Between Prompt and Publish: Community Perceptions and Practices Related to AI-Generated Wikipedia Content

Anwesha Chakraborty
University of Urbino, Italy
anwesha.chakraborty@uniurb.it

Netha Hussain
WikiWomen's Usergroup
nethahussain@gmail.com

Abstract

This project investigates how experienced Wikimedians across language communities perceive the increasing use of generative AI content on Wikipedia. While AI tools may lower the entry barrier for newcomers and enhance editorial productivity for experienced editors, they also raise concerns around misinformation, bias, and the adequacy of existing quality-control mechanisms on Wikipedia.

Wikipedia's language editions have their own communities, with each community being able to make decisions autonomously about their edition of Wikipedia without having to consult anyone externally. This provides the communities with leverage and power in terms of how to deal with generative AI content in their language edition. Therefore, it is important to know the perceptions of members from different language communities about generative AI content, to understand how these may shape the AI related policies of their communities, and to know if there are any cultural differences in terms of perceptions towards AI.

We aim to explore how experienced editors understand and engage with generative AI: what tools they use, how they evaluate AI-assisted contributions, and what opportunities or risks they associate with AI generated content on Wikipedia. This research will provide insights to support the Wikimedia movement in developing community-informed guidelines, workflows, and policies for safeguarding knowledge integrity in the age of AI. Our findings will also contribute to shaping inclusive AI governance policies in the Wikimedia movement.

Introduction

Generative AI tools such as ChatGPT and Bard are reshaping how online content is produced (Dwivedi et al., 2023; Ooi et al., 2023). While these tools are increasingly used to generate Wikipedia content (Brooks, Eggert & Peskoff, 2024), there is concern that the volume and pace of AI-generated content may compromise article quality and Wikipedia's long-standing principles of verifiability and neutrality (Vetter, Jiang & McDowell, 2025).

A recent Wikimedia-funded study explored the broader implications of generative AI tools like ChatGPT on knowledge integrity and editorial governance, examining whether Wikipedia's

current rules are sufficient to handle these new technologies. Although it suggested that existing editorial practices may be sufficient to handle some challenges posed by AI, it also acknowledged deeper, systemic risks that cannot be resolved through community guidelines alone [Research:Implications of ChatGPT for knowledge integrity on Wikipedia - Meta]. The limitation of verifiability of AI-generated content was highlighted by this research, identifying it as a major concern that might undermine Wikipedia's knowledge integrity.

Distinguishing AI-generated content from human-created text has become increasingly difficult. Existing content verification mechanisms, including community oversight and automated tools like ORES, were not designed to handle the scale and complexity introduced by generative AI. Communities are often unequipped both in terms of capacity and technology to verify content at this scale using volunteer labor alone. To date, English Wikipedia has no formal policy governing AI-generated content, but there are certain guidelines on how to handle them [Wikipedia:Artificial intelligence - Wikipedia]. This would mean that the communities have to navigate a rapidly evolving generative AI landscape with unclear expectations.

This project responds to this gap by exploring how experienced editors across language communities are currently dealing with generative AI. We ask:

1. How informed are experienced Wikipedians about generative AI?

2. What are their perceptions of the quality of AI-generated content in their languages?

3. Which AI tools (if any) do they use, and in what contexts?

4. What workflows do they adopt when integrating AI-generated content?

5. How do they address the issue of information reliability when they integrate AI into their editorial workflow?

6. What opportunities and challenges do they see in generating AI-driven content on Wikipedia?

We focus on the lived experiences and practices of Wikimedians themselves to address these research questions. Furthermore, we examine community perceptions across different languages and cultural contexts, aiming at providing practical, grounded insights that can inform future community guidelines and AI governance strategies.

By centering editor perspectives, this project aims to generate insights from practitioners that can guide community decisions and policy-making related to handling generative AI content on Wikipedia.

We have envisioned the following start and end dates for the project:

Start: October 1, 2025, **End**: August 31, 2026

Methods

We will employ a qualitative approach to data gathering and analysis, employing the following workflow:

1. Participant Selection

We will engage active editors from five language Wikipedias—English, Italian, Malayalam, Swedish and Bangla (as we have a high level of proficiency in these languages), focusing on administrators and experienced content creators. Participants will be recruited through

direct outreach (e.g., via Telegram groups, mailing lists, Village Pumps), and a short screening form will assess eligibility.

Sampling will be based on purposive and snowball strategy, continuing until thematic saturation is achieved. Diversity in experience level, regional background, and familiarity with AI will be maintained.

2. Semi-Structured Interviews

We will conduct in-depth interviews exploring the following themes: familiarity with generative AI, perceptions of AI-generated content quality, use of AI tools (if any) in content creation, verification techniques used for checking sources of AI-generated content and attitudes toward future AI integration on Wikipedia.

