To save us from a Kafkaesque future, we must democratise AI

Stephen Cave

The history of artificial intelligence is entwined with state and corporate power. It must now reflect those it has excluded

Rory Kinnear as Josef K in a stage adaptation of The Trial at the Young Vic theatre, London, 2015. Photograph: Tristram Kenton for the Guardian

Picture a system that makes decisions with huge impacts on a person's prospects – even decisions of life and death. Imagine that system is complex and opaque: it sorts people into winners and losers, but the criteria by which it does so are never made clear. Those being assessed do not know what data the system has gathered about them, or with what data theirs is being compared. And no one is willing to take responsibility for the system's decisions – everyone claims to be fulfilling their own cog-like function.

This is the vision offered to us by Franz Kafka in his 1915 novel, The Trial. In that book, Kafka tells a parodic tale of an encounter with the apparatus of an indifferent bureaucracy. The protagonist, Josef K, does not know why he has been arrested, or what the evidence against him is; no one is willing to take responsibility for the decision, or to give him a proper account of how the system works. And it ends gloomily, with Josef K utterly

defeated, resigning himself to his fate.

Fast forward 100 years and artificial intelligence and data-driven computer systems are frequently portrayed in a similar way by their critics: increasingly consequential, yet opaque and unaccountable. This is not a coincidence. There is a direct link between the trials of Josef K and the ethical and political questions raised by artificial intelligence. Contrary to the hype, this technology has not appeared fully formed in the past couple of years. As the historian Jonnie Penn has recently pointed out, it has a long history, one that is deeply entwined with state and corporate power. AI systems were developed largely to further the interests of their funders: governments, military and big business.

Most importantly, the models of decision-making that these systems sought to automate were taken directly from these bureaucracies. The two great pioneers of machine intelligence, Alan Turing and John von Neumann, both developed their prototypes in the crucible of the second world war. Under Von Neumann's oversight, the very first task in 1946 of the very first general-purpose computer, the Eniac, was running computations for the hydrogen bomb.

In other words, the "intelligence" in "artificial intelligence" is not the intelligence of the human individual – not that of the composer, the care worker or the doctor – it is the systemic intelligence of the bureaucracy, of the machine that processes vast amounts of data about people's lives, then categorises them, pigeonholes them, makes decisions about them, and puts them in their place. The problems of AI resemble those of the Kafkaesque state because they are a product of it. Josef K would immediately recognise the "computer says no" culture of our time.

Of course, there are countless ways in which AI and related technologies can be used to empower people: for example, to bring better medical care to more of us, and to provide access to many other services, from digital personal assistants to tailored online learning.

But at the same time, they risk perpetuating injustice because, for all that they are the newest and shiniest of technologies, they also embody the biases of the past – the reductionist systemic thinking and institutional biases of their origins. By default, these Kafkaesque systems will perpetuate existing forms of discrimination, and even exacerbate them – a case in point being Amazon's now-abandoned recruitment algorithm, which learned from previous records what kind of people the company usually employs, and on the basis of this downgraded new applicants whose CVs indicated they were women.

A crucial step in making the most of AI is therefore to ensure diverse voices are involved in its development and deployment. This means including those who have been excluded from the systems of power from which AI sprang, such as women; or who were colonised by them, such as much of the developing world and numerous communities in the developed world; or who were victimised by them, such as poor or disabled people.

The challenges to this are immense. A report from the World Economic Forum published in December concluded that only 22% of AI professionals globally are women (in the UK only 20%). The situation for people of colour is equally difficult: last month more than 100 researchers were denied visas for travel to Canada to attend NeurIPS, one of the most important AI conferences. Since many were travelling from Africa, this had a particular impact on the "Black in AI" meetings, which aimed to increase representation in the field.

But there is good news, too. Thanks to US-based researcher-activist groups such as the AI Now Institute and the Algorithmic Justice League, the importance of involving marginalised groups is gaining acceptance. In the UK, the newly founded Ada Lovelace Institute has as one of its three core aims to "convene diverse voices" in shaping the future of an AI society. The institute is well-placed to do that: it is independent, yet well enough connected to ensure that those voices are heard; and it can build on the established record of its founder, the Nuffield Foundation, in bringing ethics to science.

Those who have historically been failed by systems of power, such as Kafka – a German-speaking Jew living in Prague – have always been particularly well-placed to recognise their opacity, arbitrariness and unaccountability. Including those voices will therefore ensure that AI makes the future not just more efficient but also more ethical.

· Stephen Cave is executive director of the Leverhulme Centre for the Future of Intelligence at the University of Cambridge

Topics
- Artificial intelligence (AI)
- Opinion
- Computing
- Consciousness
- comment