

结构化非结构化数据：一个用于提取和查询 财务 KPIs 和指导的多代理系统

Chanyeol Choi

LinqAlpha
New York, NY, USA

Alejandro

Lopez-Lira
University of Florida
Gainesville, FL, USA

Yongjae Lee

UNIST
Ulsan, Republic of Korea

Jihoon Kwon

LinqAlpha
New York, NY, USA

Minjae Kim

LinqAlpha
New York, NY, USA

Juneha Hwang

LinqAlpha
New York, NY, USA

Minsoo Ha

LinqAlpha
New York, NY, USA

Chaewoon Kim

LinqAlpha
New York, NY, USA

Jaeseon Ha

LinqAlpha
New York, NY, USA

Suyeol Yun

LinqAlpha
New York, NY, USA

Jin Kim

LinqAlpha
New York, NY, USA

摘要

从非结构化的财务文件中提取和检索有组织且定量的见解在投资研究中至关重要，然而这一过程仍然耗时且耗费资源。实践中常规的方法严重依赖于劳动密集型的手动流程，限制了可扩展性并延迟了研究工作流。在这篇论文中，我们提出了一种高效且可扩展的方法

来准确地获取提取和检索财务见解，利用了一个由大型语言模型组成的多代理系统。我们提出的多代理系统包含两个专门的代理：提取代理和文本到 SQL 代理。提取代理自动识别非结构化财务文本中的关键绩效指标，标准化它们的格式，并验证其准确性。另一方面，文本到 SQL 代理从自然语言查询生成可执行的 SQL 语句，使用户能够准确地访问结构化数据而无需熟悉数据库模式。通过实验，我们证明了我们的系统有效且精准地将非结构化的文本转化为结构化数据，并能精确检索关键信息。首先，我们展示了我们的系统在将近 95% 的准确性下将财务文件转换为结构化数据，达到了通常由人工标注者实现的表现水平。其次，在对检索任务的人工评估中——使用自然语言查询从结构化数据中搜索信息——有 91% 的回答被人类评估者评为正确。在这两项评估中，我们的系统在各种类型的金融文档上表现良好，始终提供可靠的性能。

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR FinIR'25, Padua, Italy

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/25/07

<https://doi.org/XXXXXXX.XXXXXXX>

Keywords

财务信息提取, 多智能体系统, 文本到 SQL, 金融分析

ACM Reference Format:

Chanyeol Choi, Alejandro Lopez-Lira, Yongjae Lee, Jihoon Kwon, Minjae Kim, Juneha Hwang, Minsoo Ha, Chaewoon Kim, Jaeseon Ha, Suyeol Yun, and Jin Kim. 2025. 结构化非结构化数据: 一个用于提取和查询财务 KPIs 和指导的多代理系统. In *Proceedings of The 2nd Workshop on Financial Information Retrieval in the Era of Generative AI (SIGIR FinIR'25)*. ACM, New York, NY, USA, 9 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 介绍

投资研究关键依赖于对非结构化公司披露信息的分析。在实践中, 分析师通常通过手动扫描长篇文档或查找分散在外部分散源和内部摘要中的数据来提取关键绩效指标 (KPIs)。随着企业继续发布越来越多的收益记录、财务报告和监管文件, 这些手动流程变得越来越耗时且不切实际。为了确保及时性和覆盖面广, 从这些文本中提取和检索见解的可扩展方法现在对于做出明智决策至关重要。大型语言模型 (LLMs) 方面的最新进展为大规模处理非结构化文档开辟了新的可能性, 提供了自动化财务信息提取和检索的有希望的基础。这些模型原则上可以将复杂的文本数据转化为结构化的见解, 并支持与文档内容的自然语言交互。

然而, 尽管它们具有潜力, LLMs 在现实世界的金融工作流程中仍然不可靠。单独使用时, 它们经常误解数值指导, 提取语义上不正确的值, 或者完全省略关键指标。这些失败主要是由于布局敏感性限制以及难以区分事实信号与周围叙述的困难。为了解决这些问题, 最近的工作探索了多智能体框架, 在该框架中, 多个 LLMs 协作以验证、交叉检查和改进彼此的输出——提高准确性, 减少幻觉, 并增强金融文档分析的可靠性。

