

Artificial Intelligence: Is Johnny 5 Alive? Key Bits and Bytes from the UK's Robotics and Artificial Intelligence Inquiry

Emma L. Flett

Kirkland & Ellis International LLP

Jennifer F. Wilson*

Kirkland & Ellis International LLP

☞ Artificial intelligence; Regulation

Once the preserve of the silver screen, sci-fi novels and overactive imaginations, robots and artificial intelligence (AI) are now living among us, and will soon fundamentally change the way we live, work and play. While such cutting-edge technology represents a new and exciting chapter, it also presents new challenges—social, legal and ethical. There is a clear need to maintain some degree of control, so that the things of dreams do not transform into those of nightmares. No matter what your generation—whether you grew up with the never-blinking, all-seeing eye of HAL 9000 from the sci-fi classic *2001: A Space Odyssey* in the 1960s; R2D2 and C3PO from *Star Wars*; Johnny 5 from the 1980s cult-classic *Short Circuit*; Wallace and Gromit's "wrong trousers", the cyberdog Wendolene and contraption for a "cracking breakfast"; or Pixar's very own small robot with a big heart, Wall-E, and the seemingly good machine turned bad, AUTO—science fiction is now becoming a reality. If you believe in movies, you may even fall in love with your operating system! So, can Johnny 5 really come alive? Or will he short-circuit?

In late 2016, the House of Commons Science and Technology Committee published a report on Robotics and Artificial Intelligence which followed an inquiry with a broad focus and examined robotics and AI in the round. The report talks of the transformational impacts that AI systems are starting to have on everyday life, and touches on the social, legal and ethical issues raised by such breakthrough technology. By way of illustration, the report states the importance of ensuring that AI systems are operating as intended and that unwanted and unpredictable behaviours are not produced by either accident or malicious intervention. The report also explores issues of accountability and, taking driverless cars as an example, asks the all-important question, "if

something goes wrong, who is responsible?"—who's behind the driving wheel from a liability perspective? Unsurprisingly, Brexit makes an appearance and the report asks how the EU's General Data Protection Regulation (GDPR), under which an individual can ask for an explanation of an algorithmic decision made about them, will be transposed into English law following Brexit.

Verification and validation

The Committee states that processes are required to verify that AI technology is functioning correctly and that such validation of autonomous systems is extremely challenging. The report uses Google DeepMind as an example of ensuring safe and reliable operation. In June 2016, Google DeepMind was reported to be working with academics at the University of Oxford to develop a "kill switch": code that would ensure an AI system could "be repeatedly and safely interrupted by human overseers without [the system] learning how to avoid or manipulate these interventions". In the same month, the report states, researchers from Google, Open AI, Stanford University and UC Berkeley in the US together published a paper which examined potential AI safety challenges and considered how to engineer AI systems so that they operated as intended. There is an acknowledgement that humans need to maintain ultimate control, to avoid situations getting out of control—a sophisticated re-set button for robots with a mind of their own.

Decision-making transparency

The report refers to Google DeepMind's artificially intelligent computer programme, AlphaGo, which in 2016 won a five-match series of the ancient Chinese board game "Go" against the reigning world champion, Lee Sedol, to broach the subject of transparency. The machine was able to beat its human opponent in one match by playing a highly unusual move that prompted match commentators to assume that AlphaGo had malfunctioned. The report explains that AlphaGo cannot express why it made this move and, at present, humans cannot fully understand or unpick its rationale. For this reason it is impossible to get under a computer's skin to understand what is really going on in its head, and why. As Dr Owen Cotton-Barratt from the Future of Humanity Institute reflected, we do not "really know how the machine was better than the best human 'Go' player". However, we do know that the computer beat him at his own game!

It is currently rare for AI systems to be programmed to provide a reason for reaching a particular decision. The report makes the point that when the stakes are low—such as in a board game like "Go"—this lack of transparency does not matter. However, where the stakes are far higher, say for example in the financial sector or medical diagnostics which dice with death, an absence of transparency and full explanations could lead to a level

* Emma L. Flett is a Partner and Jennifer F. Wilson an Associate, Technology & IP Transactions, Kirkland & Ellis International LLP.

of public mistrust, since the reasoning behind the algorithmic decision is opaque. Patients, for example, may be unwilling to trust a computer by simply accepting the supposed quality of an algorithm where their treatment is concerned and may instead want a clear justification from a leading human physician in the field. On the flip side, a lack of clear explanation also makes it difficult for humans to learn from machines—the irony being that a computer could trump a leading living expert—the conundrum of artificial intelligence v real intelligence.

On this very subject, the report refers to the GDPR, which is due to come into effect across the 28 EU Member States on 25 May 2018. It will create a “right to explanation”, whereby a user can ask for an explanation of an automated algorithmic decision that was made about them. The UK Government has confirmed that the UK’s decision to leave the EU will not affect the commencement of the GDPR in May 2018. However, according to the report, whether and how this new right will be transposed into English law following Brexit remains unclear.

Minimising humans’ imbedded bias in artificial intelligence

The potential for bias and discrimination in the design of AI systems is brought to the fore in the report, no less so than with the example of Google’s photo app, which automatically applies labels to pictures in digital photo albums. The app was reported to have misclassified images of black people. According to Microsoft, the AI system built a model of the world based on training data. The report then quotes experts from the University of Nottingham who explain that “all data-driven systems are susceptible to bias based on factors such as the choice of training data sets, which are likely to reflect subconscious cultural biases”. The report seems to be agreeing with the article “Artificial Intelligence’s White Guy Problem” in the *New York Times* (25 June 2016), which states that if a system was “trained on photos of people who are overwhelmingly white, it will have a harder time recognizing non-white faces”.

