Simran Arora

Contact

E-mail: simran@cs.stanford.edu
Website: https://arorasimran.com/

Education

2019–	PhD in Computer Science, Stanford University Advisor: Christopher Ré
2015-2019	Jerome Fisher Management & Technology Program, University of Pennsylvania *BSE in Computer Science, School of Engineering* *BS in Economics (Finance Concentration), The Wharton School* Advisors: Boon Thau Loo, Vijay Kumar, Vincent Liu

Experience

2019-	PhD Student, **Stanford University**
2023-	Academic Partner, **Cartesia AI**
Fall 2023	Co-Instructor, **Stanford University** Created the new 3-Unit Systems for ML course, Taught 100+ students https://cs229s.stanford.edu/fall2023/
Fall 2023	MLSys YouTube Podcast Host, **Stanford MLSys Seminar** 16,000+ subscribers, 30,000+ monthly views
2021-2022	Research Scientist, **Facebook AI Research** Worked with Jacob Kahn, Patrick Lewis, Angela Fan, and Ronan Collobert Published our research at TACL
Summer 2017	Software Engineer, **Google**

Awards

July 2024	**ICML Spotlight Award (Top 3.5% of Papers)** Simple linear attention language models balance the recall-throughput tradeoff
December 2023	**NeurIPS Outstanding Paper Award (4 Papers in 12.3K Submissions)** DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models
December 2023	**NeurIPS Oral Award (Top 0.5% of Papers)** Monarch Mixer: A Simple Sub-Quadratic GEMM-Based Architecture
April 2023	**ICLR Spotlight Award (Top 25% of Accepted Papers)** Ask Me Anything: A simple strategy for prompting language models
February 2023	**AAAI KnowledgeNLP Workshop: Oral Award (Top 20% of Papers)** Reasoning over Public and Private Data in Retrieval-Based Systems
2019-2023	**Stanford Graduate Fellowship**
2019	**Rhodes Scholarship National Finalist**
2019	**Marshall Scholarship National Finalist**
2019	**Penn Computer Science Academic Award** One graduating CS major per year

2019	**Michele Huber and Bryan D. Giles Memorial Award** One graduating Jerome Fisher student per year
2019	**Penn Computer Science Senior Engineering Capstone Project 2nd Place**
2019	**Wharton School Summa Cum Laude (Highest Honors)**
2019	**Penn Engineering Summa Cum Laude (Highest Honors)**
July 2017	**Best Paper Runner Up: IEEE MARSS Conference** Control of multiple microrobots with multiscale magnetic field superposition
2017-2019	**University of Pennsylvania Tau Beta Pi and Eta Kappa Nu**
2016	**The University Physics Competition (Worldwide Competition), Silver**
2015	**Siemens Research Competition Semifinalist**

Publications

[1] Simran Arora, Sabri Eyuboglu, Michael Zhang, Aman Timalsina, Silas Alberti, James Zou, Atri Rudra, and Christopher Ré
Simple linear attention language models balance the recall-throughput tradeoff
International Conference on Machine Learning (ICML), 2024. **Spotlight Award** and ICML ES-FoMo, 2024. **Oral Award**

[2] Jon Saad-Falcon, Daniel Y. Fu, Simran Arora, Neel Guha, and Christopher Ré
Benchmarking and Building Long-Context Retrieval Models with LoCo and M2-BERT
International Conference on Machine Learning (ICML), 2024 and ICML ES-FoMo, 2024.

[3] Simran Arora, Sabri Eyuboglu, Aman Timalsina, Isys Johnson, Michael Poli, James Zou, Atri Rudra, and Christopher Ré
Zoology: Measuring and Improving Recall in Efficient Language Models
International Conference on Learning Representations (ICLR), 2024.

[4] Furui Cheng, Vilém Zouhar, Simran Arora, Mrinmaya Sachan, Hendrik Strobelt, and Mennatallah El-Assady
RELIC: Investigating Large Language Model Responses using Self-Consistency
Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI), 2024.

