



OPEN ACCESS

EDITED BY
Izzy Fox,
Maynooth University, Ireland

REVIEWED BY
Kaoutar Berrada,
Moulay Ismail University, Morocco

*CORRESPONDENCE
Nishanshi Shukla
✉ nishanshi.shukla@utdallas.edu

RECEIVED 01 February 2024
ACCEPTED 20 January 2025
PUBLISHED 30 January 2025

CITATION
Shukla N (2025) Investigating AI systems:
examining data and algorithmic bias through
hermeneutic reverse engineering.
Front. Commun. 10:1380252.
doi: 10.3389/fcomm.2025.1380252

COPYRIGHT
© 2025 Shukla. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Investigating AI systems: examining data and algorithmic bias through hermeneutic reverse engineering

Nishanshi Shukla*

Harry W. Bass, Jr. School of Arts, Humanities, and Technology, The University of Texas at Dallas,
Richardson, TX, United States

Considering Artificial Intelligence systems as boundary objects, which are interdisciplinary objects sustained differently by diverse fields while providing shared discourses between them, this essay summarizes the approaches of examining bias in AI systems. It argues that examining each part related to the building and working of AI systems is essential for unpacking the political play and potential insert points of biases in them. It concentrates on the critical analysis of data and algorithms as two core parts of AI systems by operationalizing hermeneutic reverse engineering. Hermeneutic reverse engineering is a framework to unpack and understand different elements of a technocultural object and/or system that contribute to the construction of its meaning and contexts. It employs a speculative imagination of what other realities can be designed and includes cultural analysis to identify existing meanings and assumptions behind the technocultural object, identifying key elements of signification, and speculating possibilities of reassembling different meanings for the object. The main results obtained by this method on AI systems is using cultural consideration and technological imagination to unpack existing meanings created by AI and design innovative approaches for AI to exert alternate/ inclusive meanings. The research perspectives presented in this article include critical examination of biases and politics within different elements of AI systems, and the impact of these biases on different social groups. The paper proposes using the method of hermeneutic reverse engineering to investigate AI systems and speculate possible alternate and more accountable futures for AI systems.

KEYWORDS

boundary objects, hermeneutic reverse engineering, critical data studies, critical algorithm studies, critical code studies, critical artificial intelligence studies

Introduction

Artificial Intelligence (AI) systems are digital technologies that learn on their own through the data they are trained on, algorithms they are modeled on, and feedback given to them. AI systems, like all other technologies (Winner, 1986), have politics and exercise power. Examples of these politics are visible all around us. Bias based on race, gender, ability, language, class, economic background, and religion, among many other indicators, perpetuating in the AI systems are not just a glitch or an error, but systemic coding of the dominant social fabric (Gebru, 2020; Broussard, 2023). AI systems systemically and structurally enact discrimination and oppression because they are neither inclusive nor do they acknowledge the people, groups

and perspectives they exclude (Broussard, 2019; Crawford, 2021), such as women, people of color, disabled people, and queer people (Buolamwini and Gebru, 2018; Broussard, 2023). The AI Systems operate on an asymmetric power dynamic wherein the groups most impacted by the results of injustices enacted by AI are often devoid of resources to design and deploy these systems (Whittaker et al., 2019, p. 9).

Two ways in which this exclusion is practiced is through a lack of documentation and the domination of hegemonic narration in expert discourses around AI systems. The exclusion practiced in documentation stems from a lack of consideration about whose data is collected, why is that data collected, which computational logics are used on that data, and what or whose purposes are served using results from that data. As Benjamin (2019, p. 1–48) states, “engineered equity” of AI systems either practices “default discrimination” to ignore marginalized people or practices “coded exposure” to overexpose the minoritized groups for extra surveillance. For example, a face detection AI failed to detect Buolamwini’s (2023, p. 13) dark-skinned face as a black woman, but the same AI instantly detected a human face when she put on a plain white mask. This happened because the “coded gaze” (Buolamwini, 2023, p. 13–21) of AI was not trained to see dark skinned faces and detected only light skin. Such errors are possible when the system is trained and coded with data and algorithms that consider light skin as the norm (Buolamwini, 2023). It fails to recognize any face that does not belong to the societal definitions of race and gender that were fed into it during its creation and re-enacts the default social discrimination (Benjamin, 2019; Buolamwini, 2023; Gebru, 2020).

