Mohan Jiang

+86-189-5582-0630 | mhjiang0408@sjtu.edu.cn | github.com/mhjiang0408

EDUCATION

Shanghai Jiao Tong University   Shanghai,China
Bachelor of Engineering in Automation,Minor in Finance   *Sept 2021 - Present*

- **GPA**:3.78

- **Related Courses**: Thinking and Methodology in Programming(93/100); Data Structure(92/100); Probability and Statistics(100/100); Artificial Intelligence(93/100)

HONOR

- Shangjun Scholarship(top 3% in Automation)

- Type-B Merit Scholarship(top 10% in SJTU)

- Zhongyuan Haiyun Scholarship(top 5% in Automation)

RESEARCH INTEREST

I have a broad interest in large language models and bio-automation labs, especially the application of LLM. I am now researching the integration and application of LLM with biology and with automated driving. I sincerely hope that I can enhance the reliability and intelligence of AI for Biology and apply my findings in practical settings.

RESEARCH PROJECT

LLM as NPCs:Human-Like and Interpretable Multi-Agent Driving Simulation Jan 2024 - July 2024
ECCV 2024 Workshop**, Core Member, Instructor: **Dequan Wang   *Shanghai,China*

- **Project Description**: Leveraging the semantic understanding and generation capabilities of Large Language Models (LLMs) through prompt engineering, this project aims to use LLMs for decision-making in autonomous vehicles, determining the next steps in acceleration and steering angles. Additionally, LLMs are endowed with different driving personalities and using natural language to define the NPCs in the simulation.

- **My Work**: Optimizing the underlying control of the simulation, performing prompt engineering for LLM decision-making, and filtering the decisions' acceleration and steering intervals using the IDM model; Designing eight different driving personalities for the LLM in the simulation and creating evaluation metrics, such as comfort score, to quantify the driving performance of each personality; Designing various scenario maps to test and demonstrate the effectiveness of the LLM in playing NPC roles, covering corner cases in different scenarios.

- **Achievements**: The trajectory control effectiveness of the simulation was improved. The eight designed driving personalities each demonstrated distinct and characteristic control effects during testing. Additionally, the NPCs driven by the LLM effectively covered corner cases.

MAC: A Live Benchmark for MLLMs in Scientific Understanding   Dec 2024 – March 2025
*Core Member, Instructor: **Dequan Wang***   *Shanghai, China*

- **Project Description**: This project presents MAC (Multimodal Academic Cover), a live benchmark designed to evaluate the scientific reasoning abilities of multimodal large language models (MLLMs). By leveraging over 25,000 image-text pairs from top-tier journals like Nature and Science, MAC challenges models to understand and reason over complex, abstract visual and textual scientific content.

- **My Work**: I led the construction of the MAC benchmark, curated the dataset from high-impact journal covers, and designed evaluation protocols. Additionally, I proposed DAD, a lightweight inference-time method that enhances MLLM performance by extending visual features with language-based reasoning, enabling more accurate scientific interpretation across modalities.

- **Achievements**: Our method achieved up to 11% improvement in cross-modal scientific reasoning on the MAC-2025 benchmark. The benchmark itself demonstrated strong scalability and adaptability, validating its role as a continually evolving platform for assessing scientific understanding in MLLMs. This work has been submitted to **COLM 2025**.

HUE: A Benchmark for Humor Expression Understanding in MLLMs　　　March 2025 – May 2025
*Core Member, Instructor: **Dequan Wang***　　　　　　　　　　　　　　　　　　　　*Shanghai, China*

- **Project Description**: This project introduces HUE (HUmor Expression Benchmark), a large-scale benchmark comprising over 260,000 comics from 400+ creators, aimed at evaluating MLLMs' ability to understand humor expression. Unlike existing benchmarks, HUE focuses on how creators convey intent through the visual-semantic design of comics, including both strip and political genres.

- **My Work**: I curated the HUE dataset, designed the homogeneous humor expression classification task, and conducted detailed analysis of model performance across styles and genres. To address reasoning deficiencies, I developed DAD, a lightweight inference-time method that augments visual input with structured reasoning to improve model understanding of nuanced humor expression.

- **Achievements**: Our method improves performance by up to 57% on political comics, the most challenging subset of HUE. The benchmark exposes clear limitations in current MLLMs and highlights the need for deeper contextual and visual-semantic reasoning. HUE sets a new standard for humor evaluation and opens new directions for advancing cognitive capabilities in MLLMs. This work has been submitted to **NeurIPS 2025 D&B Track**.

TECHNICAL SKILLS

Languages: Python, C/C++, HTML, LaTeX, MATLAB, Verilog
Frameworks: Flask
Developer Tools: Git, VS Code, Visual Studio
Libraries: opencv, numpy, TensorFlow, Pytorch, rospy, roscpp