

Wikipedia and Large Language Models: Perfect Pairing or Perfect Storm?

By Paul A. Thomas*
ORCID: [0000-0002-5596-7951](https://orcid.org/0000-0002-5596-7951)

Scheduled to be published as: Thomas, Paul A. "Wikipedia and Large Language Models: Perfect Pairing or Perfect Storm?" *Library Hi Tech News*, 2023.
<http://doi.org/10.1108/LHTN-03-2023-0056>.

Abstract

Purpose: The purpose of this paper is to explore the potential benefits and challenges of using large language models (LLMs) like ChatGPT to edit Wikipedia.

Approach: The first portion of this paper provides background about Wikipedia and LLMs, explicating briefly how each works. The paper's second section then explores both the ways that LLMs can be used to make Wikipedia a stronger site and the challenges that these technologies pose to Wikipedia editors. The paper's final section explores the implications for information professionals.

Findings: The paper argues that LLMs can be used to proofread Wikipedia articles, outline potential articles, and generate usable Wikitext. The pitfalls include the technology's potential to generate text that is plagiarized or violates copyright, its tendency to produce "original research," and its tendency to generate incorrect or biased information.

Originality: While there has been limited discussion among Wikipedia editors about the use of LLMs when editing the site, hardly any scholarship has been given to how these models can impact Wikipedia's development and quality. This paper thus aims to fill this gap in knowledge by examining both the potential benefits and pitfalls of using LLMs on Wikipedia.

Keywords

Wikipedia, large language models, ChatGPT, artificial intelligence, AI

* Paul A. Thomas is a library specialist at the University of Kansas (paulthomas@ku.edu). He holds a Doctor of Philosophy (Ph.D.) degree from the School of Library and Information Management, Emporia State University.

Introduction

Late 2022 saw an explosion of interest in "large language models" (LLMs)—a specialized term that refers to artificial intelligence (AI) models that have been trained on large corpora of text and, via neural network-powered deep learning, can generate coherent text in response to user queries (Lund and Wang, 2023). As Sun and Hoelscher (2023) note, these models are revolutionary given their ability to "process vast amounts of information and generate large amounts of coherent and informative text in real time, based on simple, even short, vague, or ambiguous prompts" (p. 1). Of the many LLMs that now exist, arguably the most-discussed models include OpenAI's ChatGPT (OpenAI, 2023), Google's BERT (Devlin *et al.*, 2019), Meta's LLaMA (Touvron *et al.*, 2023), and HuggingFace et al.'s BLOOM (Gibney, 2022).

Understandably, LLMs have been both heralded and criticized by academics and information professionals for their disruptive potential (Taecharungroj, 2023), but hardly any scholarship has been given to how these models can impact the development and quality of Wikipedia, the free encyclopedia that anyone can edit. As tools that can produce paragraphs of text in only a few seconds, LLMs could be used to rapidly expand Wikipedia; at the same time, however, there are also many pitfalls to their use. For librarians and academics interested in both using and studying Wikipedia, it is thus imperative to consider the benefits and dangers of LLMs vis-à-vis Wikipedia. This paper aims to do just that. The first portion of this paper provides background about Wikipedia and LLMs, explicating briefly how each works. The paper's second section then explores both the ways that LLMs can be used to make Wikipedia a stronger site and the challenges that these technologies pose to Wikipedia editors. The paper's final section explores the implications for information professionals.

Background: Wikipedia and Large Language Models

"Wikipedia" is the name of a free encyclopedia that allows anyone with internet access the ability to change (or "edit") its contents. Officially launched in 2001, Wikipedia was initially lambasted by critics who argued that because anyone could edit the site, its content was almost certain to be unreliable. However, these negative evaluations were, for the most part, unfounded: In 2005, the journal *Nature* concluded that Wikipedia is just as reliable as *Encyclopædia Britannica* (Giles, 2005). Further studies by *The Guardian*, the *Journal of Clinical Oncology*, *PC Pro*, the Canadian Library Association, and *Library Journal* have all found that Wikipedia is relatively reliable (Wolchover, 2011). As of February 2023, Wikipedia is 7th most-visited website in the world, according to SimilarWeb (2023), receiving around 315.1 million visits a day (<https://w.wiki/6UJt>). In terms of size, Wikipedia is a leviathan, comprising over 6.63 million articles (<https://w.wiki/6UK4>), which have been collaboratively written by 45.23 million registered editors and millions more unregistered users (<https://w.wiki/6UK7>). Today, thanks in large part to its visibility and its comprehensiveness, Wikipedia is—whether people care to admit it or not—often the go-to destination when individuals need to locate information.

