[Groq](#) is fast AI inference.

The LPU™ Inference Engine by Groq is a hardware and software infrastructure that delivers exceptional compute speed, quality, and energy efficiency.

Groq, headquartered in Silicon Valley, provides cloud and on-prem solutions at scale for AI applications. The LPU and related systems are designed, fabricated, and assembled in North America.

During the hackathon, Groq is excited to offer access to our API. This will allow you to leverage a diverse array of AI models, including Llama3-70b, Mixtral 8x7b, Gemma2, Whisper, and many others!

To get started:

1. Visit https://console.groq.com to create a free account.
2. Generate an API key at https://console.groq.com/keys.
3. Begin Building!
4. Here are links to the [documentation](#) and the [Groq API Cookbook](#).
5. If you have questions, feel free to ask a Groqster in-person at the hackathon or post any questions in the [tool-support-groq](#) hackathon channel.
6. Fill out the Groq Rate Limit Increase Form in the [tool-support-groq](#) hackathon channel if you would like increased rate limits for the duration of the hackathon.