

# 基于大语言模型的数据查询机器人在医学领域的应用



全筱筱<sup>1,2</sup>, 熊文举<sup>3</sup>, 潘军杰<sup>3</sup>, 曾华堂<sup>4,5</sup>

1. 深圳市第二人民医院 (深圳大学第一附属医院) 团委 (广东深圳 518035)
2. 武汉大学政治与公共管理学院 (武汉 430072)
3. 深圳市第二人民医院 (深圳大学第一附属医院) 信息中心 (广东深圳 518035)
4. 清华大学万科公共卫生与健康学院 (北京 100084)
5. 深圳市卫生健康发展研究和数据管理中心 (广东深圳 518028)

**【摘要】**本研究对大语言模型 (large language model, LLM)、数据查询机器人 (data query robot, DQR) 的发展历程和研究现状进行了介绍, 同时通过实证分析, 探讨了在数字医学领域中, 基于 LLM 的 DQR 的实际应用效果及其在处理医疗数据查询和分析的复杂任务中的作用, 证实了基于 LLM 的 DQR 能为非技术人员提供一个直观且便捷的工具, 显著提升医疗数据的查询效率和分析能力。此外, 本文还探讨了 LLM 和 DQR 技术在当前应用中的局限性及未来发展潜力, 为进一步的研究和应用提供参考。

**【关键词】**大语言模型; 数据查询机器人; 数字医学; 自然语言处理; 深度学习  
**【中图分类号】**R 311 **【文献标识码】**D

## Application of data query robots based on large language models in the medical field

QUAN Xiaoxiao<sup>1,2</sup>, XIONG Wenju<sup>3</sup>, PAN Junjie<sup>3</sup>, ZENG Huatang<sup>4,5</sup>

1. Youth League Committee, Shenzhen Second People's Hospital (First Affiliated Hospital of Shenzhen University), Shenzhen 518035, Guangdong Province, China
2. School of Political Science and Public Administration, Wuhan University, Wuhan 430072, China
3. Department of Information Center, Shenzhen Second People's Hospital (First Affiliated Hospital of Shenzhen University), Shenzhen 518035, Guangdong Province, China
4. Vanke School of Public Health, Tsinghua University, Beijing 100084, China
5. Shenzhen Health Development Research and Data Management Center, Shenzhen 518028, Guangdong Province, China

Corresponding author: ZENG Huatang, Email: zht22@mails.tsinghua.edu.cn

**【Abstract】**This study introduced the development history and current research status of large language model (LLM), data query robot (DQR). Meanwhile, through empirical analyses, the practical application effect of LLM-based DQR and its role in dealing with the complex tasks of medical data querying and analysis in the field of digital medicine was

DOI: 10.12173/j.issn.1004-5511.202312071

基金项目: 科技创新 2030 “新一代人工智能”重大项目 (2021ZD0114100); 深圳市科创委基础研究专项面上项目 (JCYJ20220530151209022); 国家卫生健康委医院管理研究项目 (YLZLXZ22G020)

通信作者: 曾华堂, 博士, 研究员, Email: zht22@mails.tsinghua.edu.cn

explored, and it was confirmed that LLM-based DQR could provide non-technical people with an intuitive and convenient tool to significantly improve the querying efficiency and analysis capability of medical data. In addition, this paper discusses the limitations and potential of future development of LLM and DQR techniques in current applications, providing reference for further research and applications.

**【Keywords】** Large language model; Data query robot; Digital medicine; Natural language processing; Deep learning

大语言模型 (large language model, LLM) 和数据查询机器人 (data query robot, DQR) 是当前自然语言处理领域的两大前沿技术, 受到了广泛的关注和研究。LLM 是一类基于深度学习的人工智能模型, 具备强大的自然语言理解和生成能力, 如生成式预训练变换器-3 (generative pre-trained transformer-3, GPT-3)、双向编码器表示变换器 (bidirectional encoder representations from transformers, BERT) 等<sup>[1-3]</sup>。DQR 是一类利用自然语言进行数据库查询和信息检索的智能系统, 能够以用户友好的方式帮助医务人员访问和分析庞大的医学数据, 从而提高医学研究和临床决策的效率。

LLM 和 DQR 在医学领域的应用具有重要性和紧迫性。随着电子病历、临床数据、医学文献等医学数据的爆炸性增长, 如何高效处理和分析这些数据已成为医学界的重要挑战。传统的数据处理方法难以应对这些数据的规模和复杂性, 导致信息获取的滞后和决策的延迟, 而 LLM 和 DQR 能通过自动化和智能化的方式, 高效解析和处理海量医学数据, 提高信息检索的准确性和速度, 显著提升医学研究和临床实践的效率<sup>[4]</sup>。

