Received: 4 June 2024 | Revised: 14 June 2024 | Accepted: 14 June 2024

DOI: 10.1111/jop.13569

Journal of Oral Pathology & Medicine WILEY

EDITORIAL

Unveiling the Larsen effect in the scientific literature domain: Navigating quality amidst the artificial intelligence-driven deluge

Luca D'Aniello

Department of Social Sciences, K-Synth Academic Spin-Off, University of Naples Federico II, Naples, Italy

Correspondence
Luca D'Aniello, Department of Social Sciences, K-Synth Academic Spin-Off, University of Naples Federico II, Naples, Italy.
Email: luca.daniello@unina.it

Abstract

The challenges faced by the massive increase in scientific publications draw parallels to the Larsen effect, where an amplified sound loop leads to escalating noise. This phenomenon has resulted in information overload, making it difficult for researchers to stay updated and identify significant findings. To address this, knowledge synthesis techniques are recommended. These methods help synthesize and visualize large bodies of literature, aiding researchers in navigating the expanding information landscape. Furthermore, artificial intelligence (AI) and natural language processing tools, such as text summarization, offer innovative solutions for managing information overload. However, the overuse of AI in producing scientific literature raises concerns about the quality and integrity of research. This manuscript highlights the need for balanced use of AI tools and collaborative efforts to maintain high-quality scientific output while leveraging the benefits of extensive research.

The recent explosion of scientific publications has led to the phenomenon of information overload,[1] becoming a pervasive challenge. Researchers across all disciplines face the daunting task of navigating an ever-growing sea of documents, making it increasingly difficult to keep pace and identify the most impactful findings. The number of articles published each year has grown by 8%–9% over the past several decades. In the biomedical field alone, more than 1 million papers are indexed in the PubMed bibliographic database annually, averaging approximately two per minute.[1] COVID-19, for instance, significantly impacted scientific research on a global scale. Since the beginning of the pandemic in December 2019,[2] there has been an impressive surge in COVID-19-related publications. A search conducted in January 2021 identified over 97 000 articles on the topic indexed in the PubMed database,[3] with over 23 000 published since January 2020 alone.[4] It represented the largest explosion of scientific literature ever recorded.

The research field of oral science has been inundated by an avalanche of scientific information, particularly during the COVID-19 pandemic. Numerous articles were published on using saliva as a diagnostic tool, changes in oral health and hygiene practices, and identifying oral manifestations of COVID-19 and its vaccines, such as lesions, taste alterations, and Bell's palsy. Additionally, tele-dentistry has become a crucial solution for remote dental care. To measure scientific production on these topics, a total of 142 316 documents were retrieved on PubMed using the following query: "saliva" OR "oral health*" OR "oral hygiene*" OR "oral manifestation*" OR "oral side effect*" OR "tele-dentistry." Almost 40 000 documents have been published since 2019, representing 26% of the entire collection. This search provided a rapid overview of scientific production while confirming the significant increase in literature, especially over the past 6 years.

The phenomenon of exponentially growing publications mirrors the self-perpetuating cycle of the Larsen effect. Just as the increasing volume in an audio loop provides further amplification, the surge in published research provides further research endeavors. Each new publication stimulates additional research, leading to more publications. This creates a cycle where the output (published papers) continually amplifies the input (new research), just like an audio feedback loop generates escalating noise. Scientists, eager to stay at the cutting edge of the research, are forced to delve deeper into this expanding

knowledge base, leading to even more publications. However, similar to distortion provided by the Larsen effect, the sheer volume of scientific documents is reaching a critical threshold. Scholars are requesting tools and strategies to enhance information retrieval efficiency, allowing them to quickly identify research's core contributions, limitations, and practical implications. The approaches of knowledge synthesis represent a key solution. By the 1960s, the methodologies of knowledge synthesis were widely used in several research fields, such as social science, education, and psychology.[5] Knowledge synthesis provides a complete overview of research by evaluating and summarizing all available evidence on a particular topic through advanced qualitative and quantitative methods,[5] with bibliometrics emerging as an especially valuable approach. By introducing a systematic, transparent, and reproducible review process, bibliometrics employs statistical methods to analyze science, scientists, and scientific activity.[6] The main bibliometric approach is science mapping, which aims to identify the structural and cognitive patterns within a particular research domain.[7,8] It allows researchers to synthesize and visualize large bodies of scientific literature, providing relevant insights into the organization and evolution of research fields.[9]

