LONDON JOURNAL OF SOCIAL SCIENCES

eISSN 2754-7671

AI discrimination trap? - Recommendations for combating discrimination against women through AI

Christina Schabasser*

Abstract

The current overrepresentation of men in the development of AI technologies and the associated one-sided gender perspective is one of the reasons for discrimination against women in AI applications. In this context, AI exacerbates such overrepresentation, thus presenting a new discriminatory risk against women. Today´s omnipresence of AI makes the need for fair algorithms even more crucial. Thus, this study analyzes five application areas of AI (i.e. virtual private assistance, public and private transport and safety, precision medicine, employment area, and credit rating) and discusses potential biases against women identified and proposes anti-discrimination suggestions to overcome this kind of invisibility of women in many AI applications.

Keywords: AI bias, gender bias in AI, AI discrimination, algorithmic bias, AI gender gap

https://doi.org/10.31039/ljss.2025.9.319

* University of Applied Sciences Burgenland, Austria, christina.schabasser@live.at

1. Introduction

Milestones in Artificial Intelligence development, such as autonomous, self-driving cars, best job candidate identification in recruiting, facial recognition and fingerprints to unlock phones, decision-making about what content to display on people's social media feeds, and improvements in the diagnostic phase in the medical sector, show that AI is everywhere and affects everyone (European Union Agency for Fundamental Rights, 2022; De Felice et al., 2022). While those examples only represent a small part of the areas of AI application, they show how diverse such an application is.

This article focuses on the dark sides of AI, which are typically slightly obvious. AI carries a risk of reinforcing discrimination, e.g. bias against women. A UNESCO study examining stereotyping in Large Language Models (LLMs) found clear evidence of bias against women in the LLM-generated content. For instance, the technology tends to assign more diverse, high-status jobs to men than women, with terms such as "business", "executive", "salary", and "career" being associated with men. Since LLMs are used every day by more and more people in their work, studies, and at home, these AI applications can have subtle but lasting impacts on people's perceptions, reinforcing real-world inequalities (UNESCO, 2024).

Moreover, computer scientist Andrew Ng claimed that just as electricity has fundamentally changed the world, forward-looking AI is expected to fundamentally change many areas of our lives (Lynch, 2017). AI is often described as a powerful tool that will revolutionize the world as it has become more prevalent over time worldwide. That said, it is crucial to investigate how AI can also reproduce discrimination and how more fairness can be introduced into the models. Although this paper focuses on how AI exacerbates biases against women, it is important to mention that algorithms can also uncover existing discrimination. In this context, these algorithms can promote both injustice and justice simultaneously (Heinrichs, 2022).

This paper examines five areas of application of AI: the Employment Area, Precision Medicine, Transportation, Credit Risk Assessment, and Virtual Personal Assistants, to describe the biases found in each of those areas and present solutions for any discrimination identified there.

2. Bias Against Women in AI

AI systems should be urgently re-evaluated as they reproduce patterns of racial and gender prejudice which may reinforce inequalities (West et al., 2019). When algorithms are fed by training data with historical discrimination patterns, such discrimination is maintained by the algorithm (Hoffman et al., 2020). The biases in training data are only one of the potential sources of bias in AI systems, although biased recommendations from AI systems are assumed to negatively influence human decisions. Alarmingly, humans seem to reproduce the same biases exhibited by AI, even after their interaction with the biased AI system has ended (Vicente et al., 2023).

The AI systems are often opaque, thus making it even more difficult to eliminate biases in the system. Concerning the bias against women in AI applications, one suggestion would be to promote the participation of women in design processes to reduce the systemic bias against women. As a diversity crisis shakes the AI sector and women are severely underrepresented in it (Armutat et al., 2023), it remains to be seen what role women will play in the fight for a bias-free AI world. For example, 80% of AI professors are male, the portion of women who work as AI research staff is low at Facebook (15%) and Google (10%), and the proportion of female authors at leading AI conferences is only at 18% (West et al., 2019). To address these conditions, the following sub-sections focus on describing AI biases in specific application areas and offering solutions to overcome discrimination against women in those areas.

