Headquarters
New Zealand Defence Force
Defence House
Private Bag 39997
Wellington Mail Centre
Lower Hutt 5045
New Zealand

OIA-2026-5813

28th May 2026

[redacted]

Dear [redacted]

I refer to your email of 21 May 2026 requesting, under the Official Information Act 1982 (OIA):

- *What Artificial Intelligence is your Ministry considering implementing in the next year?*
- *What aspects or parts of the business will it be implemented? Ie how will the AI be used*
- *Will it's implementation reduce the workload or staff numbers?*
- *What safety and security protocols has it had to pass to be considered acceptable for implementation?*

The only artificial intelligence (AI) tool approved for use by all New Zealand Defence Force (NZDF) personnel and staff is Copilot Chat, a Microsoft 365 application. The following AI enabled tools are used in a restricted capacity for research purposes in data processing, sensor processing, and in support of modelling activities:

- ChatGPT
- Dalle-2
- Github Copilot
- Azure Machine learning services
- Azure OpenAI services
- Microsoft Copilot
- Microsoft teams - transcription and summarisation
- AiZynthFinder
- Meta Llama2

The use of Copilot Chat and these tools augments the capability of NZDF personnel and staff, not replace them. Enclosed is a copy of the high-level risk assessment for Copilot Chat. Where indicated the names of those who have provided advice are withheld in order to maintain the effective conduct of public affairs in accordance with section 9(2)(g)(i) of the OIA. Research risk assessments are withheld in accordance with section 6(a) of the OIA.

You have the right, under section 28(3) of the OIA, to ask an Ombudsman to review this response to your request. Information about how to make a complaint is available at www.ombudsman.parliament.nz or freephone 0800 802 602.

Please note that responses to official information requests are proactively released where possible. This response to your request will be published shortly on the NZDF website, with your personal information removed.

Yours sincerely

GA Motley
Brigadier
Chief of Staff HQNZDF

Enclosure:
1. High-level risk assessment

High-level Risk Assessment:
M365 Copilot Chat (Free Version)

Reviewer(s)	s. 9(2)(g)(i) **DAIM-D (D&A Team)** DIS Cyber DIS Advisory
Date of Review	**May 28, 2025**

A. Background

Microsoft Copilot (free version) is a generative AI assistance tool built on Microsoft's secure enterprise infrastructure. This assessment evaluates the risks associated with deployment and use of Copilot Chat by members of the NZDF and contractors for NZDF purposes. This assessment explicitly focuses on the free version of Copilot that predominantly uses publicly accessible internet content and Microsoft-hosted Large Language Models (LLMs), while providing limited and controlled interactions with organisational data through user-uploaded documents stored securely in OneDrive for Business.

B. Immediate Security Concerns and Clarifications

1. Model perimeter

i. **Note:** The model perimeter is cited as excluding all internal organisational data sources (e.g., Outlook, Teams, SharePoint, OneDrive, and local drives). The free version of Copilot explicitly does not integrate with Microsoft Graph[1] and ensures that **prompts and responses remain within the M365 tenant**, avoiding external visibility or storage (see point (3) below).

2. Aggregation and use of prompts

i. **Note:** According to the official Microsoft documentation on Copilot Chat's enterprise privacy and protections, prompts and model responses used within Copilot Chat are **neither visible nor accessible to Microsoft or third parties**.

- When users interact with Copilot Chat, their prompts and responses are logged and stored within the M365 boundary for internal auditing (see 4.iii below). Organisational (NZDF) M365 administrators (i.e., admins within the service boundary) can view anonymised and deanonymised usage metrics (e.g., total active users, average daily active users, and activity per application).
- Microsoft staff only access aggregate and anonymised service-level metrics for operational purposes and assure that they do not have access to organisational data as per their Copilot Chat Privacy and Protections and Enterprise Data Protection (EDP) policies.

