

ChatGPT: Challenges to editors and reviewers

Bibha Simkhada^{1*}, Alexander van Teijlingen², Padam Simkhada^{1,3}, Edwin van Teijlingen^{3,4}

¹School of Human and Health Sciences, University of Huddersfield, UK

²Pure and Applied Chemistry, the University of Strathclyde, Glasgow, UK

³Manmohan Memorial Institute of Health Sciences, Kathmandu, Nepal

⁴Centre for Midwifery & Women's Health, Bournemouth University, Bournemouth, UK

Received:

18 December 2023

Revised:

11 January 2024

Accepted:

1 April 2024

*Corresponding author

Edwin van Teijlingen
evteijlingen@bournemouth.ac.uk

DOI: 10.3126/hprospect.v23i1.60819

Abstract

The past year saw an exponential growth in the use of machine learning using AI (artificial intelligence) and particularly Generative AI (GenAI) such as ChatGPT. The latter has seen a spectacular rise in the public debate and in the mass media. Those not involved in the development of AI were amazed by the capabilities of ChatGPT to produce text equal to the average human produced texts. There is no doubt that the adoption of AI is advancing rapidly.

To test the ability of ChatGPT in its free version, we posed simple questions about the topic we had previously published. After reading the short essay produced by ChatGPT we repeated the question whilst asking for references to be included. We were surprised by the quality of this very general piece of work.

In many UK universities there is a debate starting about students' use of ChatGPT, and how difficult it is to distinguish between work produced by the average student and that produced by AI. There is a similar problem for editors and reviewers of academic journals. It really boils down to the question: 'How can you be certain the submitted manuscript came from a human source?' However, we feel the progress of AI is not all doom and gloom. We outline some of the key problems around AI and academic publishing, but also opportunities arising from the use of AI in this area.

Keywords: ChatGPT, AI, plagiarism, Artificial Intelligence, detection, academic writing

Tweetable Abstract: Labor-saving tech for writing Background or Introduction is logical but ensuring authors get credit is crucial as AI's co-author status is uncertain.

Introduction

The year 2023 has been the year of the revelation of the abilities and power of Artificial Intelligence (AI) text-generating technology. The past year has seen a flurry of media coverage about AI generated text. This has been made possible by the recent dramatic advancements in Large Language Models (LLMs), with ChatGPT (Generative Pre-Trained Transformer) being the most well-known one, and others such as Bard and Galactica [1-2], LLMs contain deep neural networks with extremely large parameter sets that are trained on vast datasets in a very resource-intensive process. LLMs can recognize, summarize, translate, predict, and generate texts and other content. In the media, ChatGPT especially has been highlighted as a treat to some traditional forms of written assessments used in higher education. With several newspaper stories of students using ChatGPT to 'write' their essays [3-5].

Our experiment

In order to test the ability of AI to produce work in the area of our academic expertise, we devised a test for ChatGPT to assess the quality of its work. We wanted to assess ChatGPT's capability to produce content within a particular academic domain. Specifically, the experiment aimed to determine how effectively ChatGPT could simulate the type and caliber of work expected from an undergraduate student or manuscripts submitted to academic journals. Therefore, we did an experiment using the following three sets of key words on ChatGPT: labour migrant and health issues and Nepal. The text in Box 1 was automatically generated using those instructions.

BOX 1: AI instructions: labour migrant AND health issues AND Nepal

Nepal is a country that has seen a significant number of its citizens migrating to foreign countries for employment opportunities. These labor migrants often face various health issues that can have long-term consequences for their well-being.

One of the major health concerns faced by labor migrants from Nepal is the risk of infectious diseases. Many migrants work in environments with poor sanitation and hygiene, making them more susceptible to diseases such as tuberculosis, hepatitis, and HIV/AIDS. Furthermore, many labor migrants from Nepal work in industries such as construction, where they are exposed to hazardous materials and work conditions that can lead to occupational illnesses.

Mental health is also a significant issue for labor migrants from Nepal. These individuals often leave their families and support systems behind and are subjected to long hours of work, isolation, and discrimination. This can lead to feelings of loneliness, depression, and anxiety, which can have long-term effects on their mental health.

Finally, labor migrants from Nepal also face significant challenges when accessing healthcare in their host countries. They may not be familiar with the healthcare system, face language barriers, or lack access to healthcare services due to their immigration status. This can result in delayed diagnosis and treatment of health issues, which can have serious consequences.

