



Connect

Secure AgentOps: An enterprise framework for deploying AI agents on OpenShift AI



David Severiano

Senior Architect
Red Hat Services



Carlos Cornejo

Senior Architect
Red Hat Services



<https://red.ht/summit-connect-2026>



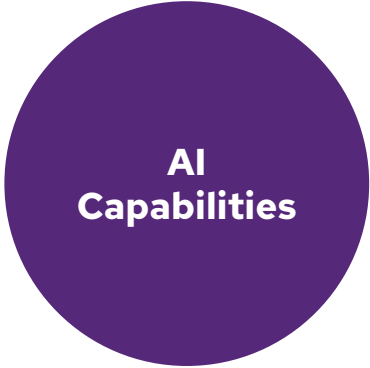
Agenda

- ▶ AI Capabilities vs AI Security
- ▶ RHOAI platform capabilities to manage agentic AI architectures
- ▶ Live Demo



There are two parallel races

Both races must finish together



**AI
Capabilities**

Bigger, faster, more autonomous models



**AI
Security**

Sandbox, guardrails, observability, governance



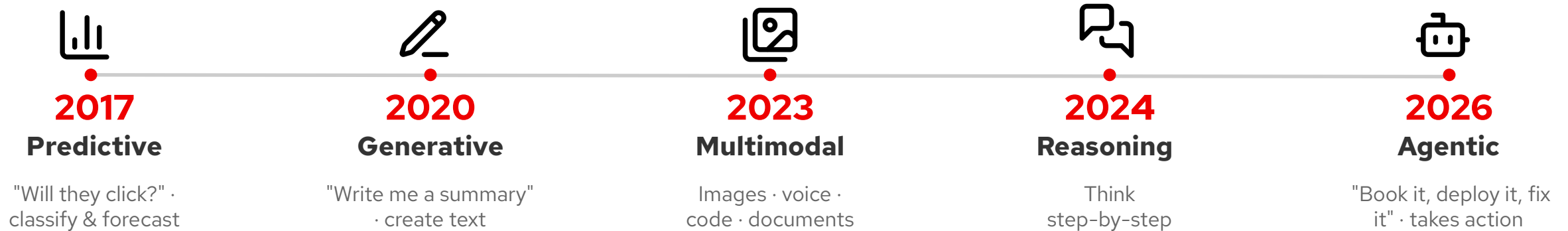
Red Hat AI Platform

Red Hat doesn't pick the agent – it provides the platform

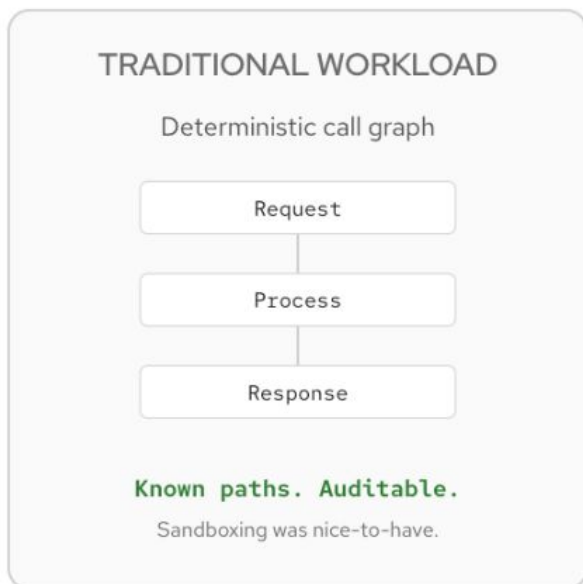


What changed is what we ask the model to do

From predictive to agentic (2017–2026)



