

COVER WOMEN: A comparative study of Wikipedia's front page content from a gender and intersectional perspective with volunteer-driven insights and their newsroom guidelines

Núria Ferran-Ferrer, Miquel Centelles Velilla & Laura Fernández
Faculty of Information and Communication, Universitat de Barcelona

Abstract

This proposal presents a research project that will look into the most popular Wikipedia page. This page, known as the main page, or front page from a communication perspective, will be analysed across the seven longest-standing Wikipedia editions: English, German, Catalan, French, Portuguese, Italian, and Spanish. Grounded in a gender and intersectional perspective, this study will delve into the daily content, newsroom guidelines (principles and standards that guide the dissemination of information), and volunteer community insights. The examination will employ communication theories like gatekeeping and agenda-setting. Beyond academic research, our goal is to actively contribute to editing communities by addressing the daily challenges and needs in crafting front-page content.

Introduction

Despite Wikipedia being a key player in the public sphere and having a transformative impact on information dissemination, Wikipedia grapples with **persistent gender bias in both editing and content** (Antin et al., 2011; Bear and Collier, 2016; Wagner et al., 2016;

Hinnosaar, 2019; Minguillón et al., 2021; Ferran-Ferrer, Boté-Vericad, et al., 2023) **alongside additional prejudices** (Redi et al., 2021; Beytía et al., 2022). Bias in contributions perpetuates imbalances in content coverage and discourages diversity, which further exacerbates the issue (Worku et al., 2020).

Scholars highlight the need for a comprehensive understanding of Wikipedia's knowledge production culture to address these biases and make Wikipedia more robust, reliable, and transparent (Menking and Erickson, 2015). Reducing the gender and other intersectional biases necessitates more than acknowledging Wikipedia as a mirror of societal biases—it involves addressing the platform's deeper logic embedded in its techno-scientific project (Ford and Wajcman, 2017).

We have selected the most popular Wikipedia page for analysis. This page, commonly referred to as the main page, or front page from a communication perspective, is accessible in all language editions of the global encyclopedia, and we will conduct our study on it. We will research into the possible gender and intersectional bias in its daily content, in its newsroom guidelines (principles and standards

that govern the dissemination of information), and in the insights from the volunteer community who decide which information gets disseminated to the public on the main page. This research will utilize communication theories such as gatekeeping which examines the process by which information is filtered, selected, and ultimately presented to the public (Barzilai-Nahon, 2009) and agenda-setting (McCombs and Shaw, 1972) which studies the effect.

Therefore, the research questions that we address are:

- **Research Question 1 (RQ1):** What insights do interviews with volunteer gatekeepers (editors of the main page of Wikipedia) provide on **decision-making, biases, and strategies affecting the visibility of gender and intersectionality-related content on Wikipedia's front page**, particularly regarding how their preferences and interests, shape the topics featured?
- **RQ2:** How does **gatekeeping impact gender gaps** in content representation on digital platforms, specifically in the peer production of knowledge (**decision-making system on suitable content and what is not**) within newsrooms or **editorial policies**, and why is understanding this phenomenon crucial for addressing gender disparities?
- **RQ3: How does agenda setting influence the selection of frames and sentiment** adopted by Wikipedia pages concerning specific issues or events, and how does it shape the focus and intensity of user edit activity within Wikipedia?

- **RQ4:** How prevalent is **gender and intersectional bias in the content** featured on Wikipedia's front pages?

This research is necessary to draw further attention to the need for systemic change within the platform's newsroom/editorial practices to address disparities in gender and diversity representation in online knowledge and foster a more inclusive and diverse digital information landscape.

