



Australian Government
Australian Signals Directorate

cyber.gov.au

ASD Essentials overview

AI considerations within the Essentials

 Essentials

AI within the Essentials

The Essentials framework applies regardless of whether an organisation consumes, develops or deploys artificial intelligence (AI). The series is technology-agnostic and focuses on security outcomes that apply to AI-enabled and traditional systems.

This publication explains considerations that apply to AI across the ASD Essentials series. It also highlights AI-specific considerations when implementing the principles and actions in *ASD Essentials for enterprise IT*, *ASD Essentials for cloud* and *ASD Essentials for OT*.



Considerations for all environments

The following considerations apply across enterprise IT, cloud and OT environments. Organisations should consider AI as a source of cyber security risk, a capability that can support cyber defence and an asset that needs to be identified and governed.

AI as a threat

AI has not fundamentally changed how adversaries gain access to and move through an environment, although AI systems themselves introduce new attack surfaces. AI has made existing attacks faster, cheaper and available to more people, weakening some of the assumptions organisations have relied on for years. AI can help adversaries conduct reconnaissance, identify vulnerabilities and develop exploits or malware more quickly. Synthetic voice and video can also be used to impersonate trusted people, including to deceive help desks or people who approve payments or access.

AI systems can themselves be targeted. For example, adversaries can use indirect prompt injection through emails, documents or web content processed by AI agents with access to tools or data. Other risks include tampered models obtained from public sources and sensitive data being provided to unapproved AI services.

The response lies in layered protection. Where an individual safeguard can be circumvented, an attacker who defeats one measure should encounter another. The implementation sequence in each ASD Essentials series chapter establishes these layers based on the threats, systems and consequences relevant to that environment.

AI as a solution

AI-enabled security capabilities can help organisations process large volumes of data, correlate events across systems, identify vulnerabilities, triage alerts and synthesise cyber threat intelligence. This can help defenders respond more quickly to emerging threats and focus their effort on activities that require human judgement.

These benefits depend on effective governance. AI-enabled security products are themselves AI assets and should be assessed, configured and monitored accordingly. Organisations should use these capabilities to support the actions in the ASD Essentials series rather than replace them.

AI can support analysis and decision-making, but accountability for security outcomes remains with people. Organisations should validate AI-generated outputs before using them to inform significant decisions.

AI as an asset

AI systems are assets within the scope of the ASD Essentials series. Depending on how an organisation uses AI, these assets can include:

- internally developed or fine-tuned models
- infrastructure used to host or train models
- externally provided AI services
- AI-enabled platforms and applications
- training and fine-tuning data
- prompts, outputs and inferences generated during operation
- interfaces, connectors and application programming interfaces (APIs)
- accounts and non-person identities used by AI agents and services.

AI tools can be given identities, permissions and access of their own. Organisations should identify and govern these in the same way as other identities and access pathways.

When scoping domains, organisations should consider an AI asset's function, connectivity, criticality, access and the consequence of its compromise. The relevant ASD Essentials series chapter then determines how the associated principles and actions apply.

Figure 1 shows how the components of an AI system can inform its risk profile and domain mapping.

