

groq

groq

2022-25

MARKETING CAMPAIGNS AND SOCIAL MEDIA

UX/UI

EVENTS AND SPACES

SWAG+

PRODUCTS AND PACKAGING

MARKS

STYLE GUIDE AND DESIGN SYSTEMS

PRESENTATIONS

CONTENT

Jake Louderback

415.301.0213

2jlouderback@gmail.com

jlouderbackcreative.com

linkedin.com/in/jakelouderback/

MARKETING CAMPAIGNS AND SOCIAL MEDIA

F1 PARTNERSHIP CONCEPT



INFERENCE CAMPAIGN 24: A/B TESTING

Inference
processors
for AI.

All your training has
led to this moment.

groq™

LEARN MORE

Inference defines
AI success.

Evolving modern
workload solutions.

groq™

LEARN MORE

INFERENCE

Where production
meets profitability.

groq™

LEARN MORE

groq™

INFERENCE

Where training ends and business begins.

Talk with a Groq Solutions Specialist

AI systems for modern workload solutions.

Inference uses input data to solve real-world challenges, enabling businesses to compete in a world, abundant with data, demanding real-time insights with an accelerated time-to-production. Training was the necessary first step and investment when building your AI strategy. Now, it's time to turn data into profits, operationalizing your production-ready workloads and models to help with real-world and real-time decision making.

It's time for inference.

What's your role?

I DEPLOY AI SYSTEMS.

I AM AN AI DEVELOPER.

I NEED AI INSIGHTS

Software-defined
Hardware

GroqWare™ Suite

ENTANGLEMENT

U.S. ARMY

Learn about our
software-defined hardware.

Learn about our easy-to-use
software suite.

Learn about our customer's inference
performance advantage.

INFERENCE CAMPAIGN 25: A/B TESTING



OPEN FOR BUSINESS

Available Capacity
for Fast AI Inference

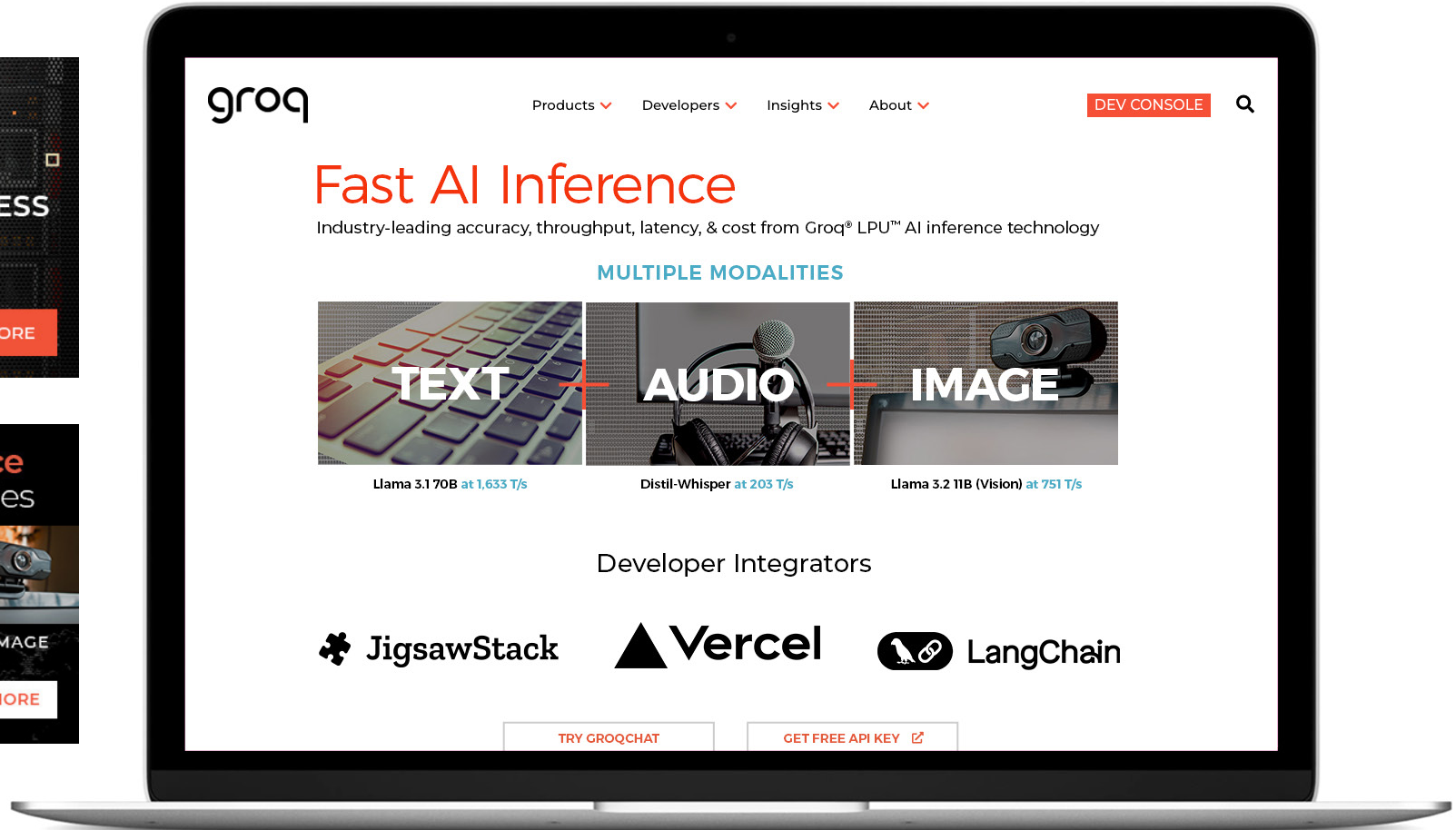
[LEARN MORE](#)

Fast AI Inference
for Multi-Modalities



TEXT + AUDIO + IMAGE

groqcloud™ [LEARN MORE](#)



The laptop screen displays the Groq website. The header includes the Groq logo, navigation links (Products, Developers, Insights, About), a red 'DEV CONSOLE' button, and a search icon. The main heading is 'Fast AI Inference' in red, followed by a subheading about industry-leading accuracy, throughput, latency, and cost. Below this is a section titled 'MULTIPLE MODALITIES' featuring three columns: 'TEXT' with a keyboard image and 'Llama 3.1 70B at 1,633 T/s', 'AUDIO' with a microphone image and 'Distil-Whisper at 203 T/s', and 'IMAGE' with a webcam image and 'Llama 3.2 11B (Vision) at 751 T/s'. The 'Developer Integrators' section lists JigsawStack, Vercel, and LangChain. At the bottom are buttons for 'TRY GROQCHAT' and 'GET FREE API KEY' with an external link icon.

groq

Products ▾ Developers ▾ Insights ▾ About ▾

[DEV CONSOLE](#) 

Fast AI Inference

Industry-leading accuracy, throughput, latency, & cost from Groq® LPU™ AI inference technology

