Forum on Digital Ethics in Research
October 18/20, 2023

Faculty of Engineering of the University of Porto, Porto and online

DIGITAL ETHICS IN RESEARCH

What's wrong with «AI ethics» narratives

Daniela Tafani

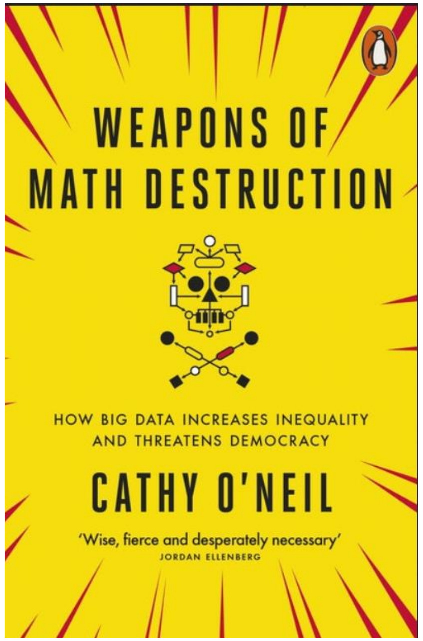
October 18, 2023

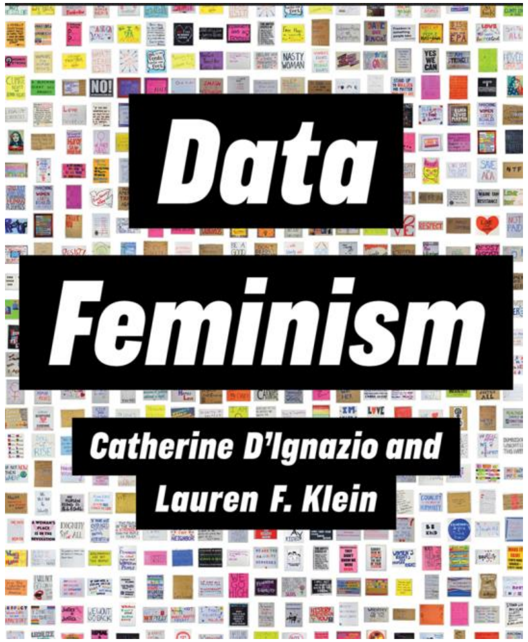
IN SUPREMÆ DIGNITATIS 1343

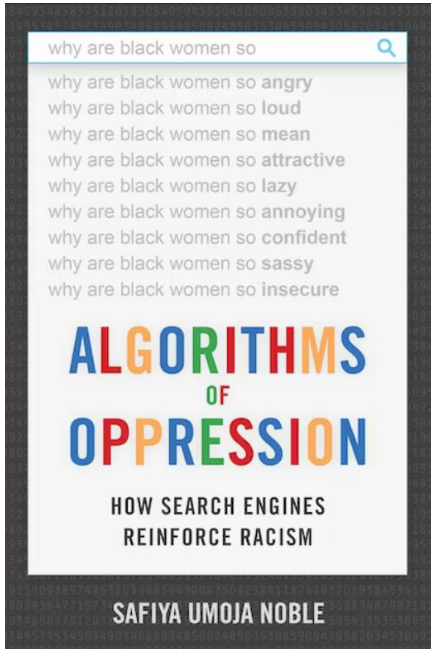
AI and decisions

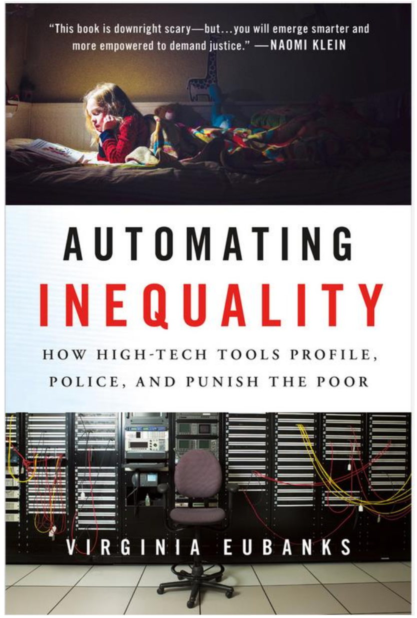
In predictive optimisation systems, machine learning is used to predict future outcomes of interest about individuals, and these predictions are used to make decisions about them.