在此背景下, 我们提出了一种高效且可扩展的方法, 能够准确地提供提取和检索的财务洞察, 利用由大语言模型组成的多代理系统。我们的系统包含两个专门的代理: 一个提取代理, 它使用经过领域调整的提示和逻辑解析并验证 KPI; 另一个是文本到 SQL 代

理, 它可以将分析师查询翻译成结构化输出上的可执行 SQL。通过明确地分解工作流程, 该系统允许有针对性的评估、模块化的调优, 并在每个阶段集成反馈循环——这使得迭代改进和稳健的错误处理成为可能。这种模块化设计不仅自动化了手动工作流, 还确保了高保真度的数据结构化以及精准且值得信赖的信息检索。我们在多样化的 SEC 文件和实际分析师任务上对该系统进行了评估, 结果表明它在准确性、可控性和可扩展性方面都显著优于传统的单一的大型语言模型基准。

2 相关工作

2.1 财务信息提取

从财务文件 (如 10-K、10-Q 和收益电话会议记录) 中提取结构化数据一直是一个长期存在的挑战。早期的方法采用基于规则的解析和模板匹配技术来识别财务实体和事件, 但这些方法在格式变化或叙述复杂性下往往不稳定 [2, 10]。监督事件驱动的抽取系统也被探索过, 例如 Malik 等人的基于分类器的框架用于定量金融事件 [8] 和 SENTiVENT, 它为经济新闻事件提取提供了细粒度标注语料库 [4]。虽然这些系统在狭窄领域内表现出色, 但它们需要大量的特征工程, 并且缺乏跨文件类型的通用性。

最近, 基于 LLM 的提取器被引入以零样本或少样本的方式捕获 KPIs [3]。虽然前景看好, 但这些模型通常缺乏领域基础, 并且倾向于虚构数值或忽略诸如财政指引、非 GAAP 指标或期间范围等细微上下文。相比之下, 我们的系统明确将特定领域的逻辑和验证步骤整合到提取管道中, 以确保在各种财务披露中的可靠性和稳健性。

2.2 大语言模型和财务文档理解

大型语言模型 (LLMs) 越来越多地应用于财务文档任务, 如摘要生成、问题回答和数据提取 [6, 9, 13, 16, 19]。虽然它们在处理一般文本方面表现出有希望的

能力，但最近的研究强调了当应用于半结构化或特定领域财务内容时的基本限制。

例如，Srivastava et al. [13] 发现大语言模型在财务问答中的算术推理方面存在问题，尤其是在汇总表格中的数值或解释财政背景时。同样，Ziegler[19] 观察到大语言模型缺乏对金融和环境术语的一致理解，通常需要外部模式知识来提取准确的见解。Shah 等人 [9] 进一步指出，尽管它们具有潜力，当前模型在多文档财务问答中无法保持一致性，特别是在引用微妙的公司披露时。

此外，幻觉、事实偏离以及无法将预测追溯到源区间仍然是持久存在的问题 [6]。当大规模语言模型在单次通过的端到端设置中应用时，这些挑战会变得更加复杂，在这种情况下，中间错误无法被审计或纠正。

我们的系统通过将职责分配给两个专门的代理来解决这些限制。提取代理使用财务逻辑确保基于事实且经过验证的 KPI 输出，而文本到 SQL 代理则能够对这些结果进行结构化查询。这种模块化设计允许迭代改进和更高的可靠性，解决了最近基于 LLM 的财务信息检索研究中发现的关键问题。

2.3 文本到 SQL 和财务查询

Text-to-SQL 系统旨在将自然语言问题翻译成可以在结构化数据上执行的 SQL 语句。虽然通用模型在 Spider 等基准测试中取得了显著进展 [11]，但在金融领域的有效性仍然有限。挑战来自于模糊的查询措辞、隐含的单位或时间范围以及特定于领域的架构结构 [15, 18]。

