IS IT TIME TO REGULATE AI?

Thanks to big increases in computing power, artificial intelligence has now become a legal and regulatory concern. **MARC BEISHON** weighs up the evidence for intervention

The biggest emerging technology and regulatory issue in 2017 was artificial intelligence (AI), although Bitcoin, cybersecurity and fake news also made big splashes. But AI is now a key issue in regulatory and policy circles, with a multitude of reports, commentary and academic papers weighing in on the principal question: does AI need regulating at this stage in its development, or at least in the near future? The battle lines have been drawn up between those advocating regulation and those who say that any such moves are premature.

In the US, there are already moves to propose legislation. A bipartisan group of senators have put forward the "Future of AI Act" that would create a federal advisory committee to examine and wrestle with the economic opportunities and impacts that emerging AI technologies will have in many aspects of American life.[1] The committee would be set up by the Department of Commerce, and would address policy on the workforce, privacy, innovation and "the development and application of unbiased AI".

Meanwhile a local bill in New York City would require the creation of a taskforce that provides recommendations on how information on automated decision systems run by agencies may be shared with the public and how agencies may address instances where people are harmed by such decision systems.[2]

Although these moves are about setting up committees, and are as much about economic growth as safety, they were no doubt prompted by alarmist talk by leading figures such as Elon Musk and Stephen Hawking. Back in 2015, they and a number of AI experts signed an open letter[3] calling for research on its societal impacts, but since then the rhetoric has ramped up. Hawking has said: "Alongside the benefits, AI will also bring dangers, like powerful autonomous weapons, or new ways for the few to oppress the many." Musk, the force behind Tesla and SpaceX, has tweeted messages such as: "If you're not concerned about AI safety, you should be. Vastly more risk than North Korea", and: "Got to regulate AI/robotics like we do food, drugs, aircraft and cars. Public risks require public oversight." He has also co-founded the non-profit OpenAI, which aims to promote "safe" development of AI technologies, and has briefed US state governors on AI.

But there are sharp disagreements with the view that we are heading for any kind of apocalyptic end with AI, and indeed that we need much in the way of even mundane regulation in any case. Typical of this "hands off" side is Andrew Burt, an executive at

data management firm, Immuta, who says that there is little agreement on what AI is, but it is certainly not a single technology or development. "Regulating an assemblage of technology we can't clearly define is a recipe for poor laws and even worse technology," he has said,[4] adding that the challenges of AI are not all new, as we have long been using automated systems in, say, financial markets. Back in 1970, Newsweek ran a cover story titled "Is privacy dead?", premised on the impact of computers. Regulating sector-specific systems such as banking algorithms is instead a reasonable strategy. The point is that many of the issues posed by AI lie in the data it uses.

GOVERNMENT RESPONSE

Governments have so far issued reports that largely agree with this more relaxed position. One of the last outputs from the Obama administration's Office of Science and Technology Policy (OSTP) was a report, "Preparing for the future of artificial intelligence",[5] which said that considerable progress has been made on what is known as "narrow AI", which is about applications such as playing strategic games, language translation, self-driving vehicles and image recognition. The type of AI that worries people more, known as "general AI", is about AI that exhibits apparently intelligent behaviour at least as advanced as a person and is the stuff of dystopian visions of the future, but the report found little here to trouble policymakers, at least at the back end of 2016.

For regulation, the report talked about specific risks that AI systems may introduce, such as ensuring the safety of autonomous vehicles and aircraft. The report says that "broad regulation of AI research or practice would be inadvisable at this time", but government agencies will need appropriate technical expertise at a senior level. However, it did also discuss the use of algorithms in applications such as criminal justice and cybersecurity, although it pointed out that they can be both positive (machine learning can help rebuff cyber-attacks) and negative (biased learning can discriminate against certain groups).

There is probably more concern at government level about the impact of automation on the economy and jobs, which was addressed in a companion report from OSTP, "Artificial intelligence, automation, and the economy".[6]

In the UK, a government-commissioned report, "Growing the artificial intelligence industry in the UK", by Wendy Hall and Jérôme Pesenti, published in autumn 2017,[7] focused as the title suggests on the business side of AI, noting that AI could add an additional $814 billion to the UK economy by 2035, increasing the annual growth rate of GVA (gross value added) from 2.5% to 3.9%. The report rightly recognises Alan Turing, the pioneer of computer science, as inspiring much development of AI, and that "the UK has an exceptional record in key AI research". But it has little to say about drawbacks.