More than a decade after these systems were first introduced, **the myth of the objectivity of algorithmic decisions** has been debunked; the biases they reproduce, the stereotypes they perpetuate, and the harm and injustice they cause are well documented.


WEAPONS OF MATH DESTRUCTION
HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY
CATHY O'NEIL
'Wise, fierce and desperately necessary'
JORDAN ELLENBERG




why are black women so
why are black women so angry
why are black women so loud
why are black women so mean
why are black women so attractive
why are black women so lazy
why are black women so annoying
why are black women so confident
why are black women so sassy
why are black women so insecure
ALGORITHMS
OF
OPPRESSION
HOW SEARCH ENGINES REINFORCE RACISM
SAFIYA UMOJA NOBLE


"This book is downright scary—but…you will emerge smarter and more empowered to demand justice." —NAOMI KLEIN
AUTOMATING
INEQUALITY
HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR
VIRGINIA EUBANKS

Three fallacies

1. **The fallacy of AI functionality:** when an 'artificial intelligence' system is given a particular task, it is assumed that the system is actually capable of doing it, even if the system is inadequate for the task or the task is not possible at all.

2. **The fallacy of examples drawn from the future or science fiction.**

3. **First step fallacy:** Dreyfus quoted an analogy made by his brother, the engineer Stuart Dreyfus: “It was like claiming that the first monkey that climbed a tree was making progress towards landing on the moon”.

For references, s. D. Tafani, *Do AI systems have politics? Predictive optimisation as a move away from the rule of law, liberalism and democracy*, postprint, forthcoming in "Ethics & Politics", https://doi.org/10.5281/zenodo.10229060

