Australian Government
Department of Home Affairs

Department of Home Affairs submission to the Inquiry into Adopting Artificial Intelligence (AI)

Select Committee on Adopting Artificial Intelligence

10 May 2024

Table of Contents

1. Overview

The Department of Home Affairs (Home Affairs) welcomes the opportunity to provide a submission to the Select Committee on Adopting AI. Home Affairs is the lead on national security policy for critical technologies, including AI. Home Affairs is also responsible for central coordination, and strategy and policy leadership in relation to cyber and critical infrastructure, immigration, border security, counter-terrorism, the protection of our sovereignty and citizenship and social cohesion. Relevantly, Home Affairs assesses the risks and opportunities presented by critical and emerging technologies within this remit.

AI has already been extensively adopted across the Australian economy and its development and use will likely accelerate. AI will have a transformative impact on the economy, including in the health; manufacturing; retail and finance; mining and agriculture; and education sectors.

However, without appropriate guardrails, AI has the potential to exacerbate existing national security risks, with potential for new and unknown risks into the future. Governments, industry and civil society in Australia and internationally have already commenced efforts to mitigate many of these risks, including guidance, ethical standards and regulatory reform. This submission provides an overview of the use cases of AI tools in Government and the national security risks posed by the adoption of AI in Australia if not appropriately managed. It does not examine the use of AI in military or defence domains.

2. Trends and opportunities of AI

Products and services that utilise AI are already broadly in use across the Australian economy. These can be broadly summarised as:

- Automated or assisted decision making: A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. These systems are designed to operate with varying levels of autonomy.
- Content curation or recommender systems: systems that prioritise content or make personalised content suggestions to users of online services. Algorithmic content moderation typically involves varying applications of either matching or prediction models, which contribute to decisions and governance outcomes on specific content or accounts
- Generative AI: sophisticated machine learning algorithms used to predict an output – such as images or words – based on an input, such as a sequence of words. These models typically recognise patterns in data and produce sophisticated answers based on those patterns.

The development of these products and services is rapidly accelerating. Significant investment by industry and governments is driving unprecedented advancements in AI. Many experts predict that there will soon be AI systems which will be able to outperform humans across a range of tasks and that artificial general intelligence (AGI) – AI which, while not sentient will have human-like intelligence and the ability to self-direct learning – will be realised within this decade.

This development will bring significant opportunities across the economy to increase productivity, service delivery and research and development. As technologies mature, government will increase automation and machine learning into core business roles such as risk, strategy, resource allocation and delivery. AI is an attractive and scalable solution to improving our service offerings to the public with increased sophistication. Home Affairs has deployed AI for predictive analytics, network graphing, natural language processes, robotic process automation, computer vision and text and image-based generation.

As AI development matures, Home Affairs continues to monitor opportunities to leverage 'rules as code' and 'digital twin' capabilities to enable better policy decision making. Trusted in-house generative AI capabilities may also be able to support policy development; analyse open source information, protect Department systems from cyber threats and improve accessibility and use of corporate information on internal departmental systems. For instance, AI-enabled technologies incorporated into the automated decision making SmartGate systems at airports provide efficiencies immediately experienced by the public. However, deployment of these use cases will need to be carefully considered to ensure they are fit for purpose and do not raise unacceptable risks.

In addition to supporting the development of AI tools, Home Affairs continues to provide national security policy advice to support the development of a whole of government AI Policy and National AI Assurance Framework, being developed by the Digital Transformation Agency and the Department of Industry Science and Resources. Home Affairs' input to these work programs is informed by lessons learned in the development and use of its own AI tools. The development of these policies and frameworks will support the safe and responsible adoption of AI across the economy, with government leading by example.

3. Risks and harms of AI

This submission highlights the national security risks of AI, in line with Home Affairs remit as Australia's lead national security policy agency. The following provides a high level assessment of the types of risks, developed following desktop research, engagement with domestic and international experts and with advice from across Government including the National Intelligence Community.

It is important to distinguish between the safety risks posed by AI (such as the social or psychological harms to individuals and communities), the security of AI models themselves (for example, AI being hacked, poisoned or altered without authorisation) and the risks AI poses to Australia's national security (the broader landscape in which and AI model operates). Home Affairs' Portfolio responsibilities extend to each of these harms.

3.1. Foreign interference

As AI in increasingly adopted and embedded across our digital ecosystem, it will provide enhanced and potentially transformative approaches to the conduct of espionage and foreign interference, Australia's principle national security threats.

