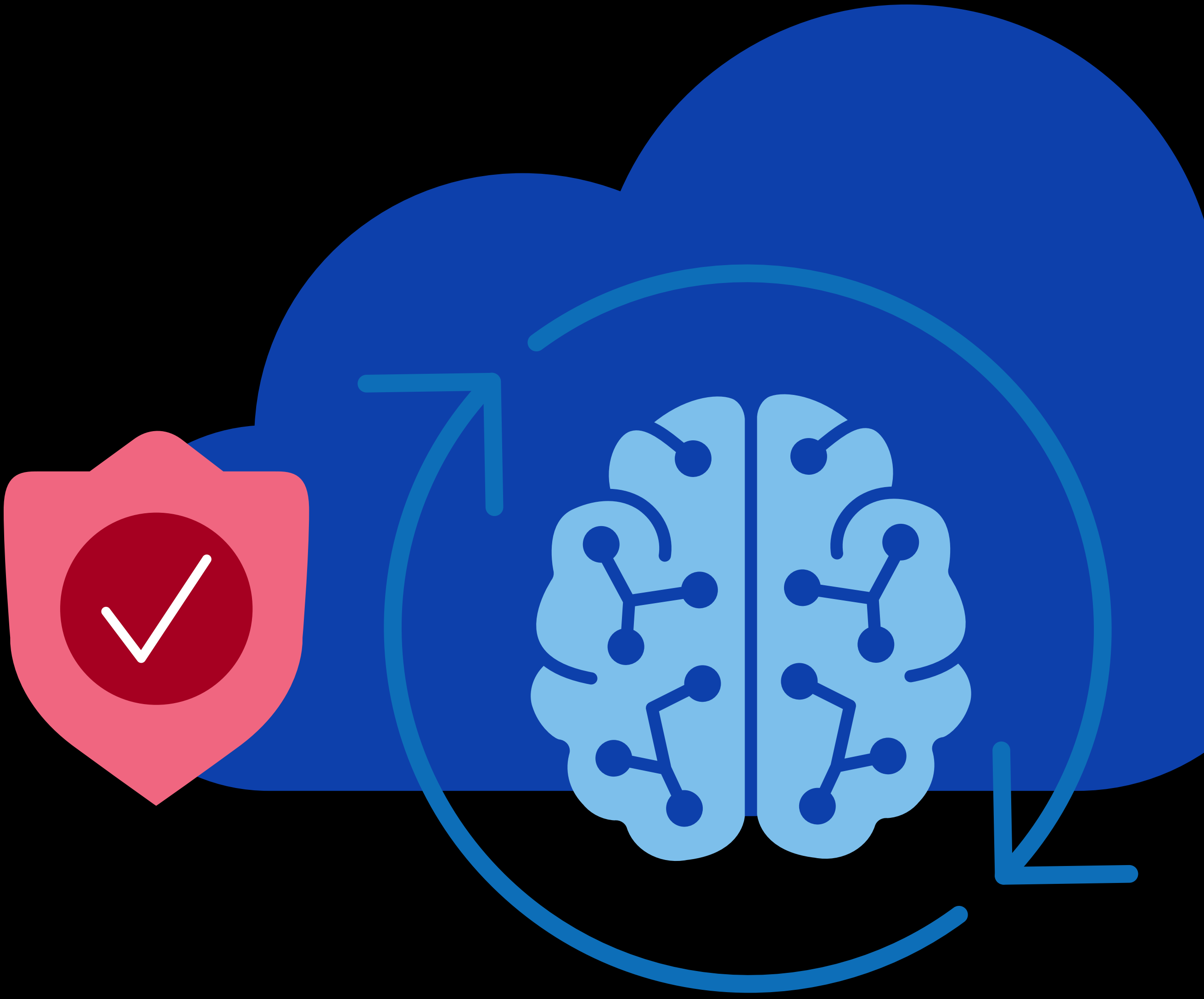
SOLUTION OVERVIEW

Protect and scale AI investments in the cloud

Secure AI model inputs and outputs against attacks, tampering, or leaks with F5 on Google Cloud.



Key benefits

Accelerate AI deployments
Connect and provision AI seamlessly across multiple environments, including Google Cloud, on premises, and edge.

