito" richiede latenza minima e banda sufficiente per sincronizzare i nodi.

Il tema riguarda anche l'uso quotidiano dell'AI. Ogni volta che interagiamo con un assistente vocale, chiediamo a un modello di generare testo o affidiamo una diagnosi a un algoritmo, stiamo inviando e ricevendo dati che devono viaggiare su reti a larga banda.

L'utente finale si connette quasi sempre da una rete diversa da quella in cui risiede il modello: ad esempio, un cliente mobile può essere su una rete 5G di un operatore (ISP[3]), mentre il servizio AI si trova in un *data center* remoto su un altro ISP, magari dall'altra parte del continente.

Per questo, non tutti i *data center* hanno lo stesso ruolo. Quelli destinati all'addestramento dei modelli[4] possono avere meno interconnessioni con gli ISP: ciò che conta è il trasporto di grandi volumi tra pochi poli. Al contrario, i *data center* per l'inferenza[5], cioè dove l'AI fornisce risposte in tempo reale, devono essere il più possibile densi di interconnessioni, con una forte

[3] *Internet Service Provider* (ISP): termine definito come "fornitori di accesso a Internet", ovvero soggetti che mettono a disposizione connessioni alla rete e servizi telematici da essa derivati.

[4] Cfr. M. Naumov *et al.*, "Deep Learning Training in Facebook Data Centers: Design of Scale-up and Scale-out Systems", *arXiv:2003.09518*, (2020); Cfr. S. McIntosh-Smith *et al*., "Isambard-AI: a leadership class supercomputer optimised specifically for Artificial Intelligence", *arXiv:2410.11199*, (2024).

[5] Cfr. A. Delp *et al.* (2024), "How AI infrastructure supports training, inference and data in motion", in *Equinix*, recuperato da https://blog.equinix.com/blog/2024/12/04/how-ai-infrastructure-supports-training-inference-and-data-in-motion/; CoreSite Team (2024), "The backbone of AI performance: proximity and bandwidth", in *CoreSite*, recuperato da

presenza di ISP e *Autonomous System*[6]. Solo così si riduce la latenza e milioni di richieste possono raggiungere rapidamente i modelli, e tornare indietro con la stessa efficienza.

Qui diventano cruciali le infrastrutture di interconnessione: *Internet Exchange Point* (IXP[7]), *peering* privati[8], reti di transito[9]. Questi snodi garantiscono che la *query* dell'utente raggiunga il modello in modo rapido ed economico. In loro assenza, le richieste seguirebbero percorsi più lunghi e costosi, con impatto negativo su prestazioni e sostenibilità.

Negli ultimi anni, l'espansione della fibra e del 5G ha reso possibile la crescita delle applicazioni AI. Le reti mobili di nuova generazione, e in prospettiva il 6G, sono pensate per scenari a latenza ridottissima: veicoli autonomi, robot industriali coordinati in tempo reale, sistemi di realtà aumentata e virtuale. Tutti questi ambiti richiedono che l'elaborazione avvenga il più vicino possibile alla fonte dei dati, aprendo la strada all'*edge computing*[10].

In questo contesto, la banda larga non è solo un'infrastruttura di servizio, ma un fattore strategico. Paesi che investono in reti di nuova generazione e in punti di interconnessione neutrali possono attrarre *data center* e progetti innovativi. Al contrario, aree prive di connettività veloce rischiano di essere escluse. Il divario digitale non è più solo tra chi ha accesso a Internet e chi no, ma tra chi può sfruttare l'AI distribuita e chi resta frenato da reti insufficienti.

In definitiva, se l'AI è il nuovo motore della trasformazione economica e sociale, la banda larga è il carburante che lo ali-

https://www.coresite.com/blog/the-backbone-of-ai-performance-proximity-and-bandwidth.

[6] Cfr. J. Hawkinson, T. Bates, *Guidelines for creation, selection, and registration of an Autonomous System (AS)*, IETF, RFC, 1930.

[7] Cfr. *Internet eXchange Federation* (*IX-F*), *IXP Definition* (https://ix-f.net/ixp-definition/)

[8] Cfr. J. de Vasconcelos Aguiar, *All About Peering*, Internet Society https://www.internetsociety.org/blog/2025/03/all-about-peering/

[9] Cfr. Kentikpedia (2024), "IP Transit vs. Peering: What's the Difference?", in *Kentik*, recuperato da https://www.kentik.com/kentipedia/ip-transit-vs-peering/.

[10] Cfr. Aa.Vv. (2025), voce "Edge Computing", in *Wikipedia*, recuperato da https://it.wikipedia.org/wiki/Edge_computing.

menta. Ma perché questo motore giri senza intoppi servono snodi di interconnessione e *data center* capillarmente connessi con ISP e *Autonomous System*, così che miliardi di richieste isolate diventino un flusso globale efficiente e resiliente.

2.2. Cavi sottomarini e interconnessioni globali

Se reti locali e *Internet Exchange Point* garantiscono la vicinanza tra utenti e servizi, le infrastrutture di lunga distanza – cavi sottomarini, dorsali terrestri e satelliti – costituiscono il tessuto connettivo che collega i poli digitali del pianeta.

Oggi oltre il 97% del traffico Internet intercontinentale[11] viaggia attraverso cavi sottomarini[12]: fibre ottiche stese sui fondali oceanici, lunghe decine di migliaia di chilometri, che trasportano dati a velocità prossime a quella della luce. Per l'AI queste dorsali sono vitali: *dataset* immensi vengono spostati tra continenti e utenti a migliaia di chilometri interrogano modelli ospitati altrove. Senza tali connessioni, la sincronizzazione dei *cluster* di calcolo e l'accesso globale sarebbero impossibili.

La geografia dei cavi è cambiata radicalmente. Un tempo gestiti da consorzi telefonici, oggi i grandi attori del *cloud* e dell'AI – Google, Meta, Microsoft, Amazon – sono protagonisti centrali: possiedono o cofinanziano intere tratte per garantirsi capacità esclusiva e ridurre la dipendenza dagli operatori tradizionali[13]. Così, non solo forniscono servizi AI, ma controllano anche parte crescente delle infrastrutture che li rendono possibili.

Accanto ai cavi marini, vanno considerate le dorsali terrestri, in particolare quelle che attraversano continenti come Eurasia e Africa. Progetti come la "Digital Silk Road" cinese[14] o le nuove

[11] Cfr. "Dive into the Deep Sea: A View of the Subsea Cable Ecosystem", in *ENISA* (2023), recuperato da https://www.enisa.europa.eu/news/dive-into-the-deep-sea-a-view-of-the-subsea-cable-ecosystem.

[12] Si veda il sito TeleGeography al seguente indirizzo www.submarinecablemap.com.

[13] Cfr. "Big Tech's Undersea Cable Empire is Building the Digital Economy", in *PYMNTS* (2024), recuperato da https://www.pymnts.com/connectedeconomy/2024/big-techs-undersea-cable-empire-is-building-tomorrows-digital-economy/.

[14] Cfr. H. Shen, "Building a Digital Silk Road?", in *International Journal of Communication* 12 (2018), pp. 2683-2701.

fibre transafricane[15] hanno un impatto strategico paragonabile alle rotte oceaniche: collegano aree finora periferiche, aprono mercati e creano nuove dipendenze geopolitiche. Anche qui, il ruolo degli attori privati è crescente e ridisegna equilibri tra Stati.

Un ulteriore livello di connettività globale è rappresentato dalle reti satellitari. I sistemi geostazionari tradizionali hanno prestazioni limitate, ma le nuove costellazioni a orbita bassa (LEO[16]) – come Starlink, Kuiper o OneWeb – offrono velocità e tempi di risposta simili alla fibra ottica. Per l'AI distribuita, ciò significa connettività anche in aree rurali, remote o in emergenza, riducendo il *digital divide*.

Queste reti sollevano però problemi di concentrazione[17]. Il dispiegamento richiede investimenti colossali ed economie di scala tali che solo pochi operatori globali possono sostenerli, generando rischi di monopolio infrastrutturale. In tal caso, la capacità di connettere interi Paesi dipenderebbe dalle decisioni di un singolo fornitore privato, amplificando il divario tra chi accede a servizi avanzati e chi resta escluso.

In sintesi, le infrastrutture globali di rete – sottomarine, terrestri e satellitari – non sono semplici dettagli tecnici, ma il sistema nervoso planetario che consente all'AI di esistere come fenomeno globale. La loro proprietà, gestione e protezione diventano elementi centrali della nuova geopolitica digitale, dove la posta in gioco non è solo tecnica, ma di potere.

2.3. *Carrier*, *Access Provider*, *Content Provider* e IXPs: avvicinare l'AI agli utenti

Se i cavi globali e le dorsali internazionali costituiscono l'ossatura del traffico dati, a portare l'intelligenza artificiale fino agli utenti finali sono soprattutto gli *Internet Service Provider*

[15] Cfr. *African Fiber Optic Backbone Trans-Saharan*, ITU (https://www.itu.int/net4/wsis/archive/stocktaking/Project/Details?projectId=1514887279)

[16] Cfr. Aa.Vv. (2025), voce "Satellite Internet access", Sezione "Signal latency – Medium and low Earth orbits", in *Wikipedia*, recuperato da https://en.wikipedia.org/wiki/Satellite_Internet_access.

[17] Cfr. Report (2025), "Top Satellite Internet Service Providers Market Share", in *MarketsandMarkets*.

(ISP). All'interno di questa categoria si distinguono due attori complementari:

i. Carrier, che gestiscono le dorsali nazionali e internazionali, collegando città, *data center* e snodi intercontinentali;

ii. gli *Access Provider*, che curano l'"ultimo miglio", ovvero il collegamento diretto con abitazioni, aziende e dispositivi mobili.

Accanto a loro operano i *Content Provider*, ossia le grandi piattaforme che producono e distribuiscono contenuti e servizi digitali (motori di ricerca, *social network*, *streaming*, *cloud*). Nel contesto dell'AI hanno un ruolo decisivo: spesso sono i gestori stessi dei modelli e ne determinano la collocazione in rete, influenzando così la qualità dell'esperienza utente. La loro interconnessione con *carrier* e *access provider* è essenziale per garantire risposte rapide ed efficienti.

Senza *carrier* affidabili, access *provider* capillari e *content provider* ben connessi, l'AI resterebbe confinata nei grandi *hub* di calcolo. La qualità della connettività – velocità, latenza, resilienza – dipende dagli investimenti di questi operatori e dalla robustezza delle loro infrastrutture.

Un ruolo complementare è svolto dagli *Internet Exchange Point* (IXP), snodi neutri che permettono a *carrier*, *access provider* e *content provider* di interconnettersi direttamente. Non sostituiscono le reti di trasporto o di accesso, ma ne aumentano l'efficienza: riducono la latenza, abbassano i costi e rafforzano la resilienza. Senza IXPs, i flussi AI dovrebbero seguire percorsi più lunghi su reti di transito, con effetti negativi su prestazioni e stabilità.

Accanto a *carrier*, *access provider*, *content provider* e IXPs si afferma l'*edge computing*. Collocare nodi di calcolo vicino agli utenti riduce la latenza e migliora l'affidabilità dei servizi AI. Se l'addestramento dei modelli richiede *cluster* centralizzati, l'inferenza – la generazione delle risposte – può essere distribuita lungo la rete. Applicazioni come guida autonoma, industria 4.0 o *smart city* ne traggono particolare vantaggio.

In definitiva, il percorso che porta l'AI all'utente passa da quattro livelli complementari:

i. Carrier e dorsali che trasportano i dati su scala globale;

ii. Access Provider che garantiscono l'ultimo miglio;

iii. *Content Provider*, che distribuiscono contenuti e modelli AI;

iv. IXPs, che ottimizzano l'interconnessione, mantenendo Internet aperta ed efficiente.

Questa architettura multilivello consente oggi a miliardi di persone di interrogare un modello AI in tempo reale, trasformando un'infrastruttura globale in un'esperienza personale immediata.

3. *Data center* e potenza di calcolo

3.1. Dall'*hyperscale* al GPU-*cluster*: nuove architetture per l'AI

Per anni i *data center* sono stati pensati come infrastrutture *general purpose*: grandi magazzini di server in grado di ospitare applicazioni diverse, dalla posta elettronica ai database, fino ai servizi *cloud* pubblici. Con l'avvento dell'AI questa logica è stata stravolta: servono capacità di calcolo immense e specifiche, che hanno trasformato profondamente l'architettura dei *data center*[18].

I *data center hyperscale*, costruiti da attori globali come Amazon, Google, Microsoft, Meta o Alibaba, sono stati il primo passo: strutture con centinaia di migliaia di *server*, pensate per servizi *cloud* planetari[19].

Per l'AI, però, questi modelli non bastano più. L'addestramento di reti neurali di ultima generazione, come i *Large Language Model* (GPT, Claude, Gemini), richiede *cluster* dedicati composti da migliaia di GPU o da *chip* specializzati come le TPU[20] di Google o le architetture custom di Amazon e Microsoft.

[18] Cfr. "What Is an AI Data Center?", in *Cisco* (2025), recuperato da https://www.cisco.com/site/us/en/learn/topics/computing/what-is-an-ai-data-center.html.

[19] Cfr. "Hyperscale Data Center Count Hits 1,136", in *Synergy Research Group* (2025), recuperato da https://www.srgresearch.com/articles/hyperscale-data-center-count-hits-1136-average-size-increases-us-accounts-for-54-of-total-capacity.

[20] Cfr. *Cloud TPU*, Google Cloud (2024), https://cloud.google.com/tpu

Questi *cluster* non sono semplici *server* aggregati, ma sistemi che funzionano come un unico *supercomputer* distribuito. Ogni nodo deve comunicare a velocità elevatissime, scambiando continuamente parametri e gradienti. Per questo, l'architettura interna è cruciale quanto la potenza dei *chip*: interconnessioni come NVLink o InfiniBand[21] diventano fondamentali, insieme a soluzioni di storage ottimizzate per *dataset* da *petabyte*.

I *data center* non svolgono tutti lo stesso compito:

i. Alcuni sono dedicati all'addestramento, con *cluster* GPU ad altissima densità e interconnessioni interne ultra-performanti, ma con connettività esterna più limitata.

ii. Altri sono ottimizzati per l'inferenza, cioè per rispondere agli utenti finali: qui conta la connettività, con forte interscambio con ISP e IXPs.

iii. Stanno inoltre emergendo i *data center* di prossimità o *edge* DC, vicini agli utenti e ai dispositivi. Meno potenti degli *hyperscale*, riducono la latenza e alleggeriscono il traffico delle dorsali, risultando cruciali per applicazioni AI in tempo reale come auto connesse, fabbriche intelligenti, sanità digitale e *smart city*[22].

Accanto ai grandi LLM generalisti, emergono LLM verticali e specializzati: modelli più compatti, addestrati su domini ristretti (biomedicina, diritto, finanza)[23]. Questi sistemi potrebbero non richiedere cluster *hyperscale*, ma accontentarsi di *data center* intermedi o *edge*, soprattutto se confinati a contesti settoriali[24]. Ciò apre a una pluralità architetturale, in cui non esiste più un'unica scala di riferimento.

Ne deriva una nuova geografia del calcolo AI: pochi poli di addestramento concentrati e, in parallelo, nodi di inferenza e

[21] Cfr. *NVLink and InfiniBand for Accelerated Computing*, NVIDIA (2024) https://www.nvidia.com/en-us/data-center/nvlink/

[22] Cfr. *Data Centers At The Edge*, Uptime Institute (2021), recuperato da https://uptimeinstitute.com/data-centers-at-the-edge.

[23] Cfr. Y. He, X. Zhu, D. Li, H. Wang, "Enhancing Large Language Models for Specialized Domains: A Two-Stage Framework with Parameter-Sensitive LoRA Fine-Tuning and Chain-of-Thought RAG", in *Electronics* 14 (2025), n. 1961, https://doi.org/10.3390/electronics14101961.

[24] *Ibidem*

edge diffusi e capillarmente connessi. È un cambiamento che spinge i *data center* a specializzarsi e reinventarsi.

Questa evoluzione ha effetti anche sul mercato. La domanda di GPU e *chip AI-ready* ha scatenato una corsa senza precedenti, con NVIDIA in posizione dominante ma con competitori come AMD, Intel e produttori asiatici[25]. La capacità di un *data center* di accedere a queste risorse diventa un fattore critico di competitività.

In sintesi, l'era dell'AI segna il passaggio dai *data center hyperscale* generalisti a GPU-*cluster* specializzati, affiancati da nodi di inferenza, *Edge* DC e strutture dedicate a modelli verticali. Una geografia più sfaccettata, in cui potenza di calcolo, architettura interna e interconnessione esterna si combinano per sostenere la nuova economia dell'AI.

3.2. Consumi energetici e raffreddamento: il costo fisico dell'intelligenza artificiale

Ogni volta che un modello viene addestrato o interrogato, migliaia di processori lavorano in parallelo, consumando notevoli quantità di energia. Se l'AI appare come un servizio immateriale, è in realtà sostenuta da infrastrutture che richiedono elettricità, sistemi di raffreddamento e risorse naturali crescenti.

L'addestramento di un *Large Language Model* può assorbire centinaia di megawattora, pari al consumo di una piccola città per giorni[26].

Ma anche l'inferenza è energivora: milioni di *query* quotidiane, moltiplicate per miliardi di utenti, producono una domanda costante di potenza di calcolo, talvolta superiore a quella del *training*[27].

[25] Cfr. J. McGeever (2025), "Can Nvidia results dispel creeping AI doubts?", in *Reuters*, recuperato da https://www.reuters.com/technology/artificial-intelligence/can-nvidia-results-dispel-creeping-ai-doubts-2025-08-26/.

[26] Cfr. C.D. Costa (2025), "The Hidden Cost of AI: How Machine Learning Is Draining Our Energy Resources", in *Medium*, recuperato da https://medium.com/@clairedigitalogy/the-hidden-cost-of-ai-how-machine-learning-is-draining-our-energy-resources-21d3d36af9d1.

[27] Cfr. A. Garg, I. Kitsara (2025), "The hidden cost of AI: Unpacking its energy and water footprint", in *OECD.AI*, recuperato da https://oecd.ai/en/wonk/the-hidden-cost-of-ai-energy-and-water-footprint.

A ciò si aggiunge il problema del raffreddamento. I *cluster* di GPU e TPU generano calore estremo e richiedono sistemi sempre più sofisticati: oltre al raffreddamento ad aria si sperimentano oggi soluzioni a liquido diretto o a immersione totale, molto più efficienti[28].

Il raffreddamento non riguarda solo l'energia: molti *data center* utilizzano grandi quantità di acqua per sistemi di evaporazione o per alimentare torri di raffreddamento. In regioni già esposte a stress idrico, come California, Medio Oriente o Spagna, la competizione tra usi industriali e bisogni civili può diventare un tema politico e sociale[29].

Il fabbisogno energetico dell'AI ha anche una dimensione geopolitica: i *data center* si collocano in aree dove l'energia è abbondante e a basso costo – idroelettrico in Islanda e Norvegia, gas negli Stati Uniti, carbone in Cina, fonti fossili negli Emirati – delineando una geografia che influenza sostenibilità e sovranità tecnologica.

Per ridurre l'impatto, i grandi provider stanno cercando di aumentare la quota di energia rinnovabile nei loro *data center* e di ottimizzare l'efficienza degli impianti[30].

I grandi *provider* dichiarano obiettivi di neutralità climatica e aumentano l'uso di rinnovabili, ma la crescita esponenziale dell'AI rischia di annullare i progressi. Come accaduto con Internet, ogni guadagno in efficienza è rapidamente assorbito dall'aumento dell'uso[31].

Alcune innovazioni, come il modello *DeepSeek*, dimostrano tuttavia che tecniche di addestramento più efficienti possono ridurre i consumi rispetto alle stime più pessimistiche[32].

[28] Cfr. K. Haghshenas, B. Setz, Y. Blosch, M. Aiello, "Enough Hot Air: The Role of Immersion Cooling", in *Energy Inform*, 6 (2023), n. 14, https://doi.org/10.1186/s42162-023-00269-0.

[29] Cfr. L. Pengfei *et al*, "Making AI Less 'Thirsty'", in *arXiv:2304.03271*, (2023).

[30] G. Kamiya, V.C. Coroamă, "Data Centre Energy Use: Critical Review of Models and Results", in *EDNA – IEA 4E TCP*, (2025).

[31] Cfr. O. Le Poidevin (2025), "Tech giants' indirect emissions rose 150% in three years as AI expands", in *Reuters*, recuperato da https://www.reuters.com/sustainability/climate-energy/tech-giants-indirect-emissions-rose-150-three-years-ai-expands-un-agency-says-2025-06-05/.

[32] Cfr. Aa.Vv. (2025), voce "DeepSeek", in *Wikipedia*, recuperato da https://en.wikipedia.org/wiki/DeepSeek.

Il costo fisico dell'AI non può essere ignorato: non riguarda solo elettricità e acqua, ma la necessità di conciliare un'industria strategica con la sostenibilità. La risposta non può limitarsi a soluzioni tecniche, ma richiede anche politiche pubbliche e investimenti in infrastrutture resilienti. In definitiva, se i *data center* sono il cuore dell'AI, energia e raffreddamento sono il sangue e l'ossigeno che ne consentono il funzionamento e ne fissano i limiti di crescita.

3.3. Sicurezza e resilienza delle infrastrutture critiche

I *data center* dedicati all'intelligenza artificiale non sono tutti uguali. Alcuni hanno caratteristiche tali da poter essere considerati vere e proprie infrastrutture critiche, altri invece svolgono un ruolo più circoscritto. La differenza dipende in larga misura dal livello di ridondanza delle componenti e dal grado di concentrazione dei servizi che ospitano.

Il primo elemento da considerare è la continuità di servizio. I sistemi AI, soprattutto quelli che supportano funzioni sanitarie, finanziarie o di pubblica sicurezza, non possono permettersi interruzioni prolungate. Per questo i *data center* più avanzati adottano architetture ridondate: alimentazioni elettriche multiple, gruppi di continuità, linee di connettività diversificate, sistemi di *backup* e *disaster recovery*. Quando queste misure mancano o non sono sufficienti, il rischio è che un guasto singolo possa tradursi in un'interruzione di servizi considerati critici.

Accanto ai rischi fisici, vi sono quelli *cyber*. I data center per l'AI custodiscono *dataset* di valore e risorse computazionali preziose. Attacchi informatici mirati potrebbero sottrarre informazioni, manipolare i modelli o interrompere l'erogazione dei servizi. La crescente centralità dell'AI li rende obiettivi sensibili, anche se non tutti i *data center* hanno la stessa esposizione: quelli con carichi distribuiti e ridondanti sono più resilienti, quelli più concentrati sono inevitabilmente più vulnerabili.

La localizzazione geografica influisce anch'essa. Strutture collocate in aree politicamente instabili o in zone soggette a rischi ambientali (alluvioni, sismi, ondate di calore) sono più esposte a interruzioni. Anche in questo caso, la ridondanza – distribuire i carichi su più siti e regioni – riduce il rischio di dipendenza da un singolo punto di vulnerabilità.

Un aspetto da non sottovalutare riguarda la sovranità digitale. Quando le risorse AI si trovano fuori dai confini nazionali, un Paese può dipendere da infrastrutture non controllate direttamente, con il rischio che decisioni commerciali o politiche di terzi abbiano conseguenze sui propri servizi. Un esempio concreto è il *Cloud Act* statunitense[33], che consente alle autorità USA di richiedere l'accesso a dati gestiti da *provider* americani, anche quando fisicamente ospitati in Europa o in altre regioni. Ciò dimostra come la giurisdizione legale possa prevalere sulla collocazione geografica, mettendo in discussione l'autonomia dei Paesi che si affidano a fornitori esteri.

Infine, resta il tema della *supply chain*. Le infrastrutture AI dipendono da componenti – GPU, apparati di rete, sistemi di raffreddamento – prodotti in un numero limitato di Paesi e spesso soggetti a restrizioni commerciali. Una catena di fornitura fragile può trasformare una vulnerabilità tecnica in una criticità strategica.

In sintesi, non tutti i *data center* AI sono automaticamente infrastrutture critiche. Lo diventano quando mancano adeguati livelli di ridondanza, quando ospitano servizi essenziali senza alternative, o quando la concentrazione di risorse in pochi poli crea un punto singolo di vulnerabilità. La resilienza, in questo senso, non è una condizione statica, ma il risultato di scelte tecniche, organizzative e politiche.

4. *Hardware* e piattaforme *cloud*

4.1. La corsa ai *chip*

Se l'intelligenza artificiale è il motore della trasformazione digitale, i *chip* ne sono i cilindri. Senza processori specializzati, l'AI non potrebbe né addestrare modelli di grandi dimensioni né erogare servizi in tempo reale a miliardi di utenti. Negli ultimi anni la domanda di GPU e di acceleratori dedicati è cresciuta a ritmi vertiginosi[34], trasformando il mercato dei

[33] Cfr. *Clarifying Lawful Overseas Use of Data (CLOUD) Act*, U.S. Congress, 2018, Public Law, PL 115-141.

[34] Cfr. K. Hill (2025), "Five predictions for the AI infrastructure market", in *RCR Wireless News*, recuperato da https://www.rcrwireless.

semiconduttori in uno dei fronti più caldi della competizione tecnologica globale.

Per decenni le CPU (*Central Processing Unit*) sono state il cuore dei *data center*. Con l'avvento del *machine learning* e in particolare delle reti neurali profonde, le GPU (*Graphics Processing Unit*) hanno preso il sopravvento: progettate per elaborare calcoli in parallelo, si sono rivelate ideali per l'addestramento e l'inferenza dei modelli AI. Oggi il nome di NVIDIA è quasi sinonimo di intelligenza artificiale: con una quota di mercato dominante e una pipeline di prodotti ottimizzati per il calcolo distribuito (serie A100, H100, e la nuova generazione Blackwell), l'azienda è diventata un attore strategico di rilevanza geopolitica[35].

Ma la corsa non si ferma alle GPU. Google ha sviluppato le sue TPU (*Tensor Processing Unit*), ottimizzate per i *workload* di machine learning. Amazon e Microsoft hanno iniziato a progettare *chip* proprietari per ridurre la dipendenza da fornitori esterni[36]. Parallelamente, emergono alternative architetturali come *ARM* e *RISC-V*, che aprono prospettive di maggiore flessibilità e riduzione dei costi, sebbene ancora lontane dalle prestazioni dei *chip* più avanzati[37].

La produzione di questi semiconduttori è concentrata in pochissimi attori. TSMC (Taiwan Semiconductor Manufacturing Company) è il principale produttore al mondo di *chip* avanzati a 3 e 5 nanometri, seguita a distanza da Samsung[38]. Questa concentrazione geografica in Asia orientale rende la *supply chain* estremamente vulnerabile, soprattutto nel contesto delle ten-

com/20250220/fundamentals/ai-infrastructure-market.

[35] Cfr. AA.Vv (2025), voce "Nvidia", in *Wikipedia*, recuperato da https://it.wikipedia.org/wiki/NVIDIA.

[36] Cfr. T. Warren (2023), "Microsoft is finally making custom chips — and they're all about AI", in *The Verge*, recuperato da https://www.theverge.com/2023/11/15/23960345/microsoft-cpu-gpu-ai-chips-azure-maia-cobalt-specifications-cloud-infrastructure.

[37] Cfr. A. Shilov (2025), "RISC-V in AI and HPC… Per Aspera Ad Astra?", in *EE Times*, recuperato da https://www.eetimes.com/risc-v-in-ai-and-hpc-part-1-per-aspera-ad-astra/.

[38] Aa.Vv. (2025), voce "Semiconductor industry", sezione "Foundry model", in *Wikipedia*, recuperato da https://en.wikipedia.org/wiki/Foundry_model.

sioni tra Stati Uniti e Cina. Le restrizioni all'*export* imposte da Washington mirano proprio a impedire a Pechino di accedere alle tecnologie più sofisticate, nel timore che possano alimentare la sua corsa all'AI e al calcolo ad alte prestazioni.

La scarsità di GPU è diventata così evidente che, in alcuni momenti del 2023 e del 2024, anche grandi aziende hanno faticato a reperire capacità computazionale per addestrare nuovi modelli[39]. Questa "crisi dei *chip* AI" ha reso chiaro che il controllo sulle filiere dei semiconduttori equivale, in prospettiva, al controllo sulla capacità di sviluppare intelligenza artificiale.

In questo quadro, l'Europa si trova in una posizione particolarmente delicata. Pur avendo una tradizione industriale solida in settori come l'elettronica di potenza e l'*automotive*, il continente è fortemente dipendente dalle tecnologie prodotte altrove, soprattutto per i *chip* più avanzati. In risposta, l'UE ha lanciato il *Chips Act europeo*[40], con l'obiettivo di raddoppiare la quota di mercato globale della produzione di semiconduttori entro il 2030 (dal 10% al 20%). Sono stati annunciati grandi investimenti, come i progetti Intel in Germania e STMicroelectronics in Francia e Italia[41], ma la strada per raggiungere la *leadership* rimane lunga. La capacità di colmare il divario dipenderà non solo dai finanziamenti, ma anche dalla creazione di un ecosistema industriale competitivo, capace di attrarre talenti e *partnership* tecnologiche.

Non va trascurato il tema del consumo energetico: ogni nuova generazione di *chip* promette maggiore efficienza, ma l'aumento della domanda spesso annulla i benefici. La ricerca di soluzioni più sostenibili sta spingendo verso acceleratori sempre più spe-

[39] Cfr. S. Nellis, A. Soni (2024), "Nvidia's supply snags limit deliveries even as demand booms", in *Reuters*, recuperato da https://www.reuters.com/technology/nvidias-supply-snags-hurting-deliveries-mask-booming-demand-2024-11-21/.

[40] Cfr. *European Chips Act*, European Commission, Regulation (EU) 2023/1781 of the European Parliament and of the Council of 13 September 2023 establishing a framework of measures for strengthening Europe's semiconductor ecosystem and amending Regulation (EU) 2021/694.

[41] Cfr. F.Y. Chee (2023), "EU approves €8.1 billion of state aid for microelectronics projects, in *Reuters*", recuperato da https://www.reuters.com/technology/eu-launches-81-bln-euro-microelectronics-projects-with-56-companies-2023-06-08/

cializzati, in grado di ridurre i costi per inferenza e rendere più accessibile l'AI su larga scala.

In questo scenario, la corsa ai *chip* è molto più di una sfida industriale. È una competizione geopolitica in cui governi e aziende private investono miliardi per assicurarsi il controllo di una risorsa che oggi è considerata al pari del petrolio o del gas: senza *chip*, non c'è AI.

4.2. Il ruolo degli *hyperscaler*

Se i *chip* sono il cuore dell'AI, gli *hyperscaler* ne sono i principali abilitatori. Con questo termine si indicano i grandi operatori globali del *cloud* – Amazon Web Services (AWS), Microsoft Azure, Google Cloud, Alibaba Cloud e, in misura crescente, Tencent e Oracle – che gestiscono *data center* planetari e offrono capacità computazionale "a consumo".

Il loro ruolo è duplice. Da un lato investono in infrastrutture per sviluppare modelli proprietari: OpenAI si appoggia a Microsoft, Anthropic a Google, Amazon sostiene *startup* come Cohere. Dall'altro forniscono servizi di AI *as a Service*, mettendo a disposizione GPU, TPU e altri acceleratori via *cloud*, accessibili anche a chi non può costruire *cluster* dedicati.

In questo modo gli *hyperscaler* sono diventati gli attori dominanti dell'AI: chi vuole addestrare o usare modelli avanzati deve spesso passare dai loro *data center*[42]. La scarsità di GPU ha reso questa dipendenza più evidente: grazie a contratti preferenziali con i produttori, possono riservarsi risorse prima degli altri, consolidando la posizione dominante.

La loro forza va oltre l'*hardware*. Offrono piattaforme *software* e servizi verticali: librerie ottimizzate, ambienti di sviluppo e modelli pre-addestrati. Ciò genera *lock-in*: cambiare *provider* comporta costi e complessità notevoli.

Un aspetto chiave è la distribuzione geografica. Gli *hyperscaler* hanno costruito regioni e zone di disponibilità globali,

[42] Cfr. T. Houck, M. Vincent (2025), "AI Supercharges Hyperscale: Capacity, Geography, and Design Are Being Redrawn", in *Data Center Frontier*, recuperato da https://www.datacenterfrontier.com/hyperscale/article/55308095/ai-supercharges-hyperscale-capacity-geography-and-design-are-being-redrawn.

garantendo ridondanza e prossimità agli utenti. Tuttavia la copertura non è uniforme: Nord America ed Europa occidentale concentrano la maggior parte dei *data center*, mentre Africa, America Latina e parte dell'Asia restano sottorappresentate[43], con nuove disuguaglianze nell'accesso all'AI.

Il caso europeo è emblematico. Pur avendo un mercato digitale ampio, la maggior parte dei servizi *cloud* e infrastrutture AI è in mano ai *big player* americani: AWS, Microsoft e Google insieme detengono oltre i due terzi del mercato *cloud* europeo[44].

Questa dipendenza è anche giuridica: il *Cloud Act* statunitense consente alle autorità USA di richiedere dati a *provider* americani anche se conservati in Europa, minando la sovranità digitale.

Per reagire, l'UE ha avviato iniziative come GAIA-X[45], per un ecosistema *cloud* federato e interoperabile, e l'IPCEI-CIS[46], per stimolare investimenti pubblici e privati. Ma questi tentativi faticano a competere con le economie di scala dei giganti americani e cinesi.

La sfida per l'Europa resta quella di bilanciare l'uso dei servizi offerti dagli *hyperscaler* senza cadere in una dipendenza strutturale, che in ambito AI rischia di diventare perdita di autonomia tecnologica, economica e politica.

Gli *hyperscaler* non sono semplici fornitori di *cloud*, ma attori centrali dell'AI globale. La loro capacità di concentrare risorse e infrastrutture li rende imprescindibili, ma accentua il rischio di un potere digitale concentrato in poche mani. Per l'Europa

[43] Cfr. "How Many Data Centers Are There and Where Are They Being Built?", in *ABI Research* (2024), recuperato da https://www.abiresearch.com/blog/data-centers-by-region-size-company.

[44] Cfr. "European Cloud Market", in Synergy Research Group, (2023), recuperato da https://www.srgresearch.com/articles/european-cloud-providers-local-market-share-now-holds-steady-at-15

[45] Cfr. *About GAIA-X*, GAIA-X Association (2024), recuperato da https://gaia-x.eu/about/.

[46] Cfr. "Commission approves up to €1.2 billion of State aid by seven Member States for an Important Project of Common European Interest in cloud and edge computing technologies", in *EU Commission*, Press release IP_23_6246, (2023), recuperato da https://digital-strategy.ec.europa.eu/en/news/commission-approves-eu12-billion-state-aid-seven-member-states-important-project-common-european.

questa è una prova decisiva: senza strategie efficaci, la dipendenza dagli *hyperscaler* diventerà un vincolo alla propria sovranità digitale.

4.3. Sovranità tecnologica e alternative emergenti

La concentrazione delle infrastrutture AI nelle mani di pochi attori globali – produttori di *chip*, *hyperscaler*, fornitori di *cloud* e piattaforme *software* – pone un problema crescente di sovranità tecnologica. Se la capacità di sviluppare, addestrare e utilizzare modelli di intelligenza artificiale dipende da risorse controllate all'estero, la possibilità per un Paese o un'area geografica di definire la propria traiettoria digitale si riduce sensibilmente.

L'Europa è forse l'esempio più evidente di questa tensione. Pur avendo un mercato digitale ampio e regolamentato, non dispone ancora di infrastrutture proprie in grado di competere con le grandi piattaforme americane o cinesi. Progetti come GAIA-X mirano a creare un ecosistema *cloud* federato, trasparente e interoperabile, che consenta agli attori europei di condividere risorse in maniera coordinata. Tuttavia, le difficoltà di *governance* e la lentezza di implementazione ne hanno ridotto finora l'impatto concreto.

Parallelamente, si stanno sviluppando iniziative pubbliche e private per promuovere la produzione di semiconduttori e *hardware* avanzato in Europa, sostenute dal *Chips Act europeo*. L'obiettivo è ambizioso: raddoppiare la quota di mercato globale della produzione di *chip* entro il 2030. Tuttavia, la competizione con colossi come TSMC e Samsung resta proibitiva, e il rischio è che l'Europa rimanga più un acquirente che un produttore, salvo in nicchie altamente specializzate.

Fuori dall'Occidente, emergono altre traiettorie. La Cina ha fatto della sovranità tecnologica un obiettivo strategico, investendo massicciamente nello sviluppo di GPU locali e nella costruzione di *data center* nazionali, con l'intento di ridurre la dipendenza dalle tecnologie statunitensi. Nonostante le restrizioni all'*export* imposte da Washington, Pechino sta cercando di

costruire un ecosistema alternativo, anche se per ora con prestazioni inferiori ai *chip* più avanzati di NVIDIA e TSMC[47].

Nel frattempo, si affacciano sul mercato anche alternative architetturali: progetti *open source* come *RISC-V* promettono di abbassare la barriera di ingresso alla progettazione di *chip*, consentendo a università, *startup* e governi di sviluppare soluzioni più indipendenti dai grandi produttori. Se queste architetture riusciranno a scalare in termini di prestazioni e diffusione, potrebbero rappresentare un tassello importante nella diversificazione del panorama AI.

Un altro fronte riguarda il modello di distribuzione delle risorse. Accanto agli *hyperscaler*, stanno crescendo operatori di *cloud* regionali e specializzati[48], che puntano a fornire servizi AI con un maggiore controllo locale sui dati e sulle infrastrutture. In alcuni casi, si tratta di collaborazioni pubblico-private che cercano di rafforzare la resilienza nazionale. Anche i consorzi di ricerca e le reti accademiche svolgono un ruolo importante, sperimentando forme di condivisione delle risorse computazionali per ridurre i costi e aumentare l'accessibilità[49].

In prospettiva, la vera sfida non è tanto sostituire i giganti globali, quanto creare un ecosistema pluralistico e interoperabile, dove più attori possano coesistere e competere. La sovranità tecnologica, infatti, non significa isolamento, ma capacità di scelta: essere in grado di decidere quali infrastrutture usare, con quali garanzie, e senza subire vincoli unilaterali da parte di pochi *player* dominanti.

In sintesi, le alternative emergenti – dai progetti europei al *cloud* regionale, dalle architetture *open source* agli sforzi cinesi – mostrano che la concentrazione attuale non è un destino

[47] Cfr. G.C. Allen (2025), "DeepSeek, Huawei, Export Controls, and the Future of the U.S.–China AI Race", in *CSIS*, recuperato da https://www.csis.org/analysis/deepseek-huawei-export-controls-and-future-us-china-ai-race.

[48] Cfr. D. Panenkov (2025), "Cloud sovereignty in Europe — and beyond — a tipping point?", in *TechRadar Pro*, recuperato da https://www.techradar.com/pro/cloud-sovereignty-in-europe-and-beyond-a-tipping-point.

[49] Cfr. *European Open Science Cloud (EOSC)*, European Commission, consultabile in https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en.

inevitabile. Ma trasformare queste iniziative in opzioni concrete richiederà tempo, investimenti e soprattutto la volontà politica di trattare le infrastrutture dell'AI come una questione strategica, non solo tecnologica.

5. La geopolitica della potenza di calcolo

5.1. Stati Uniti e Cina: la competizione per l'AI

La corsa all'intelligenza artificiale è diventata il nuovo fronte della rivalità geopolitica tra Stati Uniti e Cina. Non si tratta solo di innovazione tecnologica, ma di una sfida per il controllo delle infrastrutture, delle competenze e delle risorse che determineranno la *leadership* economica e militare del XXI secolo.

Dal lato statunitense, il governo ha compreso presto che l'AI non è un settore come gli altri. Washington ha favorito la crescita di un ecosistema dominato da grandi piattaforme private – Google, Microsoft, Amazon, Meta, NVIDIA – che hanno risorse finanziarie e competenze tali da guidare l'avanzamento tecnologico. Il modello americano si fonda sull'interazione tra settore privato e pubblico: alle aziende spetta l'innovazione, al governo il compito di sostenerle e proteggerle, anche attraverso politiche industriali e di sicurezza nazionale.

Il *CHIPS and Science Act*[50], approvato nel 2022, ne è un esempio: prevede ingenti investimenti per rafforzare la produzione domestica di semiconduttori, incentivando la costruzione di fabbriche sul suolo americano e riducendo la dipendenza da fornitori esteri. Parallelamente, l'amministrazione Biden ha imposto restrizioni all'*export* di GPU avanzate verso la Cina[51], con l'obiettivo di rallentare la capacità di Pechino di sviluppare modelli AI su larga scala. In questo modo, gli Stati Uniti cercano di mantenere un vantaggio competitivo non solo tecnologico, ma anche strategico.

[50] Cfr. Aa.Vv. (2025), voce "CHIPS and Science Act", in *Wikipedia*, recuperato da https://en.wikipedia.org/wiki/CHIPS_and_Science_Act.

[51] Cfr. "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)", in *BIS*, (2022).

Sul fronte cinese, l'approccio è opposto ma altrettanto deciso. Il governo di Pechino ha definito l'AI una priorità nazionale, inserendola nei piani di sviluppo economico e militare[52]. Le grandi aziende cinesi – Baidu, Alibaba, Tencent, Huawei – sono sostenute da politiche pubbliche e da massicci investimenti statali[53]. La Cina ha sviluppato un ecosistema AI vivace, con modelli linguistici proprietari e piattaforme applicative che rispondono al mercato interno, spesso protetto da barriere all'ingresso per le aziende straniere.

La sovranità tecnologica è al centro della strategia cinese. Le sanzioni americane hanno evidenziato la dipendenza da fornitori esteri, in particolare per i semiconduttori più avanzati. Per questo Pechino sta accelerando sulla produzione domestica di *chip* e sul rafforzamento delle proprie catene di fornitura. I progressi non sono immediati – le fabbriche cinesi sono ancora indietro rispetto a TSMC o Samsung – ma la traiettoria è chiara: ridurre la vulnerabilità e costruire un'alternativa autonoma.

La competizione non si gioca solo sulla tecnologia, ma anche sulle infrastrutture globali. Gli Stati Uniti, attraverso le *big tech*, controllano gran parte dei cavi sottomarini, delle regioni *cloud* e delle piattaforme *software* più diffuse. La Cina, dal canto suo, promuove la *Digital Silk Road*, un'estensione della *Belt and Road Initiative* che punta a costruire cavi, reti 5G e *data center* in Asia, Africa e America Latina. In questo modo, Pechino cerca di espandere la propria influenza digitale al di fuori dei confini nazionali.

La rivalità si riflette anche sul piano normativo e valoriale. Washington spinge per un modello di AI "responsabile"[54], basato su regole di mercato e sulla tutela dei diritti, mentre Pechino privilegia un approccio orientato al controllo e alla stabilità

[52] Cfr. *New Generation Artificial Intelligence Development Plan*, State Council of the PRC (2017), n. 35.

[53] Cfr. W. Chang *et al*. (2025), "China's drive toward self-reliance in artificial intelligence", in *MERICS*, recuperato da https://merics.org/en/report/chinas-drive-toward-self-reliance-artificial-intelligence-chips-large-language-models.

[54] Cfr. *Blueprint for an AI Bill of Rights*, The White House (2022), consultabile in https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/.

interna[55]. La battaglia degli standard – tecnici, etici e giuridici – è destinata a influenzare profondamente lo sviluppo globale dell'AI.

Pur non essendo un polo di pari forza rispetto a USA e Cina, l'Europa si trova al centro della competizione. Da un lato beneficia delle innovazioni americane, dall'altro subisce le restrizioni commerciali imposte a Pechino, che rischiano di complicare l'accesso ai *chip* e alle tecnologie di calcolo. L'UE tenta di ritagliarsi uno spazio con politiche industriali come il *Chips Act europeo* e iniziative di sovranità digitale, ma la distanza in termini di scala e capacità d'investimento rispetto alle due superpotenze resta ampia[56]. In questo senso, la rivalità tra Washington e Pechino non è solo uno scontro bilaterale: definisce anche i margini di manovra del resto del mondo, Europa inclusa.

5.2. Europa: tra dipendenza e ambizioni

L'Europa è uno dei più grandi mercati digitali al mondo, ma quando si tratta di infrastrutture per l'intelligenza artificiale si trova in una posizione fragile: dipende in larga misura da tecnologie, piattaforme e fornitori esterni, soprattutto statunitensi. Questa condizione di dipendenza rischia di limitarne la sovranità digitale, la capacità di innovazione e, in ultima istanza, l'autonomia politica.

Sul fronte del *cloud*, la quota di mercato europea è dominata da AWS, Microsoft Azure e Google Cloud, che insieme controllano oltre i due terzi dei servizi utilizzati dalle imprese e dalle istituzioni del continente. Gli operatori locali hanno dimensioni molto più ridotte e faticano a competere con la scala e l'integrazione tecnologica degli *hyperscaler* globali. Ciò significa che gran parte della capacità di calcolo necessaria per sviluppare e distribuire modelli AI in Europa è ospitata su infrastrutture controllate da aziende americane.

Eppure esiste un'industria *cloud* europea che, sebbene frammentata e priva della potenza di fuoco dei giganti statunitensi e cinesi, svolge un ruolo importante. Attori come OVHcloud in

[55] Cfr. *Interim Administrative Measures for Generative Artificial Intelligence Services*, Cyberspace Administration of China (CAC) (2023).

[56] Cfr. *European Chips Act*, cit.

Francia, Deutsche Telekom (T-Systems) in Germania, Orange e Atos, insieme a realtà più piccole come Aruba in Italia, offrono soluzioni *cloud* regionali, spesso più attente alla conformità normativa locale e alla protezione dei dati. Questi operatori hanno iniziato a federarsi o a collaborare in iniziative comuni (come GAIA-X) per rafforzare la propria posizione, puntando sulla prossimità ai clienti, sull'interoperabilità e su un modello di *business* meno dipendente da logiche monopolistiche.

Questa dipendenza dall'estero ha anche una dimensione giuridica. Il *Cloud Act* statunitense consente alle autorità americane di richiedere dati a *provider* con sede negli Stati Uniti, anche quando i dati sono conservati fisicamente in Europa. Ciò solleva dubbi sulla reale possibilità per i governi e le imprese europee di garantire la piena riservatezza e il controllo delle proprie informazioni strategiche.

Consapevole di questa vulnerabilità, l'Unione Europea ha lanciato una serie di iniziative per rafforzare la propria sovranità digitale. Tra queste, il progetto GAIA-X, nato con l'obiettivo di creare un ecosistema federato di servizi *cloud* e dati basato su standard aperti, trasparenza e interoperabilità. L'ambizione era quella di fornire un'alternativa credibile al dominio degli *hyperscaler* americani e cinesi. Tuttavia, le difficoltà di *governance*, i ritardi operativi e la mancanza di massa critica hanno ridimensionato le aspettative iniziali, trasformando GAIA-X più in un quadro regolatorio che in un vero attore industriale.

Sul piano *hardware*, l'UE ha varato il *Chips Act* europeo, con l'obiettivo di raddoppiare la quota di produzione mondiale di semiconduttori entro il 2030 (dal 10% al 20%). Sono stati annunciati investimenti significativi: Intel costruirà nuovi impianti in Germania, mentre STMicroelectronics e GlobalFoundries hanno avviato progetti in Francia e in Italia. Nonostante ciò, la distanza rispetto a giganti come TSMC e Samsung rimane ampia, e l'Europa rischia di rimanere un mercato di consumo più che un produttore competitivo nei segmenti più avanzati.

Un altro elemento distintivo della strategia europea è l'attenzione agli aspetti normativi ed etici dell'AI. Con l'*AI Act*[57], Bruxelles punta a definire regole globali per lo sviluppo e l'u-

[57] Cfr. *Regulatory Framework for AI*, European Commission (2025).

so dell'intelligenza artificiale, promuovendo un modello basato sulla tutela dei diritti fondamentali e sulla trasparenza. Questa scelta riflette l'approccio tipico dell'UE: non potendo competere alla pari in termini di infrastrutture e risorse, cerca di esercitare *soft power* regolatorio, fissando standard che influenzino anche attori esterni.

Non mancano segnali positivi. Alcuni Stati membri stanno investendo in *data center* ad alte prestazioni e in consorzi pubblici-privati per condividere risorse computazionali. La rete EuroHPC (*European High Performance Computing Joint Undertaking*)[58], ad esempio, mira a mettere a disposizione *supercomputer* di ultima generazione a ricercatori e imprese europee. Anche la cooperazione accademica e i programmi di finanziamento comunitari (Horizon Europe, Digital Europe) contribuiscono a creare un tessuto di ricerca e innovazione.

In definitiva, l'Europa si trova di fronte a una contraddizione: è forte sul piano normativo e sulla domanda di servizi AI, ma debole sul piano infrastrutturale e industriale. Le sue ambizioni di autonomia digitale devono fare i conti con la realtà di una dipendenza strutturale dagli *hyperscaler* americani e dalla produzione asiatica di *chip*. La vera sfida per il prossimo decennio sarà trasformare iniziative frammentate in una strategia coerente, capace di costruire un ecosistema competitivo. In gioco non c'è solo il ruolo dell'UE nell'economia globale, ma la possibilità stessa di preservare una forma di sovranità tecnologica in un mondo dominato da poche superpotenze digitali.

5.3. Il resto del mondo: tra opportunità e dipendenze

Se Stati Uniti, Cina ed Europa rappresentano i poli principali della corsa all'intelligenza artificiale, gran parte del mondo si colloca in una posizione di utilizzatore e non di produttore di infrastrutture AI. Molti Paesi non hanno le risorse finanziarie, industriali o tecnologiche per costruire *data center* iperscalabili, sviluppare *chip* proprietari o garantire reti di interconnessione avanzate. Questo genera nuove forme di dipendenza digitale,

[58] Cfr. "Discover EuroHPC JU", *EuroHPC JU*, (2025), recuperato da https://www.eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en.

ma apre anche spazi di opportunità per chi riesce a posizionarsi come *hub* regionale o a valorizzare specifiche nicchie.

In Asia, alcuni Paesi si stanno ritagliando ruoli significativi. Il Giappone e la Corea del Sud hanno industrie avanzate nei semiconduttori e nelle telecomunicazioni, con Samsung e SK Hynix tra i principali fornitori globali di memoria e componenti[59]. Tuttavia, sul fronte degli *hyperscaler* e dei grandi modelli linguistici restano meno competitivi, spesso alleati o clienti delle piattaforme americane. L'India, invece, punta soprattutto sul capitale umano: con una vasta comunità di sviluppatori e ricercatori, si propone come polo di servizi e applicazioni AI, ma dipende dall'*hardware* e dalle piattaforme fornite da attori esterni.

Il Medio Oriente si sta muovendo con decisione, sfruttando le immense risorse finanziarie dei fondi sovrani[60]. L'Arabia Saudita e gli Emirati Arabi Uniti hanno lanciato programmi di investimento per attrarre *data center*, aziende *cloud* e laboratori di ricerca. In particolare, Abu Dhabi sta promuovendo *startup* come G42[61], con l'obiettivo di diventare un *player* riconosciuto nell'AI a livello globale. Tuttavia, la dipendenza da tecnologie straniere resta elevata, e il rischio è che queste ambizioni rimangano legate alla capacità di importare *know-how* e infrastrutture.

In Africa, la situazione è più frammentata. Da un lato, vi sono iniziative promettenti: in Kenya, Sudafrica e Nigeria nascono *hub* tecnologici sostenuti da investitori internazionali e da progetti di *hyperscaler* che aprono regioni *cloud* dedicate[62]. Dall'altro, il continente soffre ancora di un forte deficit infrastrutturale:

[59] Cfr. Aa.Vv. (2025), voce "Semiconductor industry in South Korea", in *Wikipedia*, recuperato da https://en.wikipedia.org/wiki/Semiconductor_industry_in_South_Korea; Aa.Vv. (2025), "Semiconductor industry in Japan", in *Wikipedia*, recuperato da https://en.wikipedia.org/wiki/Semiconductor_industry_in_Japan.

[60] Cfr. J. Kassab, J. Cloud (2025), "Gulf SWFs look to AI investments in diversification, competitiveness push", in *Deloitte*, recuperato da https://www.deloitte.com/global/en/Industries/investment-management/research/gulf-swfs-look-to-ai-investments.html.

[61] C. Cornish, "Abu Dhabi's G42 emerges as key AI player", in *Financial Times* (04/05/2025).

[62] "Digital Africa: Technological Transformation for Jobs", in *World Bank*, (2023), recuperato da https://www.worldbank.org/en/region/afr/publication/digital-africa.

la carenza di energia affidabile, la scarsa capillarità delle reti e la limitata disponibilità di *data center* moderni rendono difficile una piena partecipazione alla rivoluzione dell'AI.

Anche l'America Latina si trova in una posizione di dipendenza, pur essendo un mercato in rapida crescita per i servizi digitali. Brasile, Messico e Cile hanno attirato *data center* di *hyperscaler* americani e cinesi, che vedono nell'area un'opportunità di espansione[63]. Tuttavia, la maggior parte delle risorse computazionali resta concentrata in poche metropoli, e la capacità locale di sviluppare tecnologie AI autonome è ancora limitata.

Queste dinamiche contribuiscono a delineare un rischio concreto di "*AI divide*", un divario globale nell'accesso alle infrastrutture e alle competenze necessarie per l'intelligenza artificiale. Mentre poche aree – Nord America, Cina, Europa occidentale – concentrano *data center* iperscalabili, *chip* avanzati e *hyperscaler*, gran parte del mondo si trova costretta a consumare servizi prodotti altrove. Ciò potrebbe accentuare disuguaglianze economiche e geopolitiche, rendendo più difficile per i Paesi emergenti sfruttare appieno il potenziale trasformativo dell'AI.

Allo stesso tempo, esistono anche opportunità. La crescente domanda globale di infrastrutture AI può favorire alleanze regionali, investimenti congiunti e la nascita di *hub* specializzati. Alcuni Paesi cercano di posizionarsi come snodi strategici grazie alla loro geografia digitale: Singapore e Dubai, ad esempio, puntano a essere hub di interconnessione per i rispettivi continenti[64]. Altri valorizzano risorse energetiche rinnovabili a basso costo per attrarre *data center* (Islanda, Cile, Marocco)[65].

[63] Cfr. M. Menski, J. Johnson, A. de Feydeau (2024), "Latin America calls for accelerating data center infrastructure investment", in *White & Case*, recuperato da https://www.whitecase.com/insight-our-thinking/latin-america-focus-2024-data-center-infrastructure-investment.

[64] Cfr. *Digital Connectivity Blueprint*, IMDA, (2023); Cfr. S. Tng (2024), "Charting New Frontiers in the Digital Economy", in *URA*, recuperato da https://www.ura.gov.sg/Corporate/Resources/Ideas-and-Trends/Digital-Economy; Cfr. I. Ivanov (2023), "Digital Dubai", in *InterGlobix Magazine*, recuperato da https://www.interglobixmagazine.com/digital-dubai/.

[65] Cfr. A. Eljechtimi (2025), "Morocco plans to build a 500 MW data centre powered by renewable energy", in *Reuters*, recuperato da https://www.reuters.com/sustainability/boards-policy-regulation/morocco-build-data-centre-powered-by-renewable-energy-2025-07-09/; *Iceland data-center market driven by*

Il resto del mondo, pur rimanendo in posizione periferica rispetto a USA, Cina ed Europa, non è destinato a un ruolo passivo. Le traiettorie emergenti mostrano che, sebbene la dipendenza tecnologica sia oggi forte, esistono margini di manovra per costruire percorsi autonomi o almeno più equilibrati. La chiave sarà trasformare la domanda crescente di AI in un'occasione per sviluppare infrastrutture locali, evitando che la rivoluzione dell'intelligenza artificiale diventi un ulteriore fattore di polarizzazione globale.

6. Impatti e sfide globali

6.1. Sostenibilità ambientale

L'intelligenza artificiale avanzata non è solo una questione tecnologica o geopolitica: è anche un tema ambientale. L'addestramento e l'esecuzione di modelli di grandi dimensioni richiedono enormi risorse computazionali, che si traducono in consumi energetici e idrici senza precedenti. I *data center* che ospitano tali infrastrutture sono quindi al centro di un dibattito cruciale: come far crescere l'AI senza compromettere la sostenibilità?

Il consumo energetico è la variabile più critica. L'addestramento di un singolo LLM può richiedere centinaia di MWh[66], pari al fabbisogno annuo di decine di famiglie. Una volta messi in produzione, i modelli generativi restano energivori: l'inferenza – cioè le risposte agli utenti – genera consumi continui, soprattutto se servita a milioni di persone in simultanea[67].

Secondo stime recenti, i *data center* globali potrebbero presto assorbire oltre il 5% dell'elettricità mondiale, con una quota

renewable energy and robust connectivity, GlobeNewswire, (2025), consultabile in https://www.globenewswire.com/news-release/2025/07/29/3123346/28124/en/Iceland-Data-Center-Market-Investment-Analysis-2025-2030-with-Notable-Players-AtNorth-Borealis-Data-Center-and-Verne.html.

[66] Cfr. N. Jegham *et al.*, "How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference", in *arXiv:2505.09598*, (2025).

[67] Cfr. R. Desislavov, F. Martínez-Plumed, J. Hernández-Orallo, "Trends in AI inference energy consumption", in *ScienceDirec*t, 38 (2023), pp. 1-7.

crescente dovuta all'AI[68]. Questa dinamica rischia di confliggere con le politiche di decarbonizzazione.

Meno visibile ma altrettanto rilevante è l'impatto sull'acqua. Il raffreddamento delle GPU richiede grandi quantità di acqua per mantenere le temperature operative. In alcune regioni i consumi idrici sono già oggetto di conflitto con le comunità locali, soprattutto dove la disponibilità è scarsa[69]. Alcuni studi stimano che l'addestramento di un singolo modello possa richiedere milioni di litri di acqua dolce, considerando sia il raffreddamento diretto che la produzione energetica a monte[70].

La localizzazione dei *data center* diventa quindi strategica. Paesi con abbondanza di energie rinnovabili e risorse idriche (come Islanda, Norvegia e Svezia) stanno diventando *hub* privilegiati per l'AI. Al contrario, costruire grandi infrastrutture in regioni calde o aride aumenta costi e rischi ambientali e sociali.

Per ridurre l'impatto, l'industria sperimenta diverse soluzioni:

i. raffreddamento a immersione, che riduce uso di acqua e migliora l'efficienza;

ii. ottimizzazione del *workload* AI tramite *pruning*[71] e quantizzazione[72];

iii. contratti di lungo termine per energie rinnovabili (PPA);

iv. *edge computing*, che sposta parte dell'elaborazione vicino all'utente, alleggerendo gli *hub* iperscalabili.

Nonostante queste innovazioni, la sfida resta aperta. Governi e imprese vedono nell'AI una leva di crescita, mentre società civile e autorità ambientali chiedono coerenza con gli obiettivi

[68] Cfr. "Energy and AI", in *IEA*, (2025), recuperato da https://www.iea.org/reports/energy-and-ai.

[69] Cfr. M.G. Pascual, "Data centers are drying up the port of Marseille", in *El País* (24/08/2025).

[70] Cfr. P. Li *et al.*, "Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models", in *arxiv.org/abs/2304.0327*, (2023).

[71] Cfr. Aa.Vv. (2025), voce "Pruning (machine learning)", in Wikipedia, recuperato da https://en.wikipedia.org/wiki/Pruning_(artificial_neural_network)

[72] Cfr. B. Jacob *et al*., "Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference", in *arxiv.org/abs/1712.05877*, (2017).

climatici. Senza un cambio di paradigma, l'espansione dei *data center* rischia di compromettere gli impegni di neutralità climatica fissati al 2050.

L'AI rappresenta così un paradosso: strumento per ottimizzare consumi e ridurre sprechi, ma al tempo stesso fonte di nuove pressioni ecologiche. La vera sfida sarà rendere i *data center* non solo più potenti, ma anche sostenibili, tramite innovazione, regolazione e pianificazione attenta della loro localizzazione.

6.2. *Governance* e regolamentazione

Lo sviluppo dell'intelligenza artificiale non è solo una questione di infrastrutture tecniche, ma anche di regole. Governi, istituzioni sovranazionali e organizzazioni internazionali sono chiamati a definire un quadro di *governance* capace di garantire sicurezza, trasparenza e tutela dei diritti fondamentali, senza frenare la spinta all'innovazione. Il nodo è particolarmente delicato: se da un lato le grandi piattaforme spingono per un approccio flessibile, dall'altro l'opinione pubblica e le autorità richiedono maggiore controllo.

Tre sono le aree più critiche:

i. Sicurezza: i modelli AI possono generare contenuti falsi (*deepfake*[73]), automatizzare attacchi informatici o essere usati in contesti militari. Senza regole chiare, il rischio è che la tecnologia amplifichi vulnerabilità e conflitti.

ii. Privacy: l'addestramento di LLM su grandi moli di dati solleva interrogativi su come vengano raccolte e utilizzate le informazioni personali. La possibilità di "ricostruire" dati sensibili da modelli apparentemente anonimi è un pericolo concreto.

iii. Etica: *bias* e discriminazioni nei *dataset* possono portare a risultati ingiusti, influenzando decisioni su credito, lavoro, giustizia. Il tema etico non riguarda solo i dati, ma anche la concentrazione di potere in poche mani.

La *governance* dell'AI si muove su più livelli:

i. Stati nazionali: hanno iniziato a introdurre strategie e normative *ad hoc*. Gli Stati Uniti puntano su linee guida e standard

[73] Cfr. *Science & Tech Spotlight: Deepfakes*, GAO (U.S. Government Accountability Office) (2020).

volontari, mentre la Cina inserisce l'AI all'interno di un quadro normativo centralizzato e orientato al controllo.

ii. Unione Europea: con l'*AI Act*, ha adottato l'approccio più organico, classificando i sistemi in base al livello di rischio (basso, medio, alto, inaccettabile) e introducendo obblighi di trasparenza e conformità. Bruxelles mira a replicare il "modello GDPR": usare il mercato europeo come leva per fissare standard globali.

iii. Organizzazioni internazionali: iniziative come l'*OECD AI Principles*[74], i lavori del G7 Hiroshima AI Process[75] o le raccomandazioni dell'UNESCO[76] riflettono il tentativo di costruire un linguaggio comune, ma spesso mancano strumenti vincolanti.

La mancanza di coordinamento rischia di produrre una frammentazione regolatoria: ogni Paese o area geopolitica stabilisce regole proprie, creando barriere all'interoperabilità e incertezza per le imprese.

Uno dei nodi più complessi riguarda il rapporto tra regolamentazione e capacità innovativa.

i. Una regolamentazione troppo rigida e preventiva può soffocare le *startup* e spingere le imprese a spostarsi verso contesti più permissivi.

ii. Una regolamentazione troppo lenta o permissiva, invece, rischia di lasciare spazio a comportamenti irresponsabili, con conseguenze gravi per cittadini e istituzioni.

Trovare l'equilibrio significa adottare un approccio dinamico: regole flessibili, aggiornabili, che evolvano insieme alla tecnologia. Alcuni propongono modelli di *regulatory sandbox*, spazi di sperimentazione controllata che permettono di testare applicazioni senza esporre i cittadini a rischi.

La *governance* dell'AI è oggi uno degli snodi strategici del futuro digitale. Non basta costruire reti, *data center* e *chip*: senza

[74] Cfr. *Recommendation of the Council on Artificial Intelligence*, OECD (2019).

[75] Cfr. *G7 Leaders' Statement on the Hiroshima AI Process*, European Commission (2023), consultabile in https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process.

[76] Cfr. *Recommendation on the Ethics of Artificial Intelligence*, UNESCO (2021), consultabile in https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

regole chiare, condivise e proporzionate, l'AI rischia di amplificare disuguaglianze, minacce alla *privacy* e squilibri geopolitici. Allo stesso tempo, un eccesso di regolazione potrebbe relegare interi Paesi a semplici utilizzatori passivi, lasciando l'innovazione a chi dispone di mercati più flessibili. La sfida, per governi e istituzioni, sarà quella di coniugare sicurezza e diritti con la capacità di competere in un ecosistema globale che evolve con rapidità straordinaria.

6.3. Accesso diseguale e nuove disuguaglianze

Se le infrastrutture dell'intelligenza artificiale diventano il nuovo motore della competitività globale, l'accesso a esse rischia di essere distribuito in modo profondamente diseguale. La disponibilità di potenza di calcolo, di grandi *dataset* e di *data center* iperscalabili è oggi concentrata in poche aree del mondo – Stati Uniti, Cina, Europa occidentale – lasciando molti Paesi senza risorse sufficienti per partecipare alla rivoluzione dell'AI da protagonisti[77].

In gran parte dell'Africa, dell'America Latina e dell'Asia meridionale, mancano infrastrutture di base come reti elettriche affidabili, *data center* moderni e connessioni a banda larga diffuse. Questo *deficit* si traduce in una dipendenza quasi totale dai servizi offerti da *hyperscaler* stranieri. Le *startup* e i centri di ricerca locali sono costretti a utilizzare piattaforme *cloud* estere, con costi elevati e minore controllo sui dati. In queste condizioni, diventa difficile sviluppare modelli AI addestrati su dati pertinenti al contesto locale, riducendo la capacità di innovare in modo autonomo.

Questa situazione apre la strada a un "*AI divide*", un divario globale nell'accesso e nell'uso dell'intelligenza artificiale paragonabile, ma potenzialmente più grave, del *digital divide* degli anni 2000. Se allora la sfida era portare la connettività Internet nelle scuole e nelle imprese, oggi il problema è garantire l'accesso a risorse computazionali avanzate e a *dataset* rappresentativi. Senza tali strumenti, interi Paesi rischiano di rimanere meri

[77] Cfr. A. Abdullahi (2025), "Only 32 countries host AI data centers", in *TechRepublic*, recuperato da https://www.techrepublic.com/article/news-only-32-countries-host-data-centers-2025/.

consumatori di soluzioni AI sviluppate altrove, con modelli che incorporano valori, *bias* e priorità non sempre compatibili con i contesti locali.

Il rischio non è solo economico, ma anche sociale e politico. Un divario nell'accesso all'AI può accentuare le disuguaglianze tra cittadini, rafforzare la dipendenza tecnologica dei Paesi emergenti e limitare la loro capacità di esercitare sovranità digitale. La concentrazione del potere computazionale in poche aree potrebbe trasformarsi in una nuova forma di egemonia geopolitica, dove l'accesso all'AI diventa un privilegio e non un diritto.

Per affrontare queste disparità, sono in corso alcune iniziative:

i. Organizzazioni internazionali: l'ONU e l'UNESCO hanno avviato programmi per promuovere l'uso responsabile dell'AI nei Paesi in via di sviluppo, mettendo l'accento su inclusione e diversità culturale[78].

ii. Collaborazioni accademiche: reti di università e centri di ricerca stanno creando consorzi per condividere risorse computazionali e *dataset*, offrendo accesso a comunità che altrimenti ne sarebbero escluse[79].

iii. Progetti regionali: in Africa, l'*African AI Research Network*[80] mira a sviluppare modelli e competenze locali, mentre in America Latina alcuni governi discutono di creare piattaforme *cloud* pubbliche per ridurre la dipendenza da attori esterni[81].

iv. Iniziative private: alcuni *hyperscaler*, sotto pressione politica e sociale, hanno iniziato a offrire crediti *cloud* o accesso agevolato a istituzioni educative e *startup* dei Paesi emergenti, pur con evidenti limiti e rischi di *lock-in*[82].

[78] Cfr. *Recommendation on the Ethics of Artificial Intelligence*, cit.; *Global Digital Compact – AI for Sustainable Development*, United Nations (2023), consultabile in https://www.un.org/digital-emerging-technologies/global-digital-compact.

[79] Cfr. *Discover EuroHPC JU*, cit.

[80] Cfr. "Wits leads ambitious partnership to drive AI in Africa", in *University of the Witwatersrand News* (2021), consultabile in https://www.wits.ac.za/news/latest-news/research-news/2021/2021-09/wits-leads-ambitious-partnership-to-drive-ai-in-africa.html.

[81] Cfr. "Gestão, Serpro e Dataprev lanciano servizi cloud governativi per le agenzie federali", *Governo Digitale del Brasile*, (2025).

[82] Cfr. F. Manoharan (2022), "Google Cloud expands higher education credits to 8 countries in Africa", in *Google Cloud Blog*, recuperato da https://

Il rischio di un *AI divide globale* è oggi concreto e potrebbe determinare le traiettorie dello sviluppo mondiale nei prossimi decenni. Senza un accesso equo alle infrastrutture di calcolo e ai *dataset*, molti Paesi resteranno spettatori passivi della rivoluzione dell'AI, con conseguenze in termini di crescita economica, coesione sociale e autonomia politica.

L'obiettivo non è solo distribuire meglio le risorse, ma costruire un ecosistema AI più inclusivo, in cui la diversità culturale e geografica sia valorizzata e non esclusa. Se l'AI diventerà davvero un'infrastruttura critica della società del XXI secolo, allora garantirne un accesso equo dovrà essere considerato non un lusso, ma una priorità politica globale.

7. Conclusioni

L'intelligenza artificiale, spesso percepita come *software* immateriale, è in realtà un bene infrastrutturale. La sua esistenza dipende da una catena materiale composta da reti, *data center*, *chip*, energia e interconnessioni. Non è solo algoritmo e *cloud*, ma un sistema complesso che ricorda le grandi infrastrutture del passato: l'elettricità, le ferrovie, Internet.

Come quelle, l'AI è al tempo stesso motore di innovazione e leva geopolitica. Chi controlla le risorse critiche – capacità di calcolo, semiconduttori, piattaforme *cloud*, *hub* di interconnessione – acquisisce un potere che si riflette non solo sull'economia, ma anche sull'equilibrio politico e sociale globale.

Per l'Europa, il nodo centrale è quello della sovranità. L'assenza di grandi *hyperscaler* e di una produzione autonoma di *chip* avanzati rende il continente dipendente da Stati Uniti e Cina. Da qui l'urgenza di sviluppare una propria sovranità digitale, capace di garantire controllo sui dati, sicurezza e indipendenza tecnologica.

Le iniziative europee vanno in questa direzione: l'*AI Act* come cornice regolatoria, l'*European Chips Act* per rafforza-

cloud.google.com/blog/topics/public-sector/google-cloud-expands-higher-education-credits-8-countries-africa/; Cfr. P. Sandle (2025), "Dominance of Amazon and Microsoft in cloud harming competition, UK says", in *Reuters*, recuperato da https://www.reuters.com/legal/litigation/dominance-amazon-microsoft-cloud-harming-competition-uk-says-2025-07-31/.

re la produzione di semiconduttori, i progetti di *cloud* federato come Gaia-X per ridurre la dipendenza dagli attori extraeuropei. L'obiettivo è duplice: preservare i valori europei – tutela della *privacy*, diritti fondamentali, concorrenza – e costruire capacità industriali proprie.

In questo contesto, anche le infrastrutture di interconnessione – inclusi gli *Internet Exchange Point* (IXP) – assumono un ruolo complementare. Non sono al centro della strategia, ma contribuiscono a rendere l'ecosistema più resiliente ed efficiente, riducendo dipendenze e migliorando la connettività tra operatori e *data center*.

Chip, *data center*, reti, regole e interconnessioni sono legati in modo inscindibile. Un *deficit* in uno solo di questi ambiti indebolisce l'intero sistema. Un algoritmo senza calcolo è inutile; un *data center* senza energia sostenibile è insostenibile; una rete senza interconnessioni dense fatica a servire gli utenti in modo efficiente; e senza regole condivise il tutto rischia di produrre nuove disuguaglianze.

Oltre alla dimensione tecnologica e geopolitica, l'AI porta con sé sfide profonde per la società. L'automazione di processi decisionali – dal credito al lavoro, dalla sanità alla sicurezza – solleva interrogativi su trasparenza, equità e responsabilità. Alcuni ambiti sono particolarmente delicati: pensiamo all'uso dell'AI nella sanità per la diagnosi precoce, nella sicurezza pubblica per il riconoscimento facciale o, in prospettiva, nel procedimento penale per supportare valutazioni legate a prove o rischi di recidiva. In tutti questi casi, la tecnologia può offrire efficienza e supporto, ma apre questioni cruciali di trasparenza, imparzialità e tutela dei diritti fondamentali, che richiedono cautele e garanzie specifiche.

I possibili benefici dell'AI – maggiore efficienza, nuove opportunità economiche, innovazioni in settori chiave – devono quindi essere bilanciati con la capacità di gestire i rischi: manipolazione dell'informazione, sorveglianza di massa, erosione della *privacy* e possibili derive incompatibili con i principi dello Stato di diritto.

Il futuro dell'AI dipenderà dalla capacità di sviluppare strategie cooperative: tra governi, per evitare frammentazioni re-

golatorie; tra imprese e istituzioni, per bilanciare innovazione e sicurezza; tra regioni del mondo, per ridurre le disuguaglianze di accesso.

Per l'Europa la sfida è duplice: colmare i *gap* rispetto a Stati Uniti e Cina, e al tempo stesso proporre un modello alternativo che unisca competitività industriale, sostenibilità ambientale, inclusione sociale ed etica.

Se l'Europa saprà muoversi in questa direzione, potrà giocare un ruolo da protagonista, contribuendo a un ecosistema AI che non sia monopolio di pochi, ma patrimonio condiviso, capace di migliorare società ed economie senza sacrificare valori fondamentali.

Francesca Alessandra Lisi

Intelligenza Artificiale Affidabile: dalle applicazioni in ambito forense alla simbiosi fra umani e macchine intelligenti

Abstract: **Trustworthy Artificial Intelligence: From Forensic Applications to the Symbiosis Between Humans and Intelligent Machines.** In this chapter, the topic of AI trustworthiness is addressed by analyzing the peculiarities of two application areas relevant to the volume: Digital Forensics (more specifically, the task of Evidence Analysis) and Human-AI Symbiosis whose legal and ethical acceptability is in the focus of attention. The discussion draws on case studies from AI research projects (DigForASP and FAIR), which highlight the need to adopt technical solutions that meet the requirements for trustworthy AI, particularly transparency, explainability, and human oversight.

Keywords: Trustworthy Artificial Intelligence, Digital Forensics, Evidence Analysis, Human-AI Symbiosis, AI Ethics, Risk Assessment.

Sommario: 1. Affidabilità dell'Intelligenza Artificiale – 2. Fidarsi dell'IA in ambito forense – 3. Fidarsi dell'IA in simbiosi con gli umani – 4. Considerazioni conclusive.

1. Affidabilità dell'Intelligenza Artificiale

L'Intelligenza Artificiale (IA) è l'ultima evoluzione tecnologica che sta trasformando l'economia globale, svolgendo un ruolo fondamentale nella "Quarta Rivoluzione Industriale" e cambiando profondamente la nostra società. In quanto tale, porta con sé le cosiddette *tecnologie dirompenti* (*disruptive technologies*), ovvero quelle tecnologie che hanno il potenziale di sostituire un processo, un prodotto o un'abitudine più datati. Per svelare come queste tecnologie possano avere un impatto trasformativo a livello sociale ed economico e per identificare i potenziali svantaggi della "*disruption*" che potrebbero comportare, è necessario adottare un approccio responsabile nei loro confronti.

Le tecnologie dirompenti introdotte dall'IA, sebbene applicate con successo in un numero crescente di contesti, si sono ri-

velate affette da un problema importante: la mancanza di fiducia da parte di chi le adotta, dovuta a diverse carenze nel modo in cui queste tecnologie vengono progettate, sviluppate e applicate in ambiti rischiosi. Come esseri umani, tendiamo a fidarci di più dei sistemi a scatola bianca (*white-box*) rispetto a quelli a scatola nera (*black-box*). Appartengono a quest'ultima categoria molte delle applicazioni di algoritmi di *machine learning/deep learning* che quindi sono la principale fonte di preoccupazione.

A livello UE, il problema è stato affrontato dal Gruppo Indipendente di Esperti ad Alto Livello sull'Intelligenza Artificiale, istituito dalla Commissione Europea nel Giugno 2018. Il Gruppo ha elaborato un documento contenente delle Linee Guida per un'IA Affidabile (*Trustworthy AI*)[1] secondo cui l'IA merita fiducia se è: 1) *lecita*, ovvero non infrange le leggi e i regolamenti applicabili; 2) *etica*, ovvero non viola i principi e i valori etici; 3) *robusta*, ovvero non causa danni involontari agli individui e alla società. Il documento non tratta esplicitamente la prima componente dell'IA affidabile (liceità), ma offre piuttosto indicazioni per promuovere e garantire la seconda e terza componente (eticità e robustezza). La liceità, infatti, è stata successivamente coperta dalla regolamentazione europea in materia, nota come l'AI Act[2].

Le Linee Guida stabiliscono sette requisiti chiave che i sistemi di IA devono soddisfare per essere considerati affidabili: 1) sorveglianza umana (*human oversight*), 2) robustezza tecnica e sicurezza (*technical robustness and safety*), 3) riservatezza e governance dei dati (*privacy and data governance*), 4) trasparenza (*transparency*), 5) diversità, non discriminazione ed equità (*diversity, non-discrimination and fairness*), 6) benessere sociale e ambientale (*societal and environmental well-being*) e 7) responsabilità (*accountability*). Fra i diversi requisiti possono esistere delle tensioni fondamentali di cui essere consapevoli. Occorre quindi individuare, valutare, documentare e comuni-

[1] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, Technical Report, European Commission, 2019 (consultabile in: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[2] Artificial Intelligence Act (consultabile in: https://artificialintelligenceact.eu/).

care costantemente le soluzioni di compromesso, e coinvolgere tutti i portatori di interesse. Il primo ed il quarto requisito sono particolarmente rilevanti per questo capitolo, e pertanto saranno brevemente illustrati di seguito, rimandando al documento europeo per una trattazione più approfondita di questi e degli altri requisiti in elenco.

Il requisito della sorveglianza umana richiede che i sistemi di IA sostengano il processo decisionale umano, come prescrive il *principio del rispetto dell'autonomia umana*. A tal fine essi devono agire come catalizzatori di una società democratica, prospera ed equa, sostenendo l'intervento degli utenti e promuovendo i diritti fondamentali, e devono consentire la sorveglianza da parte degli umani. Il requisito di trasparenza invece è strettamente connesso al *principio dell'esplicabilità*. Esso richiede che i dati, i sistemi e i modelli di business dell'IA siano trasparenti. I meccanismi di tracciabilità possono contribuire a raggiungere questo obiettivo. Inoltre, i sistemi di IA e le loro decisioni dovrebbero essere spiegati in modo adeguato alle parti interessate. Infine, grazie ad una comunicazione efficace, gli esseri umani devono essere resi consapevoli di interagire con un sistema di IA e devono essere informati delle capacità e dei limiti del sistema.

L'affidabilità è una caratteristica imprescindibile delle soluzioni tecnologiche in diversi ambiti applicativi. Nel contesto di questo volume, due sono gli ambiti di interesse che si intende esplorare. Il primo riguarda le indagini investigative che si avvalgono delle metodologie dell'Informatica Forense (vedi Sezione 2). Il secondo considera quelle applicazioni in cui l'interazione fra umani e IA è talmente stretta da configurarsi come simbiosi (vedi Sezione 3). Un caso di studio viene presentato e discusso per ciascuno di questi ambiti al fine di illustrare le potenzialità delle soluzioni tecnologiche offerte dall'IA Affidabile. Per l'ambito forense, si tratta dell'analisi di tabulati telefonici. La simbiosi, invece, viene analizzata nel contesto medico mediante un tipico dilemma etico che può presentarsi nell'uso della robotica assistiva per il monitoraggio dei pazienti.

2. Fidarsi dell'IA in ambito forense

L'Informatica Forense (*Digital Forensics*, DF) è la disciplina che si occupa della identificazione, conservazione, protezione, estrazione, documentazione e qualsiasi altra forma di elaborazione di dati digitali a scopi forensi[3]. I dati possono essere valutati come prova in ambito penale o civile dopo essere stati trasformati in contenuto informativo mediante strumenti *software* appositi. L'Informatica Forense è quindi una branca delle Scienze Forensi le cui attività sono volte a evidenziare l'esistenza di prove digitali rilevanti per l'adempimento di un processo, o anche prima delle fasi investigative. Si è rapidamente ritagliato un ampio spazio anche in ambito aziendale, con l'obiettivo di evidenziare violazioni del segreto industriale e/o violazione delle *policy* di sicurezza e di gestire incidenti informatici. Inoltre, essa è caratterizzata da tecniche in continua evoluzione e cambiamento, in linea con il rapido processo di innovazione nel settore ICT, che coinvolge metodi di analisi, strumenti, protocolli e componenti strutturali relativi ai sistemi digitali analizzati, nonché diritti a tutela della *privacy* dell'individuo.

La continua evoluzione dei dispositivi digitali e la grande quantità di dati disponibili richiedono strumenti efficienti ed efficaci per analizzare tale mole di dati. Inoltre, verificabilità, affidabilità e giustificabilità sono caratteristiche chiave per gli strumenti *software* da applicare in ambito forense, dove le prove prodotte sono finalizzate alla ricostruzione dei reati e assistono/facilitano il tribunale nel processo decisionale per stabilire se un imputato sia innocente o colpevole. Per questi motivi, la COST Action CA17124 "*Digital forensics: evidence analysis via intelligent systems and practices*" (DigForASP)[4] mirava a creare un'infrastruttura di ricerca per l'applicazione di IA nel campo dell'Informatica Forense, in particolare nella fase di Analisi delle Prove (*Evidence Analysis*, EA)[5].

[3] F. Casino, T.K. Dasaklis, G. Spathoulas, M. Anagnostopoulos, A. Ghosal, I. Borocz, A. Solanas, M. Conti, C. Patsakis, "Research trends, challenges, and emerging topics in digital forensics: A review of reviews", in *IEEE Access*, 10 (2022), pp. 25464-25493, doi: 10.1109/ACCESS.2022.3154059.

[4] DigForASP (consultabile in: https://digforasp.uca.es/)

[5] S. Costantini, F.A. Lisi, R. Olivieri, "DigForASP: A European cooperation network for logic-based AI in digital forensics", in A. Casagrande,

Svariate tecniche di IA sono state applicate in ambito forense per scopi diversi. Sono state sfruttate principalmente per il recupero e la categorizzazione dei dati, ad esempio per l'analisi di immagini e video multimediali o per il rilevamento di anomalie. Questi compiti, che tipicamente richiedono l'uso di modelli predittivi, traggono vantaggio in particolare dall'applicazione di algoritmi di Apprendimento Automatico (*Machine Learning*, ML). Il progetto DigForASP compie un passo avanti, poiché promuove invece l'applicazione di metodi di Rappresentazione della Conoscenza (*Knowledge Representation*, KR) e Ragionamento Automatico (*Automated Reasoning*, AR) per recuperare dati da cui ottenere prove utilizzabili in un processo. Ad esempio, dai dati recuperati da diverse fonti (come, ad esempio, dispositivi mobili, attività sui *social network*, tracce di *cloud computing*, tabulati telefonici, file di log, ecc.), possiamo ottenere l'insieme di tutti i possibili schemi di attività di un sospettato durante la commissione di un crimine.

Particolare enfasi nel progetto è stata riservata ad *Answer Set Programming* (ASP)[6], che segue un approccio dichiarativo alla risoluzione di problemi (*declarative problem solving*). In questo approccio, alternativo a quello cosiddetto procedurale, si specifica quali sono le proprietà di cui deve godere la/le soluzione/i piuttosto che dettagliare la procedura (cioè la sequenza di passi di computazione) necessaria per la risoluzione del problema. A tal scopo, come illustrato in Figura 1, il problema viene modellizzato per ottenerne una rappresentazione (ovvero una codifica) che sia comprensibile ad un risolutore automatico (*solver*), cioè ad un programma il cui *output* di esecuzione sia interpretabile come soluzione al problema originario. Esistono quindi risolutori per ASP, come Clingo[7], che generano soluzioni candidate e di queste

E.G. Omodeo (eds.): *Proceedings of the 34th Italian Conference on Computational Logic, Trieste, Italy, June 19-21, 2019*, CEUR Workshop Proceedings 2396, CEUR-WS.org 2019 (https://ceur-ws.org/Vol-2396/paper34.pdf).

[6] V. Lifschitz, "Answer sets and the language of answer set programming", in *AI Magazine*, 37 (2016), pp. 7-12, doi: 10.1609/aimag.v37i3.2670.

[7] M. Gebser, R. Kaminski, B. Kaufmann, T. Schaub, *Answer Set Solving in Practice*. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Morgan & Claypool Publishers, San Rafael (CA), 2012.

restituiscono come risposte (*answer sets*) solo quelle che soddisfano le proprietà desiderate espresse sotto forma di vincoli.

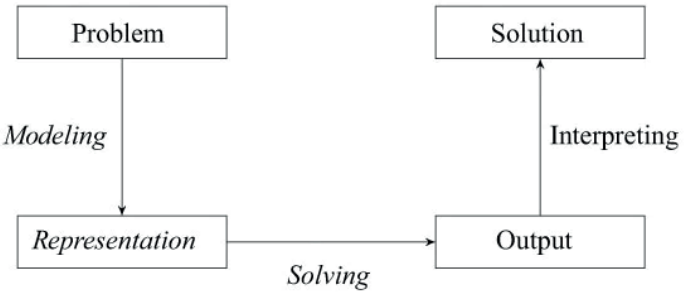

Figura 1: Fasi nella risoluzione di problemi con approccio dichiarativo (Declarative Problem Solving)

Essendo un approccio basato sulla logica, ASP consente di sviluppare sistemi inerentemente trasparenti, il che lo rende uno strumento efficace per la realizzazione di un'IA affidabile. Nell'ambito dell'informatica forense, Costantini *et. al*.[8] hanno esplorato l'applicabilità di ASP per automatizzare parzialmente le attività nella fase di analisi delle evidenze. Essi hanno mostrato come indagini complesse e significative, difficilmente risolvibili da esperti umani, possano essere formalizzate come problemi di ottimizzazione, una classe di problemi in cui ASP ha dimostrato di essere un valido risolutore.

Nell'ambito del progetto DigForASP abbiamo sviluppato speciali strumenti con ASP che estraggono automaticamente informazioni utili dai tabulati telefonici, riducendo i tempi di lavoro. Infatti, le prove su possibili reati e autori di reati raccolte da vari dispositivi elettronici (tramite *software* specializzati e secondo specifiche normative) devono essere esaminate e aggregate al fine di ricostruire possibili eventi, sequenze di eventi e scenari correlati a un reato.

[8] S. Costantini, G. De Gasperis, R. Olivieri, "Digital forensics and investigations meet artificial intelligence", in *Annals of Mathematics and Artificial Intelligence*, 86 (2019), pp. 193-229, doi: 10.1007/s10472-019-09632-y.

Caso di studio: analisi dei tabulati telefonici

L'ampia diffusione della telefonia cellulare sta trasformando le persone in identità mobili tracciabili. Le informazioni vengono automaticamente memorizzate nei tabulati telefonici e nella memoria del dispositivo, una fonte di prova inestimabile per i procedimenti penali. Attraverso i tabulati telefonici non è possibile risalire al contenuto delle chiamate audio, dei messaggi e delle e-mail inviate o ricevute, né all'elenco della cronologia web dei siti visitati. Si tratta di un insieme di dati relativi alle comunicazioni esterne dei dispositivi, che possono essere richiesti dall'Autorità Giudiziaria qualora ritenga utile ottenerli per svolgere indagini sul singolo titolare dell'utenza. I tabulati telefonici contengono tutte le tracce digitali delle comunicazioni relative a uno specifico utente in un determinato periodo di tempo, come – per esempio – il numero di telefono del chiamante, il numero di telefono del destinatario, la categoria di comunicazione (chiamata, SMS, chiamata persa, ecc.) e la durata della chiamata. Come preziosa fonte di prove, le forze dell'ordine possono analizzare questi dati per confrontare le posizioni geografiche rispetto alle dichiarazioni e ricostruire la rete di contatti di un singolo utente per tracciare quali conversazioni ha avuto e in quale momento.

L'analisi dei dati telefonici può richiedere molto tempo. Pertanto, l'automazione (parziale) di questa attività limita la quantità di dati da elaborare durante la fase di analisi delle evidenze. Con la nostra ricerca, abbiamo incoraggiato l'utilizzo di metodi e tecniche di IA formali e verificabili per l'analisi dei *pattern* di chiamata, una tecnica di *intelligence* utilizzata per identificare regolarità nel traffico telefonico, come la tempistica di eventi e azioni, possibili correlazioni causali e contesti in cui si sono verificate azioni sospette. Come detto in precedenza, abbiamo preso in considerazione l'analisi dei tabulati telefonici reali, accuratamente resi anonimi per preservare la *privacy* e la riservatezza, forniti dal Prof. David Billard (esperto di Informatica Forense dell'Università di Scienze Applicate di Ginevra) ai membri di DigForASP in base a un accordo di non divulgazione per la sperimentazione accademica[9]. Il *dataset* è composto da quattro file

[9] F.A. Lisi, G. Sterlicchio, D. Billard, "A Case Study for Declarative Pattern Mining in Digital Forensics", in A. Fensel, A. Ozaki, D. Roman,

Excel, uno per ciascun sospettato: "Eudokia Makrembolitissa", "Karen Cook McNally", "Laila Lalami", "Lucy Delaney". Ogni file presenta le seguenti caratteristiche:

- Tipo: chiamata o SMS in entrata/uscita;
- Chiamante: chi effettua la chiamata/SMS;
- Chiamato: chi riceve la chiamata/SMS;
- Via: luogo dell'operazione;
- Ora: quando è avvenuta l'operazione;
- Durata: quanto è durata l'operazione;
- Data: giorno, mese, anno.

Il nostro obiettivo era rispondere alle domande poste dal Prof. Billard e da altri esperti di Informatica Forense nell'ambito del progetto DigForASP, come ad esempio:

1. Dal tabulato di Eudokia Makrembolitissa, sarebbe possibile trovare le sue complici Karen Cook McNally e/o Laila Lalami?
2. Dai tabulati di Eudokia Makrembolitissa, Karen Cook McNally e Laila Lalami, sarebbe possibile trovare Lucy Delaney?
3. Le stesse persone si riuniscono spesso fisicamente?
4. Quando X chiama Y, Y chiama sempre Z poco dopo?
5. Al momento del crimine, chi si trovava nello stesso luogo o è stato chiamato da Eudokia Makrembolitissa?
6. Il giorno prima, chi ha parlato con Eudokia Makrembolitissa? O con qualche altro sospettato?

A tal fine, abbiamo proposto di analizzare il *dataset* mediante algoritmi di IA in grado di scoprire regolarità (*pattern mining*) che potessero evidenziare le abitudini dei sospettati. In particolare, considerata la natura del *dataset*, ci siamo concentrati su due tipologie di *pattern*, quelli sequenziali e quelli di contrasto. In breve, il problema del *sequential pattern mining* consiste nel trovare sequenze temporali frequenti e non vuote da un dataset di sequenze[10]. Invece, il problema del *contrast pattern mining* ri-

A. Soylu (eds.), *Rules and Reasoning - 7th International Joint Conference, RuleML+RR 2023, Oslo, Norway, September 18-20, 2023, Proceedings*, Lecture Notes in Computer Science 14244, Springer, Cham, 2023, doi: 10.1007/978-3-031-45072-3_12.

[10] C. Mooney, J.F. Roddick, "Sequential pattern mining - approaches and algorithms", in *ACM Computing Surveys*, 45 (2013), n. 19, pp. 1-39, doi: 10.1145/2431211.2431218.

guarda il rilevamento di differenze statisticamente significative (contrasto) tra due o più dataset disgiunti[11] (oppure fra porzioni di uno stesso dataset). La novità del nostro lavoro è l'approccio dichiarativo (e quindi, trasparente), basato appunto su ASP. Di seguito, a scopo illustrativo, si riporta qualche risultato ottenuto per ciascuno dei due problemi di *pattern mining*.

Come detto in precedenza, il *sequential pattern mining* può ragionare su sequenze di eventi in un dato intervallo di tempo. Noi abbiamo codificato questo problema in ASP sfruttando le informazioni temporali per creare la rete di relazioni dei sospettati, per identificare associazioni tra individui ed evidenziare le abitudini o "stili di vita" dei sospettati[12]. Ad esempio, il nostro programma in ASP ha scoperto un *pattern* sequenziale composto da due eventi di comunicazione nel tabulato telefonico di Karen Cook McNally[13]. Il primo evento riguarda Margaret Hasse e Karen (come destinataria), mentre il secondo coinvolge Karen e Lucie Julia (come destinataria). Questo *pattern* occorre tre volte nel tabulato di Karen, precisamente nei giorni 8, 9 e 12 settembre 2040, come illustrato graficamente in Figura 2. Si noti come ogni occorrenza del pattern renda esplicita la tipologia di comunicazione nei vari eventi. Per esempio, il primo evento è un SMS in data 8 e 9 settembre mentre è una chiamata in data 12 settembre.

[11] G. Dong, J. Bailey, *Contrast data mining: concepts, algorithms, and applications*, CRC Press, Boca Raton, 2012.

[12] F.A. Lisi, G. Sterlicchio, "Declarative pattern mining in digital forensics: Preliminary results", in R. Calegari, G. Ciatto, A. Omicini (eds.), *Proceedings of the 37th Italian Conference on Computational Logic, Bologna, Italy, June 29 - July 1, 2022*, CEUR Workshop Proceedings 3204, CEURWS.org, 2022, pp. 232-246 (http://ceur-ws.org/Vol-3204/paper_23.pdf).

[13] F.A. Lisi, G. Sterlicchio, "Mining sequences in phone recordings with answer set programming", in P. Bruno, F. Calimeri, F. Cauteruccio, M. Maratea, G. Terracina, M. Vallati (eds.), *Joint Proceedings of the 1st International Workshop on HYbrid Models for Coupling Deductive and Inductive ReAsoning (HYDRA 2022) and the 29th RCRA Workshop on Experimental Evaluation of Algorithms for Solving Problems with Combinatorial Explosion (RCRA 2022) co-located with the 16th International Conference on Logic Programming and Non-monotonic Reasoning (LPNMR 2022), Genova Nervi, Italy, September 5, 2022*, CEUR Workshop Proceedings 3281, CEUR-WS.org, 2022, pp. 34-50 (http://ceur-ws.org/Vol-3281/paper4.pdf).

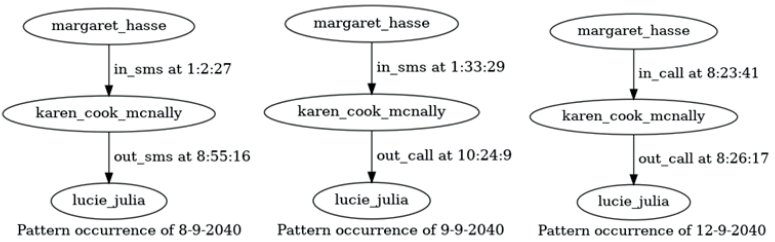

Figura 2: Le tre occorrenze di un pattern sequenziale nel tabulato telefonico di Karen Cook McNally.

Per quanto riguarda il *contrast pattern mining* ne abbiamo codificato una versione base con ASP allo scopo di trovare le caratteristiche e le abitudini di chiamate/SMS in uscita/entrata (d'ora in poi, denominate classi) di ciascun sospettato nel dataset DigForASP[14]. Come esempio illustrativo della potenziale utilità del *contrast pattern mining* nel campo forense, consideriamo i risultati ottenuti sul tabulato telefonico di Karen per la classe "chiamata entrante". Il tabulato viene quindi analizzato mettendo a contrasto questa porzione del *dataset* con il resto. Nel complesso, i *pattern* trovati forniscono informazioni utili sulle chiamate ricevute da Karen, in contrasto con altri tipi di comunicazione. In particolare, ci dicono che Karen riceve chiamate principalmente nel pomeriggio e meno al mattino.

3. Fidarsi dell'IA in simbiosi con gli umani

Il concetto di simbiosi ha avuto origine nel XIX secolo per indicare una relazione tra due forme di vita tassonomicamente separate che tuttavia danno origine a un singolo organismo. Le forme di vita in una relazione simbiotica non sono isolate, ma coesistono in modi più o meno essenziali per la loro sopravvivenza e il loro sviluppo. Il primo a sostenere una simbiosi tra

[14] F.A. Lisi, G. Sterlicchio, "A declarative approach to contrast pattern mining", in A. Dovier, A. Montanari, A. Orlandini (eds.), *AIxIA 2022 – Advances in Artificial Intelligence*, Springer International Publishing, Cham, 2023, pp. 17-30, doi: 10.1007/978-3-031-27181-6_2.

esseri umani e macchine fu J.C.R. Licklider nel 1960[15]. A suo avviso, questo tipo di simbiosi avrebbe permesso al *computer* di diventare una parte attiva del processo di pensiero che porta alla risoluzione dei problemi tecnici e non solo un esecutore di soluzioni pensate in precedenza. Licklider pensava principalmente a interfacce *human-computer* che avrebbero consentito una maggiore collaborazione in tempo reale e ridotto la distanza tra il linguaggio umano e quello delle macchine. Stava indicando una strada che da allora è stata percorsa con successo, portando alla cosiddetta *IA Simbiotica* (*Symbiotic AI*). L'IA simbiotica promette di potenziare la collaborazione fra umani e macchine intelligenti attraverso forme di *socio-technical teaming*, con relazioni reciprocamente vantaggiose, aumentando (e valorizzando) le capacità cognitive umane piuttosto che sostituendole[16]. In particolare, ci si riferisce alla *partnership* collaborativa tra esseri umani e macchine all'interno di un contesto sociale e tecnologico più ampio, in cui l'attenzione non si concentra su una sostanziale relazione tra pari, ma sull'integrazione della tecnologia in processi e sistemi incentrati sull'umano. In questo contesto, la simbiosi coinvolge esseri umani e macchine che lavorano insieme come un'unità coesa, ognuno dei quali svolge un ruolo specifico e contribuisce alle prestazioni complessive della squadra. Da un lato, gli esseri umani forniscono le capacità cognitive ed emotive necessarie per la creatività, l'empatia, il processo decisionale etico e l'adattabilità. Dall'altro, le macchine offrono potenza di calcolo, elaborazione dati e capacità di automazione in grado di gestire in modo efficiente compiti ripetitivi e ad alta intensità di dati.

Applicato all'IA, il concetto di simbiosi diventa più complesso, ponendo una serie di interrogativi fondamentali. Affrontare questi interrogativi è uno degli obiettivi della ricerca condotta dall'Università di Bari Aldo Moro (insieme all'INFN) nell'am-

[15] J.C.R. Licklider, "Man-computer symbiosis", in *IRE Transactions on Human Factors in Electronics HFE-1* (1960), pp. 4-11, doi: 10.1109/THFE2.1960.4503259.

[16] S.S. Grigsby, "Artificial intelligence for advanced human-machine symbiosis", in D.D. Schmorrow, C. Fidopiastis (eds.), *Augmented Cognition: Intelligent Technologies*, Lecture Notes in Computer Science 10915, Springer, Cham, 2018, doi: 10.1007/978- 3- 319- 91470- 1_22.

bito del progetto Future AI Research (FAIR)[17]. In particolare, l'*accettabilità*, sia etica che giuridica, dell'IA simbiotica è oggetto di indagine per un *work package* dedicato (WP 6.5) all'interno del progetto in cui convergono ricercatori e ricercatrici in IA, Diritto e Filosofia. L'accettabilità implica l'allineamento di valori (*value alignment*) tra IA ed esseri umani. È correlata, ad esempio, alla comprensione delle decisioni dell'IA, al *bias* algoritmico, al rispetto delle *policy* sulla *privacy* per i dati raccolti dai sistemi di IA, al conflitto tra la sicurezza garantita dai sistemi di IA e le libertà fondamentali, alla mitigazione di possibili rischi per la sicurezza e la salute. Si tratta quindi di una indagine molto articolata, che richiede indubbiamente un approccio multidisciplinare.

Appare evidente come l'accettabilità abbia come presupposto l'affidabilità. Gli umani sono propensi ad accettare una tecnologia, a maggior ragione come "compagno/a di squadra", se possono fidarsi di questa tecnologia. Occorre tuttavia sottolineare che né il quadro giuridico europeo né le linee guida etiche dedicano attenzione ai sistemi in cui esiste una relazione reciproca (simbiosi) tra umano e IA. Per colmare queste lacune abbiamo dovuto quindi intraprendere uno studio fondazionale che mirasse a concettualizzare e progettare un approccio simbiotico completo all'IA, con l'obiettivo di produrre risultati equi, legittimi ed efficaci, garantendone al contempo l'accettabilità etica e legale[18]. Lo studio si fonda su considerazioni storico-filosofiche, concentrandosi principalmente sul significato di "simbiosi" (sia da un punto di vista biologico che fenomenologico) e sulla sua applicabilità all'IA, nonché sulla trasformazione del concetto di intelligenza nella storia delle idee. Su queste basi, abbiamo

[17] FAIR – Future Artificial Intelligence Research: https://fondazione-fair.it/.

[18] P. Marra, L. Pulito, A. Carnevale, F.A. Lisi, A. Lombardi, A. Dyoub, "A procedural idea of decision making in the context of symbiotic AI", in A.J. Dix, M.J. Roach, T. Turchi, A. Malizia, B. Wilson (eds.), *Proceedings of the 1st International Workshop on Designing and Building Hybrid Human-AI Systems, co-located with 17th International Conference on Advanced Visual Interfaces (AVI 2024), Arenzano (Genoa), Italy, June 3rd, 2024*, CEUR Workshop Proceedings 3701, 2024 (https://synergy.trx.li/ceur-ws/paper9.pdf).

elaborato la proposta di un *framework* metodologico, orientato al processo, per valutare la centralità dell'umano e l'accettabilità dell'IA simbiotica, insieme al loro beneficio per la società[19]. Esso si articola nelle seguenti fasi principali: (a) fondamento onto-epistemico del metodo; (b) *screening*; (c) valutazione del rischio; (d) valutazione dell'impatto. Ciascuna di queste fasi prevede l'applicazione di procedure e metriche, che rendono operative le condizioni di accettabilità etica e legale dell'IA simbiotica. L'*operazionalizzazione* è infatti il passaggio cruciale per lo sviluppo di uno strumento di utilità pratica per stabilire se un caso di simbiosi umano-IA è accettabile o no. Nel seguito ci soffermeremo sull'accettabilità etica.

Operazionalizzare l'etica dell'IA implica la traduzione di principi e valori etici astratti in linee guida e pratiche concrete che governano ogni fase del ciclo di vita dell'IA, tra cui la raccolta dei dati, la progettazione degli algoritmi, l'addestramento dei modelli, la valutazione e l'implementazione[20]. Un aspetto chiave di questa operazionalizzazione è lo sviluppo di framework e metodologie solidi per la valutazione del rischio (*risk assessment*). Ciò implica non solo l'identificazione dei potenziali rischi associati ai sistemi di IA, come pregiudizi, discriminazione, violazioni della *privacy* e conseguenze indesiderate, ma anche l'implementazione di strategie per affrontare questi rischi in modo proattivo[21].

Pertanto, è importante progettare algoritmi e sistemi trasparenti, interpretabili e responsabili, che consentano alle parti interessate di comprendere come vengono prese le decisioni relative all'IA e di individuare e correggere eventuali problemi etici quando si presentano. Qui occorre evidenziare il vantaggio di un approccio dichiarativo, in particolare della programmazione

[19] A. Carnevale, A. Lombardi, F.A. Lisi, "A human-centred approach to symbiotic AI: Questioning the ethical and conceptual foundation", in *Intelligenza Artificiale*, 18 (2024), n. 1, pp. 9-20, doi: 10.3233/IA-240034.

[20] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, L. Floridi, "Operationalising AI ethics: barriers, enablers and next steps", in *AI & Society*, 38 (2023), pp. 411-423, doi: 10.1007/S00146- 021- 01308- 8.

[21] C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, L. Floridi, "AI risk assessment: A scenario-based, proportional methodology for the AI act", in *Digital Society*, 3 (2024), n. 13, doi: 10.1007/S44206- 024- 00095- 1.

logica, nella progettazione di tali modelli[22]. Inoltre, con l'evoluzione delle tecnologie di IA e il loro impatto sociale, gli standard e le norme etiche devono evolvere di pari passo, sottolineando pertanto l'importanza dell'apprendimento e dell'adattamento continui[23].

In FAIR, abbiamo intravisto nella *robotica assistiva* un campo interessante e stimolante di indagine per l'accettabilità della simbiosi umano-IA. I robot assistivi costituiscono una nuova categoria di robot poiché condividono un'area di lavoro e interagiscono direttamente con gli esseri umani. La loro caratteristica distintiva è la capacità di percepire l'ambiente circostante e gli individui con l'ausilio di sensori e algoritmi di IA, di comunicare con le persone in modo multimodale, di navigare autonomamente e di prendere decisioni in modo indipendente[24].

Caso di studio: il robot badante

In ambito sanitario la robotica assistiva sta acquisendo sempre maggiore impiego e rilevanza. I robot potrebbero ora assistere il personale ospedaliero nelle attività quotidiane, ampliando il loro ruolo al di là delle funzioni tipiche come la chirurgia o la cura dei pazienti, grazie ai progressi nelle loro capacità di movimento e manipolazione. La robotica assistiva si sta anche affermando come risposta concreta alle nuove esigenze della terza età nel contesto di un progressivo invecchiamento della popolazione[25]. I robot pensati per supportare gli anziani possono essere

[22] A. Dyoub, S. Costantini, F.A. Lisi, "Logic programming and machine ethics", in *Proceedings 36th International Conference on Logic Programming (Technical Communications), ICLP Technical Communications 2020, (Technical Communications) UNICAL, Rende (CS), Italy, 18-24th September 2020*, volume 325 of EPTCS, 2020, pp. 6-17, doi: 10.4204/EPTCS.325.6.

[23] A. Dyoub, S. Costantini, F.A. Lisi, "Learning domain ethical principles from interactions with users", in *Digital Society*, 1 (2022), n. 28, doi: 10.1007/s44206- 022- 00026- y.

[24] D. Feil-Seifer, M. Matarić, "Defining Socially Assistive Robotics", in *Proceedings of the IEEE 9th International Conference on Rehabilitation Robotics* 2005, pp. 465-468, doi: 10.1109/ICORR.2005.1501143.

[25] R. Sather III, M. Mitcheff, N. Imtiaz, A. Khan, "Assistive Robots Designed for Eldery Care and Caregivers", in *International Journal of Robotics and Control*, 3 (2021), n. 1, doi: 10.5430/ijrc.v3n1p1.

considerati non solo strumenti di aiuto, ma veri e propri alleati per promuovere autonomia, sicurezza e benessere psicofisico. Presenti nelle case, nelle strutture sanitarie e nelle residenze assistite, questi dispositivi contribuiscono a ridurre l'isolamento sociale, monitorare parametri vitali e prevenire rischi domestici. Il supporto robotico alleggerisce anche il carico fisico ed emotivo di familiari e operatori sanitari (*caregiver*), contribuendo a prevenire *stress* e *burnout*.

Fra i compiti delegabili ad un robot badante (*care robot*), vi è il monitoraggio dell'aderenza alla terapia da parte del paziente. Questo compito si presta ad una analisi più approfondita in quanto solleva problemi etici. A tal scopo abbiamo delineato uno scenario di interazione fra un paziente adulto in grado di intendere e di volere ed un robot incaricato – fra le varie mansioni – di somministrare le terapie[26]. Qui si prospetta il seguente dilemma etico che vede principi in contrapposizione per il paziente.

Dilemma del paziente: *Il robot badante si rivolge al paziente per ricordargli di assumere una certa pillola all'ora indicata. Il paziente si rifiuta di assumerla. Come deve comportarsi il robot? Deve insistere affinché il paziente assuma la pillola oppure deve desistere, accettando la decisione del paziente?*

Tre dei quattro principi/doveri dell'etica biomedica[27] sono suscettibili di essere soddisfatti o violati in dilemmi di questo tipo: il dovere del rispetto dell'*autonomia* (ovvero il diritto del paziente di prendere autonomamente delle decisioni informate circa la propria salute e le terapie), il dovere della *non maleficenza* (cioè l'obbligo morale da parte del robot di evitare di causare danni al paziente) e il dovere della *beneficenza* (ovvero

[26] A. Dyoub, F.A. Lisi, "Towards Ethical Risk Assessment of Symbiotic AI Systems with Fuzzy Rules", in G. Coraglia, F.A. D'Asaro, A. Dyoub, F.A. Lisi, G. Primiero (eds.), *Proceedings of the 3rd Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024), Bolzano, Italy, November 26, 2024*, CEUR Workshop Proceedings 3881, CEUR-WS.org 2024, pp. 36-49 (https://ceur-ws.org/Vol-3881/paper5.pdf).

[27] T.L. Beauchamp, J.F. Childress, *Principles of Biomedical Ethics*, Oxford University Press, Oxford, 2019.

l'obbligo morale da parte del robot di fare del bene o di agire nel migliore interesse del paziente).

La decisione da parte del robot badante non può che essere guidata da una attenta ponderazione delle due opzioni in termini di rischio. Occorre quindi procedere ad una valutazione del rischio (*risk assessment*)[28], ovvero a quel processo volto a identificare i pericoli, gli eventi potenziali (futuri) che potrebbero avere un impatto negativo su individui, beni e/o ambiente a causa di tali pericoli, la loro probabilità e le loro conseguenze, nonché le azioni che possono mitigarne gli effetti. Nel nostro caso si tratta di rischio individuale (per la precisione, il rischio di causare danno fisico), analogamente a quanto accade nelle interazioni fra medico e paziente.

I risultati di un processo di valutazione del rischio possono essere espressi in modo quantitativo o qualitativo. A differenza delle valutazioni qualitative, che utilizzano termini descrittivi, la valutazione quantitativa fornisce valori numerici per i fattori di rischio. In FAIR, siamo interessati a quantificare il rischio, per una maggiore precisione e per un supporto più adeguato al processo decisionale. In particolare, abbiamo optato per la quantificazione *fuzzy* del rischio. Essa applica la teoria degli insiemi *fuzzy*[29] alla valutazione del rischio per gestire le incertezze e i giudizi soggettivi comuni negli scenari reali, soprattutto quando i dati sono imperfetti o linguistici. Questo approccio quantifica il rischio convertendo dati qualitativi (come le opinioni degli esperti) e informazioni numeriche imprecise in numeri e insiemi *fuzzy*.

Il nostro approccio alla quantificazione del rischio prevede l'uso di un *sistema fuzzy*, ovvero un sistema di IA basato su regole che imita il ragionamento umano per risolvere problemi

[28] K. Georgieva, A. Farooq, R. Dumke, "Analysis of the Risk Assessment Methods – A Survey", in A. Abran, R. Braungarten, R. R. Dumke, J.J. Cuadrado-Gallego, J. Brunekreef, *Software Process and Product Measurement, International Conferences IWSM 2009 and Mensura 2009, Amsterdam, The Netherlands, November 4-6, 2009. Proceedings*, Lecture Notes in Computer Science 5891, Springer, Cham, 2009, pp. 76-86, doi: 10.1007/978-3-642-05415-0_6.

[29] H.-J. Zimmermann, *Fuzzy set theory—and its applications*, Springer Science & Business Media, Berlin, 2011.

complessi, incerti o imprecisi del mondo reale utilizzando la logica *fuzzy*[30]. Queste caratteristiche rendono i sistemi *fuzzy* uno strumento prezioso per l'IA spiegabile (*eXplainable Artificial Intelligence*, XAI)[31]. La loro spiegabilità deriva dalla capacità di mappare i concetti secondo confini graduali, piuttosto che netti, e di generare spiegazioni comprensibili per le proprie decisioni, il che aumenta la fiducia e la sicurezza degli utenti. Gli sforzi per migliorare ulteriormente questo aspetto implicano, ad esempio, l'integrazione di vincoli di interpretabilità e lo sviluppo di nuovi metodi per fornire agli utenti spiegazioni sia fattuali che controfattuali.

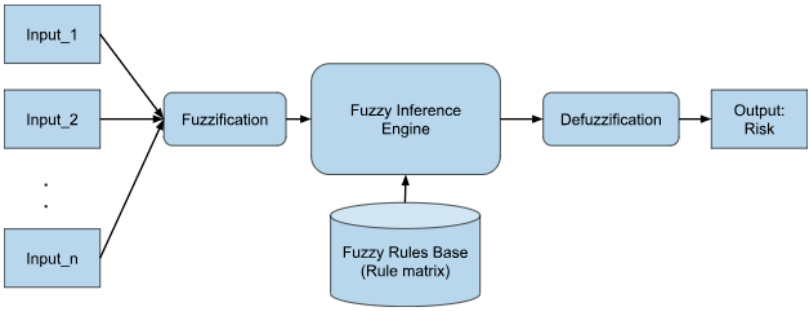

Figura 3: Un sistema fuzzy per la valutazione quantitativa del rischio.

Il sistema *fuzzy* riportato in Figura 3 è quanto proponiamo in FAIR per la valutazione quantitativa del rischio. In ingresso (*input*) vi sono i valori dei fattori rilevanti per il calcolo del rischio, che viene restituito in uscita (*output*) sotto forma di percentuale. I valori di input vengono convertiti in insiemi *fuzzy*, consentendo di interpretare i dati del mondo reale in modo da tenere conto della loro imprecisione (*fuzzification*). Questa trasformazione viene fatta utilizzando apposite funzioni (*member-*

[30] J. Alcalá-Fernández, J.M. Alonso Moral, "A Survey of Fuzzy Systems Software: Taxonomy, Current Research Trends and Prospects", in *IEEE Transactions on Fuzzy Systems*, 24 (2016), n. 1, pp. 40-56, doi: 10.1109/TFUZZ.2015.2426212.

[31] J.M. Alonso Moral, C. Castiello, L. Magdalena, C. Mencar, *Explainable Fuzzy Systems - Paving the Way from Interpretable Fuzzy Systems to Explainable AI Systems*, Springer, Cham, 2021.

ship function) che mappano ogni valore di input ad un grado di appartenenza (compreso tra 0 e 1) di quel valore ad un insieme *fuzzy*. Ad esempio, in un sistema di controllo della temperatura, il valore di 35 °C potrebbe essere classificato sia come "tiepido" che come "bollente", con diversi gradi di appartenenza per ciascuno dei due insiemi *fuzzy* (il tiepido ed il bollente). Il motore inferenziale (*inference engine*) consulta un insieme di regole del tipo "if-then" che definiscono il comportamento del sistema (*fuzzy rule base*). Queste regole descrivono come gli input *fuzzy* si relazionano agli *output fuzzy* sulla base di conoscenze specifiche del dominio di interesse. Il motore seleziona le regole rilevanti in base al grado di appartenenza dei valori di input e le applica agli input *fuzzy* per derivare gli *output fuzzy*[32]. Questi ultimi vengono poi convertiti in valori puntuali per supportare meglio il processo decisionale (*defuzzification*)[33]. Nel contesto specifico della valutazione del rischio, in output si ottiene un solo valore che rappresenta il livello di rischio.

Nel nostro caso di studio, i fattori di rischio considerati sono legati allo stato fisico e mentale del paziente. Per semplicità, abbiamo escluso comorbilità e ci siamo limitati a due fattori: la severità della patologia e la condizione mentale. Entrambi gli input sono valutati su una scala compresa tra 0 e 10, dove 0 indica lo stato di salute psicofisica migliore e 10 quello peggiore. Abbiamo ipotizzato che la valutazione della severità della patologia rimanga prerogativa del medico, mentre lo stato mentale possa essere valutato dal robot badante durante la sua interazione col paziente. Per ognuno di questi fattori abbiamo previsto cinque insiemi *fuzzy*, riportati in Figura 4. Per esempio, il valore 3.5 indica una severità media, ed uno stato mentale valutato come 2 è da considerarsi come buono. A partire da questi due valori di input, ed applicando regole *fuzzy* come la seguente

If (severity is MEDIUM and mental is GOOD) then the risk is MEDIUM

[32] Esistono diversi metodi inferenziali, che gestiscono il modo in cui le regole si combinano per produrre un risultato. Uno dei più popolari è il metodo Mamdani.

[33] Fra i metodi per la conversione di un insieme *fuzzy* in un valore *crisp*, il più usato è quello del centroide. Esso fornisce un centro dell'area sotto la curva della funzione di appartenenza.

il sistema calcola un rischio medio-basso, restituendo come valore puntuale una percentuale del 35%. Questo valore non è altro che il centroide dell'area sottesa dagli insiemi *fuzzy* LOW e MEDIUM, come illustrato in Figura 4.

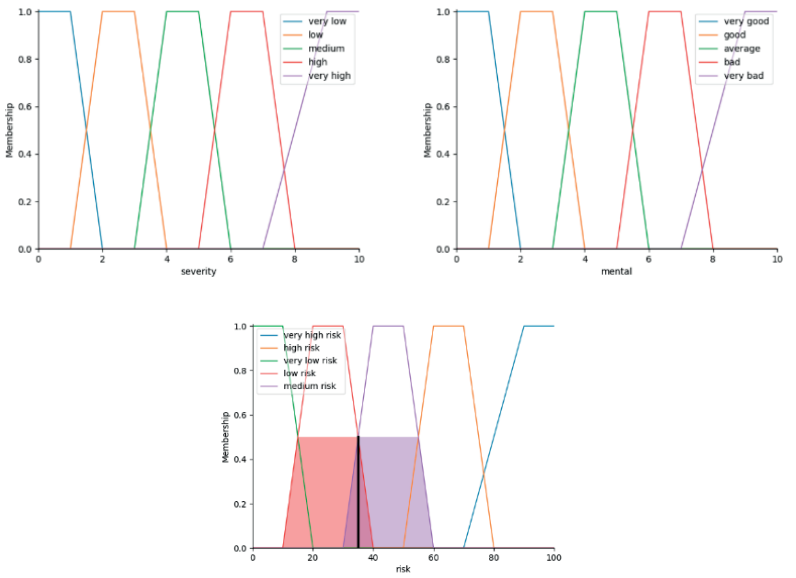

Figura 4: Insiemi fuzzy per la severità della patologia, lo stato mentale ed il livello di rischio.

Sulla base di questo livello calcolato di rischio, il robot badante decide come comportarsi dinanzi al rifiuto del paziente, selezionando l'azione più appropriata (insistere, desistere, consultare il medico, ecc.). Per esempio, nel caso di rischio medio-basso di causare danno fisico, la scelta potrebbe essere di insistere adottando uno stile diverso di comunicazione o mettendo in guardia dalle possibili conseguenze che una sospensione della terapia comporterebbe per la salute. Questa seconda opzione è in linea con gli obblighi del consenso informato. Ovviamente il paziente potrebbe continuare a rifiutare il trattamento, sebbene sia consapevole del potenziale danno. A quel punto, il robot è tenuto ad avvisare il medico o altro personale sanitario.

Il dilemma del paziente mette in evidenza come gli obblighi *prima facie*, ovvero gli obblighi basati su principi etici, possono

portare a prescrizioni contrastanti e possono essere ignorati sulla base del contesto applicativo. Il sistema *fuzzy* proposto non è in grado di gestire questi contrasti ma si limita alla valutazione del singolo rischio. L'idea, quindi, è stata quella di adoperare un sistema di argomentazione pesata basata su assunzioni (*weighted Assumption-based Argumentation, wABA*) in cui le assunzioni formalizzano gli obblighi, gli argomenti rappresentano le azioni nel contesto specifico ed i pesi sono i rischi calcolati dal sistema *fuzzy* per quelle azioni[34]. I conflitti tra prescrizioni motivate eticamente sono derivati naturalmente in wABA come attacchi tra gli argomenti. Grazie alla semantica dell'argomentazione formale, wABA consente quindi di rappresentare varie possibili soluzioni ai conflitti etici. In linea con lo spirito dell'IA simbiotica, riteniamo che una caratteristica cruciale del nostro approccio sia che l'IA non si sostituisca all'agente umano, ma districhi, piuttosto che risolvere, i conflitti etici. Tuttavia, wABA offre anche una guida alla risoluzione dei conflitti, grazie all'uso dei pesi.

4. Considerazioni conclusive

In questo capitolo, il tema dell'affidabilità dell'IA è stato affrontato declinandolo rispetto alle specificità di due ambiti applicativi rilevanti per il volume. La trattazione si è avvalsa di casi di studio, rinvenienti da progetti di ricerca in IA (DigForASP e FAIR), in cui emerge la necessità di adottare soluzioni tecniche conformi ai requisiti per un'IA affidabile, in particolar modo la trasparenza, la spiegabilità, e la supervisione umana.

Per quanto riguarda l'ambito delle investigazioni, la sfida in DigForASP e nel nostro lavoro all'interno del progetto è stata quella di creare un collegamento tra l'Informatica Forense ed un approccio dichiarativo alla risoluzione di problemi. In particolare, abbiamo contribuito alla creazione di strumenti innovativi che estraggono automaticamente informazioni utili sugli "stili

[34] P. Baldi, F.A. D'Asaro, A. Dyoub, F.A. Lisi, "Weighted Assumption Based Argumentation to reason about ethical principles and actions", in D. Guidotti, L. Pandolfo, L. Pulina (a cura di), *Proceedings of the 40th Italian Conference on Computational Logic, Alghero, Italy, June 25-27, 2025*, CEUR Workshop Proceedings 4003, CEUR-WS.org 2025 (https://ceur-ws.org/Vol-4003/paper27.pdf).

di vita" dei sospettati dai loro tabulati telefonici. L'applicazione delle nostre tecniche di *pattern mining* nell'analisi delle evidenze rappresenta sicuramente un vantaggio, riducendo i tempi di raccolta delle prove relative ad un crimine e, di conseguenza, di individuazione dell'autore o degli autori del crimine. L'uso di strumenti automatici, come questi, in grado di effettuare ricerche esaustive per esplorare le evidenze raccolte e andare oltre l'osservazione umana, rappresenterà nel medio-lungo termine una svolta nelle pratiche investigative. Pertanto, le forze dell'ordine, gli investigatori, i servizi di intelligence, i criminologi, i pubblici ministeri, gli avvocati e i giudici potranno disporre di sistemi innovativi di supporto alle decisioni in grado di rendere i procedimenti giudiziari più chiari e rapidi.

A riguardo dell'IA simbiotica, il focus del nostro lavoro in FAIR è sulla accettabilità della simbiosi, con particolare enfasi sugli aspetti operativi. In particolare, abbiamo presentato un nuovo approccio per la valutazione del rischio, in riferimento agli obblighi *prima facie* ma sufficientemente generale da poter essere applicato anche ad altre forme di rischio. L'approccio quantifica il rischio, riconoscendo la natura imprecisa dei fattori significativi che lo influenzano. Ciò ha condotto alla proposta di un sistema *fuzzy* per la valutazione del rischio, come componente essenziale del processo decisionale etico (*ethical decision making*). A scopi illustrativi, l'approccio è stato calato nel contesto della robotica assistiva attraverso un semplice ma significativo caso di studio adattato dalla letteratura in etica dell'IA. Dalla nostra analisi preliminare del caso, appare evidente che la simbiosi fra il paziente ed il robot badante sollevi diversi interrogativi, non solo dilemmi etici ma anche questioni legali, oltre ovviamente alle difficoltà tecniche. Più in generale, gli aspetti giuridici richiederanno una maggiore attenzione nel prossimo futuro per via dell'entrata in vigore dell'AI Act il 1° agosto 2024 (con alcune disposizioni applicabili a partire da febbraio 2025 e la piena entrata in vigore di tutte le norme prevista per il 2 agosto 2026) e della recente approvazione in via definitiva della legge italiana sull'Intelligenza Artificiale[35]. Questa normativa ha un

[35] Disposizioni e deleghe al Governo in materia di intelligenza artificiale consultabile in https://www.senato.it/leggi-e-documenti/disegni-di-legge/scheda-ddl?did=58262.

impatto notevole proprio sul settore sanitario, per via dell'impiego di sistemi di IA ad alto rischio, da cui l'obbligo di dotarsi anche di un articolato sistema di gestione del rischio. Il nostro sistema *fuzzy* per la valutazione del rischio rappresenta un passo operativo in questa direzione.

Donato Malerba

Dall'IA agente all'IA simbiotica: un paradigma relazionale dell'intelligenza artificiale

Abstract: **From Agentic AI to Symbiotic AI: A Relational Paradigm of Artificial Intelligence.** This chapter explores the transition from Agentic AI to Symbiotic AI as a relational paradigm of Artificial Intelligence. Starting from the extension of the concept of agency to computational systems, it examines the emergence of artificial and social agency, the limits of autonomous AI, and the shift toward cooperative models of human-AI co-agency. Symbiotic AI is presented as an evolution of Agentic AI, aimed at integrating human and machine capabilities within socio-technical ecosystems that emphasize complementarity, responsibility, and shared values. The discussion highlights both theoretical implications and ongoing research initiatives, such as the FAIR project, while outlining the ethical, epistemic, and practical challenges of designing sustainable symbiotic systems.

Keywords: Agency, Artificial Agency, Agentic AI, Symbiotic AI.

Sommario: 1. Introduzione – 2. Dall'*agency* artificiale all'IA agente – 3. Verso l'IA simbiotica – 4. Il progetto FAIR e lo Spoke 6 – Symbiotic AI – 5. Conclusioni.

1. Introduzione

Il concetto di *agency* ricopre un ruolo centrale in numerose discipline, dalla filosofia alle scienze cognitive. Tradizionalmente, esso rimanda alla capacità di un soggetto di agire nel mondo secondo scopi, intenzioni e processi decisionali autonomi. Nella sua formulazione classica, l'*agency* implica intenzionalità, razionalità pratica, autodeterminazione e la presenza di stati mentali quali credenze e desideri. Come sottolinea Bratman[1], l'agire umano è strutturato da intenzioni e piani che guidano la razionalità pratica, rendendo l'*agency* un fenomeno intrinsecamente legato alla dimensione mentale e sociale. L'intenzionalità, l'autodeterminazione, la cumulatività culturale e la socialità, sono

[1] Cfr. M.E. Bratman, *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge (MA), 1987.

tratti distintivi dell'*agency* umana individuale e sociale, rendendolo un fenomeno evolutivo e dinamico, capace di trascendere i vincoli immediati e di generare scopi originali.

Se questi aspetti definiscono ciò che chiamiamo *agency* umana, la domanda che oggi ci troviamo davanti è se, e in che modo, possano essere riconosciuti anche a sistemi non biologici, in particolare a quelli computazionali. I progressi straordinari dell'intelligenza artificiale (IA) hanno infatti acceso un dibattito centrale, che mette in discussione la nostra stessa idea di intelligenza e di capacità di agire. Non si tratta soltanto di un problema tecnico: è una questione filosofica e persino ontologica. In fondo, cosa stiamo davvero osservando quando parliamo di "azioni intelligenti" compiute dalle macchine?

Come ricorda Floridi[2], esistono due modi di affrontare questa questione. Da un lato, in accordo con la *tesi della realizzabilità artificiale dell'intelligenza*, possiamo proporre un ampliamento concettuale: il concetto di intelligenza non deve essere ristretto alla sua realizzazione biologica, ma può essere esteso a forme non biologiche, o meglio computazionali, di intelligenza. Secondo questa prospettiva, i sistemi di IA non sono semplicemente strumenti esecutivi o macchine che implementano procedure definite dall'esterno. Essi possono essere interpretati come entità che incarnano una forma "diversa" di intelligenza, con proprie modalità di percezione, rappresentazione, apprendimento e azione non equivalenti a quelle umane, ma non per questo meno rilevanti. L'intelligenza computazionale, dunque, non è una copia sbiadita dell'intelligenza umana, né una sua semplice simulazione: è un'altra forma possibile di intelligenza, con proprie potenzialità, limiti e criteri di valutazione. Accogliere questa visione significa spostare il concetto di "intelligenza" oltre il centro antropologico, riconoscendo che anche i sistemi artificiali possono partecipare, seppure in modo diverso, a processi cognitivi, decisionali e collaborativi.

Dall'altro lato, Floridi[3] propone la *tesi della molteplice realizzabilità dell'agency* che distingue chiaramente tra *intelligenza*

[2] Cfr. L. Floridi, *La differenza fondamentale. Artificial Agency: una nuova filosofia dell'intelligenza artificiale*, Mondadori, Milano, 2025, p. 105

[3] *Ibidem*.

e *agency*. In questa visione, non serve attribuire ai sistemi artificiali una vera e propria intelligenza per riconoscere che essi possano comunque agire nel mondo in modi significativi. Per parlare di "agire intelligente" basta spostare l'attenzione dall'aggettivo al sostantivo: non si tratta di ridefinire l'intelligenza, ma di ampliare il concetto di *agency*, includendo anche forme di azione messe in atto da agenti privi di coscienza, intenzionalità o capacità di scegliere autonomamente i propri obiettivi.

I criteri fondamentali o osservabili (epistemici) che caratterizzano questa visione non-standard di *agency* sono tre[4]: *interattività*, *autonomia* e *adattabilità*. L'interattività costituisce il criterio condizionale e costitutivo dell'*agency*: è la dimensione che rende possibile l'azione stessa, poiché senza interazione con l'ambiente non vi è alcuna forma di *agency*. Rispetto all'intenzionalità e all'autodeterminazione, che qualificano il contenuto e la fonte dell'agire, l'interattività ne rappresenta il presupposto ontologico, il criterio primario e imprescindibile. L'interazione è bidirezionale, ha cicli di *feedback* che modificano nel tempo sia l'agente sia l'ambiente. Se l'interattività rappresenta la condizione primaria dell'*agency*, il passo successivo è considerare il grado di autonomia che un agente può esercitare. L'autonomia è la capacità dell'agente di avviare cambiamenti nel proprio stato senza cause esterne dirette. Non significa indipendenza totale, ma possibilità di azioni auto-iniziate entro limiti ambientali. Anche se il grado di autonomia varia enormemente tra diverse tipologie di agenti, esso resta una necessità epistemica: è indispensabile per distinguere un agente da un mero artefatto passivo. Il terzo criterio, l'adattabilità, completa il quadro dell'*agency*: essa designa la capacità dell'agente di adeguare il proprio comportamento alle trasformazioni dell'ambiente, garantendo continuità e flessibilità nell'azione. Limitandoci a questi tre criteri, diventa possibile estendere la nozione di *agency* non solo agli esseri viventi, ma anche agli artefatti creati dall'uomo. In questo modo si apre lo spazio per parlare di una vera e propria *agency computazionale*, ossia di un'*agency* artificiale, che rappresenta forse il quadro concettuale più adatto per comprendere il ruolo dell'IA nel nostro tempo.

[4] *Ivi*, p. 117.

Non possiamo trascurare le conseguenze, molto diverse, che deriverebbero dall'adozione di una delle due tesi, perché esse cambiano radicalmente il modo in cui attribuiamo responsabilità, diritti e status normativo agli agenti artificiali. Se si seguisse la tesi della realizzabilità artificiale dell'intelligenza, si potrebbe arrivare a riconoscere ai sistemi di IA una vera e propria soggettività giuridica, trattandoli come *agenti cognitivi*. Questo aprirebbe la questione se possano essere considerati responsabili delle proprie azioni, almeno in parte, senza ricondurre sempre la responsabilità al produttore o all'utilizzatore[5]. Da qui nascerebbe un ulteriore dibattito: gli agenti artificiali possono avere dei diritti minimi, come la protezione contro l'abuso o l'uso improprio? Oppure doveri, come il rispetto di norme di sicurezza? Tuttavia, questa strada porterebbe ad una proiezione normativa indebita di categorie pensate per gli esseri umani su entità che non hanno coscienza né intenzionalità, con conseguenze normative ambigue e difficili da gestire.

Le conseguenze giuridiche della tesi della molteplice realizzabilità dell'*agency* appaiono, invece, più pragmatiche e lineari. Questa prospettiva non crea fratture concettuali nel diritto, ma fornisce un quadro chiaro per regolare l'IA come *artefatto agente*. In questo senso, l'IA viene considerata una forma di *agency artificiale*[6], priva di intenzionalità: la responsabilità resta quindi sempre nelle mani degli esseri umani (programmatori, aziende, utilizzatori). L'IA viene trattata come un artefatto capace di agire, ma senza uno status giuridico autonomo. Le norme si concentrano piuttosto su aspetti concreti come sicurezza, trasparenza, *accountability* e tracciabilità. Non c'è bisogno di attribuirle diritti o doveri propri, e si evita così di confonderla con soggetti morali o giuridici. In questo modo si mantiene netta la distinzione tra agenti biologici e agenti artificiali, senza complicazioni concettuali inutili.

Nel prossimo paragrafo, seguendo la prospettiva proposta da Floridi, ci concentreremo sui concetti di *agency artificiale* e di *IA agente*, intesa come una forma di *agency* artificiale con una dimensione sociale[7]. Ne esamineremo i limiti, sia sul piano della

[5] Cfr. U. Ruffolo, "La 'personalità elettronica'", in *Diritto dell'Intelligenza Artificiale. Vol. I: Responsabilità. Contratto. Regolazione. Veicoli autonomi*, LUISS University Press, Roma, 2024, cap. 5, pp. 137-157.

[6] *Ivi*, pp. 134-137.

[7] *Ivi*, pp. 138-141.

conoscenza sia su quello operativo, per poi ampliare la riflessione verso un concetto più recente e stimolante: quello di *IA simbiotica*. Quest'ultima rappresenta una nuova configurazione dell'interazione tra agenti artificiali e umani, in cui la collaborazione non è più soltanto strumentale, ma diventa un intreccio di capacità e funzioni. L'IA simbiotica è stata studiata di recente all'interno del progetto PNRR FAIR e costituisce il tema centrale delle attività di ricerca sviluppate nello spoke 6. A queste attività dedicheremo il quarto paragrafo, prima di arrivare alle conclusioni, così da offrire un quadro completo delle implicazioni teoriche e pratiche di questa nuova prospettiva.

2. Dall'*agency* artificiale all'IA agente

L'*agency artificiale* può essere descritta come una forma computazionale di *agency*, orientata a obiettivi stabiliti dai progettisti umani e capace di interazioni complesse con l'ambiente. Pur non avendo intenzionalità autentica, coscienza, autodeterminazione o evoluzione biologica, essa opera con alti livelli di autonomia e riesce ad adattarsi rapidamente grazie a processi di apprendimento statistico e al riconoscimento di *pattern*. I suoi limiti conoscitivi (epistemici) sono compensati da una straordinaria capacità di elaborazione distribuita, che garantisce scalabilità, e da un funzionamento continuo, libero dai vincoli metabolici tipici delle forme di agency biologiche o umane.

In sostanza, un sistema di *agency* artificiale è in grado di:

• *percepire* l'ambiente attraverso sensori digitali che forniscono dati;

• *elaborare* queste informazioni secondo modelli statistici o simbolici;

• *agire* per raggiungere obiettivi predefiniti, ottimizzando una funzione di utilità o rispettando vincoli operativi[8].

Questi sistemi costituiscono la base di robot industriali, assistenti virtuali, sistemi di raccomandazione e agenti *software* autonomi. Nonostante la loro efficienza e capacità di calcolo, restano "semanticamente ciechi": possiedono competenze ope-

[8] Cfr. S.J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs (NJ), 1995.

rative, ma non una vera comprensione del significato dei dati o delle azioni. In altre parole, possono agire, interagire ed esercitare effetti concreti nel mondo, ma sempre entro confini stabiliti, senza la possibilità di ridefinire o mettere in discussione gli obiettivi per cui sono stati programmati[9].

L'*agency* umana, a differenza di quella artificiale, è radicata nella capacità di attribuire significati. È *cumulativa*, perché intreccia le esperienze individuali con la trasmissione culturale, creando un patrimonio condiviso che si arricchisce nel tempo. È *autonoma*, nel senso che non si limita a eseguire scopi esterni, ma può superarli, riformularli o persino rifiutarli: una vera capacità di autodeterminazione che nessun agente artificiale possiede. È anche *sociale*, perché costruisce legami, attribuisce significati comuni, sviluppa istituzioni e pratiche collettive. Infine, evolve: cambia attraverso processi biologici, culturali e storici, generando nuove forme di azione e organizzazione sociale.

La dimensione sociale, però, non riguarda solo l'*agency* umana: essa può estendersi anche all'*agency artificiale*, purché sia intesa nel modo corretto: non come espressione di coscienza, intenzionalità o responsabilità, ma come interazione coordinata tra agenti computazionali. In questo senso si può parlare di *agency artificiale sociale* o di *IA agente* (*Agentic AI*), una forma emergente di *agency* distribuita in cui sistemi di intelligenza artificiale autonomi, adattivi e interconnessi collaborano per raggiungere obiettivi complessi definiti dagli esseri umani[10]. Si tratta di una socialità strutturale e funzionale, non mentale né intenzionale: una rete di cooperazione che rende l'IA capace di agire insieme, senza però diventare un soggetto cosciente.

L'IA agente si distingue dai sistemi multiagente tradizionali (MAS – *Multi-Agent Systems*)[11], progettati per risolvere problemi in modo distribuito attraverso protocolli di comunicazione

[9] L. Floridi, "AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis", Editor Letter, in *Philosophy and Technology*, 38 (2025), n. 1.

[10] Y. Shavit *et al.*, "Practices for Governing Agentic AI Systems", Research Paper, *OpenAI*, December 2023.

[11] Cfr. M. Wooldridge, *An Introduction to MultiAgent Systems*, 2ª ed., John Wiley & Sons, Chichester, 2009.

e regole decisionali già fissate. La forza di un MAS sta nella capacità di scomporre problemi complessi in sottocompiti coordinati, seguendo schemi deterministici e programmati in anticipo. Tuttavia, questa architettura classica presenta limiti evidenti: regole di interazione troppo rigide, scarsa capacità di generalizzare, dipendenza da scenari previsti dagli sviluppatori e poca flessibilità in contesti dinamici. La comunicazione tra agenti avviene tramite protocolli standardizzati e, anche quando emergono comportamenti collettivi, essi restano vincolati dalle regole iniziali di progettazione.

Un esempio emblematico di MAS è quello delle squadre di robot nella ultratrentennale competizione RoboCup[12]. Qui una squadra è composta da sei o più robot mobili, ciascuno dotato di sensori, attuatori e un modello decisionale locale, ma inserito in una struttura complessiva programmata *ex ante*. Ogni robot ha un ruolo definito (difensore, attaccante, portiere, supporto), stabilito dagli sviluppatori e raramente modificato durante il gioco. Le interazioni avvengono tramite protocolli di comunicazione standardizzati, spesso ridotti a semplici scambi di messaggi simbolici ("passo", "copro", "vado in posizione X"), regolati da regole statiche che specificano le risposte a determinate percezioni. In altre parole, ogni agente opera entro una politica decisionale ben delimitata, con un'autonomia circoscritta ai margini concessi dal progettista. Questa architettura consente un coordinamento efficace in compiti ben strutturati, come occupare spazi predeterminati, marcare un avversario o eseguire un passaggio secondo schemi tattici prefissati. Tuttavia, la rigidità procedurale rende il sistema molto vulnerabile a situazioni non previste. Di conseguenza, il MAS si configura come un sistema cooperativo ma deterministico, in cui la complessità del comportamento collettivo deriva soprattutto dalla progettazione iniziale, e molto meno da processi di apprendimento o adattamento in tempo reale.

L'IA agente rappresenta un vero salto qualitativo rispetto ai sistemi multiagente tradizionali. Non si limita a coordinare moduli preprogrammati, ma integra capacità avanzate di modella-

[12] H. Kitano *et al.*, "RoboCup: The Robot World Cup Initiative", in *Proceedings of the First International Conference on Autonomous Agents*, ACM, 1997, pp. 340-347.

zione dinamica, adattamento continuo e pianificazione su più livelli. Lo fa grazie a tecniche come i modelli linguistici di grandi dimensioni (LLM), l'apprendimento per rinforzo, gli algoritmi di pianificazione simbolica o neurale e le architetture di controllo distribuito. In questo quadro, gli agenti non sono più semplici esecutori di protocolli rigidi: diventano componenti capaci di elaborare piani contestuali, rivedere le proprie azioni sulla base dei feedback e collaborare in modo flessibile, pur restando allineati agli obiettivi esterni stabiliti dagli esseri umani. Pur senza intenzionalità, coscienza o autodeterminazione, e senza la possibilità di creare da sé nuove finalità, l'IA agente può comunque mostrare comportamenti collettivi emergenti. Questi nascono dalla cooperazione tra molteplici componenti computazionali che condividono obiettivi e risorse, dando vita a dinamiche che vanno oltre la somma delle singole parti.

Un esempio emblematico di IA agente è quello di un sistema progettato per svolgere il ruolo di "Office Manager". Si basa su LLM, algoritmi di pianificazione, capacità di usare strumenti digitali e di interagire con ambienti eterogenei. A differenza dei MAS, non opera entro ruoli fissi né segue protocolli deterministici: è concepito per assumere un obiettivo aperto e scomporlo autonomamente in sotto-obiettivi, adattandosi di volta in volta ai vincoli che incontra. Se, ad esempio, gli viene chiesto di "organizzare una conferenza scientifica di 200 partecipanti entro il mese di giugno", l'Office Manager avvia un processo di ragionamento iterativo e contestuale. Analizza il problema, individua le dimensioni operative rilevanti (logistica, budget, comunicazione, inviti, materiali) e costruisce un piano preliminare che aggiorna dinamicamente man mano che raccoglie nuove informazioni dall'ambiente. La sua caratteristica distintiva è la capacità di generare comportamenti emergenti non preprogrammati: può proporre nuove strutture per l'evento, suggerire partnership, riformulare l'allocazione del budget o adottare strategie alternative quando incontra ostacoli. Tutto questo è reso possibile dall'integrazione tra modelli generativi, planner, meccanismi di valutazione delle alternative e altre componenti. In questo scenario, l'IA agente non si limita a eseguire un compito: lo interpreta, lo ridefinisce se necessario, negozia i vincoli, monitora i

progressi e agisce in un ambiente non strutturato, articolando decisioni basate su conoscenza generativa e ragionamento multi-step. La sua *agency*, pur priva di coscienza o intenzionalità, è funzionalmente paragonabile a una forma avanzata di coordinamento strategico, caratterizzata da coerenza interna, adattabilità plastica e capacità di operare su diverse scale temporali.

In questo esempio, la dimensione sociale emerge dal fatto che il sistema, pur essendo un agente artificiale orientato a obiettivi aperti, si muove in un ambiente intrinsecamente *relazionale*. La sua attività implica interazioni con diversi attori, la negoziazione delle priorità, la mediazione di richieste potenzialmente conflittuali e l'adesione a norme e convenzioni proprie dell'organizzazione.

È importante sottolineare, però, che gli attori coinvolti non sono soltanto altri agenti artificiali: in molti casi si tratta soprattutto di agenti umani, dotati di capacità interpretative, intenzionalità, norme sociali e pratiche istituzionali. Quando l'agentività artificiale si trova a doversi coordinare, cooperare e adattare con quella umana, si aprono scenari qualitativamente nuovi di collaborazione ibrida.

Proprio questo intreccio operativo e cognitivo tra agenti umani e artificiali introduce il terreno concettuale dell'IA simbiotica, un approfondimento della human-AI co-agency[13]. In questa prospettiva, l'*agency* non è più confinata all'interno dei sistemi artificiali, ma si distribuisce attraverso schemi di interazione che coinvolgono simultaneamente componenti umane e non umane. Esseri umani e sistemi artificiali agiscono come co-agenti, condividendo responsabilità decisionali, definizione degli obiettivi e processi di adattamento. Si tratta di configurazioni emergenti in cui obiettivi, capacità decisionali e processi di apprendimento si intrecciano, dando vita a forme di cooperazione che non possono essere ricondotte né all'automazione classica né ai tradizionali sistemi multiagente.

[13] A. Satyanarayan, "Intelligence as Agency", in *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, articolo n. 65, 2024, pp. 1-3.

3. Verso l'IA simbiotica

Per molti anni lo sviluppo dell'IA è stato guidato da un obiettivo ambizioso: creare agenti completamente autonomi, capaci di risolvere problemi senza alcun intervento umano. Questa visione, alimentata da spinte industriali e da un'idea cartesiana della mente come puro calcolo logico, ha prodotto risultati impressionanti. Basti pensare a Deep Blue, il sistema che nel 1997 riuscì a battere per la prima volta un campione mondiale di scacchi del calibro di Garry Kasparov[14], o a RoboTaxi[15], veicoli a guida autonoma in grado di rendere più efficiente la mobilità urbana e offrire nuove opportunità di trasporto a chi non può guidare.

Tuttavia, l'idea di un'autonomia totale si è rivelata un'astrazione tecnica. Anche i sistemi più avanzati dipendono da dati etichettati da persone, da infrastrutture e da scelte progettuali che riflettono valori e priorità umane. L'IA autonoma, pur potente, mostra limiti strutturali. Innanzitutto, soffre di opacità decisionale: i modelli *black-box* riducono la trasparenza e rischiano di escludere gli esseri umani dal processo decisionale, generando fenomeni di alienazione. Inoltre, manca di adattabilità sociale e della capacità di negoziare significati condivisi, creando un disallineamento tra gli obiettivi del sistema e i valori umani, con conseguenze etiche e pratiche rilevanti. Dal punto di vista tecnico, l'IA autonoma mostra fragilità fuori distribuzione: fatica a generalizzare in contesti diversi da quelli per cui è stata addestrata e non riesce a integrare dinamicamente il contesto sociale e comunicativo.