The politics of AI are visible through the biases they exhibit, societal disparities they reflect, the impact they create, and disciplines from which they emerge (Benjamin, 2019; Eubanks, 2019; Keyes, 2018). These politics are visible through the popular discourses within the fields that influence its technological practices. Keyes (2018) showcases how a conventional binary understanding of “gender” within tools of Automatic Gender Recognition (AGR) and the field of Human Computer Interaction (HCI) operationalizes non-inclusive and harmful technologies for transgender people (Keyes, 2018). Thus, to not limit the world view and formulate inclusive AI systems, it is essential in our analysis to incorporate knowledge and practices from different fields. In a study on meaningful digital connections and digital inequality, Katz and Gonzalez (2016) showcase the effectiveness of multilevel research. This approach accounts for influences of and toward technological adoption and engagement at different levels like individual, family and community (Katz and Gonzalez, 2016). Such a multilevel approach can be adopted into the study of the power and politics of AI systems by questioning their components individually and examining the effects of AI use and outputs at different levels of human existence, namely, the level of the individual, the family and the community.

This investigation of AI systems can be conducted considering them as boundary objects (Star, 2015) in the fields of critical data studies, critical algorithm studies, critical code studies, and feminist science and technology studies, and using the tools and approaches given by each of these fields to investigate different facets of existence and the execution of various AI systems. This includes the data used to train and test them, the algorithmic logics used to make them, the programming code used to execute them, and the impact created by using them (Crawford, 2021). As in the case of most digital

technologies that use AI, the systems’ data and algorithms cannot be accessed as they are proprietary information (Bucher, 2018). But one way in which the biases of technologies can be understood, identified and examined is by using hermeneutic reverse engineering (Balsamo, 2011, p. 13–17). This essay summarizes different approaches outlined by the aforementioned fields to unpack the power and politics of AI as executed by each of its parts and then advocates for the use of hermeneutic reverse engineering (Balsamo, 2011, p. 13–17) as a method to investigate AI systems.

Hermeneutic reverse engineering

Hermeneutic Reverse Engineering, as suggested by Anne Balsamo in *Designing Culture* (Balsamo, 2011, p. 13–17), is a framework for constructing meaning around existing technocultural assemblages or a network of objects that hold both technical and cultural significance (Balsamo, 2011, p. 13–17). It is a systematic process which includes steps for cultural analysis and technological reverse engineering to identify key elements of the system that construct meaning for its existence and output (Balsamo, 2011, p. 13–17). Identifying key elements helps in understanding implicit assumptions and formative structures within the system (Balsamo, 2011, p. 13–17). These signifying elements are then interpreted and highlighted in different socio-cultural contexts to elucidate how they operate differently in different scenarios (Balsamo, 2011, p. 13–17). This process leads to an exploration of different meanings, contexts, and outcomes that can be created from them (Balsamo, 2011, p. 13–17). The iterative steps of this process are: observation and description, analysis, interpretation, articulation, rearticulation, prototype, assessment, iteration, production, reflection, and critique (Balsamo, 2011, p. 17).

Using hermeneutic reverse engineering to understand AI systems helps in questioning the biased hegemony present in current AI systems through the lenses of different fields in which they exist. Hermeneutic reverse engineering for AI involves using technological imagination to reconstruct AI that is based in intersectional feminist thought of acknowledging and understanding interlocking systems of structural oppression as stated by the “Combahee River Collective Statement” (Combahee River Collective, 1978, p. 362). Rather than casting AI systems as mysterious, the agency can be shifted, and the AI systems can be unpacked by reverse engineering the black box they seem to present (Bucher, 2018). The learning process of the algorithms along with the processes of establishing their agency and exercising control can be understood through this (Crawford, 2021). Their functioning can also be reimagined if placed in the context of counter narratives that remain hidden in the default hegemonic view (Abbate, 2012). This process can be used to understand the difference in the intended/ unintended and avoidable/ unavoidable biases and politics existing within the data and algorithms of AI systems (Broussard, 2019).

