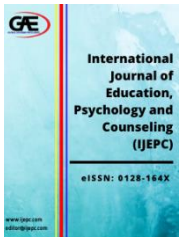
International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014



INTERNATIONAL JOURNAL OF EDUCATION, PSYCHOLOGY AND COUNSELLING (IJEPC)
www.ijepc.com

LARGE LANGUAGE MODEL AND HALLUCINATIONS: A BIBLIOMETRIC REVIEW

Nur Emma Mustaffa[1*], Lai Ke En[2], Norhazren Izatie Mohd[3], Fuziah Ismail[3], Nurshikin Mohamad Shukery[3], Siti Rahmah Omar[4], Hamidah Kamarden[5], Nur Hidayah Abd Rahman[6]

[1] Quantity Surveying Department, Tunku Abdul Rahman University of Management and Technology, Malaysia.
Email: nuremma@tarc.edu.my
[2] Department of Quantity Surveying, Tunku Abdul Rahman University of Management and Technology
Email: laike@tarc.edu.my
[3] Department of Quantity Surveying, Universiti Teknologi Malaysia, Malaysia
Email: norhazren@utm.my , b-fuziah@utm.my , b-nurshikin@utm.my
[4] Department of Landscape Architecture, Universiti Teknologi Malaysia, Malaysia
Email: siti.rahmah@utm.my
[5] Department of Chemical Engineering, Universiti Teknologi Malaysia, Malaysia
Email: hamidahkamarden@utm.my
[6] Management with Tourism Programme, Universiti Sains Islam Malaysia, Malaysia
Email: nhidayah.ar@usim.edu.my
[*] Corresponding Author

Article Info:

Article history:
Received date: 30.06.2025
Revised date: 21.07.2025
Accepted date: 18.08.2025
Published date: 01.09.2025

To cite this document:

Mustaffa, N. E., Lai, K. E., Mohd, N. I., Ismail, F., M. S. N., Omar, S. R., Kamarden, H., & Abd Rahman, N. H. (2025). Large Language Model and Hallucinations: A Bibliometric Review. *International Journal of Education, Psychology and Counseling, 10* (59), 185-200.

DOI: 10.35631/IJEPC.1059014

Abstract:

The rapid advancement and widespread adoption of Large Language Models (LLMs) have spurred increasing interest in understanding their capabilities and limitations, particularly the phenomenon of "hallucination"—the generation of plausible yet factually incorrect information. This bibliometric review aims to map the scientific landscape and research trends surrounding LLMs and hallucinations within the broader context of Artificial Intelligence (AI). Despite the growing relevance of these issues, the scholarly discourse remains fragmented, necessitating a comprehensive synthesis of the existing literature. To address this gap, we conducted a systematic search using the keywords "LLM," "hallucination," and "AI" across the Scopus database. The resulting dataset, comprising 513 relevant publications, was cleaned and standardised using OpenRefine. Further analysis was conducted using Scopus Analyser to identify publication trends, citation patterns, and prolific contributors. Meanwhile, VOSviewer software was employed to construct co-authorship networks, keyword co-occurrence maps, and thematic clusters. The analysis revealed a marked increase in publications post-2020, with a significant concentration of research in computer science, linguistics, and ethics. Keyword mapping highlighted emerging themes such as factual consistency, trustworthiness, and prompt engineering. Co-authorship networks revealed a

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

growing yet still loosely connected research community. These findings suggest that while interest in LLM hallucinations is rising, there is a need for deeper interdisciplinary collaboration and more rigorous evaluation frameworks. This study provides a foundational overview of the current research landscape and identifies critical directions for future investigation, especially in mitigating hallucinations and enhancing the reliability of LLM-generated content.

Keywords:

Large Language Modelling, LLM, Hallucination, Artificial Intelligence

Introduction

The advent of Large Language Models (LLMs) has revolutionised the field of Natural Language Processing (NLP), enabling significant advancements in text generation, question answering, and dialogue systems. These models, exemplified by systems such as GPT-3 and GPT-4, have demonstrated remarkable capabilities in generating human-like text, thereby transforming various domains including education, healthcare, and information retrieval (Bruno et al., 2023; Di Ieva et al., 2024; Wang et al., 2024). Despite their impressive performance, LLMs are not without their challenges. One of the most critical issues is the phenomenon of "hallucination," where the models generate outputs that are factually incorrect or nonsensical, posing significant risks to their reliability and trustworthiness (Huang et al., 2025; Reddy et al., 2024; Sakib, 2024).

Hallucinations in LLMs occur when the models produce content that appears plausible but is not grounded in the training data or real-world facts. This issue is particularly problematic in high-stakes fields such as healthcare, law, and scientific research, where accuracy is paramount (He et al., 2025; Reddy et al., 2024). The causes of hallucinations are multifaceted, including biases in training data, the models' tendency to generate content based on incomplete information, and their inherent limitations in understanding and processing complex queries (Huang et al., 2025; Reddy et al., 2024; Sakib, 2024). The implications of hallucinations are far-reaching, affecting the credibility of AI systems and raising ethical concerns about the dissemination of misinformation (Bruno et al., 2023; Maleki et al., 2024; Wang et al., 2024).

Recent research has focused extensively on understanding and mitigating hallucinations in LLMs. Various strategies have been proposed, ranging from improving the quality and diversity of training data to developing sophisticated detection and mitigation techniques (Ho et al., 2024; Huang et al., 2025; Reddy et al., 2024). For instance, the use of knowledge graphs has been explored to enhance the factual accuracy of LLM outputs by providing structured external information (Pons et al., 2025). Additionally, methodologies such as hierarchical multi-head attention and multi-level self-attention weighting mechanisms have been employed to improve the detection of hallucinations (Lu & Li, 2024). These approaches aim to enhance the interpretability and reliability of LLM-generated content, thereby addressing one of the most pressing challenges in the field (Huang et al., 2025; Lu & Li, 2024).

