

POLICY BRIEF 03

In response to the United Nations Global Digital Compact (Zero Draft, 1 April 2024)

An Intersectional Approach to Gender and Artificial Intelligence

Larissa Reichel, Poojika Amarakoon, Peppi Bruns, Denise Brancaccio, Giovanni Capone, Pauline Prochnow



This policy brief is part of a six-part series produced by students participating in the Erasmus+ blended-intensive program “Digital Constitutionalism and the UN Global Digital Compact” hosted at the University of Bremen.

Between March and June 2024, around 30 students and ten instructors worked to build and put to work transversal skills in a transnational research-based learning and policy-engaged learning setting. As part of the program, students worked together for a week at the University of Bremen and participated in the European Dialogue on Internet Governance (EuroDIG) 2024 policy conference.

Executive Summary

The pressing issue of bias and discrimination against minorities in Artificial Intelligence (AI) is addressed through this policy brief by applying an intersectional approach to gender. The term “Intersectionality” is the development of social identities in counteracting different individual experiences and social power dynamics, whereby it is a mutually constitutive relationship.¹ The identified problem of using incomprehensive datasets in AI development and the adherence to traditional gender perspectives in standard-setting continues to reflect in the field amplifying societal bias, particularly against women and other marginalized groups, despite existing regulatory frameworks. Drawing evidence from academic research and industry reports, it derives key recommendations on technical and normative measures to foster intersectionality, curb societal bias, and promote equity. Hence, it emphasizes the need for coordinating the technical community, especially AI developers, and the advisory board to the United Nations Global Digital Compact in understanding and mitigating these biases in the long run for the benefit of society and the economy.

.

Introduction

The technological evolution of recent years is demonstrating the increasing role that AI is taking on in society and in the world of work, but being an innovation still under development, its regulatory framework is still vague and incomplete. To better understand what is meant by Artificial Intelligence, it is useful to report two definitions. The first is given by the US company IBM, which defines it as: “[...] technology that enables computers and machines to simulate human intelligence and problem-solving capabilities.”² and the second by the European Union, according to which: “Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.”³

.

¹ Shields, S. (2008). *Gender: An Intersectionality Perspective*. *Sex Roles*. 59. 301-311. 10.1007/s11199-008-9501-8.

² IBM.com, *What is Artificial Intelligence?*, <https://www.ibm.com/topics/artificial-intelligence>

³ *The European Commission’s High - Level Expert Group on Artificial Intelligence, A definition of AI: Main capabilities and Scientific Disciplines, 2018, https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf*

There are more and more cases of discrimination against marginalised groups found in AI, which replicate and perpetuate gender and ethnic stereotypes despite attempts at regulation at the European level with the AI Act (European Union, 13/03/2024), which seems to play a role in moving in the direction of increasingly regulated and transparent AI. These discriminations seen in AI can be partly attributed to the lack of an intersectional perspective, which considers "The interconnected nature of social categorizations such as race, class, and gender, regarded as creating overlapping and interdependent systems of discrimination or disadvantage; a theoretical approach based on such a premise."⁴.

Gender bias in leadership: "73% of business leaders believe having more women leadership is important for mitigating gender bias in AI, only 33% currently have a woman in charge of decision-making for AI strategy."
Facial recognition disparities: "Error rates vary from 35% for darker-skinned women, to 12% for darker-skinned men, 7% for lighter-skinned women, and less than 1% for lighter-skinned men."
Bias in recruitment: Amazon's resume evaluation AI preferred male candidates due to historical male dominance in tech roles.
Racial bias in image generation: Associating dark-skinned people with low-paying jobs.

AI replicates and reinforces human biases, including gender biases, being trained through the use of inadequate datasets, the error of which is present in their data collection and interpretation. "While 73% of business leaders believe having more women leadership is important for mitigating gender bias in AI, only 33% currently have a woman in charge of decision-making for AI strategy."⁵.

⁴ *Oxford English Dictionary, Intersectionality, meaning and use;*
https://www.oed.com/dictionary/intersectionality_n?tab=meaning_and_use#335059764

⁵ *Lin, S. Downie, A 2024, No silver bullet; Closing the gender gap in the era of generative AI,*
<https://www.ibm.com/blog/no-silver-bullet-closing-the-gender-gap-in-the-era-of-generative-ai/>

According to a National Institute of Health publication, based on a study on intersectional gender inequality, facial recognition systems work better on faces belonging to men than women and on light rather than dark skin. "Error rates vary from 35% for darker-skinned women, to 12% for darker-skinned men, 7% for lighter-skinned women, and less than 1% for lighter-skinned men."⁶.