To address these health issues faced by labor migrants from Nepal, it is essential to provide them with adequate healthcare services, including preventative measures such as vaccinations and education on basic hygiene and sanitation practices. Furthermore, it is necessary to develop policies that protect the rights of these workers and ensure that they have access to adequate working conditions, including proper health and safety standards. Finally, it is crucial to support the mental health needs of labor migrants through counseling services and support networks.

Upon reviewing the contents of Box 1, our primary concern is that the text appears to be fairly appropriate. In other words, it is what we might expect from an undergraduate student who is below average in terms of writing proficiency. In our opinion, this essay reflects a reasonably well-informed opinion on the topic at hand. However, we were surprised by the lack of references in the text. Consequently, we repeated the request to ChatGPT, asking them to add references. This time, we used the same keywords: "labour migrant," "health issues," and "Nepal," along with the instruction to add in-text citations.

BOX 2 AI instructions: labour migrant AND health issues AND Nepal AND in-text citation

Labor migration is a significant issue in Nepal, with a large number of Nepali citizens migrating to foreign countries for employment opportunities (Dahal, 2017). These labor migrants often face various health issues that can have long-term consequences for their well-being. One of the major health concerns faced by labor migrants from Nepal is the risk of infectious diseases (Dahal, 2017). Many migrants work in environments with poor sanitation and hygiene, making them more susceptible to diseases such as tuberculosis, hepatitis, and HIV/AIDS. Furthermore, many labor migrants from Nepal work in industries such as construction, where they are exposed to hazardous materials and work conditions that can lead to occupational illnesses (Choudhury & Karmacharya, 2015).

Mental health is also a significant issue for labor migrants from Nepal. These individuals often leave their families and support systems behind and are subjected to long hours of work, isolation, and discrimination (Sharma & Thapa, 2019). This can lead to feelings of loneliness, depression, and anxiety, which can have long-term effects on their mental health.

Finally, labor migrants from Nepal also face significant challenges when accessing healthcare in their host countries. They may not be familiar with the healthcare system, face language barriers, or lack access to healthcare services due to their immigration status (Dahal, 2017). This can result in delayed diagnosis and treatment of health issues, which can have serious consequences.

To address these health issues faced by labor migrants from Nepal, it is essential to provide them with adequate healthcare services, including preventative measures such as vaccinations and education on basic hygiene and sanitation practices. Furthermore, it is necessary to develop policies that protect the rights of these workers and ensure that they have access to adequate working conditions, including proper health and safety standards (Choudhury & Karmacharya, 2015). Finally, it is crucial to support the mental health needs of labor migrants through counseling services and support networks (Sharma & Thapa, 2019).

References:

1. Choudhury, N., & Karmacharya, B. M. (2015). Occupational health problems of migrant workers in the construction industry of Nepal. *Journal of occupational health*, 57(1), 64-71. <https://doi.org/10.1539/joh.14-0148-FS>
2. Dahal, G. P. (2017). Health and well-being of Nepali migrant workers abroad. *International Journal of Migration, Health and Social Care*, 13(1), 1-14. <https://doi.org/10.1108/IJMHSC-06-2016-0024>
3. Sharma, S., & Thapa, P. (2019). Migration and mental health: A study of Nepali migrant workers in the United Arab Emirates. *International Journal of Mental Health*, 48(4), 357-374. <https://doi.org/10.1080/00207411.2018.1491819>

We noted that the three references in Box 2 are slightly older than we would have liked. In addition, the reference list does not include any of the 30-odd papers we have published on this very topic [6]. This might be partly explained by the fact that the LLM was only trained on data existing before 2021. Furthermore, although the references are appropriately listed in alphabetical order, we noticed that one of the three journal names is provided in a different style from the other, as it does not have opening capitals [7]. Of course, this is a very common mistake. This brought us to the question 'How can we identify the use of AI in academic work?'

Spotting the use of AI: plagiarism detection

Our little experiment in Boxes 1 and 2 made us realize that spotting the use of AI used to produce academic texts is going to be very difficult. Current plagiarism detection relies on matching the similarity of students'/researchers' work to existing texts. This works well on human-written text which has been copied. Such plagiarism-detection software can it quickly and automatically mark plagiarism, whilst providing evidence that such text has been plagiarised, i.e. the sources copied. This system works in the case of online LLMs where a pla-

gism checking tool can be provided a backdoor into the logs of all conversations kept on various websites such as ChatGPT. However, while it takes a wealth of resources to train these models, they can quite readily be installed on a home computer and yield no evidence to any servers a plagiarism detector might have access to.

