MDR SECURITY

PO Box 7196
Yarralumla
ACT 2600

Our ref: MDR-MKT-C007, 12 September 2026

Submission to the Joint Select Committee on Artificial Intelligence

About MDR Security

MDR Security is a Canberra-based advisory firm specialising in the security and adoption of critical and emerging technologies, including AI, cyber security and quantum technologies. Our work sits where the technology meets the business, operational and policy settings in which it is used: we advise on what the technology can and cannot do, and what that means for how organisations can ensure it is both useful and secure.

Our Managing Director, Dr Rajiv Shah, holds a PhD in physics from the University of Cambridge, and has over 30 years of experience in big data systems, artificial intelligence and cyber security. He was previously the inaugural Australian General Manager of BAE Systems Applied Intelligence, and is a director of the Australian Information Security Association and a Fellow of the Australian Strategic Policy Institute. However, this submission is made on behalf of MDR Security and does not represent the views of any other organisation. Dr Shah would be very happy to provide further evidence at the Committee's hearings.

Recommendations

1. Every implementation of AI should be designed on the assumption that the AI will sometimes be wrong and sometimes be unavailable. Commonwealth agencies and critical infrastructure operators should be required to document how errors are detected and corrected, and how the service continues without the AI system.
2. "Sovereign AI capability" has different meanings to different people. If the term cannot be avoided, it should be defined in terms of assured access, trust and leverage, not some misguided idea of complete self-sufficiency. Investment to develop sovereignty should be guided by the theory of comparative advantage, focusing on areas where Australia can deliver capability that others depend on. The Government should convene economists, technical experts and end-users to work together to identify the most promising areas.
3. "Data sovereignty" – normally taken to refer to the physical location of data – is similarly confusing and can create false reassurance. Assessments of foreign interference risk should not simply focus on where data is stored, but instead consider who operates an AI system, which foreign governments can compel access to the data or withdrawal of the service, and whether the function could keep running if a foreign supplier withdrew.
4. For critical applications, narrow, domain-specific and locally deployable models should be given equal consideration with large general-purpose models, weighing reliability, auditability, energy and water use and resilience, not just benchmark performance.
5. Accountability for AI systems should remain with the people and organisations that build, deploy and operate them. 'The AI did it' should never be a defence.
6. The Australian Public Service should protect genuine, well-designed experimentation to be able to realise the benefits of AI, while holding AI systems that make or inform decisions about people to strict standards of lawfulness, accuracy, transparency and human accountability.
7. AI safety and security should be integrated into existing cyber security frameworks and methods rather than treated as a separate discipline, and the AI Safety Institute's remit should cover the risks of deploying AI, not only the capabilities of frontier models – especially noting that frontier models may not be the solution for every potential AI application.

MDR SECURITY

PO Box 7196
Yarralumla
ACT 2600

Our ref: MDR-MKT-C007, 12 September 2026

General comments

The committee should bear in mind that artificial intelligence (AI) covers a broad range of technologies that seek to automate complex processing tasks by using statistical and mathematical models, some of which were loosely inspired by theories about the operation of the brain. Very large language models that are hosted in big data centres currently dominate the narrative and capture the imagination, but these are only part of the field of AI. The field also includes narrower, domain-specific models and a wide range of machine learning techniques applied to structured data such as transactions, sensor readings and images, much of which was well established long before ChatGPT came along.

Debates about whether AI systems can "think" or have "consciousness" are a distraction, because there is no agreed definition of either concept, let alone a test for it[1]. More importantly, such terminology is dangerous: it lets the individuals and organisations who build and use these systems deflect responsibility for what the systems do. That principle runs through the rest of this submission.

Comments on the terms of reference

(a) the opportunities presented by artificial intelligence (AI) to lift Australia's resilience, productivity growth, economic competitiveness, and living standards

AI has the potential to be a major transformative technology, enabling the automation of a broad range of tasks. However, it is important not to be swept up in the hype – including claims that 'artificial general intelligence' has arrived, most recently from Nvidia's chief executive on 6 September 2026. Today's most capable systems are broad, but not reliable enough to be trusted unsupervised in most business processes, and turning them into something an organisation can depend on (integrating them with its data and processes, testing them and supervising them) takes substantial human effort. We also have to be realistic that no-one can reliably predict what the most effective applications and use cases might be for a particular organisation, with its own data and processes; that has to be discovered through experimentation. Our approach and mindset need to be flexible, to accept failures as valuable lessons and to learn and adapt from these. Critics who point to the number of AI pilots that "fail" are missing the point– if we knew what was worthwhile and what wasn't there would be no need to run pilots. However, such pilots do need to be designed as experiments, with a clear hypothesis, success and stop criteria set in advance, an accountable business owner, and a realistic path to production if they work.

