

| AI INFRASTRUCTURE

AI Infrastructure That Works Across Every Cloud You Run.

emma's AI capabilities let teams provision GPU compute, connect AI workloads across clouds, and deploy inference environments from governed templates – all from one control plane, without hiring platform engineers for each provider.

[🏠 CIO / CTO](#)[⚙️ PLATFORM ENGINEERS](#)[⚙️ ML / AI TEAMS](#)[📅 FINOPS](#)[🔒 CISO](#)

Table of Contents

THE PROBLEM

01	Why AI Infrastructure Stalls	3
02	Four Infrastructure Bottlenecks	3

THE SOLUTION

03	emma's Take: One Operating Layer	4
04	How emma Solves It: Five Capabilities	4

WHY EMMA

05	emma at a Glance	5
06	Why emma for AI Infrastructure	5
07	What Makes emma Structurally Different	5
08	About emma	8

DISCLAIMER

This solution guide is provided for informational purposes only. All figures, statistics, and third-party references are cited from publicly available sources. emma reserves the right to update platform features and interfaces.

©2026 emma Technologies S.à r.l. All rights reserved.

Why AI Infrastructure Stalls

For many companies, AI adoption stalls because the infrastructure underneath is fragmented, slow, and impossible to operate across providers. That's the gap emma closes.

CORE CHALLENGE

AI infrastructure today isn't one system, it's five disconnected ones.

When you actually try to ship AI in production, GPU availability, data location, latency requirements, and compliance rules all pull in different directions. No single provider has the right capacity, in the right region, at the right price, under the right jurisdiction for every stage of the pipeline.



Four Infrastructure Bottlenecks

This leads to several issues:



GPU Access Is Fragmented

Every cloud provider has its own provisioning workflow – for both VMs and Kubernetes clusters. If your team doesn't have dedicated platform engineering per provider, you're spending weeks on just getting the environments ready.



Workloads Don't Connect Cleanly

Connecting GPU workloads across providers is manual and breaks easily. 59% of organizations report bandwidth issues (Flexential State of AI Infrastructure, 2025) and only 49% say their networks can support AI bandwidth and latency requirements (Broadcom 2026 State of Network Operations).



GPU Infrastructure Runs Unobserved

When GPU metrics are scattered across multiple provider consoles or third-party tools, engineers context-switch constantly. GPU health, utilization, and memory consumption should be visible in the same place everything else is operated.



Inference Runs Ungoverned

Every team builds inference environments from scratch. No shared templates, no audit trail, no consistent access controls. What's running in production and how is often unclear.

One Operating Layer for All of AI Infrastructure

AI teams need portable GPU compute, connected workloads, and governed deployment. Still, AI infrastructure shouldn't be five toolchains stitched together.

emma becomes the operating layer that makes AI infrastructure work as one system across every cloud in the estate.

How emma Solves It

Instead of adding another layer of complexity, emma collapses it. It integrates five capabilities – GPU compute at two levels, integrated observability, cross-cloud networking, and governed inference deployment – under a single governance model that each address the infrastructure bottlenecks blocking AI at scale.



GPU Compute – Multi-Provider

Provision GPU VMs across AWS, Azure, GCP, Nebius, and emma's own cloud from one dashboard. Preconfigured GPU environments come ready with drivers, CUDA, and ML dependencies, so teams can start running workloads immediately, without setup or compatibility issues.



GPU Managed Kubernetes (mk8s)

Fully managed GPU Kubernetes clusters across AWS, Azure, and GCP, provisioned in minutes. Clusters launch with pre-validated NVIDIA drivers and CUDA images. No cluster lifecycle management, no driver conflicts.



GPU Observability – VMs & Clusters

GPU metrics built directly into emma at two levels. For GPU VMs: utilization, vRAM usage, and vRAM utilization visible in the same interface as CPU, memory, and network. For mk8s clusters: nine GPU metrics in the monitoring tab – utilization, memory, power draw, temperature, and clock speeds. Zero additional tooling, no agents, no separate dashboards.



Cross-Cloud AI Networking

On-demand, high-bandwidth connectivity between AI workloads across providers through emma's private networking backbone. Private networking is built into the platform, which makes egress costs predictable.



