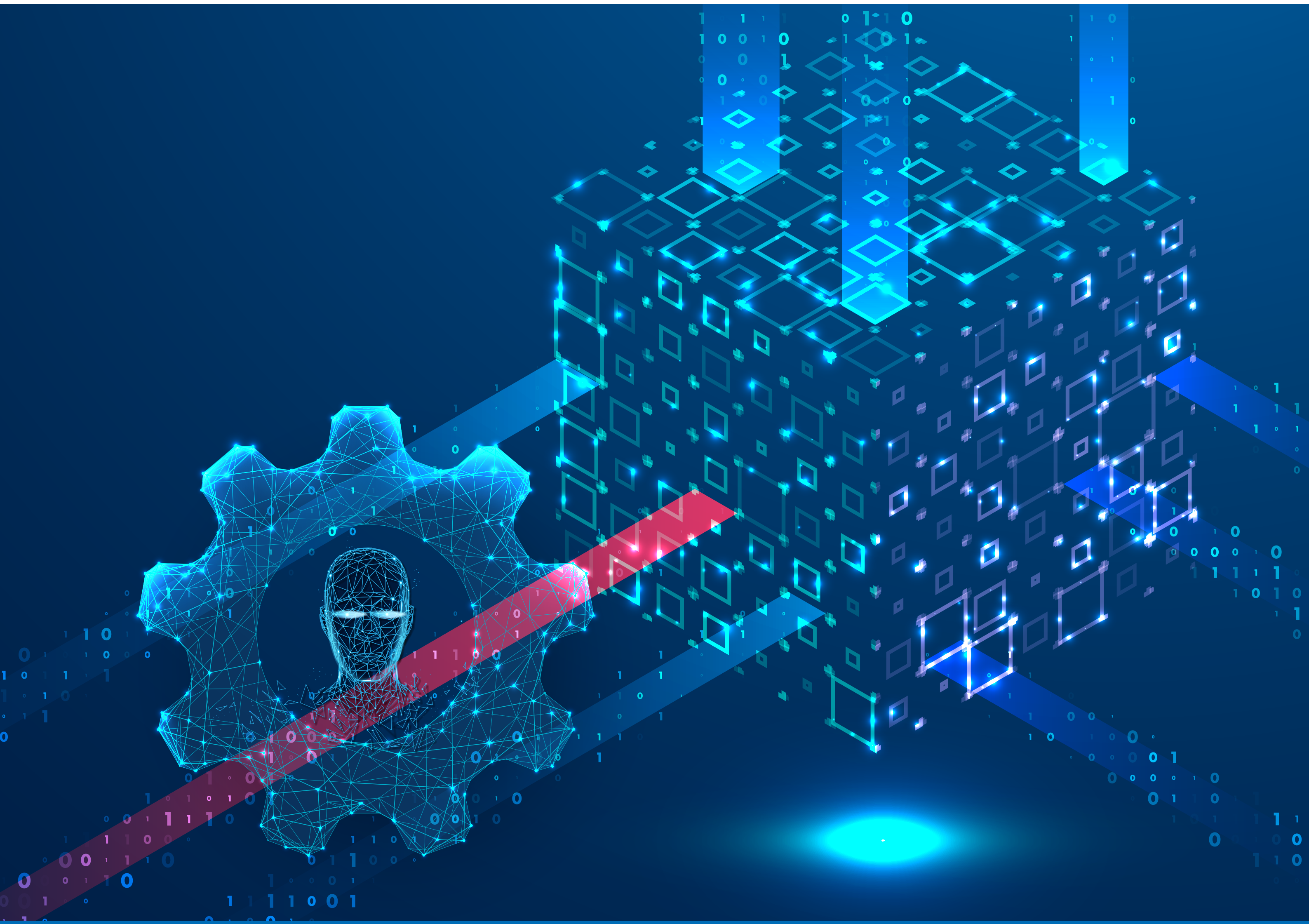
Artificial intelligence for small business

Managing cyber security risks



Australian Government

Australian Signals Directorate

ASD AUSTRALIAN SIGNALS DIRECTORATE

ACSC Australian Cyber Security Centre

CyberWardens.

Te Tira Tiaki
Government Communications
Security Bureau

National Cyber
Security Centre
NEW ZEALAND

Table of contents

Introduction

Small business owners often seek ways to stay ahead of the competition and drive growth. One technology making a big impact is artificial intelligence (AI).