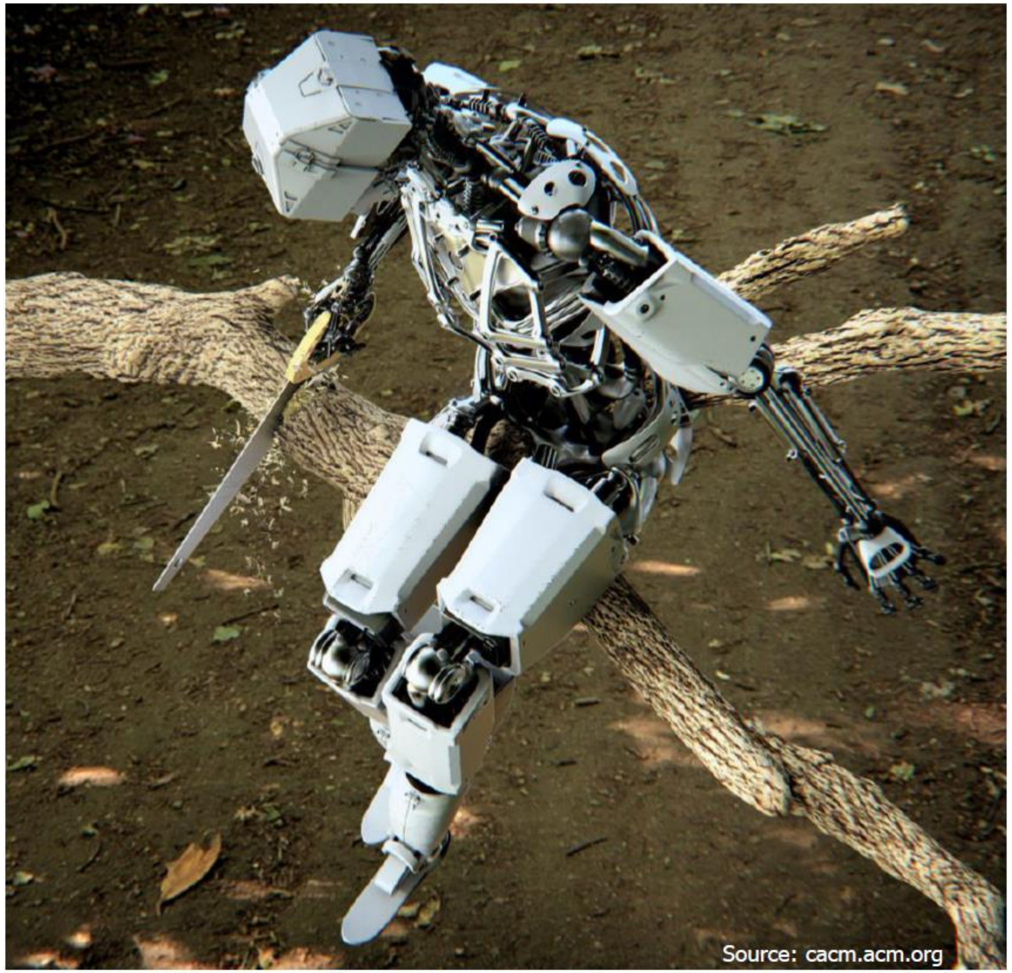
AI ethics

Taken seriously, AI ethics would require a set of conditions, none of which are currently fulfilled.

1. **From an ethical point of view**, it would be necessary to identify normative ethics that does not allow the existence of genuine moral dilemmas - and thus contains the criteria for the solutions to all apparent moral conflicts - and that would be shared broadly enough to make its implementation in machines publicly admitted.

2. **From the metaethical point of view**, it would be necessary to address the question of the translatability into computational terms of the normative ethics adopted, or at least of a coherent subset thereof.

3. **First and foremost, it should be possible to implement the non-moral requirements of AI ethics**: moral judgement requires at least
   a. being capable of acting, not merely according to laws, but also according to the representation of laws
   b. logical reasoning,
   c. a genuine understanding of language,
   d. the ability to distinguish a causal connection from a mere correlation,
   e. the whole family of intuitions and reasoning procedures included in human common sense.



Ernest Davis and Gary Marcus. 2015. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Commun. ACM* 58, 9 (August 2015), 92-103.

Moral judgement would require artificial general intelligence (AGI) and currently no one has any realistic idea of how to implement it.

Therefore, even if we overlook the hard questions of conscience and freedom and set aside the issue of empathy, strictly limiting ourselves to the goals of AI moral reasoning, it is actually obvious, for those not adhering to an animistic conception, that **ML systems are constitutively incapable of making moral judgments**.

Therefore, it should come as no surprise that a recent report on artificial intelligence states that, as the size of models increases, biases also increase and research on AI ethics, which has “exploded” since 2014, has produced many metrics of bias, but with no decrease in that bias.

