Wikipedia and Artificial Intelligence

As generative artificial intelligence continues to permeate all aspects of culture, the people who steward Wikipedia are divided on how best to proceed.

During a recent community call, it became apparent that there is a community split over whether or not to use large language models to generate content. While some people expressed that tools like Open AI's ChatGPT could help with generating and summarizing articles, others remained wary.

The concern is that machine-generated content has to be balanced with a lot of human review and would overwhelm lesser-known wikis with bad content. While AI generators are useful for writing believable, human-like text, they are also prone to including erroneous information, and even citing sources and academic papers which don't exist. This often results in text summaries which seem accurate, but on closer inspection are revealed to be completely fabricated.

Amy Bruckman is a regents professor and senior associate chair of the school of interactive computing at the Georgia Institute of Technology and author of Should You Believe Wikipedia?: Online Communities and the Construction of Knowledge. Like people who socially construct knowledge, she says, large language models are only as good as their ability to discern fact from fiction.

"Our only recourse is to use large language models, but edit it and have someone check the sourcing," Bruckman told Motherboard.

It didn't take long for researchers to figure out that OpenAI's ChatGPT is a terrible fabricator, which is what tends to doom students who rely solely on the chatbot to write their essays. Sometimes it will invent articles and their authors. Other times it will name-splice lesser-known scholars with more prolific ones, but will do so with the utmost confidence. OpenAI has even said that the model "hallucinates" when it makes up facts—a term that has been criticized by some AI experts as a way for AI companies to avoid accountability for their tools spreading misinformation.

"The risk for Wikipedia is people could be lowering the quality by throwing in stuff that they haven't checked," Bruckman added. "I don't think there's anything wrong with using it as a first draft, but every point has to be verified."

The Wikimedia Foundation, the nonprofit organization behind the website, is looking into building tools to make it easier for volunteers to identify bot-generated content. Meanwhile, Wikipedia is working to draft a policy that lays out the limits to how volunteers can use large language models to create content.

The current draft policy notes that anyone unfamiliar with the risks of large language models should avoid using them to create Wikipedia content, because it can open the Wikimedia Foundation up to libel suits and copyright violations—both of which the nonprofit gets protections from but the Wikipedia volunteers do not. These large language models also contain implicit biases, which often result in content skewed against marginalized and underrepresented groups of people.

The community is also divided on whether large language models should be allowed to train on Wikipedia content. While open access is a cornerstone of Wikipedia's design principles, some worry the unrestricted scraping of internet data allows AI companies like OpenAI to exploit the open web to create closed commercial datasets for their models. This is especially a problem if the Wikipedia content itself is AI-generated, creating a feedback loop of potentially biased information, if left unchecked.

One suggestion posted to Wikipedia's mailing list drew attention to the idea of using BLOOM, a large language model released last year under the new Responsible AI License (RAIL) that "combines an Open Access approach to licensing with behavioral restrictions aimed to enforce a vision of responsible AI use." Similar to some versions of the Creative Commons license, the RAIL license enables flexible use of the AI model while also imposing some restrictions—for example, requiring that any derivative models clearly disclose that their outputs are AI-generated, and that anything built off them abide by the same rules.

Mariana Fossatti, a coordinator with Whose Knowledge?—a global campaign focused on enabling access to knowledge on the internet across geographic locations and languages—says large language models and Wikipedia are in a feedback loop that introduces even more biases.

"We have this massive body of knowledge in more than 300 languages," Fossatti told Motherboard. "But of course these 300 different languages are very unequal also.

English Wikipedia is much more rich in content than others and we are feeding AI systems with this body of knowledge."

AI isn't exactly new to Wikipedians—automated systems have long been used on the site to perform tasks like machine translation and removing vandalism. But there are longtime volunteers who are less open to the idea of expanding AI use on the platform.

In a statement from the Wikimedia Foundation, the nonprofit said that AI represents an opportunity to help scale the work of volunteers on Wikipedia and Wikimedia projects.

"Based on feedback from volunteers, we're looking into how these models may be able to help close knowledge gaps and increase knowledge access and participation," a Wikimedia Foundation spokesperson told Motherboard in a statement. "However, human engagement remains the most essential building block of the Wikimedia knowledge ecosystem. AI works best as an augmentation for the work that humans do on our project."

As of this writing, the draft policy includes a point that explicitly states that in-text attribution is necessary for AI generated content. Bruckman doesn't see some of the issues that come with large language models as much different than deliberate and malicious attempts to edit Wikipedia pages.

"I don't think it's that different from vandalism fighting," Bruckman added. "We have strategies for fighting that. I think that unreviewed AI generated content is a form of vandalism, and we can use the same techniques that we use for vandalism fighting on Wikipedia, to fight garbage coming from AI."

In a recent email to the Wikimedia Foundation listserv, Selena Deckelmann, chief product and technology officer at the organization, noted that complex issues exist between volunteers and foundation staff around unfinished technical migrations that affect community decision making among volunteers.

"We must be able to choose maintenance and technical migration areas for prioritization and then be ok with not doing work on others in order to complete some of these big projects," Deckelmann said in the email obtained by Motherboard.

But until then, Bruckman says it's important for editors and volunteers to remain vigilant.

“Content is only as reliable as the number of people who have verified it with strong citation practices,” said Bruckman. “Yes, generative AI does not have strong citation preferences, so we have to check it. I don't think we can tell people ‘don't use it’ because it's just not going to happen. I mean, I would put the genie back in the bottle, if you let me. But given that that's not possible, all we can do is to check it.”