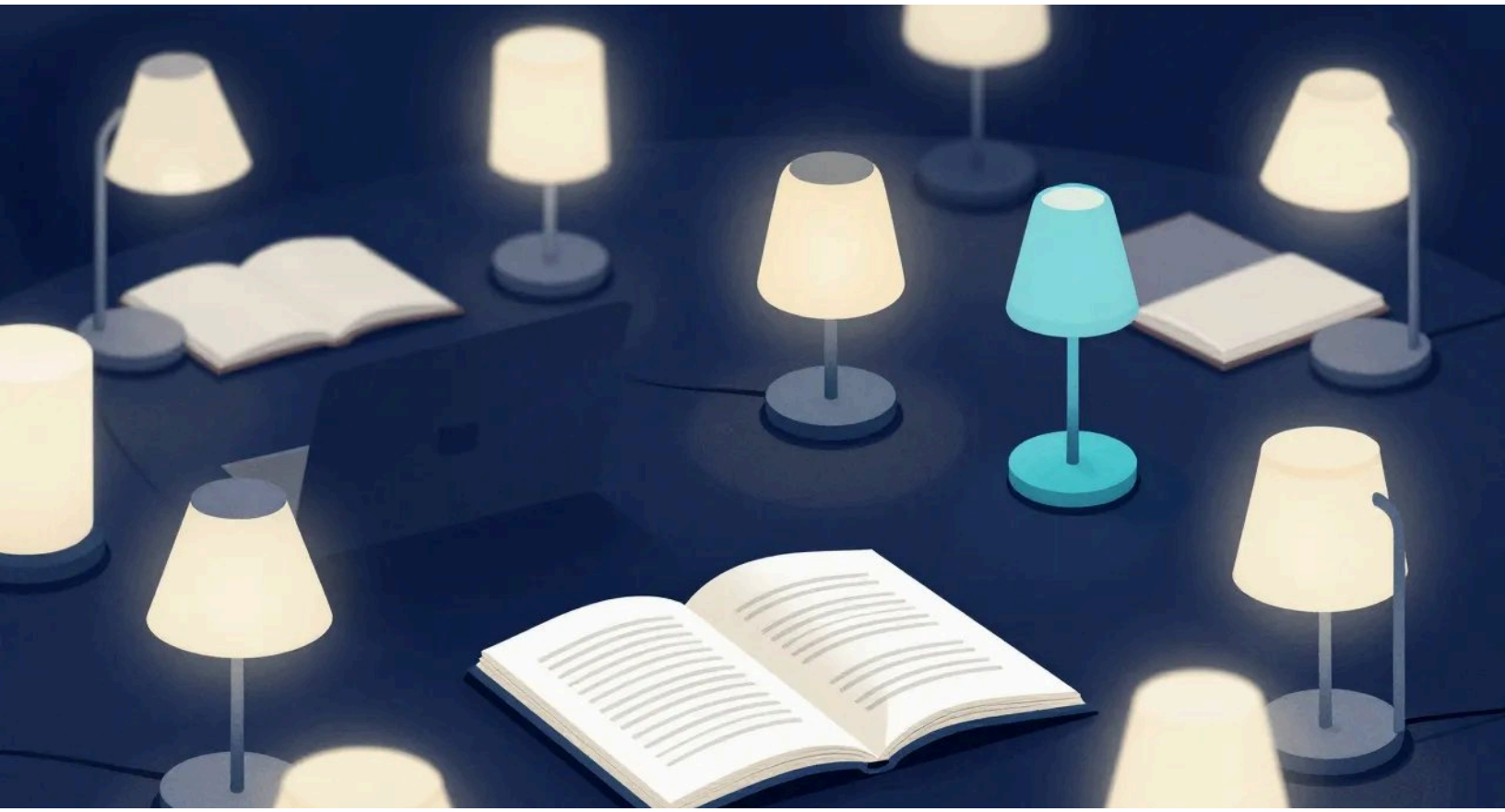


EXECUTIVE PAPER · AI

Your Employees Already Use AI. Set the Rules.

Most staff already paste work into AI tools, often through personal accounts the company cannot see. Chief executives and their legal and security leaders should replace bans and silence with a short policy people can follow.

AUTHOR

Chris Watkins

SERVICE

AI

FOR

CEO, COO, General Counsel, CISO · 12 min read

AT A GLANCE

The problem. Three in four knowledge workers used AI at work in 2024, and most of them brought their own tools. In 2025 data from 222 companies, about a third of ChatGPT use ran through personal accounts and 39.7% of interactions with AI tools involved sensitive data. IBM found that breaches at organizations with high levels of unsanctioned AI cost about $670,000 more.

