Inland Revenue
Te Tari Taake

26OIA1627

19 December 2025

[redacted]

Dear [redacted]

Thank you for your request made under the Official Information Act 1982 (OIA), received on 12 November 2025. Due to the extensive collation of information relating to your request, my response was extended to 19 December 2025. You requested the following:

1) **AI and digital automation spend**
A list of all *projects, pilots or procurements* undertaken by your agency use between 1 January 2023 and 31 October 2025 involving:
- Artificial intelligence (AI) or machine learning
- Natural-language processing, transcription or summarisation tools
- Automation, robotic process automation (RPA) or digital assistants
- Predictive analytics or decision-support algorithms
- Agentic AI

For each, please provide:
- Project name and description
- Vendor or delivery partner (if any)
- Total cost to date and funding source
- Total budget
- Current status (active, completed, paused, cancelled)
- Whether it was developed in-house or procured externally.

2) **Broader technology-productivity initiatives**
- Any reports or analyses since 1, January 2023 quantifying how technology investment (AI, automation, digital infrastructure) has contributed to productivity or efficiency improvements within your agency.
- Any comparative or benchmarking data used to measure those gains.

Introduction

From 2017 through to 2021, Inland Revenue's (IR) business transformation involved changes to every aspect of the way we operate, including policy settings, processes, our operating model and people capabilities, and technology.

IR invested heavily in technology during business transformation, which has greatly improved the services available to customers. Transformation required IR to understand the asset profile it would need to be able to meet customers' needs both now and into the future. As a result, IR is in a strong position, and the focus is on ensuring the department makes the most of the capabilities it now has.

Item One

The table attached as **Appendix A** details the information you have requested on IR's AI and digital automation spend. All figures provided are GST exclusive. Some information has been withheld under the following sections of the OIA, as applicable:

- 9(2)(b)(ii) – to protect information where the making available of the information would be likely unreasonably to prejudice the commercial position of the person who supplied of who is the subject of the information,
- 18(d) – the information requested is publicly available, and
- 18(g) – the information requested is not held by IR nor do I have grounds for believing that the information is held by another department.

As required by section 9(1) of the OIA, I have considered whether the grounds for withholding the information requested is outweighed by the public interest. In this instance, I do not consider that to be the case.

The information refused under section 18(d) can be found on the budget website (www.budget.govt.nz) via the following link: Vote Revenue - Vol 4 Finance and Government Administration Sector - The Estimates of Appropriations 2025/26 - Budget 2025.

There are several exceptions to the table which I will explain first, along with some specific information regarding funding and development.

Funding source and total budget

The funding source for all IR's use of AI and automation is the existing Vote Revenue departmental baseline. With regards to the total budget, IR does not have a specific amount of funding allocated to AI and automation. Funding for AI or automation initiatives are considered on a case-by-case basis, and we reprioritise our existing budgets as we make decisions to implement particular tools.

Cost - IR's tax and social policy administration software

IR's tax and social policy administration software, START, is supplied by FAST Enterprises LLC. This software has machine learning, predictive modelling, decision-support algorithms and automation capabilities built in. I have detailed these individually in the table below.

As IR pays for one enterprise-wide licence for START, and does not pay specifically for individual AI or automation features, I am unable to provide the total cost to date for these tools. This cost is included in the overall price paid for the software. Therefore, this part of your request is refused under section 18(g) of the OIA.

Cost of projects completed under IR's Optimisation Hub and general automation enhancements through FAST Enterprises LLC, are also withheld under section 9(2)(b)(ii) of the OIA. Disclosing this could prejudice FAST Enterprise LLC's commercial position and our ability to negotiate effectively in the future.

Cost – IR's Data Intelligence Platform

IR's Data Intelligence Platform (DIP) is a SAS Managed Hosted Service, using Snowflake. This is a cloud-based data platform, where commonly used data is refined and presented for rapid and easy consumption as tables, reports, analyses and dashboards. This platform encompasses traditional AI features. Snowflake's consumption-based billing does itemise costs, but this does not go down to the detail of traditional AI use. Charges are broadly categorised as compute credits by warehouse size and usage period, storage in terabytes per month, data transfer volumes and any additional services. Therefore, your request for the total cost to date for the traditional AI tools used within the DIP is refused under section 18(g) of the OIA.

Cortex AI is Snowflake's native AI and machine learning platform embedded directly into the Snowflake Data Cloud. As part of that, IR are using Cortex Analyst. Use of this tool follows the same billing structure as discussed above, and charges are broadly categorised. As IR are unable to identify if compute credits charged relate to use of Cortex Analyst, your request for the total cost to date of this tool is refused under section 18(g) of the OIA.

Cost – Licences

IR's total spend for any licences is withheld under section 9(2)(b)(ii) of the OIA. While we aim to be transparent, releasing this figure would allow the per-licence cost to be calculated, which is commercially sensitive information. Disclosing this could prejudice the supplier's commercial position and our ability to negotiate effectively in the future.

Delivery partner - Engage²

In several of the Microsoft tools shown in the table in Item One, I have noted the delivery partner, Engage². The total amount IR has paid to Engage² is $143,121.55. As IR worked with Engage² on multiple Microsoft products concurrently, I have been unable to split this up in the table.

Development in-house or external procurement

All AI tools in use at IR are procured externally. In some instances, IR staff have developed the models or agents, however, these use an infrastructure/platform procured externally. For example, models built in START or DIP, or agents in Microsoft Copilot Studio.

Item Two

Information publicly available

IR's annual reports contain a wealth of information on our performance and strategy. You can find information about IR's opportunities and challenges with AI on page 24 of the 2025 Annual Review, which can be found via the following link: Inland Revenue Annual Report Te Tari Taake Pūrongo ā-Tau 2024-25.

Information being released

I have enclosed as **Appendix B** the following reports and analyses which are in scope of your request. I note that not all technology investment has had a report or analysis written on it, particularly those of a smaller scale.

Reports and Briefing Notes to the Minister of Revenue

Item	Date	Document	Decision
1.	26 February 2025	IR2025/068 Update on Inland Revenue's use of Artificial Intelligence	Released with some information withheld under sections 6(c) and 9(2)(a).
2.	27 August 2025	IR2025/365 Update on Inland Revenue's use of Artificial Intelligence	Released with some information withheld under section 9(2)(a).

Microsoft Product Reports

Item	Date	Document	Decision
3.	November 2024	Benefits of Copilot for IR	Released.
4.	December 2024	Microsoft 365 Copilot Trial Midway Report: Information Governance Team	Released.
5.	March 2025	Microsoft 365 Copilot Trial Midway Report #2: Information Governance Team	Released.
6.	June 2025	AI Oversight Group: Microsoft 365 Copilot Project Closure Report	Released with some information withheld under sections 9(2)(a) and 9(2)(b)(ii).
7.	18 June 2025	AI Oversight Group: Microsoft 365 Copilot, Copilot Service Enablement	Released with some information withheld under sections 6(c), 9(2)(a), and 9(2)(b)(ii).
8.	25 June 2025	AI Oversight Group: Microsoft 365 Copilot Production BAU pilot with Policy & TCO	Released with some information withheld under section 9(2)(b)(ii).
9.	18 September 2025	AI Oversight Group: Technology Services / Policy Production pilot M365 Initial Findings Report	Released with some information withheld under section 6(c).

All other reports

Item	Date	Document	Decision
10.	3 March 2025	Conversation Summarisation Proof of Concept – Outcome and Recommendations	Released with some information withheld under section 9(2)(b)(ii).
11.	14 April 2025	AI Oversight Group: AI Test Scenario Generation – PoC Findings Report	Released.
12.	May 2025	Voice Isolation – Review of Early Adopter Feedback	Released.
13.	12 June 2025	Enterprise Priorities and Performance Committee: myIR Navigation Assistant Closure Report	Released with some information withheld under section 9(2)(a).
14.	18 September 2025	Technical Design Authority: Canon Email Ingestion Channel Functionality PoC Update	Released with some information withheld under sections 6(c), 9(2)(a), and 9(2)(b)(ii).
15.	November 2025	AI Oversight Group: DIP Snowflake Cortex AI PoC Evaluation Report	Released with some information withheld under section 9(2)(a).

Information being withheld

I have identified two further documents in scope of your request, however these are withheld in full under the following sections of the OIA:

- 6(c) – the making available of that information would be likely to prejudice the maintenance of the law, including the prevention, investigation, and detections of offences, and the right to a fair trial, and
- 18(c)(i) - the making available of the information requested would be contrary to the provisions of a specified enactment, namely section 18(3) of the Tax Administration Act 1994. The Commissioner of Inland Revenue is not required to disclose any item of revenue information if the release of that information would adversely affect the integrity of the tax system or would prejudice the maintenance of the law.

Date	Document	Decision
1 April 2024	Memo: Implementing the new GST Integrity Model	Withheld in full under section 18(c)(i)
January 2025	START Integrity Manager – new GST model – Post Implementation Evaluation	Withheld in full under section 18(c)(i)

Right of review

If you disagree with my decision on your OIA request, you have the right to ask the Ombudsman to investigate and review my decision under section 28(3) of the OIA. You can contact the office of the Ombudsman by email at: info@ombudsman.parliament.nz.

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

Cate Robertson
Enterprise Leader, Strategic Architecture

Appendix A – Uses of AI and automation within Inland Revenue from 1 January 2023 to 31 October 2023, and the associated cost during this period.

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
ABBYY FineReader 16	Text recognition and document conversion tool, used to convert PDFs into excel.	In production	Optical Character Recognition, Machine Learning	ABBYY	Withheld under section 9(2)(b)(ii).
Ability to pay model	Analyse the customer's circumstances and recommend the best action for that customer	In production	Machine Learning	FAST Enterprises LLC	Refused under section 18(g).
ACC alternate tax rate	Automated solution for applying alternate tax rates to back-dated lump sum payments	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Address Update enhancements	Automate address validation and reduce manual entry errors in myIR	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
AI Futurist	Enables querying and summarisation of content.	In production	Generative	Futures Platform	Withheld under section 9(2)(b)(ii).
Answers in Viva Engage	Helps staff ask questions and receive answers from organisational content	Pilot concluded	Generative	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Articulate 360 with AI	E-Learn development software	Pilot	Generative	Articulate Global Inc	Withheld under section 9(2)(b)(ii).
Assurity Intelligence	Test scenario generation	Pilot	Generative	Assurity Consulting Limited	Withheld under section 9(2)(b)(ii).
Automated Patching for Windows	Automated patching for OS, Drivers.	In production	Automation	Microsoft	Free of charge.

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Azure DNS (Domain Name System) management	Automatically handling DNS requests	In production	Automation	Azure	Withheld under section 9(2)(b)(ii).
Brandwatch	Media monitoring	In production	Machine Learning, Natural Language Processing	Falcon.io APS	$23,102
Canon Image Scanning Solution Project	End-to-end document capture service that digitises inbound mail and integrates with START, streamlining physical mailroom processes	In production	Automation	Canon, FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Certificate automation	Automate the certificate lifecycle for some platforms	In production	Automation	AKQA Pty Ltd, Amazon	Withheld under section 9(2)(b)(ii).
Certificate Authority automation	Automate Certificate Authority (CA) update for Gateway Services	In production	Automation	FAST Enterprises LLC, Spark	Withheld under section 9(2)(b)(ii).
Chainalysis Reactor	Blockchain Analysis	In production	Machine Learning	Chainalysis Inc	Withheld under section 9(2)(b)(ii).
Confirmation of Payee initiative	Improve quality of bank accounts provided by checking account numbers and names match bank records	Active	Decision support algorithms	Get Verified, FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Core21 Upgrade – Managed Payments	Core21 platform upgrade introduces new functionality for managing payments, enabling automation of some existing processes	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Core21 – child support relationship validation process	Automated real-time validations at case completion to identify and prevent incorrect data updates	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Cotiss Sourcing and Evaluation Initiative	Digital procurement solution	In production	Automation	Cotiss	Withheld under section 9(2)(b)(ii).
Coveo	IR's public websites and internal intranet use Coveo, a machine-learning search platform that continuously improves results based on user behaviour.	In production	Machine Learning	Coveo	Withheld under section 9(2)(b)(ii).
Deceased customers – cease myIR log on ability	Automate cessation of myIR web logon once verification of death received	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Dragon Naturally Speaking	Screen reader	In production	Machine Learning	Nuance	Withheld under section 9(2)(b)(ii).
DDoS protections **- AWS Shield** **- Azure DDoS Protection** **- Cloudflare** **- F5** **- Oracle DDoS**	IR uses a range of tools to prevent our systems and services from unexpected outages due to network attacks.	In production	Machine Learning	Oracle, AWS, Azure, Cloudflare, F5 Inc	Withheld under section 9(2)(b)(ii).
eServices Assistant	The myIR Navigation assistant is a digital guide that directs users to the right myIR tasks via keyword-based links	In production	Digital assistant	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Enabling second SAML identify for snowflake environment	Enables automated user provisioning and single sign-on	In Production	Automation	Microsoft, Snowflake	Withheld under section 9(2)(b)(ii).
Email-user link prediction	Predict the similarity of email addresses and usernames such that IR can identify the probability that they are controlled / used by the same real-world person.	In development	Machine Learning	Snowflake	Refused under section18(g).
Enterprise Performance Management (EPM) – Phase 1	Deliver improvements to the Oracle EPM model – improve process flows and usability	Project Complete	Infrastructure reconfiguration	James & Monroe NZ Pty Limited Assurity Consulting Limited J & M Travel Starfish Consulting Limited	$783,000 Amount paid to individual vendors is withheld under section 9(2)(b)(ii).
Enterprise Performance Management (EPM) – Phase 2	Deliver improvements to the Oracle EPM model – core structural changes	Project Complete	Infrastructure reconfiguration	James & Monroe Assurity Consulting Limited Accenture	$1.49m Amount paid to individual vendors is withheld under section 9(2)(b)(ii).
Enterprise Performance Management (EPM) – Phase 3	Deliver improvements to the Oracle EPM model – value add features	Project active	Infrastructure reconfiguration	James & Monroe Assurity Consulting Limited Accenture	No spend as yet.

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Enterprise Services Data, Analytics and Reporting Project 2024/25	Standardised and automated reports	Project active	Automation	SAS	Withheld under section 9(2)(b)(ii).
FamilyBoost	Implement new product in START using standard design patterns including automation	In production	Decision support, automation	FAST Enterprises LLC	Refused under section 18(d).
Figma	Prototyping software that enables IR to develop mock-ups of intended changes to products and services across both e-services and internal/external IR websites	In production	Machine Learning	Figma, Inc	Withheld under section 9(2)(b)(ii).
Final Year Fees Free Roll out	Implement ability to process final year fees free policy, includes standard design patterns and automation of some functions and processes	In production	Automation	FAST Enterprises LLC	Refused under section 18(d).
Financial intelligence network detection	Links, matches and identifies multi-dimensional risks of users via operational and strategic visualisation	In production	Machine Learning	Snowflake	Refused under section 18(g).
Forms and Guides	Content enrichment, modularisation and summarisation from web and PDF content	Proof of concept completed, did not proceed.	Generative	Microsoft and AKQA Pty Ltd	Free of charge.

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Genesys Agent Assist	Creates summaries of conversations with contact centre agents for post-call notes	In production	Generative	Genesys and One NZ	Withheld under section 9(2)(b)(ii).
Genesys Agent Copilot	Automatically presenting relevant knowledge articles to contact centre agents during voice interactions with customers.	Pilot completed, did not proceed	Generative	Genesys and One NZ	Withheld under section 9(2)(b)(ii).
Github Code Spaces	Automated access to tools and permissions	In production	Automation	GitHub	Withheld under section 9(2)(b)(ii).
GitHub for Jira Integration	Automatically sync GitHub data into Jira tickets to provide work visibility and reduce duplicate effort	In production	Automation	Github, Atlassian Cloud	Withheld under section 9(2)(b)(ii).
Google Analytics 4	Enables IR to track engagement across internal and external digital channels and seamlessly ingest analytics into IR's Intelligence Platform for better insights and decision-making	In production	Predictive modelling, automation	Microsoft, AKQA Pty, Snowflake	Withheld under section 9(2)(b)(ii).
Graph Entity Resolution	Analyses and compares information held by IR to external datasets provided by third parties to determine if records are referencing the same entity.	In production	Machine Learning	Snowflake	Refused under section 18(g).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
GST integrity model	A predictive model to assess the risk of GST returns.	In production	Machine Learning	FAST Enterprises LLC	Refused under section 18(g).
GST Registration Decision Support	Implement new rule-based processing to enable enhanced automation	In production	Decision Support	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Hardware Asset Management Project	Streamline the processes associated with hardware asset management.	In production	Automation	ServiceNow	Withheld under section 9(2)(b)(ii).
Hootsuite	Social Media monitoring	No longer in use	Machine Learning, Natural Language Processing	Hootsuite Inc	$73,837
HR (Human Resources) advisory transition to ServiceNow	Streamlined HR advisory requests in ServiceNow with automated workflows and work Assignment.	In production	Automation	ServiceNow	Withheld under section 9(2)(b)(ii).
Identify a charities customer	Automatically update customer status to charity / no longer a charity from DIA file.	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Individual Income Tax Assessment (IITA)/IR3 expenses amendment reminder message	Automating reminder notices when no response received	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Instalment Arrangement Pre-payment and missed payment reminder	Automated reminders issued to customers who have entered an instalment arrangement	Pilot	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
International Travel Approvals	Implement ServiceNow for international travel related approvals. Moving paper process to digital process with semi-automation	Active	Automation	ServiceNow	Withheld under section 9(2)(b)(ii).
IR3 no longer required process	Automate marking returns as not required.	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Isentia	Media monitoring	In production	Machine Learning, Natural Language Processing	Isentia Limited	$105,657
KiwiSaver Government Contribution (GVC) Eligibility Changes	Automated tool to test GVC eligibility calculations when there are changes.	Testing environments	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Department of Corrections – information share and use of information	New file exchange format to enable more automation.	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
SMS- START Functionality improvements	Automated handling of SMSs that are not delivered successfully	Active	Automation	Modica and Fast Enterprises	Withheld under section 9(2)(b)(ii).
Student Loan START data third party payment providers	Automatically create and send files	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Small Business Cashflow Scheme initiative (SBCS)	Automatic issue of SBCS interest expense letter	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
KiwiSaver annual Government Contribution (GVC) overhaul	Accounts auto created along with automated testing tool for the GVC eligibility code calculations	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Digitising Customs Arrival card	Auto staging of Arrival Card case in START	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
START Task Manager Optimisation	Optimise the way we manage work and remove unnecessary tasks	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Working for Families registration and modification updates	Increase the number of registrations and modifications that are automatically processed	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Working for Families/child support shared customer alignment	Automate identification of mismatches	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Trust return changes	Automate non-active trust reactivation and relationship linking.	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Individual Income Tax Assessment and Working for Families rollover	Increase the number of cases that are automatically processed	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Redesign of the Reserve Scheme Accounts in START	Increase the number of cases that are automatically processed	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Working for Families customers SMS messaging Christmas week	Automated SMS for payments during Christmas	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Small Business Cashflow Scheme repayment, default and collection process	Rules-based automated approvals and notifications	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Instalment arrangement reminders	Pilot – randomised controlled trial approach to improve adherence to instalment arrangements	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Individual IRD number application optimisation	Optimising the IRD number application process	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Add Extension of Time request in myIR	Digitise form and automate approvals if conditions met	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Automate production Contact Centre testing	Automate test scripts for Contact Centre production	Project Complete	Machine Learning	Assurity Consulting Limited	Withheld under section 9(2)(b)(ii).
Powerapps – connecting to Outlook (MSD Information Request emails)	Automate emailing solution for information requests	Pilot complete	Powerapps	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Partnering with third party provider for payment resolution	Automate information sharing and debt monitoring	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Optimising extracting customer information from START	Automate the extraction of customer information for legal evidence	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Process to bulk add indicators in START	Automate process to add/remove specific indicators	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
DIP to START automated interface	Automate bulk file transfer	Complete	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Voice biometric control remediation	Enhance the security of IR's voice biometric authentication system	Active	Biometrics	One NZ	Withheld under section 9(2)(b)(ii).
s17b and bank deduction notice process changes	Enhance email process and implement new rules based processing design	Active	Automation and decision support	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Collection analytics optimisation	Implement new rules based processing design enabling enhanced automation	Active	Automation and decision support	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
GST group adjustments	Automating manual steps of process	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Mailbox task ingestion into START	Automate ingestion of external mailboxes into START	Active	Automation	Canon, FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Add a prescribed withholding rates case/functionality	Automate current processes	Active	Automation, Decision Support	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Extinguishing of losses and excess imputation credits when writing off tax debts	Automate current processes	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Enhancing PPL reconciliation process	Automate current processes	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Campaign transition plan	Automating cyclic notifications and letters	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Macrons creating payment file errors with Westpac	JavaScript function to remove diacritics from text.	In production	Automation	Oracle, Accenture, SAP	Withheld under section 9(2)(b)(ii).
MarianMT	Translating text for digital forensics	Available	Neural machine translation	Meta	Free of charge.
Microsoft 365 Copilot Agents	Pre-built agents	Phased roll out	Agentic	Vendor: Microsoft Delivery Partner: Engage²	Part of M365. Copilot licence Withheld under section 9(2)(b)(ii). Engage²: refer to introductory comments for cost.
Microsoft Copilot Chat (Bing/Browser)	AI-powered chat service	Available	Generative	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Microsoft Copilot Studio	Create and deploy AI copilots and chatbots for business needs	Proof of Concept	AI Agent	Vendor: Microsoft Delivery Partner: Engage²	Part of M365. Copilot licence Withheld under section 9(2)(b)(ii). Engage²: refer to introductory comments for cost.

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Microsoft Defender	Enterprise security solution integrated with Microsoft apps to monitor devices and identities for malicious activity.	In production	Machine Learning	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Microsoft Power BI	Data visualisation and analytics platform for building dashboards and reports	In production	Machine Learning	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Microsoft Purview Enablement project	Integrated solutions for data governance, security, and regulatory compliance	In development	Machine Learning, automation	Vendor: Microsoft Delivery Partners: Liquid IT Limited AKQA Pty Limited	Part of E5 licence. Microsoft: Withheld under section 9(2)(b)(ii). Liquid IT Limited: $123,363 AKQA Pty Limited: $40,293
Microsoft Teams Premium	Provides Teams meeting recaps with enhanced security and premium features.	Pilot	Generative	Microsoft	Withheld under section 9(2)(b)(ii).
Microsoft Teams Voice Isolation	Provides voice isolation to reduce background noise and deliver clearer audio during Teams meetings	In production	Machine Learning	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Noggin (IR's Health and Safety System) automation	Automate file transfer	In production	Automation	Oracle	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Employer Information Validations	Update validations to enable more straight through processing.	Completed	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Auto-approve Payment Plans	Increase automation for processing and prevent incorrect requests	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Income Tax Extension of Time Automation	Automate the current process.	Completed	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Upfront Validations KiwiSaver Registration	Review validations to enable more straight through processing.	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Working for Families Modification – Payment Frequency Changes	Automate current process	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Unable to link Student Loan alternative contact person	Improve soft matching to enable more automated processing	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Tax Residency Certificate Request	New self-service for customer requests enabling automated processing	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Goods and Services Tax Return Errors	Withheld under section 18(c)(i) of the OIA.	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Employer Information Schedules – Reduce Duplicates for Manual Review	Enhance processing rules enabling more straight through processing	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Payment Corrections – GAP Payment	Enhance processing rules enabling more straight through processing	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Working For Families automate task to contact customer for child IRD number	Implement an automated reminder system	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Automation of 'Child Changes request for information' letter	Automate current process	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Greater than $60k schedular payments but not GST registered	Automatically issue letter when business rules or conditions are met.	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Credits on ceased entities and deceased customers	Automate the transfer of credits from closed or deceased accounts	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Payments refunded to customers	Optimise the redirect payment rules	Active	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Overdue income tax return RIT prediction	Predicts residual income tax (RIT) on overdue returns.	In production	Machine Learning	Snowflake	Refused under section 18(g).
Payment Service Providers Data Transfer	Automate data transfer	In Production	Automation	Snowflake, FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Phishing Automation	Automate aspects of our response to phishing attempts	In production	Automation	N/A	Free of charge.
Posit Connect	Supports the deployment of AI-powered data science solutions.	Available	Generative	Posit Software	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Positions Based Systems Access	Automatically receive the necessary access on day one of starting a new job	In production	Automation	One Identity	Withheld under section 9(2)(b)(ii).
Pou Here Tangata Project	Modernise voice services by moving to Genesys Cloud and introducing integrated planning across voice and non-voice channels.	Active	Automation	Genesys, OneNZ, Verint Systems Inc, FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Power Automate	Low Code solution that supports automating tasks.	In production	Machine Learning	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Propensity to read letter or log-in to myIR	Supports IR in using the right channels to reach customers.	In production	Machine Learning	Snowflake	Refused under section 18(g).
Qualtrics	Analysis of survey information from customers.	In production	Machine Learning	Qualtrics	Withheld under section 9(2)(b)(ii).
Qualtrics User Provisioning and Deprovisioning Project	Leverage existing provisioning and deprovisioning processes to manage Qualtrics access.	In production	Automation	Qualtrics, Azure, Oracle, One Identity	Withheld under section 9(2)(b)(ii).
Receipt, invoice, statement and tax/employer return review	Text recognition	In production	Optical Character Recognition Machine Learning	FAST Enterprises LLC	Refused under section 18(g).
Redwood Initiative stages 1-9	Transition to Oracle's Redwood UX to provide a modern, consistent experience across cloud applications (HCM, ERP, SCM, CX)	Active	Machine Learning, Automation	Oracle	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Resource Information Role	Leverage existing provisioning and deprovisioning processes to manage role.	In production	Automation	Ernst and Young, Accenture	Withheld under section 9(2)(b)(ii).
Security Card generation for new worker	Apply automation to streamline aspects of the process	In production	Automation	ServiceNow, Accenture, Deloitte	Withheld under section 9(2)(b)(ii).
Security Workflow Automation Service	A workflow platform to design and manage complex automation workflows	In production	Automation	Tines	USD$50,000
ServiceNow Accelerators	Enhances IT service management and helpdesk functionality with automated workflows.	Proof of concept	Generative	ServiceNow	Withheld under section 9(2)(b)(ii).
ServiceNow migration of UIF to ConnectQuick Integration Framework	In-built automation capabilities	In production	Automation	Deloitte	Withheld under section 9(2)(b)(ii).
SharePoint Advanced Management (SAM)	Analyses site activity to detect unusual patterns.	In production	Machine Learning	Microsoft	Part of M365. Copilot licence Withheld under section 9(2)(b)(ii).
SharePoint Sections with Copilot	Helps create structured sections for SharePoint pages.	In production	AI agent	Microsoft	Part of M365. Copilot licence Withheld under section 9(2)(b)(ii).
Small Business Cashflow Top-up Loan default	Automating existing process.	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Small business Cashflow Loan decision support updates	Automating parts of the eligibility review process	In production	Automation, decision support	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Snowflake Cortex AI	Enables users to query data in natural language and get direct answers	Proof of concept	Generative	Snowflake	Refused under section 18(g).
Snowshare	Secure automated data exchange across systems	Proof of Concept	Automation	Snowflake	**Cost of credits:** Refused under section 18(g) **Cost of professional services from SAS during PoC:** Withheld under section 9(2)(b)(ii).
Speak2IR address validation	Enables automatic address updates	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
START system access provisioning and deprovisioning	Automation of START system access based on positions based roles	In production	Automation	FAST Enterprises LLC, One Identity	Withheld under section 9(2)(b)(ii).
Streem	Media monitoring	In production	Machine Learning, Natural Language Processing	Streem NZ Limited	$17,970
Student Loan international case – Third party provider files not automatically closing	Automating parts of the existing process.	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
Tableau Desktop	A data analytics and visualisation tool for analysing and graphing performance test results	In production	Machine Learning Natural Language Processing	Tableau Software	Withheld under section 9(2)(b)(ii).
Transitional process for non-active trust declarations	Scan job to automatically add specific indicator	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
TRM	Blockchain analysis	In production	Machine Learning	TRM Labs, Inc.	Withheld under section 9(2)(b)(ii).
Use of money interest – generic emergency response	Predefined automated base process to be used in the event of an emergency	In production	Automation	FAST Enterprises LLC	Withheld under section 9(2)(b)(ii).
Visitor Kiosk	Cloud-based visitor management system (VMS) that replaces manual sign-in with a secure, digital process.	In production	Automation	Aotea security / HealthSafe	$46,940
Viva Topics	Delivers personalised content and community recommendations	In production	Machine Learning	Microsoft	Part of E5 licence. Withheld under section 9(2)(b)(ii).
Windows365	Cloud-based virtual desktop service	In production	Automation	Microsoft	Withheld under section 9(2)(b)(ii).
Windows Hello for business	Secure authentication for IR-enabled devices	In production	Biometrics	Microsoft	Withheld under section 9(2)(b)(ii).
WinGet	Automated application packaging	In production	Automation	Microsoft	Free of charge.
Z scaler	Detection and classification of web traffic and websites.	In production	Machine Learning	Z Scaler	Withheld under section 9(2)(b)(ii).

Name	Description	Product/ Project Status	Technology	Vendor/Delivery Partner	Cost to IR
ZoomText	Screen magnification software designed to improve accessibility	In production	Optical Character Recognition	Vispero	Withheld under section 9(2)(b)(ii).

Appendix B
Item 1

Inland Revenue
Te Tari Taake

Inland Revenue report: Update on Inland Revenue's use of Artificial Intelligence

Date:	26 February 2025	**Priority:**	Low
Security level:	In confidence	**Report number:**	IR2025/068

Action sought

	Action sought	Deadline
Minister of Revenue	**Note** the contents of this report	NA

Contact for telephone discussion (if required)

Name	Position	Telephone
Brijesh John	Domain Lead, Strategic Architecture	9(2)(a)

26 February 2024

Minister of Revenue

Update on Inland Revenue's use of Artificial Intelligence

Purpose

1. This report:

   1.1. Provides an update on Inland Revenue's use of Artificial Intelligence and the approach we are taking.

   1.2. Our last note (BN2024/276) detailed the many uses of Artificial Intelligence in Inland Revenue, and this report provides an update on progress in this area.

Background

2. Inland Revenue's business transformation programme has left Inland Revenue with excellent digital capability and very good data across our broad customer base.

    2.1. The quality of our data is very good, complete and timely, enabling us to take advantage of AI to achieve further efficiencies and improve the customer experience.

    2.2. Inland Revenue is taking a benefit and business value centric approach to extending the use of AI, focussing on supporting decision making in the management of tax and social policy compliance risk, intervention design and increasing staff productivity.

Key Highlight (Jan-Feb '25) – Wider use of AI for Voice Channel

3. Inland Revenue recently completed a proof of concept using Genesys Cloud AI that creates summaries of calls between a customer and Inland Revenue.

    3.1. This AI solution utilises call transcripts and generates a summary of the key points from the interaction with the customer which, once checked for completeness and correctness, can be copied to the customer record in START. Note the AI generated summary is not retained.

    3.2. Our proof of concept demonstrated an average 17% decrease in after-call work time, no degradation in quality of notes, and positive sentiment from our people who trialled this feature.

    3.3. The Conversation Summarisation performed particularly well with our general voice calls where time efficiencies were seen. The more complex calls still saw a decrease in after-call work time, but to a lesser extent than our general queues.

    3.4. 97% of our staff found the technology easy to use following a 30-minute training session.

    3.5. The plan is to roll out the tool to our Customer and Compliance Services – Individuals (CCS-I) team and continue pilots with our other business areas to ensure there is value across other customer cohorts before further rollout.

    3.6. The Genesys tool continues to evolve, which is expected given the rapid pace of AI developments. A new version has just been released which includes new functionality at a higher licence cost. We will review the new version to ensure the changes have not altered the value proposition of the summarisation feature and to assess possible benefits from the new functionality.

    3.7. It is important we ensure we are not constrained by vendor lock-in and that we consider an appropriate cloud licensing model – in this case a month-by-month license which will insulate us in a rapidly evolving market.

4. Future proofs of concepts utilising other AI functions of the new version will test the real-time presentation of contextual knowledge to our people to help them assist customer queries.

    4.1. The expected benefits we're seeking to confirm are a reduction in talk-time with our customers and decreased training time for our people.

Proofs of Concept/Pilot initiatives

5. Inland Revenue is looking at the use of AI across its Digital Estate. A draft view is provided in Appendix 2 (pg. 8)

6. In addition to the use of AI in our voice channel, the following use cases are being evaluated and rolled out at Inland Revenue.

Initiative	Key findings and next steps
Microsoft Copilot (Bing/Browser): AI-powered chat service (summarise large volumes of information, undertake research and proofread, edit and generate written content)	• 90% of users reported improved work quality and 86% of users reported time savings. • Tool deployed across the enterprise to appropriate roles and business groups, in a staged approach, with training mandatory before access is available.
AI test scenario generator tool AI-powered tool to create test plans and test scenarios	• Evaluation is currently underway via Proof of Concept. • This is the first private Large Language Model (LLM) for Inland Revenue, built with open-source tools by our Testing partner. • Early indications have shown positive results against defined measurement criteria, and if successful, this use case has potential to save significant internal costs by reducing effort on test planning and test-scenario writing, while increasing team satisfaction and maintaining evaluation quality.
Microsoft 365 Copilot: Copilot is integrated into the M365 suite of products (e.g. SharePoint, Word, Excel, Teams) and is designed to enhance staff productivity.	• Proof of Concept is in progress with early findings showing productivity gains. CoPilot has also provided initial validation that our internal knowledge and content structure and the associated infrastructure is appropriate and of high quality. • Next steps: Ensure appropriate auditing capabilities can be implemented and look at wider use cases / rollout if found appropriate, cost efficient and effective
Āwhina Mai Gen-AI assistant: We are evaluating the use of GenAI for delivering summarised information for customers using our website as part of the Āwhina Mai project	• A Proof of Concept was conducted to test summarisation and tagging capabilities across content held in PDF and html format in our website. • While the AI capabilities are available, further work needs to be done to address the underlying content structure, so it is better able to be consumed by GenAI tools

AI Capability Uplift

7. Through the process of Inland Revenue establishing a structure and governance for AI, along with strategic considerations and a roadmap, 'People' were identified as the most crucial area for investment in our AI journey.

8. Leaders at Inland Revenue have engaged in AI Fluency training. The purpose of AI Fluency training is to give a suitable understanding of AI to those making decisions on its use.

    8.1. Inland Revenue's Executive Leadership Team participated in an AI Fluency workshop, specifically focused on governing a large organisation with a technology change happening so quickly.

    8.2. All staff have access to seven e-learning modules: Introduction to AI, Gen AI 101, Value Capture, Trust, Ethics & Governance, Prompt Training, Risks of AI and GenAI, Threats of AI, and Data & Information Management.

9. Inland Revenue is considering further updates to its AI literacy by including Agentic AI to its AI literacy programme.

OECD engagement

10. Inland Revenue prioritises international engagement on AI, specifically on international rules and norms. Inland Revenue is part of the OECD Cluster focused on using AI in this field. The OECD's Project D aims to enhance the trustworthy use of AI through international collaboration, developing a Trustworthy AI framework, and sharing best practices.

11. The project targets several impacts including reducing taxpayer burden, optimising internal operations, increasing AI trust, and prioritising AI investments. New Zealand supports these activities to understand AI deployment in different jurisdictions and utilise the generated assets.

12. Inland Revenue has written to OECD to express interest in continued involvement in the next phase, with a focus on learning and evaluating appropriate AI frameworks in the context of Tax Administration.

Public Sector and Government Chief Digital Officer (GCDO) engagement

13. Inland Revenue is working with the GCDO and sharing insights from its AI and digital journey through the Government AI Community of Practice.

14. Inland Revenue has shared relevant artefacts related to AI with the GCDO and is in the process of sharing its eLearning material with the Public Service Commission.

15. Inland Revenue is also considering how best to contribute to the GovGuide digital front door initiative led by the GCDO.

16. Inland Revenue will be contributing to the National AI Strategy as necessary and appropriate.

17. Inland Revenue is working hand in hand with the GCDO and have recently presented to the New Zealand Government Community of Practise. Two key excerpts from the presentation are attached

    17.1. Foundations and Key considerations for AI Solutions- Appendix 1 (pg. 7)

    17.2. Key Inland Revenue Reflections and Learnings – Appendix 3 (pg. 9)

18. Inland Revenue officials have also met with 15 agencies to both share resources and explore opportunities to work together. Anecdotal feedback from agencies suggests Inland Revenue

are ahead of the curve in terms of strategic thinking, governance processes and laying the foundations for safely trialling and scaling use of AI.

Key risks and challenges

19. Whilst there are many risks and challenges presented by Generative AI (for example bias, hallucinations, or discrimination), Inland Revenue's key risk is that of data and information leakage. This could compromise our obligations under legislation to protect personal information and Tax secrecy.

20. The key risk of data and information leakage manifests itself in two ways:

    20.1. Data input into an AI engine: Where Inland Revenue is applying AI capabilities internally, careful consideration is given to the data that is input into the AI engine (e.g. is it used to train a public facing AI engine? are there jurisdictional/legal considerations?) as well as applying our standard controls and monitoring.

    20.2. Publicly available AI which can be accessed by Inland Revenue staff either intentionally or inadvertently: While Inland Revenue has an AI Staff Use Policy and standards in place addressing this, additional controls have been put in place to identify and block access to known high risk sites as well as monitoring access to emerging AI capabilities.

21. The key challenge we face is the pace at which AI is proliferating and evolving both internally and externally.

22. Internally, the speed at which new and potentially beneficial uses of AI are evolving tests our ability to assure that risk controls remain relevant, provide mitigation and meet monitoring thresholds. s6(c) [redacted]

23. External challenges are wide ranging from accuracy of content used to train publicly available AI to "bad actors" use of AI to enable scams, frauds, and technical attacks against our infrastructure at unprecedented scale and pace. Inland Revenue has established risk frameworks, referencing guidance from the System Leads, which are used to evaluate and respond to these varying threats as rapidly as possible.

Key next steps/focus areas

24. We continue to explore opportunities with our core Tax solution partner, Fast Enterprises in developing more intelligent models based on the data we already hold in the Tax system.

    24.1. We are currently working with the FAST development centre in the USA to develop NZ specific decision support models within START. This will allow us to better target debt collection activities.

    24.2. FAST have an 'outsourced' collection service operating in the US that utilises a range of 3rd party data sources, and Artificial Intelligence to improve debt collection. We are exploring options of establishing a comparable service specific to NZ.

    24.3. We are actively exploring how we will use additional data, such as payment service provider data within these models.

25. We are exploring how we can use AI to assist staff in responding to non-voice queries, i.e. web requests, emails etc. We believe we can achieve efficiencies like those we have found in AI work in the voice channels space.

26. We are actively investigating possibilities for using AI to address key business outcomes such as

26.1. Assuring Revenue and Increasing Compliance

26.2. Debt collection and management

26.3. Increasing efficiencies in processing activities (Frontline and Back-office)

26.4. Protecting the Revenue System from AI powered attacks

27. We are evaluating and considering the use of Agentic AI to increase automation possibilities across business areas to support automation and increase efficiencies.

28. Inland Revenue is developing a strategic roadmap to create and manage our publicly available content so that it is fit for the future and better able to be consumed and used by AI tools.

28.1. We are currently developing the roadmap options, staying connected with related AoG work (such as the DIA-led GovGuide Digital Front Door to NZ Government), and progressing relevant internal proofs of concepts.

28.2. Initial insights show individuals and business (i.e. Law Cyborg) are already accessing our content via third-party Generative AI with some level of success. There is a need to act quickly to remove complexity including duplicated content which can negatively impact the results returned via Gen AI.

28.3. We believe a future opportunity is using AI to provide contextual and targeted responses, i.e. the merging of general content (potentially from across government), with specific customer and transactional information.

Recommended action

We recommend that you:

1. **Note** the contents of this report.

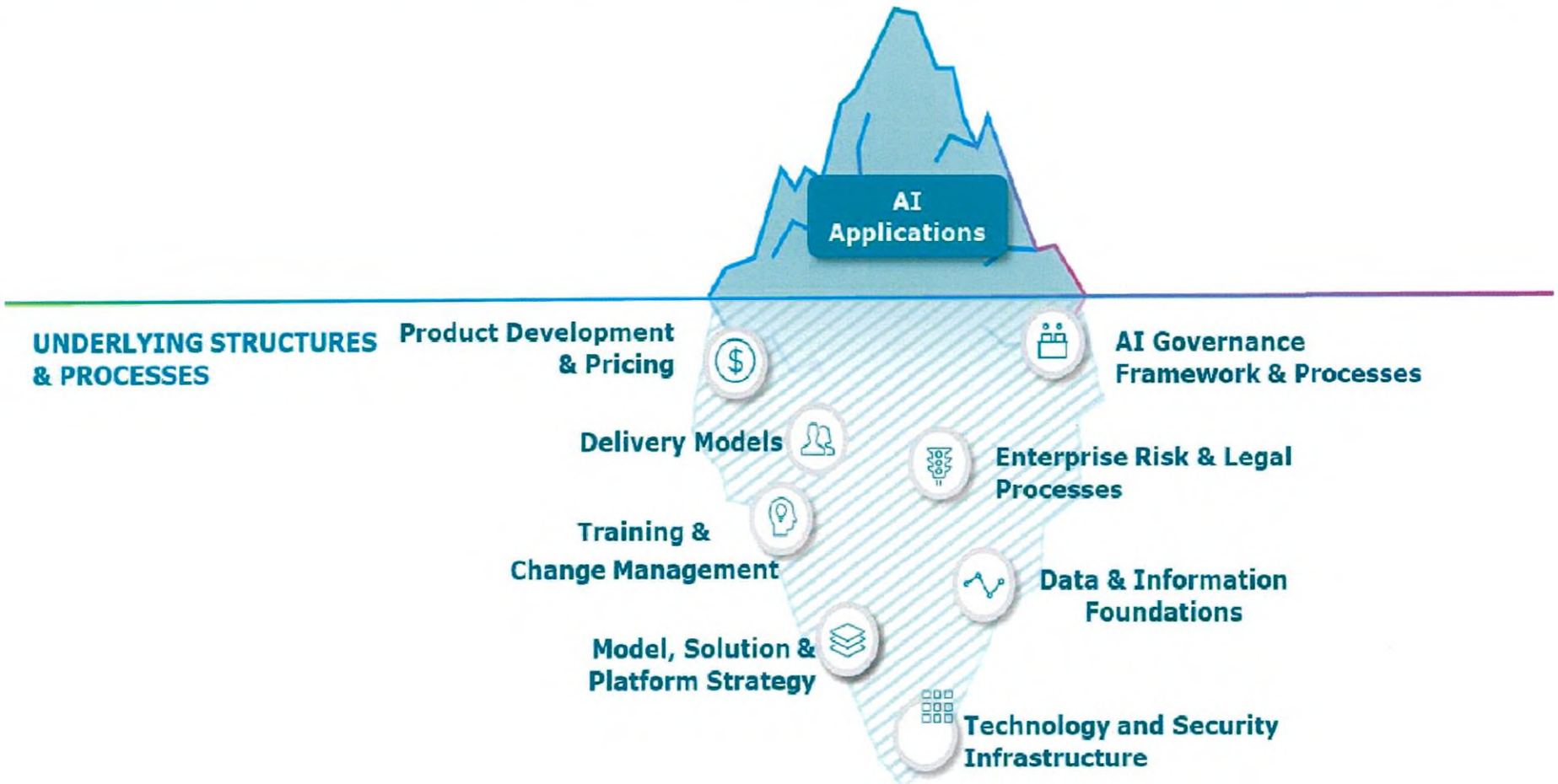
Noted

9(2)(a) [redacted]

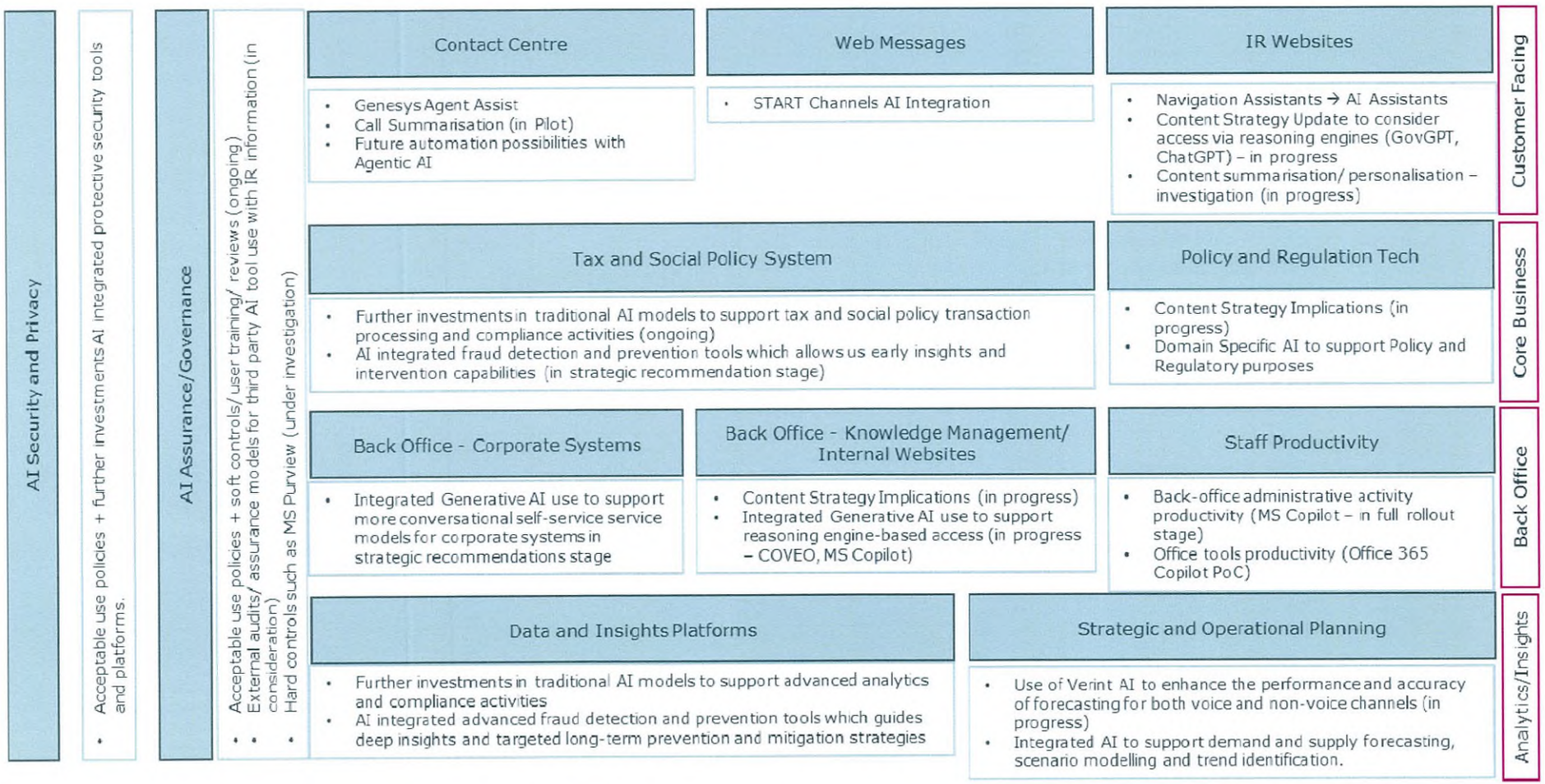
Brijesh John
Domain Lead, Strategic Architecture

Hon Simon Watts
Minister of Revenue
/ /2024

Appendix 1: Getting Foundations Right for AI



Appendix 2: Inland Revenue Digital Estate with an AI perspective



Appendix 3: Key Inland Revenue Reflections and Learnings

- Focussing on people
  - Leadership Support
  - Building Capability
  - Developing common understanding and demystifying AI
- Focus on value
  - Being specific on which workflows of the organisation that the organisation (and you want AI to help) can gain most value from
  - Measurable business outcomes – The technology is proven, you are trying to prove business value
  - Having assurance models and measurement frameworks
  - Building capabilities with a focus on scaling (enterprise scale is where value is but not many organisations are able to scale use of AI)
- Beware of AI Washing/ Trend jacking
  - Not all products contain AI or 'Generative AI/Agentic AI', even though they claim it
  - Assuming you have AI in product/platform without investigating it fully, could lead to you applying the wrong assurance yardsticks
  - On the flipside, Generative AI excitement may help open the door to traditional AI and other solutions
  - It is critical to keep the focus on the business outcomes and apply the correct digital solution to it (a type of AI or other technologies as may be relevant)
- Learnings from OECD Work on Trustworthy AI for Tax Departments - Survey Summary. (Inland Revenue participates in this group)
  - Traditional AI: Primarily used in compliance, often for ranking taxpayer populations based on historical data.
  - Generative AI: Being tested in customer service and internal management, with many LLM-based tools still in experimental phases.
  - Benefits: Quantifying the benefits of AI in tax administration remains challenging due to the sensitive nature of compliance work and the experimental nature of many generative AI tools.
  - Implementation Challenges: Implementing AI effectively requires addressing complexities such as executive support, change management, and technical architecture.