Examining the politics of data

The field of critical data studies understands that one of the ways in which AI produces and reproduces power and politics is through datasets that are partial and biased (Chadarevian and Porter, 2018; Miceli et al., 2022; Crawford, 2021). After collection, data is often disconnected from its history, people, and context of collection (Gebru, 2020; Chadarevian

and Porter, 2018). This disconnection masks underlying subjective meanings of data and wrongfully brands the quantification of data as purely objective (Gitelman and Jackson, 2013). The discriminatory decisions made automatically by algorithmic systems are based on dominant ideologies (Eubanks, 2019), and stem from the loss of context of the original data. Because of this loss of context, the data remains partial and creates problems, like (i) people losing control over their data and how its continuous analysis impacts them (Radin, 2017; Willse, 2015); (ii) algorithms built on the data producing results that are stereotypically biased (Musto, 2016); (iii) algorithmic outputs reproducing power relations (Jefferson, 2020); and (iv) reinforcing and multiplying the bias by using prejudiced outputs as inputs for other algorithms and tasks (Buolamwini and Gebru, 2018).

“Big data” enacts power and politics. While the technical rhetoric around big data is that it is objective and neutral, big data, which is essentially huge collections of data, is subjective and biased (Gitelman and Jackson, 2013). It has partial and contextual stories encoded in numbers, and not overarching generalized truths (Gitelman and Jackson, 2013; Roberge and Seyfert, 2016). Datasets like ImageNet, a collection of human faces, is a collection of images annotated and labeled by humans, and thus organized in political taxonomies laden with implicit assumptions, ideologies, subjectivity, and hierarchical classification (Crawford and Paglen, 2021).

Databases are cultural narratives, that is, they are networked and subjective accounts at individual and cultural levels (Paul, 2007). Their formation is not merely based on data they possess, but also around whose data it is, whose stories it represents, how that data is organized, who organized it, and what meanings can be derived from them (Paul, 2007). The different aspects of examining data of AI systems includes investigating different stages of the data life cycle, such as data collection, categorization, translation, annotation, labeling, storage, use, and access. If data is not directly and freely available, the only way to understand the possibilities of what kind of data might have been used to train the algorithmic model, is to analyze the patterns in output of the algorithmic system (Chun, 2021).

Examining the politics of algorithms and code

The field of critical algorithm studies identifies that another way in which AI produces and reproduces power and politics is through the politics of machine learning algorithms which lie in the realities they create. According to Taina Bucher (2018, p. 1–18), the realities created by algorithms are the result of “programmed sociality” coupled with algorithmic decision making. Bucher’s (2018, p. 1–18) notion of “programmed sociality” refers to the use of computation for the purposes of influencing societal actors and functions. This process of influencing is performed in two parts: (i) how the algorithm is built (that is, the decisions made while building it); and (ii) by enabling the algorithm to make certain kinds of decisions during its execution (Bucher, 2018 p. 1–18).

The algorithmic processes are political as they give only certain outputs and encourage only certain kinds of scenarios to take place. They create biased realities which represent differential power equations among different societal groups (Bucher, 2018). These realities can be studied by questioning which groups are included and/or considered while designing algorithmic realities, and which groups are excluded

and/ or overlooked (Benjamin, 2019). An example of this is apparent in case studies around Google Image Search and Google Photos. For Noble (2018, p. 1–14), searching the term ‘three black teens’ gave mugshots of Black teenagers, while the search term ‘three white teens’ gave wholesome photographs. Gebru (2020, p. 21–22) writes about a Google Photos incident where Black people were misclassified as gorillas. Such misclassifications are not arbitrary and are rooted in racist and discriminatory history (Gebru, 2020). Noble (2018, p. 1–14) uses the concept of “technological redlining” to explain the practices of enforcing and maintaining power by these digital methods of oppression by attaching racist and sexist connotations to different search terms.

