Zhilin(Hadlay) Zhang

hadlay.zhang@gmail.com linkedin.com/in/Zhilin-Zhang
github.com/Hadlay-Zhang hadlay-zhang.github.io

Research Interests

My research focuses on **Multimodal Learning** and **LLM-based Multi-Agent Systems**. I aim to develop collaborative multi-agent systems that enhance multimodal reasoning capabilities to decompose and solve complex real-world tasks, such as **scientific workflows** and **medicine**.

Education

New York University, New York, USA **Sep 2024 – Present**

Master of Science in Computer Science, Overall GPA: 3.852/4.0

Courses: *Machine Learning (A), Operating System (A), Programming Languages (A), Intro to Java (A), Algorithmic Machine Learning and Data Science (A-), Deep Learning (A-), Computer Vision (A)*

Tongji University, Shanghai, China **Sep 2020 – Jul 2024**

Bachelor of Engineering in Computer Science

Courses: *Data Structure (A), Software Engineering (A), Data Mining (A), Machine Learning (A), Computer Vision (A), Natural Language Processing (A)*

Honors and Awards: *Scholarship for Undergraduate Freshmen (2020-2021); Undergraduate Third-Class Scholarship (2021-2022, 2023-2024)*

Publications and Preprints

† denotes equal contribution[1]. **C** = Conference, **P** = Preprint.

[C3] **Zhilin Zhang**†, Xiang Zhang†, Jiaqi Wei, Yiwei Xu, Chenyu You. *PosterGen: Aesthetic-Aware Multi-Modal Paper-to-Poster Generation via Multi-Agent LLMs.* IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2026 (Findings). [paper][code]

[C2] **Zhilin Zhang**, Jie Wang, Ruiqi Zhu, Xiaoliang Gong. *Efficient Bilinear Attention-based Fusion for Medical Visual Question Answering.* International Joint Conference on Neural Networks (IJCNN) 2025. [paper]

[C1] **Zhilin Zhang**, Fangyu Wu. *Enhanced Textual Feature Extraction for Visual Question Answering: A Simple Convolutional Approach.* 6th International Conference on Computer Vision and Computational Intelligence 2025. [paper]

[P1] Jie Wang, Yichen Wang, **Zhilin Zhang**, Kaidi Wang, Zhiyang Chen. *SentiXRL: An advanced large language Model Framework for Multilingual Fine-Grained Emotion Classification in Complex Text Environment.* Under review. [paper]

Research Experience

Research Intern, Stony Brook University **May 2025– Present**

mentored by Prof. Chenyu You *AI4Research, Multi-agents*

- Proposed **PosterGen**, the first aesthetic-aware multi-agent framework capable of generating visually appealing conference posters directly from research papers (⭐ **200+**).
- Designed a specialized multi-agent workflow (Parser, Curator, Layout, and Stylist) that mirrors professional human design pipelines and integrates narrative flow, spatial balance, and cohesive typography.
- Achieved performance significantly surpassing SoTAs in VLM-as-Judge aesthetic design metrics, delivering a tool that substantially alleviates the manual design burden for researchers.

Graduate Research Assistant, University College London **Oct 2024– May 2025**

mentored by Dr. Yujian Gan *Multi-agents; Text2SQL*

- Developed multiple agents with a layered query parsing architecture and managed robust inter-agent communication to accurately interpret natural language queries and provide reliable feedbacks on UK Census flow data.
- Led the generation of intermediate training data for fine-tuning LLMs by implementing two strategies: rewriting basic APIs and directly generating APIs based on metadata.
- Fine-tuned LLMs (e.g., Llama-3.1) and conducted comprehensive testing and evaluations.

[1] Google Scholar: https://scholar.google.com/citations?user=vWTBvyQAAAAJ

Undergraduate Thesis, Tongji University **Sep 2023– Aug 2024**

mentored by Prof. Xiaoliang Gong (Tongji University) *Vision and Language*

- Proposed a fusion framework **OMniBAN** by integrating Orthogonality Loss, Intra-modal Multi-head Self-Attention, and Cross-modal Bilinear Attention Networks to efficiently fuse visual (from **BiomedCLIP**) and textual (from **BioBERT**) representations for MedVQA.
- Conducted extensive comparative experiments on VQA-RAD and SLAKE datasets, and demonstrated that OMniBAN achieves competitive accuracy while reducing computational complexity (around **3.8× fewer FLOPs** and **1.5× fewer parameters**) compared to Transformer-based Co-Attention fusion models.

Undergraduate Research Assistant, Tongji University **Jun 2022– Oct 2022**

Mentored by Prof. Zhongpan Zhu (SRIAS) *Multi-modal Re-Identification*

- **Second inventor** for the patent: "*Intelligent Unmanned Parking Method based on Spatio-Temporal Vehicle Re-Identification*" Patent Link
- Proposed a solution for smart parking based on vehicle re-identification, spatial-temporal correlation, sensorless payment, and the architecture of the automated parking system.

Teaching Experience

Teaching Assistant (Spring & Fall 2025, Spring 2026) **Jan 2025– May 2026**

CS-GY 6233 Intro to Operating Systems (New York University) *C/C++; Operating System*

- Guided by Prof. Omar Mansour.

Additional Skills

- **Standard Tests:** TOEFL **110** (R29, L30, S23, W28), GRE **324** (V154, Q170)
- **Programming Languages:** Python, C/C++, Lean, PHP, Shell
- **Tools & Libraries:** Git, NumPy, Scikit-learn, Pytorch, Keras, LLaMA Factory, LangChain, LangGraph

Academic Services

- **Journal Reviewer:** Pattern Recognition, EAAI