Australian Government
Australian Signals Directorate

ASD AUSTRALIAN SIGNALS DIRECTORATE
ACSC Australian Cyber Security Centre

An Introduction to Artificial Intelligence

First published: November 2023

Introduction

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, the Australian Signals Directorate (ASD) is expanding its AI guidance to help 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.

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 such as setting alarms or searching the internet. For a more detailed explanation of what AI is, refer to ASD's *Convoluted Layers: An Artificial Intelligence Primer* publication.

There are already plenty of examples of AI having a positive impact on Australia. The Spark system, developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), is an excellent example. Spark, an AI-enabled system, is capable of designing custom fire propagation models to predict bushfire spread. It empowers decision-making, planning, response and research processes by providing a realistic simulation of how a bushfire could progress while assisting 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, individuals 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 outlined below.

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. As the quality of training data affects the performance of the AI model, a malicious actor that alters this training data could influence the AI to make poor or incorrect decisions.

Adversarial inputs

Once an AI system is in operation, malicious actors may be able to provide it with specially crafted inputs/prompts to force it to make a mistake, such as by generating sensitive or harmful content.

Abuse of generative AI

Generative AI can allow malicious actors to easily create convincing scams, disinformation or abusive material. For example, by generating fake voice and/or video clips of individuals to influence public opinion or to harass individuals.

Identifying vulnerabilities

AI systems can be used to automate data collection and analysis, potentially reducing the effort and skill required by malicious actors to find vulnerabilities to exploit in applications, thereby allowing for quicker and more effective target selection.

Privacy concerns

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

What mitigations exist?

There are steps that individuals and organisations can take to engage with AI securely. Answering the questions below can help individuals and organisations understand how they can use AI securely while mitigating some of its risks.

For individuals

When using AI systems, particularly generative AI, ASD recommends applying the same basic security principles as when using any online service.

Individuals should ask themselves:

- Does this AI system have a good reputation?
- Do I need to share this information with the AI system?
- How will the AI system use my information? What does its privacy policy say?
- What can I do to ensure the output of the AI system 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 information technology and evaluate the specific benefits, risks and consequences for their organisation.

Organisations should ask themselves:

- Does our organisation understand the AI system, including the risks it poses?
- Is the AI system secure-by-design?
- Have we identified cyber supply chain risks associated with the AI system?
- How will the AI system affect our organisation's privacy and data protection obligations?
- Who is accountable for oversight and/or if something goes wrong with the AI system?

Further information

The *Information Security Manual* is a cyber security framework that organisations can apply to protect their systems and data from cyber threats. The advice in the *Strategies to Mitigate Cyber Security Incidents*, along with its Essential Eight, complements this framework.

For more information on engaging with AI securely and responsibly, consider *Australia's Artificial Intelligence Ethics Framework*. This is a voluntary framework produced by the Department of Industry, Science and Resources allowing for organisations to commit to ethical AI practices.

For more information on best practices when implementing or considering the use of AI, the Digital Transformation Agency has released *Interim guidance on government use of public generative AI tools*.

Contact details

If you have any questions regarding this guidance you can write to us or call us on 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.

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

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

Australian Government

Australian Signals Directorate