Extended: COVER WOMEN II: Bias-aware approaches for equitable representation on Wikipedia's front page

Miquel Centelles
Universitat de Barcelona

Núria Ferran-Ferrer
Universitat de Barcelona

Abstract

In this research, we propose a comparative, interdisciplinary study to address gender and intersectional bias in the content featured on Wikipedia's main page. Building on the findings of the initial COVER WOMEN project, this extended study (COVER WOMEN II) explores how the concept of notability is constructed and applied across different language editions of Wikipedia, and how this influences the visibility of marginalized individuals.

We aim to develop a bias-aware metric of notability based on Wikidata, and integrate it into a prototype recommendation system to support equitable editorial decision-making on Wikipedia. Our research combines qualitative analysis of editorial guidelines and interviews with Wikipedia editors, quantitative modelling using structured and unstructured data, and participatory design methods to ensure community alignment.

By producing open-access tools, datasets, and inclusive editorial resources, the project seeks to support the Wikimedia ecosystem in fostering fairer content representation. The outcomes are designed to inform policy, guide editors, and ultimately contribute to a more inclusive and equitable digital knowledge infrastructure.

Introduction

This research project builds upon the findings of the **initial *COVER WOMEN* study**, funded by a Wikimedia research grant, which examined gender and intersectional bias in the representation of individuals featured on Wikipedia's front page across seven linguistic editions during a decade (2014-24). Through the analysis of thousands of biographies and nearly a hundred editorial policy guidelines, as well as interviews with volunteer editors responsible for front-page content of each of the seven Wikipedia editions, the study identified systemic patterns of exclusion and unequal visibility. It highlighted how editorial practices—shaped by community norms and technical infrastructures—tend to reproduce existing inequalities, particularly in terms of gender and other intersecting dimensions.

The present proposal deepens and expands this work by focusing on the notion of *notability* as a key mechanism in content selection. By developing an objective and bias-aware notability metric and integrating it into decision-making tools, this project aims to provide concrete solutions to promote more equitable representation across Wikimedia platforms.

a) Why this problem matters to Wikimedia

This research is **deeply aligned with the Wikimedia Movement Strategy's commitment**

to Knowledge Equity. By focusing on the systemic underrepresentation of women and marginalized groups **not only on Wikipedia's most visible space—the Main Page—but also within editorial decision-making processes such as article deletion discussions**, this project directly challenges the historical structures of privilege that shape knowledge production. It addresses **how notability is defined and who gets to be deemed worthy of inclusion**, providing concrete, data-driven tools to support a more inclusive editorial culture. In doing so, it seeks to dismantle both technical and social/cultural barriers to equitable participation, empowering editors from diverse backgrounds and helping ensure that the visibility and preservation of knowledge on Wikipedia reflects a broader, more just representation of global experience.

This proposal aims to address the **lack of objective, statistically and mathematically accountable data on gender and intersectional bias** in the selection of notable individuals featured on Wikipedia covers. Additionally, we aim to create **a decision-making tool for highlighting individuals based on an objective calculation of their notability**. The outcome of this notability measure can be integrated into a recommendation tool for suggesting individuals for inclusion on covers, Wikimedia projects and support arguments to avoid article deletion, thereby **assisting Wikipedia editors in their decision-making process**.

The design and implementation of the recommendation tool, as well as the overall conception of the project, benefit from the knowledge that the workgroup has acquired about Wikidata. This knowledge of **Wikidata includes its data model based on a knowledge graph and its ability to address the imbalance present in Wikipedia regarding gender equality and intersectionality.**

b) Tackling bias for a more inclusive Wikipedia
Wikipedia grapples with persistent gender bias in both editing and content (Antin et al., 2011; Bear & Collier, 2016; Wagner et al., 2016; Hinnosaar, 2019; Minguillón et al., 2021) alongside additional prejudices (Redi et al., 2021; Beytía et al., 2022). Reducing the gender and other intersectional gaps 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 & Wajcman, 2017). By contrast, it is essential to acknowledge that Wikidata's potential biases are no greater than those present in the real world (Zhang & Terveen, 2021). An analysis of Wikidata entries for male and female MEPs reveals equal amounts of property-value pairs, contradicting earlier studies that found Wikipedia content related to women emphasized family and relationships. Differences related to real-world disparities suggest that the structured data of Wikidata might be less prone to bias.

c) Research Questions:
Therefore, the **four research questions** that we address are:

- **Research Question 1 (RQ1):** How can the criterion of notability be measured objectively within the Wikimedia projects, specifically Wikipedia?

