

Terms-we-Serve-with: A critical feminist intervention in coming to terms with algorithmic systems

Bogdana Rakova
bogdana@mozillafoundation.org
Senior Trustworthy AI Fellow, Mozilla Foundation

Abstract

In working to expand our vocabulary and capacity to transform algorithmic harms and injustice into positive human experience, it is critical to consider the contractual agreements between people and technology companies. For example, Terms-of-Service (ToS) agreements, cookie policies, content moderation policies, privacy policies, limitations of liability clauses, non-disclosure agreements, and other kinds of user agreements. The small print and legalese in contractual agreements often fail to provide people with meaningful consent and contestability in cases of algorithmic failures, risks, harms, or injustice (Vincent, 2021; Fiesler, Beard, and Keegan, 2020; Vaccaro et al., 2015). Instead of contributing to extractive power structures and information asymmetries, what if contractual agreements could become a living socio-technical artifact that empowers trust through equitable community-driven participation and oversight in the lifecycle algorithmic systems? This question is at the core of the Terms-we-Serve-with (TwSw) framework for evolving and enacting social norms and agreements between people and companies developing AI-driven products and services. The goal of the TwSw intervention proposal is achieving improved transparency and human agency in AI beyond debiasing, explainability, and ethics. Furthermore, it inspires a new legal and regulatory approach to

accountability based on a taxonomy of sociotechnical harms and risks (Shelby et al. 2022). We hope that many critical feminist interventions will emerge during engaging with the TwSw framework and will provide meaningful steps towards centering work around the lived experiences of members of communities affected by complex algorithmic systems.

Who are the people engaged along the entire lifecycle of design, development, and deployment of AI, from the material resource extraction to the value that is created and experienced by people? What frictions exist and need to be made visible? How could critical interventions in AI rather than algorithmic predictions made by AI contribute to improved algorithmic justice outcomes? We argue that there is a need to center algorithmic repair, critical theory, critical design, and design justice, in facilitating a roundtable discussion on what critical feminist interventions could meaningfully address these questions. Furthermore, new kinds of accountability mechanisms could empower answering algorithmic justice questions in ways that lead to real world outcomes. Based on these goals, the TwSw is a feminist-inspired intervention for improving the contractual agreements between people and technology companies. It is a roundtable engagement framework and tools for enacting social norms and agreements centered on five dimensions. These are: (1) co-constitution - participatory development and co-design of user agreements (DiSalvo, Clement, and Pipek, 2012; Hagan, 2020), (2) addressing friction (Raji et al., 2022; Costanza-Chock, 2020), (3) informed refusal (Cifor et al., 2019; Benjamin, 2016), (4) complaint and contestability which enables people to report potential concerns (Ahmed, 2021; Kaminski and Urban, 2021), and (5) disclosure-centered mediation when harm occurs (Davis, Williams, and Yang, 2021; Costanza-Chock, Raji, & Buolamwini, 2022).

Case Study: Large language models and gender-based violence prevention

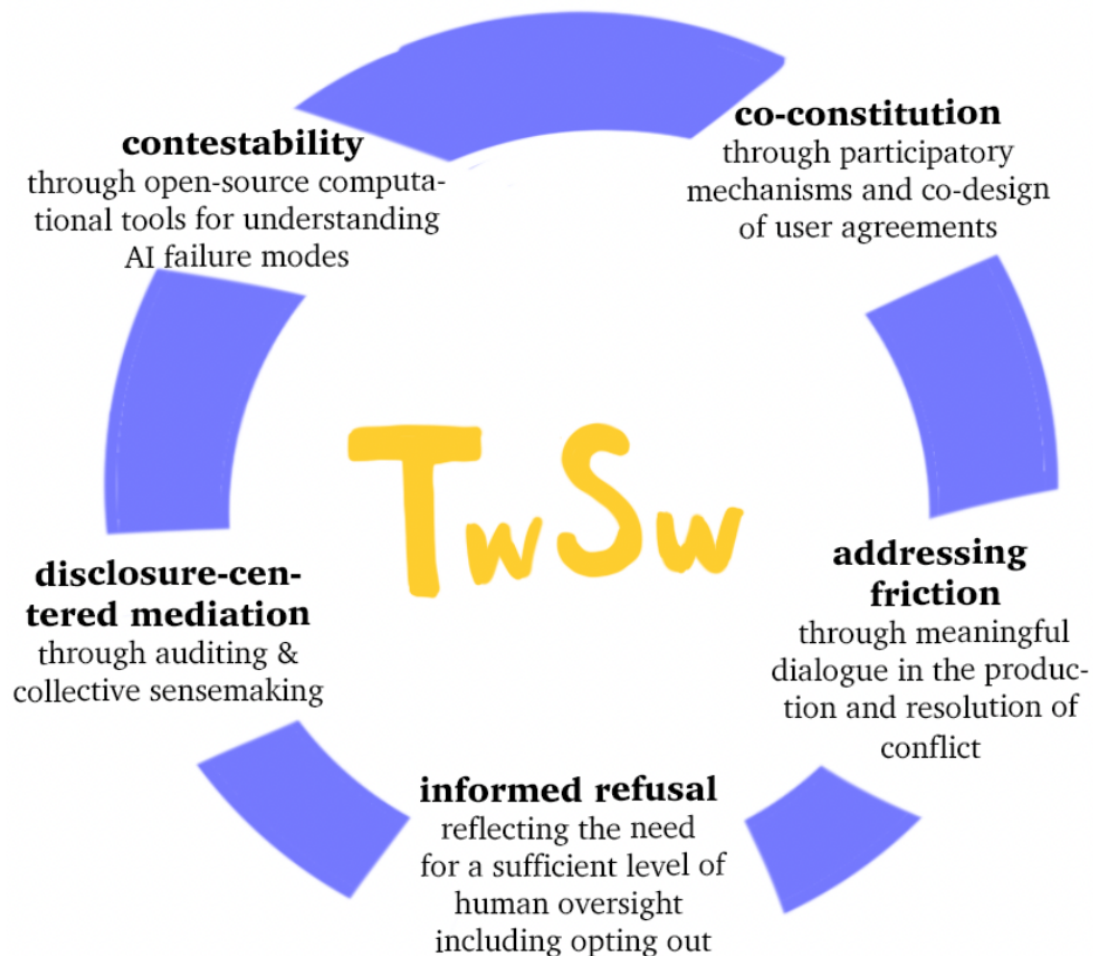
We inform our analysis of critical feminist interventions through a collaboration with a South African gender-based violence reporting AI startup - “Kwanele - Bringing Women Justice” (Kanwele, 2022). Kwanele’s team wanted to engage with the TwSw framework in determining ways to incorporate AI in a manner that aligns with their mission, values, and the needs of the communities they serve. Kwanele aims to help women and children report and prosecute crimes involving gender-based violence (GBV). The team is developing an AI chatbot leveraging the latest advancement in large language models to guide users in reporting GBV cases and answer any questions related to South Africa's legislation. Recognizing the broader social context, Kwanele sees the chatbot as embodying three roles: (1) a legal analyst, helping make the legalese within government regulations easier to understand; (2) a crisis response social worker, guiding people to report GBV and seek help; and (3) a mental health therapist, conversing with victims in a psychologically and potentially physically vulnerable state.