MULTIPLE MODALITIES



TEXT

Llama 3.1 70B at 1,633 T/s



AUDIO

Distil-Whisper at 203 T/s



IMAGE

Llama 3.2 11B (Vision) at 751 T/s

Developer Integrators

 JigsawStack

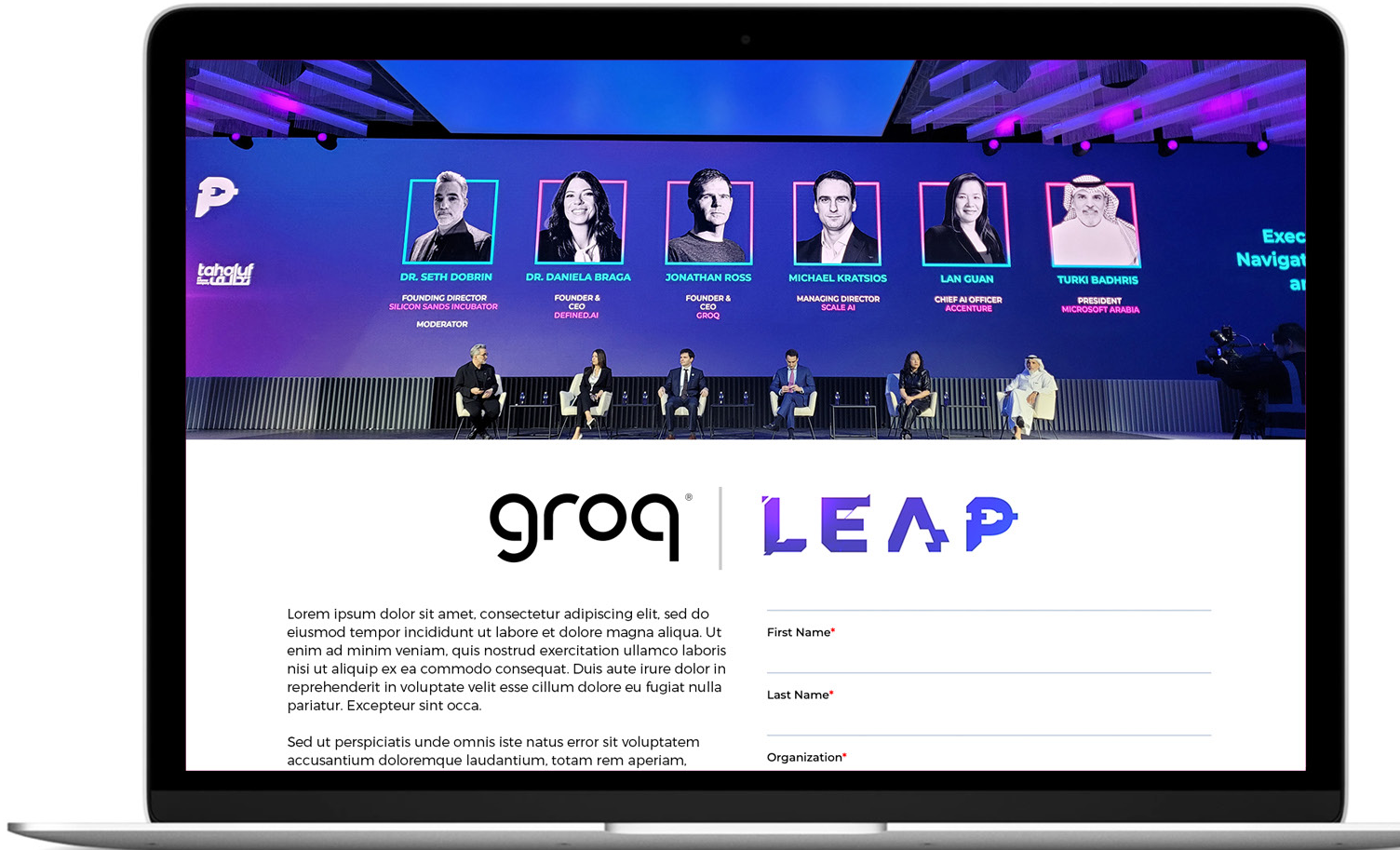
 Vercel

 LangChain

[TRY GROQCHAT](#)

[GET FREE API KEY](#) 

LEAP CAMPAIGN: A/B TESTING



DEVELOPERS: HACKATHONS, DISCORD AND THE PLAYGROUND



groqday

For Developers, by Developers.

Featuring demos of cutting edge software tools for developer velocity.

WATCH VIDEO

VIRTUAL EVENT RECAP



groqday

FREE VIRTUAL EVENT
Tuesday, 10/4 | 1pm PST

An exclusive look at the software suite and developer tools of a leading AI accelerator.

REGISTER



groqday **SAVE THE DATE**

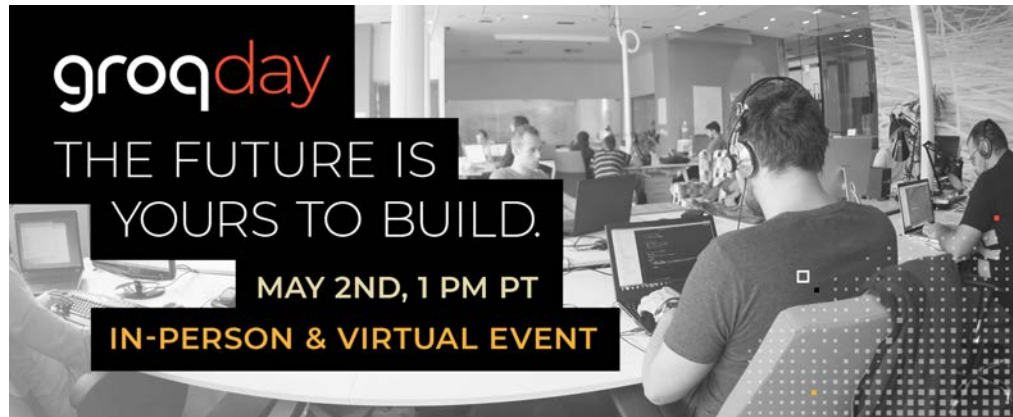
Virtual Event | Tuesday, 10/4 | 1pm PST

See demos of our software tools, designed for developer velocity:

- Groq SDK**
- GroqFlow™ Tool Chain**
- GroqView™ Visualization & Profiler and Groq API**

Register at **groq.link/groqday3**

DEVELOPERS: HACKATHONS, DISCORD AND THE PLAYGROUND



Start Building on Our Developer Console Today

- Increased rate limits during hackathons
- Leading openly-available models across text, audio, and vision modalities
- 1,000+ hackathon projects Powered by Groq
- 1M+ developer community

