

数据论文

大语言模型在电池科研全流程应用的测评与
无机固态电解质综合数据库构建*

吴思远 李泓†

(中国科学院物理研究所, 北京 100190)

(2025 年 4 月 29 日收到; 2025 年 6 月 20 日收到修改稿)

大语言模型的出现极大地推动了科学研究的进步. 以 ChatGPT 为代表的语言模型和 DeepSeek R1 为代表的推理模型, 为科研范式带来了显著变革. 尽管这些模型均为通用型, 但它们在电池领域, 尤其是固态电池的研究中, 展现出强大的泛化能力. 本研究系统地筛选了 2024 年及之前重点期刊中的 5309268 篇文章, 精准提取了 124021 篇电池相关文献. 同时, 我们全面检索了欧洲专利局与美国专利局 2024 年及以前的申请与授权专利, 共计 17559750 篇, 从中筛选出 125716 篇电池相关专利. 利用这些文献与专利, 对语言模型的知识储备、实时学习、指令遵从和结构化输出能力进行了大量实验. 通过多维度的模型评估与分析发现: 当前的大语言模型在信息分类和数据提取等的精度基本达到了研究生水平, 语言模型在内容总结和趋势分析方面也展现出强大的能力. 同时, 我们也发现模型在极少数情况下可能出现数值幻觉问题. 而在处理电池领域海量数据时, 模型在工程应用方面仍存在优化空间. 我们根据模型的特点和以上测试结果, 利用模型提取了无机固态电解质材料数据, 包括离子电导率数据 5970 条、扩散系数数据 387 条、迁移势垒数据 3094 条, 此外还包括 1000 多条化学、电化学、力学等数据, 涵盖了无机固态电解质所涉及的几乎所有物理、化学、电学性质, 这也意味着大语言模型对科研的应用已经从辅助科研转向主动促进科研发展阶段. 本文数据集可在中国科学院凝聚态物质科学数据中心查看, 网址 <https://cmpdc.iphy.ac.cn/literature/SSE.html> (DOI: <https://doi.org/10.57760/sciencedb.j00213.00172>).

关键词: 固态电池, 人工智能, 自然语言处理, 材料数据库

PACS: 07.05.Kf, 82.47.Aa

DOI: 10.7498/aps.74.20250572

CSTR: 32037.14.aps.74.20250572

1 引言

语言模型的出现极大地改变了当前的科研范式. 以 ChatGPT^[1] 为代表的语言模型和以 DeepSeek R1^[2] 为代表的推理模型, 将自然语言的应用从文学创作领域扩展到了科学研究领域. 尽管这些模型是通用型的, 但它们在科学研究领域却展现出了强大的泛化能力. 语言模型的发展历程也充满了重要的里程碑. 2017 年, 谷歌推出了基于注意力机

制的 Transformer 架构^[3], 这种架构通过高度并行化和捕获长距离信息的特点, 取代了传统的循环神经网络 (RNN) 和长短记忆网络 (LSTM), 使用注意力值来实现不同信息的捕获. 2018 年, OpenAI 与谷歌公司相继提出了两种新型模型: 利用自回归训练的生成式预训练 Transformer 架构模型 (GPT)^[4] 和利用掩码训练的双向注意力编码表示模型 (BERT)^[5], 从而开启了 Transformer 架构模型无监督学习预训练的新时代. 2019 年, OpenAI 推出了 GPT-2^[6], 通过将有监督学习转化为无监督学习

* 国家重点研发计划 (批准号: 2022YFB2502200) 和国家自然科学基金 (批准号: 2239304) 资助的课题.

† 通信作者. E-mail: hli@iphy.ac.cn

预训练, 并采用多任务的有监督微调来提升模型性能, 进一步增强了模型的能力. 2020年, OpenAI 又推出了 GPT-3^[7], 这次引入了 prompt 策略和对比学习, 通过提示词或例子利用正向传播调整注意力值, 取代了传统的反向传播更新模型参数的方法, 从而显著提升了模型的泛化能力. 随着思维链方法^[8]的引入, 研究者能够利用模型进行更长时间的推理, 这为科学研究提供了新的解决方案, 推动了推理模型如 GPT-o1^[9] 和 DeepSeek R1^[2] 的发展. 继 2022 年 ChatGPT^[1] 出现后, 科研人员更加关注到语言模型带来的强大功能. 这些发展不仅改变了科研范式, 也为科学研究提供了更加强大的工具和方法.

语言模型和推理模型在电池领域的应用潜力巨大, 但其具体能力和应用方式尚未被充分探索. 如何利用语言模型加速电池科研进程, 以及电池领域中哪些基础性科研问题尚未得到深入研究, 导致先进模型无法有效支持电池发展, 仍是亟待解决的关键问题.

大语言模型对电池科研全流程的影响既包括在电池科研过程中原本由人工对文本的分析、处理等与语言直接交互的过程, 也包括一些涉及到的实验操作、计算模拟的输入和命令行等间接接受到结构化输出或者函数调用所影响的科研过程. 针对这些挑战和问题, 本文从文本分类、信息抽取(结构化输出)、文章总结、文章分析与优化、知识库问答、个性化推荐等多个维度, 对几种主流模型(包括 BERT^[5], GPT-2^[6], Qwen2^[10], GLM^[11], Gemini^[12], ChatGPT o3 mini^[9] 和 DeepSeek R1^[2]. 注: 这些为 2025 年 3 月最新最强模型) 进行系统性研究. 通过二次预训练、指令微调、prompt 优化和挂载知识库等多种优化方法, 我们旨在提升模型在上述能力方面的性能, 特别是 DeepSeek R1 671B (简称 DeepSeek R1) 和使用 DeepSeek R1 蒸馏的 70B 模型 (简称 DeepSeek R1 70B), 从而开发出更适合电池领域应用的方法, 为语言模型在电池研究中的有效应用铺平道路.

最后, 基于以上测试结果, 设计了固态电池核心材料——无机固态电解质综合数据库的构建流程, 通过优化语言模型的工作流, 实现了语言模型构建物质科学数据库的新范式. 本工作主要考核目标与设计的实验示意如图 1 所示.

