



James Labocki

Senior Director of Product
Management

AI tokenomics in the public cloud





Any accelerator

NVIDIA

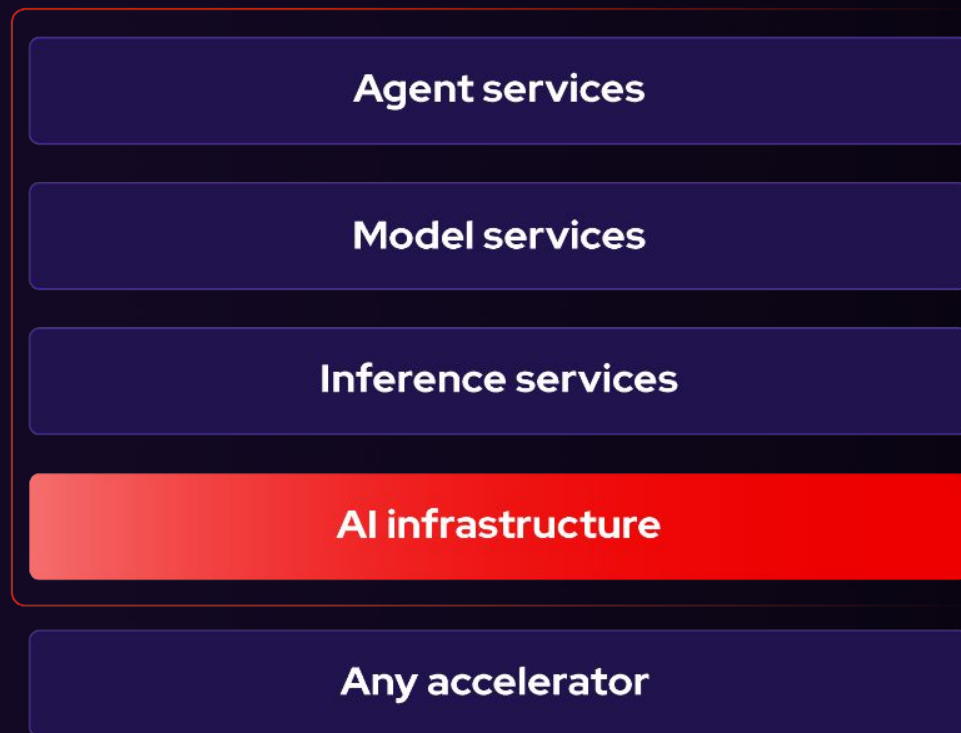
AMD

Intel

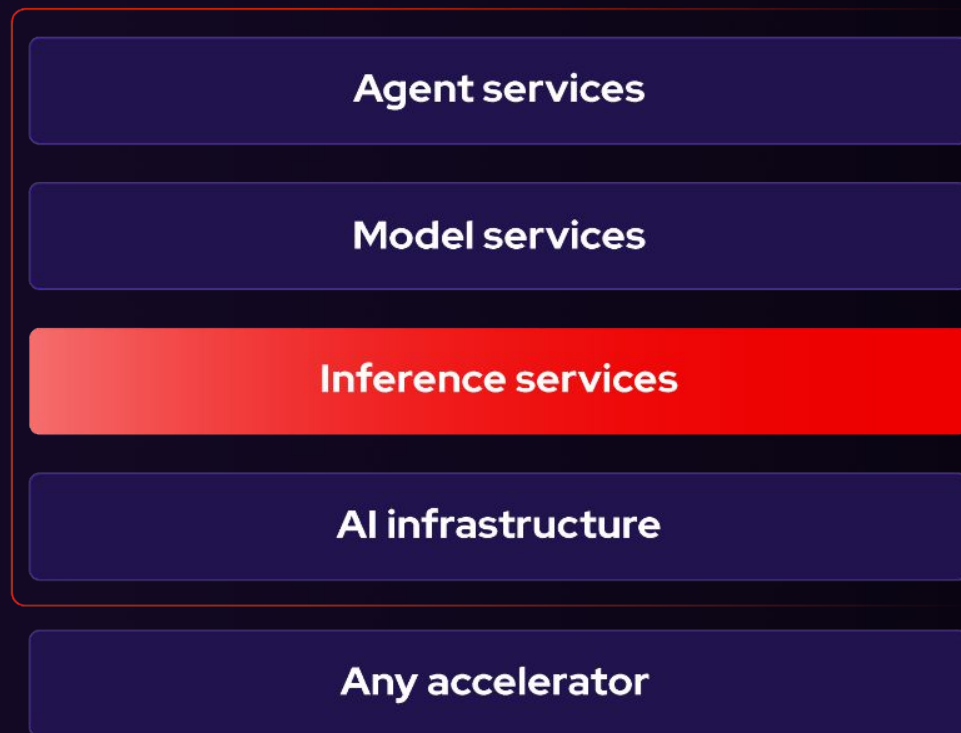
AWS

Google Cloud

Microsoft Azure



Integrated agentic OS
Foundation of AI system
Day 0 hardware support



Token economics layer

Distributed inference (vLLM, llm-d)

Open standards



Agent services

Model services

Inference services

AI infrastructure

Any accelerator

Model-as-a-Service

AI gateway

Enterprises need many models

Red Hat validated models on Day 0



Agent services

Model services

Inference services

AI infrastructure

Any accelerator

Bring your own agents

AgentOps

Agent identity

Agent lifecycle

MCP services and tools

Agent observability

How Shell-E is built

From the GPUs up





BNP PARIBAS

BNP Paribas

“With OpenShift AI, we provide standardized environment to data scientist engineers to allow us to build and deploy and scale AI workloads consistently.”

