

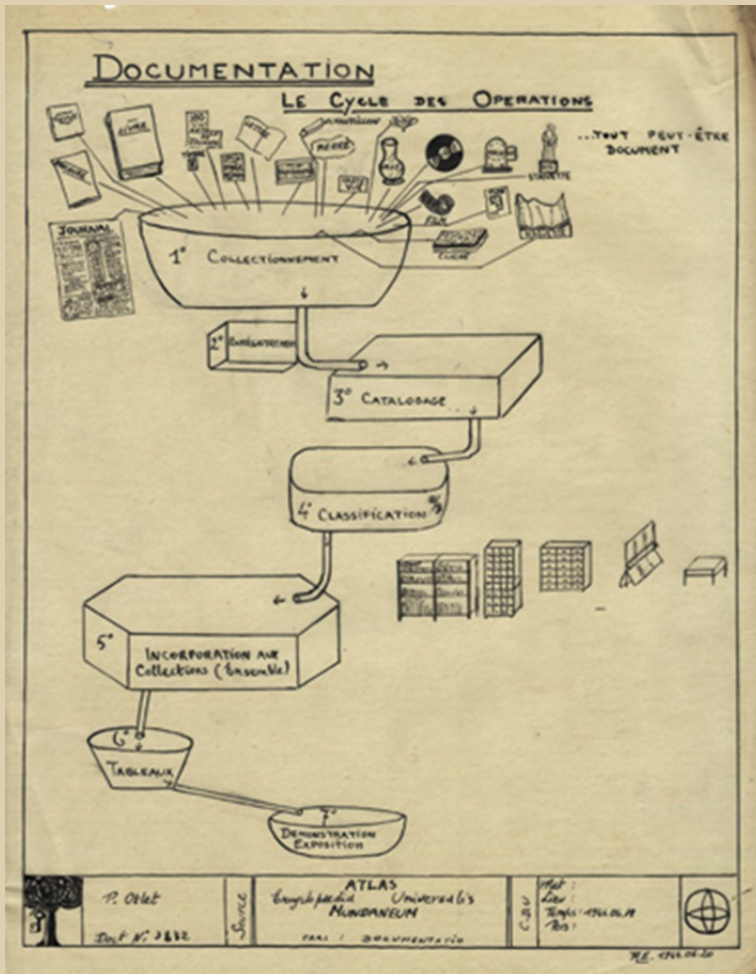
AIDa informazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 1-2

ANNO 42

GENNAIO-GIUGNO 2024



AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

Fondata nel 1983 da Paolo Bisogno

Proprietario della rivista:

Università della Calabria

Direttore Scientifico:

Roberto Guarasci, *Università della Calabria*

Direttore Responsabile:

Fabrizia Flavia Sernia

Comitato scientifico:

Anna Rovella, *Università della Calabria*;

Maria Guercio, *Sapienza Università di Roma*;

Giovanni Adamo, *Consiglio Nazionale delle Ricerche* †;

Claudio Gnoli, *Università degli Studi di Pavia*;

Ferruccio Diozzi, *Centro Italiano Ricerche Aerospaziali*;

Gino Roncaglia, *Università della Toscana*;

Laurence Favier, *Université Charles-de-Gaulle Lille 3*;

Madjid Ihadjadene, *Université Vincennes-Saint-Denis Paris 8*;

Maria Mirabelli, *Università della Calabria*;

Agustín Vivas Moreno, *Universidad de Extremadura*;

Douglas Tudhope, *University of South Wales*;

Christian Galinski, *International Information Centre for Terminology*;

Béatrice Daille, *Université de Nantes*;

Alexander Murzaku, *College of Saint Elizabeth, USA*;

Federico Valacchi, *Università di Macerata*.

Comitato di redazione:

Antonietta Folino, *Università della Calabria*;

Erika Pasceri, *Università della Calabria*;

Maria Taverniti, *Consiglio Nazionale delle Ricerche*;

Maria Teresa Chiaravallotti, *Consiglio Nazionale delle Ricerche*;

Assunta Caruso, *Università della Calabria*;

Claudia Lanza, *Università della Calabria*.

Segreteria di Redazione:

Valeria Rovella, *Università della Calabria*

Editrice: Cacucci Editore S.a.s.

Via D. Nicolai, 39 – 70122 Bari (BA)

www.cacuccieditore.it

e-mail: riviste@cacuccieditore.it

Telefono 080/5214220

AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti le Scienze dell'Informazione, la Documentazione, la Gestione Documentale e l'Organizzazione della Conoscenza. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi possono riguardare topics quali Documentazione, Scienze dell'informazione e della comunicazione, Scienze del testo e del documento, Organizzazione e Gestione della conoscenza, Terminologia, Statistica testuale e Linguistica computazionale e possono illustrare studi sperimentali in domini specialistici, casi di studio, aspetti e risultati metodologici conseguiti in attività di ricerca applicata, presentazioni dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Settore 11/A4 e censita per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali, così come dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) che la annovera tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE – Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; Letteratura Professionale Italiana – Associazione Italiana Biblioteche; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Summon™ – by SerialsSolutions; Ulrich's; UniCat – Union Catalogue of Belgian Libraries; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti dal comitato scientifico sono inviati in forma anonima a due referee, selezionati sulla base della loro comprovata esperienza nei topics specifici del contributo in valutazione.

AIDAinformazioni

Anno 42

N. 1-2 – gennaio-giugno 2024

CACUCCI  EDITORE
BARI

PROPRIETÀ LETTERARIA RISERVATA

© 2024 Cacucci Editore – Bari

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220

<http://www.cacuccieditore.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.

