"Too Soon" To Count? The impact of gender and race on perceived notability

Mackenzie Lemieux*, Rebecca Zhang, Francesca Tripodi
Washington University School of Medicine in St. Louis, University of North Carolina at Chapel Hill
*Corresponding Author

Abstract

Gender and racial bias lead to inequality and lack of diversity in academia. English-language Wikipedia, one of the most frequently accessed websites on the internet, perpetuates this inequality with an abysmally low proportion of biographies dedicated to women and people of color. While research has explored the extent of gender bias and the barriers to women's inclusion on Wikipedia, very little research has focused on the problem of racial bias on the encyclopedia. While advocacy groups try to independently improve representation on Wikipedia, nothing has been done from an institutional level to address the pervasive bias. Applying a combination of web-scraping, deep learning, natural language processing, and content analysis of academic biographies nominated for deletion, we demonstrate how Wikipedia's notability criteria are unequally applied across race and gender. We find that online presence is an important predictor of notability for white male academics but not for female and non-white academics. A deeper dive into the mechanisms behind this inequality highlights a subjective application of the Wikipedia shortcut, WP:TooSoon, as well as the importance of "weak ties."

Extended Abstract

Katie Bouman is an American computer scientist and engineer who developed parts of the code used to capture the first image of a black hole (Elfrink, 2019). Shortly thereafter, a Wikipedia page was created about Bouman, but was just as quickly it flagged for deletion citing that since Bouman was not even an assistant professor she certainly was not "notable as a scientist." Despite extensive third-party coverage of her discovery around the globe (Harrison, 2019, Mervosh, 2019), other editors agreed claiming "WP:TooSoon applies." "WP: TooSoon" is a technical label used to indicate that the subject of a Wikipedia biography lacks sufficient coverage in independent, high-quality news sources (WP:Too soon). However, Bouman *had* received extensive media coverage in the New York Times, CNN, The Washington Post, The Atlantic, and more (Mervosh, 2019, Lou and Ahmed, 2019, Bever, 2019, Koren, 2019, Maguire, 2019, Boyer, 2019). Our analysis reveals that Bouman's qualifications were being merited against her perceived professional accomplishments, not Wikipedia's criteria for inclusion. This apparent discrepancy led us to ask whether editorial evaluation of academic biographies on Wikipedia is truly rooted in Wikipedia guidelines (online coverage), or whether other factors, such as perceived career stage, race, and gender, are the primary tools of assessment.

Gender inequality on Wikipedia has been well documented. Women represent less than twenty percent of all English-language biographies (WP:WikiProject Women in Red; Tripodi 2021). Female academics are less likely to be recognized on Wikipedia than their male counterparts across all fields of study (Lou et. al YEAR; Schellekens et al, 2019; Vitulli, 2017). Women's pages display a negative linguistic bias, and the hyperlink structure of women's pages puts them at a disadvantage in terms of page visibility and traffic compared to men's (Wagner et. al, 2016).

To try and improve gender representation on the site, Wikipedians and Wikimedia have created advocacy groups, like Women in Red and Art+Feminism. Wikipedia editing events aim to inspire new writers to create more biographies about women and minorities. Despite success, this effort has been stymied by targeted deletions of biographies of women and people of color (Kramer, 2019). These deletions emphasize the need to continue to improve the depth of research on Wikipedia's gender bias as well as its' understudied racial bias (Smith, 2015). While recent work has uncovered that women are more likely to be nominated for deletion (Tripodi, 2021), we

still lack a concrete understanding of how the Wikipedia rules that structure notability are applied (Gauthier and Sawchuk, 2017). More research is needed to understand how to fix the glaring inequalities on Wikipedia with respect to both gender and race (Boboltz, 2017).