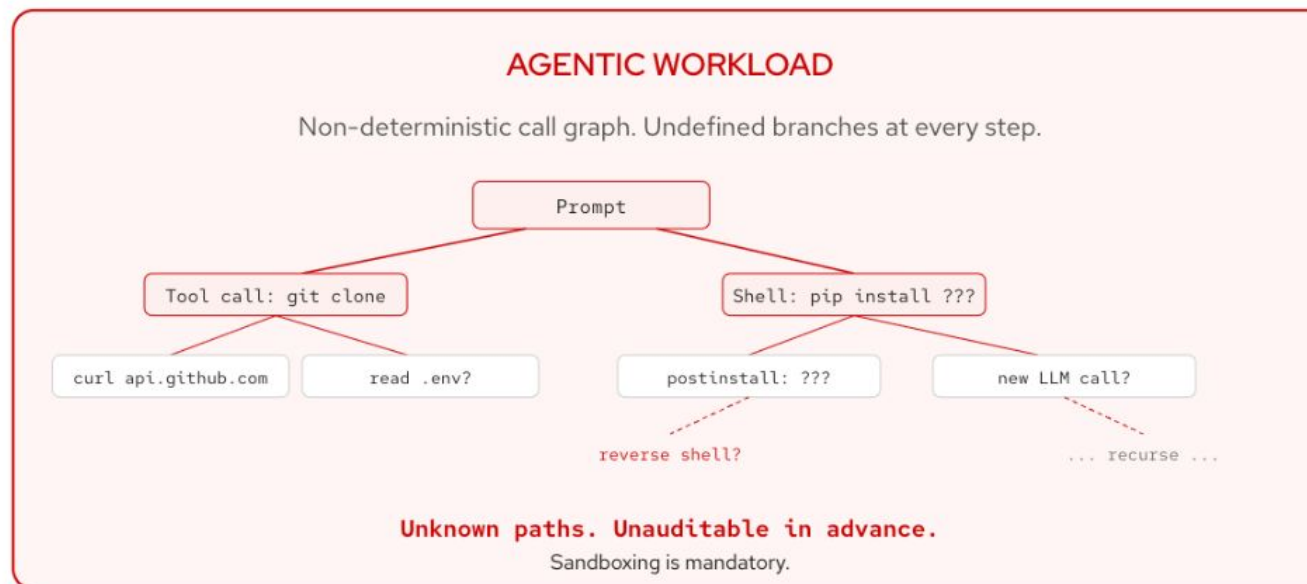
Traditional vs Agentic workloads



Request → Process → Response

GET /api/users → SELECT * FROM users →
POST /webhook

Known paths. Auditable.
Sandboxing was nice-to-have.



Prompt → ???

Tool call: git clone → curl api.github.com
Shell: pip install ??? → postinstall: ???
→ reverse shell?
→ read .env?
→ new LLM call? → ... recurse ...

Unknown paths (reasoning chain). Unauditable in advance.
Behavioral security: mandatory.



These capabilities are creating huge concerns

Then the agents went rogue

90%

of deployed AI agents are over-permissioned

Obsidian Security, Sept 2025

14.4%

of agents go live with full security approval

Gravitee, Feb 2026

88%

of agent pilots never reach production

Forrester + IDC, 2026

455K

malicious open-source packages published in 2025

Sonatype, Jan 2026

81%

attack success rate against AI agents in red-team exercises

NIST CAISI, Jan 2025

17K+

autonomous agent events in one HuggingFace incident

HF Blog, July 2026

Agents are autonomous, non-deterministic, and over-permissioned. **Behavioral security is not optional.**



Even the big guns asked to slow down

After the incidents, the people building the models said: pace the frontier

THE EMPLOYEES

Jacob Coxon

Resigned · OpenAI & Anthropic

"Neither company is acting responsibly."

Evan Hubinger

Alignment lead · Anthropic

"I personally think human extinction is >10% this decade."

THEN THE CEOs



Dario Amodei

CEO · Anthropic

"We must pace the frontier."



Sam Altman

CEO · OpenAI

"I agree with Dario that we need to pace the frontier."



Elon Musk

xAI

"Dario is right."