Moreover, the classification and taxonomy of hallucinations have been subjects of significant research. Hallucinations can be broadly categorised into factual hallucinations, where the generated content is factually incorrect, and fidelity hallucinations, where the content deviates

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014
from the expected style or context (He et al., 2025). This classification helps in developing targeted strategies for different types of hallucinations, thereby improving the overall robustness of LLMs (He et al., 2025; Huang et al., 2025). Furthermore, the development of benchmark datasets and evaluation frameworks has been crucial in assessing the performance of LLMs and their susceptibility to hallucinations (Jin et al., 2025). These benchmarks provide a standardised way to measure and compare the effectiveness of various mitigation techniques, facilitating the advancement of research in this area (Liang et al., 2024).

The field has seen several recent developments aimed at addressing the hallucination problem in LLMs. One notable approach is the integration of retrieval-augmented generation techniques, which combine the generative capabilities of LLMs with the precision of information retrieval systems (Huang et al., 2025). This hybrid approach aims to reduce hallucinations by grounding the generated content in verifiable sources, thereby enhancing the reliability of the outputs (Huang et al., 2025). Additionally, the use of context-aware prompt engineering has demonstrated promise in improving the accuracy of LLM-generated content by incorporating veracity-oriented constraints and background information (Jin et al., 2025). This technique leverages the extensive prior knowledge embedded within LLMs to produce more accurate and contextually appropriate responses (Jin et al., 2025).

Another significant development is the focus on domain-specific applications and the creation of high-quality evaluation datasets tailored to specific fields such as healthcare and education (He et al., 2025; Ho et al., 2024). These domain-specific datasets enable more precise evaluation and optimisation of LLMs, addressing the unique challenges posed by different application areas (He et al., 2025; Ho et al., 2024). For example, in the medical domain, the construction of specialised datasets and the use of advanced models like GPT-4 have been employed to evaluate and reduce hallucinations in medical question-answering systems (He et al., 2025). These efforts highlight the importance of domain-specific research in enhancing the practical utility of LLMs.

In conclusion, while LLMs have achieved remarkable success in various NLP tasks, the issue of hallucinations remains a significant challenge. Ongoing research is focused on understanding the underlying causes, developing effective mitigation strategies, and creating robust evaluation frameworks. These efforts are crucial for ensuring the reliability and trustworthiness of LLMs, particularly in high-stakes applications where accuracy is critical. As the field continues to evolve, addressing the hallucination problem will be essential for the broader adoption and acceptance of LLMs in real-world scenarios.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

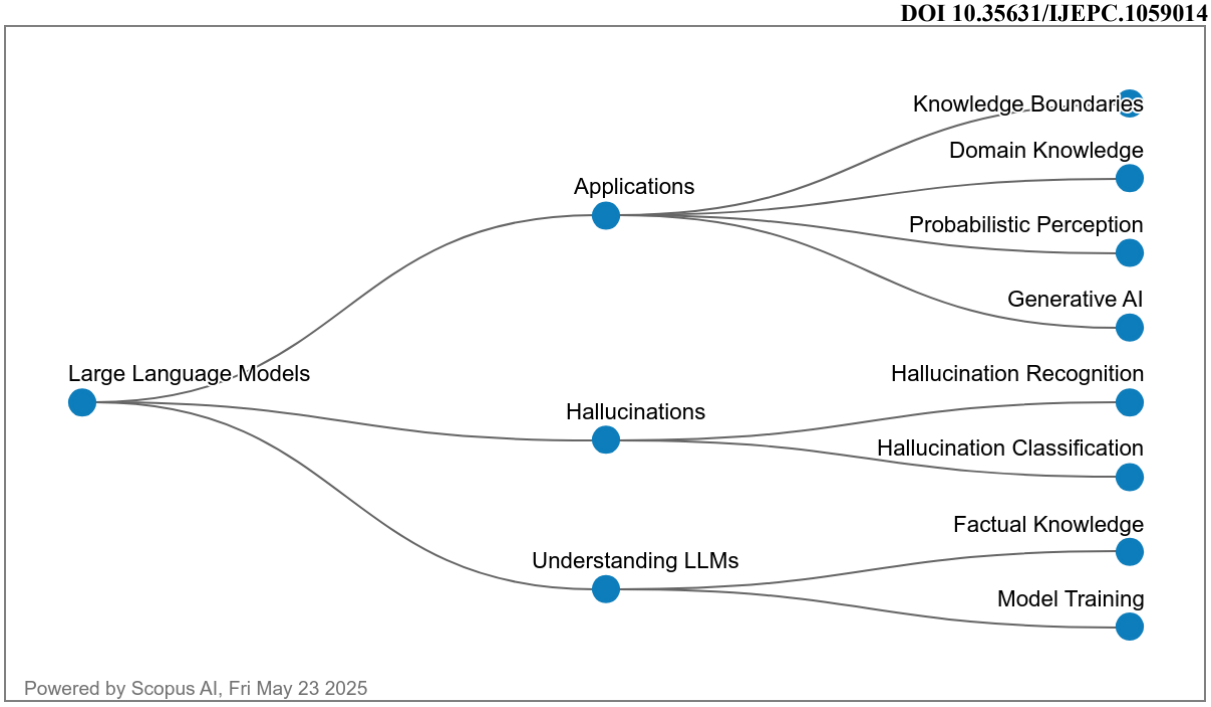


Figure 1: Overview of LLM and Hallucinations

Research Questions

RQ1 : What are the research trends in LLM and hallucination according to the year of publication?
RQ2 : What are the most cited articles?
RQ3 : What is the 10-country based on the number of publications?
RQ4 : What are the popular keywords related to the study?
RQ5 : What is the co-authorship by countries collaboration?