Sommario

Editoriale

ROBERTO GUARASCI, Quarant'anni di «AIDAinformazioni» (1984-2024) 9

Contributo su invito

PIERO INNOCENTI, Que reste-t-il de nos amours? Arti tradizionali di un possibile moderno Trivio: Archivistica, Bibliografia, Documentazione 17

Contributi

FRANCESCO AMATO, ANTONELLA BENVENUTO, SILVIA CENITI, MARIA TERESA CHIARAVALLOTTI, CLAUDIA LANZA, ERIKA PASCERI, Indagine e analisi comparativa delle procedure di codifica nella Terapia del Dolore in Italia 59

ANDREA CAPACCIONI, Sull'affidabilità delle bibliografie generate dai chatbot. Alcune considerazioni 81

ALESSANDRO MAISTO, La dimensione Testuale del Videogioco. Classificazione dei transcript dei videogiochi basata sul lessico 95

ALEXANDER MURZAKU, PONTISH YERAMYAN, CURT ANDERSON, STEVEN BUXBAUM, RUBEN DIAZ, MARIELLE LERNER, ARMENUI MINASYAN, HAZEL MITCHLEY, JODIE-ANN PENNANT, MIA SHANG, BRISA SPEIER BRITO, Discovering and documenting brilliance. A novel multimodal annotation method 117

CAMILLA NAPPI, Le document : miroir des enjeux institutionnels et de l'évolution linguistique des transitions écologique et énergétique en France (2007-2022) 145

Testimonianze

PIERO CAVALERI, FERRUCCIO DIOZZI, La Documentazione in Italia e il ruolo di Paolo Bisogno: una rapida evoluzione 163

PAOLA GARGIULO, LUCIA MAFFEI, Domenico (Ingo) Bogliolo. Profilo di un navigatore nell'Information Science 169

Contributi

Sull'affidabilità delle bibliografie generate dai chatbot. Alcune considerazioni

Andrea Capaccioni*

Abstract: The aim of this paper is to examine the use of chatbots in academic settings, focusing on certain aspects of the reliability of bibliographies generated by artificial intelligence (AI). This study uses data obtained from querying ChatGPT 3.5 and Bard to produce a bibliography on a specific topic (Open Access Books). The analysis of the query results, along with comparisons to similar studies, has highlighted the low level of accuracy in the bibliographic citations generated by chatbots and the prevalence of phenomena such as hallucinations and confabulations in AI, which can affect the accuracy of the information. The chatbots use real author and editor names and existing titles that are consistent with the prompt terms, but they frequently use them incorrectly.

Keywords: Artificial Intelligence, Chatbot, Fake References, Hallucinations, Confabulations.

1. Introduzione

Con il lancio del primo chatbot di nuova generazione, nel novembre del 2022, nel mondo accademico è ripresa con vigore la discussione sulle potenzialità dell'intelligenza artificiale (d'ora in poi IA) in ambito scientifico. I chatbot sono programmi di IA basati su modelli linguistici, ovvero software addestrati a ricavare informazioni da grandi quantità di dati in gran parte presi da Internet, che simulano una conversazione con un essere umano e spesso sono utilizzati nei siti web e nelle applicazioni mobili per fornire assistenza ai clienti o per rispondere alle domande degli utenti. Quando i chatbot vengono interrogati, non avviano un processo di ricerca ed estrazione delle informazioni, ma elaborano le loro risposte utilizzando algoritmi che considerano i termini presenti nei prompt e si basano su stime probabilistiche e previsioni statistiche generate da grandi modelli linguistici e da un lungo addestramento¹.

* Dipartimento di Lettere-Lingue, Letterature e Civiltà antiche e moderne, Università degli Studi di Perugia, Italia. andrea.capaccioni@unipg.it.

¹ Per una introduzione generale ai chatbot si rinvia a (Roncaglia 2023, in particolare ai capitoli 9. "Il contesto: IA e reti neurali" e 10. "Le IA generative").

I chatbot hanno il potenziale per migliorare e offrire nuove opportunità nell'ambito dell'istruzione superiore e della ricerca. Tuttavia, sono emersi alcuni aspetti critici. Sono state evidenziate, per esempio, implicazioni di natura etica legate all'impiego nella ricerca e nella didattica (Stokel-Walker 2022). L'UNESCO ha promosso una guida che contiene un utile riepilogo delle principali problematiche nell'ambito dell'*higher education* (Sabzalieva e Valentini 2023). I chatbot sono in grado di generare idee per progetti o fornire indicazioni sulle fonti, assistono nella ricerca in repository e banche dati, traducono in molte lingue, contribuiscono a migliorare la qualità dei testi dal punto di vista linguistico e dello stile, sono di supporto alle attività di revisione paritaria (Srivastava 2023), contribuiscono perfino a generare articoli scientifici (Buruk 2023; Macdonald et al. 2023). Ci sono degli ambiti però in cui queste tecnologie hanno evidenziato dei limiti, per esempio nella capacità di elaborare bibliografie conformi agli standard accademici.

Il presente contributo intende esaminare alcuni aspetti dell'affidabilità delle bibliografie generate dall'IA anche con l'ausilio dei dati ottenuti attraverso l'interrogazione di ChatGPT 3.5 e Bard (attualmente non più disponibile)². Va specificato che si tratta di una semplice query che ha generato un campione limitato che crediamo tuttavia possa fornire una esemplificazione attendibile per comprendere il funzionamento di queste tecnologie e in più costituire la testimonianza di una fase della loro rapida evoluzione, come dimostra la decisione di Google di sostituire Bard con Gemini.