- **Research Question 2 (RQ2):** What role can Wikidata play in the development of this mathematical model, particularly regarding its enhanced representation of gender and other real-world intersections, as well as its foundation in ontology and factual data?

- **Research Question 3 (RQ3):** How can the objective measurement of notability be applied to the recommendation of

articles about individuals for inclusion on Wikipedia's main page?

- **Research Question 4 (RQ4):** What cultural shifts are necessary to implement a decision-making mechanism based on objective measures of notability for the inclusion of individuals on Wikipedia's main page?

Date: Specify the start and end time of your proposal. Proposed work can start as early as July 1, 2025 and should conclude by the grant's end time June 30, 2027.

Related work

The issue of gender and intersectional bias in Wikipedia has drawn increasing scholarly attention over the past decade, positioning Wikipedia as both a mirror of societal inequalities and a battleground for epistemic justice. Foundational studies have highlighted the platform's persistent gender gap in both content and participation, with less than 20% of biographies featuring women and an even smaller percentage of female or gender-diverse editors (Minguillón et al., 2021; Tripodi, 2023). These disparities are compounded by broader systemic patterns that marginalize non-Western, racialized, and working-class individuals.

Our research builds on and contributes to a growing body of work that examines Wikipedia through the lens of gatekeeping (Barzilai-Nahon, 2009) and agenda-setting theories (McCombs & Shaw, 1972). Prior studies have largely focused on article content or editor demographics. However, few have critically analyzed Wikipedia's *Main Page*—the platform's most visible and symbolically powerful space—as a site where representational decisions are made daily. The *Cover Women* project addresses this gap by examining how editorial guidelines and volunteer practices contribute to patterns of inclusion and exclusion across seven Wikipedia language editions (Ferran-Ferrer et al., 2024).

Beyond content analysis, our work in COVER WOMEN introduces a triangulated approach combining editorial policy review, qualitative interviews with volunteer editors, and large-scale Wikidata-enhanced analysis of 22,924 biographies featured on Wikipedia's front pages between 2013 and 2023. This approach reveals that decisions around notability, quality, and neutrality are not value-neutral, but reflect culturally embedded biases—often unintentionally—through implicit editorial norms and selection criteria.

Our findings resonate with existing research that critiques the systemic and structural barriers embedded in Wikipedia's architecture (Ford & Wajcman, 2017; Menking & Rosenberg, 2021). Yet our study moves a step further by proposing a measurable, reproducible model for assessing *notability* using Wikidata properties, which can inform inclusive recommendation systems for content selection. This methodological innovation seeks to address both the descriptive (what is happening?) and normative (what should be done?) aspects of representation on Wikipedia.

Various articles have addressed bias in notability on Wikipedia and analyzed the importance of designing and implementing solutions to mitigate these inequalities, including the design of computational tools and quantitative data analysis. Taraborelli and Ciampaglia (2010) emphasize the significance of community deliberation dynamics in content inclusion decisions, suggesting that using tools to promote diversity among active voices during these discussions could reduce the biases

present in these decisions. Wagner et al.(2016) (2016) analyze metadata structures that influence article visibility, proposing that tools could be designed to adjust the presentation of these articles and improve their representativity. Margolin et al. (2016) discuss the use of socio-technical data to forecast page enablement, which could inspire the creation of tools that analyze patterns in article creation and help identify topics that require greater coverage. Ferran-Ferrer et al. (2022), in their discussion on recommendations for more equitable inclusion, suggest that tools could be designed to facilitate contributions from diverse voices and track the inclusion of biographies of women. Laouenan et al. (2022) present a database of notable individuals that employs deduplication and verification techniques to minimize biases in notability by recording information about the diversity present in the articles. Ford et al. (2023) analyze how care work is not recognized as notable on Wikipedia, proposing a data analysis approach to highlight contributions in traditionally undervalued areas. Finally, Lemieux et al. (2023) use machine learning techniques and natural language processing to analyze pages of academics nominated for deletion, indicating that this type of analysis could inform the development of tools that help identify and address racial and gender biases in biographies.

