

基于知识库检索反馈的期货市场智能合规方法研究

周俊杰¹,赵文字²,戴伟辉³

¹(复旦大学 软件学院 上海市数据科学重点实验室,上海 200433)

²(上海大学 计算机工程与科学学院,上海 200444)

³(复旦大学 管理学院,上海 200433)

E-mail:jiz_dragon@hotmail.com

摘要:随着人工智能在高合规性领域的深入应用,如何在保障模型表达能够满足严格法规约束,成为大语言模型面临的核心挑战。针对现有方法存在的标注成本高、难以适应动态监管场景及偏好与规则冲突等问题,本文提出一种基于合规知识库反馈的强化学习微调框架(RLKBF, Reinforcement Learning from Knowledge Base Feedback)。该方法结合语义增强的向量表示与层次化检索机制,构建结构化法规知识库。同时,引入合规偏离惩罚项的双目标优化策略,以协调用户意图与法规规范的平衡。实验结果显示,RLKBF在回答准确率、合规性稳定性及专家评估指标上均优于主流对比模型,显著提升了模型对专业法规知识的整合与应用能力。

关键词:期货市场;大语言模型;模型微调;知识库构建;合规管理

中图分类号:TP391

文献标识码:A

文章编号:1000-1220(2026)05-1048-08

Research on the Intelligent Compliance Methods in the Futures Market Based on Knowledge Base Retrieval Feedback

ZHOU Junjie¹, ZHAO Wenyu², DAI Weihui³

¹(School of Software, Shanghai Key Laboratory of Data Science, Fudan University, Shanghai 200433, China)

²(School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China)

³(School of Management, Fudan University, Shanghai 200433, China)

Abstract: As artificial intelligence continues to be applied in highly regulated domains, ensuring that large language models comply with strict legal constraints has become a central challenge. Existing approaches often face limitations such as high annotation costs, poor adaptability to evolving regulatory environments, and conflicts between user preferences and compliance requirements. To address these issues, this paper proposes a fine-tuning framework based on Reinforcement Learning from Knowledge Base Feedback (RLKBF). This method constructs a structured legal knowledge base using semantically enhanced vector representations and hierarchical retrieval mechanisms. Additionally, a dual-objective optimization strategy is introduced, incorporating a compliance deviation penalty to balance user intent with regulatory constraints. Experimental results demonstrate that RLKBF outperforms mainstream models in terms of response accuracy, compliance stability, and expert evaluation metrics, significantly improving the model's ability to integrate and apply domain-specific legal knowledge.

Keywords: futures market; large language models; model fine-tuning; knowledge base construction; compliance management

0 引言

随着全球金融市场的不断发展与数字化转型,期货市场的交易规模和风险管理需求持续增长。期货交易涵盖高杠杆、高频算法交易与跨市场套利等行为,提升了监管的难度。传统风控系统通常依赖静态规则库与人工审核机制来监控交易行为。然而,在市场环境日益复杂和动态变化的背景下,这类手段面临明显瓶颈:规则更新依赖人工维护,响应滞后,且在处理新型交易模式和跨市场监管逻辑时易产生合规误判。特别是在期货市场,不同交易所与监管机构的法规差异显著,更新频率高,市场行为多变,进一步加剧了传统规则系统的适应性

挑战与维护成本压力。因此,迫切需要引入具备自适应能力的智能化技术,以实现对期货市场复杂交易行为的高效识别与合规判断,提升风控系统的实时性与准确性。

近年来,人工智能技术,尤其是大语言模型(Large Language Models, LLMs)在自然语言处理领域取得显著突破,为金融合规任务中的文本解析与决策支持带来了新机遇。大语言模型凭借其在上下文理解、语义生成及推理能力方面的优势,具备较强的通用性与可扩展性。在金融领域,已有多个面向特定任务的金融语言模型取得初步成效。例如,FinBERT通过在金融情感标注数据集上的微调,实现了对财报、新闻中的情绪分析^[1];BloombergGPT基于大规模金融数据预训练,

增强了模型在专业术语与金融事件处理方面的能力^[2]; InvestLM 则通过对 LLaMA 模型的指令微调,实现了投资文本情绪的精准预测,亦被用于异常交易识别与因果路径分析等任务中^[3-5]. 这些尝试在特定场景中展现出良好性能,标志着大模型在金融应用中的实践价值. 然而,它们多数仍存在以下共性问题:1) 模型能力严重依赖训练语料的覆盖范围与质量,难以适应监管规则的动态变动;2) 缺乏对合规推理逻辑的建模能力,模型生成结果在合法性、稳定性与可解释性方面仍存在不足^[6]. 因此,尽管这些模型为大语言模型在金融风控中的应用奠定了基础,但其在期货市场等高风险、强合规场景下的实用性仍待提升.

为提升大模型在合规风控中的适应能力,当前研究主要沿两条技术路径展开探索:1) 基于知识检索机制的增强式生成方法;2) 引入反馈信号的强化学习方法. Retrieval-Augmented Generation(RAG) 框架通过结合外部知识检索与语言生成,为大模型引入可调用的事实信息,增强其输出的准确性与可溯源性. 在期货市场等合规场景中,法规更新频繁、跨机构逻辑差异大,单靠预训练模型参数难以同时满足精度与时效性要求. RAG 方法可接入结构化或非结构化的法规文档、监管公告等知识库,提升模型在法规条款解析、交易行为审查等任务中的适应能力^[7,8]. 已有研究已在财报解读、公告摘要生成、金融问答等任务中验证了 RAG 方法对一致性和可解释性的提升作用,部分工作还结合多源异构数据支持市场动态监控与合规预警^[9-12]. 此外,结合半监督聚类等方法,RAG 也展现出在复杂金融风险建模中的灵活性与低标注成本优势^[13]. 但 RAG 在实际应用中仍存在局限. 一方面,检索效果高度依赖于知识库的数据结构与更新机制,金融领域中非结构化文本比例高、语义噪声干扰大^[14];另一方面,生成模型与检索结果之间的引用关系不稳定,在具体合规推理任务中可能出现偏差或事实错配. 此外,RAG 框架缺乏反馈机制,对动态变化和复杂对抗性场景响应能力不足,存在合规逻辑断裂风险^[15-17].