Informed consent will be obtained from all the interviewees. All interviews will be anonymized and the primary data gathered would be made accessible only to the authors of the study.

3. Thematic Analysis

We will use Clarke and Braun's (2017) approach to thematic analysis to code and interpret interview transcripts. This flexible method allows emergent patterns and editor attitudes to surface without imposing predefined categories.

4. Further Content Analysis

We will identify AI-related discussions from: a) different language Wikipedia talk pages (such as English and Italian), b) Diff blog, c) Wikimedia-I and d) Wikimedia Signpost. After that, we will perform content analysis on that dataset to let the key frames and themes emerge, while also describing the exact volume of data analysed (White and Marsh 2006). Our aim is to include voices from non-English, non-European languages as well, hence we are actively on the lookout for such discussions among Bangla and Malayalam Wikipedians (the two Indian languages that we have native fluency in), although our preliminary assessment indicates that such content is sparse.

Expected output

At the end of the study, we will know how Wikipedians perceive the use of AI generated content on Wikipedia, how they evaluate quality of articles that are likely created using generative AI and their thoughts about the possible ways by which generative AI can be used to make the workflow easier for Wikipedia editors, while at the same time knowing from them what challenges they face due to generative AI content in their daily work, focusing among other issues on misinformation.

This project will generate multiple outputs aimed at both scholarly and Wikimedia community audiences. The insights gathered will contribute to academic research on AI and online collaboration, while also supporting Wikimedia affiliates and volunteers in shaping informed AI governance. The following outputs are envisioned:

A peer-reviewed academic journal article detailing the full findings of the study and contributing to scholarship on AI in digital knowledge communities.

A comprehensive research report on Meta-Wiki, offering an accessible summary of results and key takeaways for Wikimedia contributors, with opportunities for community feedback.

A targeted summary for Wikimedia affiliates and hubs, offering insights to inform local policy discussions and strategy development.

A blog post for Diff, sharing the core findings in a concise, engaging format for the broader Wikimedia community.

A conference presentation at Wikimania 2026 (pending acceptance), allowing for live engagement with Wikimedians interested in AI and content quality.

We will also be available to present the findings at Wikimedia Research events and local events of the communities in which the authors are a part of.

Risks

Participant familiarity with AI may vary: If initial participants lack sufficient knowledge, we will expand the sample size to ensure meaningful input.

Self-disclosure of AI use: Editors may underreport AI use due to community skepticism. We will ensure a non-judgmental, confidential environment.

Community impact plan

Our work aims to equip Wikimedia communities with evidence-based insights about generative AI's current and potential uses. We will:

1. Post findings to Meta-Wiki inviting feedback on talk pages
2. Share summaries tailored for language-based affiliates
3. Present findings at Wikimania 2026 (pending abstract acceptance), Wikimedia Research events and local events.
4. Translate select outputs into local languages as needed
5. Collaborate with Wikimedia affiliates and organized communities for outreach and engagement

Evaluation

Success will be evaluated by:

1. Completion and publication of a peer-reviewed article
2. Community engagement via Meta-Wiki (e.g., comments and discussions)
3. Use of findings by Wikimedia's organized communities for decision making surrounding AI generated content
4. Citation of our research in future AI or Wikipedia policy-related work

Budget

Our project will be run entirely by the authors in their capacity as independent researchers, although both of them are affiliated to Universities in Europe. Therefore, we do not ask for institutional overhead. We ask for funding only to compensate for our time, a total of 8,571 USD.

One (or both) of the researchers will be traveling to Wikimania 2026 at their own expense to disseminate the research findings there.

The participants of the research will be offered a 20 USD gift card each as a gift voucher of their choice. The expenses associated in purchasing the gift card will be borne by the researchers.

Link to the full budget here.

References

Brooks, C., Eggert, S., & Peskoff, D. (2024). The Rise of AI-Generated Content in Wikipedia. arXiv preprint arXiv:2410.08044.

Burtch, G., Lee, D., & Chen, Z. (2024). The consequences of generative AI for online

knowledge communities. Scientific Reports, 14(1), 10413.

Clarke, V., & Braun, V. (2016). Thematic analysis. *The Journal of Positive Psychology, 12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper:"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, 71, 102642.

Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., ... Wong, L. W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. Journal of Computer Information Systems, 65(1), 76–107. https://doi.org/10.1080/08874417.2023.2261010

Vetter, M.A., Jiang, J. & McDowell, Z.J. An endangered species: how LLMs threaten Wikipedia's sustainability. AI & Soc (2025). https://doi.org/10.1007/s00146-025-02199-9

White, M. D. & Marsh, E. E.(2006). Content analysis: A flexible methodology. *Library trends, 55*(1), 22-45.