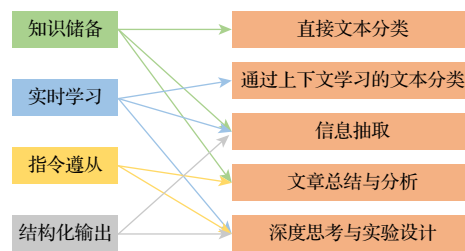


图 1 本工作主要考核目标(左侧)与设计的实验(右侧)
Fig. 1. The main assessment objectives (left) and the designed experiment (right) of this work.

2 评测方法

2.1 文本分类

文本分类是自然语言处理中的基础问题, 旨在快速筛选与特定主题相符或相近的内容. 常见的文本分类方法是将文本向量化后映射到 $[0, 1]$ 的空间范围内, 投影值越接近 1, 表示与主题的相关性越高, 越接近 0 则表示与主题无关. 随着 GPT-3^[7] 尤其是 DeepSeek R1^[2] 的出现, 使用提示 (prompt) 方式进行文本分类已成为主流方法. 本研究将比较 BERT 和 GPT-2、微调的 Qwen2-7B-Instruct DeepSeek R1 直接指令输出的效果, 其实验内容具体如下.

1) BERT 和 GPT-2: 使用 arXiv 的摘要等信息作为预训练数据集 (2023 年 10 月版本), 分别采用掩码预训练和自回归方式对模型进行二次预训练, 并利用文献自带的 arXiv 分类标签进行分类任务训练. 实验结果显示, BERT 和 GPT-2 在准确性上表现相当, 最终选择 BERT 并在人工标记的 10^4 条电池数据上进行微调, 微调后作为模型进行比较.

2) 微调的 Qwen2-7B-Instruct: 使用与 BERT/GPT-2 相同的数据集进行二次预训练, 并进行零样本的指令微调 (剔除本次测试的原位固态化), 以优化模型性能.

3) DeepSeek R1: 直接利用 DeepSeek R1 判断文本是否属于无机固态电解质或原位固态化技术, 并将两者结果分别与上述 BERT 和 Qwen2-7B-Instruct 进行比较.

2.2 信息抽取

信息抽取是自然语言处理中的另一项核心任务. 在 GPT-3^[7] 出现之前, 基于 BERT 的模型是

主要的信息抽取工具. GPT-3 引入了上下文对比学习和提示词的方法, 将传统的信息提取转化为对话形式, 提升了抽取效率和准确性. 本研究将展示使用小模型指令微调和 DeepSeek R1 的 prompt 方式进行信息抽取.

微调的 Qwen2-7B-Instruct: 在二次预训练 Qwen2-7B-Instruct 模型的基础上, 使用人工标记的 40452 条化学式数据进行指令微调, 以优化模型在特定任务上的表现.

DeepSeek R1: 直接使用 DeepSeek R1 模型并使用精心设计的提示词, 提取文本中的化学式、对应的离子电导率及其测量条件, 展示其在复杂信息抽取任务中的能力.

2.3 文章总结与分析

本部分对比分析了 DeepSeek R1, OpenAI o3 mini 和 Gemini 最新推理模型在文献总结与概括方面的性能. 研究方法是通过选取特定时间段内新增的固态锂离子电池领域的文献进行总结, 这些文献均通过 2.1 节所述的电池 BERT 模型产生. 研究重点在于考察各模型在以下方面的表现: 1) 对文献内容的概括能力; 2) 对最新研究进展的分析; 3) 对当前热点文章的识别; 4) 新材料研究的总结; 5) 对上述内容的综合思考和展望能力.

2.4 时间尺度研究动态分析

基于 2.3 节的研究, 选择 DeepSeek R1 进行跨时间尺度内容的趋势分析, 使用日韩四家企业 (LG、三星、松下和丰田) 近 20 年的固态锂电池专利申请和授权数据 (数据同样由 2.1 节的 BERT 产生), 开展跨时间线分析, 着重研究 DeepSeek 是否能够根据时间线揭示该企业研究方向的变化及其未来可能的研究内容.

2.5 个性化推荐

基于 2.3 节和 2.4 节的实验, 尝试利用 DeepSeek R1 70B 针对每位科研人员的潜在研究兴趣, 首先使用排序模型 (bge-m3^[13]) 将当天最新的文献向量化并根据研究人员的兴趣进行匹配并排序, 再利用 R1 生成固定模板的文献日报.

2.6 知识库问答

除了模型本身的回答能力外, 我们还使用了

65366 篇 PDF 提取的文本数据作为外部知识库, 并结合 DeepSeek R1 70B 模型进行知识问答任务. 通过对比挂载知识库与未挂载知识库的 DeepSeek R1 70B 模型, 考察其在回答电池专业问题上的差异.

2.7 更深层次思考

除了使用知识库外挂的方法来解决模型在专业知识储备不足的问题外, 我们还通过优化思维链过程并让模型自我迭代学习来提升模型回答的全面性和专业性. 将 2.6 节搭建的知识库使用 DeepSeek R1 70B 进行主动学习思考, 通过延长模型的思考时间和迭代学习来提升其回答的全面性.

3 评测结果与讨论

3.1 文本分类

在 124021 篇电池相关文献中, BERT 模型筛选出 6925 篇无机固态电解质文献, 经人工核查后, 最终确认为 7000 篇. 其精确率为 5949/6925 (约 0.859), 召回率为 5949/7000 (约 0.850), 最终 F1 分数为 0.854. 相比之下, DeepSeek R1 在 BERT 的基础上进行了优化 (阈值设置为 0.05 以便对比), 筛选出 6558 篇无机固态电解质文章, 其精确率为 5788/6558 (约 0.883), 召回率为 5788/7000 (约 0.823), 最终 F1 分数同样为 0.854; 同时我们测试了 QwQ 32B 的结果, 筛选出 7376 篇无机固态电解质文章, 其精确率为 5760/7376 (约 0.781), 召回率为 5760/7000 (约 0.827), 最终 F1 分数同样为 0.801. 比较发现, 监督学习的 BERT 与采用 prompt 策略的 DeepSeek R1 相比, DeepSeek R1 在精确率上表现优于 BERT, 但召回率略低. 整体来看, 两者准确性相当, 两者都优于 QwQ 32B. 值得注意的是, 对于无机固态电解质, 有一些属于复合固态电解质, 这可能会对结果产生一些潜在的影响; 同时发现 DeepSeek R1 也会将一些匹配无机固态电解质的正极/负极判断为无机固态电解质, 这表明从 prompt 设计的角度来看, DeepSeek R1 的准确性仍有提升空间. 对于 QwQ 32B 来说, 其准确性相对低一些, QwQ 在处理指定格式上略弱于 DeepSeek R1.

