

Fortanix Confidential AI for AI Model Owners

Deploy Into Enterprise On-Premises Environments Without Losing Your IP.

Vast amount of enterprise sensitive data lives on-premises, blocked by data sovereignty and compliance laws. Frontier AI models, meanwhile, live in the cloud. The result is a hard gap between where the best AI runs and where the most critical data lives. Both sides want to close this gap. Enterprises need frontier models on their own infrastructure to unlock data they can never expose, while also minimizing inference latency.. AI labs want enterprise revenue, but not at the cost of model weights, which represents years of research and millions in investments.

The obstacle is inference. When a model runs, its weights are loaded into GPU memory in plaintext, fully exposed. Traditional security covers data at rest and in transit but stops at the moment that matters most-- in use, during inference. Anyone with host access can walk away with the IP. The core challenge for model owners is how to deploy frontier AI on-premises to accelerate revenue growth while keeping model IP completely protected.

Protect Your IP with Hardware-Based Proof

Fortanix Confidential AI is a joint Fortanix & NVIDIA solution that leverages Confidential Computing Technology for deployment of AI models on-premises inside hardware-enforced Trusted Execution Environments (TEEs). Local deployments enable faster response times, while running AI IP inside the TEEs ensures that model weights are protected from the underlying infrastructure, OS, and privileged administrators. The built-in composite cryptographic attestation, across GPU and CPU, verifies the exact hardware and software environment to confirm that the enclave is genuine and untampered with. Only then is the model securely decrypted inside the TEE and ready for secure execution. If there is a slightest deviation from the verified attestation, the encryption key is revoked, and the AI workload is terminated.

Benefits



Unlock New Revenue Streams

Unlock regulated markets, support sovereign AI initiatives, and deliver low-latency, real-time performance



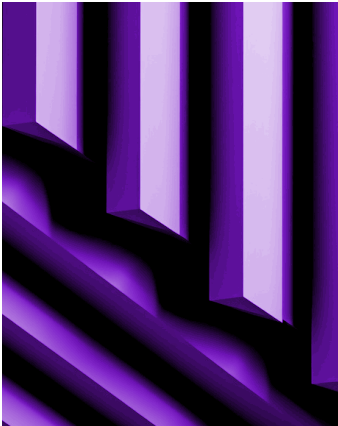
Scale Deployments Without Exposure

Package model weights inside an encrypted, attested CVM image once and deploy to any number of enterprises



Gain Trust Through Compliance

Offer verifiable, auditable proof of end-to-end AI workload protection



“Our models represent years of proprietary research and engineering. Working with Fortanix on NVIDIA Confidential Computing infrastructure lets us give organizations in government, healthcare, and finance the ability to run our models on their own servers, on their own data.”

