AI Beyond the Hype

AI and Ethics

Chris Rees MA, FBCS, CITP

Immediate Past President, BCS – The Chartered Institute for IT

Edinburgh, 31 May 2019

Prof Timo Minssen has discussed a number of ethical issues related to AI in a medical context:

- What happens when AI makes a mistake and things go wrong?
- How existing medical regulations will cope with AI
- The tension between patient data privacy and the need for AI to learn from large scale patient data
- The effect of proprietary systems
- Whether AI systems should have legal personalities
- The effect on employment and the potential for unequal wealth distribution
- Bias and discrimination

As he noted, some, indeed many of these issues have wider ramifications than just the medical field. I shall look at some of the same issues, and one or two others, in a broader, societal context.

Most of the ethical issues that arise with the implementation of AI systems are not unique to AI. I recently spoke in a debate at the WCIT in London with Professor Richard Harvey, Gresham Professor of IT and Professor of AI at the University of East Anglia on the motion, "There is no such thing as AI ethics, just ethics". It became apparent in the debate that there are ethical issues which are specific to AI or which AI renders materially more acute. Today I want to look at the main

ethical issues that arise in relation to AI, and reflect on the extent to which they are legal issues too.

I shall discuss six ethical topics:

1. Bias
2. Explainability
3. Harmlessness
4. Responsibility for harm
5. The effect on employment and on society
6. AIs impersonating humans

I shall also discuss the question, Does ethics matter in AI?

I shall not talk about the trolley problem, because it is boring, old hat, not a real problem even for automated vehicles and was clearly set out by Thomas Aquinas in the 13th century (the doctrine of Double Effect). Nor shall I discuss the ethics of Artificial General Intelligence because I don't think it's going to happen.

1. BIAS.

Bias is definitely not unique to AI. AI systems are biased because we are biased. Bias in AI systems arises from two main causes, unwitting bias in the minds of the engineers who design and build the systems, and secondly bias in the data.

The vast majority of these engineers in Europe and the USA are young, white males. Nothing wrong with that perhaps but they may not even be aware that they are designing systems that work really well for white males but significantly less well for black females. Famously a very sharp, young black researcher at the MIT Media Lab called Joy Buolamwini was offended when Google's facial recognition system identified her as a gorilla, and did not recognise her as a human being until she put on a white mask. So she constructed an experiment with about 300 photos of light skinned males and similar numbers of light skinned females, dark skinned

males and dark skinned females. The system was close to 100% accurate with the first set, light-skinned males, in the 90% range for the second and third and an appalling 66% for the last, dark skinned females.

AI systems learn from the datasets they are trained on, so any biases in the training dataset will be "learned" by the AI. Biases will be embedded in the datasets because we are biased, in many ways, some of which we notice, others we may not be aware of. Then they go on learning from datasets they operate on, which themselves will have embedded biases. The larger the dataset, the greater the incidence of bias.

Does this matter? Not in all applications. In machine translation, you are interested in the quality of translation into the target language. Google translate copes well with translating into French, "time flies like an arrow" offering "le temps file comme une flêche." But not so good with "Fruit flies like a banana" - "les mouches des fruits comme une banane." However gender bias can creep in here too. Turkish has genderless pronouns. Google and other automatic translation engines translate "o bir mühendis" as "he is an engineer", "o bir doktor" as "he is a doctor", "o bir hemşire" as "she is a nurse", and "o bir aşçi" as "she is a cook". This is offensive rather than critical.

But bias certainly does matter in many ways. As many in this audience will be aware, judges in several American states use an AI system called Compas to determine whether to grant bail to alleged offenders and in Wisconsin to help the judge decide the length of a sentence. The system relies on a number of indicators, which do not include race. However it does take into account where the offender or alleged offender lives, and given the racial distribution of populations in American cities, geography becomes a proxy for race. So a black accused who may well not re-offend, given his record, is more likely to be denied bail than a white man with a comparable record. Compas, developed by Northpointe (now

Equivant) is proprietary and Equivant will not divulge how it works. Perhaps they cannot. This is surely unethical.

