Title:

PROJECT DOCUMENT: AI Regulation A3

Author:

OPMCSA

Regulating risks associated with AI

There are known risks to unregulated AI

Generative AI could be used to spread misinformation and disinformation and manipulate social institutions. A recent example pictured the Pentagon after a bomb, which resulted in a temporary decline in the stock market.

The issue of large players in the search or recommendation space producing output that privileges their own products and interests could be exacerbated by generative AI.

Algorithmic decision making can amplify bias and discrimination. In sectors with potentially large impacts on people's lives, for example, access to public services, hiring, or lending, this could entrench structural inequalities and disadvantage.

Some uses of AI are incompatible with human rights protections, for example real-time biometric surveillance.

Training data and data created in the use of AI may violate data privacy expectations and IP.

Designing regulation is challenging

This is a rapidly evolving space. Regulation needs to be technologically neutral so as not to become out of date.

One issue is how to monitor, audit, and enforce compliance with regulations. In general, independent oversight engenders trust and credibility, but there are budgetary considerations to creating a new oversight or auditing body. Additionally, any monitoring or auditing of compliance will require a highly technical workforce, which, in practice may mean recruiting from industry and in turn could create conflicts of interest.

Regulation should also consider proportionality in the cost of compliance. This could mean having different requirements for small and medium size enterprises (SMEs) than for larger organisations, or making funding available to support SMEs in complying.

Finally, with regard to generative AI, there is the question of "guardrails", or preventing harmful content from being created. To date, the large companies creating these tools have attempted to curtail responses that provide guidance on criminal activity or promote some kinds of offensive ideas. As AI tools become more ubiquitous, however, it is probably not compatible with democratic principles that the selection of guardrails is wholly left to large corporate entities. Moreover, any guardrails – whether arising from within the private sector or elsewhere – can potentially be subverted.

There are additional considerations in the Aotearoa New Zealand context

The risks of unregulated AI and challenges in designing fit for purpose regulation on the other all apply in NZ. However, there are additional considerations in our context.

A related concern is ensuring that AI applications developed overseas and trained on international data are fit for purpose in our context. This will be especially important for domains such as health where demographics and place, conceived broadly, have large impacts.

Any regulation will need to meet obligations under Te Tiriti o Waitangi, as well as being consistent with a recent Supreme Court finding that Tikanga Māori is common law. The United Nations Declaration on the Rights of Indigenous Peoples 2007 also has implications. Finally, regulations will need to ensure AI products respect Māori data sovereignty.

As a relatively small economy, NZ doesn't have the market power to incentivise suppliers to comply with overly onerous regulation that is not in place elsewhere. To the degree that it is important for NZ government, businesses, and consumers to have access to various AI applications, NZ may need to harmonise any AI regulations with those of other countries. Conversely, NZ may be able to benefit from products designed for stricter regulations in other jurisdictions.

International consensus is emerging

Although relatively few countries have drafted regulation, many have published principles around how AI should operate within their societies. These include Australia's Artificial Intelligence Ethics Framework, the US's AI Bill of Rights NIST, and Singapore's Model AI Governance Framework. Supra-national institutions such as the OECD and WHO have also published their own guidance or principals on the ethical use of AI. Though each set of principles reflects its own context, consistent themes are apparent in the principles published by the countries and institutions which are generally aligned with NZ values. These principles are summarised below.

Transparency People should know when and how AI is involved in a decision that affects their lives. Relatedly, individual decisions made by or with assistance from AI should be *explainable*.

Accountability A person or organisation must take responsibility for the outcomes of AI, whether the outcomes are intended or not. Relatedly, an individual decision or outcome from an AI process should be *contestable*.

Robust, secure, and safe systems System should not pose harm to people at any point during its life cycle, not only in normal use but with possible misuse or in adverse conditions. Among other things, this requires protecting *privacy* and avoiding discrimination.

Fairness The use of AI should not lead to discrimination against individuals or groups. Equity implications should be considered in the design of a system.

These are the principles with the strongest international consensus – and are aligned with the OECD AI Principles to which NZ has already committed. They are also consistent with the AI Forum's Trustworthy AI in Aotearoa principles, which consider the local context including te Tiriti and a focus on wellbeing. As such, they represent a reasonable starting point for thinking about NZ's approach to AI risk management, but are not comprehensive and will need to be adapted to include unique local considerations such as Māori data sovereignty.

Rapid developments in AI – exemplified recently by publicly accessible generative AI – will make the application of currently accepted principles more challenging and may even require new principles.

Selected resources

OMPCSA online hub of resources

Rapid Response Information Report – Generative AI: Language models and multimodal foundation models. Produced by Australia's Chief Scientist, March 2023.

Blueprint for an AI Bill of Rights: Making automated systems work for the American people. White House Office of Science and Technology Policy, October 2022.

Māori data governance Model. Te Kāhui Raraunga, June 2023.

Dawson D and Schleiger E, Horton J, McLaughlin J, Robinson C, Quezada G, Scowcroft J, and Hajkowicz S (2019) Artificial Intelligence: Australia's Ethics Framework. Data61 CSIRO, Australia.

