



KubeCon



CloudNativeCon

Europe 2026

#KubeCon #CloudNativeCon

Your AI Is Vulnerable

**How KitOps Turns KServe
Into a Zero-Trust AI Platform**



About Us

Brad Micklea

CNCF Maintainer: KitOps & ModelPack
CEO of Jozu (DevOps Platform for Secure AI)

Ex-GM of AWS API Gateway
Ex-VP of Red Hat Developer Tools

LinkedIn Bio



Gavrish Prabhu

Kserve and Envoy AI Gateway Maintainer
Founding Engineer Nutanix AI

LinkedIn Bio



KServe: Inference Platform

(CNCF Incubating)



KubeCon

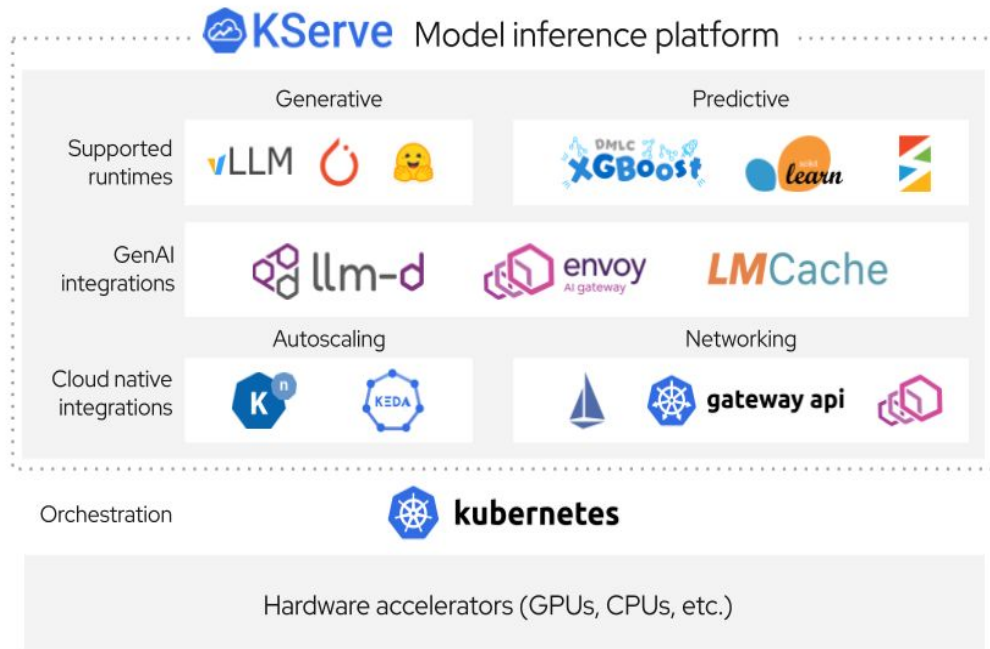


CloudNativeCon

Europe 2026

What it Does

- Unified platform for both Generative and Predictive AI inference
- Encapsulates the complexity of autoscaling, networking, health checking, and server configuration
- Scale-to-zero, canary rollouts, and A/B testing out of the box



KServe Makes Serving Easy You Need to Add Security



KubeCon



CloudNativeCon

Europe 2026

You need to:

- Verify the model / agent served is the same one that was approved
- Validate signatures or attestations before loading weights
- Track which dataset, training run, or pipeline produced the artifact
- Provide provable / audit-ready evidence for chain-of-custody

Zero-Trust for Inference

The same principles you apply to containers, applied to AI workloads.

Never Trust the Workload

Every AI workload must **prove its identity and integrity** before serving.

Verify at Deploy Time

Build-time scans are necessary but not enough.

Verify again at deployment.

Fail Closed

Missing signature?
Missing attestation?

Pod does not start.