据此, 本文旨在深入探讨 LLM 和 DQR 在医学领域的应用, 分析基于 LLM 的 DQR 在处理医疗数据查询和分析复杂任务中的应用效果。

## 1 大语言模型技术在医学领域的应用进展

### 1.1 发展历程

LLM 的发展历程可追溯至早期的自然语言处理研究, 但其爆发式增长始于近年来深度学习技术的迅猛发展。最早的语言模型采用传统的基于规则和统计的方法, 但在规模和准确性上受到了严重的制约。随着深度学习技术的兴起, 特别是递归神经网络 (recurrent neural network, RNN) 和长短时记

忆网络 (long short-term memory, LSTM) 的引入, 语言模型的性能开始有了显著提升<sup>[5]</sup>。

然而, 真正引领 LLM 革命的是 Transformer 架构的提出, 尤其是 2018 年发布的 BERT 模型<sup>[6]</sup>。BERT 的关键创新在于使用了大规模的无监督预训练数据, 通过自注意力机制实现对文本的双向编码, 使得模型能够更好地理解上下文和语境。BERT 的出现引领了 LLM 的新时代, 随后相继推出的 GPT-2、T5、GPT-3 等模型规模更大、性能更强, 极大地推动了自然语言处理领域的发展<sup>[7-10]</sup>。值得注意的是, BERT 模型自 2018 年推出以来, 已衍生出多个增强版本, 如 RoBERTa、ALBERT 等, 这些版本通过优化模型结构和训练过程, 进一步提高了模型在各类自然语言处理任务中的性能。同时, GPT 系列也在不断优化, 最新的 GPT-4 模型相较于 GPT-3, 在理解深度、生成质量和多样性上都有了显著的提升。这些模型的迭代, 不仅推动了自然语言处理技术的前沿进展, 也为医学领域的应用提供了更强大的工具<sup>[11]</sup>。

### 1.2 关键技术与工作原理

LLM 的关键技术包括预训练和微调。在预训练阶段, 模型通过处理大规模文本数据来学习语言的语法、语义和世界知识, 形成通用的语言表示。预训练通常采用遮蔽语言模型 (masked language model, MLM) 任务, 模型需要根据上下文来预测被遮蔽的词语。在微调阶段, 模型通过在特定任务上进行有监督训练, 如文本分类、命名实体识别等, 来适应特定应用领域的需求<sup>[12-13]</sup>。

LLM 的工作原理基于神经网络, 特别是 Transformer 架构。Transformer 模型包括多层编码器和解码器, 通过自注意力机制实现文本序列的编码和解码。模型通过多头注意力机制来捕捉不同位置和关系的信息, 同时通过前馈神经网络来进行特征的映射和变换。这种架构使得模型能够在处理长文本序列时保持较好的性能<sup>[14]</sup>。

### 1.3 应用现状

LLM 的应用已经涵盖各个领域<sup>[15-16]</sup>，特别是在医学领域取得了较大的成就<sup>[17-18]</sup>，被用于自动化生成医学文本摘要、疾病诊断、药物发现、临床决策支持等任务。例如，GPT-3 和 BERT 等模型能自动提取医学文献中的关键信息，帮助医生更快速地了解最新的医学研究成果。同时，它们还可用于分析患者的临床数据，提供个性化的诊断和治疗建议<sup>[19-20]</sup>。在疾病诊断方面，利用 BERT 模型进行电子病历分析，可以自动提取患者的病史和诊断信息，辅助医生进行准确诊断<sup>[11]</sup>。药物发现是 LLM 的重要应用领域之一。通过对海量医学文献和临床试验数据的分析，GPT-3 和 BERT 可以快速筛选出潜在的药物靶点，加速新药研发进程；同时利用 LLM 预测药物的副作用和相互作用，可为药物安全性评估提供重要参考。临床决策支持方面，LLM 也展现了强大的能力<sup>[21]</sup>。GPT-3 能够生成个性化的治疗建议，结合患者的病史和最新医学研究，提供更加精准的治疗方案。一项研究中，利用 GPT-3 生成的治疗建议在临床试验中表现出较高的准确性和实用性，有助于提高临床决策的质量<sup>[12]</sup>。

