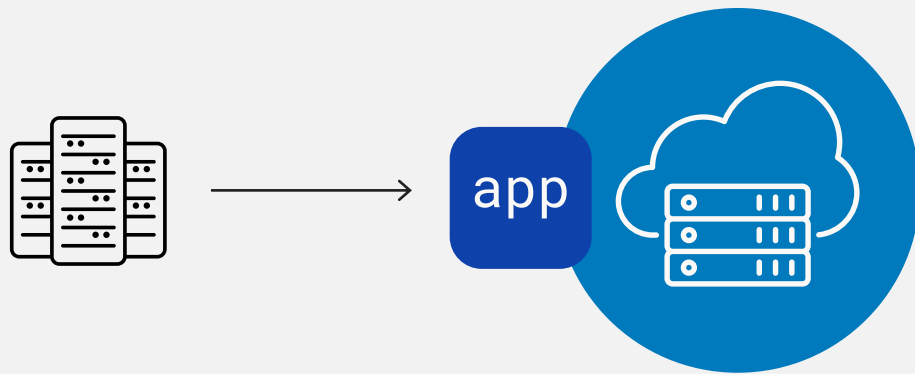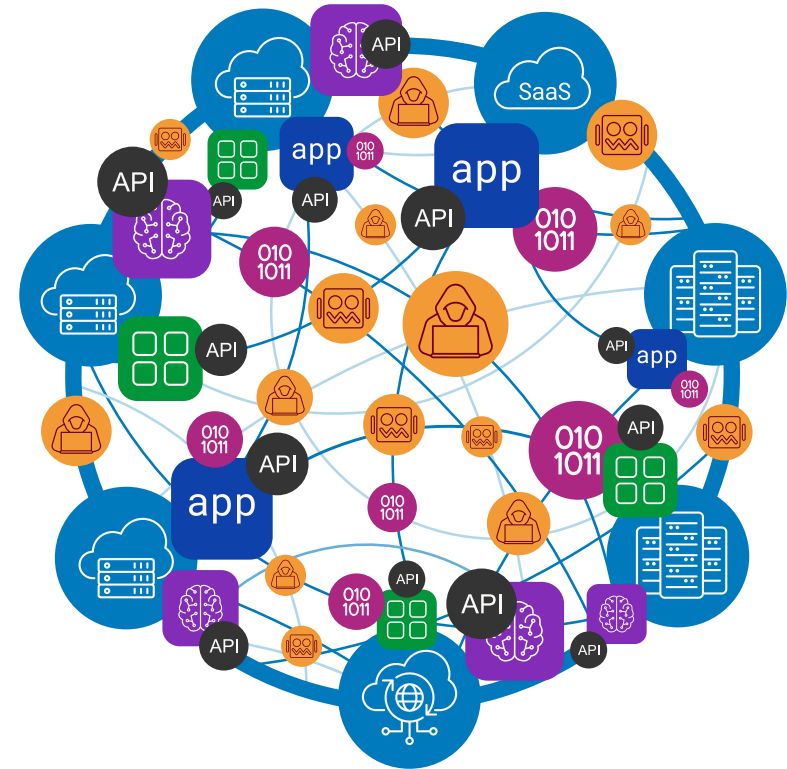# AI in Cybersecurity : From Smarter Load Balancing to Intelligent Defense

