Generative AI-powered Chatbot for New Wikipedians

José Emilio García
UNAM

Pepe Flores
Wikimedia Mexico/R3D

Saiph Savage
Northeastern University/UNAM

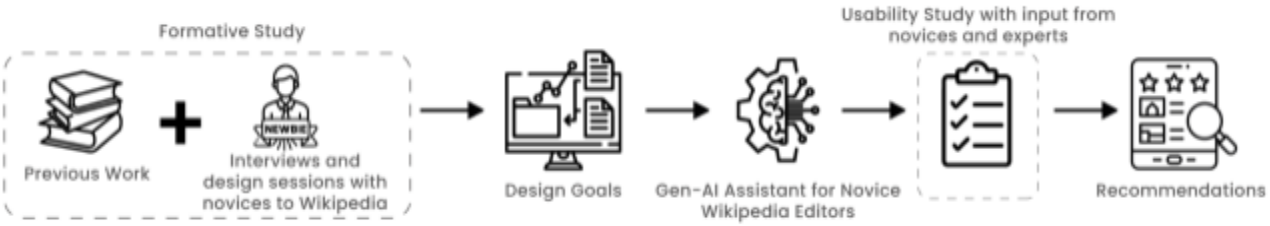

Figure 1: Overview of our study.

Abstract

This research will study how generative AI can assist new Wikipedia editors, especially from Latin America, in contributing to articles about Latin American women. Initially, through interviews and participatory design sessions, we will capture novice editors' needs and expectations for AI support. Leveraging these insights, we will create and deploy a generative AI-powered chatbot tailored for new Latin American Wikipedians. The chatbots will guide novices in editing Wikipedia content, with all content produced by humans. To evaluate the chatbot we will conduct a usability study and real-world deployment. The usability study will collect feedback and evaluations from both novice and expert Wikipedians to assess the effectiveness and limitations of this chatbot, focusing on its ability to guide newcomers in different editing tasks. Based on the results from the usability study, we will propose design recommendations to enhance AI tools for supporting novice editors and closing content gaps about the Global South, particularly Latin America. We will implement some of these recommendations into the chatbot and then conduct a real world deployment where we will measure how the chatbot can help novices to cover gender information gaps on Wikipedia. Our research aims to make Wikipedia more inclusive, while underlining the potential and challenges of using generative AI to engage new editors and enrich global cultural narratives.

Introduction

The sustainability of open collaborative systems, such as Wikipedia, is significantly influenced by their ability to recruit and retain contributors [13]. This is a pressing concern for Wikipedia, which has been facing a documented decline in its editor base [13]. While initiatives like ART+Feminism and offline edit-a-thons have shown success in attracting new collaborators [6], ensuring their continued participation has been challenging [6]. The complexities involved in editing Wikipedia, from understanding its editorial guidelines to technical aspects of article formatting, can be daunting for newcomers [14], leading to a low retention rate.

Figure 2. The chatbot answers a question regarding adding missing information to an article.

In this context, generative AI technologies, such as AI-powered chatbots, present a promising avenue for enhancing the participation of new Wikipedians [9]. By offering personalized guidance on editing practices, as well as spell and grammar checks, these tools have the potential to lower the entry barriers for novice editors. However, there is a notable gap in our understanding of how new Wikipedia editors, particularly those from underrepresented regions, envision the use of generative AI to aid their contributions. Furthermore, the actual efficacy of these AI-powered tools in supporting new editors remains to be fully explored.

To address these challenges, this research will embark on a formative study beginning with interviews and participatory design sessions with new Wikipedians from Latin America. These interactions aim to capture their perceptions and expectations of generative AI support tools. Drawing from these insights, we will then design and deploy a generative AI chatbot tailored to the needs of these novice editors. A significant focus of our study will be on guiding novices in editing articles about Latin American women, an area where Wikipedia suffers from substantial knowledge gaps. By targeting this subject matter, we strive to not only enhance the tool's effectiveness but also contribute to a more inclusive and diverse representation of global histories and cultures on Wikipedia.

To evaluate the chatbot's effectiveness, we will conduct a usability study and a real-world deployment, where novice editors use the AI-enhanced chatbot for Wikipedia editing guidance. We will collect feedback from both novices and experienced Wikipedians on the chatbot's advice and the novices' content creation. This feedback is crucial for understanding the limitations and strengths of generative AI in supporting the writing and participation of new editors.