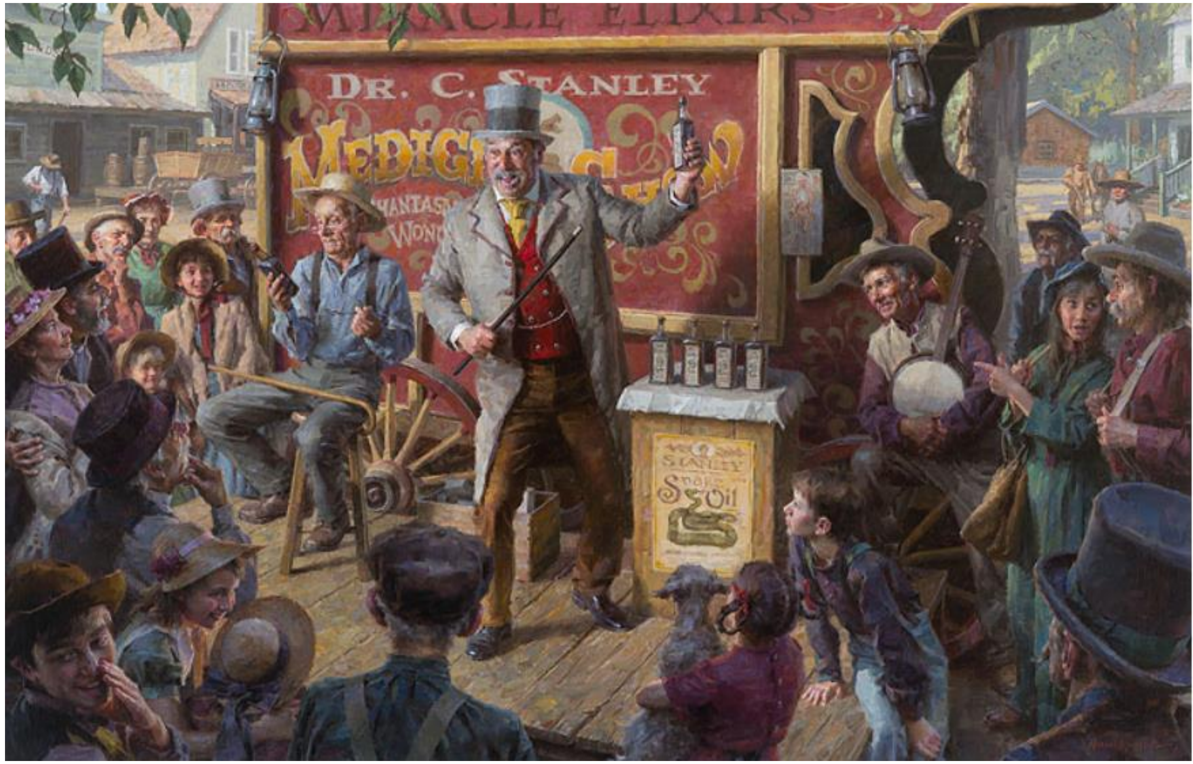
AI and magical thinking

Artificial intelligence is the subject of a constellation of narratives– i.e. of ideas that are spread in the form of stories– which bear three features typical of magical thinking:

1. the tendency to imagine certain objects of technology in anthropomorphic terms;

2. the magicians’ move of showing a result or an effect, while at the same time concealing its concrete causes and costs;

3. the belief that the future behavior of each individual person can be predicted (a belief which, like astrology, is grounded on refined mathematics and a hybrid mixture of superstition and science).



D. Tafani, *What’s wrong with “AI ethics” narratives*, in «Bollettino telematico di filosofia politica», 2022, https://commentbfp.sp.unipi.it/daniela-tafani-what-s-wrong-with-ai-ethics-narratives/.

Animism and dehumanisation

The shift from a figurative sense to a literal sense of language also takes place with “deep learning” systems and “artificial neural networks”, whereby the use of biological metaphors to describe the operations of machines blurs the difference between machines and organisms. Conversely, the computational metaphor that mistakenly assimilates the brain to a computer legitimize, among other “powerful and false ideologies that serve to diminish human and worker rights”, the idea that human beings can be programmed like machines, and therefore governing humans can be equated to a form of cybernetics

A.T. Baria, K. Cross, *The brain is a computer is a brain: Neuroscience’s internal debate and the social significance of the computational metaphor*, 2021, https://arxiv.org/abs/2107.14042