Jeremiah Agboola | F5 Solutions Engineer
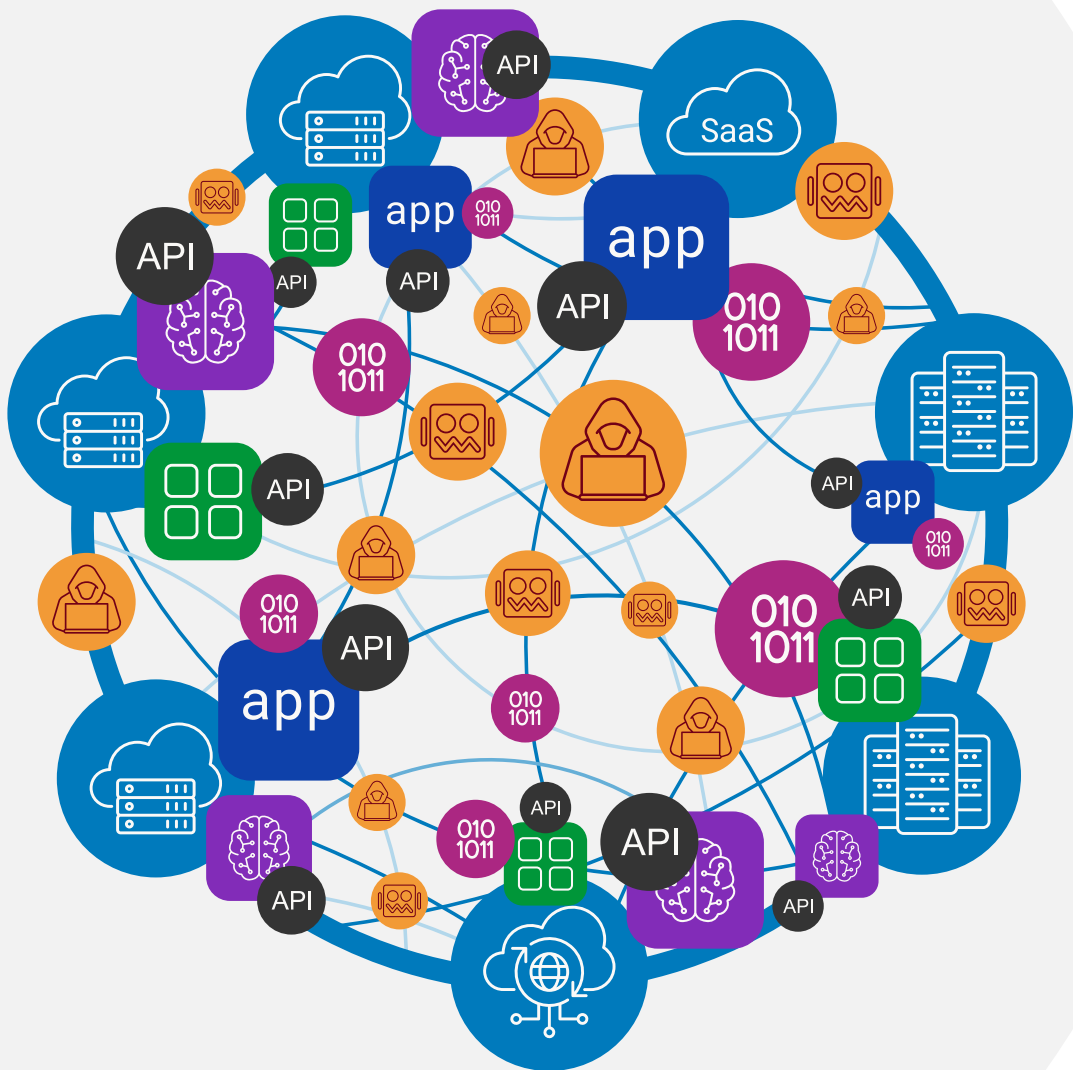
What many CIOs thought would happen

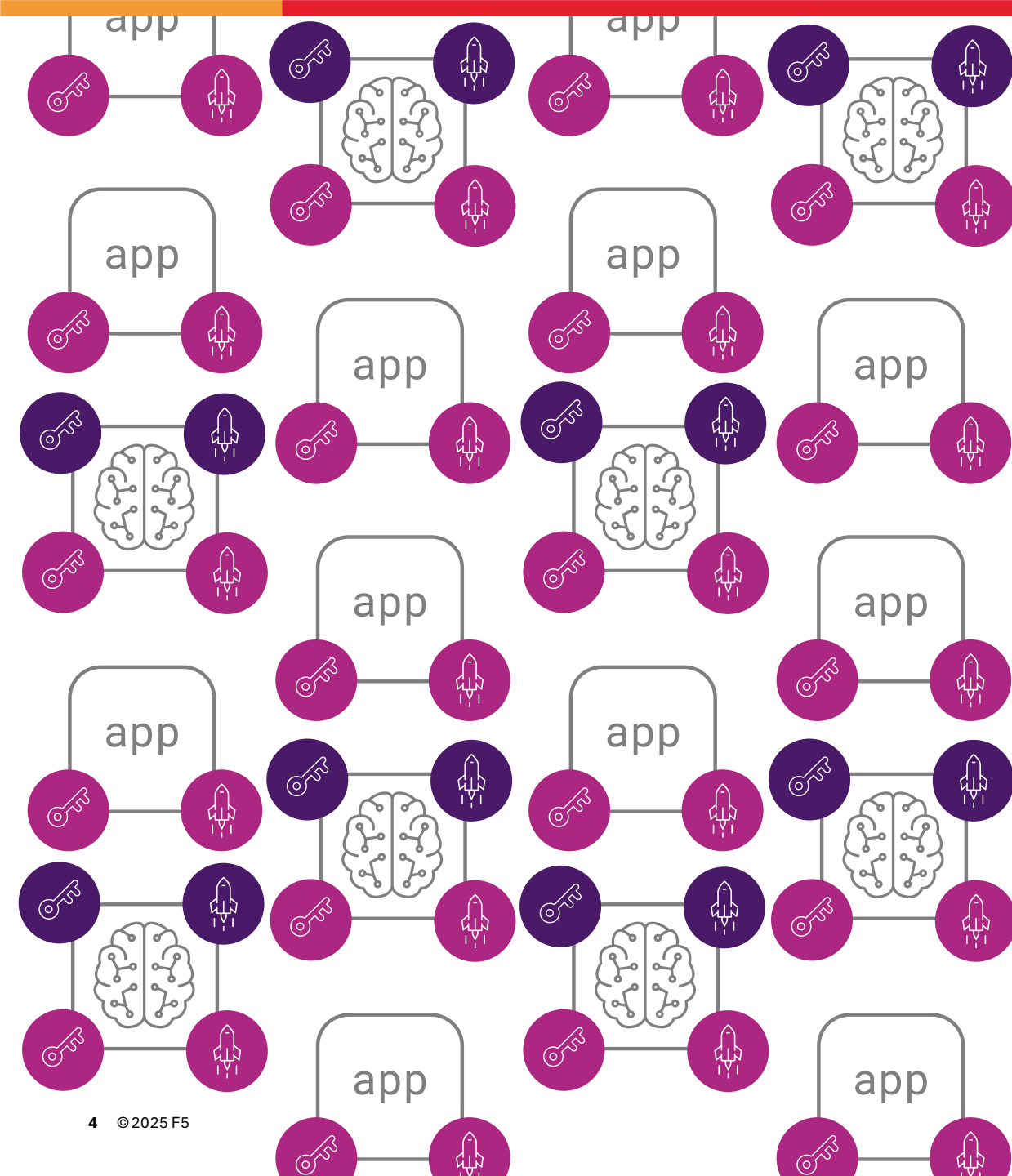# The hybrid, multicloud world CIOs face today

