Australian Government
Defence

Policy Settings for Responsible Use of Artificial Intelligence in Defence

Responsible use of AI at all stages of the technology lifecycle

Acknowledgement of Country

Defence acknowledges the Traditional Custodians of the lands, seas and air in which we live, work and train. We pay our respects to their Elders past and present. We also pay our respects to the Aboriginal and Torres Strait Islander men and women who have contributed to the defence of Australia in times of peace and war.

Contents

Introduction

The strategic and operational environments in which the Australian Defence Force (ADF) and the Department of Defence ('Defence') operate are becoming increasingly challenging. The 2024 National Defence Strategy recognised that increasing strategic competition is a primary feature of Australia's security environment.

Artificial intelligence (AI), and other new and emerging technologies, are becoming an integral feature of this strategic competition.

As a general purpose technology, AI is permeating government and commercial institutions, infrastructure, products and services. AI has the potential for immense positive benefit across all sectors of our society and economy. In the national defence context, AI has particular potential for driving greater accuracy, efficiency, speed and safety in defence functions and ADF operations. Public trust in the adoption of AI in defence activities is critical to delivering that potential.

Defence personnel must address both AI's opportunities and risks as the technology evolves. Harnessing the benefits of AI for national defence will require robust governance to assure responsible use and to identify, assess and manage risks.

These policy settings state Defence's commitments to lawfulness, adherence to values-based principles, and proportionate controls in its responsible use of AI. As the technology evolves, Defence will make adjustments to its governance and controls to ensure its use of AI remains responsible and compliant with Australia's domestic law and international legal obligations.

Purpose and scope

These policy settings support the responsible use of AI in Defence, including to deliver the National Defence Strategy. Defence will use AI responsibly through an informed, risk-based approach, which preserves individual accountability for AI-enabled decisions and outcomes and maintains compliance with Australia's domestic law and international legal obligations.

> These policy settings implement Australia's commitments under the Responsible AI in the Military Domain's 2024 Blueprint for Action and 2023 Call to Action, and the 2023 Political Declaration on the Responsible Use of Military AI and Autonomy. These settings will be updated as AI technology, applications and Government policy evolve.

Harnessing AI for the Defence mission

Through faster and better informed decisions, AI technologies offer opportunities to achieve decision advantage in civilian and military functions. AI, applied appropriately in concert with human judgement, offers the ability to provide asymmetric advantage across joint warfighting and enabling functions.

Defence is investigating opportunities to integrate AI in support of ADF and civilian personnel, commanders and decision-makers, such as to:

- make sense of increasingly vast volumes of data to improve the quality and timeliness of strategic, operational and tactical decision making,
- improve productivity and accuracy by taking on dangerous, repetitive or expensive tasks (reducing the risk to personnel, and freeing them to focus on more productive activities),
- protect Defence systems and platforms from evolving cyber threats,
- optimise logistics and maintenance in support of military operations ,
- enhance force protection through time-sensitive processing of threat data, and
- enable trusted autonomy to improve the effectiveness, accuracy, duration and safety of military effects, including options to reduce risk to personnel.

Application

These settings apply to research, design, development, deployment, use and decommissioning of AI for use by Defence. It includes all AI technologies and sub-technologies, including frontier models, and any capabilities supported by AI for both warfighting and enabling functions.

These policy settings apply to the Department of Defence (including the Defence Intelligence Organisation and the Australian Geospatial-Intelligence Organisation), the Australian Defence Force, and the Australian Submarine Agency. The Australian Signals Directorate (ASD) applies the *Ethical AI in the ASD* framework and principles to its use of AI.

Definition of AI

The Australian Government has adopted the 2023 Organisation for Economic Co-operation and Development definition of an AI system:

"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."

Obligations for Responsible use of AI

Defence's responsible use of AI technologies will be achieved by meeting three policy requirements:

1. **Lawfulness:** Defence will use all technologies in compliance with Australia's domestic law and international legal obligations.
2. **Adherence to values-based principles**: Defence will ensure use of AI is:
   a) underpinned by individual accountability,
   b) bounded by consideration of impacts on humans,
   c) explainable,
   d) reliable and secure, and
   e) designed to mitigate unintended bias and unintentional harm.
3. **Proportionate controls:** Technology use will be risk-based. Risks will be managed through proportionate control measures.

