Yuxuan Xia

xyxchengdu@gmail.com|+1 (805) 895-6836|xyx-creator.github.io|github.com/xyx-creator

Education

University of California, Santa Barbara — **Santa Barbara, California**
*Master of Electrical and Computer Engineering (**Computer Engineering**)* — *Sep. 2024 – Mar. 2026 (Expected)*

- GPA: 3.76/4.00
- Relevant Courses: Matrix, Tensor, Convex Optimization, Robust Machine Learning, LLMs, AI for Science

Southwest Jiaotong University & University of Leeds — **Chengdu, China / Leeds, UK**
Bachelor of Science in Computer Science — *Sep. 2020 – Jul. 2024*

- Average Score: 88/100 (Southwest Jiaotong University)
- Graduated with First Class Honours (University of Leeds)

Research Interests

Multimodal Learning, Large Language Models, Trustworthy AI, Graph Representation Learning

Publications

Preprints & Under Review

- **Y. Xia**, S. Wang, and P. Li. *SDCD: Structure-Disrupted Contrastive Decoding for Mitigating Hallucinations in Large Vision-Language Models.* Submitted to the **Association for Computational Linguistics (ACL) 2026**. arXiv:2601.03500.
- Z. Wang, B. Xu, **Y. Xia**, and P. Li. *VEGAS: Mitigating Hallucinations in Large Vision-Language Models via Vision-Encoder Attention Guided Adaptive Steering.* Submitted to the **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2026**. arXiv:2512.12089.

Published Conference Papers

- **Y. Xia***, H. Wu*, and G. Chen*. (*co-first authors*) *Joint Median Channel Prior and Edge Enhancement for Single Image Dehazing.* In *Proceedings of the 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, Haikou, China, 2023, pp. 326–332. DOI: 10.1109/PRAI59366.2023.10331999.
- Y. Xin, Z. Zhou, **Y. Xia**. *Scene Separation and Data Selection: Temporal Segmentation Algorithm for Real-Time Video Stream Analysis.* In *Proceedings of the IJCAI-ECAI 2022 Workshop on Spatio-Temporal Reasoning and Learning (STRL)*, Vienna, Austria, July 2022. doi:10.48550/arXiv.2308.00210.

Work Experience

Peng Li's Lab — **Santa Barbara, California**
Student Researcher — *Mar. 2025 – Present*

- Spearheaded research on **Trustworthy AI**, focusing on mitigating object hallucinations in Large Vision-Language Models (LVLMs) through both internal attention intervention and external decoding strategies.
- Proposed **VEGAS**, an adaptive steering mechanism that calibrates visual attention weights during inference. Achieved a **20-point CHAIR improvement** by suppressing linguistic priors.
- Developed **SDCD**, a contrastive decoding framework that isolates structural biases via spatially shuffled negative views, achieving a **6% F1 score improvement** on the POPE benchmark.
- Currently conducting research on **AI Agents**.

Handshake AI — **Remote**
AI Data Annotator (CS Fellow) — *Sep. 2025 – Nov. 2025*

- Developed and curated **multimodal reasoning datasets** for **LLM instruction tuning**, focusing on graph problems. Designed and integrated graph-based visuals to align with textual reasoning tasks.
- Authored **step-by-step reasoning annotations** that model human problem-solving processes, illustrating intermediate states, algorithmic invariants, and detailed solution logic for each problem.
- Improved LLMs' **algorithmic reasoning** performance through expert validation and dataset refinement.

Aotu Technology Co., Ltd. **Chengdu, China**
Algorithm Engineer *Apr. 2024 – Jul. 2024*

- Designed and developed the **TDOA** (Time Difference of Arrival) localization algorithm, which precisely located signal sources in urban environments using a multi-sensor network.
- Used MATLAB for comprehensive simulation validation and employed C++ for field test implementation, achieving critical technological breakthroughs from theory to practical application.
- Ensured 5 m localization accuracy in simulations for sensor distances over 2 km, and achieved 23 m drone signal localization accuracy in field tests.

Projects

Graph-Augmented Large Language Models *Sep. 2025 – Dec. 2025*

- Developed a LLaVA-style **graph-augmented LLM** by integrating a GraphCLIP encoder with a 2-layer projection module to inject graph embeddings as **soft prompts** into a frozen LLM.
- Enabled **node classification** and **open-ended graph QA**, improving the model's structured reasoning capability.
- Achieved comparable performance on GLBench zero-shot node classification using only 10k training pairs, trailing the SOTA model (trained on 360k pairs) by only **3%**.

Intelligent Healthcare AI Agent Project *Apr. 2025 – Jul. 2025*

- Designed and implemented AI agents using LangChain for healthcare, integrating local large models via Ollama and remote APIs (OpenAI, BaiLian) for versatile medical intelligence.
- Developed multi-turn conversation memory management with MongoDB persistence.
- Constructed a retrieval-augmented generation pipeline by vectorizing medical documents, creating custom data loaders and splitters, and setting up vector-based search for instant clinical knowledge access.
- Added custom tool functions for mathematical operations to the AI agent for enhanced functionality.

ZO-Optimized Classification Neural Network *Mar. 2025 – May 2025*

- Implemented **zero-order optimization** for neural network, enabling effective training of gradient-agnostic models.
- Applied feature selection, feature reuse, and iterative pruning to accelerate training and reduce memory usage.
- Demonstrated only a **1%** or less drop in classification accuracy while achieving a **90%** reduction in memory consumption compared to the Adam optimizer.

Optimization for Tensorized Transformer Training *Oct. 2024 – Dec. 2024*

- Refined the traditional transformer architecture by integrating a **Block-Term Tensor Decomposition** technique, replacing the multi-head attention layer with a more efficient multi-linear attention mechanism.
- Used custom CUDA operators outperform Torch's tensor multiplication, achieving **8× performance** increase.
- Enhanced training efficiency by over 30%, maintaining comparable perplexity levels for language modeling.

Skills

Programming Languages: JavaScript, Java, Python, C/C++, MATLAB, HTML, CSS, Node.js

Machine Learning Engineering: CUDA, PyTorch, NumPy, Matplotlib, PyTorch Geometric (PyG), Transformers, Scikit-learn, OpenCV, Pandas

Software Engineering: Vue.js, Bootstrap, React.js, Django, Spring Boot, Qt (C++ and PyQt)

Natural Languages: Chinese (native), English (fluent)

Awards & Scholarships

- Provincial 1st Prize in the 16th Chinese Collegiate Computing Competition
- 2nd Class Scholarship, Southwest Jiaotong University