FEMINIST MEDIA STUDIES
https://doi.org/10.1080/14680777.2023.2263659

The invisible women: uncovering gender bias in AI-generated images of professionals

Anna M. Gorska[a] and Dariusz Jemielniak[b]

[a]Human Resources Department, Kozminski University, Warsaw, Poland; [b]Management in Networked and Digital Environments (MINDS) Department, Kozminski University, Warszawa, Poland

ABSTRACT

This study explores gender bias in AI-generated images of professionals, focusing on the visual representation of male and female professionals in law, medicine, engineering, and scientific research. Using a sample of 99 images from nine popular text-to-image generators, we conducted a survey of 120 respondents who assessed the perceived gender of the images. Our findings reveal a significant gender bias, with men represented in 76% of the images and women in only 8%. This bias persists across all four professions and varies between different AI image generators. The results highlight the potential of AI to perpetuate and reinforce gender inequalities, suggesting the need for more intersectional and inclusive approaches in AI design and research. It further underscores the necessity of diversifying the design process and redistributing power in decision-making procedures to challenge existing biases in AI. Our study emphasizes the need for further action to address gender bias in AI-generated images and highlights the importance of adopting a more intersectional and inclusive approach in future research, considering factors such as race, class, and ability. This commentary aims to raise awareness of the current issues with AI-text to image generators and encourages the development of more inclusive and equitable AI technologies.

KEYWORDS

Gender bias; artificial intelligence bias; AI generated images; artificial intelligence; professional representation

Introduction

The interplay between technical aspects, like algorithms, and social constructs, such as gender, significantly impacts the perception and function of modern technology, particularly in Artificial Intelligence (AI) (Joy Buolamwini and Timnit Gebru 2018; Yunghe Feng and Chirag Shah 2022; Ariane Hanemaayer 2020). Despite initial hopes that AI could overcome human limitations, including gender prejudices, research indicates that AI systems are prone to biases (Naomi Ellemers 2018; Feng and Shah 2022; Safiya Umola Noble 2018). Given that AI systems are a product of human creation, they inherently embody the limitations and constraints characteristic of their developers.

Noble (2018) argues that algorithm creators import a variety of values onto AI, many of which inadvertently support sexism, racism, and uncorroborated meritocratic beliefs. The predominantly white and Asian male demographic in technology centers such as Silicon

CONTACT Anna M. Gorska amg@kozminski.edu.pl Human Resources Department, Kozminski University, Jagiellonska 57/59, Warsaw, Poland

Valley often dismisses issues of gender, race, class, and sexual identity as archaic, leading to denial of algorithmic sexism and racism (Mar Hicks 2021; Safiya Umola Noble and Sarah Roberts 2019). This mindset obstructs interventions aimed at dismantling discriminatory barriers and inadvertently perpetuates established prejudiced practices (Noble and Roberts 2019). As AI's societal impact grows, addressing these biases becomes critical to prevent the perpetuation of social inequality and exclusion.

AI's advancements have led to the rise of image generation tools that, through simple, no-code, and natural language prompts produce art-award-winning pieces. These generators, increasingly used in digital communication, can inadvertently perpetuate stereotypes, reinforcing inequalities, and even amplifying discrimination (Lizhen Liang and Daniel Acuna 2020; Escudero Pérez Jimena 2020). Given their potential to influence perceptions of reality and social phenomena, such as gender roles, it's critical to scrutinize their embedded biases. These generators are rapidly becoming popular topics of academic study. However, no study so far has investigated gender bias in AI image generators in the context of the workplace.

This commentary explores gender bias in AI-generated images of professionals in the workplace. We compare nine popular generators and examine their depiction of four prestigious occupations—law, medicine, engineering, and scientific research—to study the gender representation in these AI-created images.

AI gender inequalities

The common belief that algorithms and AI are inherently objective is misguided. Technology has always embodied the biases and stereotypes of its creators and users (Buolamwini and Gebru 2018; Hicks 2021). Hicks (2021) argued that those who control these systems do so at the expense of marginalized groups. In AI, this issue is exacerbated by the lack of clear regulations, guidelines, and the opaque nature of the results (Noble 2018). The data used to train algorithms may underrepresent certain groups, and system designers may inadvertently embed their assumptions into their creations (Noble and Roberts 2019), while these can frame societies' perception of an entire group or community (Noble 2018). As a result, as Judy Wajcman (2000) and Ariene Hanemaayer (2020) argue, technology and AI can not only reproduce inequalities but also reinforce existing power structures.

IT and new media often reflect masculine perspectives, which leads to biased platforms and products (Trang Thi Quynh Dinh and Janne Tienari 2022). This results in further inequality and gender segregation in education and professions (Estrella Gomez-Herrera and Sabine Köszegi 2022), which further reproduces inequalities.

