Artificial Intelligence – time to get regulating?

The buzz regarding the potential for artificial intelligence ("AI") to revolutionise our lives is inescapable. Development of AI technology is a huge growth area, and investors are banking on an "AI boom" in everything from cybersecurity and healthcare[1]. The capabilities and achievements of AI in some areas are certainly astonishing – self-driving cars are no longer theoretical but a reality, and AlphaGo is now arguably the strongest Go player in history[2]. But the picture isn't all rosy, which the Economist has recently described as a 'Techlash' against the digital giants. As with any technology, there are negative as well as positive effects of AI. Applications of AI in social media can help us find long-lost friends, but those same channels can be manipulated to disseminate fake news and influence our decisions. Are these and other similar worries matters of public concern that warrant a societal response? AI applications, whether it's a smart city, logistics management or build to order (BTO) and just-in-time manufacturing can be optimised to increase efficiency but who should take responsibility when automated processes cause harm?

In light of these emerging externalities, many, Elon Musk included, have called for structured regulation of AI to manage and control the risks. But is regulation the answer? We explored some of the current issues and debates with Karen Yeung, Interdisciplinary Professorial Fellow in Law, Ethics and Informatics at the University of Birmingham.

A common concern often voiced about regulation of new technologies is that if we intervene with regulation too early, we might impede the potential of the innovation by imposing controls, constraints and additional expense. This can deter investment and discourage development of the technology. On the other hand, if we delay in taking action because we are concerned about inhibiting development, then it may become too late to do anything about the risks and harms that have already been generated. Karen agrees this so-called 'tech control dilemma' (or Collingridge dilemma, as it is sometimes called) presents a real challenge for AI. She also warned, however, that pitting innovation and regulation against one another in those terms creates a false dichotomy. In her experience there are both beneficial innovations, which we should encourage, as well as unhelpful or potentially harmful innovations, the effects of which we should aim to mitigate.

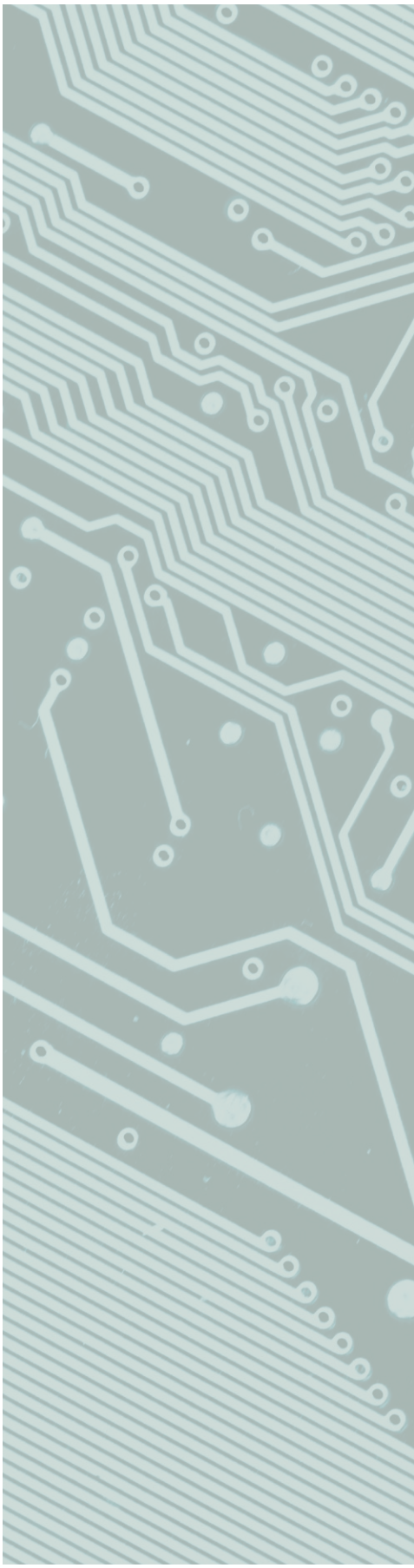
> “A serious challenge is managing for bias in the underlying data on which most AI algorithms are built.”

