|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学生姓名 | |  | | 学号 | |  | |
| 选题学院 | |  | | 专业班级 | |  | |
| 课题中文名称 | | 引入质疑者角色的多代理协作框架优化与性能提升研究 | | 课题英文名称 | | Performance Enhancement of Multi-Agent Debate Framework with Questioner Role | |
| 指导教师 | |  | | 题目类型 | |  | |
| 进行方式 | |  | | | | | |
| 学期 | 第二学期（春季） | | 起始周 | 1 | 结束周 | | 16 |
| 课题来源 | |  | | | | | |
| 目的要求 | | 本课题旨在研究引入“质疑者”角色对多代理协作框架评估能力的影响，通过分析“质疑者”在识别答复中的疏漏、逻辑漏洞和偏见等方面的作用，探索如何优化多代理协作辩论中的角色设置，提升框架的评估全面性和一致性。具体要求如下：   1. 明确质疑者在多代理协作框架中的定位与任务。 2. 探讨质疑者质疑内容的选取策略，包括但不限于答复中的疏漏、逻辑漏洞、偏见和任务符合性。 3. 设计并实现质疑者角色的具体提示和交互逻辑。 4. 通过实验验证质疑者的引入对整体评估性能的提升效果。 5. 提供优化建议，为未来多代理协作评估框架设计提供参考。 | | | | | |
| 主要内容 | | 多代理协作框架近年来在提升文本生成任务评估质量方面显示出了巨大潜力。尤其是《CHATEVAL: TOWARDS BETTER LLM-BASED EVALUATORS THROUGH MULTI-AGENT DEBATE》这篇文章提出了通过多代理辩论改进大型语言模型评估能力的设想，该框架模拟了人类评估者协作辩论的过程，从而在评估的深度和广度上接近人类水平。  在此基础上，本课题进一步探讨多代理协作框架中的角色优化问题，特别是引入一个“质疑者”角色的可能性。质疑者的核心任务是专注于发现协作者讨论中的潜在问题，例如答复中的疏漏、逻辑漏洞以及偏见问题。通过对讨论内容进行批判性分析，质疑者可以提出针对性的质疑和修正意见，从而弥补框架在多视角协作中的不足。  本课题将从以下几个方面展开：   1. **研究背景与问题定义：** 结合现有文献，分析当前多代理协作框架的优势与不足，尤其是其在评估全面性和一致性方面可能存在的问题。 2. **质疑者角色的任务设计：** 基于提示工程的方法，为质疑者角色构建明确的任务描述，使其能够有效识别疏漏与偏见，并在协作过程中提出具有建设性的改进意见。 3. **质疑者的行为模式设计：** 探讨质疑者在多代理协作框架中的交互逻辑，包括如何与其他协作者进行信息交换、何时介入讨论以及如何优化其质疑行为。 4. **实验验证：** 通过构建对比实验（如无质疑者框架 vs. 有质疑者框架），在开放性问题回答和文本生成评估任务中验证质疑者角色的有效性。分析其对评估准确性、一致性和覆盖率的影响。 5. **结果分析与优化建议：** 通过实验结果总结质疑者角色的实际作用，讨论其可能的副作用（如过度质疑或偏离任务目标），并提出改进建议。   本课题的研究不仅可以丰富多代理协作框架的角色设置，还能为基于大型语言模型的自动化评估提供新的思路和技术支持。 | | | | | |
| 预期目标 | | 1. 设计并实现适合多代理协作框架的质疑者角色提示和交互逻辑。 2. 通过实验验证质疑者角色对框架评估全面性和一致性的显著提升。 3. 提供系统化的优化策略，为多代理协作框架的设计与应用提供理论与实践支持。 4. 发表相关研究成果，推动LLM自动化评估方法的发展。 | | | | | |
| 经费 | |  | | | | | |
| 参考资料 | | ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate, ICLR 2024 | | | | | |
| 实验加工  条件 | |  | | | | | |