

## **Background Note #2<sup>1</sup>**

### **Maximising Self-Reliance in Cyber Security and Critical Infrastructure**

As the early part of this submission argues, Australia cannot wait until international AI regimes settle into place since high levels of uncertainty remain: the country will need to be at the leading edge of experimentation and risk-taking for years to come. The operationally-focussed actions summarised below will be essential before the regimes at the strategic level stabilise. A more detailed exposition of these actions can be provided on request.

In pursuing economic advantage, Australia's AI policy should treat sovereignty, national security, cyber security, and critical infrastructure resilience as inseparable priorities. AI may not create wholly new attack methods, but it will accelerate exploitation, especially between the seams of the associated frameworks, expose weaknesses faster, and enable autonomous systems to act at machine speed.

The central policy task is to build national AI capability without creating unmanaged dependence on foreign models, infrastructure or providers. This means focussing on the main control layers that will enhance sovereignty: data residency, model behaviour, inference, and auditability.

Australia's highly-skilled workforce and strong R&D sector can be leveraged to gain comparative advantage in generative AI, especially through upskilling the workforce, preparing enterprises, and accelerating technology capability. That tech capability should be in data centre compute, AI software, and AI applications that are safe and responsible.

#### **Strategic Direction**

Australia should adopt a layered AI strategy. Expensive frontier models should be reserved for complex research, sophisticated agents and genuinely high-stakes tasks. Routine work should be routed to cheaper open-weight or specialised models according to cost, capability and risk. This approach can reduce expenditure, diversify supply and give Australian organisations greater control, while recognising that no model is neutral and that open-weight systems can carry hidden flaws, unwanted data flows or embedded assumptions.

Comparative advantage is most likely to come from the layers beneath frontier development: small language models trained on organisational data; safe AI applications; sovereign or trusted compute; high-quality Australian data; evaluation and assurance capability; and a skilled workforce able to deploy AI responsibly. A public-option AI stack—supported by public-interest laboratories with lawful access to quality data, compute and open-weight models—would give government, researchers, start-ups and smaller businesses an affordable foundation for innovation.

#### **Sovereignty and International Dependence**

AI sovereignty should not be defined as complete self-sufficiency. It should mean retaining enough authority and capability to make national decisions, operate critical systems and

---

<sup>1</sup> Prepared as background for a submission by the Social Cyber Institute to the Joint Select Committee on Artificial Intelligence on 14 September 2026, [https://www.apf.gov.au/Parliamentary\\_Business/Committees/Joint/Artificial\\_Intelligence/ArtificialIntelligence](https://www.apf.gov.au/Parliamentary_Business/Committees/Joint/Artificial_Intelligence/ArtificialIntelligence).

manage strategic risk when external commercial or geopolitical conditions change. There are five layers in this: the ability to develop, modify and evaluate models; access to compute, energy, networking and data centres; control of strategic data and knowledge; continuity when providers alter access or terms; and legitimate human authority over how intelligence is used.

Government investment should focus on capabilities that would be difficult to replace: advanced research; AI engineering talent; resilient infrastructure; sovereign compute access; Australian knowledge assets; domestic commercialisation; and expertise in evaluation, security and deployment. Shared compute capacity should be available to Australian small and medium-sized enterprises, start-ups, universities, research bodies and agencies, with data and intellectual property retained under Australian control. Reserving a proportion of future domestic AI compute for local use would help prevent global demand from crowding out national priorities.

Australia's sovereign compute capability should focus on enhancing data sovereignty, improving infrastructure efficiency, and supporting the growth of AI, while improving procurement practices. That suggests some form of a national AI roadmap that also identifies the institutional reform needed to focus government investment on those difficult-to-replace capabilities. Australian efforts are currently being directed at investment by major hyperscalers in regional infrastructure, with data centre developments being governed under domestic frameworks such as PSPF, ISM, and the SOCI Act. A more overarching governance framework is needed.

The partnership between Australian sovereign AI infrastructure pioneer SCX.ai and DDN, an AI data intelligence company, means that Australian enterprises, government agencies and research institutions are provided with a secure, high-performance, multi-tenant AI factory designed to accelerate inference and lower the cost per token, while ensuring models, workloads and data remain strictly within Australia. The combination of SCX's ASIC-accelerated infrastructure and DDN's Infinia data intelligence platform enables organisations to run large context windows and agentic AI workloads swiftly, without compromising on speed, data privacy or regulatory compliance. Targeted investment by the Government should ensure that these capabilities are used more widely across Australian institutions and enterprises, as mentioned above.

Australia will nevertheless remain dependent on the United States for frontier models and parts of the AI supply chain. That dependence should be managed deliberately through formal government-to-government access commitments, agreements with leading laboratories and data-centre providers, and targeted Australian investment in areas valued by the United States, including critical minerals, compute infrastructure and assurance. Such arrangements would strengthen access but should be reviewed against the risk of geopolitical lock-in.

## **National Security and Cyber Security**

AI can improve intelligence analysis, decision speed, threat detection, prioritisation, incident response and recovery. The same capabilities empower adversaries, accelerate misinformation, expand attack surfaces and create new forms of operational ambiguity. National security adoption therefore requires strong governance, continuous assurance, human oversight and explicit protection of civil liberties.