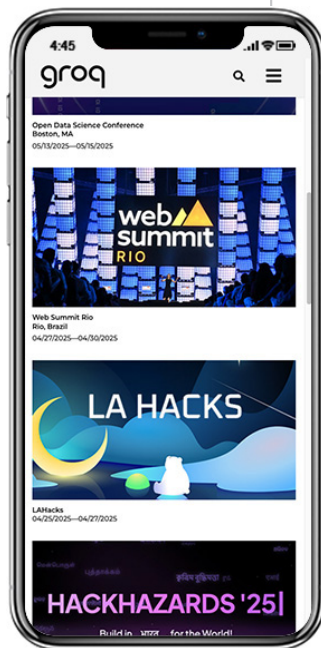


FREE API KEY



JOIN OUR DISCORD

INTEGRATORS



groq®



CONGRATULATIONS

WINNING SUBMISSION BUILT ON GROQ



groqcloud™



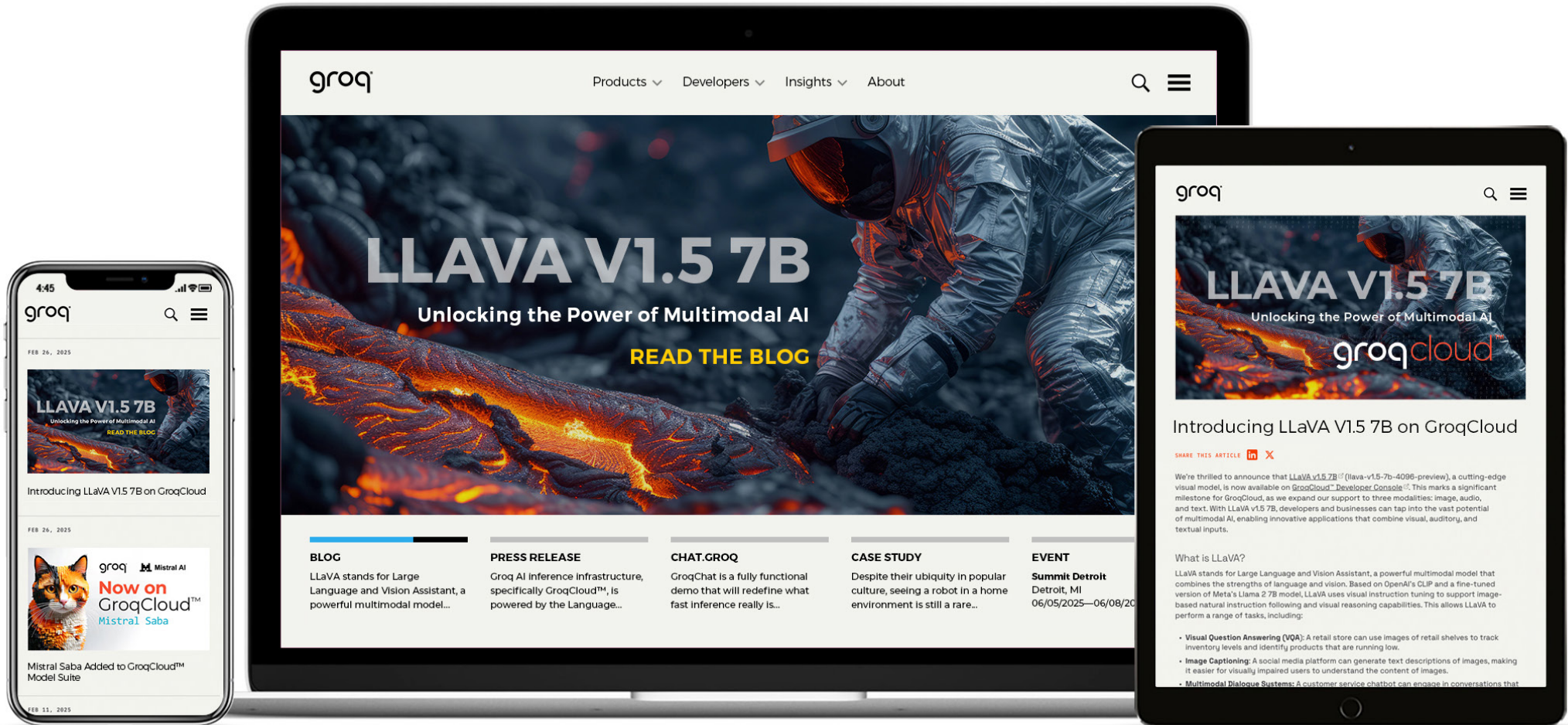
Fastest LLM Performance for AI Applications

- Purpose-built for AI Performance at Scale
- Accelerating & Simplifying Time-to-production
- Domestic Supply Chain & Available to Ship
- Synchronous Hardware & Software Ecosystem

groq

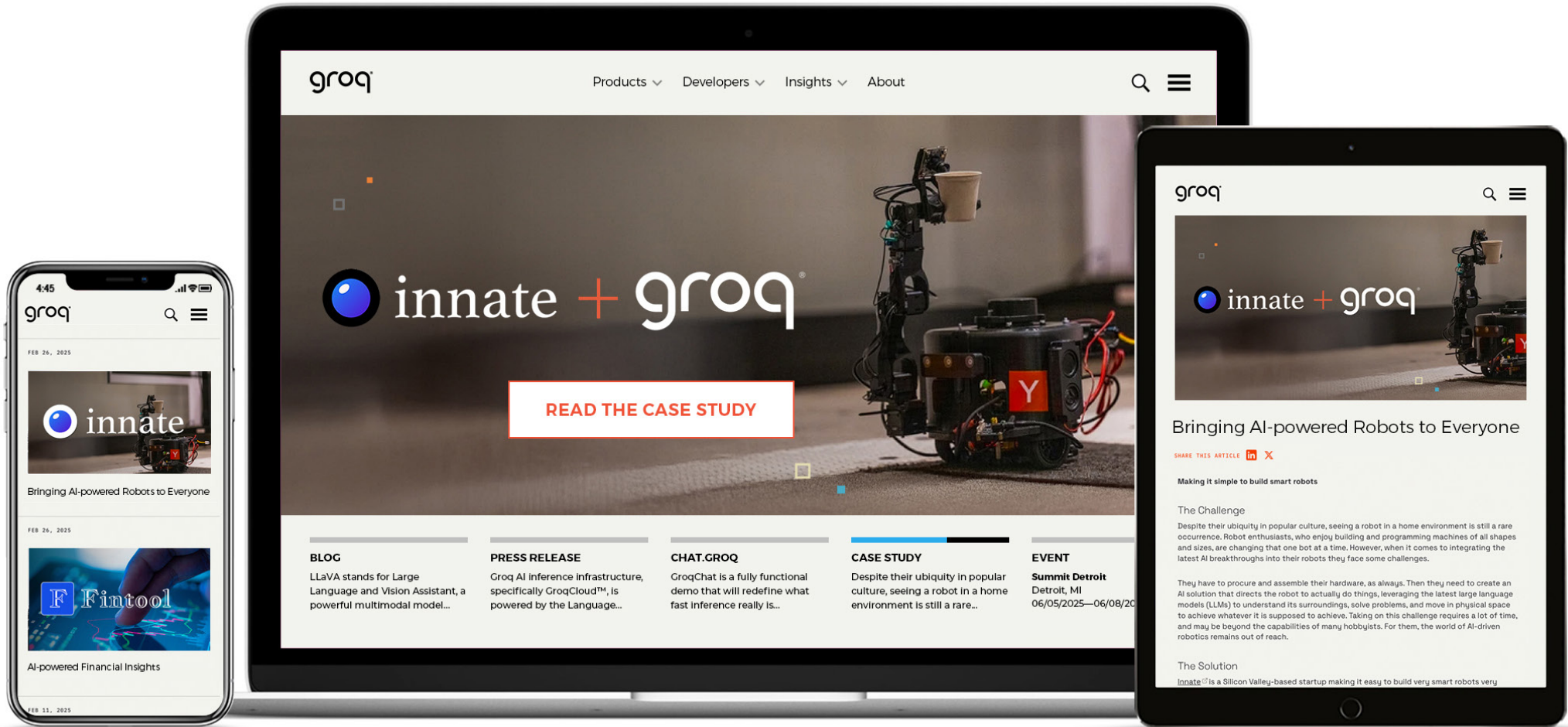
UX/UI

The 2024 site design focused on the theme of **"Making It Real"**. That year Groq achieved record-breaking product performance, revenue and brand recognition—beginning to re-shape the world through the application of AI technologies, tools and services delivered at unprecedented speed. This marked the transition from start-up contender to enterprise dominance.



