

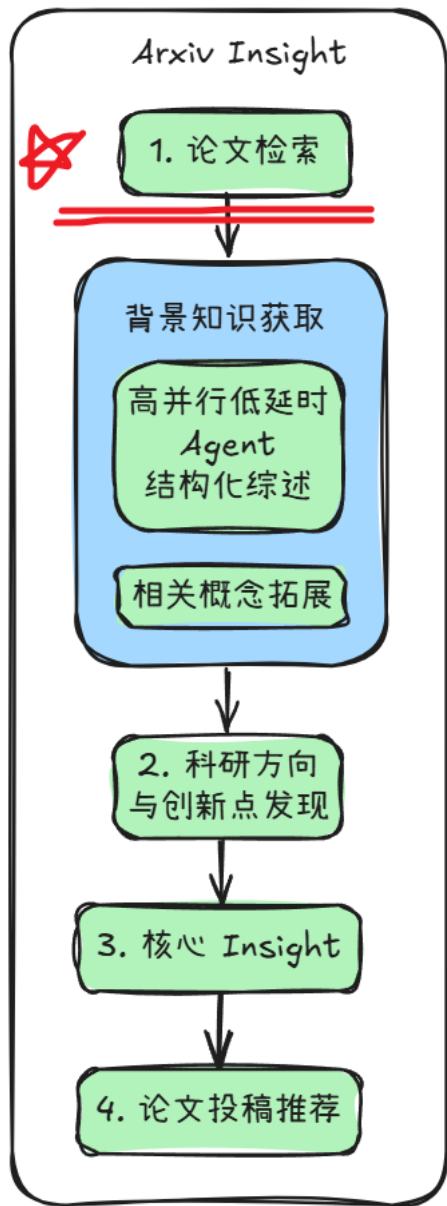


ArXiv Insight: 多 LLM 驱动的论文语义检索

网址: axsight.top



一、Overview



arXiv Insight

多 LLM 驱动的论文语义检索

Q 随意描述主题，尽可能详细（支持中文）

已发表会议

标题翻译

搜索

更多选项 ^

年份（默认 10-25）

选择会议（默认全选）

- ☐ AAAI
- ☐ CVPR
- ☐ ECCV
- ☐ ICCV
- ☐ ICDM
- ☐ ICIP
- ☐ ICLR

豫ICP备2025106078号

arXiv Insight 专注检索，并提供周边服务

当前功能：

1. 强大的语义检索能力（Small LM）
2. 收藏夹：分类、拖拽及导出
3. AI Chat：包括@功能，#选择Prompt

即将上线（Agent模式）：

1. 给定主题 -> 结构化综述
2. 旧概念 + 新热点 -> Insight



一、语义检索及重排



arXiv Insight

多 LLM 驱动的论文语义检索

基于大语言模型（LLM）进行学术文献综述的合成

arXiv预印本

标题翻译

搜索

更多选项

2. arxiv、会议、期刊可选

按相关度排序

1. LLMs4Synthesis: 利用大规模语言模型进行科学综合

2024-09 - arxiv

提出LLMs4Synthesis框架，利用大语言模型生成高质量科学综述，定义九项质量标准，优化综述生成和评价过程

4. LLMs是否擅长撰写文献综述？评估大型语言模型的文献综述撰写能力

2024-12 - arxiv

提出评估大语言模型撰写文献综述能力的框架，发现模型仍存在生成虚假引用的问题。

2. LLM4SR: 大型语言模型在科学研究中的应用综述

2025-01 - arxiv

本论文首次系统综述了大语言模型在科研过程中的应用，涵盖四个关键阶段，提出挑战及未来方向。

5. 评估LLMs和预训练模型在多样化数据集上的文本摘要性能

2025-02 - arxiv

评估了BART、FLAN-T5、LLaMA-3-8B和Gemma-7B在多个数据集上的表现，使用ROUGE等指标衡量生成摘要能力。

