Is AI guilty of pre-crime?

In the 2002 film 'Minority Report' (based on the 1956 book by Philip Dick), PreCrime, a specialised police department set in the future, apprehends criminals based on foreknowledge provided by three psychics.

I draw a parallel with discussion of AI, where new 'ethical' rules are proposed for machine but not necessarily human decisions[1]; and AI in effect stands accused of pre-crime.

This conclusion is neither warranted, nor I would argue ethical, since it could forego the benefits of innovation, including potentially lives saved. Further, the 'open internet' supported innovation because it allowed innovation without permission, we should maintain this principle with AI.

Focussing on AI, or automated decisions, represents a category error. After all, the humble thermostat makes automated decisions regarding when to heat a home; whilst humans make decisions that may be neither explained nor explainable. We should focus on specific problems in specific contexts, not AI *per se*.

We should fear, and seek to improve, ourselves; not the routine application of AI to automate tasks and improve services and decisions.

Is AI a helpful policy category?

Discussion of ethics and other policy questions in relation to AI presumes that AI is a distinct and useful category for policy purposes. This is far from obvious.

AI can involve recognition (does an image include a cat?), prediction (which way will upper atmospheric winds blow tomorrow?) and decisions (should a car apply its breaks?).

AI can be used to automate existing tasks, and to perform new tasks that may not have been possible or economically viable without AI. AI is also expected to be widely applicable throughout the economy and society, like stream, electricity and computing it is a general-purpose technology.

However, these characteristics do not of themselves raise fundamentally new questions. AI is a tool, like a shovel, a bulldozer or an autopilot – that extends and substitutes for human capabilities. Further, to the extent that AI makes 'decisions', these may have an existing human or machine counterpart.

AI depends on data and models, but so do many pre-AI systems including accounting, insurance, fraud detection, search and social networks. Data and privacy raise a set of policy issues, though these appear distinct from AI *per se* (AI is likely, however, to change the balance of costs and benefits in relation to data policy).

AI is unquestionably topical, much as steam was over a century ago; but we did not have stream ethics or a law of steam (though specific applications did lead to the development of regulation, for example, in relation to rail safety). General horizontal frameworks, such as competition law, also apply to all areas of the economy, including AI.

AI is not an obvious category requiring the application of a distinct regulatory or ethical approach.

---

[1] For example, the New York University AINOW Institute report "AINOW 2017 Report", the House of Lords Select Committee report "AI in the UK: ready, willing and able?" (April 2018), the French strategy for artificial intelligence (March 2018) "AI for humanity" and the European Commission communication "Artificial intelligence for Europe" (April 2018).

Consideration of ethics alongside other policy approaches

It has been proposed that ethics apply to AI, and ethical principles have been applied in other contexts, in particular in relation to human biology including assistive reproductive and genetic technologies.

However, as a rule the policy approach applying to technology and innovation is to allow open innovation 'without permission' coupled with consumer choice, and to apply targeted intervention where externalities or other 'market failures' arise (some of these interventions may be general and would not need to change with a new technology).

The appropriate place of ethics alongside these frameworks for making trade-offs deserves consideration[2], but it would arguably be a mistake to apply a code of 'ethics' in an overarching manner to a wide class of innovation.

It is in general recognised that new technology is neither good nor bad, it is up to us to adapt to maximise the benefits and minimise harm. It is also recognised that it is impossible to foresee many of the consequences, even the main consequences, of a new technology (not only is the development of the technology uncertain, it requires entrepreneurial exploration and consumer response before the viable applications and their implications unfold). We learn and adapt.

When it comes to public policy, there are three broad decision processes: first principles analysis (e.g. the presumption that removing price distortions leads to a more efficient allocation of resources), cost-benefit analysis where we do not have a strong prior as to what would be efficient and political choice.

Ethics, at least as applied to human biology, differs from an economic welfare approach in that certain things are proscribed. A particularly sensitive area relates to altering the human germline – introducing a change to human DNA which is inheritable:

> *"The possibility of human germline engineering has long been a source of excitement and unease among the general public, especially in light of concerns about initiating a 'slippery slope' from disease-curing applications toward uses with less compelling or even troubling implications."*[3]

In relation to possible engineering of the human germline, we are dealing with a question fraught with risks, and which raises a fundamental question about what it means to be human. But is AI, at least the AI that is been used and developed today and in the foreseeable future in this category?