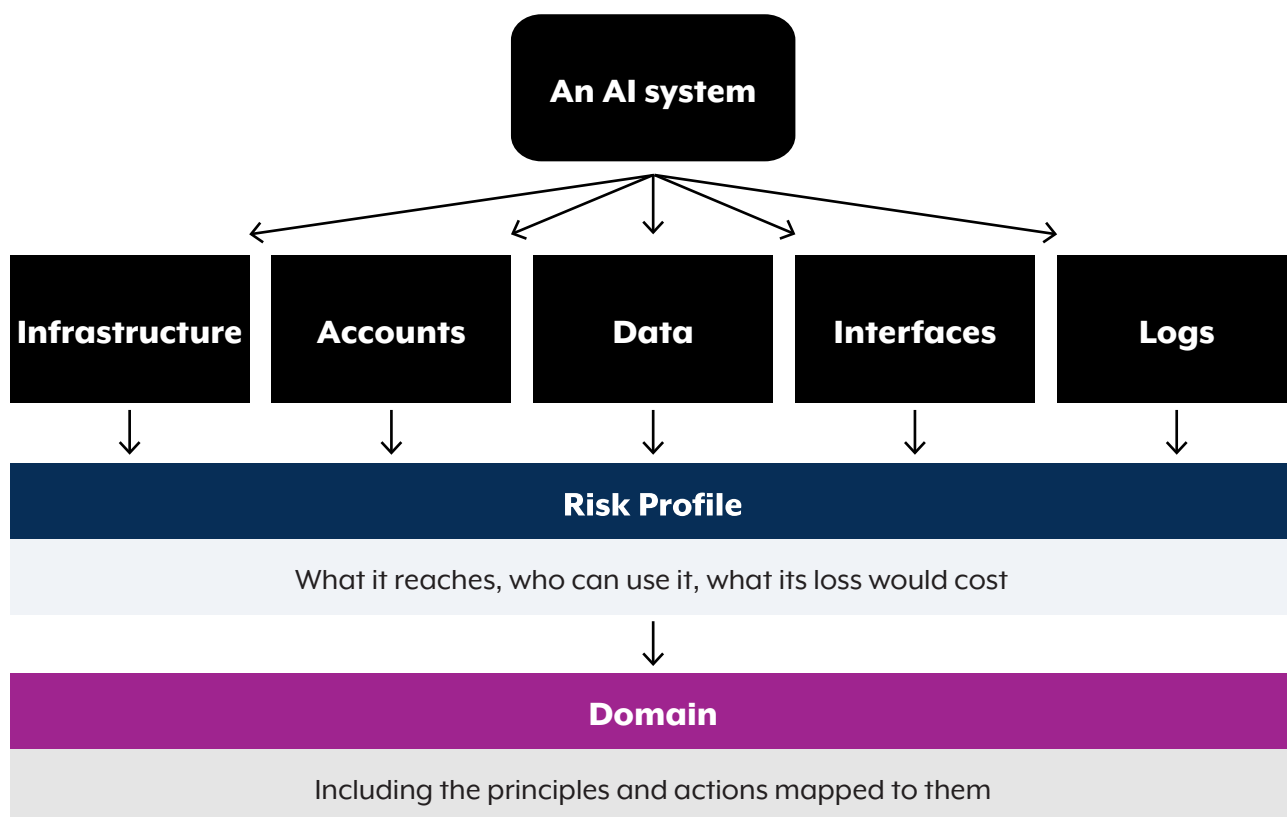


Figure 1. AI system asset, risk profile and domain mapping

Executive AI accountability

AI introduces governance considerations addressed through ASD's *Information Security Manual* **Executive artificial intelligence accountability** principle.

Under this principle, the board of directors or executive committee is accountable for ensuring AI is secure, controllable, human-supervised and used in an ethical and accountable manner.

Organisations should establish and maintain executive accountability for the use, governance and risk management of AI. This should be supported by processes for approval, risk assessment, oversight and ongoing review of AI use cases.

This ensures AI-related technical controls operate within a governance framework that recognises AI as a distinct area of organisational risk.



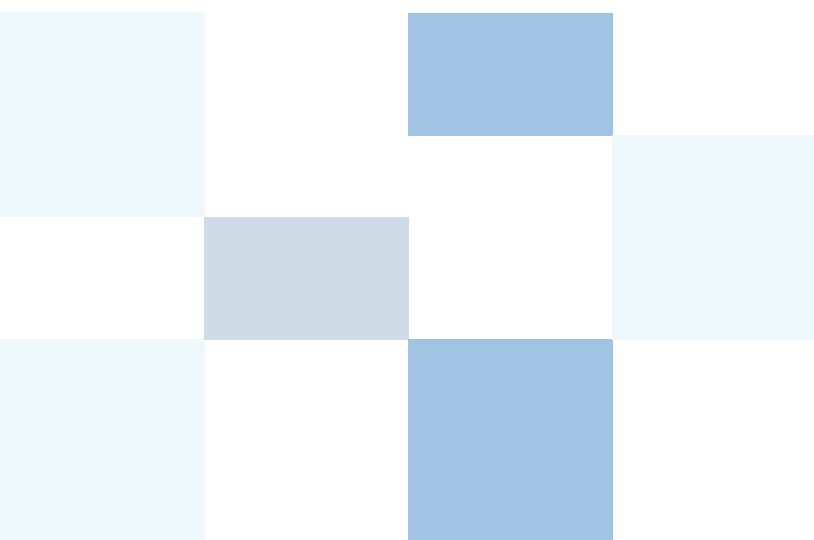
Enterprise IT considerations

AI used in enterprise IT can interact with organisational data, identities, applications, administrative functions and externally accessible services. Organisations should consider these relationships when identifying AI assets and applying the principles and actions in *ASD Essentials for enterprise IT*.

AI systems supporting core organisational outcomes are likely to be scoped within critical processes and services, while AI embedded within general enterprise platforms may belong within business systems. Public-facing AI services and APIs are likely to be relevant to boundary services. AI agents and their non-person identities are relevant to identity services, while AI tools capable of administrative actions belong within privileged environments.

