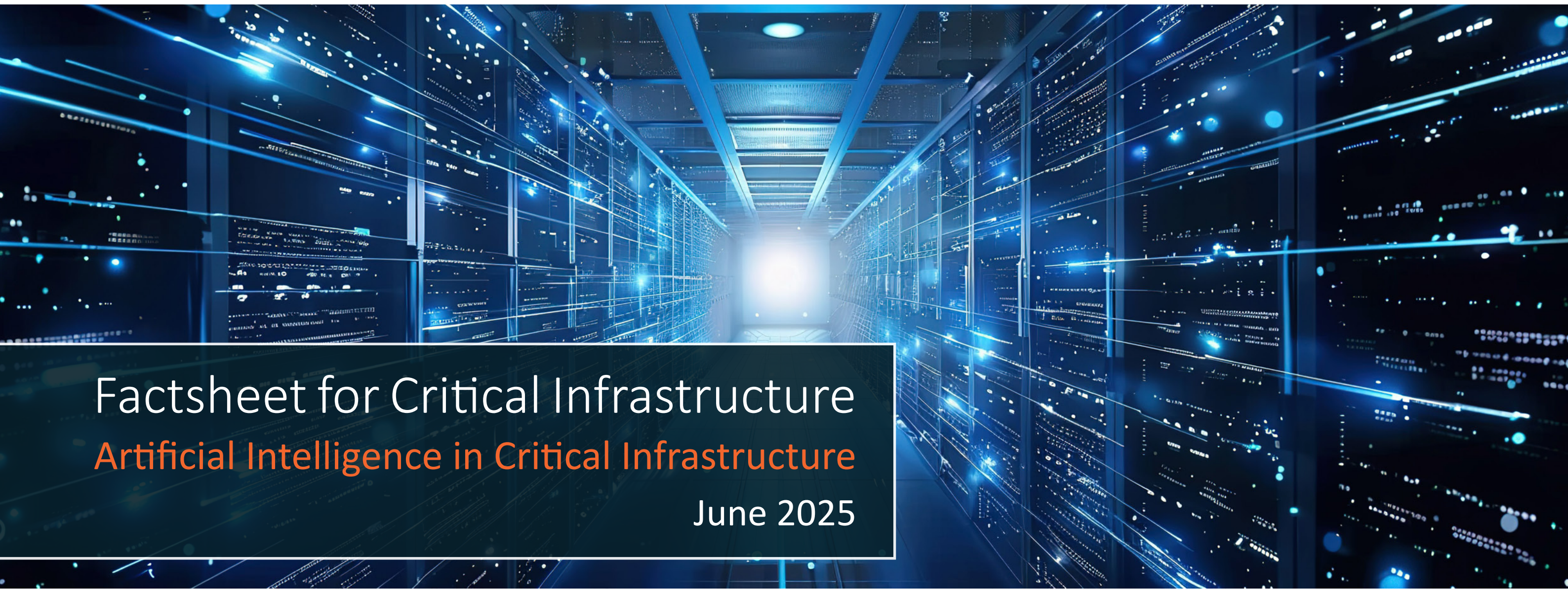
Australian Government
Department of Home Affairs

CYBER AND INFRASTRUCTURE SECURITY CENTRE



Factsheet for Critical Infrastructure

Artificial Intelligence in Critical Infrastructure

June 2025

Artificial Intelligence[1] (AI) has the potential to transform the critical infrastructure environment, delivering wide-ranging benefits across all sectors. AI technologies are distinctive in their speed of innovation, scope of impact and variability of deployment. However, critical infrastructure owners and operators that use (or are considering using) AI systems must also consider its implications for security and safety.

This factsheet provides critical infrastructure owners and operators with an understanding of these risks, their obligations under the *Security of Critical Infrastructure Act 2018* (SOCI Act) and some potential risk mitigations.

What are the risks in AI?

The security and resilience of Australian critical infrastructure is vital and the emergence of AI introduces new risks - an AI-related cyber-security incident could be catastrophic for Australia's economy, security and wider society.

Whilst critical infrastructure owners and operators need to consider their context-specific threats, this factsheet highlights three broad categories of AI risk to consider.[2]

1. Attacks using AI

AI technology will almost certainly make elements of cyber intrusion operations more effective and efficient, which may lead to an increase in the frequency and intensity of cyber threats.[3]

AI can be used to automate, enhance, plan or scale physical attacks (e.g. attacks on physical infrastructure by autonomous systems such as drones) and cyber compromises of systems. AI significantly lowers the technical barriers for carrying out a variety of steps in offensive cyber operations, including malware development and reconnaissance.