[5] Daniel Y. Fu, Simran Arora, Jessica Grogan, Isys Johnson, Sabri Eyuboglu, Armin W. Thomas, Benjamin F. Spector, Michael Poli, Atri Rudra, and Christopher Ré
Monarch Mixer: A Simple Sub-Quadratic GEMM-Based Architecture
Advances in Neural Information Processing Systems (NeurIPS), 2023. **Oral Award**

[6] Boxin Wang, Weixin Chen, Hengzhi Pei, Chulin Xie, Mintong Kang, Chenhui Zhang, Chejian Xu, Zidi Xiong, Ritik Dutta, Rylan Schaeffer, Sang T. Truong, Simran Arora, Mantas Mazeika, Dan Hendrycks, Zinan Lin, Yu Cheng, Sanmi Koyejo, Dawn Song, and Bo Li
DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models
Advances in Neural Information Processing Systems (NeurIPS), 2023. **Oral Award** and **Outstanding Paper Award**

[7] Simran Arora, Brandon Yang, Sabri Eyuboglu, Avanika Narayan, Andrew Hojel, Immanuel Trummer, and Christopher Ré
Language Models Enable Simple Systems for Generating Structured Views of Heterogeneous Data Lakes
Proceedings of the VLDB Endowment (PVLDB), 2023.

[8] Simran Arora, Avanika Narayan, Mayee F. Chen, Laurel Orr, Neel Guha, Kush Bhatia, Ines Chami, Frederic Sala, and Christopher Ré
Ask Me Anything: A simple strategy for prompting language models
International Conference on Learning Representations (ICLR), 2023. **Spotlight Award**

[9] Avanika Narayan, Laurel Orr, Ines Chami, Simran Arora, and Christopher Ré
Can Foundation Models Wrangle Your Data?
Proceedings of the VLDB Endowment (PVLDB), 2022.

[10] Simran Arora, Patrick Lewis, Angela Fan, Jacob Kahn, and Christopher Ré
Reasoning over Public and Private Data in Retrieval-Based Systems
Transactions of the Association for Computational Linguistics (TACL), 2023.

[11] Simran Arora, Sen Wu, Enci Liu, and Christopher Re
Metadata shaping: A simple approach for knowledge-enhanced language models
Findings of the Association for Computational Linguistics (ACL), 2022.

[12] Laurel Orr, Megan Leszczynski, Simran Arora, Sen Wu, Neel Guha, Xiao Ling, and Christopher Re
Bootleg: Chasing the Tail with Self-Supervised Named Entity Disambiguation
Conference on Innovative Data Systems Research (CIDR), 2021.

[13] Simran Arora, Avner May, Jian Zhang, and Christopher Ré
Contextual Embeddings: When Are They Worth It?
Proceedings of the Association for Computational Linguistics (ACL), 2020.

[14] Qizhen Zhang, Akash Acharya, Hongzhi Chen, Simran Arora, Ang Chen, Vincent Liu, and Boon Thau Loo
Optimizing Declarative Graph Queries at Large Scale
Proceedings of the 2019 International Conference on Management of Data (SIGMOD), 2019.

[15] Edward Steager, Denise Wong, Jeremy Wang, Simran Arora, and Vijay Kumar
Control of multiple microrobots with multiscale magnetic field superposition
International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS), 2017.
Best Paper Runner Up

[16] B. P. Mason, M. Whittaker, J. Hemmer, Simran Arora, A. Harper, S. Alnemrat, A. McEachen, S. Helmy, J. Read de Alaniz, and J. P. Hooper
A temperature-mapping molecular sensor for polyurethane-based elastomers
Applied Physics Letters (APL), 2016.

Workshop Papers and Preprints

[17] Simran Arora, Aman Timalsina, Aaryan Singhal, Benjamin Spector, Sabri Eyuboglu, Xinyi Zhao, Ashish Rao, Atri Rudra, and Christopher Ré
Just read twice: closing the recall gap for recurrent language models, ICML ES-FoMo 2024.

[18] Michael Zhang, Aaryan Singhal, Benjamin Frederick Spector, Simran Arora, and Christopher Ré
Low-rank Linearization of Large Language Models, ICML ES-FoMo 2024.

[19] Jerry Liu, Jessica Grogan, Owen Dugan, Simran Arora, Atri Rudra, and Christopher Ré
Can Transformers Solve Least Squares to High Precision?, ICML ES-FoMo 2024.

[20] Sabri Eyuboglu, Dylan Zinsley, Jon Saad-Falcon, Simran Arora, Atri Rudra, James Zou, and Christopher Ré
Towards smaller language models via layer looping, ICML ES-FoMo 2024.

[21] Megha Srivastava, Simran Arora, and Dan Boneh
Optimistic Verifiable Training by Controlling Hardware Nondeterminism, ICML ES-FoMo 2024.

[22] Simran Arora and Christopher Ré
Can Foundation Models Help Us Achieve Perfect Secrecy?, AAAI PPAI Workshop, 2023.