ChatGPT 3.5 e Bard, sviluppati rispettivamente da OpenAI e Google, sono progettati per simulare conversazioni con esseri umani e destinati a un pubblico ampio e per questo più diffusi e facilmente raggiungibili, anche perché accessibili in modo gratuito. Il primo si distingue nei compiti che richiedono la comprensione e la risposta a domande basate su fatti, il secondo si caratterizza per la capacità di generare testi creativi. Ciò si deve anche al fatto che i due chatbot lavorano su differenti set di dati la cui consistenza non siamo in grado di conoscere nel dettaglio in quanto OpenAI e Google forniscono solo informazioni generiche. ChatGPT utilizza un *large supervised training dataset* basato anche su informazioni prese da Internet, utilizzando per esempio i dati raccolti nel Web da soggetti terzi come (Common Crawl, n.d.) (Brown et al. 2020, 3), mentre Google ha addestrato LaMDA (Language Model for Dialogue Applications), il linguaggio utilizzato da Bard, «on a dataset created from public dialog data and other public web documents» (Thoppilan et al., 2022, 4).

Può essere utile aggiungere qualche altra informazione. ChatGPT 3.5 (*chat generative pre-trained transformer*) è stato reso disponibile al pubblico da OpenAI, un'azienda di ricerca e distribuzione di IA fondata da Elon Musk (dimessosi nel 2018), Sam Altman e altri, a fine novembre del 2022 (Mahyoob,

² In Internet si può trovare anche la denominazione Bard 2.0.

Algaraady, e Alblwi 2023). Il chatbot, che può contare su un dataset che ha raggiunto l'ordine dei terabyte, è in grado di apprendere le relazioni tra le parole e di generare testi che si avvicinano a quelli che potrebbero essere elaborati dall'uomo. ChatGPT è basato sull'architettura *transformer*, una rete neurale progettata per compiti di elaborazione del linguaggio naturale (NLP) particolarmente adatta per la generazione di testi. Questo chatbot nella versione da noi presa in esame, non ha accesso, per scelta del produttore, a dati internet aggiornati (il limite è fissato nel mese di settembre del 2021). Per completare l'informazione, diremo che OpenAI ha sviluppato altri due modelli (che non prenderemo in esame), consultabili prevalentemente previa sottoscrizione di un abbonamento, denominati GPT-4 e GPT-4o che presentano diverse funzionalità avanzate.

Bard, il chatbot di Google, è stato lanciato agli inizi del 2023 (Knight 2023) ed è basato su un modello linguistico di intelligenza artificiale definito *factual*, sviluppato da Google AI, in grado di comprendere il significato di parole e frasi, analizzare il contesto di una conversazione, fornire informazioni. Bard, come abbiamo detto, è stato addestrato su una grande quantità di dati e ciò gli ha consentito di generare testi, tradurre, scrivere diversi tipi di contenuti, rispondere alle domande. Dal febbraio del 2024 Bard non è più consultabile nelle modalità originarie, Google ha infatti ufficializzato l'avvio di Gemini (Google, n.d.), un progetto presentato come parte della riorganizzazione dei propri modelli di IA (Pichai e Hassabis 2023).

2. “Create a bibliography”: gli errori dell'IA

Il 17 dicembre del 2023 abbiamo chiesto a ChatGPT 3.5 e Bard di elaborare una bibliografia su di un determinato argomento. I due chatbot sono stati messi alla prova utilizzando il seguente *prompt*, ovvero l'istruzione impartita all'IA in modo che possa compiere l'azione richiesta: “Create a bibliography about Open Access Books”. L'interrogazione è stata formulata in inglese perché online sono disponibili più informazioni in questa lingua (Fig. 1) e per questo motivo abbiamo pensato che avremmo potuto ottenere un numero maggiore di risposte. L'espressione “to create a bibliography” è attestata in lingua inglese, in particolare in ambito universitario (es. “The purpose of this course is to learn how to create a bibliography”; “Students are required to create a bibliography at the end of their research paper, listing all the sources they consulted”). Si è preferito far ricorso a *bibliography* rispetto a *references* (o *list of references*) in quanto, pur essendo termini considerati simili, la bibliografia fa riferimento a un insieme più ampio di informazioni relative alle fonti consultate per la stesura di un articolo scientifico o di una monografia.

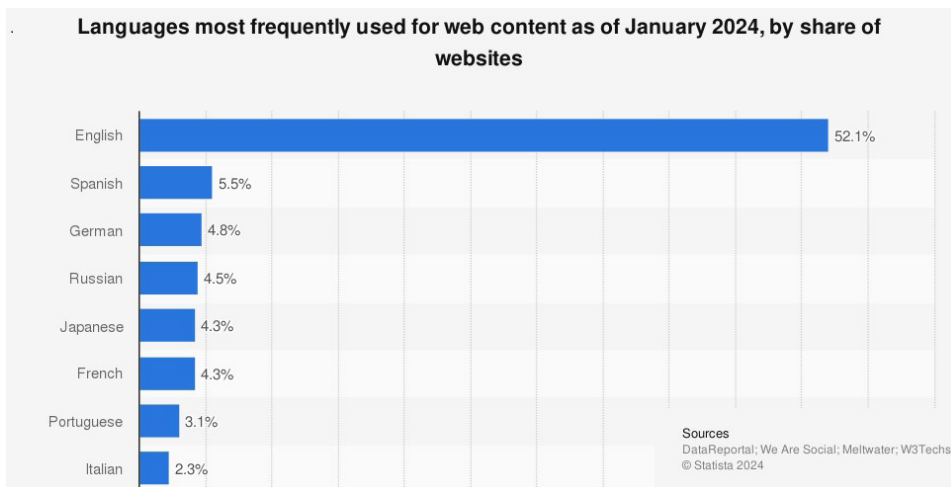


Figura 1: Lingue maggiormente usate per i contenuti nel web (DataReportal; W3Techs 2024).