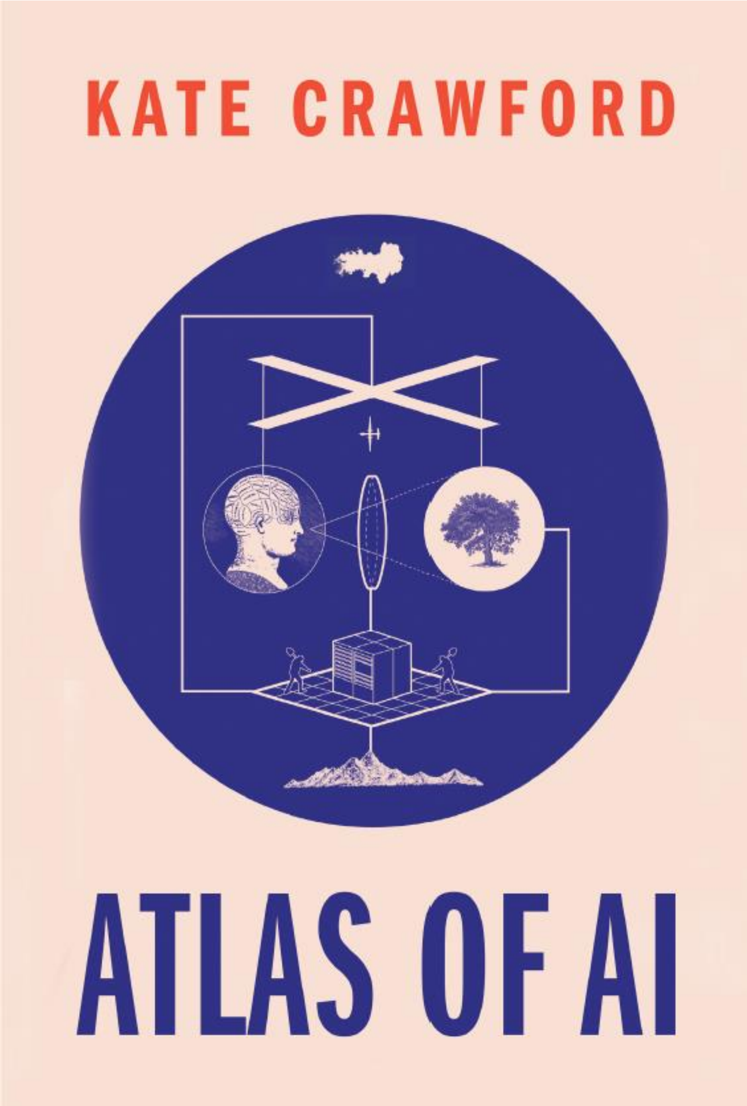
Machine Learning systems

ML systems are deliberately presented in anthropomorphic terms by exploiting, both the first and also the second characteristic of magical thinking: that of showing a result, or an effect, while concealing the material elements of the process and its side effects.

Generally, only three conditions are identified that made ML by its very definition:

1. the exponential growth of computing power per cost unit,
2. the enormous amount of data available in digital form,
3. algorithms.

Artificial intelligence is presented as self-made, with algorithms that “learn” by themselves, extracting value from data, the “new oil” or “new gold,” according to metaphors that imply (thus imposing it as a truism) that data are natural and raw.



ML SYSTEMS

1. Computing power
2. data available in digital form
3. algorithms.

Other “extractions”

4. Earth
5. Energy
6. Labor

K. Crawford, *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven and London, Yale University Press, 2021.

Predictive optimization systems

In an increasing number of areas, judgments and decisions that have major effects on people's lives are now being entrusted to Machine Learning (ML) systems that do not work. In areas such as pre-trial risk assessment, financial services, education, social services and recruitment, the employment of ML systems in assessment and decision making has led to unfair, harmful and absurd outcomes, as documented in an extensive body of literature, with consequences that can rumble on for a long time, sometimes years, in the lives of victims.

Such flaws are not occasional and cannot be prevented by technical interventions. Like faith in the predictions of astrology, faith in these algorithmic predictions vanishes as soon as the modern scientific criteria of communicability and reproducibility are applied.

Closer examination reveals that such systems are unreliable in predicting individual events and actions, to the extent that some researchers have suggested using a lottery rather than ML systems to choose between eligible individuals when resources are scarce and it is not possible to use simple computational methods with relevant and explicit variables. If gender predicts lower pay and skin colour predicts the likelihood of being stopped by the police, then in the transition from prediction to decision such social profiling becomes self-fulfilling, legitimising the biases embedded in the initial statistical description by virtue of the supposed objectivity of the algorithm. Prediction thus produces what it purports to predict.

