There's a Second Company Inside Your Company

Shadow AI — your employees already run on tools you don't know about, can't see, and never approved.

Kymata Labs Research · An independent research institution studying how AI systems actually behave in production.

~12 min read

Tags · Shadow AI · Enterprise Risk · Data Governance · AI Security

While leadership debates an AI strategy, the workforce has already built one — unsanctioned, running in the dark, pasting source code, customer records, and trade secrets into consumer AI tools every day, on accounts nobody audits. Drawing on Samsung's May 2023 generative-AI ban after engineers leaked semiconductor source code across three incidents in roughly twenty days, Cyberhaven telemetry showing [3] 39.7% of AI interactions now involve sensitive data (up from an 11%-confidential benchmark across 1.6 million workers in 2023), Microsoft and LinkedIn's finding [5] that 75% of knowledge workers use AI and 78% bring their own tools while 52% won't admit it, Software AG's finding [6] that 50% use shadow AI and 46% would refuse to stop if banned, Harmonic Security's measurement [7] that 45% of ChatGPT prompts already route through ungoverned personal accounts, and IBM's 2025 figure [8] that shadow-AI breaches cost roughly $670K more, this paper argues the biggest AI risk to most companies is not the model they are cautiously evaluating — it is the dozen they are unknowingly already using, with no policy, no audit, and no record of what has already left the building.

The argument, in moves

The thesis is short and the evidence converges from four independent directions. The moves:

1. **The behaviour is already universal.** Microsoft and LinkedIn's 2024 survey of 31,000 people across 31 countries found 75% of knowledge workers already use AI at work and 78% bring their own tools [5]. This is not a future risk; it arrived a while ago.
2. **The exposure is the tools already in use, not the one under review.** For most companies the model being cautiously evaluated in a committee is not the real exposure — the exposure is the dozen tools already in daily use, running with no policy, no audit, and no record of what has already left the building.
3. **Sensitive data is pouring in, and accelerating.** Cyberhaven instruments what employees actually paste, not what they report, and reads 39.7% of AI interactions as involving sensitive data — roughly a quadrupling from an 11%-confidential benchmark in 2023 [3, 4].
4. **A ban does not end the behaviour; it moves it.** 46% of shadow-AI users say they would refuse to stop even if banned [6], and 45% of ChatGPT prompts already route through personal accounts that bypass governance entirely [7].

5. **A policy is not a control, and the gap has a price.** 97% of organizations with an AI-related breach lacked proper AI access controls, 63% had no AI governance policy, and shadow-AI breaches cost about $670,000 more [8].

Takeaway: the second company inside your company is already running and carrying your data; the only real choice left is whether you can see it clearly enough to govern it.

The flagship case: Samsung banned generative AI after engineers leaked source code

The strongest single piece of evidence is not a survey; it is an incident at one of the most security-conscious manufacturers on earth. In roughly three weeks in the spring of 2023, Samsung's own semiconductor engineers pasted confidential data into ChatGPT across three separate incidents [1]. The leaked material included source code from a facility-measurement database, equipment program code, and a recording of an internal meeting. On May 2, 2023, Samsung banned generative-AI tools on company devices and networks and capped any remaining prompts at 1,024 bytes [2].

Its warning to staff is the part that should travel: data sent to ChatGPT sits on external servers the company cannot retrieve or delete. Some of the fastest engineers at one of the most security-conscious manufacturers on earth did this — not out of malice, but because the tool was useful and the friction to reach it was zero. The early movers drew the obvious conclusion. In the first months of 2023, JPMorgan, Amazon, Bank of America, Citigroup, Deutsche Bank, Wells Fargo, Goldman Sachs, and Verizon all restricted ChatGPT inside their walls. None of these are organizations that underspend on security; they acted because the exposure was never about a single careless employee — it had become the default behaviour of an entire workforce.

The evidence: four lenses, one convergent picture

Four independent studies measure the same thing, and the number isn't reassuring: a flagship leak, telemetry from millions of workers, the most-cited workplace survey in the field, and the hard breach economics. What makes the thesis hard to dismiss is not any single study but that the same shape recurs across sources that share almost nothing else.

Lens	Finding	Source
The flagship leak	Samsung engineers leaked source code in three incidents in ~20 days; ban May 2, 2023	Bloomberg / TechCrunch, 2023
The telemetry	39.7% of AI interactions involve sensitive data, up from an 11%-confidential paste benchmark across 1.6M workers (2023)	Cyberhaven, 2026 / 2023
The survey	75% of knowledge workers use AI; 78% bring their own tools; 52% won't admit it for key tasks	Microsoft & LinkedIn, 2024
The majority	50% use shadow AI; 46% would refuse to stop if banned; 45% of prompts via personal accounts	Software AG, 2024; Harmonic, 2025
The economics	Shadow-AI breaches cost ~$670K more ($4.63M vs $3.96M); a factor in 20% of breaches	IBM, 2025

The telemetry: sensitive data is pouring into AI, and it's accelerating. Cyberhaven's current reading is stark: 39.7% of AI interactions involve sensitive data [3]. Held against the original benchmark — in 2023, with telemetry from 1.6 million workers, 11% of everything employees pasted into ChatGPT was confidential,

with source code and client data among the top leaking categories [4] — the trend line runs in a single direction. As the tools grew better and more embedded, the share of sensitive material flowing through them roughly quadrupled.