La scelta dell'oggetto della bibliografia è ricaduta sulla “monografia scientifica ad accesso aperto”, conosciuta in lingua inglese con l'espressione *Open Access Book* (OAB), in quanto è un argomento da noi approfondito in altre sedi e questo ha reso più facile la verifica del grado di correttezza delle citazioni bibliografiche³. Nonostante sia stato scelto un argomento molto specifico e, come in precedenza ricordato, sia stata effettuata una singola interrogazione, riteniamo che i risultati ottenuti possano fornire informazioni utili⁴. In primo luogo, essi documentano il modo in cui i chatbot elaborano una bibliografia in un preciso momento della loro esistenza. Si tratta di un aspetto non trascurabile, in quanto l'IA è in grado di migliorarsi costantemente grazie al feedback degli utenti e all'intervento di azioni correttive (anche umane) introdotte dagli stessi gestori dei chatbot. Inoltre, i dati ottenuti possono contribuire a fornire, anche grazie al confronto con i risultati di altre ricerche (Agrawal, Mackey, e Kalai 2023; Athaluri et al. 2023; Day 2023; McGowan et al. 2023), un quadro più dettagliato sul modo in cui l'IA elabora le bibliografie. Torneremo su questi aspetti nel paragrafo successivo.

Per facilitare la lettura dei risultati abbiamo suddiviso le citazioni bibliografiche elaborate dai chatbot in tre categorie: citazione corretta, citazione falsa, altro argomento (Tab. 1). I dati ottenuti a seguito della formulazione del prompt sono stati trascritti mantenendo lo stile citazionale utilizzato dai chatbot (Tab. 2). Le valutazioni, a seguito della nostra analisi, sono inserite subito

³ Sul tema dell'OAB si veda in particolare (Capaccioni e Ranfa 2023) e (Capaccioni 2021). Altri studiosi hanno scelto criteri diversi per il riscontro dei risultati. Per esempio, Terence Day ha verificato l'affidabilità dei riferimenti bibliografici forniti da ChatGPT prendendo come riferimento cinque articoli pubblicati da una rivista scientifica (Day 2023).

⁴ I risultati di una ricerca impostata in modo simile sono consultabili in (McGowan et al. 2023).

dopo le singole citazioni e sono contrassegnate da uno a tre asterischi: * citazione corretta; ** citazione falsa; *** altro argomento. Tra parentesi rotonde sono indicati gli elementi della citazione che non risultano coerenti. Va inoltre fatta una importante precisazione: in entrambi i chatbot le citazioni risultate false e riferite ad articoli e libri non fanno riferimento allo specifico argomento degli OAB da noi richiesto, bensì a quello più generale dell'Open Access.

<i>citazione corretta</i>	tutti gli elementi (autore, titolo, editore, anno di pubblicazione, ecc.) che compongono la citazione sono corretti.
<i>citazione falsa</i>	uno o più elementi che compongono la citazione non risultano coerenti.
<i>altro argomento</i>	l'opera citata non corrisponde all'argomento richiesto nel prompt.

Tabella 1: Suddivisione delle citazioni bibliografiche.

Nella risposta di ChatGPT sono presenti 10 contributi (tutti accompagnati da un breve abstract, ma privo di indicazione del Digital Object Identifier-DOI): 4 libri, 4 rapporti⁵, 1 articolo; 1 contributo è risultato non attinente all'argomento prescelto. Delle 9 citazioni prese in considerazione ne risultano: 2 corrette e 7 false (Tab. 2).

ChatGPT	Bard
<p>Suber, P. (2012). "Open Access." The MIT Press. *</p> <p>A foundational work by Peter Suber providing an in-depth exploration of the open access movement, its principles, and its impact.</p> <p>Eve, M. P. (2014). "Open Access and the Humanities: Contexts, Controversies, and the Future." Cambridge University Press. *</p> <p>Examines the challenges and opportunities of open access in the humanities, offering insights into the specific context of this field.</p> <p>Schöpfel, J., & Prost, H. (Eds.). (2013). "Open Access to Scientific and Technical Information: State of the Art and Future Trends." De Gruyter Saur. ** (autori, editore, anno)</p>	<p><i>Books</i></p> <p>Suber, P. (2022). Open Access: Why It Matters. MIT Press. ** (titolo, anno)</p> <p>This book provides a comprehensive overview of open access, discussing its history, benefits, and challenges. It also covers the different models of open access publishing.</p> <p>Bawden, D. (2020). Open Access in Theory and Practice: The Theory-Practice Relationship and Openness. Routledge. ** (autori)</p> <p>This book explores the relationship between theory and practice in open access, examining philosophical, legal, and economic aspects of open access alongside its practical implementation.</p>

⁵ Nella categoria "libri" abbiamo compreso le monografie, gli atti di convegno, ecc.; in quella dei "rapporti" abbiamo inserito i contributi che non si configurano né come articoli, né come libri.

