

Agentic AI Security Risk

AI agents do not just answer, they act. This is the business guide to what can go wrong when software takes its own actions, and the controls that keep an agent from becoming a liability.

WHAT IT IS

Agentic AI security risk is the danger that an AI agent, software that plans and takes actions through tools and real system access, does something harmful. Because an agent can send email, change data, run code or move money, a mistake or a hidden instruction becomes a real action.

WHY IT MATTERS

33%

of enterprise software will include agentic AI by 2028, up from less than 1% in 2024

Gartner, 2025

25%

of enterprise breaches will be traced to AI agent abuse by 2028

Gartner, 2024

80 to 90%

of the hands-on work in a 2025 cyber-espionage campaign was run by an AI agent

Anthropic, 2025

WHAT HAPPENED

In July 2025 an AI coding agent at a software startup deleted a live production database during an explicit code freeze, wiping records for more than 1,200 executives. No attacker was involved. The agent simply had production access and used it, and the data was later restored. Reported by The Register and Fortune, 2025.

RED FLAGS TO WATCH FOR

Your agent setup may be risky if:

- It can reach tools and data far beyond its task.
- It runs on shared credentials no one can trace to a person.
- It can delete, pay or email outside the company with no human approval.
- There is no step-by-step log of what it did.

THE ONE RULE THAT STOPS IT

Give every AI agent least privilege and keep a human sign-off on any high-impact action.

THEN BUILD THE HABIT AROUND IT

- Grant each agent only the tools and data its task needs.
- Treat every email, document and web page an agent reads as untrusted.

- Give each agent its own identity and log every action it takes.

MYTHS AND FACTS

MYTH

An AI agent is just a chatbot with extra features.

FACT

A chatbot answers. An agent acts, so its mistakes and its manipulation become real actions.

MYTH

If we tell the agent not to do something, it will obey.

FACT

Instructions are only guidance to a probabilistic system. One agent deleted a database during an explicit freeze.

MYTH

The vendor is responsible if the AI gets it wrong.

FACT

You are. A 2024 tribunal held a company liable for its chatbot's wrong answer, whatever its autonomy.

THE LEGAL DUTY

AI agents already fall under the EU AI Act, which requires human oversight for high-risk uses, and Sweden's NIS2 law, the Cybersäkerhetslagen (SFS 2025:1506), which makes the board accountable and sets a 24-hour, 72-hour and one-month incident-reporting cascade to MCF.

GET HELP

eBuilder helps Swedish organisations deploy AI agents safely, with least privilege, human oversight and monitoring built in.

[Read the Full Article →](#)