为了解决这个问题，最近的研究引入了特定于金融的文本到 SQL 数据集，如 FinSQL [18] 和 BookSQL [5]，这些数据集提供了基于会计或财务报告的训练语料库。然而，即使进行了领域适应，模型仍然经常生成语法正确但语义错误的查询，例如混淆同比增长 (YoY) 和季度环比增长 (QoQ) 值、误识别列类型，或者违反财政约束 [12, 15]。

我们的框架通过将模式感知提示、约束验证和语义一致性检查嵌入到专门的文本转 SQL 代理中来缓

解这些问题。这种模块化设计允许对单元解释、时间对齐和执行正确性进行细粒度控制，从而为动态金融数据集提供强大且用户友好的查询界面。

2.4 多代理架构用于模块化推理

近期基于大语言模型的多智能体系统的研究表明，任务分解和智能体协调可以显著提高推理鲁棒性、模块化和可扩展性 [1, 7, 14]。与依赖单一的大语言模型流水线不同，多智能体框架将子任务分配给具有明确职责和交互协议的专业智能体。这种结构允许中间输出被验证、调试，并进行迭代改进。

例如，AgentNet [17] 探索了在复杂规划环境中协调 LLM 代理的去中心化架构，而 LLM-Coordination [1] 分析了诸如心智理论等认知结构如何增强协作任务完成。李等人 [7] 和 Tran 等人 [14] 的综述强调，多代理设计提供了减少幻觉、显式记忆共享和更好的长期任务对齐等好处。

3 方法

本节介绍了我们用于自动财务 KPI 提取和查询的代理系统。我们首先概述了分析师在处理财务文件时面临的手动工作流程挑战。然后，我们展示了一个模块化的双代理架构，包括 (1) 一个提取代理，它将原始文件转换为结构化 KPI 记录，以及 (2) 一个文本到 SQL 代理，它支持对这些记录进行自然语言查询。最后，我们描述了支持可扩展部署和集成的技术架构。

3.1 动机：手动提取 KPI 效率低下

财务分析师经常从 10-K、10-Q、收益发布和投资者记录等文件中提取关键绩效指标 (KPIs)。如图 2 所示，这一过程涉及手动识别相关陈述、解释财务术语、规范化数字（例如，将“22 - 24%”转换为中间值 23%），并将它们转移到下游工具，如电子表格或数据库中。这条流水线速度慢，容易出错，并且经常缺乏标准化——特别是在处理范围、单位和模糊时间段时。

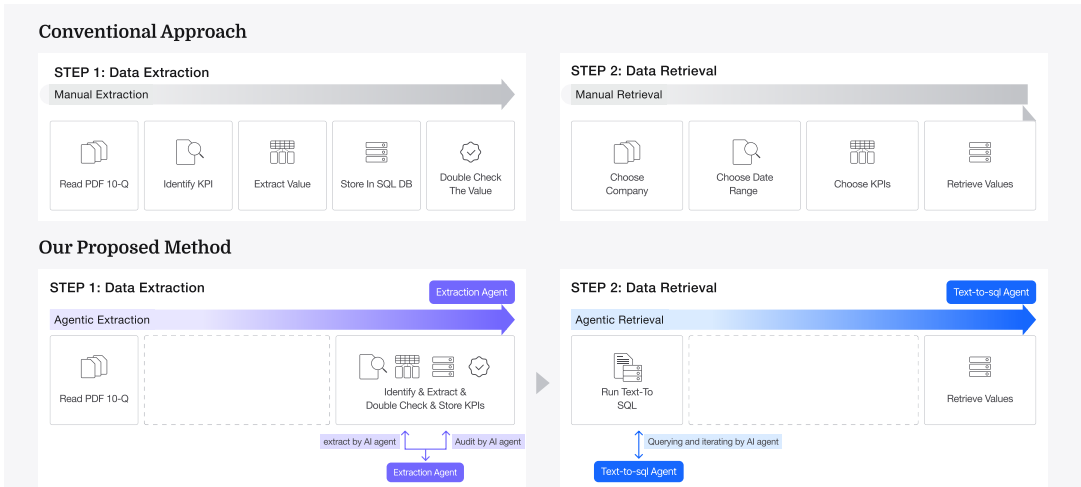


图 1: 传统手动管道与我们提出的代理系统在财务 KPI 提取和检索方面的比较。虽然传统方法（顶部）依赖于在提取和检索步骤中劳动密集型的人工过程，我们的方法（底部）使用两个代理——提取代理和文本到 SQL 代理——自动化工作流程，协作交付可扩展、准确且端到端的财务数据结构化和查询。

