

DISCUSSION PAPER:

# Responsible AI

Ensuring AI works reliably, fairly,  
transparently and safely for everyone.

DECEMBER 2023

## ABOUT

The present paper was prepared by Dr Helen Lu (University of Auckland), Dr Karaitiana Taiuru (Taiuru & Associates), Dr Paul Brown (University of Waikato), Dr Paul Geertsema (University of Auckland), Dr Daniel Wilson (University of Auckland), Dr Amanda Williamson (University of Waikato), Prof. Richard Green (University of Canterbury), Dr Benjamin Liu (University of Auckland).

The views and opinions expressed in this discussion paper are solely those of the authors listed above, and do not reflect the views of any other organisation or person, whether affiliated with the authors or otherwise.

## CIRCULATED BY

The Artificial Intelligence Researchers Association

Website: [www.airesearchers.nz](http://www.airesearchers.nz)

Email: [hello@airesearchers.nz](mailto:hello@airesearchers.nz)

# CONTENTS

## INTRODUCTION

<b>Introduction</b> .....	4-5
<b>Why It Matters?</b> .....	6
<b>Potential Pitfalls of AI Use</b> .....	7-10
Performance .....	7
Biases .....	8
Privacy Breaches .....	9
Vendor Risks .....	10
<b>High Level Principles</b> .....	11
<b>Translating Principles into Practices</b> .....	12-13

## CHALLENGES AND OPPORTUNITIES

<b>Challenge 1: Many SMEs are Yet to Fully Embrace Digitalisation</b> .....	14-15
<b>Challenge 2: Lack of Awareness of AI Usage &amp; Risks</b> .....	16
<b>Opportunity 1: Responsible AI by Design</b> .....	16
<b>Opportunity 2: Regulatory Flexibility</b> .....	17
<b>Opportunity 3: Te Tiriti o Waitangi and AI</b> .....	18-19

## RECOMMENDATIONS

<b>New Zealand SMEs</b> .....	20-21
<b>Government Policy</b> .....	22

## APPENDICES

<b>Essential AI Concepts Explained</b> .....	23
<b>Resources for Definitions &amp; Principles</b> .....	24

# Introduction

**Artificial Intelligence (AI)<sup>[1]</sup> is transforming our lives. Imagine a day without auto-correct or Google maps. Beyond these everyday applications, AI powers groundbreaking advances in fields like medicine<sup>[2]</sup> and agriculture<sup>[3]</sup>. With the emergence of Generative AI applications such as ChatGPT, DALL-E, Stable Diffusion and MidJourney, many powerful AI tools are now within everyone's reach.**

This rapid progress is both exciting and potentially concerning. While AI wields immense power, it can bring unintended harms: for example, misinformation at scale, unintended privacy breaches, exacerbating existing biases and AI-assisted cyber-attacks. Many companies have grappled with AI missteps, leading to damaged reputations, security issues, lawsuits and financial losses. To navigate these challenges and ensure AI is used for the benefit of society, industry leaders have adopted Responsible AI (RAI) practices. At the same time, law makers are being urged to expedite the introduction of comprehensive AI regulations to create a level playing field.

**“  
It is about ensuring  
AI works reliably,  
fairly, transparently,  
and safely for everyone.  
”**

RAI goes beyond compliance. It is about ensuring AI works reliably, fairly, transparently, and safely for everyone. This calls for commitment from senior leadership, RAI awareness across the organisation, concrete procedures and measurable outcomes.

This report aims to promote responsible AI innovation and adoption in New Zealand. We delve into the unique challenges and opportunities presented to our Small and Medium-size Enterprises (SMEs) and offer practical advice for organisations and regulators.

[1] Appendix A contains brief definitions of technologies in acronyms for non-technical readers

[2] Arnold, C. Inside the nascent industry of AI-designed drugs. *Nat Med* 29, 1292–1295 (2023).  
<https://doi.org/10.1038/s41591-023-02361-0>

[3] “Artificial intelligence in the agri-food sector: Applications, risks and impacts”, Panel for the Future of Science and Technology, European Parliament, March 2023

## In the journey towards responsible AI innovations and adoption, New Zealand faces two primary challenges:

**1** SMEs employ more than 90% of workers in New Zealand, but many SMEs are yet to fully embrace digitalisation.<sup>[4]</sup> Digitalisation is a necessary foundation for SMEs to harness AI's potential. Our businesses can't realise the full benefits of AI without a robust digital infrastructure and a data-driven mindset.

