Yingjie (Ellen) Zhang

github.com/yuran986 | yingjie.zhang@duke.edu | (919) 423-7398 | linkedin.com/in/yingjie-zhang-8405b4359

Education

Duke University, Master of Science in Computer Science Aug. 2025 – May. 2027

- Coursework: Theory and Algorithms for Machine Learning, Introduction to Deep Learning, Data Science

Shanghai Jiao Tong University, Bachelor of Engineering in Computer Science and Technology Sep. 2021 – Jun. 2025

- GPA: 87.41/100; Coursework: Artificial Intelligence, Thinking and Approach of Programming (C++)

Internship

Software Engineer, Alibaba Group (Amap) – Xiamen, China May. - Aug. 2025

- Built a **multi-agent** map theme generation system on the **Autogen** framework with **Swarm**-based orchestration, leveraging **multimodal LLMs** (e.g., Gemini, GPT-4), achieving **3rd** place in the Agent Show competition.
- Designed and implemented a high-level intent recognition agent to accurately capture user intent, incorporating advanced **prompt engineering** techniques applied across the agent ecosystem.
- Developed a frontend color-palette workflow embedded in the Theme Color Generation agent via **Autogen Studio** (**TypeScript**/**React**).

Machine Learning Engineer, Meituan – Shanghai, China Jun. - Sept. 2024

- Developed and implemented a multilingual text matching system with **SQL**-based data extraction, focusing on detecting merchant self-evaluation patterns across review platforms with a user base of **700M+**.
- Compressed the 120B **Longcat-Prime** model into 7B **Qwen2-7B-Instruct** by building an efficient **knowledge distillation** pipeline, leveraging **supervised fine-tuning** and custom labeling for optimized transfer.
- Improved matching precision by **6.14** percentage points (from 91.2% to 97.34%) over the existing regex-based solution and maintained inference efficiency to support **12k+ daily model calls**.
- Coordinated and guided collaboration between the NLP and Friday LLM platform teams, driving smooth project integration and on-time delivery.

Projects & Research Experiences

Retriever-Augmented Generation (RAG) Web Application Dec. 2025 - Present

- Designing an end-to-end RAG system that dynamically indexes user-uploaded documents for Q&A.
- Implementing FastAPI backend for file ingestion, document parsing (PDF/Text), and real-time vector embeddings using LangChain + Chroma.

Review of Large-Scale EEG Models for Meditation State Recognition – ICASSP 2026 (Under Review) Mar. – Aug. 2025

- Reviewed **large-scale EEG models** (LEMs) along with their **self-supervised pretraining strategies**.
- Evaluated five representative **Transformer**-based LEMs (EEGPT, BENDR, BIOT, LaBraM, Gram) for meditation recognition under various experimental settings, revealing model–task alignment effects.
- Analyzed potential performance degradation caused by continued pre-training on SEED emotion dataset.

Meditation EEG Signal Analysis with Self-Supervised Learning – BIBM 2024 Mar. – Sep. 2024

- Developed a robust experimental framework for EEG data acquisition.
- Proposed and implemented the **Multi-view Spectral-Spatial-Temporal Masked Autoencoder** (MV-SSTMA), yielding substantial gains over 7 baselines (+**6.49**% F1 Subject-Dependent, +**5.86**% Cross-Subject).

Skills

Languages: Python (Proficient; PyTorch, TensorFlow, scikit-learn), C/C++ (Intermediate), MATLAB (Familiar)

Tools: SQL, Linux, Git, Jupyter Notebook, LaTeX, Markdown, Docker