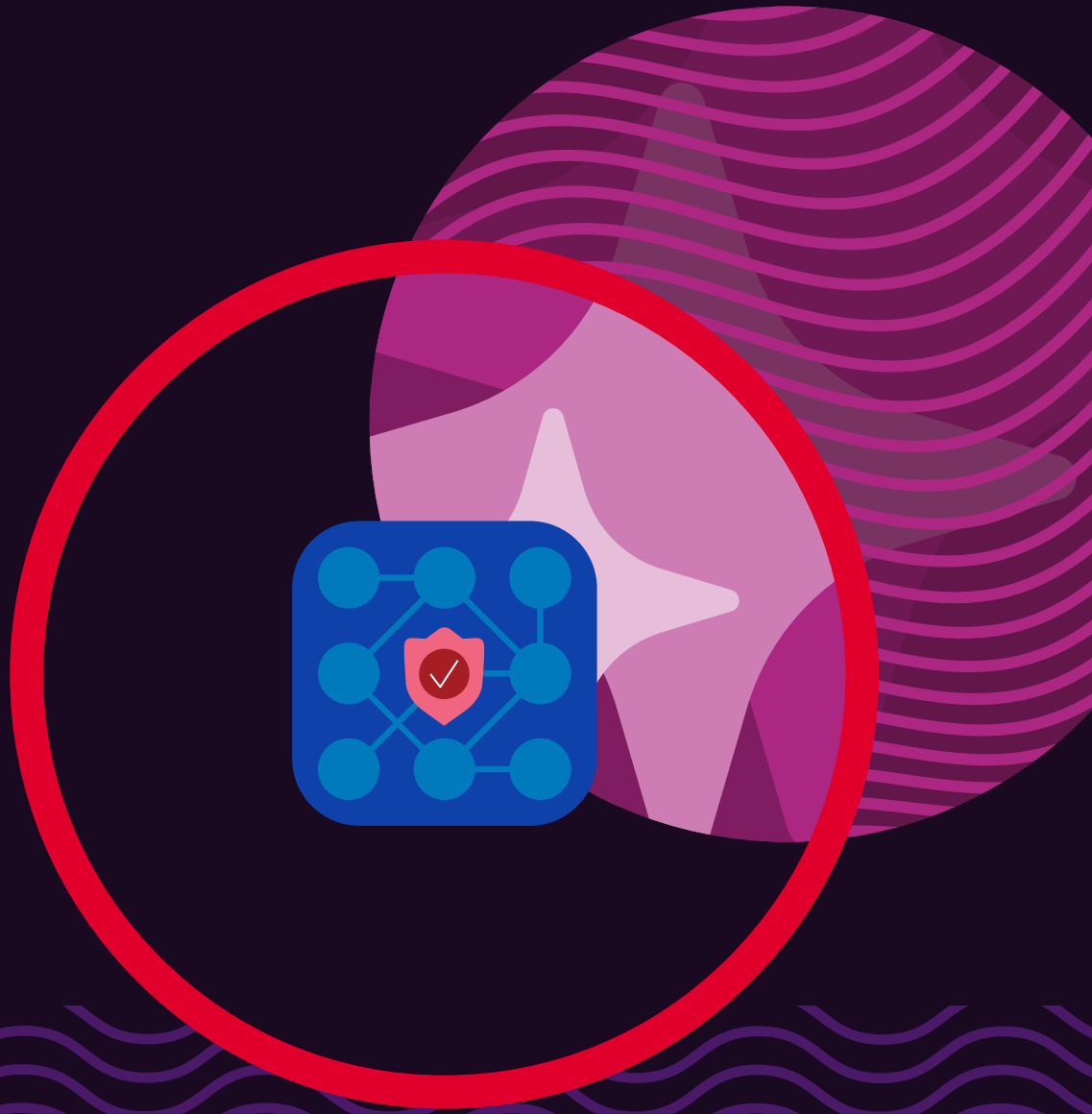
Google Cloud

Secure your AI applications with advanced API protection



Discover, protect, and optimize APIs powering your AI applications at scale with F5 and Google Cloud.