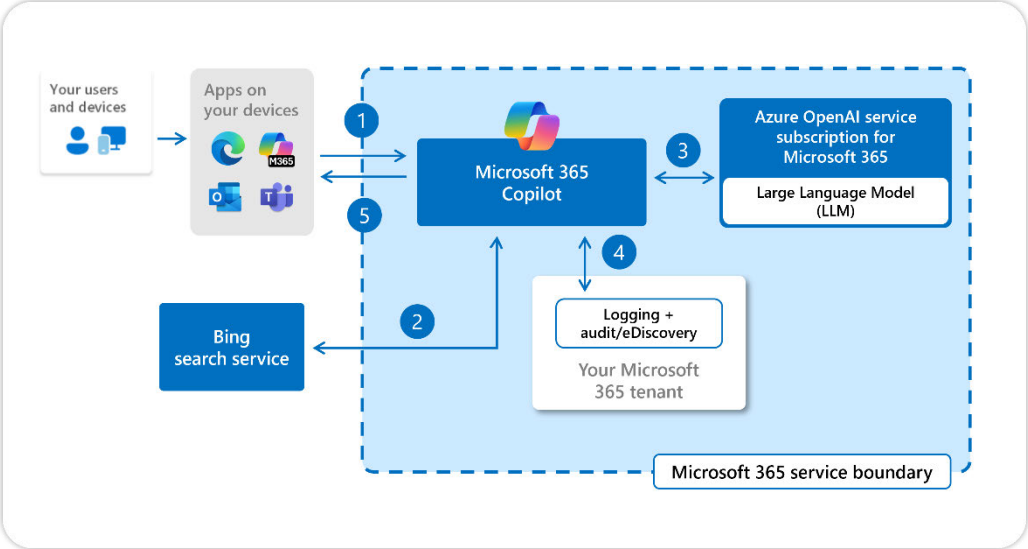
[1] Microsoft Graph provides Application Programming Interfaces (APIs) to enable access to data and services within an organisation's Tenant (e.g., an API to query data from SharePoint).

*(Refer to: *"M365 Copilot Privacy and Protections"* Microsoft Documentation).*

ii. **Note:** Microsoft explicitly states that prompts and responses **are not stored or used to train foundation models**, including those maintained by Microsoft or OpenAI. Further, when enabled, **web search queries are anonymised and stripped of user and tenant identifiers** prior to external transmission (e.g., Bing queried) via a process referred to as **"(web/prompt) grounding"** *(Step 2, Figure A).*

- That is, when a user submits a prompt, both the query (prompt) and response remain within the service boundary. The response a user receives is formulated using publicly available data and the LLMs themselves. There is no reaching into NZDF data/documents, so classified queries would be answered in terms of what is publicly available or presumed by the model. Prompt grounding, which Microsoft implements by default, means any data leaving the service boundary does NOT include:
    a. The users entire prompt (key words are used only);
    b. Files uploaded from Onedrive;
    c. Web pages or PDFs summarised by Copilot Chat;
    d. User or tenant identifiers.

iii. Figure A: M365 Copilot Chat Architecture Diagram *("M365 Copilot Privacy and Protections").*



3. Data access within DIE boundaries

i. **Note:** The free version of Copilot Chat does not integrate with Microsoft Graph and thus **cannot autonomously access or interact** with any documents stored within the Defence Information Environment (DIE). That is, Copilot Chat responses consist of publicly available web content and LLM-generated responses, bar user-uploaded data/documents (see point (4) below).

4. User-uploaded data & documents

i. **Note:** Users can upload data/documents in the free version of M365 Copilot Chat, provided they have access to OneDrive for Business, **which NZDF possesses**. However, **uploaded files are securely stored within the user's OneDrive for Business account inside of a 'Microsoft Copilot Chat Files' folder**, which ensures that any uploaded content remains within the enterprise's M365 tenant and is governed by existing data protection policies. Therefore, the risk of data exposure or unauthorised access remains minimal, provided OneDrive governance and security controls are sufficiently managed.

ii. **Note:** Copilot Pages is accessible to users with a Microsoft Entra ID (formerly Azure active dir.) and either SharePoint or OneDrive storage and ***does not*** require a paid M365 Copilot license. Any content created by Copilot Pages is stored in a user-owned SharePoint Embedded container (one per user). Copilot pages reside in the ecosystem as *.loop* files. These can extend user permissions to others via file sharing settings. However, **conditional access can be enabled to fully block users from opening *.loop* files**.

