6 June 2023

Australian Government

Department of Industry, Science and Resources

Consultation hub

RE: Supporting Responsible AI

The discussion paper on supporting responsible AI available on the consultation hub is an excellent resource that does well to provide an overview of the topic along with an analysis of the existing international and domestic landscape as it relates to the opportunities and challenges presented by AI. However, it falls short in one very specific but important way. In fact, that way is a way many government attempts at regulation fail. That is, it suggests an internal self-assessment-based approach to managing "medium risk" ("high impact") scenarios that include the use of AI-enabled applications involving the use of sensitive personal (financial or health) information and where decision making (such as access to emergency services or medical treatment) is critical. The government should either re-categorise these types of scenarios as being high risk, or preferably fill the void of monitoring in the "medium risk" category with external assurance requirements.

Background

Page 33 of the 'Safe and responsible AI in Australia' paper available on the Consultation Hub website refers to "medium risk" as involving "high impacts that are ongoing and difficult to reverse" with example/indicative use cases involving:

- AI-enabled application that preliminarily assesses a business loan applicant's creditworthiness
- Use of generative AI in educational settings to assess the performance of teachers and students
- Use of AI-enabled chatbots to direct citizens to essential or emergency services
- AI-enabled applications in hiring and employee evaluation processes
- Use of AI to generate patient records in care settings

There is an abundance of potential AI applications that would fall into this category.

Presumably based on some kind of fear of receiving criticism based on over regulation, the paper suggests a "comprehensive and specific self-assessment" which alone could be considered paradoxical to many yet is multiplied when combined with a monitoring system based on "special internal frequent monitoring" truly seems undercooked for this level of risk and given the inherent rush of innovation and inability of regulation to keep pace with technological change. It is common for any business of size to undergo an external/independent audit of their financials (even private companies) so why does the government shy away from at least encouraging external/independent validation of AI applications as a way of creating transparency and building trust? Of most concern are those applications that may have a "high impact" in sectors that are otherwise highly regulated such as critical infrastructure (e.g., financial services, energy, communications, transport etc).

Possible solutions

Section 4 of the discussion paper presents an approach based on a broad set of general regulations, sector-specific regulation, and self-regulation. Self-regulation has proven inadequate to achieve cyber resilience. General regulations appear to have utility but are not specific enough in terms of the use of AI in industry specific applications. Therefore, sector-specific regulation is necessary and should be preceded by regulators taking an informative and consultative approach to developing standards in their sector for medium and high-risk activities that are commensurate with the risk and application. It is likely that regulators will need to rely on expertise from relevant groups such as the National Science and Technology Council, the National AI Centre and the nascent Responsible AI Adopt program centres (or similar).

The paper also refers to the International Organization for Standardization (ISO) and its ISO/IEC JTC/1 SC42 and the IEEE which are developing international standards in AI. One of these is ISO/IEC DIS 42001 which provides a management system for AI in a similar way that ISO/IEC 27001 does for information security, privacy and cyber security, ISO 9001 does for Quality, ISO 45001 does for OHS/WHS, and ISO 14001 does for the environment. These are opt-in independent assurance mechanisms that are not perfect, but they are risk-based, can be augmented by sector specific guidance, can be used to build trust with external parties, and can demonstrate continual improvement in a way that self-assessment and internal monitoring alone will never achieve.

Bans should be applied sparingly because they deter innovation and forgo the opportunity to achieve better outcomes than we do today. Designing required attributes like security into AI applications based on standards is an essential part of a risk management approach, along with providing external assurance. A risk management approach leaving the adoption of AI applications in "high impact" scenarios to self-assessment and internal monitoring alone is simply not enough.

Kind regards,

A Robinson

Andrew Robinson
Co-founder & Chief Security Officer
(ASD IRAP, ISO 27001 LA, CISSP, CISM)
6clicks

andrew@6clicks.com
Melbourne, Australia
6clicks.com