The Ethics of Artificial Intelligence

Chris Rees

▶ **To cite this version:**

Chris Rees. The Ethics of Artificial Intelligence. Leon Strous; Roger Johnson; David Alan Grier; Doron Swade. Unimagined Futures – ICT Opportunities and Challenges :, AICT-555, Springer International Publishing, pp.55-69, 2020, IFIP Advances in Information and Communication Technology, 978-3-030-64245-7. 10.1007/978-3-030-64246-4_5 . hal-03194715

HAL Id: hal-03194715
https://inria.hal.science/hal-03194715

Submitted on 9 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

The Ethics of Artificial Intelligence

Chris Rees

Past President of the British Computer Society – The Chartered Institute for IT in the UK
3 Newbridge.Square, Swindon,.SNl .18Y, UK
chris.rees@ficino.org

Abstract. This chapter focuses on the ethics of narrow, as opposed to general AI. It makes the practical as well as the philosophical case for discussion of AI ethics. It considers ethical charters, then discusses the principal ethical issues: bias, explainability, liability for failure, harmlessness, the ethical use of data, whether AIs should have legal personality, the effects on employment and society, and AIs impersonating humans. A case study is presented of AI in personal insurance. It makes the case for regulation of AI and discusses the challenges of enacting regulation. It draws conclusions, that the benefits of AI are so valuable that the ethical risks must be managed, or the benefits may be lost because of the loss of public trust. There are grounds for optimism, notably the public consciousness of the issues, the engagement of governments and the amount of private and public investment in ethical research.

Keywords: Ethics, charters, bias, explainability, liability, harmlessness, data-ethics, legal-personality, employment, regulation.

1 Introduction

Artificial intelligence theories and technologies are not new. The concepts were first elaborated by Alan Turing [1] in 1950. Since then AI has gone through periods of hope, when new developments appeared to offer exciting new possibilities, followed by so-called "winters", when those hopes faded, together with the investment and much of the research.

However the technology has exploded in popularity in the last 20 years, for three main reasons: first, dramatic increases in computing power and storage and corresponding reductions in cost, particularly cloud computing and storage, secondly the growth in the internet, providing access to the huge datasets which AI requires, and thirdly the development of new AI techniques, particularly artificial neural networks and machine learning.

Until recently, few considered that AI posed particular ethical challenges beyond those posed by any other computing technique. However that has changed radically. Not only in academic and professional circles but in the quality press and even the popular press, articles appear frequently, even on the front pages, raising ethical issues in the application of AI. The concerns commonly centre on the ethical risks and the

threats to privacy posed by AI systems, even where they are developed and applied for entirely laudable ends.

In this chapter I shall focus on the ethics of narrow AI, not Artificial General Intelligence (AGI). All current implementations of AI are narrow, in the sense that they are applied to a narrowly focused domain, such as diagnosing cancer or playing chess. They can often do that better than any human, but they cannot then turn their attention to stacking pallets in a warehouse or translating from French to English.

There is much discussion in the literature of AGI, that is AI capable of doing what humans do, turning its hand, as it were, to any task, and far exceeding human capability, with the flexibility of the human mind and using common sense. The concept leads to the notion of "singularity", the point at which, when an AI is implanted in the brain, it is impossible to tell where the AI stops and the brain starts. Gurus like Ray Kurzweil, Elon Musk, and the late Stephen Hawking all predict it is coming, in 20 or 50 years, estimates vary. The alternative position is that it is never going to happen, or at least not for a very long time. Certainly, it is not today's problem. Undoubtedly AI will become smarter, more capable, more effective, but the route to AGI is not a continuum. In this chapter I shall make the case for the discussion of the ethics of narrow AI, consider current ethical charters, discuss the principal risks which arise in relation to the ethics of AI, illustrate some of them through a short case study in the insurance industry, and consider the case for regulation. Finally I shall draw some overall conclusions.

2 Why Discuss the Ethics of AI?

AIs are artefacts, things. They have no ethics, or put another way, they are ethically neutral. It is important that we do not attribute agency to artefacts, a topic that will be further discussed in section 4.6 below in relation to the question of giving AIs legal personality. When we talk about ethics, we are talking about human ethics, the ethics of those who design, develop, deploy and use AI systems. Ethics has been a subject of philosophical debate at least since Aristotle's *Nicomachean Ethics* in 350 BCE, and of course it was the subject of extensive discourse in the much older Hebrew Bible, the Upanishads and other ancient scriptures. There is nothing new about identifying ethical issues in society or in relation to IT in particular.

