

10-9-2023

## Power Imbalances in Society and AI: On the Need to Expand the Feminist Approach

Eva Johanna Gengler

*Friedrich-Alexander-Universität Erlangen-Nürnberg; FemAI - Center for Feminist Artificial Intelligence GmbH, eva.gengler@fau.de*

Marco Wedel

*Technische Universität Berlin, marco.wedel@tu-berlin.de*

Alexandra Wudel

*FemAI - Center for Feminist Artificial Intelligence GmbH, alexandra.wudel@fem-ai.com*

Sven Laumer

*Friedrich-Alexander-Universität Erlangen-Nürnberg, sven.laumer@fau.de*

Follow this and additional works at: <https://aisel.aisnet.org/wi2023>

---

### Recommended Citation

Gengler, Eva Johanna; Wedel, Marco; Wudel, Alexandra; and Laumer, Sven, "Power Imbalances in Society and AI: On the Need to Expand the Feminist Approach" (2023). *Wirtschaftsinformatik 2023 Proceedings*. 37.

<https://aisel.aisnet.org/wi2023/37>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik 2023 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Power Imbalances in Society and AI: On the Need to Expand the Feminist Approach

## Research in Progress

Eva Johanna Gengler<sup>1,3</sup>, Marco Wedel<sup>2</sup>, Alexandra Wudel<sup>3</sup>, and Sven Laumer<sup>1</sup>

<sup>1</sup> FAU Erlangen-Nürnberg, Institute of Information Systems, Nuremberg, Germany  
{eva.gengler,sven.laumer}@fau.de

<sup>2</sup> TU Berlin, Chair for Work, Technology and Participation, Berlin, Germany  
{marco.wedel}@tu-berlin.de

<sup>3</sup> FemAI – Center for Feminist Artificial Intelligence GmbH, Neunkirchen a.B., Germany  
{eva.gengler,alexandra.wudel}@fem-ai.com

**Abstract.** Recent advances in Artificial Intelligence have brought novel opportunities for businesses, societies, and individuals alike, yet they also raise complex questions on inequitable power distribution. We see contemporary AI systems, that reinforce power imbalances and disadvantage marginalized, underrepresented, and underprivileged people. Current approaches to advancing AI, such as Ethical, Fair, or Trustworthy AI, have not included the effects of power in their considerations. As feminism has a long history of doing so, we introduce an intersectional and inclusive feminist approach to shape AI in a more equitable way. We approach this by building on recent Information Systems and interdisciplinary research as well as on evidence from expert interviews in focus groups, which we conducted in 2022 and 2023. Our study reveals that utilizing the feminist approach could be effective firstly, to shape AI systems and secondly, to change prevailing power structures in societal systems to become more equitable.

**Keywords:** *artificial intelligence, power imbalances, feminist philosophy of technology, socio-technical-systems, beyond AI Ethics*

## 1 Introduction

In our societies we have a long history of an inequitable distribution of power. Recently, Artificial Intelligence (AI) has become power, too, which reinforces traditional power structures (Kim, 2021; Zajko, 2022). Therefore, marginalized, underrepresented, and underprivileged people (MUUP) frequently do not benefit from AI, but are disadvantaged even further. This needs to be recognized, challenged, and changed. Consequently, complex questions around accountability, discrimination, and bias have been raised (Berente et al., 2021) and it becomes necessary to reflect upon the interdependencies between societal and AI systems.

Our research is motivated by numerous cases of biased and discriminatory AI (Feuerriegel et al., 2020). This includes applications in hiring (Jago & Laurin, 2022), lending (Bertrand & Weill, 2021), and recidivism prediction (Angwin et al., 2016; Chouldechova, 2017; Malek, 2022). Certain groups of people have been structurally disadvantaged including women, people of color, and gender queer people (Buolamwini & Gebru, 2018; Chouldechova, 2017; Jethwani et al., 2022). These and other groups face numerous disadvantages in our societies, and recently these have been mirrored by AI systems.

Structural power imbalances influence the outcomes of AI, further perpetuating the inequities of the analog world. Shifting existing power relations in favor of creating an inclusive society, which is increasingly influenced by AI, becomes a twofold challenge: One must continuously improve societal systems of the analog world, while simultaneously shaping a nexus of AI systems that does not recreate the deeply rooted structural deficiencies of the modern era of 20<sup>th</sup> century societies or counteract the positive achievements reached to this day.