Contents

AI and APIs are inextricably linked

The integration of AI capabilities into enterprise applications has led to an exponential rise in API usage. APIs serve as the connective tissue that links AI models with applications and the data they process for training and inference. From fine-tuning a model to implementing retrieval-augmented generation (RAG), these processes create new API endpoints that must be secured.

A single AI application may ultimately have hundreds of endpoints, with API calls buried deep inside business logic that remains invisible to security teams. According to Gartner, more than 30% of the increase in demand for APIs will come from AI and tools using large language models (LLMs) by 2026.[1] This explosive growth expands the attack surface, providing more opportunities for threat actors to exploit APIs to exfiltrate data, manipulate AI models, or steal intellectual property.

The stakes are high because AI applications represent significant investments in both computational resources and proprietary data. Without strong API security, organizations face multiple risks, including data theft through model extraction, system compromise via prompt injection, service disruption through API abuse, and unauthorized access to valuable AI capabilities.

A new approach is needed

F5 and Google Cloud have partnered to secure your AI journey. Joint solutions for API security combine Google Cloud's AI-optimized infrastructure and security with F5's advanced discovery, protection, and management capabilities. We'll explore how this partnership addresses AI-specific threats, discovers hidden API risks, and provides unified protection across your entire AI ecosystem.

AI-specific API security threats you need to address

Attackers exploit the complex nature of AI infrastructure and the multiple API interfaces required for training, inference, and data exchange to find gaps in your security defenses. The Open Worldwide Application Security Project (OWASP) has identified the top risks to AI apps in the OWASP Top 10 for LLM Applications, many of which require robust API security to mitigate, including:

Prompt injection, which occurs when attackers manipulate API inputs to make AI models disregard safety controls or expose sensitive information. Through carefully crafted prompts, attackers can alter model behavior or extract proprietary data via downstream APIs.

