Xiang Li

+1 765 637 5476 ▪ li2068@purdue.edu ▪ Google Scholar ▪ Linkedin Profile

CAREER SUMMARY

Ph.D. student specializing in the end-to-end optimization of Large Language Model applications. Expertise in designing high-performance RAG pipelines, from novel memory retrieval architectures to low-latency LLM inference. Focused on bridging the gap between complex model architectures and efficient, secure, scalable production deployments.

EDUCATION

Purdue University, West Lafayette — 2021–2027 (expected)

- Ph.D. in Computer Engineering. GPA. 3.87
  - Advisor: Saurabh Bagchi
  - Research Area: System for Machine Learning, Efficient LLM Serving

Purdue university, West Lafayette — 2015–2019

- B.S. Computer Engineering. GPA. 3.89

WORK EXPERIENCE

ML Algorithm Intern- Futurewei Technology Inc. San Jose, CA — Aug 2025 - Dec 2025

- Developed a novel memory architecture utilizing Graph Neural Networks to address the limited context window and degraded attention result in long-horizon embodied task planning.
- Designed a dynamic experience graph mechanism to consolidate past agentic experiences with current observations, facilitating more robust state representation and decision-making.

ML Research Intern- Houston Methodist Research Institute. Houston, TX — Jun 2023 – Aug 2023

- Built a 3D group equivaraint CNN-based segmentation model for non-contrast CT brain scans, reaching 87% soft tissue segmentation accuracy and reducing diagnostic latency for ischemic stroke detection.
- Partnered with physicians to translate AI models into clinical pipelines, accelerating imaging-based diagnosis.

Control Software Engineer- Cummins Inc. Columbus, IN — Jun 2019 – Jul 2021

- Engineered dual-encryption bootloader security for Cummins ECM, eliminating unauthorized firmware tampering across 250k+ vehicles in production.
- Collaborated with cross-functional teams to deliver model-based control software for critical automotive components, ensuring reliability in high-volume manufacturing.
- Tuned and validated 30+ OBD parameters for RAM truck compliance with U.S. emission standards, strengthening regulatory approval and system robustness.

Graduate Teaching Assistant-Purdue University. West Lafayette, IN — Aug 2021 – July 2025

- Teaching assistant for Purdue ECE embedded software senior design course ECE 47700.
- Teaching assistant for Purdue ECE embedded system design course ECE 36200.
- Teaching assistant for Purdue First Year Engineering course ENG 131,132.

Graduate Research Assistant-Purdue University. West Lafayette, IN — Summer 2022, Summer 2024

- Engineered a unified scheduling system for heterogeneous edge devices and hybrid networks, optimizing resource allocation across diverse computing nodes.
- Established a testbed integrating Private 5G and Wi-Fi 6 to evaluate network performance and reliability under industrial constraints.

RECENT PROJECTS

Purdue ECE Senior Design Website [link]

- Developed an end-to-end web application within the React framework for the instruction team of the senior design course.
- Engineered a modular architecture to synthesize diverse course needs into a unified system, including journaling, inventory, archiving, etc.

TECHNICAL SKILLS

Languages. C, Python, R, JS ▪ **Machine Learning.** PyTorch, TensorFlow, LLM Inference Optimization ▪ **H/W Platforms.** Jetson AGX, Jetson NX, Jetson Nano, Arduino Uno, Raspberry Pi 4 (baremetal) ▪ **Web Technologies.** socketio, Wireshark, React

COURSEWORK

Graduate level. Estimation Theory, Random Process, Computational Modeling, Deep Learning, Reinforcement Learning Theory

PUBLICATIONS

[NIPS@DL4C] Deep-Reproducer: From Paper Understanding to Code Generation
Pengcheng Chen, Ning Yan, Zihan Zhao, Yixiao Lin, Huaibo Chen, Yue Hu, Qinbo Bai, **Xiang Li**, Masood Mortazavi (2025)

[arXiv] Ascendra: Dynamic Request Prioritization for Efficient LLM Serving.
Xiang Li*, Azam Ikram*, Sameh Elnikety, Saurabh Bagchi. arXiv preprint (2025)

[ESOC] Enhanced Brain Tissue Segmentation In Non-Contrast CT For Ischemic Stroke Diagnosis
Xiang Li, Saurabh Bagchi, John Volpi, Stephen Wong, Kelvin Wong (2025)

[arXiv] HopTrack: A Real-time Multi-Object Tracking System for Embedded Devices
Xiang Li, Cheng Chen, Yuan-yao Lou, Mustafa Abdallah, Kwang Taik Kim, Saurabh Bagchi. arXiv preprint (2024)

[arXiv] Dynamic DAG-application scheduling for multi-tier edge computing in heterogeneous networks
Xiang Li, Mustafa Abdallah, Yuan-Yao Lou, Mung Chiang, Kwang Taik Kim, Saurabh Bagchi. arXiv preprint (2023)

[SRDS] DAG-based task orchestration for edge computing
Xiang Li, Mustafa Abdallah, Shikhar Suryavansh, Mung Chiang, Kwang Taik Kim, Saurabh Bagchi (2022)

ACHIEVEMENTS

Tinker Research Grant, Thinking Machine Lab. 2025
- Awarded $5000 credits for research project uses Tinker.

Honor Society Eta Kappa Nu, Purdue University. 2021
- Selected for being in the top 10% of graduate students.

Eli Shay Scholarship, Purdue University 2017

SERVICE

Reviewer, BHI 25' 2025
Student Administrator, USENIX ATC 24' 2024
Artifact review committee, OSDI 24' 2024
Artifact review committee, USENIX ATC 24' 2024
Reviewer, IEEE Networking Letter 2023