Most enterprise IT principles apply to AI in the same way they apply to other technologies, although implementation may differ. When applying these principles:

- **Asset identification** should include AI systems, models, training data, model artefacts and inference endpoints.
- **Vulnerability management** should address AI-specific risks such as prompt injection, adversarial inputs, model extraction and compromised model weights or training data, in addition to traditional vulnerabilities.
- **Secure system lifecycle** should ensure a manipulated model cannot exceed the access it has been granted, recognising that prompt injection cannot be completely prevented.
- **Cyber supply chain security** should verify the integrity and provenance of model weights, training data and other AI components before use. Organisations should also approve AI tools and integrations before they access organisational data.
- **Supplier cyber security assurance** should establish where AI providers retain prompts and outputs and whether organisational data is used to train models.
- **Secure configuration management** should cover model access controls, inference endpoints, connected data sources, interaction limits and safety settings.



- **Trustworthy software** should address unapproved AI tools, browser extensions, locally run models and agents, including agents capable of executing code.
- **Identity, credential and access management** should apply to users and machine identities, including service principals and managed identities used by AI platforms.
- **Least privilege access** should limit AI agents and integrations to the minimum data and permissions required and address overshared data that AI assistants may surface.
- **Secure administration** should ensure AI tools capable of administrative actions operate through privileged environments and that their actions are recorded.
- **Centralised event logging** should capture relevant prompts, outputs, retrieval activities and administrative actions while accounting for the sensitivity and volume of AI interaction data.
- **Baselined high-risk access activities** should account for the normal behaviour of AI agents and their identities so anomalous activity can be detected.
- **Content filtering** should address content consumed and generated by AI systems.
- **Regular and proven backups** should include the models, configurations and retrieval data stores needed to restore AI-enabled services.
- **Cyber security awareness training** should cover approved AI use, the handling of organisational data, AI-enabled social engineering, out-of-band verification of unexpected requests and the limitations of AI-generated outputs.



Cloud considerations

AI used in cloud environments introduces additional considerations because responsibilities for services, platforms and infrastructure can be distributed between the organisation and its cloud service providers.

Organisations should obtain executive or delegate approval before enabling SaaS or platform as a service (PaaS) AI functions. High-risk SaaS or PaaS services that store sensitive data should be subject to the executive approval arrangements in *ASD Essentials for cloud*.

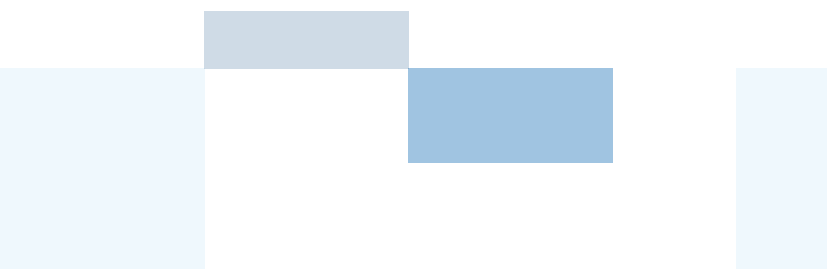
AI capabilities should be recorded in the organisation's shared responsibility register. The register should identify which AI controls the organisation operates, which capabilities the provider supplies for the organisation to configure and which outcomes the provider delivers for the organisation to verify through assurance.

AI systems supporting core organisational outcomes are likely to be scoped within critical processes and services, while AI embedded within general-purpose software as a service (SaaS) may belong within business systems. AI APIs, connectors and integrations are likely to be relevant to boundary services. AI agents and their non-person identities are relevant to identity services, while AI capabilities with access to SaaS administration or cloud control planes belong within privileged environments.

Where an organisation operates its own infrastructure to host or train AI models, it should also consider whether *ASD Essentials for enterprise IT* better reflects the responsibilities it holds for that infrastructure.

Most cloud principles apply to AI in a similar way they apply to other cloud technologies, although implementation may differ. When applying these principles:

- **Asset identification** should include AI services, agents and connectors and bring unsanctioned AI applications under organisational governance.
- **Supplier cyber security assurance** should establish where AI providers retain prompts and outputs, whether customer data is used to train models and the regions in which AI processing occurs.
- **Cyber supply chain security** should assess and approve AI applications, connectors and consent grants before they access organisational data.
- **Identity, credential and access management** should provide each AI agent with an owned non-person identity that is constrained to its intended context.
- **Least privilege access** should limit AI agents and connectors to the minimum data and permissions required and address overshared data that AI assistants may surface.
- **Secure administration** should treat AI agents and automation with access to SaaS administration or cloud control planes as privileged and record their actions.



- **Secure configuration management** should include tenant-level AI settings, such as web grounding, plugins, connectors and data access scopes, in approved configuration baselines.
- **Vulnerability management** should address weaknesses in AI applications and integrations built on platform services, including susceptibility to prompt injection.
- **Cryptographic protection** should retain customer control of the keys protecting the most sensitive data processed or stored by AI services where the service supports it.
- **Content filtering** should prevent sensitive data flowing to unsanctioned AI services and inspect content retrieved by AI systems for malicious instructions.
- **Centralised event logging** should capture AI prompts, outputs and agent actions and ensure relevant provider logging is enabled and retained for investigation.

