

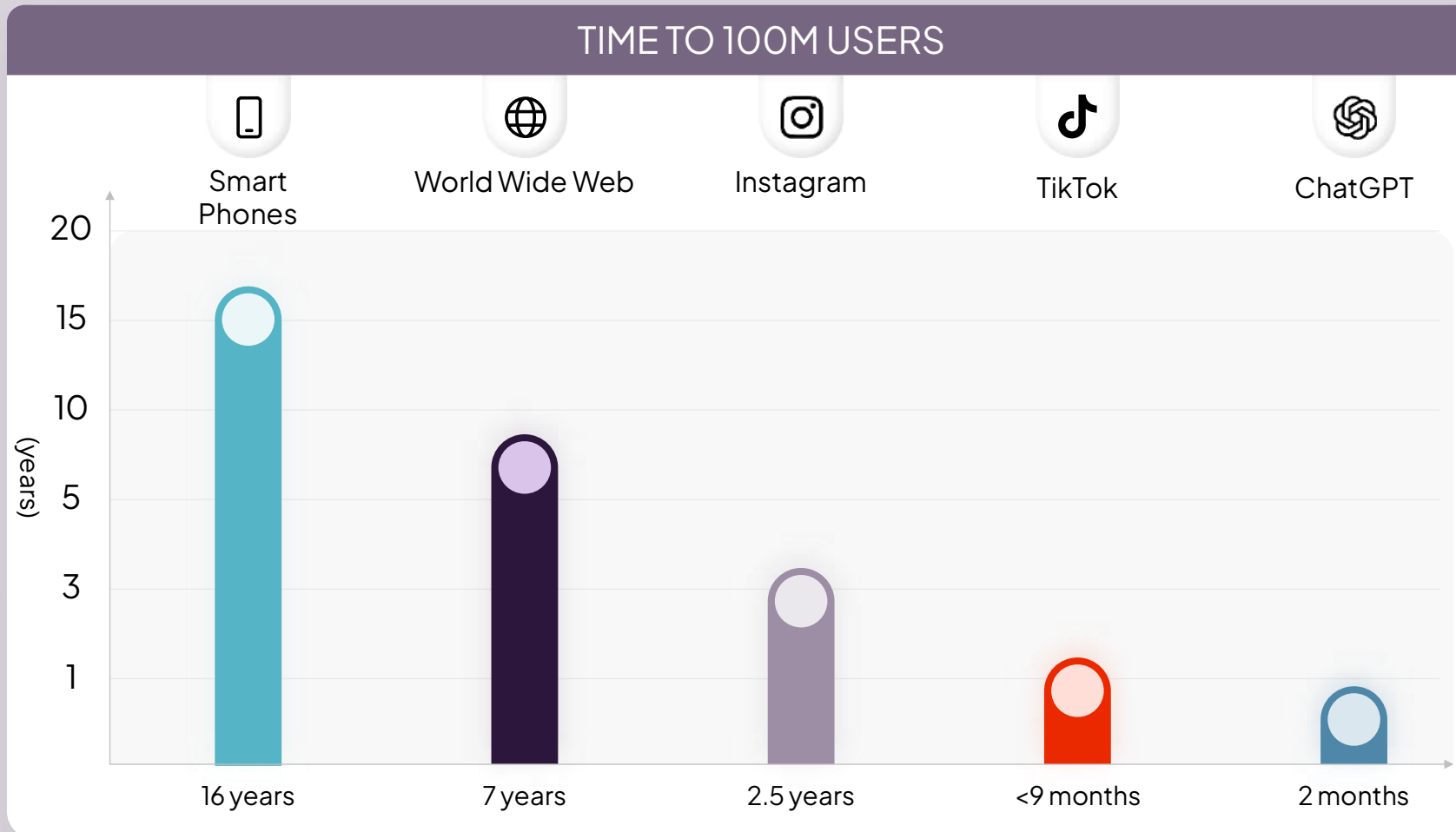
A man in a striped shirt is pointing at a whiteboard with sticky notes in a meeting. The whiteboard has various diagrams and notes, including the word 'Process' and 'Plan'. Other people are visible in the background, some looking at the whiteboard. The scene is brightly lit, suggesting an office or meeting room environment.