**2** Many SMEs may be unaware of the extent of AI usage within their operations. The easy accessibility of AI tools, especially Generative AI, means employees might be using them without the knowledge of their managers and without clear policy guidance. Furthermore, instead of training AI models in house, many SMEs are likely to primarily rely on third-party solutions, which can further obscure the degree of AI integration in daily tasks. This reliance on external providers may also raise concerns about data privacy, security, and potential dependencies on vendors, underscoring the importance of clear RAI guideline.

## Despite facing challenges, New Zealand has the opportunity to learn and benefit from global innovations in RAI and from regulatory experiences elsewhere:

**1** SMEs can integrate responsible AI practices into their core strategies, governance and risk management systems in the early design stage of AI projects. Embedding these principles can lead to more trustworthy and responsible AI use.

**2** New Zealand's nascent AI regulatory landscape allows it to draw from best practices of more established AI regulatory regimes.

**3** New Zealand has the potential to position itself as one of the leaders in culturally appropriate regulation through aligning AI usage with Te Tiriti o Waitangi.

[4] Digitalisation refers to the use of digital technologies to change a business model, providing new revenue and value-producing opportunities. It encompasses the move from traditional processes to digital ones, and the use of digital data to inform decision-making.

# RESPONSIBLE AI

## Why It Matters?

**Responsible AI (RAI) can be defined as the practice of developing, deploying and using AI to benefit everyone reliably, fairly, transparently, and safely.<sup>[5]</sup>**

RAI is a necessity because AI failures are frequent and can cause unforeseen harms. In a 2021 global survey of over 1,000 large firms, conducted by BCG and MIT Sloan, nearly a quarter of respondents reported experiencing AI failures. These ranged from technical issues to unexpected outcomes that put individuals and communities at risk.<sup>[6]</sup> Since Generative AI emerged in 2022, more executives are becoming aware of the risks within their organization as employees experiment with this new form of AI.

In light of these challenges, and with the need to maintain reputation and avoid regulatory backlash, leading technology companies have an inherent interest in promoting and adopting RAI. Driven by a commitment to build trust with customers and the wider society, these companies are among the first to embark on a RAI journey.

AI mishaps have also caught the attention of regulators. While lawmakers are developing AI-specific legislation, the future of AI regulation remains highly uncertain due to divergent regulatory visions among nations. The European Union (EU) adopts a cautious and risk-based approach in its proposed AI Act.<sup>[7]</sup> While this draft received approval from the European Parliament in June 2023, it is still to be debated and passed by the European Commission and the Council. In contrast to the EU, no other major jurisdiction has bespoke, overarching AI legislation.<sup>[8]</sup>

Meanwhile, the majority of organisations do not develop AI applications in-house. A common question from this group is whether they should wait for the passage of laws to gain clarity before implementing RAI programmes. We believe that they will be better off adopting a forward-looking RAI programme, instead of a reactive approach. Such an approach is viable since technological advances do not change our fundamental ethical principles. Without a proactive RAI programme, both managers and employees risk overlooking blind corners and making missteps. Furthermore, a robust RAI programme can build in flexibility to adapt to future regulatory changes.

---

[5] This definition is consistent with those made by leading consulting firms and government agencies. See Appendix B for a list of definitions of RAI principles.

[6] "To Be a Responsible AI Leader, Focus on Being Responsible," MIT Sloan Management Review and Boston Consulting Group, September 2022.

[7] In the proposed AI Act, certain high-risk AI uses are banned, while the requirements of control and disclosure for permitted AI uses increase with the risk levels.

[8] We do not provide a detailed summary of global regulations related to data and AI. Refer to "Why is regulating AI such a challenge?" Office of the Prime Minister's Chief Science Advisor, New Zealand.  
<https://www.pmcsa.ac.nz/2023/07/13/why-is-regulating-ai-such-a-challenge/>

# Potential pitfalls of AI use

**AI failures are common and can have unforeseen harmful effects on employees, customers and communities. In this section, we highlight potential pitfalls, using examples related to performance, biases, privacy breaches and vendor risks.**

## PERFORMANCE

Carefully developed AI models tend to perform well in their pilot stage. However, once deployed to analyse new data, their performance often degrades. This occurs because the underlying distribution of the new data differs from that used in the pilot stage, a common problem in AI projects known as “data drift”.



**Significant dips in performance warrant scrutiny and may necessitate re-training the AI model with new data.**



To mitigate the impact from data drift, the AI system’s performance must be continuously monitored. Significant dips in performance warrant scrutiny and may necessitate re-training the AI model with new data.