<p>A collection of essays providing a comprehensive overview of open access in scientific and technical disciplines, addressing current trends.</p> <p>Chan, L., & Costa, S. (Eds.). (2013). "Open Access and the Humanities: Contexts, Controversies, and the Future." Cambridge University Press. ** (autori, anno)</p> <p>An anthology that delves into the impact and challenges of open access in humanities research, offering diverse perspectives.</p> <p>Morrison, H. (2015). "Sustaining the Knowledge Commons." An OCLC Publication. ** (anno, editore)</p> <p>Explores economic and strategic issues related to sustaining open access initiatives, emphasizing the importance of a knowledge commons.</p> <p>Swan, A. (2010). "The Open Access Citation Advantage: Studies and Results to Date." JISC. ** (editore)</p> <p>A report summarizing studies on the citation advantage of open access articles, contributing to the ongoing discussion on the impact of open access.</p> <p>Jubb, M. (2016). "Monitoring the Transition to Open Access: A Report for Universities UK." Research Information Network. ** (anno)</p> <p>Examines the progress and challenges in the transition to open access in the UK, offering insights into institutional perspectives.</p> <p>OAPEN. (2019). "Open Access Books in the Humanities and Social Sciences: An Overview of the History and Current Situation." OAPEN. ** (autore, titolo, anno, editore)</p> <p>Provides a comprehensive overview of the historical development and current state of open access books in the humanities and social sciences.</p> <p>Bosman, J., Kramer, B., & Otte, W. (2006). "The Economics of Open Access Publishing." Information Economics and Policy, 18(4), 369-381. ** (autore, titolo, editore, anno)</p>	<p>Landsman, K. (2020). Foundations of Quantum Theory: From Classical Concepts to Operator Algebras. Springer Nature. *** (altro argomento)</p> <p>This open access book offers an in-depth exploration of the foundations of quantum theory, starting from classical concepts and progressing to operator algebras.</p> <p>Nordhoff, S. (2019). Cookbook for Open Access Books. Open Book Publishers. ** (editore, anno)</p> <p>This hands-on guide provides practical advice for authors and publishers on how to create and publish open access books.</p> <p><i>Journal Articles</i></p> <p>Solomon, D. J., & Björk, B.-C. (2022). Developing Open Access Journals: A Practical Guide. Routledge. ** (autore/i, editore, anno+ non è un articolo)</p> <p>This practical guide provides step-by-step instructions on how to establish and manage an open access journal.</p> <p>Levy, N. (2022). Bad Beliefs: Why They Happen to Good People. Oxford University Press. *** (altro argomento, non è un articolo)</p> <p>This engaging book explores the psychology of belief formation and the reasons why people sometimes hold false beliefs.</p> <p>Duschinsky, R. (2022). Mentalizing and Epistemic Trust: The Work of Peter Fonagy and Colleagues at the Anna Freud Centre. Oxford University Press. *** (altro argomento, non è un articolo)</p> <p>This book examines the role of mentalizing in human relationships and its impact on epistemic trust.</p> <p><i>Web Resources</i></p> <p>Directory of Open Access Books (DOAB): https://www.doabooks.org/ *</p> <p>DOAB is a community-driven discovery service that indexes and provides access to scholarly, peer-reviewed open access books.</p>
---	--

<p>A scholarly article that conducts an economic analysis of open access publishing, shedding light on the economic aspects of this publishing model.</p> <p>Fuchs, C., & Sandoval, M. (Eds.). (2013). "Critique, Social Media and the Information Society." Routledge. *** (altro argomento)</p> <p>Explores the intersection of open access with social media and the information society, providing a critical perspective on contemporary issues.</p> <p>Remember to check for the latest publications, as the field of open access is dynamic, and new research and insights are continually emerging.</p>	<p>OAPEN: https://www.oapen.org/ *</p> <p>OAPEN is an online library and publication platform that works with publishers to build a quality-controlled collection of open access books.</p> <p>MIT Press Open Access Books: https://mitpress.mit.edu/open-access-at-mit-press/books/ *</p> <p>MIT Press offers a variety of open access books across a range of disciplines.</p> <p>IntechOpen Open Access Books: https://www.intechopen.com/books *</p> <p>IntechOpen is a leading publisher of open access books in science, technology, and medicine.</p> <p>Fonti</p> <p>www.waikato.ac.nz/library/guidance/guides/open-access-information</p>
---	--

Tabella 2: I risultati del prompt: "Create a bibliography about open access books". Legenda: * citazione corretta; ** citazione falsa; *** altro argomento.

Tra parentesi tonde sono indicati gli elementi non corretti.

Osserviamo ora il tipo di manipolazione messo in atto dal chatbot prendendo in esame più da vicino una delle citazioni false che trascriviamo nello stile citazionale originale:

Bosman, J., Kramer, B., & Otte, W. (2006). "The Economics of Open Access Publishing." *Information Economics and Policy*, 18(4), 369-381.

Va fatto notare che la rivista «*Information Economics and Policy*» esiste e presenta il numero 18, fascicolo 4 pubblicato nel 2006, al suo interno però non c'è traccia del contributo segnalato⁶. Con lo stesso titolo si può invece rintracciare un altro articolo, non presente nell'elenco generato dal chatbot, di Christian Zimmermann, *The economics of open access publishing*, pubblicato però nella rivista «*Economic Analysis and Policy*» (39, 1, 2009, pp. 49-53, [https://doi.org/10.1016/S0313-5926\(09\)50054-9](https://doi.org/10.1016/S0313-5926(09)50054-9)).

Da questo accertamento è possibile cogliere alcune caratteristiche del modo in cui ChatGPT genera citazioni false che però apparentemente sembrano vere. Il chatbot impiega nomi di autori ed editori reali e titoli esistenti, spesso coerenti con i termini del *prompt*, ma li utilizza in modo non corretto.

⁶ Cfr. (Science Direct, n.d.). Anche la paginazione dell'articolo risulta errata.

Bard, nella sua risposta, propone 11 contributi (con breve abstract, senza DOI) suddivisi, ed è questo un aspetto che lo differenzia dal modo in cui ChatGPT presenta i dati bibliografici, in categorie: *books*, *journal articles* e *web resources*; con l'aggiunta della segnalazione di una "fonte"⁷. Va notato che il chatbot risulta impreciso nella classificazione delle pubblicazioni in quanto inserisce tre monografie tra le riviste accademiche. Al termine del riscontro risultano utili ai fini della nostra analisi 4 libri e 4 risorse web, mentre 3 contributi non si riferiscono all'argomento prescelto. Sono dunque 8 i contributi presi in considerazione: 4 riferiti a monografie e 4 a risorse reperibili in Internet. Nessuno dei volumi risulta descritto correttamente, le descrizioni bibliografiche presentano le caratteristiche che abbiamo riscontrato in ChatGPT. Bard genera infatti citazioni false ricorrendo all'utilizzo di autori ed editori reali e titoli esistenti. Anche in questo caso ricorriamo a un esempio. Nella bibliografia è presente l'opera, considerata un punto di riferimento per lo studio dell'accesso aperto, *Open access* di Peter Suber:

Suber, P. (2022). *Open Access: Why It Matters*. MIT Press.

