Yibin (Leon) Liu

Shenyang, China | yibin.leon.liu@outlook.com | 13830039815 | Website | Github | LinkedIn

Education

Northeastern University, Shenyang, China *Sept 2022 – June 2026*
Bachelor of Artificial Intelligence, College of Information Science and Engineering

- Achievements: 86.5/100
- Programs: 985, 211, Double 1st-Class

Global Innovation Exchange - UW & Tsinghua *Jul 2024 – Oct 2024*
Access Computing Summer Program, AI & HCI
A collaboration between the University of Washington, Tsinghua University, and Microsoft, focused on AI and HCI innovation. Full scholarship awarded.

Selected Courses & GPA:
Fundamentals of Machine Learning (99), Digital Signal Processing (96), Computing Theory (95), Discrete Mathematics (91), Image Processing and Computer Vision (93), Intelligent Optimization Algorithm (90), Fundamentals of Programming C (93), Object-Oriented Programming C++ (90), Scientific Computing with MATLAB (92), The Mathematical Basis of Artificial Intelligence (87), Data Structures (86), Algorithm Design and Analysis (86)

Research Interests

My research interests focus on grounding language in action and perception, specifically how computational models can learn from real-world data, uncover the underlying causal relationships, and use this understanding to perform reasoning and decision-making. Ultimately, I aim to explore how these models interact more effectively with the real world and generate positive impacts. Building on this foundation, I am also keen to investigate how these technologies can be applied to control robotic systems and collaborate with humans on specific tasks.

Research Experience

Research Assistant at Pervasive HCI Lab, Tsinghua University *Jun 2024 – Jan 2025*
Advisor: Postdoc. Nan Gao, Associate Professor Chun Yu

- Conducted research in family education using LLM and HCI to infer behaviors and mental states, promoting self-awareness and well-being.

Research Intern at NEUIR, Northeastern University *Oct 2023 – Apr 2024*
Advisor: Associate Professor Zhenghao Liu

- Research on Faithful Reasoning of LLM, RAG, QA, and NLP tasks, focusing on enhancing model reasoning capabilities.

Publications

Self-Guide: A LLM Reasoning Enhancement Method Based on Self-Guided Planning. (Journal of Chinese Information Processing) 2024

Yibin Liu, Zhenghao Liu, Yukun Yan, Shi Yu, Shuo Wang, Liner Yang, Yu Gu, Ge Yu, Huimin Chen

The Homework Wars: Exploring Emotions, Behaviours, and Conflicts in Parent-Child Homework Interactions (Under Review of IMWUT 2025) Under Review

Nan Gao, ***Yibin Liu***, Xin Tang, Yanyan Liu, Chun Yu, Yun Huang, Xuhai "Orson" Xu, Jun Wei, Yuanchun Shi

Academic Activities

- **Academic Service:** Reviewer for Chinese CHI 2024
- **Talks:** 2024.08, "Retrieval-Augmented Generation Modeling" for Mingtong Weilai (Beijing) Digital Health Science & Technology Research Institute.

Awards

- 2024.05 Finalist in Mathematical Contest in Modeling (MCM/ICM), **Top 2% of 10,387 teams**
- 2023.10 National Level Third Prize in RoboCup China Competition, Simulation 3D League
- 2023.10 National Level Second Prize in FIRA SimuroSot China Competition
- 2023.11 Future Technology Taihu Scholarship
- 2023.09 Excellent Student Scholarship at Northeastern University

Projects

Autoregressive Language Model *github.com/10-OASIS-01*

- Developed a beginner-friendly autoregressive Transformer-based language model, covering all steps from data processing to model training, evaluation, and inference. Employed manual tokenization with regular expressions and Byte Pair Encoding (BPE). Optimized for local GPU training and integrated with Wandb for real-time experiment tracking.
- Tools Used: Python, PyTorch, Wandb, Hugging Face

Bencao RAG Medical Intelligent Assistant *github.com/10-OASIS-01*

- Developed a medical knowledge question-answering system that integrates context awareness, internet access, knowledge graphs, and retrieval-augmented generation (RAG) to provide accurate and personalized medical information. Built a user-friendly interface with Streamlit for seamless interaction. Features include basic and context-aware medical Q&A, internet-enhanced answers, document-based advice, and a knowledge graph for structured medical data.
- Tools Used: Python, RAG, LLM, LangChain, Streamlit, Neo4j, Knowledge Graph

BPE Tokenizer *github.com/10-OASIS-01*

- Developed a flexible and minimalistic tokenizer based on the Byte Pair Encoding (BPE) algorithm, optimized for both small and large-scale datasets. Designed the tokenizer to closely approximate the behavior of models like GPT-4, suitable for processing datasets such as OpenWebText or WikiText-103.
- Tools Used: Python, NumPy

Lane Detection Pipeline *github.com/10-OASIS-01*

- Developed a pipeline to detect lane lines on the road using computer vision techniques applied to video input. Outputs include annotated video frames with detected lanes, curvature, and vehicle offset information for autonomous driving systems.
- Tools Used: Python, OpenCV, NumPy

Technologies

Languages: Python, C++, C, HTML/CSS, JavaScript, SQL, MATLAB/Simulink, LaTeX

Technologies: PyTorch, Hugging Face, scikit-learn, ROS, OpenCV, NumPy, Git, RAG, Linux, SLAM