So why should we consider the ethics of artificial intelligence specifically? There are not only good philosophical reasons to discuss it, but at a practical level we should consider it because of the overarching risk that if AI comes to be seen by the public as unethical, they may lose trust in it and the benefits would be lost. There are precedents. There is no scientific evidence that there is anything wrong with genetically modified foods, but the European and particularly the British public lost trust in them [2] in 2003-4 and rejected them. In the UK this was despite a statement in 2004 by Margaret Beckett MP, then Secretary of State for Environment, Food and Rural Affairs, in the House of Commons saying *inter alia* that "There was no scientific case for ruling out all GM crops or products". And after the fraudulent linking of the MMR vaccine with autism [3] by the disgraced former medical doctor, Andrew Wakefield, vaccination rates for

measles, mumps and rubella have dropped in most countries, dangerously so in some, leading to a rise in deaths from measles, particularly among children.

To quote the EU AI High Level Expert Group [4], "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."

3 Current Ethical Charters.

One way to demonstrate ethical principles and earn trust is by publishing an ethical charter. There are many ethical charters for AI in the market. Indeed there is a risk of corporations "charter shopping" until they find a set that suits their purpose. However the basis for regulation and the safe, proper development of AI has been formulated and published by the OECD as The Principles of AI [5]. 44 governments have signed up to these principles, including all the G20, and including some countries which are not members of the OECD. They don't have the force of law, but they are influential. They are set out below.

AI should be:

1. **Human-centred:** AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.
2. **Fair:** AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
3. **Transparent:** There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
4. **Safe:** AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
5. **Accountable:** Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

These are fine principles, which should inform regulation and be at the heart of those engaged in AI, whether as developers or users. However it is not always so.

4 Principal risks which arise in relation to the ethics of AI

The principal ethical issues and potentially associated risks which I shall discuss are these:

1. Bias
2. Explainability
3. Liability for failure

4. Harmlessness
5. The ethical use of data
6. Should AIs have legal personality?
7. The effects on employment and society
8. AIs impersonating humans

4.1 Bias

Why are AI systems biased? Because we are biased, all of us. We are aware of some biases, not others. Not all are bad. We tend to read newspapers whose views reflect our own. We like people who are like us. Many people are biased against people unlike themselves, foreigners, immigrants, people of different colours or religions. There are many other examples.

Bias gets embedded in AI systems in different ways. For example, the vast majority of AI engineers are young, white males. They may not perceive that the systems they build have a bias and, for example, work better for white males than black females. AI's learn bias from biased data in the training dataset. To the extent that the data reflects the biases in the population, or a segment of the population, the data is biased and so the AI systems will learn that bias and carry it through into live operation. Because of their speed and ubiquity, the bias is spread far and fast.

Does this matter? Not always. In machine translation, you are interested in the quality of translation into the target language. Gender bias can creep in here too. Turkish has genderless pronouns. Some automatic translation engines [6] 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"
"o bir aşçi" as "**she** is a cook"

This is perhaps offensive rather than critical.

Yet gender and racial bias does matter in all sorts of ways. Facial recognition technology (FRT) is one application where it often arises. FRT has been widely used by police forces, in the UK for instance by the Metropolitan Police and the South Wales Police. However it is controversial, because of current inaccuracy, particularly with certain racial groups, as well as raising concerns over privacy. By way of example, its use by South Wales Police was challenged in the High Court [7] of England by Ed Bridges. He lost the case, the court finding inter alia that the current legal regime is adequate to ensure the appropriate and non-arbitrary use of FRT. However this was disputed by the Information Commissioner who expressed reservations about the adequacy of the legal framework. And recently Lord Clement-Jones, Chairman of the Lords Select Committee which produced the report on "AI in the UK: Ready, Willing and Able?" [8] (HL Paper 100), has introduced a Private Member's Bill in the House of Lords which seeks to make it a criminal offence to use FRT for overt surveillance in public places and to require the government to review its use within a year. Such bills seldom become law, but the bill may put pressure on the government to act.

In the USA, IBM's and Microsoft's facial recognition technologies confused dark-skinned people with gorillas. In 2018 Joy Buolamwini, a researcher in AI at the M.I.T.

Media Lab, was not recognised as a human being by the algorithm until she put on a white mask. She conducted an experiment [9] in which she ran the Microsoft AI on 385 photos of light-skinned males, and comparable sets of light-skinned females, darker-skinned males and darker-skinned females. The algorithm got 99% of the light-skinned males right, 93% of the light-skinned females right, 88% of the darker-skinned males and only 65% of the darker-skinned females. When she published her paper, IBM and Microsoft quickly changed their algorithms.