Why it persists. Companies either ban AI, which staff work around, or say nothing, which staff read as permission. Neither tells an employee which tool to open, what they may paste into it, or who answers for the output.

The recommendation. Publish a one-page policy with four parts: a list of approved tools on enterprise accounts, a table of which data classes may go into them, a rule that the person who uses the output is accountable for checking it, and a request path that answers within a week. Measure use through single sign-on and network logs, and report it to leadership each quarter.

The Decision in Front of You

Your employees are already using AI on company work. The decision is whether that use happens in tools you have contracted, configured, and can audit, or in personal accounts you cannot see.

Microsoft and LinkedIn's 2024 Work Trend Index, a survey of 31,000 people in 31 countries, found that 75% of knowledge workers used AI at work and that 78% of those users brought their own tools [1]. Telemetry tells the same story. Cyberhaven's 2026 AI Adoption and Risk Report, drawn from data movements of hundreds of thousands of employees at 222 companies, found that 32.3% of ChatGPT use and 24.9% of Gemini use came through personal accounts, with higher rates for Claude and Perplexity [2]. It also found that 39.7% of all interactions with AI tools involved sensitive data, about once every three days for the average employee [2].

Written policies, where they exist, are not changing behavior. A June 2026 survey of 1,250 office workers at companies with revenue above $500 million, run by Wakefield Research for PagerDuty, found that 66% had used AI tools they believed company policy did not allow [3]. Of the same group, 34% had entered customer data into public AI tools and 31% had entered financial information or confidential documents [3].

The cost shows up in breaches. IBM's 2025 Cost of a Data Breach study, run by the Ponemon Institute across 600 breached organizations, found that one in five had a breach involving shadow AI, meaning AI tools used without the organization's approval or knowledge [4]. Breaches at organizations with high levels of shadow AI cost an average of $670,000 more than at those with little or none, and 63% of

breached organizations had no AI governance policy or were still writing one [4]. Gartner predicted in November 2025 that more than 40% of organizations will have security or compliance incidents caused by unauthorized AI by 2030 [5].

What's Happening

The failures fall into two groups

Most AI incidents at work fall into two groups. Data goes out, when an employee puts confidential material into a tool the company has not approved. Bad output comes back, when an employee or a customer-facing system presents invented content as fact. A policy has to cover both, and the cases below show each.

The pattern is older than the current tools. In 2023 Samsung banned generative AI on company devices after engineers uploaded internal source code to ChatGPT [6].

Four cases from 2024 and 2025

NSW Reconstruction Authority, 2025. Between March 12 and 15, 2025, a contractor to the New South Wales government agency uploaded a spreadsheet with more than 12,000 rows of data on applicants to a flood recovery program into ChatGPT [7]. The file held names, addresses, phone numbers, and some health information for up to 3,000 people [7]. The agency said the upload was not authorized, found no evidence of third-party access, and disclosed the incident in October 2025, about six months later, after which it issued staff guidance on unapproved AI tools [7]. The lesson is the order of events. The guidance arrived after the breach, and the agency then had to admit that public AI tools are hard to monitor.

Deloitte Australia, 2025. Australia's Department of Employment and Workplace Relations paid Deloitte AU$440,000 for an independent review of a welfare compliance system [8]. After publication in July 2025, an academic found fabricated citations, references to reports that did not exist, and a quote attributed to a Federal Court judge that the judge never wrote [8]. The corrected report disclosed that Deloitte had used a GPT-4o tool chain hosted in the department's own Azure tenancy, and the firm agreed to repay the final installment of the contract [8]. The tool was approved and the data stayed inside the client's environment. The failure happened at the review step, which no tool choice replaces.

Johnson v. Dunn, 2025. On July 23, 2025, a federal judge in Alabama publicly reprimanded three lawyers at Butler Snow, a firm of more than 400 attorneys, removed them from the case, and referred them to the state bar [9]. One had used general-purpose ChatGPT to generate citations for two motions; five did not exist, and no one checked them before filing [9]. The firm had adopted an AI policy in June 2023 that required written approval for generative AI tools and independent verification of their work [9]. The policy was correct and had no effect, because no approved tool was easier to reach than ChatGPT and nothing checked the citations before they reached the court.

That case followed Mata v. Avianca in 2023, where a New York court fined two lawyers and their firm $5,000 for filing fake cases produced by ChatGPT [10]. A public database maintained by the researcher Damien Charlotin had logged 2,262 court and tribunal decisions involving AI-generated hallucinations at the time of writing [11].

