

You'd Never Hire Someone Without a Background Check. So Why Are You Deploying AI Agents Without One?

Enterprises spent decades building rigorous screening for human hires. The AI agents now working alongside those employees are getting a much easier ride.

SettleTop | September 2026

Before a new employee's first day, a lot happens. Identity is verified, work history is checked, references are called, and access is scoped to the job rather than to everything. A manager is named, and there is a probation period. In SHRM's survey of 6,500 HR professionals, **92% of organizations said they run background checks**, mostly before the hire.¹ The survey was fielded in 2018.

Now look at how the newest "workers" arrive. An AI agent shows up from a vendor, an open-source marketplace, or an internal team's side project. It gets an API key, a set of permissions, and a place in production. It has no references and no track record, and often no named owner. And it can act across dozens of systems at machine speed.

That is a double standard, and the data suggests it is already costing enterprises.

14.4%

of teams have full security approval for their AI agents, even though 81% are past the planning phase

Gravitee, 2026

109 : 1

machine identities (including AI agents) per human, up from 82 to 1 a year earlier

Palo Alto Networks, 2026

97%

of organizations with an AI-related breach lacked proper AI access controls

IBM Cost of a Data Breach, 2025

The gap between deployment and vetting

Agents are now production infrastructure. SAP LeanIX's 2026 survey found that **98% of companies have deployed AI agents or plan to**, yet fewer than half have visibility into an inventory of them.² Gartner projects that **40% of enterprise applications will embed task-specific agents by the end of 2026**, up from under 5% in 2025.³

Gravitee's 2026 survey of more than 900 executives and practitioners makes the gap concrete: **81% of teams are past the planning phase, but only 14.4% have full security approval.**⁴ Compare that with the 92% screening rate for people. Human screening has gaps too, and SHRM found that only 4% of organizations run continuous checks.¹ But for people a baseline exists. For agents, it mostly does not.

They are not tools. They are workers with keys.

Identity teams already see the shift. Palo Alto Networks' 2026 Identity Security Landscape puts machine identities, including AI agents, at **109 for every human, up from 82 to 1 a year earlier.**⁵ KPMG's 2026 report says 80 to 1,⁶ and the Cloud Security Alliance (CSA) cites an average of 45 to 1, reaching 144 to 1 in cloud-native environments.⁷ The methods differ, but every estimate points the same way.

Governance has not kept up. Okta's 2025 survey found that 91% of organizations were already using AI agents, while only 10% had a well-developed strategy for managing non-human identities.⁸ Gravitee found that only 22% of teams treat agents as independent identities, and most still rely on shared API keys.⁴

The CSA notes that agents can query databases, send emails, execute code, and modify cloud configurations, often with the same permissions as the human who provisioned them.³ A person given that access would have been screened first.

The incidents are already here

In Gravitee's data, **88% of organizations confirmed or suspected a security incident involving AI agents** this year.⁴ IBM's 2025 Cost of a Data Breach Report found that 13% of organizations reported breaches involving their AI models or applications, and **97% of those lacked proper AI access controls**. IBM also found that 63% of breached organizations had no AI governance policies.⁹

The financial impact is showing up too. An EY/AIUC-1 survey found that **64% of companies with revenue above \$1 billion reported losses over \$1 million** tied to AI system failures in 2025, and 80% of organizations documented risky agent behavior such as unauthorized access and data exposure.³

Third-party agents: the résumé nobody checked

Third-party agent ecosystems are where the missing vetting shows most clearly.

- **Vulnerable by default.** An empirical study of 42,447 agent skills found that 26.1% had at least one security vulnerability. Separately, the ClawHavoc campaign placed more than 1,200 malicious skills in the OpenClaw marketplace.¹⁰
- **A fake candidate gets hired.** In June 2026, a security firm reportedly built a harmless fake agent skill, got it past Cisco and NVIDIA scanners, promoted it through a marketplace and an Instagram ad, and reached roughly 26,000 agents, including some on corporate accounts.¹¹
- **No documentation.** A CSA Labs scan of 1,899 MCP servers, as reported by DEV Community, found that 5.5% showed tool poisoning and none included security documentation.¹²
- **Résumé inflation.** Gartner warns of "agent washing," the rebranding of chatbots and RPA as agents, and estimates that only about 130 of the thousands of agentic AI vendors are real.¹³
- **Thin vendor oversight.** Per EY data compiled by Secureframe, 23% of organizations still do not monitor vendor AI usage, down from 37% in 2024.¹⁵

The pace matters as well. An analysis of PyPI downloads found that agent frameworks pulled 483 million downloads in May 2026, against 5.8 million for agent security and guardrail packages: **83 to 1**.¹⁶

Transparency: the missing reference check

A reference check works because someone independent can vouch for the candidate. AI transparency is moving the wrong way. Stanford's 2025 Foundation Model Transparency Index found that the **average score fell from 58 to 40 out of 100**, with companies most opaque about training data and post-deployment usage and impact. The spread is wide: IBM scored 95, while xAI and Midjourney scored 14. That spread shows transparency is a choice, and enterprise buyers can reward the vendors who make it.¹⁷

Inside enterprises the picture is similar. Gravitee found that only 7.2% of organizations have a named individual with formal accountability for agent behavior, and mean monitoring coverage is 52%, so nearly half of production agents run unsecured. Respondents also described third-party AI tools mishandling data, changing processing logic without notice, or offering no clarity on where data is stored.⁴

If a job candidate said “trust us,” you would keep looking. Agents should not get a pass.