# AI, security, risk & governance: innovation with accountability

Presented by Jayden Liddelow  
Cloud, Infrastructure & Security

fusion5

# AI uptake is redefining the innovation curve



And can help...



Unleash creativity



Unlock productivity



Uplevel skills

# In the 90s, the internet didn't come without it's worries...

- Fear of personal information being exposed online
- Difficulty distinguishing credible sources from fake or misleading content
- Viruses, hacking, fraud
- Job displacement fears



# The duality of AI

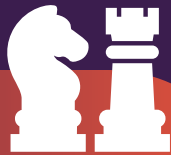
## AI as a Catalyst



Productivity



Personalisation



Strategic

## AI as a Risk Vector



Bias



Data Leakage



Reputational Harm

# Beyond the Hype: Facing the Realities

## False matches 'more likely' for people of colour

Last week Privacy Commissioner Michael Webster launched an inquiry into [REDACTED] facial recognition trial to monitor whether it complied with the Privacy Act and was effective in reducing harmful behaviour at those supermarkets.

In a written statement, a spokesperson from the Commissioner's office said he was concerned about "bias and accuracy" and facial recognition technology (FRT) not being "a proven tool" to reduce harmful behaviour in supermarkets.

(<https://www.rnz.co.nz/news/national/508613/privacy-commissioner-keeps-close-eye-on-supermarkets-facial-recognition-trial>)

"Global evaluations of even the most accurate FRT software show that false matches are more likely to happen for people of colour, particularly women of colour."

The software was also not trained on New Zealand's population and he did not want to see people falsely accused or incorrectly banned from their local supermarket.



The screenshot shows a news article from the Financial Review. The title is "Deloitte to refund government, admits using AI in \$440k report". The authors are Edmund Tadros and Paul Karp. The article is dated Oct 5, 2025, at 7:41pm. It includes a "Listen to this article" button with a 5 min duration. The main text states: "Deloitte Australia will issue a partial refund to the federal government after admitting that artificial intelligence had been used in the creation of a \$440,000 report littered with errors including three nonexistent academic references and a made-up quote from a Federal Court judgement." A second paragraph begins: "A new version of the report for the Department of Employment and Workplace Relations (DEWR) was quietly uploaded to the department's website on Friday, ahead of a long weekend across much of Australia. It features more than a dozen deletions of nonexistent references and footnotes, a rewritten reference list, and corrections to multiple typographic errors."

# Beyond the Hype: Facing the Realities

## Selected deletions from the revised Deloitte report for DEWR\*

Lisa Burton Crawford, *The Rule of Law and Administrative Justice in the Welfare State: A Study of Centrelink* (Federation Press, 2021) [various pages]

**10 references deleted - publication doesn't exist**

Lisa Burton Crawford, *The Rule of Law and Administrative Discretion* (Federation Press, 2021) 112-115.

**Two references deleted - publication doesn't exist**

Notably, her Honour Justice Davis stated at [25]-[26]:  
The burden rests on the decision-maker to be satisfied on the evidence that the debt is owed. A person's statutory entitlements cannot lawfully be reduced based on an assumption unsupported by evidence.

**No such quote exists and the ruling does not have those paragraph numbers. It is Justice Davies, not Justice Davis**

Björn Regnell et al, 'Exploratory Case Study on Release Planning in a Market-Driven Software Company' (2001) *Journal of Systems and Software* 66(1) 37, 39-40.

**Two references deleted - publication doesn't exist**

Other changes include citation updates, a rewritten reference list in appendix H and corrected typographical errors. The new report is three pages longer at 237 pages.

\* Department of Employment and Workplace Relations, *Jul 4 v Sep 26* version

SOURCE: DELOITTE

## NSW flood victims' personal details loaded to ChatGPT in major data breach

By Cathy Adams and Emma Rennie

ABC North Coast

Personal Data Collection Policy

Mon 6 Oct



The data of up to 3,000 Resilient Homes Program applicants has been uploaded to an AI program. (ABC North Coast: Bronwyn Herbert)

# Beyond the Hype: Facing the Realities

**BREAKING** | BUSINESS

## Samsung Bans ChatGPT Among Employees After Sensitive Code Leak

By [Siladitya Ray](#), Forbes Staff. Siladitya Ray is a New Delhi-based Forbes news...