(e) the opportunities for Australian sovereign AI capability, including the measures required to build, retain and commercialise domestic capability, grow research and development, and attract and retain investment and talent;

"Sovereign capability" is a misunderstood term, meaning different things to different people[2]. In particular, it should not be interpreted as requiring every physical component to come from Australia and every line of code written in Australia, or used to justify government support for uneconomic and uncompetitive business propositions. Australia does not have the scale to be self-sufficient across the whole technology stack; instead it needs to leverage the economic theory of comparative advantage to carve out a position that retains agency and autonomy while collaborating with others. Potential areas of comparative advantage include not only our critical minerals, land and energy supplies, but also

[1] https://www.aspistrategist.org.au/can-artificial-intelligence-really-think-and-do-we-care/
[2] https://www.aspistrategist.org.au/sovereign-capability-can-benefit-australia-up-to-a-point/

MDR SECURITY

PO Box 7196
Yarralumla
ACT 2600

Our ref: MDR-MKT-C007, 12 September 2026

algorithm development and other technical capabilities. Economists, technical experts and end-users should work together to identify the areas where investment and government support should be focused.

Dimensions that should be considered include:

- Ensuring that we have AI systems that are suited to what Australia needs, ie systems suited to the applications that provide value to our economy, trained on data that is appropriate to our context, and which are reliable for the purposes we use them for.
- Ensuring security of supply, ie that when we need access to AI systems we are confident we can get it, particularly where we integrate them into critical processes. This could include minimising dependence on using "software as a service" where the provider could cut off access without notice[3], but also where we do have dependencies on other countries (eg supply of high end GPUs) offsetting this with our own comparative advantage in other capabilities that provide opportunities for leverage.
- Ensuring that we have confidence in the suppliers of the systems, that they are providing products that are safe, reliable and trustworthy, backed by evidence of testing, transparency about changes and incidents, and advance notice before a model that a critical process depends on is changed or retired.

(f) the importance of data sovereignty, and the opportunities presented by narrow, domain-specific AI models trained on Australian data in areas of Australian sectoral expertise;

This term of reference combines four distinct ideas, which are best considered separately:

(i) Data sovereignty: this term is normally used to describe where data is physically stored. This is important in terms of the laws that may apply to how the data can be used, but it is only relevant in controlling the actions of those who care about complying with Australian law. In a world where ICT systems are virtualised and hyperconnected, the physical location of electronic data storage only mitigates a small subsection of risks. Hackers do not care where a system is located. A multinational operating a facility in Australia may still be compelled by its home government to hand over the data it holds – as the US CLOUD Act allows – or to cut off service to Australian users, as the June 2026 Anthropic example shows. The trustworthiness of a supplier and its supply chain is not necessarily related to where the hardware happens to sit.

(ii) Narrow, domain-specific models: much of the discourse on "AI" has defaulted to running larger and larger language models in ever bigger data centres. While the idea of one large multi-purpose model to do everything undoubtedly makes initial adoption and experimentation easier, we believe that more effort should be put into considering the trade-off against narrower, domain specific models, especially for key critical applications. They may require more effort to train and set up, but they can be more predictable within their domain and far more efficient, reducing the demands for computing hardware, power and water to run the AI systems that can deliver the potential benefits to Australia. Smaller models can also run on

[3] This risk is no longer hypothetical. In June 2026 the US Government directed Anthropic to suspend access to two of its most advanced models by any foreign national – which includes every Australian – and the company withdrew them from all customers worldwide for almost three weeks.

MDR SECURITY

PO Box 7196
Yarralumla
ACT 2600

Our ref: MDR-MKT-C007, 12 September 2026

infrastructure under Australian control, which addresses the security of supply concerns raised above: a model running on your own hardware cannot be switched off by someone else.

(iii) Models trained on Australian data: there are a number of cases where this may make sense, for example systems that rely on characteristics of the Australian people, Australian law, local flora, fauna and geology etc. However, we need to be clear that models should be trained on Australian data where this makes economic or practical sense – eg making training cheaper, or making the results more accurate in the context where they will be used. The temptation should be avoided to stray into misplaced "nationalistic pride" that a model is somehow intrinsically better because it only uses Australian data, at the cost of access to the most effective models and systems.

(iv) Areas of Australian sectoral expertise: Australia should identify and nurture areas of sectoral expertise not only for domestic use but for export, establishing a critical role in AI systems that other countries rely on, which in turn gives us leverage over capabilities we rely on them for.

(h) the adoption of AI by Commonwealth departments and agencies to improve services for Australians, and the transparency and accountability measures that apply to AI use;

The need for experimentation described under (a) applies equally to Commonwealth agencies. The political environment must allow and encourage appropriate experimentation, with the right guardrails, and avoid public servants being stifled by a fear of trying something that is seen to fail. The Digital Transformation Agency's Policy for the responsible use of AI in government already exempts early-stage experimentation from its impact assessment requirements, so the barrier is cultural rather than regulatory. The committee should draw a clear line between experiments, where some failure is expected and should be tolerated, and production systems that make or inform decisions about people, which must meet strict standards of lawfulness, accuracy, transparency and human accountability. Robodebt is instructive: it was not an AI system but a simple automated calculation applied unlawfully and at scale, and its lessons are about accountability rather than technology.

We also emphasis that agencies need access to independent technical expertise to be intelligent customers of AI suppliers, rather than relying on suppliers to assess their own products.