The misclassifications and discriminatory biases seen in algorithmic results are not arbitrary. They are rooted in racist and discriminatory history (Gebru, 2020) and are reflections of implicit biases embedded within the institutions that designed these systems (Noble, 2018). They create structurally discriminatory systems coded with societal inequalities and inequities (Katz, 2020). This happens because the only social and human context that AI systems have is the way in which they are programmed, which includes the data they are fed, and the algorithm and code which they use to make sense of that data (Katz, 2020). AI systems are based on social assumptions that they reify and reproduce and are neither neutral nor objective (Bucher, 2018).

Machine learning algorithms exist in multiplicities, that is every time they are executed, they calculate possibilities of various results and then decide the best option for the given input and function at hand (Bucher, 2018; Roberge and Seyfert, 2016). So, each time a decision is made on how to process the input and how to select and display the output, it is a political move (Roberge and Seyfert, 2016). Exploring the political economy of AI reveals that the primary goal of AI systems is not to serve its users, but in fact to serve the commercial goals of the companies that build them (Noble, 2018; Benjamin, 2019).

The algorithm can also be questioned using Marino’s (2020) critical code analysis. While most of the code of AI systems is hidden, codes for certain generic foundation models which can be fine-tuned further, are open source and available on websites like Hugging Face and GitHub. The method of critical code analysis considers the source code of an algorithmic system as a social text whose meaning develops and transforms depending on readers and context (Marino, 2020). This is done using tools of semiotics, cultural studies and critical theories to unpack meanings of codes that are contingent on context and evolve based on the functional use of that code (Marino, 2020). Reading code critically means unpacking the significance of code’s symbolic structures, their effects and their execution, within the cultural moment in which they were developed and deployed (Marino, 2020). Analyzing open-source programming is useful to critically analyze the codified sentiment of the power and politics of AI systems. Thus, most of the black box politics of the algorithm of AI systems can be interrogated by closely observing the output of the AI system for different prompts, to understand hidden meanings of the output of AI systems and questioning the reasons behind them (Bucher, 2018).

Findings and research perspectives

The main findings of this paper include understanding the politics of data, algorithms, and code. Data is a political tool that is always subjective and partial. Questioning the subjectivity, context, collection, categorization practices and storage of data using the framework of

hermeneutic reverse engineering helps us to understand the contribution of data in the power dynamics of AI systems. This enables us to unveil a hegemony of power that prevails in the data and analyze its implications. It helps us to understand the politics of representation within data and how the privilege of different societal groups is reflected in that. Questioning the data makes it possible to interpret what data and whose cultural narrative is missing and the reasons behind that. It also explains, to an extent, the biases embedded in the algorithmic models trained on this data.

Algorithms and code are also political tools and examining the structurally constructed reality, political economy, black boxing, and available programming of AI systems using the framework of hermeneutic reverse engineering builds a non-technosolutionist narrative of code from a non-hegemonic and intersectional feminist standpoint to question algorithmic bias acting within it. It reveals the power play within algorithms and opens the possibility of creating and imagining alternative non-discriminatory realities. To understand these alternative realities, it is important to pay attention to the tensions of fairness at the intersection of individual and group needs (Binns, 2020) and to explore ways to improve fairness in machine learning systems by mitigating discrimination without collecting sensitive data (Veale and Binns, 2017). Some practices that can be employed with the speculative imagination of hermeneutic reverse engineering include actionable AI audits that lead to the reduction of biased results in industrial AI applications (Raji and Buolamwini, 2019), and the compilation of actionable strategies based on alignments and disconnections between AI practitioners and fairness literature (Holstein et al., 2019).

This paper proposes three research perspectives, which include: (i) The comparative exploration of algorithmic biases in various AI systems to better understand their cultural and social impacts. (ii) An examination of how these biases affect different social groups and to test alternative approaches for further analysis. (iii) A participatory approach involving users in the design of AI systems and evaluating the effectiveness of strategies put in place to mitigate bias and promote greater algorithmic equity. These research perspectives aim to deepen the critical examination of AI systems by exploring algorithmic biases and power dynamics across different contexts. They propose a range of empirical studies, including case studies on diverse AI systems, international comparisons of biases, and the impact of biases on marginalized groups. Additionally, they suggest developing alternative analysis methods, participatory design approaches, and longitudinal studies of AI systems' evolution, as well as evaluating ethical challenges and bias remediation strategies.