One approach to improve societal systems, by identifying i.e., the patriarchy as a focal point of intervention to overcome power imbalances, oppression, and exploitation in modern societies, is feminism (Allen, 1998; Oksala, 2017). For this paper we define patriarchy as a system of social structures, and practices which oppresses, dominates, and exploits MUUP (Walby, 1989). A society is patriarchal if it promotes male privilege “by being male-dominated, male-identified, male-centered, and control-obsessed” (Johnson, 2007, p. 5). Power (to control others) is at its core. In that, the (mis)use of power, the patriarchy is harmful to women and men alike (Johnson, 2007). It is these structures that are reproduced in today's AI systems (Siapka, 2022; Weinberg, 2020).

Feminism seeks to challenge and change societal systems towards equality, freedom, and justice for all e.g., through intersectionality, interdisciplinarity, and evidence-based activism (Ferguson, 2017). It includes aspects of diversity such as gender, age, religion and belief, disability, sexual identity, ethnicity, and appearance (Ferguson, 2017). While it is beyond the scope of this paper to describe the many faces and various facets of feminism, its overall impact can be regarded as one of the most influential contributions to the creation of more inclusive and equitable societal systems in the post-modern era (Carbado et al., 2013; Crenshaw, 1989, 1991). In this study, we decisively incorporate the intersectional and inclusive feminist approach. Based on that deduction and considering the twofold challenge described above, a question arises: Can the success of the intersectional feminist approach in shaping societal systems be translated into the world of AI to model a nexus of AI systems that does not recreate the structural deficiencies and power imbalances of the patriarchy still and predominantly enshrined in the analog world? Following this idea, the overall assumption of our research is that by building an effective feminist approach for AI, it can a) diminish the recreation of patriarchal power distribution in AI systems and b) reduce the negative influence of such systems on the analogue world and its societal systems.

This research in progress constitutes first theoretical principles and results, by asking:

**RQ:** *Can the feminist approach shaping societal systems in the analog world be utilized to shape AI systems?*

Berente et al. (2021) call for less siloed research on AI and to realize this through the boundary-spanning and interdisciplinary nature of Information Systems (IS). Consequently, this research in progress builds on existing IS research (Berente et al., 2021; Chouldechova, 2017; Feuerriegel et al., 2020; Mikalef et al., 2022; Mikalef & Gupta, 2021), and on research concerning societal approaches to governing and designing AI in a social way (Riedl, 2019; Shneiderman, 2020), computer science (Floridi, 2019; Hagendorff, 2020; Mittelstadt, 2019) as well as social science research (Arrieta et al., 2020; Carbado et al., 2013; Johnson, 2007; Wellner & Rothman, 2020). It extends the discussion through an empirical study (expert interviews in focus groups) on the benefits of a feminist approach to shaping AI. We will illustrate our preliminary findings of how this can challenge and change prevailing power structures (PPS) and thus, enable true participation of and equity for MUUP.

The paper is structured as follows: We motivate our research question in section 2 by providing an overview of current societal approaches to AI and the intersectional feminist approach. We then illustrate our research method in section 3 and present our preliminary results in section 4. Our paper concludes with the outlook in Section 5.

## 2 Theoretical Background

### 2.1 Societal Approaches to Shape AI

Due to the numerous fields in which AI is utilized, the importance of an impact assessment of AI systems in research, politics, and business is imperative. When analyzing the current discussions on AI's impact on societies and humanity, a variety of different debates emerge: e.g., Ethical AI (Hagendorff, 2020; Mittelstadt, 2019), Fair AI (Binns, 2017; Feuerriegel et al., 2020), Trustworthy AI (European Commission, 2019; Floridi, 2019), Responsible AI (Arrieta et al., 2020; Mikalef et al., 2022; Peters et al., 2020), or Human-Centered AI (Riedl, 2019; Shneiderman, 2020).

