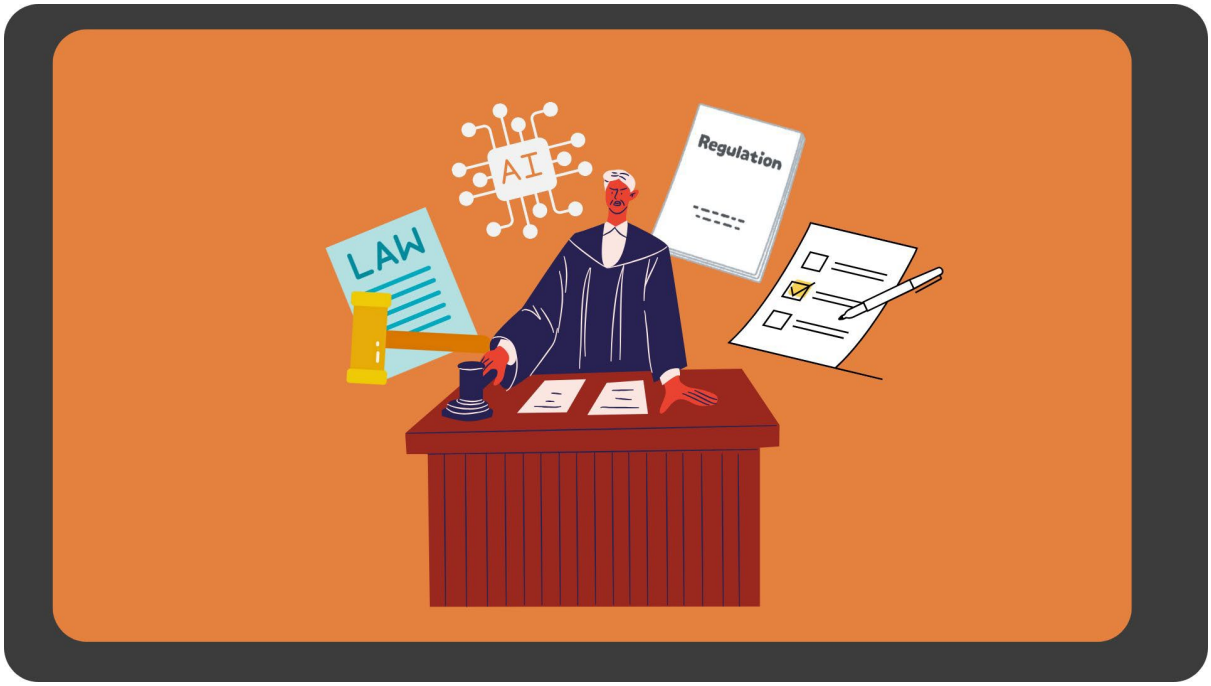
Government guidance: For educators using AI agents

Article #5 of AI in Education Article Series: March 2025



Several foreign governments have begun to regulate the use of AI, with some enforcing strict controls and others pushing for open innovation (Pattison, 2025). Here in New Zealand, no specific laws on AI have been introduced, although some existing laws apply (see the Fair Trading Act 1986, the Human Rights Act 1993 and the Privacy Act 2020) (Pattison, 2025). As of March 2025, the government has released other sources of guidance in the place of specific laws.

This article is the **fifth in a series titled "AI in Education"**, aimed at education providers interested in AI. The intention is for this series to act as a beginner's guide to the use of AI in education, with a particular focus on AI agents. This series is being developed as part of a project to develop an AI agent for learner oral assessment, funded by the Food and Fibre Centre of Vocational Excellence. We invite you to follow along as we (Scarlatti) document our learnings about this exciting space.

Within this article, we summarise three pieces of **government guidance on AI use**: from the Office of the Privacy Commissioner; from the joint work by the Department of Internal Affairs, the National Cyber Security Centre and Statistics New Zealand (2023a); and from the Ministry of Education (2023). At the end, we demonstrate how this guidance may apply to an AI agent for learner oral assessment.

Guidelines released by the New Zealand Government

The Office of the Privacy Commissioner's guidance

In 2023, the Privacy Commissioner issued comprehensive guidance on the use of AI for *people throughout New Zealand.* This guidance sets out several expectations for agencies wanting to use AI. For example, it suggests that you:

- **Review whether a generative AI tool is necessary and proportionate given potential privacy impacts and consider whether you could take a different approach** – Organisations should weigh up the benefits and risks of using AI before adopting this technology.
- **Have senior leadership approval based on full consideration of risks and mitigations** – Make sure that upper leadership is on board with rolling out AI.
- **Conduct a privacy impact assessment (PIA) before using these tools** – PIAs should gather feedback from impacted communities, including Māori and ask the AI provider how privacy protections have been integrated.
- **Be transparent, tell people how, when, and why the tool is being used** – Being upfront and giving users information about the AI tool you are using will help to maintain trust and your organisation's social license to use AI.
- **Engage with Māori about potential risks and impacts to the taonga of their information** – The Commissioner recommends being proactive in your engagement with Māori.
- **Develop procedures about accuracy and access by individuals to their information** – Have procedures in place to ensure that collected information is accurate and that your organisation can respond to requests from individuals wanting to access their information and correct it.
- **Ensure human review before acting on AI outputs to reduce risks of inaccuracy and bias** – To mitigate the risk of acting on inaccurate information, a person should review the output of the AI tool and assess the risk of re-identification of gathered information.
- **Ensure that personal information is not retained or disclosed by the AI tool** – They strongly advise against inputting any personal information into these tools unless the provider explicitly confirms that the information will not be retained or disclosed by the provider. Alternatively, they suggest stripping input data of any information that would enable users to be re-identified.

