

Automating Inequality and the Governance of AI in Social Welfare

Munongedzi Mabhoko, Clarkson University, MSc Computer Science

Introduction

Virginia Eubanks's *Automating Inequality* (2018) remains a foundational intervention in debates over automation, governance, and social justice. Her concept of the "digital poorhouse" illustrates how contemporary welfare technologies reproduce the logics of older institutions that stigmatized and disciplined the poor. Through her analysis of Indiana's welfare modernization, Los Angeles's homelessness triage system and Allegheny County's child welfare risk model, Eubanks highlights how automated systems can deepen inequality under the guise of efficiency. The force of her work lies in revealing how technology embodies political choices. Yet as Jo Ann Oravec (Oravec, 2018) has argued, discourses on artificial intelligence (AI) and automation often oscillate between hyperbolic optimism and dystopian alarm, neglecting the nuances of sociotechnical failure. A rigorous academic assessment must therefore both support and challenge Eubanks. Supporting her critique of governance gaps and ethical blind spots, while also exploring how technical safeguards, interpretability methods and participatory governance might mitigate risks. The central question is not whether welfare systems should be automated but under what conditions automation can be legitimate, accountable and socially just.

The Indiana Case: Automation as Exclusion

Indiana's welfare automation project initiated under Governor Mitch Daniels in 2006, sought to streamline eligibility determinations through call centers and digital workflow software. A \$1.3 billion contract with IBM promised modernization and fraud reduction but the results were disastrous. Benefit denials skyrocketed often for failure to cooperate a bureaucratic code that shifted system errors onto applicants (Eubanks, 2018).

Eubanks is persuasive in showing how rigid automation amplified harm. Minor documentation lapses, lost forms or misunderstandings became grounds for exclusion with devastating consequences for vulnerable families. Automation here reproduced the poorhouse logic, punishment through procedural hurdles and denial of care. Yet to interpret Indiana as evidence that automation itself is inherently harmful risks overstating the case. The failure was less about AI or intelligent reasoning than about brittle process automation, weak grievance mechanisms and procurement incentives that rewarded scale over accountability. As the Indiana litigation showed, governance, not technology alone was decisive in producing harm. Properly designed automation with uncertainty handling, human review, and rights of redress might have improved efficiency without eroding dignity. The lesson is not that automation is doomed but that automation without governance is dangerous.

The Los Angeles Case: The Ethics of Triage

Los Angeles's Coordinated Entry System (CES) represents a different application of automation. Using a standardized questionnaire (the VI-SPDAT), the system generated vulnerability scores to allocate scarce

housing resources. In principle, CES sought to make allocation more transparent and less subject to arbitrary discretion.

Eubanks argues that the system codified scarcity rather than solving it. By ranking the unhoused, CES rendered structural failures and chronic underinvestment in affordable housing into seemingly neutral scores. In doing so, it normalized the inevitability of leaving many without assistance. This analysis resonates with broader critiques of AI hyperbole, which often frames technological mediation as a solution to fundamentally political problems (Oravec, 2018).

Still, CES cannot be dismissed outright. In contexts of scarcity some allocation mechanism is unavoidable. Standardized tools, however blunt, can reduce individual bias in triage. Moreover, Los Angeles has recently acknowledged flaws in VI-SPDAT, moving toward community-vetted tools that incorporate stakeholder input (National Alliance to End Homelessness, 2021). This shows that algorithmic systems can evolve through participatory redesign. The ethical challenge is not triage itself but whether triage is coupled with commitments to expand resources and whether the criteria reflect the values of affected communities.

The Allegheny County Case: Prediction and Surveillance

Allegheny County's Family Screening Tool (AFST) represents one of the most ambitious uses of predictive analytics in U.S. welfare. By combining administrative data from welfare, justice and health systems, AFST produces a risk score for child maltreatment to aid hotline screeners.

Eubanks views AFST as a troubling extension of the digital poorhouse. Families most entangled with public systems accumulate more data trails which become proxies for risk. In effect, surveillance itself becomes a liability, disproportionately targeting poor and minority households. This critique aligns with concerns that predictive systems amplify structural inequalities by encoding existing distributions of state oversight. Yet evaluations of AFST complicate this narrative. Studies found that the tool modestly improved accuracy in identifying children later found to be in need of services and narrowed racial disparities in case openings (Chouldechova, 2018). AFST was implemented as decision support not replacement, screeners retained discretion. Allegheny also established community advisory processes and independent audits, unusual in the U.S. context. While these measures do not erase ethical risks, they suggest that predictive tools, when carefully governed, can be integrated with safeguards. The case illustrates that automation is not destiny, its outcomes are contingent on governance structures, transparency and oversight. Predictive tools could theoretically help allocate scarce investigative resources and flag risks earlier, provided they remain advisory rather than determinative. The critique should therefore be about governance safeguards, not prediction per se.