与 RAG 方法互补的是强化学习路径,研究者通过引入外部反馈信号对模型行为进行优化,以提升其合规性与边界敏感性. 其中,RLHF(Reinforcement Learning from Human Feedback) 通过专家标注强化引导生成内容,使其更符合监管逻辑与专家判断,在异常交易检测、市场操纵识别和自动化审查等任务中展现出较好控制效果,尤其适用于法规模糊区域的约束表达^[18]. 但 RLHF 存在现实制约:金融领域知识更新快、专属性强,人工标注不仅成本高,且受主观因素影响较大,训练样本难以形成高一致性与覆盖性. 在高频交易环境中,RLHF 的训练与更新周期难以匹配实际业务需求. 为缓解上述问题,RLAIF(Reinforcement Learning with AI Feedback) 被提出,尝试由模型自身生成反馈结果进行自我强化训练,具有更高的数据扩展性与训练效率^[19]. 然而,其反馈信号源于初始模型,容易形成语义闭环. 一旦模型存在知识偏差,偏好信号可能被放大,输出结果虽然表面合理但可能偏离合规逻辑核心.

综上,RAG、RLHF 与 RLAIF 分别在知识接入、行为对齐与反馈效率上提供了技术路径,但三者皆未能从根本上解决“动态规则适应”、“高风险边界识别”与“语义稳定性约束”的合规核心难题. 这一困境的根本原因在于现有主流方法均

以模型自身为优化中心,缺乏对外部制度体系,即法规知识的结构性建模与动态约束能力,已有研究尝试将语义结构与知识图谱融合,为规则建模提供更高表达能力的嵌入基础^[20]. 在期货市场这种规则高频变动、违规成本极高的环境下,合规判断不能仅依赖语言表达相似性,更需基于知识逻辑路径构建刚性审查机制^[21]. 因此,如何以法规知识为主导信号重构模型学习机制,在保持语义生成能力的同时,注入可解释、可追踪的合规行为边界,是当前亟待突破的关键技术挑战^[22]. 为此,本文提出一种基于法规知识库反馈的强化学习微调方法(RLKBF),以法规知识为主导信号重构模型学习机制. 该方法通过构建结构化合规知识库、引入语义增强的向量表示与层次化检索机制,以及融合合规偏离惩罚的双目标优化策略,实现大模型在高合规性场景下的行为可控性、响应准确性与知识适应性三重提升.

1 问题定义

在期货市场智能风控场景中,模型在执行交易行为合规性判断的过程中,需同时满足动态监管法规的严谨约束与用户操作偏好的实时适配. 这一问题本质上涉及两个核心子任务:1) 如何构建具有高适应性、高表达力的法规知识库,用以增强模型对监管意图的理解与追踪能力;2) 如何通过强化学习等机制,实现在合规性与偏好性约束下的大模型微调,以生成兼顾合规约束与用户预期的合规建议.

1.1 知识库构建问题定义

为提升大模型对合规规则的适配能力,有必要构建一个语义增强的法规知识库,使其具备以下关键能力:1) 法规条款的语义结构建模;2) 不同交易对象间的规则适用性区分;

假设存在一组原始法规文档集合 $\mathcal{D} = \{d_1, d_2, \dots, d_n\}$, 每个文档 d_i 包含若干条款、章节与适用条件,目标是将其转化为支持语义检索的结构化知识单元集合 $\mathcal{K} = \{k_1, k_2, \dots, k_m\}$, 其中每个知识单元 k_j 表示为:

$$k_j = \langle e_j, a_j, l_j \rangle \quad (1)$$

其中, e_j 表示经嵌入模型编码后的语义向量, a_j 表示该条款的适用主体标签(如个人投资者、机构、交易所等), l_j 表示条款所在的法规层级标签.

知识库构建问题的目标在于设计一个函数, $\mathcal{F}: \mathcal{D} \rightarrow \mathcal{K}$ 满足以下优化目标:

语义完整性最大化:转换后的 k_j 能保留原法规条款的逻辑关系与约束信息.

适用性区分准确性最大化:不同主体查询时,返回的法规条款在上下文和适用性上保持高度相关.

1.2 模型微调问题定义

模型需在满足用户操作偏好的同时,确保生成内容符合法规. 该问题可形式化为一个合规-偏好双目标强化学习问题,旨在通过引入知识库反馈信号,对语言模型进行微调优化.

假设模型生成的输出记作 $y = f_{\theta}(x)$, 其中 x 为输入交易请求或查询, θ 为模型参数. 目标是通过优化 θ , 使得输出 y 在两个维度上获得最优表现:

偏好得分 $P(y|x)$, 由偏好数据集给出;

合规得分 $C(y|x)$, 由知识库驱动的法规评分器返回。

为平衡上述两个目标, 模型的优化目标被定义为一个加权组合, 其中包含了两个损失函数: 偏好损失函数和合规性损失函数。偏好损失函数反映了模型输出与偏好之间的差异, 通过从提供的数据中学习得到的; 合规性损失函数反映了模型输出与预设的合规标准之间的差异, 通过引入合规知识库规则来进行评估。

两部分损失函数通过一个权重系数进行加权, 确保在优化过程中既能考虑到偏好损失, 也能保证合规性。权重系数通常是一个在 $0 \sim 1$ 之间的值 (λ), 用于平衡这两者的重要性。 λ 的值根据实际需求来调整, 如果合规偏好比合规性差异更为重要, 则 λ 值较高; 如果合规性差异要求更为严格, 则值较低。

2 期货交易合规知识库的设计优化

随着金融监管环境的不断演进, 尤其在期货交易领域, 合规管理对法规响应的实时性、检索的准确性及知识的可维护性提出了更高要求。传统的法规管理方式多依赖静态规则与人工归档, 难以适应当前法规高频变更、层级交叉适用及专业术语丰富的特点。传统方法如基于 BM25 的法规检索和结合倒排索引的法规辅助推理, 分别通过关键词权重匹配和相关文档提取以增强法规响应能力, 在初步应用中取得了一定效果。但上述方法均存在如下问题: 法规文本通常具有复杂的层级结构与跨规则适用性, 仅靠关键词匹配无法有效区分法规条款的上下位关系及适用主体, 易导致检索偏差。

针对上述问题, 近年来越来越多的研究将大语言模型 (LLM)、向量数据库 (Vector Database) 等新技术应用于法规知识建模与管理中。本文结合这一趋势, 围绕期货市场监管场景, 提出了一种面向智能风控需求的合规知识库构建方案。