For references, s. D. Tafani, *Predictive optimisation systems do not work, and they infringe rights. So why do we use them?*, «Journal of Law, Market & Innovation», II, 3, 2023, https://ojs.unito.it/index.php/JLMI/article/view/8890.

“AI ethics” narratives

Since predictive optimization systems do not work, i.e. they are not predictive, the problem of making them unbiased should seem irrelevant: at most, you would have an unbiased system that does not work.

The nonsense of decision making based on automated statistics has anyway been presented by tech companies as a problem of single and isolated biases, amendable by algorithmic fairness, i.e., by technical fulfilment. Fearing a blanket ban, Big Tech has financed, in an evident conflict of interest, a discourse on AI ethics, as a regulatory capture, with the aim of making a merely self-regulatory regime seem plausible. The function of this discourse on AI ethics is to protect and legitimise a surveillance advertising business model. Since the framing of the discourse is determined by its function, AI ethics is peddled within the perspective of technological determinism and solutionism, within the “logic of the fait accompli”

In recent years, ‘AI ethics’ narratives (and their fungible variants, such as ‘value alignment’ or ‘algorithmic fairness’ or ‘AI safety’ narratives) have been widely recognised as mere ‘ethics washing’ and regulatory capture, i.e. as a tool of distraction, to avoid legal regulation, while continuing business as usual.

For all decisions that significantly affect people's lives, the only rational governance of predictive optimisation systems is the same as for any other dangerous, non-functioning product: to ban their use and sale, and to consider claims of the existence of such systems for commercial purposes as misleading advertising .

D. Tafani, *Predictive optimisation systems do not work, and they infringe rights. So why do we use them?*, «Journal of Law, Market & Innovation», II, 3, 2023, https://ojs.unito.it/index.php/JLMI/article/view/8890.

“Self driving” cars and the trolley dilemma

In the contemporary debate about self-driving cars, the trolley dilemma is presented as an unsolved moral problem and as a legal case not yet covered by any law: it is argued that to cope with rare, unavoidable accidents it is necessary to program self-driving vehicles in advance, so that they will choose who to run over, in cases where a fatal injury is unavoidable and where it is certain that each of the alternative maneuvers undertaken by the vehicle will resulting in killing a different victim.

Posing the trolley problem as if it were relevant to existing self-driving cars is like trying to solve the problem of a broken dishwasher that keeps flooding the whole house, through an ethics of dishwashers which will make the dishwasher fair, so that it will be able to decide whose room should be flooded.

Despite repeated announcements over the past decade of the imminent commercialization of self-driving cars, such vehicles still collide against one-third of cyclists and all the cars that are, albeit slowly, in their way. Moreover, evidence is emerging about autopilot being programmed to shut off vehicle control, in case of imminent crash, just one second prior to the impact, so as to blame human drivers. Additional evidence is also emerging regarding self-driving technologies’ failure to detect children on the road.

In the dilemma about self-driving vehicles regarding whom to sacrifice in cases of unavoidable accidents, the question itself is entirely misplaced, both from a legal and from a factual point of view.

AI as technology and “AI” as speech act

We need to distinguish between

1. **artificial intelligence (AI) as a technology with practical application:** “as a technology, AI exists somewhere on a spectrum from, practically, at one end, expert systems, path planners, and practical reasoning systems […] through to, theoretically, at the other end, Alan Turing’s “imaginable digital computers which would do well in the imitation game” or John Haugeland’s synthetic intelligence (i.e., machine intelligence that is constructed but not necessarily imitative)”;

2. **“artificial intelligence” (“AI”) as a speech act with conventional force: “**a social constructor that stems largely from science fiction with computers and robots having hugely overblown capabilities and a tendency to the apocalyptic”. “People have been, and are being, “encouraged” to think about artificial intelligence wrongly”. **Companies “are leveraging “AI” to exert control without responsibility”**.

Peter R Lewis, Stephen Marsh, Jeremy Pitt, *AI vs «AI»: Synthetic Minds or Speech Acts*, in «IEEE Technology and Society Magazine», 2021, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9445758

Thank you.
Any questions?

daniela.tafani@unipi.it