On the other hand, the term "large language model" (often abbreviated "LLM") refers to a complex language model that uses neural networks to analyze and reproduce human language. To do this, LLMs are first "trained" on corpora of millions—and, in some cases, billions—of words. The models then use deep learning to identify linguistic patterns in the training data; by processing and analyzing this data, LLMs can consequently generate human-like responses to text-based inquiries (Roose, 2023). Although LLMs have been around for some time now, it was only near the end of the beginning of the 2020s that they truly captured the zeitgeist. Much of this interest was engendered by the free release of OpenAI's impressive ChatGPT in late 2022 (Atlas, 2023), which in turn resulted in an increasing number of academics beginning to question

the wider impact of this technology (Taecharunroj, 2023). While some scholars have argued that LLMs will lead to increased productivity, others have argued that the technology raises serious ethical issues (Dwivedi, *et al.*, 2023).

Wikipedia and Large Language Models: Uses and Issues

As LLMs have grown in popularity, there has been limited but growing discussion among Wikipedia editors as to whether LLMs should or should not be used when editing the site. As with all things, the answer to this question is complicated, with the technology having the potential to both "massively improve [and] disrupt Wikipedia", as Wikipedia itself notes (<https://w.wiki/6WXw>). This section will thus examine the ways that LLMs can be used to improve Wikipedia and the issues that such use entails.

Potential Uses

Proofreading

As AI models that have been trained on billions of words, LLMs are remarkably effective editing tools (Atlas, 2023). In particular, LLMs like ChatGPT have proven themselves capable of identifying grammatical mistakes, misspellings/typos, and syntactical errors (Atlas, 2023; Sun and Hoelscher, 2023). In addition, many of these models are also capable of suggesting changes to text that will enhance clarity or how phrases can be rewritten to ensure tonal consistency (Sun and Hoelscher, 2023). Because Wikipedia has, in the past, been criticized for the quality of its prose (Rosenzweig, 2006) and the readability of certain articles (Lucassen *et al.*, 2012), LLMs could come to serve as a sort of grammatical panacea in this regard (Liu *et al.*, 2021)

Article Outlining

LLMs can also be used to generate article outlines (Atlas, 2023). For a practical demonstration consider the following: Upon being prompted to "provide an outline in the style of Wikipedia" for an article on the "Abgig Obelisk" (a topic for which I created an article without the assistance of an LLM on March 10, 2023), ChatGPT generated the following:

- Introduction
- Description
 - Physical features of the Abgig Obelisk
 - Materials used in the construction of the Abgig Obelisk
 - Inscriptions on the Abgig Obelisk
- History
 - Origins of the Abgig Obelisk
 - Early history of the Abgig Obelisk
 - Rediscovery and preservation of the Abgig Obelisk
- Significance
 - Cultural significance of the Abgig Obelisk
 - Religious significance of the Abgig Obelisk
 - Political significance of the Abgig Obelisk
- Controversies
 - Disputes over ownership of the Abgig Obelisk
 - Cultural and religious conflicts surrounding the Abgig Obelisk
 - Vandalism and damage to the Abgig Obelisk
- See also
- References

With a few exceptions, this outline is remarkably similar to the way in which I laid out my article (https://en.wikipedia.org/wiki/Abgig_obelisk). For editors who struggle with how best to organize an article, LLMs can thus be a major boon.

Wikitext Generation

For many Wikipedia neophytes, one of the most confusing aspects of the site can be Wikitext, or the markup language used to code text, templates, and tables. Luckily, many LLMs were trained on data that at least partially included Wikitext, meaning that these models can often help editors parse and generate usable markup language; in fact, Wikipedia user JPxG (2023) has documented the ways in which ChatGPT, after being presented with a sample of Wikitext, can quickly generate code that, were it implemented, would create and rotate tables, format text, and generate userboxes, among other things.

Potential Pitfalls

Plagiarism and Copyright Violation

When adding content to Wikipedia, editors must ensure that their textual contributions do not violate copyright and that they are properly sourced. The text generated by LLMs, however, often lacks citations, and because these models are trained to generate output on a probabilistic basis, they can potentially "create" content that either closely paraphrases or reproduces in full non-free text (Eliot, 2023). As such, editors who add LLM-generated text to Wikipedia without first checking to make sure that text is properly sourced and adequately paraphrased run the risk of being accused of plagiarism and/or copyright infringement.

