

ALGORITHMIC BIAS

Gema Sánchez Medero (Complutense University of Madrid)

Definition

Algorithmic bias refers to systematic errors or distortions embedded in automated decision-making systems that produce discriminatory or inequitable outcomes. These biases are not mere technical malfunctions but structural manifestations rooted in the data used to train algorithms and the sociocultural assumptions encoded in their design. Rather than operating as neutral tools, algorithmic systems tend to reproduce and amplify pre-existing social inequalities—related to gender, age, geographic origin, or cultural capital—affecting domains such as candidate selection, deliberation processes, and political representation within organizations and institutions.

Context

Algorithmic bias has emerged as a critical concern in scholarly and policy debates surrounding the digitalization of democratic institutions, and particularly the internal governance of political parties. As parties increasingly adopt automated tools for candidate selection, membership engagement, and deliberation, the risk of embedding and amplifying structural inequalities through algorithmic design becomes a pressing analytical and normative challenge.

In the domain of candidate selection, predictive algorithms trained on historical data tend to reproduce patterns of political success that reflect past distributions of power rather than normatively desirable criteria of representation. This dynamic aligns with what scholars have termed *statistical discrimination*: the use of group-level proxies—gender, age, geographic origin, or educational background—as predictors of individual suitability, producing outcomes that systematically disadvantage members of underrepresented groups (Eubanks, 2018). The apparent neutrality of data-driven optimization obscures the normative choices embedded in the design of these systems, naturalizing exclusionary criteria as if they were objective.

In deliberative settings, the deployment of natural language processing (NLP) and sentiment analysis tools introduces additional layers of bias. As Bender et al. (2021) have extensively documented, large language models encode the sociocultural assumptions and power asymmetries present in their training corpora, rendering them far from neutral interpreters of political discourse. When such tools are used to aggregate or evaluate membership preferences, discursive registers that deviate from dominant political idioms—whether due to class, ethnicity, regional identity, or ideological heterodoxy—risk being systematically misread or devalued. This produces what Noble (2018) has described as

algorithmic oppression: the reinforcement of social hierarchies through ostensibly technical systems.

The political consequences of these dynamics are compounding. Drawing on feedback loop theory, scholars have demonstrated that when digital systems consistently render certain voices invisible, affected members progressively disengage from participatory platforms (O'Neil, 2016). This behavioral withdrawal generates datasets that are even less representative of marginalized groups, reinforcing the underlying bias in subsequent algorithmic iterations. The result is a self-reinforcing cycle of exclusion that undermines the pluralist foundations of intra-party democracy, narrowing the range of perspectives that shape collective decision-making and concentrating effective political agency among those whose profiles and discursive practices already align with algorithmic defaults.

Related Concepts

Algorithm; Algorithmic censorship; Cybercratic centralism; Digital activism; Digital engagement; Digital exclusion; Digital governance; Digital party; Intra-party democracy; Political participation.

References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishers.

Suggested citation: Sánchez Medero, Gema (2026), “Algorithmic bias”, *Dictionary on Digitalization and Democracy*, RELINK² Cost Action 23114. <https://relink2.eu/>