

# Validate Production Readiness and TCO with CoreWeave ARENA

Replace assumptions with reproducible performance and cost evidence from a production-scale AI lab

## Evaluate real workloads

Experience CoreWeave ARENA in a hands-on, invite-only program.

Apply now



## When theoretical benchmarks simply aren't enough

CoreWeave AI Ready, Native Applications (CoreWeave ARENA) is a structured, time-bounded evaluation program for pioneers that need to go beyond synthetic benchmarks to understand actual production behavior. By running real workloads on production-grade infrastructure, CoreWeave ARENA produces reproducible performance, operational, and cost evidence.

### What benchmarks miss

- GPU infrastructure operating under real conditions
- AI-native orchestration for distributed multi-node workloads
- End-to-end operational visibility during sustained execution
- High-throughput data movement reflecting real data paths
- Direct access to CoreWeave infrastructure experts
- Consistent, reproducible outputs suitable for decision-making

### The real insights to be gained

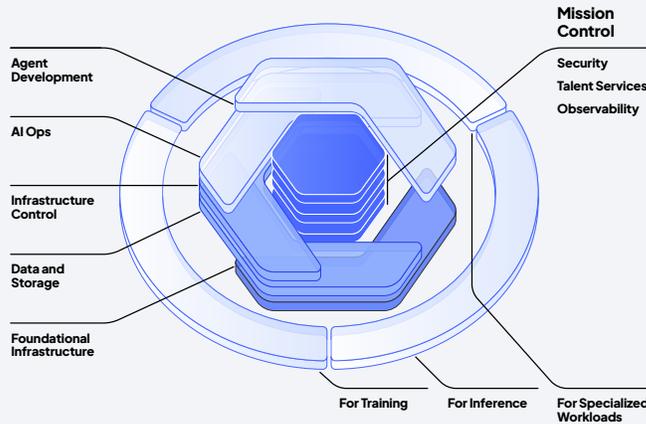
- Will this workload scale reliably beyond initial tests and pilots?
- How do performance and cost behave under sustained, multi-node load?
- What operational signals matter when things slow down or fail?
- Are we ready to move this workload into production?

## PROOF YOU CAN TRUST



### Validation across critical areas

- Reproducible performance and cost results
- Operational behavior captured under real load
- Evidence tied to defined success criteria
- Clear recommendation on next steps



**FIGURE 1**

CoreWeave ARENA evaluates how the layers of the CoreWeave Cloud Platform behave together when running real workloads.

## Understand performance and cost under real load

CoreWeave ARENA evaluates how AI cloud, orchestration, data movement, and operations behave together beneath a real workload—beyond raw benchmark performance. Evaluations run against pre-defined success criteria so your teams can understand performance, cost, and operational readiness before models hit production.

### DESIGNED FOR EVALUATION

### What's tested

**Workloads:**

Training, inference, and RL/agentic workloads

**Orchestration:**

CoreWeave Kubernetes Service (CKS), with SLURM via SUNK where applicable

**Operations and observability:**

Mission Control dashboards and optional Mission Control Agent

**Data movement:**

High-throughput paths between storage and GPUs, accelerated with CoreWeave LOTA™

**Infrastructure:**

Production-grade GPU clusters and high-performance networking

### What's proven

**Operational observability:**

End-to-end visibility into utilization, bottlenecks, communication overhead, and scheduling effects

**Orchestration and scheduling:**

Validation of workload execution on CKS, with SLURM via SUNK supported where applicable

**Data ingest and sustained throughput:**

Testing of realistic data paths and sustained transfer rates from storage to GPUs, including scenarios powered by CoreWeave LOTA™

**Multi-node scaling and networking behavior:**

Confirmation of scaling limits, straggler effects, and inter-node communication patterns under production-like load



**Key outputs:**

- Reproducible performance and cost summary tied to defined success criteria
- Captured operational views from Mission Control during evaluation runs
- Recommendation and next-step plan

BUILT FOR TRANSPARENCY



**Proven at production scale**

- SemiAnalysis ClusterMAX™ 2.0 Platinum rating
- MLPerf training and inference results
  - Up to 96% goodput
  - Up to 20% higher MFU

## Hands-on testing under real conditions

CoreWeave ARENA is a productized evaluation program built on the same infrastructure and operating model used in production—designed to go beyond ad-hoc POCs and synthetic benchmarks.

Capabilities that matter for adoption	CoreWeave ARENA	General-purpose cloud POC	GPU marketplace POC
<b>Real workloads, not synthetic</b> (your models + pipelines)	Built-In	Not standardized	Not standardized
<b>Production-like behavior</b> (multi-node scaling + comms + scheduling effects)	Built-In	Not standardized	Limited
<b>Operational control and diagnostics under real load</b> (Mission Control operating standard)	Built-In	Limited	Limited
<b>End-to-end evaluation</b> (compute + orchestration + data movement)	Built-In	Not standardized	Limited



# How **General Intuition** de-risked world-model training with **CoreWeave ARENA**

General Intuition used CoreWeave ARENA to evaluate world-model training workloads with non-standard scaling behavior before committing to production infrastructure.

## DEEP INSIGHTS, REAL VALUE

“

On other providers, our distributed jobs were failing on a daily basis, sometimes multiple times per day. On CoreWeave, long-running training jobs were far more stable.

**Florian Laurent,**  
*Infra Lead / Member of  
Founding Team*

## Real workloads

Through CoreWeave ARENA, the team ran their actual world-model training pipelines—not simulations—on production-grade GPU clusters, observing distributed training behavior under sustained, multi-node load.

## Operational visibility

Direct access to CoreWeave infrastructure engineers and operational visibility during runs helped the team understand performance, scaling dynamics, and cost behavior.

## Proven at scale

The result was a faster transition from evaluation to production training, significantly higher performance, and reduced infrastructure friction for a small research team without a dedicated infrastructure function.

## Faster path to production

CoreWeave has supported similar evaluation-to-production journeys for teams such as IBM and Mistral, where real workload evidence and operational transparency were critical to scaling confidently.

## Evaluate your workload

Apply to evaluate your AI workloads in CoreWeave ARENA and generate performance, cost, and operational evidence before production.

[Apply to ARENA](#)



### CoreWeave Mission Control™

Explore the unified operating standard teams rely on during CoreWeave ARENA evaluations.

[Read →](#)



### ClusterMax 2.0 ratings

Read an Independent validation of CoreWeave's production-scale AI infrastructure vs competitors.

[Read →](#)



### How to Really Measure TCO for AI Cloud

Get a practical framework for translating evaluation results into infrastructure and budget decisions

[Watch →](#)