Governed Inference Workflows

Platform teams can publish reusable templates for inference environments. Development teams can use those templates to spin up production-ready environments in minutes, with RBAC and instance limits.



Unified GPU Cost Visibility

All teams get one view for cross-provider GPU cost attribution at the VM and cluster level. FinOps teams can see every GPU dollar across every provider before spending gets out of hand.

emma at a Glance



5 GPU providers

under one provisioning workflow – no per-cloud rebuild



3 hyperscalers

for managed GPU Kubernetes – AWS, Azure, GCP



1 control plane

for GPU VMs, mK8s, observability, governance, and networking



Certifications

ISO 27001 & SOC 2 certified

Why emma for AI Infrastructure

No mid-market-accessible platform today combines GPU compute provisioning across multiple providers, managed GPU Kubernetes across hyperscalers, integrated GPU observability at both VM and cluster level, cross-cloud networking for AI workloads, and templated inference deployment under a single governance platform with unified cost visibility.

What Makes emma Structurally Different

DIFFERENTIATOR	WHAT IT MEANS FOR YOUR TEAM
One dashboard, five providers	GPU compute provisioned across AWS, Azure, GCP, Nebius, and emma cloud without rebuilding per-cloud workflows.
Managed GPU Kubernetes across hyperscalers	GPU K8s clusters across AWS, Azure, and GCP – governed, pre-validated, production-ready in minutes.
GPU observability, built-in	VM-level and cluster-level GPU metrics within the emma interface. No agents, no external tools, no context switching.
Private networking backbone	AI workloads connected across clouds through emma's own infrastructure, not the public internet.
Governance at the point of deployment	Every GPU VM, every inference deployment inherits RBAC, and audit trails – automatically.
Sovereignty-ready by design	Luxembourg HQ and EU cloud providers supported.
No commercial bias	emma is vendor-neutral across all supported providers; no preference for any hyperscaler.



BOOK A DEMO

Run AI Anywhere and Everywhere.

See GPU provisioning, cross-cloud networking, and inference deployment from a single control plane – in a live 30-minute demo.

 GPU Provisioning

 Cross-Cloud Networking

 Inference Deployment

 BOOK A DEMO →

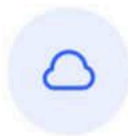
 LEARN MORE →

About emma

The cloud operations platform for distributed infrastructure

Founded in 2021 in Luxembourg, emma is built to help organizations operate data-intensive cloud operations across hybrid and multi-cloud environments more efficiently.

The emma platform provides a unified operating layer across hyperscalers, regional European cloud providers, AI-optimized infrastructure, and on-prem environments, enabling deployment, governance, and cost control from a single interface. With emma, organizations simplify cloud operations, reduce vendor lock-in, and ensure compliance with sovereignty and other regulatory requirements.



Multi-Cloud & Unified Operations

Managing workloads distributed across public, private, and hybrid clouds in a consistent way, providing unified visibility and control.



Cost & Financial Management (FinOps)

Tracking usage and spend, optimizing costs, and budgeting across multi-cloud environments – aligning cloud spending with business value.



Governance & Risk Management

Defining and enforcing policies, guardrails, and controls to manage risk, compliance, and business alignment in cloud environments.



Automation & Lifecycle Management

Automating provisioning, scaling, deployments, and operations to increase agility and reduce manual effort.



Security & Compliance

Ensuring secure configurations, access controls, threat detection, posture management, and compliance with standards and regulations.



Service & Performance Management

Monitoring performance of applications and cloud services, ensuring SLAs, reliability, and resource efficiency.



2021

Founded & HQ in Luxembourg



17+

Cloud providers supported



~90

Cloud engineers at your service



<4h

Response SLA

CERTIFICATIONS & FRAMEWORKS

emma operates under internationally recognized security and compliance frameworks, including ISO-certified security management and SOC 2 Type II audited controls, with data protection aligned to GDPR and resilience aligned with NIS2 and DORA.



NEXT STEPS

Contact us

🌐 emma.ms

✉ info@emma.ms

📍 19-21 Rte d'Arlon, 8009 Strassen, Luxembourg

©2025 emma Technologies S.à r.l. All rights reserved.

This publication is for informational purposes only. All third-party trademarks are the property of their respective owners.