4. 标题自动翻译

3. LLM4SR: 使用大型语言模型自动执行文献回顾的简单工具

2024-07 - arxiv

提出LLAssist工具，利用大语言模型和自然语言处理技术自动化文献综述过程，减少时间和精力消耗。

6. 利用大型语言模型增强系统综述：使用GPT-4和Kimi

2025-04 - arxiv

研究评估了GPT-4和Kimi在学术文献综述中的表现，发现其性能受数据量和问题复杂度影响。

功能特性包括但不限于：

1. 更快的检索速度（相比同类平台）
2. 更优的语义匹配
3. 更符合国人习惯的翻译及总结
4. 更多的论文来源（arxiv + 多数顶会顶刊）

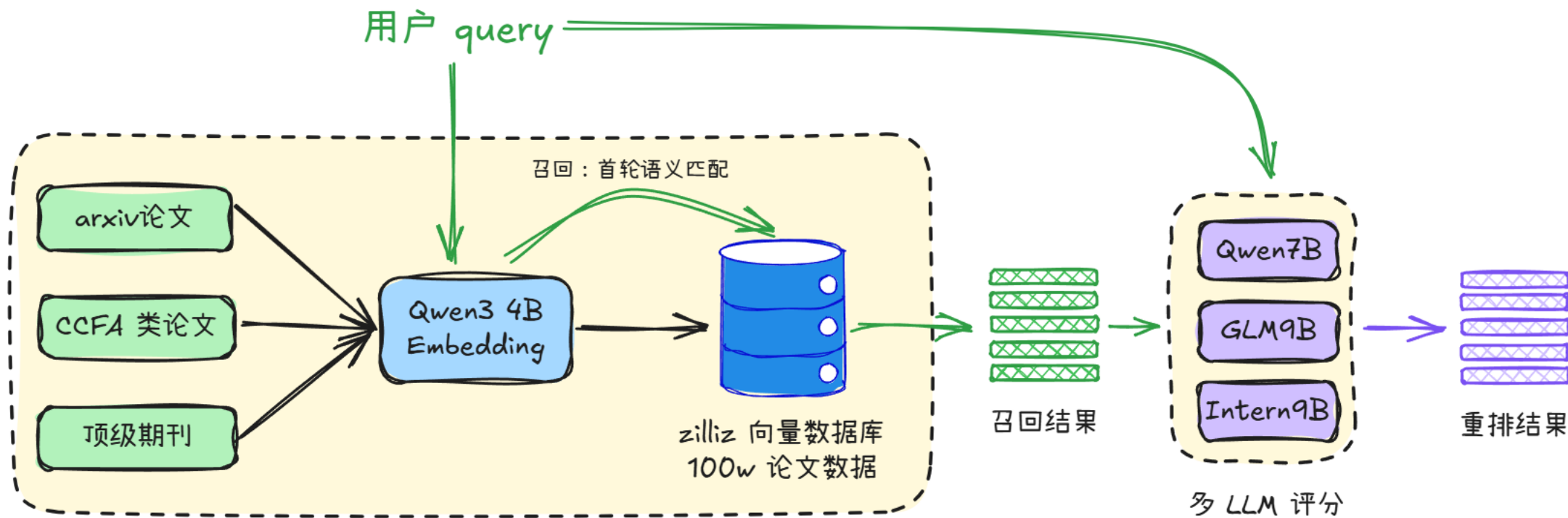
5. 点击跳转谷歌学术

6. 提供摘要自动总结（效果极佳），方便速览

7. 点击卡片提供原始摘要翻译及作者等详细信息



一、语义检索及重排 - 原理



当前常见的 ReRanker 模型在论文检索场景下效果极差，而使用多 LLM 进行评分后重排序得到的结果极佳，但较为依赖 Prompt



二、收藏夹管理

点击网站左上角按钮即可展开收藏夹，并进行收藏夹管理



文件夹

未分类 删除文件夹内容临时存储 8

LLM 合成综述 8

Agent 4 Retrieval 7

Agent for Survey 1

Agent 4 Rerank 8

LLM 4 ReRank (Zero) 10

写作质量评估 2

1. LLM4SR: 关于科学研究中大型语言模型的...
arxiv | 发表: 2025-01 | 收藏: 2025-5-1

2. LitLLM: 科学文献综述工具包
arxiv | 发表: 2024-02 | 收藏: 2025-5-1

3. 让我们直奔主题: 基于大型语言模型的学术论...
arxiv | 发表: 2024-06 | 收藏: 2025-5-1

4. 拉特瑞夫: 一种用于大型语言模型系统化审查...
arxiv | 发表: 2025-01 | 收藏: 2025-5-1

8

重命名文件夹

设为默认文件夹

删除文件夹

下载本文件夹论文

研究中大型语言模型的...