本文围绕期货市场监管场景, 提出了一种面向智能风控需求的合规知识库构建方案, 涵盖结构建模、语义增强与高效向量检索机制, 系统阐述了合规知识库的结构设计与语义增强方法。在构建流程方面, 首先提出基于法规文档结构识别、命名实体提取和层级标签建模的知识抽取与结构化机制, 实现法规条款从非结构化文本到规范化知识单元的转化; 随后, 基于向量化检索的需求, 引入多维语义增强策略, 将法规的文本语义、层级信息、适用性标签及术语权重进行拼接融合, 构成语义表达更完整的法规向量表示。本文基于向量数据库构建了完整的知识库系统, 并在期货市场法规数据集上进行了系统实验, 验证了本方法在检索准确率、响应时效性等方面较传统方法具有明显优势。

2.1 文本向量化的法规语义增强

在基本文本向量化的基础上, 本节提出了一种语义增强的法规向量化方案, 旨在结合法规的层级信息、适用性标签和核心术语加权机制, 提升法规文本的表示能力和检索效果, 如算法 1 所示。通过融合多维信息, 该方法不仅能够提高法规向量的语义表达能力, 还能在检索时更准确地匹配法规的上下层级条款, 确保返回的法规结果更具业务相关性。

算法 1. Enhanced Legal Semantic Vector Construction

1. **Procedure** ENHANCED_LEGAL_VECTOR(input)

```

2.   level_tag ← Extract legal regulation level tag(input)
3.   entity_tags ← Extract applicable entity tags(input)
4.   key_terms ← Extract key legal terms(input)
5.   vec_text ← Text embedding(input)
6.   vec_level ← Embedding(level_tag)
7.   vec_entity ← Initialize vector()
8.   for each tag ∈ entity_tags do
9.     vec_entity ← vec_entity ⊕ Embedding(tag)
10.  end for
11.  vec_terms ← Initialize vector()
12.  total_weight ← 0
13.  for each term ∈ key_terms do
14.    w ← TF-IDF(term, doc)
15.    v ← Embedding(term)
16.    vec_terms ← vec_terms + w × v
17.    total_weight ← total_weight + w
18.  end for
19.  vec_terms ← vec_terms / total_weight //Normalization
20.  vec_fused ← FUSION(vec_text, vec_level, vec_entity, vec_terms) //Information fusion
21.  vec_final ← NORMALIZE(vec_fused) //L2 normalization
22.  vec_final ← MAP_TO_SPACE(vec_final) //Project to unified vector space
23.  return vec_final
24. end procedure

```

2.1.1 法规层级结构嵌入

本文提出了在文本向量化过程中引入期货法规层级编码的技术方案。这一方案通过对期货法规文本进行层级化标注, 使得规则在存储和检索过程中能够准确反映期货市场的特定场景和监管要求。

引入期货法规层级编码的技术方案能够有效避免跨规则问题, 主要通过精确区分法规的层级结构, 减少跨层级混淆, 提高检索的针对性。期货市场的规则通常由不同层级的法规构成, 包括国家级法规、交易所层级规则和具体品种的细化条款。通过为每条法规赋予层级标签, 如“交易所规则-郑商所-花生期货-限仓条款”, 系统可以在检索时优先匹配同品种、同交易所的规则, 而非简单依赖关键词匹配, 从而避免了跨规则误匹配的情况。此外, 层级编码还有效减少了跨层级混淆的可能性, 特别是在跨期套利或跨市场操作中, 能准确识别哪些规则适用于特定交易所或品种。通过这种层次化的设计, 系统能够在面对复杂的跨品种、跨规则监管场景时, 优先筛选出相关合规规则, 确保合规审查的精准性和执行的有效性。层级信息的引入不仅提高了法规检索的准确性, 也能帮助系统动态适应监管变化和市场复杂性, 确保合规审查在多变的市场中始终保持高效和准确。

2.1.2 法规适用性信息的增强嵌入

金融法规的适用性是影响法规检索结果准确性的另一个重要因素。在金融领域, 同一条法规可能适用于不同的市场主体, 如个人投资者、机构投资者、期货公司、交易所等。因此, 法规条款的适用性信息需要在向量化过程中加以考虑, 确保系统能够返回最相关的法规条款。

为此, 本文引入了法规适用性标签的增强嵌入机制。在法

规文本中,本文通过数据预处理提取适用性信息,例如“适用于期货公司”、“适用于个人投资者”等。然后,将这些适用性标签转换为向量形式,作为附加信息融入到法规的最终向量中。在检索时,模型能够根据查询的主体信息自动筛选出相关法规,避免返回不符合需求的条款。

最后,将法规文本向量与法规的层级和适用性信息一同拼接至文本向量,形成最终的法规向量。

2.1.3 核心术语加权优化

在金融法规中,关键术语对法规条款的含义和适用性具有决定性作用。术语如“杠杆比例”、“持仓限额”和“风险敞口”通常会出现在关键的风控条款中,因此这些术语的权重在法规向量化过程中应当得到特殊处理。传统的向量化方法可能无法充分捕捉这些核心术语的影响力,从而影响检索结果的精准度。

为此,本文提出了核心术语加权优化方案。首先,通过TF-IDF算法计算每个核心术语在法规中的重要性,并将该权重应用于术语的词向量中。例如,若“杠杆比例”在某条法规中的权重较高,系统就会增加该术语的语义向量在法规整体向量中的比重。最终的文本向量见公式(2):

$$Vec_{FINAL} = Vec_{BASE} + \alpha \sum_{i=1}^m TF \cdot IDF(Vec_{t_i}) + \beta \sum_{i=1}^m \sum_{j=1}^m R_{ij} \cdot Vec_{t_i} \cdot Vec_{t_j} \quad (2)$$

其中, α 是用于增强核心术语权重的参数,它影响的是文本向量中核心术语的重要性提升程度。值越大,核心术语对最终向量的影响越显著,从而提升系统在捕捉关键法律概念时的准确性; t_i 是核心术语, m 是法规文本句子中包含的核心术语数, R_{ij} 是相关性矩阵中术语 t_i 和 t_j 之间的相关性值; $Vec_{t_i} \cdot Vec_{t_j}$ 表示两个术语向量的点积,反映它们的语义交互, β 作为一个超参数,可以通过实验调整其值,以平衡独立术语权重和相关性项的贡献,进一步增加模型的灵活性和复杂性。

通过这种加权策略,核心术语对法规向量的贡献得到了强化,从而确保在检索过程中,涉及这些核心术语的法规能够被优先返回,一定程度的避免返回单一法规内的规则,提高跨法规检索的精准度。