OT considerations

AI used in OT requires additional consideration because cyber security decisions can affect safety, operational continuity and the physical processes supported by OT.

AI can help adversaries understand industrial protocols and control logic more quickly, increasing the value of exposed OT system information. OT environments can also have constraints on how quickly systems can be patched or changed, increasing reliance on compensating controls and isolation.

AI-enabled monitoring and analytics products may depend on vendor cloud services. Organisations should ensure these products do not create pathways or dependencies that impact OT isolation capability. Vital OT, and the logging and detection capabilities that support it, should remain available when systems outside the OT environment are unavailable.

Organisations should also ensure automated actions recommended or taken by systems with input from AI cannot compromise safety functions.

AI capabilities can be embedded in OT products, vendor services and supporting platforms, as well as in models and data that influence operational decisions or control functions. Organisations should map these assets according to their role in the OT environment, connectivity, criticality and potential impact on safe and continuous operations.

Models and decision-support systems that provide input to control are likely to be relevant to critical processes and services. AI tools used on engineering or operator workstations belong within privileged environments, consistent with the treatment of devices within OT as privileged.

Most OT principles apply to AI in similar ways they apply to other technologies, although implementation may differ. When applying these principles:

- **Asset identification** should include AI models, training data and AI capabilities embedded in OT products or vendor cloud services.
- **Business criticality rating identification** should record where vital OT depends on AI capabilities, including capabilities located outside the OT environment.
- **Cyber supply chain security** should verify the integrity and provenance of AI models and model updates before they are introduced into OT environments.
- **Supplier cyber security assurance** should establish whether suppliers use OT telemetry or data to train or improve models and where that data is held.
- **System exposure minimisation** should prevent OT system information, such as configurations and drawings, being disclosed to AI services outside the organisation's control and require third parties to provide equivalent protection.
- **Secure configuration management** should cover models that provide inputs to control systems, constrain outputs to safe operating limits and detect unauthorised changes. Safety systems should enforce those limits independently.
- **Network segmentation and segregation** should ensure AI platforms have no pathway capable of changing safety systems.

- **Operational technology isolation** should ensure vital OT does not depend on AI services outside the OT environment and remains safe throughout isolation.
- **Remote access to operational technology** should ensure AI-enabled vendor services that collect telemetry or provide remote support connect through the OT demilitarised zone rather than directly into OT.
- **Regular and proven backups** should include the models, configurations and data required to rebuild AI-enabled OT functions and should be held per site alongside logic and project files etc.
- **System recovery assurance** should include the ability to recover AI-enabled OT functions without supplier support.
- **Continuous cyber security improvement** should include isolation exercises that confirm AI-dependent functions fail safely and operators can sustain operations without them.
- **Centralised event logging** should capture AI recommendations, automated actions and operator overrides and remain available during isolation.
- **Cyber security awareness training** should cover the risks of sharing OT information with AI tools and when operators should override AI recommendations.

Looking forward

The application of the ASD Essentials series to AI will continue to evolve as AI capabilities, threats and assurance practices mature. While the series' core principles are intended to remain stable, AI-specific guidance will be updated through supporting publications.

Organisations are encouraged to engage with ASD on emerging AI security considerations through consultation channels, the [Information Security Manual](#) and other guidance available at cyber.gov.au.

**For more information, or to report a
cyber security incident, contact us:**

cyber.gov.au

1300 CYBER1 (1300 292 371)

Disclaimer

The material in this guide is of a general nature and should not be regarded as legal advice or relied on for assistance in any particular circumstance or emergency situation. In any important matter, you should seek appropriate independent professional advice in relation to your own circumstances.

The Commonwealth accepts no responsibility or liability for any damage, loss or expense incurred as a result of the reliance on information contained in this guide.

Copyright

© Commonwealth of Australia 2026

With the exception of the Coat of Arms, the entity's logo, third party content, and where otherwise stated, all material presented in this publication is provided under a Creative Commons Attribution 4.0 International license

<https://creativecommons.org/licenses/by/4.0/>

For the avoidance of doubt, this means this license only applies to material as set out in this document.



The details of the relevant license conditions are available on the Creative Commons website as is the full legal code for the CC BY 4.0 license

<https://creativecommons.org/licenses/by/4.0/legalcode.en>

Use of the Coat of Arms

The terms under which the Coat of Arms can be used are detailed on the Department of the Prime Minister and Cabinet website

<https://www.pmc.gov.au/resources/commonwealth-coat-arms-information-and-guidelines>