综上，LLM 不仅提高了医学数据处理的效率和准确性，还为医学研究和临床实践提供了新的方法和工具，极大地推动了医学领域的发展。

## 2 数据查询机器人在医学领域的应用进展

### 2.1 传统医学数据查询方法的局限性

传统医学数据查询方法主要依赖于人工搜索和专业数据库查询，这些方法在一定程度上能够满足医学研究和临床实践的需求，但也存在明显的局限性。首先是信息过载问题，医学领域的文献、临床数据和病例数量庞大，医生和研究人员需要花费大量时间来查找相关信息，传统查询方法往往不能有效应对海量信息，导致信息的遗漏和滞后<sup>[22]</sup>。其次，传统查询方法通常需要用户具备医学领域的专业知识，以便正确构建查询语句和解释检索结果，这对于非专业人士来说是一项巨大的挑战<sup>[23]</sup>。更主要的是人工查询和数据库查询通常耗时耗力，限制了医生和研究人员的工作效率，尤其是在需要迅速获取信息的紧急情况下，查询效率低下。

### 2.2 数据查询机器人技术的优势及其在医学领域的应用潜力

DQR 则在上述方面展现出了明显的优势。首先，DQR 利用自然语言处理技术，能够快速解析用户的查询需求，并在海量数据中迅速找到相关信息，显著提高了查询效率<sup>[22]</sup>。其次，DQR 在准确性方面表现出色，通过智能算法和深度学习模型，能够精确匹配查询内容，减少信息遗漏和错误。同时，DQR 还具有很高的用户友好性，不需要用户具备专业的医学知识，只需用自然语言输入查询内容即可得到准确结果，降低了使用门槛，使得非专业人士也能方便使用<sup>[1]</sup>。此外，DQR 还能根据用户的需求和兴趣，提供定制化的智能推荐和个性化服务，这不仅帮助用户发现重要信息，还能提供更加贴合个人需求的查询结果<sup>[24]</sup>。总的来说，DQR 的引入为医学领域提供了一种更高效、更智能、更个性化的信息处理方式，有望在未来的医疗研究和实践中发挥越来越重要的作用。

### 2.3 数据查询机器人在医学领域中的研究现状

目前，许多机构和研究团队致力于开发具有医学领域专长的 DQR，以解决医学信息获取的难题，DQR 现已应用于医学文献检索、临床数据分析、疾病诊断等多个方面，为医学研究和医疗实践提供了有力支持，其主要工作流程示意图如图 1 所示。