AI advancements are augmenting traditional methods of foreign interference and will enable tailored foreign information and manipulation and interference (FIMI) activities to audiences, including vulnerable Australians, diaspora communities, businesses and against our democratic institutions. With the assistance of AI, FIMI can be created and disseminated at unprecedented speed and scale, in multiple languages; and often at a low cost. Foreign governments could use AI to create coordinated and inauthentic influence campaigns that are designed to foster widespread misinformation, incite protests, exacerbate cultural divides and weaken social cohesion, covertly promote foreign government content, target journalists or dissidents and influence the views of Australians on key issues. AI could lower the barriers for non-sophisticated actors to engage in manipulation and increases the threat of the public being targeted.

Key publically observed tactics employed by foreign actors include using AI generated voice, image, video and text to convey messages on social media platforms such as X (formerly Twitter) and Meta. Several publicly available reports on FIMI activities on Australian social media indicate that threat actors could have used AI as part of their toolkits. The Australian Strategic Policy Institute have released a number of reports that have identified the use and possible use of AI in FIMI activities in Australia. Internationally, Australia's partners have observed the clandestine use of AI in a number of influence campaigns. This use obscures the link back to the foreign government, which makes it difficult for a general audience to discern the origin of the content.

These harms can have national security ramifications, erode public trust in AI and trust in our democratic institutions and damage social cohesion. AI also has potential to be used to combat these risks, particularly in cybersecurity and the information environment. Misinformation will remain a consistent threat to democracy and trust in government institutions enabled by all AI technology domains, not only generative AI. There is a need to target and prioritise resources. Home Affairs will continue to work through the Counter Foreign Interference Coordination Centre (CFICC) to engage other government stakeholders on policy responses to mitigate the foreign interference risks posed by AI

3.2. Data security and integrity

AI will amplify the amount and type of data being collected as commercial incentives drive AI developers to collect more data to support the development of more mature language models. Hostile actors will be motivated to seek and aggregate data they steal or obtain from data breaches to enhance models they develop. AI capabilities trained on personal and sensitive data have potential to accelerate adversaries' efforts to erode our technological advantage and to target our networks, systems and people.

AI also presents the risk of minority groups and small communities being misrepresented in AI models. Under representation in underlying training datasets could result in disparities and unconscious systematic bias between the quality of services, or excessive scrutiny from authorities between majority and minority groups. Tailored guidance or requirements for the integrity of data sets used in AI-supported decision making may be appropriate to mitigate this risk.

3.3. Cyber security

The adoption of AI across the Australian economy will significantly increase the cyber threat surface. Access to AI will make it easier for malign actors and cybercriminals to undertake malicious cyber activity. Whilst the emergence of AI models designed to identify and mitigate against cyber vulnerabilities will improve cyber resilience, it will also facilitate identification of cyber vulnerabilities.

Noting the ease of deployment of AI, non-technical actors will more simply be able to generate malicious code which will increase the risk of cyber security incidents including large scale data breaches and ransomware attacks. Tools designed for malicious activities such as the creation of phishing emails at scale are already available for purchase on the Dark Web. Malicious cyber actors may easily deploy and develop increasingly convincing and personable phishing and scam campaigns. The growing demand for subscription-based criminal models, commonly referred to as 'crime-as-a-service,' will be further enhanced in accessibility and efficiency through the integration of AI. AI may also support increased outsourcing of malicious cyber activity from nation states to criminal syndicates, particularly should more sophisticated capabilities become readily accessible.

3.4. Dissimination of harmful content

Access, generation and dissemination of malicious content is made simpler with AI. Future generations of AI may help close the knowledge gap for non – sophisticated actors in the planning of complex attacks, including for the development of bioweapons, synthetic drugs or explosive devices. AI will lower the barriers for non-sophisticated actors, equipping them with previously unattainable capabilities such as access to instructions to develop bioweapons, synthetic drugs or explosive devices. AI models can be created or repurposed to enable malicious activity. This may include accessing and generating illegal content faster and more easily than previously possible, and jailbreaking public AI models to produce content that breaches safety controls.

3.5. Democracy and trust in institutions

International evidence and assessment of the practical impacts of AI on democracy and trust in institutions is still foundational. This is, in part, due to the constantly evolving nature of AI and a paucity of research, including research specific to the Australian context. What is clear, however, is that rapid technological change such as AI will challenge historic strengths for Australian democracy including strong institutions, information integrity and social inclusion, with erosion of trust cutting across all three.