> It is possible that through use of AI, our environment will become so smart and pre-emptive that all of our choices will be structured and manipulated in ways that will ultimately reduce our capacity to make authentic free independent choices.

Considerable public attention has been given to risks that Al might pose that are more existential in nature. For Karen, this is not primarily a concern about fears that we will create some kind of general Al taking over the world. Rather, "*it is possible that through use of AI, our environment will become so smart and pre-emptive that all of our choices will be structured and manipulated in ways that will ultimately reduce our capacity to make authentic free independent choices*." We are already seeing the impact of this through the echochamber of social media, dissemination of fake news and misinformation, all of which have the potential to shape our environment and influence our collective beliefs and actions.

On that basis, the case for regulation in theory sounds convincing. But will it be achievable in practice? AI is a universal technology that crosses borders and disciplines. Establishment and implementation of a regulatory framework for this complex area is no simple task, and in the view of some, almost impossible. Karen does not share this view, noting that China has done an extraordinarily effective job at regulating access to the internet from mainland China, providing an illustration of what can be done with focused efforts and resources, although she hastened to add that she was certainly not advocating state censorship along the lines adopted in China. Nonetheless, Karen remarked, "*I don't really buy the argument that it's completely unreasonable to regulate AI or that it's too big a task to take on. It's certainly challenging, and I think that technological mechanisms for ensuring that certain standards are implemented will be vital, given the scale on which these technologies operate, but I don't buy the argument that it's too difficult for us to get a handle on it. Where there's a will there's a way*."

A critical aspect of any regulatory framework for AI will be allocation of responsibility for AI based outcomes. In Karen's view, the assessment is relatively straightforward where the AI decisions are subject to meaningful human review – the human reviewer takes ultimate responsibility. It becomes more difficult when you fully automate the decision, like who gets a loan, or who gets a job interview, and so on, if these decisions are not subject to human review. In the classic example of a self-driving car which is programmed to take decisions based on an in-built risk assessment process, who takes responsibility for the outcome of those decisions? The data scientist that designed the relevant algorithms? The car owner? The manufacturer? The car occupants? Karen observes that these debates are challenging our intuitive and long understood social conventions about how we should attribute responsibility. She explained that the behaviour of AI systems that are capable of learning is emergent and therefore unpredictable, and that the AI might be harnessed by bad actors for malign purposes, or simply used in contexts for which it was never intended. Arguably therefore, software developers and others in the supply chain should not be responsible where they could not have reasonably foreseen a particular outcome. The difficulty with that position is that the resultant loss may then lie with the innocent victims, a position that Karen finds untenable. "*The solution,*" she said, "*may be to think about allocation of risk in different ways than current social conventions dictates. For example, in the case of driverless cars, maybe the right answer is to have a mandatory insurance scheme that would bear the risk such that none of the software developer, nor the manufacturer, nor the victim, bears liability*."



Is regulation the right tool to achieve this? If allocation of liability is essentially the result of a social contract, is there a risk that in regulating the development and use of AI, we impose social norms which, on the plus side, might embody certain ideals, but which fail to keep pace with the changes in fast developing technologies and which only serve to reinforce biases? Karen points out that the claim that 'technology outpaces law' does not imply that we should therefore forgo attempts to mitigate against the serious and genuine risks that these technologies may generate. She also explained that "*regulation is meant to promote certain kinds of objectives and values, so to the extent that bias is just another word for embodying particular values then I think that's unavoidable. The important thing is that those objectives and values should be articulated and transparent, and subject to democratic deliberation*." In her view, a more serious challenge is managing for bias in the underlying data on which most AI algorithms are built. She said that we see historic forms of discrimination inherent in the underlying data and these biases

are then replicated due to the way the algorithms operate. Karen referred to one study that found that men were shown high paying job ads six times more often than they were shown to women because historically women have been statistically less likely to apply for high paying jobs than men. The algorithm that generated the ads was based on its analysis of historic data, which showed that women were not placed in high paying jobs and thus *inferred* from these historic patterns that women are thus less interested than men in high paying jobs. She explained that, "*this kind of bias is really problematic because our society has historically discriminated against certain vulnerable groups. I'm not sure whether you can in fact correct for that kind of historic bias that is embedded into our social structures. The data available to us includes these inherent biases and if we tried to correct it there wouldn't be any data upon which to train an algorithm.*"