Working with the TwSw, Kwanele derived practical strategies for using AI centering the needs of the vulnerable communities they serve. During a participatory roundtable discussion using the TwSw framework, we discussed interventions that fell into the following themes: (1) improving communication and engagement in user agreements; (2) clear pathways for escalation of algorithmic injustice; (3) a complaint handling process-based approach that encompasses - confirmation, recognition, acknowledgement, and follow up with impacted users; (4) compassion-centered approach to the user interface through which people interact with AI; and (5) improved feedback loops between product teams and frontline workers who process user

reports of algorithmic injustice. The critical feminist interventions that emerged during this workshop are a step towards centering work around the lived experiences of members of communities affected by AI chatbot systems.

At its heart, the socio-technical TwSw intervention leverages a relational approach to algorithmic justice centered on the need for a radical restructuring of the “take-it-or-leave-it” ToS agreement. We hope that it will lead to real world systems change by empowering new social imaginaries in how we come to terms with algorithms.

Table 1. Terms-we-Serve-with (TwSw) socio-technical framework dimensions for evolving and enacting social norms and agreements in AI



Acknowledgement

A related paper in collaboration with Megan Ma and Renee Shelby is currently in peer review.

References

Ahmed S (2021) *Complaint!* Duke University Press.

Benjamin R (2016) Informed refusal: Toward a justice-based bioethics. *Science, Technology, & Human Values* 41(6): 967-990.

- Bergram K, Bezençon V, Maingot, P, Gjerlufsen T and Holzer A (2020) Digital nudges for privacy awareness: From consent to informed consent? Proceedings of the 28th European Conference on Information Systems (ECIS2020).
- Cifor M, Garcia P, Cowan TL, Rault J, Sutherland T, Chan A, Rode J, Hoffmann AL, Salehi N and Nakamura L (2019) Feminist Data Manifest-No. Available at: <https://www.manifestno.com/> (accessed 30 April, 2023).
- Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. The MIT Press.
- Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022, June). Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 1571-1583).
- DiSalvo C, Clement A and Pipek V (2012) Participatory design for, with, and by communities. In: Simonsen J and Robertson T (eds.) *The Routledge International Handbook of Participatory Design*. Routledge, pp.182-209.
- Davis JL, Williams A, and Yang MW (2021) Algorithmic reparation. *Big Data & Society* 8(2).
- Fiesler C, Lampe C and Bruckman AS (2016) Reality and perception of copyright terms of service for online content creation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1450-1461).
- Fiesler C, Morrison S and Bruckman AS (2016) An archive of their own: A case study of feminist HCI and values in design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2574-2585).

- Fiesler C, Beard N and Keegan, BC (2020) No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 187-196).
- Hagan M (2020) Legal design as a thing: A theory of change and a set of methods to craft a human-centered legal system. *Design Issues* 36(3): 3-15.
- Kaminski, M. E., & Urban, J. M. (2021). The right to contest AI. *Columbia Law Review*, 121(7), 1957-2048.
- Kanwele. (n.d.). Home. Available at: <https://kwanelesouthafrica.org/>. (accessed 21 February 2023).
- Raji ID, Kumar IE, Horowitz A and Selbst A (2022) The fallacy of AI functionality. In *Proceedings of 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 959-972).
- Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., ... & VIRK, G. (2022). Identifying Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. arXiv preprint arXiv:2210.05791.
- Vaccaro K, Karahalios K, Sandvig C, Hamilton K and Langbort C (2015) Agree or cancel? Research and terms of service compliance. In *2015 CSCW Workshop on Ethics for Studying Sociotechnical Systems in a Big Data World*.
- Vincent J (2021) Facebook bans academics who researched ad transparency and misinformation on Facebook. *The Verge*, 4 August. Available at: <https://www.theverge.com/2021/8/4/22609020/facebook-bans-academic-researchers-ad-transparency-misinformation-nyu-ad-observatory-plugin>. (accessed 21 February 2023).