The recommendations are grouped into four categories – improving access to data (and here there is advice that policy should support open data and trusted frameworks); improving skills; maximising AI research; and supporting the uptake of AI. The report does say: "Resolving ethical and societal questions is beyond the scope and the expertise of this industry-focused review" – so it is a rather one-sided report. The authors do refer to a Royal Society and British Academy report, also published last year, "Data management and use: governance in the 21st century".[8] From this they conclude that, "While AI will generate some specific challenges, it would not be helpful to see AI governance as something unrelated [to] and separate [from] broader data governance." This is in line with the idea that AI is essentially a data management issue.

However, there is a House of Lords Select Committee on Artificial Intelligence that is looking at wider issues in the UK, such as what the public needs to know about AI.[9] In December it heard from a number of witnesses on questions such as, "What assessment has the government made of existing regulations related to AI to ensure that they are fit for purpose, now and in the future?" and "Is the AI community right to be concerned that scaremongering over AI could damage trust in the technology?"

> **"It would not be helpful to see AI governance as separate from data governance."**

Hall and Pesenti also call for the establishment of an AI council to promote growth and coordination in the sector. The idea of some kind of coordinating body is a common theme in other countries and reports and papers – there is feeling that AI does require something more than existing regulatory and policy bodies. The authors also recommend "guidance on how to explain decisions and processes enabled by AI" – which is related to the concerns of the Lords committee and a big topic alone (see the panel, p22). The EU's General Data Protection Regulation (GDPR) has been said to contain a "right to explanation" that would apply to AI systems. However, a paper by Sandra Wachter and colleagues[10] says that "there are several reasons to doubt the existence, scope, and feasibility of a 'right to explanation' of automated decisions... We argue that the GDPR does not, in its current form, implement a right to explanation, but rather what we term a limited 'right to be informed'." The paper includes discussion of how EU member states such as France and Germany have made such distinctions in their own laws.

Last year, European members of parliament asked the European Commission to propose rules on robotics and AI, in particular legislation to clarify liability issues for self-driving cars and robots.[11] Studies, such as one on civil law rules in robotics, have been published.[12] But much of the recent output from Europe has been more on economic opportunities, and launching projects on high-performance computing and quantum technology. Andrus Ansip, the Commission's VP for the digital single market, has said though: "We are now

ACCOUNTABILITY: SOME EXPLAINING TO DO

A paper titled "Accountability of AI under the law: the role of explanation", asks: "How can we take advantage of what AI systems have to offer, while also holding them accountable?"[13] The authors focus on one important tool, which is called "explanation". This is about the reasons or justifications for a particular outcome from a decision, and in certain cases these are expected under legal systems (which they outline in the US context). Law apart, if there is reason to believe an error has or will occur in decision making, we may want explanations if it is believed there were inadequate inputs, inexplicable outcomes or distrust in a system's integrity.

AI systems are like "black boxes" here in that we won't need to know how they work as we can design and test an explanation regulation independently. But this will work best where its terms are determined in advance, i.e. ex-ante (such as training a health system to know when someone has diabetes). It gets more challenging for ex-post scenarios – in such cases, "AI systems may struggle; unlike humans, they cannot be asked to refine their explanations after the fact without additional training data."

The authors say that demanding explanations from AI systems is reasonable, and that we should start by asking of our AI systems what we ask of humans. "Doing so avoids AI systems from getting a 'free pass' to avoid the kinds of scrutiny that may come to humans, and also avoids asking so much of AI systems that it would hamper innovation and progress."

However, there is also a question of need. "Just as with requirements around human explanation, we will need to think about why and when explanations are useful enough to outweigh the cost. Requiring every AI system to explain every decision could result in less efficient systems, forced design choices, and a bias towards explainable but suboptimal outcomes." We probably don't want a smart toaster to say what it did, but "may be willing to accept the monetary cost of an explainable but slightly less accurate loan approval system for the societal benefit of being able to verify that it is nondiscriminatory".

The authors note two other approaches for accountability. One is empirical evidence, which is a measure of a system's overall performance. Examples are observing that an autonomous aircraft landing system has fewer safety incidents than human pilots, or that the use of a clinical diagnostic support tool reduces mortality.

The other is a theoretical guarantee. For example, "We trust our encryption systems because they are backed by proofs... Similarly, if there are certain agreed-on schemes for voting and vote counting, then it may be possible to design a system that provably follows those processes."

working on a European strategy on robotics and AI, planned for early next year. This will look at how best to promote AI to benefit Europe's people and businesses, our society and economy. It will also address ethical, legal and socioeconomic aspects." Undoubtedly the GDPR, the e-privacy regulation and measures to strengthen cybersecurity will go some way to addressing concerns.