These requirements apply to the research, design, development, deployment, use and decommissioning of all AI technologies for use by the Department of Defence (including the Defence Intelligence Organisation and the Australian Geospatial-Intelligence Organisation), Australian Defence Force and Australian Submarine Agency.

> Responsible use of AI is the responsibility of all personnel and must be considered at all stages of the technology life cycle.

1. Lawfulness

Defence's domestic and international legal obligations

Defence's use of AI must be lawful. Use of AI technologies in Defence must be compliant with Australia's domestic law and international legal obligations, including under international humanitarian law and international human rights law. This applies throughout the research, design, development, deployment, use and decommissioning of AI technologies.

Personnel with responsibilities across the technology lifecycle must consider legal obligations that may apply to the inclusion of AI in a technology, or its application. Legal considerations may arise at all stages of decision-making to acquire and use AI.

> Human judgement and accountability is central to lawful, legitimate and responsible use of AI, and designated Defence personnel (Accountable Officers) will always be accountable for its use, decisions and outcomes.

Accountable Officers

Responsible use of AI in Defence requires an identifiable Defence official to be specifically accountable for the actions, outcomes and decisions of AI-enabled functions. All Defence officials are accountable for their own contribution. Accountability cannot be transferred to a technology, and humans will always be accountable for the use, decisions and outcomes of AI technologies in Defence.

For the purpose of these policy settings, the Accountable Officer is the official that is accountable for the capability in its present state in the technology life cycle.[1] As the AI capability moves through the life cycle, the Accountable Officer will change, however all officials will remain accountable for the decisions or contributions they made in any stage and for their contribution to the development or use of the AI.

As the Accountable Officer changes during the technology life cycle, for example, when an AI capability moves from acquisition to introduction to service, the transfer of accountability must be documented.

Accountable Officers delegate authorities and responsibilities through the chain of command, in accordance with the *Public Governance Performance and Accountability Act 2013*, and related Defence policies and doctrine. The delegation of authority or responsibility does not change the accountability of the Accountable Officer. Delegations must be appropriately documented.

Accountable Officers will monitor, address and report on how the risks of unintended bias or unintentional harm are detected, corrected and mitigated, with careful consideration of reasonably anticipated impacts.

AI in weapon systems

Specific legal obligations apply to AI in weapon systems. Australia is required by Article 36 of Additional Protocol 1 to the Geneva Conventions to conduct legal reviews of all new weapons, means and methods of warfare to ensure that their employment in some, or all circumstances, is lawful. Defence personnel must seek legal advice at the earliest opportunity during study, development, acquisition or adoption of AI in a weapon system to determine whether and at what stage/s an Article 36 review is necessary, as required throughout the technology life cycle.

[1] Examples of relevant Defence appointments would be capability managers, capability sponsors or design authorities.

2. Adherence to values-based Principles

Defence will apply values-based principles to ensure that our use of AI technology is in line with Australia's high legal and ethical standards and public expectations. Defence has developed these principles for the Defence mission, informed by and aligned with the 2024 *Policy for responsible AI in government* and the values articulated in the *2019 Australia's AI Ethics Framework*.

