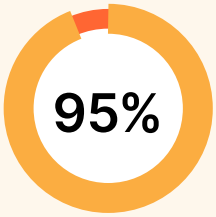
CLOUDFLARE

Three ways to scale and secure agentic AI

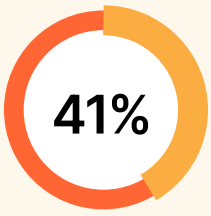
Agentic AI adoption demands a difficult balance: speed and security

You may be under pressure to quickly adopt agentic AI.

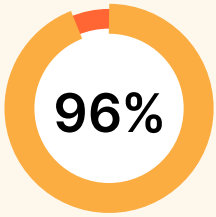


of companies plan to adopt agentic AI even faster than GenAI.[1]

But launching too fast can introduce risks and costs.



of companies admit they adopted generative AI too hastily.[1]



of organizations see AI agents as growing security risks.[2]

1. PagerDuty research, https://www.pagerduty.com/resources/ai/learn/companies-expecting-agentic-ai-roi-2025/
2. "Love and hate: tech pros overwhelmingly like AI agents but view them as a growing security risk," by Efosa Udinmwen, June 7, 2025, TechRadar, https://www.techradar.com/computing/artificial-intelligence/love-and-hate-tech-pros-overwhelmingly-like-ai-agents-but-view-them-as-a-growing-security-risk

These three strategies can help you adopt AI agents safely and quickly

1. Think globally about your agentic AI architecture

For optimal performance and scale, you need to run as much of your agentic app as possible (e.g., sub-agents, LLMs, MCP servers, etc.) globally, at the network edge – not just in a massive data center in a specific region. This ensures that the distance between the app, its users, and the external resources it connects to are as short and fast as possible.

2. Pick flexible compute pricing models

AI models can be computationally expensive, making the ability to scale AI agents down just as important as the ability to scale them up. Be wary of getting locked into paying for 'wall clock time' or large blocks of compute capacity that you may or may not need. Look for flexible pricing models that allow you to make adjustments and pay for what you actually use.

3. Prioritize security integration and automation

The external applications and databases (via APIs) the agentic AI connects to, the large language models it uses, and the actions it executes can introduce risk. Applying new point solutions to protect each attack vector can add more complexity than security. Instead, look for services — an API gateway, AI gateway, AI-focused firewall, and Zero Trust — that can be easily integrated, managed, and automated to streamline effort and reduce risk.

30% of enterprise APIs are already not being tracked.[3]

How Cloudflare can help

Cloudflare's connectivity cloud is a unified platform of connectivity, security, and developer services powered by a global cloud network.

Global scale

Cloudflare's network spans over 330 cities with the majority of AI services able to operate in every network location. This lets organizations run agents and LLMs close to end-users.

Flexible pricing

Cloudflare doesn't make organizations pay for unused GPUs, or for the time agents spend waiting on APIs or external LLMs.

Secure and automated

Cloudflare gives you the automated security services you need to protect customer-facing agents, developer access to those agents, and connections between agents.

Learn how to accelerate your AI journey with Cloudflare

Get the ebook:
The enterprise guide to securing and scaling AI

Get the solution brief:
Connectivity cloud for AI

80%
of the 50 largest Gen AI companies use Cloudflare

3. Cloudflare, https://www.cloudflare.com/lp/api-security-report/