Ball of Fire

ADC 1.0 — Data Centers

ADC 2.0 — Public Cloud

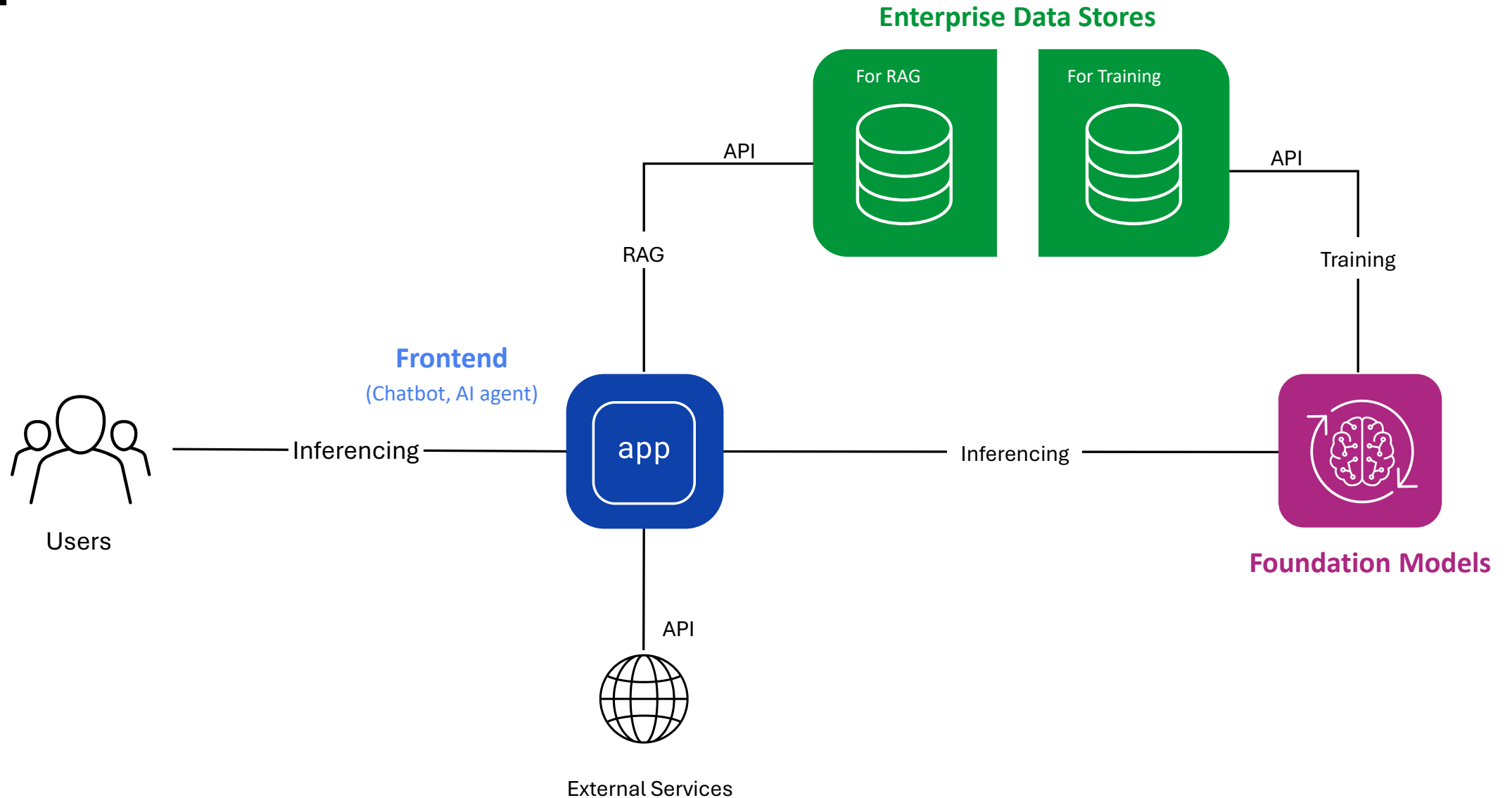ADC 3.0 — Hybrid Multicloud Distributed AI Apps AI Security

