

Dr. Alexandra Chen

✉ a.chen@stanford.edu ☎ +1 (650) 555-0142 📍 Stanford, CA
🌐 alexchen.io 🐙 github.com/alexchen 🔗 [linkedin.com/in/alexchen](https://www.linkedin.com/in/alexchen) 🎓 Google Scholar

Summary

Machine learning researcher with 8+ years of experience in natural language processing and large language models. Published 25+ papers at top venues (NeurIPS, ICML, ACL, EMNLP) with 3,800+ citations. Led research teams developing foundation models deployed to millions of users. Seeking faculty positions in computer science.

Education

Ph.D.

in Computer Science

Stanford University

2016 – 2021

Stanford, CA

- Thesis: “Efficient Transformers for Long-Range Sequence Modeling”
- Advisor: Prof. Christopher Manning Committee: Profs. Percy Liang, Dan Jurafsky
- Stanford Graduate Fellowship (SGF) recipient

M.S.

in Computer Science

Carnegie Mellon University

2014 – 2016

Pittsburgh, PA

- Focus: Machine Learning and Language Technologies
- GPA: 4.0/4.0 Dean’s List all semesters

B.S.

in Mathematics and Computer Science

MIT

2010 – 2014

Cambridge, MA

- Summa Cum Laude Phi Beta Kappa
- Undergraduate thesis on spectral graph theory (awarded departmental prize)

Professional Experience

Senior Research Scientist

Google DeepMind

2023 – Present

Mountain View, CA

- Lead a team of 8 researchers developing next-generation language models
- Designed novel attention mechanisms reducing compute by 40% at scale
- Core contributor to Gemini model family; responsible for efficiency improvements
- Mentored 4 PhD interns, all placed at top research labs or faculty positions

Research Scientist

Meta AI (FAIR)

2021 – 2023

Menlo Park, CA

- Developed sparse mixture-of-experts models for efficient inference

- Led the retrieval-augmented generation (RAG) project for factual consistency
- Published 8 papers; 2 received spotlight/oral designations at NeurIPS
- Open-sourced model checkpoints downloaded 500K+ times

Research Intern
OpenAI

Summers 2018, 2019
San Francisco, CA

- Contributed to GPT-2 training infrastructure and evaluation framework
- Developed techniques for reducing hallucination in text generation

Selected Publications

1. **A.**
Chen, J.
Smith, R.
Patel. “FlashLinear: Sub-Quadratic Attention via Structured State Spaces.” *NeurIPS 2024*. (**Oral, top 0.5%**)
2. **A.**
Chen, M.
Johnson. “Scaling Laws for Mixture-of-Experts Language Models.” *ICML 2024*.
3. **L.**
Wang, **A.**
Chen, K.
Lee. “Retrieval-Augmented Generation with Learned Relevance.” *ACL 2023*. (**Best Paper Award**)
4. **A.**
Chen, T.
Brown, S.
Narang. “Efficient Long-Context Transformers via Sliding Window Attention.” *NeurIPS 2022*. (**Spotlight**)
5. **A.**
Chen, C.
Manning. “Linear Transformers with Gated Retention.” *ICLR 2021*.

Awards & Honors

- ACL Best Paper Award, 2023
- MIT Technology Review 35 Under 35, 2024
- Stanford Graduate Fellowship, 2016–2021
- NSF Graduate Research Fellowship, 2016
- CRA Outstanding Undergraduate Award, 2014
- Phi Beta Kappa, MIT Chapter, 2014

Technical Skills

Languages: Python, C++, CUDA, Julia, Rust
Frameworks: PyTorch, JAX, TensorFlow, Hugging Face, DeepSpeed, Megatron-LM
Infrastructure: Kubernetes, Slurm, TPU Pods, multi-node GPU training (up to 2048 GPUs)
Specialties: Large language models, efficient attention, distributed training, NLP

Academic Service

- **Area Chair:** NeurIPS 2024, ICML 2024, ACL 2023
- **Reviewer:** NeurIPS, ICML, ICLR, ACL, EMNLP, JMLR, TACL (100+ reviews total)
- **Workshop Organizer:** “Efficient NLP” at EMNLP 2023; “Scaling Laws” at ICML 2024
- **Invited Talks:** Google, Microsoft, Meta, Amazon, Apple, NVIDIA, 15+ universities