The Value of Human Intelligence in the AI Era: An Examination Through Wikipedia

Kai Zhu
Bocconi University

Abstract

This project investigates the value of human curation in knowledge production in the age of AI, using Wikipedia as a case study. As large language models like ChatGPT can now generate encyclopedia-style content instantly, we ask: what unique role does Wikipedia's human-curated knowledge still play? Through two studies, we analyze how AI-generated articles are shaped by the presence or absence of pre-existing Wikipedia content and whether human editorial work adds measurable value beyond AI's capabilities. The findings will inform strategies for maintaining the integrity, relevance, and impact of Wikipedia in an AI-driven information ecosystem.

Introduction

The rapid advancement of Large Language Models (LLMs) such as GPT-4, Claude, and others has revolutionized our capacity to generate encyclopedia-like articles on virtually any subject within seconds. This unprecedented capability raises a fundamental question: In an era where artificial intelligence can produce detailed expositions almost instantaneously, what role remains for human curation of knowledge?

Wikipedia stands as one of humanity's most ambitious collaborative intellectual projects—a continuously evolving, multilingual repository of information curated by thousands of volunteer editors worldwide. Its persistence and growth in the digital age suggest that human intelligence still plays a vital role in knowledge creation and organization. However, the exact nature of this role in relation to emerging AI capabilities requires systematic investigation.

This research proposal addresses two central questions. First, to what extent do AI-generated texts depend on pre-existing human-curated content from Wikipedia? This question explores the relationship between human knowledge organization and machine learning outcomes. Second, what unique value, if any, does human intelligence contribute to the knowledge creation process that AI cannot readily replicate? This deeper question examines whether human curation provides distinctive benefits beyond efficient information processing.

By examining these questions through empirical studies, we aim to clarify the complementary relationship between human and artificial intelligence in knowledge ecosystems.

Date: July 1, 2025 - June 30, 2026

Related work

Recent scholarship has focused extensively on the impact of Large Language Models on Wikipedia, with relatively less attention paid to how Wikipedia's human curation shapes LLM outputs. Lyu et al. (2025) found that Wikipedia articles whose content overlaps with ChatGPT

experienced greater declines in editing and viewership after ChatGPT's launch, suggesting selective user disengagement when AI provides comparable content. Huang et al. (2025) identified that certain English Wikipedia articles have been influenced by LLMs, with approximately 1-2% impact in specific categories, as evidenced by increasing frequency of LLM-favored words like "crucial" and "additionally."

Other research has explored potential benefits of AI on content production. Zhu and Walker (2025) investigated how machine translation technology impacts multilingual content creation on Wikipedia. Using Wikipedia's integration of Google Translate in 2019 as a natural experiment, they found that this technological improvement substantially increased translation volume by 149% while maintaining quality standards and reader engagement.

Our research addresses a gap in the literature by examining how human-curated Wikipedia content shapes AI language model outputs across languages. While Wikipedia's role as a primary training source for most LLMs is well established, the impact of human curation processes on model performance remains less understood. By empirically analyzing how AI-generated content depends on human-curated Wikipedia articles, our study provides key insights into the symbiotic relationship between human editorial work and artificial intelligence in knowledge production.

Methods

Our research is structured into two interconnected studies.

In the first study, we will select a range of topics that have corresponding Wikipedia articles in multiple languages. A key design element involves choosing languages where articles were established before the AI model's knowledge cutoff and those where articles emerged afterward. For each topic, we will prompt an LLM, such as ChatGPT, to generate an encyclopedic entry in the target languages using a standardized prompt. The resulting AI-generated texts will be compared to the respective human-curated Wikipedia articles. This comparison will utilize embedding-based techniques to calculate semantic similarity and measure textual overlap with metrics such as cosine similarity. Complementary qualitative evaluations will assess structure, depth, and accuracy. Our hypothesis is that languages with pre-existing curated articles will yield AI outputs that more closely mirror human-generated content, thereby suggesting that AI models rely significantly on curated data from their training sets.

