Trustworthy AI in Aotearoa

AI Principles

March 2020

forum
NEW ZEALAND

Te Kāhui Atamai Iahiko o Aotearoa

Trustworthy AI in Aotearoa

March 2020

Artificial intelligence (**AI**) can drive significant economic and social benefits for New Zealand. But it also introduces a range of risks and challenges to New Zealand society that cannot be overlooked.

To help maintain public trust in the development and use of AI in New Zealand, the Law, Society and Ethics Working Group of the AI Forum has published a set of guiding principles for "Trustworthy AI in Aotearoa New Zealand" (**the AI Principles**). Those principles are designed to provide high-level guidance for anyone involved in designing, developing and using artificial intelligence in New Zealand (**AI stakeholders**), with the goal of ensuring New Zealanders have access to trustworthy AI.

A key focus of the group has been to make sure the AI principles are simple, succinct and user friendly.

PURPOSE OF THE PRINCIPLES

The AI Principles aim to:

- start a conversation about the importance of ethical and legal considerations in AI design, development, deployment and operation
- raise awareness that AI ethical and legal issues need to be identified and addressed early on
- set the groundwork for more detailed, practical guidance that will define and inform good practice in AI design, development and implementation in New Zealand.

A FIRST STEP

The AI Principles are designed to be very much a first step in developing more comprehensive Trustworthy AI guidance for New Zealand.

It is hoped the AI Principles can be operationalised through a follow-up set of detailed and practical guidelines to help ensure adequate consideration of the legal and ethical implications of AI at each stage of the AI lifecycle. It is expected that such guidelines would help AI stakeholders understand what they should be planning, doing and reviewing at each state of the AI lifecycle.

We also recognise the complex issues associated with AI cannot be reduced to a simple "tick-box" activity. As such, this document is intended to provide to set of overarching first principles for AI stakeholders to refer to when faced with challenging questions relating to AI.

ALIGNMENT WITH EMERGING AI ETHICAL NORMS

We have drawn upon the common themes emerging from the growing body of published AI ethical principles, including the OECD Recommendation on Artificial Intelligence, the European Commission's Ethics Guidelines for Trustworthy AI, the iTech Principles, the Montréal Declaration and Singapore's Proposed Model AI Governance Framework.

However, we have aimed to distil the language and concepts used elsewhere into something more accessible and relevant to New Zealand.

MORE THAN ETHICS

A common feature of many sets of AI principles is a focus on ethical considerations only.

We believe it is important to remind AI stakeholders that existing laws and regulations apply to AI just as they do to any other form of technology. AI does not exist in a legal vacuum - among others, existing laws relating to consumer protection, privacy, liability, intellectual property and human rights will all continue to apply. The AI Principles are no substitute for compliance with those legal regimes.

Human rights law plays a key role. It provides a ready-made, internationally tested and legitimate framework of civil, political, economic, cultural and social values, addressing both individual and collective concerns.

THE ROLE OF GOVERNMENT

The AI Principles are designed to assist everyone in the AI ecosystem, including in both the private and public sectors. However, we recognise that Government has additional obligations and a broader role to play in ensuring AI and other emerging technologies serve the long-term public good of New Zealand, including in meeting its obligations under Te Tiriti o Waitangi.

Government regulation and regulators have an important role to play here. Self-regulation in the form of ethical principles or standards may fill a gap where the law is incomplete or out of date, but they are no substitution for democratically-mandated rules backed up by the force of law.

We note also that public sector agencies' AI-related decisions are subject to additional public and international law obligations. Those demand a higher level of scrutiny, transparency and protection of human rights than is currently legally demanded of businesses. Drawing out the implications of these requirements in the AI context is an important task we hope the Government will undertake, with support from organisations like the AI Forum, which is a neutral, professional and inclusive group of thought leaders, experts and practitioners.

CONCLUSION

The overall intention of these AI principles is not to provide an exhaustive list of AI legal and ethical principles. Instead, we have aimed to provide a useful reference point that summarises key considerations in the design, development, deployment and operation of AI systems.

We believe AI stakeholders that design, develop and use AI systems in accordance with these principles will be better able to manage many of the identified risks and untintended consequences of AI. As a result, the public will be more likely to trust AI, ultimately enhancing the ability of all New Zealanders to enjoy the benefits of AI.

Trustworthy AI in Aotearoa
The AI Principles

The following principles aim to foster both innovation and trust in the design, development and deployment of artificial intelligence (**AI**) in Aotearoa New Zealand.

1. FAIRNESS AND JUSTICE

Designers, developers and users of AI systems (**AI stakeholders**) must respect:

- **Applicable laws** in New Zealand and other relevant jurisdictions
- **Human rights** recognised under domestic and international law
- **Rights of Māori** articulated in Te Tiriti o Waitangi
- **Democratic values** including the electoral process and informed public debate
- **Principles of equality and fairness** so that AI systems do not unjustly harm, exclude, disempower or discriminate against individuals or particular groups.

2. RELIABILITY, SECURITY AND PRIVACY

AI stakeholders must ensure AI systems and related data are reliable, accurate and secure and the privacy of individuals is protected throughout the AI system's life cycle, with potential risks identified and managed on an ongoing basis.

3. TRANSPARENCY

The operation and impacts of an AI system should be transparent, traceable, auditable and generally explainable to a degree appropriate to its use and potential risk profile so outcomes can be understood and challenged, particularly where they relate to people.

4. HUMAN OVERSIGHT AND ACCOUNTABILITY

AI stakeholders should retain an appropriate level of human oversight of AI systems and their outputs. Technologies capable of harming individuals or groups should not be deployed until stakeholders have determined appropriate accountability and liability.

5. WELLBEING

Where appropriate, AI stakeholders should design, develop and use AI systems to promote, as much as possible, the wellbeing of New Zealand's people and environment in areas such as health, education, employment, sustainability, diversity, inclusion and recognition of the unique values of Te Ao Māori.

LAW, ETHICS & SOCIETY WORKING GROUP 6 CO-LEADS

Frith Tweedie, Head of Digital Law, EY Law Limited, (Co-lead W6)

Frith has over 18 years' experience helping New Zealand, Australian and UK clients with their privacy, emerging technology, IP and e-commerce law. Frith is currently leading the development of an interactive website detailing AI policy and legislation around the world. She is a member of the Executive Council of the AI Forum and the Australia New Zealand Advisory Board of the International Association of Privacy Professionals (IAPP).

Dr Sean Welsh, University of Canterbury, (Co-lead W6)

Sean is the author of Ethics and Security Automata (Routledge, 2018) and Ethik in KI und Robotik (Hanser, 2019). An English version of the latter, Ethics in AI and Robotics, is forthcoming from Springer later this year. Sean has a PhD in philosophy from the University of Canterbury. He is currently working on an edtech start-up.