

尹海森

☎ : 18080945965 | ✉ : 1627162415@qq.com

👤 : <https://hisensen.github.io/> (含数字人以及 demo, 电脑观赏更佳)



教育背景

北京邮电大学 (硕士) - 通信工程专业 信息与通信工程学院 2024-2027

北京邮电大学 (本科) - 通信工程专业 信息与通信工程学院 2020-2024

核心项目

蚂蚁集团 (Aigc 算法实习)

2026.05 - 至今

- 生成与自动评测: 基于 Qwen-Image-2512 训练 rank-32 商品/人物 LoRA, 构建 API 与自托管模型的混合路由层 (失败自动回退、成本与延迟埋点); 设计四维自动评测流水线——商品一致性 (DINOv2 实例级检索)、文字正确性 (OCR)、指令遵循 (checklist 式 VLM 问答)、营销价值, 并以 200 张人工标注验证各维度与人工评分的一致性。
- 偏好数据与 Reward Model: 构建 1,200 组 chosen/rejected 多模态偏好数据; 采用 Qwen3-VL-4B LoRA + 标量奖励头与 Bradley-Terry 损失训练 Reward Model, 在按商品隔离的测试集达到 72.4% pairwise accuracy; 完成 Rules、VLM 以及 Hybrid Reward 的消融, Best-of-N Top-1 人工命中率由 45% 提升至 60%。

清华大学 AIR 智能产业研究院 (强化学习算法实习生)

2024.11 - 2025.06

- GUI Agent 表示设计与 SFT (移动端): 将 Android view hierarchy XML 经剪枝、语义抽取与状态归一化重构为 UTG (界面转移图) 表示, 单屏上下文由数千 token 压缩至数百; 自研 code-as-action 动作 DSL, 把 AitW 的坐标手势标注转换为可执行代码, 使模型学习语义元素而非绝对坐标。基于 Qwen-7B 完成 SFT, 在 AitW 评测动作准确率, 并在 AndroidWorld 可交互环境完成端到端任务验证。
- 过程奖励建模: 针对任务级 0/1 结果奖励在低成功率场景下正样本极稀缺的问题, 尝试从 ORM 到 PRM, 设计「规则式过程奖励 + 结果锚定」的标签构造方案, 以 LoRA + reward head 将规则信号蒸馏为可泛化的 PRM, 替代只能在有标注环境中计算的规则函数; 基于 PPO 系统对比 GRPO、DPO 等主流强化学习的方案, 在 WebArena 测试平台进行测试。

AIGC 内容生成自媒体博主 (独立开发, 全网 40w 粉丝商业验证)

2024.07 - 至今

- 人物 LoRA 训练与保真一致性: 自建数据集训练人物 LoRA, 以 OpenPose ControlNet 约束动作、IP-Adapter 注入身份特征, 基于 Stable Diffusion 1.5 实现人物形象与动作一致的视频流渲染, 抑制逐帧生成常见的脸部漂移与服饰跳变。全自动内容生产管线: 自研改写 ComfyUI 开源工作流, 打通文案生成、视频渲染、字幕配置与一键发布。
- 商业闭环验证: 全网累计 40w 粉丝, 上述管线支撑内容规模化日更产出; 社群/私域转化 + 自研 App 应用落地, 独立跑通「选题与内容生产 → 多平台分发 → 私域沉淀 → 转化变现」完整链路, 累计会员 2000 名。

论文发表 (通信顶级会议两篇, 一篇在投)

- PIMRC (一作): 面向无人机多智能体路由, 利用 LLM 以训练回报为反馈信号迭代优化 MADDPG 奖励函数, 替代人工试错式奖励塑形, 改善模型收敛效果及复杂网络环境下的训练稳定性, 减少稀疏奖励引发的训练震荡。
- ICC (一作): 将 MADDPG 升级为 MAPPO, 引入逐步过程奖励以缓解稀疏回报下的信用分配困难、经验过滤及并行训练机制提升采样效率, 加速冗余路由任务收敛, 提升动态拓扑下的决策效率与鲁棒性。
- GLOBECOM (一作): 引入 Diffusion 建模路由决策分布以刻画多峰近优解, 设计截断式去噪采样策略减少反向扩散步数, 在生成质量与推理开销间做显式权衡, 兼顾路由质量的同时降低推理时延。

技能与奖项

国家奖学金; 全国大学生数学竞赛一等奖; 连续 3 年 “互联网+” 大创赛北京市一等奖; 全国大学生计算机设计大赛全国二等奖; 北京市 “三好学生”; 英语六级 口语 A; 华为昇腾图像和自然语言处理 (优秀营员); 个人集 (<https://hisensen.github.io/>)