程浩

(+86)156-9563-5438 | chhao.hfut@gmail.com | github.com/Cb1ock | zhihu.com

教育经历

合肥工业大学 | 人工智能，计算机与信息学院 2023.09–2026.06

研究方向：多模态表征学习、多模态音视频理解 国家奖学金、*CET-6*

安徽理工大学 | 土木工程，土木与建筑学院 2019.09–2023.06

实习经历

北京智谱华章科技股份有限公司 | AI 院 | **语音算法实习生** | 2025.06–**至今**

GLM-ASR-9B 模型在线上 Chat 场景的优化

- **背景：** GLM-ASR 在线上语音识别任务中存在数字与标点错误率高、专有名词识别弱、抗噪能力不足等问题，影响用户体验。
- **数据质量保障：** 与产品及数据团队协作构建 hard case 数据集，设计**自动化样本筛选**与**多场景标注文档**；针对用户意图理解**完善数据清洗流程**，确保训练样本的高质量与多样性。
- **模型优化：** 负责模型 SFT 与 RL 后训练。SFT 阶段将 WER 降至 **2.67%**；RL 阶段采用 **DPO** 方法对齐人类偏好，基于 Gemini-2.5-Pro 自动构建正负样本对，使模型更贴近人类表达习惯；在自建的 HumanEval Arena 上相较 SFT 模型 Elo 分数为 **1790 vs 1309**。
- **核心亮点：** 在 **GLM-4** 架构**引入日志数据**作为 SFT 输入，虽然使早期 loss 升高，但**显著提升**了收敛稳定性与 loss 表现；在内部 Benchmark 中相较无 log 版本取得 **2.78 vs 4.69 %** 的表现，并在真实业务测试中**均优于**无 log 版本，并**成功通过**测试组验收，现已服务于公司某对话 app。

会议场景 ASR 优化与系统构建

- **背景：** 会议场景下多人交替发言及超长音频，导致 VAD 与 GLM-ASR 的组合在识别准确率上表现不佳。
- **多说话人方案：** 调研并实现基于 **SOT** 的多说话人处理方法，提升多人交替发言识别的准确度。
- **无限延拓转录：** 针对 GLM-ASR 的 30s 音频窗口限制，独立设计**音频时长无限延拓方案**，无需重新预训练 audio encoder，即可实现对超长音频的连续识别，并可避免 **VAD 切分**导致的语义破坏以及参数难调问题。
- **效果与对比：** 结合 SOT 方案后，在会议场景测试集上将 WER 降至 9.36；与**业界 SOTA 系统**（如 Qwen3-ASR-Flash/Qwen3 Omni + Qwen3-ASR-Toolkit、飞书妙记）对比，具有较强竞争力。
- **系统落地：** 完成 **Meeting ASR Toolkit** 与 **Benchmark** 框架的代码实现，覆盖从数据处理到模型推理的全链路；方案已部署至公司某 Agent 产品核心算法中。

科研经历

CCF-A **类会议** ACM Multimedia 2025 Oral：

"VAEmo: Efficient Representation Learning for Visual-Audio Emotion with Knowledge Injection" 第一作者

- 主导高效视听情绪模型搭建（共享编码器 + 同步时序 PE + 两阶段预训练与 MLLM 知识注入），并创新性地提出无监督的**表征域迁移**方案；以 **33M** 参数在多项 AVER 任务上较 80M+ SOTA 提升 **2–10%**，完成从方案到实现与实验闭环。

多模态人格 Benchmark 采集构建

- 主导**需求设计**、**人格激发与评判标准**等 Benchmark 方案，打造高并发采集与帧级入库平台（Java Socket + C++/JNI 编解码），显著提升传输效率与系统稳定性；统一流程与标签规范，支撑大规模多模态数据采集。

其他研究经历

- 将音视频同步性作为先验知识，将两个时间序列切分后，使用改进的 DTW 距离计算两序列之间的差异，并以最小化该距离为优化目标；在相同实验条件下，该研究已在多个数据集上获得性能提升。
- 深入探索多模态大模型，如 QWen 系列、LLaVA 等在音视频表征学习领域的应用

技能图谱

- **编程语言：** Python（熟练）、C/C++（熟练）、Shell
- **工具链：** PyTorch、Transformers、vLLM、Verl、LLaMA Factory、Git、Markdown、LaTeX、wandb
- **课程基础：** 概率论与统计、矩阵论、运筹学、信息论、CS231n、CS336
- 具备丰富的**多节点大模型训练经验**
- 负责实验室与学院的 GPU 服务器管理与运维，擅长**技术文档撰写与环境配置排障**