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

Related work

To contrast the feasibility of this proposal with seven language editions of Wikipedia, **we have already conducted a micro project with a sample of the English and Spanish Wikipedia to assess the viability of the global project.**

That is:

- a) If there are **open and formalized recommendations and guidelines that determine which contents are published** on the main page and if the publication criteria can be analysed.
- b) At the same time, we were interested in seeing if with **data wrangling techniques we could work with the biographies published on all Wikipedia main pages and analyse them from a gender and intersectional perspective** using the properties of Wikidata.
- c) Finally, we highlight the ease of contacting the community that performs gatekeeping tasks, and we begin to prepare the relevant questions to understand the decision-making process, editorial practices, and identify the issues that may be relevant to understanding the phenomenon.

The results of this previous trial work, with two language editions, will be published soon (Ferran-Ferrer et al., 2024). The trend is not encouraging if we take into account that bias in contributions perpetuates imbalances in content coverage and discourages diversity, which further exacerbates the issue (Worku et al., 2020).

To address this, scholars stress the importance of understanding Wikipedia's knowledge production culture to tackle its gender gap (Menking and Erickson, 2015). Addressing this issue requires delving into the foundational principles driving the platform's techno-scientific project (Ford and Wajcman, 2017; Geiger, 2017), necessitating the recognition and dismantling of exclusionary practices (Menking and Rosenberg, 2021).

Communication theories like gatekeeping and agenda-setting provide valuable frameworks for understanding Wikipedia's potential biases. Gatekeeping theory, focusing on information filtering processes, is applied to scrutinize stories selected for the Front Page, which attracts millions of readers monthly (Barzilai-Nahon, 2009; Wikimedia, 2023). Gatekeeping theory has previously been applied to Wikipedia by researchers to further understand biases in content selection and presentation (Li and Farzan, 2020) and to advocate for a reorganization of online spaces to democratize content and encourage dialectical gatekeeping that could reduce racial and other disparities (Ezell, 2021). Additionally, drawing from agenda-setting theory, we examine how Wikipedia's main page influences viewers and shapes news hierarchy, including its agenda-building power (McCombs and Shaw, 1972; Ren and Xu, 2023). Agenda setting can impact the choices of frames and sentiment adopted by Wikipedia pages regarding a particular issue or event (Lee, 2018) and it can play a role in shaping the focus and intensity of

user edit activity in Wikipedia (Mahabir et al., 2018).

This study goes beyond affirming Wikipedia's reflection of reality to delve into its systemic challenges (Ford and Wajcman, 2017). It analyses not only main page content selection but also newsroom guidelines, including interviews with gatekeepers, to enhance understanding and address systemic issues.

Methods

This research proposal outlines a study on gender representation and biases on Wikipedia's main page, the most visited Wikipedia page, the main page (or front page from a communication perspective), which got 46.8 billion visits last November on the English edition (Wikimedia, 2023). We will do a **comparative analysis across seven longest-standing Wikipedia editions**, English, German, Catalan, French, Portuguese, Italian, and Spanish, all of them born in 2001, employing a **mixed-methods approach**.

Grounded in **gender and intersectionality**, the study will analyse daily content, editorial/newsroom guidelines, and insights from volunteer communities using communication theories like gatekeeping (Barzilai-Nahon, 2009) and agenda-setting (McCombs and Shaw, 1972).

Our aim is not only academic research, but also active contribution to editing communities by addressing daily challenges in crafting front-page content. Therefore, in the project's work team, we have already included seven working groups of Wikipedia users involved in gender for each language edition and the chapters of all the Wikipedias analysed in this project (See Table 2).

The first stage of the project will be:

- a) To conduct a scoping review, a **systematic literature review** using the SALSA Framework (Grant and Booth, 2009) to analyse the academic publications from 2001 to 2024. This review **will concentrate on examining Wikipedia within the framework of a communication ecosystem.**

Then, we will employ a triangulation methodology.

