

AI for better or for worse, or AI at all?

Kobi Leins

The ideal subject of totalitarian rule is not the convinced Nazi or the dedicated communist, but people for whom the distinction between fact and fiction, true and false, no longer exists.

Hannah Arendt

What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people.

Joseph Weizenbaum

When I was a little girl, I was taught a song about a ball of white string, in which the white string could fix everything — tie a bow on a gift, fly a kite, mend things. The second verse of the song was about all the things that string cannot fix — broken hearts, mend friendships — the list goes on. In all of the research I have been doing about Artificial Intelligence (AI), its governance and what it can do, this song has frequently come to mind.

Many authors and researchers are doing the equivalent of repeatedly singing the first verse of the song, about all the things that AI can do, without contemplating where AI cannot effectively or, more importantly, should not, be used. Probably now nearing the height of the Gartner-hype cycle, AI is often

misleadingly touted as being able to fix practically everything. And although it is true that AI will expedite many business processes and result in new ways of acquiring and creating wealth through more sophisticated use of data, for the everyday citizen, those benefits are not always apparent.

The reality of what AI can do is very different from what is conveyed, particularly by industry. In some instances, in fact, AI is breaking things in a way and at a speed that is unprecedented. It is impossible for businesses and governments to simultaneously maximise public benefits, service levels, market competition and profitability. Profitability is almost inevitably being prioritised in a neoliberal context, at the expense of democracy, individual freedoms and the voices of civil society and citizens. Many are voicing these concerns, but these concerns are yet to be actively addressed at all stages of any contemplation of the use of AI.

AI includes a series of component parts; both software and hardware. AI may include: data-based or model-based algorithms; the data, both structured and unstructured; machine learning, both supervised and unsupervised; the sensors that provide input and the actuators that effect output. AI is complicated and encompasses many things. For ease of reference in this chapter, the term “AI” includes all of these components, each of which may require different considerations and limitations. The consideration of each of these pieces is beyond the scope of this brief chapter, but this complexity is important to hold in mind when considering applications of AI.

This chapter will contemplate the current popular dichotomy between techno-utopians (those who think that technology — including AI — will save the world) and techno-

dystopians (those who think technology will destroy it). I conclude that there needs to be a greater space in the discourse for questions, challenge and dissent regarding the use of technology without being dismissed as a techno-dystopian. In fact, these voices are required to ensure safe and beneficial uses of AI, particularly as this technology is increasingly embedded in physical systems and affects not only our virtual, but also our physical worlds.

AI is not inherently good or bad – but it does have a past and a context

The overly simplistic and popular dichotomy often posed is between those who are techno-optimists and techno-dystopians. The reality is far more complex, and creating a notion of “friends” or “enemies” of technology does not foster helpful dialogue about the risks and dangers of developing and using certain applications of AI. Differing interests and profoundly powerful market forces are shaping the conversation about AI and its capabilities. What Zuboff coins “surveillance capitalism” is far too profitable and unregulated to genuinely contemplate human rights, civil liberties or public benefit.¹

Every technology has a history and a context.² A prominent example from Winner’s book, *The Whale and the Reactor*, involves traffic overpasses in and around New York, designed by Robert Moses. Many of the overpasses were built low, which prevented access by public buses. This, in turn, excluded low-income people, disproportionately racial minorities, who depended entirely on public transportation. Winner argues that politics is built into everything we make, and that historical moral questions asked throughout history — including by

Plato and Hannah Arendt — are questions relevant to technology: our experience of being free or unfree, the social arrangements that either foster equality or inequality, the kinds of institutions that hold and use power and authority. The capabilities of AI, and the way that it is being used by corporations and by governments, continue to raise these questions today. Current systems using facial recognition or policing tools that reinforce prejudice are examples of technology that builds on politics. The difference is that, in non-physical systems, politics are not as easy to identify as in an overpass, although they may be similarly challenging to rectify after they have been built, and equally, create outcomes that generate power and control over certain constituents.

The beginning of AI as a discipline

Often cited as the birth of AI as an academic discipline is the 1956 Dartmouth conference, which ran over eight weeks, and which was a product of its time.

