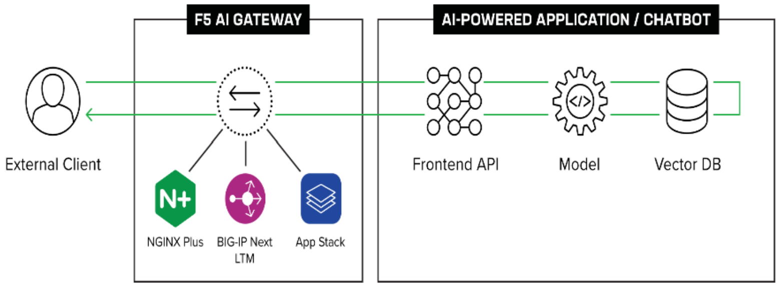
Protect and Optimize Your AI Applications with F5 and Trace3

AI represents the next wave in the evolution of applications but also presents serious challenges that could easily surpass the benefits. Given that incoming prompts to AI applications could be attacks on resources or designed to exfiltrate data, securing the AI applications becomes paramount. Flexibility in deployment and LLM/SLM support are just as important as time-to-market and managing costs.

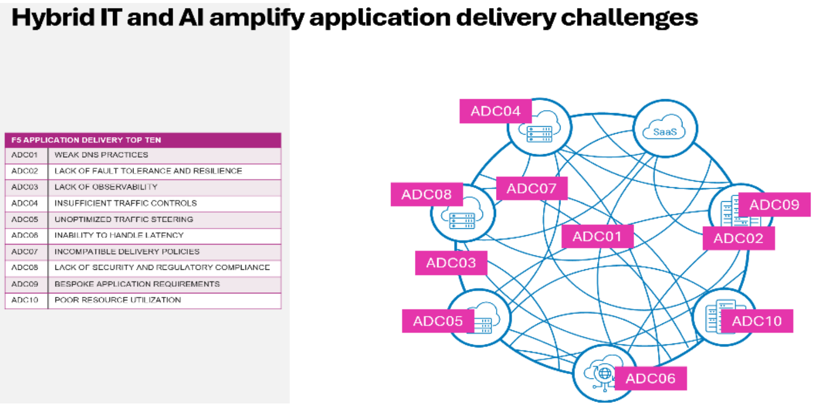
OVERVIEW

F5 AI Gateway enables organizations to confidently deploy AI applications anywhere. Easily ensure security, scalability, and reliability for your AI implementation. AI Gateway inspects inbound prompts and outbound responses to prevent unexpected outcomes or critical data leakage. Customizable observation, protection, and management of AI interactions helps improve the usability of AI applications and simplifies compliance.

AI applications require an evolved security solution to mitigate attacks via incoming prompts and to stop the leakage of sensitive data or hallucinations. Visibility over all transactions ensures that cost management, governance, and compliance are built into day-to-day operations. mapping across billions of dependencies.



AI Gateway inspects, identifies, and blocks inbound attacks such as prompt injections, insecure output handling, model denial-of-service, sensitive information disclosure, and model theft. For outbound responses, AI Gateway identifies and scrubs PII data and prevents hallucinations.



Hybrid IT and AI amplify application delivery challenges

F5 APPLICATION DELIVERY TOP TEN	
ADC01	WEAK DNS PRACTICES
ADC02	LACK OF FAULT TOLERANCE AND RESILIENCE
ADC03	LACK OF OBSERVABILITY
ADC04	INSUFFICIENT TRAFFIC CONTROLS
ADC05	UNOPTIMIZED TRAFFIC STEERING
ADC06	INABILITY TO HANDLE LATENCY
ADC07	INCOMPATIBLE DELIVERY POLICIES
ADC08	LACK OF SECURITY AND REGULATORY COMPLIANCE
ADC09	BESPOKE APPLICATION REQUIREMENTS
ADC10	POOR RESOURCE UTILIZATION