To address this issue, we explore how notability criteria on Wikipedia are applied across gender and race for biographies of academics. To do so, we web-scraped the articles for deletion archive (AfD) to compile biographies of academics nominated for deletion from 2017 to 2020 (n=843). Next, we determined each biography subject's online presence score by the extent of their online news coverage using natural language processing software. This online presence score was used as an objective and quantitative surrogate for Wikipedia's existing notability criteria (e.g., WP: Notability; WP: Search Engine Test). Finally, we compared the online presence score for biographies that were deleted from Wikipedia versus those that remained published.

While online presence serves as a meaningful metric for predicting page stability for white male academics, we find that notability criteria are implemented idiosyncratically for women and non-white academics (Figure 1.). Moreover, we find that the label "Too Soon" is more frequently applied to women and is inconsistent with Wikipedia guidelines for how "Too Soon" should be applied. This suggests that notability criteria are implemented subjectively, allowing for implicit biases to drive inclusion decisions. Our work emphasizes the need for advocacy networks, like Women in Red, to monitor pages miscategorized as non-notable.
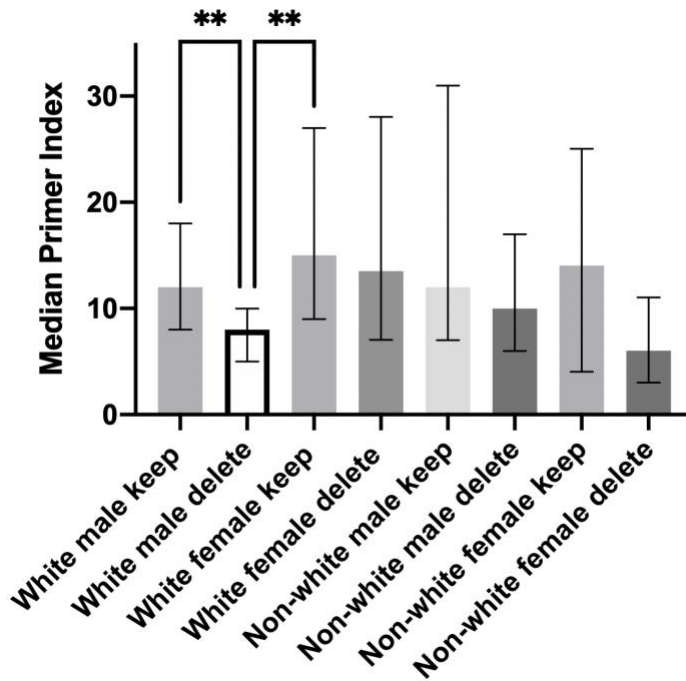
Figure 1. A bar chart comparing median Primer Indices across race and gender, showing median plus 95% confidence interval.

Citations

Adams J and Brückner H (2015) Wikipedia, sociology, and the promise and pitfalls of Big Data. Big Data & Society 2;2.

Adams J, Brückner H and Naslund C (2019) Who counts as a notable sociologist on Wikipedia? Gender, Race, and the "Professor Test". Socius 5: 1–14.

Altheide D (2000). Tracking discourse and qualitative document analysis. Poetics 27:4, 287-299

Ambekar et al. (2009) Name-ethnicity classification from open sources. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining 40-49.

Berger M (2020) Women are systematically excluded from global coronavirus coverage, experts say. The Washington Post. Available from: https://www.washingtonpost.com/world/2020/10/08/women-experts-excluded-global-coronavirus-coverage/ (accessed 5 November 2021).

Berterolo et al. (2020). Racial and ethnic imbalance in neuroscience reference lists and intersections with gender. bioRxiv. Available at: https://en.wikipedia.org/wiki/Help:Find_sources (accessed 1 June 2020).

Bever L (2019) Katie Bouman helped the world see a black hole. fans want 'a rightful seat in history' for her. The Washington Post, WP Company. Available from: https://www.washingtonpost.com/science/2019/04/11/katie-bouman-helped-world-see-black-hole-fans-want-rightful-seat-history-her/ (accessed 12 December 2021).