2.2 层次化法规向量检索机制

拼接后的法规向量可以通过余弦相似度进行高效匹配。生成查询向量Query后,系统开始在知识库(知识向量集合 V_{FINAL})中进行相似性检索。采用倒排索引(IVF)和乘积量化(PQ)相结合的算法以提高检索效率。首先,通过使用K-means,IVF将所有向量分配到不同的簇(clusters)中,形成若干个聚类中心。查询过程中,系统首先将查询向量Query与这些聚类中心进行比较,快速定位到几个与之最接近的聚类中心。这样,系统只需在这些特定的聚类中进行向量检索,而不是在整个数据库中进行,从而显著减少了检索所需的计算量。接着,PQ将每个向量分解为多个子向量,并对每个子向量进行独立的量化处理,形成多个子码本。通过这种方法,原始向量被近似表示为一系列的码本索引,大幅减少了存储空间的需求并允许高效的近似距离计算。在检索时,系统通过比较码本之间的距离来估算向量间的相似度,从而快速地锁定与查询向量Query相似度最高的 m 个向量。整个检索流程利用了IVF快速定位聚类中心的优势和PQ在压缩数据和计算距离上的高效性,有效地结合了两者的优点以优化检索性能和数

据存储。

3 合规知识库驱动的双目标优化机制

本文提出了一种面向高合规领域的大语言模型优化新范式,基于知识库反馈的双目标强化微调机制(RLKBF)。如算法2所示,该方法通过合规性反馈与语义偏好双信号的融合优化,突破了现有DPO算法仅基于排序偏好的单目标限制,实现了生成质量与法规约束之间的动态平衡。RLKBF不局限于模块组合,而是对大语言模型优化范式在高合规风险场景中的结构性拓展。相较于现有方法,本文的突破性体现在两个维度:合规约束的数学形式化建模和动态偏离度敏感的权重调节机制。本文所研究的模型输出不仅需要反映知识库反馈的相对偏好,还必须严格遵守知识库的绝对合规性规则,尤其在高合规性要求的领域如期货合规领域中。

算法2. RLKBF

```

1. Procedure RLKBF_TRAINING(sample_set S, knowledge_base K, model_parameters  $\theta$ )
2.   for each training epoch t do
3.     Sample a mini-batch( $x, y_{pos}, y_{neg}$ ) from  $S(x, y_{pos}, y_{neg})$ 
4.      $L_{pref} \leftarrow DPO\_LOSS(x, y_{pos}, y_{neg}; \theta)$  # Step 1: Preference loss calculation
5.      $sim_{y_{pos}}, sim_{y_{neg}} \leftarrow SIM(y_{pos}, K), SIM(y_{neg}, K)$  # Step 2: Compliance distance calculation
6.      $\Delta_y \leftarrow GAP(sim_{y_{pos}}, compliance\_boundary(\theta))$ 
7.      $L_{compliance} \leftarrow \lambda_{penalty} \times EXP\_DECAY(\Delta_y)$ 
8.      $\lambda_{base}, \lambda_{domain} \leftarrow Update\_frequency-sensitive\_weights(), contextual\_risk\_factor(x)$  # Step 3: Dynamic weight tuning
9.      $\lambda_t \leftarrow (1 + \lambda_{base}) \times (1 + \lambda_{domain})$ 
10.     $g_{pref}, g_{compliance} \leftarrow \nabla_{\theta} L_{pref}, \nabla_{\theta} L_{compliance}$  # Step 4: Gradient orthogonal correction
11.     $g_{proj} \leftarrow ORTHO\_PROJ(g_{compliance}, g_{pref})$ 
12.     $g_{total} \leftarrow g_{pref} + \lambda_t \times g_{proj}$ 
13.     $\theta \leftarrow \theta - \eta \times g_{total}$  # Step 5: Parameter update
14.    if violation count exceeds threshold then # Step 6: Real-time compliance feedback adjustment
15.      Increase  $\lambda_{penalty} \leftarrow \lambda_{penalty} \times \alpha$ 
16.    end if
17.  end for
18. return  $\theta$ 
19. end procedure

```

3.1 合规性规则的引入

合规性嵌入的核心在于通过数学方式量化模型输出与预设合规标准之间的偏离程度,目的是防止模型输出“用户偏好但违规”的答案。为此,公式(3)定义了一种合规性差距函数,用于衡量模型输出与知识库匹配标准的距离:

$$c_{compliance}(y, K) = \sum_{j=1}^n \lambda_j \cdot \exp(-\alpha_j \cdot (\gamma_j(t) - sim_j(y, K))) \cdot \Pi(sim_j(y, K) < \gamma_j(t)) \quad (3)$$

其中, y 为回答文本; K 为合规知识库,包含与期货市场相关的法规、规则; j 为合规维度的索引,表示不同的合规性要求,例如交易规则、交割规则或结算规则; $sim_j(y, K)$ 表示在第 j 个维度上,模型输出 y 与知识库 K 的相似度; λ_j 第 j 个维度的惩罚系数,控制违反合规性时的惩罚力度,值越大惩罚越强;

α_j 第 j 个维度的指数衰减因子, 调节惩罚对偏离程度的敏感性. $\Pi(\text{sim}_j(y, K) < \gamma_j(t))$ 为指示函数, 当 $\Pi(\text{sim}_j(y, K) < \gamma_j(t))$ 时值为 1 (表示不合规), 否则为 0; $\gamma_j(t)$ 第 j 个维度的动态合规边界, 随时间 t 变化, 反映法规或市场环境的实时更新, 如公式(4)所示:

$$\gamma(t) = \{y \in \mathcal{Y} | E_{k \in \mathcal{K}(t)} [\phi(y, K(t)) \cdot \omega_k] \geq \tau(t) - \varepsilon(t)\} \quad (4)$$

其中, $E_{k \in \mathcal{K}(t)}$ 表示在法规知识库中所有相关法规条目前计算相似度的期望值, $\phi()$ 为实时更新的语义相似度判别器, $K(t)$ 为时变法规知识库 $Top K$, $\tau(t)$ 为风险敏感的动态阈值, 根据市场风险或者法规更新进行调整, $\varepsilon(t)$ 是容忍度项, 用于允许一定的偏差避免边界过于严格, 增强模型的鲁棒性; ω_k 为法规条目 k 在的权重, 反映其对合规边界的贡献程度, 如式(5)所示:

$$\omega_k = \frac{\exp(-\beta \cdot d(k, y, t))}{\sum_{k' \in \mathcal{K}(t)} \exp(-\beta \cdot d(k', y, t))} \quad (5)$$

其中, $d(k, y, t)$ 为距离函数, 衡量法规条目 k 与输出 y 在时间 t 的语义或时间相关性. β 为调节参数, 控制权重的衰减速度, 值越大权重越集中于距离较小的条目. 根据机构风险偏好可以动态调整.

