

The Inherent Politics of AI as a Matter of Justice

Laila Wegner

Eindhoven University of Technology
l.wegner@tue.nl

Introduction

Within the research field of algorithmic fairness, two dominant subfields stand out: ‘formal algorithmic fairness’, characterized by a local focus and technocentric reforms (Green, 2022), and ‘structural accounts’. Structural, here, captures the focus on the socio-technical structures of AI, as well as on structural injustices (e.g., Green, 2022; Kasirzadeh, 2022; Lin & Chen, 2022). Although both subfields address AI-related injustices, they tend to position themselves as opposing fields. I argue that this creates a conceptual gap and therefore, neither formal nor structural approaches, alone or together, provide a sufficient account to address AI-related injustices. To bridge this gap, I propose a third perspective, inspired by the inherent politics of technology (Winner, 1980), and provide an illustration based on predictive AI.

In essence, formal and structural approaches to algorithmic fairness differ in the locus of intervention, the level of abstraction, and their methods. Formal approaches tend to treat algorithms as technical artifacts and focus on mathematical bias mitigation. They work at a concrete, actionable level, grounded in legal frameworks and intervene through technocentric reform strategies. Structural approaches, by contrast, situate algorithms within broader socio-technical contexts, drawing on social science, political philosophy, and critical theory to address more complex and structural concerns of justice. Interventions here aim at wider societal reforms rather than technical adjustments. Furthermore, structural approaches tend to dismiss formal algorithmic fairness as overly simplistic, thereby failing to address structural injustices and creating illusory solutions susceptible to abuse (e.g., Lin and Chen 2022). However, while structural approaches excel at identifying limitations of formal approaches, they tend to lose nuance about the concrete role of the algorithms. For example, Kasirzadeh (2022) illustrates that a distributive fairness metric applied to a loan algorithm does not remedy the underlying injustices that caused the

dependence on a loan in the first place. This is certainly true. However, whether the loan was approved by a formally ‘fair’ algorithm or by a human would not make a meaningful difference. Thus, the system's nature (sociotechnical or not) does not necessarily change the evaluation. By discussing AI more broadly, structural approaches risk bypassing the question of whether, and how, AI-related injustices are distinctive from structural injustices in society.

To examine the interplay of the technological details as well as the structural complexities, I propose investigating the inherent politics of AI. Winner (1980) emphasized that inherently embedded structures of power and authority can only be identified based on a close investigation of the technology itself. Furthermore, he introduces two ways in which technology is political. The first and widely cited case relates to politically shaped technologies, capturing technology designed to reflect specific social or political attitudes. However, more important for my purposes is the second case, in which Winner (1980) argues that certain technologies in themselves have political qualities. This applies to technologies that require, or at least are strongly compatible with, particular social and political structures, rendering these politics inherent to the technology in question. The requirement of specific politics (e.g., safety measures and surveillance) makes them constitutive of the technical system (e.g., nuclear energy) and is thus inherent to the technology in question. While politically shaped technologies become political due to the particular way they were designed, inherently political technologies cannot be redesigned to suit different political values or social contexts.

As will become important in my subsequent analysis, Winner (1980) further specifies the origin of the inherent politics and highlights that “there is a further distinction to be made between conditions that are internal to the workings of a given technical system and those that are external to it” (p. 130, emphasis added). For AI, inherent but externally caused characteristics encompass the exploitative social and

political environment on which AI is built, such as the vast amount of natural resources, data, and massive human labor. Being of great importance, these issues are well captured by structural approaches to algorithmic fairness (e.g., Bolte and van Wynsberghe 2024; Crawford 2021). Yet, a discussion of the inherent and internal politics of AI has been largely overlooked. These politics constitute specific AI techniques, such that altering them would fundamentally change AI technologies as we know them. One example of this feature is AI's reliance on machine learning to extract common patterns. Thus, to reveal the inherent and internal politics of AI, it is essential to have a solid understanding of the technology in question. This makes it inevitable to focus on specific AI techniques rather than AI as a sociotechnical system.