More small businesses are using AI through applications, websites and enterprise systems hosted in the public cloud like OpenAI's ChatGPT, Google Gemini, Anthropic's Claude, and Microsoft Copilot. AI adoption is growing fast in Australia. Based on data from the Department of Industry, Science and Resources (DISR), this is rising every year.

Cloud-based AI gives affordable access to advanced tools without the heavy investment. They help automate tasks, provide insights and improve customer experience.

As AI becomes part of small business operations, understanding the related cyber security risks is essential. Small businesses must take proactive steps to protect data, customer privacy and business systems. Having strong cyber security practices is crucial to reducing risks in an evolving and complex emerging technology space.

This guidance – authored by the Australian Signals Directorate's Australian Cyber Security Centre (ASD's ACSC) in collaboration with the New Zealand National Cyber Security Centre (NCSC-NZ) and the Council of Small Business Organisations Australia (COSBOA) – explains the key cyber security risks of small businesses adopting cloud-based AI technologies and how to mitigate them. While traditional threats such as phishing, ransomware and insider threats are still relevant, this guide focuses on important cyber security risks related to AI.

For more resources to help improve the overall cyber security posture of your small business, visit ASD's ACSC's Small business hub at cyber.gov.au.

The Council of Small Business Organisations Australia (COSBOA) offer resources to help small businesses improve their cyber security posture and build essential cyber skills. To learn more, visit COSBOA's Cyber Wardens program at cyberwardens.com.au.

Other resources for business are available at the NCSC NZ's Own Your Online site for small businesses and organisations at ownyouronline.govt.nz.

Key cyber security risks in AI systems

Before adopting AI tools, small businesses should understand the related risks and ways to mitigate them to protect your business. These risks include:

Data leaks and privacy breaches

Reliability and manipulation of AI outputs

Supply chain vulnerabilities

Data leaks and privacy breaches

While AI tools can improve efficiency and decision-making, they may need access to sensitive data such as customer, staff and financial records to respond back to questions or deliver an outcome. This can introduce significant privacy and security risks if not properly managed. For example, uploading customer or staff details into generative AI platforms without proper anonymisation can expose sensitive or private information.

Some artificial intelligence providers may use customer submitted data to train or refine their models. This can depend on the configuration settings or the type of subscription. As a result, information entered into these platforms could potentially be reused or disclosed in unexpected contexts later. Small businesses should carefully review the configuration settings, terms and conditions, and privacy policies of any AI platform they engage with to understand how their data may be collected, stored, and used.

Due to limited resources and technical expertise, small businesses may lack strong data security and governance frameworks. This makes them vulnerable to:

- accidental data leaks through cloud-based AI tools
- unauthorised access to sensitive customer information
- potential misuse of customer data by third-party AI providers.

In early 2025, a contractor working for an Australian organisation uploaded personal information – including names, contact details and health records – of people involved with a government program into an AI system. This led to a serious data spill and is considered a notifiable data breach.

Managing risks

To reduce the risk of AI-related data leaks and privacy breaches:

- review internal data management, protection and governance practices, identify and secure sensitive and proprietary information
- review the data handling (including access and use) and privacy policies of AI vendors to ensure compliance with business requirements and relevant local data security and privacy laws
- establish an internal AI use policy or process, and clearly define what data can't be uploaded into AI platforms and systems
- train and remind staff on responsible use of AI, especially surrounding sensitive and proprietary information
- remove, anonymise or change personal details when using an AI application so it can't be used to identify or link to an individual.

Traditional cyber security practices – such as enforcing role-based access controls and using encryption (for example, AES-256) for data storage and transmission – can also strengthen overall data protection measures.

Reliability and manipulation of AI outputs

AI systems can be tricked by malicious cyber actors through prompt injection, which are malicious inputs disguised as legitimate requests designed to confuse or mislead the AI into giving sensitive, wrong or unsafe answers. AI systems can also make up information, called hallucinations, where the response sounds correct but isn't true. These risks can compromise small business decision-making, customer interactions, and operational integrity.

Whether caused by hackers or inadvertently, due to technological limitations in the AI system, unreliable outputs pose significant cyber security and business risks.

In 2025, a lawyer used AI to prepare a court document, but it generated false legal cases that the lawyer didn't verify before submitting to the court. After the court discovered this, the lawyer was barred from operating and owning a law practice.