A list of conference participants, mainly white, wealthy and educated men, meet:

proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.³

Following this conference, developments in AI moved ahead in fits and starts. The quiet periods are now retrospectively referred to as “AI winters”.⁴ More recently, successes in playing games and facial recognition, among other advances, have attracted considerable attention, both positive and negative. The most rapid advances, however, have been in the growth of companies such as Google and Facebook, which configure and use data derived from AI for the benefit of tracking their users and providing them with particular features.

Corporate use of AI

Many companies use AI to deliberately pursue anti-regulatory approaches to existing legal structures governing business.⁵ These structures exist to ensure that society receives tax to improve the physical world for their citizens; to ensure that safety and privacy are upheld; and to express protections and norms (ideally) created by democratically elected representatives. Some companies using “disruptive” AI technologies are deliberately avoiding these regulations.

No ethical limitations will ever prevent these companies from a model designed to deliberately pursue anti-regulatory approaches. In German, the word for disruptive is *Vorsprung*,⁶ which literally translates to “jump over”. In effect, these companies are jumping over existing regulatory frameworks to pursue market dominance. In some jurisdictions, legal action is being taken to try to mitigate these approaches. In Australia, in May 2019, over 6,000 taxi drivers filed a class action lawsuit for lost income against Uber for its deliberate attempt to “jump over” existing laws regulating taxi and limousine licensing. “It is not acceptable for a business to place itself above the law

and operate illegally to the disadvantage of others,” said Andrew Watson, a lawyer with the claimants’ firm Maurice Blackburn.⁷

Facebook, which collects, stores and uses people’s private data, was recently fined US\$5 billion for breaches related to the Cambridge Analytica scandal.⁸ At the time of writing, Facebook’s market capitalisation was approximately US\$584 billion.⁹ Facebook’s share price increased after the fine to a value that would have covered the fine, most probably due to investor relief that it was clear no further regulatory responses were imminent. The fine, which represents about three months of Facebook’s revenue, also shows that regulators are toothless, unserious, or even worse, both.¹⁰

Use of AI to avoid public governance

It has already been suggested that the use of AI in particular circumstances is not accidental, but rather often a deliberate method to avoid introspection and traditional governance through inexplicable decision-making processes.

*“governance-by-design” — the purposeful effort to use technology to embed values — is becoming a central mode of policymaking, and ... our existing regulatory system is fundamentally ill-equipped to prevent that phenomenon from subverting public governance.*¹¹

Mulligan and Bamberger raise four main points. First, governance-by-design overreaches by using overbroad technological fixes that lack the flexibility to balance equities and adapt to changing circumstances. Errors and unintended consequences result. Second, governance-by-design often

privileges one or a few values while excluding other important ones, particularly broad human rights. Third, regulators lack the proper tools for governance-by-design. Administrative agencies, legislatures and courts often lack technical expertise and have traditional structures and accountability mechanisms that poorly fit the job of regulating technology. Fourth, governance-by-design decisions that broadly affect the public are often made in private venues or in processes that make technological choices appear inevitable and apolitical.

Each of these points remain valid. Use of AI by governments and corporates alike often masks the underlying political agenda that use of the technology enables. In the case of what has been coined “Robodebt” in Australia, the Federal Government’s welfare department, Centrelink, used an algorithm to average a person’s annual income gathered from tax office data over 26 fortnights, instead of individual fortnightly periods, to calculate whether they were overpaid welfare benefits. Recipients identified as having been overpaid were automatically sent letters demanding explanation, followed by the swift issuance of debt notices to recover the amount.

This method of calculation resulted in many incorrect debts being raised. Recipients of aid in Australia are one of the most vulnerable groups in Australia, and even raising these debts without consultation or human interaction arguably caused profound detrimental effects.¹² Senator Rachel Siewert, who chaired the Senate Estimates hearings into Robodebt, noted that, “[t]here were nine hearings across Australia, and what will always stick with me, is that at every single hearing, we

heard from or about people having suicidal thoughts or a severe deterioration in mental health upon receiving a letter.”¹³

The practical onus at law is on any creditor (in this case, Centrelink) to prove a debt, not for the debtor to disprove it. Automating debt-seeking letters challenges a fundamental law derived from long-standing principles of procedural fairness.¹⁴ This is the type of automation that, even if rectified, causes irreversible damage, and is in and of itself a form of oppression.