为了进一步比较 DeepSeek 在零样本任务下实时学习的能力, 我们使用文章自带的关键词对 Qwen2-7B-Instruct 进行了微调, 并将其与 Deep-

Seek R1 进行对比. 我们选择原位固化技术作为筛选目标, 选取了 2024 年与电池相关的 15967 篇文章作为测试集, 并在微调后的 Qwen2-7B-Instruct 和使用 prompt 的 DeepSeek R1 上分别进行了测试. 经过人工初步筛选, 共筛选出 478 篇与原位固化技术相关的论文. 微调后的 Qwen2-7B-Instruct 模型在该测试集上的精确率为 194/478(约为 0.406), 召回率为 194/233(约为 0.833), F1 分数为 0.576. DeepSeek R1 模型的精确率为 454/478(约为 0.950), 召回率为 454/695(约为 0.653), F1 分数为 0.774. 值得注意的是, DeepSeek R1 的召回率较低的主要原因在于其对固化技术的理解可能与人工 SEI(固体-电解质界面)混淆, 这导致了少量通过反应生成人工 SEI 层但并非原位固化的论文被误判为符合筛选标准.

我们的评估显示, DeepSeek R1 模型在科学领域分类任务中表现出较高的准确性, 其 F1 评分总体维持在 0.7—0.8 之间, 性能等同于相关专业博士研究生的水平(准确性与我们通过博士生大量标记并微调的分类模型相当). 然而, 在某些冷门领域(如本文测评的原位固态化技术), 可能受到训练数据匮乏的影响, 其准确性相对于前面的固态电解质由 0.8—0.9 下降至 0.5—0.6, 但是仍然可以采取一些 prompt 工程来优化模型. 尽管如此, R1 表现仍然优于通过有监督微调获得的 Qwen2-7B-Instruct 模型(其准确性在 0.4—0.5 之间). 具体数据结构详见表 1 与表 2.

表 1 无机固态电解质文献分类准确性

Table 1. Accuracy of literature classification for inorganic solid electrolytes.

	MatElab模型	DeepSeek R1	QwQ 32B
精确性	0.859	0.883	0.781
召回率	0.850	0.823	0.827
F1	0.854	0.854	0.801

表 2 原位固化技术文献分类准确性

Table 2. Accuracy of literature classification for *in-situ* solidification technology.

	微调后的Qwen2-7B-Instruct	DeepSeek R1
精确性	0.406	0.950
召回率	0.833	0.653
F1	0.576	0.774

3.2 信息抽取

我们选择 3.1 节的无机固态电解质数据集, 并

使用微调的 Qwen2-7B-Instruct 模型来提取化学式. 经测试, 我们微调的模型在化学式识别准确性方面表现出显著提升(在 MatElab 平台上查看每天的识别结果^[14]), 但这种提升建立在大量有监督数据微调的基础上, 其泛化能力有限, 仅能识别化学式. 为此, 还采用 DeepSeek R1 进行相同的化学式提取任务, 结果发现其在 10604 条数据中识别出 4298 条与无机电解质离子电导率相关的数据. 受工作量限制, 选择前 750 条进行人工核对. 在这 750 条数据中, 模型的准确提取率为 692/750. 经分析, 存在问题的 58 条数据主要源于以下 3 个方面: 首先, 模型未能正确识别指代关系, 例如将上文中作者缩写为 LLZO-0.1 Al 的体系误判为其他相关缩写或者仅给出缩写; 其次, 模型在温度识别上存在错误, 常将材料测量离子电导率时的温度与材料合成时的温度混淆; 最后, 模型将文章引用的内容误认为是本文的研究内容. 尽管如此, DeepSeek R1 在结构化信息提取的准确性上已超越微调的小模型. 为进一步提升模型准确性, 后续可以考虑采用强化学习或 prompt 优化等方法进行改进.

3.3 文章总结与分析

由于 ChatGPT o3-mini 存在上下文限制, 选择 2025 年 2 月 26 日新增的 22 篇固态锂电池文献, 并使用相同的提示词让 o3-mini, Gemini 和 DeepSeek R1 进行总结, 要求它们给出概述、最新进展、热点文章、新材料、启发性思考和总结与展望来考察模型的指令遵从能力(见图 2). 这些总结结果在补充材料(online). 通过对比三者的内容, 发现 o3-mini 的表现最差, 其概述、最新进展和总结与展望的建议过于笼统, 未能有效结合当日文献, 仅能达到速览文献的作用, 无法进一步提供启发性思考. 虽然 Gemini 2.0 和 DeepSeek R1 在这方面表现较好, 但 Gemini 的总结不如 DeepSeek R1 具体且有针对性. 总的来说, DeepSeek R1 在上下文长度和总结分析方面均优于 Gemini 和 o3-mini, Gemini 次之, 而 o3-mini 由于上下文长度限制, 其总结分析未能有效结合当前内容, 表现最差.

3.4 时间尺度研究动态分析

利用训练的电池 BERT 模型筛选出日韩四家企业固态锂电池相关专利, 并结合 3.3 节的结果选取 DeepSeek R1 作为分析依据. 考虑到 LG 和松

下企业的专利数量较多 (超过 DeepSeek R1 单次上下文处理极限即 64000 个 token), 因此我们将 LG 和松下企业的专利各自切割成了三部分, 以确保分析内容在 DeepSeek R1 的单次上下文处理能力范围内. 基于 DeepSeek R1 的分析结果 (图 3, 详见补充材料 (online)), 三星和 LG 的技术发展路线从聚合物逐步过渡到无机固态, 并最终朝向全固态方向发展; 丰田一开始主要专注于硫化物相关技术, 后期则转向全固态结构设计; 松下则从初期

的多种体系探索转向目前主攻的卤化物体系. 结合作者对相关企业及其技术发展趋势的了解, 这一发展趋势具有较高的可信度. DeepSeek R1 详细分析结果见补充材料 (online).

