

# AI Opener: OpenAI's Sutskever in Conversation With Jensen Huang

In a fireside chat at GTC, NVIDIA's founder and CEO and OpenAI co-founder Ilya Sutskever discussed GPT-4, ChatGPT, deep learning's future and how it all began.

Author: Rick Merritt

Like old friends catching up over coffee, two industry icons reflected on how modern AI got its start, where it's at today and where it needs to go next.

Jensen Huang, founder and CEO of NVIDIA, interviewed AI pioneer Ilya Sutskever in a fireside chat at GTC . The talk was recorded a day after the launch of GPT-4, the most powerful AI model to date from OpenAI, the research company Sutskever co-founded.

They talked at length about GPT-4 and its forerunners, including ChatGPT. That generative AI model, though only a few months old, is already the most popular computer application in history.

Their conversation touched on the capabilities, limits and inner workings of the deep neural networks that are capturing the imaginations of hundreds of millions of users.

Compared to ChatGPT, GPT-4 marks a "pretty substantial improvement across many dimensions," said Sutskever, noting the new model can read images as well as text.

"In some future version, [users] might get a diagram back" in response to a query, he said.

"There's a misunderstanding that ChatGPT is one large language model, but there's a system around it," said Huang.

In a sign of that complexity, Sutskever said OpenAI uses two levels of training.

The first stage focuses on accurately predicting the next word in a series. Here, "what the neural net learns is some representation of the process that produced the text, and that's a projection of the world," he said.

The second "is where we communicate to the neural network what we want, including guardrails ... so it becomes more reliable and precise," he added.

While he's at the swirling center of modern AI today, Sutskever was also present at its creation.

In 2012, he was among the first to show the power of deep neural networks trained on massive datasets. In an academic contest, the AlexNet model he demonstrated with AI pioneers Geoff Hinton and Alex Krizhevsky recognized images faster than a human could.

Huang referred to their work as the Big Bang of AI .

The results "broke the record by such a large margin, it was clear there was a discontinuity here," Huang said.

Part of that breakthrough came from the parallel processing the team applied to its model with GPUs.

"The ImageNet dataset and a convolutional neural network were a great fit for GPUs that made it unbelievably fast to train something unprecedented," Sutskever said.

That early work ran on a few GeForce GTX 580 GPUs in a University of Toronto lab. Today, tens of thousands of the latest NVIDIA A100 and H100 Tensor Core GPUs in the Microsoft Azure cloud service handle training and inference on models like ChatGPT.

"In the 10 years we've known each other, the models you've trained [have grown by] about a million times," Huang said. "No one in computer science would have believed the computation done in that time would be a million times larger."

"I had a very strong belief that bigger is better, and a goal at OpenAI was to scale," said Sutskever.

Along the way, the two shared a laugh.

"Humans hear a billion words in a lifetime," Sutskever said.

"Does that include the words in my own head," Huang shot back.

"Make it 2 billion," Sutskever deadpanned.

They ended their nearly hour-long talk discussing the outlook for AI.

Asked if GPT-4 has reasoning capabilities, Sutskever suggested the term is hard to define and the capability may still be on the horizon.

"We'll keep seeing systems that astound us with what they can do," he said. "The frontier is in reliability, getting to a point where we can trust what it can do, and that if it doesn't know something, it says so," he added.

"Your body of work is incredible ... truly remarkable," said Huang in closing the session. "This has been one of the best beyond Ph.D. descriptions of the state of the art of large language models," he said.

To get all the news from GTC, watch the keynote below.

Original URL: <https://blogs.nvidia.com/blog/2023/03/22/sutskever-openai-gtc/>