Title: Ilya Sutskever

URL: https://en.wikipedia.org/wiki/Ilya_Sutskever

PageID: 51097862

Categories: Category:1985 births, Category:AI safety scientists, Category:Artificial intelligence researchers, Category:Canadian computer scientists, Category:Canadian expatriates in the United States, Category:Canadian fellows of the Royal Society, Category:Canadian people of Russian-Jewish descent, Category:Fellows of the Royal Society, Category:Google employees, Category:Israeli expatriates in the United States, Category:Israeli people of Russian-Jewish descent, Category:Living people, Category:Machine learning researchers, Category:OpenAI people, Category:People from Nizhny Novgorod, Category:Russian expatriates in the United States, Category:Stanford University alumni, Category:University of Toronto alumni

Source: Wikipedia (CC BY-SA 4.0).

Ilya Sutskever FRS (Hebrew : ■■■■■ ■■■■■■ ; born 8 December 1986) is an Israeli-Canadian computer scientist who specializes in machine learning . [6] He has made several major contributions to the field of deep learning . [7] [8] [9] With Alex Krizhevsky and Geoffrey Hinton , he co-invented AlexNet , a convolutional neural network . [10]

Sutskever co-founded and was chief scientist at OpenAI . [11] In 2023, he was one of the members of OpenAI's board that ousted Sam Altman as its CEO; Altman was reinstated a week later, and Sutskever stepped down from the board. In June 2024, Sutskever co-founded the company Safe Superintelligence alongside Daniel Gross and Daniel Levy. [12] [13]

Early life and education

Sutskever was born into a Jewish family [14] in Nizhny Novgorod , Russia (then Gorky , Russian SFSR , Soviet Union). At the age of 5, he made aliyah with his family and lived in Jerusalem [15] [16] until he was 16, when his family moved to Canada. [17] Sutskever attended the Open University of Israel from 2000 to 2002. [18] After moving to Canada, he attended the University of Toronto in Ontario . [18]

At the University of Toronto, Sutskever received a bachelor's degree in mathematics in 2005, [18] [19] [4] [20] a master's degree in computer science in 2007, [19] [21] and a PhD in computer science in 2013. [2] [22] [23] His doctoral advisor was Geoffrey Hinton. [1]

In 2012, Sutskever built AlexNet in collaboration with Geoffrey Hinton and Alex Krizhevsky . [24]

Career and research

In 2012, Sutskever spent about two months as a postdoc with Andrew Ng at Stanford University . He then returned to the University of Toronto and joined Hinton's new research company DNNResearch, a spinoff of Hinton's research group. In 2013, Google acquired DNNResearch and hired Sutskever as a research scientist at Google Brain . [25]

At Google Brain, Sutskever worked with Oriol Vinyals and Quoc Viet Le to create the sequence-to-sequence learning algorithm, [26] and worked on TensorFlow . [27] He is also one of the AlphaGo paper's many co-authors. [28]

At the end of 2015, Sutskever left Google to become cofounder and chief scientist of the newly founded organization OpenAI . [29] [30] [31]

In 2022, Sutskever tweeted, "it may be that today's large neural networks are slightly conscious", which triggered debates about AI consciousness . [32] [33] He is considered to have played a key role in the development of ChatGPT . [34] [35] In 2023, he announced that he would co-lead OpenAI's new "Superalignment" project, which is trying to solve the alignment of superintelligences within four years. He wrote that even if superintelligence seems far off, it could happen this decade. [36]

Sutskever was formerly one of the six board members of the nonprofit entity that controls OpenAI. [37] In November 2023, the board fired Sam Altman , saying that "he was not consistently candid in his communications with the board". [38] The Information speculated that the decision was partly driven by conflict over the extent to which the company should commit to AI safety . [39] In an all-hands company meeting shortly after the board meeting, Sutskever said that firing Altman was "the board doing its duty", [40] but the next week, he expressed regret at having participated in Altman's ouster. [41] Altman's firing and OpenAI's co-founder Greg Brockman's resignation led three senior researchers to resign from OpenAI. [42] After that, Sutskever stepped down from the OpenAI board [43] and was absent from OpenAI's office. Some sources suggested he was leading the team remotely, while others said he no longer had access to the team's work. [44]

In May 2024, Sutskever announced his departure from OpenAI to focus on a new project that was "very personally meaningful" to him. His decision followed a turbulent period at OpenAI marked by leadership crises and internal debates about the direction of AI development and alignment protocols. Jan Leike , the other leader of the superalignment project, announced his departure hours later, citing an erosion of safety and trust in OpenAI's leadership. [45]