重命名

删除

移动到文件夹 (可拖拽)

未分类

LLM 合成综述

Agent 4 Retrieval

Agent for Survey

下载论文 (8 篇)

将自动下载所有 PDF 论文, 并将非 PDF 链接写入 txt 文件, 全部打包为 zip。

0%

取消

开始下载

耐心就是你所需要的! 一种用于执行科学文献综述的代理系统

Patience is all you need! An agentic system for performing scientific literature review

2025-03 arxiv

作者
David Brett, Anniek Myatt

中文摘要
大型语言模型 (LLMs) 在提供跨学科问答支持方面的应用日益广泛。这些模型本身在回答基本问题方面已显示出潜力, 但一旦需要专业知识或问题较为复杂, 它们就会迅速失效。科学研究通常涉及查找相关文献、从这些文献中提炼相关信息以及分析这些发现如何相互支持或反驳。这些信息通常包含在研究文章的全文中, 而不仅仅是摘要中。这些文章中的陈述往往需要更广泛的上下文才能完全理解。我们构建了一个基于 LLM 的系统, 用于执行科学文献中封装信息的搜索和提炼, 并使用先前发布的文献基准中的生物学相关问题来评估我们的基于关键词的搜索和信息提炼系统。我们证明了稀疏检索方法在不使用密集检索的情况下, 其结果接近最先进水平, 无需承担其相关基础设施和复杂性开销。我们还展示了如何增加文献综述生成中相关文档的覆盖率。

原文摘要
Large language models (LLMs) have grown in their usage to provide support for question

三、AI 对话



1. 默认以当前检索结果的所有摘要为上下文对话

arXiv Insight

多 LLM 驱动的论文语义检索

基于大语言模型（LLM）进行学术文献综述的合成

arXiv预印本

标题翻译

搜索

更多选项

按相关度排序

1. LLMs4Synthesis:
利用大规模语言模型进行科学综合

2024-09 - arxiv

提出LLMs4Synthesis框架，利用大语言模型生成高质量科学综述，定义了九项评估标准，提升了综述质量和可靠性。

4. LLMs是否擅长撰写文献综述？评估大型语言模型撰写文献综述的能力

2. LLM4SR: 大型语言模型在科学研究中的应用综述

2025-01 - arxiv

基于大语言模型的科学过程被系统综述，涵盖四个关键阶段，提出未来研究方向。

5. 穿透杂音：LLMs在系统文献综述中高效筛选信息的能力

3. LLAssist: 使用大语言模型自动执行文献回顾的简单工具

2024-07 - arxiv

提出LLAssist工具，利用大语言模型和自然语言处理技术自动化文献综述，提高效率

6. 科学大型语言模型及其在科学发现中的应用

Claude4-Sonnet

有哪些文献是针对使用 LLM 进行小型综述合成以帮助用户快速了解一个领域的?

删除 复制

基于检索到的文献，我为您整理了专门针对使用LLM进行小型综述合成以帮助用户快速了解领域的相关研究

核心框架与工具

综合性框架

[1] LLMs4Synthesis 提出了专门的LLMs4Synthesis框架，这是最直接针对您需求的研究！该框架利用大语言模型生成高质量科学综述，并定义了九项评估标准来提升综述质量和可靠性。