3.5 个性化推荐

基于 3.3 节与 3.4 节的结果, 我们在 MatElab 测试了 DeepSeek R1 和 70B 的每日新增电池文献和近 24 h 电池相关新闻的推荐功 (图 4), 为作者



图 2 DeepSeek R1, ChatGPT o3 min 和 Gemini 2.0 对同样内容的总结比较

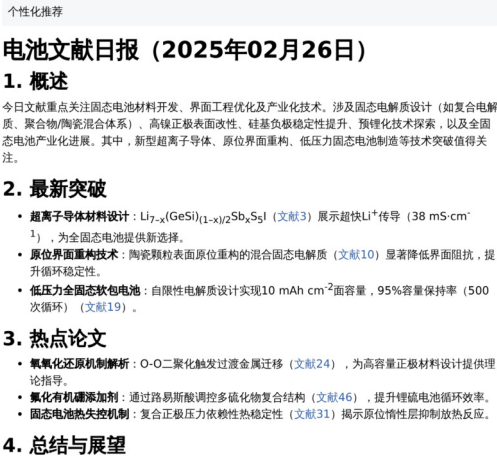
Fig. 2. Comparison of summaries generated by DeepSeek R1, ChatGPT o3 min, and Gemini 2.0 on the same content.



图 3 DeepSeek R1 分析日韩 4 家企业过去 20 年固态电池专利

Fig. 3. Analysis of solid-state battery patents from four Japanese and South Korean companies over the past 20 years by DeepSeek R1.

提供个性化的文献和新闻日报 (分析内容见补充材料 (online)). 经测试, DeepSeek R1 在整体文本总结和筛选与作者兴趣相符的文献准确性方面表现出色, 但在数值处理上仍存在改进空间: 一方面会生成文献中未包含的电池性能数据, 另一方面会出现多篇文献的内容交叉输出 (即将文献 A 和文献 B 的内容混合输出为新内容). 我们发现, 特别是在多篇文献合并分析时, 语言模型会将一些不存在文章的数值结果输出, 这些数据可能是大模型微调的语料数据, 其出现是强化学习所带来的幻觉问题, 也会将两篇文章的内容交叉混合, 这意味着模型在处理多篇文章或者较长上下文时错误率会上升. 这些问题都需要在后续得到优化和改进, 详细内容见第 4 节.



3.6 知识库问答

DeepSeek R1 模型于 2025 年 1 月进行了更新, 其之后的数据库无法进一步更新. 因此, 为了提升模型在专业知识回答能力, 采用检索增强生成 (RAG)^[15] 技术是必要的. 我们通过外挂 65366 篇电池相关 PDF 提取文本构建的知识库, 对 DeepSeek R1 70B 在不同情况下的性能进行了测试. 如图 5 所示, 当我们提出“ LiCoO_2 正极改性策略”这一问题时, 外接知识库提供的答案显著更为丰富和具体 (完整回答参考补充材料 (online)). 这表明, 增加知识库有助于提升模型对具体问题的针对性回答能力. 然而, 外接知识库也存在一些待优化之处: 当知识库规模过大时, 其所需内存可能增加服

电池新闻日报 2025年02月26日

1. 概述

今日新闻重点关注固态电池技术突破、材料界面工程创新及电池产业化动态. 其中, 固态电池领域迎来多项性能与产业化进展, 欧盟发布电池法规新动向, 中国企业 (如宁德时代、CALB) 及国际厂商 (如奔驰、ION Storage Systems) 在技术合作与产能布局上表现活跃.

2. 最新进展

- 固态电池性能突破:** ION Storage Systems 宣布其固态电池实现 25 倍容量提升并完成 1,000 次循环测试, 标志着固态电池在高循环稳定性上的重要进展¹.
- 锡泡沫电极材料:** 新型锡泡沫电极技术成功缓解锂离子电池膨胀问题, 有望提升能量密度与寿命².
- 电池回收创新:** Altium 开发高效电池回收工艺, 结合 AI 优化金属提取效率, 降低环境影响³.
- 换电模式拓展:** 宁德时代与滴滴成立换电合资企业, 推动电动汽车换电标准化⁴.

3. 热点新闻

中国地区

- 宁德时代:** 与滴滴合资布局换电网络, 同时与大众深化换电车型合作⁵.
- CALB:** 宣布投资 20 亿欧元在葡萄牙建设欧洲首座电池工厂, 强化全球供应链⁶.
- Voyah 汽车:** 启动第三代固态电池技术研发, 目标提升能量密度与快充性能⁷.

图 4 DeepSeek R1 分析每日电池文献 (左图) 和电池新闻 (右图)

Fig. 4. Analysis of solid-state battery literatures (left subfigure) and news (right subfigure) by DeepSeek R1.

DeepSeek R1 70B 未挂载知识库回答

LiCoO_2 (钴氧化物) 是一种常用的正极材料, 但其存在循环稳定性差、热稳定性低和钴资源昂贵等问题. 为改善这些缺陷, 研究人员提出了多种改性策略, 主要包括以下几种:

1. 表面改性

- 目的:** 提高材料的热稳定性和防止直接锂化.
- 方法:**
 - 氧化物涂层:** 如 Al_2O_3 、 ZrO_2 、 SiO_2 等, 通过溶胶-凝胶法或化学气相沉积法涂覆.
 - 磷酸盐涂层:** 如 Li_3PO_4 , 通过溶液法或 hydrothermal 法制备.
 - 其他化合物涂层:** 如 $\text{LiCoO}_2@\text{C}$ 、 $\text{LiCoO}_2@\text{graphene}$ 等, 通过物理或化学方法结合.
- 效果:** 改善热稳定性, 降低钴溶解, 提高循环性能.

2. 掺杂改性

- 目的:** 通过引入其他金属或非金属元素, 改善材料的结构稳定性和电子导电性.
- 方法:**
 - 金属掺杂:** 如 Mg、Al、Fe、Mn 等, 通过固相合成或溶液法引入.
 - 非金属掺杂:** 如 F、S、P 等, 通过类似方法引入.
- 效果:** 提高循环稳定性, 优化电化学性能.

