



Does the Responsible AI Guidance Adequately Protect Māori in Predictive Workplace Safety Regulation?

Vol 3; edition 2; July 2026

Author: Nicola Knobel

Email: nicolaknobel@icloud.com

DOI: <https://doi.org/10.26686/nzjhsp.v3i2.10605>

Abstract

Artificial intelligence (AI) is increasingly embedded in the design and delivery of public services in Aotearoa New Zealand, particularly in high-risk regulatory environments such as workplace health and safety. This paper evaluates the sufficiency of the Responsible AI Guidance for the Public Service in upholding Te Tiriti o Waitangi, applying three intersecting frameworks: Critical Tiriti Analysis, really responsive regulation theory, and post-colonial legal analysis. Drawing on the Woolworths warehouse strike in Australia and WorkSafe New Zealand's engagement with AI-enabled tools through enforceable undertakings, the paper critiques the Guidance's reliance on non-binding, soft-law principles. Across all five Critical Tiriti Analysis criteria, the Guidance scores either Silent or Poor. It lacks enforceable protections, Māori governance structures, and procedural fairness mechanisms. The paper proposes a series of reforms to embed co-governance, Māori Data Sovereignty, and legally mandated safeguards into public sector AI infrastructure.

Keywords: artificial intelligence, Te Tiriti o Waitangi, Māori Data Sovereignty, workplace health and safety, algorithmic bias, soft law, really responsive regulation

I Introduction

AI is reshaping how public services are delivered in Aotearoa New Zealand, including in high-risk regulatory contexts such as workplace health and safety. Predictive analytics, wearable surveillance technologies, and algorithmic triage systems are being introduced to improve harm prevention, monitor compliance, and prioritise enforcement. While such tools promise efficiency and proactive intervention, their use raises significant legal, ethical, and cultural concerns, particularly for Māori workers, who are overrepresented in high-risk industries and often excluded from decision-making about the tools that govern them.

This paper investigates whether the Responsible AI Guidance for the Public Service (Department of Internal Affairs, 2022) provides sufficient protection for Māori in the context of AI-supported safety and enforcement systems. While the Guidance makes rhetorical commitments to fairness, transparency, and Te Tiriti o Waitangi, it remains non-binding, lacks enforcement mechanisms, and offers no clear pathway for Māori co-governance, data sovereignty, or structural redress. As such, it leaves open the risk that AI tools may reinforce, rather than challenge, systemic inequities while appearing neutral or objective.

To assess these concerns, the paper applies three intersecting analytical frameworks: the Critical Tiriti Analysis (CTA) tool developed by Came and Baker (2022), Baldwin and Black's (2008) theory of really responsive regulation, and a post-colonial legal lens drawing on Smith (2012). These frameworks are applied to a critical reading of the Guidance, contextualised with the Health and Safety at Work Act 2015 (HSWA), the Public Service Act 2020, and Indigenous data governance principles articulated by Te Mana Raraunga (2018).

II Background

A AI and the Public Sector

New Zealand's public service is governed by the Public Service Act 2020, which affirms the Crown's commitment to Te Tiriti o Waitangi and the expectation that agencies will actively partner with Māori. A 2024 government survey identified 108 current or planned AI use cases across 37 agencies, with

applications spanning trend analysis, service optimisation, and research synthesis (New Zealand Government, 2024). The Responsible AI Guidance for the Public Service, issued by the Department of Internal Affairs in 2022, sets out six high-level ethical principles: fairness and human-centredness; transparency and explainability; privacy, ethics, and security; accountability and oversight; inclusivity and Treaty of Waitangi commitment; and robustness and reliability. The Guidance was developed alongside the Algorithm Charter for Aotearoa New Zealand (New Zealand Government, 2020). Both instruments reflect a soft-law approach, remaining non-binding and leaving implementation at the discretion of individual agencies (Dentons, 2024). A 2025 framework under the National AI Strategy similarly favours voluntary standards over enforceable protections (Blythe & Taylor, 2025).