By 2028, 80% of all enterprise apps will use AI

This transition is rewriting the stack – delivery and security included

- Traditional requirements
- AI-specific requirements

Source: F5 Corporate Strategy, 2025.

Simple AI Reference Architecture

Enterprise Data Stores

For RAG

For Training

API

API

RAG

Training

Frontend
(Chatbot, AI agent)

app

Inferencing

Inferencing

Users

API

External Services

Foundation Models

AI applications = Modern applications

Built as **modern applications**...

Containerized workloads in Kubernetes

Built with APIs

High degree of automation

Product mindset with fast iterations

...with **unique challenges**

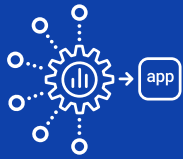Data loss prevention and PII compliance

Non-deterministic outputs

Cost control for LLMs

New threat vectors and attack surfaces

What does AI actually look like in today's enterprise?

Challenges enterprises face when deploying AI

Explosion of data pipelines

Growing demands for real-time AI workloads

Increased data security, compliance, and governance

Real AI use cases we are beginning to rely on today

- Autonomous vehicles
- Real-time medical imaging detection
- Predictive cybersecurity threat detection
- AI-driven customer service interactions
- Personalized media content tailored to users

Enterprises are now architecting to support the transition from independent generative AI apps to agentic AI workflows

	Generative AI		Agentic AI
Definition	Generates content and recommendations to augment human processes	→	Autonomously initiates and completes tasks
Architecture	AI deployed as standalone apps, activated on-demand, isolated within specific solutions	→	Deeply integrated across apps and protocols, uses event-driven autonomic processes
Deployment	Primarily static API calls, manual initiation through protocols	→	Dynamic, self-initiating agents utilize API-driven automation and real-time contextual processing
Integration	Limited integration, point solutions and isolated data flows	→	Broad integration, multi-system orchestration, proactive interoperability across platforms
Examples	Drafting a marketing email or summarize a meeting transcript	→	Automatically identify leads, builds personalized marketing cadence, and monitors responses

As enterprises scale AI in production, delivery and data ingestion challenges throttle performance and reduce ROI

Latency bottlenecks

Accelerating cloud egress costs

Inability to scale

Multi-tenant security & isolation
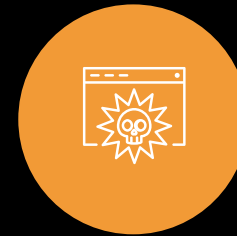
Operational complexity