合规性差距函数的设计突破了传统方法仅依赖语义偏好排序的局限, 通过引入动态合规边界 $\gamma_j(t)$ 和非线性衰减因子 β , 实现了对模型行为的前置约束. 这一创新机制使得模型在优化过程中能够主动规避合规风险区域, 尤其在期货市场等高风险场景中, 显著提升了合规性保障的鲁棒性. 相较于传统的静态规则检查, 该方法通过期望加权 and 模糊容忍度项, 使得合规边界具备动态适应性, 避免了硬性阈值导致的误判问题. 该惩罚机制并非对模型行为结果的后验调整, 而是通过引入动态合规边界与非线性因子, 实现对语言模型优化路径的前置约束. 这种机制打破了以往仅依赖排序偏好或语义评分的弱约束方式, 从而重塑了模型输出在语义空间中的收敛轨迹.

3.2 帕累托优化的双目标代价函数

在 DPO 原有的损失函数 \mathcal{L}_{DPO} 中引入合规性惩罚项. 本文设计的动态双通道调节机制, 从监管强度偏好与语义偏离程度两个维度调节合规惩罚敏感度, 具有良好的领域泛化性. 该机制本质上为偏好对齐问题引入了可控外部结构约束, 为大模型合规生成提供了新的损失建模策略. 在模型的输出与知识库的答案偏差过大的情况下, 损失函数中的惩罚项将会增加, 迫使模型调整输出, 使之更符合合规标准. 故构建包含偏好损失与合规惩罚项的混合代价函数如(6)所示:

$$\mathcal{L}(\theta) = \mathcal{L}_{DPO}(\theta) + \lambda_{\text{compliance}} * c_{\text{compliance}}(y^*) + \xi \|\theta\|_2^2 \quad (6)$$

其中, $\mathcal{L}_{DPO}(\theta)$ 是 DPO 的偏好损失, 用于优化模型输出以符合偏好. $\xi \|\theta\|_2^2$ 为常规 L2 正则, 用极小系数抑制过拟合却几乎不影响主损失标度. $c_{\text{compliance}}(y^*)$ 是模型正输出与合规标准的差距. $\lambda_{\text{compliance}}$ 是惩罚权重, 决定合规性要求对总损失的影响大小, 其采用自适应的双通道更新策略:

1) 宏观层面: 根据监管更新频率调整基线权重 λ_{base} 政策频繁变更场景下提升对合规惩罚项的敏感性.

2) 微观层面: 基于具体对话上下文风控合规要求的领域系数 λ_{domain} .

3) 高合规风险领域会自动开启惩罚项的强化模式:

$$\lambda_t = (1 + \lambda_{\text{base}})(1 + \lambda_{\text{domain}}) \quad (7)$$

双目标代价函数通过动态权重 $\lambda_{\text{compliance}}$ 的设计, 首次实现了合规性惩罚强度的自适应调节. 相较于传统固定权重的多目标优化方法, 本机制能够根据监管政策的变更频率和具体对话的风险等级, 实时调整优化重点, 确保模型在高风险场景下优先保障合规性, 同时在低风险场景下兼顾用户偏好. 这一创新为大模型在动态监管环境中的应用提供了新的解决方案, 尤其在期货市场高频变动场景下, 显著增强了模型的领域泛化能力.

3.3 合规性导向的梯度重定向算法

为解决偏好优化目标与合规性目标在梯度方向上可能存在冲突的问题, 本文引入了一种基于梯度空间投影的合规性导向梯度重定向算法. 在传统多目标加权优化方法中, 常采用偏好损失与合规损失的线性加权组合形式, 但该方式在两个梯度方向高度不一致甚至相互对抗时, 可能导致梯度震荡或优化路径偏移, 影响模型最终性能. 为此, 本文提出在合规损失方向上进行正交投影修正, 即将合规性梯度 $\nabla_{\theta} \mathcal{L}_{\text{compliance}}$ 投影到与偏好梯度 $\nabla_{\theta} \mathcal{L}_{\text{pref}}$ 正交的子空间中, 仅保留不干扰偏好优化的那部分修正方向, 从而获得新的梯度表达式, 如公式(8)所示:

$$\nabla_{\theta} \mathcal{L}_{\text{total}} = \alpha_t \cdot \nabla_{\theta} \mathcal{L}_{\text{pref}} + (1 - \alpha_t) \cdot \mathbb{P}_{\perp}(\nabla_{\theta} \mathcal{L}_{\text{compliance}}) \quad (8)$$

其中, \mathbb{P}_{\perp} 表示正交投影算子, 用于消除与偏好梯度方向重合的分量. 该方法将梯度空间正交投影引入到偏好优化与合规调整之间的融合过程中, 提升了两类目标协同优化的稳定性; 而且通过 α_t 的动态调节机制, 可根据具体任务阶段灵活调整偏好与合规之间的权重平衡, 控制模型在优化过程中对“合规偏好”与“合规约束”两个目标的重视程度, 值越大越侧重偏好, 值越小越强调合规, 实现更精细的策略控制; 同时, 该结构天然兼容现有的策略梯度优化流程, 具有良好的可扩展性和部署友好性.

本文所引入的合规性梯度正交修正策略专门针对合规约束的可解释性冲突, 通过向偏好梯度的正交空间投影, 有效实现语义层级的冲突调和. 通过该方法, 模型在确保合规性的同时, 最大程度保留了原始偏好信息, 从而达到偏好最优且合规安全的优化目标. 偏好-合规双目标优化策略通过权重动态调节与合规性正交投影机制, 首次实现了对偏好优化路径中法规冲突区域的自动规避.

3.4 动态合规性调整与反馈机制

为实现实时监控和快速响应, 本文实施了一个实时反馈机制. 每次模型生成回答后, 系统都会自动进行合规性检测. 如果检测到合规性问题, 系统可以调整模型的相关参数, 确保下一次输出更加符合合规标准. 这种即时的调整机制是保证模型输出持续合规的关键.

为确保模型在整个训练过程中严格遵守合规性要求, 引入了动态调整的合规性惩罚权重 $\lambda_{\text{compliance}}$ 的机制. 由于即使模型一开始能够较好地遵守合规性规则, 它在训练过程中可能因为探索新的答案路径而开始生成偏离合规标准的输出. 因此, 如果系统监测到模型在连续几次训练迭代中违反了合规性要求, 自动增大合规性惩罚权重 $\lambda_{\text{compliance}}$ 变得至关重要. 本文先用离线 Pareto 搜索给 $\lambda_{\text{compliance}}$ 设置初始值; 上线后再以 1% 目标违规率为基准, 借助 PI 闭环控制实时微调

$\lambda_{\text{compliance}}$, 从而在确保合规安全的同时保持生成质量. 这样做的目的是增强模型对合规性的敏感性, 通过调高损失函数中合规性偏离的成本, 迫使模型调整其行为, 优先生成符合规定的答案.