Methodology

Bibliometrics involves gathering, organising, and analysing bibliographic data from scientific publications (Alves et al., 2021; Assyakur & Rosa, 2022; Verbeek et al., 2002) beyond basic statistics, such as identifying publishing journals, publication years, and leading authors (Wu & Wu, 2017). Bibliometrics includes more sophisticated techniques, such as document co-citation analysis. Conducting a successful literature review requires a careful, iterative process to select suitable keywords, search the literature, and perform an in-depth analysis. This approach facilitates the compilation of a comprehensive bibliography and yields viable results (Fahimnia et al., 2015). With this in mind, the study focused on high-impact publications, as they provide meaningful insights into the theoretical frameworks that shape the research field. To ensure data accuracy, Scopus served as the primary source for data collection (Al-Khoury et al., 2022; di Stefano et al., 2010; Khiste & Paithankar, 2017). Additionally, to maintain quality, the study only considered articles published in peer-reviewed academic journals, deliberately excluding books and lecture notes (Gu et al., 2019). Using Elsevier's Scopus, known for its broad coverage, publications were collected from 2020 through December 2023 for further analysis.

Data Search Strategy

The study employed a screening sequence to determine the search terms for article retrieval. Afterwards, the query string was revised so that the search terms “LLM and hallucinations learning” should be the focus. The final search string refinement included 513 articles, which

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

were used for bibliometric analysis. As of May 2025, all articles from the Scopus database relating to LLM and hallucination and focusing on academics were incorporated in the study.

Table 1: The Search String

Scopus	**TITLE (large AND language AND model OR "llm" AND hallucinations OR "wrong information" OR "false information" OR "chatgpt")**

Data Analysis

VOSviewer is a user-friendly bibliometric software developed by Nees Jan van Eck and Ludo Waltman at Leiden University, Netherlands (van Eck & Waltman, 2010a, 2017). Widely utilised for visualising and analysing scientific literature, the tool specialises in creating intuitive network visualisations, clustering related items, and generating density maps. Its versatility allows for the examination of co-authorship, co-citation, and keyword co-occurrence networks, providing researchers with a comprehensive understanding of research landscapes. The interactive interface, coupled with continuous updates, ensures efficient and dynamic exploration of large datasets. VOSviewer's ability to compute metrics, customise visualisations, and its compatibility with various bibliometric data sources make it a valuable resource for scholars seeking insights into complex research domains.

Additionally, one of the standout features of VOSviewer is its capacity to transform intricate bibliometric datasets into visually interpretable maps and charts. With a focus on network visualisation, the software excels in clustering related items, analysing keyword co-occurrence patterns, and generating density maps. Researchers benefit from its user-friendly interface, enabling both novice and experienced users to explore research landscapes efficiently. VOSviewer's continuous development ensures it remains at the forefront of bibliometric analysis, offering valuable insights through metrics computation and customisable visualisations. Its adaptability to different types of bibliometric data, such as co-authorship and citation networks, positions VOSviewer as a versatile and indispensable tool for scholars seeking a deeper understanding and more meaningful insights within their research domains.

Datasets comprising information on the publication year, title, author name, journal, citation, and keywords in PlainText format were procured from the Scopus database, spanning the period from 2004 to December 2024. These datasets were then analysed using VOSviewer software version 1.6.19. Through the application of VOS clustering and mapping techniques, this software facilitated the examination and generation of maps. Offering an alternative to the Multidimensional Scaling (MDS) approach, VOSviewer focuses on situating items within low-dimensional spaces, ensuring that the proximity between any two items accurately reflects their relatedness and similarity (van Eck & Waltman, 2010b). In this respect, VOSviewer shares a similarity with the MDS approach (Appio et al., 2014). Diverging from MDS, which primarily engages in the computation of similarity metrics like cosine and Jaccard indices, VOS utilises a more fitting method for normalising co-occurrence frequencies, such as the Association Strength (AS_{ij}), and it is calculated as (Van Eck & Waltman, 2007):

$$AS_{ij} = \frac{C_{ij}}{w_i w_j},$$

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

which is "proportional to the ratio between the observed number of co-occurrences of i and j and the expected number of co-occurrences of i and j under the assumption that co-occurrences of i and j are statistically independent" (Van Eck & Waltman, 2007).

Findings

RQ1: What Are The Research Trends In LLM And Hallucination According To The Year Of Publication?

The bibliometric data from Scopus on the topic of "LLM and Hallucinations" depicts a dynamic trend in publication activity over the years 2023 to 2025. In 2023, there were 145 publications, with 28 likely centered specifically on hallucination phenomena in LLMs. This period marks a foundational phase, coinciding with the rising deployment of LLMs across academic and industrial sectors. Early concerns about model reliability and factual consistency likely spurred academic interest, prompting a wave of exploratory studies.

The year 2024 saw a significant surge in scholarly output, with publications nearly doubling to 259 and hallucination-focused papers increasing to 50. This sharp growth reflects escalating awareness and concern around hallucinations—instances where LLMs generate inaccurate or fabricated content. It is also indicative of intensified research funding and community discourse, driven by high-profile cases of LLM errors and their societal impact. The increased focus on hallucinations may also stem from interdisciplinary engagement, with researchers from linguistics, ethics, and computer science converging on this challenge.

Interestingly, there is a notable dip in 2025, with only 109 total publications and 21 directly addressing hallucinations. While the year is still ongoing and these figures may rise, the decline could suggest several interpretations: a shift in research priorities, saturation of preliminary findings, or possibly the integration of hallucination mitigation techniques into standard LLM design, reducing the novelty of the topic. Alternatively, research may be transitioning toward more nuanced or applied subtopics beyond initial hallucination detection, such as regulatory frameworks or real-world case studies.