Drops in model accuracy related to data drift can result in “biased” recommendations. For instance, an AI-model trained to diagnose breast cancer might yield varying accuracy levels among different ethnic groups.<sup>[9]</sup> If trained on a dataset scant in observations of a particular ethnic group, these individuals may be more likely to be assigned as “low risk” incorrectly. The impact might be fewer follow-up scans for the false negatives, and potentially more undiagnosed/untreated cancer cases, worsening health inequality for an already disadvantaged group.

New Zealand organisations are likely to use AI models trained on international data, through systems or applications purchased from vendors. The model performance might deteriorate on New Zealand data in domains where demographics play an important role (e.g. healthcare).

To protect disadvantaged groups, it is prudent to maintain human oversight, monitoring model performance and benchmarking it across diverse groups. Model underperformance in certain groups invites a closer investigation and remedies.

[9] Mittermaier, M., Raza, M.M. & Kvedar, J.C. Bias in AI-based models for medical applications: challenges and mitigation strategies. *npj Digit. Med.* 6, 113 (2023). <https://doi.org/10.1038/s41746-023-00858-z>; Dressel J, Farid H. The accuracy, fairness, and limits of predicting recidivism. *Sci Adv.* 2018;4(1):eaao5580. Published 2018 Jan 17. doi:10.1126/sciadv.aao5580

## BIASES

We'd like to highlight the distinction between a model's performance degradation (often manifested as low accuracy) and its inherent biases. It is important to understand that accurate models can produce biased recommendations. For example, AI models deployed by hiring platforms to score and rank candidates could exacerbate existing unfairness, because they may be trained on historical data that reflects structural biases.

Such models offer "accurate" recommendations based on their training – they have learned to systematically discriminate against certain groups using patterns from past data. Addressing these biases isn't straightforward. Even if we remove data related to gender, age or race, these models pick up hints from other data points. Variables like the types of verbs used in CVs, schools, addresses, hobbies, and clubs can hint at such personal

details. For example, Amazon had to abandon an AI-hiring algorithm used to filter applicants CVs. Even with direct gender information excluded, the model deduced gender from other CV details and systematically downgraded women's applications.<sup>[10]</sup>

Nevertheless, an important step towards RAI is to exclude data that shouldn't influence a decision when training a model, even if it does not offer a complete fix. For instance, while age may be pertinent when assessing car insurance risk, ethnicity likely should not be.

However, consider a scenario where an AI-model for car insurance risk assessment assigns higher risk scores for individuals based on their country of birth. Does this necessarily imply a bias in the model? Not always. Various factors may correlate both with the country of birth and with heightened risk of driving accidents. This is where Explainable AI plays an important role in upholding the fairness principle of RAI.

Explainable AI tools can reveal important variables driving an AI's recommendations. For decisions with significant implications, it is imperative to

continuously assess these important variables throughout an AI project's lifecycle. The challenge then becomes distinguishing between "reasonable" variables and potentially "biased" ones. A good rule of thumb is: if you can comfortably explain how certain variables should affect a decision

“  
**Explainable AI tools  
 can reveal important  
 variables driving an  
 AI's recommendations.**  
 ”

without the fear of negative publicity, they are likely to be "reasonable". In the car insurance premium context, such variables might include years of driving experience, experience in specific countries like New Zealand, experience of driving on the left/right side of road, age, etc. Conversely, referencing ethnicity or country of birth as reasons for higher insurance premiums can be contentious.

To assess if specific information might introduce biases into AI algorithms, organisations should promote open discussions in a team with diverse backgrounds.

[10] "Robot recruiters: can bias be banished from AI hiring?" The Guardian. Sun 26 Mar 2023. <https://www.theguardian.com/technology/2023/mar/27/robot-recruiters-can-bias-be-banished-from-ai-recruitment-hiring-artificial-intelligence>



## PRIVACY BREACHES

AI models require vast amount of data. Improper collection, storage, or usage of personal or sensitive information, especially without user consent, can lead to privacy breaches, potential litigation, reputation damage and financial losses.

With the rise of large language models (LLMs) like ChatGPT, privacy concerns have intensified. It has been reported that many employees use ChatGPT for work-related tasks without clear organisational guidelines. Users might provide sensitive information during their interactions with ChatGPT. If such data are used for future training or temporarily stored for monitoring the system, there is a risk it could be obtained by other parties.

Consider a few examples.

- *A financial planner enters a client's name and their financial details into ChatGPT to draft a letter. Later, someone could potentially ask ChatGPT about the client's investments if that data were used to train subsequent iterations of ChatGPT.*
- *In April 2023, Samsung employees used ChatGPT to debug confidential source code. As a result, Samsung later cautioned its staff against entering personal or proprietary data into ChatGPT or similar platforms.*
- *If employees use ChatGPT to draft a strategy presentation based on meeting notes, the company's strategy might be exposed to its competitors.*

It is important to note that ChatGPT is one of many LLMs available. Regardless of which LLM you choose, it's crucial to thoroughly understand its privacy policy. While some providers may use your data to improve their models, others could temporarily retain it for operational, debugging, or quality assurance purposes. In all cases, the data can be susceptible to cyber threats. Thus, we emphasise the importance of organisations establishing clear guidelines and promoting responsible LLM usage among their staff.



