

Australia's AI Ethics Principles

Source: Department of Industry, Science and Resources — industry.gov.au/publications/australias-ai-ethics-principles

Date published: 7 November 2019 · Date updated: 2 December 2025

Note on currency: on 21 October 2025 the Department published its *Guidance for AI Adoption*, which the source page describes as an updated and simplified evolution of these 8 AI Ethics Principles (and of the 10 guardrails in the Voluntary AI Safety Standard). The principles below remain published for reference, but readers should treat the newer *Guidance for AI Adoption* as the current primary framework.

The Artificial Intelligence (AI) Ethics Principles guided businesses and governments to responsibly design, develop and implement AI. The principles were part of the Australian Government's commitment to make Australia a global leader in responsible and inclusive AI.

These voluntary principles were designed to ensure AI is safe, secure and reliable, aiming to achieve safer, more reliable and fairer outcomes for all Australians, reduce the risk of negative impact on those affected by AI applications, and help businesses and governments practice the highest ethical standards when designing, developing and implementing AI.

AI Ethics Principles at a glance

- **Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment.
- **Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals.
- **Fairness:** AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.
- **Privacy protection and security:** AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.

- **Reliability and safety:** AI systems should reliably operate in accordance with their intended purpose.
- **Transparency and explainability:** There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.
- **Contestability:** When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.
- **Accountability:** People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

AI Ethics Principles in detail

Human, social and environmental wellbeing

Throughout their lifecycle, AI systems should benefit individuals, society and the environment. AI system objectives should be clearly identified and justified, and AI systems that help address areas of global concern — like the United Nations' Sustainable Development Goals — should be encouraged. Ideally, AI systems should be used to benefit all human beings, including future generations.

AI systems designed for legitimate internal business purposes, like increasing efficiency, can have broader impacts on individual, social and environmental wellbeing. Those impacts, both positive and negative, should be accounted for throughout the AI system's lifecycle, including impacts outside the organisation.

Human-centred values

Throughout their lifecycle, AI systems should respect human rights, diversity, and the autonomy of individuals. Machines should serve humans, and not the other way around: AI systems should enable an equitable and democratic society by respecting, protecting and promoting human rights, enabling diversity, respecting human freedom and the autonomy of individuals, and protecting the environment.

All people interacting with AI systems should be able to keep full and effective control over themselves. AI systems should not undermine the democratic process, and should not undertake actions that threaten individual autonomy, like deception, unfair manipulation, unjustified surveillance, and failing to maintain alignment between a disclosed purpose and true action.

Fairness

Throughout their lifecycle, AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups. AI systems should be user-centric and designed in a way that allows all people interacting with them to access the related products or services, including appropriate consultation with affected stakeholders and equitable access and treatment.

This is particularly important given concerns about the potential for AI to perpetuate societal injustices and have a disparate impact on vulnerable and underrepresented groups. Measures should be taken to ensure AI-produced decisions are compliant with anti-discrimination laws.

Privacy protection and security

Throughout their lifecycle, AI systems should respect and uphold privacy rights and data protection, and ensure the security of data. This includes proper data governance and management for all data used and generated by the AI system, and appropriate data anonymisation where used.

Appropriate data and AI system security measures should be in place, including identification of potential security vulnerabilities and assurance of resilience to adversarial attacks, accounting for unintended applications and potential abuse risks with appropriate mitigation.

Reliability and safety

Throughout their lifecycle, AI systems should reliably operate in accordance with their intended purpose, including being reliable, accurate and reproducible as appropriate. AI systems should not pose unreasonable safety risks, and should adopt safety measures proportionate to the magnitude of potential risks.

AI systems should be monitored and tested to ensure they continue to meet their intended purpose, with identified problems addressed through ongoing risk management, and responsibility clearly identified for ensuring an AI system is robust and safe.

Transparency and explainability

There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them. Responsible disclosures should be provided in a timely manner, with reasonable justifications for AI system outcomes, including information that helps people understand key factors used in decision-making.

This principle also aims to ensure people can find out when an AI system is engaging with them, regardless of the level of impact, and are able to obtain a reasonable disclosure regarding the AI system.

Contestability

When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system. Knowing that redress for harm is possible when things go wrong is key to ensuring public trust in AI, with particular attention paid to vulnerable persons or groups.

There should be sufficient access to the information available to the algorithm, and inferences drawn, to make contestability effective. Where decisions significantly affect rights, there should be an effective system of oversight that makes appropriate use of human judgment.

Accountability

Those responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled. Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after design, development, deployment and operation.

The organisation and individual accountable for a decision should be identifiable, and consider the appropriate level of human control or oversight for the particular AI system or use case. AI systems that have a significant impact on an individual's rights should be accountable to external review, including providing timely, accurate and complete information for independent oversight bodies.

This document reproduces publicly published policy text from the Department of Industry, Science and Resources for reference purposes on the YouMimic AI Ethics page. Refer to the source URL above for the current, authoritative version and for the Department's newer Guidance for AI Adoption.