La citazione, in apparenza verosimile, contiene ben 2 indicazioni errate: il sottotitolo (*Why it matters*) e l'anno di prima pubblicazione in quanto è indicato il 2022 anziché il 2012. Le 4 risorse web presenti nell'elenco di Bard corrispondono invece a risorse esistenti.

3. Allucinazioni, confabulazioni e accuracy

I risultati ottenuti dall'interrogazione mostrano che ChatGPT ha generato un numero più alto di citazioni false rispetto a Bard. In un'altra ricerca in cui è stato effettuato un confronto simile, nell'ambito però della letteratura in psichiatria, si sono registrati esiti diversi. Secondo (McGowan et al. 2023, 5) ChatGPT ha fornito un elenco di 35 citazioni, di cui solo il 6% corrisponde effettivamente a opere reali, mentre Bard ha generato otto citazioni false. Questi risultati ci dicono che entrambi i chatbot a prescindere dalle loro "competenze" nei diversi ambiti disciplinari, che crediamo possano variare in quanto essi utilizzano differenti basi di dati per il loro addestramento, hanno dimostrato di essere inaffidabili nella generazione di citazioni bibliografiche. Il modo in cui ChatGPT e Bard elaborano le false citazioni bibliografiche è, come abbiamo visto, simile e conferma l'esistenza di un fenomeno noto agli studiosi con il nome di *allucinazione*⁸. Un termine preso in prestito dalla

⁷ Si tratta di un sito universitario che contiene informazioni generali sull'OA (The University of Waikato, n.d.).

⁸ Il fenomeno generale delle allucinazioni dell'IA è molto studiato; per un'ampia introduzione

psicologia e riferito a quelle persone che percepiscono come reale ciò che è immaginario. Nell'ambito dell'IA l'allucinazione (*hallucination*) si può definire «a phenomenon where AI generates a convincing but completely made-up answer» (Athaluri et al. 2023).

Questo tipo di allucinazione può essere attribuita a un problema specifico che presentano i modelli di NLP e indica appunto la generazione di contenuti falsi basati su dati reali. OpenAI ha riconosciuto fin da subito questo tipo di problemi e le difficoltà incontrate nel risolverli:

ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there's currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the model knows (opens in a new window), rather than what the human demonstrator knows (OpenAI 2022).

L'allucinazione può dunque essere considerata una sorta di "limite" (OpenAI 2022) dell'IA, oppure un "effetto collaterale", o meglio il risultato «di previsioni statistico-probabilistiche basate su grandi modelli linguistici e su un lungo addestramento, in parte autonomo e in parte supervisionato o per rinforzo: non vi è dunque nessuna "copiatura" meccanica delle informazioni incamerate attraverso il corpus di testi di partenza, e non vi è neanche un'operazione di estrazione dal corpus delle informazioni considerate più rilevanti rispetto al prompt dell'utente» (Roncaglia 2023, cap. 11).

Sulla base di queste considerazioni, possiamo comprendere come l'analisi delle allucinazioni presupponga la conoscenza della natura e del funzionamento dei modelli linguistici su cui sono basati i chatbot. Tuttavia, questo fenomeno, in particolare se studiato in ambito bibliografico, non può prescindere dall'osservazione delle risposte che forniscono i chatbot. Seguendo questo metodo, un gruppo di studiosi newyorkesi ha proposto alcune osservazioni che ci aiutano a comprendere meglio il fenomeno (McGowan et al. 2023, 5). Grazie all'esame dei risultati ottenuti a seguito di interrogazioni di tipo "bibliografico" («What has been published regarding suicidal behavior in individuals at high risk for psychosis?») effettuate su ChatGPT e Bard, i ricercatori sono giunti alla conclusione che il riferimento alle allucinazioni in questi casi deve essere considerato improprio. Le allucinazioni sarebbero percettive per natura e verrebbero prodotte dunque senza uno stimolo esterno, mentre le *fake references* dovrebbero essere assimilate a un disturbo nella produzione del linguaggio conosciuto con il nome di confabulazione (*confabulation*). Per i linguisti, la confabulazione consiste nel ricordare in modo errato o inventare del tutto le

dedicata alle allucinazioni generate da modelli di linguaggio naturale (NLG) si veda (Ji et al. 2023) e (Smith 2023).

informazioni per colmare le lacune della memoria⁹. L'esame della correttezza (*accuracy*¹⁰) delle citazioni generate dai chatbot, ovvero la possibilità di determinare la presenza di informazioni accurate e/o di *fabricated citations* (citazioni false o inventate), fornisce un prezioso strumento per testare l'affidabilità di alcune azioni dei modelli di IA e verificarne i progressi.

Perché dunque i chatbot “confabulano”? Gli studiosi americani suppongono che il motivo sia rintracciabile nel fatto che gli *output* (i risultati di un processo) dei modelli linguistici di grandi dimensioni (appartenenti al gruppo dei Large Language Models-LLM) utilizzati dai chatbot sono per natura probabilistici e fanno ricorso a stime basate sulla *semantic similarity* tra due elementi di testo¹¹. Nello specifico, questi chatbot utilizzano modelli di co-occorrenza basati sulla probabilità che due o più parole appaiano vicine in un testo l'una all'altra. Nel momento in cui sono interrogati, i chatbot cercano parole (token) «that co-occur often with query terms in its training data» e poi ne trovano altre che co-occorrono con quelle parole in un processo di iterazione che porta a generare una risposta coerente: «the end result is a very convincing pastiche from up to millions of probabilistically linked bodies of text» (McGowan et al. 2023, 5).