Also, we have precedents regarding the application of recommendation systems to identify notable entities for inclusion in Wikipedia articles. Since 2009, with Korsgaard and Jensen (2009) and Yuan et al. (2009), we have located a dozen works on the application of recommendation systems for articles that are considered necessary or important for editing on Wikipedia. It is worth noting that few works focus on promoting biographical articles about notable people or celebrities; this is the case with Lin et al. (2010).

Regarding the purpose of the recommendations, we observe an interest in improving articles, not always the inclusion of articles about non-existent entities. Thus, Wulczyn et al. (2016) present a system to recommend articles that are missing in one language, based on the interests of editors. In line with this, Yuan et al. (2009) developed SuggestBot, a system focused on recommending articles that need improvements, rather than articles that do not exist, reflecting a need for diversification in the functionalities of recommendation systems.

In terms of recommendation models, the articles address different approaches. Zhang et al. (2014) present a graph model for predicting editor interest in articles, which is based on an integration of social and editing properties; on the other hand, Wulczyn et al. (2016) design a recommendation system that employs a machine learning model to enhance the growth of Wikipedia in various languages, implementing a regression model to predict the importance of articles.

Regarding the data used by recommendation systems, the use of internal Wikipedia data is predominant, as it serves as a foundation for personalizing editing experiences and facilitates the customization of recommendations. Generally, external data is based on the prior activity of editors and interactions within the platform (Yuan et al., 2009; Haisu Zhang et al., 2014; Moskalenko et al., 2020). However, the incorporation of external data can significantly enhance the quality and relevance of recommendations, especially in contexts involving translations or articles that already exist in other languages. From the external side, Wikidata is applied in the works of Wulczyn et al. (2016) and AlGhamdi et al. (2021; 2022), due to its privileged link with Wikipedia. Both highlight that Wikidata provides structured data to enhance recommendations, allowing for

increased relevance and personalization of the recommended articles.

It has been recurrently pointed out that the developed recommendation systems are not without limitations. Concerns about data scarcity are common; for instance, Alghamdi et al. (2021) emphasize that high inequality in participation can negatively affect the quality of recommendations. Schwarzer et al. (2016, 2017) also highlight that the difficulty in assessing the effectiveness of recommendation systems can complicate their acceptance and improvement within the community. Finally, the importance of making recommendations explainable is mentioned by Alghamdi et al. (2022), which underscores the need for systems to provide clear information on why an article is recommended, thereby fostering trust among editors in the recommendations.

By addressing the intersection between data infrastructure, editorial culture, and representational politics, our research contributes to filling a critical gap in Wikipedia studies: how the most visible elements of the platform—the Main Page and its editorial gatekeepers—shape global perceptions of knowledge and who is deemed worthy of visibility.

Methods

This research adopts a triangulated methodological approach, combining qualitative and quantitative methods, as well as user-centered design techniques. The methodological framework is structured around the project's four research questions and is designed to both conceptualize and test mechanisms for evaluating and recommending notability in a bias-aware manner. (See Cover Women II Time Line in the Annex).

RQ1 – Conceptualization of Notability and Editorial Policy Analysis (Qualitative)

To understand how Wikipedia's editorial policies shape gender and intersectional representation, we will:

- Conduct a **scoping review** of Wikipedia's editorial guidelines (e.g., notability, neutrality, reliable sources, deletion policies).
- Perform **content analysis** and **critical discourse analysis** of these policies to identify how bias might be embedded or reproduced.
- Create a **conceptual map** outlining the mechanisms of bias and develop an **academic paper** on the conceptualization of notability, drawing from a multidisciplinary perspective (communication, data science, gender studies, etc.).

RQ2 – Quantitative Assessment of Notability (Quantitative)

To define and assess objective measures of notability:

- Apply **inferential statistical methods** to analyze structured data (from Wikidata) and unstructured data (from Wikipedia articles), focusing on features such as occupation, references, and interlinkage.
- Develop a **notability vector and scoring model** that reflects inclusive and bias-aware standards.
- Generate and release an **open-access dataset** and produce an **academic paper** presenting the model and its potential for fairer representation.