3. 纳米结构化

DeepSeek R1 70B 挂载知识库回答

LiCoO_2 作为锂离子电池的重要正极材料, 在高压和高循环次数下往往面临容量衰减和结构不稳定的问题. 为了改善这些缺陷, 研究者提出了多种正极改性策略, 主要包括以下几种方法:

1. 表面涂层

表面涂层是通过在 LiCoO_2 表面涂覆其他材料以改善其电化学性能的常用方法. 常见的涂层材料包括氧化铝 (Al_2O_3)、二氧化硅 (SiO_2)、磷酸锂 (Li_3PO_4)、钛酸锂 (Li_4TiO_4) 等. 这些涂层材料能够有效减少正极与电解液之间的副反应, 抑制钴的溶解, 提高电池的循环稳定性和倍率性能.

例如:

- Al_2O_3 涂层:** 通过溶胶-凝胶法在 LiCoO_2 表面形成 Al_2O_3 涂层, 显著提高了其循环性能和界面稳定性 (DOI: 10.1149/2.107311jes).
- SiO_2 涂层:** 在 LiCoO_2 中添加 SiO_2 形成 $\text{LiCoO}_2/\text{SiO}_2$ 复合材料, 优化了其结构稳定性和电化学性能 (DOI: 10.1016/j.electacta.2005.09.021).

2. 电解液添加剂

在电解液中添加特定的化学物质, 可以在正极表面形成稳定的保护膜, 减少电解液的分解, 改善电池性能. 常用的添加剂包括 $\text{LiPF}_6/\text{C}_6\text{O}_8$ 、 LiPO_2F_2 、4-(三氟甲基)苯甲腈 (4-TB) 等.

例如:

图 5 DeepSeek R1 70B 未挂载知识库 (左图) 和挂载知识库 (右图) 在回答同一问题的差异

Fig. 5. DeepSeek R1 70B without knowledge base mounted (left) and with knowledge base mounted (right) showing differences in responses to the same question.

务器负担. 因此, 在 RAG 工程中, 减少知识库中的非必要内容并优化搜索方法是必要的. 此外, 外挂知识库对使用者提问也有较大要求, 事实上, 当我们提问“钴酸锂正极改性策略”, 该方法在检索准确性上 (相较于 LiCoO_2 正极改性策略) 显著降低. 总之, 外挂知识库会提升语言模型在专业知识回答的准确性和缓解模型幻觉的问题, 但是仍存在知识库构建成本高和上下文过长对资源消耗更高的问题, 这些均需要进一步优化.

3.7 DeepResearch 提升模型回答全面性

除了挂载知识库的方法外, 通过上下文让模型具有实时学习和不断改进的策略既可以考察模型指令遵从和结构化输出能力, 也可以提升模型回答专业问题的全面性, 我们使用图 6 的方法让模型自我迭代与学习来提升问题的准确性. 该方法通过多次迭代并延长模型的思考时间, 让某些自我学习和优化检索内容从而提升问题回答的全面性, 从结果来看对问题的回答更加全面, 然而由于思考时间增加, 模型输出耗时也有提升, 详细的回答可以参考补充材料 ([online](#)).

4 无机固态电解质综合数据库构建

4.1 无机固态电解质数据库现状

当前的材料学数据库主要以计算数据库为主,

实验数据由于成本问题相对匮乏^[16]. 当前的电池数据库主要是以计算为主的大量数据 (如中国科学院物理研究所^[17,18]、上海大学^[19] 等通过高通量键价法计算的离子传输数据库), 或者是少量的人工整理的实验数据库 (如日本东北大学李昊等^[20] 整理的固态电解质数据库共 745 条、Hargreaves 等^[21] 安排了 20 多位相关领域研究人员整理了 820 条数据). 即使安排了大量人力, 当前的实验数据库仍然存在数据量少、种类不全的问题, 仅提供了离子电导率和扩散势垒, 缺乏综合全面的电池数据库. 利用大语言模型提取数据成为一种有效的手段.

结合以上分析, 发现大模型在材料学的主要幻觉存在两种形式: 一是会输出文章中不存在的数据或者结论, 这些数据和结论可能源自于模型预训练或者微调时的语料; 二是同时处理多篇文章或者较长的上下文时, 会将不相关的内容混合, 例如将文章 1 的实验方法和文章 2 的结论混合输出, 这可能是较长上下文所带来的幻觉问题. 总之, 如果能够有效缓解以上问题, 就可以通过大语言模型从文献中提取并生成数据库, 有效解决实验数据库缺失问题.

4.2 数据库处理方法

我们根据以上测评结果, 对大模型可能存在的问题进行和流程优化和重新设计 (图 7): 使用小模型初筛潜在的文献, 将一些可能存在幻觉的文章

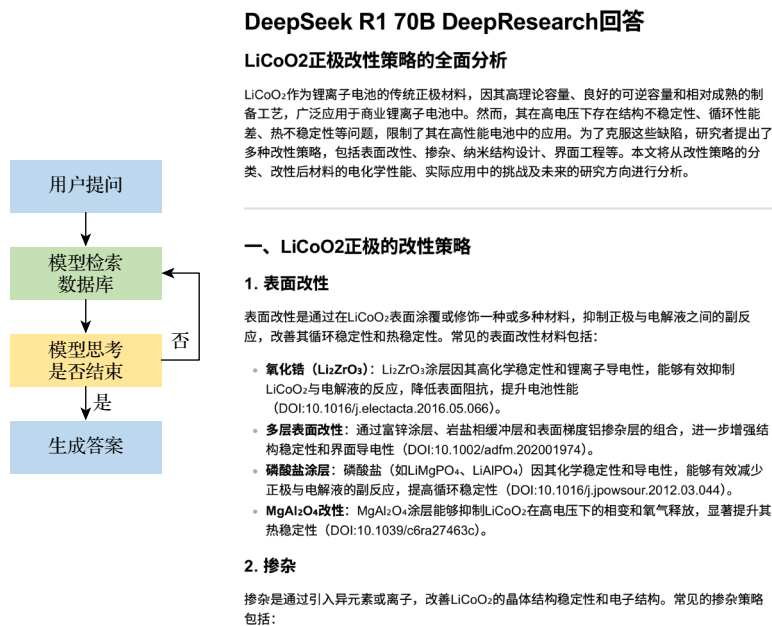


图 6 DeepSeek R1 70B Deep Research 流程 (左图) 和优化后输出 (右图)

Fig. 6. Workflow of deep research of DeepSeek R1 70B (left) and answer by it (right).

