

LETTER TO EDITOR

Hallucination: A key challenge to Artificial Intelligence-Generated writing

Jazlan Jamaluddin, Nadia Abd Gaffar, Nor Shazatul Salwana Din

Jamaluddin J, Gaffar NA, Din NSS. Hallucination: A key challenge to Artificial Intelligence-Generated writing. *Malays Fam Physician*. 2023;18:68.
<https://doi.org/10.51866/lte.527>

Keywords:

Artificial intelligence,
Hallucinations, Delusions,
Medical writing, Biomedical
research

Authors:

Jazlan Jamaluddin

(Corresponding author)
MD, MMed
Klinik Kesihatan Sauk, Jalan Besar
Lenggong, Sauk, Kuala Kangsar,
Perak, Malaysia.
Email: jazlanjamaluddin@gmail.com

Nadia Abd Gaffar

MD, MMed
Klinik Kesihatan Tanjung Malim
Jalan Besar, Tanjung Malim, Perak,
Malaysia.

Nor Shazatul Salwana Din

MD, MMed
Klinik Kesihatan Bota Kiri, Jalan Bota
Kiri, Bota Kiri, Perak Tengah, Perak,
Malaysia.

Open Access: This is an Open Access article licensed under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original author(s) and source are properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

Dear editor,

We read the article by Lee et al. with great interest.¹ Artificial intelligence (AI) has become an indispensable tool not only in academic writing, particularly among students, trainees and other related individuals, but also in clinical and clerical work, especially in resource-limited and rural clinics, where it continues to aid clinicians in drafting letters, memos, dossiers and other documents. With the increasing use of AI, including but not limited to systems such as ChatGPT, in various facets of daily life, it is crucial to address the issue of 'hallucination' in AI-generated content. This issue has been gaining prominence in the field of AI this year after the launch of ChatGPT in late November 2022.²

In AI, hallucination (also known as artificial hallucination or confabulation) occurs when responses given are not factually accurate or do not correspond to any real-world data.^{2,3} It arises when a generative model is asked to produce text, image or other content without strict constraints or clear guidance. Generative models are trained on vast amounts of data and learn to generate content based on patterns they have observed in such data. When they generate content without specific guidelines, they may draw upon unusual or unexpected patterns they have learnt, leading to outputs that seem strange, surreal or disconnected from reality. This issue can occur owing to errors in AI systems' training data, overfitting of AI systems and a lack of understanding of the underlying biomedical concepts. This becomes especially problematic when citing references or sources that are non-existent or inaccurate. While AI tools can be used for citation and reference management, many of the references generated are often fabricated.⁴ When using AI-generated content, authors need to exercise vigilance when AI references external sources. The output provided by AI often depends on the prompts given by its users, the version of model used and its ability to access the internet. It is not enough to simply rely on AI's ability to generate citations. Authors should take the time to find and read through the referenced articles, ensuring their accuracy and relevance. This additional step is essential to maintain the integrity of the academic and professional work produced with the assistance of AI. Unfortunately, there have been increasing instances of authors simply referencing articles without reading them or, worse, referencing non-existent articles. This phenomenon has been reported worldwide and poses a significant challenge to the credibility of scholarly publications and even official documentation.^{3,5,6}

As Lee et al. are part of the editorial team of this esteemed journal, it is important to also emphasise the proactive role of editors and reviewers in addressing this issue.² When evaluating manuscripts submitted to the journal, editors and reviewers must double-check not only the content but also the references provided. The easiest way to confirm the validity of a reference is by visiting the webpage link provided. In addition to manual verification, there are various software tools available that can help confirm whether AI has been used in the writing of an article, most of which have been included in plagiarism detection software. These tools can assist in identifying potential instances of hallucinations in AI-generated content. Such software can be valuable in the fight against misinformation and inaccuracies in academic and professional discourses.

Just as hallucinations in human cognition are acknowledged and addressed as part of World Mental Health Day every year on October 10, it is equally important to recognise and address the phenomenon of hallucinations in AI-generated content. Notably, AI hallucinations are not actual hallucinations in the psychological sense, as they are not experienced by AI

tools themselves but rather a result of models' pattern recognition and generation processes. Researchers and developers are continually working to improve AI systems and reduce the occurrence of such hallucinatory outputs while making them more useful and aligned with human intent. However, managing hallucinations in AI is akin to managing cognitive distortions in human thinking. Authors must take the responsibility to corroborate the facts provided by AI and ensure the accuracy of their references. As the use of AI continues to grow in various aspects of life, including academia and professional work, it is imperative to address the issue of hallucinations in AI-generated content. In the realm of clinical papers and research, although there exists a genuine requirement for a validated AI tool designed specifically for referencing and citation purposes, authors, editors and reviewers all have roles to play in mitigating this problem and upholding the standards of accuracy and integrity in publications. Human touch remains important when using AI, and AI-generated content should be used with caution. While advancements in AI hold promise for increased reliability, complete replacement of humans by AI remains improbable at this point in time.

Acknowledgements

In the same manner as in the article by Lee et al., ChatGPT (version 3.5; OpenAI) was used to generate the preliminary text of this article, which was then revised by the authors. The references were manually cited using referencing software and checked by the authors to avoid hallucinations by AI.

Author contributions

JJ conceptualised the study, reviewed the literature and wrote the manuscript. NAG and NSSD interpreted the relevant literature and revised the manuscript. All authors approved the final version of this manuscript.

Conflicts of interest

None.

Funding

None.

References

1. Lee PY, Salim H, Abdullah A, Teo CH. Use of ChatGPT in medical research and scientific writing. *Malays Fam Physician*. 2023;18:58. doi:10.51866/cm0006
2. Ang T, Choolani M, See K, Poh K. The rise of artificial intelligence: addressing the impact of large language models such as ChatGPT on scientific publications. *Singapore Med J*. 2023;64(4):219. doi:10.4103/singaporemedj.SMJ-2023-055
3. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagaadda V, Dave T, Duddumpudi RTS. Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*. 2023;15(4):e37432. Published 2023 Apr 11. doi:10.7759/cureus.37432
4. Gravel J, D'Amours-Gravel M, Osmanlliu E. Learning to fake it: limited responses and fabricated references provided by ChatGPT for medical questions. *Mayo Clin Proc Digit Heal*. 2023;1(3):226–234. doi:10.1016/j.mcpdig.2023.05.004
5. Jansz J, Tran HW, Sweiss NJ. Pitfalls in the Diagnosis and Management of an Unusual Presentation of Clinically Amyopathic Dermatomyositis: A Case Report Written With the Assistance of ChatGPT. *Cureus*. 2023;15(7):e41879. Published 2023 Jul 14. doi:10.7759/cureus.41879
6. Milmo D, agency. Two US lawyers fined for submitting fake court citations from ChatGPT. *The Guardian*. Published June 23, 2023. Accessed October 1, 2023. <https://www.theguardian.com/technology/2023/jun/23/two-us-lawyers-fined-submitting-fake-court-citations-chatgpt>