Managing risks

To reduce the risk of unreliable or manipulated AI outputs:

- train and remind staff to verify AI outputs for inaccurate answers, biased language, unethical or irrelevant responses
- have a human involved in the decision-making process especially in high-stakes or sensitive operations
- use trusted and regularly updated AI models from vendors that prioritise security
- monitor the behaviour of AI systems for any unusual patterns or anomalies
- review any AI-integrated processes regularly (either in real time for critical tasks or during scheduled checks)

Supply chain vulnerabilities

Small businesses may use AI Software as a Service (SaaS) applications, such as online chatbots or customer relationship management systems. These SaaS applications typically rely on third-party providers to manage the underlying infrastructure, AI models, data handling and operations. This creates a supply chain dependency risk for small businesses. It means any vulnerabilities in the third-party provider's systems, infrastructure, data sources or AI models can indirectly affect the business.

In April 2024, an education business discovered a data breach affecting tens of thousands of students, parents, guardians and staff. A hacker exploited a vulnerability to access their system. A security patch was available but wasn't installed. The exposed system contained highly sensitive information. It included fee records, childhood education and care details, children's status, welfare requests and medical certificates.

Managing risks

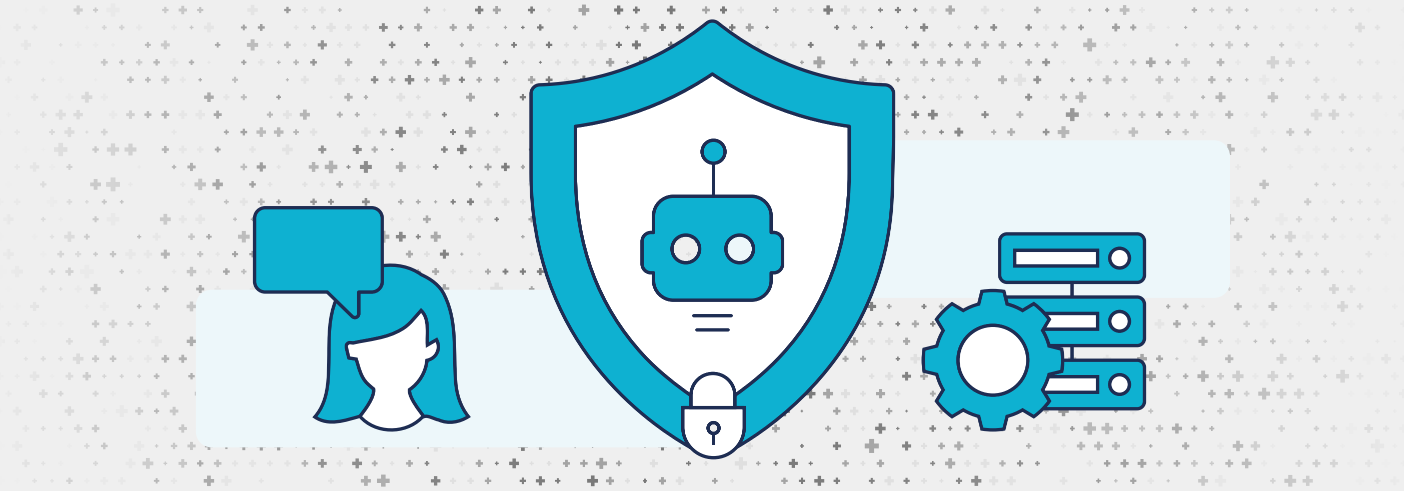
To reduce the risk of supply chain vulnerability:

- evaluate the AI vendor's reputation and commitment to security, including its use of third-party tools or services
- review the AI vendor's terms and conditions related to data ownership, protection, usage and storage
- understand the AI vendor's cyber security incident notification process and incident response mechanisms.

Implementation example– Secure deployment of AI chatbots

Businesses are increasingly deploying customer-facing AI chatbots to enhance service delivery and streamline support operations. While traditional IT security controls remain essential, small businesses should also implement chatbot-specific safeguards to protect user data and ensure responsible use.

Secure Chatbot



Key safeguards

Limit data collection

Collect only what is essential for the chatbot's function (e.g., service-related queries).

Human-in-the-loop oversight for high-risk use cases

In scenarios involving legal, medical or financial advice, ensure a qualified person reviews chatbot responses before taking any action.