The second study will evaluate the added value of human intelligence in knowledge generation by testing the AI's performance under three different conditions. In the first condition, the AI will have full access to a comprehensive, human-curated Wikipedia article. In the second condition, it will be provided only with the reference list from that article, simulating a scenario where only raw sources are available. In the third condition, the AI will first generate a draft article based solely on the references and then refine its output after receiving both the draft and the complete curated text. We will assess outcomes through tasks that measure factual retrieval accuracy, the precision of responses in a Q&A format, and the AI's ability to synthesize insights. Performance will be evaluated using automated similarity metrics as well as expert assessments. Should the AI underperform in the conditions lacking full access to curated text, it would underscore the indispensable role of human curation.

Together, these two studies offer complementary perspectives that address our research questions. While the first study examines the relationship between existing Wikipedia content and AI outputs across languages (investigating dependency), the second study directly measures the value added by human curation beyond raw information access (investigating unique contribution). This dual approach allows us to triangulate findings: the cross-linguistic comparison reveals how human knowledge organization shapes AI performance, while the controlled experiment isolates specific mechanisms through which human curation improves knowledge quality. The findings will provide robust insights into both the extent to which AI relies on human-curated knowledge and the unique value that human curation adds to the knowledge ecosystem—foundational questions for understanding the future relationship between artificial and human intelligence in knowledge production.

Expected output

We anticipate multiple research outputs from this project. Foremost, two scientific publications will be produced to disseminate the findings. We also intend to release datasets and source code, thereby facilitating reproducibility and further research in the area.

The primary beneficiaries of our work include the Wikimedia community, academic researchers interested in the dynamics of human-AI collaboration, and policymakers concerned with digital knowledge governance. The insights gained may also inform the development of improved AI systems that better integrate human curatorial processes.

Risks

One potential challenge is that existing embedding techniques may not fully capture the nuances of editorial quality and accuracy. Additionally, variations in Wikipedia article quality and structural differences across languages could introduce unforeseen variability. To mitigate these risks, we will employ a robust combination of automated metrics and human expert evaluations, and carefully select topics that are well-documented and broadly recognized.

Community impact plan

The outcomes of this research have the potential to offer valuable insights for a range of audiences beyond the academic community. We plan to engage with Wikimedia editors, volunteer developers, and affiliates by presenting our findings in accessible formats, such as in public talks and online workshops. By directly addressing the role of human curation in the era of AI, our work will contribute to policy discussions regarding editorial practices and knowledge governance within the Wikimedia ecosystem. We are committed to collaborating with these communities to ensure that our research not only advances scholarly understanding but also has practical implications for how Wikipedia and similar projects evolve in the future.

Evaluation

Our project will be evaluated through a combination of quantitative and qualitative measures. Success in Study 1 will be determined by the degree of semantic similarity between AI-generated content and human-curated Wikipedia articles, as measured by established embedding techniques and corroborated by expert reviews. In Study 2, improvement in task

performance across conditions will be a key indicator.

We will also assess the robustness of our findings through cross-language comparisons and by testing across diverse subject matters. External feedback from the Wikimedia community and peer-reviewed publications will serve as additional metrics of impact and success.

Budget

Our budget has been developed using the Wikimedia Foundation's budget template, and it reflects the detailed costs associated with data collection, computational resources, and dissemination activities. A full breakdown of the budget is available via the following link: https://docs.google.com/spreadsheets/d/10hl6IgboaKHXFAyE1eYjAS5XNuOzaUjGMMYf6jRDXbI/edit

References

Lyu, L., Siderius, J., Li, H., Acemoglu, D., Huttenlocher, D., & Ozdaglar, A. (2025). Wikipedia Contributions in the Wake of ChatGPT. arXiv preprint arXiv:2503.00757.

Huang, S., Xu, Y., Geng, M., Wan, Y., & Chen, D. (2025). Wikipedia in the Era of LLMs: Evolution and Risks. arXiv preprint arXiv:2503.02879.

Zhu, K., Walker, D. (2025). Machine-assisted Content Creation on Peer Production Platforms. Working Paper.