
Nikita Sutskever

Senior AI Engineer | Machine Learning Architect | AI Researcher

 San Francisco, CA |  nikitasutskever@aiworks.io |  +1 (415) 555-9021

 www.nikitasutskever.com | GitHub: github.com/nikita-ai | LinkedIn: linkedin.com/in/nikitasutskever-ai

Professional Summary

Innovative and impact-driven **Senior AI Engineer** with over **12 years** of experience building, scaling, and deploying AI/ML systems across **FAANG companies (Facebook, Apple, Amazon, Netflix, Google)** and high-impact startups. Expert in machine learning architecture, LLMs, computer vision, NLP, reinforcement learning, and MLOps. Proven track record solving global-scale problems—from fraud detection and autonomous systems to climate AI and personalized healthcare.

Core Competencies

- **AI/ML:** Deep Learning, LLMs, Transformers, GANs, CNNs, RL, GNNs
 - **Languages & Frameworks:** Python, PyTorch, TensorFlow, JAX, Scala, Go
 - **Cloud & Infra:** AWS, GCP, Azure, Kubernetes, Airflow, MLflow, Ray
 - **Data:** SQL, Spark, Snowflake, Pandas, NumPy, Apache Kafka
 - **Tooling:** Docker, GitOps, CI/CD, HuggingFace, DVC, LangChain
 - **Research:** Paper writing, benchmarking, model optimization (quantization, pruning), graph-based AI
-



Work Experience

Google DeepMind – Mountain View, CA

Senior AI Research Engineer | Mar 2022 – Present

- Led development of **AlphaCarbon**, a generative model for protein folding with faster inference than AlphaFold.
- Created knowledge-aware LLMs for scientific QA over 20M+ PubMed articles.
- Built scalable distributed training infrastructure with 4x speedup on TPUv4 pods.

Meta (Facebook AI Research - FAIR) – Menlo Park, CA

Staff AI Engineer | Oct 2018 – Feb 2022

- Architected **RealTimeDet**, Meta's production object detection system for Reels and Marketplace.
- Reduced model latency by 60% using quantized vision transformers on mobile devices.
- Co-authored Meta's fairness-aware recommendation engine for 3B users.

Amazon Web Services (AWS AI Labs) – Seattle, WA

Senior Applied Scientist | May 2015 – Sep 2018

- Designed and deployed **anomaly detection models** for AWS Shield (DDoS protection).
- Built AutoML pipeline powering over 500 enterprise ML workloads.
- Led AWS Sagemaker's distributed training integration with Horovod + EFA.

Netflix – Los Gatos, CA

Machine Learning Engineer II | Jul 2013 – Apr 2015

- Optimized personalization algorithms for dynamic thumbnails and content ranking.
- Developed **adaptive bit-rate prediction models** for 200M+ global users.

Apple – Cupertino, CA

AI/ML Intern, Siri Team | Summer 2012

- Built a voice intent classifier improving Siri's accuracy on African English accents by 35%.
-

Flagship Projects

ClimateNet (2021 – Present)

- Co-founded open-source AI project to predict extreme climate events.
- Uses satellite imagery + temporal transformers to model hurricane trajectories.
- Deployed to African climate agencies for early flood warnings.

LLMGuard

- Built a security layer for enterprise LLMs to detect prompt injection, hallucination risks, and leakage.
- Adopted by 20+ enterprise clients and integrated into LangChain plugins.

MediVision AI

- Deep learning system for retinal disease detection using fundus images.
- Deployed in 15 rural hospitals across India and East Africa.

SafeChat AI – Open-source

- Contextual moderation and real-time toxicity filtering for chatbots and generative AI apps.
 - Hugging Face model with 10M+ downloads.
-

Education

Stanford University – M.S. in Artificial Intelligence (2011 – 2013)

- Advisor: Prof. Fei-Fei Li | Focus: Vision + Language models
- Thesis: *Neural Scene Graphs for Multimodal Reasoning*

Egerton University, Kenya – B.Sc. in Computer Science (2007 – 2011)

- Valedictorian | ACM ICPC Finalist | Founded AI Student Club
-

Awards & Recognition

-  *Google AI Impact Challenge Finalist* – 2023
 -  *Best Paper, NeurIPS Workshop on AI for Climate* – 2022
 -  *Meta AI Top Innovator Award* – 2021
 -  *Stanford AI Fellowship Recipient* – 2012
-

Talks & Publications

-  *"Towards Safer Large Language Models for Enterprise Use"* – ACL 2024
 -  *"Engineering AI for the Planet"* – TEDx Nairobi 2023
 -  *"Scalable Transformers for Video Understanding"* – CVPR 2021
 -  Speaker at Google Next, re:Invent, NeurIPS Expo, and PyData Global
-

Open Source Contributions

-  Maintainer – [mlops-cookbook](#) (35K+ stars)
 -  Contributor – HuggingFace Transformers, DeepSpeed, LangChain
 -  Author – **“AI Engineering from First Principles”** (O’Reilly, 2025 release)
-

Professional Affiliations

- Member – Association for Computing Machinery (ACM)
 - Member – IEEE Computational Intelligence Society
 - Contributor – OpenAI Safety Community
-

References

Available upon request or at www.nikitasutskever-ai.com