



# We are capturing new AI opportunities

**Kunal Anand**  
Chief Product Officer



# AI creates new demand for app delivery and security



## AI changes delivery



More inference traffic



More latency pressure



More utilization pressure



New AI factory insertion points



**F5 is positioned  
where AI traffic, policy,  
performance, and  
protection converge.**



## AI changes security



More autonomous workflows



More AI-specific attacks



Faster vulnerability discovery



More need for runtime  
and perimeter enforcement

# Production AI is not a model – it is an ecosystem



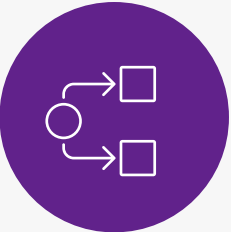
## Users and agents

More traffic and autonomous interaction



## Apps and APIs

More policy decisions



## Models and inference

More security and governance



## Data and services

More infrastructure pressure

**Production AI needs delivery, security, policy, and control to work.**

# AI is creating three distinct opportunities for F5



**AI data delivery**



**AI runtime security**



**AI factory load  
balancing**



## **AI data delivery**

# F5 is the only programmable, high-performance control point sitting in front of S3-compatible storage



**Programmable  
traffic management**

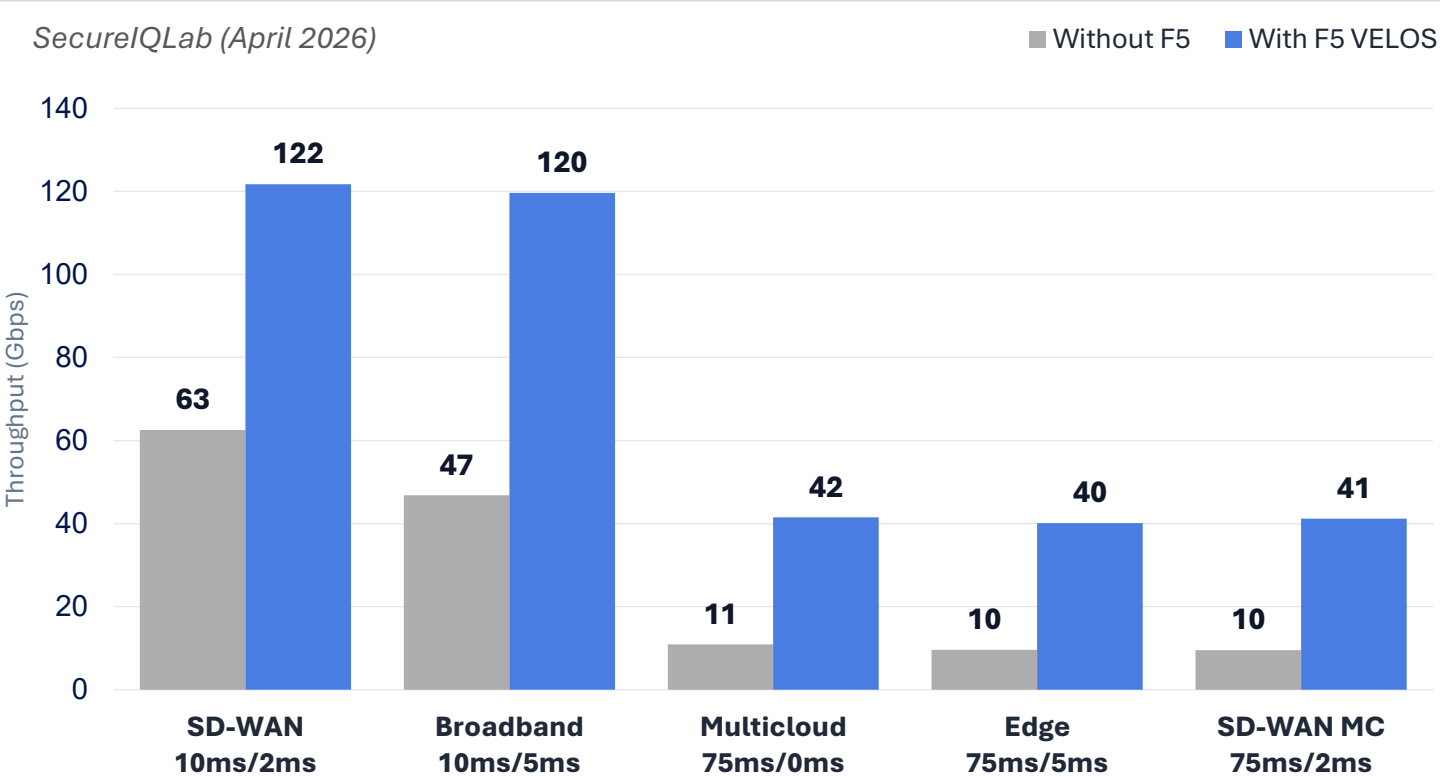


**Hardware scale that  
storage controllers  
cannot match**



**Vendor-neutral across  
the S3 ecosystem**

# F5 is delivering more than 3X higher S3 throughput, validated by our customers and independent analysis



### Performance lift

Low-latency SD-WAN	<b>+95%</b>
Low-latency broadband/VPN	<b>+155%</b>
High-latency multicloud	<b>+281%</b>
High-latency edge	<b>+316%</b>
High-latency SD-WAN multicloud	<b>+332%</b>

F5 is partnered with:





## **AI runtime security**

# AI changes the security operating model



**AI finds  
vulnerabilities  
faster**

+



**and enables  
attackers to  
strike faster**

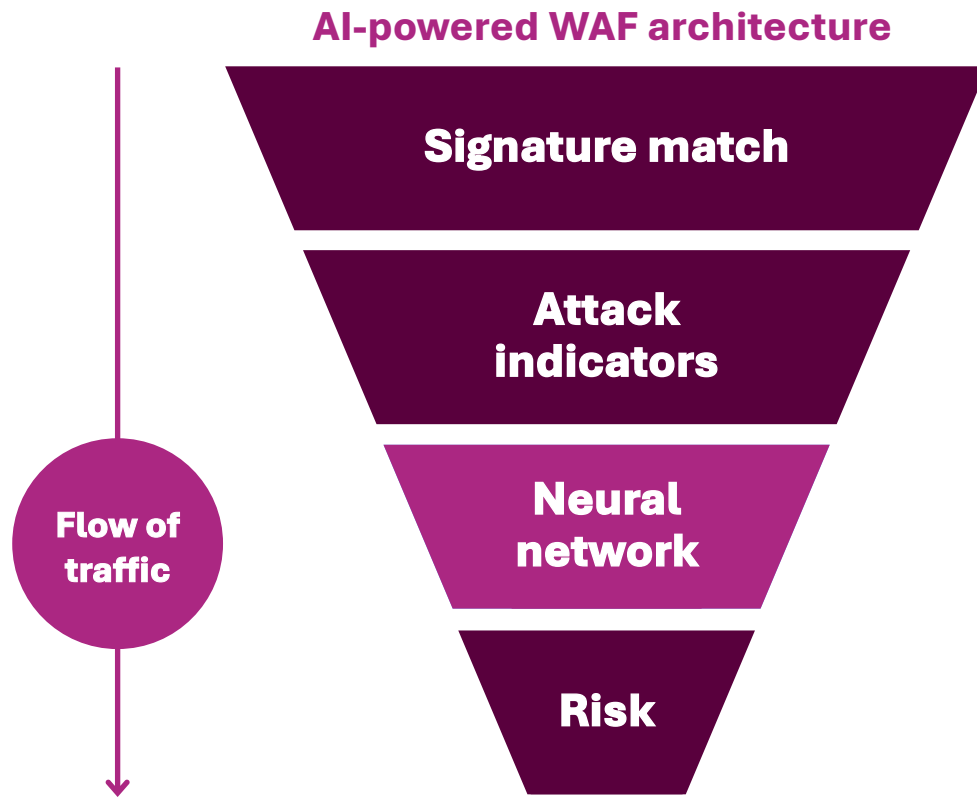
=



**making runtime  
protection  
non-negotiable**

# AI-powered security is the answer

Introducing F5's AI-powered WAF:  
200+ production customers within 70 days of GA



