

Jonathan Li

(510) 953 1453
✉ jonathanlingjieli@gmail.com
📄 [jonathanlli.github.io](https://github.com/jonathanlli.github.io)

Education

- May 2021 **Master of Science in Computer Science, Stanford University.**
GPA: 3.87/4.3
Research Assistant at Hazy Research, Shah Lab
TA for CS229 Machine Learning, CS230 Deep Learning
- May 2019 **Bachelor of Science in Computer Science with University Honors, Carnegie Mellon University.**
GPA: 3.68/4.0, *Dean's List Fall 2015, Spring 2016, Fall 2017*
Carnegie Scholarship
School of Computer Science College Honors
Honors Thesis: Investigating the effectiveness of deep reinforcement learning techniques on steering T-cell differentiation
TA for 15-150 Functional Programming, 15-381 Artificial Intelligence

Experience

- 2021-present **Principal Machine Learning Research Engineer, SambaNova Systems, Inc..**
- Summer 2020 **Software Engineering Intern, Google, Inc..**
- Summer 2017 **Software Engineering Intern, Xiaomi Inc..**

Preprints

- **Composition of Experts: A Modular Compound AI System Leveraging Large Language Models (2024)**
Swayambhoo Jain, Ravi Raju, Bo Li, Zoltan Csaki, [Jonathan Li](#), Kaizhao Liang, Guoyao Feng, Urmish Thakker, Anand Sampat, Raghu Prabhakar, Sumti Jairath
<https://arxiv.org/abs/2412.01868>
- **Climbing the WOL: Training for Cheaper Inference (2020)**
Zichang Liu, Zhaozhuo Xu, Alan Ji, [Jonathan Li](#), Beidi Chen, Anshumali Shrivastava
<https://arxiv.org/abs/2007.01230>

Publications

- **SnapStream: Efficient Long Sequence Decoding on Dataflow Accelerators**
ISC High Performance (2026)
Jonathan Li, Nasim Farahini, Evgenii Iuliugin, Magnus Vesterlund, Christian Häggström, Guangtao Wang, Shubhangi Upasani, Ayush Sachdeva, Rui Li, Faline Fu, Chen Wu, Ayesha Siddiqua, John Long, Tuowen Zhao, Matheen Musaddiq, Håkan Zeffner, Yun Du, Mingran Wang, Qinghua Li, Bo Li, Urmish Thakker, Raghu Prabhakar
<https://arxiv.org/pdf/2511.03092>
- **Synthetic Document Question Answering in Hungarian**
Vision Language Models For All Workshop @ CVPR (2025)
Jonathan Li, Zoltan Csaki, Nidhi Hiremath, Etash Guha, Fenglu Hong, Edward Ma, Urmish Thakker
<https://arxiv.org/abs/2505.23008>
- **LLMs Know What to Drop: Self-Attention Guided KV Cache Eviction for Efficient Long-Context Inference**
Workshop on Sparsity in LLMs @ ICLR (2025)
Guangtao Wang, Shubhangi Upasani, Chen Wu, Darshan Gandhi, Jonathan Li, Changran Hu, Bo Li, Urmish Thakker
<https://arxiv.org/abs/2503.08879>
- **Training Domain Draft Models for Speculative Decoding: Best Practices and Insights**
Workshop on Scalable Optimization for Efficient and Adaptive Foundation Models @ ICLR (2025)
Fenglu Hong, Ravi Raju, Jonathan Lingjie Li, Bo Li, Urmish Thakker, Avinash Ravichandran, Swayambhoo Jain, Changran Hu
<https://arxiv.org/abs/2503.07807>
- **Constructing Domain-Specific Evaluation Sets for LLM-as-a-Judge**
Best Paper Award First Workshop on Customizable NLP @ EMNLP (2024)
Ravi Raju, Swayambhoo Jain, Bo Li, Jonathan Li, Urmish Thakker
<https://arxiv.org/abs/2408.08808v2>
- **SambaLingo: Teaching Large Language Models New Languages**
Fourth Workshop on Multilingual Representation Learning @ EMNLP (2024)
Zoltan Csaki, Bo Li, Jonathan Li, Qiantong Xu, Pian Pawakapan, Leon Zhang, Yun Du, Hengyu Zhao, Changran Hu, Urmish Thakker
<https://arxiv.org/abs/2404.05829>
- **Pipemare: Asynchronous Pipeline Parallel DNN Training**
Fourth Conference on Machine Learning Systems (2021)
Bowen Yang, Jian Zhang, Jonathan Li, Christopher Ré, Christopher Aberger, Christopher De Sa
https://proceedings.mlsys.org/paper_files/paper/2021/file/9412531719be7ccf755c4ff98d0969dc-Paper.pdf

- **Mongoose: A Learnable LSH Framework for Efficient Neural Network Training**
Ninth International Conference on Learning Representations (2021)
Beidi Chen, Zichang Liu, Binghui Peng, Zhaozhuo Xu, Jonathan Lingjie Li, Tri Dao, Zhao Song, Anshumali Shrivastava, Christopher Ré
<https://openreview.net/forum?id=wWK7yXkULyh>
- **Multicopter UAV state prediction through multi-microphone side-channel fusion**
IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (2017)
Hendrik Vincent Kooops, Kashish Garg, Munsung Kim, Jonathan Li, Anja Volk, Franz Franchetti
<https://ieeexplore.ieee.org/abstract/document/8170401>