

## 引言

存在问题	优化建议	工程文件引用（如需）
<b>背景交代不充分：</b> 缺少对研究数据集和任务背景的具体信息支撑，如未提及FinDER数据集规模等。	<b>补充背景细节：</b> 增加FinDER数据集的介绍及规模数据，以体现问题重要性。例如说明FinDER包含大量真实金融问答（三元组数量）并强调复杂查询比例。	金融AI选题评价.pdf 第1页：“FinDER由金融领域专家构建，包含5,703个真实金融问答三元组。”
<b>现有方法局限性描述不足：</b> 对当前金融问答系统存在的问题着墨不够，未充分引出多步检索和数值计算的必要性。	<b>突出研究动机：</b> 在引言中增加对传统单步检索RAG系统不足之处的描述，指出面对跨段落、多文档查询时性能不足，从而引出本研究引入多步检索和计算模块的动机。	金融AI选题评价.pdf 第1页：现有金融RAG系统主要针对简单查询，复杂查询时“仍存在性能不足的现象”。
<b>创新点陈述不清晰：</b> 引言对本文的贡献和创新缺乏明确列举，读者难以抓住论文亮点。	<b>明确列出贡献：</b> 在引言末尾以梳理方式陈述本文创新点，例如：① 提出多步检索推理机制提升跨段落问答能力；② 引入数值计算模块处理定量问题；③ 构建开源可复现的金融RAG系统等，确保每一点清晰可见。	(无)
<b>支撑引用不足：</b> 引言大量基于内部资料描述，缺少权威文献佐证；可能存在引用未闭环（引用了内部报告但未正式列入参考文献）。	<b>加强引用规范：</b> 引入至少1-2篇相关领域文献（如RAG方法或FinDER数据集论文）的引用来支撑背景和动机，避免只有内部资料。检查参考文献列表，移除未引用文献，补全缺失的引用信息，确保引言中的数据和论断都有出处。	(例如可引用 FinDER 数据集的公开论文和 RAG 原始论文)
<b>逻辑顺序有跳跃：</b> 引言可能直接进入方法细节或结果，缺少层层递进的铺垫。	<b>理清引言结构：</b> 调整段落顺序，遵循“背景现状→问题挑战→拟解决方案→贡献摘要”的逻辑。避免过早涉及技术细节或结论，在引言中保持宏观概述，将具体实现和结果留到后续章节。	(无)
<b>术语缩写未解释：</b> 部分专业术语/缩写（如“RAG”等）在引言首次出现时未作解释，可能导致读者困惑。	<b>补充术语解释：</b> 对于首次出现的缩写，添加英文全称和简要解释。例如说明RAG是Retrieval-Augmented Generation（检索增强生成）的方法，以确保读者理解关键概念。	(无)
<b>论文结构未概述（可选）：</b> 引言结尾缺少对全文结构的概览介绍。	<b>增加章节概览：</b> 在引言末尾增加一段，简要说明各章节内容安排，例如“第二章介绍方法，第三章为实验设计与结果，第四章给出结论与展望”等，提升论文可读性。	(无)