Phishing (cyber attacks)

Generative AI can be used to create highly personalised and realistic emails, text messages or social media outreach; scrape and analyse large amounts of personal data; or automate the sending of malicious content. The use of AI to support phishing operations enables malicious actors to more effectively plan and carry out cyber-attacks.[4]

[1] Artificial Intelligence refers to digital systems capable of performing tasks commonly thought of as requiring human intelligence, such as writing meaningful sentences, solving equations, creating art, and navigating obstacles.

[2] Department of Homeland Security, Safety and Security Guidelines for Critical Infrastructure Owners and Operators (2024).

[3] National Cyber Security Centre (NCSC), Impact of AI on cyber threat from now to 2027 (2025).

[4] Australian Signals Directorate (ASD), Engaging with artificial intelligence (2024).

The information contained in this document is general in nature and does not constitute legal advice. Readers are encouraged to obtain legal advice that applies to their particular circumstances. The Commonwealth of Australia does not guarantee the accuracy, currency or completeness of any information in this document.

2. Attacks on AI

As AI systems become more widely adopted in critical infrastructure, they are set to become an attractive vector for exploitation.

Attacks can include data poisoning, model poisoning, and compromise of the underlying system or infrastructure that is hosting the AI capability. Compromise of AI systems that perform or support mission critical functions carries the potential for the most immediate and significant impact. Incorporating AI into any existing structures will increase the cyber-attack surface, providing malicious actors with new channels for compromise.

The novelty, complexity, and lack of operator experience associated with the use of AI in critical infrastructure networks and systems will further increase the risk.

Data poisoning

Data poisoning is the external manipulation of training data to intentionally encourage the AI model to produce inaccurate or biased information. This can occur during the development of an AI model by modifying training data, inserting new data or relying on already poisoned data. Additionally, AI models that utilise feedback to refine outputs remain vulnerable to data poisoning. For example, an AI model can be undermined by malicious actors deliberately providing negative or misleading feedback on otherwise accurate outputs, which results in the AI model learning from incorrect patterns and producing erroneous or biased outputs.[5]

3. Reliability of AI

Relying on AI systems that fail to perform their intended functions could result in catastrophic disruption to critical infrastructure, especially when AI is integrated with sensitive mission critical systems.

Any deficiencies when deploying or maintaining AI systems heightens the risk of malfunctions or unintended consequences for critical infrastructure, as do misconceptions about the capabilities and limitations of AI. These risks are compounded by challenges that may arise when testing and evaluating AI systems, especially generative AI systems.

Reliability issues

Reliability issues arise when AI systems misinterpret information (often referred to as 'AI hallucination', which is internal to the AI model), fail to perform common sense reasoning, or cease to maintain a contextually relevant knowledge base. For example, AI has been observed citing non-existent legal precedence and company policies; failing basic mathematical calculations; and providing inaccurate medical information.[6]

What can be done to mitigate these risks?

Entities can take a number of actions to mitigate the risks presented by AI systems through integrating AI-specific security considerations into existing organisational security frameworks. Ensuring human oversight of AI systems and using a trusted vendor can limit the impacts from both foreign ownership, control and influence (FOCI), and the actions of malicious actors. Once an AI system is embedded in critical infrastructure, maintaining strong cybersecurity hygiene, safeguarding data security and conducting ongoing monitoring and evaluation should remain a priority.

1. Consider the AI deployment environment and ensure human monitoring and control

Entities should undertake risk assessments throughout the lifecycle of any AI product, considering its necessity in light of the potential benefits and harms of its adoption. Operators should be aware of the environment that AI systems operate in, including the systems they interface with to fully understand the risks. When incorporating an AI system into operations, avoid reliance on AI output data for critical systems and ensure a human has oversight of any decision making that is informed by an AI system.

Under the SOCI Act, owners, operators and direct interest holders of critical infrastructure assets are required to establish, maintain, and comply with a critical infrastructure risk management program (CIRMP), including consideration of cyber and information security risks. Responsible entities have broad discretion in how they approach the management of hazards. However, entities are strongly encouraged to consider AI-specific frameworks when assessing AI risks, such as ISO/IEC 42001:2023 or the National Institute of Standards and Technology (NIST) AI Risk Management Framework.

[5] ASD, Engaging with artificial intelligence (2024).
[6] AI Action Summit, International AI Safety Report (2025).