In the case of Wisconsin v. Loomis, the defendant Eric Loomis was found guilty for his role in a drive-by shooting. Pre-trial, Loomis answered a series of questions that were then entered into Compas. The trial judge gave Loomis a long sentence partially because of the "high risk" score the defendant received from this risk-assessment tool. Loomis challenged his sentence, because he was not allowed to assess the algorithm. The state supreme court ruled against Loomis, reasoning that knowledge of the algorithm's output was a sufficient level of transparency. A legal matter or an ethical one too? Both in my view.

This leads to consideration of another important ethical issue, which has legal consequences as well as more general ones, for instance in finance.

2. EXPLAINABILITY. The most popular forms of AI today are based on deep learning via artificial neural networks. It is a characteristic of such systems that they cannot explain how they reach their decisions. Nor can their designers or developers. This is known as the "black box" problem. They continue to learn, once trained and the problem persists. It is a principle of the Common Law that a judge must explain his decisions. If he is relying on a black box system to make or support his decision, he cannot fulfil that obligation. A legal and an ethical problem.

A comparable problem arises if a bank or mortgage company is relying on a black box AI to determine whether to grant you a loan or a mortgage. If it denies your application and cannot tell you why, then you cannot change your application to enhance your chance of success, for instance by increasing your deposit, because you would not know whether this would meet the system's objections. Many human beings and human institutions refuse to divulge their reasons, but these AI systems cannot do so. A uniquely AI-specific ethical problem. Work is going on to

solve it and make them transparent, for instance at MIT Lincoln Labs, but so far no general solution is available as far as I know.

3. HARMLESSNESS.

In 1942 Isaac Asimov published his laws of robotics in *I Robot*. The first was "A robot may not injure a human being or, through inaction, allow a human being to come to harm". AIs are tools and any tool can be used for good or ill. A knife can be used to cut bread or stab someone. AI can be used for the benefit of mankind or to hurt people. The use of AI-driven drones (let alone stray ones wandering around Gatwick airport) as **Lethal Autonomous Weapons Systems** (LAWS) is already controversial. Many at the UN argue for them to be banned like chemical and biological weapons. Some nations, including the UK, argue against it. Undoubtedly all the major powers are developing such systems and a number probably deploying them already. Could a LAWS abort a mission if its target had entered a hospital or hidden in a group of children? A drone operated by a human could and would be aborted in such circumstances. I doubt whether an AI system could make such a sophisticated judgement. Similarly would an AI guided drone be able to decide whether to crash land in a populated or less populated area? Again possible but I doubt it. Unethical? Yes, but ethics is not always the first consideration in military thinking. Definitely an issue in international law.

Secondly facial recognition, an AI technology, can be and is being used by repressive, authoritarian regimes to facilitate persecution. The most egregious example in the world is the widely publicised use of facial recognition technology by the Chinese government to identify, arrest and incarcerate over a million Uighurs in Xinjiang inn Western China for no misdemeanour other than being Uighurs and Moslems. Totally unethical. We were there in 2016 and it pains us to witness it. The one positive note in this sorry tale is that the Trump administration is considering blacklisting five Chinese companies (including Megvii, Zhejiang

Dahua Technology Co., and Hangzhou Hikvision Digital Technology Co.) which supply these systems to the Chinese government.

More generally the ethical issues raised by the use of facial recognition technology are gaining wide notice. It is useful and convenient that my PC recognises my face so that I don't need to use a password to access the machine; more significantly it has enabled police to identify and arrest elusive criminals including the suspect in a mass shooting in an Annapolis MD newspaper office last June. But San Francisco has banned its use by the police and public authorities and Ed Bridges is suing South Wales Police over its use of the technology without his permission. He will argue that it is an unlawful violation of his privacy, and his rights to free expression and protest and that it breaches data protection and equality laws.