Part of the problem in regulating AI is that if we rely on AI to make decisions we do not always know how the AI system reached that decision, and so it becomes difficult to explain the process behind how certain decisions were reached. This is critical in the context of any regulatory framework, where, absent strict liability, the ability and opportunity to give reasons and justifications for taking certain actions is central to the allocation of liability. In Karen's view, the importance of explainability varies according to context. How and why an algorithm concluded that it should recommend the purchase of a book or similar item is probably of little consequence to most people, whereas in contexts such as the provision of legal advice, medical diagnosis, and parole releases, explainability becomes extremely important and the decision-makers need to be able to offer an explanation that they can defend. As Karen highlighted, in those highly consequential contexts "*it's unlikely to be acceptable just to say, 'well the machine said so!'* ". But in order to do this Karen suggested that we need to get much better at a formal mathematical verification systems and testing. Her view is that, alongside any regulatory framework, we need to develop robust methods for verifying the validity of the outcomes, and that these methods need to be available and accessible to professionals in all sectors, not just data scientists and coders.

In light of these issues concerning the influence of social conventions on the development of regulation, we asked Karen if we are leaning towards AI regulation on a piecemeal basis, with individual countries developing their own standards and approaches, having regard to individual countries', culture, customs, and existing legislative frameworks. Karen responded that in an ideal world we would have global cooperation on some core baseline principles, whilst allowing scope for divergence where that is culturally and politically legitimate and appropriate, but it seems unlikely that we will be able to coordinate a truly global approach. For example, as between the UK and the US, the US approach to regulation of risks such data privacy and maintenance of individual freedom of choice, is much less robust than the European approach. On that basis she does not realistically see that there will be "*any serious regulatory collaboration across the Atlantic.*"

Karen's views are consistent with others working in this space. Various themes are emerging with respect to the shape of AI regulation concerning transparency, accountability, obligations to manage for bias in the algorithm or underlying data and provision of mechanisms for testing and verifying AI based outcomes. Nesta has gone one step further by putting together a suggested set of standards for the use of AI by public sector organisations[3]. These include requiring that any use of AI is accompanied by a description of its function, objectives and intended impact, ensuring that where AI is deployed, a human being takes responsibility for the outcomes of AI decision making, and publishing risk assessments for mitigating potential biases.

So Elon Musk may be right that AI represents a public risk, and it may now be time to put serious thought into nature of the externalities generated by AI in order to develop a regulatory framework to manage those externalities. As Karen emphasised, even if those risks are unlikely to materialise in the form of James Cameron's The Terminator, "*AI has the potential to erode our autonomy and our freedom in ways that we might not even notice, if it is allowed to continue unchecked, unexamined and unregulated.*"

“
AI has the potential to erode our autonomy and our freedom in ways that we might not even notice, if it is allowed to continue unchecked, unexamined and unregulated.
”

Richard Diffenthal
Partner, London
T +44 (20) 7296 5868
richard.diffenthal@hoganlovells.com

Helen McGowan
Associate, London
T +44 (20) 7296 5581
helen.mcgowan@hoganlovells.com

Professor Karen Yeung
Birmingham Law School
Interdisciplinary Professorial Fellow in Law, Ethics and Informatics
T +44 (0)121 414 6298
pa-ipflawandethics@contacts.bham.ac.uk

Karen Yeung is the University of Birmingham's first Interdisciplinary Chair, taking up the post at the University of Birmingham in the School of Law and the School of Computer Science in January 2018. She has been a Distinguished Visiting Fellow at Melbourne Law School since 2016. She is a Fellow of the University's recently established Institute for Global Innovation (IGI), leading the 'Responsible Artificial Intelligence' challenge theme which supports over 50 researchers from a wide range of disciplines from across the University.