These concepts generally reference a value-based approach, which in itself is subject to controversial debates on how to create universal ethical norms, principles and frameworks alongside local, culturally variable identities and traditions (Ess, 2020). However, they do not take into account the socio-structural practices and social relations of the underlying societies, groups, and individuals that shape, interact or are confronted by AI (Weinberg, 2020). Socio-technical, human-centered AI debates, such as Ethical AI, do in their normative nature neither question nor take into account the status quo of underlying societal systems and therefore risk translating undesired aspects of their social constitutionality and habitude into the digital world (Hagendorff, 2020; Mittelstadt, 2019; Siapka, 2022). Catherine D'Ignazio and Lauren F. Klein (2020) elaborate: They present the widely used concepts *ethics*, *bias*, *fairness*, *accountability*, *transparency*, and the *understanding of algorithms* as concepts that secure power, “[b]ecause they locate the source of the problem in individual or technical systems” (D’Ignazio & Klein, 2020, p. 60).

## 2.2 The Intersectional and Inclusive Feminist Approach

Existing societal approaches to AI currently do not address the phenomenon of mirrored patriarchal structures and practices in AI models and systems (Michelfelder et al., 2017; Wellner & Rothman, 2020). Feminism contextualizes prevailing power imbalances that currently tend to foster discrimination from the non-digital into the digital world (Weinberg, 2020). Thus, we introduce a feminist approach.

Feminism has a long history of challenging and changing our societies towards more equity and justice. Over the centuries, feminism has evolved. Today there is not one understanding of feminism, but a plurality of feminisms. Feminism has its origin in fighting for equal rights of women and men, however, one of the dominant contemporary directions of feminism is inclusive and intersectional (Ferguson, 2017). This feminism envisions equity for all people, while focusing on power imbalances and MUUP (Ferguson, 2017). This is the concept we use as *feminism* in this paper.

Intersectionality, as one of feminism's greatest achievements, is an open and relentlessly critical approach to power (Ferguson, 2017). It grasps the complex nature of the patriarchal functioning in our societies that impacts people in diverse ways. Intersectionality incorporates the variety of MUUP: e.g., gender, age, religion and belief, disability, sexual identity, ethnicity, and appearance (Ferguson, 2017). Thus, it demonstrates how individuals and groups can face discrimination through more than one attribute. This overlap of more than one disadvantaged feature, potentially results in extensive discrimination (Crenshaw, 1989). E.g., Black women are misclassified by face recognition systems to a high extent, because they are women and at the same time Black (Buolamwini & Gebru, 2018; Perkowitz, 2021). Thus, they face far stronger disadvantages than both Black men and White women (Buolamwini & Gebru, 2018). As AI reflects traits that are devalued in the patriarchal system, e.g., female people of color particularly suffer from discrimination of AI systems (Buolamwini & Gebru, 2018; Crenshaw, 1989). Therefore, MUUP must be at the core, when deciding upon and building AI. Intersectional and inclusive feminism can be used to analyze and change AI's functioning on the AI system level, as well as on the societal system level.

## 3 Research Method

The sketched discourses and considerations mentioned above are the guiding deductions for a working hypothesis that has been the starting point for this research in progress. The initial reflections upon the theme have been derived by way of qualitative deliberations informed by past research projects, literature review (Webster & Watson, 2002), and horizon scanning (Cuhls, 2019). The resulting working hypothesis and the research questions have recently been – throughout 2022 and 2023 – being tested empirically by way of monthly online roundtables serving as expert focus groups. This participatory form of data collection, based on semi-structured interviews, stimulates discussions between different groups of people and diverse opinions. It is a moderated discourse procedure in which a small group of individuals is inspired to discuss a specific topic based on an input (Kamberelis & Dimitriadis, 2013; Morgan, 1996; Schulz et al., 2012).

We have deliberately emphasized diversity in the composition of participants in terms of MUUP, multi-dimensionality, trans-disciplinarity, and intersectionality, to bring in and discuss different perspectives (Sundermeier et al., 2020). The experts' lenses have been shaped through professional occupations in numerous domains, e.g.: Human Rights, IS, Political Science, and Policy. The pool of experts has been established as a research effort over the last three years through professional networks and by way of individual recruitment (e.g., conferences, publications, social media, research projects). It currently consists of about one hundred individuals (and counting).

Eight English speaking focus groups sessions were conducted between July 2022 and March 2023. The group size averaged ten people. The sessions were scheduled for 1h 15min and conducted in an online video call format. Participants joined the sessions from around the globe.