There are many other examples of gender and racial bias. In 2019, Amazon shut down an AI-driven human resources system project [10] because it was perpetuating its male gender bias, by being trained on its recruitment records.

Judges in several American states use an AI system called the Correctional Offender Management Profiling for Alternative Sanctions tool (Compas) [11] 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 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 [12]. Compas is proprietary and Equivant, the company that markets it, will not divulge how it works, asserting that it is a trade secret. Perhaps they cannot divulge how it reaches its conclusions because they do not know.

Perhaps the most conspicuous application of FRT is its use by the Chinese Communist Party in Xinjiang in Western China, to identify and confine some 1.8 million Uighur people in so-called re-education camps. The technology does not need to be very accurate as Uighurs are Turkic people, with features quite unlike those of the Han Chinese. This policy has been widely reported in the Western press [13], and has led to American sanctions [14] on the companies supplying the FRT, but with no apparent effect on the policy to date.

4.2 Explainability

Unlike traditional software programs, AIs based on neural networks cannot explain how they reach their conclusions. Nor can their developers. If a bank is using AI to determine whether to grant you a loan and they decline, but cannot explain why, that is unfair and unethical. You would not know what you had to do to qualify. Similarly with insurance if the insurer declines the risk without explanation. Explainability is the one ethical issue that is unique to AI – discussion of other ethical issues typically goes back to Aristotle.

To return to Compas, in Loomis v. Wisconsin the trial judge gave Eric Loomis a six year sentence for his role in a drive-by shooting, partially because of the "high risk" score the defendant received from Compas [15]. Loomis appealed against his sentence, on the grounds that he was not allowed to assess the algorithm. The state Supreme Court ruled against Loomis [16], reasoning that knowledge of the algorithm's output was a sufficient level of transparency. This is surely unethical. It is a principle of the Common Law that a judge must explain her/his decision.

There is a lot of work going on to solve this problem. The most likely route appears to be an external audit approach, comparable to financial audit. But right now, there is no general solution available.

4.3 Liability for failure

What happens when things go wrong? The question most frequently arises in relation to automated vehicles (AVs) – self-driving cars and commercial vehicles. But the question does not only apply to them.

As far as automated vehicles are concerned, should it be the AV or the 'driver'? In fact in the UK there is an answer. Under the Automated & Electrical Vehicles Act 2018, (AEVA) [17], the insurer (or the owner, if the vehicle is not insured) is liable if the AV causes damage, death or injury. The insurer then has right of recourse against the manufacturer of the vehicle or the developer of the faulty component. The injured party typically needs recompense quickly. The insurers can afford to wait to recover their costs where appropriate until the post-crash investigation has revealed the root cause.

Cover can be voided if the owner has tampered with the system or failed to update safety-critical software. But what if the vehicle's software has been hacked? And if there is a fleet of vehicles, who is responsible for the software updates? If the insurer escapes cover, who is liable? The individual 'driver'? There are unresolved problems in this field, and as yet no case law to resolve them.

There are many other applications of AI where the same issue arises – who is liable when things go wrong? For instance, what if an AI-controlled medical device implanted in the human body fails? Is the surgeon who implanted it liable, or the hospital, or the manufacturer? What about off-road vehicles like tractors? The list goes on. All of these issues of liability exhibit both ethical and legal concerns. There is no case law. In human resource situations in the UK, the Equality Act 2010 [18] will bite, there are similar laws in other countries. Otherwise it is likely that suppliers will seek to decline the consequences of failure by contract, though normally they cannot do so for death or injury.

Two aspects of such risks which can scarcely be overemphasised are the importance of protecting such AIs from failures in cybersecurity, and from inadequate testing. For instance, if a number of automated vehicles were hacked, they could be turned into a potent weapon – cars, buses and trucks have all been already used as weapons in many cities, when driven by human terrorists.

Testing is a particularly difficult task with AI. In simple terms, this is first because they are typically agglomerations of large numbers of software components, which may never have been tested together, even if the individual components have been tested. Secondly, because they often use publicly available open source code, whose testing status may not be clear. Thirdly, the range of use cases for which test scenarios need to be constructed may be vast, for instance for AVs. When manufacturers claim that an automated vehicle has been driven for several million miles, it says nothing about the effectiveness of the testing regime. If the testing of an AI system is inadequate or defective, then its implementation would be unethical.

4.4 Harmlessness

AIs should be harmless. In *I Robot* [19] in 1942, long before Turing's work, Isaac Asimov formulated his Three Laws of Robotics. The first was "A robot may not injure a human being or, through inaction, allow a human being to come to harm". He later added a fourth: "A robot may not harm humanity, or, by inaction, allow humanity to come to harm."

