

RSIS Commentary is a platform to provide timely and, where appropriate, policy-relevant commentary and analysis of topical and contemporary issues. The authors' views are their own and do not represent the official position of the S. Rajaratnam School of International Studies, NTU. These commentaries may be reproduced with prior permission from RSIS and due recognition to the author(s) and RSIS. Please email to Mr Yang Razali Kassim, Editor RSIS Commentary at RSISPublications@ntu.edu.sg.

Regulating Artificial Intelligence: An Ethical Approach

By Teo Yi-Ling

Synopsis

As the use of artificial intelligence (AI) becomes more pervasive in everyday life, questions are being asked about where the legal and ethical bright lines are – or should be – in regulating its use.

Commentary

ARTIFICIAL INTELLIGENCE or AI has been developing incrementally for the last six decades, but is currently in an explosive period of growth. Key force multipliers have been the proliferation of big data, increasingly powerful computing capabilities, and complex developments in algorithms. AI is now the main differential in the transport, manufacturing, retail, and healthcare industries, deployed in automating jobs that are by nature repetitive and prone to risk.

AI applications for processes relating to asset and risk management, fraud detection, compliance, due diligence, document review, and research are being increasingly deployed in the finance and legal services industries. The United States and Israeli militaries are also investing heavily in the creation of autonomous weapon systems. Such disruption by AI in these contexts is largely justified on the bases of consistency, risk mitigation, and increased efficiencies.

Narrow vs General AI

For clarity, it is useful to distinguish between “narrow” AI (function-specific applications like automated systems and image recognition) and “general” AI (applications that appear to demonstrate sentient behaviour). With “general” AI, a key concern is the apparent surrender of human autonomy to algorithmic functioning.

The question here is where accountability would then lie in situations where AI makes decisions that are opaque to human scrutiny, contrary to human welfare. Fully autonomous weapon systems are an example – these seek out targets, decide on engagement, and carry out that decision, without human intervention.

Opponents of such systems argue for a ban on these, on the bases of unpredictability, susceptibility to algorithmic bias, and ability to cause unintended harm.

This debate cannot only happen within stakeholder silos, as this would be inefficient and inherently biased. It needs to take place among government authorities, civil society organisations, universities, and industry bodies.

As we look to the potential benefits that AI is intended to deliver – sustainable economic growth; smarter living; new job opportunities, a more connected society; security; increased national resilience – a collective debate is critical to ensure that the most ethically appropriate decisions are made towards achieving these goals.

Rationalisation and Regulation

The use of AI has accelerated ahead of our ability to reckon with the ethical and legal issues it is raising. Calls for regulation are being heard worldwide, and in this respect, governments appear to be treading carefully. The challenge is calibrating the balance between encouraging and chilling AI innovation. In rationalising the existence of AI in our lives, the notion *primum non nocere* – “first, do no harm”, is helpful. This can be unpacked as follows:

Firstly, AI’s development must be guided by human welfare. Justice, equality, safety, accountability, respect for privacy, and transparency are some values it must reflect. Secondly, it must enhance our human existence by serving our needs complementarily. Notwithstanding the efficiencies that AI contributes, humans still master the emotional, creative and intellectual domains.

Thirdly, AI’s range and depth of perception, and ability for subjective nuance must become sympathetic with ours, for it to be appropriately and sensitively harnessed in certain contexts that have the potential for harm. Examples are image recognition, the administration of criminal justice, and military targeting.

Ethical Decision-Making, Then Law

The call for a unique set of laws to establish regulatory bright lines has been heard for some time, as current legal frameworks are deemed insufficient to accommodate the existence of AI and its potential consequences. In this respect, the European Commission spent over a year since January 2017 working on a framework of legislative and non-legislative measures, at the behest of the European Parliament.

At the heart of the debate is the legal personality of AI. The law recognises the concept of the artificial legal person – incorporated entities granted a set of rights, powers and obligations for conducting a range of activities, separate from the natural legal persons operating them.

With this arrangement, there are still the elements of control and accountability – human actors operating the artificial platform of the entity. This could be an analogue for the management and control of the development and manifestation of AI. However, given the manner in which legal systems are structured, the law is always caught on the back foot in providing answers for novel situations.

When it comes to determining causation of harm and assessing liability, it is an exercise in shutting the stable door after the horse has bolted. Law and regulation may not be the best approaches to start with.

Alternative Approach

Ahead of constructing any legal framework, the following approach is posited as an alternative:

- Deeply engage with the contextual scenarios presented by the current landscape of AI development and deployment: where the innovation is happening; where the disruptions are taking place and the attendant net effects; who the main actors are.
- Debate the values that we place on AI innovation and its goals in and of themselves, in contrast to how such innovation benefits humankind: its strengths, weaknesses, the opportunities it provides, and the threats it presents.
- Clarify the philosophical basis (or bases) that justifies and drives our conclusions on either side of the argument: do we call on utilitarianism, adopt an attitude of moderation and temperance, act only with the welfare of humans in mind, employ a rationally egalitarian approach, or be persuaded by absolute virtue?
- Clarify where the competing loyalties to the stakeholders in these contexts lie, and subsequently prioritising them.

The above model for discussion and eventual decision-making is the Potter Box model, originally developed by Ralph B. Potter, Jr., professor of social ethics emeritus at Harvard Divinity School. He designed it as a systematic discipline for the purpose of resolving ethical dilemmas.

The genius of the model is its applicability to almost any dilemma requiring ethical decision-making. Whatever decision emerges at the end is a result of the decision-makers' fully-considered rationalisation.

Teo Yi-Ling is a Senior Fellow with the Centre of Excellence for National Security (CENS), S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University.