RQ3 – Recommendation System for Inclusion (Quantitative + User-Centered Design)

To operationalize notability into actionable tools:

- **Model a system** based on the previously defined notability metrics using **bias-aware algorithmic design**.
- Develop and implement a **prototype recommendation tool** that integrates data from Wikidata and other linked structured datasets.
- Use **knowledge graph** technologies and **machine learning**, where applicable, to support scalable recommendations.
- Deliver an **integration guide**, a **recommendation system**, and an **evaluation report** analyzing the potential of these tools to mitigate bias in Wikipedia's front page selections and deletion discussions.

RQ4 – Cultural and Editorial Shifts for Implementation (Qualitative)

To explore adoption within the Wikipedia community:

- Conduct a **comparative analysis** between community-based and algorithmic content suggestion systems.
- Facilitate **testing sessions** in real scenarios (e.g., featured articles, deletion debates).
- Engage in **co-creation sessions**, **focus groups**, and **interviews** with editors to understand perceptions and gather feedback on inclusivity, fairness, and cultural acceptability.
- Output will include **policy recommendations**, an **editorial communication and diversity toolkit**, and **academic publications** supporting inclusive content governance.

This multimethod strategy ensures that the project not only proposes theoretical advancements but also delivers actionable tools and insights for the Wikimedia ecosystem.

Expected output

The COVER WOMEN II project will produce a range of outputs designed to benefit both academic and non-academic stakeholders, including Wikimedia editors, policy-makers, and the broader research and open knowledge communities. The outputs are structured to align with each research question and reflect the multidimensional nature of the project.

a) Scientific Publications

Audience: Scholars in communication studies, data science, gender studies, information science, and digital humanities.

Outputs:

- Academic paper on the **conceptualization of notability**, with a focus on how Wikipedia editorial policies influence gender and intersectional representation.
- Academic publication on **bias-aware notability assessment**, presenting the statistical model and its implications.
- A third academic paper discussing the **recommendation system** and the sociotechnical dimensions of implementing it within a collaborative knowledge platform.
- A fourth paper outlining **cultural shifts and editorial governance recommendations**, based on co-creation sessions and real-world testing.

Suggested journals and venues: *New Media & Society*, *Information, Communication & Society*, *Feminist Media Studies*, *Wikimania*, *International Conference on Web and Social Media (ICWSM)*, WikiWorkshop and others.

b) Open Datasets and Tools

Audience: Data scientists, developers, researchers, and Wikimedia technical communities.

Outputs:

- An **open-access dataset** including the variables and metadata used to assess notability across biographies.
- A **bias-aware notability scoring model** that can be reused and adapted for future studies or tool development.
- A **recommendation system prototype** for suggesting individuals for inclusion on Wikipedia's front page or other high-visibility spaces.

Benefits: These resources offer reproducibility, transparency, and potential integration into Wikimedia's content governance or community-developed tools.

c) Decision-making Support Tools

Audience: Wikipedia editors, Wikimedia Foundation policy makers, and affiliates.

Outputs:

- An **evaluation report** on Wikidata's capability to reduce content bias in front-page selection and deletion discussions.
- An **integration guide** detailing how to embed the notability model and recommendation system within Wikimedia workflows.
- A **communication and diversity toolkit** for editors, designed to support inclusive editorial decisions. This toolkit will include editorial policies and guidelines with gender and diversity perspectives.

Benefits: These outputs provide practical and actionable support to enhance equity and **visibility on Wikipedia.**

e) Events and Community Engagement

Audience: Wikimedia communities, volunteers, and advocacy groups.

Outputs:

- **Co-creation workshops** and **focus groups** with editors to test the tools and discuss inclusion strategies.
- Presentations at **Wikimedia events** (e.g., Wikimania), conferences, and public webinars to disseminate findings and encourage adoption.

Benefits: Engaging the community ensures transparency, relevance, and increases the likelihood of long-term impact and tool integration.

Risks

Five potential risks are envisaged for the proposed research project, which are described below, along with the strategies to implement in order to mitigate them.