From Hyperbole to Governance

Taken together, these cases highlight a central paradox. Automation in welfare is frequently justified with rhetoric of modernization and efficiency, yet its failures often stem from governance gaps rather than technical limits. Predictive tools in social services raise uncomfortable questions, should scarce housing or investigative attention be allocated through algorithms? If scarcity itself is unjust, prediction risks legitimizing injustice. Yet, without such tools, allocation may be even more biased. The problem lies less in automation

than in austerity and underinvestment. As Oravec (Oravec, 2018) observes, AI discourse has long been marked by overstatement, minimizing the importance of sociotechnical failures. In welfare contexts, this rhetorical framing can depoliticize scarcity, mask inequity and disempower citizens. To move beyond this binary of hype and dystopia, welfare automation must be treated as civic infrastructure. This entails rigorous safeguards that increase interpretability, reliability, and safety. Technical methods exist, model documentation tools such as *Model Cards* (Mitchell, 2019) and *Datasheets for Datasets* (Gebru, 2018) promote transparency. Uncertainty quantification techniques such as conformal prediction allow systems to abstain when risk estimates are unreliable. There are also counterfactual explanations that provide affected individuals with meaningful grounds for contestation. Automated reasoning frameworks, including formalized assurance cases, can ensure that welfare algorithms respect policy monotonicity, preventing paradoxical outcomes such as additional evidence of need lowering eligibility.

Yet technical methods alone cannot substitute for ethical design. Governance must guarantee rights to notice, explanation and appeal. Participation of marginalized groups is not optional but essential. Los Angeles's current efforts to replace VI-SPDAT with community-vetted assessments show how stakeholder voice can reshape design. Similarly, Allegheny's advisory processes demonstrate the value of embedding participatory oversight into system governance.

Stakeholders and Values in Tension

These cases also illuminate competing stakeholder perspectives. Governments emphasize consistency and cost containment, corporations seek contracts and reputational gains, frontline workers balance discretion with workload, communities prioritize dignity, privacy, and equity. Without structures for negotiation, automation privileges the values of the most powerful actors. Participatory governance can help rebalance these tensions. Algorithmic impact assessments, community advisory boards and citizen review panels can bring affected voices into decision-making. Rights to contest and amend scores ensure that clients remain active agents not passive data points. Embedding these processes requires treating welfare automation not as proprietary software but as public infrastructure subject to democratic accountability.

Conclusion

Eubanks is right to warn that automation often reproduces the disciplinary logics of the poorhouse (Eubanks, 2018). Indiana, Los Angeles and Allegheny reveal the risks of exclusion, depoliticization and surveillance. Yet they also demonstrate that outcomes vary with governance. Indiana collapsed under rigid automation without recourse. Los Angeles exposed the politics of scarcity but is now experimenting with participatory redesign. Allegheny shows how predictive tools can be cautiously deployed with oversight, though not without ethical risk.

Ultimately, Eubanks reminds us that technology is never neutral, yet to fully grasp the role of automation one must see both its harms and its potential. Every algorithm embeds political choices and recognizing this truth means holding policymakers, technologists, and the community accountable, not to abandon automation. The challenge is not whether to automate, but how. If automation is paired with interpretability tools, robust rights protections and genuine participatory governance, it can serve justice rather than austerity. Without these conditions, the digital poorhouse will persist. With them, welfare automation might instead become a site of civic innovation, reflecting values of equity, dignity, and accountability.

References

- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- Chouldechova, A., Putnam-Hornstein, E., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. *Proceedings of the Conference on Fairness, Accountability, and Transparency*.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. *Communications of the ACM*, 64(12).
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*.
- National Alliance to End Homelessness. (2021). *Assessing vulnerability: An evidence-based review of VI-SPDAT*. Washington, DC.
- Oravec, J. A. (2018). Artificial intelligence, automation, and social welfare: Some ethical and historical perspectives on technological overstatement and hyperbole. *Ethics and Social Welfare*.