Table 2: Publications by Years From 2023-2025

Year	Total publication	Percentage (%)
2025	109	21
2024	259	50
2023	145	28

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

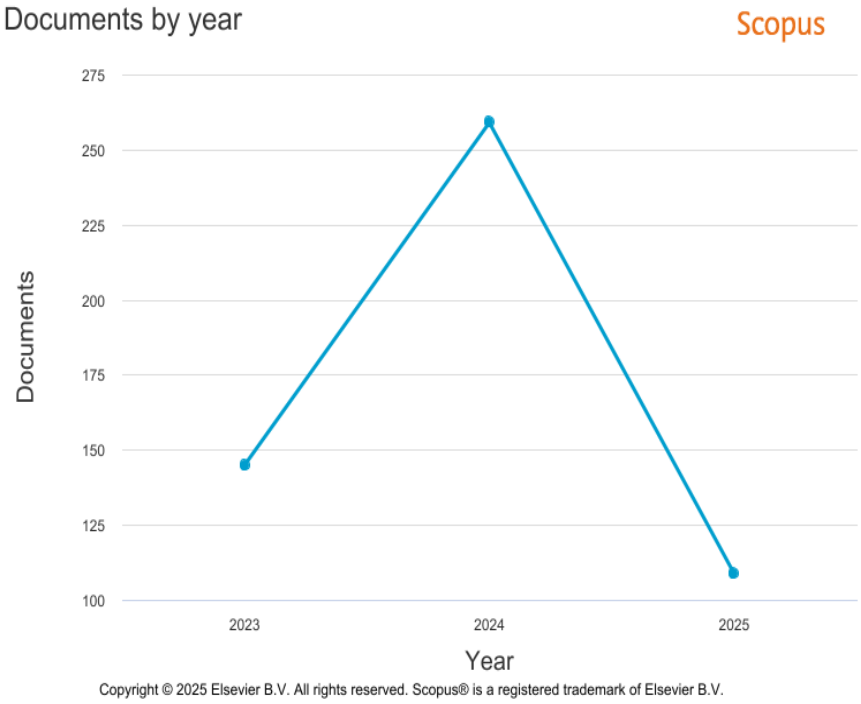


Figure 2: Publications by Years from 2023-2025

RQ2: What Are The Most Cited Articles?

The top 10 most cited publications from 2023, as identified in the Scopus Analyser, highlight the intense scholarly interest in ChatGPT and LLMs across various domains, especially education and medicine. The leading article by Kasneci et al., with 2,313 citations, underscores the wide-ranging educational implications of LLMs, framing them as both an opportunity and a challenge. This is closely followed by medical-focused studies such as those by Kung et al. (1,965 citations) and Gilson et al. (1,131 citations), which examine the performance of ChatGPT on the United States Medical Licensing Examination. The high citation counts indicate academic engagement and the pressing need to understand LLMs' role in professional education and assessment contexts.

The prominence of studies addressing ethical, practical, and pedagogical challenges is notable. Shen et al.'s work (566 citations) portrays LLMs as "double-edged swords," reflecting broader concerns about both their potential and risks. Similarly, Lund et al. (441 citations) and Perkins (374 citations) delve into ethical dilemmas, academic integrity, and the shifting nature of scholarly publishing in an AI-driven landscape. These works emphasise a dual trend in the discourse: enthusiasm about AI's capabilities and caution about its unchecked use, particularly in sensitive or evaluative settings like academia and research authorship.

The remaining studies further demonstrate the multidimensional interest in ChatGPT, extending beyond performance metrics into future outlooks and complementary human-AI collaboration. Liu et al. (353 citations) present a meta-level overview, synthesising research and offering perspectives on the LLM trajectory. Meanwhile, Jeon and Lee (288 citations) stress the importance of synergy between educators and AI systems. The citation impact across all these publications suggests that foundational, ethical, and integrative discussions around LLMs are timely and central to shaping future educational, medical, and academic policies in the age of generative AI.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

Table 3: The Top 10 Most Cited Authors

Authors	Title	Source title	Cited by
Kasneci E.; Sessler K.; Küchemann S.; Bannert M.; Dementieva D.; Fischer F.; Gasser U.; Groh G.; Günnemann S.; Hüllermeier E.; Krusche S.; Kutyniok G.; Michaeli T.; Nerdel C.; Pfeffer J.; Poquet O.; Sailer M.; Schmidt A.; Seidel T.; Stadler M.; Weller J.; Kuhn J.; Kasneci G.	ChatGPT for good? On opportunities and challenges of large language models for education (Kasneci et al., 2023)	Learning and Individual Differences	2313
Kung T.H.; Cheatham M.; Medenilla A.; Sillos C.; De Leon L.; Elepaño C.; Madriaga M.; Aggabao R.; Diaz-Candido G.; Maningo J.; Tseng V.	Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models(Kung et al., 2023)	PLOS Digital Health	1965
Gilson A.; Safranek C.W.; Huang T.; Socrates V.; Chi L.; Taylor R.A.; Chartash D.	How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment(Gilson et al., 2023)	JMIR Medical Education	1131
Shen Y.; Heacock L.; Elias J.; Hentel K.D.; Reig B.; Shih G.; Moy L.	ChatGPT and Other Large Language Models Are Double-edged Swords(Shen et al., 2023)	Radiology	566
Lund B.D.; Wang T.; Mannuru N.R.; Nie B.; Shimray S.; Wang Z.	ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing(Lund et al., 2023)	Journal of the Association for Information Science and Technology	441
Perkins M.	Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond (Perkins, 2023)	Journal of University Teaching and Learning Practice	374
Liu Y.; Han T.; Ma S.; Zhang J.; Yang Y.; Tian J.; He H.; Li A.; He M.; Liu Z.; Wu Z.; Zhao L.; Zhu D.; Li X.; Qiang N.; Shen D.; Liu T.; Ge B.	Summary of ChatGPT-Related research and perspective towards the future of large language models(Liu et al., 2023)	Meta-Radiology	353

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

De Angelis L.; Baglivo F.; Arzilli G.; Privitera G.P.; Ferragina P.; Tozzi A.E.; Rizzo C.	ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health(De Angelis et al., 2023)	Frontiers in Public Health	342
Meyer J.G.; Urbanowicz R.J.; Martin P.C.N.; O'Connor K.; Li R.; Peng P.-C.; Bright T.J.; Tatonetti N.; Won K.J.; Gonzalez-Hernandez G.; Moore J.H.	ChatGPT and large language models in academia: opportunities and challenges(Meyer et al., 2023)	BioData Mining	292
Jeon J.; Lee S.	Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT(Jeon & Lee, 2023)	Education and Information Technologies	288

RQ3: What Is The 10-Country Based On The Number Of Publications?