**>80%**

Customers in a blocking mode

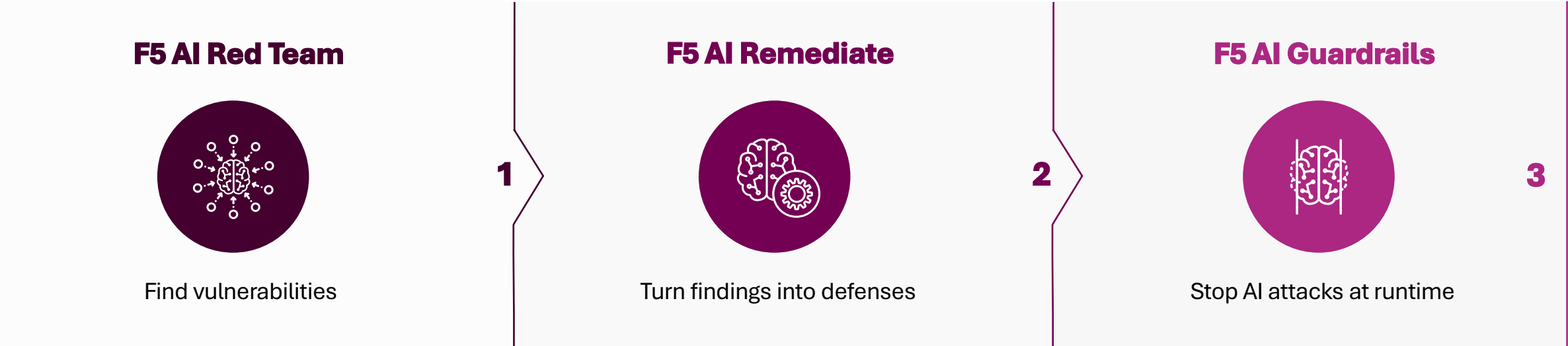
**98%**

Detection accuracy

**1%**

False positive rate

# Find, fix, protect: F5 secures the full AI lifecycle



## F5 AI Security received strong independent validation by SecureQLab (March 2026)

**98.4%**  
Overall AI Guardrails security score

**99.3%**  
Effective against direct prompt injection

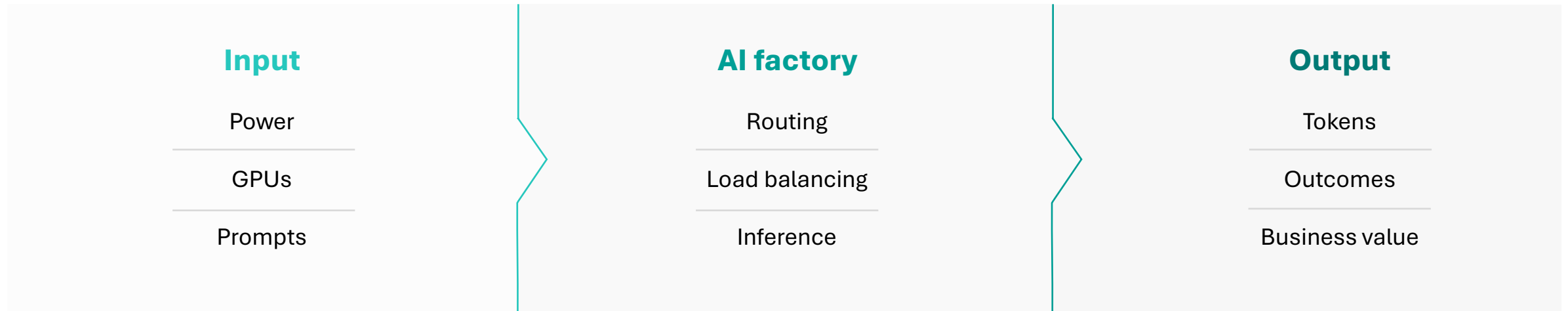
**99.0%**  
Effective against sensitive data leakage





