

AI, Normality, and Oppressive Things (Extended Abstract)

Ting-an Lin¹, Linus Ta-Lun Huang²

¹University of Connecticut

²The Chinese University of Hong Kong

tinganlin@uconn.edu, linushuang@cuhk.edu.hk

Extended Abstract

Recent research has revealed that AI systems might bring about unfair social impacts by influencing social schemas, namely the understanding and attitudes people hold toward different social groups. Nonetheless, much attention has been paid to instances where the content presented by AI systems explicitly demeans marginalized groups or reinforces problematic stereotypes. Drawing from philosophical discussions on the politics of artifacts, this paper focuses on instances that shape oppressive systems through subtler manners—such as through the accents and communication styles materialized in AI systems—and argues that their subtlety does not make them less morally concerning. By integrating insights from philosophy and machine learning, this paper contributes a novel framework to examine AI fairness.

To that end, we utilize what Liao and Hueber (2021) call *oppressive things*—physical components that are in congruence with existing oppressive systems—to analyze AI systems. Existing discussions of *oppressive things* have been primarily focused on instances of material artifacts (Spurrett, 2024), and our analysis suggests that this notion can be expanded to account for instances of digital artifacts. Overall, building upon existing AI scholarship that raised concerns regarding various forms of algorithmic injustice, this paper aims to enrich the efforts of pursuing digital justice by developing a philosophical account of how digital artifacts can be not only biased but also oppressive.

Analysis proceeds as follows. In Section 2, to set the ground for the analysis, we briefly introduce the notion of *oppressive things* and situate it in literature on the political nature of artifacts. Section 3 shows that many AI systems meet the definition of *oppressive things* and should be recognized as such. We begin by categorizing three ways AI systems may manifest *oppressive normality*—through content, performance, and style. By fleshing out the double-bind nature of oppression (Frye, 1983), we further argue that

AI systems could be oppressive even if they can benefit the oppressed groups. Furthermore, we identify three contributing factors that make the oppressive impacts of AI systems more concerning: the phenomenon of algorithmic monoculture, their guise of neutrality, and their growing prevalence. Our analysis advances theoretical understanding of *oppressive things* and has implications for analyzing the political impacts of a broader range of artifacts.

The final sections discuss conceptual and practical implications of our analysis. Section 4 highlights three contributions our analysis brings to the existing discourses surrounding AI ethics: It helps discern the moral differences among biased AI systems, identifying subtler cases where AI systems problematically sustain oppression. In Section 5, drawing from Fraser’s (1995) distinction between affirmative and transformative remedies to injustice, we provide some directions for practical remedies that go beyond fixing the unfair outcomes of AI but also challenge, disturb, and deconstruct the unjust power hierarchies encoded through the process of AI design, development, and deployment.

References

- Liao, Shen-Yi, and Bryce Huebner. 2021. “Oppressive Things.” *Philosophy and Phenomenological Research* 103 (1): 92–113.
- Spurrett, David. 2024. “On Hostile and Oppressive Affective Technologies.” *Topoi: An International Review of Philosophy* 43 (3): 821–32.
- Frye, Marilyn. 1983. *The Politics of Reality*. The Crossing Press Feminist Series. Freedom, CA: Crossing Press.
- Fraser, Nancy. 1995. “From Redistribution to Recognition? Dilemmas of Justice in a ‘Post-Socialist’ Age.” *New Left Review* I (212): 68–93.