As enterprises deploy AI, new AI-specific threats expand the attack surface

Hallucinations

Data exfiltration

Abuse of large language models

Agentic malware

Prompt injection

The growing challenges of scaling AI workloads and data management

Performance bottlenecks

High costs of poor efficiency

Limited scalability

Security and compliance risks

Inconsistent availability

Lack of acceleration

Data accessibility

Lack of intelligent Traffic Management

Multicloud and hybrid complexity

F5 Application Delivery and Security Platform (ADSP)
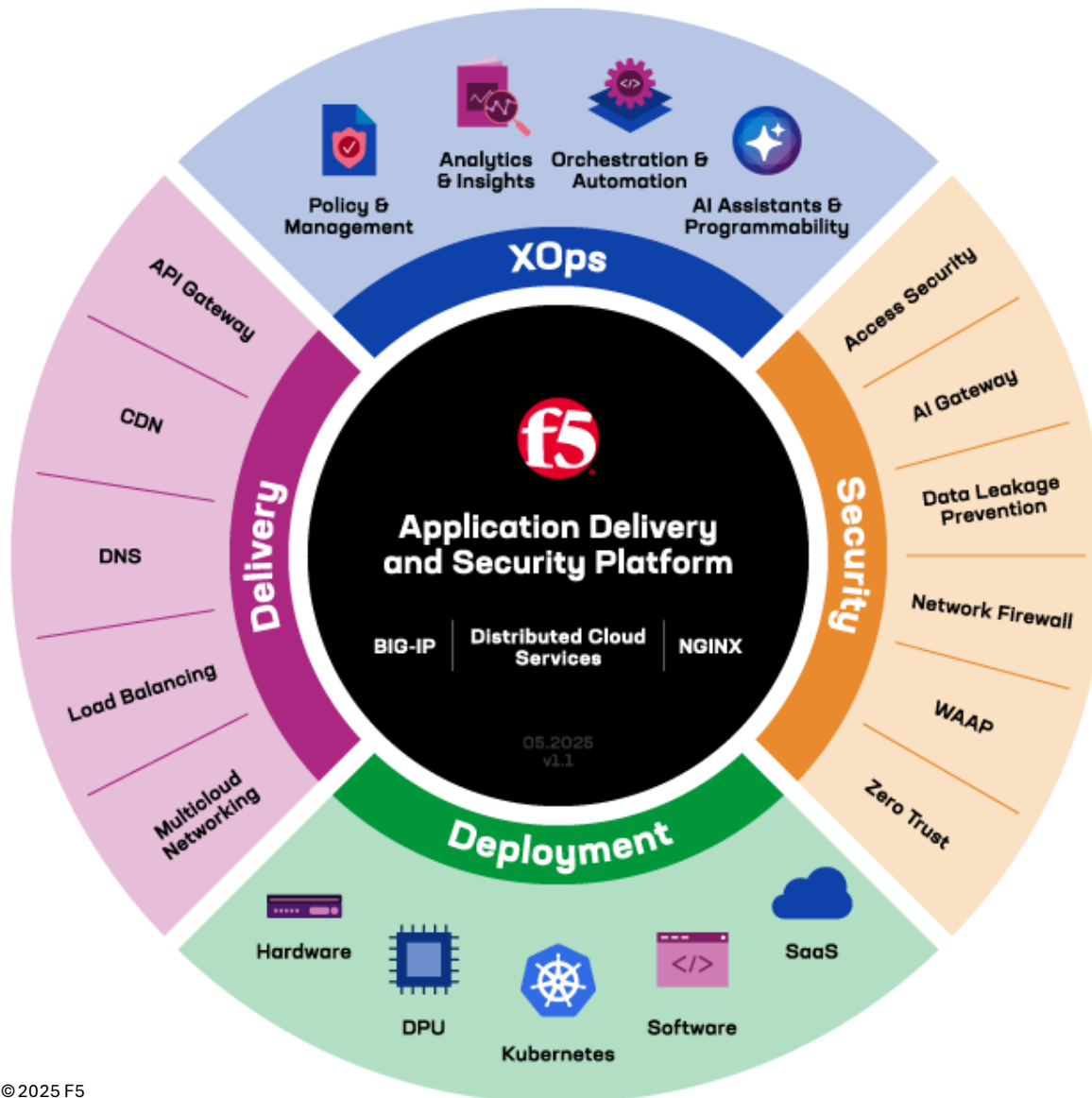
1 Complete delivery for every app

2 Complete security for every app

3 Deployable anywhere and in any form factor

4 X Operations across NetOps, SecOps and DevOps

F5 enterprise AI delivery and security solutions offers a unified, scalable solution to protect and optimize AI workloads

F5 leverages decades of application delivery expertise with enterprise-grade AI security to optimize performance, scalability, and protection for AI workloads.

AI Data Delivery
Ensure fast, secure, and reliable data pipelines to fuel AI model training at scale.

AI Factory Load Balancing
Minimize costs and maximize GPU utilization while streamlining AI workloads for efficient infrastructure.

AI Runtime Security
Safeguard AI applications, APIs, and models against emerging threats like data leakage, API exploitation, shadow AI, and malicious injections.

AI Multicloud Networking
Enable seamless and secure data and model connectivity across distributed AI environments.

