First published: 6 July 2026

Defending against AI-enabled cyber attacks: Guidance for government, critical infrastructure and large enterprises

The Australian Signals Directorate's Australian Cyber Security Centre (ASD's ACSC) has developed this guidance to help government, critical infrastructure providers and large enterprises defend against cyber attacks that involve the use of artificial intelligence (AI). Malicious actors may use AI to discover vulnerabilities, develop exploits within hours, target many victims at once - especially for internet-facing services such as websites, exploit a chain of lower severity vulnerabilities to cause high severity impact, and rapidly progress an attack from initial compromise to data theft, extortion or disruption.

Recommended actions - immediate

Appoint a named executive to be accountable for the risks to the organisation from malicious actors using AI, and the risks to the organisation from their use of AI.

At an executive committee or board of directors meeting, increase risk appetite for internet-facing services being briefly unavailable due to rapid vulnerability remediation, more frequent patch windows, and a bad patch or mitigation.

Update vulnerability management processes to rapidly patch or mitigate vulnerabilities of all severities in internet-facing services, automatically if the risk of a bad patch or mitigation is acceptably low or can be rolled back easily.

At an executive committee or board of directors meeting, reduce risk appetite for compromise of internet-facing services that are not vendor-supported, and significantly increase the urgency of removing or replacing them.

Identify all internet-facing services daily, remove all that are not necessary, progress removing or replacing all that are not vendor-supported and implement enhanced monitoring and compensating controls in the meantime.

At an executive committee or board of directors meeting, allocate suitable and sufficient personnel and resources to address the risks from AI, including to perform the recommended actions in this publication.

Recommended actions - short-term

Grant accounts used by internet-facing services the minimum access and privileges required, and implement network segmentation and segregation between each internet-facing service, and between internet-facing services and the corporate network.

Develop, deploy, enforce and monitor secure configuration baselines for internet-facing services to disable or remove unnecessary functionality, insecure settings, and unsupported or weak communication protocols.

Verify the implementation effectiveness of patching, network segmentation and segregation, least privilege accounts, secure configuration baselines, and other technical security controls for internet-facing services.

Centrally capture and promptly analyse security logs from internet-facing services to identify cyber security incidents, using automated triage and automated response actions that are pre-approved, narrowly-scoped, and reversible.

Prepare and exercise incident response plans to rapidly remediate cyber security incidents, including restoring from backups that are recent, tested, and protected from unauthorised access, modification and deletion.

Prepare and exercise business continuity and disaster recovery plans to support resilient business operations during cyber security incidents, including testing the ability to rapidly block network traffic and reset all user passwords.

Recommended actions - medium-term

Consider using software-as-a-service cloud services for internet-facing services, from secure and transparent cloud service providers that are responsible for rapidly remediating vulnerabilities on behalf of the organisation.

Use software from suppliers with a demonstrated commitment to security and transparency, to reduce the number of vulnerabilities requiring remediation, and to ensure that patches or mitigations are provided promptly.

Use software from suppliers that securely provide adequately tested patches separate to optional feature updates, so that automatic patching can be enabled with low risk of supply chain attacks and faulty/incompatible patches.

Consider using AI, in a secure, controlled and human-supervised manner, to find and confirm vulnerabilities in code you build for internet-facing services, before release into production and periodically for code in production.

Consider using AI, in a secure, controlled and human-supervised manner, to assist with analysing security logs from internet-facing services, and performing penetration tests and vulnerability assessments of internet-facing services.

Consider which other computing devices, based on a risk assessment, would benefit from timely patching, network segmentation and segregation, least privilege accounts, secure configuration baselines, and prompt log analysis

Recommended actions - long-term

Progress to a modern defensible architecture aligned to a layered architecture with clear traceability of architectural designs to business, security and resilience objectives, using zero trust principles and Secure by Design practices.