B The Limits of Soft Law

Baldwin and Black's (2008) theory of really responsive regulation argues that effective regulatory design must adapt to social context, institutional capacity, and power imbalances, particularly where the scale of potential harm is significant and unevenly distributed. In the context of public sector AI, a responsive approach would require more than high-level principles. It would demand Treaty-based co-governance, algorithmic impact assessments, and enforceable mechanisms tailored to protect those most at risk.

The current soft-law approach creates real gaps. A 2025 Public Service Association survey found that 55% of public sector members reported AI use in their workplaces, yet only 12% had access to a formal AI policy (Radio New Zealand, 2025). Māori engagement in AI development or governance was almost entirely absent. Without enforceable safeguards, these technologies risk becoming tools of regulatory overreach, amplifying surveillance and exclusion under the guise of innovation (Prime Minister's Chief Science Advisor, 2023).

C WorkSafe New Zealand and AI

WorkSafe NZ was established in 2013 following the Pike River Mine tragedy and is governed by the HSWA 2015. Its core mandate encompasses education, support, and enforcement, with statutory powers of inspection, investigation, and prosecution directed toward employers. WorkSafe is also expected to uphold equity and Te Tiriti o Waitangi obligations as part of its mandate (WorkSafe New Zealand, 2022).

Māori workers experience disproportionately higher rates of work-related harm. WorkSafe NZ data shows that Māori workers have a serious injury rate 31% higher than non-Māori workers, with Māori lodging 103 claims per 1,000 full-time equivalents compared to 83 per 1,000 for European workers (WorkSafe New Zealand, 2023). These disparities are acute in forestry, where Māori constitute 34% of the workforce and 18 Māori forestry workers died in Te Tairāwhiti alone between 2003 and 2022.

While there is no formal public statement about WorkSafe's use of AI, the agency has endorsed AI-enabled tools through enforceable undertakings with Ecostore (WorkSafe New Zealand, 2023a) and Claymark (WorkSafe New Zealand, 2023b), introducing predictive analytics, real-time monitoring, and wearable sensor technologies. These cases illustrate how algorithmic tools can become embedded in regulatory systems rapidly, and without meaningful Māori oversight.

D The Woolworths Case Study

In late 2024, over 1,500 Woolworths workers across Australia launched a 17-day strike in response to an AI-powered Coaching and Productivity Framework. The system tracked workers' movements via wearable scanners and warehouse software, generating real-time performance metrics that triggered coaching or disciplinary action for deviations from productivity targets, including slowdowns attributable to safe lifting practices or legally required breaks (Barnes, 2024; Sadler, 2024). The AI model could not distinguish between inefficiency and safety-conscious behaviour.

The dispute (Woolworths Group Ltd v United Workers Union, 2024) was resolved with a negotiated settlement that removed automatic disciplinary consequences tied to AI performance scores, secured a 12.84% wage increase over three years, and included commitments to consult workers on future technology use. The strike cost an estimated AUD \$140 million in lost sales. Woolworths' ongoing use of the system in a softened form underscored how such tools become entrenched even after contested rollout.

For Aotearoa, this case is directly relevant. The Australian Senate Select Committee on AI (2024) warned that predictive AI systems trained on historical data often reflect systemic bias and disproportionately affect vulnerable and Indigenous communities. Cebulla et al. (2023) similarly

identify that AI in workplace safety can amplify risk through biased assumptions, particularly when used for surveillance and predictive profiling. In industries where Māori workers are heavily represented, the deployment of equivalent systems without Treaty-aligned governance raises urgent concerns about procedural fairness and accountability.

III Analytical Frameworks

This paper applies three intersecting lenses. First, Came and Baker's (2022) Critical Tiriti Analysis (CTA) framework evaluates whether tino rangatiratanga, Māori governance, and equitable protections are structurally embedded in the Guidance, moving beyond surface-level Treaty references to ask whose worldviews shape policy and whether redistribution and redress are built into implementation. Second, Baldwin and Black's (2008) really responsive regulation theory examines whether non-binding guidance is fit for purpose in high-risk technological contexts. Third, a post-colonial legal lens, drawing on Smith (2012), explores how AI may entrench colonial dynamics through digital surveillance, data extraction, and exclusion from power, recognising that Aotearoa's legal institutions were built in ways that privileged Crown authority over Māori tino rangatiratanga.