## 方法

存在问题	优化建议	工程文件引用（如需）
<b>章节结构不清晰：</b> 方法部分缺少合理的小节划分，所有内容堆积在一起，读者不易抓住重点。	<b>细分方法章节结构：</b> 根据系统流程将方法划分章节，例如“2.1 系统总体架构”、“2.2 文本检索模块”、“2.3 多步检索策略”、“2.4 数值计算模块”等，每部分聚焦一种模块或步骤，层次清晰。	（无）
<b>缺少流程概览：</b> 未提供系统整体流程的描述或图示，读者难以理解方法各部分如何衔接。	<b>增加方法概览：</b> 在方法开头增加整体框架介绍，必要时附加流程图。例如描述从用户查询出发，经过文档检索→多步推理→数值计算→答案生成的完整流程，为后续细节铺垫。	（无）
<b>检索模块细节不足：</b> 对基础检索方法的说明不充分，如未交代使用何种检索算法和参数。	<b>补充检索实现：</b> 说明检索模块采用了何种技术。例如：“本研究使用混合检索策略，结合 BM25 算法和句向量检索（基于预训练SBERT模型）提高召回率。”可提及主要参数设置（如BM25的ngram范围等）以增加可解释性。	src/retrieval/retriever.py: 实现了 HybridRetriever，将 BM25 与 SBERT (all-MiniLM-L6-v2) 向量检索相结合。
<b>多步检索机制不明确：</b> 论文未充分解释多步检索是如何实现的，如查询迭代策略和终止条件。	<b>详述多步检索策略：</b> 增加对多步检索算法的说明。例如阐明系统如何利用初步检索结果生成后续查询，是否采用 Planner/Refiner 机制，以及设置何种停止策略（如检索轮次上限或相关性阈值）。确保读者理解多步检索的工作流程和创新之处。	金融AI选题评价.pdf 第1页：“多步检索允许系统像分析员一样逐步深入…反复迭代，直到汇集足够证据支持最终回答。”
<b>数值计算模块交代不详：</b> 针对需要计算的问答缺少说明系统如何提取并计算数值答案。	<b>介绍数值计算模块：</b> 补充说明在检索到相关证据后，如何通过数值模块提取关键数值并执行运算（如同比增长计算）。可以说明模块由“数值抽取”和“数值计算”两个步骤组成，保障读者了解其功能和作用。	金融AI选题评价.pdf 第1页：“在推理过程中融入显式的算术计算模块，也能提升数值类问答的准确度。” src/calculator/: extract.py 提供数值提取功能，compute.py 实现同比增长等运算。
<b>算法引用缺失：</b> 方法中涉及的经典算法或模型（如 BM25、SBERT 等）未引用原始论文或来源。	<b>补充理论引用：</b> 在描述 BM25 公式、SBERT 模型等关键方法时，加上相应引用（如 BM25 原始论文、SBERT 论文），以彰显研究的严谨性和与现有工作的衔接。	（无）
<b>关键参数和设置未说明：</b> 对实验再现重要的设置如随机种子、检索 top K 等未在方法或实验部分提及。	<b>增加参数说明：</b> 明确文中使用的重要参数。例如说明“为了保证结果可复现，所有实验固定随机种子为42，检索阶段取 top-5 文档进行回答生成”等。这有助于他人理解实验条件。	scripts/smoke.py: 默认随机种子为42（第121-124行），当 subset_size<=0 时设为20（第142-145行），以及检索采用 Recall@5 即 top_k=5（第149-151行）。

存在问题	优化建议	工程文件引用（如需）
<b>内容衔接不够连贯</b> （若存在）：各模块描述之间过渡生硬，读者可能难以跟随。	<b>增强段落衔接：</b> 在各小节开头或结尾增加过渡句，解释模块之间的关系。例如，在多步检索部分开头先承接说明它与基础检索的区别，在数值模块部分说明其在多步检索结果基础上处理定量问题。确保方法章节形成一个完整的逻辑链条。	（无）

## 实验设计与结果分析

存在问题	优化建议	工程文件引用（如需）
<b>实验设置描述不完整：</b> 缺少对数据集和实验环境的详细说明，例如未明确FinDER数据集的规模、来源或划分，以及实验所用模型、硬件等信息。	<b>补充实验背景：</b> 在实验设计开头交代数据集和环境。说明FinDER数据集的基本情况（问答对数量、复杂查询比例等）以及训练/测试集划分方法。如有使用预训练模型或特定硬件（GPU）也应简要说明，确保实验具有可理解的基础。	金融AI选题评价.pdf 第1页：FinDER数据集由真实金融文件构建，包含5,703个问答三元组，涵盖大量复杂查询。
<b>缺少对照与消融实验：</b> 仅给出了单一方案的结果，未比较基线方法或拆解新模块的作用，难以量化方法改进。	<b>增加对比实验：</b> 引入基线和增量实验以验证方法有效性。例如报告“无多步检索/无数值模块”的基线性能，与完整模型表现对比；或分别移除多步检索、数值模块观察指标变化。通过对照实验清晰展示每个创新点带来的性能提升。	（无）
<b>评价指标定义不明确：</b> 文中直接给出Recall、Exact Match等指标结果，但未解释这些指标含义或计算方法。	<b>阐明评价指标：</b> 在结果呈现前添加指标说明。例如：“本文采用Recall@5（检索前5命中相关答案的比例）和Exact Match（答案完全匹配率）作为主要评价指标。”确保读者了解指标意义，再呈现具体数值。	README.md 第33-36行：项目评估指标包括Recall@5和Exact Match等基础问答指标。
<b>结果图表支撑不足：</b> 实验结果的表格/图缺少必要信息，或图表尚未完整（例如缺少误差项、统计显著性等）。	<b>完善结果呈现：</b> 确保每个表格和图都有清晰的标题和注释，列明指标单位和含义。如果条件允许，报告多次实验的平均值及标准差，或进行简单显著性检验，以增强结果说服力。另外，检查图表是否有遗漏（如需要展示趋势的曲线图或案例图示），视需要补充。	（无）