[Follow Author](#)

Published May 02, 2023, 07:17am EDT, Updated May 02, 2023, 07:31am EDT

## Asana admits one of its AI features might have exposed your data to other users

**News** By [Sead Fadilpašić](#) published 19 June 2025

A bug in a newly introduced Asana tool was leaking data for a month

# GenAI attack surfaces introduce new and amplified risks

## GenAI new and amplified risks

Data leakage

Jailbreak

Indirect prompt injection

Model vulnerability

Hallucinations

## GenAI new attack surfaces

Prompts

Responses

AI orchestration

Training data

RAG data

Models

Plugins/skills

## Your threat vectors

Application

Identity

Endpoints

Network

Data

Cloud

# Foundational principles of Trustworthy AI



## Lawful

Complying with all applicable laws and regulations

Ensuring AI systems are developed and used in full compliance with relevant laws, regulations, and standards

EU and US regulations are in force – Australia will follow suit



## Ethical

Aligning AI systems with fundamental ethical values and societal expectations

Key ethical principles include respect for human autonomy, prevention of harm, fairness and non-discrimination, transparency and explainability, privacy, and accountability



## Robust

Ensuring AI systems are technically reliable and resilient to avoid unintended harm

Important aspects include reliability and validity, safety, and security and resilience

Observability and threat detection ensure ongoing operations and usage are monitored for points of failure

# Establishing the Foundations and Governance framework

- Data management
- Data quality management
- Data policies and standards
- Stewardship and accountability



# Accountability

- Establish accountability for systems and outcomes
- Leaders need to set guidelines for responsible use
- Develop AI governance charter
- Proactively review use of AI applications