IV Critical Tiriti Analysis of the Guidance

The CTA framework uses a five-point Likert scale: Silent (no reference), Poor (tokenistic), Fair (limited engagement), Good (clear inclusion and equity focus), and Excellent (Māori leadership is central). Table 1 applies this framework across five evaluative criteria.

Table 1. Critical Tiriti Analysis of the Responsible AI Guidance for the Public Service

CTA Question	What to look for	Rating	Finding
Whose worldviews shape the Guidance?	Are Māori worldviews, tikanga, and mātauranga Māori central to design?	Poor	Built on Western ethical paradigms. No discussion of how Māori define harm, fairness, or risk in AI contexts.
Are Māori positioned as co-governors?	Were Māori collectives meaningfully involved in co-design and shared decision-making?	Poor	No evidence of substantive Māori involvement in development. Reflects consultation rather than genuine partnership.
Is tino rangatiratanga reflected through data sovereignty?	Do Māori have control over data about them, including through Māori Data Sovereignty mechanisms?	Silent	No acknowledgement of Māori Data Sovereignty. No reference to Te Mana Raraunga. No collective consent requirements.
Does the Guidance include equity, redress, or redistribution?	Are there mechanisms to acknowledge systemic inequity and redistribute power to achieve equitable outcomes?	Poor	Fairness is referenced but no mandatory impact assessments, no redress mechanisms, no obligations to monitor discriminatory outcomes.
Are there monitoring and accountability mechanisms?	Does the policy include built-in accountability to Māori, with enforceable oversight if breaches occur?	Silent	Non-binding. No Māori oversight body. No reporting obligations. No enforcement if Te Tiriti is breached or algorithmic harm results.

Note. CTA ratings based on Came and Baker (2022) five-point scale.

Taken together, this analysis confirms that the Guidance fails to meet Treaty-based expectations of partnership, protection, and participation. It does not embed Māori governance, uphold tino rangatiratanga, or provide mechanisms for Māori to shape, scrutinise, or contest AI systems that may have material impacts on their lives, safety, or autonomy. This reflects a broader pattern of Crown policymaking that acknowledges Te Tiriti in form, but not in function (Came, 2020).

V Māori Data Sovereignty and Algorithmic Bias

Māori Data Sovereignty refers to the right of Māori to control, access, and govern data about themselves, their whānau, hapū, and iwi, in accordance with Te Tiriti and tikanga Māori (Te Mana

Raraunga, 2018). It is a collective right grounded in the understanding that data is a taonga linked to whakapapa and tino rangatiratanga. The Guidance is entirely silent on these principles.

This silence creates concrete risks. AI tools used in the public sector are frequently trained on historical datasets that were not designed with Māori self-determination in mind. Where past enforcement actions disproportionately targeted Māori businesses or workers, a predictive model may conclude that Māori sites are higher-risk. Where workplace incidents involving Māori are more likely to be recorded or escalated, the data may associate Māori presence with non-compliance. Where Māori workers under-report safety issues due to cultural or power dynamics, the model may wrongly infer that Māori workplaces are lower risk and fail to intervene. Without critical Māori oversight, AI tools risk encoding racism into enforcement in ways that are difficult to detect and harder to dismantle (Wichmann, 2023).

Most AI tools also operate using Western, individualistic, risk-reduction frameworks that ignore tikanga Māori. An algorithm might assess non-compliance based on deviation from a scripted safety protocol, without recognising that kaupapa Māori approaches may be equally effective but undocumented. Risk defined solely in physical or economic terms excludes mana, whakapapa, and relational wellbeing. Without Māori Data Sovereignty, there is no mechanism to contest these framings, as control over data flows, AI logic, and decision points remains with the Crown (Roberts, 2023).