存在问题	优化建议	工程文件引用（如需）
<b>结果分析不深入：</b> 对实验结果只做罗列，缺乏针对性的分析和解释。	<b>加强结果讨论：</b> 对主要发现进行解释。例如，当多步检索比基线Recall提高时，分析其原因（如覆盖更多相关证据段落）；数值模块提升了定量查询的准确率时，应指出该模块有效弥补了什么不足。通过将定量结果与方法特点关联，凸显论文结论的意义。	（无）
<b>缺少案例或误差分析：</b> 论文未展示具体问答案例或错误分析来直观说明方法优劣。	<b>补充质性分析：</b> 选取1-2个具有代表性的复杂提问案例进行深入分析。展示系统的检索和推理过程（可引用多步检索的中间结果），对比正确答案与模型答案，分析错误原因或成功关键。这样的案例能帮助读者理解算法如何处理实际问题，增进说服力。	outputs/<run_id>/multistep_traces.jsonl：记录了多步检索每轮查询及检索结果，可供提取典型案例进行分析说明。
<b>可复现性说明不足：</b> 没有提及实验代码开放或复现实验的要点，使他人难以重复研究。	<b>强调复现细节：</b> 在实验部分说明为保障结果复现所做的工作。如注明“所有实验代码和配置已在仓库公开，包含运行脚本与参数设置”，并提供项目仓库链接或附录引用。这体现研究的开放性，方便读者验证结果。	（无）

## 结论

存在问题	优化建议	工程文件引用（如需）
<b>总结内容不充分：</b> 结论过于简短笼统，只重复了引言内容，未突出研究所得的具体成果。	<b>充实结论要点：</b> 精炼总结全文贡献与发现。重申研究解决的问题和实现的主要创新，给出关键结果数据（例如性能提升了多少个百分比），突出本研究相较于现有工作的改进程度。使结论部分展现论文的价值精华。	（无）
<b>未呼应研究目标：</b> 结论中未明确回答引言提出的研究问题或目标，导致首尾呼应欠缺。	<b>首尾呼应：</b> 在结论中明确指出本研究达成了引言所定目标。例如：“本文针对复杂金融查询提出的多步检索+计算方法，经验证显著提高了检索召回率和问答准确率，成功应对了引言中提出的跨段落推理和数值计算挑战。” 这样确保论文在结论处对最初的问题做出回应。	（无）
<b>结论出现新信息：</b> 结论段落包含正文未提及的新方法或数据（例如提出新的想法却未在论文中展开）。	<b>确保内容闭环：</b> 删除或调整任何在正文中未详细讨论就出现在结论的新内容。结论应仅总结已有论述。若确需提及新的展望或设想，应明确其属于未来工作展望，而非本研究结论的一部分。	（无）
<b>缺少展望或局限讨论：</b> 对研究的局限性和未来工作没有交代，结论显得不够完整。	<b>增加未来展望：</b> 在结论末增加对后续工作的简短展望。例如指出本研究的局限（如仅在单一数据集上验证），以及未来可以如何改进（如引入更智能的LLM Agent策略、扩展到更多样的金融语料等），为读者提供延伸思考空间。	（无）

存在问题	优化建议	工程文件引用（如需）
<b>复现性与开放性未提及：</b> 结论中没有强调代码开放和结果可复现，这一研究亮点可能被忽略。	<b>强调开放价值：</b> 在结论中重申本项目已开源且可复现，邀请业界和学术界使用与检验。本可作为论文贡献的一部分，体现研究对社区的贡献和影响力。	（无）