[11] LatteReview 提出了多代理系统框架，利用大语言模型自动化系统综述过程，特别适合提高效率和准确性。

实用工具包

[3] LLAssist 开发了简单易用的工具，专门利用大语言模型和自然语言处理技术自动化文献综述，显著提高效率。

[17] LitLLM 提出基于检索增强生成(RAG)的工具包，使用大语言模型进行文献综述，能够减少时间和人工成本。

Shift + Enter 换行, @选择论文, #选择Prompt

Claude4-Sonnet

GLM-4.5

GLM-4.5 355B 参数, 激活 32B

Kimi-K2

KIMI - 1000B 参数, 激活 32B

DeepSeek-V3

最新 DeepSeek-V3-0324

DeepSeek-R1

最新 DeepSeek-R1-0528

GPT-4.1

速度 ↑ 通用能力 ↑

gemini-2.5-flash

速度 ↑↑↑ 通用能力 ↑

✓ Claude4-Sonnet

速度 ↓ 编程能力 ↑↑↑

主流模型均免费使用



三、AI 对话 - @ 功能

2. 使用 @ 选中单个检索结果进行单篇对话

主页面全部论文

- [1] LLMs4Synthesis: 利用大规模语言模型进行科学综合 / LLMs4Synthesis: Leveraging ...
- [2] LLM4SR: 大型语言模型在科学研究中的应用综述 / LLM4SR: A Survey on Large Lang...
- [3] LLAssist: 使用大语言模型自动执行文献回顾的简单工具 / LLAssist: Simple Tools for ...
- [4] LLMs是否擅长撰写文献综述? 评估大型语言模型撰写文献综述的能力 / Are LLMs Goo...
- [5] 穿透杂音: LLMs在系统文献综述中高效筛选的潜力 / Cutting Through the Clutter: The...
- [6] 科学大型语言模型及其在科学发现中的应用综述 / A Comprehensive Survey of Scienti...
- [7] SciLitLLM: 如何适应LLM以理解科学文献 / SciLitLLM: How to Adapt LLMs for Scien...
- [8] SciAssess: 评估LLM在科学文献分析中的能力基准 / SciAssess: Benchmarking LLM ...
- [9] 大型语言模型在自动化学术论文评审中的应用: 一项调查 / Large language models for...

@



注意:

单篇 @ 对话时自动提取全文内容作为上下文, 但若选中多篇仍使用摘要 (否则效果较差且成本过高)

@ 单篇文章并进行详细问答

Qwen3-235B-A22B v

这篇文章使用的大模型是什么

删除 复制



根据您的论文内容, 该研究使用了以下两种大语言模型 (LLMs):

1. **GPT-4**: 由OpenAI开发的**专有 (proprietary) 模型**, 在研究中作为高性能基准和"黄金标准"使用。
 - 用于生成高质量的科学综述作为参考 (gold standard)。
 - 作为自动化评估器 (LLM evaluator) 来评估其他模型生成的综述质量。
2. **Mistral-7B**: 一个**开源 (open-source) 模型**, 是本研究重点优化和提升的对象。
 - 论文提出的 **LLMs4Synthesis 框架** 主要目标就是通过监督微调 (SFT) 和基于AI反馈的强化学习 (RLAIF) 等技术, 来提升Mistral-7B这类开源模型在科学综述生成任务上的表现, 使其接近GPT-4的水平。

简单来说, **GPT-4** 是用来设定标准和进行评估的“老师”, 而 **Mistral-7B** 是被优化和训练的“学生”。

删除 复制

[1] LLMs4Synthesis: 利用... x

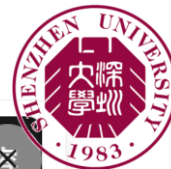
Shift + Enter 换行, @选择论文, #选择Prompt





三、AI 对话 - @ 功能

@ 收藏夹并总结收藏夹主题



3. 使用 @ 选中收藏夹 or 收藏夹单篇文献进行对话

- 收藏夹：LLM 合成综述
- 收藏夹：Agent 4 Retrieval
- 收藏夹：Agent for Survey
- 收藏夹：Agent 4 Rerank
- 收藏夹：LLM 4 ReRank (Zero)
- 收藏夹：写作质量评估
- 【未分类】[1] 面向任务的查询重写与强化学习
- 【未分类】[2] 通过引导搜索增强强化学习
- 【未分类】[3] 探索-Go：利用探索实现深度强化学习中的泛化
- 【未分类】[4] DR-RAG：将动态文档相关性应用于问答的检索增强生成