- b) **In-depth interviews** with voluntary editors of the front page from all seven Wikipedia editions to ascertain decision-making processes, biases, and strategies that influence content visibility related to gender and other intersectionalities. The interviews will be conducted in person or online and in the native languages of the volunteer participants. We plan to make around five interviews by language edition. Contacts with the volunteers will be obtained through discussion pages related to editing the main page, as well as from user groups participating in the project, such as calls from the same chapters to their networks. The interview transcriptions will be coded and analysed using qualitative data analysis software, and a specific codebook will be generated to facilitate the coding. **This methodological approach will address RQ1 and RQ3.**
- c) **Newsroom guidelines:** We will apply **content analysis to main-page, or front-page editorial guidelines**, for each language edition, and we will explore what leads the decision-making of the gatekeepers who determine story prominence. The content of these guidelines will be coded and analysed

using qualitative data analysis software, and a specific codebook will be generated to facilitate the coding.

This research strategy will tackle RQ2.

The analysis of the **qualitative approach to agenda setting and gatekeeping practices (RQ1-3)** will be conducted independently with two codebooks, one for the interviews and one for the editorial policies. However, each codebook will encode elements specific to gatekeeping and agenda setting to obtain evidence that corresponds to the theoretical framework.

- d) **Main-page content quantitative analysis:** We will scrutinize the content (biographies) on the front page in each of the seven language editions for ten years, with **data wrangling**. To do so, first, **we will identify the sections of the main page that are consistently present across all Wikipedias and are easily comparable.** Wikipedia's front pages regularly feature changing content, offering a snapshot of current events, featured articles, and useful links. It's important to note that volunteers maintain these main pages and may evolve in format and content over time.

For each language edition, a unique method will be employed to retrieve the content and data of its main page from the past ten years, as the URLs of previous main pages cannot be obtained from the dumps. **Quantitative analysis will begin by scraping through the open-source tool OpenRefine to reconcile the URIs found in the sections of Wikipedia covers in both language editions.** This process will

enrich them with specific properties from Wikidata to obtain values of the selected properties for study: like P21 (sex or gender), P106 (occupation), P172 (ethnic group), P103 (native language) and others. OpenRefine, utilized in various contexts and applications, is essential for this research as it enables the preparation and analysis of vast amounts of data.

This method will respond to RQ4.

Table 1 offers a comprehensive overview of the research proposal.

Expected output

The specific research outputs that we envision for our proposed project include, but are not limited to:

- **Scientific publications:**
We will draft scientific publications for each research question and assess whether the approach is comparative across all editions or if it is better to separate them by smaller communities, editorial process typologies, etc. This will be determined once the study is completed to decide on the best dissemination approach.
- The dataset, emerging from RQ4, will be made available as downloadable dumps, and will be accessed via public APIs and a SPARQL endpoint.
- Participation at least at these conferences:
 - Wikiworkshop
 - Wikimania
 - WikiWomen Camp
 - Each user group and chapter will participate in national or regional events with Women Cover results.
- **Tools to support the editorial tasks** of gatekeeping, namely:

- **Guidelines for content selection on front pages** that are attuned to intersectionality and gender diversity;
- **Bots and AI assistants that facilitate the content selection process for front pages**, with a focus on acknowledging intersectionality and gender differences. Both tools will be developed with a focus on considering the collaborative environment and consensus-driven approach characteristic of Wikipedia.
- **Resources aimed at enhancing the archiving and curation of main page content** across all Wikipedia editions outlined in this proposal.

For each output, explain who the primary intended audience for the output is and what benefit, if more specifics are available, they can gain by receiving the output. If you have specific publication venues, conferences, and so on in mind, please list these.

Risks

This project poses the challenge of working with **seven different editions of Wikipedia**. Although in the research team, we are used to investigating Wikipedia in English, Spanish, and Catalan, we are now facing a new, more multicultural and multilingual challenge. Furthermore, when working with different editions of Wikipedia, editorial policies, front page archive systems, etc., will follow different procedures that we will need to understand and find ways to address to satisfy the research questions. For this challenge, we count on the support of chapters and user groups, which not only prevent the challenge from causing fear but also add more excitement, if possible, and more diversity to the proposal.