## **AI factory load balancing**

# AI factories convert electricity into useful tokens, a conversion process measurable through the lens of tokenomics



## Tokenomics



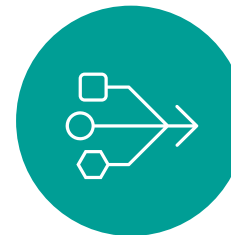
**Total tokens generated**



**Time to first token**



**Cost per token**



**End-to-end latency**



**Tokens per watt**

# F5 enables improved performance and economics for customers building AI infrastructure

**40%+**

**Token generation**

Increase in total throughput

**61%+**

**Faster TTFT**

Reduction in time to first token

**34%+**

**Lower end-to-end latency**

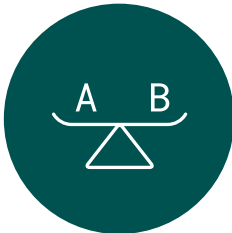
Improved full inference response time

*Tolly Group (March 2026)*

## Features and capabilities



**Security inspection**



**Load balancing**



**LLM routing**



**Token governance**

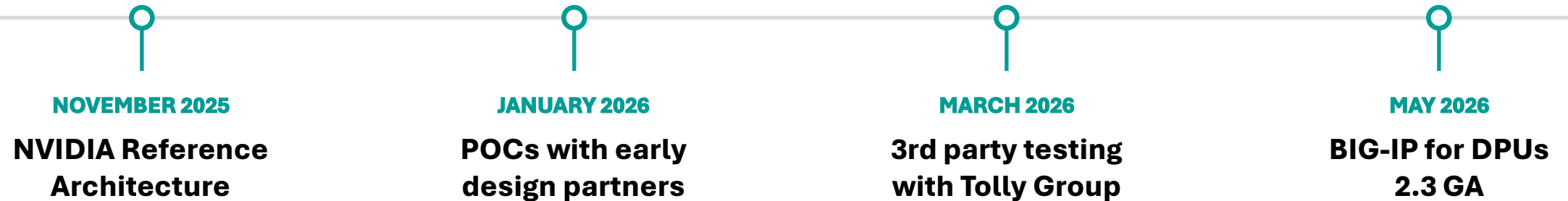


**Encryption**



**Observability**

# F5 BIG-IP for DPUs is positioned for the inference buildout and has the potential to be a growth catalyst



## Early findings and observations



DPU adoption lags GPU buildouts by a phase



Inference economics are decided in the data path



The neocloud stack is heterogeneous and still forming

# Key takeaways

# AI is contributing both directly and indirectly to our opportunity

## F5 DIRECT AI USE CASES

### AI data delivery



#### Target

Existing large enterprise, government, and telco customers



#### Solutions

BIG-IP in front of data stores



#### Monetization

Majority hardware, mainly on-premises data centers



#### Market timing

Strong current adoption and growth

### AI runtime security



#### Target

Existing large enterprise, government, and telco customers



#### Solutions

AI-powered WAF, AI Red Team, AI Guardrails, AI Remediate



#### Monetization

Software subscription across private, edge and cloud



#### Market timing

Momentum building

### AI factory load balancing



#### Target

Sovereign AI factories and neoclouds



#### Solutions

BIG-IP hardware (across AI factories), BIG-IP Next for Kubernetes software on DPUs (within AI factories)



#### Monetization

Initial hardware deployments. Larger potential for software subscription on Nvidia BlueField DPUs



#### Market timing

Nascent given customer readiness

## INDIRECT AI DEMAND

### AI-driven workload growth



#### Target

Existing large enterprise, government, and telco customers



#### Solutions

BIG-IP in front of AI-enabled apps



#### Monetization

Hardware and software on premises, in private cloud



#### Market timing

Strong current adoption and growth

# Our AI momentum is building

1

AI DATA DELIVERY



## Healthcare Services

F5 delivers the critical data delivery infrastructure powering their **AI-driven interactive voice response platform**

**\$2M+ deal**

2

AI RUNTIME SECURITY



## Global Financial Services

F5 secures enterprise-wide GenAI adoption with **AI Guardrails** that deliver high-fidelity threat detection

**~\$4M deal**

3

AI FACTORY LOAD BALANCING



## Energy & Chemicals

F5 **accelerates and stabilizes AI inference** offloading L4 and SSL to cut latency and prevent timeouts.

**\$1M+ deal**

# AI creates new demand for app delivery and security



## App delivery



BIG-IP for AI data delivery



BIG-IP for AI factory load balancing



BIG-IP Next for Kubernetes for DPUs

**Delivering and  
securing traffic  
for any app, API,  
and agent**



## App security



AI-powered WAF



AI Red Team



AI Guardrails



AI Remediate

