MARLUC

CASE STUDY · MARLUC SYNOPSIS

Australian Treasury Copilot trial

IMPLEMENTED 2024 · GOVERNMENT AND PUBLIC SERVICES · AUSTRALIA · BAND D · PARTLY PROVED

VERSION 1.0 · PUBLISHED 15 NOVEMBER 2026 · UPDATED 15 NOVEMBER 2026 · FACTS CHECKED 15 NOVEMBER 2026

Treasury gave 218 staff Microsoft 365 Copilot for 14 weeks in 2024. Its published evaluation found help with basic administrative tasks, lower use than expected, and no clear evidence of better work outcomes.

ABOUT THIS SYNOPSIS

This is MarLuc's summary and grading of publicly available sources, for information only. It is not investment, legal, financial or professional advice, and should not be relied on for any decision. Readers must validate it against the original sources and, where appropriate, challenge its conclusions.

Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date. Check the original source before relying on any figure.

The impact band, proof marker and SEAM view are MarLuc's opinion. The organisation is not a MarLuc client and has not reviewed or endorsed this synopsis. Names and trademarks belong to their owners.

NOTICE WORDING LAST UPDATED 16 NOVEMBER 2026

AT A GLANCE

AI used: Generative AI assistant: Microsoft 365 Copilot.

Implemented: 2024. Grade: band D, Task support, Partly proved.

Headline result, as reported: Help with basic administrative tasks; no clear evidence of better work outcomes, per Treasury's own evaluation

1 · The case

The facts, as reported by the 4 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

Australian Government Department of the Treasury. Australian central government department. 218 staff took part in the trial; Treasury procured and deployed 275 Microsoft 365 Copilot licences in total. Canvas block: 5. People and skills. Also 4. Operations and process, 8. Governance and risk.

Business challenge

The challenge. In November 2023 the Prime Minister announced a six-month Australian Government trial of Microsoft 365 Copilot, coordinated by the Digital Transformation Agency. The trial set out to test staff sentiment, productivity, safe and responsible adoption and unintended consequences. Treasury took part on a later timetable because of the 2024-25 Budget.

Who implemented it

Microsoft (Software vendor): Supplied the assistant.

Technology (SEAM pillar E)

SEAM types: T2 Embedded AI in existing software. Deployment: Vendor SaaS. Vendors and products: Microsoft, Microsoft 365 Copilot.

Where it was built in: Inside Microsoft Office products, including Word, PowerPoint, Excel, Outlook and Teams, within Treasury's protected IT environment and limited to its internal systems.

What changed

Before: staff worked in Microsoft Office without a generative AI assistant. After: participants used Copilot for summaries, meeting minutes, finding information and first drafts; most used it two to three times a week or less, and the evaluation found no evidence of improved workflows. [1] [2] [3] [4]

What it cost

Licence and trial costs: Not reported. Treasury procured and deployed 275 licences. The evaluation estimated that a mid-level (APS6) staff member would need to redirect about 13 minutes a week to higher value tasks to offset the licence cost.

What it saved

Time saved was not measured. In the post-trial survey, 61 per cent said Copilot reduced time on low-value or low-priority tasks and 49 per cent agreed it improved the speed of completing tasks, against 85 per cent who expected that before the trial. Among responding managers, 80 per cent reported no impact on the timeliness of their team's work and 59 per cent no impact on its efficiency. [1] [2] [3] [4]

Governance and control

Treasury required a permissions audit, mandatory training with a knowledge test and agreement to policies before use, and the use cases stressed checking outputs for errors. A project team and an AI Working Group ran the trial. The evaluation found there were no clear guidelines on when and how to disclose Copilot use, and that onboarding was limited by trial resources.

S2 · Accountable owner · Claimed

The report describes a Copilot trial project team and a Treasury AI Working Group responsible for implementation, policy and evaluation; no named senior owner is reported. [1]

A1 · Safe use · Claimed
Participants had to pass a mandatory training module and agree to follow relevant policies and guidance; the evaluation found there were no clear guidelines on when and how to disclose Copilot use. [1]

A2 · Human in the loop · Claimed
All use cases emphasised reviewing outputs for errors and ensuring accuracy, by the report's account. [1]

A3 · Monitoring and assurance · Claimed
Treasury's Australian Centre for Evaluation ran surveys, focus groups and an issues log and published the results; it describes the work as an internal evaluation. [1] [2]

A4 · Vendor AI · Claimed
Copilot was limited to Treasury's internal systems under its privacy and security restrictions, which the evaluation says limited the product's quality. [1]

A5 · Agent controls · Not reported
No agents described.

A6 · Standards and regulation · Claimed
The trial was part of the whole-of-government trial coordinated by the Digital Transformation Agency; the report says the agency's draft policy on responsible use of AI was not available to participants during the trial. [1]

What worked, what did not, what is unproven

WORKED

– Basic administrative tasks: finding and summarising information, meeting minutes, knowledge management and drafting content.