VI Gaps in Worker Protections

The Guidance is silent on AI systems used for regulatory enforcement or workplace surveillance. It does not address how its principles of fairness, accountability, and explainability apply when AI is used to monitor worker performance, assess risk, or guide compliance actions. This stands in contrast to the European Union's AI Act (Artificial Intelligence Act, Regulation (EU) 2024/1258), which explicitly categorises workplace surveillance and enforcement applications as high-risk (De Freitas et al., 2024). New Zealand's Guidance does not differentiate between back-end automation and high-stakes, rights-impacting systems.

The Guidance also fails to address procedural fairness. There is no obligation for agencies to inform individuals when an algorithm has been used in a decision about them, provide clear explanations, or create pathways for challenge. A worker flagged by an AI system as non-compliant or high-risk may never know how the determination was made, on what data it was based, or how to contest it. This is inconsistent with natural justice principles and existing administrative law protections under the New Zealand Bill of Rights Act 1990, section 27 (Brownlie, 2020).

VII Pathways Forward

To move beyond aspirational language and ensure AI use in the public sector upholds Te Tiriti o Waitangi, Aotearoa must adopt stronger, enforceable mechanisms. Four interlinked reforms are proposed.

A Enact Binding Legal Duties for High-Risk AI

Drawing from the EU AI Act, AI systems used to inform or automate decisions in workplace safety enforcement, policing, and welfare should be classified as high-risk, with corresponding legal obligations including mandatory algorithmic impact assessments, transparency around data sources and system design, and independent review requirements. Such duties could be incorporated into the HSWA 2015, with AI-specific impact assessments required under the existing duty to eliminate or minimise foreseeable risk (s 36), and into the Public Service Act 2020 via expanded stewardship obligations under s 14.

B Mandate Māori Co-Governance

Te Tiriti requires more than consultation. Any public sector use of AI must embed Māori governance at every stage, including design, implementation, oversight, and review. This should involve Māori-led AI Ethics Boards to review proposed public sector systems involving Māori data, legislative recognition of Māori Data Sovereignty including collective consent rights for iwi and hapū, and adoption of Te Mana Raraunga principles as a required foundation for state data projects. Canada's First Nations Information Governance Centre and the OCAP® principles (Ownership, Control, Access, and Possession) offer a structural model that could be adapted within a Te Tiriti framework (Seokki, 2023). Public Service Act 2020 ss 14 and 16 provide an existing legislative pathway for embedding specific Māori governance requirements.

C Procedural Rights for Affected Workers

Where AI informs regulatory or employment-related decisions, individuals must be guaranteed the right to know when an algorithm has been used in a decision about them, access to clear and non-technical explanations, the right to contest or appeal AI-informed decisions, and obligations on agencies to monitor and report on equity outcomes. For Māori workers, procedural fairness must also encompass cultural safety, reflecting manaakitanga and the right to be treated with dignity. These protections align with the Crown's duty of protection under Te Tiriti and with natural justice obligations under the New Zealand Bill of Rights Act 1990 (Stats NZ, 2023).

D Establish Māori-Led Audit and Accountability Infrastructure

The current lack of enforceability requires dedicated AI oversight. This could involve establishing a Commissioner for Algorithmic Systems with investigatory powers and binding rulings, amending the Privacy Act 2020 to include collective Māori data rights beyond individual privacy (aligning with information privacy principles 1 and 3), and requiring all public agencies to report annually on AI systems used, equity impacts, and Māori engagement. Independent Māori review panels should conduct pre-deployment Māori impact assessments modelled on cultural impact assessments, with ongoing cultural audits of systems in use. Just as major infrastructure projects require iwi-led assessment, algorithmic systems used for regulation and surveillance warrant the same level of scrutiny (Te Mana Raraunga, 2018; Knott & Taiuru, 2024).