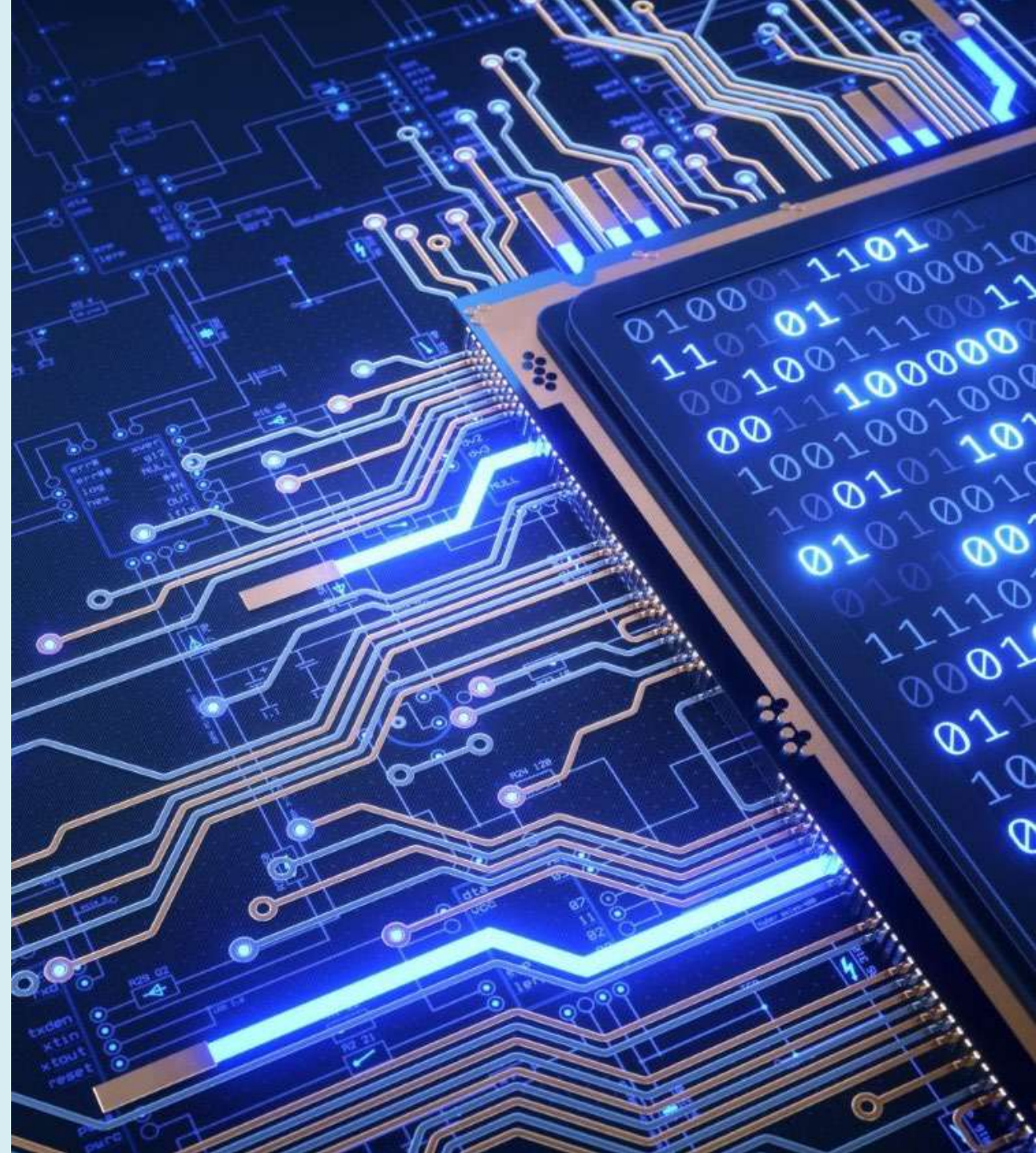
# Human Oversight

- Retain the human in the loop
- Empower employees on AI principals
- Implement AI champions
- Create channels for employee feedback