2.1. Employment Area

To identify the best applicants, Amazon developed an AI-based software with the aim of identifying the best applicants by reviewing applicant's resumes. This software would utilize the historical applicant data in organizational applicant tracking systems to later develop hiring decisions. The aim was to free the recruitment process from dependence on human subjectivity. There is often a tendency among recruiters and HR managers to hire people who are similar to them. But as it turned out, the algorithm was biased against women and routinely suggested men as the better job candidates. In particular, the recruiting machine learning engine did not evaluate candidates for software developer jobs and other technical positions in a gender-neutral manner.

By analyzing data about the past, the algorithm can make predictions about the future. If these past data contain biased judgments, these biases will also be transferred to the models. For example, imagine a hiring algorithm that selects candidates by comparing them with an employer's current workforce. If a hiring company has employed a very small number of women as computer programmers in the past, the algorithm simply reproduces the previously existing biases (Hoffman et al., 2020; Iriondo, 2018; Ore et al., 2020); Kim 2019).

An important cornerstone for change is learning from bias. Data and analytics tools used in recruitment uncover bias against women. For example, a case study company attempted to find the reasons for bias by surveying employees and found that women face sexual harassment in certain areas and that women are excluded from important training or mentoring opportunities. The company demonstrated an approach that promotes equality in the workplace rather than undermining it.

To bring about change, constant questioning is necessary. Employers should ask the developers of algorithms hard and uncomfortable questions instead of blindly trusting their feeling of neutrality to ensure that they may use AI responsibly to promote equality in the employment area. Last but not least, Human-machine cooperation should be considered. AI tools should be seen as supporting tools for new team members rather than as a threat to the current recruiter's job (Kim, 2018).

2.2. Precision Medicine

Precision medicine is concerned with finding personalized prevention and treatment strategies by considering differences in genes, environment, and lifestyle throughout life. Evidence of gender differences has been identified in chronic diseases such as diabetes, cardiovascular disorders, neurological diseases, mental health disorders, cancer, and autoimmunity, as well as physiological processes such as brain aging and sensitivity to pain. Therefore, a desirable bias concerning the differential consideration of men and women seems to be a necessary condition in Precision Medicine to unlock a reasonable diagnosis and tailored treatment recommendations.

Unfortunately, differences observed in sex and gender in health and well-being are often accompanied by stigma, stereotypes, and misrepresentation of data. Misdiagnosis of female patients is typically due to the lack of data regarding female health conditions, which makes AI systems unable to make accurate diagnoses (Cirillo et al., 2020). To offer a promising perspective for women regarding the use of AI for precision medicine, some suggestions are now made.

Clinical research should focus on data about both males and females, although studies with women are expected to be more complex due to the variability of their hormone levels. There is a need to increase sensibility for potential gender biases by including the efforts necessary to understand the specific parameters used to make clinical conclusions. When using AI in precision medicine, the question "Personalized medicine is fine, but for whom?" should always be at the forefront. Researchers who address large clinical data sets and apply that knowledge to individual patients need to know how to identify biases. It is necessary to employ diverse datasets to train algorithms to prevent biases (Cirillo et al., 2020; Hill 2023; Carnevale et al., 2023; Hajirasouliha et al.,2020; Yogeshappa, 2024).

2.3. Transportation

When considering AI in transport, many may think of autonomous vehicles that perform all driving tasks without human intervention (Asha et al., 2024). The question is whether gender-based preferences and safety issues are taken into account when using AI in the transportation sector. Safety systems in cars (seatbelts, headrests, and airbags) are designed predominantly based on data from crash tests with dummies that consider men's physiques and their seating positions. Women's breasts and pregnancies are not taken into account when designing car safety systems. The result shows that compared to men, women are 47% more likely to be heavily insured and 17% more likely to die than men in a similar accident (Niethammer, 2020).

Furthermore, driving routines can be different for women and this has to be considered in the planning process. In London, bus passengers were asked to pay every time they boarded a new bus, which particularly affected women. Only the introduction of the ’hopper fare’ brought relief, meaning that one ticket could be used for two trips within one hour now (Schuß et al., 2021). To address these and further biases against women in transportation, some suggestions are made. The requirements and preferences of female passengers in cars have to be addressed.