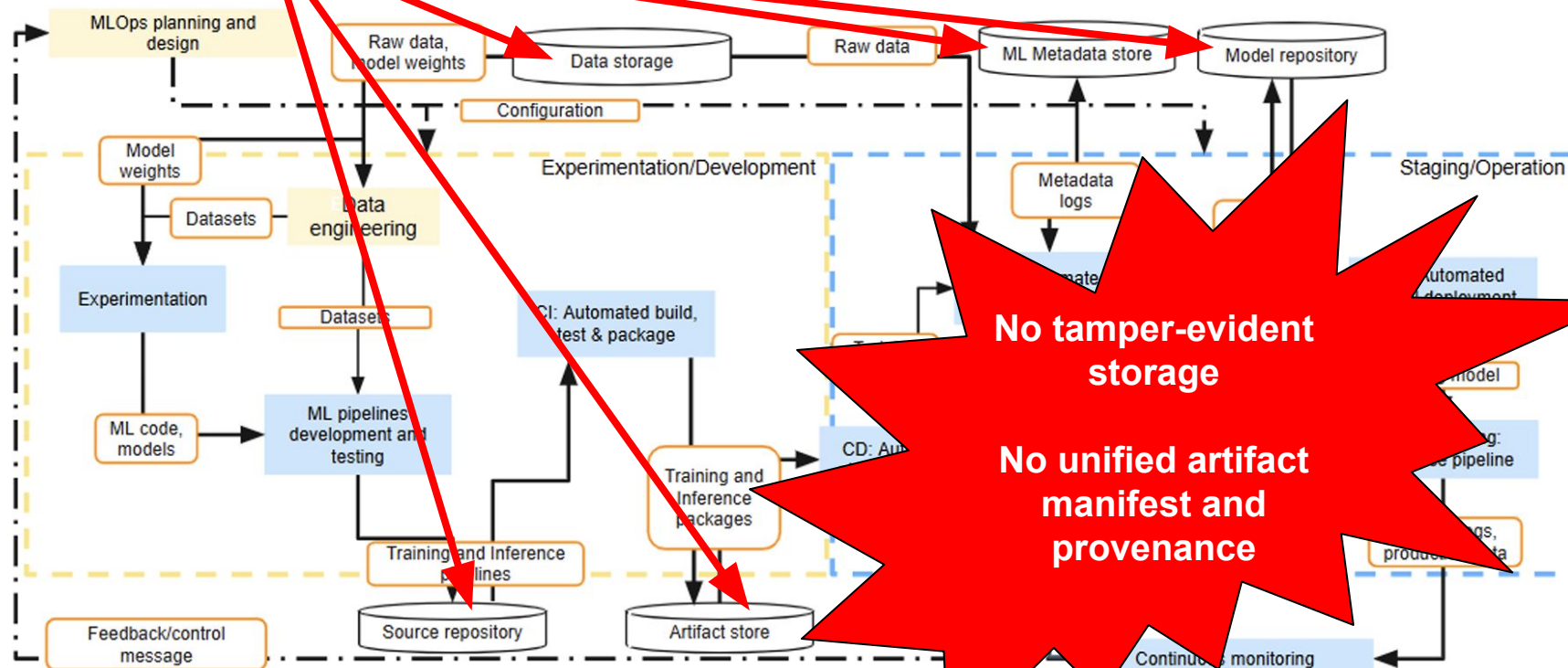
For microservices you have git...but not for AI

- Single repository; unified versioning for all artifacts
- Compatible with tools
- Easy to pull

“Git LFS – even though it stands for Large File storage – was just never meant for the type of large files we handle in AI, which are not just large, but *very very* large 😊”

- *Julien Chaumond, Hugging Face CTO*

Typical: 5 disconnected versioning locations 🤔



KitOps: OCI-native packaging for AI/ML artifacts

(CNCF Sandbox)

KitOps ModelKits bundle model weights, datasets, prompts, and config into an OCI artifact. Stored in the registry you already run.

What goes in...

- Any model, agent, or MCP server
- Training/validation datasets
- Serving code and configs
- Prompts & skills
- (Anything else you need)

What you get...

- SHA-256 digests on every layer
- Works with any registry
- Proveable chain-of-custody
- Standard OCI distribution
- Content addressable storage

KitOps in a >\$30B Logistics Company

A >\$30B logistics company, with 3000 offices in 90 countries is standardizing on KitOps for their AI/ML workloads.

They Use KitOps Because They Need:

- Freedom for teams to choose model and agent development tools
- Provenance and chain-of-custody across the AI project lifecycle
- Compatibility with on-prem and all clouds
- Open standards (KitOps, ModelPack, Kubernetes); no vendor lock-in
- Secure packaging compatible with Kubernetes

Tamper-evident packaging that's easy to use

Development

Data Scientists use notebooks, experiment trackers, training tools, etc... as usual.