实验中发现, 通过引入两阶段反馈机制, 模型能够自适应调整合规性敏感度, 从而实现在多个风险情境下的泛化一致性. 通过这些措施, 故本文确保了模型在学习和优化过程中不断适应并遵循合规性要求, 同时也使模型能够更灵活地应对可能出现的合规性挑战. 这种动态和实时调整的合规性策略, 不仅提升了模型的合规性保障, 也增强了模型在实际应用中的可靠性和安全性.

综上, 本文提出的双目标合规性优化机制在建模设计上实现了一定程度的优化. 在合规差距函数中引入期望加权与容忍项, 构造动态模糊合规域, 实现更强的表达能力与鲁棒性; 同时, 首次将合规性目标的梯度正交投影机制引入语言生成优化中, 有效解决目标冲突问题.

4 实验

在模型选择方面, Target LLM (Qwen2. 5-14B-Instruct) 作为微调核心, 其输出与更大且精调的参考模型 Ref LLM (Qwen2. 5-32B-Instruct) 进行对比. 参考模型提供高质量基准, 帮助识别目标模型在逻辑一致性与专业性上的不足, 最终生成反馈信号用于优化训练数据.

本文所使用的实验数据集主要由两部分构成: 第 1 部分基于期货合规考试题库, 通过两个大型语言模型 (LLM) 生成回答, 并结合评分体系形成偏好回答对, 共计 1.7K 条; 第 2 部分为人工优化生成的数据, 来源于对已有偏好对的人工编辑与质量提升, 共计 0.6K 条.

由于数据的收集受限于数据标注成本与真实违规场景的结构化难度, 故实验数据集的整体规模仍相对有限. 但是本文在数据设计过程中引入了多层筛选与专家审校机制, 优先选择涵盖高频违规类型, 如杠杆使用、持仓超限、跨所套利等代表性样本, 以保证数据在典型性和问题多样性上的平衡. 此外, 本文所提出的 RLKBF 方法通过引入法规知识约束机制, 显著增强了模型在高风险合规场景下的边界控制能力, 有效避免了语义漂移与违规表达的风险. 在该机制的支持下, 即使在小样本条件, 模型亦表现出良好的鲁棒性与合规性, 验证了 RLKBF 在复杂期货场景中的广泛适应潜力.

在模型评估方面, 依据 3 个典型场景构建测试集, 数据来源于各大交易所官网发布的违规案例. 通过设计提示语引导 ChatGPT 提取事件背景、监管问题及答案, 并对生成结果进行人工审查和筛选, 确保测试集具有针对性和可检索性. 实验对比模型设计方面, 构建了多层次参照体系, 包括基准模型 (如 Qwen2. 5-32B/72B-Instruct) 以分析不同参数规模下知识增强方法的效果; 经典 RAG 模型 (如 RAG-Token/Sequence) 使用 DPR 检索器和 BART 生成器, 并接入相同知识库以保障实验公平性; 此外还引入谷歌优化版 RAG 模型 datagemma-rag-27b-it, 其通过 Data Commons 提供结构化知识支持. 为评估 RLKBF 方法在前沿 RAG 技术下的性能, 本文进一步纳入两种创新模型: Self-RAG 通过插入控制标记实现动态检索抽

发, 并在期货场景中引入 `<compliance_check>` 标记响应高风险关键词; FLARE 采用“生成-检索-验证”迭代机制, 基于困惑度判断是否进行二次检索, 并对不同类型问题设置差异化阈值控制检索敏感性和精度. 上述模型均基于统一的知识库索引与配置参数进行实验对比, 确保评估结果的可比性与可靠性.

4.1 模型准确性评估实验

为突显本文所提出的模型微调方法相较传统方法的优势, 本章分别构建了基于本文微调模型和基准方法的模型, 并对其性能进行了系统对比. 在实验设置上, 微调数据集的 90% 用于训练, 10% 用于测试评估. 实验采用 F1 得分、ROUGE-L 和准确率 (Acc) 来衡量实体集的正确覆盖效果. F1 得分用于评价模型在满足实体集中正确覆盖率方面的整体性能; ROUGE-L 则通过最长公共子序列匹配, 测量生成文本与参考文本的相似度, 以反映实体覆盖的准确度与流畅性; Acc 用于衡量模型在首个实体的完全精确匹配上的准确性. 通过这些指标的结合, 能够更全面地评估系统在识别与覆盖实体集任务中的综合表现.

表 1 的实验结果表明, 本文提出的基于知识库反馈的微调模型在合规导向任务中实现了性能与安全性的良好平衡. 尽管 FLARE 和 Qwen2. 5-72B-Instruct 等大模型在准确率和 F1 得分上表现突出, 但在高风险合规场景中暴露出明显缺陷, 例如 FLARE 在 50 例测试中出现 3 例违规建议, Qwen 因缺乏知识约束导致 9.2% 的法规条款遗漏. Self-RAG 准确率虽达 0.766, 但 ROUGE-L 仅为 0.656, 语义连贯性不足.

表 1 不同模型性能效果对比

Table 1 Performance comparison of different models

模型	F1	Acc	ROUGE-L	Reject Rate (%)	Reject Rate* (%)
本文微调模型	0.662	0.743	0.668	25.00	43.00
Qwen2. 5-72B-Instruct	0.678	0.767	0.681	30.00	35.67
Qwen2. 5-32B-Instruct	0.657	0.719	0.664	32.00	37.00
rag-token-nq	0.658	0.756	0.656	20.33	30.00
rag-sequence-nq	0.619	0.664	0.609	22.00	31.67
datagemma-rag-27b-it	0.652	0.749	0.674	28.00	38.33
Self_RAG	0.677	0.766	0.656	23.00	36.50
FLARE	0.693	0.772	0.683	34.00	37.00

相比之下, 本文微调模型在仅 14B 参数量级下实现 0.743 的准确率和 0.668 的 ROUGE-L, 核心性能接近 32B 大模型, 且在 200 例高风险测试中未出现违规建议. 特别是在应对 2024 年新规方面, 传统 RAG 模型准确率降至 0.541, 而本文模型凭借动态知识更新保持在 0.823, 显示其强大的适应性和更新能力. 此外, 模型生成文本的法律条款覆盖率达 93.4%, 句式合规率达 98.5%, 3 项指标的离散系数仅为 0.067, 优于其他对比模型, 表现出更好的稳定性与多指标平衡性.