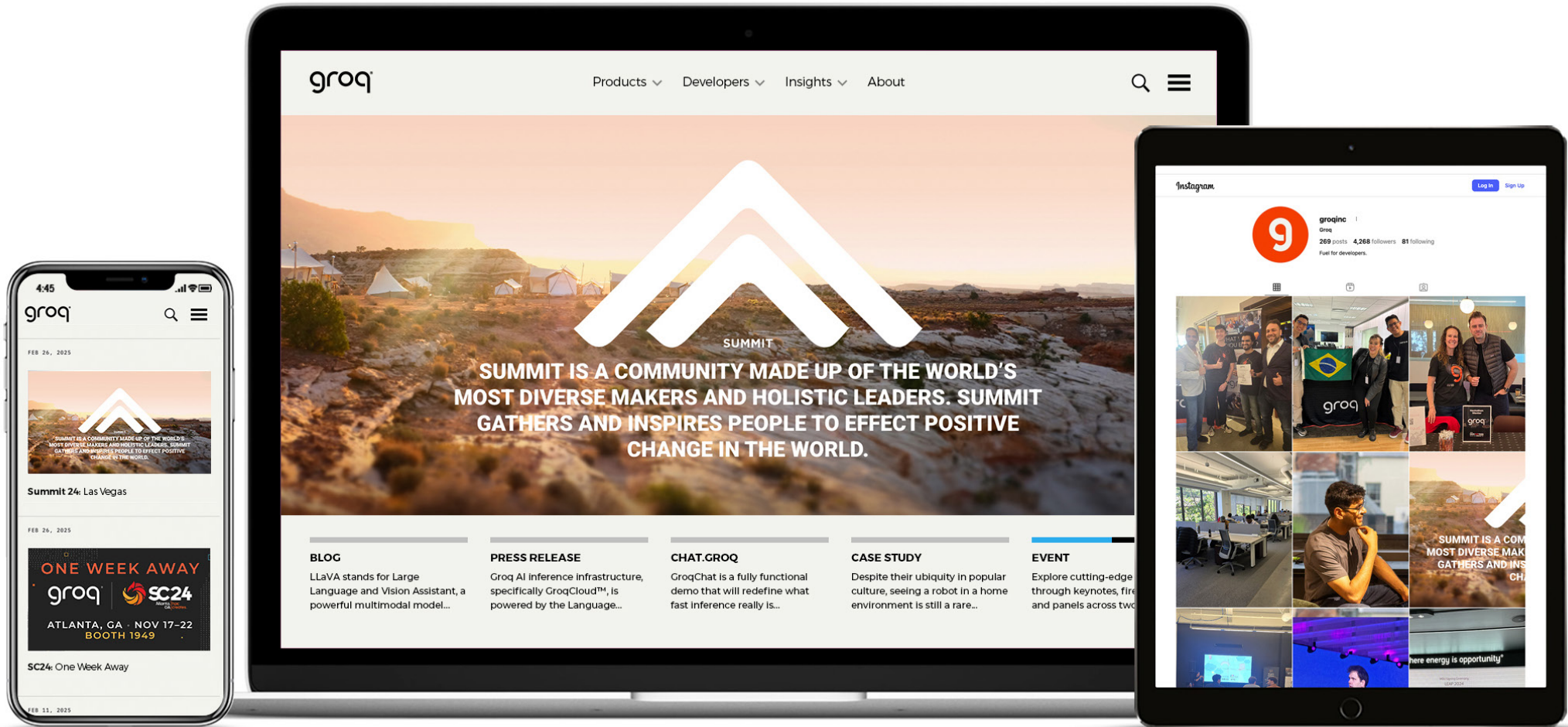
GROQ.COM: HOME > CASE STUDY

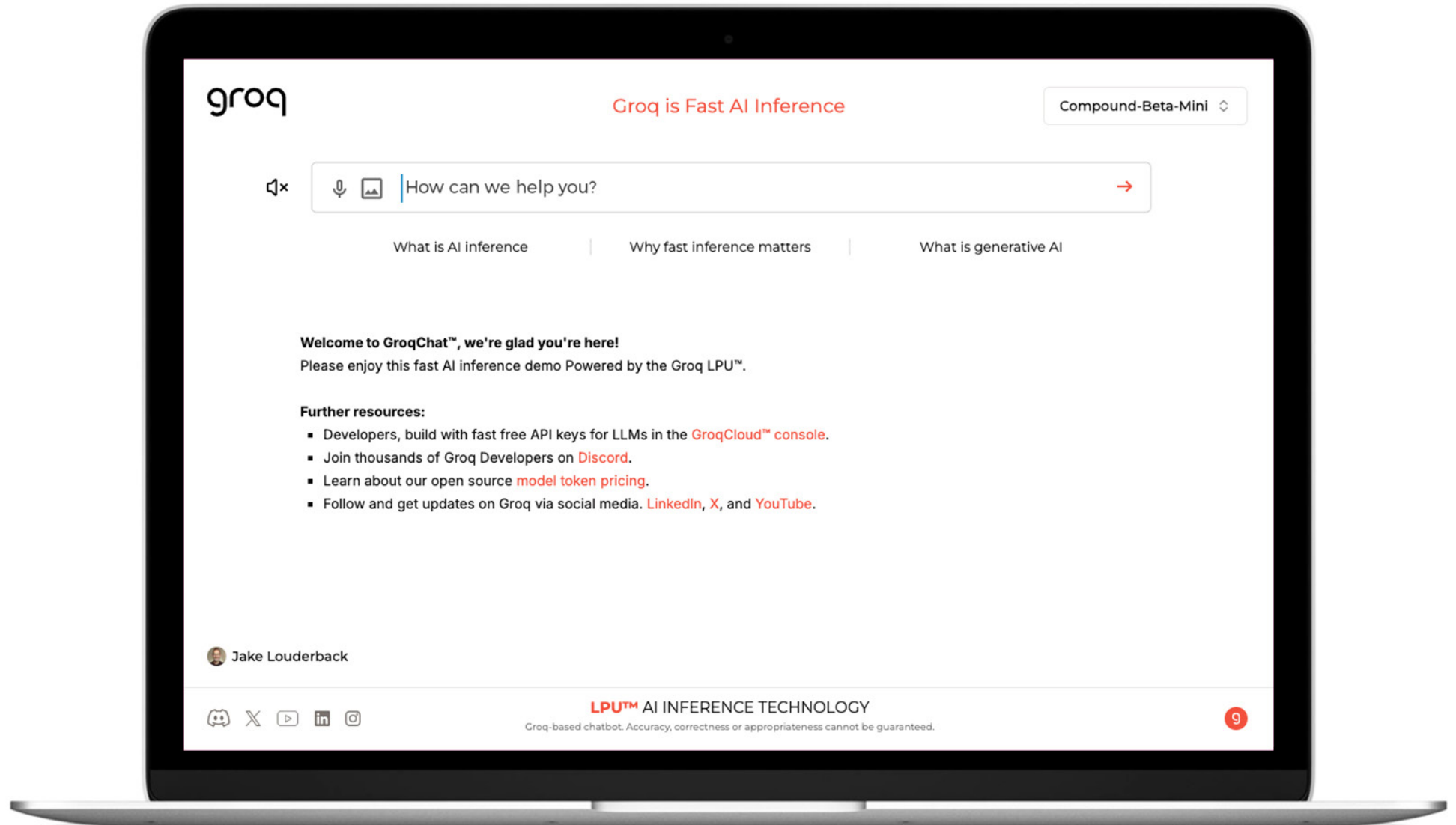
The 2024 site design focused on the theme of **"Making It Real"**. That year Groq achieved record-breaking product performance, revenue and brand recognition—beginning to re-shape the world through the application of AI technologies, tools and services delivered at unprecedented speed. This marked the transition from start-up contender to enterprise dominance.