AI capabilities can be used to undermine trust in democratic institutions and encourage discrimination and social division within open democracies like Australia. There is a risk that foreign interference could have implications for Australia's electoral integrity, and reduce Australians' trust in their democratic institutions. Political ads and campaigns that utilise AI to generate false content could spread narratives that may sway public perception of candidates and their positions on certain issues. AI capabilities can also be used to interfere in the voting process, such as circulating realistic disinformation about where and how to vote, or to discourage voters from showing up to polling locations. If left unchecked, the use of AI could gain prominence in political and electoral advertising and have a great impact on the voter's right to make informed decisions. This potential prominence and its impact is more concerning when foreign interference plays a role with an aim to interfere in Australia's democratic institutions.

Generative AI could provide malicious actors with the ability to rapidly produce immense volumes of content at low cost and without regard for accuracy, which could pose a major threat to democratic representation, accountability and trust, particularly during election periods:[i]

- **Threat to representation:** Generative AI allows anyone – from passionate citizens to malicious actors – to create unique letters, emails and social media posts that skew elected officials' perceptions of constituent sentiment, undermining genuine representation.
- **Threat to accountability:** AI-generated information operations and smear campaigns could unfairly influence perceptions of elected representatives, undermining elections as a mechanism of accountability since the basis for people's vote is factually dubious.
- **Threat to trust:** A proliferation of false and misleading information may make people sceptical of the entire information ecosystem, in turn eroding the trust that fuels civic engagement, political participation and confidence in institutions, and potentially exacerbating polarisation.

There are emerging examples of innovative responses to these threats. For example, industry collaborations such as the Content Authenticity Initiative are seeking to add digital watermarks or 'fingerprints' to identify digital provenance, and help audiences decide what they can believe. It is hoped that the same technology which was so disruptive to trust could help restore it.

The opportunities and risks associated with the development and uptake of AI have clear implications for Australian democratic resilience. In October 2023, Home Affairs supported an academic roundtable convened by the Human Technology Institute at UTS. The participants identified seven dimensions along which AI impacts, and has the potential to harm, core democratic principles:

- **Connectedness — Polarisation**

  A more connected society has greater democratic resilience than a polarised one. Current AI technologies are have the potential to push democracies toward greater polarisation.

- **Transparency — Opacity**

  Transparency is a key pillar of democracy. A large proportion of current AI technologies are characterised by opaqueness including how and what data is collected, how decisions are made, who is responsible for those decisions and where the technology is leading democracies.

- **Decentralisation — Consolidated Control**

  Democracy, by definition, decentralises power to the public by way of elections, the rule of law, a free press and other principles and processes. The current trajectory of AI technology is leading to greater control by small number of players.

- **Democratising Voices — Narrowing Voices**

  Liberal democracy defends and benefits from a plurality of voices, whereas AI technology has the potential to be exclusionary or biased against certain voices in its design and application.

- **Truth and Deliberation — Deception**

  Democracy is founded on the ability to have an equal stake in the future, which requires access to factual information in order to make informed decisions. The current trajectory of AI is increasing the prevalence of deceptive material in the information environment.

- **Public Good — Private Gain**

  Liberal democracy as a system has at its foundation the aim of providing good to the broadest element of the public without undermining the rights of the minority. AI technologies are often seen to further private gain rather than the public good.

- **Information Engagement — Information Transmission**

  Democracy is based on deliberation of ideas, a discussion of what the public values and what it wants. Current AI technologies are predominantly about the supply and transmission of information rather than the deliberation of it.

4. Fostering responsible AI adoption across society, industry and infrastructure

Appropriate national security risk mitigation settings act as an important enabler to promote the safe and responsible adoption of AI across the economy and unlock economic opportunities. Risk mitigations will increase trust in AI technology, accelerate safe and responsible adoption of AI across the economy and reduce the costs of harms. Where trade-offs are necessary, a consistent whole-of-government response is required to assess the appropriate balance required when implementing risk mitigations.

International partners offer different models to consider:

- The US' *Executive Order on Artificial Intelligence* mandates that AI responds to national security vulnerabilities, and directs the development of a national security memorandum that directs further actions on AI and security. Led by the Department of Homeland Security and Chief AI Officer, there is clear policy leadership to ensure AI is deployed with national security considerations, including in supply chain security and protecting critical infrastructure. This is in addition to the establishment of an AI Security Centre which will lead the development of best practice, evaluation methodology and risk frameworks with the aim of promoting the secure adoption of new AI capabilities across the national security landscape.