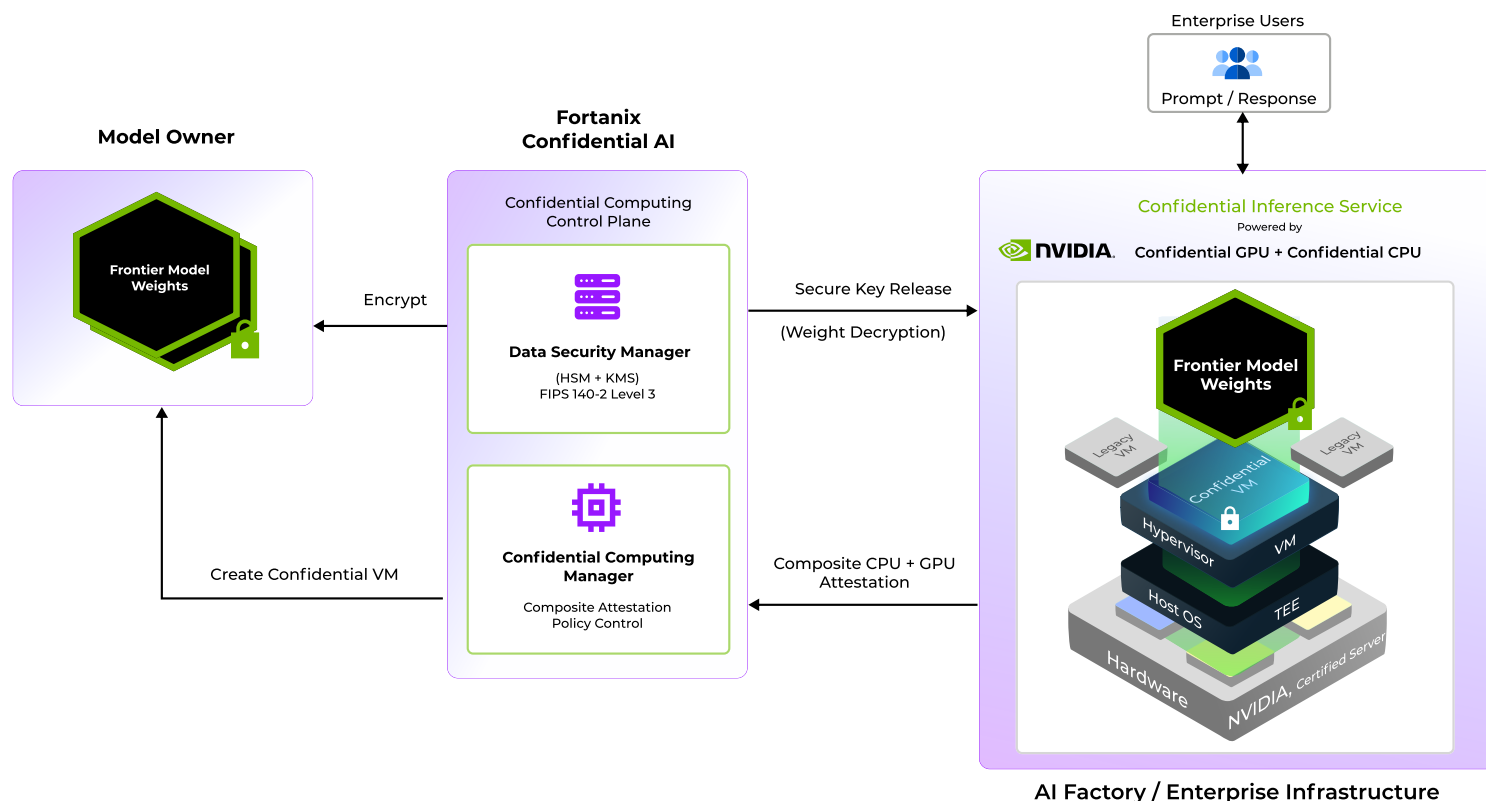
Kuba Abramczyk,

Forward Deployed Engineer, ElevenLabs

ElevenLabs

How It Works

- ✓ **Encrypt & Package:** Model weights, code, and configurations are encrypted with a non-exportable key stored in Fortanix DSM (Data Security Manager), a FIPS 140-2 Level 3 Hardware Security Module, available on-prem and as SaaS. The encrypted model is packaged into a Confidential VM image for distribution.
- ✓ **Composite Attestation:** Fortanix CCM (Confidential Computing Manager) manages the TEE, enforces attestation policies, and verifies the exact hardware CPU + GPU and software environments, confirming the enclave is genuine and untampered with before issuing a certificate.
- ✓ **Secure Key Release:** Only after both attestation checks pass are encryption keys released exclusively to the verified TEE. The frontier model decrypts and runs inside the enclave. No admin, cloud operator, or privileged insider can access the weights.



Supported Hardware

- GPU: NVIDIA Hopper, Blackwell Confidential Computing
- CPU: Intel TDX, AMD SEV-SNP
- Deployments: On-premises, Sovereign AI Factory, Public Cloud.

About Us

Fortanix is the global leader in data and AI security and a pioneer of Confidential Computing, delivering a unified platform to protect sensitive data, AI models, and applications across on-premises and multi-cloud environments at rest, in transit, and in use. Built on hardware-enforced security, Fortanix enables workloads to run in tamper-proof, isolated enclaves, protecting against data leakage, model extraction, and unauthorized access, even from privileged insiders.