**We emphasise  
the importance  
of organisations  
establishing clear  
guidelines and promoting  
responsible LLM usage  
among their staff.**



For a formal assessment of AI-related privacy risks, we recommend that organisations seek professional advice. In New Zealand, the Privacy Act 2020 governs how organisations and businesses can collect, store, use and share personal information. In addition, the Privacy Commissioner has issued guidance on Artificial Intelligence and the Information Privacy Principles.<sup>[11]</sup> This guidance offers advice on how to mitigate privacy risks associated with AI use.

[11] <https://www.privacy.org.nz/publications/guidance-resources/ai/>

## VENDOR RISKS

Many organisations, especially SMEs, do not have the capability to build AI tools in-house. SMEs often rely on third-party AI tools and inherently assume vendor-related risks. Even if a third-party AI tool causes harms due to the vendor's fault, it still can damage the reputation of the organisation using it.

We highlight risks arising from using third-party AI tools versus using in-house AI models, accompanied with practical advice:

### *Limited visibility of the AI tool's training data, fairness and accuracy.*

- Perform thorough due diligence on the vendor's RAI practices. Ensure their guidelines and methodologies align with yours and test the tool with your own data.

### *Added concerns about data privacy and security.*

- Ensure contracts clearly define data ownership, access rights and data handling procedures that align with regulations and best RAI practices.
- Encrypt sensitive data before sharing with third parties and only provide essential data.<sup>[12]</sup>

### *Restrictions on AI model customisation and improvements:*

- Custom features may incur additional costs.
- Feedback-based refinements of AI-models might be slower and depends on vendor's schedule.

### *Exit risk:*

- Ensure contingency plans are in place. Whether transitioning by choice or due to a vendor's sudden discontinuation, risks include data retrieval challenges, financial costs and potential service disruptions.

**“  
SMEs must carefully  
evaluate potential  
vendors and remain  
proactive in managing  
third-party AI tools.  
”**

Given these risks, SMEs must carefully evaluate potential vendors and remain proactive in managing third-party AI tools.

[12] When data is encrypted, the actual sensitive content is scrambled and unreadable without a decryption key. However, to ensure usability in various applications, encrypted data can be associated with non-sensitive identifiers. This allows for operations like matching predictions back to specific identities without directly exposing the sensitive content.

# High-level Principles

Implementing a RAI programme begins with establishing the right foundation. For any organisation, the first step is to make sure everyone understands and speaks the same “RAI language”. Two key questions to ask are: What does RAI mean to us? And which fundamental RAI principles should steer our team members in their decisions and actions?

While larger organisations might have the capacity to tailor their RAI principles, SMEs could benefit from adopting a set of well-established RAI guidelines.

In recent years, intergovernmental organisations, governments, industry groups and technology companies have put forward ethical guidelines for AI. While the specifics differ, they typically converge on several key themes. We provide a summary of these common themes in this section.

These RAI principles are deeply rooted in core human values. Therefore, the evolving landscape of AI innovation and regulation should not materially alter these RAI principles. By adhering to these guidelines, organisations equip themselves with the flexibility to meet future AI regulations.

1

## *Human Control*

AI should empower instead of subordinate humans; use human-centric designs and ensure human oversight over AI.

2

## *Transparency & Explainability*

AI should operate transparently. Those impacted by AI-assisted decisions should be informed about the use of AI and how decisions are made. AI-assisted decisions, to the extent possible, need to be explainable.

3

## *Accountability*

People should be held accountable for how AI systems impact individuals and society based on their roles. This includes creating mechanisms for the redress for harms caused by AI.

4

## *Reliability, Safety, Security & Privacy*

AI systems should be technically robust and safe, resistant to misuse and protect privacy.

5

## *Fairness & Inclusiveness*

AI systems should ensure equitable distribution of benefits and costs, prevent bias and discrimination, and promote equal opportunity.

6

## *Sustainability*

AI should be designed and used for the benefit of all (including impacts on the natural environment and biodiversity).



# Translating Principles to Practices

To bring the overarching RAI principles into day-to-day operations, we need individuals to make decisions and act in ways consistent with these principles. This calls for a RAI programme led by a senior executive and buy-in at all levels and across all teams.