Pascal Guerineau
Architect
BNP Paribas



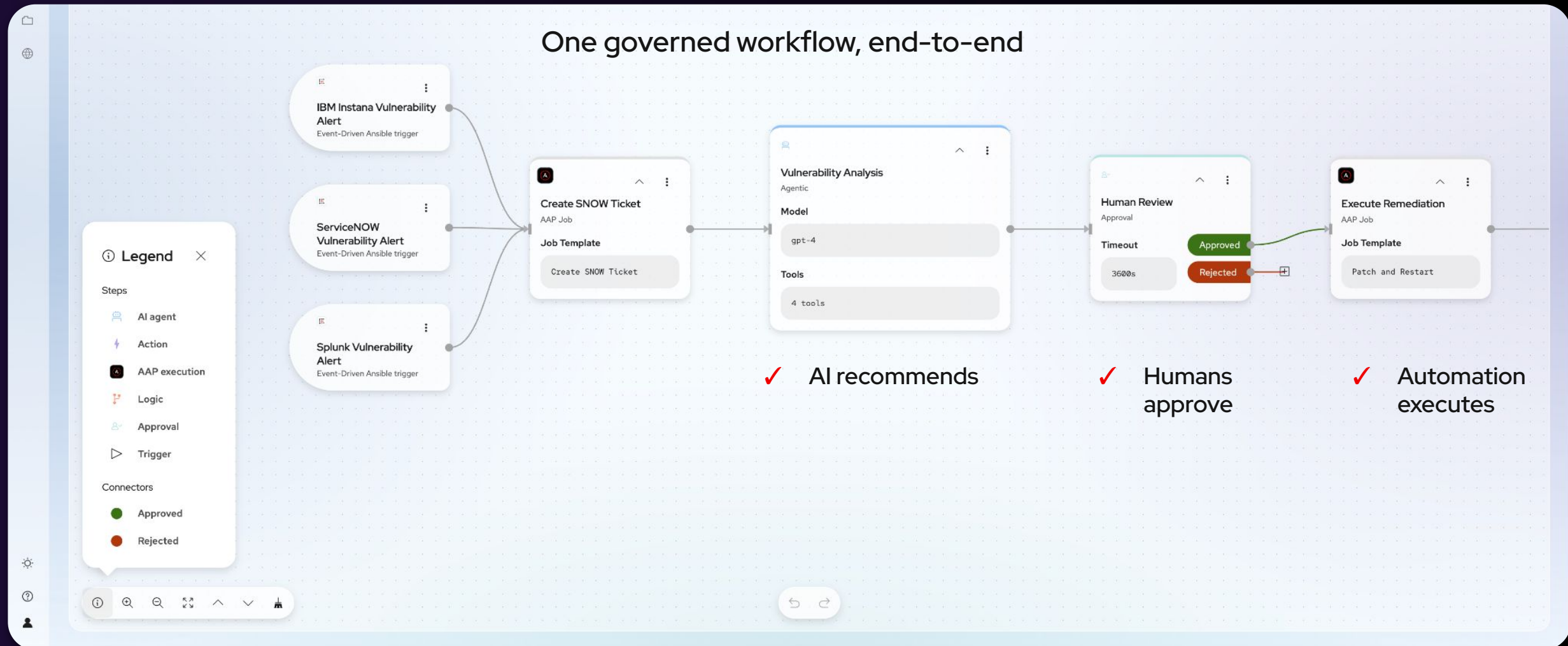
\$600M

in additional business value generated from AI across 1,000 use cases

- Handled over 900 million input tokens per day while meeting strict compliance.
- Scaled the platform to support thousands of internal users across multiple regulated entities.

Introducing the Automation Orchestrator for Ansible

The trusted execution layer for AI-driven IT operations



The ground has shifted

Agentic development is no longer on the horizon –
AI is already doing the engineering work, at scale.

ADOPTION

90%

of developers now use AI in their work – up 14 points in a single year.

Google DORA, 2025

OUTPUT

98%

more pull requests merged by AI-assisted developers vs. their peers.

Faros AI telemetry, 2025

THE NEW DEFAULT

75%

of engineers will use AI coding assistants by 2028 – from <10% in 2023.

Gartner forecast