The immense volume of scientific data has created a growing demand for innovative statistical techniques, particularly in the field of natural language processing (NLP). These techniques include methodologies for extracting knowledge by analyzing textual data.[10] Among these methodologies, text summarization has emerged as an innovative solution to manage the phenomenon of information overload,[11] helping scholars to pinpoint the information they need.[12] Text summarization automatically identifies relevant information from one or more documents using two main approaches: extractive summarization, which involves selecting and extracting key sentences directly from the text, and abstractive summarization, which rewrites and condenses the input text into a shorter version while preserving the main ideas and reducing redundancy.[13,14]

Relying on a combination of NLP, statistical, and machine learning techniques, artificial intelligence (AI) plays a crucial role in bringing academic research to a transformative era. Addressing the challenges posed by information overproduction, AI provides significant improvements in information retrieval and knowledge synthesis methodologies. The introduction of AI tools such as ChatGPT, Gemini, and other AI chatbots is revolutionizing scientific knowledge acquisition.[15] However, their current overuse, particularly for writing articles, is leading to a concerning trend. A recent study, which analyzed the most frequent keywords commonly used by these chatbots for text rephrasing and writing, identified over 60 000 AI-assisted documents in the Dimension database, representing over 1% of all scientific articles published worldwide last year.[16]

Given this information, which highlights the near-essential and indispensable use of AI, two major considerations, as noted by Gray,[16] remain pertinent. First, determining whether these tools are primarily used for purely stylistic purposes is challenging without detailed analysis. The rise in publications suggests that AI is increasingly used for more than just stylistic polishing. Second, the models behind the AI tools themselves may be implicated. As the more model-generated text is used as training data for future text generations, there is a predicted risk of "model collapse." This occurs when artificially generated text outweighs real text, leading to increasingly low-quality results.[17]

Any tool that helps scholars produce more publications in academia is highly attractive,[18] but the consequences might not be entirely positive. Although some researchers openly use and acknowledge AI tools, others may use them excessively, resulting in low-value research. The already challenging task of sifting through numerous published papers to find relevant studies will only become more difficult if AI tools like ChatGPT increase output. According to Daniel Hook, chief executive of Digital Science, a research–analytics firm in London, this could produce poor-quality papers and compromise research integrity.[19]

The surge in scientific literature presents a double-edged sword. On the one hand, it offers a wealth of benefits. The vast research findings facilitate cross-disciplinary collaboration, as scientists from different fields can draw upon and integrate knowledge from diverse perspectives. Moreover, this big knowledge repository allows for the rapid distribution of findings, sharing advancements and innovations, and accelerating scientific progress on a global scale. However, this growth also comes with drawbacks. Information overload is a primary concern, as researchers struggle to keep up with the sheer volume of new publications. This can lead to difficulties in finding quality research from less rigorous studies. The pressure to publish frequently—the well-known "publish or perish" aphorism—incentivizes researchers to prioritize quantity over quality, potentially compromising scientific integrity with redundant or marginally significant studies. Additionally, the ease and efficiency of AI tools contribute to a considerable increase in scientific literature, often at the expense of quality. This publication surge may lack the rigorous peer review and originality of high-quality research. The peer-review system is facing a collapse under the weight of an ever-increasing number of submissions, leading to longer review times and a potential compromise in the thoroughness and quality of reviews.[20,21] This strain exacerbates the problem of information overload, as more articles are published without rigorous scrutiny. Despite its role in knowledge production, the peer-review system appears compromised.[22] Greene[23] discusses this issue, stating: "During the past few years, the world's biomedical journals have been struck by a small but consistent rash of falsified peer reviews, a scientific crime spree that is pushing editors and publishers to greater vigilance to secure the integrity of the research they publish." As a result, more academic journals across all disciplines are removing the peer-review system, emphasizing a more careful selection of scientific articles for publication while trying to counter the phenomenon of publishing poor-quality documents.