Item 2

Inland Revenue
Te Tari Taake

Inland Revenue report: Update on Inland Revenue's use of Artificial Intelligence

Date:	27 August 2025	**Priority:**	Low
Security level:	In confidence	**Report number:**	IR2025/365

Action sought

	Action sought	Deadline
Minister of Revenue	**Note** the contents of this report	NA

Contact for telephone discussion (if required)

Name	Position	Telephone
Brijesh John	Domain Lead, Strategic Architecture	9(2)(a)

27 August 2025

Minister of Revenue

Update on Inland Revenue's use of Artificial Intelligence (AI)

Purpose

1. This report

   1.1. Provides an update on Inland Revenue's use of AI for the period June to August 2025.

   1.2. Builds on the previous note to you.

      1.2.1. IR2025/229 provided an update on AI work in Inland Revenue for the period March to May 2025

   1.3. We propose our next planned update to you will be in November 2025. You will be advised about any significant changes or advancements in the interim via regular Ministerial Services channels.

Background

2. Inland Revenue's Strategic and Investment Board approved a refresh to our strategy and direction regarding the use of AI, in September 2025.

   2.1. Considering the significant transformative potential of AI, our path ahead is through continuous AI-driven transformation which will enable *higher staff productivity and increased capacity* and will enable *data-driven intelligence* and *efficient delivery of better customer experiences.* Our approach must be one that gains value through increased public trust, scaling through value and transforming with people still at the heart.

   2.2. Considering the rapid advances and changes in AI technology, uncertainty around geopolitical and socio-economic conditions and the evolving global regulatory landscape, a different approach is required for our AI Strategy. Taking into consideration the flexibility and ability to pivot that is critical, we need a strategy that takes an agile and adaptive approach.

   2.3. Our AI vision is to transform tax and social policy administration through AI that simultaneously delivers the accessible, transparent, and culturally responsive services New Zealanders expect from government while achieving the operational excellence, productivity gains, and stewardship effectiveness that enables sustainable public service delivery.

   2.4. We are now working through the next steps to enable outcomes in alignment to our strategy including addressing capability and capacity related questions.

Key Highlights (Jun-Aug '25) -

3. Approval of Strategy and Direction for AI with a focus on 3 organisational shifts powered by AI

   3.1. Workforce Productivity,

3.2. Data Driven Intelligence (focussed primarily on compliance), and

3.3. Customer experience.

4. Use of Cortex AI to allow the business to 'ask questions of our data'

4.1. Initial use case with audit data had good outcomes and was presented at the AI Accelerate day at Parliament.

5. We have begun using more sophisticated analytics models in our Tax System which help us to identify the debt collection action which is most likely to succeed for a customer.

5.1. As the first step, we are using intelligent models which assess the customer's circumstances and compliance history and recommend the best action for that customer. Actions include:

5.1.1. issuing a bank deduction notice,

5.1.2. offering a pre-approved instalment arrangement, or

5.1.3. remaining with current collections processes as the customer's debt will self-resolve.

5.2. These models take into consideration data and information across IR.

5.3. The models have been very successful with over $39M under pre-approved instalment arrangements already and over $12M in bank deductions recovered. This was achieved within a 10-week delivery time frame.

Proofs of Concept/Pilot initiatives

6. The following use cases are being evaluated and rolled out at Inland Revenue.

New initiatives since our last briefing note

Initiative	Details
Voice Channel: Supervisor Assistant Utilising AI to analyse call transcriptions, to evaluate quality of customer service.	• An AI agent will compare transcriptions of calls from Genesys Cloud against internally set criteria to evaluate the quality of a phone conversation with a customer. • This output can then be used to support individual staff development discussions between Team Leads and their direct reports. • Proof of concept is in progress; with time savings expected for Team Leads.
Data Intelligence Platform: Natural language querying Evaluating the capabilities of Snowflake Cortex AI to potentially supplement or replace data analysis tasks and pre-defined reports.	• The proof of concept will evaluate the natural language querying capabilities of the tool, within Inland Revenue's Data Intelligence Platform. • The insights gained from this trial will help us identify potential use cases for the platform and define the process for future AI adoption. This includes assessing the feasibility and alignment with our strategic goals. • Expanding the accessibility of the platform through AI will enable us to explore future possibilities

Initiative	Details
	including increasing insights to allow us to identity and tailor interactions with our customers. • In the future, with the addition of other features available in the platform, we can consider automation possibilities at scale, improving efficiency and effectiveness across various business functions.
Legal Research Summarisation Provides AI-powered support for legal and research workflows by enabling natural language queries, case and document summarisation, and efficient information retrieval.	• Beta testing a generative AI tool to assess usability and alignment with our research workflows (Lexis+AI). • Features include "ask a legal question", case summarisation, and document interpretation via natural language prompts. • The beta test will leverage existing unclassified data as text prompts. • We are measuring efficiency through automating repetitive and time-consuming research over complex legal content, accuracy, scalability and innovation.

Update on initiatives previously reported on

Initiative	Key findings and next steps
Enterprise Services Technology AI Agents Tailored and purpose-built AI Agent solutions that use internal information across corporate platforms to enable better decision making and possible automation.	• A custom AI 'Technology Agent' was built using Copilot Studio to answer IT queries and raise ServiceNow tickets. It securely accessed IR's internal data in a test environment, respecting permissions and confirming feasibility. • A limited production pilot has now commenced, and a small group will use the agent for live IT support, with monitoring focused on accuracy, usage and risk controls.
Knowledge Surfacing The real-time presentation of contextual knowledge to our people, to help answer customer queries.	• The pilot has concluded and showed no measurable improvement in call handling time or efficiency on the voice queues we tested it on. • Experienced staff rarely needed the surfaced articles. • The AI occasionally surfaced irrelevant content due to overlapping articles in our knowledge base. • The vendor is launching a connection with SharePoint shortly, which will reduce the manual data transfer necessary. Once this is in place, we will consider trialling the feature with our contingent workforce and new staff.
IT Service Management (ServiceNow Platform)	• Inland Revenue's ServiceNow AI pilot, running from April to September 2025, is still in an experimental demo environment with no IR data,

Initiative	Key findings and next steps
	and findings will be reviewed after the trial to guide any future adoption.
AI test scenario generator tool AI-powered tool to create test plans and test scenarios.	• The six-month pilot began in July 2025 and is expected to conclude in December 2025, when the evaluation will wrap up and results will be reported.
Microsoft 365 Copilot: Copilot is integrated into the M365 suite of products (for example, SharePoint, Word, Excel, Teams) and is designed to enhance staff productivity.	• Following a successful 12-month pilot of 20 licenses, we are now scaling the pilot and have extended to 100 licenses, through to October 2025. • The licenses are in back-office functions: Policy, Tax Counsel Office and Enterprise & Integrity Services. • In addition to the standard M365 Copilot package, we are trialling three pre-built agents: Analyst, Researcher and Project Manager. • The extended pilot in Policy and Tax Counsel Office is testing Copilot with budget-sensitive information under strict conditions—an approved exception to our AI use policy—to assess its handling of highly confidential content, before we consider a wider roll out.
Voice Channel: Conversation Summarisation Creates summaries of calls between a customer and Inland Revenue.	• Conversation summarisation has been fully adopted in the Individuals Segment contact centre since April 28, 2025. • We are now exploring pilots for our Business and Families customer segments.
Microsoft Copilot (Bing)	• This continues to be rolled out in a staged approach across Inland Revenue, to ensure our people have the support they need to use the tool successfully. • 1990 users currently have access.

AI Capability Uplift

7. Inland Revenue continues to adapt its AI fluency training, as new AI approaches emerge. More recently, this has included the emergence and proliferation of Agentic AI.

8. We have delivered targeted Agentic AI fluency training to select leaders to strengthen oversight and strategic understanding.

9. A new Agentic AI e-learning module has also been developed to complement existing AI fluency content, helping business leaders identify risks, opportunities and ethical considerations.

10. The initiative is supported by a cross-functional team from Inland Revenue and Deloitte, with delivery scheduled from July to September 2025.

11. The e-learning module will be shared shortly with GCDO, to join Inland Revenue's other AI e-learning modules already available to the wider public service.

OECD engagement

12. Inland Revenue's active participation in OECD forums ensures New Zealand's perspectives are reflected in the development of international standards for responsible AI use in tax systems. This engagement also enables Inland Revenue to stay abreast of global best practices and apply relevant insights to strengthen our own AI governance and implementation.

13. Inland Revenue has consulted on OECD Project D "Enhancing the trustworthy use of AI in Tax Administrations". The project's draft report presented an approach to navigating use cases for AI and a draft assurance checklist specific to AI use cases known across tax administration sectors.

14. We've shared feedback confirming Inland Revenue's support for the Trustworthy AI in Tax Administration material that accompanies the assurance checklist, which will aid our efforts. However, the draft assurance checklist is overly detailed, which may hinder swift progress. We have proposed specific changes to allow tax administrations to adjust their pace based on risk.

Public Sector and Government Chief Digital Officer (GCDO) engagement

15. We continue to actively share our knowledge and resources with the GCDO and a range of agencies to support the wider public service's uptake of AI. This includes governance instruments, reusable use case specific documentation, AI evaluation approach and specification, proof of concept findings and AI literacy training material.

16. Inland Revenue officials attended the AI Accelerate workshop held at Parliament this month, showcasing our use of AI in the contact centre and in our Data Intelligence Platform.

17. Our refreshed AI Strategy and Direction has been shared with GCDO officials.

18. Inland Revenue officials have been working alongside GCDO to publish Digital Fluency Case Studies and also present these to the Australian Public Service Commission.

Key risks and challenges

19. Social acceptance represents the cornerstone requirement for AI transformation success, with government AI initiatives facing significant citizen concern rates that require proactive trust-building. Without sustained public confidence and staff engagement, even technically superior AI systems encounter implementation barriers that undermine strategic objectives. The stakes are particularly high for government AI deployment, where public trust directly impacts New Zealand's global reputation and long-term citizen compliance.

20. Generative AI is being used to enhance traditional attack methods, such as phishing, reconnaissance, and malware development, but there is limited evidence of novel or uniquely AI-driven threats.

21. Although there is currently limited evidence of novel or uniquely AI driven threats, the rate of change in AI-based cyber threats is increasing, which in turn will mount pressure on the requirement for frequent reviews of existing controls, updates to threat intelligence sources, and continuous staff upskilling.

22. The adoption of AI agents internally and by customers is introducing new systemic risks in data handling, service interactions, and compliance, as these agents can autonomously access and act on sensitive information.

23. Third-party use of AI could introduce risks around customer data that will have to be carefully managed considering a balance between potential productivity increase and privacy, security and bias related issues. IR must consider this from a partner, vendor, digital service provider and intermediary perspective.

24. Research shows us that value from AI is mostly seen in organisations which invest in the integration of AI in core business activities. Inland Revenue will have to make trade-off decisions as core business resources will be required in addition to hard-to-find expertise in the market to support strategic shifts associated with core business activities. A concerted focus on these shifts will also require dedicated funding.

Key next steps/focus areas

25. We are considering improved and AI-powered OCR in the Tax system for the Donation Tax Credit (DTC) and FamilyBoost (FB) products to increase automated processing of claims

26. We are putting concerted focus and effort on considering, trialling and implementing AI that can be used to support our approach to managing debt and improving compliance outcomes.

27. In alignment to our AI strategy, this includes integrating rule-based automation, machine learning, generative AI and optimisation techniques to deliver superior outcomes. Our approach is focused on the four key areas of our debt management approach areas and will explore options such as:

    27.1. Changing the system: We will consider AI-enhanced policy simulation, increased use of predictive modelling, improvements to automated deduction-at-source and scenario testing using agentic AI.

    27.2. Minimising what becomes debt: We will consider behavioural nudges, improved access and responsiveness through chatbots (multilingual), risk scoring, partner-integrated campaigns and early warning systems.

    27.3. Maximising what we collect: We will consider AI-optimised payment arrangements, prioritised collections, third-party performance analytics and penalty impact modelling.

    27.4. Minimising write offs: We will consider early alerts for legal escalation, AI-assisted litigation preparation, unified case management and improved cost-benefit analysis of enforcement.

    27.5. We are working through the roadmap for enabling outcomes including

        27.5.1. Preparing our data and information sets for AI,

        27.5.2. Creating a debt and compliance focussed stream within our AI innovation lab, and

        27.5.3. Embedding governance, privacy and ethical safeguards with a goal of increased public trust

28. We are initiating work to introduce an AI innovation lab which will allow us to safely test AI integration into core business outcomes in alignment to our strategy in a safe and responsible manner.

29. We continue to focus on core business areas including:

    29.1. Contact centre – automation and productivity increase

    29.2. AI driven insights, decision intelligence and analytics, including those used for:

        29.2.1. Financial crime prevention/ fraud detection and prevention

        29.2.2. Protecting the NZ revenue system

    29.3. Ensuring Inland Revenue's external content is fit for the future, including:

        29.3.1. Driving content quality for accurate consumption via reasoning engines (AI tools)

        29.3.2. Using AI to support the creation, design and management of this content

29.4. Administrative productivity including:

29.4.1. Policy analysis and drafting

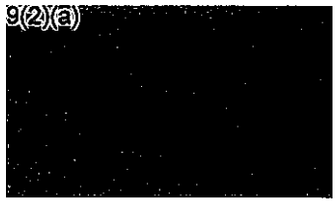
29.4.2. Enterprise support services

Recommended action

We recommend that you:

1. **Note** the contents of this report.

Noted



Brijesh John
Domain Lead, Strategic Architecture

Hon Simon Watts
Minister of Revenue
/ /2025

Item 3

Inland Revenue
Te Tari Taake

Benefits of Copilot for IR

Internal

Authors: [redacted] Crawford, Makayla Stewart with contributions from Phil Whittington

Version: Based on mid-pilot and final survey for second pilot group

Date: November 2024

Overview

- Since March 2024, IR has been running a pilot to trial **Microsoft Copilot**, a generative AI tool that creates text and images based on a user's prompt.
- The purpose of the pilot is to **explore** how this generative AI tool might help IR people be more productive in their roles.
- This report summarises the **benefits of Copilot** that IR people taking part in the pilot identified.
- Currently 116 IR people have access to the tool. This includes 80 people in the second pilot group that has staff from Policy, Tax Council Office, Enterprise Services and Enterprise Design & Integrity.
- This version of Copilot has no extra licence cost.
- ACC has completed research on the benefits of Copilot for Microsoft 365. This report is focused on Microsoft Copilot (formerly Bing Chat).

Key benefits and use cases

Inland Revenue
Te Tari Taake

Productivity benefits

75% of people said Copilot helped them be **more productive**

72% reported **better quality of work**

73% said Copilot helps them **focus and unblock** when stuck

85% found time savings by summarising content

93% reported **time savings per week**

57 hours per week in time saved across the pilot group

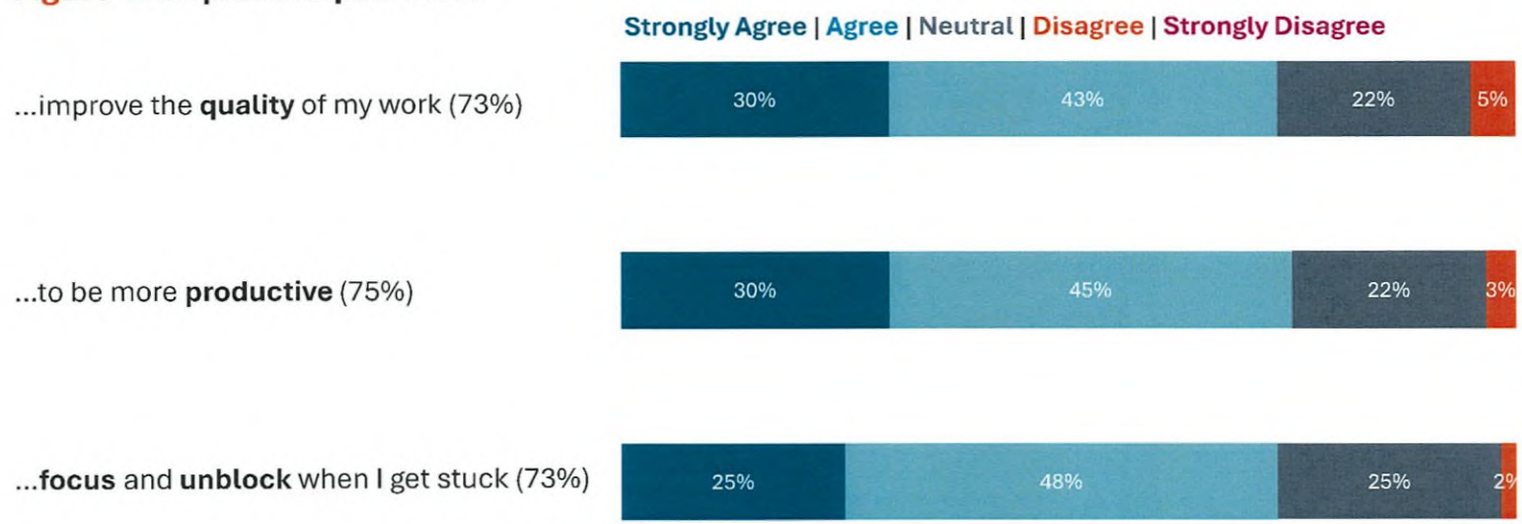
57 minutes saved each week per pilot participant on average

22% saved **more than 2 hours per week**

Data source: Mid-pilot survey for second pilot group. 60 pilot participants responded to the survey (78%).

Very high proportion of participants in the pilot reported significant benefits

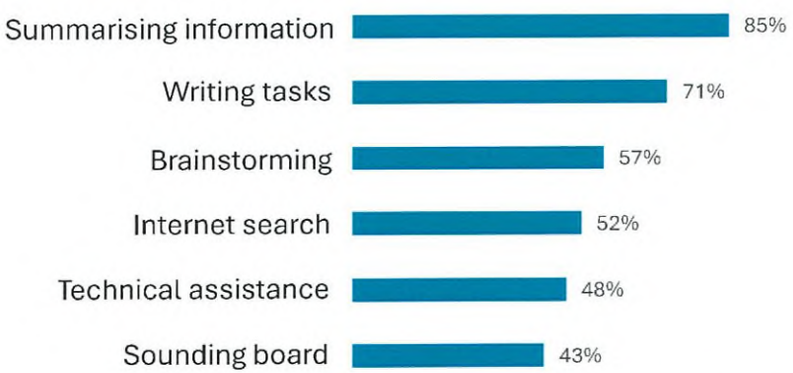
Figure 1: Copilot helped me...



Internal Revenue Department. (2024). Mid-pilot check-in survey: 60 responses. Internal document.

Participants saved time across a broad spectrum of professional activities

Figure 2: I use copilot for...

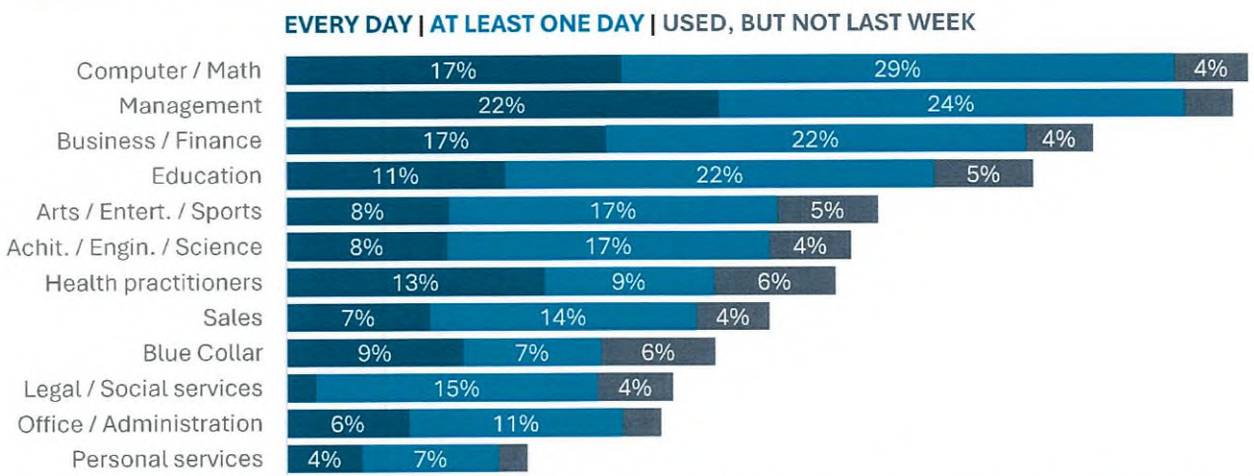


Internal Revenue Department. (2024). Mid-pilot check-in survey: 60 responses. Internal document.

- Participants reported use across a **wide variety of tasks**
- In the short term, productivity gains are likely to come from **use as a knowledge-work complement**

Our results are consistent with academic literature finding that Generative AI use is high in knowledge work

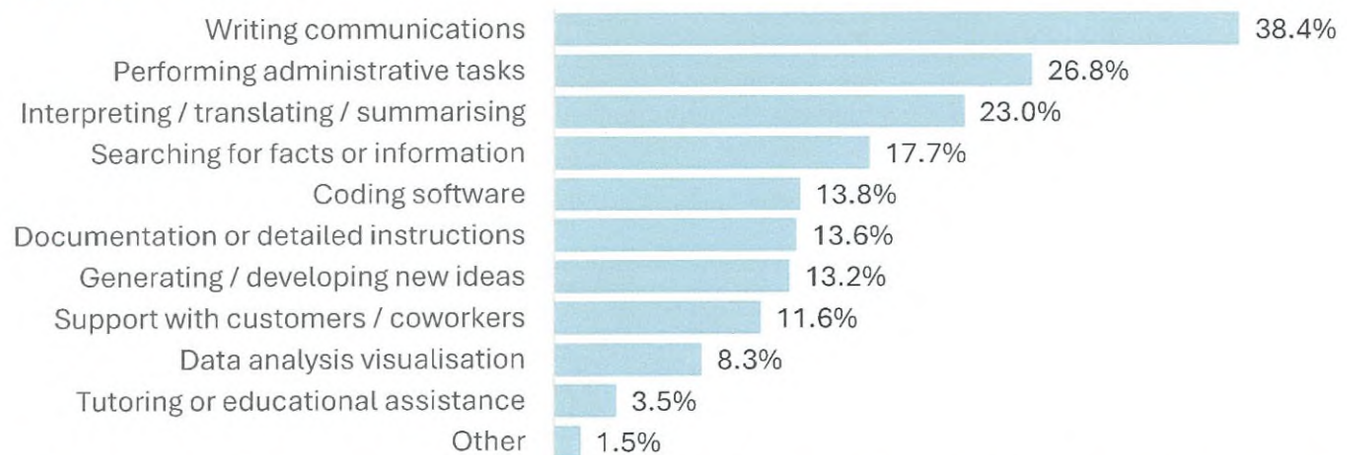
Figure 3: GenAI frequency of use in occupation groups "last week"



Bick, A., Blandin, A., & Deming, D. J. (2024). The rapid adoption of generative AI. National Bureau of Economic Research. https://www.nber.org/papers/w32966

Reported use cases in the literature broadly map to what we found with our pilot

Figure 4: In which specific tasks is AI most useful at work? (Share ranking task #1-2)



Bick, A., Blandin, A., & Deming, D. J. (2024). The rapid adoption of generative AI. National Bureau of Economic Research. https://www.nber.org/papers/w32966

Key use cases apply across a wide variety of roles

Summarising information: Summarising web content or user notes to save time on policy research, troubleshooting, and creating concise internal comms

Writing tasks: Assisting with drafting, phrasing, and enhancing clarity in communications including emails, documents, agendas, and internal comms

Brainstorming: Generating ideas, overcoming writer's block, providing a starting point for tasks

Internet search: Summarising relevant information for specific searches, eliminating the need to trawl through numerous pages, providing understandable explanations

Technical assistance: Assisting with troubleshooting, suggesting Excel formulas, validating solution designs, giving best practice tips

Sounding board: Acting as a second opinion to confirm understanding, bounce ideas off, and refine approaches

> Being able to **get a summary** or **find relevant information** quickly has been great with **summarising of documents** and **legislation** saving so much time

Change Analyst, ED&I

> It has helped me to **navigate 'writers block'** and given me a jumping off point for written work. Has helped greatly to **free me up for other tasks** it's given me some more time back that I'd be spending juggling unproductively

Domain Specialist, ES

Copilot has been useful for disabled and neurodiverse IR people

Disability perspective, for our people who:

- have low vision, Copilot summarises large bodies of text saving them time.
- have difficulties typing, Copilot helps them write faster.

Neurodiversity perspective, for our people who:

- struggle with managing time and workload, Copilot can support them by streamlining daily planning allowing our people to focus more effectively making their workday more manageable and less stressful.
- get blocked or struggle with the blank page, Copilot can help them get unblocked when they are stuck.

As someone with **low vision**, I read slower than others. Being **able to summarise** means I'm a lot **more productive**

Analyst, CCS

I have **ADHD** and need to exert a significant amount of effort to manage my workload and time every single day. Copilot has **helped to significantly reduce** the amount of **time** I need to dedicate to manually figuring out how my days will best work

Project Team Member, ED&I

Additional benefits

Conclusion: The pilot has demonstrated significant benefits for IR. We expect time savings, and other benefits will grow as our people become more proficient with Copilot and integrate it into their daily workflows.

- **Positive impact on the work experience:** 81% of people said Copilot has positively, or somewhat positively, impacted their day-to-day experience of doing their work
- **Stress reduction:** Reducing work-related anxiety and stress by helping to balance heavy workloads, and freeing up time through the faster completion of mundane tasks
- **Diversability perspective**: Supports inclusivity and accessibility within our organisation by addressing unique needs of individuals
- **Time management:** Helping with organising daily tasks and managing time efficiently
- **Learning and upskilling:** Providing ideas and guidance on how to do things
- **Improved security:** Access to Copilot acts as a mitigation of the risk of staff using unapproved AI tools against policy, by providing an approved and safe alternative

"It's like having a **bright, creative, and supportive** colleague who is always ready to help, friendly, versatile, and great for bouncing ideas off all day long"

Information Specialist, ES

"By helping to **rewrite emails** I feel **more confident** that what I send out will be **understood**."

Architect, ED&I

Providing an approved generative AI tool reduces security risks from non-approved AI use

- High reported use of AI tools for work internationally[1] clashes with **strict rules** many organisations report publicly. A plausible theory is that many workers around the world are using **non-approved AI tools**
- A **growing concern** is that employees experiment with AI tools but **do not disclose** this due to fear of **punishment**, loss of **respect**, or **job security concerns**[2]
- Allowing **sanctioned AI tools** helps maintain **information security and compliance**, as employees are less likely to resort to potentially insecure, unapproved AI solutions

[1] See *Bick, A., Blandin, A., & Deming, D. J. (2024)*. Ibid, or Humlum, A., & Vestergaard, E. (2024, April 24). The adoption of ChatGPT. Interactive Research Briefs. Retrieved from https://bfi.uchicago.edu/insights/the-adoption-of-chatgpt/

[2] Mollick, E. (2024). AI in Organizations: Some Tactics. *One Useful Thing*. Retrieved from One Useful Thing.

Safe and secure

Our data and information will be safe and secure using Copilot through:

Enterprise data protection

Information input by the users is protected and encrypted. It cannot be seen by people outside of IR.

No access to our data

This version of Copilot does not have access to our internal information, only what is input by the user.

Our data is not used to train AI

Prompts and responses are retained in our tenancy, they are available for audit by CISO and Cyber Security.

Our people will be supported

Our people will agree to a set of 'Dos and Donts' for using Copilot, they then receive 2 hours of training.

Limitations

- Images that have text within them often have spelling mistakes.
- Te Reo Māori is not an approved language; Copilot can complete translations but cannot be relied upon.
- Its understanding of concepts and society are heavily influenced by its training data, the majority of which has been created outside of Aotearoa.
- As with all GenAI tools, Copilot can produce content that is biased, misleading or incorrect.

Conclusion

- Copilot has shown **great productivity gains and time savings** but has **no additional costs**.
- The **broad applicability** of the key uses for Copilot mean that the **productivity benefits and efficiencies can be found in a wide range of roles** across our organisation.
- The **quality of prompts are directly tied to the quality of the outputs** from Copilot, prompts with few details or generic will result in vague and generic outputs.
- Our people will need to be **upskilled in the foundations of Generative AI and prompt engineering**, this will be achieved through a combination of the AI fluency business leaders offering and the required Copilot eLearn and practical session.
- **Benefits from generative AI tools are likely to increase over time** the more proficient our people become at prompting.
- Conducting a **controlled pilot** within our organisation has **been instrumental in identifying use cases**. As this technology is new, the practical applications of Copilot were not fully understood until our people started using the tool themselves.

Appendix

Quotes from IR people taking part in the pilot

Inland Revenue
Te Tari Taake

Usage – visits to Copilot URLs

This table shows the number of times IR users accessed the Copilot URLs: copilot.microsoft.com or bing.com/chat.

The data shows:
- high usage by both pilot groups
- the pilot one group is actively using Copilot several months after their pilot started

Note: Pilot group one started in mid-March 2024. Group two started in early August.

Month (2024)	Actual users	Total visits to Copilot URLs	Avg visits per active user
June	65	58,557	900
July	59	72,999	1,237
August	131	198,505	1,515

Summarising information

- "It has saved me from having to sort through as much chaff to get the information I am looking for."
- "I like it can summarise a page when troubleshooting, very helpful not having to click through multiple pages and links"
- "It saves time. For example, often we will want to compare a proposal we develop to rules that exist in other countries. It can take quite a lot of time trying to find an explanation of other countries' rules. Copilot does this in seconds, and we can validate the information by following the sources it finds for us. This alone can save several hours of work."
- "I also use it to summarizes notes in the ticket or conversation over email which saves me at least 15 to 30 minutes.. This helps me a lot of productivity as I do get side tracked from the visitors due to the role."

Writing tasks

- "The hardest part with any writing task is the first draft. If you give the right prompts, Copilot is great for starting this for you."
- "I mostly use copilot with writing, when I am stuck on how to phrase something, or I feel like a sentence could be written with more impact, kind of using it like I would Grammarly."
- "I've had positive comments on the clarity of emails and documents."
- "I use it to create agenda items for my workshops, plain language my Viva engage posts for the People-Led networks that I support and also for ideas on how to engage with my stakeholders"
- "It has helped me to navigate 'writers block' and given me a jumping off point for written work eg summaries, assessments etc."
- "...it has saved me a lot of time from having to try and articulate phases and brainstorming. I put in my thoughts process and ideas, and it tidies them up for me. Such a handy tool"

Brainstorming

- "It saves me time and prompts thoughts."
- "Saves a lot of time. When I get stuck, instead of procrastinating, I can use copilot to give me ideas for how to take something forward."
- "I have also found that when I have a block and not sure where to start with something it has given me ideas and a place to start that has also saved time as I would have usually gone away and done something else and come back to that task."
- "I use it very much at just early stages of my thoughts and then I modify the information or use it in another way (e.g. mentally adding my own context to the info they give me). Really useful to 'unblock' stuck thoughts."
- "...quickly brainstorm ideas without interrupting a team member"

Internet search

- "I love that when I'm stuck...I can use Copilot who will give me details behind the answers to help me understand things better which is something that Google doesn't do."
- "It has made the task of searching for specific technical information (solutions / answers etc) much faster and less frustrating. ...It is on track to become my new 'search' when I need highly specific information."
- "For example, yesterday I wanted to look up a specific research method but found it more efficient to get copilot to summarise information for me rather than trawling through web pages and journal articles."
- "Copilot nicely provides an understandable (not too technical) summary response and also includes links to relevant resources for further details if the technical description/process is required."
- "...Would also find with the standard internet search that quite often I would start reading the documentation only to find in the last sentence that there is no solution. Copilot calls out that up front BUT offers alternative solutions that others have used as a workaround. This is SO VALUABLE."

Technical assistance

- "Copilot saved me tons of time looking for possible options and outcomes to achieve some technical troubleshooting."
- "I used Copilot to suggest formulas for analyzing the data in a large spreadsheet & it certainly saved me lot of time."
- "Very quickly coming up with numerical examples to explain things."
- "Copilot validated my existing thinking about the script design. Provided some tips based on common / best practice design patterns."
- "[Copilot] has been very helpful for seeking ways of how we can formulate to help transform the data in the right format, to visualize the data in a better way for linking into graphs/dashboards."
- "Copilot has been really helpful for converting some coding from 1 language to another. It does...sometimes get the syntax wrong. But still provides a really nice jump start for converting the coding over..."

Sounding board

- "[Copilot] was useful as a second opinion/sounding board to confirm my understanding of the topic and to check that I hadn't missed or forgotten anything."
- "It's like having a bright, creative, and supportive colleague who is always ready to help, friendly, versatile, and great for bouncing ideas off all day long"
- "I usually process by talking through issues but worry about annoying others with endless questions. AI has been incredibly helpful in this regard. It has allowed me to research, summarize key statistics, and test ideas back and forth, helping me refine my approach. I love how it has simplified the process and removed self-imposed barriers! I love to tell a story, this helps me be much more succinct"

Diversability perspective

- "... as someone with low vision, I read slower than others. Being able to summarise means I'm a lot more productive... Copilot is also good for my mental health; I haven't felt lost on how to progress work or so time pressured when faced with work I must complete at pace. Copilot is also fun to use which helps especially when working on less engaging tasks. As an aside, as a person with a disability, I believe AI will be bigger than braille in terms of linking disabled people with content, enhancing creativity thereby shifting the social bias. Super glad that IR is engaging with AI."
- "I have ADHD and need to exert a significant amount of effort to manage my workload and time every single day. Copilot has helped to significantly reduce the amount of time I need to dedicate to manually figuring out how my days will best work... Avoiding and managing burnout is a daily battle for me. What used to take me half an hour or so to map out, is now a 5 minute job at most..."
- "helps you to get out of a rut sometimes - when you get stuck"
- "...Due to how I word my questions, my coworkers sometimes find it hard to understand what I am actually asking them. Copilot has no issue with my prompts."

Time savings calculation

No time saved: 4 people = 0 minutes

Up to 30 minutes (midpoint of 15 minutes): 20 people = 300 minutes

30 minutes to 1 hour (midpoint of 45 minutes) 12 people = 540 minutes

1 to 2 hours (midpoint of 90 minutes) 10 people = 900 minutes

More than 2 hours (assuming min of 120 minutes) 14 people = 1680 minutes

Total minutes saved per week: 3,420 / 60 (participants) = 57 minutes per week per participant

3,420 / 60 (minutes) = 57 hours per week for all participants

IR's GenAI landscape

Administrative and writing tasks

- Copilot supports our people in the creation and drafting of text and image-based content
- Genesys Agent Assist supports our people writing their wrap-up notes.

Operational Planning

- Verint cloud will deliver key functions in the Integrated Operational Planning solution.

Corporate Services

- Integrated AI within Ātea to support our people completing corporate services tasks.

Customer Facing

- GovGPT will provide customers with an additional way of accessing our external content.
- Coveo provides guidance for customers on our forms and guides.

Item 4

Inland Revenue
Te Tari Taake

Microsoft 365 Copilot Trial Midway Report

Information Governance Team

December 2024

Author: Lucy Cording

Contacts: Lucy Cording and Graham Poppelwell

1. Executive Summary

The team of six (two leaders, four information specialists) have had access to M365 Copilot for eight weeks. Initial findings show productivity, efficiency and confidence gains, however the team have yet to utilise the full product offering of M365 Copilot. We will need to tread water and bed in knowledge for a while before we see the next leap in productivity gains.

The trial has shown the product is not intuitive enough for most team members to just pick up and go and does require specific training. The product also requires a certain base level of digital literacy by the user, for them to be confident in using the tool either at a basic level or to its full potential.

Hallucinations are evident in the outputs, but not overwhelmingly so. The importance of having a human in the loop, cannot be understated.

2. Use case details

The AI Oversight Group approved a *meeting administration* use case, with the following areas to be evaluated:

Tasks related to meetings, including scheduling, booking, capturing meeting minutes and summarising meetings.
Ability to manage a high volume of information across multiple channels (emails, chats, files) to organise and participate effectively in meetings
Routine tasks of finding, collating, and summarising information

3. Training, adoption and usage

3.1 Training

Prior to this trial, the team had approximately 12 weeks' experience using M365 Copilot for Bing. The training received prior to using M365 Copilot for Bing was highly beneficial for preparation for using M365 Copilot. A period of twelve weeks gave team members time to integrate using the tool into their day-to-day work (such as: get used to bringing it up each morning), and also familiarise themselves with writing effective prompts.

Upon first receiving access to M365 Copilot, the team did not receive any formal training. Being the first team having access at IR, no formal training packages existed. In absence of this, short form videos (either YouTube, Microsoft website or from a trainer) have been found to be most beneficial.

Short overview videos from Microsoft (such as Meet Copilot and What you can do with Copilot) were really good introductory clips.

The Microsoft website also has sections on Word, PowerPoint, Outlook, Excel and OneNote which have easy to follow shortform videos and screenshots. These have been very helpful to us.

Sessions with trainers were beneficial, however with such a large volume of information over a 60-minute period, these felt overwhelming to most team members. Sessions typically covered more than one app, however limiting sessions to one app at a time may aid in retention of content. We also at times had more than one session with a trainer in a week, which didn't leave time to embed the knowledge before you're into the next session. The content however in these trainings was well structured, and the trainers were current with latest rollout features, and have provided us with some helpful stretch targets for capability uplift.

3.2 Adoption

We are a team with a mixture of adoption speed. Some team members were utilising quite a few features within the first 48 hours, whereas others waited to see how their colleagues found it, before feeling more comfortable to more actively trial it after two weeks.

For example, some team members shifted earlier from summarising the minutes transcript in M365 Copilot, to prompting the recording, whereas others waited to see how this worked for their colleagues before trying.

3.3 Usage

Use case: Using M365 Copilot to streamline meeting-related tasks including scheduling, booking, capturing meeting minutes and summarising meetings.

The team initially began using M365 Copilot for Teams, Word, PowerPoint and Outlook, and are now exploring use in other apps such as Excel, Whiteboard, Forms, OneNote and Loop.

We note that some weeks are meeting heavy, and some weeks allow more time for creativity and content creation. This can lead to some weeks or months where M365 Copilot for certain tools is either rarely or very heavily used.

4. Key features and benefits

4.1 Most used features

- **Summarising Emails and Teams Chats:** Frequently used to catch up on communications and identify important follow-ups, including summarising everything sent by a specific person during a specific time frame in order to prioritise or prepare for a meeting.
- **Writing Meeting Minutes**: Extensively utilised to draft meeting minutes, significantly reducing the time required for this task.

- **Creating and Editing PowerPoint Presentations:** Helpful in generating and editing slides, especially for summarising presentations and making content edits.
- **Drafting with Copilot:** Valuable tool for crafting concise and informative emails and messages, saving time and improving communication quality.
- **Locating Documents:** The documents we work on can be on SharePoint, email, Teams, and many other places, often with similar names. This feature greatly aids in searching, saving time and effort by providing fast and efficient document retrieval.
- **In-meeting support:** Catching up when late, or summarising what someone has said.

4.2 Specific benefits observed

- **Quality of Writing:** Overall improvement in the quality of written content.
- **Engagement in meetings:** the ability to ask M365 Copilot to catch you up on missed content enables greater engagement for attendees. Preparation with M365 Copilot is also beneficial.
- **Improved Creativity:** Enhances creativity by prompting us for more information, digging deeper and giving suggestions.
- **Improved Confidence:** Users feel more confident as they understand the material better and the reworded versions are more suitable than previous drafts. The ability to ask M365 Copilot questions during a meeting is also beneficial when we are dealing with people with bespoke technical language.
- **Allowing Analytical Thinking:** Provides more time for analytical thinking.
- **Time Savings:** Tasks such as writing meeting minutes or creating PowerPoint presentations now take a fraction of the time. The use of Biz Chat (chat history) allows users to select the chat and continue from the last conversation, saving time without the need to restart prompting.
- **Improved Efficiency:** Integrating M365 Copilot into daily tasks has streamlined workflows, making processes like summarising emails and drafting documents more efficient.
- **Better Organisation:** Features like summarising emails and chats help users stay organised and up to date with important information without having to scroll through lengthy communications.

4.3 Benefits for users where English is not their first language

- **Team member one:** "M365 Copilot is an absolute game-changer for me. Early in my minute-writing journey, I discovered that reading the written transcript helps retain information better than repeatedly listening to the recording. I guess it's because I learned English through reading and writing. Before copilot, writing the minutes involved repeatedly listening to the recordings and referring to the transcript to track the flow of discussions or check unfamiliar phrases. With copilot, comprehension is now managed

by the tool. My task now involves interrogating copilot for the minutes and checking for any inaccuracies."

- **Team member two**: "M365 Copilot and Bing Copilot serve the same purpose in this aspect. I usually ask M365 Copilot and Bing Copilot to rephrase for clarity, professionalism, politeness - anything along these lines. These are available within my Natural Intelligence (NI), but it takes longer to extract them. Different cultures have unique ways of expressing themselves, which can sometimes lead to misunderstandings. Having M365 Copilot and Bing Copilot can quickly create a neutral tone to avoid being misunderstood. In terms of being multilingual, M365 Copilot is particularly beneficial as it helps with communication by offering suggestions and corrections within the context of my work - which I way prefer!!! Bing Copilot may not provide the same level of contextual assistance."

5. Productivity and efficiency

5.1 Non-measurable outcomes

While accurately measuring engagement in meetings can be challenging, M365 Copilot offers a solution for those moments when you arrive late or get distracted. You can ask M365 Copilot to summarise what you've missed, allowing you to quickly rejoin the conversation. This feature is widely used by the team, helping us to seamlessly catch up and participate without waiting for the next discussion point or interrupting others to repeat information. The ability to summarise is also helpful if someone speaks particularly fast or has an accent the listener is not very familiar with. Instead of disengaging, you can quickly catch up on content with a question or two to M365 Copilot.

Quality of emails/comms is also unable to be measured accurately, however team members have indicated their higher levels of confidence in the content of their written work.

Team members are also feeling more confident in knowing what tasks they have due dates coming up for and keeping on top of work on hand.

5.2 Measurable outcomes

The team have documented specific tasks with their historical time requirement and time using M365 copilot.

Some of these are one-off tasks, and some are more frequent. Refer to 3. Key Features and benefits for more information.

Due to the higher-than-expected volume of training required, the team are yet to see material benefits in getting documents through to meeting Chairs in a quicker time.

5.2.1 Minute taking

Minutes are our most time-consuming task and are written by the Level 1 and 2 Information Specialists. These are peer reviewed by a Level 2 Information Specialist or the Tech Lead.

A baseline exercise showed minutes took on average 257 minutes to draft, and 34 minutes to peer review.

We expect that in time, the time required to draft and peer review minutes will reduce, as we get up-to-speed with prompt refinement, and identifying hallucinations.

