

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: 4.26/4.33.
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. .
Bachelor Thesis: Deep-Learning-Based Partial Differential Equation Solvers (*Top 10 Thesis in School of EECS*)

Recent Research Interests

ML for math and science: Mathematical reasoning, code generation, differential equation solving.

Principled scaling of LLMs: Inference scaling laws, context scaling of LLMs, etc.

Selected Publications & Manuscripts (* denotes equal contribution)

Inference scaling laws and mathematical reasoning of LLMs

- [1] **Inference Scaling Laws: An Empirical Analysis of Compute-Optimal Inference for Problem-Solving with Language Models**, [arXiv](#), Yangzhen Wu, Zhiqing Sun, **Shanda Li**, Sean Welleck, Yiming Yang .
- [2] **CMU-MATH Team's Innovative Approach Secures 2nd Place at the AIMO Prize**, [CMU ML Blog](#), Yangzhen Wu, Zhiqing Sun, **Shanda Li**, Sean Welleck, Yiming Yang .

Long context Transformers

- [3] **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 .
- [4] **Learning a Fourier Transform for Linear Relative Positional Encodings in Transformers**, [AISTATS 2024](#), Krzysztof Choromanski*, **Shanda Li***, Valerii Likhoshesterov, Kumar Avinava Dubey, Shengjie Luo, Di He, Yiming Yang, Tamas Sarlos, Thomas Weingarten, Adrian Weller.
- [5] **Functional Interpolation for Relative Positions Improves Long Context Transformers**, [ICLR 2024](#), **Shanda Li**, Chong You, Guru Guruganesh, Joshua Ainslie, Santiago Ontanon, Manzil Zaheer, Sumit Sanghai, Yiming Yang, Sanjiv Kumar, Srinadh Bhojanapalli.

Deep-learning-based partial differential equation solvers

- [6] **Is L^2 Physics-Informed Loss Always Suitable for Training Physics-Informed Neural Network?**, [NeurIPS 2022](#), Chuwei Wang*, **Shanda Li***, Di He, Liwei Wang.
- [7] **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 .

Selected Awards and Honors

- Jul 2024 **Rank 2/1161 in the first progress prize of Artificial Intelligence Mathematical Olympiad.**
- 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 China in the field of AI, SenseTime.

Work Experiences

- May 2024 **Google Research (New York)**, *Research intern*, Host: [Nikunj Saunshi](#).
– Aug 2024
 - Researched on looped Transformers for mathematical reasoning with language models.
 - Explored techniques for pretraining looped Transformers or finetuning existing standard models into looped ones.
- Jun 2023 **Google Research (New York)**, *Research intern*, Host: [Srinadh Bhojanapalli](#).
– Aug 2023
 - Researched on length generalization and long context Transformers.
 - Proposed a new method which matches the long-context performances of baselines with $0.36\times$ parameters.
 - Published the work in ICLR 2024 as the first author.
- Mar 2021 **Microsoft Research Asia**, *Research intern*, Host: [Guolin Ke](#).
– Jun 2021
 - 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 work in NeurIPS 2021 as a co-first author.

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

Inference Scaling Law of Large Language Models and Second-Prize Winning Solution of AIMO.

- International Seminar on Foundational Artificial Intelligence (FAI-Seminar) Aug 2024
- NLP reading group, MiniMax Aug 2024

Professional Service

Reviewer.

- Conference: ICML 2022-2024; NeurIPS 2022-2024; LoG 2023-2024; ICLR 2024-2025; AISTATS 2024-2025.
- Workshop: M3L@NeurIPS 2023, 2024; BGPT@ICLR 2024; MATH-AI@NeurIPS 2024

Teaching Assistant.

- Probability and Statistics (A), Peking University Spring 2022
- Advanced Introduction to Machine Learning (10-715), Carnegie Mellon University Fall 2024

Skills

Programming: Python (Pytorch, Jax), C/C++, \LaTeX .

Languages: Chinese, native speaker; English, proficient (TOEFL 108/120, Speaking 26/30).