

THE HIDDEN POLITICS OF GENERATIVE AI

Professor
Martin Oliver

Prof. Martin Oliver



IEUC Advocate in Residence and Professor in
Education Technology



*Framing questions
about AI ethics as
if they are technical
questions masks
their politics, hiding
them from scrutiny
and challenge.*

It might seem flippant to ask if a bridge can be racist. Given the rapid spread of generative Artificial Intelligence (AI), though, this question turns out to be quite important.

The story of how these two things are connected starts with road building on Long Island, but it was Langdon Winner's analysis of this project that drew attention to the social consequences it had:

The two hundred or so low-hanging overpasses on Long Island were deliberately designed to achieve a particular social effect. Robert Moses [...] had these overpasses built to specifications that would discourage the presence of buses on his parkways. [...] The reasons reflect Moses's social-class bias and racial prejudice. [...] Poor people and blacks, who normally used public transit, were kept off the roads because the twelve-foot tall buses could not get through the overpasses. One consequence was to limit access of racial minorities and low-income groups to Jones Beach, Moses's widely acclaimed public park. (Winner, 1980)

This example illustrates a wider theoretical point in the study of science and technology: the way that we design artefacts and infrastructure reflects people's assumptions about what is normal, desirable, appropriate, and so on. These designs encode these assumptions into the material culture that surrounds us, affecting all our subsequent interactions with them. To use Bruno Latour's phrase (1990), "technology is society made durable".

Optimistic designers and engineers have used this principle as a way to intervene in the world. Alvin Weinberg (1967), for example, asked

optimistically:

To what extent can technological remedies be found for social problems without first having to remove the causes of the problem? It is in this sense that I ask, "Can technology replace social engineering?" [...] The Technological Fix accepts man's intrinsic shortcomings and circumvents them or capitalizes on them for socially useful ends. [...] One does not wait around trying to change people's minds: [...] if people insist on driving autos while they are drunk, one provides safer autos that prevent injuries even after a severe accident. (Weinberg, 1967)

It wasn't long before the problems with this approach became apparent, however, and the idea of the 'technological fix' was rapidly debunked by critical research. One important point is that designs are always ideals, and require contextualisation in practice. This means that there are always unintended consequences as designed artefacts are incorporated into daily lives – in some cases, these lead to what Tenner (1997) describes as 'revenge effects', where the supposedly beneficial features of a new technology foster harmful rather than beneficial behaviours. (Applying this to Weinberg's example above, for example: maybe safer cars mean people feel it's ok to drive while drunk and so do it more frequently.)

To illustrate some of these points in the context of education, the development of Wikipedia has generally been seen as positive, making it easier to share trustworthy information widely. The principle of requiring independently verifiable sources on Wikipedia pages is one of the things differentiating it from sites like Reddit or Quora, where people share their opinions freely: the evidence or warrant for the claims

made has to be provided.

Broadly this is seen as a good thing to do, enacting the kinds of standards educationalists and academics value. However, as Ford & Wajcman (2017) showed, the unintended consequence of this is a form of conservatism that entrenches existing privilege. Minoritised groups and experiences are under-researched and under-documented, which means evidence isn't available to use as a supporting citation, which means that topics relevant to these groups are structurally excluded. Their analysis showed, for example, how attempts to write about subjects from India, Malaysia and South Africa were blocked because independent sources weren't available and interviews with local experts could not be independently verified; and similarly, that Wikipedia's coverage entrenched historical neglect of women as scientists, inventors and poets but provided extensive documentation of their roles in pornography.

This example is part of a much wider pattern. Technology may be society made durable, but it is not necessarily the society we want to see, nor even the whole of the society we have. For example, design has often been done by men for men, ignoring women's different experiences, preferences and bodies, with consequences that can be fatal (Perez, 2019). More invidiously, many of the measures we take as foundational – such as census categories and medical diagnoses – were created as forms of governance, often as mechanisms of Imperial control (Bowker & Star, 2000). The trick to obscuring this fact lies in presenting them as if they were purely technical concerns – what Ferguson, in the context of international development, called “anti-politics machines”, framing social issues as technical problems, then posing

technical solutions to them.

In this perspective, the “development” apparatus in Lesotho is not a machine for eliminating poverty that is incidentally involved with the state bureaucracy; it is a machine for reinforcing and expanding the exercise of bureaucratic state power [...]. By uncompromisingly reducing poverty to a technical problem, and by promising technical solutions to the sufferings of powerless and oppressed people, the hegemonic problematic of “development” is the principal means through which the question of poverty is de-politicized in the world today. [...] The way it all works out suggests an analogy with the wondrous machine made famous in Science Fiction stories - the “anti-gravity machine,” that at the flick of a switch suspends the effects of gravity. In Lesotho, at least, the “development” apparatus sometimes seems almost capable of pulling nearly as good a trick: the suspension of politics. (Ferguson, 1994)

Bringing these discussions up to date, we can see similar issues with generative AI systems such as ChatGPT. These technologies are examples of Large Language Models (LLM), created to generate text or images. The way that they do this is through machine learning – the development of statistical algorithms that ‘learn’ patterns in existing dataset and then generalise them to new prompts or data. This ‘learning’ involves passing data through neural networks or other statistical algorithms to see what the output will be; this process can be ‘supervised’ (example inputs and desired outputs are given to the system) or ‘unsupervised’ (the system is left to find patterns in a data set), and the generative predictions are ‘reinforced’ if they produce

desirable outcomes and suppressed if they don't.