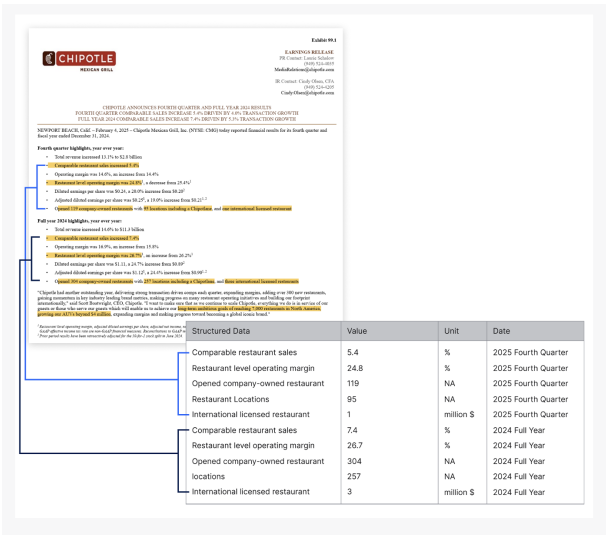


图 2: 从收益公告中手动提取 KPI 的过程 (例如 Chipotle)。分析师必须解析文本、解决数值范围，并手动输入结构化数据。

3.2 系统概述：双代理架构

为了解决这些低效问题，我们提出了一种基于两步多代理架构的模块化系统，如图?? 所示。该系统将提取和查询组件分离成专门的代理，每个代理都旨在优化财务准确性、可控性和可解释性。

管道按如下步骤进行：

- (1) **文档保存**：包括 PDF、HTML 收益页面和记录在内的原始输入被存储。
- (2) **布局与解析**：文档会经过 OCR 处理（如有必要）并按逻辑部分分割。
- (3) **预处理**：文本已归一化；表格和数字跨度已被检测到。
- (4) **嵌入与结构化**：领域特定嵌入模型识别与 KPI 相关的片段。
- (5) **步骤 1 – 提取代理**：关键绩效指标、财政期间和元数据被提取并验证。
- (6) **步骤 2 – 文本到 SQL 代理**：分析师查询通过一个模式感知模型被转换成可执行的 SQL。

