



12 March 2026

[REDACTED]
[REDACTED]
[REDACTED]

Dear [REDACTED]

Thank you for your request made under the Official Information Act 1982 (OIA), received on 6 March 2026. You requested the following as a follow up to the published OIA responses about Inland Revenue's use of AI:

1. *IRD's OIA response says Engage was paid \$143,121.55 as a delivery partner across multiple Microsoft AI products. Engage is also a Microsoft delivery partner. Was Engage involved in any advisory, scoping, or assessment role before or alongside its delivery work at IRD? If so, how was that conflict managed?*
2. *IRD's OIA response says Assurity Consulting conducted a proof of concept for a product it was also proposing to deliver, and subsequently built IR's private large language model as a testing partner. How did IRD manage the conflict of interest in Assurity assessing its own product's suitability?*
3. *IRD's OIA response confirms the AI Oversight Group reviewed and signed off on Assurity's assessment of Assurity's own product. What steps did the AI Oversight Group take to independently verify that assessment?*
4. *IRD used Microsoft Copilot to generate a summary of Copilot's own benefits as part of its procurement process. How does IRD ensure objectivity when vendor-generated material forms part of its decision-making?*
5. *Can IRD confirm the GST Integrity Model has never produced a false positive that led to an investigation of a compliant taxpayer?*
6. *IR documents record failures of AI tools to correctly handle Te Reo Māori and technical terms. Were these failures addressed before the tools were deployed at scale?*

Question one

Inland Revenue (IR) uses the Microsoft Office365 suite of tools as our core productivity platform to support communication, collaboration and information management. Use of Office365 has been in place for over a decade, and SharePoint was implemented as a collaboration platform in 2018. Our approach to AI adoption over the past 12-18 months has been to explore the AI offerings of our existing toolsets first, before exploring any new or 'AI Native' tools. This formed the basis for IR to explore the use of Microsoft Copilot, as a natural expansion of the already widely used Office365 toolset.

As part of exploring the potential use of M365 Copilot, we consulted with Microsoft about potential vendors with Microsoft specialist certification who could support our work in this area. As a long-term trusted partner of IR, we felt it appropriate to speak to them about getting access to the right level of specialist knowledge in this new and emerging area.

Microsoft provided us with a list of several vendors, from which we chose Engage². Engage² was engaged through a Marketplace subscription agreement, which aligns with standard practice permitted by Government Procurement rules. Direct sourcing via Marketplace is a routine and approved approach for engaging certified professional services providers.

Engage²'s role was to support an initial pilot (proof of concept/value) of Copilot Studio, a second pilot of M365 Copilot and the subsequent wider rollout of M365 Copilot. Engage² were not involved in any evaluation or decision making (procurement, strategic or licencing related). All decisions remained with IR through our existing governance mechanism.

IR has not identified any conflict of interest in its engagement with Engage². IR has a multi-stage governance process for technology investment within which probity oversight is included.

Question two

IR has a long-standing relationship with Assurity Consulting Limited, and they are contracted to deliver a range of testing related services to IR. IR contracts with Assurity Consulting Limited on an outcomes basis with a requirement under the contract to bring efficiencies to the table. Where any efficiencies involve the use of technology, automation, or AI, IR will test these to ensure any efficiencies can be delivered safely and securely, and deliver the outcome proposed. This agreement with Assurity Consulting Limited is reviewed regularly as part of our standard commercial processes.

Utilising AI capability to more efficiently generate software testing scenarios is an example of the innovative approach we expect and encourage our partners to deliver. As part of this, Assurity Consulting Limited developed a software testing scenario generation solution. While the work was performed by Assurity Consulting Limited, evaluation and all subsequent decisions to adopt this approach were led by IR staff.

Question three

While the work to build the solution was performed by Assurity Consulting Limited, IR's Testing team led the assessment of the value of the solution, and subsequent recommendation of next steps to the AI Oversight Group.

Question four

IR approved organisational-wide use of Microsoft Copilot Chat (formerly known as Bing Chat) in November 2024. In the event IR staff used Microsoft Copilot Chat to research the capabilities of M365 Copilot prior to the procurement of licences, this is acceptable under IR's Artificial Intelligence Staff Use Policy.