Generation of "Original Research"

Another problem with LLMs is that, when asked to generate text on a given subject, they will often produce responses that either lack reliable sources (Zhong, 2023) or which distill down a variety of sources to reach a novel conclusion that has not been published. The issue here is that Wikipedia is epistemologically citational and thus forbids the inclusion of both "original research" (i.e., "facts, allegations, and ideas ... for which no reliable, published sources exist") and the "synthesis of published material" (i.e., the combination of "material from multiple sources to reach or imply a conclusion not explicitly stated by any source") (<https://w.wiki/6WXg>). Wikipedia's prohibition of original research exists to ensure that all the content on the site has already been vetted by experts. LLM-generated text, however, often includes assertions that have not been vetted, meaning that this text is often intrinsically unfit for inclusion in Wikipedia.

Generation of Incorrect and Biased Information

The final issue with using LLMs to generate Wikipedia content is that these models are often known to output blatantly wrong or patently biased information—and often, they do so in language that is confident and assertive, which can easily fool an incredulous user (Borji, 2023; Ji *et al.*, 2023; Marr, 2023). What is more, when asked to provide references on a given topic, some LLMs will respond with a list of sources that might look legitimate but which are actually non-existent (Gravel *et al.*, 2023). These sorts of erroneous outputs—which are colloquially known as "hallucinations" (Azamfirei *et al.*, 2023; Borji, 2023; Marr, 2023)—in and of

themselves pose a considerable risk for a world that is increasingly marred by disinformation, systemic bias, and "fake news", but the situation is even graver when these outputs are considered in the context of Wikipedia: After all, the site is extremely popular and is used by millions of people every day. If hallucinations—in the form of either erroneous "facts" or fabricated references—are uncritically added to the site, they will almost certainly be disseminated the world over, resulting in the widespread circulation of misinformation (cf. Thomas, 2021). This issue is made even more pressing given how fast LLMs can produce text, which means that Wikipedia could easily be swamped by a deluge of erroneous or biased LLM-generated content.

Implications for Information Professionals

Given that Wikipedia and LLMs are, in their own ways, resources that facilitate the dissemination of information, the overlap of these two topics is arguably of great interest to the information profession. Thanks to their technological capabilities, LLMs can pair nicely with Wikipedia, helping to make the site stronger and more readable, but uncritical widespread use of LLMs on Wikipedia could be something of the perfect storm, resulting in the addition of text that plagiarizes, violates copyright, lacks acceptable sourcing, or contains misinformation. It is a complex situation with far-reaching ramifications, but luckily, for information professionals who recognize the significance of the issue, there are several ways to mitigate any damages while clarifying the benefits: First, information professionals should increasingly discuss the benefits and pitfalls of using LLMs on Wikipedia, both amongst themselves and with the community members that they regularly help. Second, information professionals are also encouraged to move to the "front lines" (so to speak) by actively becoming Wikipedia editors. By moving from talk to action, information professionals can actually demonstrate appropriate LLM use, critically analyze content that may have been created using LLMs, and ultimately ensure that LLM-generated misinformation is not broadcast the world over.