To apply the inherent politics of technology to the field of algorithmic fairness and justice, some conceptual steps are necessary. As criticized by Liao and Huebner (2021), Winner does not sufficiently account for the relation of politics and (in)justice. They argue that Winner „fails to address the interaction between material artifacts and oppressive systems such as racism“ (Liao and Huebner, 2021, p. 100). Addressing this gap, they develop the concept of ‘oppressive things’ in which unfair social dynamics manifest in physical artefacts. In the context of AI, Lin and Huang (2026) recently argued that generative AI becomes an oppressive thing by manifesting oppressive normality in its content, performance, and style. This insightful contribution starts from a premise similar to mine. Yet, the identified qualities are not inherent and internal as outlined above. These features do not necessarily follow from the technical qualities of AI, and with sufficient effort and political will, they could theoretically be changed while using the same technical foundations. Furthermore, there might well be more subtle politics inherent to AI which do not yet warrant the conclusion that it constitutes oppression – though this doesn't render them trivial either.

I make a more modest claim and extend Winner's approach by conceptualizing the inherent politics of technology as ‘justice-sensitive’. While these politics are not necessarily (un)just in themselves, they may become so in the interaction with a social context where they can be evaluated against specific standards of justice. These politics can be *either* positive *or* negative, as opposed to politics that are *inherently* positive or negative, such as oppressive politics. Thus, conceptualizing inherent politics as justice-sensitive is, on the one hand, a less strong claim than identifying oppressive things, as it is agnostic to different justice theories and more descriptive about the politics itself. However, on the other hand, it is a more foundational claim, as the identified politics are not flexible in their design and thus convey grounding limitations to approaches to enhance algorithmic fairness. If these politics are rejected on the grounds of justice, no design intervention could change this premise. This would have subsequent implications for formal algorithmic

fairness, as well as for broader approaches such as value-sensitive design and calls to democratize AI. Therefore, exploring the inherent, justice-sensitive politics of AI provides a conceptual gain by identifying underexplored ways in which AI contributes to (in)justice, as well as a practical gain by devising better interventions based on the specific impact of the AI technique.

To illustrate the outlined account, I focus on predictive AI, which is widely used by governments and in the private sector to predict human behavior, for example, in welfare support and crime sentencing. At the present moment of research, I analyze the inherent, internally caused politics of predictive AI and its justice-sensitivity. For example, by the extraction of patterns, predictive AI optimizes for similarity by recommending similar candidates to past ones. In some cases, the focus on similarity as a decision factor might be morally well-justified. For example, in court, similar cases serve as precedents against which new cases are assessed to determine whether a similar sentence is warranted. Conditional on the absence of reinforced structural biases, such similarity may provide a reasonable heuristic for ensuring equal treatment. However, in the hiring process, the focus on similarity to former employees might reproduce privileges and stereotypes, becoming an obstacle to diversity. The focus on similarity here rests on a normative assumption that it is desirable to conform to the majority. Focusing on the inherent politics of predictive AI brings these considerations to the forefront by asking whether these politics provide a reasonable rationale in the respective case.

References

- Bolte, L., & van Wynsberghe, A. 2025. Sustainable AI and the third wave of AI ethics: A structural turn. *AI Ethics*, 5, 1733–1742. <https://doi.org/10.1007/s43681-024-00522-6>
- Crawford, K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Green, B. 2022. Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness. *Philosophy & Technology*, 35(4). <https://doi.org/10.1007/s13347-022-00584-6>
- Kasirzadeh, A. 2022. Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy. In *AIES '22: Proceedings of the AAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery (ACM) (pp. 348–356). <https://doi.org/10.1145/3514094.3534188>
- Liao, S., & Huebner, B. 2021. Oppressive Things*. *Philosophy and Phenomenological Research*, 103(1), 92–113. <https://doi.org/10.1111/phpr.12701>
- Lin, T., & Huang, L. T.-L. 2026. AI, Normality, and Oppressive Things. *Minds and Machines*, 36(2), 26. <https://doi.org/10.1007/s11023-026-09778-9>
- Lin, T.-A., & Chen, P.-H. C. 2022. Artificial Intelligence in a Structurally Unjust Society. *Feminist Philosophy Quarterly*, 8(3/4). <https://doi.org/https://doi.org/10.5206/fpq/2022.3/4.14191>
- Winner, L. 1980. Do Artifacts Have Politics? *Daedalus, Modern Technology: Problem or Opportunity?*, 109(1), 121–136.