CASE STUDY

Mistral AI unlocks 2.5x faster training speeds

MISTRAL AI_

INDUSTRY	HEADQUARTERS	USE CASES
Artificial intelligence	Paris, France	Compute, SUNK

AI INFRASTRUCTURE AT SCALE

Challenge

Since launching in 2023, Mistral AI has taken the AI scene by storm, proving that the next generation of groundbreaking AI innovation doesn't have to be led from Silicon Valley. In an era long dominated by North American hyperscalers, tech giants, and well-funded AI labs, Mistral AI has emerged as a bold new entrepreneurial force and a major player in today's AI revolution.

By releasing powerful, open-source large language models (LLMs) at unprecedented speed, Mistral AI has not only made waves in the developer community but also redefined what's possible for lean, fast-moving startups. Its mission is bold and clear: to deliver LLMs that empower greater access, more ownership, and better flexibility. That empowers labs and developers alike with models that are reliable, adaptable, and easy to integrate.

But that ambitious vision brings intense pressure. Mistral AI knew it needed to move faster than the market and scale without compromise. They knew they needed to grow at a pace that demanded a **highly specialized, deeply collaborative infrastructure partner** from day one in order to effectively tackle the following challenges:

— **Limited time and resources:** To stay competitive as a young startup, Mistral AI needed to get its models to market fast. That meant its developers needed to stay focused on developing exceptional models, not firefighting infrastructure challenges.

— **Reliability and resiliency:** For streamlined organizations like Mistra AI, saving on costs is absolutely essential. Mistral AI's teams required a partner that could ensure their clusters stayed healthy—and when there was an issue, that it could be rapidly resolved to minimize potential downtime and any impact on their bottom line.

— **Rapid growth at scale:** Growth is a major goal of any startup, but with rapid growth comes the need to quickly adjust to exponential increases in scale. Mistral AI needed a partner that could match its acceleration without delay or bottlenecks.

> "CoreWeave is one of the few providers that has real experience at very large scale for exactly what we do, so large language model training."
>
> **Timothée Lacroix**
> *CTO, Mistral AI*

Solution

AI WITHOUT COMPROMISE

> “[Our models] were trained 100% on CoreWeave infrastructure. Not being on that infrastructure would’ve delayed us by at least a few months.”
>
> **Timothée Lacroix**
> *CTO, Mistral AI*

Mistral AI chose to partner with CoreWeave for a reason that is as bold and focused as their own mission: **CoreWeave is purpose-built for AI**. Every layer of our platform—from data center architecture to application services—is meticulously engineered to deliver the highest levels of performance, efficiency, and scalability for modern AI workloads.

Unlike traditional cloud platforms that stretch to meet a wide range of use cases, the CoreWeave AI Cloud platform is fine-tuned for AI acceleration. For Mistral AI, this meant they had access to a partner that could support their demand for speed, was flexible enough to meet their aspirations, and was capable and experienced enough to let innovation soar without compromise.

That’s exactly why Mistral AI has consistently chosen to partner with CoreWeave since signing their first contract for H100s in 2023, standing up H200 and GB200 clusters over the last two years that have exceeded their highest expectations and gotten their AI models to market at a lightning-fast pace.

Mistral AI was able to tackle its critical challenges and accelerate innovation by using the following solutions:

01 **Slurm on Kubernetes (SUNK):** Our SUNK solution facilitates greater workload fungibility and resource sharing between training and inference workloads by running both use cases on the same cluster. With SUNK, Mistral AI’s engineering teams could collaborate without compromise, improving workflows and reducing bottlenecks across training and inference.

02 **Observability with ultra-granular metrics:** Our observability platform provides detailed metrics and data for Mistral AI clusters out-of-the box, with no additional setup or extra charges required. Mistral AI can now easily visualize entire fleets of GPUs in one place, complete with full analytical detail about every node. Plus, Mistral AI can see real-time ingress and egress traffic throughput from each node in a cluster to external Internet endpoints, such as external model weight data sources, to identify under-optimized workloads.

03 **Automated health checks and 24/7 support:** CoreWeave Mission Control continuously monitors the health of Mistral AI’s nodes, ensuring job interruptions and issues are promptly flagged and resolved as quickly as possible. Plus, our FleetOps, CloudOps, and Data Center Technician teams all work in tandem 24/7 to monitor for signs of deterioration across our fleets and our AI Cloud platform environments.

04 **NVIDIA GB200 NVL72:** Mistral AI was among the first to gain access to NVIDIA GB200 NVL72 racks. This opportunity unlocked an unprecedented level of computing power their teams could leverage to run training and inference jobs—accelerating both overall productivity and time-to-market. With the support of CoreWeave infrastructure, Mistral AI was able to train models 2.5x faster than previously experienced on NVIDIA H200s—and also vastly exceeded the performance, efficiency, and speed they experienced on NVIDIA H100s.

UNLEASH AI'S POTENTIAL

2.5x faster training
On NVIDIA GB200s

100% CoreWeave infra
Trained on CoreWeave Cloud

24/7 service
Dedicated engineering support

> **"What we don't see with CoreWeave are all of the many interruptions or issues that we would see elsewhere."**
>
> **Timothée Lacroix**
> *CTO, Mistral AI*

Results

Mistral AI's meteoric rise isn't just based on their ability to move fast and rapidly scale. It's also about their ability to strategically choose the right partner—one who is ready and able to grow and scale with their constantly growing needs. Mistral AI's partnership with CoreWeave allows their teams to focus more exclusively on **model innovation**, without burning time or resources on inevitable AI infrastructure challenges. That intense focus, combined with best-in-class compute, helped propel Mistral AI into the global spotlight with game-changing results:

Less maintenance, more focus:

CoreWeave's strong foundation and expert knowledge of AI infrastructure swiftly takes care of maintenance tasks and massively reduces interruptions, giving Mistral AI's teams the ability to focus their efforts on AI innovation. It also allows them to engage in faster experimentation, reduce the time required for training cycles, and achieve faster time-to-market.

Cost-efficient scaling:

CoreWeave infrastructure enables efficiency that unlocks faster speeds for Mistral AI, resulting in higher productivity and critical cost savings without sacrificing their ability to rapidly innovate at scale or introducing unnecessary strain or burnout to their teams.

A long-lasting, strategic partnership:

CoreWeave and Mistral AI's partnership goes above and beyond simply providing plug-in-and-play infrastructure. Mistral AI benefits from a hands-on relationship with CoreWeave expert engineers determined to provide 24/7 support. In return, CoreWeave benefits from continually developing and adapting its infrastructure capabilities to the specific needs of an AI partner like Mistral AI, driving further innovation and efficiency.

Take the next step

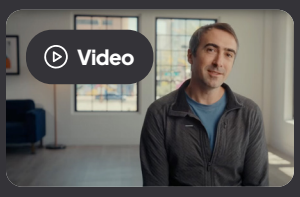
Unleash the full potential of the most ambitious AI innovations.

Contact Us

Powering innovation

See how CoreWeave's AI cloud delivers more.

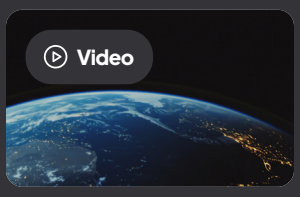
Read more →



Mistral video

Hear directly from the CTO of Mistral AI.

Watch →



IBM boosts performance

IBM acheived greater than 80% faster speed.

Watch now →