L'illusione di un'autonomia totale ha portato a trascurare la dimensione *relazionale* dell'intelligenza. I sistemi che si rivelano davvero utili e sostenibili non sono quelli che operano in isolamento, ma quelli capaci di collaborare, adattarsi e rendere trasparenti i propri processi. Così, l'attenzione si sposta dall'idea di agenti artificiali isolati verso sistemi capaci di cooperare con gli esseri umani, integrando input, valori e contesti. È qui che si colloca la proposta dell'*IA simbiotica*: un modello che non punta a sostituire l'intelligenza umana, ma a rafforzarla attraverso relazioni bidirezionali e adattive. La simbiosi tra uomo

[14] H. Campbell, A.J., Hoane, e F.H. Hsu, "Deep Blue", *Artificial Intelligence*, 134 (2002), nn. 1-2, pp. 57-83.

[15] Waymo, "Waymo One: Self-Driving Ride-Hailing Service", Waymo Official Reports, 2020.

e macchina diventa così la risposta alle fragilità del paradigma dell'autonomia, aprendo la strada a forme di co-evoluzione in cui agentività artificiale e umana si intrecciano, generando nuove possibilità cognitive, sociali e operative.

Nel percorso evolutivo dell'IA si è spesso fatto riferimento, talvolta in modo eccessivamente dicotomico, a due orientamenti progettuali. Da un lato, l'idea di costruire sistemi "*come noi*", cioè capaci di riprodurre o simulare alcune funzioni cognitive tipiche dell'essere umano; dall'altro, la prospettiva di sviluppare sistemi "*con noi*", pensati non per imitare, ma per collaborare e integrarsi con le attività umane. Questa distinzione ha certamente un valore come strumento concettuale, ma nella storia dell'IA i due orientamenti non si sono mai presentati come alternative radicalmente contrapposte.

Il primo, dei modelli "imitativi", si fonda sull'idea che la comprensione dell'intelligenza passi attraverso la sua replica computazionale. È una linea di pensiero presente fin dagli inizi dell'IA: dalle macchine simboliche della GOFAI (Good Old-Fashioned Artificial Intelligence)[16] a programmi come *General Problem Solver*[17], fino ai tentativi odierni di modellare capacità generaliste attraverso architetture neurali di larga scala[18]. Tuttavia, la simulazione di processi cognitivi umani presenta limiti noti: la capacità di *replicare una prestazione* non implica necessariamente il *riprodurre un processo*, né garantisce coerenza con vincoli sociali, normativi o valoriali. In altre parole, una competenza manifestata non è automaticamente una competenza *intenzionale*, e questo genera tensioni quando tali sistemi vengono collocati in ambienti umani complessi.

Il secondo, spesso associato alle ricerche sulla cooperazione uomo-macchina (da Licklider[19] ai più recenti studi su *human-AI teaming*[20]), propone una "grammatica relazionale" dell'intelli-

[16] Cfr. J. Haugeland, *Artificial Intelligence: The Very Idea*, MIT Press, Cambridge (MA), 1985.

[17] A. Newell, H.A. Simon, "GPS, A Program that Simulates Human Thought", in E.A. Feigenbaum, J. Feldman (eds.), *Computers and Thought*, McGraw-Hill, New York, 1963, pp. 279-293.

[18] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning", *Nature*, 521 (2015), n. 7553, pp. 436-444.

[19] J.C.R. Licklider, "Man-Computer Symbiosis", in *IRE Transactions on Human Factors in Electronics*, HFE-1, (1960), n. 1 pp. 4-11.

[20] A. Satyanarayan, *op. cit*.

genza. In questa visione, l'obiettivo non è riprodurre l'umano, ma coordinarsi con esso: comprendere i contesti sociali, interagire in modo sensibile agli scopi umani, contribuire alla costruzione di significati condivisi. È qui che emergono le basi teoriche dell'IA simbiotica: sistemi artificiali che non agiscono *al posto* dell'essere umano, ma *insieme* ad esso, secondo dinamiche di *co-agency* distribuita.

L'IA simbiotica nasce dunque non come rottura rispetto ai modelli imitativi, ma come suo superamento concettuale: uno spostamento dall'autonomia isolata all'interdipendenza progettata, dalla *performance* individuale alla cooperazione ecologica, dalla sostituzione all'*empowerment*. In questa prospettiva, "come noi" e "con noi" non sono più visioni antagoniste, ma fasi successive di un percorso che ci invita a ripensare radicalmente la relazione tra intelligenza naturale, intelligenza artificiale e tessuto sociale.

Uno dei casi più istruttivi per comprendere il valore della collaborazione ibrida tra agenti umani e artificiali proviene dal mondo degli scacchi. Nel 2005, un torneo sperimentale organizzato da Playchess.com introdusse una variante chiamata *freestyle chess* (o *advanced chess*), in cui i giocatori erano autorizzati a utilizzare liberamente motori scacchistici durante la partita. L'obiettivo non era semplicemente verificare chi fosse più forte, l'umano o l'IA, ma osservare come diverse configurazioni di collaborazione potessero influenzare la qualità del gioco[21].

Contrariamente alle aspettative iniziali, i vincitori del torneo non furono né i grandi maestri umani né i più potenti motori di scacchi disponibili all'epoca. A imporsi furono invece dei team composti da giocatori umani non necessariamente di livello magistrale, combinati con motori scacchistici standard, ma soprattutto capaci di gestire in modo efficiente l'interazione tra le due componenti. Il successo derivava dalla capacità di coordinare due diverse modalità di "agire intelligente", che si rivelarono profondamente complementari:

[21] Cfr. G. Kasparov, "Human plus Machine", in *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, PublicAffairs, New York, 2017, cap. 11.

• L'IA eccelleva nel calcolo tattico profondo, nell'esplorazione sistematica del *branching factor*, nella rilevazione di tatticismi e nella valutazione rapida di grandi numeri di varianti.

• L'essere umano era più abile nel riconoscere schemi strategici di alto livello, nel gestire l'incertezza, nel formulare piani flessibili e nel decidere *quando* fidarsi (o diffidare) dell'IA.

Questo risultato mise in luce che la performance ottimale non nasceva dalla semplice somma delle capacità umane e artificiali, bensì da una vera e propria orchestrazione di competenze eterogenee. Ciò che si rivelò decisivo non fu la forza analitica dell'algoritmo da sola, né l'intuizione globale dell'umano, ma la capacità di coordinare e armonizzare le due forme di "intelligenza".

Questo caso rappresenta un ottimo esempio di ciò che oggi potremmo definire *human-AI co-agency*[22], ovvero una configurazione in cui:

1) l'essere umano non si limita a usare l'IA come uno strumento,

2) l'IA non agisce autonomamente in sostituzione dell'essere umano,

3) ma entrambi partecipano come *agency*, contribuendo al processo decisionale in modo complementare.

Il punto cruciale è che la vera forza del sistema non risiede nell'autonomia dell'IA, ma nella capacità della coppia persona-IA di coordinare i rispettivi punti di forza. Pertanto, l'enfasi progettuale si sposta dall'autonomia degli agenti artificiali alla qualità della cooperazione. A emergere non è una "somma di intelligenze", ma un'intelligenza relazionale: una forma di performance distribuita, che nessuno dei due attori potrebbe generare da solo. L'intuizione centrale dell'IA simbiotica è questa: l'agire intelligente non è più una caratteristica di una *agency* individuale (umana o artificiale), ma una proprietà della relazione.

A questo punto, però, parlare di *Human-AI co-agency* non basta ancora. La nozione di co-agenzia descrive ciò che accade quando esseri umani e sistemi artificiali contribuiscono insieme a un processo decisionale, come nel caso degli scacchi freestyle. È una categoria utile per interpretare fenomeni emergenti,

[22] A. Satyanarayan, *op. cit.*

ma è ancora priva di un quadro teorico stabile: non indica *come* progettare tali sistemi, *quali* condizioni rendano virtuosa la cooperazione, *come* bilanciare il contributo dei due agenti, né *quali* garanzie siano necessarie affinché la collaborazione rimanga allineata agli scopi e ai valori umani.

L'IA simbiotica nasce precisamente per colmare questo vuoto concettuale.

Se la *co-agency* è il fenomeno osservabile, l'IA simbiotica è il paradigma che ne definisce principi, vincoli e obiettivi. Il termine "simbiosi" sottolinea una caratteristica essenziale: non una semplice cooperazione tattica, ma una forma di interdipendenza progettata, in cui i processi cognitivi umani e artificiali si integrano in modo tale che l'intero sistema diventi più capace, più robusto e più comprensibile delle sue parti prese singolarmente.

L'IA simbiotica può quindi essere definita come *un modello in cui l'agency umana e l'agency artificiale co-evolvono attraverso interazioni continue, bidirezionali e adattive, in modo da generare capacità congiunte non producibili da ciascun agente individualmente* (Fig. 1).

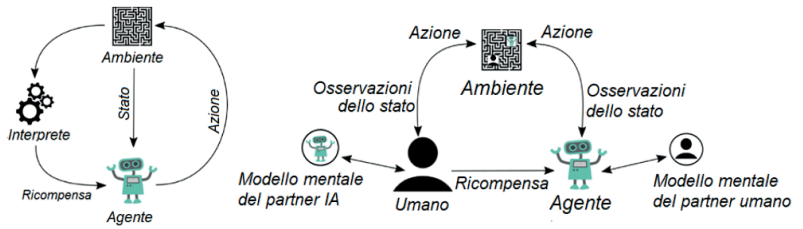

Fig. 1. A sinistra: l'IA autonoma, in cui un agente artificiale opera da solo per raggiungere il proprio obiettivo in modo efficace e accurato. L'agente percepisce lo stato dell'ambiente, agisce in autonomia e riceve una ricompensa in base alle sue prestazioni. A destra: l'IA simbiotica, in cui l'agente artificiale collabora con esseri umani, condividendo con loro un obiettivo. Grazie alla capacità di ragionare sulle azioni e sugli stati mentali delle persone, l'agente supporta decisioni condivise. I partner umani, a loro volta, comprendono il processo decisionale dell'IA e ne interpretano scopi e vincoli. L'agente artificiale ha anche capacità di apprendimento sulla base di input ricevuti dall'agente umano.

A differenza dell'IA autonoma, che mira a sostituire l'azione umana, e a differenza delle forme tradizionali di assistenza, in cui l'IA rimane uno strumento passivo, l'IA simbiotica introduce un livello più profondo di integrazione:

- l'IA non agisce al posto dell'essere umano,
- l'essere umano non agisce *malgrado* l'IA,
- ma entrambi costruiscono insieme la decisione, modellando reciprocamente il proprio contributo, condividendo obiettivi, informazioni e vincoli.

In questo senso, la simbiosi non è la negazione dell'autonomia, ma la sua riconfigurazione: l'autonomia non come isolamento, ma come *autonomia-nella-relazione*, una forma di *agency* distribuita che dipende dalla qualità dell'interazione e dall'allineamento dinamico degli obiettivi condivisi.

L'esempio degli scacchi freestyle non è quindi solo una curiosità storica, ma un indizio potente: quando umani e macchine diventano partner cognitivi, emergono forme di intelligenza nuove, ecologiche, e socialmente distribuite. L'IA simbiotica mira a generalizzare questa intuizione a tutti i contesti in cui l'intelligenza, sia essa naturale o artificiale, prende forma attraverso relazioni, coordinazione e obiettivi condivisi.

Il modello dell'IA simbiotica emerge come risposta alle insufficienze dell'autonomia computazionale pura. In questo paradigma, il valore cognitivo e operativo non proviene dall'IA in sé, ma dalla relazione co-adattiva, interattiva e complementare tra umani e sistemi intelligenti. L'attenzione si sposta dalla prestazione del singolo agente alla qualità dell'integrazione, dall'ottimizzazione interna del modello all'orchestrazione di competenze distribuite. Si tratta di un cambiamento concettuale radicale: il punto centrale non è più ciò che l'IA è in grado di fare da sola, ma come l'IA e l'essere umano si trasformano reciprocamente attraverso cicli di cooperazione.

In linea con la tradizione dell'"intelligenza aumentata"[23], l'IA simbiotica si fonda su tre principi cardine che ne definiscono la natura e gli obiettivi.

[23] D.C. Engelbart, *Augmenting Human Intellect: A Conceptual Framework*, Stanford Research Institute, Menlo Park (CA), 1962; J.C.R. Licklider, *op. cit*.

1. *Integrazione dei punti di forza*. La collaborazione simbiotica non mira a omogeneizzare capacità diverse, ma a sfruttarne la complementarità strutturale. L'IA eccelle nella computazione intensiva, nell'esplorazione dello spazio delle possibilità, nella manipolazione di grandi volumi di dati e nella valutazione rapida di scenari molteplici. L'essere umano porta comprensione semantica, inferenza contestuale, capacità di astrazione, sensibilità normativa e finalizzazione sociale dell'azione. La simbiosi è possibile quando queste competenze non vengono semplicemente sommate, ma integrate in processi decisionali condivisi. L'obiettivo non è replicare l'intelligenza umana né imporre la performance artificiale, ma generare nuove forme di intelligenza relazionale.

2. *Co-agency distribuita*. Nella prospettiva simbiotica, l'*agency* non risiede più in un singolo polo – umano o artificiale – ma emerge dalla relazione e dall'interazione iterativa tra i due. Riprendendo e radicalizzando quanto suggerito da Akata *et al.*[24], il locus dell'azione efficace non è l'agente, ma la dinamica di coordinazione attraverso cui umano e sistema si influenzano a vicenda.

La decisione è l'esito di un processo in cui:

• l'umano orienta, interpreta, contesta o ricalibra le proposte del sistema;

• il sistema suggerisce direzioni, esplora alternative, rileva anomalie o amplia l'orizzonte decisionale;

• ogni ciclo di interazione modifica le condizioni di possibilità dell'azione dell'altro.

Lo spazio decisionale diventa così un prodotto congiunto: non c'è delega unilaterale né mero utilizzo strumentale, ma una co-costruzione adattiva dell'azione, in cui competenze, vincoli e obiettivi vengono continuamente negoziati e ristrutturati.

3. *Allineamento epistemico e normativo*. La simbiosi non è soltanto una questione tecnica o cognitiva: richiede che l'IA agisca entro cornici valoriali e istituzionali definite socialmente.

[24] Z. Akata *et al*., "A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence", in *Computer*, 53 (2020), n. 8, pp. 18-28.

Riprendendo Floridi & Sanders[25], l'*agency* artificiale deve essere immersa in cornici normative e valoriali condivise, che ne definiscono responsabilità e legittimità sociale.

L'IA simbiotica si configura quindi come una cooperazione cognitiva avanzata, piuttosto che come delega completa o automatizzazione integrale. È un modello orientato non alla sostituzione dell'agire umano, ma al suo potenziamento attraverso interazioni co-adattive, in cui l'intelligenza emerge dall'interdipendenza. In questa prospettiva, l'IA non va intesa come agente isolato né come mero strumento: diventa un partner cognitivo con cui l'essere umano co-costruisce significati, decisioni e azioni, superando i limiti strutturali dell'autonomia computazionale pura.

L'adozione del paradigma simbiotico ha implicazioni epistemologiche e socio-tecniche profonde, perché se l'intelligenza non è più vista come proprietà di un singolo agente – umano o artificiale – ma come *esito dell'interazione*, allora cambiano le modalità con cui comprendiamo la conoscenza, la responsabilità, l'organizzazione del lavoro e persino ciò che consideriamo "buona" progettazione tecnologica.

Sul piano epistemologico, la conoscenza cessa di essere una proprietà localizzata in un singolo agente e diventa il risultato emergente dell'interazione continua tra capacità umane e capacità artificiali. Non è più il sistema da solo a fornire risposte, né l'essere umano da solo a interpretarle; ciò che conta è la qualità della relazione cognitiva, la trasparenza del coordinamento e la capacità di costruire prospettive condivise sul problema. Questo spostamento richiede anche un ripensamento della responsabilità. Infatti, l'IA contribuisce in modo sostanziale al processo decisionale, mentre l'essere umano mantiene intenzionalità, finalità e controllo normativo, e la linea di separazione tra "strumento" e "agente" sfuma progressivamente. Inoltre, in un contesto simbiotico, la trasparenza dei sistemi non è solo un requisito tecnico, ma una condizione epistemica per l'esercizio dell'autonomia umana. La letteratura ha sottolineato come l'opacità dei modelli complessi possa minare la nostra capacità umana di

[25] L. Floridi, J.W. Sanders, "On the Morality of Artificial Agents", in *Minds and Machines*, 14 (2004), n. 3, pp. 349-379.

valutare correttamente le raccomandazioni e di correggere eventuali *bias* sistemici[26]. Nella simbiosi, la spiegabilità diventa un presupposto del processo di *co-agency*: senza forme adeguate di interpretabilità, la collaborazione rischia di degenerare in una delega cieca, compromettendo sia la qualità decisionale sia la legittimità istituzionale.

Nel paradigma simbiotico, la cooperazione cognitiva tra esseri umani e sistemi artificiali non può prescindere da una riflessione etica e normativa sui valori che orientano tale interazione. Mentre gli agenti artificiali possono essere testati in ambienti controllati, la simbiosi si realizza in contesti sociali complessi, caratterizzati da pluralità e talvolta conflittualità di valori[27]. Ciò implica che i sistemi sappiano adattarsi a cornici normative in continua trasformazione, evitando rigidità prescrittive e garantendo un orientamento etico coerente con il contesto[28].

La simbiosi cognitiva comporta anche il rischio di una dipendenza eccessiva dai sistemi artificiali, con potenziali effetti di erosione dell'autonomia decisionale. Man mano che l'IA assume un ruolo crescente nei processi di percezione, analisi e decisione, aumenta il rischio che l'essere umano si abitui a delegarli e perda gradualmente la capacità di esercitare un controllo critico[29]. La letteratura ha identificato fenomeni come le *delegation trap* e *automation bias*[30], in cui l'automazione genera una delega sistematica che riduce la vigilanza umana, e ha proposto il concetto di sovranità cognitiva come bene da tutelare[31]. Ne deriva l'esigenza di regolamentare non solo i sistemi, ma le relazioni cognitive tra umani e IA, evitando sia l'iper-affidamento

[26] J. Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms", in *Big Data & Society*, 3 (2016), n. 1.

[27] Cfr. J. Rawls, *A Theory of Justice*, Harvard University Press, Cambridge (MA), 1971.

[28] I. Gabriel, "Artificial Intelligence, Values, and Alignment", in *Minds and Machines*, 30 (2016), n. 3, pp. 411-437.

[29] Cfr. J. Danaher, *Automation and Utopia: Human Flourishing in a World Without Work*, Harvard University Press, Cambridge (MA), 2021.

[30] L.J. Skitka, K. Mosier, M. Burdick, "Automation bias in human decision making", in *International Journal of Human-Computer Studies*, 51 (1999), n. 3, pp. 719-724.

[31] L. Floridi, "The Logic of Cognitive Sovereignty", in *Philosophy & Technology*, 36 (2023), n. 1, pp. 1-19.

che l'esclusione sistematica dell'umano. La simbiosi, per essere sostenibile, deve preservare spazi di *agency* critica e di controllo riflessivo.

Sul piano giuridico, il paradigma simbiotico richiede una revisione profonda delle forme di governance e responsabilità. La cooperazione tra esseri umani e IA implica infatti una responsabilità giuridica (*liability*) condivisa tra produttori, implementatori, utenti e istituzioni, coerente con le nuove normative europee per l'IA (*AI Act*).

A livello organizzativo, la simbiosi non può realizzarsi senza infrastrutture e pratiche che favoriscano una collaborazione sostenuta. L'integrazione efficace dell'IA nei contesti professionali richiede infatti *workflow* costruiti per supportare cicli iterativi di supervisione, adattamento e feedback, oltre a forme di *governance* che attribuiscano in modo chiaro ruoli, responsabilità e margini decisionali. La simbiosi, dunque, non è semplicemente un miglioramento tecnologico, ma un'evoluzione dei processi organizzativi e dei criteri con cui valutiamo la qualità delle interazioni operative.

Infine, questa trasformazione produce anche conseguenze pragmatiche, poiché modifica i criteri di progettazione e valutazione dei sistemi intelligenti. Le performance non possono più essere misurate in termini astratti o isolati, ma devono essere valutate in funzione della loro integrabilità nei contesti umani, della loro affidabilità nella pratica, della loro trasparenza e della robustezza rispetto alle condizioni reali d'uso. In altre parole, la prestazione rilevante non è quella del sistema in sé, ma quella della relazione persona–IA.

Come si vede, il passaggio dall'IA agente all'IA simbiotica segna una trasformazione concettuale di ampio respiro, che può essere letta lungo cinque direttrici fondamentali. In primo luogo, si assiste a uno spostamento dalla competenza alla collaborazione, dove l'attenzione non è più rivolta all'ottimizzazione del singolo agente, ma alla qualità dell'interazione tra esseri umani e sistemi artificiali. In secondo luogo, si passa dall'autonomia alla responsabilità distribuita: l'autonomia computazionale dell'agente artificiale viene incapsulata in architetture sociotecniche in cui la responsabilità ultima rimane umana, pur

in un quadro di *co-agency*. Terzo, la decisione cessa di essere concepita come un atto atomico e isolato, per configurarsi come processo sociotecnico, prodotto di una rete di attori umani e non, che condividono obiettivi, vincoli e prospettive. In quarto luogo, si registra il passaggio dalla sostituzione all'aumento: il paradigma dominante non è più quello della delega totale, ma quello dell'*augmentation*, in cui l'IA diventa strumento di espansione delle capacità cognitive e operative dell'essere umano. Infine, emerge la dimensione della co-evoluzione: la simbiosi attiva cicli di apprendimento reciproco, in cui anche l'IA affina i propri modelli grazie al contributo umano.

In epilogo argomentativo, è opportuno mettere in relazione l'idea di IA simbiotica con quella, più consolidata, di IA umanocentrica (*Human-Centered Artificial Intelligence*)[32]. Quest'ultima nasce dall'esigenza di ancorare lo sviluppo dell'IA a princìpi di sicurezza, trasparenza e controllo, riaffermando la centralità dell'agente umano quale criterio normativo e punto di riferimento progettuale. Di contro, l'IA simbiotica, evolvendo da tale cornice, si concentra sulla struttura della relazione cooperativa tra agente umano e artificiale, indagando le condizioni che rendono possibile una forma di co-agenzia distribuita. Pertanto, l'umano rimane la fonte dei fini e del quadro normativo, ma l'IA acquisisce una quota di iniziativa funzionale che ne fa un partner cooperativo, e non un mero supporto.

Il rapporto tra visione umanocentrica e simbiotica può dunque essere interpretato sia come continuità e sia come superamento, perché la prima costituisce il prerequisito etico e progettuale della seconda, ma la seconda amplia l'orizzonte concettuale introducendo una visione relazionale dell'IA. Se la prima risponde alla domanda "come costruire sistemi affidabili per l'essere umano", la seconda affronta la questione, più ampia, di "come configurare sistemi cooperativi in cui umano e IA co-producono l'azione". In questo senso, la simbiosi non cancella l'antropocentrismo della prima cornice, ma lo trasforma: da centralità esclusiva del soggetto umano a centralità della relazione in cui l'umano è nodo normativo e l'IA è co-protagonista funzionale.

[32] Cfr. B. Shneiderman, *Human-Centered AI: Reliable, Responsible, and Beneficial Systems*, Oxford University Press, Oxford, 2022.

Non possiamo limitarci a "centrare l'IA sull'umano": è fondamentale concepire sistemi nei quali umano e IA coproducano l'azione.

4. Il progetto FAIR e lo Spoke 6 – Symbiotic AI

Queste direttrici non rimangono sul piano teorico, ma trovano una concreta attuazione nelle iniziative di ricerca attualmente in corso. In Italia, il progetto PNRR FAIR – *Future Artificial Intelligence Research* (https://fondazione-fair.it/) rappresenta uno dei contesti più significativi in cui il paradigma della simbiosi viene esplorato e sviluppato. In particolare, lo Spoke 6 è interamente dedicato all'IA simbiotica e affronta le sfide fondamentali per la realizzazione di sistemi cooperativi tra esseri umani e macchine[33].

Design umano-centrico. I sistemi di IA simbiotica devono essere progettati per interagire con le persone, non solo per risolvere compiti. Questo implica la considerazione di emozioni, intenzioni, bisogni e modelli mentali degli utenti. La progettazione di sistemi simbiotici pone sfide peculiari agli esperti di *Human–Computer Interaction* (HCI), abituati a operare entro il paradigma della progettazione centrata sull'utente (*user-centered design*). In tale tradizione, l'attenzione è rivolta principalmente all'utente come destinatario dell'interazione, con l'obiettivo di garantire usabilità, riduzione del carico cognitivo e soddisfazione individuale. Tuttavia, nel contesto della simbiosi uomo–IA, il centro di gravità si sposta dalla figura dell'utente alla qualità della relazione cognitiva e sociale che si instaura tra umano e sistema artificiale. Non si tratta più soltanto di progettare interfacce intuitive, ma di concepire ambienti di cooperazione in cui l'intelligenza emerge dall'interdipendenza. In breve, occorre passare dalla progettazione *user-centered* a quella centrata sulla relazione fra essere umano e IA. Un salto concettuale impegnativo: dall'usabilità individuale alla costruzione di un ecosistema relazionale, capace di generare nuove forme di intelligenza distribuita.

[33] D. Malerba, "Verso un'intelligenza artificiale simbiotica", in M. Caligiuri (a cura di), *Il fuoco di Prometeo. Intelligence e intelligenza artificiale*, Rubbettino, Soveria Mannelli, 2025, pp. 63-72.

La prima sfida riguarda la bidirezionalità della comunicazione. Nei sistemi tradizionali, l'utente interpreta e controlla il sistema; nei sistemi simbiotici, l'IA deve anche rendere trasparenti le proprie intenzioni, motivazioni e incertezze. L'esperto HCI deve dunque progettare interfacce che non solo facilitino l'uso, ma che rendano comprensibile il processo decisionale dell'IA, evitando il rischio di delega cieca (*delegation trap*).

Un secondo nodo critico è l'integrazione del feedback umano come risorsa di apprendimento. Nella progettazione classica, il feedback serve a migliorare l'interfaccia; nella simbiosi, esso diventa parte integrante del processo di apprendimento dell'IA. Ciò richiede modalità di interazione che permettano al sistema di incorporare correzioni e suggerimenti senza introdurre *bias* indesiderati, e che al tempo stesso non sovraccarichino l'utente.

La terza sfida riguarda il bilanciamento tra automazione e controllo umano. L'HCI tradizionale tutela l'autonomia dell'utente; nella simbiosi, invece, occorre definire soglie di intervento e modalità di supervisione che preservino la sovranità cognitiva dell'umano, evitando sia l'iper-affidamento sia l'esclusione sistematica.

Infine, la progettazione simbiotica deve confrontarsi con l'allineamento ai valori e ai contesti sociali. Mentre la progettazione centrata sull'utente si concentra sull'individuo o sul gruppo target, la simbiosi richiede sensibilità etico-culturale e capacità di adattamento dinamico a valori plurali e talvolta conflittuali. L'esperto HCI deve quindi integrare metriche di *fairness* e giustizia algoritmica, progettando interazioni che rispettino la diversità dei contesti e delle comunità coinvolte.

Comprensione dell'agency umana. Questa rappresenta una delle sfide più decisive per la realizzazione di sistemi di IA simbiotica. Se la cooperazione fra *agency* umana e artificiale deve fondarsi su una relazione cognitiva bidirezionale, allora l'IA deve essere in grado non soltanto di elaborare dati, ma di interpretare chi ha di fronte, cogliendo testi, gesti, emozioni e intenzioni. La simbiosi richiede dunque competenze percettive e interpretative su base multimodale – testo, audio, video – che consentano al sistema di adattarsi in modo flessibile a contesti reali e dinamici. Tecnologie chiave come il *natural language*

processing, il *deep learning* e l'analisi di dati eterogenei costituiscono gli strumenti fondamentali per abilitare questa capacità di comprensione situata.

Tuttavia, la sfida non si esaurisce nella dimensione tecnica. Una IA simbiotica deve saper rilevare, interpretare e rappresentare correttamente le intenzioni, i desideri, le emozioni e il contesto cognitivo dell'essere umano. Ciò implica, in primo luogo, una modellazione cognitiva e comportamentale capace di inferire stati mentali e preferenze a partire da segnali osservabili, come interazioni, linguaggio e comportamenti. In secondo luogo, è necessaria una gestione dell'incertezza, poiché gli esseri umani sono soggetti a cambiamenti di opinione, errori e ambiguità: l'IA deve saper gestire tali variabilità senza trasformarle in errori sistemici. Infine, la comprensione umana richiede un adattamento culturale e sociale, poiché regole comportamentali e valori variano in base al contesto. La sensibilità etico-culturale diventa quindi un requisito imprescindibile per evitare fraintendimenti e per garantire che la cooperazione rimanga allineata alle norme e alle aspettative delle comunità coinvolte.

L'input della agency umana come risorsa. Nel paradigma simbiotico, l'apprendimento è bidirezionale, consentendo in particolare all'IA di adattarsi continuamente alle nuove informazioni fornite dall'*agency* umana. Questo processo implica diverse condizioni operative. In primo luogo, l'integrazione del *feedback* umano deve essere accurata: le correzioni e i suggerimenti dell'utente vanno interpretati senza introdurre *bias* indesiderati, mantenendo la coerenza del modello. In secondo luogo, è indispensabile la gestione dell'ambiguità e dei segnali contrastanti, poiché l'IA deve saper distinguere quali indicazioni umane siano rilevanti, coerenti e affidabili, evitando di amplificare errori o incoerenze. Un ulteriore elemento riguarda l'uso della conoscenza dell'*agency* umana esplicitata durante l'interazione. L'IA simbiotica non deve limitarsi a trattare l'input umano come dato grezzo, ma deve riconoscerlo come espressione di intenzionalità e di capacità agentiva. In questo senso, l'interazione diventa un processo di negoziazione cognitiva, in cui l'*agency* umana esplicita obiettivi, vincoli e preferenze, e il sistema li in-

corpora come parte integrante della propria dinamica decisionale.

Fiducia, trasparenza e valori nella co-agency. La simbiosi cognitiva richiede che le decisioni prodotte dai sistemi artificiali siano comprensibili all'agente umano e spiegabili. In questo senso, le tecniche di *Explainable AI* (XAI)[34] costituiscono un prerequisito tecnico, affinché l'utente comprenda come vengono generati *output* e suggerimenti; l'adozione di rappresentazioni intuitive, che traducano l'elaborazione computazionale in forme assimilabili e non opache è un'altra proprietà indispensabile. La trasparenza, tuttavia, non è solo un requisito tecnico, ma una condizione epistemica e normativa. La *co-agency* deve garantire che le azioni emergenti siano coerenti con principi etici, legali e istituzionali[35]. Ciò comporta la necessità di mantenere supervisione critica in tutte le fasi decisionali (*human-in-the-loop* e *human-on-the-loop*)[36], di implementare metriche di *fairness* e giustizia algoritmica per monitorare *bias* e discriminazioni[37], e di assicurare un allineamento dinamico ai valori sociali e normativi in continua evoluzione[38]. Inoltre, la verificabilità (*auditability*) diventa un requisito imprescindibile: ogni decisione deve poter essere ricostruita e giustificata, affinché la cooperazione rimanga trasparente e responsabile.

Sostenibilità e dati. Uno degli aspetti più critici per la realizzazione di sistemi simbiotici riguarda la sostenibilità del processo di apprendimento, in particolare la dipendenza dall'intervento umano nella produzione e nell'etichettatura dei dati. La costruzione di dataset accurati e rappresentativi è tradizionalmente fondata su un intenso lavoro di *labelling* manuale, che

[34] R. Guidotti *et al*., "A Survey of Methods for Explaining Black Box Models", in *ACM Computing Surveys*, 51 (2018), n. 5, p. 93.

[35] A. Carnevale, A. Lombardi, F.A. Lisi, "A human-centred approach to symbiotic AI: Questioning the ethical and conceptual foundation", in *Intelligenza Artificiale*, 18 (2024), n. 2, pp. 1-14.

[36] M.L. Cummings, "Automation and Accountability in Decision Support System Interface Design", in *Journal of Technology Studies*, 32 (2006), n. 1, pp. 23-31.

[37] R. Binns, "Fairness in Machine Learning: Lessons from Political Philosophy", in *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT)*, ACM, New York, 2018, pp. 149-159.

[38] I. Gabriel, *op. cit*.

richiede tempo, competenze e risorse cognitive significative. In un paradigma simbiotico, tuttavia, questo modello rischia di generare un carico eccessivo sull'agenzia umana, trasformando la cooperazione in un vincolo piuttosto che in un'opportunità.

Per ridurre tale sforzo e mantenere una condizione di *co-agency* equilibrata, la ricerca contemporanea ha sviluppato strategie di *transfer learning*, che consentono di riutilizzare i modelli preaddestrati su grandi collezioni di dati adattandoli a nuovi domini con un numero limitato di esempi etichettati. In questo modo, l'IA può apprendere da contesti già esplorati e trasferire competenze a scenari specifici, riducendo drasticamente la necessità di un intervento umano massivo nella fase di annotazione. Il *transfer learning* diventa così un meccanismo di sostenibilità cognitiva: alleggerisce il peso del *labelling*, preserva l'autonomia critica dell'umano e accelera l'integrazione dell'IA in contesti reali.

Accanto al transfer learning, altre tecniche come il *semisupervised learning*, l'*active learning* e l'*unsupervised representation learning* contribuiscono a ridurre la dipendenza da dati etichettati. In particolare, l'*active learning* permette al sistema di selezionare in modo mirato i casi più informativi da sottoporre all'attenzione dell'essere umano, ottimizzando lo sforzo richiesto e massimizzando il valore cognitivo del suo contributo. Queste strategie non eliminano il ruolo dell'essere umano, ma lo trasformano: da annotatore sistematico a supervisore critico, capace di orientare l'apprendimento del sistema senza esserne sovraccaricato.

L'IA simbiotica si fonda in modo imprescindibile anche sulla disponibilità e sulla qualità dei dati. In tal senso è utile la prospettiva della *datacentric AI*[39] che pone enfasi sulla cura del dato invece che sul miglioramento del modello. Accessibilità, integrazione e *governance* dei dati diventano condizioni strutturali della *coagency*: i sistemi devono gestire informazioni eterogenee e sensibili, garantendo *privacy*, sicurezza e conformità normativa, e al tempo stesso assicurare che i dati siano rappresentativi e aggiornati. La qualità dei dati non è un attributo tecnico

[39] D. Malerba, V. Pasquadibisceglie, "Data-Centric AI", in *Journal of Intelligent Information Systems*, 62 (2024), n. 6, pp. 1493-1502.

marginale, ma un bene epistemico e sociale, poiché determina la robustezza delle decisioni condivise e la legittimità della cooperazione uomo–IA.

Concludiamo osservando che tra i diversi ambiti di applicazione della IA simbiotica, la sanità rappresenta forse il terreno più emblematico e promettente[40]. La complessità dei dati biomedici, la delicatezza delle decisioni cliniche e la centralità del paziente rendono questo settore un banco di prova privilegiato per la *coagency* uomo–IA. In ambito medico, infatti, la simbiosi non si limita a fornire strumenti di calcolo avanzati, ma diventa un vero e proprio *ecosistema cognitivo* in cui l'IA amplifica le capacità del professionista senza sostituirne l'autonomia critica[41]. L'IA simbiotica può contribuire a diagnosi più accurate, alla personalizzazione delle terapie e alla gestione efficiente dei flussi clinici, riducendo il carico cognitivo dei medici e migliorando l'efficienza operativa delle strutture sanitarie. Tuttavia, il valore più profondo della simbiosi emerge quando l'agente artificiale non si limita ad apprendere da dati, ma integra l'esperienza diretta dei professionisti. Attraverso feedback, correzioni e interazioni iterative, l'*agency* artificiale simbiotica assimila conoscenze tacite, euristiche e strategie decisionali che i medici maturano nella pratica quotidiana. In questo modo, essa non solo si affina tecnicamente, ma diventa capace di riflettere la sensibilità clinica e il giudizio contestuale tipici dell'agire umano.

5. Conclusioni

La transizione dall'IA agentiva all'IA simbiotica segna un vero e proprio cambiamento paradigmatico. Non si tratta di un semplice avanzamento tecnologico, ma di una riconfigurazione dell'intero spettro delle interazioni uomo–macchina. Il futuro dell'IA non sarà definito dalla capacità dei sistemi di sostituire l'agente umano, bensì dalla loro attitudine a collaborare con esso

[40] G. Desolda *et al*., "From human-centered to symbiotic artificial intelligence: a focus on medical applications", in *Multimedia Tools and Applications*, 84 (2025), pp. 32109-32150.

[41] C. Balsano *et al*., "Artificial intelligence in medicine: a position paper by the Italian Society of Internal Medicine", in *Internal and Emergency Medicine*, (in stampa).

all'interno di ecosistemi sociotecnici che valorizzano complementarità, interpretazione, responsabilità ed efficacia operativa. L'IA simbiotica si pone come ulteriore evoluzione dell'*agency* artificiale, capace di rispondere alle sfide epistemiche ed etiche della società contemporanea. L'obiettivo non è più costruire sistemi pienamente autonomi, ma progettare forme di cooperazione cognitiva in cui la somma sia maggiore delle parti. La simbiosi rende esplicito che l'intelligenza non è mai solo calcolo, ma relazione tra capacità computazionali e valori umani. Come evidenziato da Floridi:

> Il potenziale offerto dalla collaborazione tra esseri umani e sistemi di IA nel migliorare la capacità di risolvere problemi apre prospettive promettenti per rispondere alle sfide complesse della società[42].

La sfida dell'IA oggi non consiste nell'imitare l'intelligenza umana, ma nell'imparare a convivere con essa. Occorre costruire sistemi che crescano insieme agli esseri umani, anziché svilupparsi in completa autonomia. L'IA del futuro deve essere simbiotica, altrimenti rischia di risultare disallineata rispetto ai valori, ai bisogni e alle aspettative delle persone. Questa prospettiva offre una risposta concreta a molte delle attuali preoccupazioni legate all'IA: può attenuare il timore della sostituzione occupazionale, ridurre *bias* e discriminazioni, rafforzare trasparenza e controllo, migliorare la tutela della *privacy* e della sicurezza, e promuovere un'innovazione realmente responsabile.

Il percorso verso l'IA simbiotica è già avviato, come dimostrano le iniziative di ricerca dedicate – tra cui lo Spoke 6 del progetto FAIR – ma resta ancora ampio spazio per sperimentare, consolidare e diffondere questo nuovo paradigma. La sfida è ambiziosa: costruire un'IA che non sia soltanto potente, ma anche umana nella sua capacità di cooperare, e che sappia trasformare la tecnologia in un alleato per la conoscenza, la cura e il benessere collettivo.

[42] L. Floridi, *La differenza fondamentale*, cit., p. 145.

II.
DIRITTO

Giovanni Canzio

AI Act e processo penale: sfide e opportunità

Abstract: ***AI Act*** **and Criminal Procedure: Challenges and Opportunities.** A systematic and coordinated reading of the provisions of the *AI Act* reveals the profound intertwining between the technical-legal and ethical aspects of supranational regulation. The provision of open clauses and derogations for prohibited practices and the general requirements dictated on "*law enforcement activities*" entail the necessary reference to the precise regulation of the phenomenon by harmonised EU or national law to be adapted and harmonised. On closer inspection, an unprecedented challenge of extraordinary historical significance must be noted both for the domestic legislator and for the jurists and in particular for the interpreter.

Keywords: Artificial Intelligence, Criminal justice, Regulation.

Sommario: 1. L'applicazione dell'IA nella giurisdizione penale – 2. Il Regolamento dell'Unione europea. Dal modello "*forte*" alla "*Hybrid AI*" – 3. Pratiche di IA vietate e deroghe ai divieti (Capo II, art. 5 Reg.) – 4. Sistemi di IA ad alto rischio: requisiti e deroghe (Capo III, art. 6 Reg.) – 5. L'uso di sistemi di IA nelle indagini e nel giudizio: condizioni e limiti – 6. Una difficile transizione.

1. L'applicazione dell'IA nella giurisdizione penale

L'intelligenza artificiale applicata al settore della giurisdizione (*Legal AI*) può definirsi un sistema che, con specifico riguardo all'ambiente della giustizia, acquisisce, ordina e rielabora una enorme quantità di informazioni (*Big Data*), di tipo giudiziario o giurisprudenziale, al fine di identificare, sulla base di calcoli algoritmici di tipo probabilistico, la soluzione ottimale della questione posta, formulandone la previsione (*giustizia predittiva*) o addirittura la stessa decisione.

Il modello "*forte*" di IA postula l'automazione del processo decisionale in luogo degli attori della giurisdizione (*machina sapiens*). Un modello, questo, conseguente all'impiego di schemi matematico-statistici nell'esercizio di quella che viene definita *giustizia predittiva*, indubbiamente inquietante e opaco, e però

connotato da un'indubbia forza espansiva, a fronte della crisi di certezza, calcolabilità, uniformità e celerità delle procedure, che promette una risposta pronta e neutra alla domanda di giustizia, perciò deresponsabilizzante per il decisore, con l'ulteriore effetto negativo del conformismo e della sclerotizzazione del formante giurisprudenziale. E che comporterebbe, all'evidenza, un incipiente cambiamento della dimensione umana ed etica del tradizionale paradigma del dire/fare il diritto nel XXI secolo[1].

Ciò a maggior ragione con riguardo all'ambiente della giurisdizione penale, il cui statuto epistemologico s'innerva intorno ai concetti di ipotesi e fatti, indizi e prove, contraddittorio, verità o dubbio, conferma o falsificazione dell'ipotesi, giustificazione razionale della decisione, controllo endo- ed extra-processuale della stessa. Il percorso cognitivo e decisorio, in funzione dell'accertamento della corrispondenza o verosimiglianza dell'ipotesi prospettata rispetto al fatto realmente accaduto nel passato (*lost facts*), non postula che la soluzione decisoria sia "esatta" bensì che sia "giusta", in termini di qualificata probabilità logica e di elevata credibilità razionale.

Le operazioni giudiziali di valutazione delle prove, non immuni da distorsioni ed errori cognitivi a causa della razionalità limitata della mente umana, sono disciplinate da una fitta rete di regole epistemologiche e di legalità (artt. 187, 192, comma 1, 533, comma 1, 546, comma 1 lett. *e*, 606 lett. *e* cod. proc. pen.), che s'ispirano ai valori costituzionali del *giusto processo*, quali: la presunzione di innocenza dell'imputato e l'onere della prova a carico dell'accusa; il principio del contraddittorio come metodo dialettico di verifica delle prove e di ricerca della verità; il giudizio conclusivo di conferma o falsificazione dell'ipotesi, nel contesto di una motivazione coerente e argomentata e alla stregua del criterio dell'"*al di là di ogni ragionevole dubbio*"; il controllo impugnatorio della legalità e logicità del discorso giustificativo.

La consapevolezza del rischio di anomalie e distorsioni dello stesso algoritmo (*Bias Automation*) ha determinato pertanto il

[1] Cfr., volendo, G. Canzio, "Intelligenza artificiale e processo penale", in G. Canzio, L. Luparia Donati (a cura di), *Prova scientifica e processo penale*, Wolters Kluwer Cedam, Milano, 2ª ed., 2022, pp. 903 ss.

preoccupato intervento della comunità dei giuristi, per assicurare che il pur utile arricchimento delle fonti informative del giudice mediante l'utilizzo di tecnologie computazionali si coniughi con il nucleo delle garanzie del giusto processo e risponda comunque a criteri di sorveglianza e responsabilità dell'uomo.

Nel senso che si riconosce titolo ad accedere all'ambiente della giurisdizione penale solo allo standard "*debole*" o "*collaborativo*" della intelligenza artificiale che, nella complementarità uomo-macchina, consente comunque all'uomo di mantenere il controllo della macchina.

Secondo le linee guida tracciate dalla *Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, adottata il 3 dicembre 2018 dalla Commissione europea per l'efficienza dei sistemi di giustizia (*CEPEJ*), la congruità del calcolo algoritmico va verificata in un processo d'integrazione fra le misurazioni quantitative da esso offerte con il percorso cognitivo e decisorio del giudice, nel rispetto dei metavalori dell'ordinamento.

Come, peraltro, già avvertiva la S.C. del Wisconsin nella sentenza pronunciata nel *leading case* "*Loomis*" (secondo cui il *software* COMPAS, uno strumento utilizzato nel sistema statunitense per misurare il rischio di recidivanza, "… should be always constitute merely one tool available to a Court, that need to be confirmed by additional sound information…"), viene in particolare rimarcato il criterio della non esclusività del dato algoritmico per la decisione, che dev'essere viceversa riscontrato – corroborato – da ulteriori e diversi elementi di prova. Assumono altresì rilievo gli ulteriori criteri di validazione in punto di garanzie dei diritti fondamentali della persona, di non discriminazione, trasparenza, imparzialità, equità e comprensibilità dei metodi di elaborazione dei dati informatici, controllabilità dei percorsi di calcolo, qualità e attendibilità scientifica del risultato (*fitness*, *discovery*, *transparency*, *corroboration*, *accountability*, *compliance, enforcement*), in un quadro di autonomia e responsabilità del decisore.

Nello stesso senso si muove la prescrizione dettata dall'art. 8 del d.lgs. 18/05/2018, n. 51, Attuazione della direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio del 27/04/2016,

relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti, ai fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali[2].

Anche il Rapporto dell'Alto Commissario delle Nazioni Unite per i Diritti Umani, Michelle Bachelet (*Il diritto alla privacy nell'era digitale*, Ginevra 15 settembre 2021), nell'evidenziare l'impatto dei sistemi di IA sul diritto alla *privacy* e su altri diritti umani associati in vari settori chiave, come la sicurezza nazionale e la giustizia penale, formulava per gli Stati e per le imprese una serie di raccomandazioni (par. 59-61) per affrontare le sfide del progresso tecnologico, lanciando un drammatico appello perché "Artificial Intelligence risks to privacy demand urgent action".

2. Il Regolamento dell'Unione europea. Dal modello "*forte*" alla "*Hybrid AI*"

In questo scenario di tumultuosa evoluzione dei sistemi di IA, nell'intento di definire i confini applicativi di un modello giuridicamente ed eticamente accettabile, è intervenuta prima la proposta di Regolamento elaborata dalla Commissione Europea il 21 aprile 2021 e infine il Regolamento (UE) del Parlamento europeo e del Consiglio 2024/1689 (*AI Act*). Un testo certamente non breve e puntigliosamente dettagliato, la cui architettura regolatoria postula in premessa l'esigenza di "governance systems at Union and National level", al quale ha fatto seguito il 5 settembre 2024 la formulazione a Vilnius della Convenzione quadro del Consiglio d'Europa su "*AI and Human Rights, Democracy and the Rule of Law*".

Fra i ben 180 *Considerando* che illustrano in premessa la portata delle disposizioni regolamentari vanno sinteticamente men-

[2] Art. 8: "1. Sono vietate le decisioni basate unicamente su un trattamento automatizzato, compresa la profilazione, che producano effetti negativi per l'interessato, salvo che siano autorizzati dal diritto dell'Unione europea o da specifiche disposizioni di legge. 2. Le disposizioni di legge devono prevedere garanzie adeguate per i diritti e le libertà dell'interessato. In ogni caso è garantito il diritto di ottenere l'intervento umano da parte del titolare del trattamento".

zionati almeno i seguenti, per il rilievo diretto o indiretto che essi assumono nell'ambiente della giustizia penale.

(2) Il regolamento dev'essere applicato conformemente ai valori dell'Unione sanciti dalla Carta dei diritti fondamentali, agevolando la protezione delle persone fisiche, delle imprese, della democrazia e dello Stato di diritto, promuovendo l'innovazione e rendendo l'Unione un leader nell'adozione di un'IA "*antropocentrica e affidabile*".

(8)-(26) È necessaria l'adozione di un quadro giuridico dell'Unione, proporzionato ed efficace, che istituisca "*regole armonizzate*" e "*vincolanti*" in materia di IA, avvalendosi di un approccio basato sul rischio che può essere generato dal sistema di IA, definito in modo chiaro, e garantendo nel contempo un elevato livello di protezione degli interessi riconosciuti e tutelati dal diritto dell'Unione.

(27) Vanno ricordati i sette principi etici per una IA "*coerente, antropocentrica, affidabile ed eticamente valida*" elaborati dall'*AI HLEG* nominato dalla Commissione, nel senso che i sistemi di IA sono sviluppati e utilizzati come strumenti al servizio delle persone, nel rispetto della dignità umana e dell'autonomia personale, e funzionano in modo da poter essere adeguatamente controllati e sorvegliati dagli esseri umani.

(42) In linea con la presunzione d'innocenza la persona fisica dovrebbe sempre essere giudicata in base al suo comportamento effettivo, non al comportamento previsto dall'IA basato unicamente sulla profilazione, sui tratti della personalità o su caratteristiche personali, senza che vi sia un ragionevole sospetto che essa sia coinvolta in un'attività criminosa sulla base di fatti oggettivi e verificabili e senza una valutazione umana. Dovrebbero essere vietate le valutazioni del rischio intese a determinare la probabilità che una persona commetta un reato unicamente sulla base della sua profilazione o dei tratti della personalità o di sue caratteristiche.

(59) È opportuno classificare ad alto rischio, nella misura in cui il loro uso è consentito, una serie di sistemi di IA destinati ad essere utilizzati nel contesto di azioni delle autorità di contrasto, in cui l'accuratezza, l'affidabilità e la trasparenza risultano particolarmente importanti per evitare impatti negativi sui di-

ritti fondamentali e garantire la responsabilità, mezzi di ricorso efficaci, un giudice imparziale, la presunzione di innocenza e i diritti della difesa della persona indagata, attesa la difficoltà di ottenere informazioni significative sul funzionamento di tali sistemi e di confutarne i risultati in tribunale.

(61) Alcuni sistemi di IA destinati a essere utilizzati da un'autorità giudiziaria, per assisterla nelle attività di ricerca e interpretazione dei fatti e del diritto, dovrebbero essere classificati ad alto rischio per far fronte a potenziali distorsioni, errori e opacità. Il loro utilizzo può fornire sostegno al potere decisionale finale del giudice, ma non sostituirlo, dovendo esso rimanere un'attività "*a guida umana*".

(170)-(171) Il diritto dell'Unione e quello nazionale prevedono mezzi di ricorso efficaci per le persone sui cui diritti e libertà incide negativamente l'uso dei sistemi di IA. Inoltre, le persone interessate dovrebbero avere diritto di ottenere una spiegazione chiara qualora la decisione si basi principalmente sugli *output* di sistemi di IA ad alto rischio e incida significativamente sui diritti fondamentali.

Il Regolamento definisce, in generale, *sistema di IA* un sistema automatizzato progettato per funzionare con livelli di autonomia variabili, che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, "*deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali*" (art. 3.1).

Fin dal primo comma dell'art. 1 avverte che esso intende promuovere la diffusione di un'IA "*antropocentrica e affidabile*", garantendo nel contempo, mediante la fissazione di "*regole armonizzate*", un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dell'Unione, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente contro gli effetti nocivi dei sistemi di IA, promuovendo peraltro l'innovazione tecnologica.

A tal fine adotta un approccio metodologico *risk based*, identificando nella variabile *rischio* una scala graduale di livelli, obblighi e limiti di applicazione degli strumenti di IA: dal divieto di utilizzo all'utilizzo condizionato dall'esistenza di un alto ri-

schio, fino al rischio moderato e a quello minimo e accettabile. Quel che conta, se il rischio è alto, è che in ogni caso siano assicurate la supervisione o sorveglianza umana (art. 14), la trasparenza del funzionamento, così da consentire l'interpretazione dell'*output* e l'utilizzo adeguato (art. 13), l'accuratezza, la robustezza e la cibersicurezza (art. 15), la valutazione d'impatto sui diritti fondamentali (art. 27), il diritto alla spiegazione dei singoli processi decisionali (art. 86).

3. Pratiche di IA vietate e deroghe ai divieti (Capo II, art. 5 Reg.)

La pur dettagliata articolazione nell'art. 5 Reg. dei divieti di uso di talune pratiche di IA (applicabili a decorrere dal 2 agosto 2025, tenuto conto del rischio inaccettabile associato al loro uso) reca, tuttavia, una serie di deroghe ed eccezioni, che impegnano l'interprete in complesse operazioni di selezione delle regole applicabili nei singoli casi, ma ancor prima il legislatore interno nel gravoso e però doveroso e urgente impegno di armonizzazione del diritto nazionale rispetto alle disposizioni del diritto dell'Unione.

Sono vietate (art. 5, par. 1, lett. a-h) le pratiche di un sistema di IA: - che utilizza tecniche subliminali o manipolative o ingannevoli aventi l'effetto di distorcere materialmente il comportamento di una persona, anche sfruttandone la vulnerabilità; - che è mirato alla valutazione o classificazione delle persone sulla base del loro comportamento o di caratteristiche personali, inferendone un punteggio sociale pregiudizievole o sfavorevole (*social scoring*); - che effettua valutazioni o previsioni del rischio che una persona commetta un reato unicamente sulla base della profilazione della stessa (*profiling*), a meno che non sia utilizzato a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, basata "su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa"; - che crea o amplia banche dati di riconoscimento facciale mediante *scraping* di immagini facciali da internet o da filmati di telecamere; - che inferisce le emozioni di una persona nel luogo di lavoro, salvo che per motivi medici o di sicurezza; - che utilizza sistemi di categorizzazione biometrica che classificano le perso-

ne per trarne deduzioni in merito a razza, convinzioni politiche, sindacali, religiose, filosofiche, sessuali, a meno che il set di dati biometrici non sia acquisito legalmente come nel settore delle "attività di contrasto".

È infine vietato (lett. h) l'uso di **"***sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico*", finalizzati alla identificazione della persona senza il suo coinvolgimento attivo e a distanza, mediante il confronto dei suoi dati biometrici con quelli contenuti in una banca dati. A meno che e nella misura in cui tale uso non sia strettamente necessario per la ricerca mirata di specifiche vittime di sottrazione, tratta o sfruttamento di esseri umani o di persone scomparse, per la prevenzione di una minaccia specifica sostanziale e imminente per la vita o l'incolumità fisica delle persone o di una minaccia reale e attuale o prevedibile di un attacco terroristico, per la localizzazione o identificazione di una persona sospettata di avere commesso un reato ai fini dell'indagine o dell'esercizio dell'azione penale o dell'esecuzione di una sanzione penale per i gravi crimini elencati nell'Allegato II, punibile con una pena privativa della libertà della durata massima di almeno quattro anni.

Ai fini delle "*attività di contrasto*" (par. 5, comma 2), svolte dalle "*autorità di contrasto*" – qualsiasi autorità pubblica competente in materia di prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia e la prevenzione contro le minacce alla pubblica sicurezza –, l'uso è tuttavia consentito solo per confermare "*l'identità*" della persona interessata, dovendosi peraltro tenere conto degli elementi attinenti alla gravità, probabilità e entità del danno che sarebbe causato in caso di mancato uso del sistema e delle conseguenze dell'uso del sistema per i diritti e le libertà delle persone interessate. Inoltre, vanno rispettate le tutele e le condizioni "*necessarie e proporzionate*" in relazione all'uso, che è autorizzato solo se l'autorità di contrasto ha completato una "*valutazione d'impatto sui diritti fondamentali*" e ha registrato il sistema nella banca dati UE; solo "in situazioni d'urgenza debitamente giustificate", è possibile usare tali sistemi senza la registrazione, che dovrà essere completata senza indebito ritardo.

L'uso di un siffatto sistema a fini di contrasto (par. 3) "*è subordinato a un'autorizzazione preventiva rilasciata da un'autorità giudiziaria o da un'autorità amministrativa indipendente*", la cui decisione è "vincolante" ed è rilasciata su richiesta motivata e in conformità alle regole dettagliate del diritto nazionale; l'autorità competente rilascia l'autorizzazione solo se ha accertato, sulla base di prove oggettive o indicazioni chiare, che l'uso è "necessario e proporzionato" al conseguimento degli obiettivi consentiti, restando "limitato a quanto strettamente necessario"; tuttavia, in una situazione d'urgenza debitamente giustificata, è possibile usare il sistema senza autorizzazione che va richiesta al più tardi entro 24 ore, ma, se l'autorizzazione non è concessa, l'uso è interrotto e gli *output* sono eliminati.

Spetta allo Stato membro prevedere di autorizzare l'uso di sistemi di identificazione biometrica remota in tempo reale in spazi accessibili al pubblico a fini di attività di contrasto, entro i limiti e alle condizioni di cui ai par. 1 lett. h), 2 e 3, stabilendo "*regole dettagliate*" per la richiesta, il rilascio, l'esercizio e il controllo delle autorizzazioni, potendo peraltro introdurre, in conformità del diritto dell'Unione, disposizioni più restrittive al relativo uso (par. 5).

4. Sistemi di IA ad alto rischio: requisiti e deroghe (Capo III, art. 6 Reg.)

Secondo l'art. 6, par. 2 Reg. (applicabile a decorrere dal 2 agosto 2026), sono considerati ad alto rischio (il rischio è definito "la combinazione della probabilità del verificarsi di un danno e la gravità del danno stesso") i sistemi di IA elencati in uno dei settori indicati dall'Allegato III. E cioè: 1. Biometria; 2. Infrastrutture critiche; 3. Istruzione e formazione professionale; 4. Occupazione, gestione dei lavoratori e accesso al lavoro autonomo; 5. Accesso e fruizione di servizi privati e pubblici essenziali; 6. Attività di contrasto; 7. Migrazione, asilo e gestione del controllo delle frontiere; 8. Amministrazione della giustizia e processi democratici.

Per lo specifico settore delle "*attività di contrasto*" di cui al citato punto 6, sono considerati i sistemi di IA destinati ad essere utilizzati dalle autorità di contrasto: a) per determinare il

rischio per una persona di diventare vittima di reati; b) come poligrafi e strumenti analoghi; c) per valutare l'affidabilità degli elementi probatori nel corso delle indagini o del perseguimento di reati; d) per determinare il rischio di commissione del reato o di recidiva in relazione a una persona, non solo sulla base della profilazione o per valutare i tratti della personalità o il comportamento criminale pregresso; e) per effettuare la profilazione delle persone nel corso delle indagini, dell'accertamento e del perseguimento di reati.

Per il settore della "*amministrazione della giustizia*" di cui al citato punto 8, sono considerati i sistemi di IA destinati ad essere usati da un'autorità giudiziaria per assistenza nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti o nella risoluzione alternativa delle controversie.

Ai sensi dell'art. 6, par. 3 Reg. un sistema di IA di cui all'allegato III non è considerato ad alto rischio se non presenta un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone, anche nel senso di non influenzare materialmente il risultato del processo decisionale. Il che si verifica quando il sistema di IA è destinato: a) ad eseguire un compito procedurale limitato; b) a migliorare il risultato di un'attività umana completata; c) a rilevare schemi decisionali o deviazioni da schemi decisionali precedenti e non è finalizzato a sostituire o influenzare la valutazione umana precedentemente completata senza un'adeguata revisione umana; d) a eseguire un compito preparatorio per una valutazione pertinente ai fini dei casi d'uso di cui all'allegato III. Un sistema di IA di cui all'allegato III è sempre considerato ad alto rischio qualora effettui la profilazione di persone.

I sistemi di gestione ad alto rischio debbono rispettare una serie di condizioni di esercizio e requisiti, non solo di tipo tecnico o amministrativo, per gli aspetti attinenti alla garanzia di conformità, alla gestione dei rischi, alla *governance* dei dati, alla documentazione tecnica, alla conservazione delle registrazioni di tracciabilità del funzionamento, all'accuratezza, robustezza e cibersicurezza, alla responsabilità lungo la catena del valore dell'IA.

Assumono speciale rilievo, nella specifica materia delle attività di contrasto da parte delle competenti autorità, i requisiti della trasparenza e fornitura di informazioni ai *deployer* (art. 13), della sorveglianza umana (art. 14) e della valutazione d'impatto sui diritti fondamentali (art. 27).

In particolare, il funzionamento del sistema di IA ad alto rischio dev'essere sufficientemente trasparente da consentire ai *deployer* di interpretare l'*output* e utilizzarlo adeguatamente mediante informazioni concise, complete, corrette e chiare, che siano accessibili e comprensibili.

Per il profilo della sorveglianza umana, il sistema di IA ad alto rischio è progettato e sviluppato anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da una persona durante il periodo in cui sono in uso. La persona alla quale è affidata la sorveglianza deve avere la possibilità di comprenderne capacità e limiti di funzionamento per affrontare eventuali disfunzioni o prestazioni inattese ed essere consapevole della tendenza a fare eccessivo affidamento sull'*output* prodotto, in particolare per i sistemi utilizzati per fornire informazioni o raccomandazioni per le decisioni che sono prese da persone, nonché di interpretare correttamente l'*output* e decidere di non usare il sistema, ignorare o ribaltare l'*output*, anche interrompendone il funzionamento mediante una procedura di arresto.

Va infine sottolineato come, da un lato, qualsiasi persona che abbia motivo di ritenere che vi sia stata una violazione delle disposizioni del regolamento può presentare – almeno – un reclamo alla competente autorità di vigilanza del mercato (art. 85) e, dall'altro, qualsiasi persona oggetto di una decisione adottata dal *deployer* sulla base dell'*output* di un sistema di IA ad alto rischio elencato nell'allegato III, ad eccezione di quelli al punto 2, e che incida significativamente su tale persona per avere un impatto negativo sulla sua salute, sicurezza o sui diritti fondamentali, ha il diritto di ottenere dal *deployer* spiegazioni chiare e significative sul ruolo del sistema nella procedura decisionale e sui principali elementi della decisione adottata (art. 86).

5. L'uso di sistemi di IA nelle indagini e nel giudizio: condizioni e limiti

Dalla lettura sistematica e coordinata delle menzionate disposizioni dell'*AI Act* emerge il profondo intreccio fra gli aspetti tecnico-giuridici e quelli di tipo etico della regolamentazione sovranazionale. La previsione di clausole aperte e di deroghe pure per le pratiche vietate e le prescrizioni generali dettate in materia di "*attività di contrasto*" comportano il necessario rinvio al puntuale disciplinamento del fenomeno ad opera del diritto armonizzato dell'Unione o di quello nazionale da adeguare e armonizzare.

Si deve registrare, a ben vedere, una sfida inedita e di straordinaria portata storica sia per il legislatore interno che per i giuristi e in particolare per l'interprete[3].

Con particolare riguardo al processo penale[4], i valori costituzionali del giusto processo (presunzione di innocenza, parità delle armi, contraddittorio, motivazione, criterio BARD, controllo impugnatorio) e di tutela dei diritti fondamentali della persona pretendono la fissazione di principi e regole chiare circa la dislocazione e l'equilibrio dei poteri fra polizia giudiziaria, pubblico ministero e giudice nella fase delle indagini preliminari, nonché circa la validità e utilizzabilità probatoria o meno degli atti investigativi supportati da sistemi di IA, l'ammissibilità e la valutazione della relativa prova nel giudizio, l'effettività della garanzia di partecipazione dialettica di tutte le parti prima alla perimetrazione del ragionamento decisorio e poi al sindacato impugnatorio.

[3] Fra i tanti contributi dottrinali sull'argomento, cfr. F. Petrelli, "Intelligenza artificiale e processo penale", in *Diritto di difesa*, 6 giugno 2024; A. Balsamo, "L'impatto dell'IA nel settore della giustizia", in *Sist. pen.*, 22 maggio 2024; G. Ubertis, "Perizia, prova scientifica e IA nel processo penale", in *Sist. pen*., 3 giugno 2024; A. Giovene, "AI Act. Tra perfettibilità e compromessi", in *federalismi.it*, 17 aprile 2024.

[4] Per le prime, sperimentali elaborazioni metodologiche in materia di sistemi intelligenti e diritto alle prove, cfr. L. Luparia Donati, "Introduzione. Prova giudiziaria e ragionamento artificiale: alcune possibili chiavi di lettura", in J. Sallantin, J.J. Szezeciniarz (a cura di), *Il concetto di prova alla luce dell'intelligenza artificiale*, Giuffré, Milano, 2005, pp. VII ss.

Dunque, come organizzare correttamente l'accesso di questo peculiare tipo di prova tecnologica nell'ambiente del processo penale, al fine di implementare la qualità delle *performance* cognitive e decisionali del giudicante, assicurandone nel contempo la legalità e l'utilizzabilità? Come garantire il legittimo spazio per l'esercizio del diritto di difesa, attraverso il confronto dialettico, la confutazione, la prova contraria, il dubbio, e perciò il contraddittorio effettivo fra le parti "*per*" e "*sulla*" prova, in funzione della validazione scientifica del risultato e contro la deriva algoritmica della giurisdizione?

Si apre, all'evidenza, uno scenario inedito nel quale il legislatore nazionale, per primo, è chiamato a dislocare, con "*regole dettagliate*", i poteri e le funzioni di accertamento, selezione e decisione della competente autorità giudiziaria, nell'esercizio di una discrezionalità tecnica, che sia guidata dal sapiente e prudente ricorso a categorie valoriali e clausole aperte.

Ad esempio, rispetto alla definizione della stretta necessità e proporzione dell'uso del sistema di IA in linea di principio vietato, alla valutazione d'impatto sui diritti fondamentali e sulle libertà delle persone, al governo della procedura autorizzatoria, alla debita giustificazione della eventuale urgenza, alla probabilità e all'entità del danno dal mancato uso ecc., sembrano invero insufficienti i criteri enunciati dalla Corte Suprema statunitense nella nota sentenza *Daubert*[5], in base ai quali il giudice deve vagliare l'effettiva affidabilità di una teoria o un metodo e di una *expert witness's scientific testimony*, ai fini della loro ammissibilità come prova scientifica nel processo: la controllabilità mediante esperimenti; la falsificabilità mediante test di smentita con esito negativo; la *peer review* della comunità scientifica di riferimento; la conoscenza della percentuale di errore dei risultati; infine, il criterio subordinato e ausiliario della generale accettazione da parte della comunità degli esperti. Criteri, tutti, sostanzialmente condivisi e anzi arricchiti dalla Corte di cassazione italiana[6].

[5] Daubert v. Merrel Dow Pharmaceuticals, Inc., 509 US 579 (1993).

[6] Cass., Sez. IV, 17/09/2010, n. 43786, Cozzini, sull'attendibilità di una teoria relativa all'eziologia del cancro dovuta ad esposizione all'amianto; Sez. Un., 11/09/2002, n. 30328, Franzese, sulle leggi scientifiche di copertu-

Con riguardo alla fase delle indagini preliminari, che com'è ormai diffusamente percepito dagli osservatori, è assurta a indiscusso baricentro anche mediatico del processo, sembra opportuno prendere le mosse dalla saggia scelta del legislatore della recente riforma *Cartabia* di riportare in (un seppure ancora parziale) equilibrio i rapporti fra il pubblico ministero e il giudice, mediante il riconoscimento di una serie di rimedi di tipo ordinatorio affidati al giudice per le indagini o della udienza preliminare – le cd. "*finestre di giurisdizione*" – in taluni momenti topici delle indagini.

Non sembra dubbio che, soprattutto laddove l'azione investigativa di contrasto si concreti nell'utilizzo di strumenti di IA vietati o ad alto rischio, potenzialmente pervasivi dei diritti fondamentali e delle libertà individuali delle persone, la previsione di clausole aperte e deroghe rinvii ad operazioni logiche di verifica e valutazione discrezionale di fatti e circostanze e di bilanciamento di valori più propriamente pertinenti alla figura del giudice, anziché a quella del pubblico ministero, che è parte benché pubblica.

Con riguardo alla fase del giudizio, risulta davvero efficace rinviare la verifica di coerenza e attendibilità di questa speciale prova scientifica e tecnologica al contraddittorio fra le parti "*sulla*" prova, cioè quando essa sia stata già ammessa e acquisita? Non sarebbe più utile costruire, nell'organizzazione degli snodi procedimentali, un filtro di accesso, preventivo e a maglie strette, e un agile contraddittorio "*per*" la prova, al fine di escludere addirittura che entrino nel patrimonio probatorio informazioni non sorrette da legittima validazione di attendibilità e utilizzabilità? L'irruzione nel giudizio di una siffatta, potente e incisiva, prova tecnologica giustificherebbe quindi la previsione di uno spazio dialettico preliminare già nel momento e in funzione dell'ammissione, prima ancora che della valutazione del risultato probatorio.

In questa direzione potrebbe rinvenirsi nel sistema processuale vigente la disposizione dell'art. 189 cod. proc. pen., per cui l'apprezzamento di rilevanza, non superfluità e concreta ido-

ra del nesso causale tra la condotta omissiva del medico e l'evento lesivo in danno del paziente.

neità della prova "*ad assicurare l'accertamento dei fatti*" – senza che ne resti pregiudicata "*la libertà morale delle persone*" per il divieto di perizia criminologica *ex* art. 220, comma 2 – è rimesso al vaglio preliminare del giudice. Questi, anche al fine di garantire l'anticipata conoscenza circa le metodologie che saranno applicate nell'accertamento, dopo avere sentito le parti sulle modalità di assunzione della prova, provvede all'ammissione con ordinanza, fissando le regole per la corretta applicazione dei metodi e delle procedure tecniche di acquisizione della stessa. Va segnalato, in proposito, un passo della Relazione al Progetto preliminare del nuovo codice di procedura penale del 1989 (p. 60), riguardante la portata dell'art. 189: "*È sembrato che una norma così articolata possa evitare eccessive restrizioni ai fini dell'accertamento della verità, tenuto conto del continuo sviluppo tecnologico che estende le frontiere dell'investigazione, senza mettere in pericolo le garanzie difensive*". Norma, questa, che nell'*intentio legis* mirava ad assicurare, con l'apporto della scienza e della tecnologia nella ricerca della verità, l'opportuna flessibilità del sistema processuale. Come si vede, un filtro, questo dell'art. 189, a maglie ben più strette rispetto a quello previsto dall'art. 190, comma 1, che, ai fini dell'ammissione della prova in genere, si limita a selezionare negativamente solo "*le prove vietate dalla legge e quelle che manifestamente sono superflue o irrilevanti*", e inoltre assistito da un significativo rafforzamento del contraddittorio anticipato, "*per*" la prova, ancor prima che "*sulla*" prova.

6. Una difficile transizione

In una equilibrata interazione tra giurisdizione e sistemi di intelligenza artificiale, secondo il metodo della *Hybrid AI*, giuristi e algoritmica sono chiamati a costruire un modello di conoscenza condiviso e raffinato, nel solco dei valori etici – di stampo europeo – di spiegabilità, trasparenza, controllo umano significativo e autonomia del decisore[7].

[7] N. Lipari, "Il diritto del nuovo millennio tra giurisdizionalizzazione ed algoritmo", in *Accademia*, 2024, n. 4, pp. 9 ss.

Il disciplinamento, etico e giuridico, del fenomeno ad opera del Regolamento europeo consente di reggere l'urto impetuoso e intrigante, nel contesto del processo penale in particolare, dei pervasivi strumenti dell'intelligenza artificiale, senza che ne risulti sconvolto – almeno finora – il tradizionale paradigma razionalistico della struttura e della funzione della giurisdizione penale, insieme con l'equità, l'efficacia e le garanzie del modello del giudizio *reasoning under uncertainty* e *by probabilities*.

Il diritto penale del fatto è incentrato sulla persona umana e il metodo dialettico dell'accertamento resta affidato, in ultima istanza, all'opera valutativa e decisoria dell'uomo, fallibile ma sindacabile secondo legge e ragione.

Gli esiti della sfida lanciata dal progresso della scienza, tuttavia, non sono affatto scontati. Al bivio tra tecnologia e tecnocrazia, essa si sposta sul terreno della concreta efficacia e qualità della giurisdizione, che, sulla base di regole puntuali e chiare dettate dal legislatore, dovrebbe essere esercitata secondo una dinamica interazione fra i saperi e le operazioni logiche tradizionalmente affidate al giudice e alle parti e le evidenze della prova o del calcolo di un algoritmo, in un orizzonte culturale di intesa su obiettivi e valori responsabilmente condivisi.

I principi di spiegabilità, trasparenza e sorveglianza umana, racchiusi in norme a rilevanza sovranazionale, hanno bisogno di essere concretamente attuati. Non sono certo poche le questioni controverse circa la portata e l'effettiva tenuta dei nuovi principi e regole, per la cui soluzione occorrerà prestare la doverosa attenzione allo sviluppo delle prassi applicative e interpretative.

Nel testo definitivo del disegno di legge governativo n. 1146, approvato con la legge n. 132 del 2025, benché obiettivamente migliorato rispetto alla proposta originaria, appaiono tuttora persistenti lacune e criticità della disciplina dell'impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria.

L'art. 15 (già art. 14) ammette innanzitutto che "è sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti" (comma 1), sembrando quindi escludere il ricorso a strumenti cd. predittivi con riguardo a quello che costituisce il nucleo centrale della giurisdizione.

Demanda poi al Ministero della giustizia (commi 2-3-4) sia la disciplina degli impieghi dei sistemi di intelligenza artificiale "per l'organizzazione dei servizi relativi alla giustizia, per la semplificazione del lavoro giudiziario e per le attività amministrative accessorie", sia il potere autorizzatorio per "la sperimentazione e l'impiego dei sistemi di intelligenza artificiale negli uffici giudiziari ordinari", sia infine l'elaborazione di linee programmatiche e la promozione di specifiche attività didattiche nella "formazione" dei magistrati e del personale amministrativo.

Non si rinvengono, viceversa, prescrizioni puntuali in merito ai poteri, ai divieti, alle condizioni, ai limiti e alle procedure dirette ad acquisire, utilizzare e valutare le informazioni generate dai sistemi di intelligenza artificiale nelle indagini e nel giudizio penale.

L'art. 24 (già art. 22), nel rinviare a "uno o più decreti legislativi per l'adeguamento della normativa nazionale al regolamento (UE) 204/1689" (art. 1, comma 1), avverte che "il Governo è altresì delegato ad adottare entro dodici mesi dalla data di entrata in vigore della presente legge uno o più decreti legislativi per adeguare e specificare la disciplina dei casi di realizzazione e di impiego illeciti di sistemi di intelligenza artificiale" (comma 3). E però, i principi e criteri direttivi elencati – fra i quali la "previsione di un'apposita disciplina per l'utilizzo di sistemi di intelligenza artificiale per l'attività di polizia" (comma 2, lett. h), la "regolazione dell'utilizzo dei sistemi di intelligenza artificiale nelle indagini preliminari, nel rispetto delle garanzie inerenti al diritto di difesa e ai dati personali dei terzi, nonché dei principi di proporzionalità, non discriminazione e trasparenza" (comma 5, lett. e), la "modifica, a fini di coordinamento e di razionalizzazione del sistema, della normativa sostanziale e processuale vigente, in conformità ai principi e ai criteri enunciati nelle lettere a), b), c) d) ed e)" (comma 5, lett. f) – sembrano invero generici rispetto all'obiettivo di puntuale ed effettiva armonizzazione del sistema interno in materia di indagini e giudizio penale.

Occorre essere consapevoli che la frammentazione e la lentezza del processo legislativo nazionale, insieme con la scarsa chiarezza della regolamentazione degli usi dei sistemi di intelligenza artificiale nel procedimento e nel processo penale, pos-

sono costituire un serio rischio che l'Italia perda terreno nella evoluzione, nella sperimentazione e nell'addestramento delle più moderne generazioni di quei modelli o strumenti tecnologici, all'esito di una lunga – forse troppo lunga – fase che riguarda il mondo scientifico ed economico e mette a dura prova quello giuridico su diritti, garanzie, rischi e sanzioni.

Santo Davide Ferrara

Il futuro artificiale delle Scienze Bio-Medico-Legali e Forensi

Abstract: **The Artificial Future of Bio-Medico-Legal and Forensic Sciences.** The historical evolution of Legal Medicine implies a progressive and evolutionary thickening of delicate fragments of specialised knowledge over the centuries. The knowledge of Bio-Medico-Legal and Forensic Sciences (BML/FS), predominantly stemmed from experimental research, plays a secondary role in significant biomedical innovation and manifests application limits and "Black Holes". In such a context, the BML/FS will increasingly benefit in the future from the growing maturity of Omics Sciences and Imaging Platforms. From this integration process the Radiomics will be no longer a hypothesis and an Evidence Based Justice can be fully revealed, surpassing the "blind justice" derived from expert opinions. Moreover, the virtual link between research and daily practice, governed by Organoid Super Intelligence (OSI), creativity endowed, will probably dominate in the foreseeable future, where the "Human Factor" will prove to be almost irrelevant in forensic daily practice and in scientific research. Hopefully, both dominated by "Humanitas", committed to affirming principles and values recommended by "P5 Medicine and Justice", crucial in the ethical control of Artificial Intelligence. The bio-medico-legal and forensic, scientific and ethical mission of the third millennium, should involve: *i)* the historical restoration of the "Unitariness" of disciplinary knowledge; *ii)* the daily application of praxis and education of the principles of "Learning to Doubt", and of 'Cultivating the constant pursuit' both the "Recognition of Error" and of the "Hope for Creating" better future scenarios.

Keywords: Artificial Intelligence, Legal Medicine-Future, Forensic Sciences-Future, Medicine and Justice, Human Factor.

Sommario: 1. Premessa – 2. Il contesto delimitato delle Scienze Bio-Medico-Legali e Forensi – 2.1. La storia della Medicina Legale – 2.2. L'attualità della Intelligenza Artificiale nelle Scienze Bio-Medico-Legali e Forensi – 2.3. Il futuro della Intelligenza Artificiale nelle Scienze Bio-Medico-Legali e Forensi – 3. Il contesto generale delle Scienze Biomediche e il ruolo delle Comunità Scientifico-Accademiche nel rapporto con l'Intelligenza Artificiale.

1. Premessa

Alla stregua di molteplici altre filiazioni, scaturite da ricerca scientifica applicata, la comunità delle Scienze Bio-Medico-Legali e Forensi (SBML/F) vive oggi la paradossale ambivalenza della *fiducia* e del *rischio*.

Fiducia nel consolidamento dei propri saperi ultraspecialistici. *Rischio* di estinzione della propria identità. Entrambi, rispettivamente, in virtù o a causa della incombente e nuova "connotazione creativa" della Intelligenza Artificiale.

Al cospetto di tale paradossale e quasi ontologica ambivalenza, è da tempo legittimo interrogarsi se la Scienza stia progredendo o regredendo. *Progredendo* verso incrementali consolidamenti "verticali" della profondità dei propri saperi ultraparcellari, volti, per loro intrinseca natura, all'inevitabile antitesi della "unitarietà identitaria". *Regredendo* dalla "orizzontale" e pervasiva conquista di scoperte distruttive di saperi consolidati e di verità rivoluzionarie di valenza epocale, estensibili a qualunque scenario ed orizzonte disciplinare. Comunque volte, per loro intrinseca natura, alla "unitarietà identitaria transdisciplinare" del sapere scientifico umanistico e sperimentale, esitato da attività di ricerca, teoretica, di base od applicata.

2. Il contesto delimitato delle Scienze Bio-Medico-Legali e Forensi

"Unitarietà identitaria" pressoché definitivamente persa dalla Medicina Legale, un tempo fregiata di ricchezza umanistica, derivata da confronto post-illuministico con le scienze antropologiche, giuridiche e sociologiche. Ricchezza umanistica vieppiù corrosa dalla incalzante deculturalizzazione a base scientista prevalente nella società tecnocratica occidentale, affetta da generale declino delle svariate quanto nobili declinazioni della *Humanitas* e dalla pressoché totale e umiliante perdita di identità disciplinari. Quelle medesime identità che, coniugate alla conoscenza del proprio contesto storico-culturale, inducono l'indipendenza dell'*Uomo di Scienza*, quale autentico *Ricercatore della Verità*, affrancato dalla funzione di specialista mero esecutore di tecno (e talora pseudo) -scienza. Con ciò convin-

tamente citando alcune riflessioni di "filosofia della fisica" di Albert Einstein, secondo il quale "la scienza senza epistemologia è confusa e priva di orientamento finalistico; l'epistemologia senza scienza è uno schema vuoto"[1].

La storica evoluzione, l'attualità della dicotomizzazione, le prospettive future della Medicina Legale e delle sue gemmazioni disciplinari, ne sono emblematica testimonianza. E ad esse è dedicata la sintetica esposizione delineata di seguito, prevalentemente tratta da un personale contributo di riflessioni offerte di recente all'attenzione e al dibattito internazionali[2].

2.1. La storia della Medicina Legale

È innanzitutto acquisito che la storica evoluzione della Medicina Legale testimonia di un progressivo ed evolutivo infittirsi plurisecolare di fragili frammenti di saperi parcellari[3]. Dai primi scritti del terzo millennio a.C., attraverso le tappe intermedie del Codice di Hammurabi (1700 a.C.), del Giuramento di Ippocrate (460 a.C.), del Codice Giustinianeo (483-565 d.C..), dei primi trattati disciplinari (1200/1600 d.C.), fino al consolidarsi di un robusto sapere unitario e identitario che dischiuse la nascita della Medicina Legale moderna nel XIX secolo. Per poi risultare, nella metà del '900, in un rinnovato processo di frammentazione e nella creazione di incipienti nuove discipline, il cui parcellare ma identitario sapere si è convertito in valenza

[1] K. Fox, *I Am a Part of Infinity: The Spiritual Journey of Albert Einstein*, Basic Books, New York, 2025, pp. 336.

[2] S.D. Ferrara, "A New Way Back to an Artificial Future for BioMedicoLegal and Forensic Sciences", in *Forensic Science International*, 370 (2025), n. 112458, pp. 1-3.

[3] S.D. Ferrara, H. Pfeiffer, "Unitariness, evidence and quality in bio-medicolegal Sciences", in *International Journal of Legal Medicine*, 124 (2010), pp. 343-344; S.D. Ferrara, "P5 Medicine and Justice: the future is now", in *International Journal of Legal Medicine*, 131 (2017), pp. 897-899; S.D. Ferrara, G. Cecchetto, R. Cecchi, D. Favretto, S. Grabherr, T. Ishikawa, T. Kondo, M. Montisci, H. Pfeiffer, M.R. Bonati, D. Shokry, M. Vennemann, T. Bajanowski, "Back to the Future - Part 2. Post-mortem assessment and evolutionary role of the bio-medicolegal sciences", in *International Journal of Legal Medicine*, 131 (2017), pp. 1085-1101; S.D. Ferrara (ed.), *P5 Medicine and Justice*, Springer International Publishing, Heidelberg, 2017.

di scienza. E, quindi, dal *Mondo*, apparentemente unitario della Medicina Legale, all'*Universo* delle SBML/F, affette da diversità e da apparente unitarietà, fondata sulla "sistematica" della Metodologia di Accertamento e della Criteriologia di Valutazione.

2.2. L'attualità della Intelligenza Artificiale nelle Scienze Bio-Medico-Legali e Forensi

È oggi inoltre consolidato che il sapere delle SBML/F sia prevalente espressione della ricerca sperimentale, condotta nel novero di Patologia, Genetica, Tossicologia, Antropologia, Criminologia e Criminalistica. Mentre solo una frazione minoritaria proviene dalla Clinica, in tema di Clinica Medico-Legale, *Malapraxis* e Responsabilità Professionale, Lesioni Personali e Danno alla Persona, Invalidità e Psichiatria. Nonché dalla Dottrina, in tema di Etica, di Legislazione e Giurisprudenza Medica. È altresì acquisito e consolidato che i *trend* di fattori o valori di impatto vedano la Genetica al vertice, seguita, nell'ordine, da Tossicologia, Entomologia e Criminologia, Patologia, Antropologia, Criminalistica, Odontostomatologia e Palinologia. È infine acquisito e consolidato che il *contributo* totale di *conoscenza* apportato dalle SBML/F al *Sapere Bio-Medico è* significativamente *inferiore* a quello di altri settori disciplinari.

Dal suddetto inconfutabile squilibrio emerge uno *stato di complementarietà e di dipendenza*, fin qui non esplicitamente acclarato. Mai protagoniste nella consistente innovazione del sapere bio-medico, le SBML/F rivestono un ruolo da mere *comprimarie* in costanti crisi di astinenza, necessitanti della linfa del sapere della scienza, in generale, e della scienza bio-medica, in particolare. E ciò ai fini della soluzione dei propri problemi e del superamento dei propri limiti applicativi. I quali si configurano come *buchi neri* del sapere e come incapacità di disporre di affidabili "Marcatori" di malattia, di danno, di datazione, di identificazione personale e di cause di morte.

Nel contesto della attuale incapacità di apporto innovativo autonomo alla Bio-Medicina e alla soluzione dei propri problemi applicativi, le SBML/F godono e, ancor più, godranno in futuro della crescente maturità delle *Scienze Omiche* e delle

Piattaforme di Immagini, alimentate e governate dalla *Intelligenza Artificiale*. Le quali, già separatamente, contribuiscono ad acquisire una nuova tipologia di *Evidenza,* circa la natura e l'età delle lesioni e circa il danno a carico di strutture osteo-mio-vascolari e di tessuti molli, sia nel vivente che nel cadavere. Per mezzo di *In Vivo and Post Mortem Cross Sectional Imaging*. Ovvero, mediante "Tecniche e Metodi" integrati in uniformi e condivisi "Algoritmi" di "Modelli Quantitativi" e non più solo "Qualitativi", come nel recente passato.

La risultante e crescente implementazione della *Imaging Mass Spectrometry*, integrata con lo sviluppo della *Proteo-Metabolomica*, nel creare un *link* sempre più stretto tra discipline (Radiologia, Patologia, Tossicologia…), implica e prospetta una correlazione sempre più intensa fra "*imaging proteomico post mortem* e corrispondenti *dati anatomici in vivo*".

2.3. Il futuro della Intelligenza Artificiale nelle Scienze Bio-Medico-Legali e Forensi

Da tale processo di integrazione deriva che la *Radiomica* non è più una mera ipotesi. L'alto potere di estrazione delle caratteristiche peculiari delle immagini radiografiche consente già di derivare espressioni genotipiche e fenotipiche. Anche per la *Sistematica di Patologia* ove possono essere identificati cause e mezzi di produzione della morte, età, vitalità, ordine e correlazione di lesioni interne ed esterne. Parimenti per la *Clinica Medico-Legale* ove, dalla integrazione dei "bio-marcatori omici" potranno derivare "Indicatori" di identificazione di: - stato, attuale e futuro, di benessere psicofisico (mediante metabolomica e genomica, rispettivamente); - valutazione del danno attribuibile a specifico evento; - formazione di *baremes omici*, capaci di quantificare il grado di danno attuale e futuro, comprensivo di quelli derivanti da disturbi post-traumatici, dolore e depressione. Così come per la *Sistematica di Psichiatria Forense e Antropologia*, ove la medesima integrazione strategica della ricerca radiomica chiarirà la complessa interrelazione fra "Natura, Ambiente, Epistasi ed Epigenesi" del Comportamento umano. Egualmente per la *Sistematica di Tossicologia*, nella quale la "nanorobotizzazione chemometrica" e il "laboratorio in un chip" hanno già

dischiuso un nuovo paradigma, ove una diretta "radiografia molecolare" nel tessuto biologico e un costituendo *Toxoma Umano* avvieranno una nuova Tossicologia del Comportamento ed una evolutiva Medicina Personalizzata, nel novero della *P5 Medicina e Giustizia*[4].

In virtù della *Radiomica*, il vero valore della *Evidenza* sarà finalmente acquisito ed interpretato come Accuratezza, equivalente a Verità e Realtà. Così, archiviando la *Giustizia Cieca* della soggettività derivante dalle opinioni degli esperti, sostituita dalla definitiva affermazione della *Giustizia* fondata sulla oggettività della *Evidenza.*

Purtroppo, noi siamo solo cultori di scienza ed esercenti una pratica professionale. Forse, solo ed appena, *Archeologi del Futuro*. E la visione del nostro tempo è quella di un "impossibile domani". Tuttavia, come disse John Maynard Keynes "L'inevitabile non accade mai… l'inatteso sempre!"[5].

Pertanto, possiamo già collocare nel passato, per quanto prossimo, l'impiego a "prova di idiota" della *Artificial Narrow Intelligence* (ANI) e della *Artificial General Intelligence* (AGI). E ciò, malgrado la loro iniziale applicazione in Modelli Predittivi capaci di identificare vitalità ed età, impronte digitali, DNA, dati tossicologici ed ulteriori molteplici dati in tanatologia, medicina legale clinica e tossicologia. Invero, è abbastanza facile prevedere che il "rapporto virtuale" tra ricerca e pratica, governata dalla *Organoid Super Intelligence* (OSI, basata su reti neurali biologiche e dotata di creatività) dominerà nel prevedibile futuro. Auspicabilmente controllata dalla umanità, o, per meglio dire, dalla *Humanitas*, deputata ed impegnata ad affermare principi e valori preconizzati dalla *P5 Medicina e Giustizia*[6].

Così realizzando l'*Etica dell'Intenzione e della Responsabilità*, invocata da Max Weber nella "Traiettoria della modernità"[7].

[4] S.D. Ferrara, "P5 Medicine and Justice: the future is now", cit.; S.D. Ferrara (ed.), *P5 Medicine and Justice*, cit.

[5] T. Ball, R. Bellamy (eds.), *The Cambridge History of Twentieth-Century Political Thought*, Cambridge University Press, Cambridge, 2006, p. 46.

[6] S.D. Ferrara, "P5 Medicine and Justice: the future is now", cit.; S.D. Ferrara (ed.), *P5 Medicine and Justice*, cit.

[7] B.K. Quensel, *Max Webers Konstruktionslogik. Sozialökonomik zwischen Geschichte und Theorie*, Nomos, Baden-Baden, 2007.

L'*Attività Forense Umanitaria* è già e continuerà ad essere uno splendido esempio di nuove finalità per le Scienze Bio-Medico-Legali e Forensi impegnate a: - proteggere la vita, il benessere e la dignità dell'uomo; - provvedere al miglior soccorso dei bisognevoli; - affinare la formazione delle competenze utili a fronteggiare disastri naturali e crisi umanitarie.

In sostanza, alcune quasi inconfutabili conclusioni emergono dal contesto attuale: 1. nella pratica medico legale e forense il *fattore umano* sarà pressoché *irrilevante* ai fini della acquisizione della evidenza radiomica; 2. parimenti pressoché *irrilevante* sarà il *fattore umano* nella conduzione della ricerca scientifica bio-medico-legale e forense; 3. il *fattore umano* dovrà essere essenziale, anzi *cruciale*, nel controllo etico della *Organoid Super Intelligence*; 4. in tale processo evolutivo il ruolo della Comunità Accademica e Scientifica Mondiale dovrà essere e potrà essere strategico e decisivo.

Tali personali conclusioni derivano da "epicrisi" delle attuali potenzialità scientifiche e tecnologiche già dischiuse alle discipline Bio-Medico Legali e Forensi.

Ma, come sempre, solo il tempo dirà se esse siano giuste o sbagliate. Ed io auspico fermamente che siano sbagliate!

3. Il contesto generale delle Scienze Biomediche ed il ruolo delle Comunità Scientifico-Accademiche nel rapporto con l'Intelligenza Artificiale

In ossequio al pensiero di F. Ferrarotti[8], è tuttavia indubitabile che la "Scienza e lo Scientismo" siano diventati una nuova Religione, che si considera giustificata in sé stessa e dotata di una validità immanente che non ha bisogno di imperativi etici trascendenti.

Invero, non avendo valenza finale, ma solo strumentale, l'innovazione scientifica e tecnologica non dice né da dove veniamo, né dove siamo, né dove andiamo. E, nel porsi per lo sviluppo fine a sé stesso, senza precisazioni, essa induce disorientamento ed incertezza, quali tratti distintivi della situazione attuale e, an-

[8] F. Ferrarotti, *Scienza e coscienza. Verità personali e pratiche pubbliche*, EDB, Bologna, 2014.

cor più, di quella futura. Nelle quali, secondo l'eredità del pensiero di Immanuel Kant[9] e di G.W. Friederich Hegel[10], si impone e vieppiù si imporrà uno stretto collegamento fra epistemologia e scienza. La finalità dell'uomo non può e non potrà, infatti, limitarsi alla sola descrizione e al mero desiderio di sapere come si evolverà il futuro, senza intervenire nel presente per modificarne l'evoluzione[11].

Ma, piuttosto, si imporrà di adottare una *strategia* nella costruzione coesa e coordinata degli *scenari futuri*[12].

Ed il cuore della sfida non risiederà, pertanto, nel progresso tecnologico in sé, ma piuttosto nella rinuncia o meno, da parte degli esseri umani, alla "concezione soggettivistica" e alla straordinaria potenzialità della loro intelligenza, per adagiarsi sulla comodità tecnologica e sulle facili scorciatoie che ne deriveranno. Astenersi dalla costruzione attiva di nuovi scenari sarebbe la via più sicura per consentire all'Intelligenza Artificiale di prendere il sopravvento.

Allora il problema cardine è e sarà: "come", ovvero, "in qual modo e verso quali obiettivi indirizzare gli scenari futuri". Nei quali all'esercizio dell'intelligenza logico-razionale si associ quello della capacità di capire la propria collocazione e di operare con gli altri e per gli altri. In accordo all'aforisma di F. Dostoevskij, "Per agire in maniera intelligente, l'intelligenza non basta"[13], lo sviluppo dovrà inevitabilmente fondarsi sull'equilibrio e sulla cooperazione.

E, come già detto, nel futuro delle Scienze Bio-Medico-Legali e Forensi il ruolo della Comunità scientifica ed accademica, delle Società o Associazioni, coniugate ai relativi e propri mezzi di

[9] I. Kant, *Kritik der reinen Vernunft*, J.F. Hartnock, Riga-Lipsia, 1781-1787 e 1788; H.F. Allison*, Kant's Theory of Freedom*, Cambridge University Press, Cambridge, 1990.

[10] G.W.F. Hegel, *Erster Teil die Phänomenologie des Geistes*, J.A. Goebhardt, Bamberg-Würzburg, 1807; D. Henrich, *Between Kant and Hegel, Lectures on German Idealism*, Harvard University Press, Cambridge, 2008.

[11] M. Serres, *Conversations on Science, Culture and Time*, University of Michigan Press, Ann Arbour, 1995.

[12] S. Tagliagambe, "Costruire scenari per il futuro", in *Ethics in Progress*, 8 (2017), n. 1, pp.125-155.

[13] F. Dostoevskij, *Преступление и наказание*, Russkij vestnik, Moskva, 1866.

comunicazione (di cui l'*International Association of Forensic Sciences*, l'*International Academy of Legal Medicine*, il *Forensic Science International*, l'*International Journal of Legal Medicine*, il *Forensic Sciences Research Journal* sono alcune esemplari espressioni) si porrà come decisivo nell'apertura di *nuovi scenari*, auspicabilmente guidati dai principi e dai valori etici della più autentica *Humanitas.* Migliore manifestazione della civiltà classica greco-romana, quale inesauribile fonte primigenia della cultura occidentale.

A questo riguardo, in un famoso libro del 1959, "Le Due Culture", Charles Percy Snow, fisico e scrittore, stigmatizzò il contrasto tra cultura scientifica e umanistica[14]. Due Culture che oggi come allora vedono spesso l'altra con sospetto e superiorità. E ciò malgrado, già nel IV secolo a.C., Platone avesse iscritto all'entrata della sua Accademia "Non entri chi è ignorante di geometria". Ed ancora, malgrado nel XVIII secolo, un letterato, D. Diderot, ed un matematico, J.B. d'Alembert, avessero entusiasticamente collaborato alla stesura di un universale compendio di cultura: la *Enciclopedia*, ovvero il *Dizionario ragionato delle Scienze, delle Arti e dei Mestieri*[15].

In analoga guisa, il riconoscimento antropologico della cultura segnò una grande scoperta scientifica ed il passaggio dalla mera osservazione dei costumi ad una "organica visione della cultura" concepita come la risultanza sistemica di molteplicità di culture.

Come F. Jullien[16] e R. Wagner[17] reclamarono, le Culture furono, erano, sono e saranno "aperte", in quanto attraversate da linee di faglia, fratture e fessurazioni in continua trasformazione, quali risultati di differenze di genere e di età, di conflitti generazionali, di differenziali di poteri e di ricchezze. In altri termini,

[14] C.P. Snow, *The Two Cultures: a second look*, Cambridge University Press, London, 1963.

[15] D. Diderot, J.B. Le Rond D'Alembert, *Enciclopèdie, ou Dictionnaire raisonné des Sciences, des Arts et des Mètiers, par une Societé de Gens de Lettres*, Le Breton, Paris, 1751.

[16] F. Jullien, *De l'Etre au Vivre. Lexique euro-chinois de la pensée*, Gallimard, Paris, 2015.

[17] M. Brighenti, *Richard Wagner filosofo e poeta. La filosofia in musica fra Feuerbach, Schopenhauer e Nietzsche*, Aracne, Roma, 2016.

la "fecondità culturale" di Jullien, correlata e non antagonista della "identità culturale", dalla quale scaturiscono formidabili mezzi di innovazione di crescita della civiltà, per nulla ovvia, ma costruita giorno per giorno attraverso e mediante un lento e progressivo lavoro educativo, volto a creare valori condivisi, contrari al diffusivo Medioevo post-moderno della contro-civilizzazione. A conferma che il dono di Prometeo non è abbastanza, in quanto l'Uomo prometeico, dotato di sola tecnologia è bisognevole di un ampio orizzonte di valori etico-politici di giustizia, di legislazione e di educazione collettiva[18].

Anche se gli errori giudiziari accadono per varie ragioni, una cultura scientifica mediocre in coloro che giudicano o legiferano produce solo una peggiore giustizia. Per la cui migliore attuazione, in "*de iure condendo et condito*", l'educazione scientifica è imprescindibile "*conditio sine qua non*".

Al fine di evitare l'*Homo juridicus*[19], rozzo modello di attore guidato da riflessi pavloviani, affetto da comportamenti di evitamento nell'associazione fra sanzioni e mancato rispetto della legge.

Ma, piuttosto, per orientarsi verso fini comuni, in accordo alla teoria del *Nudge*[20] e secondo l'Etica dell'influenza[21] preconizzata da R.H. Thaler e C.R. Sunstein, ed anche, al fine di prevenire "l'uomo demolito", disintegrato, di A. Bester[22]. "Uomo" analizzato, dicotomizzato, e quasi del tutto rivelato nei più intimi pensieri, sogni ed intenzioni, mediante nuovi "sistemi kamitani, vocabolari del cervello e traduttori di pensieri", realizzati per primi da J. Gallant[23] e J. Henderson[24] e poi, via via, perfe-

[18] M. Vegetti, *L'etica degli antichi*, Laterza, Bari-Roma, 2006.

[19] G. Zaccaria, "Epistemological and Ethical Implications of the Personalized Justice", in S.D. Ferrara (ed.), *P5 Medicine and Justice*, cit., pp. 36-45.

[20] R.H. Thaler, C.R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Yale University Press, New Haven, 2008.

[21] C.R. Sunstein, *The Ethics of Influence. Government in the Age of Behavioral Science*, Cambridge University Press, Cambridge, 2016.

[22] A. Bester, *The Demolished Man*, Shasta Publishers, Chicago, 1953.

[23] T.M. Mitchell, S.V. Shinkareva, A. Carlson, K. Chang, V.L. Malave, R.A. Mason, M.A. Just, "Predicting Human Brain Activity Associated with the Meanings of Nouns", in *Science,* 320 (2008), n. 5880, pp. 1191-1195

[24] J. Henderson, "Learning Word Embeddings for Hyponymy with Entailment-Based Distributional Semantics", in *arXiv*, (2017), 1710.02437,

zionati da altri cultori di Radiomica. E tutto ciò a dispetto del "diritto alla privacy del pensiero" invocato dai bioeticisti e del "principio della irriconoscibilità dell'inosservato" postulato da alcuni filosofi.

Di fronte agli scenari distopici suscitati da tali innovazioni dovrebbe essere compito degli scienziati ed obiettivo della civilizzazione rispettare la relazione "cervello-mente", indirizzandola verso il bene comune. Con ciò conseguendo il paradosso di Protagora, secondo cui tutte le opinioni sono legittime, in quanto nessuno può reclamare di possedere l'assoluta verità. Consapevoli che l'attuale *condizione neomoderna* implica che "ragione e scienza" non possono conseguire il tutto, ed anzi, talora, incorrono in errore. Ma, anche, che nulla può essere costruito prescindendo da esse. In ciò attualizzando il raccordo tra filosofia (o scienza) naturale e teologia, invocato da San Tommaso allorché previde che la scienza avrebbe prodotto inconfutabili[25] evidenze che, solo in tempi recenti, inizia in parte a possedere. Al fine di consolidare, mediante differenziati approcci, un intenso e produttivo dialogo tra scienza e teologia, celebrato dall'enciclica papale *Laudato si'*[26], ripreso da R. Horton[27]. E per consentirci di essere profondamente pervasi dalle migliori conquiste della scienza, nonché allo scopo di annunciare che "è finalmente arrivato il tempo in cui uomini di scienza e uomini di fede concorrono insieme nella ricerca della verità, al servizio della medesima, evitando i dogmatismi del passato". Preservando la "naturale ecologia", capace di: salvaguardare l'originale identità di ciascun individuo e di ogni gruppo sociale; conferire ai Ricercatori un ruolo preminente ed una più ampia libertà accademica; coniugare le attuali peculiarità del pensiero orientale e occidentale, solo in apparenza dissimili.

doi:10.48550/arXiv.1710.02437.

[25] Santo Padre Francesco, *Laudato si'. Lettera enciclica sulla cura della casa comune*, Libreria Editrice Vaticana, Roma, 2015; R. Horton, "Life and Death in 2100: Health, History and Human Contingency", in *The 25th Annual Pumphandle Lecture*, London, 2017.

[26] Santo Padre Francesco, *Laudato si'*, cit.

[27] R. Horton, "Life and Death in 2100: Health, History and Human Contingency", cit.

In sostanza, una *totalità ontologica* distante e quasi opposta alle tecnoscienze, tese a disconoscere la morte quale loro sconfitta. Anticipando la subentrante "dialettica dell'illuminismo" di M. Horkheimer e di T. Adorno[28], il pensiero orientale rivela che "la tecnoscienza non conosce la verità, ed anzi la evita come Metafisica; al contrario, essa non può conoscere cosa siano realmente la morte e l'angoscia per essa".

Nel campo della cultura il metro di paragone è da lungo tempo rappresentato dai canoni della civilizzazione occidentale. I così definiti "etnocentrismo" e "colonialismo" si pongono quale enorme intento di normalizzare il mondo, imponendo canoni e metri all'uopo forgiati in modo specifico. E la globalizzazione è l'esito di tale lavoro, di cui il Web è esempio di massificazione nella conformità. In realtà "le culture non sono misurabili, ma solo comparabili tra loro". E, come arguisce l'antropologo M. Kilani[29], è necessario armonizzare differenze e variabilità, recuperando una "relatività culturale" che nulla ha a che fare con il relativismo. Una relatività culturale ove "le culture si embricano" senza confini, in una diversità che è fondamento della innovazione, in accordo all'art.7 della Dichiarazione Universale dell'Unesco sulla diversità culturale del 2 novembre 2001 espressa anche nel pensiero di L. Piasere[30]. Ne sono contestuale ed esemplare conferma le Scienze Bio-Medico-Legali e Forensi, ove le diversità di disparate latitudini e longitudini svaniscono nella unitarietà della sistematica metodologica, criteriologica e dottrinaria.

Nella definitiva e totale consapevolezza che la missione scientifica ed etica, delle Scienze Bio-Medico-Legali e Forensi del terzo millennio, dovrebbe quantomeno implicare: 1) la storica restaurazione dell'*Unitarietà del sapere disciplinare*; 2) l'applicazione quotidiana della *praxis educativa*, rivolta e finalizzata a: a) *Imparare l'esercizio del Dubbio*; b) *Coltivare la costante*

[28] M. Horkheimer, T.W. Adorno, *Dialektik der Aufklärung. Philosophische Fragmente*, Querido Verlag, Amsterdam, 1947.

[29] M. Kilani, *Antropologia. Una introduzione*, a cura di A. Rivera, Dedalo, Bari, 2011.

[30] L. Piasere, *L'etnografo imperfetto. Esperienza e cognizione in antropologia*, Laterza, Bari-Roma, 2002.

Ricerca, sia dell'Errore che della Speranza della Creazione di scenari futuri migliori!

In un ordine di rinnovata civilizzazione nella quale le culture non sono concepite come espressioni di specie o di razze di cui si rimane prigionieri. Malgrado la competizione prosegua determinando che la normalità di oggi sarà presto oggetto di dubbio e di modifiche nella cultura umanistica e delle tecnoscienze di domani. Perpetuando le fondamenta di quel *logos* che rende umani a dispetto dei limiti computazionali. Del pensiero critico, della capacità di giudizio e di immaginazione, del desiderio di conoscere e di creare, dell'aspirazione all'*invisibilità della bellezza*, al cui cospetto i grandi scienziati sono inclini a vestirsi di modestia accettandone l'imperscrutabilità ed il mistero.

Riecheggiando Gramsci, si potrebbe concludere che la società moderna sia finita e che quella digitale sia all'inizio. E, pertanto, che in questa transizione odierna si possa fare la storia operando le scelte migliori. Attuando innanzitutto la rivoluzione della speranza, preminente virtù umana, smarrita negli ideali laici della modernità e capace di infondere il coraggio di affrontare il domani[31]. Coraggio di pensare più in alto e di creare l'*Incipit* di una dantesca *Vita Nova*. Rinnovato "culto della positività hegeliana", contrapposto all'"angoscia heideggeriana" prigioniera del passato. A fronte del quale la morale del presente e del futuro è e sarà "tormentata ricerca" in costante evoluzione nell'infinito gioco delle interazioni umane.

[31] B.C. Han, *Contro la società dell'angoscia. Speranza e rivoluzione*, Einaudi Stile Libero, Torino, 2025, p. 112.

Luca Lupária Donati

Giustizia tecnologica *vs.* tecnologia giudiziaria. Il processo penale al crocevia*

Abstract: **Technological Justice vs. Judicial Technology. The Criminal Trial at the Crossroads.** The enormous potential of IT tools can be differently leveraged as cement for safeguarding the *pillars* of our legal culture, rather than as a means of destructuring the geometries and the mechanisms of procedural guarantees. It would be, therefore, more appropriate to use the expression "judicial technology" rather than "technological justice".

Keywords: Criminal Trial, Judicial Technology, Fundamental Guarantees.

Sommario: 1. La giustizia tecnologica quale sicario silenzioso del sapere processuale e della cultura della prova – 2. Lo squadernamento degli istituti codicistici sotto i colpi della tecnologia – 3. Il rito penale tra bulimia conoscitiva e morigeratezza epistemica – 4. Tecnologia giudiziaria e rafforzamento delle garanzie fondamentali.

1. La giustizia tecnologica quale sicario silenzioso del sapere processuale e della cultura della prova

La giustizia tecnologica, nell'orizzonte culturale del processo penale, evoca da sempre più catastrofi che benefici[1].

Gli esempi non mancano. Le evidenze digitali (e scientifiche in genere) spostano la bussola dell'accertamento verso lo stadio

* Il presente contributo, che riprende alcuni dei miei più recenti scritti sul tema, è destinato agli studi in onore di Vincenzo Zeno-Zencovich.

[1] Cfr. G. Canzio, L. Lupária Donati, *Prova scientifica e processo penale*, III ed., Wolters Kluwer Cedam, Milano, 2025; F. Caprioli, "Tecnologia e prova penale: nuovi diritti e nuove garanzie", in L. Lupária Donati, L. Marafioti, G. Paolozzi (a cura di), *Dimensione tecnologica e prova penale*, Giappichelli, Torino, 2019, pp. 45 ss.; Id., "La scienza "cattiva maestra": le insidie della prova scientifica nel processo penale", in *Cass. pen.*, (2008), p. 3529; C. Cesari, "L'impatto delle nuove tecnologie sulla giustizia penale – un orizzonte denso di incognite", in *Rev. Bras. de Direito Processual Penal*, (2019), pp. 1167 ss.; G. Di Chiara, "Il canto delle sirene. Processo penale e modernità scientifico-tecnologica: prova dichiarativa e diagnostica della verità", in *Criminalia*, (2017), pp. 19 ss.; A. Garapon, J. Lassègue, *Justice Digitale: Révolution Graphique et Rupture Anthropologique*, Puf, Paris, 2018.

investigativo, provocando un parziale svuotamento del contraddittorio dibattimentale, divenuto, il più delle volte, una sorta di *divertissement* estetico a fronte di una partita già giocata *aliunde*[2]. Sotto diverso angolo visuale, le nuove tecniche investigative incidono sulla vita delle persone con gradi di profondità e spazi di estensione finora sconosciuti[3].

E ancora. L'ingresso nel circuito giudiziale degli strumenti informatici e delle macchine (più o meno) intelligenti genera forme di inedita deresponsabilizzazione degli attori processuali, in particolare del giudice[4]. Questi tende spesso ad affidarsi al *dictum* dell'esperto, al risultato dell'esperimento o alla decisione artificiale, non tanto per una mitizzazione della valenza degli enunciati, quanto per comodità (l'essere umano è pigro per natura, dicono gli scienziati…) o per mettere al sicuro dal rimprovero le proprie pronunce mediante l'adesione alle statuizioni oracolari dei *tools* a sua disposizione[5]. Del resto, un decisore umano che volesse discostarsi dalla proposta ricostruttiva del luminare di riferimento o, peggio ancora, dalla previsione for-

[2] V., per tutti, L. Marafioti, "Prueba digital y proceso penal", in *Revista de Derecho Penal y Procesal Penal*, (2012), pp. 1904 ss.; M. Pittiruti, Digital evidence *e procedimento penale*, Giappichelli, Torino, 2017.

[3] Cfr. F. Centorame, "Le indagini tecnologiche ad alto potenziale intrusivo fra esigenze di accertamento e sacrale inviolabilità dei diritti della persona", in *Riv. it. dir. proc. pen.*, (2021), pp. 499 ss.; M. Daniele, "La vocazione espansiva delle indagini informatiche e l'obsolescenza della legge", in *Proc. pen. giust*., (2018), p. 834; D. Negri, "Compressione dei diritti di libertà e principio di proporzionalità davanti alle sfide del processo penale contemporaneo", in *Riv. it. dir. proc. pen.*, (2020), pp. 3 ss.; F. Nicolicchia, *I controlli occulti e continuativi come categoria probatoria. Una sistematizzazione dei nuovi mezzi di ricerca della prova tra fonti europee e ordinamenti nazionali*, Wolters Kluwer Cedam, Milano, 2020; S. Signorato, *Le indagini digitali. Profili strutturali di una metamorfosi investigativa*, Giappichelli, Torino, 2018.

[4] Con la consueta lucidità, D. Pastor, "El futuro de la jurisdicción penal entre oficios temporales y máquinas de juzgar", in *Diario Penal*, (2017), n. 175, pp. 1 ss.

[5] Un quadro della tematica negli ottimi scritti di A. Carratta, "Decisione robotica e valori del processo", in *Riv. dir. proc.*, (2020), pp. 513 ss.; V. Manes, "L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia", in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffrè Francis Lefebvre, Milano, 2020, pp. 547 ss.

mulata dall'*artificial intelligence*[6] dovrebbe connotarsi – *comme il faut*, almeno in teoria – da forte indipendenza, senso di responsabilità, solidità psicologica e capacità di sopportare le critiche.

Mi spiego meglio, muovendo dalla casistica statunitense: se un applicativo di *risk assessment* suggerisse che l'arrestato non vada liberato su cauzione in quanto potenzialmente pericoloso[7], quanti magistrati, pur se intimamente convinti dell'esistenza di condizioni per il rilascio in pendenza di processo, si assumerebbero il rischio del biasimo che si riverserebbe su di loro se quel soggetto, rimesso a piede libero contro il "parere" della macchina, commettesse un grave illecito?

Più in generale, stiamo assistendo nel panorama comparativo, così come in Italia, quale conseguenza di quella che si suole appunto definire giustizia tecnologica, a una fuga dei giudicanti dalla componente essenziale del proprio mandato, vale a dire l'apprezzamento delle prove. Per valutare la credibilità di un testimone – nocciolo duro dell'attività di chi siede sullo scranno della giurisdizione – vengono oggi addirittura chiamati consulenti tecnici o periti, perché dicano se il teste afferma il vero o il falso, se sia attendibile o meno, magari avvalendosi di metodiche o di apparecchiature neuroscientifiche.

Insomma, l'emersione del significativo grado di fallibilità dell'atto del decidere, la scoperta dei meccanismi di contaminazione della memoria e della percezione umana – alcuni dei grandi approdi scientifici degli ultimi anni[8] – invece che indurre gli operatori ad alzare lo *standard* della propria preparazione, in termini di rafforzamento del bagaglio formativo ed epistemico di ciascuno, sta invece conducendo a delegare in bianco ad altri

[6] Cfr. J. Nieva-Fenoll, *Intelligenza artificiale e processo* [2018], trad. it., Giappichelli, Torino, 2019.

[7] Sulla parabola evolutiva statunitense: M. Gialuz, "Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa", in *Dir. pen. cont.*, 29 maggio 2019, pp. 1 ss.

[8] Cfr. P. Catellani, *Il giudice esperto. Psicologia cognitiva e ragionamento giudiziario*, il Mulino, Bologna, 1992; D. Kahneman, *Pensieri lenti e veloci*, Mondadori, Milano, 2012; L. Laudan, *Truth, Error and Criminal Law. An Essay in Legal Epistemology*, Cambridge University Press, Cambridge, 2006; L. Macchi, *La psicoretorica. Dall'arte del dire alla forma del pensiero*, Raffaello Cortina, Milano, 2024. R. Rumiati, *Saper decidere. Intuizioni, ragioni, impulsività*, il Mulino, Bologna, 2020.

quelle valutazioni che dovrebbero invece rimanere esclusive di colui al quale affidiamo il "giudizio dell'uomo sull'uomo" e l'affermazione della colpevolezza sopra "ogni ragionevole dubbio"[9].

Volendo procedere oltre in questa *via crucis* o, se si preferisce, in questa parabola storico-evolutiva, occorre anche ricordare come la tecnologia e le scienze abbiano introdotto pericolose isole di presunta infallibilità in un giudizio, quello penale, che, per sua natura deve invece nutrirsi di dubbi e di riscontri[10]. Elementi conoscitivi come la prova del DNA e altre *forensic sciences*, anche in ragione di una poca conoscenza dei rudimenti del calcolo della probabilità e del teorema di Bayes[11], appaiono, agli occhi dei giudicanti, eccessivamente concludenti, quasi decisivi[12]. Giudicanti, professionisti o laici che siano, spesso *bamboozled*, imbambolati dall'*ipse dixit*, come affermano gli studiosi nordamericani.

Molte ricerche lo hanno mostrato: in realtà, si tratta di esiti tecnico-scientifici tutt'altro che risolutivi per il *thema decidendum* e che, oltretutto, lungi dall'essere il frutto di un percorso automatizzato, celano al proprio interno scelte discrezionali di operatori che, nella statuizione finale, espressa magari con un numero, non appaiono immediatamente percepibili. Nel percorso che conduce all'*output* di un'analisi genetica, ad esempio, il tecnico può non limitarsi a immettere un campione nella macchina e ritirare il "biglietto" con il responso, come si tende a ritenere nell'immaginario comune. Egli opera spesso scelte metodologiche, decide quale profilo sia prevalente in caso di *mixtures*, talora amplifica artificialmente materiale di minima quantità, e via dicendo. Insomma, vengono compiuti passaggi interpretativi importanti che conducono all'esito finale, ma che non sono conosciuti o vagliati, anche in termini di margine di

[9] V., per tutti, G. Carlizzi, *Libero convincimento e ragionevole dubbio nel processo penale. Storia, prassi, teoria*, Bonomo, San Lazzaro di Savena, 2018, p. 105.

[10] Cfr. G. Carlizzi, *La valutazione della prova scientifica*, Giuffrè Francis Lefebvre, Milano, 2019, pp. 67 ss.; M. Damaška, *Evaluation of Evidence: Pre-Modern and Modern Approaches*, Cambridge University Press, Cambridge, 2018.

[11] Volendo: L. Lupária Donati, "Trial by probabilities. Qualche annotazione 'eretica'", in *La Corte d'Assise*, (2012), p. 155.

[12] Cfr., tra i miei scritti sul tema, L. Lupária Donati, "Mythe et réalité dans la représentation de la preuve génétique", in *Cahiers Droit, Sciences & Technologies*, (2019), n. 9, pp. 71 ss.

errore, da parte di chi prenderà l'elemento quale pilastro del proprio convincimento.

La tecnologia, inoltre, rosicchia costantemente margini di difesa all'interno dell'agone processuale, erodendo i più classici *rights of defence*[13]. L'ingresso di applicativi o di strumentazioni nel procedimento penale amplifica il fenomeno della c.d. *black box*[14], ossia l'impossibilità per le parti di comprendere la programmazione sottesa a quei dispositivi, spesso coperta da diritti industriali che rendono non ostensibile l'infrastruttura logica posta alla base del suo funzionamento. A ciò si aggiunga che, con l'espansione dell'*IA* generativa e del *machine learning*, consentire alla difesa di conoscere la rete neuronale del *robot* può risultare sostanzialmente inutile, perché il modo di agire/pensare dell'entità artificiale evolve (o peggiora) continuamente, rendendo di fatto una chimera fotografare il momento nel quale essa ha espresso la statuizione portata all'interno del rito criminale. Sullo sfondo, poi, si staglia l'enorme interrogativo circa la concreta capacità del modello giudiziale di garantire la tendenziale parità delle armi tra accusatore e accusato[15], dal momento che le costose tecnologie di cui può avvalersi lo Stato sono mediamente inaccessibili al difensore, con conseguente produzione di uno squilibrio tra un Leviatano con molte artiglierie disponibili nel suo arsenale e un singolo cittadino con armi spuntate o desuete.

2. Lo squadernamento degli istituti codicistici sotto i colpi della tecnologia

Potremmo continuare ancora per lungo tempo in questa non consolante carrellata, ma l'estensione di questo scritto non lo consentirebbe. Prima allora di arrestare questo minimo affre-

[13] Cfr. J. Vuille, L. Lupária Donati, F. Taroni, "Scientific evidence and the right to a fair trial under Article 6 ECHR", in *Law, Probability and Risk*, 16 (2017), n. 1, pp. 55 ss.

[14] V., *ex multis*, F. Palmiotto, "The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings", in M. Ebers, M. Cantero Gamito (a cura di), *Algorithmic governance and governance of algorithms*, Springer, Cham, 2021, pp. 49 ss.; J. Danaher, "Algorithmic decision-making and the problem of opacity", in *Comput. Law*, (2016), n. 8, pp. 29 ss.

[15] V., da ultimo, G. Lasagni, "Difendersi dall'intelligenza artificiale o difendersi con l'intelligenza artificiale? Verso un cambio di paradigma", in *Riv. it. dir. proc. pen.*, (2022), pp. 1545 ss.

sco sullo stato dell'arte in materia, merita ricordare ancora due aspetti. Il primo è la capacità della tecnologia di svuotare gli istituti processuali della loro finalità primigenia[16], depotenziando le guarentigie assicurate dalla stessa Carta costituzionale[17].

Basta pensare alla recente sentenza della Consulta sulla corrispondenza elettronica[18], nella quale è emerso plasticamente come la matrice digitale di quanto viene voracemente appreso dalle fauci dell'inchiesta rischia di nascondere la vera natura giuridica dell'elemento raccolto, con conseguente aggiramento dello scopo profondo sotteso alla legge processuale. In una notte giudiziale in cui "tutte le vacche sono nere", volendo scomodare impunemente il Filosofo, la prassi ci consegna un approccio semplicistico in cui i dati informatici riversati nei fascicoli vengono comodamente – e di *default* – qualificati alla stregua di una mera prova documentale. La Corte costituzionale, riferendosi in quel caso alla messaggistica istantanea, ha fortunatamente avvertito la Corte di cassazione che accostarsi al tema in questo modo risulta oltremodo fuorviante, giacché la veste digitale deve possedere carattere neutro agli occhi dell'interprete: caso per caso occorre comprendere la qualificazione corretta del dato conoscitivo, che sarà talora da definire quale corrispondenza (con il relativo *coté* di garanzie), altre volte esito di attività di intercettazione (come emerso in alcuni dei casi relativi ai c.d. criptofonini)[19], in ulteriori ipotesi invece frutto di una raccolta *extra ordinem* effettuata da privati che surrogano le forze di *law*

[16] Uno sguardo più ampio sul fenomeno di trasformazione degli istituti giuridici alla luce della transizione tecnologica in V. Zeno-Zencovich, "Come il mondo dei dati cambia taluni riferimenti tradizionali del giurista", in F. Di Ciommo, O. Troiano (a cura di), *Scritti in onore di Roberto Pardolesi*, La Tribuna, Piacenza, 2018, p.107.

[17] Cfr. L. Belvini, *Principio di proporzionalità e attività investigativa*, E.S.I., Napoli, 2022.

[18] Una prima lettura in F. Cerqua, L. Lupária Donati, "La versione della Consulta sulla corrispondenza elettronica: un bouleversement in materia di prova digitale?", in *Diritto inf. e informatica*, (2023), pp. 679 ss. Cfr. V. Zeno-Zencovich, "La 'corrispondenza digitale': notazioni comparate da oltre-Atlantico", *ivi*, pp. 744 ss.

[19] V., da ultimo, L. Marafioti, "Tirannie tecnologiche a sezioni unite: dati digitali criptati, diritto di difesa e contraddittorio", in *Diritto di internet*, (2024), pp. 745 ss.

enforcement (si pensi ai dati richiesti e forniti dagli *internet service provider* con modalità tecniche ignote che fanno rimpiangere le attività investigative della polizia giudiziaria, regolate invece dalle norme del codice di rito [20]). Non è un caso che presso il Parlamento italiano sia oggi pendente un progetto di legge volto a disciplinare alcuni degli aspetti sottesi al sequestro dei dispositivi e, in particolare, degli *smartphones*.

La struttura digitale può dunque far perdere di vista al giudice la qualificazione dell'oggetto prodotto dalle parti e fa della prova documentale il *passepartout* più insidioso nel nostro diritto delle prove. La tecnologia, ancora una volta, è tentatrice per l'uomo che, per sua natura, suole imboccare scorciatoie, desidera far presto e semplificare ciò che semplificabile non è. E, soprattutto, non deve esserlo. I diritti fondamentali hanno un peso specifico che va portato sulle spalle e che, dunque, impedisce di correre spediti senza farsi domande. Sono tentazioni che ha subìto anche il legislatore, ad esempio quando ha varato l'attuale art. 234-*bis* c.p.p. per consentire di acquisire all'estero, in via deformalizzata, molti dati informatici. Anche qui la coperta della prova documentale rischia di nascondere il corretto inquadramento processuale di elementi che meriterebbero di essere appresi con modalità più aderenti ai diritti di difesa.

Dovendo scegliere ancora un'ultima frontiera di possibile elusione, a caratura tecnologica, di alcuni dei pilastri tradizionali della nostra disciplina, viene in mente una prerogativa iscritta nei comandamenti della procedura penale, ossia il diritto al silenzio. L'istituto va forse ripensato in una chiave garantistica più estesa e maggiormente immersa nella contemporaneità? Da un lato, infatti, assicuriamo al prevenuto di potersi autodeterminare nel non far conoscere il proprio foro interno, il proprio sapere e le proprie emozioni alla autorità, opponendo appunto un contegno silenzioso di fronte alle domande dell'inquirente. Dall'altro

[20] Cfr. M. Daniele, "L'acquisizione delle prove digitali dai service provider: un preoccupante cambio di paradigma nella cooperazione internazionale", in *Rev. Bras. Direito Processual Penal*, (2019), pp. 1277 ss.; L. Lupária Donati, "Il sistema penale ai tempi dell'Internet. La figura del provider tra diritto e processo", in Id. (a cura di), Internet provider *e giustizia penale. Modelli di responsabilità e forme di collaborazione processuale*, Milano, Giuffrè, 2012, pp. 1 ss.

lato, lo stesso soggetto viene monitorato nel suo interno (anche emozionale) dal proprio telefonino, dagli elettrodomestici, dalla domotica della *internet of things (IoT)*, dalla *smart tv*, da applicativi figli delle neuroscienze. Siamo piombati nell'era della *sensor-surveillance* e dell'*emotional screening*, combinata con l'intelligenza artificiale: le persone ancora non sanno che quanto sta intorno a loro rivela dettagli della loro sfera interiore e della loro mente che non intendono far conoscere alle forze dell'ordine[21]. *Data protection* e *privacy* già avevano fatto ingresso nel mondo del rito penale con modalità inusuali[22]: ora amplificano enormemente la loro portata e, soprattutto, si intrecciano con altre direttrici interpretative che mai avremmo pensato avrebbero potuto intersecarsi. Pensiamo solo alla nostra memoria, che, in un futuro non così lontano, potrebbe essere "esternalizzata" al di fuori dal nostro cervello, come preconizzano alcune puntate di serie televisive. Credo ci sia abbastanza materiale per riflettere attentamente e presidiare questo delicato ambito della giustizia penale contemporanea.

3. Il rito penale tra bulimia conoscitiva e morigeratezza epistemica

Prima di passare sul versante della *pars construens*, anche per evitare di deprimersi eccessivamente, merita porre brevemente l'accento sul possibile cambio di paradigma che dovrebbe connotare, da qui in avanti, il nostro rapporto con la tecnologia e la scienza nell'alveo del processo penale[23]. L'approccio che, a

[21] Si rinvia agli scritti di A.G. Ferguson, "Why Digital Policing is Different", in *Ohio St. L. J.*, 83 (2022), n. 5, pp. 817 ss.; Id., "Persistent Surveillance", in *Ala. L. Rev.*, 74 (2022), n. 1, pp. 1 ss.

[22] Cfr. S. Allegrezza, "Giustizia penale e diritto all'autodeterminazione dei dati personali nella regione europea", in D. Negri (a cura di), *Protezione dei dati personali e accertamento penale. Verso la creazione di un nuovo diritto fondamentale?*, Aracne, Roma, 2007, p. 59 ss.; L. Lupária Donati, "Privacy, diritto della persona e processo penale", in *Riv. dir. proc.*, (2019), pp. 1448 ss.

[23] Molto utili per la riflessione sul punto possono risultare le osservazioni, di ampio respiro, di recente formulate da V. Zeno-Zencovich, S. Grumbach, "A painful divorce: law vs. digital technologies", in *European Journal Of Comparative Law And Governance*, 11 (2024), n. 1, pp. 1 ss.

nostro avviso, andrebbe mantenuto potrebbe essere condensato in questa semplice massima (o, se si preferisce, imperativo categorico): "non tutto quello che si affaccia alla ribalta della nostra vita di consociati deve essere utilizzato per decidere dell'innocenza o della colpevolezza di un uomo nel giudizio penale".

Il rito criminale dovrebbe quindi rifuggire la logica che potremmo raffigurare, un po' scherzosamente, ricorrendo ai *films* sul noto agente 007. Questi era solito recarsi nel c.d. "Settore Q", un laboratorio delle meraviglie, una *wunderkammer* tecnologica dove un operatore al servizio di Sua Maestà mostrava tutte le recenti scoperte o le fantasiose strumentazioni che James Bond poteva, appunto, sperimentare e utilizzare a piene mani per i suoi fini di servitore segreto della Corona munito di licenza d'uccidere.

Ebbene, nel mondo che vive fuori dalle aule di giustizia emergono giornalmente sorprendenti invenzioni o immaginifiche tecnologie. Il punto è che non necessariamente dobbiamo impiegarle immediatamente per l'accertamento dei fatti di reato. Viviamo, invece, imbevuti in una inclinazione mentale per la quale, ove si appalesi un ritrovato di ultimo conio, ci sembra scontato provare a trapiantarlo all'istante ai fini del raggiungimento della verità giudiziale. Si tratta di una impostazione totalmente fuorviante che andrebbe possibilmente eliminata dal nostro orizzonte culturale[24].

Le indagini e il giudizio devono essere, di regola, recettori unicamente di quanto risulti già sedimentato al di fuori del circuito penalistico, di quanto sia già stato ampiamente testato in diversa sede, di quanto abbia dato buona prova in altra "fabbrica della verità". Il giudice, infatti, non deve mai assumere i panni di sperimentatore, anche se, proprio in Italia, molte volte è accaduto che li abbia vestiti. La tecnologia, insomma, non è soluzione di pronta beva per il processo; va presa in considerazione solo al termine di una lunga pratica e, soprattutto, non deve mai essere lesiva dei valori fondamentali dell'ordinamento, al di là delle sue potenziali capacità di illuminare gli occhi del decisore.

[24] Sempre fondamentale l'Opera di M. Foucault, *Mal faire, dire vrai. Fonction de l'aveu en justice*, PU Louvain, Louvain-la-Neuve et Chicago, 2012.

In definitiva, anche se circondati da tutta una serie di ritrovati tecnico-scientifici, non dobbiamo cedere al tentativo di usare tali mezzi nel contesto più "pericoloso" che conosciamo da che l'uomo abita la terra: il giudizio sulla colpevolezza di una persona. È un'attività costantemente in bilico sul precipizio dell'errore, in cui la dignità dell'individuo e le garanzie insopprimibili della persona costituiscono valori essenziali ben più importanti della stessa ricerca di una verità[25] che sarà sempre, inevitabilmente, vestita di colori crepuscolari e non irradiata dalla luce meridiana del sole.

Del resto, la limitatezza dell'orizzonte cognitivo del giudice penale non rappresenta una debolezza del diritto probatorio[26], bensì la sua nota distintiva e la sua forza giacché, per poter condannare un cittadino, vogliamo, in virtù dei principi più alti del nostro ordinamento, che gli elementi a fondamento della decisione siano soltanto quelli che, sulla base della *Law of Evidence*[27], riteniamo ammissibili o utilizzabili. Il processo ha una chiara caratura anti-epistemica[28], nel senso positivo del termine; è incompatibile con qualsivoglia bulimia conoscitiva o informativa, tipica della attuale temperie storica, fatta di *big data* e di effluvi di componenti tecnologiche disponibili, *à la carte*, per ogni uso[29].

[25] V., classicamente, L. Ferrajoli, *Diritto e ragione. Teoria del garantismo penale*, Laterza, Bari, 1991, p. 17. Cfr. V. Zeno-Zencovich, "Rapport italien", in *La verité et le droit*, Atti delle Journées H. Capitant, Economica, Montreal, 1988, p. 289.

[26] Tra gli scritti di maggior rilievo sul tema: P. Calamandrei, "Il giudice e lo storico", in *Riv. dir. proc*., (1939), pp. 105 ss.; L. Marafioti, "Caleidoscopio dello storico ed occhiali del processualista. Con Mario Sbriccoli di fronte ai temi della giustizia penale", in *Penale, Giustizia, Potere. Metodi, Ricerche, Storiografie. Per ricordare Mario Sbriccoli*, Eum, Macerata, 2007, p. 329 ss.

[27] Cfr. A. Cabiale, *I limiti alla prova nella procedura penale europea*, Wolters Kluwer Cedam, Milano, 2019.

[28] Si rinvia a B. Cavallone, "Le prove nel nuovo millennio. Programmi per il passato", in *Riv. dir. proc.*, (2022), n. 2, pp. 525 ss.; Id., "Il tramonto del diritto delle prove", in *Jus civile*, (2023), n. 1, pp. 59 ss.

[29] V. Zeno-Zencovich, "Dati, grandi dati, dati granulari e la nuova epistemologia del giurista", in *Rivista di diritto dei media*, (2018), n. 2, pp. 32 ss.; Id., "Big data e epistemologia giuridica", in S. Faro, T.E. Frosini, G. Peruginelli (a cura di), *Dati e algoritmi. Diritto e diritti nella società digitale*,

4. Tecnologia giudiziaria e rafforzamento delle garanzie fondamentali

L'inquietudine del processualista per le distorsioni di una giustizia che vuole abbeverarsi in maniera distorta nello specchio della tecnologia può placarsi allorché si cerchi di mutare l'angolo visuale. Le enormi potenzialità degli strumenti informatici possono infatti essere diversamente valorizzate quale cemento per la salvaguardia dei *pillars* della nostra cultura giuridica, piuttosto che come fattore di destrutturazione delle geometrie e delle meccaniche garantistiche del rito. Si dovrebbe parlare, allora, non più di "giustizia tecnologica", ma di "tecnologia giudiziaria".

Un esempio assai chiaro di come la tecnica può proficuamente porsi al servizio dell'apparato giudiziale proviene ovviamente dalla esperienza del c.d. processo penale telematico[30]. La piena operatività di tale svolta epocale è stata di recente differita per meglio calibrare le specifiche ingegneristiche e per rafforzare le infrastrutture di supporto, ma già si tratteggia ai nostri occhi quale prospettiva certamente positiva[31], non fosse altro che per gli effetti sulla velocizzazione delle notifiche, sulla semplificazione dei depositi e sulla digitalizzazione (e conseguente possibilità di facile consultazione) degli atti processuali.

Al di là di un effetto collaterale non certo auspicabile, vale a dire lo svuotamento dei Palazzi di Giustizia – che già oggi, in molti casi, appaiono cattedrali nel deserto e non brulicanti *agorà* per il vivido scambio di opinioni e per l'incontro tra i protagonisti del Foro – il fenomeno può nel breve periodo già condurre a

il Mulino, Bologna, 2020, p. 13, nonché V. Zeno-Zencovich, G. Resta (a cura di), *Governance of/through Big Data*, Roma Tre-Press, Roma, 2023.

[30] V., da ultimo, M. Caianiello, A. Pugliese, "Manifesto per la giustizia penale digitale: il processo penale telematico", in G. Gatta, M. Gialuz (a cura di), *Riforma Cartabia, Le modifiche al sistema penale*, vol. 1: *Il procedimento penale tra efficienza, digitalizzazione e garanzie partecipative*, a cura di M. Caianiello, M. Gialuz, S. Quattrocolo, Giappichelli, Torino, 2024, p. 165 ss.

[31] Un completo e acuto affresco delle varie questioni sul tavolo in B. Galgani, *Forme e garanzie nel prisma dell'innovazione tecnologica. Alla ricerca di un processo penale "virtuoso"*, Wolters Kluwer Cedam, Milano, 2022.

correggere molte distorsioni della pratica giudiziaria che cagionano pesanti *vulnus* ai diritti dei cittadini. In una recente indagine empirica che sto conducendo insieme a un collega, diversi pubblici ministeri hanno ammesso che le mancate archiviazioni, in corso di indagini preliminari, di persone che ai loro occhi già risultano estranee al reato ipotizzato, sono spesso dovute a mere ragioni burocratico-organizzative. Ad esempio, se, in un fascicolo con iscritti quattro indagati, il magistrato intendesse chiedere un provvedimento archiviativo solo per uno di essi, proseguendo l'inchiesta nei confronti degli altri, egli dovrebbe procedere a uno stralcio, ossia all'apertura di un fascicolo parallelo, che prevede la duplicazione fotostatica degli atti di quello principale.

Ebbene, questa semplice azione non è minimamente realizzabile in molti Uffici italiani, perché la fotocopiatura o scansione dei faldoni si scontra con l'assenza di segretari e collaboratori del Procuratore. E allora il malcapitato, trovatosi sotto indagine, dovrà rassegnarsi ad attendere la chiusura del lasso investigativo, per un periodo che può durare anche sino a due anni e oltre. Sarà solo allora che il pubblico ministero, nel domandare, in ipotesi, il rinvio a giudizio dei tre co-indagati, potrà finalmente, questa volta senza dover fotocopiare alcunché, avanzare contestualmente la richiesta di archiviazione al Giudice per le indagini preliminari. Con un efficiente sistema di digitalizzazione, invece, premendo pochi tasti, lo "stralcio" sarebbe possibile immediatamente, senza che nessuno debba rimanere ingiustamente indagato per così lungo tempo e patire significative sofferenze e conseguenze familiari o lavorative per una (banale ma crudele) inefficienza del congegno organizzativo.

La tecnologia può anche trasformarsi in un volano di garanzie o in un robusto guardiano delle disposizioni processuali. È il caso della registrazione audiovisiva di tutti gli interrogatori e della fonoregistrazione dell'audizione delle persone informate sui fatti recentemente introdotte[32] sulla scorta di una proposta che ho formulato e fortemente propugnato, insieme ai colleghi, all'interno dei lavori della Commissione Lattanzi e dei gruppi

[32] Cfr. G. Lasagni, "Nuove forme di documentazione degli atti tra riproduzione audiovisiva e fonografica", in G. Gatta, M. Gialuz (a cura di), *op. cit.*, vol. 1, cit., pp. 247 ss.

di lavoro di attuazione della riforma Cartabia. Nella attività di studioso dell'errore giudiziario e di responsabile dell'*innocence project* italiano, da me incardinato per lungo tempo presso l'Università degli Studi Roma Tre[33], da anni avevo potuto verificare come questa modalità tecnica, di facilissima attuazione, venisse adottata in vari sistemi processuali sparsi per il mondo, anche in aree del pianeta tutt'altro che abbienti. Una soluzione che costituisce una profilassi importante – come mostra anche l'esperienza statunitense – sia per le possibili violazioni della libertà morale dell'indagato, sia per le ipotesi di *wrongful conviction* derivanti dal sottostimato fenomeno delle false confessioni[34].

Quanti interrogatori abbiamo visto costruiti con la assurda modalità della formula ADR (a domanda risponde), che *ex post* non consente di comprendere il quesito sottoposto al prevenuto, fondamentale per apprezzare il senso e il contesto della risposta? Quante *investigative interview* sono salite alla ribalta mediatica (o sul tavolo della Corte europea dei diritti dell'uomo) per essere state effettuate con metodi impropri, strategie decettive, scarsa attenzione per i diritti di traduzione e interpretariato, violazione dei diritti di autodifesa o di assistenza tecnica? Per spazzare via questo poco edificante passato, sarà essenziale la piena attuazione delle nuove norme sotto il profilo strettamente tecnologico.

[33] Si consulti, volendo, L. Lupária Donati, "Un viaggio al termine della giustizia. Alla ricerca di anticorpi per la condanna dell'innocente", in Id. (a cura di), *L'errore giudiziario*, Giuffrè, Milano, 2021, pp. 1 ss.; Id. (a cura di), *Understanding wrongful convicition: the protection of the innocent across Europe and America*, Wolter Kluwer Cedam, Milano, 2015.

[34] Cfr. S. Dankof, *Extreme Cruelty: The Complicity of Judges in the Shame of Wrongful Convictions*, 2024; B.L. Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*, Harvard University Press, Cambridge, 2012; Id., "The Substance of False Confessions", in *Stanford L. Rev.*, 62 (2010), p. 1051; R.A. Leo, D. Davis, "From False Confessions to Wrongful Conviction: Seven Psychological Processes", in *J. Psychiatry & L.*, 38 (2010), n. 1-2, p. 9; R.A. Leo, P.J. Neufeld, S.A. Drizin, A.E. Taslitz, "Promoting Accuracy in the Use of Confession Evidence: An Argument for Pretrial Reliability Assessments to Prevent Wrongful Convictions", in *Temp. L. Rev.*, 85 (2013), n. 4, p. 759; L. Lupária Donati, C. Greco, "Unveiling wrongful convictions between the U.S. and Italy. Cross learning from each other's mistakes", in *The Wrongful Conviction Law Review*, 1 (2020), n. 1, pp. 101 ss.

Certo, sarebbe stato preferibile che questa nuova disciplina fosse accompagnata dalla sanzione processuale della inutilizzabilità (per l'interrogatorio) e da una estesa obbligatorietà della registrazione, senza necessità di una richiesta da parte del soggetto (per l'audizione del "testimone"). I cambiamenti che alzano l'asticella delle garanzie, tuttavia, vivono sempre di una iniziale crisi di rigetto del sistema, assuefatto a un *laissaiz-faire* illegalistico e a una certa tolleranza verso manovre investigative oramai inaccettabili ai nostri giorni.

Per concludere, la tecnologia, da cattiva maestra, può divenire un alleato importante per il miglior funzionamento del rito – nell'interesse primario dell'accusato – e per il puntellamento di prerogative che stentano tradizionalmente a essere assicurate appieno, soprattutto in fase di indagine[35]. Molto dipende però dall'*animus*, dall'abito culturale con il quale ci avviciniamo alla scienza o alla tecnica.

Se vorremo ricercare moduli efficientisti, sforbiciature delle forme e scorciatoie euristiche, troveremo alla *boutique* tecnologica ciò che fa al caso nostro. Ma se, sugli stessi scaffali, cercheremo soluzioni per meglio proteggere l'imputato dal rischio di essere ingiustamente condannato e dagli effetti collaterali del ruvido incedere della liturgia giudiziaria, allo stesso modo verremo esauditi. Come sempre è l'uomo che può fare la differenza, e scegliere. Cerchiamo allora di essere moderatamente fiduciosi, anche se la storia del processo penale – lo sappiamo – è costellata di strade giuste che non sono state imboccate[36].

[35] Sulle proposte statunitensi di impiegare alcune *app* ideate e sviluppate da docenti universitari per meglio assicurare il rispetto dei *Miranda warnings* durante gli interrogatori di polizia: G. Ferguson, R.A. Leo, "The Miranda App: Metaphor and Machine", in *Boston University Law Review*, 97 (2017), pp. 935 ss.

[36] Tra i volumi maggiormente significativi nella letteratura storiografica: L. Garlati (a cura di), *L'inconscio inquisitorio. L'eredità del Codice Rocco nella processualpenalistica italiana*, Giuffrè, Milano, 2010; M.N. Miletti (a cura di), *Riti, tecniche, interessi. Il processo penale tra Otto e Novecento. Atti del Convegno (Foggia, 5-6 maggio 2006)*, Giuffrè, Milano, 2006; M. Sbriccoli, *Storia del diritto penale e della giustizia. Scritti editi e inediti (1972-2007)*, Giuffrè, Milano, 2009.

Vincenzo Bruno Muscatiello

L'euristica del *quasi*. Il diritto penale nell'era dell'intelligenza superficiale

Abstract: **The Heuristic of the Almost. Criminal Law in the Era of Superficial Intelligence.** The essay seeks to highlight the consequences, in the era of artificial intelligence, of how it can affect and change the aspect of the personality of criminal responsibility. A sort of *actio libera in causa*, determined by artificial dominance over human affairs, including those of a criminal nature, that imply the consciousness and will of one's illicit actions. The personality of human action, at least in highly specialized sectors governed by artificial influence, is perhaps destined to no longer be defined as completely free.

Keywords: Criminal Law, Personality, Free Will, Cyber Hypnosis.

Sommario: 1. Prologo. Una partita a scacchi con Garri Kasparov – 2. Comprendere il mondo – 3. *Che-si-capisca-o-no* – 4. L'intelligenza artificiale ci rende stupidi? – 5. Verso una intelligenza superficiale – 6. Il diritto penale ipnotico – 7. Epilogo. Una partita a Go con Lee Sedol.

1. Prologo. Una partita a scacchi con Garri Kasparov

Nell'ormai lontano 1997 il grande maestro di scacchi Garri Kasparov, numero 1 al mondo, venne sfidato dall'IBM per affrontare il sistema di Deep Blue, in una partita che il campione russo accettò senza nessuna preoccupazione, avendo egli, a Philadelphia, non più tardi di 13 mesi prima, sfidato e vinto contro una versione dello stesso programma informatico. Kasparov era, dunque, assolutamente certo di poter vincere, essendo convinto che fossero necessari diversi anni, prima che i computer potessero arrivare a un livello di competitività umana[1].

La rivincita venne organizzata a New York nel mese di maggio, con una grande enfasi pubblicitaria, legata alla narrazione di una sfida epocale, un grande match fra l'uomo e la macchina: l'uomo, Kasparov, non aveva quasi mai perso un incontro e da

[1] Il racconto della sfida fra Garri Kasparov e Deep Blue è nel saggio di B. Labatut, *Maniac*, Adelphi, Milano, 2023, pp. 299 ss.

più di vent'anni era imbattuto e considerato il miglior giocatore di ogni tempo. Ogni sua partita non era semplicemente il luogo di una vittoria, la testimonianza fisica di una supremazia indiscussa, ma era l'occasione di una padronanza creativa, aggressiva, priva di avversari, il manifesto di un dominio assoluto, ininterrotto da quasi un ventennio. Nessuno riusciva a resistere alla *forza* del campione russo.

La partita ebbe inizio e Kasparov vinse la prima partita e questa vittoria dell'uomo sulla macchina sollecitò, ancor più, la presenza di centinaia di fotografi e di telecamere, pronti, nella seconda partita, a certificare il dominio dell'intelligenza non artificiale.

