SOLUTION OVERVIEW

Securing models on Red Hat OpenShift AI with F5 Distributed Cloud Services

As AI model serving deployments become increasingly complex, protecting sensitive data and ensuring secure API interactions are critical for data security and integrity. F5® Distributed Cloud Services, part of the F5 Application Delivery and Security Platform, works with Red Hat OpenShift AI to provide robust API security and protection for serving AI models, addressing critical challenges including unauthorized access, data exfiltration, and prompt manipulation.

Key benefits

Enhanced security for AI model serving
Protects API interactions with advanced capabilities like API discovery, schema validation, bot mitigation, sensitive data redaction, and real-time threat detection. Ensures compliance and safeguards sensitive data across multicloud environments.

Streamlined AI operations
Offers centralized, real-time dashboards for monitoring, analytics, and policy adjustments, enabling teams to identify and resolve issues faster and ensure seamless model serving.

Optimized traffic and resource management
Improves performance with intelligent traffic management, balancing workloads across hybrid environments to minimize latency, maximize resource efficiency, and maintain service reliability.

Accelerated AI model deployment
Speeds up time-to-market with a unified platform that integrates data science workflows, application development, and IT operations, fostering collaborative innovation without compromising governance or security.

Security challenges when serving AI models across multiple environments

AI model serving has expanded into multiple environments, from public clouds to private data centers. This distributed approach can strain resource management practices, introduce versioning difficulties, and complicate compliance efforts. As data and models move between different infrastructures, understanding where and how they are deployed becomes a significant task.

Security concerns increase when models draw on sensitive data or interact with external protocols. Unauthorized access, malicious prompt manipulation, and data leaks pose serious risks to model integrity and reliability. As regulations around data protection evolve, organizations face added pressure to maintain strict control over every endpoint and data pathway.

Maintaining stable performance and consistent oversight is just as challenging. Models often require high levels of processing power and complex communication channels, making real-time monitoring essential. Without clear visibility into these workloads, organizations may encounter unpredictable behavior, delayed responses, or vulnerabilities that lead to breaches or loss of service.

Securing AI model serving APIs on Red Hat OpenShift AI with F5

When deployed together, F5 Distributed Cloud Services and Red Hat OpenShift AI address the security and performance challenges associated with AI model serving. F5's suite offers protection features such as API discovery, schema validation, and threat detection, which reduce the risk of data breaches and mitigate vulnerabilities in real time. Meanwhile, Red Hat OpenShift AI provides a consistent container environment that simplifies the packaging and deployment of those models at scale.

This coordination minimizes overhead by centralizing policy control and traffic management across multiple environments. F5 Distributed Cloud Services integrates with Red Hat OpenShift AI to offer dashboards that show API traffic, usage statistics, and emerging threats. These insights support decisions about resource allocation or which requests need deeper scrutiny, helping teams limit unauthorized activity and maintain system health.

Because the solution focuses on automation and streamlined workflows, teams can adopt new models quickly without sacrificing security. F5 and Red Hat provide a platform that adapts to changing requirements, letting organizations respond to usage spikes, new compliance mandates, or updates to the models themselves. With this approach, security standards do not impede innovation; instead, they enhance the reliability and predictability of AI initiatives.

Key benefits

Reduced operational complexity
Simplifies lifecycle management with automated policy enforcement, configuration updates, and performance optimization, reducing manual intervention and freeing teams to focus on strategic AI projects.

Future-ready multicloud support
Secures and scales AI models across hybrid and multicloud environments with consistent infrastructure, governance, and observability—enabling flexibility while managing compliance.

Strengthening AI model serving security

By integrating F5 Distributed Cloud Services with Red Hat OpenShift AI, organizations create a solid foundation for handling the demands of AI model serving across Hybrid multicloud environments. F5 security capabilities, including threat detection and traffic rule enforcement, work in conjunction with OpenShift AI's container orchestration to guard against risks such as unauthorized access and data misuse. This alignment gives teams better oversight of how models and data are used, improving reliability and reducing the response time to potential incidents.

