Algorithmic Harm, Governance, and Ethics Situating *Weapons of Math Destruction*

Munongedzi Mabhoko Clarkson University mabhokm@clarkson.edu

Introduction

Algorithms have moved from the background of data processing into the center of decision-making in areas as varied as hiring, policing, health care, and politics. Their widespread adoption raises a central question. Who benefits, who is harmed, and who is accountable when automated systems govern human lives? The tendency to make reality conform to analytical frameworks, rather than adapting frameworks to the complexity of reality, is a recurring issue across technical and policy domains. This mindset emerges when mathematical models are treated as inherently descriptive of social systems, leading analysts to reframe phenomena in terms of optimization, game theory, or efficiency metrics, even when these perspectives miss the lived dynamics at play. Organizations increasingly deploy analytics to monitor productivity through keystroke counts, call times, or biometric signals. The emphasis on measurable efficiency misses intangible aspects of work like creativity, collaboration, and trust, treating labor as data flows rather than human effort embedded in relationships. Over time, such approaches gain prestige because they signal rigor, creating intellectual cultures that reward formal abstraction over explanatory adequacy. Policy responses shaped by this outlook often double down on model refinement when outcomes disappoint, as seen in how urban housing initiatives sometimes rely on models that predict supply-and-demand balances through pricing curves, while ignoring how displacement, cultural identity, and informal economies shape housing stability. This reduces deeply social questions of community life into equations of affordability and growth instead of questioning whether different methods of inquiry are needed. The result is analysis that may appear precise but risks overlooking the very forces that drive social reality. Addressing this challenge requires methodological humility within governance frameworks, recognizing that mathematical models may provide operational utility in specific domains but cannot serve as universal instruments of explanation or prediction. A policy discourse confined to quantitative abstraction risks reifying existing inequities, since the very parameters that structure these models are determined by those in positions of institutional authority. When control over data, metrics, and optimization criteria is concentrated, algorithmic systems not only fail to capture the full spectrum of social dynamics but also perpetuate and legitimize bias under the guise of technical neutrality. Effective governance therefore calls for epistemic pluralism, incorporating qualitative, participatory, and relational approaches to ensure that complex human systems are not reduced to abstractions that primarily serve dominant interests.

Opaque, Scale, Harm

Cathy O'Neil's *Weapons of Math Destruction* (O'Neil 2016) remains one of the most influential interventions in this discussion. Although intended for a general readership, it provided a conceptual vocabulary, Weapons of Math Destruction (WMDs) that captured how predictive models reproduce inequality rather than neutralize it. This essay does not treat O'Neil's book as an isolated text but situates it within the wider body of work on governance and ethics. My aim is to show how her framing intersects with later research and to highlight why it remains central as we explore how societies should regulate algorithmic power. WMDs are characterized as systems that combine opacity, scale, and harm. These three attributes together distinguish them from models that are simply imperfect. A flawed spreadsheet might miscalculate, but a flawed predictive model applied at scale can alter access to jobs, credit, or parole in ways that reinforce systemic disadvantage.

This framing has been influential because it shifts the debate away from narrow technical accuracy toward broader social consequences. Later scholars extend this perspective by showing that discrimination is not accidental but structural. Systems that are opaque and scalable often amplify existing inequities precisely because they embed historical data patterns into present decisions (Moussawi, Modaresnedzhad, and Deng 2025). This framework is therefore less about single failures and more about feedback loops that reproduce injustice.

Justice and Recidivism

One of the most consequential areas of application is criminal justice. Predictive policing tools direct law enforcement toward certain neighborhoods, while risk assessment scores shape sentencing and parole decisions. These tools claim neutrality but draw heavily on proxies such as geography, family background, or prior arrest history that are tightly linked to racial and economic disparities. As feedback loops emerge, more patrols generate more arrests, which then confirm the system's view of an area as high risk. Later research reinforces this dynamic, noting that racial bias in AI often operates through these hidden correlations (Oliver 2025). For governance, the concern is not simply transparency but whether the structural assumptions behind these models can ever be justified. If risk assessment tools amplify inequality by design, improving accuracy does little to solve the underlying harm.