3.3 提取代理

提取代理通过结合神经推理和基于规则的逻辑的七阶段过程，将原始财务文件转换为结构化、机器可读的关键绩效指标记录。图 3 的顶部显示了其详细步骤。

- **原始数据解析**：布局感知解析器将文件分割成逻辑部分。必要时应用 OCR。

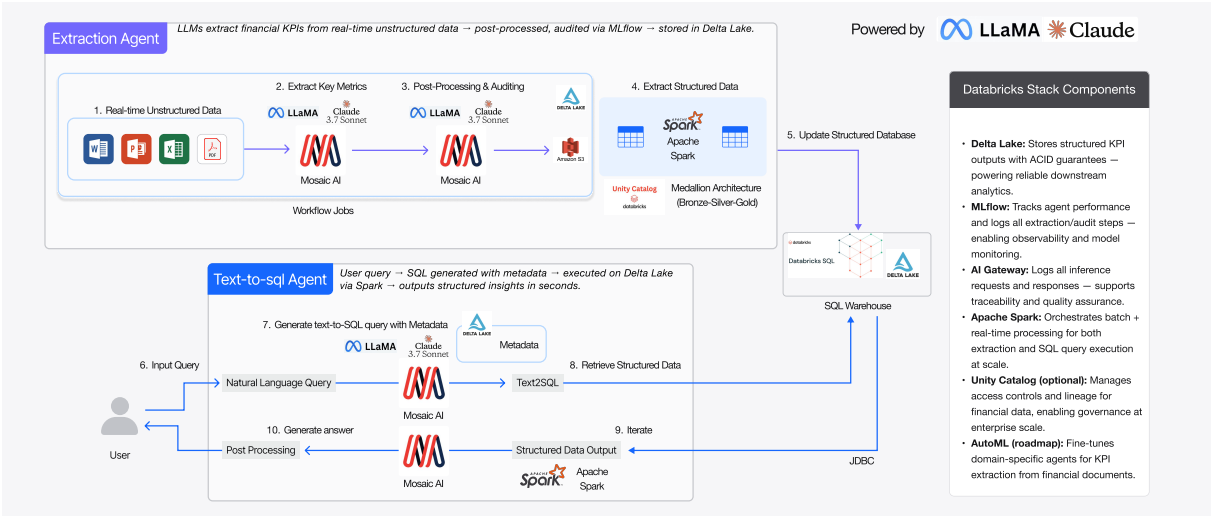


图 3: 我们多智能体系统在财务文档理解方面的概述。提取代理处理非结构化文档，使用大语言模型（LLaMA、Claude）提取结构化的 KPI，并通过 Apache Spark 将其存储在 Delta Lake 中。文本到 SQL 代理通过生成元数据感知的 SQL 查询来实现自然语言查询，并通过 Databricks SQL 检索结果。该系统建立在 Databricks 堆栈上，包括使用 MLflow 进行审计和 AI Gateway 进行可追溯性。

- **实体和指标检测**: 域调优提示使用跨度级别和分类法感知的模板识别候选 KPI 跨度。
- **上下文跨度抽取**: 时间段和限定词（例如，“指导”，“非 GAAP”）等周围上下文与数值一同被提取。
- **初步结构化**: KPI 区间映射到{指标，值，单位，周期，限定词}模式。
- **域规则注入**: 启发式方法标准化范围，解决单位问题，消除财政期间的歧义，并处理调整后的指标。
- **质量管理驱动的验证**: 一致性通过针对性问题进行验证；不一致的输出会被标记或纠正。
- **最终结构化与输出**: 记录以 JSON 格式发出，包含文档元数据和置信度分数。

3.4 文本到 SQL 代理

Text-to-SQL 代理使分析师能够对结构化的 KPI 数据库提出自然语言问题。它将诸如“第四季度的运营利润率是多少？”之类的查询翻译成可执行的 SQL，处理

财务语义和模式变化。图 3 的下半部分显示了其详细步骤。

- **查询理解**: 解析目标指标、时间约束、聚合逻辑和限定词。
- **模式扩充**: 将提示上下文与模式元数据、列类型、别名和示例行注入。
- **SQL 生成**: 使用大型语言模型（例如 Claude 3.7, GPT-4）并结合金融提示和少量示例。
- **约束验证**: 检查语法、单位一致性、时间对齐和限定词准确性。
- **执行反馈**: 测试查询，并在输出不成立或为空时重新排序或生成。
- **解释生成**: 将 SQL 转换为自然语言摘要以支持可解释性。

3.5 整体流程

整个管道在 Databricks 上实现，并支持可扩展、可审计的财务文档处理。

- **文档摄取**: 安全的 API/文件投放摄入到统一的数据湖中。

- **解析与预处理**: 布局感知解析与通过 PDF/HTML/OCR 管道的规范化。
- **大型语言模型编排**: 使用 Claude 3.7 通过基于提示的工作流进行结构化提取。
- **数据存储**: 经验证的 KPI 记录存储在具有模式保证的 Delta Lake 表中。
- **文本到 SQL 接口**: 分析员使用映射到当前模式的自然语言界面查询数据。
- **监测**: 基于规则和 MLFlow 的跟踪确保了可靠性、版本控制和可审计性。

