# Yake WEI/卫雅珂

Ph.D candidate (**will graduate in 2026**)

Gaoling School of Artificial Intelligence

Renmin University of China

yakewei@ruc.edu.cn

Google Scholar

Homepage

## EDUCATION

**Renmin University of China** — Beijing, China

Ph.D Candidate in Artificial Intelligence — Sep. 2021 - Present

Advisor: *Prof.* Di Hu

**Carnegie Mellon University** — Pittsburgh, USA

Visiting Scholar — Dec. 2023 - Aug. 2024

Advisor: *Prof.* Fernando De la Torre Frade

**University of Electronic Science and Technology of China** — Chengdu, China

B.E. in Computer Science and Technology — Sep. 2017 - Jun. 2021
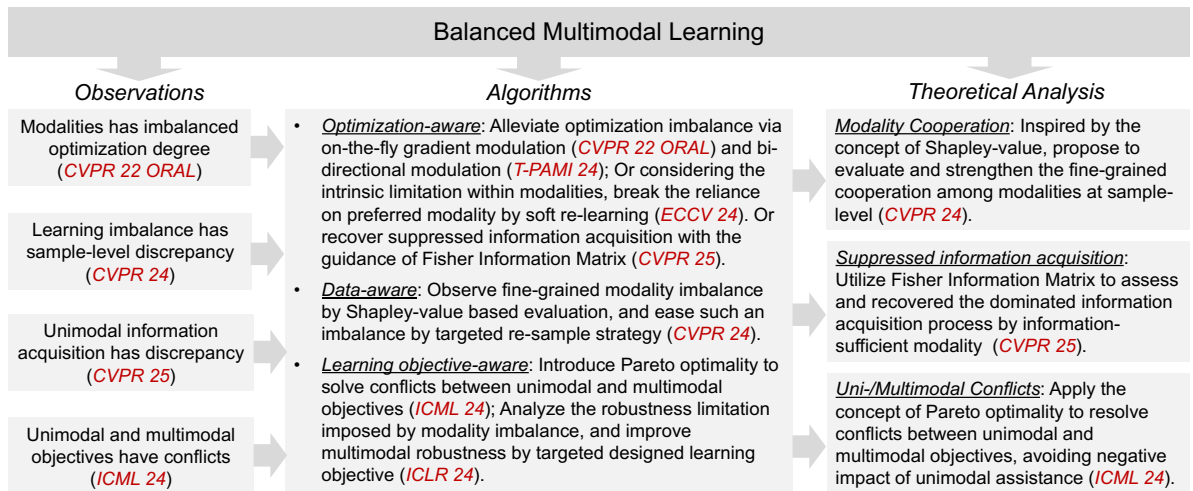
## SLECTED AWARDS AND SCHOLARSHIPS

- **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 a **better solution for building Multimodal LLMs, caring for modality-specific characteristics**. For now, we provide a new PEFT pipeline for MLLMs, ensuring both unimodal and cross-modal adaptation.

**Part 2.** In the paper presented at CVPR 2022 (ORAL), **introduce the research topic of "Balanced Multimodal Learning" for the first time**. 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.



Balanced Multimodal Learning

**Observations**

Modalities has imbalanced optimization degree (*CVPR 22 ORAL*)

Learning imbalance has sample-level discrepancy (*CVPR 24*)

Unimodal information acquisition has discrepancy (*CVPR 25*)

Unimodal and multimodal objectives have conflicts (*ICML 24*)

**Algorithms**

- *Optimization-aware*: Alleviate optimization imbalance via on-the-fly gradient modulation (*CVPR 22 ORAL*) and bi-directional modulation (*T-PAMI 24*); Or considering the intrinsic limitation within modalities, break the reliance on preferred modality by soft re-learning (*ECCV 24*). Or recover suppressed information acquisition with the guidance of Fisher Information Matrix (*CVPR 25*).
- *Data-aware*: Observe fine-grained modality imbalance by Shapley-value based evaluation, and ease such an imbalance by targeted re-sample strategy (*CVPR 24*).
- *Learning objective-aware*: Introduce Pareto optimality to solve conflicts between unimodal and multimodal objectives (*ICML 24*); Analyze the robustness limitation imposed by modality imbalance, and improve multimodal robustness by targeted designed learning objective (*ICLR 24*).

**Theoretical Analysis**

*Modality Cooperation*: Inspired by the concept of Shapley-value, propose to evaluate and strengthen the fine-grained cooperation among modalities at sample-level (*CVPR 24*).

*Suppressed information acquisition*: Utilize Fisher Information Matrix to assess and recovered the dominated information acquisition process by information-sufficient modality (*CVPR 25*).

*Uni-/Multimodal Conflicts*: Apply the concept of Pareto optimality to resolve conflicts between unimodal and multimodal objectives, avoiding negative impact of unimodal assistance (*ICML 24*).

## Paper list

[1] **Yake Wei**, Yu Miao, Dongzhan Zhou, and Di Hu. Moka: Multimodal low-rank adaptation for mllms, 2025.

[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. In *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. In *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. In *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. In *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. In *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. In *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. In *The Twelfth International Conference on Learning Representations* **(ICLR)**, 2024.

[11] Ruotian Peng, Haiying He, **Yake Wei**, Yandong Wen, and Di Hu. Patch matters: training-free fine-grained image caption enhancement via local perception. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* **(CVPR)**, 2025.

[12] Haotian Ni, **Yake Wei**, Hang Liu, Gong Chen, Chong Peng, Hao Lin, and Di Hu. Reviving the cooperation dynamics in multimodal transformer. In *International Conference on Machine Learning* **(ICML)**, 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* **(PR)**, 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 *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)

**Conference Reviewer:**

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022-2025
- IEEE International Conference on Computer Vision (ICCV), 2023/2025
- European Conference on Computer Vision (ECCV), 2022/2024
- The AAAI Conference on Artificial Intelligence (AAAI), 2023-2025