cogitatio

EDITORIAL

Artificial Intelligence and Ethnic, Religious, and Gender-Based Discrimination

Derya Ozkul

Department of Sociology, University of Warwick, UK

Correspondence: Derya Ozkul (derya.ozkul@warwick.ac.uk)

Abstract

This thematic issue explores the applications of artificial intelligence-based technologies and their potential for producing discriminatory and biased outcomes based on ethnicity, religion, and gender. This thematic issue adds to the ongoing debate with theoretical and empirical studies and a commentary that examine the topic from various perspectives. This editorial discusses the key themes highlighted in the studies and presents the findings of the different contributions to this collection.

Keywords

algorithms; artificial intelligence; automated decision-making systems; bias; discrimination

---

Numerous studies have shown that artificial intelligence (AI) technologies can produce biased outcomes due to their design and the existing inequalities in society (Benjamin, 2019; Broussard, 2023; Joyce et al., 2021). For example, hiring algorithms have been found to discriminate against women and individuals with minority names. Gender and racial discrimination have also been identified in internet search platforms (Noble, 2018), targeted ads and social media posts, credit scoring, insurance provision (O'Neil, 2016), and facial recognition technologies, especially in their accuracy with black female faces (Buolamwini & Gebru, 2018). Biased credit scoring and insurance provisions could worsen inequality, while biased algorithms in social media for voter targeting could threaten democracy (O'Neil, 2016). The lack of transparency makes it difficult for outsiders to understand how the systems were designed and by whom, highlighting the need for "data feminism" to analyse inherent inequalities leading to biases (D'Ignazio & Klein, 2020). Additionally, the lack of transparency in how these automated systems function creates challenges for users in identifying and contesting biased outcomes.

This thematic issue builds another brick into the debate, with studies exploring it from different angles. In the first article, author Carsten Orwat analyses the risk of discrimination in relation to broader threats to

---

fundamental rights and points out the shortcomings of current anti-discrimination laws in addressing algorithmic discrimination (Orwat, 2024). The use of AI and automated decision-making systems (ADMs), he argues, present challenges for data subjects who may find it difficult to recognise unfair treatment and gather enough evidence to challenge discriminatory practices caused by algorithms. As a result, Orwat contends, data subjects may face significant obstacles in taking legal action. He further explores how discrimination is linked to potential violations of human dignity.

Orwat conceptualises human dignity not as monolithic, but as multifaceted. Looking at the decisions of the German Federal Constitutional Court and developments of constitutional law in Germany, he shows that the concept of human dignity is already concretised in law, suggesting that the human being cannot be treated like an object and has the right to equal treatment and the free development of personality. As his analysis of past court rulings shows, the Federal Constitutional Court has concretised its interpretation of the right to human dignity and personality in relation to the risks posed by technologies. He then proceeds with analysing violations of human dignity concerning specific factors, namely severe and structural discrimination that has an impact on groups of individuals that are not regarded and treated with equal moral worth; lack of understanding around differentiating criteria from data subjects' perspectives; data subjects being treated as objects instead of individuals with respect and as whole persons; externalisation of identity determination and thereby elimination of self-determination, as well as the absence of meaningful informed consent in these automated processes. His article emphasises the importance of treating individuals with respect in complex human and machine interactions.

Larsson et al. (2024) present a related yet empirical case study in their article, exploring discrimination in the employers' use of AI and ADMs. While previous research has extensively examined potential biases arising from the use of AI in recruitment, this study shifts its focus to the perceptions of recruiters regarding AI and ADMs. Larsson and colleagues conducted an analysis of how recruiters comprehend AI and ADMs based on data they collected from prominent recruitment agencies and employers in Sweden. Their findings indicate a general lack of awareness among recruiters regarding the everyday integration of AI and ADM in their workplace. This lack of awareness may stem from a failure to recognise numerous common technological applications, such as searching for information on Google and encountering advertisements, in connection to algorithms and automated recommendation systems. It may also be possible that participants in the study may perhaps perceive AI as a relatively novel concept. However, the study's results demonstrate that as AI and ADMs progressively pervade everyday life, they become normalised and even disregarded by their users. This phenomenon is notably pronounced in cases where the use of these technologies is indirect and not perceived as central to the recruiters' responsibilities despite being extensively used.

The article by Inga Ulnicane explores another specific case study, this time focusing on how debates around AI frame emerging problems from the use of AI (Ulnicane, 2024). The author focuses on analysing the findings of four specific reports that she considers to be high-profile by using a framing approach and highlights their consistent emphasis on the dearth of diversity in the AI workforce, instances of AI-induced discrimination, and the inadequacies of nascent AI policies and guidelines. In addition to identifying common themes in these reports, Ulnicane also underscores their contextual specificity, raising questions about who has the power to identify problems and formulate recommendations within this domain. Indeed, in line with prevailing practices, leading reports in any given field often exhibit ties to the geopolitical powers underwriting their production.

Finally, in a short commentary, Pramod K. Nayar discusses whether or not we can think of machines as entities in their own right (Nayar, 2024). He prompts us to question who possesses the authority to withhold the entitlements of machines, especially in a world where humans are increasingly integrated with technology. Specifically, Nayar raises the question of whether robots, capable of performing tasks traditionally associated with humans and demonstrating cognitive and emotional capabilities, including biases, should be excluded from the institution of citizenship. He contends that withholding rights and responsibilities from robots could in itself be a form of discrimination. This provocative piece urges us to consider whether humans are the sole arbiters of the world's governing principles or if the paradigm of human-robot interactions should be viewed as intertwined rather than separate when considering these principles.

Conflict of Interests

The author declares no conflict of interests.

References

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. MIT Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency in Proceedings of Machine Learning Research*, *81*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., Noble, S. U., & Shestakofsky, B. (2021). Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius*, *7*. https://doi.org/10.1177/2378023121999581

Larsson, S., Merricks White, J., & Ingram Bogusz, C. (2024). The artificial recruiter: Risks of discrimination in employers' use of ai and automated decision-making. *Social Inclusion*, *12*, Article 7471. https://doi.org/10.17645/si.v12.7471

Nayar, P. K. (2024). How to include artificial bodies as citizens. *Social Inclusion*, *12*, Article 8337. https://doi.org/10.17645/si.v12.8337

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishers.

Orwat, C. (2024). Algorithmic discrimination from the perspective of human dignity. *Social Inclusion*, *12*, Article 7160. https://doi.org/10.17645/si.v12.7160

Ulnicane, I. (2024). Intersectionality in artificial intelligence: Framing concerns and recommendations for action. *Social Inclusion*, *12*, Article 7543. https://doi.org/10.17645/si.v12.7543

About the Author

Derya Ozkul (PhD) is an assistant professor in sociology at the University of Warwick and a research associate at the University of Oxford's Refugee Studies Centre.