OFFERING PROCESS

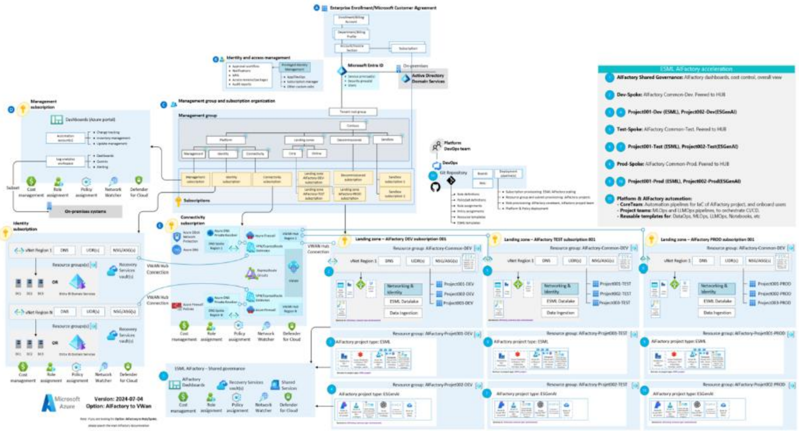
Assessment Phase

- Baseline Use Case, Application Architecture, Security, and Visibility of AI workloads
- Architect Azure Cloud Landing Zone and Operations for AI workloads
- Determine Application/Network security and architecture topology dependencies
- Operational Review and Recommendations

OFFERING PROCESS (Con't)

Identify Phase

- Pillars: application, security, infrastructure, network, cloud)
- Use Cases: Security, AI Application, Infrastructure, Network
- Opportunities to reduce tool sprawl, inconsistent tooling experience across hybrid/cloud environments
- Bring relevant Persona together
- Scope for Architecture improvements
- Cloud Cost optimization
- Self-Healing/ Auto Remediation of issues
- Automation of AI observability – deploy, configure, and maintenance
- Pipeline automation using observability data
- Automated Security/ Vulnerability Identification



Enterprise Enrollment/Microsoft Customer Agreement
Identity and access management
Management group and subscription organization
Management subscription
Identity subscription
Connectivity subscription
Subscriptions
Microsoft Azure
Version: 2024-07-04
Option: AIFactory to VWan
Landing zone – AIFactory DEV subscription 001
Landing zone – AIFactory TEST subscription 001
Landing zone – AIFactory PROD subscription 001
Resource group: AIFactory-Common-DEV
vNet Region 1
DNS
UDR(s)
NSG/ASG(s)
Networking & Identity
ESML Datalake
Data Ingestion
Project001-PROD
Project002-PROD
Project003-PROD
Project001-TEST
Project002-TEST
Project003-TEST
ESML AIFactory acceleration
AIFactory Shared Governance: AIFactory dashboards, cost control, overall view
Dev-Spoke: AIFactory Common-Dev. Peered to HUB
Project001-Dev (ESML), Project002-Dev(ESGenAI)
Test-Spoke: AIFactory Common-Test. Peered to HUB
Project001-Test (ESML), Project002-Test(ESGenAI)
Prod-Spoke: AIFactory Common-Prod. Peered to HUB
Project001-Prod (ESML), Project002-Prod(ESGenAI)
Platform & AIFactory automation:
- CoreTeam: Automation pipelines for IaC of AIFactory project, and onboard users
- Project teams: MLOps and LLMOps pipelines, to orchestrate CI/CD.
- Reusable templates for: DataOps, MLOps, LLMOps, Notebooks, etc

POC Phase

- High touch deployment, and set up of PoC targeted Landing Zones for AI use cases in Azure
- DevOps Strategy and Automation Operations Support
- Security and Compliance controls Implementation
- AI Operations and Infrastructure Deployment Training
- AI Gateway Logging and Monitoring Integration with Native Cloud tooling
- F5 Distributed Cloud console integration
- F5 AI Gateway Training and Operations Support
- Cloud Cost optimization

VALUE

Trace3 and F5 have joined forces to help enterprises confidently deploy, secure, and scale AI workloads in Microsoft Azure. Together, we deliver a comprehensive solution that aligns cloud infrastructure, enterprise security, and AI governance under a unified strategy.

At the core of this partnership is Trace3's AI Landing Zone Program, built on Microsoft's Enterprise-Scale Landing Zone (ESLZ) best practices. This framework accelerates cloud adoption by providing a pre-architected, compliance-ready foundation tailored for AI and machine learning workloads. It includes policy-based governance, automated deployment pipelines, and integration with Azure-native services to simplify operations and improve time-to-value.

F5 enhances this foundation with its AI Gateway, delivering critical security, observability, and traffic control capabilities for AI services operating in hybrid or multi-cloud environments. As enterprises expose LLM-based services and AI APIs to internal and external consumers, the F5 AI Gateway helps ensure consistent enforcement of data privacy, API security, and request throttling policies, while providing deep traffic insights into model behavior and usage

This strategic partnership between Trace3 and F5 equips organizations to move beyond proof-of-concepts and into **enterprise-scale AI deployments** with the confidence that operations, compliance, and security are built-in—not bolted on.