Moreover, the differences in preferences of passengers in relation to convenience, cost, and safety have to be considered. The participation of women in VR scenarios with traditional-sized car models to raise concerns and develop common solutions has to be taken into account. Women should be included in the planning activities of public transportation. Last but not least, a pluralism approach which considers a variety of viewpoints in research, should be adopted (Asha et al. 2024; Schuß 2021).

2.4. Credit Risk Assessment

The process of lending money requires extensive attention. Financial institutions can use AI to assess the risk of credit applicants by differentiating between "good" and "bad" applicants and offering a risk-based interest rate based on this data (Kelly et al., 2021).

In this regard, gender gaps are reported in accessing formal loans (Demirguç-Künt et al., 2015). The challenge is to minimize bias in AI applications in the area. Credit scoring models that rely on traditional data sources such as credit histories, asset ownership, and formal income are biased because women have historically been excluded from credit markets (Chioda et al., 2024).

When a population is over- or under-represented in a training dataset – the so-called sampling bias – the algorithms are unfair. For example, a digital credit app in a market reported how men are more likely to own smartphones than women. Such a situation makes the algorithm rely more on men's data than women's since men are overrepresented in the training dataset. Labeling biases exist when loan applicants' occupations are labeled as "doctor" or "nurse" instead of "health care professional." While the terms "doctor" and "nurse" provide information about gender, the neutral term "health care professional" hides information about gender (Kelly et al., 2021).

Some suggestions to bring more fairness to AI systems in this application area include granting that AI enables a fair and equal chance of having a good predicted credit score for male as well as female loan applicants. AI makes it equally likely for male and female credit applicants with a good predicted credit score to actually have a good credit score. Sensitive attributes are excluded from the decision-making process. A clear definition of fairness in this context has to be made (Kelly et al., 2021; European Union Agency for Fundamental Rights, 2022). Alternative data sources such as digital footprints could be used to assess the creditworthiness of population groups with no or limited credit history (Chioda et al., 2024).

2.5. Virtual Personal Assistants (VPAs)

AI-powered VPAs are widespread and growing in number. Amazon, for example, hires an average of 14.2 new employees every day to work on its Alexa and Echo VPA systems. The most common VPAs available on the market today include Alexa (Amazon), Cortana (Microsoft), and Siri (Apple) (Ni Loideain et al., 2018). Most of the leading voice assistants today are female. This was not always the case if thinking back to speaking car navigation systems that gave very brief instructions almost in a male voice. A BMW 5 Series from the late 1990s was one model that was equipped with a female navigation voice. Interestingly, this car model was recalled in Germany because drivers did not want to be given instructions by a woman (West M. et al., 2019).

Today´s VPSs are mostly clearly connoted with female persons with female names and voices, thus reiterating the impression that users prefer to give instructions and orders to a female rather than a male. The female voice can evoke unconscious characteristics such as helpfulness, caring behavior, and submissiveness. The design decisions of these VPAs reproduce gender stereotypes and portray women as subordinate to men (Ni Loideain et al., 2018). When female virtual assistants perform administrative jobs such as taking notes, sending messages, reminding, seeking information, and managing calendars, this reinforces traditionally gendered perceptions. Virtual assistants reflect the "feminization of AI", where the division of labor between men and women is normalized (Sutko, 2020). The described gender assignments of VPA technologies can cause social harm. The damage is particularly great regarding younger users of AI VPAs. Because, as research shows, gender stereotypes (e.g. brilliance being a male quality) develop in humans as early as the age of six (Ni Loideain et al., 2018).

Since many preferences, assumptions, and preconceptions of their creators are entering the virtual assistants (Manasi et al., 2022) and AI professions are a male-dominated sector, there might be less hope of changing AI discrimination against women by virtual assistants. An important measure to reduce biases against women in this application area of AI is to place more attention on improving and increasing the representation of women in STEM (science, technology, engineering, and math) subjects as well as in design teams as well as decision-making positions in the area of new technologies (Ni Loideain et al., 2018).

