Response to Call for Submissions: Safe and responsible AI in Australia discussion paper.

Based upon our experience in designing risk management frameworks for AI-based capabilities and advising Defence industry on the lawful design and use of AI capabilities in a military context, including authoring the Trusted Autonomous Systems' Responsible for AI in Defence Toolkit, we offer the following specific observations.

Civilian principles must inform military principles

We assess that a large number of AI uses will be 'dual-use', or capable of being interconnected with military systems, and that the outlier use cases will be exclusively military, based on our experience in developing the Responsible AI for Defence Toolkit.

For example, the use of AI in facilitating at SmartGates has interoperability considerations that may be directly relevant to Australia's military forces in the following potential situations:

- military biometrics operations outside of Australia; and
- general military intelligence operations in the event of a call out of Australia's military forces to support a law enforcement response to counter-terrorism operations under Part IIIAAA of the *Defence Act 1903* (Cth).

Standards adopted in civilian environment will inevitably impact the standards applicable in a military environment, and complementarily, novel (dual-)use applications of AI in the military environment will be required to meet civilian standards. While many military capabilities have very specific testing, evaluation, validation and verification standards (and ultimately legal certification standards in the case of weapons, means and methods of warfare in particular), many of the design choices that defence industry face, in relation to future legal compliance for their proposed capability use-cases, are the same as those of civilian capabilities. The increasingly ubiquitous nature of AI capabilities also necessitates that many legislative considerations impacting the design and development of military AI capabilities will also impact civilian ones.

We also note the concurrent legislative reform underway that requires consideration and nesting within any adopted responsible AI regulatory strategy. These legislative reforms also impact the use of these novel capabilities by the military (such as for example the existing Privacy Act limitations in relation to the use and collection of biometric information by the ADF); and thus, we consider that the integration of military responsible AI policies with general AI use policies must be conducted in parallel.

Industry requires guidance on design principles at the earliest stages

We have analysed over 70 different AI principles, regulatory models and States, regional and international organizations during our project co-authoring the Responsible AI for Defence Toolkit, which included assessing Australia's key allies' frameworks and the OECD's AI Principles. We note that, despite the absence of a set of Australian AI principles that apply to military practice there is still a requirement for clearly articulating which standards applicable to civilian use are expected to apply to the Australian military application of AI.

The standards, morals and ethics of the Australian public are also those expected of the Australian Defence Force.

Examples like the US' NIST Risk Management Framework Playbook, which provides significantly more detail about how to operationalize the AI principles into the design of AI capability, are commendable in providing more information to designers and developers. They provide useful points of reference to assist in guiding industry design legally compliance AI systems, however, there is insufficient detail in these playbooks to provide significant guidance to ensure compliance with existing legal frameworks, nor are they capable of addressing the specifics of compliance with existing Australian legal requirements (nor designed for that purpose).

Further, we consider that there is a need to consider the interoperability of any chosen framework, noting that other regions and States are somewhat more advanced in considering issues relating to privacy than the current Australian legal framework (such as the European Union's draft AI Act and GDPR regime, for example).

The provision of design requirements, incorporated at the earliest possible stage of development and innovation, will ensure that Australian industry is not later hampered by regulatory compliance requirements, but rather, empowered by their consideration as part of the design process. The value-sensitive design movement, led by States such as the Netherlands, provide useful examples of the utility of such approached.

The early incorporation of ethical and legal considerations into AI capabilities is critical to the integration of these considerations in the final product. Every decision undertaken by a software developer designing a system to make recommendations for human implementation require some kind of value judgements as to what is included, excluded, or the design approach taken to achieve a particular outcome. Equally, the ability to ensure sufficient transparency measures are incorporated or embedded into the system design ensures a higher likelihood of success in understanding the system when it is eventually released for commercial or governmental use.

Voluntary frameworks do not provide surety and mandatory frameworks do not stifle *wanted* innovation

We consider that the approach of regulation that has been adopted through the NSW Government's AI Assurance Framework specifically as it relates to the application of a mandatory risk assessment to the adoption of all AI contracts is a good first step; however, the design of this approach requires a high degree of self-assessment by designers and developers, who may in some cases be – understandably– myopic about the promise of their designs, as compared to the potential risk of their novel capabilities.

We consider a risk-based assessment process, incorporating a breadth of factors that apply to many different technological uses for AI, can be used as a mandatory requirement for product certification prior to use; and thus, the flexibility of the risk-based approach does not stifle innovation. Rather, in applying a risk-based methodology, appropriate resources can be allocated to mitigating the greatest risks with particular AI use cases, ensuring better product safety and compliance, as well as ultimately having greater efficacy for industry.

However, we consider the requirement for such standards to be adopted voluntarily is not sufficient. There is a need for consistency in approach, and minimization of the risk of AI development outside of design specifications and standards without adhering to the expected standards. Equally, given the proliferation of standards worldwide, there is risk of adopting standards that do not comply with Australia's unique legislative requirements or values, if a voluntary system is proposed. Further, aligned to the 'greenwashing' phenomenon, we consider there is risk in not mandating a standard, that consumers could be misled in relation to the compliance of an AI system, if a specific obligation is not mandated.

If a voluntary code is the preferred approach for commercial entities, then we consider that a mandatory code should be the minimum requirement for any governmental acquisitions, noting the recent Robodebt Royal Commission findings demonstrate the risk that may arise to citizens when software systems are applied without appropriate mandatory safeguards.

Responsible AI principles must apply to use context and not only capabilities

Finally, context of use is critical to understanding risk in the regulation of AI. In the same way that the European Union Artificial Intelligence Act draft legislation proposes to introduce a number of 'red-line' uses of AI, we consider it is useful to consider the use of AI in a holistic way as use in one context may amount to a red-line use in another.

This is particularly relevant in 'dual-use' situations involving military and civilian use of AI capabilities. For example, suppose a hypothetical situation where the Australian Border Force (ABF) commences the use of an AI-enabled capability to support maritime surveillance activities for Australian law enforcement purposes. The use of the AI capability by the ABF would be primarily regulated by Australia's legal frameworks; but that same capability could also be utilized in support of an Australian Defence Force military operation occurring outside Australia. The ADF's use would be subject to international agreements and/or the laws applicable during an armed conflict (thus altering the extent to which all Australian laws might apply). The legal frameworks applicable to each use case are potentially vastly different.

Risk mitigation approach ensures broadest applicability

We consider that adopting a risk mitigation approach is consistent with best practice regulatory approaches developed in other fields (such as cyber security); and commend the consideration of some of the current commentary on the US approach to regulating cybersecurity requirements to aid in identifying the types of regulatory approaches that can meaningfully be implemented in regulating novel technology such as AI (see for example, Dempsey's commentary on performance-based versus regulation). We concur with the assessment that any adopted regulatory strategy requires a combination of means-based, performance-based and management-based solutions; and thus, operationalising such a system could be best achieved through a risk mitigation strategy.

Further information

We welcome any opportunity to discuss this submission further, on the details previously provided.

Dr Lauren Sanders, CSC | Director, Responsible Innovation Legal
Ms Penny Saultry, AM, CSC, | Director, Responsible Innovation Legal