Regarding the diversity of Wikipedia editions that we will cover, there is a **technical risk that some Wikipedia editions may not have a formal archiving procedure**, resulting in periods when front pages have not been stored, etc. From a small sample, we have found that Catalan Wikipedia, for example, does not store the main page systematically overtime, and we will explore potential solutions to access its content. At the very least, if we cannot analyse all the front page content, we will still be able to examine the editorial policies determining what is published on the most visited page and the subjectivities and values volunteers bring when selecting topics. For all cases, after the research is performed, we will share a **guide on how to archive front pages sustainably and effectively**.

Another aspect that could pose a **risk is the potential for fewer participants than anticipated, as is common in any research requiring interview participation**. This may result in extended deadlines and other challenges. We rely on user groups to minimize this risk and help us find connections and candidates for interviews with the voluntary individuals who edit the homepage. Another potential risk will be the potential reluctance of participants to be interviewed orally, either face-to-face (if possible) or via video call. In these cases, we will take the risk of adapting the interview to a written format in order to prioritize the well-being and comfort of the participants, besides the potential loss derived from replacing the classical, more interactive and open, methodology of the oral personal interview.

In addition, we are mindful of the **potential risks of harming participants by revealing their identity, and therefore we will follow the ethical guidelines and protocols established by the ethics committee of the University of Barcelona**. Participants will be provided with

information about the project and the potential risks of participation, they will have the opportunity to ask questions to the researchers, and they will later be asked to sign an informed consent form. Participants will have the right to withdraw their consent and request that their data be deleted at any time during the research.

Finally, as one of our methodologies necessitates the examination of data derived from extensive periods of front pages, **access to reliable and robust services provided by the free/open resources for Wikimedians becomes essential**. Services such as [Wikimedia Cloud Services](#) are crucial for running scripts, while [PAWS](#) is indispensable for Jupyter notebooks, and OpenRefine supports data refinement processes.

Community impact plan

The project aligns with the Wikimedia Movement's 2030 strategy by focusing on delivering knowledge as a service and addressing equity in knowledge and communities overlooked by structures of power and privilege.

Furthermore, Cover Women project will involve a team of 5 researchers, professors from the University of Barcelona, one from the UOC, and a PhD student, with a multidisciplinary perspective, as we have individuals from the fields of communication, semantic web, digital humanities, and computer science. Additionally, this project proposal has been designed according to the needs of various activist groups regarding gender equality on Wikipedia, as well as with the boards of the chapters involved in each language edition. See Table 2 to anticipate the impact on communities we will reach. These users are groups of Wikipedia users who work to achieve a better Wikipedia by introducing a gender perspective. Since we are working with 7 different editions of Wikipedia, we have

considered that having a user group of female editors for each edition and a representation from each chapter's board would be interesting to achieve the project's objectives and meet the real needs of the communities.

This project will provide:

- a) **Decade-long insights into gender and intersectionality content** representation on Wikipedias' front pages. (RQ4)
- b) Beyond descriptive stats, we'll reveal **bias trends**. (RQ4)
- c) **Editorial strategies for gatekeeping and agenda setting**. (RQ1-3)
- d) **Guidelines for ethical content selection using AI and bots**. (RQ3)
- e) **Technical guidance to enhance data archives** on main pages. (RQ4)
- f) Collaborative work with volunteers ensures inclusivity, integrating advocate perspectives for a **consensus-driven approach**.(RQ1)

Built on in-depth interviews and stable and lasting collaboration with Wikipedia chapters and user groups, this work addresses gender identity under-representation.

We will utilize Wikipedia's consensus-based decision-making approach to address our research questions. This method prioritizes addressing the legitimate concerns of its editors and finding a middle ground, all while adhering to Wikipedia's established policies and guidelines. In this context, it is crucial to consider that consensus naturally evolves among editors as they make changes, the importance of quality arguments in determining consensus, the allowance for consensus to evolve based on new evidence, and the acknowledgment of decisions beyond the scope of editor consensus. This methodology underscores Wikipedia's emphasis on collaboration, incremental progress, and

communal harmony in managing a large crowd-sourced encyclopedia.