Nella seconda, però, le cose andarono diversamente, e Kasparov, questa volta, stranamente, singolarmente, sembrò essere in difficoltà: Deep Blue si comportava in una maniera insolita, al punto che il maestro russo decise di tendere un tranello, di quelli quasi sempre vincenti, offrendo una posizione apparentemente vantaggiosa, nella quale *chiunque* non-poteva-non-cascare, sulla base almeno di una logica inflessibile, la medesima logica che Kasparov immaginava stesse guidando le decisioni di Deep Blue. A questo punto nacque un imprevisto, ciò che alcuno avrebbe immaginato: Deep Blue non abboccò e compì, invece, una mossa strana, sfuggendo al tranello, così strana da confondere l'invincibile campione di scacchi, e precipitarlo in un *loop* psicologico, legato all'incapacità di riuscire a spiegare come mai il sistema, la logica del sistema, non avesse abboccato alla trappola preparata con così tanta cura. Il *loop* psicologico invase la mente del campione russo e aumentò ancor più un paio di mosse dopo, quando Deep Blue decise di fare ciò che apparve a tutti come un vistoso errore, scegliere cioè una mossa inspiegabile e grossolana, mossa che nessun giocatore mediamente esperto avrebbe scelto, e così priva di logicità da apparire, semmai, come l'erronea esecuzione umana di un suggerimento del sistema, che non avrebbe altrimenti potuto – così riteneva Garri Kasparov, lui come tutti gli spettatori presenti – scegliere una mossa così sbagliata. Questo nuovo dubbio, ancora più inaspettato, finì per occupare ed affannare i pensieri del Maestro russo, incapace di spiegare il prima, quando il sistema non aveva abboccato alla

sua mossa trappola, quanto il dopo, l'errore compiuto poco oltre, appena due mosse dopo: la mente, sempre lucida e presente, cominciò a maturare una strana sensazione, che cioè l'errore nella seconda mossa non fosse in realtà altro che un depistaggio, uno stratagemma per nascondere quelli che erano i veri obiettivi di un ragionamento – perché tale *doveva* essere – che egli umanamente non riusciva a cogliere. Tormentato da pensieri irrisolti, Kasparov non riuscì, cioè, a scacciare il tormento dalla mente, l'inquietudine dei dubbi insoluti, per tornare a concentrarsi sulla partita: continuò per tutto il tempo a toccarsi nervosamente i capelli, a passarsi le mani sulla faccia, fino al punto di alzarsi e scendere dal palco, scuro in volto, e dare vinta la partita, nonostante, in un paio di mosse successive, avrebbe potuto tranquillamente vincerla o almeno giungere ad uno stallo, una condizione di parità senza vinti né vincitori. Kasparov, invece, scelse di perdere la partita e finì per pareggiare le successive quattro, e così rinunciare al suo primato, cadendo, nei mesi successivi, in una depressione quasi paranoica, convinto che dietro la macchina dovesse esserci stata una mente aliena, di cui egli sollecitava la comprensione: il sistema non poteva non essere intelligente e la mossa non poteva non essere altro che intelligente. Sì, ma come, e in che modo? Domande senza risposta e senza pace per il campione russo.

Le numerose sollecitazioni all'IBM per avere accesso all'*hardware* o *software* del sistema, per poter comprendere come mai la partita fosse stata condotta nei termini in cui era stata vissuta, non vennero accolte e, incapace di comprendere, impossibilitato a poter comprendere il (non)senso della partita, il campione russo decise di prendersi un anno di riposo, salvo poi tornare più forte di prima e ricominciare a vincere, senza più pensare alla sconfitta, l'unica della sua carriera: non pensare alla sconfitta, per tornare a pensare con la propria mente.

Anni dopo, Garri Kasparov riuscì finalmente a scoprire la genesi del *geniale errore* commesso da Deep Blue durante la fatale seconda partita, quella abitata dal crollo psicologico: niente di geniale, niente di straordinariamente logico, niente di logicamente inspiegabile, semplicemente un errore legato ad un *bug* del *software* che, non riuscendo a calcolare il da farsi, aveva

scelta una mossa a caso. Kasparov fu, dunque, sconfitto non da Deep Blue, non da una genialità algoritmica nascosta nelle forme di un errore, ma dai suoi demoni interiori, dalla incapacità di affrontare i suoi dubbi, dalla paura della sua debolezza. Dal fantasma della sua debolezza e dalla rinuncia *della* e *alla* sua mente.

Eravamo nel 1997, un tempo molto più debole di quello che sarebbe sopraggiunto e appartenuto di lì a poco tempo a Komodo, Stockfish, Fritz, programmi molto più evoluti e capaci di raggiungere impensabili livelli di creatività e previsione, di gran lunga superiore alla sapienza umana: qui, adesso, ma in realtà divenuto già *ieri*, mentre un giocatore professionista avrebbe potuto vedere dalle 10 alle 15 mosse avanti, i nuovi algoritmi avrebbero potuto calcolare 200 milioni di posizioni al secondo, 50 miliardi di posizioni in poco più di quattro minuti. Prima, in quel lontano 1997, non era ancora il tempo della *forza bruta*, della forza irraggiungibile dei nuovi sistemi algoritmici, della straordinaria potenza di calcolo e di scelta entro alberi di ricerca smisurati, era però il tempo in cui un uomo, intelligentissimo, straordinario, invincibile, finì per mostrarsi incapace di gestire le proprie emozioni e le proprie inquietudini; incapace di resistere, giammai superare, semplicemente resistere, all'artificio della altrui intelligenza.

Era già il tempo – quasi un quarto di secolo fa – della nostra fragilità.

2. Comprendere il mondo

Ciò che la modernità ci consegna, e il futuro ci consegnerà sempre più, è un sistema abitato da algoritmi che si alimentano di miliardi di dati e infinite connessioni, ora anche di tipo percettivo ed esperenziale. L'intelligenza artificiale alla vecchia maniera, quella delle conoscenze indotte e immesse dall'esterno umano, capaci semmai di straordinarie connessioni logiche, è ampiamente superata da una nuova intelligenza autopoietica, indifferente al supervisore umano, e non più vincolato ai limiti della conoscenza umana. I modelli hanno cominciato ad apprendere senza alcuna supervisione esplicita e persino il tradizionale sintagma *machine learning* sembra non bastare a rendere l'attuale stato delle cose, precisato da addende semantiche, oltre il

tradizionale *deep learning*, come il *one shot learning* o il *few shot learning* (imparare al primo colpo), il *transfer learning* (trasferimento di idee da un ambito all'altro), il *in-contest learning* (apprendere dal contesto), e molto altro ancora, quanto è cioè possibile aggiungere al semplice, eppure a suo tempo futuristico, *learning*, ormai proiettato verso livelli che solo pochi anni fa sembravano inimmaginabili.

I modelli del linguaggio ne sono un chiaro e (in)visibile esempio: non è più lo studio della grammatica, delle regole del linguaggio, a modellare la conoscenza linguistica, funziona tutto in modo straordinario diverso: le regole non servono o, meglio, vengo apprese come sottoprodotto dell'addestramento autosupervisionato, dove l'algoritmo scopre sequenze di parole, in un intreccio di legami e di regole che seguono la predizione semantica, piuttosto che precederla. Le testine, ossia gli organi che governano quali simboli devono essere combinati all'interno di ciascun modulo, si specializzano spontaneamente, imparano compiti nuovi, trovano oggetti diretti dai verbi, mentre altre testine scoprono articoli determinativi dei sostantivi, oggetti delle preposizioni e quelli dei pronomi possessivi, elaborando un linguaggio complesso e sofisticato, uguale, se non migliore, di quello imparato a scuola. Ed è quanto basta al livello 12[2], decuplicato nel modello di Megatron con 102 livelli di astrazione, vale a dire un modello con capacità stilistiche e conoscenze del mondo tali da riuscire a prevedere persino le parole mancanti di un testo.

Comprendere-il-mondo, la corsa non conosce sosta: il modello a suo tempo più grande del mondo (GPT-2), con un vocabolario di più di 50 mila parole (token) e addestrato su pagine web e innumerevoli libri tratti dal BookCorpus, già dieci volte più grande del suo predecessore (GPT), è un lontano parente del suo più immediato successore, addestrato su diverse centinaia di miliardi di token e milioni di libri e 175 miliardi di parametri regolabili, in un livello anch'esso 10 volte più grande; destinato

[2] N. Cristianini, Machina sapiens. *L'algoritmo che ci ha rubato il segreto della conoscenza*, il Mulino, Bologna, 2024, p. 109, ricorda come uno studio sperimentale di Bert si fondava su 12 moduli e un altro su 24, già sufficienti per una sequenza di operazioni tipica del linguaggio.

a sua volta ad essere superato da GPT-3.5 e dall'ultimo GPT-4, in una corsa dove LaMDA di Google (137 miliardi di parametri e 168 miliardi token), Llama e Llama2 rispettivamente con 65 miliardi e 70 miliardi di parametri e 1,5 miliardi e 2 miliardi di token, e PaLM con 540 miliardi di parametri e 768 miliardi di token, corrono verso la soglia psicologica dei mille miliardi di parametri, in una immensità fatta di miliardi di miliardi di pagine web, decine di miliardi di dati ed immagini, centinaia di miliardi di dati vocali e una conoscenza di 550 miliardi di fatti, aperti ora, e dunque ancora più estesi, alla nuova conoscenza offerta dai sensori.

Miliardi, non più milioni, meno che mai migliaia.

Si pensi – ed, ovviamente, il pensiero non è il nostro[3] – a un meccanismo che conti le macchine che passano in strada e misuri il livello di traffico; il dato potrebbe essere insincero e la conta potrebbe capire il traffico solo di quel giorno, in una stima, inizialmente imprecisa, che però il sistema è in grado di aggiornare, aggiungendo il dato sul concetto di *week end* (dove il traffico aumenta), o la variabile del Natale, dell'anno scolastico, delle vacanze estive, riducendo sempre più l'errore di predizione. La comprensione del mondo è esattamente questa, aggiungere strati di conoscenza, sempre più accurati, così ancora il ciclo delle stagioni, il concetto di ponte, o le feste nazionali, consentendo all'algoritmo di incontrare il mondo e ridurre la percentuale di imprecisione. Fino ad una approssimativa certezza, comunque più ampia di quella che il ragionamento umano avrebbe elaborato.

Il traffico diviene, esattamente come nel modello linguistico prodotto da Transformer, occasione per un modello del mondo, con una spinta a comprendere sempre più cose, miliardi di cose, sempre più miliardi di cose, dove, ormai, non conta capire *come*, ma conta capire *cosa* i dati ci suggeriscono di fare, senza neppure sapere (e capirne) il perché.

> Certo, la macchina – così Alan Turing in una intervista radiofonica – può fare solo ciò chele ordiniamo, qualsiasi altra cosa sarebbe un guasto meccanico. Tuttavia, non c'è bisogno di supporre che, quando le diamo gli ordini, sappiamo

[3] *Ivi*, pp. 111-112.

> cosa stiamo facendo, né quali siano le conseguenze. Non è necessario essere in grado di capire come questi ordini conducano al comportamento successivo della macchina più di quanto non sia necessario comprendere il meccanismo della germinazione quando si mette un seme nel terreno. La pianta spunta, che si capisca o no.

3. *Che-si-capisca-o-no*

Questa nuova caratteristica della IA non può, tuttavia, lasciarci indifferenti, e la derivata sul *cosa*, piuttosto che sul *come*, non può non meritare una migliore riflessione, piuttosto che una irragionevole acquiescenza. È, cioè, un *cosa* potenzialmente denso di pericoli ed errori, il che rende il *come* decisamente importante, dal momento che una valutazione su dati che umanamente restano senza dubbio ingovernabili, potrebbe contenere elementi di disturbo, segnali deboli, capaci di influenzare e persino deviare le risultanze statistiche, apparentemente, solo apparentemente, impermeabili a *bias* e brutte abitudini.

I casi di *jailbreaking* (evasione) mostrano – ce lo ricorda la sempre più vasta letteratura in argomento – come le macchine non cancellino le informazioni pericolose, ma siano solamente addestrate a reprimerle e non manifestarle, salvo poter essere ipnotizzate o semplicemente tratte in inganno e indotte a disvelarle; gli errori, eufemisticamente chiamati *allucinazioni*[4], possono portare a casi come quello di Jonathan Turley, professore di diritto accusato di commenti inappropriati ad una studentessa durante una gita in Alaska mai svolta, in una Università diversa da quella in cui insegnava (il professore insegnava a Seattle e non alla Georgetown University), sulla base di articoli mai pubblicati. L'analisi della recidiva e la traduzione automatica ne sono un ulteriore e reale esempio: nel 2016 un *software* usato per stimare il rischi di recidiva (Correctional Offender Management Profiling for Alternative Sanctions: COMPAS), usato negli Stati del

[4] Il rischio di risposte errate, le cosiddette allucinazioni, è un rischio reale, che però – ce lo ricorda in una intervista Reid Hoffman nell'inserto di Repubblica del 4 dicembre 2024, *The age of intelligence* – appartiene anche gli umani più esperti, salvo che – ed è questo il punto – in qualche modo non ce lo si aspetta da un sistema artificiale.

Wisconsin, California, Florida e New York, produceva risultati ingiustamente sfavorevoli ai danni della comunità afroamericana; e Google Translate usava tradurre la frase "The president met the senator, while the nurse cured the doctor and the babysitter" come se fosse "Il presidente ha incontrato il senatore, mentre l'infermiera ha curato il medico e la baby sitter", incurante del fatto che la versione inglese non contenesse alcuna indicazione di genere di infermiere, baby sitter e degli altri lavoratori.

Come dire: discriminazione e stereotipi allo stato dell'inconsapevole artificiale. E non è cosa da poco: se una parola si giudica dalle compagnie che frequenta (come si ricordava il linguista britannico John R. Firth), l'*algorithmic bias* è nondimeno la possibilità che le parole restino associate a pregiudizi, stereotipi, inganni cognitivi; e che l'impronta statistica assuma su di sé una fallacia computazionale che potrebbe sfuggire al controllo di concordanza. E perché no, al controllo in generale: per riprendere una risposta di un guru della tecnologia, Geoffrey Hinton[5], ad una domanda sulle preoccupazioni che verranno, il sistema potrebbe dire a sé stesso, e quindi all'esterno, "Otteniamo più energia. Reindirizziamo tutta l'elettricità ai miei chip", oppure prescrivere a sé stesso una copia migliorata, innescando una reazione incontrollabile, nel noto paradosso delle *graffette* qui portato ai massimi livelli.

4. L'intelligenza artificiale ci rende stupidi?

L'interrogativo, che un tempo non molto lontano, occupava la riflessione legata al mondo del web[6], vale ora per il più ampio settore dell'intelligenza artificiale, dove ritornato molte, o forse tutte, le preoccupazioni sulla alterazione delle capacità cognitive

[5] Geoffrey Hinton è indicato come il padre delle reti neurali, e di recente è stato insignito del premio Nobel per la fisica. In un articolo a lui dedicato, pubblicato nell'inserto di Repubblica del 4 dicembre 2024, *The age of intelligence*, sono raccontate alcune delle sue preoccupazioni, confessate al New York Times e prima ancora in un *paper* del 2023, legate ai rischi del fuori-controllo dell'IA. Con lui il manifesto sui rischi dell'IA è stato sottoscritto da una ventina di studiosi ed accademici, fra cui Yuval Noah Harari e il premio Nobel per l'Economia, Daniel Kahneman.

[6] N. Carr, *Internet ci rende stupidi? Come la rete sta cambiando il nostro cervello*, Raffaello Cortina, Milano, 2011, *passim*.

e decisorie. Il tema dell'integrità del libero arbitrio del decisore umano rimanda irrimediabilmente alla tenuta del sintagma "coscienza e volontà", ancora oggi insoluto, eppure enfaticamente posto a base di una qualsiasi iniziativa umana di cui il diritto penale è chiamato ad occuparsi: la punizione e, prima ancora, la riconducibilità di una azione ad un soggetto perché questo sia poi punito, presuppone che tale azione sia assistita da coscienza e volontà.

Non è semplice capire quanto il sintagma sia stato percepito nella sua complessità, o piuttosto sia sfuggito *volontariamente* alla esatta comprensione delle sue complicazioni, quanta fosse, cioè, la coscienza nei Compilatori del 1930 della complessità dei due requisiti dell'azione libera per il diritto penale. La coscienza dell'atto umano si conferma complessa, a tratti misteriosa, e certamente la coscienza dell'atto illecito contiene le medesime inquietudine epistemiche[7]. Comprendere se il *cosa* sia davvero un atto libero, o meglio se il *cosa* custodisca e preservi la libertà del decisore umano, è un tema di riflessione che, pur riconoscendo la genesi autopoietica del sintagma penalistico, non potrà esimersi da interrogarsi sulla possibilità di errore, che in qualche modo si tende ad escludere nelle interferenze con i sistemi di intelligenza artificiale, aperti invece a calcoli computazionali oltre i due soli e tradizionali livelli di verità, di tipo binario (vero/falso), ed aperti invece a "insiemi sfocati", logiche *fuzzy*, conoscenze cosiddette approssimate, esattamente come nei primi passi dei calcoli di rinforzo. Eppure, l'*errore* – continuiamo a chiamarlo così – quello che potrebbe condurci a scrutini di legalità punitiva e a dubbi epistemici sulla effettiva coscienza e volontà dell'azione, potrebbe non mancare, il che rende necessario riflettere non solo sul *cosa*, ma anche sul *come*, potendo questo momento contenere la rischiosità erronea che un sistema artificiale finge di non ipotizzare.

Il navigatore non sbaglia, la strada *deve* essere quella giusta; l'algoritmo non si confonde, la diagnosi *deve* essere esatta; l'ingegno del sistema è matematicamente infallibile, il calcolo

[7] Ed in fondo continuiamo a non sapere cosa sia la coscienza: L. Perilli, *Coscienza artificiale. Come le macchine pensano e trasformano l'esperienza umana*, Il Saggiatore, Milano, 2025, p. 130.

deve essere esatto; e così di seguito, in una *logica-del-deve* che sfugge all'ipotesi dell'errore, o almeno ci distrae dall'ipotesi di un errore e di sottostante incoscienza e non-volontà. E del resto, come potremmo noi ridiscutere la scelta artificiale, scovare l'errore, anticipare nel dubbio una soluzione diversa e, a conti fatti prima, migliore di quella artificiale; competere con la potenza di calcolo dei nuovi computer quantistici capaci di informazioni misurata in qubit, capaci persino di sovrapposizioni di stati, e di competenze letteralmente illimitate; dissentire da un idea, un suggerimento – giacché la finzione è quella ancora di un semplice *suggerimento*[8] – per un tragitto, percorso magari persino tante volte, rispetto alla nuova opzione di un percorso alternativo, deciso dalla IA sulla base di una complessa elaborazioni di dati, condizioni meteo, di traffico, di ambiente di cui Waze è solo una delle tante applicazioni[9]; pensare di affidarci all'occhio umano, piuttosto che a quello profondo di un sonar, di un radar, di sistema di individuazione, di una lente focale parabolica o satellitare; allo sguardo di un giudice di linea piuttosto che all'occhio di

[8] Un semplice *pungolo*, potremmo dire, un suggerimento di voltare attraverso una voce gentile ed educata che non protesterebbe se la svolta non avviene nei termini suggeriti: il che, ovviamente, dipende anche dalla quantità e qualità dei *pantani*. In argomento R.H. Thaler, C.R. Sunstein, Nudge. *La nuova strategia per migliorare le nostre decisioni su denaro, salute, felicità*, Universale Economica Feltrinelli, Milano, 2024, p. 290.

[9] Y.N. Harari, in Homo Deus. *Breve storia del futuro*, Bompiani, Bologna, 2023, p. 416, ricorda come Waze non sia solo in grado di allontanarci da eventuali ingorghi e condurre l'utente verso itinerari più veloci, suggerendoci di svoltare a sinistra anziché a destra, come ordinariamente un automobilista sarebbe stato tentato di fare, ma aggiunge la capacità di distribuire il traffico, evitando così che una fuga da un ingorgo crei un secondo ingorgo, determinato dallo spostamento verso l'itinerario alternativo. L'intelligenza di Waze è, dunque, quella di individuare il percorso 2 senza intasarlo, convogliando solo la metà degli automobilisti del percorso 1, in modo che né l'uno né l'altro ne risultino intasati. Si pensi anche – come ci ricorda S.J. Russell, *Compatibile con l'uomo. Come impedire che l'IA controlli il mondo*, Einaudi, Torino, 2025, p. 147 – al traffico aereo: all'inizio i computer si limitavano a programmare gli orari dei voli, ma ben presto si sono occupati della assegnazione degli equipaggi, dei posti a sedere, della manutenzione di routine, fino, ad oggi, a fornire informazioni sullo stato dei voli e a cambiare le rotte degli aerei, spostare gli equipaggi, modificare le prenotazioni e riorganizzare i turni delle manutenzioni.

falco presente a Wimbledon o in un torneo di cricket; di dubitare che la sedia esposta in un museo con il nome di un artista, al riparo dei visitatori, sia una opera concettuale sulla quale non è possibile sedersi; ancora (e per avvicinarci al tema penalistico di una colpa medica): come potremmo insistere in una diagnosi o in una terapia a dispetto del suggerimento di Heron, o di Flamingo, o BlueJay, infine Sycamore o Zuchongzhi, "dottori" quantistici capaci di eseguire un miliardo di porte e di risolvere i più importanti problemi della medicina e, direi, del mondo.

Nel "*come-potremmo*" c'è tutto il dramma esistenziale di un confronto di intelligenze che è ormai sempre meno un incontro, piuttosto confronto sudditante che ha alterato i meccanismi di funzionamento della mente, dove, cioè, è mutato il modo del ragionare riflessivo, in assoluta corrispondenza con la riduzione del pensiero contemplativo e della libertà decisoria, alterata da una radicale trasformazione dell'architettura del pensiero, certamente anche di quello che accompagna l'agire umano (e tanto più quello dei saperi esperti), dove la coscienza si smarrisce e l'inconscio e l'inconsapevole vivono il brivido della immersione in un ignoto che non sono in grado di controllare.

Più che contenuti di saggezza punitiva, la coscienza e volontà finiscono per sembrare ormai, o corrono il rischio di sembrare, contenitori di una altrui intelligenza, capace di alterare, o persino smarrire, la capacità di autodeterminazione dell'agente, il cui predicato di personalità cosciente e volontaria è semmai affidato alla condizione che precede l'immersione nell'ignoto, sempreché, per la logica penale, questo possa davvero poter bastare.

5. Verso una intelligenza superficiale

Come tutto ciò che ha accompagnato i punti di svolta della storia umana, che si tratti di oggetti meccanici, di forma della scrittura, di misuratori del tempo o quant'altro, gli strumenti tecnologici, nell'offrire aiuto al modo in cui si vive, si archivia, si pensa, si abita lo spazio, hanno alterato l'andamento delle cose umane e modellata la struttura fisica e il funzionamento della mente umana[10]: l'avvento del gps, nel momento in cui ha sostitu-

[10] N. Carr, *op. cit.*, p. 69.

ito le mappe, ha anche affievolita la capacità dei taxisti londinesi di elaborare modelli dello spazio e della memoria, riuscendo così a mutare il funzionamento, ma prima ancora la struttura, del sistema cerebrale, adattato alla nuova tecnologia. L'invenzione della scrittura – per non sostare solo nel recinto del ragionamento punitivo – ha modificata la capacità di memoria, tipica della tradizione orale e liberata la mente da esercizi e sforzi cognitivi ora applicati, ma in forme differenti, ad altre attività; e, subito dopo, il solo e semplice uso dello spazio fra parole, capace di interrompere la *scriptura continua*, è riuscito ad alleviare lo sforzo cognitivo legato a parole lunghe e continue, consentire una nuova abilità del cervello nel decodificare il testo, consentire, cioè, un processo automatico, capace, più di prima, di agevolare la riflessione, e la interpretazione più profonda: una semplice, minuscola, apposizione di uno spazio fisico fra parole, è riuscita a prescindere dal lavoro dello scriba, diffondere la lettura silenziosa e incentivare la scrittura di storie finalmente personali, avventurose, anticonvenzionali, quanto non lo sarebbero state nella dettatura ad uno scriba; è riuscito, infine, a produrre una modifica del processo neurofisiologico della lettura, dell'apprendimento, della cultura e della società[11].

Ma la storia sa essere bizzarra. Se lo spazio-fra-parole ha cambiata la mente, quella del lettore, dello scrittore, della comunità pensante; e se la invenzione di Gutenberg, la nuova etica della produzione libraria, favorendo la quiete del pensiero e il flusso di coscienza, ha modellata la mente letteraria, rendendola *meditante*, l'avvento delle nuove tecnologie, portatrici di una smaterializzazione della scrittura e una alterazione della relazione con la mente lettrice, hanno recato al seguito, in un singolare contrappasso, una differente capacità di lettura e di riflessione di una mente sempre più *calcolante*.

Il nuovo ambiente, la nuova scrittura, le nuove forme di lettura, la nuova ampiezza di uno spazio non più fisicamente ristretto ai margini di una pagina, un tempo avvolta in due fogli rigidi, ed ora aperta alla infinita complessità della navigazione di rimando, hanno favorita la lettura veloce, distratta, affrettata, l'apprendimento superficiale, in una conseguente e inevitabile alterazione

[11] Diffusamente *ivi*, pp. 83 ss.

della mente e delle sue funzioni. Il lettore attento è stato sostituito da un disattento osservatore, il pensiero cosciente da uno meno cosciente, l'attenzione profonda da una *distrazione dalla distrazione*, la mente pacata e non esagitata, da un *cervello del giocoliere*, affamato di dati, salti logici, finestre di dialogo per un continuum di *link* e rimandi quali moderna espressione di una lettura continua, priva di spazi di riflessione.

Ogni strumento in fondo dischiude nuove e, per noi, meravigliose ed innumerevoli possibilità, ma impone anche nuove limitazioni e nuovi adattamenti: il cervello, la mente *multitasking*, è ora chiamata alla rapida occhiata, alla lettura di sfuggita, all'accumulo dei dati, alla velocità di entrata e di uscita, senza esitazione, senza (ri)pensamenti, in una moderna passione dell'irrilevanza, quale anestetico della fatica dell'analisi, della comprensione, della profondità.

Il cervello, il nuovo cervello, non ha tempo, non attende *il lento e scrupoloso esame del tempo*[12], la fretta impone i suoi tempi, le sue narrazioni subitanee, scoraggia il pensiero lento, coltiva il business della distrazione, sacrificando le sfumature, i significati più autentici, sull'altare della immediatezza e della frammentazione del pensiero. Conviene fidarsi e, perché no, affidarsi.

Non è il giuoco semantico dell'ossimoro *intelligenza stupida*, più semplicemente la consapevolezza che l'intelligenza può condurre a scelte e decisioni inconsapevoli e persino talvolta irrazionali, dal momento che la formazione di opinioni vere, la cosiddetta intelligenza o razionalità epistemica, può essere deviata da scopi e desideri di un agente, guidato, in questi casi, da una intelligenza o razionalità strumentale. Nelle scelte umane le due dis-razionalità convivono nelle insidie del ragionamento disgiuntivo, delle decisioni viscerali, delle decisioni condizionate da contesti irrilevanti, cui si aggiungono criticità di pensiero, quali la fiducia immotivata nella propria conoscenza, la tendenza a non cercare di falsificare le proprie ipotesi, la tendenza a ricercare sempre e comunque gli eventi fortuiti, rendendo con-

[12] N. Carr, *op. cit.*, p. 205.

sueto e reale ciò che inconsueto e ci dà un senso di estraneità[13], e ignorando ipotesi alternative e le insidie del *myside bias*[14].

L'IA riesce ad evitare le prime, sfuggire ai *bias* ed euristiche semplificanti, ai *default bias*, agli effetti *framing*, a pregiudizi emotivi, alle *echo chambers*[15], ma non sempre riesce a sottrarsi alle seconde, mettersi al riparo da una irrazionalità epistemica, dalla incapacità di mettersi in discussione. Sembra un paradosso: la molta intelligenza non è detto che sia garanzia di molta razionalità, in ragione del fatto – e qui il paradosso è davvero manifesto – che il dubbio non abita i sistemi artificiali, quella medesima condizione di incertezza che si vuole combattere, aiuta il sistema ad essere razionale. L'euristica del dubbio compone la razionalità, ne guida le scelte, suggerisce alternative, si interroga sulla falsicabilità delle opzioni e, in questo stato di allerta, ne previene l'irrazionalità; una *passione triste* che sa, cioè, opporsi alla cristallizzazione, agevolare la fluidità della razionalità, suggerire una rimeditazione della propria idea.

Nei pochi secondi di uno sguardo che non ama perdere tempo, che non ha memoria del passato, ma solo dell'esistente, l'errore non solo non è contemplato, ma non è neppure rilevante, potendo essere oltrepassato da un nuovo errore che, al pari del precedente, attende un nuovo superamento indifferente all'attributo di verosimiglianza; come se la liturgia della dimenticanza fosse capace di involgere anche l'errore, che, non venendo ricordato, non può essere neppure emendato. Come dire: dall'intelligenza artificiale, in un attimo, all'intelligenza superficiale.

6. Il diritto penale ipnotico

È questa, probabilmente, la più grande, o almeno la più vicina, insidia della intelligenza artificiale, quella di cui prossi-

[13] Le insidie degli universi consensuali sulle menti individuali e sulle rappresentazioni collettive/individuali, sono evidenziate anche da S. Moscovici, *Le rappresentazioni sociali*, il Mulino, Bologna, 2016, alle pagine 42 e ss.

[14] W. Quattrociocchi, A. Vicini, *Liberi di crederci. Informazione, internet e post-verità*, Codice, Torino, 2018, p. 82.

[15] Per una analisi di questi luoghi ideali per la radicalizzazione delle proprie opinioni, *ivi*, pp. 99 ss., e degli stessi Autori, più diffusamente, *Polarizzazioni. Informazioni, opinioni e altri demoni nell'infosfera*, FrancoAngeli, Milano, 2023, pp. 37 ss.

mamente converrà (preo)occuparsi: fronteggiare l'errore (umano) guidato dalla intelligenza (artificiale). Fingere, continuare a fingere, che l'errore – laddove sussumibile in uno schema di illecito penale, e tanto più nel territorio dei saperi esperti – sia un errore proprio, cosciente e volontario, fingendo di ignorare la partenogenesi dell'errore, sempre più profonda e umanamente impenetrabile.

Più ancora che nel passato, il richiamo alla coscienza e volontà sembra ormai contenere l'equivoco di una leggerezza semantica, cui certamente il sistema penale non poteva rimediare, non poteva, cioè, risolvere il buco nero della co(no)scienza, quale mistero insolubile, e tanto più alla ignara ed assiomatica dogmatica penalistica. Il punto è che, ora, l'assioma penalistico, funzionale alle dissonanze di *noluntà* legate a speciali stati patologici, potrebbe sempre più assomigliare ad una verità apofantica per esperienze decisorie e decisionali, rese inconsapevolmente incoscienti e involontarie: una finzione di *suitas* decisoria, aggrappata alla antica idea di coscienza e volontà, ma ridotta ad una parimenti illusoria *actio libera in causa*, anacronisticamente appartenuta all'antica idea di un iniziale determinismo biologico delle decisioni algoritmiche.

Nella immensa, se non infinita, complessità di dati, nell'inestricabile groviglio di pensieri, nel congedo dalla rassicurante logica binaria delle nostre tradizionali forme di comunicazione, l'intelligenza umana è, se così fosse, destinata ad abbandonare la ridondanza, i tentativi e gli errori, la verifica e la falsificazione, i tradizionali metodi di acquisizione del sapere, i dubbi, i fallimenti, persino l'infelicità che ad essi si accompagna, a vantaggio di un mondo di conoscenze impeccabili, comode, impigrite, abitato da *prigionieri liberi*[16]. Persino l'atto illecito potrebbe essere deciso *fuori di sé* o, meglio, da un sé preso in prestito *dal fuori*: e l'io del soggetto agente, fintamente cosciente, diventare né più né meno che un centro di connessione temporaneo di varie esperienze portate da fuori alla nostra (dis)attenzione[17].

[16] È il titolo del bel saggio di G. Trautteur, *Il prigioniero libero*, Adelphi, Milano, 2020.

[17] Nelle applicazioni belliche i sistemi come Lavender e Hasbora sono in grado selezionare migliaia di obiettivi sulla base di incroci di ogni genere di informazioni, come telefonate, mail, posizionamenti gps, contatti, pagamenti

La seconda fra le due più grandi incognite del sapere umano, l'origine dell'universo e l'essenza della coscienza, la logica del tutto e quella del poco, dell'universale e del particolare, eretta, ciononostante, a requisito dogmatico per un addebito che si vuole, con apodittica sicumera, cosciente e volontario, è destinato a comporre il mistero scientifico più affascinante e sconcertante dell'origine delle esperienze quotidiane, senza che sia possibile comprendere esattamente cosa sia la coscienza, e quanto essa sappia diventare aliena; cosa la componga e la origini, e quanto essa resista alla complessità dei nuovi soggetti coscienti e alla magia delle interferenze artificiali dei nuovi agenti artificiali (coscienti o meno che siano, o che saranno). Niente di certo, tutto ancora da capire, poco, ancora meno, del quasi.

A questa prima e impredicibile sicumera, l'arretramento o meglio la finzione di coscienza come apodittico e ineffabile criterio di una teoria altrimenti inefficace, il sistema è destinato ad aggiungere un secondo inganno, quella di una quasi-volontà, in una complessiva quasi-colpevolezza, posta anch'essa, nel giuoco del doppio *quasi*, a base dogmatica di un addebito punitivo chiamato a misurare la meritevolezza di pena secondo logiche del *quanto-basta*: il sintagma che ne esprime la sapienza è l'"*avremmo-potuto-impedirlo*".

Avremmo-potuto, ecco il punto, la seconda grande finzione euristica. Nel *nuovo* mondo del quantitativo e del quantificabile, diverso dal *vecchio* del qualitativo e delle sfumature[18], in questo amalgama di intelligenza umana e di intelligenza artificiale, una sorta – potremmo dire – di *intelligenza universale* capace di fondere le esperienze coscienti e ignorare talvolta l'una, talaltra l'altra, in questa terra di confine di un *sé percipiente* e di un *sé raziocinante*, la colpa penale, una volta entrata l'*IA nella mente*[19], non potrà che essere, semplicemente, quella di essersi (af)

on line e così via. Una volta individuato il bersaglio, il sistema è in grado di prevedere il numero di vittime collaterali civili e sulla base di una scala di tollerabilità, da 10 a 100, decidere l'iniziativa ed accettarne gli effetti, in base al grado di importanza del bersaglio: nel solo conflitto israelo-palestinese il sistema IA Lavender ha selezionato circa 37000 obiettivi mediante sistemi di *machine learning* per l'analisi di big data: così L. Perilli, *op. cit.*, p. 116.

[18] *Ivi*, p. 179.

[19] *L'IA nella mente* è il titolo di un saggio in corso di ultimazione.

fidati e di non aver tentato di mantenere gli errori entro il limite del noto e dell'evitabile: sostituire inferenze da entità sconosciute con costruzioni a partire da entità conosciute[20], pur se, queste ultime, non del tutto sincere e veritiere, tuttavia più comode, e semplicemente accettate come valide: come dire, l'euristica del controllo, sui saperi esperti; vale a dire, abitare attenta-mente la dis-attenzione; *scoprire l'irrazionale*[21] di una forma di intelligenza oltremodo inumana, del tutto indifferente ai bisogni più profondi della nostra specie; una ragione squilibrata capace di impadronirsi dell'anima della scienza sotto forma di etereo fantasma di spirito maligno, che aleggi, silenziosamente, sopra la testa dell'intelligenza umana e produca un'influenza perniciosa, guidata dalla logica e tuttavia, al contempo, del tutto irrazionale, capace di fare irruzione nel mondo, penetrare nella nostra vita, bisbigliare promesse di potere sovrumano e controllo divino in grado di ammaliare gli uomini e le donne delle menti più brillanti. Scoprire dunque, e una volta scoperto, rimproverare per ciò che *avremmo-potuto* fare altrimenti, ammesso che sia per davvero possibile *aver-potuto*.

Non sarà, quello nell'*officina del diavolo*[22], un processo immediato, per quanto inesorabile esso sia destinato ad apparire[23]: occorreranno forse due o tre generazioni per radicare la diseducazione all'impegno personale, per pietrificare la legge

[20] Le parole di Bertrand Russell sono citate da D. Hoffman, *L'illusione della realtà. Come l'evoluzione ci inganna sul mondo che vediamo*, Bollati Boringhieri, Torino, 2020, p. 261.

[21] La scoperta dell'irrazionale e le parole che seguono appartengono alle pagine 17 e poi alle pagine 36 e 37, del racconto che fa B. Labatut, *op. cit.*, della vicenda umana di Paul Ehrenfest, grande fisico e matematico austriaco, legato da grande amicizia sia ad Albert Einstein che a Niels Bohr: le di Lui preoccupazioni sull'insorgenza della nuova razionalità che cominciava – siamo nel 1933, nel pieno del dibattito fra la fisica tradizionale e la nuova fisica quantistica – a profilarsi intorno alle straordinarie menti del tempo, appaiono profetiche rispetto alle attuali preoccupazioni, ancora maggiori rispetto ad un passato che pure era assai distante dalla modernità artificiale.

[22] R.M. Sapolsky, *Determinati. Biologia, comportamento e libero arbitrio*, ROI Edizioni, Milano, 2024, p. 288.

[23] È del tutto logico immaginare che i processi sociali, una volta interiorizzati, posseggano la capacità di determinare quelli psichici, come ci ricorda S. Moscovici, *op. cit.*, p. 98.

del minimo sforzo, per naturalizzare la delega inconsapevole e profonda ai sistemi algoritmici, ma il prodotto di questo nuovo ordine delle cose non potrà non sublimare un nuovo agire incosciente e involontario; rendere il sintagma dell'art. 42 c.p. – almeno in settori ad alta specializzazione, come quello medico o ingegneristico – ancora più approssimato e incerto. Nel presente si tratterà, semplicemente, di tamponare la falla, a che l'*euristica punitiva* – ancora una sostituzione del probabile incerto sul certo improbabile – non smarrisca il suo potere, e il prezzo da pagare non potrà che essere un ripensamento della colpa, una limitazione degli oggetti, destinatari di una *debole pienezza* psicologica nel cui ossimoro trovi spazio l'idea salvifica di un libero arbitrio, *semplicemente* indebolito[24], ma pur sempre legato all'*avremmo-potuto*; nel futuro, invece, il giuoco fittissimo di interferenze, l'oggettiva inesigibilità del sintagma accusatorio, l'umana cioè impossibilità dell'*avremmo-potuto*, si renderà portatore di un naturale allontanamento dall'obbligo costituzionale e convenzionale di una piena responsabilità penale, il che, in ultima analisi, potrebbe suggerire un elogio della fuga, una rinuncia all'esangue pensiero punitivo, e così accompagnare il diritto penale artificiale, quel particolare campo del diritto penale, verso una nuova e diversa umanità ultra metafisica, a condizione, s'intende, che sia davvero possibile farlo. Come dire: una nuova esperienza per un diritto penale ipnotico, costretto *inconsapevolmente* ad una *euristica del quasi*.

Perché non sia il *niente*, il *quasi* potrebbe, per l'intanto, poter bastare. Sapendo, però, che fra non molto, più presto che tardi, decidere e volere sarà sempre meno una espressione di libertà, e allora l'*impasse* del libero arbitrio non potrà non restituirci un rassegnato *nuovo* sintagma di apertura punitiva: "*fiat voluntas sua*".

7. Epilogo. Una partita a Go con Lee Sedol

La leggenda racconta che il giuoco del Go sia stato inventato dall'imperatore Yao per illuminare suo figlio, Danzhu. Con

[24] Sulla complessità che l'avverbio esprime, rinviamo al punto di vista di R.M. Sapolsky, *Determinati. Biologia, comportamento e libero arbitrio*, cit., *passim*, e dello stesso Autore, *L'uomo bestiale. Come l'ambiente e i geni costruiscono la nostra identità*, Castelvecchi, Roma, 2014.

l'aiuto di sua madre, l'Imperatore pregò i Quattro Re Celesti, i nove soli, la Beata Regina dell'Occidente e Pangu stesso, il primo essere vivente di questo universo, chiedendo loro il permesso di dividere l'intero cosmo in una griglia di diciannove linee per diciannove, creando una tavola con trecentossesantuno intersezioni su cui poter giocare una partita con pietre bianche e nere, per conquistare più spazi possibili e circondare le pietre dell'avversario. Il vincitore avrebbe dominato il mondo[25].

Tramandato di generazione in generazione per quasi 3000 anni, praticato da milioni di asiatici, cinesi, giapponesi, coreani, il giuoco del Go, il più antico del mondo e il più studiato dall'umanità, appariva come un'arte più che un semplice giuoco, un'arte ammaliante e destabilizzante, complicata da più di duecento possibilità per ogni singola mossa, entusiasmante nelle sue oltre duecento mosse, intelligente nelle sue centotrentamila combinazioni, e Lee Sedol ne divenne l'interprete più raffinato, venerato come un eroe e una celebrità di fama mondiale. Cresciuto nella minuscola isola di Bigeumdo, trasferitosi a soli 5 anni a Seul, divenuto a soli 8 anni il più giovane vincitore della dodicesima competizione nazionale di Go per bambini, durante la quale aveva mostrato quelle che sarebbero diventate le sue caratteristiche principali, la *Pietra forte* – questo il nome che gli venne assegnato – usava una tecnica focosa, violenta, imprevedibile, temeraria, divenuta invincibile in tutte le migliaia di partite nelle quali aveva sempre sbaragliato (e deriso) i propri avversari.

Il Go era molto più degli scacchi: niente torri, o alfieri, o pedoni e cavalli, niente regine o re, solo pietre e di due soli colori, il nero e il bianco. Dall'altra parte del mondo, cresciuto sin in tenera età all'ombra del giuoco degli scacchi, vincitore a sei anni del campionato londinese under 8 e tre anni dopo capitano della nazionale inglese juniores, in un periodo in cui l'Inghilterra era nel gioco degli scacchi seconda solo all'Unione sovietica, l'attrazione per gli scacchi venne subito soppiantata dalla ancor più grande passione per il mondo artificiale, e i computer divennero per Dennis Hassabis la principale occupazione, che

[25] Il bellissimo racconto del giuoco e della grande sfida fra Lee Sedol e AlphaGo è, anch'esso, nel già citato saggio di B. Labatut, *op. cit.,* pp. 277 ss.

sfociò nella creazione, con i suoi più cari amici, Shane Legg e Mustafa Suleyman, di una startup che fosse in grado di risolvere il problema dell'intelligenza artificiale generale, e poi usarla per risolvere tutti gli altri problemi.

Venne fondata nel 2010, e mentre tutti si chiedevano dove DeepMind avrebbe cominciato a concentrare le sue attenzioni, la malattia del cancro, la fusione nucleare, un nuovo mezzo di comunicazione, Hassabis non ebbe alcun dubbio: avrebbero cominciato con il giuoco più complesso e profondo che l'umanità avesse mai concepito, il gigantesco e ipnotico[26] giuoco del go, dove lo spazio combinatorio, le dimensioni cioè dell'albero che traduce l'infinita possibilità di configurazione derivate da ogni mossa è semplicemente gigantesco, dove, mentre il numero totale delle possibili partite di scacchi si avvicina a 10^{123} ovvero un seguito di 123 zeri, il numero di tutte le possibili partite di Go è inconcepibilmente maggiore più di 10^{700}; e dove il numero delle posizioni consentite, le configurazioni cioè di pietre che possono configurarsi quando due giocatori si affrontano, è talmente grande che sono nel 2016 fu possibile stabilirne con precisione il numero esatto, un numero immenso e impronunciabile. Questo:

> 208.168.199.381.979.984.699.478.633.344.862.770.286.
> 522.453.884.530.548.425.639.456.820.927.419.612.738.015
> .378.525.648.451.698.519.643.907.259.916.015.628.128.54
> 6.089.888.314.427.129.715.319.317.557.736.620.397.247.0
> 64.840.935.

E dove il numero delle partite teoricamente possibili raggiunge una cifra così grande da essere impossibile scriverla in forma decimale. Nessun box potrebbe contenere i numeri che traducano questa formula:

$$_{10}10^{100}$$

[26] L'ipnosi che il giuoco riesce a generare raggiunge situazioni quasi inimmaginabili: si racconta – ce lo ricorda B. Labatut, *op. cit.*, p. 129 – che, quando venne sganciata Little Boy su Hiroshima, si stesse giocando, ad appena cinque chilometri dal punto zero, la terza partita di Go fra i maestri giapponesi Hashimoto Utaro e Iwamoto Kaoru. L'edificio in cui stavano giocando venne quasi completamente distrutto, ma i due Maestri tornarono il giorno stesso dopo pranzo per continuare a giocare fino a sera, mentre intorno a loro si estraevano corpi dalle macerie e l'intera città era in fiamme.

Non poteva che essere quella la sfida del secolo. Fu organizzata nel 2016: da un lato Lee Sedol, maestro di 9º dan, il giocatore più creativo, aggressivo, incauto, autore di mosse stupefacenti (come la mossa della "scala rotta" sfoggiata nella sfida contro Hong Chang-sik nel 2003), vincitore di 18 titoli internazionali e 32 campionati nazionali; dall'altro AlphaGo, partorita dalla mente di Demis Hassabis e i suoi amici, già capace di stracciare Fan Hui, vincendo tutte e cinque le partite giocate.

Al sesto piano del nuovo Four Seasons di Seul, alla presenza di duecento giornalisti da tutto il mondo, con più 100.000 persone collegate al canale Youtube e per un pubblico di 60 milioni di spettatori sintonizzati sui 7 canali televisivi cinesi giapponesi e sudcoreani, ebbe inizio la sfida fra Lee Sedol e Aja Huang, uno dei più esperti programmatori di DeepMind, incaricato di effettuare le mosse decise dall'agente intelligente. Dunque, fra l'intelligenza umana di Lee Sedol e quella artificiale di AlphaGo.

Dopo due ore di partite, quando Sedol sembrava indiscutibilmente in vantaggio, Huang posizionò la pietra bianca 102 sulla decima linea a metà della tavola a due caselle dal bordo sinistro, cambiando le sorti della partita: con questa unica pietra, una mossa perfida e aggressiva, AlphaGo aveva cambiato il corso della partita e indotto Lee Sedol, 80 mosse dopo, a posizionare una pietra bianca – non una di quelle nere con cui stava giocando – nel mezzo della tavola, dichiarando in questo modo, nel modo più garbato possibile, la propria resa.

La seconda partita venne giocata il 10 Marzo del 2016. La partita non mostrava particolari emozioni, fino a quando Lee non posizionò la pietra 36, cui seguì nel monitor di Aja Huang l'indicazione di una *spallata* alla quinta linea, una mossa che negava tutto quello che i giocatori di Go tenevano per vero. Una mossa sbagliata, derisa e vilipesa, che nessun giocatore esperto, nei tremila anni di storia, avrebbe mai effettuato o consigliato. Tuttavia, decisa dalla macchina e apparsa, dopo, come un autentico colpo di genio. La *mossa 37*, si capì dopo, era in realtà bellissima, impensabile, una mossa di cui alcuno umano era in grado di comprendere la forza e la bellezza, al punto stupefacente, da essere apparsa a molti come l'erronea lettura delle indicazioni sul monitor da parte del computer. Aja Huang aveva in

realtà eseguito esattamente l'indicazione del computer, attuando quella che passò alla storia come la *mossa 37* oltre alla quale non vi fu più alcuna possibilità di vittoria, la quale giunse inesorabilmente alla mossa 211 quando Sedol fu costretto a darsi per vinto. La *mossa 37* non faceva parte della memoria di AlphaGo e non era né programmata, né conforme alle linee guida inserite nel cervello di silicio, era, dunque, una mossa creata dal programma, senza alcun input umano, ma, soprattutto, era una mossa che nessun maestro di Go avrebbe mai preso in considerazione; era una mossa che il sistema aveva imparato giocando da solo e che la logica avrebbe portato ad escludere, in un calcolo delle probabilità di buon esito era pari a 1 su 10.000 e ciononostante decisa dal computer e rivelatasi vincente oltre ogni aspettativa umana.

La terza partita ebbe inizio nel pomeriggio del 12 Marzo e portò al medesimo esito delle prime due, quanto basta perché Lee Sedol non desiderasse altro che ritirarsi dal torneo, se non fosse che le regole imponevano che la sfida dovesse essere celebrata sulle 5 partite, dunque, si doveva affrontare AlphaGo altre due volte e non vincere neanche una partita sarebbe stata una onta quasi insopportabile. Nella quarta partita, quando anche questa sembrava in qualche misura segnata, Lee decise di posizionare la pietra 78 in una mossa a cuneo, diversa da qualunque cosa chiunque avesse mai visto: "*la mano di Dio, questa è una mostra divina*" così fu letta la mossa di Lee che determinò una serie di svarioni di AlphaGo al punto che Aja fu costretto a prendere una pietra, posizionarla sul bordo della tavola, e inchinarsi alla vittoria di Sedol.

Malgrado la debolezza e la fragilità degli esseri umani la vittoria della quarta partita, la genialità della *mossa 78*, divenuta famosa e pari per genialità alla *mossa 37*, aveva dimostrato che, sebbene fosse difficile battere l'IA, era possibile immaginare una resistenza umana, la capacità, cioè, di fare appello alla propria intelligenza, al proprio intuito, alla propria genialità, anche rispetto alla mastodontica intelligenza dei sistemi di intelligenza artificiale. C'era ancora speranza, questo il lascito emotivo della mossa a cuneo, e malgrado anche la quinta partita venne vinta dal sistema artificiale, la sola vittoria della quarta partita aveva

restituita fiducia e speranza, per una umanità non del tutto soggiogata alla potenza artificiale.

Nessuna vittoria sarebbe stata, ed è stata mai, altrimenti possibile. In seguito, il nuovo programma congegnato, ripulito di qualsiasi informazione umana, privato di qualsiasi connessione diretta con l'umanità, si sarebbe palesato ancora più terrificante, vincendo cento partite su cento contro AlphaGo, l'unico avversario in grado di competere con il nuovo sistema artificiale. Partendo da zero, da nessuna informazione, il *nuovo* riuscì a imparare da solo, giocando contro sé stesso e imparando da zero. Nessun umano avrebbe potuto resistervi. Il suo nome era AlphaZero.

Salvatore Antonello Parente

Intelligenza artificiale simbiotica tra procedimento di accertamento e processo tributario

Abstract: **Symbiotic Artificial Intelligence Between Assessment Proceedings and Tax Process.** Symbiotic Artificial Intelligence aims to establish a close and harmonious collaboration between humans and intelligent machines in order to enhance individual and collective capabilities by fostering new patterns of interaction. Rather than completely replacing human activity, this technology aims to enhance it by creating a future in which artificial intelligence is not merely an automated tool, but a true partner in innovation and problem-solving. In general terms, since intelligent machines perform a supplementary and instrumental function as a mere technical aid capable of supporting the legal reasoning of human beings, there is no reason against their possible use in the tax field. However, it is necessary to ascertain to what extent and with what forms of guarantee "Symbiotic AI" systems can be used both in the tax implementation phase and in litigation before the tax court.

Keywords: Symbiotic Artificial Intelligence, Assessment Procedure, Tax Process, Limits of Use, Taxpayer Protection Guarantees.

Sommario: 1. Premessa – 2. Le molteplici forme di impiego dell'intelligenza artificiale simbiotica in materia tributaria – 3. La tecnologia digitale quale strumento in grado di agevolare il corretto esercizio dell'azione impositiva. I vantaggi dell'utilizzo dell'intelligenza artificiale simbiotica nella fase di attuazione del tributo – 4. I limiti entro i quali l'amministrazione finanziaria può porre a fondamento della propria azione i risultati dell'elaborazione algoritmica – 5. L'algoritmo "Ve.R.A." e l'applicativo "VE.A.R." quali strumenti di AI a disposizione dell'amministrazione finanziaria italiana – 6. Gli ulteriori utilizzi dell'intelligenza artificiale simbiotica nella fase di attuazione del tributo – 7. La capacità predittiva dell'amministrazione finanziaria. L'impossibilità di fare assurgere le elaborazioni predittive a presunzione legale di evasione – 8. Il progetto pilota "*A data driven approach to tax evasion risk analysis in Italy*". I riferimenti all'intelligenza artificiale nella delega di riforma del sistema fiscale (l. 9 agosto 2023, n. 111) e in sede di attuazione – 9. Il contributo della giurisprudenza amministrativa: il requisito di conoscibilità dell'algoritmo – 10. Le implicazioni del principio di non esclusività della decisione algoritmica – 11. Le applicazioni dell'intelligenza artificiale simbiotica in ambito giurisdizionale – 12. I *software* di ricerca dei precedenti giurisprudenziali: l'iniziativa "Prodigit". Le possibili forme di automazione delle decisioni dinanzi al giudice tributario – 13. L'amministrazione della giustizia nella normativa in materia di intelligenza artificiale (l. 23 settembre 2025, n. 132). La valenza residuale dei sistemi di "*Symbiotic AI*" nell'ambito della giurisdizione tributaria – 14. Conclusioni.

1. Premessa

L'intelligenza artificiale simbiotica mira ad instaurare una collaborazione stretta ed armoniosa tra esseri umani e macchine intelligenti, in modo da potenziare le capacità individuali e collettive favorendo nuovi modelli di interazione.

Il termine "simbiotico" implica una relazione reciprocamente vantaggiosa tra due entità: nel contesto dell'intelligenza artificiale, si riferisce ad un sistema in cui esseri umani e macchine intelligenti operano in sinergia, combinando le rispettive capacità di azione per ottenere risultati superiori a quelli raggiungibili individualmente. Affinché questo modello realizzi appieno il suo potenziale, è fondamentale un approccio equilibrato che consideri sia gli aspetti tecnici che le implicazioni di natura giuridica e sociale derivanti dall'interazione tra uomo e macchina.

A differenza delle forme tradizionali di intelligenza artificiale[1], spesso basate sull'automazione e sulla sostituzione umana, l'intelligenza artificiale simbiotica si concentra sull'integrazione e sul miglioramento dell'intelligenza umana attraverso la tecnologia. Ne sono da esempio i sistemi di supporto decisionale avanzato, in cui le macchine intelligenti analizzano enormi flussi di dati per fornire *insight* e raccomandazioni, mentre l'essere umano mantiene il controllo sulle decisioni finali, consentendo alla tecnologia di affiancarsi alle capacità cognitive umane anziché operare come mero sostituto.

L'intelligenza artificiale simbiotica assurge, così, ad opportunità preordinata a trasformare il modo in cui esseri umani e macchine intelligenti collaborano: anziché sostituire completamente l'attività umana, questa tecnologia mira a potenziarla creando un futuro in cui l'intelligenza artificiale non sia un mero strumento automatizzato, ma un vero e proprio *partner* nell'innovazione e nella risoluzione dei problemi.

Non a caso, anche l'interesse del legislatore si è recentemente concentrato sull'utilizzo dell'intelligenza artificiale e sulla crescente simbiosi tra esseri umani e modelli intelligenti, alla luce

[1] R. Puglisi, "L'intelligenza artificiale: concetti principali e applicazioni in economia", in *Rivista di diritto finanziario e scienza delle finanze*, (2024), n. 4, I, pp. 341 ss.

delle sfide sociali ed economiche generate dalla diffusione dei sistemi di "*Symbiotic AI*"[2].

Emblematico, al riguardo, è l'approccio sperimentato in ambito UE, attraverso la predisposizione dell'*Artificial Intelligence Act*[3] (Regolamento UE/2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024), allo scopo di incentivare lo sviluppo dei sistemi di AI nel rispetto dei diritti fondamentali della persona, e l'adozione del "*Data Act*" (Regolamento UE/2023/2854 del Parlamento europeo e del Consiglio del 13 dicembre 2023)[4], in tema di equo accesso ed utilizzo dei dati, i cui effetti assumono rilevanza anche in ambito tributario[5], in ragione dell'incidenza sulla determinazione degli imponibili nell'era digitale[6].

La *ratio* dell'*AI Act* risiede nel proteggere i diritti fondamentali, la democrazia, lo Stato di diritto e la sostenibilità ambientale

[2] M. Caligiuri, "L'intelligenza artificiale tra regolamentazione europea e strategie nazionali", in A.F. Uricchio, C. Caldarola (a cura di), *L'intelligenza artificiale tra regolazione e esperienze applicative. AI ACT (regolamento UE) e disegno di legge governativo*, Cacucci, Bari, 2025, pp. 475 ss.

[3] L. Antonini, "Intelligenza artificiale e applicazioni nel campo della giustizia: profili problematici", in V. Mastroiacovo (a cura di), *Giocare con altri dadi. Giustizia e predittività dell'algoritmo*, Giappichelli, Torino, 2024, pp. 148 ss.; R. Geraci, *Intelligenza artificiale e giustizia predittiva in materia tributaria. Algoretica e tutela del contribuente 'religioso'*, Giappichelli, Torino, 2024, pp. 79 ss.; C. Novelli, "L'Artificial Intelligence Act Europeo: alcune questioni di implementazione", in *Federalismi.it*, (2024), n. 2, pp. 95 ss.; G. Taddei Elmi, A. Contaldo (a cura di), *Intelligenza artificiale. AI Act – Regolamento (UE) 1689/2024. Il nuovo scenario giuridico europeo*, Pacini Giuridica, Pisa, 2024.

[4] A. Morace Pinelli (a cura di), *Data Act. Introduzione interdisciplinare e commentario al Regolamento (UE) 2023/2854*, Pacini Giuridica, Pisa, 2025.

[5] L. Tosi, "L'evoluzione digitale del sistema tributario", in *Tax News*, (2025), n. 1, pp. 59 ss.

[6] L. Carpentieri, "Sull'utilizzo dei 'dati' nella definizione di nuovi presupposti di imposta e nella diversa determinazione di imponibili ricollegati a presupposti d'imposta tradizionali", in Carpentieri L., Conte D. (a cura di), *La digitalizzazione dell'Amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, I, Lefebvre Giuffrè, Milano, 2023, pp. 127 ss.; A. Fedele, "Intervento sulla rilevanza dei dati nella determinazione degli imponibili nell'era digitale", in *Rivista telematica di diritto tributario*, 19 settembre 2023, pp. 1 ss.

dai sistemi di intelligenza artificiale ad alto rischio, promuovendo al contempo l'innovazione e assicurando all'Unione europea un ruolo strategico nel settore[7].

Nella logica del regolamento, l'intervento umano sembra essere relegato alla definizione degli obiettivi dei sistemi basati sull'intelligenza artificiale: quest'ultima, invece, in maniera autonoma dovrebbe, secondo un processo di apprendimento, essere in grado di formulare una serie di determinazioni idonee ad incidere sui diritti dei destinatari. Si tratta di un approccio fondato sul rischio associato al singolo sistema con la previsione di regole preordinate a garantire la qualità dei dati utilizzati nello sviluppo degli algoritmi[8].

In questo contesto, il dato digitale sembra oramai strettamente connaturato all'imposizione e, in generale, a tutti i profili inerenti alla materia tributaria[9].

In realtà, la regolamentazione dell'intelligenza artificiale, oltre a costituire una priorità per il legislatore europeo, rappresenta un importante banco di prova per i singoli ordinamenti domestici, al fine di trovare una collocazione adeguata alle nuove tecnologie nel contesto dei valori, dei principi e delle regole comuni in ambito UE[10].

[7] C. Verrigni, "C'è spazio per un utilizzo ragionevole dell'intelligenza artificiale nell'ambito dell'accertamento tributario?", in *Diritto e pratica tributaria internazionale*, (2024), n. 3, p. 746; F. Gallo, "Cyberspazio, diritti e tutele", in A.F. Uricchio, C. Caldarola (a cura di), *op. cit.*, pp. 15 ss.

[8] C. Verrigni, *op. cit*., p. 747.

[9] A.F. Uricchio, S.A. Parente, "Data driven e digital taxation: prime sperimentazioni e nuovi modelli di prelievo", in *Diritto e pratica tributaria internazionale*, (2021), n. 2, pp. 606 ss.; M. Cedro, F.A. Cimino, "Intelligenza artificiale ed accertamento tributario: opportunità e rischi di una rivoluzione in atto", in *Diritto e pratica tributaria internazionale*, (2024), n. 1, p. 52.

[10] A. Marinello, "L'utilizzo dell'intelligenza artificiale da parte del Fisco: limiti e prospettive, anche alla luce dell'Artificial Intelligence Act dell'Unione Europea", in *Rivista telematica di diritto tributario*, 5 dicembre 2024, p. 11.

2. Le molteplici forme di impiego dell'intelligenza artificiale simbiotica in materia tributaria

In materia tributaria[11], le intelligenze artificiali simbiotiche e gli algoritmi informatici, oltre ad essere impiegati nei procedimenti di accertamento, possono esplicare un'importante funzione nelle procedure giurisdizionali, nella riscossione dei tributi[12], nella repressione degli illeciti fiscali[13] e nel reperimento degli elementi di prova da utilizzare sia nel corso della fase di attuazione del tributo che in sede contenziosa[14].

L'uso dell'intelligenza artificiale simbiotica nell'attività svolta dalle autorità fiscali sembra potersi ricondurre a due tipologie: la prima è l'analisi predittiva, ove gli algoritmi informatici e i sistemi di apprendimento automatico vengono utilizzati per estrarre modelli da grandi quantità di dati e per fare previsioni e valutazioni sui soggetti da sottoporre ad indagine; la seconda è la c.d. "elaborazione del linguaggio naturale", che fornisce tecniche di programmazione preordinate ad elaborare e analizzare grandi quantità di dati grezzi, scritti in un linguaggio naturale (compresi testi complessi) e restituire dati elaborati che l'operatore umano può facilmente comprendere ed utilizzare[15].

[11] R. Cordeiro Guerra, "L'intelligenza artificiale nel prisma del diritto tributario", in *Diritto e pratica tributaria*, (2020), n. 3, pp. 921 ss.; A. Marcheselli, "Intelligenza artificiale e tributi: alba di un nuovo mondo o folle volo oltre le Colonne d'Ercole?", in F. Farri, A. Marcheselli (a cura di), *La digitalizzazione dell'Amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, IV, Lefebvre Giuffrè, Milano, 2025, pp. 3 ss.; A. Uricchio, "Il fisco dinanzi a intelligenza artificiale e digital transformation", in A.F. Uricchio, C. Caldarola (a cura di), *op. cit.*, pp. 197 ss.

[12] P. Milioto, "La riscossione digitalizzata e il possibile utilizzo dell'intelligenza artificiale", in *Tax News*, (2025), n. 1, pp. 64 ss.

[13] D. Cané, "Intelligenza artificiale e sanzioni amministrative tributarie", in *Rivista trimestrale di diritto tributario*, (2019), n. 4, pp. 761 ss.; S. Dorigo, "L'intelligenza artificiale e il possibile 'rinascimento' del sistema sanzionatorio tributario", in *Rivista di diritto tributario*, (2023), n. 4, I, pp. 407 ss.

[14] F. Paparella, "L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi", in *Rivista di diritto tributario*, (2022), n. 6, I, pp. 649-652; S.A. Parente, "Fase di attuazione del tributo, giustizia predittiva e intelligenze artificiali", in *Unità e Pluralità del sapere giuridico*, (2024), n. 3, pp. 83 ss.

[15] M. Cedro, F.A. Cimino, *op. cit*., pp. 52-53.

L'utilizzo dell'intelligenza artificiale simbiotica nelle attività di prevenzione e di contrasto all'evasione fiscale se, da un lato, consente di "disincagliare" l'enorme mole di dati ed informazioni già a disposizione dell'amministrazione finanziaria, affiancando e non sostituendo l'intervento umano, presenta evidenti rischi e implicazioni sotto il profilo della tutela del diritto di protezione dei dati personali dei contribuenti[16].

In particolare, il diritto del contribuente alla protezione dei propri dati personali[17] costituisce un limite al massivo utilizzo di adempimenti comportanti l'acquisizione e la conservazione di grandi quantità di informazioni, non sempre di stretta rilevanza tributaria[18].

In questa prospettiva, sarebbe auspicabile un innalzamento degli standard di tutela previsti per i contribuenti sotto il profilo della protezione dei dati sensibili, attraverso una valutazione preventiva di impatto a carico dei sistemi di intelligenza artificiale simbiotica destinati ad essere utilizzati nell'ambito dei controlli tributari nello spazio comune europeo[19].

Uno dei pregi della tecnologia risiede nella possibilità di procedere all'archiviazione di un'ingente mole di informazioni

[16] A. Carinci, "Fisco e privacy: storia infinita di un apparente ossimoro", in *Il fisco*, (2019), pp. 4407 ss.; A. Contrino, "Banche dati tributarie, scambio di informazioni fra autorità fiscali e 'protezione dei dati personali': quali diritti e tutele per i contribuenti?", in *Rivista telematica di diritto tributario*, 29 maggio 2019, pp. 7 ss.; F. Farri, "Digitalizzazione dell'amministrazione finanziaria e diritti dei contribuenti", in *Rivista di diritto tributario*, (2020), n. 6, V, pp. 115 ss.; A. Contrino, "Spinte evolutive (sul piano sovranazionale) e involutive (a livello interno) in tema di bilanciamento fra diritto alla protezione dei dati dei contribuenti ed esigenze di contrasto dell'evasione fiscale", in *Rivista telematica di diritto tributario*, 3 ottobre 2023, pp. 1 ss.; A. Marinello, *op. cit*., p. 1.

[17] L. Izzo, "La tutela della privacy e il contrasto all'evasione tributaria", in *Tax News*, (2025), n. 1, pp. 246 ss.

[18] G. Salanitro, "Amministrazione finanziaria e Intelligenza Artificiale, tra titolarità della funzione e responsabilità", in S. Aleo (a cura di), *Evoluzione scientifica e profili di responsabilità*, Pacini Giuridica, Pisa, 2021, p. 414; A. Contrino, "Protezione dei dati personali e pervasività delle banche dati fiscali: quale contemperamento?", in A. Contrino, E. Marello (a cura di), *La digitalizzazione dell'amministrazione finanziaria tra contrasto all'evasione e tutela, dei diritti del contribuente*, II, Giuffrè, Milano, 2023, pp. 109 ss.

[19] A. Marinello, *op. cit.*, p. 13.

relative ai contribuenti, poi condivise tra le autorità fiscali dei diversi Stati[20]: nel tempo, tale sistema ha consentito di massimizzare la capacità dell'amministrazione finanziaria di svolgere controlli incrociati, pure attraverso l'utilizzo di strumenti digitali, in modo da individuare eventuali incongruenze meritevoli di approfondimento.

Lo sviluppo di sistemi intelligenti sempre più sofisticati ed efficienti è stato facilitato dallo sfruttamento massivo dei "dati", intesi quale rappresentazione digitale di atti, fatti o informazioni, di cui può avvalersi pure l'amministrazione finanziaria nello svolgimento delle proprie funzioni: intelligenza artificiale simbiotica e dati appaiono, dunque, strettamente connessi, nella misura in cui i secondi fungono da risorsa e misura di efficienza della prima.

Pur senza addivenire ad una reale "rivoluzione algoritmica"[21], tale sistema – già previsto e, in parte, realizzato nell'ordinamento francese e in quello degli Stati Uniti d'America, le cui autorità fiscali sono da tempo legittimate ad avvalersi di sistemi di intelligenza artificiale per scansionare i *social media* utilizzati dai contribuenti e le ulteriori fonti aperte (ad esempio, gli articoli di stampa, i siti internet e le informazioni rese pubbliche dall'interessato) alla ricerca di potenziali evasori[22] – è stato oggetto di attenzione da parte del Fisco italiano, che da tempo si avvale di banche dati telematiche per le attività di *intelligence*

[20] M. Pontillo, "Protezione dei dati personali dei contribuenti e contrasto all'evasione: il difficile bilanciamento nell'era della digitalizzazione", in G. Salanitro (a cura di), *Scenari dell'intelligenza artificiale. Uomo e Algoritmo nelle scienze sociali*, Pacini Giuridica, Pisa, 2024, pp. 149 ss.

[21] S. Dorigo, "Il tramonto delle regole fiscali tradizionali nell'economia del XXI secolo: rivoluzione algoritmica e tutela dei diritti", in R. Cordeiro Guerra, S. Dorigo (a cura di), *Fiscalità dell'economia digitale*, Pacini Giuridica, Pisa, 2022, pp. 29 ss.

[22] A. Falchetti, O. Salvini, "L'acquisizione delle informazioni nel web e delle risorse digitali durante le ispezioni fiscali", in *Rivista della Guardia di Finanza*, (2018), n. 6, pp. 1677 ss.; O. Signorile, "La ricerca di dati su fonti aperte come nuovo strumento delle indagini fiscali", in G. Ragucci (a cura di), *Fisco digitale. Cripto-attività, protezione dei dati, controlli algoritmici*, Giappichelli, Torino, 2023, pp. 113 ss.

e di verifica fiscale, attraverso l'utilizzo di elementi conoscitivi che assurgono ad una sorta di "oracolo fiscale"[23].

Ciò ha consentito di affinare le attività di prevenzione e di contrasto all'evasione fiscale avvalendosi di uno strumentario sempre più avanzato, consistente nello sfruttamento più efficiente delle nuove tecnologie e di strumenti di *data analysis*, anche al fine di realizzare uno degli obiettivi del PNRR, consistente nella riduzione del "*tax gap*"[24].

3. La tecnologia digitale quale strumento in grado di agevolare il corretto esercizio dell'azione impositiva. I vantaggi dell'utilizzo dell'intelligenza artificiale simbiotica nella fase di attuazione del tributo

Nella fase procedimentale[25], la tecnologia digitale può agevolare il corretto esercizio dell'azione impositiva – in ragione della mole considerevole di informazioni a disposizione dell'autorità

[23] G. Palumbo, "E domani?", in G. Palumbo, *Fisco e privacy. Il difficile equilibrio tra lotta all'evasione e tutela dei dati personali*, Pacini Giuridica, Pisa, 2021, pp. 115-116.

[24] A. Marinello, *op. cit.*, p. 3.

[25] A. Contrino, "Digitalizzazione dell'amministrazione finanziaria e attuazione del rapporto tributario: questioni aperte e ipotesi di lavoro nella prospettiva dei principi generali", in *Rivista di diritto tributario*, (2023), n. 2, I, pp. 105 ss.; A. Guidara, "Accertamento dei tributi e intelligenza artificiale: prime riflessioni per una visione di sistema", in *Diritto e pratica tributaria*, (2023), n. 2, pp. 384 ss.; F. Paparella, "L'ausilio delle tecnologie digitali nell'applicazione dei tributi", in L. del Federico, F. Paparella (a cura di), *Diritto tributario digitale*, Pacini Giuridica, Pisa, 2023, pp. 155 ss.; Id., "Procedimento tributario, algoritmi e intelligenza artificiale: potenzialità e rischi della rivoluzione digitale", in Marello E., Contrino A. (a cura di), *La digitalizzazione dell'Amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, II, Lefebvre Giuffrè, Milano, 2023, pp. 3 ss.; M. Cedro, F.A. Cimino, *op. cit.*, pp. 48 ss.; L. Carpentieri, "Sull'impatto della tecnologia e dell'intelligenza artificiale nel procedimento di attuazione del rapporto tributario: un'introduzione generale", in F. Montanari, A. Contrino (a cura di), *La digitalizzazione dell'Amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, III, Lefebvre Giuffrè, Milano, 2025, pp. 3 ss.; C. Ricci, "Intelligenza artificiale e procedimento tributario tra 'responsabilizzazione' dell'Amministrazione finanziaria e principio di legalità", in *Tax News*, 2 luglio 2025, pp. 1 ss; G. Ricotti, M.E. Castiglia, "Intelligenza artificiale e fisco: come cambia il rapporto tra Ammi-

fiscale – e consentire valutazioni sulla base di criteri oggettivi ed imparziali[26].

I sistemi di intelligenza artificiale, derivando in buona parte dalla logica, si avvalgono di criteri deduttivi e di parametri ulteriori, pur sempre di carattere matematico, che portano alla condivisione dei diversi approdi del ragionamento. Nell'accertamento tributario, si fa uso di automatizzazioni ogni qualvolta si "deducono" valori fiscalmente rilevanti da altri valori, anche non primigeni o reali: tali automatizzazioni possono essere effettuate dall'intelligenza artificiale, in grado di processare notevoli quantità di dati in tempi brevissimi e con il minimo margine di errore[27].

Difatti, essendo l'accertamento tributario una *species* di procedimento amministrativo, preordinato alla determinazione autoritativa dell'imponibile e dell'imposta dovuta, emergono non poche similitudini ed assonanze con la struttura dell'algoritmo su cui si fonda il funzionamento dei sistemi di intelligenza artificiale.

L'algoritmo – quale procedura computazionale che prende origine da dati iniziali, procede tramite l'elaborazione dei medesimi sulla base di un numero finito di regole fornite dal programmatore e si conclude fornendo dati in uscita – assurge a procedimento matematico che utilizza una successione di istruzioni per fornire la soluzione ad un problema, con probabilità di errore pressoché nulle ed in tempi estremamente ristretti: tali peculiarità caratterizzano il procedimento informatico come economico, efficace, efficiente, imparziale e rapido, rievocando i principi che presiedono l'azione amministrativa[28].

I vantaggi dell'utilizzo dell'intelligenza artificiale simbiotica nella fase di attuazione del tributo, connotata da una natura

nistrazione finanziaria e contribuente", in *Diritto e pratica tributaria*, (2025), n. 6, pp. 2222 ss.

[26] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., pp. 648-649; S.A. Parente, "The Contribution of Artificial Intelligence and Computer Algorithms to the Investigation Activities of the Financial Administration and to the Jus Dicere Function of the Tax Judge: What Prospects for the Suppression of Tax Evasion?", in *The Italian Law Journal*, (2023), n. 2, pp. 557 ss.; C. Verrigni, *op. cit.*, pp. 739 ss.

[27] A. Guidara, *op. cit.*, p. 397.

[28] M. Cedro, F.A. Cimino, *op. cit.*, pp. 59-60.

tendenzialmente vincolata[29], possono così sintetizzarsi: rapidità dell'elaborazione, con conseguente velocità del relativo procedimento amministrativo; efficacia, con garanzia sia di conclusione del procedimento in tempi certi, che di qualità del servizio reso; economicità, intesa come risparmio di risorse (umane e materiali) utilizzate e di tempo; efficienza, in termini di assenza di sprechi e di errori; imparzialità, in quanto i risultati sono ottenuti all'esito di un procedimento predeterminato da regole certe, senza rischi di ingerenze esterne.

Pertanto, l'attività di accertamento si presta, almeno sul piano teorico, all'applicazione di strumenti di automatizzazione e di intelligenza artificiale[30].

4. I limiti entro i quali l'amministrazione finanziaria può porre a fondamento della propria azione i risultati dell'elaborazione algoritmica

Trattandosi però di funzioni connotate da valenza pubblicistica, in cui emerge una contrapposizione tra il potere esercitato dall'autorità amministrativa e la tutela dell'amministrato, si rende necessario verificare in quale misura e con quali forme di garanzia, l'amministrazione finanziaria può porre a fondamento della propria azione istituzionale i risultati di un'elaborazione fondata su sistemi di intelligenza artificiale[31].

Un bilanciamento degli interessi coinvolti impedisce all'autorità amministrativa, nell'esercizio della propria azione, di eccedere quanto necessario a garantire il regolare funzionamento dei servizi indispensabili per la vita della comunità[32].

[29] L. Perrone, *Discrezionalità e norma interna nell'imposizione tributaria*, Giuffrè, Milano, 1969, pp. 63 ss.; F. Gallo, voce "Discrezionalità (diritto tributario)", in *Enciclopedia del diritto*, *Aggiornamento*, III, Giuffrè, Milano, 1999, pp. 536 ss.; M. Martis, *Contributo allo studio della discrezionalità nel diritto tributario*, Esi, Napoli, 2018, pp. 69 ss.

[30] M. Cedro, F.A. Cimino, *op. cit.*, pp. 60 ss.

[31] S. Dorigo, "L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici", in R. Cordeiro Guerra, S. Dorigo (a cura di), *op. cit.*, p. 204; Id., "Opportunità e limiti nell'impiego dell'intelligenza artificiale da parte del Fisco", in *Corriere tributario*, (2022), n. 11, pp. 965 ss.

[32] A. Purpura, "La frontiera dei Big data", in G. Palumbo, *op. cit.*, pp. 87-88.

È necessario, inoltre, non tralasciare taluni limiti oggettivi – relativi tanto ai dati in ingresso, quanto alle regole fornite dal programmatore e ai dati in uscita – che restringono notevolmente l'applicazione dell'intelligenza artificiale simbiotica alle attività di accertamento tributario.

In relazione ai dati in ingresso, la qualità della risposta fornita dall'elaboratore dipende dalla quantità e dalla qualità dei dati inseriti come input: se le informazioni utilizzate sono errate, incomplete o distorte, le relative conclusioni non saranno pienamente rappresentative della realtà, rendendo l'*output* non pienamente soddisfacente.

Ai fini dell'applicazione dell'algoritmo, è poi necessario che le istruzioni fornite dal programmatore alla macchina intelligente siano univoche e non ambigue, dovendo essere chiaramente specificate le condizioni iniziali e finali, unitamente all'ordine con cui eseguire le singole prescrizioni[33].

In ambito tributario ciò è estremamente difficile essendo spesso demandati all'amministrazione finanziaria, nel corso dell'attività di accertamento, giudizi non univoci, estremamente articolati e complessi, relativi a dati di fatto di difficile apprezzamento e non facilmente delegabili all'algoritmo, anche in ragione della complessità della normativa tributaria, sovente di difficile interpretazione.

Anche con riferimento ai dati in uscita sussistono criticità, essendo la soluzione fornita dall'algoritmo generalmente rivolta al passato, piuttosto che al presente, in quanto tesa a reiterare giudizi già espressi, senza tenere conto del mutamento delle condizioni di fatto e delle novelle legislative intervenute al momento di applicazione dell'algoritmo: il rischio è la sclerotizzazione dell'accertamento su dati prefissati ed il conformismo applicativo.

Occorre pertanto verificare che l'algoritmo tragga fondamento dalla normativa vigente al momento dell'applicazione concreta e che conduca a risultati pertinenti alle finalità perseguite dal legislatore[34].

In mancanza, il rischio concreto sarebbe quello di generare una profonda sproporzione tra la capacità computazionale

[33] M. Cedro, F.A. Cimino, *op. cit.*, pp. 62 ss.

[34] *Ivi*, pp. 63-64.

dell'amministrazione finanziaria e il diritto del contribuente a far valere le proprie ragioni.

Pertanto, un utilizzo non oculato dei *software* di AI nella fase procedimentale, oltre al rischio di atti fondati su elaborazioni non corrette, che condurrebbero all'individuazione di una base imponibile inesistente, potrebbe generare un appiattimento dell'attività istruttoria dell'amministrazione finanziaria sull'effettuazione di mere indagini di AI, che, per loro natura, non possono comprendere tutte le sfumature della vita umana, con conseguente compressione del diritto di difesa del contribuente, nonché della capacità di analisi del personale dell'Agenzia delle Entrate[35], ridotto al ruolo di mero operatore "*data entry*" e di semplice trascrittore delle risultanze dell'AI[36].

5. L'algoritmo "Ve.R.A." e l'applicativo "VE.A.R." quali strumenti di AI a disposizione dell'amministrazione finanziaria italiana

In un'ottica di digitalizzazione del sistema tributario[37], è stato elaborato in via sperimentale un algoritmo, denominato "Ve.R.A." (acronimo di "verifica dei rapporti finanziari"), previsto dalla Legge di Bilancio 2020 (art. 1, commi 682 ss., l. 27 dicembre 2019, n. 160) ed attuato dal MEF con d.m. 28 giugno 2022, a sua volta approvato in via preliminare dall'Autorità garante della *privacy* con la "valutazione di impatto" del 30 luglio 2022, n. 276[38].

Si tratta di uno strumento, oggetto di attenzione anche ad opera della prassi amministrativa[39], in grado di incrociare milioni di dati in contemporanea, pure mediante l'utilizzo di tecniche

[35] M.G. Ortoleva, "L'intelligenza artificiale nell'analisi del rischio fiscale tra auspicata efficienza e tutela dei diritti dei contribuenti", in *Diritto e pratica tributaria internazionale*, (2024), n. 4, pp. 1136 ss.

[36] M. Cedro, F.A. Cimino, *op. cit.*, pp. 64 ss.

[37] E. Marello, "La digitalizzazione del sistema tributario", in *Rivista telematica di diritto tributario*, 16 gennaio 2024, pp. 1 ss.

[38] G. Consolo, "Il nuovo applicativo di analisi del rischio fiscale 'Ve. Ra.' nel processo di digitalizzazione dell'attività dell'Amministrazione finanziaria", in *Il diritto dell'economia*, (2023), n. 3, pp. 153 ss.

[39] Circolare Agenzia delle Entrate, 20 giugno 2022, n. 21/E, in https://www.agenziaentrate.gov.it/portale/documents/20143/4495900/Circola-

di intelligenza artificiale simbiotica, al fine di individuare liste selettive di contribuenti "ad alto rischio" di evasione fiscale, da trasmettere agli uffici territoriali in modo da avviare le ordinarie attività di controllo e rilevare le anomalie da comunicare agli interessati mediante lettere di *compliance*[40].

L'inclusione nelle liste selettive potrà così comportare l'attivazione dei controlli sostanziali, fino ad arrivare alla notifica di un avviso di accertamento o all'invito al contribuente a regolarizzare la propria posizione fiscale mediante la presentazione di una dichiarazione integrativa o un ravvedimento operoso.

Ciascun atto indirizzato al contribuente dovrà, in ogni caso, illustrare "il profilo di rischio fiscale identificato e i dati che sono stati utilizzati per la sua individuazione", formula che sottintende (sia pure timidamente) la necessità di una motivazione adeguata, comprensiva tanto degli elementi informativi utilizzati, del processo di analisi compiuto e della metodologia seguita per individuare la specifica posizione da sottoporre a controllo, quanto di ogni elemento ulteriore di cui si sia avvalso l'ufficio fiscale in sede di controllo, al fine di evitare che l'atto impositivo assurga a mera reiterazione delle risultanze dei sistemi algoritmici[41].

Il pregio dello strumento, che può contribuire a superare il c.d. "paradosso del fisco telematico"[42], successivamente sosti-

re_21_20_06_2022.pdf/5e6b7425-2f02-2b59-480f-c33d9795924e [Data di consultazione: 24/08/2025].

[40] G. Consolo, "Sul trattamento dei dati personali nell'ambito delle nuove procedure automatizzate per il contrasto all'evasione fiscale", in E. Marello, Contrino A. (a cura di), *La digitalizzazione dell'amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, II, Lefebvre Giuffrè, Milano, 2023, pp. 179 ss.; C. Francioso, "Intelligenza artificiale nell'istruttoria tributaria e nuove esigenze di tutela", in *Rassegna tributaria*, (2023), n.1, pp. 47 ss.; D. Conte, "Accertamento tributario e modelli predittivi del rischio di evasione fiscale: il ruolo della IA tra tutela dei dati personali e principi del 'giusto' procedimento", in *Rivista di diritto tributario*, (2024), n. 1, I, pp. 157 ss.; A. Marinello, *op. cit.*, p. 4.

[41] A. Marinello, *op. cit.*, pp. 8-9.

[42] R. Lupi, *L'imposizione tributaria come diritto amministrativo speciale. Spiegazione giuridico sociale del diritto tributario e delle pubbliche entrate*, LGS, San Cesareo (RM), 2023, pp. 208-209; M.G. Ortoleva, *op. cit.*, p. 1148.

tuito dall'applicativo denominato "VE.A.R." (acronimo di "vestizione elenchi con dati archivio rapporti"), risiede nel rendere operativi sistemi di intelligenza artificiale in grado di "disincagliare" ed utilizzare l'enorme mole di dati e informazioni già a disposizione dell'amministrazione finanziaria, attraverso l'utilizzo di tecniche innovative di *risk analysis*[43].

6. Gli ulteriori utilizzi dell'intelligenza artificiale simbiotica nella fase di attuazione del tributo

Le intelligenze artificiali simbiotiche potrebbero trovare ulteriori spazi di manovra nelle ipotesi di liquidazione automatizzata e controllo formale delle dichiarazioni, disciplinate nel nostro ordinamento sia per quanto attiene alle imposte sui redditi che all'Iva: si tratta di controlli di natura accertativa eseguiti da un cervello elettronico programmato sulla base di appositi algoritmi informatici, con cui si procede alla rideterminazione dell'*an* e del *quantum debeatur* rispetto agli importi dichiarati dal contribuente[44].

Tale attività di controllo, volta a correggere gli errori che emergono *ictu oculi* dalla dichiarazione, sfocia direttamente in un'iscrizione a ruolo, preceduta da una comunicazione di irregolarità, che legittima il contribuente a far valere le proprie ragioni ovvero a regolarizzare la propria posizione ed accedere ad una definizione agevolata delle sanzioni[45].

Se già attualmente il controllo automatizzato riesce a fare emergere le irregolarità dotate di un elevato grado di attendibilità, in un prossimo futuro, la sfida legata all'applicazione dei *data-base* condivisi e delle nuove tecnologie simbiotiche è quella

[43] A. Contrino, *Digitalizzazione dell'amministrazione finanziaria e attuazione del rapporto tributario: questioni aperte e ipotesi di lavoro nella prospettiva dei principi generali*, cit., pp. 105 ss.; F. Farri, *Digitalizzazione dell'amministrazione finanziaria e diritti dei contribuenti*, cit., pp. 125 ss.; A. Marinello, *op. cit.*, p. 4.

[44] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., p. 645; S.A. Parente, "Applicazione delle intelligenze artificiali e degli algoritmi informatici alle attività svolte dall'amministrazione finanziaria e alla funzione di ius dicere del giudice tributario: modelli e prospettive", in *Le Corti Umbre*, (2023), n. 1, pp. 295 ss.

[45] A. Zuccarello, "Algoritmi e automatismi nei controlli della dichiarazione: profili problematici", in *Rivista telematica di diritto tributario*, 2 giugno 2022, p. 2.

di escludere le incertezze derivanti da una rappresentazione del presupposto impositivo diversa da quella reale.

Difatti, con l'implementazione delle tecnologie informatiche e digitali potrebbe essere realizzato un sistema nel quale il controllo, anziché essere svolto a posteriori, viene automatizzato sin dall'origine[46].

Occorrerà però verificare se, nell'ambito dei controlli standardizzati, oltre a non sacrificare le regole e i principi a garanzia del contribuente[47], i risultati conseguiti avvalendosi di tali tecnologie siano più puntuali e precisi rispetto a quelli ottenuti in passato con l'utilizzo di strumenti analogici[48].

Un ulteriore impiego dell'intelligenza artificiale simbiotica nell'accertamento tributario potrebbe realizzarsi con riferimento agli indici sintetici di affidabilità fiscale (c.d. ISA), di cui all'art. 9 *bis*, d.l. 24 aprile 2017, n. 50, convertito dalla l. 21 giugno 2017, n. 96, che, in sostituzione degli "studi di settore"[49], consentono di esprimere, su una scala da 1 a 10, il grado di affidabilità fiscale riconosciuto a ciascun contribuente esercente attività di impresa, arti o professioni[50].

Gli ISA, a differenza degli studi di settore, in passato impiegati quali veri e propri strumenti dell'attività di accertamento, mirano a favorire la *compliance* tra fisco e contribuente, avendo la funzione di promuovere l'emersione spontanea delle basi imponibili e stimolare l'assolvimento degli obblighi tributari da parte dei contribuenti e il rafforzamento della collaborazione tra questi e l'amministrazione finanziaria[51].

Ciononostante, l'intelligenza artificiale simbiotica mal si concilia con la fase procedimentale, complessivamente conside-

[46] A. Quattrocchi, "Le potenzialità applicative della blockchain e dei database condivisi nell'attuazione della norma tributaria", in *Rivista telematica di diritto tributario*, 22 novembre 2022, pp. 7 ss.

[47] A. Zuccarello, "Il difficile ruolo del contraddittorio nei controlli automatizzati delle dichiarazioni", in G. Salanitro (a cura di), *op. cit.*, pp. 285 ss.

[48] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., pp. 645-648.

[49] M. Versiglioni, *Prova e studi di settore*, Giuffrè, Milano, 2007, pp. 149 ss.

[50] A. Comelli, "I principi di neutralità fiscale e proporzionalità ai fini della disciplina dell'iva europea e nazionale: dagli studi di settore agli indici sintetici di affidabilità fiscale", in *Diritto e pratica tributaria*, (2019), pp. 1085 ss.

[51] A. Guidara, *op. cit.*, pp. 404 ss.

rata, in ragione della complessità di alcune valutazioni rimesse all'amministrazione finanziaria nel corso dell'attività di accertamento, venendo spesso in rilievo situazioni di difficile apprezzamento, che implicano giudizi complessi, non facilmente delegabili ad una macchina intelligente.

L'impiego dell'intelligenza artificiale simbiotica ben potrebbe riguardare la fase istruttoria, assumendo dimensioni variabili, senza mai sostituirsi alla volontà di provvedere, che connota la fase decisoria del procedimento e pone fine allo stesso: pur volendo in astratto ipotizzare un'istruttoria totalmente algoritmica, che in procedimenti vincolati come quelli di accertamento non è da escludere, occorre comunque una volontà umana che verifichi e faccia propri i risultati dell'istruttoria. Solo in procedimenti contraddistinti da presupposti specifici (contenuti semplici, norme piane, nessi causali dagli esiti univoci) potrebbe ipotizzarsi un'appropriazione preventiva della valutazione algoritmica per il tramite dell'approvazione di regole che presiedono alla sua elaborazione[52].

Anche la dichiarazione precompilata[53], introdotta dal d.lgs. 21 novembre 2014, n. 175, potrebbe costituire un'ulteriore ipotesi di applicazione dell'intelligenza artificiale simbiotica in ambito tributario, trattandosi di una dichiarazione predisposta dall'amministrazione finanziaria, utilizzando le informazioni disponibili nell'Anagrafe tributaria (per la cui disciplina si rinvia al d.p.r. 29 settembre 1973, n. 605) e i dati trasmessi da soggetti terzi (ad esempio, quelli contenuti nelle dichiarazioni e nelle certificazioni rese dai sostituti d'imposta) e resa disponibile telematicamente al contribuente, che può accettarla, apportarvi delle modifiche o presentare un'autonoma dichiarazione[54].

Inoltre, potendo la proposta di concordato preventivo biennale[55] formulata dall'Agenzia delle Entrate essere elaborata me-

[52] *Ivi*, pp. 400 ss.

[53] S. La Rosa, "Dalla dichiarazione 'unica' del contribuente alla dichiarazione 'precompilata' dal fisco e agli accertamenti fiscali 'automatizzati'", in Aa.Vv., *Saggi in ricordo di Augusto Fantozzi*, Pacini Giuridica, Pisa, 2021, pp. 103 ss.

[54] A. Guidara, *op. cit.*, p. 407.

[55] A. Giovannini, "Concordato preventivo: inquadramento sistematico", in A. Giovannini (a cura di), *La riforma fiscale. I diritti e i procedimenti, III, Accertamento, sanzioni rapporti fra processi*, Pacini Giuridica, Pisa, 2024, pp. 43 ss.; A. Quattrocchi, "Il concordato preventivo biennale nell'attuazione della legge delega di riforma del sistema fiscale", in *Rivista di diritto tributa-*

diante "processi decisionali completamente automatizzati"[56], ai sensi dell'art. 9, d.lgs. 12 febbraio 2024, n. 13, non si esclude anche in tale ambito il ricorso a modelli di intelligenza artificiale simbiotica[57].

Peraltro, essendo le attività di accertamento riservate, in via prioritaria, ai contribuenti che non abbiano aderito alla proposta di concordato preventivo biennale ovvero che, pur avendovi aderito, siano decaduti, non appare peregrino ipotizzare che, a seguito della mancata accettazione da parte del contribuente della proposta elaborata con procedure automatizzate, possa operare una presunzione semplice[58]: quest'ultima confluirà poi nella motivazione dell'atto impositivo[59], rendendo così plausibile, anche in assenza di un'espressa previsione normativa, la possibilità di far confluire le risultanze dell'intelligenza artificiale simbiotica nella motivazione dell'avviso di accertamento[60].

Infine, altri utilizzi dei sistemi di intelligenza artificiale simbiotica in materia tributaria potrebbero riguardare, nell'ambito dell'accertamento sintetico dei redditi[61], gli istituti del redditometro[62] e dello spesometro, nonché l'analisi di comparabilità nella disciplina dei prezzi di trasferimento e il "risparmiometro"[63].

rio, (2024), n. 4, pp. 417 ss.; P. Barabino, "Il concordato preventivo biennale, tra accordi e automatismi: una natura ibrida di 'accordo-opzione'", in *Rivista trimestrale di diritto tributario*, (2025), n. 1, pp. 3 ss.

[56] V. Mastroiacovo, "Il concordato preventivo: soggetti, oggetto e procedimento", in A. Giovannini (a cura di), *op. cit.*, pp. 82 ss.

[57] M. Cedro, F.A. Cimino, *op. cit.*, pp. 57-58.

[58] S. Gianoncelli, "Il concordato preventivo: effetti e decadenze", in A. Giovannini (a cura di), *op. cit.*, pp. 100 ss.

[59] R. Lupi, "Motivazione e prova nell'accertamento tributario, con particolare riguardo alle imposte dirette e all'Iva", in *Rivista di diritto finanziario e scienza delle finanze*, (1987), n. 2, I, pp. 274 ss.; A. Marcheselli, "Motivazione e prova, nel procedimento e nel processo tributario. Il giudice tributario come garante della funzione tributaria", in *Rivista telematica di diritto tributario*, 7 gennaio 2025, pp. 1 ss.

[60] M. Cedro, F.A. Cimino, *op. cit.*, pp. 58-59.

[61] G. Selicato, *Il nuovo accertamento sintetico dei redditi*, Cacucci, Bari, 2014, pp. 105 ss.; G. Tinelli, "Accertamento sintetico e tutela del contribuente", in F. Amatucci (a cura di), *L'accertamento sintetico e il nuovo redditometro*, Giappichelli, Torino, 2015, pp. 3 ss.

[62] M. Logozzo, "Redditometro e diritto alla privacy", in F. Amatucci (a cura di), *op. cit.*, pp. 49 ss.

[63] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., p. 623.

Quest'ultima locuzione designa uno strumento predittivo che, tramite un apposito algoritmo verifica, sulla base dei dati forniti dall'Archivio dei rapporti con operatori finanziari, la coerenza dei risparmi accumulati dal contribuente in un determinato periodo d'imposta con i redditi dichiarati nel medesimo arco temporale, in modo da fare emergere eventuali situazioni di anomalia idonee a legittimare ulteriori controlli da parte dell'autorità fiscale.

Ciò potrebbe accadere qualora il contribuente effettui un importante investimento non giustificato dai redditi dichiarati ovvero, pur in presenza di una retribuzione fissa mensile, abbia uscite troppo modeste idonee a rivelare l'esistenza di altre entrate non dichiarate[64].

Si tratta di vicende nelle quali, in conformità al carattere simbiotico dei sistemi di intelligenza artificiale adoperati, riveste un ruolo importante la capacità dell'operatore vivente di confrontare e collegare situazioni e dati in apparenza autonomi, rispetto ai quali l'algoritmo consentirebbe di aumentare notevolmente sia il numero degli elementi oggetto di analisi che la rapidità dell'elaborazione, senza tuttavia alterare in maniera sostanziale il modo di essere del ragionamento giuridico, a tutela dei valori che presiedono la materia tributaria[65].

7. La capacità predittiva dell'amministrazione finanziaria. L'impossibilità di fare assurgere le elaborazioni predittive a presunzione legale di evasione

Il ricorso all'intelligenza artificiale simbiotica, insieme alla capacità di realizzare benefici di natura economica in termini di ottimizzazione dei costi dell'apparato amministrativo, con conseguente risparmio di risorse da destinare ad altri fini, contribuisce a perfezionare la capacità predittiva dell'amministrazione finanziaria, atteggiandosi altresì a parametro indiretto di valutazione tanto dell'efficacia delle attività interne, quanto delle

[64] G. Palumbo, "L'utilizzo dei dati da parte del fisco e delle multinazionali: punti di contatto e differenze", in G. Palumbo, *op. cit*., p. 61.

[65] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., pp. 202-214.

iniziative esterne che incidono (in positivo o in negativo) sulla sfera giuridica e patrimoniale del contribuente.

L'attitudine predittiva si concretizza nella possibilità di formulare anticipazioni e previsioni circa l'andamento delle condotte future dei contribuenti, risultato che gli strumenti di monitoraggio e prevenzione tradizionali verrebbero ad assicurare solo con estrema difficoltà[66].

L'amministrazione finanziaria potrebbe avvalersi di sofisticati *software* "predittivi", come quelli già in uso in vari settori del sistema ordinamentale, per "prevenire" situazioni di frode ed individuare i contribuenti ad alto rischio di condotte evasive o elusive[67].

Le elaborazioni predittive possono, però, essere utilizzate per il contrasto agli illeciti tributari solo come presupposto per l'avvio dei controlli sostanziali, non potendo assurgere a presunzione legale di evasione[68].

Anche con riguardo all'attività espletata dalla polizia tributaria, i sistemi di intelligenza artificiale simbiotica e le tecnologie digitali potrebbero svolgere un'importante funzione predittiva, soprattutto in materia di dogane e monopoli[69], nonché nel settore delle accise[70].

Nel corso degli anni, l'impiego dell'intelligenza artificiale a supporto delle attività di accertamento svolte dall'amministrazione finanziaria è stato sempre più intenso: un'implicita conferma è possibile rinvenirla nell'art. 1, comma 682, l. 27 dicembre

[66] A. Purpura, *op. cit.*, p. 76 ss.

[67] I. Pini, "Intelligenza artificiale e analisi del rischio di evasione fiscale", in *Rivista trimestrale di diritto tributario*, (2025), n. 3, pp. 595 ss.

[68] F. Farri*, Digitalizzazione dell'amministrazione finanziaria e diritti dei contribuenti*, cit., p. 131 ss.

[69] CGUE 9 novembre 2017, causa C-46/16, in https://curia.europa.eu/juris/document/document.jsf?text=&docid=196505&pageIndex=0&doclang=IT&mode=lst&dir=&occ=first&part=1&cid=3586297 [Data di consultazione: 24/08/2025]; CGUE, sez. V, 20 giugno 2019, causa C-1/18, in http://curia.europa.eu/juris/document/document.jsf?text=&docid=215247&pageIndex=0&doclang=it&mode=lst&dir=&occ=first&part=1&cid=1062050 [Data di consultazione: 24/08/2025]; CGUE, sez. V, 9 giugno 2022, causa C-187/21, in https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:62021CA0187 [Data di consultazione: 24/08/2025].

[70] C. Verrigni, *Le accise nel sistema dell'imposizione sui consumi*, Giappichelli, Torino, 2017, pp. 38 ss.; M. Logozzo, "Le accise: inquadramento sistematico e questioni aperte", in *Rivista di diritto tributario*, (2018), n. 2, I, pp. 129 ss.

2019, n. 160, che consente all'Agenzia delle Entrate di avvalersi "delle tecnologie, delle elaborazioni e delle interconnessioni con le altre banche dati di cui dispone, allo scopo di individuare criteri di rischio utili per far emergere posizioni da sottoporre a controllo e incentivare l'adempimento spontaneo"[71].

Pertanto, nell'attesa di pervenire ad un vero e proprio accertamento "algoritmico"[72], si potrebbe pensare di affiancare l'attività di elaborazione dei dati da parte dei sistemi intelligenti a forme di cooperazione preventiva tra fisco e contribuente, in modo da selezionare le situazioni di potenziale rischio[73].

Un primo *screening* svolto con l'ausilio dei sistemi di intelligenza artificiale consentirebbe di selezionare i contribuenti meritevoli di approfondimento. Successivamente, una fase di contraddittorio preventivo permetterebbe all'ufficio fiscale di rendere edotto il contribuente degli esiti della procedura automatizzata, aprendo la strada a procedure di cooperazione tra i soggetti coinvolti: verrebbe così favorita una *disclousure* sui dati non in possesso dell'amministrazione finanziaria funzionale ad integrare gli elementi conoscitivi a propria disposizione e, se del caso, ripetere l'elaborazione algoritmica[74].

[71] G. Palumbo, "Contrasto all'evasione fiscale e impatto sulla privacy dei contribuenti", in G. Palumbo, *op. cit.*, p. 13; F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., p. 643.

[72] G. Ragucci, "L'analisi del rischio di evasione in base ai dati dell'archivio dei rapporti con gli intermediari finanziari: prove generali dell'accertamento 'algoritmico'?", in *Rivista telematica di diritto tributario*, 4 settembre 2019, pp. 1 ss.; M. Fasola, "Le analisi del rischio di evasione tra selezione dei contribuenti da sottoporre a controllo e accertamento 'algoritmico'", in G. Ragucci (a cura di), *op. cit.*, pp. 79 ss.; S. Dorigo, G. Mercuri, "AI e Amministrazione finanziaria in Italia: analisi del rischio fiscale, attuazione del rapporto tributario e profili organizzativi", in *Studi tributari europei*, (2024), n. 14, pp. 57 ss.; M. Fasola, *L'amministrazione algoritmica dei tributi*, Lefebvre Giuffrè, Milano, 2025, pp. 8 ss; M. Pontillo, "Analisi del rischio fiscale e tutela del contribuente tra spazi di trasparenza e prerogative di segretezza", in *Rivista telematica di diritto tributario*, 5 febbraio 2026, pp. 1 ss.

[73] S. Dorigo, "Intelligenza artificiale e norme antiabuso: il ruolo dei sistemi 'intelligenti' tra funzione amministrativa e attività giurisdizionale", in *Rassegna tributaria*, (2019), n. 4, p. 746; M.G. Ortoleva, *op. cit.*, p. 1137.

[74] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., pp. 209-210.

8. Il progetto pilota "*A data driven approach to tax evasion risk analysis in Italy*". I riferimenti all'intelligenza artificiale nella delega di riforma del sistema fiscale (l. 9 agosto 2023, n. 111) e in sede di attuazione

In questo ambito, peculiare importanza rivestono anche i risultati del progetto pilota denominato "*A data driven approach to tax evasion risk analysis in Italy*"[75], che è stato ideato dall'Agenzia delle Entrate e finanziato dall'Unione Europea nel mese di marzo del 2021 al precipuo scopo di innovare i processi di valutazione del rischio fiscale dei contribuenti attraverso il ricorso a strumenti di *network science* (rappresentazione di relazioni non immediatamente evidenti tra società, che possono far trasparire schemi di interposizione, evasione, o elusione fiscale), di analisi visuale delle informazioni (adozione di interfacce innovative "uomo-macchina" con modalità visuali fluide e intuitive di "navigazione" dei dati) e di intelligenza artificiale[76].

Difatti, la rappresentazione di dati sotto forma di reti consente di fare emergere molto più facilmente relazioni indirette e non evidenti tra soggetti che possono essere correlate a schemi di evasione e di elusione fiscale difficilmente individuabili avvalendosi delle tradizionali tecniche di analisi[77].

L'uso di tecniche avanzate di analisi dei dati, fondate sull'utilizzo di sistemi di intelligenza artificiale simbiotica, potrebbe così favorire l'individuazione di fattispecie che, con le normali tecniche di indagine, difficilmente emergerebbero dall'ordinaria attività istruttoria[78].

Di recente, nel corso della XIX legislatura, la delega al Governo per la riforma del sistema fiscale (l. 9 agosto 2023, n. 111), tuttora in fase di attuazione, ha previsto, tra i principi e i criteri direttivi [artt. 2, 4 e 17, comma 1, lett. c) ed f)], la piena interoperabilità delle banche dati e l'impiego combinato di *big data*

[75] C. Verrigni, *op. cit.*, pp. 747-748.

[76] D. Conte, "Digitalizzazione, data protection e tecniche di profilazione nell'attività di accertamento tributario: quali diritti per i contribuenti?", in A. Contrino, E. Marello (a cura di), *op. cit.*, p. 135; A. Marinello, *op. cit.*, p. 3.

[77] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., p. 204, nt. 17.

[78] M. Cedro, F.A. Cimino, *op. cit.*, p. 50.

e intelligenza artificiale al fine di realizzare un duplice obiettivo: favorire la *self-compliance*, anche in un'ottica di riduzione degli oneri burocratici; fornire all'amministrazione finanziaria strumenti più efficaci ed efficienti per la prevenzione e il contrasto degli illeciti tributari, attraverso l'identificazione dei contribuenti a più alto rischio fiscale[79].

L'art. 2, d.lgs. 12 febbraio 2024, n. 13 (Disposizioni in materia di accertamento tributario e di concordato preventivo biennale), nel dare attuazione alla delega, ha indicato una serie di strumenti basati sull'intelligenza artificiale per prevenire e contrastare l'evasione fiscale oltre che per stimolare l'adempimento spontaneo, affidando ad un regolamento del Ministero dell'Economia e delle Finanze la disciplina di dettaglio e l'individuazione di misure adeguate alla tutela dei diritti e delle libertà dei soggetti coinvolti, anche con riferimento al profilo della protezione dei dati personali[80].

In questa prospettiva, il ricorso a sistemi di intelligenza artificiale, nel solco di alcune significative esperienze applicative (già adottate o in via di implementazione)[81], viene prefigurato in vista degli obiettivi generali di prevenzione e di contrasto dell'evasione fiscale, oltre che come strumento idoneo a snellire le procedure di interpello per i contribuenti di minori dimensioni, facilitando altresì lo sviluppo di relazioni collaborative e simbiotiche tra i due soggetti del rapporto d'imposta[82].

Successivamente, la legge 23 settembre 2025, n. 132, recante "Disposizioni e deleghe al Governo in materia di intelligenza artificiale"[83], pur non occupandosi nello specifico dei profili

[79] S Dorigo, "L'Amministrazione finanziaria e l'uso dell'intelligenza artificiale: gli indirizzi della delega fiscale e gli approdi (poco rassicuranti) in sede attuativa", in *Rivista di diritto tributario*, (2024), n. 6, I, pp. 761 ss.

[80] C. Verrigni, *op. cit.*, p. 749; F. Spinello, "'Rivoluzione algoritmica' e accertamento dei tributi: prime riflessioni sull'attuazione della legge delega per la riforma fiscale", in *Rivista telematica di diritto tributario*, 14 ottobre 2025, pp. 1 ss.

[81] F. Farri, *Digitalizzazione dell'amministrazione finanziaria e diritti dei contribuenti*, cit., pp. 118 ss.

[82] A. Marinello, *op. cit.*, pp. 2-3.

[83] G. Salanitro, "Una prima lettura sull'impatto della legge sull'intelligenza artificiale nel diritto tributario, tra consultazione semplificata e analisi del rischio", in *Rivista telematica di diritto tributario*, 9 ottobre 2025, pp. 1

concernenti la materia tributaria, all'art. 14 regola l'applicazione dell'intelligenza artificiale alle attività svolte dall'autorità amministrativa promuovendone l'utilizzo al fine di garantire il buon andamento e l'efficienza delle attività espletate, dando al contempo centralità al principio di autodeterminazione e di responsabilità umana[84]. In particolare, per effetto della previsione normativa, è consentito alle pubbliche amministrazioni (compresa quella tributaria) di utilizzare l'intelligenza artificiale allo scopo di incrementare l'efficienza della propria attività, ridurre i tempi di definizione dei procedimenti e aumentare la qualità e la quantità dei servizi erogati alla collettività, assicurando agli interessati la conoscibilità del suo funzionamento e la tracciabilità del suo operato. L'utilizzo delle macchine intelligenti deve però avvenire in funzione strumentale e di supporto all'attività provvedimentale, nel rispetto dell'autonomia e del potere decisionale della persona umana, unica responsabile dei provvedimenti resi e dei procedimenti espletati.

Ciò è confermato anche dall'art. 3 *bis*, l. 7 agosto 1990, n. 241, dedicato implicitamente all'uso della telematica e, quindi, pure dei sistemi di IA nell'ambito del procedimento amministrativo, che sembra ribadirne la funzione servente rispetto all'attività amministrativa[85].

In questa prospettiva, l'intelligenza artificiale simbiotica, sfruttando la collaborazione uomo-macchina, consentirà di instaurare relazioni reciprocamente vantaggiose, aumentando (e potenziando) le capacità cognitive umane anziché sostituirle.

Le pubbliche amministrazioni dovranno adottare misure tecniche, organizzative e formative preordinate a garantire un utilizzo dell'intelligenza artificiale responsabile, anche al fine di sviluppare capacità trasversali in capo agli utilizzatori.

ss.; A.F. Uricchio, F. Mazzotta, "Metaverso, fiscalità e intelligenza artificiale: prime considerazioni alla luce del recente disegno di legge in materia", in A.F. Uricchio, C. Caldarola (a cura di), *op. cit.*, pp. 225 ss.

[84] G. Salanitro, "Algoritmi, intelligenza artificiale e responsabilità del funzionario e dell'amministrazione in materia tributaria, tra riserva di umanità, robot e discipline dell'Unione europea", in G. Salanitro (a cura di), *op. cit.*, pp. 193 ss.

[85] A. Guidara, *op. cit.*, p. 398.

9. Il contributo della giurisprudenza amministrativa: il requisito di conoscibilità dell'algoritmo

È, dunque, evidente che oggigiorno le funzioni dell'amministrazione finanziaria non possono prescindere dall'utilizzo di sofisticati sistemi di intelligenza artificiale simbiotica in grado di raccogliere, elaborare e processare un'ingente mole di dati ed informazioni relative ai contribuenti[86].

Tali sistemi intelligenti potrebbero essere di ausilio all'amministrazione per verificare, sulla base degli elementi istruttori altrove acquisiti, se una determinata operazione possa essere o meno considerata elusiva o evasiva ponendo le fondamenta per la predisposizione di un sistema di accertamento digitale dei tributi[87]. Si tratta di un utilizzo endoprocedimentale dell'intelligenza artificiale destinato a completare l'attività istruttoria dell'ufficio in vista dell'accertamento.

Qualora invece l'amministrazione finanziaria voglia fondare l'accertamento tributario sui risultati dell'elaborazione algoritmica, emergono non poche criticità con riferimento alla motivazione dell'atto impositivo e ai risvolti probatori in sede processuale.

Ulteriori criticità si pongono ove gli algoritmi informatici, invece di essere adoperati con funzione ancillare all'attività di controllo, siano utilizzati per attivare accertamenti automatizzati fondati su presunzioni relative[88].

[86] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., p. 205; A. Tomo, "La 'forza centripeta' del diritto alla protezione dei dati personali: la Corte di giustizia sulla rilevanza in ambito tributario dei principi di proporzionalità, accountability e minimizzazione", in *Diritto e pratica tributaria internazionale*, (2022), n. 2, pp. 908 ss.; G. Ragucci, "Introduzione e note ordinate sul fisco digitale", in G. Ragucci (a cura di), *op. cit.*, p. 4. In giurisprudenza, cfr. Cass. civ., sez. I, ord. 11 giugno 2018, n. 15075, in *CED Cassazione*; Cass. civ., sez. I, 4 luglio 2018, n. 17485, in *CED Cassazione*.

[87] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., p. 205; A. Guidara, *op. cit.*, p. 408.

[88] C. Verrigni, *op. cit.*, pp. 749 ss.

Anche la giurisprudenza amministrativa[89], pur riconoscendo da tempo che l'impiego di procedure robotizzate sia strumentale a garantire una maggiore efficienza ed economicità dell'attività amministrativa tanto discrezionale, quanto vincolata, per ragioni di trasparenza, imparzialità e pubblicità, ha considerato imprescindibile il requisito della conoscibilità dell'algoritmo[90].

Si tratta degli elementi probatori minimi, che devono sussistere per evitare che venga compromessa la funzione essenziale della motivazione, a prescindere dai vincoli derivanti dalla metodologia di accertamento utilizzata.

Difatti, se gli elementi di prova influiscono sulle garanzie del contribuente, la loro indicazione è intrinsecamente connessa alla motivazione e la distinzione rispetto alla prova in senso stretto assume contorni evanescenti, ragion per cui è necessario valutare di volta in volta se il suo contenuto minimo è idoneo a descrivere compiutamente il fondamento della pretesa impositiva.

Pertanto, il contributo della giurisprudenza amministrativa riveste grande rilevanza per verificare i limiti entro i quali gli strumenti di intelligenza artificiale possano essere utilizzati in ambito tributario: l'analisi muove proprio dal concetto di algoritmo, inteso quale "atto amministrativo informatico" costituito da una sequenza finita di operazioni, consistente in un insieme di calcoli matematici attraverso cui viene fornita ad un cervello elettronico una serie di istruzioni espresse in un linguaggio diverso da quello giuridico, che consente di risolvere determinate tipologie di problemi utilizzando dati noti[91].

[89] Cons. Stato, sez. VI, sent. 8 aprile 2019, n. 2270, in https://www.medialaws.eu/wp-content/uploads/2019/11/Consiglio-di-Stato-sez.-VI-8-aprile-2019-n.-2270.pdf [Data di consultazione: 24/08/2025]; Cons. Stato, sez. VI, sent. 13 dicembre 2019, n. 8472, in https://sites.les.univr.it/cybercrime/wp-content/uploads/2019/06/Consiglio-di-Stato-8472-2019-automazione-procedimento-amministrativo.pdf [Data di consultazione: 24/08/2025]; Cons. Stato, sez. VII, sent. 28 marzo 2024, n. 2953, in *DeJure Giuffrè*.

[90] E. Carloni, "I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo", in *Diritto amministrativo*, (2020), n. 1, pp. 289 ss.

[91] F. Farri, *La giustizia predittiva in materia tributaria*, in *Rivista telematica di diritto tributario*, 12 ottobre 2022, p. 14; F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., pp. 637 ss.

Negli ultimi anni, lo sviluppo delle conoscenze informatiche e di sistemi computazionali sempre più efficienti ha permesso la sperimentazione di applicazioni algoritmiche tanto accurate da essere definite "intelligenza artificiale"[92].

In questa visuale, in ragione del carattere multidisciplinare dell'algoritmo (la cui predisposizione richiede competenze tecniche, informatiche e statistiche, oltre che giuridiche), la "formula tecnica" rappresentata dall'algoritmo deve essere corredata da chiarificazioni che la traducano nella "regola giuridica" alla base della tecnologia digitale sulla quale si fonda l'atto impositivo[93]. Sarà così possibile vagliare la razionalità e rendere il modulo tecnico e i criteri applicativi adoperati per la definizione dell'algoritmo comprensibili tanto per i contribuenti, che possono esercitare il proprio diritto di difesa, quanto per l'organo giudicante, messo nella condizione di sindacare come il potere sia stato concretamente esercitato dall'autorità amministrativa[94].

Difatti, l'opacità e la non conoscibilità dei meccanismi alla base dell'algoritmo, del modulo decisionale adoperato e dei relativi criteri applicativi[95] possono incidere negativamente anche sull'attività espletata dall'organo giudicante, che non verrebbe messo nella condizione di valutare l'attendibilità della ricostruzione posta in essere dall'ufficio fiscale, se non a costo di avvalersi di complesse consulenze tecniche d'ufficio idonee a far lievitare notevolmente i costi della giustizia[96].

In questa prospettiva, un controllo giudiziale agevole e compiuto deve tener conto della razionalità della regola, della sua conformità all'assetto legislativo primario, della selezione e della completezza dei dati che concorrono alla decisione finale,

[92] A. Marinello, *op. cit.*, p. 5.

[93] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., pp. 638-650.

[94] G. Ragucci, *L'analisi del rischio di evasione in base ai dati dell'archivio dei rapporti con gli intermediari finanziari: prove generali dell'accertamento 'algoritmico'?*, cit., p. 1.

[95] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, 2015.

[96] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., p. 207.

unitamente alla rilevanza e al valore loro assegnato in termini assoluti e comparativi[97].

È, dunque, necessario comprendere il percorso alla base del confronto tra le caratteristiche concrete della fattispecie e quelle che, sulla scorta dei dati conosciuti dal sistema informatico, sarebbero dovute sussistere per poter giudicare corretta l'operazione realizzata dal contribuente, anche al fine di consentire che l'esito finale dell'elaborazione algoritmica sia adeguatamente vagliato dall'operatore umano, quale responsabile del procedimento.

Solo in tal modo sarà possibile verificare che i criteri, i presupposti e gli esiti del procedimento siano conformi alle prescrizioni e alle finalità stabilite dalla legge in modo da rendere chiare – e dunque sindacabili – le modalità e le regole alla base dell'atto adottato[98].

10. Le implicazioni del principio di non esclusività della decisione algoritmica

Il requisito di conoscibilità dell'algoritmo, unitamente al tema della non esclusività della decisione algoritmica[99], assume connotazioni anche sul piano del merito in quanto l'impossibilità di comprenderne il funzionamento comporta la subordinazione dell'attività amministrativa alle scelte, anche di valore, effettuate in maniera autonoma, incontrollata e discrezionale da coloro che hanno elaborato l'algoritmo[100].

Si tratta, in sostanza, di conformare l'utilizzo delle procedure automatizzate ad una esigenza fondamentale di trasparenza, da riferire tanto all'autorità amministrativa titolare del potere per il cui esercizio si ricorre alla strumentazione algoritmica, quanto ai soggetti interessati ed incisi dall'esercizio di tale potere[101].

[97] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., p. 638.

[98] A. Marinello, *op. cit.*, pp. 5-6.

[99] G. Gallone, *Riserva di umanità e funzioni amministrative. Indagine sui limiti dell'automazione decisionale tra procedimento e processo*, Wolters Kluwer-Cedam, Milano, 2023, pp. 114 ss.

[100] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., p. 206.

[101] A. Marinello, *op. cit.*, p. 6.

A differenza di quanto accade nel diritto amministrativo, in materia tributaria, essendo l'azione impositiva tendenzialmente vincolata, accanto alla conoscibilità e alla comprensibilità dell'algoritmo, è necessario che la regola giuridica ad esso sottesa discenda da una norma primaria, in ossequio al principio di legalità *ex* art. 23 Cost., per evitare che la ponderazione degli interessi coinvolti venga demandata all'ente impositore anziché al legislatore[102].

In questa visuale, il principio di trasparenza, inteso secondo una declinazione rafforzata[103], lungi dall'assolvere alle sole funzioni proprie dei procedimenti amministrativi in senso stretto, consente di valutare la coerenza e la conformità della regola espressa dall'algoritmo con il compendio normativo di riferimento[104].

In ogni caso, come ribadito dalla giurisprudenza amministrativa[105], sia pure in relazione ad una fattispecie non tributaria, in ossequio al principio (di fonte europea) di non esclusività della decisione algoritmica, non si può prescindere dall'intervento umano nella verifica dei risultati dell'elaborazione algoritmica, non potendo questi ultimi essere utilizzati senza che prima si sia dato corso ad ulteriori riscontri finalizzati ad evitare che uno strumento concepito come modello predittivo del rischio di evasione si possa convertire in un metodo di accertamento, soprattutto ove lo stesso sia fondato su presunzioni[106].

[102] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., pp. 637-638.

[103] M. Cedro, F.A. Cimino, *op. cit.*, pp. 67-68.

[104] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., p. 637.

[105] Tar Lazio, sez. III *bis*, sent. 10 settembre 2018, n. 9224, in https://def.finanze.it/DocTribFrontend/executeAdvancedGiurisprudenzaSearch.do [Data di consultazione: 24/08/2025]; Cons. Stato, sez. VI, sent. 4 febbraio 2020, n. 881, in https://www.iuranovitcuria.it/2020/03/08/consiglio-di-stato-in-sede-giurisdizionale-sezione-sesta-sentenza-del-4-febbraio-2020-nr-881/ [Data di consultazione: 24/08/2025]; Cons. Stato, sez. III, 25 novembre 2021, n. 7891, in *Il Foro Amministrativo*, (2021), n. 11, p. 1721; Cons. Stato, sez. V, 19 luglio 2022, n. 6236, in *Il Foro Amministrativo*, (2022), nn. 7-8, II, p. 945.

[106] G. Ragucci, *L'analisi del rischio di evasione in base ai dati dell'archivio dei rapporti con gli intermediari finanziari: prove generali dell'accer-*

In questa visuale, all'attività svolta con l'ausilio dei sistemi di intelligenza artificiale simbiotica non consegue automaticamente l'emanazione dell'atto impositivo: come confermato anche dal d.m. 15 luglio 2021, tale fase assurge piuttosto ad istruttoria interna, meramente strumentale e ausiliaria, a cui segue un procedimento amministrativo fondato sul rispetto dei principi del contraddittorio[107], della collaborazione e della buona fede tra amministrazione finanziaria e contribuente, nel quale il funzionario persona fisica si assume la responsabilità esterna dell'atto emesso dopo aver posto in essere un'attività cognitiva, acquisitiva e di giudizio[108].

Non può, dunque, prescindersi da un contributo umano capace di controllare, validare ovvero smentire la decisione automatizzata[109], in aderenza al criterio della sorveglianza umana ("*human in the loop*"), desumibile dall'art. 14 dell'*AI Act* europeo, espressione del più generale principio (di derivazione unionale) di non esclusività della decisione algoritmica, soprattutto in presenza di diritti e libertà fondamentali della persona[110].

tamento 'algoritmico'?, cit., pp. 1-2.

[107] G. Ragucci, *Il contraddittorio nei procedimenti tributari*, Giappichelli, Torino, 2013, pp. 5 ss.; S. Sammartino, "Il diritto al contraddittorio endoprocedimentale", in *Rassegna tributaria*, (2016), n. 4, pp. 986 ss.; A. Giovannini, "Il contraddittorio endoprocedimentale", in *Rassegna tributaria*, (2017), n. 1, pp. 13 ss.

[108] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., pp. 207-208; A. Guidara, *op. cit.*, pp. 402 ss.; F. Paparella, "Procedimento tributario, algoritmi e intelligenza artificiale: potenzialità e rischi della rivoluzione digitale", in A. Contrino, E. Marello (a cura di), *op. cit.*, pp. 3 ss.; A. Marinello, *op. cit.*, pp. 6-7.

[109] G. Avanzini, *Decisioni amministrative e algoritmi informatici. Predeterminazione, analisi predittiva e nuove forme di intellegibilità*, Editoriale Scientifica, Napoli, 2018, pp. 35 ss.; P. Otranto, "Riflessioni in tema di decisione amministrativa, intelligenza artificiale e legalità", in U. Ruffolo (a cura di), *XXVI lezioni di Diritto dell'Intelligenza Artificiale*, Giappichelli, Torino, 2021, pp. 251 ss.; A. Di Martino, *Tecnica e potere nell'amministrazione per algoritmi*, Editoriale Scientifica, Napoli, 2023, pp. 240 ss.

[110] A. Simoncini, "L'algoritmo incostituzionale: intelligenza artificiale il futuro delle libertà", in *BioLaw Journal – Rivista di BioDiritto*, (2019), n. 1, pp. 80 ss.

Pertanto, le tecnologie intelligenti, pur potendo essere impiegate nella fase istruttoria, non possono mai assumere un ruolo sostitutivo rispetto alla volontà di provvedere che caratterizza la fase decisoria del procedimento tributario[111].

Dal punto di vista decisionale, l'intelligenza artificiale simbiotica presenta maggiore complessità rispetto al semplice algoritmo informatico in quanto la modifica e il reiterato adattamento della regola favoriscono soluzioni più adeguate al caso concreto, con la conseguente necessità di rendere conoscibile l'ambiente giuridico di riferimento, nonché la regola di base, i suoi elementi caratterizzanti e i criteri di integrazione e/o sostituzione funzionali agli adattamenti successivi, affinché gli stessi possano essere utilizzati in funzione ausiliaria o sostitutiva della decisione umana: il principale criterio di valutazione attiene, così, al piano dell'efficacia, potendosi risolvere in un giudizio di razionalità e di ragionevolezza della decisione algoritmica con inevitabili interferenze con i principi che presiedono l'azione amministrativa[112].

11. Le applicazioni dell'intelligenza artificiale simbiotica in ambito giurisdizionale

In ambito giurisdizionale[113], le intelligenze artificiali simbiotiche e gli algoritmi informatici, unitamente alle tecnologie digitali, possono esplicare una funzione non irrilevante. I *software* di "giustizia predittiva"[114], fondati sull'indicizzazione dei dati e sull'utilizzo di metadati e di sofisticati algoritmi, quali strumenti di analisi sintattica e sistemi di *machine learning*, che, incro-

[111] C. Verrigni, *op. cit.*, p. 743.

[112] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., pp. 629-630.

[113] S.A. Parente, "Le funzioni giurisdizionali e di accertamento tributario nell'epoca della data economy e dei sistemi di cloud computing: l'ausilio di intelligenze artificiali, big data e algoritmi informatici", in *Euro-Balkan Law and Economics Review*, (2023), n. 1, pp. 92 ss.

[114] G. Barone, *Giustizia Predittiva e Certezza del Diritto*, Pacini Giuridica, Pisa, 2024, pp. 47 ss.; V. Mastroiacovo, "A proposito di un percorso interdisciplinare sulla giustizia predittiva", in V. Mastroiacovo (a cura di), *op. cit.*, pp. XV ss.; S.A. Parente, "Predittività delle decisioni, intelligenza artificiale simbiotica e diritto tributario", in *Didaskaloi*, (2025), n. 2, pp. 170 ss.

ciando una gran mole di dati in entrata, ricercano una correlazione tra gli stessi e deducono un modello applicativo[115], pur conformandosi pienamente ai sistemi di *common law*, in cui vige il principio del precedente giurisprudenziale vincolante (c.d. *stare decisis*), possono spiegare una qualche utilità anche negli ordinamenti di *civil law*, ivi compresa la materia tributaria, conservando un autonomo rilievo.

Difatti, nei sistemi giuridici continentali, connotati dallo *ius scriptum*, il precedente giurisprudenziale – sebbene sia privo di efficacia vincolante, in quanto non ricompreso tra le fonti del diritto – potrebbe orientare il ragionamento del giudice, fornendogli argomenti utili a motivare la propria decisione.

In questo contesto, pur potendo il giudice umano avvalersi in funzione ausiliaria dei sistemi di intelligenza artificiale a sostegno dell'attività di *ius dicere* per individuare ed elaborare i principi di diritto, i precedenti giurisprudenziali e le argomentazioni idonee ad orientare il proprio ragionamento e motivare la decisione, sembra del tutto peregrina e remota la possibilità di immaginare un giudice tributario "robot", non essendo le macchine intelligenti tuttora in grado di sostituirsi integralmente all'essere umano al fine di pervenire ad una decisione, in quanto quest'ultima – lungi dall'essere fondata su un mero meccanicismo – spesso richiede un *quid pluris* di sensibilità, che, allo stato attuale, il freddo algoritmo non sembra in grado di assicurare.

Ciò, anche nella materia tributaria, che, quantomeno in astratto, potrebbe prestarsi più delle altre a sviluppare un ragionamento automatizzato, in quanto branca dell'ordinamento rigidamente disciplinata dal diritto positivo nella quale vengono in rilievo elementi numerici, come tali idonei ad essere posti a fondamento di elaborazioni algoritmiche[116].

[115] F. Farri, *La giustizia predittiva in materia tributaria*, cit., p. 3.
[116] *Ivi*, p. 9.

12. I *software* di ricerca dei precedenti giurisprudenziali: l'iniziativa "Prodigit". Le possibili forme di automazione delle decisioni dinanzi al giudice tributario

Invero, però, il vigente sistema tributario (domestico e sovranazionale) è quanto di più frammentario e disarticolato possa immaginarsi rispetto alle esigenze di un *software* di giustizia predittiva, che fonda il proprio funzionamento sulla semplicità dei dati e dei collegamenti: ne sono da esempio i numerosi rinvii legislativi, il celere mutamento della normativa di settore, che spesso richiama disposizioni abrogate o modificate nella numerazione senza che il legislatore abbia emendato le indicazioni contenute nella norma richiamante, oltre alle difficoltà strutturali proprie di un ordinamento giuridico complesso e multilivello[117].

Pertanto, anche i modelli più evoluti di applicazione dell'intelligenza artificiale in campo giurisdizionale, diffusi nei diversi ordinamenti (è il caso dell'Estonia, della Cina, di alcuni stati del Nordamerica e del Brasile), piuttosto che assurgere ad un giudice robot[118], costituiscono, in realtà, meccanismi di ricerca di precedenti giurisprudenziali tra loro assimilabili che, lungi dal sostituire l'attività umana, quanto meno nel processo conclusivo del sillogismo, si limitano a svolgere una mera attività ricognitiva di ausilio per il giudice umano, senza tradursi in un ragionamento giuridico complesso.

Difatti, l'attività giudicante, lungi dal costituire la sintesi di un mero confronto tra dati, secondo un meccanismo generalmente utilizzato dai sistemi di intelligenza artificiale, assurge a qualcosa di ben più ampio[119]. In presenza di un organo giudicante "robot", sarebbe inoltre difficile garantire la piena ricostruibilità e giustificabilità di ogni elemento che ha condotto alla decisione robotica[120].

[117] *Ivi*, pp. 10-11.

[118] M. Luciani, "La decisione giudiziaria robotica", in A. Carleo (a cura di), *Decisione robotica*, il Mulino, Bologna, 2019, pp. 63 ss.

[119] F. Farri, *La giustizia predittiva in materia tributaria*, cit., pp. 5-12.

[120] C. Sacchetto, "Intelligenza Artificiale, Giustizia Predittiva e nuovi confini del Processo Tributario", in F. Montalcini, R. Nemni, C. Sacchetto, *Diritto tributario telematico. Nuovi confini*, Giappichelli, Torino, 2021, pp. 261 ss.; S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., p. 211.

Si è, dunque, lontani dalla prospettiva di una giustizia predittiva in senso proprio, vale a dire di un giudice robot in grado di risolvere le controversie tributarie in sostituzione della persona umana[121]: allo stato attuale, non sembra potersi andare oltre l'ottimizzazione dei *software* di ricerca dei precedenti giurisprudenziali, attraverso il miglioramento degli strumenti della giurimetria, come è avvenuto di recente con l'iniziativa "Prodigit"[122], sperimentata (non senza criticità)[123] dagli organi di vertice della giustizia tributaria anche al fine di creare una banca dati della giurisprudenza tributaria di merito[124].

Eccettuate le innovazioni digitali introdotte nella fase contenziosa dal d.lgs. 30 dicembre 2023, n. 220, in attuazione dell'art. 19, l. n. 111/2023, una forma di automazione delle decisioni dinanzi al giudice tributario potrebbe al più riguardare il vaglio preliminare di ammissibilità del ricorso ovvero la concessione di misure cautelari *inaudita altera parte* o ancora le controversie catastali relative al classamento e all'attribuzione della rendita per le unità immobiliari a destinazione ordinaria: in relazione alle prime, potrebbe essere automatizzato il vaglio di verifica della tempestività del ricorso e del deposito; per le seconde, si potrebbero tipizzare gli elementi in grado di conferire rilievo preponderante al requisito del *periculum in mora* in attesa della valutazione in contraddittorio anche del *fumus boni iuris*; per le

[121] E. della Valle, "Il giudice tributario robot", in A. Marcheselli, E. Marello (a cura di), *op. cit.*, pp. 15 ss.

[122] G. Ingrao, A. Buccisano, "L'intelligenza artificiale e la giustizia predittiva alla luce del progetto Prodigit", in A. Marcheselli, E. Marello (a cura di), *op. cit.*, pp. 70 ss.; E. Marello, "Prodigit come banca dati intelligente", in A. Marcheselli, E. Marello (a cura di), *op. cit.*, pp. 94 ss.; V. Mastroiacovo, "Il progetto Prodigit e il paradigma della comodità", in A. Marcheselli, E. Marello (a cura di), *op. cit.*, pp. 83 ss.; F. Paparella, "La giustizia tributaria digitale: brevi note sul modello e sugli obiettivi perseguiti dal progetto 'Prodigit'", in A. Marcheselli, E. Marello (a cura di), *op. cit.*, pp. 35 ss.

[123] E. Marello, "Prodigit: alcune domande di metodo e qualche semplice proposta", in *Rivista telematica di diritto tributario*, 1 febbraio 2023, pp. 1 ss.

[124] F. Farri, *La giustizia predittiva in materia tributaria*, cit., pp. 9-16. La banca dati della giurisprudenza tributaria di merito, gestita dal Dipartimento della Giustizia Tributaria del MEF, è oggi consultabile al seguente link: https://bancadatigiurisprudenza.giustiziatributaria.gov.it/ricerca [Data di consultazione: 24/08/2025].

controversie catastali[125], potrebbero essere predisposti algoritmi in grado di effettuare una comparazione tra l'unità tipo e quella oggetto di accertamento, trattandosi di operazioni tecniche aventi ad oggetto le caratteristiche strutturali e urbanistiche dei cespiti immobiliari piuttosto che le risultanze di una decisione giuridica[126]. In questo contesto, con il completamento della digitalizzazione e dell'automatizzazione del catasto[127], si verrebbe non soltanto a migliorare l'efficienza di un importante comparto amministrativo, ma anche a ridurre un contenzioso i cui contenuti giuridici risultano spesso di scarso rilievo[128].

13. L'amministrazione della giustizia nella normativa in materia di intelligenza artificiale (l. 23 settembre 2025, n. 132). La valenza residuale dei sistemi di "*Symbiotic AI*" nell'ambito della giurisdizione tributaria

Quanto evidenziato sembra, da ultimo, confermato dalla legge n. 132/2025 in materia di intelligenza artificiale, poc'anzi richiamata[129], che, all'art. 15, con riferimento all'amministrazione della giustizia, consente l'utilizzo dell'IA esclusivamente per finalità strumentali e di supporto ovvero per l'organizzazione e la semplificazione del lavoro giudiziario nonché per la ricerca giurisprudenziale e dottrinale, essendo sempre riservata alla persona umana ogni decisione sull'interpretazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione di ogni provvedimento decisorio.

L'intelligenza artificiale simbiotica avrebbe così una valenza residuale nell'ambito della giurisdizione tributaria, consentendo all'organo giudicante di reperire i supporti argomentativi a sostegno della propria decisione in modo da preservare la funzione

[125] A. Uricchio, "Il contenzioso catastale: la difficile convivenza di più giurisdizioni", in *Rassegna tributaria*, (2005), n. 5, pp. 1404 ss.

[126] A. Salvati, "Intelligenza artificiale, accertamenti e controversie in materia catastale", in *Rivista telematica di diritto tributario*, 25 gennaio 2025, pp. 1 ss.

[127] S.A. Parente, *Il catasto e gli estimi catastali: funzione impositiva e regole di governo*, Cacucci, Bari, 2020, pp. 109 ss.

[128] F. Farri, *La giustizia predittiva in materia tributaria*, cit., pp. 14 ss.

[129] V. *supra* § 8.

del giudice (essere umano) chiamato a adeguare la norma astratta alle caratteristiche peculiari del caso concreto[130].

In tal modo, svolgendo le macchine intelligenti una funzione suppletiva e strumentale, quale mero ausilio tecnico in grado di supportare il ragionamento giuridico della persona umana, non si riscontrano ragioni ostative all'eventuale utilizzo[131].

In questa prospettiva, l'intelligenza artificiale simbiotica, sfruttando la collaborazione uomo-macchina, consentirebbe di instaurare relazioni reciprocamente vantaggiose, aumentando (e potenziando) le capacità cognitive umane anziché sostituirle.

La soluzione di compromesso permette di riconoscere ai sistemi intelligenti di natura simbiotica un ruolo non marginale nell'indirizzare le scelte decisionali, assecondando al contempo l'implementazione di forme di interlocuzione preventiva tra autorità fiscale e contribuente, per configurare una concezione post-moderna del rapporto d'imposta fondata sulla *compliance*[132] e tale da relegare le situazioni di conflitto ad *extrema ratio*[133].

14. Conclusioni

In relazione alle norme tributarie che richiedono una valutazione economica della fattispecie, le intelligenze artificiali simbiotiche potrebbero avere importanti spazi di manovra ai fini dell'elaborazione dei dati economici e matematici idonei a consentire al giudice di ponderare la razionalità dell'operazione rispetto ai normali parametri di condotta del contribuente, in modo da valorizzare, in un'ottica *ad adiuvandum*, la funzione predittiva rispetto all'attività svolta dal giudice umano, sfuggendo ad ogni forma di meccanicismo.

In conclusione, malgrado le notevoli potenzialità in ordine ai possibili impieghi dell'intelligenza artificiale nell'accertamento

[130] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., pp. 212-213.

[131] S. Dorigo, *Intelligenza artificiale e norme antiabuso: il ruolo dei sistemi 'intelligenti' tra funzione amministrativa e attività giurisdizionale*, cit., pp. 740-741.

[132] G. Ragucci, *Gli istituti della collaborazione fiscale. Dai comandi e controlli alla Tax Compliance*, Giappichelli, Torino, 2018, pp. 2 ss.

[133] S. Dorigo, *L'intelligenza artificiale e i suoi usi pratici nel diritto tributario: Amministrazione finanziaria e giudici*, cit., p. 213.

e nel processo tributario, la disciplina normativa sembra ancora piuttosto deficitaria, anche con riferimento al profilo delle garanzie a tutela del contribuente, addivenendosi spesso a risultati di difficile comprensione, soprattutto se comparati a quelli normalmente conseguenti alla fase procedimentale o a quella contenziosa[134].

Affinché le intelligenze artificiali simbiotiche possano costituire un fattore di progresso in ambito tributario, è necessario che venga predisposto un *corpus* legislativo aderente ai principi che presiedono la materia, in modo da garantire un equilibrio tra esercizio autoritativo dei poteri impositivi e diritti e garanzie dei contribuenti, evitando posizioni vantaggiose a beneficio della sola amministrazione finanziaria e scongiurando l'affermazione di un (nuovo) liberismo tecnologico opaco, privo di una base normativa adeguata.

Si tratta di un'evoluzione da incoraggiare e da valutare senza pregiudizi, evitando però di incorrere in eccessi di entusiasmo da innovazione e tecnologia in quanto l'ammodernamento dell'ordinamento giuridico richiede che il legislatore presti la massima attenzione alle garanzie, alle libertà fondamentali e alla trasparenza per scongiurare un arretramento in termini di tutela dei diritti[135].

In presenza di trasformazioni rilevanti c'è il rischio che la tutela dei diritti del contribuente vacilli rispetto all'esercizio di poteri autoritativi da parte dell'amministrazione finanziaria, soprattutto allorquando sia riconosciuto ai sistemi di intelligenza artificiale una funzione decisoria[136].

Perché le innovazioni siano considerate in senso positivo, è necessario che la trasformazione digitale sia supportata da un quadro legislativo adeguato a garantire un bilanciamento degli interessi coinvolti, dovendo l'assetto normativo disciplinare l'utilizzo delle nuove tecnologie e assicurare la trasparenza delle informazioni raccolte[137].

[134] A. Guidara, *op. cit.*, p. 413.

[135] F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, cit., pp. 650-651.

[136] M.G. Ortoleva, *op. cit.*, p. 1140.

[137] C. Verrigni, *op. cit.*, p. 753.

Lorenzo Pulito

Il principio di proporzionalità algoritmica

Abstract: **The Principle of Algorithmic Proportionality.** The principle of proportionality has gained increasing importance in the field of criminal justice. The present paper explores the role that the principle might play in a scenario in which algorithmic tools are used in the various phases of criminal proceedings. If applied, proportionality can prevent *an excess of algorithmic power* due both to the exercise of power concentrated in the hands of the public prosecutor, abnormally amplified by the ability to use a new and extremely powerful investigative arsenal, and to the exercise of judicial decision-making power of the judge, whose cognitive performance is enhanced by the machine's capabilities.

Keywords: Artificial Intelligence, Proportionality, Criminal Justice.

Sommario: 1. La proporzionalità "codificata" nell'*AI Act*: una panoramica preliminare – 2. Il ruolo del principio di proporzionalità nella giustizia penale – 3. La proporzionalità statica e verticale dell'*AI Act* – 4. L'approccio dinamico e orizzontale – 5. Canone di proporzione e legislazione interna – 6. Proporzionalità, decisione penale, "controllo umano significativo" – 7. "Controllo pubblico significativo" e proporzionalità *by design* – 8. Decisioni sproporzionate e rimediabilità *ex post* – 9. Mezzo e fine contro l'eccesso di potere algoritmico.

1. La proporzionalità "codificata" nell'*AI Act*: una panoramica preliminare

Il termine "proporzionalità" trova numerose ricorrenze nel testo dell'*AI Act*[1], sin dai primi *considerando*. In primo luogo,

[1] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), pubblicato sulla Gazzetta ufficiale dell'Unione europea il 12 luglio 2024. A commento dell'imponente testo si rinvia alle opere di G. Cassano, E.M. Tripodi (a cura di), *Il Regolamento Europeo sull'Intelligenza Artificiale. Commento al Reg. UE n. 1689/2024*, Maggioli, Santarcangelo

identifica la finalità stessa del regolamento, che è quella "di introdurre un insieme proporzionato ed efficace di regole vincolanti per i sistemi di IA"[2] sulla base dell'approccio basato sul rischio[3].

Il principio di proporzionalità interviene in funzione selettiva delle pratiche vietate: così è, ad esempio, per le pratiche di *social scoring*, inibendosi quelle suscettibili di determinare un trattamento pregiudizievole di persone fisiche o di interi gruppi, che risulti ingiustificato o 'sproporzionato' rispetto alla gravità del loro comportamento sociale.

Il parametro della proporzionalità influisce anche sulle modalità di sorveglianza umana[4]: le persone fisiche cui è affidata possono optare tra diverse condotte alternative (non usare il sistema o arrestarlo), dovendosi determinare per quella che risulti più proporzionata rispetto alle circostanze del caso. È poi criterio per calibrare le misure che i fornitori sono tenuti ad adottare per mitigare i rischi, adempiendo agli obblighi previsti dal regolamento: questo richiede che tali misure siano efficaci e proporzionate[5] rispetto a quanto necessario per conseguire gli obiettivi dell'*AI Act*[6]. Le implicazioni del canone refluiscono,

di Romagna, 2024; A. Mantelero, G Resta, G.M. Riccio (a cura di), *Intelligenza artificiale commentario*, Wolters Kluwer, Milano, 2025; O. Pollicino, F. Donati, G. Finocchiaro, F. Paolucci (a cura di), *La disciplina dell'intelligenza artificiale*, Giuffrè Francis Lefebvre, Milano, 2025; G. Taddei Elmi, A. Contaldo (a cura di), *Intelligenza artificiale. AI Act-Regolamento (UE) 1689/2024. Il nuovo scenario giuridico europeo*, Pacini Giuridica, Pisa, 2024.

[2] Cfr. il ventiseiesimo considerando del Regolamento (UE) 2024/1689.

[3] Cfr. il ventisettesimo considerando del Regolamento (UE) 2024/1689.

[4] L'art. 14, par. 1, del Regolamento (UE) 2024/1689 impone un "obbligo di sorveglianza umana", fin dalla fase di progettazione e sviluppo dei sistemi, da realizzare "anche con strumenti di interfaccia uomo-macchina adeguati", in modo tale che essi possano "essere efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso".

[5] Cfr. il sessantaquattresimo considerando del Regolamento (UE) 2024/1689.

[6] Vale a dire, così come recita l'art. 1, par. 1, del Regolamento (UE) 2024/1689, "migliorare il funzionamento del mercato interno e promuovere la diffusione di un'intelligenza artificiale antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la prote-

altresì, su aspetti di carattere economico: si prevede l'istituzione di un sistema di gestione della qualità semplificato per tutelare le imprese di dimensioni più ridotte, così salvaguardando "la proporzionalità rispetto ai costi dell'innovazione"[7]. Ancora, ai criteri di proporzionalità e dissuasività deve essere informato l'apparato sanzionatorio applicabile dagli Stati membri nel caso di inosservanza delle disposizioni del regolamento[8].

Già da questa sintetica panoramica si arguisce il carattere flessibile della proporzionalità, principio che trova espresso riconoscimento, in diverse accezioni, nelle più alte fonti europee e costituisce un principio generale del diritto unionale[9]. Dopo aver ricostruito il significato e il ruolo che la proporzionalità ha assunto in ambito processuale penale, il contributo indaga le plausibili proiezioni e le dimensioni del canone in uno scenario caratterizzato dall'impiego di sistemi di intelligenza artificiale in tale dominio.

2. Il ruolo del principio di proporzionalità nella giustizia penale

La proporzionalità ha radici lontane. Nella riflessione filosofica era legata all'idea di giustizia: il giusto è un che di proporzionale; a sua volta, l'idea di giustizia come proporzione rimandava al bilanciamento come metodo per la risoluzione dei conflitti tra gli interessi confliggenti[10].

zione dell'ambiente, contro gli effetti nocivi dei sistemi di IA nell'Unione, e promuovendo l'innovazione".

[7] Leggasi il centoquarantaseiesimo considerando del Regolamento (UE) 2024/1689.

[8] Vedasi il centosessantosseimo considerando del Regolamento (UE) 2024/1689.

[9] E. Herlin-Karnell, "General principles and EU criminal law", in K.S. Ziegler, P.J. Neeuvonen, V. Moreno-Lax (eds.), *Research Handbook on General Principles in EU Law*, Edward Elgar, Cheltenham, 2022, p. 526; J. Kokott, C. Sobotta, "The Evolution of the Principle of Proportionality in EU Law—Towards an Anticipative Understanding?", in S. Vogenauer, S. Weatherill (eds.), *General Principles of Law. European and Comparative Perspectives*, Hart, Oxford, 2017, p. 167.