Defend your reputation
Prevent sensitive data exposure, protect against AI-specific threats, and maintain high-quality outputs in real time.

Simplify operations and reduce overhead
Centralize management and deploy policies consistently across hybrid multicloud environments.

Optimize for better performance and cost
Gain access to AI-specific metrics that facilitate easier threat monitoring and efficient optimization efforts.

Security struggles to keep pace with AI adoption

Organizations across industries are racing to deploy large language models (LLMs), AI-powered chatbots, and machine learning systems to drive innovation and gain a competitive advantage. However, highly distributed infrastructure, multi-modal data flows, and generative capabilities make AI apps harder to protect from a host of new threats, including prompt injection, model jailbreaking, data poisoning, and denial-of-service (DoS) attacks.

Operational and infrastructure challenges

Beyond direct cyber threats, infrastructure complexity compounds security risks. AI services are resource-intensive, highly distributed across multiple environments, and have dynamic scaling needs due to massive data flows and unpredictable traffic spikes. These characteristics make it difficult to apply security policies consistently while maintaining app performance.

Visibility gaps and a lack of AI-specific metrics, such as token usage and model performance, also hamper security operations. Because AI systems produce varied outputs, security teams struggle to correlate malicious prompts to harmful outcomes and identify threats with precision.

A clear framework for AI security

Google's Secure AI Framework (SAIF) provides a comprehensive approach to addressing these security challenges for AI systems at scale:

1. **Expand strong security foundations** by adopting proven controls for AI and secure-by-default infrastructure.
2. **Extend detection and response** by monitoring and correlating AI inputs/outputs and integrating AI events with existing SecOps.
3. **Automate defenses** by using AI-powered security tools that can match the scale and speed of AI-powered attacks.
4. **Harmonize platform-level controls** to ensure consistent security across hybrid multicloud AI deployments.
5. **Adapt controls to adjust mitigation tactics** through real-time, continuous learning enhanced by threat intelligence and user feedback.
6. **Contextualize AI system risks** with end-to-end assessments that factor in business impacts and organizational objectives.

Key features

Securely connect multiple environments
Enable encryption at the network or application layers and seamlessly connect diverse data sources over a private backbone.

Ensure AI responsiveness
Automatically verify AI service functionality to prevent traffic routing to impaired AI instances.

Protect AI without degrading performance
Distinguish between legitimate users and automated attacks, safeguarding AI services from abuse and resource exhaustion.

Get unified observability
Surface metrics and logs directly to Google Cloud's observability suite to enable end-to-end AI request tracing.

Comprehensive AI security through a strategic partnership

F5 and Google Cloud operationalize SAIF principles through the F5® Application Delivery and Security Platform (ADSP). Together, Google Cloud services and the F5 ADSP optimize the performance and security of AI workloads across environments.

Foundational to delivering comprehensive AI security, the F5 ADSP leverages:

- F5® AI Guardrails: Define and observe how your AI models and agents interact with users and data, and defend against attackers.
- F5® NGINX Plus®: Advanced load balancing, AI-aware traffic management, and container-native security for Kubernetes environments.
- F5® Distributed Cloud Services: SaaS-delivered web app and API security, multicloud networking, and app delivery.

Through this strategic partnership, organizations can accelerate AI innovation while ensuring secure, seamless access to data across every environment.

Connect and deliver AI apps securely across clouds

Multicloud AI requires access to enterprise data in on-premises systems, private clouds, and edge locations. Distributed Cloud Services and Google Cloud Vertex AI make it simple to connect and provision separate data sources across environments. The F5® Global Network enables a private connectivity backbone that avoids exposing data to the public Internet.