Evaluation

The measure of success could include:

- **Addressing the needs of user groups within Wikipedia, such as Wikimujeres, Women in Red, etc., who edit with a gender perspective** and are involved in the project, by facilitating content reviews on front pages. Subsequently, delivering guidelines and strategies based on their feedback to mitigate biases present in these pages.
- **4 publications in peer-reviewed journals**.
- Conference and event participation in different countries to **raise awareness about biases in the content featured on Wikipedia's main page and create engaging guidelines to encourage adherence**.
- **Media coverage** by delivering press release notes to national media but also to all chapters to generate interest in different countries.
- **Following the ethical standards and gender-sensitive approach** demonstrating integrity and responsibility.

Budget

[Budget details](#)

Response to reviewers and meta-reviewers

Thank you sincerely for reviewing our research proposal. Your insightful feedback has greatly enriched our project. We truly appreciate your valuable support and guidance.

1) Firstly, **we want to clarify that the project proposal “Cover Women” we are presenting has not yet been implemented and has not received any funding.** We have an ongoing project funded by the Ministry of Science and Innovation of the Government of Spain that ends in August 2024, which is called [Women and Wikipedia](#). In this Spanish project, we are studying female participation in the fourth largest (by the number of registered editors), Spanish, Catalan, Basque, and Galician Wikipedias, representing a wide range of cultures in Europe and Latin America. Aside from the Spanish language, Spain has four other official languages, so the research project seeks to develop a comprehensive approach that also includes Euskal Wikipedia, Viquipèdia, and Galipedia. We are interested in examining these Wikipedias not only because they belong to the same country but also because they have smaller communities, a characteristic that appears to be related to the development of a less hostile environment for women editors. Therefore, with the "Cover Women" proposal, we will focus on a different matter, and we will study the WP main page, employing communication theories, and including seven global scope Wikipedias, outside our country. From our perspective, it's a completely different project from the current project [Women and Wikipedia](#), although it aligns with our concern, which is the gender and diversity gap on Wikipedia.

That being said, as we have explained in the Related Work section, **to test the technical feasibility of the proposal, especially the part concerning the analysis of front-page content, we conducted a trial using both the English and Spanish versions of Wikipedia.** In this way, we were able to identify the most suitable method for conducting comparative analyses of front-page sections across different Wikipedias, given that each of them structures and stores front-page content differently. Therefore, this

preliminary work instils confidence in project execution and mitigates risks, while the tests conducted on two large-volume Wikipedia editions pave the way for embracing the remaining proposed editions. **This first explanation tries to answer the observation of the Official Reviewer xHbs21 and the Meta Review Chairs13.**

2) Secondly and regarding the proposed **scope of the project**, the idea is to generate academic publications, but what motivates us is to create guides and recommendations for the editorial community responsible for the front page on how to make content decisions without gender gaps and while respecting diversity. In our research, we strive to collaborate with entities that share our aspirations for making social changes. In fact, we see research as a tool in service to society and with an impact on its needs.

3) In third place, regarding the question about why does it need the WMF **to fund just one year of a PhD student**, well, we would like the WMF to be able to fund all the years that a doctoral student is working on their thesis, but the funding of this call is only for one year. **This addresses the inquiry posed by Official Reviewer zHbs31.**

4) In fourth place, **we have chosen these editions of Wikipedia because they are the first ones created and represent linguistic communities with whom we already have previous connections** through previous projects. Additionally, within the team, we cover the necessary linguistic needs to attempt to address the languages of the interview participants. At the same time, these editions will greatly enrich the analysis we have conducted so far on two Wikipedias, English and Spanish. We will see if the problem is common or if there are good practices in any of

them, such as affirmative actions in gatekeeping strategies, which could be extrapolated to the rest. **Methodologically, to address comparative analysis of cases, we will use the same codebook** for interviews from all Wikipedia language editions and the same shared codebook for all content analysis of editorial policies. Finally, in terms of sharing results and dissemination, we believe that preparing an academic article comparing content and editorial lines could also be very interesting.