Use of social media to influence elections in the United States and Brexit have been extensively covered, but more recently, similar tools were used in the 2019 Australian Federal election when advertisements started to appear in social media feeds regarding a proposed (and utterly fictitious) death tax by the opposition Australian Labor Party.¹⁵ Although the impact of this input on the Australian election is not completely clear, it is known that nearly half of Gen Z obtain their information from social media alone, and that 69% of Australians are not interested in politics.¹⁶ These kinds of audiences are prime targets for deliberately placed misleading social media advertisements, based on algorithmically generated categorisations.

Warnings about AI use from the past

If we could agree on clear parameters about how we build, design, and deploy AI, all of the normative questions that humans have posed for millennia remain.¹⁷ No simple instruction set, ethical framework, or design parameters will provide the answers to complex existential and philosophical questions posed since the dawn of civilisation. And the very use of AI, as we have seen, can be a way of governing and making decisions.

Joseph Weizenbaum was the creator of Eliza, the first chatbot named after the character in *Pygmalion*. It was designed to emulate a therapist through the relatively crude technique of consistently asking the person interacting to expand on what they are talking about and asking how it made them feel. A German-American computer scientist and professor at MIT, Weizenbaum was disturbed by how seriously his secretary took the chatbot, even when she knew it was not a real person.

Weizenbaum's observations led him to argue that AI should not be used to replace people in positions that require respect and care.¹⁸ Weizenbaum was very vocal in his concerns about the use of AI to replace human decision-making. In an interview with MIT's *The Tech*, Weizenbaum elaborated, expanding beyond the realm of mere artificial intelligence, explaining that his fears for society and its future were largely because of the computer itself. His belief was that the computer, at its most base level, is a fundamentally conservative force — it can only take in predefined datasets in the interest of preserving a status quo.

More recently, other writers have expressed similar concerns.¹⁹ Promises being made by technology companies are increasingly being questioned in light of repeated violations of human rights, privacy and ethical codes. The United Nations has found that Facebook played a role in the generation of hate speech that resulted in genocide in Myanmar.²⁰ The Cambridge Analytica scandal has, by Facebook's own estimation, been associated with the violation of the privacy of approximately 87 million people.²¹ Palentir, an infamous data-analytics firm, is partnering with the United Nations World

Food Program to collect and collate data on some of the world's most vulnerable populations.²² Many questions are being raised about the role and responsibility of those managing the technology in such situations, not just in times of conflict,²³ but also in times of peacekeeping,²⁴ including a United Nations' response to identify, confront and combat hate-speech and violence.²⁵

Weizenbaum was also a product of his time, having escaped Nazi Germany with his family in 1936. His knowledge, and personal experience, of the use of data to make the Holocaust more efficient and effective inevitably shaped his perception and thinking about the wider use of data and computers. The first computers of IBM were used to create punch cards outlining certain characteristics of German citizens.²⁶ IBM did not enable the Holocaust, but without IBM's systematic punch card system, the trains would not have run on time, and the Nazis would not have been anywhere near as clinically "efficient" at identifying Jews. For this reason, Germans still resist any national census, and they also resist a cashless society such as that which Sweden has wholeheartedly adopted.

Hubert Dreyfus expressed similar concerns. His book, *What Computers Can't Do*, was ridiculed on its release in 1972, another peak of hype about computing.²⁷ Dreyfus' main contention was that computers cannot "know" in the sense that humans know, using intuition, contrary to what IT industry marketing would have you believe. Dreyfus, as a Professor of Philosophy at Berkeley, was particularly bothered that AI researchers seemed to believe they were on the verge of solving many long-standing philosophical problems within a few years, using computers.