Leading tech companies have abundant resources and can devise a comprehensive RAI programme. They often hire AI ethics executives, design customised RAI tools and roll out extensive training.<sup>[13]</sup> Their leadership in RAI is partially stimulated by their experiences of AI failures and the realisation that trust is essential.

An effective RAI programme in large organisations typically has three pillars:

- 1. A governance structure to ensure accountability**
- 2. A set of measurable RAI standards**
- 3. Tools, training and practices**

**“**  
**The debate surrounding an RAI programme is not about its necessity or timing; it is about how to effectively implement it immediately.**  
**”**

In contrast, SMEs might lack the resources or the sense of urgency to closely follow the footsteps of these tech-leaders. However, the proliferation of Generative AI tools and third-party AI solutions make AI risks an imminent concern for all organisations. The debate surrounding an RAI programme is not about its necessity or timing; it is about how to effectively implement it immediately.

---

[13] World Economic Forum issued a series of white papers on Responsible Use of Technology containing case studies of Microsoft, IBM and Salesforce.  
[https://www3.weforum.org/docs/WEF\\_Responsible\\_Use\\_of\\_Technology\\_2021.pdf](https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_2021.pdf)  
<https://www.weforum.org/whitepapers/responsible-use-of-technology-the-ibm-case-study/>  
<https://www.weforum.org/whitepapers/responsible-use-of-technology-the-salesforce-case-study/>

## SMEs can consider the following steps to incorporate RAI principles into its daily operations:

### *Designate one senior executive to be responsible for RAI*

### *Assemble a cross-functional team with diverse backgrounds*

- It's a misconception to view RAI solely as a task for risk management. RAI should be proactive, not reactive. For instance, tech teams can mitigate potential AI harms during the design phase.
- Maintain an up-to-date inventory of all AI applications and ensure regular reviews.
- Identify and evaluate potential risks associated with current AI uses.
- Design RAI standards and put measures in place to monitor the performance of high risk AI usage.
- Incorporate RAI risk assessment and evaluation into existing governance and control routines, or ensure the existing routines are adapted to eliminate RAI blind spots.
- Document risk assessments and evaluations for high risk AI uses.
- Establish a clear reporting line and encourage reporting of RAI-related issues.

### *Encourage RAI by design*

- Encourage employees to participate in webinars, workshops, or short courses related to AI and RAI.
- Practise RAI reasoning using high risk cases (ask, is this right?) in company-wide training.

### *Prioritise transparency*

- Make sure AI-assisted decision-making processes are disclosed to those affected.
- Offer explanations of how AI works.

### *Māori participation*

- Allow Māori participation and partnership when using Māori data and/or applying AI technologies to Māori populations or environments.

# Challenges and Opportunities

## CHALLENGE 1

### Many SMEs are yet to fully embrace digitalisation

New Zealand lags behind many other OECD countries in digitalisation. In 2022, New Zealand is ranked 27th out of 63 countries compared to an overall ranking of 19th in 2018.<sup>[14]</sup> Many SMEs in New Zealand are not fully digitalised, especially when it comes to data-centric IT infrastructure and data analytics. This is particularly true if they operate in traditional manufacturing or service industries instead of being digital natives.

“  
**Many SMEs in New Zealand are not fully digitalised, especially when it comes to data-centric IT infrastructure and data analytics.**  
”

---

[14] IMD. 2022. 'World Digital Competitiveness Rankings'. IMD Business School. 2022.  
<https://www.imd.org/centers/world-competitiveness-center/rankings/world-digitalcompetitiveness/>

## Some prevalent issues include:

### **1** *Siloed Data*

Instead of integrated data lakes or warehouses, many SMEs manage data in isolated spreadsheets and databases, making data retrieval, formatting and analysis cumbersome. Consistent, real-time and high-quality data is essential for training and operating AI models effectively. A data-centric data infrastructure enables rapid development and deployment of new AI projects with low cost. In contrast, with traditional and siloed data infrastructure, only extremely profitable AI projects can justify the high recurring<sup>[15]</sup> cost of data collection and cleaning.

### **2** *Limited digital presence and digital processes*

Limited digital presence and digital processes. Many SMEs have not fully integrated digital tools and platforms into their operations. AI can automate many digital processes, leading to cost savings. Without digitalisation, these processes remain manual, and the full cost-saving benefits of AI cannot be realised.

### **3** *Low productivity in data insights generation*

Even when data is available, SMEs may struggle to generate actionable insights due to a lack of advanced analytical tools or expertise.

### **4** *Culture and decision-making based on hunches*

Business decisions often rely on intuition or past experiences, which may not always be accurate or optimal. A culture accustomed to digital tools is more likely to adopt and harness the benefits of AI.

