

Hehai Lin

📍 Zhuhai, Guangdong ✉ linhh29@mail2.sysu.edu.cn ✉ hehailin11@gmail.com ☎ +86 13413304520

Research Interests

- Multimodal learning
- Large vision-language models (LVLMs)
- Logic reasoning

Education

RA	Nanyang Technological University , College of Computing and Data Science • Supervisor: Prof. Wenya Wang	2024 – now
BS	Sun Yat-sen University , School of Artificial Intelligence • GPA: 3.9/4.0, Rank: 6/71	2020 – 2024

Publications

I worked on several research projects during my undergraduate studies with different professors. Although these projects spanned various fields, I gained valuable experience and learned practical skills from them.

Human-Computer Interaction

- An Adaptive Training Tool for Critical Paper Reading. Kangyu Yuan*, **Hehai Lin***, Shilei Cao*, Zhenhui Peng, Qingyu Guo, and Xiaojuan Ma. UIST 2023 (CCF-A)
Learning to read scientific papers critically, which requires first grasping their main ideas and then raising critical thoughts, is important yet challenging for novice researchers. The traditional ways to develop critical paper reading (CPR) skills, e.g., checking general tutorials or taking reading courses, often can not provide individuals with adaptive and accessible support. In this paper, we first derive user requirements of a CPR training tool based on literature and a survey study (N=52). Then, we develop CriTrainer, an interactive tool for CPR training. It leverages text summarization techniques to train readers' skills in grasping the paper's main ideas. It further utilizes template-based generated questions to help them learn how to raise critical thoughts.

Data Mining

- Spatial-aware Semi-supervision for Arable Land Quality Evaluation. **Hehai Lin**, Wei Liu, Mengting Li, Kangyu Yuan, Zhao Liu, Huaijie Zhu, Jianxing Yu, and Jian Yin ICONIP 2024 (CCF-C)
we introduce SAFE, a Spatial-Aware Framework for arable land quality Evaluation that integrates convolutional neural network (CNN) and graph neural network (GNN) components into a deep learning-based architecture. This framework is designed to discern spatial local features and capture topological relationships within tabular data. Additionally, we incorporate a self-supervised regularization technique utilizing contrastive learning into the architecture training objective.

Multimodal Learning

- Self-Correction is More than Refinement: A Learning Framework for Language and Visual Reasoning Tasks. Jiayi He*, **Hehai Lin***, Yi Fung, Qingyun Wang, Sha Li, Heng Ji Under review (AAAI 2025)
While Vision-Language Models (VLMs) have shown remarkable abilities in visual and language reasoning tasks, they invariably generate flawed responses. Self-correction that instructs models to refine their outputs presents a promising solution to this issue. Previous studies have mainly concentrated on Large Language Models (LLMs), while the self-correction abilities of VLMs, particularly concerning both visual and linguistic information, remain largely unexamined. This study investigates the self-correction capabilities of VLMs during both inference and fine-tuning stages. We introduce a Self-Correction Learning (SCL) approach that enables VLMs to learn from their self-generated self-correction data through Direct Preference Optimization (DPO) without relying on external feedback, facilitating self-improvement. This study emphasizes that self-correction is not merely a refinement process; rather, it should enhance the reasoning abilities of models through additional training, enabling them to generate high-quality responses directly without further refinement.

- Multi-view Analysis for Modality Bias in Multimodal Misinformation Benchmarks. **Hehai Lin**, Hui Liu, Shilei Cao, Hao-liang Li, Wenya Wang Under review (WWW 2025)
Numerous multimodal misinformation benchmarks exhibit bias toward specific modalities, allowing detectors to make predictions based solely on one modality. Training detectors on such datasets can significantly degrade performance in real-world applications. While previous research has quantified modality bias at the dataset level or manually identified spurious correlations between modalities and labels, these approaches lack meaningful insights at the sample level and struggle to scale to the vast amount of online information. In this paper, we investigate the design for automatically recognizing modality bias at the sample level. Specifically, we introduce three views, namely modality benefit, modality flow, and modality causal effect, to quantify samples' modality contribution based on different theories. To verify their effectiveness and discover the pattern of bias, we conduct a human evaluation on two benchmarks Fakeddit and MMFakeBench, and compare the performance of each view and their ensemble multi-view analysis. The experimental result indicates that multi-view analysis yields the highest performance and is aligned with human judgment in most samples. We further discuss the sensitivity and consistency of each view.

Honors & Awards

- **JinDao Scholarship** Jun. 2024
- **Third-class Scholarship for outstanding Students of Sun Yat-sen University (top 30%)** 2022 - 2023
- **Honorable Mention in the ECV2023 Cross-Camera Person Tracking Competition (7th place)** Jun. 2023
- **The Fourth Place in the Second Intelligent Network Competition (top 5%)** Oct. 2022
- **Second-class Scholarship for outstanding Students of Sun Yat-sen University (top 15%)** 2021 - 2022
- **First-class Scholarship for outstanding Students of Sun Yat-sen University (top 5%)** 2020 - 2021

Copyright of Computer Software

- **Identification and Track of Contacts based on Surveillance Camera Network.** Dec. 2021 - Oct. 2022
Build a system to re-identify COVID-19 patients in all cameras and track them to find all contiguities. In this project, I undertake the part of Multiple Objectives Tracking.

Academic Services

- Conference Reviewer: COLING 2025, WWW 2025

Technologies

Languages: Chinese (Native), English (Fluent)

Coding: C/C++, Python, Matlab, Shell

Tools: Pytorch, Docker, LaTeX, Markdown