In 2015, Amazon discovered a bias in its system for evaluating candidates' resumes. Male candidates were systematically preferred over women for roles in IT and software development. The AI had been trained taking into account the CV models presented to society in the last 10 years, in which due to historical and social factors men were predominant in the tech industry.

A UNESCO study has shown how gender stereotypes in today's society are also reflected in AI-generated texts and images. In fact, it emerges from this work that men are associated with more intellectual jobs with a higher status (e.g., engineers, teachers, doctors) than women, who are associated with more menial jobs, such as janitors and cooks. In addition, terms such as "husband" and "love" are used to describe women, while men are associated with terms such as "wealth" and "adventure."

Recently, the *New York Times* talked about how AI is unable to properly generate images depicting black people, lightening their skin and ignoring instructions given to it. In addition, although this is also the case for white women, AI tends even more to associate darker-skinned people with low-paying jobs, such as housekeepers and dishwashers.

In addition, AI is also used in the medical field, to make diagnoses, and in safety, with regard to predictive policing activities. Even in these areas, the results reveal the presence of biases, which prevent the effectiveness of the systems. The discrimination implemented by artificial intelligence is therefore much wider than one might think, and its effects have a significant impact on people's lives. The need to understand and counter these issues has already been recognized by the Charter of Feminist Demands from the Global South and The Declaration Of Feminist Digital Justice, which call for the intersectional vision to be recognized as fundamental among the principles of the Global Digital Compact.

⁶ Zou J, Schiebinger L. *Ensuring that biomedical AI benefits diverse populations.* *EBioMedicine.* 2021 May;67:103358. doi: 10.1016/j.ebiom.2021.103358. Epub 2021 May 4. PMID: 33962897; PMCID: PMC8176083

Critique of the current situation and alternative options

Overview of the current situation:

GDC Zero Draft mentions the aim of empowering women and girls and the importance of gender.

GDC and advisory boards have a very traditional feminist approach which arises through the structural conflict approach of patriarchy, in their sections on gender and matters are addressed somewhat vaguely.

Serious issues with AI that can reinforce and amplify human bias and inequality, including gender.



Gender_Bias, © Science.org

Why the current implementations are ineffective:

People who are part of two or more marginalized groups have unique experiences and these experiences need to be accounted for in their complexity, not as separate issues.

Again, there are serious issues with inequality in AI. This is proof that change is needed, yet this is not addressed enough.

AI is currently viewed through fairness, accountability, transparency, and ethics (FATE)⁷ where it fails to incorporate the intersectional approach.

Alternative ways of implementation:

The United Nations introducing a more direct and intersectional approach that should be included in the GDC.

The GDC influencing the technical community to focus on the gender and intersectional aspects of AI.

The feasibility aspect:

The proposed policies are feasible, with regard to the two separately addressed parties. The technical community has the resources and capabilities to implement concrete technical solutions, and the United Nations bodies have the power to create soft laws and incentives that increase the political will.

The key here is for both addressed parties to have the desire to put these changes into action. The biggest motivator should be the established fact that gender equality will benefit society and the economy in the long run.⁸ That being said, it might be challenging to persuade people of this. Politics plays a part in this, as even the most basic gender equality issues might be challenging to include.

⁷ Bauer, G. R., & Lizotte, D. J. (2021). *Artificial Intelligence, Intersectionality, and the Future of Public Health*. *American journal of public health*, 111(1), 98–100. <https://doi.org/10.2105/AJPH.2020.306006>

⁸ Lagerlöf, N. (2003). *Gender Equality and Long-Run Growth*. *Journal of Economic Growth*, 8, 403-426 ; Maceira, H. (2017). *Economic Benefits of Gender Equality in the EU*. *Intereconomics*, 2017, 178-183.

The aim of empowering women and girls and the importance of gender is mentioned in the GDC Zero Draft Principles, more specifically in section 7(d) which states the following:

“(d) Gender equal: The full, equal, and meaningful participation and leadership of women and girls in the digital space is essential to close gender digital divides and advance sustainable development. Our cooperation will empower women and girls, mainstream gender perspectives and prioritize the elimination of sexual and gender-based violence online”.