(如文章并无相应数据) 排除^[22], 随后利用 DeepSeek V3-0324(经过少量样本测试, V3-0324 与 R1 最初版的准确率几乎一致, 但是 V3 没有思维链内容, 速度更快, 我们选择 V3-0324 来提取数据, 少量异常数据由 R1 提取) 同时使用两套 prompt 提取数据——第 1 个 prompt 在小模型基础上分析潜在的性质, 仅输出可能的性质, 不提取数据; 第 2 个 prompt 在第 1 个的基础上进行限制性提取数据, 最大程度降低了幻觉产生的可能性. 最后, 我们安排一名专业的博士后来修正和剔除异常值 (如无意义的“较高”、“较低”的值), 这部分绝大部分工作是通过正则匹配完成, 少量为博士后手工修正. 整个过程共消耗博士后约 5 天的工作量 (1 天使用代码优化 DeepSeek V3 接口处理数据过程; 1 天通过正则等修正数据值; 2 天时间修正数据和剔除不合理值; 1 天时间基于 Gemini 2.5 Pro 设计并搭建网页, 不包括代码和模型运行耗时).

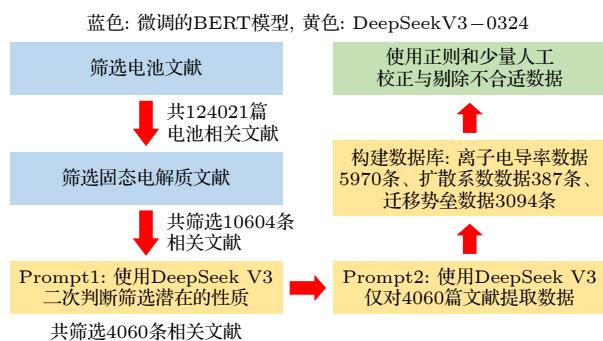


图 7 无机固态电解质数据库构建流程

Fig. 7. Workflow of constructing inorganic solid electrolyte database.

4.3 数据库内容

根据这套流程, 对上述 10604 条文献数据进行过滤, 将开放获取的 4060 篇 PDF 文章利用大模型进行信息提取, 包括离子电导率、扩散系数、迁移势垒、晶体结构、合成方法、颗粒形貌等近 30 种性质进行提取, 汇总出离子电导率数据 5970 条、扩散系数数据 387 条、迁移势垒数据 3094 条, 此外还包括 1000 多条化学、电化学、力学等数据, 涵盖了无机固态电解质所涉及的几乎所有物理、化学、电化学性质, 是当前最大最全的无机固态电解质综合数据库^[23], 同时也体现了大语言模型对科研的应用已经从辅助科研转向主动促进科研发展阶段.

4.4 数据库示例

我们以 Li-Si-P-S-Cl 为例, 检索出 76 条记录, 这些记录可能是不同的实验合成条件, 或者不同的测量方法. 通过这些数据, 可以比较不同实验方法和合成条件下固态电解质性能差异. 以这 76 条数据为例, 同样是 $\text{Li}_{9.54}\text{Si}_{1.74}\text{P}_{1.44}\text{S}_{11.7}\text{Cl}_{0.3}$, Zhang 等^[24] 实验结果是 1.5 mS/cm(室温), Li 等^[25] 测量值为 25 mS/cm(原文为 298 K 数据为 (21 ± 4) mS/cm), Xu 等^[26] 测量值为 16 mS/cm(室温)、Li 等^[27] 测量值为 3 mS/cm (25 °C), Wu 等^[28] 的测量值则为 3.1 mS/cm (室温). 以上结果说明固态电解质的离子电导率受到合成方法、测量方法和颗粒粒径等多种因素影响, 未来需要进一步探究.

5 总结与展望

本研究通过文本分类、信息抽取、文章总结、文章分析与优化、知识库问答和个性化推荐等多个维度, 探讨了语言模型在电池科研过程中的应用. 通过对比实验发现, 虽然小模型在单一任务的准确性上相对优于 DeepSeek R1, 但其需要投入较高的人力标注成本. DeepSeek R1 引入的推理过程显著提升了科研效率, 但其仍继承了原有语言模型的部分问题, 如幻觉现象, 同时在工程实现方面仍有优化空间. 通过外挂知识库和多次迭代并不断让模型学习专业知识来改善回答问题的方法可以改善模型回答专业问题的幻觉问题和提升回答问题的准确性, 但是也增加了模型对算力的要求. 总体而言, DeepSeek R1 代表的推理模型在科学研究领域具有里程碑意义, 其在固态电池信息分析与总结能力已达研究生水准. 未来研究的重点应放在如何优化 DeepSeek R1 的输出质量, 利用私有数据提升其性能, 同时在工程层面提升推理速度. 为了进一步证实这一结论, 我们优化了流程, 通过语言模型和文献数据库汇总出无机固态电解质综合数据库^[23]. 通过数据库发现即使是同一个材料, 不同的合成方法、颗粒粒径、测量方法所得到的实验数据可能也不一样, 未来需要进一步探究影响结果的因素.

对于科研人员特别是研究生而言, 未来考核科研能力的重点将着重放在逻辑思考能力上, 对逻辑过程的思考加上语言模型强大的指令遵从能力, 将有力提升科研效率. 此外, 未来人工智能发展的方

向会从通用性转向为私人助理型——通过个人数据库和知识库结合模型的泛化能力, 每个人都将在已有通用模型的基础上, 通过模型强大的上下文学习和指令遵从能力来构建属于自己的个性化模型, 从而提升科研效率.

感谢中国科学院物理研究所超导实验室 SC02 组在超导文献上的数据贡献, EX4 组在非晶数据的贡献, T03 组吴泉生研究员在拓扑标签的贡献, 怀柔研究部李国栋研究员在热电标签与标准化的贡献, A01 组在催化标签的贡献, A06 组在超快电镜的标签贡献, 部分训练数据来源于中国物理学会和物理研究所的“四刊”数据. 模型测试数据来源于物理所官网实验室介绍、科研进展中整理的数据库, 该数据由科技处胡颖老师提供, 在测试过程的标签选择了未在训练集使用的标签. 本文的 DeepSeek 是通过微软 Azure 等平台获取, 数据库网页是使用 Gemini2.5 Pro 构建并搭建部署在数据中心服务器上, 作者对数据中心翁红明研究员表示感谢.