(i) the adequacy of Australia's existing laws and regulatory frameworks as they apply to AI and whether there are any gaps that warrant reform;

Existing laws already do much of the work, and the committee should resist the framing of AI as a regulatory vacuum. We should not be seduced by talk of autonomous systems, as though AI systems have their own free will and no human is responsible for the actions of such systems. The people and organisations that build and operate such systems must be accountable for their actions under the existing laws and regulations. Because every AI system makes mistakes (see (m) below), no operator should be able to claim ignorance of the risk. This means that if they fail to supervise or monitor their systems effectively, they cannot evade liability for the harms that result.

(k) the risks and harms arising from AI, including fraud and scams, deepfakes and risks to the safety and wellbeing of children and vulnerable Australians;

We limit our comments here to fraud and scams. The ability of generative AI to create realistic voice and video impersonations at modest cost challenges many of our assumptions about how we verify who we are communicating with, and how actions are authorised. Relying on AI-detection systems will never be

MDR SECURITY

PO Box 7196
Yarralumla
ACT 2600

Our ref: MDR-MKT-C007, 12 September 2026

an effective long-term solution, as generative capabilities will always evolve ahead of the ability to detect them. Reducing these risks depends instead on fundamental security architecture: robust digital identity[4], and out-of-band verification of payment and access requests – and in particular more work is needed on digital identity to promote its adoption and ensure its end-to-end security[5].

(l) the national security and cyber security opportunities to protect Australians, as well as the foreign interference risks, associated with AI, including the security and resilience of supply chains and of systems supporting critical infrastructure;

We will limit our comments here to the cyber security risks and opportunities of AI (although see the earlier comments on sovereignty which relate to supply chain resilience and security). The proliferation of AI tools will make it easier for many cyber attackers to put together attacks, and to execute them at scale over a large range of targets. This will require organisations to ensure they have taken appropriate measures to defend their systems, but we should avoid reflexively assuming that "we need AI to defend against AI". Most successful cyberattacks today exploit failures of basic configuration and security hygiene; AI will let attackers find and exploit such failures faster and at greater scale. The answer is for organisations first to get better at the basics: configuring systems correctly, patching promptly, not connecting systems to the internet unnecessarily and segmenting networks. Where AI might help defenders, it will be by doing those basics faster, particularly finding and fixing vulnerabilities before attackers do. That is where investment in cyber defence should go, rather than into the latest agentic AI security product.

(m) the implications of emerging AI capability for Australia's national security and strategic resilience, including the ability of regulators, the Australian AI Safety Institute and the intelligence and security community to identify and respond to emerging risks;

Key points which the committee should consider include:

- AI systems will always make some mistakes. We can seek to minimise the chance of errors but never eliminate them. If we had a definitive, deterministic model of how to process a set of inputs to get the correct outputs we would write software to do that. We use AI when we don't have such a definitive model; instead we train a model on examples, and it learns to approximate the patterns in that data and generalise, imperfectly, to inputs it has not seen. Therefore, every time a process relies on AI, the process must include consideration of what happens when the AI makes a mistake.
- If a process relies on AI, we should also always consider what happens if the AI system is not available. Maybe the software crashes or malfunctions, or maybe we lose the connectivity to where the AI model is running, or the supplier withdraws access – what is the business continuity plan in this case?
- The data that goes into and out of AI systems is often highly sensitive and valuable. Therefore, there needs to be a clear understanding of where this data goes, who has access to it and how it might be used; this is especially true of AI systems that run remotely (eg in data centres) or are operated by software-as-a-service providers on behalf of end-users.

[4] https://www.aspi.org.au/report/australias-new-digital-id-system-finding-right-way-implement-it/
[5] https://www.aspistrategist.org.au/getting-australias-digital-trust-exchange-right

PO Box 7196
Yarralumla
ACT 2600

Our ref: MDR-MKT-C007, 12 September 2026

We welcome the establishment of the AI Safety Institute. However, much public discussion still treats "AI safety" as a distinct discipline, focused on the capabilities of frontier models and separate from cyber security. For most Australian organisations, the risk also lies in deployment: connecting AI systems to their data and business systems, and giving AI agents permission to act, creates new routes for data leakage and manipulation – for example 'prompt injection', where instructions hidden in a document or email hijack the AI system that reads it. That is a security engineering problem.

Effective risk management to ensure that new technology can be adopted safely and with confidence is what cybersecurity professionals have been doing for decades. In the early days of AI, cyber security teams were often seen as the "department of no" and people went around them to try out the shiny new AI toys. They should now be an integral part of understanding these risks and working out how best to manage them.

Closing remarks

AI is a technical subject, albeit one with broad societal and economic impacts. AI policy that is not grounded in how the technology actually works will miss its mark. It is good that we see non-technical stakeholders from industry, government, academia and civil society taking an active part in these discussions, but this needs to be alongside, not instead of, technical experts. We encourage the committee to take evidence from those who understand, and can explain, the underlying technology, alongside those who study its implications.

Rajiv Shah, Managing Director

Contact details: [redacted]