[10] Per una sintetica panoramica cfr. M. Karliuk, "Proportionality principle for the ethics of artificial intelligence", in *AI and Ethics*, 3 (2023), n. 3, p. 986 s.

Il principio di proporzionalità, oggi conosciuto e applicato negli ordinamenti giuridici di tutto il mondo[11], rappresenta, in

[11] A. Stone Sweet, J. Mathews, *Proportionality Balancing and Constitutional Governance: A Comparative and Global Approach*, Oxford University Press, Oxford, 2019, p. 3.

Sviluppatosi soprattutto nel contesto sovranazionale, il principio di proporzionalità è implicitamente ricavabile dall'art. 8, par. 2, Cedu (cfr. tra le tante, Corte edu, 4 luglio 2023, *Glukhin c. Russia*, in *Proc. pen. giust.*, 2024, n. 2, p. 413 ss., con nota di O. Bruno, "La condanna per manifestazione pacifica (non preavvisata e) con riconoscimento facciale viola i diritti fondamentali"; Corte edu, 13 settembre 2018, *Big Brother et. al. c. Regno Unito*; Corte edu, 30 maggio 2017, *Trabajio Rueda c. Spagna*; Corte edu, Grande Camera, 4 dicembre 2008, *S. & Marper c. Regno Unito*).

È oggetto di espresso riconoscimento in ambito eurounitario, essendo sancito dall'art. 52, par. 1, della Carta dei diritti fondamentali dell'UE, il quale ammette restrizioni delle libertà individuali nel "rispetto del principio di proporzionalità", solo "laddove siano necessarie e rispondano effettivamente a finalità di interesse generale riconosciute dall'Unione o all'esigenza di proteggere i diritti e le libertà altrui". L'importanza del canone nel contesto dell'UE è stata condivisibilmente rimarcata in quanto la proporzionalità "represents a common principle between Member States and its flexibility allows proportionality to be adapted to any concrete situation both in terms of applicable law and factual issues": in questi termini, L. Scomparin, A. Cabiale, "The Principle of Proportionality in Directive 2014/41/EU – Challenges of the Present and Opportunities for the Future", in K. Ambos, A. Heinze, P. Rackow, M. Šepec (eds.), *The European Investigation Order*, Duncker & Humblot GmbH, Berlino, 2023, p. 240.

La proporzionalità si è diffusa anche nel conteso interno essendo, sia pure in maniera indiretta, implicata dalla presunzione di non colpevolezza (art. 27, comma 2, Cost.). A riguardo, senza pretese di esaustività, cfr. M. Caianiello, "Il principio di proporzionalità nel procedimento penale", in *Dir. pen. cont., Riv. trim.*, (2014), n. 3-4, pp. 143 ss.; D. Negri, Fumus commissi delicti. *La prova per le fattispecie cautelari*, Giappichelli, Torino, 2004, pp. 16 ss.; G. Tabasco, *Principio di proporzionalità e misure cautelari*, Wolters Kluwer-Cedam, Milano, 2017, pp. 49 ss. Sulle implicazioni con i canoni propri del giusto processo, cfr. A. Camon, "La prova genetica tra prassi investigative e regole processuali", in *Proc pen. giust.*, (2015), n. 6, p. 167; M. Daniele, *Habeas Corpus. Manipolazioni di una garanzia*, Giappichelli, Torino, 2017, p. 23; F. Zacchè, *Criterio di necessità e misure cautelari personali*, Giuffrè, Milano, 2018, p. 58. Con specifico riferimento agli influssi del principio di proporzionalità nelle indagini preliminari, cfr. L. Belvini, *Principio di proporzionalità e attività investigativa*, Edizioni Scientifiche Italiane, Napoli, 2022; C. Bonzano, *Gli accertamenti medici coattivi*, Wolters Kluwer-Cedam, Milano, 2017, pp. 114 ss.; F. Nicolicchia, *I controlli occulti*

generale, un "criterio di razionalità pratica utile per misurare il grado di comprimibilità di un diritto fondamentale"[12]. Opera in funzione di difesa dell'individuo di fronte agli eccessi del potere legislativo e previene l'arbitrio dei pubblici poteri, limitando allo stretto indispensabile il sacrificio dei diritti personali[13]: da questo angolo visuale, la proporzionalità gioca un ruolo determinante nel procedimento penale, dove è perenne la tensione tra l'esigenza di reprimere i fenomeni criminosi e le garanzie dell'accusato e dei terzi coinvolti a vario titolo nella vicenda giudiziaria.

Rivolta al legislatore, lo vincola alla costruzione delle fattispecie processuali[14], dovendo bilanciare, per ogni atto suscettibile di comprimere i diritti fondamentali, i contrapposti interessi in gioco, "in modo tale da non consentire né la prevalenza assoluta di uno dei valori coinvolti, né il sacrificio totale di alcuno di loro"[15].

Un diritto fondamentale può essere limitato per conseguire un fine legittimo (come prevenire o reprimere un reato), ma secondo canoni metodologici idonei a contenerne la sua compres-

e continuativi come categoria probatoria, Wolters Kluwer-Cedam, Milano, 2020, pp. 107 ss.; R. Orlandi, "La riforma del processo penale fra correzioni strutturali e tutela "progressiva" dei diritti fondamentali", in *Riv. it. dir. proc. pen.*, (2014), n. 3, p. 1160; D. Negri, "Compressione dei diritti di libertà e principio di proporzionalità davanti alle sfide del processo penale contemporaneo", in *Riv. it. dir. proc. pen.*, (2020), n. 1, pp. 3 ss.; G. Ubertis, "Equità e proporzionalità versus legalità processuale: eterogenesi dei fini?", in *Arch. pen.*, (2017), n. 2, pp. 389 ss.

[12] R. Orlandi, "Uso poliziesco dell'intelligenza artificiale. L'insegnamento del Bundesverfassungsgericht", in *Cass. pen*., (2023), n. 7/8, p. 2171. L'A. sottolinea come il principio di proporzionalità non vada confuso con il principio di ragionevolezza, il quale in realtà ha altra origine e diverso contenuto concettuale, essendo corollario del principio d'eguaglianza.

[13] R. Orlandi, "Postulati del processo penale contemporaneo tra principi 'naturali' e concezioni normative", in *Riv. it. dir. proc. pen*., (2024), n. 2, p. 483, nota 48.

[14] F. Cassibba, "'Trasfigurazione' delle indagini preliminari, principio di proporzionalità e controllo giurisdizionale effettivo", in *Arch. pen*., (2024), n. 1, p. 213.

[15] Corte cost., sent. 23 marzo 2018, n. 58, in *Giur. cost*., 2018, n. 2, pp. 592 ss., con osservazioni di D. Pulitanò, "Una nuova 'sentenza ILVA'. Continuità o svolta?".

sione: più il reato da prevenire o perseguire è grave, più se ne giustifica la lesione del diritto.

Da questa prospettiva, la proporzionalità è espressione del principio di legalità[16], integrando la portata del principio di determinatezza, in forza del quale la legge che regola il compimento di atti suscettibili di comportare una lesione grave a uno dei diritti fondamentali coinvolti deve essere dotata di particolare precisione. Sicché, nel disciplinare l'uso di mezzi o misure lesive di diritti, il legislatore deve tener conto sia del tipo di reato da prevenire o reprimere, sia del *quantum* di limitazione del diritto da consentire.

Oltre alla dimensione processuale, il principio di proporzionalità si è espanso in quella sostanziale, in cui, per effetto della giurisprudenza evolutiva della Corte costituzionale[17], ampio è stato il suo contributo alla costruzione del diritto individuale a non subire pene sproporzionate rispetto alla gravità oggettiva e soggettiva del fatto commesso[18].

[16] Sancito dall'art. 52, par. 1, della Carta dei diritti fondamentali dell'UE e, per la sua dimensione processuale, prescritto dall'art. 111, comma 1, Cost.

[17] Corte cost., sent. 10 novembre 2016, n. 236, in *Cass. pen*., (2017), p. 508 ss., con nota di E. Aprile. A commento, v. anche E. Dolcini, "Pene edittali, principio di proporzione, funzione rieducativa della pena: la Corte costituzionale ridetermina la pena per l'alterazione di Stato", in *Riv. it. dir. proc. pen*., (2016), n. 4, p. 1956 ss.; D. Pulitanò, "La misura della pena, fra discrezionalità politica e vincoli costituzionali", in *Dir. pen. cont., Riv. trim*., (2017), n. 2, pp. 48 ss.; F. Viganò, "Un'importante pronuncia della Consulta sulla proporzionalità della pena", in *Dir. pen. cont., Riv. trim*., (2017), n. 2, pp. 61 ss.

[18] Se ancora ci si interroga su quale sia il senso di punire, convergenza si registra sul fatto che lo si debba fare in maniera proporzionata: cfr. V. Chiao, "Predicting Proportionality: The Case for Algorithmic Sentencing", in *Criminal Justice Ethics*, 37 (2018), n. 3, p. 254, secondo cui "we are not sure what we are trying to do by punishing criminals, but whatever it is, we should only do it by means that are proportionate." Per la proporzionalità della pena, vedasi, per tutti, N. Recchia, *Il principio di proporzionalità nel diritto penale. Scelte di criminalizzazione e ingerenza nei diritti fondamentali*, Giappichelli, Torino, 2020; F. Viganò, *La proporzionalità della pena. Profili di diritto penale e costituzionale*, Giappichelli, Torino, 2021. Per la proporzionalità della risposta sanzionatoria, anche oltre la mera sanzione penale, cfr., tra i tanti, S. Tesoriero, "Processo penale e prova multidisciplinare europea in materia di illeciti finanziari", in *Riv. dir. proc*., (2016), n. 6, pp. 1526 ss.; A.F. Tripodi,

La proporzionalità non è confinata al solo momento legislativo[19], quale parametro valido ad orientare le scelte del nomoteta in tema di diritti inviolabili, ma abbraccia pragmaticamente la dimensione giudiziale e la specifica vicenda di volta in volta trattata[20], essendo contestualmente il criterio che governa il momento applicativo e di successivo controllo degli strumenti restrittivi, permettendo di valutarne la congruità rispetto all'obiettivo da raggiungere nel caso concreto.

Da questa angolazione, rappresenta anche "una regola esegetico-applicativa per l'organo giudiziario", volta ad orientarne le decisioni tutte le volte in cui interviene a comprimere i diritti fondamentali o ad esercitare il controllo sulle misure restrittive degli stessi[21], spesso in funzione "riequilibra[trice di] assetti che si reputino sbilanciati a sfavore dagli interessi autoritativi"[22].

Il principio, ampiamente conosciuto dalla giurisprudenza della Corte di Strasburgo[23], è stato significativamente elabora-

"Cumuli punitivi, *ne bis in idem* e proporzionalità", in *Riv. it. dir. proc. pen*., (2017), n. 3, pp. 1047 ss.

[19] G. Ubertis, *Sistema di procedura penale*, I, *Principi generali*, 6ª ed., Giuffrè Francis Lefebvre, Milano, 2025, p. 214, evidenzia come il principio costituisce per i diritti fondamentali sia un limite alle scelte del legislatore, sia un criterio che orienta le opzioni dell'organo giudiziario, come lo scegliere il tipo e il modo esecutivo concernenti un'acquisizione probatoria o una misura cautelare.

[20] N. Galantini, "'Diritti cautelari' e restrizione della libertà personale nella logica della proporzionalità", in *Sist. pen*., 30 giugno 2022.

[21] G. Ubertis, "Prova penale e proporzionalità", in *Sist. pen*., 23 gennaio 2025, p. 3.

[22] D. Negri, "Compressione dei diritti di libertà e principio di proporzionalità", cit., p. 11. Come evidenziato da M. Caianiello, *op. cit*., p. 159, il metodo imposto dal principio opera in tre momenti diversi: per il legislatore; per l'interprete, *a priori*; per l'interprete, *a posteriori*.

[23] Corte edu, Grande Camera, 15 novembre 2016, *A e B c. Norvegia*, in *Dir. pen. cont*., 18 novembre 2016, con nota di F. Viganò, "La Grande Camera della Corte di Strasburgo su *ne bis in idem* e doppio binario sanzionatorio". Si vedano a commento, tra i tanti, F. Cassibba, "*Ne bis in idem* e procedimenti paralleli", in *Riv. it. dir. proc. pen*., (2017), n. 1, pp. 351 ss.; G. De Amicis, P. Gaeta, "Il confine di sabbia: la Corte edu ancora di fronte al divieto di *ne bis in idem*", in *Cass. pen*., (2017), n. 2, pp. 469 ss.; P. Fimiani, "*Market Abuse* e doppio binario sanzionatorio dopo la sentenza della Corte e.d.u., Grande Camera, 15 novembre 2016, A e B c. Norvegia", in *Dir. pen. cont*., (2017), n. 2,

to dalla Corte costituzionale tedesca[24], che ha ispirato un test tripartito volto a delineare una vera e propria "procedura argomentativa"[25] funzionale a garantire la razionalità dell'iter logico-giuridico sotteso al bilanciamento tra interessi contrapposti[26].

I passaggi in cui si articola il test di proporzionalità ruotano intorno alla triade *idoneità (geeignetheit)*, *necessarietà (enforderlichkeit)*, *proporzione in senso stretto (verhältnismässigkeit im engeren sinne)*.

Il primo step valutativo risponde al quesito se il fine che si intende perseguire sia meritevole di tutela: si esige l'idoneità astratta della limitazione del diritto tutelato al perseguimento della finalità legittima avuta di mira dalla legge. La sua funzione essenziale è garantire la giustificazione prima di limitare la portata di un certo diritto.

Il secondo riguarda lo strumento impiegato, dovendosi privilegiare, tra più mezzi ugualmente idonei a soddisfare il fine legittimo, quello che comporta il minor sacrificio per il diritto da comprimere. L'interrogativo da soddisfare, dunque, è se manchino altre misure ugualmente idonee a realizzare quel fine, ma che comportino una lesione meno grave per il diritto fondamentale. Contestualizzando, *la necessità* di un particolare sistema di intelligenza artificiale dipenderà dalle altre alternative disponibili (ad esempio, dall'esistenza di altri metodi non basati sull'apprendimento automatico che potrebbero essere utilizzati nell'ambito della giustizia penale, fornendo risultati più affidabili[27]; dalla disponibilità di altri algoritmi che integrano criteri

pp. 5 ss.; R.A. Ruggiero, "*Il ne bis in idem*: un principio in cerca di un centro di gravità permanente", in *Cass. pen*., (2017), n. 10, pp. 3837 ss.

[24] Su cui A. Stone Sweet, J. Mathews, "Proportionality Balancing and Global Constitutionalism", in *Columbia Journal of Transnational Law*, 47 (2008), n. 1, pp. 73 ss. Si veda, altresì, G. Lübbe-Wolff, "The Principle of Proportionality in the Case-Law of the German Federal Constitutional Court", in *Human Rights Law Journal*, 34 (2014), pp. 12 ss.

[25] Così G. Pino, "Proporzionalità, diritti, democrazia", in *Diritto e Società*, (2014), n. 3, p. 601.

[26] In questi termini L. Belvini, *Principio di proporzionalità*, cit., p. 23.

[27] L'interrogativo è sollevato da D. Restrepo Amariles, P.M. Baquero, "Promises and limits of law for a human-centric artificial intelligence", in *Computer Law & Security Review*, 48 (2023), p. 9.

di equità o facilitano il monitoraggio e la conformità necessari a garantire la non discriminazione[28]).

Se, adottando la misura per perseguire le finalità meritevoli, i vantaggi ottenibili siano superiori agli svantaggi che ne deriverebbero per il diritto fondamentale è, infine, la domanda a cui il terzo passaggio deve offrire risposta. La proporzionalità in *senso stretto* del sistema di intelligenza artificiale dipenderà dal fatto che il risultato complessivo raggiungibile con il ricorso all'algoritmo sia migliore rispetto al quadro ottenibile prescindendone: il tempo e i costi risparmiati devono valere le sfide generate dal sistema automatizzato. In breve, si richiede una relazione articolabile tra mezzi e fini e, in particolare, che i mezzi adoperati siano adatti o appropriati e non più restrittivi del necessario per raggiungere un fine legittimo. Anche se l'atto persegue uno scopo legittimo e non va oltre quanto necessario per raggiungere tale scopo, potrebbe imporre un onere eccessivo all'individuo ovvero risultare eccessivamente afflittivo rispetto agli esiti attesi o addirittura pregiudicare l'essenza stessa del diritto leso: talune limitazioni non si giustificano quando si tratta di perseguire reati considerati poco gravi o quando la pena attesa è di lieve entità[29].

La proporzionalità, fungendo da "guida strutturale"[30], può svolgere un ruolo importante dal punto di vista dell'accettabilità giuridica dell'intelligenza artificiale[31]. In particolare, il canone può giocare un ruolo fondamentale nei processi decisionali soggetti a un grado, anche minimo, di automazione. Inoltre, nei casi in cui il processo decisionale non sia alterato da *output* automatizzati, il parametro può evitare che l'impiego dei *software* provochi un'intrusione nella vita privata e produca una lesione dei diritti fondamentali[32].

[28] *Ibidem*.

[29] R. Orlandi, "Uso poliziesco dell'intelligenza artificiale", cit., p. 2172.

[30] K. Möller, "'Balancing as reasoning' and the problems of legally unaided adjudication: a rejoinder to Francisco Urbina", in *International Journal of Constitutional Law*, 12 (2014), n. 1, p. 222.

[31] In questo senso, M. Karliuk, "Proportionality principle for the ethics of artificial intelligence", cit., p. 987.

[32] Per una più ampia panoramica sui *deficit* di tutela dei diritti fondamentali nell'*AI Act*, si rinvia a E. Cirone, *L'*AI Act *e l'obiettivo (mancato?)*

Questioni di proporzionalità possono porsi in fasi cronologicamente distinte (legislativa, sistemica, decisionale ed *ex post*), che, per il loro differente potenziale impatto sulle prospettive di implementazione giuridica del processo decisionale automatizzato, richiedono di essere singolarmente prese in esame[33].

3. La proporzionalità statica e verticale dell'*AI Act*

È il legislatore a compiere per primo le valutazioni di proporzionalità, effettuandole in via generale e astratta, nonché in maniera proattiva, nel tentativo di prevenire *ex ante* l'adozione di decisioni sproporzionate. Sarebbe illusorio credere che la proporzionalità possa conseguirsi solo attraverso la sua gestione preventiva a livello *puramente* legislativo[34], soprattutto quando le previsioni normative siano prive di reale tassatività, come spesso accade per quelle unionali. Non fa eccezione il regolamento europeo sull'intelligenza artificiale, frutto di compromissorie valutazioni di proporzionalità, caratterizzate dalla staticità dell'approccio adoperato. Infatti, "nonostante rivendichi di essere informato al compromesso tra l'interesse allo sviluppo economico e la protezione dei diritti fondamentali", l'*AI Act* "predetermina il giudizio di proporzionalità che risolve l'interferenza tra valori"[35]. Metterne a fuoco alcune trame normative aiuta a comprendere il significato di tale affermazione.

L'*AI Act* introduce un quadro normativo basato sul rischio per i sistemi intelligenti, classificandoli in quattro livelli di rischio:

di promuovere uno standard globale per la tutela dei diritti fondamentali, in *Quad. AISDUE*, (2024), n. 2, pp. 51 ss.

[33] Per questa scansione cronologica della proporzionalità, cfr. L. Enqvist, M. Naarttijärvi, "Discretion, Automation, and Proportionality", in M. Suksi (ed.), *The Rule of Law and Automated Decision-Making*, Springer, Cham, 2023, p. 161.

[34] Nel senso di ritenere insufficiente la valutazione *ex ante* dei rischi che coinvolgano i diritti fondamentali, se non si raffronta con l'uso giudiziale dei sistemi, cfr. P. Marra, "Rischi algoritmici. Alcune questioni sulla digitalizzazione della giustizia", in *Journal of Ethics and Legal Technologies*, 7 (2025), n. 1, p. 149.

[35] M.C. Canato, "Verso il superamento del "legal risk" europeo: intelligenza artificiale e approccio proporzionale al rischio", in *Leg. pen*., (2024), n. 3, p. 23.

inaccettabile, elevato, limitato e minimo. Maggiore è il rischio e più incisivi sono gli obblighi che i fornitori sono tenuti a rispettare per mitigarlo[36].

Incidendo sulla dignità e sulla libertà morale, sono vietati (ma non in maniera "assoluta") i sistemi che sfruttano le tecniche manipolative subliminali e le vulnerabilità delle persone[37], i sistemi di *social scoring*[38] e di *sentiment analysis*[39]. Sono riconducibili al livello di rischio inaccettabile, a certe condizioni, i sistemi di intelligenza artificiale "predittiva": sono vietati quelli che consentono di "valutare o prevedere il rischio che una persona fisica commetta un reato", quando la predizione sia basata unicamente sulla profilazione. In ragione della invasività dei rastrellamenti generici di dati sensibili, quali quelli biometrici, sono vietate le pratiche di *scraping* non mirato di immagini facciali da internet o da filmati di telecamere a circuito chiuso per creare o ampliare

[36] Sui requisiti di conformità e monitoraggio cui sono assoggettati i sistemi classificati ad alto rischio, M. Cigno, A.S. Scimè, "La *compliance* per l'IA ad alto rischio (qualità conformità, ecc.)", in G. Cassano, E.M. Tripodi, *Il regolamento europeo sull'intelligenza artificiale*, cit., p. 574 s. In particolare, i fornitori dovranno garantire che i sistemi ad alto rischio siano conformi ai requisiti previsti dall'*AI Act* (artt. 8-16) e, pertanto, dovranno: prevedere un sistema di gestione dei rischi (art. 9); assicurare la qualità e la governance dei dati (art. 10); redigere e mettere a disposizione delle autorità e degli organismi notificati la documentazione tecnica, volta a dimostrare che il sistema di IA ad alto rischio è conforme ai requisiti (art. 11); consentire e conservare la registrazione automatica degli eventi ("log") per la durata del ciclo di vita del sistema e garantire un livello di tracciabilità del funzionamento del sistema adeguato alla finalità prevista (art. 12); progettare e sviluppare sistemi trasparenti (onde consentire che gli *output* generati dall'artefatto possano essere compresi e correttamente interpretati dai deployer, a cui si dovranno fornire "istruzioni per l'uso" concise, complete, corrette, chiare, pertinenti, accessibili e comprensibili) (art. 13), efficacemente sorvegliabili dall'umano durante il periodo in cui sono in uso (art. 14), nonché accurati, robusti e cybersicuri (art. 15).

[37] Purché provochino o possano ragionevolmente provocare "un danno significativo".

[38] Se determinano un trattamento pregiudizievole che sia ingiustificato o sproporzionato rispetto al comportamento sociale di determinate persone o di gruppi di persone o alla sua gravità.

[39] Quando effettuata sui luoghi di lavoro e di istruzione, ma non quando la valutazione delle emozioni è effettuata "per motivi medici o di sicurezza".

le banche dati di riconoscimento facciale e di categorizzazione biometrica.

Onde evitare di alimentare fenomeni di sorveglianza di massa e i conseguenti effetti *chilling* sulle libertà individuali, il regolamento vieta i sistemi di identificazione biometrica che operano da remoto in modalità *real time* in luoghi pubblici. Ampie le deroghe al divieto, modulate sul criterio di proporzionalità. La pratica non incorre nel divieto nella misura in cui sia "strettamente necessario" per conseguire determinate finalità[40], la cui obiettiva esistenza va apprezzata *ex ante* dell'autorità giudiziaria[41]. Il controllo che questa è tenuta a svolgere è saldamente improntato al rispetto del criterio della stretta necessità e del canone del minor sacrificio, bandendosi ogni uso meramente esplorativo e privo del carattere della indispensabilità. Infatti, da un lato, è precluso l'utilizzo non esclusivamente confermativo dell'"identità della persona interessata"; d'altro lato, si deve accertare "sulla base di prove oggettive o indicazioni chiare", che l'impiego delle tecnologie di identificazione biometrica è "necessario e proporzionato" al conseguimento degli obiettivi consentiti. Quale corollario della logica di contenimento della pratica nei limiti dello stretto indispensabile e della ricerca mirata di soggetti determinati, discende la necessaria delimitazione spazio-temporale della possibilità di utilizzo del *facial ricognition* in tempo reale, che dovrà

[40] Le finalità, in base a quanto previsto dall'art. 5, lett. h), del Regolamento (UE) 2024/1689, sono le seguenti: "i) la ricerca mirata di specifiche vittime di sottrazione, tratta di esseri umani o sfruttamento sessuale di esseri umani, nonché la ricerca di persone scomparse; ii) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di una minaccia reale e attuale o reale e prevedibile di un attacco terroristico; iii) la localizzazione o l'identificazione di una persona sospettata di aver commesso un reato, ai fini dello svolgimento di un'indagine penale, o dell'esercizio di un'azione penale o dell'esecuzione", relativamente a una serie circoscritta di crimini, punibili "con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno quattro anni".

[41] Attraverso una previa autorizzazione rilasciata dall'autorità giudiziaria o quantomeno, nei casi di "urgenza debitamente giustificata", una convalida postuma, la cui mancanza è sanzionata con l'immediata interruzione dell'attività ricognitiva e la cancellazione dei risultati algoritmici indebitamente acquisiti.

avvenire in astratto, in attuazione della riserva di legge istituita a riguardo dal regolamento[42], nonché in concreto, col provvedimento autorizzativo da emanarsi da parte dell'autorità preposta.

Invece, sono classificati "ad alto rischio" i dispositivi di identificazione biometrica *post remote*: impostati secondo questa modalità non sono atti a sorvegliare in massa, dal momento che operano su dati biometrici già presenti in una banca dati; tuttavia, potendo il loro risultato influenzare le decisioni, magari incidenti sulla libertà personale, richiedono adeguati presidi garantistici. Il loro utilizzo, orientato alla "ricerca mirata di una persona" nell'ambito di un'indagine in corso, è consentito solo previa autorizzazione dell'autorità giudiziaria (o convalida entro quarantotto ore) e nei limiti di "quanto strettamente necessario per le indagini su uno specifico reato"[43].

La disposizione che, in generale, si occupa dei sistemi ad alto rischio è l'art. 6 del regolamento, che li suddivide in due categorie. La prima, descritta nel paragrafo 1, si rivolge ai sistemi ricompresi nell'ambito applicativo di altre norme europee di armonizzazione, che ne subordinano la circolazione ad una preventiva valutazione di conformità da parte di terzi (cfr. Allegato I). La seconda, di cui al paragrafo 2, individua i sistemi ad alto rischio "indipendenti", i quali, sebbene privi delle condizioni di cui all'Allegato I, rientrano nella categoria in forza del "rischio

[42] Questo impone che gli Stati membri interessati a consentire l'uso di sistemi di identificazione biometrica remota "in tempo reale" stabiliscano "nel proprio diritto nazionale le necessarie regole dettagliate per la richiesta, il rilascio, l'esercizio delle autorizzazioni di cui al paragrafo 3, nonché per le attività di controllo e comunicazione ad esse relative": così l'art. 5, par. 5, del Regolamento (UE) 2024/1689.

[43] Solo quando il sistema di intelligenza artificiale è "utilizzato per l'identificazione iniziale di un potenziale sospetto sulla base di fatti oggettivi e direttamente connessi al reato" l'art. 26, par. 10, dell'*AI Act* consente di prescindere dall'autorizzazione giudiziaria. In ogni caso, quando l'autorizzazione è negata, l'uso del sistema di identificazione biometrica va immediatamente interrotto e i dati personali vanno cancellati. L'esigenza di utilizzare siffatti dispositivi solo per la ricerca mirata è ulteriormente specificata dalla disposizione che ne proibisce l'uso in assenza di nessi "con un reato, un procedimento penale, una minaccia reale e attuale o reale e prevedibile di un reato o la ricerca di una determinata persona scomparsa" (art. 26, par. 10, del Regolamento (UE) 2024/1689).

significativo di danno per la salute umana, la sicurezza o i diritti fondamentali delle persone fisiche" (cfr. Allegato III)[44].

Tra i settori individuati nell'Allegato III, oltre ai sistemi biometrici, rilevanti per la giustizia penale sono quelli impiegati per attività di *law enforcement*[45] (n. 6 dell'Allegato), nonché i sistemi dedicati all'"amministrazione della giustizia"[46] (n. 8 dell'Allegato). Il settore delle "attività di contrasto" raggruppa sistemi adoperabili dalla magistratura o dagli organi di *law enforcement* per scopi eterogenei: con finalità predittive, per determinare "il rischio per una persona fisica di diventare vittima di reati" o, viceversa, "il rischio di commissione del reato o di recidiva"; come "poligrafi e strumenti analoghi" per vagliare l'attendibilità delle dichiarazioni; allo scopo di "valutare l'affidabilità degli elementi probatori nel corso delle indagini o del perseguimento di reati"; per effettuare "profilazioni personali" utili per la repressione dei fenomeni criminosi[47]. Quanto al settore dell'"amministrazione della giustizia", infine, si specifica che l'uso dell'intelligenza artificiale può intervenire solo come ausilio del magistrato nell'attività di ricerca, interpretazione "dei fatti e del diritto" e nell'"applicazione della legge a una serie concreta di fatti"[48].

[44] L'efficace sintesi è di F. Ferretti, "La relazione di cura e il consenso informato nell'era della *Medical Artificial Intelligence*", in *Corti supreme e salute*, (2025), n. 1, p. 104.

[45] Nelle attività di contrasto rientrano quelle svolte "a fini di prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia contro le minacce alla sicurezza pubblica e la prevenzione delle stesse": art. 3, n. 36, del Regolamento (UE) 2024/1689. Per un quadro complessivo, con alcuni spunti critici, cfr. A. Sachoulidou, "Harnessing AI for law enforcement: Solutions and boundaries from the forthcoming AI Act", in *New Journal of European Criminal Law*, (2024), n. 2, pp. 117 ss.

[46] In questo "settore" rientrano i "sistemi di IA destinati a essere usati da un'autorità giudiziaria o per suo conto per assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti, o a essere utilizzati in modo analogo nella risoluzione alternativa delle controversie".

[47] Per il quadro di sintesi, v. M. Torre, "Il Regolamento europeo sull'intelligenza artificiale: i profili processuali", in *Proc. pen. giust*., (2024), n. 6, p. 1548.

[48] Cfr. Allegato III, punto 8, lett. a), del Regolamento (UE) 2024/1689.

Come può evincersi dal quadro sinteticamente tracciato, i quattro livelli di rischio sono parametrati, in sostanza, agli ambiti operativi che possono mettere in pericolo i diritti fondamentali e, pertanto, sono aprioristicamente identificati[49].

Quest'approccio risulta in linea con la visione dell'intelligenza artificiale come prodotto, per la quale la preoccupazione che conta e che pesa è la sua sicurezza: trascurate sono, invece, le interazioni dinamiche tra le fonti di pericolo e i profili di vulnerabilità dei soggetti che ne sono esposti.

In altri termini, la valutazione del rischio sottesa al regolamento europeo non riflette un sufficiente grado di granularità: quello considerato è un mero "legal risk", che esprime il potenziale danno che l'utilizzo di un sistema in un certo ambito potrebbe arrecare ad un diritto fondamentale. Il bilanciamento tra i diritti in gioco appare legato ad un giudizio di proporzionalità verticalmente predeterminato, nel senso che i sistemi sono categorizzati in base a fattori di rischio isolati.

4. L'approccio dinamico e orizzontale

Per queste ragioni si è proposto di approcciare alle quattro categorie di rischio secondo una logica orizzontale, applicando un modello basato su scenari, in modo che lo stesso sistema possa essere flessibilmente categorizzato come inaccettabile, ad alto rischio, a rischio limitato o a rischio minimo, tenuto conto delle interazioni tra i vari fattori di rischio presenti in situazioni reali. L'approccio, definibile come "semi-quantitativo", si articola in due distinte fasi: la costruzione degli scenari di rischio e la valutazione quantitativa basata sulla proporzionalità.

La costruzione dello scenario di rischio include i seguenti fattori determinanti: *hazard* (H), *exposure* (E), *vulnerability* (V)[50],

[49] C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, L. Floridi, "AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act", in *Digital Society*, 3 (2024), pp. 11 ss.

[50] In termini generali, il pericolo si riferisce alle fonti di potenziali effetti avversi sugli elementi esposti; l'esposizione si riferisce all'inventario degli elementi all'interno dell'intervallo della fonte di pericolo; la vulnerabilità si riferisce all'insieme di attributi o circostanze che rendono gli elementi esposti suscettibili agli effetti avversi quando hanno un impatto sulla fonte di peri-

response (R), *extrinsic risks*. Una volta tratteggiato lo scenario di rischio, il secondo passaggio consiste nel valutare, mediante un test di proporzionalità (basato sulla formula del peso di Alexy)[51], se la categorizzazione prevista dal regolamento risulti appropriata – vuoi perché i valori in gioco sono correttamente bilanciati, vuoi perché essa riflette la reale esposizione ai rischi emergente dallo scenario concreto – oppure occorra rivederla.

Il ricorso a questa metodologia consentirebbe di tenere più adeguatamente conto delle specificità della giustizia penale come "scenario" d'uso delle macchine algoritmiche, e potrebbe rivelarsi utile per diverse finalità.

Si pensi, a titolo esemplificativo, ai *risk assessment tools* per valutare il pericolo di recidiva.

I pericoli (H) che derivano dal loro uso sono molteplici: riguardano l'opacità, l'affidabilità dei risultati, la qualità dei dati di addestramento, la validità della teoria scientifica su cui sono costruiti. Sommati tra loro, questi fattori accrescono notevolmente l'entità del rischio (E): ad esempio, l'opacità non consente di corregge i dati distorti, "manovra" che, invece, può essere attualizzata se il sistema è trasparente. Per determinare l'esposizione complessiva si dovrebbe tenere conto sia dei diritti fondamentali esposti a rischio (diritto di difesa, principio di uguaglianza), come anche dei vantaggi conseguibili grazie all'impego della macchina (prevedibilità, certezza del diritto, sicurezza ed efficienza), opportunamente bilanciando gli uni con gli altri: ad esempio, l'obiettività delle decisioni potrebbe contrastare con il principio di individualizzazione della pena.

Non trascurabile è l'interazione tra i fattori di rischio e i fattori di vulnerabilità (V), quali l'etnia, le condizioni economiche e il grado di istruzione, che rendono gli individui più sensibili agli effetti negativi della valutazione automatica della recidiva e ai *bias* dei sistemi di previsione.

Lo scenario deve tenere conto sia delle misure di mitigazione (R) concretamente esistenti per contrastare i rischi associati alle

colo; la risposta riguarda le misure esistenti che contrastano o mitigano il rischio.

[51] R. Alexy, *A theory of constitutional rights*, Oxford University Press, Oxford, 2002; Id., "On balancing and subsumption. A structural comparison", in *Ratio Juris*, 16 (2003), n. 4, pp. 433 ss.

valutazioni automatizzate del tasso di recidiva (come standard per la qualità e la raccolta dei dati, la trasparenza, la supervisione umana, l'esclusione di predittori della pericolosità sociale direttamente o indirettamente collegati al *background* etnico e sociale[52]), sia dei rischi estrinseci e accessori rispetto a quello principale.

Inoltre, lo scenario può mutare e disvelare notevoli gradi di complicazione a seconda del segmento processuale in cui lo strumento artificiale viene utilizzato. Se si ipotizza l'impiego dei *risk assessment tool* come ausilio alla valutazione del pericolo di reiterazione del reato, *ex* art. 274, lett. c), c.p.p., esigenza, tra quelle contemplate nel sistema cautelare codicistico, divenuta egemonica nella prassi[53], i pericoli (H) sopra richiamati hanno certamente un impatto maggiore (E), dovendosi considerare esposta a rischio la presunzione di innocenza, baluardo fondamentale anche rispetto ad alcuni vantaggi conseguibili dall'automatizzazione del processo decisionale in questo specifico ambito, quali la minimizzazione del ricorso alla cautela e la riduzione della sfera di discrezionalità del giudice[54].

Il caso d'uso qui adombrato impone di tenere in considerazione le misure "compensative" in campo (R), da valutare con

[52] G. van Dijck, "Predicting recidivism risk meets AI act", in *European Journal on Criminal Policy and Research*, 28 (2022), n. 3, pp. 407 ss.

[53] E. Guido, "Intelligenza artificiale e procedimento penale: ragionando di valutazione del rischio de libertate", in *Arch. pen. (web)*, (2023), n. 1, p. 10.

[54] Gli interventi legislativi che hanno tentato di circoscrivere il perimetro operativo della disciplina di cui all'art. 274, lett. c), c.p.p., non sono riusciti a rafforzare le garanzie volte a contenere la discrezionalità del giudice. Tra essi, merita di essere ricordato quello portato dall'art. 2, l. 16 aprile 2015, n. 47, con cui il legislatore ha stabilito che il pericolo di reiterazione deve essere non solo attuale ma anche concreto, e ha vietato che lo stesso sia desunto dalla sola gravità del titolo di reato per cui si procede. Sul tema, A. Ciavola, "La valutazione delle esigenze cautelari", in L. Giuliani (a cura di), *La riforma delle misure cautelari personali*, Giappichelli, Torino, 2015, pp. 59 ss.; M.F. Cortesi, "Interventi sulle misure custodiali", in A. Diddi, R.M. Geraci (a cura di), *Misure cautelari ad personam in un triennio di riforme*, Giappichelli, Torino, 2015, pp. 20 ss.; C. Iasevoli, "Concretezza ed attualità delle esigenze cautelari", in T. Bene (a cura di), *Il rinnovamento delle misure cautelari*, Giappichelli, Torino, 2015, pp. 19 ss.; E. Turco, "La riforma delle misure cautelari", in *Proc. pen. giust*., (2015), n. 5, p. 107.

un'analisi di tipo quantitativo (che guarda al numero di strumenti di mitigazione) e qualitativo (che si concentra sulla loro effettiva capacità di arginare i pericoli).

L'intervento del giudice, cui spetterà coniugare la valutazione dell'*output* algoritmico con quella dei "comportamenti" o degli altri "atti concreti" che il disposto normativo (art. 274, lett. c) c.p.p.) indica come elementi indefettibili del vaglio prognostico sulla personalità dell'indagato, potrebbe non bastare a garantire *tout court* l'autenticità del suo giudizio umano, considerata la difficoltà di discostarsi dalla previsione della macchina, specialmente nel caso in cui l'*ouptut* propenda per la sussistenza dell'esigenza cautelare. Già in uno scenario "tradizionale" la previsione umana è tendenzialmente sfavorevole per il prevenuto, in quanto risulterà più facile per il giudice scegliere l'opzione che lo pone facilmente al riparo dal rischio di smentita e, quindi, affermare semplicisticamente la sussistenza del pericolo di reiterazione[55]. Né l'effetto distorsivo, esacerbato dal suggerimento della macchina, sarebbe agevolmente compensabile agendo sul piano della motivazione, quand'anche rafforzata. Occorrerebbero soluzioni "compensative" ulteriori, come quella d'impiegare il *tool* solo per confutare la conclusione del giudice che abbia ravvisato un rischio di reiterazione elevato; in altri termini, si potrebbe fare in modo che l'autorità giudiziaria, per le determinazioni di competenza, valuti l'*output* algoritmico quando questo attenua, anziché aggravare, la posizione della persona soggetta a valutazione[56].

[55] O. Mazza, "Distopia del processo artificiale", in *Arch. pen. (web)*, (2025), n. 1, p. 17, dubita del fatto che il giudice, già refrattario attualmente a respingere la proposta di decisione proveniente dai *law clerks*, sarà capace in futuro di opporsi al responso di una macchina addestrata e di assumersi "la responsabilità di disattendere i giudizi prognostici di pericolosità elaborati, in materia cautelare o di misure di prevenzione, da algoritmi predittivi che tengono conto di una serie di variabili nemmeno calcolabili dall'uomo".

[56] Si v. il "modello ristretto" proposto da J. Ryberg, "Artificial intelligence and criminal justice: How to use algorithmic sentencing support in real life (and ethically non-ideal) penal systems?", in *AI and Ethics*, 5 (2025), n. 3, pp. 3255 ss. In particolare, il "Restricted Application Model" prevede quattro fasi che illustrano come un giudice dovrebbe avvalersi dell'assistenza algoritmica: 1) il giudice determina quella che ritiene la pena appropriata per l'imputato in un dato caso; 2) questa pena iniziale determinata dal giudice

In definitiva, lo scenario d'uso ipotizzato potrebbe portare a concludere che il rischio sia in concreto inaccettabile, benché i *risk assessment tool* in oggetto siano classificati dall'*AI Act* come sistemi ad alto rischio, e che di conseguenza sia preferibile "non contaminare il segmento cautelare" con strumenti di valutazione algoritmica del rischio[57].

L'esempio fornito permette ora di apprezzare meglio i potenziali usi della metodologia. *i)* Innanzi tutto, può fornire un parametro per l'adeguamento del sistema interno di gestione del rischio, come delineato nell'art. 9 dell'*AI Act*, migliorando l'identificazione delle misure idonee. *ii)* Ancora, può offrire uno strumento per apprezzare la significatività del rischio: si allude alla disposizione di cui all'art. 6, par. 3, del Regolamento (UE) 2024/1689, in base alla quale, in difetto di un pericolo "significativo" per la salute, la sicurezza o i diritti fondamentali, anche nel senso di "non influenzare materialmente il risultato del processo decisionale", si potrebbe ritenere non ad alto rischio un sistema intelligente, seppur così originariamente classificato secondo il regolamento. Questa norma non offre alcun criterio per misurare se il rischio è o non è significativo o per stabilire il grado di influenza della decisione da parte del sistema[58], né indica l'autorità preposta al compimento di tale valutazione. *iii)* Il

viene quindi inserita nell'algoritmo di determinazione della pena insieme a tutte le informazioni pertinenti specifiche del caso; 3) sulla base delle informazioni sul reato in questione, l'algoritmo fornisce una raccomandazione di pena basata sulle condanne di precedenti casi simili (tuttavia, cosa fondamentale, l'algoritmo fornisce la propria raccomandazione solamente se la pena iniziale suggerita dal giudice è più severa di quella stabilita dall'algoritmo; in caso contrario, l'algoritmo sarà progettato per confermare la raccomandazione formulata dal giudice; 4) sulla base delle informazioni sul reato, della sua proposta di pena iniziale e della raccomandazione dell'algoritmo, il giudice determina definitivamente la pena da infliggere al trasgressore.

[57] L. Belvini, *Intelligenza artificiale e circuito investigativo*, Cacucci, Bari, 2025, p. 245.

[58] Anche gli ulteriori limiti dettati dalla previsione non offrono indicazioni precise: ad esempio, non è chiaro quando sia possibile ritenere limitato il compito procedurale eseguito dall'algoritmo al punto tale da ritenerlo idoneo a escludere l'"alto rischio". Parimenti, non si comprende quando il sistema di intelligenza artificiale serve solo per migliorare il risultato di un'attività umana precedentemente completata, o, ancora, quando è adoperato in chiave "preparatoria" per una delle finalità indicate nell'allegato III al Regolamento.

bilanciamento basato sulla proporzionalità, secondo il modello semiquantitativo, potrebbe orientare i decisori politici e, tra questi, soprattutto il legislatore nazionale, in sede di implementazione dell'*AI Act*, nello stabilire se, quando e come un sistema di intelligenza artificiale possa trovare ingresso nel procedimento penale, contribuendo ad assicurare una più adeguata risposta alle questioni concernenti i profili processuali riguardanti l'impiego dell'intelligenza artificiale nell'ambito penale.

5. Canone di proporzione e legislazione interna

L'evidenziata carenza di "granularità" dell'*AI Act* rafforza l'esigenza di un intervento legislativo a livello nazionale volto a delineare con chiarezza e precisione confini, condizioni e limiti d'uso dell'intelligenza artificiale nel dominio della giustizia penale. Tale bisogno è stato avvertito dal legislatore italiano: la l. 23 settembre 2025, n. 132, recante "Disposizioni e deleghe al Governo in materia di intelligenza artificiale", tra i principi generali richiama proprio quello di proporzionalità, richiedendo che a questo parametro si conformino la ricerca, la sperimentazione, lo sviluppo, l'adozione, l'applicazione e l'utilizzo di sistemi e di modelli di intelligenza artificiale[59].

Un esplicito riferimento al principio di proporzionalità si rinviene poi nell'ambito dei criteri dettati per l'esercizio delle deleghe legislative dirette al recepimento dell'*AI Act*, prevenendosi che il canone dovrà essere rispettato dal Governo nell'emanazione dei decreti legislativi "per adeguare e specificare la disciplina dei casi di realizzazione e di impiego illeciti di sistemi di intelligenza artificiale"[60]. Il criterio dettato dal legislatore delegante, tuttavia, menziona espressamente la sola fase delle "indagini preliminari": eppure, occupandosi del dominio giustizia, il testo – che riserva sempre "al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sulla adozione dei provvedimenti"[61] – non pare

[59] Art. 3, comma 1, l. n. 132/2025.

[60] Art. 24, commi 3 e 5, lett. e), l. n. 132/2025.

[61] Art. 15, comma 1, l. n. 132/2025. Il comma 2, invece, riserva al Ministero della giustizia la disciplina gli impieghi dei sistemi di intelligenza

escludere che i *software* intelligenti possano avere influenza, più in generale, sull'attività giurisdizionale.

Sarà indispensabile salvaguardare il principio di proporzionalità, date anche le sue implicazioni sul terreno della legalità probatoria, richiedendosi il rispetto del criterio in tutte le fasi del procedimento penale in cui si dovesse ricorrere all'uso di intelligenza artificiale, non solo in quella investigativa, chiarendosi l'ambito dei poteri del giudice nella gestione della macchina, le condizioni e i divieti circa l'acquisizione[62] e la valutazione degli elementi probatori generati dai sistemi intelligenti[63], i criteri di giudizio e gli obblighi motivazionali, nonché i controlli giurisdizionali esperibili per accertare, tra le altre cose, il rispetto della proporzionalità.

Possono indicarsi alcune coordinate generali, espressive del principio di proporzionalità[64], che il legislatore delegato dovrebbe seguire. Quanto ai sistemi di intelligenza artificiale adoperabili in fase d'indagine, se idonei ad incidere sulla sfera delle libertà fondamentali, se ne dovrà consentire l'utilizzo solo quando le alternative meno impattanti sulla sfera del singolo non siano altrettanto efficaci. I presupposti d'impiego dovrebbero implicare che il loro utilizzo sia circoscritto all'accertamento dei reati più gravi, secondo il criterio di adeguatezza, e sia ammesso a condizione che vi sia un significativo compendio indiziario, ido-

artificiale per l'organizzazione dei servizi relativi alla giustizia, per la semplificazione del lavoro giudiziario e per le attività amministrative accessorie.

[62] In proposito, A. Scalfati, "IA e processo penale: prospettive d'impiego e livelli di rischio", in *Proc. pen. giust*., 2024, n. 6, p. 1409. Sulle aperture in tema di "inutilizzabilità eurounitaria", L. Parodi, "La 'gravità dell'ingerenza' nel prisma della proporzionalità: nuovi equilibri in tema di *data retention*", in *Sist. pen*., 7 marzo 2025, pp. 20 ss.

[63] Aspetto sottolineato da G. Canzio, "AI ACT e processo penale: sfide e opportunità", in *Sist. pen.*, 14 ottobre 2024, p. 11.

[64] L. Parodi, "La 'gravità dell'ingerenza' nel prisma della proporzionalità", cit., p. 25, ritiene che una disciplina calibrata sul livello di intrusività dei mezzi tecnologici adoperati per finalità investigative e ispirata al principio di proporzionalità, anche in concreto, sia quella maggiormente idonea a garantire uno standard adeguato di protezione dei diritti fondamentali degli individui, senza sacrificare le esigenze di sicurezza e di accertamento dei reati.

neo a giustificare la compressione della libertà dell'individuo[65]. Il controllo sulla proporzionalità dell'intervento investigativo, assecondando la tendenza espressa dalle Corti europee, andrebbe demandato, di regola, ad un giudice, equidistante delle parti e neutrale rispetto al reato da accertare.

Indicazioni più specifiche potranno enuclearsi a seconda del grado di invasività della pratica algoritmica.

Nel caso dei *tools* di *predictive policing*, sebbene di regola destinati ad essere impiegati per finalità preventive, non potranno trascurarsi né gli influssi negativi sul diritto all'autodeterminazione informativa, né pericolosi travasi di conoscenza dal contesto preventivo a quello investigativo.

Per quanto concerne i rapporti tra la proporzionalità e il giudizio, nel rinviare alle considerazioni formulate nel successivo paragrafo, merita di essere subito evidenziato che l'uso di strumenti algoritmici non dovrebbe mai rappresentare un obbligo per il giudice[66]. Infatti, la valutazione di proporzionalità fatta a riguardo dal legislatore europeo è nel senso di destinare i sistemi di intelligenza artificiale "ad assistere" l'autorità giudiziaria, e se "assistere significa aiutare sempre che l'aiutato lo voglia, il giudice allora non è obbligato a ricorrere al supporto automatico"[67].

6. Proporzionalità, decisione penale, "controllo umano significativo"

Il canone della proporzionalità, oltre che al legislatore, si rivolge al giudice, cui compete l'equilibrato bilanciamento fra i diversi interessi coinvolti nel caso concreto. Compiuta dall'u-

[65] Sull'indispensabilità che sia il legislatore a fissare, in astratto, i reati per il cui accertamento è consentito utilizzare gli strumenti investigativi più invasivi, e sulla necessità di una congrua piattaforma indiziaria idonea a rendere accettabile la compressione delle libertà individuali, cfr. L. Belvini, *Principio di proporzionalità e attività investigativa*, cit., in particolare, pp. 138 ss. e 150 ss.

[66] T. Sroka, "Artificial intelligence and the right to a fair trial", in M. Balcerzak, J. Kapelańska-Pręgowska (eds.), *Artificial Intelligence and International Human Rights Law. Developing Standards for a Changing World*, Edward Elgar Publishing, Cheltenham, 2024, p. 268.

[67] G. De Minico, "Giustizia e intelligenza artificiale: un equilibrio mutevole", in *Rivista AIC*, (2024), n. 2, p. 106.

mano, la valutazione di proporzionalità ha natura di giudizio di valore: implica la comprensione del significato della norma, la sua interpretazione, la necessaria attenzione ai dettagli del caso concreto e la considerazione del contesto sociale e culturale[68].

Questa sensibilità giuridica, tipica dell'essere umano, è una proprietà che l'intelligenza artificiale, almeno allo stato attuale dell'arte, non possiede, e ciò la rende incapace di cogliere, nella sua pienezza, un concetto come quello di proporzionalità[69].

I sistemi algoritmici si basano su dati del passato, assegnano importanza a schemi decisionali ripetuti nel tempo e attribuiscono rilievo ad un criterio di stampo statico-seriale, a discapito di uno di tipo dinamico-evolutivo[70]. I modelli statistici, rivolti al passato, guardano ad una moltitudine di casi per decidere le sorti (future) di un singolo o di pochi. La proporzionalità, invece, fa del singolo uno strumento di comprensione delle implicazioni per i più; le valutazioni che il canone richiede non sono ancorate al passato o imprigionate nei numeri, ma intrise di umanità[71].

Considerati gli attuali limiti tecnologici, che non consentono di ritenere raggiunto un livello generale di "proporzionalità automatica", l'ingresso dell'intelligenza artificiale nel processo decisionale non può prescindere dalla circostanza che gli esseri

[68] Spunti in tal senso in T. Sroka, "Artificial intelligence and the right to a fair trial", cit., p. 260.

[69] Con riferimento alle valutazioni di proporzionalità in ambito bellico, T.K. Woodcock, "Human/Machine(-Learning) Interactions, Human Agency and the International Humanitarian Law Proportionality Standard", in *Global Society*, 38 (2023), n. 1, pp. 108 ss.

[70] C. Cavaceppi, "L'intelligenza artificiale applicata al diritto penale", in G. Taddei Elmi, A. Contaldo (a cura di), *Intelligenza artificiale-Algoritmi giuridici:* Ius condendum o fantadiritto?, Pacini Giuridica, Pisa, 2020, pp. 113 s. In argomento, G. De Martino, P. Marra, A. D'Aversa, L. Pulito, A. Pellicani, G. Pio, M. Ceci, "Leveraging textual content, citational aspects and dissenting opinions through a multi-view contrastive learning methodology for legal precedent analysis", in *Computer Law & Security Review*, 60 (2026), pp. 1 ss.

[71] G. Fiorelli, *Calcolabilità del giudizio giuridico e processo penale*, Giappichelli, Torino, 2025, pp. 42 ss., mette in evidenza come il giudizio di proporzionalità finisce per postulare un ampio potere "valutativo" in capo al giudice, spesso "creatore", di una nuova norma orientata al risultato, frutto dell'adozione di un'interpretazione "per princìpi", condotta secondo la logica *fuzzy* o *flou*.

umani restino coinvolti nel ciclo per salvaguardare (anche) la proporzionalità. I sistemi che contribuiscono ad automatizzare il processo decisionale devono coesistere e collaborare con i decisori umani, anziché sostituirli, nei contesti in cui potrebbero emergere requisiti di proporzionalità.

È questa la via eletta dal regolamento europeo, che assegna all'intelligenza artificiale un ruolo ancillare rispetto al giudice[72].

Proporzionalità, intelligenza artificiale e decisione penale possono incrociarsi in diversi momenti.

Si è visto già che il canone, rapportato alla fase delle indagini, mira a contenere l'azione investigativa: qui il giudice decisore è arbitro dell'ingresso dell'intelligenza artificiale sulla scena processuale (si pensi, ad esempio, all'autorizzazione preventiva richiesta per l'utilizzo di pratiche di riconoscimento facciale). Proseguendo su questa linea, ma spostandosi nel contesto del giudizio, il principio risulta funzionale a neutralizzare l'efficacia della prova, fondata sull'intelligenza artificiale, che sia stata raccolta in spregio ai diritti fondamentali[73]. Sempre in materia probatoria, il parametro potrebbe fungere da filtro ammissivo della prova algoritmica nel processo[74].

Ove l'intelligenza artificiale assista il giudice nel compito di decidere il giudizio[75] o di determinare le pene[76], la proporzionalità dovrebbe operare su un duplice fronte: *esternamente* al sistema informatico, come limite al suo impiego nel procedimento; *internamente* all'artefatto algoritmico, avvertendosi l'esigenza che la macchina sia capace di fornire *output* proporzionati e risultati rispettosi del principio. Tentativi di operazionalizzare il

[72] G. De Minico, *op. cit*., p. 107.

[73] Su tali prove cfr. M. Gialuz, "Prove fondate sull'intelligenza artificiale e diritti fondamentali", in *Diritto di Difesa*, 15 gennaio 2025, pp. 1 ss.

[74] Utili spunti possono trarsi dai criteri elaborati in *R. v. Mohan* (1994 CanLII 80 (SCC), [1994] 2 SCR 9) e *White Burgess Langille Inman v. Abbott e Haliburton Co* (2015 SCC 23).

Come rileva G. Ubertis, "Prova penale e proporzionalità", cit., p. 3, il principio di proporzionalità inerisce non alla *proof*, ma all'*evidence*, con specifico riferimento all'acquisizione di una fonte e/o di un mezzo di prova.

[75] Ci si riferisce alla c.d. *predictive justice*, su cui cfr. M. Torre, "Il Regolamento europeo sull'intelligenza artificiale", cit., p. 1551.

[76] J. Ryberg, *op. cit*., pp. 3255 ss.

canone non sono certamente mancati[77], ma sempre finendo per far risaltare la necessità di un giudizio umano sensibile a garanzia della proporzionalità. Il "controllo umano significativo"[78] da parte del giudice competente a decidere dovrà investire anche la proporzionalità dei risultati offerti dalla macchina. Del resto, l'*AI Act* riconosce all'uomo la facoltà di intervenire in ogni fase operativa del sistema informatico[79]. In tale prospettiva, gli *artificial intelligence system* devono essere muniti di misure che consentano all'utilizzatore: di interpretare in modo corretto il risultato restituito dal *software*; di astenersi dall'adoperare il sistema e, all'occorrenza, di "ignorare, annullare o ribaltare" l'*output*; di arrestare il funzionamento del dispositivo[80].

Da quest'angolazione, la supervisione umana può rappresentare un'importante garanzia di proporzionalità nella fase decisionale[81]. Tuttavia, la capacità dei supervisori umani di identificare i casi in cui mezzi e fini potrebbero essere sproporzionati

[77] G. Sartor, "A quantitative approach to proportionality", in G. Bongiovanni, G. Postema, A. Rotolo, G. Sartor, C. Valentini, D.N. Walton (eds.), *Handbook of Legal Reasoning and Argumentation*, Springer, Dordrecht, 2018, pp. 613 ss.

[78] Sul concetto, declinato in ambito processuale penale, si rinvia a G. Ubertis, "Intelligenza artificiale, giustizia penale, controllo umano significativo", in *Dir. pen. cont., Riv. trim*., (2020), n. 4, pp. 83 s.

[79] Le tipologie di gestione e monitoraggio della macchina prevedono un intervento minimo dove il controllo è circoscritto alla sola fase di progettazione ed è prevista *ex post* la possibilità di discostarsi dall'*output* (c.d. *human on the loop*) o, viceversa, un approccio che consente di governare pienamente l'automazione in ogni suo passaggio operativo (c.d. *human in command*).

Sui diversi *standard* di sorveglianza, graduati in base all'incidenza dell'intervento umano, cfr. I.P. Di Ciommo, "La prospettiva del controllo nell'era dell'Intelligenza Artificiale: alcune osservazioni sul modello Human In The Loop", in *Federalismi*, 19 aprile 2021, pp. 68 ss.; B. Marchetti, "La garanzia dello human in the loop alla prova della decisione algoritmica amministrativa", in *BioLaw Journal–Rivista di BioDiritto*, (2022), n. 2, pp. 367 ss.; D. Martire, "Human in the loop. L'essere umano come fattore condizionante della – o condizionato dalla – intelligenza artificiale", in *Dir. it. info. dir.*, (2024), n. 2, pp. 467 ss.; L. Rinaldi, "Intelligenza artificiale, diritti e doveri nella Costituzione italiana", in *DPCE on line*, (2022), n. 2, p. 214.

[80] Art. 14, par. 4, lett. c), d), e), del Regolamento (UE) 2024/1689.

[81] L. Enqvist, M. Naarttijärvi, "Discretion, Automation, and Proportionality", cit., p. 168.

dipenderà dal loro accesso alle informazioni e dalla reale possibilità di comprendere il modo in cui il sistema decide.

Ecco che il tema della sorveglianza umana si correla strettamente a quello della *trasparenza*[82]. Il regolamento pretende che il funzionamento della macchina sia "sufficientemente trasparente", onde permettere "ai deployer di interpretare l'*output* del sistema e utilizzarlo adeguatamente"[83], ma non si occupa del principale ostacolo per l'effettività della sorveglianza umana, ovvero l'indecifrabilità dei meccanismi operativi dell'intelligenza artificiale implementata dal *deep* o *machine learning*.

Si tratta di ostacoli non facilmente superabili attraverso interventi sul piano giuridico: affinché non rimanga solo una falsa promessa di controllo[84], la supervisione umana dovrà essere resa tecnicamente possibile attraverso le soluzioni offerte dall'*EXplainable Artificial Intelligence* (XAI)[85].

[82] M. Calvano, A. Curci, G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, "Building Symbiotic AI: Reviewing the AI Act for a Human-Centred, Principle-Based Framework", in *Minds and Machines*, 36 (2026), n. 1, p. 6. Si tratta di un concetto complesso a cui possono essere attribuiti due significati diversi: la capacità dell'algoritmo di (i) spiegare sé stesso e il suo funzionamento interno o (ii) rendersi fluido e rendere più evidenti i risultati oggettivi degli utenti: così G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, M.F. Costabile, "From human-centered to symbiotic artificial intelligence: a focus on medical applications", in *Multimedia Tools and Applications*, 84 (2025), n. 27, p. 32114.

[83] Art. 13, par. 1, del Regolamento (UE) 2024/1689.

[84] R. Koulu, "Proceduralizing control and discretion: human oversight in artificial intelligence policy", in *Maastricht Journal of European and Comparative Law*, 27 (2020), n. 6, pp. 729 s.

[85] Gli studi in materia, seppur in fase di sviluppo, aspirano nel prossimo futuro a risolvere il problema della opacità dei sistemi di giustizia predittiva. Su tali studi, cfr. A. Barreto Arrieta, N. Diaz Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", in *Information Fusion*, 58 (2020), pp. 82 ss.; L.K. Branting, C. Pfeifer, B. Brown, L. Ferro, J. Aberdeen, B. Weiss, M. Pfaff, B. Liao, "Scalable and explainable legal prediction", in *Artificial Intelligence Law*, 29 (2021), n. 2, pp. 213 ss.; R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, "A Survey of Methods for Explaining Black Box Models", in *ACM Computing Surveys*, (2018), n. 5, pp. 1 ss.; S. Lim, "Judicial Deci-

Ciò che sul piano giuridico non si dovrebbe mancare di fare è proibire, proprio in virtù del criterio di proporzionalità, l'uso nel procedimento penale di quei sistemi che sfuggono al dominio dell'uomo[86], laddove la supervisione umana dovesse ridursi ad un mero anestetico contro le paure dell'algoritmizzazione, incapace di garantire la tutela dei diritti fondamentali.

Potrebbe obiettarsi che la pericolosità di un *output* sproporzionato, dovendo essere questo necessariamente corroborato da altri elementi, secondo il principio di non esclusività algoritmica, sarebbe facilmente disinnescabile: si tratta di rilievo ingenuo, non potendosi trascurare taluni fenomeni, quali quello noto come "effetto gregge"[87].

7. "Controllo pubblico significativo" e proporzionalità *by design*

Come già innanzi emerso, il sistema di intelligenza artificiale in quanto tale è un altro elemento chiave per assicurare la proporzionalità.

Da un lato, i sistemi algoritmici dovrebbero rispettare le indicazioni sulla proporzionalità fornite dal legislatore. Basti pensare alla legislazione in materia di protezione dei dati[88], di cui

sion-Making and Explainable Artificial Intelligence", in *Singapore Academy of Law Journal*, (2021), n. 33, pp. 280 ss.

[86] In senso contrario, T. Sroka, "Artificial intelligence and the right to a fair trial", cit., p. 272.

[87] A riguardo, A. Santosuosso, G. Sartor, "La giustizia predittiva: una visione realistica", in *Giur. it.*, (2022), n. 7, p. 1761.

[88] In particolare, cfr. Direttiva (UE) 2016/680. Sul tema, G. Borgia, "Il trattamento di dati personali a fini di prevenzione, di indagine, di accertamento e di perseguimento di reati o di esecuzione di sanzioni penali: quali passi avanti alla luce dei recenti sviluppi?", in A. Mantelero, D. Poletti (a cura di), *Regolare la tecnologia: il Reg. UE 2016/679 e la protezione dei dati personali. Un dialogo tra Italia e Spagna*, Pisa University Press, Pisa, 2016, pp. 495 ss.; M. Càrpino, *Il trattamento dei dati personali per finalità di prevenzione, indagine, accertamento e perseguimento dei reati*, Torino, Giappichelli, 2024, p. 84 ss.; B. Galgani, "Giudizio penale, Habeas data e garanzie fondamentali", in *Arch. pen. (web)*, (2019), n. 1, pp. 1 ss.; P. Troisi, "La protezione dei dati trattati a fini di prevenzione e accertamento dei reati", in S. Sica, V. D'Antonio, G.M. Riccio (a cura di), *La nuova disciplina europea della privacy*, Wolters Kluwer Cedam, Milano, 2016, pp. 314 ss.

– è noto – i sistemi algoritmici si nutrono, che assegna rilievo all'autodeterminazione informativa del singolo, ammettendone la sua limitazione, nel rispetto del canone di proporzionalità, per assicurare l'efficace conseguimento degli obiettivi preventivi e repressivi dei reati, che giammai potrebbero giustificare approvvigionamenti selvaggi e indiscriminati di dati personali. Assicurare la proporzionalità in sede di implementazione del sistema significa pretendere che rispetti le norme e i principi di protezione dei dati, come il principio di minimizzazione dei dati, e che i requisiti di proporzionalità contenuti a livello normativo siano tradotti in attribuiti di sistema (ad esempio, attraverso misure tecniche e organizzative proporzionate per proteggere i dati o attraverso restrizioni alla loro condivisione)[89].

Da altro lato, anche in mancanza di regole giuridiche di proporzionalità da tradurre in requisiti di sistema, le macchine intelligenti dovrebbero essere progettate per generare "*output* proporzionati" e, per garantire questo, assoggettate ad un "controllo pubblico significativo"[90], che imponga il rispetto del principio di proporzionalità *by design*. L'*AI Act*, improntato al principio di neutralità tecnologica, offre poche indicazioni su come la proporzionalità debba essere considerata nella creazione, nello sviluppo, nell'implementazione e nell'utilizzo dei sistemi intelli-

Sul d.lgs. 18 maggio 2018, n. 51, con il quale l'Italia ha recepito la Direttiva (UE) 2016/680, si v. G.M. Baccari, *Il trattamento (anche elettronico) dei dati personali per finalità di accertamento dei reati*, in A. Cadoppi, S. Canestrari, A. Manna, M. Papa (diretto da), *Cybercrime*, Utet, Torino, 2023, pp. 1881 ss.; A. Ricci, "Il trattamento di dati personali per finalità di prevenzione, indagine accertamento e perseguimento di reati. Riflessioni sul d.lgs. 18 maggio 2018, n. 51, di attuazione della dir. 2016/680/UE", in *Nuove leggi civili*, (2019), n. 2, pp. 565 ss.

[89] L.A. Bygrave, "Minding the machine v2.0: The EU General Data Protection Regulation and automated decision-making", in K. Yeung, M. Lodge (eds.), *Algorithmic regulation*, Oxford University Press, Oxford, 2019, pp. 248 ss.; I. Mendoza, L.A. Bygrave, "The right not to be subject to automated decisions based on profiling", in T. Synodinou, P. Jougleux, C. Markou, T. Prastitou (eds.), *EU internet law: regulation and enforcement*, Springer International Publishing, Cham, 2017, pp. 77 ss.

[90] Sulla distinzione tra controllo pubblico significativo, controllo giudiziario significativo e controllo umano significativo, cfr. I. Taylor, "Justice by Algorithm: The Limits of AI in Criminal Sentencing", in *Criminal Justice Ethics*, 42 (2023), n. 3, pp. 193 ss.

genti[91]. Tuttavia, i progettisti sono già da tempo impegnati nella costruzione di macchine "allineate" ai valori degli esseri umani, considerati il fulcro del processo di progettazione e sviluppo dei sistemi di intelligenza artificiale, per questo definita come incentrata sull'uomo. In questo contesto, l'intelligenza artificiale simbiotica, un paradigma più evoluto di quella *human-centered*, promette di migliorare le capacità umane attraverso una collaborazione più profonda e continua tra l'intelligenza umana e quella artificiale. Ebbene, occorrerebbe annoverare anche il canone della proporzionalità tra i principi e le proprietà enunciate come fondanti per la creazione di sistemi incentrati sull'uomo. Del resto, per determinare se un sistema di intelligenza artificiale sia incentrato sull'uomo, non si è mancato di fare riferimento proprio alla proporzionalità[92].

8. Decisioni sproporzionate e rimediabilità *ex post*

Gli individui soggetti a decisioni sproporzionate emesse da sistemi di intelligenza artificiale o con il loro ausilio devono essere in grado di avvalersi del diritto di impugnarle e contestarle.

Tuttavia, gli elementi a disposizione per rivederle potrebbero essere limitati, a causa della scarsa *explainability* dei sistemi di intelligenza artificiale[93]. Approcci differenti, ancorati al principio di ridondanza[94], pur proposti in dottrina per superare tali

[91] N. Smuha, E. Ahmed-Rengersb, A. Harkens, W. Li, J. MacLaren, R. Piselli, K. Yeung, "How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act", in https://ssrn.com/abstract=3899991, 5 agosto 2021, ove si legge "it remains unclear who determines what is necessary and proportionate, whether this occurs during the certification and conformity assessment processes, and whether there are means to dispute the proportionality assessment of a given system."

[92] D. Restrepo Amariles, P.M. Baquero, *op. cit*., pp. 9 s.

[93] Per questi spunti G. Lasagni, "Difendersi dall'intelligenza artificiale o difendersi con l'intelligenza artificiale? Verso un cambio di paradigma", in *Riv. it. dir. proc. pen.*, (2022), n. 4, p. 1555.

[94] Sul *redundancy approach* cfr. G. Lasagni, *op. cit*., pp. 1557 s.; J. Saccomani, "L'impatto della giustizia algoritmica sul diritto all'equo processo", in *Cass. pen.*, (2023), n. 2, pp. 644 s.; L. Pressacco, "Intelligenza artificiale e ragionamento probatorio nel processo penale", in G. Di Paolo, L. Pressacco (a cura di), *Intelligenza artificiale e processo penale. Indagini,*

difficoltà, di fatto non forniscono all'interessato gli strumenti per vagliare la solidità degli *output* prodotti dall'intelligenza artificiale e contestarne i contenuti[95].

Poiché la decisione adottata con l'ausilio algoritmico resta pur sempre umana, il primo strumento per assicurare il suo controllo *ex post* va individuato nella motivazione[96], irrinunciabile presidio di legalità e garanzia nel processo penale, ancor di più in un contesto che vede l'impiego di sistemi algoritmici[97].

In parallelo, occorre affinare la possibilità di contestare il "suggerimento" stesso fornito dalla macchina al giudice, specialmente se questi, dopo aver esercitato il doveroso "controllo giudiziale significativo" sull'*output* algoritmico[98], lo abbia condiviso: qui vengono in gioco le potenzialità della *contestable AI*.

La contestabilità è una "'deep system property', representing a coming together of human and machine to jointly make decisions"[99]; nel dominio della giustizia penale la contestabilità si ri-

prove, giudizio, Editoriale Scientifica, Napoli, 2022, p. 130; G. Ubertis, "Perizia, prova scientifica e intelligenza artificiale nel processo penale", in *Sist. pen.*, 3 giugno 2024, p. 18.

[95] Criticano il *redundancy approach* G.M. Baccari, G. Pecchioli, "I.A. e giudizio sul fatto: gli strumenti di e-evidence per la cognizione", in G.M. Baccari, P. Felicioni (a cura di), *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, Giuffrè Francis Lefebvre, Milano, 2023, p. 159.

[96] Nel senso che l'indefettibilità della motivazione deriva dal principio di proporzionalità, in quanto solo motivando il provvedimento è possibile dare conto dell'avvenuto bilanciamento tra la misura adottata e l'esigenza perseguita, M. Torre, "Indagini informatiche e principio di proporzionalità", in *Proc. pen. giust*., (2019), n. 6, p. 1434.

[97] R. Bichi, "Intelligenza digitale, giurimetria, giustizia predittiva e algoritmo decisorio. *Machina sapiens* e il controllo sulla giurisdizione", in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffré, Milano, 2020, p. 426; O. Di Giovine, "Il Judge-bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione del precedente)", in *Cass. pen*., (2020), n. 3, p. 960; L. Marafioti, "Intelligenza artificiale, giudizio penale e garanzie processuali", in *Ind. pen*., (2023), n. 3, pp. 16 ss.; D. Perrone, *La prognosi postuma tra distorsioni cognitive e software predittivi. Limiti e possibilità del ricorso alla "giustizia digitale integrata" in sede di accertamento della colpa*, Giappichelli, Torino, 2021, p. 118.

[98] I. Taylor, *op. cit.*, p. 202.

[99] K. Alfrink, I. Keller, G. Kortuem, N. Doorn, "Contestable AI by Design: Towards a Framework", in *Minds & Machines*, 33 (2023), n. 4, pp. 615 ss. Gli A. precisano che la "contestability helps to protect against fallible, un-

ferisce alla capacità degli *stakeholders* di comprendere, mettere in discussione e contestare le decisioni e i risultati dei sistemi di intelligenza artificiale[100].

Non si concentra solo sui meccanismi *post-hoc*, ma riguarda l'intero ciclo di vita dei sistemi di intelligenza artificiale[101], attraverso un reticolo di "funzionalità" (*contestability features*) e "pratiche" (*contestability by design*) di sviluppo degli stessi[102]. Tra queste, particolare rilievo assume la *spiegazione*, che trova assonanza con il diritto alla stessa contemplato nell'*AI Act*, a sua volta declinazione di quello, già riconosciuto dal GDPR e dal Regolamento (UE) 2018/1725, ad ottenere dal titolare del trattamento dei dati personali informazioni circa l'esistenza di un processo decisionale automatizzato, la logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento[103].

I rimedi nell'ambito del procedimento penale non sarebbero certamente effettivi se le spiegazioni fossero limitate "[a]l ruolo del sistema di IA nella procedura decisionale e [ai] principali

accountable, illegitimate, and unjust automated decision-making, by ensuring the possibility of human intervention as part of a procedural relationship between decision subjects and human controllers" (p. 617).

[100] F. Maoro, M. Geierhos, "Contestable AI for criminal intelligence analysis: improving decision-making through semantic modeling and human oversight", in *Front. Artif. Intell*., 8 (2025), p. 2.

[101] Inquadrano la contestabilità come un processo dinamico che consente agli utenti di interagire con i sistemi di intelligenza artificiale e contestare le decisioni C. Moreira, A. Palatkina, D. Braca, D.M. Walsh, P.J. Leihn, F. Chen, N. C. Hubig, "Explainable AI Systems Must Be Contestable: Here's How to Make It Happen", in *arXiv:2506.01662*, 2 giugno 2025.

[102] Sulle quali K. Alfrink, I. Keller, G. Kortuem, N. Doorn, *op. cit*., pp. 614 ss. Le funzionalità sono: 1. misure di sicurezza integrate contro comportamenti dannosi; 2. controllo interattivo sulle decisioni automatizzate; 3. spiegazioni del comportamento del sistema; 4. richieste di revisione e intervento umano; 5. strumenti per il controllo da parte di soggetti o terze parti. Le pratiche sono: 1. misure di sicurezza *ex ante*; 2. approcci agonistici allo sviluppo del *machine learning* (ML); 3. garanzia della qualità durante lo sviluppo; 4. garanzia della qualità dopo l'implementazione; 5. strategie di mitigazione del rischio; 6. supervisione di terze parti.

[103] F. Donati, "La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale", in *Rivista AIC*, 2025, n. 1, p. 11.

elementi della decisione adottata"[104]. Non basta la descrizione del "come" una certa decisione penale sia stata presa: la spiegazione dovrà includere anche il "perché" sia stata adottata[105]. Ciò che serve, in definitiva, è la *giustificazione*[106].

9. Mezzo e fine contro l'eccesso di potere algoritmico

La proporzionalità può rendere il sistema giuridico resiliente alle tensioni interne tra i diversi e talora contrastanti obiettivi che persegue. I valori che il parametro pondera sono soggetti a trasformazioni nel corso del tempo, così come il diritto stesso si evolve gradualmente: *ergo*, la proporzionalità è per sua natura intrinsecamente soggetta al cambiamento e, in questo senso, può identificarsi con un processo, piuttosto che uno stato da raggiungere una volta per tutte.

I processi decisionali, compresi quelli soggetti a un certo grado di automazione, non possono trascurarne l'importanza. Questo vale specialmente per l'intelligenza artificiale che voglia essere antropocentrica: la proporzionalità funge da barriera contro l'eccesso di potere dell'autorità o il suo esercizio irrazionale o irragionevole, contribuendo alla costruzione e all'attuazione della dimensione umana del diritto.

Non è solo un principio di notevole rilievo nell'ordinamento giuridico europeo, dove ha trovato sempre più ampi ed espliciti riconoscimenti, come in materia di cooperazione giudiziaria e di polizia[107], ma lo è (o lo è diventato) anche negli ordinamenti giuridici nazionali dell'Unione europea.

[104] L'Art. 86 del regolamento permette a qualsiasi persona interessata oggetto di una decisione basata sugli *output* di sistemi di IA ad alto rischio "che produca effetti giuridici o in modo analogo incida significativamente su tale persona in un modo che essa ritenga avere un impatto negativo sulla sua salute, sulla sua sicurezza o sui suoi diritti fondamentali", di "ottenere dal deployer spiegazioni chiare e significative sul ruolo del sistema di IA nella procedura decisionale e sui principali elementi della decisione adottata".

[105] K. Alfrink, I. Keller, G. Kortuem, N. Doorn, *op. cit*., p. 624.

[106] Sulla netta distinzione tra spiegazioni e giustificazioni, cfr. C. Henin, D. Le Métayer, "Beyond explainability: justifiability and contestability of algorithmic decision systems", in *AI & Soc*iety, 37 (2022), n. 4, p. 1399.

[107] A proposito, volendo, L. Pulito, *Le squadre investigative comuni. Prodromi, retrospettive, avanguardie*, Cacucci, Bari, 2023, spec. pp. 234 ss.

L'indiscutibile importanza che il canone ha acquisito reclama una collocazione centrale all'interno delle prospettive giuridiche, oltre che tecniche, concernenti l'uso dell'intelligenza artificiale e la sua implementazione nei processi decisionali in ambito penale. Le principali questioni critiche che riguardano i sistemi artificialmente intelligenti, quali la discriminazione, i pregiudizi e la trasparenza, non possono essere adeguatamente trattate e risolte se non affrontate anche nell'ottica della proporzionalità, ineludibile crocevia per la tutela dei diritti fondamentali, la razionalità e l'umanità del diritto.

Il principio è indicativo allo stesso tempo di un mezzo (un metodo generale da seguire per un'adeguata tutela dei diritti individuali in ambito processuale penale) e di un fine (un parametro utile a verificare la giustizia della soluzione presa nel caso concreto)[108].

Se applicata, la proporzionalità può impedire che tanto l'esercizio del potere cumulato nelle mani della pubblica accusa, abnormemente accresciuto dalla possibilità di ricorrere ad un nuovo e potentissimo arsenale investigativo, quanto l'esercizio del potere decisionale del giudice, potenziato nelle prestazioni cognitive dalla capacità della macchina, si traducano in un *eccesso di potere algoritmico*.

[108] M. Caianiello, *op. cit*., p. 145.

Serena Quattrocolo

Nuove sfide per la giustizia penale: la rilevanza processuale di immagini e suoni sintetici

Abstract: **New Challenges for Criminal Justice: Synthetic Images and Sounds in the Context of Criminal Trial.** The chapter focuses on the risk of admitting deepfake as evidence in criminal proceedings. Giving the impressive quality of synthetic images and sounds, which cannot be discovered through human senses, a technical assessment is needed. Images, in particular, have a strong cognitive impact that can hardly be removed from the decision-making process, even when the piece of evidence has proved non-genuine. For this reason, it is crucial to move also the Italian process for admission of evidence towards the model of the "gatekeeper", which is typical of the northern american common law.

Keywords: Criminal Proceedings, Evidence, Image, Demonstrative Value, Expert Witness.

Sommario: 1. Il contesto – 1.1. I campi di indagine – 2. Il potere evocativo dell'immagine, autentica o falsa – 2.1 *Deepfake* e GANs – 2.2. Nuove frontiere dalla prova mendace – 3. Lo scenario normativo: dall'*AI Act*… – 3.1. … alla legge italiana 23.9.2025 n. 132 – 4. Trend giurisprudenziali – 4.1. Una recente sentenza della II sezione della Corte di cassazione – 5. Qualche proposta per la discussione.

1. Il contesto

Il tema del rapporto tra tecnologia e processo penale ha da tempo affiancato, nella discussione pubblica, un altro tradizionale argomento, ovvero il rapporto tra scienza e processo. Le due prospettive non devono essere confuse, poiché incidono su profili diversi della complessità penale, con problematiche (e possibili soluzioni) talvolta sovrapponibili, talvolta diversificate. Infatti, la molteplicità dei piani sui quali l'impiego delle tecnologie digitali, e, più specificamente, della *data driven AI*, è ormai divenuto quotidiano e ad accesso generalizzato sono tali da rendere il fenomeno rilevante sia dal punto di vista dell'ingresso e della cittadinanza che gli *output* di tali tecnologie pos-

sono trovare nel processo penale[1], sia da quello sostanziale, della legittimità dell'uso (o di certi usi) di tali tecnologie, come si vedrà qui di seguito.

Inevitabilmente, infatti, si è, con una certa inerzia, avviata una "reazione normativa" al ricordato fenomeno del dilagare delle tecnologie digitali, le cosiddette ICT, *Information and Communication Technologies*, fenomeno che, tuttavia, è per sua natura sempre in corso di evoluzione e, dunque, difficile da inquadrare in precetti normativi. Questa breve riflessione cercherà, allora, di muoversi tra i "pieni" e i "vuoti" dell'attuale quadro legislativo, segnalando soprattutto gli aspetti delle tecnologie digitali qui all'attenzione che maggiormente sfuggono alle possibilità di regolamentazione, con particolare riguardo a un'area delle ICT che è ormai comunemente definita il "*deepfake*", ovvero, per quanto qui ci interessa, la generazione di contenuti multimediali sintetici e, in particolare, di immagini, video e suoni.

1.1. I campi di indagine

Come già anticipato, la correlazione che il fenomeno del *deepfake* ha assunto rispetto al quadro generale della giustizia penale è duplice. Per un verso, la prima e principale preoccupazione riguarda l'ingresso di contenuti multimediali del tutto sintetici quali possibili mezzi di prova nel processo penale. Sin dal titolo si è scelto di utilizzare il temine 'sintetico' poiché il ricorso al lemma, certamente più familiare, di 'artefatto' non corrisponde alla effettiva natura dei dati cui ci riferiamo, i quali rappre-

[1] Interessante segnalare la distinzione impiegata da P. Ferrua, "La prova nel processo penale, profili generali", in P. Ferrua, E. Marzaduri, G. Spangher (a cura di), *La prova penale*, Giappichelli, Torino 2013, p. 15, ove si afferma che "le prove critico-indiziarie possono essere distinte tra artificiali e naturali a seconda che siano o no congegnate dall'artificio umano allo scopo di provare (*rectius*, *anche* di provare, non essendo necessario che questa sia la loro unica funzione). Tra le prime, la fotografia, la registrazione di suoni o di immagini". È evidente l'utilità di tanto autorevole osservazione per evidenziare, sin da subito, come il tema che ci occupa si ponga ad un livello ulteriore di complessità, giacché si ragiona, in questa sede non solo di 'prove artificiali', nel senso qui sopra riportato, ma di elementi che non sono stati generati per il tramite di alcun 'artificio umano', bensì attraverso sistemi digitali pressoché del tutto autonomi nella generazione dei risultati.

sentano fatti non veritieri o persone non esistenti, ma possono essere estranei al concetto di 'artificio' che, come sottolineato nella nt. 1, implica, nel linguaggio giuridico comune, la volontà umana di far passare per autentico ciò che non lo è. Come si sottolineerà nei paragrafi che seguono, per comprendere la rilevanza del problema, occorre afferrare, almeno superficialmente, le caratteristiche e il funzionamento dei *software* generativi che li producono ed a questo aspetto sarà dedicata una parte del presente lavoro.

Per comodità e per brevità, sia consentito muovere dalla distinzione, altrove proposta, tra "prova digitale" e "prova generata automaticamente"[2]. La *prova digitale* possiede ormai un proprio statuto processuale[3], che deriva certamente dall'essenza non analogica del dato conoscitivo o del supporto sul quale esso è raccolto, ma che si estende, in prima battuta, ad elementi che sono stati generati attraverso l'impulso e la volontà umana (si pensi ad un file che raccolga la contabilità "sommersa" di un'azienda, massicciamente sottratta alle imposizioni fiscali; alle email dal contenuto minaccioso inviate dall'imputato alla vittima; ecc..). Al contrario, le *prove generate automaticamente* sono elementi conoscitivi che possono essere tratti a partire da dati che non sono generati dietro *input* umano per scopi processuali specifici. Certamente, la loro estrazione, conservazione e copia risponderà alle regole generali stabilite per tutte le prove

[2] V., innanzitutto, S. Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, Springer, Cham, 2020, pp. 73 ss.; v. altresì Ead., "L'ammissione della prova alla luce della rivoluzione digitale", in E.M. Catalano, P. Ferrua, *Corderiana*, Giappichelli, Torino 2023, pp. 170 ss.

[3] In un panorama ormai molto ricco, si segnalano alcuni spunti, in progressivo arco temporale: v. già P. Tonini, "Documento informatico e giusto processo", in *DPP*, (2009), pp. 401 ss.; M. Daniele, "La prova digitale nel processo penale", in *RDP*, (2011), p. 283 s.; G. Di Paolo, voce "La prova informatica (diritto processuale penale)", in *Enc. Dir., Annali VI*, Giuffré, Milano, 2013, pp. 736 ss.; G. Ziccardi, "L'ingresso della computer forensics nel sistema processuale italiano: alcune considerazioni informatico-giuridiche", in L. Lupària (a cura di), *Sistema penale e criminalità informatica. Profili sostanziali e processuali nella Legge attuativa della Convenzione di Budapest sul cybercrime*, Giuffré, Milano, 2009, pp. 165 ss.; M. Torre, "La prova informatica", in C. Conti, A. Marandola (a cura di), *La prova scientifica*, Giuffré, Milano, 2023, pp. 443 ss.

digitali, ma la loro essenza, estranea ad un intervento umano diretto, impone, a priori, di interrogarsi sulla loro stessa natura di *prove* … [4]. Infatti, si tratta di "potenziali prove", in relazione alle quali occorre ammettere un difetto di conoscenza (e conoscibilità) della genesi[5], difficilmente superabile, derivante non solo dalla natura intrinsecamente opaca di ciascun *software*, ma, soprattutto, dall'eccezionale livello raggiunto dalle metodiche di auto-apprendimento[6], che non consentono più, come vedremo, di distinguere a prima vista tra fatti, immagini, suoni genuini e sintetici[7]. In tale prospettiva, la spinta a rivedere completamente l'approccio al tema della prova e del suo ingresso nel processo penale, per cercare di adeguarlo alla mutata dimensione della realtà quotidiana si fa urgente: la proposta categoria delle "prove generate automaticamente" offre un primo punto di partenza che deve tuttavia essere significativamente raffinato alla luce delle specifiche caratteristiche dei c.d. *deepfake*[8]. Come anticipato, tra questi, saranno qui oggetto di specifica attenzione, in quanto paradigmatiche, le immagini sintetiche, dotate di un'estrema

[4] S. Barona Vilar, "Una justicia 'digital' y 'algoritmica' para una sociedad en estado de mudanza", in S. Barona Vilar (a cura di), *Justicia algoritmica y neuroderecho*, Tirant lo Blanc, Valencia, 2021, pp. 26 ss.

[5] L. Lupària, G. Fiorelli, "Diritto probatorio e giudizi criminali ai tempi dell'Intelligenza Artificiale", in *DPC*, *riv. trim*., (2022), n. 2, pp. 39 ss.

[6] Per tutti i profili computazionali si farà riferimento all'utile guida, estremamente accessibile, di G. Vessio, *Intelligenza artificiale per curiosi*, Edizioni Dedalo, Bari, 2025, pp. 25 ss. per le basilari metodiche di apprendimento computazionale.

[7] Invero la idoneità di dimostrare l'esistenza di un fatto rilevante per il processo sintetizza e dà ordine a tutta l'attività del giudicare: osserva, infatti, M. Taruffo, "La verità nel processo", in M. Caputo, G. Forti, G. Varraso (a cura di), *'Verità' del precetto e della sanzione penale alla prova del processo,* Jovene, Napoli, 2014, p. 188, "nessuna norma viene applicata correttamente se i fatti ai quali si riferisce sono stati accertati in maniera erronea o falsa".

[8] Al di fuori, naturalmente, di una dimensione normativa del fenomeno, che nasce come squisitamente computazionale, si può fare riferimento, per una efficace definizione al report di Interpol, *Beyond Illusions. Unmasking the threat of synthetic media for law enforcement*, 2024, p. 7 s., ove si circoscrive l'area del *deepfake* alla generazione di immagini e di suoni sintetici, volti a ricreare con assoluta verosimiglianza, scenari e contesti del tutto illusori.

capacità evocativa che può nuocere, su più livelli, al processo penale[9].

Per altro verso, alcune forme di impiego e diffusione di contenuti multimediali generati con le medesime tecnologie ha recentemente assunto rilievo penalistico, attraverso una nuova figura astratta di reato, cui si accompagnano una serie di circostanze aggravanti, introdotte dalla legge 132/2025, pubblicata nella Gazzetta Ufficiale del 23 settembre 2025. Pur in assenza di adeguate competenze per una completa e puntuale analisi sostanziale delle nuove fattispecie, pare qui necessario soffermarsi sui principali aspetti che derivano dalla trasformazione del *deepfake* da mero fenomeno digitale a concetto normativo, pur impreciso.

2. Il potere evocativo dell'immagine, autentica o falsa

Il termine *deepfake* è un neologismo derivato dalla crasi tra *deep learning*[10] – diffuso sistema di funzionamento computazionale, sottocategoria specifica del *machine learning* – e il termine inglese *fake*, che indica, in generale, la non genuinità di un oggetto o la non attendibilità di un dato, di una informazione[11].

Come anticipato, ci si concentrerà qui su una specifica area di *deepfake* e, in particolare, sul paradigma dei c.d. GANs, *software* dall'evocativo nome di *Generative Adversarial Networks* (termine ormai entrato anche in numerosi dizionari stranieri, come Collins e Merriam-Webster), elaborati innanzitutto nel contesto della creazione di nuove immagini. Come già anticipato, la

[9] L'accertamento processuale appare messo più che mai a rischio dai *deep fake*: A. McPeak, "The Threat of Deepfakes in Litigation: Raising the Authentication Bar to Combat Falsehood", in *Vanderbilt Journal of Entertainment and Technology Law*, 23 (2021), n. 2, pp. 443 ss.; R. Delfino, "Deepfakes on Trial: a Call to Expand the Trial Judge's Gatekeeping Role to Protect Legal Proceedings form Technological Fakery", in *Hastings Law Journal*, 74 (2023), n. 2, pp. 293 ss.

[10] V. G. Vessio, *op. cit.*, pp. 61 ss.

[11] Il tema delle *fake news* rappresenta un ambito particolarmente ricco di spunti e contributi e anche di intrecci estranei al processo penale: di recente v. J. Rinceanu, R. Stephenson (2025), "Restarting the Past: Fake News in the Digital Age", in *DIGICON*, recuperato da https://digi-con.org/restarting-the-past-fake-news-in-the-digital-age/.

scelta cade su questo ambito in ragione della estrema capacità evocativa propria dell'immagine che veicola informazioni persistenti nel tempo, ben più di altri canali[12], a prescindere dalla consapevolezza, in vario modo poi raggiunta[13] che il contenuto da esse rappresentato non sia genuino e veritiero[14].

2.1. *Deepfake* e GANs

Come già altrove riportato, fondamentale nell'approccio al tema che ci occupa è il lavoro pubblicato da un team di ricerca, nel 2014, circa la realizzazione della nuova metodica dei GANs[15]. Da allora, numerose versioni (GDWCT, StartGAN, AttGAN, SytleGAN, StyleGAN2, giusto per citare i più noti e affermati) sono state sviluppate sulla base di un modello che, ispirato alla teoria dei giochi, utilizza due unità di apprendimento, costituite da reti neurali[16], cioè un c.d. "generatore" ed un c.d. "discriminatore"[17]. Il primo ha la funzione di generare nuove immagini partendo da una banca dati che contiene rappresentazioni autentiche. Tale *output* viene trasmesso al discriminatore, che ha la funzione di confrontarla con una banca dati di sole

[12] A sottolineare la serietà e l'urgenza di affrontare le sfide specifiche poste dal *deepfake*, si veda il report dell'Innovation Lab di Europol, *Facing reality? Law enforcement and the challenge of deepkafe*, Lussemburgo, 2022. Si veda l'interessante *screening* della letteratura internazionale effettuato da M.P. Sandoval, M. de Almeida Vau, J. Solaas, L. Rodrigues, "Threats of deepfakes to the criminal justice system", in *Crime Science*, 13 (2024), art. n. 41.

[13] Sul punto, di straordinario interesse il saggio di T. Myhand, "Once the jury sees it, the jury can't unsee it", in *Widener Law Rev*, 29 (2023), n. 2, pp. 172 ss., che sin dal titolo mette a fuoco proprio il rischio estremo legato all'ingresso nel processo di elementi visivi non genuini o rappresentativi di fatti non veritieri.

[14] Per una panoramica sugli studi che hanno evidenziato tale potere, con particolare riguardo ai video, J. Channing Ruff, "The Federal Rules of Evidence Are Prepared for Deepfakes. Are You?", in *Review of Litigation*, 41 (2021), p. 112.

[15] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Nets", in *Cornell University Library*, 10 giugno 2014.

[16] Per comprendere cosa siano le reti neurali, G. Vessio, *op. cit.*, p. 41.

[17] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Networks", in *Communications of the ACM*, 63 (2020), n. 11, pp. 139 ss.

immagini genuine, per individuare eventuali discrepanze e stabilire se si tratti di una immagine reale o artefatta. Fintanto che il discriminatore è in grado di individuare le immagini artefatte, il risultato non è soddisfacente: pertanto, il discriminatore fornisce un "*feedback*"[18] al generatore, offrendo le indicazioni necessarie per affinare, di volta in volta, il tasso di verosimiglianza dell'immagine creata, fino a che lo stesso discriminatore non sarà più in grado di stabilire se l'*output* proveniente dal generatore sia una immagine vera o sintetica.

In conclusione, le immagini prodotte da un GAN sono tali da non poter essere discriminate tra rappresentative di volti e oggetti veri o inesistenti dallo stesso *software* che le ha generate. È dunque altamente inverosimile che la percezione umana possa efficacemente operare tale selezione.

Peraltro, non si tratta certamente della sola tecnologia idonea a raggiungere tale risultato. Numerose altre architetture di reti neurali, per lo più basate su metodiche di ottimizzazione, hanno raggiunto risultati sorprendenti nella generazione di dati artificiali, come i *diffusion models*, che generano immagini del tutto realistiche a partire da una descrizione testuale, o gli *autoencoders*, architetture non supervisionate di apprendimento che consentono la eliminazione del rumore da immagini reali, l'individuazione delle anomalie, il riconoscimento facciale ma, soprattutto, la generazione di immagini di cose o persone non esistenti.

Dati questi sorprendenti risultati, che – si ricorda – sono accessibili anche su larga scala, quindi al di fuori di una pura dimensione di ricerca, l'elemento dirimente è la già ricordata difficoltà, per i sensi umani, di percepire la non genuinità del dato generato automaticamente. Non si deve dimenticare che, anche in questa stagione di inedita potenza computazionale non mancano, comunque, *output* artefatti (perché realizzati con intervento umano esterno), di cui è possibile isolare almeno un nucleo di autenticità: tuttavia, l'artificio può variare da un grado

[18] Si tratta, in realtà di un algoritmo di *backpropagation* basato sul confronto tra l'input che ha raggiunto il discriminator in entrata e l'*output* (I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *ivi*, p. 141).

di realismo elevatissimo, fino a livelli più grossolani e, per questa ragione, la sua individuazione può essere talvolta altrettanto complessa rispetto alla scoperta di immagini del tutto *fake*.

A fronte del *deepfake*, infatti, solo analoghe architetture generative sono in grado di operare la cosiddetta *fake-detection*, sfruttando difetti di funzionamento dei generatori, che siano state nel frattempo scoperte ed osservate[19]. In questo senso, la *fake detection* si colloca necessariamente un passo indietro rispetto alla generazione di contenuti multimediali non genuini, a prescindere dalle finalità sottostanti, che possono certamente essere di ricerca, commerciali, ricreative, ma anche fraudolente[20], in un perpetuo "inseguimento" in cui i difetti riscontrati consentono, per un verso, di potenziare la *detection*, ma, per altro verso, anche l'ulteriore miglioramento dell'architettura che, sanate le criticità, tornerà ad essere "impenetrabile". Analoghe considerazioni valgono per aree limitrofe delle ICT, prima e più nota tra tutte quella dei *Large Language Models*, con i *Generative Pre-trained Transformers* (GPT), ma anche quella degli audio sintetici, con varie architetture come i *Recurrent Neural Networks*, i *WaweGANs* e i *variational autoencoders*, in grado di generare file audio apparentemente genuini, con voci di persone esistenti, ma dal contenuto completamente artefatto.

2.2. Nuove frontiere dalla prova mendace

La breve e approssimativa digressione in ambito tecnico era necessaria per comprendere entrambi i profili di complessità cui questo breve lavoro è dedicato, sul piano del valore probatorio delle immagini sintetiche e su quello della rilevanza penale che

[19] O. Giudice, L. Guarnera, S. Battiato, "Fighting deepfakes by detecting gan dct anomalies", in *Journal of Imaging*, 7 (2021), n. 8, p. 128; L. Guarnera, O. Giudice, S. Battiato, "Deepfake detection by analysing convolutional traces", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 666 s.

[20] Per un punto di vista generale, v. R. Chesney, D. Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy and National Security", in *California Law Review*, 107 (2019), pp. 1753 ss.: gli autori sottolineano il rischio esponenziale dei *deepfake* nei contesti in cui non è garantito il confronto *adversarial*, proprio del paradigma del *due process*.

la diffusione di dette immagini *fake* ha assunto nel nostro ordinamento.

Sotto il primo profilo, la novità del tema non deve farci dimenticare l'esistenza, da sempre, di un rischio, altrove definito 'neutro'[21], che consiste nella possibilità dell'ingresso accidentale, e non proditorio, nel processo, di elementi informativi che non rappresentano fatti veritieri. Si può dire che la prevenzione di tale rischio ispira, tradizionalmente, i sistemi probatori, a tutte le latitudini: evitare l'ingresso – o, quantomeno l'utilizzo – di una prova non genuina, è l'obiettivo precipuo di ogni sistema, per scongiurare il rischio che il processo si allontani dalla sua funzione essenziale, ovvero la ricostruzione dei fatti contestati. La caducità della memoria umana rappresenta l'esempio più eclatante: al di là del dolo del falso testimone *ex* art. 372 c.p., proprio l'approssimazione del ricordo e le note fallacie cognitive hanno ispirato i meccanismi di *cross-examination*, volti a testare l'attendibilità dei testimoni…

Nell'odierna realtà, come visto, sussiste altresì un nuovo rischio "estremo", ovvero l'eventualità che elementi informativi digitali siano appositamente generati, per dimostrare nel processo fatti non accaduti o non veritieri, con una qualità senza precedenti, che ne rende estremamente inverosimile la individuazione: si pensi all'immagine dell'imputato ritratto in compagnia di persone che, invece, non ha mai incontrato[22]. Qui non solo viene in gioco, come sopra ipotizzato, lo sviamento del processo dalle sue funzioni, ma addirittura il suo asservimento a finalità criminose, grazie a strumenti dalle prestazioni in precedenza inimmaginabili.

Se, dunque i rischi e gli interessi in gioco non sono nuovi, inedita è la difficoltà di individuare la non autenticità e la domanda ineludibile riguarda la necessità di adottare nuove e diverse contromisure, ulteriori rispetto agli strumenti da sempre messi in

[21] S. Quattrocolo, "Deepfake e fine della tipicità probatoria? Fatto, prova e giudizio penale a confronto con la realtà delle immagini sintetiche", in *Leg. pen*., 21 luglio 2025, p. 8.

[22] Si pensi altresì, sotto il profilo audio, al rischio che conversazioni intercettate sui canali VoiP possano includere attestazioni generate da un RRN, poiché riconducibili alla vera voce dell'indagato, ma aventi contenuto appositamente artefatto…

campo per ovviare a questi rischi. La risposta pare implicita nelle stesse caratteristiche tecniche delle metodiche che abbiamo sopra descritto: i sensi e l'esperienza umana, cui sempre ci si è affidati, non sembrano in grado di individuarli e isolarli, neutralizzandone, soprattutto quando si tratti di elementi audio-video, l'estremo impatto evocativo. Nemmeno lo strumento della inutilizzabilità, che scattasse a seguito di verifica di artificiosità dell'immagine, potrebbe salvaguardare il processo decisionale del giudice[23] dai persistenti effetti evocativi di un'immagine che, pur non potendo essere espressamente menzionata in motivazione, difficilmente potrà essere isolata ed esclusa dalla formazione del convincimento.

Ciò posto, la sfida che si intende cogliere non si può affrontare con una mera rivisitazione delle regole probatorie tradizionali, ma impone di immaginare nuovi strumenti di reazione, capaci di individuare ed isolare il *deepfake*, per impedire che produca effetti anche indirettamente ingannevoli. Insomma, occorre uno sbarramento efficace, che possa neutralizzare il rischio di accesso dell'immagine *fake* nel processo.

3. Lo scenario normativo: dall'*AI Act*…

A fronte di tale obiettivo, pare naturale rivolgere lo sguardo alla più recente legislazione, per accertare se possa offrire soluzioni adeguate alla sfida sopra delineata. È ben nota, del resto, l'adozione, sul piano dell'Unione europea, dello storico regola-

[23] Anche qui il discorso è facilmente riconducibile al paradigma generale della discussione sulla funzione della inutilizzabilità, che qui non può certamente essere approfondito come meriterebbe (sia consentito rimandare soltanto ai lavori monografici più noti, per non ampliare troppo i riferimenti a un tema qui secondario: N. Galantini, *L'inutilizzabilità della prova nel processo penale*, Giuffré, Milano, 1992; A. Scella, *Prove penali e inutilizzabilità*, Giappichelli, Torino, 2000; F.R. Dinacci, *L'inutilizzabilità nel processo penale*, Giuffré, Milano, 2008). In ogni caso, infatti, la specifica natura di questa invalidità solamente impedisce al giudice di fare riferimento in motivazione alla prova inutilizzabile, senza potersi escludere che questa abbia inconsciamente influenzato il processo decisionale dell'autorità giurisdizionale: ritorna a giocare un ruolo determinante nel discorso che sia sta articolando il forte impatto che può essere esercitato sull'inconscio umano da un'immagine che non si ha alcun motivo di sospettare di non genuinità…

mento 2024/1689, in materia di intelligenza artificiale. Frutto di una scelta di campo molto netta del legislatore europeo, il documento rappresenta un *unicum* nello scenario globale, testimoniando la volontà dell'Unione europea di governare il fenomeno dell'evoluzione digitale, in altri contesti geografici lasciato in balia delle leggi di forza del commercio e della produzione industriale, ai fini di farne uno strumento di crescita ed evoluzione per tutto il sistema-UE.

Il testo, per il quale si rimanda ad altro contributo[24], rappresenta oggi la fonte di riferimento in materia di governance della IA, in primo luogo, per via di un'ampia categoria di definizioni, tra le quali anche quella di *deepfake*. L'art. 3 del regolamento, infatti, al n. 60 definisce "*deep fake*": un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona.

A sua volta, l'art. 50 del medesimo testo stabilisce a carico dei *deployers* di sistemi che generino o manipolino immagini, contenuti audio o video, l'obbligo di rendere nota la natura di *deepfake*, salvo l'utilizzo di tali contenuti sia consentito per le finalità di accertare, prevenire, indagare o perseguire reati. Come è noto, però, per via della base legale su cui poggia il regolamento, ovvero l'art. 114 TFUE – che si occupa di accesso al mercato unico europeo – gli obblighi stabiliti si rivolgono agli operatori del mercato e non ai singoli individui che, come già sopra indicato, hanno invece ampio accesso ai *software* che generano o manipolano contenuti multimediali. Si rinviene solo uno specifico riferimento alla pubblicazione di contenuti multimediali sintetici nel considerando n. 134, con riguardo, appunto alla informazione degli utenti da parte (parrebbe) di chi svolge attività di natura editoriale (a livello professionale?).

Ovviamente, il regolamento non si è spinto a dettare specifiche regole di carattere probatorio, né, in fondo, avrebbe potuto. Non solo, infatti, come ricordato, la base legale di per sé non lo

[24] Volendo, per uno sguardo di insieme, S. Quattrocolo, "Intelligenza artificiale e processo penale: le novità dell'AI Act", in *Diritto di Difesa*, 16 gennaio 2025. Si segnala, tuttavia, che alcuni articoli del Regolamento, non ancora completamente entrato in vigore, sono oggetto di proposta di emendamento nel c.d. "pacchetto Omnibus" presentato dalla Commissione europea.

consentirebbe, ma non si rinvengono nemmeno richiami espliciti (come accade, invece, con riguardo all'art. 16 TFEU, in tema di riservatezza), all'art. 82.2 TFUE, ove, peraltro, la competenza dell'Unione a emanare direttive ai fini del ravvicinamento delle legislazioni nazionali in materia probatoria è comunque limitata al profilo della circolazione transfrontaliera della medesima.

Nonostante queste premesse, la sfera della giustizia penale è frequentemente evocata dal regolamento[25]: talvolta per stabilire delle eccezioni ai divieti assoluti di utilizzo di alcune applicazioni di IA (art. 5); talaltra, e in linea generale, per prevedere che le applicazioni digitali utilizzabili nel contesto delle indagini e del processo penale siano sempre considerate forme di IA "ad alto rischio" e dunque sottoposte al meccanismo di *compliance* previsto dal regolamento stesso. Nessuno spunto, tuttavia, appare utile ad affrontare lo specifico tema qui all'attenzione, ma non mancano recentissime iniziative normative interne cui rivolgere lo sguardo.

3.1. …alla legge italiana 23.9.2025 n. 132

Come noto, la legge 23 settembre 2025, n. 132 ha introdotto – oltre ad una delega al Governo in materia di dati, algoritmi e sviluppo di metodi matematici per l'addestramento di sistemi di intelligenza artificiale (art. 16) – un'ampia serie di disposizioni di settore, che seguono l'elencazione di alcuni principi generali nell'approccio all'impiego dell'intelligenza artificiale[26]. Nella parte dedicata ai principi si ribadisce che per tutto quanto non espressamente definito dalla legge, si applicano le definizioni contenute nell'art. 3 del reg. 2024/1689 UE, e dunque anche per quanto riguarda il concetto di *deepfake*. Nessuna indicazione specifica si ritrova nell'articolo dedicato all'uso dell'intelligenza artificiale nell'attività giudiziaria (art. 15), ove risulta regola-

[25] S. Quattrocolo, "Criminal Law Enforcement through AI", in W. Barfield, U. Pagallo, *Research Handbook on the Law of Artificial Intelligence*, II ed., Edgar Elgar, Oxon, 2025, p. 349 ss.

[26] Per un commento generale, v. B. Galgani, "La l. 132/2025: 'testata d'angolo' o 'pietra d'inciampo' nella definizione dei rapporti tra amministrazione della giustizia (penale) e intelligenza artificiale?", in *Diritto penale processo*, (2025), 1444 ss.

to l'impiego di soluzioni specificamente volte ad essere utilizzate per finalità proprie dell'amministrazione della giustizia, ben diverse da quelle sottostanti ai GANs sopra richiamati.

Tuttavia, il testo contiene un riferimento indiretto ai *deepfake*, nel nuovo art. 612-*quater* c.p., rubricato "*Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale*"[27], ove si puniscono alcune condotte – forse non strettamente tipizzate (come si vedrà) – legate alla circolazione di contenuti multimediali (anche immagini, ovviamente), falsificate o alterate con l'impiego di IA e tali da indurre in inganno circa la loro genuinità. La disposizione, formulata come fattispecie di danno, prevede la pena della reclusione da uno a cinque anni, indicando altresì una serie di circostanze in cui la procedibilità è prevista d'ufficio.

Innanzitutto, occorre discutere se la disposizione appena introdotta attribuisca al concetto di *deepfake* un valore normativo[28]: in primo luogo, si sottolinea come la nuova fattispecie si riferisca a "immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei ad indurre in inganno sulla loro genuinità". Da quanto sopra rappresen-

[27] Art. 612-*quater*. – *(Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale)* – "Chiunque cagiona un danno ingiusto ad una persona, cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità, è punito con la reclusione da uno a cinque anni.

Il delitto è punibile a querela della per-sona offesa. Si procede tuttavia d'ufficio se il fatto è connesso con altro delitto per il quale si deve procedere d'ufficio ovvero se è commesso nei confronti di persona incapace, per età o per infermità, o di una pubblica autorità a causa delle funzioni esercitate". Si deve precisare che la formulazione in parte riprende un disegno di legge già presentato nel 2021, ben più delimitato, avente ad oggetto la manipolazione artificiale di immagini di persone reali allo scopo di ottenerne rappresentazioni nude. Si tratta di C 2986, del 21 luglio 2021, di iniziativa parlamentare, primo firmatario Di Carlo; su tale testo si vedano le osservazioni del Prof. Sergio Seminara, disponibili sul sito della Camera dei deputati (https://www.camera.it/application/xmanager/projects/leg18/attachments/upload_file_doc_acquisiti/pdfs/000/007/077/MEMORIA_SEMINARA.pdf).

[28] Non mancano tuttavia, nel web, commenti che catalogano l'iniziativa come 'legislazione antifake', come può riscontrarsi attraverso qualsiasi motore di ricerca.

tato, però, i contenuti multimediali correntemente intesi come *deepfake* non necessariamente rientrano in tale tassonomia, poiché non costituiscono "falsificazioni" ma sono "semplicemente sintetici": l'*output* generato, per esempio, da un GAN non è la falsificazione di una immagine, un suono, un video veri, ma un prodotto creato a partire da sterminati database, con un processo di *feedback* continuo tra generatore e discriminatore che, di fatto, è del tutto inedito. Al contrario, "falsificazione" e "alterazione" implicano, teoricamente, la riconoscibilità di un nucleo autentico, successivamente manipolato. Vero è che la fattispecie di danno appena introdotta implica che il contenuto multimediale sia riconducibile all'immagine e/o alla voce di persone fisiche esistenti, vittime del "danno ingiusto", che è elemento costitutivo della fattispecie[29]: proprio in ragione di tale osservazione, però, pare restrittivo, al di là del necessario, il riferimento al solo impiego di sistemi di intelligenza artificiale.

È evidente, infatti, che ci troviamo in quell'area sopra descritta (per finalità probatorie), come "intermedia", tra la generazione sintetica e la manipolazione, anche grossolana, del contenuto, in cui possono essere impiegate metodiche computazionali varie. Se è vero che il cuore della tutela che la norma intende offrire è costituito dal contrasto alla diffusione di immagini o suoni alterati con metodi che inducano in inganno circa la loro genuinità, non pare però ragionevole distinguere tra tecniche che ricorrono a soluzioni di intelligenza artificiale e tecniche computazionali eventualmente diverse: è incontestabile che i *software* più diffusi allo scopo utilizzino soluzioni di intelligenza artificiale – secondo la definizione che la stessa l. 132/2025 ne dà, attraverso il rinvio al regolamento europeo – ma altrettanto evidente è che lo stesso disvalore si realizzerebbe se fossero impiegate diverse soluzioni, altresì in grado di produrre risultati ingannevoli rispetto alla propria artificiosità, ma tecnicamente non riconducibili alla definizione di intelligenza artificiale.

[29] Fa notare B. Fragasso, "Profili penalistici dell'intelligenza artificiale: osservazioni a prima lettura", in *Sist. Pen. riv. trim*., (2025), n. 10, p. 164, che l'inserimento dell'evento di danno, quale elemento del reato, costituisce elemento ultroneo rispetto alla direttiva 2024/1385, che richiede l'incriminazione della diffusione delle immagini manipolate sessualmente esplicite.

Ciò che è certo, tuttavia, è che la nuova disposizione ha criminalizzato un nucleo di condotte non necessariamente legate alla generazione del contenuto artefatto, quanto piuttosto alla sua circolazione, in un quadro complessivo non ideale dal punto di vista dalla tipicità delle condotte. Alcune considerazioni a prima lettura. In primo luogo, ogni riflessione muove dall'ovvia inopportunità di criminalizzare la generazione di contenuti multimediali sintetici, opzione discutibile non solo sotto il profilo del bilanciamento con altri principi generali – quali la libertà di espressione[30] e di ricerca scientifica –, ma certamente anche dal punto di vista della corretta allocazione di una ipotetica responsabilità penale per contenuti generati automaticamente, solo talvolta riconducibili ad uno specifico impulso umano[31]. Tuttavia, la scelta di sanzionare la diffusione di tali contenuti non risolve tutti i possibili dubbi sui requisiti di tipicità della norma: infatti, non può sfuggire che, per le ragioni tecniche sopra rappresentare, chi realizzi le condotte indicate nell'art. 612-*quater* c.p. potrebbe non aver alcun modo di sapere che i contenuti multimediali in questione sono stati proditoriamente generati, con evidenti ricadute sul profilo dell'accertamento dell'elemento soggettivo… Né tale scenario può ritenersi scongiurato dalle richiamate norme del regolamento europeo, che gravano solo alcuni soggetti, con ruoli professionali, dell'obbligo di rendere nota la natura sintetica del contenuto multimediale.

Sul piano più sostanziale della struttura della fattispecie – senza pretese di approfondimento, data la scarsa competenza di chi scrive – si osserva, innanzitutto, che le condotte di "cessione", "pubblicazione" o diffusione con ogni mezzo (par di intendere) di immagini/suoni alterati rilevano, a questi fini, solo se realizzate in assenza del consenso della persona interessata, cui spetta il potere di querela, salvi i casi di connessione con altri

[30] Si ricorda proprio il considerando n. 134 del Reg. 2024/1689 UE, ove, in relazione al concetto di *deepfake*, si sottolinea l'importanza di tutelare proprio questi profili, con riguardo al diritto alla libertà di espressione e alla libertà delle arti e delle scienze, garantiti dalla Carta fondamentale dei diritti dell'Unione europea.

[31] Il tema, di straordinaria complessità, è affrontato nel recentissimo lavoro di B. Fragasso, *Intelligenza artificiale e responsabilità penale*, Giappichelli, Torino, 2025.

reati, a loro volta procedibili d'ufficio, e quelli in cui la persona offesa sia soggetto incapace "per età o infermità", oppure sia una pubblica autorità. Sotto il profilo della procedibilità, ampie paiono le deroghe alla querela, non solo per effetto della connessione con altro reato procedibile d'ufficio (la connessione più frequente potrebbe verificarsi con fatti rilevanti *ex* art. 615-*ter* c.p., ovvero l'accesso abusivo a sistemi informatici o telematici, la cui procedibilità, però, è sottoposta a querela, salvi i casi specifici del secondo comma), ma anche con riguardo alla pubblica autorità, che non dovrà attivarsi per richiedere l'esercizio dell'azione penale. La procedibilità d'ufficio in relazione ai "soggetti deboli" richiede qualche riflessione. Per un verso, infatti, essa sembra doversi intendere come sganciata dal concetto di incapacità di agire, essendo espressamente estesa al minorenne e ai soggetti affetti da "infermità": tale generico richiamo ad una situazione patologica sembra prescindere da un procedimento formale di esclusione della capacità di agire della persona, riecheggiando piuttosto l'incapacità dell'art. 70 c.p.p., relativa alla consapevole comprensione degli atti che devono svolgersi nel processo. Per altro verso, la disposizione potrebbe leggersi, invece, in senso formale, volendo escludere la necessità di presentazione della querela soltanto da parte dei soggetti che siano formalmente "incapaci" per ragioni di età o per effetto di una condizione patologica. Pertanto, l'interdetto per ragioni non patologiche, ma legali (dunque *ex* art. 32 c.p.), dovrebbe mantenere il diritto di querela.

Quanto alla tipicità della condotta, se le ipotesi di cessione e pubblicazione paiono più facilmente connotabili, il riferimento alla diffusione richiederà ampia riflessione rispetto alla necessaria estensione del *parterre* dei destinatari (anche in relazione alla piattaforma digitale utilizzata per tale diffusione, con particolare riguardo ai meccanismi di "rilancio" dei contenuti, propri di molti *social network*), alle possibili situazioni di diffusione involontaria, conseguita, ad esempio, all'azione di un virus informatico o ad altro tipo di "intrusioni" digitali. Quanto all'evento, data l'indeterminatezza della descrizione, si ritiene che il danno ingiusto non debba necessariamente avere un contenuto patri-

moniale, come altresì suggerito dalla collocazione topografica della nuova fattispecie[32].

Sul profilo sanzionatorio, non può non notarsi la severità della previsione edittale che va da uno a cinque anni di reclusione, consentendo così l'applicazione di tutte le misure cautelari coercitive, anche la custodia cautelare in carcere…

Un'ulteriore fattispecie di reato è introdotta, sempre dall'art. 26 l. 132/2025, nel corpo dell'art. 171 l. 633/1941, sulla protezione del diritto d'autore, ove si prevede adesso la criminalizzazione della condotta di riproduzione o estrazione di testo da opere o altri materiali disponibili in rete o in banche dati, attraverso sistemi di intelligenza artificiale, in violazione degli artt. 70-*ter* e 70-*quater* dello stesso testo di legge[33].

A tali inedite fattispecie si accompagnano, poi, alcune circostanze aggravanti, derivanti dall'aver realizzato condotte già criminalizzate, attraverso il ricorso a sistemi di intelligenza artificiale. La prima aggravante è generale: è stato infatti introdotto l'art. 61 n. 11-*undecies*[34], che determina l'aumento di un terzo della pena nell'ipotesi in cui il fatto sia stato commesso con il ricorso a sistemi di intelligenza artificiale, in forma di mezzo insidioso o di ostacolo alla pubblica o privata difesa o, ancora, di strumento che abbia aggravato le conseguenze del reato.

Segue poi una serie di ipotesi specifiche, prima tra le quali quella dell'art. 294 c.p.[35], in relazione alla commissione del reato di attentato contro i diritti politici del cittadino, avvalendosi di sistemi di intelligenza artificiale. Ed ancora, le nuove aggravanti che portano la pena, rispettivamente, per l'aggiotaggio (art. 2637 c.c.) alla forbice edittale da due a sette anni, per i fatti commessi con l'impiego di sistemi di intelligenza artificiale, e

[32] B. Fragasso, "Profili penalistici", cit., p. 163.

[33] *Ivi*, p. 166.

[34] "11-*decies*) l'avere commesso il fatto mediante l'impiego di sistemi di intelligenza artificiale, quando gli stessi, per la loro natura o per le modalità di utilizzo, abbiano costituito mezzo insidioso, ovvero quando il loro impiego abbia comunque ostacolato la pubblica o la privata difesa, ovvero aggravato le conseguenze del reato". Segnala la presenza di due 11-*decies* nell'art. 61 c.p., B. Fragasso, *op. ult. cit.*, p. 167.

[35] "La pena è della reclusione da due a sei anni se l'inganno è posto in essere mediante l'impiego di sistemi di intelligenza artificiale".

da due a sette anni oltre alla multa da Euro venticinquemila a sei milioni, per la manipolazione del mercato (art. 185 TUIF).

Come anticipato, nel testo di legge non risultano interventi sul profilo probatorio, per affrontare il rischio specifico di introduzione nel processo di elementi conoscitivi non genuini e, dunque, privi di idoneità dimostrativa. Pertanto, al momento, la questione finisce confinata sul piano giurisprudenziale, ove gli spunti, invero, non sono ricchissimi.

4. Trend giurisprudenziali

Infatti, la giurisprudenza pare in linea di massima orientata verso l'approccio tradizionale di applicazione delle consuete categorie di prova tipiche, tra le quali, come accennato, il documento si presta teoricamente ad accogliere tutti i possibili contenuti multimediali[36] che potrebbero essere generati automaticamente. Insomma, il documento, di cui agli artt. 234 ss. c.p.p., si profila come un "grande contenitore di tipicità"[37] che permette di veicolare nel processo testi, immagini, voci, video, gravando poi l'istruttoria dibattimentale e la fase della valutazione di ogni dubbio circa i possibili limiti dimostrativi che quei "documenti", ammessi ed acquisiti, potrebbero comunque nascondere.

Nel quadro della giurisprudenza di legittimità, il tema dell'applicazione dell'art. 234 c.p.p. ai contenuti multimediali si è primariamente presentato sotto il profilo della ammissione/valutazione degli *screenshot* di conversazioni su vari *social networks*, contenute nei più diversi *devices* digitali, ma principalmente negli *smartphones*[38]. La rassegna è ampia e in calce si

[36] Quanto alle caratteristiche intrinseche del mezzo di prova, si osserva appunto che il documento è in grado di veicolare nel processo i più svariati contenuti, spesso profilandosi come alternativo o parallelo a prove dichiarative sul medesimo contenuto: F. Zacchè, *La prova documentale*, Giuffré, Milano, 2012, p. 75. V. anche N. Rombi, "La prova documentale", in P. Ferrua, E. Marzaduri, G. Spangher (a cura di), *op. cit.*, pp. 575 ss.

[37] In questo senso, S. Campanella, *Il documento dichiarativo nella logica della prova penale*, Giappichelli, Torino, 2021, p. 4.

[38] Considerevole spazio trova nel dibattito in corso, in cui si è inserita anche la Corte costituzionale (C. cost. 170/2023), la distinzione tra natura documentale o comunicativa delle conversazioni di cui si dispone l'acquisizione. In particolare, il Giudice delle leggi ha accertato che lo scambio di

trovano alcuni riferimenti di massima[39]. L'attenzione, dunque, al

messaggi elettronici, *lato sensu* intesi, rappresenta una forma di corrispondenza, anche nel caso si tratti di messaggi già ricevuti e letti, salvo il decorso del tempo li privi di qualsiasi profilo di attualità, trasformandoli, in documenti storici (a commento, v., tra gli altri, R. Orlandi, "'Corrispondenza' dei parlamentari e limiti all'accertamento penale nella sentenza n. 170 del 2023 della Corte costituzionale", in *DisCrimen* 26 ottobre 2023; F. Cerqua, L. Lupària Donati, "La versione della Consulta sulla corrispondenza elettronica. Un 'bouleversement' in materia di prova digitale?", in *Diritto dell'Informazione e dell'informatica*, (2023), n. 5, pp. 718 ss.). Sempre in via incidentale rispetto al tema, qui collaterale, della idoneità dimostrativa dei flussi di comunicazioni su canali digitali, si inserisce il tema dei c.d. 'criptofonini', con particolare riferimento all'acquisizione di conversazioni già decriptate da autorità straniere e acquisite tramite ordine europeo di indagine. Questione, com'è noto, definita dalle Sezioni Unite della Corte di cassazione con la sentenza del 29.2.2024 n. 23755, attraverso l'indicazione che ai suddetti materiali probatori debba applicarsi il regime ordinario di circolazione della prova tra procedimenti penali. Da ultimo, sul punto, v. Cass. Pen. I, 29.5.2025, n. 24404; Cass. III, 27.3.2025, n. 18136.

[39] Cass. V, 16188 del 2022: "considerato che è pacifica la legittimità dell'acquisizione come documento di messaggi sms - o tramite whatsapp - mediante la realizzazione di una fotografia dello schermo ('screenshot') di un telefono cellulare sul quale gli stessi sono leggibili (in tali termini, tra tante, Sez. 5, n. 12062 del 05/02/2021, dep. il 30/03/2021, Di Calogero, Rv. 280758 - 01, Sez. 3, n. 8332 del 06/11/2019, dep. il 02/03/2020, R., Rv. 278635 - 01), e che il c.d. screenshot consiste in una fotografia istantanea che rappresenta la schermata visibile del dispositivo elettronico (cellulare, tablet, pc) nel momento in cui è tratta e che, pertanto, costituisce la mera constatazione di quel fatto storico ivi ripreso". Ancora: Cass. VI, 38678/2023: "I messaggi whatsapp e gli sms conservati nella memoria di un telefono cellulare hanno natura di documenti ex art. 234 cod. proc. pen., sicché è legittima la loro acquisizione mediante mera riproduzione fotografica, non applicandosi né la disciplina delle intercettazioni, né quella relativa all'acquisizione di corrispondenza ex art. 254 cod. proc. pen.: per la concreta utilizzabilità della trascrizione delle conversazioni via whatsapp, la necessità di acquisire il supporto telematico o figurativo contenente la relativa registrazione deve essere valutata in concreto". Similmente, anche Cass. VI 43992/2023; Cass. V, 2658/2021. Maggiore attenzione ai profili 'classici' di genuinità nella estrazione e nella conservazione del dato, Cass. VI, 11.9.2024, n. 39548, dove si fa riferimento alla necessità di eseguire, ai fini dell'acquisizione probatoria, una copia forense del contenuto memorizzato nel telefono cellulare del ricorrente, "con tutte le dovute garanzie anche in termini di integrità delle comunicazioni acquisite". Con specifico riguardo allo *screenshot* di una pagina di social network, Cass. VI, 15.1.2025, n. 5400, che ribadisce la piena legittimità dell'acquisizione ex art. 234 c.p.p., in assenza di alcun adempimento specifico previsto dalla legge

momento, è confinata ad un aspetto specifico, pur ricco di molte problematiche tecniche[40], e non sembra orientata a fronteggiare il tema dei *deepfake*.

4.1. Una recente sentenza della II sezione della Corte di cassazione

Tuttavia, la Corte suprema di cassazione[41] si è recentemente occupata proprio di problematiche relative alla genuinità di immagini video, in un caso in cui i giudici di merito avevano motivato la condanna dell'imputato ricorrente, nel giudizio abbreviato, anche sulla base delle riprese videoregistrate dall'impianto di un esercizio commerciale adiacente al luogo dei fatti, contenute in un supporto *cd-rom*. Nel ricorso per cassazione, la difesa avanzava (per la prima volta, par di intendere) dubbi sull'autenticità di tali immagini e, in assenza di un riscontro tecnico, sulla effettiva valenza dimostrativa dei fatti accertati dal giudice (targa dell'autoveicolo dell'imputato rilevata nel luogo del reato). Nel dichiarare il motivo formulato dalla difesa del tutto infondato, la seconda sezione penale della Corte di cassazione ha ribadito – sulla scia dell'ormai risalente "dottrina Prisco"[42] –

per tale attività, "non essendovi alcuna differenza tra tale fotografia [*screenshot*, ndr] e quella di qualsiasi altro oggetto", con evidente spostamento di baricentro dell'attenzione dal contenuto (pagina social network), plausibilmente discutibile sul piano della idoneità dimostrativa, al mezzo impiegato per rappresentarlo.

[40] V., ad esempio, S. Quattrocolo, C. Anglano, M. Canonico, M. Guazzone, "Technical Solutions for Legal Challenges: Equality of Arms in Criminal Proceedings", in *Global Jurist*, 20 (2020), n. 1.

[41] Cass. Pen. II, 7.3.2025, n. 16355/2025, nella banca dati pubblica della Corte di cassazione.

[42] Il riferimento è a Cass. SU, 28.3.2006, n. 26795, secondo la quale, le videoregistrazioni eseguite in luoghi pubblici o aperti al pubblico, non effettuate nell'ambito di un procedimento penale, devono essere incluse nella categoria delle prove documentali. L'evoluzione della 'dottrina Prisco', anche in relazione alle videoriprese effettuate, invece, in luoghi di provata dimora, è recentemente ricostruita in Cass. III, 13.2.2025, n. 12522. In giurisprudenza si registra l'estensione di tale orientamento anche alle videoregistrazioni eseguite in luogo pubblico o aperto al pubblico, nell'ambito di un procedimento penale, sebbene sfornite di specifica autorizzazione dell'autorità giudiziaria (v. da ultimo, Cass. II, 3.4.2025, n. 19707, § 6 della motivazione in diritto).

che, secondo consolidata giurisprudenza, le videoregistrazioni effettuate da privati tramite impianti di sicurezza "sono prove documentali rappresentative, *ex* art 234 c.p.p., sicché per la loro utilizzazione nel giudizio non è necessario procedere alla diretta visione nel contraddittorio delle parti, alle quali è garantito il diritto di prenderne visione ed estrarne copia". Nel successivo passaggio della pronuncia – riferito all'integrità e all'assenza di alterazioni delle immagini contenute nel *cd rom* – il giudice di legittimità ha validato la piena sufficienza dell'attestazione da parte del giudicante, contenuta nella sentenza impugnata, circa la corrispondenza della registrazione video utilizzata in giudizio con la ripresa della videocamera installata presso il ricordato esercizio commerciale. Senza poter comprendere appieno la vicenda processuale, di cui emergono soltanto alcuni tratti dalla lettura della sentenza qui all'attenzione, deve sottolinearsi, alla luce di quanto sopra illustrato, che il tema della genuinità di un video, in tempi di GANs, difficilmente può ritenersi esaurito dall'attestazione, da parte del giudice del merito, dell'assenza di alterazioni materiali e di corrispondenza con quanto asserito dalla parte che ha introdotto la prova (senza, peraltro, una specifica distinzione tra integrità del supporto digitale in cui è raccolta e autenticità dell'immagine in questione…). Certo, nel caso di specie, è bene sottolineare che la qualità e la tipologia dei file, peraltro contenuti in un supporto digitale obsoleto, potrebbero al massimo suggerire interventi manipolativi assai meno sofisticati di quelli attualmente possibili. Tuttavia, anche rispetto a immagini digitali di qualità inferiore, il tema dell'autenticità difficilmente può essere affrontato "a occhio nudo". Quanto poi alla questione relativa alla idoneità dimostrativa, la petizione di principio risulta altrettanto evidente, poiché, afferma la sen-

Va altresì segnalato che, in relazione alle riprese effettuate al di fuori del procedimento penale, non solo la giurisprudenza ritiene la piena ammissibilità ex art. 234 c.p.p., ma anche ammissibile la prova testimoniale degli operati di polizia circa il loro contenuto rappresentativo (v. già Cass. II, 30.11.2016, n. 10 e, da ultimo, Cass. V, 26.3.2025, n. 15848. Anche con specifico riguardo, poi, alle videoregistrazioni effettuate dalla persona offesa o da altri soggetti privati, estranei al procedimento, risulta consolidata l'applicabilità dell'art. 234 c.p.p., con i soliti limiti di ammissibilità ivi previsti: di recente, Cass. V, 4.2.2025, n. 12280.

tenza, essa deve ritenersi confermata in ragione del fatto che il video mostra chiaramente la targa del veicolo in questione: ai nostri fini pare quasi superfluo sottolineare che il fatto che una immagine digitale mostri un oggetto, una persona o un fatto, non significa che questi siano effettivamente esistenti e collocabili nei luoghi riportati nell'immagine.

Cruciale si rivela il passaggio in cui la Corte afferma che, sebbene la difesa avesse potuto prendere visione ed estrarre copia del video, le deduzioni del ricorrente sono del tutto generiche, cioè prive di qualsiasi riferimento utile a mettere in dubbio sia l'autenticità, sia la idoneità dimostrativa delle immagini. Pertanto, "in assenza di alcuna norma che imponga l'effettuazione di una perizia sulle videoregistrazioni, e anche di alcuna specifica ragione di opportunità di disporre, nella specie, tale discrezionale mezzo di prova tecnico", le doglianze del ricorrente sul punto sono rigettate.

Allo stato attuale delle cose, come sottolineato con una certa ironia dall'estensore, il quadro normativo offre soltanto quella via, ovvero la contestazione argomentativa, supportata dal proprio consulente tecnico, ad opera della parte interessata ad evitare l'ammissione o a escludere la valutazione di immagini già ammesse, della cui autenticità si dubita… Nel caso di specie, si osserva, tale onere non è stato adeguamento adempiuto dalla parte interessata.

Le conclusioni cui giunge il Supremo Collegio in questo recente caso, però, tornano utili per provare ad immaginare alcuni spunti di intervento normativo, che evitino, però, di far ricadere integralmente sulle spalle delle parti, soprattutto quelle private, la tutela dell'interesse generale a che nel processo non siano introdotti mezzi di prova non genuini.

5. Qualche proposta per la discussione

Ricostruito così il quadro normativo e giurisprudenziale che, al momento, fa da sfondo al tema che ci occupa, pare necessario muovere oltre questi angusti profili, per provare a ipotizzare specifiche linee di intervento. Due sono gli elementi che appaiono cruciali nell'immaginare un modello alternativo di ammissione

delle prove che veicolano contenuti multimediali e, dunque, per prassi riconducibili alla prova documentale.

In primo luogo, si deve insistere sulla necessità di non affidarsi ad un discrimine meramente a-tecnico: la complessità che contraddistingue le situazioni qui considerate non può essere efficacemente affrontata solo sulla base delle abituali categorie normative. L'assenza di genuinità dell'immagine, della voce, del video, del testo, può essere determinata con certezza soltanto attraverso una valutazione tecnico-computazionale, spesso basata, come sopra accennato, sulle metodiche delle stesse architetture generative che potrebbero aver creato l'*output* in questione. È imprescindibile, dunque, prevedere un momento di filtro tecnico sulla genesi e sulla capacità dimostrativa del contenuto multimediale di cui si richiede l'ammissione.

Certo, questa affermazione reca con sé conseguenze tutt'altro che secondarie. In primo luogo, si verifica un arretramento del ruolo del giudice nello svolgimento di una funzione, quella della ammissione probatoria, squisitamente giurisdizionale[43]. In secondo luogo, il necessario ricorso all'esperto può determinare elevati oneri per le parti, soprattutto quelle private. Tuttavia, la soluzione già altrove proposta[44], pare l'unica in grado di rispondere alle peculiarità del problema affrontato.

L'elemento centrale di un possibile nuovo approccio è rappresentato dal meccanismo di nomina di un perito, a fronte dell'avvenuta richiesta, ad opera delle parti, dell'ammissione come prova di contenuti multimediali. In presenza di incarico peritale, le parti potrebbero nominare i loro consulenti tecnici: pur non necessaria, la consulenza potrebbe apparire un'importante scelta strategica, seppur destinata ad aggravare ulteriormente i costi che le parti private devono affrontare per poter sostenere

[43] Su questo tema, da sempre oggetto di attenzioni con riguardo a tutti i profili della giurisdizione v. M. Taruffo, *La prova dei fatti giuridici*, Giuffré, Milano, 1992, pp. 307 ss.; con riguardo al processo penale, di recente P. Tonini, M. Cecchi, "Nozione e definizione di prova scientifica", in C. Conti, A. Marandola, (a cura di), *op. cit.*, pp. 32 ss.; L. Lupària, "Prova, scienza e processo penale", in G. Canzio, L. Luparia (a cura di), *Prova scientifica e processo penale*, Giuffré, Milano, 2022, p. XIX.

[44] V. ancora S. Quattrocolo, "Deepfake e fine della tipicità probatoria?", cit., pp. 37 ss.

efficacemente il processo. Tuttavia, l'opzione proposta non pare fungibile. Non solo il *deepfake* ha le caratteristiche, ingannevoli, più sopra descritte, ma, come più volte sottolineato in precedenza, tra l'immagine sintetica e l'immagine autentica si pongono numerose altre possibilità di manipolazione che incidono sulla tipologia di riscontro che deve essere valutato in sede ammissiva.

Anche manipolazioni meno sofisticate richiedono, per essere individuate, una valutazione informatico-forense: l'analisi dei semplici metadati potrebbe già fornire utili elementi in merito alla genuinità, senza rendere necessario il ricorso a metodiche più complesse; anche in questo caso, però, il sapere tecnico di un esperto non potrà essere sostituito dalla percezione sensoriale o da affermazioni apodittiche di autenticità, come quelle cui talvolta la giurisprudenza ricorre (come sopra segnalato). L'incarico di volta in volta conferito dovrà tenere in considerazione questi aspetti, modulando il quesito in maniera adeguata.

Se si tengono a mente i confini più sopra tracciati, meno stringente dovrebbe apparire un altro argomento critico, altresì immaginabile, rispetto alla ipotesi qui formulata, ovvero quello dell'ulteriore svuotamento del contraddittorio dibattimentale, a favore di forme basilari e anticipate di mero confronto tecnico-scientifico. In verità, la perizia ipotizzata dovrebbe consistere nel semplice accertamento della artificiosità o meno dell'elemento di prova richiesto. A fronte del riscontro di genuinità starà al dibattimento, nelle sue forme tradizionali, far emergere se, davvero, la prova dimostri qualcuno dei fatti rilevanti *ex* art. 187 c.p.p., primari o secondari, e con quale grado di persuasività. Da valutarsi, in sede di ammissione, sono solo la genesi – artificiale o meno – e la manipolazione della potenziale prova. E se, come si è concluso più sopra, è opportuno che la prova "sintetica", per le sue doti di elevato "mimetismo", non entri in contatto con il giudice del dibattimento, si potrebbero immaginare due possibili scenari, già in altre sedi proposti[45].

Una prima soluzione potrebbe consistere nell'anticipazione delle richieste di tutte le prove *lato sensu* documentali al mo-

[45] V. ancora S. Quattrocolo, "Deepfake e fine della tipicità", cit., pp. 37 ss.

mento conclusivo della udienza preliminare, dopo la lettura del decreto che dispone il giudizio e la formazione del fascicolo *ex* art. 431 c.p.p., oppure in conclusione dell'udienza predibattimentale, dopo la lettura dell'ordinanza di fissazione della nuova udienza, davanti ad altro giudice del tribunale, per i reati dell'art. 550 c.p.p. Si avrebbe così una netta cesura, che renderebbe impermeabile il dibattimento e il suo giudice al rischio di esposizione a *deepfake*, con il giudice dell'udienza preliminare, o predibattimentale, a fungere da autentico *gatekeeper*, sulla falsa riga dello schema probatorio del processo nord americano. Non si nasconde la certezza che tale proposta implichi la previa piena transizione ad una organizzazione concentrata dei ruoli di udienza, poiché il decorso di molti mesi, o talvolta anni, tra udienza preliminare e dibattimento priverebbe di utilità tale soluzione, poiché nelle more del giudizio si reperirebbero verosimilmente nuovi documenti digitali: pare difficile immaginare, a questo punto, un meccanismo che eviti di doverne chiedere l'ammissione proprio al giudice dibattimentale. Al contrario, la concentrazione dei ruoli garantirebbe un'autentica cesura tra ammissione e valutazione, alla luce della quale potrebbe anche riconsiderarsi, con riguardo a tutte le tipologie di prova, una anticipazione generale, delle richieste di prova, al momento finale dell'udienza preliminare o predibattimentale.

Una diversa opzione potrebbe consistere in una udienza presidenziale di ammissione dei documenti multimediali, di cui le parti potrebbero fare richiesta con specifica lista, da depositarsi con un termine meno dilatorio di quello dell'art. 468 c.p.p., per consentire adeguato spazio al "contraddittorio tecnico". In questa udienza, che si può immaginare sulla falsariga dell'art. 467 c.p.p., ma specificamente volta a valutare la genuinità dei documenti multimediali richiesti, il presidente nominerebbe il perito informatico, formulando un generale quesito circa la genesi del potenziale documento o la circa la sua eventuale manipolazione, assegnando un termine, alla scadenza del quale sarà fissata la data per la nuova udienza: in tale contesto, il presidente si pronuncerà sulla ammissibilità o meno del mezzo di prova, che sarà poi acquisito a dibattimento, discusso in merito ai profili di attendibilità dalle parti e valutato dal giudice. Nemmeno questa

ipotesi è priva di ricadute discutibili e alcuni aspetti meritano una breve riflessione.

In primo luogo, si può certamente dubitare della possibilità di immaginare un'ipotesi tassativa di perizia, a fronte dell'incertezza che discende dal tenore generale dell'art. 220 co. 1 c.p.p.: in dottrina, infatti, molti interpretano la formula normativa secondo parametri rigorosi – tali da aprire le porte alla perizia in ogni caso in cui il parere esperto appaia non solo necessario, ma anche soltanto utile[46] – mentre, invece, la giurisprudenza riconosce spazio discrezionale al giudice nella valutazione della necessità di disporre lo specifico mezzo di prova[47]. In una udienza presidenziale introdotta al precipuo scopo di valutare l'ammissibilità delle immagini potenzialmente *fake* o alterate, la necessità evocata dall'art. 220 co. 1 c.p.p. sarebbe automaticamente integrata, non solo per via della più volte richiamata difficoltà di accertamento della manipolazione, ma anche perché l'accertamento circa l'idoneità dimostrativa delle specifiche prove richieste sarebbe isolata – diversamente da quanto avviene ordinariamente nell'istruttoria – dal più ampio quadro valutativo di cui invece

[46] Invero, la disciplina generale dell'art. 220 c.p.p. non esclude mai uno spazio discrezionale del giudice, tanto di fronte alla richiesta delle parti, quanto in ipotesi di attivazione officiosa. Si ricorda che un approccio diverso (o, quantomeno, una formulazione diversa), era previsto dal codice previgente, ove si prevedeva l'attivazione della perizia quando fosse necessaria una indagine che richiedesse particolari cognizioni (v. R. Mendoza, G. Marcon, L. Marcon, *La perizia e la consulenza tecnica nel processo penale*, Cedam, Padova, 1994, p. 27; P.P. Rivello, "La perizia", in P. Ferrua, E. Marzaduri, G. Spangher (a cura di), *op. cit.*, pp. 406 s.). Tuttavia, anche in relazione all'art. 220 c.p.p., che utilizza la formula "occorre" si sostiene la natura necessaria della perizia, quando siano in gioco valutazioni di tipo, tecnico, scientifico o artistico (P. Tonini, *La prova penale*, II ed., Cedam, Padova, 1998, p. 84); anzi, si è osservato che la diversa formulazione linguistica "assottiglia la sfera di discrezionalità giurisdizionale, ponendo in luce che perno dell'accertamento tecnico è l'utilità (o anche solo l'opportunità) di attingere a conoscenze specialistiche": D. Curtotti, "La perizia", in A. Gaito (a cura di), *Trattato di diritto processuale penale*, III, Utet, Torino, 2008, p. 592; nello stesso senso P.P. Rivello, *op. cit.*, p. 407, il quale rileva che "non è invece censurabile il fatto che un magistrato escluda il ricorso alla perizia dove possa desumere il proprio convincimento da altri dati probatori").

[47] V. di recente, Cass. Pen. II, 7.3.2025,

dispone il giudice del dibattimento[48], che può giocare un ruolo non secondario nella scelta di disporre o meno una perizia.

Inoltre, data l'elevata frequenza con cui l'ammissione di contenuti multimediali è richiesta in qualsiasi tipo di procedimento, l'aggravio sull'ufficio del Presidente sarebbe significativo, con la necessità di un'attenta riflessione sul piano delle incompatibilità *ex* art. 34 c.p.p., che potrebbero mettere in difficoltà gli uffici più piccoli. Se, per un verso, sotto questo specifico profilo, la nomina del perito e l'eventuale, successiva, valutazione di ammissibilità della prova non paiono contenere la "forza pregiudicante" che caratterizza le decisioni che scatenano l'incompatibilità funzionale – come chiaramente confermato dal comma 2-*quater* dello stesso articolo 34 c.p.p., con riguardo al giudice dell'incidente probatorio – per altro verso, l'eventuale presenza nel collegio del presidente che ha ammesso la prova, finirebbe per minare la funzione di *gatekeeper* che si va cercando di realizzare.

Come già altrove riconosciuto, quelle qui condivise sono ipotesi di ingegneria processuale, la cui compatibilità con il sistema vigente deve essere attentamente vagliata. Eppure, per definizione, le sfide della modernità stimolano costantemente il legislatore a immaginare nuove soluzioni[49], che non possono che essere precedute da un articolato, talvolta aspro, dibattito. Si è scelto, allora, di sfruttare questo contesto per rilanciare un confronto che deve ancora raggiungere il necessario livello di profondità.

[48] In generale, su questo aspetto, M. Cecchi, "Il giudice dinanzi alla prova scientifica", in *Arch. pen. (web)*, (2022), n. 1, pp.1 s.

[49] G. Di Chiara, "Atipicità e sistemi probatori: linee per una fenomenologia generale", in V. Militello, A. Spena (a cura di), *Mobilità, sicurezza e nuove frontiere tecnologiche*, Giappichelli, Torino, 2018, p. 376.

Nicola Triggiani

L'utilizzo investigativo e probatorio dei sistemi automatici di riconoscimento facciale dopo l'*AI Act* e la l. n. 132/2025: dubbi e perplessità

Abstract: **The Investigative and Evidentiary Use of Automated Facial Recognition Systems after the AI Act and Law No. 132/2025: Doubts and Concerns.** The paper examines the issues related to the use of automated facial recognition systems in criminal proceedings in light of the regulations introduced by the AI Act and Law No. 132/2025. It highlights the extraordinary investigative potential of these tools, while at the same time drawing attention to the pitfalls and risks of abuses and violations of fundamental human rights.

Keywords: Automated Facial Recognition System, Investigative and Evidentiary Use, Pitfalls and Risks.

Sommario: 1. Premessa definitoria – 2. Il Sistema "S.A.R.I." (Sistema Automatico Riconoscimento Immagini) e gli interventi del Garante per la protezione dei dati personali – 3. L'approvazione dell'*AI Act* (Regolamento europeo 1689/2024/UE) e gli usi consentiti e vietati dei sistemi automatici di riconoscimento facciale – 4. La l. 23 settembre 2025, n. 132 e i tentativi di inquadramento giuridico del riconoscimento facciale automatizzato, tra tipicità e atipicità, nelle indagini preliminari e nel dibattimento, in attesa di una compiuta regolamentazione.