The survey: they're using it, and over half are hiding it. Microsoft and LinkedIn's 2024 Work Trend Index surveyed 31,000 people across 31 countries: 75% of knowledge workers use AI at work and 78% bring their own AI tools, a habit the report labels "BYOAI" [5]. The figure that should change how you read your own org chart is the next one: 52% are reluctant to admit they use AI for their most important tasks. Roughly half the workforce is doing its highest-stakes thinking with a tool it won't name.

The majority: shadow AI is the norm, and they won't give it up. Software AG's October 2024 study of roughly 6,000 knowledge workers across the US, UK, and Germany found that 50% use shadow AI — tools their company never issued — and 46% of them would refuse to stop even if their employer banned the tools outright [6]. Harmonic Security analyzed 176,460 prompts in Q1 2025 and found 45% of ChatGPT prompts came through personal accounts that bypass corporate governance entirely [7]. Bolt the front door and the traffic simply finds the window.

The economics: shadow-AI breaches cost more, and the controls aren't there. IBM's 2025 Cost of a Data Breach report, drawn from roughly 600 breached organizations, priced the gap: breaches involving shadow AI cost about $670,000 more — $4.63M versus $3.96M — and shadow AI was a factor in 20% of breaches [8]. The governance picture underneath turns out worse than the cost: 97% of organizations with an AI-related breach lacked proper AI access controls, and 63% had no AI governance policy at all. This is not exotic; it is the default condition of most companies, now with a price tag attached.

How we got here: no one ordered this, everyone built a piece of it

Shadow AI was never a strategy and never a rebellion. It accumulated as the sum of a million small and entirely reasonable choices. A capable tool arrived, free or nearly so, with no install, no procurement, and no ticket to file. It made hard work fast, collapsing the distance between "I'm stuck" and "I'm unstuck" from hours to seconds. Inside that newly tiny gap, an employee under deadline does the most human thing available: pastes in the problem and takes the answer.

Leadership, meanwhile, did the responsible-sounding thing — convened the committee, commissioned the risk review, debated the policy. The result is a structural mismatch in tempo. Adoption moves at the speed of an individual's next task; governance moves at the speed of a quarterly steering meeting. By the time the official AI strategy clears its final approval, the unofficial one has been running for two years and touched almost everything.

This is why two kinds of companies are now forming, and only one of them can see itself. The open question was never whether your employees use AI — that is settled at 75%. What still varies is whether they have a sanctioned way to do it. One kind of company offers an approved, monitored, enterprise-grade path and meets the demand out in the open. The other bans the tools, declares the matter closed, and drives the identical behaviour onto personal accounts and personal phones it will never see. The resulting risk lands unevenly: it pools in the companies that mistook a prohibition for a control, which are precisely the ones least equipped to measure what is leaving.

What it means: the same evidence, read by three people who own the problem

Exposure is not destiny. The same studies that diagnose shadow AI also point toward the response: meet the demand in the open, instrument the flow of data, then govern what you have made safe to govern. What that requires depends on where you sit.

For employees

- Assume anything you paste into a consumer AI tool has left the building for good — Samsung's own warning was that it cannot be retrieved or deleted.
- Source code, customer records, contracts, and anything under NDA are the categories that leak most. Treat the paste field like an outbound email to a stranger.
- If the sanctioned tool is slower, say so, loudly. Quietly routing around it is exactly how the second company grows.

For security & legal

- 97% of AI-related breaches involved organizations without proper AI access controls — assume you are in that 97% until you have measured otherwise.
- Instrument the actual flow of sensitive data into AI tools. Telemetry exists precisely because policy text is invisible to the paste field.
- Bans displace rather than stop, and 45% of prompts already route through personal accounts. Provide a safe path, or accept that you are mostly blinding yourself.

For leadership

- A policy is not a control. 63% of breached orgs had no AI policy, 97% with breaches lacked the controls, and 46% of users say a ban won't stop them anyway.
- Sanction and instrument fast. Every week of committee debate is another week the workforce governs itself, off the record.
- Price the inaction. A shadow-AI breach runs about $670K higher, and shadow AI shows up in one breach out of five, so the cost of moving stays well below the cost of not knowing.