We are all products of our time and have a story. My great-uncle was kept in Sachsenhausen in solitary confinement for being a political objector to participating in World War II. I was partially raised in Germany during the tumultuous late 1980s, and the tensions and histories are etched into some of my earliest memories. I returned to Germany in the late 1990s to watch the processing of history on the front page of every newspaper of the day. Watching today how data is being collected, shared, traded and married with other data using AI, knowing its potential use and misuse, it is difficult for me not to agree with Weizenbaum that there are simply some tasks for which computers are not fit. Beyond being unfit, there are tools being created and enabled by AI that are shaping our elections, our categorisation by governments, our information streams, and least importantly, our purchasing habits. My own history shapes my concern about how and why these systems and tools are being developed and used, a history that an increasing proportion of the world do not remember or prefer to think of as impossible in their own time and context.

Conclusion

Although AI as we understand it was conceived and developed as early as 1956, we are only just coming to understand the implications of rapid computation of data, enabled by AI, and the risks and challenges it poses to society and to democracy. Once data is available, AI can be used in many different ways to affect our behaviours and our lives. Although these conversations are starting to increase now, Weizenbaum and Dreyfus considered these issues nearly five decades ago. Their warnings and writings remain prescient.

The question should not be AI for better or for worse, but rather, whether AI is at all suitable, or compatible with democracy, for certain functions. Not all problems can be solved by computers. Like the ball of string, we must consider the things that AI cannot do, as well as the things it can do. Most importantly, civil society needs to be educated and empowered to be able to ask whether AI should be used at all in certain circumstances.

Consideration of these questions requires interdisciplinary teams with an understanding of context and history, enabling the asking of questions and listening to, at times, complicated and uncomfortable answers. Further, corporations, boards and shareholders need to understand that all profitability is linked to trust, which requires inclusion of divergent views, and the asking of these questions. It is not only end users who need to be protected and prioritised, but also civil society and citizens.

Technological solutions, including AI, are part of a bag of potential solutions to human problems. They are not the only ones, and they are not always the most effective. It is important to take a problem-based approach where problems are considered holistically, and not with technology or AI as the automatic solution. It is time to include diversity in gender, race, sexuality and professional experience in our approaches to problems, not because it is mandated, but because it is necessary to approach any problem effectively. Finally, we need to ensure that technology does not become a proxy for democratic governance, nor undermine democratic elections. All of these issues require entirely human solutions.

Acknowledgement. Enormous thanks to Dr Emma Nyhan for her kind review of this chapter.

Endnotes

- 1 Leins is currently an Honorary Senior Fellow at the University of Melbourne, Australia. The views expressed in this article are the individual views of the author. Enormous thanks to Dr Emma Nyhan for her kind review.
- 1 Zuboff S (2019). *The Age of Surveillance Capitalism*. Profile Books.
- 2 Winner L (1986). *The Whale and the Reactor*. University of Chicago Press.
- 3 McCarthy J et al. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Stanford University. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- 4 Crevier D (1993). *AI: The Tumultuous Search for Artificial Intelligence*. Basic Books.
- 5 Horan HH (2019). Uber's path of destruction. *American Affairs*, 3(2). <https://americanaffairsjournal.org/2019/05/ubers-path-of-destruction/>
- 6 Many readers would be familiar with one of Audi's slogans: "Vorsprung durch Technik" — disruption through technology.
- 7 Xu VX, Australian taxi drivers sue Uber over lost wages in classaction-lawsuit. *New York Times*, 3 May 2019, <https://www.nytimes.com/2019/05/03/technology/australia-uberdrivers-class-action.html>
- 8 The Facebook-Cambridge Analytica data scandal broke in early 2018 when it was revealed that the persona data of millions of Facebook\ users had been taken without their consent and used to target political advertising at them. It has been described as a watershed moment in the public understanding of personal data. See Kang C, F.T.C. approves Facebook fine of about \$5 billion, *New York Times*, <https://www.nytimes.com/2019/07/12/technology/facebook-ftc-fine.html>
- 9 Facebook market cap. *YCharts*. https://ycharts.com/companies/FB/market_cap
- 10 Patel N, Facebook's \$5 billion FTC fine is an embarrassing joke. *The Verge*, 12 July 2019, <https://www.theverge.com/2019/7/12/20692524/facebook-five-billion-ftc-fine-embarrassing-joke>

