

# 王晨瑜

+86-18292065293

wangchenyu241@mailsucas.ac.cn

HomePage



## 基本信息

出生年月：2001年06月      性别：女      籍贯：陕西省西安市      政治面貌：中共党员  
中国科学院大学电子信息硕士研究生（预计 2027 年毕业），在上海微系统所（SIMIT）开展研究，现于上海 AI Lab 实习。研究聚焦医疗大模型的规则监督学习、检索增强临床决策与 Agent 工具安全、开展迁移学习与无监督领域自适应研究。

## 教育经历

|                      |                 |    |                   |
|----------------------|-----------------|----|-------------------|
| 中国科学院大学              | 电子信息            | 硕士 | 2024.09 – 2027.06 |
| 上海微系统与信息技术研究所（SIMIT） | 导师：李宝清研究员、涂敏研究员 |    | 上海                |
| 西安理工大学               | 电气工程及其自动化       | 本科 | 2019.09 – 2023.07 |
| 电气工程学院               | 导师：王开艳副教授       |    | 西安                |

## 科研与实习经历

上海人工智能实验室（Shanghai AI Lab）实习      2026.03 – 至今  
多模态智能体算法实习生    高峰科研任务部 · 多模态前沿

[1] Guideline-as-Oracle：规则监督的眼科分诊智能体      2026.03 – 2026.08

- 数据与训练：针对医学对话标注成本高的问题，将 70 条眼科指南规则转化为 3,000 条多轮训练对话；以规则分配标签，结合类别重采样与角色掩码完成 9B 模型全参数 SFT，并探索 Outcome GRPO。
- 评测结果：在 201 例内部规则参照评测中，一致率由 61.7% 提升至 74.1%，急症召回率由 9.5% 提升至 69.0%；通过第二随机种子与第二患者模拟器复核增益，考察训练与交互条件变化下的稳定性。

[2] 转移性结直肠癌 MDT 决策预测与检索增强建模      2026.08 – 至今

- 数据与协议：以会前病历预测五类 MDT 处置决策，完成 9,779 次会诊的去标识化、筛选与患者级隔离；冻结 1,247 次会诊 / 610 名患者测试集，对比经典机器学习、检索与大模型方案。
- 方法与结果：将相似病例检索、纵向病史与 XGBoost 概率先验融入大模型提示，测试准确率 64.1%、宏 F1 0.527；通过信息源消融与跨厂商模型复核，评估检索、历史与统计先验的贡献。

[3] 医学 Agent 工具安全：DSH 组件静态审计与风险归因      2026.08 – 至今

- 审计框架：针对仓库级扫描混淆共享与工具私有风险的问题，设计依赖闭包归因与四级证据评分，覆盖 12 个安全域；完成 207 个仓库、12,830 个组件、11,124 个入口的静态审计。
- 验证与复现：结合语法树解析与函数内污点分析核验风险证据，将需人工处理条目由 4,794 条降至 153 条，归并定位 24 处代码缺陷；固定源码及解析器版本，两次干净环境重跑结果字节一致。

- 模型方法: 针对时变漂移与跨设备分布偏移, 提出单阶段无监督适应框架, 联合分层 MK-MMD、CORAL、双层中心正则与多头分类; 以相对响应前端消除乘性增益, 设计条件对抗变体 UnifiedGas-C。
- 实验结果: UCI Drift 按目标域最优轮次回溯报告, 固定源 / 序贯迁移平均准确率 **81.10% / 84.36%**; Twin Arrays 采用统一前端、固定末轮五种子协议, 全周期输入的 UnifiedGas-C 与 CDAN 均达 **100%**。

 竞赛经历

---

ECCV 2026 SLoMO Workshop | 长视频理解与描述生成

任务背景: MovieQA 根据电影内容回答人物关系与剧情问题, 考察长程叙事理解; AD 根据视频片段生成面向视障观众的画面描述文本, 要求人物、动作与场景信息准确, 并满足对应片段的口述时长约束。

团队: ucas\_xlab AD 专项赛道亚军、主赛道季军 (官方最终榜单)

- AD Special | 亚军: 基于 Qwen3.5-35B-A3B 零样本推理, 融合时序上下文与密集帧采样, 在口述时长预算内生成场景描述; 以视觉内容覆盖与描述时长约束指导生成, Private AD Score **49.80**。
- AD Main | 季军: 以基准描述为锚点, 按风险筛选待改写片段, 并结合视觉证据与双判官盲评筛选候选; 控制描述替换对原有正确内容的影响, Private AD Score **53.47**。
- MovieQA Main | 第 4 名: 结合多尺度时序采样与字幕 / ASR 对齐, 检索支持剧情回答的跨片段证据; 依据问题类型融合多模型候选与共识, Private Accuracy **64.40%**。
- MovieQA Special | 第 5 名: 基于冻结的 Qwen3-VL-8B, 构建 8 / 64 / 128 帧多时序视图; 通过加权 Jaccard 共识选择聚合不同采样视图的候选答案, Private Accuracy **51.02%**。

 论文与预印本

---

- [1] Guideline-as-Oracle: Zero-Annotation Training of an Ophthalmic Telephone Triage Agent ([AAAI, under review](#), 第一作者)
- [2] UnifiedGas: End-to-End Unsupervised Domain Adaptation for Drift-Robust Gas Classification ([IEEE Sensors Journal, under review](#), 第一作者)

 专业技能

---

- 大模型与 Agent: 熟练使用 Python / PyTorch; 具备 SFT、GRPO 探索、RAG 与多 Agent 协作经验。
- 迁移学习与评测: 熟悉无监督领域自适应、特征工程、模型校准及对照实验; 使用 scikit-learn / XGBoost。
- AI 辅助科研: 熟练使用 Claude Code、Codex 开展算法实现、调试与论文复现。

 相关荣誉

---

- 中国科学院大学三好学生 (2026); 西安理工大学校级一等奖学金 (2020、2021、2022)。
- 西安理工大学创新成果三等奖 (2022); 优秀学生干部 (2020、2021、2022)。