数据可用性说明

本工作的所有方法与相关数据已经全部公开在中国科学院物理研究所电子实验平台 (<https://in.iphy.ac.cn/eln/#/recusertype>) 和中国科学院凝聚态物质科学数据中心 (网址 <https://cmpdc.iphy.ac.cn/literature/SSE.html>. DOI: <https://doi.org/10.57760/sciencedb.j00213.00172>), 同时在补充材料 (online) 中提供了新手指导以便重复本工作全部实验内容.

参考文献

- [1] ChatGPT Website <https://chatgpt.com/>
- [2] DeepSeek-AI 2025 [arXiv: 2501.12948](https://arxiv.org/abs/2501.12948) [cs.CL]
- [3] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, Polosukhin I 2017 [arXiv: 1706.03762](https://arxiv.org/abs/1706.03762) [cs.CL]
- [4] Radford A, Narasimhan K, Salimans K, Sutskever I 2018 [OpenAI Blog](https://openai.com/blog)
- [5] Devlin J, Chang M W, Lee K, Toutanova K 2018 [arXiv: 1810.04805](https://arxiv.org/abs/1810.04805) [cs.CL]
- [6] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I 2019 [OpenAI Blog](https://openai.com/blog)
- [7] Brown T B, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D M, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D 2020 [arXiv: 2005.14165](https://arxiv.org/abs/2005.14165) [cs.CL]

- [8] Wei J, Wang X Z, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E, Le Q, Zhou D 2022 [arXiv: 2201.11903](https://arxiv.org/abs/2201.11903) [cs.CL]
- [9] OpenAI o1 Hub <https://openai.com/zh-Hans-CN/o1/>
- [10] Yang A, Yang B S, Hui B Y, Zheng B, Yu B W, Zhou C, Li C P, Li C Y, Liu D Y H, Huang F, Dong G T, Wei H R, Lin H, Tang J L, Wang J L, Yang J, Tu J H, Zhang J W, Ma J X, Xu J, Zhou J R, Bai J Z, He J Z, Lin J Y, Dang K, Lu K M, Chen K Q, Yang K X, Li M, Xue M F, Ni N, Zhang P, Wang P, Peng R, Men R, Gao R Z, Lin R J, Wang S J, Bai S, Tan S N, Zhu T H, Li T H, Liu T Y, Ge W B, Deng X D, Zhou X H, Ren X Z, Zhang X Y, Wei X P, Ren X C, Fan Y, Yao Y, Zhang Y C, Wan Y, Chu Y F, Cui Z Y, Zhang Z Y, Fan Z H 2024 [arXiv: 2407.10671](https://arxiv.org/abs/2407.10671) [cs.CL]
- [11] Zeng A H, Xu B, Wang B W, Zhang C H, Yin D, Zhang D, Rojas D, Feng G Y, Zhao H L, Lai H Y, Yu H, Wang H N, Sun J D, Zhang J J, Cheng J L, Gui J Y, Tang J, Zhang J, Sun J Y, Li J Z, Zhao L, Wu L D, Zhong L C, Liu M D, Huang M L, Zhang P, Zheng Q K, Lu R, Duan S Q, Zhang S D, Cao S L, Yang S X, Tam W L, Zhao W Y, Liu X, Xia X, Zhang X H, Gu X T, Lü X, Liu X H, Liu X Y, Yang X Y, Song X X, Zhang X K, An Y F, Xu Y F, Niu Y L, Yang Y T, Li Y Y, Bai Y S, Dong Y X, Qi Z H, Wang Z Y, Yang Z, Du Z X, Hou Z Y, Wang Z H 2024 [arXiv: 2406.12793](https://arxiv.org/abs/2406.12793) [cs.CL]
- [12] Gemini 2.5 Pro Gemini 官网 <https://gemini.google.com/app>
- [13] Chen J L, Xiao S T, Zhang P T, Luo K, Lian D F, Liu Z 2024 [arXiv: 2402.03216](https://arxiv.org/abs/2402.03216) [cs.CL]
- [14] MatElab 平台 <https://in.iphy.ac.cn/eln/#/recusertype>
- [15] Lewis P, Perez E, Piktus A, Petroni A, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W T, Rocktäschel T, Riedel S, Kiela D 2000 [arXiv: 2005.11401](https://arxiv.org/abs/2005.11401) [cs.CL]
- [16] Wu S Y, Wang Y Q, Xiao R J, Chen L Q 2020 *Acta Phys. Sin.* **69** 226104 (in Chinese) [吴思远, 王宇琦, 肖睿娟, 陈立泉, 2020 物理学报 **69** 226104]
- [17] 离子输运数据库 <http://e01.iphy.ac.cn/bmd>
- [18] Xiao R J, Li H, Chen L Q 2015 *Sci. Rep.* **5** 14227
- [19] He B, Chi S T, Ye A J, Mi P H, Zhang L W, Pu B W, Zou Z Y, Ran Y B, Zhao Q, Wang D, Zhang W Q, Zhao J T, Adams S, Avdeev M, Shi S Q 2020 *Sci. Data* **7** 151
- [20] Yang F L, dos Santos E C, Jia X, Sato R, Kisu K, Hashimoto Y, Orimo S, Li H 2024 *Nano Mater. Sci.* **6** 256
- [21] Hargreaves C J, Gaultois M W, Daniels L M, Watts E J, Kurlin V A, Moran M, Dang Y, Morris R, Axandra M, Thompson K, Wright M A, Beluvalli-Eshwarappa P, Blanc F, Collins C M, Crawford C A, Duff B B, Evans J, Gamon J, Han G P, Leube B T, Niu H J, Perez A J, Robinson A, Rogan O, Sharp P A, Shoko E, Sonni M, Thomas W J, Vasylenko A, Wang L, Rosseinsky M J, Dyer M S 2023 *npj Comput. Mater.* **9** 9
- [22] Wu S Y, Zhu T N, Tu S J, Xiao R J, Yuan J, Wu Q S, Li H, Weng H M 2024 *Chin. Phys. B* **33** 050704
- [23] 无机固态电解质材料数据库 <https://cmpdc.iphy.ac.cn/literature/SSE.html>
- [24] Zhang Y, Xie M X, Zhang W, Yan J L, Shao G Q 2020 *Mater. Lett.* **266** 127508
- [25] Li Y X, Daikuhara S, Hori S, Sun X Y, Suzuki K, Hirayama M, Kanno R 2020 *Chem. Mater.* **32** 8860
- [26] Xu R C, Wu Z, Zhang S Z, Wang X L, Xia Y, Xia X H, Huang X H, Tu J P 2017 *Chem. Eur. J.* **23** 13950
- [27] Li X L, Peng W X, Tian R Z, Song D W, Wang Z Y, Zhang H Z, Zhu L Y, Zhang L Q 2020 *Electrochim. Acta* **363** 137185
- [28] Wu F, Fitzhugh W, Ye L H, Ning J X, Li X 2018 *Nat. Commun.* **9** 4037