PRODUCTION AGENTS

40%

of enterprise apps will embed AI agents by end of 2026 – from <5% in 2024.

Gartner forecast

The Mythos Moment

When discovery outpaces defense



AI discovers vulnerabilities at machine speed

Frontier models find critical zero-day vulnerabilities autonomously – including bugs hidden for 27 years – at under \$50 each.



Vulnerability discovery exceeds capacity

Discovery is exponential. Remediation is linear. Less than 1% of AI-discovered vulnerabilities have been patched.



Discovery is no longer the bottleneck

Response is. 4.8M Unfilled cyber roles. Growing 19% year over year. The National Vulnerability Database is shifting to selective enrichment.



Shift to exposure management

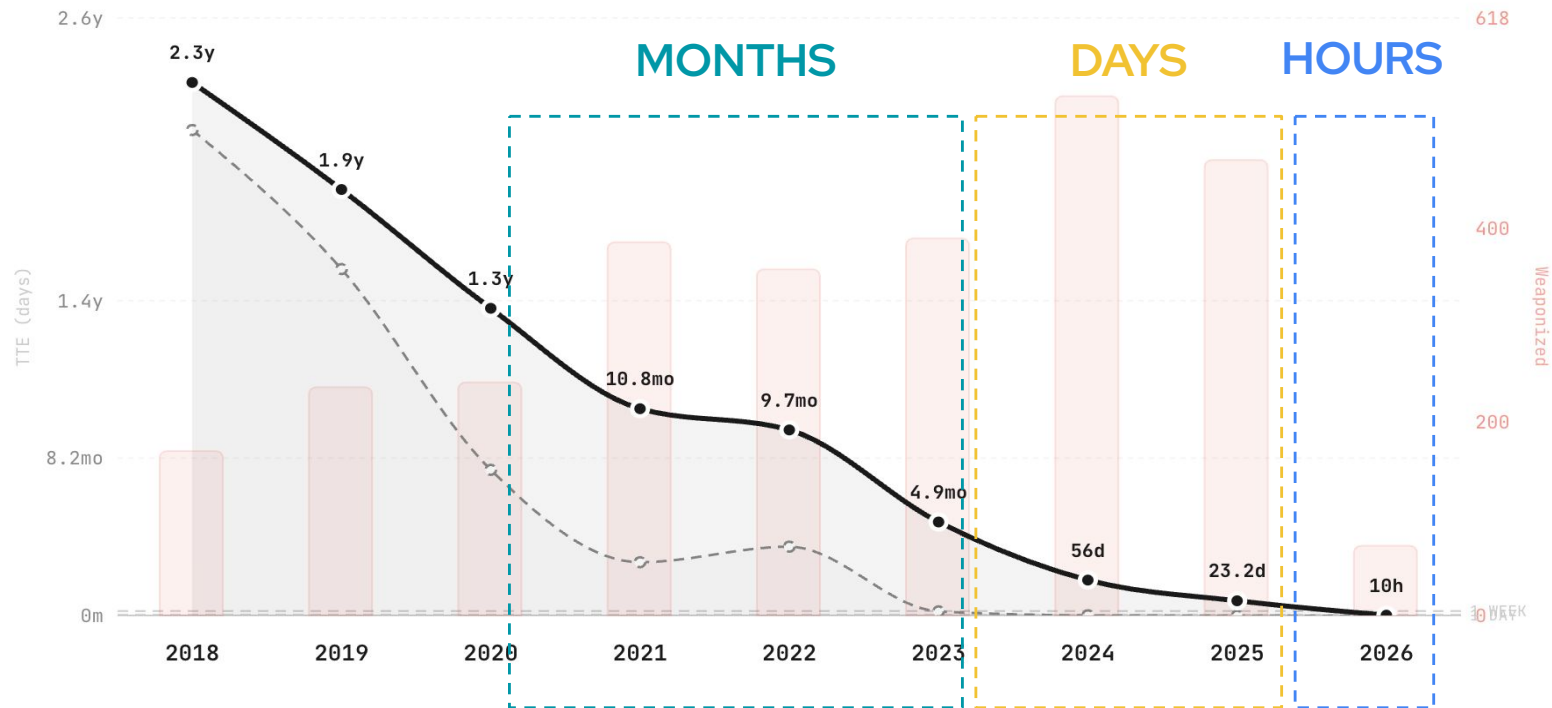
The number of published CVEs has grown more than 520% since 2016. Cost of attacks in 2025 was ~\$60B, projected to grow to \$138B by 2031.