1. Data Availability and Quality
Risk: The availability and quality of data from Wikipedia, Wikidata, and linked external sources may be limited or biased, affecting the reliability and validity of the research findings.
Mitigation: Implement robust data validation and cleaning procedures. Diversify data sources to include a broader range of perspectives and reduce reliance on potentially biased datasets.

2. Methodological Challenges
Risk: Applying inferential statistical methods and developing objective notability metrics may

be complex and require significant technical expertise.
Mitigation: Ensure the research team includes experts in statistical analysis, data science, and machine learning. Conduct thorough pilot studies to refine methodologies and address potential challenges early on.

3. Algorithmic Bias
Risk: The development and implementation of "bias-aware" algorithms may inadvertently introduce new biases or perpetuate existing ones.
Mitigation: Employ rigorous testing and validation procedures to identify and mitigate algorithmic biases. Involve diverse stakeholders in the design and evaluation of algorithms to ensure fairness and inclusivity.

4. Community Engagement
Risk: Resistance from the Wikipedia community to the proposed changes or recommendations based on the research findings.
Mitigation: Engage with the Wikipedia community throughout the research process through co-creation sessions, focus groups, and interviews. Communicate research findings transparently and address concerns proactively.

5. Ethical Considerations
Risk: The research may inadvertently reveal sensitive information about individuals or groups, raising ethical concerns related to privacy and representation.
Mitigation: Adhere to strict ethical guidelines and obtain informed consent from participants. Anonymize data and protect the privacy of individuals and groups mentioned in the research.

Community impact plan

The *COVER WOMEN II* project aims to generate a tangible and lasting impact within the Wikimedia community, reaching beyond academic circles. Through a set of concrete actions and open, participatory strategies, we intend to ensure that our findings and tools are meaningful, adoptable, and reusable by Wikimedia volunteers, developers, affiliates, and user groups.

1. Collaboration with Wikipedia Editor Communities

We will organize co-creation workshops and focus groups with editors from different language editions of Wikipedia. These sessions will serve to:

- Test the recommendation system based on objective notability scoring and gather feedback on usability and effectiveness.
- Foster dialogue on inclusion strategies and the visibility of underrepresented biographies on the front page and in deletion discussions.
- Tailor the tools and proposals to align with each community's norms, workflows, and cultural contexts.
- Collaboratively develop editorial policies that promote inclusive communication and diversity, grounded in both empirical findings and shared community values.

2. Dissemination through Wikimedia Events and Networks

We plan to present the project's findings and tools at Wikimedia events such as Wikimania, WikiWorkshop, and local user group meetups. These venues will:

- Support direct dialogue with active contributors and decision-makers.

- Promote early adoption of the tools and encourage feedback-driven improvement.
- Build bridges between researchers, technical contributors, and policy advocates.

3. Development of Open and Transparent Tools for the Technical Community

All tools developed during the project—including the recommendation system and the notability scoring model—will be fully open source, with public repositories and clear documentation to support transparency and reuse.

Additionally, the **criteria, design assumptions, and editorial policies used in the algorithmic model will be developed transparently**, allowing the community to **inspect, challenge, and iteratively improve them over time**. This ensures methodological openness and fosters long-term trust and adaptability.

These resources will:

- Enable volunteer developers to adapt or expand the tools across different Wikimedia use cases.
- Allow integration into existing community-led solutions such as bots, gadgets, or moderation workflows.

4. Creation of Educational and Editorial Support Resources

We will develop a **Communication and Diversity Toolkit** to support inclusive editorial practices. This toolkit will include:

- Best practices for equitable representation of biographies, with a focus on intersectional visibility.
- Editorial guidance for front-page inclusion and deletion discussions.
- Training materials that promote awareness of structural bias and foster inclusive community dialogue.

5. Strengthening Collaboration with Wikimedia Affiliates and User Groups

We will collaborate with Wikimedia affiliates and user groups to help adapt and implement the project outcomes in local contexts. This includes:

- Translating and localizing project materials.
- Organizing tailored training sessions and presentations.
- Supporting ongoing dialogue on equity-focused editorial governance.

