Shanda Li

✉ shandal@cs.cmu.edu
⌂ lithiumda.github.io
⦿ LithiumDA

Education

Aug 2022 – present
Machine Learning Department, School of Computer Science, Carnegie Mellon University,
Ph.D. student in Machine Learning. QPA: 4.33/4.33.
Research advisor: Prof. Yiming Yang.

Aug 2018 –Jul 2022
Turing Class, School of EECS, Peking University,
B.S. in Computer Science (Summa Cum Laude) with a minor in Mathematics. GPA: 3.78/4.00.
Research advisor: Prof. Liwei Wang and Prof. Di He.
Bachelor Thesis: Deep-Learning-Based Partial Differential Equation Solvers (*Top 10 Thesis in School of EECS*)

Recent Research Interests

Algorithmic reasoning: Reasoning with LLM; learning graph algorithms with GNN.

Principled understanding of LLM: Representational power, inductive bias, mechanistic interpretability.

Efficient scaling of Transformers: Context scaling and length generalization of Transformers.

Work Experiences

Jun 2023 – Aug 2023
Google Research (New York), *Research intern*, Host: Srinadh Bhojanapalli.
○ Researched on length generalization and long context Transformers.
○ Probed the long context learning capability of existing techniques on zero/few-shot and fintuning settings.
○ Proposed a new method which matches the long context performances of baselines with $0.36\times$ parameters.
○ Submitted the work to ICLR 2024 as the first author (with rating 8/6/6 on OpenReview).

Mar 2021 – Jun 2021
Microsoft Research Asia, *Research intern*, Host: Guolin Ke.
○ Researched on accelerating attention with relative positional encodings (RPE) for long sequences.
○ Designed efficient RPE-based attention with $O(n \log n)$ complexity in sequence lengths via Fast Fourier Transform.
○ Published the result in NeurIPS 2021 as a co-first author.

Publications

(* denotes equal contribution)

[1] **Stable, Fast and Accurate: Kernelized Attention with Relative Positional Encoding**, *NeurIPS 2021*,
Shengjie Luo*, **Shanda Li***, Tianle Cai, Di He, Dinglan Peng, Shuxin Zheng, Guolin Ke, Liwei Wang, Tie-Yan Liu .

[2] **Your Transformer May Not be as Powerful as You Expect**, *NeurIPS 2022*,
Shengjie Luo*, **Shanda Li***, Shuxin Zheng, Tie-Yan Liu, Liwei Wang, Di He.

[3] **Is L^2 Physics-Informed Loss Always Suitable for Training Physics-Informed Neural Network?**, *NeurIPS 2022*, Chuwei Wang*, **Shanda Li***, Di He, Liwei Wang.

[4] **Learning Physics-Informed Neural Networks without Stacked Back-propagation**, *AISTATS 2023*,
Di He, **Shanda Li**, Wenlei Shi, Xiaotian Gao, Jia Zhang, Jiang Bian, Liwei Wang, Tie-Yan Liu .

[5] **Functional Interpolation for Relative Positions Improves Long Context Transformers**, *ArXiv Preprint*,
Shanda Li, Chong You, Guru Guruganesh, Joshua Ainslie, Santiago Ontanon, Manzil Zaheer, Sumit Sanghai, Yiming Yang, Sanjiv Kumar, Srinadh Bhojanapalli.

[6] **Can Vision Transformers Perform Convolution?**, *ArXiv Preprint*,
Shanda Li, Xiangning Chen, Di He, Cho-Jui Hsieh.

[7] **Learning a Fourier Transform for Linear Relative Positional Encodings in Transformers**, *ArXiv Preprint*,
Krzysztof Choromanski*, **Shanda Li***, Valerii Likhosherstov, Kumar Avinava Dubey, Shengjie Luo, Di He, Yiming Yang,
Tamas Sarlos, Thomas Weingarten, Adrian Weller.

Selected Awards and Honors

Jun 2022	**Excellent College Graduate in Beijing**, *Top 1%*,	Beijing Municipal Commission of Education.
Jun 2022	**Top 10 Bachelor Thesis**,	School of EECS, Peking University.
Nov 2021	**SenseTime Scholarship**, *30 undergraduates per year in the field of AI*,	SenseTime.
Nov 2020	**First Prize**, *National University Mathematical Contest*,	Chinese Mathematical Society.
Sep 2017	**First Prize**, *31st Chinese Chemistry Olympiad (Preliminary)*.	

Invited Talks

Stable, Fast and Accurate: Kernelized Attention with Relative Positional Encoding.
- Mini Research Symposium of CFCS and Turing Class, Peking University — Dec 2021

Your Transformer May Not be as Powerful as You Expect.
- International Joint Conference on Theoretical Computer Science — Aug 2022

Is L^2 Physics-Informed Loss Always Suitable for Training Physics-Informed Neural Network?.
- Turing Student Research Forum, Peking University — Jun 2022
- Machine Learning+X Seminar, Brown University — Oct 2022

Professional Service

Conference Reviewer: ICML 2022, 2023; NeurIPS 2022, 2023; LOG 2023; ICLR 2024; AISTATS 2024.
Workshop Reviewer: M3L@NeurIPS 2023.
Teaching Assistant: Spring 2022, Probability and Statistics (A), Peking University.

Skills

Programming: Python (Pytorch, Jax), C/C++, LATEX.
Languages: Chinese, native speaker; English, proficient (TOEFL 108/120, Speaking 26/30).