5) In light of the **gender gap and gatekeeping inquiries**, restructuring research questions 2 and 3 in the Introduction, along with providing detailed explanations in the Methodology section, has significantly enhanced the clarity surrounding the imperative to investigate the functioning of gatekeeping in peer production and its implications for content biases. **This response to the query raised by Official Reviewer NGzx27.**

6) **In response to Official Reviewer NGzx27's inquiry** regarding clarifying the **method used to achieve the objectives outlined in the impact plan**, we have included the corresponding research question alongside each outcome of the plan. Additionally, since research questions in the methodology section are associated with methodological approaches and techniques, we believe we have provided a clearer explanation of how the impact will be realized.

7) Lastly, **we have addressed NGzx27's comment** by providing **clarification on the selection process for the sample of seven language editions of Wikipedia** in the Methodology section. This research proposal focuses on examining gender representation and biases on Wikipedia's main page, which is the most visited page on Wikipedia. Our study will involve a comparative analysis across the seven oldest Wikipedia editions: English,

German, Catalan, French, Portuguese, Italian, and Spanish, all established in 2001. Therefore, the criterion of the age of creation of these editions allows us to work with languages that the research team members are familiar with.

References

You can choose which format you use for showing the references. References will not be counted against your word limit.

Antin, Judd, Yee, Raymond, Cheshire, Coye, and Nov, Oded. (2011). Gender differences in Wikipedia editing. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, 11–14.
<https://doi.org/10.1145/2038558.2038561>

Barzilai-Nahon, Karine. (2009). Gatekeeping: A critical review. *Annual Review of Information Science and Technology*, 43(1), 1–79.
<https://doi.org/10.1002/aris.2009.1440430117>

Bear, Julia B., and Collier, Benjamin. (2016). Where are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia. *Sex Roles*, 74(5), 254–265.
<https://doi.org/10.1007/s11199-015-0573-y>

Beytía, Pablo, Agarwal, Pushkal, Redi, Miriam, and Singh, Vivek K. (2022). Visual gender biases in wikipedia: A systematic evaluation across the ten most spoken languages. *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 43–54.
<https://ojs.aaai.org/index.php/ICWSM/article/view/19271>

Ezell, Jerel M. (2021). Empathy Plasticity: Decolonizing and Reorganizing Wikipedia and Other Online Spaces to Address Racial Equity.

Ethnic and Racial Studies, 44(8), 1324–36.
<https://doi.org/10.1080/01419870.2020.1851383>

Ferran-Ferrer, Núria, Boté-Vericad, Juan-José, and Minguillón, Julià. (2023). Wikipedia gender gap: A scoping review. *El Profesional de La Información*, 32(5), e320617.
<https://doi.org/10.3145/epi.2023.nov.17>

Ferran-Ferrer, Núria, Fernández, Laura, and Centelles, Miquel. (2024). Wikipedia's Front page 10 years evolution: Analysis of gender and intersectionalities on content, newsroom guidelines and volunteer-driven insights in the Spanish and English editions. *Forthcoming*.

Ford, Heather, and Wajcman, Judy. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, 47(4), 511–527.
<https://doi.org/10.1177/0306312717692172>

Geiger, R Stuart. (2017). Beyond Opening up the Black Box: Investigating the Role of Algorithmic Systems in Wikipedian Organizational Culture. *Big Data & Society* 4(2), 205395171773073.
<https://doi.org/10.1177/2053951717730735>

Grant, Maria J., and Booth, Andrew. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108.
<https://doi.org/10.1111/j.1471-1842.2009.00848.x>

Hinnosaar, Marit. (2019). Gender inequality in new media: Evidence from Wikipedia. *Journal of Economic Behavior & Organization*, 163, 262–276.
<https://doi.org/10.1016/j.jebo.2019.04.020>