– 63 per cent of post-trial survey respondents reported some positive impact on work processes.

– An unexpected benefit for accessibility and inclusion, including for neurodivergent and part-time staff.

– The technical implementation was smooth, with relatively few issues.

DID NOT

– Use was lower than expected: most participants used Copilot two to three times a week or less.

– 59 per cent said Copilot helped with little to none of their weekly workload.

– Copilot was judged unsuitable for more complex tasks, and some participants reported inaccurate or fictional output.

– One in five post-trial respondents reported negative sentiment towards using Copilot at Treasury; none had before the trial.

– Training was limited and many support resources were underused.

UNPROVEN

– Time savings: the evaluation says the data collected did not quantify them, while judging them likely to offset licence costs.

– Work outcomes: the evaluation did not find clear evidence that Copilot improved them during the trial.

– Whether Treasury went on to a wider rollout: Not reported.

Timeline

Nov 2023 Prime Minister announces a six-month government trial of Microsoft 365 Copilot [1]

May 2024 Treasury's trial starts for 218 participants [1]

Aug 2024 Formal trial period closes after 14 weeks [1]

2 · Grade and SEAM assessment

MarLuc's opinion, made on the evidence above. How cases are graded: marluc.co.uk/case-studies/how-we-rate

Level of change and impact

Claimed: Not applicable. Grade: D, Task support. Proof: Partly proved. Level of change: Task support.

Task support (Level 1, band D). The evaluation rated progress on improved workflows and new approaches to problems as "No evidence of progress".

Participants reported that Copilot helped with basic administrative tasks, and most managers saw no effect on their teams' timeliness or efficiency. The evaluation did not measure time saved and found no evidence of improved workflows, so this is task support with the benefit partly evidenced.

The SEAM view and lessons learnt

The tool worked technically and helped individuals with routine tasks. The gaps sat between the technology and the people using it: expectations were set high, training was thin, and security restrictions limited what the product could do, by the evaluation's account. For a mid-sized or larger organisation, the transferable point is that licences alone change little; use cases, training and a way to measure time saved need to be in place first.

1. Give licences to staff with a clear use case and set realistic expectations.
2. Fund role-specific training and count staff time as part of the cost.
3. Measure time saved directly and include a comparison group.

Evidence

Score 2 of 4. Official evaluation, published: an internal evaluation by the Australian Centre for Evaluation, which sits within Treasury. It lists its own limits: a 14-week trial, voluntary self-reported data, volunteer participants and no robust counterfactual.

Evaluation method. Official evaluation, no comparison group: pre-trial and post-trial surveys of participants, pulse surveys, a manager survey, focus groups, case studies and an issues log; self-reported, with no counterfactual.

Sources disagree. The report's table of methods gives 136 post-trial survey responses (59 per cent); its appendix says 164 respondents completed the survey (71 per cent), with 150 complete and verifiable responses. iTnews reports that the Digital Transformation Agency's wider evaluation suggested an average time saving of around an hour a day, a far more optimistic figure than Treasury's.

Status. Trial closed 23 August 2024; participants kept their licences unless they indicated

3 · Original sources

Read these in full. To challenge a fact or a grade, use the contact form at marluc.co.uk or the link on https://marluc.co.uk/case-studies/australian-treasury-copilot

[1] Australian Centre for Evaluation: Evaluation of a trial of generative artificial intelligence (Copilot) in The Department of the Treasury, evaluation report (February 2025)
https://evaluation.treasury.gov.au/sites/evaluation.treasury.gov.au/files/2025-02/evaluation-generative-artificial-intelligence.pdf

[2] Australian Centre for Evaluation: Evaluation of a trial of generative AI (Copilot) in The Treasury, publication page
https://evaluation.treasury.gov.au/publications/evaluation-generative-artificial-intelligence

[3] iTnews: Treasury M365 Copilot review estimates 13-minute efficiency gain needed to justify licence cost
https://www.itnews.com.au/news/treasury-m365-copilot-review-estimates-13-minute-efficiency-gain-needed-to-justify-licence-cost-614905

[4] The Register: After Copilot trial, government staff rated Microsoft's AI less useful than expected
https://www.theregister.com/2025/02/12/australian_treasury_copilot_pilot_assessment

Australian Government Department of the Treasury website: https://treasury.gov.au

Version 1.0 · Published 15 November 2026 · Updated 15 November 2026 · Facts checked against sources 15 November 2026 · Downloaded 26 November 2026. This PDF may be out of date: the latest version, and a way to report an error or a better source, is at https://marluc.co.uk/case-studies/australian-treasury-copilot

VERSION HISTORY

v1.0 15 November 2026 First published