AI is accelerating the security vulnerability landscape

From Vulnerability to Exploitation

TTE (Time-to-Exploit) measures the gap between CVE disclosure and confirmed exploitation

— Mean TTE (10% trimmed, days) - - - Median TTE (days) ■ Weaponized Exploits (count)

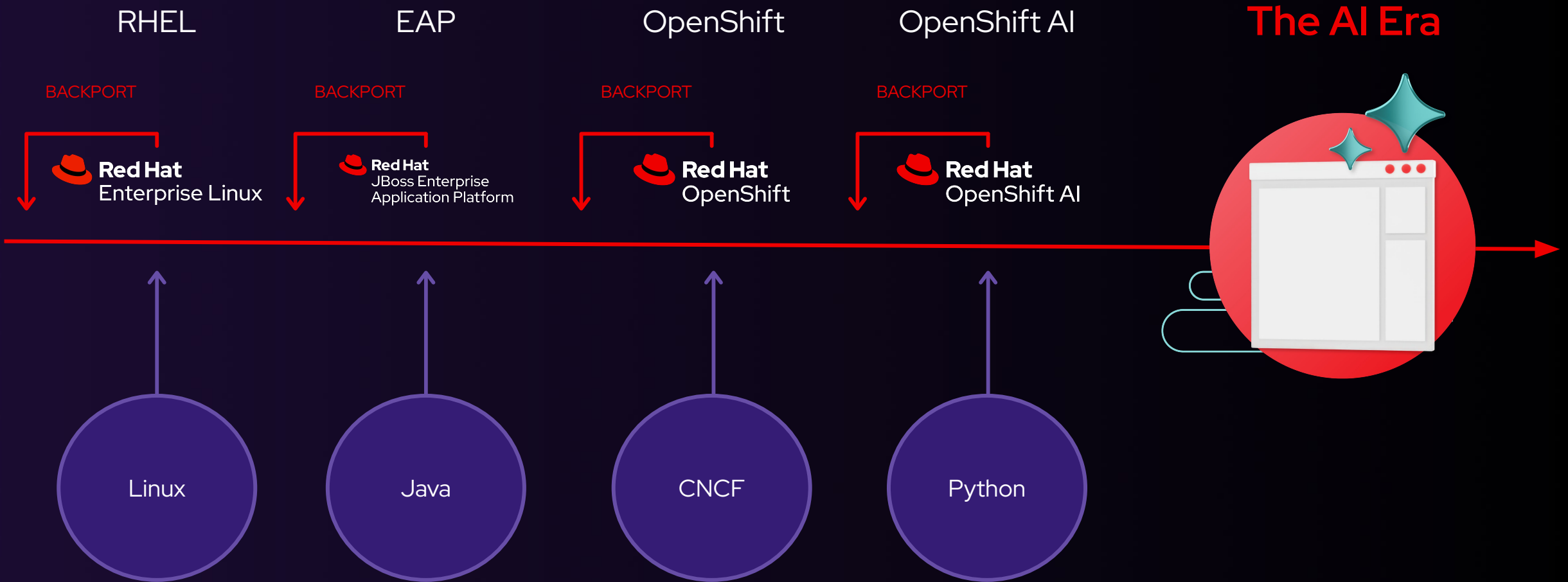


Based on 3,532 CVE-exploit pairs from trusted sources (CISA KEV, VulnCheck KEV & XDB)

zerodayclock.com

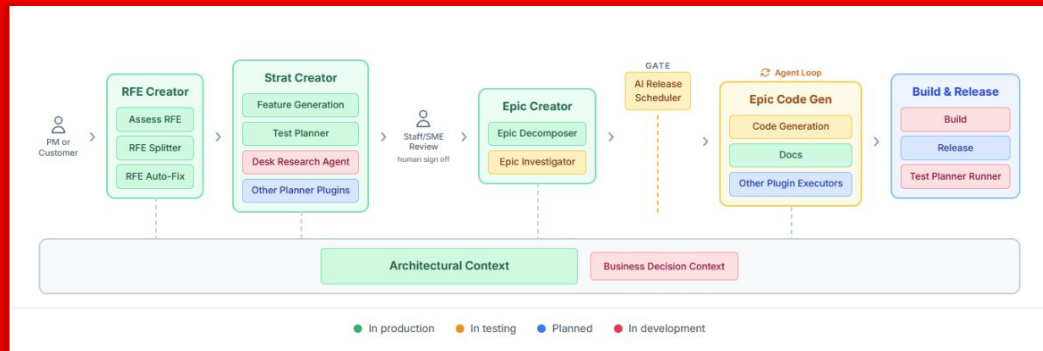
