

CUSTOMER USE CASE

Groq + Orq.ai: Enabling LLM Ops with Fast AI Inference

groqcloud™

EMERGENCE EMERGENT ADAPTIVE DYNAMIC
PREDICTIVE PRESCRIPTIVE ACCELERATOR AI
DETERMINISM LPU CHIP NEURAL NETWORK
COMPUTE EFFICIENT SCALABLE HIGH-PER
EARNING LANGUAGE MODEL LARGE INTELLI
GENT AUTONOMOUS PREDICTIVE ANALYTIC
SYSTEM ARCHITECTURE DESIGN ENGINEERIN
G BREAKTHROUGH INNOVATION DISRUPTIO
N INFORMATION INTEGRATION PRECISION QUANTI
ZATION TENSOR GRAPH NODE EDGE FLOW STREA
MING THRESHOLD BOUNDARY FRONTIER HORIZO
N VELOCITY AGILITY RESILIENCE ADAP

orq.ai

Orq.ai is the end-to-end platform for serious software teams to control GenAI at scale. Build, ship, and scale LLM applications – all in one place.

The Challenge

Development teams across businesses of all sizes are working hard to tap into the power of the latest Large Language Models (LLMs) and technologies. Their goal: deliver LLM-powered solutions that transform businesses, lower operating costs, and deliver bottom-line value. These may be new AI-native startups, existing software as a service (SaaS) companies, or AI teams within large enterprises. While the goal is ambitious, it does not come without its challenges. It's fairly easy to hook up the latest LLM into a software application. However, teams lack the proper tooling needed to manage the entire software development cycle of LLM-based applications. Coordinating requirements between business and software experts, managing the entire AI project workflow, building resilience into LLM apps, handling security, and monitoring performance are all critical to success. This takes a lot of time and resources to get right. AI teams and their business clients would love to focus on their core differentiators, but unfortunately, a lot of their capacity is sucked up by these important tasks.

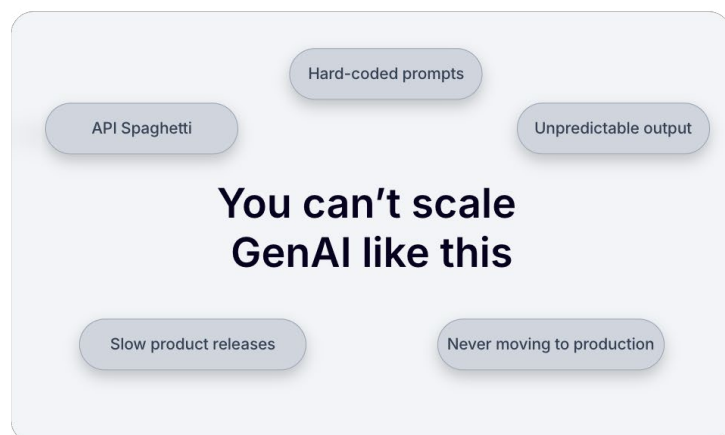
The Solution

Orq.ai has built an end-to-end Generative AI Collaboration Platform to support these AI teams in managing the LLM product development workflow. With Orq.ai's platform, teams build reliable LLM-based products from the ground up, run them at scale, control output in real-time, and optimize performance continuously. Orq.ai's solution is middleware for Generative AI, sitting on top of foundational AI technologies and helping software developers and business experts get a firm grip on core workflows needed to make LLM-based projects a success. The platform provides developers, product teams, and business experts with the flexibility to try different LLMs and prompt configurations, easily update business logic, control app output with guardrails and monitor the performance of their LLM app's performance in real-time – all in one platform that every team member can access.

groq®

Teams struggle to scale GenAI in their products—Orq.ai fixes that.

DevOps ≠ LLMOps



GenAI Lifecycle Management



Orq.ai provides its customers access to a wide range of LLMs running on GroqCloud™. Its customers especially prize Groq-hosted models for their exceptional speed. Orq.ai clients are developing a variety of LLM-based solutions, most of which require fast response time. Groq delivers ultra-low latency for a number of models across modalities, meaning developers using Orq.ai have a choice of leading models, all offering the speed they need to be successful. Groq integrates easily and seamlessly with Orq.ai, delivering rock-solid reliability alongside superior performance.

GroqCloud integrates easily and seamlessly with Orq.ai, delivering rock-solid reliability alongside superior performance.

The Opportunity

GenAI holds the potential to transform virtually every business. Orq.ai, working with Groq, has the opportunity to become a ubiquitous Generative AI operating system whose composable platform helps teams build and deploy GenAI solutions with a high degree of flexibility and control. Combining the advantages of Orq.ai and GroqCloud opens up a world of possibilities for teams looking to harness the power of Generative AI.