Developers and designers should definitely include variations in the design, language, and tone across cultures. Cultural sensitivity should gain high attention. Ethical designs and thoughtful decision-making should be used to reduce unconscious bias in VPAs. Major players in the market should critically examine the marketing of their products and engage with both the female voice of the VPAs and the responses of their VPAs to avoid portraying stereotypical and heteronormative female characterizations. Policy documents on AI and related technologies published by the EU, U.S., and UK (amongst others), should be reworked and edited to meet the ethical standards even more to include considerations on not just how AI technologies may produce or reproduce social biases, but whether they include social biases within their very design (Ni Loideain et al., 2018; Singh et al., 2024).

Conclusion

Gender biases were identified in all the AI application fields analyzed (i.e. virtual private assistance, public and private transport and safety, precision medicine, employment area, and credit rating). This study aimed to raise awareness of AI-led biases against women to move toward fair AI development. Since AI is inevitable in our everyday lives, it is highly important to pay attention to these discriminating AI technologies to avoid the exacerbation of gender stereotypes. This study presents recommendations for all fields affected by AI bias, including but not limited to

- analyze the variety of training data for AI algorithms.
- focus research on data about both males and females, despite data from females are being expected to be more complex.
- revisit and amend the policy positions of the EU, U.S., and UK (amongst others) as presented in their policy documents on AI and related technologies.
- promote women's involvement in STEM subjects.
- promote participation of women in design processes.

References

Armutat, S.; Mauritz, N., Prädikow, L.; Schulte, M.; Wattenberg, M. (2023). Fit für KI? – Genderspezifische Unterschiede in der Wahrnehmung, dem Verständnis und in den Weiterbildungswünschen bezüglich Künstlicher Intelligenz. https://doi.org/10.57720/3734

Asha, A.Z.; Sultana, S.; He, H.; Sharlin, E. (2024). "Shotitwo First!": Unraveling Global South Women's Challenges in Public Transport to Inform Autonomous Vehicle Design. Proceedings of the 2024 ACM Designing Interactive Systems Conference, 3193-3209. https://doi.org/10.1145/3643834.366155

Carnevale, A.; Tangari, EA; Iannone. A; Sartini, E. (2023). Will Big Data and personalized medicine do the gender dimension justice? AI & Society, 38(2), 829-841. https://doi.org/10.1007/s00146-021-01234-9

Chioda, L.; Gertler, P.; Higgins, S.; Medina, P. (2024). Equitable AI Challenge: Improving access to credit with gender-differentiated credit scoring algorithms: Executive Summary. https://www.usaid.gov/sites/default/files/2024-06/AI%20Executive%20Summary_Credit%20Scoring_1_0.pdf

Cirillo, D.; Catuara-Solarz, S.; Morey, C.; Guney, E.; Subirats, L.;Mellino, S.; Gigante, A.A.; Valencia, A.; Rementeria, M.J.; Chadha, A.S.; Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digital Medicine, 3*, 81. https://doi.org/10.1038/s41746-020-0288-5

De Felice, F.; Petrillo, A.; Luca, C.; Baffo, I. (2022). Artificial Intelligence or Augmented Intelligence? Impact on our lives, rights and ethics. Procedia Computer Science, 200, 1846-1856. https://doi.org/10.1016/j.procs.2022.01.385

Demirguç-Künt, A.; Klapper, L.; Singer, D.; Van Oudheusden, P. (2015). The global findex database 2014: Measuring financial inclusion around the world. World Bank Policy Research Working Paper 7255. https://thedocs.worldbank.org/en/doc/681361466184854434-0050022016/original/2014GlobalFindexReportDKSV.pdf

European Union Agency for Fundamental Rights (2022, December 8). Bias in Algorithms – Artificial Intelligence and Discrimination. https://fra.europa.eu/en/publication/2022/bias-algorithm

Hajirasouliha, I.; Elemento, O. (2020). Precision medicine and artificial intelligence: overview and relevance to reproductive medicine. Fertility and Sterility, 114, 908-913. https://doi.org/10.1016/j.fertnstert.2020.09.156