Through these combined efforts, *COVER WOMEN II* not only seeks to analyze existing content selection mechanisms on Wikipedia, but also to co-develop practical, participatory, and evidence-based tools that empower the community to promote more equitable, inclusive, and visible representation across Wikimedia projects.

Evaluation

The success of *COVER WOMEN II* will be evaluated through a combination of qualitative and quantitative indicators that reflect the project's dual aim: to advance scholarly understanding of bias and representation in Wikipedia, and to co-develop practical tools that empower Wikimedia communities to promote inclusion and diversity.

1. Research Outputs and Academic Impact

- Completion and peer-reviewed publication of at least four academic papers aligned with the project's research questions.

- Presentation of research findings at major international conferences (e.g., Wikimania, WikiWorkshop).
- Citation and uptake of project outputs in scholarly and open knowledge literature.

2. Tool Development and Technical Deliverables

- Release of a bias-aware notability scoring model and an open-access dataset.
- Development and successful deployment of a functional prototype of the recommendation system.
- Public availability and documentation of all tools and datasets in open repositories, ensuring transparency and reproducibility.
- Adoption or adaptation of the tools by Wikimedia technical contributors or community projects.

3. Community Engagement and Capacity Building

- Number and diversity (by language, gender, region) of participants involved in co-creation sessions, focus groups, and testing phases.
- Incorporation of community feedback into iterative improvements of tools and editorial policies.
- Creation and dissemination of a Communication and Diversity Toolkit, with evidence of uptake by user groups or affiliates. This toolkit will include newly developed editorial guidelines and policy recommendations specifically designed to counteract the gender and intersectional biases identified during the front-page content analysis phase. These resources will support communities in making more inclusive, equitable decisions around content visibility and selection.

4. Cultural and Policy Impact

- Emergence or revision of editorial practices or guidelines in line with project recommendations (e.g., on notability or inclusive communication).
- Engagement from Wikimedia affiliates in applying the project's outputs in local or linguistic contexts.
- Positive reception and critical engagement by editors, as reflected in surveys, interviews, or online discussions.

5. Sustainability and Long-term Integration

- Inclusion of project tools or methods in Wikimedia workflows (e.g., bots, deletion discussions, featured content selection).
- Maintenance or further development of tools beyond the project's timeline by the community or interested affiliates.
- Potential for scalability or adaptation of the notability model to other Wikimedia content domains (e.g., media files, categories).

Overall Success Criteria Success will be measured not only by the completion of deliverables, but also by the degree to which the project contributes to a more equitable, transparent, and participatory content governance culture in Wikimedia. We will consider the project impactful if it leads to measurable improvements in representation, promotes ongoing community-led innovation, and strengthens the Wikimedia movement's capacity to address structural bias.

Budget

Budget details

References

AlGhamdi, Kholoud, Shi, Miaojing, & Simperl, Elena. (2021). Learning to Recommend Items to Wikidata Editors. In Andreas Hotho, Eva Blomqvist, Stefan Dietze, Achille Fokoue, Ying Ding, Payam Barnaghi, Armin Haller, Mauro Dragoni, & Harith Alani (Eds.), *The Semantic Web – ISWC 2021* (Vol. 12922, pp. 163–181). Springer International Publishing. https://doi.org/10.1007/978-3-030-88361-4_10

Alghamdi, Kholoud, Shi, Miaojing, & Simperl, Elena. (2022). *Exploring and Eliciting Needs and Preferences from Editors for Wikidata Recommendations* (arXiv:2212.01818). arXiv. https://doi.org/10.48550/arXiv.2212.01818

Antin, Judd, Yee, Raymond, Cheshire, Coye, & 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., & 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, & 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

Ferran-Ferrer, N., Miquel-Ribé, M., Meneses, J., & Minguillón, J. (2022). The Gender Perspective in Wikipedia: A Content and Participation Challenge. *WWW - Companion Proc. Web Conf.*, 1319–1323. Scopus. https://doi.org/10.1145/3487553.3524937

Ferran-Ferrer, Núria, Fernández, Laura, & 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, Pietsch, Tamson, & Tall, Kelly. (2023). Gender and the invisibility of care on Wikipedia. *Big Data & Society*, *10*(2), 20539517231210276. https://doi.org/10.1177/20539517231210276