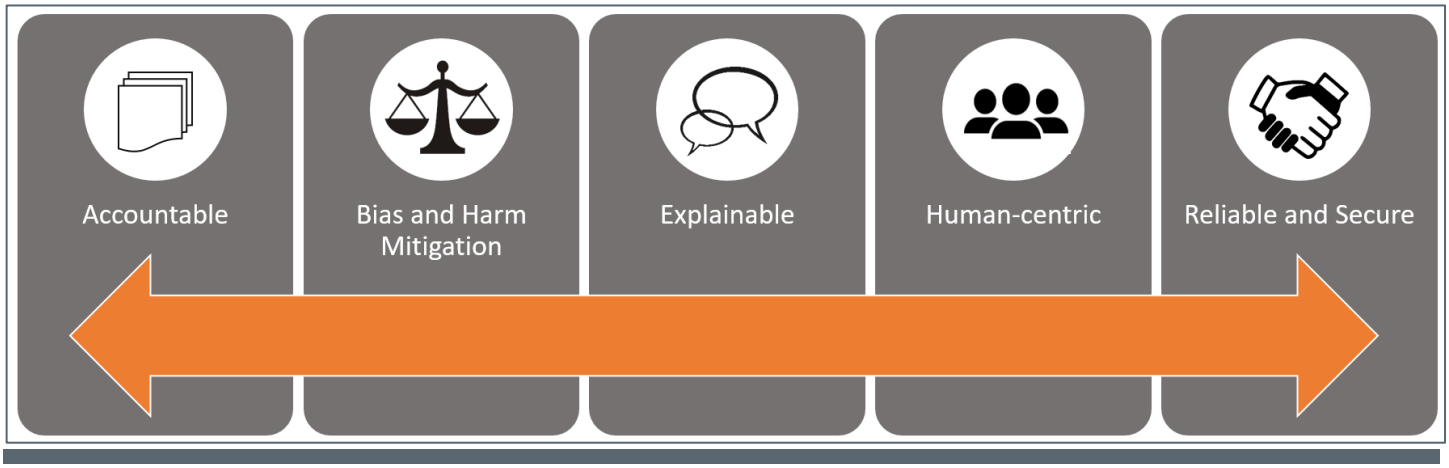

Figure 1: Values-based principles must be applied throughout the life-cycle, from research to decommissioning and post-impact review.

Accountable

1. Defence personnel must exercise informed judgement and care in the research, design, development, deployment, use and decommissioning of AI technologies, and all outcomes will be attributable to an identifiable Defence official.

Bias and Harm mitigation

2. AI technologies must be designed to enable the detection, correction and mitigation of unintended bias and unintentional harm, where it can be reasonably anticipated.

Explainable

3. The function and relationship between inputs and outputs of AI technologies must be traceable, and the technologies able to meaningfully accommodate human-machine interactions.

Human-centric

4. All AI technologies and their use must be considered in light of their planned, likely and potential impact on humans at all levels of involvement.

Reliable and Secure

5. AI technologies must perform reliably against well-defined use cases, with resilience to change or interference, including mitigations and provisions for the risk of failure.

3. Proportionate controls

Control Measures

Defence will apply risk-based control measures to AI technologies proportionate to likely and potential consequences, including unintended outcomes. These risk-based controls consist of layers of policies, processes, training, and procedures, including ongoing assurance and after-action evaluation. Defence uses these to identify, assess and mitigate risks. They must be in place at every stage of the technology lifecycle.

Control measures include robust Test and Evaluation, Post-Incident/Impact Reviews, and appropriate reporting and oversight. Control measures require a clear chain of custody and responsibility for AI technologies within Defence. Accountable Officers must clearly document and accept an AI technology's risks, limitations and controls. These must be communicated before transferring responsibility to another Accountable Officer. This communication must include the nature, functions and limitations of an appropriate control model.