Note: Further information on what each of these expectations involves can be found in the Privacy Commissioner's guidance and their website (including a valuable step-by-step guide on how to create a PIA).

The Department of Internal Affairs, National Cyber Security Centre and Statistics New Zealand's joint guidance

In comparison to the Privacy Commissioner's Guidance above, this joint guidance was created specifically for *people working in the public service* (2023a). It places these within two groups, which we describe below.

They strongly recommend that you:

- **Don't use GenAI tools for data classified at SENSITIVE or above** – Due to the risk that these datasets could be compromised and the impact this could have on the public service, economy and wider society.
- **Don't input personal information to GenAI tools if they are external to your environment –** As you risk compromising personal information and losing people's trust and confidence in the government.

They also recommend that you:

- **Avoid inputting personal data into GenAI tools in your network** – Do not include personal information (including clients) when using GenAI unless all potential risks have been mitigated, or it is not possible to deidentify data.
- **Prevent AI from being used as a shadow IT (IT that is not supported or approved by your workplace) -** There are added security and privacy risks if shadow IT is being used in your workplace.
- **Avoid inputting any information into GenAI tools that would be withheld under the Official Information Act (OIA)** – If any information that is withheld under the OIA is accessed or used inappropriately, it could be very damaging to the public's trust and confidence in the government.
- **Avoid using GenAI for business-critical information, systems or public-facing channels** – As GenAI can 'hallucinate' information and has the potential to perpetuate bias and misinformation, avoid using these agents in certain contexts.

Note: A later summary of the above guidance put forward '10 dos for the trustworthy use of generative AI' in the public service (2023b). This appears to be intended to be more accessible:

1. **Govern the use of Gen AI robustly** – Their guidance suggests obtaining senior approval for decisions regarding AI, developing an AI policy and sharing it with the Government's Chief Privacy Officer.
2. **Assess and manage for privacy risk** – Protect privacy by undertaking a Privacy Impact Assessment to use and test AI.
3. **Assess and control for security risk** – Undertake security risk assessments and, if possible, opt out of tools that retain your data for training.
4. **Consider Te Tiriti o Waitangi** – Engage with Māori when AI tools are being used for Māori data and when they may impact Māori.
5. **Use AI ethically and ensure accuracy** – Be aware of the limitations of AI, how these tools can perpetuate bias and misinformation, and the importance of checking outputs for accuracy to avoid harm.
6. **Be accountable** – Ensure that those making decisions about using and applying AI are accountable and that decision-makers have the relevant authority.
7. **Be transparent, including to the Public** – Be open about when and why AI is being used. Have processes in place to respond to requests to access and correct information.
8. **Exercise caution when using publicly available AI** – Understand the risks when using publicly available AI such as issues around quality, security and intellectual property.
9. **Apply the Government's procurement principles –** Abide by the procurement rules when sourcing AI tools.
10. **Test safely –** Allow time for teams to learn how to use and trial AI. During training, use low-risk datasets and review outputs for accuracy before they are implemented.

Ministry of Education's Guidance

The Ministry of Education has guidance aimed at *teachers and students* wanting to use AI (2023). They outline four important considerations:

- **Make sure to check the output** – As AI can 'hallucinate' (give answers that seem plausible but do not make sense), you should review and check the tool's answers.
- **Consider cultural bias** – AI models are built from data from across the globe, meaning that most models are built on dominant cultures and languages. As such, these tools may not accurately reflect Indigenous knowledge and are likely to not have comprehensive knowledge of Mātauranga Māori, Te Reo, Pasifika languages and Polynesian cultures.
- **Do not use personal data** – As some AI tools use the prompts users have asked to continue training the model, the Ministry recommends avoiding inputting personal data into the tool.
- **Look over the terms and conditions** – They recommend reading over the terms and conditions to check whether the AI model will use your prompts or the data you provide to train the model.

Note: In addition, the Ministry recommends that each provider has internal discussions about AI in their organisation and come up with their own policies on using this technology. Examples of education providers' AI policies can be found on Netsafe's website.

Shared themes in the Government's guidance

Below, we synthesise some *example* shared themes and explore their relevance to our work.