Although the media have reported recently on the use AI plagiarism detection software [8-9], AI content detection is currently weak. This is partly because it can be used to further refine LLMs until they no longer produce AI-detectable content, but possibly more importantly because of the so-called false positives. The latter means that the current AI detection software flags so much material written by humans (i.e. clearly non-AI generator content) as AI generated and does not provide evidence for their assertion. This is further compounded by the fact that LLMs don't tend to give the same answer twice to the same question and the 'creativity' of the responses can be selected in many available LLMs (this is known as the temperature parameter), though selecting the temperature is not an available feature in ChatGPT at this time.

Again, we did an experiment, this time using the beginning part of the American Constitution, which we know has been written by hand. Here, we tested the original text of the USA Constitution (up to Section 2 of article 1 due to word limits of the free software) in writer.com, and zeroGPT.com which gave responses of 64 % and 94.86 % AI written whilst copyLeaks.com correctly identified it as "human text". This can lead to the situation of honest students being unfairly penalized for having used LLMs to write their work when they are not, and dishonest students using LLMs to produce their work and when confronted, claiming they have not and that the burden to prove it should lay, quite rightly so, with the university/publisher.

Ultimately, however, as with every new disruptive technology LLMs are useful tools which can be the writer's friend rather than enemy, we already use technologies that may seem like cheating to researchers of the previous century. For instance, we have used several tools that were not available to researchers even a few decades ago, including Google search and Microsoft Word's predictive text (neither ChatGPT nor any other LLM has produced any of our text, except where examples are noted in Boxes 1 & 2).

Outside of the question of plagiarism, one of the main problems with LLMs are that they are language models, not databases of facts and often present wrong data very confidently, which leads people to believe they are authoritative sources. This can be problematic as a study-aid, and creates wider problem of propagating mis- and disinformation which is then more widely presented and spread.

Therefore, they seem to produce better answers to non-quantitative questions such as seen above. An illustration of the latter would be when asked in 2023 to give the pKa (acidity) of oleic acid, ChatGPT gave the answers of 4.8, 9.85 and 10.5 which are all potentially correct, but each in a different setting [10-12], though the typical textbook definition of pKa would only accept a value around 4.8. To a non-chemist, a pKa of 10.5 may not immediately appear wrong (or require the justification that this would not be typically what is meant by pKa and would only occur under certain conditions) and they may use this text to post misinformation in other places around the internet. Interestingly, and illustrating the rapid development of LLMs, in February 2024 ChatGPT, just months after the first draft of this Viewpoint, ChatGPT4 was answering this questions much better!

The problems for editors and reviewers

We fear it is also a threat to journal editors and reviewers, who will find it much more difficult to determine whether a submitted paper is based on original work done by the authors. Maintaining consistency, style, and expressions can be tricky when working with AI-generated text, specifically across multiple sections of a document. There might be a more basic issue for both editors and reviewers, namely the sheer volume of papers generated using AI models such as ChatGPT and which may lack originality and expression in writing.

If we agree that it is not desirable for academic papers to be written by AI, the next question we need to ask is whether it is acceptable for non-native English speakers to use AI to improve the readability of their academically sound but poorly written papers. Should we set a limit on the proportion of AI use that is acceptable? We do recognize that improving readability increases the likelihood of acceptance by the journal editor.

It is important to note that most journal editors and reviewers are volunteers who are generally unpaid, poorly trained and doing the reviewing and editorial work in their spare time [13]. For journal editors, it is difficult to find enough reviewers for each manuscript and chase them to ensure timely submission of their comments. For reviewers, it is already challenging to assess whether genuine researchers have conducted the research to the highest standards or have cut corners. It is an additional burden to consider whether the manuscript under review is mostly written by a machine.

A further challenge includes 'lazy' reviewers running a submitted paper through an AI system for comments, insights, similarities with existing literature, etc. This brings with it risks of AI systems being based on biased data, or at least dominant views, which may result in rejecting innovative papers and potentially blocking the publication of genuinely new insights and solutions. Moreover, AI systems may not be able to understand the subtleties of, for example, different social science theories or technical differences in drug delivery systems in the human body. There is also the risk that reviewers may use AI to break the double-blind anonymous peer reviewing of their allocated paper to find out who the authors are or to which research group/institution they belong. All this may undermine the peer review process.

We do also see some potential benefits, AI systems can help editors save time finding the most appropriate available peer reviewers, based on information available on their publications, web presence, social media postings, etc. We also think AI can be trained to assess the quality of submitted peer reviews, which would give editors greater or lesser confidence in certain reviews and may reduce the impact of potential biases or personal preferences of reviewers.