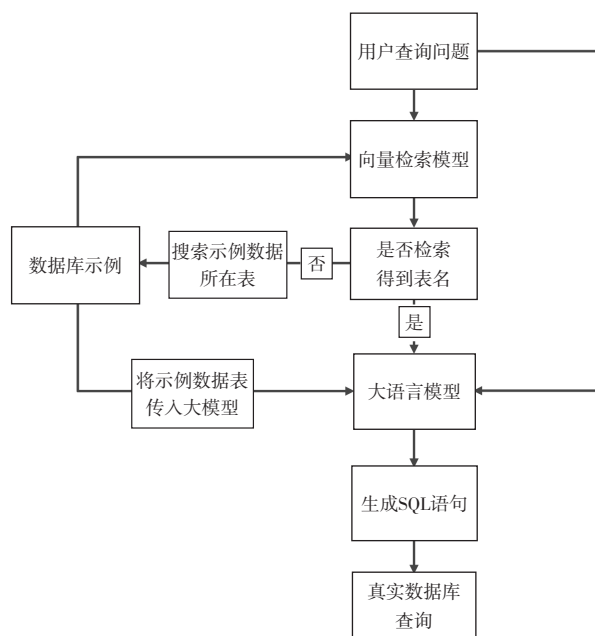


图1 数据查询机器人工作流程图

Figure 1. Workflow of data query robot

### 3 大语言模型与数据查询机器人的设计与实现

#### 3.1 大语言模型与数据查询机器人的整合方法

本文通过 OpenAI 提供的 API 接口访问 GPT-3 模型，通过示例以呈现针对特定医学领域的重训练过程。首先，项目团队筛选和预处理了大量的医学文献，标准如下：①选择过去 5 年内在 PubMed 和 Medline 等权威数据库中发表的经过同行评审的研究论文和临床试验报告，确保数据的时效性和科学性；②文献内容需与数字医学、临床诊断支持系统及自然语言处理应用相关，重点涵盖临床诊断、医疗数据管理等领域，以支持 LLM 在医学数据查询中的应用。其次，利用 OpenAI 提供的 API 接口，将这些经过预处理的医学数据集输入到 GPT-3 模型中。考虑到直接训练整个模型的成本和复杂性，本团队采用微调策略，即只调整模型的最后几层，以适应特定的医学任务和查询需求。该策略不仅节省了计算资源，也保证了模型能够快速适应医学领域的特殊需求。为了进一步提升模型的性能和准确性，还引入了强化学习 (reinforcement learning, RL) 技术。具体来说，通过设定一系列基于查询准确性和响应时间的奖励机制，训练模型在提供医学信息查询服务时，能够更加精准且迅速地给出回答。此外，本团队还探索了检索增强生成 (retrieval-augmented generation, RAG) 技术，该方法利用专门的检索系统在医学数据库中查找最相关信息，并将信息作为输入，辅助 GPT-3 生成更加准确和详细的回答。这种结合检索和生成的方法，可极大提高模型处理复杂医学查询的能力。

#### 3.2 在医学数据查询中的工作流程与技术细节

模型整合完成后，LLM 与 DQR 在医学数据查询中的工作流程和技术细节包括以下几个方面 (图 2)：首先，机器人接收用户的自然语言查询，并利用自然语言处理技术解析成机器可以理解的查询任务；其次，模型根据解析后的查询任务，在医学数据集中进行数据检索和分析，以识别出相关的文献、病例或临床数据，并从中提取必要信息；最后，机器人将查询结果以用户易于理解的方式呈现，通常包括自然语言文本或图表形式，

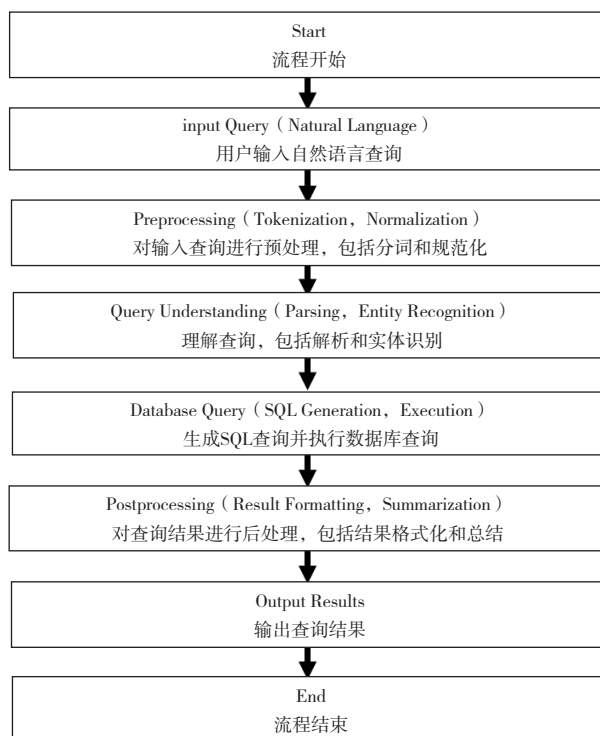


图2 医学数据库查询流程图

Figure 2. Workflow of medical database query

包括文献摘要、疾病诊断建议等，为医生和研究人员提供了有效的信息支持。

#### 3.3 性能评估与对比分析

##### 3.3.1 评价指标体系的建立

为了深入理解基于 LLM 的 DQR 在医学领域的性能，本研究建立了包含准确率 (accuracy)、召回率 (recall) 和精确度 (precision) 的评估指标体系，并选择了包括基于支持向量机 (support vector machine, SVM) 和基于 BERT 模型的自然语言处理工具等常见的医学查询方法进行对比。其中，准确率指模型正确返回相关医学信息的比例；召回率指模型能够检索到所有相关医学信息的比例，而精确度则衡量模型在排除无关信息方面的表现并以 F1 分数作为具体的评价指标。

##### 3.3.2 模型评估结果

使用两个数据集对模型性能进行评估，第一个数据集由 PubMed 数据库中随机抽取的 50 000 条医学研究文献的摘要组成，第二个数据集则包含从某综合性医院的电子病历系统中提取的 20 000 个真实临床病例报告。这些数据集用于模拟实际医学数据查询场景，以评估 GPT-3 模型的性能。

在实验中，GPT-3 模型通过微调进行训练，学习率设置为  $2 \times 10^{-5}$ ，批次大小为 16，共训练 5

轮, 实验在配置有 NVIDIA RTX 3080 GPU 的计算机上完成。实验结果显示, 在医学研究文献摘要数据集中, GPT-3 模型达到了 92% 的查询准确率、94% 的召回率以及 93% 的 F1 分数; 在临床病例报告数据集中, GPT-3 模型实现了 89% 的准确率、91% 的召回率和 90% 的 F1 分数。相比之下, 基于 BERT 的模型在医学研究文献摘要数据集中的准确率为 85%, 召回率为 87%; 而基于 SVM 的方法准确率和召回率分别为 80% 和 82%。

