THREATS LIBRARY

JULITH GmbH
Awareness-as-a-Service

Data leaks & shadow IT

Most data leaks in 2026 do not come from hacking, but from publicly configured cloud storage, forgotten SaaS trials, and ChatGPT inputs. We show how shadow IT is identified and channelled constructively.

min read: 7 min **Updated:** 14 March 2026 **Risk:** High risk
Source: awareness-as-a-service.com/en/resources/threats/data-leaks-shadow-it

What are data leaks caused by shadow IT?

Shadow IT describes all hardware and software solutions that employees use in a professional context without the knowledge or approval of the IT department. This includes personal Dropbox accounts used for file transfers, unauthorised SaaS subscriptions (Trello, Notion, Monday.com), AI assistants (ChatGPT, free-tier Copilot), and app downloads on company devices.

The underlying problem is not malicious intent, but a gap between need and IT supply: when approved tools are too slow, too expensive, or too cumbersome, employees use what works. The result is an uncontrolled proliferation of corporate data across external services that have not been assessed, are not encrypted to enterprise standards, and are not registered in the organization's data inventory.

Data leaks in this context frequently arise not from attackers but from misconfiguration: an S3 bucket set to "public"; a Notion document accidentally made shareable via link; internal strategy documents entered as a ChatGPT prompt. These leaks are often open for weeks before anyone notices.

At a glance

01

Misconfiguration overtakes hacking

Publicly accessible cloud storage and incorrectly configured sharing settings cause more data leaks than active attacks. The danger comes from inside.

02

AI chatbots as a data protection risk

Content entered into commercial AI assistants may be used for training purposes. Internal calculations, customer data, and source code do not belong in public AI services.

03

Shadow IT is a symptom, not a problem

Where shadow IT grows, an appropriate approved tool is missing. IT security that only prohibits without offering alternatives shifts the problem but does not solve it.

How to recognise shadow IT and data leaks

Publicly accessible object storage buckets
S3, Azure Blob, or GCS buckets with insecure configuration that can be found with a simple internet search.

Employee accounts registered with personal email
SaaS services registered under a personal email address and used for work data - outside any IT governance.

Internal documents entered into AI chatbots
Employees copy draft contracts, financial data, or customer data into ChatGPT, Gemini, or similar services to improve text or answer questions.

Forgotten test domains and demo environments
Subdomains or staging environments set up for tests but never shut down, still containing corporate data.

Unapproved browser extensions
Productivity extensions that read keystrokes or page content can exfiltrate credentials or sensitive page content.

File sharing via consumer cloud
"I just shared it on Dropbox" or "the link expires in 7 days" - but who else has access, and where is that file stored now?

How to protect yourself

For employees

- **Use approved tools** and actively ask IT if no suitable solution exists - rather than finding one independently.
- **Do not enter internal data into public AI services:** Customer names, contract details, financial data, and source code have no place in commercial chatbots.
- **Choose sharing settings deliberately:** Before sharing a link, check: is it really accessible only to the intended recipient? Or to "anyone with the link"?
- **Report unapproved apps and extensions** rather than simply installing them. IT can evaluate the need and provide a secure alternative.

For administrators

- **Shadow IT discovery:** Analyse CASB (Cloud Access Security Broker) or proxy logs to identify unapproved SaaS services.
- **Cloud configuration audits:** Regularly scan S3, Azure, and GCP for publicly accessible resources (tools: Prowler, ScoutSuite, Cloudmapper).
- **Approved SaaS catalog:** A well-maintained list of approved tools significantly reduces the incentive for shadow IT.
- **AI usage policy** for ChatGPT, Copilot, and similar: Which data classifications may be used in which AI services? Enterprise instances (Azure OpenAI, Microsoft 365 Copilot with a data processing agreement) are preferable to consumer versions.
- **DLP rules** for known file-sharing sites and AI services - block or alert on uploads of sensitive data.

Real cases

CASE 01 · SOFTWARE COMPANY · DE · Q4/2025

A developer configured an S3 bucket for internal deployment artefacts and accidentally set the permission to public - a widespread misconfiguration. The bucket contained environment variables including database passwords and API keys. An automated scanner found it within hours.

Damage: database compromised, customer data exfiltrated · **Detection:** bug bounty report three days after the misconfiguration · **Lesson:** Bucket policies as infrastructure-as-code with mandatory review; Cloud Security Posture Management (CSPM) would have alerted immediately.

CASE 02 · MANAGEMENT CONSULTANCY · CH · Q1/2026

Several consultants used ChatGPT (free tier) to polish client presentations. In doing so, they entered client names, revenue figures, and strategic recommendations as context. The data was thereby in OpenAI's training pipeline.

Damage: data protection breach (notifiable under the Swiss DSG), contractual risk with clients · **Detection:** internal compliance review · **Lesson:** An AI policy was missing. Microsoft 365 Copilot with an enterprise data processing agreement would have been the compliant alternative.

What to do if it happens?

THE FIRST 15 MINUTES

1. **Close access immediately:** Set the public bucket to private, deactivate the sharing link, revoke SaaS access.
2. **Determine the scope:** Which data was affected? How long was access open? (Analyse access logs.)
3. **Check GDPR/DSG notification requirements:** For personal data, a 72-hour notification window for supervisory authorities may apply.
4. **Notify affected individuals and clients** if their data was compromised - early and transparently.
5. **Involve IT Security and Compliance** - do not act alone, to avoid delaying escalation.
6. **Forensic analysis:** Was the data actually accessed? Check access logs, CDN logs, and monitoring data.

Frequently asked questions

May I use ChatGPT for work?

That depends on the company policy and data sensitivity. Public versions without a data processing agreement are not suitable for internal, confidential, or personal data. Enterprise versions (Azure OpenAI, Microsoft 365 Copilot) with appropriate agreements are more privacy-compliant.

What is a CASB and do we need one?

A Cloud Access Security Broker sits between users and cloud services, enforcing policies: which SaaS services are permitted, which data may be uploaded. For organizations with around 50 or more employees and substantial SaaS usage, it is worth evaluating.

How do I find out which shadow IT tools my employees are using?

DNS logs and proxy data give clues about unapproved services. Alternatively, CASB solutions and shadow IT discovery features in modern SASE platforms provide structured overviews.

Are forgotten SaaS trials a genuine risk?

Yes - especially if they were populated with corporate data and are no longer actively monitored. Data in a forgotten trial account can persist for years and is no longer under the organization's control.

Related topics

Shadow IT and uncontrolled data sharing overlap strongly with insider threats and weak credentials. Addressing these three areas together covers a large proportion of the non-technical attack surface.