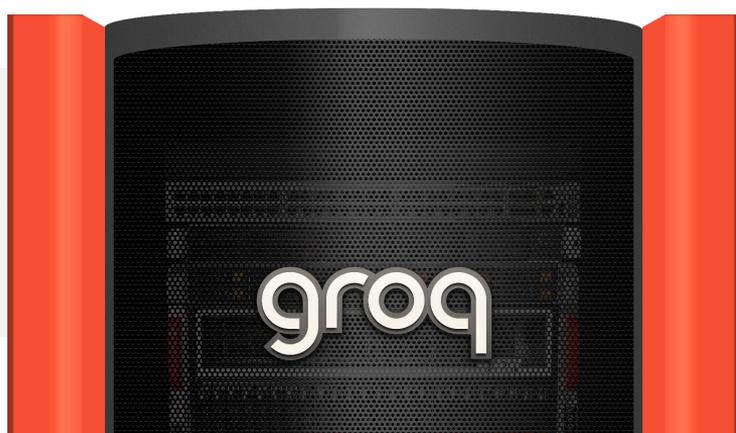


Hello.



The AI
Ecosystem

Who is
Groq?

Show
& Tell

Getting
Started

Let's
Talk

Hello.



The Current AI Ecosystem

Risks & Challenges

- ≡ Migration to deployment
- ≡ Hardware **not optimized** for inference
- ≡ Hardware manufactured **outside of US**
- ≡ High cost of real-time performance **at scale**

Who is Groq?

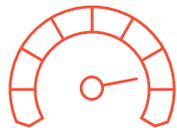
☰ WE ARE

- >>> an AI-born
- >>> software-first platform
- >>> made to run real-time
- >>> AI inference solutions at scale

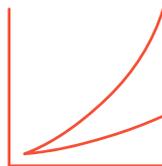


groq™

Why We're Different



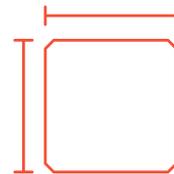
Performance



Pace



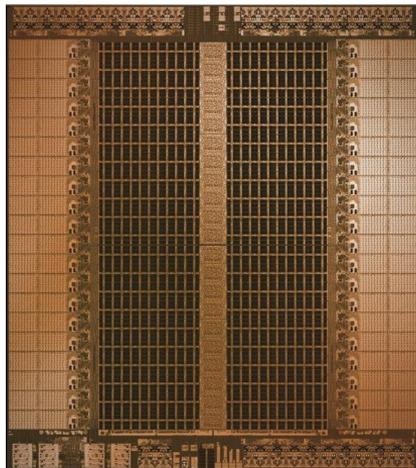
Predictability



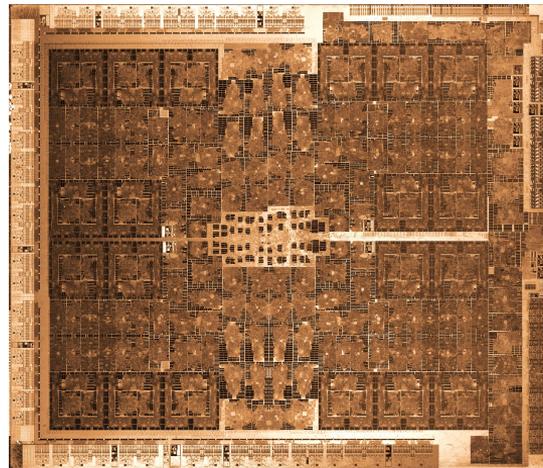
Pinpoint

Simply
Better

Language Processing Unit™
Accelerator
GroqChip™ 1



Typical Graphics
Processor
(GPU)



It's Sort of
Like This



Compile It Fast

Model Name	Model Release Date	Time from Download To Compiling on Groq
Llama 65B	Feb 24th	5 days
Llama 13B	Feb 24th	4 days
Llama 7B	Feb 24th	5 days
Vicuna 13B	April 3rd	0 days (same as Llama 13B)
Vicuna 33B	June 21st	5 days
Llama-2 70B	July 18th	5 days
Code Llama 34B	August 24th	4 days

Run It Faster

July 18th

Model released

Llama-2 70B
released

July 24th

Model compiling
five days after
1st download

10T/s per
user initial
performance

July 29th

Performance five
days after 1st
compile

65T/s per user

Aug 3rd

Performance 10
days after 1st
compile

100T/s per
user

Aug 31st

Performance 38
days after 1st
compile

240T/s per
user

AS OF

Oct 3rd

Performance 72
days after 1st
compile

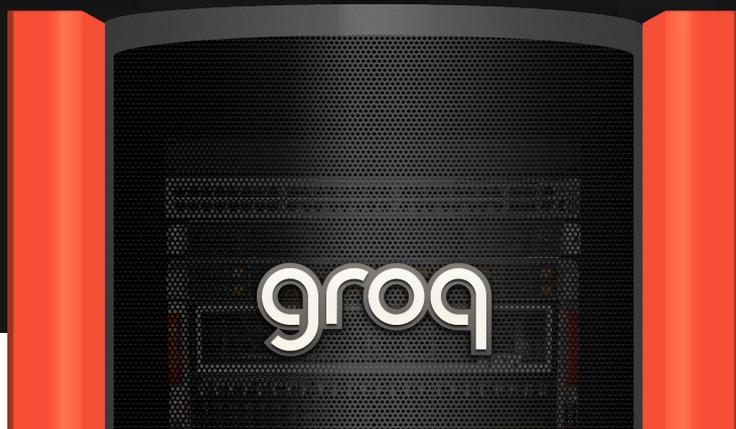
288T/s per
user



WHAT CAN YOU IMAGINE?

SENTIMENT ANALYSIS | CUSTOMER SERVICE | CYBERSECURITY | CONTENT GENERATION |
TRANSLATION AND LOCALIZATION | COMPETITIVE INTELLIGENCE | LEGAL AND COMPLIANCE

Welcome back.



Products: Seamlessly Scalable



GroqChip™ 1



GroqCard™



GroqNode™



GroqRack™

☰ Partnering, **Simply.**

- 1 | Llama-2 workload,
deployed on Groq
- 2 | Analysis to understand
workload needs
- 3 | Deploying solutions

☰ Let's Talk.

What is a Language Processing Unit™ Inference Engine?

What is a kernel-less compiler?

What does Groq make possible?

Assembled in North America

groq™

Thank You!



groq™

