NEURAL SEVEN
ADVISORY

Your IAM Assumes Humans. Your AI Agents Didn't Get The Memo

You're probably not going to get taken down by a lone insider with a dodgy USB stick.
If something bites you, it's more likely to be an AI agent glued into three systems at once, quietly pushing changes while everyone's in the Monday leadership meeting.

We've moved past "AI as a chatbot". The industry data backs that up. McKinsey's State of AI work points out that the vast majority of organisations now use AI in at least one function, and about a quarter are already scaling agentic systems, not just experimenting.

PwC's agent survey says around 79% of executives report AI agents running somewhere in their business.

So this isn't hypothetical. If you haven't got agents yet, you probably will. If you already have them, they're probably deeper in your stack than you think.

Why IAM starts to wobble when agents show up

Most identity stacks were built around a simple picture: a person logs in, gets some roles, clicks around, logs out. IAM tracks the human.

Agents don't fit that picture. One agent might act for several people, jump across SaaS apps, hit internal APIs, and keep running overnight. Identity practitioners are increasingly blunt about it: IAM, as we've used it for years, doesn't cope well with non-human actors behaving like this.

Under the hood, four things tend to break:

- Inventories rarely include agents properly, so you don't actually know what's live.
- Permissions often end up broad and permanent because that's faster than designing granular scopes.
- Delegation models are fuzzy; agents impersonate humans instead of working with clear authority grants.

- Audit trails don't tell a clean story of who or what took a given action, and on whose behalf.

That would be concerning even in a quiet threat landscape. It's worse when identity is already the main attack path. Various reports now put 70–80% of security incidents down to identity failures: compromised credentials, misconfigured accounts, unmanaged access. CyberArk's survey found 93% of organisations had two or more identity-related breaches in the past year, with machine identities driving much of the growth.

So you've got identity under pressure, and then you start wiring in autonomous agents. It's not hard to see why people talk about an "identity crisis" around AI.

The shadow AI problem: the agents you didn't sign off on

There's another twist. Not all AI in your organisation is sanctioned.

Shadow AI is the label being used for AI tools, agents and plugins staff use on their own, browser extensions, SaaS assistants, and personal accounts hooked up to corporate data. Shadow AI studies paint a pretty clear picture: big chunks of the workforce are using AI tools at work, and only a minority of organisations have mature AI governance in place. One 2026 report notes that 63% of organisations lack formal AI governance and that shadow AI is involved in a significant share of breaches.

Salesforce's workforce survey found 67% of employees using AI tools at work, but only 18% of organisations with formal AI security policies. That's a huge gap.

From a board or owner's perspective, shadow AI is where governance quietly breaks. You can't track risk properly if you don't know which tools are in play, what data they touch, or how they're authorised. AI governance pieces aimed at boards now call shadow AI one of the fastest-growing attack surfaces, not because the tools are evil, but because they're invisible.

What AI agents actually do to your risk story

Strip away the buzzwords and you're left with some very practical problems.

An agent updates a production config. Something goes wrong. You dig into logs and find a generic service account or a long-lived API token that three different workflows share. There's no obvious link to a named agent, a specific user, or a policy that explains why that change was allowed.

Or you see an agent querying customer data across regions. The agent was given broad read rights "for flexibility" and now it's creating exports that don't sit comfortably with your privacy obligations or regional data controls.

Maybe you discover that MFA was turned off for a class of non-human identities because “agents can’t click push prompts”, leaving a hole in your authentication story. Some IAM guidance explicitly warns that organisations disable MFA to let agents through, which undermines identity protection at exactly the wrong time.

And sitting around all that, you’ve got staff quietly using third-party AI tools with sensitive documents, because they’re trying to get work done. Shadow AI reports describe that pattern in detail: more tools than anyone realises, more data flowing out than anyone intended.

None of this is exotic. It’s day-to-day risk drift.