Moffatt v. Air Canada, 2024. British Columbia's Civil Resolution Tribunal held Air Canada liable for negligent misrepresentation after the chatbot on its website told a grieving customer he could claim a bereavement fare retroactively, which contradicted another page on the same site [12]. The airline argued that the chatbot was effectively a separate entity responsible for its own statements, and the tribunal rejected that argument [12]. The award was small, and the ruling puts the company on the hook for what its AI tells customers.

SIDEBAR: WHY A BAN FAILS

A ban removes the approved option and leaves the personal one. Staff who have found AI useful keep using it on their phones and home laptops, where the company has no contract, no retention controls, and no logs. Samsung's 2023 ban followed a leak [6]. The surveys above show that bans and unread policies move use out of sight without stopping it [2][3].

Why the Usual Approach Fails

Companies tend to pick one of three positions. Each fails for a specific reason.

Silence. Leadership says nothing, and employees take that as permission. They choose tools by convenience, use free tiers whose terms may allow the provider to keep or train on inputs, and decide for themselves what is safe to paste. IBM found that only 34% of organizations with AI governance policies perform regular audits [4].

A ban. The company blocks AI sites on its network and tells staff not to use AI. Usage moves to personal devices. The company loses visibility and keeps the risk. The PagerDuty survey found that 86% of workers believed their organization had AI policies, while two-thirds had used tools they thought were not allowed [3].

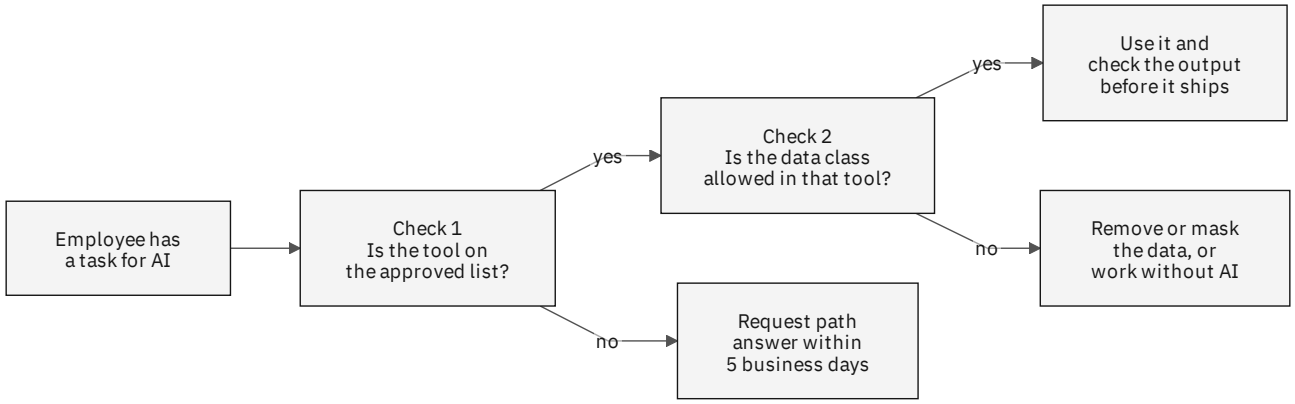
A long policy without tools. Legal writes a careful ten-page document and publishes it on the intranet. It does not name an approved tool, so the easiest path is still a personal account. Johnson v. Dunn shows the result: a sound written rule, an unapproved tool, and no verification [9].

The common root cause is that each approach tells employees what they may not do and gives them nothing they can use instead.

The Framework: The One-Page AI Rule

The One-Page AI Rule fits on a single page so that employees will read it. It has four parts, each with an owner.

Exhibit 1. Every AI task passes two checks before work starts, and a request path handles everything else



Source: author's framework.

Approved tools. Name the tools staff may use, on enterprise accounts the company controls. The contract and configuration should cover five points: the provider does not train on your data, retention is set by you, sign-in runs through your single sign-on, administrators can see usage logs, and data stays in a region you accept. Pay for enough seats that the approved tool is easier to reach than a personal account. The CIO or CTO owns this list.

Data classes. Most companies already classify data for security or privacy. Map each class to where it may go. Exhibit 2 shows a typical mapping. The general counsel and the CISO own it together.

