Australian Government
Australian Signals Directorate

ASD AUSTRALIAN SIGNALS DIRECTORATE
ACSC Australian Cyber Security Centre

An introduction to Artificial Intelligence

Content Complexity
MODERATE

Artificial Intelligence (AI) is an emerging technology that will play an increasingly influential role in the everyday life of Australians. In response to the rising interest and discussion around AI, highlighted globally at the UK's 2023 AI Safety Summit, the Australian Signals Directorate (ASD) is expanding its AI guidance to help Australian individuals and organisations engage with AI systems in a secure way. The purpose of this publication is to provide readers with an understanding of what AI is and how it may impact the digital systems and services they use. In addition to this publication, readers are also encouraged to read ASD's Ethical AI framework, which outlines ethical principles for the implementation and use of AI.

What is AI?

AI is the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. Modern AI is usually built using machine learning algorithms. These algorithms find complex patterns in data, which can be used to form rules.

For example, voice assistants, such as Siri, use natural language processing and machine learning to understand a user's voice and correctly perform tasks like setting alarms or searching on an internet browser. For a more detailed explanation of what AI is, refer to ASD's 'Convoluted Layers: An Artificial Intelligence Primer'.

There are already plenty of examples of AI having a positive impact on Australia. The CSIRO's Spark system is an excellent example. CSIRO developed Spark, an AI-enabled system, which is capable of designing custom fire propagation models to predict bushfire spread. It will empower decision-making, planning, response and research processes by providing a realistic simulation of how a bushfire could progress and assist in improving crisis preparedness and community safety.

Considerations for engaging with AI

Like all emerging technologies, AI presents both opportunities and risks. To take advantage of the benefits of AI securely, users and organisations should take some time to understand what risks apply to them and how those risks can be mitigated. Some common AI related risks are:

Data poisoning

- An AI model is built by training it on a large amount of data. For example, an AI that can discern objects within a photo would be trained on a large number of pictures. The quality of training data affects the performance of the trained AI. If an attacker alters this training data, they could influence the AI to make poor or incorrect decisions. This is known as data poisoning.

Adversarial example

- Once the AI is in operation, attackers may be able to provide the AI model with specially crafted inputs or prompts to force the AI to make a mistake. This is commonly known as an adversarial example attack.

Scams empowered by generative AI

- Generative AI will enable attackers to create convincing scam messages, for example, with fake voice and video clips of public figures.

Identifying vulnerabilities

- AI systems will be used to automate data collection and analysis, potentially reducing the effort and technical skill required by attackers to find vulnerabilities to exploit. This could allow for quicker and more effective target selection.

Privacy concerns

- Often data collected from you is anonymised to protect your privacy. When anonymised correctly, it should require a substantial effort to re-identify an individual. With the emergence of AI, there are concerns that, by leveraging AI's capability to work across large data sets, attackers may be able to re-identify individuals in pools of anonymised data.

For more technical information on AI-specific threats visit MITRE ATLAS.

What mitigation considerations exist?

There are steps that government, industry and end users can take to engage with AI securely. Answering the questions posed below will help individuals and organisations understand how they can benefit from using AI while still mitigating its risks.

For individuals

When using AI technologies, particularly generative AI, ASD recommends applying the same basic security principles as they would when using any online tool. Ask yourself:

- Does this system have a good reputation?
- Do I need to share this information?
- How will the system use my information? What does their privacy policy say?
- What can I do to ensure the output is accurate and appropriate for use?

For further guidance on how individuals can stay secure online, refer to ASD's Personal Security Guides.

For organisations

Organisations considering using or developing AI systems should approach their decision-making as they would with any other IT technology, and evaluate the specific benefits, risks, and consequences for their organisation. Some of the key questions to ask are:

- Has your organisation implemented ASD's Essential Eight?
- ASD's Essential Eight framework provides mitigations for organisations for various cyber threats, including AI-enabled threats.
- Does your organisation understand the AI system, including the risks it poses?
- Is the system secure-by-design?
- Have you assessed the system's supply chain risks? See ASD's guidance on Cyber Supply Chain Risk Management and our supply chain exercises on Exercise in a Box.
- How will the system affect your organisation's privacy and data protection obligations? See ASD's Guidelines for Database Systems.
- Who is accountable for oversight or if something goes wrong with the system? See ASD's Guidelines for Cyber Security Incidents.

What's next?

As an emerging technology, developments in AI are rapid. ASD will continue to engage with AI technologies and publish guidance to inform and protect Australians.

ASD will release additional guidance on cyber security mitigation strategies for engaging with AI systems for individuals and organisations, in cooperation with international partners. Additionally the UK's National Cyber Security Centre is planning to release secure AI system development guidelines, which has been co-sealed by ASD and other international partners.

Further information

For further guidance on engaging with AI securely and responsibly, consider the following resources:

- For more information on how to approach using AI systems ethically, visit ASD's Ethical AI framework.
- The Australian Government has also developed 8 AI Ethics Principles. This is a voluntary framework for organisations (in addition to legislative requirements) to commit to ethical AI practices.
- For more information on best practice advice for Government agencies implementing or considering the use of AI/ML, The Digital Transformation Agency has released interim guidance for generative AI.