Rethinking Racial Inequality in Artificial Intelligence: A Critical Reflection through Cybernetics

Mia Damen-Soko, University of Sussex

Abstract

In my second year at the University of Sussex, I wrote an essay for the AI, Culture and Society module exploring how algorithms and data collection practices can reproduce racial inequality, and what could be done to make AI systems fairer and more accountable. Drawing on Frantz Fanon, Simone Browne and Shoshana Zuboff, I connected discriminatory outcomes to historical data, institutional decisions and commercial priorities. Returning to that work, I question whether its broad account of bias sufficiently explained the relationships between these different sources of harm. This reflection develops that question through cybernetics, which examines how the consequences of an action return as information that shapes later action. Predictive policing provides a focused example, as decisions on where officers patrol influence the records used to guide subsequent deployment. Although my original essay recognised this feedback, I had not consistently followed its implications either for the structure of my argument or the remedies I proposed. I reconsider the essay's strengths, its assumption of a recoverable neutrality and the reassurance offered by terms such as transparency and accountability. The reflection leads towards a more focused research practice, attentive to how institutional decisions produce the evidence used to justify them, and to whether those affected possess the power to challenge that process.

Where this thinking began

This summer, I undertook a research internship at the University of Cambridge where I examined processes of **racialisation** and bias within live facial recognition

technologies used in British policing. As I began thinking more deeply about the relationship between race and technology, I returned to an essay I had written in my second year, both for inspiration and to consider how my approach to these questions had developed.

The original essay was submitted in 2025 for my AI, Culture and Society module as a final assignment, and asked how **algorithms** and data collection perpetuate racial inequality, and what changes might make these systems fairer. At the time, I took an algorithm to mean a set of instructions for processing information. Returning to the essay now, I am struck by how much work I asked 'bias' to do, using it to connect problems in training data, designers' decisions and the unequal consequences of an algorithm's use without always explaining how those concerns related to one another. The systems I considered used records to inform decisions such as where police should patrol or whether a face matched an identity, allowing me to question the assumption that a decision becomes impartial because a computer participates in making it.

I introduced this argument through Pygmalion, the sculptor whose creation comes to life in Ovid's *Metamorphoses* (Ovid, 2018, Book 10), presenting artificial intelligence as a mirror that reflects and amplifies its makers' imperfections. While I still value that image, returning to the essay has made me question whether I followed the histories entering the machine more carefully than the consequences that emerged from its decisions. My reflection therefore develops a question that the original work raised but did not fully pursue: how algorithms help produce the conditions they subsequently appear only to observe.

What I was trying to understand

My starting point was that algorithmic discrimination could not be explained as technical mistakes alone, because records already carry histories of unequal treatment while institutions determine what should be measured and how predictions influence people's lives. I understood this relationship as '**sociotechnical**', meaning that technology and its surrounding social arrangements must be examined together, and I approached its racial implications through the work of Frantz Fanon and Simone Browne. Fanon's account of **epidermalisation** examines how imposed racial meanings

shape embodied identity, with his ‘racial epidermal schema’ describing the experience of encountering oneself through racist assumptions imposed by others (1986: 112). This helped me understand racialisation as a process through which assigned meanings organise people’s treatment, while Browne’s *Dark Matters* (2015) connected surveillance, the organised monitoring of people, to histories of anti-Black racism. Her discussion of **digital epidermalisation** extended these questions into technologies that encode and classify bodies (2015: 109–10), showing why an apparently technical classification could determine whose identity was accepted and whose presence became suspicious. Alongside Shoshana Zuboff’s account of **surveillance capitalism**, which describes the extraction of human experience as data for prediction and profit (2019), these perspectives helped me place algorithms within broader structures of power. Although that breadth was valuable, my examples accumulated faster than I developed their implications, establishing several places where inequality appeared without always explaining the relationships through which it was reproduced and sustained.