此架构支持跨新文档类型的扩展性、对模式漂移的鲁棒性以及第三方系统或人工介入工作流的集成。

4 实验

4.1 实验设置

为了评估我们多代理框架的有效性,我们模拟了从 SEC 文件 (10-K、8-K、收益发布) 中提取 KPI 并对其进行自然语言查询的真实世界分析师工作流程。我们的评估使用:

- **文档**: 1,200 份来自上市公司的 SEC 文件。
- **查询**: 衍生自对冲基金分析师日志的 312 个自然语言问题。

每份文档均由经过培训的财务分析师手动标注,以识别真实的关键绩效指标 (KPI), 包括收入、营业利润、自由现金流和前瞻性指引。注释进行了交叉验证以确保一致性,并在适用情况下与公认会计原则 (GAAP) 定义保持一致。我们比较了三种配置:

- **基线 A (GPT-4 零样本)**: 直接提示, 不带模式或约束。
- **基线 B (Claude 3.7 + 领域提示)**: 带有模式知识的结构化提示。
- **我们的 (Claude 3.7 + 规则注入)**: 带有基于规则的验证和语义约束的完整管道。

4.2 组件级评估和端到端管道比较

我们首先单独评估每个代理, 以衡量其对整体准确性的贡献, 见表 2 和 3。接下来, 我们在三个配置中评估组合系统, 见表 1。结果显示, 结构化提示和基于规则的验证显著提高了抽取和查询的准确性, 同时降低了错误率并增加了吞吐量。

提示工程影响。· 特定领域提示调优 (Claude 3.7) 将抽取 F1 值提升了 12.3 个百分点, 比 GPT-4 减少了 60% 的单元错误率, 展示了架构意识提示的价值。

规则注入效果。· 增加验证和启发式方法显著减少了语义和数值上的不一致, 特别是在处理时间段不匹配以及 GAAP 与非 GAAP 误解释方面。

查询可靠性。· 约束验证的 SQL 生成相比零样本基线在查询执行准确性上提升了 13.5 个百分点, 反映出用户意图与受模式约束的 SQL 之间有更好的对齐。

系统效率。· 我们的完整框架通过最小化误差传播和减少手动规范化开销, 实现了 2.4 倍的吞吐量提升。

4.3 错误案例示例

- **GPT-4 零样本**: 将“运营利润率指导值为 15-17%”解读为实际第四季度数值 (16.0%)。
- **克劳德 (无规则)**: 正确提取了 16.0%, 但未能将其标记为指导。
- **克劳德 + 规则**: 将识别的范围作为指导并将其从 GAAP 特定查询中排除。

我们的模块化、规则增强型多智能体架构能够实现稳健且可解释的金融问答。通过将管道分解为提取、验证和 SQL 生成, 我们获得了对准确性、可解释性和系统级效率的控制。这些改进对于实际的分析师工作流程特别有价值, 在这种情况下, 数值精度和审计性至关重要。

表 1: 整体管道（提取代理与文本到 SQL 代理）在不同配置下的比较

Config.	Extraction F1	Query Accuracy	Unit Error Rate	Throughput Gain
GPT-4o (naive zero-shot)	72.4%	81.3%	18.5%	1.0×
Claude-3.7-Sonnet (domain prompt)	84.7%	89.6%	7.2%	1.5×
Claude-3.7-Sonnet + Rule Injection	91.2%	94.8%	2.3%	2.4×

表 2: 个体智能体的表现：提取代理

度量	值
Metric Precision	95.3%
Structuring Accuracy	91.6%
Unit Error Rate	2.3%
QA Match Rate	92.3%
Schema Compliance	100.0%

表 3: 个体代理性能：文本到 SQL 代理

度量	值
Intent Classification Accuracy	95.6%
SQL Syntax Validity	98.2%
Semantic Constraint Pass Rate	91.3%
Execution Success Rate	96.4%
Top-1 SQL Accuracy (Human Eval)	91.2%
Explanation Relevance	90.5%