From our usability study, we will then propose design recommendations for generative AI tools

for novice Wikipedians, incorporating some into our chatbot for a real-world deployment. In this phase we will examine the chatbot's effectiveness in helping novices address Wikipedia's knowledge gender gaps, focusing on the Global South and Latin America. Fig 1 illustrates our research, including our formative and usability studies. Fig 2 shows a visual representation of our envisioned AI-enhanced chatbot.

Overall, we will investigate how generative AI can aid beginner Wikipedians, pinpointing future opportunities for enhanced collaboration between human mentors and AI, or areas where human guidance excels. Our goal is to discover how to best combine human insight with AI's capabilities to improve the support for novice Wikipedians. Through our study, we will create advanced, inclusive technologies that elevate the contributions of novices, thereby enriching Wikipedia's content diversity for the benefit of future generations.

Date: June 1, 2024 - April 30, 2025.

Related work

We recently created a platform for Latin-American crowdworkers that makes use of AI-enhanced tools, including an Intelligent Assistant in the form of a chatbot [5]. We found these tools to be exceptionally useful to our target group's needs, particularly, to help them build a supportive community, and to offer them assistance to complete their tasks. We believe a chatbot can also be integrated to Wikipedia to help newcomers successfully navigate the editing process.

Regarding Spanish Wikipedia, where we plan to implement the chatbot, one tool is currently in use: SeroBOT, focused on fighting vandalism [16]. It is possible to build upon its experience, and complement its work with our chatbot. As a result, Spanish Wikipedia would have tools to both support maintenance tasks, and facilitate participation of newcomers.

Our chatbot could complement another closely related tool that is already in use: SuggestBot [4]. It recommends articles to users based on their profile, previously edited articles, specific categories or WikiProjects. In this way, new editors can easily find articles to work on. After choosing an article from SuggestBot, our chatbot could help newcomers easily perform their first edits. As a result, the tasks of helping newcomers what to edit, and then how to do it would be assisted by bots. Our chatbot will also be able to build upon SuggestBot's experience and recommend articles. The main difference being the chatbot's suggestions will be based on predefined lists that help address the gender gap in regards to content, e.g., the WikiProject Women writers/Missing articles page.

Finally, we have previously worked on bots to recruit volunteers to Wikipedia [7]. This experience has provided us with insights on both newcomers' motivation and participation, and their engagement with virtual assistants. We can leverage this experience, along with our knowledge of previous bot implementations on Wikipedia [18], to successfully integrate our virtual assistant and aid newcomers in navigating the editing process.

Methods

Our research will create a generative AI chatbot to aid new Wikipedia editors, starting with a formative study for design insights, then a usability study with both new and experienced editors to evaluate the effectiveness of the chatbot. This usability study will help to highlight the integration of generative AI into Wikipedia's editing process, offering design recommendations for future generative AI tools. Insights from the usability study will be applied

to improve the chatbot, which will then be deployed in a real-world setting to examine its impact on reducing gender information gaps on Wikipedia.

Formative Study

The first phase of our research will involve conducting a formative study to understand the specific needs and challenges faced by novices when editing Wikipedia articles. This study will consist of semi-structured interviews with individuals from Latin America who are interested in contributing to Wikipedia but may feel intimidated by the editing process or unsure where to start. Our goal is to identify tasks where generative AI offers the most benefit, such as editing, citing, or discussion guidance. We'll apply culture theory to shape our findings and design advice [xx], ensuring our AI tools are adapted to the Latin American context. We'll leverage our ties with Mexico's Wikipedia community, especially those formed during women's edit-a-thons [6], to find and recruit study participants.

Implementation of Generative AI Chatbots

Based on the insights gained from the formative study, we will create a generative AI chatbot designed to guide novices through the editing process. This AI assistant will be tailored to the specific needs highlighted from our formative study. It will likely provide step-by-step guidance, answer questions in real-time, and offer feedback on proposed edits. The assistants will be equipped with generative AI technologies to understand and respond to user queries effectively. In the subsequent sections, we provide a more detailed exploration of our generative AI chatbot. Figure 2 illustrates a practical example of our chatbot's expected functionality.