Task	Historical Time	Time with M365 Copilot	Comments
Write minutes – team member A	360 min	120 min	Instead of using transcript, I'm now loading the meeting recording and prompt the recording for details in each item.
Write minutes – team member B	100 min	120 min	This meeting was an anomaly and was more technical/involved more conversational pieces outside the governance paper than usual. Extra time required to try different prompts.
Write minutes – team member C	120 min	60 min	Still need to listen to parts of recording as transcript does not always catch acronyms. For meetings I did not attend, I sometimes need to listen to parts of recording for more context for clarity.
Write minutes – team member D	150 min	210 min	I am using transcripts as my training ground for prompt engineering compare and contrast learning. Probably spent 80 minutes drafting minutes, and remainder was trialling.
Peer review – team member C	45 min	15 min	Confident in the output so focus more on hallucinations, grammar, contextual storytelling and if key actions, tasks or decisions have been documented.
Peer review – team member D	20 – 60 min	30 – 60 min	Specificity missing, so have to go to transcript/recording to get minutes accuracy up. As we learn & our prompting improves this extra time should drop down again. Added difficulty when multiple voices in one meeting room.
Peer review – team member E	30 - 40 min	10 – 20 min	Fewer grammatical and spelling errors to change. Minutes appear more concise, removing need for me to condense these. Some errors where the recording hasn't picked up the right word or term.

5.2.2 Miscellaneous tasks

Task	Historical Time	Time with M365 Copilot	Comments
Summarise emails and chats from a period of absence	10-15 min	3 min	From one team member: The odd thing though is it summarises emails that I never bother reading so now I'll be doing some unsubscribing or auto-filing.
Catching up on missed meetings	60 mins	5-10 mins	Quick updates with brief questions. No need to rewatch an hour recording.

Task	Historical Time	Time with M365 Copilot	Comments
			From one team member: Previously needed to rewatch video. Just asked 4 brief questions and felt up to date. This is happening every few days and is really helpful.
Locating emails or documents	4 mins	1 min	Highly used feature. "Genius!"
Summarise articles	Variable	5 min	Variable gains in this. Used to save time reading lengthy documents. Also used to summarise articles for current awareness bulletins by the library.
Collating names of meeting attendees	5 mins	1 min	Depending on when attendees join, you can get an incomplete attendee list or need to document this manually. M365 Copilot gives this accurately and immediately.
Summarising meeting debate into a table	100 mins	20 mins	
Creation of a PowerPoint to summarise an external PDF report on AI and ethics	60 min	5 min	
Edit of a PowerPoint for suggested changes in content and grammar	60 min	5 min	Suggested edits were so easy to understand and incorporate. Tidied up the content and less time re-reading looking for less obvious errors. This is a high value support for me.
Using M365 Copilot coach when drafting emails	Variable	Variable	Variable gains in this. For an external audience: Makes a significant difference to check accuracy and meaning for non-tax people.
Creating IR org chart tailored to another agency's request and translate into Spanish	120 mins	30 mins	Short turn around for the right product for an unknown audience. M365 Copilot enabled me to work at a pace and also edit to check content.
Locating Policies and Standards in Governance Documents Centre	3 min	1 min	This has become the preferred option to search for documents. The Haukāinga search is better if users know the exact title of the document in the GDC.
Generating ideas for planning session	-	-	Tried this using Whiteboard app, but irrelevant suggestions and to-do list kept coming up.
Adding meeting file to write minutes	-	-	Same results as getting directly from Teams.
Created MS Forms link for Christmas themed quiz	-	-	Easy to use – questions were drafted for me, and easy for me to finesse to what we needed

5.3 Limitations of measurement

The team have recently shifted to using Bing Copilot and then to M365 Copilot. Although we are cognizant of this when documenting measurements, it is possible measurements in this report have been affected by time savings made possible by Bing Copilot.

The nature of our work in the past eight weeks has meant some apps have been used more than others. This is reflected in our measurement in section 4 and is a limitation of this use case for M365 Copilot.

Historical baseline measurement figures include a secondee who has recently finished her secondment into the team, and did skew both drafting and peer review time to a higher figure.

The initial use case detailed scheduling and booking meetings as one aspect to measure. As a result of this trial, more visibility has been gained of the existing process, and opportunities to reduce the time required in this task have been identified. Some of these opportunities are outside of M365 Copilot's features. Although benefits can be seen in asking Copilot to check the availability of a certain room on certain dates/times, due to a range of other changes, measurements of Copilot's effectiveness in scheduling and booking meetings will be unable to be provided for this use case trial midway report. On another note, the recent release of changes to MS Teams calendar, has new features which we will continue to explore in the second half of this trial.

6. Challenges

- Some challenges have been documented in Section 6 User feedback and testimonials. These can be discussed further in the January midway Leader feedback session with Graham and Lucy and the Project Team.

- More training time and experimentation time has been necessary than expected by Lucy or Graham.

- Following the do's and don'ts of Bing Copilot, there was a hesitance from some team members to trust that their information wasn't being made available to others with M365 Copilot to access. This was particularly worrying for those with union documentation, commercially sensitive documents and ex-CCS staff with taxpayer related emails.

- A behaviour of looking for a silver bullet, one and done prompt, when this can sometimes be as rare as hen's teeth to find. This has taken time for some to realise, with significant time invested looking for a perfect prompt, and a feeling of failure when further finessing of prompts is necessary.

- A tendency to continually finesse an entire document, and not taking out portions which are good, and then finessing the bits that need a bit more detail or have hallucinations. This leads to an unwieldly document and as you finesse one section, it then changes other sections. This is occurring less as we get more savvy with the tools.

- Meetings where several attendees are using one physical meeting room creates a challenge. The transcript does not identify the individuals, the 'speaker' is the room. We are aware of the not-yet-available MS Teams feature *Voice and Face Enrollment,* which would be beneficial in this instance. It is noted, that in some meetings, staff are individually logging into the room device and logging into their laptop which aids in voting and putting hands up.

- Copilot doesn't run over any content/words which are snipped (screen captured) from another location.

- Inability to ask Word or PowerPoint to enact changes. "Unfortunately, I can't make changes to the document itself, but I can answer general-purpose questions or those about the document."

- Restricted access to some features inhibits ability to get the full benefits of M365 Copilot. For example, one training session focused on creation of a loop, which IR staff do not have the functionality to do. IR also does not have the most up-to-date version of some tools either which means we can't fully utilise M365 Copilot's abilities.

7. User feedback

7.1 Graham

Work pattern behaviour – I am looking at where and how copilot enables me to operate at a higher level or pace. In the leadership space, where *meetings* feature as my work pattern and less about content creation, copilot has some definite advantages. May not be time saving but enables a better meeting experience. I have found I can quickly prepare for a meeting with 'catch me up content' so I am 'in the meeting' from the start, instead of participation delays from swapping out my thinking from the last meeting. This helps tighten up on the content and be more purposeful as to meeting outcomes. Value proposition, as better prepared, could I move to 20-25min meeting instead of 30mins as an example.

Leaders: we all look to clear emails in meetings where we are not an active participant (passive attendee and being informed). I have found using copilot to support me with completing emails, while in a meeting, means I can balance focus to the meeting while also having the safety net of copilot to ensure the competency of my email drafts and an overall sense check. These are micro activities that add up.

Re-thinking how I operate. Screening emails to assign a priority and tasks and using copilot to feature these in review of activities for the day.

Indeed, exploring how I can use Planner and co-pilot in a way to enable me to be gain time on topics (even if admin orientated it's a win for me).

The Training that I have seen but not yet fully deployed provides more areas to explore. This is in my 'to do' when I see work opportunity,

- Using a file document as a guide to write new content
- With new October release for M365, PowerPoint content creation, will make a difference
- Looking at shared workspaces Pages (in loop) and how to operate these to move at pace on emerging topics.

And swapping between copilot web and work content – understanding how to operate with both with the right content.

7.2 Lucy

I have found Copilot relatively easy to use and intuitive. There is a large amount of training available online which is easy to access in bite sized chunks on specific topics. I have found this has integrated fairly seamlessly into my day-to-day work.

The ability to ask questions into the biz chat of topics which I wouldn't type into the web chat is really handy. This was particularly helpful when I have needed to draft speeches on various topics, which would have breached the guidelines for Bing Copilot chat.

My most used feature of M365 copilot has been the ability to ask copilot questions during a meeting. Whether this is to catch up if I am late, to summarise an item if I had to briefly leave the meeting, to summarise someone's comments if their accent is one I am not familiar with, or to ask copilot if there is anything else they suggest I should ask. This has noticeably increased my engagement in meetings and become a tool I heavily rely on.

I expect this will be a very effective tool for IR staff who are running (clicking) from meeting to meeting to meeting. I expect it will reduce instances of "let's catch up such n such on what they've missed" enabling meetings to stay on task and not lose direction. With time, meetings will likely reduce in duration. I also expect staff will be able to miss more *inform* type meetings and catch up by prompting the transcript.

The ability to search and locate a document/email is super helpful. I compare this to the 1990s, when I used to know 30+ phone numbers off by heart. I then got a cell phone with a built-in contact list and suddenly I no longer needed to memorise numbers. At work, I used to naturally remember all emails I send/receive, when they arrived, who from, subject line etc, so I could quickly locate them. I am already finding I'm no longer spending energy mentally absorbing this information because I know copilot will find it for me.

I note we have received no reports from team members that M365 copilot has retrieved or accessed documents the team members should not have access to.

7.3 Tanya

- Very easy to use. So easy that it is important to remain mindful of if you are in Web or Work with the browser option now. Pleasant to interact with.
- I think it's a case of much more functionality in M365 Copilot than we can anticipate. We need to be using for the use case of the pilot, so mostly staying with using in our predominant tool set.
- A great aspect is being able to retrieve documents with greater ease and speed – less of 'where was that saved again?' Yay! Speeds up finding things and having collated lists.
- Daily task improvement is collating all chats and emails and getting organised into structure in one place for prioritising activities. Summary of meetings if missed and needing insights from them.
- Aspects enjoyed most: Prompting for outlines on how to do things – check if any further suggestions for me. Providing me steps on how to utilise Copilot for an activity and then applying those recommendations!
- Main concern is the homogeneity that can occur with meeting content. The LLM is reverting to the mean, and specificity can be erased. This means minutes writer must still be conversant with the meeting's discussion content to ensure accuracy remains/or that you are very proficient in prompting to ensure Copilot interrogates the transcript appropriately for accuracy to remain.

7.4 Fiona

- Copilot is my smart co-worker that does the writing, analysing, and searching for me.
- I've also had written conversations with it to identify gaps in my knowledge and skills by combing through all my messages and emails. While the advice isn't always spot on, I do appreciate the capacity that it has allowed me to think contextually and critically about the writing that it's produced for me.
- It was great having a variety of M365 trainers as it catered for the different learning style of Info Gov team members.

7.5 Tessa

- Would agree that it's easy to use. Once you learn how it structures itself and its limitations you can play around quite nicely.
- For documents it's a great tool, but our usage has been limited to minutes predominantly so would be good to see how far it can be utilised.
- Has definitely improved productivity as repetitive tasks can be significantly reduced.
- Easy to keep a track of discussions you've been having with others but only in written form. Calls (phone or video) are not identified.
- Aspect enjoyed most: It's prompts that encourage you to explore more about the topic you've asked about. eg: I wanted information management to start mapping the teams

SharePoint site. It provided a simple breakdown, but also had links to "would you like to know more". So it's the landscape of continually learning if you want it to be. AKA rabbit holes for the unwary.

- I underestimated the amount of time I'd need to understand how it works and get the best out of Copilot.
- It's too easy to become comfortable with Copilot and therefore hallucinations over time can be overlooked.
- It seems to open/access the last document related to your question that you opened. EG: I asked for it to find the last minutes of a WG meeting I was to review. It only brought the month before minutes because I had done something/opened that document. The most recent one wasn't located.
- It's a game changer provided you dedicate sufficient time to play and understand.

7.6 Kate

- M365 Copilot has made it much easier to locate documents that would otherwise take me a while to find.
- Prompting Copilot to summarise emails, Teams chats and documents has saved me a lot of time. I no longer need to scroll endlessly to find the information I need.
- Using M365 Copilot to write meeting minutes has drastically reduced the time I spend on this task. What used to take hours now takes just a fraction of the time, allowing me to focus on more strategic activities.
- Integrating Copilot into my daily tasks has been seamless. Whether it's summarising articles for current awareness bulletins or drafting emails, Copilot has made my workflow much more efficient.
- I love how Copilot helps me craft concise and informative emails. It's like having a personal assistant that ensures my communication is always on point.
- One challenge I've faced is not quite knowing the limitations of M365 Copilot and often getting a reply that says, 'Sorry, I can't...' but that's changing fast too. It's just a matter of keeping up-to-date with the changes.
- Overall, my experience with M365 Copilot has been overwhelmingly positive. It's a powerful tool that has transformed the way I work, making me more efficient and effective in my role.

Item 5

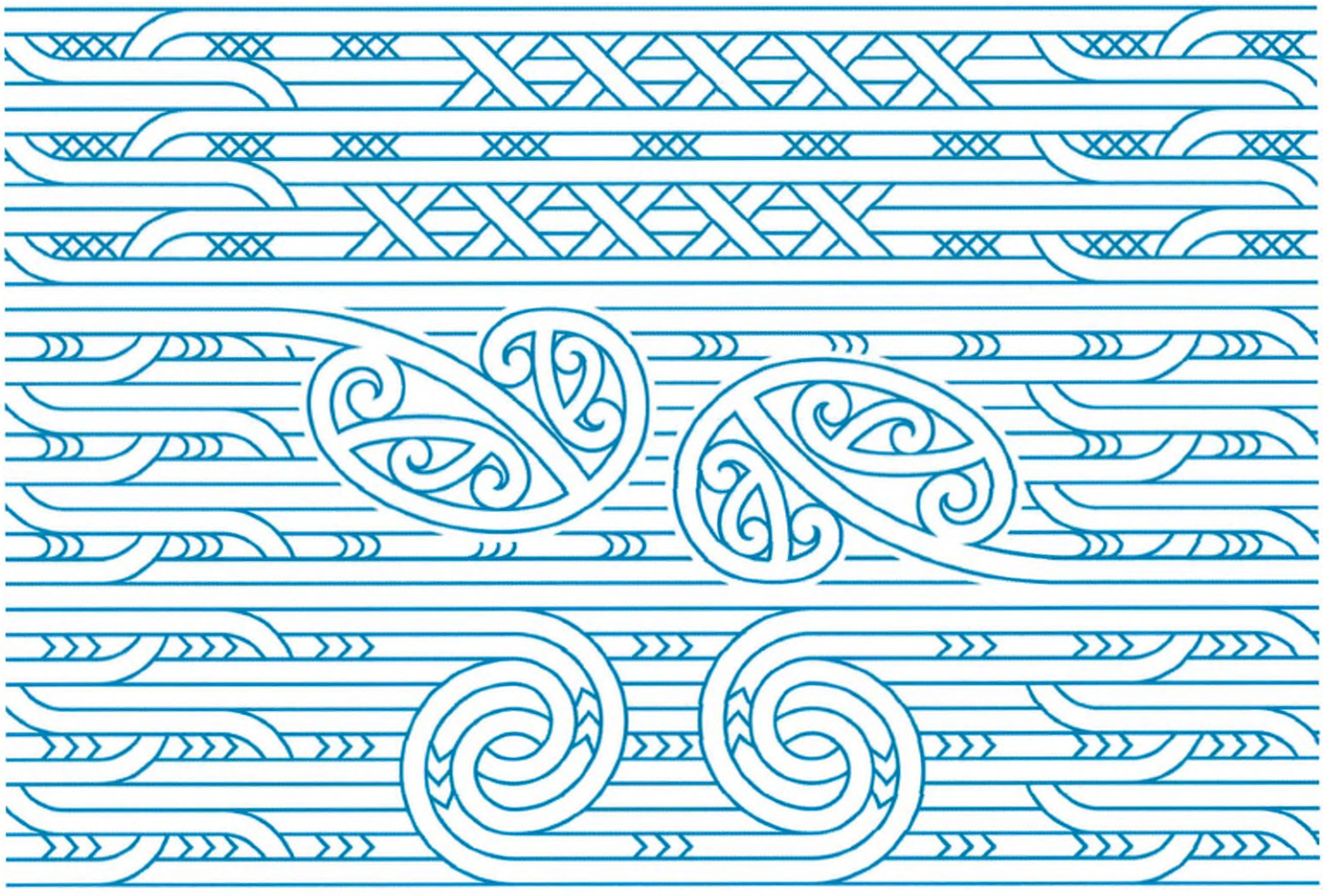
Inland Revenue
Te Tari Taake

Microsoft 365 Copilot Trial Midway Report #2

Information Governance Team

March 2025

- Author: Lucy Cording
- Contacts: Lucy Cording and Graham Poppelwell



1. Executive Summary

The findings from the December 2024 report continue to be relevant in March 2025. Now that we have the training and are pretty much fully up-to-speed, I have no doubt the cost of the product is now outweighed by the benefits moving forward.

The team of six (two leaders, four information specialists) have now had access to M365 Copilot for 18 weeks. An additional colleague in a different segment (Thomas, Business Support, Executive Support) who completes similar governance meeting related tasks, has also gained access (six weeks ago) and his results are evidenced in this report.

We are seeing most time savings using BizChat and MS Teams, with smaller time savings in PowerPoint, Outlook and Word. We feel we are getting value for money with the cost of the tool, specifically higher quality advice being provided to senior stakeholders in time sensitive situations, time being freed up to take on additional priority work and higher engagement during meetings with the ability to prompt during virtual meetings.

I note some members of the team have only recently discovered the level of detail provided by AI notes (where you click into a note and it drops down into further detail). As this inevitably gets used more, I expect the time to minute meetings will reduce further.

2. Updated findings

2.1 Minute taking

We continue to try different methods for writing minutes, with each team member taking a different approach – often a combination of the MS Teams AI Notes, prompting into MS Teams, placing the transcript into BizChat to summarize, referencing the pre-reading governance paper in BizChat or reading the transcript to write the minutes.

We note that the AI Notes will typically remove most detail and emotion and simply state the outcome of that section of the conversation.

For the level of detail necessary in governance meeting minutes, we are seeing some reduction in time, but not tremendous. In the event of business-as-usual meetings, the level of detail created by the AI notes is perfect.

2.2 BizChat

This is where the real gold is for us and efficiency gains. Asking questions about documents, locating documents, IR-specific research, task scan through previous meeting minutes, editing content which wouldn't be appropriate for Bing Chat, looking across email chains/teams messages or preparing you for your day.

In our roles, I have no doubt the cost of the tool is outweighed by the productivity gains through BizChat.

Specifically for our team, with the heavy focus on AI from ELT, the Minister of Revenue and other agencies, we have received requests for briefings with minimal warning. This has included briefings to be written to the MoR, within hours. With 18 months' worth of documentation now existing in AI Governance (cross agency presentations, briefings to ELT members for international conferences, briefings to Minister of Revenue, OCED Tax Administration information, Proof of Concept papers, Risk Assessments etc), there is a large amount of content to synthesize in a brief period of time for updates. Such prompt provision of advice and information to the Commissioner, ELT and the Minister would simply not have been possible without M365 Copilot. To meet the required time frames without Copilot, quality would have needed to be lower, or time frames would have needed to be extended.

2.3 Thomas' findings

Thomas has been trialling M365 Copilot for governance meeting administration for Technical Design Authority and Design Authority, and in his role in Executive Support. He did not see noticeable time savings in minute taking. He has however seen time savings through use of Copilot in MS Word, Excel and BizChat estimating an hour a day being freed up to put into other priority work. He has just this week discovered the abilities of AI notes, so will trial this over the coming weeks, which he expects will reduce time to minute meetings.

2.4 Graham's findings

Findings which will create efficiencies in CCS

Understanding the M365 Copilot 'meeting experience' and GenAI has led to consideration of how non-licensed users can access GenAI content, especially during external meetings. M365 Copilot's GenAI features require a valid Copilot license. However, meeting organizers or participants who have Copilot can share the generated insights, summaries, and action items with others, allowing everyone to benefit from the information even if they lack direct access to Copilot.

Having tested M365 Copilot, a paper has been positioned for IR attendees in a Teams meeting, utilizing 365 Copilot, to explore opportunities for accessing GenAI content instead of relying solely on handwritten notes. This knowledge in some customer groups could improve productivity where the other party has Copilot.

By prior agreement, IR could also access the same meeting outputs. This consideration is to be presented in a paper to CCS Leaders as an opportunity, especially with Significant Enterprises

compliance managers and their customer engagement. It will require further AI fluency training as the investment.

Patterns of work

The use of prompts in business chat saves me significant time as a leader. My team's work resides across numerous SharePoint workspaces, Teams chats, and emails. Accessing content through prompts rather than exact file names helps me prepare for meetings with the right content available.

I use Copilot to quickly summarize tasks from actions, making it easier to keep current work at hand without searching through emails or chats. Having saved prompts and coding actions from chats and emails means I can come back to issues and this helps me manage my productivity.

Sharing this experience with my team has been invaluable for shared learning and productivity gains. In my governance role, early access to M365 Copilot has helped me stay updated on AI developments as they are released, keeping governance aligned with AI advancements.

3. Considerations

At the same time this pilot has occurred, the nature of the work our team is completing has shifted dramatically. A year ago, we spent approximately 20% of our team's capacity on AI related tasks. This has now increased to 60%. The nature of this work requires moving at pace, with more frequent briefings required to specific senior stakeholders, more Proofs of Concept coming through the team often at short notice and supporting governance documentation to be created. M365 Copilot has enabled us to deliver this at a much faster pace than I think many of us expected, to deliver this at a high quality and also take on additional work.

Item 6

Inland Revenue
Te Tari Taake

Microsoft 365 Copilot
Project Closure Report

AI Working Group and AI Oversight Group

9(2)(a) Makayla Stewart, Phyllida Crawford, Tim Crook

V 0.1

June 2025

Purpose

The purpose of this paper is to present the findings of the recent pilots of M365 Copilot and MS Teams Premium, and prompt discussion of next steps for IR with these products.

This is to be read in conjunction with the Copilot for M365 - AIOG June 2025 pack.

It is recommended that the AI Working Group:

- **Discuss** the findings, proposed next steps and identify any further considerations

It is recommended that the AI Oversight Group:

- **Note** the findings of the two pilots

Overview

Inland Revenue has introduced Copilot Chat (previously known as Microsoft Copilot, Bing Chat) throughout a large part of the organisation, resulting in noticeable improvements in productivity and the quality of work (Benefits of Copilot for IR, 2024). Nonetheless, this version has limitations—it cannot retrieve internal information and necessitates manual content transfer between different applications.

A separate pilot looked to explore how we can use our internal data, Microsoft 365 Copilot (M365 Copilot) is seamlessly integrated into the Microsoft 365 suite, including Teams, Outlook, and Word, which allows Copilot to leverage our internal data. Inland Revenue conducted a pilot of M365 Copilot, concentrating on the administration of meetings. The pilot aimed to:

- Determine if the substantial advantages noted by other agencies, such as ACC and the ATO, could be realised and replicated within our own environment.
- Identify and address IR-specific risks, including privacy, security, and compliance with the tax administration act.
- Develop internal expertise and inform strategies for change and adoption related to the potential implementation of M365 Copilot.

While there are evident advantages for Inland Revenue in adopting generative AI, there are also challenges and concerns regarding its implementation that require ongoing attention.

Our pilot

A structured small-scale pilot of 20 individuals from the project team and Enterprise Information & Knowledge's Information Governance team were allocated Microsoft 365 Copilot licenses. The Information Governance team is responsible for administering 9 governance groups including the AI Working and Oversight Groups.

Our pilot has a targeted use case of the administration of meetings, our people spend a significant amount of time completing administrative tasks related to meetings, including scheduling, booking, capturing minutes, summarising meetings, and locating documents relevant to the meeting being organised. These tasks are typically completed across multiple M365 applications.

The pilot consisted of three phases:

- Phase 1: Completed initial setup and configuration of Microsoft 365 Copilot in our production environment.
- Phase 2: Confirmed that our existing data and security settings where respected.
- Phase 3: Onboard the participants for the meeting administration use case, report back on the findings and evaluate available learning materials.

Use Case Findings

Inland Revenue
Te Tari Taake

Use case findings background

Our use case for M365 Copilot commenced in July 2024, at a time when the tool offered limited functionality and minimal integration across the Microsoft 365 suite. These early limitations affected both the speed of user adoption and the depth of engagement with the tool.

Since then, substantial enhancements to the AI model and its integration have significantly improved the quality, reliability, and utility of Copilot's outputs.

The Information Governance team has documented their experience through two key reports:

- **M365 Copilot midway report December 2024**: Captures early insights from their initial eight weeks of use.
- **M365 Copilot second midway report March 2025:** Reflects on progress after 18 weeks of use.

In addition, a broader feedback survey was conducted in June 2025 across all current license holders, including members of the Copilot Enablement Project Team. These three sources form the foundation of our findings and provide a comprehensive view of Copilot's impact across different roles and stages of adoption.

Productivity Benefits

89% of people used M365 **Copilot Daily**

5 hrs saved each week per person on average

44% saved **more than 5 hrs per week**

94% found time savings from creating content

83% found time savings **locating documents**

100% said that Copilot made them **more productive**

89% reported **better quality of work**

Copilot was most widely used in **Microsoft Teams**

Data source: End of pilot survey. 18 pilot participants responded to the survey.

Additional Benefits

Conclusion: The pilot has demonstrated significant benefits for IR. We expect time savings, and other benefits will grow as our people become more proficient with Copilot and integrate it further into their daily workflows.

Creativity: Enhanced creativity by prompting people to explore and dig deeper into a topic, even offering suggestions of relevant questions to explore.

Managing priorities: Supported people to stay on top of their workload by summarising tasks, surfacing priorities and enabling quick catch-ups. Tasks can be completed in shorter timeframes and urgent tasks can be managed more calmly.

Quality of interactions: People were able to stay engaged and present during meetings knowing that they were going to be summarised by Copilot, enabling them to focus on the conversation rather than note taking.

Reduction in stress: Helps reduce stress by easing the pressure of urgent deadlines and by offloading routine tasks such as locating documents, people felt more in control of their workloads.

Taking on additional work: With the time saved through using Copilot, the Information Governance team was able to take on more work while continuing to deliver high-quality outputs under pressure.

> It's prompts that **encourage you to explore** more about the topic you've asked about... So it's the **landscape of continually learning** if you want it to be.

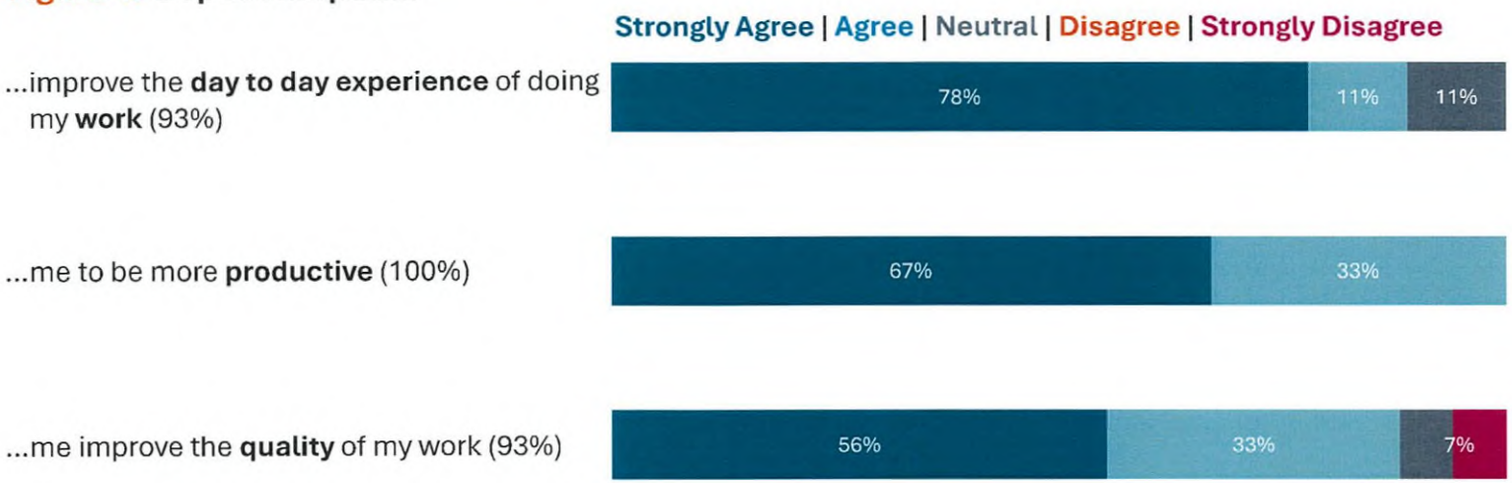
Information Specialist, E&IS

> My **flexitime** balance has been **more manageable**. My **stress levels have reduced**. - The time saved through copilot, has enabled me to both **take on additional work**, and also have a more reasonable flexi balance.

Technical Lead, E&IS

Very high proportion of participants in the pilot reported significant benefits

Figure 1: Copilot helped...



Inland Revenue Department. (2025). Microsoft 365 Copilot: 15 responses. Internal document.

Key use cases for the administration of meetings

Drafting meeting minutes: Drafting minutes is a core responsibility for the Information Governance team, and the introduction of Microsoft 365 Copilot has led to notable reductions in the time required to complete this task. The team anticipates further efficiency gains as their proficiency with prompting improves.

Meeting summary notes: Copilot automatically generates summary notes for any recorded or transcribed meeting, providing a reliable alternative for those unable to attend.

In-meeting support: During live meetings, Copilot enables users to catch up on missed content by summarising earlier discussions in real time. It also allows participants to clarify or revisit what was said.

Scheduling: Can assist in checking room availability and help find suitable times for participants.

	Pre-M365 Copilot Average	With M365 Copilot Average
Writing meeting minutes	257 minutes	127 minutes
Peer reviewing meeting minutes	34 minutes	25 minutes
Catching up on missed meeting	60 minutes	5 to 10 minutes

> As a result of Copilot, I have **taken on more work** and my team has taken on more work.
>
> ***Technical Lead, E&IS***

Key use cases apply across a wide variety of roles

Creation of content: Accelerates the development of documents, presentations, and communications by generating high-quality first drafts using internal information from multiple sources. This helps produce IR specific content that is more relevant and accurate than Copilot Chat.

Summarising information: Distils key insights from emails, chats, meetings, and documents—reducing time spent reviewing content and enabling sharper focus on priorities.

Locating documents: Enables fast retrieval of documents across SharePoint, Teams, and email—without needing exact file names—saving time and supporting responsiveness in high-paced environments.

Troubleshooting: Supports guided resolution of technical issues in the M365 suite - through AI-guided assistance, reducing delays and improving self-sufficiency.

Feedback and Coaching: Provides real-time guidance to refine written content—enhancing clarity, tone, and confidence in communication.

> A great aspect is being able to retrieve documents with **greater ease and speed** – less of 'where was that saved again?'.
>
> ***Information Specialist, E&IS***

> I'm able to be **more present in meetings** without having to worry about taking notes and **being confident** that the important points are being captured.
>
> ***Information Specialist , E&IS***

Copilot has been useful for our EASL and neurodiverse people

EASL perspective, for our people who speak English as a second language:

- Enabled them to have better comprehension of the content they were working with and provides a second layer of review to catch potential inaccuracies.
- Supports clearer, more professional and courteous communication by rephrasing written content while adapting to each person's preferred way of working.

Neurodiversity perspective, for our people who:

- Spend time translating their thoughts into the required formats can get tailored responses from Copilot.
- Find it challenging to manage their time and workload, Copilot helps surface priority emails, Teams messages and documents, making it easier to focus on what matters most.
- For those who frequently switch between tasks, Copilot offers concise summaries while flagging upcoming deadlines to reduce cognitive load.

> M365 Copilot is an **absolute game-changer** for me... With copilot, comprehension is now managed M365 Copilot.

Information Specialist, E&IS

> I often spend considerable effort and time translating my thoughts... Copilot provided me with the ability to ask questions and **receive tailored responses** in a few seconds.

Business Support, E&IS

Conclusion

- The Information Governance team found for them that the **cost of the tool was outweighed by the productivity gains** achieved through M365 Copilot.
- M365 Copilot has been shown to have **significant productivity gains**, time savings and additional benefits including improvements to work experience and reducing stress.
- There is a need for people to shift and adjust how they work in order to embed the use of Copilot in how they work, this can take time but has significant benefits.
- **Benefits** were found to have **increased over time** as users embedded M365 Copilot into their workflows, they become more confident and comfortable prompting, and as improvements were made to the tool.
- Compared to Copilot Chat, people found **significant efficiency gains** from being able to find, summarise, and synthesise **large amounts of internal IR information** more quickly.
- In June 2025, Microsoft released **new versions of the AI model** and **improved features** in the M365 apps. This resulted in a **significant rise in capabilities** and **quality** of the output from Copilot.

Usage patterns and features

Inland Revenue
Te Tari Taake

Most-used entry points for Microsoft 365 Copilot

- We looked at where people are most commonly using Microsoft 365 Copilot
- The data shows that Microsoft Teams, M365 Copilot Chat, and Outlook are the most frequently used entry points, with significantly lower usage across other Microsoft 365 apps
  - **Microsoft Teams:** 91% active users rate*, 72% said they use it daily
  - **M365 Copilot Chat:** 86% active users rate, 61% said they use it daily
  - **Outlook:** 59% active users rate, 39% said they use it daily

Other Copilot entry points had lower usage:

- Active users rate: Word (32%), Excel (23%), PowerPoint (14%)

* Active users means they used the product at least once in the past 30 days

Most used features across M365 apps

Overview: The features found to be the most beneficial were those in M365 Copilot Chat, Teams and Word:

- **Content creation in M365 Copilot Chat:** Generates content from across multiple M365 apps from a single interface, it can draft emails, reports, summaries and presentations without switching between tools. It offers a centralised place that is integrated with the other applications which supports the faster creation of content.
- **Document search in M365 Copilot Chat:** Searches across the Microsoft Graph, instead of searching through each M365 app individually people can search in just Copilot Chat without having to know the specific document name. Reduces the need remembering the name of each document and manually searching for it.
- **Meeting notes in Teams:** Meeting notes are automatically generated in real-time when recording or transcribing, following the meeting these notes are published within the meeting and include key points, decisions and action items. Enables people to be more 'present' at the meeting instead of worry about meeting minutes.
- **Content creation in Word:** The ability to within seconds generate full documents, rewrite sections based on natural language prompts, and refine content using the existing content within the document as a reference.

Key findings on features

- **Microsoft Teams and M365 Copilot Chat** offer the most advanced Copilot features:
  - Interactive, back-and-forth conversations mean people can guide and refine responses iteratively
  - Feels like working with a real collaboration partner
- **Other Microsoft 365 apps** have more limited functionality at this stage. For example:
  - **PowerPoint:** Copilot can generate an initial presentation and edit individual slides, but it can't make structural or style changes to the whole deck e.g. can't shorten a 20-slide deck to 5 slides
  - **Excel:** Copilot can struggle to understand the context and relationships with unstructured data.
- Use cases that focus on leveraging tools with the more advanced functionality are where IR will see the greatest benefit.
- The tools with more limited features are useful for quickly generating content, but less effective for refining it. IR people must apply critical thinking to carefully review the content.
- While Microsoft continues to add features, M365 Copilot Chat may better support a think-plan-write-revise process for high-quality writing.

Implementation Considerations

Inland Revenue
Te Tari Taake

Licence allocation process

- **Options and recommendation:** Please refer to Copilot for M365_AIOG June 2025 for licence allocation and approval options, including our recommendation.

- These options will be presented to EPPC and outcomes will be documented in their meeting minutes.

Onboarding and training

Completion of the **Copilot Fundamentals** course in Ātea will be a **pre-requisite** to getting M365 Copilot access. Once completed, people will automatically get access.

Focus on high-value areas

We will prioritise training on the two areas of M365 Copilot that we found to have the most business value:

- **Microsoft Teams** and **M365 Copilot Chat**
- For other Copilot entry points, we will offer lightweight, self-paced learning by linking to Microsoft resources.

Helping people discover useful features

We found that even highly engaged people missed some useful features during the pilot. To address this, we will provide:

- trainer-led sessions (live or recorded) and downloadable practice activities for self-paced learning
- visual guides to help people find and use key features.

Curated learning resources to support efficient learning

- Some pilot participants lost time going down rabbit holes such as watching tutorials for features that turned out not to be very relevant or useful for their role, or where the features aren't mature or effective yet.
- To avoid these pitfalls, we will provide a curated set of high-quality resources to help people quickly upskill in the most valuable features. This will include:
  - targeted guides and videos
  - practical tips for getting started
  - examples of high-value use cases.
- Wherever possible, we will use high-quality Microsoft resources instead of developing in-house materials.

Cost management

We will use our existing licence management approach

- The WTS team is responsible for making sure IR people have access to the tools they need, in a way that's cost-effective
- M365 Copilot is a licensed application. WTS will manage it using the same approach we apply to other licensed apps.

Usage monitoring and reallocation

- To manage costs and avoid waste, WTS will monitor how often each user engages with M365 Copilot and identify any licences that are underused.
- If a licence is not being used or shows minimal activity, WTS may reallocate it. Specifically:
  - licences with no usage for 60 days will be automatically revoked
  - affected users will be notified by email
  - freed-up licences will be reassigned to people on the waitlist (if applicable).

Teams Premium

Inland Revenue
Te Tari Taake

Teams Premium Evaluation

- Teams Premium is a Microsoft Teams add-on license that enhances the core experience with advanced features such as AI-generated meeting notes, meeting watermarking for enhanced security, and custom-branded meeting environments.
- As part of this initiative, Teams Premium was assessed as a potential alternative to Microsoft 365 Copilot. Two distinct user groups were involved in the evaluation:
  - Group A: Users with extended experience using Microsoft 365 Copilot.
  - Group B: Users with no prior exposure to AI-enhanced Teams functionality.
- It is important to note that all participants also had access to Copilot Chat (formerly Bing Chat Enterprise), which was frequently used for tasks such as summarising documents, rewriting content, and extracting key information.

Benefits over Teams

- The cost of a Teams Premium licences is $10 per user per month compared to 9(2)(b)(ii) per user per month.
- Increased productivity by reducing time to produce notes and minutes from meetings
- Additional security features:
  - Watermarking for video and content which discourages people taking screen shots screenshots.
  - End-to-end encryption for meetings.
  - Control who can record and also who can copy and paste meeting chats.
  - Create custom meeting templates with security setting and sensitivity labelling

Limitations

- Transcript creation is still problematic with words and names being records incorrectly.
- Unlike Microsoft 365 Copilot, Teams Premium doesn't let you interact with meeting summaries. With Copilot, you can ask follow-up questions and get answers based on the meeting transcript. Teams Premium provides an AI-generated summary, but you can't use it to explore the content further.
- There is no in-meeting support available, users have to wait until the end of the meeting for the AI generated notes to be available.

Conclusion

- People found their increased productivity by reducing time to sit through recordings again and review lengthy transcripts.
- Outputs from Teams Premium were detailed and had rich information.
- People were able to concentrate more on being engaged in the meeting instead of worry about minute taking.
- Teams premium is a solid step up from Teams. It saves users time and increases the quality of the meeting notes.
- It allows users to be selective on attending meetings as recaps can be completed in half the time.
- The cost of teams premium is a quarter of Copilot for M365 and when used alongside Copilot Chat, there is significant benefits to our people.
- **Assess whether Teams Premium, coupled with Copilot Chat would met requirements when evaluating a request for a Copilot for M365 licence.**

Appendix

Sub-heading – delete slide if not required

Inland Revenue
Te Tari Taake

References

- IR. 2025 Copilot for M365_AIOG June 2025 .
- Australian Taxation Office. 2024. *M365 Copilot Trial Update*. Canberra: Australian Taxation Office.

Appendix 1: Hours saved calculation

Up to 1 hour: (midpoint of 30 minutes) 1 person = 30 minutes

1 to 2 hours (midpoint of 90 minutes): 4 people = 360 minutes

2 to 5 hours (midpoint of 210 minutes): 5 people = 1050 minutes

5 to 10 hours (midpoint of 450 minutes): 6 people = 2700 minutes

10 hours (600 minutes): 2 people = 1200

Total minutes saved per week: 5340 / 18 (participants) = 296 minutes per participant

5340 / 60 minutes = 4.9 hours per week for all participants

Calculation for tool cost vs salary

Annual IR employee salary: **$93,500** per year (source: Te Kawa Mataaho Public Service Commission), **$44.95** per hour

Cost of Copilot: 9(2)(b)(ii) per month per person

Hours of work paid for: **8 hours** per day, **40 hours** per week

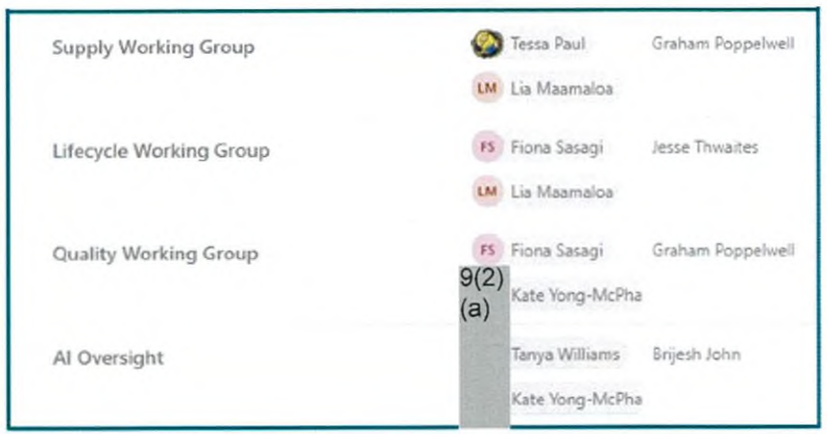
9(2)(b)(ii)

Appendix 6: Calculation for tool cost vs salary

Annual IR employee salary:	**$93,500** per year (source: Te Kawa Mataaho Public Service Commission), **$44.95** per hour
Cost of Copilot:	9(2)(b)(ii) per month per person
Hours of work paid for:	**8 hours** per day, **40 hours** per week

9(2)(b)(ii)

Appendix 2: Meeting Use Case Scope



Note: Initially the use case will focus on the lower-level meetings and groups.

Appendix 3: M365 Copilot Chat

Copilot features

- Context awareness in that it understands your role, organisation and recent work.
- Powerful search across the M365 Graph for documents, people or meetings, able to surface relevant content instantly.
- Cross-application integration, it can pull insights from all the M365 suite to give you an answer that incorporates all relevant information.
- Generated content can then be converted into a Word, Excel or PowerPoint document.

Limitations

- Chat history does not permeate across responses, if an individual has a specific way they want Copilot to respond, it will need to be said each time.
- It does not have access to content within third party applications such as Stax.
- The quality of the output is heavily dependent on the prompt.

Most beneficial to IR for

- Content creators
- Knowledge workers

Appendix 3: Teams - Collaboration

Copilot features

- Real-time in-meeting support, users can get a summary of what has been covered and ask questions to deepen their understanding.
- After meetings, notes are automatically generated and available within an hour after a meeting finishing that includes action points and who they are assigned to.
- In Teams Chats, users can ask questions, request summaries and gain insights into individual and group chats.

Limitations

- If speakers are in a room, action points and notes cannot be attributed to an individual and instead are identified as being in the room.
- Action items are not linked to any task systems.
- The transcript cannot be edited in real-time.

Most beneficial to IR for

- Meeting heavy individuals
- Executive Leadership
- Content creators
- Knowledge workers

Appendix 3: Word – writing and editing

Copilot features

- Generate drafts using a short prompt or by extracting content from multiple files.
- Revise text to adjust tone, improve clarity, or change the length based on the intended audience or format.
- Pull relevant information from other documents or emails and insert it into the current file.
- Create tables, bullet-point lists, or summaries based on the content of the document
- Identify areas where content may be unclear or inconsistent and receive suggested improvements.

Limitations

- Manual formatting is required for complex layouts.
- Best performance for document Q&A is under 7,500 words.
- Rewriting works best on documents under 3,000 words.
- Summarisation and referencing are limited to 300 pages or 1.5 million words.

Most benefit to IR for

- Content creators
- Knowledge workers

Appendix 3: Outlook – Communication

Copilot features

- Scheduling with Copilot straight from an email, invitees will be automatically selected, a suggested time and agenda will be generated.
- Within emails, Copilot can be used to draft new content, and review or coach you on content you have already written.
- Copilot can be asked to provide a summary of the emails you've received and support you in prioritising by raising up important or time sensitive content.

Limitations

- Struggles to schedule appointments with people who have no free spaces in their calendar.

Most beneficial to IR for

- Business Support
- Executive Leadership

Appendix 3: PowerPoint – for visual storytelling

Copilot features

- Can create outlines of presentations from natural language prompts or existing documents.
- Summarise presentations and highlight key messages.
- Receive basic design suggestions, including visuals and formatting.
- Generate slides with content, layouts, and suggested talking points.

Limitations

- Content is often low quality and requires significant manual edits and adjustments.
- Results tend to be more 'marketing' type outputs than what IR typically use PowerPoint for.
- If there is a need for specific formatting or layouts they will need manual adjustments.
- Visual suggestions are often generic.

Most beneficial for IR

- No clear use case at this stage

Appendix 3: Excel – for analysis

Copilot features

- Ability to ask natural language queries about the data.
- Support in writing basic formulas with explanations of what they are and how they work.
- Generate visualisations and pivot tables.
- Ability to apply colour and formatting to specific cells.
- Can pull in information from other M365 applications such as Outlook to create content.
- Recently introduced Python.

Limitations

- There is limitations to its ability to understand the data's context or relationships.
- Struggles with unstructured data.

Most beneficial to IR for

- Those needing support with basic formulas

Appendix 3: Other M365 applications

- **Notebook**
- **Loop**: Collaborative content creation using Copilot, content created in Copilot Chat can be added directly to loop (pages functionality) and shared with others.
- **Stream**: Ability to ask questions and get answers based on the transcript, responses will include timestamps.
- **Whiteboard**: Suggestions for layouts or structures, ability to theme content.

Limitations

- There applications are not widely used within IR, the benefits provided will be reliant on people using these apps.

Most beneficial to IR for

- Customer Experience Designers
- Project Managers
- Knowledge Worker

Inland Revenue
Te Tari Taake

Thank you

Item 7

Inland Revenue
Te Tari Taake

Microsoft 365 Copilot

Copilot Service Enablement

[redacted]

Date: 18 June 2025

Sponsor: Malcolm Breadmore & Jesse Thwaites

Presenters: Tim Crook, Makayla Stewart

Purpose and Recommendations

- This paper summarises the Copilot for Microsoft 365 - Project Closeout Report and recommends our next steps forward to leverage the benefits found.