A collaborative effort from all the communities is necessary to achieve a balanced environment where scientific publishing thrives. The landscape is becoming increasingly challenging for scholars across all disciplines who rely exclusively on conventional tools and approaches, such as database filters and traditional search methods. These methods struggle to keep pace with the rapid growth of scientific publications, leading to several issues: significantly more time and effort required for information synthesis and retrieval, a higher risk of

missing crucial information, and a potential decline in the quality and impact of research. Thus, it is imperative for researchers, including clinicians and students in oral science, to adopt and integrate AI and advanced statistical tools into their research practices. By using these technologies, researchers can remain competitive, well-informed, and capable of producing high-quality research that contributes meaningfully to the field. Nevertheless, it is essential to remember that these tools should assist the researcher, not replace them entirely. The ever-expanding body of knowledge will be readily accessible and effectively utilized, maximizing the benefits of this information production and guiding scientific progress toward a brighter future of high-quality research.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/jop.13569.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Luca D'Aniello https://orcid.org/0000-0003-1019-9212

REFERENCES

1. Landhuis E. Scientific literature: information overload. *Nature*. 2016;535(7612):457-458.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395:497-506. doi:10.1016/S0140-6736(20)30183-5
3. Becker RC, Bode C. Medical publishing during the COVID-19 pandemic: then and now. *J Thromb Thrombolysis*. 2021;51:1101-1106.
4. Brainard J. Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? *Science*. 2020;13(10):1126. doi:10.1126/science.abc7839.
5. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Eval Health Prof*. 2002;25(1):12-37.
6. Cuccurullo C, Aria M, Sarto F. Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics*. 2016;108(2):595-611.
7. Börner K, Chen C, Boyack K. Visualizing knowledge domains. *Annu Rev Inf Sci Technol*. 2003;37:179-255.
8. Noyons ECM, Moed HF, van Raan AFJ. Integrating research performance analysis and science mapping. *Scientometrics*. 1999;46(3):591-604.
9. Antons D, Breidbach CF, Joshi AM, Salge TO. Computational literature reviews: method, algorithms, and roadmap. *Organ Res Methods*. 2023;26(1):107-138.
10. Mohammad M, Seyed AM. A novel extractive multi-document text summarization system using quantum-inspired genetic algorithm: MTSQIGA. *Expert Syst Appl*. 2021;171:114555. doi:10.1016/j.eswa.2020.114555
11. Aria M, Cuccurullo C, D'Aniello L, Misuraca M, Spano M. Text Summarization of a scientific document: a comparison of extractive unsupervised methods. *Proceedings of the 16th International Conference on Statistical Analysis of Textual Data*. Vol 1. VADISTAT Press/Edizioni Erranti; 2022:67-73.
12. Zhang C, Zhao L, Zhao M, Zhang Y. Enhancing keyphrase extraction from academic articles with their reference information. *Scientometrics*. 2022;127(2):703-731.
13. Moratanch N, Chitrakala S. A survey on abstractive text summarization. Paper presented at: The 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT). 2016.
14. Radev DR, Hovy E, McKeown K. Introduction to the special issue on summarization. *Comput Linguist*. 2002;28(4):399-408. doi:10.1162/089120102762671927
15. Ilegbusi PH. The integration of artificial intelligence (AI) in literature review and its potentials to revolutionize scientific knowledge acquisition. *AfricArXiv*. 2024. doi:10.21428/3b2160cd.50b471d6
16. Gray A. ChatGPT "contamination": estimating the prevalence of LLMs in the scholarly literature. 2024 *arXiv Preprint arXiv:2403.16887*.
17. Shumailov I, Shumaylov Z, Zhao Y, Gal Y, Papernot N, Anderson R. The curse of recursion: training on generated data makes models forget. 2023 *arXiv Preprint arXiv:2305.17493*.
18. Prillaman M. Is ChatGPT making scientists hyper-productive? The highs and lows of using AI. *Nature*. 2024;627:16-17.
19. Conroy G. How ChatGPT and other AI tools could disrupt scientific publishing. *Nature*. 2023;622(7982):234-236.
20. Iyengar KP, Jain VK, Ish P. Publication surge in COVID-19: the flip side of the coin! *J Orthop Traumatol Rehabil*. 2021;13(2):180-182.
21. Nguyen VM, Haddaway NR, Gutowsky LF, et al. How long is too long in contemporary peer review? Perspectives from authors publishing in conservation biology journals. *PLoS One*. 2015;10(8):e0132557. doi:10.1371/journal.pone.0132557
22. Bell K, Kingori P, Mills D. Scholarly publishing, boundary processes, and the problem of fake peer reviews. *Sci Technol Hum Values*. 2024;49(1):78-104.
23. Greene J. Fraudsters strike peer review: stolen passwords, fake reviews threaten biomedical literature. *Ann Emerg Med*. 2015;65(4):A13-A15.

How to cite this article: D'Aniello L. Unveiling the Larsen effect in the scientific literature domain: Navigating quality amidst the artificial intelligence-driven deluge. *J Oral Pathol Med*. 2024;1-3. doi:10.1111/jop.13569