- 11 Mulligan DK & Bamberger K (2018). Saving governance by design. *106 California Law Review*, 697. <https://doi.org/10.15779/Z38QN5ZB5H>
- 12 Karp P & Knaus C (2018). Centrelink robo-debt program accused of enforcing “illegal” debts. *The Guardian*. <https://www.theguardian.com/australia-news/2018/apr/04/centrelink-robo-debtprogram-accused-of-enforcing-illegal-debts>
- 13 Siewert R (2019). What I learned about poverty and mental health chairing the robo-debt enquiry. *Crikey*. <https://www.crikey.com.au/2019/05/31/siewert-centelink-robo-debt-suicide/>
- 14 Carney T (2018). The new digital future for welfare: Debts without legal proofs or moral authority? *UNSW Law Journal Forum*. <http://www.unswlawjournal.unsw.edu.au/wpcontent/uploads/2018/03/006-Carney.pdf>
- 15 Murphy K et al. (2019). “It felt like a big tide”: how the death tax lie infected Australia’s election campaign. *The Guardian*. <https://www.theguardian.com/australianews/2019/jun/08/it-feltlike-a-big-tide-how-the-death-tax-lie-infected-australias-electioncampaign>
- 16 Fisher C et al. (2019, 12 June). Digital news report: Australia 2019. *Analysis & Policy Observatory*. <https://apo.org.au/node/240786>
- 17 Roff H (2019). Artificial intelligence: Power to the people. *Ethics and International Affairs*. <https://www.ethicsandinternationalaffairs.org/2019/artificial-intelligence-power-to-the-people/>
- 18 Weizenbaum J (1976). *Computer Power and Human Reason*. WH Freeman.
- 19 Broad E (2018). *Made by Humans*. Melbourne University Publishing.
- Broussard M (2018). *Artificial Intelligence: How Computers Misunderstand the World*. MIT Press.
- O’Neil C (2016). *Weapons of Math Destruction*. New York Crown Publishers.
- Goodman EP & Powles J (forthcoming). Urbanism under Google: Lessons from Sidewalk Toronto. *Fordham Law Review* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3390610
- 20 UN: Facebook had a “role” in Rohingya genocide. *Aljazeera*, 14 March 2018, <https://www.aljazeera.com/news/2018/03/facebook-rolerohingya-genocide-180313161609822.htm>. Mozur P, A genocide incited on Facebook, with posts from Myanmar’s military. *New York Times*, 15 October 2018, <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>

- 21 Kang C & Frenkel S (2018). Facebook says Cambridge Analytica harvested data of up to 87 million users. *New York Times*. <https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html>
- 22 Palantir and the UN's World Food Programme are partnering for a reported \$45 million. *Privacy International*, 6 February 2019. <https://www.privacyinternational.org/news/2684/palantir-and-unsworld-food-programme-are-partnering-reported-45-million>
- 23 McDougall C (2019). Autonomous weapons systems: Putting the cart before the horse. *Melbourne Journal of International Law*, 20(1).
- 24 Digital Blue Helmets, United Nations, <https://unite.un.org/digital-bluehelmets/>
- 25 United Nations launches hate speech strategy [video]. *SABC Digital News*, <https://www.youtube.com/watch?v=V8DJkGEpddg>
- 26 Black E (2001). *IBM and the Holocaust: The Strategic Alliance between Nazi Germany and America's Most Powerful Corporation*. Crown Publishers. Discussed in Preston P (2001). Six million and counting. *The Guardian*. <https://www.theguardian.com/books/2001/feb/18/historybooks.features>
- 27 Dreyfus HL (1972). *What Computers Can't Do*. Harper & Row. Dreyfus wrote an updated edition in 1992 entitled *What Computers Still Can't Do*.