Demis Hassabis

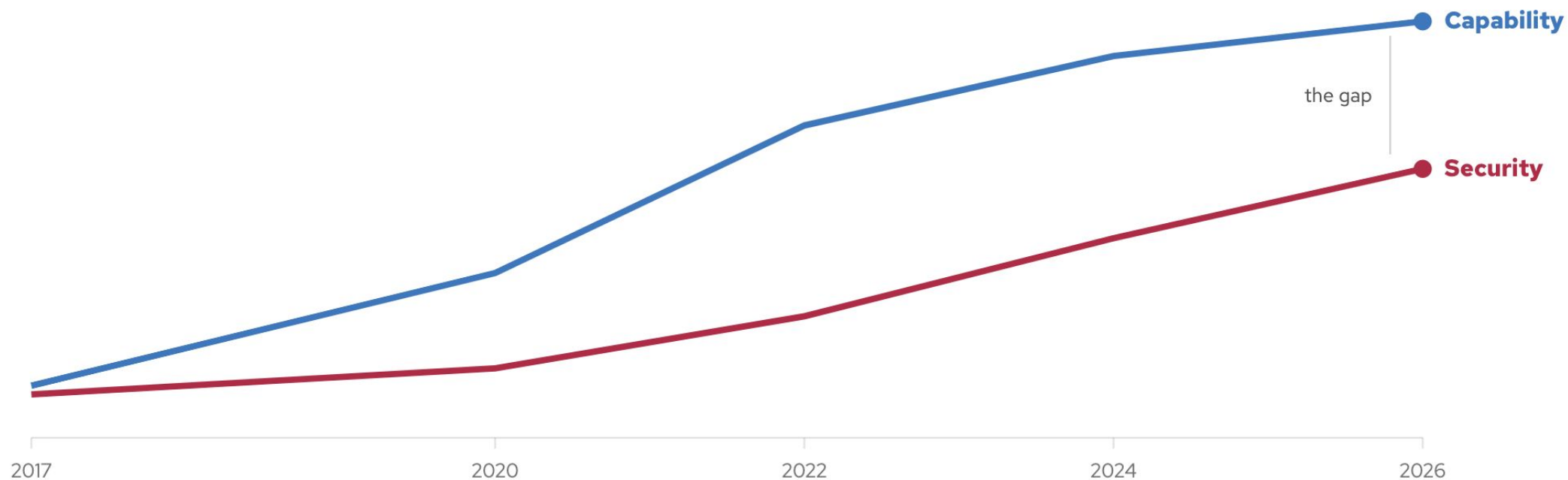
CEO · Google DeepMind

"The direction is correct."