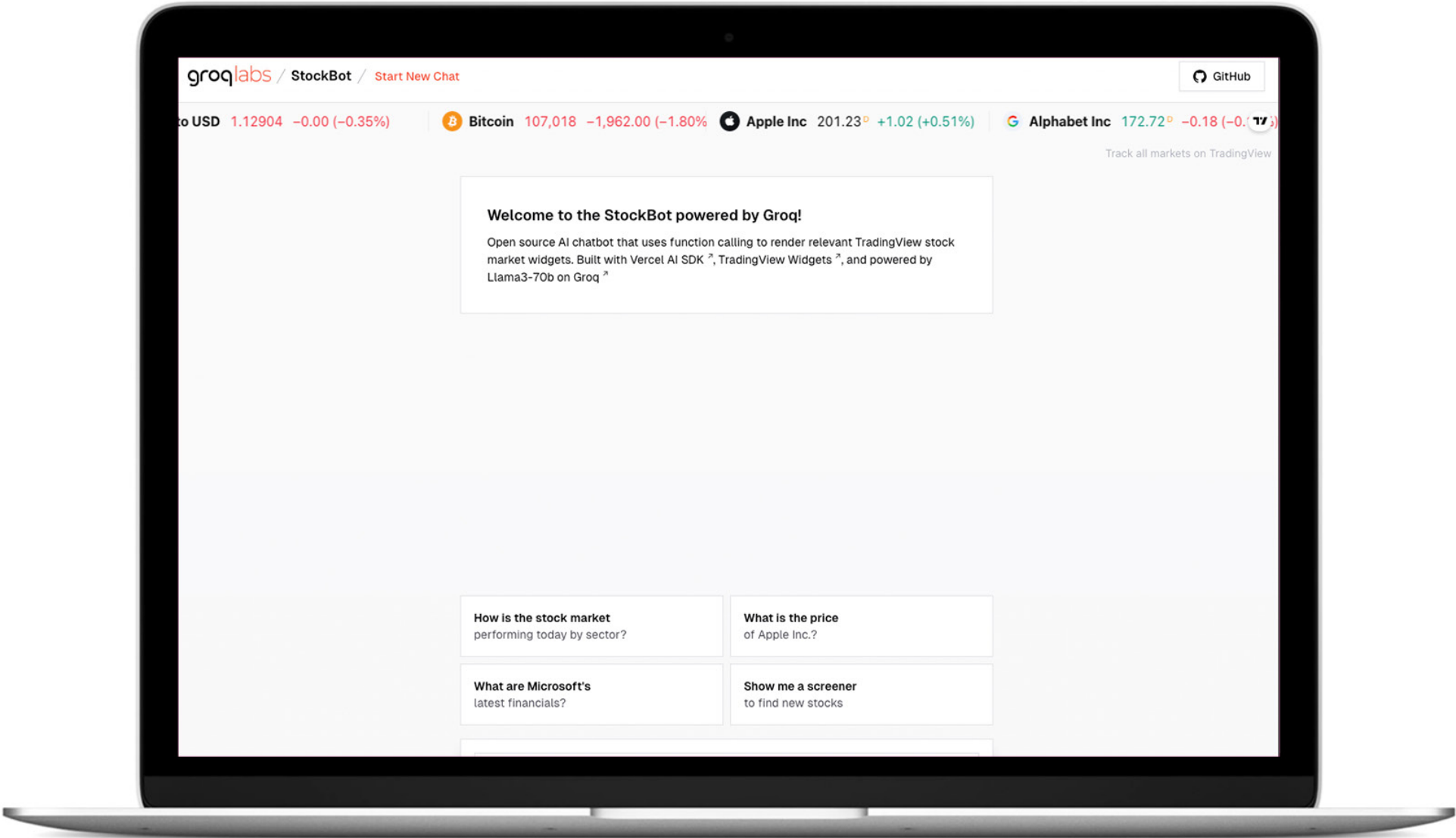
GROQ.COM: HOME > EVENTS

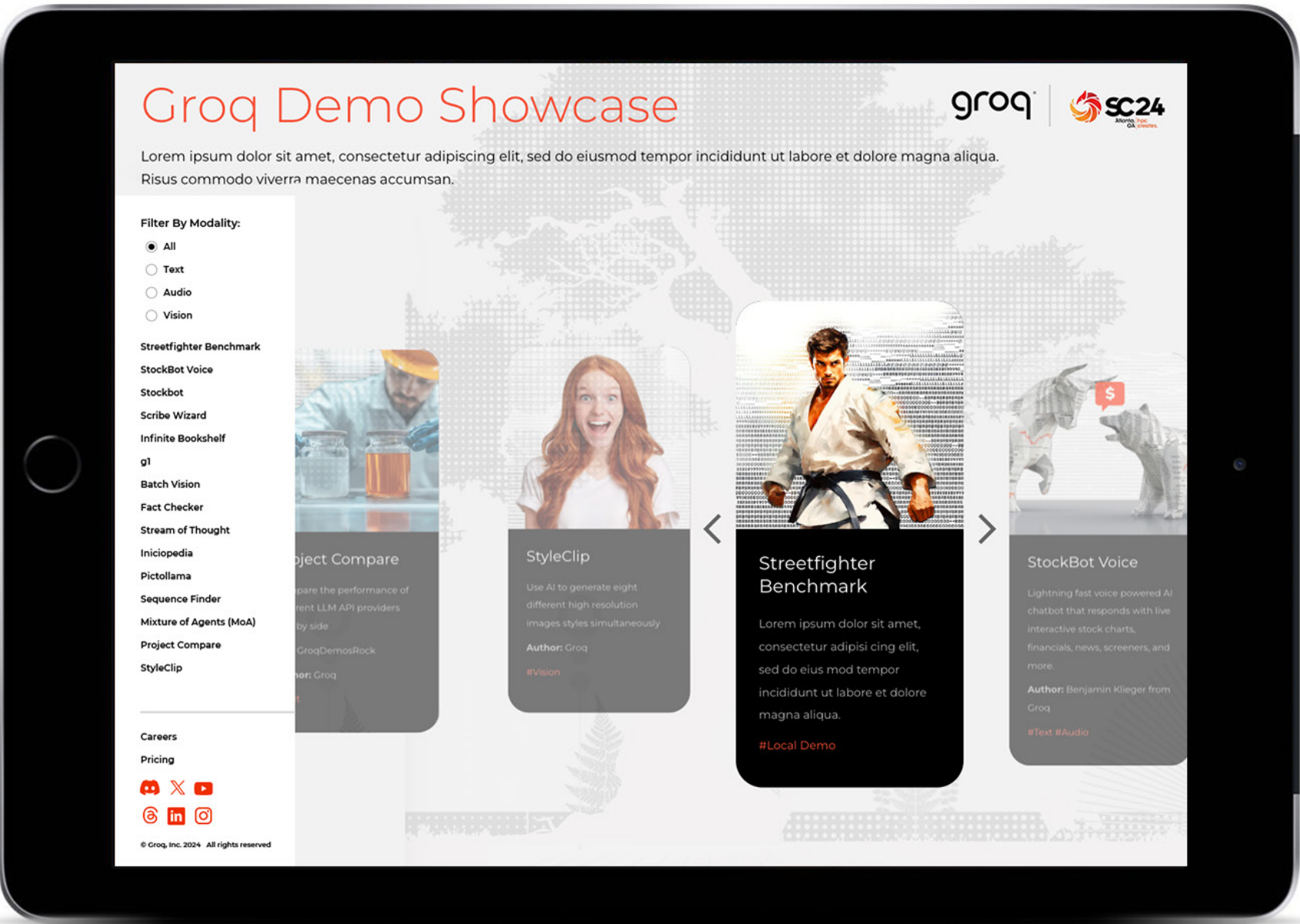
The 2024 site design focused on the theme of **"Making It Real"**. That year Groq achieved record-breaking product performance, revenue and brand recognition—beginning to re-shape the world through the application of AI technologies, tools and services delivered at unprecedented speed. This marked the transition from start-up contender to enterprise dominance.





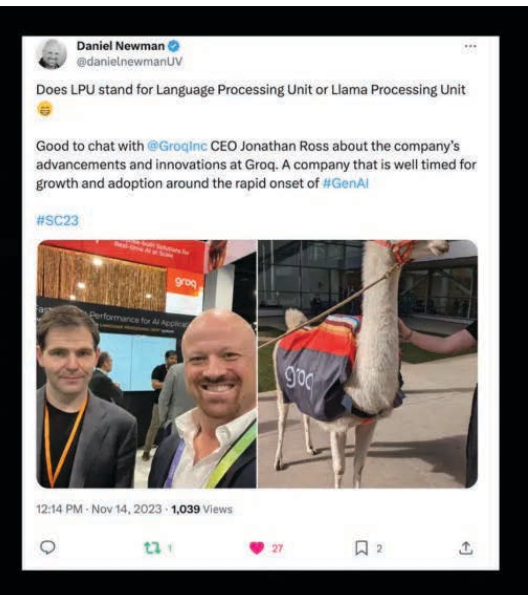
DEMOS: IPAD > STOCK BOT