Ethics, Transparency and Accountability Framework for Automated Decision-Making. UK Government, May 2021.

OECD AI Principles Overview May 2019. OECD, available at OECD Policy Observatory.

Comparison of global approaches to AI risk management

	EU	US	Other	NZ
Dedicated AI legislation?	AI Act in in the works and expected to become law later in the year; legislation will require creation of standards Existing GDPR covers algorithmic decision making and targeted ads, and Digital Services and Digital Market Acts target transparency and fair market competition	White House has produced AI Bill of Rights and other guidelines; these have no mechanism to compel compliance Algorithm Accountability Act before both chambers of Congress; not clear whether has political legs to progress Some state legislatures have passed algorithm accountability legislation; NYC has imposed requirements on the use of algorithms in hiring/promotion Some existing legislation has implications for AI e.g. around fair trading practices, anti-discrimination	China: Has taken a "vertical" approach; individual pieces of legislation on algorithmic recommendations, deep synthesis, and generative AI. Legislation around generative AI and deepfakes has created a compulsory registry; it is expected this will be a part of future legislation for different AI Canada: Digital Charter Implementation Act (Bill C-27) passed second reading in lower house; long process to go before adopted Australia: Consultation process for creating legislation launched June 2023, including rapid evidence review and a report to inform public submissions UK: no dedicated legislation. AI white paper describes possible future regulation in context of wider strategic approach to AI.	No dedicated legislation. Some existing legislation has implications for AI e.g. Privacy Act, Human Rights Act, as well as Te Tiriti o Waitangi. Algorithm Charter has been adopted by most government agencies.
AI for human processes/socioeconomic decisions *AI in hiring, educational access, and financial services approval*	GDPR requires human in the loop for significant decisions. "High-risk" AI applications AI Act would need to meet quality standards, implement risk management system, and perform conformity assessment	AI Bill of Rights and associated Federal Agency Actions have created patchwork oversight for some of these applications. Notable gap even in Algorithm Accountability Act is that some sectors are out of scope, including public services NYC requires impact assessment of hiring and promotion decisions that involve algorithms; in practice, these requirements are poorly defined, and deadlines have been repeatedly pushed back as a result	Canada: Directive on Automated Decision Making applies to government services and imposes requirements around transparency when there are AI components in decision making UK: Piloting Algorithmic Transparency Reporting Standard in some parts of government	Algorithm Charter has been adopted by most government agencies
AI in consumer products *AI in medical devices, partially autonomous vehicles, and planes*	AI Act considers AI implemented within products that are already regulated under EU law to be high risk; new AI standards to be incorporated into current regulatory process.	Individual federal agency adaptations, such as by FDA for medical devices; DOT for automated vehicles; CPSC for consumer products		Existing laws such as Consumer Guarantees Act, Human Rights Act, Privacy Act apply as relevant
Chatbots	AI Act would require disclosure that a chatbot is an AI (i.e., not a human).	California BOT Act makes it an offense to pretend to be a person to sell products of influence elections.		Existing laws such as Fair Trading Act 1986, Human Rights Act, Privacy Act 2020apply as relevant, as does sector-specific regulation (e.g. in the financial services sector)
Social media recommender algorithms *Newsfeeds and group recommendations on social media*	Digital Services Act creates transparency requirement; also enables independent research and analysis		China: Consumer must be informed that an algorithm has been used	NZ Code of Practice for Online Safety and Harms applies as relevant if a company has adopted it
Algorithms on e-commerce platforms *Algorithms for search or recommendation of products and vendors*	Digital Markets Act restricts self-preferencing algorithms in digital markets		China: Prohibits use of personal info in price setting	
Foundation models/generative AI *DALL-E; ChatGPT*	Draft proposals of the EU AI Act consider quality and risk management requirements.		China: Output must be true, unbiased, and conform with state ideology; developers responsible for all content produced (even by a different end user). Developers responsible for ensuring training data are unbiased, objective and accurate	Privacy Act applies; Office of the Privacy Commissioner released guidance on this in May 2023.
Facial recognition	AI Act will include restrictions on remote facial recognition and biometric identification. Data Protection Authorities have fined facial recognition companies under GDPR.	NIST's AI Face Recognition Vendor Test program contributes efficacy and fairness information to the market for facial recognition software.		Office of the Privacy Commissioner is exploring a code of practice on biometrics.
Targeted advertising *Algorithmically targeted advertising on websites and phone applications*	Meta has been fined under GDPR for using personal user data for behavioural ads. The Digital Services Act bans targeted advertising to children and certain types of profiling (e.g., by sexual orientation). It requires targeted ads have explanations and users have control over what ads they see.	Individual federal agency lawsuits have slightly curtailed some targeted advertising. This includes the DOJ and HUD, who successfully sued Meta for discriminatory housing ads and an FTC penalty against Twitter for using security data for targeted ads.		Existing law such as Unsolicited Electronic Messages Act 2007 and Privacy Act 2020 apply as relevant.