No, though there may be a future in which AI raises similarly challenging ethical questions to germline engineering. Yuval Harari framed a challenging question at the conclusion of his book Homo Deus:[4]

> *"What will happen to society, politics and daily life when non-conscious but highly intelligent algorithms know us better than we know ourselves?"*

There is also concern, amongst some, that an out of control AI could destroy humanity – perhaps even in error in pursuit of a narrow objective.

---

[2] Jean Tirole, Economics for the common good, Princeton Press, 2017.

[3] Baltimore *et al*, A prudent path forward for genomic engineering and germline gene modification, Science, Vol 348 (6230), April 2015. *http://arep.med.harvard.edu/pdf/Baltimore_Science-2015.pdf*

[4] Yuval Harari, Homo Deus – A Brief History of Tomorrow, Harvill Seckler, 2016.

COMMUNICATIONS CHAMBERS

However, we should distinguish AI as a tool for recognition, prediction and automation (including of 'decisions') from more existential questions regarding the future of humanity.

We should not, as we are inclined to do, over attribute intelligence and human agency to machines. We may also dislike what we see, as developments in AI hold a mirror up to humanity, exposing our prejudices and weaknesses. But these things are challenges for us, rather than flaws in AI.

There is a role here for ethics and philosophy, but it would be both a category and policy error to apply a set of ethical principles to AI generally. AI is not guilty of pre-crime.

Reducing bias and discrimination

There is understandable concern not to build bias into AI in relation to protected categories such as race or gender. However, this is nothing new, and society has evolved both in terms of the categories that are protected and the means by which discrimination can be reduced.

The first line of defense against bias in relation to AI-based or augmented 'decisions' is that they will not be exempt from existing law. New ethical principles are not required for this. Further, whilst AI may include 'black-box' elements, so to large extent do human decisions (even to those making them). A check against hidden bias, for humans and AI, is to scrutinise outcomes; and for that one needs transparency of outcomes for human and machine-based decisions.

There is also technical work going into AI, both to remove discrimination[5], which may prove

easier than for humans; and as a check on human decisions.[6]

Bias in human decisions may also be promoted via side-payments – we call this corruption. I was struck by the comment of an Uber driver, previously a mini-cab driver in London, who told me that he preferred Uber because an algorithm decided whether he got a pick-up, not a cab dispatcher, and the algorithm wasn't corrupt. We should not lose sight of the virtues of automated decisions.

Right to explanation

An idea in relation to AI is that one should have a right to an explanation, or to demand that a human rather than a machine reach a decision. But human decisions are not transparent, and explanations are partial or may not be offered at all.[7]

There are domains where we consider that machines, or even for that matter human experts, should not decide or provide a pivotal input to a decision. An example is the criminal justice system, where, in the UK, there is a right to trial by a jury of one's peers, which arguably derives from Magna Carta.

In other areas, assertion of a right to explanation or a human decision may involve trade-offs that are arguably unethical from a broader societal point of view.

An example is the application of deep learning to medical imagery. Not only may AI interpret such images at low cost, but it may offer diagnoses that were previously either not contemplated or possible.

---

[5] Chiappa and Gillam, Path-Specific Counterfactual Fairness, February 2018. https://arxiv.org/abs/1802.08139

[6] Erel, Stern, Tan and Weisbach, Selecting Directors Using Machine Learning, March 2018. NBER Working Paper No. 24435. http://www.nber.org/papers/w24435

[7] Pande, Why we shouldn't fear the 'black box' of AI (in Healthcare and Everywhere), New York Times, January 2018. https://www.nytimes.com/2018/01/25/opinion/artificial-intelligence-black-box.html

COMMUNICATIONS CHAMBERS

For example, the identification of cardiovascular disease from an image of the retina.[8] Another example of is an AI system that can listen to emergency calls and assess risk in relation to heart attacks. But, explaining what patterns it spots and how it weights them is not part of the software's design. Maaløe tells The Verge that Corti's team knows that certain words "have a higher impact on the final output than others," but he says this analysis is necessarily "imprecise."[9]

Let's assume there are applications of deep learning that are clinically proven but allow limited scope for explanation (the challenge of explanation in relation to technology is not limited to AI, but applies also, for example, to general anesthesia which has been used for over 150 years, though the molecular mechanism of action for general anesthetics is poorly understood[10]).