F5 Generative and Agentic AI Reference Architecture

RETRIEVAL AUGMENTED GENERATION

ENTERPRISE EVENT & DATA STORES

FRONT-END APPLICATIONS

End Users

AI FACTORY

AI AGENTS

MODEL INFERENCE

MODEL TRAINING

DOWNSTREAM SERVICES

AI Agents   Databases   Websites   Queues

© 2025 F5
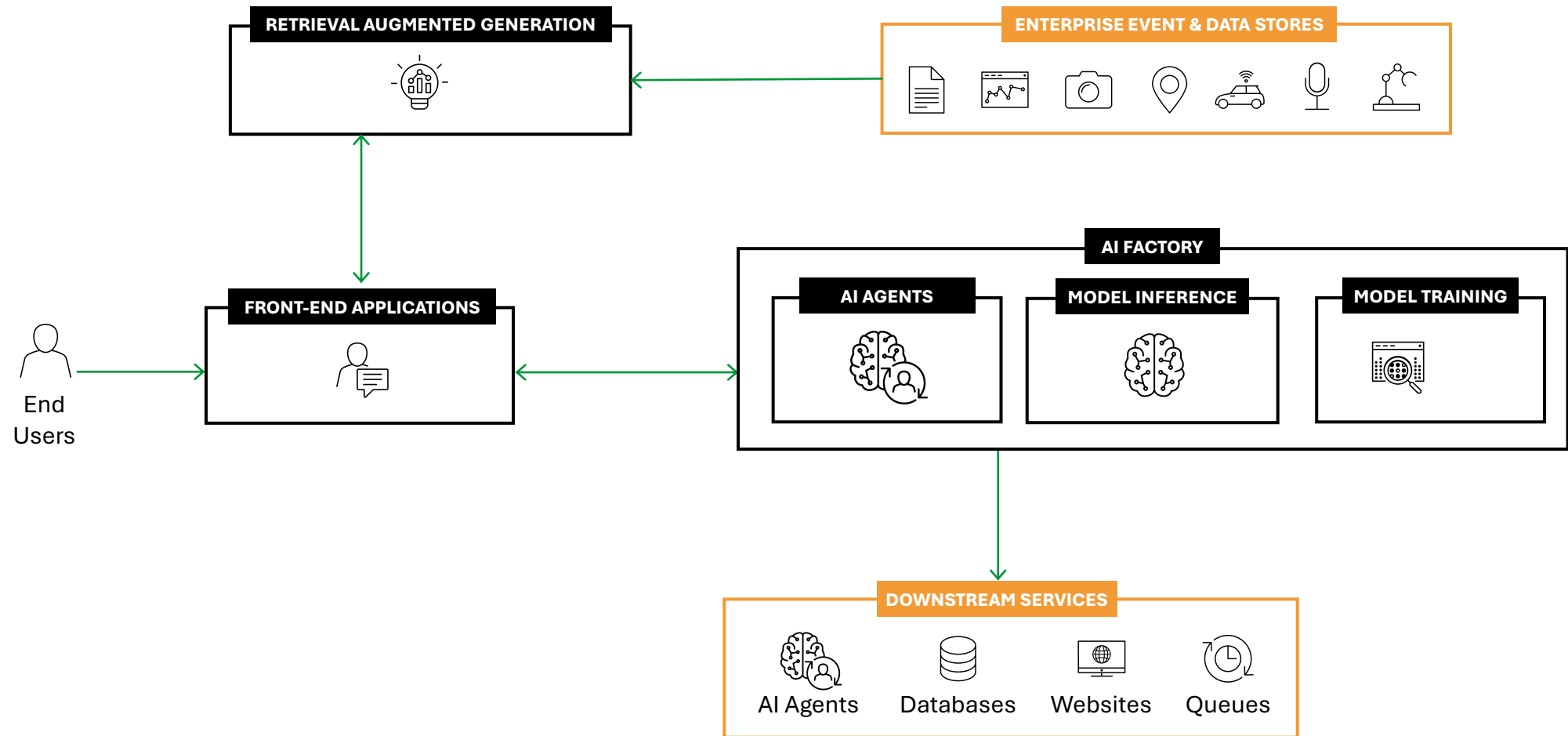
F5 is the cornerstone of scalable, resilient, and secure AI architectures

RETRIEVAL AUGMENTED GENERATION

Secure **AI Object Storage**

ENTERPRISE EVENT & DATA STORES

Secure and load balance the **Web and API** "front door"

Secure and load balance **AI APIs**

Secure and load balance **AI Data Delivery**

AI FACTORY

FRONT-END APPLICATIONS

AI AGENTS

MODEL INFERENCE

MODEL TRAINING

End Users

Secure **APIs and AI guardrails**

Secure and scale **Cluster Ingress**

Optimized **Resource Utilization**

DOWNSTREAM SERVICES

AI Agents   Databases   Websites   Queues

Data path

External Data

Safeguard AI applications, APIs, and models while keeping them connected across highly distributed environments

RETRIEVAL AUGMENTED GENERATION

ENTERPRISE EVENT & DATA STORES

Secure and load balance the **Web and API** "front door"

Secure and load balance **AI APIs**

FRONT-END APPLICATIONS

End Users

AI FACTORY

AI AGENTS

MODEL INFERENCE

MODEL TRAINING

LLM Observability, Model Security and Model Routing

Secure external **API Access**

DOWNSTREAM SERVICES

AI Agents   Databases   Websites   Queues

F5 delivers data end-to-end: secure and scalable data delivery

Efficiency

Smart load balancing & security
Directs traffic to the optimal storage for performance, security, and availability

Seamless data flow
BIG-IP LTM ensures efficient, secure routing from external sources to local storage

Optimized S3 routing
BIG-IP DNS enables global resiliency for highly critical S3 storage environments

Scale

High-throughput traffic management
Optimize TCP and HTTPS flows for seamless object storage access