Designing a Generative AI-Enhanced Chatbot for Novice Wikipedians

The chatbot, designed as a web plugin, will simplify the editing process for newcomers by providing support and guidance directly within Wikipedia's interface, as illustrated in Fig 2. We expect that our chatbot will have backend prompt engineering by connecting with OpenAI's API on our backend server. Through this we will be able to leverage the capabilities of Large Language Models (LLMs), and based on prior research [5, 7], we expect that our chatbot will be designed to:

- Provide step-by-step guidance to newcomers throughout the editing process, making it more accessible and less daunting.
- Answer inquiries regarding Wikipedia's editing guidelines, ensuring that edits meet the platform's quality standards.
- Explain the procedural steps to follow when edits are contested, fostering a constructive and collegial atmosphere for discussion in the Wikipedia Talk pages.
- Recommend articles for editing, with a special emphasis on addressing the gender gap by suggesting topics related to women from Latin America. To do so, we will incorporate lists of articles that need editing from the WikiProject Women writers/Missing articles page, as well as lists from the WikiProject Women in Red initiative, prioritizing women from Latin America. This integration aims to systematically identify and fill information voids about these women on Wikipedia. We plan to use the Wikipedia API to supply the LLM with articles about Latin American women, enabling it to recommend those that are lacking in information.

The chatbot will provide real-time feedback to maintain Wikipedia's quality standards, guiding users on guidelines and editing rules to minimize extensive searches. This aims to lighten the workload for less experienced editors by offering instant clarifications.

In situations with contested edits, the chatbot will guide newcomers on the etiquette of Wikipedia discussions and encourage constructive dialogue, without engaging in discussions itself. Its advice on handling these scenarios aims to be invaluable to newcomers, potentially reducing stress.

Lastly, by suggesting articles to edit, particularly those that highlight women from Latin America, our chatbot aligns with Wikipedia's ongoing efforts to close the gender content gap. Drawing on lists like those from the WikiProject Women writers/Missing articles page, it will contribute to the WikiProject Women in Red initiative, which aims to improve Wikipedia's coverage of women.

Overall, we expect that our generative AI-enhanced chatbot will represent a significant step forward in making Wikipedia editing more accessible, engaging, and equitable for newcomers. Through real-time assistance and guidance, we aim to foster a more inclusive and diverse community of editors, thereby enriching the wealth of knowledge available on Wikipedia.

Usability Studies and Real World Deployment

To evaluate our generative AI-enhanced chatbots we will conduct a usability study and a real world deployment.

Usability Study

Our usability study will examine the viewpoints of both novice and experienced Wikipedians with our chatbot.

For the case of novices, we will be specifically focusing on:

- *Usability for Novices:* We will assess how easily novice users can use our chatbot to assist them with various Wikipedia editing tasks, like creating articles, expanding content, enhancing readability, updating facts, correcting errors, citing sources, organizing content, adding multimedia, linking (wikifying), removing vandalism, translating, categorizing, and participating in discussions. They'll work in personal sandboxes, enabling us to evaluate their success and the chatbot's support level in these activities.
- *Guidance and Original Content Creation:* Our study will evaluate how well the chatbot helps novices create original content on Wikipedia. We will analyze if the chatbot encourages them to add new information across various editing tasks.
- *Improvement in User Experience:* Lastly, we will evaluate how the chatbot enhances novices' overall user experience on Wikipedia, particularly focusing on novices' perspectives on editing content more effectively and enjoyably.

In addition to focusing on novice users, our study will also capture the perspectives of expert Wikipedians. Specifically, we will study:

- *Accuracy of Guidance:* Experts will assess the precision of the chatbot's advice, verifying its accuracy for novices. For each editing task assigned to novices and the corresponding guidance offered by the chatbot, expert Wikipedians will rate the accuracy of the chatbot's advice using a five-point Likert scale.

- *Adherence to Wikipedia's Guidelines:* For each editing task, expert Wikipedians will evaluate if novice-created content adheres to Wikipedia's guidelines, rating compliance on a five-point Likert scale.

Feedback from both novices and experts will guide the design improvements for our AI chatbot, emphasizing a "human in the loop" system where experts step in to help novices when AI advice falls short. Figure 1 shows an overview of our study including user studies.