Boboltz S (2017) Editors are trying to fix Wikipedia's gender and racial bias problem. HuffPost, HuffPost. Available from: https://www.huffpost.com/entry/wikipedia-gender-racial-bias_n_7054550 (accessed 12 December 2021).

Boyer M (2019) Katie Bouman's actual contributions to the historic Black Hole Image. Telegraph India, Telegraph India. Available from: https://www.telegraphindia.com/opinion/katie-bouman-s-actual-contributions-to-the-historic-black-hole-image/cid/1688787 (accessed 12 December 2021).

Callahan ES and Herring SC (2011) Cultural bias in Wikipedia content on famous persons. J. Am. Soc. Inf. Sci., 62: 1899-1915.

Center for the Digital Future at USC Annenberg (2017) Surveying the Digital Future: The 15th annual study of the impact of digital technology on Americans. Available from: http://www.digitalcenter.org/wp-content/uploads/2013/10/2017-Digital-Future-Report.pdf (accessed 15 January 2022)

Crenshaw K (1989) Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. University of Chicago Legal Forum 1:8.

Du Bois WEB (1940). Dusk of dawn: An essay toward an autobiography of a race concept. Print.

Elfrink T (2019) Trolls hijacked a scientist's image to attack Katie Bouman. they picked the wrong astrophysicist. The Washington Post, WP Company. Available from: https://www.washingtonpost.com/nation/2019/04/12/trolls-hijacked-scientists-image-attack-katie-bouman-they-picked-wrong-astrophysicist/ (accessed 12 December 2021).

Ezell JM (2021) Empathy plasticity: decolonizing and reorganizing Wikipedia and other online spaces to address racial equity. Ethnic and Racial Studies 44:8, 1324-1336.

Gauthier M and Sawchuk K (2017) Not notable enough: feminism and expertise in Wikipedia. Communication and Critical/Cultural Studies, 14:4, 385-402.

Gibbs K (2014) Diversity in STEM: What it is and why it matters. Scientific American Blog Network, Scientific American. Available from: https://blogs.scientificamerican.com/voices/diversity-in-stem-what-it-is-and-why-it-matters/ (accessed 12 December 2021).

Gibson EM (2020) How Support of Early Career Researchers Can Reset Science in the Post-COVID19 World. Cell 181(7):1445-1449.

Granovetter MS (1973) The Strength of Weak Ties. The American Journal of Sociology 78:6, 1360-1380.

Harrison S (2019) How Katie Bouman shook wikipedia. Slate Magazine, Slate. Available from: https://slate.com/technology/2019/04/katie-bouman-wikipedia-page-deletion-black-hole.html (accessed 12 December 2021).

Harrison S (2019) How the sexism of the past reinforces Wikipedia's gender gap. Slate Magazine, Slate. Available from: https://slate.com/technology/2019/03/wikipedia-women-history-notability-gender-gap.html (accessed 12 December 2021).

Hill Collins, Patricia. (1990). Black feminist thought : knowledge, consciousness, and the politics of empowerment. Boston :Unwin Hyman.

Humaniki: Wikimedia Diversity Dashboard tool (n.d.) Humaniki | Wikimedia Diversity Dashboard Tool. Available from: https://whgi.wmflabs.org/ (accessed 12 December 2021).

Kassova L (2020) The Missing Perspectives of Women in COVID-19 News. The Bill & Melinda Gates Foundation. Available from: https://www.iwmf.org/wp-content/uploads/2020/09/2020.09.16-FULL-COVID-REPORT.pdf (accessed 15 October 2021)

Khurana R (2021) How to make artificial intelligence more democratic. Scientific American, Scientific American. Available from: https://www.scientificamerican.com/article/how-to-make-artificial-intelligence-more-democratic/ (accessed 12 December 2021).