AI can also be used for **criminal purposes**, to take over an autonomous vehicle and turn it into a weapon, to attack infrastructure facilities such as telephone networks and power grids, and to gather the data to increase the effectiveness and frequency of spear phishing. There are many other comparable concerns in IoT. Obviously unethical, indeed criminal. The risks need to be assessed in such situations. What this concern emphasises is the crucial importance of cybersecurity in relation to AI, well covered in an excellent report, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* published in February 2018 by a group of scholars who met at Oxford University the previous February.[1]

They identified that the growing use of AI systems would lead to changes in the landscape of threats:

> Expansion of existing threats. The costs of attacks may be lowered by the scalable use of AI systems to complete tasks that would ordinarily require human labour, intelligence and expertise. A natural effect would be to

[1] https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

expand the set of actors who can carry out particular attacks, the rate at which they can carry out these attacks, and the set of potential targets.

Introduction of new threats. New attacks may arise through the use of AI systems to complete tasks that would be otherwise impractical for humans. In addition, malicious actors may exploit the vulnerabilities of AI systems deployed by defenders. So AIs attacking AIs.

Change to the typical character of threats. We believe there is reason to expect attacks enabled by the growing use of AI to be especially effective, finely targeted, difficult to attribute, and likely to exploit vulnerabilities in AI systems.

In response to this changing threat landscape they made four high-level recommendations:

1. Policymakers should collaborate closely with technical researchers to investigate, prevent, and mitigate potential malicious uses of AI.
2. Researchers and engineers in artificial intelligence should take the dual-use nature of their work seriously, allowing misuse-related considerations to influence research priorities and norms, and proactively reaching out to relevant actors when harmful applications are foreseeable.
3. Best practices should be identified in research areas with more mature methods for addressing dual-use concerns, such as computer security, and imported where applicable to the case of AI.
4. We should actively seek to expand the range of stakeholders and domain experts involved in discussions of these challenges.

4. **RESPONSIBILITY**.

Timo identified the issues arising when the use of AI in a medical context goes wrong. Of course this issue is not limited to the medical field, and is of undoubted

concern to lawyers at this conference. It may be argued that it is a legal rather than ethical issue but in my view it is both. It is most commonly cited in relation to autonomous vehicles (AVs), which, like doctors, and human drivers, can kill people. As far as AVs are concerned, the UK parliament passed a law last July, the Automated and Electric Vehicles Act 2018, which has yet to be implemented. It assigned clear responsibility to the insurer in the case of an AV causing harm or death, assuming that the AV was insured. If not, responsibility lies with the owner. This should ensure that injured parties are compensated relatively quickly, while investigation of the root cause can take its time. The insurer is entitled under the Act to have recourse later to the manufacturer or developer of a failing component, sub-assembly or software module. The Act also has provisions for, for example, failure by the owner to install safety-critical software updates issued by the manufacturer and unauthorised modifications to the software by the user. The Act does not cover the situation where an AV has been hacked and taken over by a hostile party, as I discussed before. Conceptually however, the Act addresses the issue of the absence of a driver who can be held to account. Nevertheless the ethical and legal issues of liability in other domains rest open as far as I know, though work is being undertaken by the EU Commission on potential updates to the Product Liability Directive to take more recent developments like AI into its ambit. Whether and how this will affect us of course depends somewhat on Brexit.

5. Effect on employment and the distribution of the benefits of AI.

Will AI create new jobs or destroy old ones? Undoubtedly both. What are the ethical implications? Like industrial revolutions before it, going back to the first industrial revolution, the so-called 4th industrial revolution will first destroy jobs, replacing human operators by faster, cheaper, more efficient machines that don't take holidays, demand a pay rise or fret about Brexit. In due course they will create the demand for new functions to be performed by humans, that we cannot now even imagine. The problem is the sequence and the unemployment that it will cause in the interim. We cannot estimate the scale of either phenomenon. Some of

AI's most ardent advocates say that there is nothing to worry about. This is wrong. At the other extreme, some have exaggerated the dangers and predicted an employment Armageddon. This is unhelpful. We should be concerned. To displace people from work without helping them to find new work is unethical and harms society.

