

INTRODUCING

UPC AI FACTORY

Run enterprise AI on infrastructure
you actually control.

A **GPU-first, fully managed private cloud** where your data, models, and inference never leave your boundary — at a fixed monthly cost you can plan around.

↓ **42%**

LOWER TCO FOR
AI WORKLOADS

<100ms

P95 INFERENCE
LATENCY

↑ **40%**

FASTER AI
TRAINING

100%

DATA INSIDE YOUR
TENANT BOUNDARY

— WHY AI STALLS IN PRODUCTION

- Prompts, embeddings, and fine-tuned models sitting on shared infrastructure
- GPU, egress, and API charges that make every invoice a surprise
- Spiky inference latency from noisy neighbours on shared GPUs
- No single view of GPU utilisation or cost per million tokens

— WHAT UPC AI FACTORY CHANGES

- **Isolated tenant vDCs** per business line, with private links to your data
- **One fixed monthly charge** — no egress, no metered surprises
- **Dedicated GPUs** with MIG partitioning for training and inference
- **FinOps dashboards** showing cost per million tokens by model



AI Router

OpenAI-compatible API so existing apps point at your private models.



Model Registry

Catalogue, version, and govern every model in one place.



Managed Vector DB

RAG pipelines co-located with GPU and storage for low latency.



GPU Workbench

Ready-made dev environments for data science teams.

12

WEEKS TO
PRODUCTION

WEEKS 1–2

Discovery & assessment

Map workloads and data flows, baseline cost and latency, agree compliance guardrails.

WEEKS 3–8

Parallel deployment

Stand up tenant vDCs and GPU pools, re-host models, validate side by side.

WEEKS 9–12

Cutover & optimise

Shift production traffic, tune scheduling, switch on FinOps and observability.

— ABOUT UNITEDLAYER

UnitedLayer® is a global private cloud provider with over 20 years of managed services delivery, trusted by enterprises and public agencies including the City & County of San Francisco, Blackstone, and Granicus. Our portfolio spans **United Private Cloud®**, **UnitedSecure™**, **UnitedConnect®**, and **UnitedEdge®** — delivered with built-in observability, cloud management, and agentic orchestration, and backed by compliance across SOC 1/2, ISO 27001, HIPAA, PCI, GDPR, FedRAMP, and FINRA. Recognised by ISG as a **Leader in Private/Hybrid Cloud** for four consecutive years.

20+ yrs

MANAGED SERVICES
DELIVERY

99.999%

PLATFORM
AVAILABILITY

100+

IAAS & PAAS SERVICES

ISG 2026

LEADER,
PRIVATE/HYBRID CLOUD

Ready for enterprise AI without compromise?

Book a walkthrough, or ask us to scope a private AI factory for your workloads.

UPC AI FACTORY



EMAIL

contact@unitedlayer.com



PHONE

+1-888-853-7733

UNITEDLAYER.COM