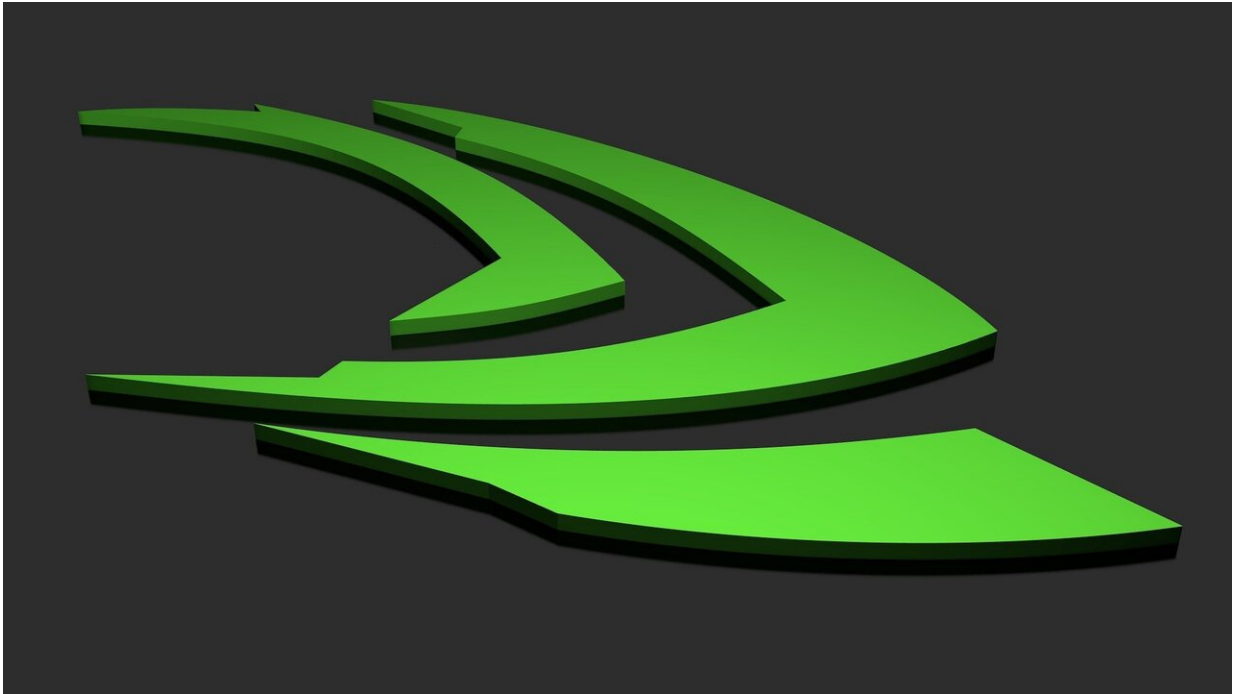# How Nvidia became an AI giant

June 22 2024, by Sarah Parvini

Credit: Pixabay/CC0 Public Domain

It all started at a Denny's in San Jose in 1993. Three engineers—Jensen Huang, Chris Malachowsky and Curtis Priem—gathered at the diner in what is now the heart of Silicon Valley to discuss building a computer chip that would make graphics for video games faster and more realistic. That conversation, and the ones that followed, led to the founding of Nvidia, the tech company that soared through the ranks of the stock market to briefly top Microsoft as the most valuable company in the

S&P 500 this week.

The company is now worth over $3.2 trillion, with its dominance as a chipmaker cementing Nvidia's place as the poster child of the artificial intelligence boom—a moment that Huang, Nvidia's CEO, has dubbed "the next industrial revolution."

On a conference call with analysts last month, Huang predicted that the companies using Nvidia chips would build a new type of data center called "AI factories."

Huang added that training AI models is becoming a faster process as they learn to become "multimodal"—able to understand text, speech, images, video and 3D data—and also "to reason and plan."

"People kind of talk about AI as if Jensen just kind of arrived like in the last 18 months, like 24 months ago all of a sudden figured this out," said Daniel Newman, CEO of The Futurum Group, a tech research firm. "But if you actually go back in time and listen to Jensen talking about accelerated computing, he's been sharing his vision for more than a decade."

The Santa Clara, California-based tech company's invention of the graphics processor unit, or GPU, in 1999 helped spark the growth of the PC gaming market and redefined computer graphics. Now Nvidia's specialized chips are key components that help power different forms of artificial intelligence, including the latest generative AI chatbots such as ChatGPT and Google's Gemini.

Nvidia's GPUs are a key factor in the company's success in artificial intelligence, Newman added.

"They took an architecture that was used for a single thing, to maybe

enhance gaming, and they figured out how to network these things," he said. "The GPU became the most compelling architecture for AI, going from gaming, rendering graphics and stuff, to actually using it for data. ... They basically ended up creating a market that didn't exist, which was GPUs for AI, or GPUs for machine learning."

AI chips are designed to perform artificial intelligence tasks faster and more efficiently. While general-purpose chips like CPUs can also be used for simpler AI tasks, they're "becoming less and less useful as AI advances," a 2020 report from Georgetown University's Center for Security and Emerging Technology found.

Tech giants are snapping up Nvidia chips as they wade deeper into AI—a movement that's enabling cars to drive by themselves, and generating stories, art and music.

"Jensen basically has made AI digestible and then Apple will make it consumable," Newman said.

The company carved out an early lead in the hardware and software needed to tailor its technology to AI applications, partly because Huang nudged it into what was still a nascent technology more than a decade ago.

"Nvidia has been working on different portions of this problem for more than two decades now. They have a deep innovation engine that goes all the way back to the early 2000s," said Chirag Dekate, a VP analyst at Gartner, a tech research and consulting firm. "What Nvidia did two decades ago is they both identified and they nurtured an adjacent market where they discovered that the same processors, same GPUs that they were using for graphics could be shaped to solve highly parallel tasks."

At the time, he said, AI was only in its infancy. But Nvidia's

understanding that GPUs would be central to the development of AI was "the fundamental breakthrough that was needed," Dekate said.

"Until then, we would have been, I would say, in the analytic Dark Ages," he said. "The analytics were there, but we could never bring these AI elements to life."

Analysts estimate that Nvidia's revenue for the [fiscal year](#) that ends in January 2025 will reach $119.9 billion—about double its revenue for fiscal 2024 and more than four times its receipts the year before that.

"My hypothesis is the kind of exponential growth that we're seeing with Nvidia today is potentially a pattern that we're going to see replicated more frequently in the decades to come," he said. "This is the Golden Age if you will...this is the best time to be an AI engineer."