5 结论

我们提出了一种模块化、多代理框架，用于从非结构化的财务文档（如 10-Ks、8-Ks 和收益记录）中提取和

查询结构化信息。通过将流水线分解为专门的提取代理和文本到 SQL 代理，我们的系统解决了财务信息检索中的关键挑战——包括文档布局变化、单位标准化以及自然语言查询中的语义歧义。

通过提示工程和特定领域验证，我们的架构在准确性和鲁棒性方面都优于基线 LLM 设置，使分析师能够以显著减少的手动工作量访问高保真度的关键绩效指标数据。该框架的关注点分离也允许进行有针对性的调整、迭代改进以及更轻松的审计。

未来的工作包括扩展对多语言文件的支持，集成更丰富的时态逻辑，并启用跨文档的长期推理。我们相信这种方法为在金融研究工作流程中部署生成式 AI 提供了一个可扩展且值得信赖的基础。

A 提示工程示例

本节比较了提取代理中使用的两种提示策略：一种是简单的零样本提示，另一种是包含模式、单位归一化和周期锚定的领域感知提示。

B 消融研究：基于规则的验证

如表 5 所示，禁用基于规则的验证会导致单元和周期错误显著增加。

表 4: 提示工程比较：零样本与领域感知

Zero-Shot Prompt	Domain-Aware Prompt
Prompt Extract key financial figures from the text below.	Prompt Extract financial KPIs from the earnings text using the schema below. Output in JSON. Normalize units and link metrics to fiscal periods.
Text In Q1 2024, revenue grew 12% YoY to \$4.3 billion, beating consensus by \$150 million.	Text Same as left. Schema includes Period, Revenue (B), Revenue_YoY_Growth, and Consensus_Delta (M).
Output (Expected) { Revenue: "4.3B", Growth: "12%", Period: "Q1 2024" }	Output (Expected) { "Period": "Q1 2024", "Revenue": {value: 4.3, unit: "B"}, "Revenue_YoY_Growth": "12%", "Consensus_Delta": {value: 150, unit: "M"} }
Notes Lacks structure; prone to inconsistent formatting and unit mismatches.	Notes Improves consistency, reduces hallucinations, enables downstream parsing and SQL conversion.

提示 A：天真零样本提示（无智能体迭代）

提示：

提取以下文本中提到的所有财务指标，并以 JSON 对象的形式返回它们。

示例输入文本：

表 5: 消融实验：没有基于规则的验证

度量	带有规则	无规则	δ 错误
Extraction Precision	91.2%	82.6%	-8.6%
Unit Error Rate	2.3%	12.9%	+10.6%
Period Misalignment Rate	3.1%	15.4%	+12.3%
QA Match Accuracy	90.4%	78.7%	-11.7%
SQL Validity Pass Rate	94.8%	86.0%	-8.8%

2024 年第四季度的营业利润率是 14.6%，比去年的 14.4%有所上升。收入增长了 15.2%，达到 25.2 亿美元。公司预计在 2025 财年的营业利润率将在 15 – 17% 之间。

大语言模型输出：“json ” 营业利润率：“14.6%”，“收入”：“25.2 亿美元”“

References

[1] Shreyas Agashe, Yifan Fan, Aidan Reyna, and Xiaoyuan Edward Wang. 2023. LLM-Coordination: Evaluating and Analyzing Multi-Agent Coordination Abilities in Large Language Models. *arXiv preprint arXiv:2310.03903* (2023). <https://arxiv.org/abs/2310.03903>

[2] Mario Costantino and Paolo Coletti. 2008. *Information Extraction in Finance*. WIT Press.

[3] Mario Costantino and R. G. Morgan. 1997. Natural language processing and information extraction: Qualitative analysis of financial news articles. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 3870–3875.

[4] Gilles Jacobs and Veronique Hoste. 2022. SENTiVENT: enabling supervised information extraction of company-specific events in economic and financial news. *Language Resources and Evaluation* 56, 4 (2022), 1121–1150.