Ford, Heather, & 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

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

Korsgaard, Thomas Rune, & Jensen, Christian D. (2009). Reengineering the Wikipedia for Reputation. *Electronic Notes in Theoretical Computer Science*, *244*, 81–94. https://doi.org/10.1016/j.entcs.2009.07.040

Laouenan, M., Bhargava, P., Eyméoud, J. B., Gergaud, O., Plique, G., & Wasmer, E. (2022). A cross-verified database of notable people,

3500BC-2018AD. *Scientific Data*, *9*(1). Scopus. https://doi.org/10.1038/s41597-022-01369-4

Lemieux, Mackenzie Emily, Zhang, Rebecca, & Tripodi, Francesca. (2023). "Too Soon" to count? How gender and race cloud notability considerations on Wikipedia. *Big Data & Society*, *10*(1), 20539517231165490. https://doi.org/10.1177/20539517231165490

Lin, Demiao, Jin, Jianming, & Xiong, Yuhong. (2010). *Wikipedia-based Online Celebrity Recognition*. https://shiftleft.com/mirrors/www.hpl.hp.com/techreports/2010/HPL-2010-2.pdf

Margolin, Drew B., Goodman, Sasha, Keegan, Brian, Lin, Yu-Ru, & Lazer, David. (2016). Wiki-worthy: Collective judgment of candidate notability. *Information, Communication & Society*, *19*(8), 1029–1045. https://doi.org/10.1080/1369118X.2015.1069871

McCombs, Maxwell E., & Shaw, Donald L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*(2), 176–187.

Menking, Amanda, & 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, & 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

Moskalenko, Oleksii, Parra, Denis, & Saez-Trumper, Diego. (2020). *Scalable Recommendation of Wikipedia Articles to Editors Using Representation Learning*. http://www.toinebogers.com/workshops/complexrec2020/Moskalenko.pdf

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

Schwarzer, Malte, Breitinger, Corinna, Schubotz, Moritz, Meuschke, Norman, & Gipp, Bela. (2017). Citolytics: A Link-based Recommender System for Wikipedia. *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 360–361. https://doi.org/10.1145/3109859.3109981

Schwarzer, Malte, Schubotz, Moritz, Meuschke, Norman, Breitinger, Corinna, Markl, Volker, & Gipp, Bela. (2016). Evaluating Link-based Recommendations for Wikipedia. *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, 191–200. https://doi.org/10.1145/2910896.2910908

Taraborelli, Dario, & Ciampaglia, Giovanni Luca. (2010). Beyond Notability. Collective Deliberation on Content Inclusion in Wikipedia. *2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop*, 122–125. https://doi.org/10.1109/SASOW.2010.26

Tripodi, Francesca. (2023). Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, *25*(7), 1687–1707. https://doi.org/10.1177/14614448211023772

Wagner, Claudia, Graells-Garrido, Eduardo, Garcia, David, & 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

Wulczyn, Ellery, West, Robert, Zia, Leila, & Leskovec, Jure. (2016). Growing Wikipedia

Across Languages via Recommendation. *Proceedings of the 25th International Conference on World Wide Web*, 975–985. https://doi.org/10.1145/2872427.2883077

Yuan, Y. Connie, Cosley, Dan, Welser, Howards T., Xia, Ling, & Gay, Geri. (2009). The Diffusion of a Task Recommendation System to Facilitate Contributions to an Online Community. *Journal of Computer-Mediated Communication*, *15*(1), 32–59. https://doi.org/10.1111/j.1083-6101.2009.01491.x

Zhang, Charles Chuankai, & Terveen, Loren. (2021). Quantifying the Gap: A Case Study of Wikidata Gender Disparities. *17th International Symposium on Open Collaboration*, 1–12. https://doi.org/10.1145/3479986.3479992

Zhang, Haisu, Zhang, Sheng, Wu, Zhaolin, Huang, Liwei, & Ma, Yutao. (2014). Predicting Wikipedia editor's editing interest based on factor graph model. *2014 IEEE International Congress on Big Data*, 382–389. https://ieeexplore.ieee.org/abstract/document/6906806/

Annex

Cover Women II Time line