ModelKits are created automatically (SDK or CLI)

Deployment

ModelKits are compatible with pipelines and the **Kubernetes** ecosystem.

Build Phase: Data Scientists



KubeCon



CloudNativeCon

Europe 2026

PHASE	TASKS	ARTIFACTS	KITOPS MODELKIT
1. Build	Select base model Curate dataset Run training	Base model Training dataset Training scripts Hyperparameters Training run metrics	Base model Training dataset & scripts Hyperparameters Training Metrics

- Team can use KitOps ModelKits with any tool (freedom)
- AI projects are shared through their existing OCI registry (easy)
- During handoffs teams can pull only what they need (`kit --filter=prompts`)

Verify Phase: ML Eng & Compliance



KubeCon



CloudNativeCon

Europe 2026

PHASE	TASKS	ARTIFACTS	KITOPS MODELKIT
2. Verify	Human review Compliance data extract	Human approval Compliance docs Evaluation results	Base Model Training dataset & scripts Hyperparameters Compliance docs <u>ATT</u>: Human approval <u>ATT</u>: Evaluation results

- Add verification artifacts to the ModelKit, or as attestations
- Human approvals should be signed attestations attached to ModelKits
- Remember: anything added has provable chain-of-custody; references don't

Deploy Phase: SRE and Platform Eng



KubeCon



CloudNativeCon

Europe 2026

PHASE	TASKS	ARTIFACTS	KITOPS MODELKIT
3. Deploy	Optimization for serving Deploy pipeline	Quantization configs Optimized weights Performance benchmarks RAG pipeline config LoRA adapter config Deployment script	Base Model Training dataset & scripts Hyperparameters Compliance docs ATT: Human approval ATT: Evaluation results Quantization configs Optimized weights RAG pipeline config LoRA adapter config Deployment script

Provenance for all artifacts
Unified package and version
Signing and attestations for compliance / security

Level 1: Verify Before You Serve (Automatic)