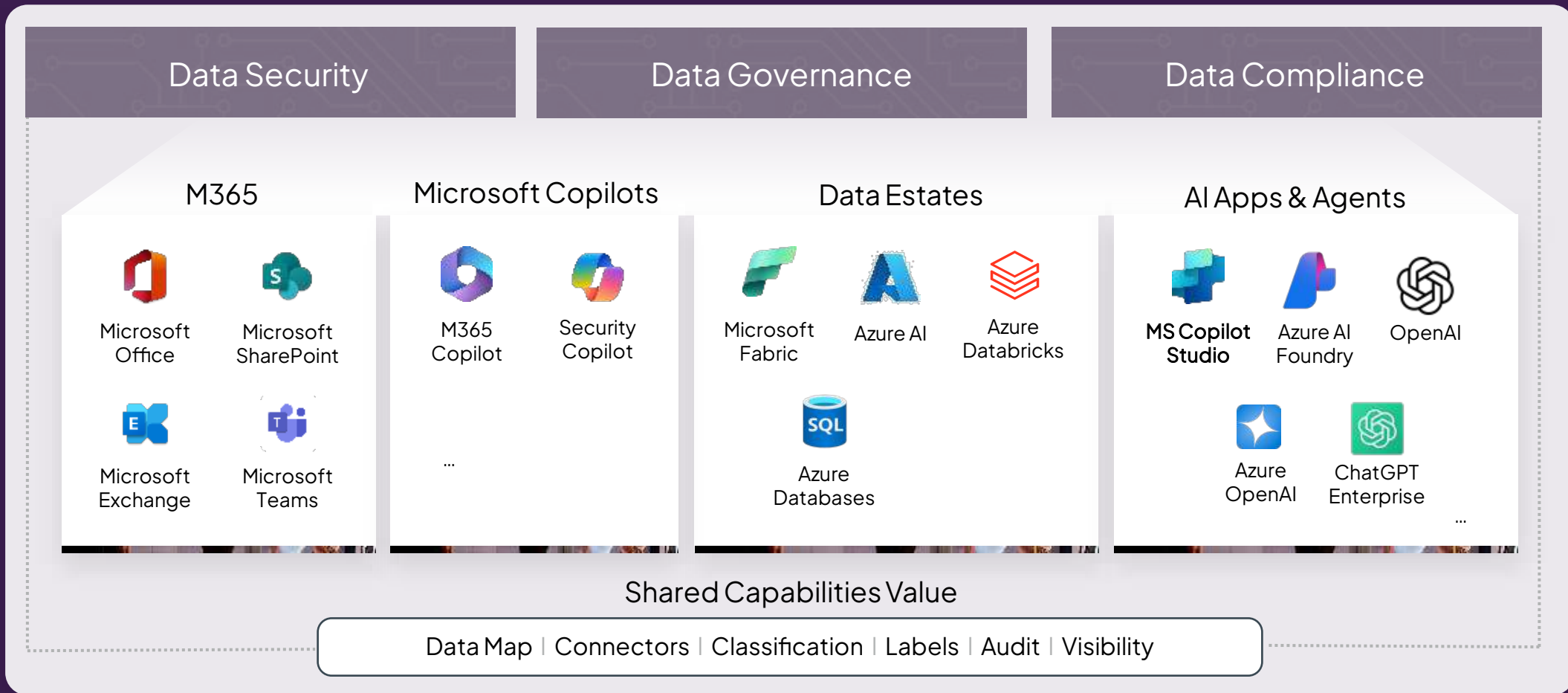


# Technical Robustness and Safety

- Secure by design AI
- Reliability and accuracy
- Resilience against attacks
- Fail-safe mechanisms
- Robustness to data quality issues



# The tools you need are ready



# AI Solutions with the safety of governance



Visibility & Centralised  
Governance



Integration and  
Interoperability



Data Classification &  
Protection



Versioning, Auditing & AI  
Traceability



Compliance & Regulation  
Alignment



## Microsoft Purview

Microsoft Purview Preview

Try the new Microsoft Purview

Home

AI Hub

Analytics

Policies

Alerts

Activity explorer

## AI Hub

Discover and secure all AI activity in Microsoft Copilot and other generative AI apps. Keep your data safe and stay on track with industry regulations. [Learn more](#)

### Get started

- Activate Microsoft Purview Audit**  
Get insights into user interactions with Microsoft Copilot experiences.
- Install Microsoft Purview browser extension**  
Detect risky user activity and get insights into user interactions with other generative AI apps.
- Onboard devices to Microsoft Purview**  
Protect sensitive data from leaking to other generative AI apps.
- Extend your insights for data discovery**  
Discover sensitive data in user interactions with other generative AI apps.

### Recommendations

**Data Security Investigations**

#### Protect sensitive data referenced in Copilot responses

In the last 30 days, 2,932 unprotected files across 12 SharePoint sites were referenced in Copilot responses. Start a data investigation or take steps avoid potential oversharing of sensitive data.

[View details](#)

**New AI regulations**

#### Get guided assistance with new AI regulations

Stay on track with with new AI, such as ISO 42001 and NIS 2, interactions, we've identified regulations.

[View details](#)

**Data security investigations**

#### Protect sensitive data referenced in Copilot responses

Sensitivity labels help protect files by controlling user access to data. Copilot Microsoft 365 honors the sensitivity label on files and only shows the user the files they already have access to.

