Generative AI and sustainability of information production of Wikimedia projects

Paul Keller
Open Future Foundation /
University of Amsterdam

dr Alek Tarkowski
Open Future Foundation

Abstract

This research critically examines whether Wikimedia's existing sharing model remains effective and fit for purpose in the face of these AI-driven pressures. It aims to investigate mechanisms that can further support sustainable production of information on Wikimedia projects. The rise of generative AI presents significant sustainability challenges for free knowledge commons like Wikimedia. Large-scale data scraping by AI models extracts considerable value from these projects, often without fair reciprocity, straining infrastructure and questioning the viability of established knowledge-sharing frameworks. The study aims to identify actionable pathways for ensuring the continued vitality of the free knowledge ecosystem in the AI era.

Introduction

The rise of AI, particularly generative AI models, poses profound challenges and opportunities for Wikimedia projects and the knowledge commons in general. These new digital technologies challenge current ways of producing, preserving and accessing broadly understood information. Wikimedia and other organizations that steward collective sources of information are at risk of being disintermediated by AI-driven solutions. Yet, we may still very much need them as sources of ground truth and as continuous sources of new information. This points toward a crisis of sustainable information production, as free knowledge paradoxically empowers the concentrations of power that it was meant to challenge [1]. In face of these shifts in the information ecosystem, there is a need to seek new strategies and tools for ensuring sustainability of information production.

In recent years, WMF and other movement entities started addressing this challenge, which has been identified as one of key external trends since 2023 [2,3,4]. In April 2024, WMF published data on how crawlers impact the operations of Wikimedia projects, pointing to the most practical aspect of the issue: the burden on Wikimedia infrastructure due to increased automatic traffic [5]. Scraping of Wikimedia at massive scale is a prime example of how value is extracted from the commons, with few reciprocal mechanisms. A Wikimedia convening on AI, organized in 2024, recognized research on "empowering knowledge sharing communities in the AI era" as one of the priorities [6].

This is a sign of fundamental shifts in the free knowledge ecosystem that need to be analysed, as they put in question the knowledge sharing frameworks that underpin Wikimedia projects. This also requires a critical review of this sharing framework, based on copyleft licensing, and including elements like the Wikimedia Enterprise project, offering a voluntary payment mechanism. Understanding the complex interplay of AI, copyright, and economic incentives is essential for developing strategies that sustain the production of information on Wikimedia projects.

This research project seeks to answer the following questions:

1. Is Wikimedia's sharing framework fit for purpose, in face of challenges caused by generative AI development? In particular, are copyleft mechanisms effective?
2. What new mechanisms, including copyright-based solutions, payment schemes and other approaches, increase sustainability of Wikimedia projects?
3. Are there relevant examples of such mechanisms from other free knowledge initiatives?

Date: 1 July 2025 - 30 March 2026.

Related work

At Open Future we have been investigating challenges related to shifts in the knowledge ecosystem, and framed them as cases of the Paradox of Open [1]. With the "Shifting Tides" study we have investigated attitudes towards these changes among leaders in the Open / Digital Commons movement [7]. In 2024, we used the Pol.is platform to organize an Alignment Assembly on AI and the Commons, a

form of action research providing a deeper look into the attitudes of people engaged in the knowledge commons [8]. A study by Kacper Szkalej and Martin Senftleben looked at the impact of copyleft licensing on generative AI development [9].

Researchers have been investigating alternative licensing and data governance frameworks [10], and broader changes in the media ecosystem [11]. While the Wikimedia Foundation has been conducting work to address these challenges, we are not aware of research that investigates these issues, as they relate to Wikimedia.

Paul Keller is co-director at Open Future and fellow at the Institute for Information Law (IViR), University of Amsterdam. He is an expert in open licensing and the design of knowledge sharing frameworks.

Alek Tarkowski is co-director at Open Future and holds a PhD in sociology. In recent years, he has been studying shifting attitudes towards sharing, in the free knowledge movement.

Methods

Our research project will be based on following research methods:

1. **Expert interviews.** These will be semi-structured interviews with approximately 10 experts representing diverse perspectives, including members of the Wikimedia movement, copyright and policy specialists, and AI researchers and developers working with knowledge commons. The primary aim of the interviews will be to map potential solutions that ensure sustainable information production.
2. **Qualitative analysis of potential solutions.** This analysis will support the interviews and offer a mapping of

existing solutions, developed in other free knowledge communities.

3. **Focus group.** We will organize an online focus group to confirm the findings of our analysis in a participatory manner.