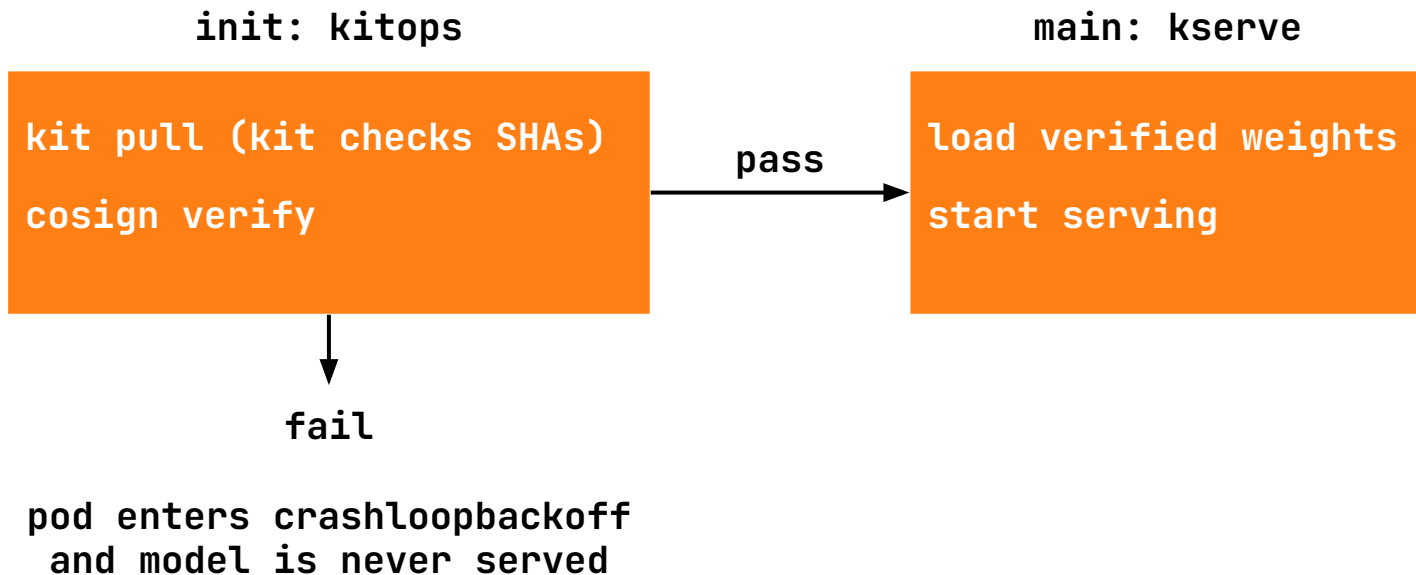


KubeCon



CloudNativeCon

Europe 2026



How You Can Add Security to AI Serving

Level 2: Apply Policies Before Admission

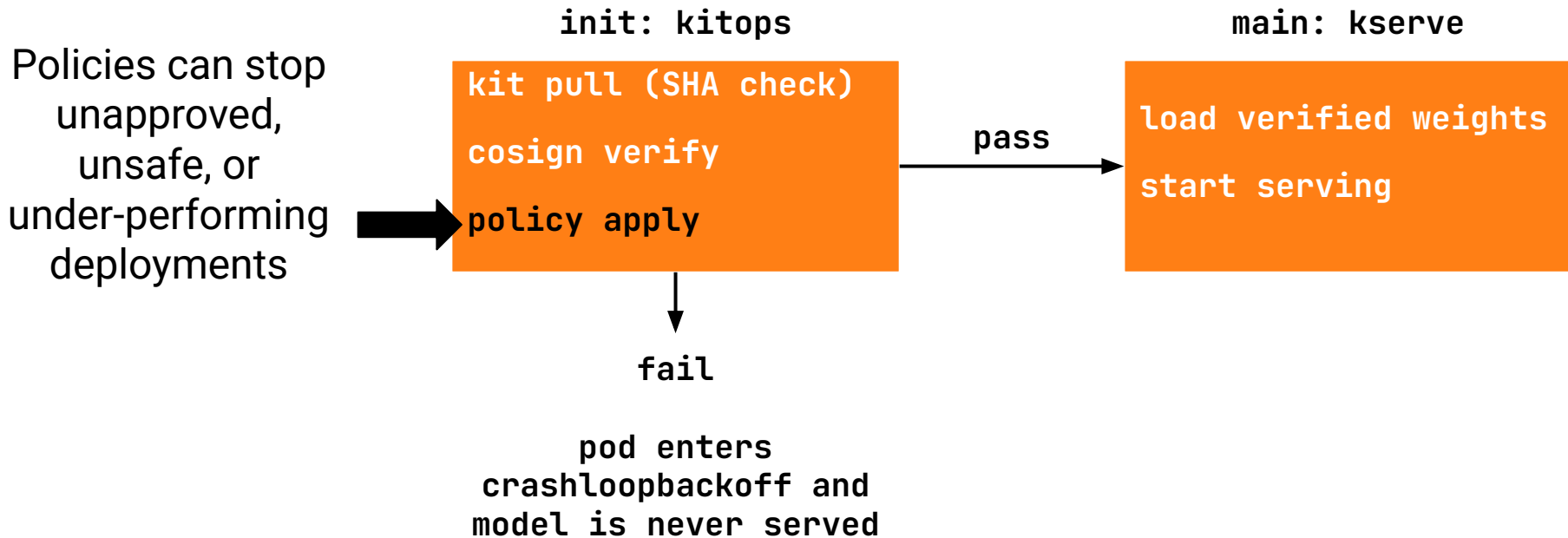


KubeCon



CloudNativeCon