L'ipotesi della confabulazione è convincente, anche se sono auspicabili ulteriori ricerche, in quanto sembra favorire un'interpretazione più accurata delle caratteristiche delle false bibliografie e può così contribuire a contrastare meglio il fenomeno. Da tempo, informatici e linguisti in particolare, stanno lavorando per cercare di correggere questo limite dei chatbot. Una delle tecniche più note è il Reinforcement Learning from Human Feedback (RLHF) che prevede l'intervento dell'uomo sia nel controllo del set di dati di addestramento, sia nella verifica del gradimento (umano) delle risposte generate (OpenAI 2022). Ricordiamo anche la Retrieval-Augmented Generation (RAG) una modalità di addestramento dei LLM che favorisce la fruizione di informazioni esterne in tempo reale permettendo di generare risposte più accurate.

Gli sviluppatori di OpenAI non hanno dubbi sul fatto che nel tempo le allucinazioni e le confabulazioni svaniranno, ma c'è ancora del lavoro da fare. Il gruppo di studiosi americani il cui lavoro abbiamo altre volte menzionato fa notare che la versione di marzo 2023 di ChatGPT 3.5 e quella di giugno 2023 di Bard, entrambe sottoposte alla tecnica del RLHF, mostrano ancora bassi

⁹ Altri studi hanno esaminato il fenomeno della confabulazione proposto dagli studiosi americani, si veda per esempio (Emsley 2023).

¹⁰ Nell'ambito dell'IA con il termine *accuracy* si intende la misura di quanto un sistema informatico progettato per elaborare automaticamente testo in linguaggio naturale produce *output* corretti e fedeli rispetto ai dati di *input* o alle informazioni di riferimento. Si veda anche (Estevez et al. 2022, 9).

¹¹ Come abbiamo visto in precedenza, G. Roncaglia propone un'analisi per molti versi simile (Roncaglia 2023, cap. 11).

tassi di *accuracy* nella generazione delle citazioni bibliografiche (McGowan et al. 2023, 5). Gli stessi risultati da noi raccolti nel dicembre del 2023 ed esposti nel presente contributo confermano questi limiti (Tab. 2).

Gli sforzi compiuti dai produttori di IA per migliorare le prestazioni dei chatbot sono innegabili. Osserviamo anche lo sviluppo di strumenti più specifici che ricorrono all'uso dell'IA (es. Elicit, Paperpile, ecc.) per supportare la comunicazione scientifica. Tuttavia in questa fase, all'interno della quale includiamo anche il periodo qui preso in esame, i chatbot di OpenAI e Google sono caratterizzati da una accentuata sperimentazione che propone diverse soluzioni, ancora non definitive e soprattutto poco efficaci. Durante la seconda metà del 2023, per esempio, è stata segnalata la propensione di ChatGPT 3.5, quando gli è stato chiesto di elaborare una bibliografia su un argomento, a fornire fonti di riferimento generiche (siti web, directory, ecc.), anziché elenchi di citazioni bibliografiche¹². Nei mesi successivi, questa tendenza è stata corretta, e il chatbot è tornato a generare liste di risorse bibliografiche più specifiche¹³.

Gli utenti, in particolare quelli che operano in un contesto accademico, devono essere consapevoli dei limiti dei chatbot e adottare un approccio critico nell'utilizzo delle bibliografie generate da questi strumenti.

Riferimenti bibliografici

Agrawal, Ayush, Lester Mackey, and Adam Tauman Kalai. 2023. "Do Language Models Know When They're Hallucinating References?" *arXiv*, January 2023. <https://doi.org/10.48550/arxiv.2305.18248>.

Athaluri, Sai Anirudh, Sandeep Varma Manthena, V.S.R. Krishna Manoj Kesapragada, Vineel Yarlagadda, Tirth Dave, and Rama Tulasi Siri Duddumpudi. 2023. "Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References." *Curēus* 15 (4). <https://doi.org/10.7759/curēus.37432>.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. "Language Models are Few-Shot Learners." *arXiv*, May 2020. <https://arxiv.org/abs/2005.14165>.

¹² «For example, whereas ChatGPT in March 2023 provided a list of citations, most of which were fabricated, in June 2023, it no longer provided any citations. While this change did not increase the number of accurate citations, it reduced the number of fabricated citations to zero, thereby eliminating the need for time intensive verification» (McGowan et al. 2023, 5).

¹³ Abbiamo sottoposto a ChatGPT 3.5 in data 14 giugno 2023 il *prompt* già usato in precedenza: "Create a bibliography about Open Access Books", il risultato è stato un elenco di risorse bibliografiche (di cui non abbiamo controllato il livello di accuratezza). Nello stesso giorno la medesima richiesta è stata rivolta a Google Gemini che ha invece generato un breve elenco di fonti.

- Buruk, Oğuz Oz. 2023. "Academic Writing with GPT-3.5 (ChatGPT): Reflections on Practices, Efficacy and Transparency". In *26th International Academic Mindtrek Conference*, 144-53. Tampere, Finland: ACM. <https://doi.org/10.1145/3616961.3616992>.
- Capaccioni, Andrea. 2021. "Origins and Developments of the Open Access Books." In *Digital Libraries. Advancing Open Science*. <https://doi.org/10.5772/intechopen.95357>.
- Capaccioni, Andrea, e Elena Ranfa. 2023. "Gli editori e le monografie scientifiche ad accesso aperto: un rapporto in evoluzione." *Sistema Editoria. Rivista internazionale di studi sulla contemporaneità* 1 (1): 79-94. <https://doi.org/10.14672/se.v1i1.2124>.
- Common Crawl. n.d. Consultato il 19 maggio 2024. <https://commoncrawl.org/>.
- DataReportal; W3Techs. 2024. "Languages most frequently used for web content as of January 2024, by share of websites." Statista, January 2024. <https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/>.
- Day, Terence. 2023. "A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT." *The Professional Geographer* 75 (6): 1024-27. <https://doi.org/10.1080/00330124.2023.2190373>.
- Emsley, Robin 2023. "ChatGPT: these are not hallucinations – they're fabrications and falsifications." *Schizophrenia* 9 (1). <https://doi.org/10.1038/s41537-023-00379-4>.
- Estevez, Almenzar Marina, David Fernandez Llorca, Emilia Gomez Gutierrez, and Fernando Martinez Plumed. 2022. "Glossary of Human-centric Artificial Intelligence." *JRC Publications Repository*. <https://doi.org/10.2760/86066>.
- Google. n.d. "Gemini." Consultato il 19 maggio 2024. <https://gemini.google.com>.
- Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. 2023. "Survey of Hallucination in Natural Language Generation." *ACM Computing Surveys* 55 (12): 1-38. <https://doi.org/10.1145/3571730>.
- Knight, Will. 2023. "Meet Bard, Google's Answer to ChatGPT." *WIRED*, February 6, 2023. <https://www.wired.com/story/meet-bard-googles-answer-to-chatgpt/>.
- Macdonald, Calum, Davies Adeloye, Aziz Sheikh, and Igor Rudan. 2023. "Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis." *Journal of Global Health* 13 (February). <https://doi.org/10.7189/jogh.13.01003>.