DEVELOPING AN AI INDEX

It is a reasonable point that there is not enough information about what constitutes AI and where it is going. A project that is tracking this is Stanford University's One Hundred Year Study on Artificial Intelligence, and part of this is the AI Index, the first edition of which was published in November 2017.[14] Initially, it is focused mostly on the US but the trends it reports show just how quickly AI has expanded in the past few years:

- Volume of activity – the number of AI papers has increased by more than 9x since 1996, particularly those in computer science. The number of students at Stanford taking AI and machine learning (ML) courses has risen sharply since 2010; this is also reflected in other US colleges. Attendance at AI conferences has also zoomed up in the past 2-3 years and the research focus has shifted from symbolic reasoning to ML and "deep learning".
- Industry – the number of venture-backed US private companies developing AI systems has zoomed up by 14x since the year 2000. The share of jobs requiring AI skills in the US has increased 5x. Canada and the UK have even higher numbers. The number of robots imported has grown rapidly since the recession of 2008.
- Public interest – a metric here is the "sentiment" of articles referencing AI. Positive articles, perhaps surprisingly, outweigh negative ones.
- Technical performance – here the Stanford researchers report how well AI systems perform. In object detection, the best AI system now outperforms humans, but progress seems to have stalled in asking a system to give open-ended answers to questions about images, where people do much better. AI is doing better in language processing, but took a dip recently in German to English translation (and vice versa). AI is approaching human levels in finding the answer to a question within a document, and now equals people in recognising speech from phone call audio.

Also reported are milestones, such as in 2017 when a program called Libratus defeated four top human poker players in a tournament, and an AI system trained on a data set of 129,450 clinical images was capable of classifying skin cancer at a level of competence comparable to dermatologists.

The authors note that important areas are not yet covered in this inaugural index, such as in technical performance where there are no benchmarks for some metrics. They also say: "Tracking areas that have traditionally lacked concrete measurements may also facilitate a more sober assessment of AI progress. Progress is typically tracked consistently when good progress has been made. As a result, this report may present an overly optimistic picture."

They would also like to include more international data, impacts on vertical industries, and impacts of AI on society. "Issues associated with societal risks resulting from AI are left unaddressed... In future, we hope to provide metrics that help ground discussions of AI safety and predictability, fairness of AI algorithms, privacy in an era of AI, ethical implications of increased automation, and other topics." In other words, like much of AI, we are in the early days of developing a comprehensive index.

AN AI ROADMAP

Among the recent crop of papers trying to make sense of AI is one by Ryan Calo at the Washington School of Law, "AI policy: a primer and roadmap".[15] As he says: "I am a law professor with no formal training in AI. But my longstanding engagement with AI has provided me with a front row seat to many of the recent efforts to assess and channel the

impact of AI on society." This puts him in a position to stand back from the technology and set out a roadmap for policy that addresses many of the concerns voiced by commentators (not least that we do lack policy tools to make sense of the rapidly emerging technologies). Calo's roadmap comprises these components:

Justice and equity. "Perhaps the most visible and developed area of AI policy to date involves the capacity of algorithms or trained systems to reflect human values such as fairness, accountability, and transparency," he says. Outlined are ways in which applications can treat people unequally, such as in image recognition of ethnic minorities, price discrimination against certain groups, and using crime "heat maps" that can result in certain groups being subject to more harassment. Calo says there are two policy questions – what constitutes best practice in minimising bias (antidiscrimination laws, consumer protection, industry standards) and how to ensure that the risks and benefits of AI are evenly distributed across society.

Also in this category is what he terms "consequential decision making", which is related to inequality but a consequence of using AI, especially by government in areas such as criminal justice where there are existing rules and process guarantees. Calo say these systems need to be understood before new ones are designed, and objectives and values imported into a machine need context. "AI here can say what will happen but not why," he says, noting an example – when should a person be taken off a life-support system?

An interesting point is that striving for transparency and fairness should not come at the expense of efficiency otherwise an AI process could take too long and could actually decrease its accuracy, according to researchers. So there are trade-offs.

Use of force. This is a special case and concerns the debate over taking life and whether there can still be meaningful human control with, say, autonomous weapons. "Policymakers must work toward a framework for responsibility around AI and force that is fair and satisfactory to all stakeholders," he writes.

Safety and certification. Given that AI systems can control vehicles and provide sensitive services that normally require human training and certification, Calo asks how standards can be met. Just how safe should a driverless car be? That's more a policy than a technology question, and then how would one set a standard for a given safety threshold? Indirect harm can also be caused by technology, such as cognitive radio systems (a telecoms example) that inadvertently cut off emergency communications. An AI system could also spread disinformation, but is this just free speech? There is already such a concern in the conduct of elections. There is scope for regulatory standards here and also legal liability.

