

Yibin (Léon) Liu

📍 Shenyang, China ✉ kelvin.lau.cn@outlook.com 📞 13830039815 🌐 Website 🐙 Github in LinkedIn

Education

Northeastern University, Shenyang, China *Sept 2022 – June 2026*

Bachelor of Artificial Intelligence, College of Information Science and Engineering

- Achievements: 85.6/100 (Junior year)
- Programs: 985, 211, Double 1st-Class

Global Innovation Exchange - University of Washington *Jul 2024 – Oct 2024*

Access Computing Summer Program, AI & HCI

Awarded ¥6400 scholarship to participate in this program, a collaboration between the University of Washington, Tsinghua University, and Microsoft, focused on AI and HCI innovation.

Research Interests

My research interests lie in finding a harmonious balance between **Language data**, **Interaction**, and **Society**. By applying NLP and DL to this interdisciplinary area, my goal is to develop Autonomous Systems that are intelligent, human-centered, and socially responsible.

Research Experience

Research Assistant at Pervasive HCI Lab, Tsinghua University *Sept 2023 – Present*

Advisor: Dr. 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 *Sept 2022 – Sept 2023*

Advisor: Associate Professor Zhenghao Liu

- Research on Faithful Reasoning of LLM, RAG-sft, 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

ActiveRAG: Revealing the Treasures of Knowledge via Active Learning (arXiv) 2024

Zhipeng Xu, Zhenghao Liu*, Yibin Liu, Chenyan Xiong, Yukun Yan, Shuo Wang, Shi Yu, Zhiyuan Liu, Ge Yu*

Academic Activities

- **Academic Service:** Reviewer for Chinese CHI 2024
- **Talks:** 2024.08, “Retrieval-Augmented Generation Modeling” for Mingtong Weilai (Beijing)

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

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

Character-Level Language Modeling

github.com/10-OASIS-01

- Implemented character-level language models using Transformer, RNN, and GRU architectures, trained on the tiny Shakespeare dataset. Focused on predicting the next character in a sequence to generate text stylistically resembling Shakespeare's writing.
- Tools Used: Python, PyTorch

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, Java, JavaScript, SQL, MATLAB

Technologies: PyTorch, Hugging Face, scikit-learn, OpenCV, NumPy, Git, MySQL, Linux