Recommendation	OPC	DIA, NCSC StatsNZ	MoE	Relevance to Scarlatti's AI agent for learner oral assessment
Avoid inputting personal data	✓	✓	✓	We need to collect learner's names for tutors to be able to identify them and save their grades. However: • We do not use AI to process student names. The agent sends the learners assessment answers to OpenAI, but it does *not* send their names. These are instead processed on Scarlatti's side. • The assessments we are using for this pilot do not ask the student to provide any personal information in their answers (e.g., employer name, their location). • If a student unexpectedly provided personal information in their answers, our subscription to OpenAI does not give permission for OpenAI to use inputted data to train their models, only to retain data for 30 days for security reasons (see their policy, also mentioned in our reference list). • Despite the above, we acknowledge there are misconceptions (and therefore concerns) about OpenAI's use of data. We therefore also provide learners with the option of not undertaking the assessment by AI, and if they do use it, we tell them explicitly that they do not need to provide personal information to the agent.
Check and review outputs	✓	✓	✓	Reviewing outputs (e.g., transcripts, recordings, AI provisional grades) is important to ensure that the assessment agent is giving learners accurate grades and useful feedback that aligns with course materials and assessment rubrics. This involves running internal tests, piloting the agent with tutors and learners, and having safeguards in place during these pilots to ensure tutors are checking outputs.
Engage with Māori and consider cultural impacts	✓	✓	✓	We want our assessment agent to benefit a diverse range of users, and we certainly do not want to perpetuate harmful bias or inequalities. However, *larger* pilots are likely required to detect any bias confidently. We strongly suggest that funding is needed to test how a similar agent would perform when used in te reo Māori, and to test whether learners' assessment outcomes are impacted by accent or language.

Conduct a Privacy Impact Assessment	✓	✓	NA	The first action in this work was to submit an application to Scarlatti's internal ethics committee. On approval, we then undertook a review of both Aotearoa government guidance (this article), a number of international pieces of guidance, and key AI thought leaders in Aotearoa. Common themes and "so what's" were identified for each stage of the project. This process was taken as a substitute for a PIA. We would encourage any education providers that use such an agent to undertake their own PIA.
Be transparent	✓	✓	NA	Before a learner undertakes their assessment with the AI oral agent, we provide them with an information sheet. This includes an overview of the project, why the education provider is interested in using AI, what an AI assessment agent is, what would happen with their information, and how the pilot would affect the rest of the course. We also make it clear that the AI agent is not intended to replace in-class tutoring. Learners and tutors are given our contact details so that they can come to us with any concerns. Before starting the assessment, the learner is asked to give their consent to participate.
Obtain senior leadership approval	✓	✓	NA	Our pilot partners (education providers piloting the AI agent for learner oral assessment) have obtained senior leadership approval to conduct these pilots. This involved considering the appropriateness of this technology for their goals, programmes, assessments and learners.

Scarlatti's take

The above government recommendations are primarily focused on the legal, reputational, and ethical risks of AI usage. What appears to be missing is a more optimistic view on the potential for AI to improve the efficiency, effectiveness and inclusiveness of education (as well as other public services). AI could present substantial opportunities to improve learning outcomes, especially for learners who are not currently well-served by the system.

Questions that we are asking for our own AI agent:

- What will government regulation of AI look like in future in Aotearoa, and how will this impact the use of AI in education?

Interested in following our journey into AI?

- Sign up to receive our next article directly to your inbox.
- Contact the Scarlatti team to share your thoughts or questions.

References

Department of Internal Affairs, National Cyber Security Centre & Stats NZ. (2023a). Initial Advice on Generative Artificial Intelligence in the public service. Retrieved January 8, 2025, from https://www.digital.govt.nz/assets/Standards-guidance/Technology-and-architecture/Generative-AI/Joint-System-Leads-tactical-guidance-on-public-service-use-of-GenAI-September-2023.pdf

Department of Internal Affairs, National Cyber Security Centre & Stats NZ. (2023b). Summary of system leaders' guidance for use of gen-AI across the New Zealand public service. Retrieved January 8, 2025, from https://www.digital.govt.nz/assets/Standards-guidance/Technology-and-architecture/Generative-AI/Joint-System-Leads-tactical-guidance-on-public-service-use-of-GenAI-summary-September-2023.pdf

Ministry of Education. (2024, November 25). *Generative AI: Guidance and resources for education professionals on the use of artificial intelligence in schools.* Retrieved January 8, 2025, from https://www.education.govt.nz/school/digital-technology/generative-ai

Office of the Privacy Commissioner. (2023). Artificial intelligence and the information privacy principles. Retrieved January 8, 2025, from https://privacy.org.nz/assets/New-order/Resources-/Publications/Guidance-resources/AI-Guidance-Resources-/AI-and-the-Information-Privacy-Principles.pdf

OpenAI. (2023). Data Controls in the OpenAI platform. Retrieved March 31, 2025, from https://platform.openai.com/docs/guides/your-data

Pattison, C. (2025). AI playbook for New Zealand. Capability Collective. Retrieved March 18, 2025, from https://www.linkedin.com/posts/craig-pattison-4353b337_ai-playbook-for-new-zealand-activity-7305427400241160192-ZCRG/?utm_source=share&utm_medium=member_desktop&rcm=ACoAAB0RivcBwjdGQ-wOH39QjJNgo8BQiVc8Ar0