In essence, while AI is a tool of immense potential, its efficacy is maximised when it's layered upon a digitalised foundation. Without this foundation, SMEs might not be able to reap the full benefit of AI.

[15] To keep machine learning models up to date, new data need to be regularly collected and cleaned to re-train the model.

## CHALLENGE 2

# Lack of Awareness

Many SMEs in New Zealand are unaware of the full extent of their AI usage. This often stems from AI tools being embedded in third-party software and employees using generative AI without specific guidelines. Moreover, there is a general lack of knowledge about the associated risks of AI. Even among larger companies surveyed by Forsyth Barr, only 10% expressed concerns over algorithmic biases in their AI applications.<sup>[16]</sup>

## OPPORTUNITY 1

# Responsible AI by Design

Being a late adopter of full digitalisation and AI provides organisations with the chance to incorporate responsible practices directly into their decision-making processes.

This strategic position empowers SMEs to:

- Integrate responsible practices into their AI plans from the outset. This ensures that ethical considerations are not an afterthought but a foundational element of their AI strategy. This allows them to avoid some of the challenges larger companies face when trying to adapt existing AI systems to meet ethical standards.
- Ensure future adaptability to regulations through a foundational commitment to RAI. Given the dynamic regulatory landscape, SMEs rooted in RAI principles are better positioned, as social values tend to evolve more slowly than technology.

---

[16] "A guide to AI and implications for New Zealand", Forsyth Barr, February 2022



## OPPORTUNITY 2

# Regulatory Flexibility

The global regulatory landscape for AI is marked by significant uncertainty and rapid transformation. The EU is taking the lead in making a comprehensive AI Act effective by late 2023, likely setting a benchmark in AI regulation. In contrast, the United States have only passed state-level and city-level laws addressing specific aspects of AI. In June 2023, the Australian government initiated public consultations on AI regulations.<sup>[17]</sup>

In New Zealand, the situation offers unique flexibility. While there are no AI-specific laws, the Algorithm charter for Aotearoa New Zealand shows a movement towards a commitment to responsible AI use in the public sector.<sup>[18]</sup> The wide acceptance of this charter by government agencies, and the continued development of the charter, underline New Zealand government's proactive approach to the responsible use of algorithms.

It is crucial to note that AI is regulated in New Zealand by existing laws. These include the Privacy Act, the Human Rights Act, the Fair Trading Act, the Harmful Digital Communications Act, among others – estimated to cover around 80% of AI-related issues but with gaps.<sup>[19]</sup>



**In a dynamic global AI landscape, does a smaller economy like New Zealand benefit from crafting bespoke AI regulations, or should it align with the dominant regulatory regimes?**



In New Zealand, the current regulatory focus is on data privacy under the Privacy Act 2020. This focus is highlighted by the Privacy Commissioner's recent guidance in relation to Generative AI and AI in relation to the Privacy Act 2020.

But herein lies the opportunity: In a dynamic global AI landscape, does a smaller economy like New Zealand benefit

from crafting bespoke AI regulations, or should it align with the dominant regulatory regimes? Harmonising New Zealand's AI regulations with more stringent frameworks from larger jurisdictions could serve as a dual strategy – ensuring protection for its citizens while making its AI solutions internationally compliant and thus, more exportable.

[17] <https://consult.industry.gov.au/supporting-responsible-ai>

[18] <https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/>

[19] "Reimagining Regulation for the Age of AI: New Zealand Pilot Project", Word Economic Forum, June 2022

## OPPORTUNITY 3

# Te Tiriti o Waitangi (Treaty of Waitangi) and AI

### Leadership in Cultural Appropriate AI Regulation

Knowledge from Indigenous nations can help inform culturally appropriate AI regulation through Indigenous knowledge and practices. In Aotearoa New Zealand, the incorporation of Māori cultural values into AI deployment is paramount.<sup>[20]</sup> AI systems must respect principles embracing a Te Tiriti lens for any AI algorithmic development, deployment that uses or produces Māori data, or that makes decisions about Māori and should adopt in conjunction with Māori Data sovereignty principles.<sup>[21]</sup>



**AI systems must respect principles embracing a Te Tiriti lens for any AI algorithmic development, deployment that uses or produces Māori data, or that makes decisions about Māori and should adopt in conjunction with Māori Data sovereignty principles.**



---

[20] <https://link.springer.com/article/10.1007/s00146-023-01636-x#Sec3>

[21] Recently, Māori Data Sovereignty principles have been extended to issues regarding Governance, privacy, and algorithms (see a list of references in Appendix B).

