

陈逸涵

chenyihan20241@iscas.ac.cn | 14759977748 | y1hanchen.github.io

教育背景

中国科学院软件研究所 · 中文信息处理实验室 · 电子信息硕士

2024.09 - 至今

主要研究方向: NLP、LLMs、Responsible AI 导师: 何笨 教授

南开大学 · 软件学院 · 软件工程学士

2020.09 - 2024.06

IELTS 7.0 / CET-6 560

工作经历

阿里巴巴 · 淘天集团 / 阿里妈妈 / 广告技术部 · 大模型应用算法实习生

2026.05 - 2026.08

• **广告文本内容安全**: 优化原粗召回涉政与色情词包, 微调 RoBERTa 构建粗排模块, 并以电商数据与通用黑样本微调 Qwen3.5-0.8B 构建精排模块; 在 RT 小于 2 秒约束下, 全链路误拦截率由 3.8% 降至 0.11%。

• **AIGC 广告图片合理性审核**: 将 20 余个人审标签归纳为图文不匹配、文字错误、人体畸形、违反常识与物理规律四类问题, 微调 Qwen3.5-2B 构建端到端模型; 上线后人审成本降低 50% 以上。

• **前后效果对比审核自动决策**: 针对夸大虚假中的功效前后对比违规点, 建设基于 VQ-VAE 的多中心粗召回模块与基于 LLM 的精排模块, 在 99% 召回率下达到 95% 精度。

• **Game Agent 比赛**: 基于科斯定理设计统一估值函数, 以双边收益上界搜索多资产最优交易组合, 并加入缓存、跨阶段预热与熔断机制, 最终获得比赛第一名。

科研经历

学术同行评审 AIGC 检测

2025.05 - 2026.01

• **研究背景**: 针对现有 AIGC 检测工具过度依赖语言风格、难以区分合规 AI 润色与实质性内容生成的问题, 提出关注内容而忽略风格的检测范式。

• **研究方法**: 构建覆盖 6 种人机交互模式、规模 30 万以上的细粒度数据集; 微调 ModernBERT 构建 CoCoDet, 引入代价敏感边界损失与多任务学习, 在人类文本误报率 1.31% 的同时达到 98.2% 平均 F1。

项目经历

文档结构化解析系统

2025.01 - 2025.09

• **基准构建与赛事主办**: 扩展 READoc 数据集并优化评估管线, 系统评估 20 余种 Pipeline 工具、MLLM 与专家模型的性能和推理速度, 支撑 CCKS 2025 文档结构化比赛。

• **模型部署与优化**: 统一部署实验室 OCR 服务, 通过分页、后处理和文本密度动态切分解决超长文档、跨页表格与高分辨率图像问题, 并针对嵌套图片等错误对 dots.ocr 进行 LoRA 微调。

领域内容安全审核系统

2025.10 - 2026.01

• **数据管线**: 基于推理不确定性筛选困难样本, 与行业数据融合构建多源数据集; 利用闭源模型扩展训练数据, 并基于字音字形规则进行对抗性扩增。

• **模型训练**: 对 Qwen3-4B 进行 SFT 蒸馏与 GRPO 两阶段训练, 设计准确性、格式约束与推理依据三维奖励函数, 测试集 F1 超过 90%。

专业技能

• **大模型技术**: 熟练使用 Python 与 Linux; 熟悉 Transformer 架构及大模型 SFT、RLHF 流程, 具备主流训练与推理框架的实战经验。

• **大模型应用**: 熟悉模型评估选型、RAG、Embedding 与 Reranker, 具备检索、重排及多模态模型训练和部署经验。

• **英语水平**: CET-6 560; IELTS 7.0, 具备英文文献阅读与学术写作能力。

论文发表

• [ACL 2026 Main] Yihan Chen, Jiawei Chen, Guozhao Mo, Xuanang Chen, et al. CoCoNUTS: Concentrating on Content while Neglecting Uninformative Textual Styles for AI-Generated Peer Review Detection.

• [Submitted to EMNLP 2026] Tianyun Zhong, Guozhao Mo, Yanjiang Liu, Yihan Chen, et al. Beyond Isolated Dots: Benchmarking Table Construction as Deep Knowledge Extraction.