

Recommendation: Prepare a New Zealand AI Safety Strategy

Key points

- While New Zealand has an AI Strategy for adoption and investment, it has no stated plan for managing the risks of increasingly capable AI.
- We recommend the government prepare a national strategy for managing the risks posed by AI, like New Zealand has for cyber security and disaster resilience.
- For New Zealanders to adopt AI with confidence, the risks need visible, credible management.

Why it matters

New Zealand's [AI Strategy](#), released in July 2025, is an economic strategy. It is designed to reduce barriers to adoption and give businesses confidence to invest. It is not, and was never intended to be, a plan for managing the risks of increasingly capable AI.

New Zealand has no official stated position on AI safety: no assessment of the risks we face, no assignment of responsibilities across government, and no plan. When issues arise, from AI-enabled cyberattacks to deepfakes disrupting our election, each agency will respond ad hoc, or not at all.

Surveys consistently show low public trust in AI and strong support for the government taking risks seriously. For example, [a global KPMG study](#) found that only 34 percent of New Zealanders are willing to trust AI, 81 percent believe regulation of AI is required and 89 percent want laws to combat AI-generated misinformation.

The government's current AI strategy identifies concerns about the risks as a barrier to uptake. But without a clear plan to manage the current and emerging risks of AI, New Zealand will remain vulnerable given the pace of AI development.

What we're proposing

We recommend a Cabinet-endorsed national AI Safety Strategy (a companion to the existing AI Strategy), that sets out:

- The government's assessment of the risks advanced AI poses to New Zealand, now and over the coming decade.
- A vision and clear objectives for what New Zealand will do about them.
- Roles and responsibilities across departments, regulators and the intelligence community.
- A prioritised work programme, and how New Zealand will engage internationally.
- A regular review cycle, so the strategy keeps pace with the technology.

The opportunity for New Zealand

An AI Safety Strategy is a natural complement to New Zealand's existing AI strategy. The Government wants New Zealanders to adopt AI with confidence, which requires visible, credible management of the risks.

A strategy would give the Government a shared framework, so decisions on AI safety are coherent and responsibilities are clear across departments and regulators.

It would position New Zealand as a serious, trustworthy participant in international AI governance. This is increasingly valuable as access to advanced AI models could be shaped by trust between governments.

How it could be done

New Zealand already does this for comparable challenges: we have national strategies or action plans for cyber security, biosecurity and disaster resilience. Like these challenges, AI safety is serious, evolving and cross-cutting, and warrants the same treatment.

The work could be led by the Department of the Prime Minister and Cabinet (given the national security dimensions) or the Ministry of Business, Innovation and Employment (which owns the AI Strategy), developed with public consultation, and published within 12 months.

The cost is modest: a strategy requires policy effort, but not major new spending.

This is not a proposal for AI regulation. A strategy is a statement of the government's assessment, intentions and priorities – the same instrument New Zealand already uses for cyber security and disaster resilience. The specifics of any regulatory intervention would be considered on their own merits through normal policy and legislative processes.

International precedent

Comparable countries increasingly pair their AI growth agendas with dedicated safety frameworks and institutions. Australia has now embedded its AI safety and AI growth thinking into its institutions – establishing its AI Safety Institute and a broader national framework including centralised government capability in an [Office of AI](#) within the Department of the Prime Minister and Cabinet.

New Zealand was [the last OECD country to publish an AI strategy](#) and it is yet to establish an AI safety institute. New Zealand cannot afford the same delay on safety: other countries are already assessing and managing these risks while New Zealand has yet to start.

Frequently asked questions

Don't we already have an AI strategy? Yes. And it is explicitly an adoption and investment strategy, built around a light-touch approach and creating a regulatory environment that supports AI adoption. It does not set out a clear plan for how the government will manage the risks. The two strategies address different questions and would work together.

Won't a safety strategy deter investment? The opposite. Uncertainty and risk deter investment, and a clear statement of the government's approach to managing AI risks reduces both. A strategy is not regulation; it creates no obligations on business.

How much will it cost? Very little. This is primarily a policy exercise using existing policy resources.