Inequality is more specifically mentioned in section 27 of the GDC, which states the following: “We must urgently eliminate and prevent technology-facilitated gender-based and sexual violence, hate speech, discrimination, information manipulation and disinformation, cybercrime, cyberbullying and online child sexual exploitation and abuse. We acknowledge our collective responsibility to establish and maintain robust risk mitigation and redress measures that also protect privacy and freedom of expression.”

Furthermore, section 49(c) states the following: “Call on international and national standard-setting organizations to collaborate to promote the harmonization of AI standards that uphold safety, reliability, environmental sustainability, gender equality and human rights (SDGs 5, 7, 9, 10, 12, 16 & 17)”.

Section 49(c) demonstrates the link between the GDC and the technical community and establishes the important role of international and national standard-setting organizations in the fight to promote equality in AI. As can be seen from the sections above, while the GDC mentions gender equality and equality in general, an intersectional approach is lacking. Equality-related issues with AI are not explicitly mentioned, although section 49(c) implicitly refers to them by stating that “harmonization of AI standards that uphold...gender equality” is needed.

This policy brief is addressed not only to the United Nations systems that affect the wording of the GDC but also to the technical community. Herein lies the challenge when it comes to the wording of the GDC. On the one hand, Compacts such as the GDC should be kept somewhat vague to increase flexibility and global cooperation. On the other hand, the GDC has the potential to create more impactful change, especially within the technical community, if the issues highlighted in this brief are explicitly addressed. We argue that in order for the technical community to implement the needed concrete technical solutions, the United Nations systems and the GDC should use their influence and focus on the matter and thus create soft laws and incentives that increase political will. This would include addressing the

issue of inequality in AI directly, keeping in mind intersectional perspectives.

Recommendations

As we want to reach more than one target group with our Policy Brief, we have divided our recommendations into two sections: *technical and normative*.

The section on technical recommendations is explicitly aimed at the technical community, i.e. all those actively involved in the development of AI-generated software. The other section covering normative recommendations is aimed at the general public (but also includes the technical community and the UN systems to some extent). However, to implement our ideas and recommendations in concrete terms, we explicitly address them to the UN so that everyone involved can take a leading role when it comes to executing the suggested normative recommendations in particular.

Technical Recommendations

Technical recommendations are aimed at parties involved in the development, implementation, and application of AI systems. These recommendations are more concrete in nature and offer specific pathways that can be taken to ensure a more intersectional perspective to fostering gender equality within these technologies.

- Retain a "Human-in-the-Loop"⁹, to ensure that decision-making processes are verified by human actors
- Achieve counterfactual fairness¹⁰ by expanding the use of methods and forms of media or data to develop AI¹¹
- Identify and watch out for sensitive areas.
 - ❑ Measure accuracy levels differently among certain populations and people from overlapping disadvantaged groups.
 - ❑ Make the development and training process transparent to the UN systems.
 - ❑ Have diverse development teams in order to achieve diverse databases.
 - ❑ Focus on high diversity in training samples and more adequate tests before the introduction of AI into systems.

Normative Recommendations

The following recommendations address more broad conceptual changes needed to facilitate intersectional gender equality. To make these recommendations more tangible, examples will be featured. The general recommendation for the UN systems is that the topic should be directly addressed in the GDC and further regulatory measures.

- Address the societal background that is responsible for biases
 - ❑ e. g. further intersectional efforts in all sectors of research, etc. that aim to improve gender equality
 - ❑ e. g. employ continuous awareness of intersectional experiences in daily life and actively work to mitigate them
- Aim for equity and proactively approach cases of inequality
 - ❑ e. g. offer disadvantaged employees of the UN systems additional support so that they can equally participate in regulation making
 - ❑ e. g. emphasize the importance of gender intersectionality through education and training
 - ❑ e. g. improving opportunities for women in STEM professions
- Specific recommendations for the GDC:
 - ❑ Mention intersectionality in context to gender equality
 - ❑ Emphasize the benefits of a diverse database
 - ❑ Encourage guidelines for a responsible approach by AI developers
 - ❑ Encourage intersectional approaches to equality (e. g. through mentioning economic and societal benefits)

⁹ Arambepola, N., & Munasinghe, L. (2021). *Human in the loop design for intelligent interactive systems: A systematic review*.

¹⁰ introduced by Pearl, J. (2009). *Causality : Models, Reasoning and Inference (2nd ed)*. Cambridge University Press. <http://site.ebrary.com/id/10697730>; applied to AI e. g. Wu, Y., Zhang, L., & Wu, X. (2019, August).