Scaling operations becomes more effective when the platform balances workloads and routes traffic to optimal resources. This practice supports continuous deployments by allowing quick launches of new models and features without disrupting ongoing activities. Rigorous checks on compliance and security also occur automatically whenever changes take place, preserving consistency in how policies are applied across all environments.

Bringing these strengths together creates a framework where security and operational excellence reinforce each other. F5 and Red Hat provide organizations a predictable roadmap for expanding AI capabilities, whether that means exploring new use cases or adjusting existing workflows. As AI continues to influence strategic decisions, this combined approach helps teams remain prepared for emerging challenges and maintain a strong position in a competitive landscape.

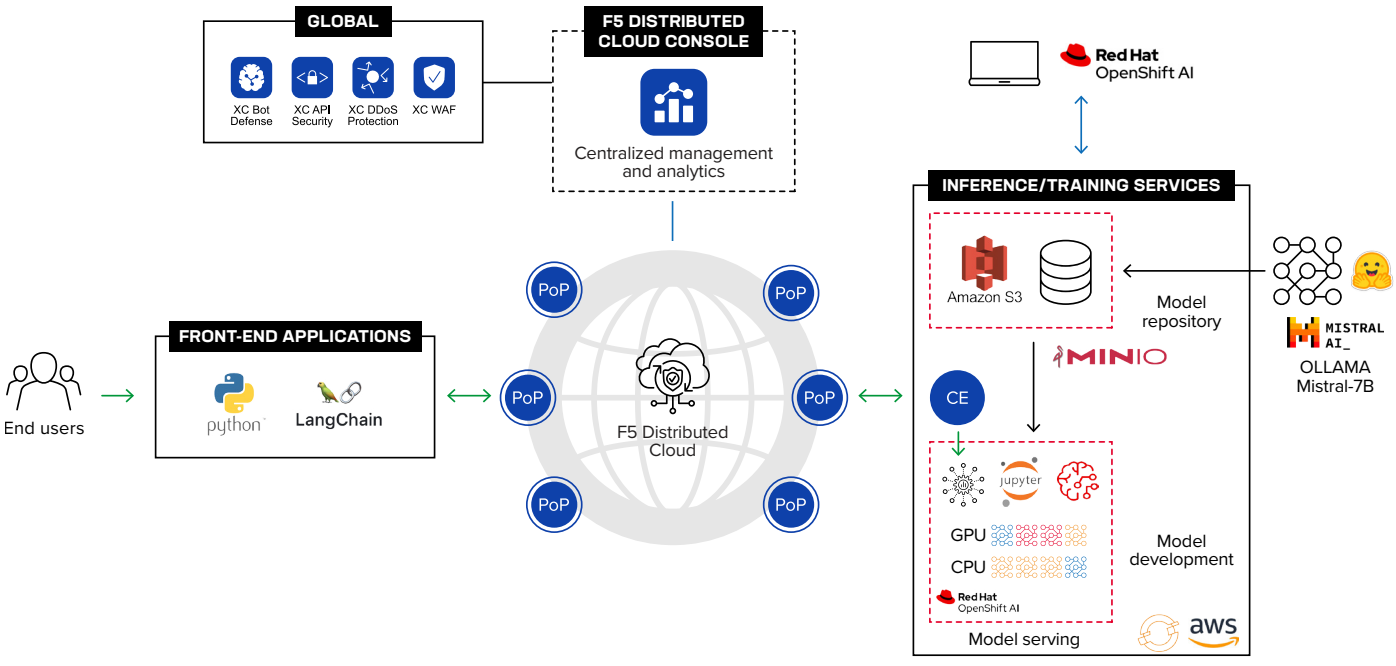

Figure 1: As AI model serving deployments become increasingly complex, protecting sensitive data and ensuring secure API interactions are critical. F5 Distributed Cloud Services provides robust API security for Red Hat OpenShift AI, addressing critical challenges including unauthorized access and data exfiltration.

Next steps

Experience F5 Distributed Cloud Services with a free enterprise trial.

Find out how F5 products and solutions paired with our technology alliance partners can enable you to achieve your goals.