Today there are two ways in which these laws are being breached. First, through the malicious use of AI. AI, like any tool, is ethically neutral and dual use. A knife can be used to cut cake and stab someone. It can be used for good and ill. Why take the risk of burgling a house if you can use technology to steal "from the comfort of your own home", to quote that hackneyed marketing phrase. AI can maximise the effectiveness of such theft, by reducing the cost and increasing the volume of spear phishing attacks, in which detailed information about the victim, harvested from a number of sources, is used to gain his confidence, so that he imports a virus or trojan. Gathering the information is expensive and laborious. AI reduces the cost and effort.

A wide range of such threats were analysed in a report published in 2018, on the malicious use of artificial intelligence [20]. The report highlighted the potential for AI systems to attack other AI systems, for AI to enable higher speed, lower cost and higher frequency attacks on a wide variety of systems including automated vehicles and utilities, and the need to plan and prepare counter-measures. Today AI is being widely used both to attack and defend systems, including AI systems.

The second way is through Lethal Autonomous Weapons Systems (LAWS). Drones are useful tools. They can be used for crop inspection, distributing aid to disaster victims, searching for crashed planes under water and so on. However they can also be weaponised, and if configured in swarms, become an even more effective offensive weapon system. The use of autonomous, i.e. AI-guided drones constitutes a significant ethical issue and, in terms of international law, a legal issue. It is expressly prohibited by the Geneva Convention [21]. A human operator can react to changing circumstances in the target – if he has moved into a hospital or among a group of children for instance, and abort the mission. Could an AI make such a sophisticated judgment?

The British Government has decided not to develop or deploy LAWS [22], even if the enemy does. That policy could change

What if, for instance, the American, Russian, Chinese and perhaps Israeli militaries are developing them, and may be willing to deploy them? And when human-guided drones are already being deployed to great effect, not just by the USAF in Pakistan and Afghanistan but by rival militias in the Libyan and Syrian civil wars, can LAWS be far behind?

4.5 The Ethical Use of Data

All AI applications depend on large datasets. That gives rise to privacy issues. There is a tension and a trade-off, for instance, between the use of medical data for the public good and the protection of personal data. You can readily anonymise data in a dataset, for instance removing name, address and other identifying characteristics. However it

has been shown [23] that if you have two datasets of similar type, then an overlap of less than 20% enables you to de-anonymise them.

Netflix discovered this as long ago as 2009, when it released anonymised movie reviews penned by subscribers [24]. By crossmatching those snippets with reviews on another website, data sleuths revealed they could identify individual subscribers and what they had been watching. A gay customer sued for breach of privacy; Netflix settled.

Ways are being found around these problems. Synthetic data is artificially generated, usually by funnelling real-world data through an algorithm which adds noise to construct a new data set without personal information. The resulting data set captures the statistical features of the original information without being a giveaway replica. This dataset can then be used to train the AI, or provide the data on which it is to operate.

There are other concerns too, of companies misusing the data in unethical ways. For instance using Crispr technology the 23andMe company [25] sells kits with which users can send off samples of their spit for genetic analysis to companies, either to discover more about their ancestry or their ancestry and their future health. This raises concerns that we could lose control of profoundly personal data or unearth unwelcome family secrets. The science of genetics has a long history of abuse by eugenicists, obsessed with the idea of breeding out "inferior" intelligence or ensuring racial "purity". As new Crispr technology opens up a world where embryos might be edited, genetic data needs to be handled more carefully than ever. 23andMe has never suggested it could detect intelligence in people's genes. However, companies such as GenePlaza allow users to upload their genetic data and claim to show how comparatively intelligent they are. Meanwhile, members of the alt-right in the USA have shared their 23andMe results on social media and boasted about their white European ancestry.

From a societal point of view, there is a further risk in this situation, that the benefits of AI will accrue disproportionately to a few, technically capable, wealthy individuals while the mass of the population loses out. Prof. Shoshana Zuboff has pointed out in her book, *The Age of Surveillance Capitalism* [26] that Google and the other huge IT companies use AI to create a new form of capitalism, which she termed 'Surveillance Capitalism', which they dominate and in which individuals willingly if unknowingly surrender their rights to their personal data. She argues that people are very willing to give up their private information in return for perceived benefits such as ease of use, navigation and access to friends and information. The agency we can actively assert over our own futures is fundamentally usurped by predictive, data-driven AI systems. Engaging with the system of surveillance capitalism, and acquiescing to its demands for ever deeper incursions into everyday life, involves much more than the surrender of information: it is to place the entire track of one's life, the determination of one's path, under the purview and control of the market, just as Pokémon Go players are walked, lit by their glowing screens, straight through the doors of shops they didn't even know they wanted to visit, after the company sold virtual locations to the highest bidders, including McDonalds and Starbucks.

