Duality

CUT AI COSTS ON NEOCLOUD WITHOUT EXPOSING YOUR DATA

Run your LLM workloads on lower-cost Neocloud GPUs, with your prompts and data processed inside Trusted Execution Environments at the Neocloud's premises.

Challenge: Lower-Cost GPUs Come With a Trust Question

AI workloads are expensive to run, and Neocloud providers offer GPU capacity for less than the usual options. Moving your LLM there can cut your AI cost.

But a Neocloud is a provider you may not know, and its site is often in another region. Sending it prompts, customer records or internal knowledge means trusting its operators and its infrastructure with your most sensitive data, and many security teams will not sign off on that.

The Solution: Run in a TEE at the Neocloud, and Verify It Yourself

Duality Technologies runs your LLM inside a Trusted Execution Environment (TEE) at the Neocloud's premises. Your Duality node attests the enclave before anything is released, and your apps and clients then connect straight into it, so prompts and data are processed inside the TEE and protected while in use.

The result: you use the latest models on lower-cost capacity, and the Neocloud operators and cloud vendor have no access to your prompts or data inside the TEE.

Key Benefits

Lower AI Cost

Run LLM inference on lower-cost Neocloud GPU capacity.

Protected While in Use

Prompts and data are processed inside a hardware-isolated enclave.

Verify Before You Send

You attest the enclave yourself, so trust rests on evidence.

Any Region, Same Controls

The same protection applies wherever the Neocloud site is.

HEADQUARTERS
5 Marine View Plaza, Hoboken, NJ 07030

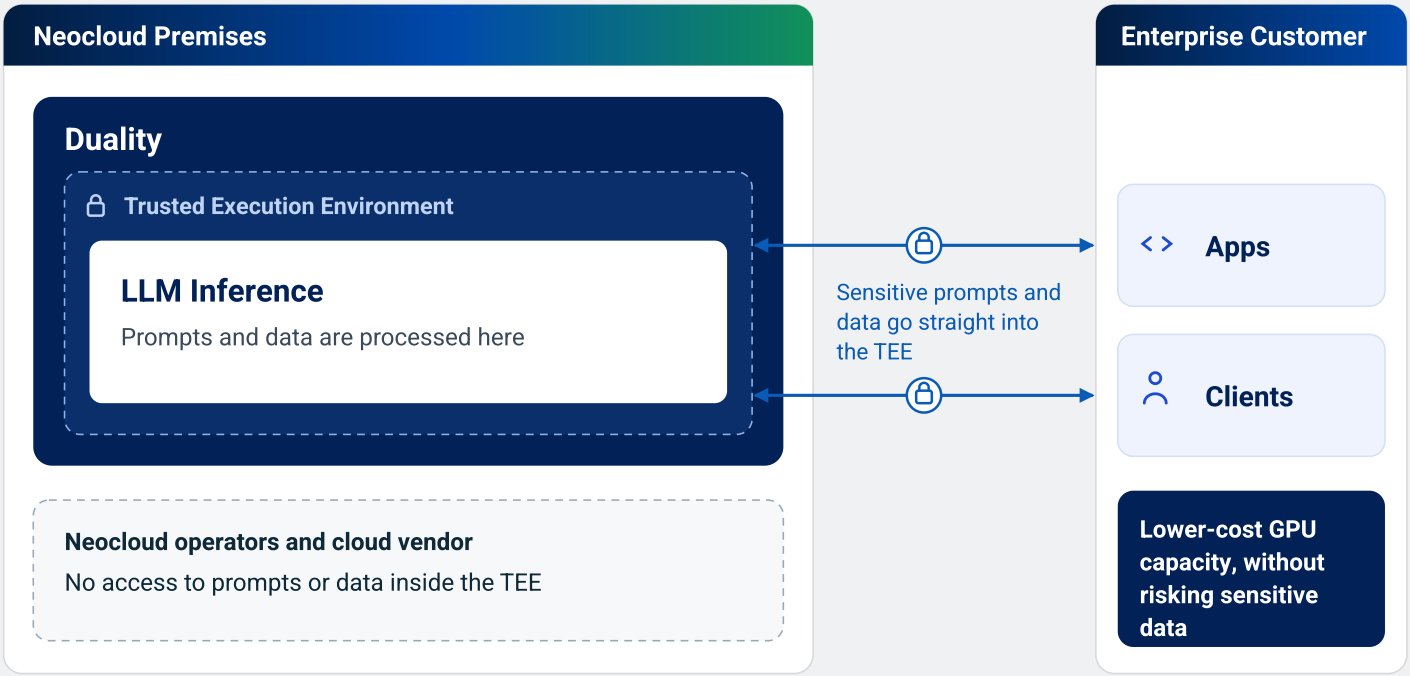
CONTACT US
info@dualitytech.com

LINKEDIN
Duality Technologies

WEBSITE
dualitytech.com

CONFIDENTIAL LLM ON NEOCLOUD

How It Works: A Hosted Model That Does Not Expose Your Data



1 Host the TEE at the Neocloud

The Neocloud runs a confidential-GPU TEE with Duality at its own premises, in any region.

2 Verify Before You Send

Your Duality node attests the enclave at start and periodically, and nothing is released until it passes.

3 Connect Straight to the TEE

Your apps and clients connect with mutual-TLS certificates you issue, and the connection ends inside the enclave.

4 Run the Model, Keep the Data

The LLM processes prompts and data inside the TEE, and responses return only to your authorized clients.

HEADQUARTERS
5 Marine View Plaza, Hoboken, NJ 07030

CONTACT US
info@dualitytech.com

LINKEDIN
Duality Technologies

WEBSITE
dualitytech.com

Trust Controls, and Where It Fits

Built-In Trust Controls

	Feature / Benefit	With Duality
01	**Customer-Held Attestation**	You attest the enclave at start and periodically, and your users can verify it in the browser with no Duality software on their side.
02	**Customer-Held Client Trust**	Mutual-TLS client certificates are issued, rotated and revoked by you. The certificate authority key never leaves your node.
03	**Isolated Model Service**	A kernel-level mechanism blocks outbound connections from the application. Only the listed platform channels cross the TEE boundary.
04	**Protected Models**	Run your own sensitive model at the Neocloud while protecting the intellectual property of the model. The model is packaged encrypted, and its key stays in a KMS or Secret Manager you choose.
05	**Governed Release**	Policy decides which applications may be used. Nothing is released to the enclave until it is attested.
06	**Visible Operation**	See lifecycle and trust state, plus CPU, GPU and memory use for each service. Prompts are never recorded.

Where It Fits

Cost-Driven AI

Your AI bill is dominated by GPU capacity, and a Neocloud offers it for less.

Capacity in Another Region

The GPUs you need are at a provider or site you do not control.

Sensitive Prompts

Banks, telecom providers and insurers sending customer or proprietary data to a model.

Deployment

Duality runs the TEE at the Neocloud's premises, including TEEs with GPUs, and a Duality node on your side verifies it. Duality is not a model vendor or a cloud: the model, the Neocloud and your data stay yours.

HEADQUARTERS
5 Marine View Plaza, Hoboken, NJ 07030

CONTACT US
info@dualitytech.com

LINKEDIN
Duality Technologies

WEBSITE
dualitytech.com