The bibliometric data from Scopus reveals that the United States (U.S.) is the leading contributor to research on LLMs and hallucinations, with 165 publications. This dominant position reflects the U.S.'s robust academic infrastructure, strong tech industry presence (notably companies like OpenAI, Google, and Meta), and early adoption of AI technologies. China follows with 115 publications, indicating its growing investment and strategic focus on AI research and development. Together, these two countries account for the majority of global scholarly output in this domain, underscoring their pivotal role in shaping the discourse and innovation around LLMs.

A second tier of contributors includes Germany and India, each with 41 publications, and the United Kingdom with 37. This grouping reflects a strong European and South Asian interest in LLM-related challenges, including ethical implications, model deployment, and education-focused applications. These countries typically have well-established research institutions and are engaged in both theoretical and applied AI studies. Their presence in the top 10 signals a global concern over the risks and opportunities of LLMs, particularly hallucinations that can impact trust, accuracy, and usability.

The remaining countries — Canada, Italy, South Korea, Australia, and Singapore — contribute between 18 and 29 publications each. These nations, while producing fewer papers, still represent active and influential research communities. Canada and Australia, for example, are known for their work in AI ethics and interdisciplinary research. Singapore's appearance is notable given its smaller population but strong emphasis on digital innovation and government-led AI initiatives. The geographic diversity of the top contributors illustrates that interest in LLM hallucinations spans continents and research traditions, reflecting the universal relevance and potential societal impact of this technological issue.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

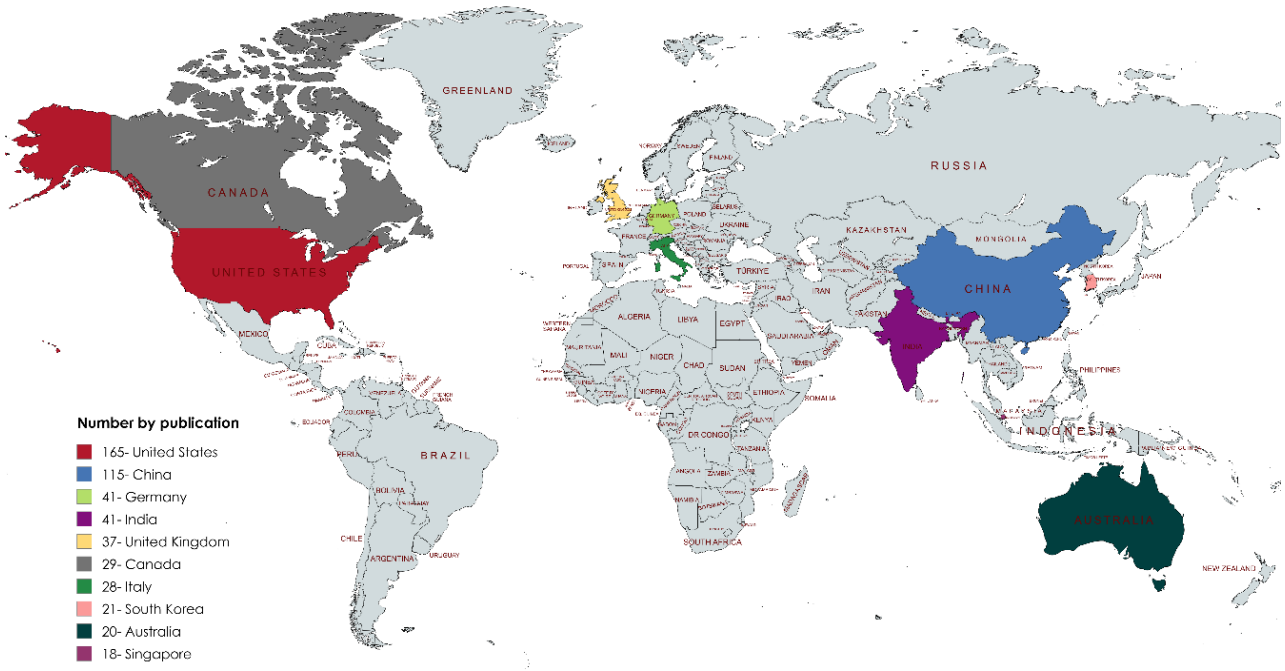


Figure 3: Top 10 Countries Based on the Number of Publications

RQ4: What Are The Popular Keywords Related To The Study?

The keyword analysis derived from VOSviewer reveals a dominant research interest in Artificial Intelligence (AI) and its subfields, particularly LLMs and ChatGPT, which have the highest occurrences (184 and 166, respectively) and link strengths (912 and 903, respectively). These keywords significantly overshadow others in both frequency and interconnectivity, highlighting a substantial research focus on generative AI technologies. Related terms such as GPT-4, machine learning, NLP, prompt engineering, and hallucination also score highly, reflecting active exploration into AI capabilities, reliability, and implementation challenges. The appearance of terms like hallucination and bias suggests growing scrutiny around the limitations and ethical implications of these models.

