Wenshuo Zhang

wzhangeb@connect.ust.hk 86 187 0678 8065 https://zhang373.github.io/ Wenshuo ZHANG

Objective

I am looking for **long-term** collaboration on **research projects** alongside my Ph.D. studies.

I am open to topics and projects around NLP/LLM, especially on Human-LLM Alignment in Dialog/Agentic Systems for coding or X, emphasizing (1) regularizing user input to meet LLM requirements, (2) fine-tuning LLMs to address user limitations, and (3) data augmentation and evaluation to support both.

Education

The Hong Kong University of Science and Technology *Sep 2024 – Jul 2028*
Ph.D. Student in Computer Science and Engineering with Full Scholarship *CGA 3.66*

Xi'an Jiaotong University *Sep 2020 – Jul 2024*
B.S. in Automation (like ECE), Faculty of Electronic and Information Engineering *GPA 3.91*
Dual B.S. in Accounting, Management School *GPA 3.83*

The Hong Kong University of Science and Technology (Guangzhou) *Sep 2023 – Feb 2024*
Visiting Student in Data Science and Analysis, Information Hub

The University of Hong Kong *Jan 2023 – May 2023*
Exchange Student in Electrical and Electronic Engineering

Leading Experience

NeuroSync: Intent-Aware Code-Based Problem Solving and Alignment via Direct LLM Understanding Modification - Research, Submitted *Jan 2025 - May 2025*
VisLab@HKUST & WVisdom@UWaterloo, UIST (CCF-A)

- Resolved the misalignment (caused by "dual-directional ambiguity") between user intent and LLM-generated code for non-programmers by creating NeuroSync. This system pioneers a "direct intent-task matching" paradigm, externalizing the LLM's understanding (inferred coding tasks) as an interactive graph. Users can visually inspect and refine tasks prior to code generation, minimizing useless iterations, reducing waiting times, and enhancing coding efficiency.
- Designed a multi-agent system to simulate user-LLM interactions for synthetically generating high-quality training data, addressing critical data bottlenecks; developed a distillation pipeline that utilizes this data to fine-tune a small language model, enabling rapid and computationally efficient extraction of the LLM's understanding; and created an intent-aware graph simplification algorithm that dynamically reduces visual complexity by focusing on task nodes relevant to the user's intent.

Neuro Network and Distillation Based Ocean Data Compression for Visualization Acceleration - Project, Delivered *Jul 2024 - Feb 2025*
Centre for Ocean Research in Hong Kong and Macau (CORE) & Vislab@HKUST

- Part of "Study of the regional earth system for sustainable development under climate change in the Greater Bay Area"(60 million HKD form RGC@HK, 1.4 million for us)
- Developed a machine learning-based compression framework to enable real-time, interactive VR visualization of large-scale scientific data. This innovation addressed significant latency issues by compressing daily data files from over 6 GB to a highly efficient 6.75 MB model, reducing data processing and rendering time to just 0.75 seconds on Nvidia 3070.
- Engineered a data compression pipeline inspired by Neural Radiance Fields (NeRF), which utilized tiny neural networks for the spatial compression of individual data time steps. A Knowledge Distillation (KD) algorithm was then implemented to aggregate these multiple temporal models into a single, compact "distilled" model, ensuring data integrity while minimizing storage for deployment on edge computing platforms.

ParaGAN: A Parameter-Controlled Adversarial Generative Model for Cervical Cells Image Augmentation - Research *Oct 2023 - Feb 2024*
Med Data Lab@HKUST(GZ)

- Designed and implemented ParaGAN, a novel generative model for medical data augmentation. The key

innovation was an architecture that uses a contrastive loss function to decouple specific image attributes (such as cell area) in the latent space, aiming to create distinct and controllable generation channels.

- Engineered a multi-stage training framework to enable parameter-controlled image synthesis. This involved training a generator-discriminator pair to learn the data distribution, followed by a controller component that maps explicit, user-defined parameters to the decoupled latent space to guide the generation process.
- While precise parameter control proved challenging, the core architecture aims to significantly outperform baselines, i.e., generating higher-quality and more diverse images than standard GANs and Diffusion Models to boost the performance of downstream tasks like classification.

Publications

NeuroSync: Intent-Aware Code-Based Problem Solving and Alignment via Direct LLM Understanding Modification
First Author, UIST 2025 (CCF-A)

Branch Explorer: Leveraging Branching Narratives to Support Interactive 360° Video Viewing for Blind and Low Vision Users
Third Author, UIST 2025 (CCF-A)

Understanding Human Engagement with AI Characters: A Preliminary Investigation into AI Talk Shows
Fourth Author, Under Review @ CSCW 2025 (CCF-A)

UAV swarm positioning and scheduling method from the perspective of pure orientation passive positioning
First Author, IEEE ITNEC 2023

Laboratory Cooperation

Vislab, The Hong Kong University of Science and Technology *Sep 2024 - Jun 2028*
Dr. Huamin QU, Dr. Linping YUAN

WVisdom, University of Waterloo *Feb 2025 - May 2025*
Remote, Dr. Jian ZHAO

Med Data Lab, The Hong Kong University of Science and Technology (Guangzhou) *Oct 2023 - Feb 2024*
Dr. Zhao CHEN, Dr. Lei CHEN

CVMI Lab, The University of Hong Kong *Jan 2023 - Jul 2023*
Dr. Xiaojuan QI

Key Laboratory of Intelligent Network and Network Security, Ministry of Education, Xi'an Jiaotong University *Oct 2022 - Nov 2022*
Dr. Pinghui WANG

Skills

Languages: Python, C++, C, JavaScript

IELTS: 7.0 (Listening 8.5, Reading 7.5, Writing 6.0, Speaking 6.0)

PG Course: Knowledge Discovery in Database; Advanced Topics in Multimodal Machine Learning; Computational Theory; Machine Learning; Visual Computing for Visual Content Creation.

Service

Reviewer: LBW CHI 2025

Teaching Assistant: COMP 1029C - C Programming Bridging Course

Invited Talks

Oral Presentation on NeuroSync: Intent-Aware Code-Based Problem Solving and Alignment via Direct LLM Understanding Modification. ***UIST 2025, Korea*** Busan, 2025.10

Guest Talk on NeuroSync: How to do better intent-aware coding by chatting with Hangzhou, 2025.07

LLM. ***ChinaVis 2025, China***

Guest Talk on NeuroSync: Intent-Aware Code-Based Problem Solving and Alignment via Direct LLM Understanding Modification. ***Nanjing University, China*** Suzhou, 2025.07

Award

Post Graduate Studentship(PGS), The Hong Kong University of Science and Technology	2024.08
Outstanding Graduate Student in 2024, Xi'an Jiaotong University	2024.05
Outstanding Student in 2023, Xi'an Jiaotong University	2023.05
Fung Scholar and Fung Scholarship (Together with World Level University Students like MIT), the University of Hong Kong	2023.05
First Prize of National University Student Mathematical Modeling Competition in Shaanxi Competition Region, Xi'an Jiaotong University	2023.05
Outstanding Student in 2022, Xi'an Jiaotong University	2022.05
Jinyi and Academic Scholarships, Xi'an Jiaotong University	2022.05
Excellent Student Cadre in 2021, Xi'an Jiaotong University	2021.05
Jianbing Alumni Inspirational Scholarship (Top 2 out of over 300 people), Xi'an Jiaotong University	2021.05