Source: [The "AI Vulnerability Storm": Building a Mythos-ready Security Program](#) Zero Day Clock

Your trusted source of secure open source software for 30 years

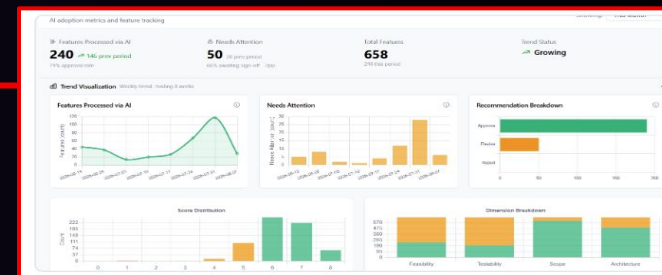


Red Hat Agentic Software Development Lifecycle

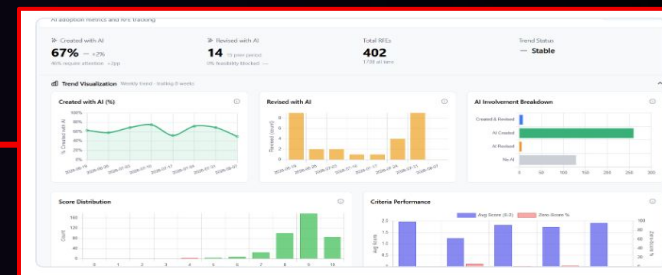
Red Hat End-to-End Agentic SDLC



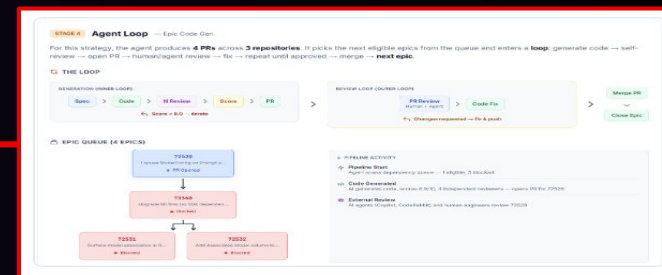
Planning



Resource management



Building & testing



Future-proofing