Koren M (2019) The Dark Saga of Katie Bouman. The Atlantic, Atlantic Media Company. Available from: https://www.theatlantic.com/science/archive/2019/04/katie-bouman-black-hole/587137/ (accessed 12 December 2021).

Kramer K (2019) Female scientists' pages keep disappearing from Wikipedia – what's going on? Chemistry World, Chemistry World. Available from: https://www.chemistryworld.com/news/female-scientists-pages-keep-disappearing-from-wikipedia-whats-going-on/3010664.article (accessed 12 December 2021).

Lam SK, Uduwage A, Dong Z, et al. (2011) WP:Clubhouse?: an exploration of Wikipedia's gender imbalance. Proceedings of the 7th International Symposium on Wikis and Open Collaboration, 1-10.

Lou M and Ahmed S (2019) That image of a black hole you saw everywhere? thank this grad student for making it possible. CNN, Cable News Network. Available from: https://www.cnn.com/2019/04/10/us/katie-bouman-mit-black-hole-algorithm-sci-trnd/index.html (accessed 12 December 2021).

Lucky S and Rubin J (2017) Cultivating Your Academic Online Presence. Evidence Based Library and Information Practice 12.2.

Luo W, Adams J and Brückner H (2018) The Ladies Vanish? Comparative Sociology 17(5): 519–556.

Luyt B (2012) The inclusivity of Wikipedia and the drawing of expert boundaries: an examination of talk pages and reference lists. Journal of the American Society for Information Science and Technology 63(9): 1868–1878.

Luyt B and Tan D (2010) Improving Wikipedia's credibility: reference and citations in a sample of history articles. Journal of the American Society for Information Science and Technology 61(4): 715–722.

Maguire D (2019) Trolls target the scientist who became the face of the Black Hole Photo. ABC News, ABC News. Available from: https://www.abc.net.au/news/2019-04-15/black-hole-photo-katie-bouman-trolls/11006820 (accessed 12 December 2021).

Matei SA and Dobrescu C (2011) Wikipedia's "Neutral Point of View": Settling Conflict through Ambiguity. The Information Society 27(1): 40–51.

McDonough M (2021) Women are writing themselves back into history on Wikipedia. LX, LX. Available from: https://www.lx.com/social-justice/women-are-writing-themselves-back-into-history-on-wikipedia/34029/ (accessed 12 December 2021).

Mervosh S (2019) How Katie Bouman accidentally became the face of the Black Hole Project. The New York Times, The New York Times. Available from: https://www.nytimes.com/2019/04/11/science/katie-bouman-black-hole.html (accessed 12 December 2021).

Mostaghimi A and Crotty BH (2011) Professionalism in the Digital Age. Annals of Internal Medicine 154:560-562.

Press AL and Tripodi F (2021) Media-ready Feminism and Everyday Sexism. Albany, NY: State University of New York.

Primer. Available from: https://primer.ai (accessed 6 June 2021).

Rainone C (2020) 'The Humanity of Blackness' Missing From History Classes: How to Transform Black History Education in Schools. NBC Philadelphia. Available from: https://www.nbcphiladelphia.com/news/national-international/how-to-transform-black-history-education-in-schools/2450465/ (accessed 6 June 2021).

Ray V (2019) A theory of racialized organizations. American Sociological Review 84:1, 26-53.

Robitzski D (2017) To keep AI from 'eating a table,' scientists make it read Wikipedia. LiveScience, Purch. Available from: https://www.livescience.com/60520-artificial-intelligence-reads-wikipedia-to-learn-english.html (accessed 12 December 2021).

Sadeque S (2021) Wikipedia doesn't think this black female scientist is notable enough for a page. The Daily Dot. Available from: https://www.dailydot.com/irl/wikipedia-clarice-phelps/ (accessed 12 December 2021).

Salam M (2019) Most Wikipedia profiles are of men. this scientist is changing that. The New York Times, The New York Times. Available from: https://www.nytimes.com/2019/07/19/science/jess-wade-wikipedia-science-women.html (accessed 12 December 2021).