- Canada's proposed Artificial Intelligence and Data Act (AIDA) constitutes a compliance regime intended to manage high-impact AI technologies, characterised as those which may cause individual or collective harms or propagate biases. AIDA would establish an AI and Data Commissioner to monitor compliance with a number of regulatory requirements: Human oversight; transparency; fairness; safety; accountability; and robustness. These requirements would apply to any business based or operating in Canada which: designs or develops AI, makes AI available for use or manages the operations of an AI system. Enforcement measures for non-compliant entities include monetary penalties or the prosecution of regulatory or true criminal offences. AIDA would also provide ministerial powers to proscribe, block or remove AI systems where an imminent threat was identified.
- The EU's *Artificial Intelligence Act* (the AI Act) proposes a horizontal legislative instrument. The instrument would be applicable to any AI system developed, marketed or accessed within the EU. Based on a technology-neutral definition, the AI Act will prescribe various requirements, obligations or prohibitions based on the assessed risk of AI systems. This is represented as a risk pyramid whereby: AI systems which present an unacceptable risk are prohibited; those AI systems which constitute a high risk are highly regulated; systems assessed as being of limited risk have requirements around transparency; and, low or minimal risk AI systems have no obligations placed upon them.

Domestically, achieving the right balance will require collaboration across government, industry and civil society that spans technical, policy and ethical considerations. Responsibility for mitigating risk and promoting safe and secure use of AI should be equitably distributed across the economy.

Developers of AI play an important role in this ecosystem. It is critical that the safety of users and security of systems is considered at the outset, rather than being retrofitted once harms occur. The industry-led development and adoption of internationally aligned standards for secure by design should be encouraged to achieve more consistent security baselines for AI products and services.

Businesses and individuals using AI systems also hold responsibility to ensure AI is used safely without creating risks for consumers or the broader community. Understanding the security implications of AI models and systems, particularly in high risk settings such as critical infrastructure, will enable the secure use of AI within Australia.

Government intervention is necessary where the risks of AI development and use are considered unacceptable. Market incentives do not currently promote transparency regarding AI algorithms and do not provide users with recourse to influence development to mitigate harms. Some initiatives that Government could lead to better support the market include:

- **Developing and implementing ethical frameworks and guidelines**: AI should be guided by ethical frameworks and guidelines that define and operationalise the core values and principles that should inform the design, development, and use of AI, such as fairness, accountability, transparency, privacy, and human oversight. Australia has already developed and adopted the Australian AI Ethics Framework, and should continue to implement and update it, and to engage and align with other national and international ethical frameworks and guidelines for AI.
- **Establishing and enforcing legal and regulatory standards**: AI should be subject to legal and regulatory standards that ensure the compliance and liability of the actors involved in the AI value chain, and the protection and promotion of the rights and interests of the affected stakeholders. Australia should continue to apply laws and regulations that are relevant to the use and governance of AI, and review and update them. Furthermore, Australia should develop and adopt new policies that are specific and appropriate to the AI domain.
- **Building and strengthening capacities and competencies**: AI should be supported by capacities and competencies that enable the responsible and beneficial use and governance of AI, such as technical, ethical, and social skills, and education, awareness, and empowerment. Australia has already invested and supported various initiatives and programs that aim to build and strengthen the AI capacities and competencies of its citizens, this should continue and expand into the future.

- **Fostering and sustaining a culture of accountability and reliability**: Australia should aspire to create a transparent and diverse AI ecosystem and community. To this end, it has already implemented and supported various platforms and mechanisms, such as the Australian AI Ethics Framework led by DISR.

Within its own Portfolio, Home Affairs will continue to deliver support to critical infrastructure owners and operators to better understand and deploy AI. The Cyber and Infrastructure Security Centre (CISC) has a suite of existing measures which can be applied to uplift the resilience of critical infrastructure to AI enabled threats. These measures are enabled by four separate but overlapping legislative schemes which are administered by Home Affairs:

- The *Security of Critical Infrastructure Act 2018* (SOCI Act) implements a series of preventative and responsive measures designed to provide a common baseline of security across all critical infrastructure sectors and support a collective uplift in the security and resilience of critical infrastructure.
- The *Aviation Transport Security Act 2004* (ATSA) and *Maritime Transport and Offshore Facilities Security Act 2003* (MTOFSA) protect Australia's civil aviation and maritime infrastructure, and offshore oil and gas facilities, from acts of unlawful interference (primarily terrorism).
- Part 14 of the *Telecommunications Act 1997* includes obligations for telecommunications carriers for carriage service providers to protect their networks from unauthorised access and interference. These entities are also obliged to notify the Government of proposed changes that may impact the security of their networks

Home Affairs will work closely with government, industry and international partners in developing policy responses that can adapt to frequent technological changes, and embedding human oversight. Potential policy responses to the national security risks of AI will benefit from the flexibility of not being tied to particular technical specifications. With unknown future technological developments that increase the capabilities of AI and reduce levels of human oversight and intervention, there is potential for new, unknown national security threats to emerge.