Although the gendered division of labor and the entanglement of work and gender are not new, AI systems add a major multiplier. Since a given profession is important in career choices, and the feeling of inadequacy and not matching the dominant image of a professional may outweigh personal interests and skill match, it is important that gendered representations of occupations refute rather than reinforce biases and stereotypes.

The images of feminine and masculine occupations are reflected in biases observed in AI (Liang and Acuna 2020). For instance, an analysis of Google's image search algorithm found that women were significantly underrepresented in

images of jobs such as CEO, programmer, and biologist, and overrepresented in lower-paying positions such as nurse and primary school teacher (Feng and Shah 2022).

The underrepresentation of women in certain occupations in images used by AI systems reinforces the gender pay gap and the underrepresentation of women in positions of power and prestige. Moreover, such biases are particularly harmful for women who are already underrepresented in certain fields, as they reinforce stereotypes and discourage young women from pursuing these careers (Sylvia Walby 2011). A gendered nature of AI and socio-technological systems perpetuates intersectional inequalities and contributes to advancing algorithmic control and gender imbalance (D'Ignazio and Klein 2020; Noble 2018).

Method

To explore gender bias in AI-generated images, we studied the most prestigious professions per the 2018 Global Teacher Status Index. We analyzed 99 images of professionals (doctors, lawyers, engineers, scientists) from nine popular and accessible AI text-to-image tools, aiming to ascertain gendered patterns in these AI representations.

The images were then analyzed using a survey of 120 respondents, who were asked to evaluate the perceived sex of the entity portrayed in each image. The survey respondents were a diverse group, with 32% identifying as women and the majority (73%) being between the ages of 18 and 24. The survey consisted of 101 questions, 99 of which were designed to distinguish the gender of the AI-generated image. Respondents were given four options to describe the image: 1) male, 2) female, 3) ambiguous, and 4) inhuman. They could choose only one option, and the occupation of the AI-generated image was not specified.

Results

The analysis revealed a significant gender bias in AI-generated images of professionals, with men being represented in 76% of the images and women in only 8%. The smallest representation of women was among doctors (7%) even though in reality women constitute almost a half of doctors (OECD 2021).

The results showed that the gender bias was present across all four professions examined in the study, with men being overrepresented in each profession. When the "inhuman" and "ambiguous" answer choices were excluded from the analysis, the gender bias became even more pronounced, with 91% of all images being identified as male.

In terms of differences between the AI image generators, there was significant variation in their portrayal of gender. While seven of the nine generators primarily generated images that were perceived as human, Hypnogram and StarryAI were outliers, with their representations of professions often being perceived as inhuman and/or ambiguous. Interestingly, the only AI generator that presented more women than men was HotPotAI, with women represented in 67% of the generated images. However, in the majority of the other generators, men constituted over 75% of all AI-generated images when only responses evaluating images as male, or female were considered.

Discussion

This study's findings contribute to the growing literature on gender bias in AI technologies, offering valuable insights into how AI-generated images of professionals can perpetuate and reinforce gender inequalities. Our results indicate that gender bias is a widespread issue in AI-generated images of professionals, with men being overrepresented and certain professions and AI image generators being especially susceptible to this bias. This bias significantly impacts how women and men are portrayed across professions, leading to the normalization of stereotypical views and reinforcing existing gender inequalities.

Our findings highlight the need for a more intersectional and inclusive approach in future research on AI-generated images, considering how gender intersects with other power and oppression dimensions, such as race, class, and ability. This approach would help ensure that marginalized groups' experiences, including women of color, transgender individuals, and people with disabilities, are adequately addressed within the context of feminist technoscience studies.

Echoing previous research on AI bias in search engines (Feng and Shah 2022; Ellemers 2018; Noble 2018), our findings suggest that further action is required to address gender bias in AI-generated images. Although some tech companies have attempted to include more diversity in their image search results (Feng and Shah 2022), it is evident that more needs to be done.

However, tech companies, driven by commercial interests, may not always be incentivized to rectify their algorithms towards more equitable outcomes. The commercial interests of tech companies often hinder their motivation to enhance algorithmic equity. The technological architecture, which mirrors and magnifies societal biases, poses significant challenges for marginalized groups.

It should be noted that increasing "representation" or "inclusion" is not a solution, as these approaches may become superficial or tokenistic (Sasha Costanza-Chock 2020; Noble 2018), while as Costanza-Chock (2020) argues we should strive for diversification in the design process itself and create "design justice" to be driven by marginalized communicates. As C O'Neil (2016) wrote "blind spots reflect the judgments and priorities of its creators," thus, the change in AI should not just be about representation and inclusion but also about the redistribution of power in design and decision-making processes. As our reliance on AI tools grows, we must prioritize the reimagining of information design, access, and knowledge generation towards a more equitable paradigm.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Anna M. Górska is Assistant Professor in Human Resources Department and Director of Research Center Women and Diversity in Organizations, Kozminski University. Her works have been published in international academic journals such as *Gender, Work and Organization, Gender in Management, Information & Management, Gender, Work and Organization, Feminist Media Studies, Gender in Management*. Her recent book "Gender and Academic Career Development in Central and Eastern Europe" has been published by Routledge. She has participated as a principal investigator and

researcher in multiple research grants funded by Polish National Center of Science, EU funds and Norwegian Funds. She has also been awarded scholarships and awards, including Fulbright Junior Research Award. She has gained her international experience as a visiting scholar at Columbia Business School and ESCP Business School.