4.6 Should AI's have legal personality?

I discussed, in section 4.3 above, AIs that let people down or cause accidents. The question of liability for failure leads on to the question whether AIs should have legal personality. In 2017 Saudi Arabia granted Sophia, a "female" robot, legal personality [27]. No other jurisdiction has followed this example.

The law in Common Law and Roman Law jurisdictions, and many others, recognises natural persons – real people, and corporate persons – limited companies, partnerships and government entities. The latter have legal personality, they can sue and be sued. Essentially the concept is that these legal persons are controlled by natural persons. Should machines, robots, AIs have legal personality?

The case of animals has useful parallels. The issue of the legal personhood of chimpanzees was considered by a New York court in 2015 in Nonhuman Rights Project, Inc. v. Stanley [28], where a writ of habeas corpus was filed by Nonhuman Rights Project, an NGO, seeking the release of Hercules and Leo, two chimpanzees confined in a laboratory at Stony Brook University.

The NGO argued that for the institution of habeas corpus, the law does not define the notion of a person. Given the lack of any precedent concerning the application of habeas corpus to anyone other than a human, the court decided to consider the issue of its application to a chimpanzee. An *amicus curiae* brief was filed in the case by the Center for the Study of The Great Ideas, arguing that under New York law, legal personality is held by humans and certain public and private entities, but the legal personality vested in such non-human entities is justified because they are composed of humans. Thus personhood should not be extended to cover animals.

In its judgment, the court refused to recognise the personhood of chimpanzees because they are neither capable of bearing legal responsibility for their actions, and also are not capable of performing obligations. The court also pointed out that it is the capacity to assume rights and obligations, and not the physical resemblance to humans, that is decisive for recognising the legal personality of a being.

On exactly the same grounds, one cannot argue that a robot equipped with AI has a free will which could lead to commission of prohibited acts with the aim of achieving its own ends. Thus it cannot be ascribed a degree of fault, such as negligence or recklessness. Nor is it possible to hold it liable for damages for its errors, for example as in the case of an accident caused by an autonomous car or malpractice by surgical robots.

The European Patent Office has refused to grant a patent to an AI invented by Dabus, an AI [29]. They said that AI systems or machines do not have any legal personality comparable to natural or legal persons. They can neither be employed nor can they transfer any rights to a successor in title. "Since an AI system or a machine cannot have rights, it cannot be considered to own its output or own any alleged invention and it cannot transfer any rights thereto."

In summary, granting legal personhood would be a bad idea: "My AI just caused you damage. Oh dear, go ahead and sue it." Companies have capital and therefore can pay damages if they lose a case. If a company has little capital, you take care before you contract with it. Robots do not have financial resources.

4.7 AIs impersonating humans

Famously Alan Turing devised the Imitation Game, now commonly known as the Turing Test, whether a machine (then a teletype) could convince a human being on the other side of an opaque screen that it is another human being. Arguably no machine has comprehensively passed the test. However at the Google developer conference in 2018, Sundar Pichai, the CEO, demonstrated Duplex [30], an AI that convincingly called a beauty parlour and a restaurant to make a hair appointment and a table booking respectively. The AI successfully negotiated quite complex conversations, including saying "Ah ha" at appropriate points and correcting a misunderstanding by one of its interlocutors. 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. The EU High Level Expert Group has stated that such behaviour contravenes one of their Principles [31], namely transparency. It also contravenes the OECD Principles discussed above.

4.8 The effects on employment and society

Is AI going to replace us all and abolish work and jobs by doing what we do more efficiently and at lower cost? The answer is no, but it will replace many job functions, and not just repetitive tasks. That will lead to some existing roles becoming redundant. New roles will be created. The difference this time is that it is not just the physical functions that are being replaced but the mental ones. Very few professions and occupations are immune: maybe philosophers and priests, not lawyers or software developers. As noted previously, some medical functions but not yet the role of doctor.

There are two fallacies in this discussion, as Daniel Susskind has pointed out in his book, *A World Without Work: Technology, Automation and How We Should Respond* [32]. First is the "Lump of Work" fallacy. It is not the case nowadays that there is a given lump of work and if AI does some of it there is correspondingly less for humans to do. The amount of work to be done has grown year by year at least since the industrial revolution, even if in the agricultural Middle Ages it was pretty static. Job functions and indeed jobs have been continually destroyed and created, e.g. domestic servants (other than for the super-rich) and punch card operators have gone but there were no data scientists or web designers 50 years ago.

