

# Yake WEI/卫雅珂

Ph.D candidate (**will graduate in June 2026**)  
Gaoling School of Artificial Intelligence  
Renmin University of China

[yakewei@ruc.edu.cn](mailto:yakewei@ruc.edu.cn)

[Google Scholar](#)

[Homepage](#)

## EDUCATION



**Renmin University of China**  
Ph.D Candidate in Artificial Intelligence  
Advisor: *Prof. Di Hu*

Beijing, China  
Sep. 2021 - Present



**Carnegie Mellon University**  
Visiting Scholar  
Advisor: *Prof. Fernando De la Torre Frade*

Pittsburgh, USA  
Dec. 2023 - Aug. 2024



**University of Electronic Science and Technology of China**  
B.E. in Computer Science and Technology

Chengdu, China  
Sep. 2017 - Jun. 2021

## SELECTED AWARDS AND SCHOLARSHIPS

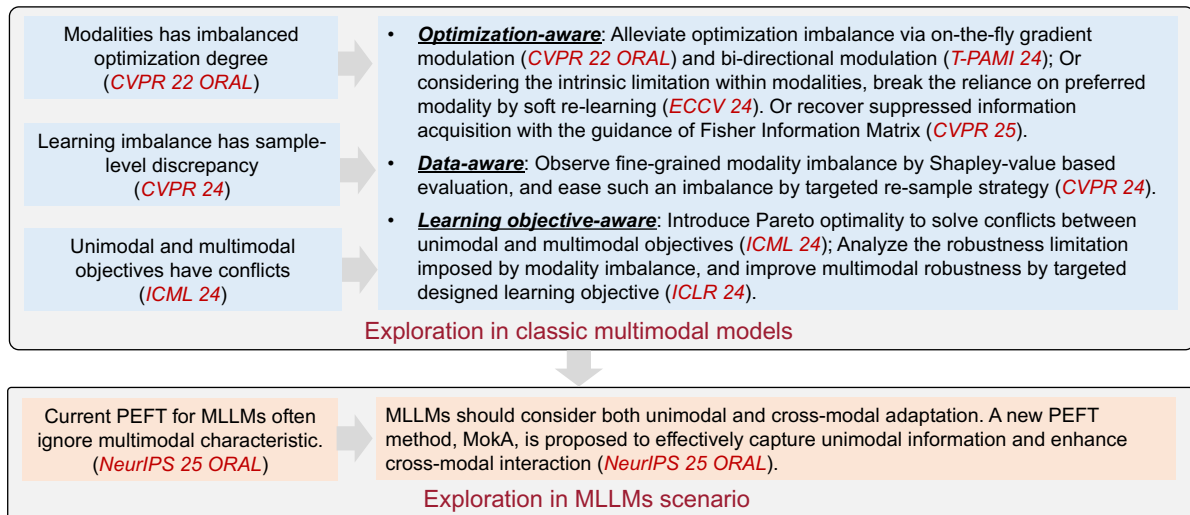
- **Baidu Scholarship** (10 Ph.D students worldwide), 2024.
- **China National Scholarship for Ph.D student** (highest student honor in China), 2024.
- Outstanding Graduate Award (highest honor for graduates set by Sichuan province), 2021.
- Outstanding Graduate of University of Electronic Science and Technology of China, 2021.

## RESEARCH INTERESTS

Interested in the inherent learning mechanism of perceiving, formulating, and understanding the environment with heterogeneous information from multiple modalities, *e.g.*, *vision*, *sound*, *text*.

**Part 1.** Focus on providing **better solutions for building Multimodal LLMs, caring for modality-specific characteristics**. For now, we have provided a new PEFT pipeline for MLLMs, *MokA*, which ensures both unimodal and cross-modal adaptation.

**Part 2.** In the paper presented at CVPR 2022 (ORAL), **introduce the research topic of “Balanced Multimodal Learning”**. Highlight a pervasive issue in multimodal learning, where information utilization of certain modality can be undesirably suppressed by others. Then conduct a series of systematic studies to alleviate this issue.