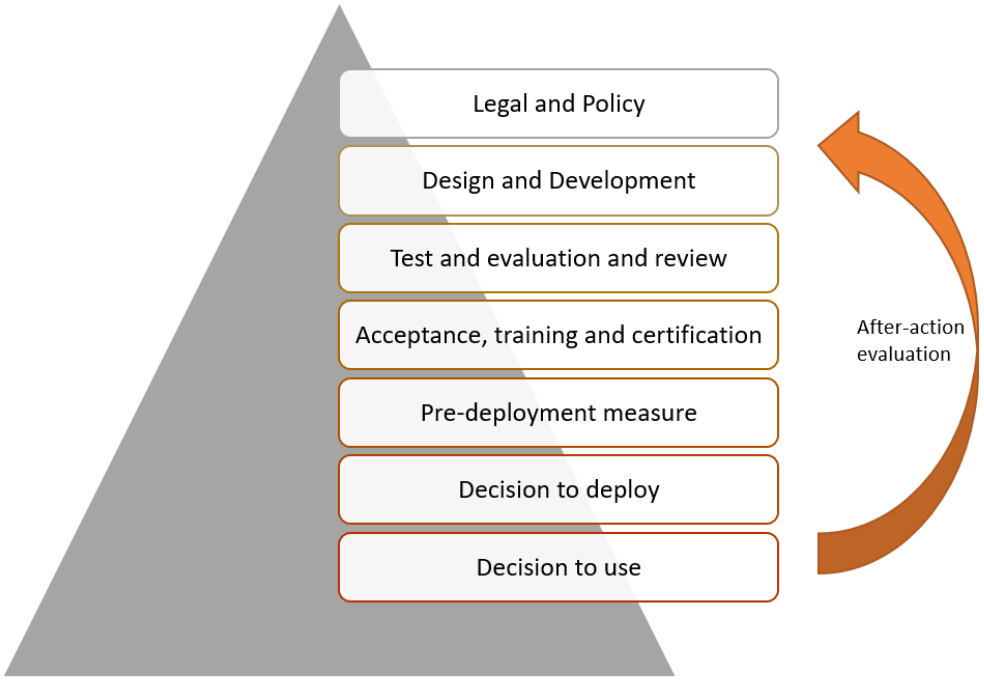

Figure 2: Control measures comprise layers of policies, processes and procedures that enact risk assessment and mitigations.

Control Models

Control models describe how a user and AI technology interact with each other, how human control is enacted throughout a system's operations, and how an AI technology interacts with other AI technologies, systems and machines. Control models continue to evolve as AI technologies advance. Examples may include Direct Human Oversight and Supervision Models (also known as Human-In-The-Loop/Human-On-The-Loop), Human-Machine Teaming, and Machine-Machine-Teaming.

The purpose, capabilities and limitations of individual AI technologies will guide what control model is most appropriate. Ensuring an appropriate control model is applied is the responsibillity of the Accountable Officer. This includes ensuring that users are trained and supported in the technology's use, commensurate with the risk.

Governance and Oversight

Defence is responsible for defending Australia and its national interests in order to advance Australia's security and prosperity. Defence is accountable to the Government in its exercise of this unique function. Defence will meet the obligations of these policy settings by adjusting and strengthening existing governance and controls, or establishing new mechanisms as needed.

External oversight and accountability of Defence includes:

- The Joint Standing Committee on Foreign Affairs, Defence and Trade considers and reports on matters relating to defence referred to it by either House of the Parliament or a Minister. The Committee also inquires into matters raised in annual reports of the Department of Defence.[2]
- The Senate Standing Committees on Foreign Affairs, Defence and Trade (the Legislation Committee and References Committee) conduct public inquiries, review legislation and scrutinise government operations, expenditure and policy actions, including through the Senate estimates process.
- Defence intelligence agencies are subject to oversight by the Parliamentary Joint Committee on Intelligence and Security and the Inspector-General of Intelligence and Security.[3]
- The Independent National Security Legislation Monitor (INSLM) reviews the operation, effectiveness and implications of counter-terrorism and national security legislation, including legislation that engages Defence equities.
- The Auditor-General and the Australian National Audit Office provide independent audit reporting to Parliament on Defence's performance.

As AI technologies and safety standards continue to evolve, Defence, through the Defence Artificial Intelligence Centre, will adapt and develop its AI governance and oversight arrangements to maintain responsible use.

> Defence has established a Defence Artificial Intelligence Centre through which it governs, manages and accelerates Defence's use of AI. As a central hub of expertise, its functions include developing mechanisms to monitor and assure compliance with Government and Defence policies for the responsible use of AI, AI strategy, workforce professionalisation, decentralised development, innovation and engagement, including engagement with Whole of Government initiatives (such as the National AI Centre).

[2] The Committee's *Inquiry into the Department of Defence Annual Report 2022-23* considered AI and Autonomous Weapons related issues. The Committee's *Inquiry into the Department of Defence Annual Report 2023-24* has identified Uncrewed/Autonomous systems, AI and their integration into the Joint Force as a key theme of inquiry.
[3] The Inspector-General of Intelligence and Security issued a *Preliminary Inquiry – Use of AI by Intelligence Agencies* in 2024.

Related Australian Government documents

Department of Defence (2024), Defence Data Strategy 2.0 - Decision Advantage in the Data Age.

Australian Signals Directorate (ASD) (2023), Ethical Artificial Intelligence in the ASD.

Department of Industry, Science and Resources (2019), Australia's AI Ethics Principles.

Department of Industry, Science and Resources (2024), Voluntary AI Safety Standard.

Digital Transformation Agency (2024), Policy for the responsible use of AI in government.

Related international documents

Responsible AI in the Military Domain Summit (2023), *Call to Action*

Responsible AI in the Military Domain Summit (2024), *Blueprint for Action*

Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy (2023)