The issue with generative AI is that the training datasets that have been used for many language models are based on publicly available data – such as emails from Enron employees released as part of the fraud case against the company, or posts on Reddit, in spite of recognition that Reddit's users are unrepresentative of the wider population and include many toxic communities (Baack, 2024). The consequence of this is that LLMs such as ChatGPT have been trained to be racist, sexist, homophobic and so on; they have encoded the social issues that are present in the data sets most conveniently available to developers. It should be no surprise when outputs from these systems reflect this toxic heritage. The response has been to try and filter the worst of these outputs, and also to help the systems 'learn' to be less offensive by reinforcing better patterns within the system. However, both of these responses require extensive human intervention, sifting through data sets to remove toxic content – work that is at best boring, and at worst traumatising, and which is typically outsourced as casual workers in India or African countries, keeping these harms geographically distant from the California-based companies that benefit the most from these developments.

Similar issues arise in related areas of AI, such as facial recognition – where algorithms trained to detect patterns from images of faces are used to create or classify new images. These technologies build upon eugenic histories designed to establish and police racial differences (Taylor et al, 2023). Like many new technologies, they are then tested out on, and disproportionately used to police, people from minoritised groups (Eubanks, 2018).

While there has been outcry about the ethics of such systems, and a massive proliferation of policies and guidelines intended to govern the risks and harms associated with their use, critical research suggests that most of the policy work is ineffective. It typically focuses on issues where technical fixes can be or have already been developed (Hagendorff, 2020), or else is never really intended to be used, serving instead as a form of 'ethics-washing' so that universities and companies appear benevolent while allowing problematic practices to continue unchecked (Ochigame, 2022).

Since many of these new technologies, up to and including generative AI, encode inequality and then hide their politics, what then can be done? What, for example, might anti-racist AI look like?

Very few of us have the power to change how major international companies are operating, meaning that we will need to rely on modest forms of resistance. At an individual level, some of these systems are easier to refuse or opt out of than others – surveillance is often indiscriminate, making it hard to avoid, for example. Constructively, systems could be trained on smaller, better curated data sets rather than relying on common, problematic 'general' data sets (boyd & Crawford, 2012); this would enable tailored generative AI to be created, which might reflect the practices and values of the communities represented in that specific training data. Rather than waiting and hoping that 'general' LLMs like ChatGPT will learn to be anti-racist, it may be necessary to create small-scale alternatives in partnership with specific communities.

While we advocate and hope and wait for this to happen, we can continue the important work of drawing attention to this as an ongoing,

cumulative form of structural inequity. Encoding prejudice and injustice in material and technical systems sustains them, making them harder to dismantle. Framing questions about AI ethics as if they are technical questions masks their politics, hiding them from scrutiny and challenge. Developing awareness of these issues, sharing examples when we see it happening and developing our ways of naming and understanding the problem are all important forms of work, which we can do now. They will not solve the problem, but they can provide the groundwork for a better relationship with technologies in the future.

References

- Baack, S. (2024). [A critical analysis of the largest source for generative ai training data: Common crawl](#). In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, 2199-2208.
- Bowker, G. C., & Star, S. L. (2000) *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT press.
- boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Ferguson, J. (1994). *Anti-politics machine: Development, depoliticization, and bureaucratic power in Lesotho*. University of Minnesota Press.
- Ford, H., & Wajcman, J. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social studies of science*, 47(4), 511-527.
- Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Boston: Houghton Mifflin Harcourt.
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, 30(1), 99-120.
- Latour, B. (1990) *Technology is Society Made Durable*. *The Sociological Review*, 38 (1), 103-131
- Ochigame, R. (2022). The invention of 'ethical AI': How big tech manipulates academia to avoid regulation. In Phan, T., Goldenfein, J., Kuch, D., & Mann, M. (Eds), *Economies of Virtue – The Circulation of 'Ethics' in AI*, 49-56. Amsterdam: Institute of Network Cultures.
- Perez, C. C. (2019). *Invisible women: Data bias in a world designed for men*. London: Vintage.
- Taylor, S. M., Gulson, K. N., & McDuie-Ra, D. (2023). Artificial intelligence from colonial india: Race, statistics, and facial recognition in the global south. *Science, Technology, & Human Values*, 48(3), 663-689.
- Tenner, E. (1997) *Why things bite back: Technology and the revenge of unintended consequences*. Vintage.

Weinberg, A. M. (1967). Can technology replace social engineering? *American Behavioral Scientist*, 10(9), 7-9.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 121–136.

Focus On

EDI ©

Disclaimer: Focus of EDI © is a brand of the Institute for Equity, University Centre. The author of this article retains copyright, and views expressed are not necessarily those of the Institute.

Citation: Oliver, M. (2025). The Hidden Politics of Generative AI, Focus on EDI © – Issue 8, London: Institute for Equity, University Centre.