Accelerated packet processing
Leverage best in class FPGA to accelerate packet processing for traffic management

Crypto offload for speed
BIG-IP LTM offloads encryption to best-in-class hardware on rSeries and VELOS, boosting performance

Security

Robust DDoS protection
BIG-IP AFM defends against volumetric and targeted attacks

Secure traffic management
BIG-IP LTM ensures efficient, secure routing from external sources to local storage

End-to-end data protection
Safeguards AI workloads with policy-driven security and threat mitigation

AI architectures are becoming geographically distributed, driving the need for secure, optimized, multicloud networking

AI Models

As AI becomes the new control plane for operations, F5 AI Assistant empowers your teams to manage app and API delivery and security
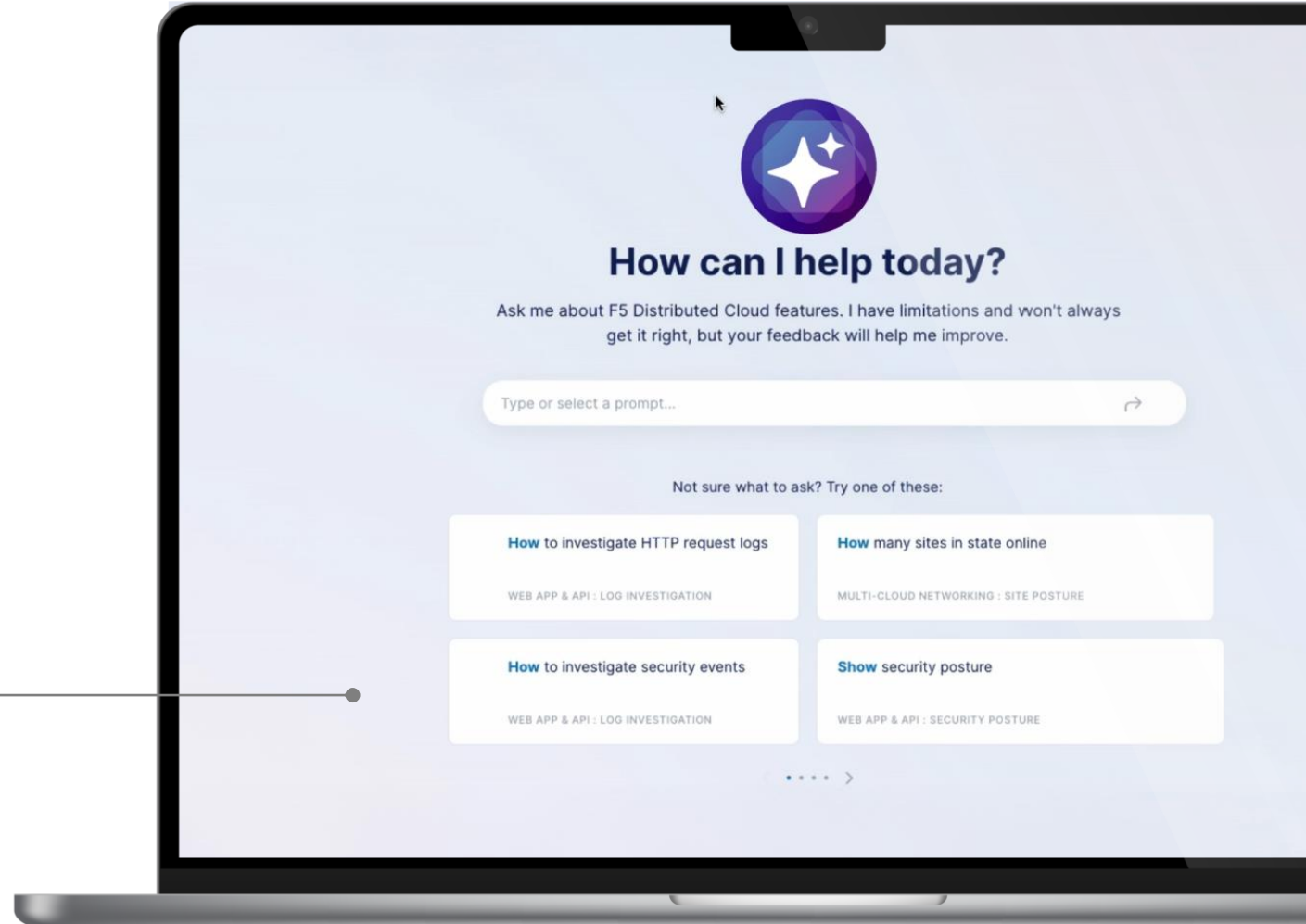
Intelligent summarization

Actionable recommendations

Customizable programmability (e.g., iRules Code Generation)

Predictive insights

F5 AI Assistant is available across the F5 Application Delivery and Security Platform