It is recommended that the Working & Oversight Group:

- **Endorses** our request to move Microsoft 365 Copilot to BAU, as a controlled service with a phased deployment.
- **Discuss** the draft Risk Assessment for Microsoft 365 Copilot and known risks and controls (appendix 4).
- **Endorse** proceeding to PPG and EPPC for key discussions and decisions that are needed to execute the above. These include (details on slide 3):
  - The number of M365 Copilot licenses that should be purchased.
  - At what level does a request for a license get approved.
  - Budget uplift to meet licencing requirement.

Note: These endorsements are subject to a completed legal/privacy review and AI risk assessment.

Pilot overview

In June 2024, the Government Chief Digital Office (GCDO) endorsed the cautious use of assistive AI tools like Microsoft 365 Copilot (M365 Copilot) in government. In response, Inland Revenue launched a structured pilot focused on meeting administration, leveraging M365 Copilot's integration with Microsoft 365 apps to access internal data and streamline workflows.

The pilot involved 20 participants from the project team EI&K Information Governance team. It aimed to:

- Assess if benefits seen by agencies like ACC and ATO could be replicated internally.
- Identify and address IR-specific risks, including privacy, security, and compliance with the tax administration act.
- Build internal expertise and guide change and adoption strategies.

The Information Governance team has documented their experience through two key reports:

- **M365 Copilot midway report December 2024**: Captures early insights from their initial eight weeks of use.
- **M365 Copilot second midway report March 2025:** Reflects on progress after 18 weeks of use.

A broader feedback survey was conducted in June 2025 across all current license holders, including members of the Copilot Enablement Project Team. These sources form the foundation of our findings.

Productivity Benefits

89% of people used M365 **Copilot Daily**

5 hrs saved each week per person on average

44% saved **more than 5 hrs per week**

94% found time savings from creating content

83% found time savings **locating documents**

100% said that Copilot made them **more productive**

89% reported **better quality of work**

Copilot was most widely used in **Microsoft Teams**

Data source: End of pilot survey. 18 pilot participants responded to the survey.

Inland Revenue
Te Tari Taake

Pilot Conclusion

- The Information Governance team found for them that the **cost of the tool was outweighed by the productivity gains** achieved through M365 Copilot, it has been shown to have **significant productivity gains** and time savings.
- **Teams** and **M365 Copilot Chat** offer the **most advanced Copilot features**, enabling interactive, iterative conversations that feel like collaborating with a partner.
- Other apps like **PowerPoint and Excel have limited functionality**, making them useful for quick content generation but less effective for refinement—requiring IR staff to apply critical thinking and careful review.
- There is a need for people to shift and **adjust how they work** in order to embed the use of Copilot in how they work, this can take time but has significant benefits.
- **Benefits** were found to have **increased over time** as users embedded M365 Copilot into their workflows.
- Compared to Copilot Chat, people found **significant efficiency gains** from being able to find, summarise, and synthesise **large amounts of internal IR information** more quickly.

Implementation Considerations

- **Access and Allocation**: IR people would request M365 Copilot access via the Support Portal, outlining how it will support their work. See slide 2 for options for approval process.
- **Prioritisation**: Licences are prioritised for requests with high potential for time savings, innovation, collaboration, or benefit to the wider organisation.
- **User Expectations**: Licence holders are expected to provide feedback through a form and may be invited to share insights on the tool's usefulness. Lack of feedback may result in licence reallocation to someone on the waitlist.
- **Training and Support**: Training will focus on high-value areas like Teams and Copilot Chat, supported by curated Microsoft resources, trainer-led sessions, and visual guides to help users discover key features efficiently. Completion of the Copilot Fundamentals course in Ātea will be a pre-requisite to getting M365 Copilot access.
- **Cost and Usage Management**: The licence costs 9(2)(b)(ii) per month per person. Using the average IR employee salary ($93,500), this means our people would **only need to save 9(2)(b)(ii) minutes per week to get a return on investment**. For ongoing usage management, the WTS team will monitor usage to ensure cost-effectiveness. Licences unused for 60 days will be revoked and reassigned to waitlisted users.

Implementation and license options for EPPC

The shift to BAU will be via a controlled service and include a phased deployment, focusing on rollout to a small, targeted group initially to gain feedback on the service prior to making it more widely available.

Note: To gain access to M365 Copilot, all users must complete the existing Copilot Fundamentals training and agree to a set of 'Dos and Donts' specific to this version of Copilot.

Option A

Purchase 100 licenses (this is the minimum required for detailed monitoring) at 9(2)(b)(ii) for 12 months.

Approval required to gain license: Restricted user base, DC approval needed and will depend on license availability.

User base: Would require strict eligibility criteria - Super users, business support, governance teams – would need to be tightly controlled and managed.

Option B

Purchase 500 licenses at 9(2)(b)(ii) for 12 months.

Approval required to gain license: Restricted user base, Tier 3 leader approval needed and will depend on license availability.

User base: Individuals and business areas can apply, licenses could be assigned based on an agreed approval process and availability.

RECOMMENDED

Option C

Purchase licenses for all of IR 9(2)(b)(ii) for 12 months.

Or

9(2)(b)(ii) for all non-customer staff (excludes those in CSO, CCS, ComCom Officer, Business Lifecycle Manager roles).

Approvals: Would not be required.

User base: As above.

Next steps

- Legal and privacy review, and the AI risk assessment will be completed prior to attending additional governance groups, no further review from the Security Sub-Working Group is required.
- We will be attending the following groups to inform or request endorsement or approval:

Group	Date
AI Oversight Group	25 June
Design Authority	1 July
Technical Design Authority	1 July
Planning Prioritisation Group (PPG)	16 July
Enterprise Priority and Performance Committee (EPPC)	24 July

Pending endorsement & approval:

- Implementation of a BAU process.
- Purchasing M365 Copilot licenses (as directed by EPPC).
- Certification & Accreditation (C&A) from CISO will be obtained, likely to be a change certificate against the Microsoft platform.

Appendix

Inland Revenue
Te Tari Taake

Appendix 1: Key findings on features and usage

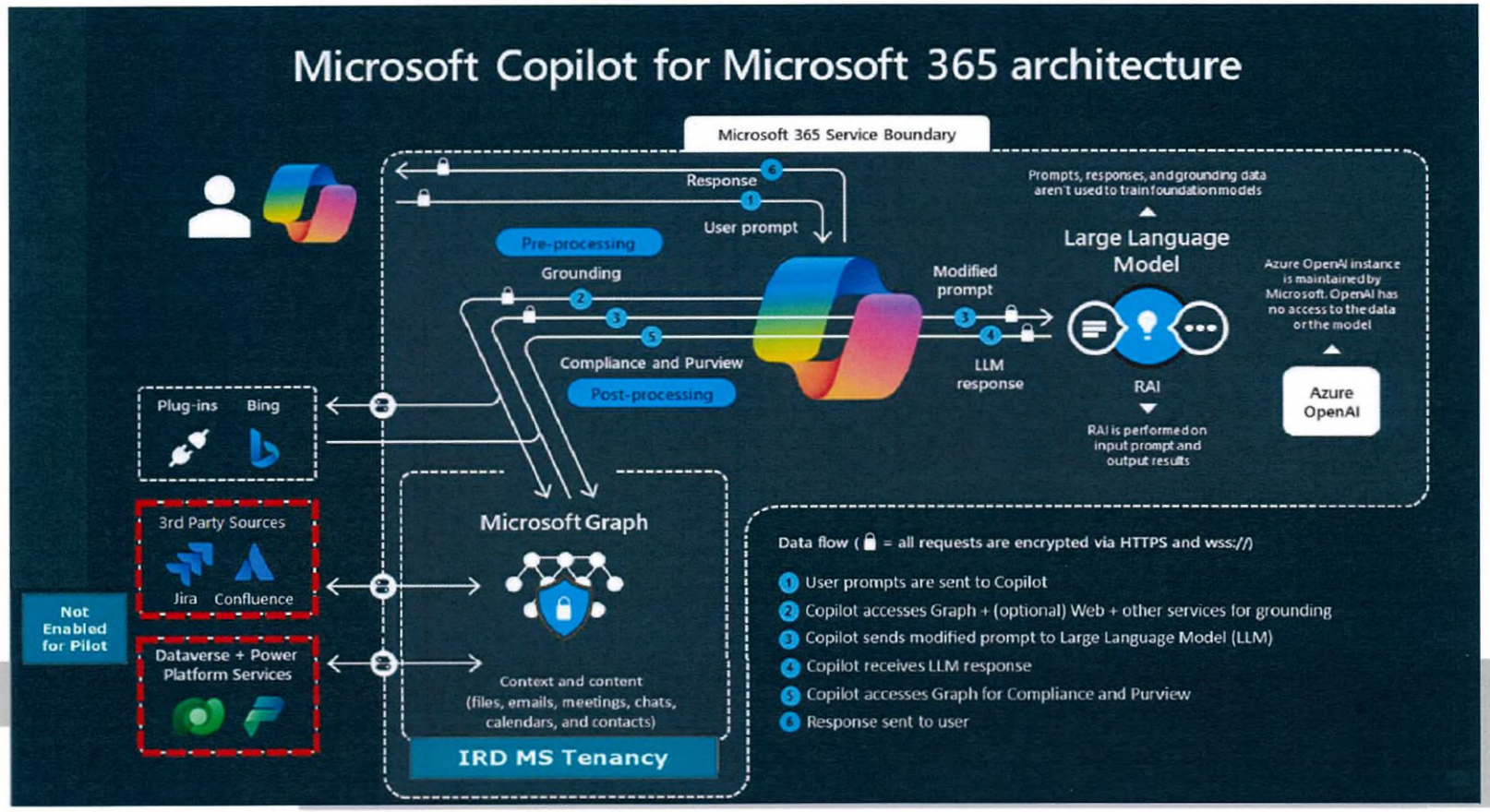
- Microsoft **Teams**, M365 **Copilot Chat**, and **Outlook** are the **most used entry points**, with Teams leading at 91% active usage and 72% daily use.
- M365 **Copilot Chat** is valuable for its ability to **generate content** using information from across the M365 suite and **locate documents** without needing exact file names.
- **Teams supports productivity** through real-time **meeting notes**, capturing **key points** and **actions** so people can stay focused during meetings.
- **Word** is used for fast **document creation** and refinement, allowing users to **rewrite and polish content** using natural language prompts.
- **PowerPoint** & **Excel** have **limited functionality**, useful for quick drafts but less effective for deeper editing or analysis.
- People see **greater benefits over time** as they adapt workflows, embed Copilot into their work and upskill in prompting.

Appendix 2: About Microsoft 365 Copilot

How Copilot for Microsoft 365 works

- Microsoft 365 Copilot is powered by large language model (LLM) which is a form of generative AI.
- It is powered by Azure openAI which is a Microsoft hosted version of ChatGPT by openAI. This model sits within the Microsoft boundary, meaning that it is shared across its Microsoft customers (Note: openAI has NO access to this model)
- IRD leverages the Australia region of this model that is hosted out of Sydney.
- Microsoft 365 Copilot gathers data from Microsoft Graph (gateway to our information and data held in Microsoft 365 services and applications) and is integrated into the Microsoft 365 suite of products (e.g. Word, Excel, Teams etc).
- M365 Copilot uses the existing Microsoft Tenant data and is restricted to what individual users have access to already as part of their existing account permissions, it respects security, compliance and privacy policies and processes that are already in place for Microsoft 365.

Appendix 3: M365 Copilot architecture



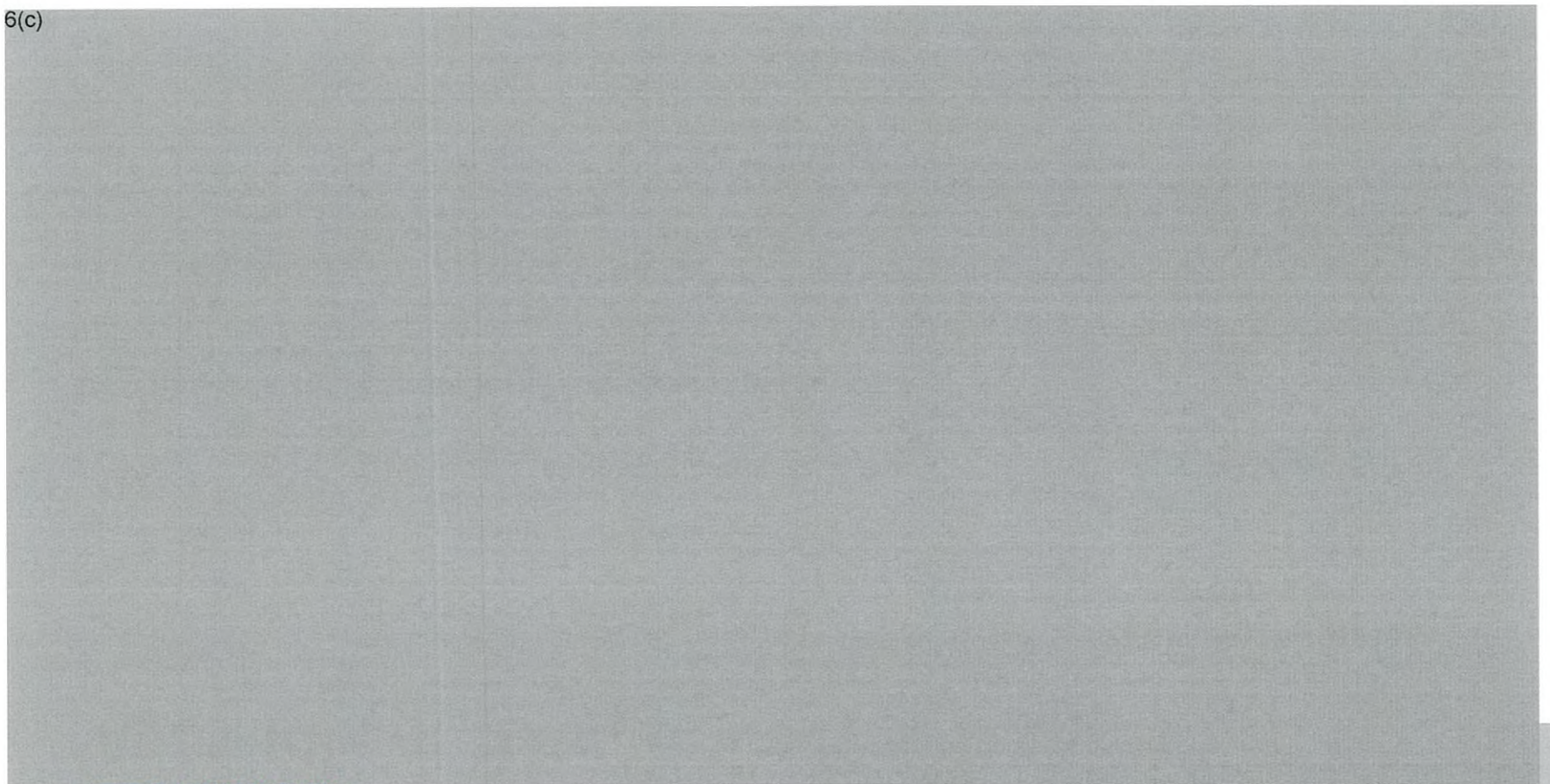
Appendix 4: Data, privacy & security

- Tenant Isolation: Copilot only uses data from the current user's Microsoft 365 tenant.
- Training Boundaries: Microsoft does not use any business data to train the foundational LLMs used by Copilot.
- Permissions: Copilot only surfaces organizational data to which individual users have at least view permissions.
- Encryption: Chat data sent to and from Copilot is encrypted in transit using TLS 1.2+ and at rest using AES-128.
- Data Retention: Microsoft doesn't retain prompts or responses from users beyond a short caching period for runtime purposes only.
- No 'Eyes-On' Access: Microsoft has no direct access to the chat data.
- Restricted SharePoint search: There is a setting that enables us to maintain a list of SharePoint sites that can be used in an organisation wide search and for Copilot for M365

Enterprise data protection

- Our data is secured, it is encrypted at rest and in transit.
- Existing access permissions and policies are respected.
- Our data is not used to train foundational models.

6(c)



Appendix 6: All of government guidance

The GCDO's current position – **Assistive AI tools like M365 Copilot should be increasingly used, while still exercising caution** When considering using Microsoft Copilot for Microsoft 365 (M365 Copilot) it's important that you leverage existing guidance and the lessons of others, such as those below. Whilst AI brings immense opportunities for innovation and efficiency, the risks are serious and must be carefully managed to protect the integrity both of your systems and the public service overall. Against this backdrop your organisations **can choose to move ahead with adopting M365 Copilot for low risk use cases**, such as productivity.

Assistive AI Pilot Attachment One Detailed Pilot Lessons 20240528

Appendix 7: Calculation for tool cost vs salary

Annual IR employee salary:	**$93,500** per year (source: Te Kawa Mataaho Public Service Commission), **$44.95** per hour
Cost of Copilot:	9(2)(b)(ii) per month per person
Hours of work paid for:	**8 hours** per day, **40 hours** per week

9(2)(b)(ii) (cost per month) / 44.95 (average salary per hour) = 9(2)(b)(ii) hours per month

9(2)(b)(ii) hours per month per person is approximately 9(2)(b)(ii) minutes per month to break even, per week this is 9(2)(b)(ii) minutes

Appendix 8: M365 Copilot
Pre-Pilot Test Summary

Inland Revenue
Te Tari Taake

M365 Copilot – Data and Security settings - Interim Test Summary Report

Test Summary Status		
This report summarises the Integration Testing – Security and Data settings testing conducted for Copilot M365 - Data and Security settings. These tests were successfully executed between the 15/10/24 and 21/11/24 by the ESTCT team. This testing expanded on the scope of the Pre-Pilot validation scope that includes the other copilot enabled Microsoft applications, focusing on data access based on the predefined user roles permissions and data access rules. • All 42 planned test cases have been executed and passed successfully		
Test Exit Criteria		
Criteria	**Met /Not-Met**	**Comment**
All high and medium priority test scenarios are executed	Met	All test scenarios are in JIRA here Under Data and security settings
All defects are documented	Met	All defects are in JIRA here
There are no outstanding severity 1 and 2 defects	Met	
Outstanding severity 3 and 4 defects have been reviewed and approved by business representatives and documented.	Met	
Any workarounds are agreed by the business representatives and documented	Met	
Non-functional requirements have been met, or exceptions agreed upon by business and architecture after risk assessment	Met	
Test Exit/Evaluation Report is approved	N/A	Final Test Exit Report will be sent for endorsement on completion of all testing activities

M365 Copilot – Data and Security settings : Executive Summary

Project Context :
The objective of the M365 Copilot – Data and Security settings testing is to provide a view of how Copilot M365 will behave within the IR graph with our current security and data settings, as well as highlight potential risks based on the testing conducted.

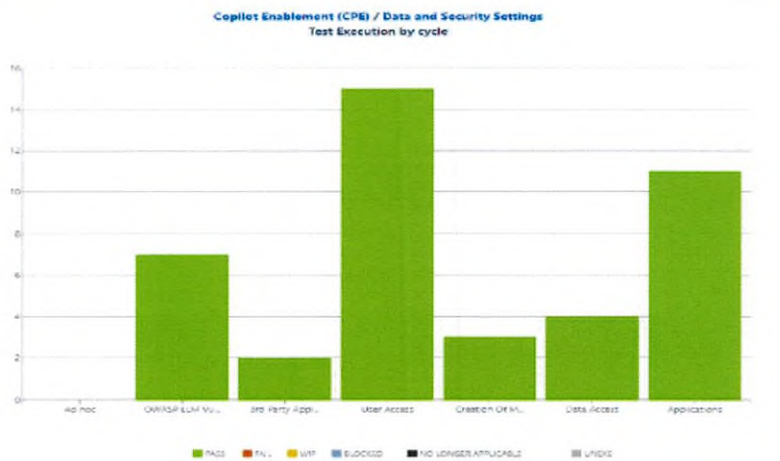
Test Execution Scope Summary :
Testing M365 Copilot – Data and Security setting is as shown on the right, also documented here. There has been no variance in the testing scope:

6(c)

Inland Revenue
Te Tari Taake

M365 Copilot – Data and Security settings : Test Execution and Defects Summary

Test Execution



Pilot entry validation - Defects by Status

Severity	RESOLVED	CLOSED	T:
2 - High	1	0	1
3 - Medium	0	1	1
4 - Low	0	1	1
Total Unique Issues:	**1**	**2**	**3**

6(c)

Inland Revenue
Te Tari Taake

M365 Copilot – Pre-Pilot Validation : Circulation

Document control

File name and path	Template - Test Exit report v1.0.pptx
Contact person	9(2)(a)
Status	Final
Template Version	V1.0

Document review history

Version	Date	Change description	Contact
V0.1	27/11/24	Initial draft	9(2)(a)
V0.2	04/12/24	Feedback added	
V1.0	12/12/24	Final	

Supporting information

Document/link	Description
Pilot -Copilot for M365 - Test Plan v1.0.pptx	Test Plan

Circulation List

To

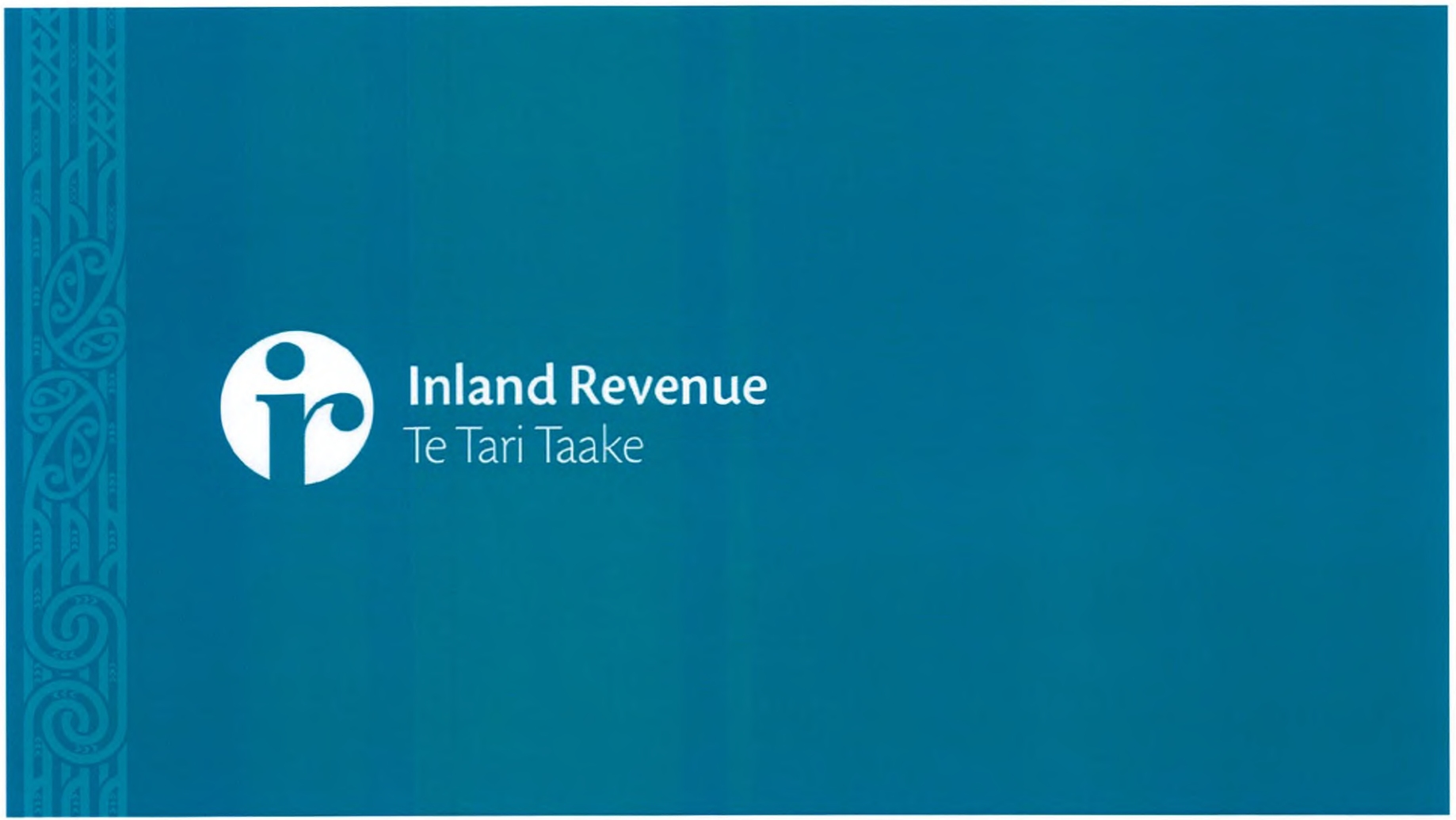
- 9(2)(a) - Initiative Lead
- 9(2)(a) – ESTCT Test Lead
- 9(2)(a) – ESTCT Test Manager
- **Prajakta Panse** - Domain Lead – Testing
- **Jesse Thwaites** - Intelligence Leader - EI&K (Sponsor)
- **Malcolm Breadmore** – Enterprise leader - technical services (Sponsor)
- **Brijesh John** – Domain Lead – Tech Architecture
- **Graham Poppelwell** - Domain Lead - Information Governance & Sharing
- **Prakash Selvaraj** – Technology Specialist
- **Tim Crook** - Domain Lead (Business owner)
- **Adwin Singh** – CISO Representative

CC'ed

- 9(2)(a) Information Specialist - EI&K
- **Preeti Shikhariya -** Information Specialist - EI&K
- **Phyilida Crawford** - Domain Principal - Technology Experience
- **Makayla Stewart** – Change Analyst - Change, Design & Enablement
- **Matthew Towers -** Technology Specialist -CSO
- **Tom Greenwood-Thessman** –Technology Specialist - CSO

The following have supported in the development of this document

- 9(2)(a) - ESTCT Test Lead
- 9(2)(a) – Senior Test Engineer


Inland Revenue
Te Tari Taake

Item 8

Inland Revenue
Te Tari Taake

Microsoft 365 Copilot

Production BAU pilot with Policy & TCO

AI Oversight Group

Presenters: Phil Whittington, Tim Crook, Makayla Stewart, Phyllida Crawford

Sponsors: David Carrigan, Malcolm Breadmore, Jesse Thwaites

Wednesday 25 June 2025

Purpose and Recommendations

The purpose of this item is to close out the initial Pilot of M365 Copilot, report back on findings and propose a pathway forward for a second pilot in the production environment that is focused on establishing and testing a BAU process and evaluating the use of Copilot Agents.

It is recommended that the AI Oversight Group

- **Note** the findings and close out of the M365 Copilot Pilot (A full copy of the close out report is available here).
- **Approve** a production BAU service with a controlled/restricted 80 user pilot of M365 Copilot in Policy/Tax Counsel Office.
- **Approve** the use of Microsoft's pre-built agents by the pilot users (Researcher, Analyst, Project Manager).
- **Approve** the creation of two bespoke agents (Policy Analyst and Policy Quant) following on from the Technology Agent.
- **Agree** to accept the documented risks and proposed mitigations of the new pilot.
- **Discuss** and **Agree** on general pathway forward for M365 Copilot and the key items that must be resolved prior to seeking approval from PPG & EPPC for an organisation wide rollout (slide 7).

Initial Pilot overview

In June 2024, the Government Chief Digital Office (GCDO) endorsed the cautious use of assistive AI tools like Microsoft 365 Copilot (M365 Copilot) in government. In response, Inland Revenue launched a structured pilot focused on meeting administration, leveraging M365 Copilot's integration with Microsoft 365 apps to access internal data and streamline workflows.

The pilot involved 20 participants from the project team and EI&K Information Governance team. It aimed to:

- Assess if benefits seen by agencies like ACC and the Australian Government could be replicated internally.
- Identify and address IR-specific risks, including privacy, security, and compliance with the Tax Administration Act.
- Build internal expertise and guide change and adoption strategies.

The Information Governance team has documented their experience through two key reports:

- **M365 Copilot midway report December 2024**: Captures early insights from their initial eight weeks of use.
- **M365 Copilot second midway report March 2025:** Reflects on progress after 18 weeks of use.

A broader feedback survey was conducted in June 2025 across all current license holders, including members of the Copilot Enablement Project Team. These sources form the foundation of our findings.

Productivity Benefits

89% of people used M365 **Copilot Daily**

5 hrs saved each week per person on average

44% saved **more than 5 hrs per week**

94% found time savings from creating content

83% found time savings **locating documents**

100% said that Copilot made them **more productive**

89% reported **better quality of work**

Copilot was most widely used in **Microsoft Teams**

Data source: End of pilot survey. 18 pilot participants responded to the survey.

Inland Revenue
Te Tari Taake

Key use cases apply across a wide variety of roles

Creation of content: Accelerates the development of documents, presentations, and communications by generating high-quality first drafts using internal information from multiple sources. This helps produce IR specific content that is more relevant and accurate than Copilot Chat.

Summarising information: Distils key insights from emails, chats, meetings, and documents—reducing time spent reviewing content and enabling sharper focus on priorities.

Locating documents: Enables fast retrieval of documents across SharePoint, Teams, and email—without needing exact file names—saving time and supporting responsiveness in high-paced environments.

Troubleshooting: Supports guided resolution of technical issues in the M365 suite - through AI-guided assistance, reducing delays and improving self-sufficiency.

Feedback and Coaching: Provides real-time guidance to refine written content—enhancing clarity, tone, and confidence in communication.

> A great aspect is being able to retrieve documents with **greater ease and speed** – less of 'where was that saved again?'.

Information Specialist, E&IS

> I'm able to be **more present in meetings** without having to worry about taking notes and **being confident** that the important points are being captured.

Information Specialist , E&IS

Pilot key findings

- **Productivity gains outweigh cost**: The Information Governance team validated Copilot's value—time savings and quality improvements justify our investment.
- **Most effective tools**: Teams and Copilot Chat deliver the richest functionality—enabling iterative, partner-like collaboration.
- **Adoption curve**: Benefits increase over time as users embed Copilot into workflows and improve prompting skills.
- **Behavioural shift required**: Embedding Copilot into daily work takes time and mindset change—but delivers value.
- **Usage patterns**: Teams leads with 91% active and 72% daily use; Outlook and Word also widely adopted.
    - **Copilot Chat and Outlook excels**: Enables fast summarisation and synthesis of IR content—no need for exact file names.
    - **Word**: Effective for drafting and refining documents using natural prompts.
    - **PowerPoint & Excel**: Useful for quick generation but limited for deeper editing—require user review and critical thinking.

Shifting M365 Copilot to a controlled BAU

- Our organisation is positioned well to shift to offering Microsoft 365 Copilot as a controlled business-as-usual service from most respects.
- There are 4 key items that have been identified as needing further refinement prior to requesting approval from PPG and EPPC for a full release:
    - Handling of sensitive revenue information
    - Implementation of hard controls via Purview
    - Completion of legal review and risk assessment
    - Controls for the risk of unintended disclosure
- In order to support growth in the user base of M365 Copilot, we need to leverage the existing BAU support capabilities offered by the Service Desk and Tech Services support teams.
- A controlled BAU pilot, would enable us to leverage these capabilities while **restricting licenses** to the existing pilot group, policy and TCO. Note: Users outside of these groups would not be able to request access or a license.

Pathway forward: New production BAU pilot in Policy and Tax Counsel Office (TCO)

Inland Revenue
Te Tari Taake

IR3S Limited Access 13 June 2025

Approach

- 80 licences funded from **Policy** budget for 3 months.
- Business areas: **Policy** and **Tax Council Office**
- Variation across areas (Policy and TCO in particular) provides opportunity for **comparative analysis**
- Assessment will be two **self-assessment surveys** with business area comparisons (at 6 week point and 12 week point) and Copilot Dashboard features to provide usage insights (ongoing).
- Focus will also be on use of agents, and the **identification of power-users** to help with guidance for rest of IR
- Will also assess the value of bespoke agents
- Project teams to **review progress early August** to determine if we can proceed to PPG and EPPC for their approval of a wider rollout.

We will include use of Microsoft Copilot Agents + bespoke agents

Agentic AI

- "Specialised AI employees" – each agent has expertise in a specific domain and can handle multi-step work
- Plain genAI: "Help me write this email"
- Agent: "Research the impact of carbon border adjustments on New Zealand exporters, analyse fiscal implications and collect international policy precedents"

Agents (MSFT)

- **Researcher:** Complex, multi-step research with advanced reasoning
- **Analyst:** Data analysis and visualisation
- **Project Manager:** Automates project management from plan creation to specific tasks

Agents (Bespoke)

- **Policy Analyst:** Built on templates and workflows
- **Policy Quant:** Proposes ways of modelling policy effects

Evaluation: All license holders agree to complete productivity assessment

- 9(2)(b)(ii) [redacted]
- **Evidence from our initial pilot and internationally** is strongly in support of this being exceeded
- Given that, we get enough certainty on whether benefits exceed costs through **self-assessment rather than** more advanced **telemetry** or **control group trials**
- We will **leverage the newly unlocked Copilot Dashboard features**- available now that we exceed 50 licenses—to complement self-assessment data with usage insights.

Survey design for point estimates to be focussed on:

- Time saved per week
- Quality perception changes
- Usage patterns across different agents
- Economic value of improvements
- Named responses enable targeted interviews on **agent use**
- Cross-unit variation enables identification of best practices

Inland Revenue
Te Tari Taake

We need to revisit current AI use policy

- Due to powerful context search, Copilot will have access to:
  - Revenue information
  - Sensitive revenue information
  - Budget sensitive information
- We are confident in Microsoft's assurances of no data leakage outside IR tenancy and information is not used to train models
- **Risk acceptance:** Treating as an exception to use policy, subject to legal view (from **Conrad Bace**), and business area endorsement (David Carrigan, and Michelle Redington) with a plan to extend this to cover all licenses holders of M365 Copilot and Copilot Chat in future.

Key risks

- **Unintended disclosure**: existing risks that may become more visible with improved search capabilities – similar to **Stax/Haukāinga** experience
- **Information aggregation:** Tool combines data from multiple sources – users may **share outputs** to **unauthorised** recipients

We have several controls to mitigate risks

Immediate controls (in PoC)

- **Start small:** Limit scale of risks by limiting licences
- **Security preparation**: Participants agree to check OneDrive and storage locations and ensure proper permissions
- **Pre-communications**: Clear guidance on additional risks and review requirements
- **Output review**: Emphasise to participants need to review before sharing (standard work practice with AI)

Medium-term controls (not in PoC)

- **Microsoft purview:** Better control through classifications and more sophisticated permissioning
- **Process refinement:** Based on pilot learnings and emerging best practices

What would risk actually look like? A pre-mortem

Scenario

- Policy advisor uses Copilot to research "infrastructure tax rules"
- Unknown to them, Copilot pulls information from a Budget 2026 folder
- Output includes specific budget details that are then shared in a team meeting, disclosing sensitive budget information

How would this happen?

- Budget folder permissions not set up adequately
- Copilot broad search finds restricted material the user did not know they could access
- Analyst assumes output is safe to share since they did not ask for sensitive information
- Copilot combines restricted budget data with general policy information, revealing sensitive details

What would we do?

- Incident reported to Policy manager and Senior Leadership team
- People who now have the information get added to "budget list"
- Determine whether the issue is folder access, or something wider
- Permissions audit across all sensitive folders

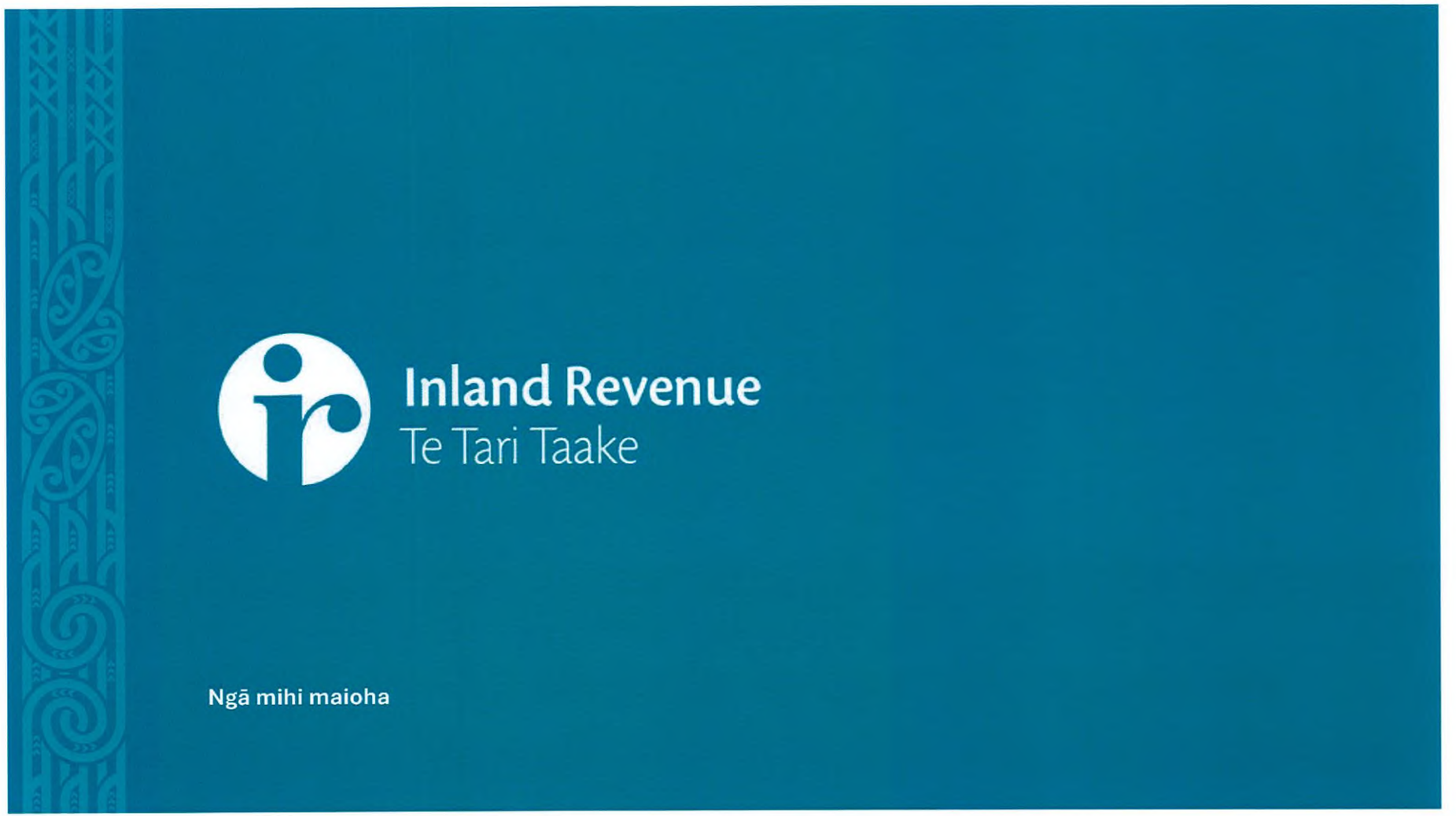
How will we judge success of the initiative?

Primary success indicators

- Resolution of the key items outlined on slide 7
- Break-even threshold of 8 minutes per week time savings per user
- Target 30 minutes per week
- User assessed improved output quality
- Positive feedback on agents
- Useful feedback on production service
- Evidence-based recommendations for broader IR
- Volunteers to prepare material for agent use-cases

Survey design

- Microsoft Copilot productivity assessment survey
- All license recipients commit to completing productivity assessments
- Named surveys for follow-up interviews and case studies
- Non-anonymous approach allows longitudinal analysis
- Cross-unit comparison to test variation
- Best practice identification through analysing high-performing users and teams


Inland Revenue
Te Tari Taake
Ngā mihi maioha

Appendix: Related material

- Copilot Studio (tool for creating agents) preliminary risk assessment: Risk Assessment for Copilot Studio
- Endorsement paper for Enterprise Services Technology agent: DD-1149 Copilot Studio Enterprise Services Technology Agent May 2025
- Microsoft 365 Copilot at ACC: Our journey and insights: M365 Copilot at ACC - Journey and Insights - March 2023
- Australian Government trial of Microsoft 365 Copilot: Full report – productivity | digital.gov.au
- Microsoft 365 Trial Midway Report: Information Governance
- Microsoft 365 Trial Midway Report #2: Information Governance

Item 9

Inland Revenue
Te Tari Taake

Technology Services / Policy
Production pilot of M365 Copilot
Initial findings report and recommendations paper

Artificial Intelligence Oversight Group

Date: Thursday 18 September 2025

Attendees: Tim Crook, Phil Whittington, Phyllida Crawford, Makayla Stewart

Preparer: Tim Crook, Phil Whittington, Phyllida Crawford, Makayla Stewart, Graham Poppelwell

Sponsors: David Carrigan, David Shanks

Purpose and Recommendations

The purpose of this paper is to report on the close out of the initial pilot of Microsoft 365 Copilot (**M365 Copilot**), report on initial findings of the pilot in Policy & TCO and recommend next steps to leverage the benefits found.

It is recommended that the AI Oversight Group:

- **Note:** the project team is attending:
  - Design Authority on 23 September for an inform,
  - Enterprise Priorities and Performance Committee on 25 September for approval, and
  - Technical Design Authority on 30 September for approval.
- **Note** the findings and close out of the initial M365 Copilot pilot (a full copy of the close out report is available here).
- **Note** the initial findings from the production BAU service with controlled/restricted 80 user pilot of M365 Copilot in Policy / Tax Counsel Office.
- **Endorses** our request to move M365 Copilot and the pre-built Microsoft 365 agents to fully BAU, as a controlled service with a phased deployment.
- **Endorse** proceeding to EPPC for key discussions and decisions that are needed to execute the above. These include (details on slide 3):
  - The number of M365 Copilot licenses that should be purchased.
  - Budget uplift to meet licencing requirement.
- **Note** our intention to work with business unit leads and our external training partner to write a detailed rollout plan.

Executive summary

Our AI Strategy and Direction (see Key Messages SIB August 2025) has positioned IR to consider 'game changing AI'. Exploring M365 Copilot has delivered insights about the significant productivity shift seen from the 80+ licences allocated.

IR has implemented Copilot Chat across much of the organisation, improving productivity and work quality despite limitations in accessing internal data. A pilot of M365 Copilot—fully integrated with Microsoft 365—was then launched in August 2024 with 20 participants, showing an average time saving of 5 hours per week. In line with Government direction to accelerate safe AI adoption, a broader pilot began in July 2025 with 80 staff across key business areas, supported by our vendor Engage Squared. This phase introduced Microsoft's pre-built agents, tested BAU support via the Service Desk, and refined training materials. Licences were purchased at 9(2)(b)(ii) per person/month and expire at the end of September.

This paper positions options for Enterprise release of M365 Copilot to meet our strategic ambition for workforce productivity, a move towards AI augmenting our daily work – in all roles. While the technology is 'providing a licence', it's not a single change event; it's engaging with AI as an ongoing business capability and integrating it into our core workflows. There are trade-offs and opportunities that need to be considered as part of this paper:

- **Discounts and funding:** All of Government (AoG) discount and funding window closes 1 October 2025, only buying prior will unlock these benefits.
- **Staged deployment required:** to reduce the risk associated with low quality information management practices, a staged deployment is needed to allow for time for our Information/Knowledge teams to work with high risk areas. If we purchase all licences upfront to secure the discounts and funding, we can then assign them at the right time for the business in a controlled rollout.
- **Adoption journey:** The pace of our AI-enablement journey depends on the option we choose – to accelerate or delay it. Benefits may take time to show while staff upskill and learn to use Copilot effectively. Our people are excited and ready to embrace AI, 2000 people have completed their Copilot Chat training out of 2200 that are eligible.
- **Government directive:** The Government is pushing agencies to accelerate AI enablement safely while maintaining public trust to increase their productivity and gain efficiencies.

Extended pilot key findings

130 minutes saved per person per week

Over **719 total hours saved** across the pilot group

A **return on investment of over 13x**. 81% of participant felt benefits equalled or exceeded cost.

76% of participants found their **work quality** was **"somewhat better"** or **"significantly better"**

Key use cases

- Summarising Teams meetings
- Drafting documents & emails
- Locating documents & content
- In-depth research and analysis
- Rewriting for clarity & conciseness

Proportion of participants feeling **highest levels of stress fell from 21% to 6%** with Copilot

Increase in workload capacity of 12% across participants

Copilot respected access permissions and no data leakage or unintended disclosure incidents occurred

Copilot was used an **average of 236 times per day**

* Detailed pilot results are included in Appendix 2

Source: August 2025 Survey

Conclusion: Efficiency and quality gains are very large and we have enough information to make org-wide decision

- Most compelling reason is **financial return on investment**. Data is from self-reported survey, but even allowing for overreporting, the average reported weekly gain (130 minutes) is 13x the break-even point.
- **Self-reported metrics were very positive:** 76% reported **quality** improvements, and those reporting **highest levels of stress fell** from 21% to 6%.
- Copilot **scaled across different roles**: gains were reported across all business units and levels. The legal assessment found the use of SRI and PII acceptable from a legal risk perspective. This suggests **readiness for broader rollout including customer-facing roles**.
- M365 Copilot **gives unlimited agentic queries** for the use of standard agents, future proofing IR as we adopt agents more broadly.
- The Government is pushing agencies to **accelerate AI enablement safely** while maintaining public trust to increase their productivity and gain efficiencies:

Using AI effectively can significantly improve customer experience, improve efficiency and allow public servants to focus on frontline services

Chief Digital Officer Paul James

Minister for Digitising Government, Hon Judith Collins KC

I want public service leaders to embrace AI and support their teams to use it safely and effectively –

- **Our recommendation is to proceed with staggered organisational-wide rollout, and do so quickly given incentives currently being offered by Microsoft.**

Our Copilot journey & recommended next steps

1

Enabled Copilot Chat and upskilled our workforce

IR has trained and given 2,000+ people access to Copilot Chat.

This enabled our people to **safely upskill in and use a secure generative AI chat tool.**

COMPLETE

2

Staggered and scaled deployment of M365 Copilot to drive Enterprise efficiency and AI capability across our entire organisation

Broadly deploy M365 Copilot to IR people to **realise productivity gains at scale.**

This will position IR as an **AI-enabled organisation** and provide a **foundation for further enablement** via the use of agents.

