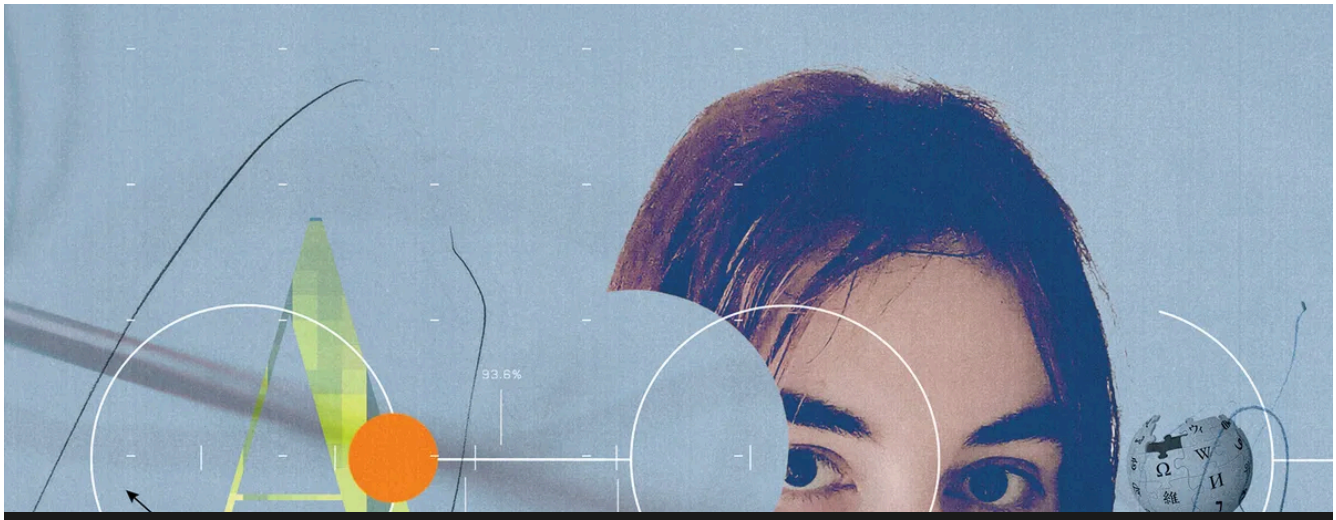
INTERVIEW AI

How Wikipedia’s Editors Keep the Online Encyclopedia Human-Written › AI-generated text is harder to spot, but that hasn’t deterred them

BY ELIZA STRICKLAND 21 OCT 2026

Eliza Strickland is IEEE Spectrum's features editor. She also covers AI and biomedical engineering.



This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. Privacy Policy

ACCEPT

DENY NON-ESSENTIAL

MANAGE PREFERENCES

ELLIA LEBLEU IS A FOUNDING MEMBER OF THE WikiProject AI Cleanup, a volunteer effort launched in 2023 to find and remove AI-generated material from Wikipedia that breaks the website's guidelines. When the cleanup began, AI writing was often betrayed by conspicuous clichés. Today, the project's volunteers are developing tools to detect subtler forms of machine-generated content and debating AI's role in the world's most widely used encyclopedia.

When did you realize Wikipedia had a problem with AI-generated material?

Ellia Lebleu: It was around December 2023. A few other editors and I were chatting on Wikipedia's Discord server, and we had all seen very strange articles with the same odd characteristics. They repeated phrases like "stands as a testament" or "rich cultural heritage." They sometimes had conclusion sections, or sections called "Why is this important?" That was very typical of ChatGPT in 2023 and early 2024 but very odd for Wikipedia articles—and very easy to spot. Then I decided, let's create an official project on Wikipedia so more people can deal with this.

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising.
Privacy Policy

Lebleu: By 2025, the signs of AI writing were becoming subtler and more structural. For example, AI loves the “rule of three,” negative parallelisms, and other structures of speech that aren’t conclusive by themselves but become meaningful as they accumulate. We also look beyond the article. If someone writes five articles of thousands of words each in 5 minutes, there is reason to suspect AI.

Does the AI problem just appear in the main text, or also in citations?

Lebleu: Fake or misleading citations have become a major worry. Sometimes an AI hallucinates a citation entirely, although that’s much rarer today because of retrieval-augmented generation that allows the AI to find sources. More often, it finds a real source but gets its content completely wrong. AI citations that don’t actually support what is being written are extremely commonplace.

English Wikipedia recently prohibited editors from using LLMs to generate articles. How much AI-generated content is still getting through?

Lebleu: I’ve been working with [AI-detection company] Pangram to estimate the proportion of AI-generated content. I tested it on Wikipedia content from before ChatGPT and found no false positives, so its results provide a very conservative lower bound. In late 2025, it flagged perhaps 20 percent of new articles and drafts as AI-generated. In early 2026, that went down to 15 or 16 percent. But we don’t yet know

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising.
Privacy Policy

something that works at scale and filters out obvious cases, it's a good first step. The Wikimedia Foundation is also working on a tool called Paste Check, which can flag text pasted directly from some AI chatbots.

And we've seen Claude sometimes say, "I can't write a Wikipedia article; it would go against Wikipedia's policies." If all major AI companies implemented that kind of prompt refusal, it wouldn't solve the problem entirely, because there are still open-weight models that can be tweaked to avoid restrictions. But it would help with people who use commercial models and aren't especially familiar with the technology.

Can you tell if the rule prohibiting AI writing is working?

Lebleu: It's still a little early to tell. We see possible effects, such as prompt refusal and the apparent slight drop in AI-generated content. But again, we don't yet know if that is because models are getting better or because beginners are learning the rules. The vast majority of AI-generated content is submitted by beginner editors, so we've made many efforts to make the rule very visible in welcome messages, editing help, and advice.

What if AI becomes capable of writing truly reliable Wikipedia articles?

Lebleu: That's one of the big questions we're facing. When the guideline was passed, it was for very practical reasons: We were flooded with AI content, we didn't have

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising.
Privacy Policy

have a basic human-written corpus on which to train AI systems. The fact that

Wikipedia is still that, and will remain that for at least the near future, makes it a precious resource that isn't replicated anywhere else.

But will the people who read and contribute to Wikipedia want it to become partially AI-generated if that makes it more comprehensive? We could preserve Wikipedia's identity as a human-written corpus while also offering an AI-assisted option.

Has Wikipedia reckoned with the idea of "Google Zero"—a possible future in which websites no longer get traffic from Google search?

Lebleu: If everything goes through AI search and websites plummet to zero visibility, how do we expose Wikipedia and show that it can be edited and that people can still contribute? Wikipedia is frequently cited in [Google's] AI Overviews, but users may only see an AI answer citing Wikipedia. They may not realize that Wikipedia is a living project they can edit and contribute to.

The Wikimedia Foundation is working on an attribution framework that could give AI users more information about the article [being cited]: when it was last edited, how comprehensive it is, whether editors have flagged problems, and perhaps even a call to action. In the future, readers might even contribute through AI models: "The AI cited Wikipedia, but there seems to be an error. Can I make an edit request and add this other source I found?"

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. Privacy Policy

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising.
Privacy Policy