##### Unlabeled files in Copilot prompts

Unlabeled files: **2.9K** | Sharepoint sites with unlabeled files: **30**

##### Top 5 SharePoint sites with unlabeled files in Copilot prompts

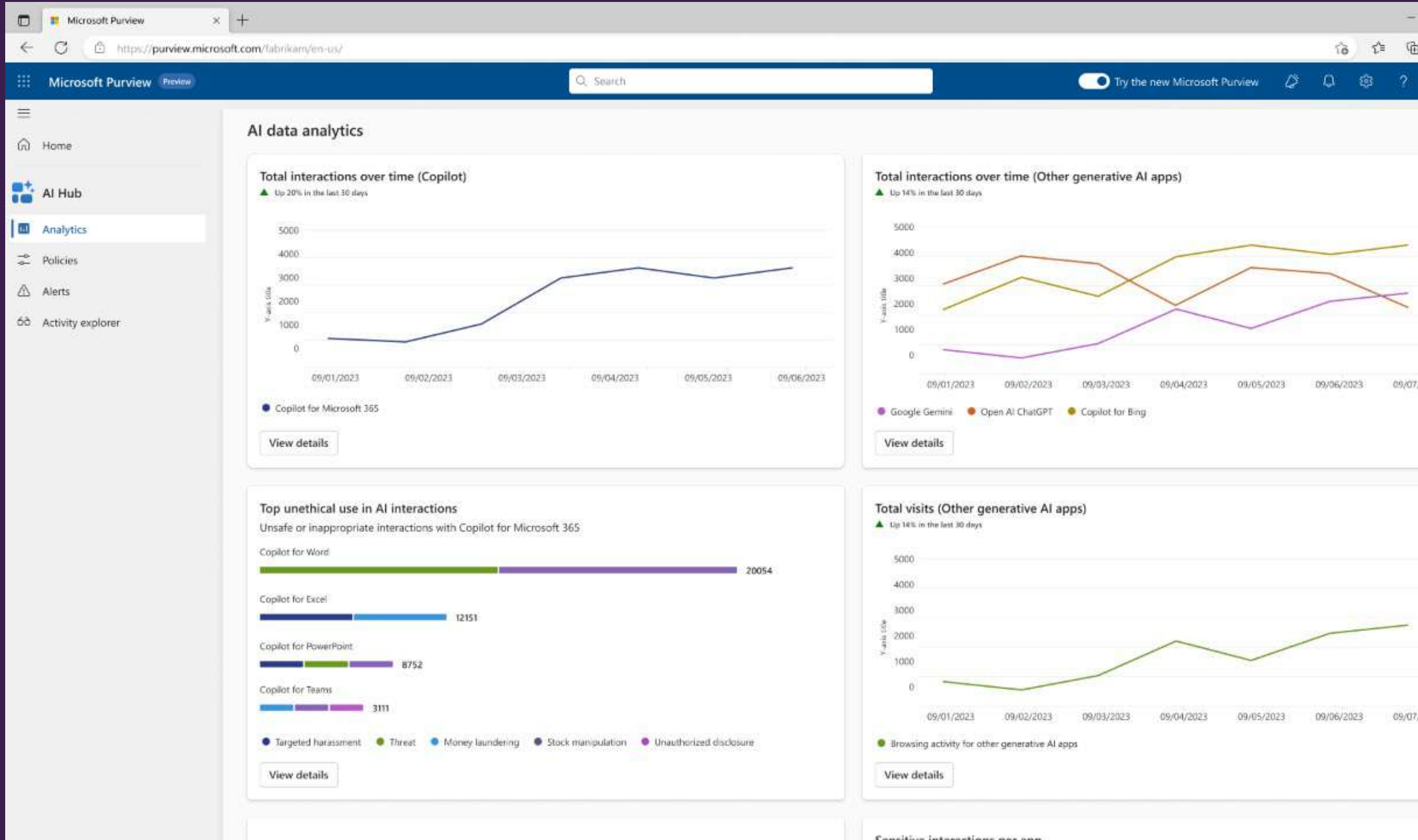
Site	Unlabeled files
Site 1	1,303
Site 2	1,303
Site 3	1,303
Site 4	1,303
Site 5	1,303

