Zhuo **Zhi**

(+44) 7419907662 | zhuo.zhi.21@ucl.ac.uk | zhuozhi-ucl.github.io/

Education

University College London (UCL)

UK

PhD in Electronic and Electrical Engineering — *Oct. 2021 - Oct. 2025 (expected)*

- Research interests in machine learning: multimodal learning (robustness and fusion theory), computer vision, LLM, MLLM, video understanding.

Harbin Institute of Technology (HIT)

China

M.S. in Instrument Science and Technology — *Sep. 2019 - June 2021*

- Research direction: Multi-sensor fusion, integrated navigation, multimodal learning.

Shandong University (SDU)

China

B.S. in Automation — *Sep. 2015 - June 2019*

- Core courses: Automatic Control Theory, Signal and System, Computer Control System.

Employment

Samsung R&D Institute UK

Research Intern — *July 2025 - now*

- Research topic: real-time explainable emotion analysis in streaming videos.
- Build the industry's first streaming video explainable emotion analysis dataset and benchmark.
- Build an efficient omni-MLLM to support explainable emotion analysis from streaming video, achieving 5fps on A40 GPU.

Huawei London Research Center

Research Intern — *Jun. 2024 - Jun. 2025*

- Propose two solutions for long video understanding: the LLM-based agent system and end-to-end MLLM.
- VideoAgent2 achieves 1st rank on Ego4D EgoShema Challenge Leaderboard and accepted by NeurIPS SEA 2025.
- A patent is approved and a product line award is received for Videoagent2.

Publication

- **Zhi, Z.**, Wu, Q., Li, W., Li, Y., Shao, K. and Zhou, K., 2025. VideoAgent2: Enhancing the LLM-Based Agent System for Long-Form Video Understanding by Uncertainty-Aware CoT. NeurIPS SEA.
- **Zhi, Z.**, Feng, C., et al., 2025. TFAR: A Training-Free Framework for Autonomous Reliable Reasoning in Visual Question Answering. TMLR.
- **Zhi, Z**., Liu, Z., Wu, Q. and Rodrigues, M., 2025. Wasserstein Modality Alignment Makes Your Multimodal Transformer More Robust. TMLR.
- **Zhi, Z**., Liu, Z. et al., 2024. Borrowing Treasures from Neighbors: In-Context Learning for Multimodal Learning with Missing Modalities and Data Scarcity. Neurocomputing.
- **Zhi, Z**., Liu, D. and Liu, L., 2022. A performance compensation method for GPS/INS integrated navigation system based on CNN–LSTM during GPS outages. Measurement.
- **Zhi, Z.**, Elbadawi, M., Daneshmend, A., Orlu, M., Basit, A., Demosthenous, A. and Rodrigues, M., 2022, July. Multimodal diagnosis for pulmonary embolism from EHR data and CT images. In 2022 44th Annual International Conference of the IEEE EMBC (pp. 2053-2057). IEEE.
- Feng, C., Liu, Z., **Zhi, Z**., Bogunovic, I., Gerner-Beuerle, C. and Rodrigues, M., 2024. PROSAC: Provably Safe Certification for Machine Learning Models under Adversarial Attacks. AAAI.

Project & Research

VideoAgent2: Enhancing the LLM-based agent system for long-form video understanding by uncertainty-aware CoT

Research intern in Huawei — *2025*

- Designed a specialized CoT process based on plan-adjust mode to enhance LLM reasoning and decision-making in long video understanding.
- Introduced uncertainty-guided CoT reasoning, which mitigates noise and illusion in the system while requiring no additional parameters.
- Proposed the VideoAgent2 pipeline, which integrates innovative designs such as general context acquisition and specialized tools.
- VideoAgent2 achieves 1st rank on Egoschema leaderboard (https://eval.ai/web/challenges/challenge-page/2238/leaderboard/5535)

TFAR: A Training-Free Framework for Autonomous Reliable Reasoning in Visual Question Answering.

TMLR — *2025*

- Proposed a training-free multimodal reasoning framework that integrates external vision models with uncertainty quantification into an MLLM to solve the problem (1) CoT-based multimodal reasoning often demands costly data annotation and fine-tuning and (2) agentic approaches relying on external tools risk introducing unreliable output from these tools.

Wasserstein Modality Alignment Makes Your Multimodal Transformer More Robust

TMLR *2024*

- Proposed a Wasserstein distance-based regularization method to adjust the modality alignment degree during the fine-tuning phase.
- The proposed method is tested on both 2-modalities and 3-modalities tasks. The results show that it achieves a good balance of modality alignment and heterogeneity which significantly improves model performance and robustness.

In-Context Learning for Multimodal Learning with Missing Modalities and Data Scarcity

Neurocomputing *2023*

- Proposed a retrieval-augmented in-context learning method to improve the robustness of multimodal models under missing modalities and data scarcity.
- The proposed approach is evaluated under both multimodal medical and vision-language tasks to show the SOTA performance.

Honors & Awards

2021 **UCL EPSRC DTP Research Studentship**,
2020 **First-class Special Scholarship for Graduate Student**, (Top 18%, 20/110)
2019 **First-class Scholarship for Outstanding Undergraduate Student**, (Top 5%, 6/120)
2018 **Shandong University Outstanding Undergraduate Research Assistant** , (Top 5%, 2/40)
2018 **First Prize of National College Student Smart Car Competition** , (Top 0.5%, 10/2000)
2018 **Special Prize of China Engineering Robot Competition** , (Top 1.5%, 3/200)
2017 **First Prize of National College Students Electronic Design Competition**, (Top 1%, 20/2000)

Teaching & Service

2023 **Teaching assistant**, Integrated Machine Learning Systems at UCL, Dept. EEE
2022 **Teaching assistant**, Applied Machine Learning Systems at UCL, Dept. EEE
2023 **Teaching assistant**, Machine Learning Masterclass at the UK Defence Science and Technology Laboratory
2023 **Organiser**, Weekly Machine Learning Reading Group at UCL Dept. EEE
2024 **Mentor**, Third-year undergraduate student project at UCL Dept. EEE
2024 **Mentor**, Master student project at UCL Dept. EEE
now **Reviewer**, AAAI, IEEE-JBHI, IEEE-TIM, IEEE sensor journal, Measurement, WACV