[SUCCESS STORY]

Imaginario

Scaling AI applications with compliance and efficiency using Code Capsules

- Deployment went from manual to streamlined via platform automation
- GPU costs dropped by ~90% with event-based scaling
- Cloud flexibility expanded to include client-owned AWS accounts
- Compliance became automated, secure, and region-aware

> Code Capsules helped us scale deployments across multiple global clients without increasing operational burden. More importantly, we cut our GPU costs by 90% while staying fully compliant.
>
> CTO

Customer overview

Imaginario is an AI startup serving major media clients with a compute-heavy web app that requires dedicated cloud deployments per customer, often within client-controlled environments—creating complex compliance and infrastructure demands.

The challenge

Imaginario needed a scalable, compliant way to deploy GPU-heavy, single-tenant AI apps across global cloud environments without overloading their engineering team.

Isolation

Each client required a dedicated, single-tenant setup

Cloud flexibility

Deployments had to work in both Imaginario's and clients' AWS accounts

Compliance at scale

Global regions were needed to meet US and EU data laws

DevOps burden

Manual deployments and 24/7 GPU use strained the team

The solution

CodeCapsules enabled Imaginario to simplify and scale their AI infrastructure with secure, single-tenant deployments across global AWS environments.

Specific deployments

Launch single-tenant environments across any AWS account or region

PaaS simplicity, full control

Maintain infrastructure ownership with the ease of a managed platform

Smart GPU scaling

Auto-scale GPU nodes to cut costs and avoid idle resource usage

Streamlined compliance

Ensure security and data sovereignty with isolated, client-specific setups