@



Claude4-Sonnet v

总结一下我这个收藏夹内的主题内容

删除 复制

收藏夹主题总结：LLM在科学文献综述中的应用

您的收藏夹聚焦于一个非常前沿且实用的研究领域：**大型语言模型 (LLM) 在科学研究和文献综述中的应用**

核心主题分析

这个收藏夹围绕着以下几个关键维度展开：

1 理论综述与框架

- [1] LLM4SR 和 [5] 跨学科视角 提供了宏观的理论视角，系统性地分析了LLM在科学研究四个关键阶段的作用
- 这些论文为整个领域提供了理论基础和发展方向

2 实用工具开发

- [2] LitLLM工具包 和 [7] LitLLM代理系统 都基于RAG（检索增强生成）原理
- [3] Papers-to-Posts 采用Plan-Draft-Revise流程

收藏夹：LLM 合成综述 x

[4] LitLLM-Review 使用多智能体框架

Shift + Enter 换行，@选择论文，#选择Prompt





三、AI 对话 - # 功能

4. 使用 # 获取预设 Prompt （配合 @ 使用）

单篇文章详解

详细的结构化解读，仅支持单篇来源是 arxiv 的文章（或 ACL 系列文章），gemini 效果更佳。

【Agent 4 Retrieval】[1] ... x

#



调出外部工具和执行任务

Claude4-Sonnet v

3. 论文方法

使用预设 Prompt 配合 @ 详解单篇检索结果

3.1 过去方法的问题

传统IR系统存在以下根本性限制：

1. **静态语料库依赖**：只能从预定义语料库中过滤和排序现有信息项，无法操作、合成或生成新内容
2. **功能局限性**：
 - 仅能进行信息项过滤，缺乏内容操作能力
 - 无法执行任务（如预订、购买）
 - 缺乏个性化和上下文适应能力
3. **交互模式受限**：主要是反应式的，缺乏主动的意图推理和任务解决能力

Motivation：LLM驱动的AI代理的出现为IR带来了革命性机会，使系统不仅能够反应式地过滤信息，还能主动进行用户意图推理、外部工具使用和最终任务解决。

3.2 整体框架

【Agent 4 Retrieval】[1] ... x

Shift + Enter 换行，@选择论文，#选择Prompt





三、AI 对话 - # 功能



5. 预设 Prompt 详情

帮我详细解释一下这篇文章，包括以下部分

1. 论文核心概念 🗨️ (对论文核心 insight 的简要总结)
2. 论文内名词解释 🤖 (对论文中出现多次，或者比较重要的名词的详细解释)
3. 论文方法 📄
 - a. 过去方法的问题 (顺便引出方法的 motivation)
 - b. 整体框架 (整个论文的方法部分的核心流程的超详细说明，需要保证通过说明可以完整复现出整个方法，包括细节，公式流程，变量说明)
 - c. 核心难点解析 (将方法中比较复杂的部分或者比较关键的部分在这里进行更加直白易懂的解释)
4. 实验结果与分析 📊
 - a. 实验设置 (数据集，模型，指标，超参数设置，对比方法等的内容)
 - b. 实验结果 (该方法指标提升了多少，以及其他相关效果或正面的评价)
5. 结论 💡
 - a. 论文的贡献
 - b. 论文的限制 (论文在哪些方面有问题)
 - c. 未来的方向 (未来可能的发展方向)

注意：

1. 在最开头加上一个一级标题作为文章的简单标记，让我在后续回顾的时候能根据标记快速回想起这篇文章的特点，例如 GQE-PRF：基于伪相关反馈的生成式查询扩展
2. 请你对 x. 这类使用二级标题，对 x.x 使用三级标题，除了上述说明外其他所有内容都不要使用标题加粗，但可以使用序号进行罗列
3. 注意对于较长的公式你需要将其分为多行公式，这样更清晰，方便我理解 (即公式内容不变，但从某个运算符进行切分并展示为多行)
4. 我是一个刚入门的小白，所以你需要尽可能详细的解释关键部分