AI will create new work functions that we cannot even envisage now. However what Susskind calls the "Lump of Work Fallacy Fallacy" is important too. We cannot assume that the work that AI will create will be best done by humans. It may be work for AIs. Almost certainly some of it will be. We have no way of predicting the speed with which these changes will take place. It is likely that the destructive force of new technology will precede the constructive phase. It usually does.

What is to be done? The key is retraining. There are functions that AI will struggle to touch such as user interface design, and jobs requiring empathy and physical care

like nursing. Who will fund such training? It will need to be some combination of government, companies and individuals.

A good example of how it can be done is AT&T's Workforce 2020 project [33]. AT&T recognised in 2013 that the company was not going to need the thousands of technicians they had, who could repair wires up poles and down holes, as it would all be fibre. But they would need an army of software engineers that would cost a fortune to hire and train. So they instituted a major retraining programme for the technicians they already had, which has been fully supported by top management and the workforce themselves, working with universities and training companies. It has been very successful. The project saved the company a huge amount of money in redundancy and recruitment costs and wound up with a happier, more productive and secure workforce.

There will be those who are unable or unwilling to learn the necessary skills to thrive in the new environment. What is to happen to them? Unless we address these ethical issues as a society, the resulting unemployment and inequality could lead to societal unrest.

5 AI in personal insurance – a case study

The insurance industry is exploring and slowly taking up AI [34] – in most cases slowly because of the technical challenge of grafting new technology onto legacy systems. This has both benefits and risks for the public and constitutes an interesting case study in the ethical implications of AI.

Benefits include:

- More precise risk assessments, enabling previously uninsurable customers to obtain cover, e.g. because of age or location.
- Greater efficiency in a variety of labour-intensive processes, such as onboarding a new customer. Such efficiencies should lead to lower premiums.
- Better claims management, e.g. detecting fraud – by identifying from social media that a claimant was not where they said they were when the incident occurred, again reducing costs.
- It could enable them to offer novel advisory services, like suggesting safer driving routes or healthier exercise regimes (known as 'nudging').

On the other hand, there are ethical risks:

- Hyper-personal risk assessments could leave some individuals uninsurable, e.g. identifying the potential for cancer by analysing sources which could indicate such a propensity, of which the individual is unaware. The principle underlying the insurance business model has always been the spreading of individual risk among a large population. Such risk assessments go against that principle.
- The use of large datasets may affect privacy.
- Insurers could use AI to model the minimum benefit it would take for customers to renew.
- New forms of advisory service, such as 'nudging' could be intrusive.

Clearly there are ethical as well as legal concerns here.

6 The Case for Regulation

Given the risks I have described, is there a need for regulation? Yes, even the industry is recognising that. Some say that the big companies are trying to mould potential regulation to their business models.

Regulation is difficult to formulate in a fast-developing field like AI. There is always a risk that government will introduce regulations based on a view of the technology which goes rapidly out of date. For instance in the UK, many of the strictures on the use of emails for marketing are caught, not by GPDR but by the Privacy and Electronic Communications Regulations (PECR) [35] which sits alongside GPDR. PECR were promulgated in 2003, but modified after lobbying by the mail order industry based on their then model of postal marketing. In the age of email, the constraint on an email to a customer being classed as marketing and therefore illegal unless consent has been given, even if the purpose of the email is to seek consent, is out of date but still in force.

The British and other governments, the EU, and the OECD (see section 3 above) as well as the Nolan Committee on Standards in Public Life in the UK have all recognised the need for regulation. In its report on Artificial Intelligence and Public Standards [36], published in February 2020, the Nolan Committee concluded that the UK does not need a new regulator for AI but noted that: "Honesty, integrity, objectivity, openness, leadership, selflessness and accountability were first outlined by Lord Nolan as the standards expected of all those who act on the public's behalf. Artificial intelligence – and in particular, machine learning – will transform the way public sector organisations make decisions and deliver public services. Demonstrating high standards will help realise the huge potential benefits of AI in public service delivery. However, it is clear that the public need greater reassurance about the use of AI in the public sector." They highlighted explainability and data bias as two key ethical concerns in relation to the use of AI in the public sector. I would argue that these principles should apply equally to AI in the private and third sectors.

The challenge is to regulate the development and deployment of AI in such a way as to protect the public and the individual without inhibiting innovation, to enact regulations quickly enough to have an impact in the near term and to avoid the regulations being hijacked by the giant corporations.