The model under evaluation was never the real risk

None of this is an argument against AI at work — the productivity is real and the demand is universal. It is an argument for ending the convenient fiction that the shift hasn't already happened. The second company inside your company will keep running whether or not leadership acknowledges it. The only real choice left is whether you can see it clearly enough to govern it. Find the second company while its leaks are still its problem and not yet yours.

Frequently asked questions

We have a policy against this. Doesn't that cover us?

A policy is not a control. IBM's 2025 breach report found that 63% of breached organizations had no AI governance policy at all, and the more uncomfortable figure sits right beside it: 97% of organizations that suffered an AI-related breach lacked proper AI access controls [8]. A line in the handbook stops nobody. The Software AG data is blunter still: of the employees using unsanctioned AI tools, 46% said they would keep using them even if their employer banned them outright [6]. You can write the rule, but the second company never reads the handbook.

Isn't blocking the tools the obvious fix?

Blocking is the obvious fix that happens not to work. The demand behind shadow AI is real and large: 75% of knowledge workers already use AI at work, and 78% bring their own tools [5]. Block the sanctioned path and the behaviour doesn't die — it migrates to personal accounts and personal phones, where you have no visibility at all. Harmonic Security measured exactly that displacement, finding that 45% of ChatGPT prompts came through personal accounts that bypass corporate governance entirely [7]. Banning the tools without offering a safe alternative simply converts a problem you can see into one you can't.

How much does getting this wrong actually cost?

IBM put a number on it. In 2025, breaches involving shadow AI cost roughly $670,000 more — $4.63M versus $3.96M — and shadow AI was a factor in 20% of all breaches studied [8]. That is not a rare tail event; it is a recurring line item that shows up in one breach out of every five. Worse, the dollar figure is only the part you can measure. The trade secret now sitting on a third-party server, the source code an engineer pasted to debug it on a Tuesday afternoon — none of that shows up cleanly on the incident report.

Our data went into a consumer AI tool. Can we get it back?

No, and that was the heart of Samsung's warning. When it banned generative-AI tools in May 2023, the company told staff plainly that data transmitted to services like ChatGPT is stored on external servers, which makes it impossible to retrieve and delete [1]. Once a fragment of source code or a customer record crosses your perimeter into a consumer tool, custody is gone. There is no recall button. The only durable control is preventing the paste in the first place, because chasing it afterward leads nowhere.

Is this really an enterprise problem, or just a few careless employees?

It's structural, and the biggest names reached that conclusion early. In the first months of 2023, JPMorgan, Amazon, Bank of America, Citigroup, Deutsche Bank, Wells Fargo, Goldman Sachs, and Verizon all restricted ChatGPT inside their walls. None of these are organizations short on security budget or sophistication. They moved because the exposure is a property of how the workforce now operates rather than the carelessness of a few individuals. Any security model that assumes employees will refrain from pasting sensitive data into the most useful tool they have ever been handed is quietly betting against human nature, and losing.

References

1 Bloomberg (2023). "Samsung Bans ChatGPT, Generative AI Use by Staff After Leak." May 2, 2023. (The three-incidents-in-twenty-days detail traces to The Economist Korea / Economist economy reporting.) https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak

2 TechCrunch (2023). "Samsung bans use of generative AI tools like ChatGPT after April internal data leak." May 2, 2023. https://techcrunch.com/2023/05/02/samsung-bans-use-of-generative-ai-tools-like-chatgpt-after-april-internal-data-leak/

3 Cyberhaven (2026). "Sensitive data flowing into AI tools." Current telemetry: 39.7% of AI interactions involve sensitive data. https://www.cyberhaven.com/blog/sensitive-data-flowing-into-ai-tools

4 Cyberhaven (2023). "4.2% of workers have pasted company data into ChatGPT." Telemetry from 1.6 million workers; 11% of pasted content was confidential. https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt

5 Microsoft & LinkedIn (2024). "AI at Work Is Here. Now Comes the Hard Part." 2024 Work Trend Index Annual Report. Survey of 31,000 people across 31 countries. https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part

6 SecurityWeek / Software AG (2024). "The Shadow AI Surge: Study Finds 50% of Workers Use Unapproved AI Tools." Survey of ~6,000 knowledge workers across the US, UK, and Germany, October 2024. https://www.securityweek.com/the-shadow-ai-surge-study-finds-50-of-workers-use-unapproved-ai-tools/

7 Harmonic Security (2025). "The AI Tightrope: Balancing Innovation and Exposure." Analysis of 176,460 prompts, Q1 2025: 45% of ChatGPT prompts came through personal accounts. https://www.harmonic.security/resources/the-ai-tightrope-balancing-innovation-and-exposure

8 IBM (2025). "Cost of a Data Breach Report 2025." Conducted by Ponemon Institute across ~600 breached organizations. https://www.ibm.com/reports/data-breach