

AI Gateway

Product Overview

SOLUTION OVERVIEW FOR AKAMAI CONNECTED CLOUD

What is an AI Gateway?

Businesses building AI-powered applications face new risks when coding against LLM APIs like ChatGPT, Gemini, and Claude—spanning security, cost control, and observability. An AI gateway serves as your central control plane to manage usage and minimize these risks.

Why use an AI Gateway?

- **Virtualize access to LLMs**
Control which LLMs teams use and enable rapid model adoption without code changes. Eliminate API key risks with virtual access points. Dynamically route based on capabilities and costs, with failover protection for reliability.
- **Manage costs**
Set hierarchical limits and warnings—for example, \$10,000 monthly for departments and \$100 daily for applications—to prevent accidental overspend during development.
- **Enforce security guardrails**
Block PII leaks and prompt injection attacks. Filter toxic content and unwanted elements in prompts and responses based on your requirements.
- **Observability**
Real-time logging, analytics, and reporting for usage patterns, costs, and LLM monitoring across all organizational levels and applications.

Inside Akamai AI and API Manager

The AI Gateway is one capability of Akamai AI and API Manager (AAM)—a single platform spanning API management, instant MCP servers, and the MCP Gateway. It combines self-serve productivity with organizational governance for an enterprise-grade experience.

- **Self-serve access with governance controls**
Create team hierarchies with enforced policies and cost controls. Team members build virtual applications that access LLMs within company policy.
- **Dynamic AI model routing**
Configure LLM failover and dynamic load balancing. Test new LLMs without changing application code.
- **Cost and usage controls**
Define cost, token, and request limits by day or month at multiple levels. Trigger warnings at configurable thresholds. Enable individual developer limits to prevent overspend.
- **Firewall your AI usage**
Akamai AI Firewall powers guardrails including PII detection and prompt injection protection.
- **Observability**
Dashboard logging plus native integrations with preferred providers for real-time AI usage analysis.
- **Bank-grade security**
SOC 2 Type 2 certified with OAuth, JWT, mTLS, and fine-grained authorization policies.