Real World Deployment Study

A key aspect of this research is to study how our chatbot can help bridge gender knowledge gaps on Wikipedia. We plan to conduct a real-world deployment of our chatbot. The deployment will pay particular attention to the chatbot's impact on helping novices create new content and articles about women in Latin America. By comparing the volume of edits and new articles created during workshop and edit-a-thon deployments with historical Wikipedia trends and statistics, we will quantify the chatbot's effectiveness in addressing these knowledge gaps.

Our real world deployment will also assess the chatbot's impact on editorial discussions, focusing on its effectiveness in guiding novices through contested edits and studying ways it could improve discussions without overshadowing human editors. We aim to validate the chatbot's usefulness while carefully evaluating the risks of integrating AI into Wikipedia's editorial workflow. This represents a key advancement in understanding AI's role in Wikipedia's operations.

In conducting this usability study and its following real-world deployment, our goal is to evolve our generative AI-powered chatbot into a more effective resource. This research aims to enhance the editing experience for newcomers to Wikipedia, foster a more inclusive and thorough knowledge base on the platform, and provide insights into how AI can streamline the encyclopedia's editorial process.

Expected output

Main outputs that will directly benefit Wikipedia:

- Enhanced introduction of newcomers to the editing process.
- Increased editor participation and retention.
- Better understanding of Wikipedia's processes and guidelines by new editors.
- Reduce the gender gap in Wikipedia articles.

We also plan to disseminate our findings through:

- Publications in HCI scientific conferences: ACM Conference on Human Factors in Computing Systems (CHI), ACM Conference on Computer-Supported Cooperative Work, Avances en Interacción Humano Computadora (Mexican CHI)
- White papers and blog posts shared with the Wikipedia community.
- Edita-thons.
- Workshops with Wikipedians.

Risks

Content Bias

We acknowledge that LLMs, despite their potential for innovation, can perpetuate and amplify existing biases [2]. A significant concern arises when considering how these biases may manifest within the editorial guidance offered to Wikipedia contributors. To address this issue, we will conduct a usability study. It will engage both experts and novices in evaluating the

guidance provided by an LLM-enhanced chatbot, in order to identify and flag these biases to better understand them.

However, understanding that bias detection and mitigation are complex processes, we are committed to maintaining a 'human in the loop' approach. This strategy could help to ensure that human editors retain ultimate control over the content that is published on Wikipedia [17, 11]. By integrating human judgment with the capabilities of LLMs, we believe that editors will be better equipped to recognize and address biases. This collaborative process not only enhances the accuracy and fairness of the information on Wikipedia, but also empowers our community of editors by placing them at the forefront of decision-making. As a result, it can safeguard the integrity and diversity of content for future generations [8].

Complexity

It is crucial that our chatbot simplifies, rather than complicates, processes for newcomers. Conducting user studies to design and assess our AI-enhanced tool is key to ensuring it eases the editing process for novices.

Editorial Integrity

Recognizing the potential of LLMs to generate instances where the output is wrong or misleading, we will focus on integrating mechanisms to train the new editors to verify that the output can be verified with information from a reliable source. Additionally, we plan to ensure that the wider community of Wikipedia editors is informed about which articles have been edited with the assistance of LLMs. This step will allow experienced editors to scrutinize AI-suggested edits more closely, further safeguarding against the propagation of inaccuracies. It is crucial to emphasize that our generative AI tool is designed not to produce content for Wikipedia, but to serve as a support system, offering guidance to human editors who will be the primary content creators. Our research will assess the effectiveness of this guidance, taking into account the perspectives of both expert and novice Wikipedians.

Wikipedia Editing Rules

Finally, since LLMs might output believable but wrong information, we will consider the implementation of a retrieval augmented generation (RAG) system. It can direct the LLM to a predefined source to answer users' queries. By defining Wikipedia itself as our reputable source for its guidelines and processes, the chatbot can have the newest and most accurate information at its disposal. Our usability study will assess the accuracy of the AI's guidance from experienced Wikipedians' viewpoints, identifying situations where human expertise is necessary to complement or enhance AI advice.