Heinrichs, B. (2022). Discrimination in the age of artificial intelligence. AI & Society, 37, 143–154. https://doi.org/10.1007/s00146-021-01192-2

Hill, T. (2023, October 2). Bias in AI for precision medicine. https://www.reprocell.com/blog/bias-in-ai-for-precision-medicine

Hoffman, S.; Podgurski,A. (2020). Artificial Intelligence and Discrimination in Health Care. Yale Journal of Health Policy, Law, and Ethics, 19 (3), Case Legal Studies Research Paper No. 2020-29. Available at SSRN: https://ssrn.com/abstract=3747737

Iriondo, R. (2018, October 11). Amazon Scraps Secret AI Recruiting Engine that Showed Biases Against Women. https://www.ml.cmu.edu/news/news-archive/2016-2020/2018/october/amazon-scraps-secret-artificial-intelligence-recruiting-engine-that-showed-biases-against-women.html

Kelly, S.; Mirpourian, M. (2021). Algorithmic Bias, Financial Inclusion, and Gender. New York: Women's World Banking. https://www.womensworldbanking.org/wp-content/uploads/2021/02/2021_Algorithmic_Bias_Report.pdf

Kim, P. (2018). Big Data and Artificial Intelligence: New Challenges for Workplace Equality. University of Louisville Law Review, Forthcoming. https://ssrn.com/abstract=3296521

Lynch, S. (2017, March 11). Andrew Ng: Why AI Is the New Electricity. https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity

Manasi, A.; Panchanadeswaran, S.; Sours, E.; Lee, S. (2022). Mirroring the bias: gender and artificial intelligence. *Gender, Technology and Development*, *26*(3), 295–305. https://doi.org/10.1080/09718524.2022.2128254

Niethammer, C. (2020, May 27). AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators. https://www.forbes.com/sites/carmenniethammer/2020/03/02/ai-bias-could-put-womens-lives-at-riska-challenge-for-regulators/

Ni Loideain, N.;Adams, R. (2018, November 9). From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of EU Data Protection Law. King's College London Dickson Poon School of Law Legal Studies Research Paper Series, https://ssrn.com/abstract=3281807 or http://dx.doi.org/10.2139/ssrn.3281807

Ore, O.; Sposato, M. (2022). Opportunities and risks of artificial intelligence in recruitment and selection", International Journal of Organizational Analysis, 30 (6), 1771-1782. https://doi.org/10.1108/IJOA-07-2020-2291

Schuß, M.; Wintersberger, P.; Riener, A. (2021). Security Issues in Shared Automated Mobility Systems: A Feminist HCI Perspective. Multimodal Technologies and Interaction, 5(8), 43. https://doi.org/10.3390/mti5080043

Singh, S.; Kumar, A.; Bose, S. (2024). Behind The Feminine Facade: Gender Bias in Virtual Assistants and Its Effect on Users. Journal of Ecohumanism, 3 (4), 351-356. http://dx.doi.org/10.62754/joe.v3i4.3592

Sutko, D. M. (2020). Theorizing femininity in artificial intelligence: a framework for undoing tech nology's gender troubles. Cultural Studies, 34(4), 567–592. https://doi.org/10.1080/09502386. 2019.1671469

UNESCO (2024, July 5). Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes. https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes

Vicente L; Matute H. (2023). Humans inherit artificial intelligence biases. Scientific Reports, 13(1), article number 15737. https://doi.org/10.1038/s41598-023-42384-8

West, S.M.; Whittaker, M.; Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. https://ainowinstitute.org/wp-content/uploads/2023/04/discriminatingsystems.pdf

West, M.; Kraut, R.; Chew, H.E. (2019). I'd blush if I could: closing gender divides in digital skills through education. https://unesdoc.unesco.org/ark:/48223/pf0000367416

Yogeshappa, V. (2024). AI-driven Precision medicine: Revolutionizing personalized treatment plans. International Journal of Computer Engineering and Technology, 15 (5), 455-474, https://doi.org/10.5281/zenodo.13843057