Inside the black box: the regulation of AI

Pouyan Afshar
9th Floor Wentworth Chambers

Much has recently been written about AI. The acronym and the fears that it elicits are omnipresent. So is content about regulating AI. Debate has raged for a number of years among entrepreneurs, regulators and commentators. In recent times, governments have weighed into the debate and legislators have begun to busily draft new laws to 'regulate AI'. This article seeks to do three things: first, to describe what 'AI' is and is not, so that the subject matter of the proposed regulation is exposed with clarity; second, to identify where regulation has or may be effected for it to be *effective*. Finally, to make some observations about the current and future regulation of AI.

What AI is and is not

Some of the commentary on AI regulation is posited on the basis of an incorrect definition of AI. The problem is fundamental, given that the promulgators of regulation first need to understand the cause of the harm they seek to minimise.

Like all emerging concepts, AI is not easy to understand, let alone define. Having said that, the differences between AI and software are at least, at a general level, tolerably clear. AI is technology that enables machines and computers to use unstructured data from various sources to perform tasks that humans perform. It consumes data in the form of language, is commanded by language, and produces results in language. In contrast, software is the application of rules, including business rules, imposed by code to structured data. Software consumes data from databases, is commanded by code (or algorithms), and produces results in whatever form the software requires: images, reports, tables etc.

The differences may be illustrated thus. In a recent speech on the regulation of AI, the following example was put forward to highlight some of the dangers of AI:

> It isn't fanciful to imagine that credit providers using AI systems to identify 'better' credit risks could (potentially) unfairly discriminate against those vulnerable consumers. And with 'opaque' AI systems, the mechanisms by which that discrimination occurs could be difficult to detect. Even if the current laws are sufficient to punish bad action, their ability to prevent the harm might not be.[1]

The conduct described is problematic and potentially unlawful. But *what* was being described may not be, and probably is not, AI; it is probably software. Such software exists and has existed for years. It is used to determine the fate of credit applications, set insurance premiums, or select suitable candidates for open job vacancies. It is the application of specific business decisions (through code) to data. In the context of lending, such software invariably discriminates against those with question marks as to their creditworthiness; it is meant to do so. The code is opaque, and its aims are not readily ascertainable. The perpetrator(s) of the conduct, be they natural or juridical persons, may well be culpable of or liable for unlawful conduct. But the function is one of software, divined, designed, and implemented by humans, and not AI.

An AI application would not apply predetermined rules to structured financial data to discriminate (fairly or otherwise) against or in favour of consumers. It would burrow into the *unstructured* data, including, but not limited to, data held by a lending institution, to conclude, independently of software or code, the indicia of creditworthiness, and it would apply that criteria to a credit application. No business rules imposed by code determine the outcome. If they do, the technology is not AI; it is likely software that uses business rules (defined in certain circumstances as 'algorithms') to determine the most advantageous outcome for the business that uses it. In contrast, AI technology *itself* concludes who ought to be given credit and who should be denied it based on criteria it has determined to be appropriate.

In the same speech, another example was given of an AI investment manager used by a 'provider' to support a related party product. Again, the technology that is described is likely not AI, but software. An AI investment manager's decision-making may lack transparency and accountability and may not allow the means of oversight, but *it* makes those decisions rather than being at the behest of business rules imposed by code.

What is to be regulated

Once the mist over the definition of AI is lifted, it is possible to identify three aspects of AI that may be susceptible to regulation. (It must be said at this juncture that the debate is, much like any substantive debate, more detailed and nuanced – and capturing it is beyond the scope of this short article.)

The first is the information AI receives and on which it is trained. Some regulation already exists on the use of information. The *Privacy Act 1988* (Cth) regulates the use of personal information in Australia; similar laws exist in other jurisdictions, for example, the European Union *General Data Protection Regulation*. Copyright laws protect the substance of materials that are fed to AI; in recent lawsuits, the *New York Times* and others have claimed that OpenAI, the makers of ChatGPT, breached copyright. The use of information may also be protected by contractual terms imposing confidentiality or by court order. There may be some scope to add further protections to the use of information to protect against the sheer voraciousness of AI when it comes to consuming data. Some changes to the privacy laws and regulations are proposed. But this is not where regulation could best be deployed.

The second aspect of AI that may be regulated is its output. Asking AI or causing it, through language commands, to do something – say, crack a code to a computer to access confidential or other protected information or to implement a money laundering scheme – is already regulated, but some other uses of AI may not be. In that construct, AI as a technology is but a tool for a natural person or corporation to engage in conduct that may be lawful or unlawful. Regulation may be necessary to attribute liability or culpability for the autonomous actions of AI. Is the company that owns the rights to an AI investment manager that decides to use data to which it has access to commit insider trading responsible for that unlawful conduct? Or are its directors? Did they aid and abet the commission of that unlawful conduct? Did they even know the AI had made those decisions?