References

- Atlas, S. (2023), *ChatGPT for higher education and professional development: A guide to conversational AI*, University of Rhode Island DigitalCommons@URI, South Kingstown, RI.
- Azamfirei, R., Kudchadkar, S.R. and Fackler, J. (2023), "Large language models and the perils of their hallucinations", *Critical Care*, Vol. 27 No. 1, pp. 1-2.
- Borji, A. (2023), "A categorical archive of ChatGPT failures", *arXiv*, doi: [10.48550/arXiv.2302.03494](https://doi.org/10.48550/arXiv.2302.03494).
- Devlin, J., Chang, M-W., Lee, K. and Toutanova, K. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding", *arXiv*, doi: [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805).
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... and Wright, R. (2023), "So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy", *International Journal of Information Management*, Vol. 71, doi: [10.1016/j.ijinfomgt.2023.102642](https://doi.org/10.1016/j.ijinfomgt.2023.102642).
- Eliot, L. (2023), "Legal doomsday for generative AI ChatGPT if caught plagiarizing or infringing, warns AI ethics and AI law", *Forbes*, available at: <https://www.forbes.com/sites/lanceeliot/2023/02/26/legal-doomsday-for-generative-ai-chatgpt-if-caught-plagiarizing-or-infringing-warns-ai-ethics-and-ai-law/?sh=5a5d8dff122b> (accessed 29 March 2023).
- Gibney, E. (2022), "Open-source language AI challenges big tech's models", *Nature*, available at: <https://www.nature.com/articles/d41586-022-01705-z> (accessed 29 March 2023)
- Giles, J. (2005), "Internet encyclopaedias go head to head", *Nature*, Vol. 438 No. 7070, pp. 900-901, doi: [10.1038/438900a](https://doi.org/10.1038/438900a).
- Gravel, J., D'Amours-Gravel, M., Osmanlliu, E. (2023), "Learning to fake it: Limited responses and fabricated references provided by ChatGPT for medical questions", *medRxiv*, doi: [10.1101/2023.03.16.23286914](https://doi.org/10.1101/2023.03.16.23286914).
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023), "Survey of hallucination in natural language generation", *ACM Computing Surveys*, Vol. 55 No. 12, 1-38.
- JPxG (2023), "User:JPxG/LLM demonstration", Wikipedia, available at https://en.wikipedia.org/w/index.php?title=User:JPxG/LLM_demonstration&oldid=1135546141 (accessed 29 March 2023).
- Liu, Y., Medlar, A. and Glowacka, D. (2021), "Can language models identify Wikipedia articles with readability and style issues?", in *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, Association for Computing Machinery, New York City, pp. 113-117.
- Lucassen, T., Dijkstra, R., Schraagen, J.M. (2012), "Readability of Wikipedia", *First Monday*, Vol. 17 No. 9, doi: [10.5210/fm.v0i0.3916](https://doi.org/10.5210/fm.v0i0.3916).
- Lund, B. and Wang, T. (2023), "Chatting about ChatGPT: How may AI and GPT impact academia and libraries?", *Library Hi Tech News*, doi: [10.1108/LHTN-01-2023-0009](https://doi.org/10.1108/LHTN-01-2023-0009).
- Marr, B. (2023), "ChatGPT: What are hallucinations and why are they a problem for AI systems", available at <https://bernardmarr.com/chatgpt-what-are-hallucinations-and-why-are-they-a-problem-for-ai-systems/> (accessed 22 March 2023).
- OpenAI (2023), "GPT-4 technical report", *arXiv*, doi: [10.48550/arXiv.2303.08774](https://doi.org/10.48550/arXiv.2303.08774).

- Rosenzweig, R. (2006), "Can history be open source? Wikipedia and the future of the past", *The Journal of American History*, Vol. 93 No. 1, 117-146.
- Roose, K. (2023), "How does ChatGPT really work?", *The New York Times*, available at <https://www.nytimes.com/2023/03/28/technology/ai-chatbots-chatgpt-bing-bard-llm.html> (accessed 29 March 2023).
- SimilarWeb (2023), "Top Websites Ranking", available at <https://www.similarweb.com/top-websites/> (accessed 29 March 2023).
- Sun, G.H. and Hoelscher, S.H. (2023), "The ChatGPT storm and what faculty can do", *Nurse Educator*, [doi:10.1097/NNE.0000000000001390](https://doi.org/10.1097/NNE.0000000000001390).
- Taecharunroj V. (2023), "'What can ChatGPT do?' Analyzing early reactions to the innovative AI chatbot on Twitter", *Big Data and Cognitive Computing*, Vol. 7 No. 1, [doi: 10.3390/bdcc7010035](https://doi.org/10.3390/bdcc7010035).
- Thomas, P.A. (2021), "Reverting hegemonic ideology: Research librarians and information professionals as 'critical editors' of Wikipedia", *College & Research Libraries*, Vol. 82 No. 4, pp. 567-583, [doi:10.5860/crl.82.4.567](https://doi.org/10.5860/crl.82.4.567).
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E. and Lample G. (2023), "LLaMA: Open and Efficient Foundation Language Models", *arXiv*, [doi:10.48550/arXiv.2302.13971](https://doi.org/10.48550/arXiv.2302.13971).
- Wolchover, N. (2011), "How Accurate Is Wikipedia?", *Live Science*, available at <https://www.livescience.com/32950-how-accurate-is-wikipedia.html> (accessed 22 March 2023).
- Zhong, Q., Tan, X., Du, R., Liu, J., Liao, L., Wang, C., Sun, R., Tang, Z., Ren, J., Mebrahtu, C., Zeng, F. (2023), "Is ChatGPT A reliable source for writing review articles in catalysis research? A case study on CO₂ hydrogenation to higher alcohols", Preprints, [doi:10.20944/preprints202302.0292.v1](https://doi.org/10.20944/preprints202302.0292.v1).