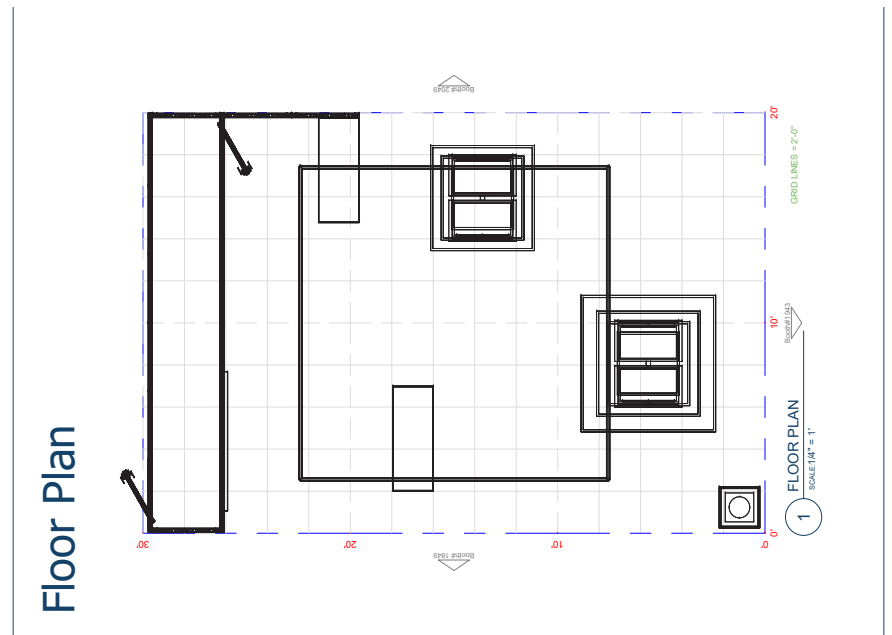


EVENTS, SPACES AND EXPERIENCES

SUPER COMPUTE 22



SUPER COMPUTE 23



PORTABLE WALLS



AI SUMMIT 22

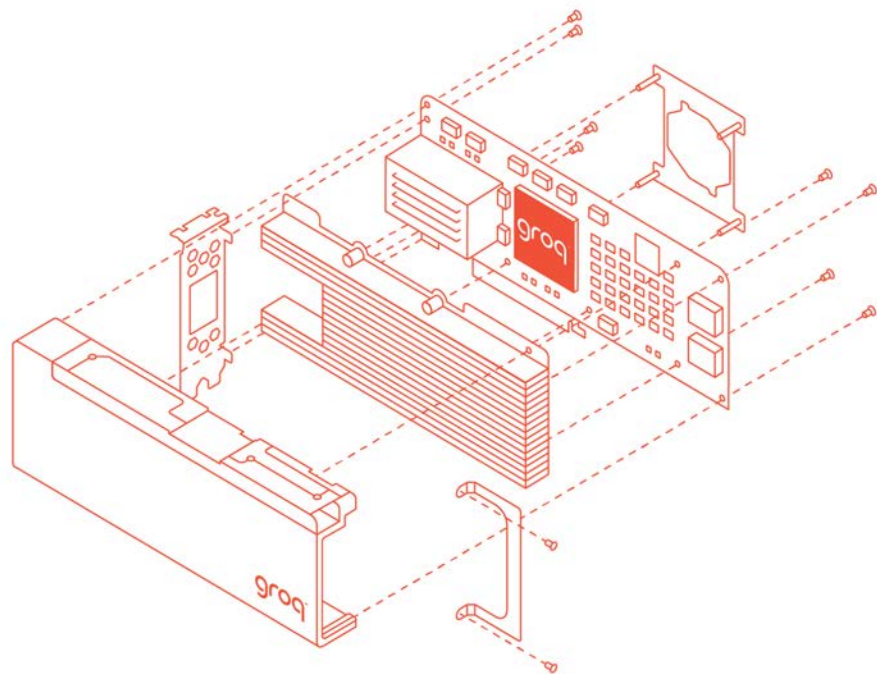


SWAG+



```
curl https://api.groq.com/openai/v1/chat/completions -s \  
-H "Content-Type: application/json" \  
-H "Authorization: Bearer $GROQ_API_KEY" \  
-d '{  
  "model": "deepseek-r1-distill-llama-70b",  
  "messages": [{  
    "role": "user",  
    "content": "Explain why we vibe with fast AI inference"  
  }]  
'
```







