



个人简介

香港中文大学多模态（语音、长文本、连续多图）人工智能方向博士生，具备丰富的工业级大模型研发经验，曾于微软亚洲研究院、腾讯、阿里巴巴达摩院（X-PLUG）和美团（Longcat-Omni）等核心团队从事实习，深入参与全模态大模型音频推理等前沿技术攻坚。博士期间于J-STSP、ACM-MM、ACL等顶级期刊/会议发表学术论文共14篇（一作9篇）。

教育经历

- 博士 | 香港中文大学** 2019年8月至2026年7月(预期)
系统工程与工程管理系（信息科学专业），导师：**蒙美玲教授** **Fellow of ISCA and IEEE** 香港
研究方向：多模态（语音、文本、图片）人工智能
- 学士 | 南京大学** 2015年9月至2019年6月
物理学院（声学专业和电子信息专业双学位），导师：**卢晶教授** 南京
研究方向：音频信号处理

科研经历

- 长篇视觉叙事任务的跨模态理解与生成评估 | 微软亚洲研究院 & 香港中文大学**
面向长篇视觉叙事（多图、长文）的细粒度跨模态语义分析与开放式生成评估。
 - 单模态表征**：针对长篇视觉内容中特征复杂、数据稀缺等难点，引入参数高效微调（LoRA+Attention），将通用表征映射至长篇叙事领域，有效捕捉复杂的语义结构和上下文依赖关系。
 - 跨模态对齐**：针对开放式跨模态生成任务中缺乏“唯一标准答案”的痛点，创新性地构建了自适应参考系（动态主题和多图锚定），精准量化了视觉输入流与语言生成流在语义及时序上的一致性。
 - 多模态融合**：针对多模态任务中信息稀释及模态冲突等问题，设计了基于不确定性和动态路由的多专家融合策略，大幅提升了复杂多模态场景下评估系统的性能和鲁棒性（多个公开基准集SOTA）。
- 全模态大模型音频推理 | 美团（Longcat-Omni团队）**
面向全模态大模型的复杂音频推理能力评估与后训练对齐（六个月）。
 - 能力解耦与评估体系**：针对目前评测集中任务冗余与覆盖度不足的难点，创新性提出基于共识聚类的细粒度能力解耦方案，构建了最大化能力覆盖且最小化冗余的综合评估基准，为模型迭代提供了更精准的依据和优化方向。
 - 后训练对齐**：针对数据稀缺和模态惰性等，提出了均衡采样、迭代蒸馏、数据质检和GRPO对齐的全链路后训练方案，最终在各音频推理基准集上实现了显著且一致的性能提升（超越Qwen3-Omni-30B-A3B-Thinking）。
- 语音文本对话大模型 | 阿里巴巴达摩院（X-PLUG团队）**
大规模数据处理和模块化语音文本大模型的基线搭建（三个月）。
 - 大规模数据工程**：搜集和构建大规模语音文本数据（约10万小时语音或18B词元），合作完成emotion2vec项目。
 - 模块化基线搭建**：基于离散词表的范式模块化构建Speech LLM，增强模型的情绪感知和个性化对话能力。
- 情感识别 | 腾讯（光子实验室）**
基于上下文感知和跨模态注意力微调的情感识别（六个月）。
 - 上下文依赖建模**：引入上下文滑动窗机制以捕捉对话交互中情绪表达的连贯性与动态变化。
 - 多模态注意力融合**：针对语音与文本模态间的异构性与互补性，设计了基于跨模态注意力的特征融合结构，最终在公开情感数据集IEMOCAP上实现SOTA。

1. Constraint, Calibration, and Coordination (C³): Factual Grounding and Evidential Fusion for Holistic and Reliable Multimodal Neurocognitive Disorder Detection. (*under review*), 2026.
2. Detecting Neurocognitive Disorders through Analyses of Topic Evolution and Cross-modal Consistency in Visual-Stimulated Narratives. *IEEE Journal of Selected Topics in Signal Processing*, 2025. [JCR: Q1](#) [📄](#)
3. Generate, Align & Predict (GAP): Detecting NCDs via Cross-modal Consistency in Narratives. *ACM MM*, 2025. [CCF: A](#) [📄](#)
4. On the Within-class Variation Issue in Alzheimer's Disease Detection. *Interspeech*, 2025. [CCF: B](#) [📄](#)
5. Emotion2vec: Self-Supervised Pre-Training for Speech Emotion Representation. *ACL*, 2024. [CCF: A](#) [📄](#)
6. Not all errors are equal: investigation of speech recognition errors in AD detection. *ISCSLP*, 2024. [📄](#)
7. A Hierarchical Regression Chain Framework for Affective Vocal Burst Recognition. *ICASSP*, 2023. [CCF: B](#) [📄](#)
8. Leveraging Pretrained Representations With Task-Related Keywords for AD Detection. *ICASSP*, 2023. [CCF: B](#) [📄](#)
9. AD Detection using Pretrained Acoustic Representations: A Comparison Study. *NCMMSC*, 2022. [📄](#)
10. Spoofing-aware speaker verification by multi-level fusion. *Interspeech*, 2022. [CCF: B](#) [📄](#)
11. Context-aware Multimodal Fusion for Emotion Recognition. *Interspeech*, 2022. [CCF: B](#) [📄](#)
12. Dual Dropout Ranking of Linguistic Features for Alzheimer's Disease Recognition. *APSIPA ASC*, 2021. [📄](#)
13. A Comparative Study of Acoustic and Linguistic Features Classification for AD Detection. *ICASSP*, 2021. [CCF: B](#) [📄](#)
14. Development of the CUHK Elderly ASR System for AD Detection using the DementiaBank Corpus. *ICASSP*, 2021. [CCF: B](#) [📄](#)
15. 基于传声器阵列确定声源信息的方法、装置及电子设备. 专利: *CN110148422B*, 2021 (本科). [📄](#)

学术奖项和服务

• 荣誉奖项:

- Best Paper Award of ACM Multimedia MCHM Workshop
- Champion of The ACII Affective Vocal Bursts Competition
- 南京大学优秀毕业论文
- 全国大学生数学建模竞赛二等奖
- 教育部国家奖学金

• 学术服务:

- 审稿人: T-ASLP (期刊)、ACL、ICASSP、INTERSPEECH、COLING等
- 合作讲师: 研究生课程“对话式人工智能系统”(ASR部分)

专业技能

- **编程语言与工具:** Python, C++, Bash, Git, Docker, FFmpeg, SoX
- **训练框架:** PyTorch, HuggingFace, DeepSpeed, Megatron-LM, vLLM
- **研究领域:** 多模态学习, 语音信号处理, 强化学习