Exhibit 2. Most work data may go into approved enterprise tools; regulated and secret data needs a specific approval

Data class	Examples	Approved enterprise AI tools	Personal or free AI accounts
Public	Published marketing, public filings, job postings	Allowed	Allowed
Internal	Process documents, meeting notes, internal email	Allowed	Not allowed
Confidential	Contracts, pricing, financials before release, source code	Allowed in tools approved for this class	Not allowed
Restricted	Customer personal data, health data, payment card data, privileged legal material	Only in a tool approved for that specific data set, after legal and security review	Not allowed
Secrets	Passwords, API keys, private keys	Not allowed in any AI tool	Not allowed

Source: author's recommendation.

Output responsibility. The person who sends, files, publishes, or deploys AI-assisted work is accountable for it. They check every citation, figure, quotation, and legal or medical statement against a primary source before it leaves their hands. Where a client contract or court rule requires disclosure of AI use, they disclose it. Customer-facing AI systems need an owner who tests their answers against company policy before launch and after each change; the Air Canada ruling makes clear the company is answerable for what they say [12]. Each department head enforces this for their team.

Request path. Staff who need a tool that is not on the list submit a short form: the tool, the task, the data class involved. Security and legal answer within five business days with yes, no, or yes with conditions. Publish the decisions so the next person can see them. A slow answer sends people back to their personal accounts.

Standards give the policy a structure auditors recognize

The NIST AI Risk Management Framework, AI 100-1, organizes AI risk work into four functions: Govern, Map, Measure, and Manage [13]. Its 2024 Generative AI Profile, NIST AI 600-1, lists risks specific to generative systems, including invented content and leakage of confidential information, with suggested actions for each [14]. ISO/IEC 42001:2023 is a certifiable management system standard for organizations that build or use AI [15]. The One-Page AI Rule is the employee-facing layer of that work. The approved list, the data class table, and the request log become audit evidence.

In the European Union, Article 4 of the AI Act, as amended by the 2026 Digital Omnibus, requires providers and deployers of AI systems to take measures to support AI literacy among staff who operate or use those systems [16]. A short policy, with training on what it means, is a direct way to show those

measures exist.

Costs, Risks, and Tradeoffs

The policy costs money for enterprise licenses, staff time in security and legal to run the request path, and some friction for employees used to free tools. Exhibit 3 compares it with the alternatives.

Exhibit 3. A short enforced policy costs more than silence and less than a breach

Option	What it costs	What it risks	Mitigation
Silence	Nothing up front	Data in personal accounts with no logs; shadow AI breach premium of about $670,000 [4]	None; this is the default the policy replaces
Ban	Blocking and enforcement effort	Use moves to personal devices; the company loses visibility and talent [2][3]	Keep bans only for specific tools with unacceptable terms
One-Page AI Rule	Enterprise seats; request-path staffing; training	Staff still paste restricted data or skip output checks	Single sign-on logs, data loss prevention rules on AI sites, spot checks of shipped work
Output checks	Reviewer time on every AI-assisted deliverable	Rushed checks miss invented citations, as in Deloitte and Johnson v. Dunn [8][9]	Require a named reviewer for client, court, and regulator deliverables
When this is wrong	A firm whose contracts or regulators forbid any third-party AI processing, such as some defense or classified work	An approved cloud tool may itself breach those terms	Use AI models hosted inside your own environment, or approve no tools and enforce the ban with device controls

Source: author's analysis.

How to Act

Exhibit 4. A one-year roadmap

Window	Actions	Owner	Milestone
First 30 days	Measure current AI use from single sign-on, web proxy, and expense data. Pick one or two approved tools and sign enterprise terms. Publish the one-page policy with the data class table.	CIO with general counsel and CISO	Usage baseline on file; policy published; approved tools available to every employee who needs them
First 90 days	Open the request path with a five-business-day target. Add data loss prevention rules for restricted data and secrets on AI sites. Train managers on the policy. Inventory customer-facing AI systems and name an owner for each.	CISO with HR and department heads	90% of requests answered within five business days; share of AI traffic on approved accounts measured monthly; every customer-facing AI system has an owner
First 365 days	Require named review of AI-assisted client, court, and regulator deliverables. Map policy evidence to NIST AI RMF and ISO/IEC 42001. Report AI use, requests, and incidents to leadership quarterly.	General counsel with risk or compliance	At least 80% of AI traffic on approved enterprise accounts; zero restricted-data incidents in AI tools; quarterly report delivered

Source: author's recommendation.

The 80% target is the author's suggested starting point. Set your own after the 30-day baseline shows where you begin.