> **Indirect harm can also be caused by technology such as cognitive radio systems.**

Meanwhile there are already systems such as autopilots that do not require a certificate like one awarded to a human pilot. But Calo says there are new systems such as surgical robots that will be able to do things that no human can do – so should they be certified in some way?

A SELECTION OF AI DATA POINTS AND RESOURCES

- AI is a term coined as long ago as 1956. Narrow AI refers to systems that perform particular tasks while general AI is about the ability for machines to reason like a human. Machine learning gives computers the ability to learn without being explicitly programmed, while deep learning is about mimicking neural networks.
- A paper by Urs Gasser and Virgilio Almeida has a model for AI governance. It has three interacting layers that sit between society and AI systems: social and legal, ethical and technical.[16] They also say: "Absent an AI-specific international legal framework, a global oversight body... could be the curator of global principles and emerging norms for AI systems."
- There are a number of groups taking action about AI:
  – Algorithmic Justice League (ajlunited.org), which highlights algorithmic bias
  – Future of Life Institute's AI stream (see futureoflife.org)
  – Computer Science for All Movement (csforallconsortium.org)
  – AI4All initiative (ai-4-all.org)
  – The Partnership on AI has been set up by Microsoft, Google and others.
- "Humans are defenceless in information environments that are grossly corrupted," said Stuart Russell, from UC-Berkeley, speaking at last year's AI for Global Good Summit.[17] How we might protect the integrity of our information environment is a problem likely to prove very difficult to solve as AI advances. "The use of malware has been a catastrophe," said Russell, highlighting that direct theft from bank accounts now exceeds $100 billion a year. "I think by solving malware we will set up a paradigm for how we can start to think about controlling the misuse of AI. But we have to do this soon."
- Emerging countries such as India have been looking closely at the benefits of AI but as an article in a special AI issue[18] of MIT Technology Review notes: "Automation could hit India particularly hard because much of its high-tech economy involves relatively routine work that is prime for computers to take over." Also in the issue are articles on how AI can detect how people are feeling, how to root out hidden biases in AI, and "Don't let regulators ruin AI".
- Many great ideas in AI languished in textbooks because we did not have the computer power to apply them, an article in Science notes.[19] Now, one idea, improving neural networks not through teaching, but through evolution, is revealing its potential. Five papers from Uber demonstrate the power of so-called neuroevolution to play video games, solve mazes, and even make a simulated robot walk.
- Some AI apps just seem to be for the good – a system that listens in on emergency calls in Denmark can diagnose heart attacks from voices and other background sounds better than dispatchers can.[20]

Privacy and power. This is more familiar ground for data protection and communications regulators. As Calo says, AI is intimately tied to the availability of data, and can engage in pattern recognition that if captured by companies and governments can track and predict your every move. It's deep thought about big data. Serious policy issues concern the lack of insight that consumers have about sharing information and this is the domain of the current reform of the e-privacy directive in Europe.

But there is the genuinely frightening prospect of everyone being identified by facial recognition when in public. The power of knowing so much about someone could also put them at the mercy of cleverly targeted messaging. Much of this is now a reality.

It is also about an issue he calls data parity – where a handful of very large companies monopolise information and smaller companies find it increasingly hard to compete as data is not shared (and this is also an issue for government-held data). This is also a privacy question, as "firms will, and already do, invoke consumer privacy as a rationale for not permitting access to their data. This is partly why the AI policy community must maintain a healthy dose of scepticism toward 'ethical codes of conduct' developed by industry."

This is not to say we should "run roughshod over privacy in pursuit of data parity... The hard policy question is how to incentivise technical, legal, social and other interventions that safeguard privacy even as AI is democratised."

Taxation and displacement of labour. The final item on Calo's list is how we deal with the replacement of jobs by robots, driverless trucks and so on, and whether there should be taxation of AI systems (which Bill Gates, for one, is in favour of).

Calo says these headings are by no means exhaustive and there are also issues that cut across them. One is what institutional arrangements we might set up and what expertise is needed to manage AI policy, which if done at all is currently piecemeal. "One overarching policy challenge is how best to introduce expertise about AI and robotics into all branches and levels of government so they can make better decisions with greater confidence." Governments can also influence AI through their procurement decisions. He also notes various calls for policymakers to open up AI systems to more legal scrutiny given that most are proprietary.