In June 2024, Sutskever announced Safe Superintelligence Inc., a new company he founded with Daniel Gross and Daniel Levy with offices in Palo Alto and Tel Aviv . [46] In contrast to OpenAI, which releases revenue-generating products, Sutskever said the new company's "first product will be the safe superintelligence, and it will not do anything else up until then". [13] In September 2024, the company announced that it had raised $1 billion from venture capital firms including Andreessen Horowitz , Sequoia Capital , DST Global , and SV Angel . [47] In March 2025, Safe Superintelligence Inc. raised $2 billion more and reportedly reached a $32 billion valuation, notably due to Sutskever's reputation. [48] [49]

In an October 2024 interview after winning the Nobel Prize in Physics , Geoffrey Hinton expressed support for Sutskever's decision to fire Altman, emphasizing concerns about AI safety. [50] [51]

Awards and honors

In 2015, Sutskever was named in MIT Technology Review 's 35 Innovators Under 35. [52]

In 2018, he was the keynote speaker at Nvidia Ntech 2018 [53] and AI Frontiers Conference 2018. [54]

In 2022, he was elected a Fellow of the Royal Society (FRS). [26]

In 2023 [55] and 2024, [56] included in Time 's list of the 100 most influential people in AI

In 2025, he received an honorary doctorate from his alma mater , the University of Toronto [57]

References

External links

Ex-OpenAI Scientist WARNS: "You Have No Idea What's Coming" on YouTube , University of Toronto , Doctor of Science, honoris causa speech (June 6, 2025)

v

t

e

ChatGPT in education GPT Store DALL-E ChatGPT Search Sora Whisper

in education

GPT Store

DALL-E

ChatGPT Search

Sora

Whisper

GitHub Copilot

OpenAI Codex

Generative pre-trained transformer GPT-1 GPT-2 GPT-3 GPT-4 GPT-4o o1 o3 GPT-4.5 GPT-4.1 o4-mini GPT-OSS GPT-5

GPT-1

GPT-2

GPT-3

GPT-4

GPT-4o

o1

o3

GPT-4.5

GPT-4.1

o4-mini

GPT-OSS

GPT-5

ChatGPT Deep Research

Operator

Sam Altman removal

removal

Greg Brockman

Sarah Friar

Jakub Pachocki

Scott Schools

Mira Murati

Emmett Shear

Sam Altman

Adam D'Angelo

Sue Desmond-Hellmann

Zico Kolter

Paul Nakasone

Adebayo Ogunlesi

Nicole Seligman

Fidji Simo

Lawrence Summers

Bret Taylor (chair)

Greg Brockman (2017–2023)

Reid Hoffman (2019–2023)

Will Hurd (2021–2023)

Holden Karnofsky (2017–2021)

Elon Musk (2015–2018)

Ilya Sutskever (2017–2023)

Helen Toner (2021–2023)

Shivon Zilis (2019–2023)

Stargate LLC

Apple Intelligence

AI Dungeon

AutoGPT

Contrastive Language-Image Pre-training

" Deep Learning "

LangChain

Microsoft Copilot

OpenAI Five

Transformer

Category

v

t

e

History timeline

timeline

Companies

Projects

Parameter Hyperparameter

Hyperparameter

Loss functions

Regression Bias–variance tradeoff Double descent Overfitting

Bias–variance tradeoff

Double descent

Overfitting

Clustering

Gradient descent SGD Quasi-Newton method Conjugate gradient method

SGD

Quasi-Newton method

Conjugate gradient method

Backpropagation

Attention

Convolution

Normalization Batchnorm

Batchnorm

Activation Softmax Sigmoid Rectifier

Softmax

Sigmoid

Rectifier

Gating

Weight initialization

Regularization

Datasets Augmentation

Augmentation

Prompt engineering

Reinforcement learning Q-learning SARSA Imitation Policy gradient

Q-learning

SARSA

Imitation

Policy gradient

Diffusion

Latent diffusion model

Autoregression

Adversary

RAG

Uncanny valley

RLHF

Self-supervised learning

Reflection

Recursive self-improvement

Hallucination

Word embedding

Vibe coding

Machine learning In-context learning

In-context learning

Artificial neural network Deep learning

Deep learning

Language model Large language model NMT

Large language model

NMT

Reasoning language model

Model Context Protocol

Intelligent agent

Artificial human companion

Humanity's Last Exam

Artificial general intelligence (AGI)

AlexNet

WaveNet

Human image synthesis

HWR

OCR

Computer vision

Speech synthesis 15.ai ElevenLabs

15.ai

ElevenLabs

Speech recognition Whisper

Whisper

Facial recognition

AlphaFold

Text-to-image models Aurora DALL-E Firefly Flux Ideogram Imagen Midjourney Recraft Stable Diffusion