VIII Conclusion

This paper has evaluated the Guidance for the Public Service through a Te Tiriti o Waitangi lens, using Critical Tiriti Analysis to assess whether Māori rights, tino rangatiratanga, and structural equity are meaningfully embedded in current AI governance. The findings are unambiguous: across all five CTA criteria, the Guidance scores either Silent or Poor. Māori worldviews are not embedded in its design. Co-governance and tino rangatiratanga are absent. There are no legally mandated protections for Māori data or decision-making authority. Equity and redress mechanisms are not required. There is no external oversight, enforcement, or monitoring, rendering the Guidance effectively voluntary and leaving Māori communities exposed to algorithmic harm without procedural safeguards or collective control.

The Woolworths strike and the AI-driven safety tools endorsed through enforceable undertakings at Ecostore and Claymark demonstrate that algorithmic systems can rapidly become entrenched in workplace regulation. Without clear legal parameters and Tiriti-aligned governance, agencies like WorkSafe may unintentionally replicate the same harms under the guise of innovation, particularly in industries with high Māori and Pacific workforce representation.

A balanced, Treaty-honouring approach to AI governance is both possible and necessary. The reforms proposed here would move Aotearoa from soft law and ethical guidance toward a rights-based, Tiriti-aligned regulatory model. Public trust in government use of AI will not be built on principles alone. It requires enforceable protections, transparent systems, and a power-sharing approach that recognises Indigenous rights not only as a matter of justice, but as a constitutional obligation. The regulatory choices made now will shape not just the future of AI, but the nature of Crown-Māori relations in the digital age.

Use of artificial intelligence: no artificial intelligence was used in researching or writing this article.

References

Artificial Intelligence Act, Regulation (EU) 2024/1258.

Australian Government, Department of Industry, Science, Energy and Resources. (2019). Artificial intelligence ethics framework. <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework>

Australian Senate. (2024). Select Committee on Adopting Artificial Intelligence (AI): Final report. Parliament of Australia.

Baldwin, R., & Black, J. (2008). Really responsive regulation. *Modern Law Review*, 71(1), 59–94. <https://doi.org/10.1111/j.1468-2230.2008.00681.x>