What is to be done? Retraining is key. There are jobs that AIs cannot touch – user interface design is an example. Who should fund the retraining – government, corporations, individuals? Probably some combination. There are some excellent examples of corporate endeavours in this area. AT&T's Future Ready initiative is a $1 billion, web-based, multiyear effort that includes online courses; collaborations with Coursera, Udacity and leading universities; and a career centre that allows employees to identify and train for the kinds of jobs the company needs today and will need. By 2020 AT&T will have re-educated 100,000 employees for new jobs with cutting-edge skills. BT has a similar programme and the government sponsored Institute of Coding, supported by BT, IBM, Cisco, Microsoft and others as well as 25 universities and my own institute, the BCS, is a very positive initiative in this direction. We cannot sit idly by.

There is another economic and societal ethical issue relating to AI – the risk that the benefits will accrue disproportionately to a privileged few while the costs fall on those lower down the social scale. There is no quick fix for this risk but it is one that policy makers need to have in mind.

6. **Is AI impersonating a human unethical?**

Famously Alan Turing devised the Imitation Game, now known as the Turing test, whether a machine could convince a human being that it is another human being. Until recently no machine has passed the Turing test, though many human beings have failed it. However at the Google developer conference last year, Sundar Pichai, the CEO, demonstrated Duplex, an AI that convincingly called a beauty parlour and a restaurant to make a hair appointment and a table booking

respectively. Neither receptionist realised that they were talking to a machine. It was so realistic. This technology is now live and available from Google. Although the audience at the conference applauded the demonstration, the reaction on social media was that this was unethical. Not to identify that the machine is a machine is unethical. [As Karmen Turk has said,] the EU High Level Expert group has stated that this contravenes one of their principles.

7. Does ethics matter in AI?

Yes, profoundly. If the public concludes that AI or the use of AI is unethical, it will lose public trust. There are many examples of this happening, with or without scientific justification. GM Foods and the Boeing 737 Max are obvious examples. To quote the EU AI High Level Expert Group,

> Trustworthiness is a prerequisite for people and societies to develop, deploy and use AI systems. Without AI systems – and the human beings behind them – being demonstrably worthy of trust, unwanted consequences may ensue and their uptake might be hindered, preventing the realisation of the potentially vast social and economic benefits that they can bring.
>
> They wrote, "To help Europe realise those benefits, our vision is to ensure and scale Trustworthy AI."

The Group identified four ethical principles to which AI should adhere. These are:

(i) Respect for human autonomy

(ii) Prevention of harm

(iii) Fairness

(iv) Explicability

They noted that while many legal obligations reflect ethical principles, adherence to ethical principles goes beyond formal compliance with existing laws. This is a cogent argument.

Ethics is often seen as a bunch of "don'ts". Certainly, there are unethical threats that need to be countered. But ethics should be seen as a positive. We want to buy from companies that we perceive as ethical. People want to work for ethical employers. Academics want to conduct research in an ethical way. Being ethical should be and increasingly is seen as an important element of an organisation's stance and strategy. And ethical considerations need to be front of mind in the development and use of AI at every stage, from conception through design and build to deployment.

To conclude, I have considered six ethical issues in relation to AI: Bias, Explainability, Harmlessness, Responsibility for harm, the effect on Employment and Society, and AIs Impersonating Humans. I have argued that ethics really matter if AI is to be trusted and the benefits of AI are to accrue to society.

How does this affect the law? Self-evidently any law, including regulation, should be based on ethical principles. Drafting good regulation is difficult in such a fast-moving domain; excessive regulation runs the risk of inhibiting innovation. But the EU has shown itself to be adept at it and should continue to set the standard.

In this context it is encouraging to note that 42 OECD countries have just signed an accord to support a global governance framework for AI[2] "Recommendations of the Council on Artificial Intelligence – Principles for Responsible Stewardship of Trustworthy AI". It has no force of law but it's a good start if the signatories act on it.

So for the lawyers here today, the message is clear, we must consider ethics at every stage, whether advising clients or developing or deploying AI in your practices or in the administration of justice, and that consideration must extend from the Senior Partner or Chief Executive all the way down the organisation. In all matters relating to AI, ethics must come first.

[2] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449