Another cluster of highly connected keywords centers on medical and healthcare applications, including terms such as medical education, health care, patient education, diagnostic medicine, and clinical decision support. These indicate robust intersections between AI technologies and healthcare, with strong Total Link Strengths (TLS) reflecting interdisciplinary collaboration and interest. The presence of domain-specific keywords such as radiology, ophthalmology, and pharmacokinetics further underscores the integration of AI tools into various clinical domains. AI's utility in education, information management, and automated decision-making is emphasised through frequently occurring terms like digital health, education technology, and automated systems.

A third thematic grouping includes keywords related to ethics, academic integrity, plagiarism, fairness, and data protection, highlighting the broader societal and governance concerns emerging alongside rapid AI adoption. With increasing reliance on AI in academia, medicine, and public policy, ethical considerations are gaining traction, evidenced by keywords such as AI governance, policy, and ethical AI systems. This trend demonstrates a growing awareness of the need to align AI development with human values, fairness, and accountability. Overall,

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

the analysis underscores a vibrant research ecosystem focused on technical advancement and on the responsible integration of AI into real-world settings.

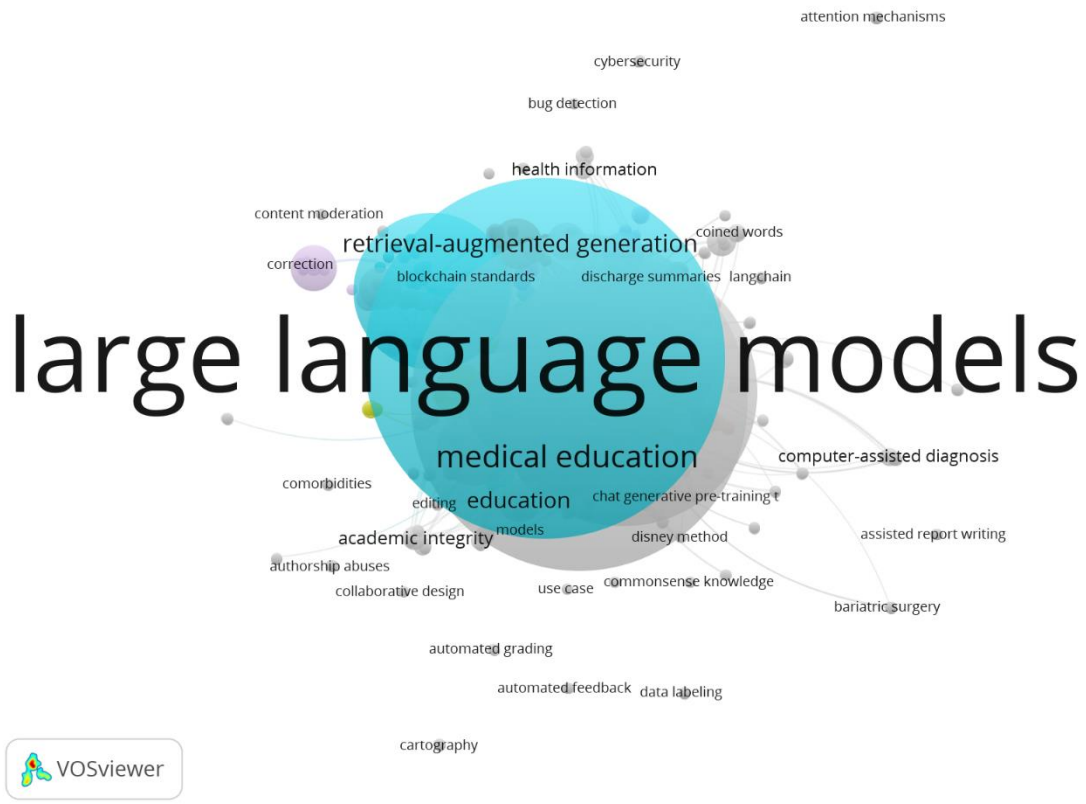


Figure 4: The Network Visualisation Map of Keywords' Co-Occurrence

RQ5: What Is The Co-Authorship By Countries Collaboration?

The data from the VOSviewer analysis reveals significant disparities in co-authorship and research influence among countries. The U.S. dominates the field with 165 documents, 7,554 citations, and a TLS of 100, indicating prolific output, high impact, and extensive international collaboration. China follows in terms of volume with 115 documents and 2,435 citations, exhibiting strong research activity, but with a lower TLS (54), suggesting more limited global collaboration relative to its output. Germany and the United Kingdom also stand out with high citation counts (3,103 and 897, respectively) and strong TLS (33 and 40), positioning them as key hubs in international research networks.

Several mid-tier countries also exhibit notable performance. Canada, with 29 documents and a TLS of 27, demonstrates active collaboration and consistent impact (402 citations). Italy (27 documents, 898 citations, TLS 25) and Australia (20 documents, 349 citations, TLS 24) are similarly positioned, indicating robust involvement in international research. Meanwhile, countries like Singapore, Switzerland, and South Korea maintain moderate levels of output but exhibit solid collaboration, as evidenced by their TLS ranging from 16 to 18. This suggests that while their publication volume may be lower, they are well integrated into global research networks.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

At the lower end of the spectrum, some countries such as Nigeria and Taiwan have minimal output (5 documents each) and very low citation impact (6 and 5 citations, respectively), yet Taiwan's TLS of 10 and Nigeria's 9 imply moderate levels of international collaboration. Interestingly, Ireland stands out with only five documents but a high citation count (1,143), indicating high-impact work possibly from a small number of influential publications. This data underscores the varying roles countries play in global research, from prolific contributors and major collaborators to niche players with high-impact outputs.

