

Groq und HUMAIN bieten neue offene Modelle von OpenAI vom ersten Tag an

Palo Alto, Kalifornien und Riad, Saudi-Arabien (ots/PRNewswire) -

Weltweit verfügbar mit Echtzeitleistung, niedrigen Kosten und lokalem Support in Saudi-Arabien

Groq, der Pionier der schnellen Inferenz, und HUMAIN, ein PIF-Unternehmen und Saudi-Arabien führender KI-Dienstleister, gaben heute die sofortige Verfügbarkeit der beiden offenen Modelle von OpenAI auf GroqCloud bekannt. Mit der Markteinführung werden gpt-oss-120B und gpt-oss-20B mit vollem 128K-Kontext, Echtzeitantworten und integrierten serverseitigen Tools auf der optimierten Inferenzplattform von Groq vom ersten Tag an live bereitgestellt.

Groq unterstützt seit langem die Open-Source-Bemühungen von OpenAI, einschließlich des groß angelegten Einsatzes von Whisper. Diese Markteinführung baut auf dieser Grundlage auf und bringt ihre neuesten Modelle mit globalem Zugang und lokaler Unterstützung durch HUMAIN in Produktion.

„OpenAI setzt einen neuen Hochleistungsstandard für Open-Source-Modelle“, sagte Jonathan Ross, CEO von Groq. „Groq wurde entwickelt, um solche Modelle schnell und kostengünstig auszuführen, so dass Entwickler sie vom ersten Tag an nutzen können. Die Zusammenarbeit mit HUMAIN stärkt den lokalen Zugang und die Unterstützung im Königreich Saudi-Arabien und ermöglicht es den Entwicklern in der Region, intelligenter und schneller zu bauen.“

„Groq bietet die unübertroffene Inferenzgeschwindigkeit, Skalierbarkeit und Kosteneffizienz, die wir benötigen, um modernste KI in das Königreich zu bringen“, sagte Tareq Amin, CEO von HUMAIN. „Gemeinsam ermöglichen wir eine neue Welle saudischer Innovationen – angetrieben von den besten Open-Source-Modellen und der Infrastruktur, um sie weltweit zu skalieren. Wir sind stolz darauf, die Führungsrolle von OpenAI im Bereich der Open-Source-KI zu unterstützen.“

Gebaut für volle Modellfähigkeit

Um die neuen Modelle von OpenAI optimal zu nutzen, bietet Groq erweiterten Kontext und integrierte Tools wie Codeausführung und Websuche. Die Websuche hilft bei der Bereitstellung relevanter Informationen in Echtzeit, während die Codeausführung logisches Denken und komplexe Arbeitsabläufe ermöglicht. Die Plattform von Groq bietet diese Fähigkeiten vom ersten Tag an mit einer vollen 128k Token-Kontextlänge.

Unübertroffenes Preis-Leistungs-Verhältnis

Der speziell entwickelte Stack von Groq bietet die niedrigsten Kosten pro Token für die neuen Modelle von OpenAI und gewährleistet gleichzeitig Geschwindigkeit und Genauigkeit.

gpt-oss-120B läuft derzeit mit 500+ t/s und gpt-oss-20B läuft derzeit mit 1000+ t/s auf GroqCloud.

Groq bietet die neuesten offenen Modelle von OpenAI zu den folgenden Preisen an:

- gpt-oss-120B: 0,15 USD / M Eingabe-Token und 0,75 USD / M Ausgabe-Token
- gpt-oss-20B: 0,10 USD / M Eingabe-Token und 0,50 USD / M Ausgabe-Token

Hinweis: Für eine begrenzte Zeit werden Tool-Aufrufe, die mit den offenen Modellen von OpenAI verwendet werden, nicht berechnet. Weitere Informationen finden Sie unter groq.com/pricing.

Global vom ersten Tag an

Die globale Präsenz von Groq mit Rechenzentren in Nordamerika, Europa und dem Nahen Osten gewährleistet eine zuverlässige und leistungsstarke KI-Inferenz, wo immer Entwickler tätig sind. **Durch GroqCloud sind die offenen Modelle von OpenAI nun weltweit** mit minimaler Latenzzeit verfügbar.

Informationen zu Groq

Groq ist die KI-Inferenzplattform, die die Preisleistung neu definiert. Die maßgeschneiderte LPU und die Cloud wurden speziell entwickelt, um leistungsstarke Modelle sofort, zuverlässig und zu den niedrigsten Kosten pro

Token auszuführen – ohne Kompromisse. Mehr als 1,9 Millionen Entwickler vertrauen auf Groq, um schnell zu entwickeln und intelligenter zu skalieren.

Informationen zu HUMAIN

www.humain.ai

Kontakt:

pr-media@groq.com HUMAIN, ein PIF-Unternehmen, ist ein globales Unternehmen für künstliche Intelligenz, das umfassende KI-Fähigkeiten in vier Kernbereichen anbietet: Rechenzentren der nächsten Generation, Hochleistungsinfrastrukturen und Cloud-Plattformen, fortschrittliche KI-Modelle, einschließlich der weltweit fortschrittlichsten arabischen multimodalen LLMs, und transformative KI-Lösungen, die tiefe Einblicke in die Branche mit der Umsetzung in der Praxis verbinden.

Das End-to-End-Modell von HUMAIN dient sowohl Organisationen des öffentlichen als auch des privaten Sektors, um in allen Branchen exponentielle Werte zu erschließen, die Transformation voranzutreiben und die Fähigkeiten durch Synergien zwischen Mensch und KI zu stärken. Mit einem wachsenden Portfolio an branchenspezifischen KI-Produkten und der Kernaufgabe, die IP-Führerschaft und die weltweite Vorherrschaft von Talenten voranzutreiben, ist HUMAIN auf globale Wettbewerbsfähigkeit und nationale Differenzierung ausgerichtet.

www.humain.ai Logo - https://mma.prnewswire.com/media/2725167/Groq_Logo_No_Trademark.jpg

View original content: <https://www.prnewswire.com/news-releases/groq-und-humain-bieten-neue-offene-modelle-von-openai-vom-ersten-tag-an-302522720.html>

Diese Meldung kann unter <https://www.presseportal.ch/de/pm/100102574/100933876> abgerufen werden.