詹沿

18857187718 | gannchuukaju@gmail.com
2001-01 | 男 | 2027届

教育经历

北京大学 985 211 双一流 2024年09月 - 2027年06月
软件工程 硕士 全日制 北京
- 主修课程： 机器学习 自然语言处理 大语言模型开发与应用 人工智能实践 算法分析与设计
- 绩点：3.71/4.0

北京林业大学 211 双一流 2020年09月 - 2024年07月
数字媒体专业 本科 北京
- 年级排名：2020-2021学年 2/37 | 2021-2022学年 3/37 | 2022-2023学年 1/37 | 总排名 2/37
- 校级荣誉：优秀学生奖学金 文体优秀奖学金 学术优秀奖学金 三好学生 优秀学生组织干事

个人简介

- 热爱人工智能，熟练掌握基本机器学习算法及大语言模型相关理论和算法
- 能熟练使用Python、PyTorch等编程语言和深度学习框架进行数据分析处理与模型训练
- 了解LLM相关技术，如RLHF、PPO、GRPO、DPO、VERL、Scaling Law、RAG、Agent、vLLM、Chain-of-Thought、MoE、Agent等
- 熟悉自然语言处理模型、大语言模型及多模态模型，如 BERT、GPT、LLaMA、ChatGLM、Qwen、DeepSeek等
- 熟悉CNN、ResNet、RNN、LSTM、GRU、Seq2seq、Transformer等经典深度学习算法
- 学习能力强，有良好的沟通能力，具备团队合作精神和较高执行力

实习经历

完美世界（北京）软件科技发展有限公司 2023年06月 - 2023年12月
NLP算法实习生 基于LLM的广告语识别 北京
- 项目背景：为防止玩家微信交流群中出现竞品广告，导致客户流失，需及时识别竞品、恶意及游戏相关广告。
- 解决方案：采用RAG+LoRA技术，结合检索增强与大模型泛化能力，通过NNI框架优化超参数，并使用vLLM部署提升效率。
- 项目成果：F1-score值提升至85%，成功应用于小时级别的离线批量广告语识别任务。

深圳市腾讯计算机系统有限公司 2025年10月 - 2026年06月
大模型算法研究实习生 PCG社交平台架构中心 QQ算法组 深圳
- 项目背景： QQ海外版聊天系统需要实现类似微信“边写边译”的快速翻译功能，要求在欧亚语种翻译准确、推理速度快，且能够适应较短口语化句子的翻译场景。现有的混元翻译模型在亚洲语言存在短板，需独立制作数据集并训练翻译小模型。
- 解决方案： 以高质量影视对白与语音对话数据为主、文学数据为辅，构建横跨欧亚的13种语言互译数据100万条，涉及数据筛选与清洗、缺失语种蒸馏补齐、NER实体修复等规范化。同时对各家4B级别小模型进行翻译能力与推理延迟测评，最终选择训练基座为gemma-4-E4B进行SFT。
- 项目成果：模型域内性能提升约10个百分点，域外FLORES-200与WMT25（GEMBA/XCOMET指标）通用翻译测评均与Hy-MT2-30B-A3B持平。

项目经历

科研审稿大模型 2024年08月 - 2025年01月
- 项目背景：从OpenReview爬取并聚合论文与评论数据，目标是生成结构化的、包含总结、优缺点分析、改进建议和评审结果的统一格式review。
- 解决方案：通过GPT聚合多条review数据，使用QLoRA微调LLaMA-3-8B-Instruct-8K+PI模型学习格式化，并进行DPO训练确保输出一致性和准确性。
- 项目成果：经DPO训练后，模型生成的review与人工review的观点匹配率超越GPT-4审稿人效果，提升了自动化审稿的准确性。