IR's Artificial Intelligence Staff Use Policy includes principles with which approved use cases of AI tools are required to comply with. One principle is Human Oversight: *"We will retain human oversight to assess for unintended consequences and act on this information. This includes understanding limitations and identifying and managing bias in the implementation of AI tools and in the explanations and decisions made by AI tools"*.

The Policy also details several responsibilities of users of AI:

- *"Users must consistently review and confirm the accuracy of any generated AI/LLM/Agent output (including but not limited to text, audio, or visual media) or intelligent system output"*
- *"Users must review AI generated output for potential bias, ethical concerns, and unintended messaging"*
- *"Users must acknowledge that any information produced in part or whole by an AI system will be attributable to the Inland Revenue staff working with that information, and ultimately Inland Revenue will be responsible for the quality and outcomes of that information."*

In the event an IR staff member used Microsoft Copilot Chat to research the capabilities of M365 Copilot, they would be required to review and confirm the accuracy of the answer through primary sources, for example Microsoft's website. With regards to the performance of the tool, IR used material from many sources to support the decision making for initiating a small-scale proof of concept or pilot. Wider deployment decisions used IR's findings from the proof of concept as a key input. All decisions remained with IR.

Question five

IR's GST Integrity Model was introduced in March 2019. This model is not an AI tool in the context of solutions that use a Large Language Model (LLM) to support decision making. It is more correct to call the GST Integrity Model a predictive scoring model and is by nature risk based. This is a solution that is an integral part of our Core Tax system (START) that was deployed as part of IR's Business Transformation Programme. While it contributes to identification and commencement of investigations, the specific model results (i.e. false positive or otherwise) are not in themselves determinative in this regard. Specialised compliance staff will review model outputs alongside other data and information available/requested before commencement of any investigation.

It is also worth noting that how this model is configured is regularly reviewed for how it is operating.

Question six

As mentioned in the response to question four, IR staff are required to have human oversight of all AI outputs to review and confirm for both accuracy, and applicability to the specific context they are being used in. With regards to the rollout of Genesys Conversation Summarisation, IR staff complete a 30-minute training module followed by a knowledge check.

As part of this knowledge check, IR staff agree to a set of "must dos" which include:

- Review the summary: Always review the AI-generated summary to ensure it accurately reflects the conversation.
- Edit as required: Edit the summary if it contains inaccurate or incomplete information, or if it would not make sense to someone else reading it.
- Ensure completeness: Ensure all key points and essential details from the conversation are included in the summary.

In the event the natural language processing model doesn't accurately pick up a non-English word, or a technical term, the staff member will identify this as they do their review, prior to lodging the notes into the customer record. In this way, conversation Summarisation is deployed as an efficient support to our human customer service officers to more efficiently handle call wrap up. It is not intended to be a replacement for these staff.

Genesys Cloud has a dictionary management tool that allows IR to add IR-specific jargon, terms, and commonly used phrases. This includes commonly used phrases, technical terms, greetings, etc. that may not exist in the dictionary 'out of the box'. This ensures that the generated summaries are more relevant and precise and reflect the unique language and terminology used at IR. We are continuously updating our IR-Genesys Dictionary so it can recognise jargon specific to IR (for example 'FamilyBoost' or 'KiwiSaver'), and terms unique to New Zealand (for example 'RealMe'). We have also added some frequently used Te Reo Māori words into the dictionary. This dictionary was updated throughout our pilot phases of evaluating this tool. This ensured a number of the most common phrases were included before the scaled deployment commenced. There may be terms which have not yet been added to the dictionary which could result in inaccurate transcripts, however this is not seen as a "failure", as the IR staff member will always review the summary before lodging into the customer's record.

A copy of the Genesys Learning module has previously been released under the OIA and is available on our website (<https://www.ird.govt.nz/-/media/project/ir/home/documents/oia-responses/november-2025/2025-11-14-agency-approved-ai-tools-usage-guidelines-licensing-privacy-assessment.pdf?modified=20260114233945>). Please refer to pages 12 to 53.

Publishing of OIA response

We intend to publish our response to your request on Inland Revenue's website (ird.govt.nz) as this information may be of interest to other members of the public. This letter, with your personal details removed, may be published in its entirety. Publishing responses increases the availability of information to the public and is consistent with the OIA's purpose of enabling more effective participation in the making and administration of laws and policies and promoting the accountability of officials.

Thank you again for your request.

Yours sincerely



Patrick O'Doherty
Enterprise Leader, Digital