The approaches proposed in this paper are: (i) Compare Algorithmic Biases Across Contexts: Study how algorithmic biases differ based on cultural and geographical factors by comparing AI systems used in different countries. (ii) Evaluate Impact on Vulnerable Groups: Investigate how biases in AI systems affect marginalized or vulnerable social groups by conducting field studies and surveys to assess their experiences. (iii) Develop Alternative Critical Analysis Methods: Explore and test other methodologies, such as network or content analysis, to improve the detection and understanding of biases in AI systems. (iv) Investigate Participatory Design: Examine how involving end-users in the design process of AI systems can help minimize biases and promote fairness by organizing co-design workshops. (v) Conduct Longitudinal Studies: Track AI systems over time to observe how biases change and how updates to these systems influence existing power dynamics.

Conclusion

The findings of this essay support and illustrate the need to re-examine and reimagine AI systems to avoid bias and inequality. This can be seen in the following connections established from the arguments of this paper: (i) Critical analysis: The results on biases in data and algorithms highlight the importance of conducting critical analysis to understand and correct these biases. (ii) Systems reimagining: The biased realities revealed by the study support the conclusion that alternative approaches are needed to create more just AI systems. (iii) Research perspectives: The problems identified by the results encourage continuous exploration and development of new methods for a better design of AI systems.

Therefore, considering AI systems as boundary objects (Star, 2015) and critically examining them through hermeneutic reverse engineering prompts us to be speculative and work alongside existing technologies to seek all other possible realities than can be created. It entails unpacking data, algorithms and code that make AI systems, and extends to imagining alternative futures by exploring different decisions that make a particular system, how to alter these decisions, and what else can exist if some of those decisions are altered. It involves questioning what is normal within a particular system, and who falls outside that norm (Whittaker et al., 2019, p. 27)? This approach also asks: What would the system look like if the underprivileged and marginalized groups were the ones being overrepresented and responsible for designing the AI system (Geburu, 2020, p. 264)? Though a complete cultural analysis and reverse engineering of AI systems is not possible owing to their protected propriety, vast and interconnected resources; even partial analysis might lead toward a technological reimagination of AI that exposes its underlying biases.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