科研项目

TrustRoboReward: Preference-Ordered Isotonic Score Editing for Multi-Paradigm Robot Reward Models 2025年07月 - 2026年05月
共一 NeurIPS 2026在投
- 项目背景：RoboReward 主要采用 1-5 分轨迹进度打分，缺少 pairwise 偏好监督和视频场景理解维度；直接引入多范式监督又会产生 pointwise score 与 pairwise preference 之间的冲突，影响奖励模型训练稳定性。本研究面向具身任务构建更通用、稳定的多范式多模态奖励模型。
- 实验方案：基于 RoboReward 视频样本构建 Score-A、Score-B、Pair-A、Pair-B 四类监督数据，分别覆盖轨迹进度打分、Video-QA 答案质量打分、轨迹视频偏好比较和 VQA 答案偏好比较。针对 score-pair 反转冲突，引入 POISE 一致性校正方法，将 pairwise 标签建模为偏序约束，并通过 isotonic regression / PAVA 校正 pointwise 分数；对少量 pairwise 环和传递性冲突进行修复，并将清洗后的数据用于 SFT 与 GRPO 训练。
- 研究意义：通过多范式监督扩展奖励模型的评价维度，并通过显式一致性校正降低训练语料中的冲突信号，提升模型评价稳定性和可信度。实验中训练语料score-pair冲突可降至0%，同参数模型综合指标达到78%，较RoboReward-4B提升10%，测试时score-pair一致率提升至72%。

SLCA-GRPO：Resolving Cross-Segment Credit Misattribution in Tool-Calling RL 2025年09月 - 2026年05月
一作 NeurIPS 2026在投
- 项目背景：Tool-Calling 智能体输出同时包含结构化工具调用段与自然语言总结段，标准 GRPO 将轨迹级优势统一广播至所有 token，容易导致总结质量对工具调用 token 产生跨段优势污染，进而造成错误归因与训练不稳定。
- 解决方案：提出SLCA-GRPO框架，对工具段与总结段奖励独立归一化并分别路由，将执行奖励路由至工具调用 token、偏好奖励路由至总结 token，并基于Schema约束的LLM模拟器实现可扩展API交互训练与利用Golden ToolCalling构建分层稠密奖励。
- 项目成果：在7B模型上相比标准 GRPO、ToolPO 和 RLTR，域内评测提升 +2.53pp，BFCL 提升 +1.36pp，τ^2-Bench 提升 +9.15pp，有效抑制冗余工具调用；论文投稿后尝试将训练方式用于agent类数据，并将分段评分思路迁移至业务Eval系统，搭建工具调用与总结质量独立评分体系。

Length-Adaptive Decoding for Masked Diffusion Machine Translation 2026年01月 - 2026年05月
一作 EMNLP 2026在投
- 项目背景：掩码扩散语言模型在机器翻译中需预先确定目标端长度，长度过短会导致信息遗漏，过长则容易产生重复、幻觉或冗余。现有解码研究主要关注token揭示顺序，对目标长度选择这一关键瓶颈探索不足。本研究旨在无训练条件下，提升扩散式机器翻译长度自适应能力与翻译完整性。
- 实验方案：提出训练无关的 Entropy-Valley（EV）长度自适应解码方法，通过候选长度构造与全掩码预测熵打分，为扩散式机器翻译动态选择目标长度。在 LLaDA-8B SFT 的英中、中英、英德任务上，与固定比例长度解码、以参考译文长度作为理论上限的解码结果进行对比，验证该方法能够在不依赖参考译文长度的情况下逼近理想长度选择效果，并结合自动指标、人工评测和跨模型实验分析其稳定性。
- 研究意义：证明了目标长度选择是固定画布掩码扩散机器翻译中的核心性能瓶颈，且无需新增训练目标、长度预测器或外部模型，即可显著提升翻译充分性与覆盖率。该方法为扩散语言模型在机器翻译等长度敏感生成任务中的解码优化提供了轻量、可迁移的解决方案。

荣誉奖项

第八届中国国际“互联网+”大学生创新创业大赛北京赛区复赛高教主赛道一等奖 2022.08
全国大学生数字媒体科技作品及创意竞赛全国总决赛二、三等奖 2021.11
2021全国建模大赛年度竞赛二等奖 2021.12

其他

- **语言：** 英语（CET-4），英语（CET-6）