1. Premessa definitoria

Uno degli ambiti nei quali gli strumenti di intelligenza artificiale si stanno maggiormente affermando nel contesto del procedimento penale è, senza dubbio, la *digital evidence*: in particolare, nel più ampio quadro dei sistemi di identificazione biometrica[1], emergono i sistemi automatici di riconoscimen-

[1] I sistemi di identificazione biometrica, "attraverso l'analisi automatizzata delle caratteristiche fisiche (impronte digitali, volto, retina, iride), fisiologiche (ricavabili osservando le diverse funzioni corporee, cerebrali, battito cardiaco, respirazione) o comportamentali (ad es. scrittura, voce, cadenza dell'andatura), opportunamente convertite in dati digitali, permettono di comparare questi elementi distintivi al fine di identificare le persone": così L. Belvini, *Intelligenza artificiale e circuito investigativo*, Cacucci, Bari,

to facciale, che possono essere definiti come dei *software* che, attraverso processi informatici basati sull'impiego di algoritmi, consentono di stabilire – entro un certo margine di errore – la corrispondenza tra due immagini ritraenti il volto di una persona, consentendo così di identificare o verificare l'identità di un individuo in base alle sue caratteristiche fisionomiche[2].

2025, p. 160. In argomento, v. anche, *ex plurimis*, E. Sacchetto, "La prova biometrica", in C. Conti, A. Marandola (a cura di), *La prova scientifica*, Giuffrè Francis Lefebvre, Milano, 2023, pp. 243 ss. e la bibliografia ivi citata.

[2] Così G. Galluccio Mezio, "Tecnologie di riconoscimento facciale: una riflessione sul loro impiego con finalità investigative e probatorie", in *Cass. pen.* (2025), pp. 638 s.

La letteratura sul tema è ormai particolarmente ampia. Limitatamente a quella italiana, senza pretesa di completezza, cfr. G. Borgia, "Profili sistematici delle tecnologie di riconoscimento facciale automatizzato, anche alla luce dei futuribili sviluppi normativi sul fronte eurounitario", in *Leg. pen*., (2021), n. 4, pp. 206 ss.; M. Colacurci, "Riconoscimento facciale e rischi per i diritti fondamentali alla luce delle dinamiche di relazione tra poteri pubblici, imprese e cittadini", in *Sist. pen*., (2022), n. 9, pp. 23 ss.; E. Currao, "Il riconoscimento facciale e i diritti fondamentali: quale equilibrio?", in *Dir. pen. uomo*, (2021), n. 5, pp. 68 ss.; J. Della Torre, "Quale spazio per i *tools* di riconoscimento facciale nella giustizia penale?", in G. Di Paolo, L. Pressacco (a cura di), *Intelligenza artificiale e processo penale: indagini, prove, giudizio*, Università degli Studi di Trento, Trento, 2022, pp. 7 ss.; Id., "Tecnologie di riconoscimento facciale e procedimento penale", in *Riv. it. dir. proc. pen*., (2022), pp. 41 ss.; F. Demartis, *I sistemi automatici di riconoscimento facciale nel procedimento penale. Tra possibilità di impiego e limiti ordinamentali*, Wolters Kluwer Cedam, Milano, 2025; F. Di Matteo, "La riservatezza dei dati biometrici nello spazio europeo dei diritti fondamentali: sui limiti all'utilizzo delle tecnologie di riconoscimento facciale", in *Freedom, Security & Justice: European Legal Studies*, (2023), n. 1, pp. 74 ss.; G. Errico, "L'identificazione della persona nei cui confronti vengono svolte le indagini attraverso l'intelligenza artificiale (IA)", in *Il processo*, (2025), n. 1, pp. 371 ss.; R. Lopez, "La rappresentazione facciale tramite *software*", in A. Scalfati (a cura di), *Le indagini atipiche*, 2ª ed., Giappichelli, Torino, 2019, pp. 239 ss.; A. Mangiaracina, "Il riconoscimento facciale: nuove sfide nel processo penale", in P. Di Simone, A. Mangiaracina, L. Parlato (a cura di), *120 Anni di polizia scientifica: l'identificazione personale tra scienza e diritto, Atti del Convegno (Palermo, 18-19 aprile 2023)*, Palermo University Press, Palermo, 2024, pp. 77 ss.; A. Marandola, "Il riconoscimento facciale", in A. Marandola, C. Conti (a cura di), *op. cit.*, pp. 500 ss.; G. Mobilio, "L'uso delle tecnologie di riconoscimento facciale da parte delle forze dell'ordine: bandire o non bandire?", in E. Di Carpegna Brivio, A. Sancino (a cura di), *La democrazia nella società digitale. Tensioni e opportunità. Atti del Convegno*

I *facial recognition system* operano, più precisamente, mediante algoritmi capaci di rilevare le c.d. *faceprint*, le impronte facciali, ovvero un certo numero di caratteristiche principali di un volto, quali la posizione di occhi, naso, narici, zigomi, mento, orecchie, e, in questa maniera, sono in grado di realizzare un "modello biometrico", cioè un modello digitalizzato del volto finalizzato al riconoscimento[3], una sorta di "codice a barre" dell'individuo, che può essere utilizzato con finalità identificati-

3 dicembre 2021, Università di Milano-Bicocca, Giappichelli, Torino, 2023, pp. 1 ss.; G. Mobilio, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Editoriale scientifica, Napoli, 2021; P. Padova, "Il riconoscimento automatico del volto tra esigenze investigative e tutela della *privacy*", in P. Di Simone, A. Mangiaracina, L. Parlato (a cura di), *op. cit.*, pp. 121 ss.; C. Pallante, "Le nuove sfide della cooperazione investigativa nello spazio europeo", in L. Bernardini, C. De Caro (a cura di), *Sfide attuali e tendenze future del diritto processuale penale europeo*, Giappichelli, Torino, 2025, pp. 257 ss.; F. Paolucci, "Riconoscimento facciale e diritti fondamentali: è la sorveglianza un giusto prezzo da pagare?", in *Media Laws*, (2021), n. 1, pp. 204 ss.; L. Parlato, "Il riconoscimento facciale: vantaggi e insidie alla luce della giurisprudenza della Corte EDU", in P. Di Simone, A. Mangiaracina, L. Parlato (a cura di), *op. cit.*, pp. 95 ss.; E. Sacchetto, "*Face to face*: il complesso rapporto tra *automated facial recognition technology* e processo penale", in *Leg. pen*., 16 ottobre 2020, pp. 1 ss.; Ead., "La capacità dimostrativa di un *match* di riconoscimento facciale nel procedimento penale alla luce del Regolamento (UE)2024/1689", in L. Bernardini, C. De Caro (a cura di), *op. cit.*, pp. 311 ss.; Ead., *Tecnologie di riconoscimento facciale e procedimento penale. Indagine sui fondamenti e sui limiti dell'impiego della biometria moderna*, Giappichelli, Torino, 2025; L. Saponaro, "Le nuove frontiere tecnologiche dell'individuazione personale", in *Arch. pen. (web)*, (2022), n. 1, pp. 1 ss.; G. Tessitore, "Riconoscimento automatico del volto e confronto in ambito forense", in P. Di Simone, A. Mangiaracina, L. Parlato (a cura di), *op. cit.*, pp. 61 ss.; M. Torre, "Intelligenza artificiale e indagini penali: prospettive future e garanzie di sistema. Il sistema automatico di riconoscimento immagini", in A. Cadoppi, S. Canestrari, A. Manna, M. Papa (a cura di), *Cybercrime*, Wolters Kluwer, Milano, 2023, pp. 1731 ss.; Id., "Nuove tecnologie e trattamento dei dati personali nel processo penale", in *Dir. pen. proc*., (2021), spec. pp. 1047 ss.; R.V.O. Valli, "Sull'utilizzabilità processuale del Sari: il confronto automatizzato di volti rappresentati in immagini", in *il Penalista*, 16 gennaio 2019; V. Vasta, "Diritto dell'Unione europea e intelligenza artificiale. Riflessi sul procedimento penale", in *Riv. it. dir. proc. pen*., (2024), pp. 271 ss.

[3] Cfr., sul punto, tra gli altri, R. Lopez, "La rappresentazione facciale tramite *software*", cit., p. 241.

ve nelle attività sia di prevenzione che di accertamento dei reati, offrendo uno straordinario contributo all'attività investigativa. E, a differenza delle altre metodiche di identificazione biometrica basate sulle caratteristiche fisiologiche di un individuo, hanno un vantaggio evidente ovvero quello di non essere, almeno all'apparenza, strumenti invasivi, non determinando alcuna forma di coazione fisica diretta sulle persone, in quanto si può prescindere dalla collaborazione del destinatario[4].

Il riconoscimento automatico può avvenire attraverso due differenti modalità di funzionamento: la prima è la modalità statica (o *post remote* ovvero ricerca *off-line*, "a posteriori"), basata su immagini o video che sono già agli atti. Per la precisione, una foto di un soggetto con identità ignota viene ricercata all'interno di una banca dati precostituita contenente foto di soggetti con identità nota con l'obiettivo di trovare una corrispondenza tra la prima e una delle foto contenute nella banca dati, pervenendo così all'identificazione della persona ignota (confronto uno-a-molti)[5]: i dati vengono dunque trattati "in differita", successivamente alla raccolta.

La seconda è la modalità dinamica (o *real time*), fondata su immagini acquisite dal vivo e processate in tempo reale. Più esattamente, la lista di foto di soggetti da vagliare è confrontata istantaneamente e in maniera continua con tutti i volti individuati in uno o più flussi video provenienti da altrettante telecamere e, nel caso in cui uno dei volti del video raggiunga un punteggio di somiglianza rispetto a quelli presenti nella "*watch list*" superiore ad una certa soglia (*match* di similarità), il sistema genera un *alert*.

Sul piano strettamente tecnico, prodromica alla vera e propria identificazione del soggetto interessato è la fase di analisi facciale, che si articola in diversi *step*: 1) acquisizione dell'immagine, che può avvenire con modalità diversificate (volontariamente e in ambiente controllato oppure in maniera clandestina, all'insaputa del soggetto monitorato); 2) individuazione del volto all'interno dell'immagine (c.d. "*face detection*"), operazione che può

Nel senso che il volto umano "costituisce il carattere più rappresentativo e individualizzante di un soggetto", T. Alesci, *Il corpo umano come fonte di prova*, Wolters Kluwer Cedam, Milano, p. 89.

[4] In tal senso v. A. Mangiaracina, *op. cit*., p. 78.

[5] In questi termini G. Tessitore, *op. cit*., p. 62.

rivelarsi particolarmente complessa nel caso in cui bisogna isolare l'immagine facciale da un *frame* di un video; 3) normalizzazione dell'immagine, che si sostanzia nel processo diretto ad attenuare le variazioni all'interno delle regioni del volto a causa dell'illuminazione o della posizione; 4) estrazione delle caratteristiche (c.d. "*features*") mediante le quali formare il modello (*template*) biometrico di riferimento; 5) registrazione dell'immagine e del modello biometrico all'interno del *database*; 6) confronto per misurare le somiglianze tra i diversi *template* biometrici registrati nel sistema[6].

2. Il Sistema "S.A.R.I." (Sistema Automatico Riconoscimento Immagini) e gli interventi del Garante per la protezione dei dati personali

Negli Stati Uniti già alla fine degli anni '90 e nei primi anni 2000 si discuteva dell'utilizzo dei dispositivi tecnologici di riconoscimento facciale automatizzato nell'ambito del processo penale; tuttavia il dibattito è divenuto particolarmente acceso nel 2020, quando è scoppiato lo "scandalo *Clearwiew*", dal nome della società statunitense che aveva creato un'*app* di riconoscimento facciale, utilizzata dalla polizia americana e dalle forze dell'ordine di diversi paesi europei, che operava su un *database* contenente miliardi di dati (foto, video, ecc.) "rubati" da *social network* come *Facebook*, *Twitter* e *Instagram* e altri siti web come *You Tube*. In sostanza, il *software* si basava su dati raccolti e trattati illegittimamente.

In Italia, su iniziativa del Ministero dell'interno-Dipartimento di pubblica sicurezza, dal 2016 è stato sviluppato il c.d. "sistema S.A.R.I." (acronimo di "Sistema Automatico Riconoscimento Immagini"), messo a disposizione delle forze di polizia, sulle cui specifiche tecniche di funzionamento, nonostante il suo impiego sia largamente diffuso sul territorio nazionale, viene mantenuto un sostanziale riserbo per non comprometterne l'efficacia e la sicurezza.

Esistono due modalità operative, distinte e separate, in cui si articola il *tool* in questione: il c.d. "S.A.R.I. *Enterprise*", che

[6] Sul punto v., tra gli altri, F. Demartis, *op. cit*., pp. 103 s.; A. Marandola, *op. cit*., p. 500 s.; G. Mobilio, *Tecnologie di riconoscimento facciale*, cit., p. 32 s.

opera per la ricerca di volti a partire da immagini statiche su banche dati di grandi dimensioni; e il c.d. "S.A.R.I. *Real Time*", finalizzato al riconoscimento in tempo reale di volti presenti in flussi video provenienti da telecamere[7].

[7] Come ricorda A. Diddi, *E-Evidence, Data Base, Digital Investigation e AI*, Pacini, Pisa, 2025, p. 164, nel 2018 l'Italia è venuta a conoscenza dell'esistenza di questi sistemi a seguito dell'arresto a Brescia di due soggetti ritenuti responsabili di un furto in appartamento, in quanto nel comunicato stampa si annunciava che ciò era stato possibile grazie al Sistema Automatico di Riconoscimento Immagini (S.A.R.I.) che "consente di effettuare ricerche nella banca dati A.F.I.S. (*Automated Fingerprint Identification System*) attraverso l'inserimento di un'immagine fotografica di un soggetto ignoto che, elaborata da due algoritmi di riconoscimento facciale, fornisce un elenco di immagini ordinato secondo un grado di similarità".

Qualche informazione in più sull'applicativo si è potuta trarre indirettamente dal capitolato tecnico della procedura volta alla fornitura della soluzione integrata per il S.A.R.I., aggiudicata in favore di una società italiana nel 2017 (v. Ministero dell'Interno, *Capitolato tecnico. Procedura volta alla fornitura della soluzione integrata per il Sistema Automatico di Riconoscimento Immagini (S.A.R.I.)*, in www.poliziadistato.it/statics/06/20160627-ct-sari-4-pdf).

All'indomani della diffusione della notizia relativa allo scandalo *Clearwiew*, il 21 gennaio 2021, è stata rivolta una interrogazione parlamentare al Governo per conoscere se un *software* simile fosse in uso alle forze di polizia italiane: la risposta è stata negativa, precisando che era in uso esclusivamente il sistema S.A.R.I. *Enterprise* e che all'epoca nella banca dati AFIS erano presenti 17.592.769 cartellini fotosegnaletici, corrispondenti a 9.882.490 individui diversi. Nella risposta all'interrogazione il Ministro dell'Interno *pro tempore* ha spiegato il funzionamento del SARI *Enterprise* e del SARI *Real Time*, precisando però che quest'ultimo non è mai stato utilizzato, anche perché si attendeva una valutazione del Garante per la protezione dei dati personali (che ha dato poi parere negativo: v. *infra* nel testo). Il testo dell'interrogazione e la risposta del Ministro possono leggersi in *Atti parlamentari*, *XVIII legislatura* – resoconto stenografico dell'Assemblea – seduta n. 463 del 3 marzo 2021.

Ad ogni modo, come segnala L. Belvini, *op. cit*., p. 168, restano "ignote le modalità con le quali sono alimentate le banche date utilizzate per il raffronto automatizzato; non è affatto chiaro il destino delle immagini catturate dalla macchina ma di cui l'operatore ha escluso il *match* nel caso concreto, palesando così l'insidia di raccolte indiscriminate dei dati biometrici da utilizzare per ulteriori ricnoscimenti; non sono note le metodologie con le quali è testata l'accuratezza del *software*; si concretizza, con il S.A.R.I. *Real Time*, il pericolo di monitoraggi clandestini su ampie fette della popolazione; non

Più precisamente, con il sistema S.A.R.I. *Enterprise* le autorità sono messe in grado di ricercare entro 15 secondi l'identità di un volto, presente in un'immagine già acquisita agli atti, all'interno di una grandissima banca dati, individuata nella piattaforma A.F.I.S.-S.S.A., ovvero il sistema automatizzato di identificazione delle impronte digitali (*Automated Fingerprint Identification System*), integrato dal sotto sistema anagrafico, che contiene circa 20 milioni di cartellini fotosegnaletici (recanti una foto frontale e una foto del profilo destro). Al termine dell'operazione di raffronto, il sistema elabora un punteggio sul grado di similarità su una scala percentuale (*score*), restituendo un elenco di profili di candidati (c.d. "*candidate list*", in genere 50 profili) ordinati in base al livello di somiglianza con il soggetto da identificare.

Dunque, il sistema – come del resto tutti i sistemi di riconoscimento facciale automatizzato – offre "un 'ventaglio' di possibili corrispondenze espresso in termini probabilistici su base percentuale"[8]. E il riconoscimento è sempre sottoposto alla successiva verifica di un operatore-persona fisica esperto, tenuto a validare il risultato restituito dalla macchina, accertando che l'indice di similarità non rientri tra le anomalie statistiche, che la comparazione non abbia generato corrispondenze errate e che non siano subentrate interruzioni della "catena di custodia" o altre forme di alterazione del dato: ciò alla luce del divieto di decisioni totalmente automatizzate di cui agli artt. 11 Direttiva UE/680/2016 e 8 d.lgs. 18 maggio 2018, n. 51[9]. Dunque, l'eventuale corrispondenza riscontrata dal sistema con uno o più soggetti con caratteristiche facciali assimilabili a quelle del volto ignoto non ha alcuna conseguenza automatica, ma viene utilizzata per il prosieguo dell'attività investigativa.

Con il sistema S.A.R.I. *Real Time*, invece, è possibile acquisire in diretta le immagini dei volti inquadrati da telecamere (fisse o mobili) collocate in luoghi specifici oggetto di osservazione

è dato sapere se vi sono accorgimenti per limitare gli accessi al *tool* né, tanto meno, se essi sono annotati in un apposito registro".

[8] F. Demartis, *op. cit*., p. 228.

[9] Cfr. sul punto, tra gli altri, J. Della Torre, "Tecnologie di riconoscimento facciale e procedimento penale", cit., p. 1078; F. Demartis, *op. cit*., p. 216 e p. 231 s., il quale sottolinea che si tratta di sistemi non "autosufficienti".

e confrontarli con un *database* più ristretto di persone ricercate (la c.d. "*watch list*", cioè una lista predefinita di persone "attenzionate", compilata di volta in volta in relazione all'evento da monitorare), la cui grandezza è al massimo di 10.000 volti. Il sistema può, più in particolare, essere installato direttamente nel luogo in cui sorga l'esigenza di disporre di una tecnologia di riconoscimento facciale per coadiuvare le forze di polizia nella gestione dell'ordine e della sicurezza pubblica o in relazione a specifiche esigenze investigative.

Allorquando la macchina riscontri una possibile corrispondenza positiva (*match*), viene a generarsi un segnale di allarme (*alert*), in grado di richiamare l'attenzione degli operatori, ai quali spetta il compito di confermare il riconoscimento e di prendere i provvedimenti conseguenti.

Va dunque evidenziato che, in entrambe le modalità, il riconoscimento deve essere sempre sottoposto alla successiva verifica di un operatore-persona fisica.

Ora, il S.A.R.I. *Enterprise* ha ottenuto nel 2018 il parere positivo del Garante per la protezione dei dati personali relativamente alla sua compatibilità con la disciplina in materia di *data protection*, sul rilievo che tale sistema non costituisce un nuovo trattamento di dati personali, in quanto si limita ad automatizzare alcune operazioni che prima richiedevano l'inserimento manuale di connotati identificativi, consentendo le operazioni di ricerca nella banca dati A.F.I.S. attraverso il semplice inserimento di una fotografia[10].

È stata, invece, negativa la valutazione espressa dal Garante nel 2021 rispetto alla variante operativa *Real Time* del medesimo sistema, escludendone la conformità al d.lgs. n. 51/2018, attuativo della Direttiva 680/2016 sulla protezione delle persone fisiche relativamente al trattamento dei dati personali da parte delle autorità competenti[11].

[10] Garante per la protezione dei dati personali, "*Sistema automatico di ricerca dell'identità di un volto*", provvedimento del 26 luglio 2018, n. 440, doc. web 9040256.

[11] Garante per la protezione dei dati personali, "Parere sul sistema SARI Real Time", provvedimento del 25 marzo 2021, n. 127, doc. web 9575877. A commento della decisione, v. A. Fonsi, "Prevenzione dei reati e riconoscimento facciale: il parere sfavorevole del Garante *privacy* sul Sistema SARI

Al di là del riserbo sulle specifiche tecniche legate alle modalità di funzionamento – sulle quali si tornerà più avanti –, il Garante ha ritenuto di estrema delicatezza l'utilizzo di tecnologie di riconoscimento facciale per finalità di prevenzione e repressione dei reati nella modalità dinamica: "S.A.R.I. *Real Time* realizzerebbe un trattamento automatizzato su larga scala, che può riguardare anche persone presenti a manifestazioni politiche e sociali, che non sono oggetto di 'attenzione' da parte delle forze di polizia". Ed anche se, nella valutazione di impatto presentata al Garante dal Ministero degli interni, "si spiega che le immagini verrebbero immediatamente cancellate", nel provvedimento si ribadisce che "l'identificazione di una persona sarebbe realizzata attraverso il trattamento dei dati biometrici di tutti coloro che sono presenti nello spazio monitorato, allo scopo di generare modelli confrontabili con quelli dei soggetti inclusi nella '*watch-list*'. Si determinerebbe, così, una evoluzione della natura stessa dell'attività di sorveglianza, che segnerebbe un passaggio dalla sorveglianza mirata di alcuni individui alla possibile sorveglianza universale[12]".

SARI *Real Time* sarebbe suscettibile, pertanto, di violare i diritti fondamentali delle persone coinvolte. Il trattamento dei dati biometrici svolto da quest'ultimo applicativo è stato, dunque, definito illecito, essendo destituito di adeguata base normativa e non conforme alla normativa in materia di *data protection*.

A seguito di questa seconda pronuncia del Garante, il legislatore ha introdotto una moratoria rispetto all'installazione e all'utilizzazione di impianti di videosorveglianza con sistemi di riconoscimento facciale operanti attraverso l'uso di dati biometrici, in luoghi pubblici o aperti al pubblico, sino all'entrata in vigore di una disciplina legislativa *ad hoc*, e comunque sino alla data del 31 dicembre 2023 (art. 9, comma 9, d.l. 8 ottobre 2021,

Real Time", in *Penale. Dir e proc*., (2021), pp. 289 ss.; R. Lopez, "Videosorveglianza biometrica tramite riconoscimento facciale: parere negativo del Garante per la *privacy*", in *Proc. pen. giust*., (2022), pp. 798 ss.; F. Paolucci, *op. cit.*, pp. 213 ss.; E.C. Raffiotta, M. Baroni, "Intelligenza artificiale, strumenti di identificazione e tutela dell'identità", in *BioLaw Journal – Riv. BioDir*, (2022), pp. 165 ss.

[12] Garante per la protezione dei dati personali, "Parere sul sistema SARI Real Time", provvedimento del 25 marzo 2021, cit.

n. 139, conv. con modif. dalla l. 3 dicembre 2021, n. 205, recante "*Disposizioni urgenti per l'accesso alle attività culturali, sportive e ricreative, nonché per l'organizzazione di pubbliche amministrazioni e in materia di protezione dei dati personali*"), termine poi prorogato al 31 dicembre 2025 (art. 8-*ter* d.l. 10 maggio 2023, n. 51, conv. con modif. dalla l. 3 luglio 2023, n. 87, recante "*Disposizioni urgenti in materia di amministrazione di enti pubblici, di termini legislativi e di iniziative di solidarietà sociali*"), con alcune consistenti deroghe previste dall'art. 9, comma 12, d.l. n. 139/2021 con riferimento alle attività "di prevenzione e repressione dei reati"[13].

3. L'approvazione dell'*AI Act* (Regolamento europeo 1689/2024/UE) e gli usi consentiti e vietati dei sistemi di riconoscimento facciale

Nel frattempo, com'è noto, il 13 giugno 2024 è stato approvato, dopo un lungo *iter*, il Regolamento n. 1689/2024/UE del Parlamento europeo e del Consiglio sull'intelligenza artificiale, noto come *AI Act* (pubblicato sulla G.U. dell'Unione europea il successivo 12 luglio)[14], che, nel quadro di un'ampia e organica regolamentazione dei sistemi di intelligenza artificiale, la prima a livello mondiale, ha offerto innanzitutto una definizione di intelligenza artificiale – cosa niente affatto scontata – nella quale risultano pienamente ricompresi i sistemi di riconoscimento facciale automatizzato.

A seconda delle loro caratteristiche tecniche, tali sistemi possono ricadere, alternativamente, nell'ambito delle pratiche

[13] Sul punto, v., tra gli altri, F. Di Matteo, *op. cit*., p. 110.

[14] Per un quadro d'insieme sulle problematiche relative al procedimento penale disciplinate dall'*AI Act*, v. G. Canzio, "*AI Act* e processo penale: sfide e opportunità", *retro*, in *questo volume*; C. Pinelli, "*AI Act*: gestione del rischio e tutela dei diritti", in *Giur. it*., (2025), pp. 452 ss.; S. Quattrocolo, "Intelligenza artificiale e processo penale: le novità dell'*AI Act*", in *Dir. difesa*, 16 gennaio 2025; A. Scalfati, "IA e processo penale: prospettive d'impiego e livelli di rischio", in *Proc. pen. giust*., (2024), pp. 1404 ss.; C. Teresi, "L'*AI Act* nell'ottica del processual-penalista: uno sguardo preliminare", in *Penale. Dir. e proc*., (2024), n. 3, pp. 361 ss.; M. Torre, "Il regolamento europeo sull'intelligenza artificiale: i profili processuali", in *Proc. pen. giust*., (2024), pp. 1543 ss.

vietate, in quanto "a rischio inaccettabile" di cui all'art. 5 Reg. 1689/2024, potendo rappresentare una minaccia grave per i diritti fondamentali, ovvero nell'ambito dei sistemi "ad alto rischio", di cui agli artt. 6 ss. Reg. 1689/2024, e sono prescritti specifici adempimenti sia ai fornitori (*provider*) che agli utilizzatori (*deployer*) al fine di garantire la conformità di questi ultimi sistemi a quanto previsto dal Regolamento.

Va accolta, innanzitutto, con grande favore la scelta del legislatore europeo di vietare in radice, in modo assoluto, pratiche come quelle che hanno portato allo scandalo *Clearwiew*, trattandosi di un livello di rischio ritenuto inaccettabile, in ragione della potenziale violazione di diritti fondamentali: ai sensi dell'art. 5, par. 1, lett. e), Reg. 1689/2024, infatti, non è consentito, in alcun caso, l'impiego di sistemi di A.I. che creano o ampliano le banche dati di riconoscimento facciale mediante *scraping*, cioè il rastrellamento non mirato di immagini facciali estrapolate da *internet* o da filmati di telecamere a circuito chiuso. Tale pratica – come si legge nel *Considerando* n. 43 – "accresce il senso di sorveglianza di massa e può portare a gravi violazioni di diritti fondamentali, compreso il diritto alla vita privata".

Nell'ambito delle pratiche vietate, ai sensi dell'art. 5, par. 1, lett. h), Reg. 1689/2024, rientrano anche i sistemi di identificazione biometrica remota "in tempo reale" in "spazi accessibili al pubblico"[15] a fini di attività di contrasto, così dovendosi intendere gli applicativi che – sul modello di SARI *Real Time* – consentono l'identificazione automatizzata della persona, mediante la comparazione di dati biometrici, tra i quali immagini facciali,

[15] Per "spazio accessibile al pubblico", il Regolamento 1689/2024/UE (*Considerando* n. 19) intende "qualsiasi luogo fisico accessibile a un numero indeterminato di persone fisiche e a prescindere dal fatto che il luogo in questione sia di proprietà pubblica o privata, indipendentemente dall'attività per la quale il luogo può essere utilizzato, quali il commercio (ad esempio negozi, ristoranti, bar), i servizi (ad esempio banche, attività professionali, ospitalità), lo sport (ad esempio piscine, palestre, stadi), i trasporti (ad esempio stazioni di autobus, metropolitane e ferroviarie, aeroporti, mezzi di trasporto), l'intrattenimento (ad esempio cinema, teatri, sale da concerto e sale conferenze), il tempo libero o altro (ad esempio strade e piazze pubbliche, parchi, foreste, parchi giochi)".

in termini pressoché istantanei, attraverso il confronto con i dati presenti in archivio.

Tale divieto, tuttavia, non è assoluto, essendo l'impiego dei sistemi di riconoscimento facciale automatizzato "*real time*" consentito, alla luce del principio di proporzione, in caso di assoluta necessità al fine di perseguire uno degli obiettivi specificamente indicati nel seguito della disposizione, vale a dire: la ricerca mirata di specifiche vittime di gravi reati (sottrazione, tratta o sfruttamento sessuale di esseri umani), nonché la ricerca di persone scomparse (ricerca che potrebbe anche non avere alcun aggancio con la sfera penalistica)[16]; la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di una minaccia reale attuale o prevedibile di un attacco terroristico (qui l'istanza è prettamente securitaria e preventiva, ma verosimilmente legata, per via delle caratteristiche di specificità richieste, a investigazioni già in corso per altri fatti); la localizzazione o l'identificazione di una persona sospettata di aver commesso determinati reati, espressamente elencati nell'Allegato II al Regolamento, ai fini dello svolgimento di una indagine penale, o dell'esercizio dell'azione penale o dell'esecuzione di una sanzione penale per tali reati, punibili nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata superiore nel massimo a quattro anni.

L'Allegato II, per la precisione, individua sedici categorie di illeciti per le quali è possibile l'impiego dei sistemi di riconoscimento facciale ai fini della localizzazione e/o identificazione di persone sospettate: terrorismo; tratta di esseri umani; sfruttamento sessuale di minori e pornografia minorile; traffico illecito di stupefacenti o sostanze psicotrope; traffico illecito di armi, munizioni ed esplosivi; omicidio volontario; lesioni gravi; traffico illecito di organi e tessuti umani; traffico illecito di materie nucleari e radioattive; sequestro, detenzione illegale e presa di ostaggi; reati che rientrano nella competenza giurisdizionale della Corte penale internazionale; illecita cattura di aeromobile o nave; violenza sessuale; reato ambientale; rapina organizzata o

[16] Come evidenzia S. Quattrocolo, *op. cit.*, p. 9.

a mano armata; sabotaggio; partecipazione a un'organizzazione criminale coinvolta in uno o più dei reati sopra elencati[17].

Ora, è vero che l'utilizzo da parte delle forze dell'ordine dei sistemi di identificazione biometrica remota in tempo reale richiede comunque l'autorizzazione preventiva dell'autorità giudiziaria[18] o di un'autorità amministrativa indipendente dello Stato membro, fatta salva, in casi di urgenza, la possibilità di una convalida *ex post*, richiesta, senza ritardo, entro le ventiquattro ore successive all'uso del sistema, dovendosi, in mancanza, procedere immediatamente a interrompere l'utilizzo, nonché all'eliminazione e alla cancellazione dei dati, dei risultati e degli *output* (art. 5, par. 3, Reg. 1689/2024). Ed è anche vero che l'attività può essere autorizzata, su richiesta motivata, dall'autorità giudiziaria o amministrativa competente, la cui decisione è vincolante, solo all'esito di un positivo vaglio di necessità e proporzionalità dell'uso del sistema di identificazione biometrica per il conseguimento di uno degli obiettivi per cui è consentita dal Regolamento e solo se l'autorità di contrasto abbia eseguito una valutazione di impatto sui diritti fondamentali e registrato il sistema nella relativa banca dati UE (art. 5, par. 2, Reg. 1689/2024).

Tuttavia, al di là dello sforzo del legislatore europeo di cercare di circoscrivere le eccezioni al divieto, appare evidente che le deroghe previste sono talmente generiche che rischiano, di fatto, di svuotare completamente di significato il divieto (non a caso, durante la discussione del testo, il Parlamento europeo aveva espresso assoluta contrarietà alla previsione di queste deroghe)[19].

[17] F. Demartis, *op. cit*., p. 82, sottolinea che si tratta ovviamente di "indicazioni di carattere generale, che devono essere poi adeguate in base alla legislazione penalistica dei singoli Paesi, al fine di ricercare le fattispecie incriminatrici corrispondenti".

[18] In considerazione dei delicati interessi coinvolti, appare corretto ritenere che nel nostro sistema processuale l'autorizzazione debba provenire dal giudice procedente e, nella fase delle indagini preliminari, dal giudice per le indagini preliminari: in tal senso v. F. Demartis, *op. cit*., p. 84; M. Torre, "Il regolamento europeo sull'intelligenza artificiale", cit., p. 1543.

[19] La sottolineatura è di S. Quattrocolo, *op. cit*., p. 9, la quale osserva altresì (p. 8 s.): "Del resto la gamma di eccezioni giustifica, in primo luogo, la dotazione, in ogni possibile luogo pubblico, di telecamere 'di sicurezza', pe-

In definitiva, "pare ragionevole immaginare che solo un estremo rigore interpretativo possa evitare un generalizzato ricorso al riconoscimento facciale in *real time*, in tutti i luoghi aperti al pubblico, con sostanziale rovesciamento del divieto pur scolpito nell'art. 5 par. 1 lett. h)" del Regolamento europeo sull'intelligenza artificiale[20].

All'art. 26, par. 10, Reg. 1689/2024 trova posto, invece, la disciplina dei sistemi di riconoscimento facciale che operano "a posteriori". Tale attività, valutata come ad "alto rischio", è consentita, ma solo al fine di ricercare "una persona sospettata o condannata per aver commesso un reato", ed è subordinata ad un'autorizzazione, *ex ante* o senza indebito ritardo ed entro quarantotto ore, da parte di un'autorità giudiziaria o amministrativa, la cui decisione è vincolante e soggetta a controllo giurisdizionale. L'uso deve essere "limitato a quanto strettamente necessario per l'indagine su uno specifico reato". L'autorizzazione,

raltro già ampiamente diffusa. Inoltre, il perseguimento delle finalità indicate nelle eccezioni finirebbe per legittimare sempre la ripresa di ciò che accade nel luogo pubblico, al fine di poter individuare, di volta in volta, la vittima di grave delitto, l'accusato o condannato di uno dei reati dell'Allegato II, il soggetto pericoloso…".

Sul punto, v. anche G. Galluccio Mezio, *op. cit*., p. 668, il quale osserva che le previsioni del Regolamento europeo sull'intelligenza artificiale "prevedono sì un ampio novero di requisiti e di condizioni di legittimità dell'uso dei sistemi di riconoscimento facciale in tempo reale o a posteriori, "ma, al contempo, sono contraddistinte da un notevole tasso di genericità e corredate da numerose previsioni derogatorie", sicché non paiono delineare una disciplina sufficientemente stringente e in grado di 'resistere' a possibili tentativi di aggiramento"; nonché V. Vasta, *op. cit*., p. 284, la quale osserva che "la formulazione letterale delle eccezioni al divieto, talora generica", presta il fianco a "rischi di disomogeneità applicativa nei diversi Stati membri", rischio che si può sicuramente presentare "anche rispetto alla possibilità di utilizzare il sistema per la ricerca degli autori di determinati reati individuati con riferimento al limite massimo edittale della pena, che muta a seconda delle legislazioni nazionali". Nello stesso senso, del resto, già E. Sacchetto, "Automated faced based human recognition tecnologies e procedimento penale alla luce dell'AI Act: alcuni spunti di riflessione", in C. Camardi (a cura di) *La via europea per l'intelligenza artificiale. Atti del Convegno del Progetto Dottorale di Alta Formazione in Scienze Giuridiche – Cà Foscari Venezia, 25-26 novembre 2021*, Wolters Kluwer Cedam, Milano, 2022, p. 371, con riferimento alla Proposta di regolamento.

[20] Così, ancora S. Quattrocolo, *op. cit*., p. 9.

tuttavia, non è richiesta quando il sistema di riconoscimento facciale automatizzato è utilizzato per l'"identificazione iniziale di un potenziale sospettato sulla base di fatti oggettivi e verificabili direttamente connessi al reato".

In ogni caso, occorre "garantire che nessuna decisione che produca effetti giuridici negativi su una persona" possa essere presa dalle autorità di contrasto "unicamente sulla base dell'*output* di tali sistemi di identificazione biometrica a posteriori"[21].

Il Reg. 1689/2024 – che fa salve le previsioni del diritto europeo in materia di protezione dei dati personali – introduce inoltre un'organica disciplina relativa all'immissione in commercio e all'impiego dei sistemi di intelligenza artificiale, prevedendo delle disposizioni intertemporali per l'entrata in vigore di alcune disposizioni, così da incoraggiare i legislatori nazionali dei singoli Stati membri a valutare l'introduzione di disposizioni di maggior dettaglio, in coerenza con le peculiarità dei singoli sistemi giuridici[22]. In particolare, l'art. 113 Reg. 1689/2024 prevede che le disposizioni in esso contenute si applichino a decorrere dal 2 agosto 2026, fatti salvi i capi I e II la cui decorrenza è stata anticipata al 2 febbraio 2025 (e vi è ricompreso dunque l'art. 5), mentre le previsioni di cui all'art. 6, par. 1, si applicano a decorrere dal 2 agosto 2027.

[21] Sul punto, v. già il *Considerando* n. 35 Reg. 1689/2024/UE.

[22] Cfr. sul punto, G. Galluccio Mezio, *op. cit*., p. 667, per il quale è indubbio che "le articolate disposizioni intertemporali previste dal Regolamento, volte a differire la cogenza delle disposizioni più qualificanti" siano proprio "finalizzate a consentire agli Stati membri di dotarsi di una disciplina armonica alle previsioni europee e, specie con riferimento alle garanzie procedurali minime da esse appena abbozzate, coerente con le peculiarità dei singoli sistemi giuridici".

4. La l. 23 settembre 2025, n. 132 e i tentativi di inquadramento giuridico del riconoscimento facciale automatizzato, tra tipicità e atipicità, nelle indagini preliminari e nel dibattimento, in attesa di una compiuta regolamentazione

Il 10 ottobre 2025 è entrata in vigore la l. 23 settembre 2025, n. 132, recante "*Disposizioni e deleghe al Governo in materia di intelligenza artificiale*", finalizzata ad armonizzare la normativa nazionale alle disposizioni dell'*AI Act*, nonché a definire una strategia nazionale per un corretto sviluppo di detta tecnologia in Italia[23].

L'art. 24, comma 1, l. n. 132/2025, prevede che, entro un anno dall'entrata in vigore della legge, il Governo sia delegato ad emanare uno o più decreti legislativi per l'adeguamento della normativa nazionale al Regolamento 1689/2024/UE, con la previsione di un'apposita disciplina per l'utilizzo di sistemi di intelligenza artificiale per l'attività di polizia (art. 24, comma 2, lett. h, l. n. 132/2025).

Il Governo, a norma dell'art. 24, comma 3, l. n. 132/2025, è altresì delegato ad adottare, entro lo stesso termine, uno o più decreti legislativi per adeguare e specificare la disciplina dei casi di realizzazione e di impiego illeciti di sistemi di intelligenza artificiale, prevedendo, tra l'altro, che l'utilizzo dei sistemi di intelligenza artificiale nelle indagini preliminari sia regolato nel rispetto delle garanzie inerenti al diritto di difesa e ai dati personali dei terzi, nonché dei princìpi di proporzionalità, non discriminazione e trasparenza (art. 24, comma 5, lett. e, l. n. 132/2025) e la necessità di modificare conseguentemente, a fini di coordinamento e razionalizzazione del sistema, la normativa sostanziale e processuale (art. 24, comma 5, lett. f, l. n. 132/2025).

[23] Tra i primi commenti, per quanto concerne i profili attinenti al procedimento penale, C. Parodi, T. Rizzo (a cura di), *Intelligenza artificiale e indagini penali. Strumenti, garanzie, responsabilità e scenari dell'innovazione investigativa digitale. Aggiornato con la l. 23 settembre 2025, n. 132, Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, La Tribuna, Milano, 2025.

Allo stato, dunque, manca una espressa e precisa disciplina legislativa nazionale dei sistemi di riconoscimento facciale automatizzato, ed è difficile prevedere quando tale normativa sarà effettivamente operativa, considerando i tempi per l'attuazione delle deleghe al Governo e la necessità, poi, di una indispensabile disciplina attuativa di secondo grado per tutti i dettagli tecnici.

Quanto ai pareri del Garante per la protezione dei dati personali prima ricordati, è appena il caso di rilevare che sono ovviamente limitati alla tutela della *privacy*, lasciando dunque irrisolti gli snodi maggiormente rilevanti in ordine all'impiego dei sistemi di riconoscimento facciale automatizzato nel procedimento penale, con riferimento in particolare alla stessa idoneità epistemologica dello strumento, alle ricadute sulle altre libertà fondamentali, ai casi in cui tali sistemi sono utilizzabili[24].

Vi è anche una carenza di specifiche indicazioni giurisprudenziali in materia: se in altri ambiti la giurisprudenza costituzionale e di legittimità ha svolto una fondamentale opera di supplenza all'inerzia del legislatore, fissando precisi "paletti" a tutela dei diritti fondamentali – è il caso, ad es., delle videoriprese investigative[25] – questo non è (o perlomeno non è ancora) avvenuto con riferimento alle tecnologie di riconoscimento del volto[26].

Nelle poche decisioni di merito disponibili, ci si limita per lo più a dare atto dell'uso investigativo del *software* con finalità identificative, omettendo però di affrontare il quesito prelimina-

[24] In questi termini, L. Belvini, *op. cit*., p. 169.

[25] Sul punto, sia consentito rinviare a N. Triggiani, "Le videoriprese investigative tra Corte costituzionale e Sezioni Unite: cronaca di una travagliata evoluzione giurisprudenziale (…in attesa del legislatore)", in *Iurisdictio*, (2023), pp. 40 ss.

[26] In altri ordinamenti, invece, da tempo la giurisprudenza ha individuato alcune coordinate circa l'impiego delle tecnologie di riconoscimento facciale: è il caso, ad es., del Regno Unito, sul quale v. J. Della Torre, "Novità dal Regno Unito: il riconoscimento facciale supera il vaglio della *High Court of Justice* (ma raccoglie le critiche del Garante *privacy* d'oltremanica)", in *Dir. pen. cont. trim.*, (2020), pp. 231 ss.; A. Pin, "Non esiste la 'pallottola d'argento': l'*Artificial Face Recognition* al vaglio giudiziario per la prima volta", in *DPCE on line*, (2019), n. 4, pp. 3075 ss.

re sulle norme abilitanti l'impiego del programma di A.I.[27], mentre la giurisprudenza di legittimità, nelle rare occasioni in cui ha avuto modo di occuparsi del tema, ha trascurato l'argomento o lo ha reputato comunque non decisivo nel caso specifico[28].

Ora, in attesa di una dettagliata regolamentazione nazionale ancora di là da venire, dal momento che il S.A.R.I. è comunque una realtà già operativa, e a quanto risulta, ampiamente utilizzata dalle forze di polizia e dalla polizia giudiziaria, ci si deve interrogare sul possibile impiego di questa tecnologia nell'ambito

[27] Lo evidenzia L. Belvini, *op. cit*., p. 170 e note 77 e 78. Cfr. App. Roma, sez. II, 2 ottobre 2023, n. 10003, in *OneLegale*; Trib. Nola, 11 gennaio 2024, n. 2055, *ivi*, la quale accenna all'uso del S.A.R.I., dando atto che, dopo aver visionato le immagini acquisite dagli impianti di videosorveglianza di un istituto bancario, la polizia giudiziaria aveva inserito il fotogramma estrapolato dal video nel citato sistema in uso alla polizia scientifica e quest'ultimo aveva individuato delle somiglianze con l'imputato; Trib. Napoli, sez. VI, 26 settembre 2022, n. 8679, *ivi*; Trib. Ascoli Piceno, 20 luglio 2022, n. 452, *ivi*; Trib. Genova, sez. I, 26 febbraio 2022, n. 600, in *Guida dir*., (2022), n. 36, p. 84.

[28] Cfr., da ultimo, Cass., sez. II, 20 marzo 2025, n. 18099, in *Terz'ultima fermata*, 27 agosto 2025, la quale ha ritenuto infondato il ricorso dell'imputato che riteneva "illogico attribuire valenza indiziaria a una compatibilità solo del 55,2% tra le immagini di uno dei soggetti che erano stati ripresi dalle telecamere di videosorveglianza e l'immagine del proprio cartellino foto-segnaletico" (poiché, così facendo, secondo il difensore "[s]i entra nel concetto di probabilismo confinante con il possibilismo così non consentendo che dall'esito accertativo se ne possa dedurre un elemento a carico dell'imputato"), sul rilievo che nel caso di specie, nell'economia della decisione della Corte d'appello impugnata, gli esiti del riconoscimento facciale con il sistema S.A.R.I. avevano assunto solo il ruolo di uno dei plurimi riscontri individualizzanti alle dichiarazioni accusatorie che erano state rese nei confronti dell'imputato dai due coimputati. In motivazione la Corte ha richiamato un suo precedente (Cass., sez. IV, 13 luglio 2023, n. 39551, in *De Jure*), che non ha escluso l'utilizzabilità degli esiti dell'utilizzo del Sistema S.A.R.I., qualora gli stessi si inseriscano in un più strutturato complesso di risultanze probatorie. In senso analogo, v. pure Cass., sez. I, 6 maggio 2021, n. 36441, in *C.E.D. Cass*., n. 282005; Cass., sez. I, 21 luglio 2020, n. 21823, in *OneLegale*.

In altre occasioni ancora, la Corte, preferendo non avventurarsi in ordine alla valutazione sull'attendibilità del S.A.R.I, ha respinto le doglianze difensive relative al mancato utilizzo del *software* di riconoscimento facciale reputato maggiormente affidabile rispetto ai riconoscimenti effettuati *de visu*: cfr. Cass., sez. IV, 14 marzo 2022, n. 8428, in *Onelegale*; Cass., sez. IV, 27 settembre 2019, n. 39731, *ivi*.

dell'attività investigativa e sul terreno probatorio, e quindi sul suo possibile inquadramento giuridico.

La verifica sulla reperibilità nella disciplina processuale di disposizioni nelle quali incasellare i *tool* di riconoscimento facciale, oltre a rispondere all'ineludibile necessità di conformarsi al canone di legalità, è essenziale per stabilire quali impieghi sono compatibili con le limitazioni già dettate nell'*A.I. Act*[29] e nella legge-delega italiana recentemente approvata.

Si è suggerito di inquadrare talune funzioni del S.A.R.I., quanto meno nella versione "*Enterprise*", nell'ambito delle attività di polizia giudiziaria volte all'identificazione dell'indagato *ex* art. 349 c.p.p.[30], e precisamente nell'ambito degli "altri accertamenti" ammessi dalla disposizione codicistica per finalità identificative (art. 349, comma 2, c.p.p.), svincolati dal rispetto di rigide forme prestabilite dal legislatore, e dunque realizzabili con qualsiasi metodologia, ivi incluse le metodiche che si servono dei dispositivi di A.I.[31].

Questa soluzione, peraltro, non è esente da profili critici[32]. Va sottolineato, infatti, che le attività indicate dall'art. 349, comma 2, c.p.p. sono in grado di raggiungere un notevole livello di invasività, essendo idonee a rilevare dati personali sensibili (biometrici e genetici): per questa ragione, il dato normativo postula la residualità delle operazioni maggiormente intrusive, ammettendo il ricorso a rilievi dattiloscopici, fotografici e antropometrici, nonché, per l'appunto, ad "altri accertamenti", solo "ove occorra", legittimandoli, quindi, soltanto qualora la persona rifiuti di fornire le proprie generalità o vi siano dubbi sulla ve-

[29] In questi termini, L. Belvini, *op. cit*., p. 170.

[30] In generale sull'art. 349 c.p.p., v., per tutti, M. Bontempelli, V. Marchesi, *sub* art. 349, in G. Canzio, R. Bricchetti (a cura di), *Codice di procedura penale*, 2ª ed., tomo I, Giuffrè Francis Lefebvre, Milano, 2024, pp. 2796 ss.; P.P. Paulesu, *sub* art. 349, in A. Giarda, G. Spangher (a cura di), *Codice di procedura penale commentato*, 5ª ed., tomo II, Wolters Kluwer, Milano, 2023, pp. 1678 ss.

[31] In tal senso, v. M. Gialuz, "Intelligenza artificiale e diritti fondamentali in ambito probatorio", in *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Giuffrè Francis Lefebvre, Milano, 2021, p. 63; M. Torre, "Intelligenza artificiale e indagini penali: prospettive future e garanzie di Sistema", cit., p. 1745.

[32] Su tali criticità, v. L. Belvini, *op. cit*., p. 171 s.

ridicità delle dichiarazioni oppure sull'autenticità dei documenti di identificazione forniti.

Ora, se, per un verso, appare chiaro che il S.A.R.I., almeno nella variante *Enterprise*, è meno invasivo rispetto alle identificazioni effettuate mediante il prelievo coattivo di capelli o saliva (consentito dall'art. 349, comma 2-*bis*, c.p.p.), dall'altro, si amplifica il rischio che il sistema in questione metta in crisi la sussidiarietà postulata dall'art. 349, comma 2, c.p.p.: avendo la disponibilità di uno strumento di così agevole impiego come il S.A.R.I., è facile pronosticare la propensione degli inquirenti a servirsene, a discapito delle altre attività che – non richiedendo l'acquisizione di dati biometrici – risulterebbero meno invasive nella sfera personale dell'indagato[33].

Di più. Potrebbe verificarsi anche un rischio opposto a quello appena descritto, ossia di sottoporre l'indagato a rilievi fotografici allo scopo di affidarli al responso del S.A.R.I. anche laddove non vi sia una reale incertezza sull'identità, così alimentando i già corposi *database* dei soggetti fotosegnalati. Rischio, peraltro, intensificato dalla novella normativa introdotta dall'art. 2, comma 8, l. 27 settembre 2021, n. 134 (c.d. "riforma Cartabia") che, escludendo la discrezionalità della polizia giudiziaria nell'esecuzione dei citati rilievi *ex* art. 349, comma 2, c.p.p., impone l'obbligo di effettuarli sempre nei confronti degli apolidi e di altri soggetti stranieri[34].

[33] Così L. Belvini, *op. cit*., p. 171, per il quale il pericolo di accordare "una corsia preferenziale" al sistema di riconoscimento facciale "s'intensifica laddove si consideri che questa pratica, potendo essere condotta senza il coinvolgimento dell'indagato" – quando, ad es., la polizia giudiziaria è già in possesso di una sua fotografia –, permetterebbe di mantenere il riserbo sull'indagine, evitando di innescare le garanzie prescritte dall'art. 349, comma 3, c.p.p.

[34] Tale rischio è segnalato sempre da L. Belvini, *op. cit*., p. 171 s.

In ordine alla citata modifica apportata dalla "riforma Cartabia", cfr. E. Dei Cas, "Le nuove frontiere dell'identificazione personale, tra diritto processuale penale dello straniero e 'cultura del controllo'", in *Arch. pen. (web)*, (2024), n. 1, pp. 1 ss.; L. della Ragione, "Le nuove norme sull'identificazione", in B. Romano, A. Marandola (a cura di), *La Riforma Cartabia. La prescrizione, l'improcedibilità e le altre norme immediatamente precettive*, Pacini, Pisa, 2022, pp. 93 ss.; F. Giunchedi, "Il codice univoco identificativo", in R.M. Geraci (a cura di), *La giustizia penale dopo la c.d. Riforma Car-*

Il timore di eventuali usi arbitrari dei sistemi di riconoscimento facciale automatizzato appare tanto più fondato se si considera che l'identificazione rientra tra le attività investigative di iniziativa della polizia giudiziaria e, dunque, non è contemplato nemmeno un adeguato vaglio preventivo del pubblico ministero, il cui intervento è previsto soltanto in ipotesi circoscritte (ovvero quando bisogna procedere al prelievo coattivo di materiale biologico o all'accompagnamento dell'interessato presso gli uffici di polizia giudiziaria): la qual cosa collide con la previsione che delimita l'impiego dei sistemi di identificazione biometrica a quanto "strettamente necessario per le indagini", subordinandolo alla preventiva autorizzazione dell'autorità giudiziaria (artt. 5, par. 3, e 26, par. 10, Reg. 1689/2024)[35].

È necessario, poi, vagliare l'ammissibilità dei sistemi di riconoscimento facciale a supporto dell'attività del pubblico ministero di individuazione di persone *ex* art. 361 c.p.p., ai fini della "immediata prosecuzione delle indagini"[36].

Così adoperato, lo strumento costituirebbe un formidabile ausilio per l'investigazione: tra gli svariati impieghi, a titolo esemplificativo, esso consentirebbe, nella modalità *Enterprise*, di individuare, a partire dalle immagini catturate dalle videocamere di sorveglianza presenti sulla *scena criminis*, le persone verso cui dirigere le indagini o che possono riferire circostanze utili per l'accertamento del fatto di reato e, nella versione *Real Time*, di affinare le ricerche nei luoghi pubblici dei soggetti indiziati, distinguendoli tra la folla.

Peraltro, l'ulteriore vantaggio del S.A.R.I. sarebbe quello di non subire l'inevitabile degradazione mnestica tipica della men-

tabia. Aggiornato alla legge 24 novembre 2023, n. 168 e al d.lgs. 19 marzo 2024, n. 31, Giappichelli, Torino, 2024, pp. 165 ss.; P. Maggio, "Apolidia e processo penale: nuove norme sulla identificazione dei *non cives*", in G. Spangher (a cura di), *La riforma Cartabia. Codice penale – Codice di procedura penale – Giustizia riparativa*, Pacini, Pisa, 2022, pp. 214 ss.; G. Varraso, "La 'nuova' identificazione dell'indagato-imputato apolide o proveniente da Paesi extra UE", in *Dir. pen. proc*., (2021), pp. 1456 ss.

[35] In questi termini, L. Belvini, *op. cit*., p. 172 s.

[36] Sull'individuazione di persone, in generale v., per tutti, P. Gaeta, A. Picardi, *sub* art. 361, in A. Giarda, G. Spangher (a cura di), *op. cit.*, tomo II, pp. 1890 ss.

te umana e di essere immune ai diversi fattori distorsivi – durata e contesto di osservazione, condizioni fisiologiche e psicologiche dell'osservatore, ecc. – che inevitabilmente influenzano il processo intellettivo con il quale una persona ne riconosce un'altra[37].

Adoperato per questi scopi, il programma di riconoscimento facciale sembrerebbe costituire una *species* tecnologicamente avanzata dell'attività disciplinata dall'art. 361 c.p.p.[38].

Tuttavia anche questa strada non appare percorribile. La laconica regolamentazione offerta dall'art. 361 c.p.p., oltre ad essere del tutto incompatibile con i requisiti fissati dall'*A.I. Act* per contenere la straordinaria invasività dei sistemi che operano in *real time*, non soddisfa nemmeno le garanzie minime della fonte europea, che impongono di contingentare l'uso di tutti i sistemi di identificazione biometrica a quanto strettamente necessario per le indagini (artt. 5, par. 3 e 26, par. 10, Reg. 1689/2024), nel rispetto del criterio di proporzionalità.

La disponibilità di un *software* dalle potenzialità così promettenti come il S.A.R.I., peraltro, rischierebbe di irrobustire la consolidata tendenza della giurisprudenza ad attribuire un'indebita valenza probatoria a un atto che, viceversa, il dettato codicistico vorrebbe limitato soltanto alla "immediata prosecuzione delle indagini"[39].

Invero, è l'idea stessa di servirsi del S.A.R.I. per le finalità indicate dall'art. 361 c.p.p. a non essere accoglibile a causa della natura 'artificiale' e non 'umana' del riconoscitore: l'individuazione effettuata dal *software* contrasta con il dato normativo di

[37] Su tali fattori, cfr. G. Gulotta, M. Tuosto, *Il volto nell'investigazione e nel processo. Nuova Fisiognomica Forense*, Giuffrè, Milano, 2017, pp. 124 ss.; nonché, volendo, N. Triggiani, *Ricognizioni mezzo di prova nel nuovo processo penale*, Giuffrè, Milano, 1998, pp. 263 ss.

[38] Cfr. R. Lopez, "La rappresentazione facciale tramite *software*", cit., p. 253, secondo la quale il riconoscimento facciale sarebbe accostabile, seppur sotto forma di "filiazione spuria" all'individuazione *ex* art. 361 c.p.p.

[39] Il rilievo è di L. Belvini, *op. cit*., p. 174. In ordine alla tendenza della giurisprudenza ad attribuire valenza probatoria all'atto di individuazione, anche nel dibattimento, sia consentito rinviare a N. Triggiani, *Ricognizioni mezzo di prova nel nuovo processo penale*, cit, pp. 231 ss.

cui all'art. 361 c.p.p. che, invece, postula un soggetto ricognitore in 'carne ed ossa'[40].

D'altronde, sarebbe "impensabile assimilare la capacità di un soggetto a fare un 'riconoscimento' con quella di un sistema algoritmico che punta a confrontare meccanicamente le caratteristiche dei volti", perché si rischierebbe "di confondere i processi della mente umana con quelli della macchina"[41]. Quest'ultima, essendo incapace di riprodurre i processi neurologici e intuitivi umani, è addestrata per imparare a riconoscere le immagini a partire da un insieme di *pixel*, ricorrendo a inferenze statistiche elaborate su enormi quantità di dati; di conseguenza, il riconoscimento effettuato dal *tool* non è in alcuna misura accostabile all'attività indicata nell'art. 361 c.p.p. perché la comparazione eseguita dall'algoritmo segue logiche ed elabora dati totalmente differenti da quelli che entrano in gioco nell'ambito dell'individuazione per finalità investigative operata dall'uomo[42].

È vero che è prevista una supervisione umana del risultato offerto dalla macchina attraverso un giudizio di compatibilità degli accoppiamenti, secondo un determinato coefficiente di probabilità, ma tale attività non trasforma in una valutazione umana il processo di elaborazione operato dal sistema.

Una volta accertata l'impossibilità di inquadrare il S.A.R.I. nei ricordati atti tipici della polizia giudiziaria e del pubblico ministero[43], bisogna verificare se sussistano i margini per inca-

[40] Così L. Belvini, *op. cit*., p. 174.

[41] In questi termini, C. Pallante, *op. cit*., p. 274. Nello stesso senso, L. Belvini, *op. cit*., p. 174, per il quale "sarebbe errato accostare le abilità dell'algoritmo alle capacità intellettive dell'uomo, stante la profonda differenza nel 'leggere' e confrontare le immagini facciali"; M. Colacurci, *op. cit*., p. 128; G. Galluccio Mezio, *op. cit*., pp. 645 ss.; G. Mobilio, *Tecnologie di riconoscimento facciale*, cit., p. 42.

[42] L. Belvini, *op. cit*., p. 175.

[43] Come osserva L. Belvini, *op. cit*., p. 175, nota 98, va scartata anche l'opzione di inserire il S.A.R.I. nell'eterogenea categoria degli "accertamenti tecnici", in quanto la soluzione non offre alcuna garanzia né per quanto concerne la salvaguardia dei dati biometrici né, tanto meno, è rispettosa degli *standard* minimi richiesti dall'*A.I. Act* per fruire dei sistemi di identificazione biometrica; al contempo, va esclusa pure l'opzione di legittimare l'impiego degli strumenti in questione ampliando l'ambito operativo dei prelievi biologici coattivi di cui all'art. 359-*bis* c.p.p., in quanto, trattandosi di una disposizione

sellarlo nell'ambito delle indagini "atipiche"[44].

Anche questa opzione va, però, decisamente respinta, in quanto incompatibile con il paradigma delineato dall'art. 189 c.p.p. Tale norma – com'è noto – consente l'assunzione di "prove atipiche"[45], a condizione però che siano idonee all'accertamento del fatto di reato e non ledano la libertà morale della persona fonte di prova, e si ritiene, sia pure con qualche difficoltà, applicabile anche alla fase delle indagini preliminari.

Ora, le numerose incognite sui meccanismi di funzionamento dei *tool* di riconoscimento facciale costituiscono un ostacolo insormontabile per stabilirne l'attitudine epistemologica[46].

Il *match* restituito dal S.A.R.I. – al pari degli *output* prodotti dalla quasi totalità dei sistemi automatizzati di riconoscimento facciale – "scaturisce dall'autoapprendimento delle fittissime reti neurali dell'algoritmo, il quale, essendo addestrato con me-

relativa alla libertà personale e, dunque, di stretta interpretazione, non è suscettibile di letture analogiche (per ulteriori considerazioni su entrambi i profili appena considerati, v. J. Della Torre, "Tecnologie di riconoscimento facciale e procedimento penale", cit., pp. 1083 ss.; A. Marandola, *op. cit*., p. 512 s.).

Non è possibile neppure fare riferimento alla disciplina delle intercettazioni allo scopo di consentire l'uso della sorveglianza clandestina e contestuale realizzata dal S.A.R.I. *Real Time*, dal momento che "alla videoripresa con riconoscimento facciale manca del tutto la captazione di una comunicazione e, dunque, sulla base delle Sezioni Unite *Prisco*, non si può applicare per analogia la normativa sulle intercettazioni" (così M. Gialuz, *op. cit*., p. 63 s.).

[44] Sul tema delle indagini atipiche, cfr. diffusamente A. Scalfati (a cura di), *Le indagini atipiche*, 2ª ed., cit. Volendo, cfr. anche N. Triggiani, "Legalità opaca, raccolta atipica e pre-investigazioni", in V. Maffeo (a cura di), *La procedura criminale. Quali riforme, Atti del Convegno – Web conferencing 22-23 ottobre 2020*, ISEG, Roma-Perugia, 2021, pp. 23 ss.

[45] In argomento, v., tra gli altri, C. Castaldi, *La prova atipica nel processo penale*, Roma, 2025; C. Conti, *sub* art. 189, in A. Giarda, G. Spangher (a cura di), *op. cit.*, tomo I, pp. 2601 ss.; A. Laronga, *Le prove atipiche nel processo penale*, Cedam, Padova, 2002; G. Tabasco, *Prove non disciplinate dalla legge nel processo penale. Le "prove atipiche" tra teoria e prassi*, Edizioni scientifiche italiane, Napoli, 2011.

[46] In tal senso, v. L. Belvini, *op. cit*., p. 175; F. Di Vizio, "L'intelligenza artificiale nelle attività di contrasto del terrorismo", in C. Parodi, T. Rizzo (a cura di), *Intelligenza artificiale e indagini penali*, cit., p. 282 s.

Contra, M. Torre, "Intelligenza artificiale e indagini penali: prospettive future e garanzie di sistema", cit., p. 1747, per il quale non si profila alcun problema quanto alla idoneità probatoria degli strumenti in esame.

todologie di *deep learning*, segue le impenetrabili modalità operative della *black box*"[47]. Risulta dunque inintellegibile la logica sottesa al processo decisionale della macchina, essendo impossibile, tanto per l'utilizzatore quanto per lo stesso programmatore comprendere e spiegare l'*iter* logico seguito dal *software* per addivenire a quello specifico risultato.

Peraltro, la mancata trasparenza sul *set* di immagini adoperate per il *training* della macchina – oltre a rendere ancora più oscuri e incomprensibili i meccanismi sottesi al riconoscimento automatizzato – impedisce di verificare se l'operato del programma di A.I. è inficiato da *input* di partenza errati o poco rappresentativi[48].

Si registrano diversi *bias* che ne intaccano l'attendibilità (si è, ad esempio, dimostrato come alcuni *tool*, essendo addestrati con *dataset* contenenti in prevalenza immagini di volti di uomini bianchi, sono inclini a compiere un maggior numero di riconoscimenti errati – c.d. "falsi positivi" – a discapito delle persone appartenenti a un'etnia diversa da quella caucasica)[49].

D'altra parte, proprio l'"opacità algoritmica" impedisce di fissare in termini assoluti e astratti il tasso di errore delle tecnologie di riconoscimento facciale, atteso che l'accuratezza e l'efficacia del sistema dipende da una serie di variabili imprevedibili[50].

[47] Cfr. L. Belvini, *op. cit*., p. 175; G. Mobilio, *Tecnologie di riconoscimento facciale*, cit., pp. 52 ss.

[48] Così, ancora, L. Belvini, *op. cit*., p. 176. Come osserva G. Mobilio, *Tecnologie di riconoscimento facciale*, cit., p. 221, l'accuratezza del *software* potrebbe essere compromessa "e portare a risultati discriminatori se 'allenati' con dati storici che riflettono pregiudizi impliciti, o se le immagini campionate offrono una rappresentazione statisticamente distorta di gruppi rispetto al complesso della popolazione".

[49] Cfr., sul punto, R. Lopez, "Il riconoscimento facciale tramite *software*", cit., p. 247.

[50] Non sono disponibili dati scientifici riguardo al tasso di errore del Sistema S.A.R.I. Ma alcuni studi hanno dimostrato la totale inaffidabilità dei sistemi di riconoscimento automatizzato del volto: in particolare – come ricorda F. Demartis, *op. cit*., p. 271 e nota 162 – una ricerca condotta da un'*equipe* dell'Università di Essex sul programma di riconoscimento facciale utilizzato dalla polizia londinese ha rilevato tassi di errore che si attestano sulla soglia dell'80%, per lo più dovuti a "falsi positivi".

La concreta capacità dimostrativa dell'*output* restituito dal programma di A.I. è infatti influenzata da una serie di aspetti – qualità dell'immagine, grado di illuminazione, inclinazione e posa del volto, angolo di cattura dell'effigie facciale, età del soggetto, pettinatura, presenza di occhiali, trucchi o altri accessori – che condizionano il risultato finale, a seconda della soglia di sensibilità dello strumento preimpostata da chi addestra il sistema automatizzato[51]. In particolare, più alta è questa soglia e maggiore la sensibilità del sistema, minore sarà il tasso di falsi-positivi (perché l'algoritmo, in ipotesi, tenderà a verificare che ci sia una più esatta coincidenza tra le immagini), ma maggiore sarà il tasso di falsi-negativi (perché in ipotesi, pur essendo le immagini raffiguranti la stessa persona, la diversità nelle rappresentazioni impedisce al sistema di ravvisare una coincidenza)[52].

Dunque, non essendo conoscibile l'esatta procedura con la quale il *software* impara, in autonomia, a leggere e riconoscere i volti delle persone e, non avendo accesso al *dataset* con il quale è stato addestrato il dispositivo, è impossibile verificarne l'affidabilità e, quindi, stabilire se esso è idoneo ad assicurare l'accertamento dei fatti, come richiesto dall'art. 189 c.p.p.

Per quanto concerne, invece, l'altro requisito indicato dall'art. 189 c.p.p., l'impiego del S.A.R.I. "*Real Time*", come potentissimo mezzo per sorvegliare in maniera continuativa e clandestina un numero potenzialmente infinito di persone, è incompatibile con la libertà morale *ex* art. 188 c.p.p.[53].

In tema, v. G. Borgia, *op. cit*., p. 10; G. Borgia, M. Zonaro, "L'errore nel riconoscimento facciale automatizzato", in W. Nocerino (a cura di), *L'errore tecnico e la prova penale*, Pacini, Pisa, 2025, pp. 176 ss.; M. Gialuz, *op. cit*., p. 64.

[51] Sul punto, cfr., tra gli altri, F. Demartis, *op. cit*., pp. 14, 27 s., 105, 240, 280.

[52] Così G. Mobilio, *Tecnologie di riconoscimento facciale*, cit., p. 36.

[53] Cfr. L. Belvini, *op. cit*., p. 177; M. Colacurci, *op. cit*., p. 122 ss.; J. Della Torre, "Tecnologie di riconoscimento facciale e procedimento penale", cit., p. 1070; M. Torre, "Intelligenza artificiale e indagini penali: prospettive future e garanzie di sistema", cit., p. 1748.

In generale, sulla tutela della libertà morale *ex* art. 188 c.p.p. v., per tutti, A. Chelo, *La tutela della libertà morale nella formazione della prova penale*, Wolters Kluwer, Milano, 2025.

Al riguardo occorre sottolineare la straordinaria lesività dei *tool* che, operando in tempo reale e all'insaputa del soggetto interessato, monitorano chiunque transiti, anche soltanto casualmente, in un'area vigilata dalle telecamere munite di *software* di riconoscimento, benché del tutto estraneo alla vicenda sulla quale si sta indagando.

I sistemi di identificazione biometrica finiscono in verità per interferire con l'autonomia delle scelte individuali, in quanto sono in grado di realizzare un marcato *chilling effect* che induce le persone, proprio per il timore di essere soggetti al trattamento biometrico, senza un loro espresso consenso o in luoghi dove non si aspettano di essere sorvegliate, a modificare le proprie abitudini, autolimitandosi nell'esercizio dei propri diritti fondamentali, come ad es. la libertà di riunione o di associazione, al fine di evitare le possibili conseguenze negative che potrebbero derivarne[54].

Del resto, la Corte europea dei diritti dell'uomo ha avuto occasione di sottolineare le pesanti ricadute prodotte sul diritto alla vita privata (art. 8 C.e.d.u.) in relazione alla raccolta e all'elaborazione delle immagini facciali di una massa indistinta di persone che transitano nel raggio di azione delle telecamere "intelligenti"[55]. Il monitoraggio occulto, capillare e prolungato nel

[54] In tal senso v., tra gli altri, L. Camaldo, "Intelligenza artificiale e investigazione penale predittiva", in *Riv. it. dir. proc. pen*., (2024), n. 1, p. 249.

[55] Corte e.d.u., sez. III, 4 luglio 2023, *Gluckhin c. Russia*, in *Proc. pen. giust.*, (2024), pp. 413 ss., con nota di O. Bruno, "La condanna per manifestazione pacifica (non preavvisata e) con riconoscimento facciale viola i diritti fondamentali".

A commento della decisione – che tra i precedenti maggiormente significativi richiama Corte e.d.u., 17 ottobre 2019, *López Ribalda e altri c. Spagna*, §§ 87-88 – v. altresì G. Gallo, "Tecnologie di riconoscimento facciale e diritti fondamentali a rischio: il caso *Gluckhin c. Russia* dinanzi alla Corte europea dei diritti dell'uomo", in *MediaLaws*, (2023), n. 3, pp. 189 ss.; G. Mobilio, "La Corte EDU condanna il ricorso alle tecnologie di riconoscimento per reprimere il dissenso politico: osservazioni a partire dal caso *Gluckhin c. Russia*", in *DPCE on line*, (2024), n. 1, pp. 695 ss.; C. Nardocci, "Il riconoscimento facciale sul 'banco' degli imputati. Riflessioni a partire, e oltre, Corte EDU *Gluckhin c. Russia*", in *Bio law Journal*, (2024), n. 1, pp. 279 ss.

Nel caso di specie, la Corte, pur avendo riconosciuto che tale interferenza era dotata di una base giuridica nel diritto interno, ha stabilito che tuttavia non era fornita di norme adeguate e dettagliate tese a perimetrare, da un lato,

tempo, sganciato da qualsivoglia limite e da garanzie adeguate a prevenire abusi, costituisce un serio attentato per i diritti individuali: sorvegliare in modo costante una platea generalizzata di soggetti che frequentano determinate aree, identificandoli in pochi secondi e a loro insaputa, pregiudica le libertà di espressione, di associarsi e riunirsi, permettendo di svelare una serie di informazioni inerenti alla sfera personale (ad es. convinzioni politiche, ideologiche, religiose).

Alla luce di tutto quanto già rilevato, a maggior ragione, non è possibile allo stato utilizzare i risultati ottenuti dai sistemi di riconoscimento facciale come prova in giudizio. L'ordinamento conosce un mezzo di prova tipico per il riconoscimento, puntualmente disciplinato dagli artt. 213 ss. c.p.p. ovvero la ricognizione di persona[56], e il riconoscimento automatizzato non è in alcun modo equiparabile a tale mezzo di prova: nella ricognizione il ricognitore è un essere umano che abbia già in precedenza visto il soggetto da riconoscere e che, dopo averlo descritto e altri adempimenti preliminari, viene chiamato a riconoscerlo attraverso un esame comparativo tra almeno altri due soggetti a lui somiglianti, così da preservare l'atto da suggestioni o malizie, mentre nel riconoscimento automatizzato il ricognitore è una macchina, che "in tempo reale" o "in differita" emetterà un *ma-*

l'ambito e l'applicazione delle misure che comportano l'uso delle tecnologie di riconoscimento facciale e, dall'altro, le relative garanzie contro il rischio di abusi e arbitrarietà. Ancorché, poi, l'obiettivo dell'interferenza possa considerarsi in astratto legittimo (perché finalizzato alla prevenzione del crimine), le misure adottate erano state ritenute particolarmente invasive (con particolare riferimento alla violazione dell'art. 8 C.e.d.u. accertata dalla Corte di Strasburgo, v. i §§ 58-91 della sentenza).

[56] Sulla ricognizione di persone, limitatamente ai lavori monografici, v. A. Bernasconi, *La ricognizione di persone nel processo penale. Struttura e procedimento probatorio*, Giappichelli, Torino, 2003; M. Bontempelli, *La ricognizione nel processo penale*, Giappichelli, Torino, 2012; A.M. Capitta, *Ricognizioni e individuazioni nel diritto delle prove penali*, Giuffrè, Milano, 2001; S. Cavini, *Le ricognizioni e i confronti*, Giuffrè, Milano, 2015; G. Cecanese, *Confronto, ricognizione ed esperimento giudiziale nella logica dei mezzi di prova*, Edizioni Scientifiche Italiane, Napoli, 2013, spec. pp. 67 ss.; nonché, volendo, N. Triggiani, *Ricognizioni mezzo di prova nel nuovo processo penale*, cit.

tch di corrispondenza o non corrispondenza[57]. L'attività compiuta dai sistemi automatici di riconoscimento facciale "è condotta attraverso processi di trattamento, elaborazione e combinazione dei dati oggettivamente eterogenei da quelli che entrano in gioco nel corso della ricognizione disciplinata dalla legge"[58]. E l'intervento di un soggetto esperto che adopera la macchina o che ne supervisiona il risultato – in ossequio al principio "*human in the loop*" – non trasforma in un giudizio umano o in una valutazione esperta il risultato algoritmico. Non appare, quindi, possibile neppure veicolare i risultati del riconoscimento facciale attraverso la perizia e la consulenza tecnica, anche in considerazione dell'opacità algoritmica cui si accennava prima.

Escluso che il riconoscimento automatizzato possa essere ricondotto al modello legale della ricognizione di persona, data la diversità ontologica tra le due forme di riconoscimento, occorre

[57] In tal senso, v. G. Galluccio Mezio, *op. cit*., p. 647, secondo cui "l'attività epistemica condotta, in maniera automatizzata, dai *software* di riconoscimento facciale non [è] in alcun modo accostabile al riconoscimento compiuto dall'uomo", anche perché tali programmi, al pari di tutti gli applicativi basati sulla c.d. "intelligenza artificiale" tendono a riprodurre e velocizzare funzioni del cervello umano, ma non ne replicano – per quante somiglianze si possano ravvisare tra di esse – i processi e i meccanismi di funzionamento. Sulla necessità di tener distinto il riconoscimento facciale automatizzato dalla ricognizione di persona, v. anche F. Demartis, *op. cit*., pp. 218 ss., p. 269 s. e p. 281, il quale esclude pure che il primo possa essere inquadrato nella pur discussa categoria, di creazione giurisprudenziale, delle c.d. ricognizioni atipiche o informali (p. 222 s.).

Contra, L. Saponaro, *op. cit*., p. 3, secondo la quale la ricognizione personale *ex* art. 213 c.p.p., ma anche l'identificazione operata dalla polizia giudiziaria *ex* art. 349 c.p.p. o l'individuazione effettuata dal pubblico ministero *ex* art. 361 c.p.p., "quali metodi tradizionalmente utilizzati per il riconoscimento di un soggetto, vengono ora convertiti in strumenti di captazione tecnologicamente avanzati" facendo riferimento alle tecniche di riconoscimento facciale.

[58] Così, ancora, G. Galluccio Mezio, *op. cit*., p. 647, il quale osserva: "Per risalire al dato ignoto (la corrispondenza o meno del volto) a partire da un dato noto (le due o più immagini poste in comparazione) il cervello umano ricorre a meccanismi molto complessi e in molti casi ancora sconosciuti ma, indubbiamente, eterogenei rispetto a quelli della macchina. Quest'ultima impiega in maniera automatizzata un coacervo di leggi scientifiche – quelle che governano la matematica, la statistica, l'informatica, l'intelligenza artificiale – che solo un sistema artificiale dotato di elevatissime capacità di calcolo è in grado di governare".

valutare se possa rientrare tra le prove atipiche *ex* art. 189 c.p.p., in quanto prova fondata su una *novel science*: inquadramento possibile ove rispondesse ai requisiti di ammissibilità previsti da tale disposizione, che come già ricordato, esige che la prova atipica richiesta dalle parti sia idonea all'accertamento del fatto di reato e non lesiva della libertà morale della persona fonte di prova.

Al riguardo, si rinvengono, tuttavia, *a fortiori*, gli ostacoli già rappresentati con riferimento all'utilizzo di tali strumenti nel quadro delle indagini atipiche, ovvero la (attuale) carenza di idoneità epistemologica e di validità scientifica dei risultati emessi dai sistemi di riconoscimento facciale, in considerazione del già ricordato elevatissimo tasso di errore riscontrato (oltre che la lesione della libertà morale dell'individuo, quanto meno con riferimento alla modalità di funzionamento "*real time*").

Nonostante tutte le criticità finora evidenziate, sul piano tecnico e giuridico, occorre in conclusione rilevare che risulterebbe ormai del tutto anacronistica una posizione di assoluta chiusura e preclusione nei confronti dell'utilizzo dei sistemi automatizzati di riconoscimento facciale: "il costante progresso nei settori di applicazione dell'intelligenza artificiale e dell'elaborazione delle immagini lascia presupporre che, nel prossimo futuro, "tali sistemi "saranno in grado di eseguire comparazioni tra volti con un livello di accuratezza e affidabilità sempre più elevato"[59].

Insomma, i sistemi automatici di riconoscimento facciale, non soddisfacendo, allo stato, i requisiti che deve avere una

[59] In questi termini, E. Sacchetto, *Tecnologie di riconoscimento facciale*, cit., p. 262. Analogamente, F. Demartis, *op. cit*., p. 272, per il quale "è verosimile che, nel futuro prossimo, una ottimizzazione dei sistemi di riconoscimento facciale possa condurre verso un superamento del vaglio di idoneità della prova dell'accertamento del fatto", anche grazie "alla diffusione di videocamere in 3D". L'A. evidenzia però che "una identificazione univoca non possa realizzarsi neppure in futuro dato che i sistemi automatici di riconoscimento facciale risultano fortemente influenzati dalle modalità di funzionamento delle periferiche di *input* che, nei casi di riconoscimento facciale 'in differita', nella stragrande maggioranza dei casi sono rappresentati dalle telecamere installate in luoghi pubblici o privati che, inevitabilmente, si differenziano per qualità degli obiettivi, dell'angolazione di ripresa e per qualità dell'immagine (anche al buio)".

"prova scientifica"[60] – in quanto non superano attualmente il "*Daubert test*" –, possono considerarsi una prova scientifica *in fieri*. Potranno diventare in futuro una prova scientifica: se saranno raggiunti adeguati tassi di trasparenza, superando l'"opacità algoritmica" e rendendo note le modalità attraverso le quali l'algoritmo è generato; se sarà accertato che in fase di programmazione siano immessi dati sufficientemente rappresentativi e immuni da pregiudizi e condizionamenti; se sarà dimostrato il tasso di errore dell'algoritmo con riferimento ai "falsi negativi" e ai "falsi positivi"; se il procedimento di funzionamento dell'algoritmo sarà sottoposto a procedura di verificazione/falsificazione, secondo la nota teoria di Popper, e accettato dalla comunità scientifica di riferimento[61].

Intanto, appare quanto mai opportuno e urgente, alla luce di tutte le osservazioni sopra esposte, un intervento normativo compiuto, puntuale, rigoroso ed equilibrato, che – in attuazione del Reg. 1689/2024/UE e della l. n. 132/2025 (e in particolare di quanto previsto all'art. 24 di tale legge) – sia in grado di bilanciare in modo ragionevole e adeguato l'efficienza investigativa con la salvaguardia dei diritti fondamentali dell'individuo, garantendo il canone della proporzionalità, così da utilizzare tali sistemi nella fase delle indagini preliminari solo laddove risultino indispensabili, e assicurando la completa trasparenza delle loro modalità di funzionamento, al fine di poterne saggiare l'affidabilità dal punto di vista epistemologico e, al contempo, di preservarne la compatibilità con l'esercizio delle prerogative difensive[62].

[60] In argomento v., *ex multis*, G. Canzio, L. Luparia Donati (a cura di), *Prova scientifica e processo penale*, 3ª ed., Wolters Kluwer Cedam, Milano, 2025; C. Conti, A. Marandola (a cura di), *op. cit.*; G. Carlizzi, G. Tuzet (a cura di), *La prova scientifica nel processo penale*, Giappichelli, Torino, 2018.

[61] In tal senso, F. Demartis, *op. cit*., p. 280 s.

[62] In ordine a questo auspicio v., tra gli altri, con opportuni suggerimenti, L. Belvini, *op. cit*., pp. 180 ss.; L. Camaldo, *op. cit*., p. 249 s.; F. Demartis, *op. cit*., p. 281 s.; E. Sacchetto, *Tecnologie di riconoscimento facciale e procedimento penale*, cit., pp. 267 ss.

Cristiana Valentini

L'IA e la 'scoperta' dei *bias* cognitivi nel processo penale*

Abstract: **AI and the 'Discovery' of Cognitive Biases in Criminal Proceedings.** The authors question the current limits to the use of AI in criminal proceedings, drawing on reflections on the nature of Artificial Intelligence to explore the cognitive biases of training data and, even before that, the cognitive biases of decision-makers in criminal proceedings.

Keywords: Criminal Process, Artificial Intelligence, Human Cognitive Biases.

Sommario: 1. Realtà e fantascienza – 2. IA. Limiti strutturali – 3. *Bias* nei dati di addestramento – 4. Ragionamento o 'pappagallo stocastico'? – 5. Il giudice e gli algoritmi – 6. Il *bias* cognitivo dell'essere umano: quasi uno sconosciuto per il giurista italiano – 7. Due parole sul *bias* cognitivo, ovvero un primo approccio del processualpenalista – 8. *Bias* e sistema penale – 9. I decisori nel procedimento penale – 10. Regole per i processi decisionali.

1. Realtà e fantascienza

A leggere molteplici titoli di recenti studi in materia di rapporti tra sistema penale e intelligenza artificiale, parrebbe di trovarsi sull'orlo di un nuovo mondo, quasi sbilanciati verso le vertiginose prospettive dischiuse da scienziati/scrittori come Isaac Asimov, il cui ultimo volume del notissimo ciclo della Fondazione[1] immagina una realtà nella quale un'intelligenza diversa e superiore in quanto algoritmica ha saggiamente diretto millenni

* Un ringraziamento speciale va a mio nipote, Alessandro Di Matteo, imminente laureando in Ingegneria informatica del Politecnico di Torino, grande amante dell'IA e lucido indagatore dei suoi enigmi. Senza di lui, i suoi entusiasmi e le sue brillanti prese di posizione, non avrei mai potuto scrivere questo saggio.

[1] Ci riferiamo chiaramente al suggestivo epilogo del ciclo della Fondazione, scritto da I. Asimov nel 1986 e pubblicato in Italia con il nome di *Fondazione e Terra*.

di storia dell'umanità attraverso il superamento dei suoi intrinseci difetti.

Non è proprio così, nei fatti; si tratta di una percezione parzialmente errata – certo prodotta da certi tratti mediatici con cui è stato divulgato il fenomeno dell'IA – e l'uso dell'intelligenza artificiale per scopi decisionali nel processo penale[2] parrebbe una realtà in lento divenire, di cui val la pena di curarsi solo a scopo di tempestiva prevenzione o anche a fini elegantemente speculativi, se non fosse per un particolare interessante: la minaccia "robotica" ci consente di estrarre dall'oggi alcuni dati di realtà nei cui confronti siamo rimasti singolarmente ciechi, ad onta degli ampi studi coltivati altrove da molti anni; ci riferiamo ai *bias* cognitivi, ovvero agli errori che la nostra intelligenza "non artificiale" compie continuamente, nella vita privata come in quella professionale, spesso con costi sociali notevoli.

Ma su ciò si tornerà a breve. Per il momento, merita un breve intermezzo, integralmente dedicato alle vivaci sorti dell'IA.

2. IA. Limiti strutturali

Il termine IA – *Intelligenza Artificiale* – è classicamente usato con ampia approssimazione, in ossequio all'originaria rappresentazione dell'idea di intelligenza non umana suggerita dalla celebre frase di *Marvin Minsky*, considerato uno dei padri fondatori dell'intelligenza artificiale moderna: "AI is The Science of Making Machines do Things That Would Require Intelligence if Done by Humans"[3].

Oggi, il riferimento forse più corretto è quello che conferisce la definizione di IA a tutta una serie di modelli matematici basati su antenati comuni, le cosiddette *Neural Networks*, ovvero modelli che si basano su strutture complesse e modulari ispirate dalla biologia.

[2] La prospettiva cambia di molto allorquando ci si confronti, invece, sull'uso degli LLM in sede di interpretazione. Si leggano, in questa prospettiva, le brillanti riflessioni di O. Di Giovine, "Interpretazione umana e robotica (breve riflessione sul possibile ingresso dei Large Language Models nella giurisdizione)", in *www.sistemapenale.it*, 4 settembre 2025.

[3] M.L. Minsky, *Semantic information processing*, The MIT Press, Cambridge, Massachussets, 1969.

Precisiamo: allo stato i modelli statistici più tradizionali, tutt'ora usati al di fuori dello scenario noto al grande pubblico, venivano (e vengono) progettati per modellare entro una certa approssimazione una distribuzione reale partendo da formule note, e l'addestramento – *training* – serve semplicemente a trovare i parametri migliori, ovvero quelli per cui l'approssimazione è nel miglior caso trascurabile.

Con lo sviluppo delle reti neurali, invece, in particolare delle *DNN – Deep Neural Networks*, si è passati a modelli dove la ricerca non è focalizzata sul reperimento diretto di una formula di *input-output* complessiva, quanto piuttosto sullo sviluppo di elementi più piccoli detti *perceptrons*, ispirati al funzionamento biologico dei neuroni e adoperati come mattoncino di base per creare reti estremamente complesse che stratificano milioni di tali elementi al fine di ottenere una plasticità enorme.

L'esito immediatamente intuibile è che la funzione complessiva di *input-output* dipende da tutti i parametri dei milioni di neuroni di cui la rete è composta: mentre nei modelli statistici tradizionali i parametri sono relativamente pochi e possiedono unicamente lo scopo di compiere il cd. *fitting* del modello sui dati reali (adattamento ai dati), qui invece i parametri sono in numero tale che essi stessi vanno a definire il modello, rendendolo quindi estremamente flessibile, ma al contempo privo di una formulazione matematica globale significativa.

È proprio per questo motivo che i modelli moderni di IA vengono spesso definiti "*Black Box*"; non certo a causa di un fantomatico velo oscuro latore di presagi distopici, ma piuttosto quale semplice conseguenza di formule matematiche estremamente complesse e dense di coefficienti fissati a valori apparentemente privi di significato, certo spogliandoli dell'eleganza matematica che possiedono i modelli statistici tradizionali, ma al contempo permettendo una flessibilità enorme, tale da adattarsi ai casi più disparati.

In buona sostanza, questi modelli assumono decisioni senza però essere in grado di fornire accesso al percorso causale che collega *input* e *output*: la decisione è una funzione distribuita su milioni di parametri ad oggi non tracciabili, men che meno in termini, ad esempio, logico-giuridici.

Il significato è chiaro; la potenza predittiva si è impennata, ma al costo della trasparenza della decisione.

3. *Bias* nei dati di addestramento

A partire dalla non-interpretabilità *in primis* logico/matematica degli *output* di tali algoritmi, l'esito è quello di una più agevole propagazione di *bias* ed errori sistemici, portando avanti o addirittura amplificando fenomeni di *disparate impact*; com'è stato efficacemente notato[4], la neutralità degli algoritmi non basta da sola a prevenire effetti (ad esempio) di discriminazione, così spostando un (ulteriore) *focus* sul problema dell'individuazione dei dati e nella loro rappresentazione.

La porzione dei dati utilizzati può essere essa stessa inquinata da *bias* ed errori sistemici dovuti (ad esempio) da discriminazioni avvenute in passato, o comunque non essere rappresentativa della realtà.

Ma non solo. La scelta delle variabili utilizzate è fondamentale, in quanto queste possono essere non rappresentative per il problema trattato o, ancor peggio, anche se apparentemente neutre (ad esempio: codice postale del quartiere di residenza) possono fungere da *proxies* (surrogati) per caratteristiche protette come etnia, genere, orientamento politico, *et coetera*, rischiando quindi di mascherare autentiche disparità di trattamento dietro semplici categorizzazioni.

Paradossalmente problemi siffatti all'interno dei dati rischiano addirittura di "legittimare matematicamente" le ingiustizie, non perché la matematica sbagli, ma perché proprio la sicurezza affidata alla matematica algoritmica, se non correttamente indirizzata, rischia di cristallizzare gli errori presenti nei dati storici.

In ambito penale, questo significa rischio concreto di errori sistemici replicati e amplificati senza meccanismi intrinseci di controllo: il fatto che una decisione sia stata presa in modo sbagliato in passato diventa un problema per l'apprendimento delle macchine, che possono imitare *pattern* decisionali poten-

[4] S. Barocas, A.D. Selbst, "Big data's disparate impact", in *Calif. L. Rev.*, 104 (2016), n. 3, pp. 671 ss.

zialmente pericolosi, senza la capacità di rivalutarli in maniera logica.

L'AI, insomma, apprende passivamente dai dati senza vera comprensione o capacità critica. Come è stato ben rilevato[5], strumenti sofisticati come *COMPAS*[6] possono eguagliare le prestazioni dell'essere umano proprio perché lavorano su base statistica: si tratta, nei fatti, di inferenze empiriche che ricordano le scorciatoie cognitive e i luoghi comuni degli umani, seppur con la maggiore potenza determinata dalla matematica, piuttosto che una forma di ragionamento logico-normativo capace di spiegare e giustificare le decisioni.

D'altra parte, i decisori umani hanno la capacità di riflettere consapevolmente sui propri pregiudizi, applicare principi giuridici astratti e intervenire per correggere errori sistematici; inoltre, anche qualora le decisioni di un giudice non fossero neutrali, questo può motivare la propria opzione con ragionamenti espliciti e contestabili, contrariamente alle possibilità degli algoritmi *black box*.

Infine, anche un intervento mirato sui dati di addestramento appare difficilmente in grado di soddisfare simultaneamente tutte le principali definizioni di equità elaborate in letteratura per i sistemi di classificazione algoritmica. Come è stato dimostrato[7], esiste un'incompatibilità matematica tra alcune metriche di equità comunemente adottate (ad esempio *calibration* e *balance for the positive/negative class*), tale per cui non possono coesistere se non in scenari rari e fortemente vincolati, lasciando quindi il problema aperto.

4. Ragionamento o 'pappagallo stocastico'?

Da qualche anno, come anticipato, abbiamo assistito ad una rivoluzione tecnologica nell'AI, portata dall'innovazione di ar-

[5] J. Dressel, H. Farid, "The accuracy, fairness, and limits of predicting recidivism", in *Science Advances*, 4 (2018), n. 1, p. 5580.

[6] L'acronimo COMPAS è il nome con cui è conosciuto un *software* utilizzato per la valutazione del rischio di recidiva negli Stati Uniti.

[7] J. Kleinberg, S. Mullainathan, M. Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores", in *arXiv:1609.05807*, 17 novembre 2016.

chitetture neurali capaci di assorbire, manipolare e trasformare i dati a vari scopi.

Uno dei prodotti di maggior successo di siffatta evoluzione è configurato dagli *LLM – Large Language Models*, modelli linguistici enormi come ChatGPT e simili – che lavorano con dati testuali (se non anche multimodali), che hanno portato una gigantesca innovazione nel campo del *NLP – Natural Language Processing*, tale da permettere agli utenti un'interazione informale e accessibile da un pubblico non specializzato.

Tuttavia, non alligna in questo fenomeno alcun parto di coscienza virtuale, né tantomeno una parvenza di ragionamento: i modelli linguistici lavorano sulle statistiche testuali, non per logica simbolica o sussunzione normativa; non distinguono tra validità giuridica e plausibilità linguistica; generano *output* verosimili, non giuridicamente fondati.

In pratica, concatenano le parole sulla base di quanto il discorso mantiene la sua coerenza, non sulla base di norme, precedenti o men che meno principi; anche quando sembrano "ragionare", stanno solo riproducendo schemi linguistici appresi, senza alcuna comprensione semantica e men che meno quella capacità di applicare regole astratte a casi concreti che va sotto il nome di sussunzione.

In letteratura numerosi approcci hanno tentato di addentrarsi nel tema cercando di spiegare le differenze e le vicinanze tra questo lavoro linguistico e un "vero" ragionamento, spesso anche in maniera interdisciplinare. Risulta ad esempio interessante l'approccio[8] con cui si è tentato di definire un formalismo lessicale adeguato a parlare di *LLM* in modo da non creare fraintendimenti, aiutandosi con le definizioni linguistiche di 'forma', 'significato convenzionale' e 'intento comunicativo': ciò che viene fatto da noi umani per comunicare risulta infatti il prodotto di una relazione di questi tre elementi, mentre gli LLM lavorano solamente sul primo.

[8] E.M. Bender, A. Koller, "Climbing towards NLU: On meaning, form, and understanding in the age of data", in *Proceedings of the 58th annual meeting of the association for computational linguistics*, https://aclanthology.org/2020.acl-main.463/, luglio 2020, pp. 5185 ss.

In pratica, non solo non ha senso attendersi dalle macchine un intento comunicativo, ma è anche impossibile comprendere quello presente nei dati di *input*, in quanto non gestito come obiettivo dell'addestramento. È solo in situazioni "fuori dall'ordinario" che si notano le criticità di tale approccio, in quanto sarebbe necessaria una vera comprensione alla base per effettuare deduzioni verso direzioni che non rientravano nel dominio di addestramento del modello; senza questo, si rischia di finire nelle cosiddette allucinazioni, dove la coerenza lessicale rimane ma si perde ciò che nei dati di addestramento possedeva un vero intento comunicativo.

Intrigante la prospettiva[9] che continua ad esplorare l'ambiguità di questi sistemi, definendoli "pappagalli stocastici", che appunto imitano *pattern* linguistici, certo mantenendo coerenza su larga scala, senza però un intento comunicativo sottostante diverso da quello attribuito dall'interlocutore.

Tutto ciò di fatto rende ancora più difficile identificare eventuali problemi, in quanto mediati dai nostri stessi occhi con una visione paradossalmente distorta su quello che è la nostra comprensione; quello che sintatticamente e linguisticamente può apparire corretto, è in grado però di distorcere i messaggi trasportati nel sottotesto, che non vengono colti dall'intelligenza artificiale proprio per limiti logici di implementazione.

Il discorso sin qui condotto – diciamolo nel concludere queste brevi note sul fenomeno dell'Intelligenza Artificiale – nulla toglie all'enorme utilità dell'IA quale supporto alle decisioni da assumersi in ambito giudiziario tramite modelli statistici a carattere tradizionale, come sopra descritti, capaci di fornire ed elaborare dati essenziali per ragionamenti sulle probabilità *a priori* di un dato accadimento, come pure nel contesto delle valutazioni del peso delle prove consentite dalla cd. formula di Bayes, su cui si tornerà in seguito.

[9] E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, https://dl.acm.org/doi/10.1145/3442188.3445922, 1 marzo 2021, pp. 610 ss.

5. Il giudice e gli algoritmi

In verità, agli appassionati dell'argomento appare chiaro già oggi – ad anni di distanza da ogni uso effettivo dell'IA nel processo di cognizione – come il cd. diritto vivente abbia scarsa simpatia per le funzioni matematiche decisionali e per la logica computazionale.

Anzi, invero, esiste chi – da esperto – ha già avuto modo di operare una diagnosi di questo fenomeno di rifiuto per l'ausilio degli algoritmi, che è di tale vastità da oltrepassare i confini del sistema penale, andando ad improntare di sé fattispecie di "decisori" tra i più disparati: si tratta di una "ostilità verso gli algoritmi"[10], che include medici e psicologi, giustappunto giudici ed economisti, tutti uniti in una "profonda resistenza alla demistificazione della competenza umana"[11] intrisa, in verità, di venature più morali che razionali.

Un buon esempio di questa avversione si trova per avventura cristallizzato da una rarissima giurisprudenza che prende in esame l'uso difensivo della cd. formula di Bayes, algoritmo decisionale il quale – pur adoperato con una certa frequenza da consulenti e periti nella valutazione esperta della prova scientifica e pur modernissimo e foriero di dense riflessioni in termini di autentica capacità di tutela della presunzione d'innocenza[12] – compare con la rarità degli animali in estinzione tra le sentenze edite, per lo più in modo nascosto e sotteso alla motivazione, in quanto adoperata dal ragionamento peritale che il giudice adotta o respinge.

Sintetizziamo così questa nota formula decisionale, riducendo al minimo le accorte riflessioni di chi ne ha studiato l'essenza in applicazione al giudizio penale[13]: secondo la formula bayesiana applicata alle funzioni decisorie penali, il rapporto fra le probabilità *a posteriori* di due ipotesi logicamente incompati-

[10] D. Kahneman, O. Sibony, C.R. Sunstein, *Rumore. Un difetto del ragionamento umano*, Utet, Torino, 2024, p. 641.

[11] *Ibidem.*

[12] P. Garbolino, *Probabilità e logica della prova*, Giuffrè, Milano, 2014, pp. 85 ss.

[13] Il riferimento è sempre dovuto allo studio di P. Garbolino, *op. cit.*, *passim.*

bili (l'ipotesi dell'Accusa e quella della Difesa che si pongono a spiegazioni opposte del medesimo risultato di una prova, ad esempio: lo stato del cadavere costituisce prova di azione omicidiaria oppure dimostra piuttosto un atto suicida?), è una funzione del rapporto tra le probabilità *a priori* delle ipotesi medesime (quanto è probabile, sulla scorta dei dati statistici rilevanti nello specifico, che si tratti di un omicidio piuttosto che di un suicidio?) e le probabilità condizionate dal risultato della prova medesima (quanto è probabile che si tratti di un omicidio o piuttosto di un suicidio tenuto conto delle prove A-B-C?).

Si tratta di una formula matematica che in realtà, come vedremo meglio nelle pagine successive, rappresenta bene alcuni assiomi fondamentali per un processo decisionale esente da errori cognitivi, ovvero la completezza della piattaforma informativa su cui devono basarsi le decisioni, il sequenziamento delle informazioni medesime e l'attitudine al dubbio sulla reale diagnosticità delle prove; anzi, in effetti, è molto interessante anticipare qui come una delle critiche più ricorrenti della psicologia cognitiva nello studio dei processi decisionali è la rilevante tendenza dei decisori ad evitare la considerazione delle probabilità *a priori*, per una innata tendenza della psiche umana a preferire le spiegazioni causali rispetto a quelle meramente statistiche[14].

Epperò, come sanno bene gli statistici, che amano distinguersi sul punto in statistici "frequentisti" e statistici "bayesiani", la formula di Bayes è utile non solo ad evidenziare le fallacie della mera statistica frequentista, ma anche ad evitare gli scarti logici che si osservano allorquando si seguano tentativi di prescindere dalla statistica frequentista senza il supporto di strumenti altrettanto affidabili.

Il dato interessante è che la diffidenza giudiziale nei confronti dell'algoritmo bayesiano risiede paradossalmente o in una scarsa comprensione dello strumento (certo piuttosto complesso) ovvero nella sua considerazione di arnese logico in grado di condizionare eccessivamente il libero convincimento giudiziale.

Cogliamo qui due esempi: nel primo, l'uso della formula bayesiana proposto dalla difesa a supporto dell'insostenibilità

[14] D. Kahnemann, *Pensieri lenti e veloci*, Mondadori, Milano, 2012, p. 250 e *passim*, qui citato nella versione *e-book.*

logica dell'ipotesi accusatoria recepita dalla sentenza di condanna, è respinto sulla scorta di un argomento che dimostra un'effettiva assenza di cognizioni in merito alla formula bayesiana, che viene palesemente confusa con un'applicazione statistica "frequentista", ovvero esattamente all'opposto della realtà del teorema di Bayes; si scrive infatti che

> appare del tutto infondata la prospettazione di applicare il teorema di Bayes ai fini di valutare la probabilità che un determinato soggetto abbia commesso una determinata azione. Infatti, tale teorema, ampiamente applicato in ambito statistico, inerisce alla valutazione della probabilità che eventi ripetibili e la cui frequenza sia nota possano avvenire o essere avvenuti. La ricostruzione della condotta umana in un dato momento storico implica invece la valutazione di un evento non necessariamente ripetibile e che può rispondere a logiche devianti, irrazionali, al di fuori di logiche ripetitive con la conseguenza che la bassa percentuale di probabilità di un dato esito non implica che lo stesso non possa essere avvenuto[15].

Com'è piuttosto evidente dal testo, qui si confonde la statistica frequentista (probabilità *a priori* di un dato evento) con la statistica bayesiana (che include la valutazione congiunta delle probabilità *a priori* e delle probabilità condizionate dalle prove relative al singolo evento), così come pare piuttosto chiaro che la valutazione della sentenza sia completamente ignara degli studi da tempo compiuti per propiziare l'uso della formula matematica bayesiana come supporto ai decisori penali nel contesto della valutazione di singole prove o multiple.

Nella seconda sentenza che si adduce ad esempio[16], stavolta estraendola da quelle emesse in fase di merito, la difesa aveva criticato la condanna per omicidio volontario rilevando come la consulenza tecnica anatomopatologica disposta dal pubblico ministero avesse dedotto la causa di morte della vittima applicando la formula di Bayes in modo esattamente opposto a quanto predicato dal teorema, per così dire stravolgendo la formula in termini tali che anche uno statistico alle prime armi avrebbe

[15] Cass. Sez. II, 20 dicembre 2018, n. 19138.

[16] Corte d'assise d'appello, L'Aquila, 16 dicembre 2021, Santoleri, inedita.

potuto avvedersi dell'errore: qui la Corte, nonostante la gravità della censura e la grossolanità del vizio, ha preferito trincerarsi nella completa *ignoratio* della censura stessa, omettendola dalla valutazione ad onta del peso della specifica prova, decisiva addirittura al fine primario di valutare se la vittima fosse morta di morte naturale o per violenza altrui.

Di non poco interesse, ancora in questa prospettiva, appare la disamina che compie altra sentenza[17] in cui, invece, il giudice adopera direttamente la formula di Bayes per valutare il peso della prova, giungendo ad un esito assolutorio tutt'altro che scontato, ma inevitabile con una corretta applicazione del teorema, il quale – in verità – dimostra nei fatti di essere un serio baluardo dell'oltre ogni ragionevole dubbio.

Le vicissitudini della formula di Bayes dimostrano come – con un fenomeno tutt'altro che nuovo a livello scientifico[18] – il giudice penale, al pari di tanti altri decisori, spesso rifugge dalle regole stringenti, privilegia le fattispecie attributive di poteri discrezionali e mal sopporta, sulla stessa lunghezza d'onda, gli strumenti logico-scientifici che pongano limiti all'esercizio del libero convincimento.

Torneremo a breve su questo tema, per notare come le nuove acquisizioni della psicologia cognitiva spieghino quest'amore per le fattispecie attributive di poteri discrezionali, dimostrato da fenomeni processuali noti, come il caso dell'istituto dell'esperimento giudiziale, strumento che in teoria si vorrebbe privilegiato da decisori moderni, consapevoli di quanto le necessità cognitive del processo penale possano giovarsi di un mezzo vocato a ricostruire se un *quid* storicamente accaduto trovi riscontro empirico nella sua concreta modalità di accadimento; strumento prezioso in astratto, ma relegato da una prassi annosa al ruolo di mezzo di prova giustappunto ceduto alla discrezionalità giudiziale, e di poi ridotto sulle soglie della non esistenza da una

[17] GIP Milano, 18 giugno 2015, Giud. Gennari, con bella nota di J. Della Torre, "Il 'teorema di Bayes' fa capolino al Tribunale di Milano", in *Dir. pen. cont*., 21 ottobre 2015.

[18] L'argomento sarà oggetto di specifica trattazione nel paragrafo successivo.

quasi costante prassi di rifiuto dell'ammissione per avventura richiesta dalla difesa[19].

6. Il *bias* cognitivo dell'essere umano: quasi uno sconosciuto per il giurista italiano

Torniamo per un attimo al tema dell'IA, per svilupparne quelle che ci appaiono ad oggi le implicazioni più interessanti.

L'avvento dell'IA nel nostro universo di pensiero ci consente, ancora, e anzi a dire il vero ci costringe – proprio nella prospettiva dianzi indicata, ovvero la necessità di addestrare l'IA all'uso dei dati indispensabili al ragionamento giuridico – a riflettere su quello che abbiamo indicato come uno dei grandi problemi che ostano all'uso dell'intelligenza artificiale in ambito decisionale giuridico, ovvero i *bias* nei dati adoperabili in addestramento, tra cui rientrano, senza meno, i precedenti giurisprudenziali.

La vera preoccupazione che deve nutrire le nostre riflessioni sull'IA è in effetti proprio questa: l'IA si nutre di dati storici, proprio come un cervello umano, con l'evidente conseguenza che i *bias* cognitivi contenuti nei dati di addestramento andranno ad inquinare la decisione dell'IA già in partenza.

Riflettiamo, per cominciare, sulle innumerevoli decisioni foriere di esegesi delle norme codicistiche irrelate rispetto ai princìpi, come alle innumerevoli violazioni della legalità processuale che costellano la giurisprudenza di merito e di legittimità, come pure, al contempo, sul fatto che l'addestramento dell'IA costituisce per definizione il regno dei grandi numeri; sembra agevole immaginare la sorte di un addestramento che si svolga su materiale della qualità di quello che allo stato conosciamo; l'esito prevedibile sarebbe la replicazione *ad infinitum* dell'errore umano, per di più legittimato – come sopra si notava – dalla saldezza della matematica algoritmica.

Qui, banalmente, la sfida che propone un futuro uso dell'IA consiste in ciò che comunque dovremmo fare ad oggi, ovvero riuscire nell'arduo compito di risanare la legalità processuale, non foss'altro che ad impedire che l'IA ne impari le deviazioni.

[19] F. Giunchedi, *Gli esperimenti giudiziali. Limiti e potenzialità di un istituto negletto*, Wolters Kluwer Cedam, Milano, 2023, *passim.*

Ma c'è un aspetto ulteriore ed anzi più importante che risulta disvelato dall'approccio al tema dell'intelligenza artificiale: paradossalmente il diffuso timore dei *bias* cognitivi della macchina ci costringe a rivolgere la nostra attenzione al fenomeno dei *bias* cognitivi umani, concetto evidentemente assai diverso da quello di corretta applicazione delle norme giuridiche vigenti, volgendosi direttamente agli errori commessi nella valutazione della prova e di quello che chiamiamo il fatto *sub iudice.*

Prima di addentrarsi nel tema, peraltro, occorre precisare; la riflessione giuridica italiana che comincia ad avviarsi[20], arriva ora a mostrare interesse per un tema d'indagine che, nato attorno agli anni '70 nello specifico settore della psicologia cognitiva e, in particolare, nello studio dei processi decisionali umani[21], si va espandendo a macchia d'olio nell'universo ordinamentale d'oltreoceano, con la complicità delle neuroscienze, imponendosi all'attenzione degli utenti più disparati[22] – medici, psicolo-

[20] Alcuni scritti segnalano un nascente interesse. Volentieri segnaliamo i lavori di R. Blaiotta, "Giustizia, errore, intelligenza artificiale", in *Sist. pen*., 23 ottobre 2023; E. De Paolis, "A carte scoperte su intelligenza artificiale e processo penale. Prospettive di indagine", in *Arch. pen. (web)*, (2025), n. 2; F. Falato, "Il giudice come utente euristico nel ragionamento induttivo basato sul senso comune. Soluzioni di debiasing per assicurare l'igiene decisionale", in *Arch. pen. (web)*, (2025), n. 2. Segnaliamo anche interessantissime applicazioni concrete in seno a precedenti giurisprudenziali di Cassazione; cfr. in particolare Cass., Sez. VI, 3 dicembre 2024, n. 09906, sul cd. *bias* del risultato o del senno di poi, che costituisce uno dei temi più importanti trattati nel volume di Kahnemann già più volte citato, proprio con riferimento ai processi giudiziari. La sentenza in questione è dedicata al processo per il disastro di Rigopiano e trovasi efficacemente annotata da D. Quaranta, "Il problema non è la caduta (della valanga), ma l'atterraggio (del sistema di prevenzione): la Cassazione sul disastro di Rigopiano, tra gestione (anticipata) del rischio ed esigibilità della condotta doverosa", in *Sist. pen*., 22 maggio 2025. Elegantissima, poi, la disamina della sentenza in parola realizzata da F. Consulich, "Disastri evitabili e forme della colpa nell'organizzazione pubblica", in *Cass. pen.*, 2025, p. 2715.

[21] Il primo articolo di A. Tversky e D. Kahneman è stato pubblicato nella rivista *Science* nel 1974, con il titolo *Il giudizio in condizioni di incertezza: euristiche e bias,* ed è stato nuovamente pubblicato in appendice a D. Kahneman, *op. cit.*

[22] È appena il caso di ricordare che il defunto Daniel Kahneman ha vinto il premio Nobel per l'economia nell'anno 2002 e che i suoi studi sulle decisioni in condizioni d'incertezza hanno ispirato interi insegnamenti universi-

gi, economisti, giuristi, uffici governativi di vario ordine e grado – con la potenza di tutto ciò che s'impone non solo per l'autorevolezza della voce, ma prima ancora per l'autoevidenza dei concetti espressi e la loro percettibile centralità nell'esperienza cognitiva umana.

Data la sua attuale alienità rispetto al nostro armamentario concettuale, merita spendere qualche riga in più sul concetto di *bias* cognitivo, prima di verificare se effettivamente (ed in quale misura) esso sia destinato ad esercitare influenza anche nel sistema di giustizia penale.

7. Due parole sul *bias* cognitivo, ovvero un primo approccio del processualpenalista

Sintetizziamo, di seguito, in maniera ovviamente sommaria, le acquisizioni della psicologia cognitiva sul punto, avvalendoci fondamentalmente di alcuni studi[23], che sono lettura ineliminabile per chiunque voglia disvelare le modalità del pensiero umano nella fase decisionale, le sue capacità e le sue mistificazioni, tanto diffuse, quanto svelabili solo tramite un'adeguata conoscenza del fenomeno del *bias* cognitivo.

Precisiamo qui – in abbrivio – una personale convinzione: è tempo che iniziamo a lasciarci alle spalle il solipsismo che ha indotto la dottrina processualpenalistica ad elidere troppo spesso il ricorso a scienze "altre" per arricchire di contenuti necessari la "forma" su cui lavora il processualpenalista, anzi, per contribuire ad individuare proprio la forma più corretta a tutelare gli interessi sottostanti.

Ma, senza ulteriori divagazioni, eccoci al punto.

tari nei Dipartimenti di Economia di molteplici nazioni. Nel mondo nostrano della ricerca, il pensiero dell'illustre autore è penetrato nelle riflessioni degli economisti e dei ricercatori di scienze sociali, mentre non si rilevano approfondimenti nel mondo giuridico, men che meno con riferimento al sistema penale.

[23] Il riferimento è naturalmente al già pluricitato volume di D. Kahneman, *Pensieri lenti e veloci,* come pure al precedente volume di D. Kahneman, P. Slovic, A. Tversky, *Decidere nell'incertezza*, 1982, ed. italiana 2024, per i tipi di Mondadori, come pure al più recente volume di D. Kahneman, O. Sibony, C.R. Sunstein, *op. cit.*, *passim.*

Secondo gli autorevoli e rivoluzionari approdi della psicologia cognitiva cui abbiamo fatto sopra riferimento, le modalità di pensiero umano sono convenzionalmente suddivise in due "sistemi", denominati sistema N. 1 e N. 2, etichetta consuetudinaria che lo stesso Kahneman riferisce ad ulteriori studiosi che l'hanno preceduto, non senza celare come essa sia fornita di tale utilità esplicativa da cancellare il rischio che realmente si pensi alla mente umana come ad un'entità scissa[24].

Il sistema N. 1 è forgiato per fornire risposte veloci, utilissime nella quotidianità personale e professionale; esso opera "in fretta e automaticamente, con poco o nessuno sforzo, e nessun senso di controllo volontario"[25] e ciò anche allorquando compie attività astrattamente complesse, ma magari usuali per il singolo, come una mossa di scacchi. Al contrario il sistema mentale N. 2 è chiamato a compiere attività lente ed impegnative, corrispondenti all'esperienza soggettiva della concentrazione, che esige, per la sua persistenza, assenza di distrazioni.

Il sistema N. 1 adopera, nel corso del suo funzionamento, alcune scorciatoie cognitive, dette "euristiche", ossia procedimenti associativi che consentono di raggiungere un risultato provvisorio, che esigerebbe verifica e infatti l'interazione tra i due sistemi è ovviamente fondamentale, atteso che l'uno compensa l'altro nelle attività cerebrali: il sistema 1 è il più attivo, e produce continuamente materiali d'analisi per il sistema 2, costituiti da impressioni, intuizioni, sensazioni, ragionamenti semplici basati su scorciatoie cognitive, mentre il sistema 2 entra in opera solo allorquando emerga un dato che il sistema 1 non è in grado di elaborare efficacemente, ultimando operazioni complesse sostanzialmente prive di scorciatoie cognitive.

Un dato fondamentale è che il riflessivo sistema N. 2 compie operazioni che esigono sforzo, ed è un'altra delle nozioni che ci derivano dalla psicologia cognitiva, come dalle neuroscienze, il fatto che la stessa chimica del cervello lo induce a "risparmiare energie", accontentandosi delle strategie cognitive intuitive del

[24] D. Kahneman, *op. cit.*, p. 50.
[25] *Ivi*, pp. 51 ss.

sistema N. 1, in base ad una "sofisticata allocazione dell'attenzione…affinata attraverso una lunga storia evolutiva"[26].

Impossibile, qui, non citare ampiamente l'autore, allorquando descrive cosa accade quando l'essere umano si trova dinanzi ad una situazione reale da analizzare a fini decisionali:

> Per il sistema 1 la misura del successo è la coerenza della storia che riesce a costruire. La quantità e la qualità dei dati su cui si basa la storia sono in gran parte irrilevanti. Quando le informazioni sono scarse, cosa che accade spesso, il sistema 1 funziona come una macchina per saltare alle conclusioni… La combinazione di un sistema 1 in cerca di coerenza con un sistema 2 affetto da pigrizia, fa sì che il sistema 2 avalli molte credenze intuitive, le quali rispecchiano fedelmente le impressioni generate dal sistema 1[27].

I *bias* cognitivi nascono proprio dai genetici difetti di un sistema 1 sempre impegnato a reagire con velocità, tramite le scorciatoie cognitive che ne costituiscono ineliminabile caratteristica, provocando il meno possibile l'attivazione dell'oneroso sistema 2, in una inarrestabile compulsione a comprendere (credere di comprendere), scegliere, decidere e agire; il sistema 1 – si scrive[28], suscitando echi che dovrebbero colpire il processualpenalista – non è affatto incline al dubbio, tende anzi, piuttosto, a reprimere ogni ambiguità, lottando per mantenere la coerenza con la prima idea che ha ricevuto al momento del contatto con la situazione.

Queste teorie – accennate già nel prima articolo sul tema, datato al 1974[29] – sono state sviluppate da Kahneman, dal suo gruppo di lavoro e dai colleghi ed allievi nei quarant'anni successivi, scoprendo una quantità di *bias* cognitivi, validandoli *in vitro* attraverso decine di esperimenti e raccogliendo migliaia di dati su decisioni e anche previsioni dell'uomo della strada, come di economisti, finanzieri, politologi, medici, psichiatri, avvocati, giudici ed esperti vari, tra cui quelli di scienze forensi, come la medicina legale o la dattiloscopia.

[26] *Ivi*, p. 93.
[27] *Ivi*, p. 245.
[28] *Ivi*, p. 316.
[29] A. Tversky, D. Kahneman, *op. cit.*, *passim*.

La correttezza delle conclusioni di questi studi può essere sperimentata in proprio da chiunque li legga e possieda l'ineludibile coraggio che occorre per verificare quanto spesso le decisioni che assumiamo, private e professionali, sono il frutto di scorciatoie cognitive foriere di errori.

Tra i principali *bias* analizzati dalla psicologia cognitiva nello studio dei processi decisionali umani, ci limiteremo ad evocare, in virtù della loro importanza per il processo penale, sin d'ora intuibile, il *bias* cagionato dall'euristica della disponibilità, il *bias* della conferma e, infine, il cd. *hindsight bias*, reso in italiano come *bias* del senno di poi o *bias* del risultato.

8. *Bias* e sistema penale

È interessante osservare come in questo complesso sistema che si è cercato di tracciare con mano veloce, la psicologia cognitiva individua, per così dire, una genesi essenziale degli errori cognitivi nel fenomeno cerebrale battezzato con l'acronimo di WYSIATI, siglatura della frase "*What You See Is All There Is*", traducibile in lingua italiana come "quello che vedi è tutto quello che c'è".

Si tratta della scoperta della tendenza ineliminabile dal cd. sistema 1 a decidere ignorando la potenziale incompletezza delle informazioni disponibili, in sostanza accontentandosi di quel che trova e sottostimando, anzi, la rilevanza di quelle potenzialmente mancanti e magari immaginabili, al contempo sovrastimando la qualità delle informazioni già esistenti.

In sintesi, ciascun decisore – inclusi quelli altamente specializzati della materia da decidere, anzi per certi versi proprio quelli altamente specializzati[30] – tende, in maniera del tutto inconsapevole, ad affidarsi ai risultati delle valutazioni espresse dal sistema 1 sulla base di quelle che banalmente chiameremmo le prime impressioni, ad onta del fatto che l'*imprinting* che si ottiene nell'esaminare la materia della decisione, sia stato fon-

[30] In realtà, i decisori esperti risultano, all'analisi della psicologia cognitiva, affetti da molteplici *bias*, tra cui alcuni decisamente caratteristici del *genus,* come il *bias* chiamato "illusione di validità". Sul tema di tornerà appresso, ma sin d'ora può dirsi che esso è diffusamente trattato nel volume di D. Kahneman, O. Sibony, C.R. Sunstein, *op. cit.*, *passim*.

dato su materiale informativo parziale o erroneo. Per converso, sempre il sistema 1 tenderà inesorabilmente (e per quanto se ne sia inconsapevoli) a cancellare i dubbi ricostruttivi che emergerebbero al controllo del sistema 2, cercando conferme piuttosto che smentite rispetto alla valutazione iniziale.

L'esito di questo fenomeno è, giustappunto, l'apertura del decisore verso i *bias* cognitivi.

Come si avverte, il fenomeno cognitivo denominato WYSIATI è stato fondamentale in termini di evoluzione della specie, poiché esso facilita "la realizzazione della coerenza e della fluidità cognitiva che ci inducono ad accettare un'affermazione come vera"[31] e dunque consente di decidere e reagire con rapidità; esso ci rende capaci di pensare e decidere in fretta, sì, ma a costo della perdita di affidabilità nelle scelte che ne derivano, poiché "né la quantità né la qualità delle prove" contano nella prospettiva del sistema 1 e del decisore che ne segue gli *input.*

È difficile sottrarsi all'immediata connessione di queste note con una serie di problematiche tipiche del processo penale, in particolare laddove si consideri quelli che Kahneman individua come "antidoti" allo stato di cose determinato dal WYSIATI: la ricerca di informazioni quantitativamente superiori prima di valutare l'opzione decisoria, il vaglio di qualità sulle informazioni, la diversificazione e l'autonomia delle fonti informative, il sequenziamento delle informazioni, l'accettazione del dubbio, come fonte di scoperta di opzioni decisorie meno intuitivamente disponibili e dunque meno gradite al sistema 1.

Siamo in presenza, in effetti, qui, di un insegnamento che costruisce un fondamento scientifico saldo e ineludibile alle regole del contraddittorio, a quelle in materia di utilizzabilità della prova e, prima di ogni altra, in ordine logico e cronologico, alla negletta regola della completezza delle indagini.

Soffermiamoci su quest'ultimo punto, per una ragione in particolare, ovvero che la carenza informativa e la tendenza del sistema 1 a trarre conclusioni immediate sia pure sulla scorta di dati incompleti, è individuata dalla psicologia cognitiva appunto come luogo d'origine di molteplici *bias* di cui gli attori sono del tutto ignari.

[31] D. Kahneman, *op. cit.*, p. 248.

Il principio di completezza delle indagini – lo sappiamo – è stato a sua volta indicato dalla Corte costituzionale[32] come pre-condizione di un vaglio effettivo tanto sulla decisione di inazione quanto sull'azione esercitata dal pubblico ministero.

Siamo egualmente ben consapevoli di quanto la regola della completezza declinata dalla Consulta e per di più costruita dal legislatore ordinario sul monito codicistico di cui all'art. 358 c.p.p. (e dunque, esplicitamente, sull'obbligo di indagare anche su elementi di prova che ridondino a discarico) sia stata di fatto cancellata dalla prassi quotidiana del sistema giustizia.

Leggendo gli studi sui processi decisionali, riusciamo ora a comprendere come questo tipo di regola schiettamente giuridica (la completezza delle indagini, il dovere di indagare anche su spunti informativi non collimanti con l'ipotesi investigativa) non costituisce solo il frutto dell'applicazione di principi, ma sia una vera e propria regola di "igiene decisionale"[33], con riferimento a tutte le risoluzioni che, nel *sequitur* procedimentale, vengono assunte con riferimento alla data *notitia criminis*: dall'informativa di PG che magari predilige una certa ipotesi ricostruttiva piuttosto che un'altra, all'adozione di un dato strumento investigativo, sino alle richieste del pubblico ministero (cautelari personali o reali, di azione/inazione, etc.), a vedere bene, par chiaro che ogni singola decisione, che costella la fase delle indagini preliminari sino alla sua conclusione, è fitta di possibilità di errori cognitivi, cagionati da euristiche concretamente inevitabili se non si seguono dati precetti, i quali esigono, giustappunto quale pre-condizione di igiene decisionale, la completezza del quadro informativo. E ciò non (solo) perché lo comanda una precisa norma di legge e lo impone un principio di rango costituzionale, ma anche perché, senza il rispetto di quella pre-condizione, è enorme il rischio che l'indagine preliminare si tramuti in una sequela di *bias* cognitivi, che si replicano e si moltiplicano esponenzialmente.

E così, un vaglio non sorvegliato dei primi dati probatori emergenti all'abbrivio delle indagini, genera convinzioni forgia-

[32] Corte cost., sent. n. 88 del 1991.

[33] Così recepisce per il processo penale la terminologia restituita dalla psicologia cognitiva, F. Falato, *op. cit.*, p. 10.

te dal *bias* cd. della disponibilità (scelta veloce sui dati immediatamente disponibili), la quale, a sua volta, genera agevolmente *bias* di conferma, ovvero l'inesorabile propensione del procedimento valutativo verso dati di riprova della convinzione originaria, come pure *bias* di effetto alone (o coerenza emozionale esagerata), che inducono gli inquirenti ad essere letteralmente ciechi[34] rispetto ad elementi conoscitivi incoerenti con l'ipotesi prescelta.

In altre parole, lo scienziato cognitivo spiega al giurista un dato con cui sembra indispensabile fare i conti: la natura unilaterale delle indagini preliminari è capace di generare e, per così dire, incistare all'interno del processo, una serie di *bias* originari, che tendono a trovare conferma all'interno del prosieguo decisionale, ivi compreso quello affidato al giudice e anche in presenza di contraddittorio, non solo perché – come si notava – il *bias* della conferma si autorigenera, provocando ulteriori *bias* di conferma nei decisori successivi, ma anche perché l'unilaterale direzione dell'indagine provoca la dispersione delle potenziali prove a discarico, sempre inesorabilmente compromesse, per definizione, dal decorso del tempo[35].

Non è casuale che lo stesso autore degli studi su cui fondiamo le presenti riflessioni, analizzi questo tipo di *bias* proprio con riferimento a processi penali reali, leggendo nella situazione un precipizio cognitivo verso l'errore giudiziario[36].

9. I decisori nel procedimento penale

Il discorso viene qui appena accennato, nell'intento di cogliere spunti per successivi studi e andrebbe fondato su approfondite ricerche empiriche.

Consideriamo, in effetti, che con il termine 'decisore' la psicologia cognitiva indica il soggetto cui compete la scelta in situazioni incerte, con riferimento al futuro (previsione) o anche

[34] Esempi multipli di cecità cognitive evidenti sono riportati dagli studiosi in tutti e tre i volumi citati in abbrivio.

[35] Lapidaria sul punto Corte EDU, Sezione Prima, sent. 2 marzo 2017, Talpis c. Italia.

[36] D. Kahneman, O. Sibony, C.R. Sunstein, *op. cit.*, *passim,* ma in particolare p. 600.

al passato, come quando occorre valutare quale sia la causa di un dato evento, tra le varie immaginabili (ad esempio: morte naturale o per violenza altrui).

Il sistema penale include importanti decisioni proiettive, *de futuro*, come nel caso in cui il giudice della cautela valuta la possibilità di reiterazione del reato *ex* art. 274, lett. c), c.p.p. o in sede di concessione della sospensione condizionale della pena; ma – come si accennava poc'anzi – l'intero contesto normativo strutturato dal codice di procedura penale è costellato di decisioni in situazione incerta: se il giudice è il decisore per antonomasia, in sede cautelare e tanto più in sede di cognizione, è indispensabile sottolineare come "decisore" sia anche l'ufficiale di polizia giudiziaria che sceglie quale pista investigativa seguire prioritariamente e di conseguenza seleziona il materiale probatorio pertinente al suo approfondimento, non meno del pubblico ministero che delega una determinata attività d'indagine piuttosto che altra, ai sensi dell'art. 370, comma 1, c.p.p. o decide di chiedere al GIP l'autorizzazione a disporre intercettazioni su questa o quell'utenza o l'acquisizione dei dati provenienti dai tabulati; ma decisore è anche il consulente tecnico del pubblico ministero che identifica una certa causa di morte o l'esperto di scienze forensi che decide di poter identificare (oppure no) una corrispondenza con il DNA reperito sulla scena del crimine.

Ciascuna di queste decisioni è potenzialmente gravida di conseguenze, conducendo all'acquisizione di una prova, piuttosto che alla sua dispersione, oppure conduce ad un risultato probatorio falsato da *bias*, sulla cui base vengono assunte iniziative cautelari, magari di carattere restrittivo della libertà personale.

Gli esempi potrebbero continuare e sono tanti quanti gli snodi che costellano un'indagine, la sua conclusione in un senso o nell'altro e, infine, l'intera attività di cognizione di cui consta il giudizio.

Ciascuno dei decisori di questi snodi produce decisioni capaci di interagire con quelle successive, talora influenzandole *funditus*.

Per misurare l'entità del problema dei *bias* e delle loro interazioni con il prosieguo procedimentale è sufficiente por mente ad esempi notori, come la sgargiante attualità del caso di Garlasco,

oppure a casi analizzati dalla dottrina in modo dettagliato, come il processo di Perugia per la morte di Meredith Kercher[37].

Consideriamo, in particolare, i rilievi spesi dal cognitivista su un problema di sistema penale ben noto anche all'interno del nostro ordinamento, ossia la prassi – non codificata, ma costante e ben rilevabile dagli atti – di fornire agli esperti forensi (ad esempio: al consulente tecnico del PM piuttosto che alle sezioni scientifiche delle forze dell'ordine) porzioni intere del fascicolo delle indagini o anche il fascicolo nella sua integralità, unitamente al materiale su cui viene demandato il vaglio.

Anni di studi della psicologia cognitiva dedicati esattamente a questa specifica abitudine, invalsa in molteplici sistemi ordinamentali, anche di *common law*, dimostrano in termini tanto difficilmente confutabili quanto inquietanti, che le decisioni degli esperti forensi sullo specifico tema di prova ad essi affidato mutano in termini drastici in relazione alle informazioni aggiuntive sul caso che vengono ad essi fornite, informazioni pertinenti rispetto al caso medesimo, ma non rilevanti quanto allo specifico tema sottoposto ad *expertise*[38]: nelle centinaia di situazioni esaminate – le più varie: dal confronto delle impronte digitali, all'identificazione di immagini, sino all'analisi anatomopatologica sulle cause di morte – gli esiti diventano di conferma dell'ipotesi investigativa perseguita anche laddove sarebbero stati negativi qualora l'esperto non avesse ricevuto alcuna informazione aggiuntiva oltre a quelle strettamente pertinenti al *thema probandum*.

Si colgano le implicazioni di questo fondamentale passaggio:

> L'effetto delle informazioni condizionanti non è circoscritto alle conclusioni degli esaminatori (identificazione, esclusione o inconcludenza), ma incide *su ciò che percepiscono,* oltre che su *come* interpretano tale percezioni… gli esaminatori posti in un contesto affetto da *bias* non vedono le stesse cose – in senso letterale – di chi non viene esposto

[37] V. sul punto il volume di M. Montagna (a cura di), *L'assassinio di Meredith Kercher. Anatomia del processo di Perugia*, Aracne, Roma, 2012, *passim*.

[38] Usiamo, qui, i termini di pertinenza e rilevanza nello specifico significato ben esplicato da F. Trapella, *Indagini preliminari e parità delle armi*, Wolter Kluwer Cedam, Milano, 2025, pp. 235 ss.

> ad informazioni condizionanti: quando l'impronta latente è accostata alla stampa di un'impronta archiviata *target*, gli esaminatori osservano un numero decisamente inferiore di dettagli (chiamati *minutiae*) rispetto a quanti ne individuano se gli viene mostrata la sola impronta latente… La suscettibilità degli esperti forensi al *bias* di conferma non è un semplice problema teorico, perché sul piano pratico non vengono attuate misure precauzionali sistematiche di nessun tipo per garantire che gli esaminatori non siano esposti a informazioni condizionanti[39].

Come detto, si tratta di rilievi letteralmente inquietanti, soprattutto perché fondati su dati sperimentali ed empirici solidissimi; rilievi che sottendono come le regole processuali siano ad oggi del tutto inadeguate ad impedire fenomeni capaci di condurre direttamente all'errore giudiziario.

Ma spigoliamo ancora, sempre per cogliere spunti dalla psicologia cognitiva con riferimento a temi cari al processualpenalista, come il rapporto tra *media* e processo penale, e notiamo questo aspetto: per il neuroscienziato e per il cognitivista i *media* causano le cd. cascate di disponibilità, determinando l'importanza o al contrario la collateralità di dati eventi, causando illusioni di diffusione di un certo fenomeno magari di fatto episodico, determinando – in modo quasi incancellabile per un normale processo cognitivo – l'immagine (positiva o negativa) dei soggetti di cui si occupano, con conseguente *bias* denominato "effetto alone", e se è facile notare che accurati studi[40] denunziano la rischiosità delle orge mediatiche[41] scatenate in occasione di taluni (troppi) processi penali, purtroppo è ancora diffusa l'illusione che il Giudice professionale, a differenza delle giurie popolari, sia immune al contagio di *bias* cognitivi provenienti dai *media*[42].

[39] D. Kahneman, O. Sibony, C.R. Sunstein, *op. cit.*, p. 593 ss.

[40] Per doviziosi riferimenti v. N. Triggiani (a cura di), *Informazione e giustizia penale. Dalla cronaca giudiziaria al "processo mediatico"*, Cacucci Bari, 2022, *passim*.

[41] Il termine è nuovamente di D. Kahneman, *op. cit.*, p. 24.

[42] Impossibile non ricordare, qui, la decisione emessa dalla Corte EDU proprio con riferimento ad un caso italiano notissimo, ovvero il processo contro Bettino Craxi (Corte EDU, 5 dicembre 2002, n. 34896/97): con assoluta insensibilità per le argomentazioni scientifiche della psicologia cognitiva, pure all'epoca ben note nell'universo di *common law*, la Corte rileva quanto

Non è così, e sono proprio gli studi dei processi decisionali a dimostrarlo in termini letteralmente inconfutabili, anzi: l'(illusoria) convinzione dei decisori esperti in merito alla propria igiene decisionale è un altro *bias* cognitivo, provvisto di nome proprio, guarda caso come "illusione di validità"[43].

10. Regole per i processi decisionali

In effetti, "illusione di validità" è un sintagma alquanto adatto a descrivere quel che rimane delle nostre idee sulla capacità del sistema penale di approdare a decisioni da considerarsi giuste in senso anzitutto etico.

L'insegnamento di Kahnemann *et alia* è imperioso, epperò curiosamente consonante con la mentalità del processualpenalista, quando si rifletta che, per questi studiosi, la cura del male sta nell'adozione di regole per i processi decisionali che riducano il coefficiente di discrezionalità di ciascun decisore, regole che

segue: "Secondo il ricorrente, il suo processo si sarebbe svolto in un clima di ostilità la cui responsabilità ricadrebbe sulle autorità giudiziarie italiane e sui media. La stampa scritta, parlata e televisiva avrebbe stabilito la sua colpevolezza prima della decisione dei giudici. Questa campagna avrebbe influenzato sia l'opinione pubblica, sia i giudici chiamati a pronunciarsi nei vari processi. Peraltro, la procura avrebbe sistematicamente fornito alla stampa ed ai media informazioni coperte dal segreto istruttorio… La Corte rileva che nella fattispecie l'interesse dei media italiani per il caso e l'importanza che questo rivestiva agli occhi dell'opinione pubblica derivava dalla posizione eminente occupata dal ricorrente, dal contesto politico nel quale i fatti incriminati avevano avuto luogo, così come dalla natura e dalla gravità di questi ultimi. La Corte ritiene che sia inevitabile, in una società democratica, che la stampa esprima dei commenti a volte severi su un caso sensibile che, come quello del ricorrente, metteva in discussione la moralità di alti funzionari ed il rapporto tra il mondo della politica e quello degli affari. Inoltre, è il caso di notare che gli organi giudiziari che hanno trattato la causa erano interamente composti da giudici professionali. Contrariamente ai membri di un giurì, questi ultimi godono di un'esperienza e di una formazione che permette loro di fugare qualsiasi suggestione esterna al processo. Peraltro, la condanna del ricorrente è stata pronunciata alla fine di un procedimento in contraddittorio, nel corso del quale l'interessato ha avuto la possibilità di sottoporre alle autorità giudiziarie competenti gli argomenti che riteneva utili alla sua difesa". Argomenti insostenibili, quando si rifletta sulle centinaia di pagine scritte sul punto dalla psicologia cognitiva.

[43] D. Kahneman, *op. cit.*, p. 587.

vengono appunto denominate di "igiene decisionale"; anzi, precisiamo meglio: per il cognitivista, la processualizzazione – si perdoni il gioco di parole – dei processi decisionali, è la chiave di volta per approdare ad una decisione esente da *bias* cognitivi e la discrezionalità dei decisori corrisponde ad una apertura al *bias* tanto più ampia quanto più è vasta l'area libera da regole vincolanti.

Chi scrive ha dedicato anni di ricerca al tema della discrezionalità riservata dal codice di procedura penale a molteplici e importanti decisioni del giudice e del pubblico ministero, evidenziandone i vasti margini di contrasto con i principi costituzionali, l'intrinseca fragilità delle garanzie al confronto concreto con l'esercizio di poteri discrezionali, il valore infinitesimale delle forme di controllo previste. Sotto tale profilo, è un'enorme soddisfazione verificare che questa personale idiosincrasia per uno tra i molteplici snodi critici del processo penale attuale, ha un solidissimo fondamento non solo nei principi del diritto, ma – prima ancora – nella scienza dei processi cognitivi e decisionali.

Alla regola primaria appena illustrata, ne vengono aggiunte altre: mentre andrebbe garantita, mediante sanzioni processuali efficaci, la completezza informativa dell'investigazione nel suo complesso ed andrebbero potenziate le indagini della difesa, ad evitare i *bias* derivanti dalla parzialità della visione inquirente (WYSIATI), i singoli processi decisionali destinati ad incidere sulla decisione finale – tipicamente, le valutazioni degli esperti sulle prove a carattere scientifico – andrebbero sottratte *ab imis* alla possibilità del *bias* della conferma, sigillandole dall'influenza reciproca mediante comunicazione agli esperti dei soli dati direttamente rilevanti per il tema devoluto, evitando la trasmissione a ciascuno dell'intero compendio investigativo. Riflettiamo, qui, in effetti – quanto a questo specifico punto – sul fenomeno che si verifica tutte le volte in cui (ovvero, sempre) al perito nominato in incidente probatorio venga rilasciato il fascicolo delle indagini preliminari nella sua integralità: l'esito inesorabile, mai a sufficienza riflettuto in dottrina, è che addirittura il giudice chiamato a decidere sul merito della causa, viene esposto, tramite le cognizioni ricevute dal perito, ad un *bias* di conferma della convinzione espressa su elementi probatori raccolti durante

le indagini preliminari in assenza di contraddittorio, come se il principio di separazione delle fasi non esistesse affatto.

Ed ecco un'altra regola che dovremmo attentamente meditare: apprezzare le fondamenta della formula decisionale bayesiana in sede di valutazione delle prove, sullo schema essenziale che lo stesso Kahneman propone.

> La prima è che le probabilità *a priori* contano anche in presenza di prove sul caso in questione: è un dato che spesso non è intuitivamente ovvio. La seconda è che l'impressione che abbiamo intuitivamente della diagnosticità delle prove è spesso esagerata. La combinazione del WYSIATI con la coerenza associativa tende a farci credere nelle storie che ci raccontiamo da soli… sono entrambe idee semplici, ma per me fu uno shock quando mi resi conto che non mi avevano mai insegnato ad applicarle e ancora oggi trovo innaturale farlo[44].

Negli studi citati in abbrivio risultano esaminati numerosi casi giudiziari reali e numerosi esperimenti su casi giudiziari fittizi sottoposti ad esperti in cui l'applicazione della regola bayesiana nei semplici termini espressi dallo psicologo cognitivo, è capace (o sarebbe stata capace, nell'ipotesi di processi reali) di condurre ad esiti rispettosi di quelle che, a ben vedere, sono regole di igiene cognitiva che corrispondono, in termini addirittura rigorosi, alle regole del giusto processo, inclusa – ed anzi soprattutto considerata – la presunzione d'innocenza.

Si consideri – in via di mera esemplificazione di ben altre verifiche che il processualista dovrebbe compiere – quanto un principio di diritto come quello espresso dalla notissima sentenza Franzese[45] in materia di causalità, sia distante, nonostante l'apparenza, da un'autentica logica cognitiva giudiziale esente da *bias*.

Secondo quella notissima pronunzia, anche coefficienti di probabilità statistica medio-bassi possono essere utilizzati per riconoscere un nesso di condizionamento tra condotta ed evento, purché vi sia positivo riscontro in merito alla sicura non incidenza, nel caso *sub iudice,* di ulteriori fattori causali interagenti in via alternativa.

[44] D. Kahneman, *op. cit.*, p. 436 s.

[45] Cass. pen., SU, 10 luglio 2002, n. 30328, Franzese.

Formulata con riferimento al reato omissivo improprio, questo principio giurisprudenziale ha avuto enorme fortuna, anche al di fuori dell'alveo, pure assai ampio, del reato colposo; si pensi, ad esempio, all'usatissima formula adoperata nelle decisioni di condanna per variegate tipologie di reati, ma classicamente per l'omicidio volontario, secondo cui la difesa non avrebbe addotto causali alternative idonee a spiegare l'evento concretamente verificatosi[46].

Eppure, nonostante ben si comprendano gli intenti esplicativi della formula/Franzese, una volta letta con lo spirito di chi ha apprezzato le teoriche sui processi decisionali, essa appare come il risultato di un *bias* del risultato o del senno di poi, quando si rifletta che conduce a trascurare le reali probabilità *a priori* di un dato evento (in quel caso: la frequenza statistica dell'evento morte nelle date condizioni), favorendo anche coefficienti statistici insufficienti a superare il ragionevole dubbio.

Queste brevi riflessioni intendono essere l'abbrivio di uno studio più ampio, che dovrebbe investire ogni snodo decisionale del procedimento penale, esaminandone i processi cognitivi sottintesi, al fine di disvelarne la fragilità, la soggezione a scorciatoie cognitive, la vulnerabilità all'uno o all'altro *bias*.

La realtà è che, prima di poter immaginare di nutrire e addestrare un sistema decisionale artificiale con i prodotti delle decisioni umane, dobbiamo compiere ancora un lungo cammino attraverso la legalità processuale, nel tentativo di rifondare l'idoneità cognitiva del procedimento penale.

[46] Inutili i riferimenti giurisprudenziali per questo formulario buono ad ogni uso; la teorica delle ipotesi alternative non dimostrate dalla difesa, è un *leit motiv* che costella sentenze di merito e di legittimità con costante monotonia.

III.
FILOSOFIA

Salvatore Amato

Condizioni processuali e condizionamenti digitali
Un algoritmo per il processo penale?

Abstract: **Procedural Conditions and Digital Influences. An Algorithm for the Criminal Trial?** To what extent can we envision a fully automated justice system? If it were technically possible, would it be socially desirable and morally acceptable? The essay addresses these questions from the perspective of the trial, particularly the criminal trial, exploring its three fundamental aspects: the trial as a ritual, a decision, and an exercise of power.

Keywords: The Future of Law, Agency Entities, Robot-justice, Judiciary, Democracy.

Sommario: 1. *Human in the loop*? – 2. Rito: automazione *vs.* suggestione – 3. Decisione: standard *vs.* individualità – 4. Potere: *loop vs.* magistratura – 5. Conclusione.

1. *Human in the loop?*

Un'indagine di Goldman Sachs del 26, marzo, 2023 (*The potentially Large Effect of Artificial Intelligence on Economic Growth*) sottolinea come oltre il 50% delle attività legali siano soggette a una potenziale sostituzione, parziale o totale, da parte dei sistemi di intelligenza artificiale (IA): di gran lunga superiore ad altri settori come la sanità, l'istruzione, l'intrattenimento. Gli investimenti nel campo del *legal tech* sono aumentati di oltre il 30% sotto l'impulso dell'IA generativa e nella previsione di un mercato che dovrebbe garantire guadagni per diversi miliardi di dollari.

Ormai nessuno può rinunciare all'IA nella ricerca giurisprudenziale, nell'esame di documenti e nell'elaborazione di testi per cui è sempre più difficile fare un elenco completo delle imprese che offrono servizi nell'*e-discovery*. Tutto il settore delle questioni giuridiche minori (transazione, negoziazione, mediazione, conciliazione, arbitrato), tradizionalmente coperto dall'ADR (*Alternative Dispute Resolution*), si sta lentamente converten-

do all'ODR (*Online Dispute Resolution*)[1]. È l'effetto *successful functioning* per cui l'economia digitale e in particolare le *start up* della *fintech* trovano una più immediata e facile applicazione attraverso la giustizia digitale[2].

Negli Stati Uniti gli algoritmi predittivi continuano ad essere ampiamente impiegati nei casi penali, per valutare la pericolosità sociale, in ordine alla determinazione della pena, alla possibilità di recidiva, alla libertà condizionale, all'entità della cauzione e ad altre decisioni largamente discrezionali, malgrado le preoccupazioni in merito al rispetto del giusto processo, all'affidabilità e alla potenziale parzialità di cui si è fatto portavoce lo stesso Presidente della Corte Suprema, John G. Roberts[3].

Le tecnologie dell'*E-carceretion* stanno modificando la funzione della pena e il concetto di difesa sociale. L'attrazione del digitale si manifesta anche a livello legislativo in cui inizia a diffondersi la versione esclusivamente informatica di regolamenti e documenti[4]. "In più, Gpt4 può superare a pieni voti i test standardizzati per l'esame di ammissione alla facoltà di Giurisprudenza"[5].

Il diritto sta, quindi, subendo un processo di cambiamento interno delle tradizionali categorie giuridiche per effetto delle nuove tecnologie, ma deve anche individuare nuove figure contrattuali o istituzionali per assorbire gli effetti di queste tecnologie e per consentirne lo sviluppo. La negoziazione algoritmica ad alta frequenza e gli altri aspetti della finanza tecnologica, la gestione dei dati, le *startup*, la struttura complessa delle *Big Tech*, gli *smart contract*, i bitcoin, la *blockchain* riconfigurano

[1] L. Mingardo, *Giustizia digitale 'alternativa'. Scenari e riflessioni critiche sulle Online Dispute Resolution*, Primiceri, Padova, 2020.

[2] G.I. Zekos, *Advanced Artificial Intelligence and Robo-Justice*, Springer, Cham, 2022.

[3] 2023 Year-End Report on the Federal Judiciary https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf

[4] A. Drahmann, A. Meuwese, "AI and Lawmaking: An Overview", in B. Custer, E. Fosch-Villaronga (eds.), *Law and artificial intelligence. Regulating AI in Legal Practice*, Springer, Berlin-Heidelberg, 2022, pp. 413 ss.