Community impact plan

In order to contextualize our tools, we decided to implement our chatbot in Spanish Wikipedia. We have reached and included Pepe Flores from Wikipedia Mexico. With his help, we will be able to connect with the local Wikipedia community, and take its specific needs into account. He will be fundamental to facilitate our chatbot's adoption, improve communication with regional editors, and disseminate our findings locally.

Pepe Flores has worked on recruiting women to contribute to Wikipedia. His efforts will help us tackle the gender gap by both editing more articles about women from the Global South, and integrating them into the editing process.

Evaluation

To evaluate our generative AI tool, we will conduct a usability study and a real world deployment.

Usability Study

The usability study will examine the effectiveness and accuracy of our chatbots. It will analyze the labor outputs generated by participants, as well as garnering insights into the perceptions of both novices and expert Wikipedians regarding the chatbot's guidance, particularly, in terms of its accuracy and practicality. We will also measure participants' time completion times, which can help to understand the chatbot's impact on streamlining the editing process. We also plan to delve into the novices' reception of the system's recommendations, employing precision and recall techniques to quantify satisfaction and the relevance of the chatbot's recommendations. The findings of our usability study will help us to generate design recommendations for creating AI-enhanced chatbots for novice Wikipedians.

Real World Deployment

We will conduct a real world deployment to study how our AI enhanced tool can help to address gender information gaps on Wikipedia with regards to Latin American women. We will deploy our tool to assess its impact on the creation of articles and content by novice Wikipedians. This will be studied during workshops and edit-a-thons led by Pepe Flores, without compensating participants for their involvement in the deployment.

Additionally, we will study the system's influence within editorial discussions, specifically examining whether the guidance provided on navigating contested edits proves useful. This evaluation will extend to exploring alternative methods by which the chatbot could enhance the discussion process, without supplanting the invaluable role of human editors. Such an inquiry not only seeks to validate the utility of the chatbot, but also to assess the potential risks and boundaries associated with integrating AI into Wikipedia's editorial ecosystem. Ultimately, our chatbot represents a significant step forward in understanding the multifaceted implications of AI incorporation into Wikipedia's operational processes.

Budget

https://docs.google.com/spreadsheets/d/1D5zuouHJUYVJmYxRoRfmfqURXOsP4QnEZd3xqvaKM5o/edit?usp=sharing

Response to reviewers and meta-reviewers

Regarding Functionality

The chatbot will now offer step-by-step guidance for the editing process, and answer questions about Wikipedia's guidelines. It will not participate in discussions, but only explain the discussion process to newcomers and guide them on how to participate in discussions. We have also included a usability study and a real world deployment to evaluate our tool's functionality.

Regarding Large Language Models

Large Language Models (LLMs) have the potential to significantly improve the effectiveness of chatbots through enhanced interactions. We are optimistic that a chatbot powered by an LLM will improve the experience for newcomers contributing to Wikipedia.

Initially, we will employ GPT-4 for our prototypes, given our familiarity and success with it. However, we are also open to exploring open-source alternatives in the future, like those provided by Hugging Face.

Furthermore, the integration of Retrieval Augmented Generation (RAG) could provide

substantial benefits [12], enabling the chatbot to access the most current and accurate information on Wikipedia's guidelines and processes. This approach ensures that the chatbot is always informed by reliable sources.

Regarding the Gender Gap

We address the gender gap by increasing the number of articles on women from Latin America, and by integrating women to the editing process. Including Pepe Flores, a key member of the Wikipedia communities in Latin America will help immensely to achieve this goal.

Regarding Costs

We have reserved funds to use the GPT-4 API. Therefore, no specialized hardware will be required. We will later consider open source APIs, so the chatbot might be easier to implement and maintain. We will migrate to these open source APIs likely for our real world deployment to facilitate that our tool can be more widely used. We also plan to use UNAM's servers to host the tool and further reduce costs.

Apart from this, we are aware of techniques to reduce the costs of using APIs, such as LLM Cascade, Prompt Adaptation, and LLM Approximation [3]. Additionally, research indicates that large pretrained models, such as GPT-3, exhibit high efficiency after their initial training phase. For instance, generating 100 pages of content with a trained GPT-3 model would consume approximately 0.4 kW/hr of energy, translating to minimal energy costs of just a few cents [1].