Questions to Ask Your Team

1. What share of our employees' AI use runs through accounts we control, and how do we know?
2. Which AI tools have we approved, and do their contracts bar the provider from training on our data?
3. If an employee needs a new AI tool tomorrow, how long until they get an answer?
4. Which of our data classes may go into AI tools, and is that written down where staff will see it?
5. Who checks AI-assisted work before it goes to a client, a court, or a regulator?
6. Which AI systems answer our customers directly, and who tested their answers last?
7. If a contractor uploaded customer data to a personal AI account last month, how would we find out?

Closing

Your employees decided to use AI some time ago. A ban sends that use into personal accounts and silence leaves it there. Publish a one-page rule that names approved enterprise tools, states which data may go into them, holds the person who ships the work responsible for checking it, and answers tool requests within a week. Then measure how much use moves onto the accounts you control and report it to leadership every quarter.

Sources

[1] Microsoft. "Microsoft and LinkedIn release the 2024 Work Trend Index on the state of AI at work." 2024. https://blogs.microsoft.com/blog/2024/05/08/microsoft-and-linkedin-release-the-2024-work-trend-index-on-the-state-of-ai-at-work/.

[2] Cyberhaven Labs. "2026 AI Adoption & Risk Report." 2026. https://info.cyberhaven.com/hubfs/Webflow_Resources/Cyberhaven-AI-Risk-Report-2026.pdf.

[3] PagerDuty. "PagerDuty Report Finds Two-Thirds (66%) of Office Professionals Have Used Unauthorized AI Tools at Work." Survey conducted by Wakefield Research. 2026. https://www.pagerduty.com/newsroom/shadow-ai-workplace-survey-2026/.

[4] IBM. "IBM Report: 13% Of Organizations Reported Breaches Of AI Models Or Applications, 97% Of Which Reported Lacking Proper AI Access Controls." Cost of a Data Breach Report 2025, conducted by Ponemon Institute. 2025. https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls.

[5] Infosecurity Magazine. "Gartner: 40% of Firms to Be Hit By Shadow AI Security Incidents." Coverage of a Gartner prediction. 2025. https://infosecurity-magazine.com/news/gartner-40-firms-hit-shadow-ai.

[6] Bloomberg. "Samsung Bans Staff's AI Use After Spotting ChatGPT Data Leak." 2023. https://bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak.

[7] iTnews. "NSW gov contractor uploaded Excel spreadsheet of flood victims' data to ChatGPT." 2025. https://www.itnews.com.au/news/nsw-gov-contractor-uploaded-excel-spreadsheet-of-flood-victims-data-to-chatgpt-620815.

[8] The Register. "Deloitte refunds Aussie gov after AI fabrications slip into $440K welfare report." 2025. https://www.theregister.com/2025/10/06/deloitte_ai_report_australia/.

[9] EDRM. "When AI Policies Fail: The AI Sanctions in Johnson v. Dunn and What They Mean for the Profession." Johnson v. Dunn, No. 2:21-cv-1701 (N.D. Ala. July 23, 2025). 2025. https://edrm.net/2025/07/when-ai-policies-fail-the-ai-sanctions-in-johnson-v-dunn-and-what-they-mean-for-the-profession/.

[10] Bloomberg Government. "Phony ChatGPT Brief Leads to $5,000 Fine for NY Lawyers." Mata v. Avianca, Inc. (S.D.N.Y. 2023). 2023. https://news.bgov.com/daily-tax-report-state/chatgpt-phony-legal-filing-case-gets-lawyers-a-5-000-fine.

[11] Charlotin, D. "AI Hallucination Cases." Database. https://www.damiencharlotin.com/hallucinations/.

[12] McCarthy Tétrault. "Moffatt v. Air Canada: A Misrepresentation by an AI Chatbot." Moffatt v. Air Canada, 2024 BCCRT 149. 2024. https://www.mccarthy.ca/en/insights/blogs/techlex/moffatt-v-air-canada-misrepresentation-ai-chatbot.

[13] National Institute of Standards and Technology. "AI Risk Management Framework." NIST AI 100-1. 2023. https://www.nist.gov/itl/ai-risk-management-framework.

[14] National Institute of Standards and Technology. "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile." NIST AI 600-1. 2024. https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence.

[15] International Organization for Standardization. "ISO/IEC 42001:2023 Information technology: Artificial intelligence: Management system." 2023. https://www.iso.org/standard/42001.

[16] European Commission, AI Act Service Desk. "Article 4: AI literacy." Regulation (EU) 2024/1689, consolidated text as amended. 2026. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-4.

The views in this paper are the author's own and do not represent any employer. Scenarios are hypothetical. This paper is not legal, financial, or compliance advice.