

Recommendation: Establish a New Zealand AI Safety Institute

Key points

- New Zealand has no dedicated government capability to monitor the risks of advanced AI and advise Ministers and regulators on them.
- We recommend the government establish a small, expert AI Safety Institute to act as the government's centre of technical and policy expertise on AI safety.
- This would give New Zealand independent domestic capability and a seat in the growing international network of AI safety institutes, with access to assessments and research we could never produce alone.

Why it matters

Advanced AI is developing at extraordinary speed, creating [a range of new risks](#) – from scams and cyberattacks to threats to elections and critical infrastructure.

New Zealand currently lacks the government capability to monitor these risks, evaluate leading AI models and provide expert advice to Ministers.

Without this capability, the government will need to rely on ad hoc advice, offshore assessments, and the goodwill of other nations to share their research and expertise.

Other countries have closed this gap by establishing government centres for expertise on AI safety.

What we're proposing

We recommend the Government establish a small, independent government body made up of a team of technical and policy experts. The Institute would have a mandate to:

- Act as the government's centre of expertise on AI safety.
- Monitor and assess AI capabilities and risks as they develop.
- Provide independent, expert advice to Ministers, government departments and regulators.
- Evaluate significant AI systems and developments relevant to New Zealand.
- Connect New Zealand into the [international network of AI Safety Institutes](#).

This is not a proposal for a new regulator. The Institute would have no enforcement, licensing or compliance powers, and would impose no obligations on business. It would be an advisory and monitoring capability.

The opportunity for New Zealand

AI safety institutes are fast becoming standard institutional practice among comparable countries.

An institute would help to build New Zealand's domestic base of AI safety expertise, rather than relying entirely on researchers working offshore.

New Zealand signed the *Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity*, affirming the following:

"We affirm the unique role of AI safety institutes and other relevant institutions to enhance international cooperation on AI risk management and increase global understanding in the realm of AI safety and security. Through our AI safety institutes or other relevant institutions, we plan to share best practices and evaluation datasets, as appropriate, and collaborate in establishing safety testing guidelines."

New Zealand has an opportunity to follow through on that stated intention and join the growing international network of AI Safety Institutes. This would give New Zealand a seat at the tables where AI safety standards and assessments are being shaped, early visibility of emerging risks and access to the technical work of far larger institutes, at modest cost.

The alternative is that New Zealand remains exposed and unprepared for a growing range of risks associated with AI.

How it could be done

An institute does not require a large new agency. Options range from a dedicated expert unit hosted within an existing department (as Australia has done), through to a small standalone Crown entity with statutory recognition.

The core requirements are:

- A clear mandate.
- Independence of advice (the ability to give Ministers frank assessments).
- Dedicated funding.
- The standing to represent New Zealand in the international network of AI safety institutes.

The right vehicle can be settled through detailed further advice. The important decision is to commit to establishing the function with the right mandate.

International precedent

[Australia](#) established its AI Safety Institute in early 2026, and has since gone further, announcing a new [framework for managing AI](#), including an Office of AI within the Department of the Prime Minister and Cabinet.

The UK, US, Japan, Korea, Canada, Singapore and others operate AI Safety Institutes or equivalents, which are networked and share evaluations and research.

New Zealand is increasingly conspicuous by its absence.

Frequently asked questions

How much will it cost? A credible institute could be established for a few million dollars annually. This is a small fraction of the [\\$1.6 billion that online threats cost New Zealanders in 2024](#) alone.

Isn't this what the National Cyber Security Centre (NCSC) already does? The NCSC's work is valuable, but there is more to do. The risks of AI are much broader than cybersecurity. Assessing emerging threats from leading AI models requires dedicated AI expertise and evaluation capability that goes beyond cybersecurity skills. There is currently no government body doing this work.

How could a small New Zealand institute realistically evaluate leading AI models? Establishing an institute would ensure we have local expertise, but New Zealand doesn't have to do this alone. We would join a network of existing AI safety institutes and gain access to technical information we couldn't produce alone.