Ilya Sutskever
AI visionary

"He's got a very strong moral compass."

A pioneer of ChatGPT and other AI systems that are changing society.

By Nicola Jones

As a teenager, Ilya Sutskever knocked on Geoffrey Hinton's door at the University of Toronto in Canada and asked for a job. "He said he was cooking fries to make money over the summer, and he would rather be working for me doing AI," says Hinton, who is often recognized as the godfather of modern artificial intelligence (AI).

Hinton gave the ambitious student some papers to read, and Sutskever came back wondering why the authors hadn't worked out what seemed, to him, obvious solutions. "All his instincts were correct," says Hinton. He credits Sutskever with being a visionary pioneer in deep learning and large language models (LLMs) — the roots of conversational AI bots, such as ChatGPT, that have captivated the world this year. "It's not just intelligence" that sets him apart, says Hinton. "It's the urgency with which he gets on with things."

Sutskever became the chief scientist at OpenAI in San Francisco, California, where he has had a central role in developing ChatGPT. But he's also worried about the future of AI. This July, he shifted focus to co-lead OpenAI's four-year 'superalignment' project, which the company said will use 20% of its computing power to study how "to steer and control AI systems much smarter than us".

A tension between safety and the commercial incentive to move fast might have contributed to a bewildering drama this November, in which Sutskever played a key part in firing and then rehiring OpenAI's chief executive, Sam Altman. After the upheaval, Sutskever declined to talk to *Nature*.

To some, Sutskever's vision is admirable. "He's got a very strong moral compass," says Hinton. "He really is very concerned about AI safety." But others say that focusing on how to control yet-to-arrive AI systems distracts from the real and present dangers of the technology. It "puts intervention way off on the horizon", says Sarah Myers West, managing director of the AI Now Institute, a policy-research organization in New York City. Instead, she says, we need "to tackle near-term harms", such as AI systems reinforcing biases in their training data, or potentially leaking private information.

The lack of transparency about AI systems is also a concern. OpenAI and some other companies keep their code and training data private. Sutskever says that in the long term, closed systems will be the responsible course to avoid letting others make powerful AIs. "At some point, the capability will become so vast that it will be obviously irresponsible to open-source models," he said this April.

Sutskever, who was born in the Soviet Union in 1986, has always been a precocious learner, taking university-level classes in coding as a teenager in Israel. After his family moved to Canada, he began working with Hinton on deep learning in 2003. In 2012, Sutskever and another of Hinton's students built AlexNet, a neural network that won a landmark image-recognition competition by a startling margin. Sutskever later moved to Google, where he helped to develop AlphaGo, which beat human champions at the complex board game Go.

JIM WILSON/*THE NEW YORK TIMES*/REDUX/EYEVINE

In 2015, Sutskever was invited to dinner with Altman and others, including billionaire entrepreneur Elon Musk. They co-founded OpenAI as a non-profit organization that year, to "benefit humanity". Sutskever saw it as a chance to take seriously the pursuit of artificial general intelligence (AGI), a system as smart as any person. "Researchers are somehow, I would say, trained to think small … But at OpenAI we took the liberty to look at the big picture," he said earlier this year.

Wojciech Zaremba, another OpenAI co-founder, credits Sutskever with pushing the company to pour more effort into its Generative Pre-trained Transformer (GPT) system after the first incarnation of GPT-1 in 2018. Unlike many others at the time, Sutskever was convinced that scaling up computing power alone would make these systems smarter. "He understood that before almost anybody else," says Hinton.

To attract the funding needed for more computing power, the team moved OpenAI from a non-profit to a 'capped-profit' model in 2019, enticing technology giant Microsoft to pour billions in cash and computing resources into the operation. That paid off: the LLM improved and ChatGPT, released in November 2022, became a sensation.

Amid this success, internal chaos erupted when Sutskever and other OpenAI board members fired Altman on 17 November this year. Many employees threatened to decamp to Microsoft with Altman — including Sutskever, who said he regretted his actions. He was removed from the board when Altman rejoined the company five days later.

Throughout, Sutskever has been consistently bold in his statements about AI. In 2022, he declared that AI might already be "slightly conscious", variously evoking awe, terror and laughter. He says publicly that AGI and even 'superintelligence' surpassing the combined intellect of humanity could be developed within years or decades. "To this day, I'm surprised at how bullish he is," says AI researcher Andrew Ng, for whom Sutskever worked as a postdoc at Stanford University in California in 2012.

But, adds Ng, Sutskever "has the admirable trait of being able to pick a direction and pursue it relentlessly, regardless of whether others agree with him or not".