Street Fighter Demo_25



PROJECT POLARIS

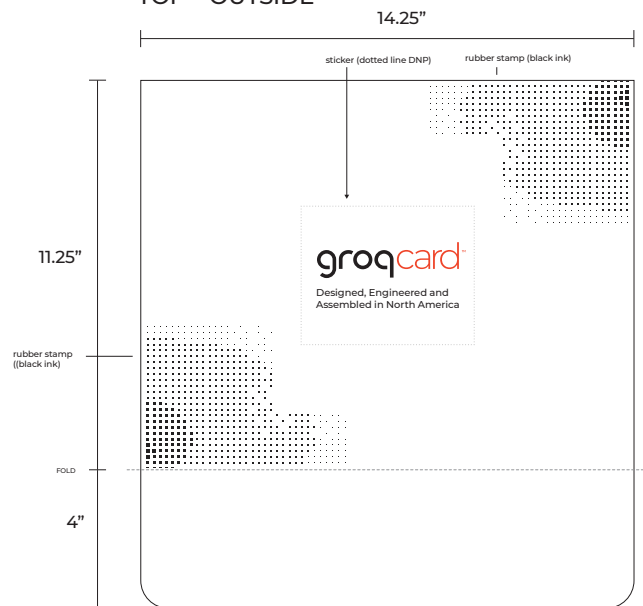
10.2024



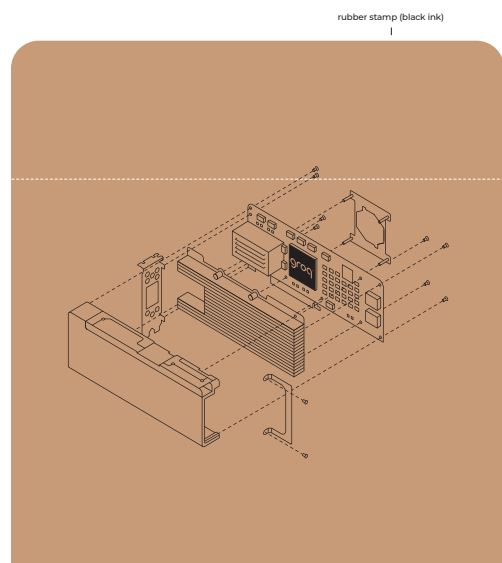


PRODUCTS AND PACKAGING

TOP—OUTSIDE



TOP—INSIDE



RACK BLADE CONCEPTS



MARKS



The Language Processing Unit represented the next leap in Groq's chip evolution—a moment that called for a visual identity as forward-thinking as the technology itself. Our goal was to design a signature mark that felt unmistakably Groq: clean, modern, and built on the foundations of our existing brand system.

The box surrounding "LPU" dovetails seamlessly with the Groq logo, creating a sense of motion as it flows upward and outward in a confident, curving slope. Its geometry echoes the precision of the chip itself, while the overall form strikes a balance between contrast and cohesion—adding a fresh layer of visual continuity to the Groq identity.





Groq Labs is the research and development powerhouse within Groq—where innovation is not just encouraged, but expected. As the engine behind new product breakthroughs, Groq Labs embodies a spirit of curiosity and bold experimentation, constantly asking, “What if...?” and pushing beyond the boundaries of what’s possible. In an industry that moves at lightning speed, this relentless drive keeps Groq ahead of the curve. The Groq Labs logo plays on the dual meaning of “Labs”—both a nod to the laboratory and to the Labrado—symbolizing tenacity, focus, and an unshakable commitment to discovery—like a dog with a bone.

PERSISTENCE, DEDICATION AND DETERMINATION



The Key to Hardware is Software

ML Agility was Groq's software design and development group. They prided themselves on their ability to work quickly and efficiently, creating custom solutions that seamlessly integrated with Groq's rapidly evolving hardware. The growing influence of this innovative team inspired an identity that, while connected to the established brand system, introduced its own distinctive elements and a sense of dynamic creativity.

MLAgil!ty

MLAgil!ty

MLAgil!ty

MLAgil!ty

groq®



Only use in black or white



Don't stretch or distort



Only use logo with ®



Don't apply any effects



No adaptations and additions



No dissections



Fonts

Montserrat

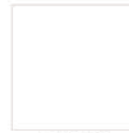
Consolas

Arial

Primary Colors



HEX #F55038



HEX #FFFFFF



HEX #F3F3F3



HEX #000000



PMS 2028C

Secondary Colors



#0A8FC8



HEX #8C1E14



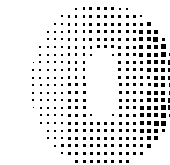
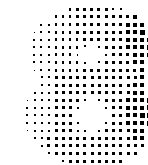
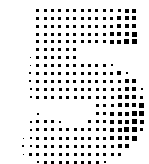
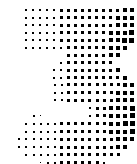
HEX #FA7828