A question already inside the essay

The opening page exposes a tension in my approach. I defined bias as the remnants of power and social context that ‘skew the neutrality of algorithms’ (Damen-Soko, 2025) while simultaneously questioning whether algorithms could be neutral. Although I intended to emphasise social inequality, my wording suggested a neutral machine beneath the distortion, as though sufficient correction might eventually reveal it. I am reluctant, however, to let that sentence represent the whole essay, which also questioned institutional power and recognised that technical accuracy could not adequately measure social harm. Overlooking those insights would exaggerate the novelty of my later thinking, when the more revealing limitation lies between recognising a problem and allowing it to reshape my argument. This becomes clearest in my discussion of feedback, where I described how intensified policing generates records that can justify continued police attention, but treated this as one mechanism among several rather than a relationship connecting data collection, prediction and institutional action. Revisiting that discussion, I realise that the question I now want to pursue was already present, although I had moved on after identifying the loop without

considering how it unsettled the distinctions between data collection, prediction and action that I analysed separately elsewhere.

Cybernetics gives me a way to remain with that question through its study of communication and control in systems, an approach developed across mathematics, engineering and biology (Wiener, 1961). A system consists of interacting parts, which here include records, software, officers and institutional decisions, while 'control' concerns how their activity is steered towards an objective. Within this framework, 'feedback' describes information about the consequences of an action returning to influence subsequent action, connecting observation and intervention over time. Wiener's (1961: 95–97) thermostat example makes this relationship tangible through a sensor that measures room temperature and a heating system that responds to the difference between the measured and desired temperatures. As heating changes the room, new measurements alter the response, producing **negative feedback** because the process counteracts a departure from the desired condition. **Positive feedback** instead amplifies a change, with 'positive' and 'negative' describing the direction of the process rather than whether its consequences are morally good or bad. For my reflection, this means that returning information does not necessarily correct an existing problem, while the choice of a desired temperature also prompts me to consider who determines an institution's objectives and definitions of success. Cybernetics consequently encourages me to follow how the parts of a system interact over time, including how human decisions can complete a **feedback loop** without any software having to rewrite its own code.

Following what happens next

Predictive policing makes these relationships concrete because it uses historical incident records to help police decide where to concentrate patrols, translating patterns in recorded activity into predictions that guide deployment. Ensign *et al*. (2018) examined this process through mathematical modelling and computer simulations, showing how incidents discovered during patrols can influence subsequent predictions and repeatedly direct officers towards the same areas. When officers spend more time in a particular neighbourhood, their presence affects which incidents they encounter and record, so the information returned to the model is partly shaped by the

deployment decision that preceded it. Those records may then reinforce the original prediction without demonstrating a corresponding difference in underlying crime, giving the system apparent confirmation through observations influenced by its own operation. As Figure 1 shows, the relationship connects observation, decision, action and the environment, with institutional objectives shaping what is recorded and how predictions are used. This provides a more precise basis for my cybernetic interpretation than simply describing the data as biased, because it requires me to examine how policing participates in producing the evidence on which future decisions depend. Ensign *et al*. also explore changes to model inputs that can prevent runaway feedback within their model, reminding me that the cycle should be examined for possible interruption rather than treated as an inevitable feature of every policing system. Distinguishing a simulated mechanism from evidence about an actual police force matters here, since I cannot assume that every institution uses the same records or responds to predictions in the same way.