Expected output

The project will produce the following outputs:

1. Scientific publication detailing the findings. Audience: scientific community interested in knowledge commons, free licensing and information production. We will aim to present the publication at at least one scientific conference.
2. Policy brief with recommendations for governance of Wikimedia projects. Audience: Leaders in various Wikimedia Movement entities, responsible for shaping policies and governance mechanisms.
3. Presentation for Wikimedia community to disseminate the findings (during an online event).

Risks

Potential risks include:

- Difficulty in ensuring broad and diverse participation in the Pol.is survey.
- The rapidly evolving nature of generative AI technology and policy, which may require adjustments to the research plan.

Mitigation strategies include:

- Developing a comprehensive recruitment strategy and leveraging existing networks to engage relevant participants in the Pol.is survey. We will

draw on our previous experience with this methodology
- Adopting a flexible research design that allows for adaptation to new developments in the field.

Community impact plan

We will engage with experts from the community and the affiliates, through the interviews and the focus group. The outcomes of this research will provide practical advice that hopefully will influence strategies of WMF and the affiliates.

Evaluation

The project can be evaluated based on following criteria:

1. **Originality**. The project proposes a novel, participatory approach to researching a core strategic challenge for the Wikimedia movement.
2. **Participation:** The participatory research component, using the Pol.is platform, can be evaluated using participation metrics, such as number of people involved, their diversity, etc.

We will consider the project a success if it provides insights and specific solutions that will help Wikimedia projects, WMF and affiliates solve the issue of sustainable information production.

Budget

Budget spreadsheet can be accessed here: https://docs.google.com/spreadsheets/d/1ICzofdDydq2IhiDBS2Dq6Hv0XT1yIJcUgJwFWLr1LwE/edit?usp=sharing

References

[1] Keller, Paul, and Alek Tarkowski. 'The Paradox of Open'. Accessed 16 April 2025. https://paradox.openfuture.eu/.

[2] Wikimedia Foundation. 'Wikimedia Foundation Annual Plan/2023-2024/External Trends - Meta'. Accessed 16 April 2025. https://meta.wikimedia.org/wiki/Wikimedia_Foundation_Annual_Plan/2023-2024/External_Trends.

[3] Wikimedia Foundation. 'Wikimedia Foundation Annual Plan/2024-2025/External Trends - Meta'. Accessed 16 April 2025. https://meta.wikimedia.org/wiki/Wikimedia_Foundation_Annual_Plan/2024-2025/External_Trends.

[4] Wikimedia Foundation. 'Wikimedia Foundation Annual Plan/2025-2026/Global Trends - Meta'. Accessed 16 April 2025. https://meta.wikimedia.org/wiki/Wikimedia_Foundation_Annual_Plan/2025-2026/Global_Trends.

[5] Mueller, Birgit, Chris Danis, and Giuseppe Lavagetto. 'How Crawlers Impact the Operations of the Wikimedia Projects'. Diff (blog), 1 April 2025. https://diff.wikimedia.org/2025/04/01/how-crawlers-impact-the-operations-of-the-wikimedia-projects/.

[6] Wikimedia Foundation. 'Artificial Intelligence/Bellagio 2024 - Meta'. Accessed 16 April 2025. https://meta.wikimedia.org/wiki/Artificial_intelligence/Bellagio_2024.

[7] Tarkowski, Alek, and Zuzanna Warso (2023). 'Shifting Tides'. Open Future (7 June 2023). https://openfuture.eu/publication/shifting-tides.

[8] Tarkowski, Alek, and Shannon Y. Hong. 'Alignment Assembly on AI and the Commons — Outcomes and Learnings'. Open Future (12 June 2024). https://openfuture.eu/publication/alignment-assembly-on-ai-and-the-commons-outcomes-and-learnings.

[9] Szkalej, Kacper, and Martin Senftleben. 'Generative AI and Creative Commons Licences - The Application of Share Alike Obligations to Trained Models, Curated Datasets and AI Output'. JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law 15, no. 3 (13 December 2024). https://www.jipitec.eu/jipitec/article/view/415.

[10] Benhamou, Yaniv, and Melanie Dulong de Rosnay. 'Open Data Commons Licenses (ODCL): Licensing Personal and Non Personal Data Supporting the Commons and Privacy'. SSRN Scholarly Paper. Rochester, NY, 12 December 2023. https://doi.org/10.2139/ssrn.4662511.

[11] Cools, Hannes, Natali Helberger, and Claes De Vreese. 'Co-Creating Research at The AI, Media, and Democracy Lab: Reflections on the Role of Academia in Collaborations with Media Partners'. Journalism, 4 February 2025, 14648849251318622. https://doi.org/10.1177/14648849251318622.