Open Source Artifacts

Software Artifacts

- https://github.com/HazyResearch/prefix-linear-attention
- https://github.com/HazyResearch/based
- https://github.com/HazyResearch/zoology
- https://github.com/HazyResearch/ama_prompting
- https://github.com/HazyResearch/evaporate
- https://github.com/HazyResearch/m2
- https://github.com/HazyResearch/ThunderKittens
- https://github.com/facebookresearch/concurrentqa

Datasets and Pretrained Model Artifacts

- https://huggingface.co/collections/
- https://huggingface.co/togethercomputer
- https://huggingface.co/datasets/stanfordnlp/concurrentqa

Invited Talks

Understanding and Improving Efficient Language Models
Princeton University PLI Group (Princeton, NJ): Winter 2024
Cornell Tech (New York, NY): Winter 2024
Microsoft AI Research (Virtual): Winter 2024
56th Annual ACM Symposium on Theory of Computing (STOC) Workshop Keynote Speaker (Vancouver, Canada): Summer 2024

Building High Throughput Data Management Systems
NeurIPS TRL Workshop Keynote Speaker (New Orleans, LA): Winter 2023

Ask Me Anything: How are Foundation Models Changing the Way We Build Data Management Systems?
Snorkel Foundation Model Summit (Virtual): Winter 2023
Apple Machine Learning Research (Cupertino, CA): Winter 2023
ICLR Spotlight Presentation (Kigali, Rwanda): Fall 2023
Stanford CRFM Research Spotlight Talk (Stanford, CA): Fall 2023

Can Foundation Models Help Us Achieve Perfect Secrecy?
IBM AI Research (Virtual): Fall 2022
MIT CSS Seminar (Virtual): Spring 2023
Stanford HAI: AI and Society (Stanford, CA): Spring 2023
Oral at KnowledgeNLP-AAAI'23 (Washington DC): Winter 2023

Metadata Shaping: A Simple Approach for Knowledge-Enhanced Language Models
Facebook AI Research Reading Group (Virtual): Summer 2021
Spotlight at Stanford HAI Data-Centric AI Workshop (Virtual): Fall 2021

Contextual Embeddings: When are they worth it?
ACL Conference (Virtual): Summer 2020
Stanford DAWN Retreat (Virtual): Summer 2020

From M&T to Research!
UPenn M&TSI (Virtual): Summer 2020
UPenn M&TSI (Philadelphia, PA): Summer 2023

Teaching

Fall 2023	**Course Co-Creator and Co-Instructor**, *CS 229S: Systems for Machine Learning* Stanford University, 3-Unit Bachelor/Graduate course. Taught 110 students and delivered half the lectures.
Fall 2023	**Instructor** *CS: 528: Machine Learning Systems Seminar* Stanford University
Spring 2019	**Course Co-Creator**, *MCIT 595: Computer Systems* University of Pennsylvania
Fall 2018	**Course Assistant**, *CIS 380: Operating Systems* University of Pennsylvania
Fall 2017 and Spring 2018	**Course Assistant**, *CIS 160: Discrete Mathematics* University of Pennsylvania

Mentorship

2023-	Xinyi ("Jojo") Zhao (Stanford CS MS, In submission NeurIPS 2024 paper)
2024-	Ashish Rao (Stanford CS Undergrad/Coterm, In submission NeurIPS 2024 paper)
2024-	Jerry Liu (Stanford CS PhD, First author paper submitted to NeurIPS 2024 ES-FoMo)
2024	Jon Saad-Falcon (Stanford CS PhD Rotator, First author paper at ICML 2024)
2023	Soumya Chatterjee (Stanford CS MS, First author paper at SIGIR REML 2023)

2023	Andrew Hojel (Stanford CS Undergrad/Coterm, VLDB paper, now Member of the Technical Staff at Essential AI)
2021	Enci Liu (Stanford CS Undergrad/Coterm, ACL paper, now ML at Apple)

Service

Reviewer	ICML (Top Reviewer Award), NeurIPS, ACL, PPAI-AAAI, NeurIPS TRL, ICLR ME-FoMo, ICML ES-FoMo
2023	Co-organized workshop on Decentralized and Collaborative Learning at MLSys
Summer 2023	I organized the Stanford NLP Group weekly meetings
2022-2023	Stanford Center for Research on Foundation Models (CRFM) Organizing Committee
2019-2022	Undergrad Mentor in the Stanford Women in STEM Mentorship Program
2018-2019	Undergrad Mentor in the Penn Women in CS Mentorship Program

Last updated: July 23, 2024 *

*CV template by Neel Guha, Daniel Fu, and Christopher Morris.