# Beyond Nvidia: the search for AI's next breakthrough

June 23 2024, by Alex PIGMAN



Vinod Khosla, founder of Khosla Ventures, speaks on a panel on the main stage during the 2024 Collision tech conference in Toronto, Canada.

For a few days, AI chip juggernaut Nvidia sat on the throne as the world's biggest company, but behind its staggering success are questions

on whether new entrants can stake a claim to the artificial intelligence bonanza.

Nvidia, which makes the processors that are the only option to train generative AI's large language models, is now Big Tech's newest member and its stock market takeoff has lifted the whole sector.

Even tech's second rung on Wall Street has ridden on Nvidia's coattails with Oracle, Broadcom, HP and a spate of others seeing their stock valuations surge, despite sometimes shaky earnings.

Amid the champagne popping, startups seeking the attention of Silicon Valley venture capitalists are being asked to innovate -- but without a clear indication of where the next chapter of AI will be written.

When it comes to generative AI, doubts persist on what exactly will be left for companies that are not existing [model](#) makers, a field dominated by Microsoft-backed OpenAI, Google and Anthropic.

Most agree that competing with them head-on could be a fool's errand.

"I don't think that there's a great opportunity to start a foundational AI company at this point in time," said Mike Myer, founder and CEO of tech firm Quiq, at the Collision technology conference in Toronto.

Some have tried to build applications that use or mimic the powers of the existing big models, but this is being slapped down by Silicon Valley's biggest players.

"What I find disturbing is that people are not differentiating between those applications which are roadkill for the models as they progress in their capabilities, and those that are really adding value and will be here 10 years from now," said venture capital veteran Vinod Khosla.

## 'Won't keep up'

The tough-talking Khosla is one of OpenAI's earliest investors.

"Grammarly won't keep up," Khosla predicted of the spelling and grammar checking app, and others similar to it.

He said these companies, which put only a "thin wrapper" around what the AI models can offer, are doomed.

One of the fields ripe for the taking is [chip design](#), Khosla said, with AI demanding ever more specialized processors that provide highly specific powers.

"If you look across the chip history, we really have for the most part focused on more general chips," Rebecca Parsons, CTO at tech consultancy Thoughtworks, told AFP.

Providing more specialized processing for the many demands of AI is an opportunity seized by Groq, a hot startup that has built chips for the deployment of AI as opposed to its training, or inference -- the specialty of Nvidia's world-dominating GPUs.

Groq CEO Jonathan Ross told AFP that Nvidia won't be the best at everything, even if they are uncontested for generative AI training.

"Nvidia and (its CEO) Jensen Huang are like Michael Jordan... the greatest of all time in basketball. But inference is baseball, and we try and forget the time where Michael Jordan tried to play baseball and wasn't very good at it," he said.

Another opportunity will come from highly specialized AI that will provide expertise and know-how based on proprietary data which won't

be co-opted by voracious big tech.

"Open AI and Google aren't going to build a structural engineer. They're not going to build products like a primary care doctor or a mental health therapist," said Khosla.

Profiting from highly specialized data is the basis of Cohere, another of Silicon Valley's hottest startups that pitches specifically-made models to businesses that are skittish about AI veering out of their control.

"Enterprises are skeptical of technology, and they're risk-averse, and so we need to win their trust and to prove to them that there's a way to adopt this technology that's reliable, trustworthy and secure," Cohere CEO Aidan Gomez told AFP.

When he was just 20 and working at Google, Gomez co-authored the seminal paper "Attention Is All You Need," which introduced Transformer, the architecture behind popular large language models like OpenAI's GPT-4.

The company has received funding from Nvidia and Salesforce Ventures and is valued in the billions of dollars.

© 2024 AFP