Employment and Economic Life

Employment represents another domain where algorithmic harm becomes visible. Screening software filters applicants using personality tests or past hiring data. These systems encode past biases, creating cycles where certain groups face persistent exclusion. Recruitment algorithms rarely pass ethical evaluation because they optimize for efficiency rather than equity (Chaudari 2025). When "successful" past employees become the template, minority candidates are disadvantaged from the outset. In the workplace, algorithmic monitoring not only intensifies surveillance but also erodes trust between employers and employees. From an ethical perspective, the issue is not only fairness in access to jobs but also the preservation of dignity once inside the workplace.

Politics and Democracy

Predictive analytics in politics shows how campaigns fracture the public sphere through microtargeting. Instead of messages circulating in shared spaces, appeals are delivered in private channels on platforms like TikTok, or Facebook where only the intended recipients can see them. The result is not only the unchecked spread of misinformation but a thinning of democratic deliberation itself. Voters are reimagined less as citizens capable of judgment and more as behavioral datasets to be segmented, nudged, and managed. This reflects a failure of imagination in governance. The optimization of persuasion has displaced the cultivation of civic life. The danger is that politics becomes a series of private manipulations masked as democratic engagement. And the same logic is already migrating into adjacent domains. If political speech can be individually priced for attention, so too can economic life, with targeted pricing strategies offering different costs or opportunities to different people based on algorithmic predictions of what they will tolerate. In both politics and markets, the risk is that algorithms will not just predict our behavior but shape the very terms on which participation occurs, deepening inequality while presenting themselves as neutral instruments of efficiency.

Governance and Ethical Implications

Scholarship since O'Neil has deepened this picture, tracing how racial hierarchies, labor dynamics, and civic fragmentation are carried forward through automated decision-making. Current governance efforts have largely centered on transparency, explainability, and technical auditing, important but limited moves that risk treating algorithms as black boxes to be opened rather than political projects to be contested.

Regulatory proposals in Europe, the United States, and elsewhere often stop at compliance checklists, emphasizing disclosure and documentation while sidestepping more fundamental questions about what purposes are being optimized, which populations bear the risks, and who has power over design. Governance at this stage remains reactive and fragmented, addressing symptoms of algorithmic harm without reconfiguring the institutional conditions that allow harm to persist. What is needed now is a shift from procedural oversight toward substantive accountability. Participatory frameworks that give affected communities a voice in shaping systems, regulatory mechanisms that can intervene in design choices rather than only audit outputs, and a willingness to confront how algorithms entrench existing political and economic arrangements. Without this deeper transformation, governance risks legitimizing the very systems it seeks to regulate, reinforcing structural injustice under the appearance of technical responsibility.

Conclusion

O'Neil's *Weapons of Math Destruction* endures because it unsettles the idea of algorithms as neutral instruments and reframes them as actors that intervene in social life. Her triad of opacity, scale, and harm captures how mathematical models can slip from analytic devices into engines

of inequality, producing outcomes that are treated as inevitable rather than contested. What later scholarship makes plain is that these dynamics are not abstract. They intersect with racial hierarchies, patterns of labor exploitation, and the slow erosion of democratic norms. The expansion of artificial intelligence into finance, health, policing, and governance only intensifies this trajectory. The risk is not simply technical error but the normalization of systems that encode particular interests and distribute their consequences unevenly, while presenting themselves as natural law. Models that rank, sort, and optimize do not just process information, they redraw the terrain of opportunity and constraint. Recognizing this means resisting the temptation to treat formal sophistication as proof of neutrality and instead acknowledging that mathematical systems always embed political choices. The task then, is not to refine the map until it finally matches the terrain, but to build forms of governance that admit the limits of abstraction and insist on accountability to the realities these systems so powerfully reshape.

References

Chaudari, D. R. 2025. “Bias in AI Recruitment Systems: An Ethical Evaluation of Algorithmic Hiring Tools.” *Bias in AI Recruitment Systems: An Ethical Evaluation of Algorithmic Hiring Tools*.

Moussawi, S., m. Modaresnedzhad, and X. Deng. 2025. “Introduction to the Minitrack on AI and Digital Discrimination.” *Introduction to the Minitrack on AI and Digital Discrimination*.

Oliver, O. N. 2025. “Anti-Blackness Bots in Artificial Intelligence.”

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. N.p.: Crown.

Tallam, K. 2025. “Decoding the Black Box: Integrating Moral Imagination with Technical AI Governance.”