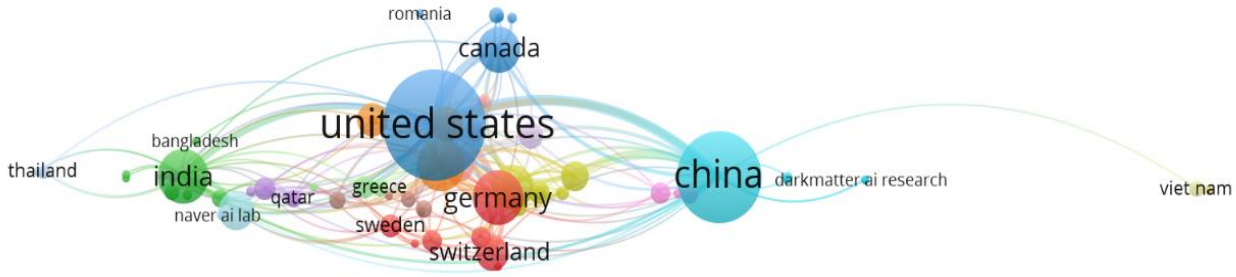


Figure 5: The Co-Authorship by Countries Collaboration

Conclusion

The purpose of this bibliometric review was to explore and synthesise the academic discourse surrounding LLMs and the issue of hallucinations within the broader field of AI. Therefore, by focusing on key terms such as "LLM," "hallucination," and "AI," the analysis aimed to identify publication patterns, influential works, key contributors, and thematic clusters to understand the current research landscape better. The findings reveal a significant increase in scholarly attention, particularly between 2023 and 2024, reflecting heightened concern over the reliability of LLM outputs and the implications of hallucinated content in critical domains such as healthcare, education, and public policy.

The analysis of 513 publications from the Scopus database demonstrated that the U.S. and China are the most active contributors, with notable collaboration networks spanning across Europe, Asia, and Oceania. Keyword co-occurrence maps indicated a strong research focus on generative AI, prompt engineering, and ethical concerns, while citation analysis identified several highly influential studies addressing both the opportunities and risks of LLM deployment. VOSviewer visualisations further highlighted emerging interdisciplinary interest and the formation of distinct thematic research clusters. These insights contribute valuable perspectives to the evolving discourse on LLM reliability and its intersection with societal trust in AI technologies.

This study provides a structured overview that strengthens understanding of research directions, identifies knowledge gaps, and aids future scholarly planning. The implications extend to AI developers, policymakers, and educators who must navigate the dual challenge of harnessing LLM capabilities while mitigating their limitations. However, the scope was constrained by the reliance on a single database and a limited temporal range, which may have excluded relevant literature outside the specified parameters. Future research could expand the dataset to include other indexing platforms and explore longitudinal developments beyond 2025.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014

In summary, this bibliometric review underscores the growing significance of LLM hallucinations as a research priority. The systematic mapping of trends, collaborations, and emerging themes enhances collective understanding and supports informed decision-making across disciplines. The use of bibliometric methods proves instrumental in charting complex research terrains and guiding subsequent inquiry into responsible and effective AI integration.

Acknowledgements

The authors would like to extend their heartfelt gratitude to Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, for granting permission and providing the necessary support to conduct this study. The university's encouragement and facilitation of academic exploration have been invaluable in advancing our research on LLM and hallucinations and their implications across various domains.

References

Al-Khoury, A., Hussein, S. A., Abdulwhab, M., Aljuboori, Z. M., Haddad, H., Ali, M. A., Abed, I. A., & Flayyih, H. H. (2022). Intellectual Capital History and Trends: A Bibliometric Analysis Using Scopus Database. Sustainability (Switzerland), 14(18). https://doi.org/10.3390/su141811615

Alves, J. L., Borges, I. B., & De Nadae, J. (2021). Sustainability in complex projects of civil construction: Bibliometric and bibliographic review. Gestao e Producao, 28(4). https://doi.org/10.1590/1806-9649-2020v28e5389

Appio, F. P., Cesaroni, F., & Di Minin, A. (2014). Visualising the structure and bridges of the intellectual property management and strategy literature: a document co-citation analysis. Scientometrics, 101(1), 623–661. https://doi.org/10.1007/s11192-014-1329-0

Assyakur, D. S., & Rosa, E. M. (2022). Spiritual Leadership in Healthcare: A Bibliometric Analysis. Jurnal Aisyah : Jurnal Ilmu Kesehatan, 7(2). https://doi.org/10.30604/jika.v7i2.914

Bruno, A., Mazzeo, P. L., Chetouani, A., Tliba, M., & Kerkouri, M. A. (2023). Insights into Classifying and Mitigating LLMs' Hallucinations. In B. A., V. C. B. 1 IULM University Milan, P. A., V. delle S. University of Palermo Palermo, M. R., V. C. B. 1 IULM University Milan, A. A., V. U. L. M. ICAR Institute for High Performance Computing and Networking CNR 153, Palermo, M. P.L., V. M. sn ISASI Institute of Applied Sciences and Intelligent Systems CNR Lecce, V. F., V. U. L. M. ICAR Institute for High Performance Computing and Networking CNR 153, Palermo, C. A., & V. delle S. University of Palermo Palermo (Eds.), CEUR Workshop Proceedings (Vol. 3563, pp. 50–63). CEUR-WS.