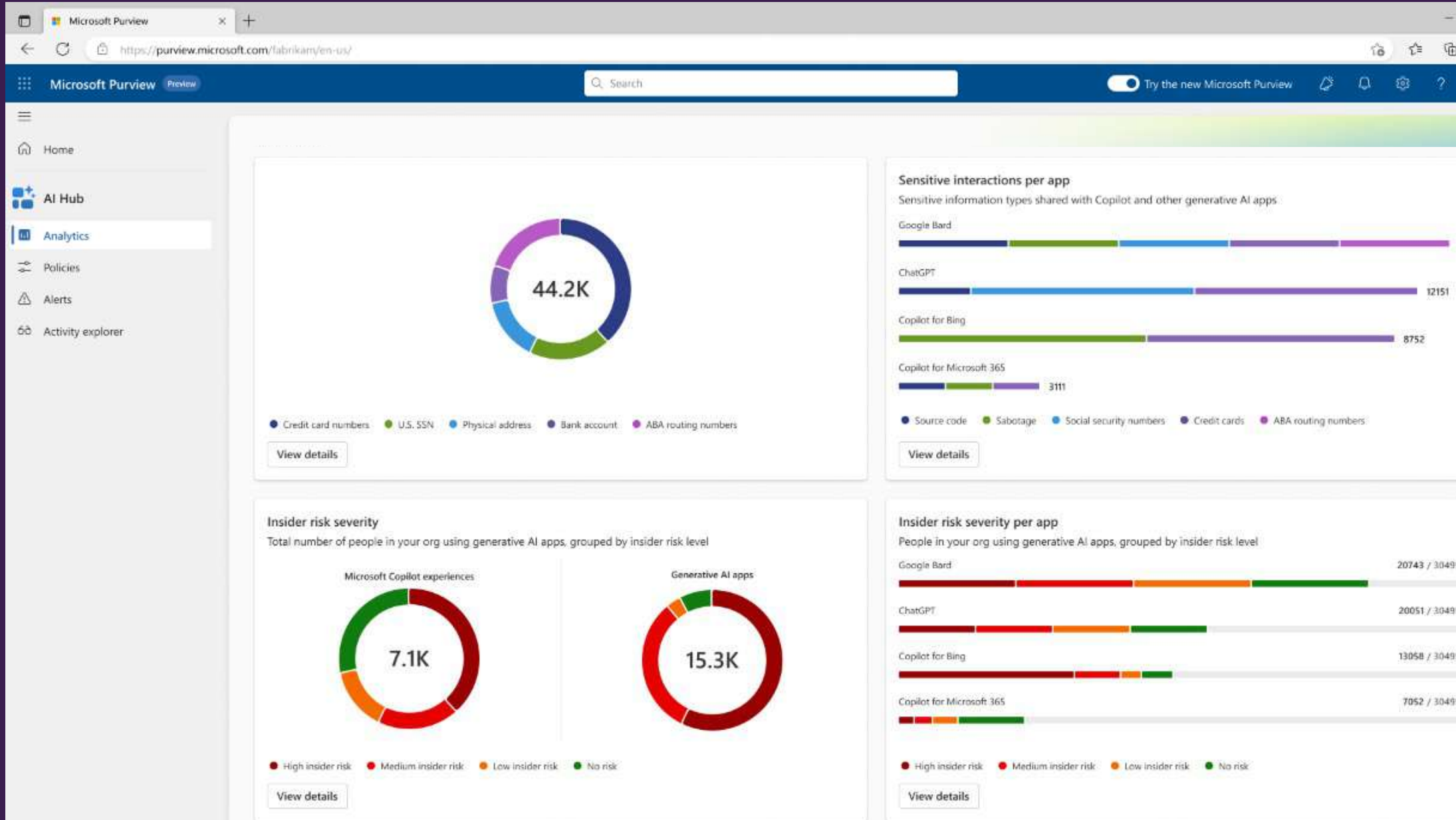
[View top 25 sites](#)

##### Steps

Set up sensitivity label policies:

- Create labels.** Go to Information Protection -> Labels. Select **Create a label** and follow wizard. Repeat process for multiple labels.
- Deploy label policy.** Go to Information Protection -> Label Policy. Select **Create a label policy**.
- Manually apply labels.** Get list of files using [Powershell script](#). Apply sensitivity labels on files using sensitivity label button in apps by opening in app or at scale using powershell.
- Create auto-labeling policy.** Go to Information Protection -> Auto-labeling. Select **Create auto-labeling policy**.







Trustworthy AI, trustworthy privacy  
and trustworthy security is core...  
...if you don't trust it, you're not  
going to use it.

SATYA NADELLA – MICROSOFT CEO  
LONDON UK AI TOUR KEYNOTE – MONDAY 21 OCTOBER 2024

fusion5



# Fusion5 Offer



## Our Services

### Cybersecurity Assessments

- Rapid Microsoft 365
- Comprehensive on-prem/cross cloud

### Data Security Envisioning Workshop

Interactive engagement designed to help organisations uncover hidden data security risks, strengthen compliance, and prepare for emerging challenges such as generative AI.

Leveraging Microsoft Purview tools and services, the workshop provides deep visibility into sensitive data, user behaviours, and regulatory gaps across Microsoft 365 and optional on-premises environments.



## How we can help secure your organization:

### FastStart for Secure Data

Structured 8-week engagement designed to help organisations establish foundational data governance and protection using Microsoft Purview.

It focuses on improving data hygiene, reducing risk exposure, and enabling secure information handling across Microsoft 365.



Turn risk into resilience –  
make security your AI  
accelerator, not a barrier.

**fusion5**