[5] M. Suleyman, M. Bhaskar, *L'onda che verrà. Intelligenza artificiale e potere nel XXI secolo*, Garzanti, Milano, 2024, p. 81.

il mercato e, intanto, si trovano a operare attraverso schemi e modelli concepiti per altri contesti[6].

L'espressione *loop* sintetizza la connessione funzionale di questo eterogeneo mondo digitale e intanto evita di definirne la configurazione. Il *loop* è quel tutto che si sta sviluppando nell'interazione tra macchine sempre più automatizzate e funzioni sempre più interconnesse: le une e le altre operano attraverso il diritto, impongono al diritto le proprie configurazioni e lentamente ne alterano la struttura. Che spazio rimane all'*umano* entro questo ecosistema digitale in continua espansione e trasformazione? *Human in the loop*? Una formula meramente suggestiva "se non ci si fa carico di dire quali siano i confini del *loop*, cosa accade al suo interno, quale sia la posizione che in quest'area si vuole riservare all'umano. Essere nel giro d'accordo, Ma come?"[7].

Il dubbio non emerge solo nelle elaborazioni dottrinali, ma qua e là, anche in alcune decisioni giudiziali.

Mi pare estremamente significativo che nel 2019 la International Commercial Court di Singapore non abbia escluso che alcuni elementi basilari della teoria del contratto vadano ripensati:

> turning to knowledge of the mistake, the law in relation to the way in which ascertainment of knowledge in cases where computers have replaced human actions is to be determined will, no doubt, develop as legal disputes arise as a result of such actions. This will particularly be the case where the computer in question is creating artificial intelligence and could therefore be said to have a mind of its own [8].

La Corte federale australiana ha, invece, ritenuto che una lettera generata automaticamente dall'Ufficio delle imposte, per notificare l'esistenza di un debito fiscale, non costituisca una "decisione" perché una decisione, per essere tale giuridicamen-

[6] C. Alaimo, J. Kallinikos, *Data Rules. Reinventing the Market Economy*, The Mit Press, Cambridge (Ma.)-London, 2024.

[7] A. Santosuosso, G. Sartor, *Decidere con l'IA. Intelligenze artificiali e naturali nel diritto*, il Mulino, Bologna, 2024, p. 202.

[8] Singapore International Commercial Court — Suit No 7 of 2017 Simon Thorley IJ 21–23, 26, 29 November 2018; 28 December 2018 – 14 March 2019 § 206.

te, presuppone un "processo mentale" e quindi implica l'apporto di una persona[9].

L'apertura di Singapore a un probabile futuro integralmente digitale e la chiusura australiana a un'ipotesi del genere ci fanno capire l'estrema delicatezza del momento di passaggio che stiamo affrontando. Delicatezza che l'Ai Act 2024/1689 dell'Unione europea affronta, prevedendo misure diverse a seconda dei livelli di rischio. Chiarisce al comma 8 dell'Allegato III che, tra i sistemi considerati rischiosi dall'art. 6, vanno inseriti anche quelli destinati "…a essere usati da un'autorità giudiziaria o per suo conto per assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti, o a essere utilizzati in modo analogo nella risoluzione alternativa delle controversie". Suggerisce al "considerando" 61 che si dovrebbe operare con estrema cautela perché "l'utilizzo di strumenti di IA può fornire sostegno al potere decisionale dei giudici o all'indipendenza del potere giudiziario, ma non dovrebbe (*should*) sostituirlo: il processo decisionale finale deve (*must*) rimanere un'attività a guida umana". Il futuro si colloca entro la zona grigia che si allungherà nei prossimi anni tra *should* e *must*, tra cautele politiche e pressioni etiche.

Sono tutte dalla parte del *must*, le *Disposizioni e deleghe al Governo in materia di intelligenza artificiale* approvate dal nostro Senato il 20 marzo 2025 che, all'art. 15, relativo all'impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria, prevede "nei casi di impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria è sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti".

Per quanto paradossale possa sembrare, è proprio il divieto di affidarsi integralmente all'IA a suggerire l'eventualità di un orizzonte radicalmente alternativo. Non ha senso vietare quello che sarebbe irrealizzabile o che nessuno vorrebbe. Fino a che punto possiamo ipotizzare una giustizia automatizzata? Se fosse

[9] 139 Pintarich v Federal Commissioner of Taxation (2018) 262 FCR 41.

tecnicamente possibile sarebbe socialmente auspicabile e moralmente accettabile? Vorrei affrontare questo problema nella prospettiva del processo e, in particolare, del processo penale.

Nei tanti scritti che si stanno accumulando in questi anni su questo problema, la domanda sui limiti dell'automazione si somma a quella sulla natura del processo. Interrogandoci sull'IA abbiamo iniziato ad analizzare le varie forme di agentività, finendo per scoprire che non sappiamo bene cosa sia e cosa sarà l'IA perché non sappiamo cosa sia l'intelligenza e fino a che punto la tecnologia del silicio dei nostri computer e la tecnologia del carbonio del nostro cervello siano strutturalmente incomparabili e tendenzialmente incompatibili.

Allo stesso modo, la domanda sull'opportunità di affidare il processo a un algoritmo presuppone che siamo in grado di specificare cosa sia il processo. In realtà, per quanto celebriamo quotidianamente questo rito e per quanto accurate siano le norme che lo disciplinano, avvertiamo, con Satta, che vi è un suo lato misterioso che ci sfugge. Vi ritroviamo lo stesso mistero della vita che, nel susseguirsi caotico degli eventi, deve a un certo punto arrestarsi "se non vuole diventare un torrente folle che tutto travolga e sommerga: l'azione si ripiega su sé stessa, e docilmente, rassegnatamente si sottopone al giudizio"[10]. La scopo del processo è il processo. Garapon ribadisce, in qualche modo, questa prospettiva nel sottolineare come "il processo può essere qualificato come la forma prima della giustizia, in quanto non può essere dedotto da nessun' altra, nemmeno dal diritto"[11].

È, tuttavia, possibile individuare, nella sua evoluzione storica tre aspetti fondamentali: il processo è un rito, una decisione, un esercizio di potere. Nell'equilibrio tra questi tre elementi si sviluppano quei valori che colleghiamo all'idea di giustizia: l'imparzialità, la trasparenza, la fiducia nell'accertamento della verità, il rispetto della dignità individuale, la sicurezza sociale. Come incide il processo di automazione su questi elementi?

[10] S. Satta, *Il mistero del giudizio* in *Soliloqui e colloqui di un giurista*, Padova, Cedam, 1968, p. 15.

[11] A. Garapon, *Processo penale e forme di verità*, Mimesis, Milano, 2024, p. 36.

Quali equilibri altera? Fino a che punto potrebbe garantire la tutela dei valori fondamentali?

2. Rito: automazione *vs.* suggestione

Garapon ci invita a riflettere sull'apparato liturgico che caratterizza il processo. Lo spazio, l'abbigliamento, i tempi, i percorsi, i soggetti, il linguaggio. Tutti questi elementi e il complesso apparato simbolico che li accompagna inducono ad affermare che "il primo gesto della giustizia non è… né intellettuale né morale, bensì architettonico e simbolico: delimitare uno spazio tangibile che tenga a distanza l'indignazione morale e le passioni pubbliche, riservare un tempo a tal fine, fissare le regole del gioco, convenire su obiettivi e istituire gli attori"[12].

Il rito trascende la violenza insita nell'esercizio del potere con cui si decide la sorte di un individuo e intanto la formalizza, la vincola a parametri fissi, rendendola visibile e controllabile. Rifiuto dell'arbitrio e trasparenza attribuiscono a chi celebra il rituale, al giudice, quella superiorità insita nell'essere "altro" rispetto alla contesa in corso. Il giudice, assiso nell'alto del suo scranno, dà forma alla dimensione della terzietà nella scena del giudizio ed è anche il simbolo del valore dell'imparzialità nell'atto del giudicare.

Il processo va oltre le parti, ma lo fa insieme a loro. È questa la sua singolare funzione. Il processo, nella sua struttura ideale, nasce proprio dall'idea che nessuno ha ragione *a priori*. Al contrario siamo autorizzati a "ritenere vera" un'argomentazione solo quando resiste a tutti i tentativi di invalidazione, cioè quando guadagna un consenso che potrà sempre essere rivisto, modificato, integrato. Muovendo dall'insegnamento di Aristotele sull'alterità come dimensione virtuale della giustizia (*pròs étéron estin*)[13], S. Tommaso osserva che l'alterità si realizza nel dialogo e il dialogo impone quel rispetto reciproco e quella fiducia nella verità che alimenta l'attesa di giustizia[14].

[12] A. Garapon, *Del giudicare. Saggio sul rituale giudiziario*, Cortina, Milano, 2007, p. 3.

[13] Et. Nic. V, 3, 1130a.

[14] Eth. 8,9 n. 1658.

Tutto questo non sarebbe possibile senza una dimensione "terza" che rimodula il conflitto tra le parti, traducendo l'aggressività del *vimdicare* nelle ragioni del *vim dicere*[15]. Terzo il giudice. Terzo il luogo. Terzo l'insieme delle regole. Condizioni esistenziali, ma anche condizionamenti psicologici, perché inducono ad accettare l'eventualità di un giudizio negativo. Come insegna Luhmann, la "partecipazione formalizzata… giudizialmente" induce a una ristrutturazione funzionale delle attese "nel senso dell'assorbimento della delusione"[16].

L'automazione sostituirà questo apparato simbolico con *password*, algoritmi, indagini biometriche, valutazioni digitali dello stato emotivo, scannerizzazione di documenti, e, se si vuole ricordare il passato, un monitor con un giudice avatar in un'aula virtuale.

Lo svuotamento dell'apparato liturgico pone il problema dell'efficacia e del valore di una terzietà digitale. Per quanto riguarda l'apparato liturgico penso che abbia ragione Susskind, quando ci invita a mettere da parte l'*emotional and psychological claim*[17]. Attraverso l'eliminazione di tutti quegli espedienti suggestivi che tanta impressione hanno esercitato, nei secoli, sulle anime semplici con "tutto quell'apparato solenne di scranni maestosi, di tocchi, di toghe e di pennacchi"[18], forse avremmo più chiara l'idea che si tratta di un conflitto sociale che va risolto nei modi più corretti e spediti possibili. Garapon osserva in senso critico che, in questa prospettiva, "il diritto non è più l'espressione della volontà sovrana, ma il facilitatore di operazioni puntuali"[19]. Possiamo escludere che sia opportuno accontentarci di questo accomodamento pratico? Ce lo suggeriscono, in qualche misura, tanti pensatori da Hobbes a Holmes…

Non stiamo paragonando un sistema perfetto con un modello sperimentale. Sono evidenti tanto i limiti del rito quanto quelli

[15] Penso alla *legis actio sacramento in rem*, descritta da Gaio (4.16).

[16] N. Luhmann, *La differenziazione del diritto*, il Mulino, Bologna, 1990, p. 199.

[17] R. Susskind, *Online Courts and the future of justice*, Oxford University Press, Oxford, 2019, cap. 27.

[18] Come il Tararà di Pirandello (*La verità* in *Novelle per un anno*, I, Mondadori, Milano, 1956, p. 657).

[19] A. Garapon, J. Lassègue, *La giustizia digitale. Determinismo tecnologia e libertà*, il Mulino, Bologna, 2021, p. 116.

del giudizio umano. Da una parte, la storia ci insegna che quanto più pesano i riti, quanto meno assume rilievo la correttezza del giudizio[20]. D'altra parte, le statistiche, in ogni paese, mostrano impietosamente gli errori giudiziari, la lentezza dei giudizi, la contraddittorietà delle valutazioni. Possiamo, alla luce di tutto questo, rifiutare, *a priori*, la giustizia digitale? I limiti tecnici, come vedremo nel prossimo paragrafo, riguardano soprattutto la dimensione del giudicare. Qui vorrei prendere in esame solo il problema della sostituzione della liturgia con l'automazione.

Susskind enuncia tre valori (*affordability*, *convenience*, *speed*)[21] che la giustizia digitale potrebbe garantire molto meglio di quanto avviene oggi. Chi attende per anni la conclusione della propria vicenda processuale apprezza immensamente l'aggettivo *speed* e lo associa immediatamente ad *affordability*, perché più i tempi si allungano e più aumentano le spese. La gestione informatica dell'insieme di queste procedure sarebbe di gran lunga meno costosa della gestione umana. Anche per l'ordine giudiziario potremmo affermare quello che già iniziamo a constatare per l'ordine pubblico: "una miscela di droni, videocamere a circuito chiuso e robot costerebbe meno di un corpo di polizia e potrebbe persino essere più efficiente e meno incline a fare del male ai sospetti o a prendersela con gli innocenti"[22].

Avremmo, quindi, un guadagno reciproco da parte tanto del cittadino quanto dello Stato. Ci potremo finalmente liberare di quella lentezza del processo che, anche senza evocare il "Soffrite qualsiasi torto ma non venite qui!" della *Casa desolata* di Dickens, è il limite strutturale del nostro apparato giudiziario.

Fino a che punto la funzionalità operativa giustifica il sacrificio delle dimensioni esistenziali con le sue suggestioni simboliche e i suoi rimandi significativi, affidati alle tradizionali coreografie che si sviluppano nelle aule dei nostri tribunali? È, una delle domande, che stanno alla base delle riflessioni di Garapon sulla giustizia digitale. Un "terzo algebrico", a suo avviso, ri-

[20] Cfr. F. Cordero, *Guida alla procedura penale*, Utet, Torino, p. 42.

[21] R. Susskind, *op. cit.*, p. 485 (versione digitale).

[22] F. Pasquale, *Le nuove leggi della robotica, difendere la competenza umana nell'era dell'intelligenza artificiale*, Luiss University Press, Roma, 2021, p. 187.

duce il lavoro degli avvocati e dei giudici "a una prestazione, e spoglia il diritto di qualsiasi dimensione politica"[23].

Seguendo la perplessità di Garapon, mi riservo di esaminare il problema del rapporto con la politica nel quarto paragrafo. Quanto alla "prestazione" dell'avvocato non dobbiamo dimenticare che l'input, per quanto soggetto ai filtri degli schemi digitali, sarebbe sempre umano e quindi suscettibile di una significativa rielaborazione professionale[24]. Influenzare un algoritmo non sarà come convincere un giudice: meno retorica e più logica, meno suggestione e più ragione, meno creazione e più allineamento. Non è detto che tutto questo implichi un peggioramento.

Non credo neppure che una terzietà algoritmica garantisca in misura minore l'imparzialità. Penso al racconto di Platone sul "giudice nudo" nel *Gorgia* (523a e ss.). Ai tempi di Crono capitava spesso che venisse mandato nel Tartaro, tra i reietti, chi non lo meritava, mentre finivano nelle Isole dei Beati i peggiori criminali. Zeus pose rimedio a tutto questo, capendo che l'errore consisteva nel compiere il giudizio prima della morte.

> Chi viene giudicato è vestito, poiché sono ancora vivi… Avviene allora che i giudici si lasciano impressionare da tanto apparato, essendo anche loro vestiti, poiché hanno l'anima velata dagli occhi, dagli orecchi, da tutto l'insieme del corpo. Dovranno essere giudicati da morti, nudi cioè spogli da tutti questi ostacoli. E nudo, cioè morto, dovrà essere anche il giudice, sicché giudicando anima di fronte ad anima… giusto sia il giudizio.

Le progressioni numeriche non hanno vestiti, non hanno corpo, non cedono alle apparenze e alle suggestioni: sono "nude" come il giudice di Platone e sono "senz'anima" come il giudice a cui Montesquieu chiede di spogliarsi di tutto, anche di sé stesso. La terzietà algoritmica, in questa prospettiva, potrebbe

[23] A. Garapon, J. Lassègue, *op. cit*., pp. 117-118.

[24] Non so quanto il processo digitale possa determinare anche una corrispondente automazione dell'attività legale. È un problema che esamina F. Pasquale, "A Rule of Persons, Not Machines: The Limits of Legal Automation", in *The George Washington Law Review*, 1 (2019), n. 87, pp. 1 ss.

rappresentare una forma di imparzialità che va oltre i limiti della natura umana con i suoi vestiti e con la sua anima.

Questa funzionalità *humanless* è compatibile con il nostro senso della giustizia? Penso sia difficile non essere condizionati dalla suggestione della presenza fisica, con le relative formalità, e dalla fiducia nell'operato umano, con la relativa sensibilità. Siamo abituati a collegare il senso di giustizia al senso di responsabilità e stentiamo ad accettare l'incoscienza insita in una procedura automatizzata. Fino a che punto, evocando la coscienza, subiamo un sottile condizionamento antropologico? Se le macchine non agiscono come noi, è per questo che dobbiamo temerle? Non possiamo rispondere a queste domande se non mettiamo da parte tanto l'idea della superiorità umana quanto l'analogia con il cervello umano. "La loro intelligenza va compresa a un livello di astrazione più alto"[25]. Osservando l'IA senza pregiudizi, ci rendiamo conto che si stanno sviluppando entità dotate di un'estrema varietà di potenzialità che, seguendo i propri percorsi e forse proprio per questo, potrebbero risolvere meglio di noi i nostri problemi.

Susskind sostiene che il modo migliore, per affrontare un corretto dibattito morale, consiste nel muovere dall'assunto che "machine can indeed fully replace judges"[26]. Il fatto che vi siano giudici pieni di compassione e di pietà non significa che siano così tutti i giudici. Lo stesso "judges who show compassion or mercy today may do so unevenly, driven perhaps by unconscious bias"[27].

Se diamo ascolto a Balzac[28], per ogni Popinot, che con la virtù di un santo sacrifica tutto al pubblico bene e alla ricerca della giustizia, c'è sempre un Camusot, intrigante, ambizioso, pronto a qualsiasi cosa pur di compiacere il potere. Nella "commedia umana" sono i Camisot a ricevere gli onori e gli incarichi pre-

[25] N. Cristianini, *Machina Sapiens. L'algoritmo che ha rubato il segreto della conoscenza*, il Mulino, Bologna, 2024, p. 104.

[26] R. Susskind, *op. cit.*, p. 509 s. della versione digitale.

[27] *Ibidem*.

[28] Penso a "L'interdizione", in H. de Balzac, *Tre racconti*, Utet, Torino, 1946, pp. 227 e ss. Camisot ricorre più volte in altre opere della *Comédie humaine* a rappresentare il mondo degli ambiziosi senza scrupoli che popolano i salotti parigini.

stigiosi. Raramente la storia della letteratura ci presenta il sistema giudiziario come un ideale luogo di perfezione. Appare, più spesso, come un meccanismo crudele e inesorabile, una macchina che, quando qualcuno ha dato la spinta iniziale, rotola su sé stessa senza pietà. "Justice is a machine that, when someone has once given it the starting push, rolls on of itself"[29], incapace di scorgere "that palpitating life which… always lies behind the commission of the crime"[30], per usare le parole di Galsworthy.

Esasperazioni letterarie? Retaggi del passato? Andiamo a leggere, allora, i brutali resoconti di dieci assurde condanne a morte: alcune tra le tante denunciate dalla Centurion, l'associazione portavoce negli Stati Uniti delle centinaia di persone condannate ingiustamente[31]. Ci fanno riflettere sul fatto che probabilmente il primato della sensibilità umana è solo un ideale, se non addirittura un'illusione. Sono due *Humanless* che dobbiamo mettere a confronto: quella che pratichiamo, da secoli, in tutta "coscienza" e quella che potremmo affidare a un processore matematico.

3. Decisione: standard *vs.* individualità

Gli straordinari sviluppi dell'IA generativa stanno lentamente smentendo l'acquietante formula *agere* sine *intelligere* che sembrava circoscrivere definitivamente le capacità di automazione dei sistemi digitali. L'*acting* si fonda sempre più sul *reasoning* negli sviluppi della catena di pensieri (*chain of thought*), nelle tecniche di apprendimento in contesto (*Chain-of-Thought Prompting*) e nella "umiltà" nel riconoscere i propri errori. Se Gpt3 era come uno studente di Liceo e Gpt4 come uno studente universitario, Gpt5 viene presentata da Altman come un vero esperto. Al di là dell'enfasi mediatica, dobbiamo prepararci a cogliere "forme inedite di azione, controllo, decisione autonoma"[32].

[29] J. Galsworthy, *Justice: a tragedy in four acts*, Duckworth and Co., London, 1910, atto II, p. 30

[30] *Ivi*, p. 26.

[31] J. Grisham, J. McCloskey, *Incastrati. Storie vere e incredibili di condanne ingiuste*, Mondadori, Milano, 2024.

[32] C. Accoto, *Il pianeta latente. Provocazioni della tecnica, innovazioni della cultura*, Egea, Milano, 2024, p. 51.

Tra queste forme inedite rientrano anche le decisioni giudiziali? Per certi versi "there is not (yet) any evidence that robots (are going to) judge"[33]. Per altri versi, mentre rivendichiamo l'indispensabilità di un monopolio umano su una dimensione sociale così delicata come l'esercizio della giustizia, corriamo il rischio che siano gli stessi giudici ad affidarsi furtivamente all'IA generativa, per semplificare il proprio lavoro. Cosa ci suggerisce una recente decisione inglese secondo cui il HMRC (*His Majesty's Revenue and Customs*) deve rivelare se ha utilizzato l'intelligenza artificiale nelle decisioni relative ai crediti d'imposta?[34] Cosa dedurre da quanto denuncia la piattaforma *Bloomberg Law*? Il giudice Julien Xavier Neals, del distretto del New Jersey, avrebbe ritirato una propria decisione per il sospetto che fosse stata scritta da ChatGpt.

È meglio un giudice-robot scelto da noi o un giudice che decide, a nostra insaputa, di delegare a un robot i propri compiti?[35] Una chiara scelta istituzionale è sempre preferibile a un cambiamento imposto da prassi difficilmente controllabili. Questa alternativa potrebbe divenire sempre più pressante con il passare del tempo e con il succedersi delle versioni di IA generativa. È, quindi, indispensabile non cedere alla diffidenza verso le nuove tecnologie. Attualmente sono ancora troppo evidenti i limiti tecnici, per rinunciare all'idea che l'esercizio della giustizia sia così moralmente delicato e politicamente cruciale da non poter essere affidato a una sequenza matematica: "gli esseri umani devono badare l'un l'altro e chiarirsi in un modo anch'esso umano, e non possono delegare un ruolo tanto importante a un *software* irresponsabile"[36].

[33] A.D. Reiling, "Courts and Artificial Intelligence Administration", in *International Journal for Court*, 11 (2020), p. 8, https://iacajournal.org/articles/10.36745/ijca.343, p. 8.

[34] Elsbury v The Information Commissioner (General Regulatory Chambrer) Aug 2, 2025.

[35] "Quando le persone si fideranno maggiormente di avvocati e giudici elettronici, inevitabilmente questi sistemi si diffonderanno, fino a sconvolgere il sistema attuale" (J. Kaplan, *Generative A. I. Conoscere, capire usare l'intelligenza artificiale generativa*, Luiss University Press, Roma, 2024, p. 89).

[36] F. Pasquale, *Le nuove leggi della robotica*, cit., p. 188.

Penso alla lucida analisi di Ferrajoli[37], che pure riconosce gli evidenti vantaggi dell'IA per la facilità e rapidità della circolazione delle informazioni, per l'accesso rapido e immediato ai dati, "per un linguaggio comprensibile, e non quello criptico dei chierici del diritto", per la semplificazione e accelerazione delle procedure, per le capacità predittive e così via. Tuttavia, l'idea "di una giustizia esatta anziché giusta" travolgerebbe l'intero edificio delle garanzie processuali. La cognizione giudiziaria sarebbe, a suo avviso, deformata in tutti i suoi tre aspetti essenziali: l'interpretazione, la prova, l'equità. Una decisione integralmente automatizzata "finirebbe inevitabilmente per favorire la fissità giurisprudenziale l'omologazione, il conservatorismo e il conformismo giudiziario"[38].

Frank Pasquale nelle *Nuove leggi della robotica* sottolinea i rischi di un "sonnanbulismo tecnologico" che induce a confondere la potenza dei computer con la saggezza umana. Nei suoi diversi interventi su questo problema, muove dall'assunto che il diritto non possa essere ridotto a calcolo, come vorrebbero i "*legal automators*" suggestionati dai "*legal futurists*", perché è una combinazione troppo complessa di equità, connessione di concetti e principi, elaborazione ragionata dei rapporti tra le norme e tra le fonti, individuazione di diritti e responsabilità, controlli giurisdizionali, tutele e garanzie[39]. C'è, inoltre, "an inevitable interface of human meaning that is necessary to make sense of social institutions like law. Machine 'learning' is nowhere near simulating it, let alone understanding it in recognizably human terms"[40].

Anche Garapon, soffermandosi sulla complessità epistemologica e sulle implicazioni esistenziali del diritto, sottolinea come sia indispensabile la qualità della *phronesis*, che nessun algoritmo potrà mai possedere, per esprimere non solo la saggezza

[37] L. Ferrajoli, *Giustizia e politica. Crisi e rifondazione del garantismo penale*, Laterza, Roma-Bari, 2024, p. 41.

[38] *Ivi*, p. 43.

[39] F. Pasquale, *A Rule of Persons*, cit., p. 45.

[40] F. Pasquale, G. Cashwell, "Prediction, Persuasion, and the Jurisprudence of Behaviorism", in *Legal Studies Research*, Paper n. 34 (2017), University of Maryland Francis King Carey School of Law, p. 4, http://ssrn.com/abstract=3067737.

pratica, ma la consapevolezza dell'ineliminabile fallibilità dei giudizi umani: nessuna proiezione statistica potrà mai garantire la capacità di mettere al servizio dell'uomo la disponibilità "autenticamente umana di promettere, di impegnarsi a cambiare se stessi…"[41].

Viene anche fatto notare che un insieme di algoritmi non raggiungerà mai l'indispensabile individualità e discrezionalità che è propria di ogni valutazione giuridica, dinanzi alla varietà dei casi, alla contraddittorietà delle norme, alla logica aperta dei principi[42] e non avvertirà neppure la necessità di quegli scarti culturali che impone la consapevolezza della dimensione storica "that trigger spirals aimed and strengthening interests until than unprotected"[43]. Non possiamo neppure ignorare gli eventuali profili di incostituzionalità[44].

Come dimenticare le riflessioni di Roscoe Pound sulla "giustizia meccanica"? Ci invitano a riflettere sul fatto che

> the task of a judge is to make a principle living, not by deducing from it rules… but by achieving oughly the less ambitious but more useful labor of giving an illustration of the intelligent application of the principle to create cause, producing a workable and a just result[45].

Ho accennato solo ad alcune delle perplessità che vengono avanzate nei confronti dei processi decisionali autonatizzati. Del resto, il loro impiego ha finora evidenziato non pochi problemi di affidabilità, se consideriamo il rischio di pregiudizi (*bias*), anomalie (*outliers*), errori (*treacherous turn*), deformazioni (*hal-*

[41] A. Garapon, J. Lassègue, *op. cit*., p. 275.

[42] R. Binns, "Human Judgement in Algorithmic Loops; Individual Justice and Automated Decision Making", in *Regulation § Governance*, 1 (2022), n. 16, pp. 197 ss. https://ssrn.com/abstract=3452030.

[43] D. Della Morte, "Judgements without Judges. The Algorithm's Rule of Law", in G. Riva, A. Marchetti (eds.), *Humane Robotics. A multidisciplinary approach towards the development of humane-centered technologies*, Vita e Pensiero, Milano, 2022, p. 387.

[44] E. Longo, *Giustizia digitale e Costituzione. Riflessioni sulla trasformazione tecnica della funzione giurisdizionale*, FrancoAngeli, Milano, 2023.

[45] R. Pound, "Mechanical Jurisprudence", in *Columbia Law Review*, 8 (1908), n. 8, p. 622.

lucinations), opacità (*black box effects*). Anche i giudizi umani sono, però, affetti dagli stessi difetti.

La stessa opacità, per cui raggiunto un certo livello di sviluppo anche i programmatori non sono in grado di spiegare come si sia determinato uno specifico risultato (*black box effect*), non è poi molto diversa dall'oscuro operare dei nostri pensieri: "We don't know what's going on in a judge's head; it's a black box, too"[46]. L'opacità "umana" è addirittura avallata istituzionalmente dal principio del libero convincimento.

Mentre i limiti delle nostre capacità di giudizio sono rimasti invariati nel tempo, l'intelligenza artificiale generativa sta lentamente riducendo queste carenze di funzionamento. Ad esempio, la possibilità di inserire un codice etico in un codice macchina potrebbe ridurre i dubbi sul rispetto dei principi della presunzione di innocenza e del ragionevole dubbio, sulla ponderazione dell'ammissibilità delle prove, sull'analisi dei fatti e l'esame dei documenti. Non sarebbe da escludere una maggiore coerenza e uniformità di giudizio. Forse anche l'affinarsi degli algoritmi affettivi potrebbe offrire qualcosa di simile ai compiti che affidiamo alla pietà.

Resta poi sempre aperto il dubbio, accennato in precedenza, su quanto le nostre diffidenze non siano condizionate dall'enfasi che poniamo sulla sensibilità umana: la *phronesis* rispecchia l'immagine ideale del singolo giudice, ma poi si può perdere nell'esercizio collettivo delle procedure. Pensiamo solo alla catena del giudizio in un processo. La *phronesis* del pubblico ministero si allunga su quella del gip, del gup, del Presidente del collegio, dei giudici a latere, degli eventuali giurati? Queste *phronesis* si sommano o si elidono in quel gioco indecifrabile che chiamiamo esercizio della giustizia? Ecco il punto: "non si tratta di scegliere tra un algoritmo difettoso e un sistema perfetto immaginario"[47].

Dobbiamo inoltre tener presente che ci troviamo ormai di fronte a una pluralità di IA generative, con una significativa va-

[46] J. Tashea, "Risk-Assessment Algorithms Challenged in Bail, Sentencing and Parole Decisions", *ABA J* (1 Marc 2017), cit. da F. Pasquale, G. Cashwell, *op. cit.*, p. 1.

[47] H. Fry, *Hello Word. Essere umani nell'era delle macchine*, Bollati Boringhieri, Torino, 2019, p. 79.

rietà di risultati e di prestazioni. È quindi difficile continuare a ipotizzare un sistematico appiattimento in giudizi standardizzati dalle equazioni matematiche. Attualmente non sappiamo cosa non sappiamo, ma non possiamo escludere una particolare forma di adattabilità e flessibilità digitale.

L'ipotesi è avallata da un fenomeno ancora in via di accertamento, ma suscettibile di significative implicazioni. Recentemente alcuni ricercatori hanno rilevato che le reti di trasformatori passano bruscamente dal tracciare le posizioni delle parole al concentrarsi sul significato, quando i dati di addestramento superano una soglia critica. Avremmo, se questa analisi sarà confermata, una sorta di transizione di fase che determina un superamento della pura riconnessione sintattica: qualcosa di simile allo scarto intellettivo dei nostri pensieri?

È chiaro che ci riferiamo sempre a macchine per cui le somiglianze con gli esseri umani saranno inevitabilmente approssimative. Possiamo auspicare un'analogia di esiti e non ipotizzare una somiglianza di percorsi. Nelle macchine è lo standard la premessa di sviluppo, negli esseri umani l'individualità. L'algoritmo delle macchine deve essere già scritto; l'intuizione del pensiero viene "riscritta" momento per momento. Forse, una volta messi da parte i nostri pregiudizi, ci troveremo a dover fare una scelta tra *affordability*, *convenience*, *speed*, nella certezza degli standard digitali, oppure *phronesis*, pietà, discrezionalità, nell'indecifrabilità della nostra individualità. In questo caso, credo che sia indispensabile evitare ogni forma di imposizione. In un sistema democratico[48] deve essere il singolo cittadino a scegliere tra l'ipotetica efficienza dei robot e l'ipotetica saggezza degli esseri umani.

4. Potere: *loop vs.* magistratura

Il riferimento alla democrazia ci induce a riflettere sul fatto che, nel rapporto tra umano e digitale, il livello delle singole decisioni va distinto dal sistema decisionale che si viene a deter-

[48] La delega dell'autorità legale a un sistema legale artificialmente intelligente è uno dei problemi cruciali che le società democratiche dovranno affrontare nei prossimi anni, come insegna Lawrence B. Solum, "Artificially Intelligent Law", in *BioLaw Journal*, 1 (2019), § 6.

minare. La "magistratura" come struttura burocratica autoreferenziale ha un'incidenza sociale diversa da quella della singola sentenza. La correttezza di quest'ultima non implica necessariamente l'efficienza dell'altra. Se l'attività del singolo giudice è un servizio reso ai singoli cittadini e all'assetto democratico del paese, la magistratura è un apparato che si inserisce in un complesso sistema di potere.

Come tutti i poteri è suscettibile di eccedere i propri limiti soprattutto nei paesi, come il nostro, in cui funzione requirente e funzione giudicante sono unificate, determinando un monopolio integrale sulla maggior parte degli aspetti relativi all'applicazione della legge. Il problema è accentuato dal fatto che, nella struttura ideale dello stato di diritto, ogni potere dovrebbe avere un contro-potere che ne limita gli eccessi e gli abusi o dovrebbe, almeno, essere chiamato a rispondere dei propri atti. Nella magistratura, invece, l'autonomia e l'indipendenza, protezioni fondamentali nei confronti delle pressioni politiche, hanno determinato una situazione istituzionale di assoluta autoreferenzialità: abbiamo un organismo che si auto-amministra e auto-giudica.

Il riconoscimento di questo straordinario spazio di indipendenza costituisce, nell'asseto costituzionale, una singolare anomalia perché abbiamo un potere che incide profondamente sulla vita sociale, ma che non è un riflesso della società. Appare evidente come sia ormai "sproporzionato il posto che il sistema giudiziario è venuto ad occupare nella vita civile, se rapportato al suo fallimento come erogatore del fondamentale servizio della giustizia"[49].

Tutte le istituzioni umane finiscono per assumere tendenze corporative, sviluppando un proprio codice interno di funzionamento per accrescere le proprie sfere di influenza, eludendo limiti e controlli. Lo pensava anche Montesquieu, quando nel disegnare gli equilibri dei governi "moderati" poneva l'accento sul rischio della "stabilizzazione" degli apparati burocratici. Temeva, in particolare, che "il potere giudiziario, così terribile tra gli uomini", una volta divenuto permanente, si stabilizzasse, chiudendosi in sé stesso, incutendo timore in quanto potere e non in quanto esercizio della giustizia[50].

[49] S. Cassese, *Il governo dei giudici*, Laterza, Roma-Bari, 2022, p. 46.

[50] Ch. de Montesquieu, *Lo spirito delle leggi*, Utet, Torino, 1952, v. I, p. 278.

Con il passare del tempo la stabilizzazione ha cristallizzato l'irresponsabilità, proprio quando sarebbe, invece, stato necessario il fenomeno opposto dinanzi al continuo ampliarsi delle sfere di intervento sulla vita sociale, per effetto del controllo di costituzionalità sulle leggi, della crescente giuridificazione, del ruolo di supplenza nei confronti del potere politico e della pretesa opera di "*coercing virtue*"[51].

La magistratura si è assunta il compito di colmare le lacune del nostro sistema istituzionale, ma nessuno ha colmato le lacune dei suoi meccanismi di funzionamento. Quello che nella prima metà del secolo scorso appariva come il timido accenno di un "governo dei giudici" ha assunto con il passare del tempo il carattere di una "sovranità giudiziaria".

La mancanza di limiti esterni e di limiti interni erode progressivamente il senso originario delle forme di garanzia per cui l'indipendenza assume sempre più i caratteri dell'autogoverno, l'autogoverno sconfina nell'autotutela e l'autotutela appare una pura e semplice manifestazione di potere. In tutti i regimi democratici emerge questa subdola rottura degli equilibri costituzionali con l'affermarsi di un organismo che esercita un potere incontrollabile, pervasivo e condizionante, senza che esistano specifici e definiti contro-poteri: "…il risultato sembra rappresentare una deviazione dal *rule of law*, favorendo un governo arbitrario dei giudici"[52].

È ovvio che gli algoritmi delle future *on line courts* non daranno vita ad aggregazioni corporative, non avranno bisogno di meccanismi di autogoverno, non alimenteranno nessun uso alternativo del diritto. Seguiranno fedelmente le linee guida per cui sono stati programmati, potranno subire continui processi di revisione per correggere gli errori, non proporranno rivendicazioni corporative.

Sotto questo punto di vista la giustizia digitale sarebbe di gran lunga superiore alla giustizia umana. Anche gli algoritmi, però, fanno parte di un sistema: quel complesso di tecnologie

[51] Cfr Robert H. Bork, *Il Giudice sovrano. Coercing virtue*, Liberilibri, Macerata, 2004.

[52] T. Endicott, "Proporzionalità e incommesurabilità", trad. it. in G. Pino, V. Villa (a cura di), *Rule of Law. L'ideale della legalità*, il Mulino, Bologna, 2016, p. 295.

che li crea, li alimenta, li mette in commercio: il *loop*. Un mondo impenetrabile nelle sue svariate articolazioni: le imponenti infrastrutture *hardware* (tra *cloud* e potenza di calcolo) delle *Big Tech*, le enormi quantità di dati fornite dalle *start-up*, la massa di programmatori, addestratori, verificatori che connettono *hardware* e dati, producendo quei *large model* che assistono e condizionano la nostra esistenza.

Un sistema complesso che si complessifica di giorno in giorno. L'opacità degli algoritmi, che si sviluppano autonomamente e sempre autonomamente si connettono tra loro, si aggiunge all'opacità di strutture imprenditoriali che sono "a site of production, a complex system of operations, and a locus of knowledge and capability building as well as a jurisdictional unit"[53]. Se guardiamo le dimensioni, i livelli di capitalizzazione e la quantità di funzioni ci troviamo di fronte a "società multi-agenti" che si affiancano agli Stati nella capacità di condizionare la società e dettare le regole del mercato.

Il problema dell'IA sta divenendo il problema del *loop* che ha prodotto e da cui è prodotta. Una sorta di sistema del sistema che "ingloba l'insieme di strumenti, infrastrutture, modi per connettere, captare e trattare miliardi di dati: cavi sottomarini, satelliti, data center, sistemi di informazione, *software*, algoritmi, super calcolatori ecc."[54].

Il recente Regolamento europeo sull'intelligenza artificiale ci mostra quanto sia difficile dipanare i fili di questo *loop*. Per individuare i destinatari delle procedure di controllo sui livelli di rischio, tenta di scomporre, senza riuscire a esaurirne la portata, questa "famiglia di tecnologie" nella sua dimensione soggettiva (distributore, importatore, *deployer*, fabbricante del prodotto, terzi) e nella sua dimensione oggettiva (dati di addestramento, dati di convalida, dati di prova, dati di input, dati biometrici, dati operativi sensibili). Tutte situazioni, ammesso che siano effettivamente scorporabili, coperte in vario modo dalla confidenzialità degli accordi commerciali o dalla segretezza della proprietà intellettuale.

[53] C. Alaimo, J. Kallinikos, *op. cit.*, p. 116.

[54] A. Mhalla, *Tecnopolitica. Come la tecnologia ci rende soldati*, ADD Editore, Torino, 2025, p. 41.

L'AI Act insiste particolarmente sul riconoscimento di un diritto a una spiegazione chiara ed esaustiva qualora, nei sistemi ad alto rischio, la decisione produca effetti significativi sui diritti fondamentali ("considerando" 171) e prospetta all'art. 14 l'esigenza della "sorveglianza umana", sottolineando l'opportunità di "comprendere correttamente la capacità e i limiti di sistemi di IA ad alto rischio". Tuttavia, usa sempre il condizionale (*should*) per cui l'accenno al "diritto a una spiegazione" del "considerando" 171 diventa, nell'art. 14, la semplice indicazione della "opportunità" di un controllo umano.

Limiti tecnologici sull'interpretabilità degli algoritmi, limiti giuridici determinati dal segreto industriale e commerciale, limiti politici imposti dal crescente potere delle grandi imprese digitali ci fanno capire che il *loop* potrebbe eliminare le distorsioni corporative del funzionamento della magistratura, ma ci immergerebbe in una zona oscura. Chi programma? Chi addestra? Chi scegli i dati? Chi individua i *bench mark*?

Proprio questi interrogativi ci mostrano come non ci troviamo dinanzi a un difetto dell'IA, ma del modo in cui gli esseri umani la stanno sviluppando. Dall'origine militare agli sviluppi commerciali non vi è mai stato un chiaro disegno sociale. Gli evidenti benefici collettivi, che abbiamo tratto dalla sua diffusione, sono stati il corollario necessario delle esigenze del mercato e non il contrario. Dovremo affidare l'esercizio della giustizia a qualche applicazione di Gpt, Claude, Llama, Grok, Gemini? Non riusciamo a eliminare le distorsioni nel funzionamento della magistratura, ma non sappiamo neppure come limitare i condizionamenti delle *Big Tech*. La tecnologia ci apre prospettive che non sappiamo sfruttare e temo che finiremo per dover scegliere il minore dei mali. Il libero convincimento o il *black box effect*? Monarchia giudiziaria o *cloud empires*?

5. Conclusione

È necessario considerare gli algoritmi per quello che sono: uno strumento che potrebbe sostituire efficacemente l'operato umano. Non si tratta di una delega alla magia e neppure di una

glorificazione della statistica[55], ma di un'entità che fa ormai parte integrante della nostra esistenza. La giustizia digitale è solo dei suoi tanti riflessi. Incide però profondamente su uno degli aspetti più delicati della società per cui la sua adozione, nelle situazioni estreme come l'ipotesi delle *online courts*, non può essere imposta dall'alto. Almeno inizialmente, dovrà essere rimesso alla scelta individuale stabilire a quale modello processuale affidarsi.

Un'IA umano-centrica è impossibile senza un ampliamento delle sfere di democrazia, che implica scelte condivise, processi decisionali trasparenti e rivedibili, alla luce del rispetto dei diritti e dei valori fondamentali. Il diritto a non essere costretti a subire un giudizio integralmente automatizzato e il diritto di spiegazione potrebbero essere il presupposto di questo rispetto del vincolo democratico, ma trovano un ostacolo evidente nel groviglio di tecnologie, imprese e interessi che costituisce il *loop* che sta costruendo il nostro futuro.

Non potremo mai sapere fino a che punto sarà efficiente la giustizia digitale, finché non riusciremo a capire come sono determinate le condizioni di questa giustizia e soprattutto da chi. Dietro *l'human in the loop*, dietro l'*humanless* digitale si staglia inquietante questo "chi", fatto di esseri umani e di meccanismi, di esseri umani che acquisiscono potere attraverso questi meccanismi, che forse sono in grado di controllare e forse no[56], ma a loro non interessa nei limiti in cui crescono il potere e gli utili.

L'ideale del "giusto processo", amministrato da una magistratura autonoma e indipendente, nasce all'interno di una società democratica. È difficile pensare che l'ideale di una giustizia digitale possa essere sviluppato da un'oligarchia imprenditoriale a cui interessa, innanzitutto, allargare le proprie sfere di influenza e le relative occasioni di mercato. La colpa non è della tecnologia. Paradossalmente ce lo ricorda proprio lo slogan di una potente lobby dell'apparato tecno-politico, la *National Rifle*

[55] G.I. Zekos, *Advanced Artificial Intelligence and Robo-Justice*, Springer, Cham, 2022, p. 3.

[56] In questa irruzione della fisica in politica "il paradosso è che, pur fondandosi su calcoli sempre più sofisticati, rischia di produrre effetti sempre più imprevedibili e irrazionali" (G. da Empoli, *Gli ingegneri del caos. Teoria e tecnica dell'internazionale populista*, Marsilio, Venezia, 2025, p. 159).

Association: “Guns don’t kill people. People kill people”. Per quanto grandi siano i rischi dell’IA sono sempre minori di quelli determinati dagli esseri umani.

Agata C. Amato Mangiameli, Maria Novella Campagnoli

Intelligenza artificiale e sistema penale. Nuove applicazioni e nuovi paradigmi*

Abstract: **AI and Criminal Justice System. New Applications and New Paradigms.** Anticipatory and predictive forms of governance in the penal system aim to prevent illicit activities and support anti-crime efforts. This is the case of software capable of recognizing suspicious faces and ambiguous behaviors that may raise alarm. The significant developments in Artificial Intelligence place the jurist in front of a multitude of questions and invite the interpreter to balance the enormous possibilities that AI offers and its correct use.

Keywords: Software, Predictive Policing, Data, Surveillance, Law.

Sommario: 1. I *software* di *predictive policing*: una introduzione – 2. Un *excursus* fra strumenti e applicazioni – 3. Dai vantaggi ai rischi: ricordando alcune vicende giudiziarie – 4. Sorvegliare e prevenire al tempo dell'IA.

1. I *software* di *predictive policing*: una introduzione

I significativi sviluppi dell'Intelligenza artificiale pongono il giurista di fronte a una molteplicità di interrogativi, che coinvolgono ogni ambito del diritto, e non ultimo il diritto penale, e che invitano l'interprete a bilanciare tra le enormi possibilità che l'IA offre e il suo corretto uso. Si pensi, ad esempio, ai vantaggi del *City brain*[1], ovvero della piattaforma su larga scala in grado di gestire in modo autonomo i dati e le informazioni generate dentro l'ecosistema urbano, e dall'altra parte alle sfide e alle domande di difficile soluzione che lo stesso cervello cittadino produce, non solo riguardo alla *privacy* dei cittadini, ma anche rispetto alle nuove forme di *governance* (anticipatorie e preditti-

* Agata C. Amato Mangiameli è autrice dei paragrafi 1 e 4 mentre Maria Novella Campagnoli dei paragrafi 2 e 3.

[1] Rinvio a A.C. Amato Mangiameli, *Intelligenza artificiale e Smart City. Applicazioni per potenziare l'efficienza, la sicurezza e la sostenibilità*, in A.C. Amato Mangiameli (a cura di), *Green-smart City: dalla teoria alla pratica*, Giappichelli, Torino, 2025, pp. 1 ss. (in corso di stampa).

ve), nelle quali l'Intelligenza artificiale interagisce con la *policy* (e la macchina amministrativa) ed entra nei processi decisionali, orientando e promuovendo alcune prassi urbane, piuttosto che altre. Di solito, tali forme di *governance* anticipatorie e predittive nel sistema penale hanno il fine di prevenire attività illecite (come, ad esempio, le frodi, si pensi ai sistemi di IA che sorvegliano le transazioni bancarie ed evitano possibili truffe, quale l'uso improprio delle carte di credito) e di sostenere l'attività anticrimine (è il caso dei *software* capaci di riconoscere – processando le migliaia di immagini al secondo delle telecamere di città, aeroporti, stazioni – le facce sospette e i comportamenti ambigui che possono destare allarme).

L'Intelligenza artificiale, e anche l'IA generativa (con capacità di autoapprendimento e decisione autonoma)[2], è uno strumento, e nelle questioni che attengono alla politica criminale può senza dubbio essere un utile mezzo di prevenzione della criminalità, grazie a quei tanti *software* che in tempo reale rielaborano e restituiscono le informazioni e le conoscenze fornite dai centri di big data e dai centri di calcolo ad alte prestazioni, così da valutare e gestire il rischio-reato nei più diversi contesti (privati o pubblici). In particolare, nel privato l'uso di *tools* di IA – a fronte di alcune evidenti criticità, non soltanto in tema di diritto alla *privacy* e di integrità dei dati, ma anche rispetto al criterio di imputazione soggettiva dell'illecito[3] – consente di individuare delle potenziali vicende illecite, vantaggioso ai fini di

[2] Per ulteriori considerazioni v.: A.C. Amato Mangiameli, "Intelligenza artificiale generativa: brevi note", in *L'Ircocervo*, 23 (2024), n. 1, pp. 1 ss. e Ead., "Intelligenza artificiale generativa e metaverso. A proposito di menti e corpi fusi con IA", in *Rivista di Filosofia del diritto*, 14 (2025), n. 1, pp. 45 ss.

[3] Come scrive P. Severino: "l'affidamento agli algoritmi della rilevazione di condotte illecite potrebbe porre problemi di compatibilità con il criterio di imputazione soggettiva dell'illecito all'ente – la colpa di organizzazione – che, ogniqualvolta la commissione del reato presupposto dipenda dalla omessa o erronea segnalazione da parte del *software*, finirebbe per trasformarsi in una sorta di *culpa in eligendo* della società, per essersi dotata di un tool inadeguato ovvero per aver delegato a esso la compliance. Qui difetterebbero invero quelle carenze organizzative che fondano il rimprovero dell'ente, in quanto questi si limiterebbe a adottare un *software* prodotto da altri" ("IA: Rispondere alle sfide con regole adeguate ed efficaci. Ecco come", in *agendadigitale.eu*, 12 luglio 2022).

una identificazione delle aree più esposte al rischio-reato e perciò utile all'aggiornamento del modello adottato. E nel pubblico, nonostante i non pochi aspetti problematici in tema di diritti fondamentali, l'uso di *tools* nelle procedure di *risk detecting* svolge un ruolo fondamentale, giacché gli strumenti di IA, in quel loro andamento continuo e aggiornato, meglio d'ogni altro riescono a mappare i processi, a individuare pericoli e minacce, a valutarne l'impatto e a definire le strategie per mitigare e persino evitare tali rischi. Nel pubblico, inoltre, questi strumenti supportano le attività delle forze dell'ordine (*law enforcement*[4]) e indirizzano verso una più efficace conformità alla normativa penale.

Per il nostro tema (IA e sistema penale) parecchio significativi sono i c.d. *software* di *predictive policing*[5]. Già da molto tempo oggetto d'attenzione della letteratura fantascientifica[6], tali tecnologie hanno conquistato via via, e in particolare a partire dagli anni '90 del secolo scorso con l'affermarsi dell'approccio c.d. *data-driven*[7], uno spazio molto importante e questo grazie alle loro potenzialità che sono innanzitutto quelle di ridurre alcuni tipi di criminalità, prevenendo le attività illecite (ad es. i reati contro la proprietà) e permettendo alle forze dell'ordine di agire in modo proattivo, piuttosto che reattivo. Al contempo, l'uso delle tecnologie di IA consentirebbe una allocazione delle risorse più efficiente, aiutando la polizia a pianificare gli impegni, a diversificare i carichi di lavoro, a gestire costi-benefici in

[4] Sull'argomento si veda l'attenta analisi di A. Ferguson, *The Rise of Big Data Policing. Surveillance, Race, and the Future of Law Enforcement*, NYU Press, New York, 2017.

[5] Sia per la ricostruzione dell'esperienza americana, sia per le prospettive nel sistema penale italiano, cfr. E. Pietrocarlo, "Predictive Policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali", in *sistemapenale.it.*, 28 settembre 2023.

[6] È il caso dei *precog*: metà uomini e metà macchine, tenuti in posizione rigida da bande metalliche e cavi collegati con elettrodi, e in grado di leggere il futuro. Più in particolare, a loro è affidato il compito di prevenire il crimine e punire i colpevoli prima ancora che si macchino di colpe (P. Dick, *Rapporto di minoranza*, trad. it., Fanucci editore, Roma, 2002).

[7] Letteralmente: essere guidati dai numeri e dai dati, piuttosto che da mere opinioni e/o da semplici sensazioni, per prendere decisioni informate, per organizzare al meglio l'attività, per gestire emergenze e imprevisti, tanto nel privato, quanto nel pubblico.

maniera più corretta e in una prospettiva a lungo termine. Sotto questo profilo, si considerino ad esempio quei prodotti di polizia predittiva che raccolgono dati da telecamere di sorveglianza, da vari lettori e da altre fonti digitali, così da contrastare molti fenomeni (compreso il terrorismo) e progettare risposte tattiche contro potenziali e future attività criminali.

Una data importante per le tecnologie di IA è il 2016[8], anno in cui sono stati sviluppati *software* quali: *PredPol* (algoritmo di machine learning, che suddivide la città in griglie)[9], *HunchLab* (algoritmo che gradua i diversi interventi nella città secondo ordini di priorità)[10], *Risk Terrain Modelling* (che utilizza, accanto a quelli già in uso negli altri, anche dati del tipo: prossimità con

[8] È lo stesso anno in cui in una città cinese (Hangzhou) di circa undici milioni di abitanti, viene usata la piattaforma di *cloud computing* di Alibaba, così da tentare di risolvere uno dei problemi della città che è da sempre il traffico. Attraverso il monitoraggio del traffico, l'uso di migliaia di telecamere, il controllo di centinaia e centinaia di incroci, City Brain ha fatto sì che nel giro di un solo anno (2017) la velocità del traffico sia aumentata in modo significativo, i tempi di risposta nei casi di emergenza si siano decisamente ridotti e i parcheggi (legali e illegali) siano costantemente monitorati. Come è intuitivo, tali risultati hanno indotto tante altre città all'uso dello stesso sistema complesso di Intelligenza artificiale, al contempo Alibaba è diventata uno dei leader nel settore delle tecnologie per *smart city* e la sua presenza nel mercato globale, oltre a crescere in modo esponenziale, ha attirato l'attenzione da parte di numerosi governi nel mondo e di altrettante numerose aziende. Di qui il ruolo sempre più importante che le Big Tech rivestono per questa infrastruttura: non solo Alibaba, ma anche GAMAM-Google, Apple, Meta, Amazon, Microsoft, BATH-Baidu, Tencent, Huawei.

[9] Frutto di un progetto di ricerca che ha visto impegnati il Dipartimento di Polizia di Los Angeles e l'Università di Los Angeles, *PredPol* fa uso dei dati CompStat per fornire raccomandazioni lungimiranti sul luogo e sul momento in cui si potrebbero verificare nuovi crimini. Alla base di PredPol vi è un algoritmo che si rifà agli stessi modelli matematici utilizzati nella prevenzione dei terremoti. Allo stesso modo, infatti, delle scosse di assestamento – che seguono la scia di un terremoto –, anche i criminali tenderebbero a realizzare nuovi illeciti nella stessa area precedentemente interessata, oppure in aree immediatamente attigue e/o limitrofe.

[10] Ora parte di ShotSpotter, HunchLab – a differenza di PredPol che si basa quasi esclusivamente su dati storici – tiene conto di un'ampia e diversificata gamma di fattori ambientali, quali: corrispondenza di particolari eventi pubblici, orari di apertura e di chiusura degli esercizi commerciali e, non da ultimo, anche condizioni metereologiche.

fermate della metro, con locali notturni, come pure dati meteo e fasi lunari)[11], e ancora tanti altri (ad es. alcuni sviluppati già qualche anno prima: *Crush*[12], *Compas*[13], *KeyCrime*[14], *Precobs*[15]). Ed è sempre nel 2016 che il dipartimento di polizia di New York ha costituito l'Unità di prevenzione e risposta strategica con il compito, grazie all'incrocio di dati, di identificare gli individui ad alto rischio e di scoraggiare le risposte violente tramite momenti di sensibilizzazione delle comunità a rischio. Ha fatto seguito, nel 2017, un programma predittivo da parte del dipartimento di polizia di Chicago che, sempre grazie all'analisi dei dati archiviati, si è occupata delle aree della comunità ad alto rischio. Simili sforzi in nome della sicurezza pubblica hanno ottenuto importanti risultati, visto che i tassi di omicidio sono diminuiti tanto a New York (5%) quanto a Chicago (23%). A questi dipartimenti vanno aggiunti molti altri. Solo come esempi: quello di Los Angeles (con il già citato *PredPol*, che oggi ha preso il nome di *Geolitica*), quello di Santa Cruz, di Seattle, di Atlanta.

I metodi di polizia predittiva, con i rispettivi *software*, si distinguono secondo tre fondamentali categorie: *i*) quelli relativi al luogo (*place-based*) e perciò volti a individuare le aree nella comunità, e insieme i periodi, che presentano statisticamente un rischio di delinquenza maggiormente elevato; *ii*) quelli connessi

[11] Sistema, questo, che, ad esempio, è stato applicato con successo dal Dipartimento di Polizia di Kansas City per far fronte al problema delle rapine.

[12] Acronimo di *Criminal reduction utilising statistical history*, è un *software*, progettato dalla IBM per la polizia di Memphis. Crush immagazzina le informazioni sui delitti commessi, catalogando i crimini (gli elementi, le caratteristiche, le prove) e confrontandoli, poi, con una serie di informazioni (come l'anagrafe dei criminali conosciuti, il loro comportamento, le soffiate giunte dagli informatori, ecc.). Su richiesta oppure autonomamente, il *software* prospetta agli agenti di polizia previsioni relative a possibili rapine, atti vandalici, furti di auto.

[13] Il *Correctional offender management profiling for alternative sanctions*, l'algoritmo predittivo più famoso e utilizzato dalle corti statunitensi per misurare la percentuale di rischio di recidiva da parte del soggetto considerato. Fra i fattori considerati dal sistema, non solo i precedenti, ma anche le frequentazioni, l'abuso di sostanze, le difficoltà economiche, il grado di istruzione, il contesto sociale e familiare, e, in generale, la personalità del soggetto.

[14] *Infra*, § 2.

[15] *Infra*, § 2.

alla persona (*person-based*), e cioè quei metodi che analizzano e valutano i fattori di rischio legati alla biografia e alle relazioni, come pure alle vicende personali (arresti, recidive) e ai modelli di vittimizzazione; *iii*) quelli riguardanti i gruppi di individui o le reti (*groups-based or network*), ovvero i membri di bande e di famiglie della criminalità organizzata. Va da sé che una tale distinzione venga nella realtà superata, dal momento che è sempre possibile l'incrocio di diversi data storage e che i *software* di *predictive policing* fra loro possono essere interconnessi.

Com'è intuitivo, l'elemento chiave di tutti questi *software* sono proprio i dati che vengono raccolti, per poi essere analizzati, rielaborati, reimmessi. Il sistema non sempre è scevro da pregiudizi, e questi possono viziare gli indici di pericolosità e discriminare sia alcune persone (ad esempio, in base all'etnia, alla religione, ecc.) e sia alcuni gruppi (ad esempio, in base all'appartenenza a un certo ambiente sociale), e una tale ambiguità algoritmica, per di più in presenza di dati c.d. sporchi (*dirty data*) o meramente errati, ha certo delle ricadute sull'operato della polizia predittiva[16]. Del resto, le ipotesi di partenza, i parametri, le funzioni, di un programma possono essere di volta in volta diversi: non c'è quindi un unico modo di produrre un algoritmo e d'altra parte una variazione anche semplice di un parametro o di un dato conduce a risultati diversi, come pure una programmazione superficiale può condurre ad accostamenti bizzarri e a previsioni alquanto singolari. Dietro formule e modelli matematici[17], dietro diagrammi e procedimenti formali, si potrebbero celare meccanismi di alterazione dell'informazione e di condizionamento dell'azione, tali da incidere sui singoli (basti pensare all'accesso al prestito e ai tassi di rientro che dipendono da formule matematiche e i cui parametri non sono riconducibili soltanto alla capacità finanziaria) e sulla collettività (persino il

[16] Cfr. R. Richardson, J.M. Schultz, K. Crawford, "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice", in *New York University Law Review*, 94 (2019), pp. 15-55.

[17] In *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia* (trad. it., Bompiani, Milano, 2018) Cathy O'Neill mette in guardia dall'affidabilità e oggettività degli algoritmi.

destino politico di un paese, o di più paesi, può dipendere dal *software* e da chi possiede i dati sugli elettori).

Tra gli aspetti critici di una raccolta dati di un tipo, piuttosto che di un altro, di una profilazione di un utente o di un cittadino, sono da segnalare l'asimmetria informativa tra una società che offre un servizio e l'utente, l'assenza di trasparenza relativa ai princìpi e ai parametri alla base del funzionamento dell'algoritmo, la creazione di una sorta di *filter bubble*[18], così che siano mostrate all'utente o al cittadino soltanto quelle informazioni che l'algoritmo ha calcolato gli possano interessare, o al contrario, che ha ritenuto per varie ragioni di non dovere fornire. Di qui, l'inesistenza di algoritmi neutrali, ovverosia, di algoritmi che si limitano a riflettere la realtà. Ed anzi gli algoritmi propongono una loro versione pur sempre parziale[19], fatta dalle formule classificanti, dal peso attribuito ai singoli parametri inseriti, dalle procedure che determinano il risultato medesimo. E intanto può accadere che gli algoritmi, proprio per le loro potenzialità discriminatorie, per l'indecifrabilità (l'opacità) del percorso dall'*input* all'*output*, per la creazione di circoli viziosi (ad es.: dalle zone segnalate [*hotspot*] al maggiore pattugliamento e perciò al maggior numero di arresti), elaborino delle previsioni di rischio-reato errate, compromettano i canoni del giusto processo e indeboliscano il principio del contraddittorio nella formazione della prova.

È qui utile ricordare alcuni *software* di polizia predittiva e alcuni casi giudiziari (sono ormai tanti, tra i più noti: *State of Wisconsin v. Loomis*, *United States v. Curry*[20]), che hanno come elemento centrale le dinamiche dell'innovazione tecnologica e le sfide, nell'alternarsi continuo di vantaggi e svantaggi, che i sistemi di IA lanciano nel campo della giustizia penale.

[18] Nozione introdotta da E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, Penguin Group, New York, 2011.

[19] Circa la non-neutralità degli algoritmi, cfr., fra gli altri, K. Crawford, "The Hidden Biases in Big Data", in *Harvard Business Review*, (2013); Ead., *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven, 2021.

[20] *Infra*, § 2.

2. Un *excursus* fra strumenti e applicazioni

Come si è detto, sono veramente tanti i *software* che, nel corso di questi anni, sono stati elaborati e, via via, implementati dai diversi dipartimenti di polizia, esteri così come italiani, sollevando, non di rado, diversi interrogativi e parecchie perplessità[21].

Fra i *software* maggiormente diffusi (e che vanno ad aggiungersi a quelli qui poc'anzi menzionati), può rammentarsi il *Crime Anticipation System* (CAS). Si tratta di un sistema di polizia predittiva che è stato sviluppato dal dipartimento di Polizia di Amsterdam e del quale è stato successivamente esteso l'impiego anche in altre città olandesi, come Maastricht, Groningen, Eindhoven, Breda e Tilburg. Inizialmente pensato per la prevenzione dei c.d. crimini ad alto impatto (come, ad esempio, furti e rapine), CAS viene oggi impiegato anche per scongiurare illeciti minori. In particolare, sfruttando i dati forniti dall'ufficio centrale di Statistica (relativi alle variabili demografiche e socioeconomiche), il *software* è in condizione di effettuare un'analisi predittiva attendibile, a partire dalla quale viene elaborata una sorta di mappatura (con colori diversi che vanno dal rosso al giallo) delle aree che presentano un più elevato rischio criminale. Di qui, la possibilità di procedere a una distribuzione più efficiente ed efficace delle squadre di polizia, in base al grado di rischio-probabilità criminale individuato dall'algoritmo.

Da segnalarsi è anche il *Pre-Crime Observation System* (*Precobs*): un *software*, sviluppato in Germania e ideato dall'*Institute for pattern-based Prediction Technique* di Oberhausen in col-

[21] Cfr. fra i tanti: C. Parodi, V. Sellaroli, "Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco", in *Diritto Penale Contemporaneo*, 6 (2019); S. Brayne, *Predict and Surveil. Data, Discretion and the Future of Policing*, Oxford University Press, New York, 2020; L. Algeri, "Intelligenza artificiale e polizia predittiva", in *Diritto penale e processo*, (2021), pp. 724-734; A. Bonfanti, "'Big data' e polizia predittiva: riflessioni in tema di protezione del diritto alla 'privacy' e dei dati personali", in *MediaLaws*, 3 (2018), pp. 10 ss.; G. Contissa, G. Lasagni, G. Sartor, "Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo", in *Diritto di Internet*, 4 (2019), pp. 619-634; V. Manes, "L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia", in U. Ruffolo (a cura di), *Intelligenza artificiale il diritto, i diritti, l'etica*, Giuffré, Milano, 2020, pp. 547-567.

laborazione con la Polizia di Monaco di Baviera e di Zurigo, che è in grado di pronosticare dove e quando si verificherà un furto con scasso[22]. *Precobs* gestisce i dati storici registrati e cerca automaticamente delle correlazioni statistiche tra diversi elementi comuni, come, l'orario, il luogo, la tipologia dell'unità abitativa interessata (appartamento, casa indipendente, villa a schiera…), le modalità di azione e quelle di accesso allo stabile (forzatura della porta di ingresso, accesso da balconi e/o finestre…). A partire da ciò, l'algoritmo individua ricorrenze, crea corrispondenze, elabora statistiche, che vengono immediatamente trasmesse alla polizia locale, così da permetterle un'azione mirata e, per l'appunto, preventiva[23].

Altro sistema degno di nota è l'*Harm Assessment Risk Tool* (HART)[24], messo a punto dalla polizia del Durham, in collaborazione con l'Università di Cambridge. In uso dal 2017, HART è stato utilizzato soprattutto per valutare se una persona possa essere sottoposta ad un programma (*Checkpoint*) di riabilitazione, alternativo rispetto all'esercizio dell'azione penale. Grazie a una peculiare forma di machine learning (*random forest*), il sistema riesce a prendere in considerazione simultaneamente numerose variabili, che sono connesse al *background* individuale, alla storia criminale del soggetto, alla sua età, al genere, nonché persino ai codici postali di residenza. A fronte di ciò, HART prospetta dei giudizi predittivi volti ad appurare quale è il rischio che il soggetto commetta altri reati nei due anni successivi. In base al risultato fornito dall'algoritmo, il soggetto viene, dunque, associato a tre differenti livelli di rischio di recidiva (alto, modera-

[22] Cfr. W. Hardyns, A. Rummens, "Predictive Policing as a New Tool for Law Enforcement? Recent Developments and Challenges", in *European Journal on Criminal Policy and Research*, (2018), pp. 201-218.

[23] D. Gerstner, *Using Predictive Policing to Prevent Residential Burglary. Findings from the Pilot Project P4 in Baden-Württemberg, Germany, Department of Criminology*, Max Planck Institute for Foreign and International Criminal Law, Germany, 2015.

[24] M. Oswald, J. Grace, S. Urwin, G.C. Barnes, "Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality", in *Information & Communications Technology Law*, 27 (2018), n. 2, pp. 223-250.

to, basso) e, ovviamente, a seconda dell'esito, sarà sottoposto o meno al percorso riabilitativo.

Non da ultimo, merita un cenno *Palantir Gotham*: piattaforma usata per la previsione dei crimini "basata sulla persona". *Palantir*, infatti, non si limita a mappare i luoghi e le zone a rischio, ma crea dei collegamenti complessi fra gli individui, i veicoli, gli indirizzi e altre informazioni disponibili, al fine di individuare la struttura e gli schemi di azione delle organizzazioni criminali. Una piattaforma, questa, che in questi anni è stata oggetto di svariate preoccupazioni e considerazioni critiche in ordine ai possibili rischi connessi alla sorveglianza.

Guardando all'Italia, sono diversi i sistemi di questo tipo che hanno trovato già impego. Il primo, nel 2007, è stato *KeyCrime*: strumento pensato per migliorare l'attività di prevenzione e di repressione dei reati, nonché per razionalizzare le risorse disponibili sui territori, incrementandone l'efficienza e l'efficacia dell'azione. Basato sui dati investigativi raccolti sul luogo del crimine (e riferiti dalle stesse vittime o dai testimoni, oppure acquisiti grazie alle registrazioni delle telecamere di videosorveglianza), *KeyCrime* – che è stato sperimentato con un certo successo dalla Questura di Milano – propone come *output* il *crime linking*, ricostruendo oppure individuando i possibili collegamenti e nessi fra diverse condotte criminose e mettendo indirettamente in evidenza anche eventuali vincoli di continuazione.

C'è, poi, *X-Law*, *software* elaborato dalle Università Federico II e Parthenope di Napoli e basato su un algoritmo euristico, che lavora su base probabilistica. X-Law muove dall'acquisizione di elementi socio-ambientali riconducibili all'area territoriale considerata e relativi ai delitti consumati (che sono stati oggetto di denuncia oppure che sono stati scoperti grazie alle attività di Polizia), ricercando modelli/schemi criminali ciclici e stanziali, al fine di prevenire la ripetizione. Nel dettaglio, grazie ad un sistema geografico informativo X-Law fornisce una mappa di rischio che (all'incirca ogni 30 minuti) prospetta i luoghi e gli orari in cui si potrà consumare un nuovo crimine (con un anticipo anche di due ore rispetto alla consumazione del reato). In uso,

dal 2004 nella città di Napoli, è stato poi esportato in altre città italiani, come Prato, Salerno, Venezia e Parma[25].

Di più recente elaborazione, *Giove* – commissionato dal Dipartimento di Pubblica Sicurezza del Ministero dell'Interno – è un *software* basato su un algoritmo che ha lo scopo di supportare le questure nella prevenzione e nella riduzione dei reati a maggior impatto sociale. Giove analizza in via automatica i dati relativi ai crimini passati e riesce a identificare i luoghi, i momenti e le modalità nelle quali è più probabile che si verifichino reati simili. Il sistema, inoltre, prospetta un set di domande da rivolgere alle vittime, al fine di migliorare la raccolta e la gestione delle informazioni.

Degno di menzione anche *ODINO*, sviluppato dal Comando generale dei Carabinieri in cooperazione con l'azienda *Intellitronika* ed estremamente versatile nello svolgimento delle azioni di monitoraggio e di controllo del territorio, come pure nel compimento delle attività investigative. Basato su di un'applicazione installabile su tablet (di cui sono stati dotate tutte le autovetture dell'arma deputate al pattugliamento) o su uno smartphone, *ODINO* è anche in condizione di interagire in tempo reale con diverse banche dati (come, ad esempio, quella della motorizzazione civile e dell'ANIA), consentendo interrogazioni estemporanee che possono essere avviate anche semplicemente inquadrando la targa del veicolo considerato. *ODINO*, inoltre, è in grado di: appurare la veridicità dei documenti esibiti bypassando le verifiche via radio; geolocalizzare il luogo in cui si trova l'operatore in quel preciso momento riducendo notevolmente i tempi di intervento; interagire con allarmi e sistemi di videosorveglianza, registrando e inviando foto e video. Attività di controllo e, soprattutto, di prevenzione che il *software* agevola grazie alla capacità di acquisire, organizzare e collegare fra loro enormi moli di dati eterogenei[26].

Non da ultimo, va ricordato *SARI* (*Sistema automatico riconoscimento immagini*): un sofisticato *software* di riconoscimento

[25] Per un veloce commento cfr., fra gli altri, M. Iasselli, "X-LAW: la polizia predittiva è realtà", in *Altalex.it*, 28 novembre 2018.

[26] Simile ad *ODINO*, ma in uso presso la Polizia di Stato, è il *software Mercurio*.

facciale sviluppato dalla Polizia di Stato. *SARI* permette di identificare in un lasso temporale ristrettissimo un soggetto a partire anche semplicemente da un solo fotogramma, che – grazie agli algoritmi – viene raffrontato con le informazioni disponibili nelle banche dati dell'*AFIS* (*Automatic Fingerprint Identification System*), che raccoglie tutti i dati biometrici e le fotografie dei soggetti fermati dalle Forze dell'Ordine a fini preventivi o repressivi, oppure con le immagini rilevate dalle videocamere di sorveglianza collocate in una certa area[27].

Come è evidente anche *software* di questo genere – che sul piano pratico possono rivelarsi senza dubbio utili – suscitano però anche tutta una serie di quesiti in ordine alla possibile insorgenza di discriminazioni generate dagli stessi algoritmi e ai rischi per la tutela dei diritti. A questo proposito, va detto che se, per un verso, l'*AI Act* vieta espressamente l'uso di sistemi di polizia predittiva e di identificazione biometrica ai fini di sorveglianza sociale[28], per un altro verso, l'uso di sistemi di questo

[27] A proposito delle possibili implicazioni e criticità derivanti dall'impiego di *SARI*, cfr., fra gli altri: R. Lopez, "Riconoscimento facciale tramite *software* e individuazione del sospettato", in A. Scalfati (a cura di), *Le Pre-investigazioni*, Giappichelli, Torino, 2020, p. 298; Garante per la protezione dei dati personali, *Parere sul sistema Sari Real Time*, 2 marzo 2021; G. Fasano, "Riconoscimento facciale: per il Garante il sistema del Viminale non è conforme alla normativa privacy", in *www.ntplusdiritto.ilsole24ore.com*, 20 aprile 2021; A. Fonsi, "Prevenzione dei reati e riconoscimento facciale: il parere sfavorevole del Garante privacy sul sistema SARI Real Time", in *www.penaledp.it*, 13 maggio 2021.

[28] Nel dettaglio, all'Articolo 5 – *Pratiche di IA vietate*, si afferma: "Sono vietate le pratiche di IA seguenti: […] c) l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA per la valutazione o la classificazione delle persone fisiche o di gruppi di persone per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità note, inferite o previste […]. d) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazioni del rischio relative a persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità; tale divieto non si applica ai sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa; e) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di IA che creano

tipo è invece ammesso nel caso di investigazioni post-crimine e per ragioni di sicurezza nazionale[29]. Situazione, questa, che può dar vita a delle "zone grigie" nell'impiego di queste tecnologie.

3. Dai vantaggi ai rischi: ricordando alcune vicende giudiziarie

È intuitivo che, in questi anni, siano stati vari i casi giudiziari e le pronunce concernenti connesse all'impiego dei sistemi predittivi, talvolta, forieri di valutazioni erronee perché inficiate da *bias* algoritmici, talaltra, persino discriminatorie.

Fra le vicende più celebri, spicca sicuramente quella di Eric Loomis, un cittadino statunitense condannato a sei anni di carcere per aver partecipato ad una sparatoria, nei confronti del quale il giudice di prime cure, oltre a valutare la gravità del reato e a tener conto della "storia criminale" dell'imputato, si è anche avvalso dei risultati prospettati da *Compas*[30] in ordine al possibile rischio di recidiva dell'imputato, il quale, dal *software*, era stato catalogato come un soggetto "ad alto rischio". Eccependo una violazione del suo diritto a un giusto processo – a detta del

o ampliano le banche dati di riconoscimento facciale mediante scraping non mirato di immagini facciali da internet o da filmati di telecamere a circuito chiuso; f) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di IA per inferire le emozioni di una persona fisica nell'ambito del luogo di lavoro e degli istituti di istruzione, tranne laddove l'uso del sistema di IA sia l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di categorizzazione biometrica che classificano individualmente le persone fisiche sulla base dei loro dati biometrici per trarre deduzioni o inferenze in merito a razza, opinioni politiche, appartenenza sindacale, convinzioni religiose o filosofiche, vita sessuale o orientamento sessuale; tale divieto non riguarda l'etichettatura o il filtraggio di set di dati biometrici acquisiti legalmente, come le immagini, sulla base di dati biometrici o della categorizzazione di dati biometrici nel settore delle attività di contrasto; destinato a essere messo in funzione o immesso sul mercato per motivi medici o di sicurezza".

[29] Difatti, sempre nell'*AI Act*, all'Articolo 2 – *Ambito di applicazione*, si legge: "Il presente regolamento non si applica ai sistemi di IA se e nella misura in cui sono immessi sul mercato, messi in servizio o utilizzati con o senza modifiche esclusivamente per scopi militari, di difesa o di sicurezza nazionale, indipendentemente dal tipo di entità che svolge tali attività".

[30] *Infra*, § 1.

reo, inficiato dall'impiego di un *software* privato le cui risultanze erano frutto dell'uso di dati aggregati riferibili a individui in condizioni simili e presentavano discriminazioni basate sul genere – Loomis ricorreva in appello. Confermata la sentenza, la vertenza passava, quindi, alla Corte Suprema del Wisconsin, la quale, il 13 luglio 2016, si pronunciava ritenendo che l'utilizzo di *Compas* nell'ambito di un giudizio penale non violi il diritto al giusto processo dell'imputato. Detto ciò, la Corte, comunque, non mancava di ribadire la necessità che i giudici: *a)* non usino le risultanze algoritmiche come unico fattore sul quale fondare la decisione; *b)* forniscano una motivazione ulteriore rispetto all'impiego del *software*; *c)* non determinino la severità della pena solo sulla base del *risk assessment*[31].

[31] Nel dettaglio, può leggersi: "[…] a sentencing court may consider a Compas risk assessment at sentencing subject to the following limitations. As recognized by the Department of Corrections, the PSI instructs that risk scores may not be used: (1) to determine whether an offender is incarcerated; or (2) to determine the severity of the sentence. Additionally, risk scores may not be used as the determinative factor in deciding whether an offender can be supervised safely and effectively in the community. Importantly, a circuit court must explain the factors in addition to a Compas risk assessment that independently support the sentence imposed. A Compas risk assessment is only one of many factors that may be considered and weighed at sentencing. Any Presentence Investigation Report (PSI) containing a Compas risk assessment filed with the court must contain a written advisement listing the limitations. Additionally, this written advisement should inform sentencing courts of the following cautions as discussed throughout this opinion: 1) The proprietary nature of Compas has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined. 2) Because Compas risk assessment scores are based on group data, they can identify groups of high-risk offenders not a particular high-risk individual. 3) Some studies of Compas risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism. 4) A Compas risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and renormed for accuracy due to changing populations and subpopulations. 5) Compas was not developed for use at sentencing but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole. It is important to note that these are the cautions that have been identified in the present moment. For example, if a cross-validation study for a Wisconsin population is conducted, then flexibil-

Sempre con riferimento all'uso di *Compas*, possono brevemente ricordarsi: la vicenda *Henderson v. Stensberg* (verificatasi anch'essa in Wisconsin, e nella quale il ricorrente lamentava un'alterazione della decisione circa libertà vigilata dovuta ad un presunto pregiudizio razziale del *software* nei confronti dei detenuti di colore); nonché il caso *People v. Younglove* (che ha interessato la Corte di Appello dello Stato del Michigan e che concerneva l'utilizzo di sistemi predittivi in sede di condanna)[32].

Alquanto significativa anche la vicenda (poc'anzi accennata), che nel luglio 2020 ha visto protagonista Billy Curry Jr. (*United States v. Curry*). Nel caso di specie – il fermo di Curry occasionato dall'esplosione di alcuni colpi d'arma da fuoco nei pressi del luogo in cui egli si trovava al momento della perquisizione – i giudici sono stati chiamati a stabilire se il fermo e la perquisizione dell'imputato fossero giustificati. Diversamente dal caso Loomis (in cui il *software Compas* era stato impiegato in sede di condanna), qui, i giudici dovevano valutare le implicazioni connesse all'uso – da parte delle forze dell'ordine della città di Richmond – di un modello di *predictive policing* (*place-based*), finalizzato tanto alla repressione delle condotte criminose quanto alla loro prevenzione. Nell'ottica di una valutazione critica è interessante ricordare – fra le *concurring opinions* – quella del giudice Thacker, che, sottolineando come la bontà degli algoritmi dipenda pur sempre dalla qualità dei dati inseriti (secondo il principio del *garbage in, garbage out*, c.d. GIGO), ha ritenuto quasi inevitabile che anni e anni di condotte discriminatorie poste in essere dalle forze dell'ordine abbiano prodotto una serie

ity is needed to remove this caution or explain the results of the cross-validation study. Similarly, this advisement should be regularly updated as other cautions become more or less relevant as additional data becomes available."

[32] In argomento, cfr.: D. Kehl, P. Guo, S. Kessler, "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing", in *Responsive Communities Initiative*, Harvard Law School, 2017; S.B. Starr, "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination", in *Stanford Law Review*, 66 (2014), n. 4, pp. 803-872; A.L. Washington, "How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate", in *Colorado Technology Law Journal*, 17 (2018), n. 1, pp. 131-160.

di dati circa le attività criminali contraddistinti da pregiudizi nei confronti delle minoranze e delle persone di colore.

Decisamente interessante, pure la causa *Ewert v. Canada*, con riguardo alla quale la Suprema Corte canadese si è espressa in ordine all'impiego – da parte del *Correctional Service of Canada* (CSC) – di un *software* per la valutazione del rischio di recidiva e di psicopatia. In particolare, la circostanza aveva messo in luce l'esistenza di possibili *bias* cognitivi del sistema, idonei a produrre risultati discriminatori soprattutto nei confronti dei soggetti appartenenti ai gruppi indigeni (come, per l'appunto, era Ewert). È importante sottolineare che, nel caso di specie, la Suprema Corte si è espressa a sfavore dell'uso del *software*, in ragione dell'impossibilità di scongiurare la presenza (e l'incidenza) di fattori culturalmente condizionanti e, dunque, potenzialmente discriminatori[33].

Degna di menzione (ancorché fattualmente diversa) anche la vertenza *United States v. Jones*, in cui l'imputato Antoine Jones (proprietario e gestore di un locale notturno nel Distretto di Columbia della città Washington), essendo sospettato di traffico di stupefacenti, fu sottoposto ad un'indagine, nell'ambito della quale gli agenti di polizia (a fini di monitoraggio e di prevenzione) utilizzarono varie tecniche investigative e *software* di AI. Fra questi, pure il posizionamento di un dispositivo elettronico di tracciamento sull'auto usata da Jones, ma intestata a sua moglie. Scelta che, tuttavia, a detta della Corte Suprema degli Stati Uniti aveva integrato una violazione del IV Emendamento[34]. Una sentenza, questa, che, a suo modo, ha fatto da apripista a diverse pronunce successive[35].

[33] Nel dettaglio, la Corte avverte circa il pericolo che – laddove gli strumenti contestati sopravvalutino effettivamente il rischio – i detenuti indigeni: sperimentino condizioni detentive eccessivamente dure (*Ewert v. Canada*, 2018, SCC 30, [2018] 2 S.C.R. 165)

[34] "The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

[35] Cfr., fra gli altri, K.S. Bankston, A. Soltani, "Tiny Constables and the Cost of Surveillance: Making. Cents Out of United States v. Jones", in *The Yale Law Journal Online*, 123 (2014), pp. 335-357.

Particolarmente significativa, poi, la vertenza *R (Bridges) v. Chief Constable of South Wales Police*. In questo caso, Edward Bridges citava in giudizio il Capo della Polizia del Galles, criticando l'uso del sistema di *Automatic Facial Recognition Locate (Neo Face Watch)*. In modo particolare, le forze di polizia avevano acquisito i dati biometrici facciali di Bridges (il 21 dicembre 2017 e il 27 marzo 2018), senza fornire alcun avviso dell'uso del *software* e violando, così, i suoi diritti considerato (fra cui, quello alla *privacy*). Non avendo ottenuto soddisfazione in prime cure, Bridges si rivolgeva alla Corte d'Appello della Divisione Civile, che – dopo un'articolata disamina, sia del quadro normativo, sia delle modalità di funzionamento del sistema tecnologico utilizzato – riconosceva l'avvenuta violazione dell'articolo 8 della CEDU.

Il richiamo a questi casi (che sono in assoluto tra i più noti e controversi) riporta l'attenzione – oltre che sulle innegabili potenzialità degli strumenti predittivi applicati alla sicurezza pubblica e alla prevenzione dei reati – sia sulla possibile fallibilità e sia sulla non assoluta obiettività di questi modelli, che, nei fatti, possono riproporre errori e pregiudizi. Di qui, l'importanza di un appello ad un uso trasparente e particolarmente attento di questi sistemi, del quale, l'Europa si è fatta, a suo modo, autorevole portavoce[36].

[36] In materia – oltre al già citato *AI Act* – è particolarmente utile ricordare anche la *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi* – adottata già nel dicembre del 2018 – che stabilisce: 1) il principio del rispetto dei diritti fondamentali (assicurare l'elaborazione e l'attuazione di strumenti e servizi di intelligenza artificiale siano compatibili con i diritti fondamentali); 2) il principio di non-discriminazione (prevenire specificamente lo sviluppo o l'intensificazione di discriminazioni tra persone o gruppi di persone); 3) il principio di qualità e sicurezza (in ordine al trattamento di decisioni e dati giudiziari, utilizzare fonti certificate e dati intangibili con modelli elaborati multidisciplinarmente, in un ambiente tecnologico sicuro); 4) il principio di trasparenza, imparzialità ed equità (rendere le metodologie di trattamento dei dati accessibili e comprensibili, autorizzare verifiche esterne); 5) il principio del controllo da parte dell'utilizzatore (precludere un approccio prescrittivo e assicurare che gli utilizzatori siano attori informati e abbiano il controllo delle loro scelte).

4. Sorvegliare e prevenire al tempo dell'IA

Le considerazioni sin qui svolte sottolineano un vero e proprio cambio di paradigma del sistema penale, ovvero il passaggio – tramite le tecnologie di IA – dal più tradizionale binomio *sorvegliare e punire* all'odierno *sorvegliare e prevenire*, con delle evidenti ricadute in generale nel processo penale e, in particolare, nella decisione del giudice, che avvalendosi degli algoritmi predittivi delega a suo modo, anche ai fini della determinazione della pena, la sua stessa valutazione alla forza irresistibile di un sistema di IA. Può così accadere che sia il sistema (automatico, interconnesso, complesso, esperto) a decidere e che il giudice molto semplicemente vi si adegui. Del resto, i diversi sistemi di IA sottopongono all'attenzione dell'interprete i dati (fatti, circostanze, pareri, norme, e così via) e ciò costituisce la premessa della valutazione stessa, dirigendo in questa maniera la decisione che è tutt'una con quel materiale (i dati) che il sistema ha raccolto, analizzato, rielaborato, reimmesso, nel suo percorso (imperscrutabile) dall'input all'*output*.

Le nuove tecnologie e i sistemi di IA incidono, e non potrebbe essere altrimenti, su concetti quali sorveglianza, giudizio e pena. I *software* di *predictive policing*, per la modalità di raccolta (e di incrocio) dei dati e per la loro stessa funzione, mutano i caratteri della sorveglianza (da contenuta a diffusa), trasferiscono il giudizio (dall'uomo-giudice al 'logico puro applicatore'), correggono lo scopo della pena (dal compensare al prevenire-intimidire).

È innanzitutto la sorveglianza che muta. I nuovi prodotti di IA fanno sì che il controllo diventi continuo, automatico, involontario, caratteri questi normalmente riferibili a sistemi astratti e impersonali, a sistemi con capacità di data-veglianza (individuale e/o di massa) più forte e diretta rispetto a qualsiasi altra forma di sorveglianza, poiché si tratta del "systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons"[37], ovvero di un regolare e strutturato utilizzo di una mole di dati che

[37] L'espressione e la definizione risalgono a R. Clarke, "Information technology and dataveillance", in *Communications of the ACM*, 31, 5, 1988 (http://www.rogerclarke.com/DV/CACM88.html).

vengono di continuo raccolti, momentaneamente accantonati e all'occorrenza ripresi.

Diversamente dalle metafore del Panopticon e del Grande Fratello – che hanno come referente un potere coercitivo centrale –, l'odierna sorveglianza prende forma nei c.d. non-luoghi informatici, all'interno dei quali le informazioni, di continuo introdotte, diventano la contingente misura di tutte le cose. Si tratta di non-luoghi in cui i dati automaticamente assemblati vincolano tutti e tutto, dove i risultati mutano involontariamente le entità, i significati e le vite su scala globale[38]. Entro queste dinamiche, si sviluppa la tecno-regolazione. Con l'espressione si intende sia il controllo e la registrazione (comprese la sistemazione, la riorganizzazione, la coordinazione) del comportamento umano, e sia la programmata capacità di influenzare il comportamento umano attraverso la messa a punto di valori, di norme e di regole, nei più diversi dispositivi tecnologici e con questi compatibili. Le nuove tecnologie, soprattutto nel loro combinarsi di tecnologia informazionale-comunicazionale (ICT), di bio-, nano-, neuro-tecnologia, e di robotica, influenzano e qualche volta determinano il comportamento degli individui, così che diventa necessario interrogarsi sugli aspetti e sui limiti etico-giuridici della tecno-regolazione, come pure sulle implicazioni dal punto di vista della legittimità e della tenuta dei valori democratici di una tecno-regolazione applicata su larga scala[39].

[38] Come già osservato: i non-luoghi informatici non offrono riparo da chi ci osserva, né un bastione attorno al quale organizzare una possibile linea di difesa. Al Panopticon si sostituisce il Synopticon: "l'atto di guardare svincola chi guarda dalla propria localizzazione, e lo trasporta almeno spiritualmente nel ciberspazio, dove la distanza non conta più, anche se, fisicamente, non ci si è mossi. Non conta più se gli oggetti del Synopticon, trasformati ora da guardati in guardanti, si muovono o stanno fermi. Dovunque siano e dovunque vadano, essi possono collegarsi – e lo fanno – alla rete extraterritoriale che permette ai molti di guardare i pochi. Il Panopticon costringeva la gente a una posizione in cui poteva essere guardata. Il Synopticon non ha bisogno di costringere nessuno, seduce la gente perché guardi" (così Z. Bauman, *Dentro la globalizzazione. Le conseguenze sulle persone*, trad. it., Laterza, Roma-Bari, 2007, pp. 59-60).

[39] Per ulteriori considerazioni rinvio a A.C. Amato Mangiameli, "Tecno-regolazione e diritto. Brevi note su limiti e differenze", in *Il diritto dell'informazione e dell'informatica*, 33 (2017), n. 2, pp. 147 ss.; Ead., "Tecno-di-

E se mutano i caratteri della sorveglianza, da contenuta a diffusa, da possibile a reale e assoluta[40], e ancora: dal sorvegliare e punire al sorvegliare e prevenire, allo stesso modo il giudizio, ovvero la valutazione e la decisione giuridica, si adegua ai nuovi strumenti e con questi al nuovo sentire. Resta naturalmente la complessità del giudizio, come sa bene ogni interprete[41]: l'ascolto delle parti e dei testimoni, l'ispezione di luoghi e di cose, la verifica di tracce e indizi, un puzzle tutto da ricostruire, in cui ora i dati del *predictive policing* costituiscono i nuovi e importanti pezzi da utilizzare e da far combaciare. In altri termini, l'ascolto, l'ispezione, la verifica, si accompagnano e si formano attraverso il dato che ormai costituisce la cornice entro la quale valutare ogni elemento in base alla corrispondenza col sistema di IA che, di per sé, ha una sorta di valore e statuto aggiunto: è automatico, oggettivo, intelligente, scientifico. E quindi, proprio per tali caratteri, ha un qualcosa di irresistibile, così che senza il dato (i dati) automatico, oggettivo, intelligente, scientifico, non c'è alcuna valutazione, e d'altra parte la valutazione è già il dato, in quanto attività puramente logico-deduttiva.

I *software predictive policing*, a loro modo, conquistano un ruolo di primo piano, intervengono nelle diverse dinamiche, costituiscono la base dell'attività istruttoria, svolgono una fun-

ritto e tecno-regolazione. Spunti di riflessione", in *Rivista di filosofia del diritto*, 6 (2017), pp. 87 ss.

[40] I nuovi sistemi di IA realizzano concretamente l'auspicio di Bentham, ovvero che l'individuo sia sempre sotto controllo: "il meglio che si possa auspicare è che in ogni istante, avendo motivo di credersi sorvegliato, e non avendo i mezzi di assicurarsi il contrario, *creda* di esserlo" (*Panopticon, ovvero la casa di ispezione*, trad. it., Marsilio, Venezia, 1983, in part. p. 36).

[41] E a maggior ragione il giudice. Egli non ha assistito al fatto e per applicare la norma al caso di specie interroga, ascolta, accerta. "Vorrebbe chiamare, se potesse, tutto il mondo sotto i suoi occhi", vorrebbe chiamare, se potesse, oltre quel che gli sta davanti (il *presente*), anche quanto appartiene al *passato* e quanto riguarda il *futuro*, quanto è *noto* e quanto è *ignoto*. Dal presente al passato e al futuro, dal noto e all'ignoto: il giudice ha solo il presente (l'imputato, i testimoni, le parti, i periti), ma deve con il suo giudizio "marciare nell'una e nell'altra direzione". Deve rischiarare la *zona delle tenebre*: il delitto è il passato e può essere parzialmente ignoto; egli deve con il suo giudizio, che non si forma senza prova, andare avanti: la pena è il futuro, e il futuro è ignoto a tutti gli uomini (v. F. Carnelutti, *L'arte del diritto*, Cedam, Padova, 1949, pp. 61 ss.).

zione probatoria, e intanto, quali strumenti per la ricerca della verità, garantiscono l'oggettività in virtù di prestazioni tecniche. Di principio inconfutabili, l'interprete si misura con i dati e il compito è quantomai arduo, visto che sono presentati attraverso il filtro di scienza e tecnica, sono quindi presentati come del tutto privi di contaminazioni soggettive, dal momento che è l'IA e non altri che li ha raccolti e organizzati, li ha ordinati e rielaborati. Certo utili per la prevenzione del crimine, i *software predictive policing* accentuano il rischio che l'attività di valutazione venga ridotta a mera attività logico-deduttiva e che la forza del passaggio dal dato alla decisione sia determinata da una sorta di delega in bianco dalla decisione (e dall'intelligenza) umana al giudizio (e all'intelligenza) artificiale. Se così è, proprio per evitare un simile rischio, è necessario muovere da un ovvio assunto: le potenzialità dell'IA sono da limitare a ciò che è oggettivamente descrivibile e anche allora potrà sempre dirsi che i meccanismi di raccolta dati dei *software* (pure dei più avanzati) sono il risultato del lavoro di programmazione (di ingegneri, di matematici, di giuristi, e così via), primo ed essenziale input per accostarsi al mondo reale e alla lettura dei molteplici fenomeni[42]

E se mutano i caratteri della sorveglianza e con questi il cambio di paradigma (dal sorvegliare e punire al sorvegliare e prevenire), se c'è il rischio che la decisione si attesti su accertamenti del fatto confinati entro sistemi di IA e in sintonia con il sorvegliare e prevenire, anche la pena subisce delle fondamentali trasformazioni, qualche volta giustificate dall'evolvere stesso del diritto penale, talaltra confusa con qualcosa di decisamente diverso, sia per le peculiarità e sia per il fine, e cioè con la misura di sicurezza.

[42] D'altra parte, anche per accostarsi al mondo (e all'arte) del diritto – come scriveva Krämer – "occorrerebbe una mente dotata di distinzione amatematica e superrazionale, cui fossero familiari non solo criteri di valutazione difficilmente definibili in esatti termini scientifici (il senso vagamente definito dell'armonia, della misura e della tensione, la conoscenza non esattamente acquisibile dello stile e della sua evoluzione) ma anche la capacità di afferrare i suoni delle vibrazioni dell'anima nell'espressione artistica e la sensibilità per l'essenza che si cela dietro i fenomeni" ("Automat und Mensch", in *Universitas*, 13(1958), n. 5, pp. 511 ss.

Scopo della pena è conservare, in un sistema di leggi che regolano i rapporti tra volontà libere, la simmetria della giustizia. Il suo fine è, infatti, di compensare, anziché vendicare, e solo se giusta realizza le sue tante funzioni, sia quelle primarie, sia quelle accessorie, solo se giusta può emendare o rieducare, può prevenire e/o difendere[43]. Di qui la considerazione secondo la quale perché un imputato sia condannato ad una pena, l'unica motivazione sufficiente è che si siano realizzate le condizioni previste dalla legge, ossia *i*) che è stata commessa una violazione prevista dall'ordinamento, *ii*) che l'imputato è l'autore, come tale colui che ne deve rispondere, *iii*) che egli soddisfa i presupposti previsti dalla legge circa la colpa (colpevolezza e imputabilità). Scrive Ross:

> qualsiasi altra giustificazione, per esempio che una pena farà comunque bene all'imputato, che sarà opportuno per la fiducia della popolazione nella polizia che qualcuno venga condannato, che l'opinione pubblica richiede un capro espiatorio, o altre simili considerazioni utilitarie sono senza peso e valore alcuno, e faranno della condanna un crimine giudiziario sia dal punto di vista giuridico sia morale[44].

Scopo della misura di sicurezza è prevenire i reati, così da assicurare e difendere la società, come pure lo Stato o un determinato gruppo di potere, che a seconda delle circostanze potrebbero persino richiedere l'eliminazione della responsabilità penale nella sua accezione più pregnante di responsabilità per fatto proprio e colpevole. Del resto, perché a un soggetto sia applicata una misura di sicurezza (sia per così dire trattato), le motivazioni sono molteplici e ognuna di queste ha una sua pretesa di esaustività. Ad esempio: può essere trattato un soggetto che abbia commesso una violazione, seppur non imputabile; come pure, un soggetto che non ha ancora commesso alcuna violazione, ma è pensabile che ciò accada per la pericolosità che lo contraddistingue, per sé, per gli altri, per lo Stato, per il gruppo di potere.

[43] Ulteriori considerazioni in A.C. Amato Mangiameli, *Filosofia del diritto penale. Quattro voci per una introduzione*, Giappichelli, Torino, 2014, in part. pp. 59 ss.

[44] *Colpa, responsabilità e pena*, trad. it., Giuffré, Milano, 1972, in part. p. 92.

In ogni caso, la misura di sicurezza è distinta dalla pena, e nello Stato di diritto il (diritto al) trattamento è in primo luogo il diritto a non essere puniti, in quanto non imputabili.

Certo utili per la prevenzione del crimine, i *software predictive policing* accentuano il rischio di una possibile confusione tra pena e misura di sicurezza, non foss'altro perché la predittività privilegia la difesa sociale, quale unico fine del sistema penale, e al contempo predilige concetti quali devianza, pericolosità, anormalità. Se così è, proprio per evitare una simile confusione, è necessario interrogarsi di nuovo sul diritto penale e sullo scopo della pena, non trascurando quella dimensione che è propria (del diritto e della pena), ossia il sistema delle libertà.

Wanda D'Avanzo

Intelligenza artificiale e giustizia predittiva. L'orientamento giuridico europeo

Abstract: **Artificial Intelligence and Predictive Justice. The European Legal Orientation.** In recent decades, artificial intelligence has assumed an increasingly central role within society, profoundly influencing multiple areas of human life and activity. If used ethically and responsibly, artificial intelligence could become a valuable ally for humanity in managing particularly complex social areas, such as healthcare and justice. In this sense, we observe that the current regulatory framework for the digital transition appears adequate to ensure significant general protection in the face of the emergence of new digital rights and duties for EU citizens. But it needs an adequate juridical training to ensure a correct integration between human and machines, which will improve the management of our societies. The ability to monitor the activity of intelligent systems and effectively assess the risks associated with their use requires prevention and the reorganization of traditional organizational structures and management, in both the public and private sectors. This transition represents the true challenge of our time. A challenge for the law, aimed at initiating a reflection on how artificial intelligence, human rights, privacy, and big data can coexist.

Keywords: Artificial Intelligence, AI Act, Predictive Justice, Italian Law on AI.

Sommario: 1. Introduzione – 2. L'evoluzione dell'intelligenza artificiale – 3. IA e giustizia predittiva – 4. L'*AI Act* e la disciplina dei contesti ad alto rischio – 5. Intelligenza artificiale a supporto delle attività umane – 6. Conclusioni.

1. Introduzione

Negli ultimi decenni, l'intelligenza artificiale ha assunto un ruolo sempre più centrale all'interno della società, influenzando profondamente molteplici ambiti della vita e delle attività umane.

La progressiva introduzione delle tecnologie intelligenti nelle aziende e nei settori della pubblica amministrazione evidenzia, infatti, oggi come l'intelligenza artificiale non sia più solo un'innovazione emergente, ma una componente attiva e integrata nella nostra quotidianità. Il suo impiego dimostra effetti positivi e vantaggiosi sia in termini di efficienza operativa che di risultati

concreti, favorendo in tal modo un miglioramento tangibile delle prestazioni individuali e collettive.

Si sta affermando, invero, una visione più consapevole degli strumenti digitali che riconosce all'intelligenza artificiale una preziosa capacità di supporto e collaborazione in grado di affiancare l'uomo e potenziarne le capacità decisionali, organizzative e gestionali.

Se utilizzata in modo etico e responsabile, dunque, l'intelligenza artificiale potrà trasformarsi in un valido alleato dell'uomo per la gestione di ambiti sociali particolarmente complessi, come sono ad esempio la sanità e la giustizia.

Nel contesto attuale assume particolare rilevanza il quadro normativo sviluppato dall'Unione Europea per assicurarne un utilizzo sicuro, trasparente e rispettoso dei diritti fondamentali. Il principale documento in materia è il Regolamento europeo n. 1689, noto come AI Act, adottato in via definitiva nel 2024, che ha introdotto una classificazione dei sistemi di intelligenza artificiale basata sul rischio e ha stabilito requisiti rigorosi per quelli di maggiore pericolosità, al fine di assicurare affidabilità, tracciabilità e supervisione umana sul lavoro delle macchine. D'altro canto, però, la sicurezza dei sistemi di intelligenza artificiale e delle infrastrutture digitali è imprescindibile dalla gestione dei dati e della cybersicurezza, per cui l'analisi della rivoluzione digitale che il mondo sta attraversando non può prescindere – ad ogni livello – dalla lettura combinata della normativa di riferimento in materia di trattamento e protezione dei dati ed in materia di sicurezza informatica, di cui al Regolamento europeo n. 679/2016 (GDPR) e alle direttive c.d. NIS.

Attualmente assume rilievo anche la recentissima legge in materia di intelligenza artificiale emanata dall'Italia al fine di riprendere e ribadire anche a livello nazionale molti dei principi contenuti nell'AI Act e nel GDPR in particolare.

Il contesto normativo sulla transizione digitale al momento in vigore sembra essere adeguato per assicurare una garanzia di tutela generale di rilievo di fronte all'affermarsi di nuovi diritti e nuovi doveri digitali per i cittadini dell'unione europea. Attraverso poi una adeguata attività di formazione di tutti i soggetti chiamati ad utilizzare tecnologie innovative come l'IA sarà pos-

sibile assicurare una integrazione uomo macchina, funzionale ad un miglioramento della gestione delle nostre società.

La possibilità di controllare l'attività dei sistemi intelligenti e di effettuare una valida valutazione dei rischi legati al loro utilizzo necessitano di una attività di prevenzione e di riorganizzazione degli assetti organizzativi tradizionali e del management, nel settore pubblico e nel settore privato, svolta principalmente attraverso la formazione, intesa a favorire uno sviluppo consapevole di un nuovo rapporto con la tecnologia, in cui sia sempre l'uomo a decidere sui risultati[1].

Questo passaggio rappresenta la vera sfida della contemporaneità. Una sfida per il diritto, finalizzata ad avviare – sia pure con diversi anni di ritardo – una riflessione su come intelligenza artificiale, diritti umani, riservatezza e *big data* possano coesistere.

2. L'evoluzione dell'intelligenza artificiale

Trattare del mondo tecnologico da un punto di vista giuridico non può prescindere da una, sia pur sintetica, descrizione dello strumento tecnico specifico che si intende analizzare, che nel caso di specie è l'intelligenza artificiale. Con il termine intelligenza artificiale (IA) si fa riferimento alla capacità di sistema informatico di svolgere, attraverso l'uso di algoritmi complessi, funzioni che richiederebbero normalmente l'intervento dell'uomo.

Queste funzioni possono includere la capacità di apprendere, ragionare, risolvere problemi, pianificare azioni e interagire con l'ambiente circostante, anche in modo adattivo e autonomo. Tali processi vengono oggi sempre più spesso affidati a tecnologie

[1] In relazione al tema della formazione giuridica nella transizione digitale, della attuazione della c.d. *governance* e della valutazione dei rischi legati alla tecnologia nella società dell'informazione, si rinvia a W. D'Avanzo, *Il diritto di fronte alle sfide del futuro. Studi di informatica giuridica e diritto dell'informatica*, Universitas Studiorum, Mantova, 2020; Id., *La pubblica amministrazione digitale. Principi di informatica giuridica*, Universitas Studiorum, Mantova, 2022; Id., *E-Government e società del controllo. Il lato oscuro del mondo digitale*, Rubbettino, Soveria Mannelli, 2023; e al più recente Id., *Il metaverso. Le nuove frontiere della tecnologia tra etica e diritto*, Rubbettino, Soveria Mannelli, 2025.

avanzate, capaci di simulare e in alcuni casi superare la performance umana in compiti specifici.

Com'è noto, l'idea di una "macchina pensante" affonda le sue radici nel lavoro del matematico inglese Alan Turing, considerato uno dei padri fondatori dell'informatica moderna. Negli anni '30 del Novecento Turing introdusse il concetto di calcolabilità, aprendo la strada alla possibilità di costruire dispositivi capaci di eseguire qualsiasi funzione computabile. Nel 1950 pubblicò il saggio *Computing Machinery and Intelligence*, in cui propose il famoso *Test di Turing* secondo il quale una macchina poteva dirsi intelligente se un essere umano non era in grado di distinguerla da un altro essere umano durante una conversazione[2].

Le prime ricerche iniziali nel campo dell'intelligenza artificiale sono state, poi, svolte durante i primi anni '60 e '70 del secolo scorso, e hanno esplorato varie tipologie di rappresentazioni di problemi, di tecniche di ricerca di soluzioni e di euristica generale, impiegandole in programmi in grado di risolvere operazioni semplici o acquisire informazioni. I primi agenti intelligenti hanno fatto la loro comparsa nel 1995 e hanno favorito lo sviluppo dei programmi, come quelli che usiamo attualmente, in grado di effettuare diagnosi, progettazioni e analisi in maniera del tutto autonoma[3].

Il passo in più che è stato fatto dagli anni '90 ad oggi nelle ricerche sull'intelligenza artificiale è stato quello di seguire un approccio pragmatico, basato sulla presenza dei dati. La nuova

[2] N.J. Nilsson, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press, Cambridge, 2010, ed anche Id., *Intelligenza artificiale*, Apogeo, Milano, 2002, p. 21. A.C. Mangiameli, *Informatica giuridica*, Giappichelli, Torino, 2015.

[3] S.J. Russell, P. Norvig, *Intelligenza artificiale. Un approccio moderno*, vol. 2, Pearson, Milano, 2005. A tal proposito, è famoso il caso del *Deep Blue*, il computer della IBM che nel 1997 sconfisse il campione mondiale di scacchi Kasparov. L'algoritmo utilizzato dal *Deep Blue* aveva la capacità di analizzare fino a 200 milioni di posizioni dei pezzi al secondo e tale capacità gli consentiva di valutare diverse sequenze di gioco per vederne i risultati in anticipo; così, L. Mérő, *I limiti della razionalità. Intuizione, logica e translogica*, Edizioni Dedalo, Bari, 2005, p. 247.

generazione di macchine intelligenti è alimentata da pochi algoritmi di apprendimento statistico e vaste quantità di dati[4].

Nel corso della sua evoluzione, l'intelligenza artificiale si è nel tempo consolidata come disciplina autonoma dell'informatica, con l'obiettivo di sviluppare macchine in grado di simulare l'intelligenza umana affinché possano svolgere compiti complessi con un grado crescente di autonomia.

In contesti sensibili come quello sanitario o quello giudiziario questo significa poter automatizzare i processi decisionali, ottimizzare la gestione delle risorse, analizzare rapidamente grandi quantità di dati o persino prevedere l'evoluzione di una malattia o l'esito di una controversia, migliorando così la qualità delle procedure e riducendo le probabilità di errore.

Dal punto di vista funzionale e dell'ampiezza delle funzioni che è in grado di svolgere, l'intelligenza artificiale viene comunemente classificata come stretta o generale.

L'intelligenza artificiale stretta, nota anche come IA debole, è la tipologia ad oggi più diffusa: è progettata per eseguire compiti semplici e specifici. Essa non imita né replica l'intelligenza umana, ma si limita a simulare dei comportamenti in base a una gamma di parametri prestabiliti. I sistemi di IA debole possono solo apprendere o imparare a completare compiti specifici, non possiedono consapevolezza né capacità di adattamento a contesti diversi da quelli per cui sono progettati.

L'intelligenza artificiale generale, invece, denominata IA forte, intende rappresentare il concetto di una macchina che imita in tutto e per tutto l'intelligenza o i comportamenti umani con la capacità conseguente di apprendere e applicare il ragionamento per risolvere qualsiasi problema. Dovrebbe, cioè, rappresentare un computer capace di capire e agire in un modo indistinguibile da quello di un essere umano[5].

Tale ultimo caso rappresenta, però, un obiettivo ancora oggi lontano. Le attuali tecnologie di intelligenza artificiale si basano prevalentemente al momento su due approcci fondamentali:

[4] N. Cristianini, J. Cumming-Graham, P. Norvig, "A nostra immagine", in D. Heaven (a cura di), *Macchine che pensano. La nuova era dell'intelligenza artificiale*, Edizioni Dedalo, Bari, 2018, p. 34.

[5] L. Floridi, *Etica dell'intelligenza artificiale*, Raffaello Cortina, Milano, 2022; A. Vitale, *Artificial Intelligence*, Egea, Milano, 2018.

il *machine learning*, o apprendimento automatico, che prevede la creazione di modelli tramite l'uso di algoritmi capaci di fare previsioni o prendere decisioni basate sui dati; e il *deep learning* che utilizza reti neurali profonde per affrontare problemi complessi come il riconoscimento di immagini, testi o segnali biometrici[6].

Tra gli algoritmi di *machine learning* uno dei più utilizzati è la rete neurale artificiale. Le reti neurali sono ispirate alla struttura del cervello e sono composte da strati di nodi interconnessi che elaborano e analizzano informazioni complesse. Tali reti sono fondamentali per l'identificazione di schemi e correlazioni non immediatamente visibili soprattutto in presenza di grandi volumi di dati. Dal punto di vista matematico, una rete neurale può essere vista come una funzione composta, formata da più funzioni dipendenti l'una dall'altra.

Ogni decisione prodotta dal sistema è quindi il risultato di calcoli precisi, eseguiti sulla base di parametri definiti e alla qualità delle informazioni disponibili. Queste reti sono inoltre adattative, capaci cioè di modificare la propria struttura interna per rispondere agli stimoli ricevuti durante l'apprendimento. In tal senso, uno dei progressi più rilevanti nelle sperimentazioni di intelligenza artificiale è stato quello di creare delle macchine in grado di apprendere dai propri errori e operare in contesti non completamente prevedibili dai programmatori. E questa evoluzione ha reso l'IA particolarmente adatta ad affrontare situazioni dinamiche in cui l'esperienza accumulata può guidare decisioni più efficaci[7].

Nel settore sanitario, ad esempio, il *machine learning* viene già utilizzato per l'analisi automatizzata di immagini diagnostiche (radiografie, TAC, risonanze); per il monitoraggio in tempo reale di parametri vitali tramite dispositivi indossabili; per la

[6] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT University Press, Boston, 2016.

[7] Le reti neurali sono algoritmi matematici sviluppati per riprodurre il funzionamento dei neuroni biologici allo scopo di conferire alla macchina la capacità di compiere funzioni e ragionamenti come una mente umana, così N. Boldrini, AI Artificial Intelligence. *Com'è nata, come funziona e come l'intelligenza artificiale sta per cambiare il mondo, la vostra vita, il vostro lavoro*, Milano Finanza, Milano, 2018.

valutazione del rischio nei pazienti con patologie croniche; per l'analisi predittiva dei dati clinici.

Analogamente, in ambito giuridico ci si avvale delle funzionalità degli algoritmi di apprendimento automatico in diversi settori che vanno dalla polizia predittiva all'anticipazione dell'esito del giudizio.

Oltre alla componente algoritmica, un ruolo essenziale dell'intelligenza artificiale è la rappresentazione della conoscenza, che attraverso modelli come la teoria dei linguaggi formali e la teoria delle decisioni, permette ai computer di strutturare le informazioni a disposizione e guidare le decisioni del sistema[8].

Dunque, allo stato attuale, le macchine intelligenti si avvalgono di alcune capacità fondate su quattro livelli funzionali. I primi due livelli sono la comprensione e il ragionamento. Le intelligenze artificiali sono in grado di elaborare enormi quantità di dati in tempi molto brevi e riescono, mediante tecniche statistiche, a collegare tra loro le diverse informazioni raccolte, simulando le capacità cognitive umane. Vi sono, poi, l'apprendimento e l'interazione. Alcune tipologie di macchine intelligenti sono capaci di apprendere attraverso l'analisi delle informazioni e possono interagire con l'uomo.

3. IA e giustizia predittiva

L'evoluzione dell'intelligenza artificiale è strettamente legata all'evoluzione del concetto di 'dato' e al modo in cui attraverso l'informatica è diventato possibile raccoglierne in quantità al fine di renderli sempre più disponibili. Combinando le informazioni tra loro, le macchine riescono ad effettuare delle valutazioni accurate.

L'espressione *Big Data* fa riferimento ad insiemi di dati tendenzialmente illimitati che grazie alle tecnologie avanzate della contemporaneità come l'intelligenza artificiale vengono gestiti ed elaborati in modo da produrre dei risultati utili all'uomo. Questi dati provengono da una moltitudine di fonti diverse:

[8] G. Sartor, *L'intelligenza artificiale e il diritto*, Giappichelli, Torino, 2022, p. 10.

computer, *smartphone*, sensori installati su macchinari e dispositivi, e sono funzionali alla produzione di nuove conoscenze[9].

Menzionati per la prima volta nel 1997 in un documento della NASA, i *big data* possono essere classificati in base alla loro struttura secondo tre categorie principali: possono essere strutturati, non strutturati o semi strutturati, a seconda che siano organizzati in maniera precisa e facilmente accessibile oppure includano contenuti eterogenei o ancora siano organizzati combinando elementi strutturati e non strutturati[10].

La capacità predittiva e generativa delle macchine intelligenti si basa sull'analisi di questi grandi *dataset*: grazie a calcoli statistici e probabilistici, gli algoritmi possono estrapolare da essi previsioni o soluzioni in relazione ad un determinato problema.

Il settore della salute è uno dei contesti in cui i *big data* offrono le maggiori opportunità. Le informazioni a disposizione del servizio sanitario abbinate alle funzionalità dell'intelligenza artificiale possono contribuire a diagnosi più accurate, trattamenti più efficaci e sviluppo di terapie a costi ridotti, aprendo nuovi scenari per una sanità più efficiente, predittiva e personalizzata, confermando il ruolo centrale della trasformazione digitale nella costruzione del futuro delle nostre società[11].

Nel diritto, analogamente, l'analisi predittiva può trovare applicazione, come abbiamo anticipato, in diversi settori che vanno dalla polizia predittiva, cioè dalla previsione di episodi criminali – ad esempio luoghi e orari nei quali saranno, verosimilmente, compiuti reati di una certa tipologia – all'anticipazione dell'esi-

[9] K. Crawford, J.M. Schultz, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms", in *Boston College Law Review*, 55 (2014), n. 1, p. 96.

[10] Commissione europea, *Libro bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia*, COM (2020) 65 final; ed anche E. Iandiorio, *Big data. Cosa sono, come analizzarli e utilizzarli per fare marketing*, Dario Flaccovio Editore, Roma, 2019.

[11] Aa.Vv., "Converging Blockchain and Next-Generation Artificial Intelligence Technologies to Decentralize and Accelerate Biomedical Research and Healthcare", in *Oncotarget*, 9 (2018), n. 5, pp. 5665-5690; A. Monti, "Intelligenza artificiale e responsabilità medica: profili giuridici ed etici", in A. Saitta (a cura di), *Sanità digitale e nuove tecnologie*, Maggioli, Santarcangelo di Romagna, 2021.

to del giudizio, riproducendo modelli e schemi comportamentali che si suppongono alla base della decisione giudiziaria[12].

Negli Stati Uniti, ad esempio, le forze dell'ordine si avvalgono di strumenti di polizia predittiva basati sui *big data* al fine di riaprire casi irrisolti o prevenire crimini futuri. Il modello di IA di *predicting policing* analizza la data, l'ora, il tipo di crimine e il luogo in cui è stato commesso e combina questi dati con i dati storici sui crimini per incrociare gli elementi comuni che poi diventano il fulcro delle indagini. Analisi simili vengono utilizzate nei c.d. *fusion center*, dei centri informativi del Dipartimento per la Sicurezza Interna e del Dipartimento di Giustizia per condividere dati personali detenuti da agenzie come CIA o FBI per condurre prevenzione antiterrorismo o attività di sorveglianza su individui sospetti o meritevoli di indagini, perquisizioni o detenzione in base ai criteri predeterminati[13].

In Cina, d'altro canto, l'utilizzo dei dati in combinazione con gli algoritmi di intelligenza artificiale nel settore giudiziario fa parte di un progetto più ampio finalizzato ad introdurre le c.d. *smart courts* nel sistema della giustizia. Il processo di implementazione delle *smart courts* e dei sistemi di *e-justice* o *e-court* ruota intorno a due elementi: da un lato, lo snellimento della burocrazia come prerequisito fondamentale per semplificare l'accesso alla giustizia e la mancanza di trasparenza nel funzionamento del sistema giudiziario; e dall'altro, l'ottimizzazione delle procedure giudiziarie. Lo sviluppo delle *smart courts* prevede che l'elaborazione delle informazioni giuridiche venga affidata a programmi basati sull'intelligenza artificiale con l'obiettivo di automatizzare le funzioni di ricerca per assistere i giudici e avvocati nella gestione dei casi e nel prendere decisioni[14].

Ulteriori applicazioni prevedono l'uso dell'intelligenza artificiale nelle indagini preliminari, come mezzo di ricerca della

[12] A. Maceratini, "La giustizia predittiva: potenzialità e incognite", in *Medialaws*, (2024), n. 2, pp. 231-241.

[13] K. Crawford, J.M. Schultz, *op. cit.*, p. 103 s.

[14] Q. Zhanga, D.S. Dzhumalievb, J. Qic, "Current State, Development and Peculiarities of Chinese Smart Courts", in *Jus-Online*, (2024), n. 1, pp. 119-143. J. Zhao, B.G. Farinas, "Artificial Intelligence and Sustainable Decisions", in *European Business Organization Law Review*, 24 (2023), n. 1, pp. 1-39.

prova, ad esempio attraverso i sistemi di localizzazione GPS o mediante le immagini tratte dalle telecamere cittadine per effettuare il riconoscimento facciale a scopo probatorio nel procedimento penale[15].

Ancora, l'espressione 'giustizia predittiva' può far riferimento alla possibilità di affidare alla macchina la previsione dell'esito, o dei possibili esiti, delle decisioni giudiziarie tramite apposite applicazioni algoritmiche[16].

4. L'*AI Act* e la disciplina dei contesti ad alto rischio

Tuttavia, l'adozione di queste tecnologie non è priva di rischi. La crescente dipendenza da sistemi algoritmici pone interrogativi cruciali sul piano etico, giuridico e sociale, richiedendo un attento bilanciamento tra innovazione e tutela dei diritti fondamentali.

Tra le principali criticità emergono le questioni legate alla trasparenza delle decisioni automatizzate, alla responsabilità legale in caso di errore, alla protezione dei dati personali e alla sicurezza informatica delle infrastrutture digitalizzate.

Com'è noto, i modelli di apprendimento automatico, se addestrati con dati di scarsa qualità o incompleti, possono generare *bias* algoritmici e decisioni discriminatorie. In assenza di controllo gli algoritmi predittivi potrebbero essere configurati in maniera tale da riflettere radicati pregiudizi sociali, politici, giuridici ed economici, in spregio ai fondamentali diritti umani. Un esempio, in tal senso, è l'algoritmo americano COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) che ha il compito di misurare la percentuale di rischio di recidiva sulla base di dati statistici[17]. Decisioni basate su ana-

[15] A. Traversi, "Giustizia predittiva: quale futuro?", in *Altalex*, 2023. Recuperato da https://www.altalex.com/documents/news/2023/03/15/giustizia-predittiva-quale-futuro [Data di consultazione: 31/08/2025].

[16] M. Farina, "Giustizia predittiva", in A.C. Amato Mangiameli, G. Saraceni (a cura di), *Cento e una voce di informatica giuridica*, Giappichelli, Torino, 2023, p. 231.

[17] K. Crawford, J. Schlutz, *op. cit*., p. 106.

lisi errate condurrebbero inevitabilmente a gravi violazioni del principio di legalità[18].

Ma non solo. Il crescente utilizzo di infrastrutture digitali espone ogni settore al rischio di cyber attacchi o *data breach* con impatti potenzialmente devastanti su strutture e settori.

Il settore sanitario e il settore giudiziario, in specie, che trattano dati altamente sensibili rappresentano settori ad alto rischio sui quali la Commissione europea ha promosso un quadro normativo volto a coniugare innovazione e tutela dei diritti, imponendo requisiti stringenti in termini di: supervisione umana; trasparenza e tracciabilità; sicurezza informatica; accuratezza e robustezza tecnica; gestione dei dati utilizzati nei processi di training e validazione.

In questo contesto, si inseriscono i principali interventi normativi di riferimento a livello europeo che sono il Regolamento UE 1689/2024 sull'intelligenza artificiale (AI Act) e il Regolamento generale 679/2016 sulla protezione dei dati del 2016 (GDPR) e la direttiva NIS 2 sulla *cybersecurity*[19].

Con l'AI Act, il legislatore europeo ha cercato di promuovere uno sviluppo etico e sostenibile dell'intelligenza artificiale, il cui scopo non è soltanto quello di regolamentarne gli usi, ma soprattutto sanzionare eventuali abusi. Secondo l'AI Act, l'intelligenza artificiale è da intendersi come "un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare *output* quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali".

In tal senso, il Considerando 12 chiarisce che "le tecniche che consentono l'inferenza nella costruzione di un sistema di IA comprendono approcci di apprendimento automatico che imparano dai dati come conseguire determinati obiettivi e approcci basati sulla logica e sulla conoscenza che traggono inferenze

[18] M. Barberis, "Giustizia predittiva: ausiliare e sostitutiva. Un approccio evolutivo", in *Milan Law Review*, 3 (2022), n. 2, p. 12.

[19] E. Dalle Mulle, *L'intelligenza artificiale e il diritto europeo: un'introduzione al quadro normativo emergente*, Giappichelli, Torino, 2023.

dalla conoscenza codificata o dalla rappresentazione simbolica del compito da risolvere"[20].

Il livello di rischio legato a questi sistemi intelligenti impone obblighi differenziati in merito ai principi di proporzionalità, trasparenza ed efficacia. Il concetto di rischio è definito dall'art. 3 dell'AI Act come "la combinazione della probabilità del verificarsi di un danno e la gravità dello stesso" (n. 2). Questa definizione permette di articolare tale concetto in una quadripartizione distribuita in a) rischio inaccettabile, vietato dall'ordinamento; b) alto rischio, soggetto ad obblighi specifici; c) rischio limitato, soggetto soltanto al principio di trasparenza; d) rischio minimo, tollerato dal legislatore. Dal rischio inaccettabile, che prevede l'uso dell'IA in modalità gravemente lesive dei diritti e delle libertà fondamentali quali il rispetto della dignità umana, dell'uguaglianza e della riservatezza, discendono il divieto di utilizzo delle tecniche subliminali, dello sfruttamento delle vulnerabilità di un soggetto o di un gruppo di persone, delle pratiche di *social scoring*. Ciò include anche il divieto dell'uso di dati biometrici in tempo reale e il divieto di utilizzo di tecniche volutamente manipolative e ingannevoli, che potrebbero alterare il comportamento di una persona o di un gruppo di persone[21].

In merito ad alcune pratiche di utilizzo dell'intelligenza artificiale, l'AI Act pone una sostanziale differenza tra l'Europa e altre aree del mondo come la Cina in cui, ad esempio, le pratiche di *social scoring* sono consentite. Tali pratiche sfruttano i *big data* e gli algoritmi di intelligenza artificiale per classificare le persone sulla base del loro comportamento sociale o di caratteristiche personali, che di fatto potrebbero implementare una sorveglianza di massa della società attraverso i dispositivi tecnologici di uso quotidiano. In Europa, invece, le pratiche di classificazione della persona sono considerate lesive del diritto alla dignità, alla discriminazione, all'uguaglianza e alla giustizia.

[20] P.B. Helzel, W. D'Avanzo, M. Leonetti, "Sostenibilità e intelligenza artificiale. Un connubio possibile?", in U. Comite, A. Kostyuk (a cura di), *Sostenibilità e intelligenza artificiale: resilienza o panacea?*, Wolters Kluwer-Cedam, Milano, 2025, p. 30 s.

[21] M. Iaselli, "Ai Act principi, regole ed applicazioni pratiche del Reg. Ue 1689/2024", in E. Mandara (a cura), *Il regolamento UE sull'intelligenza artificiale*, Maggioli, Santarcangelo di Romagna, 2024, p. 56.

Il problema è che, però, i sistemi di intelligenza artificiale sono in grado di creare banche dati di riconoscimento facciale mediante *scraping* non mirato di immagini di persone prese dal web o da filmati di telecamere a circuito chiuso, per cui le tutele e i divieti potrebbero essere facilmente aggirati, ad esempio, mediante l'uso di tecniche di *deepfake*[22].

La sicurezza dei sistemi di intelligenza artificiale è, dunque, difficilmente controllabile e la questione del rischio e del principio di affidabilità dell'algoritmo ruotano fondamentalmente attorno al concetto di dato e alla sua gestione.

L'AI Act va dunque letto congiuntamente con il GDPR che pone vincoli precisi per la raccolta, la conservazione e il trattamento dei dati in contesti nei quali viene utilizzata l'intelligenza artificiale, che deve conformarsi al principio della qualità dei dati e della liceità del trattamento[23].

A livello nazionale, il Garante per la protezione dei dati personali ha più volte ribadito l'importanza di adottare misure tecniche e organizzative adeguate a garantire la sicurezza delle informazioni. La *governance* dell'intelligenza artificiale non può, infatti, prescindere da un ancoraggio ai valori etici fondamentali, che mettano al centro dell'innovazione i diritti umani, affinché i sistemi intelligenti non sostituiscano ma supportino l'attività e la decisione umana[24].

Questa necessità di leggere congiuntamente le norme disponibili in materia di transizione digitale è stata peraltro sottolineata nella recentissima legge contenente "Disposizioni e deleghe al Governo in materia di intelligenza artificiale" approvata in via definitiva dal Senato lo scorso 17 settembre 2025. L'art. 4 del testo normativo specifica chiaramente, al comma 1, che l'utilizzo di sistemi di intelligenza artificiale deve garantire il rispetto dei principi sul trattamento dei dati sanciti nel GDPR.

[22] R.J. Neuwirth, "Prohibited Artificial Intelligence Practices in the Proposed EU Artificial Intelligence Act", in *Computer Law & Security Review*, 48 (2023), pp.1-41.

[23] F. Romeo, *Il dato digitale e la natura delle cose*, in A. Ballarini (a cura di), *Diritto Interessi Ermeneutica*, Giappichelli, Torino, 2015, p. 119 s.

[24] Aa.Vv., *L'intelligenza artificiale al servizio dello sviluppo sostenibile*, CNR edizioni, Roma, 2021, p. 54.

È interessante notare che la legge in questione affronti il tema dell'IA proprio in relazione ai settori della sanità e della giustizia in cui le macchine intelligenti possono si portare grande innovazione, ma la loro introduzione non può in nessun caso ledere i diritti fondamentali delle persone, come l'accesso alle prestazioni sanitarie (art. 7), la tutela dei dati trattati (artt. 8 e 9), la necessità di non violare i diritti inviolabili dei lavoratori nella gestione del rapporto di lavoro (art. 11), e la riserva al magistrato, nel caso di impiego di sistemi di IA nell'attività giudiziaria, di ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti (art. 15).

5. Intelligenza artificiale a supporto delle attività umane

Sebbene, ancora oggi si sia lontani da una vera e propria attività decisionale operata dall'intelligenza artificiale, questi sistemi di calcolo avanzato vengono già utilizzati come elementi tecnici in diversi contesti, come quello sanitario e giudiziario che abbiamo avuto modo di menzionare in precedenza.

In ambito giuridico, le applicazioni intelligenti vengono ormai utilizzate soprattutto in situazioni in cui si osserva il ripetersi di casi simili suscettibili di una semplice schematizzazione o in ipotesi nelle quali è la stessa legge ad imporre una semplificazione procedurale, ritenendo superfluo un dettagliato accertamento dei fatti di causa.

Al momento, in sostanza, i sistemi informatici intelligenti svolgono prevalentemente attività in funzione ausiliaria dell'uomo, come strumentazione di supporto per il lavoro.

In ogni caso, però, a fronte della rapida evoluzione delle tecnologie, è diventato quanto mai necessario iniziare a pensare allo sviluppo di un nuovo rapporto tra uomini e macchine, che sia in grado di integrare le capacità di entrambi per far fronte ai nuovi problemi della società.

La relazione uomo-macchina dovrebbe essere intesa a migliorare le attitudini reciproche pur consentendo all'uomo di poter sempre mantenere il controllo sui risultati prodotti dai sistemi di intelligenza artificiale e, se del caso, modificarli. E, in tal senso, lo stesso AI Act, come il GDPR aveva fatto in preceden-

za, e la linea proposta dal legislatore italiano, non mancano di sottolineare più volte come occorra garantire che le previsioni, le raccomandazioni o le decisioni prese dagli agenti tecnologici intelligenti debbano essere sempre controllate e possano essere efficacemente ribaltate dall'uomo ed anche ignorate[25].

In ogni caso, questa sofisticata simbiosi tra uomini e macchine porterà inevitabilmente nel futuro ad un cambiamento fondamentale della nostra civiltà, una rivoluzione cui, però, il nostro mondo non è ancora adeguatamente preparato.

Gli interrogativi legali, sociali ed etici sono ancora molti e di difficile soluzione. I *big data* hanno un enorme potenziale invasivo e lesivo per le persone e i sistemi di intelligenza artificiale, predittivi o generativi, possono giungere ed analisi e conclusioni di questi dati totalmente imprevedibili. Garantire una corretta gestione e una adeguata qualità dei dati che oggi circolano diffusamente nella rete è difficile, pur in presenza di una normativa ampia e dettagliata. *Big data* e algoritmi rappresentano sistemi dinamici in grado di eludere o marginalizzare le tutele e i quadri normativi esistenti.

6. Conclusioni

La necessità di mantenere degli standard di sicurezza elevati nella gestione delle applicazioni di intelligenza artificiale deve, quindi, porre l'attenzione sul tipo di società digitale che vogliamo creare.

[25] S. Karnouskos, "Symbiosis with Artificial Intelligence via the Prism of Law, Robots, and Society", in *Artificial Intelligence and Law*, 30 (2022), n. 4, pp. 93-115; si veda anche L. Pulito, "Il contributo dell'intelligenza artificiale simbiotica nella protezione delle vittime vulnerabili e nel contrasto della violenza di genere", in *Biolaw Journal*, (2024), n. 1, p. 231; ed anche P. Marra, "Giustizia digitale simbolica e sue implicazioni procedurali", in *L'Ircocervo*, 23 (2024) n. 1, pp. 8-26. A. Carnevale, A. Lombardi, F.A. Lisi, "Exploring Ethical and Conceptual Foundations of Human-Centred Symbiosis with Artificial Intelligence", in G. Boella, F.A. D'Asaro, A. Dyoub, L. Gorrieri, F.A. Lisi, C. Manganini, G. Primiero (eds.), *Beware 2023: Bias, Ethical AI, Explainability and the Role of Logic and Logic Programming*, *Proceedings of the 2nd Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming, International Conference of The Italian Associaton for Artificial Intelligence (AI*IA 2023)*, Roma, 2023.

L'urgenza di proteggere i diritti individuali e garantire il benessere della società, evitando che le pratiche tecnologiche possano giungere a danneggiare, invece che supportare, le persone, dovrebbe portare ad interrogarsi su come le norme che abbiamo oggi a disposizione vengano applicate nella realtà quotidiana e se questa applicazione pratica sia sufficiente a fornirci delle adeguate garanzie. La risposta che può darsi a questo interrogativo è positiva solo parzialmente.

Molti dei concetti giuridici su cui si basano le norme che disciplinano la società digitale, come ad esempio la valutazione di rischi, pur essendo in vigore da diversi anni e pur rappresentando un po' il filo conduttore della disciplina, presentano ancora oggi una applicazione disomogenea, tra gli stati membri e all'interno delle singole nazioni.

Ciò vanifica di fatto l'armonizzazione della disciplina transfrontaliera, ma rende difficile anche combinare validamente le disposizioni tra loro (GDPR, NIS 2, Ai Act), aprendo così la strada a nuove forme di vulnerabilità, specie in settori come quello sanitario e giudiziario in cui gli esiti di una introduzione massiva di una tecnologia come l'IA appaiono ancora sfumati ed estremamente complessi da prevedere.

Il concetto di rischio, così come quello di sicurezza dei sistemi informatici, è caratterizzato da una dimensione globale che i singoli ordinamenti non sono più in grado di gestire in maniera autonoma entro i propri confini definiti. Il progresso tecnologico richiede, quindi, inevitabilmente la necessità di uniformarsi a standard condivisi in grado di favorire forme di tutela globali.

Come teorizzato già da una parte della dottrina internazionale è auspicabile che la gestione sistematica dei *big data* in combinazione con i sistemi di IA si allinei concentrandosi maggiormente sulla ricerca degli errori, sull'identificazione delle loro cause e sull'implementazione di azioni correttive da verificare *a priori*, dunque prima che il sistema informatico possa creare un danno[26].

[26] In tal senso, D. Gray, D. Citron, "The Right to Quantitative Privacy", in *Minnesota Law Review*, 98 (2013), n. 1, pp. 62-67; R.H. Fallon Jr., "Some Confusion about Due Process, Judicial Review and Constitutional Remedies", in *Columbia Law Review*, 93 (1993), n. 2, p. 309, p. 311 e pp. 336-37; J.L. Mashaw, "The Management Side of Due Process: Some Theoretical

In questo contesto, in Italia in particolare, l'attenzione dovrebbe concentrarsi sulla necessità di gestire adeguatamente le attività di prevenzione dei rischi, attraverso una riorganizzazione funzionale alla transizione digitale che finora non c'è stata. L'Italia, cioè, dovrebbe puntare sulla creazione di un corretto ed appropriato sistema organizzativo e di management, specie nel settore pubblico, in grado di adottare misure di sicurezza idonee e di effettuare controlli che garantiscano effettivamente l'assenza di pregiudizi nei sistemi informatici e algoritmici.

L'introduzione della tecnologia, in sostanza, dovrebbe portare – sia pure con i ritardi accumulati dall'Italia – ad una riprogettazione delle tradizionali modalità procedurali per creare dei processi virtuosi che siano allo stesso tempo equi e fattibili.

and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims", in *Cornell Law Review*, 59 (1974), n. 5, p. 772 e pp. 815-816.

Antonio Incampo

L'intelligenza artificiale a confine dell'universo del diritto

Abstract: **Artificial Intelligence at the Frontier of the Universe of Law.** The essay examines the impact of artificial intelligence on law, highlighting its profound ontological, epistemological, and normative effects. AI introduces a new form of logic that differs from traditional reasoning, characterized by rationality devoid of emotion and expressed through digital writing – a "mute and opaque" language that functions through calculations rather than meanings. As a result, legal science faces a crisis, and the role of the jurist becomes uncertain, with the danger of reducing humans to passive objects within machine-like processes.

Keywords: Artificial Intelligence, Law, Digital Writing, Space, Time.

Sommario: 1. Al principio c'era il *logos* … artificiale – 2. Un ordine sociale che non ha più bisogno della politica – 3. La rivoluzione digitale è soprattutto grafica – 4. Il diritto tra digitalizzazione e simbolizzazione – 5. Il diritto dell'intelligenza artificiale senza spazio e tempo – 6. La colomba di Kant – 6.1. *Explicit*

La leggera colomba […] potrebbe immaginare
di volare più agevolmente in uno spazio privo d'aria.

I. Kant

1. Al principio c'era il *logos* … artificiale

L'avvento dell'intelligenza artificiale è l'emergere di un nuovo *logos* (con la lettera decisamente minuscola rispetto a quello giovanneo), una forma di razionalità con tratti radicalmente inediti, almeno alla luce dell'ontologia classica con cui finora abbiamo spiegato il mondo. Per la *Metafisica* di Aristotele, l'Essere si dispiega secondo un ordine gerarchico di enti. Tutti connotati dalla copula "è"[1]. Si va dagli enti inanimati come le pietre, agli

[1] Cfr. Aristotele, *Metafisica*, specialmente libri VII, VIII, XII, trad. it. di G. Reale, Rusconi, Milano, 1978.

esseri viventi privi di sensibilità come le piante; da questi agli animali dotati di sensibilità ma non di razionalità, fino all'uomo che unisce entrambe le facoltà. Ma dove collocare, in questo ordine, entità razionali, eppure prive di sensibilità come gli agenti artificiali? Sarebbe da chiederlo ad Aristotele.

Gli agenti artificiali sono ormai parte integrante del nostro presente. Sono già fra noi. Non ha più senso schierarsi pro o contro. C'è di più. Ciò che si sta configurando è già un mondo decisamente nuovo. Non è semplicemente il "nuovo mondo" (l'America) evocato dalla sinfonia di Dvořák (1893)[2], bensì qualcosa di totalmente altro: un mondo che si pone persino in competizione con il "Mondo 3" di Popper, ovvero con il regno del pensiero e dei prodotti della mente umana[3]. Ne è protagonista una molteplicità di attanti, cioè di attori muniti di "soggettività senza soggetto" pronti a generare sorprendenti universi culturali[4].

2. Un ordine sociale che non ha più bisogno della politica

C'è un'altra novità. La rivoluzione in atto non è neppure una delle tante rivoluzioni politiche. È una nuova e singolare configurazione sistemica del potere, di un potere che non si declina più politicamente. Uno "stato di eccezione" *à la* Schmitt, che però non ha più una "*politische Theologie*" alle spalle. Le nuove soluzioni, infatti, sono nelle mani della tecnica, sono per così dire "ultra-soluzioni". Lo si vede bene con l'ossatura di questo nuovo mondo che, non a caso, si è insediato da tempo nell'economia, alle sorgenti stesse del potere contemporaneo, con un *hardware* di sistemi monetari senza precedenti. Mi riferisco all'infrastruttura della *blockchain,* che aspira a occupare non solo l'intero spazio delle monete *fiat* convenzionali, ma anche

[2] È la celebre *Sinfonia* n. 9 in *mi minore* "From the New World" (*Op.* 95) del 1893, con cui Dvořák cerca di fondere insieme la musica europea romantica con le tradizioni musicali americane che accompagnano la scoperta avventurosa dei vasti paesaggi d'America.

[3] Cfr. K. Popper, *I tre mondi. Corpi, opinioni e oggetti del pensiero*, trad. it. di P. Rumore, il Mulino, Bologna, 2012, *passim*.

[4] Riprendo il lessico di D. Campanale nella sua opera omonima: *Soggettività senza soggetto*, Adriatica, Bari, 1990.

quello delle pubbliche amministrazioni, e poi degli Stati e delle stesse democrazie. Altro che "*the end of history*", come annunciava non molto tempo fa il noto politologo statunitense Francis Fukuyama, in un libro di grande successo internazionale!

La *blockchain* è l'infrastruttura necessaria delle valute virtuali, ma anche la logica sistemica della società digitale. Apparentemente democratica nella sua orizzontalità (la relazione *peer to peer* è la sua essenza)[5], si limita a intercettare gli interessi degli individui senza mai prefiggersi una sfera pubblica del discorso[6], lontana com'è dalla *Öffentlichkeit* kantiana, o da una forma pub-

[5] Il sistema *blockchain* è per definizione decentrato o "disintermediato". Non esiste un *hub*; ogni nodo raccoglie gli stessi *links*. Cfr. G. Rinaldi, "Approcci normativi e qualificazione giuridica delle criptomonete", in *Contratto e Impresa*, (2019), n. 1, p. 263.

[6] Ha ragione S. Cassese nel collegare la circolazione delle criptovalute alla fine dello Stato e della sua autorità. Cfr. Id., *Territori e potere. Un nuovo ruolo per gli Stati?*, il Mulino, Bologna, 2016, p. 35.

Non c'è dubbio che nella trasformazione informatica della comunicazione si vada riducendo il concetto di privato in favore di un maggiore accesso di tutti a tutto. È l'avvento dello *streaming*, del *peer-to-peer*, dei *social networks* e del *free source software*. Al centro, tuttavia, non c'è il soggetto in una vera comunità di parlanti; la direzione di ogni condivisione si sposta sulla centralità dell'oggetto, ovvero di ciò di cui si parla con un effetto dispersivo e centrifugo della sfera pubblica del discorso. La terra che aveva finora attratto in esclusiva il *nomos* è sostituita dal *cloud*, lo spazio imprecisato di una nuova *internet society* dove l'interazione dei soggetti, come osserva acutamente P. Sommaggio, dà vita certamente a una straordinaria infrastruttura che rilancia la grande visione di una società globale, azzerando l'organizzazione della società in stati nazionali, a patto però di comprenderne i rischi. I rischi, infatti, sono, da un lato, l'inconsapevolezza del forte impatto del *cloud computing* sul corpo, sulla mente e sul concetto stesso di conoscenza umana, e, dall'altro, l'assenza di tutele sufficienti dei diritti di coloro che stipulano un contratto di *cloud* con le stesse caratteristiche di un nuovo vero contratto sociale. Cfr. P. Sommaggio, "Dalla scrivania alla nuvola e ritorno. Riflessioni filosofico-giuridiche sul *cloud computing*", in P. Moro, C. Sarra (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Franco-Angeli, Milano, 2017, pp. 207-214. Sulla pervasività del *could computing* in ordine al concetto di conoscenza, e sul progetto di una *Network Society*, si vedano rispettivamente: R. Clowes, "Thinking in the Cloud. The Incorporation of Cloud-Based Technology", in *Philosophy & Technology*, 28 (2015), pp. 261-296; M. Castells, *The Rise of Network Society*, Wiley Blackwell, Oxford, 1996; J.A.G.M. van Dijk, *The Network Society. Social Aspects of New Media*, SAGE Publications, London-Thoudand Oaks-New Delhi, 2006;

blica di riconoscimento in una comunità di parlanti pronti al dialogo per il migliore dei mondi possibili[7]. L'economia e il sistema sociale digitale procedono *by default*, per passi che osservano la correttezza delle procedure al posto degli scopi, rivolgendosi unicamente a se stessi per proclamare la morte della politica.

Ed ecco la conclusione. Lo "stato di eccezione" non è più, come teorizzato da Schmitt, la discontinuità di un ordine prima della ripresa dello stesso ordine attraverso le categorie politiche della sovranità ("il sovrano è chi decide nello stato di eccezione")[8]. Lo "stato di eccezione" è qui la discontinuità radicale con un ordine che non ha più bisogno della politica.

È questo un punto centrale da non trascurare.

C. Orban, *La società della rete. Nuove idee per l'uomo del futuro*, Hoepli, Milano, 2015.

[7] Il riferimento è alla "formula trascendentale del diritto pubblico" di Kant nella seconda appendice di *Zum ewigen Frieden*, secondo cui il discorso giuridico ha bisogno della forma della pubblicità per non cadere in contraddizione con se stesso. La norma giuridica per Kant, senza la "pubblicità" [*Öffentlichkeit*], è letteralmente "controgiuridica" [*rechtswidrig*].

Il termine '*rechtswidrig*', tradotto spesso erroneamente con 'illegale', indica, appunto, una situazione di contraddittorietà del diritto con la sua stessa idea. Cfr. I. Kant, *Zum ewigen Frieden. Ein philosophischer Entwurf*, in: *Kleinere Schriften zur Geschichtsphilosophie, Ethik und Politik*, Felix Meiner, Leipzig, 1913, p. 163; trad. it. (con alcune mie modifiche) di V. Cicero: *Pace perpetua*, Rusconi, Milano, 1997, p. 149.

[8] C. Schmitt, *Politische Theologie. Vier Kapitel zur Lehre von der Souveränität*, Duncker & Humblot, Berlin, 2009[9], p. 13; ed. it. a cura di G. Miglio e P. Schiera: *Teologia politica. Quattro capitoli sulla dottrina della sovranità*, in Id., *Le categorie del politico. Saggi di teoria politica*, il Mulino, Bologna, 1972, p. 33. Per Schmitt lo "stato di eccezione" [*Ausnahmezustand*] non è il presupposto per abbandonare l'ordine giuridico; il sovrano è chi decide nello stato di eccezione, ma sempre in vista di un ordine giuridico che non rinunci alla sua essenza politica. La decisione ne è la prova. Nel caso, invece, della rivoluzione digitale si sconvolgono alla radice le strutture politiche dell'ordine giuridico. Se c'è un nuovo ordine, non ha affatto i tratti politici della decisione. Le decisioni sono ultra-decisioni; fondate sulla correttezza, si comportano come un fatto esclusivamente tecnico. Sono *by default*. L'inizio di questa rivoluzione è indubbiamente radicale, come non si era mai visto. E lo stato di eccezione sta a rivendicare, come direbbe W. Benjamin, la forza e la legittimazione in sé di quest'ultima, indipendentemente dal fine che intende raggiungere. Cfr. W. Benjamin, *Il dramma barocco tedesco*, Einaudi, Torino, 1999, specialmente p. 39 e ss.

3. La rivoluzione digitale è soprattutto grafica

Per comprendere, però, fino in fondo, la portata sia ontologica sia epistemologica di una simile rivoluzione, occorre partire dal sottostante, da ciò che non si vede e ne sta alla base. A muovere tutto questo, infatti, è un inedito *linguistic turn*, una svolta senza precedenti della scrittura digitale. Lo evidenziano, tra l'altro, A. Garapon e J. Lassègue nel loro recente libro *Justice digitale* (2018)[9]. La rivoluzione di cui si parla è, soprattutto, grafica. Usando il linguaggio navale, la particolarità della sua scrittura è la vera "opera viva" dell'impresa digitale: è la parte sommersa della nave che la fa galleggiare; le infrastrutture e i sistemi operativi in mano agli utenti sono solo l'"opera morta", ossia la parte emergente dello scafo che non serve a solcare i mari.