Europe 2026



How You Can Add Security to AI Serving

Level 3: Model Registry & Quarantine



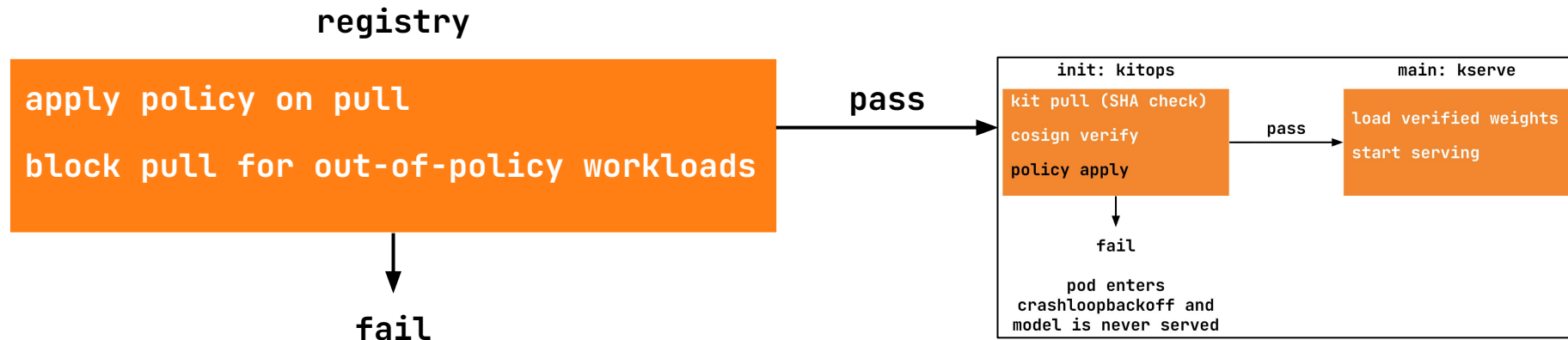
KubeCon



CloudNativeCon

Europe 2026

Quarantine policy should be more strict than admission policy.