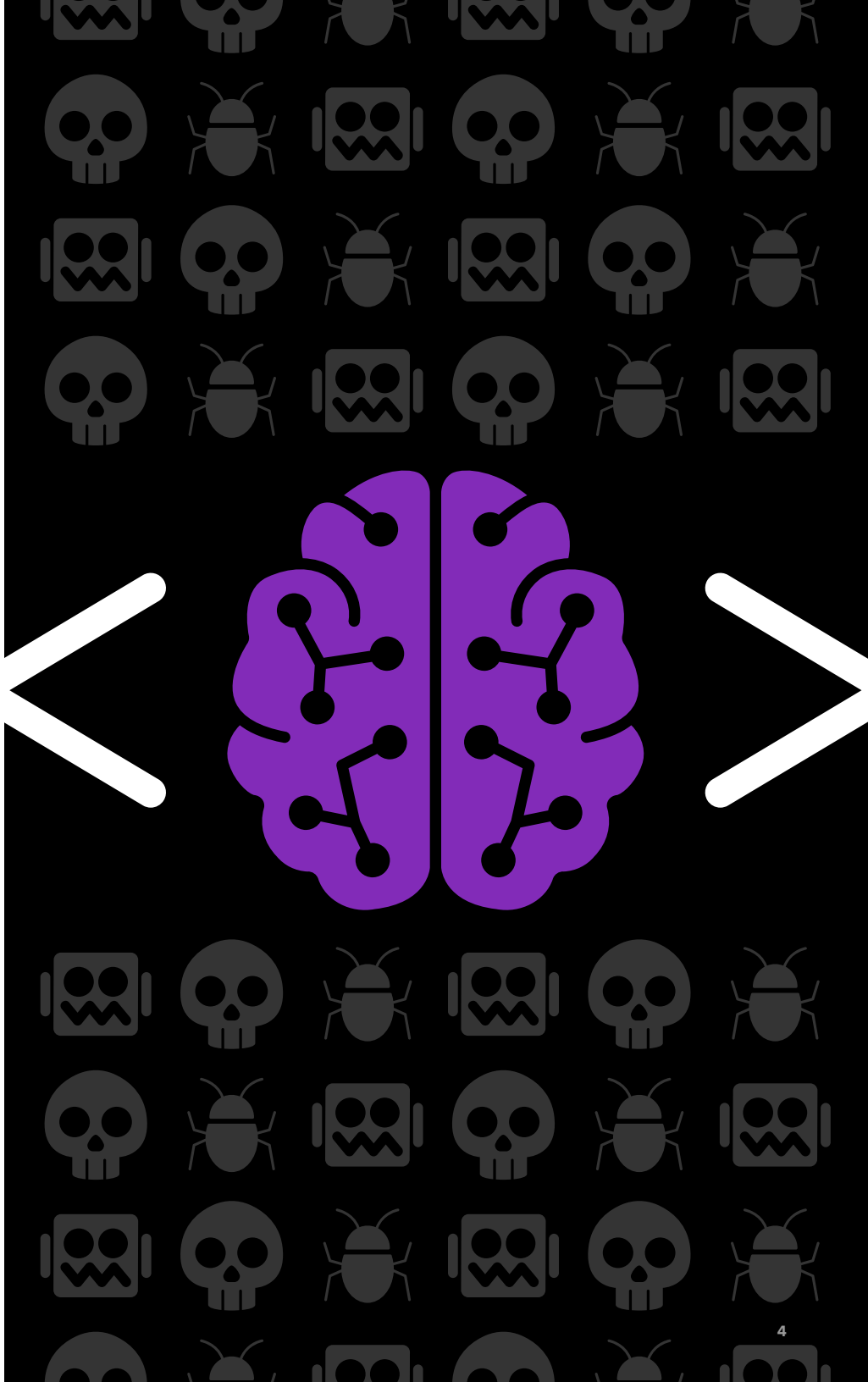
Unbounded consumption from APIs that lack proper rate limiting, allowing attackers to overwhelm AI models with requests. This leads to service degradation, outages, or high costs due to resource consumption. Models may even be extracted via carefully crafted API inputs.

System prompt leakage that occurs when AI systems accidentally reveal their internal instructions, configuration details, or operational parameters through API responses. This vulnerability can expose sensitive system information for attackers to exploit.

Shadow APIs multiply your risk

When APIs exist outside the knowledge and control of security teams, these unprotected shadow APIs are an open invitation for attackers. Often, shadow APIs are created by eager development teams building model access, training, or data collection without proper security oversight. Rapid development causes unauthorized or undocumented APIs to proliferate, creating version control chaos, integration blind spots, and incomplete organizational risk assessments.

You can't protect what you can't see. Complete visibility into API endpoints is essential to secure your AI investments.



Traditional API security falls short for AI workloads

Most API security solutions were designed around predictable, structured request-response patterns. AI applications fundamentally challenge these assumptions through several key differences:

Massive API proliferation: Where conventional applications might have dozens of API endpoints, AI systems can generate hundreds. Each AI model, data connection, and processing step typically requires its own API interface, creating complexity that overwhelms traditional discovery and management approaches.

Complex authentication patterns: AI applications often involve machine-to-machine communication between models, data sources, and processing services. These automated interactions require sophisticated authentication mechanisms that go beyond simple user-based access controls.

High-volume, variable traffic: AI API calls can involve large payloads containing training data, model parameters, or multimodal content. Traffic patterns are unpredictable, with sudden spikes during model training or inference operations that can overwhelm rate limiting designed for typical web API usage.

Buried API dependencies: Unlike traditional applications where API calls are often explicit and documented, AI applications frequently generate API requests dynamically based on model decisions or data processing needs. This makes it difficult to map and secure the complete API ecosystem.