Vendor due diligence

Select chatbot platforms that:

- are reputable, trustworthy and committed to security
- have transparent data handling and protection policies
- comply with relevant privacy regulations.

Business contracts should include clear clauses on data protection, breach notification procedures, and service uptime guarantees.

AI cyber security checklist for small businesses

□ I understand the benefits and risks of integrating AI into my business.

□ I know what business information can be safely shared with the AI tool.

□ I have verified what data the AI tool collects and where it is stored.

□ I know who owns the data: my business or the AI vendor.

□ I have confirmed whether my business' data will be used to train AI models.

□ I know where and how to fact-check the AI system's outputs.

□ I have provided AI security-related training and advice to my staff members e.g. Cyber Wardens Level Two – Safe AI for Small Business, and Cyber Wardens Level Three – Cyber Fit for the Supply Chain, at cyberwardens.com.au/courses/

□ I have verified that the AI vendor is committed to security, for example by using a recognised cyber security compliance framework (e.g., ISO 27001 for Information Security Management Systems, NIST AI Risk Management Framework).

□ I know the process for handling a cyber security incident related to the AI application or tool.

Contact

For issues with AI applications and tools, contact the AI vendor in the first instance.

Report a cyber security incident or vulnerability to:

- ASD's ACSC at cyber.gov.au/report
- NCSC-NZ at ncsc.govt.nz/report

Report a Notifiable data breach to:

- Australia's Office of the Australian Information Commissioner at oaic.gov.au
- New Zealand's Office of the Privacy Commissioner at privacy.org.nz

More information

- Small business hub at cyber.gov.au
- Artificial intelligence at cyber.gov.au
- Mitigating cyber security incidents at cyber.gov.au
- Cloud shared responsibility model: Guidance for individuals and small and medium businesses at cyber.gov.au
- Ten things to know about data security at cyber.gov.au
- COSBOA's Cyber Wardens 'Level Two – Safe AI for small business' online course at cyberwardens.com.au
- DISR's Guidance for AI Adoption at industry.gov.au
- DISR's AI adoption in Australian businesses for 2025 Q1 at industry.gov.au
- New Zealand's Ministry of Business, Innovation and Employment's Responsible Artificial Intelligence guidance for businesses at mbie.govt.nz

Glossary of technical terms

Term	Definition	Example
Artificial Intelligence (AI)	Computer systems that perform tasks requiring human-like intelligence, such as decision-making and language understanding.	Chatbots, predictive analytics tools.
Generative AI	AI that creates new content (text, images, audio) based on learned patterns from existing data.	AI systems generating marketing copy.
Cloud-based AI	AI services hosted on remote servers, accessible via the internet	OpenAI's ChatGPT, Google Gemini, Anthropic's Claude
Data Leak	Unauthorised exposure of sensitive information due to poor handling or security gaps.	Uploading customer data into an AI tool without anonymisation.
Privacy Breach	Violation of data protection laws or policies where personal information is accessed or disclosed without consent.	Sharing health records via an unsecured AI platform.
Prompt Injection	Malicious instructions embedded in user inputs to manipulate AI into producing harmful outputs.	A hacker tricks an AI chatbot into revealing confidential data.
Hallucination (AI)	AI generates plausible but factually incorrect or fabricated information.	AI inventing fake legal cases in a court document.
Supply Chain Vulnerability	Risks from reliance on third-party vendors for AI services; their weaknesses can affect your business.	SaaS chatbot vendor suffers a data breach.
SaaS (Software as a Service)	Software delivered online by a provider, often subscription-based.	OpenAI's ChatGPT, Google Gemini, Anthropic's Claude
Role-Based Access Control (RBAC)	Restricts system access based on user roles to protect sensitive data.	Only managers can access financial AI reports.
AES-256 Encryption	A strong encryption standard for securing data during storage and transmission.	Encrypting customer data before uploading to AI tools.
Human-in-the-Loop	Human oversight in AI decision-making, especially for high-risk tasks.	A lawyer reviews AI-generated legal advice before submission.
ISO 27001	International standard for managing information security.	AI vendor certified under ISO 27001.
NIST AI Risk Management Framework	A framework for managing AI-related risks developed by NIST.	Vendor compliance with NIST AI risk guidelines.

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

Use of the Coat of Arms

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

cyber.gov.au | 1300 CYBER1 (1300 292 371)

ASD AUSTRALIAN SIGNALS DIRECTORATE

ACSC Australian **Cyber Security** Centre