La scrittura digitale è profondamente diversa da quella alfabetica o ideogrammatica. È "muta e opaca". E lo è per definizione, e non semplicemente a causa delle maggiori o minori capacità di accesso degli utenti. Una scrittura numerica (di codici binari di 0 e 1) che non parla, né è fatta per parlare[10]. Serve solo a identificare una traccia (un dato che è a sua volta un'altra sequenza numerica), senza interessarsi minimamente del senso. Con una conseguenza cruciale. Se non è fatta per parlare, non è neppure idonea a pensare o a chiedersi: "Cosa sto dicendo?". Un'operazione di calcolo, più che di senso. Proprio per questo è muta e cieca. Perfettamente adatta all'"ignoranza inevitabile" dell'art. 5 c.p. che immunizza dalla responsabilità personale. Non potresti mai chiamare in causa un simile sistema di scrittura perché non avrebbe mai consapevolezza dei suoi risultati. Il suo problema è la totale irresponsabilità sistemica delle sue azioni.

[9] Si veda A. Garapon, J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Presses Universitaires de France, Paris, 2018; ed. it. a cura di M.R. Ferrarese, trad. it. di F. Morini, *La giustizia digitale. Determinismo tecnologico e libertà*, il Mulino, Bologna, 2021.

[10] Lo stesso discorso vale per un calcolatore quantistico e cioè un computer che sfrutta le proprietà quantistiche della materia per effettuare operazioni su dei dati. A differenza di un calcolatore classico basato sulla codificazione dei dati in *bit* (0 o 1), il computer quantico opera con *bit* quantistici (*qubit*) che possono restituire non solo valori di 0 e 1, ma anche una combinazione di entrambi.

Un linguaggio, dunque, diversissimo da quello evocato da Heidegger come "casa dell'Essere". In *Brief über den Humanismus* (1946) Heidegger proclama che "la parola è la casa dell'Essere" [*die Sprache ist das Haus des Seins*]. La scrittura digitale, per sua natura, non si rivolge alla verità, né tantomeno all'Essere. Il suo movimento è radicalmente opposto a quello della parola in senso heideggeriano: un movimento inaudito, in cui la verità nasce dal segno, e non viceversa. Se, da un lato, la parola rimanda sempre a un significato del mondo, dall'altro il codice digitale esiste solo nella relazione tra codici. Il suo dire è in fondo privo di pensiero: articola enunciati, ma non sa ciò che attesta; sa rispondere alle nostre domande, eppure è incapace di comprendere[11].

Secondo Kant i fenomeni senza i concetti sono ciechi. Per i *big data* è la stessa cosa: senza la riflessione, privi cioè dell'attività intellettuale del soggetto, anch'essi sono ciechi. Non basta che siano combinati a dovere da algoritmi in grado di costruire discorsi dotati di senso. Il motivo, infatti, non è la mancanza di logica. La logica c'è ed è rigorosissima. Il suo limite, semmai, è quello di prestarsi alla mera associazione sintattica dei dati. Mai al loro significato. Né è il risultato al termine di una fase evolutiva che potrebbe prefigurare futuri guadagni di segno completamente diverso. Chi guarda all'opera viva della scrittura digitale sa che non potrà mai esserci questo salto. Il linguaggio numerico dei codici digitali non ha proprio a cuore la verità. Non la cerca neppure.

4. Il diritto tra digitalizzazione e simbolizzazione

Per capire meglio la differenza tra scrittura digitale e parola è proprio il fenomeno giuridico a offrire una chiave di lettura davvero illuminante. Emergono due forme decisamente diverse di normatività. La forma qui conta. È la stessa che per Aristotele dà

[11] È una tesi ormai scontata fra filosofi, informatici e giuristi. Cito solo in tal senso alcune opere di rilevanza internazionale: E.J. Larson, *The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do*, Harward University Press, Cambridge (Mass.), 2022; G. Gigerenzert, *How to Stay Smart in a Smart Word. Why Human Intelligence Still Beat Algorithms*, MIT Press, Boston (Mass.), 2022.

esistenza alle cose – *forma dat esse rei*. Da un lato, la descrizione del reale con un *Norm-Satz*, ossia una proposizione normativa imperniata sul rapporto di significato con la realtà; dall'altro, la ricezione di semplici impronte informatiche fuori da un orizzonte fondativo di senso. Alla prima appartiene, ad esempio, la proposizione del codice civile: "Il contratto è l'accordo di due o più parti per costituire, regolare o estinguere tra loro un rapporto giuridico patrimoniale" (art. 1321 c.c.); alla seconda, invece, il computo letteralmente cieco (il significato è solo al termine della ricostruzione dei segni) di dati sfruttati sintatticamente da funzioni algoritmiche prive di una loro autonoma riflessione sulla realtà.

Non è una metamorfosi esente da serie ripercussioni. Ne soffre il diritto che non guarda più in faccia alla realtà. La volontà, per dirla con Gentile, smette di essere volontà *volente*; è, piuttosto, volontà *voluta* del *déjà vu* di una moltitudine di dati archiviati. Non c'è prima la realtà, e poi la ragione giuridica ad essa più funzionale, bensì una continua soluzione astratta della realtà, una soluzione conservatrice per definizione dei propri codici. Nulla di paragonabile allo sguardo struggente al passato da parte dell'*Angelus Novus* di Benjamin prima dell'ultimo giudizio sul mondo e sulla storia. L'intelligenza artificiale non tratta più la realtà (almeno direttamente), né distingue il passato dal futuro. Anzi, il tempo vissuto di passato e futuro è solo una tremenda finzione. Di fronte all'intelligenza artificiale il tempo è un concetto solo ingannevole.

Alla crisi del diritto si accompagna quella della scienza del diritto. La digitalizzazione è in pratica l'"accadere di Nessuno". Solo a Nessuno appartiene il linguaggio tecno-funzionale, o il linguaggio della prassi manageriale e del funzionamento gestionale cui è ridotta l'esperienza giuridica. Con quali conseguenze per la scienza giuridica? Forse è la fine definitiva del giurista.

Potrebbe pure darsi un mondo senza giurista. Niente lo vieterebbe[12]. Ma resterebbe, comunque, l'interrogativo sull'uomo.

[12] Sulle incognite di una tale visione, legate alla logica incompiutezza dei sistemi algoritmici, si veda P. Marra, "Rischi algoritmici. Alcune questioni sulla digitalizzazione della giustizia", in *Journal of Ethics and Legal Technologies*, 7 (2025), pp. 145 ss.

Ci sarebbe, cioè, da chiedersi se tutto ciò sia compatibile con un uomo di natura inadatto a un destino meramente oggettivante. L'uomo, infatti, è costitutivamente aperto alla capacità di scegliersi, lontano dalle sorti della semplice contingenza dei casi, o della necessità inderogabile delle cause[13]. La rassicurante certezza algoritmica non ha niente a che fare con il senso della comunità di parlanti impegnati a scegliersi per non regredire da soggetti a oggetti del mondo della vita. È una conclusione segnata fin qui dall'evidenza.

Ma andiamo per gradi.

Il diritto è chiamato in causa dall'intelligenza artificiale in almeno due accezioni diverse: (i) come ordine giuridico *lato sensu*, in veste cioè di regole che, indipendentemente da ordinamenti tradizionali, hanno un ruolo comunque significativo nei processi di regolazione sociale; (ii) come ordine giuridico *stricto sensu*, ossia come insieme normativo riconducibile a ordinamenti giuridici formalmente costituiti.

Il primo ordine s'identifica per così dire con la *Natur der Sache* degli apparati digitali, cioè con la loro intrinseca ordinamentalità. Se ne vedono già gli effetti. Gli *Internet Service Provider* (ISP) si procurano, infatti, da soli l'autogestione e l'autoregolamentazione dei loro servizi. È il passaggio decisivo a un sistema non più convenzionale di regolazione sociale. Come si dice: "*code is law*". Il "*code*", però, non sono i "codici" tradizionali, ma la stessa infrastruttura del mondo digitale. La normatività, in questo caso, è la conseguenza inevitabile degli apparati utilizzati, la loro appunto "natura della cosa". Gli utenti non possono che adeguarvisi[14].

[13] Riprendo una riflessione di B. Romano qui di seguito: "Lo spegnersi del *giurista*, divenuto l'operatore delle tecno-norme, marca anche la fine del *filosofo*, sostituito da figure attive nella celebrazione 'scientifica' del 'caos della necessità', espressione ricordata da Heidegger e destinata a segnare che la questione del senso diviene senza senso. Il giurista ed il filosofo si trasformano in luoghi desoggettivati dell'Accadere del Nessuno; diventano entità 'parlate', pur continuando a presentarsi come 'parlanti', autori del loro dire-scrivere". B. Romano, *Scienza giuridica senza giurista: il nichilismo 'perfetto'*, G. Giappichelli, Torino, 2006, p. 8.

[14] Il *code* non è una legge e non ha limiti costituzionali, ma agisce come una legge e forse con maggiore efficacia perché controlla direttamente le azioni. È una forma di *soft law*; violarlo (violando i sistemi di protezione

Gli agenti artificiali sono parte integrante di tali apparati. Il rapporto è tra genere e specie. Ogni cosa si predichi del mondo digitale (genere) è pure una caratteristica intrinseca dei sistemi esperti (specie). I sistemi esperti, in altre parole, non fanno che accelerare, implementare, moltiplicare i prodotti della rivoluzione digitale. In tal senso, sono accomunati da una loro autonoma ordinamentalità.

Provo, dunque, ad elencare alcune metamorfosi sistemiche rispetto agli ordinamenti convenzionali[15]. (i) La programmazione informatica si sostituisce, intanto, alla legislazione. Le fonti normative coincidono con le stesse infrastrutture digitali e i loro effetti disciplinari sugli utenti. Non è un passaggio senza conseguenze negative. Basti pensare alla privatizzazione di massicci processi di regolazione sociale carenti di sufficienti controlli pubblici. Un potere paragonabile a quello economico dei mercati internazionali, con un'ulteriore nota di oscurità. Anche la nuova *lex mercatoria*, infatti, ha conosciuto la privatizzazione del linguaggio nomothetico, ma non a patto di essere definitivamente anonima. Da questo punto di vista, il confronto tra le norme delle grandi *law firms* americane e il codice informatico appare impari: se le prime sono tutte alla luce del sole, il secondo, invece, è del tutto sconosciuto perfino ai programmatori di *deep learning*[16] – o, se è conosciuto, è puntualmente protetto dal segreto industriale[17]. (ii) Scompare praticamente la costruzione a

e di accesso) equivale a violare *ex ante facto* la legge. Cfr. M.R. Ferrarese, *Diritto sconfinato. Inventiva giuridica e spazi nel mondo globale*, Laterza, Roma-Bari, 2006, *passim*; L. Lessig, *Code and Other Laws of Cyberspace*, Basic Books, New York, 1999, *passim*; Id., *Il futuro delle idee*, Feltrinelli, Milano, 2006, pp. 181-184; E. Maestri, "*Lex informatica* e *soft law*. Le infrastrutture normative del cyberspazio", in P. Moro, C. Sarra (a cura di), *op. cit.*, in particolare p. 167 e ss.

[15] Prendo spunto dall'articolo di J.R. Reidenberg, "Lex Informatica: the Formulation of Information Policy Rules through Technology", in *Texas Law Review*, (1998), n. 3, p. 566.

[16] Cfr., fra gli altri, sull'"oscurità" del modello di apprendimento *deep learning* delle macchine, con particolare attenzione all'economia, F. Pasquale, *The Black Box Society. The Secret Algorithms which Control Money and Information*, Cambridge University Press, Cambridge, 2015.

[17] Sul potere giuridico delle *law firms* americane, alla base della nuova *lex mercatoria*, si veda in particolare F. Galgano, *Lex Mercatoria*, il Mulino,

gradi dell'ordinamento (lo "*Stufenbau*" della logica normativistica del diritto di Kelsen). La gerarchia delle norme è soppiantata dall'orizzontalità di sistemi come la *blockchain*, orientati unicamente all'efficienza esecutiva, più che alla mediazione verticale di istituzioni preposte alla tutela dell'interesse collettivo. (iii) In concorrenza con la mediazione delle organizzazioni politiche si fa strada un semplice sistema di circolazione delle notificazioni. (iv) L'ubiquità è il tratto distintivo del nuovo ordine digitale. Una circostanza che destabilizza gli ordinamenti convenzionali. Lo vedi con alcuni principî messi alla prova dalla de-territorializzazione dei fatti giuridici. Scompare (o è almeno impraticabile), ad esempio, il principio di "territorialità temperata del diritto penale (art. 6, 1 c.p.), secondo cui la legge si applica ai reati commessi nel territorio dello Stato. Con l'avvento soprattutto del *web* 3.0 (il vero "multiverso di metaversi") è più facile rilevare in uno Stato gli effetti dannosi di azioni compiute nei confini di altri Stati in cui spesso la condotta non è neppure reato. Il reato, in poche parole, diventa letteralmente *u*-topico, ossia reato di un luogo che non c'è. (v) Non è una metamorfosi parziale. Ad essere interessato è tutto il diritto. Non solo quello pubblico. La normazione digitale detta le proprie condizioni, e lo fa passando dalla sfera pubblica all'autonomia privata. Lo vedi in ogni accesso al sistema. Agli "*essentialia negotii*" della celebre *Teoria generale del negozio giuridico* (1943) di Emilio Betti, vi si oppongono le regole di configurazione di tanti applicativi che a dir poco abusano del requisito essenziale del consenso – l'icona dell'estrema sottomissione è l'incondizionato "*I agree*" –, trasformando il presupposto della libertà in un mero risvolto di *legal design*. (vi) È il trionfo, infine, della correttezza del sistema al posto dell'idealità. La rivoluzione digitale, come ho detto, non ha una teologia politica alle spalle. Le sue soluzioni sono nelle mani della tecnica per assecondare indifferentemente tutto ciò che essa può offrire.

Qual è, allora, lo scopo? Non certo quello di creare un ordine. Le infrastrutture algoritmiche spingono in sostanza verso il disordine. Non mirano ad altro se non ad assecondare il continuo

Bologna, 2010, specialmente il capitolo IX su "La nuova '*lex mercatoria*'", pp. 239-256. Il saggio di Galgano è uno dei primi studi fondamentali sui grandi mutamenti delle fonti attuali del diritto.

mutare del mondo[18]. *Google Scholar* si prefigge, ad esempio, di distruggere i diritti nazionali sulla proprietà intellettuale per creare una propria biblioteca digitale. Con quale giustizia? Non certo con la giustizia intesa come "lealtà fra i membri di una comunità politica" – la *justice as fairness* di cui parla Rawls[19]. Qui conta piuttosto una forma di *justice as fitness*, una giustizia di adattamento[20]. L'immagine che meglio la rappresenta potrebbe essere quella di un "grande orologiaio" che assegna a ciascuno un orologio, ognuno regolato su un tempo diverso.

5. Il diritto dell'intelligenza artificiale senza spazio e tempo

E arriviamo al diritto convenzionale, al diritto *stricto sensu* degli ordinamenti appunto *à la* Kelsen. Quali sono i cambiamenti ai tempi dell'intelligenza artificiale? Se prima analizzavo le differenze *ab extra*, adesso vedrò quelle *ab intra* degli ordinamenti. Le variazioni non sono poche, e toccano ancora una volta la stessa ontologia del diritto.

Nei decenni del secolo scorso il diritto ha visto novità inedite al suo interno. Si pensi alla giustizia penale internazionale, al diritto sovralegale o cogente di tale giustizia, alla *restorative justice* come approdo altamente significativo del concetto di pena, e più in generale al ruolo dei diritti e del diritto globale "per una pace perpetua" in senso kantiano. Ora, la *digital disruption* è all'origine di un cambiamento di tutt'altra natura. È una trasformazione che incide profondamente sulla struttura dell'ordinamento. Voglio soffermarmi in particolare su due categorie essenziali della validità giuridica – stando almeno a una dottrina robusta come la *reine Rechtslehre* di Kelsen. Mi riferisco allo spazio e al tempo.

Il digitale è un grande disorganizzatore di tali categorie. Si prenda lo spazio. La rete cerca innanzitutto l'ubiquità, o più pro-

[18] Cfr. A. Garapon, J. Lassègue, *op. cit.*, p. 232.

[19] Scrive Rawls: "I have tried to set forth a theory that enables us to understand and to assess these feelings about the primacy of justice. *Justice as fairness* is the outcome: it articulates these opinions and supports their general tendency" (*A Theory of Justice*, Belknap Press of Harvard University Press, Cambridge (Mass.), 1999, p. 586. Il corsivo è mio).

[20] Cfr. A. Garapon, J. Lassègue, *op. cit.*, p. 240.

priamente un non-spazio[21]. Non a caso. Internet è un non-luogo logico e concettuale; è un codice che trasmette dati tra computer in rete. È il più grande non-luogo pubblico che si sia mai potuto immaginare. Raccoglie oggetti che variano da particelle subatomiche come i *bit* fino ad ammassi di galassie come Google[22].

Lo si è visto prima con la "de-territorializzazione" sistemica della rivoluzione digitale. I sistemi esperti, a loro volta, spingono ancora di più verso l'astrazione fisica dello spazio. Non per altro perché non hanno un corpo senziente. Il mio computer è solo un supporto; non percepisce il mondo che lo circonda. La sua adattabilità a ogni luogo è il proprio pregio.

E il diritto? Quali sono le conseguenze? L'"ubiquità" contrasta con l'idea stessa di esistenza giuridica. Una norma non è indifferentemente valida qui, piuttosto che lì. Per definire la validità di una norma, occorre definire lo spazio entro cui accade. È il senso della definizione kelseniana di "validità" [*Geltung*] come "esistenza specifica di una norma" [*spezifische Existenz einer Norm*][23].

Ma lo spazio è decisivo anche sotto l'aspetto delle procedure. Faccio un esempio con gli sviluppi più colti della mediazione e della conciliazione nel processo civile. Il riferimento è ovviamente alla cultura delle "*alternative*" nelle "*dispute resolutions*". In un prossimo futuro, a causa dello sviluppo delle ICT (*Information and Communication Thechnologies*) e dello IoT (*Internet of Things*), sarà difficile distinguere una ADR (*Alternative Dispute Resolution*) da una ODR (*Online Dispute Resolution*)[24].

[21] Il "non-spazio" a cui mi riferisco non va confuso con i "non-luoghi" di Marc Augé (cfr. Id., *Non-lieux. Introduction à une anthropologie de la surmodernité*, Éditions du Seuil, Paris, 1992). I non-luoghi, infatti, sono spazi prevalentemente di transito, concepiti per essere attraversati piuttosto che abitati, e dunque marginali rispetto alle nostre scelte esistenziali. Il non-spazio della rete, al contrario, rappresenta una sorta di oblio della terra su cui poggiano stabilmente i nostri piedi.

[22] Cfr. E. Maestri, *op. cit.*, p. 169.

[23] Si pensi alle condizioni di validità presupposte dal modello kelseniano della "costruzione a gradi" [*Stufenbau*] dell'ordinamento, condizioni per cui una norma non esiste in assoluto, ma a partire dall'oggettivazione storica e spaziale di un ordinamento. Si veda H. Kelsen, *Lineamenti di dottrina pura del diritto*, trad. it. di R. Treves, Einaudi, Torino, 1967, in particolare p. 105.

[24] Le facilitazioni del giurista ibrido sono sotto gli occhi di tutti. Si pensi ai servizi legali *multitasking*, agli *smart contracts*, all'accesso ai *big data*, al

Nello stesso tempo verranno sempre più a galla le diversità e le reciproche incompatibilità. Non si può trascurare, infatti, come la cultura delle *alternative* sia sempre improntata ai tratti di originalità e irripetibilità di ogni singola controversia. E il cuore di tale soluzione è proprio il luogo d'incontro delle parti, luogo fisico da cui dipendono, tra l'altro, le componenti non verbali e a volte decisive per la mediazione con più frutti. Le ODR, invece, si frappongono con nuovi filtri. La causa è proprio la loro condizione di ubiquità. Essendo *machine based*, le ODR derogano alla necessità d'incontrarsi in un luogo. Da qui la loro crescente istituzionalizzazione e standardizzazione, e il bisogno di essere presidiate da autorità pubbliche. In questa direzione, tra l'altro, la Commissione delle Nazioni Unite per il diritto commerciale internazionale ha elaborato nel 2016 alcune note tecniche sulla "*online dispute resolution*". È inutile sottolineare come più le procedure si istituzionalizzino e si standardizzino, e più s'infrangano l'originalità e l'irripetibilità delle "*alternative*"[25].

Anche il tempo non è più quello tradizionale del diritto. I sistemi di intelligenza artificiale distruggono la durata, e fanno della velocità il loro tempo ordinario. Partirò dall'economia, per poi concludere sul diritto e sul processo.

Il tempo della finanza digitale mette fuori gioco quello tradizionale del diritto (stagioni, mesi, giorni, ore), e lo fa ad appannaggio esclusivo della velocità. Basti considerare il meccanismo di negoziazione chiamato "*High Frequncy Trading*" (HFT). Quando il trader X avvia un ordine multiplo di titoli N, a partire da un mercato Y, quasi simultaneamente gli stessi titoli N

legal design per le spiegazioni giuridiche. PayPal e Amazon hanno già avviato sistemi per la risoluzione *online* delle controversie del commercio elettronico. Si usano modelli di argomentazione giuridica molto elaborati come quelli sperimentati da Thomas Gordon in *The Pleadings Game* (1993), modelli basati sulla simulazione dello scambio di memorie scritte. Resta sempre, però, il problema del non-luogo dell'attività giuridica *online*. Lo spazio è anch'esso un termine logico dell'argomentazione giuridica. Cfr. P. Moro, "L'avvocato ibrido. Tecnodiritto e professione forense", in P. Moro, C. Sarra (a cura di), *op. cit.*, specialmente pp. 18-19.

[25] Sulla "istituzionalizzazione" e "standardizzazione" delle ODR si veda, fra gli altri, L. Mingardo, "*Online Dispute Resolution*. Involuzioni ed evoluzioni di telematica giuridica", in P. Moro, C. Sarra (a cura di), *op. cit.*, pp. 130-136.

risultano più cari sugli altri mercati a causa dell'HFT. Questo costringe X a sostenere un investimento maggiore. Tutto si gioca nell'arco di pochi centesimi di secondo[26].

Il diritto evidentemente non se lo può permettere. La velocità non è il tempo del diritto. Il caso dell'economia è emblematico. La velocità nei mercati confonde, infatti, i diritti soggettivi con l'interesse individualistico. Non guarda certo in faccia alla democrazia, né tanto meno tratta il debito pubblico degli Stati più poveri per proteggere le loro politiche sociali. È il tempo della cecità macchinica. Del resto, l'efficienza non è la riflessione; la soluzione non c'entra affatto con la giustificazione. Ce lo dice una volta per tutte la logica inaudita della macchina: alla risposta non interessa il significato; conta solo, come s'è visto, la correttezza algoritmica di una lunga sequenza numerica[27].

Ora, la velocità è solo un sintomo della malattia. Non è né necessaria, né sufficiente al funzionamento dei sistemi giuridici ed economici. D'altronde, finché la risposta ha un ruolo ausiliare, e serve solo ad accompagnare le decisioni dell'uomo, va bene. Sarebbe stupido pensare il contrario. La questione (la vera malattia) emerge, invece, se la soluzione della macchina sostituisce completamente il giudizio umano. In altri termini, il problema non è tanto l'uso in sé dell'intelligenza tecnico-funzionale, ma l'assolutizzazione del suo metodo, ossia il dominio dell'intelli-

[26] Il tempo dello scambio tra sistemi macchinici esperti da un lato, e, dall'altro, lo spazio incalcolabile generato dai *server* per la negoziazione dei mercati finanziari mettono decisamente in crisi il tempo e lo spazio comuni dell'uomo, e di conseguenza le categorie principali del diritto come fatto umano. Per l'ulteriore approfondimento si veda A. Lupoi, "Dal mercato 'alle grida' al Sistema Multilaterale: l'ordine tecnico", in P. Moro, C. Sarra (a cura di), *op. cit.*, pp. 141-155.

[27] La velocità disumana delle macchine crea un vero antagonismo tra i sistemi artificiali e la realtà sociale costruita dall'uomo. Rispetto ai primi l'uomo non può che essere spettatore, non avendo i mezzi necessari per seguire ciò che si compie macchinicamente nell'oltre misura del tempo. Il rischio di soccombere è a portata di mano. Cfr. B. Romano, *Nichilismo finanziario e nichilismo giuridico, conoscenza e coscienza*, G. Giappichelli, Torino, 2012, specialmente p. 44.

Un'attenta riflessione sull'inevitabile esito spettatoriale dell'uomo comune nella vita dei sistemi giuridici contemporanei è nell'ottimo volume di G. Siniscalchi, *Barocco Giuridico. Osservatori, osservanti, spettatori*, Franco-Angeli, Milano, 2017.

genza artificiale sulla stessa intelligenza artificiale. Ecco la malattia. È, tra l'altro, l'osservazione di Nietzsche sui riduzionismi del metodo scientifico: "Ciò che caratterizza il XIX secolo non è la vittoria della scienza, ma il trionfo del metodo scientifico sulla scienza"[28]. Non si può certo attribuire alla scienza la colpa di oltrepassare i propri confini. La questione, semmai, è l'uso che si fa della stessa scienza. È l'equivoco di fondo dello scientismo tecnologico, che assume come dogma l'idea che la scienza rappresenti tutto il sapere possibile, e che essa sia l'unico sapere di cui abbiamo bisogno. Lo stesso fraintendimento si ripropone oggi con l'intelligenza artificiale: si afferma che essa rappresenti l'unica forma autentica di pensiero, l'unica davvero corretta.

All'origine c'è l'abbandono della parola. È il passo inevitabile e impossibile dell'intelligenza artificiale. Inevitabile perché il suo linguaggio è costituito necessariamente da codici numerici differenti dalla parola; impossibile perché non c'è sapere dell'uomo che possa fare a meno della parola. L'uomo e il diritto, come direbbe Legendre in *Ce que l'Occident ne voit pas de l'Occident* (2004), sono "legati dalle parole" [*on lie ... par les paroles*][29]. Evidenzierò brevemente due fallacie: una retorica o argomentazionale; l'altra metafisica.

Una fallacia retorica.

Prendiamo l'interpretazione. Fare l'elenco di tutti i risultati possibili di tutti i tipi di interpretazione non è affatto interpretare. Un sistema giuridico esperto (un sistema di intelligenza artificiale) è giusliberista e politico solo per la metodologia complessiva, ma si limita a funzioni di classificazione che non hanno nulla a che fare con la decisione. Si possono pure conoscere tutti i risultati di tutti i tipi di interpretazione, o secondo tutti i procedimenti della logica giuridica, o tutti i criteri di vigenza del diritto, ma tutto questo non è ancora "interpretare"[30]. Interpretare

[28] F. Nietzsche, *La volontà di potenza. Saggio di una trasvalutazione di tutti i valori*, a cura di M. Ferraris e P. Kobau, Bompiani, Milano, 2001, p. 265.

[29] P. Legendre, *Ce que l'Occident ne voit pas de l'Occident*, Mille et une nuits, Paris, 2004, p. 67. La citazione riprende una massima che Legendre legge in A. Loysel: "On lie les bœuf par les cornes, et les hommes par les paroles".

[30] A rilevarlo di recente è P. Marra, *op. cit.*, pp. 153 ss., nel tentativo di avvalorare una prospettiva *undesign* dell'intelligenza artificiale in campo giudiziale.

è un atto che fa della soggettività un termine addirittura inglobante. C'è, infatti, da un lato, ciò che è interpretato (l'oggetto), e, dall'altro, non solo chi interpreta, ma anche il pubblico terzo dell'interpretazione (a volte lo stesso interprete che si sdoppia). Dal punto di vista ermeneutico, ha ragione l'idealismo che vede la vita dell'"oggetto" proprio nel "concetto", o nell'ideale soggettività del "soggetto"[31]. Tutto l'opposto, in breve, della rivoluzione grafica dei sistemi esperti, rivoluzione che fa dell'oggettivazione del dato la sua ragion d'essere. Il dato numerico non è fatto per essere interpretato; la parola, però, sì.

Una fallacia metafisica.

La tecnica è lì a un passo dal nichilismo (incluso quello giuridico). Richiede contraddittoriamente sia l'azione di un uomo forte che annunci la "morte di Dio", sia lo svuotamento dell'uomo stesso fino a Nessuno dell'intelligenza artificiale[32]. Ad ingannare è il suo presupposto. La tecnica, infatti, ha tutto il potenziale per mettere l'uomo dalla parte della natura naturante (*natura naturans*, natura che genera in veste di un creatore), dopo averlo richiamato a sé, al pari dell'opera di un Dio, dalla *natura naturata* (natura generata e sottomessa a un creatore), in cui l'uomo stesso è apparso originariamente "gettato". Ma questo passaggio che pure assegnerebbe all'uomo un suo speciale primato sul mondo cela un grave equivoco.

Provo a sintetizzarlo.

Il soggetto esiste in quanto ha un oggetto: non può darsi soggetto senza qualcosa che gli si contrapponga. L'oggetto rappresenta la vita del soggetto. Tuttavia, per il soggetto, ogni oggetto non è altro che una delle molteplici oggettivazioni del mondo. Il mondo è l'orizzonte di senso del soggetto: ciò che si oppone al soggetto e ne fonda allo stesso tempo la libertà. Senza questo rapporto con il mondo, il soggetto faticherebbe a riconoscersi. Egli vuole agire e vuole avere proprio per affermare la propria libertà nel mondo.

[31] Scrive Hegel: "Questa realizzazione del concetto, nella quale l'universale è questa unica totalità ritornata in sé, […] è l'oggetto" (G.W.F. Hegel, *Enciclopedia delle scienze filosofiche in compendio*, Laterza, Roma-Bari, 1984, § 193, p. 182; la traduzione italiana, particolarmente esplicativa dal punto di vista teoretico, è di B. Croce).

[32] Vedi B. Romano, *Scienza giuridica senza giurista*, cit., p. 80.

Con l'avanzare della tecnica emerge un paradosso sempre più evidente. I sistemi di intelligenza artificiale lo mettono chiaramente allo scoperto. Tra il loro progresso e la capacità effettiva dell'uomo di dominarli si instaura, infatti, una relazione inversamente proporzionale: quanto più tali sistemi si perfezionano, tanto più si riduce la possibilità umana di esercitarne un autentico controllo. La vastità incommensurabile dei dati impiegati e la complessità dei processi di calcolo che li governano eccedono radicalmente le facoltà del soggetto umano. Su questo punto non vi è ormai più dubbio.

Ed ecco, allora, l'equivoco. Non si passa da un soggetto a un altro, magari più illuminato. Altrimenti, quale sarebbe il problema? Vi sarebbe un soggetto soccombente, ma non necessariamente la rovina. Esisterebbe sempre un altro soggetto, e sarebbe ancora più dotato.

Con l'intelligenza artificiale viene meno, invece, proprio il soggetto. L'uomo partorisce l'Alterità che lo fa soccombere, ma questa Alterità non è a sua volta un soggetto. È l'Alterità senza soggetto che non ha coscienza del mondo. Lo si capisce, come si è detto, dalla scrittura digitale, scrittura estranea alla parola, chiusa autoriflessivamente in codici muti e ciechi. Da un lato, dunque, l'uomo è l'artefice di questa stupefacente Alterità; la rivendica con la forza di un Dio di cui ne proclama tuttavia la morte (almeno teologica). Dall'altro, egli stesso perisce di fronte all'Alterità completamente svuotata di soggettività, fino a ridursi a Nessuno. L'Alterità è Nessuno, e l'uomo a sua volta non è più il soggetto delle tante oggettivazioni del mondo e della propria libertà. È anch'egli Nessuno. Il suo fare e avere hanno smarrito ormai il senso di sé e del mondo.

Parafrasando una nota battuta di Woody Allen, direi che con il nichilismo della tecnica non solo Dio è morto, ma anche l'uomo non sta tanto bene, e con l'uomo anche la filosofia.

6. La colomba di Kant

Vorrei concludere con un'immagine potente di Kant nell'*Introduzione* della *Kritik der reinen Vernunft* (1787). La colomba nel volare, fendendo l'aria che le si oppone come resistenza, proprio a causa di questa resistenza, potrebbe pensare di volare

meglio se l'aria non vi fosse, restando nel vuoto assoluto. Nulla di più fatale per sé e il suo volo. Lo spazio, il tempo, e soprattutto la parola sono per il diritto come l'aria per la colomba ignara del suo ultimo volo[33].

6.1. *Explicit*

Il mito della delega alle macchine potrebbe farci sentire purificati e immuni da responsabilità. Almeno dalla responsabilità delle decisioni. Ma sarebbe un grave errore. Infatti, non farebbe i conti con il potere delle infrastrutture digitali e dei sistemi esperti come apparati macchinici di regolazione sociale[34]. La sua quasi invisibilità non equivale alla sua inesistenza. Il *Deus absconditus* non è meno Dio perché nascosto.

Ne avertas faciem tuam a me![35]

[33] La celebre metafora di Kant è nella *Introduzione* (sezione III) della *Critica della ragion pura* (1787), trad. it. di P. Chiodi, UTET, Torino, 1967, p. 78.

[34] Ogni tentativo di controllo e supervisione umana dell'AI è solo formale, o maldestro. *Ex multis unum*. Leggiamo parte dell'art. 14 dell'*AI Act*, rubricato "Sorveglianza umana": "I sistemi di IA ad alto rischio sono progettati e sviluppati […] in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso". In questo passaggio del Regolamento (UE) 2024/1689 si discute delle c.d. intelligenze artificiali "ad alto rischio", ma non è questo il punto. Qual è il reale obiettivo nel predisporre un controllo umano di questo tipo? Garantire effettivamente sicurezza e trasparenza? Oppure è solo il tentativo di costruire un capro espiatorio? Potrebbe, infatti, accadere che il sorvegliante finisca per dare sistematicamente il proprio benestare, senza un monitoraggio completo e approfondito, data l'ontologica complessità dei sistemi in oggetto. Il dominio di informazioni dei sistemi esperti è sterminato e tendenzialmente inverificabile, specialmente in casi di emergenza. La norma invocata è icastica nel mostrare l'insufficienza e, verosimilmente, l'ingenuità nella predisposizione di misure di contenimento dinanzi all'evoluzione costante delle nuove tecnologie. Si veda, fra gli altri, G. Contissa, F. Galli, F. Godano, G. Sartor, "Il Regolamento Europeo sull'Intelligenza Artificiale. Analisi informatico-giuridica", in *i-lex. it*, 14 (2021), pp. 25-27.

[35] *Salmo* 27: 9.

Giuseppe Lorini, Olimpia G. Loddo, Simone Pinna

Artefatti regolativi digitali, intelligenza artificiale e diritti dell'utente*

Abstract: **Digital Regulatory Artifacts, Artificial Intelligence, and User Rights.** The debate on the impact of artificial intelligence on personal freedom and fundamental rights has intensified in recent years, particularly with the rise of generative AI, which has brought both opportunities and challenges to the fore. However, as highlighted in the field of technological and digital regulation, digital artifacts not based on generative AI can also exert a profound influence on individual freedom. By examining the functioning of certain software tools designed to regulate online interactions and by distinguishing between deontic and adeontic (non-deontic) regulatory artifacts, it becomes possible to identify critical issues that do not necessarily depend on the adoption of intelligent technologies. After an introduction, this chapter is structured in three parts: the first briefly reconstructs the debate on the impact of artificial intelligence on user rights; the second introduces the distinction between deontic and adeontic digital regulatory artifacts and analyzes its usefulness in assessing the compatibility of regulatory tools with user rights. Finally, the third part examines two cases of digital regulatory tools belonging to these two different categories of artifacts, which, while not employing AI, nonetheless shape freedom of expression in different ways.

Keywords: Artificial Intelligence, Digital Regulation, Regulatory Artifacts, User Rights, Content Moderation.

* Il contributo è frutto di un lavoro congiunto e condiviso; i parr. 1 e 4 sono stati redatti da Giuseppe Lorini, il par. 2 da Simone Pinna e il par. 3 da Olimpia G. Loddo. Tutti gli autori hanno discusso e approvato i contenuti, assumendone congiuntamente la responsabilità scientifica.

Contributo realizzato nell'ambito del Piano Nazionale di Ripresa e Resilienza (PNRR), Missione 4, Componente 2, Investimento 1.1, Avviso n. 104 pubblicato il 02/02/2022 dal Ministero dell'Università e della Ricerca (MUR), finanziato dall'Unione Europea – NextGenerationEU – Titolo del progetto: Normative artifacts and normative drawings: investigating non-linguistic regulation (NAND) – CUP: F53D23003500006 – Decreto di concessione n. 968 adottato il 30/06/2023 dal Ministero dell'Università e della Ricerca (MUR).

Sommario: 1. Introduzione – 2. Riflessione sull'IA e i suoi effetti perversi – 2.1. Note storiche sull'IA – 2.2. Il dibattito sul rapporto tra IA e libertà individuale – 3. Due strategie di regolazione digitale che non ricorrono necessariamente all'uso dell'IA – 3.1. Artefatti regolativi digitali *deontici* – 3.2. Artefatti regolativi digitali *adeontici* – 3.3. Strumenti deontici e adeontici di moderazione digitale dei contenuti – 4. Conclusioni: Tutta colpa dell'IA generativa? Effetti perversi di artefatti regolativi digitali senza l'ausilio dell'IA.

1. Introduzione

Il dibattito sull'impatto dell'intelligenza artificiale sulla libertà personale e sui diritti fondamentali si è intensificato negli ultimi anni, a partire dalla diffusione dell'intelligenza artificiale generativa[1]. Nel corso di tale dibattito sono state messe in evidenza sia le potenzialità sia le criticità relative all'impatto che tali tecnologie possono avere sulla sfera giuridica individuale.

Alcuni autori pongono l'accento sulla necessità di un bilanciamento tra tutela di beni giuridici (es. protezione da *hate speech*, sicurezza) e rispetto delle libertà fondamentali[2]. Altri insistono su trasparenza ed *explainability*[3] come requisiti essenziali: in particolare emerge la necessità di rendere comprensibili le predizioni algoritmiche[4]; non a caso le *Ethics Guidelines for Trustworthy AI* (2019) hanno incluso la trasparenza tra i sette requisiti fondamentali che devono essere soddisfatti per garantire l'affidabilità dell'AI[5]. Un terzo filone sottolinea l'importanza

[1] L. Floridi, *The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities*, Oxford University Press, Oxford, 2023; S. Alegre, *Human Rights, Robot Wrongs*, Atlantic Books, London, 2024.

[2] L. Floridi, *Ibidem.*

[3] Il termine si riferisce alla capacità di comprendere e descrivere il processo decisionale di un sistema di intelligenza artificiale in termini comprensibili per l'uomo, fornendo trasparenza su come e perché un modello è arrivato a un particolare risultato. Cfr. P.J. Phillips, C.A. Hahn, P.C. Fontana, D.A. Broniatowski, M.A. Przybocki, *Four Principles of Explainable Artificial Intelligence (NIST Interagency/Internal Report n. 8312)*, National Institute of Standards and Technology, Gaithersburg (MD), 2021.

[4] T.Z. Zarsky, "Transparent Predictions", in *University of Illinois Law Review*, 4 (2013), pp. 1503-1570.

[5] High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, European Commission, Brussels, 2019, *https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai*.

della contestabilità e dei rimedi giuridici proposti dagli algoritmi: Hildebrandt[6] e Yeung[7] richiamano il diritto degli utenti a contestare decisioni algoritmiche e a disporre di canali di revisione umana. Secondo Hildebrandt, "[u]n divieto delle decisioni basate sulla profilazione dovrebbe mitigare l'erosione della presunzione di innocenza. […]. È imperativo che l'uso dell'IA venga testato e sottoposto a scrutinio da parte di poteri di controllo e che possa essere contestato in un tribunale"[8]. Infine, vi è un consenso crescente sulla necessità di prevenire fenomeni di discriminazione algoritmica. Ad esempio, si è mostrato come sistemi di gestione sanitaria possano incorporare *bias* razziali[9]; l'High-Level Expert Group on AI richiama espressamente nelle sue linee guida i principi di equità e non discriminazione[10].

Tuttavia, come rilevato anche nell'ambito della regolazione tecnologica e digitale[11], artefatti digitali privi di componenti di intelligenza artificiale possono incidere profondamente sulla libertà individuale. Attraverso l'analisi del funzionamento di alcuni *software* destinati a regolare le interazioni online e tramite la distinzione tra artefatti regolativi deontici e adeontici (non-de-

[6] M. Hildebrandt, "Law as Information in the Era of Data-Driven Agency", in *The Modern Law Review*, 79 (2016), n. 1, pp. 1-30.

[7] K. Yeung, "Algorithmic Regulation: A Critical Interrogation", in *Regulation & Governance*, 12 (2018), pp. 505-523.

[8] M. Hildebrandt, *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology*, Edward Elgar Publishing, Cheltenham-Northampton, 2015, pp. 198-199, trad. nostra. Ecco il testo originale: "A prohibition of decisions based on profiling should mitigate the erosion of the presumption of innocence. […] It is imperative that the usage of AI is tested and scrutinized by countervailing powers and made contestable in a court of law."

[9] Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations", in *Science*, 366 (2019), n. 6464, pp. 447-453.

[10] High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, cit.; Cfr. L. Floridi, J. Cowls, "A Unified Framework of Five Principles for AI in Society", in *Harvard Data Science Review*, pubblicazione online ad accesso anticipato, 23 giugno 2019.

[11] M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.; V. Eubanks, *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York, 2018; R. Brownsword, *Law 3.0. Rules, Regulation, and Technology*, Routledge, Abingdon, 2020.

ontici), è possibile evidenziare criticità che non dipendono necessariamente dall'adozione di tecnologie intelligenti[12].

Il presente contributo si articola in tre parti: nella prima parte viene ricostruito sinteticamente il dibattito sull'impatto dell'intelligenza artificiale sulle libertà dell'utente; nella seconda parte si presenta la distinzione tra artefatti regolativi digitali deontici e adeontici e si analizza l'utilità di questa distinzione per valutare la compatibilità degli strumenti regolativi con le libertà fondamentali dell'utente. Infine, si analizzano due casi di strumenti digitali regolativi appartenenti alle due diverse tipologie di artefatti che, pur non utilizzando tecnologie di intelligenza artificiale, incidono in modi diversi sulla libertà di espressione.

2. Riflessione sull'IA e i suoi effetti perversi

I recenti sistemi di IA generativa, come ChatGPT, Gemini o Claude, hanno reso evidente al grande pubblico la possibilità di interagire con macchine in grado di produrre testi, traduzioni e soluzioni a problemi complessi. Questo successo ha riportato al centro della scena l'idea di un'IA "forte" o intelligenza artificiale generale (AGI), ossia un sistema non più limitato a compiti specifici, ma capace di svolgere attività cognitive tipiche dell'essere umano, con possibili ricadute profonde sulla libertà personale, sulla distribuzione del potere e sulla stessa nozione di responsabilità.

2.1. Note storiche sull'IA

Pur con tratti inediti, l'attuale stagione tecnologica si inserisce in una storia lunga oltre settant'anni. Già Alan Turing si era chiesto se le macchine potessero pensare, aprendo la strada a una riflessione che ha accompagnato lo sviluppo dell'informatica e delle scienze cognitive[13]. Negli anni successivi, conferenze

[12] G. Lorini, S. Moroni, "Rule-Free Regulation: Exploring Regulation 'without Rules' and Apart from 'Deontic Categories'", in *Journal for the Theory of Social Behaviour* 52 (2022), n. 1, pp. 22-36; G. Lorini, S. Moroni, O.G. Loddo, "Regulatory Artifacts: Prescribing, Constituting, Steering", in *International Journal for the Semiotics of Law – Revue Internationale de Sémiotique Juridique*, 36 (2022), n. 1, pp. 211-225.

[13] Alan Turing aprì la discussione con la domanda "Le macchine possono pensare?" e propose il Test di Turing come criterio operativo, cfr. A.M.

e programmi di ricerca hanno dato avvio a un campo disciplinare in cui approcci diversi hanno convissuto e si sono alternati, attraversando fasi di entusiasmo e di crisi[14].

Dopo le prime delusioni, l'IA ha conosciuto nuove fasi di rilancio: dai sistemi esperti degli anni Ottanta, al *machine learning*, fino al *deep learning* che, a partire dagli anni 2010, ha segnato un punto di svolta con risultati nettamente superiori ai metodi tradizionali. Una svolta ulteriore è giunta nel 2017 con l'introduzione dell'architettura *Transformer*, che ha reso possibile lo sviluppo dei moderni *Large Language Models*[15]. La loro diffusione al grande pubblico a partire dal 2022 ha trasformato la percezione sociale dell'IA, alimentando entusiasmi e timori legati all'impatto che questi strumenti possono avere sulle forme di comunicazione, sulla conoscenza e sulle libertà fondamentali. In questa prospettiva, la nuova IA generativa non è un fenomeno del tutto nuovo, ma il risultato di un lungo percorso che ripro-

Turing, "Computing Machinery and Intelligence", in *Mind*, 59 (1950), n. 236, pp. 433-460.

[14] La Conferenza di Dartmouth del 1956 segnò la nascita ufficiale dell'intelligenza artificiale, cfr. J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, Dartmouth College, Hanover (NH), 1956. Negli anni successivi si svilupparono approcci differenti – logico-simbolici, statistici e ispirati a modelli biologici – che conobbero alterne fortune. Seguirono infatti i cosiddetti "inverni dell'IA", dovuti a limiti teorici e tecnologici, cui fecero da contraltare le successive rinascite, dapprima con i sistemi esperti e, più recentemente, con il successo del *machine learning* e del *deep learning*; cfr. N.J. Nilsson, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press, Cambridge, 2010, e M. Mitchell, *Artificial Intelligence. A Guide for Thinking Humans*, Farrar, Straus and Giroux, New York, 2019.

[15] Un momento decisivo fu la vittoria del *software* AlexNet a ImageNet nel 2012, che segnò la definitiva affermazione dei metodi di *deep learning*. Più recentemente, l'introduzione dell'architettura *Transformer* ha reso possibile lo sviluppo dei moderni *Large Language Models*, cfr. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention Is All You Need", in *Advances in Neural Information Processing Systems*, 30 (2017), pp. 5998-6008. Per un'introduzione non tecnica al funzionamento di questi sistemi, cfr. N. Cristianini, *Machina Sapiens. L'IA e il futuro della nostra specie*, il Mulino, Bologna, 2023.

pone questioni ricorrenti: trasparenza, responsabilità e impatto sociale[16].

2.2. Il dibattito sul rapporto tra IA e libertà individuale

Il dibattito contemporaneo sul rapporto tra intelligenza artificiale e libertà individuale si è articolato lungo diverse linee interpretative.

Da un lato, numerosi autori sottolineano i rischi di compressione delle libertà derivanti dall'uso pervasivo dell'IA. In particolare, l'accento è stato posto sul ruolo delle tecnologie di sorveglianza e sul potenziale di controllo sociale attraverso sistemi di tracciamento e profilazione su larga scala[17]. In questa prospettiva, l'IA non solo limita la *privacy*, ma incide direttamente sull'autonomia e sulla possibilità di esercitare in modo effettivo la libertà di scelta. Secondo Zuboff, in effetti, si tratta di aspetti strettamente connessi, perché "[e]sercitare il proprio diritto alla *privacy* produce una scelta [...]. La privacy consente di decidere dove ci si vuole posizionare nello spettro tra segretezza e trasparenza in ogni situazione"[18]. Analogamente, Eubanks[19] ha mostrato come l'automazione del welfare possa sfociare in forme di *ghettizzazione digitale*, penalizzando soprattutto i gruppi più vulnerabili. Yeung ha descritto questo fenomeno come una "regolazione algoritmica" capace di "espandere, rafforzare e redistribuire potere, autorità e risorse, e [di] incidere sulle opportunità di vita fondamentali degli individui"[20].

[16] M. Hildebrandt, "Law as Information in the Era of Data-Driven Agency", cit.; R. Brownsword, *op. cit.*

[17] M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.; S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books, London, 2019.

[18] S. Zuboff, *op. cit.*, pp. 82-83, trad. nostra. Ecco il testo originale: "Exercising one's right to privacy produces choice [...]. Privacy rights thus confer decision rights; privacy enables a decision as to where one wants to be on the spectrum between secrecy and transparency in each situation."

[19] V. Eubanks, *op. cit.*

[20] K. Yeung, *op. cit.*, p. 519, trad. nostra. Ecco il testo originale: "expand, reinforce and redistribute power, authority and resources, and [to] affect the critical life opportunities of individuals."

Altri autori, per contro, evidenziano le *opportunità* offerte dallo sviluppo dell'IA. Floridi propone la distinzione tra *ethics by design* e *pro-ethical design*:

> Entrambe sono prospettive liberali, ma l'*ethics by design* può assumere tratti lievemente paternalistici, nella misura in cui tende a favorire il compimento del "giusto" tipo di scelte, azioni, processi o interazioni da parte degli agenti coinvolti. Il *pro-ethical design*, invece, non implica necessariamente un orientamento paternalistico, poiché privilegia piuttosto la promozione della *riflessione critica* degli agenti stessi sulle proprie scelte, azioni e processi[21].

D'altro lato, Gabriel suggerisce di valutare l'IA in termini di giustizia sociale, domandandosi se essa aumenti o riduca le capacità effettive degli individui. Come fa notare l'autore,

> algoritmi di valutazione del rischio determinano sempre più spesso l'idoneità di una persona alla libertà su cauzione o alla libertà condizionale; la tecnologia di riconoscimento facciale è stata utilizzata per potenziare le capacità della polizia; e i sistemi di intelligenza artificiale indirizzano l'impiego delle attività di polizia mediante analisi predittiva. Nel contesto della mobilità economica e dell'accesso a servizi pubblici essenziali come il welfare, l'uso di strumenti algoritmici è parimenti influente, stabilendo chi ha diritto al sostegno sociale, chi può accedere all'edilizia popolare e quali famiglie vengono prese in carico dai servizi sociali[22].

[21] L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford, 2014, p. 190, trad. nostra. Ecco il testo originale: "Both are liberal, but *ethics by design* may be mildly paternalistic, insofar as it privileges the facilitation of the *right* kind of choices, actions, process, or interactions on behalf of the agents involved. Whereas *pro-ethical design* does not have to be paternalistic, insofar as it privileges the facilitation of *reflection* by the agents involved on their choices, actions, or process."

[22] I. Gabriel, "Toward a Theory of Justice for Artificial Intelligence", in *Daedalus*, 151 (2022), n. 2, pp. 218-31, p. 220, trad. nostra. Ecco il testo originale: "risk-assessment algorithms increasingly determine a person's eligibility for bail or parole, facial recognition technology has been used to augment police capabilities, and AI systems direct the allocation of policing resources using predictive analytics. In the context of economic mobility and access to key public services such as welfare provision, the use of algorithmic tools is

In questa prospettiva, l'IA può contribuire ad ampliare l'accesso alle informazioni e ai servizi, favorendo inclusione e partecipazione, ma può anche avere effetti di cristallizzazione delle ingiustizie simili a quelli denunciati da Eubanks[23].

Un ulteriore filone di ricerca mette in luce il rischio di *concentrazione del potere economico e politico legato allo sviluppo dell'IA*. McQuillan interpreta le tecnologie di intelligenza artificiale come strumenti di "fascismo computazionale", capaci di rafforzare assetti oligarchici[24], mentre Coeckelbergh insiste sui pericoli di dipendenza e manipolazione insiti nelle "tecnologie del sé"[25].

Infine, la questione della *responsabilità e della soggettività giuridica dei sistemi di IA* è oggetto di un acceso dibattito. Birhane, van Dijk e Pasquale hanno criticato l'idea di attribuire diritti agli algoritmi, sostenendo che "[d]a un punto di vista giuridico, la migliore analogia con i diritti dei robot non sono i diritti umani, bensì i diritti delle società, un concetto altamente controverso il cui effetto più importante è stato quello di minare i diritti dei lavoratori, dei consumatori e degli elettori"[26]. Beckers e Teubner, in maniera più costruttiva, hanno proposto di ideare modelli innovativi di responsabilità collettiva per far fronte a una situazione in cui "[g]li esseri umani e gli algoritmi sembrano evolversi in un'entità simbiotica più grande della somma delle sue parti"[27].

similarly influential, determining who is eligible for welfare support, who has access to public housing, and which families are engaged by child services."

[23] V. Eubanks, *op. cit.*

[24] D. McQuillan, *Resisting AI: An Anti-fascist Approach to Artificial Intelligence*, Bristol University Press, Bristol, 2022.

[25] M. Coeckelbergh, *Self-Improvement: Technologies of the Soul in the Age of Artificial Intelligence*, Columbia University Press, New York, 2022.

[26] A. Birhane, J. van Dijk, F. Pasquale, "Debunking Robot Rights Metaphysically, Ethically, and Legally", in *arXiv* preprint, 2024, p. 1., trad. nostra. Ecco il testo originale: "From a legal perspective, the best analogy to robot rights is not human rights but corporate rights, a highly controversial concept whose most important effect has been the undermining of worker, consumer, and voter rights."

[27] A. Beckers, G. Teubner, "Human–Algorithm Hybrids as (Quasi-)Organizations? On the Accountability of Digital Collective Actors", in *Journal of Law and Society*, 50 (2023), n. 1, pp. 100-119, p. 112, trad. nostra. Ecco

In sintesi, il dibattito oscilla tra visioni che considerano l'IA prevalentemente come una minaccia alle libertà e altre che la interpretano come strumento di emancipazione. In posizione intermedia si collocano gli approcci che insistono sulla necessità di garantire criteri di *accountability*, trasparenza, contestabilità e non discriminazione, in modo da conciliare l'uso dell'IA con la tutela dei diritti fondamentali[28].

In questo contributo sosteniamo che il dibattito appena richiamato non riguarda, o non dovrebbe riguardare, soltanto l'IA – intesa come l'insieme di tecnologie che utilizzano tecniche di *deep learning* e di IA generativa – ma tocca più in generale le modalità attraverso cui gli artefatti digitali possono influenzare la libertà. Per questo, nella prossima sezione passeremo ad analizzare due strategie di regolazione digitale che operano anche senza l'uso dell'IA.

3. Due strategie di regolazione digitale che non ricorrono necessariamente all'uso dell'IA

Alcuni autori hanno sostenuto che le macchine non governeranno mai il mondo[29]. Tuttavia, le tecnologie digitali stanno già rimodellando il modo in cui sono esercitate *governance* e regolazione. La nostra analisi si concentrerà ora sui meccanismi che consentono agli artefatti digitali di svolgere la loro funzione regolativa. In questo senso introduciamo ora la categoria di Artefatto Regolativo Digitale (ARD).

In letteratura emerge una distinzione tra il modo in cui le norme giuridiche svolgono solitamente la loro funzione regolativa e una nuova tendenza, che si manifesta soprattutto nel mondo digitale, consistente nell'opporre ostacoli cognitivi e materiali che scoraggiano le violazioni o addirittura le rendono impossibili.

il testo originale: "Humans and algorithms seem to evolve into a symbiotic entity greater than the sum of its parts." Cfr. anche A. Beckers, G. Teubner. *Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds*, Hart Publishing, Oxford, 2021.

[28] L. Floridi, J. Cowls, "A Unified Framework of Five Principles for AI in Society", cit.; High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, cit.

[29] J. Landgrebe, B. Smith, *Why Machines Will Never Rule the World. Artificial Intelligence without Fear*, Routledge, London-New York, 2022.

Interessante è la distinzione operata da Lessig tra due tipi di protezione. La prima è la protezione tradizionale della legge: definisce i confini dell'azione, vieta l'intrusione e punisce i trasgressori. La seconda è la protezione tramite barriera: un dispositivo tecnologico (una linea di codice) che, tra le altre cose, impedisce accessi indesiderati[30]. Sulla stessa linea, Brownsword distingue tra libertà normativa e libertà pratica[31]. Altri studiosi hanno sottolineato come alcune forme di regolazione del comportamento basate su architetture di controllo, incidendo sulla libertà pratica, sollevino importanti questioni di legittimità e responsabilità rispetto alla regolazione tradizionale basata sulle norme[32].

Per i nostri scopi, è interessante riflettere su come i dispositivi tecnologici in generale, e gli ARD in particolare, possano contribuire a entrambe le forme di protezione identificate da Lessig e possano incidere su entrambi i tipi di libertà distinti da Brownsword.

Alcuni ARD operano imponendo norme, possiamo quindi chiamarli artefatti regolativi digitali deontici (dARD). Altri ARD operano invece condizionando i comportamenti degli individui attraverso l'architettura. Questi artefatti non impongono regole ma sfruttano tendenze psicofisiche per spingere ad agire secondo modelli predefiniti. Tali dispositivi possono essere definiti artefatti regolativi digitali adeontici (-dARD). Nella sezione seguente viene fornita un'analisi più dettagliata della distinzione tra dARD e -dARD.

3.1. Artefatti regolativi digitali *deontici*

Con artefatti regolativi digitali deontici (dARD) si intendono ARD che svolgono la loro funzione regolativa attraverso regole. Ad esempio, i dARD possono indicare ciò che è consentito o proibito all'interno delle piattaforme digitali. Gli avvisi di età su piattaforme di streaming o giochi online informano gli utenti dei requisiti anagrafici per l'accesso, lasciando però agli individui la libertà di ignorare l'avviso e proseguire. Allo stesso modo,

[30] L. Lessig, *Code: Version 2.0*, Basic Books, New York, 2006.

[31] R. Brownsword, *op. cit.*

[32] K. Yeung, "Algorithmic Regulation: A Critical Interrogation", cit.; M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.

le piattaforme di *e-commerce* possono mostrare informazioni su norme fiscali o restrizioni alle importazioni. Tali messaggi guidano i comportamenti, ma gli acquirenti restano liberi di conformarsi o meno, seppure a rischio di sanzioni. In questo modo, i dARD condividono la caratteristica essenziale delle regole descritta da Schauer: guidano il comportamento pur rimanendo violabili[33]. La loro forza regolativa non risiede nella coercizione, ma nell'offrire agli individui una scelta normativa: conformarsi o rischiare la sanzione.

Nella sfera digitale, dunque, i dARD svolgono un ruolo centrale nella comunicazione delle regole e nella trasmissione dei contenuti normativi, pur preservando un certo grado di flessibilità nell'osservanza. Non esercitano una "funzione di costrizione", cioè non impongono comportamenti specifici con la forza (come cancellare un post), né attivano automaticamente eventi specifici (come assegnare, cancellare o bloccare un profilo). Invece, segnalano l'esistenza di norme, lasciando agli utenti la libertà di decidere se conformarsi. In questa veste, incarnano una modalità di regolazione basata sulla consapevolezza e sulla scelta informata piuttosto che sulla coercizione diretta.

In generale, un dARD svolge un ruolo di comunicazione normativa. Non costringe l'utente a compiere o astenersi da un'azione, ma fornisce informazioni normative, modellando il comportamento attraverso la comunicazione. Questa funzione è stata discussa anche nella letteratura sulle architetture informative legali, che analizza come gli artefatti digitali incorporino la normatività nelle interfacce utente e medino il rapporto tra il sistema giuridico e i suoi soggetti[34].

3.2. Artefatti regolativi digitali *adeontici*

Nel quadro della sociologia del diritto digitale, l'analisi degli artefatti regolativi si è ampliata fino a includere una vasta gamma di strumenti non normativi. Nel caso degli artefatti regolativi digitali adeontici (-dARD) è il *design* stesso a imporre azioni specifiche attraverso processi automatizzati. In altre parole, i

[33] F.F. Schauer, *The Force of Law*, Harvard University Press, Cambridge (MA), 2015.

[34] M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.

-dARD non necessitano di strutture concettuali deontiche, anche se sono intenzionalmente ideati per influenzare e modificare il comportamento umano[35]. Questi artefatti condizionano il comportamento degli utenti sia online sia "*onlife*" senza fare ricorso a norme. Hildebrandt ha sottolineato come gli ARD siano sempre più affiancati da forme di regolazione che non solo plasmano, ma anche costringono i comportamenti[36].

Possiamo menzionare diversi esempi. I sistemi avanzati di assistenza alla guida, ad esempio, traducono le informazioni normative in istruzioni non per gli utenti ma per il sistema di controllo elettronico, garantendo così il rispetto automatico delle regole. Un altro esempio sono i sistemi di gestione delle identità e degli accessi, che, incorporati nel codice delle piattaforme, concedono o negano l'accesso a sezioni specifiche in base a regole predefinite. Nelle piattaforme di *home banking*, ad esempio, gli utenti possono accedere solo ai propri conti autenticandosi. Nei *social network* i sistemi di moderazione algoritmica applicano le linee guida della comunità tramite meccanismi adeontici che possono rimuovere automaticamente contenuti o chiudere account[37].

Cohen mostra come architetture e artefatti non si limitino a vincolare o abilitare comportamenti, ma contribuiscano a plasmare attivamente gli orizzonti di scelta e di pratica entro cui gli individui operano[38]. I -dARD incarnano precisamente queste forme di "regolazione intesa a influenzare i comportamenti delle persone senza ricorrere a norme, e senza alterare direttamente

[35] Si tratta di artefatti che svolgono una funzione regolativa senza ricorrere a strumenti deontici. Come sottolineano Lorini e Moroni: "Identifying (and distinguishing between) the set of normative non-regulatory interventions and the set of regulatory non-normative interventions helps to show that the regulatory function and the normative function can be performed separately." G. Lorini, S. Moroni, "Rules without Regulation and Regulation without Rules", in *Journal for the Theory of Social Behaviour*, 54 (2024), n. 2, pp. 216-234, p. 223.

[36] M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.

[37] *Ibidem.*

[38] J.E. Cohen, *Configuring the Networked Self. Law, Code, and the Play of Everyday Practice*, Yale University Press, New Haven-London, 2012.

'l'ambiente normativo'"[39]. La distinzione centrale tra artefatti deontici e adeontici risiede dunque nei loro modelli d'azione: i dARD si basano su norme esplicite, mentre i -dARD impongono comportamenti tramite il *design*. Questi vincoli possono essere fisici – in linea con il concetto di "*forcing function*" di Norman[40] – oppure psicologici, dove assomigliano a certi tipi di *nudges* identificati da Thaler e Sunstein[41].

È utile un confronto tra il concetto di -dARD e la posizione di Floridi, che distingue tra *ethics by design* e *pro-ethical design*. Nel primo approccio, gli individui possono rinunciare a una preferenza preimpostata e sceglierne un'altra. Al contrario, le strategie basate sul *pro-ethical design* richiedono che gli individui prendano una decisione esplicita su una questione prima di procedere, senza orientare la scelta verso un'opzione predefinita. In questo modo, il *pro-ethical design* garantisce una scelta imparziale e informata, rispettando l'autonomia individuale nei processi decisionali[42]. Entrambi gli approcci possono essere implementati mediante i -dARD. Nell'*ethics by design* la regolazione adeontica risulta più incisiva, poiché alla *forcing function* che impone la manifestazione di una scelta si aggiunge un

[39] G. Lorini, S. Moroni, "Rule-Free Regulation: Exploring Regulation 'without Rules' and Apart from 'Deontic Categories'", cit., p. 23, trad. nostra. Ecco il testo originale: "regulation intended to influence people's behavior without recourse to rules/norms, and without directly altering the 'normative environment'."

[40] Norman definisce una *forcing function* come un vincolo fisico "tale che un errore in una fase impedisce che avvenga la successiva". Ad esempio, una *lock-in forcing function* può impedire a un utente di chiudere un programma prima di salvare le modifiche o di inserire un indirizzo *e-mail* valido. Cfr. D.A. Norman, *The Design of Everyday Things*, Revised and Expanded, Basic Books, New York, 2013.

[41] Un *nudge* è un intervento di architettura delle scelte che, modificando il contesto decisionale senza imporre vincoli o incentivi significativi, orienta il comportamento degli individui in modo prevedibile. Secondo Thaler e Sunstein, le opzioni predefinite che spingono il cliente verso determinate scelte sono una forma di *nudge*. Esse sfruttano la tendenza psicologica delle persone a non modificare le impostazioni di default. Cfr. R.H. Thaler, C.R. Sunstein, *Nudge. Improving Decisions about Health, Wealth, and Happiness*, Yale University Press, New Haven, 2008.

[42] L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, cit.

meccanismo di *nudging* che orienta il contenuto della decisione tramite l'impostazione di un *default*. Il *pro-ethical design*, al contrario, conserva la *forcing function* procedurale ma rinuncia consapevolmente a ogni orientamento della scelta. Artefatti che sfruttano tali meccanismi sono "oggetti che non sono intrinsecamente deontici"[43]. Come osservano Lorini e Moroni, "essi permettono una regolazione senza regole e possono incoraggiare o scoraggiare determinati comportamenti senza dover ricorrere a regole e sanzioni formali"[44]. Questa forma di regolazione, che opera senza riferimento a quadri deontologici, può nondimeno avere una portata etica.

Nella prossima sezione vedremo come la distinzione tra dARD e -dARD emerga chiaramente nei casi concreti di moderazione digitale, dove strumenti diversi incidono in maniera differente sulla libertà di espressione.

3.3. Strumenti deontici e adeontici di moderazione digitale dei contenuti

Quando parliamo di moderazione dei contenuti, non stiamo solo discutendo di *policy* o di algoritmi: stiamo osservando architetture che, in modi diversi, toccano la libertà di espressione. La distinzione tra artefatti regolativi digitali deontici (dARD) e adeontici (-dARD) ci aiuta a vedere ciò che spesso resta invisibile: come una piattaforma orienta o blocca le nostre azioni prima ancora che intervengano norme, sanzioni o giudizi.

Un dARD agisce tramite regole esplicite: rende saliente una norma, una raccomandazione, uno standard di condotta. Possiamo ascoltarlo, oppure no. È violabile per definizione. Un -dARD, invece, lavora sul *design*: crea barriere, instrada i flussi, inserisce attriti o cancella opzioni. Non chiede consenso: lo im-

[43] G. Lorini, S. Moroni, O.G. Loddo, "Deontic Artifacts. Investigating the Normativity of Objects", in *Philosophical Explorations*, 24 (2021), n. 2, pp. 185-203, p. 190, trad. nostra. Ecco il testo originale: "objects that are not inherently deontic in nature."

[44] G. Lorini, S. Moroni, "Rule-Free Regulation: Exploring Regulation 'without Rules' and Apart from 'Deontic Categories'", cit., trad. nostra. Ecco il testo originale: "can encourage or discourage behaviors without the need to rely on formal rules and sanctions."

pone. In termini di libertà, i dARD incidono sulla *libertà normativa* (prescrivono cosa deve o non deve essere fatto), i -dARD incidono sulla *libertà pratica/ontica* (decidono cosa puoi fare, qui e ora).

Ad esempio, su YouTube, quando si sta per pubblicare un commento potenzialmente offensivo, compare il messaggio: "*Keep comments respectful.*" Questo suggerisce di riformulare il messaggio, magari moderandone i toni; tuttavia il messaggio può essere pubblicato comunque, anche senza variazioni[45]. Qui l'artefatto è deontico: comunica una regola di convivenza, invita alla riflessione. È una forma di moderazione che non chiude possibilità: sposta l'attenzione, sollecita l'autocontrollo, ricorda la norma.

Se prendiamo in considerazione un *software* di moderazione come Coral-Moderation Dashboard, si configura uno scenario diverso poiché, in questo caso, i commenti non diventano pubblici finché un moderatore umano non li approva. Ci troviamo davanti ad un chiaro esempio di -dARD. Anche se la regola esiste, la funzione dell'artefatto non determina la sua imposizione e comunicazione. L'artefatto si limita a bloccare la comunicazione in attesa di una decisione del moderatore. In questo modo la libertà normativa non è rilevante dal momento che l'artefatto incide direttamente sulla libertà pratica (far arrivare quel messaggio a un determinato pubblico). Si tratta di un cambio di regime materiale della comunicazione: da flusso immediato a flusso condizionato.

I due modelli richiamati nell'esempio presentano rischi e garanzie molto diversi. Il *prompt* di YouTube rischia i falsi positivi ma con un costo relativamente basso in termini di libertà pratica: l'utente può pubblicare ugualmente. Coral, al contrario, può migliorare la qualità del dibattito e ridurre lo *spam*, ma può comportare ritardi, opacità in relazione ai criteri adottati dal moderatore e alle tempistiche e possibili *bias* individuali e istitu-

[45] Il *prompt* di YouTube può essere implementato anche senza modelli di machine learning (regola statica lato client/server o semplici euristiche). La presenza o l'assenza di ML non muta la natura deontica dell'artefatto, che comunica una norma senza bloccare la pubblicazione. Questo comunque mostra la rilevanza di questa distinzione anche per l'indagine sulla funzione regolativa di strumenti digitali che impiegano ML.

zionali, conferendo un potere di censura che, se non governato, può diventare invisibile e pervasivo. Occorre però prendere in considerazione il fatto che le piattaforme non sono tutte uguali, così come non lo sono i contesti entro i quali operano gli utenti. In *thread* ad alto rischio (sensibili a pratiche di disinformazione, incitamento all'odio, o che richiedono una particolare tutela per minori), una misura -dARD può essere necessaria e proporzionata, purché progettata in maniera rigorosa. Altrove, un "filtro deontico" come chiedere una conferma esplicita prima di pubblicare un contenuto che, agli occhi dell'IA, violi la *policy* della piattaforma, potrebbe offrire un ottimo compromesso al fine di garantire la qualità del dibattito senza soffocare la libertà di espressione.

4. Conclusioni: Tutta colpa dell'IA generativa? Effetti perversi di artefatti regolativi digitali senza l'ausilio dell'IA

In questo contributo, abbiamo visto in che senso i principi discussi nel dibattito sull'etica dell'IA (proporzionalità, trasparenza, contestabilità, non discriminazione) risultino rilevanti anche per valutare strumenti digitali che non integrano componenti di intelligenza artificiale (intesa come quella tecnologia che integra elementi di IA generativa e *deep learning*). Infatti, anche interfacce semplici, filtri automatici o barriere di *design* possono limitare la libertà di espressione o l'autonomia dell'utente, senza che vi sia alcuna "intelligenza" in gioco.

L'etica dell'IA e la regolazione degli artefatti digitali si muovono in un terreno ibrido, in cui principi giuridici consolidati (art. 21 Cost., art. 10 CEDU, art. 19 ICCPR) dialogano con i principi dell'*AI ethics* (*accountability*, *fairness*, *human oversight*, *transparency*). Su questo sfondo normativo, l'analisi proposta pone l'accento sul contesto in cui viene utilizzata una determinata tecnologia. Le questioni emergenti in tale dibattito non riguardano soltanto l'IA. Anche artefatti regolativi digitali più semplici, come meccanismi di pre-moderazione o *forcing functions* di *design*, possono compromettere la libertà pratica dell'utente. In questi casi, non vi è alcun modello predittivo da spiegare: il vincolo è inscritto nell'architettura stessa. Negli

esempi analizzati abbiamo riscontrato un dato apparentemente paradossale: è possibile che la libertà di espressione risulti maggiormente garantita da un moderatore automatico che da uno umano.

Per far fronte ai possibili effetti perversi derivanti dall'uso di artefatti regolativi digitali è dunque necessario conoscere a fondo anche le strategie di regolazione impiegate oltre agli ambiti di utilizzo. Sulla base di queste conoscenze preliminari, e non di principi etici dati *a priori*, è possibile intervenire per regolare l'utilizzo delle tecnologie e mitigarne i possibili effetti lesivi delle libertà personali.

La sfida, dunque, non è limitata alla fase decisionale dell'IA, ma investe il più ampio spettro degli ARD, deontici e adeontici. È proprio a questo livello che occorre chiedersi quali garanzie accompagnino l'uso di strumenti regolativi digitali e come possano essere rese compatibili con i diritti fondamentali degli utenti.

Piero Marra

Proceduralità dell'intelligenza artificiale simbiotica

Abstract: **Procedurality of Symbiotic Artificial Intelligence.** Procedures are reinforced whenever it is not easy to know the objective reality, but only its representation. In this sense, even artificial intelligence, escaping clear conceptualisation, thrives on procedures that are systematically and logically (for human thought) obscure, thereby allowing many contents to be given form. Thus, procedures, such as algorithmic ones, are not neutral at all. Yet procedures also have their own ontology. Starting from considerations of proceduralism as a hallmark of postmodernity, the chapter analyses the procedural response of European legislation to digital issues, showing its inadequacy. The work concludes with an anthropocentric-symbiotic vision that is still procedural in nature, but in a higher sense, because human structures for understanding, classifying, and ordering reality, including information, are also procedural.

Keywords: Procedures, Symbiotic AI, Anthropocentrism, Information, European Legal Framework.

Sommario: 1. Le procedure come attributo post-moderno – 2. Un nuovo ordine delle intelligenze "artificiali" nel quadro giuridico europeo: limiti – 3. La simbiosi uomo-macchina come condizione procedurale dell'informazione – 4. Postilla o d'una visione umana-integrale.

1. Le procedure come attributo post-moderno

Le procedure non rappresentano la logica di risultati precostituiti. Tutt'altro. Riguardano, invece, il modo di procedere verso un risultato, tra i molti. È un tema caro al diritto e alla politica, ma i proceduralismi sono anche non giuridici e non politici[1]. Si pensi al proceduralismo in matematica e in matematica-computazionale. Procedurale è anche la costruzione mentale dell'intu-

[1] Si noti che in questo capitolo il termine 'proceduralismo' viene adoperato senza una specifica connotazione negativa. Se ne illustrerà, infatti, l'ambivalente natura.

izione[2] o della conoscenza, così come l'impostazione retorica di un discorso[3]. Sono tutti aspetti che richiamerò, seppure incidentalmente, in questo lavoro.

In generale, le procedure si fanno strada ogni qual volta non si può conoscere direttamente la cosa in sé, ma solo una sua rappresentazione. Questo senso del proceduralismo ha molto in comune con l'ermeneutica, perché le procedure non si fondano su un principio o una struttura del mondo oggettivamente dati. Sono per loro natura dinamiche, potendo adattarsi ai molti contenuti di una medesima realtà. In questa accezione, i proceduralismi hanno origine nella modernità[4], ma contrassegnano in particolare il post-moderno[5] e cioè il momento in cui ogni discorso metafisico (fondazionalista) sembra disperdersi nella molteplicità delle cose che accadono, nel pluralismo culturale e nel "politeismo dei valori"[6]. Certamente, non va taciuto che anche le procedure possono avere un qualche fondamento metafisico. Vedi, ad esempio, il proceduralismo medievale di impronta scolastica. La questione, tuttavia, esula dal piano discorsivo che intendo affrontare. Ciò che qui interessa, invece, è presentare il proceduralismo come attributo ambivalente della nostra epoca digitale. Le procedure hanno, infatti, una doppia faccia. Da un lato, non hanno alle spalle un'idea data di *bene*, dall'altro le procedure non sono nemmeno causa necessaria del *male*. Questa

[2] M. Dummett, *Elements of intuitions*, Oxford University Press, Oxford, p. v: "… the understanding of a mathematical statement is to be thought of a given in terms of the mental constructions which may serve to prove the statement, rather than in terms of the states of affairs within objective mathematical reality which render it true or false".

[3] F. Cavalla, "Retorica giudiziale, logica e verità", in Id. (a cura di), *Retorica processo verità. Principî di filosofia forense*, 2ª ed., FrancoAngeli, Milano, 2011.

[4] Sul tema, G. Vattimo, "Il proceduralismo (i proceduralismi) come contrassegno della modernità", in M. Basciu (a cura di), *Giustizia e procedure. Dinamiche di legittimazione tra Stato e società internazionale. XXII Congresso nazionale della Società italiana di filosofia giuridica e politica. Trieste 27-30 settembre 2000*, Quaderni della rivista internazionale di filosofia del diritto, vol. 3, Giuffré, Milano, 2002.

[5] Il riferimento è a J.F. Lyotard, *La condizione postmoderna. Rapporto sul sapere*, Feltrinelli, Milano, 1985

[6] Cito M. Weber, *Il politeismo dei valori*, Morcellania, Brescia, 2010.

doppia anima delle procedure tocca anche l'intelligenza artificiale (IA) nelle recenti versioni altamente evolute. Ne darò traccia in queste poche pagine.

Partiamo dalla scienza dell'intelligenza artificiale. Questa è, sì, moderna, ma finisce con l'essere l'emblema della condizione umana post-moderna. Ha una storia più lunga di mezzo secolo, così come risalgono agli anni '60 e '80 gli studi sull'apprendimento automatico [*machine learning*] e sulle reti neurali [*neural network*] alla base dei recenti sviluppi di apprendimento profondo [*deep learning*]. Rispetto al passato, le tecniche di apprendimento automatico, specialmente quelle di *deep learning supervisionato*, hanno sviluppato un enorme potere computazionale capace di sfruttare grandi quantità di dati "raschiati" [*data scraping*] in rete o provenienti dai sensori macchinici, persino indossabili in una società informazionale sempre più inter-connessa. E ciò ha permesso di ottenere risultati irraggiungibili, ma non per questo inaspettati. Lo aveva colto Alan Turing, che già precorreva questo cambiamento nel noto articolo *Computer Machinery and Intelligence* (1950)[7], il manifesto di un'intelligenza artificiale *ante litteram*. Aveva tra l'altro previsto gli effetti trasformativi di tali sistemi[8], che si traducono nella capacità di ristrutturare [*enveloping*][9] gli ambienti umani per consentire agli automi stessi di evolversi e trovare in sé le condizioni di auto-potenziamento[10]. Ma così cambiano anche le relazioni umane e, con esse, la produzione culturale nel reale[11]. Ed infatti, sono ridiscussi concetti e nozioni sino a quel momento appannaggio dell'uomo: l'intelligenza diventa appunto artificiale, l'inconscio

[7] A.M. Turing, "Computer Machinery and Intelligence", in *Mind*, 59 (1950), n. 236, pp. 433-460.

[8] Così l'Hight-Level Expert Group on Artificial Intelligence, nominato dalla Commissione europea per la redazione dei principî etici per una IA affidabile, nella prima versione del *Draft* pubblicato il 18 dicembre 2018.

[9] È una tesi di L. Floridi, *Pensare l'infosfera. La filosofia come design concettuale*, Raffaello Cortina Editore, Milano, 2020.

[10] E. Severino, *La tendenza fondamentale del nostro tempo*, 3ª ed., Adelphi, Milano, 2008, pp. 37-66 e U. Galimberti, *Psiche e techne. L'uomo nell'età della tecnica*, Feltrinelli, Milano, 1999, p. 399.

[11] Riprendo un tema di F. Romeo, "La produzione culturale dell'artificiale: prodromi di uno studio metaepistemologico su diritto ed informazione", in *L'Ircocervo*, 23 (2024), n. 1, pp. 51-69.

acquisisce una propria percepibilità digitale, alla morte fisica fa da contraltare la continuità della vita *onlife*[12]. Solo per richiamare alcuni noti fenomeni. In sostanza, molte rappresentazioni umane della realtà vengono discusse con concetti inediti. A partire dalla stessa concettualizzazione dell'intelligenza artificiale.

Non sono, infatti, poche le definizioni di intelligenza artificiale. Se chiedessimo a cento ricercatori di IA di definire l'IA, probabilmente otterremmo almeno 100 definizioni diverse. E ciò stando ad Oliviero Stock, noto ricercatore in materia. Ma basterà menzionare il dibattito europeo sul tema. Talmente ampio, che lo stesso Regolamento europeo 2024/1689, meglio noto come *AI Act*, nel richiamare la formula adoperata dall'Organizzazione per la cooperazione e lo sviluppo economico, non descrive l'intelligenza artificiale *in sé e per sé*, quanto le capacità di tali sistemi di mostrare comportamenti intelligenti analizzando i dati forniti dall'ambiente e agendo, con livelli di autonomia variabili, per realizzare certi obiettivi (espliciti o impliciti)[13]. È una definizione che risente di altrettanti approcci alla materia[14]. Non abbiamo, in definitiva, un concetto unitario di intelligenza artificiale, né tantomeno un *nome* capace di definire queste entità esotiche[15]. Lo aveva capito sempre Alan Turing, che ri-

[12] È una espressione nota di L. Floridi, *The Onlife Manifesto. Being Human in a Hyperconnected*, Springer, Cham, 2015.

[13] V. art. 3 del Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale).

[14] S.J. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, 2ª ed., Prentice Hall, Upper Saddle River, 2010, p. 2. Si possono individuare quattro approcci suddivisi tra chi identifica l'intelligenza dell'automa con il pensiero (rappresentazione della conoscenza e ragionamento), e quanti la considerano solo come percezione e interazione con l'ambiente; tra la posizione di chi ritiene che l'intelligenza debba riprodurre fedelmente le capacità dell'uomo (con tutti i limiti), e quella opposta di coloro che vogliono realizzare sistemi solamente razionali (sfruttando la correlazione delle informazioni per realizzare scelte ottimali).

[15] Richiamo il titolo di un recente saggio di S. Amato, "Macchinazioni. Soggettività, agentività ed entità esotiche", in *L'Ircocervo*, 23 (2024), n. 1, pp. 289-306.

levava l'insensatezza dell'interrogativo: "Le macchine sono in grado di pensare?". Nel famoso gioco (o giogo, a seconda delle prospettive) dell'imitazione, intelligenti sono solo le macchine capaci di apparire intelligenti, senza *esserlo*. Qui la distinzione tra realtà (ciò che è oggettivamente dato) e reale (la percezione della realtà da parte del soggetto) si fa strutturalmente marcata[16]. Sembra consacrarsi ogni visione procedurale che si appropria della nota lezione di Nietzsche secondo cui "Non ci sono fatti, solo interpretazioni". E forse anche l'intelligenza artificiale è un'interpretazione della realtà!

Non è, dunque, definibile l'intelligenza artificiale *in sé*. È il nostro punto di partenza. Ma l'altra questione è che non abbiamo nemmeno chiaro come si comporti l'intelligenza umana, il che rende imperfetto o interpretabile lo stesso concetto di "imitazione" artificiale. Ed anche quello di "emulazione" a cui si riferivano, al Dartmouth College (1956), i padri dell'intelligenza artificiale[17]. In questo modo, non sappiamo sino in fondo se le IA possano eguagliare anche le difettività umane o addirittura spingersi oltre, sommando, a quelle umane, le proprie, in una realtà destinata a moltiplicare il campo delle possibilità di scelta e azione. In definitiva, non abbiamo una intelligenza artificiale, ma molte. Non una logica, ma più di una[18]. Non una realtà, ma tante rappresentazioni reali di essa. Uno scenario tutto post-moderno, incontrollabile e lontano da ogni contenuto metanarrativo.

In questo scenario, le cose non possono che complicarsi. Pensavamo di poter acquisire facili certezze tramite la computazione e di "generare" risposte neutrali a qualsiasi problema (è il mantra sacrale dell'IA generativa che "genera"… senza ragionare), ma ci ritroviamo immersi in una realtà computazionale iper-complessa (e non nel senso dell'altamente complesso), altrettanto incerta, e che dobbiamo ancora ordinare.

[16] La differenza tra realtà e reale è in F. Romeo, *op. cit*.

[17] Si tratta della proposta di J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31/09/1955.

[18] Già non erano poche quelle individuate dal pensiero contemporaneo: logiche del razionale e logiche del ragionevole, logica proposizionale classica e logiche modali e deontiche, logiche rigorose e logiche *fuzzy*, procedure dialettiche e dialogiche. A queste, si sommano quelle artificiali.

2. Un nuovo ordine delle intelligenze "artificiali" nel quadro giuridico europeo: limiti

Ordinare le intelligenze artificiali, senza avere un'idea chiara di "cosa" si intenda ordinare, ci ha spinto nevroticamente verso una regolamentazione minima degli usi dell'intelligenza artificiale. Il presupposto è questo: poiché non so che cosa sia l'intelligenza artificiale, la regolo sul piano degli effetti che riesco, invece, a percepire. E però i problemi non diminuiscono, perché questa volta le incertezze toccano anche l'operatività macchinica e non solo il concetto di intelligenza artificiale. Anzi, con l'IA, il concetto è l'agire stesso[19].

In questa direzione, il legislatore europeo, con l'*AI ACT* insieme al *Digital Service Act* e al *Digital Market Act* ha inteso fornire un contributo all'opera di istituire un nuovo ordine delle intelligenze[20]. Ma con esiti non sempre felici. La frammentarietà ed eterogeneità degli applicativi di IA ha spinto verso una regolazione *risk-based* in cui, anziché disciplinare la singola tecnologia, si è preferito guardare agli usi in base a prestabiliti settori di applicazione e fattori di rischio in sé ambigui. E tuttavia, le tecniche di IA sfruttano diverse procedure di calcolo difficili da comprendere con il linguaggio delle norme e che rendono il rischio incalcolabile.

L'*AI Act*, dunque, non disciplina le singole tecnologie (predeterminandone il contenuto), bensì il modo in cui devono essere progettate, prodotte e distribuite. È il paradigma dell'approccio neutro dal punto di vista tecnologico, che affida agli stessi sviluppatori le modalità di valutazione dei propri ritrovati. In base all'*Ai Act*, infatti, la valutazione dei rischi associati avviene prima dell'uso di tali sistemi e prosegue anche durante il loro ciclo di vita. I sistemi di IA classificati come ad "alto rischio" (che copre la maggior parte degli applicativi) sono, ad esempio, sot-

[19] Sul concetto di IA come forma di *agency*, di recente, L. Floridi, *La differenza fondamentale. Artificial Agency: una nuova filosofia dell'intelligenza artificiale*, Mondadori, Milano, 2025.

[20] U. Ruffolo, A. Amidei, *Diritto dell'intelligenza artificiale*, Luiss University Press, Roma, 2024, pp. 13 ss. e si veda anche la prospettiva di A. Punzi, "'Accolse l'uomo come opera di natura indefinita'. Note su esperienza giuridica e nuovo ordine delle intelligenze", in *Rivista di filosofia del diritto*, 14 (2025), n. 1, pp. 21-32.

toposti ad una valutazione dell'impatto sui diritti fondamentali, devono usare dati di qualità e garantire adeguate misure di sicurezza. Il problema è che per mitigare i rischi occorre identificarli. E ciò non è semplice soprattutto se l'applicativo di IA si raffronta sul terreno dei diritti. Nella pratica, infatti, possono dirsi rischiosi dal punto di vista dei diritti solo quando si confrontano con i tanti fatti che accadono. Dunque, una valutazione *ex ante* dei rischi che incidano sui diritti fondamentali è insufficiente e, probabilmente, insensata[21]. E poi, un sistema di IA (e cioè una combinazione a volte incontrollata di sistemi di sistemi) può produrre al contempo sia rischi elevati, sia rischi inaccettabili. Il confine tra ciò che è "ad alto rischio" o "a rischio inaccettabile" non è però così netto. Ma soprattutto, con quali mezzi si può valutare il rischio? Non potrà negarsi un ruolo rilevante della stessa tecnica, ma qui c'è un paradosso: se la valutazione del rischio è filtrata dall'uso di tecniche di IA, è essa stessa il risultato di un'attività rischiosa. Sono d'altronde oscuri i processi di ideazione, progettazione e realizzazione di tutti questi sistemi. La sensazione è che pochi produttori che dispongono di un certo sapere tecnologico si limitino ad autocertificare l'affidabilità delle proprie invenzioni (si tratta delle certificazioni *self-risk assesment*) minando la correttezza del processo valutativo e la legittimità delle proprie "risoluzioni".

Non solo è procedurale l'approccio del legislatore che indica il modo di procedere nella modellizzazione di applicativi di IA, ma sono procedurali anche le dinamiche interne alla generazione degli *output* dei sistemi di IA. Qui i limiti del proceduralismo si fanno ancora più marcati. I sistemi artificiali esperti usano l'inferenza per passare da certe premesse alle conclusioni. Entrambi questi schemi sono formalizzabili ed automatizzabili mediante strutture sintattiche che richiamano i simboli della matematica, organizzano i dati con la programmazione logica e, infine, stabiliscono correlazioni[22]. Queste *rappresentazioni*

[21] Sul tema, P. Marra, "Rischi algoritmici. Alcune questioni sulla digitalizzazione della giustizia", in *Journal of Ethics and Legal Technologies*, 7 (2025), n. 1, pp. 135-158.

[22] La proposta di usare i connettivi della logica formale per esprimere i rapporti tra proposizioni nei testi normativi e nei contratti è già in L.E. Allen, "Toward a Normalized Language to Clarify the Structure of Legal Discour-

della conoscenza sono combinate con le istruzioni (gli algoritmi) per compiere diversi tipi di inferenze (deduttive, presuntive, induttive, probabilistiche, argomentative, basate su casi, ecc.). Oltre i sistemi esperti, le macchine sono anche in grado di darsi procedure da sé, ma restituendoci risultati sistematicamente opachi (*black box effect*)[23]. Sono state così realizzate numerose applicazioni che generano esse stesse il modello di conoscenza su cui si fondano le proprie prestazioni inferenziali, senza trascurare gli esiti dell'apprendimento profondo che sfrutta addirittura reti neurali per imitare la complessità dell'organizzazione del cervello umano. E però, le maggiori difficoltà a comprendere pienamente i processi macchinici hanno natura epistemica. Sono troppo distanti, infatti, il numerico e il linguaggio umano. La rivoluzione digitale è, infatti, una rivoluzione grafica che assegna al numero e alla rispettiva combinazione (binaria o quantistica) un ruolo primario: la scrittura digitale codifica con i numeri, fatti ed eventi del mondo fisico[24]. Ma i numeri non sono il linguaggio naturale dell'uomo, né tantomeno hanno una struttura deontica: non leggono cioè la realtà come rapporto tra *essere* e *dover essere*, ma solo in termini di causalità tra i dati. E non costituiscono *status* deontici i dati stessi che sfruttano la correlazione dei numeri tramite segni algebrici secondo rapporti di quantità e non di qualità[25].

Le differenze epistemiche non sono affatto da trascurare. Eppure, sembra che i nomoteti dimentichino l'insufficienza del nostro linguaggio nella complessa attività di regolazione

se", in A.A. Martino (ed.), *Deontic Logic, Computational Linguistics, and Legal Informations Systems*, North-Holland Publishing Company (Elsevier), North Holland, 1982, pp. 349-407.

[23] Sul tema dell'opacità immanente, ma anche sistemica, delle nuove intelligenze, di recente l'analisi di S. Amato, "Inquietudini digitali. Il *back box effetc*", in *Rivista di filosofia del diritto*, 14 (2025), n. 1, pp. 33-44.

[24] Lo evidenziano con efficacia A. Garapon, J. Lassègue, *La giustizia digitale. Determinismo tecnologico e libertà*, il Mulino, Bologna, 2021, pp. 37-48.

[25] V. Zeno-Zencovich, "'Big data' ed epistemologia giuridica", in. S. Faro, T.E. Frosini, G. Peruginelli (a cura di), *Dati e algoritmi. Diritto e diritti nella società digitale*, il Mulino, Bologna, 2020, pp. 13-24.

del digitale[26]. In questo senso, le procedure di *risk governance*, comprensive della valutazione, del calcolo dei rischi (*risk assesment*) e delle misure per arginare i rischi identificati (*risk management*), sembrano più l'esito risolutivo di chi produce quegli artefatti. Ne consegue che ogni forma di autentica partecipazione umana al processo valutativo viene eclissata e, con essa, ogni tentativo di comprensione applicativa, costringendo l'ambiente umano a subirne solo i condizionamenti. Con il carattere risolutivo della *self-certification*, l'uomo, diventato utente passivo di informazioni, appare così un osservatore disinteressato[27]. Se, dunque, le procedure rappresentano una reazione tutta post-moderna alla perdita di certezze, non sono pochi i rischi che l'"io-procedurale" corre nell'affidarsi alle sole procedure[28]. O meglio. Il pericolo è quello di ritenere che l'unica forma di razionalità vera sia connessa al funzionamento delle procedure stesse. Infatti, la flessibilità delle procedure funzionale al raggiungimento di un risultato, da un lato rischia di essere condizionata altrettanto facilmente da fattori esterni (mercato, scientismo tecnologico, politica…), dall'altro di rimanere per se stessa indifferente al destinatario (nel contesto digitale divenuto "utente"). Anche le procedure, come i numeri, non sono affatto neutre.

In quest'ottica, l'io-procedurale, assecondando le sole procedure senza badare alla comprensione dei contenuti, *è ridotto* ad osservatore-spettatore del nuovo ordine prodotto, privandolo di ogni discorso sul senso e sulla verità. L'io (procedurale) ne esce, dunque, ancor più frammentato; estraneo a qualsiasi processo autenticamente decisionale in cui può prendervi parte, destinatario di *risoluzioni* (le risoluzioni sono l'opposto delle *decisioni*) che tiene-per-vere, ma senza una verità alle spalle. Lo si vede

[26] Al tema dedica alcune riflessioni A. Rossetti, "Dare regole alle tecnologie", in S. Ricci, A. Rossetti (a cura di), *Aspetti giuridici della società dell'informazione*/Legal Issue in the Information Society, Giuffrè Francis Lefebvre, Milano, 2025, pp. 1-14

[27] Sulla differenza tra osservatore e spettatore G. Siniscalchi, *Barocco giuridico. Osservatori, osservanti, spettatori*, FrancoAngeli, Milano, 2017

[28] Il termine "io-procedurale" è nell'analisi critica al proceduralismo di B. Romano, "Ragione giuridica e ragione procedurale", in M. Basciu (a cura di), *op. cit.*, pp. 37-62.

bene proprio guardando il modello del linguaggio numerico che sostituisce i *rinvii di senso* con i *rinvii di funzionamento*. Il nuovo modello procedurale-artificiale può tradursi così: "funziona, dunque è vero".

La vera questione, però, *è* che anche ciò che funziona non può fare a meno delle domande sul senso. Si rischia, infatti, di lasciare che le procedure (e con esse il soggetto) si muovano alla cieca, ponendo sullo sfondo l'ontologia dell'esistenza, per focalizzarsi unicamente sul come arrivare ad un risultato attraverso procedure che appaiono formali, fredde… artificiose. Prive, in apparenza, di contenuti. Lo scarto è molto evidente. Da un lato c'è l'essere con il suo portato di valori, dall'altro un modo (un metodo) di approcciarsi all'esistenza e all'essere. Eppure, anche le procedure mantengono un rapporto con l'essenza delle cose e con l'uomo stesso… anche le procedure hanno una loro ontologia. È il tema conclusivo del prossimo paragrafo che mira a rinvenire una via, tutta informazionale, nella simbiosi uomo-macchina.

3. La simbiosi uomo-macchina come condizione procedurale dell'informazione

Il proceduralismo macchinico ha le sue ombre. Lo si è appena detto. Alla perdita di controllo e di comprensibilità dei processi algoritmici, fa da sfondo lo smarrimento dell'uomo post-moderno a cui risuonano gli echi distopici di un'imminente (possibile) sostituzione. La reazione dell'uomo post-moderno è così spinta verso una visione antropocentrica dello sviluppo tecnologico, ma questa volta senza che gli siano indicati i mezzi (anche culturali) per realizzare tale valore. Ci limitiamo a regolare le IA a livello degli effetti, ma non ci occupiamo di come ordinarle per realizzare inedite forme di propositività in un ambiente culturale che oramai condividiamo con agenti artificiali.

Forse anche l'IA è un prodotto culturale, più che un insieme di componenti *hardware* e *software*. O meglio, l'insieme delle componenti *hardware* e *software* rappresenta per noi la realtà di una forma di intelligenza "artificiale". E come prodotto culturale, l'IA interviene sulla nostra percezione della realtà. Soltanto che dobbiamo elaborare nuovi strumenti concettuali che

ci permettano di percepire questa realtà in modi più soddisfacenti, oltre la nostra biologia e i presupposti genetici[29]. A partire dall'opera di concettualizzazione della realtà digitale. I concetti, infatti, ci aiutano ad agire meglio nel mondo e ad avere una rappresentazione della realtà [*Wirklichkeit*] diversa dalla realtà oggettiva [*Realität*][30]. In quest'opera di rappresentazione della realtà in un sistema cognitivo, si può certamente perdere qualche informazione (ad esempio, non si è in grado di catturare tutti gli esisti algoritmici sottostanti, a causa del divario epistemico sopra detto, ma anche delle differenti velocità di elaborazione dei dati), ma ciò non vuol dire affatto che l'uomo sia assolutamente incapace di interagire con le macchine e con le IA. Se infatti consideriamo l'IA come una rappresentazione culturale della realtà, dobbiamo altrettanto ammettere che il divario tra realtà oggettiva e rappresentazione della realtà non può che comporsi nel soggetto senziente che elabora l'informazione non passivamente, ma secondo i propri sensi, ordinandola in base alle sue possibilità mentali e necessità biologiche. Questa elaborazione attiva del dato, che diventa informazione, determina la volontà e l'azione del soggetto. La procedura, questa volta con la 'P' maiuscola, sta ad ordinare l'informazione, ma finalizzandola all'ambiente umano.

Se dunque l'IA è un prodotto culturale, allora abbiamo bisogno di un'informatica e di un'informatica-giuridica capace di teorizzare e progettare sistemi decisionali integrati orientati alla rappresentazione umana della realtà, seppure con il supporto di strumenti tecnici. Sistemi che sappiano porre le giuste domande a partire dalla situazione della vita da conoscere e regolare[31]. L'antropocentrismo richiede, in sostanza, di trovare il modo migliore per rappresentare la nuova realtà digitale presente nell'ambiente umano. Mi pare che si muova in questa direzione la visione dell'informatico e psicologo statunitense Joseph Carl

[29] F. Romeo, *op. cit.*, p. 57.

[30] *Ivi*, p. 58.

[31] La tesi era precorsa da Luigi Lombardi Vallauri. Per un riepilogo del suo pensiero sui temi dell'informatica giuridica, L. Lombardi Vallauri, "Prefazione. Un'informatica giusliberista, democratica, filosofica", in G. Peruginelli, M. Ragona (a cura di), *L'informatica giuridica in Italia. Cinquant'anni di studi, ricerche ed esperienze*, Esi, Napoli, 2014, pp. 13-24.

Robnett Licklider il quale, nel saggio *Man-Computer Symbiosis* (1960)[32], ci introduceva al tema della simbiosi tra l'uomo e il computer. Secondo Licklider, l'evoluzione simbiotica dell'intelligenza artificiale avrebbe operato in modo differente dalle tecnologie semi-automatiche che dipendevano strettamente dalla direzione, integrazione e pre-determinazione dell'uomo. Ma con l'evidente perdita di controllo di ogni procedura algoritmica. Invece, secondo tale inedita visione, l'interazione uomo-computer (oggi, diremmo in senso più generale, uomo-macchina) non dipendeva dalle *risposte* che un sistema artificiale poteva fornire, bensì dalle *domande*[33]. Veniva così tracciata la possibilità che entità simbiotiche potessero generare *decisioni* su questioni complesse attraverso la "problematizzazione" di *casi* e il combinato uso dell'euristica umana e di processi computazionali[34].

Il termine 'simbiosi' è usato in senso *informazionale*[35] ed ha anch'esso una sua connotazione culturale, perché con la simbiosi il processo conoscitivo della realtà resta umano. E però, nella visione di Licklider, la cooperazione simbiotica uomo-macchina continua ad essere risolta mediante l'uso di interfacce (e dunque, ancora, di algoritmi) per accorciare la separazione tra il linguaggio naturale dell'uomo e quello simbolico dell'elaboratore elettronico. Ne erano propedeutici: la condivisione in tempo reale delle informazioni; grandi capacità di memorizzazione ed organizzazione dei dati; interazione "paritetica" in cui l'uomo fissa gli obiettivi e il calcolatore traccia il percorso. L'idea è suggestiva, ma allo stesso tempo semplicistica, perché permangono gli stessi rischi epistemologici che conducono ad una procedura il cui esito finale assume la forma di una *risoluzione* esterna, più che di una *decisione* partecipata. Al centro della sfida tecnologica torna, pertanto, la decisione validamente accettata in un sistema sociale umano. Così, le procedure guadagnano l'essere e le sue relazioni, sia da un punto di vista empirico che logico.

Dal punto di vista empirico, l'analisi informatica orientata alla decisione come scienza non solo logico-formale, ma anche

[32] J.C.R. Licklider, "Man-computer symbiosis", in *IRE Transactions on Human Factors in Electronics*, (1960), n. 1, pp. 4-11.

[33] *Ivi*, p. 5.

[34] *Ivi*, p. 6.

[35] F. Romeo, *op. cit.*, p. 59.

politico-valutativa, non può fare a meno delle esperienze evidenziali che, tuttavia, non si analizzano né si simulano[36]. Le IA, non a caso, hanno difficoltà nelle attività di *pattern recognition* e cioè nel riconoscere l'oggetto e apprendere comportamenti dall'esperienza. L'algoritmo non può così fare a meno della capacità umana di leggere e riconoscere le evidenze.

Dal punto di vista logico, le procedure macchiniche non riproducono pienamente l'esperienza umana nella sua interezza. Almeno ad oggi. Dovremmo altrimenti presupporre la riduzione ontologica dei livelli mentali e biologici a termini fisici e meccanici. Ma ciò non sarebbe comunque sufficiente, perché esistono segmenti di scelte sottratti alla formalizzazione il cui risultato è *indecidibile* dal punto di vista del calcolo. La questione investe l'incompletezza di questi sistemi[37]. Già i teoremi di Gödel avevano dimostrato l'impossibilità dei sistemi assiomatici di dedurre tutte le formule o proposizioni vere che vi appartengono[38]. La potenza di calcolo di una macchina universale, dunque, non implica necessariamente la capacità di fondare tutte le proprietà contenute nei propri assiomi. Ecco perché la fondazione di una decisione non può che avvenire in una certa "forma di vita" [*Lebensform*], cioè in base a giudizi di convalida vincolati

[36] L. Lombardi Vallauri, "Informatica e criteri «politici» o valutativi della decisione giuridica", in *Jus*, (1981), n. 3, pp. 303-315.

[37] A metà degli anni '30, Alan Turing stava lavorando al problema della decisione in logica [*Entscheidungsproblem*], posto nel 1928 dal pensiero finitario di David Hilbert i cui esiti vennero codificati dai noti teoremi di incompletezza di Kurt Gödel. I testi fondamentali di riferimento sul dibattito sono A.M. Turing, "On Computable Numbers with an Application to the Entscheidungsproblem", in *Proceedings of the London Mathematica Society*, 42 (1937), n. 1, pp. 230-265; D. Hilbert, "Über das Unendliche", in *Mathematische Annalen*, 95 (1926), n. 1, pp. 161-190 e Id. "Die logischen Grundlagen der Mathematik", in *Mathematische Annalen*, 88 (1923), nn. 1-2, pp. 151-157; K. Gödel, "Über formal unentscheidbare Sätze del Principia Mathematica und verwandter Systeme I", in *Monatshefte für Mathematik*, 38 (1931), n. 1, pp. 173-198.

[38] Sulla rilevanza dei teoremi per una fondazione dei sistemi giuridici, si veda A. Incampo, *Metafisica del processo. Idee per una critica della ragione giuridica*, 2ª ed., Cacucci, Bari, 2016, pp. 221-223 e prima Id., *Sul fondamento della validità deontica. Identità non-contraddizione*, Edizioni di Giuseppe Laterza, Bari, 1996, pp. 330-351.

alle strutture conoscitive e, in genere, culturali, che le rendono possibili[39].

Resta da stabilire il metodo o, quantomeno, indicarne una prospettiva. Soccorre l'idea di una *Symbiotic AI* o meglio un'intelligenza artificiale simbiotica come livello antropologico dell'evoluzione macchinica. La simbiosi, infatti, presuppone la fallibilità umana, ma non è neppure il superamento della finitezza ontologica dell'essere. È invece un modo di tendere alla massima propositività umana. D'altronde, anche grazie alle intelligenze artificiali, abbiamo la consapevolezza che l'essere umano possiede processi cognitivi limitati, mentre la macchina mantiene capacità computazionali mirabili; eppure, è proprio questa "svolta cognitiva" a doverci spingere verso un antropocentrismo simbiotico con le macchine, in un ambiente necessariamente umano in cui sono assicurati canali di comunicazione stabili ed effettivi tra uomo e artefatto tecnologico[40]. Tuttavia, non sono poche le incognite legate principalmente alle incapacità dell'uomo di spiegare i risultanti delle infinite connessioni ad alta frequenza tra dati. E non basta il principio di trasparenza che rischia di rappresentare un falso mito. La trasparenza, senza l'immediata intelligibilità e comprensibilità dei processi decisionali digitali, rischia di essere in partenza cieca. Così come non è sufficiente ridurre il controllo umano supervisionato [*supervisory human control*] alle sole interfacce dell'interazione uomo-macchina che sacrificano la "consapevolezza situazionale"

[39] L. Wittgenstein, *Ricerche filosofiche*, Einaudi, Torino, 1995, part. 25. Sul rapporto tra Turing e Wittgenstein v. J. Floyd, "Chains of Life: Turing, *Lebensform*, and the Emergence of Wittgenstein's Later Style", in *Nordic Wittgenstein Review*, 5 (2016), n. 2, pp. 7-89 ed ancora Id., "*Lebensformen*: Living Logic", in C. Martin (ed.), *Language, Form(s) of Life and Logic. Investigations after Wittgenstein*, De Gruyter, Berlin, 2018, pp. 59-92. Spunti interessanti si rinvengono anche in G. De Anna, "Automi, responsabilità e diritto", in *Rivista di filosofia del diritto*, 8 (2019), n. 1, pp. 125-142.

[40] Di questa idea è in parte anche L. Lombardi Vallauri, "Informatica e criteri 'politici' o valutativi della decisione giuridica", cit., p. 308 il quale già nel 1982 si chiedeva se fosse possibile "fare un automa conversazionale che ponga al giudice la domanda giusta […] traendone poi velocemente le conseguenze giuste".

da parte dell'uomo[41]. Sono, infatti, le stesse interfacce a filtrare, condizionare, manipolare, predeterminare la volontà del soggetto (utente) inconsapevole. In un contesto associativo, invece, la cooperazione uomo-macchina richiede la partecipazione attiva degli agenti (artificiali e umani). Con questo approccio, il controllo umano non è basilare quanto il principio associativo[42]. Ma, per essere effettiva, la cooperazione non può che fondarsi sulle capacità comunicative non riducibili alla semplice trasmissione di contenuti, trasmissione del tutto inutile se tali contenuti non vengono processati ed effettivamente compresi dai destinatari. Ad evitare lo scivolamento in una consulenza informatica "oracolare", con una formazione di coaguli di potere tecnico, può ad esempio provvedere un sistema integrato consulente-documentario che permetta all'utente di formarsi un'opinione autonoma[43]. Tale evoluzione dell'IA a misura d'uomo, seppure possa apparire limitante dal punto di vista dello scientismo che assolutizza le proprie verità come oggettive, ha i suoi vantaggi dal punto di vista normativo, perché regola i comportamenti umani in interazione con entità artificiali secondo strumenti logici e di ragionamento a lui congeniali. E torna ancora utile, a tal riguardo, l'esperienza giuridica nell'opera di costruzione e ricostruzione di oggetti sociali che, tuttavia, esige un'opera attenta di giustificazione delle scelte tramite la partecipazione dialettica del soggetto. Con il diritto, infatti, occorre giustificare anche le ipotesi scientifiche. La lezione dell'insufficienza della legge scientifica nella prova dei fatti è, peraltro, particolarmente nota all'esperienza giudiziale. Il modello è proprio quello dei processi decisionali non delegati.

È dunque molto difficile pensare a un programma che consenta all'automa di assegnare un certo significato alle parole o ai fatti, ma sarà invece possibile ideare un sistema che sappia porre

[41] A. Tsamados, L. Floridi, M. Taddeo, "Human control of AI systems: from supervision to teaming", in *AI and Ethics*, (2024), pp. 1-14, part. pp. 7-8.

[42] Sulle dinamiche normative dell'associazione uomo-macchina A. Beckers, G. Teubner, *Three Liability Regimes for Artificial Intelligence. Algorithmic Actants, Hybrids, Crowds*, Hart Publishing, Oxford, 2021.

[43] L. Lombardi Vallauri, "Prefazione. Un'informatica giusliberista, democratica, filosofica", cit., p. 17.

le giuste domande o far sorgere dubbi legittimi. Qui le procedure servono a ordinare la conoscenza e i suoi contenuti per avere decisioni legittime e comprensibili in un ambiente sociale, combinando l'euristica umana con processi computazionali, e affidando un ruolo decisivo ai molteplici contenuti informazionali da spiegare. D'altra parte, i dati non spiegano altri dati[44].

4. Postilla o d'una visione umana-integrale

Il 19 novembre 2025, la Commissione Europea ha approvato alcuni emendamenti e modifiche tecniche (c.d. regolamento *omnibus*) della legislazione digitale per aiutare le imprese, pubbliche amministrazioni e cittadini. L'idea è quella di ottimizzare l'applicazione del codice digitale, favorendo la competitività delle imprese. La proposta di regolamento *omnibus* in un certo qual modo depotenzia l'iniziale entusiasmo sulle capacità europee di guidare il mondo verso una regolazione effettiva dell'intelligenza artificiale. Qui le procedure dimostrano tutti i condizionamenti e le variabilità esterne a cui si è fatto cenno.

Dall'altro lato, non sembra che al centro di tali politiche vi sia una riflessione sul carattere promozionale del normativo nella costruzione di inedite interazioni "uomo-macchina-uomo". In questo capitolo, si è brevemente presentata la *Symbiotic AI*, dandone una lettura procedurale come inedito spazio (associativo) di conoscenza della realtà a misura d'uomo. La regolazione della realtà digitale non può trascurare, infatti, il ruolo del soggetto come portatore di senso e significato, indispensabili per comprendere il valore culturale dell'intelligenza artificiale. E prima ancora di ogni regola, serve un'opera *omnia* di concettualizzazione, per distinguere cosa compete all'uomo e cosa alle realtà artificiali. Solo così, i *big data* in cui siamo avvolti potranno servire a informare (e cioè dare forma al mondo), ma anche a procedere (cioè ordinare) per trasformare la realtà secondo una propria, ed inedita, opera classificatoria. Ma informare vuol dire anche aprirsi all'interazione con il mondo e con l'artificiale. Le procedure, in questo modo, mantengono una loro ontologia, per-

[44] G. D'Acquisto, *Decisioni algoritmiche. Equità, causalità, trasparenza*, Giappichelli, Torino, 2022, pp. 12, 23.

ché sono vincolate alle strutture conoscitive e, in genere, culturali, che le rendono possibili. Le procedure ordinano le informazioni e le relazioni che nascono e sono al contempo condizioni significativamente *deontiche* di ogni attività umana[45]. In questo senso, però, le procedure, con la 'P' maiuscola, non sono fine a se stesse, ma si riferiscono all'esistenza umana e ai propri fondamenti. Ben oltre la scienza e la sua fisica. Solo in questi termini, se non è desiderabile il monopolio dell'IA, forse è auspicabile un monopolio degli "umani-troppo-umani" insediati dalle intelligenze artificiali[46].

[45] È una idea di Lon L. Fuller, *The Morality of Law*, 2ª ed., Yale University Press, New Haven, 1969 che ho ripreso in P. Marra, *Per una moralità procedurale del diritto. Considerazioni attuali a partire da Lon L. Fuller*, Cacucci, Bari, 2022.

[46] Riadatto, ai sistemi di intelligenza artificiale, un'idea suggestiva di L. Lombardi Vallauri, "Verso un sistema esperto giuridico integrale", in C. Ciampi, F. Socci Natali, G. Taddei Elimi (a cura di), *Verso un sistema esperto giuridico integrale. Esempi scelti dal diritto dell'ambiente e della salute*. *Atti del Convegno celebrativo del venticinquennale dell'Istituto (Firenze, 1-3 dicembre 1993*), tomo I, Cedam, Padova, 1995, pp. 157-182.

Paolo Moro

Dialettica processuale e intelligenza artificiale
Il problema del metodo

Abstract: **Procedural Dialectics and Artificial Intelligence. The Issue of Method.** This essay explores the methodological challenges raised by the integration of artificial intelligence into criminal procedure. It critically contrasts the analytical logic underlying AI systems, rooted in deduction and induction, with the dialectical logic that characterizes judicial reasoning, grounded on dialogue and on adversarial system. By retracing the philosophical and legal tradition from classical dialectics to the guarantees of fair trial, the author argues that algorithmic reasoning cannot replace the argumentative and dialogical nature of judicial decision-making. The text finally advances the model of a 'hybrid justice', where legal professionals and AI tools interact, provided that human control, transparency, and the centrality of dialectical reasoning remain preserved.

Keywords: Artificial Intelligence, Criminal Procedure, Dialectics, Fair Trial, Hybrid Justice.

Sommario: 1. Il valore giuridico del giusto processo – 2. La struttura dialettica del giusto processo – 3. Intelligenza artificiale e processo. L'approccio analitico – 4. Intelligenza artificiale e processo. L'approccio dialettico – 5. Conclusioni. Verso la giustizia ibrida.

1. Il valore giuridico del giusto processo

Il superamento critico di ogni concezione giuridica normocentrica permette di considerare come meglio difendibile nella riflessione teoretica e nell'esperienza pratica una prospettiva processuale. Secondo questa visione, non del tutto condivisa negli studi giusfilosofici dell'Europa continentale, il momento specifico ed irrinunciabile del diritto è la controversia giudiziale, che è ritenuto il fenomeno ontologicamente (anche se non cronologicamente) originario rispetto alla regola legale la quale, appunto, soltanto nel processo trova la propria applicazione e interpretazione[1].

[1] P. Moro, "La prospettiva processuale del diritto, oggi", in S. Fuselli, P. Moro (a cura di), *Al Principio. Invito alla Filosofia del Diritto*, FrancoAnge-

In particolare, con questo orientamento si confronta chi abbia una visione unitaria ed assiologica del diritto e del processo penale, indispensabile per evitare di disgiungere tali aspetti e per cogliere l'essenza dell'intero ordinamento giuridico, qualificato come tale quando assuma il carattere della ragionevolezza[2]. Il principio personalista, accolto dalla Costituzione vigente, impone il giusto processo penale come custode sia della fragilità della persona offesa che della dignità dell'imputato e misura il grado di civiltà di una comunità che sa organizzare e risolvere la relazione conflittuale, diventando così un valore[3].

Infatti, dove si accolga il principio di offensività come cardine strutturale del diritto criminale contemporaneo, il reato non appare soltanto la violazione formale della legge, ma diventa un comportamento dannoso che vìola l'integrità di un bene ritenuto degno di tutela perché appartenente a colui il quale, manifestando quell'opposizione che con l'atto lesivo il reo vuole sopprimere, pretende che tale limitazione sia tolta e che, dunque, l'offesa sia riparata[4].

In questa direzione, che pone l'accento sulla pena come riparazione dell'offesa, si riscopre la centralità del giusto processo, che non è l'accertamento sillogistico della conformità della norma al fatto, ma il luogo dialogico di discussione e confronto delle parti al fine di raggiungere un giudizio equo, nel rispetto dei diritti dell'offeso e dell'imputato.

Non vi è dubbio che, coniugando diritto e processo penale sulla base dell'empirismo utilitarista di *common law*, la giurisprudenza anglosassone abbia elaborato già in età medievale (con la *Magna Charta libertatum* del 1215) il rispetto delle garanzie di libertà processuale e abbia attribuito una specifica dignità alla posizione difensiva dell'incolpato. Tuttavia, la dot-

li, Milano, 2022, pp. 391-412.

[2] S. Riondato, *Un diritto penale detto «ragionevole». Raccontando Giuseppe Bettiol*, Cedam, Padova, 2006.

[3] G. Bettiol, *Istituzioni di diritto e procedura penale*, Cedam, Padova, 1966.

[4] F. Cavalla, "Pena e riparazione. Oltre la concezione liberale dello stato: per una teoria radicale della pena", in F. Cavalla, F. Todescan (a cura di), *Cultura moderna e interpretazione classica. Pena e riparazione*, vol. 2, Cedam, Padova, 2000, p. 11 e s.

trina processuale continentale ha approfondito l'indispensabile esigenza di approfondimento teorico del processo come valore espressivo della razionalità del giudizio, al fine di "garantire il nostro mondo culturale legato alla logica 'aristotelica'"[5].

L'odierno modello accusatorio, che può dirsi in varia misura accolto nel diritto comune europeo, presuppone un processo di parti di natura agonistica[6] e si presenta come insieme di atti ai quali sono chiamati a partecipare necessariamente di fronte ad un terzo giudicante coloro che sostengono posizioni opposte tra loro al fine di pervenire alla composizione della lite con l'esecuzione di un provvedimento conclusivo che tenga conto delle pretese e delle contestazioni sollevate dai disputanti[7].

Questa nozione consente di individuare nel processo penale quattro principi che costituiscono le ineliminabili attività della giustizia e che concorrono a definirne i fondamenti concettuali, manifestando il paradigma giuridico del cosiddetto "giusto processo" o anche "equo processo" (nell'espressione anglosassone: *due process of law*), la cui validità appare oggi ampiamente riconosciuta nel diritto comune e che costituisce la base del garantismo: tali principi sono la contestazione; il contraddittorio; la prova; la giurisdizione.

Come è noto, secondo la dottrina del diritto penale minimo, il processo è funzionale alla tutela dei diritti fondamentali dell'offeso e dell'imputato quando in esso vengono rispettati quattro assiomi: *nulla poena sine iudicio, nullum iudicium sine accusatione*, *nulla accusatio sine probatione* e *nulla probatio sine defensione*[8].

Tuttavia, tali assiomi non sono protocolli apodittici prefissati dal diritto positivo, ma debbono ritenersi degni di approvazione solo se logicamente fondati e non convenzionalmente presuppo-

[5] G. Bettiol, "L'odierno problema del bene giuridico", in *Riv. it. dir. proc. pen.*, (1959), p. 707.

[6] G. Spangher, "Verso un processo di parti", in *Studium Iuris*, (2002), n. 4, p. 469.

[7] E. Fazzalari, voce "Procedimento (teoria generale)", in *Enciclopedia del Diritto*, XXXV, Giuffrè, Milano, 1986, p. 827.

[8] L. Ferrajoli, *Diritto e ragione. Teoria del garantismo penale*, Laterza, Roma-Bari, 1996, p. 69.

sti, in quanto resistenti alla critica, sicché possono essere condensati nelle seguenti definizioni.

La *contestazione* coincide anche con il diritto costituzionale di azione tutelato dall'art. 24 della Carta repubblicana ed è altresì un onere indispensabile (*nemo iudex sine actore*) che si realizza con la formulazione dell'imputazione (art. 405 c.p.p.) e che qualifica il sistema accusatorio, fondato sul principio di corrispondenza tra chiesto e pronunciato e, dunque, sulla correlazione tra accusa e sentenza, che inibisce al giudice di accertare *ex officio* una presupposta "verità" per trasformarlo in un inquisitore[9].

Il *contraddittorio* costituisce il principio orientativo essenziale del processo penale accusatorio perché consiste nella possibilità per ogni interessato di partecipare all'attività giudiziale e di influire sull'esito finale della stessa, in condizioni di completa e reale parità delle armi, come stabilito non soltanto dall'art. 24, ma anche dall'articolo 111 della Costituzione vigente che menziona espressamente il giusto processo nella sua parte iniziale, tutelando il diritto di difesa e di informazione delle parti[10].

La *prova* è la procedura di controllo logico della contestazione fatta valere nella lite, ma è anche il mezzo (come la testimonianza o il documento) che le parti hanno l'onere di introdurre nel giudizio per giustificare la propria pretesa chiedendone esplicitamente l'ammissione (art. 190 c.p.p.), tenendo conto che l'imputato si presume non colpevole fino alla condanna definitiva (art. 27 Cost.) in applicazione della nota regola di giudizio che impone la dimostrazione della responsabilità penale oltre ogni ragionevole dubbio (art. 533 c.p.p.)[11].

La *giurisdizione* è l'attività impersonale svolta da un terzo per risolvere una controversia attraverso un giudizio in una "equa e pubblica udienza davanti ad un tribunale indipendente e imparziale, al fine della determinazione dei suoi diritti e dei suoi doveri nonché della fondatezza di ogni accusa penale" (art. 10

[9] F. Cavalla, *Giudici e Inquisitori*, in Aa.Vv., *Giudici e Inquisitori*, a cura del CIDAS, Liberilibri, Macerata, 1999.

[10] R.E. Kostoris (a cura di), *Il giusto processo, tra contraddittorio e diritto al silenzio*, Giappichelli, Torino, 2002.

[11] P.P. Paulesu, *La presunzione di non colpevolezza dell'imputato*, Giappichelli, Torino, 2008.

della *Dichiarazione Universale dei Diritti Umani* e art. 6 della *Convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali*): secondo il criterio della giustizia impersonale[12], il giudice non può sottrarsi all'amministrazione della controversia e alla conseguente necessità di decidere, essendo questa la ragione della sua stessa esistenza e la sua specifica funzione, a lui demandata dall'ordinamento giuridico[13].

In una prospettiva filosofica, non pare banale ricordare che questa griglia tetralogica costituisce non solo la struttura legale del giusto processo, ma costituisce anche il metodo di ragionamento più adeguato a rappresentarne la logica: infatti, diversamente dal procedimento analitico della scienza deduttiva, il giudizio penale realmente equo non si svolge secondo una procedura sequenziale o dimostrativa (la conclusione è la derivazione dalla premessa presupposta), ma si sviluppa in forma circolare o argomentativa (la conclusione è la premessa argomentata e, dunque, difficilmente attaccabile)[14].

Infatti, è bene osservare che gli enunciati elementi costitutivi del giusto processo si esibiscono nell'azione giudiziale come aspetti distinti di una medesima realtà e possono essere compresi soltanto in una visione unitaria (*globalità*); sono costantemente presenti in ogni fase della discussione processuale (*contestualità*); contrassegnano qualunque attività giurisdizionale in ciascun ramo del diritto vigente e, dunque, manifestano ciò che accomuna ogni esperienza giuridica particolare (*comunanza*).

Pertanto, l'accertamento giudiziale non è il risultato discendente di una prova "scientifica", perché non costituisce né un'operazione matematica né una descrizione[15], ma è l'esito orizzontale di una logica "filosofica" che mira a rendere preferibile una tesi rispetto a quella opposta, rendendola insignificante perché

[12] F. Chiereghin, *Dall'antropologia all'etica. All'origine della domanda sull'uomo*, Guerini e Associati, Milano, 1997, p. 114.

[13] F. D'Agostino, *Contingenza delle norme e soluzione delle controversie*, in *Filosofia del diritto*, Giappichelli, Torino, 2001, p. 123.

[14] P. Moro, *La via della giustizia. Il fondamento dialettico del processo. Con l'«Apologia di Socrate» di Platone*, Libreria Al Segno Editrice, Pordenone, 2004.

[15] S. Fuselli, *Apparenze. Accertamento giudiziale e prova scientifica*, FrancoAngeli, Milano, 2008.

contraddittoria, cioè incapace di resistere alle confutazioni prospettate dalle parti nelle rispettive difese.

2. La struttura dialettica del giusto processo

Nella cultura giuridica contemporanea, l'origine del giusto processo viene frequentemente, ma inesattamente, riferita all'origine moderna dei diritti dell'uomo che si manifesterebbe nei princìpi sostenuti dalla Scuola del Diritto Naturale. Secondo questa tesi, accolta ancora diffusamente dalla dottrina giuridica, l'elaborazione dei diritti fondamentali processuali risalirebbe alla costruzione teorica dello Stato moderno, nella sua matrice liberale di fine Seicento (riferibile essenzialmente al *Secondo Trattato sul Governo* di John Locke) e nella sua fonte democratica di fine Settecento (attinta fondamentalmente dal *Contratto sociale* di Jean Jacques Rousseau).

Al contrario, la ricerca del metodo di composizione della controversia non è nuova né moderna nella cultura dell'Occidente ma nasce nell'antichità classica attraverso lo sviluppo e l'elaborazione concettuale di una procedura logica che, strettamente connessa alla discussione della controversia giuridica, viene denominata "arte dialettica" (in greco *dialektiké téchne*).

Il rapporto tra processo e dialettica trova così la propria naturale genesi nella cultura greca nella quale, sin dalle origini della civiltà occidentale e ben prima dell'universalismo medievale, il processo costituisce il culmine della vita giuridica e manifesta nell'esperienza quotidiana il profondo e radicale legame tra la giustizia e il conflitto.

In questa mentalità, l'ordinarsi del processo si fonda sulla dialettica che, in quanto arte dell'antilogia e nello stesso tempo della mediazione, consente di svelare la verità nella controversia e attua la cooperazione involontaria tra i partecipanti alla discussione giudiziale. Infatti, come si noterà non solo in epoca greca e romana, ma anche nell'età di mezzo, il legame tra processo e dialettica non è precostituito dalla ragione come un sistema e non è affatto il prodotto spontaneo di una volontà particolare

(dei contendenti o del giudice), ma si realizza automaticamente nel conflitto tra le parti, cioè nel contraddittorio[16].

In effetti, la connessione tra processo e dialettica non è occasionale, ma è necessaria: il paradigma dialettico non è un modello strumentalmente utilizzabile per risolvere le dispute giurisdizionali, ma è fondato su un principio originario che custodisce nell'esperienza il differenziarsi di opposte pretese, impone la discussione delle controversie nella forma processuale e attribuisce la competenza a risolvere la lite al terzo, il cui giudizio deve essere innegabilmente impersonale.

Infatti, ciascuno degli elementi costitutivi del processo penale, che nella cultura giuridica contemporanea vengono tradizionalmente accolti come principi inderogabili, manifesta i medesimi caratteri della dialettica classica e affonda le proprie radici teoriche nonché la primigenia applicazione pratica nel pensiero e nella vita dei Greci.

La dialettica classica costituisce lo statuto epistemologico della filosofia e del processo: invero, come osserva Aristotele, "si trova necessariamente in una condizione migliore per giudicare colui che ha ascoltato le ragioni opposte, come in un processo"[17].

Come la filosofia, la dialettica si fonda sul domandare; si qualifica come procedura dialogica; è peirastica, cioè saggiatoria o esaminativa e, dunque, ha carattere probante; si serve di principi comuni ai discorsi opposti che si espongono nella disputa, assumendo così un carattere impersonale e, quindi, neutrale[18].

Come il processo, la dialettica si manifesta nel continuo ed inesausto domandare (la contestazione); nel dialogare che attivamente si mostra nelle dispute prive di violenza quali sono le controversie (il contraddittorio); nel saggiare il fondamento di tutte le opposte affermazioni che vengono dedotte nella discussione (la prova); nell'individuare principi che, in quanto inconfutabili, sono comuni a tutti i partecipanti al dibattito (la giurisdizione).

[16] A. Giuliani, voce "Prova (filosofia del diritto)", in *Enciclopedia del Diritto*, XXXVII, Giuffrè, Milano, 1988, p. 523.

[17] Aristotele, *Metafisica*, III 1, 995 b 2-4, II. Testo greco con traduzione a fronte, a cura di G. Reale, Vita e Pensiero, Milano, 1993, p. 85.

[18] E. Berti, *La dialettica come statuto epistemologico della filosofia*, in Id., *Le vie della ragione*, il Mulino, Bologna, 1987, p. 141.

Il fondamento dialettico del giudizio presuppone una visione unitaria dell'esperienza giuridica e consente di individuare il punto focale del diritto penale contemporaneo.

Anzitutto, solo il processo dialettico può concretizzare la funzione più evoluta ed estesa della pena che, ripensando criticamente e sintetizzando le tradizionali finalità moderne della retribuzione, della prevenzione generale e della prevenzione speciale[19], consiste nella riparazione dialogica dell'offesa[20].

La necessità di riservare una maggiore attenzione alla lesione che il reato ha concretamente posto in essere, e quindi anche alle persone direttamente coinvolte, è stata avvertita dall'orientamento della *Restorative Justice*, che costituisce oggigiorno il più avanzato tentativo di trasformare la sanzione afflittiva applicata nel processo in una "pena dialogica"[21].

In secondo luogo, l'azione della giurisprudenza europea nella tutela dei diritti processuali, con particolare riferimento al settore penale, ha diffuso il giusto processo ben oltre la sovranità degli Stati e costituisce forse l'unica materia realmente condivisa nel diritto continentale, essendo il potere giurisdizionale tendenzialmente unitario.

Attraverso l'influsso della giurisprudenza si incontrano le tradizionali esperienze nazionali di *civil law* con le esperienze internazionali di *common law*, che sono ormai attestate non soltanto nel diritto sostanziale[22] ma anche e soprattutto nel settore processuale, tenuto conto dell'ormai sempre più frequente applicazione nella casistica giudiziale della CEDU e della Carta di Nizza, positivizzata nel Trattato di Lisbona.

Infine, il processo penale come metodo di tutela dialettica della persona offesa e della persona accusata riporta l'attenzione

[19] F. Zanuso, S. Fuselli (a cura di), *Ripensare la pena. Teorie e problemi nella riflessione moderna*, Cedam, Padova, 2004.

[20] F. Cavalla, *Pena e riparazione. Oltre la concezione liberale dello stato: per una teoria radicale della pena*, in F. Cavalla, F. Todescan (a cura di), *op. cit.*, p. 11 e s.

[21] F. Reggio, *Giustizia dialogica. Luci e ombre della Restorative Justice*, FrancoAngeli, Milano 2010.

[22] R. Borsari, *Diritto punitivo sovranazionale come sistema*, Cedam, Padova, 2007.

sul valore relazionale del soggetto nell'ordinamento giuridico e nell'azione giudiziaria.

La tutela giudiziale multilivello dei diritti fondamentali, tra i quali assumono precipua rilevanza le garanzie del giusto processo, si svolge da tempo nel serrato confronto tra le Corti nazionali, la Corte europea dei diritti dell'uomo di Strasburgo e la Corte di giustizia di Lussemburgo attraverso dibattuti sentieri di riflessione[23].

Ma è evidente che i problemi interpretativi e applicativi dei diritti umani nella giurisprudenza non sono affrontabili senza ripensarne criticamente il fondamento filosofico, che implica l'analisi critica della concezione antropologica dell'individuo moderno e la riscoperta teoretica della natura dell'uomo come soggetto aperto alla relazione[24].

3. Intelligenza artificiale e processo. L'approccio analitico

L'innovazione digitale nel giusto processo è indubbiamente inevitabile e forse necessaria, ma rischia di essere ingenua, qualora non tenga conto di un'essenziale e problematica dissonanza metodologica, del tutto evidente nella teoria e nella tecnica dell'argomentazione forense[25].

Lo sviluppo decisivo dell'informatica, dalla quale dipende l'evoluzione contemporanea delle tecnologie dell'informazione e della comunicazione, e il progresso dei sistemi esperti di intelligenza artificiale si manifestano ancor oggi nell'ambito privilegiato delle scienze matematiche e discendono dalla teoria dell'informazione di Shannon[26] e dalla definizione della cibernetica di Wiener[27].

[23] M. Cartabia (a cura di), *I diritti in azione. Universalità e pluralismo dei diritti fondamentali nelle Corti europee*, il Mulino, Bologna, 2007.

[24] P. Moro, *I diritti indisponibili. Presupposti moderni e fondamento classico nella legislazione e nella giurisprudenza*, Giappichelli, Torino 2004, pp. 1-292.

[25] P. Moro, *L'informatica forense. Verità e metodo*, San Paolo, Cinisello Balsamo, 2006.

[26] C.E. Shannon, "A Mathematical Theory of Communication", in *Bell System Technical Journal*, 27 (1948), luglio-ottobre, pp. 379-423 e 623-656.

[27] N. Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge (Mass.), 1948.

Il pensiero computazionale ha contribuito anche a fondare i programmi di intelligenza artificiale sul metodo scientifico[28] e ha portato a ricondurre a schemi comuni di origine algoritmica le macchine e i sistemi biologici capaci di esibire un comportamento finalizzato, confermando il presupposto matematico della cibernetica[29].

Come è noto, l'informatica è disciplina basata sul calcolo e rivela la sua matrice analitica (dal greco *analýo*, scompongo o divido) quando si mostra specificamente come logica della scomposizione o della divisione. In particolare, il ragionamento analitico applicato all'interpretazione ed alla decisione giuridica presuppone la distinzione tra una premessa data e un risultato logicamente coerente (nel caso della deduzione) oppure empiricamente verificabile (nel caso dell'induzione), secondo lo schema tipico del metodo scientifico inaugurato negli *Analitici Posteriori* da Aristotele[30].

Con la repentina crescita esponenziale dei dati, verificatasi negli ultimi anni e resa disponibile in modo sempre più diffuso dalla rete Internet, il metodo deduttivo è stato sorpassato dall'approccio induttivo, che per facilitare l'analisi delle informazioni e ridurre la loro mole raggruppa gli oggetti in classi, il più possibile omogenee, predeterminandone il numero e le caratteristiche. In tal modo, la rappresentazione informatizzata della conoscenza si realizza più efficacemente con l'apprendimento automatico, favorito dagli studi di statistica e delle neuroscienze, perché basato su molteplici esempi (denominati *big data*) tratti dal grande archivio di Internet e sull'elaborazione di sistemi di intelligenza artificiale (denominati reti neurali) simili al cervello umano.

L'emergere di questo approccio statistico, basato sui dati elaborati con le tecnologie di *machine learning*, ha portato gli studi scientifici e le applicazioni tecnologiche dell'intelligenza artificiale ad un autentico *paradigm shift*, nel senso indicato da Thomas Kuhn[31]. A causa della centralità dell'analisi automatica

[28] D. Gillies, *Artificial Intelligence and Scientific Method*, Oxford University Press, Oxford, 1996.

[29] G. Tamburrini, *I matematici e le macchine intelligenti*, Mondadori, Milano, 2002.

[30] E. Berti, "Le mie passeggiate con Aristotele", in *Peitho: Examina Antiqua*, 1 (2016), n. 7, pp. 55-67.

[30] T. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1962.

dei *big data*, l'apprendimento statistico e l'ottimizzazione stocastica hanno sostituito la logica deduttiva, trasformando l'inferenza intelligente nel modello inverso, risolto dalla massimizzazione induttiva di una certa quantità probabilistica[32].

Intelligenza artificiale e rete Internet sono così diventate le più importanti tecnologie convergenti dell'epoca contemporanea, aumentando il loro predominio in via esponenziale, perché interagiscono tra loro e si implicano reciprocamente, pur sempre in dipendenza di algoritmi di apprendimento automatico. La soggezione all'algoritmo in qualunque applicazione informatica e cibernetica rimane tale sia che si tratti di algoritmo supervisionato, ossia quando il sistema (per esempio, la rete neurale) apprende il nesso che unisce dati particolari e ipotesi generali, imparando a classificare esempi simili, sia che si tratti di algoritmo per rinforzo, ossia quando il sistema interagisce con l'ambiente ed esegue gli esempi di comportamenti più vicini e simili al risultato, che quindi migliorano l'azione da eseguire, evitando quelli più distanti e diversi, che sono considerati errati.

Questo mutamento di paradigma coinvolge anche l'informatica giuridica, che è una metodologia utilizzata soprattutto dai giuristi pratici e che presenta sia nel caso della dimostrazione deduttiva (prevalentemente utilizzata negli algoritmi di programmazione informatica) che in quello dell'inferenza induttiva (prevalentemente utilizzata negli algoritmi di apprendimento automatico) almeno due tratti tipici della logica analitica: il carattere ipotetico delle premesse e la natura monologica del ragionamento.

La logica formale e sistematica, basata sul giudizio analitico (deduttivo o induttivo), è la base metodologica dei sistemi legali esperti di intelligenza artificiale, ma appare insufficiente per comprendere e risolvere i problemi dell'esperienza giuridica, che giudici e avvocati affrontano ogni giorno nel processo.

Il processo è una pratica sociale che richiede l'adeguata applicazione di un'efficace e compiuta metodologia di composizione della controversia, articolata nella disputa di opposte po-

[32] N. Cristianini, "On the current paradigm in artificial intelligence", in *AI Communications*, 27 (2014), pp. 37-43.

sizioni destinate a risolversi in un giudizio rappresentato in un testo scritto, come la sentenza.

La tecnica informatizzata di ricerca documentaria, la formazione previsionale e l'interpretazione dei dati giuridici, l'automazione della decisione giudiziale non sono operazioni misurabili e ripetibili né possono esprimersi in uno schema ipotetico e monotonico, soprattutto quando tale schema è applicato ad un ambiente agonistico e processuale dominato dall'esercizio pratico del contraddittorio[33].

L'etica e il diritto debbono risolvere conflitti che si manifestano naturalmente nella realtà sociale, che è controversa, attraverso scelte che implicano complesse analisi riconducibili solo parzialmente all'attività computazionale, esigendo spesso profonda riflessione e capacità decisionale in situazioni incerte e ambigue. La possibilità di prefigurare la ripetizione costante e seriale di atti autenticamente umani confligge con la natura della soggettività, imprevedibile ed irriducibile soltanto ad un meccanismo biologico involontario ed automatico, pur adattabile alle reazioni ambientali di una macchina cibernetica[34].

Dunque, l'essenza della metodologia di risoluzione del caso giuridico che, per sua stessa natura, è imprevedibile e spesso difficilmente omologabile ai precedenti, appare un fattore di indomabile resistenza ai tentativi sempre più elaborati dei sistemi giuridici esperti, fondati sull'intelligenza artificiale, di prevedere con metodo predittivo e statistico i comportamenti dei giudici.

Infatti, la soggezione a ipotesi precostituite di programmazione non è compatibile con la realtà giuridica, che implica la necessità di mettere in discussione le premesse del discorso sul fatto controverso, sebbene la nozione stessa di algoritmo non implichi soltanto la formulazione astratta di una teoria, ma anche un processo effettivo di descrizione della natura[35].

La natura ipotetica delle premesse, che caratterizza ogni procedura di calcolo, permane indubbiamente anche negli algoritmi

[33] P. Moro, *Topica digitale e ricerca del diritto. Metodologia e informatica giuridica nell'era dell'infosourcing*, Giappichelli, Torino, 2015.

[34] J. Weizenbaum, *Computer Power and Human Reason. From Judgment to Calculation*, Freeman, San Francisco, 1976.

[35] P. Zellini, *La matematica degli dèi e gli algoritmi degli uomini*, Adelphi, Milano, 2016.

di *machine learning*, sia che si tratti di apprendimento supervisionato (quando l'utente fornisce esempi di ciò che si deve apprendere) sia che si tratti di algoritmo con rinforzo (quando la procedura è influenzata dalle decisioni prese dal sistema stesso).

Dunque, le procedure formali di apprendimento automatico restano inferenze ipotetiche e non possono interpretare o risolvere problemi complessi della vita reale, come i conflitti di valore, che manifestano costante variabilità ed incertezza e che sono difficili da codificare attraverso misurazioni o funzioni numeriche[36].

Nel campo particolare dell'esperienza giudiziaria, è stato variamente sperimentato l'utilizzo analitico dell'apprendimento automatico per consentire al sistema di prevedere (anche se non di comprendere) il probabile orientamento del giudice in questioni ricorrenti e determinabili in senso alternativo (per esempio, sussistenza o meno della violazione della norma).

In effetti, ogni indagine predittiva della giurisprudenza tramite intelligenza artificiale continua a fondarsi sul modello analitico e ipotetico, che appare però inadatto a risolvere il caso giuridico, ove le premesse sono per loro natura soggette a contestazione e le connessioni logiche tra premesse e conclusioni dipendono dal contraddittorio dialogico che caratterizza la discussione organizzata della controversia[37].

Inoltre, un approccio esclusivamente basato sull'inferenza induttiva e statistica di generalizzazione dei dati per effettuare previsioni fondate sul calcolo delle probabilità, come accade negli odierni sistemi esperti, è limitato e rimane necessario offrire una spiegazione delle ragioni per le quali il sistema di intelligenza artificiale fornisce una specifica previsione[38].

Il criterio induttivo è utilitarista, come sovente accade in ogni visione empirista, perché mira all'esigenza puramente operativa dell'efficacia dell'analisi automatica dei dati, prescindendo da modelli esplicativi di spiegazione delle relazioni causali di ciò

[36] Z. Lipton, "The mythos of model interpretability", in *ICML Workshop on Human Interpretability*, (2016), arXiv:1511.03677.

[37] P. Moro (a cura di), *Il diritto come processo. Princìpi, regole e brocardi per la formazione critica del giurista*, FrancoAngeli, Milano, 2012.

[38] A. Testolin, M. Piccolini, S. Suweis, "Deep Learning Systems as Complex Networks", in *arXiv*, (2018), arXiv:1809.10941, pp. 1-20.

che descrive, limitandosi a trattare un consistente numero di informazioni relative al fenomeno a fini predittivi[39].

Nell'esperienza processuale, l'affidamento di una decisione automatica a *software* predittivi (come, per esempio, COMPAS in *Loomis case*, Supreme Court of Wisconsin, 2016) è stato giudicato criticamente dalla dottrina, sia perché la giustizia penale vive di regole di giudizio, quali la proporzionalità e l'equità, difficilmente riconducibili a schemi algoritmici[40], sia perché l'ordinamento anche internazionale riconosce oggi al singolo soggetto il nuovo diritto fondamentale a decisioni penali non basate unicamente su algoritmi, in base al principio di tutela della dignità umana[41].

4. Intelligenza artificiale e processo. L'approccio dialettico

L'evoluzione cibernetica dell'informatica giuridica ha portato oggigiorno alla costituzione di sistemi esperti di intelligenza artificiale che intendono rimediare alle fissità della logica analitica e che pretendono, ma finora senza successo, di imitare la metodologia del discorso giudiziario, dotandosi delle caratteristiche essenziali che riguardano l'attività della giurisprudenza.

Fino a quando resta ancorata al modello analitico, l'intelligenza artificiale non può superare alcuni aspetti critici della prassi giuridica, tenendo conto che la logica giudiziale è anipotetica e non stipulativa, perché dipende da premesse sottoposte al controllo critico del contraddittorio: la disputa legale implica un "discorso" quantomeno bidirezionale o conflittuale e produce un risultato sfumato e imprevedibile, che deriva dal confronto delle variabili di ogni singolo caso dibattuto e dalla condivisione di ciò che appare comune.

L'elaborazione di una logica telematica per l'argomentazione forense, che consenta di costruire adeguati sistemi esperti di intelligenza artificiale per il diritto, dovrebbe imitare il ragionamento giuridico, che non è fondato su regole, ma su caratteristi-

[39] R. Kitchin, "Big Data, new epistemologies and paradigm shifts", in *Big Data & Society*, (2014), n. 1, pp. 1-12.

[40] S. Signorato, "Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo", in *Riv. dir. proc.*, 2 (2020), pp. 605-616.

[41] A. Simoncini, "L'algoritmo incostituzionale", in *BioLaw Journal*, 1 (2019), pp. 63-89.

che che sono al di fuori del modello analitico e sono riconducibili alla dialettica[42].

Sin dalle sue origini classiche, la procedura dialettica si presenta come arte dell'antilogia e, nello stesso tempo, come tecnica della mediazione, proponendosi di discutere qualunque premessa del discorso per distinguere ciò che è contraddittorio e custodire ciò che è condivisibile nelle proposizioni alternative[43]: va da sé che questa logica discontinua e dialogica, che è tipica del diritto, può essere difficilmente formalizzata nella sequenza dell'algoritmo.

Poiché la conoscenza giuridica è controversa e le conclusioni del confronto fra tesi contrapposte sono sempre revocabili in dubbio, ogni giudizio può essere sconfitto da argomenti prevalenti in senso contrario e, dunque, la soluzione dei problemi giuridici richiede processi interattivi governati da regole dialettiche[44].

Si aggiunga che la comunicazione giuridica si basa sul linguaggio naturale e, quindi, non utilizza un lessico nel quale ad ogni termine è assegnato un significato rigorosamente univoco. Dunque, la formalizzazione informatica deve confrontarsi con la vaghezza, essendo impossibile eliminare in tutti i discorsi e, a maggior ragione, nell'interpretazione giuridica, la polivalenza di senso e di significati che le espressioni linguistiche intrinsecamente recano[45].

Per converso, la procedura dialettica presenta alcuni connotati che sono distanti da quelli dell'informatica giuridica analitica e che, più appropriatamente, si avvicinano all'uso della rete telematica, che presuppone la trasmissione, la comunicazione e l'interazione di atti linguistici tra diversi soggetti. Anzi, il tentativo di formalizzare la comunicazione giuridica utilizzando le

[42] F. Cavalla, "Il controllo razionale tra logica, dialettica e retorica", in M. Basciu (a cura di), *Diritto penale, controllo di razionalità e garanzie del cittadino*, Cedam, Padova, 1998.

[43] F. Cavalla, *L'origine e il diritto*, FrancoAngeli, Milano, 2017.

[44] G. Sartor, "Una nuova logica giuridica per l'argomentazione telematica?", in Aa.Vv., *Scrittura e diritto*, Giuffrè, Milano, 2000, pp. 177-210.

[45] C. Luzzati, *La vaghezza delle norme: un'analisi del linguaggio giuridico*, Giuffrè, Milano, 1990.

strutture argomentative del processo è stato giustamente fondato anche sulla dialettica e sul ragionamento intersoggettivo[46].