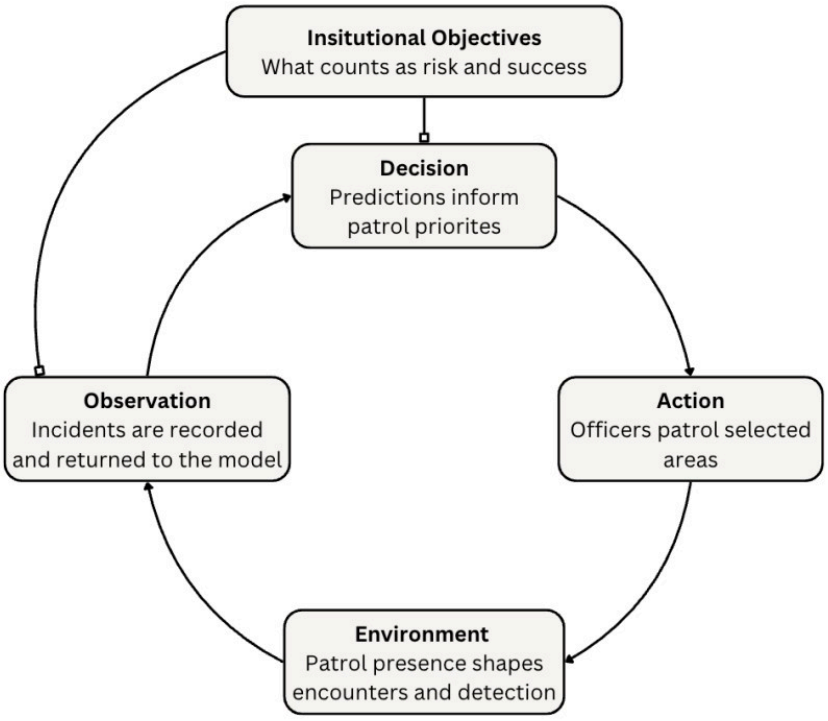


Figure 1: How policing can shape its next prediction. Solid arrows show the feedback loop; squared lines show institutional influence. Schematic informed by Ensign *et al.* (2018).

Following this loop changes what I can ask of the records, since an arrest may be documented accurately while revealing much less about the underlying distribution of offending. My earlier emphasis on identifying distorted information did not consistently distinguish between a record's accuracy and the conditions under which that record became available, even though this distinction changes how its apparent authority should be understood. The mirror begins to move when a prediction influences how a neighbourhood is policed and the resulting records appear to confirm the prediction, allowing an institution to encounter the consequences of its own intervention as apparently independent evidence.

Where police attention is already racially unequal, repeated classification and intervention can sustain associations between particular populations, places and danger. Racial judgements are renewed through routine decisions that acquire the appearance of objective knowledge. Cybernetics does not explain the origins of those judgements, but it helps me follow their reproduction without locating the whole problem inside an isolated dataset or model. This deepens the sociotechnical argument I had already begun, because the relationship between technology and society becomes something I must trace through successive decisions rather than acknowledge as a general condition surrounding the algorithm.

The reassurance of a solution

Revisiting my recommendations is consequently uncomfortable. Representative data, consent, transparency and inclusive design still matter to me, even as I recognise how naming them can make a conclusion feel more resolved than it is. Although transparency makes a system's operation understandable, it does not necessarily create accountability, since knowing why someone has been classified as a risk may leave them no better placed to challenge that classification. An institution can record a complaint without changing either its decision or the practices behind it, allowing information to return without creating any obligation to respond. Looking back, I can

see that my recommendations needed to explain how greater visibility would give those affected the power to contest decisions, rather than assuming that making a system understandable would also make it answerable. This has also made me reconsider how readily I described discriminatory outcomes as failures, because that judgement depends on whose expectations the system is understood to serve.

A model may predict recorded arrests effectively while supporting policing practices that remain unjust, meaning that improvements in its measured performance do not necessarily provide a fairer basis for deciding where officers ought to go. My emphasis on correcting bias therefore needed to extend beyond the accuracy of a prediction to consider why that prediction was being made, whose priorities it served and whether its intended purpose was itself defensible. The same question sharpens my discussion of surveillance capitalism, since organisations that reward engagement or profit may receive information about social harm without allowing it to alter what they pursue, demonstrating once again that feedback only matters when institutions are prepared to respond to it.

What I would carry forward

Finding a framework can itself be reassuring, and I need to remain attentive to what that reassurance allows me to overlook, particularly when a diagram makes complex relationships appear more coherent than lived experience. While cybernetics helps me trace the movement between observation and action, reducing people to the records generated about them would repeat the abstraction that troubled me. Fanon and Browne must therefore shape the questions, allowing histories of racialisation and unequal authority to inform a fuller cybernetic analysis of feedback and institutional control. In future research, I would concentrate on one policing context, distinguishing incidents reported by residents, those discovered by officers and arrests rather than grouping them under ‘crime data’. Tracing which records influence deployment and whether observations return to the model would allow me to establish a feedback loop through evidence, rather than infer one from the presence of an algorithm. Community research and accounts from those subjected to policing would also help define the enquiry, including how surveillance is felt and experienced beyond institutional measurements. Any proposed oversight would need to establish who participates, what

they can change and whether they can stop the system's use, leaving reform, suspension and non-deployment open to consideration. Pygmalion remains useful, although my attention has shifted towards what the creation does in the world and whether those affected can challenge it. I therefore leave the essay with a more demanding understanding of fairness, grounded in the relationship between institutional purposes, lived consequences and the power to contest them.

