



NEWS RELEASE

F5 Unleashes Next-Generation, Agentic-Ready AI Gateway to Optimize the Economics and Governance of Enterprise AI Costs

2026-08-18

Now part of the F5 AI Security Platform, new capabilities provide a unified control point for AI tokenomics, policy enforcement, and security across models, agents, and tools

SEATTLE--(BUSINESS WIRE)-- F5 (NASDAQ: FFIV), the global leader in delivering and securing every app and API, today introduced significant enhancements to the F5 AI Gateway and integrated the solution into the **F5 AI Security Platform**. The enhanced F5 AI Gateway seamlessly enforces policies on every AI request, giving enterprises a unified control plane to govern how AI models, agents, and tools are accessed and used, while optimizing the economics of AI at scale.

Enterprises have moved past AI experimentation, but governance has not kept pace. AI traffic still moves through a patchwork of standalone proxies and monitoring tools that were never built for AI, with no guardrails in between. According to F5's 2026 State of Application Strategy Report, 77% of organizations say inference, rather than model training or tuning, is now their dominant AI activity, and organizations are managing an average of seven AI models.

As inference scales, tokenomics is becoming an increasingly important consideration, with every model request carrying implications for cost, performance, and security. Yet the piecemeal approach to standalone tools to secure AI traffic leaves enterprises without consistent control over how models and agents are accessed and used.

"We're watching enterprises race to deploy AI while struggling to control it," said Kunal Anand, Chief Product Officer at F5. "Every AI request carries economic, security, and governance implications, yet most organizations are relying on fragmented tools that address only part of the problem. The result is rising costs, increased risk, and operational complexity. F5 AI Gateway, integrated into the F5 AI Security Platform, provides a single control point for managing AI across models, clouds, agents, and applications. We believe every enterprise will need an intelligent control layer for AI. F5 is building that foundation, helping customers accelerate innovation while maintaining visibility, security, and control."

F5 AI Gateway brings three critical functions together in a single integrated solution:

- Model Gateway for model access and cost optimization
- MCP Gateway for agent-to-tool governance
- AI Guardrails for prompt and response protection

Budgets, model routing policies and agent access controls are set once centrally and enforced across distributed environments wherever the models, agents and AI apps run, providing a single operations pane for AI platform ops, AI Security ops and finance teams.

Rapid adoption of AI is fueling the need for AI gateways. According to a recent Gartner® report, "AI gateways have emerged as a critical part of AI infrastructure, as enterprises need tools to support safe, efficient and controlled access to AI models and MCP servers. Adoption of AI gateways will continue to accelerate among large enterprises."¹

Bringing AI costs under control

Organizations that cannot see which teams, models, and providers are consuming tokens cannot assess the value and costs of using AI. F5 AI Gateway puts tokenomics under control: the Model Gateway function attributes every token by provider, model, team, and user, per-team budgets enforce limits as spend occurs rather than after the invoice arrives; and automated optimization — smart routing and model tiering, semantic caching, and GPU-aware load balancing — routes each request to the right model at the right cost. The solution is designed to reduce token spend by up to 60 percent, with no application changes.

Governing agents and controlling tool access

As agents multiply, so do the MCP servers they call, usually with no central registry, no per-tool authorization, and no record of what agents are doing. The MCP Gateway function of F5 AI Gateway provides fine-grained access controls that limit agents to the resources they are explicitly authorized to use, including APIs, data sources, and

RAG systems. A complete audit trail captures what was accessed, when, by which agent, and on whose behalf. MCP server registry gives teams a single source of truth for approved MCP servers, thereby removing friction for developers who would otherwise find it challenging to discover which servers and tools exist.

Protecting AI data flows and proving compliance

Personal data, health records, and intellectual property move through AI applications every day, and that data usually reaches external models uninspected while injection and jailbreak attempts go undetected. F5 AI Guardrails inspect every prompt and response, redacting sensitive data before it reaches the model, blocking injection and jailbreak attempts, and failing closed when a request cannot be evaluated. Full audit trails, SIEM export, data residency controls, and alignment with SOC 2, ISO, and HIPAA frameworks give regulated industries the evidence they need before putting AI into production.

The enforcement point for the F5 AI Security Platform

F5 introduced the F5 AI Security Platform earlier this year to give teams continuous visibility, governance, and protection across enterprise AI applications, models, agents, and the APIs connecting them. Four integrated pillars — AI governance, AI usage control, AI security testing, and AI runtime protection — plus an overarching observability layer, create a persistent security lifecycle rather than a one-time compliance exercise. F5 AI Gateway is where those pillars meet live traffic, putting policy into force at the point of interaction and controlling what AI can access, what it can expose, and, crucially, what it can cost the business. F5 AI Gateway is deployable across SaaS, hybrid SaaS, and hybrid multicloud environments, with air-gapped support planned for regulated and sovereign use cases.

Supporting resources

- [Blog:F5 AI Gateway: A single control point for all enterprise AI](#)
- [Product page:F5 AI Gateway](#)
- [F5 AI Gateway Early Access](#)

About F5

F5, Inc. (NASDAQ: FFIV) is the global leader that delivers and secures every app. Backed by three decades of expertise, F5 has built the industry's premier platform—F5 Application Delivery and Security Platform (ADSP)—to deliver and secure every app, every API, anywhere: on-premises, in the cloud, at the edge, and across hybrid, multicloud environments. F5 is committed to innovating and partnering with the world's largest and most advanced organizations to deliver fast, available, and secure digital experiences. Together, we help each other thrive and bring a better digital world to life.

For more information visit **f5.com**

Explore F5 Labs threat research at **f5.com/labs**

Follow to learn more about F5, our partners, and technologies: **Blog** | **LinkedIn** | **X** | **YouTube** | **Instagram** | **Facebook**

Forward-Looking Statements. This press release contains forward-looking statements, including regarding the expected availability, timing, functionality, and benefits of the F5 AI Gateway and F5 AI Security Platform, made under the safe harbor provisions of the Private Securities Litigation Reform Act of 1995. Actual results could differ materially due to the risks described in F5's SEC filings, including its most recent Forms 10-K and 10-Q. F5 undertakes no obligation to update these statements.

F5 is a trademark, service mark, or tradename of F5, Inc. or its affiliates in the U.S. and other countries. All other product and company names herein may be trademarks of their respective owners.

Source: F5, Inc.

1 Gartner, Market Overview for AI Gateways, Keith Guttridge, Andrew Humphreys, Shiva Varma, 13 May 2026. GARTNER is a trademark of Gartner, Inc. and/or its affiliates.

Dan Sorensen

F5

(650) 228-4842

d.sorensen@f5.com

Holly Lancaster

We. Communications

(415) 547-7054

hlancaster@wecommunications.com

Source: F5, Inc.