

The Ethics of Artificial Intelligence: Laws from Around the World

Paul Kallenbach, Vanessa Mellis, Annabelle Ritchie and Siegfried Clarke at MinterEllison give us a global view of the laws regulating artificial intelligence.

The use of artificial intelligence and machine learning (AI) driven solutions is becoming increasingly common in Australian businesses. By processing enormous amounts of information in a very short time, AI can minimise human intervention in decision making processes, and allow organisations to optimise their operations. The Minister for Industry, Science and Technology, Karen Andrews, has commented that “AI has the potential to provide real social, economic, and environmental benefits – boosting Australia’s economic growth and making direct improvements to people’s everyday lives”.

But at what cost? When does AI overstep the mark and become a tool that does more harm than good? Are we adequately assessing AI against our current policies, legal systems, business due diligence practices, and methods to protect human rights?

In this article, we examine the creation of frameworks for the ethical use of AI in Australia and globally.

Australia

Following the release of the *‘Artificial Intelligence: governance and leadership’* whitepaper by the Australian Human Rights Commission in January 2019, CSIRO’s Data61 released a discussion paper to boost conversation about AI ethics in Australia (**Discussion Paper**). The Discussion Paper focuses on civilian (as distinct from military) applications of AI and adopts the view that the key to unlocking the potential of AI is to ensure that the public have trust in AI driven solutions.

The Discussion Paper draws upon international approaches, as well as those developed by companies such as Google and Microsoft, to propose eight core principles to guide developers, industry and government in ethically deploying AI driven systems, namely:

- 1. Generate Net Benefits:** AI systems must generate benefits for people which outweigh the costs.
- 2. Do No Harm:** Civilian AI systems must not be designed to harm or deceive people and should minimise negative outcomes.
- 3. Regulatory and Legal Compliance:** AI systems must comply with all relevant laws, regulations and government obligations.
- 4. Privacy Protection:** AI systems must ensure that private data is protected and kept confidential and prevent harmful data breaches.
- 5. Fairness:** AI systems must not result in unfair discrimination against individuals, communities or groups. They must be free from training biases which may cause unfairness.
- 6. Transparency and Explainability:** People must be informed when an algorithm is being used which impacts them, and they should be informed about what information the algorithm uses to make decisions.
- 7. Contestability:** Where an algorithm impacts a person there must be an efficient process to challenge the use or output of the algorithm.

- 8. Accountability:** People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm, even if the impacts are unintended.

In addition to these eight core principles, the Discussion Paper proposes a toolkit to assist stakeholders in applying those principles. We discuss this in more detail, elsewhere in this edition of CLB, in the article *Beyond Asimov’s Three Laws: A new ethical framework for AI developers*.

Developments in Europe

The EU has established a High-Level Expert Group comprised of 52 experts on AI, including representatives from academia, civil society and industry, selected by the European Commission. The group recently published its *Ethics Guidelines for Trustworthy AI (EU Guidelines)*, following stakeholder consultation on draft guidelines. The EU Guidelines are not binding, but offer stakeholders a set of guiding principles to follow to indicate their commitment to achieving “Trustworthy AI”.

The EU Guidelines start by identifying four key ethical imperatives, which reflect the EU Charter of Fundamental Rights:

- respect for human autonomy;
- prevention of harm;
- fairness;
- explicability.

From these four ethical imperatives, the guidelines then derive seven key requirements.

Like Australia's Discussion Paper, the EU Guidelines highlight the importance of non-discrimination, promoting societal and environmental wellbeing, privacy, accountability and transparency.

However, the EU Guidelines include more of a direct focus on human rights, and human agency, as well as technical robustness and safety, where Australia's Discussion Paper focuses on regulatory and legal compliance and contestability.

Notably, although the EU guidelines are not binding, they are informed by the EU Charter of Fundamental Rights which is an instrument of EU Law that lacks an Australian counterpart. The guidelines note that the four ethical imperatives are in many respects already reflected to some extent in existing legal requirements. For example, amongst the wide-reaching provisions of the European Union's *General Data Protection Regulation*, article 22 provides that a decision with legal ramifications for a person (or which would similarly seriously affect them) cannot be based solely on automated processing or profiling, in most situations.

Other developments within the technology industry

Governments are not working in isolation on the complex ethical issues which AI presents. Silicon Valley's well known preference for self-regulation has manifested itself in AI policy focused initiatives such as the San Francisco based **Partnership on AI** which counts the tech sector's big four amongst its members. Facebook in collaboration with the Technical University of Munich has announced **\$7.5 million in funding for an independent AI ethics research centre**, while Amazon is working with CSIRO's US counterpart, the National Science Foundation to fund **research into fairness in AI**.

Microsoft

Microsoft recently released *Microsoft's Vision for AI in the Enterprise*, which outlines its approach to the use of AI. This document addresses the ethical challenges raised by AI, and states that designing trustworthy AI requires creating solutions reflecting ethical principles. Trustworthy AI, the company states, requires **"solutions that reflect ethical principles that are deeply rooted in important and timeless values."** For Microsoft this includes fairness, reliability and safety, privacy and security, inclusivity, transparency and accountability. In 2018, the tech giant established an **AI and Ethics in Engineering and Research (AETHER) committee**, to bring together senior leaders to craft internal policy and address ethics in specific issues.

Google

Google has also implemented a **set of principles governing the use of AI** and has experimented with an external AI ethics panel to offer guidance on ethical issues. The principles set out Google's objectives in assessing AI applications, include to:

- be socially beneficial;
- avoid creating or reinforcing unfair bias;
- be built and tested for safety;
- be accountable to people;
- incorporate privacy design principles;
- uphold high standards of scientific excellence; and
- be made available for uses that accord with these principles.

Google has also published a list of AI applications which it will not pursue, including technologies that cause or are likely to cause overall harm, weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people, technologies that gather or use information for surveillance

violating internationally accepted norms or technologies whose purpose contravenes widely accepted principles of international law and human rights.

The rise in AI no doubt presents incredible opportunities for governments, private actors and society more widely but it is not without serious ethical risks. This has prompted responses from organisations and nations worldwide. In Australia, following the release of the Discussion Paper we expect work in this space to continue.