With our partners, F5 simplifies and secures AI application ecosystems

© 2025 F5
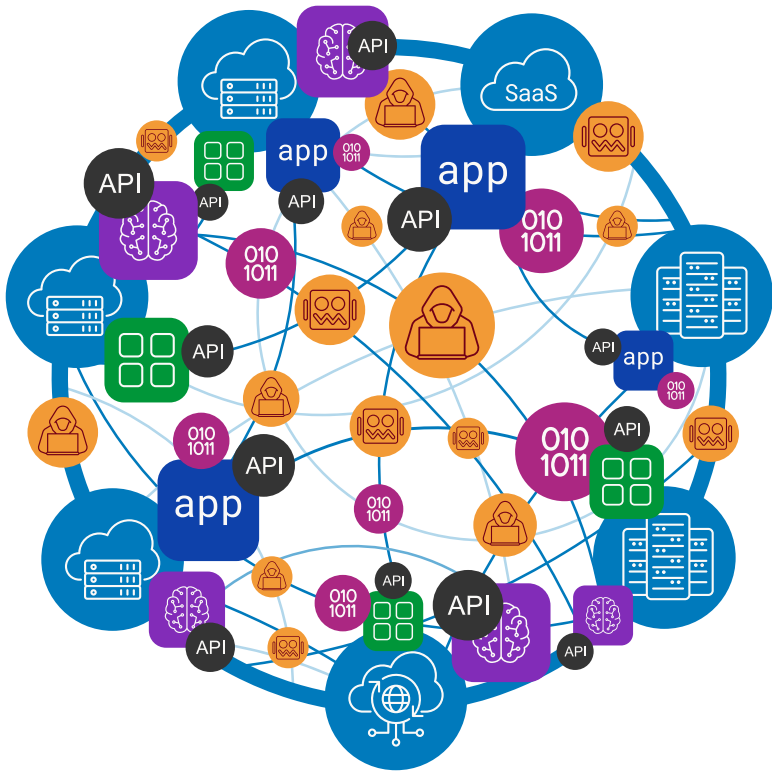
The real issue isn't AGI: it's taming your AI deployments before it's past the point of no return

1 Explosive growth of APIs creating unmanaged connectivity risks

2 Distributed AI workloads increasing network strain and complexity

3 Autonomous, agentic workflows complicating infrastructure management

4 Hybrid and multicloud deployments causing fragmented security policies

5 Real-time AI inference workloads making latency and availability critical

6 Rapid data pipeline expansion escalating compliance and governance risks

The most extensive application delivery and security platform in the industry, designed for the AI era
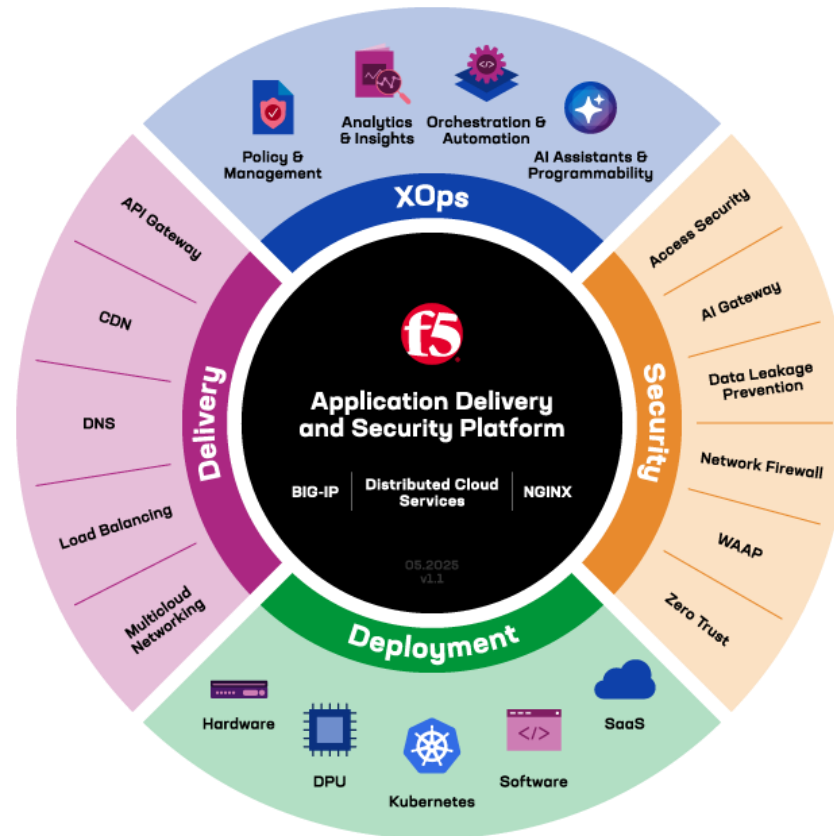
Accelerate your AI deployments at

f5.com/ai