HYPERAI
AI AS A SERVICE

EFFICIENT AI INFERENCE

NVIDIA L4 Server · 24 GB

Energy-efficient Ada Lovelace inference with native FP8 support — right-sized CPU and memory tiers for everything from chatbot serving to video pipelines, on unmetered 10 Gbps in our Tier III+ Amsterdam datacenter.

24 GB GDDR6 VRAM	**FP8** Native Tensor Core	**72 W** Power-Efficient	**Tier III+** NL Datacenter

STARTING FROM

€500 /month

REQUEST A QUOTE

sales@hyperai.ai

TECHNICAL SPECIFICATIONS

Right-sized for production inference

The NVIDIA L4 pairs fourth-generation Tensor Cores with native FP8 support at just 72W per GPU — purpose-built for serving 7B–13B language models, embeddings, video pipelines, and computer vision at low cost per request.

GPU	NVIDIA L4 (Ada Lovelace) — 24 GB GDDR6, FP8 Tensor Core, 72W TDP
Processor	AMD EPYC 9535 (High-frequency Turin architecture)
Storage	NVMe SSD — 500 GB or 1 TB depending on tier, expandable
Network	10 Gbps vPort — unmetered (fair use applies)
Precision	TF32, BF16, FP16, FP8, INT8 — optimized for inference
Location	Tier III+ Datacenter, Amsterdam Region — Netherlands

CONFIGURATIONS & PRICING

Choose your tier

All tiers include the same NVIDIA L4 24 GB GPU. Scale your vCPU cores and system RAM by selecting the tier that fits your inference workload.

Tier	vCPU Cores	System RAM	NVMe Storage	Monthly Price
Standard	8 Cores	48 GB	500 GB	**€500**
Balanced	12 Cores	64 GB	1 TB	**€600**
Performance	16 Cores	96 GB	1 TB	**€750**
Extreme	24 Cores	128 GB	1 TB	**€900**

UPGRADES, NETWORKING & SERVICES

Tailor your deployment

EXTRA NVMe STORAGE

€10 / month per 100 GB

Granular billing — expand storage as needed.

NETWORK & IPs

10 Gbps vPort — Unmetered

1× Public IPv4 included

Free IPv6 /64 on request

Additional IPv4: €2.50 / month

LOCATION

Amsterdam · Netherlands

Tier III+ datacenter with redundant power and cooling.

BACKUP SERVICE

€35 / month — includes 250 GB storage

Retention: 7 daily · 4 weekly · 12 monthly · 5 yearly | Additional backup storage: €6 / month per 100 GB

All pricing excludes VAT.

READY TO DEPLOY?

Contact our team to provision your L4 GPU server.

sales@hyperai.ai · hyperai.ai