Dariusz Jemielniak is Full Professor and head of Management in Networked and Digital Environments (MINDS) department, Kozminski University, and faculty associate at Berkman-Klein Center for Internet and Society, Harvard University. He is a corresponding member of the Polish Academy of Sciences (elected as the youngest in history in social sciences and humanities). His recent books include Collaborative Society (2020, MIT Press, with A. Przegalinska), Thick Big Data (2020, Oxford University Press), Common Knowledge? An Ethnography of Wikipedia (2014, Stanford University Press). His current research projects include disinformation, and bot detection. He currently serves on the Wikimedia Foundation Board of Trustees, as well as a vice-chair for Polish Academy of Sciences.

References

Buolamwini, Joy, and Timnit. Gebru 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR, New York City, NY, 81: 77–91.

Costanza-Chock, Sasha. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge: The MIT Press. doi:10.7551/mitpress/12255.001.0001.

D'Ignazio, C., and L. F. Klein. 2020. *Data Feminism*. Cambridge: The MIT Press.

Dinh, Trang Thi Quynh, and Janne. Tienari. 2022. "Brothers and Broken Dreams: Men, Masculinity, and Emotions in Platform Capitalism." *Gender, Work, and Organization* 29 (2): 609–625. doi:10.1111/gwao.12786.

Ellemers, Naomi. 2018. "Gender Stereotypes." *Annual Review of Psychology* 69 (1): 275–298. doi:10.1146/annurev-psych-122216-011719.

Feng, Yunghe, and Chirag. Shah 2022. "Has CEO Gender Bias Really Been Fixed? Adversarial Attacking and Improving Gender Fairness in Image Search." *Proceedings of the AAAI Conference on Artificial Intelligence*, Washington, DC, 36(11): 11882–11890.

Global Teacher Status Index. 2018. Retrieved online from: https://www.varkeyfoundation.org/media/4790/gts-index-9-11-2018.pdf.

Gomez-Herrera, Estrella, and Sabine. Köszegi 2022. *A Gender Perspective on Artificial Intelligence and Jobs: The Vicious Cycle of Digital Inequality*. Working Paper, Bruegel.

Hanemaayer, Ariane. 2020. *Artificial Intelligence and Its Discontents: Critiques from the Social Sciences and Humanities*. London: Palgrave Macmillan.

Hicks, Mar. 2021. "Sexism is a Feature, Not a Bug." In *Your Computer is on Fire*, edited by Thomas Mullaney, Benjamin Peters, Mar Hicks, and Kavita Philip, 135–158. Cambridge: The MIT Press. doi:10.7551/mitpress/10993.003.0011.

Jimena, Escudero Pérez. 2020. "'an AI Doesn't Need a gender' (But It's Still Assigned One): Paradigm Shift of the Artificially Created Woman in Film." *Feminist Media Studies* 20 (3): 393–408. doi:10.1080/14680777.2019.1615973.

Liang, Lizhen, and Daniel. Acuna 2020. "Artificial Mental Phenomena: Psychophysics as a Framework to Detect Perception Biases in AI Models." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 403–412). New York, NY, USA: Association for Computing Machinery.

Noble, Safiya Umola. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. 1st ed. New York: NYU Press. doi:10.2307/j.ctt1pwt9w5.

Noble, Safiya Umola., and Sarah. Roberts 2019. "Technological Elites, the Meritocracy, and Postracial Myths in Silicon Valley." In *RACISM POSTRACE*. UCLA. Report #: 6. Retrieved from https://escholarship.org/uc/item/7z3629nh .

OECD. 2021. *Health at a Glance 2021: OECD Indicators, Doctors (By Age, Sex and Category*. https://www.oecd.org/gender/data/the-proportion-of-female-doctors-has-increased-in-alloecd-countries-over-the-past-two-decades.htm .

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 1st ed. New York: Crown.

Wajcman, Judy. 2000. "Reflections on Gender and Technology Studies.: In What State is the Art?" *Social Studies of Science* 30 (3): 447–464. doi:10.1177/030631200030003005.

Walby, Sylvia. 2011. "Is the Knowledge Society Gendered?" *Gender, Work and Organization* 18 (1): 1–29. doi:10.1111/j.1468-0432.2010.00532.x.