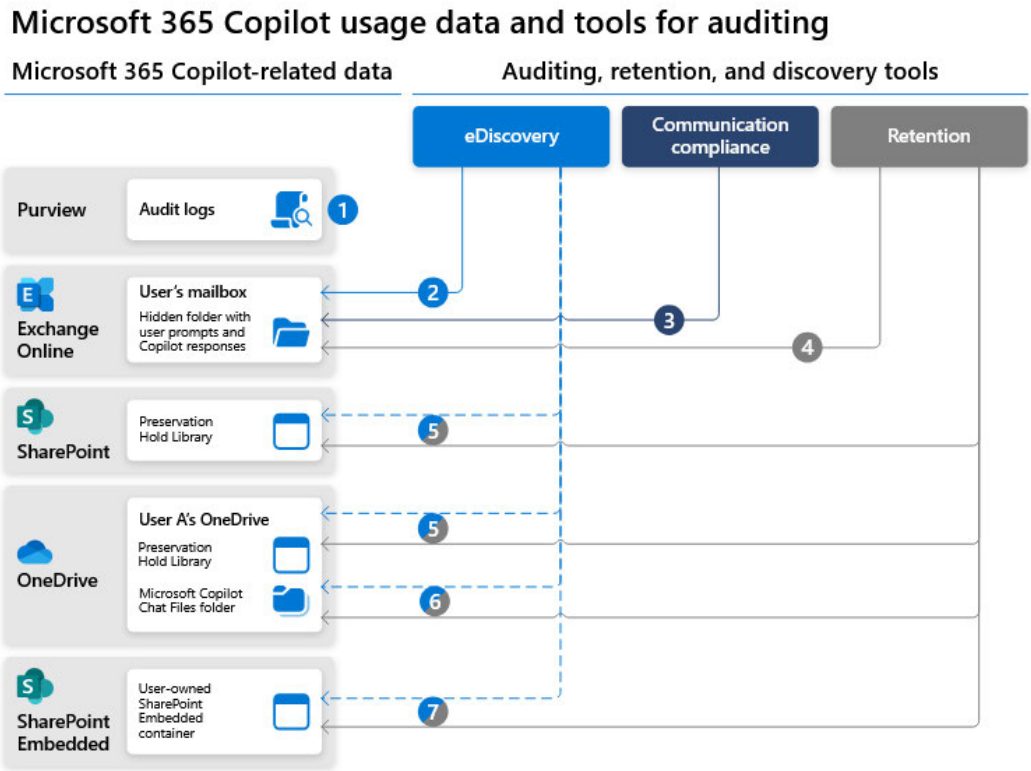
**(Refer to: "Copilot pages for IT Admins – Sep 2024 update", M365 Copilot Blog).*

iii. **Note:** Out of scope for this assessment, but it is possible to explore E5 license options for auditable Copilot trails consisting of M365 tools for auditing, retention, and discovery (Purview → Exchange Online → SharePoint → OneDrive → SharePoint Embedded [Container], *Figure B*). However, the **free version of Copilot Chat still includes some of the foundational capabilities of the Copilot Control System**.

- *[Microsoft] Purview audit logs can be used to identify how, when, and where Copilot interactions occurred, which items were accessed, and the sensitivity labels of those items. Purview eDiscovery can search for keywords in Copilot prompts and responses to identify "inappropriate, malicious, or risky behaviour". Purview Communication Compliance can detect and alert "…", like personal data or confidential information. Purview Retention Policies can maintain a copy of deleted Copilot conversations. (Paraphrased from MS Documentation).*

**(Refer to: "Where Copilot usage data is stored and how you can audit it", Microsoft Documentation).*

iv. ***Figure B: M365 Copilot usage data and tools for auditing*** *("Where Copilot usage data is stored and how you can audit it").*