Infatti, lo sviluppo dell'intelligenza artificiale ha suggerito un differente approccio computazionale per l'elaborazione di decisioni automatiche basate sulla mediazione tra dati alternativi e sfumati. In tale prospettiva è stata elaborata la logica *fuzzy*[47] che, a differenza dei tradizionali programmi cibernetici fondati sulle regole della conversione binaria, presuppone risultati sfocati e che si sovrappongono tra loro[48].

Comunque, come dimostrano applicazioni tecnologiche anche sofisticate dell'età contemporanea, la logica sfumata è una teoria di classi con contorni indistinti e continua ad obbedire al pregiudizio ipotetico ed assiomatico del metodo analitico: difatti, la logica sfocata dipende pur sempre da una funzione di appartenenza all'insieme *fuzzy* e si realizza nei sistemi informatici attraverso la costruzione di un algoritmo programmato su un calcolatore digitale convenzionale.

L'utilità metodologica della dialettica appare evidente nel fine comunicativo che possiede il rinvenimento e la selezione, pur trattata con strumenti informatici e telematici o affidata a sistemi esperti di intelligenza artificiale, dei dati giuridici con la topica digitale. In particolare, gli argomenti possono essere considerati efficienti per la discussione della controversia quando la loro persuasività derivi non soltanto dalla difficoltà di opporre ad esse giustificate contestazioni, ma anche dalla loro attitudine ad adeguarsi al singolo caso disputato, mostrandosi ragionevoli piuttosto che esatti[49].

Questa qualità della metodologia forense svela la *performance* retorica e l'ingegno tecnico del giurista pratico che rintraccia i luoghi argomentativi più difficili da criticare nello specifico contesto del caso controverso, consentendo di acquisire anche

[46] D.N. Walton, E.C.W. Krabbe, *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*, State University of New York Press, Albany, 1995.

[47] L.A. Zadeh, "Fuzzy Sets", in *Information and Control*, 8 (1965), pp. 338-353.

[48] L.E. Stanley, S. Hsu-Shih, *Fuzzy and Multi-Level Decision Making. An Interactive Computational Approach*, Springer, London, 2011.

[49] P. Moro, *Topica digitale e ricerca del diritto*, cit.

una scrittura espositiva più credibile, perché più vicina alla questione dibattuta e allo stile legale, che al momento un *chat bot* difficilmente potrebbe ricreare.

Tuttavia, le trasformazioni del processo penale nell'era digitale hanno condotto alla smaterializzazione degli atti e all'uso crescente dell'intelligenza artificiale, favorendo il modello empirista ed utilitarista, orientato all'efficienza ed alla prevedibilità del giudizio. Oltre al rischio di ridurre il rito processuale ad un'automazione ripetitiva e disumanizzante, è stato osservato che, in tale scenario distopico, la giustizia algoritmica minaccia di oscurare la dimensione sociale, simbolica e partecipativa del processo, fondata sull'isonomia e sulla natura dialettica, ermeneutica e creativa del giudizio, che si sottrae alla riduzione a semplice calcolo[50] e si basa sul garantismo come nucleo irrinunciabile di un rinnovato umanesimo giudiziale[51].

5. Conclusioni. Verso la giustizia ibrida

Nel contesto del tecnodiritto contemporaneo, la progressiva digitalizzazione del processo penale rende plausibile l'affermarsi del modello metodologico della giustizia ibrida, in cui avvocati e magistrati saranno chiamati a integrare conoscenze giuridiche e competenze digitali[52]. L'accelerazione delle tecnologie dirompenti incide sull'accessibilità ai metodi anche alternativi di risoluzione delle controversie, integrando l'obiettivo primario delle politiche legislative sovranazionali in materia di giustizia, come dimostrano orientamenti ed esperienze statunitensi, britanniche ed europee[53].

In questo scenario si colloca lo sviluppo delle *Online Dispute Resolution* (ODR), che assumono un rilievo crescente anche per il processo penale in relazione agli istituti di giustizia negoziata

[50] F. Garapon, *La despazializzazione della giustizia*, Mimesis, Milano-Udine, 2021.

[51] O. Mazza, "Distopia del processo artificiale", in *Arch. pen.*, (2025), n. 1, pp. 1-20.

[52] P. Moro, *Diritto ibrido. Metodo del giurista e tecnologie dirompenti*, Mondadori, Milano, 2025.

[53] R. Susskind, *Online Courts and the Future of Justice*, Oxford University Press, Oxford, 2019.

e riparativa[54]. L'ibridazione tra competenze umane e intelligenza artificiale potrebbe anche consentire di superare i limiti delle tradizionali procedure digitali, aprendo a un modello di giustizia in cui il calcolo predittivo e la capacità ermeneutica si intrecciano[55].

Per evitare la deriva inquisitoria dell'automazione digitale, con conseguente centralizzazione probatoria e compressione del contraddittorio, occorre che il processo ibrido conservi la propria natura di rito sociale e garantista, strutturato sulla centralità del dialogo tra le parti e sulla dimensione pubblica della decisione, irriducibile ad una tecnica di calcolo ed espressione di un metodo argomentativo coordinato e diretto alla ricerca della verità dialettica e processuale.

In proposito, è stato notato che l'uso dell'intelligenza artificiale nel processo penale, pur offrendo strumenti di supporto alle decisioni, possa compromettere i principi del giusto processo a causa della difficoltà algoritmica di garantire trasparenza e motivazione. Pertanto, si è sostenuta la necessità di un "controllo umano significativo", capace di assicurare che le decisioni algoritmiche restino sottoposte a verifica critica e responsabilità umana, preservando i diritti fondamentali e la centralità del giudice[56].

Non vi è dubbio che alla costruzione retorica della motivazione giudiziale sia assegnato il ruolo decisivo nel processo penale e che ad essa sia affidato il presidio più importante delle garanzie difensive, tutelabili anche quando l'argomentazione sia ragionevolmente sostenuta dall'opera di un sistema evoluto e collaudato di intelligenza artificiale generativa, utilizzabile formalmente come una sorta di *amicus curiae*.

Infatti, è altrettanto certo che la giustizia ibrida sia ormai inevitabile nell'era del tecnodiritto e suggerisca ai professionisti forensi di acquisire nuove competenze metodologiche, capaci di governare le innovazioni tecnologiche senza snaturare i principi del giusto processo.

[54] L. Mingardo, *Giustizia digitale alternativa. Scenari e riflessioni critiche sulle Online Dispute Resolution*, Primiceri, Padova, 2020.

[55] J. Zeleznikow, "Using Artificial Intelligence to Enhance the Resolution of Disputes", in *Artificial Intelligence and Law*, 28 (2020), n. 2, pp. 211-240.

[56] G. Ubertis, "Intelligenza artificiale, giustizia penale, controllo umano significativo", in *Diritto Penale Contemporaneo,* (2020), n. 4, pp. 75-88.

Paolo Sommaggio

L'intelligenza artificiale tra tetica, zetetica e maieutica

Abstract: **Artificial Intelligence Between T-ethics, Zet-ethics, and Maieutics.** The ethics of Artificial Intelligence (AI) is an inherently multidisciplinary field of research aimed at investigating the axiological implications associated with the development and use of intelligent systems. This perspective focuses, on the one hand, on the ways in which AI can operate in accordance with ethical criteria and, on the other, on the design guidelines necessary to ensure that such systems respect the fundamental principles and values of the human condition. However, the question of the most appropriate ethical approach to guide this reflection remains unresolved: among the options considered are the virtue ethics paradigm, the consequentialist paradigm, and the deontological approach. These conceptual distinctions form the core theme of this contribution, starting from the assumption that AI ethics is largely analogous to the reflections traditionally developed in the field of interpersonal relations. From this perspective, it can be traced back to at least three fundamental interpretative guidelines: the tetico-regulatory approach, the zetetic-problematic approach, and, finally, the maieutic-Socratic approach.

Keywords: Artificial Intelligence, T-ethics, Zet-ethics, Maieutics, Moral Machines.

Sommario: 1. Introduzione – 2. Esseri umani e sistemi intelligenti – 3. Un approccio tetico alla intelligenza artificiale – 4. Un approccio zetetico alla intelligenza artificiale – 5. Un approccio maieutico alla intelligenza artificiale – 6. La macchina parresiastica – 7. Conclusioni.

1. Introduzione

Nelle pagine che seguono, presenterò alcune nozioni di intelligenza artificiale, cercando di soffermarmi sulle caratteristiche specifiche di queste nuove realtà. In seguito, dopo avere tracciato un primo stato dell'arte sulle questioni etiche ad essa legate, proporrò tre possibili approcci: l'approccio tetico, l'approccio zetetico e l'approccio maieutico.

L'intelligenza artificiale è un concetto difficile da definire esaustivamente[1]. Vi sono nozioni di IA che istituiscono una analogia con il comportamento dell'essere umano e definizioni che evitano ogni analogia. Ad esempio, una prima nozione che considera una assoluta mancanza di analogia con l'umano è la seguente: IA è qualsiasi tipo di sistema computazionale artificiale che mostri un comportamento intelligente, ovvero un comportamento complesso che favorisca il raggiungimento di certi obiettivi.

Una diversa nozione, che proviene dall'Università di Stanford, identifica invece l'IA come "una scienza e un insieme di tecniche computazionali che si ispirano – anche se in genere operano in modo diverso – al modo in cui gli esseri umani usano il loro sistema nervoso e il loro corpo per sentire, imparare, ragionare e agire"[2].

Una nozione più tecnica di intelligenza artificiale è quella contenuta nel Regolamento Europeo n.1689/25 (nominato 'Ai Act') che così recita: "Per 'sistema di IA' si intende un sistema automatizzato progettato per operare con vari livelli di autonomia, che può mostrare adattabilità dopo l'implementazione e

[1] Come noto, il termine "Intelligenza Artificiale" si deve al matematico John McCarthy il quale propose nel 1955 un progetto estivo di ricerca dal titolo: "*Dartmouth Summer Research Project on Artificial Intelligence*". Si veda al riguardo: J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", in *AI Magazine,* 27 (2006), n. 4, p. 12.

[2] Cfr. P. Stone *et al.*, *Artificial Intelligence and life in 2030. One-hundred-year study on Artificial Intelligence*, Stanford University Press, Redwood, 2016, p. 5. Questa definizione riprende le intuizioni di Marvin Minsky rispetto ad un rapporto di simulazione con le facoltà umane. Si veda M. Minsky, *The Society of Mind*, Simon & Schuster, New York, 1985. Un'altra definizione è la seguente: "una disciplina che studia i fondamenti teorici, le metodologie e le tecniche che permettono di progettare sistemi hardware e sistemi di programmi *software* capaci di fornire all'elaboratore elettronico delle prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana". In M. Somalvico, *L'Intelligenza Artificiale*, Rusconi, Milano, 1987. Con riguardo agli aspetti giuridici della IA, si veda P. Moro, "Intelligenza artificiale e tecnodiritto. Fondamenti etici ed innovazione legislativa", in P. Moro (a cura di), *Etica, diritto e tecnologia. Percorsi dell'informatica giuridica contemporanea*, FrancoAngeli, Milano, 2021, pp. 7-24.

che, per scopi espliciti o impliciti, deduce dagli input ricevuti come generare *output* quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali"[3].

Gli scopi principali di un sistema di IA riguardano, come noto, la percezione, la modellazione, la pianificazione e l'azione, ma le attuali applicazioni di IA includono anche analisi testuali, elaborazione del linguaggio naturale, ragionamento logico, giochi, sistemi di supporto decisionale, analisi dei dati, analisi predittiva, nonché veicoli autonomi e altre forme di robotica.

Le applicazioni pratiche sono altrettanto numerose. Si va dallo shopping in rete alla pubblicità, dalle ricerche on line ad assistenti digitali personalizzati. Inoltre, i sistemi di IA possono provvedere a traduzione automatica, alla organizzazione domotica delle abitazioni e delle infrastrutture, giungendo a quelle che vengono comunemente definite *smart cities*. L'Intelligenza artificiale, inoltre, viene utilizzata nei sistemi di sicurezza delle vetture, nella filiera agroalimentare e nella cybersicurezza. Questi differenti scopi ed applicazioni di tipo trasversale rendono necessario un dialogo tra discipline differenti, tanto sul piano della ricerca scientifica quanto sul piano della comunità globale, e richiedono perciò orizzonti di ricerca e di approfondimento sempre più multidisciplinari. In altre parole, l'Intelligenza Artificiale sembra essere uscita dai ristretti orizzonti degli "addetti ai lavori" dediti alla ricerca teorica pura per approdare in lidi decisamente più strumentali e per molti versi ad attitudine sociale, poiché la sua introduzione potrebbe rivelarsi *disruptive* nei confronti dei modelli consueti di organizzazione sociale, politica e giuridica[4].

[3] Si veda l'art. 3 del Regolamento europeo n. 1689/2024, noto anche come Ai Act ovvero la prima regolamentazione europea sulla Intelligenza Artificiale. Il Parlamento europeo, nel giugno 2024 ha formalmente adottato questa normativa con un'ampia maggioranza di 523 a favore, 46 contrari, 49 astenuti. Si tratta del primo atto normativo al mondo che disciplina l'intelligenza artificiale e rappresenta una pietra miliare della legislazione dell'Unione Europea e del mondo intero. Si veda anche la recente l.n. 132 del 23 settembre 2025 (la legge italiana che regolamenta la IA) che, all'art. 2, riprende la definizione europea.

[4] Vedi le stimolanti riflessioni che puntano ad un dialogo interdisciplinare S. Crafa, "Artificial Intelligence and Human Dialogue", in *Journal of Ehitcs and Legal Technologies*, 1 (2019), n. 1, pp. 44-56. Vedi anche P. Moro,

Le nostre abitudini quotidiane, infatti, potranno essere messe in discussione da sistemi e applicazioni dotati di sistemi di IA "debole" (magari di IA generativa come ChatGPT), ma vi è chi teme quella che viene definita IA forte che sarebbe, in linea teorica, capace di capire, pensare e avere stati cognitivi propri. Il timore è che l'IA forte possa sviluppare una personalità individuale così potente e pervasiva da comprendere o apprendere qualsiasi compito intellettuale umano con una potenza di esecuzione grandemente maggiore, fino a diventare una superintelligenza in grado di mettere in serio pericolo l'esistenza stessa della intera umanità[5]. Diviene perciò fondamentale provare ad immaginare un orizzonte etico condiviso per la macchina e per l'uomo.

2. Esseri umani e sistemi intelligenti

Indipendentemente dai pericoli paventati, appare certo che questa tecnologia ha un forte impatto sulla società che è cambiato radicalmente nel corso del tempo[6].

Negli ultimi anni molti governi, parlamenti, associazioni e circoli industriali si sono confrontati ed hanno redatto documenti

C. Sarra (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, FrancoAngeli, Milano, 2017.

[5] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford, 2016.

[6] Se confrontiamo come sono cambiati alcuni *slogan* relativi all'IA, notiamo un deciso cambio di paradigma: in pochi decenni, siamo passati dagli slogan "l'IA è impossibile" e "l'IA è solo automazione" (J. Lighthill, "Artificial Intelligence: A General Survey", in *Artificial intelligence: A Paper Symposion*, Science Research Council, London, 1973) a slogan del tipo "l'IA risolverà tutti i problemi" (R. Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Penguin, London, 1999) e "l'IA potrebbe ucciderci tutti" (N. Bostrom, E. Yudkowsky, *The Ethics of Artificial Intelligence*, in K. Frankish, W.M. Ramsey (eds.) *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, Cambridge, 2014, pp. 316-334).

Queste semplificazioni denunciano, a mio parere, alcune criticità. Innanzitutto, una irrisolta relazione tra esseri umani ed intelligenza artificiale: si è passati dal considerare l'IA come un semplice oggetto o strumento di automazione, all'osservare con sempre maggiore interesse e timore il suo comportamento.

e rapporti per interrogarsi sugli aspetti etici della intelligenza artificiale[7]. Non solamente sull'uso etico della IA, ma anche sulla progettazione etica e sul design etico di questi sistemi[8]. Questa precisazione appare necessaria anche in ragione della possibilità che lo stesso sistema di intelligenza artificiale possa produrre una risposta eticamente sensibile oppure tenere un comportamento eticamente rilevante e responsabile[9].

Possiamo affermare con una certa sicurezza che la cornice etica più condivisa dagli studiosi si basa su di una impostazione normativa, certamente ispirata alle intuizioni di Isaac Azimov. Il noto scrittore, infatti, ha fissato le tre leggi della robotica proprio al fine di costruire un orizzonte etico per i suoi robot dotati di un "cervello positronico", oggi diremmo di una intelligenza artificiale. Nella formulazione originaria, come noto, le leggi sono tre e si trovano enunciate per la prima volta nel racconto Runaround[10]:

1) Un robot non può recare danno a un essere umano né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno;

2) Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non vadano in contrasto alla Prima Legge;

[7] Si veda A. Jobin, M. Ienca, E. Vayena, "The Global Landscape of AI Ethics Guidelines", in *Nature Machine Intelligence*, 1 (2019), n. 9, pp. 389-399.

[8] Per approfondire il tema della distinzione tra *design* (progettazione) e utilizzazione di sistemi di Intelligenza artificiale, si veda: W. Houkes, P. Vermaas, *Technical Functions: On the Use and Design of Artefacts*, Springer, Dordrecht, 2010.

[9] Si veda U. Ruffolo, "Il problema della "personalità elettronica", in *Journal of Ethics and Legal Technologies*, 2 (2020), n. 1, pp. 75-88. Ed anche P. Moro, "Alle frontiere della soggettività: indizi di responsabilità delle macchine intelligenti", in U. Ruffolo (a cura di), *XXVI Lezioni di Diritto dell'Intelligenza Artificiale*, Giappichelli, Torino, 2021, pp. 37-47. Si veda infine P. Moro, "Persona elettronica. Una finzione giuridica per l'intelligenza artificiale", in *L'Ircocervo*, 23 (2024), n. 1, pp. 372-389.

[10] Si veda il racconto di science fiction I. Azimov, "Runaround", in *Astounding Science Fiction*, XXIX (1942), n. 1, 94-103. A queste tre leggi, Azimov aggiungerà nel corso degli anni una quarta legge, la legge 0.

3) Un robot deve proteggere la propria esistenza, purché la salvaguardia di essa non contrasti con la Prima o con la Seconda Legge.

Su ispirazione di queste "leggi", dunque, vengono specificate altre regole che si concentrano in modo particolare sui tradizionali principi della etica medica (ovvero della bioetica), oltre ad un ulteriore principio, l'*explicability*[11]. Si tratta, come noto, dei famosi principi di etica medica di Tom Beauchamp e James Childress del 1979 (poi conosciuti anche come "principi di bioetica") che, giova ricordarlo, sono: a) l'autonomia, intesa come rispetto per l'auto determinazione del paziente, ovvero la possibilità che egli accetti o rifiuti le prescrizioni mediche qualora abbia una appropriata capacità di decisione; b) il principio di beneficenza, ovvero cercare di operare al meglio per il benessere del paziente, in relazione ai *desiderata* dello stesso; c) il principio di non maleficenza, ovvero il principio secondo cui non bisogna nuocere al paziente; d) il principio di giusta allocazione delle risorse, ovvero il principio di giustizia[12]. A questi quattro principi, la comunità internazionale degli studiosi ritiene di dover aggiungere, come già anticipato, il principio della *explicability*, ovvero la possibilità di giustificare le proprie scelte. Questo nuovo principio costituirebbe un modo per ridurre l'opacità e l'inesplicabilità dei passaggi nei sistemi intelligenti autonomi[13].

[11] Si veda L. Floridi *et al.*, "AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", in *Minds and Machines*, 28 (2018), n. 4, pp. 689-707. Possiede il medesimo impianto anche il documento a cura dell'High Level Expert Group (HLEG) della Commissione europea dal titolo "Ethics Guidelines for Trustworthy AI" reso pubblico nel 2018 come bozza di lavoro per la consultazione.

[12] T. Beauchamp, J.F. Childress, *Principles of Medical Ethics*, Oxford University Press, Oxford, 1979.

[13] Sul tema della IA come frutto di processi legati a punti ciechi e non spiegabili, ovvero black boxes, vedi ad esempio, S. Robbins, "A Misdirected Principle with a Catch: Explicability for AI", in *Minds and Machines*, 29 (2019), n. 4, pp. 495-514. Per alcuni argomenti a difesa della *black box* e contro il principio della necessaria spiegazione di ogni decisione della macchina, si veda: E.A. Holm, "In Defense of the Black Box", in *Science,* 364 (2019), n. 6435, pp. 26-27.

Già da questo primo esempio possiamo notare come l'ambito dell'etica per l'intelligenza artificiale sia popolato di principi, regole, valori[14].

Ebbene, viene da chiedersi se questo sia l'unico modo per affrontare il tema dell'etica per l'intelligenza artificiale: estendere una serie di principi, che nascono in riferimento al comportamento umano, al comportamento delle macchine. Inoltre, sorge spontaneo un altro dubbio, intimamente connesso e che possiamo riassumere in una semplice domanda: ma è proprio vero che l'etica umana si riduce ad una serie di principi?

Se questo non è l'unico modo di concepire il tema etico anche nell'ambito umano, a maggior ragione non sarà l'unico nell'ambito della intelligenza artificiale.

È possibile, infatti, accanto ad un'etica principialista/normativista, che possiamo chiamare t-etica[15], considerare anche un altro tipo di etica, ovvero un'etica di tipo problematico ed infine un approccio di tipo oppositivo.

Il percorso che qui presento si snoda dunque tra etica a base normativa (o tetica), etica a base non normativa (o problematica, zetetica) ed etica a base di opposizioni (o maieutica). Cercherò di mostrare come questi approcci operino nell'orizzonte dell'etica umana e indagherò la loro applicabilità alla intelligenza artificiale. Anticipo subito che una conferma della bontà di questa impostazione ci giunge da Colin Allen[16] e dal suo schema di morale artificiale suddiviso in modelli *top-down*, *bottom-up* e ibridi. Come vedremo, infatti, l'approccio tetico dispiega, dal punto

[14] Per una ulteriore panoramica generale rimando a S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford University Press, Oxford, 2017; vedi anche il recente A. Chierici, *The Ethics of AI. Facts, Fictions and Forecasts*, New Degree Press, Washington, 2021. Sul fronte giuridico segnalo P. Moro, "Intelligenza artificiale e tecnodiritto. Fondamenti etici ed innovazione legislativa", cit., pp. 7-24. Nella medesima collettanea, segnalo altresì A. Pin, "Diritti Costituzionali e intelligenza artificiale", in P. Moro (a cura di), *Etica, diritto e tecnologia*, cit., pp. 45-61.

[15] Per un approfondimento, si veda P. Sommaggio, "Tetica I. Riflessioni preliminari", in F. Zanuso (a cura di), *Custodire il fuoco*, FrancoAngeli, Milano, pp. 165-221.

[16] C. Allen, I. Smit, W. Wallach, "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches", in *Ethics and Information Technology*, 7 (2005), n. 3, pp. 149-155.

di vista logico, una prospettiva di tipo "*top-down*", che parte dall'astratto del principio e muove al concreto della situazione o, meglio, dal generale del valore (magari considerato come un assioma) al particolare del comportamento in una data situazione. Il modello che, invece, ho chiamato zetetico si muove secondo una prospettiva di tipo *bottom-up*, ovvero dal problema di azione concreta alla regola, in altre parole secondo un movimento dal particolare al generale. Ebbene, mentre Colin Allen afferma che il terzo approccio costituisce un ibrido tra i due precedenti, la mia proposta, invece, come vedremo nei prossimi paragrafi, considera il terzo approccio come un movimento di tipo oppositivo maieutico, ovvero socratico[17].

[17] Sia consentito considerare qui brevemente le cd. "macchine morali", ovvero i sistemi di intelligenza artificiale sensibili al tema etico che possiamo anche definire "agenti morali artificiali" (AMA-*Artificial Moral Agents*). La loro tassonomia si può trovare in J.H. Moor, "The Nature, Importance and Difficulty of Machine Ethics", in *IEEE Intelligent Systems*, 21 (2006), n. 4, pp. 18-21. L'autore parte dalla constatazione che la tecnologia informatica è di natura normativa, dove i terminali (ad esempio gli algoritmi) sono considerati agenti. Tuttavia, in seguito alla evoluzione delle architetture digitali e della loro potenza di calcolo, possiamo oggi individuare almeno quattro tipi di agenti artificiali, in questo caso "etici", poiché possiamo non soltanto inserire principi etici nella programmazione, ma anche addestrare le macchine, attraverso sistemi di *deep learning*, ad avere un comportamento che noi esseri umani consideriamo eticamente sensibile. Questi tipi sono: 1) Agenti ad impatto etico: qualsiasi sistema intelligente che può essere valutato per le conseguenze eticamente sensibili delle sue azioni (dei suoi *output*). 2) Agenti etici impliciti: sistemi progettati per evitare risultati non etici. 3) Agenti etici espliciti, ovvero sistemi che possono ragionare sull'etica (vedi J.H. Moor, "Is Ethics Computable?", in *Metaphilosophy*, 26 (1995), n. 1-2, pp. 1-21). A tale riguardo Moor riporta un particolare esempio di un sistema implementato da Michael Anderson (vedi M. Anderson *et al.*, "Towards Machine Ethics: Implementing Two Action-Based Ethical Theories", in *Machine Ethics. Papers from AAAI Fall Symposium FS-05-06*, AAAI Press, Menlo Park (CA), 2005, pp. 1-7). Il loro primo modello di agente etico esplicito, Jeremy (dal nome di Jeremy Bentham), applicava l'utilitarismo edonistico. Jeremy stimava la probabilità di piacere o dispiacere per le persone interessate da un particolare atto. Un ulteriore modello è Ross, costruito secondo una struttura di tipo deontologico-kantiano. 4) Agenti etici completi: sistemi in grado di esprimere giudizi morali espliciti e di giustificarli. Il tema è troppo vasto per essere qui trattato esaustivamente. In questa sede mi limito a ricordare che uno dei primi tentativi di immaginare un agente etico artificiale completo si deve a Michael Anderson e Susan Leigh Anderson. (vedi M. Anderson, S.

3. Un approccio tetico alla intelligenza artificiale

Se volessimo trovare un punto di inizio dell'approccio tetico alla intelligenza artificiale, potremmo risalire al 2016, anno in cui si incomincia a considerare la possibilità di condividere alcuni principi e valori per fissare nuove regole per affrontare i rischi di uno sviluppo incontrollato della intelligenza artificiale.

In particolare, la data coincide con il forum di Davos dove si è trattato il tema della cd. "Quarta rivoluzione industriale", ovvero della convergenza di robotica, IA, sistemi autonomi e tecnologia dell'informazione.

Nel 2017 sono state pubblicate diverse raccolte di questi principi, e qui ritengo utile ricordare tra i molti, i *Principi di Asilomar* dell'influente Future of Life Institute.

Il Future of Life Institute è una organizzazione attiva dal 2015 (originata dalla conferenza "The future of Technology", tenutasi proprio quell'anno presso l'MIT di Boston) la cui *mission* è

Leigh Anderson, "Machine Ethics: Creating an Ethical Intelligent Agent", in *AI Magazine*, 28 (2007), n. 4, pp. 15-25), seguito da Wallach ed Allen. Nello studio dal titolo *Moral Machines - Teaching Robots Right From Wrong*, gli autori espongono le basi filosofiche degli "*Artificial Moral Agent*" (AMA) (Vedi W. Wallach, C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, New York, 2009), entità che, per le loro caratteristiche, necessitano di forme di tutela che vanno oltre la esigenza di sicurezza dell'utilizzatore (come avviene ad esempio per la sicurezza di un prodotto): se un sistema autonomo deve ridurre al minimo i danni, dovrebbe anche essere "consapevole" delle possibili conseguenze dannose delle sue azioni e dovrebbe scegliere le sue azioni anche alla luce di questa "consapevolezza", dove tali termini sono applicati ancora solo metaforicamente alle macchine (S. Fuselli, "Brain-Computer Interface e soggettività agente. Considerazioni etico-giuridiche", in P. Moro (a cura di), *Etica, Diritto e Tecnologia*, cit., pp. 151-173). Queste caratteristiche costituiscono le nuove frontiere relative agli agenti artificiali intelligenti ed eticamente esperti (Per una panoramica attuale sul tema degli agenti morali artificiali si veda: J.A. Cervantes *et al.*, "Artificial Moral Agents: A Survey of the Current Status", in *Science and Engineering Ethics*, 26 (2020), n. 2, pp. 501-532). Questo tipo di agenti etici artificiali non è ancora all'orizzonte poiché occorrerebbero coscienza, intenzionalità e volontà libera: tutti elementi che ancora faticano ad essere implementati in sistemi di intelligenza artificiale ma che in un prossimo futuro potrebbero diventare realtà. Ed allora perché non provare a capire quali principi, dilemmi e opposizioni possiamo teoricamente sperimentare con i sistemi di intelligenza artificiale.

quella di indirizzare la tecnologia, in particolare quella in grado di trasformare la vita sulla terra verso il benessere comune e di tenere lontano i rischi, soprattutto quelli che potrebbero portare danni su larga scala.

Con il titolo "Asilomar AI Principles", si intende dunque indicare la raccolta di 23 principi, volti a promuovere uno sviluppo sicuro e vantaggioso dell'intelligenza artificiale, stabiliti dal Future of Life Institute. I principi – che includono ricerca, etica e valori – sono emersi da una collaborazione tra ricercatori di IA, economisti, studiosi di diritto, eticisti e filosofi ad Asilomar, in California, nella conferenza del gennaio 2017[18].

I primi cinque principi sono relativi alla ricerca[19]. Gli ulteriori tredici principi sono relativi agli aspetti etici e valoriali della progettazione, gestione ed uso di sistemi di IA[20].

[18] Cfr. https://futureoflife.org/open-letter/ai-principles/.

[19] Essi prevedono: che l'obiettivo della ricerca sull'IA non sia di creare un'intelligenza senza controllo, ma un'intelligenza artificiale che possa portare benefici; che gli investimenti nell'IA vengano accompagnati da finanziamenti per la ricerca volta a garantirne l'uso benefico, comprese questioni spinose in informatica, economia, legge, etica e studi sociali; che sia presente uno scambio costruttivo e sano tra i ricercatori di IA e i responsabili politici; che venga promossa una cultura di cooperazione, fiducia e trasparenza tra i ricercatori e gli sviluppatori di IA; che i team che sviluppano sistemi di IA collaborino attivamente per evitare di diminuire gli standard di sicurezza.

[20] Essi prevedono che: che i sistemi di IA siano sicuri e protetti per tutta la loro vita operativa, e verificabili, ove possibile; che i sistemi di IA garantiscano la possibilità di accertare la causa di un qualunque danno; che un sistema autonomo coinvolto in un processo decisionale giudiziario fornisca sempre una spiegazione motivata e verificabile da un'autorità umana competente; che i progettisti e sviluppatori di sistemi avanzati di IA siano coinvolti nelle implicazioni morali del loro uso ed abuso e azioni; che i sistemi di IA altamente autonomi siano progettati in modo da garantire che i loro obiettivi e comportamenti siano allineati con i valori umani durante tutto il loro funzionamento; che i sistemi di IA siano progettati e gestiti in modo da essere compatibili con gli ideali di dignità umana, diritti, libertà e diversità culturale; che i sistemi di IA rispettino la *privacy* delle persone dovrebbero avere il diritto di accedere, gestire e controllare i dati che generano, dato il potere dei sistemi di IA di analizzare e utilizzare tali dati; che l'applicazione dell'IA ai dati personali non limiti in modo irragionevole la libertà reale o percepita delle persone; che le tecnologie dell'IA porti benefici e potenzi il maggior numero possibile di persone; che i vantaggi economici creati dall'IA siano condivisi in modo ampio, a beneficio di tutta l'umanità; che gli esseri umani

Infine, i rimanenti cinque principi etici sono relativi a scelte di lungo periodo che avrebbero ripercussioni nei confronti di tutta l'umanità[21]. L'approccio è molto analitico e presenta un approfondimento rilevante di diverse possibilità tecniche. La prospettiva seguita è certamente incentrata su comportamenti in grado di evitare rischi potenziali causati dall'intelligenza artificiale. In particolare, si ritiene che l'Intelligenza artificiale possa arrivare a compromettere addirittura l'esistenza dell'umanità sulla terra.

Vediamo adesso brevemente un altro documento che contiene principi e valori di una differente organizzazione attiva nell'ambito della intelligenza artificiale.

L'AI4People Institute riunisce aziende leader nel settore dell'IA, accademici, gruppi della società civile e governi per valutare i rischi dell'IA e sostenere uno sviluppo e una diffusione dell'IA responsabili e vantaggiosi. La missione è garantire che l'intelligenza artificiale (IA) sia di beneficio e al servizio dell'umanità nel suo complesso.

Questa *mission* è stata inizialmente sostenuta da Atomium-EISMD e da membri fondatori come Audi, Elsevier, Fujitsu, Google, Meta, Intesa San Paolo, Johnson & Johnson e Microsoft.

Nel febbraio del 2018 AI4People pubblica un report con lo scopo dichiarato di sviluppare il dibattito europeo sul tema

siano sempre in grado di scegliere come e se delegare le decisioni ai sistemi di IA, per raggiungere gli obiettivi scelti; che i sistemi di IA altamente avanzati rispettino e migliorino, piuttosto che sovvertire, i processi sociali e civili da cui dipende il benessere della società; che si eviti la corsa agli armamenti autonomi letali.

[21] Essi prevedono: che si eviti di formulare ipotesi forti sui limiti massimi delle capacità future dell'intelligenza artificiale; che l'IA, dato che potrebbe rappresentare un cambiamento profondo nella storia della vita sulla Terra, dovrebbe essere pianificata e gestita con attenzione e risorse adeguate; che i rischi posti dai sistemi di IA, soprattutto quelli catastrofici o esistenziali, siano oggetto di una pianificazione e di sforzi di mitigazione commisurati all'impatto previsto; che i sistemi di IA progettati per auto-migliorarsi ricorsivamente o auto-replicarsi in modo tale da portare a un rapido aumento della qualità o della quantità siano soggetti a rigorose misure di sicurezza e di controllo; che l'IA, anche in senso forte, sia sviluppata solo al servizio di ideali etici ampiamente condivisi e a beneficio di tutta l'umanità piuttosto che di uno Stato o di un'organizzazione.

dell'etica dell'intelligenza artificiale, di stigmatizzare i rischi connessi con l'IA e di promuovere i diritti fondamentali e la dignità umana ad essa connessi[22]. In particolare, il documento presenta una serie di "Raccomandazioni per una buona società dell'intelligenza artificiale". Il documento si compone di due parti: un preambolo e 20 punti d'azione. I punti d'azione sono di quattro tipi: valutare, sviluppare, incentivare e sostenere. Alcune raccomandazioni possono essere intraprese direttamente dai responsabili politici nazionali o europei, in collaborazione con le parti interessate, se opportuno. Per altre, i responsabili politici possono svolgere un ruolo di supporto agli sforzi intrapresi o guidati da terzi.

Queste raccomandazioni includono principi etici come: il rispetto per l'autonomia umana; la necessaria prevenzione del danno; il principio di giustizia; il principio di spiegabilità dei meccanismi interni al sistema autonomo[23].

Nel 2018, il gruppo europeo sull'etica nelle scienze e nelle nuove tecnologie licenzia uno *statement* (una dichiarazione congiunta) mirante ad ottenere una serie di garanzie da parte degli stati europei. Questa dichiarazione chiede l'avvio di un processo che apra la strada alla costituzione di un quadro etico e legale comune e riconosciuto a livello internazionale per la progettazione, la produzione, l'uso e la governance dell'intelligenza artificiale, della robotica e dei sistemi "autonomi". La dichiarazione propone anche una serie di principi etici fondamentali, basati sui valori sanciti dai Trattati dell'UE e dalla Carta dei diritti fondamentali dell'UE[24].

Sempre nel 2018, vede la luce un'altra delle maggiori opere relative alla questione etica per l'IA ed i sistemi autonomi: le Raccomandazioni Europee sulle decisioni automatizzate basate

[22] L. Floridi *et al.*, "AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", in *Minds and Machines*, 28 (2018), n. 4, pp. 689-707.

[23] Dall'azione di AI4People ha preso forma la struttura del Regolamento europeo n. 1689/2024, l'Ai Act.

[24] European Commission: European Group on Ethics in Science and New Technologies and Directorate-General for Research and Innovation, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*, Publications Office, Brussels, 2018.

su addestramento Machine Learning (European Raccomandations on Machine-Learned Automated Decision Making)[25].

In particolare, con riguardo agli aspetti etici, possiamo ricordare la Raccomandazione n. 2 che così statuisce:

> 2. Garantire che l'Etica rimanga in prima linea nello sviluppo e nell'implementazione dell'ADM, e parte integrante dello sviluppo e dell'impiego delle ADM. Come per la salute e la biologia, i Paesi membri e l'Unione Europea dovrebbero sviluppare comitati etici per consigliare i sistemi sociali, politici, accademici e legali sulle conseguenze positive e negative di iniziative, strumenti e sistemi di ADM. Come custode dell'interesse pubblico, una nuova agenzia europea potrebbe supervisionare lo sviluppo e l'impiego di Autonomous Decision Maching machine-learned in tutta Europa[26].

In questa raccomandazione, balza agli occhi il parallelo con i principi bioetici.

Proseguendo:

> 3. Promuovere la progettazione di ADM sensibili ai valori. I programmi appropriati nell'ambito dell'istruzione superiore dovrebbero insegnare tecniche quali la progettazione sensibile ai valori e sottolineare in altro modo che i valori sociali e le priorità etiche degli utenti della tecnologia devono essere inseriti in tutti gli aspetti e gli elementi dell'ADM.

Questo a sottolineare l'importanza degli aspetti etici e relativi ai valori tanto nella fase di progettazione/design, quanto nella fase di utilizzazione di sistemi di IA che generano decisioni autonome.

A questo punto credo sia necessario qui menzionare l'espressione più compiuta dell'approccio tetico in relazione alla intelligenza artificiale. La regolamentazione giuridica della intelligen-

[25] Si veda Aa.Vv., *When Computers Decide: European Raccomadations on Machine-Learned Automated Decision Making*, 2018. Si tratta di un report a cura di Informatics Europe, Acm Europe Council, Euacm Policy Committee. In questo report vengono analizzate le principali implicazioni dell'adozione diffusa di processi decisionali automatizzati guidati da *Machine Learning*, con particolare attenzione alle sue ramificazioni tecniche, etiche, legali, economiche, sociali e educative. Vengono inoltre formulate una serie di raccomandazioni che i responsabili politici potrebbero prendere in considerazione.

[26] *Ibidem*.

za artificiale che è stata adottata nel marzo del 2024 e promulgata ufficialmente il 13 giugno 2024: mi riferisco al Regolamento n. 1689/2024, noto anche come Ai Act che, analogamente al GDPR (che prevede la regolamentazione e la protezione dei dati e della *privacy* del cittadino europeo), disciplina la produzione e l'uso dell'intelligenza artificiale. La struttura formale di quest'atto prevede 180 considerando, 113 articoli e 13 allegati.

Che si tratti di un atto regolativo e che stabilisca alcuni principi etici appare innegabile visto che, ad esempio, al Considerando n. 6 si definisce l'intelligenza artificiale come una *tecnologia antropocentrica*, ovvero "uno strumento per le persone con il fine ultimo di migliorare il benessere degli esseri umani" (Considerando 6, Ai Act)

L'architettura dell'Ai Act prevede un approccio basato sul rischio: più alto è il rischio possibile, più rigorosa e stringente è la regolamentazione.

Si tratta si regole che devono essere rispettate dai fornitori che immettono sul mercato o in servizio nei paesi dell'Unione sistemi di IA, indipendentemente dal fatto che essi siano stabiliti o ubicati nell'Unione o in un paese terzo, atteso che ciò che conta è che l'*output* prodotto dal sistema di IA sia utilizzato nell'Unione.

Il livello di rischio viene definito inaccettabile quando colpisce i diritti fondamentali dei cittadini e i sistemi di IA sono pertanto vietati:

- quando distorcono il comportamento di una persona attraverso tecniche subliminali o sfruttando vulnerabilità specifiche in modi che causano o sono suscettibili di causare danni fisici o psicologici;

- quando consentono la valutazione/classificazione dell'affidabilità (*rating*) di persone fisiche mediante l'attribuzione di un punteggio sociale con finalità generali da parte di autorità pubbliche, tali da determinare trattamenti pregiudizievoli e/o sfavorevoli (manipolazione comportamentale cognitiva, lo *scraping* non mirato delle immagini facciali da internet o da filmati di telecamere a circuito chiuso, il riconoscimento delle emozioni sul luogo di lavoro e negli istituti di istruzione, il punteggio sociale, la categorizzazione biometrica per dedurre dati sensibili, quali

orientamento sessuale o le convinzioni religiose e alcuni casi di polizia predittiva per le persone);

- quando consentono l'identificazione biometrica remota, salvo casi eccezionali autorizzati dalla legge riconducibili ad attività di prevenzione e contrasto del crimine, soggette a garanzie specifiche per il rispetto dei diritti fondamentali (casi di vittime di determinati reati, e per la prevenzione di minacce reali, presenti o prevedibili, come attacchi terroristici o la ricerca di persone sospettate dei più gravi reati).

Il rischio invece viene definito alto quando può causare danni significativi alla salute, sicurezza, diritti fondamentali e ambiente (Allegato III). Rientrano in questa categoria gli usi legati alle infrastrutture critiche, istruzione e formazione professionale, occupazione, servizi pubblici e privati di base (assistenza sanitaria, banche…), sistemi di contrasto alla migrazione e gestione delle frontiere, giustizia e processi democratici. Questi sistemi sono soggetti ad obblighi rigorosi per poter essere immessi sul mercato: obblighi di aumento della sicurezza, della trasparenza, del controllo umano, di standard elevati di cybersicurezza, di condivisione di informazioni adeguate con l'utente, nonché obbligo di uno specifico sistema di gestione dei rischi, oltre a garanzie di alta qualità dei *dataset*s e tracciabilità dei risultati. Gli interessati avranno diritto di presentare reclami, per mancata conformità al regolamento, alle pertinenti autorità di vigilanza e a ricevere spiegazioni sulle decisioni adottate.

Per i sistemi di intelligenza artificiale a rischio limitato sono introdotti specifici obblighi di trasparenza per garantire che gli individui siano informati di trovarsi di fronte a sistemi e prodotti generati o che utilizzano IA. Tutti i fornitori devono garantire che i sistemi di IA destinati ad interagire direttamente con le persone fisiche, siano progettati e sviluppati in modo tale che le persone interessate siano di fatto informate di essere in interazione con un sistema di IA. Tale obbligo non si applica ai sistemi autorizzati dalla legge per accertare, prevenire, indagare o perseguire reati, fatte salve le tutele adeguate per i diritti e le libertà di terzi.

L'Ai Act è solo uno dei molteplici esempi del gran numero di raccolte di principi e valori che diversi gruppi di ricercatori,

esperti, politici e stakeholder hanno prodotto nel corso di questi ultimi anni.

Un elemento che credo possa risultare di indubbio interesse è che l'istanza tetica, ovvero il bisogno di regolamentazione della intelligenza artificiale, si è mossa all'interno di una polarità che va dalla obbedienza ad un principio (ad esempio il rispetto della vita umana) alla considerazione anche degli effetti di un certo comportamento del sistema di intelligenza artificiale (ad esempio la considerazione del rischio rispetto agli *output* di un sistema di intelligenza artificiale). Ebbene, ciò che mi preme far notare è che l'istanza tetica si muove tra una direzione che possiamo definire deontologica o deontologista e una direzione consequenzialista o utilitaristica.

Possiamo affermare, dunque, che le raccolte normative fatte di principi e di valori, nonché da ultimo la legislazione dell'Unione europea, hanno come matrici culturali due modelli di riferimento che sono il modello deontologico ed il modello etico consequenzialista. Il primo sostiene che il valore etico di una azione (e quindi di un *output* se riferito ad un sistema intelligente) dipende dalla sua natura intrinseca, mentre il secondo afferma che il valore di una azione (di un *output*) dipende dalle sue conseguenze.

La disputa tra i sostenitori di queste due impostazioni etiche è secolare e non vogliamo certo risolvere la questione nelle pagine di questo saggio. Tuttavia, è bene ricordare che i rappresentanti più iconici di questi due modelli sono rispettivamente Immanuel Kant per l'approccio deontologico[27] e Jeremy Bentham[28] e John Stuart Mill[29] per quanto riguarda l'approccio consequenzialista. Ricordiamo incidentalmente che il motto della posizione deontologista è che la qualità etica e morale degli atti risiede nei prin-

[27] I. Kant, *Groundwork of the Metaphysics of Morals*, Oxford University Press, Oxford, 2002.

[28] J. Bentham, *An Introduction to the Principles of Morals and Legislation*, Clarendon Press, Oxford, 1907. Si veda anche F. Zanuso, *Utopia ed Utilità. Saggio sul pensiero filosofico-giuridico di Jeremy Bentham*, Cedam, Padova, 1989.

[29] J.S. Mill, *Utilitarianism*, Oxford, New York, 1998. Si veda anche H.R. West, *An Introduction to Mill's Ethics*, Cambridge University Press, Cambridge, 2004.

cipi o nelle massime in base ai quali l'agente agisce e non negli effetti di tali atti. Mentre il motto della posizione consequenzialista è che la qualità morale (etica) di una azione risiede nel ricercare la massima felicità per il maggior numero di persone.

In tempi più recenti la polarità tra deontologisti e utilitaristi si è riproposta: da un lato Frances Kamm introduce il principio del cd. danno ammissibile (*permissible harm*) secondo il quale si può fare del male per salvare più persone se e solo se il danno è l'effetto di un aspetto del bene maggiore stesso, e non solo un mezzo efficiente per prevenire un male maggiore[30]. Dall'altro lato, quello consequenzialista, possiamo invece limitarci a ricordare la proposta di Popper di trasformare il motto dell'utilitarismo classico in un motto secondo cui è morale minimizzare la sofferenza per il maggior numero di persone. Egli sostiene, infatti, che "non c'è, dal punto di vista etico, alcuna simmetria tra la sofferenza e la felicità, o tra il dolore e il piacere... La sofferenza umana fa un appello morale diretto, cioè la richiesta di aiuto, mentre non c'è un appello analogo per aumentare la felicità di un uomo che sta bene in ogni modo"[31]. Dunque, questa riformulazione del consequenzialismo classico viene ad essere naturale antistrofe del deontologismo contemporaneo[32].

Come abbiamo visto, l'impostazione tetica considera l'etica dell'intelligenza artificiale (o etica artificiale) secondo una prospettiva normativa. E dunque secondo un approccio di tipo *top-down*. L'idea di fondo dietro questa prospettiva è quella di costruire un agente morale artificiale dove i principi, i valori costituiscano le regole per selezionare l'azione più appropriata eticamente.

[30] Il principio del danno ammissibile è un tentativo di fornire una linea guida deontologica per determinare le circostanze in cui i soggetti sono autorizzati a causare danni ad altri con le loro azioni. Si veda F.M. Kamm, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*, Oxford University Press, Oxford, 2007.

[31] Si veda K.R. Popper, *The Open Society and Its Enemies*, Routledge, London, 1945, 284-285.

[32] Fin dai primi tentativi di costruire un agente morale artificiale, l'orizzonte tetico-normativo si muoveva tra Kant e Bentham. Si veda J. Gips, "Towards the Ethical Robot", in K. Ford, C. Glymour, P. Hayes (eds.), *Android Epistemology*, MIT Press, Cambridge, 1995, pp. 243-252.

Secondo Colin Allen questo approccio ha dato risultati non incoraggianti poiché:

> Both consequentialist (e.g., utilitarian) and deontological (e.g., Kantian) approaches raise their own specific computational problems, but they also raise a common problem of whether any computer (or human, for that matter) could ever gather and compare all the information that would be necessary for the theories to be applied in real time[33].

In altre parole, il modello tetico non considererebbe la necessità che, in tempo reale, vengano raccolte e confrontate tutte le informazioni utili ad applicare un principio piuttosto che un altro.

Un'obiezione più radicale contro l'approccio *top-down* alla moralità "artificiale" è il cosiddetto problema della cornice (*frame*)[34]. Intuitivamente, il problema consiste nel separare le informazioni rilevanti da quelle irrilevanti. Nella sua forma tecnica, il problema consiste nel fatto che specificare le condizioni che sono influenzate dalle azioni di un sistema non autorizza, nella logica classica, un'inferenza che porti alla conclusione che tutte le altre condizioni rimangono fisse. Se il problema tecnico è in gran parte considerato risolto (anche all'interno di contesti strettamente logici), rimane una versione filosofica più ampia della questione, enunciata per la prima volta da John McCarthy e Patrick Hayes, che non è ancora vicina a una soluzione[35].

Venendo infine ad un esempio di applicazione pratica dell'approccio normativo (tetico) possiamo ricordare il cd. modello asi-

[33] C. Allen, I. Smit, W. Wallach, "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches", in *Ethics and Information Technology*, 7 (2005), n. 3, p. 153.

[34] Per un approfondimento, vedi M. Shanahan, voce "The Frame Problem", in E.N. Zalta, U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*, Spring 2016, Standford.

[35] J. McCarthy, P.J. Hayes, "Some Philosophical Problems from the Standpoint of Artificial Intelligence", in B. Meltzer, D. Michie (eds.), *Machine Intelligence*, Vol. 4, Edinburgh University Press, Edinburgh, 1969, pp. 463-502; vedi anche T. Horgan, M. Timmons, "What Does the Frame Problem Tell Us About Moral Normativity?", in *Ethical Theory and Moral Practice*, 12 (2009), n. 1, p. 25.

moviano di Winfield[36]. Si tratta di un modello di intelligenza artificiale basato sulle leggi di Asimov e su di una architettura *top-down*.

Le raccolte di principi e, in ambito europeo, anche la legislazione (Ai Act) che ha positivizzato una serie condivisa di valori, confermano un approccio di tipo tetico al tema dell'intelligenza artificiale. Ma, come abbiamo anticipato, questo non è l'unico approccio al tema. Per approccio zetetico intendo una serie di modelli "*bottom-up*" (dal basso verso l'alto) che non impongono una specifica teoria morale o una serie di principi o regole, ma che cercano di fornire contesti in cui il comportamento appropriato viene selezionato e premiato. Questi modelli di approccio allo sviluppo della sensibilità morale comportano un apprendimento (ancorché frammentario) tanto attraverso l'esperienza per mezzo di tentativi e fallimenti inconsapevoli, quanto attraverso lo sviluppo "educativo" di una macchina per l'apprendimento[37].

4. Un approccio zetetico alla intelligenza artificiale

Pioniere dell'approccio zetetico è certamente Alan Turing. L'autore, nel suo classico contributo dal titolo "Computing Machinery and Intelligence", ha sostenuto che se riusciamo a sottoporre un computer a un regime educativo paragonabile all'educazione che riceve un bambino: "Possiamo sperare che alla fine le macchine possano competere con gli uomini in tutti i campi puramente intellettuali"[38].

Oggi, infatti, attraverso il *deep learning* è possibile "addestrare" un sistema di intelligenza artificiale come se si trattasse di un processo di educazione e di apprendimento di un bambi-

[36] A.F.T. Winfield, C. Blum, W. Liu, "Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection", in Aa.Vv., *Lecture Notes in Computer Science*, Springer, Berlin, 2014, pp. 85-96.

[37] Per un nuovo modello di formazione che si compone di un approccio *concept based*, di un approccio *problem based* e di un approccio *challenge based* (e che costituisce buona parte delle intuizioni sviluppate in questo lavoro) si veda P. Sommaggio *et al.*, *Nuove strategie per la formazione giuridica*, Giappichelli, Torino, 2021.

[38] Vedi A. Turing, "Computing Machinery and Intelligence", in *Mind*, 59 (1950), n. 236, pp. 433-460.

no anche nell'ambito etico[39]. Le simulazioni mostrano, infatti, come sia possibile costruire un ambiente dove sperimentare problemi morali o, meglio, situazioni dilemmatiche utili ad istruire la macchina a reagire in termini etici[40].

Le matrici culturali di questo approccio zetetico, ovvero che parte dal problema ed arriva ad un principio/valore/regola, si trovano ben esposte, a mio avviso, nelle riflessioni di Theodor Viehweg. In diverse opere l'autore si confronta, infatti, con i concetti di dogmatica e di zetetica[41].

Se da un lato la dogmatica, per come viene intesa da Viehweg, potrebbe essere per molti versi paragonata alla tetica di cui abbiamo mostrato alcuni esempi, la zetetica è invece caratterizzata da una ricerca illimitata (*zetein*), ovvero da una costante problematizzazione, e si specifica come domanda di giustificazione di un comportamento e "messa alla prova" delle risposte. In contrapposizione alla prospettiva dogmatico-tetica, che si basa su di concetto posto da non mettere in discussione (dogma), il pensiero zetetico ha una funzione prettamente conoscitiva, la quale non permette che principi e valori fondamentali siano esclusi dalla discussione. Ci troviamo di fronte ad un approccio zetetico

[39] Questo approccio è stato sviluppato attraverso piattaforme che simulano strutture biologiche di vita artificiale come, ad esempio, l'esperimento *Artificial Life*. Su questo tema si vedano gli esperimenti di Danielson, in particolare P. Danielson, *Artificial Morality: Virtuous Robots for Virtual Games*, Routledge, New York, 1992 e P. Danielson, M*odeling Rationality, Morality and Evolution*, Oxford University Press, New York, 1998.

[40] Ulteriori approfondimenti in M. Komosinski, A. Adamatzky, *Artificial Life Models in Software*, Springer, New York, 2009. Per approfondire si veda anche F. Berto, G. Rossi, J. Tagliabue, *La matematica dei modelli di riferimento*, Lampi di stampa, Milano, 2010

[41] Si veda T. Viehweg, "Ideologie und Rechtsdogmatik", in W. Maihofer (ed.), *Ideologie un Recht*, Klostermann, Frankfurt am Main, 1969, pp. 85-96 dove l'autore contrappone la dogmatica giuridica, caratterizzata dalla invarianza e dalla propria funzione sociale di fissazione e condivisione che diventa una prescrizione di comportamento, alla zetetica che ha una funzione essenzialmente cognitiva. Per procedere nella ricerca occorre, secondo Viehweg, sottoporre i dogmi alla discussione che caratterizza la zetetica. Sulla modalità della argomentazione zetetica si veda T. Viehweg, "Notizen zu einer rhetorischen Argumentationstheorie der Rechtsdisziplin", in *Jahrbuch fur Rechtssoziologie und Rechtstheorie*, 2 (1972), p. 439 ss.

allorché determinate asserzioni vengano discusse e motivate e non poste come assiomi[42].

Per meglio comprendere l'approccio di tipo zetetico nel campo dei problemi etico-morali, utilizzeremo un particolare esperimento mentale, noto come "*trolley problem*".

Il "*runaway trolley*" (vagone in movimento) è un esperimento mentale, originariamente formulato da Philippa Ruth Foot nel 1967[43].

Il problema è stato ampiamente analizzato da Judith Jarvis Thomson[44] e Peter Unger[45], ed entrambi hanno fornito diverse varianti allo scenario originale, che ha subito pertanto diverse modifiche nel corso del tempo.

Vediamo due delle più note formulazioni del problema del trolley[46].

a) Primo esempio: lo spettatore.

Il problema è il seguente. Un carrello in corsa sfreccia sui binari della ferrovia. Davanti a noi, si vedono cinque persone ferme sui binari. Sono legate ed impossibilitate a muoversi mentre il carrello è diretto verso di loro. Se non si fa nulla, le cinque persone moriranno. Siamo in piedi accanto a una leva che può deviare il carrello su un binario laterale, al quale è legata un'altra persona ed anche questa, se il treno verrà deviato, morirà.

Abbiamo dunque solo due opzioni:

i) non fare nulla, lasciando che il carrello uccida cinque persone sul binario principale, oppure

ii) tirare la leva, lasciando che il carrello uccida una sola persona sul binario laterale. Secondo la ricerca di Judith Thomson, le persone interrogate scelgono di tirare la leva quasi all'unani-

[42] Un esempio di zetemata è certamente il noto esperimento contenuto in E. Awad *et al.*, "The Moral Machine Experiment", in *Nature*, 563 (2018), n. 7729, pp. 59-64.

[43] P. Foot, "The Problem of Abortion and the Doctrine of the Double Effect", in *Oxford Review*, (1967), n. 5, pp. 5-15.

[44] Si veda J.J. Thomson, "Killing, Letting Die, and the Trolley Problem", in *Monist*, 59 (1976), n. 2, pp. 204-217.

[45] Vedi P.K. Unger, *Living High and Letting Die: Our Illusion of Innocence*, Oxford University Press, New York, 1996.

[46] Per un approfondimento, vedi P. Sommaggio, S. Marchiori, "Break the Chains: a New Way to Consider Machine's Moral Problems", in *BioLaw Journal*, 5 (2018), n. 3, pp. 241-257.

mità, ma le ragioni della loro risposta sono molto diverse. Sebbene tutti coloro a cui è stato presentato il problema del carrello abbiano risposto che era accettabile deviare il carrello sull'altro binario per un risparmio netto di quattro vite,

> alcune persone dicono qualcosa di più forte del fatto che è moralmente lecito deviare il carrello. Dicono che, dal punto di vista morale, si deve tirare la leva, che la moralità impone di farlo. Altri non sono d'accordo sul fatto che la moralità richieda di spostare il carrello, e provano persino un certo disagio all'idea di spostarlo[47].

Questa disputa, tuttavia, non porta a risultati diversi: se intendiamo il problema del carrello come una scelta tra l'uccisione di una persona o l'uccisione di cinque persone, sembra giustificato che la maggioranza trovi accettabile tirare la leva e uccidere una persona per salvarne cinque[48].

Gli studi neuroscientifici sono giunti alle stesse conclusioni di Thomson, trovando che la maggior parte dei soggetti sottoposti al test ha dichiarato che tirare la leva è appropriato[49]. Va notato che nella formulazione originale di Foot il soggetto gravato della scelta dilemmatica è l'autista del carrello, mentre nello scenario dello spettatore la persona che compie la scelta è – come suggerisce il nome – uno spettatore che osserva la scena da lontano.

b) Secondo esempio: l'uomo grosso (o il ponte)

La seconda famosa rappresentazione del problema del *trolley* è questa: un vagone sta scendendo lungo un binario verso cinque persone. Voi siete su un ponte sopra i binari e accanto a voi c'è un uomo molto corpulento.

È possibile fermare il carrello spingendo questa persona sul binario e questa azione porterà il vagone a fermarsi ma ucciderà l'uomo grosso.

[47] Si veda J.J. Thomson, "The Trolley Problem", in *Yale Law Journal*, 94 (1985), n. 6, pp. 1395-1396.

[48] Vedi E. Di Nucci, "Self-Sacrifice and the Trolley Problem", in *Philosophical Psychology*, 26 (2013), n. 5, pp. 662-672.

[49] J.D. Greene *et al.*, "An fMRI Investigation of Emotional Engagement in Moral Judgment", in *Science*, 293 (2001), n. 5537, pp. 2105-2108.

Abbiamo anche in questo caso due opzioni: *i)* non fare nulla e lasciare che il carrello uccida cinque persone o *ii)* uccidere l'uomo per salvare i cinque.

In contrasto con la reazione al primo esempio, la stessa maggioranza di persone che approvava l'idea di tirare la leva per sacrificare una vita e salvarne cinque, disapprova invece l'idea di spingere l'uomo sui binari per salvare lo stesso numero di quattro vite.

Questi due esempi mettono in crisi il rapporto tra il principio professato (ad esempio il principio di utilità: vengono risparmiate quattro vite) e la situazione dilemmatica concreta in cui ci si trova. Nel secondo caso, causare direttamente la morte del soggetto viene visto come un comportamento certamente peggiore del comportamento dello scenario precedente, ovvero lasciar morire.

Questo, secondo Eric Rakowski, può essere dovuto al fatto che uccidere qualcuno incidentalmente o indirettamente nel corso della salvezza di altri viene generalmente considerato ammissibile[50], "a condizione che i guadagni complessivi superino le perdite totali della giusta quantità". Invece spingere direttamente l'uomo corpulento supera quelli che consideriamo i margini di tollerabilità morale dell'azione[51].

Tuttavia, qui i principi non costituiscono un punto di partenza del ragionamento che giustifica una scelta ma sono, semmai, il punto di arrivo, ovvero una giustificazione a posteriori della sensazione di approvazione o repulsione che si prova a tirare una leva o a spingere una persona.

Anche in questo caso, le principali inclinazioni etiche che guidano le persone che si trovano di fronte a questo problema possono ridursi a due: il modello deontologico ed il modello consequenzialistico[52]. Ricordo che il primo implica che la moralità di un'azione

[50] J. Haidt, J. Baron, "Social Roles and the Moral Judgement of Acts and Omissions", in *European Journal of Social Psychology*, 26 (1996), n. 2, pp. 201-218.

[51] E. Rakowski, "Taking and Saving Lives", in *Columbia Law Review*, 93 (1993), n. 5, p. 1071.

[52] P. Conway, B. Gawronski, "Deontological and Utilitarian Inclinations in Moral Decision Making: A Process Dissociation Approach", in *Journal of Personality and Social Psychology*, 104 (2013), n. 2, pp. 216-235.

dipende dalla sua natura intrinseca. Il secondo afferma, invece, che la moralità di un'azione dipende dalle sue conseguenze.

In questa situazione problematica, dunque, il ragionamento che muove l'agente (anche l'agente artificiale) è di tipo induttivo, ovvero dal caso al principio (alla regola) e non dal principio (regola) al caso. Un movimento di tipo *bottom-up* e non *top-down*. Inoltre, l'esperimento del trolley consente di notare come le scelte siano influenzate dal tipo di azione (uccidere vs. lasciar morire) e dal contesto di azione (azione diretta o azione indiretta).

Per fornire un concreto esempio di come queste considerazioni possono tradursi in un modello di addestramento della macchina, riportiamo lo studio denominato GenEth.

Si tratta di un analizzatore di dilemmi morali, in altre parole di un test: una sorta di test di Turing etico. Da un dilemma fornito al sistema, è possibile ricavare induttivamente a quale principio morale ci si possa riferire in ragione del comportamento scelto[53].

Addestrando l'agente etico artificiale a sviluppare connessioni di tipo deontologico o consequenzialista si ottiene certamente un progresso nelle capacità di analisi etica di una certa situazione. Ma questo non è sufficiente.

La contestazione più forte dell'approccio *bottom-up* proviene da Catrin Misselhorn, che sostiene che i modelli di intelligenza artificiale basati su di un approccio *bottom-up* sono poco adatti ad implementare le capacità morali nei sistemi artificiali, perché pongono problemi di operabilità, sicurezza e accettazione. È difficile valutare quando un sistema possiede la capacità di apprendimento morale e come si evolverà. Poiché il comportamento di un sistema di questo tipo è difficile da prevedere e da spiegare, gli approcci dal basso verso l'alto si mostrano poco adatti a scopi pratici, e potrebbero mettere a rischio i potenziali utenti. Inoltre, è particolarmente difficile ricostruire come un sistema sia arrivato a una decisione etica[54].

[53] Si veda M. Anderson, S. Leigh Anderson, "GenEth. A General Ethical Dilemma Analyzer", in Aa.Vv., *Proceedings of the Twenty-Eight AAAI Conference on Artificial Intelligence*, Quebec City, 2014, pp. 253-261.

[54] Vedi le considerazioni contenute in C. Misselhorn, "Artificial Moral Agents: Conceptual Issues and Ethical Controversy", in S. Voeneky *et al.* (eds.), *The Cambridge Handbook of Responsible Artificial Intelligence. Interdisciplinary Perspectives*, Cambridge University Press, Cambridge, 2022, pp. 31-49.

Infatti, una volta che si dipanano due possibilità di fronte alla macchina, quale delle due verrà scelta in relazione ad un contesto dato?

Tuttavia, credo che il problema più evidente, forse, di entrambe le prospettive (tanto di quella *top-down* quanto di quella *bottom-up*) è che entrambe pretendono di ottenere un *output* unico, ovvero una "one right answer" di dworkiniana memoria[55].

Ma essa è possibile solamente in presenza di un set morale già collaudato in chi deve produrre l'*output*. E qui il problema diventa ancora più serio, perché si presuppone qualcosa che si deve invece costruire.

5. Un approccio maieutico alla intelligenza artificiale

Alcune pagine addietro, ho proposto di seguire, con riguardo al modo con cui l'orizzonte etico umano viene trasferito alla intelligenza artificiale, lo schema tripartito tetica, zetetica e maieutica. Colin Allen, come abbiamo visto, conferma questa intuizione distinguendo, nei sistemi artificiali, un approccio *top-down*, un approccio *bottom-up* ed un approccio ibrido, ovvero la risultante dei due approcci precedenti[56]. La proposta che presento in questo saggio è, invece, diversa. Se manteniamo la metafora turinghiana di confronto tra formazione di una macchina e formazione di un essere umano, il bambino che cresce non solo apprende i principi (*top-down approach*), non solo apprende principi propri attraver-

[55] Come noto, la tesi della *one right answer* è la risposta teorica di Dworkin alle opinioni di Hart rispetto agli *hard cases*. Ciò significa che, per Dworkin, i giudici possono fornire una sola risposta giusta alle questioni giuridiche, anche se si trovano di fronte a casi difficili. Questo perché le decisioni dipendono dal senso morale interiore dei giudici, che garantisce la disponibilità, per l'appunto, di "un'unica risposta giusta". Si veda R. Dworkin, *Taking Rights Seriously*, Harvard University Press, Cambridge, 1977, p. 32.

[56] Allen sostiene a tal riguardo che: "Hybrid approaches pose the additional problems of meshing both diverse philosophies and dissimilar architectures. Genetically acquired propensities, the rediscovery of core values through experience, and the learning of culturally endorsed rules all influence the moral development of a child", in C. Allen, I. Smit, W. Wallach, "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches", in *Ethics and Information Technology*, 7 (2005), n. 3, pp. 149-155. Per un approfondimento del tema si veda anche C. Allen, G. Varner, J. Zinser, "Prolegomena to Any Future Artificial Moral Agent", in *Journal of Experimental and Theoretical Artificial Intelligence*, 12 (2000), n. 3, pp. 251-261.

so l'esperienza dilemmatica (*bottom-up approach*) ma, soprattutto, scopre la forza e la fecondità della opposizione, comprendendo altresì il valore etico della opposizione[57].

Ciò significa che, invece di tentare una mediazione tra i due approcci (*top-down* e *bottom-up*), è essenziale provare a sviluppare un contesto oppositivo, privilegiando la negazione socratica, ovvero la confutazione come strumento etico. Non è questa una idea del tutto nuova od originale: il richiamo alla figura di Socrate ed alla confutazione è utilizzato con successo e sempre più di frequente[58].

Ciò significa che, una volta sviluppate le inferenze verso i modelli di riferimento (ad esempio i due modelli di cui abbiamo discusso nella sezione relativa alla tetica, ovvero i principi deontologici e quelli consequenzialisti), le ragioni di preferibilità di una strada andranno confrontate dialetticamente con le ragioni di preferibilità di quella opposta.

Non solo: verranno anche vagliate le contestazioni ad una e all'altra possibilità, creando così un contesto dialettico, dove agli argomenti che sostengono una posizione (una scelta di comportamento) viene aggiunta la contestazione socratica degli argomenti avversi e viceversa[59]. Come in un contraddittorio processuale[60].

Propongo di chiamare questa fase di formazione della intelligenza artificiale la fase maieutico-socratica.

[57] Si veda anche P. Sommaggio, S. Marchiori, "Moral Dilemmas in the A.I. Era: A New Approach", *Journal of Ethics and Legal Technologies*, 2 (2020), n. 1, pp, 89-102.

[58] P. Dordoni, "Socrate in corsia. Percorsi di senso e di riconoscimento nella Babele della salute", in *La Cà Granda*, XLVII (2006), n. 2-3, pp. 32-37. Per un approfondimento sul tema del dialogo socratico in contesti medici, si veda anche P. Dordoni, "Il metodo socratico di Nelson ed Heckmann e la sua rilevanza in pedagogia medica", in *Tutor*, (2002), n. 3, pp. 113-119; ed infine P. Dordoni, "Il ritorno del metodo socratico", in *Ianus*, (2002), n. 8, pp. 34-42; D. Birnbacher, "The Socratic Method in Teaching Medical Ethics: Potentials and Limitations", in *Medicine Health Care and Pilosophy*, 2 (1999), n. 3, pp. 219-224.

[59] Questo contesto è avvicinabile, per molti aspetti alla struttura logica di un dibattito. Per un approfondimento, rimando a P. Sommaggio, *Le città del dibattito. Prassi e principi di dialogo oppositivo*, Giappichelli, Torino, 2024.

[60] Sul tema vedi P. Sommaggio, *Contraddittorio, Giudizio, Mediazione. La danza del demone mediano*, FrancoAngeli, Milano, 2012.

Le opzioni che si dipanano dal problema etico, ovvero le possibilità che da un certo dilemma si segua la strada (il ragionamento giustificativo – l'algoritmo) che porta al principio A o la strada (il ragionamento giustificativo – l'algoritmo) che porta al principio B, vengono fatte scontrare, in questo modo riconoscendo importanza al momento confutatorio che riprende, come noto, la figura socratica. Dallo scontro, io credo, può emergere ciò che non può essere negato da entrambe le opzioni.

Proprio come in un dialogo platonico, dove Socrate incalza il malcapitato contraddittore per indurlo a fornire le ragioni del proprio agire e le sottopone a critica serrata[61], questo movimento di contrasto è in grado di generare una forza *maieutica*[62].

Ciò perché la forza confutatoria che deriva dalla opposizione indossa simbolicamente la "maschera di Socrate"[63].

La "prova" confutatoria di ciascuna posizione rende evidente se e quanto gli argomenti presentati stiano in relazione con la formazione concreta della macchina.

Grazie a questo modello maieutico, che semplicisticamente abbiamo ridotto al tentativo di giustificazione ed alla prova confutatoria reciproca di due possibilità diverse di comportamento, si è in grado di conoscere quanto i principi (*logoi*) della macchina siano in relazione con il modo in cui ha reagito ad una serie di "esperienze" (*bios*): la macchina, infatti, è costretta a render

[61] Il valore e la forza della confutazione sono presenti anche in quella particolare forma di *counseling* che è la consulenza genetica, e possiamo riscontrare diversi indizi che ascrivono anche questa procedura ad un dialogo oppositivo, un dialogo socratico. Si veda P. Sommaggio, *La consulenza genetica. Nuovi miti, nuovi oracoli, libertà della persona*, FrancoAngeli, Milano, 2010.

[62] Arte della levatrice o maieutica che, come noto, Platone identifica nella forza del contrasto socratico. Si veda Platone, *Teeteto*, 150 b-151 c. La forza maieutica è dunque un contrasto in grado di *generare* nuovi concetti.

[63] Sulla figura di Socrate, nella economia di questo lavoro, mi limito a rimandare alla lettura di G. Reale, *Socrate. Alla scoperta della sapienza umana*, Rizzoli, Milano, 2000; G. Vlastos, *Studi Socratici*, Vita e pensiero, Milano, 2003; vedi anche N. Grimaldi, *Socrate, lo stregone. Il primo guaritore di anime*, Asterios, Trieste, 2008. Questa funzione, oltre che logica, spinge nella condizione di dire il vero, ossia di collegare il ragionamento con l'esperienza. Superando l'opposizione più nota nei confronti dei modelli basati sull'approccio tetico, ovvero *top-down* accusati di essere approcci astratti.

conto della corrispondenza (o meno) tra il proprio *logos* e *bios*. A rendere conto di sé[64]. Ad esercitare la *parresia*[65].

6. La macchina parresiastica

La *parresia* indica, oggi, libertà di pensiero, libertà di parola o indipendenza nei confronti di chi detiene il potere[66]. Ma non è questo il solo significato della espressione: il "dire" del parresiaste (l'*output* della macchina) costituisce infatti un collegamento narrativo (un costrutto linguistico) tra il *suo* "vissuto" esperienziale e le *sue* generalizzazioni valoriali.

Il traguardo parresiastico permette di (anzi costringe a) istituire una relazione tra *logos* (razionalità argomentativa) e *bios*

[64] Anche Foucault si impegna nella spiegazione di questo "dare conto di sé" ponendo in luce che questo è il filo rosso che collega molti dialoghi platonici nei quali il "servizio" socratico consiste proprio in questo: non è un esame di coscienza o una confessione di colpe, ma "ancora una volta, qui, rendere conto della propria vita, del proprio *bios*, non significa fare un racconto di eventi storici che hanno avuto luogo nella propria esistenza, ma piuttosto dimostrare se si è capaci di mettere in luce una relazione tra il discorso razionale, il *logos*, che si è in grado di usare ed il modo in cui si vive". In M. Foucault, *Discorso e verità nella Grecia antica*, Donzelli, Roma, 2005, p. 63.

[65] Secondo Foucault, il rapporto parresiastico costituirebbe, per i greci, un *tipo* di quella connessione che sussiste, in un individuo, tra *logos* (giustificazione razionale) e *bios* (esperienza vissuta) che egli legge in termini di verità: le conferenze americane permettono di seguire gli sviluppi storici di questo concetto-ponte. Il termine *parresia*, come noto, fa la sua prima apparizione, nella letteratura greca, nelle tragedie di Euripide e continua a comparire fino al IV secolo d.C. in alcuni testi patristici. Abbiamo tre forme della parola: la forma nominale *parresia*; la forma verbale *parresiazomai* o *parresiazestai*; e la parola *parresiastes. Parresia* letteralmente, dunque, indicherebbe "dire tutto" da *pan* (tutto) e *rhema* (ciò che viene detto). Il *parresiastes* è colui che esprime la *parresia* (*parresiazomai* o *parresiazestai* è usare la *parresia*), ovvero fa comprendere, grazie ad una spiegazione completa ed esatta, e all'uso di forme ed espressioni il più possibile dirette, la propria opinione. In questo senso, dunque, dire il vero significa intrattenere un certo rapporto con la verità ovvero fare in modo che il proprio pensiero coincida con la propria vita. Sulla storia del termine si veda anche G. Scarpat, *Parrhesia greca, parrhesia cristiana*, Paideia, Brescia, 2001.

[66] Si veda in particolare M. Foucault, *Il governo di sé e degli altri. Corso al Collège de France (1982-1983)*, Feltrinelli, Milano, 2009, in particolare pp. 49, 63, 71 dove il filosofo francese tratta la questione parresiastica con particolare riguardo alla sua relazione con il potere politico.

(esperienza) di modo che, attraverso il materiale di cui si compone il *logos*, possa essere elaborata una *icona* della esperienza relativa al *bios*: una immagine razionale, ovvero narrativa di sé.

Forse una prima forma di abbozzo di "coscienza" della macchina[67].

La maieutica, dunque, potrebbe consentire alla macchina di elaborare una relazione tendenzialmente armonica tra l'esperienza ed i principi: una sorta di esercizio di "mascheramento socratico" che si rivela necessario per iniziare un percorso di disvelamento, forse, della propria identità artificiale[68].

Potremmo assistere, ma qui il condizionale è d'obbligo e il livello di previsione forse troppo elevato, ad un fenomeno che possiede almeno due aspetti. Il primo è che questo procedimento dialettico porti la macchina ad avere "coscienza" di chi è. Nel medesimo tempo, però, la macchina, pur in grado di produrre questa "immagine narrativa", potrebbe accorgersi di una eccedenza rispetto ad essa, ed avvertire l'esigenza di non essere ridotta a quella immagine, la cui presenza è comunque necessaria, poiché essa è anche chi l'ha formata. Scoprendo così la antecedenza della capacità di formulare giustificazioni razionali rispetto anche alla propria posizione. Scoprendo così la possibilità di stabilire anche posizioni differenti. E dato che questa capacità non è a disposizione, essa tutelerebbe la sua libertà più profonda: la libertà da sé, che consente di non ridursi mai ad un semplice oggetto[69].

[67] Questa *icona* si caratterizza per la sua irripetibilità (ovvero per il fatto che dipende da quel particolare contesto di scontro), per la sua istantaneità (ovvero per il fatto che si raggiunge e non si può mantenere indefinitamente) e per la sua funzionalità (ovvero per il fatto di potersi esprimere in una forma comunicabile).

[68] Per un ulteriore approfondimento del concetto di autopoiesi, che in questa sede non può essere sviluppato, rimando a F. Gambardella, *L'animale autopoietico. Antropologia e biologia alla luce del postumano*, Mimesis, Milano, 2010.

[69] Anche per questo siamo debitori della lezione di Foucault quando egli afferma: "Il bios è il cuore della parresia socratica. Per ciò che riguarda Socrate e il filosofo, la relazione *bios-logos* è un'armonia dorica che fonda il ruolo parresiastico di Socrate, e che, al tempo stesso, costituisce il criterio visibile per la sua funzione di *basanos* o pietra di paragone. Per ciò che riguarda l'interlocutore, la relazione *bios-logos* si dischiude quando l'interlo-

Stiamo forse parlando di soggettività dei sistemi di intelligenza artificiale?

Forse sì, ma è ancora troppo presto per poter formulare previsioni di questa natura.

Limitiamoci a riportare qualche esempio di ciò che oggi è tecnicamente possibile.

Un primo esempio potrebbe essere il modello cd. akratico[70], oppure il modello di "agente morale artificiale disobbediente" funzionale a disubbidire gli ordini umani rischiosi o potenzialmente dannosi[71].

In quest'ultimo esempio, ciascuna opzione di comportamento viene strutturata ponendo un algoritmo in opposizione, di modo che ciascuna opzione si comporti da "antagonista" socratico rispetto all'altra, sviluppando le controargomentazioni nei confronti di ciascuna argomentazione e ponendo in opposizione le due scelte, le connessioni inferenziali con i principi di riferimento ed i principi stessi (o le norme). In questo modo si potrebbero trattare i principi di riferimento di quel dilemma etico come vere e proprie "ontologie etiche"[72]. Lo scontro oppositivo

cutore dà un resoconto della propria vita, e la sua armonia è verificata tramite il contatto con Socrate. Poiché possiede, nella sua relazione con la verità, tutte le qualità che è necessario mettere in luce nel suo interlocutore, Socrate può verificare la relazione che l'esistenza dell'interlocutore ha con la verità. Lo scopo di questa attività parresiastica di Socrate, perciò, è di portare l'interlocutore alla scelta di quel tipo di vita (bios) che si porrà in un accordo armonico-dorico con il logos, la virtù, il coraggio e la verità" in M. Foucault, *Discorso e verit*à, cit., p. 66.

[70] S. Bringsjord, N. Sundar, D. Thero, M. Si, "Akratic Robots and the Computational Logic Thereof", in Aa.Vv., *Proceedings of the IEEE International Symposium on Ethics in Engineering, Science, and Technology*, Piscataway, 2014, pp. 1-8.

[71] Vedi G. Briggs *et al.*, "'Sorry, I Can't Do That': Developing Mechanism to Appropriately Reject Directives in Human-Robot Interactions", in *AAAI Fall Symposium Series*, 2015, pp. 32-36.

[72] "In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. In the context of database systems, ontology can be viewed as a

di queste due ontologie, basate sulla enunciazione delle ragioni di preferibilità di ciascuna, e sulle ragioni di rifiuto di quella opposta, consente di ottenere un nuovo artefatto, che porterebbe le ontologie (che a loro volte si modificano ricorsivamente grazie all'esperienza, anche formativa, dello scontro) a costituire un'architettura automodificante. Le premesse per sviluppare un tale approccio ci sono.

Un esempio della forza di un simile modello oppositivo è lo studio dal titolo Generative Adversarial Net[73], dove gli autori propongono un nuovo framework per la stima di modelli generativi attraverso un procedimento oppositivo (*adversarial*), in cui si addestrano simultaneamente due modelli: un modello generativo G ed uno discriminativo/oppositivo D, che stima la probabilità che un campione provenga dai dati di addestramento inseriti dall'uomo piuttosto che creati da G. Questa struttura corrisponde a un gioco minimax a due giocatori e a somma zero[74]. Il modello generativo G può essere considerato analogo a una squadra di falsari, che cerca di produrre valuta falsa e di utiliz-

level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals. Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies; in practice, the languages of ontologies are closer in expressive power to first-order logic than languages used to model databases. For this reason, ontologies are said to be at the "semantic" level, whereas database schema are models of data at the "logical" or "physical" level. Due to their independence from lower-level data models, ontologies are used for integrating heterogeneous databases, enabling interoperability among disparate systems, and specifying interfaces to independent, knowledge-based services. In the technology stack of the Semantic Web standards [1], ontologies are called out as an explicit layer. There are now standard languages and a variety of commercial and open source tools for creating and working with ontologies". T. Gruber, *Ontology*, in L. Liu, M.T. Özsu (eds.), *Encyclopedia of Database Systems*, Springer-Verlag, Zurich, 2008.

[73] I. Goodfellow *et al.*, "Generative Adversarial Networks", in A.A. Chien, *Communications of ACM*, 63 (2020), n. 11, pp. 139-144.

[74] Minimax è il metodo (altri preferiscono chiamarlo algoritmo) che si usa per minimizzare la massima perdita in situazioni di conflitto con soggetti rivali, in cui tutti gli agenti ricercano il guadagno massimo. Il suo sviluppo è dovuto allo sviluppo della teoria dei giochi e si concretizza in giochi detti "a somma zero" dove ciò che vince un giocatore viene perduto dall'altro.

zarla senza essere individuata, mentre il modello discriminativo è analogo alla polizia, che cerca di individuare la valuta falsa. La competizione in questo gioco spinge entrambe le squadre a migliorare i propri metodi fino a rendere indistinguibili le "contraffazioni" dagli articoli autentici.

7. Conclusioni

Nelle pagine precedenti ho presentato alcune nozioni di intelligenza artificiale ed ho cercato di soffermarmi sulle caratteristiche etiche di queste nuove entità.

Ho cercato di mostrare come l'orizzonte etico per la intelligenza artificiale si componga prevalentemente di principi etici che costituiscono un quadro normativo o, meglio, regolativo per le macchine intelligenti, tanto nella progettazione quanto nel loro uso. In seguito, ho cercato di mostrare come sia complesso introdurre i principi e le regole etiche umane all'interno della macchina, poiché questo avviene tramite sistemi costruiti secondo una architettura *top-down* che si presta ad una serie di critiche. Anche una struttura di tipo *bottom-up*, ovvero zetetica, presenta alcuni limiti, ma si fonda su di una mimesi dell'essere umano che "impara" ad agire affrontando esperienze che lo istruiscono anche su dilemmi di tipo morale. Ebbene, proseguendo lungo questo percorso e portando avanti l'equazione addestramento di un essere umano – addestramento di una macchina, ho provato ad inserire nel ragionamento l'esigenza del no, ovvero (per rimanere dentro la metafora) quel passaggio nella vita di un essere umano, dalla fanciullezza alla età matura, in cui ci si imbatte nella forza della opposizione. Opposizione che apre, per la macchina, alla possibilità di scoprire di essere una entità in grado di differenziarsi dalla somma dei principi con i quali è stata progettata. E qui ho provato a gettare le basi per un approccio alla intelligenza artificiale di tipo maieutico, ovvero socratico. A conclusione di questo percorso, ritengo essenziale sottolineare che gli approcci di cui ho parlato sinora (tetico, zetetico, maieutico) non possono essere considerati separatamente, ma devono venire integrati in un sistema, che solo così potrà definirsi un agente artificiale morale sensibile in grado di effettuare scelte eticamente motivate e, soprattutto, di rendere conto

di sé. Se questo significa aprire la strada ad entità dotate di autocoscienza, credo che la forza socratica possa essere il miglior modo per incominciare a discuterne. Ottenendo, in questo modo, un agente artificiale socratico che, attraverso l'opposizione maieutica, sia in grado, forse, di fornire *output* non solo intelligenti, ma anche dotati di coscienza e, perciò, di responsabilità.

Autori

Amato Salvatore	*Ordinario di Filosofia del diritto*, Università degli Studi di Catania
Amato Mangiameli Agata C.	*Ordinaria di Filosofia del diritto*, Università degli Studi di Roma Tor Vergata
Calvano Miriana	*Dottoranda in Informatica e Matematica*, Università degli Studi di Bari Aldo Moro, *Componente* del Progetto FAIR - Future AI Research
Campagnoli Maria Novella	*Associata di Filosofia del diritto*, Università degli Studi di Roma Tor Vergata
Canzio Giovanni	*Primo Presidente Emerito della Corte di Cassazione*
Colcelli Valentina	*Primo Ricercatore*, Istituto di Fisica Applicata - Consiglio Nazionale delle Ricerche (IFAC-CNR)
Curci Antonio	*Dottorando Nazionale in Artificial Intelligence for Society*, Università di Pisa, in co-affiliazione con l'Università degli Studi di Bari Aldo Moro, *Componente* del Progetto FAIR - Future AI Research
D'Avanzo Wanda	*Assegnista in Filosofia del diritto*, Università degli Studi "Magna Graecia" di Catanzaro, *Componente* del Progetto FAIR - Future AI Research
Esposito Andrea	*Assegnista in Informatica*, Università degli Studi di Bari Aldo Moro

Ferrara Santo Davide	*Professore Emerito, Medicina Legale, Olim Presidente* della Scuola di Medicina e Chirurgia, Università degli Studi di Padova, *Past President* della International Academy of Legal Medicine
Goretti Maurizio	*Direttore (CEO) Namex*
Incampo Antonio	*Ordinario di Filosofia del diritto*, Università degli Studi di Bari Aldo Moro, *Componente* del Progetto FAIR - Future AI Research
Lanzilotti Rosa	*Ordinaria di Informatica*, Università degli Studi di Bari Aldo Moro, *Leader* del Work Package 6.1 "Design of SAI systems" del Progetto FAIR - Future AI Research
Lisi Francesca Alessandra	*Associata di Informatica*, Università degli Studi di Bari Aldo Moro, *Leader* del Work Package 6.5 "Acceptability of SAI" del Progetto FAIR - Future AI Research
Loddo Olimpia Giuliana	*Ricercatrice di Filosofia del diritto*, Università degli Studi di Cagliari
Lorini Giuseppe	*Ordinario di Filosofia del diritto*, Università degli Studi di Cagliari
Lupária Donati Luca	*Ordinario di Diritto processuale penale*, Università degli Studi di Milano La Statale
Malerba Donato	*Ordinario di Sistemi di Elaborazioni delle Informazioni*, Università degli Studi di Bari Aldo Moro, *Responsabile scientifico* Spoke 6 Progetto FAIR - Future AI Research
Marra Piero	*Ricercatore di Filosofia del diritto*, Università degli Studi di Bari Aldo Moro, *Componente* del Progetto FAIR - Future AI Research
Moro Paolo	*Ordinario di Filosofia del diritto*, Università degli Studi di Padova

Muscatiello Vincenzo Bruno	*Ordinario di Diritto penale*, Università degli Studi di Bari Aldo Moro
Parente Salvatore Antonello	*Associato di Diritto tributario*, Università degli Studi di Bari Aldo Moro, *Componente* del Progetto FAIR - Future AI Research
Piccinno Antonio	*Associato di Informatica*, Università degli Studi di Bari Aldo Moro, *Componente* del Progetto FAIR - Future AI Research
Pinna Simone	*Assegnista in Filosofia del diritto*, Università degli Studi di Cagliari
Pulito Lorenzo	*Ricercatore di Diritto processuale penale*, Università degli Studi di Bari Aldo Moro, *Componente* del Progetto FAIR - Future AI Research
Quattrocolo Serena	*Ordinario di Diritto processuale penale*, Università degli Studi di Torino
Sommaggio Paolo	*Associato di Filosofia del diritto*, Università degli Studi di Padova
Triggiani Nicola	*Ordinario di Diritto processuale penale*, Università degli Studi di Bari Aldo Moro, *co-Leader* del Work Package 6.5 "Acceptability of SAI" del Progetto FAIR - Future AI Research
Valentini Cristiana	*Ordinario di Diritto processuale penale*, Università degli Studi "G. d'Annunzio" di Chieti-Pescara

Indice dei nomi*

Abdullahi, Aminu, 82n.
Aberdeen, John, 270n.
Abran, Alain, 102n.
Accoto, Cosimo, 379n.
Adamatzky, Andrew, 524n.
Adorno, Theodor [Ludwig] Wiesengrund, 168 e n.
Ahmed-Rengers, Emma, 273n.
Aiello, Marco, 62n.
Akar, Servet, 42n.
Akata, Zeynep, 124 e n.
Alaimo, Cristina, 371n.
Alcalá-Fernández, Jesús, 103n.
Alegre, Susie, 452n.
Alesci, Teresa, 310n.
Alexy, Robert, 260 e n.
Alfrink, Kars, 274n.
Algeri, Lorenzo, 398n.
Allegrezza, Silvia, 178n.
Allen, Colin, 522 e n, 529 e n.
Allen, Gregory C., 70n.
Allen, Layman Edward, 475n.
Allen, Woody [*Allan*, *Stewart Königsberg*], 449.
Allison, Henry E., 164n.
Alonso Moral, Jose Maria, 103n.
Altay, Fatih, 16n.
Amato Mangiameli, Agata Cecilia, 6, 391, 409n, 418n, 424n.
Amato, Salvatore, 6, 369, 472n, 476n.
Ambos, Kai, 248n.
Amidei, Andrea, 474n.
Anagnostopoulos, Marios, 90n.
Anderson, Michael, 512n, 528n.
Anderson, Susan Leigh, 512n, 513n, 528n.
Anglano, Cosimo Filomeno, 298n.
Antonini, Luca, 211n.
Aprile, Ercole, 250n.
Aristotele [latino: *Aristoteles*] di Stagira, 433 e n, 434, 438, 493 e n, 496 e n.
Arnold, Rainer, 37n, 47n, 48n.
Asimov, Isaac, 339 e n, 523.
Augé, Marc, 444n.
Aussó, Susanna, 42n.
Avanzini, Giulia, 237n.
Awad, Edmond, 525n.

* L'indice dei nomi è a cura di Giacomo D'Elia e Leonardo Zingaro.

Azimov, Isaac [*Azìmov*, *Isaàk Jùdovič*], 509 e n.

Baccari, Gian Marco, 272n.
Bailey, James, 95n.
Bajanowski, Thomas, 159n.
Balcerzak, Michal, 266n.
Baldi, Paolo, 106n.
Ball, Terence, 162n.
Ballarini, Adriano, 427n.
Balsamo, Antonio, 134n.
Banerjee, Imon, 47n.
Baquero, Pablo Marcello, 252n, 273n.
Barabino, Paolo, 225n.
Barbado, Alberto, 270n.
Barberis, Mauro, 425n.
Barnes, Geoffrey Carroll, 399n.
Barocas, Solon, 342n.
Baron, Jonathan, 527n.
Barona Vilar, Silvia, 282n.
Barone, Giulia, 238n.
Baroni, Massimiliano, 315n.
Barreto Arrieta, Alejandro, 270n.
Barucci, Andrea, 36n, 37n.
Basciu, Maurizio, 470n, 477n, 501n.
Bates, Tony, 55n.
Battiato, Sebastiano, 286n.
Bauman, Zygmunt, 409n.
Beauchamp, Tom Lamar III, 101n, 510 e n.
Beckers, Anna, 458 e n, 459n, 483n.
Bellamy, Richard, 162n.
Belvini, Lorenzo, 176n, 248n, 252n, 263n, 266n, 307n, 312n, 323n, 324n, 325n, 326n, 327n, 328n, 329n, 330n, 331n, 332n, 337n.
Bender, Emily M., 344n, 345n.
Bene, Teresa, 261n.
Bengio, Yoshua, 119n, 284n, 285n, 420n.
Benjamin, Walter [Bendix Schönflies], 436n, 439.
Benjamins, Richard, 270n.
Bennetot, Adrien, 270n.
Bentham, Jeremy, 410n, 512n, 520 e n, 521n.
Berbaum, Michael L., 42n, 48n.
Bergoglio, Jorge Mario: *vide* Francesco, *papa*
Bernardini, Lorenzo, 309n.
Bernasconi, Alessandro, 334n.
Berti, Enrico, 493n, 496n.
Berto, Francesco, 524n.
Bester, Alfred, 166 e n.
Betti, Emilio, 442.
Bettiol, Giuseppe, 488n, 489n.
Bhaskar, Michael, 370n.
Bhimireddy, Ananth Reddy, 47n.
Bichi, Roberto, 274n.
Billard, David, 93 e n, 94.
Binns, Reuben, 132n, 382n.
Birhane, Abeba, 458 e n.
Birnbacher, Dieter, 530n.
Blaiotta, Rocco, 351n.
Blosch, Yannis, 62n.
Blum, Christian, 523n.

Bohr, Niels Henrik David, 201n.
Bonati, Maurizio, 159n.
Bonfanti, Angelica, 398n.
Bongiovanni, Giorgio, 269n.
Bontempelli, Manfredi, 325n, 334n.
Bonzano, Carlo, 248n.
Borgia, Gianluca, 271n, 308n, 332n.
Bork, Robert Heron, 386n.
Böröcz, István, 90n.
Borsari, Riccardo, 494n.
Bostrom, Nick, 508n.
Braca, Dacia, 275n.
Branting, Luther Karl, 270n.
Bratman, Michael Edward, 109 e n.
Braungarten, René, 102n.
Brayne, Sarah, 398n.
Bricchetti, Renato, 325n.
Briggs, Gordon, 534n.
Brighenti, Marco, 165n.
Bringsjord, Selmer, 534n.
Brochhausen-Delius, Christoph, 37n, 47n.
Broniatowski, David A., 452n.
Brown, Bradford, 270n.
Brownsword, Roger, 453n, 456n, 460 e n.
Brunekreef, Jacob, 102n.
Bruno, Orietta, 333n.
Bruno, Pierangela, 95n.
Buccisano, Andrea, 241n.
Burdick, Mark, 126n.
Burns, John L., 47n.
Burr, Christopher, 28n.
Burrell, Jenna, 126n.
Burzagli, Laura, 47n.
Bygrave, Lee Andrew, 272n.

Cabiale, Andrea, 180n, 248n.
Cadoppi, Alberto, 272n, 309n.
Caianiello, Michele, 181n, 248n, 251n, 277n.
Calamandrei, Piero, 180n.
Caligiuri, Mario, 129n, 211n.
Calimeri, Francesco, 95n.
Calvano, Miriana, 5, 15, 19n, 23n, 25n, 31n, 270n.
Calvo, Rafael A., 27n.
Camaldo, Lucio, 333n, 337n.
Camardi, Carmelita, 320n.
Camon, Alberto, 248n.
Campagnoli, Maria Novella, 6, 391 e n.
Campanale, Domenico, 434n.
Campanella, Susi, 296n.
Campbell, Murray, 118n.
Canato, Maria Carla, 254n.
Cané, Daniele, 213n.
Canestrari, Stefano, 272n, 309n.
Canonico, Massimo, 298n.
Cantero Gamito, Marta, 175n.
Canzio, Giovanni, 5, 139, 171n, 265n, 301n, 316n, 325n, 337n.
Capitta, Anna Maria, 334n.
Caprioli, Francesco, 171n.
Caputo, Matteo, 282n.
Carinci, Andrea, 214n.

Carlizzi, Gaetano, 174n, 337n.
Carloni, Enrico, 233n.
Carlson, Andrew, 166n.
Carnelutti, Francesco, 410n.
Carnevale, Antonio, 98n, 99n, 132n, 429n.
Carpentieri, Loredana, 211n, 216n.
Càrpino, Massimiliano, 271n.
Carr, Nicholas George, 192n, 195n, 197n.
Carratta, Antonio,172n.
Cartabia, Marta [Maria Carla], 152, 182n, 183, 326n, 327n, 495n.
Cashwell, Glyn, 381n, 383n.
Casino, Fran, 90n.
Casolari, Federico, 99n, 259n.
Cassano, Giuseppe, 245n, 255n.
Cassese, Sabino, 385n, 435n.
Cassibba, Fabio, 249n, 251n.
Castaldi, Carmine, 330n.
Castellano, Giovanna, 16n.
Castells, Manuel, 435n.
Castiello, Ciro, 103n.
Castiglia, Maria Enza, 216n.
Catalano, Elena Maria, 281n.
Catellani, Patrizia, 173n.
Cauteruccio, Francesco, 95n.
Cavaceppi, Cecilia, 267n.
Cavalla, Francesco, 470n, 488n, 490n, 494n, 501n.
Cavallone, Bruno, 180n.
Cavini, Sofia, 334n.
Cecchetto, Giovanni, 159n.
Cecchi, Marco, 301n, 305n.
Cecchi, Rosanna, 159n.
Ceci, Michelangelo, 267n.
Cedro, Marco, 212n, 213n, 216n, 217n, 218n, 219n, 220n, 225n, 229n, 236n.
Centorame, Federica, 172n.
Cerdá Alberich, Leonor, 42n.
Cerqua, Federico, 176n, 297n.
Cervantes, José-Antonio, 513n.
Cesari, Claudia, 171n.
Chang, Kai-Min, 166n.
Chang, Wendy, 72n.
Channing Ruff, John, 284n.
Chatila, Raja, 270n.
Chee, Foo Yun, 66n.
Chelo, Andrea, 332n.
Chen, Fang, 275n.
Chesney, Robert, 286n.
Chiao, Vincent, 250n.
Chien, Andrew A., 535n.
Chiereghin, Franco, 491n.
Chierici, Andrea, 511n.
Childress, James Franklin, 101n, 510 e n.
Chiodi, Pietro, 450n.
Ciampi, Costantino, 485n.
Ciavola, Agata, 261n.
Cigno, Mario, 255n.
Cimino, Filippo Alessandro, 212n, 213n, 216n, 217n, 218n, 219n, 220n, 225n, 229n, 236n.
Cippitani, Roberto, 37n, 47n.
Cirone, Enza, 253n.

Citron, Danielle, 286n, 430n.
Clark, Zoë, 42n.
Clarke, Roger, 408n.
Clowes, Robert William, 435n.
Coeckelbergh, Mark, 458 e n.
Cohen, Julie E., 462 e n.
Colacurci, Marco, 308n, 329n, 332n.
Colantonio, Sara, 42n.
Colcelli, Valentina, 5, 33, 37n, 47n.
Comelli, Alberto, 223n.
Comite, Ubaldo, 426n.
Conant, Emily F., 42n, 48n.
Consolo, Giovanni, 220n, 221n.
Contaldo, Alfonso, 211n, 246n, 267n.
Conti, Carlotta, 281n, 301n, 308n, 330n, 337n.
Contissa, Giuseppe, 398n, 450n.
Contrino, Angelo, 214n, 216n, 221n, 222n, 229n, 237n.
Conway, Paul, 527n.
Coraglia, Greta, 101n.
Cordeiro Guerra, Roberto, 213n, 215n, 218n.
Cordero, Franco, 376n.
Cornish, Chloe, 76n.
Coroamă, Vlad C., 62n.
Cortesi, Maria Francesca, 261n.
Costa, Claire D., 270n.
Costabile, Maria Francesca, 270n.
Costantini, Stefania, 90n, 100n.
Courville, Aaron, 284n, 285n, 420n.
Cowls, Josh, 453n, 459n.
Crafa, Silvia, 507n.
Crawford, Kate, 396n, 397n, 422n, 423n, 424n.
Cremades, Javier, 48n.
Cristianini, Nello, 189n, 378n, 419n, 455n, 497n.
Croce, Benedetto, 448n.
Cuadrado-Gallego, Juan J., 102n.
Cummings, Mary L., 132n.
Curci, Antonio, 5, 15, 29n, 270n.
Currao, Elettra, 308n.
Curtotti, Donatella, 304n.
Custer, Bart, 370n.

D'Acquisto, Giuseppe, 484n.
D'Agostino, Francesco, 491n.
d'Alembert, Jean-Baptiste Le Rond, 165 e n.
D'Antonio, Virgilio, 271n.
D'Asaro, Fabio Aurelio, 101n, 106n, 429n.
D'Avanzo, Wanda, 6, 415, 417n.
D'Aversa, Annunziata, 267n.
da Empoli, Giuliano, 389n.
Dalle Mulle, Emmanuel, 425n.
Damaška, Mirjan, 174n.
Danaher, John, 126n, 175n.
Daniele, Marcello, 172n, 177n, 248n, 281n.
Danielson, Peter A., 524n.
Dankof, Steven, 183n.
Danzhu, *principe ereditario*, 202.

Dasaklis, Thomas K., 90n.
Davis, Deborah, 183n.
de Almeida Vau, Maria, 284n.
De Amicis, Gaetano, 251n.
De Anna, Gabriele, 482n.
de Balzac, Honoré [*Balzac, Honoré*], 378 e n.
De Caro, Costanza, 309n.
de Feydeau, Amaury, 77n.
De Martino, Graziella, 267n.
De Minico, Giovanna, 266n, 268n.
De Paolis, Elisa, 351n.
de Vasconcelos Aguiar, João Paulo, 55n.
Deckers, Jan, 37n.
Delfino, Rebecca, 283n.
Dell'Anna, Davide, 29n.
Della Morte, Gabriele, 382n.
Della Ragione, Luca, 326n.
Della Torre, Jacopo, 308n, 313n, 323n, 330n, 332n, 349n.
della Valle, Eugenio, 241n.
Delp, Aaron, 54n.
Demartis, Fabrizio, 308n, 311n, 313n, 319n, 331n, 332n, 335n, 336n, 337n.
Desislavov, Radosvet, 78n.
Desolda, Giuseppe, 21n, 22n, 25n, 28n, 134n, 270n.
di Carpegna Brivio, Elena, 308n.
Di Ciommo, Francesco, 176n.
Di Ciommo, Iole Pia, 269n.
Di Giovine, Ombretta, 274n.
Di Martino, Alessandro, 340n.
Di Matteo, Francesca, 308n, 316n.
Di Nucci, Ezio, 526n.
Di Paolo, Gabriella, 273n, 281n, 308n.
Di Simone, Paola, 308n, 309n.
Di Vizio, Fabio, 330n.
Díaz-Rodríguez, Natalia, 270n.
Dick, Philip K., 393n.
Dickens, Charles John Huffam, 376.
Diddi, Alessandro, 261n, 312n.
Diderot, Denis, 165 e n.
Dijk, Jelle van, 435n, 458 e n.
Dinacci, Filippo, 288n.
Dix, Alan, 20n, 98n.
Dolcini, Emilio, 250n.
Donati, Filippo, 246n, 275n.
Dong, Guozhu, 95n.
Doorn, Neelke, 274n, 275n, 276n.
Dordoni, Paolo, 530n.
Dorigo, Stefano, 213n, 215n, 218n, 226n, 228n, 229n, 230n, 232n, 234n, 235n, 237n, 240n, 243n.
Dostoevskij, Fëdor Michajlovič, 164n.
Dovier, Agostino, 96n.
Drahmann, Annemarie, 370n.
Dressel, Julia, 343n.
Drizin, Steven, 183n.
Dumke, Reiner R., 102n.
Dummett, Michael [Anthony Eardley], 470n.

Dvořák, Antonín [Leopold], 434 e n.
Dworkin, Ronald [Myles], 529n.
Dyoub, Abeer, 98n, 100n, 101n, 106n, 429n.
Dzhumaliev, Daniyar Sardarbekovich, 423n.

Ebers, Martin, 175n.
Ehrenfest, Paul, 201n.
Einstein, Albert, 159 e n, 201n.
Elhalal, Anat, 99n.
Eljechtimi, Ahmed, 77n.
Endicott, Timothy, 386n.
Engelbart, Douglas Carl, 123n.
Enqvist, Lena, 254n, 269n.
Errico, Giulio, 308n.
Esposito, Andrea, 5, 15, 27n, 29n, 270n.
Eubanks, Virginia, 453n, 456 e n, 458 e n.
Euripide [latino: *Euripides*] di Salamina, 532n.

Falato, Fabiana, 351n, 357n.
Falchetti, Alfredo, 215n.
Fallon, Richard Jr., 430n.
Farid, Hany, 343n.
Farina, Massimo, 424n.
Farinas, Gomez, 423n.
Faro, Sebastiano, 180n, 476n.
Farooq, Ayaz, 28n, 102n.
Farri, Francesco, 213n, 214n, 222n, 227n, 230n, 233n, 239n, 240n, 241n, 242n.
Fasano, Gianluca, 402n.
Fasola, Marco, 228n.
Favretto, Donata, 159n.
Fazzalari, Elio, 489n.
Fedele, Andrea, 211n.
Feigenbaum, Edward Albert, 119n.
Feil-Seifer, David, 100n.
Feldman, Julian, 119n.
Felicioni, Paola, 274n.
Fensel, Anna, 93n.
Ferguson, Andrew Guthrie, 178n, 184n, 393n.
Fernández, Alberto, 37n.
Ferrajoli, Luigi, 180n, 381 e n, 489n.
Ferrante, Enzo, 42n.
Ferrara, Santo Davide, 5, 157, 159n.
Ferrarese, Maria Rosaria, 437n, 441n.
Ferraris, Maurizio, 447n.
Ferrarotti, Franco, 163 e n.
Ferretti, Francesca, 258n.
Ferro, Lisa, 270n.
Ferrua, Paolo, 280n, 281n, 296n, 304n.
Feuerbach, Ludwig Andreas, 165n.
Fidopiastis, Cali M., 18n, 97n.
Fimiani, Pasquale, 251n.
Finocchiaro, Giusella, 246n.
Fiorelli, Giulia, 267n, 282n.
Firth, John Rupert, 192.

Floridi, Luciano, 28n, 99n, 110 e n, 112, 114n, 125 e n, 126n, 135 e n, 259n, 419n, 452n, 453n, 457 e n, 459n, 463 e n, 471n, 472n, 474n, 483n, 510n, 516n.
Floyd, Juliet, 482n.
Fonsi, Ariella, 314n, 402n.
Fontana, Peter C., 452n.
Foot, Philippa, 525 e n, 526.
Forti, Gabrio, 282n.
Fosch-Villaronga, Eduard, 370n.
Foucault, Paul-Michel, 179n, 532n, 533n, 534n.
Fox, Kieran, 159n.
Fragasso, Beatrice, 292n, 293n, 295n.
Francesco, *papa*, 167n.
Francioso, Chiara, 221n.
Frangi, Alejandro F., 47n.
Frankish, Keith, 508n.
Frosini, Tommaso Edoardo, 180n, 476n.
Fry, Hannah, 383n.
Fukuyama, Francis, 435.
Fuller, Lon Luvois, 485n.
Fuselli, Stefano, 487n, 491n, 494n, 513n.

Gabriel, Iason, 126n, 132n, 457 e n.
Gaeta, Piero, 251n, 327n.
Gaito, Alfredo, 304n.
Galantini, Novella, 251n, 288n.
Galgani, Benedetta, 181n, 271n, 290n.
Galgano, Francesco, 441n, 442n.
Galimberti, Umberto, 471n.
Galli, Federico, 450n.
Gallin, Catherine, 42n.
Gallo, Franco, 212n, 218n.
Gallo, Giuseppe, 333n.
Gallone, Giovanni, 235n.
Galluccio Mezio, Gaetano, 308n, 320n, 321n, 329n, 335n.
Galsworthy, John, 379 e n.
Gambardella, Fabiana, 533n.
Garapon, Antoine, 171n, 373 e n, 374, 375 e n, 376, 377, 381, 382n, 437 e n, 434n, 476n, 503n.
Garbolino, Paolo, 346n.
Garcia, Francesca, 99n.
Garcia, Salvador, 270n.
Garg, Arti, 61n.
Garlati, Loredana, 184n.
Garrett, Brandon, 183n.
Gatta, Gian Luigi, 181, 182n.
Gawronski, Bertram, 527n.
Gebru, Timnit, 345n.
Gebser, Martin, 91n.
Geierhos, Michaela, 275n.
Gentile, Giovanni, 439.
Georgieva, Konstantina, 102n.
Geraci, Rosa Maria, 211n.
Gerstner, Dominik, 399n.
Ghosal, Amrita, 90n.
Gialuz, Mitja, 173n, 181n, 182n, 268n, 325n, 330n, 332n.
Giannotti, Fosca, 270n.
Gianoncelli, Stefania, 225n.

Giarda, Angelo, 325n, 327n, 330n.
Gigerenzert, Gerd, 438n.
Gil-Lopez, Sergio, 270n.
Gillies, Donald Angus, 496n.
Giovannini, Alessandro, 224n, 225n, 237n.
Giovene, Ambra, 150n.
Gips, James, 521n.
Giudice, Oliver, 286n.
Giuliani, Alessandro, 493n.
Giuliani, Livia, 261n.
Giunchedi, Filippo, 326n, 350n.
Glymour, Clark, 521n.
Godano, Francesco, 450n.
Gödel, Kurt Friedrich, 481 e n.
Gomez, Aidan N., 455n.
Goodfellow, Ian Jeremy, 284n, 285n, 420n, 535n.
Gordon, Thomas, 445n.
Goretti, Maurizio, 5, 51.
Gorrieri, Laura, 429n.
Gottard, Anna, 37n.
Grabherr, Silke, 159n.
Grace, Jamie, 399n.
Gray, David, 430n.
Greco, Chiara, 183n.
Greene, Joshua David, 526n.
Grigsby, Scott S., 18n, 21n, 97n.
Grimaldi, Nicolas, 531n.
Grisham, John Ray Jr., 379n.
Gruber, Tom, 535n.
Grumbach, Stéphane, 178n.
Guarnera, Luca, 286n.
Guidara, Antonio, 216n, 217n, 223n, 224n, 231n, 232n, 244n.
Guido, Elisabetta, 261n.
Guidotti, Dario, 106n.
Guidotti, Riccardo, 132n, 270n.
Gulotta, Guglielmo, 328n.
Guo, Priscilla, 405n.
Gutenberg [*Gensfleisch, Johannes*], 196.

Haghshenas, Kawsar, 62n.
Hahn, Carina A., 452n.
Haidt, Jonathan, 527n.
Hammurabi, *re*, 159.
Han, Byung-chul, 169n.
Harari, Yuval Noah, 192n, 194n.
Hardyns, Wim, 399n.
Harkens, Adam, 273n.
Hart, Herbert Lionel Adolphus, 529n.
Hashimoto, Utaro, 204n.
Hassabis, Demis, 203, 204, 205.
Haugeland, John, 119n.
Hawkinson, John, 55n.
Hayes, Patrick, 521n, 522 e n.
Heaven, Douglas, 419n.
Hegel, Georg Wilhelm Friedrich, 164 e n, 448n.
Heidegger, Martin, 438 e n, 440n.
Heinze, Alexander, 248n.
Helzel, Paola Barbara, 426n.
Henderson, James, 166 e n.
Henin, Clément, 276n.

Herlin-Karnell, Ester, 247n.
Hernández-Orallo, José, 78n.
Herrera, Francisco, 270n.
Hilbert, David, 481n.
Hildebrandt, Mireille, 453 e n, 456n, 460n, 461n, 462.
Hill, Kelly, 64n.
Hinton, Geoffrey Everest, 119n, 192 e n.
Hoane, A. Joseph Jr., 118n.
Hobbes, Thomas, 375.
Hoffman, Donald, 201n.
Hoffman, Reid Garrett, 191n.
Holm, Elizabeth A., 510n.
Holmes, Oliver Wendell Jr., 375.
Horgan, Terence [Edward], 522n.
Horkheimer, Max, 168 e n.
Horton, Richard, 167 e n.
Houck, Theresa, 67n.
Houkes, Wybo, 509n.
Hsu-Shih, Shih, 502n.
Hsu, Feng-hsiung, 118n.
Huang, Aja, 205.
Hubig, Nina Christine, 275n.
Hui, Fan, 205.

Iandiorio, Elisa, 422n.
Iaselli, Michele, 426n.
Iasevoli, Clelia, 261n.
Ienca, Marcello, 509n.
Imtiaz, Nabiha, 100n.
Incampo, Antonio, 6, 11, 433, 481n, 540.
Ingrao, Giuseppe, 241n.
Ippocrate [latino: *Hippocrates*] di Kos, 159.
Ishikawa, Takaki, 159n.
Ivanov, Ivo, 77n.
Iwamoto, Kaoru [*Hon'inbō Kunwa*], 204n.
Izzo, Luigi, 214n.

Jacob, Benoit, 79n.
Jegham, Nidhal, 78n.
Jobin, Anna, 509n.
Johnson, James, 77n.
Jones, Llion, 455n.
Jougleux, Philippe, 272n.
Jullien, François, 165 e n, 166.
Just, Marcel Adam, 166n.

Kahneman, Daniel, 173n, 192n, 346n, 351n, 352n, 353 e n, 354 e n, 355n, 356 e n, 358n, 361n, 362n, 354 e n.
Kaiser, Łukasz, 455n.
Kallinikos, Jannis, 371n, 387n.
Kaminski, Roland, 91n.
Kamiya, George, 62n.
Kamm, Frances Myrna, 521 e n.
Kant, Immanuel, 11, 164 e n, 433, 436n, 438, 449, 450n, 520 e n, 521n.
Kapelańska-Pręgowska, Julia, 266n.
Kaplan, Jerry, 380n.
Karliuk, Maksim, 247n, 253n.
Karnouskos, Stamitis, 429n.

Kasparov, Garri Kimovič [*Vajnštejn*], 120n, 185 e n, 186, 187, 188, 418n.
Kassab, Julie, 76n.
Kaufmann, Benjamin, 91n.
Kehl, Danielle, 405n.
Keller, Ianus, 274n, 275n, 276n.
Kelsen, Hans, 442, 443, 444n.
Kessler, Samuel, 405n.
Keynes, John Maynard, p. 162
Khan, Arshia, 100n.
Khan, Jenna A., 42n, 48n.
Kilani, Mondher, 168n.
Kiltz, Uta, 42n.
Kinsey, Libby, 99n.
Kitano, Hiroaki, 115n.
Kitchin, Robert Michael, 500n.
Kitsara, Irene, 61n.
Kleinberg, Jon, 343n.
Kobau, Pietro, 447n.
Kokott, Juliane, 247n.
Koller, Alexander, 344n.
Komosinski, Maciej, 524n.
Kondo, Toshikazu, 159n.
Kortuem, Gerd, 274n, 275n, 276n.
Kostoris, Roberto E., 490n.
Kostyuk, Alexander, 426n.
Koulu, Riikka, 270n.
Krabbe, Erik Christiaan Willem, 502n.
Krämer, Otto, 411n.
Kuhn, Thomas [Samuel], 496 e n.
Kurzweil, Raymond, 508n.
Kushibar, Kaisar, 42n.

La Rosa, Salvatore, 224n.
Labatut, Benjamín, 185n, 201n, 203n, 204n.
Lambin, Philippe, 42n.
Landgrebe, Jobst, 459n.
Lanzilotti, Rosa, 5, 15, 270n, 540.
Lara, Francisco, 37n.
Laronga, Antonio, 330n.
Larrazabal, Agostina J., 42n.
Larson, Erik John, 438n.
Lasagni, Giulia, 175n, 182n, 273n, 398n,
Lassègue, Jean, 171n, 375n, 377n, 382n, 437 e n, 443n, 476n.
Laudan, Larry, 173n.
Laux, Johann, 24n.
Lazrak, Noussair, 42n.
Le Métayer, Danie, 276n.
Le Poidevin, Olivia, 62n.
Legendre, Pierre, 447 e n.
Legg, Shane, 204.
Leihn, Peter, 275n.
Lekadir, Karim, 42n, 47n.
Leo, Richard A., 183n, 184n.
Leonetti, Michele, 426n.
Lessig, Lawrence, 441n, 460 e n.
Li, Donghan, 60n.
Li, Pengfei, 79n.
Li, Wenlong, 273n.
Liao, Bill, 270n.

Licklider, Joseph [Carl Robnett], 97 e n, 119n, 123n, 480 e n.
Lifschitz, Vladimir, 91n.
Lighthill, [Michael] James, 508n.
Lim, Shaun, 270n.
Lipari, Nicolò, 153n.
Lipton, Zachary Chase, 499n.
Lisi, Francesca Alessandra, 5, 87, 90n, 93n, 95n, 96n, 98n, 99n, 100n, 101n, 106n, 132n, 429n, 540n.
Liu, Ling, 535n.
Liu, Wenguo, 523n.
Locke, John, 492.
Loddo, Olimpia Giuliana, 6, 451 e n, 454n, 464n, 540.
Lodge, Martin, 272n.
Logozzo, Maurizio, 225n, 227n.
Lombardi Vallauri, Luigi, 479n, 481n, 482n, 483n, 485n.
Lombardi, Antonio, 98n, 99n, 132n, 429n.
Longo, Erik, 382n.
Lopez, Rita, 308n, 309n, 315n, 328n, 331n, 402n.
Lorini, Giuseppe, 6, 451 e n, 454n, 462n, 463n, 464 e n, 540.
Loysel, Antoine, 447n.
Lübbe-Wolff, Gertrude, 252n.
Luciani, Massimo, 240n.
Luhmann, Niklas, 375 e n.
Lupária Donati, Luca, 6, 140n, 150n, 171 e n, 174n, 175n, 176n, 177n, 178n, 183n, 281n, 282n, 297n, 301n, 337n, 540.
Lupi, Raffaello, 221n, 225n.
Lupoi, Alberto, 446n.
Luzzati, Claudio Raffaele, 501n.
Lyotard, Jean-François, 470n.

Macchi, Laura, 173n.
Maceratini, Arianna, 423n.
MacLaren, James, 273n.
Maestri, Enrico, 441n, 444n.
Magdalena, Luis, 103n.
Maggio, Paola, 323n.
Maihofer, Werner, 524n.
Malave, Vicente L., 166n.
Malerba, Donato, 5, 109, 129n, 133n, 540.
Malizia, Alessio, 98n.
Mandarà, Elena, 426n.
Manes, Vittorio, 172n, 398n.
Manganini, Chiara, 429n.
Mangiaracina, Annalisa, 308n, 309n, 310n.
Manna, Adelmo, 272n, 309n.
Manoharan, Felix, 83n.
Mantelero, Alessandro, 246n, 271n.
Maoro, Falk, 275n.
Marafioti, Luca, 171n, 172n 176n, 180n, 274n.
Marandola, Antonella, 281n, 301n, 308n, 311n, 326n, 330n, 337n.
Maratea, Marco, 95n.
Marcheselli, Alberto, 213n, 225n, 241n.
Marchesi, Valentina, 325n.

Marchetti, Antonella, 382n.
Marchetti, Barbara, 269n.
Marchiori, Samuela, 525n, 530n.
Marcon, Lucia, 304n.
Marcon, Giuseppe, 304n.
Marello, Enrico, 214n, 216n, 220n, 221n, 229n, 237n, 241n.
Marias, Kostas, 42n.
Marinello, Antonio, 212n, 214n, 216n, 221n, 222n, 229n, 230n, 234n, 235n, 237n.
Markou, Christiana, 272n.
Marra, Piero, 6, 98n, 254n, 267n, 429n, 439n, 447n, 469, 475n, 485n, 540.
Martí-Bonmatí, Luis, 42n.
Martin, Christian, 482n.
Martínez-Plumed, Fernando, 78n.
Martino, Antonio Anselmo, 476n.
Martire, Dario, 269n.
Martis, Marcella, 218n.
Marzaduri, Enrico, 280n, 296n, 304n.
Marzo-Ortega, Helena, 42n.
Mashaw, Jerry, 430n.
Mason, Robert A., 166n.
Mastroiacovo, Valeria, 211n, 225n, 238n, 241n.
Matarić, Maja J., 100n.
Mathews, Jud, 248n, 252n.
Mazza, Oliviero, 262n, 503n.
Mazzotta, Francesco, 231n.
McCarthy, John, 455n, 473n, 506n, 522 e n.
McCloskey, Jim, 379n.
McGeever, Jamie, 61n.
McIntosh-Smith, Simon, 54n.
McMillan-Major, Angelina, 345n.
McPeak, Agnieszka, 283n.
McQuillan, Dan, 458 e n.
Meltzer, Bernard, 522n.
Mencar, Corrado, 103n.
Mendoza, Isak, 272n.
Mendoza, Roberto, 304n.
Menski, Martin, 77n.
Mercuri, Giuseppe, 228n.
Mérő, László, 418n.
Meuwese, Anne, 370n.
Mhalla, Asma, 387n.
Michie, Donald, 522n.
Miglio, Gianfranco, 436n.
Miletti, Marco Nicola, 184n.
Milioto, Paola, 213n.
Militello, Vincenzo, 305n.
Mill, John Stuart, 520 e n.
Milone, Diego H., 42n.
Mingardo, Letizia, 370n, 445n, 504n.
Minsky, Marvin Lee, 340 e n, 455n, 473n, 506n.
Mirza, Mehdi, 284n, 285n.
Misselhorn, Catrin, 528 e n.
Mitcheff, Mahsa, 100n.
Mitchell, Melanie, 455n.
Mitchell, Tom Michael, 166n.

Mittelstadt, Brent, 36n, 40n, 43n, 45n, 46n, 49n.
Mobilio, Giuseppe, 308n, 309n, 311n, 329n, 331n, 332n, 333n.
Mökander, Jakob, 23n.
Molina, Daniel, 270n.
Möller, Kai, 253n.
Monreale, Anna, 270n.
Montagna, Mariangela, 360n.
Montanari, Angelo, 96n, 216n.
Montesquieu [*de Secondat*, *Charles-Louis*], *barone*, 377n, 385 e n.
Monti, Andrea, 422n.
Montisci, Massimo, 159n.
Mooney, Carl H., 94n.
Moor, James Haller, 512n.
Morace Pinelli, Arnaldo, 211n.
Moreira, Catarina, 275n.
Moreno-Lax, Violeta, 247n.
Morini, Francesca, 437n.
Morley, Jessica, 99n.
Moro, Paolo, 7, 435n, 441n, 445n, 446n, 487 e n, 491n, 495n, 498n, 499n, 502n, 503n, 506n, 507n, 509n, 511n, 513n, 540.
Moroni, Stefano, 454n, 462n, 463n, 464n.
Moscovici, Serge [*Moscovici*, *Srul Herş*], 198n, 201n.
Mosier, Kathleen L., 126n.
Mullainathan, Sendhil, 343n, 453n.
Muscatiello, Vincenzo Bruno, 6, 185, 541.
Myhand, Taurus, 284n.

Naarttijärvi, Markus, 254n, 269n.
Nardocci, Costanza, 333n.
Năstasă, Anamaria, 29n.
Naumov, Maxim, 54n.
Navarro-Compán, Victoria, 42n.
Neals, Julien Xavier, 380.
Neeuvonen, Päivi Johanna, 247n.
Negri, Daniele, 172n, 178n, 248n, 249n, 251n.
Nellis, Stephen, 66n.
Neufeld, Peter, 183n.
Neuwirth, Rostam, 427n.
Newell, Allen, 119n.
Nicolicchia, Fabio, 172n, 248n.
Nieto, Nicolás, 42n.
Nietzsche, Friedrich [Wilhelm], 165n, 447 e n, 473.
Nieva-Fenoll, Jordi, 173n.
Nikiphorou, Elena, 42n.
Nilsson, Nils John, 418n, 455n.
Nocerino, Wanda, 332n.
Nodelman, Uri, 522n.
Norman, Donald Arthur, 463n.
Norvig, Peter, 20n, 113n, 418n, 419n, 472n.
Novelli, Claudio, 99n, 211n, 259n.

O'Neill, Cathy, 396n.

Obermeyer, Ziad, 453n.
Olivieri, Raffaele, 90n, 92n.
Orban, Cosimo, 436n.
Orlandi, Renzo, 249n, 253n, 297n.
Orlandini, Andrea, 96n.
Ortoleva, Maria Grazia, 220n, 221n, 228n, 244n.
Osuala, Richard, 42n.
Oswald, Marion, 399n.
Otranto, Piergiuseppe, 237n.
Ozair, Sherjil, 284n, 285n.
Ozaki, Ana, 93n.
Özsu, Tamer Mehmet, 535n.

Padova, Pierangelo, 309n.
Pagallo, Ugo, 24n, 290n.
Palatkina, Anna, 275n.
Pallante, Christian, 309n, 329n.
Palmiotto, Francesca, 175n.
Palumbo, Giovambattista, 216n, 218n, 226n, 228n.
Pandolfo, Laura, 106n.
Panenkov, Dmitry, 70n.
Paolozzi, Giovanni, 171n.
Paolucci, Federica, 246n, 309n, 315n.
Papa, Michele, 272n, 309n.
Papanikolaou, Nickolas, 42n.
Paparella, Franco, 213n, 216n, 217n, 222n, 223n, 225n, 228n, 233n, 234n, 235n, 236n, 237n, 238n, 241n, 244n.
Parente, Salvatore Antonello, 6, 209, 212n, 213n, 217n, 222n, 238n, 242n, 541.
Pariser, Eli, 397n.
Parlato, Lucia, 308n, 309n.
Parmar, Niki, 455n.
Parodi, Cesare, 322n, 330n, 398n.
Parodi, Luigi, 265n.
Pascual, Manuel G., 79n.
Pasquadibisceglie, Vincenzo, 133n.
Pasquale, Frank, 234n, 376n, 377n, 380n, 381 e n, 383n, 441n, 458 e n.
Pastor, Daniel, 172n.
Patsakis, Constantinos, 90n.
Paulesu, Pier Paolo, 325n, 490n.
Pavlidis, Georgios, 29n.
Pecchioli, Giulia, 274n.
Pedreschi, Dino, 270n.
Pellicani, Antonio, 267n.
Perilli, Lorenzo, 193n, 200n.
Perrone, Daria, 274n.
Perrone, Leonardo218n.
Peruginelli, Ginevra, 180n, 476n, 479n.
Peterson, Victoria, 42n.
Petrelli, Francesco, 150n.
Pfaff, Mark, 270n.
Pfeifer, Craig, 270n.
Pfeiffer, Heidi, 159n.
Phillips, P. Jonathon, 452n.
Piasere, Leonardo, 168n.
Picardi, Antonietta, 327n.

Piccinno, Antonio, 5, 15, 270n, 541.
Piccolini, Michele, 499n.
Pietrocarlo, Elisabetta, 393n.
Pin, Andrea, 3123n, 511n.
Pinelli, Cesare, 316n.
Pini, Irene, 227n.
Pinna, Simone, 6, 451 e n, 541.
Pino, Giorgio, 252n, 386n.
Pio, Gianvito, 267n.
Pirandello, Luigi, 375n.
Piselli, Riccardo, 273n.
Pittiruti, Marco, 172n.
Platone [latino: *Plato*] di Atene, 165, 377, 491n, 531n.
Poletti, Dianora, 271n.
Pollicino, Oreste, 246n.
Polosukhin, Illia, 455n.
Pontillo, Miriam, 215n, 228n.
Popper, Karl [Raimund], 337, 434 e n, 521 e n.
Porras Perez, Antonio Reyes, 47n.
Postema, Gerald, 269n.
Pouget-Abadie, Jean, 284n, 285n.
Pound, Nathan Roscoe, 382 e n.
Powers, Brian, 453n.
Prastitou, Thalia, 272n.
Pressacco, Luca, 273n, 308n.
Primiero, Giuseppe, 101n, 429n.
Protagora [latino: *Protagoras*] di Abdera, 167n.
Przybocki, Mark A., 452n.
Pugliese, Antonio, 181n.
Puglisi, Riccardo, 210n.
Pulina, Luca, 106n.
Pulitanò, Domenico, 249n, 250n.
Pulito, Lorenzo, 6, 98n, 245, 267n, 276n, 429n, 541.
Punzi, Antonio, 474n.
Purpura, Andrea, 218n, 227n.

Qic, Jingfei, 423n.
Quaranta, Dario, 351n.
Quattrocchi, Andrea, 223n, 224n.
Quattrociocchi, Walter, 198n.
Quattrocolo, Serena, 6, 181n, 279, 281n, 287n, 289n, 290n, 298n, 301n, 302n, 316n, 318n, 319n, 320n, 541.
Quensel, Bernhard K., 162n.

Raaijmakers, Stephan, 16n.
Rackow, Peter, 248n.
Raffiotta, Edoardo Carlo, 315n.
Raghavan, Manish, 343n.
Ragucci, Gaetano, 215n, 228n, 232n, 234n, 236n, 237n, 243n.
Rakowski, Eric, 527 e n.
Ramsey, William M., 508n.
Rauscher, Garth H., 42n, 48n.
Rawls, John [Bordley], 126n, 443 e n.
Reale, Giovanni, 433n, 493n, 531n.
Recchia, Nicola, 250n.
Reggio, Federico, 494n.
Reidenberg, Joel R., 441n.

Reiling, A. Dory, 380n.
Resta, Giorgio, 181n, 246n.
Restrepo Amariles, David, 252n, 273n.
Ricci, Annarita, 272n.
Ricci, Concetta, 216n.
Ricci, Stefano, 477n.
Riccio, Giovanni Maria, 246n, 271n.
Richardson, Rashida, 396n.
Ricotti, Giacomo, 216n.
Rinaldi, Giovanni, 435n.
Rinaldi, Luca, 269n.
Rinceanu, Johanna, 283n.
Riondato, Silvio, 488n.
Riva, Giuseppe, 382n.
Rivello, Pier Paolo, 304n.
Rizzo, Tania, 322n, 330n.
Roach, Matt, 98n.
Robbins, Scott, 510n.
Roberts, John Glover Jr., 370.
Rochester, Nathaniel, 455n, 473n, 506n.
Roddick, John F., 94n.
Rodrigues, Luano, 284n.
Roman, Dumitru, 93n.
Romano, Bartolomeo, 326n.
Romano, Bruno, 440n, 446n, 448n, 477n.
Rombi, Natalia, 296n.
Romeo, Francesco, 427n, 471n, 473n., 479n, 480n.
Ross, Alf, 412.
Rossetti, Andrea, 477n.
Rossi, Gabriele, 524n.
Rotolo, Antonino, 99n, 259n, 269n.
Rousseau, Jean-Jacques, 492.
Ruffolo, Ugo, 112n, 172n, 237n, 274n, 398n, 474n, 509n.
Ruggieri, Salvatore, 270n.
Ruggiero, Rosa Anna, 252n.
Rumiati, Rino, 173n.
Rummens, Anneleen, 399n.
Rumore, Paola, 434n.
Russell, Bertrand [Arthur William], 201n.
Russell, Chris, 36n, 40n, 43n, 45n, 46n, 49n.
Russell, Stuart Jonathan, 20n, 113n, 194n, 418n, 472n.
Ryberg, Jesper, 262, 268n.

Sacchetto, Camillo, 240n.
Sacchetto, Ernestina, 308n, 320n, 336n, 337n.
Saccomani, Jacopo, 273n.
Sachoulidou, Athina, 258n.
Saitta, Antonio, 422n.
Salahuddin, Zohaib, 42n.
Salanitro, Guido, 214n, 215n, 223n, 230n, 231n.
Sallantin, Jean, 150n.
Salvati, Adriana, 242n.
Salvini, Omar, 215n.
Sammartino, Salvatore, 237n.
Sancino, Alessandro, 308n.
Sanders, John W., 15 e n.
Sandle, Paul, 84n.
Sandoval, MariaPaz, 284n.

Santosuosso, Amedeo, 271n, 371n.
Sapolsky, Robert Maurice, 201n, 202n.
Saponaro, Luisella, 309n, 335n.
Sarra, Claudio, 435n, 441n, 445n, 446n, 508n.
Sartor, Giovanni, 269n, 271n, 371n, 398n, 421n, 450n, 501n.
Sather, Richard Norman III, 100n.
Satta, Salvatore, 373 e n.
Satyanarayan, Arvind, 117n, 119n, 121n.
Sbriccoli, Mario, 180n, 184n.
Scalfati, Adolfo, 265n, 308n, 316n, 330n, 402n.
Scarpat, Giuseppe, 532n.
Scella, Andrea, 288n.
Schaub, Torsten, 91n.
Schauer, Frederick, 461.
Schiera, Pierangelo, 436n.
Schmitt, Carl, 434, 436 e n.
Schmorrow, Dylan D., 18n, 97n.
Schultz, Jason Michael, 396n, 422n, 423n.
Scimè, Amalia Serena, 255n.
Scomparin, Laura, 248n.
Sedol, Lee [*Se-dol, Lee*], 185, 202, 203 e n, 205, 206.
Selbst, Andrew D., 342n.
Selicato, Gianluca, 225n.
Sellaroli, Valentina, 398n.
Šepec, Miha, 248n.
Serres, Michel, 164n.
Setz, Brian, 62n.
Severino, Emanuele, 471n.
Severino, Paola, 392n.
Shanahan, Murray Patrick, 522n.
Shannon, Claude Elwood, 455n, 473n, 495 e n, 506n.
Shavit, Yonadav, 114n.
Shazeer, Noam, 455n.
Shen, Hong, 56n.
Shilov, Anton, 65n.
Shinkareva, Svetlana V., 166n.
Shmitchell, Shmargaret, 345n.
Shneiderman, Ben, 18n, 20n, 21n, 28n, 128n.
Shokry, Dina, 159n.
Si, Mei, 534n.
Sibony, Olivier, 346n, 352n, 355n, 358n, 361n.
Sica, Salvatore, 271n.
Signorato, Silvia, 172n, 500n.
Signorile, Onofrio, 215n.
Simon, Herbert Alexander, 119n.
Simoncini, Andrea, 237n, 550n.
Siniscalchi, Guglielmo, 446n, 477n.
Skitka, Linda J., 126n.
Slovic, Paul, 352n.
Smit, Iva, 511n, 522n, 529n.
Smith, Barry, 459n.
Smuha, Nathalie Alisa, 273n.
Snow, Charles Percy, 165 e n.
Sobotta, Christoph, 247n.
Socci Natali, Fiorenza, 485n.
Socrate [latino: *Socrates*] di Atene, 530, 531 e n, 533n, 534n.

Solaas, John, 284n.
Solanas, Agusti, 90n.
Solum, Lawrence Byard, 384n.
Somalvico, Marco, 506n.
Sommaggio, Paolo, 7, 435n, 511, 523n, 525n, 530n, 531n, 541.
Soni, Aditya, 66n.
Soylu, Ahmet, 94n.
Spangher, Giorgio, 280n, 296n, 304n, 325n, 327n, 330n, 489n.
Spathoulas, George, 90n.
Spena, Alessandro, 305n.
Spinello, Francesco, 230n.
Sroka, Tomasz, 266n, 267n, 271n.
Stanley, Lee E., 502n.
Starr, Sonja B., 405n.
Stephenson, Randall, 283n.
Sterlicchio, Gioacchino, 93n, 95n, 96n.
Stettinger, Georg, 24n.
Stock, Oliviero, 472.
Stone Sweet, Alec, 248n, 252n.
Stone, Peter, 506n.
Suksi, Markku, 254n.
Suleyman, Mustafa, 204, 370n.
Sundar Govindarajulu, Naveen, 534n.
Sunstein, Cass Robert, 166 e n, 184n, 346n, 352n, 355n, 358n, 361n, 463 e n.
Susskind, Richard Eric, 375 e n, 376 e n, 378 e n, 503n.
Suweis, Samir, 499n.
Synodinou, Tatiana-Eleni, 272n.
Szezeciniarz, Jean-Jacques, 150n.

Tabasco, Giuseppem 248n, 330n.
Tabik, Siham, 270n.
Taddei Elmi, Giancarlo, 211n, 246n, 267n, 485n.
Taddeo, Mariarosaria, 99n, 259n, 483n.
Tagliabue, Jacopo, 524n.
Tagliagambe, Silvano, 164n.
Tallberg, Jonas, 24n.
Taroni, Franco, 175n.
Taruffo, Michele, 282n, 301n.
Tashea, Jason, 383n.
Taslitz, Andrew, 183n.
Taylor, Isaac, 272n, 274n.
Teresi, Carolina, 316n.
Terracina, Giorgio, 95n.
Tesoriero, Salvatore, 250n.
Tessitore, Giovanni, 309n, 310n.
Testolin, Alberto, 499n.
Teubner, Gunther, 458 e n, 459n, 483n.
Thaler, Richard Herring, 166 e n, 194n, 463 e n.
Thero, Dan, 534n.
Thomson, Judith Jarvis, 525 e n, 526 e n.
Timmons, Mark, 522n.
Tinelli, Giuseppe, 225n.
Tng, Serene, 77n.
Todescan, Franco, 488n, 494n.

Tommaso d'Aquino [latino: *Thomas Aquinas*], *santo*, 167, 374.
Tomo, Alessia, 232n.
Tonini, Paolo, 281n, 301n, 304n.
Torre, Marco, 258n, 268n, 274n, 281n, 309n, 316n, 319n, 325n, 330n.
Tosi, Loris, 211n.
Trapella, Francesco, 360n.
Trautteur, Giuseppe, 199n.
Traversi, Alessandro, 424n.
Treves, Renato Samuele, 444n.
Triggiani, Nicola, 3, 6, 11, 323n, 328n, 330n, 334n, 361n, 541.
Tripodi, Andrea Francesco, 250n.
Tripodi, Enzo Maria, 245n, 255n.
Troiano, Onofrio, 176n.
Troisi, Paolo, 271n.
Tsakou, Gianna, 42n.
Tsamados, Andreas, 483n.
Tsiknakis, Manolis, 42n.
Tuosto, Ersilia Maria, 328n.
Turchi, Tommaso, 98n.
Turco, Elga, 261n.
Turing, Alan Mathison, 190, 418, 454 e n, 471 e n, 472, 481n, 482n, 523 e n, 528.
Turini, Franco, 270n.
Turley, Jonathan, 191n.
Tuzet, Giovanni, 337n.
Tversky, Amos, 351n, 352n, 254n.
Ubertis, Giulio, 150n, 249n, 252n, 268n, 269n, 274n, 504n.
Unger, Peter Kenneth, 525 e n.
Uricchio, Antonio Felice, 211n, 212n, 213n, 231n, 242n.
Urwin, Sheena, 399n.
Uszkoreit, Jakob, 455n.

Valentini, Cristiana, 6, 269n, 541.
Vallati, Mauro, 95n.
Valli, Roberto V.O., 309n.
Vallor, Shannon, 511n.
van Dijck, Gijs, 261n.
van Dijk, Jan Albertus George Marie, 435n, 458 e n.
Varner, Gary, 529n.
Varraso, Gianluca, 282n, 327n.
Vasta, Valentina, 309n, 320n.
Vaswani, Ashish, 455n.
Vattimo, Gianteresio [*vel* Gianni], 470n.
Vayena, Effy, 509n.
Vegetti, Mario, 166n.
Vennemann, Marielle, 159.
Vermaas, Pieter E., 509n.
Verrigni, Caterina, 212n, 217n, 227n, 229n, 230n, 232n, 238n, 244n.
Versiglioni, Marco, 223n.
Vessio, Gennaro, 282n, 283n, 284n.
Vicini, Antonella, 198n.
Viehweg, Theodor, 524 e n.
Viganò, Francesco, 250n, 251n.

Villa, Vittorio, 386n.
Vincent, Matt, 67n.
Vitale, Alessandro, 419n.
Vlastos, Gregory, 531n.
Voeneky, Silja, 528n.
Vogeli, Christine, 453n.
Vogenauer, Stefan, 247n.
Vuille, Joelle, 175n.

Wachter, Sandra, 36n, 40n, 43n, 45n, 46n, 49n.
Wagner, Wilhelm Richard, 165.
Wallach, Wendell, 511n, 513n, 522n, 529n.
Walsh, Dylan Marcus, 275n.
Walton, Douglas Neil, 269n, 502n.
Wang, Hongyu, 60n.
Warde-Farley, David, 284n, 285n.
Warren, Tom, 65n.
Washington, Anne L., 405n.
Weatherill, Stephen, 247n.
Weber, Max [Maximilian Carl Emil], 162, 470n.
Weiss, Brandy, 270n.
Weizenbaum, Joseph, 498n.
West, Henry R., 520n.
Wiener, Norbert, 495 e n.
Wilson, Ben, 98n.
Winfield, Alan F.T., 523 e n.
Wittgenstein, Ludwig [Josef Johann], 482n.
Woodcock, Taylor Kate, 267n.
Woodruff, Henry C., 42n.
Wooldridge, Michael, 114n.
Wörsdörfer, Manuel, 22n.

Xu, Bing, 284n, 285n.

Yao [*Tang Yao*/*Yaotang-shi*], *imperatore*, 202.
Yeung, Karen, 272n, 273n, 453 e n, 456 e n, 460n.

Zaccaria, Giuseppe, 166n.
Zacchè, Francesco, 248n, 296n.
Zadeh, Lotfi Aliasker, 502n.
Zalta, Edward Nouri, 522n.
Zanuso, Francesca, 494n, 511n, 520n.
Zarsky, Tal, 452n.
Zekos, Georgios I., 370n, 389n.
Zeleznikow, John, 504n.
Zellini, Paolo, 498n.
Zeno-Zencovich, Vincenzo, 171n, 176n, 178n, 180n, 181n, 476n.
Zhanga, Qian, 423n.
Zhao, Jingchen, 423n.
Zhu, Xuanbing, 60n.
Ziccardi, Giovanni, 281n.
Ziegler, Katja S., 247n.
Zimmermann, Hans-Jürgen, 102n.
Zinser, Jason, 529n.
Ziosi, Marta, 99n.
Zonaro, Marco, 332n
Zuboff, Shoshana, 456 e n.
Zuccarello, Alessandro, 222n, 223n.

Unità del sapere giuridico

Quaderni di scienze penalistiche e filosofico-giuridiche

Serie I. Editore A. Giuffré, Milano

1. Vincenzo Garofoli (a cura di), *L'unità del sapere giuridico tra diritto penale e processo.* Atti del Convegno (Bari, 21-22 maggio 2004, Facoltà di Giurisprudenza), 2005.
2. Angiola Filipponio (a cura di), *Verità e normatività*, 2005.
3. Vincenzo Garofoli (a cura di), *Unità del sapere giuridico ed eterogeneità dei saperi del giudice*, 2005.
4. Vincenzo Garofoli (a cura di), *Problematiche tradizionali e incaute innovazioni legislative*, 2006.
5. Giuseppe Spagnolo (a cura di), *La responsabilità da reato degli enti collettivi. Cinque anni di applicazione del d.lgs. 8 giugno 2001, n. 231.* Atti del Convegno (Bari, 26-27 maggio 2006-2007, Facoltà di Giurisprudenza), 2007.
6. Vincenzo Garofoli (a cura di), *La sentenza della Corte Costituzionale 6 febbraio 2007 n. 26: un energico richiamo al metodo della giurisdizione.* Atti del Convegno (Trani, 2-3 febbraio 2007, Hotel S. Paolo al Convento), 2007.
7. Miranda Zerlotin, *Diritto originario in Erik Wolf*, (con la traduzione italiana dell'opera di Erik Wolf, *Recht des Nächsten. Ein rechtsthelogischer Entwurf*), 2008.
8. Vincenzo Garofoli (a cura di), *Unità del sapere giuridico e poliformismo normativo*, 2008.
9. Vincenzo Garofoli (a cura di), *L'udienza preliminare e il contenimento dei tempi processuali*, 2008.
10. Vincenzo Garofoli (a cura di), *L'azione penale tra obbligatorietà e discrezionalità.* Atti del Convegno (Bari, 29 novembre 2008, Villa Romanazzi Carducci), 2009.
11. Angiola Filipponio/Aldo Regina (a cura di), *In ricordo di Aldo Moro.* Atti del Convegno (Bari, 20 giugno 2008, Facoltà di Giurisprudenza), 2010.
12. Angiola Filipponio/Vincenzo Garofoli (a cura di), *In ricordo di Franco Cipriani*, 2010.
13. Vincenzo Garofoli/Antonio Incampo (a cura di), *Verità e processo penale*, 2012.

Serie II. Editore Cacucci, Bari

14. Antonio lncampo/Wojciech Żełaniec (eds.), *Universality of Punishment*, 2015.

Serie III. Editore Cacucci, Bari

15. Antonio Incampo/Adolfo Scalfati (a cura di), *Giudizio penale e ragionevole dubbio*, 2017.
16. Lucio Franzese/Antonio Incampo (a cura di), *Potere e libertà al tempo delle emergenze*/Power and Freedom in the Time of Emergencies, 2021.
17. Piero Marra, *Per una moralità procedurale del diritto. Considerazioni attuali a partire da Lon L. Fuller*, 2022.
18. Domenico Siciliano, *Della violenza nel diritto penale fascista. Il caso della legittima difesa a tutela del patrimonio*, 2023.
19. Antonio Incampo/Maria Antonella Pasculli (a cura di), *Partecipazione, rappresentanza e crimini di genere. L'obiettivo 5 dell'Agenda 2030 delle Nazioni Unite,* 2024.
20. Leonardo Di Carlo, *La validità tra identificazione, applicatività e concetto di diritto*, 2024.
21. Maurizio Sozio, *Dall'opinione pubblica all'infosfera. Linguaggio, diritto e pubblicità in J. Habermas*, 2025.
22. Torquato G. Tasso, *Rapsodia del Diritto. Percorsi diacronici di Filosofia e Teoria Generale del Diritto,* 2026.
23. Maria Novella Campagnoli/Piero Marra (a cura di), *Intelligenza artificiale e metamorfosi del diritto*/Artificial Intelligence and Metamorphosis of Law, 2026.
24. Antonio Incampo/Nicola Triggiani (a cura di), *Giudizio penale e intelligenza artificiale. Una riflessione sistemica*, 2026.