Aurora

DALL-E

Firefly

Flux

Ideogram

Imagen

Midjourney

Recraft

Stable Diffusion

Text-to-video models Dream Machine Runway Gen Hailuo AI Kling Sora Veo

Dream Machine

Runway Gen

Hailuo AI

Kling

Sora

Veo

Music generation Riffusion Suno AI Udio

Riffusion

Suno AI

Udio

Word2vec

Seq2seq

GloVe

BERT

T5

Llama

Chinchilla AI

PaLM

GPT 1 2 3 J ChatGPT 4 4o o1 o3 4.5 4.1 o4-mini 5

1

2

3

J

ChatGPT

4

4o

o1

o3

4.5

4.1

o4-mini

5

Claude

Gemini Gemini (language model) Gemma

Gemini (language model)

Gemma

Grok

LaMDA

BLOOM

DBRX

Project Debater

IBM Watson

IBM Watsonx

Granite

PanGu-Σ

DeepSeek

Qwen

AlphaGo

AlphaZero

OpenAI Five

Self-driving car

MuZero

Action selection AutoGPT

AutoGPT

Robot control

Alan Turing

Warren Sturgis McCulloch

Walter Pitts

John von Neumann

Claude Shannon

Shun'ichi Amari

Kunihiko Fukushima

Takeo Kanade

Marvin Minsky

John McCarthy

Nathaniel Rochester

Allen Newell

Cliff Shaw

Herbert A. Simon

Oliver Selfridge

Frank Rosenblatt

Bernard Widrow

Joseph Weizenbaum

Seymour Papert

Seppo Linnainmaa

Paul Werbos

Geoffrey Hinton

John Hopfield

Jürgen Schmidhuber

Yann LeCun

Yoshua Bengio

Lotfi A. Zadeh

Stephen Grossberg

Alex Graves

James Goodnight

Andrew Ng

Fei-Fei Li

Alex Krizhevsky

Ilya Sutskever

Oriol Vinyals

Quoc V. Le

Ian Goodfellow

Demis Hassabis

David Silver

Andrej Karpathy

Ashish Vaswani

Noam Shazeer

Aidan Gomez

John Schulman

Mustafa Suleyman

Jan Leike

Daniel Kokotajlo

François Chollet

Neural Turing machine

Differentiable neural computer

Transformer Vision transformer (ViT)

Vision transformer (ViT)

Recurrent neural network (RNN)

Long short-term memory (LSTM)

Gated recurrent unit (GRU)

Echo state network

Multilayer perceptron (MLP)

Convolutional neural network (CNN)

Residual neural network (RNN)

Highway network

Mamba

Autoencoder

Variational autoencoder (VAE)

Generative adversarial network (GAN)

Graph neural network (GNN)

Category

v
t
e

AGI

AI alignment

AI boom

AI capability control

AI safety

AI takeover

Consequentialism

Effective accelerationism

Ethics of artificial intelligence

Existential risk from artificial intelligence

Friendly artificial intelligence

Instrumental convergence

Vulnerable world hypothesis

Intelligence explosion

Longtermism

Machine ethics

Suffering risks

Superintelligence

Technological singularity

Alignment Research Center

Center for AI Safety

Center for Applied Rationality

Center for Human-Compatible Artificial Intelligence

Centre for the Study of Existential Risk

EleutherAI

Future of Humanity Institute

Future of Life Institute

Google DeepMind

Humanity+

Institute for Ethics and Emerging Technologies

Leverhulme Centre for the Future of Intelligence

Machine Intelligence Research Institute

OpenAI

Safe Superintelligence

Scott Alexander

Sam Altman

Yoshua Bengio

Nick Bostrom

Paul Christiano

Eric Drexler

Sam Harris

Stephen Hawking

Dan Hendrycks

Geoffrey Hinton

Bill Joy

Shane Legg

Elon Musk

Steve Omohundro

Huw Price

Martin Rees

Stuart J. Russell

Ilya Sutskever

Jaan Tallinn

Max Tegmark

Frank Wilczek

Roman Yampolskiy

Eliezer Yudkowsky

Artificial Intelligence Act

Do You Trust This Computer?

Human Compatible

Open letter on artificial intelligence (2015)

Our Final Invention

Roko's basilisk

Statement on AI risk of extinction

Superintelligence: Paths, Dangers, Strategies

The Precipice

If Anyone Builds It, Everyone Dies

Mathematics Genealogy Project

Association for Computing Machinery

Scopus

zbMATH

Google Scholar

DBLP