2. Using a trusted AI vendor

If procuring AI systems from vendors, entities must ensure the credibility of those vendors, including consideration of whether the AI systems are sourced from nations with robust legislation and control mechanisms that follow principles of responsible and ethical AI.[7] This includes evaluating supply chain components that support AI systems.

When procuring AI systems from vendors, entities should evaluate any FOCI risks including any that flow from the relevant jurisdiction(s) and their approach to transparency and privacy protection.[8] It is important to be aware of the data collection laws that may apply to AI systems, as well as if extrajudicial direction or terms of use agreements apply. This could include a foreign entity compelling disclosure of sensitive data.

Entities should always consider individual privacy policies, paying close attention to any provisions relating to the collection, use and disclosure of personal or sensitive information by AI systems, including for system training purposes. Under the SOCI Act, entities must establish and maintain a process or system to minimise, mitigate or eliminate supply chain hazards.

3. Maintain strong cyber security hygiene

AI systems may be subject to further attacks if an entity's broader cyber security posture is lacking. Implementing and maintaining strong, entity-wide cyber security measures[9] (e.g. strong encryption, least-privileged access, strong password requirements and multi-factor authentication) can reduce the overall attack surface, including on AI systems and any sensitive data or critical systems it supports. Ensure that employees comply with cyber security policies and any updated policy frameworks when AI and other technologies are introduced.

4. Pay close attention to data security

Entities implementing or procuring AI systems must understand how it collects, processes and stores data. As data that is input into an AI system may also be used to retrain the AI system itself, entities should provide users with clear guidance on what types of data can be introduced into AI systems.

Wherever possible, entities should remove or de-identify personally identifiable or sensitive information and recognise that residual risks remain due to the potential re-engineering of the model and subsequent data extraction by malicious actors.

Depending on the entity's context and mission requirements, the benefits and risks of hosting AI models locally rather than Software-as-a-Service (SaaS) options may need to be considered.

5. Be prepared to respond when risk becomes reality

Entities should continually log and monitor their AI systems inputs and outputs to detect any anomalies or potential malicious activity. This requires entities to have a clear understanding of their AI systems business-as-usual benchmark behaviour to determine when events are anomalous.

Entities should have a response plan in place before any incidents or errors occur. Response plans should consider how their operations may be impacted by an incident or error so that context specific contingencies can be implemented to support business continuity. Under the SOCI Act, critical infrastructure owners and operators are required to report cyber security incidents that have a relevant or significant impact, including those by or involving AI.

What other resources exist?

The Department of Industry, Science and Resources (DISR) has published the Voluntary AI Safety Standard, outlining principles for engaging with AI systems safely and reliably.

The Office of the Australian Information Commissioner (OAIC) has published guidance on privacy and the use of commercially available AI products.

The Department of Home Affairs has published guidance for entities on the SOCI Act, and the Foreign Ownership, Control or Influence (FOCI) Risk Assessment Guidance.

The Australian government has consulted with international partners on a range of resources, including the UK National Cyber Security Centre's (NCSC) guidelines for secure AI system development, and the US Department of Homeland Security's Safety and Security Guidelines for Critical Infrastructure Owners and Operators.

There are several AI-specific risk assessment tools available to entities, including ISO/IEC 42001:2023, and the US National Institute of Standards and Technology's (NIST) AI Risk Management Framework. The MITRE ATLAS collates adversary tactics and techniques against AI systems based on real-world attack observations. A group of 96 independent AI experts contributed to the International AI Safety Report, which provides comprehensive guidance on AI risks and safety.

[7] ASD, Choosing secure and verifiable technologies (2024).
[8] Department of Home Affairs, FOCI Risk Assessment Guidance.
[9] ASD, Foundations for modern defensible architecture (2025).

Australian Government
Department of Home Affairs

CYBER AND
INFRASTRUCTURE SECURITY
CENTRE

Where can I find out more?

Within the Department of Home Affairs, the Cyber and Infrastructure Security Centre (CISC) drives an all-hazards critical infrastructure risk management regime in partnership with government, industry and the broader community. The CISC assists critical infrastructure owners and operators to understand the risk environment and meet their regulatory obligations. More information can be found on the CISC website or by contacting enquiries@CISC.gov.au.

The Australian Signals Directorate (ASD) provides a range of advice at cyber.gov.au to improve cyber security, including AI-specific guidance on engaging with AI, deploying AI systems securely, and AI data security, and guidance on identifying cyber supply chain risks.