[5] Raghav Kumar, Akshay Dibbu, Shubham Harsola, et al. 2024. BookSQL: A Large Scale Text-to-SQL Dataset for Accounting Domain. *arXiv preprint arXiv:2401.12345* (2024). <https://arxiv.org/abs/2401.12345>

- [6] Hong Li, Hao Gao, and Chao Wu. 2024. Extracting financial data from unstructured sources: Leveraging large language models. *Journal of Information Systems* (2024). <https://publications.aaahq.org/jis/article/38/1/129/12345>
- [7] Xiaoxi Li, Sheng Wang, Shu Zeng, Ying Wu, and Yihan Yang. 2024. A Survey on LLM-Based Multi-Agent Systems: Workflow, Infrastructure, and Challenges. *Springer* (2024). <https://link.springer.com/article/10.1007/s10462-024-10621-2>
- [8] Hammad H Malik, Vishwas S Bhardwaj, and Herb Fiorletta. 2011. Accurate information extraction for quantitative financial events. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1253–1254.
- [9] Sarthak Shah, Sai Ryali, and Raghav Venkatesh. 2024. Multi-Document Financial Question Answering using LLMs. *arXiv preprint arXiv:2411.07264* (2024). <https://arxiv.org/abs/2411.07264>
- [10] Mudassar Sheikh and Sumali Conlon. 2012. A rule-based system to extract financial information. *Journal of Computer Information Systems* 52, 4 (2012), 88–97.
- [11] Akshay Singh, Aishwarya Shetty, Atif Ehtesham, et al. 2025. A Survey of Large Language Model-Based Generative AI for Text-to-SQL: Benchmarks, Applications, Use Cases, and Challenges. In *Proceedings of the 2025 IEEE 15th International Conference on Big Data Science and Engineering (BDSE)*. IEEE.
- [12] Yujie Song, Saif Ezzini, Xia Tang, Carl Lothritz, and Janos Klein. 2024. Enhancing Text-to-SQL Translation for Financial System Design. In *Proceedings of the 46th ACM Symposium on Applied Computing*. ACM.
- [13] Prashant Srivastava, Meghna Malik, Varun Gupta, Tapan Ganu, et al. 2024. Evaluating LLMs' Mathematical Reasoning in Financial Document Question Answering. *arXiv preprint arXiv:2402.13555* (2024). <https://arxiv.org/abs/2402.13555>
- [14] Khanh-Tung Tran, Duy Dao, Minh-Duy Nguyen, Quoc-Viet Pham, et al. 2025. Multi-Agent Collaboration Mechanisms: A Survey of LLMs. *arXiv preprint arXiv:2501.06322* (2025). <https://arxiv.org/abs/2501.06322>
- [15] Niklas Wretblad and Fredrik Gordh Riseby. 2024. Bridging Language and Data: Optimizing Text-to-SQL Generation in Large Language Models. *DIVA Portal* (2024). <https://www.diva-portal.org/smash/get/diva2:1823187/FULLTEXT01.pdf>
- [16] Xinqi Yang, Scott Zang, Yong Ren, Dingjie Peng, and Zheng Wen. 2024. Evaluating Large Language Models on Financial Report Summarization: An Empirical Study. *arXiv preprint arXiv:2411.06852* (2024). <https://arxiv.org/abs/2411.06852>
- [17] Yujing Yang, Haotian Chai, Shu Shao, Yihan Song, Shan Qi, Ruijie Rui, et al. 2025. AgentNet: Decentralized Evolutionary Coordination for LLM-Based Multi-Agent Systems. *arXiv preprint arXiv:2503.10777* (2025). <https://arxiv.org/abs/2503.10777>
- [18] Cheng Zhang, Yifan Mao, Yu Fan, Yutong Mi, Yu Gao, and Ling Chen. 2024. FinSQL: Model-Agnostic LLMs-Based Text-to-SQL Framework for Financial Analysis. *Companion of the 2024 ACM International Conference on AI in Finance* (2024).
- [19] Guilherme Gomes Ziegler. 2024. Automating Information Extraction from Financial Reports Using LLMs. *Aalto University Publications* (2024). <https://aaltodoc.aalto.fi/handle/123456789/123456>