Counterfactual fairness: Unidentification, bound and algorithm. In *Proceedings of the twenty-eighth international joint conference on Artificial*

¹¹ Sarah.C (2019). *Intersectional AI Is Essential*. *Journal of Science and Technology of the Arts*. 11. 3 8.10.7559/citarj.v11i2.665.

Acknowledgement

We extend our sincere gratitude to the academic committee of the Blended Intensive Program (BIP) for their invaluable guidance and feedback throughout this project. We also acknowledge that the BIP has been funded by the Erasmus+ program.

References

- Bauer, G. R., & Lizotte, D. J. (2021). Artificial Intelligence, Intersectionality, and the Future of Public Health. *American journal of public health*, 111(1), 98–100.
<https://doi.org/10.2105/AJPH.2020.306006>
- Buslón N., Cortés A., Catuara-Solarz S., Cirillo D., & Rementeria, M. J. (2023). Raising awareness of sex and gender bias in artificial intelligence and health. *Frontiers in global women's health*, 4, 970312. <https://doi.org/10.3389/fgwh.2023.970312>
- Dastin, J.(2018) Insight - Amazon scraps secret AI recruiting tool that showed bias against women,Reuters,<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/>
- European Commission, 2018, European Commission's High - Level Expert Group on Artificial Intelligence, A definition of AI: Main capabilities and Scientific Disciplines, https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf
- European Commission, (2020), Directorate-General for Research and Innovation, Gender & intersectional bias in artificial intelligence, Publications Office, European Commission <https://data.europa.eu/doi/10.2777/148>
- European Parliament, (2019-2024) P9_TA(2024)0138, Artificial Intelligence Act https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf
- Feminist Digital Justice, (2023), The Declaration of Feminism Digital Justice, <https://feministdigitaljustice.net>
- Gurumurthy.A and Chami.N, (n.d) Charter of Feminist Demands From the Global South, Friederich Ebert Stiftung, https://www.fes.de/public/FES/Newsletter-Bilder_IEZ/GePol/MailingsDoc/charter-of-feminist-demands-from-the-global-south.pdf
- IBM.(n.d) What is Artificial Intelligence? <https://www.ibm.com/topics/artificial-intelligence>
- Oxford English Dictionary, Intersectionality, meaning and use; https://www.oed.com/dictionary/intersectionality_n?tab=meaning_and_use#335059764
- Lagerlöf, NP. (2003) Gender Equality and Long-Run Growth. *Journal of Economic Growth* 8, 403–426 . <https://doi.org/10.1023/A:1026256917489>
- Lin, S., & Lin, S. (2024, May 15). No silver bullet: Closing the gender gap in the era of generative AI. IBM Blog. <https://www.ibm.com/blog/no-silver-bullet-closing-the-gender-gap-in-the-era-of-generativ>

e-ai/

Maceira, H. M. (2017). Economic Benefits of Gender Equality in the EU. *Intereconomics*.
<https://www.intereconomics.eu/contents/year/2017/number/3/article/economic-benefits-of-gender-equality-in-the-eu.html>

Maiei. (2023, December 14). Public perceptions of gender bias in large language models: Cases of ChatGPT and Ernie | Montreal AI Ethics Institute. Montreal AI Ethics Institute.
<https://montrealethics.ai/public-perceptions-of-gender-bias-in-large-language-models-cases-of-chatgpt-and-ernie/>

Sarah.C (2019). Intersectional AI Is Essential. *Journal of Science and Technology of the Arts*. 11. 3-8. 10.7559/citarj.v11i2.665.

Shields,S. (2008). Gender: An Intersectionality Perspective. *Sex Roles*. 59. 301-311. 10.1007/s11199-008-9501-8.

Small Z.(2023) Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History. *New York Times*,
<https://www.nytimes.com/2023/07/04/arts/design/black-artists-bias-ai.html>

Storbeck.M, (2022). Artificial intelligence and predictive policing: risks and challenges, Brussels: EUCPN, <https://eucpn.org/sites/default/files/document/files/PP%20%282%29.pdf>

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

Zou J, Schiebinger L.(2021) Ensuring that biomedical AI benefits diverse populations. *EBioMedicine*. May;67:103358. doi: 10.1016/j.ebiom.2021.103358. Epub 2021 May 4. PMID: 33962897; PMCID: PMC8176083.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8176083/>