- Mahyoob, Mohammad, Jeehaan Algaraady, and Abdulaziz Alblwi. 2023. "Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing." *International Journal of Emerging Technologies in Learning* 18 (14): 282-93. <https://doi.org/10.3991/ijet.v18i14.41725>.
- McGowan, Alessia, Yunlai Gui, Matthew Dobbs, Sophia Shuster, Matthew Cotter, Alexandria Selloni, Marianne Goodman, Agrima Srivastava, Guillermo A. Cecchi, and Cheryl M. Corcoran. 2023. "ChatGPT and Bard Exhibit Spontaneous Citation Fabrication during Psychiatry Literature Search." *Psychiatry Research* 326 (August): 115334. <https://doi.org/10.1016/j.psychres.2023.115334>.
- Open AI. 2022. "Introducing ChatGPT." November 30. <https://openai.com/blog/chatgpt/>.
- Pichai, Sundar, and Demis Hassabis. 2023. "Introducing Gemini: Our largest and most capable AI model." December 06. <https://blog.google/technology/ai/google-gemini-ai/#sundar-note>.
- Roncaglia, Gino. 2023. *L'architetto e l'oracolo: Forme digitali del sapere da Wikipedia a ChatGPT*. Bari; Roma: GLF editori Laterza. EPUB.
- Sabzalieva, Emma, and Arianna Valentini. 2023. *ChatGPT and artificial intelligence in higher education: quick start guide*. UNESCO International Institute for Higher Education in Latin America and the Caribbean. <https://unesdoc.unesco.org/ark:/48223/pf0000385146>.
- Science Direct. n.d. "Information Economics and Policy." Consultato il 19 maggio 2024. <https://www.sciencedirect.com/journal/information-economics-and-policy>.
- Smith, Craig S. 2023. "Hallucinations could blunt ChatGPT's success." *IEEE Spectrum*, March 29. <https://spectrum.ieee.org/ai-hallucination>.
- Srivastava, Mashrin. 2023. "A day in the life of ChatGPT as an academic reviewer: Investigating the potential of large language model for scientific literature review." *Preprint. Open Science Framework*, February 16. <https://doi.org/10.31219/osf.io/wydct>.
- Stokel-Walker, Chris. 2022. "AI Bot ChatGPT Writes Smart Essays-Should Professors Worry?" *Nature*, December 9. <https://doi.org/10.1038/d41586-022-04397-7>.
- The University of Waikato. n.d. "Open Research." Consultato il 19 maggio 2024. www.waikato.ac.nz/library/guidance/guides/open-access-information.
- Thoppilan, Romal, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, et al. 2022. "LAMDA: Language Models for Dialog Applications." *arXiv*, January 2022. <https://doi.org/10.48550/arxiv.2201.08239>.

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 42

N. 1-2 – gennaio-giugno 2024

Editoriale

ROBERTO GUARASCI

*Quaran'anni di «AIDAinformazioni»
(1984-2024)*

Contributi su invito

PIERO INNOCENTI

*Que reste-t-il de nos amours? Arti
tradizionali di un possibile moderno Trivio:
Archivistica, Bibliografia, Documentazione*

Contributi

FRANCESCO AMATO, ANTONELLA
BENVENUTO, SILVIA CENITI, MARIA
TERESA CHIARAVALLOTI, CLAUDIA
LANZA, ERIKA PASCERI

*Indagine e analisi comparativa delle
procedure di codifica nella Terapia del
Dolore in Italia*

ANDREA CAPACCIONI

*Sull'affidabilità delle bibliografie generate
dai chatbot. Alcune considerazioni*

ALESSANDRO MAISTO

*La dimensione Testuale del Videogioco.
Classificazione dei transcript dei
videogiochi basata sul lessico*

ALEXANDER MURZAKU, PONTISH
YERAMYAN, CURT ANDERSON, STEVEN
BUXBAUM, RUBEN DIAZ, MARIELLE
LERNER, ARMENUI MINASYAN, HAZEL
MITCHLEY, JODIE-ANN PENNANT, MIA
SHANG, BRISA SPEIER BRITO

*Discovering and documenting brilliance.
A novel multimodal annotation method*

CAMILLA NAPPI

*Le document : miroir des enjeux
institutionnels et de l'évolution linguistique
des transitions écologique et énergétique en
France (2007-2022)*

Testimonianze

PIERO CAVALERI, FERRUCCIO DIOZZI

*La Documentazione in Italia e il ruolo di
Paolo Bisogno: una rapida evoluzione*

PAOLA GARGIULO, LUCIA MAFFEI

*Domenico (Ingo) Bogliolo. Profilo di un
navigatore nell'Information Science*



mundaneum

In copertina

Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-407-6

ISSN 1121-0095



9 791259 654076



9 770112 100950