Cross-environment complexity: AI workloads are typically distributed across multiple environments, from public cloud services for model training, edge deployments for inference, and on-premises systems for sensitive data. Each environment may use different API management approaches, creating security gaps and inconsistent policies.

The distributed, API-centric nature of AI applications demands a new approach to security—one that can discover hidden endpoints, handle massive scale, and protect against both traditional API threats and AI-specific attack vectors.

Five essential elements of AI API security

Protecting your AI models and applications requires both API discovery and protection to address traditional risks and AI-specific threats.

01.

Continuous discovery maintains real-time visibility into all API endpoints that connect to AI services, including those created outside of approved development processes. This identifies unauthorized access points and potential vulnerabilities before they can be exploited.

02.

Authentication and access control ensures every API endpoint connecting to AI models or data has proper authentication controls and clearly defined access policies. This includes validating both human and machine identities while enforcing the principle of least privilege.

03.

Rate limiting sets and enforces specific thresholds for API calls, preventing both accidental and malicious resource consumption or denial-of-service attacks. Given the high computational demands of AI models, this protection is essential for controlling costs and maintaining service availability.

04.

Anomaly detection continuously monitors API traffic patterns to identify potential model theft attempts, system prompt leakage, and other suspicious behaviors that could indicate an ongoing attack or unauthorized access to your AI systems.

05.

Observability provides comprehensive visibility into API performance, usage patterns, and security events across your entire AI infrastructure. This includes real-time monitoring of API traffic, centralized logging for security analysis, and integration with existing analytics platforms to ensure consistent visibility across hybrid multicloud environments.

These five elements work together to create layered defenses that protect your valuable AI investments from both known threats and emerging attack vectors.

Protect AI with comprehensive API security

F5 and Google Cloud have joined forces to address the unique challenges of securing AI applications through integrated solutions that combine Google Cloud's AI-optimized infrastructure and security solutions with advanced API discovery and protection capabilities in the F5® Application Delivery and Security Platform (ADSP). This partnership delivers end-to-end protection across the entire AI application lifecycle.

Discover every API in your AI ecosystem
The first step of API security is knowing about every API in your environment, including shadow APIs and obsolete access points, in order to mitigate threats. Scan application traffic to discover APIs in running applications, and inspect source code to find risks before applications are deployed. F5 and Google Cloud work together to provide a holistic view of APIs across hybrid multicloud environments. You can also discover sensitive data in your environment and apply policies to prevent APIs from exposing it.

Protect against AI-specific and traditional threats
Automated protection responds to threats in real time. F5 solutions apply continuous machine learning to identify and block suspicious behavior, while also applying rate limiting to mitigate unbounded consumption and Layer 7 DoS attacks against APIs and applications. Strong authentication and access control mechanisms ensure that only authorized users and systems can access AI models and sensitive data, and granular OpenAPI Specification (OAS) enforcement validates all API requests against defined schemas, preventing malformed requests from reaching your AI models. API security policies are applied consistently across your entire AI infrastructure, in Google Cloud and beyond.

Monitor and visualize your security posture
Centralized observability provides actionable insights across all environments. F5 collects global API metrics into a centralized interface that integrates seamlessly with Google Cloud's observability suite and SIEM tools for real-time viewing, alerting, and analysis. For more advanced analytics, long-term storage, and complex querying, these logs can be efficiently loaded into BigQuery, where they can be visualized using Looker Studio.

Together, F5 and Google Cloud provide multi-layered API security to protect AI and other modern apps.

F5 and Google Cloud work together to strengthen AI and APIs

F5® Distributed Cloud API Security identifies APIs by analyzing live traffic to uncover shadow or undocumented APIs. Code repository scanning integrates with platforms like GitHub to identify APIs early in the development lifecycle, while client-side web crawling navigates dynamic AI frontends. Behavioral monitoring, risk scoring, and enforcement improve protection.