Costs for evaluating the chatbot have been included in the budget.

Regarding the Meta Review

All funding will be allocated to researchers, consultants, and students in Mexico, and does not include U.S. universities. Dr. Savage, from Mexico City and the project's sole PI, will volunteer her advisory role, not drawing a salary from the budget. She will oversee all the research of this proposal. Dr. Garcia, with his background in computer engineering and literature, will lead the creation and evaluation of the generative AI chatbot. He will also lead the publication of related research papers. An UNAM graduate student will assist in implementing the AI chatbot, help in the user studies, and paper writing. Consultant Pepe Flores will offer insights into Latin American Wikipedia communities for better recruitment and tool dissemination. Budget adjustments have necessitated salary reductions.

References

[1] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. Language models are few-shot learners. *Advances in neural information processing systems* 33, (2020), 1887-1901.

[2] Birhane, A., Kasirzadeh, A., Leslie, D., & Wachter, S. Science in the age of large language models. *Nature Reviews Physics*, 5(5), (2023), 277-280.

[3] Chen, L., Zaharia, M., & Zou, J. Frugalgpt: How to use large language models while reducing cost and improving performance. (2023). arXiv preprint arXiv:2305.05176.

[4] Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: using intelligent task routing to help people find work in wikipedia. *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces*. January

(2007), 32–41.
https://doi.org/10.1145/1216295.12163092007.

[5] De Los Santos, M., García, J. E., Chávez, N. E., Navarrete, A., Martínez, C., Savage, S., et al. La Independiente: an AI-enhanced Platform Co-Designed with Latin-American Crowd-Workers. *Avances en Interacción Humano-Computadora*, Año 8, No. 1, (2023), 6 - 10.
http://doi.org/10.47756/aihc.y8i1.1292023.

[6] Farzan, R., Savage, S., and Flores, C. Bring on Board New Enthusiasts! A Case Study of Impact of Wikipedia ART+Feminism Edit-a-thon Events on Newcomers. (2016).

[7] Flores-Saviaga, C., Savage, S., and Taraborelli, D. LeadWise: Using Online Bots to Recruit and Guide Expert Volunteers. *CSCW '16 Companion: Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. February, (2016), 257-260.

[8] Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., and Ahmed, N. K. Bias and fairness in large language models: A survey. (2023) arXiv preprint arXiv:2309.00770.

[9] Garimella, P., and Varma, V. Learning through Wikipedia and Generative AI Technologies. In *Proceedings of the 16th International Conference on Educational Data Mining*, Bengaluru, India, July, (2023), 575–577.

[10] Geertz, C. (1973). The interpretation of cultures (Vol. 5019). Basic books.

[11] Lee, N. T., Resnick, P., & Barton, G. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. (2019).

[12] Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., et al. Holistic evaluation of language models, (2022). arXiv preprint arXiv:2211-09110.

[13] Morgan, J.T., Bouterse, S., Walls, H., Stierch, S. Tea and sympathy: crafting positive new user experiences on wikipedia. In *Proceedings of the 2013 conference on Computer sup-ported cooperative work*. ACM, (2013), 839–848.

[14] Muscinat, D., Ren, Y., Johnson, J. A., and Riedl, J. Mentoring in Wikipedia: a clash of cultures. *WikiSym '11: Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, (2011), 173–182.
https://doi.org/10.1145/2038558.20385862011.

[15] Sandberg, J., and Soukas, T. Making sense of the sensemaking perspective. Its constituents, limitations, and opportunities for further development. *Journal of organizational behavior 36, S1* (2015), S6-S32.

[16] "Usuario:SeroBot." Wikipedia: The Free Encyclopedia. Wikimedia Foundation, March 16, (2024).
https://es.wikipedia.org/wiki/Usuario:SeroBOT.

[17] Vassilev, A., Booth, H., & Souppaya, M. Mitigating AI/ML Bias in Context: Establishing Practices for Testing, Evaluation. (2022).

[18] Yuqing, R., Haifeng, Z., and Robert E. K., How Did They Build the Free Encyclopedia? A Literature Review of Collaboration and Coordination among Wikipedia Editors. *ACM Trans. Comput.-Hum. Interact. 31, 1, Article 7, 48 pages*. November, (2023).
https://doi.org/10.1145/3617369