Samoilenko A and Yasseri T (2014) The distorted mirror of wikipedia: a quantitative analysis of wikipedia coverage of academics. EPJ Data Science 3(11): 1.

Schellekens MH, Holstege F and Yasseri T (2019) Female scholars need to achieve more for equal public recognition. Available at: https://arxiv.org/abs/1904.06310

Schneider J, Passant A and Decker S (2012) Deletion Discussions in Wikipedia: Decision Factors and Outcomes. Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration 17: 1-10.

Selvarajah M (2021) Canadian nobel scientist's deletion from Wikipedia points to wider bias, study finds . CBCnews, CBC/Radio Canada. Available from: https://www.cbc.ca/news/science/wikipedia-bias-1.6129073 (accessed 12 December 2021).

Sims JP, Pirtle LP and Johnson-Arnold I (2020) Doing hair, doing race: the influence of hairstyle on racial perception across the US. Ethnic and Racial Studies, 43:12, 2099-2119.

Smith JF (2015) Howard University fills in Wikipedia's gaps in black history. The New York Times, The New York Times. Available from: https://www.nytimes.com/2015/02/20/us/at-howard-a-historically-black-university-filling-in-wikipedias-gaps-in-color.html (accessed 12 December 2021).

The NCES Fast Facts Tool provides quick answers to many education questions (2018) National Center for Education Statistics (NCES) Home Page, a part of the U.S. Department of Education, National Center for Education Statistics. Available from: https://nces.ed.gov/fastfacts/display.asp?id=61 (accessed 12 December 2021).

Torres N (2016) Why do so few women edit wikipedia? Harvard Business Review. Available from: https://hbr.org/2016/06/why-do-so-few-women-edit-wikipedia (accessed 12 December 2021).

Tripodi F (2021) Ms. Categorized: Gender, notability, and inequality on Wikipedia. New Media & Society. doi: 10.1177/14614448211023772.

Vincent J (2018) AI spots 40,000 prominent scientists overlooked by Wikipedia. The Verge, The Verge. Available from: https://www.theverge.com/2018/8/8/17663544/ai-scientists-wikipedia-primer (accessed 12 December 2021).

Vitulli MA (2017) Writing women in mathematics into Wikipedia. Available at: https://arxiv.org/abs/1710.11103

Wade J and Zaringhalam M (2018) Why we're editing women scientists onto Wikipedia. Nature Careers.

Wagner C, Graells-Garrido E, Garcia D, et al. (2016) Women through the glass ceiling: gender asymmetries in Wikipedia. EPJ Data Science Journal 5; 5.

White A (2018) The history of women in engineering on Wikipedia. Science Museum Journal Group Autumn 2018, 10.

Wikimedia Foundation (2018) Gender equity report 2018/Barrier to equity. Available from: https://meta.wikimedia.org/wiki/Gender_equity_report_2018/Barriers_to_equity (accessed 10 December 2022).

Wikipedia: Notability

Wikipedia:Articles for Deletion. Available from: https://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion (accessed 15 January 2022).

Wikipedia:Help:Find sources. Available from: https://en.wikipedia.org/wiki/Help:Find_sources (accessed 22 November 2021).

Wikipedia:Notability Available. from: https://en.wikipedia.org/wiki/Wikipedia:Notability (accessed 12 December 2021).

Wikipedia:Search Engine Test. Available from: https://en.wikipedia.org/wiki/Wikipedia:Search_engine_test (accessed 12 December 2021).

Wikipedia:Systemic bias. Available from: https://en.wikipedia.org/wiki/Wikipedia:Systemic_bias (accessed 12 December 2021).

Wikipedia:Too soon. Available from: https://en.wikipedia.org/wiki/Wikipedia:Too_soon (accessed 12 December 2021).

Wikipedia:Wikiproject women in red. Available from: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red (accessed 12 December 2021).