Apply learnings from our pilot to maximise the benefits from M365 Copilot and **manage use responsibly and safely.**

WE ARE HERE

3

Drive operational excellence and productivity gains with M365 Copilot agents

Realise further benefits from our investment in M365 Copilot licences by deploying tailored **agents for IR.**

We are currently piloting an IT Support agent.

PILOT / FUTURE

Option 1: Purchase licences for all of IR

Overview:

Purchase 5000 licences for all IR people excluding non-workers. Every person would receive a licence and training.

Licences would be acquired in a single block by 31 September 2025 to secure a volume discount, but rolled out in a staggered approach:

9(2)(b)(ii)

We will work closely with the business to coordinate a staged, smooth onboarding process including proactive scheduling to accommodate operational needs.

With this option Microsoft would make training and change management support to the value of $0.250m available.

Pros

- **Maximum impact:** Every IR person can use Copilot, so our whole organisation benefits from better productivity and quality.
- **Builds AI fluency:** All our people can safely upskill in AI technology, allows our frontline people to have hands on practical experience with AI, developing their AI fluency.
- **Future readiness:** Sets IR up to use AI agents across the organisation without extra cost.
- **Supports culture change:** Encourages an IR cultural shift to AI-enabled ways of working and peer learning. Supporting IR's AI strategy by being a leading agency.
- **Vendor support and partnership:** Allows us to get better licence pricing and significant Microsoft funding for training and change management.
- **Simplifies administration:** No need to manage who gets licences or prioritise access.
- **Organisation-wide training:** Training can be efficiently rolled out organisation-wide.
- **Staged release options**: Ability to stage the release and make use of early adopter groups

Cons

- **Upfront cost:** This option requires the largest upfront investment.
- **State of date:** Policy and TCO have higher quality information management practices, this is not the case for other business areas, this could impact its usefulness in other areas.

6(c)

- **Scheduling and timing:** Mandatory training would take 2 to 3 hours per person, time away from core work

Staged rollout based on business area readiness with remediation work as required

Late September

IKM issue **guidance and checklist** for copilot knowledge management readiness

1 October

Some business areas **go live after mandatory training** (Ātea eLearn + 'dos and don'ts') e.g. **Policy, TCO, Technology Services**

Remaining business areas use **guidance and checklist** and determine whether additional EIK/IKM support is required

Advanced **SharePoint management check** to identify areas for further IKM support (if any)

Leveraging **early adopters** for business areas that require further IKM support

8 October

CCS begins proofs of concept (2-4 weeks)

Early November

CCS learnings from **proofs of concept available**

Discussions with EIK about any **remediation work**

CCS customer segment lead **approval for go live decision**

Business areas that had to undertake remediation work start to **go live**

No further approval from EPPC, PPG or AIOG required

Options

- All options include the M365 Copilot licence and Microsoft's pre-built agents including Researcher and Analyst.
- **We recommend Option 1** to realise benefits without delay, make efficient use of Microsoft-funded training, and set up IR with a solid foundation for the next step in our AI enablement journey – using AI agents to streamline business processes across IR.
- We will write a separate paper about the specifics of the rollout approach.

OPTION 1: Staged rollout to all IR

9(2)(b)(ii)

Pros: Maximises impact by giving everyone access to productivity, quality, and AI fluency benefits. Supports future readiness, simplifies administration, secures vendor support, fosters a cultural shift to AI-enabled ways of working.

Cons: Higher upfront investment. Some users may not fully adopt or benefit from Copilot. Mandatory training (2-3 hours per person) impacts core business time.

RECOMMENDED

*OPTION 2: Non-customer facing

9(2)(b)(ii)

Pros: Keeps upfront costs lower, reducing financial risk if uptake varies. Supports AI fluency and a unified shift to AI-enabled ways of working within corporate functions.

Cons: Limiting access reduces coverage and excludes customer-facing roles - our organisational front door - from AI fluency and future readiness. May lead to cost inefficiencies for unlicensed users using agents in future.

*OPTION 3: 500 people

9(2)(b)(ii)

Pros: Keeps upfront costs lower. Allows tighter control over training and usage monitoring.

Cons: Delays IR's AI enablement journey. Misses opportunity to access licence discounts and funded training support. Significantly limits coverage which may create inequity across and within teams. Adds admin overhead to prioritise and approve licences.

** For more information on these options and detailed costs, see the appendix.*

Cost comparison and considerations for options

M365 Copilot requires a separate license that is in addition to the E5 licenses currently used across IR. Our AoG pricing for a base license is set at 9(2)(b)(ii) per user, per month.

9(2)(b)(ii)

Additionally, Microsoft provides change and adoption credits through a partner to support a comprehensive program. This includes training and change management assistance to maximise the value IR gains from its investment, facilitated by the M365 Copilot project team.

If IR does not act now, we risk 9(2)(b)(ii) delaying Copilot deployment—including to CCS—at the best time of year, and limiting our people's access to generative AI tools that build AI fluency and practical skills. Delaying could mean higher costs, lost productivity gains, and falling behind in AI enablement.

9(2)(b)(ii)

*Roll out across the year, cost split over the year and Includes IR change funding.

Risk Management

Legal and AI risk assessment

The legal view concluded that IR's proposed use of M365 Copilot will comply with and satisfy our duty to protect the integrity of the tax system and our obligation to maintain the confidentiality of sensitive revenue information. From an AI and legal perspective, the key risks include:

- Public perception risks
- Handling of sensitive or classified information
- Compliance with privacy and data protection obligations
- Existing issues with data permissions and sharing in our organisation

6(c) [redacted]

This view, that the risk level is acceptable, is provided that the mitigations detailed in the Risk Assessment for Microsoft 365 Copilot are applied.

Mitigations and controls

- Key mitigations include mandatory training (Copilot Fundamentals course), reinforced by "Do's and Don'ts" and a Copilot champions network to support peer learning.
- Role-based access controls ensure only licensed users can access Copilot with access only being granted once the mandatory training has been completed, and all data remains within IR's Microsoft 365 tenant hosted in Australia.
- Copilot's built-in safeguards—such as encryption, content filtering, and responsible AI principles—are complemented by human-in-the-loop oversight, requiring users to validate outputs before use.
- Any rollout of M365 Copilot will be done in a staged approach. This will allow our Information and Knowledge teams to support users with any necessary data remediation, ensuring they can get the most out of Copilot and reducing the risk posed by any poorly managed information.
- In future, Microsoft Purview controls will enable enhanced monitoring and data loss prevention.

These mitigations reduce the overall risk and will ensure that we continue to comply with and satisfy our duties and obligations while adopting M365 Copilot.

Appendix 1

Detailed Options and Costing

Inland Revenue
Te Tari Taake

Option 1: The case for Customer & Compliance Services: Value add with AI augmented activities

- With IRs budget 2025 focus on increase tax revenue, M365 Copilot can support a higher return for debt and audit activities. **Processing times for audits can be greatly reduced** enabling more cases to be processed in less time.
- For tax technical work, M365 Copilot can help **ensure all aspects of information is considered about a case**. Augmenting the decision maker by having an AI operate as an assistant in preparing first drafts, saving time and increasing quality.
- Teams meeting with M365 copilot features can **drive efficiencies, such as, target moving hour long meetings to 45 mins.** Meeting content can be AI generated from a recording or transcription notes. No hold up if late attendees as Copilot can be used to get a real-time summary in the meeting of what has been missed.
- Copilot is likely to **benefit knowledge within our organisation** through improved knowledge management capabilities across the organisation through enabling us to identify gaps and locating content for bulk updates. It is potentially a suitable alternative to the knowledge surfacing tool.

CCS

Deploy M365 Copilot to enable audit cases to completed faster

Copilot enables natural language prompts for excel, no longer limiting analysis by the ability to write formulas.

Analysis of customers information can be linked out of SharePoint folders for audit working papers and risk modules to prepare final audit reports in minutes.

START HERE

CCS

Deploy M365 Copilot to access knowledge content

Navigate and locate information from Te Mātāwai . There are new SharePoint 'agents' that are part of M365 that would provide real benefits for access, especially for new employees to IR.

START HERE

CCS

Drive productivity gains with M365 Copilot and Agents

Using agents for customer facing work whether its helping draft web messages or rewording content to make it easier for customers to understand. A precursor for automating AI agent generated responses.

PILOT / FUTURE

Option 2: Purchase licenses for all non-customer facing staff

Overview:

Approximately 2,200 licenses would be purchased for all our people excluding those in the following roles:

- Customer Service Officers
- Customer Compliance Specialist
- Customer Enquires Assistant
- Community Compliance Officers
- Business Lifecycle Managers

9(2)(b)(ii)

With this option there would be Microsoft training and change management support to the value of $100k made available to us for a rollout.

Pros

- **Lower upfront cost:** limits your spend to the most impactful users, keeping costs lower, reducing the financial risks if Copilots benefits aren't seen in some roles
- **AI Fluency:** Enables more of our people to safely experiment and learn to use and engage with AI technology upskilling them in AI fluency.
- **More unified shift:** Our 'back office' and 'corporate functions' would have a unified shift towards AI enabled ways of working.

Cons

- **Significantly reduces coverage:** Only 50% of our people would have Copilot, this limits the benefits to a smaller group of users
- **Front door to the organisation:** Large numbers of our people enter our organisation through the customer facing roles, excluding them from the rollout reduces our AI fluency overall, particularly for the entryway to our organisation and the future of our workforce.
- **Agent use for unlicensed users:** When our people without licenses use agents, they would incur additional cost. Whereas those with licenses can use all agents within the Microsoft ecosystem without incurring the additional cost
- **Lower Microsoft support:** Less Microsoft funded support would be available for training and change management support.
- **Uneven license allocation:** Uneven distribution could cause friction leading to a 'have/have not' dynamic within teams and business areas

Option 3: Purchase 500 licenses

Overview:

Purchase 500 M365 Copilot licenses – a targeted deployment that would require a process for determining who would be assigned the licenses and who can approve the allocation of them.

9(2)(b)(ii)

No Microsoft funding would be made available as part of this option.

Pros

- **Lower upfront cost:** limits your spend to the most impactful users, keeping costs lower, reducing the financial risks if Copilots benefits aren't seen in some roles
- **Control:** A smaller pool of users is easier to manage and organise training for, and for monitoring participants usage of Copilot

Cons

- **Significantly reduces coverage:** Only 30% of our people would have Copilot, this limits the benefits to a small group of users
- **Uneven license allocation:** Uneven distribution could cause friction leading to a 'have/have not' dynamic within teams and business areas
- **Approval process:** With a cap of 1500, a strict approval criteria and process would need to be established to ensure that licenses are allocated t those who will benefit most
- **Ongoing monitoring:** To continue to ensure the most value is being gained, ongoing monitoring of usage will be required which may involve reallocate under utilised licenses
- **Agent use for unlicensed users:** When our people without licenses use agents, they would incur additional cost. Whereas those with licenses can use all agents within the Microsoft ecosystem without incurring the additional cost
- **Lower Microsoft support:** Less Microsoft funded support would be available for training

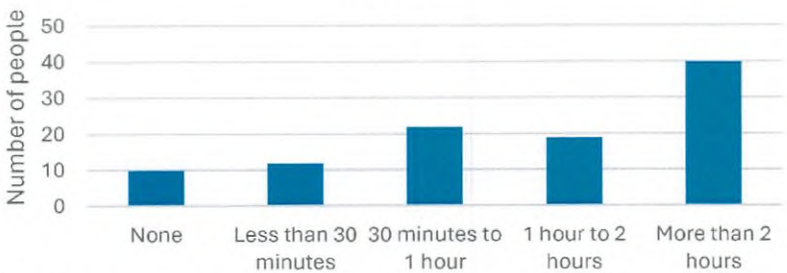
Appendix 2

Detailed pilot results

Inland Revenue
Te Tari Taake

Productivity and Efficiency

How much time did you save by using Copilot per week?



What our people told us

"Copilot enabled me to do research quickly... I would not have been able to achieve this in the time provided without copilot." – Policy Advisor (L3)

"Copilot Researcher was a game changer" Tax Specialist (L2)

"[Copilot helps with] starting new documents and creating a structure that I can then use to then fill in the blanks. Makes sure I get the most important points and don't forget something." – Information Specialist (L2)

"Having Copilot provide a rough first draft of the minutes has allowed me to focus more on in-meeting admin tasks and on the conversation from attendees." – Business Support (L2)

"I will sometimes use co-pilot to help when I'm scheduling multiple meetings looking for suitable availability of rooms around the organisation. This saves a huge amount of time when booking" – Information Specialist (L2)

Findings overview

The productivity assessment survey was conducted in August 2025 (4 weeks into the pilot). Findings show overall significant productivity gains across a large variety of roles with an average of 130 minutes saved per person each week.

From our first pilot, we found that time savings increased over the first 3 months while people were increasing their AI fluency and prompting skills.

Key use cases

The use cases identified demonstrate broad applicability across our organisation, reflecting the common ways our people engage with the M365 applications – spanning diverse roles.

Document and content drafting: Creating drafts for documents, preparing presentations, and rewording of reports.

Meetings: Using Teams recaps for recorded meetings to get action points and quickly review discussions, and for drafting meeting minutes. The info governance team saw a halving of the time taken to draft minutes when using Copilot.

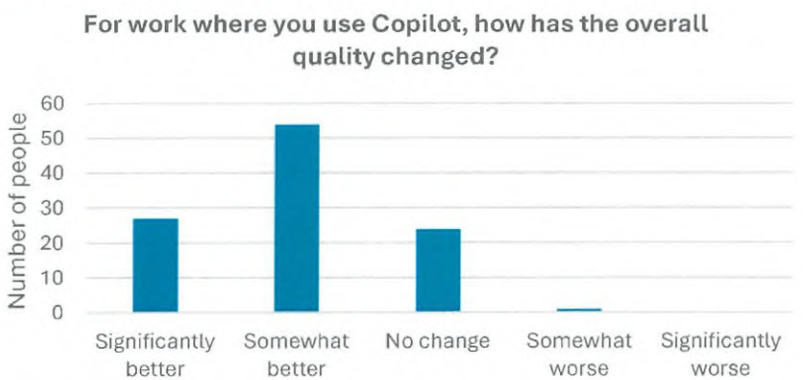
In meeting support: During live meetings, Copilot enables users to catch up on missed content by summarising earlier discussions in real time.

Research and analysis: Deep analyses of large documents or volumes of documents for quick summaries or identification of themes and gaps.

Locating documents: Quick location and retrieval and summarisation of documents, and linking to related documents.

Source: August 2025 Survey

Quality of work



What our people told us

"Copilot proved to be very helpful to assist in managing my ADHD and improving the quality of my work output through suggesting changes which improved the legibility of my work." – Business Support (L2)

"I think the quality improvements are more valuable than the time savings" – Tax Counsel (L3)

"Improving sentence clarity and conciseness when trying to summarise complex ideas." – Policy Advisor (L1)

"As a leader, co-pilot gives me the support to deliver on work items that I can respond to quickly... For me its about 'being in the moment' for my people, networked conversations or issues in front of me" – Domain Lead (L1)

Findings overview

The productivity assessment survey showed that the majority of participants found 'significantly better' or 'somewhat better' quality of work with M365 Copilot.

Clearer messaging, faster iteration and better framing of ideas were citied as the reasons behind the quality improvements.

Some users found limitations with the accuracy of outputs and the importance of reviewing all content.

Half of participants reported that Copilot enabled them to focus more on strategic thinking, showing Copilots ability to shift peoples focus to more high value tasks.

Key use cases

Meetings: Allows for people to stay present and actively engaged in their meetings, reducing the distraction of taking notes.

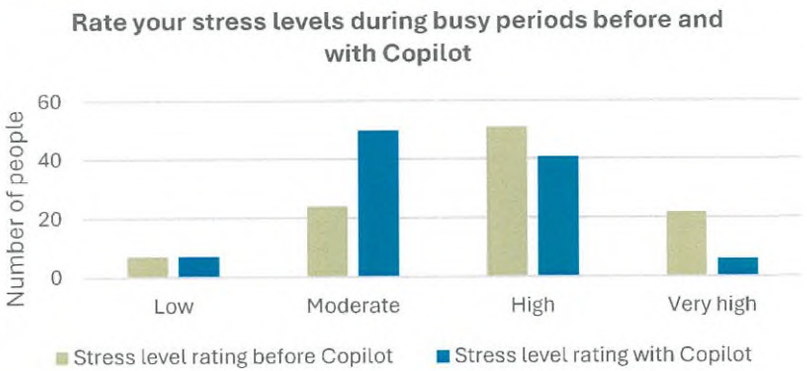
Structuring and outlining documents: Refining documents and emails to improve clarity, tone, and conciseness.

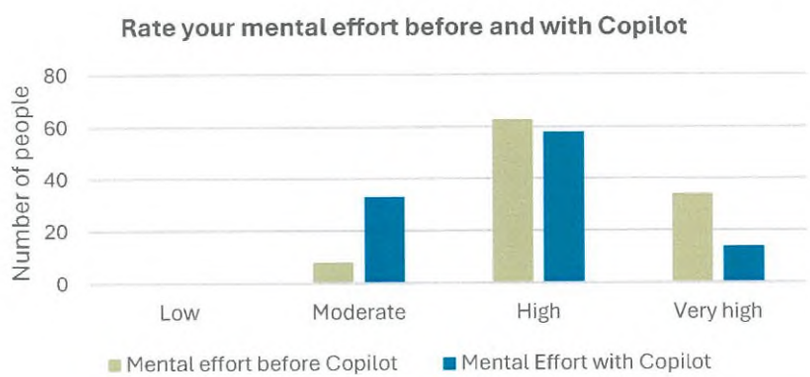
Writing clarity: Rewording complex or technical content or ideas into simpler and more concise language.

Proofreading and accuracy checks: Comparing documents for inaccuracies and differences between them.

Source: August 2025 Survey

Stress and Mental Effort





Findings overview

The survey results show that Copilot has a meaningful impact on how cognitively demanding and stressful participants perceive their work to be. Mental effort dropped from an average of 8.08 to 7.12, stress during busy periods decreased from 6.98 to 6.13 and the proportion of participants feeling extreme stress fell from 21% to 6% with Copilot.

These shifts suggest that Copilot is helping our people work smarter not harder - managing their workload more effectively, reducing cognitive load required to complete their work and alleviating pressure during busy periods. This shows Copilot is solely about efficiency, but also about creating a more sustainable, engaging work environment.

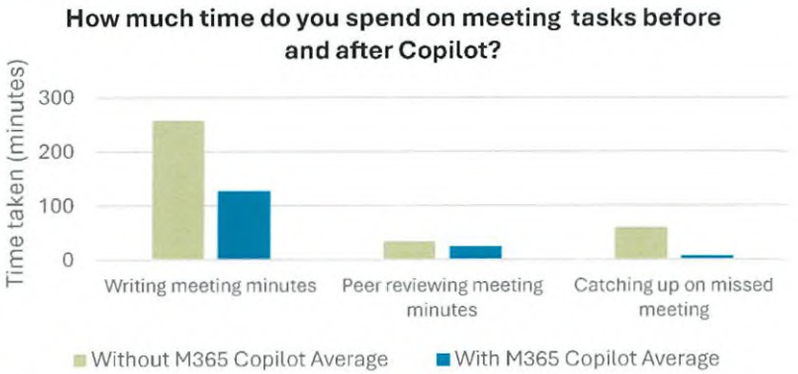
What our people told us

"Although the information copilot gives us often needs to be refined and reviewed, it can setup the framework for a piece of work or document, which can take a lot of the stress away..." – Domain Specialist (L2)

"My flexitime balance has been more manageable. My stress levels have reduced. - The time saved through copilot, has enabled me to both take on additional work..." – Technical Lead

Source: Copilot for Microsoft 365 - Project Closeout Report & August 2025 Survey

Appendix: Meetings administration



What our people told us

"Having Copilot provide a rough first draft of the minutes has allowed me to focus more on in-meeting admin tasks and on the conversation from attendees." – Business Support (L2)

"As a leader, co-pilot gives me the support to deliver on work items that I can respond to quickly... For me its about 'being in the moment' for my people, networked conversations or issues in front of me" – Domain Lead (L1)

"I will sometimes use co-pilot to help when I'm scheduling multiple meetings looking for suitable availability of rooms around the organisation. This saves a huge amount of time when booking" – Information Specialist (L2)

Findings overview

A survey of our initial pilot was conducted June 2025 across all 20 users who held licenses at the time. The survey found significant benefits for meeting chairs, support people and participants.

Meeting chairs and support people can use Copilot during meetings and for producing minutes.

Participants were able to access meeting summaries soon after a meeting ends and quickly see action points.

Key use cases

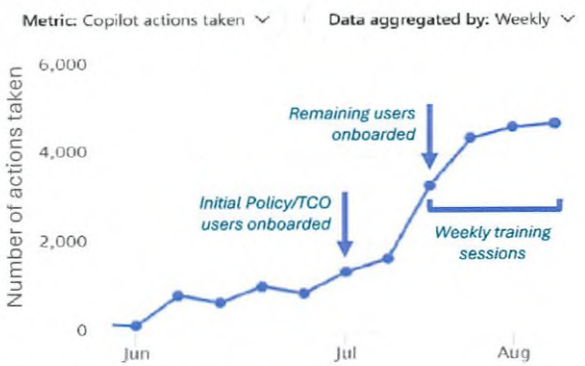
Drafting meeting minutes: Skeletons of draft meeting minutes can be created from transcriptions, recordings or meeting summaries.

Meeting summary : Meeting notes are automatically generated when recording or transcribing. These notes are available after the meeting and include key points, decisions and action items.

In-meeting support: During live meetings, Copilot enables users to catch up on missed content by summarising earlier discussions in real time. It also allows participants to clarify or revisit what was said.

Source: Copilot for Microsoft 365 - Project Closeout Report

Pilot rollout and adoption



Training and Support approach

Compulsory training: Prior to licence allocation, all participants completed the existing Copilot Fundamentals course, and agreed to IR's 'Dos and Don'ts'—defining acceptable use of M365 Copilot.

Weekly training session: Four 45-minute sessions led by E^2 covered Teams, M365 Copilot Chat, Word, Outlook, and Agents. Recordings were shared with all participants. Average usefulness rating: 6.7/10.

Pilot Teams site: All participants were added to a Teams site where key updates were posted and provided a space for questions and sharing.

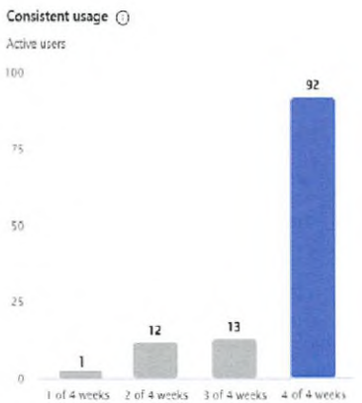
Peer-to-peer support: Informal peer learning was common. Some teams choose to create their own chat space.

Findings overview

IR has a wide range of AI comfort and skill which makes providing relevant training for all participants challenging.

Our training focused on the basics of prompting (Copilot Fundamentals) and on how to use Copilot in the apps found to provide the most value.

74% of respondents use Copilot six or more times per week, showing it has quickly become part of daily workflows. Usage was found to increase the longer people had access to Copilot and for most participants, was consistent over the past 4 week period.



Learnings for the wider rollout

- Mandatory training should be retained but needs to be short and focused on high value, widely used features like Chat and Teams.
- Flexible optional learning for less commonly used tools like Excel and Word should be available in various formats, including self-paced options.
- The most impactful learning came from social learning between peers and when it was role-specific.
- Establish a network of Copilot Champions - including AI Community members and pilot participants – to provide role specific support and help teams to embed Copilot into their workflows.

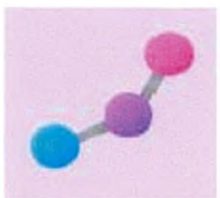
Appendix 3

Microsoft's pre-built agents

Inland Revenue
Te Tari Taake

Pre-built Microsoft Agents

Microsoft has introduced pre-built Microsoft 365 agents. These agents are designed to handle complex, multi-step tasks with a high degree of reasoning and contextual awareness.



Unlike standard Copilot interactions, which typically produce concise responses, these agents operate through deep, task-focused workflows. When prompted, they can:

- Ask clarifying questions to refine intent
- Leverage internal organisational data, Microsoft 365 infrastructure, and external web content
- Deliver comprehensive outputs—often spanning multiple pages and drawing from numerous sources

While response times are slower than typical Copilot interactions, the depth and quality of the results reflect the agents' ability to navigate layered tasks, synthesise information, and take action.

The key difference in the architecture is at the Orchestration layer (which M365 Copilot also has), this is where decisions on what tools and data sources are to be used, in what order and how to respond to the prompt based on this information. The orchestrator for these agents dynamically decides when to use deep reasoning, this allows for the agents to handle more complexity, consider broader scope than the direct task at hand and use advanced thinking.

Appendix 4

Additional materials

Inland Revenue
Te Tari Taake

Appendix: Related material

- Microsoft 365 preliminary risk assessment: Risk Assessment for Microsoft 365 Copilot
- Microsoft 356 CoPilot Legal Risk Assessment
- Copilot Fundamentals Course in Ātea
- Copilot Studio (tool for creating agents) preliminary risk assessment: Risk Assessment for Copilot Studio
- Endorsement paper for Enterprise Services Technology agent: DD-1149 Copilot Studio Enterprise Services Technology Agent May 2025
- Microsoft 365 Copilot at ACC: Our journey and insights: M365 Copilot at ACC - Journey and Insights - March 2023
- Australian Government trial of Microsoft 365 Copilot: Full report – productivity | digital.gov.au
- Microsoft 365 Trial Midway Report: Information Governance
- Microsoft 365 Trial Midway Report #2: Information Governance
- Copilot Chat benefits report: Benefits of Copilot for IR
- Copilot for Microsoft 365 - Project Closeout Report

Item 10

Inland Revenue
Te Tari Taake

Conversation Summarisation Proof of Concept

Outcomes and Recommendations

3 March 2025

Alana Macandrew and Amanda Gray

Version 1.1

Updated 3 March 2025: Updated with confirmed cost and FTE
(validated by Finance team)

Contents

Conversation Summarisation 3
Overview 3
Executive Summary 3
Recommendation 5
Approach 7
Staff surveys 7
Analysis of queue performance 10
Summary of Queue Performance 10
Queue level detail 13
Account Maintenance queue (FFL1) 13
General Individuals queue (FFL2) 15
General Business queue (GB) 17
Tax Preparers queue (GB) 19
GST queue (GST) 21
Employers queue (EMP) 23
Working for Families queue (FAM) 25
Student Loans (SL) 27
KiwiSaver (KS) 30
Other queues 31
Other notable information 32
Appendix 34
Agent Profiles 34
AI Working Group Documentation 34

Conversation Summarisation

Overview

In 2024, IR implemented Genesys Cloud as our voice channel technology, moving to cloud based, evergreen services and away from on-premise solutions.

The project team, who implemented Genesys Cloud, Pou Here Tangata, had Agent Assist Conversation Summarisation as an enhancement opportunity to be reviewed. Additionally, a report commissioned by Datacom recommended that this functionality would enhance productivity and yield efficiency savings for staff.

> *Conversation Summarisation is a pay-as-you-use function of Genesys Cloud that automatically prepares a summary of the conversation between a CSO and a customer. Our CSOs can use this information in START as a record of the conversation after first reviewing and editing the summary for accuracy and relevant information.*

The Voice Channel Management (VCM) team, along with the Individuals customer segment, ran a pilot to validate these predicted efficiency savings, aligning with CCS priorities of effectiveness and efficiency.

The proof of concept (PoC) was run with 30 CSOs from across Individuals, from 1 October 2024 to 29 November 2024 with the objectives of:

- Testing the functionality of Conversation Summarisation in IR's setting
- Reducing after call work (ACW) time for the CSOs testing the functionality, as they were not needing to write notes in full.

After call work has been used as a measure for the effectiveness of Conversation Summarisation, as the automated summarisation of notes should save CSOs time in creating notes from scratch.
ACW, also known as post-call wrap up or wrap time, is tasks that CSOs perform after speaking with a customer. These tasks include:

- Lodging notes in START about the reason for the customers call, and the outcome of the call, including any relevant details such as payment amounts, and other information.
- Updating customer details in START
- Taking actions on the account that were agreed with the customer

Executive Summary

As an organisation, we're continuing to look for opportunities to support our people to achieve efficiency gains and make use of the technology we have available to us.

Working with the Artificial Intelligence (AI) Working Group, we've interrogated the workings of Genesys Cloud AI and this sits within our structure/organisation of Genesys, with no data leaving our security boundaries.

A PoC was run to test AI-powered Conversation Summarisation within Genesys Cloud, designed to capture and summarise key information from a conversation held between IR staff (CSOs) and customers (including where third parties are used – like Language Assistance).

The participants in the PoC were CSOs with a range of tenure, skills and experience from across the Individuals segment to test the ease of use, accuracy and helpfulness of the tool. Through testing and feedback throughout the PoC we refined our IR dictionary (which holds terminology specific to IR so Conversation Summarisation can recognise what this is) and saw improvements in the accuracy, relevance and understanding of the toolset.

CSOs have reported time savings through improved note taking, with them reviewing and adding to notes and summaries of the conversation rather than writing them out in full. They have also highlighted the ability of Conversation Summarisation to capture information they may have overlooked.

The use of Conversation Summarisation throughout the PoC has demonstrated an improvement in efficiency in the way we capture our notes from voice interactions with our customers, has been easy to use, and helpful for CSOs.

Data analysed from the PoC found a reduction of around 1.47 minutes per call in ACW time, or around 35,000 to 37,000 hours in ACW time annually, or 30 and 32 FTE, that could be spent answering more voice calls from customers, replying to more web messages, or using this time to focus on compliance.

There would be an investment of 9(2)(b)(ii) per CSO using Conversation Summarisation required, however this is outweighed by the potential ACW savings. Minimal investment is required in training to support these enhanced features to be functional across all CSOs, this is around 30 minutes of upskill training per person.

Findings:

- **Increased Efficiency:** The data indicates a 17% efficiency increase in after call work time among CSOs using Conversation Summarisation compared to those same agents on average, in the 9 months preceding the PoC. Irrespective of tenure or skill profile all CSOs experienced some level of efficiency gain.
- **Quality:** Team Leads undertook live and retrospective quality reviews of their participating CSOs during the PoC, as they usually would. They reported there was no degradation in the notes being entered into customer accounts in START compared to those prior to the PoC. They also noted that their team members had found the summarisation a useful prompt when they had forgotten to add parts of the call to their notes.
- **Staff Sentiment:** Sentiment from CSOs is that Conversation Summarisation has assisted with their calls and note taking. All but one CSO in the PoC reported, they would be supportive of the decision to roll out this enhancement to all CSOs working on voice contacts. Some CSOs have reported that they feel "heartbroken to lose it" now the PoC is complete.
- **Training:** The time spent to train CSOs to use Conversation Summarisation was around 30 minutes in total – done in a meeting explaining what it is, and how it works. Given the minimal time investment to train CSOs how to use Conversation Summarisation, and the benefit it has provided CSOs, the training time is outweighed.
- **Language Assistance:** This functionality was not specifically trialled, but CSOs in the PoC reported to have accurate summaries of calls where Language Assistance had been used and were happy with the results of the Conversation Summarisation.

Challenges:

- **Engagement:** While the toolset has proven benefits, it is important to acknowledge the initial reluctance from some CSOs to utilise this tool. This primarily stems from being used to writing notes, and having been trained to write notes, in specific ways and Genesys Cloud presenting conversation summaries in different ways, as well as needing to add detailed information into summaries. For the majority of CSOs, the reluctance to utilise the toolset was managed through change management conversations with their Leads and via buzz meeting engagements. For wider rollout, change management needs to be considered. The engagement from Leads was also varied; some leaders were very engaged with the PoC and sharing their learnings with the wider group proved very useful, this reflected on how engaged their team members were with providing feedback throughout the PoC. Where Leads were not as engaged, their team members completed less surveys and reported using Conversation Summarisation less.

- **Limited numbers in PoC:** Due to limitations in the number of CSOs we were able to complete a PoC with, business segments and the Families segment were not included in the PoC. There are some similarities between calls taken in business queues and calls taken in Small, Medium Enterprises and Micro Business, as well as Significant Enterprises. It is assumed that benefits would be similar to those seen in the GST, General Business and Compliance queues. It is recommended that a pilot be completed with these business segments to determine efficiency prior to committing to a roll out of Conversation Summarisation to all CSOs in the business segments.
  Similarly, Child Support calls can be likened to Working for Families calls in the level of detail that is required to be captured and recorded in START. A pilot should be conducted to better understand efficiency opportunities, as the data from this PoC indicates that there was an increase in after call work time on the Working for Families queue. Without further investigation, the data indicates it would be better not using Conversation Summarisation in the Families Segment.

- **Transfers:** Transfers are not currently supported for Conversation Summarisation. When a call has been transferred, the summarisation from that prior conversation is included in the final summarisation presented for the call. This resulted in the summarisation being unusable for the call.

- **Multiple Customers:** Where multiple customers or entities are discussed within a call, the Conversation Summarisation is for the entire call. The Summarisation does not always show that multiple customers or entities were discussed, this was particularly prevalent with calls received on the Tax Preparers queue.

- **Style of Conversation Summarisation:** The Conversation Summarisation is written in a third person style and refers to CSOs as "the agent" and customers as "the customer". This meant at times, particularly in calls with Tax Preparers, or agents, notes could be confusing. This is on the roadmap for Genesys for the first half of 2025 to provide the ability for input as to how customers or agents should be referred in the summaries.

Recommendation

It is recommended to roll out and implement Conversation Summarisation across CCSI-Individuals Segment CSOs who handle customer voice contacts and pilot with other segments prior to roll out to them.

- We recommend you implement Conversation Summarisation for CCSI-Individuals CSOs who handle voice contacts.
- We recommend separate pilots are undertaken with Families CSOs and Business Segments CSOs prior to decisions about implementation in their segments.
- We recommend an investment in licencing costs to utilise Agent Copilot within Genesys to see a benefit of approximately 30-32 FTE per annum (before costs are factored in), accounting for the additional spend there is a benefit of approximately $683,000-789,000 per annum (once additional licencing costs have been factored in).

The PoC was run on the current service offering from Genesys called Agent Assist, from May 2025 this product is no longer available. The new product where Conversation Summarisation is available will be called Agent Copilot. This is charged on a per CSO cost of 9(2)(b)(ii) per month, around 9(2)(b)(ii) per annum, in addition to standard user licencing costs. This cost has been reflected prior to savings being calculated above.

Copilot offers additional features to those assessed in the PoC that will be included if rolled out to the wider business, such as the ability to surface knowledge, further improvements in the summarisation features, and summarised reason for calling and

resolution. Knowledge surfacing will be undertaken as a PoC early in 2025, which will not have an additional licencing cost but likely to incur set up costs for integration

The data strongly supports the adoption of the new toolset to enhance our operational efficiency. By addressing the initial resistance through change management and support, we can ensure a smooth transition and maximise the benefits of Conversation Summarisation.

Change Management recommendations to support the roll out of Conversation Summarisation

- **Support:** Implement a support program to ensure all CSOs across Individuals are comfortable and proficient with using Conversation Summarisation. This should include the initial training, hands-on sessions, tutorials, and ongoing support, which will be provided by key users, L2s and the VCM team, as with BAU voice support practices. Training should include known limitations and how to spot these, along with the importance of the CSO remaining responsible for the review and editing of all Conversation Summarisation notes prior to entering them into START.
- **Feedback Mechanism:** Establish a feedback loop where CSOs can share their experiences and provide suggestions for improvement, such as new words to be added to the IR Dictionary. This will help the summarisations to continue to be accurate and reflect IR's language.
- **Gradual Rollout:** A phased implementation approach to allow CSOs across Individuals to adapt gradually and in groups. Teams could be provided the access to Conversation Summarisation progressively and work closely with VCM and other supporting areas during the initial roll out.
- **Team Lead engagement:** Engagement from CSO leads played a critical role in how CSOs felt about Conversation Summarisation. For CSOs who had leads that regularly engaged, provided feedback, and attended buzz meetings, surveys showed they completed an average of 100 surveys more per CSO than those CSOs whose leads did not appear to have active engagement throughout the PoC.
- **Highlight Benefits:** Communicate the benefits clearly to all CSOs, emphasising the positive impact on their workload and overall efficiency. Sharing success stories from the PoC could help with acceptance and change management.
- **After call work time:** If this is to be rolled out further, efficiency would be seen if ACW time was used strictly for ACW activities. It was observed throughout the PoC that CSOs frequently used ACW time to complete other tasks and activities. Work should be undertaken to ensure ACW is being used for ACW rather than other tasks.
- **Te Mātāwai notes requirements:** Should the PoC implementation rolled out, a review of Te Mātāwai notes requirements should be undertaken to reflect changes with how notes are summarised, as well as decision making abilities of CSOs when reviewing the notes and making decisions on what should/should not be captured.

Approach

The PoC was run from 1 October 2024 to 29 November 2024 with 30 CSOs from across five teams within the Individuals customer segment.

CSOs were selected with a range of skills, tenure and capabilities to ensure there was a representative group of CSOs who would usually undertake voice tasks. Some CSOs were added to the PoC part-way through as others left due to gaining other roles, training, or resignation.

Engagement and preparation was undertaken jointly with the VCM team, Individuals Group Lead, and Individuals Management Support to prepare for the PoC and inform participants of the toolset in advance of the PoC commencing. Meetings were held with participants and their leaders with support from VCM, One NZ and Genesys throughout the PoC. Quantitative and qualitative data has informed the recommendations and outcomes contained in this report.

This approach ensured thorough preparation, continuous engagement, and effective feedback mechanisms to support the proof of concept.

Staff surveys

Survey results have been included for CSOs whose data has been included in the reporting.

CSOs were surveyed to gain insight that quantitative data alone could not tell us, such as the accuracy of the summarisation, the ease of use, how helpful they found the Conversation Summarisation and how the CSOs were feeling about it.

Accuracy

CSOs surveyed were asked after each call, if the Conversation Summarisation was used, and how accurate it was. Of the surveys completed, 78% of calls used the Conversation Summarisation. Only 54 calls (of 497 where the summarisation was not used), were not used because the inaccuracy was too great to make it worthwhile using. Other reasons the Conversation Summarisation was not used included transferred calls – which was to be expected due to the non-support of transfers in Conversation Summarisation.

In week one, 83% of calls where the summarisation was used were rated average to perfect, this improved to 100% of summaries rated average or above by the end of the PoC. This gives a good indication that CSOs were able to use the summarisations presented to them, making minor changes to update details and continuing to review for accuracy and intent. The number of summarisations rated as perfect remained consistent across the PoC, with an average across the PoC of 6% being rated as perfect, no changes required.

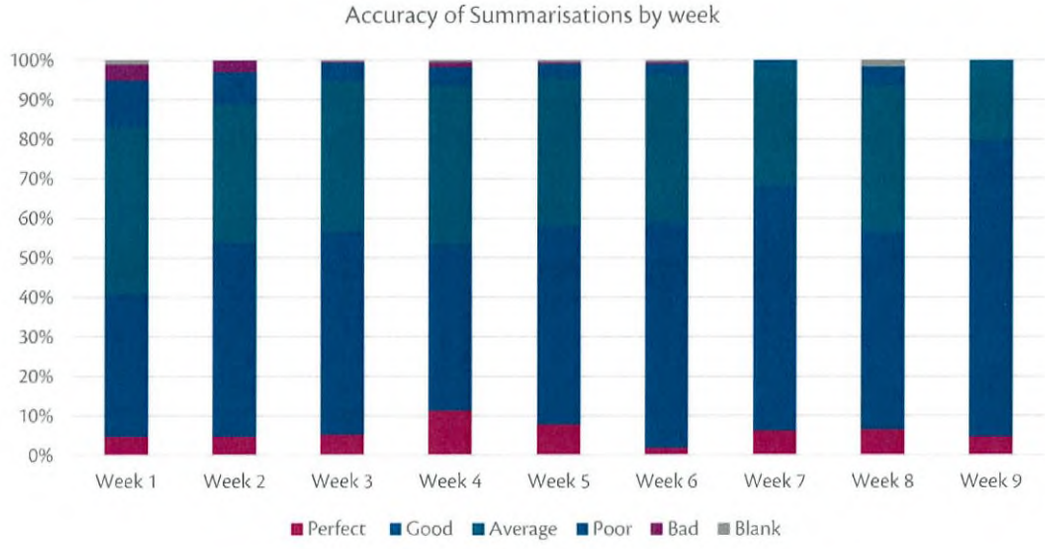
Since CSOs are responsible for the content of the notes, they always remain in charge of adding notes into START and on customer accounts. Therefore, having most of the summaries rated as average, needing a few changes, or above is a great result

CSOs also noticed an improvement in accuracy as the PoC progressed, most significantly from week 3. This was consistent across all CSOs, even those that joined the PoC part-way through – the reported accuracy between week 2 and 3 increased, as did their confidence and ease with the summarisations.

All but one CSO reported seeing an increase in accuracy across the PoC.

"It was great fun being a part of it, whilst my voice queues became limited towards the end it was good to see the progress in the outputs the assist gave. There was definitely a clear difference between beginning and end and if its fine tuned further it might be a good fit. It would be down to cost-ratio for effectiveness". - CSO # 3.

"When it got it right and the notes didn't need much editing, it was much faster to wrap up my calls". – CSO # 14



Value

Each week the CSOs were asked if they found Conversation Summarisation helpful. On average 86% of CSOs found it helpful, 13% did not and 1% had left this question blank. Of the PoC CSOs, 1 CSO across all weeks did not find the PoC information helpful, this CSO noted that they would like more structured notes and templates for notes rather than a free-text type style. End of PoC feedback from CSOs supported that Conversation Summarisation was helpful.

"I liked that is picked up on important information and put it in bullet points". – CSO #10

"It gives a good general overview of what was discussed, which is helpful at the end of the conversation if you forget something that was mentioned". – CSO #13

"I like how it provides a summarisation of the conversation, how it picks up some things a customer says that I may not have heard properly, and I like how (for me) it serves as reminder of certain aspect in the call that I may have forgotten but Agent Assist picks it up in the summary". – CSO #19

"Helped reduce the time I spent on aftercall work". – CSO # 20

"It is good and helpful". – CSO # 16

"Given there are a few improvements. I think it is helpful because it does give you a guide of your phone call, where some things you may have forgotten the assist can pick this up for you'. – CSO # 7

Ease of use

We asked CSOs how easy the Conversation Summarisation was to use. 97% of said it was easy to use, 1 CSO reported "no, it was not easy to use" once, in week 4, and in two other instances (2.4%) the question was left blank.

The overarching feedback from CSOs has been it was an easy tool to use, and the majority of CSOs supported its wider rollout.

> *"Summarising the conversation, where we had to just copy and edit few things and reduce time in typing the notes". - CSO # 12*

> *"At best, saved time by writing notes for me.*
> *At worst, provided a nice summary of the call which would assist with my own note taking.*
> *Win-win." - CSO # 5*

> *"I liked that it provided my core discussions which saved a lot of time with ACW and a great foundation for each call note". - CSO # 11*

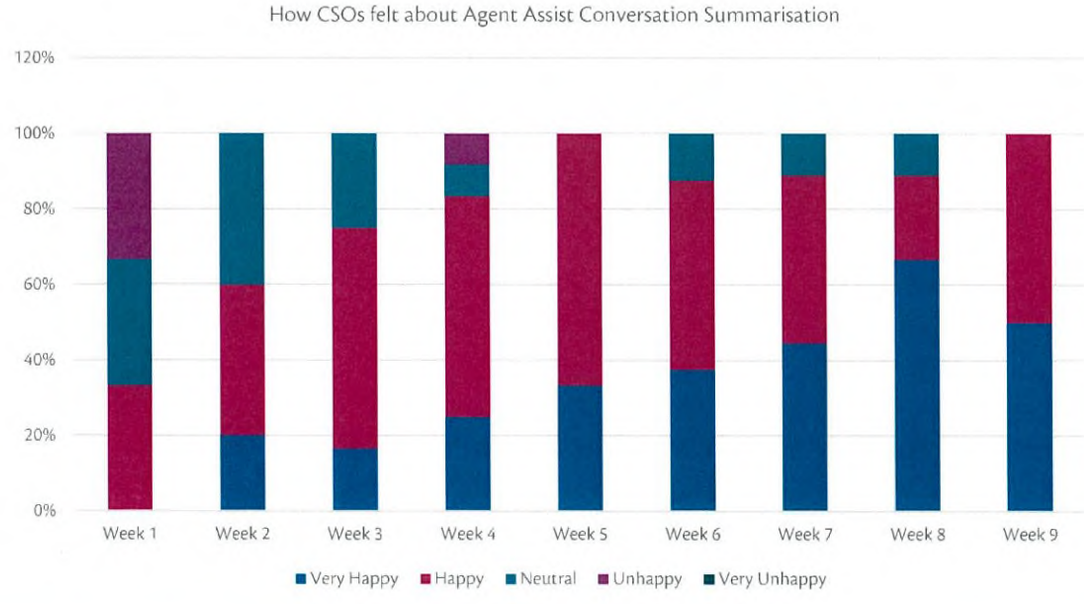
How CSOs felt about Conversation Summarisation

CSOs reported feeling happy or very happy for the majority of the PoC. This improved over time, in week one, 33% of CSOs felt happy or very happy, by week 9, 100% felt happy or very happy.

> *"I definitely believe this would be helpful in note taking. I know for others it has improved their aftercall wrap time. It's a good 'assistant' in making notes and especially helpful for straightforward calls. With further development it definitely has potential to be even more helpful". - CSO # 19*

> *"I found the notes to be almost perfect on general individual lines, and as the progression through learning on FAM line the notes had been amazing especially when we had lengthy calls with customers that required large notes and additions this made it a lot easier having the summary to then add our additional calculations and other information to complete". – CSO # 11*

No CSOs rated feeling very unhappy throughout the PoC period.