As for the existential threat to humanity, Calo is not in the Hawking/Musk camp and says they could distract policymakers from more immediate harms and challenges. He says though that while the doomsday idea of AI deciding to wipe humanity out is science fiction, there is a concern that there could be accidents or people could use AI for bad ends.

However, it is really the day to day encroachment of AI now that should be the focus. As Calo notes, the problem is not that AI "will get too smart and take over the world," as computer scientist Pedro Domingos writes, but that "the real problem is that [it's] too stupid and [has] already".

CONCLUSION

While most would agree that the most alarming predictions for AI are not likely to be a threat, there does seem to be some complacency that the gains from AI and its various manifestations far outweigh the risks, and that in any case those risks are being taken care of with regulation in data protection and privacy, and perhaps with some form of legal status.

But the European Economic and Social Committee (EESC) has called firmly for a "human-in-command" approach, a code of ethics, and a ban on autonomous weapons, and warns of developments of AI taking place "within a homogenous environment principally consisting of young, white men" and that the AI systems now being developed will not have any built-in ethical values.[21]

The EESC also says a seemingly strange thing: "Technology is not something inevitable." This point was also made by Robin Mansell in the last issue of Intermedia.[22] Writing from the digital economy perspective, and about a new type of digital divide, she says: "...the more digitally mediated benefits we have, the fewer opportunities there are for humans to exercise their control and authority." The concern about AI is also that it can exacerbate inequality and lead to less control by ordinary people and that the dominant theme is economic investment.

> **The concern about AI is also that it can exacerbate inequality and lead to less control.**

One of the best quotes that sums this up came last year from Apple CEO, Tim Cook, speaking to students at MIT: *"I'm not worried about artificial intelligence giving computers the ability to think like humans. I'm more concerned about people thinking like computers without values or compassion, without concern for consequences. That is what we need you to help us guard against. Because if science is a search in the darkness, then the humanities are a candle that shows us where we've been and the danger that lies ahead. As Steve [Jobs] once said, technology alone is not enough. It is technology married with the liberal arts married with the humanities that make our hearts sing. When you keep people at the centre of what you do, it can have an enormous impact."*

MARC BEISHON *is a long-standing technology writer, and editor of Intermedia.*

REFERENCES **1** Future of Artificial Intelligence Act of 2017. bit.ly/2D5ArtT **2** New York City (2017). A Local Law in relation to automated decision systems used by agencies. on.nyc.gov/2Dc7Cft **3** Future of Life Institute (2015). Research priorities for robust and beneficial AI. Open letter. bit.ly/2kMcltN **4** Burt A (2018). Leave AI alone. New York Times, 4 January. nyti.ms/2CIjDIZ **5** Office of Science and Technology Policy (2016). Preparing for the future of artificial intelligence. bit.ly/2J3XA4k **6** Office of Science and Technology Policy (2016). Artificial Intelligence, automation, and the economy. bit.ly/2jsebMI **7** Hall W, Pesenti J (2017). Growing the artificial intelligence industry in the UK. bit.ly/2hlshdo **8** British Academy, Royal Society (2017). Data management and use: governance in the 21st century. bit.ly/2C9TzCT **9** Lords Select Committee on AI. parliament.uk/ai-committee **10** Wachter S et al. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law 7 (2). bit.ly/2mwBBU4 **11** European Parliament (2017). Robots and artificial intelligence: MEPs call for EU-wide liability rules. Press release, 16 February. bit.ly/2r8z4or **12** European Parliament (2016). European civil law rules in robotics. bit.ly/2ilgePS **13** Doshi-Velez F, Kortz M (2017). Accountability of AI under the law: the role of explanation. Berkman Klein Center for Internet and Society. bit.ly/2EldUjf **14** Stanford University (2017). AI Index. aiindex.org **15** Calo R (2017). Artificial intelligence policy: a primer and roadmap. bit.ly/2vtyG47 **16** Gasser U, Almeida V (2017). A layered model for AI governance. IEEE Internet Computing 21 (6): 58–62. bit.ly/2Dc4rAg **17** Human-compatible AI: design principles to prevent war between machines and men. ITU. newslog.itu.int/archives/1571 **18** The artificial intelligence issue. MIT Technology Review, 2017. bit.ly/2mz4zCF **19** Hutson M (2018). Artificial intelligence can 'evolve' to solve problems. Science, 11 January. bit.ly/2D4N7R2 **20** Peters A (2018). Having a heart attack? This AI helps emergency dispatchers find out. Fast company, 11 January. bit.ly/2DkBN06 **21** European Economic and Social Committee (2017). Artificial intelligence. bit.ly/2mGOHyP **22** Mansell R (2017). Are we losing control? Intermedia 45 (3).