A right to explanation, or a human decision, may either deprive patients of a superior AI diagnosis; or deprive others of diagnosis and treatment as funds that would otherwise be available are used to fund human rather than AI based diagnoses.

Such an outcome would represent both a poor allocation of resources from an economic point of view, and an unethical outcome from a societal point of view. A right to explanation or a human based decision should not be a general right, and the trade-offs involved should be considered.

The human standard – should a regulatory Turing test apply?

The preceding examples point to a general point, namely that some of the standards proposed for AI are higher than those applied to the tasks it will automate or enhance.

A higher standard for AI will necessarily involve trade-offs where society is made worse off; namely where a more costly or inferior approach persists because it is held to a lower standard than AI.

For example, ethical standards have been suggested in relation to the decision rules of autonomous vehicles, yet human drivers are neither asked – nor likely know – what they would do in different situations. At least initially, AI should arguably be held to the same human standard required to pass a driver's test.

What is proposed might be thought of as regulatory Turing test. If a regulator did not know whether they were considering a machine or a human, would they assess that it had passed the existing standard?

Raising the bar, for people and machines

If we do want to set a higher standard, then we should raise the bar for humans and machines. Administrative decision making could be improved, be more transparent and perhaps include some right of explanation.[11] Raising the standard for human decisions might also help clean up the training data for AI.

---

[8] Poplin *et al*, Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning, Nature Biomedical Engineering, Volume 2, pages158–164, 2018. *https://www.nature.com/articles/s41551-018-0195-0*
[9] The Verge, AI that detects cardiac arrests during emergency calls will be tested across Europe this summer, April 2018. *https://www.theverge.com/2018/4/25/17278994/ai-cardiac-arrest-corti-emergency-call-response*
[10] Perkins, How does anesthesia work?, Scientific American, February 2005. *https://www.scientificamerican.com/article/how-does-anesthesia-work/*
[11] The US, for example, has the Administrative Procedures Act 1946 governs the way in which administrative agencies of the federal government of the United States may propose and establish regulations. To protect citizens, the Act also grants the judiciary oversight over all agency actions.

However, raising standards for humans and machines could result in a situation where most, or perhaps all humans, were not up to the standard; say in relation to driving motor vehicles.

The coming dilemma may therefore be, not have the machines met some ethical standard that applies to them alone, but do we allow error prone, biased and corruptible humans to continue to make a particular class of decision?

But rather than a decision that might be delegated to a committee of ethicists, this is an intensely political decision, since people (certain adults) have the right to vote and machines do not.

Conclusion

> "Once we accept our limits, we go beyond them." Albert Einstein

We should not fear the routine application of AI, and AI generally should not be found guilty of pre-crime. We should, however, take the opportunity to use AI to hold a mirror up to humanity, to improve ourselves.

AI is not a meaningful category to which new blanket rules – including ethical rules – should apply. To treat it as such would deny opportunities for innovation and use in relation to a promising new general-purpose technology.

That in turn would forego societal benefits, including potential lives saved; and that hardly seems ethical. It also risks entrenching power in groups of experts - equivalent to the psychics in Minority Report - who may unduly limit innovation, liberty and be insufficiently accountable both in terms of the trade-offs involved and politically.

Where we think higher standards should apply, we should apply them to all decisions, whether by machine or human; and that is likely to be an intensely political decision rather than a detached ethical one – since the lives of voters are directly involved. If we are going to contemplate a future challenge now, it should be this one.

Brian Williamson, April 2018[12]

---

[12] This article was first published in Medium. Brian Williamson is a Partner at Communications Chambers. The opinions offered herein are purely those of the author and do not represent a corporate opinion of Communications Chambers.