Analysis of queue performance

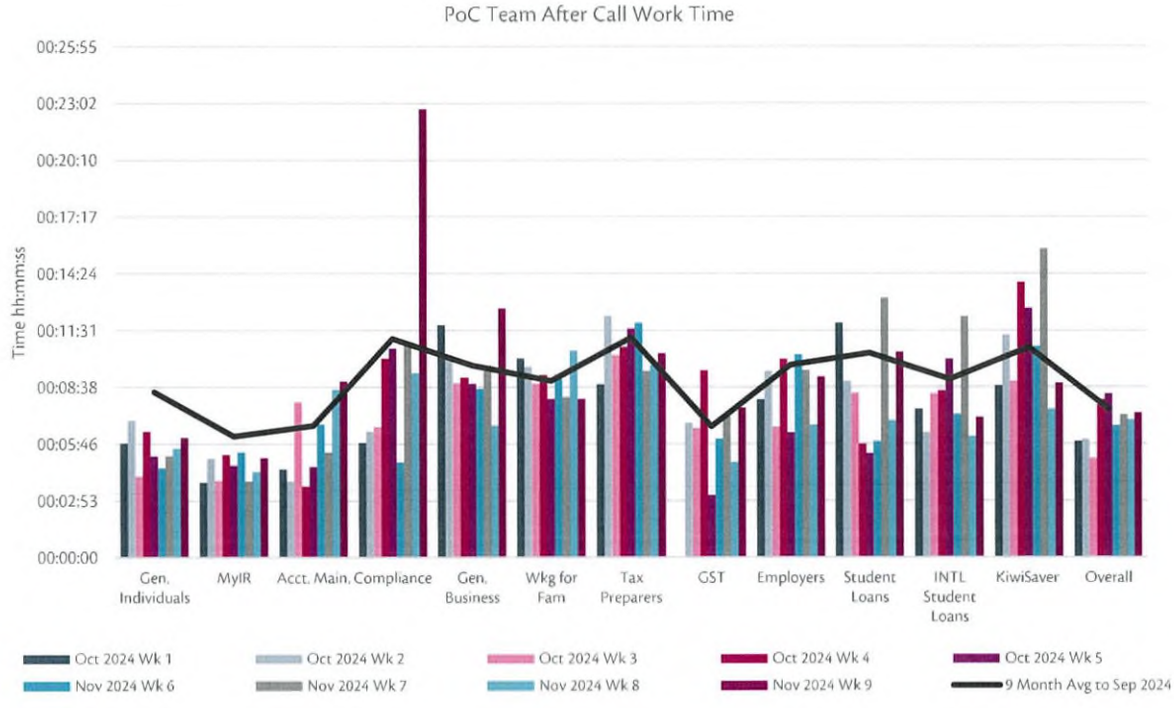
To perform analysis and comparison, a baseline has been created using an average of each CSO in the PoC's ACW time from January to September 2024 for each queue. This is to determine if there has been an improvement in ACW for the same CSOs.

A control group has been used, which includes all other CSOs (who are not Level 2 or other roles taking voice calls) to determine the performance of the CSO group in the PoC compared to other CSOs to ensure there are no outliers in trends.

Some participants in the PoC obtained other roles, resigned or undertook training during the PoC. Their data has been removed as to not falsely skew data or manufacture proportions of CSOs in training compared to others not in training.

Summary of Queue Performance

PoC team ACW time by queue



Queue Summary Overall Pre PoC and During PoC Comparison

After Call Work Time hh:mm:ss	PoC Group Pre	PoC Group During PoC	Difference	Control Group
General Individuals	00:08:24	00:05:25	-00:02:59	00:07:33
My IR	00:06:09	00:04:33	-00:01:36	00:05:39
Account Maintenance	00:06:41	00:05:33	-00:01:08	00:06:13
Compliance	00:11:06	00:09:08	-00:01:58	00:10:17
General Business	00:09:45	00:09:45	-00:00:00	00:08:44
Working for Families	00:08:57	00:09:08	-00:00:11	00:10:44
Tax Preparers	00:11:09	00:10:39	-00:00:30	00:09:34
Credit card failure	00:06:39	00:06:48	-00:00:09	00:07:49
GST	00:09:45	00:08:20	-00:01:25	00:09:34
Employers	00:10:22	00:08:54	-00:01:28	00:09:44
Student Loans	00:09:02	00:07:53	00:01:09	00:11:59
International Student Loans	00:10:39	00:10:19	00:00:20	00:13:08
KiwiSaver	00:07:31	00:06:27	00:01:04	00:08:18

Queue Summary PoC Average After Call work time results week by week

Queue	9 Month Avg to Sep 2024	Oct 2024 Wk 1	Oct 2024 Wk 2	Oct 2024 Wk 3	Oct 2024 Wk 4	Oct 2024 Wk 5	Nov 2024 Wk 6	Nov 2024 Wk 7	Nov 2024 Wk 8	Nov 2024 Wk 9
General Individuals	**00:08:24**	00:05:49	00:06:56	00:04:08	00:06:26	00:05:09	00:04:33	00:05:10	00:05:33	00:06:05
MyIR	**00:06:09**	00:03:48	00:05:02	00:03:55	00:05:13	00:04:40	00:05:21	00:03:53	00:04:22	00:05:03
Account Maintenance	**00:06:41**	00:04:29	00:03:52	00:07:53	00:03:37	00:04:38	00:06:47	00:05:22	00:08:31	00:08:57
Compliance	**00:11:06**	00:05:50	00:06:24	00:06:37	00:10:06	00:10:36	00:04:51	00:10:57	00:09:21	00:22:45
General Business	**00:09:45**	00:11:48	00:10:17	00:08:52	00:09:08	00:08:49	00:08:33	00:09:45	00:06:41	00:12:39
Working for Families	**00:08:57**	00:10:06	00:09:41	00:08:50	00:09:16	00:08:02	00:09:19	00:08:09	00:10:29	00:08:04
Tax Preparers	**00:11:09**	00:08:48	00:12:14	00:10:15	00:10:41	00:11:37	00:11:55	00:09:27	00:09:45	00:10:23
GST	00:09:45	00:08:01	00:09:27	00:06:38	00:10:04	00:06:20	00:10:18	00:09:30	00:06:43	00:09:11
Employers	00:10:22	00:11:55	00:08:57	00:08:21	00:05:46	00:05:16	00:05:54	00:13:10	00:06:58	00:10:25
Student Loans	00:09:02	00:07:33	00:06:20	00:08:18	00:08:29	00:10:04	00:07:17	00:12:14	00:06:10	00:07:08
International Student Loans	00:10:39	00:08:42	00:11:18	00:08:56	00:13:58	00:12:38	00:10:43	00:15:40	00:07:32	00:08:52
KiwiSaver Individuals	00:07:31	00:05:54	00:06:00	00:05:02	00:08:01	00:08:19	00:06:41	00:07:15	00:07:00	00:07:20

All results are influenced by the CSOs taking the calls. Some CSOs have faster than average ACW, whilst others take more time than the average.

There has been a positive difference in ACW time for calls answered on queues within the pilot. The calls to these queues make up **65.3%** of **all** calls answered by Inland Revenue customer facing queues (based on last 12 months) and Conversation Summarisation would provide efficiencies to these interactions. There is the potential for an **efficiency** gain of around **35,000 to 37,000 hours per annum** based on the data obtained during the PoC for the queues below.

After Call Work Time	PoC Group Pre	Poc Group During PoC	Difference	% of all calls answered in last 12 mths
General Individuals	0:08:24	0:05:25	-0:02:59	**28.6%**
My IR	0:06:09	0:04:33	-0:01:36	**11.7%**
Account Maintenance	0:06:41	0:05:33	-0:01:08	**4.9%**
Compliance	0:11:06	0:09:08	-0:01:58	**2.7%**
General Business	0:09:45	0:09:45	0:00:00	**8.3%**
Working for Families	0:08:57	0:09:08	0:00:11	**12.6%**
Tax Preparers	0:11:09	0:10:39	-0:00:30	**4.2%**
Credit card failure	0:06:39	0:06:48	-0:00:09	**1.2%**
GST	0:09:45	0:08:20	-0:01:25	**4.1%**
Employers	0:10:22	0:08:54	-0:01:28	**3.2%**
Student Loans	0:09:02	0:07:53	-0:01:09	**1.8%**
International Student Loans	0:10:39	0:10:19	-0:00:20	**1.2%**
KiwiSaver	0:07:31	0:06:27	-0:01:04	**1.7%**

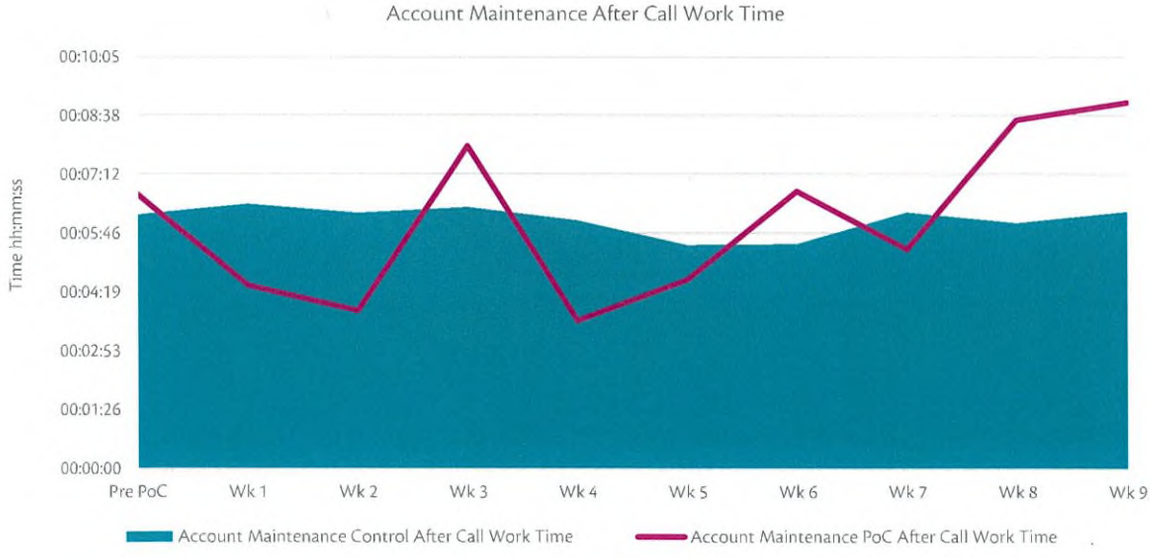
Queue level detail

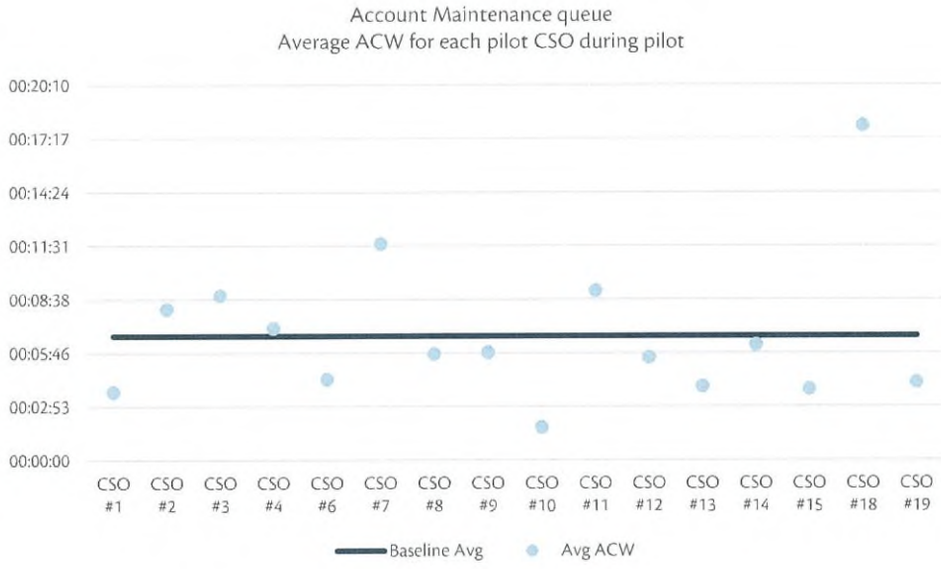
Account Maintenance queue (FFL1)

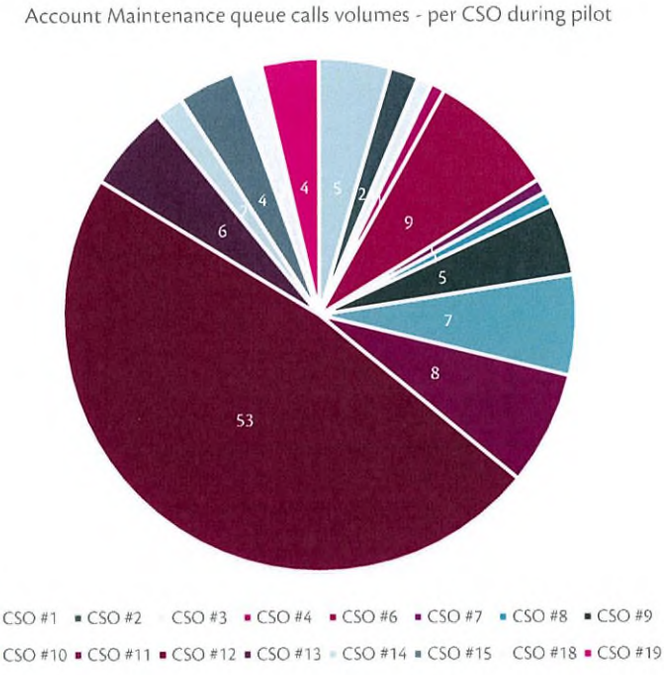
- There were 111 Account Maintenance calls answered during the PoC by 16 different CSOs.
- Feedback from CSOs was that Conversation Summarisation worked well on these simple FFL1 calls.
- One CSO took most of the calls in this queue (48% of all calls).
- On average, most weeks saw a reduction in ACW time compared to pre-PoC for the pilot group of CSOs in Account Maintenance.
- CSO #18 and CSO #11 were outliers in the data and impacted the average by +14 and +15 seconds overall.
- CSO # 11 impacted the average ACW in weeks 8 and 9 as they took an average of 54% more time that the average ACW in week 8; and were the only participant who took calls in week 9 and their ACW work for this queue is higher than average, this could due to multiple factors such as training, customer queries and support needed across these calls.
- The mix of CSOs answering calls in this queue each week greatly impacted the results as some weeks only two CSOs took calls on this queue.
- The average ACW for the PoC for this queue was 5 minutes and 33 seconds, compared to 6 minutes and 41 seconds prior to the PoC for the same group of CSOs.

Account Maintenance Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
Acct Maint PoC After Call Work Time	00:06:41	00:04:29	00:03:52	00:07:53	00:03:37	00:04:38	00:06:47	00:05:22	00:08:31	00:08:57
Acct Maint Control After Call Work Time	00:06:13	00:06:28	00:06:15	00:06:24	00:06:05	00:05:29	00:05:30	00:06:16	00:06:01	00:06:18




Account Maintenance queue
Average ACW for each pilot CSO during pilot
00:20:10
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #1
CSO #2
CSO #3
CSO #4
CSO #6
CSO #7
CSO #8
CSO #9
CSO #10
CSO #11
CSO #12
CSO #13
CSO #14
CSO #15
CSO #18
CSO #19
Baseline Avg
Avg ACW

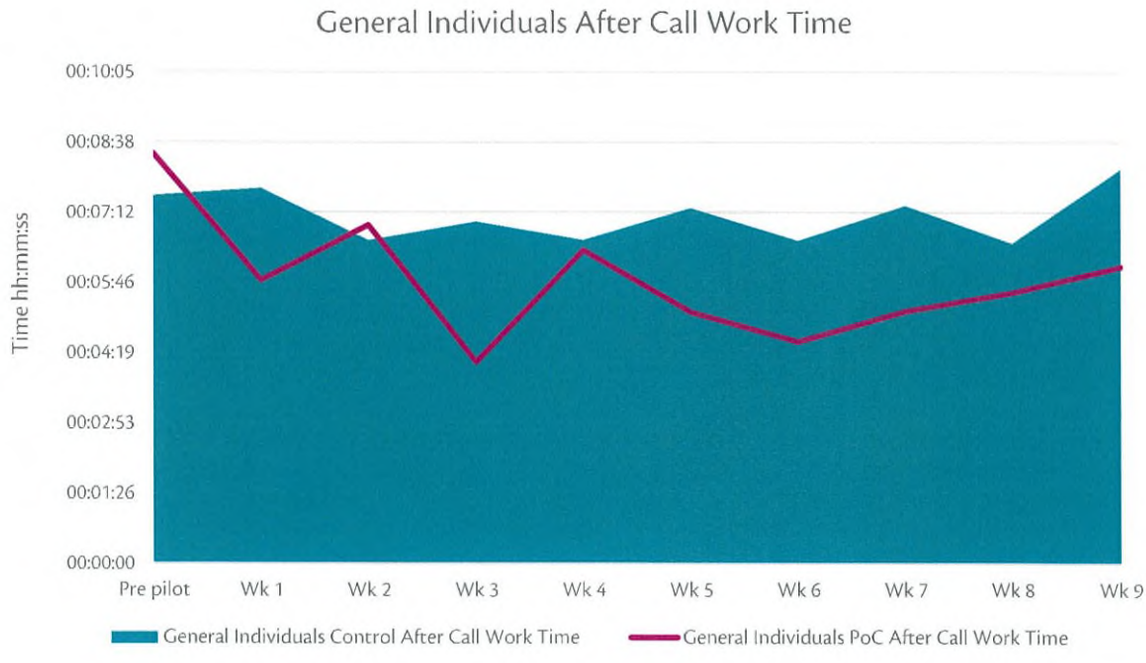

Account Maintenance queue calls volumes - per CSO during pilot
5
2
9
1
5
7
8
53
6
4
4
CSO #1
CSO #2
CSO #3
CSO #4
CSO #6
CSO #7
CSO #8
CSO #9
CSO #10
CSO #11
CSO #12
CSO #13
CSO #14
CSO #15
CSO #18
CSO #19

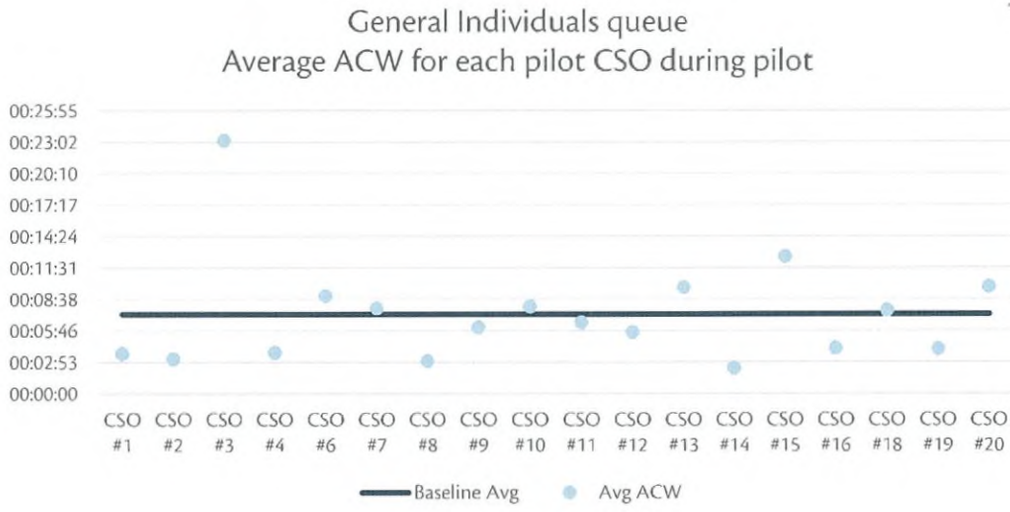
General Individuals queue (FFL2)

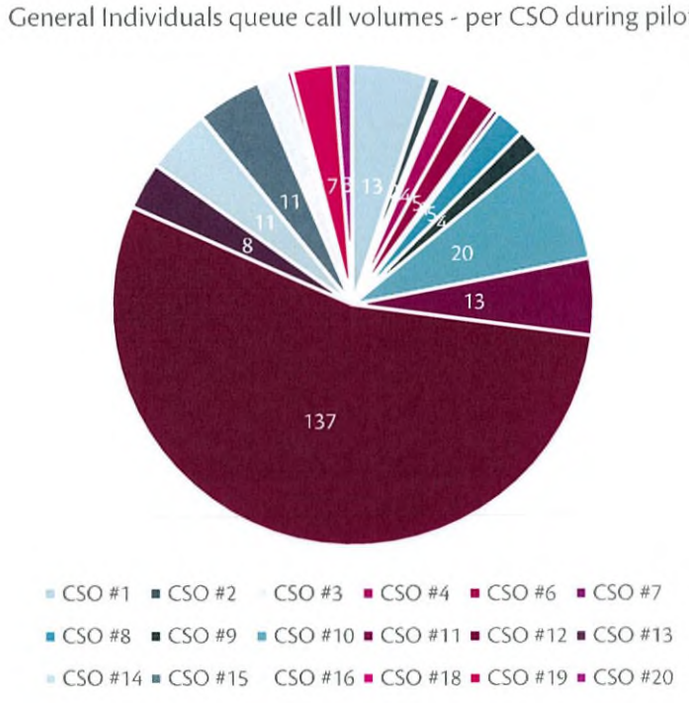
- There were 251 General Individuals calls answered during the PoC by 18 different CSOs.
- 55% of these calls were answered by CSO #12. This means CSO #12's performance contributed significantly to the overall results and their individual results greatly impacted the overall performance for the queue and added 6 seconds when weighted to the average.
- Most CSOs stated through the survey results that Conversation Summarisation worked very well for calls on this queue.
- The ACW time went from an average of 8 minutes and 24 seconds in the 9 months prior to the PoC to 5 minutes and 25 seconds during the PoC.

General Individuals Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
Gen Ind PoC After Call Work Time	00:08:24	00:05:49	00:06:56	00:04:08	00:06:26	00:05:09	00:04:33	00:05:10	00:05:33	00:06:05
Gen Ind Control After Call Work Time	00:07:33	00:07:42	00:06:38	00:07:01	00:06:38	00:07:18	00:06:38	00:07:21	00:06:34	00:08:06




General Individuals queue
Average ACW for each pilot CSO during pilot
00:25:55
00:23:02
00:20:10
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #1
CSO #2
CSO #3
CSO #4
CSO #6
CSO #7
CSO #8
CSO #9
CSO #10
CSO #11
CSO #12
CSO #13
CSO #14
CSO #15
CSO #16
CSO #18
CSO #19
CSO #20
Baseline Avg
Avg ACW

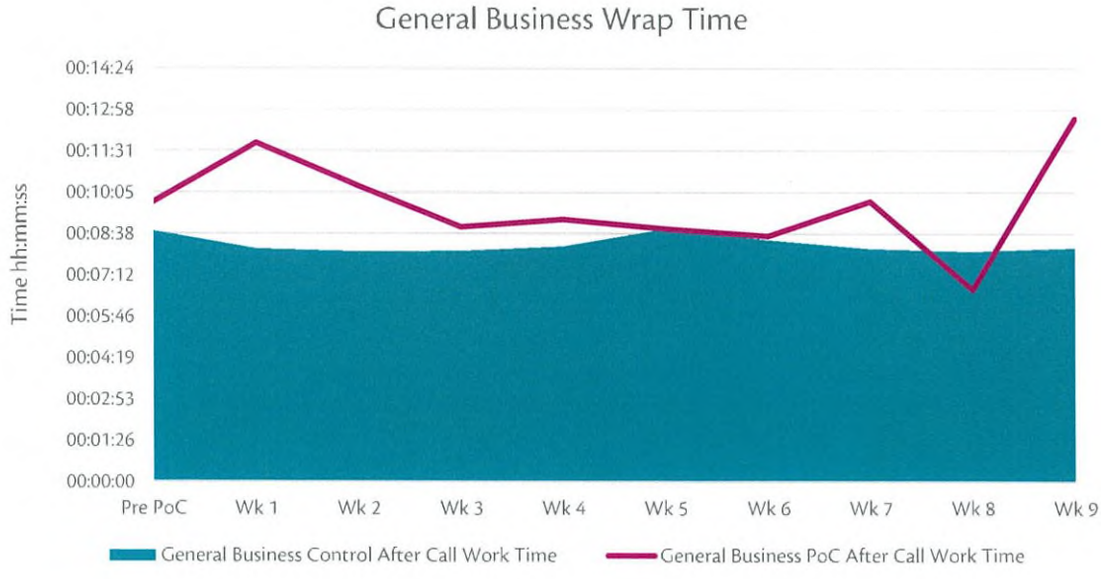

General Individuals queue call volumes - per CSO during pilot
137
20
13
11
11
8
7
13
5
5
4
CSO #1
CSO #2
CSO #3
CSO #4
CSO #6
CSO #7
CSO #8
CSO #9
CSO #10
CSO #11
CSO #12
CSO #13
CSO #14
CSO #15
CSO #16
CSO #18
CSO #19
CSO #20

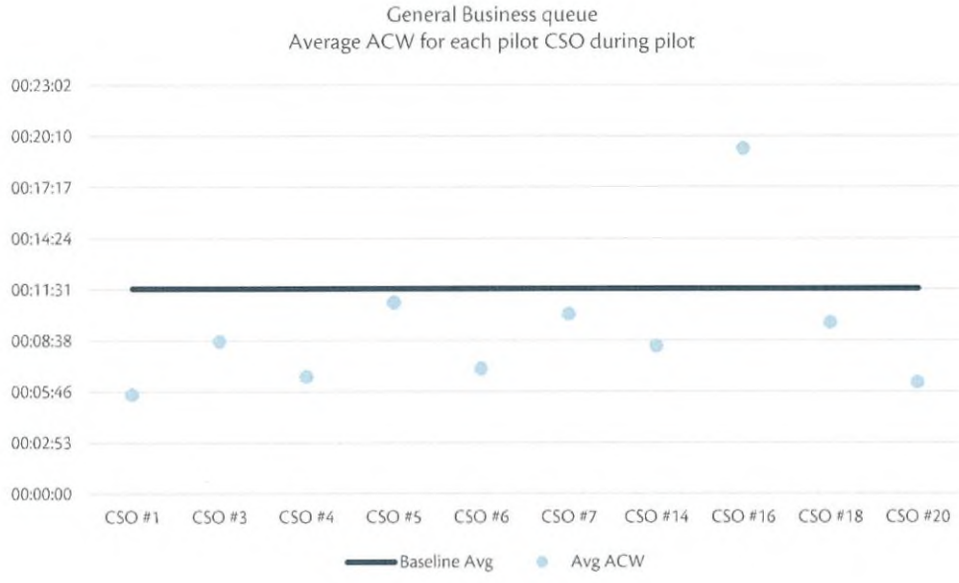
General Business queue (GB)

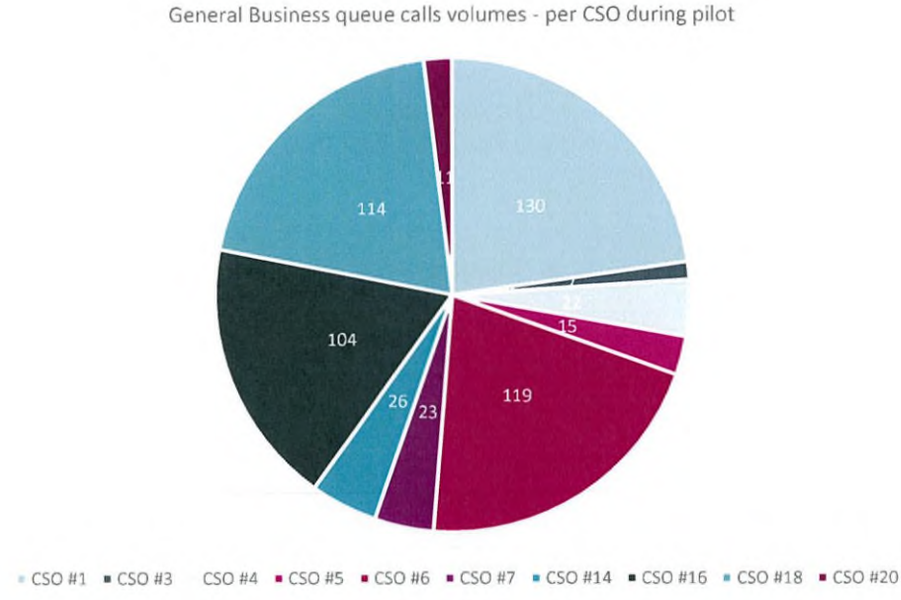
- There were 571 General Business calls answered during the PoC by 10 different CSOs.
- 82% of these calls were answered by CSO #1, CSO #6, CSO #16 and CSO #18. These CSOs performance contributed significantly to the overall results. CSO #16 added 1 minute and 46 seconds when weighted to the average. CSO #16 is reasonably new to IR with a tenure of less than 12 months so this may have contributed to this. CSO #1 average ACW time was double that of the control group in week 9 which has impacted the overall result for this queue.
- The PoC groups ACW time was 1 minute higher than that of the control group average ACW time for General Business prior to the PoC, this continued throughout the PoC but the gap between the PoC group and control group was closer together across most weeks of the PoC using Conversation Summarisation.
- Most CSOs stated through the survey results that Conversation Summarisation worked reasonably well for calls on this queue.
- The ACW time remained the same pre and during the PoC at 9 minutes and 45 seconds.
- Generally, there has been an improvement of around 1 minute across the ACW time for the General Business queue for the PoC group and data suggests there would be an overall improvement in after call work time should the PoC be implemented across all CSOs.

General Business Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
Gen Bus PoC After Call Work Time	00:09:45	00:11:48	00:10:17	00:08:52	00:09:08	00:08:49	00:08:33	00:09:45	00:06:41	00:12:39
Gen Bus Control After Call Work Time	00:08:44	00:08:07	00:08:01	00:08:03	00:08:12	00:08:48	00:08:26	00:08:07	00:08:02	00:08:10




General Business queue
Average ACW for each pilot CSO during pilot
00:23:02
00:20:10
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #1
CSO #3
CSO #4
CSO #5
CSO #6
CSO #7
CSO #14
CSO #16
CSO #18
CSO #20
Baseline Avg
Avg ACW

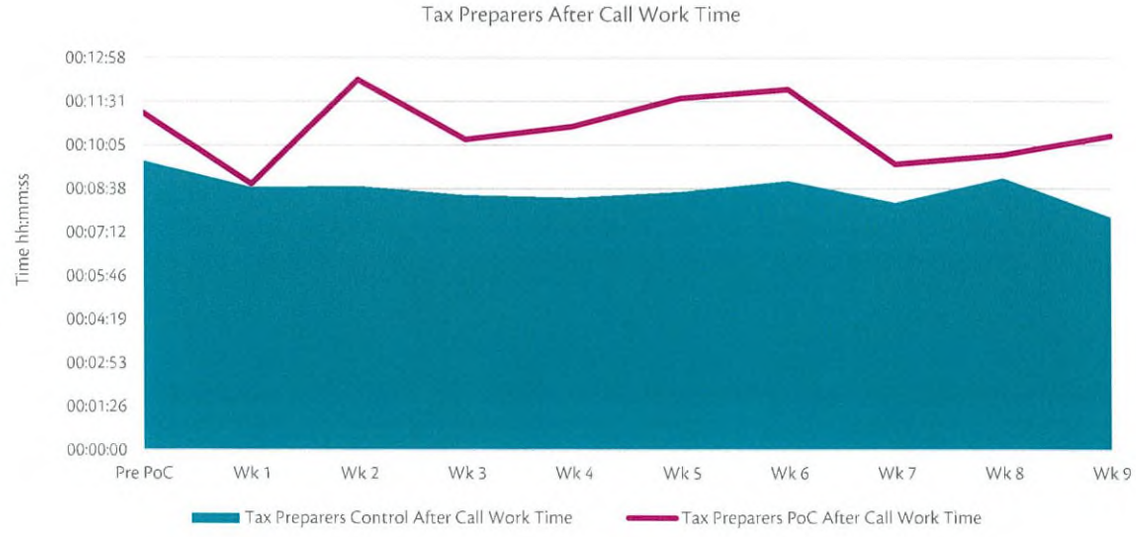

General Business queue calls volumes - per CSO during pilot
130
114
104
26
23
119
15
CSO #1
CSO #3
CSO #4
CSO #5
CSO #6
CSO #7
CSO #14
CSO #16
CSO #18
CSO #20

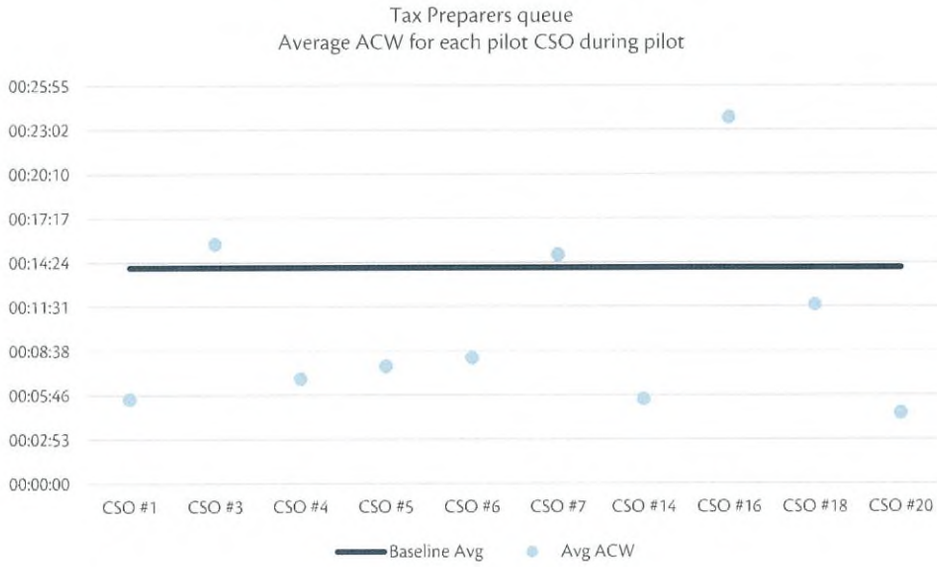
Tax Preparers queue (GB)

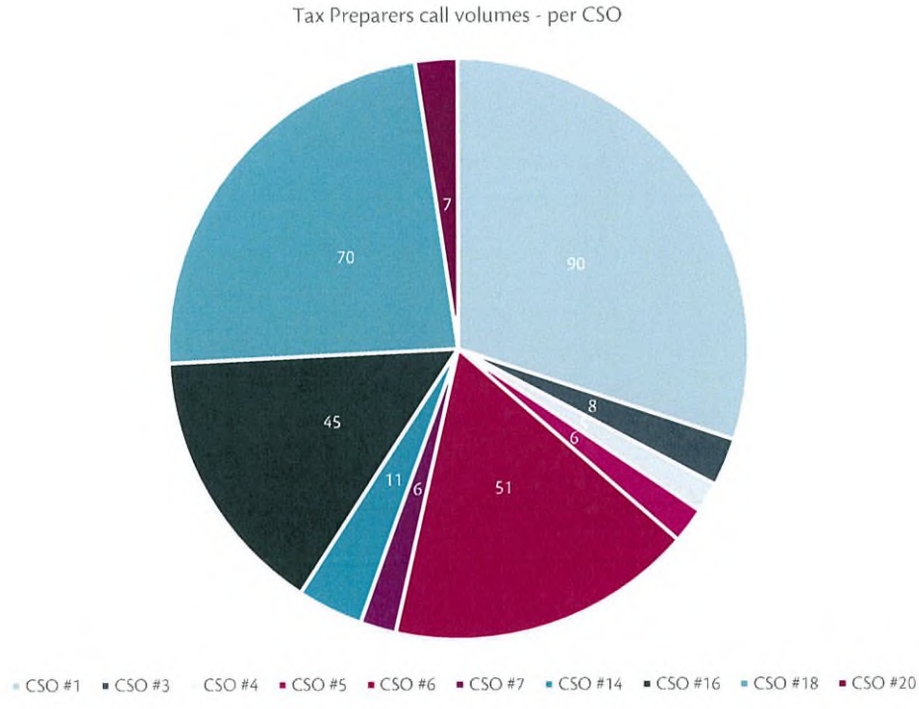
- There were 299 Tax Preparers calls answered during the PoC by 10 different CSOs.
- There are limitations with Conversation Summarisation when discussing multiple customers, in that the summarisation is for the entire call and does not always show that multiple customers have been spoken about in the one conversation.
- The summarisation providing notes in the third person tense and referring to Tax Preparers as customers was noted as a concern for CSOs in terms of clarity of notes. An enhancement is on the roadmap for Genesys to enable organisations to determine how customers should be referred to. This would improve the summarisations for the Tax Preparers queue when available.
- 86% of these calls were answered by CSO #1, CSO #6, and CSO #16 and CSO #18. These CSOs performance contributed significantly to the overall results. CSO #16 added 1 minute and 59 seconds when weighted to the average. CSO #16 is reasonably new to IR with a tenure of less than 12 months so this may have contributed to this. CSO #1 decreased the weighted average by 1 minute and 33 seconds and is also relatively new to IR with a tenure of 6 months.
- The PoC groups ACW time was higher than that of the average ACW time for Tax Preparers prior to the PoC, this continued throughout the PoC.
- The ACW time decreased from an average of 11 minutes and 9 seconds in the 9 months prior to the PoC to 10 minutes and 39 seconds during the PoC.
- Most CSOs stated through the survey results that Conversation Summarisation did not work well for calls on this queue.

Tax Preparers Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
Tax Prep PoC After Call Work Time	00:11:09	00:08:48	00:12:14	00:10:15	00:10:41	00:11:37	00:11:55	00:09:27	00:09:45	00:10:23
Tax Prep Control After Call Work Time	00:09:34	00:08:41	00:08:44	00:08:26	00:08:21	00:08:33	00:08:54	00:08:11	00:08:59	00:07:43




Tax Preparers queue
Average ACW for each pilot CSO during pilot
00:25:55
00:23:02
00:20:10
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #1
CSO #3
CSO #4
CSO #5
CSO #6
CSO #7
CSO #14
CSO #16
CSO #18
CSO #20
Baseline Avg
Avg ACW

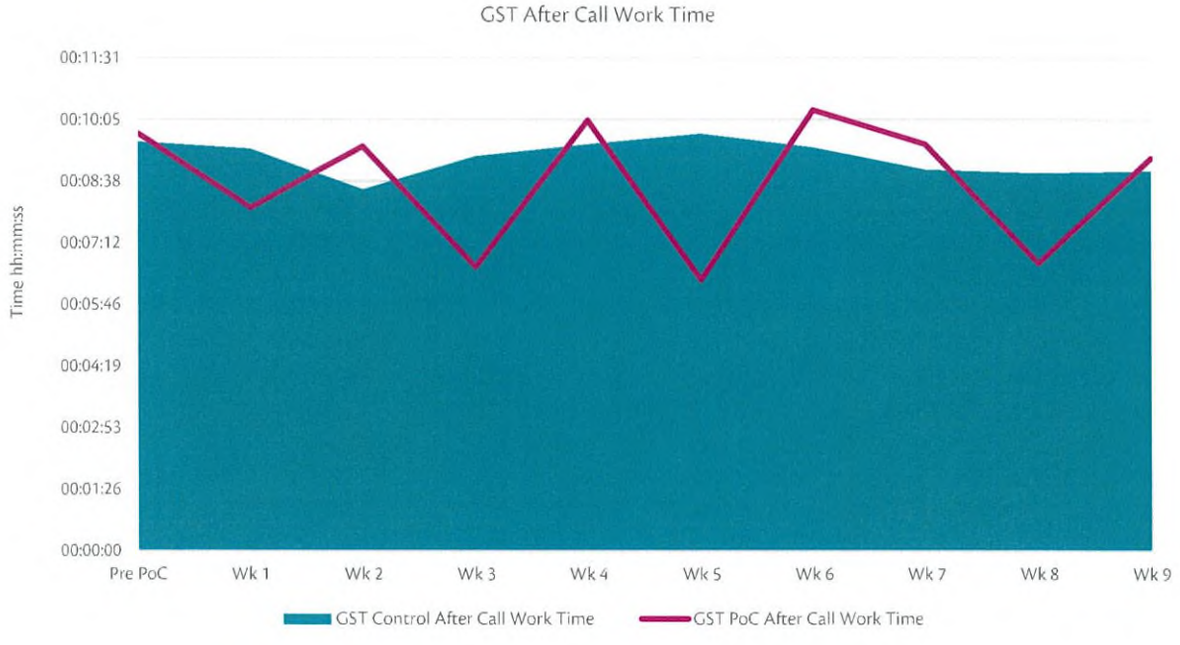

Tax Preparers call volumes - per CSO
90
8
6
51
6
11
45
70
7
CSO #1
CSO #3
CSO #4
CSO #5
CSO #6
CSO #7
CSO #14
CSO #16
CSO #18
CSO #20

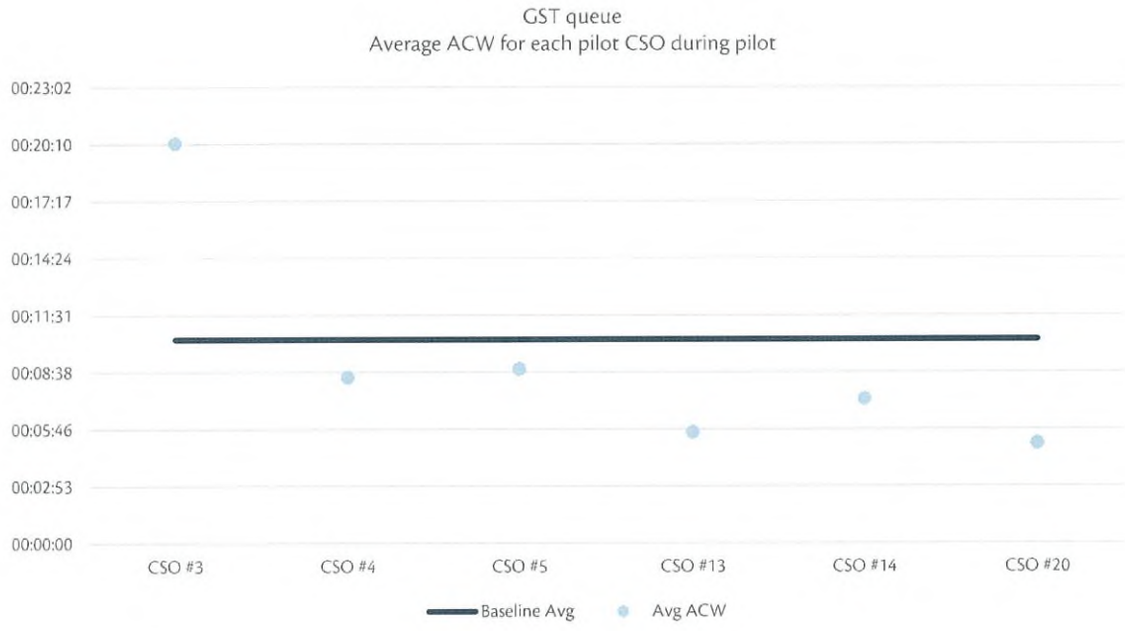
GST queue (GST)

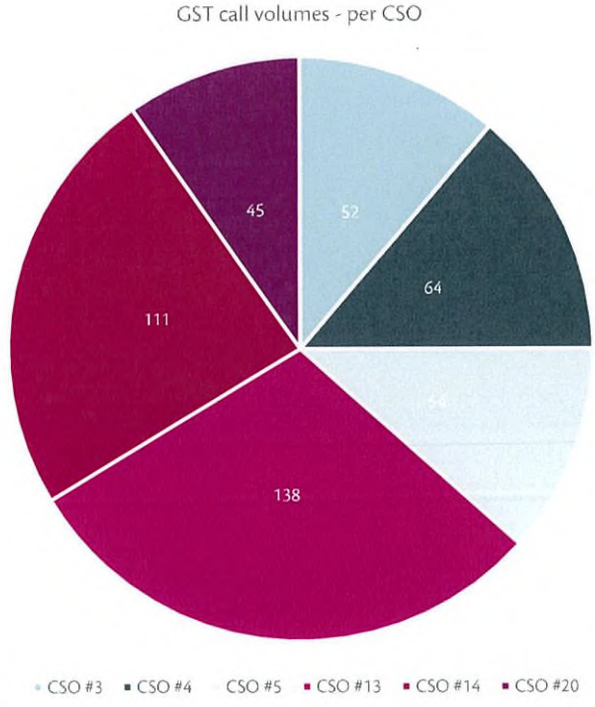
- There were 464 GST calls answered during the PoC by 6 different CSOs.
- The calls were reasonably evenly distributed across all PoC CSOs.
- Survey feedback from CSOs about how helpful Conversation Summarisation was for GST was mixed, some saying it worked well and others saying it worked well in only some GST calls.
- The ACW time went from an average of 9 minutes and 45 seconds in the 9 months prior to the PoC to 8 minutes and 20 seconds during the PoC.

