TIANYAO SHI

Purdue University, West Lafayette • (765) 123-4567 • shi0x0@purdue.edu • linkedin.com/in/tianyao-shi-507967290

RESEARCH INTEREST

My research interests span system, architecture, and machine learning. I am currently working on sustainable computing and LLM serving optimization, with a special focus on understanding and reducing the carbon impact of LLM inference.

EDUCATION

- **Purdue University, PhD Student in Electrical and Computer Engineering** 08/2024 – Now
  Advisor: Professor Yi Ding GPA: 4.0/4.0
- **Shanghai Jiao Tong University, Master in Electronic and Information Engineering** 09/2021 – 03/2024
  Advisor: Professor Xiaofeng Gao GPA: 3.69/4.0 Major GPA: 3.9/4.0
- **Shanghai Jiao Tong University, Bachelor in Computer Science** 09/2017 – 06/2021
  GPA: 88.73/100 (3.79/4.30) Major GPA: 89.65/100 (3.88/4.30)

SKILLS

- Programming languages: C/C++, Python, Linux Shell, SQL
- Software: Pytorch, Scikit-learn, XGBoost, Matplotlib, Pandas, vLLM, TensorRT-LLM, NVML, LaTeX, Tikz, Git
- Soft: Adaptability, Leadership, Technical documenting, Academic presenting, Cross-department communication

RESEARCH AND WORK EXPERIENCE

- **Sustainable LLM serving** 08/2024 – Now
  Graduate Research Assistant @ Purdue University, West Lafayette
  - Proposed a SLO-aware LLM serving framework, GreenLLM, to minimize carbon emissions by using old GPUs.
  - Wrote a profiler as a vLLM plug-in using RESTful APIs in 1.7k LOC using Python and Linux Shell.
  - Reduced the carbon footprint by up to 40.6% compared to the standard serving scheme.
- **Enterprise Customers' IT Budgets Prediction for Cloud Services**, 08/2022 – 09/2023
  Algorithmic Engineer (Internship) @ Company X, Hangzhou, China.
  - Formalized a new problem to predict enterprise customers' IT budgets for public cloud services via observed consumption records. Cooperate with salespersons to launch targeted campaigns at high-value customers.
  - Devised and implemented a two-stage framework, BSA-DaMaM, to address the coupling of high feature-missing ratio and heterogeneity in real-world data in 3k Python code.
  - Deployed the framework in production cluster, identifying $5M+ potential sales for the sales department.
- **QoS Prediction in Public Cloud** 01/2021 – 07/2022
  Algorithmic and System Engineer (Internship) @ Company X, Beijing, China.
  - Led a 10-men team, built and published Alioth-dataset through 400+ VM co-location experiments.
  - Designed Alioth, an open-source framework in Python that leverages explainable machine learning to estimate application performance degradation and detect co-location interference in public clouds with 94.71% accuracy.
- **Collaborative Recommender Systems** 12/2019 – 01/2021
  Undergraduate student researcher @ Shanghai Jiao Tong University, Shanghai, China.
  - Contributed to the visualization and writing of two papers on SIGIR and DASFAA 2021.

HONORS AND AWARDS

- B-Class Excellent Scholarship of Shanghai Jiao Tong University (twice, top 10%) 11/2018，11/2020
- Second Prize of LCCUP'20 Team Coding Contest (Ranked 95/1740) 10/2020
- Outstanding Graduate of Shanghai Jiao Tong University 06/2021
- Excellent Graduate Student Scholarship of Shanghai Jiao Tong University (6/196) 11/2023

SELECTED PUBLICATIONS

1. **Tianyao Shi**, Yanran Wu, Sihang Liu, Yi Ding, GreenLLM: Disaggregating Large Language Model Serving on Heterogeneous GPUs for Lower Carbon Emissions, arXiv preprint, arXiv:2412.20322, 2024.
2. **Tianyao Shi**, Yunlong Cheng, Zhipeng Bian, Xiaofeng Gao, Zhenli Sheng, Predicting Enterprise Customers' IT Budgets for Cloud Services, Database Systems for Advanced Applications (**DASFAA**) 2025, Accepted.
3. **Tianyao Shi**, Yingxuan Yang, Yunlong Cheng, Xiaofeng Gao, Zhen Fang, Yongqiang Yang, Alioth: A Machine Learning Based Interference-Aware Performance Monitor for Multi-Tenancy Applications in Public Cloud, the 37th IEEE International Parallel & Distributed Processing Symposium (**IPDPS**), pp. 908-917, 2023.
4. Xuehan Sun, **Tianyao Shi**, Xiaofeng Gao, Yanrong Kang, Guihai Chen, FORM: Following the Online Regularized Meta-Leader for Cold-Start Recommendation, International ACM **SIGIR** Conference, pp. 1177-1186, 2021.

ACADEMIC SERVICES

- Reviewer of IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 12/2024
- External Reviewer of IEEE International Conference on Data Mining (ICDM) 10/2023