HEX #FAB43C

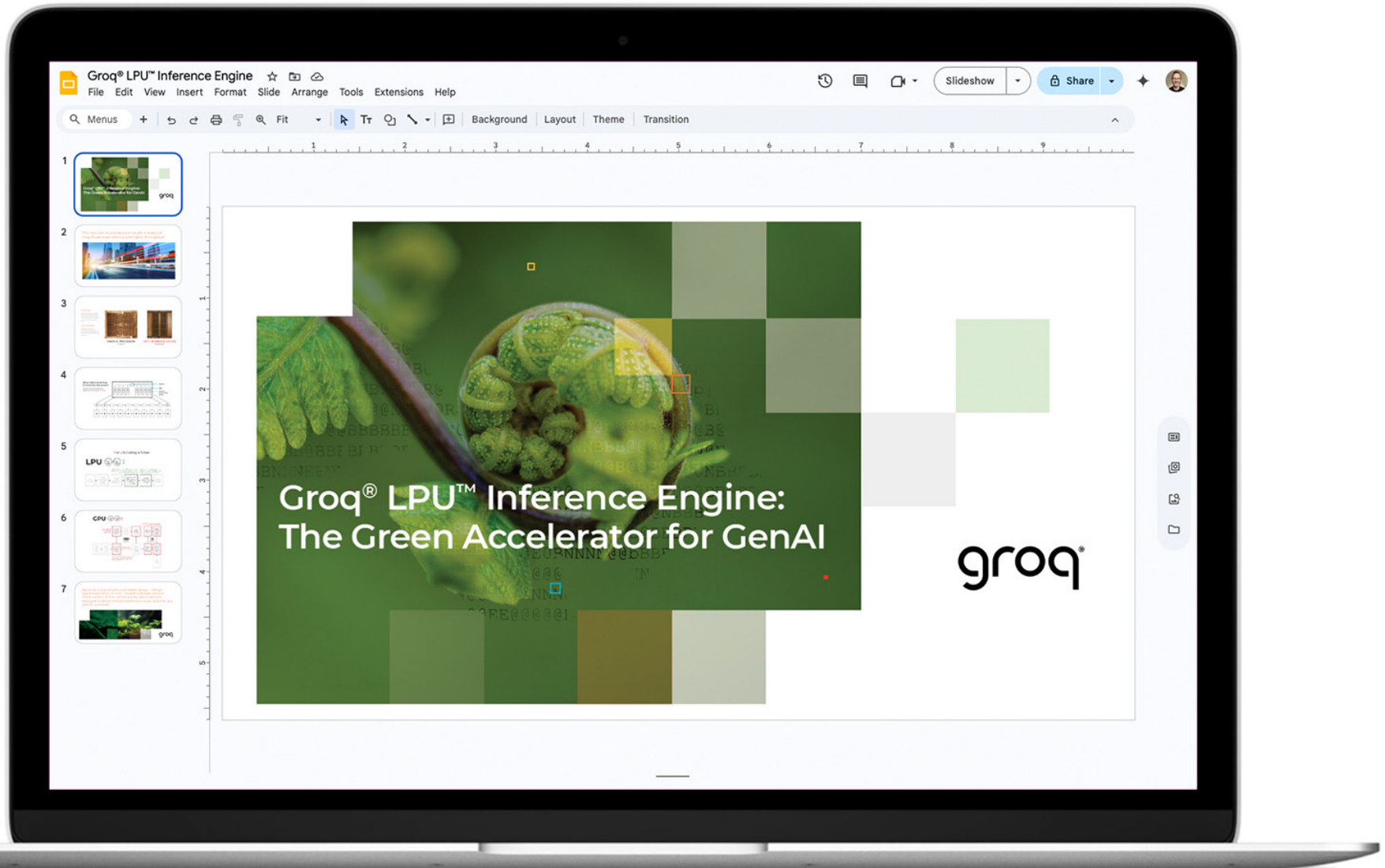


HEX #E6DCAF



PRESENTATIONS

PRESENTATIONS



CONTENT

CONTENT

groq

Products ▾ Developers ▾ Insights ▾ About ▾

DEV CONSOLE



Insights

Customer Use Cases Blog **Papers** Events Press Releases Videos In the News

White Papers



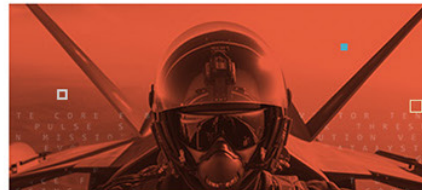
Groq Automatic Speech Recognition (ASR) API



The Future of AI Is Agentic...and Groq



What is a Language Processing Unit?



CONTENT

CUSTOMER USE CASE

Groq + Vetted: Revolutionizing the AI Shopping Assistant Experience



Leverage AI to make smarter, more informed shopping decisions with a powerful, real-time research assistant.

The Challenge

After all these years of e-commerce, researching a product purchase online is still a fairly primitive experience. You ask your favorite search engine about a product or category ("best facial cleanser", "stylish backpacks"), click through on a bunch of results, read reviews which may or may not be trustworthy, then try to figure out which question to ask next. After slogging your way through the product research process you make your purchase, but sometimes wistfully wishing you could just walk into an old fashioned brick and mortar store and talk to a human expert who would listen to you and provide an informed recommendation.

The Solution

Vetted is an AI-powered shopping assistant that walks users step-by-step through their product research and purchase journey. Users interact with Vetted via a natural language chat interface, so they can easily get the best reviews and recommendations to make informed purchase decisions quickly.

Behind the scenes, Vetted sorts through tons of data to get to the right information for any particular shopping journey. It knows where to look for the right information, processing social media threads, comments, and reviews to hone in on what's relevant. It helps users go from broad category research to specific purchase decisions, suggesting additional questions the user should ask while providing actionable recommendations and access to source material every step of the way.

The user experience is conducted in a chat interface, with a UI that's designed with user interactivity in mind. It's like having a lively conversation with that human in the store. This doesn't work unless Vetted's responses are fast, which is why Vetted is Powered by Groq fast AI inference. Low latency is critical in delivering these instant recommendations as this process involves several different models in Vetted's backend system that conduct custom research in real-time for each customer.

©2025 Groq, Inc. All rights reserved.

groq

groq.com | 1

Inference Speed Is the Key To Unleashing AI's

groq

Engagement has always been critical to exponential growth of technology. When tech is slow, we'll use it when we need it but we won't engage with it. Humans are quite impatient; research shows that when websites pages are delayed by 300 - 500 milliseconds (ms), engagement drops by around 20%¹. Conversely, when tech is fast - wow fast - we'll use it over and over again. Speed drives engagement, and engagement drives productivity, collaboration, creativity, and innovation.

The same applies to artificial intelligence (AI) and large language models (LLMs). LLMs have opened up a world of possibilities of how AI can help people across a wide range of fields up their game. Pick any role in any industry, and you'll find humans in the value loop applying their smarts, experience, training, and judgment to get things done. These "humans-in-the-loop" already use computers as tools in their work, but now LLMs present an opportunity to boost performance by orders of magnitude. Regardless of our field, we can all have access to superhuman intellect and insight at human speed. Software development, content creation, customer service, analytics, fraud and crime prevention, health care ... **AI can 10X human impact across them all (a concept we at Groq have named HumanPlus).**

But not if it is slow. You can have the best LLM system in the world, and if it's too slow people won't use it. This is why speed is the top priority for most AI applications.

The problem is, when AI developers finish training their LLMs and turn to deploying and scaling them (aka inference), the only hardware option they have is the good old graphics processor unit (GPU). GPUs are cool for training models, but for inference, they're slowpokes, leading directly to the great-model-that-no-one-uses problem.

Speed is affected by other factors, such as quality and scale. **When determining an inference strategy for a given application, business and technology leaders need to ensure it can achieve the necessary quality and scale while still maintaining a fast enough pace.** In this paper, we look deeper into each of these factors and provide a clear set of questions leaders can pose to their teams and partners to guide them to the best strategy.

LLMs and other generative AI applications have the potential to transform markets and solve big challenges, but only if they are fast enough, which depends on getting inference right. This paper helps you do that.

¹ For example, a 2017 study by Google showed that delays of 300-500ms reduce user engagement by 20%, while a 2020 study by the University of Michigan found the reduction to be 22%.

©2023 Groq, Inc. All rights reserved.

contact@groq.com | groq.com | 1



THANK YOU



Jake Louderback

415.301.0213

2jlouderback@gmail.com

jlouderbackcreative.com

linkedin.com/in/jakelouderback/