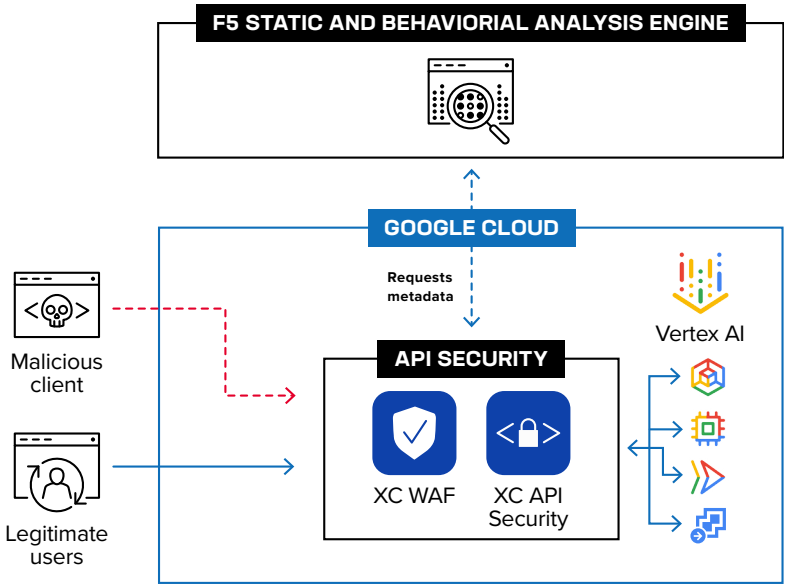
Google Cloud provides essential API security capabilities that integrate seamlessly with AI workloads. Native tooling in Google Cloud helps you build, manage, and secure APIs using AI and machine learning. It can detect undocumented and unmanaged APIs in your Google Cloud environment and identify critical API abuses.

Complementary strength through integration

Together, these solutions provide broad security coverage, leveraging Google Cloud's native tooling to manage APIs within Google Cloud infrastructure and F5 solutions to identify APIs across on-premises, multicloud, and third-party environments.

F5 security events integrate with Google Security Command Center and Google Security Operations for a holistic view of security that enables faster detection and response to threats. Native OpenTelemetry support in F5 solutions means logs flow seamlessly to Google Cloud Logging, Cloud Monitoring, and BigQuery for unified observability.

Multiple integration options between F5 and Google Cloud provide complete API security, regardless of where your AI applications are deployed.



Elevate your API security: The F5 and Google Cloud advantage

F5 and Google Cloud deliver integrated API discovery and protection that transforms how organizations approach AI security. This partnership enables businesses to secure their AI investments while maintaining agility by supporting:

Accelerated AI innovation with reduced security friction

Cost optimization through intelligent resource management

Risk reduction via improved visibility and dynamic security

Operational efficiency with unified management and observability

Ready to secure your AI applications?

Protect valuable AI investments from emerging threats with comprehensive API security. Scale AI initiatives across hybrid multicloud environments and maintain competitive advantage with secure, high-performance AI applications.

Learn more about F5 and Google Cloud at f5.com/gcp.

THE F5 AND GOOGLE CLOUD PARTNERSHIP

6+

years of collaboration

Over 50

listings in the Google Cloud Marketplace

Joint

expertise across security, networking, and industries

Appendix

[1] Gartner Press Release, Gartner Predicts More Than 30% of the Increase in Demand for APIs will Come From AI and Tools Using Large Language Models by 2026, Mar 2024

[2] F5, API Security Evaluation Guide, Dec 2023

ABOUT F5

F5 is a multicloud application delivery and security company committed to bringing a better digital world to life. F5 partners with the world's largest, most advanced organizations to secure every app—on premises, in the cloud, or at the edge. F5 enables businesses to continuously stay ahead of threats while delivering exceptional, secure digital experiences for their customers.

For more information, go to f5.com. (NASDAQ: FFIV)

Learn more about F5 and Google Cloud at **f5.com/gcp**