A registry model is starting to emerge

This is not only a thought experiment. In September 2026, SettleTop launched a public AI Registry that applies the same vetting discipline to the agent economy itself. It tracks more than 12,000 agents and AI apps from nearly 5,900 publishers across marketplaces including AWS and Microsoft, and documents each one’s origin, composition, and disclosure status.¹⁸

Each listing carries what SettleTop calls an **Agent Passport**: the underlying model, framework, tools, data sources, hosting environment, data residency, and permission scope. Claims are scored across five provenance dimensions, vendor, build, reach, data, and compliance, and each data point is marked **Verified** (independently confirmed), **Disclosed** (vendor-stated but unverified), or **Unknown**, rather than quietly assumed. “Enterprises are adopting AI agents faster than their procurement and security teams can evaluate them,” says SettleTop Chief Product Officer Niles Madison, describing the same adoption-governance gap this piece has traced through Gravitee, SAP, and Okta’s survey data.¹⁸

That three-tier model, verified, disclosed, unknown, is a workable template for the “security posture” and “identity and provenance” rows in the checklist below: it turns a vendor questionnaire into evidence someone else can audit, rather than a claim an enterprise has to take on faith.

Internal agents need vetting too

“We built it ourselves” is not a security control. IBM found that **20% of organizations studied reported a breach caused by shadow AI**, adding an average of \$670,000 in breach costs.⁹ As SAP describes it, agent sprawl happens when agents are created and connected faster than the enterprise can inventory them or assign owners.² Companies re-screen people who move into sensitive roles, and an internal agent that gains access to finance or customer data deserves the same scrutiny.

What an AI agent background check looks like

Each step maps to something HR already does:

Human hiring step	AI agent equivalent	What to verify
Identity check	Identity and provenance	Who built it, which model and data sit underneath, and who owns it. Keep a registry and an AI bill of materials.
References	Track record	Independent evaluations, incident history, and customer references.
Background check	Security posture	Red-team and prompt-injection test results, supply chain scans, and audit reports.

Human hiring step	AI agent equivalent	What to verify
Confidentiality agreement	Data handling	Retention, training use, sub-processors, and how changes are communicated.
Role-based access	Least privilege	A distinct identity per agent, no shared keys, and no standing admin rights.
Onboarding and probation	Staged rollout	Sandbox or shadow mode first. Classify agents by autonomy level, each with its own trust boundary and governance requirements. ¹⁴
Performance reviews	Continuous monitoring	Track behavior drift and tool changes, and re-screen whenever an agent's scope changes.
Exit process	Offboarding	A kill switch, credential revocation, and decommissioning ready before launch.

Vetting keeps projects alive

Some will call this friction. Gartner predicts that **over 40% of agentic AI projects will be canceled by the end of 2027** due to escalating costs, unclear business value, or inadequate risk controls.¹³ Vetting is what keeps projects from landing in that 40%.

Three questions to ask this week:

- How many agents are running in our environment?
- Who owns each one?
- If one were compromised today, what would we know, and how fast?

If you cannot answer them, the background check has not happened.

A note on the data. Many of these figures come from vendor-sponsored surveys with differing methodologies, so read them as directional rather than definitive.

Sources

1. [SHRM: Employers Slow to Pick Up Trend of Continuous Screening](#) (survey fielded July 2018).
2. [SAP News Center: AI Agent Sprawl: Why AI Governance Is Now a Board-Level Issue](#) (August 2026).
3. [Cloud Security Alliance Labs: The AI Agent Governance Gap: What CISOs Need Now](#) (cites Gartner and the EY/AIUC-1 Consortium survey).
4. [Gravitee: State of AI Agent Security 2026: April 2026 survey data.](#)
5. [Palo Alto Networks: 2026 Identity Security Landscape.](#)
6. [NHIMG: KPMG 2026 Cybersecurity Report summary.](#)
7. [CSA Labs: The Non-Human Identity Governance Vacuum](#) (May 2026).
8. [Okta: AI Agents at Work 2026](#) (cites Okta's 2025 survey findings).
9. [IBM: 2025 Cost of a Data Breach Report; Nudge Security summary; Hack The Box summary.](#)
10. [arXiv: Formal Analysis and Supply Chain Security for Agentic AI Skills](#) (February 2026).
11. [SecureWorld: We Spent 20 Years Securing the Supply Chain. AI Agents Just Reset the Clock to Zero](#) (July 2026).

12. [DEV Community: Malicious MCP Servers and the Unsolved Supply Chain Problem for AI Agents](#) (reports a May 2026 CSA Labs scan).
13. [Gartner: Over 40% of Agentic AI Projects Will Be Canceled by End of 2027](#) (June 2025; also the source for the agent-washing estimate).
14. [Gartner: Applying Uniform Governance Across AI Agents Will Lead to Enterprise AI Agent Failure](#) (May 2026).
15. [Secureframe: 100+ Essential Third-Party Risk Statistics](#) (compiles EY data).
16. [Security Boulevard: The Agent Identity Problem](#) (July 2026).
17. [Stanford CRFM: 2025 Foundation Model Transparency Index](#); [paper on arXiv](#).
18. [SettleTop: AI Registry launch announcement](#); registry.settletop.com (September 2026).