## PAPER LIST

---

- [1] **Yake Wei**, Yu Miao, Dongzhan Zhou, and Di Hu. Moka: Multimodal low-rank adaptation for mllms. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025, **ORAL (1.46% of accepted papers)**.
- [2] **Yake Wei**, Di Hu, Henghui Du, and Ji-Rong Wen. On-the-fly modulation for balanced multimodal learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2024.
- [3] **Yake Wei** and Di Hu. Mmpareto: boosting multimodal learning with innocent unimodal assistance. *International Conference on Machine Learning (ICML)*, 2024.
- [4] **Yake Wei**, Ruoxuan Feng, Zihe Wang, and Di Hu. Enhancing multimodal cooperation via sample-level modality valuation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [5] **Yake Wei**, Siwei Li, Ruoxuan Feng, and Di Hu. Diagnosing and re-learning for balanced multimodal learning. *European Conference on Computer Vision (ECCV)*, 2024.
- [6] Xiaokang Peng\*, **Yake Wei\***, Andong Deng, Dong Wang, and Di Hu. Balanced multimodal learning via on-the-fly gradient modulation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. (\* equal contribution, **ORAL**).
- [7] Chengxiang Huang\*, **Yake Wei\***, Zequn Yang, and Di Hu. Adaptive unimodal regulation for balanced multimodal information acquisition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. (\* equal contribution).
- [8] Guangyao Li\*, **Yake Wei\***, Yapeng Tian\*, Chenliang Xu, Ji-Rong Wen, and Di Hu. Learning to answer questions in dynamic audio-visual scenarios. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. (\* equal contribution, **ORAL**).
- [9] Di Hu, **Yake Wei**, Rui Qian, Weiyao Lin, Ruihua Song, and Ji-Rong Wen. Class-aware sounding objects localization via audiovisual correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2021. (**First student author**).
- [10] Zequn Yang, **Yake Wei**, Ce Liang, and Di Hu. Quantifying and enhancing multi-modal robustness with modality preference. *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.
- [11] Haotian Ni, **Yake Wei**, Hang Liu, Gong Chen, Chong Peng, Hao Lin, and Di Hu. Reviving the cooperation dynamics in multimodal transformer. *International Conference on Machine Learning (ICML)*, 2025.
- [12] Ruotian Peng, Haiying He, **Yake Wei**, Yandong Wen, and Di Hu. Patch matters: training-free fine-grained image caption enhancement via local perception. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.
- [13] Zequn Yang, Han Zhang, **Yake Wei**, Zheng Wang, Feiping Nie, and Di Hu. Geometric-inspired graph-based incomplete multi-view clustering. *Pattern Recognition*, 2024.

## SURVEYS

---

- [1] **Yake Wei**, Di Hu, Yapeng Tian, and Xuelong Li. Learning in audio-visual context: A review, analysis, and new perspective. *arXiv preprint arXiv:2208.09579*, 2022.
- [2] Qingyang Zhang, **Yake Wei**, Zongbo Han, Huazhu Fu, Xi Peng, Cheng Deng, Qinghua Hu, Cai Xu, Jie Wen, Di Hu, et al. Multimodal fusion on low-quality data: A comprehensive survey. *arXiv preprint arXiv:2404.18947*, 2024.

## INVITED PRESENTATIONS

---

- “Balanced Multimodal Learning”  
Invited talk at *Peking University, CoRe 2025*.
- “Balanced Multimodal Learning”  
Invited talk at *Global PhD Gathering, Pujiang AI Conference, 2024*.
- “Balanced Multimodal Learning”  
Invited talk at *Virginia Tech, 2024*.
- “Balanced Multimodal Learning”  
Invited talk by *TechBeat, 2024*.
- “Exploration of Audio-visual Scene Understanding and Multimodal Learning Mechanisms”  
Invited talk at *BAAI Conference, 2022*.

## PROFESSIONAL SERVICE

---

### **Journal Reviewer:**

- IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)
- IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT)
- IEEE Transactions on Multimedia (T-MM)
- Pattern Recognition

### **Conference Reviewer:**

- International Conference on Machine Learning (ICML)
- Annual Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Learning Representations (ICLR)
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)