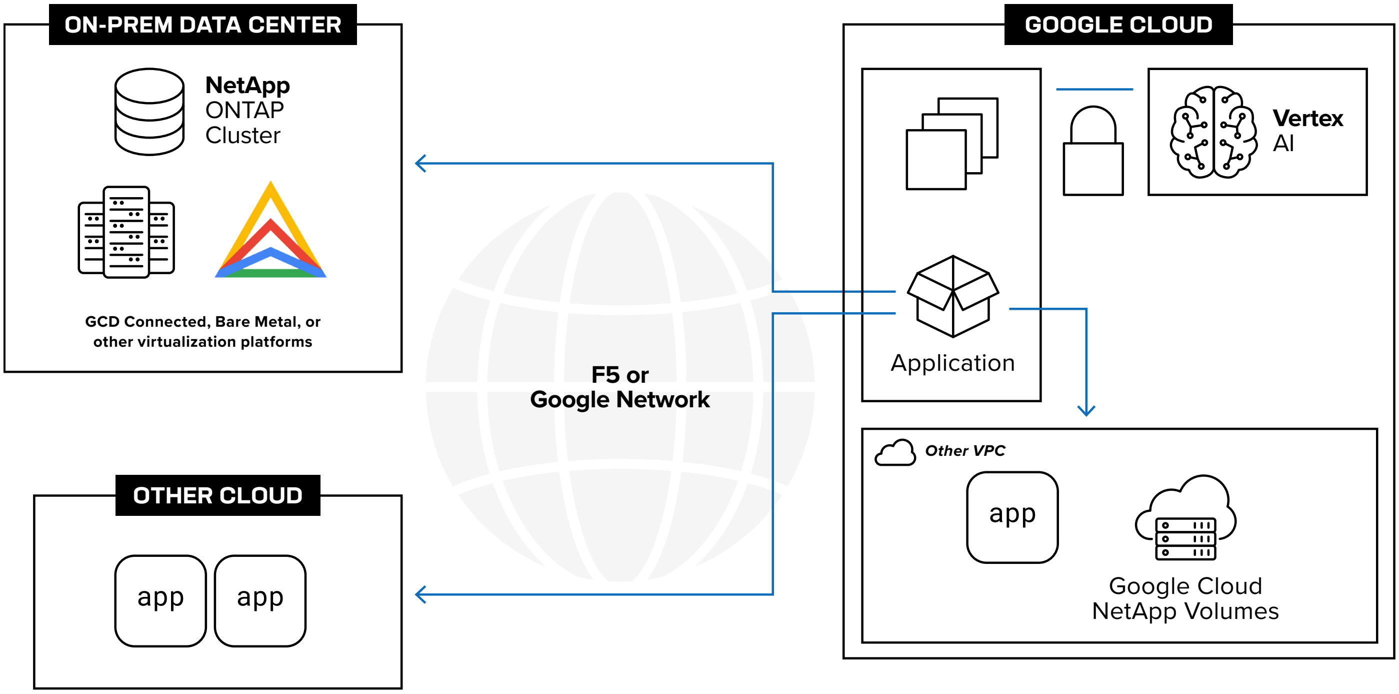


Figure 1: F5 simplifies connecting multiple environments across Google Cloud, private cloud, on premises, and at the edge.

Deploy layered protection for AI models and applications

Multiple layers of integrated protection work together to safeguard your AI systems. The Application Delivery and Security Platform filters malicious bots, prevents distributed denial-of-service (DDoS) attacks, and blocks attacks at the API ingress layer.

Google Cloud Model Armor for Vertex AI screens prompts and responses to prevent prompt injection and biased answers. AI Gateway complements Model Armor by inspecting inputs and outputs in real time and works with Google Cloud Sensitive Data Protection to redact personally identifiable information (PII) from prompts and responses.

Secure AI application delivery in Kubernetes environments

Traditional ingress controllers in Kubernetes environments often lack AI-aware capabilities, leading to suboptimal performance. NGINX Plus works with Google Kubernetes Engine (GKE) to add ingress and frictionless security capabilities compatible with AI workload needs.

These capabilities include F5® NGINX® Ingress Controller, which intelligently directs traffic to the healthiest AI model instances, and F5® NGINX® App Protect, which provides a robust web application firewall (WAF) and layer 7 DoS protection designed for containerized environments.

NGINX Plus also supports OpenTelemetry to provide comprehensive metrics, traces, and logs directly to Google Cloud's observability suite for efficient troubleshooting and optimization.

Protection for AI models, apps, and data with a joint solution

The F5 and Google Cloud partnership provides enterprise-grade AI security that addresses the unique challenges of protecting model inputs and outputs. Deploy AI applications at scale with comprehensive threat protection, operational simplicity with unified management, and AI-aware traffic management and performance optimization. Start protecting your AI investments today.

Contact F5 to get started or learn more at f5.com/gcp.