## An example of recent ethical principles developed by Dr Karaitiana Taiuru, for AI use follows<sup>[22]</sup>:

1

### *Tino Rangatiratanga*

All AI systems will embed Māori leadership, decision-making and governance at all levels of the systems life cycle from inception, design, release to monitoring.

3

### *Active Protection*

Requires AI developers and product owners to act to the fullest extent practicable, to achieve equitable outcomes for Māori. This includes ensuring that it, its agents are well-informed on the extent, and nature, of both Māori outcomes and efforts to achieve Māori equity. Ensuring Free, Prior, and Informed Consent (FPIC) is required for the use of Māori data in AI development, with robust procedures in place to prevent biases or predictions that stigmatise or harm Māori.

5

### *Mana Motuhake*

Requires that tikanga are followed throughout the whole AI development and deployment cycles, with Māori deciding what data and data uses are controlled or allowed.

2

### *Equity*

AI systems will achieve equity outcomes for Māori (individuals and collectively) across the life course and contribute to Māori development. This involves businesses and employees who are accountable to Māori in how AI models are used with Māori data and outputs that impact Māori individually and collectively, and in the active building of capacity of the Māori AI and tech workforce.

4

### *Mana Whakahaere*

Effective and appropriate stewardship or kaitiakitanga over AI systems is required. It is recognised that Māori Data is a Taonga and subject to Māori Data Sovereignty principles determined by Te Tiriti. A deep understanding of the source and intended use of data is required, so that it is not repurposed without permission or in a way that will diminish the mana of Māori. Shared and recognised Intellectual Property rights are proportioned to Māori who contribute Māori Data to the AI.

6

### *Tapu/Noa; Cultural safe practices*

No AI will be culturally unsafe or break the rules of Tapu and Noa.

[22] Source: Karaitiana Taiuru; <http://www.taiuru.maori.nz/AI-Principles>

# Recommendations for New Zealand SMEs

In this section, we answer common questions on AI adoption and risks.

## ***Q1. How can I find out what AI applications I am currently using and their RAI practices?***

- Review the software and tools your business currently utilises. Check product descriptions, user manuals, and any associated documentation for terms like “machine learning”, “automation”, “predictive analytics”, or “AI”.
- Engage with your software vendors or service providers directly can shed light on whether their solutions incorporate AI and if they adhere to responsible AI principles.

## ***Q2. How do I assess the risk related to an AI application without internal AI expertise and with a limited budget?***

- Start by researching best practices and guidelines on AI ethics and risk from reputable sources and industry associations.
- Attend AI-related events where you can ask questions and learn from shared experiences.
- Partner with local academic institutions.
- Consider contracting a consultant with RAI expertise for a brief risk assessment.

## ***Q3. How can I increase my digital presence in the market responsibly?***

- Create a transparent and user-centric website, ensuring that any AI-driven tools or features are responsibly implemented and communicated.
- Optimise your online presence using search engine optimisation (SEO) techniques that respect user privacy.
- Follow your RAI procedures and document the related due diligence exercises when choosing SEO providers.

## ***Q4. How can I improve efficiencies by streamlining and digitising more processes responsibly?***

- Begin by pinpointing repetitive or time-intensive tasks within your operations. As you introduce tools such as digital invoicing, online scheduling, or workflow automation platforms, ensure they uphold data privacy, fairness, and transparency.
- Document RAI efforts and measures in the digitalisation process.

**Q5. How can I build or revamp my data infrastructure to support my business needs efficiently and responsibly?**

- If you encounter slow data retrieval or frequent data inconsistencies, it may signal a need for an upgrade.
- When transitioning to enhanced data management platforms or cloud solutions, choose providers that uphold data privacy, user consent, and transparent data usage policies.

“

**Cultivate a data-literate culture within your team, emphasising the importance of making decisions grounded in data.**

”

**Q6. How can I make more data-informed decisions ethically?**

- Start by centralising and organising your data sources. Ensure that data collection and handling respect user consent and adhere to privacy standards.
- Cultivate a data-literate culture within your team, emphasising the importance of making decisions grounded in data. In the process of making data-driven decisions, incorporate AI ethics questions to ensure that potential implications are considered.

**Q7. How can we improve AI knowledge within the organisation, given the shortage of AI talent?**

- Leverage external expertise. Approach local universities to explore partnerships. Engage with local tech communities to find experts willing to share their insights. Encourage staff to attend webinars that provide overviews of AI trends and applications.
- Consider allocating a small budget for in-house training. Engage an AI expert to conduct a few hours of tailored training periodically, focusing on practical knowledge directly applicable to your business.
- Offer reimbursement for training courses aimed at upskilling employees.