The third and greyest aspect of the operation of AI is how it consumes data. Data, of course, is not neutral. For example, if over time the AI is fed certain historical data about the refusal of credit to members of certain communities – say, according to their socio-economic backgrounds – and is then asked to vet credit applications, it might decide to discriminate against consumers from such communities. That is perhaps a stark example, and the position will always be more nuanced. But the example highlights that *what* information AI consumes and *how* such data is processed are perhaps the least understood and most difficult areas of its operation to regulate. It is unlikely that descriptions of inputs and outputs as 'biased' clarify the problem or lead to its understanding and ultimate resolution, for most data is biased one way or the other. Further, it is unlikely that the problem will necessarily be solved if data is skewed manually to allay its inherent biases.

It follows from the above analysis that the second and third aspects of the operation of AI ought be the focus of any regulation, or further regulation. It is beyond the scope of this article to propose specific regulations or means of protecting people from the risk of harm from AI. That is a task for others.

Impacts of AI

The impacts of AI are likely to be felt across the globe. Indeed, they already are. Some impacted industry and professional bodies have begun regulating their participants' or members' use of AI to avert its impacts. Those bodies insist that misuse of AI may breach existing industry standards or professional rules of conduct.

It appears, however, that regulation may also be needed to protect against some of the wider economic and social implications of AI – for example, on many Australians' employment. While AI no doubt increases productivity, it will likely make many jobs redundant. AI can already produce reliable code, resolve complex chemical or mechanical problems, and opine on some medical questions. It is not a stretch to imagine that some roles will become wholly redundant or substantially reduced in number if more businesses adopt AI as part of their operations. Technology has historically not upended economic and social norms in the radical ways its advocates or opponents thought it would; however, the effects of such technologies ought not be underestimated. Those problems are likely to be multi-faceted. Thus, any regulation implemented as part of the fabric of employment laws and regulations will likely need to be coordinated with policy proposals to deal with the macro- and microeconomic effects of AI.

Looking ahead

On 17 January, the Australian Government published an interim response to its Responsible AI consultation held in 2023.[2] The response foreshadowed a new regulatory framework for high-risk AI applications and the development of voluntary standards and means of identification of various applications (through labelling, etc). Legislative and regulatory expansion and tightening were also foreshadowed.[3]

The response proposes a risk-based approach to regulating AI. In such a framework, the level of regulation varies based on the level of risk the relevant AI technology poses. AI may be defined as 'high risk' if its impacts are 'systemic, irreversible or perpetual'. The keen observer armed with a correct appreciation of the nature of AI (see above) may be concerned about how such a definition may be applied in practice. The response gives two examples: the use of AI in self-driving cars and robots for medical surgery. It appears then that the focus of the drafters of the response was on the final application of AI (the second aspect identified above). The references to European Union and Canadian laws that assess risk by reference to the impact of AI supports that proposition. It is suggested that the third aspect of the operation of AI also requires some attention. Otherwise, the response proposes updates to certain existing laws, the implementation of technology neutral regulations, and the setting up of various regulatory bodies.

Looking ahead, it remains unclear how any AI regulation will be implemented and whether it will work. The proof will be in the proverbial, and there are two reasons for this. First, the pace of technical advancement is breathtaking. Second, our view of AI and its implications is not yet as clear as it could be.

In extra-judicial writings on technology and the courts, the former Chief Justice of the Federal Court said:

> To a degree, the future must remain unknown. Artificial intelligence and its effect on courts, the profession and the law will change the landscape of life in ways we cannot predict.[4]

While it is perhaps uncomfortable for lawyers, those observations are prescient and correct also in relation to the broader impacts and consequences of AI. BN

ENDNOTES

[1] Keynote address by ASIC Chair Joe Longo at the UTS Human Technology Institute Shaping Our Future Symposium, 31 January 2024.

[2] Department of Industry, Science and Resources, Parliament of Australia, *Safe and responsible AI in Australia consultation: Australian Government's interim response* (Report, 17 January 2024).

[3] The need for a gap analysis was also highlighted by the NSW Bar Association's submission to the NSW Legislative Council's Inquiry into Artificial Intelligence, 26 October 2023.

[4] Allsop CJ, 'The role and future of the Federal Court within the Australian judicial system' [2017] *FedJSchol* 12.