- Barnes, T. (2024, December 4). The dispute causing empty shelves at Woolies is a test case for companies using AI and automation. The Conversation. <https://www.business-humanrights.org/en/latest-news/australia-woolworths-productivity-framework-involving-surveillance-tech-a-major-factor-in-stalled-bargaining-negotiations/>
- Blythe, L., & Taylor, L. (2025, January 31). Release of public sector AI framework. Russell McVeagh.
- Brownlie, E. (2020). Encoding inequality: The case for greater regulation of artificial intelligence and automated decision-making in New Zealand. *Victoria University of Wellington Law Review*, 51(1), 1–40. <https://doi.org/10.26686/vuwlr.v51i1.6338>
- Came, H. (2020). Critical Tiriti analysis tool: For engaging with and developing Treaty-based policy. <https://trc.org.nz/sites/trc.org.nz/files/Critical%20Tiriti%20Analysis%20tool.pdf>
- Came, H., & Baker, M. (2022). Critical Tiriti analysis: A tool for engaging with Treaty-based policy analysis in Aotearoa. *Critical Policy Studies*, 16(2), 192–209. <https://doi.org/10.1080/19460171.2021.1924262>
- Cebulla, A., Nguyen, Q., & Duong, L. (2023). Applying ethics to AI in the workplace: The design of a scorecard for Australian workplace health and safety. *AI & Society*, 38, 919–934. <https://doi.org/10.1007/s00146-022-01379-7>
- De Freitas, J., Uğuralp, A. K., Uğuralp, Z. O., & Puntoni, S. (2024). Chatbots and mental health: Insights into the safety of generative AI. *Journal of Consumer Psychology*, 34(3), 481–491. <https://doi.org/10.1002/jcpy.1393>
- Dentons. (2024, August 14). The imitation game: NZ's approach to regulating AI. <https://www.dentons.co.nz/en/insights/articles/2024/august/14/the-imitation-game-nzs-approach-to-regulating-ai>
- Department of Internal Affairs. (2022). Responsible AI guidance for the public service. <https://www.digital.govt.nz/dmsdocument/227~responsible-ai-guidance/html>
- Fair Work Commission. (2024). *Woolworths Group Limited v United Workers Union* [2024] FWC 3428. Health and Safety at Work Act 2015 (NZ).
- Hudson, M., Farrar, D., & McLean, L. (2024). Māori algorithmic sovereignty: Idea, principles, and use. *Data Science Journal*, 23. <https://doi.org/10.5334/dsj-2024-001>
- Knott, A., & Taiuru, K. (2024). Has the time come for a register of AI systems used by government agencies? Public Health Communication Centre. <https://www.phcc.org.nz/briefing/has-time-come-register-ai-systems-used-government-agencies>
- New Zealand Bill of Rights Act 1990 (NZ).
- New Zealand Government. (2020). Algorithm charter for Aotearoa New Zealand. <https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/>
- New Zealand Government. (2024). Full results: 2024 cross-agency survey for artificial intelligence (AI) use cases. Digital.govt.nz
- Prime Minister's Chief Science Advisor. (2023). AI regulation A3 – July 2023. <https://www.dpmc.govt.nz/sites/default/files/2024-01/PMCSA-23-07-03-V3-Regulatory-landscape-v3.pdf>
- Privacy Act 2020 (NZ).
- Public Service Act 2020 (NZ).
- Radio New Zealand. (2025, April 14). AI widely used in public sector, survey finds.
- Roberts, J. S., & Montoya, L. N. (2023). In consideration of Indigenous data sovereignty: Data mining as a colonial practice (arXiv:2309.10215). arXiv. <https://arxiv.org/abs/2309.10215>
- Sadler, D. (2024, December 10). Woolies workers claim victory over AI surveillance. *Information Age*.
- Seokki, L. (2023). Towards an international regulatory framework for AI safety: Lessons from the IAEA's nuclear safety regulations. *Humanities and Social Sciences Communications*, 11, Article 107. <https://doi.org/10.1057/s41599-024-02643-5>

- Smith, L. T. (2012). Decolonizing methodologies: Research and Indigenous peoples (2nd ed.). Zed Books.
- Stats NZ. (2023). Algorithm impact assessment user guide. <https://www.stats.govt.nz/assets/Uploads/Algorithm-impact-assessment-user-guide.pdf>
- Te Mana Raraunga. (2018, October). Principles of Māori data sovereignty. <https://www.temanararaunga.maori.nz/tutohinga>
- Taiuru, K. (2024). Artificial intelligence regulation from a Māori perspective. <https://taiuru.maori.nz>
- Waitangi Tribunal. (2011). Ko Aotearoa tēnei: A report into claims concerning New Zealand law and policy affecting Māori culture and identity (Wai 262). https://forms.justice.govt.nz/search/Documents/WT/wt_DOC_68356741/KoAotearoaTeneiTT2W.pdf
- Wichmann, M. (2023). The impacts of artificial intelligence on workplace health and safety in Aotearoa New Zealand [Master's thesis, Te Herenga Waka – Victoria University of Wellington].
- WorkSafe New Zealand. (2022). Statement of intent 2022/23–2025/26. <https://www.worksafe.govt.nz/about-us/publications/statement-of-intent/>
- WorkSafe New Zealand. (2023). Improving the equity gap for Māori. <https://www.worksafe.govt.nz>
- WorkSafe New Zealand. (2023a, May 17). Enforceable undertaking: Ecostore Company Limited. <https://www.worksafe.govt.nz/laws-and-regulations/enforceable-undertakings/ecostore/>
- WorkSafe New Zealand. (2023b, December 12). Enforceable undertaking: Claymark Group Limited Partnership. <https://www.worksafe.govt.nz/laws-and-regulations/enforceable-undertakings/claymark/>