Looking again

Returning to the essay has changed the point at which I consider an explanation complete, and identifying biased data or a discriminatory outcome no longer feels sufficient. I want to follow the decisions connecting them, and then follow what happens when their consequences return. Pygmalion still belongs in the argument, but my attention has shifted towards the life the creation acquires among its makers. I began by asking how to make the mirror fairer, and I now want to understand what is being done in its name, and whether those it reflects have any say in what follows. The reflection leaves me less certain that I have reached a solution, but clearer about the questions I cannot afford to leave behind.

AI use and authorship

No artificial intelligence (AI) tools were utilised in any element of this article.

List of illustrations

Figure 1: How policing can shape its next prediction. Solid arrows show the feedback loop; squared lines show institutional influence. New schematic based on my supplied flowchart, informed by Ensign *et al*. (2018), and drawn by the author using Canva.

Appendix A: Original project

The project discussed is my second-year essay, *In what way do algorithms and data collection practices perpetuate racial inequality, and what sociotechnical interventions can be employed to foster greater fairness and accountability in AI systems?*. It was submitted in

2025 at the University of Sussex for the AI, Culture and Society module.

References

Browne, S. (2015), *Dark Matters: On the Surveillance of Blackness, Durham*, NC: Duke University Press, available at https://doi.org/10.1215/9780822375302 accessed 28 August 2026.

Ensign, D., S. A. Friedler, S. Neville, C. Scheidegger and S. Venkatasubramanian (2018), 'Runaway feedback loops in predictive policing', *Proceedings of Machine Learning Research*, 81, 1–12, available at https://proceedings.mlr.press/v81/ensign18a.html, accessed 28 August 2026.

Fanon, F. (1986), *Black Skin, White Masks*, trans. C. L. Markmann, London: Pluto Press (originally published as *Peau Noire, Masques Blancs* by Éditions du Seuil in 1952; English translation first published by Grove Press in 1967), available at https://bkbcollege.in/upload/dpt_book/1645635984.pdf accessed 28 August 2026.

Damen-Soko, M (2025), *In what way do algorithms and data collection practices perpetuate racial inequality, and what sociotechnical interventions can be employed to foster greater fairness and accountability in AI systems?,* unpublished essay, University of Sussex.

Ovid (2018), *Metamorphoses: The New, Annotated Edition*, trans. R. Humphries, annotated J. D. Reed, Bloomington, IN: Indiana University Press, available at https://iupress.org/9780253033598/metamorphoses/ accessed 28 August 2026.

Wiener, N. (1961), *Cybernetics: Or Control and Communication in the Animal and the Machine*, Cambridge, MA: MIT Press (originally published by Hermann & Cie and The Technology Press in 1948), available at https://mitpress.mit.edu/9780262730099/cybernetics/ accessed 28 August 2026.

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London: Profile Books, available at https://profilebooks.com/work/the-age-of-surveillance-capitalism/ accessed 28 August 2026.

Glossary

Algorithm: Instructions for processing information or completing a task.

Cybernetics: The study of communication, control and feedback in interacting systems.

Digital epidermalisation: Browne's term for power exercised through technologies that encode and classify bodies.

Epidermalisation: The imposition of racial meanings on the body and embodied identity.

Feedback loop: A process in which consequences return as information influencing later action.

Negative feedback: Feedback that counteracts a change or deviation.

Positive feedback: Feedback that reinforces or amplifies a change.

Predictive policing: Data analysis used to inform police decisions, including where to patrol.

Racialisation: The process of assigning racial meanings to people, places or behaviours.

Sociotechnical: Concerning the interaction of technology, people, institutions and practices.

Surveillance capitalism: Shoshana Zuboff's account of extracting human experience as data for prediction and profit.

To cite this paper please use the following details: Damen-Soko, M. (2026), 'Rethinking Racial Inequality in Artificial Intelligence: A Critical Reflection through

Cybernetics', *Reinvention: an International Journal of Undergraduate Research*, Volume 19, Issue 2, https://reinventionjournal.org/index.php/reinvention/article/view/2634. Date accessed [insert date]. If you cite this article or use it in any teaching or other related activities, please let us know by emailing us at Reinventionjournal@warwick.ac.uk.

https://doi.org/10.31273/reinvention.v19i2.2634, ISSN 1755-7429 © 2026, contact reinventionjournal@warwick.ac.uk. Published by Warwick Education Academy, University of Warwick. This is an open access article under the CC-BY licence (https://creativecommons.org/licenses/by/4.0/)