There are questions too as to how to go about it:

- Principle-based or Rule-based
- Vertical (e.g. cars, pharmaceuticals, medicine) or horizontal (Facial Recognition Technology)
- Local vs international (UK/EU/USA/Australia/Japan/China, etc)

There are no easy answers, but increasing public pressure.

7 Conclusions

At the outset, a contrast was drawn between the benefits of ethical AI and the risks of unethical AI. The benefits are huge and growing for the individual and for society at large. This chapter is being written during the 2020 Coronavirus outbreak. AI is being

widely deployed in the search for medicines and vaccines which may counter the scourge. For instance, in China doctors use AI tools provided by Huawei Technologies to detect signs of Covid-19 in CT scans. In Israel, Tyto Care Ltd. offers in-home medical examinations, using AI to deliver clinical-grade data to remote doctors for diagnosis. Chinese tech giant Baidu Inc. devised an algorithm that can analyse the biological structure of the new coronavirus and made it available to scientists working on a vaccine. AI is also behind biometric identification systems being rolled out by governments to track the virus and enforce lockdown efforts, including temperature screening systems deployed throughout Beijing and CCTV cameras hooked up to facial-recognition software in Moscow. "AI is being used to fight the virus on all fronts, from screening and diagnosis to containment and drug development," says Andy Chun, an adjunct professor at City University of Hong Kong and AI adviser at the Hong Kong Computer Science Society [37]. It is critical that these benefits not be lost. But the risks – and they are risks rather than threats – must be addressed.

Despite the ethical challenges set out in this chapter, which may sound doom-laden, there are grounds for optimism. There are several reasons for this. The weight of public opinion concerned about the ethics of AI, stimulated by articles and television programmes about the issues, may move governments to act. The engagement of governments in the issues is serious. In the UK alone there are a number of government and quasi-government organisations with proper funding, devoted to defining, articulating and addressing the ethical issues. All Masters programmes in AI at British universities have to include a course on ethics. And serious money is being committed to research on the ethics of AI. For instance Steven Schwarzman, Chairman and CEO of Blackstone, has committed $350M to MIT, to be matched by the university, for the creation of the MIT Schwarzman College of Computing [38], focused on the Ethics of AI. He has also given £150M to the University of Oxford [39] for a similar purpose. There is an international consensus (at least in the West) that action is needed.

To summarise the key points,

- There are huge benefits to be derived from AI but also significant concerns, which, if not addressed, could damage public trust in the technology and put the benefits at risk.
- AI is never responsible. Its makers, owners and operators are.
- Human centring is the only coherent basis for AI ethics.
- There is increasing public and government awareness of the importance of the ethics of AI.
- Regulation is needed and there is widespread support for it, difficult as it will be to draft.

What should we do about it? As IT practitioners, we have a duty both as professionals and as members of society to engage in the debate, and to seek to inform it, concerned but not frightened.