**Q8 How should I proceed if I am using Maori data in the AI?**

- Engage in appropriate Maori consultation.
- Build meaningful partnerships with Iwi and encourage participation to allow Iwi to actively protect their taonga (including data).
- Encourage all staff to be culturally aware - not just technically literate, but socially and culturally literate.



# Government Policy Recommendations

## 1 More targeted government support for SMEs on their digitalisation and AI journey.<sup>[23]</sup>

- Launch educational workshops and webinars tailored for SMEs to shed light on the benefits and potential risks of AI. Topics should include understanding AI's capabilities, best practices for integration, and ethical considerations.
- Introduce voucher programmes to subsidise related upskilling and executive education at approved institutes. Topics could include digital transformation, data analytics, machine learning and AI management.

## 2 Address gaps in AI regulations.

- There are notable gaps in current AI regulations that require attention.<sup>[24]</sup>
- We recommend wide consultation spearheaded by a dedicated government agency to shape the future landscape of AI regulations.

---

[23] "What works for digitalisation? A review of international success" by New Zealand Institute of Economic Research March 2023. Table 1 "The short list of high-priority actions for the government"

[24] "Why is regulating AI such a challenge?" Office of the Prime Minister's Chief Science Adviser. <https://www.pmcsa.ac.nz/2023/07/13/why-is-regulating-ai-such-a-challenge/>

## APPENDIX A

# Essential AI Concepts Explained

### ARTIFICIAL INTELLIGENCE (AI)

A branch of computer science focused on creating systems capable of performing tasks that would otherwise require human intelligence.

### MACHINE LEARNING (ML)

A subset of AI where algorithms learn from and make predictions or decisions based on data without being explicitly programmed.

### GENERATIVE AI

An AI approach that allows computers to generate new content, such as images, music, or text, by training on a large dataset.

### LLM (LARGE LANGUAGE MODELS)

AI models designed to understand, generate, or complete pieces of text. They are trained on vast amounts of text data to predict the next word in a sequence, giving them the ability to generate coherent sentences.

### TRAINING

The process by which an AI or ML model learns from a dataset. It involves feeding data into the model and adjusting the model's internal parameters to minimize error and improve its predictive or classification ability.

## APPENDIX B

# Resources for Definitions & Principles

- Accenture: <https://www.accenture.com/us-en/blogs/business-functions-blog/responsible-ai-finance>
- Boston Consulting Group and MIT Sloan: Elizabeth M. Renieris, David Kiron, and Steven Mills, “To Be a Responsible AI Leader, Focus on Being Responsible,” September 2022
- Deloitte: <https://www2.deloitte.com/uk/en/pages/risk/solutions/ai-risk.html>
- McKinsey: 10 Responsible AI Guiding Principles  
<https://www.mckinsey.com/capabilities/quantumblack/how-we-help-clients/generative-ai/responsible-ai-principles>
- PwC: <https://www.pwc.co.nz/services/risk-services/what-is-responsible-ai.html>
- European Union’s Ethics Guidelines for Trustworthy AI  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- OECD AI Principles <https://oecd.ai/en/ai-principles>
- UK Policy Paper: AI regulation: a pro-innovation approach  
<https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>
- Brown, P.T., Wilson, D., West, K., Escott, K., Basabas, K., Ritchie, B., Lucas, D., Taia, I, Kusabs, N., Keegan, T.T. (2023): Māori algorithmic sovereignty: ideas, data, and use. <https://arxiv.org/abs/2311.15473>.
- Kukutai, T., Cassim, S., Clark, V., Jones, N., Mika, J., Morar, R., Muru-Lanning, M., Pouwhare, R., Teague, V., Tuffery Huria, L., Watts, D., Sterling, R. (2023a). Māori data sovereignty and privacy. Tikanga in Technology discussion paper. Te Ngira Institute for Population Research, Hamilton.
- Kukutai, T., Campbell-Kamariera, K., Mead, A., Mikaere, K., Moses, C., Whitehead, J., Cormack, D. (2023b). Māori data Governance model. Te Kāhui Raraunga.
- West, K., Wilson, D., Thompson A., & Hudson, M. (2020). Māori Perspectives on Trust and Automated Decision-Making. Report for the Digital Council.



The background is a vibrant, textured blue marbled pattern, resembling watercolor or ink wash. The colors range from deep, dark blues to lighter, almost white highlights, creating a sense of depth and movement. The texture is organic and fluid, with irregular, vein-like patterns. In the center, the letters 'AI' are prominently displayed in a clean, white, sans-serif font. The 'A' and 'I' are connected at the top, and the overall appearance is modern and minimalist.

AI