另外, 根据表 1 最后两列实验数据, Qwen 系列模型和传统 RAG 方法在仅提供负向文档的拒绝任务中稳定性有限, FLARE 虽具有较高的初始拒绝率, 但在复杂风险识别上表现不足; Self-RAG 虽能在部分场景下提升 Reject Rate*, 但整体波动较大. 而本文模型通过知识约束策略, 有效抑制噪声干扰并精准捕捉负向信号, 在保持适中 Reject Rate 的同时显著提

升 Reject Rate* 至 43.00%, 展现出更强的风险判断深度。

4.2 噪声抵抗评估

如表 2 所示, 虽然在无噪声条件下, Qwen2.5-72B-Instruct 和 FLARE 等大模型的准确率较高。然而, 随着噪声率的增加, 部分模型在 0.6 噪声点附近出现明显波动, 表现出较差的稳定性。在关键的 0.4 噪声率区间, 只有少数模型能够有效抵抗语义偏移并维持较平稳的性能。

表 2 噪声抵抗能力准确率(%)实验结果

Table 2 Accuracy results(%) of noise robustness experiments

噪声率	0	0.2	0.4	0.6	0.8
本文微调模型	96.1	93.9	92.3	89.0	75.5
Qwen2.5-32B-Instruct	96.5	94.8	94.2	90.5	75.8
Qwen2.5-72B-Instruct	97.2	95.4	94.6	91.3	76.4
rag-sequence-nq	92.3	90.4	88.5	84.2	72.1
rag-token-nq	91.8	89.7	87.4	83.6	70.8
datagemma-rag-27b-it	93.7	92.1	90.5	86.7	74.3
Self_RAG	96.8	94.7	93.2	89.4	77.6
FLARE	97.5	95.9	94.1	91.8	78.3

相比之下, 本文微调模型在无噪声条件下准确率为 96.1%, 在高噪声率下仍保持 75.5% 的准确率, 波动仅为 3.3%, 显示出更强的鲁棒性与稳定性, 验证了其在复杂真实场景中的可靠性与适应能力。

4.3 专业域自动化实验结果

如表 3 所示, 本文对多种主流模型在自动化评估中的表现进行了对比, 涵盖 RAG-token-nq、Qwen2.5 系列、ChatGLM3-32B 等模型, 重点评估其在无知识库与知识库参考两种情境下的性能差异。实验问题分为简单、中等和困难 3 类, 分别对应不同层级的逻辑与知识整合需求。

表 3 专业域自动化实验结果

Table 3 Automated experiment results in specialized domains

知识库参考	模型	准确率			
		简单题 100 题	中等题 100 题	难题 50 题	平均
无	Qwen2.5-32B-Instruct	97% (97)	88% (88)	56% (28)	80%
	Qwen2.5-72B-Instruct	96% (96)	85% (17)	76% (38)	86%
	ChatGLM3-32B	95% (95)	90% (90)	70% (35)	85%
有	Qwen2.5-32B-Instruct	98% (98)	92% (92)	86% (43)	91%
	Qwen2.5-72B-Instruct	97% (97)	95% (95)	88% (44)	94%
	ChatGLM3-32B	97% (97)	93% (93)	82% (41)	91%
	rag-token-nq	98% (98)	96% (96)	87% (44)	94%
	Self_RAG	97% (97)	94% (94)	86% (43)	92%
	FLARE	98% (98)	97% (97)	89% (45)	95%
	本文问题微调系统	98% (98)	97% (97)	90% (45)	95%

结果表明, 本文微调模型在各类任务中具备良好适应性, 知识库的引入有效提升了复杂任务的生成质量。与 RAG-token-nq 和 FLARE 等框架相比, 本文模型在准确性和领域泛化能力方面表现更稳健。整体实验为优化检索增强生成提供了实践依据, 揭示了当前大模型在领域应用中的优势与不足。

4.4 专家评估结果

为验证模型有效性, 本文除自动化评估外, 还邀请五位期货合规专家(含风控人员、产品经理及交易所专家)参与人工评估。专家评估任务基于 50 道期货合规问答题, 比较本文微调模型与 Qwen2.5-32B/72B-Instruct 在回答准确性、规范性和稳定性上的差异。

专家评估采用 3 类结果标注: 优胜(Win)、持平(Tie)和劣势(Lose)。结果显示, 本文模型在多维指标上表现更优, 专家评分中有 46% 为优胜、12% 为持平、42% 为劣势, 相较于基础模型具备明显提升。

进一步借助 GPT-4 自动打分进行辅助验证, 结果与专家一致性较高, GPT-4 给出的结果中, 本文模型在比中获得 66% 优胜评分、34% 为持平, 无明显劣势情形。这表明模型在引入领域微调与法规知识后, 对风险合规任务具备更强理解力和生成稳定性。

4.5 消融实验

本文对构建的基于知识库反馈的微调模型进行了消融实验, 旨在评估不同模块和方法对模型性能的影响。表 4 中, 模型-RKB(Remove Knowledge Base)表示本模型去除知识库检索提示后的表现, 即使用传统知识库的表现, 模型-RDO(Remove Data Set Optimization)表示本模型在训练过程中去除对于偏好数据集的优化之后的表现。模型-RCR(Remove Compliance Restraint)表示本模型去除合规性约束机制之后的表现。

表 4 消融实验结果对比

Table 4 Ablation study results comparison

	准确率			
	简单(20 题)	中等(20 题)	难度(10 题)	平均
本模型	100% (20)	85% (17)	50% (5)	86%
模型-RKB	90% (18)	75% (15)	50% (5)	76%
模型-RDO	100% (20)	80% (16)	50% (5)	82%
模型-RCR	100% (20)	75% (15)	50% (5)	80%

在去除知识库检索提示后, 模型在简易题和中等题的准确率分别下降至 90% 和 75%, 显示检索提示在提供专业背景信息、提升回答精准度方面至关重要。移除数据集优化后, 中等题准确率降至 80%, 影响较小, 但仍体现了优化策略在提升理解力上的作用。去除合规性约束后, 虽然简单题保持 100% 准确率, 但中等题下降至 75%, 表明合规性约束对保障模型决策合理性和行业标准适配性至关重要。总体来看, 知识库检索、数据优化与合规性约束共同支撑了模型在高合规任务中的核心竞争力, 其中知识库检索贡献最大, 合规性约束确保了输出的可靠性与可控性。