We used an iterative method, creating a working hypothesis or a sketched model based on literature, past research, and existing gaps, challenging it within the focus group, consolidating the outcomes afterwards, and integrating the intermediate results into the next meeting. In practical terms, this means, that e.g., a definition of feminism has been proposed to the group in order to start a discussion on the existence, benefits, and limits of varying feminist approaches. The process of each session was facilitated by one of the authors of this paper who initiated and moderated each discussion. In line with the focus group method, the moderator did not engage actively in the discussion but merely ensured an inclusive, respectful, and on-topic debate. The sessions were either recorded by video (if all participants agreed) or by minutes. The recordings were transcribed and coded in an iterative and explorative way until concepts emerged (Corbin & Strauss, 2014).

## 4 Preliminary Results

Based on the research process so far, the answer to RQ is: Yes, it is likely that the feminist approach can be utilized to shape AI systems, e.g.: *“Feminism [...] can help to be a foundation with which to analyze and address the threat and potential of AI.”* We identified power as an important concept in shaping AI systems, as participants repeatedly mentioned the perpetuation of current patriarchal power structures, e.g.: *“Again the people on top continue to get power, because they already have the power.”* Moreover, rather than focusing only on technical and individual influences on AI as emphasized by e.g., the concept of Fair AI (Feuerriegel et al., 2020), participants argued for a broader context with regard to the people behind its data, design, and context, e.g.: *“We need to encourage more marginalized people and minorities to access this field.”* Thus, the three concepts, *data & design*, *people*, and *context* evolved.

Consequently, we introduce our framework depicted in Figure 1. It illustrates the interdependence of the societal (S1) and AI system (S2): The PPS of S1 influence S2. The intersection of S1 and S2 is shaped by three conceptual areas where the social and analogue is translated into the artificial and digital. As mentioned above these are currently subsumed under the following designators 1) data and design of AI, e.g., the input data, 2) people, e.g., the diversity of people designing AI, and 3) context, e.g., the



## References

- Allen, A. (1998). Rethinking Power. *Hypatia*, 13(1), 21–40. JSTOR. <http://www.jstor.org/stable/3810605>
- Angwin, J., Larson, J., Surya, M., & Lauren, K. (2016). *Machine Bias*. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 1433–1450. <https://doi.org/10.25300/MISQ/2021/16274>
- Bertrand, J., & Weill, L. (2021). Do algorithms discriminate against African Americans in lending? *Economic Modelling*, 104, 105619. <https://doi.org/10.1016/j.econmod.2021.105619>
- Binns, R. (2017). Fairness in Machine Learning: Lessons from Political Philosophy. *Proceedings of Machine Learning Research* 81, 1–11. <http://arxiv.org/abs/1712.03586>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Carbado, D., Devon, Crenshaw, K., Mays, V., & Tomlinson, B. (2013). Intersectionality: Mapping the Movements of a Theory. *Du Bois Review: Social Science Research on Race*, 10, 303–312. <https://doi.org/10.1017/S1742058X13000349>
- Chouldechova, A. (2017). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
- Corbin, J. M., & Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (4th ed.). SAGE Publications, Inc.
- 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*, 1989(1), 139–167.
- Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, 43(6), 1241–1299. <https://www.jstor.org/stable/1229039>
- Cuhls, K. E. (2019). Horizon Scanning in Foresight – Why Horizon Scanning is only a part of the game. *Futures & Foresight Science*, 2(1), 1–21. <https://doi.org/10.1002/ffo2.23>
- D’Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press.
- Ess, C. M. (2020). Interpretative Pros Hen Pluralism: From Computer-Mediated Colonization to a Pluralistic Intercultural Digital Ethics. *Philosophy & Technology*, 33(4), 551–569. <https://doi.org/10.1007/s13347-020-00412-9>
- European Commission. (2019). *Ethics Guidelines for Trustworthy AI* (pp. 1–39). <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>
- Ferguson, K. E. (2017). Feminist Theory Today. *Annual Review of Political Science*, 20(1), 269–286. <https://doi.org/10.1146/annurev-polisci-052715-111648>
- Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI. *Business & Information Systems Engineering*, 62(4), 379–384. <https://doi.org/10.1007/s12599-020-00650-3>



- Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1(6), 261–262. <https://doi.org/10.1038/s42256-019-0055-y>
- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30, 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Jago, A. S., & Laurin, K. (2022). Assumptions About Algorithms' Capacity for Discrimination. *Personality and Social Psychology Bulletin*, 48(4), 582–595. <https://doi.org/10.1177/01461672211016187>
- Jethwani, H., Subramonian, A., Agnew, W., Bleile, M., Arora, S., Ryskina, M., & Xiong, J. (2022). Queer in AI. *XRDS: Crossroads, The ACM Magazine for Students*, 28(4), 18–21. <https://doi.org/10.1145/3538543>
- Johnson, A. G. (2007). *The Gender Knot*. Pearson Education India.
- Kamberelis, G., & Dimitriadis, G. (2013). *Focus Groups: From structured interviews to collective conversations*. Routledge. <https://doi.org/10.4324/9780203590447>
- Kim, P. (2021). AI and Inequality. In *Forthcoming in The Cambridge Handbook on Artificial Intelligence & the Law*, Kristin Johnson & Carla Reyes, eds. (2022), Washington University in St. Louis Legal Studies Research Paper No. 21-09-03. Available at SSRN: <https://ssrn.com/abstract=3938578>
- Malek, Md. A. (2022). Criminal courts' artificial intelligence: The way it reinforces bias and discrimination. *AI and Ethics*, 2(1), 233–245. <https://doi.org/10.1007/s43681-022-00137-9>
- Michelfelder, D. P., Wellner, G., & Wiltse, H. (2017). Designing differently: Toward a methodology for an ethics of feminist technology design. In *The Ethics of Technology: Methods and Approaches* (pp. 193–218). Rowman & Littlefield International.
- Mikalef, P., Conboy, K., Lundström, J. E., & Popović, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, 31(3), 257–268. <https://doi.org/10.1080/0960085X.2022.2026621>
- Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3). <https://doi.org/10.1016/j.im.2021.103434>
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Morgan, D. L. (1996). Focus Groups. *Annual Review of Sociology*, 22(1), 129–152. <https://doi.org/10.1146/annurev.soc.22.1.129>
- Niederberger, M., & Renn, O. (Eds.). (2019). *Delphi-Verfahren in den Sozial- und Gesundheitswissenschaften: Konzept, Varianten und Anwendungsbeispiele*. Springer VS.
- Oksala, J. (2017). Feminism and Power. In *The Routledge Companion to Feminist Philosophy*. Routledge.
- Perkowitz, S. (2021). The Bias in the Machine: Facial Recognition Technology and Racial Disparities. *MIT Case Studies in Social and Ethical Responsibilities of Computing*. <https://doi.org/10.21428/2c646de5.62272586>
- Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—Two Frameworks for Ethical Design Practice. *IEEE Transactions on Technology and Society*, 1(1), 34–47. <https://doi.org/10.1109/TTS.2020.2974991>
- Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies*, 1(1), 33–36. <https://doi.org/10.1002/hbe2.117>

- Schulz, M., Mack, B., & Renn, O. (Eds.). (2012). *Fokusgruppen in der empirischen Sozialwissenschaft*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-19397-7>
- Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Siapka, A. (2022). Towards a Feminist Metaethics of AI. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 665–674. <https://doi.org/10.1145/3514094.3534197>
- Sundermeier, J., Birkner, S., Ettl, K., Kensbock, J., & Tegtmeier, S. (2020). Hello Diversity! Opportunities and Challenges of Entrepreneurial Diversity in the Digital Age. *Communications of the Association for Information Systems*, 47(1). <https://doi.org/10.17705/1CAIS.04732>
- Walby, S. (1989). Theorising Patriarchy. *Sociology*, 23(2), 213–234. <https://doi.org/10.1177/0038038589023002004>
- Webster, J., & Watson, R. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26. <https://doi.org/10.2307/4132319>
- Weinberg, L. (2020). Feminist Research Ethics and Student Privacy in the Age of AI. *Catalyst: Feminism, Theory, Technoscience*, 6(2), Article 2. <https://doi.org/10.28968/cft.v6i2.32943>
- Wellner, G., & Rothman, T. (2020). Feminist AI: Can We Expect Our AI Systems to Become Feminist? *Philosophy & Technology*, 33(2), 191–205. <https://doi.org/10.1007/s13347-019-00352-z>
- Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, 16(3). <https://doi.org/10.1111/soc4.12962>