5. Data sovereignty and regulatory compliance

i. **Note:** Microsoft defines their role as a **data processor and not a controller** under the Data Protection Addendum (DPA), thus NZDF retains full ownership and control over NZDF data *("M365 Copilot Privacy and Protections", Microsoft Documentation).*

ii. **Note:** The free version of Copilot Chat is built on the same secure foundation as M365 and adheres to Microsoft's enterprise-grade security, compliance, and privacy standards, including **ISO/IEC 42001:2023 certification** (first international standard for AI management systems), **GDPR compliance** and alignment with Microsoft's Responsible AI principles, and **Data residency and encryption** within Microsoft's trusted cloud infrastructure *("ISO/IEC 42001:2023 Artificial intelligence management system", Microsoft Documentation).*

C. Aggregate (High-level) Risk Assessment

1. Risk Identification

i. **Sources:** Public web content, hosted LLMs.

ii. **Events:** Data breaches, unauthorised access, or inadvertent disclosure of sensitive information.

iii. **Vulnerabilities:** Reliance on external, publicly accessible data, potential misuse by end users, and unclear boundaries for anonymisation (grounding).

iv. **Timeframe:** Immediate/ongoing (continuous use). Continued monitoring required.

v. ***Table A: Copilot Chat Aggregate (High-level) Risk Assessment***

Probability (Likelihood)	***Impact (Severity)***	***Exposure***	***Velocity (Speed of Onset)***
Low → Mod	**High**	**Low**	**High**
Microsoft is a trusted partner with stringent security controls, but low residual risks remain (external data sources).	Potential for severe impacts **if data sovereignty, access, privacy, or governance** were compromised.	Minimal exposure to organisational data; moderate exposure through reliance on public content.	Speed of onset is immediate if the breach occurred as a result of real-time service interactions.

2. Risk Mitigations

i. **Preventative Measures:**

a. Data security/loss prevention;

b. Data encryption and residency (i.e., data remains within the boundaries of the NZDF tenant and trusted cloud infrastructure – *true by default **and** avoid autonomous tenant-specific data integrations [e.g., MS Graph] without due investigation*);

c. Compare against existing NZDF configured security controls for data loss prevention of OneDrive;

d. Anonymisation and constraints exercised when copilot accesses external applications or search engines;

e. Appropriate use and change management plans;

f. Policy, training and awareness around responsible use and best-practice prompt engineering for copilot and validation/checking outputs from copilot will be required to support personnel to avoid 'garbage in garbage out' scenarios.

ii. **Detective Measures:** Prompt/response visibility controls and stringent auditing *(– limited free license capabilities with the ability to mature over time).*

iii. **Response Measures:** Copilot control system and Enterprise Data Protection (EDP) *(– limited free license capabilities with the ability to mature over time).* Implement more robust governance and administrative controls *(– would ordinarily possess alignment with AI/RAI Governance and Use policies).*

iv. **Note: Low remaining residual risk**, particularly with respect to anonymisation process and reliance on public content.

D. Final Review and Recommendations

The reviewers of this risk assessment conclude that **Microsoft Copilot Chat (free version) presents a low to moderate risk profile within the DIE and NZDF tenant boundary *and* that more rigorous and clear governance, ownership, and mitigation strategies should be in place and validated as soon as possible to monitor and prevent the risk profile from escalating.**

Further, it would be prudent to distribute internally authored general usage guidelines for Copilot Chat and update existing NZDF guidance regarding GenAI. Specifically, it is recommended to provide a programme of training and education to ensure secure and responsible use of Copilot Chat in the immediate future.

Copilot Chat presents itself as an especially valuable and secure tool within the wider M365 ecosystem, but business ownership, active governance, continuous monitoring, and user compliance training are necessary to ensure the safe uptake and assimilation of the product on an ongoing basis.

Note: The limitations of this review include the following…

i. This paper was written largely by consulting information that is publicly available and published online by Microsoft.

ii. The NZDF tenant may have different configurations than an 'ideal' tenant. As such, NZDF should seek to validate the native auditing, compliance, and security controls before considering further action.

Acceptance of Notes, Review, and Recommendations

CDO	**Date:**
CISO	**Date:**