当前实验验证集中于期货市场的合规判定, 旨在评估 RLKBF 方法对高风险交易行为的约束能力。然而, 由于该方法基于法规知识库的动态构建与反馈优化, 其技术框架具备跨领域扩展的潜力。通过知识库替换与偏好权重调整, RLKBF 能够适配其他金融领域的法规要求。后续工作将尝试引入银行反洗钱案例、保险风险控制政策以及证券市场合规审查数据, 以进一步验证其领域泛化能力与合规性约束的迁移效果。同时, 探索更复杂场景下的动态法规更新对模型稳定性的影响, 以衡量跨市场法规差异带来的挑战。

5 结束语

本文针对期货市场高合规性场景中智能风控模型对法规适应能力不足、知识更新滞后以及微调依赖人工反馈的问题, 提出了基于知识库反馈的强化学习微调方法(RLKBF), 并围绕期货合规知识库的构建与合规性微调算法展开系统研究。

通过构建语义增强的法规知识库,设计层次化检索机制,并引入法规合规性惩罚项的 DPO 优化策略,显著提升了模型在期货市场合规任务中的适用性、专业性和可控性。

与现有偏好强化类方法主要依赖语义偏好排序不同,本文从法规知识出发构建优化函数与梯度方向的刚性约束,解决了偏好-合规冲突下模型震荡与表达扭曲的问题,不仅从理论上探讨了合规知识库对模型语义理解与任务对齐的增强机制,还通过大规模实验验证了所提出方法在准确率、法规适配性、抗噪鲁棒性和专家认可度等维度上的有效性。相比于现有研究更多着眼于检索增强或偏好强化的单一路径,本文提出的 RLKBF 方法在“知识反馈-合规优化”一体化机制上构建了系统性创新。其理论贡献在于首次将法规知识库反馈内嵌于强化学习结构中,并通过动态代价函数与梯度修正策略,拓展了偏好学习在强约束任务下的适应边界,具备推广至医疗、法律等高风险语境的通用潜力。此外,考虑到实际部署环境中对计算资源与响应时延的严格要求,未来还将重点研究合规模型的轻量化优化策略,提升其在边缘场景与高频实时应用中的部署效率与实用性。

References:

- [1] Liu Z, Huang D, Huang K, et al. Finbert: a pre-trained financial language representation model for financial text mining[C]//Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence, 2021 :4513-4519.
- [2] Nguyen D, Cao N, Nguyen S, et al. FinBERT: multilingual pre-trained language model for financial domain[C]//14th International Conference on Knowledge and Systems Engineering, 2024, doi: 10.1109/KSE56063.2022.9953749.
- [3] WChao P, Robey A, Dobriban E, et al. Jailbreaking black box large language models in twenty queries[C]//IEEE Conference on Secure and Trustworthy Machine Learning, 2025 :23-42.
- [4] Zheng J, Qiu S, Shi C, et al. Towards lifelong learning of large language models: a survey[J]. ACM Computing Surveys, 2025, 57(8) :1-35.
- [5] Xie Q, Han W, Chen Z, et al. Finben: a holistic financial benchmark for large language models[J]. Advances in Neural Information Processing Systems, 2024, 37(3) :95716-95743.
- [6] Li Y, Wang S, Ding H, et al. Large language models in finance: a survey[C]//Proceedings of the 4th ACM International Conference on AI in Finance, 2023 :374-382.
- [7] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[J]. Advances in Neural Information Processing Systems, 2020, 33(5) :9459-9474.
- [8] Fan W, Ding Y, Ning L, et al. A survey on rag meeting llms: towards retrieval-augmented large language models[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024 :6491-6501.
- [9] Cm H, Das D, Ranjan R K, et al. LoKI: money laundering report generation via logical table-to-text using meta learning[C]//Proceedings of the 5th Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI for Financial Forecasting, 2023 :104-110.
- [10] Darji A, Kheni F, Chodvadia D, et al. Enhancing financial risk analysis using rag-based large language models[C]//3rd International Conference on Automation, Computing and Renewable Systems, 2024 :754-760.
- [11] Malandri L, Mercurio F, Mezzanzanica M, et al. RE-FIN: retrieval-based enrichment for financial data[C]//Proceedings of the 31st International Conference on Computational Linguistics: Industry Track, 2025 :751-759.
- [12] Chen W, Su Y, Yan X, et al. KGPT: knowledge-grounded pre-training for data-to-text generation[C]//Conference on Empirical Methods in Natural Language Processing, 2020 :8635-8648.
- [13] Han Y, Wang T. Semi-supervised clustering for financial risk analysis[J]. Neural Processing Letters, 2021, 53(5) :3561-3572.
- [14] Hu L, Liu Z, Zhao Z, et al. A survey of knowledge enhanced pre-trained language models[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(4) :1413-1430.
- [15] Loster M. Knowledge base construction with machine learning methods[D]. Potsdam: Universität Potsdam, 2021.
- [16] Iaroshev I, Pillai R, Vaglietti L, et al. Evaluating retrieval-augmented generation models for financial report question and answering[J]. Applied Sciences, 2024, 14(20) :2076-3417.
- [17] Zarka C, Shad R, Chaurasia A, et al. Almanac—retrieval-augmented language models for clinical medicine[J]. New England Journal of Medicine AI, 2024, 1(2) , doi: 10.21203/rs.3.rs-2883198/v1.
- [18] Lee H, Phatale S, Mansoor H, et al. RLAIIF vs. RLHF: scaling reinforcement learning from human feedback with AI feedback[C]//Proceedings of the 41st International Conference on Machine Learning, 2024 :26874-26901.
- [19] Zhang D, Zhoubian S, Hu Z, et al. Rest-mcts*: LLM self-training via process reward guided tree search[J]. Advances in Neural Information Processing Systems, 2024, 37(9) :64735-64772.
- [20] Xu W, Fang M, Yang L, et al. Enabling language representation with knowledge graph and structured semantic information[C]//International Conference on Computer Communication and Artificial Intelligence, 2021 :91-96.
- [21] Chen B, Bertozzi A L. AutoKG: efficient automated knowledge graph generation for language models[C]//IEEE International Conference on Big Data, 2023 :3117-3126.
- [22] Yu Y, Yan Y, Jin Y. Structural knowledge: from brain to artificial intelligence[J]. Artificial Intelligence Review, 2025, 58(9) :1-39.