Lee, Philip T. Y. (2018). In Search of Public Agenda with Text Mining: An Exploratory Study of Agenda Setting Dynamics Between the Traditional Media and Wikipedia. In *Trends and*

Applications in Knowledge Discovery and Data Mining, edited by Mohadeseh Ganji, Lida Rashidi, Benjamin C. M. Fung, and Can Wang, 11154, 309–17. Lecture Notes in Computer Science. Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-030-04503-6_30

Li, Ang, and Farzan, Rosta. (2020). Collaboration of Open Content News in Wikipedia: The Role and Impact of Gatekeepers. In *Companion Proceedings of the Web Conference 2020*, 802–5. Taipei Taiwan: ACM.
<https://doi.org/10.1145/3366424.3383568>

Mahabir, Ron, Croitoru, Arie, Crooks, Andrew, Agouris, Peggy, and Stefanidis, Anthony. (2018). News Coverage, Digital Activism, and Geographical Saliency: A Case Study of Refugee Camps and Volunteered Geographical Information. *PLOS ONE* 13(11), e0206825.
<https://doi.org/10.1371/journal.pone.0206825>

McCombs, Maxwell E., and Shaw, Donald L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176–187.
<https://academic.oup.com/poq/article-abstract/36/2/176/1853310>

Menking, Amanda, and Erickson, Ingrid. (2015). The Heart Work of Wikipedia: Gendered, Emotional Labor in the World's Largest Online Encyclopedia. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 207–210.
<https://doi.org/10.1145/2702123.2702514>

Menking, Amanda, and Rosenberg, Jon. (2021). WP:NOT, WP:NPOV, and Other Stories Wikipedia Tells Us: A Feminist Critique of Wikipedia's Epistemology. *Science Technology and*

Human Values, 46(3), 455–479. Scopus.
<https://doi.org/10.1177/0162243920924783>

Minguillón, Julià, Meneses, Julio, Aibar, Eduard, Ferran-Ferrer, Núria, and Fàbregues, Sergi. (2021). Exploring the gender gap in the Spanish Wikipedia: Differences in engagement and editing practices. *PLOS ONE*, 16(2), e0246702.
<https://doi.org/10.1371/journal.pone.0246702>

Redi, Miriam, Gerlach, Martin, Johnson, Isaac, Morgan, Jonathan, and Zia, Leila. (2021). *A Taxonomy of Knowledge Gaps for Wikimedia Projects (Second Draft)* (arXiv:2008.12314). arXiv.
<http://arxiv.org/abs/2008.12314>

Ren, Ruqin, and Xu, Jian. (2023). It's not an encyclopedia, it's a market of agendas: Decentralized agenda networks between Wikipedia and global news media from 2015 to 2020. *New Media & Society*, 146144482211496.
<https://doi.org/10.1177/14614448221149641>

Wagner, Claudia, Graells-Garrido, Eduardo, Garcia, David, and Menczer, Filippo. (2016). Women through the glass ceiling: Gender asymmetries in Wikipedia. *EPJ Data Science*, 5(1), Article 1.
<https://doi.org/10.1140/epjds/s13688-016-0066-4>
Wikimedia. (2023). *Wikimedia Statistics: All wikis*.
<https://stats.wikimedia.org/#/all-projects>

Worku, Z., Bipat, T., McDonald, D.W., and Zachry, M. (2020). Exploring Systematic Bias through Article Deletions on Wikipedia from a Behavioral Perspective. *ACM International Conference Proceeding Series*. Scopus.
<https://doi.org/10.1145/3412569.3412573>.