It's true that plagiarism detection software currently in use mainly focuses on identifying the quantity of overlap between submitted

text and existing literature. However, with the help of AI technology, it is possible to train systems to identify more nuanced forms of plagiarism or the absence of originality in a submission. This could greatly improve the accuracy and effectiveness of plagiarism detection.

Conclusion

The world is changing, university staff may have to move away from using essays as a key assessment tool. An LLM can be a valuable resource for students and researchers to fact-check and generate ideas. Journal editors can start requiring a statement from potential authors to declare if they have used LLMs or even to highlight sections written by LLMs. Journal editors can consider requesting a statement from potential authors to disclose if they have used Language Model Models (LLMs) or highlight sections written by LLMs. In the long term, it is logical to embrace labor-saving technologies that assist aspiring authors in writing a Background or Introduction section for their paper. However, it is crucial to ensure that authors receive proper credit, and it remains uncertain whether AI can, should, or will be considered a co-author.

Acknowledgments

Some of the text in this paper has been created with the aid of AI; the relevant text has been clearly marked in Boxes 1 and 2 with single line spacing.

References

1. Neto JAR. ChatGPT and the Large Language Models (LLMs). Medium, 2023 (online: <https://medium.com/chatgpt-learning/chatgpt-and-the-large-language-models-llms-c69bfd48347>).
2. Ornes S. The Unpredictable Abilities Emerging From Large AI Models. Quantamagazine 2023 (online: <https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/>).
3. Wehner G. UK college student uses AI to write high-scoring essay, earns high grade: report, Fox Business 10 April 2023, (online <https://www.foxbusiness.com/technology/uk-college-student-ai-write-high-scoring-essay-report>).
4. Weale S. Lecturers urged to review assessments in UK amid concerns over new AI tool, The Guardian 13 January 2023 (online <https://www.theguardian.com/technology/2023/jan/13/end-of-the-essay-uk-lecturers-assessments-chatgpt-concerns-ai>).
5. Reich R. Now AI can write students' essays for them, will everyone become a cheat?, The Guardian 28 November 2022, (online: <https://www.theguardian.com/commentisfree/2022/nov/28/ai-students-essays-cheat-teachers-plagiarism-tech>).
6. Wasti, S.P., Shrestha, A., Atteraya, M.S., GC, V.S. (2023) Migrant workers' health-related research in Nepal: A bibliometric study, *Dialogues in Health* 3: 100147, <https://doi.org/10.1016/j.dialog.2023.100147>.
7. Choudhury N, Karmacharya B-M. Occupational health problems of migrant workers in the construction industry of Nepal. *Journal of Occupational Health*, 2015;57(1): 64-71. <https://doi.org/10.1539/joh.14-0148-FS>
8. Bowman E. A college student created an app that can tell whether AI wrote an essay. NPR, 2023 (online: <https://www.npr.org/2023/01/09/1147549845/gptzero-ai-chatgpt-edward-tian-plagiarism>).
9. Pinkstone J. University plagiarism detector will catch students who cheat with ChatGPT, The Telegraph 4 April 2023, (online: <https://www.telegraph.co.uk/news/2023/04/04/chatgpt-cheat-students-plagiarism-universities-ai/>).
10. van Teijlingen A, Swanson HWA, Lau KHA, Tuttle T. Constant pH Coarse-Grained Molecular Dynamics with Stochastic Charge Neutralization. *Journal of Physical Chemistry Letters*, 2022;13(18): 4046–4051. <https://doi.org/10.1021/acs.jpcllett.2c00544>.
11. Vila-Viçosa D, Teixeira VH, Baptista AM, Machuqueiro M. Constant-pH MD Simulations of an Oleic Acid Bilayer. *Journal of Chemical Theory & Computation*, 2015;11 (5): 2367–2376. <https://doi.org/10.1021/acs.jctc.5b00095>.
12. Grünewald F, Souza PCT, Abdizadeh H, Barnoud J, de Vries AH, Marrink SJ. Titratable Martini Model for Constant pH Simulations. *Journal of Chemical Physics*, 2020;153(2): 024118. <https://doi.org/10.1063/5.0014258>.
13. van Teijlingen E, Thapa D, Marahatta SB, Sapkota JL, Regmi P Sathian B. Editors and Reviewers: Roles and Responsibilities, In: Wasti, S.P., van Teijlingen, E., Simkhada, P.P., Hundley, V. with Shreesh, K. (Eds.) *Academic Writing and Publishing in Health & Social Sciences*, Kathmandu, Nepal: Himal Books, 2022: 32-37.