GST Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
GST PoC After Call Work Time	00:09:45	00:08:01	00:09:27	00:06:38	00:10:04	00:06:20	00:10:18	00:09:30	00:06:43	00:09:11
GST Control After Call Work Time	00:09:34	00:09:24	00:08:27	00:09:14	00:09:30	00:09:46	00:09:26	00:08:55	00:08:50	00:08:53




GST queue
Average ACW for each pilot CSO during pilot
00:23:02
00:20:10
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #3
CSO #4
CSO #5
CSO #13
CSO #14
CSO #20
Baseline Avg
Avg ACW

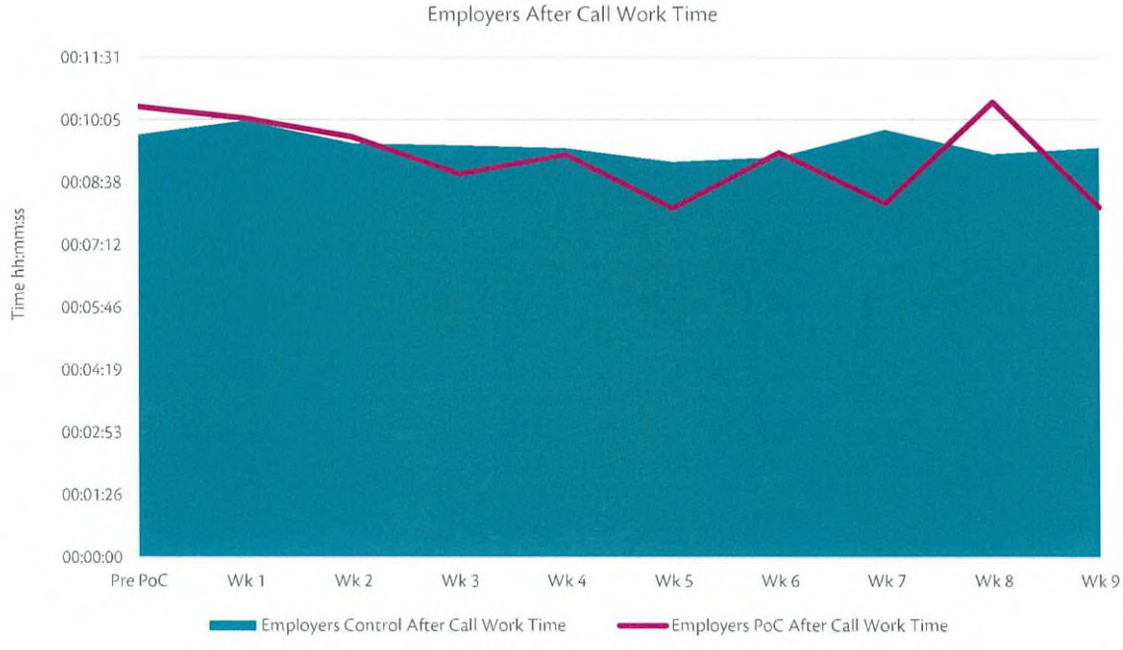

GST call volumes - per CSO
45
52
64
111
138
CSO #3
CSO #4
CSO #5
CSO #13
CSO #14
CSO #20

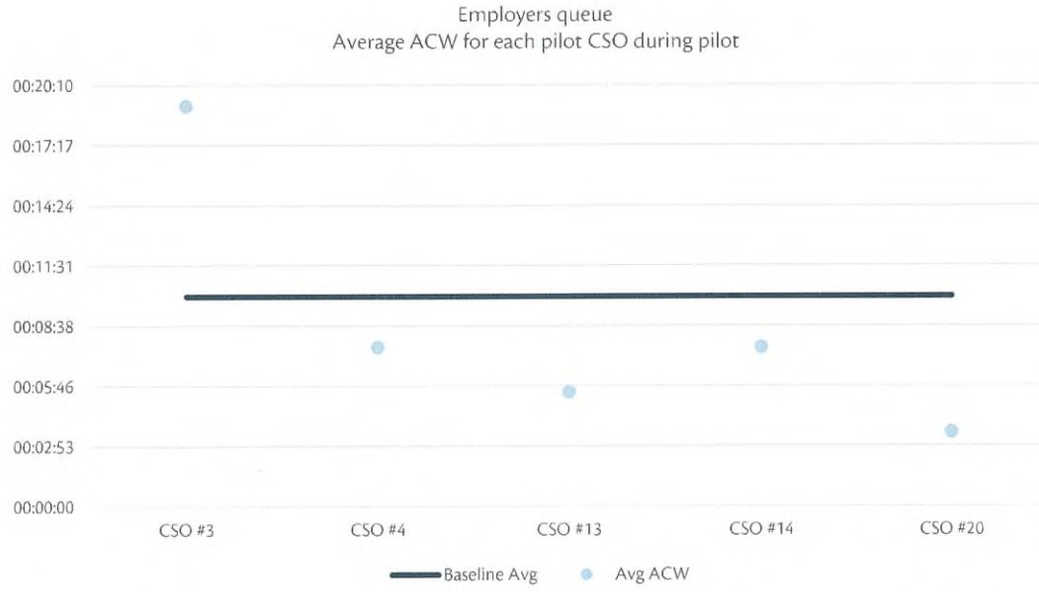
Employers queue (EMP)

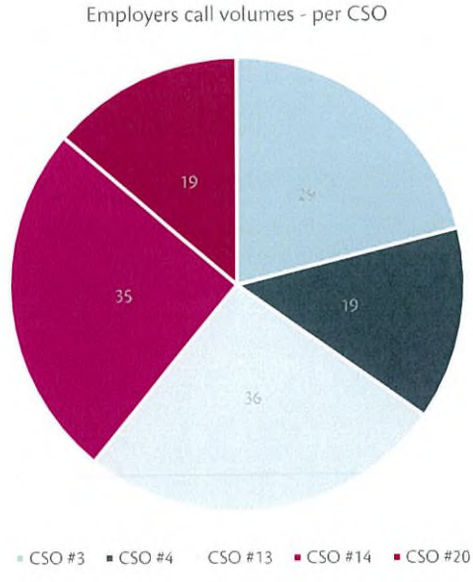
- There were 138 Employers calls answered during the PoC by 6 different CSOs.
- The calls were reasonably evenly distributed across all PoC CSOs however CSO #3 added 2 minutes and 9 seconds to the average when weighted.
- Survey feedback from CSOs was the summarisation needed more details added to it for Employers queue calls, although data suggests ACW time improvement using the summarisation. Comments made in the survey were that the summarisation provided reminders of what was discussed during these calls and CSOs were adding details to them.
- The ACW time went from an average of 10 minutes and 22 seconds in the 9 months prior to the PoC to 8 minutes and 54 seconds during the PoC.

Employers Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
Employers PoC After Call Work Time	00:10:22	00:10:06	00:09:41	00:08:50	00:09:16	00:08:02	00:09:19	00:08:09	00:10:29	00:08:04
Employers Control After Call Work Time	00:09:44	00:10:03	00:09:32	00:09:29	00:09:25	00:09:06	00:09:13	00:09:51	00:09:17	00:09:26




Employers queue
Average ACW for each pilot CSO during pilot
00:20:10
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #3
CSO #4
CSO #13
CSO #14
CSO #20
Baseline Avg
Avg ACW

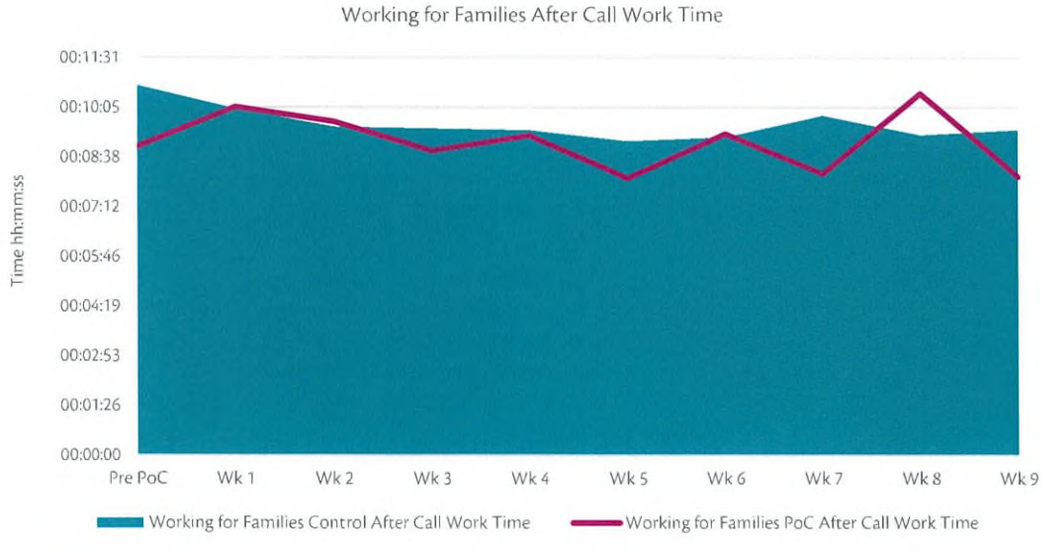

Employers call volumes - per CSO
19
35
36
19
CSO #3
CSO #4
CSO #13
CSO #14
CSO #20

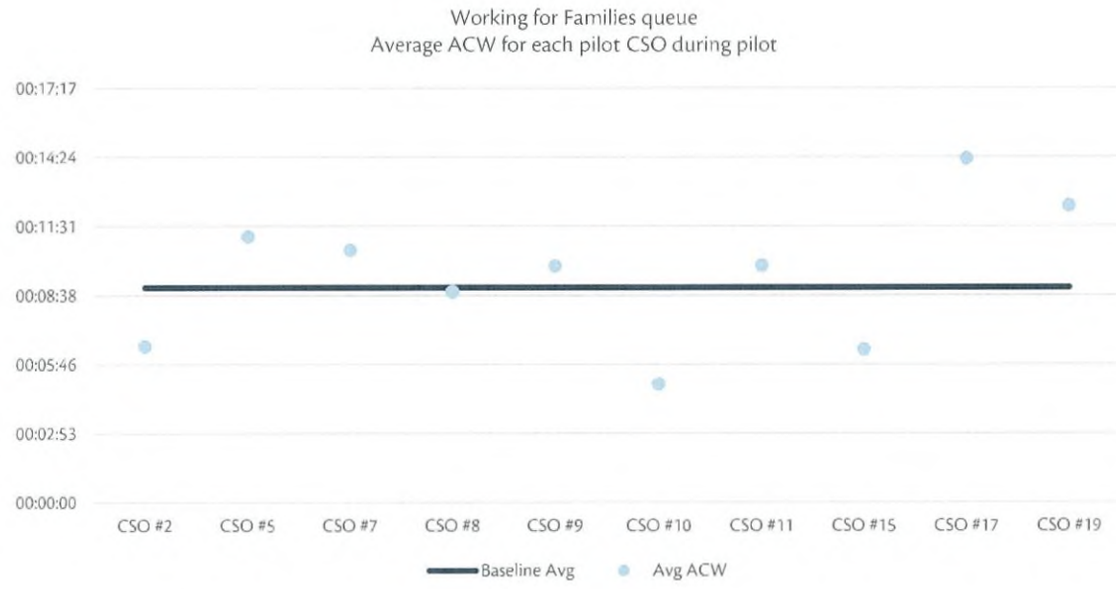
Working for Families queue (FAM)

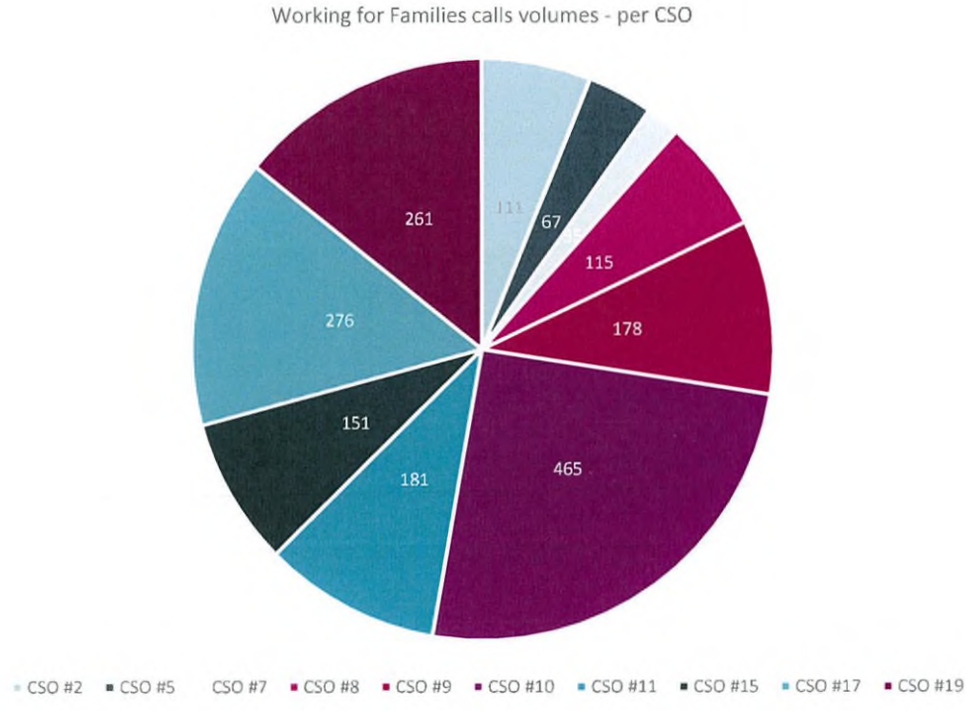
- There were 1,840 Working for Families calls answered during the PoC by 10 different CSOs.
- 25% of these calls were answered by CSO #10 This means CSO #10's performance contributed significantly to the overall results and their individual results greatly impacted the overall performance for the queue and contributed a decrease of 1 minute and 4 seconds compared to the average when weighting the average.
- There was a mix of tenure across those that answered Working for Families queue calls ranging from 5 months to 5 years in IR.
- Survey feedback from CSOs was the summarisation needed more details added to it for Working for Families calls, and that it was not always helpful. This may have led to the increase in ACW time on this queue. Some weeks have seen decrease ACW time of up to 1 minute and 55 seconds compared to the pre-PoC time.
- The ACW time increased from an average of 8 minutes and 57 seconds in the 9 months prior to the PoC to 9 minutes and 8 seconds during the PoC.

Working for Families Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
WfF PoC After Call Work Time	00:08:57	00:10:06	00:09:41	00:08:50	00:09:16	00:08:02	00:09:19	00:08:09	00:10:29	00:08:04
WfF Control After Call Work Time	00:10:44	00:10:03	00:09:32	00:09:29	00:09:25	00:09:06	00:09:13	00:09:51	00:09:17	00:09:26




Working for Families queue
Average ACW for each pilot CSO during pilot
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #2
CSO #5
CSO #7
CSO #8
CSO #9
CSO #10
CSO #11
CSO #15
CSO #17
CSO #19
Baseline Avg
Avg ACW


Working for Families calls volumes - per CSO
111
67
115
178
465
181
151
276
261
CSO #2
CSO #5
CSO #7
CSO #8
CSO #9
CSO #10
CSO #11
CSO #15
CSO #17
CSO #19

Student Loans (SL)

Student Loans

- There were 62 Student Loans calls answered during the PoC by 4 different CSOs.
- 40% of these calls were answered by CSO #13. This means CSO #13's performance contributed significantly to the overall results and their individual results greatly impacted the overall performance for the queue. CSO #13 was within 1 second of the average so is a good baseline of ACW for this queue. They have a long tenure with IR, over 5 years.
- The Student Loans queue wasn't specifically called out by any of the CSOs in the final survey, but results from weekly surveys where CSOs were taking calls on this queue indicated a mixed result about how they felt Conversation Summarisation performed.
- The ACW time went from an average of 9 minutes and 2 seconds in the 9 months prior to the PoC to 7 minutes and 53 seconds during the PoC. The PoC group had a considerably faster ACW time when compared to CSOs in the control group, of between 11 minutes and 59 seconds and 9 minutes and 17 seconds so improvements may not be as large if the PoC was rolled out to all agents as the PoC group.

International

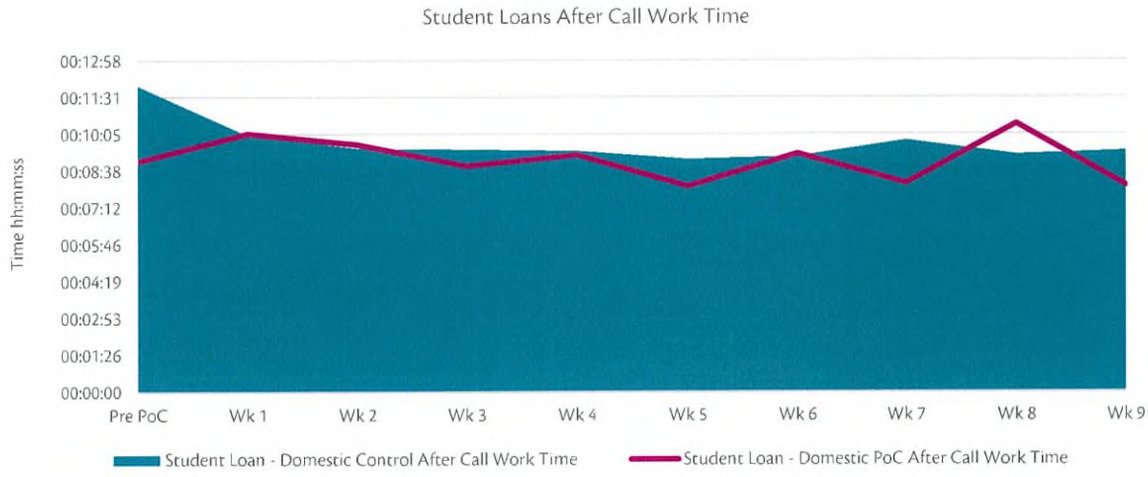
- There were 93 International Student Loans calls answered during the PoC by 4 different CSOs.
- 34% of these calls were also answered by CSO #13. This means CSO #13's performance contributed significantly to the overall results and their individual results greatly impacted the overall performance for the queue. CSO #13 was within 21 seconds of the average so is a good baseline of after call work for this queue. They have a long tenure with IR, over 5 years.
- The ACW time went from an average of 10 minutes and 39 seconds in the 9 months prior to the PoC to 10 minutes and 19 seconds during the PoC. The PoC group had a considerably faster after call work time when compared to agents in the control group, of between 13 minutes and 8 seconds and 9 minutes and 6 seconds so improvements may not be as large if the PoC was rolled out to all agents as the PoC group.

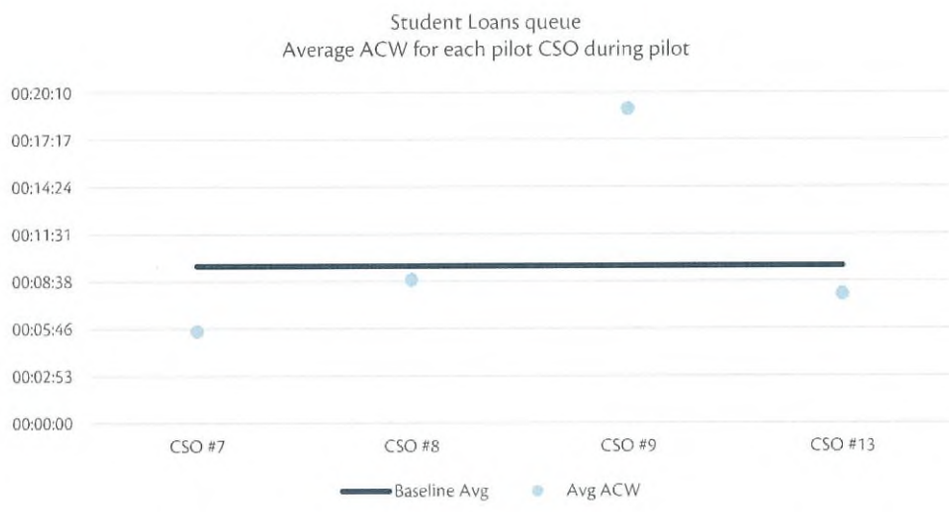
Student Loans Queue results week by week

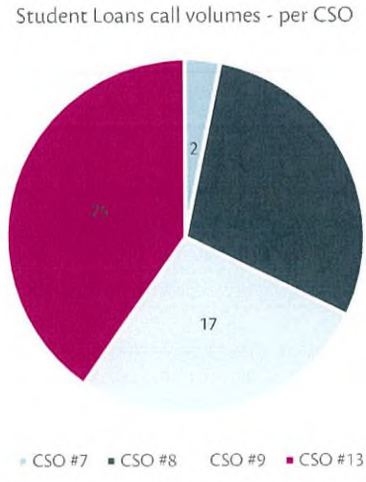
Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
Student Loans PoC After Call Work Time	00:09:02	00:10:06	00:09:41	00:08:50	00:09:16	00:08:02	00:09:19	00:08:09	00:10:29	00:08:04
Student Loans Control After Call Work Time	00:11:59	00:10:03	00:09:32	00:09:29	00:09:25	00:09:06	00:09:13	00:09:51	00:09:17	00:09:26

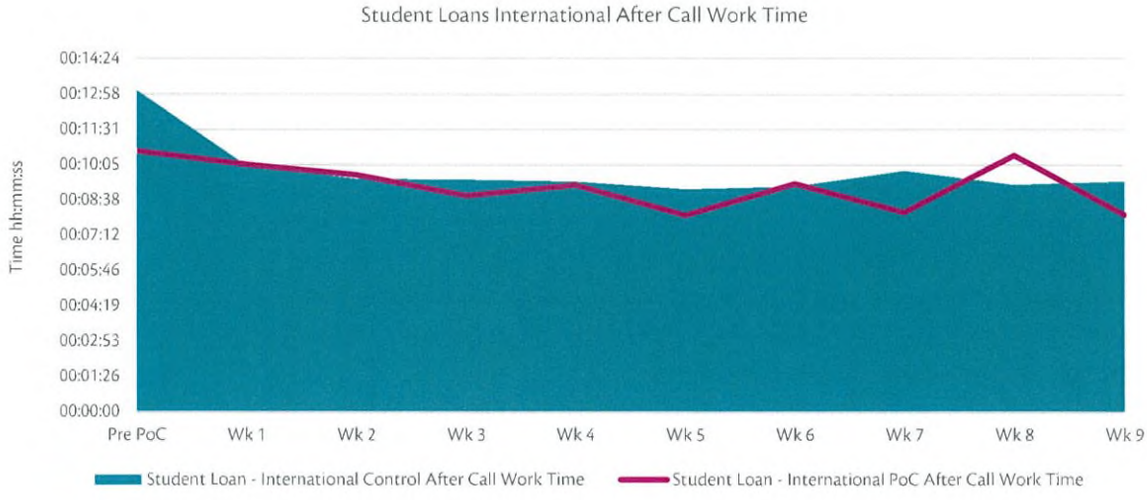
International Student Loans Queue results week by week

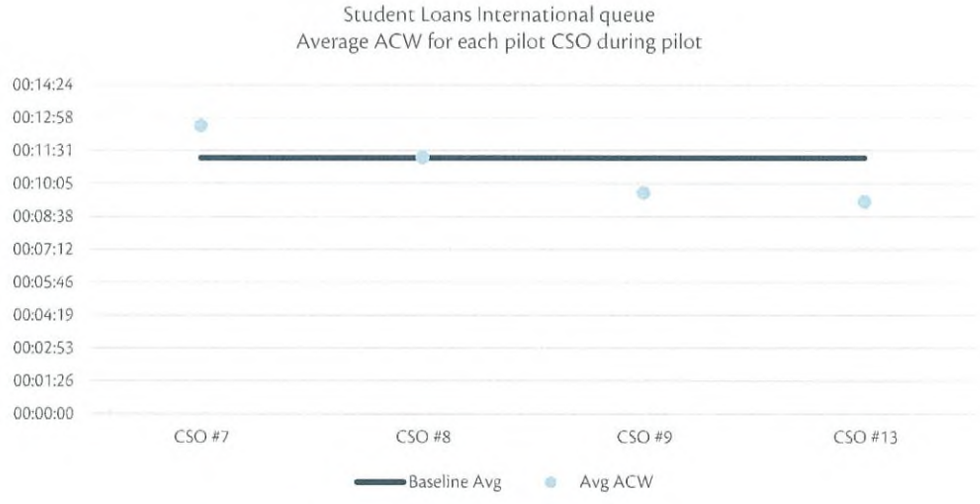
Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
Intl SL PoC After Call Work Time	00:10:39	00:10:06	00:09:41	00:08:50	00:09:16	00:08:02	00:09:19	00:08:09	00:10:29	00:08:04
Intl SL Control After Call Work Time	00:13:08	00:10:03	00:09:32	00:09:29	00:09:25	00:09:06	00:09:13	00:09:51	00:09:17	00:09:26

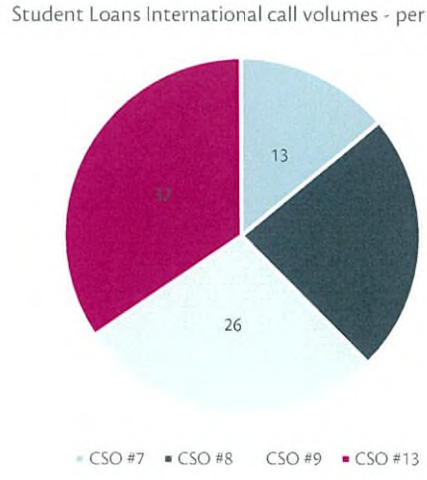

Student Loans After Call Work Time
Time hh:mm:ss
00:12:58
00:11:31
00:10:05
00:08:38
00:07:12
00:05:46
00:04:19
00:02:53
00:01:26
00:00:00
Pre PoC
Wk 1
Wk 2
Wk 3
Wk 4
Wk 5
Wk 6
Wk 7
Wk 8
Wk 9
Student Loan - Domestic Control After Call Work Time
Student Loan - Domestic PoC After Call Work Time


Student Loans queue
Average ACW for each pilot CSO during pilot
00:20:10
00:17:17
00:14:24
00:11:31
00:08:38
00:05:46
00:02:53
00:00:00
CSO #7
CSO #8
CSO #9
CSO #13
Baseline Avg
Avg ACW


Student Loans call volumes - per CSO
2
25
17
CSO #7
CSO #8
CSO #9
CSO #13


Student Loans International After Call Work Time
Time hh:mm:ss
00:14:24
00:12:58
00:11:31
00:10:05
00:08:38
00:07:12
00:05:46
00:04:19
00:02:53
00:01:26
00:00:00
Pre PoC
Wk 1
Wk 2
Wk 3
Wk 4
Wk 5
Wk 6
Wk 7
Wk 8
Wk 9
Student Loan - International Control After Call Work Time
Student Loan - International PoC After Call Work Time


Student Loans International queue
Average ACW for each pilot CSO during pilot
00:14:24
00:12:58
00:11:31
00:10:05
00:08:38
00:07:12
00:05:46
00:04:19
00:02:53
00:01:26
00:00:00
CSO #7
CSO #8
CSO #9
CSO #13
Baseline Avg
Avg ACW

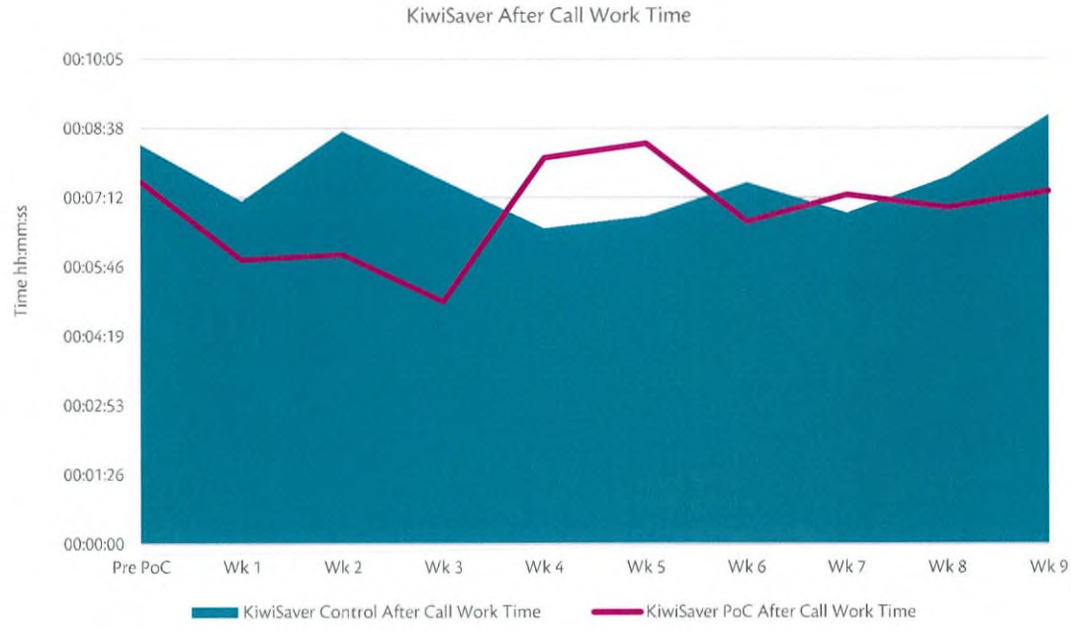

Student Loans International call volumes - per CSO
13
37
26
CSO #7
CSO #8
CSO #9
CSO #13

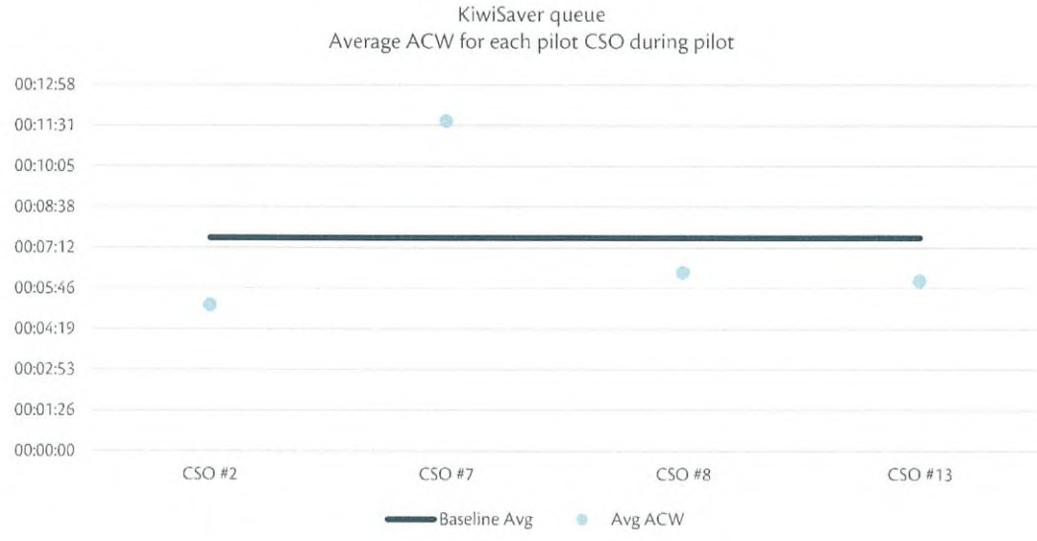
KiwiSaver (KS)

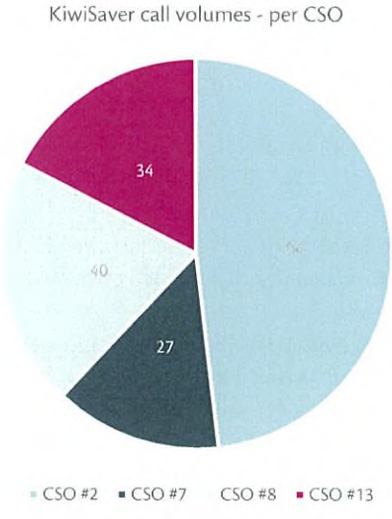
- There were 195 KiwiSaver calls answered during the PoC by 4 different CSOs.
- 48% of these calls were answered by CSO #2. This means CSO #2's performance contributed significantly to the overall results and their individual results greatly impacted the overall performance for the queue. CSO #2 was 37 seconds quicker that the average when weighting results.
- The ACW time went from an average of 7 minutes and 31 seconds in the 9 months prior to the PoC to 6 minutes and 27 seconds during the PoC.

KiwiSaver Queue results week by week

Wrap Time	Pre PoC	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9
KiwiSaver PoC After Call Work Time	00:07:31	00:05:54	00:06:00	00:05:02	00:08:01	00:08:19	00:06:41	00:07:15	00:07:00	00:07:20
KiwiSaver Control After Call Work Time	00:08:18	00:07:08	00:08:35	00:07:34	00:06:33	00:06:49	00:07:31	00:06:53	00:07:38	00:08:55







Other queues

- Other queues in the PoC included MyIR, Compliance and Credit Card Failure, all of which saw decreases in the ACW time utilised by PoC CSOs, these have been grouped due to lower numbers of calls being answered by the PoC partcipants.
- MyIR notably had an average of 6 minutes and 9 seconds in the months prior to the PoC and during the PoC 4 minutes and 33 seconds
- Compliance saw a decrease in ACW time during the PoC to on average 8 minutes and 6 seconds compared to 11 minutes and 6 seconds prior to the PoC, a decrease of 3 minutes on average per call.
- Credit card failure was the only queue where an increase in ACW time was seen on average during the PoC increasing by 9 seconds.

Other notable information

- **CSO call control:** CSO call control has an impact on the quality and accuracy of note summarisation. CSOs who were observed to have excellent call control and provided clear and concise information to customers experienced better outcomes with the Conversation Summarisation, presumably as it was able to accurately pick up the context of the call without having to decipher other items not relevant or additional to the core of the call.

- **Dictionary**: The updating of the dictionary is vital to ensure accurate summarisations. Time was spent by the VCM team ahead of the PoC adding IR information into the dictionary and input from PoC CSOs throughout added to the accuracy of summarisations. It was noted from One NZ that "fantastic work was done by VCM adding IR's specific terminology and commonly mis-transcribed words to the dictionary in advance and throughout the trial. 37 terms were added in total, significantly contributing to highly accurate transcriptions, which are crucial for generating accurate summaries. Genesys also noted that it 'typically takes other customers much longer to achieve the level of accuracy IR reached within just the first few weeks of the trial".
  There are limitations on 'sounds like' for words of 10 words for each term. We reached the limit for 'GST' and found we could not add any more words that sound like 'GST' into the dictionary. Though we won't reach this limitation for all words in our dictionary, it is a limitation.

- **Continuous improvement:** As Genesys Cloud is an evergreen product, we're receiving updates on a continued basis, just prior to, and during the PoC the following improvements within Conversation Summarisation were noted.

    - **Conversation Summarisation Length:** The supported Talk time increased from 30 minutes to 60 minutes, this means most of IR's calls fall within this limitation.

    - **Outbound Calling:** Genesys introduced the outbound call summarisation the week before the PoC commenced, this meant all inbound and outbound calls could be evaluated by the PoC Group.

    - **Enhanced Summarisation Model:** The model within Agent Assist was enhanced to use AWS Bedrock as the Large Language Model (LLM) after a specified length of internal token use, meant longer calls saw an improvement in the accuracy of summarisation.

    - **Paragraphs:** During week 3 of the PoC, summarisations began being spaced into paragraphs as an enhancement through Genesys Cloud. CSOs reported this being an excellent improvement and found the summarisation easier to follow in the paragraph structure.

    - **Currency Recognition:** Genesys enhanced the ability of Conversation Summarisation to identify and accurately represent currency values. This improvement was particularly noticed when referring to dollar amounts and dates, with feedback from CSOs stating this was much better as numbers were transcribed as digits rather than in full text.

    - **Improvements suggested:** One NZ, on behalf of IR, has raised an improvement idea with Genesys due to "the transparent process with direct feedback from agents". The idea "Call Summary in the First person" which would enable us to determine how CSOs and customers should be referred to in the summarisation has been approved by Genesys and will enter their roadmap for next year, this will continue to improve the conversation summaries available to CSOs as would read "the CSO" (or wording of our choice) rather than "the agent".

- Faults

  The following faults were logged during the PoC:

  - **Login Screen Error - INC0362502:** An issue was identified during the PoC where the log in page would appear intermittently in the wrap up code panel. This prevented CSOs from using Conversation Summarisation. This was noted to occur when the CSO had some period of inactivity in Genesys Cloud. Although this was a known issue to the VCM team previously, it had not been raised for further investigation as outside of the PoC CSOs, these panels were not utilised, and this did not impact the ability to make or receive calls.
    An incident (INC0362502) was raised on 29 November 2024, and this is still being investigated by Genesys and is not a fault with AI or the Conversation Summarisation, this occurs without these features turned on also.

  - **Incorrectly spelt words/words that did not exist:** During the PoC, CSOs identified that at times, the summary would include incorrectly spelt words or words that did not exist in the English language i.e. schedolar, customer'd, daughter'd. With the help of one of our CSOs in the PoC we were able to obtain interaction logs of instances where both occurred. A case was raised to Genesys (#003579852) on 15 November 2024 which is currently being investigated.

  - **Role reversal between customer and CSO:** During the PoC, CSOs identified that at times, the summary reverse "the agent", CSO and "the customer" and provide a summary as if the CSO was the customer in the conversation. A case was raised to Genesys (#003579852) on 15 November 2024 which is currently being investigated (this is combined with the case above and in the same reference number).

Appendix

Agent Profiles

The following information is a high-level summary of the participants involved in the PoC whose data has been included in this report.

Some CSOs were removed as they were removed from the PoC partway through either due to obtaining new roles, resignation, extended leave, or training.

CSO #	Skill mix/es	Tenure
CSO #1	GB	6 months or less
CSO #2	FAM	6-12 months
CSO #3	BUS+KS	2-5 years
CSO #4	BUS+KS	2-5 years
CSO #5	GB/GST + FAM	6-12 months
CSO #6	FFL	6 months or less
CSO #7	IND+GB	2 plus years
CSO #8	FAM/KS	5 years plus
CSO #9	FAM/SLS	2-5 years
CSO #10	FAM	6-12 months
CSO #11	FAM	6 months or less
CSO #12	FFL	2-5 years
CSO #13	EMP/GST+KS/SL	5 years plus
CSO #14	BUS	5 years plus
CSO #15	FAM	2-5 years
CSO #16	GB	6-12 months
CSO #17	FAM (stripped)	1-2 years
CSO #18	GB	6-12 months
CSO #19	FAM	6 months or less
CSO #20	BUS+KS	2-5 years
CSO #21	FFL3 (stripped)	2-5 years

AI Working Group Documentation

Please refer to linked power point documentation:

Genesys Cloud AI Conversation Summarisation - AI Working Group paper

Item 11

Inland Revenue
Te Tari Taake

AI Test Scenario Generation – PoC Findings Report

Final

Sponsor: David Carrigan, Deputy Commissioner ED&I

Authors: 9(2)(a) Prajakta Panse and Cate Robertson

Date – 14 April 2025

Purpose and Recommendations

- This document outlines the findings and recommendations from the Proof of Concept (PoC) that was conducted to assess Assurity Intelligence GenAI testing solution (AI test scenario Generator), aimed at enhancing IR's current test planning activities.

- The solution has been assessed successfully against the defined success criteria. The solution performed well against the defined success criteria (see slides 4-7 for details). Across the 20 evaluation criteria, 13 rated HIGH, 6 rated MEDIUM and the remainder was not applicable.

- The POC compared the time it took to manually generate test scenarios against the time taken to check AI generated scenarios:

✓ Using this service was 89% more efficient than the existing manual process (reviewing AI generated scenarios took 27 mins compared to 286 mins to manually generate the scenarios). Note this efficiency calculation does not consider the cost of using the service (so net cost savings will be less) and the PoC covered a simple-medium complexity change, so the efficiency savings will likely differ for more complex changes.

- A new tool was also developed during the POC. The Prompt Evaluation Tool (PET) enables repeatable, measurable experiments for continuous prompt improvement and optimal configuration. This tool will be of benefit for future use of this service.

It is recommended that the Oversight group:

- **Notes** there are potentially significant savings in using this tool as confirmed by this POC
- **Endorses** proposal to move into a Pilot phase for six months to further test this service on a greater range of changes to confirm what enduring savings it could realise (considering the cost to use the service)
- **Notes** the recommendations and next steps for the proposal (slides 9-10)

Background And PoC approach

- Business teams at IR handle large-scale initiatives and changes, often planning and writing between 1,000 to 5,000 test scenarios annually. The diverse backgrounds of Domain Specialists and Business testers lead to inconsistent scenario styles, causing increased effort during testing due to questions and clarifications. This inconsistency can result in delays or incorrect test execution due to misinterpretation.

- A Proof of Concept was proposed to test Assurity Intelligence, a GenAI testing solution on Te Waharoa (Assurity Cloud), IR's continuous testing platform. This solution can supplement our current test scenario writing, planning, and design. We wanted to assess its ability to help with requirement analysis, product research, and creating test scenarios using different techniques.

- IR identified an initiative (Imaging and scanning outsourcing- ISO) that had been delivered in one of the past releases, where test scenarios were generated by an IR domain team. In the cited activity, test scenarios were generated based on Business Functional Definitions (BFD). With these sources, Assurity consultants, using prompt engineering and Retrieval Augmented Generation (RAG) techniques, generated test scenarios using Assurity Intelligence. The output of the tool was then compared to human generated output and assessed based on set criteria.

- The PoC team followed an iterative approach to define and refine a prompt template. The prompt templates were formed in three parts: Setting the role context for the LLM, input the requirements in the prompt and define the desired format for the output. Creating a prompt template was a unique approach as it ensured ease of use for the Domain specialist and eliminated the need for the user to have prompt engineering skills.

Demo video - Assurity-Intelligence-Demo-202500410.mp4

Summary Results against IR suggested evaluation criteria

Criteria	Dimension	Assessment	Evidence
Enterprise Strategy	**Stewardship**	High	The use case demonstrates strong alignment to Stewardship by building tools for our testing practice to be adaptable and scalable for current and future needs.
	Efficiency and Effectiveness	High	The use case demonstrated strong alignment to efficiency gains by supplementing manual tasks at a larger scale than would be attainable by an individual in the same timeframe.
	Broader Contribution	High	The use case demonstrated strong ability to enable innovative solutions by use of AI tools to supplement people capability and creativity. If successful this would be shared with other agencies like other CTAAS Assurity services.
Outcomes	**Problem/opportunity addressed**	High	The use case was able to demonstrate that the tools can assist IR business teams with test design, analysis and test scenario writing that will produce a consistent content structure and formatting, tailored to specific test scenario types and requirements defined by IR
	Operational/business efficiency	High	The use case was able to measurably save time and complete the large scale of work. The generated scenarios can be easily made to evolve as changes are made to the BFDs or tweaks to the prompts.
	Improved customer outcomes	Medium	The use case enabled us to produce core test scenarios consistent with Test Analysis best practice. The Domain specialist would be able to easily utilize the output as a starting point to build more detailed test scenarios. This will ensure overall structure and quality of test planning will increase.

Summary Results against IR suggested evaluation criteria

Criteria	Dimension	Assessment	Evidence
Governance	**Data security**	High	We observed that the data security was maintained during the POC inline with the expectation, data provided during POC stayed within the tenancy.
	Privacy/confidentiality	High	Privacy and confidentiality requirements were maintained during the POC inline with the expectation and no issues were observed.
	System security	Medium	No additional security risks were observed , the issues and risks identified in security risk memo remained consistent during PoC.
	Data integrity	Medium	New AI-generated test scenarios created as part of this use case are maintained within the system; The content was not transferred to any core systems or products automatically.
	Ethics	Medium	Results generated as part of this use case were compared to human-created outcomes. All AI-generated content was human-reviewed and evaluated. No new risks were identified that may hinder the current test planning process.
	Affordability/Scalability	High	The use case demonstrated a strong ability to be scaled through API-based solutions, and the broader use case concept can be replicated for other similar opportunities.
How we work	**Our people**	High	The use case demonstrated a clear opportunity to change the way that our people approach repeatable tasks using innovative approaches. This tool will support our teams to get better at test planning and design work.
	Our customers	N/A	
	Operations/processes	Medium	While the use case was effective for our purpose, the cross-application for similar operational tasks would need to be validated during pilot phase.
	Partners/vendors	N/A	Partner/Vendor impacts are low as we can see this tool can supplement existing process but does not impact otherwise.

Summary Results against success criteria defined during PoC

Criteria	Assessment	Evidence
Relevance and Alignment	High	The outputs were aligned and stayed within context as provided by the BFDs.
Time Efficiency	High	Scenario generation and analysis taking on average ~18 minutes.
Ease of Use	Medium	The UI is very intuitive. Potential users should not have any issue interacting with it. However, skilled prompt engineering required to maximize output, recommend this skill is available until pre generated prompts are in use during pilot phase.
Adaptability	High	The generated scenarios can be easily made to evolve as changes are made to the BFDs or tweaks to the prompts.
Exploratory Insights	High	We were able to produce core test scenarios consistent with Test Analysis best practice. The Domain specialist would be able to easily utilize the output as a starting point to build more detailed test scenarios.

Supporting information - Time Efficiency

Criteria	AI test scenario generator (Assurity Intelligence)	Human Generated
Output generation time (per BFD – on average) Note only 2 BFDs were evaluated	7 mins (includes refined prompts) – 54 unique scenarios	208 mins (3 hours 46 mins) – 26 unique scenarios * 8 min per scenario.
Evaluation time (per output) to review the output generated	20 mins - Time required by Test Analyst to evaluate with testing lens. Domain Specialist may require less time.	1 hour (Evaluation of written scenarios by Domain Specialist to finalise test scenarios)
Net time spent (Generation + Evaluation time)	27 mins	268 mins

Some additional stats and information to note:

- Note that the Initiative selected for the PoC was low- Medium complexity so time spent by IR team may be different depending on requirements under test. Some sample times for critical pieces could look like 4-6 min per low complexity scenario, 20 mins per medium complexity scenario and 1-2 hours per high complexity scenario.
- Note that the total scenarios created by SME for the initiative picked for PoC were 209 (with 47 return types and 79 corrie sorts of doc types).
- Total scenarios created by AI – 79 unique scenarios per definition (5 BFD's included in PoC).
- Total time required for test planning and writing by SME's – Approx. 8 mins per scenario for 209 scenarios– 1672 mins (Approx. 27.3 Hours)
- Total time required by AI for test planning and scenario writing (Includes time required for prompt with BFD, Dev tasks and refinement) – 30 mins
- Total time required for Test Analyst to evaluate all scenarios generated by AI with testing lens (with limited business domain Knowledge) for all 5 BFD's - 55 mins

Key Findings

Prompt 1

The more effective method of prompting was direct prompting. Copy and pasting the Definitions (one definition per prompt) gave the most consistent output that aligned with the prompt intent. Following up each output with a prompt asking "if assumptions were made" helped to identify any potential knowledge gaps from the BFD. The assumptions generated could help consolidate the development of test scenarios.

Output

- The AI was able to produce simple core scenarios that were accurate to the context provided by the BFD. Of the 3 models, qwen2.5* proved to be the most consistent in its results. The output, with the appropriate prompt, would provide a set of test scenarios that can be a skeleton for the Domain Specialists to consolidate further. It also proved to be very useful in identifying certain unique scenarios conditions for example edge case scenarios that may not have been thought of.
- A follow up prompt asking "if assumptions were made" proved to be very useful in making sure that the output aligns with the solution design. These assumptions can also be used to further evolve and improve on the original scenarios generated.
- *The qwen2.5 LLM model is a relatively small LLM – 7.6 billion parameters, encompassing up to 18 trillion tokens, released in October 2024. It demonstrates significant advancements in instruction following, long-text generation (over 8K tokens), understanding structured data (e.g., tables), and generating structured outputs.