Appendix

Table 1. Research project overview

Research Questions	Methodology	Methods and techniques	Outputs
RQ1-RQ4	Systematic bibliographic review/Scoping review	SALSA Method	*Open Data set * Theoretical background for the project
RQ1&3	Quantitative and qualitative approaches	In-depth interviews with voluntary editors of the front page from all seven Wikipedia editions Content codification of the interview transcriptions and analysis with a qualitative data analysis software, and a specific codebook. Collaborative work with volunteers ensures inclusivity, integrating advocate perspectives for a consensus-driven approach	*Academic and conference papers *Editorial strategies guidelines for gatekeeping and agenda setting * Guidelines for ethical content selection using AI and bots.
RQ2	Quantitative and qualitative approaches	Preparation of the corpus of guidelines to codify for each language edition. Content analysis to front-page editorial policies (main page editing guides) through a code book designed ad hoc and qualitative data analysis software. Collaborative work with volunteers ensures inclusivity, integrating advocate perspectives for a consensus-driven approach	*Academic and conference papers *Editorial strategies guidelines for gatekeeping and agenda setting.
RQ4	Main-page content	Data from the seven language edition main pages will be wrangled.	*Open Data set *Academic and conference papers *Decade-long and bias trend report into gender

	quantitative analysis.		and intersectionality content * Technical guidance to enhance data archives on main pages.
--	------------------------	--	---

Table 2. Contributors and Collaborators Across Seven Language Editions of Wikipedia

Name	Affiliation	Country	Rol in the proposal Cover Women	Wikipedia language edition
Núria Ferran-Ferrer	* Associate Professor at UB Universitat de Barcelona *Director of EDI Office UB *Director of Information and Communication PhD Programme, UB *Board Member of Amical Wikimedia *Member Committee WMF Grants in NW Europe	Catalonia, Spain	Co-IP	All
Miquel Centelles	* Lecturer at UB *Coordinator of Master’s Degree in Digital Humanities at Universitat de Barcelona *Research collaborator at Women and Wikipedia group.	Catalonia, Spain	Co-IP	All
Laura Fernández	*Postdoctoral researcher at Universitat de Barcelona *Researcher at Women and Wikipedia research group.	Catalonia, Spain	Co-IP	All
Juan-José Boté	Lecturer professor Information Science and Media Studies	Spain	Researcher	All

Florencia Claes	*Lecturer at Rey Juan Carlos University. *Head of Free Culture at Rey Juan Carlos University. *President of Wikimedia España	Spain	Advisor Spanish Wikipedia	Spanish
Ester Bonet	Member WikimujeresUG	Spain	Advisor from WikiMujeres	Spanish//Catalan
Núria Ribas	President of the Catalan Wikimedia Board	Spain	Advisor Catalan Wikipedia	Catalan
Michael David MILLER	Associate Librarian at McGill University Vice-President of Wikimedia Canada Board Member of WikiFranca	Québec-Canada	Advisor French Wikipedia	French
Natacha Rault	Director and founder of les sans pagEs	Switzerland	Advisor from Les sans pagEs	French
Daniela Brugger	Coordinator of who writes his_tory?	Germany	infodaniela brugger.ch	German
Lilli Iliev	Head of politics & public sector, Wikimedia Deutschland e. V.	Germany		German
Laurie Bridges	Founding member and leadership team for #EveryBookItsReader. Core Committee for WikiLibCon 2022. Librarian at Oregon State University, USA	United States	Advisor English Wikipedia	English

Rosie Stephenson-Gooknight	Wikimedia Foundation Board of Trustees Visiting Scholar, Northeastern University (Boston) Co-founder, Women in Red		Women in Red	English
Lilian	*Librarian at Universidade de São Paulo*Member of GLAM Bibliotecas da USP	Brazil	Advisor Portuguese Wikipedia	Portuguese
Ana Bragança	Co-founder and co-coordinator, Wiki Editoras Lx	Portugal	Advisor Wiki Editoras Lx	Portuguese
Irene	long-time editor and has been an administrator in itwik	Italy	Advisor Italian Wikipedia	Italian
Camelia Boban	WikiDonne project in Italian Wikipedia founder WikiDonne User Group co-founder	Rome, Italy	Advisor WikiDonne	Italian