The report ends this rather disturbing paragraph with a warning from an expert at Microsoft, who emphasises “a need to be mindful of the philosophies, morals and ethics of the organisations ... creating the algorithms that increasingly we rely on every day”.

Privacy and consent

The Committee explains that during the course of its inquiry, there were reports in the media about Google DeepMind working with NHS hospitals to improve patient diagnoses and care. Media commentary focused not just on the work that was underway—such as building an app that helps clinicians detect cases of acute kidney injury, or using machine learning techniques to identify common eye diseases—but also on DeepMind’s access to sensitive patient data: namely, how much data the company could

access, whether patient consent had been obtained, and the ownership of that data. The report states that such concerns are not new and warns that the anonymisation and re-use of data is an issue that urgently needs to be addressed.

Accountability and liability

The report suggests that questions of accountability and liability are particularly pertinent. To date, these have predominately been discussed in the context of driverless cars and lethal autonomous weapons systems—accountability for the latter is particularly of critical importance. The key question preoccupying not just the Committee but the legal world is: “if something goes wrong, who is responsible?” Who shall we point the finger at, mechanised or otherwise? The need for a level of accountability for algorithms is clear, i.e. the people writing the algorithm and the AI need to be held accountable for the outcome. Those responsible for the input need to be responsible for the output—like an old-fashioned function machine.

Perhaps unsurprisingly, the report focuses on autonomous cars and liability. It refers to the Law Society’s point that situations may arise in which a driverless car takes action that causes one form of harm in order to avoid other harm. The Law Society states that this raises issues of civil, and potentially even criminal liability as well as the ownership of that liability, whether the manufacturer of the vehicle, the software developers (and the ratio of responsibility where there are multiple developers), the owner of the vehicle and a myriad of other questions—suggesting that this is just the tip of the iceberg. The need for accountability makes it clear that the numerous people behind AI technologies should not be able to wash their hands (or pincers!) when things take a turn for the worst.

The report includes a timely proposal from the Government (Centre for Connected & Autonomous Vehicles, *Pathway to Driverless Cars: Proposals to support advanced driver assistance systems and automated vehicle technologies* (July 2016)) addressing liability for automated vehicles:

“Our proposal is to extend compulsory motor insurance to cover product liability to give motorists cover when they have handed full control over to the vehicle (i.e. they are out-of-the-loop). And, that motorists (or their insurers) rely on courts to apply the existing rules of product liability — under the Consumer Protection Act, and negligence — under the common law, to determine who should be responsible.”

We need to recognise that we are living in a world where an autonomous system—which is not flesh and blood, has no heartbeat and cannot breathe—can, in fact, make life-changing decisions. We then need to consider how

we can address this brave new world that AI technology presents within the confines of the law, which in many respects is playing catch-up.

Governance—standards and regulations

The Committee’s report highlights the fact that, as AI is applied in wider and increasingly diversified fields, suitable governance frameworks will be needed in order to build public trust and confidence. The report quotes Innovate UK’s view, as follows:

“Appropriate legal and regulatory frameworks will have to be developed to support the more widespread deployment of robots and, in particular, autonomous systems. Frameworks need to be created to establish where responsibilities lie, to ensure the safe and effective functioning of autonomous systems, and how to handle disputes in areas where no legal precedence has been set.”

On the other side of the coin, the Committee’s inquiry took views from those who emphasised that a balance needed to be struck on the grounds that efforts to introduce a governance regime could curtail innovation and hold back desirable progress by stifling technological advancement in the area.

The National Endowment for Science Technology and the Arts (Nesta) is referred to in the report more than once, and on this particular issue it notes that there are moves “in both the public and private sectors to set up ethical frameworks for best practice”. Such initiatives are being developed at the company level (e.g. Google DeepMind’s ethics board); at an industry-wide level (e.g. the Institute of Electrical and Electronics Engineers global

initiative on “Ethical Considerations in the Design of Autonomous Systems”) and at the European level (e.g. Committee on Legal Affairs, European Parliament, Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), May 2016).

Conclusion

Essentially, the report finds that advances in robotics and AI hold the potential fundamentally to reshape the way we live and work. Robots and autonomous systems will change life as we know it. It laments the fact that the Government needs a reboot, as it does not yet have a strategy for developing the new skills citizens will need to flourish in a world where AI is more prevalent, or responding to the social and ethical dilemmas it poses. As such, the Committee calls for a Commission on Artificial Intelligence to be established at the Alan Turing Institute to examine the social, ethical and legal implications of recent and potential developments in AI.

The Committee concludes that the UK is well placed to provide this type of intellectual leadership. Much of the significant progress in this field, such as improved automated voice recognition software, predictive text keyboards on smartphones and autonomous vehicles, has in fact been driven by UK-based technology start-ups, founded by graduates of UK universities, as well as universities themselves. Whether such industries will have the funding to survive and throw the UK a lifeline in a post-Brexit world remains to be seen—keep those fingers crossed, whether real or robotic! But one thing is certain: robotics is the future. Arnie was right when he said, “I’ll be back!”