



ETHICS AND AI A MORAL CONUNDRUM

Charlotte Walker-Osborn, Partner and Head of Technology Sector (Eversheds Sutherland (International) LLP) and Callum Hayes, a regional lawyer on secondment from ARM Limited, give us an insight into some of the key ethical issues surrounding AI technologies.

Undoubtedly, there is huge potential for advancement and benefit to society arising out of current and future application of artificial intelligence (AI) technology. However, advances in AI raise complex legal and ethical questions. If using AI technology causes unintended negative consequences based on bias, then this may affect the adoptability of future AI technologies. Therefore, ethics is fundamental to the success of AI.

AI is redefining increasingly complex tasks that may have previously only been achievable by humans – for example, driverless cars and certain robots. Facebook's chief technology officer, Mike Schroepfer, explained that the 'power of AI technology is it can solve problems that scale to the whole planet,' including climate change and food shortages (Will Knight, *MIT Technology Review*, 2016).

Although full autonomy is (largely) not yet a reality, key ethical issues must be considered now so we are ready for future technological development before laws lag too far behind the technology. In the UK, the prime minister has launched a new advisory body at the World Economic Forum in Davos to seek a 'safe and ethical' artificial intelligence (BBC, 2018). Numerous other countries are looking at

ethical issues too. So, how should this be achieved?

How do we align the aims of autonomous AI systems with our own?

The classic dilemma we often read or hear about is how AI in autonomous transportation will decide 'who to save'. If we are driving ourselves, it is accepted that our instinct may be to save ourselves – for example, swerving away from an oncoming car onto a pavement.

Key ethical issues must be considered now so we are ready for future technological development before laws lag too far behind the technology.

Of course, this may cause more damage to others than if our own vehicle had taken the impact.

Whilst there seems to be forgiveness/acceptance of this human survivalist instinct, AI technology is mooted to be able to save thousands of lives a year by being able to make the right choice to (overall) cause the least impact to human lives. However, this only works if countries, insurance models and users of transportation accept the same models and associated choices.

What if different countries wish to apply different ethical standards / models?

If drivers prefer to choose a vehicle which puts their interests first, can manufacturers ethically build this? And how will countries and laws create a fair ethical playing field and compensation rules? This 'dilemma' can just as easily be applied to an autonomous fleet of ambulances in a city choosing where to dispatch themselves or to anti-defence AI technology choosing which missiles to favour shooting down.

These are challenging ethical decisions. If AI has at its fingertips, more information on the likely outcomes and still 'chooses' 'unethically', how will this be viewed?

How do we align the aims of autonomous AI with the right ethical programming and teaching?

If countries already value life and the treatment of behaviour (criminal or otherwise) differently, how can we define a uniform approach just because AI is involved? Precisely how an AI system

should react and make ethical decisions is even harder because it is very often fact specific, and the AI often has the ability to call upon similar fact specific examples to make its choices but its choices may not be viewed with the same understanding as human flaw.

A balance needs to be struck between certainty and flexibility. Programming an AI system may offer the certainty that an AI system will act and align with our aims. However, flexibility (in the form of greater autonomy for a system to think for itself) may be required in a time pressured environment because quick decision making is required and human involvement may slow down this process.

How do we prevent learning algorithms from acquiring unethical biases?

Whilst AI systems may promise to be more ethical and be better decision makers than

and with differing approaches across the world?

This shows that we must be careful about the data that we give to AI systems to analyse and, at times, we need to clearly define what its goals are to avoid it drawing unethical outcomes.

How should humans treat machines that can think and act autonomously?

Touching specifically upon machine learning, there is AI now that can learn how to play chess in less than four hours and play autonomously and beat a chess champion. What would happen if most machines become smarter than us and use that for more than just playing games?

Imagine a medical AI robot is tasked with caring for patients. If the robot dispenses the wrong medication to a patient this could have negative consequences. On a narrow view of ethics,

We need to clearly define what its goals are to avoid it drawing unethical outcomes.

us, there is still a question around how we manage that learning and prevent systems taking decisions which are unintended by their creators and are difficult to foresee. For example, in 2015, a well-known technology provider's learning algorithm mistakenly identified photographs of a certain racial classification as gorillas. This was not intended and illustrates how important it can be for AI systems to be created or trained in a manner so as not to misunderstand the world.

In a similar manner, if a learning algorithm is used to find the best interview candidates for a particular advertised role, what if many more men have applied for the role than women? This information in itself is not harmful or unethical. However, if an employer is trying to improve gender diversity in the workplace but the model does not account for this, there may be a situation where no women or few females are interviewed. There are of course difficult ethical considerations anyway for positive discrimination but how do you develop AI technology to cope with this

it seems hard to imagine how a robot could be held to the same level of accountability as a human. Yet, if the robot has the potential to damage someone then should it be treated/punished in the same way as a human? Any other outcome seems unjust and could lead to more unethical behaviour as robots could be used as a vehicle to commit crimes without any deterrence. Should we make a machine accountable or its creator/programmer, or the person who applies its intelligence to a use which may be improper? What happens where a robot is given its own citizenship (as has happened in at least one country already)?

We need to keep our minds open to the fact that AI machines in the future could have the same capacity as humans, and need to address how to make them accountable and deter unethical behaviour.

How do we prevent ethics shopping?

Assuming we can reach agreement on appropriate ethical guidelines, we also need to ensure that this is regulated effectively, both in law and at a policy level.

If this is not done carefully, then it could lead to ethics shopping.

The problem of ethics shopping is driven by differences between countries who have different laws and approaches towards the regulation of AI. Policymakers at a European Union level are concerned about how the ethical, societal and legal challenges are being handled by different member states.

They have described the work to provide solutions to these challenges as a 'patchwork of disparate initiatives' (European Commission, March 2018). The European Union wants to create ethical guidelines to assist the development of AI based on the EU's Convention on Human Rights.

For example, some countries, such as the UK, USA and Germany, are prioritising rules and, at times ethical guidelines, for self-driving cars on public roads, but others have not yet dealt with it. The risks with an uncoordinated, unbalanced approach towards the regulation of AI and ethics could give rise to resources and development being targeted and relocated to countries which have 'easier' or lower ethical standards. Conversely, it could lead to more powerful countries shaping what is considered ethical behaviour for AI, which can lead to bias in itself.

Ethical choices

The ethical issues with AI are broad and becoming increasingly important to tackle. Humans must define one or more ethical frameworks for AI technologies to help AI make ethical choices. It is clear that a joined-up discussion is needed and a consensus required across different industries and countries to ensure a coordinated approach (at least on certain topics).

The accelerating pace of developing AI technology in our lives make this discussion more pressing as the technological innovations will not halt or slow down and the ethical challenges are only likely to grow.

The positive impact that AI could have on society means there must be a societal imperative to get this right.