DATA PAPER

Evaluation of the application of large language models in the entire process of battery research and development of a comprehensive database for inorganic solid electrolyte*

WU Siyuan LI Hong[†]*(Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China)*

(Received 29 April 2025; revised manuscript received 20 June 2025)

Abstract

The emergence of large language models has significantly advanced scientific research. Representative models such as ChatGPT and DeepSeek R1 have brought notable changes to the paradigm of scientific research. While these models are general-purpose, they have demonstrated strong generalization capabilities in the field of batteries, especially in solid-state battery research. In this study, we systematically screen 5309268 articles from key journals up to 2024, and accurately extract 124021 papers related to batteries. Additionally, we comprehensively search through 17559750 patent applications and granted patents from the European Patent Office and the United States Patent and Trademark Office up to 2024, identifying 125716 battery-related patents. Utilizing these extensive literature and patents, we conduct numerous experiments to evaluate the structured output capabilities of knowledge base, contextual learning, instruction adherence, and language models. Through multi-dimensional model evaluations and analyses, the following points are found. First, the model exhibits high accuracy in screening literature on inorganic solid-state electrolytes, equivalent to the level of a doctoral student in the relevant field. Based on 10604 data entries, the model demonstrates good recognition capabilities in identifying literature on in-situ polymerization/solidification technology. However, its understanding accuracy for this emerging technology is slightly lower than that for solid-state electrolytes, requiring further fine-tuning to improve accuracy. Second, through testing with 10604 data entries, the model achieves reliable accuracy in extracting inorganic ionic conductivity data. Third, based on solid-state lithium battery patents from four companies in South Korea and Japan over the past 20 years, this model proves effective in analyzing historical patent trends and conducting comparative analyses. Furthermore, the model-generated personalized literature reports based on the latest publications also show high accuracy. Fourth, by utilizing the iterative strategy of the model, we enable DeepSeek to engage in self-reflection thinking, thereby providing more comprehensive responses. The research results indicate that language models possess strong capabilities in content summarization and trend analysis. However, we also observe that the model may occasionally experience issues with numerical hallucinations. Additionally, while processing a large number of battery-related data, there is still room for optimization in engineering applications. According to the characteristics of the model and the above test results, we utilize the DeepSeek V3-0324 model to extract data on inorganic solid electrolyte materials, including 5970 ionic conductivity entries, 387 diffusion coefficient entries, and 3094 migration barrier entries. Additionally, it includes over 1000 data entries related to chemical, electrochemical, and mechanical properties, covering nearly all physical, chemical, and electrochemical properties related to inorganic solid electrolytes. This also means that the application of large language models in scientific research has shifted from auxiliary research to actively promoting its development. The datasets presented in this paper may be available at the website: <https://cmpdc.iphy.ac.cn/literature/SSE.html> (DOI: <https://doi.org/10.57760/sciencedb.j00213.00172>).

Keywords: solid state battery, artificial intelligence, natural language process, materials database**PACS:** 07.05.Kf, 82.47.Aa**DOI:** 10.7498/aps.74.20250572**CSTR:** 32037.14.aps.74.20250572

* Project supported by the National Key Research and Development Program of China (Grant No. 2022YFB2502200) and the National Natural Science Foundation of China (Grant No. 2239304).

[†] Corresponding author. E-mail: hli@iphy.ac.cn

大语言模型在电池科研全流程应用的测评与无机固态电解质综合数据库构建

吴思远 李泓

Evaluation of the application of large language models in the entire process of battery research and development of a comprehensive database for inorganic solid electrolyte

WU Siyuan LI Hong

引用信息 Citation: *Acta Physica Sinica*, 74, 160701 (2025) DOI: 10.7498/aps.74.20250572

CSTR: 32037.14.aps.74.20250572

在线阅读 View online: <https://doi.org/10.7498/aps.74.20250572>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

人工智能赋能量子通信与量子传感系统

Research progress of artificial intelligence empowered quantum communication and quantum sensing systems

物理学报. 2025, 74(12): 120301 <https://doi.org/10.7498/aps.74.20250322>

“机器微纳光学科学家”：人工智能在微纳光学设计的应用与发展

“Machine micro/nano optics scientist” : Application and development of artificial intelligence in micro/nano optical design

物理学报. 2023, 72(11): 114204 <https://doi.org/10.7498/aps.72.20230208>

基于去噪概率扩散模型的蝠鲼流场智能化预测

Intelligent prediction of manta ray flow field based on a denoising probabilistic diffusion model

物理学报. 2025, 74(10): 104701 <https://doi.org/10.7498/aps.74.20241499>

叠层相干衍射成像算法发展综述

Review of development for ptychography algorithm

物理学报. 2023, 72(5): 054202 <https://doi.org/10.7498/aps.72.20221889>

固态电解质电池锂枝晶生长机械应力-热力学相场模拟研究

Mechanical stress-thermodynamic phase-field simulation of lithium dendrite growth in solid electrolyte battery

物理学报. 2023, 72(22): 220201 <https://doi.org/10.7498/aps.72.20230824>

量子人工智能中的对抗学习

Adversarial learning in quantum artificial intelligence

物理学报. 2021, 70(14): 140302 <https://doi.org/10.7498/aps.70.20210789>