NS: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abbate, J. (2012). *Recoding gender: women's changing participation in computing*. Cambridge, Massachusetts: The MIT Press.
- Balsamo, A. (2011). *Designing culture: the technological imagination at work*. Durham and London: Duke University Press.
- Benjamin, R. (2019). *Race after technology: abolitionist tools for the new Jim code*. Newark, New Jersey: Polity Press.
- Binns, R. (2020). "On the apparent conflict between individual and group fairness," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, Spain: FAccT. 514–524.
- Broussard, M. (2019). *Artificial unintelligence: how computers misunderstand the world*. Cambridge, Massachusetts: The MIT Press.
- Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability Bias in tech*. Cambridge, Massachusetts: The MIT Press.
- Bucher, T. (2018). *If ... Then: Algorithmic power and politics*. New York: Oxford University Press.
- Buolamwini, J. (2023). *Unmasking AI: My mission to protect what is human in a world of machines*. New York: Random House.
- Buolamwini, J., and Gebru, T. (2018). "Gender shades: intersectional accuracy disparities in commercial gender classification," in *Proceeding of the 2018 Conference on fairness, accountability and transparency*, Spain: FAccT. 77–91.
- Chadarevian, S., and Porter, T. (2018). Introduction: scrutinizing the data world. *Hist. Stud. Nat. Sci.* 48, 549–556. doi: 10.1525/hsns.2018.48.5.549
- Chun, W. H. K. (2021). *Discriminating data: correlation, neighborhoods, and the new politics of recognition*. Cambridge, Massachusetts: The MIT Press.
- Combahee River Collective (1978). "The Combahee River collective: a black feminist statement" in *Capitalist patriarchy and the case for socialist feminism*. ed. Z. R. Eisenstein (New York: Monthly Review Press), 362–372.
- Crawford, K. (2021). *Atlas of AI*. New Haven, Connecticut: Yale University Press.
- Crawford, K., and Paglen, T. (2021). "Excavating AI: the politics of images in machine learning training sets," in *AI & Society* 36, 1399–1116. doi: 10.1007/s00146-021-01301-1
- Eubanks, V. (2019). *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press, PICADOR.
- Gebru, T. (2020). "Race and gender: data-driven claims about race and gender perpetuate the negative biases of the day" in *The oxford handbook of ethics of AI*. eds. M. D. Dubber, F. Pasquale and S. Das (New York: Oxford University Press), 252–269.
- Gitelman, L., and Jackson, V. (2013). "Introduction" in *Raw data is an oxymoron*. ed. L. Gitelman (Cambridge, Massachusetts: The MIT Press), 1–14.
- Holstein, K., Wortman Vaughan, J., Daumé, H. III, Dudik, M., and Wallach, H. (2019). "Improving fairness in machine learning systems: what do industry practitioners need?" in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI 2019)*, 1–16. Association for Computing Machinery: New York, NY
- Jefferson, B. (2020). *Digitize and punish: racial criminalization in the digital age*. Minneapolis, Minnesota: University of Minnesota Press.
- Katz, Y. (2020). *Artificial whiteness: politics and ideology in artificial intelligence*. New York: Columbia University Press.
- Katz, V., and Gonzalez, C. (2016). Toward meaningful connectivity: using multilevel communication research to reframe digital inequality. *J. Commun.* 66, 236–249. doi: 10.1111/jcom.12214
- Keys, O. (2018). "The Misgendering machines: trans/HCI implications of automatic gender recognition," in *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), Association for Computing Machinery: New York, NY.
- Marino, M. C. (2020). *Critical code studies*. Cambridge, Massachusetts: The MIT Press.
- Miceli, M., Posada, J., and Yang, T. (2022). Studying up machine learning data: why talk about bias when we mean power? *Proc. ACM Hum.-Comput. Interact.* 6, 1–14. doi: 10.1145/3492853
- Musto, J. (2016). "Trafficking, technology, and 'data-driven' justice" in *Control and protect: collaboration, Carceral protection, and domestic sex trafficking in the United States (Berkeley and Los Angeles, California: University of California Press)*, 68–85.
- Noble, S. U. (2018). *Algorithms of oppression: how search engines reinforce racism*. New York: New York University Press.
- Paul, C. (2007). "The database as system and cultural form: anatomies of cultural narratives" in *Database aesthetics – Art in the age of information overflow*. ed. V. Vesna (Minneapolis, Minnesota: University of Minnesota Press), 95–109.
- Radin, J. (2017). Digital natives: how medical and indigenous histories matter for big data. *Osiris* 32, 43–64. doi: 10.1086/693853
- Raji, I. D., and Buolamwini, J. (2019). "Actionable auditing: investigating the impact of publicly naming biased performance results of commercial AI products," in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AI/ES)*, 429–435. Association for Computing Machinery, New York, NY.
- Roberge, J., and Seyfert, R. (2016). "What are algorithmic cultures?" in *Algorithmic cultures: Essays on meaning, performance and new technologies*. eds. J. Roberge and R. Seyfert (London: Routledge, Taylor & Francis), 1–25.
- Star, S. L. (2015). "Misplaced Concretism and concrete situations: feminism, method, and information technology" in *Boundary objects and beyond: working with Leigh Star*. eds. G. C. Bowker, S. Timmermans, A. E. Clarke and E. Balka (Cambridge, Massachusetts: The MIT Press), 143–167.
- Veale, M., and Binns, R. (2017). Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. *Big Data Soc.* 4:2. doi: 10.1177/2053951717743530
- Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., and Rea, S. (2019). *Disability, bias, and AI (MSR-TR-2019-38)*. New York: AI Now Institute.
- Willse, C. (2015). "Governing through numbers: HUD and the Databasing of homelessness" in *The values of homelessness: managing surplus life in the United States*. ed. C. E. Wayne (Minneapolis, Minnesota: The University of Minnesota Press), 109–138.
- Winner, L. (1986). "Do artifacts have politics?" in *The whale and the reactor: a search for limits in the age of high technology* (Chicago, Illinois: University of Chicago Press), 19–39.