Moving from “let’s see what happens” to “we’ve got a plan”

There isn’t a silver bullet. But you can make your environment a lot more predictable with some comparatively straightforward moves.

One big shift the Cloud Security Alliance and others keep coming back to: treat AI agents as first-class identities. Not as shared service accounts. Not as invisible runtime glue. As named actors with owners, permissions and lifecycle.

That means you can answer basic questions:

- Which agents exist?

- Who's responsible for each one?
- Which systems and data can they touch?

From there, the better patterns all look suspiciously like good IAM hygiene, just applied to a new kind of "user".

Instead of giving agents human credentials, you give them delegated tokens with tight scope, so it's clear they're acting on behalf of someone, within a particular policy. Instead of long-lived permissions, you start moving towards short-lived access that's granted per task and revoked quickly. Instead of letting agents do anything they can technically do, you define boundaries: routine tasks they can run, sensitive operations that need human approval, and a small set of actions that stay manual for now.

And rather than logging everything and hoping, you build enough visibility to answer three questions for any agent interaction: who is this agent, who authorised it to act, and what exactly was it permitted to do? That's not my framing; it's straight out of recent agentic IAM work.

It's not perfect, but it's a clear step up from "the bot has admin access, let's see how it goes".

Why this isn't just plumbing

If you sit on a board or own a business, it's easy to file all of this under "technical detail". That would be a mistake.

Identity-driven incidents are already where most of the pain lives. Various compilations of breach data put identity issues behind the overwhelming majority of attacks, and phishing alone triggers a huge share of human-associated breaches. IDC research sponsored by Commvault recently found that 90% of decision-makers say they need stronger identity management capabilities to handle risks from agentic AI.

At the same time, AI budgets are going up. PwC and Deloitte note that executives expect AI agents to lift productivity and decision speed, and most plan to increase AI spend as a result. VentureBeat summarised survey findings along similar lines: nearly two-thirds of organisations using agents report productivity gains, more than half report cost savings.

So you've got three forces colliding:

- Strong business appetite for AI agents.
- Identity under pressure.
- Shadow AI creeping in around the edges.

If you don't treat identity and AI governance as part of the same conversation, you'll still get the costs. You just won't get the trust.

What to ask your team on Monday

You don’t need a full framework on day one. Start by asking a few blunt questions and listening carefully to how specific the answers are.

- **Where are AI agents actually running, and who owns each one?**
  Recent AI agent governance notes suggest that more than 90% of organisations lack full visibility into their agent identities, which makes it hard to claim you’re really in control.
- **What data can each agent touch, and how was that scope decided?**
  CIO-oriented security checklists stress the need for a concrete inventory of systems and data sensitivity before any agent is switched on, not after an incident prompts a review.
- **Are our agents sitting on broad, standing permissions, or using short-lived, delegated access?**
  Agentic IAM guidance leans strongly towards just-in-time, per-task access for agents; if the answer sounds closer to “we gave it admin and left it there”, you’ve got work to do.
- **How do we see what an agent actually did last week?**

Board-level AI security questions now include whether management can reconstruct which actions came from humans and which came from agents, with clear attribution; if you can't get that story from your logs, response and recovery will be shaky.

- **What's our plan for shadow AI — the tools and plugins we didn't formally approve?**
  Shadow AI studies show unmanaged tools as a major emerging risk, and several governance guides recommend a specific shadow AI audit to map usage before trying to write policy.

Listen for detail. Names. Systems. Timeframes. If you only hear aspirations and vendor buzzwords, it's not a signal to kill innovation. It's a sign you don't yet have enough structure around identity and AI to defend your choices.

Used the right way, AI agents are genuinely useful. PwC's survey suggests that about two-thirds of organisations using them see productivity go up and a solid slice report cost savings and better customer experiences.

The trick is making sure the story you tell about those gains doesn't fall apart under basic questions about who your agents are, what they can do, and how you'd cope if one went off script.