References

All references in this chapter were accessed on 16-17 April 2020

1. "*Computing Machinery and Intelligence,* Alan Turing, Mind 59, 433-460 1950 https://tinyurl.com/y8c2juyl
2. *GM food and crops: what went wrong in the UK?* US National Library of Medicine, National Institutes of Health EMBO Report, May 2004, https://tinyurl.com/ycvg4p8t
3. Wakefield's article linking MMR vaccine and autism was fraudulent, The BMJ, 6 January 2011, https://tinyurl.com/ybntmdzd
4. *Ethics Guidelines for Trustworthy AI,* European Commission, December 2018, https://tinyurl.com/y37czxtc
5. The OECD AI Principles, May 2019 https://www.oecd.org/going-digital/ai/principles/
6. e.g. Bing Translator https://www.bing.com/translator.
7. Judgment of the High Court citation no [2019] EWHC 2341 (Admin), Haddon-Cave LJ and Swift J, https://tinyurl.com/yb3g2btl
8. *AI in the UK: Ready, Willing and Able?* House of Lords Select Committee on Artificial Intelligence Report of Session 2017–19, 16 April 2018, https://tinyurl.com/y8yelom9
9. Gender shades : intersectional phenotypic and demographic evaluation of face datasets and gender classifiers (PhD thesis). MIT. https://tinyurl.com/y3xxuye9
10. Amazon scraps secret AI recruiting tool that showed bias against women, Reuters, October 2018, https://tinyurl.com/y8eelatr
11. https://www.equivant.com/practitioners-guide-to-compas-core/
12. The accuracy, fairness, and limits of predicting recidivism, Science Advances, January 2018, https://tinyurl.com/u2saony
13. For instance: Absolutely No Mercy': Leaked Files Expose How China Organized Mass Detentions of Muslims, New York Times, 16 November 2019, https://tinyurl.com/vavm8d3 *What happens when China's Uighurs are released from re-education camps,* The Economist, 5 March 2020, https://tinyurl.com/tcqyvz9
14. China's 'Abusive' Facial Recognition Machine Targeted By New U.S. Sanctions, Zak Doffman, Forbes, 8 October 2019, https://tinyurl.com/yays44th
15. A Popular Algorithm Is No Better at Predicting Crimes Than Random People, The Atlantic, January 2018, https://tinyurl.com/ycef9mqv
16. STATE of Wisconsin, Plaintiff–Respondent, v. Eric L. LOOMIS, Defendant–Appellant. FindLaw, July 2016, https://tinyurl.com/y6wgq8x5
17. *Automated and Electric Vehicles Act 2018,* legislation.gov.uk, https://tinyurl.com/y99d8dpg
18. Equality Act 2010, legislation.gov.uk, https://tinyurl.com/3y6kuja
19. *I, Robot,* Isaac Asimov, originally published as a series of short stories between 1940 and 1950. Compiled into a book by Gnome Press, September 1950, https://tinyurl.com/ybzxpeur
20. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, Arxiv, February 2018, https://tinyurl.com/y9cvemk7
21. Amidst new challenges, Geneva Conventions mark 70 years of 'limiting brutality' during war, UN News, August 2019, https://tinyurl.com/y7akhqev
22. The United Kingdom and lethal autonomous weapons systems, Article 36, April 2018, https://tinyurl.com/yb8bauz5
23. *Who's Watching? De-anonymization of Netflix Reviews using Amazon Reviews*, Maryam Archie, Sophie Gershon, Abigail Katcoff, and Aaron Zeng, MIT, May 2018, https://tinyurl.com/yau6d57m

24. Financial Times, *The promise of synthetic data*, Anjana Ahuja 4 February 2020, https://tinyurl.com/y88flar4
25. Financial Times, *Anne Wojcicki: 'This is the way the world is going'* April 10, 2020, https://tinyurl.com/ybuw7yzq
26. *The Age of Surveillance Capitalism,* Shoshana Zuboff, Profile Books Ltd, 31 December 2019, https://tinyurl.com/ya9bdf8r
27. Everything You Need To Know About Sophia, The World's First Robot Citizen, Zara Stone, Forbes, 7 November 2017, https://tinyurl.com/y7zo2na9
28. *Matter of Nonhuman Rights Project, Inc. v Stanley*, New York State Law Reporting Bureau, 29 July 2015, https://tinyurl.com/ycqk2ztz
29. *EPO refuses DABUS patent applications designating a machine inventor*, European Patent Office, 20 December 2019, https://tinyurl.com/wshmc9u
30. Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone, Google AI Blog, 8 May 2018, https://tinyurl.com/yasguzo5
31. *Ethics Guidelines for Trustworthy AI,* High-Level Expert Group on Artificial Intelligence (AI HLEG), 8 April 2019, https://tinyurl.com/y37czxtc
32. *A World Without Work: Technology, Automation and How We Should Respond,* Daniel Susskind, Allen Lane, 14 January 2020, https://tinyurl.com/y8sk28xr
33. *AT&T's Talent Overhaul,* John Donovan and Cathy Benko, Harvard Business Review, October 2016, https://hbr.org/2016/10/atts-talent-overhaul
34. *Snapshot Paper – AI and Personal Insurance*, Centre for Data Ethics and Innovation, 12 September 2019, https://tinyurl.com/y8f69qvu
35. *Guide to Privacy and Electronic Communications Regulations,* Information Commissioner's Office, https://tinyurl.com/jpkw4kr
36. *Artificial Intelligence and Public Standards,* the Committee on Standards in Public Life (The Nolan Committee), February 2020, https://tinyurl.com/ya5o5ysk
37. The Virus Gives AI a Chance to Prove It Can Be a Force for Good, Bloomberg Businessweek, 7 April 2020, https://tinyurl.com/ybm45b62
38. The MIT Stephen A. Schwarzman College of Computing aims to address the opportunities and challenges presented by the ubiquity of computing — across industries and academic disciplines — perhaps most notably illustrated by the rise of artificial intelligence, MIT, October 2018, https://computing.mit.edu/
39. University announces unprecedented investment in the Humanities, University of Oxford, 19 June 2019, https://tinyurl.com/ydbbam7j