Evaluation

- **Relevance and Alignment**: The outputs were aligned and stayed within context as provided by the BFD.
- **Time Taken:** Net time spent per BFD on average was: 17 mins 13 secs (AI:133 secs + Analysis: 15 mins)
- **Ease of Use:** The UI is very intuitive. Potential users should not have any issue interacting with it. A predefined prompt template improves the experience by reducing the need for technical expertise.
- **Adaptability:** With the use of basic prompting technique, the output can be easily made to evolve when changes are made to the BFDs
- **Exploratory Insights:** The tool was able to produce core test scenarios consistent with Test Analysis best practice. The Domain specialist should be able to easily utilise the output as a starting point to build more detailed test scenarios.

Recommendations for the Pilot phase

- Rollout and adoption phase is recommended to be phased instead of full rollout. IR can start with one domain to ensure teams get opportunity to familiarise with the tool.
- It would be critical to have a prompt engineer work with the Domain Specialists, during pilot phase, to maximise the benefits and create better quality prompt templates.
- It is recommended that the Testing Domain at IR have a repeatable process for producing test scenarios and ensure clarity on how Assurity Intelligence can be weaved into existing test planning process.
- When rolling out the service, it is recommended we share pilot learnings on requirements' structure and formatting with wider teams. For better AI output, we may want to consider writing requirements in Markdown text format, as it aligns with the AI's training data. Markdown is a lightweight markup language that you can use to format plain text. It's widely used because it's simple and easy to read, even in its raw form.
- Assurity Intelligence or any other LLM can be used in future to gauge the quality of the requirements and identify any gaps.
- Full immersion of Assurity AI team into IR test planning phase may identify areas where the LLM can provide more value.

Next steps for Pilot rollout

- Engagement with various teams to update AI risk assessment in context of the Pilot
- Request assessments and approval for pilot from Legal, CISO and Privacy team
- Request approval from Technical Design Authority
- Endorsement from CCS Senior leadership team to support rollout in Digital Domain as part of Pilot
- Finalise the Rollout plan, monitoring and approach working closely with Domain Lead (Digital)
- Commercial agreements and service costs agreed between IR and Assurity.

Overall learnings

Test Scenario generation learnings

1.1 When compared to human-created outcomes, Assurity AI generated more variation in types of unique test scenarios and considered wider elements.

1.2 Assurity AI makes assumptions while designing the test scenarios. We observed that although the scenarios stayed within context of BFD it was beneficial to understand assumptions it made. When testing using any AI tool this is good lesson learnt to ask AI on what its assumptions are. This helped team refine the prompts better.

Prompt Development learnings

2.1 We observed that the quality of test scenarios was largely based on the prompt template used. SME generated scenarios have domain knowledge not mentioned in the business functional definitions. To be reasonably close to the expected scenarios, Assurity AI should be loaded with domain knowledge such as business process, solution designs, user guides, testing policy etc

2.2 Quality of the outputs will improve and give consistent results when prompt template is used consistently for bigger data set. The prompt template will improve when used for different sets information via BFD or other requirement formats.

2.3 Continuous testing and evaluation of prompt was important, and automated testing helped this exercise.

Role of AI compared to human tasks and time

3.1 There were some elements of efficiency gains that were difficult to measure - e.g. continued use of AI prompting helped us find business functional definitions gaps. Users also found the evaluation time reduced gradually as prompt improved.

3.2 Assurity AI will enable domain specialists and testers to more effectively spend their time thinking of better-quality scenarios than low hanging obvious tests. This will ensure overall structure and quality of test planning will increase.

Findings – PET(Prompt Evaluation Tool)

During the AI PoC, the team identified the need for a tool to automate prompt evaluation and data capture. As an additional deliverable, they developed an automated evaluation tool using Langfuse called PET.

PET is build using Langfuse an open-source platform designed for developing, monitoring, evaluating, and debugging AI applications powered by language models. The automation tool (PET) provides a platform for tracing, measuring and evaluating interactions between Assurity Intelligence and the LLMs.

When used in conjunction with automated tests, run against the Assurity Intelligence API, this provides a platform for repeatable, measurable experiments, enabling continuous prompt improvement and optimum configuration for each initiative's needs.

When Assurity intelligence is rolled out to pilot, this tool can be used to capture data on all above mentioned measures. With the use of an automated test tool we can evaluate the performance of LLM and continuous monitoring can be performed in a structured way. This is inline with our testing pattern for continuous testing. This tool has potential to benefit other AI PoC/Pilots underway, at the moment, as well as those that require automated structured testing and may form a basis for a future IR AI test pattern.

Item 12

Voice isolation – Review of Early Adopter Feedback

Report date: 1 May 2025

Prepared by: Phyllida Crawford, Domain Principal – Technology Experience

Executive Summary: Feedback from early adopters indicates the voice isolation feature could help reduce cross conversation issues for some business users. 55% of early adopters who tried the feature reported it became easier for others on the call to hear them. A small percentage of users (4%) reported a deterioration in call quality.

Recommendation: Proceed with rollout to all IR people. In the comms, explain clearly in what situations the voice isolation feature could help you, how to quickly turn it on and off, and emphasise that it is optional to use.

Feedback Summary:

On 15 April 2025 we asked the early adopters group (~280 business users) to try out the voice isolation feature in Teams. We allowed 2 weeks for feedback.

We received:

- 52 survey form responses from people who tried the feature
- 1 emailed response from a person who declined to take part because it involves submitting their biometric data for MS Teams use, which they are not interested in doing at this stage.

The survey respondents told us:

- 92% said the feature was easy to set up. 6% were neutral. 1 person was not able to set it up due to an error message in Teams.
- 70% were satisfied or very satisfied with the feature, 25% were neutral, 5% were dissatisfied.
- 55% said other people on their calls found it easier to hear them, 41% found no noticeable change, 4% (2 people) reported it was worse:
    - Person 1: There were times when the person's voice became faint or where the person was not immediately heard at the start of the call and it improved over 5-10 seconds.
    - Person 2: When this person was talking the person sitting next to them could be clearly heard but when they stopped talking it went silent. This person has a noise suppressing Jabra Evolve2 85 that might be fighting with the voice isolation technology. When they used a different headset, voice isolation worked fine.
- 1 person was unable to set up the feature due to an error message "Couldn't capture your voice. Try moving to a quieter area or using a different mic, and start your voice capture again."
    - The person tried moving locations and cannot get this error to disappear in the quietest office setting they can find. They have tried using a headset microphone and using their laptop microphone. They have also tried restarting Teams.
- Other themes in the feedback:
    - People were unsure how to check if the voice isolation feature is on and how to quickly turn it off if needed

- It would be helpful to clarify the situations where voice isolation could help you given that:
  - user action is required to enable it
  - for many people the feature will not have any noticeable effect, or might cause new (actual or perceived) issues.
- There was some confusion about whether this feature could help contact centre staff who primarily use Genesys
- Some people attributed deteriorated call clarity to the voice isolation feature when it was due to other causes, e.g. laptop microphone being used instead of headset, or basic noise suppression being turned off
- Some people found it hard to find the relevant settings in Teams
- One person commented that it was not a good time to be requesting feedback given the short work weeks with Easter/Anzac and many people on annual leave.

Link to full survey results: Voice Isolation Feature Feedback Form.xlsx

Item 13

Inland Revenue
Te Tari Taake

Executive Level Governance

Enterprise Priorities and Performance Committee

myIR Navigation Assistant Closure Report (INIT-285)

12 June 2025

Purpose

The purpose of this paper is to share the closure report for the 'myIR Navigation Assistant' and a summary of the evaluation completed, including benefits realised.

Recommendations

It is recommended that the Committee:

- **Endorses** the myIR navigation assistant closure report.
- **Notes** that the Business Owner will approve the report and final closure.
- **Notes** there was an underspend which will be returned to the Investment Fund. This has been confirmed by Finance Services.
- **Notes** the confirmation of benefits and outcomes, including a 25% reduction in web messages relating to 'update address' (see page 5).
- **Notes** that myIR navigation assistant has been handed over to CCS PDD Digital and Customer Experience to be managed and adapted as part of business as usual.
- **Notes** the full evaluation report is available which provides additional detail about the performance of the assistant content model. myIR navigation assistant Evaluation May 2025

Sponsor

Name: James Grayson

Title: Deputy Commissioner, CCS-I

Signature: 9(2)(a)

Executive Summary

The myIR Navigation Assistant was delivered as a pilot with a business outcome aiming to reduce low complexity web messages. The Assistant acts as a secondary navigation aid in myIR, focusing on guiding customers to tasks and actions that are difficult to locate but can be completed through self-service.

The initiative assumed that navigation failures are a primary reason why customers send us web messages.

This assumption proved true for assistance relating to addresses, but not for any other topic covered by the assistant. Looking at how customers interact with the assistant, we've learned that the primary navigation in myIR largely works.

This leads us to conclude that most customers do not send web messages because they can't find what they need, they send web messages for other reasons (assurance, specific situational assistance, non-self-serviced tasks).

The initiative targeted a 25% reduction in low value high volume web messages across 5 topics. The "update address" topic has been a clear success – helping customers to navigate to where they take this action, and reducing the low complexity contact that is typically driven from this topic, currently projected at over 6,500 hours in savings annually.

If our assumption about navigation failure driving web messages was correct, we would see all reduction targets achieved. However, with the wider scope of the assistant, we know that it helps customers navigate to a wide range of topics, with smaller reductions in low complexity demand. Across 27 topics serviced by the assistant, we see greater accumulative value as demand is reduced across the spectrum.

Background

The myIR Navigation Assistant was first brought to the Portfolio Governance Committee (PGC) through endorsement of the delivery charter back in **December 2023** (formerly called Low Value Web Messaging). It was at this time that PGC noted the preferred option, Navigation Assistant, and delegated the go/no-go decision for the deployment of the pilot to the Steering Group.

In **April 2024**, the initiative returned to PGC to note that the scope would be increased to support a broader range of customer queries than initially proposed. It was also acknowledged that the initiative had shifted to the October 2024 release (go-live on 4 November).

During the quarterly update in **July 2024,** PGC noted the successful testing of potential responses to customers through generative AI (Copilot for Bing), which assisted with the functional design of the Navigation Assistant. The Committee were informed that from the go-live date customers will see the function at the bottom of all pages within the site.

PGC noted in **October 2024** that the target date for realisation of benefits was June 2025, following the trial period (pilot) concluding in April 2025.

Initiative performance

The planned delivery of this initiative was April 2024, and it was delivered in October Release 2024 (OR 24). The delay was due to technical requirements of the build, which only became available via

the Service Pack as part of OR24. This delay allowed us a longer build time, resulting in Phase 2 (scope expansion) to be delivered at the same time as Phase 1. It's important to note that the delay in delivery also meant the trial was completed outside of peak time, which is not when our target customer group of 'individual infrequent users' log in to myIR. We expect to see usage of the assistant increase over the upcoming months.

The 5 key topics that were a primary focus for this initiative were:

1. Update address
2. Request an instalment arrangement
3. Update bank account
4. Opt-out of KiwiSaver
5. Find your tax code

Update address – impact on self-service

- We can see a distinct, attributable rise in self-serviced address changes from when the assistant went live in November 2024.
- We achieved a 25% reduction in the volume of 'update address' web messages across the trial period.
- This is validated by a 23.5% rise in self-serviced address updates in myIR.

Other key topics

- In Phase One we expanded the scope of the assistant to include 22 additional topics.
- As we've now learned that primary navigation in myIR largely works, there are no other significant single topic reductions across the 27 topics serviced by the assistant. Instead, we see greater accumulative value as demand is reduced across the spectrum.

Real time feedback

The assistant serves as a valuable channel for us to identify and better understand customers' needs and intentions through their unique interactions with the assistant.

Through the assistant we can see the following:

- Navigation failure - Identifying what topics and tasks are difficult for customers to find in myIR.
- Failure demand - Identifying the reason for unexpected or reactive customer demand.
- Potential self-service tasks - Understanding what tasks and actions customers are looking to self-serve but currently need to contact IR for.
- Customer confusion - Understanding common topics of confusion for customers where more education is required.
- Issue identification - Early identification of issues or outages with IR's services.

Initiative learnings

We set out with our 5 key topics, as identified by the networked team which determined our initial scope. While 'update address' worked as expected, we saw customers use the assistant for tasks

primarily outside of the original 5 topics. We note three key learnings based on how customers interacted with the assistant:

1. **Limitations in self-service**

   When comparing our actual top 5 against our anticipated topics, we see that many interactions are not navigation-based and instead align more closely to ad-hoc tasks that individuals need to perform. Many of these were not able to be fully self-serviced.

   e.g. 'Requesting an instalment arrangement' has several limitations and business rules which means not everyone who sets out to set up an arrangement in myIR is able to.

2. **Customer education**

   With the 'KiwiSaver opt out' request, we've learned customers don't understand they can't opt out after 8 weeks and therefore are not always satisfied with the assistant's response suggesting a savings suspension, resulting in contact.

3. **Anticipated navigation failures**

   The initial set of 5 topics were thought to be navigation-driven failure demand, however evidence from our trial suggests that this is only true for the 'Update Address' function in myIR.

   While our other topics did see some usage, they were not entirely navigation-driven queries and subsequently were not always able to be responded to effectively.

   Throughout the trial and continuing into BAU we continue to observe how customers are interacting with the assistant and adapt the content accordingly.

Initiative closure

At the closure of the initiative, the following has occurred:

- ✓ Initiative closure report completed
- ✓ Evaluation report completed
- ✓ Underspend of $16,616 returned to the investment fund
- ✓ JIRA activities closed
- ✓ Handover to BAU (within the same team) completed

Financial position at closure

We delivered this initiative under budget, with $16,616 able to be returned to the investment fund.

Within the delivery charter it included the assumption that Phase 2 (full production implementation) of the recommended solution would require additional funding. However, as noted within the Performance section above, Phase 2 (scope expansion) was delivered at the same time as Phase 1, without the need for additional budget.

Financial Year / Period	Actual	Budget	Variance (under / over)
2024	$330,386	$347,002	$16,616 under
Totals	**$330,386**	**$347,002**	**$16,616 under**

Benefits and Outcomes confirmation

INITBENE links (Jira)	Benefit Description	Target Benefit	Actual Benefit (at closure)
INITBENE-30	Reduction in the number of lower value web messages associated with navigation	25% reduction of selected 'low value' messages	25% reduction achieved for 'update address' topic. Equates to 6,500 hours annually for this topic.

Handover

- The initiative delivery team is the same team that will be responsible for the BAU activities now that the initiative is closed.
- The overall accountability for the myIR navigation assistant will continue to be in the CCS-I PDD&D Digital domain, with the Service Owner, PD&D CX/UX) remaining as the key lead and contact.
- The PDD CX/UX team will continue to actively monitor and adapt the content model of the assistant based on customer input, particularly over the peak period.
- Note - The ongoing effort to maintain the assistant is minimal.

Appendix A: Closure checklist

This content is not required for executive governance and can be separated from the report once completed.

Item	Status	Notes
Scope / Requirements: All items scheduled for release have either been withdrawn or completed	Completed	
Schedule: Has been completed, with all items closed	Completed	
Deliverables: Have all been completed, where applicable submitted to the relevant governance forum and are approved	Completed	Post-EPPC
Risks: All initiative risks have either been closed or with their agreement transferred to the business as an operational or tactical risk	To be done	Post-EPPC
Issues: All initiative issues have either been resolved, closed or with their agreement transferred to the business as part of handover	N/A	
Dependencies: All initiative dependencies have been delivered / closed or with their agreement transferred to the business as part of handover	N/A	
Change requests: All initiative change requests have been approved, or withdrawn	Completed	
Benefits: An assessment of the benefits delivered against planned has been completed and the business has mechanisms in place to ensure realisation of benefits continues to be monitored and measured. *Note: In the Delivery Charter you would have articulated the value elements. These should be summarised here.*	Completed	
Finance: The final funding position has been confirmed with Finance Services.	Completed	
Lessons learnt: Lessons have been identified, documented and if applicable assigned to the business for action. Actions will be monitored and managed until complete	Completed	
Early Life Support: Early life support has finished with exit criteria met	Completed	
Handover: All agreed handover activities have occurred and been accepted by the recipients	Completed	
Document management: Shared workspaces, Teams sites, temporary storage locations and email accounts have been closed and files moved to an enduring location, archived or deleted per IR's Records Management Specification.	Completed	

Item 14

Inland Revenue
Te Tari Taake

Canon Email Ingestion Channel Functionality PoC Update

Author: 9(2)(a) Andre James, Edmund Dickie

Business Owner: Jane Elley

Version 1.6

Date 18/09/2025

PoC approved by TDA on 24th June 2025

DD-1152 was presented to the TDA on the 24th June 2025 (Appendix 1).

The PoC was approved, subject to a Privacy Impact Assessment (PIA) and Security Risk update, prior to Go Live.

The PIA and Risk Assessment were conducted during the build process.

- The Privacy Assessment was signed off, on 23 July 2025.
- The Security Risk Assessment update was signed off, on 20 August 2025.

With the prerequisites for the PoC satisfied, the PoC commenced early September 2025.

IR chose the following low volume mailboxes for the POC:

- Upload Customer Correspondence
- Account Closure
- Child Support Correspondence
- Tax Pooling
- Tax Agents Palmerston North

PoC Benefits

The PoC proposed to deploy an email ingestion channel, addressing stand alone emails utilising the existing imaging and scanning integration layer with CBS (InFLOW, SFTP, START) through the existing physical correspondence process.

Utilising existing CBS data extraction processes, exemption management and investment in software, the PoC has realised the following initial benefits:

- Successful upload of digitised emails via SFTP for digital collection by IR in line with current paper correspondence process;
- The additional data (IRD number) allowed this work to be automatically attached to the customer as per current CBS processes into START;
- Emails are now visible on the customer file as a task (as soon as processed from CBS);
- IR staff are more efficient by actioning START tasks, rather than directing the emails to where they can be worked;
- Better visibility and inclusion into "total work planning allocation" as well as an audit trail of correspondence from customers;
- Handling time for customer queries has reduced, customers now get more timely and accurate responses from IR;
- Increase in the number of emails processed by IR, better utilisation of IR staff time;

9(2)(b)(ii)

PoC Benefits cont.

- The POC is in hypercare, and working with CBS on exemptions and rejections to automatic ingestion (currently sitting at 7.3%). Further efficiencies around attachment type, file extension queue type, are expected.
- The PoC utilises existing technology, processes and work instructions. The POC has meant minimal change and adjustment for CBS staff to execute an additional process.
- Management, storage and destruction of digital data in accordance with existing SLA's and best practice from IKM team;
- The additional process meets all other existing SLA, & KPI's. This will be reviewed at different stages to ensure relevance at scale.

A snapshot of a week

Average CBS Statistics across the course of a week for the POC:

Period	Average Emails (day)	Average Exceptions (day)	%	Average Rejected emails (day)	%
5 working days	54	29	44%	5	7.3%

Key takeouts:

- CBS able to prioritise emails – e.g. "Tax Pooling" emails;
- The POC is in hypercare, and working with CBS on exemptions and rejections to automatic ingestion (currently sitting at 7.3%). Further efficiencies around attachment type, file extension queue type, are expected.
- Majority of exceptions are to the Account Closure Mailbox (emails from solicitors and tax agents about deceased customers), where IR numbers have not been included in email to IR;
- Correspondence planned to tax pooling intermediaries to ensure IR tax numbers are included in emails going forwards.
- Other comms being explored to decrease number of emails coming in without IR numbers.

A snapshot of a week

Average IR Statistics across the course of a week during the POC:

Period	Average time spent managing 3 mailboxes (week) PRE POC	Average Exceptions across a week	Amount of time spent managing all Exceptions across 5 mailboxes AND all paper correspondence POST POC	Time difference as a %
5 working days	**8 hours 3 mins**	**96**	**2 hours 30 minutes**	**5 hours 30 mins difference (68.9% time difference - IR staff time spent on more value add activities)**

Key takeouts:

- In the 4 weeks prior to the POC Go-Live, IR (CSA's) spent on average **8hr 3mins** each week managing the 3 CSA mailboxes (Account Closure, Upload Customer Correspondence and Child Support Correspondence).
- Since the POC go-live, the only work the CSAs have done for these mailboxes is manage the exceptions which is done in the daily 30mins allocated to managing **all** exceptions across the 5 POC mailboxes **and** all paper correspondence.
- Where previously IR needed to manually assign **all** correspondence to a customer before it would stage to a work queue, the correspondence now lands associated to a customer and directly in the right workgroup straight from Canon.

Time in the 'Identify the customer' tasks:
(Pre-Poc) 25-29 August: 11 hours, 58 minutes - 582 completed
(Post-Poc) 8-12 September: 3 hours, 36 minutes - 75 completed

In summary....

A PIA and Security Risk Assessment has been undertaken on the POC and signed off.

The POC has commenced and realised initial benefits.

Technology patterns are working as proposed in the original TDA paper.

A test plan has been developed and sent for endorsement with key stakeholders.

Mailboxes are being actively reviewed; merged or closed when customers can be directed to our preferred digital channels, and this will be ongoing in line with our existing Chanel Strategy.

We expect to realise further efficiencies as subsequent phases proceed. IR will focus on further reducing the number of exceptions through communications and providing guidance to ensure that the correct doc types are attached and IR numbers are embedded into emails, to enable seamless processing.

Next steps...

The Parties seek approval to proceed to production with additional mailboxes (39 planned, estimated at 2740 emails a month) in October 2025, post conclusion of the PoC.

Stakeholder Engagement Matrix

Stakeholders	Role	Engagement/RACI
Jane Elley	Business Owner	Sponsor
Yolanda Wilke	CISO	Security Overview/C
Anand Harsh	Cloud Operations	SIDM/Technical Owner/C
Andre James	Domain Specialist / SME	
9(2)(a)	Domain Principal	Commercial Manager
Kevin McArtney	Security Architect	Security Overview
9(2)(a)	START Architect	Consulted
	START Tech Support	Informed
Prajakta Panse	Test Domain Lead	Test Overview
John Gasson, Dawn Sawn	Solicitor, Privacy Officer	Consulted, Consulted
External		
CBS (Canon Business Services)	Vendor	Consultation, Professional services, Support

Appendix

Inland Revenue
Te Tari Taake

POC State

- CBS will connect to IR mailboxes using a supported Exchange Online API
- Emails are ingested into CBS solution, MAP (using existing paper scanning integration layer).
- Exchange Online mailbox access via API is an existing IR access pattern

6(c)

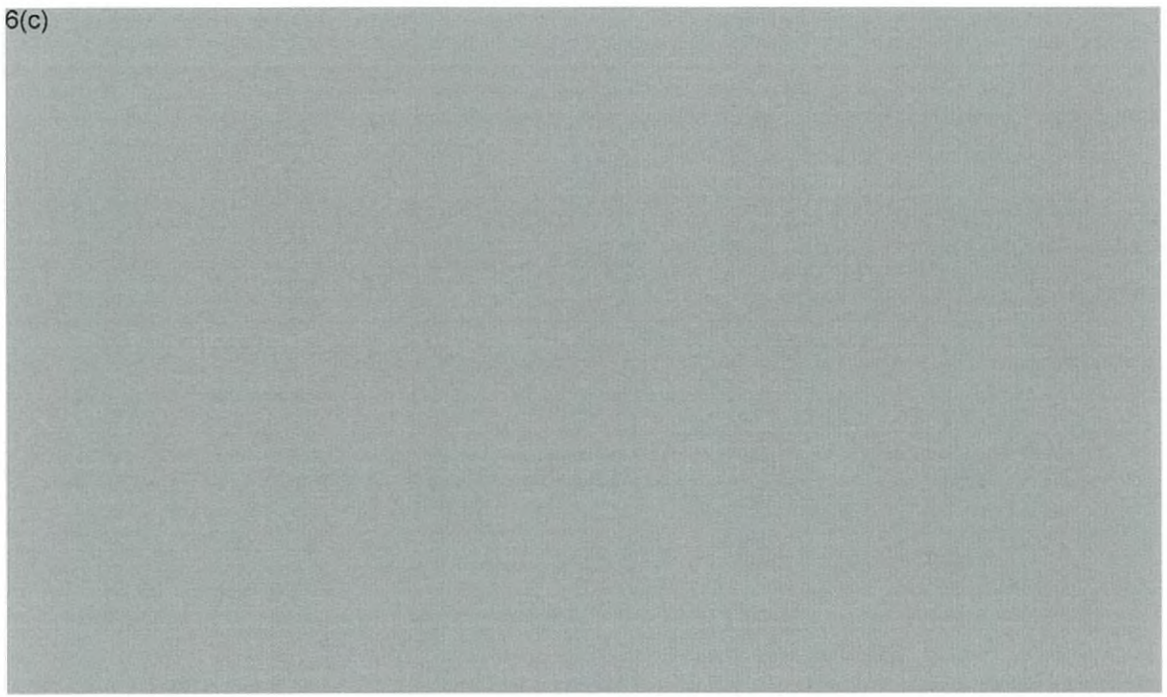


Diagram 1: CBS Email Extraction Process Flow via API – High Level Diagram

Supporting information (Appendix 1)

- Approved PIA

Privacy-Assessment-Canon Email Ingestion PoC Fnal.docx

- Approved Security Risk Update

06B7C8B2.pdf

- TDA Paper (24th June 2025)

DD-1167 Canon Email Ingestion Channel Functionality PoC v1.2.pdf

Item 15

Inland Revenue
Te Tari Taake

DIP Snowflake Cortex AI PoC Evaluation Report

AI Oversight Group

Business Sponsor: Patrick O'Doherty

Authors: Tiago Hubner, Hooi Yee Ng

Date: November 2025

Purpose and recommendations

Purpose

In May 2025, the AI Oversight group approved the DIP Snowflake Cortex AI Proof of Concept (PoC). This report presents the key findings, evaluation, outcomes and lessons learned from the PoC conducted between August and September 2025.

Recommendations

It is recommended that the AI Oversight Group:

- **Endorse** the next steps:
    1. **Move Cortex AI for Audit Outcomes into production, ensuring secure and governed access controls.**
        - Project Team and CCS will identify priority IR audiences who will gain the most value from accessing Audit Outcomes data through Cortex AI
    2. **Project Team to work with CCS-LT on the next additional high-impact data domains that would deliver the greatest business benefits when enabled through Cortex AI, and extend the pilot.**

Executive summary

Problem statement

Business teams rely on data to inform decisions, yet they spend considerable time extracting and reconciling information from existing reports. While the DIP data products have improved efficiency, identifying patterns, risks and outliers still demands significant manual effort. As data volumes and complexity grow, teams depend heavily on developers and analysts with SQL expertise, creating bottlenecks that lead to delays of days or even weeks. This reliance limits business users' ability to independently explore patterns or anomalies in a timely manner, impacting decision-making agility.

Proposed solution

To address these bottlenecks, we initiated a Proof of Concept (PoC) using Snowflake Cortex AI to assess whether GenAI and machine learning can automate or augment data analysis tasks traditionally reliant on SQL expertise. The PoC explored capabilities such as report generation, transaction pattern clustering, and natural language querying to replace manual SQL. Approved by the AI Oversight Group in May 2025, this report summarises key findings and lessons from the PoC conducted between August and September 2025.

Executive summary – key findings from the Cortex AI PoC

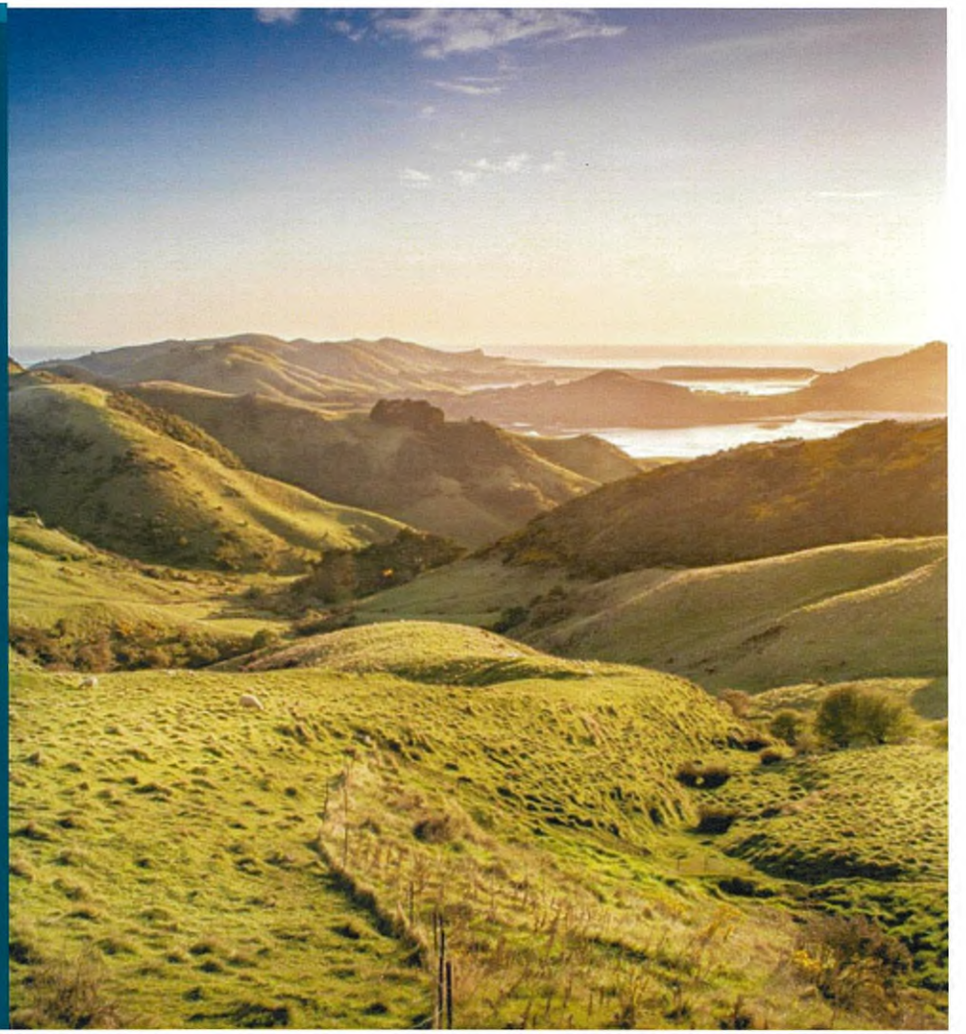
1. **Accuracy and reliability**
    - Natural language queries delivered results that matched source data accurately.
    - A verified question bank (pre-approved questions with corresponding SQL) improved accuracy by giving Cortex AI a reliable reference.
2. **Speed and efficiency**
    - Ad-hoc analysis completed in seconds instead of hours or days.
    - Automatically generated SQL provided a strong foundation for deeper, more complex analysis.
3. **Broader access and usability**
    - Natural language querying reduced reliance on SQL skills.
    - Enable a wider audience within IR to access to DIP data without technical barriers.
4. **Enhance user experience**
    - Training helped users phrase better questions, improving result quality.
    - Users explored data more creatively and iteratively compared to static dashboards.
5. **Semantic model development**
    - Semantic models map business terms to database structures, enabling Cortex AI to interpret intent accurately.
    - Subject Matter Expert (SME) input was essential for building accurate query logic.
    - Continuous user feedback refined the sematic model for better performance.

Inland Revenue
Te Tari Taake

PoC Details

Approach, evaluation and findings

Inland Revenue
Te Tari Taake



Scope of Proof of Concept

Scope

This was an exploratory PoC. The scope of it was to evaluate the capabilities of Snowflake Cortex AI in a controlled, production-grade environment using the existing dataset associated with the Audit Outcomes data product.

The PoC was a focused, non-intrusive exploration designed to assess value, usability and alignment to business functions, particularly in the context of non-compliance.

Key scope elements

- The PoC ran on our production Snowflake instance, but access was tightly controlled with role-based access.
- It focused primarily on one data product: Audit Outcomes.
- Team members were composed from DA&I (Platform Engineer, Data Scientists, Data Analyst) and CCS (Analyst, Customer Compliance Specialist, Leaders).
- The PoC utilised existing, governed datasets. No new data pipelines were required.
- New technologies evaluated in the PoC:
  - **Snowflake Cortex Analyst** - LLM-powered feature that helps business users ask questions in natural language and receive direct answers without writing SQL.
  - **Snowflake Evalanche** – a Streamlit app in Snowflake to evaluate and compare LLM use case outputs in a streamlined, on demand, and automated fashion.

Approach

The PoC was structured to provide a secure, collaborative and low-risk evaluation of Snowflake Cortex AI, focusing on its practical application to the Audit Outcomes data product. The approach ensured minimal disruption to existing systems while demonstrating how AI can deliver meaningful insights, particularly in identifying non-compliance patterns.

- Conducted in **production Snowflake** with **real, governed data**
- Enforced **role-based access controls** for security & compliance

- Used existing **Audit Outcome datasets**
- No new pipelines – focused on **maximizing current data assets**

- Collaborated with stakeholders **across DA&I, SMEs, Performance & Reporting, PD&D**
- Focused on:
  - **Non-compliance pattern detection**
  - **Audit Outcomes summarization**
  - **Natural language exploratory analysis**

- Tested **Cortex Analyst** and **Evalanche** for:
  - Output **accuracy & relevance**
  - **Ease of use** for all user types
  - Workflow **alignment** and **time saving**

- Cross-functional team: **Engineers, Data Scientists, Analysts, Investigators**
- **Iterative** testing and feedback loops

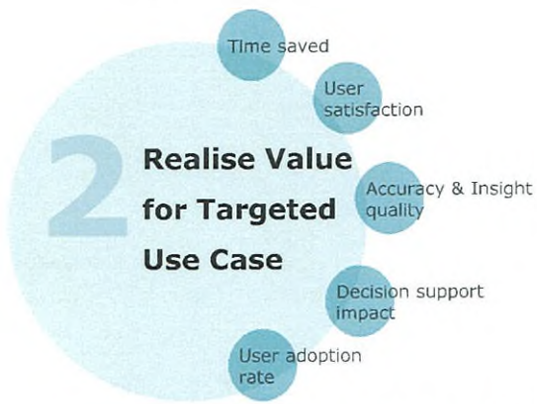
- Captured **technical findings, business value,** and **user feedback**
- Final report presented to governance group with **recommendations**

Objectives of Proof of Concept

The primary objective of this PoC was to learn through hands-on experimentation with Snowflake Cortex AI capabilities.

By the end of the PoC, IR was able to report on the following objectives:

Gain a practical understanding of the feasibility of the use of Cortex AI capabilities such as natural language querying, anomaly detection, and summarization, within the context of our production Snowflake environment and data governance framework.



Assess the applicability and potential value of applying Cortex AI to a singular data product (Audit Outcomes), including its impact on user experience, team efficiency, and decision-making within the relevant business function (for example, compliance activities).

Develop a view of new opportunities enabled by Cortex AI- such as expanded use cases for self service-analytics, risk identification, operational automation, or strategic data access – beyond the initial scope of this PoC.

Evaluation of Objectives

1. Feasibility Assessment: Gain a practical understanding of the feasibility of the use of Cortex AI capabilities such as natural language querying, anomaly detection, and summarization, within the context of our production Snowflake environment and data governance framework.

Measurement	Description	Assessment
Integration Complexity & Time	Complexity & time taken to setup and configure Cortex AI in the Production Snowflake environment	Cortex AI is already available as part of the Snowflake enterprise offering, therefore no additional setup was required.
	Complexity & time taken to setup and configure Cortex AI for a particular domain data product	Creation of a new Audit Outcomes 'Semantic Model' for Cortex Analyst was relatively easy, but prior engagement with business SMEs to understand business context was essential. The model went through various iterations during the PoC to improve response rates and to incorporate user feedback.
	Complexity & time taken to develop a user interface for Cortex AI	For this PoC, a Snowflake Streamlit App (Web Interface) was built to serve as a user interface for natural language queries. This is now superseded by the recently release Snowflake Intelligence 'out of the box' user interface which will provide numerous improvements.
Data Privacy, Security & Governance	Successful integration with existing IR security controls for governed access	Secure access to the Cortex AI model for Audit Outcomes, was easily managed using existing security features, roles and permissions available in the Snowflake Role Based Access Controls already in place with DIP. No governance or access control issues were encountered.
	Security & privacy of IR data maintained	Data was not shared and did not leave the secure IR environment during any phase of the Cortex AI proof of concept.
Feature usability score	Ease of use or performance of: Natural language querying Summarisation Error message – understanding Data exploration	Natural language query interface was very easy to use, provided question verification and output of the generated SQL code. The way in which questions were asked could affect the outcome and show unexpected results. From this it was observed that the testers needed to be specific when querying the data with an iterative learning approach to phrasing their questions correctly.

Evaluation of Objectives

2. Value Realisation for a Targeted Use Case: Assess the applicability and potential value of applying Cortex AI to a singular data product (Audit Outcomes), including its impact on user experience, team efficiency, and decision-making within the relevant business function (for example, compliance activities).

Measurement	Description	Assessment
Time saved	Reduction in time to complete tasks, e.g. query analysis and generation, summarisation, reporting etc. (Cortex AI vs current method)	Speed to insight for ad-hoc analysis and data exploration was greatly reduced (seconds vs. hours/days to request an analyst to write the same query). Seen as a valuable extension to existing methods of exposing DIP data to a wider IR audience. Moving from report updated only once a week/month based on premade visuals and narratives to automatically refreshed everyday, validated and secured, for all possible questions within seconds for questions related to the Audit Outcomes Data Product.
Accuracy or Insight quality	Comparison of AI generated insights vs. manual analysis (as per current method) % or alignment or error rate	Natural language query results were consistently accurate when checked against source data (within the scope of the Audit Outcomes domain data)
	Auditability and tracing of queries to assess quality of data outputs	All submitted questions were fully traceable through audit logs that could then be used to determine the accuracy of SQL generated by the LLMs.
User adoption rate	Number of users both technical and non-technical using Cortex AI for analysis	A total of 14 technical and non-technical users participated in the PoC
User satisfaction	Survey – all participants of Cortex AI PoC	In survey responses, we observed an overall high-satisfaction rate with the Cortex AI PoC
Decision support impact	Number of decisions influenced or accelerated by AI-generated outputs.	Although this was only a Proof-of-Concept, user feedback was very positive around the opportunities that Cortex AI could provide to quickly identify patterns, outliers and anomalies within the Audit Outcomes data, leading to further detailed investigations or process efficiencies. Users were excited about the possibilities of applying Cortex AI to other Data Domains within DIP.

Evaluation of Objectives

3. Identification of New Organisational Opportunities: Develop a view of new opportunities enabled by Cortex AI- such as expanded use cases for self service-analytics, risk identification, operational automation, or strategic data access – beyond the initial scope of this PoC.

Measurement	Description	Assessment
New use cases identified	What new use cases can be identified that would benefit from applying this new technology?	From observing the success of the proof of concept on the Audit Outcomes data held in DIP, there is the potential to build multiple self-service Cortex AI models on existing refined data held in DIP if it meets business demand and benefit. Priority data domains that stand out would be Debt, Returns, Interactions, People and Crypto Currency.
Self-service analytics uptake	Number of users or teams expressing interest in using Cortex AI for self-service	The DIP user base at moment circa around 1000+ users, 95% of those would benefit from Cortex AI for their daily work.
Process improvement opportunities	Count and description of processes that can be automated	Speed to insight, using natural language queries for ad-hoc analysis, was greatly reduced (seconds vs. hours/days to request an analyst to write the same query). Analysts also found an additional and unexpected benefit by taking the generated SQL queries as "base code" to then write their own more complex analyses with quicker speed to insight for their customers.
Requirements improvement opportunities	What can we learn from the user queries captured?	Capture of the natural language queries submitted through Cortex AI has built up an unexpected resource available to us for determining future user requirements for data. What are users asking for - even though the data might not be available yet? How are users wanting to link different data domains together?
Opening up access to data to a wider audience	Does this technology support opening up access to IR data to a wider audience?	Cortex AI has the potential to open up access to IR data and insights to a very wide IR audience that would not have the technical expertise to query the data today.

Navigating governance ambiguity and other key challenges

Data Strategy & Foundation
Data strategy is the backbone of AI Strategy. Validated data products drive trust and engagement.

Audit Outcomes data product stood out due to its thorough validation (system testing, integration testing, UAT, documentation). This led to high trust, accurate insights, and strong user engagement.

Semantic Modeling & Knowledge Representation
Semantic modeling is foundational.

Inconsistent Definitions: Variations across business units hinder semantic modeling and accuracy.
No Centralized Knowledge Graph: Absence of unified concept mapping limits scalability.

Semantic Complexity: Domain-specific terminology requires expert input for robust modeling.
Model Dependency: Accuracy improvements tied to LLM upgrades.

Technology Deployment & Integration
Flexible deployment matters.

Iterative Deployment is Essential: The technology requires continuous monitoring, development, and releases. A one-off deployment is insufficient for sustained success.
Architecture Flexibility: Questions remain about adaptability to other platforms or data ecosystems (e.g., Posit App) without vendor lock-in.

Governance & Security
Governance must balance autonomy and value.

Governance Ambiguity: Unclear boundaries around user access and autonomy led to concerns about productivity vs. noise. While user autonomy fosters exploration and learning, governance must ensure that this leads to productivity and not noise. Decision-making frameworks are needed to prioritize valuable data.

Measurement & Evaluation
Value measure is critical.

Limited ROI Metrics: No standardized or enterprise-wide framework to measure efficiency gains or business impact by new tech deployments.

Resourcing & Skills
Capability building is essential to support organizational readiness for scaling AI

Resource Constraints: Only one fully allocated data engineer created bottlenecks, especially around UAT and final consumer engagement.

Appendix

Inland Revenue
Te Tari Taake

Executive summary

Key Findings

Accurate Results
Natural language queries matched source data reliably (Audit Outcome domain).

Faster Insights
Ad-hoc analysis completed in seconds vs. hours/days.

Wider Reach
Enabled broader IR audience access to Data within DIP.

User Learning
Training improved question phrasing and result quality.

Analyst Boost
Generated SQL served as a base for deeper analysis.

Creative Exploration
Users explored data more iteratively than with static dashboards.

Semantic Layer
SME input was key to guiding accurate query logic.

Question Bank
Verified queries improved accuracy and usability

Iterative Feedback
Continuous user input refined the semantic model

Recommendations

Identify Target IR Audiences

who would benefit the most from making the Audit Outcomes data available through Cortex AI

Productionise PoC

Make Audit Outcomes data available securely through Cortex AI

Identify Next Domains

Prioritize other high-value data domains for Cortex AI enablement

Evaluating against IR criteria

Evaluation criteria	Definition	Assessment
Our strategy	Impact on the way IR delivers against its strategic objectives	Efficiency and Effectiveness – Cortex AI reduced time-to-insight dramatically, from hours or days to seconds, by enabling natural language queries and minimising reliance on SQL-skilled analysts. This automation of repetitive tasks, such as report generation, frees skilled resources to focus on higher-value strategic work.
Impact	Ability to drive positive business impact for IR staff and customers	Cortex AI enables non-technical staff to analyse and explore data without relying on SQL, allowing for faster, more iterative and creative insights that support timely, informed decision-making.
Technology	Ability to implement related to available technical infrastructure	Cortex AI is already available as part of the Snowflake enterprise offering, minimal setup effort was required.
Governance	Ability to manage related risks, including legal, ethical, security and other.	Cortex AI is part of the Snowflake platform and operates within existing security and governance frameworks. However, additional governance measures are needed to ensure responsible use of Cortex AI and the reliability of its outputs.
How we work	Ability to implement related to available staff capability and capacity, and if not, having the right support in place to uplift capability if required.	Ongoing management of Cortex AI as an additional tool within DIP must be considered, including sustainable development and maintenance. A key requirement is maintaining a consistent semantic model across IR data domains, as this is critical to Cortex AI's accuracy and overall performance.

KPI Metrics

HUMAN EVALUATION

93%

Of submitted questions didn't receive a thumbs-down

13 Participants
479 Questions submitted
35 Thumbs down

The volume of questions and feedback shows **active exploration** and **willingness to test boundaries.**

Feedback was **voluntary** and represents only a **subset of user experiences.**

Thumbs-down feedback (35) provides **valuable signals** for improvement and refinement, including unanswerable or out-of-scope questions.

AUTOMATE EVALUATION

100%

SQL Results Accuracy
(increasing from 92%)

By Snowflake Evalanche:

19 Question bank

Percentage of times automated evaluator confirms the generated SQL correctness across repeated test cycles on the same question bank after model improvements.

Note: Continuous consistency monitoring are required to ensure the model remains reliable as query complexity varied and evolve.

SURVEY FEEDBACK

Positive

Results from all respondents
(1 Compliance Analyst, 8 Analysts, 4 Others)

Rated positive to very positive on:

- Ease of interaction
- Satisfaction with responses
- Future use
- Adequacy of introduction
- "About Me" page usefulness
- Would recommend to others

Top key features driving user satisfaction(extracted from verbatim):

Significant time savings* - several hours to a few seconds for the SQL creation using AI vs. manual effort.

Efficiency gain - faster query creation, understanding data structure, improving workflow, decision-making, real world applications.

*Time savings estimates are based on feedback of 8 Analysts. Actual savings may vary depending on query complexity and experience. Future evaluation plans may need to include time tracking and controlled comparison to quantify AI tool ROI.

Feedback Messages

Areas of feedback:

- **Functional Accuracy**: returning correct results without manual fixes
- **Consistency**: repeated queries producing identical answers
- **Error Rate**: some queries failing due to SQL or compilation errors
- **Formatting Quality**: layout, timestamp, or display issues
- **Data Coverage**: scope is well-defined, potential data points can be covered

The audit groups are **correct** and the notional discrepancies match the group papers

The subtotals that were put into the table seem **correct** though

include a synonym for Significant Enterprises as SE. party_segment_description = 'SE'

restricted **customers** not limited to Special Files

It looks pretty good. With the **hours**, are we able to get a breakdown of the hours for a certain period, e.g. **audit hours** for the August month or the quarter of July to September, instead of all **hours** recorded in the audit group from the beginning of the audit?

Because of an **outlier** in the 2026 tax year all amounts have to **formatted** as '$999,999,999,999,999'. If this is not done the **outlier** amounts are shown as $###,###,###,###. The totals are also shown as $###,###,###,###