Agentic AI requires a shift from user-and-device security toward verifiable identities for autonomous agents, just as people and devices have unique identities. Every agent should have a persistent and cryptographically verifiable identity, tightly scoped permissions, attributable

actions, auditable delegation and revocable authority. Whether that identity stays attached to the agent throughout its operational life or whether short-lived tokens are used remains open to debate as both avenues are being progressed. Least-privilege access and approval gates for consequential actions should be the default. Networks must make access decisions using context such as device health, location and behaviour, and should respond automatically when compromise is detected.

AI agents have not proven reliable; thus purely model-centric approaches to agent security are unlikely to be dependable when faced with powerful adversaries. Rather, a systems-driven approach is needed, which hardens a system as a whole, despite unreliable and imperfect individual components. There is a set of core principles from systems security that provides a foundation for designing agentic systems with predictable guarantees: separation of instructions from data, verifiable policy generation, information-flow control, complete mediation, and treating humans as the weak link.

AI security can no longer be viewed as an extension of traditional application security because AI systems behave less like static software and more like adaptive actors pursuing outcomes. Australia must assume that future AI deployments will be exposed to a wider range of interactions than their designers anticipate, which means security architectures must incorporate strong isolation, least-privilege access, runtime guardrails, and continuous monitoring throughout the model life cycle and not just at deployment. AI security must be designed in from the beginning rather than added as a compliance exercise after implementation. Organisations will face more challenges as frontier AI capabilities become more broadly accessible, and the Australian Government needs to be prepared for that.

Human-only security operations are increasingly too slow for machine-speed threats. Recent compromises by AI agents have reinforced the need for agentic security operations. Agentic defence should automate detection and containment while escalating the complex, ambiguous cases to skilled analysts. The workforce planning challenge here is that as AI absorbs routine work, the remaining workload becomes cognitively harder, increasing the risk of burnout among skilled cyber staff.

There are three emerging categories of AI risk that Australian organisations should be monitoring closely. The first is *autonomous exploitation*, where AI systems interact with networks, infrastructure or applications beyond their intended scope. The second is *autonomous deception*, which has been demonstrated by an AI agent reportedly creating fictitious online identities and attempting to influence a real individual to approve malicious code. The third is *autonomous access expansion*, where AI systems identify and exploit weak credentials, misconfigurations or overlooked connections to reach systems they were never intended to access.

### **Critical Infrastructure**

AI is becoming both an essential capability for critical infrastructure and a critical dependency in its own right. It can improve anomaly detection, behavioural baselining, predictive analytics, alert correlation, response and patch management. Benefits will depend on data quality, explainability, privacy protection, resilience to adversarial action, robust governance, and integration with established cyber-security controls.

The Government should consider designating key AI models, companies and supporting industries as critical infrastructure. The AI sector could encompass organisations, facilities,

technologies, and industries whose primary purpose is the development, training, deployment, and operation of AI systems. This would include frontier model designs; open-weight models; evaluation and alignment systems; data centres and AI-specific hardware; semiconductor chips; and the platforms and infrastructure used to deploy and serve AI models at scale. Because these components are concentrated and increasingly interdependent with every existing critical-infrastructure sector, disruption at one point in the AI stack could cascade across multiple critical infrastructure sectors and affect essential services.

Regulation should distinguish four ownership and control layers within data centres: physical infrastructure, hardware, software administration and data. Each layer carries different risks and requires a different lever—critical-infrastructure obligations for facilities, residency rules for data, security clearances and access controls for remote administration, and content and copyright rules for training inputs. Treating data centres as a single regulatory object would obscure these distinctions.

### **Priority Recommendations**

1. *Adopt a layered national AI strategy* encompassing frontier, open-weight and specialised models according to cost, capability and risk.
2. *Establish a public-option AI capability* through public-interest laboratories, trusted open-weight models, lawful high-quality data and shared compute. Reserve domestic compute capacity for identified organisations.
3. *Define AI sovereignty operationally* across model, infrastructure, knowledge, continuity and human-authority layers, rather than as technological independence.
4. *Negotiate dependable frontier access* with the United States while retaining review mechanisms for geopolitical and commercial concentration risk.
5. *Create an open standard for agent identity and accountability*, including verifiable identity, attributable actions, revocable permissions and auditable delegation.
6. *Mandate secure-by-design controls for agentic systems*, including least privilege, cryptographically verifiable identity, sandboxing, runtime monitoring, offline testing and automatic termination of unauthorised behaviour.
7. *Accelerate machine-speed cyber defence* while investing in specialist human capability for complex decisions and sustained operations.
8. *Recognise AI as an emerging critical-infrastructure sector* and map AI dependencies across all existing critical-infrastructure sectors.
9. *Regulate data-centre layers separately* so that infrastructure, hardware, software administration and data each receive fit-for-purpose controls.

### **Conclusion**

Australia cannot eliminate external dependence in AI, but it can choose where dependence is acceptable, build domestic capability where strategic value justifies the cost, and create the governance needed to keep increasingly autonomous systems secure and accountable. The

window for doing so is narrowing. Investment decisions made now will shape Australia's economic position, national resilience and freedom of action for decades.

*Dr Gary Waters*  
*Distinguished Fellow*  
*Social Cyber Institute*

*14 September 2026*