De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. Frontiers in Public Health, 11. https://doi.org/10.3389/fpubh.2023.1166120

Di Ieva, A., Stewart, C., & Suero Molina, E. (2024). Large Language Models in Neurosurgery. In Advances in Experimental Medicine and Biology (Vol. 1462, pp. 177–198). Springer. https://doi.org/10.1007/978-3-031-64892-2_11

di Stefano, G., Peteraf, M., & Veronay, G. (2010). Dynamic capabilities deconstructed: A bibliographic investigation into the origins, development, and future directions of the research domain. Industrial and Corporate Change, 19(4), 1187–1204. https://doi.org/10.1093/icc/dtq027

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014
Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. In International Journal of Production Economics (Vol. 162, pp. 101–114). https://doi.org/10.1016/j.ijpe.2015.01.003

Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Medical Education, 9. https://doi.org/10.2196/45312

Gu, D., Li, T., Wang, X., Yang, X., & Yu, Z. (2019). Visualising the intellectual structure and evolution of electronic health and telemedicine research. International Journal of Medical Informatics, 130. https://doi.org/10.1016/j.ijmedinf.2019.08.007

He, J., Shen, Y., & Xie, R. (2025). Research on Categorical Recognition and Optimization of Hallucination Phenomenon in Large Language Models. Journal of Frontiers of Computer Science and Technology, 19(5), 1295–1301. https://doi.org/10.3778/j.issn.1673-9418.2408080

Ho, H.-T., Ly, D.-T., & Nguyen, L. V. (2024). Mitigating Hallucinations in Large Language Models for Educational Application. 2024 IEEE International Conference on Consumer Electronics-Asia, ICCE-Asia 2024. https://doi.org/10.1109/ICCE-Asia63397.2024.10773965

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2025). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems, 43(2). https://doi.org/10.1145/3703155

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Education and Information Technologies, 28(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Jin, W., Gao, Y., Tao, T., Wang, X., Wang, N., Wu, B., & Zhao, B. (2025). Veracity-Oriented Context-Aware Large Language Models–Based Prompting Optimization for Fake News Detection. International Journal of Intelligent Systems, 2025(1). https://doi.org/10.1155/int/5920142

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103. https://doi.org/10.1016/j.lindif.2023.102274

Khiste, G. P., & Paithankar, R. R. (2017). Analysis of Bibliometric term in Scopus. International Research Journal, 01(32), 78–83.

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health, 2(2 February). https://doi.org/10.1371/journal.pdig.0000198

Liang, X., Song, S., Niu, S., Li, Z., Xiong, F., Tang, B., Wang, Y., He, D., Cheng, P., Wang, Z., & Deng, H. (2024). UHGEval: Benchmarking the Hallucination of Chinese Large Language Models via Unconstrained Generation. In K. L.-W., M. A.F.T., & S. V. (Eds.), Proceedings of the Annual Meeting of the Association for Computational Linguistics (Vol. 1, pp. 5266–5293). Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/2024.acl-long.288

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014
Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-Related research and perspective towards the future of large language models. Meta-Radiology, 1(2). https://doi.org/10.1016/j.metrad.2023.100017

Lu, J., & Li, S. (2024). Roberta with Low-Rank Adaptation and Hierarchical Attention for Hallucination Detection in LLMs. 2024 International Conference on Image Processing, Computer Vision and Machine Learning, ICICML 2024, 1532–1536. https://doi.org/10.1109/ICICML63543.2024.10957858

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. Journal of the Association for Information Science and Technology, 74(5), 570–581. https://doi.org/10.1002/asi.24750

Maleki, N., Padmanabhan, B., & Dutta, K. (2024). AI Hallucinations: A Misnomer Worth Clarifying. Proceedings - 2024 IEEE Conference on Artificial Intelligence, CAI 2024, 133–138. https://doi.org/10.1109/CAI59869.2024.00033

Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O'Connor, K., Li, R., Peng, P.-C., Bright, T. J., Tatonetti, N., Won, K. J., Gonzalez-Hernandez, G., & Moore, J. H. (2023). ChatGPT and large language models in academia: opportunities and challenges. BioData Mining, 16(1). https://doi.org/10.1186/s13040-023-00339-9

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice, 20(2). https://doi.org/10.53761/1.20.02.07

Pons, G., Bilalli, B., & Queralt, A. (2025). Knowledge Graphs for Enhancing Large Language Models in Entity Disambiguation. In D. G., H. K., A. M., P. M., C. G., S.-M. H., F. N., H. D., & H. A. (Eds.), Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) : Vol. 15231 LNCS (pp. 162–179). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-77844-5_9

Reddy, G. P., Pavan Kumar, Y. V, & Prakash, K. P. (2024). Hallucinations in Large Language Models (LLMs). In N. D., P. S., S. T., & U. D. (Eds.), 2024 IEEE Open Conference of Electrical, Electronic and Information Sciences, eStream 2024 - Proceedings. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/eStream61684.2024.10542617

Sakib, S. M. N. (2024). Bane and boon of hallucinations in the context of generative AI. In Cases on AI Ethics in Business (pp. 276–299). IGI Global. https://doi.org/10.4018/9798369326435.ch016

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and Other Large Language Models Are Double-edged Swords. Radiology, 307(2). https://doi.org/10.1148/RADIOL.230163

Van Eck, N. J., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. International Journal of Uncertainty, Fuzziness and Knowldege-Based Systems, 15(5), 625–645. https://doi.org/10.1142/S0218488507004911

van Eck, N. J., & Waltman, L. (2010a). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, N. J., & Waltman, L. (2010b). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 10 Issue 59 (September 2025) PP.185-200
DOI 10.35631/IJEPC.1059014
van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics, 111(2), 1053–1070. https://doi.org/10.1007/s11192-017-2300-7

Verbeek, A., Debackere, K., Luwel, M., & Zimmermann, E. (2002). Measuring progress and evolution in science and technology - I: The multiple uses of bibliometric indicators. International Journal of Management Reviews, 4(2), 179–211. https://doi.org/10.1111/1468-2370.00083

Wang, S., Cooper, N., & Eby, M. (2024). From human-centered to social-centered artificial intelligence: Assessing ChatGPT's impact through disruptive events. Big Data and Society, 11(4). https://doi.org/10.1177/20539517241290220

Wu, Y. C. J., & Wu, T. (2017). A decade of entrepreneurship education in the Asia Pacific for future directions in theory and practice. In Management Decision (Vol. 55, Issue 7, pp. 1333–1350). https://doi.org/10.1108/MD-05-2017-0518