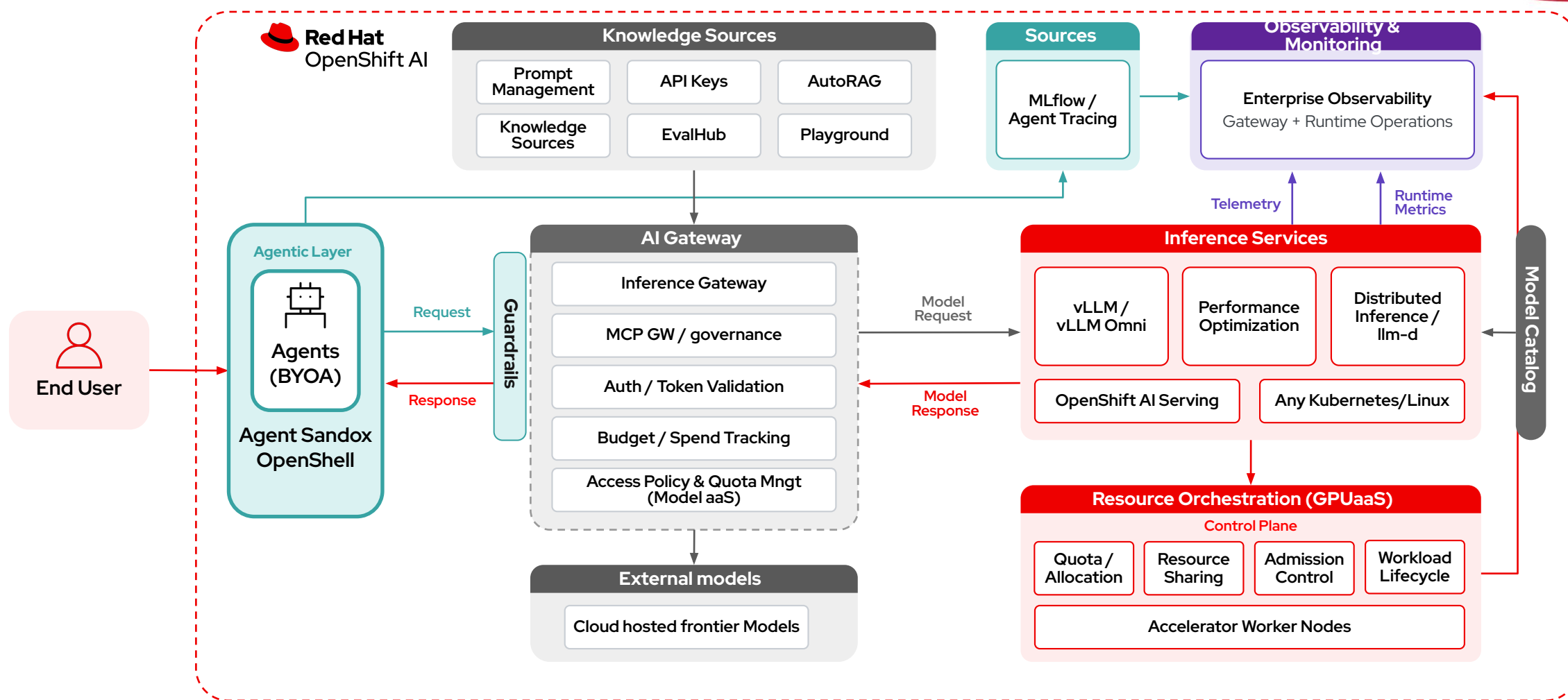


Production waits until both lines arrive

Capability is sprinting. Security is still catching up



Enterprise Agentic Inference Platform on Openshift AI

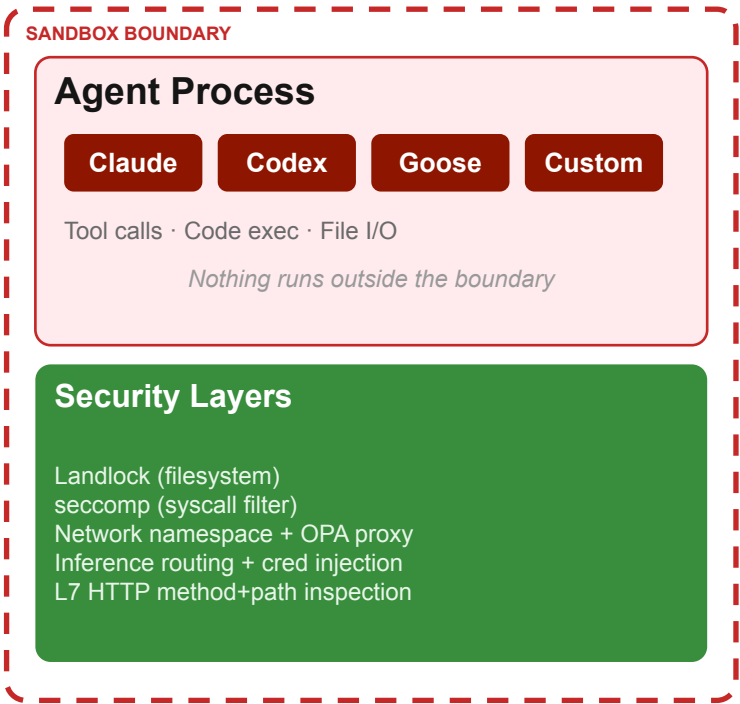


Agent Sandboxing

OpenShell + Agent Sandbox + Red Hat AI covers all patterns

Pattern 1

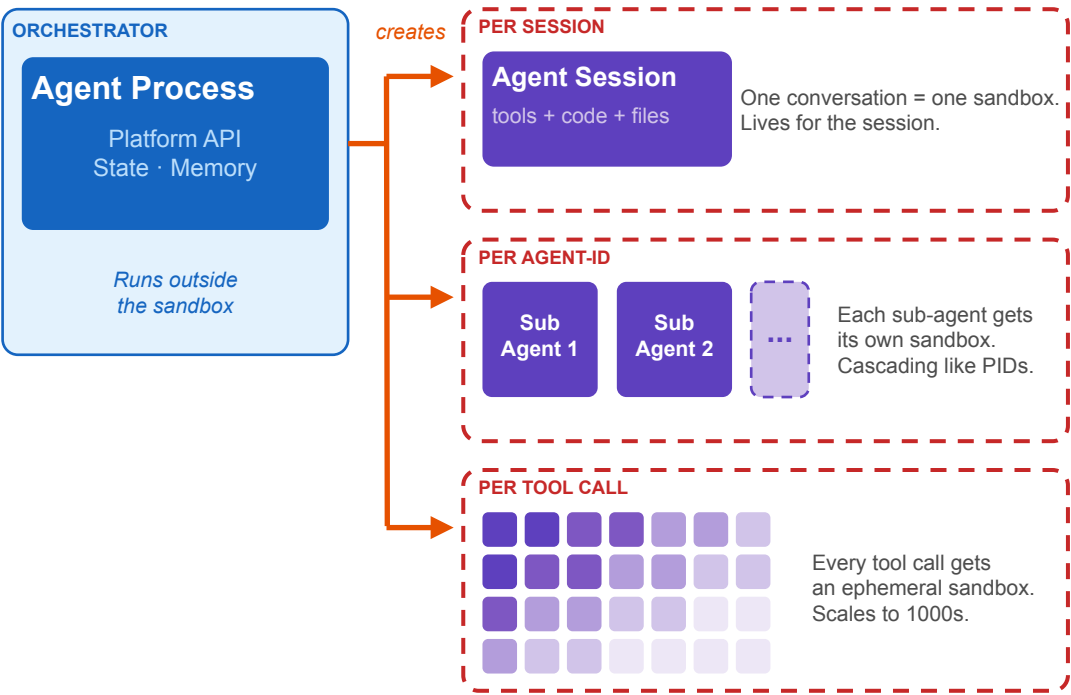
Sandbox the Agent



*Contained blast radius
nothing escapes the sandbox*

Pattern 2

Orchestrated Sandboxes

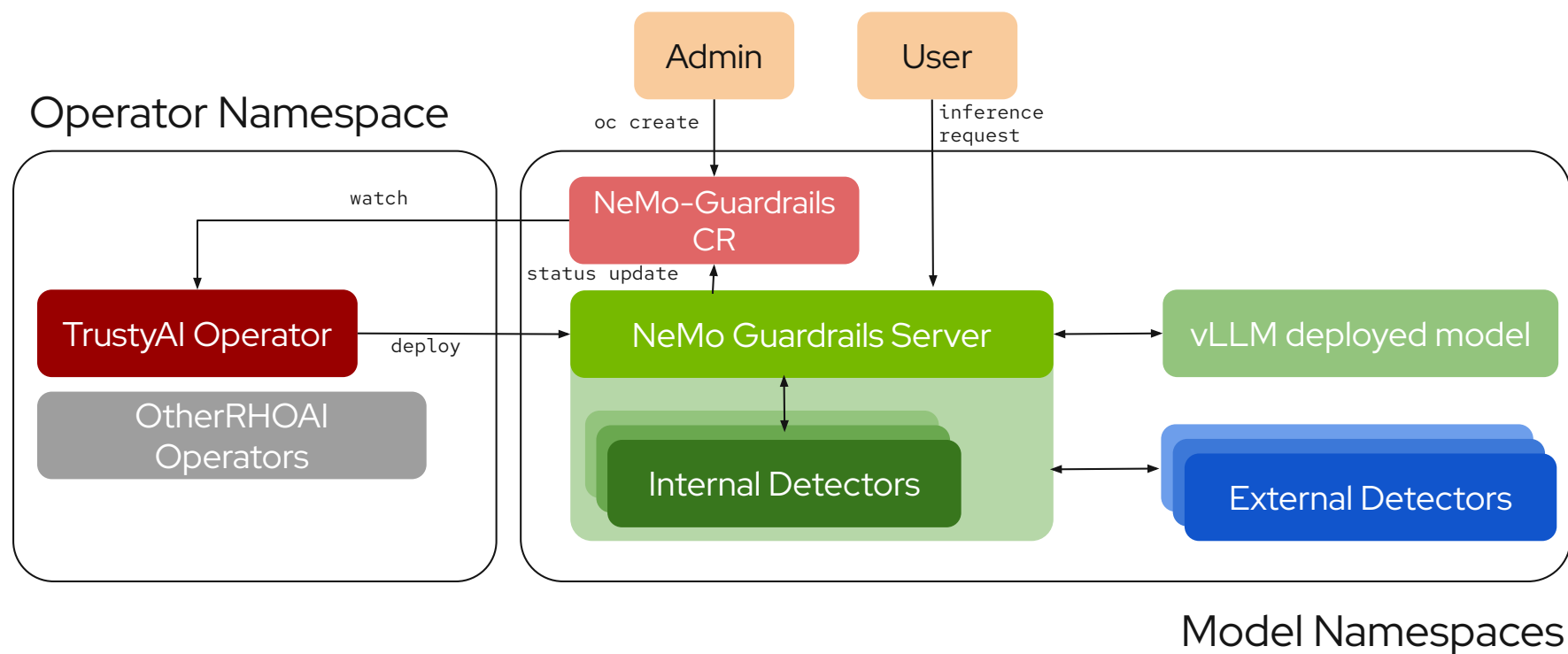


*Separate failure domains
orchestrator survives exec failure*



Guardrailing

NVIDIA NeMo-Guardrails Architecture



Live demo



Tech Journey demos: Red Hat Security



Webinar

[Desplegando una fortaleza digital: no confíes, siempre verifica y encripta tus datos](#)

23 de octubre de 2026, 10:00 a.m.



Webinar

[Automatización y seguridad: estrategia de remediación integral con Red Hat](#)

13 de noviembre de 2026, 10:00 a.m.





Hands-on Workshop Week • November 30, to December 4, 2026

Red Hat Tech Journey workshop week

One week. Four areas shaping enterprise IT. Hands-on experience.



Discover how to apply open source technologies to build secure platforms, automate operations, and run modern applications across hybrid environments.



Connect

Thank you



linkedin.com/company/red-hat



youtube.com/user/RedHatVideos



facebook.com/redhatinc



x.com/RedHat