ModelKit can't be pulled.
Acts as "quarantine" for
dangerous workloads

Demo: Zero-Trust with KitOps + KServe

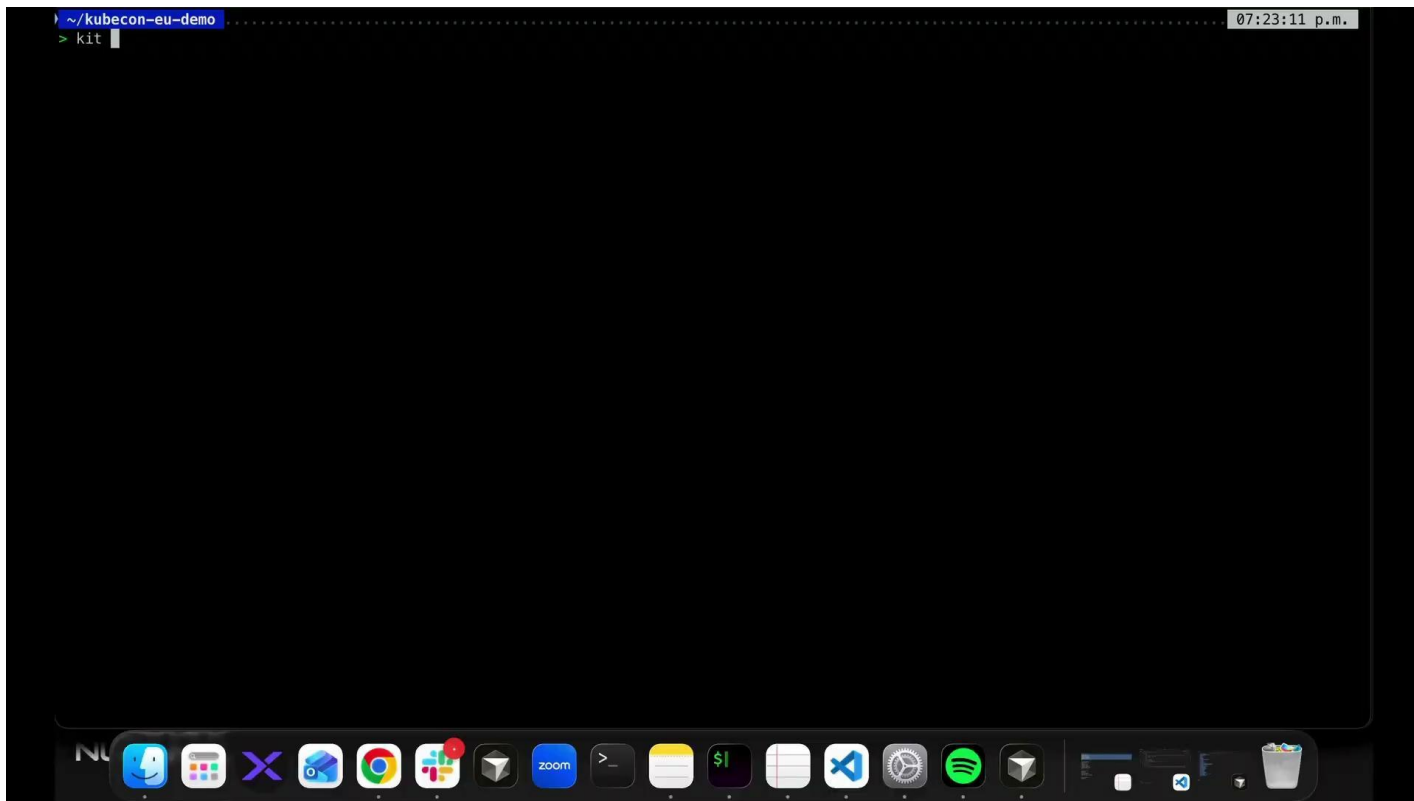


KubeCon



CloudNativeCon

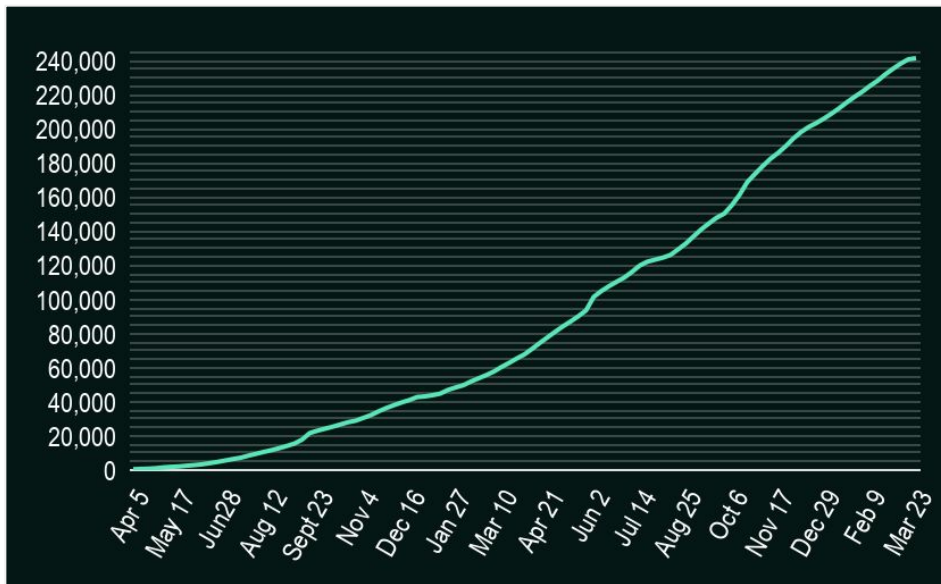
Europe 2026



KitOps Traction: >240,000 Downloads

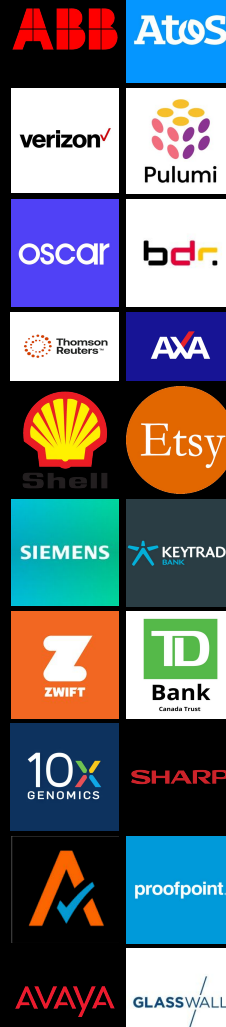


KitOps Downloads



"We see KitOps as the gate between development and production. Controlling and securing that transition is critical."

- DevOps Lead
US Dept. of Veterans Affairs
Digital Health Group



Harden KServe Next Week



KubeCon



CloudNativeCon

Europe 2026

Day 1: Secure Packaging & Signing

- Install KitOps CLI / SDK
- Package projects as ModelKits
- Sign with Cosign
- Push to your registry

Day 2: Verify

- Add verification to **init-container**
(Pin ModelKit references by digest)

Days 3 - 7:

- Capture approvals as attestations
- Add security scanning for models, agents, and datasets to your CI
- Add admission policy to deployments
 - Check for approvals
 - Check for scanned and safe models
- Pull all chain-of-custody information from ModelKits into compliance docs



KubeCon



CloudNativeCon

Europe 2026

#KubeCon #CloudNativeCon

Thank You



Learn More



KubeCon



CloudNativeCon

Europe 2026

Brad Micklea



Gavrish Prabhu



KitOps