此外, 为了评估模型的扩展性, 研究团队还在不同规模的数据集中对模型进行测试, 包括从 1 万条记录的小规模数据集到 5 万条记录的大规模数据集。结果表明, 随着训练数据的增加, 模型的准确率和召回率均有所提高, 证明了模型在处理大规模数据时具有良好的适应性和扩展性。这些发现不仅证实了基于 GPT-3 的医学查询机器

人在提高查询准确性和效率方面的显著优势, 也展示了其在医学领域应用的广泛潜力。

### 3.4 实际应用中的典型案例与实施效果

在医学领域, 为验证本项目团队构建的 DQR 的实际应用效果, 本研究选择了 6 个典型案例进行测试。这些案例来源于实际临床环境中的真实需求, 涵盖了不同类型的医学查询任务, 包括疾病发病率查询、药物使用情况查询、病床使用情况查询、患者检查结果查询、医学影像报告查询以及门诊排班查询。每个案例均使用项目团队开发的模型进行实际操作, 以评估模型在不同场景下的查询准确率、响应速度和用户满意度。通过这些案例, 旨在展示该模型在处理真实世界的医学数据查询中的有效性, 并为未来的模型优化和应用提供实证基础, 6 例典型案例汇总详见表 1。

表 1 基于大语言模型的数据查询机器人在实际应用中的典型案例与实施效果

Table 1. Typical cases and implementation effects of LLM-based data query robots in practical applications

案例编号	查询内容	查询形式	查询系统	结果类型	查询时间	实施效果	评价
案例1	查询过去五年中45岁以上男性糖尿病发病率	自然语言查询	医院数据库	数据统计	5 s	简化了医疗人员获取和处理医疗数据的工作负担, 提高了查询效率和准确性	精确度高且时间短
案例2	查询某种药物X在过去一年的使用情况	自然语言查询	药物管理系统	数据统计	5 s	简化了药师获取药物使用数据的操作流程, 辅助药物管理和决策	准确快速
案例3	查询当前住院病床使用情况	自然语言查询	医院管理系统	资源管理	3 s	快速反馈病床使用情况, 支持医院管理者的资源调配和管理决策	实时性强
案例4	查询患者张某最近一次体检血液检查结果	自然语言查询	检验数据库	检查结果	4 s	提供清晰准确的检查结果, 支持医生评估患者健康状况和制定治疗方案	准确详细
案例5	查询患者李某最近一次MRI影像报告	自然语言查询	影像数据库	影像报告	6 s	快速检索并解析MRI报告, 提供诊断信息, 提高医生工作效率	提高效率
案例6	查询门诊科室医生排班情况	自然语言查询	门诊管理系统	就医服务	2 s	提供便捷的查询和预约服务, 提升就医便利性和效率	服务优化

## 4 挑战与展望

LLM 和 DQR 在医学领域的应用仍面临挑战。首先, 医学数据的质量和标注对机器人性能有着决定性的影响。由于医学数据复杂多样, 并且需要高度专业的标注, 因此如何获取高质量的医学数据并有效标注的方法有待进一步探讨。其次, 法律和隐私也是机器人应用过程中需要特别关注

的问题。在处理医学数据时, 必须确保数据处理的安全性、合法性以及患者的隐私保护。最后, 机器人还需具备良好的领域适应性, 因为医学领域包括了众多子领域, 相关术语和知识差异巨大, 机器人需要能够适应不同医学领域的需求, 以便更有效地服务于医学研究和实践。

为了应对 LLM 和 DQR 在医学领域应用时面临的挑战和问题, 可以考虑以下几种解决方案和

发展方向。首先,使用数据合成和增强技术来改善医学数据的质量和数量,以及提供更多的标注数据,这将有助于加速模型的训练并提高其性能。其次,研发更先进的隐私保护技术来确保患者数据的安全性和合法性,可能的技术包括数据脱敏和安全传输等。最后,开发领域自适应方法使机器算法能够根据不同医学领域的特点进行调整和适应,从而提高其在多个领域中的性能。

未来,LLM 和 DQR 有望在医学诊断和决策支持方面发挥更重要的作用,能为医生提供更准确的信息和支持,有助于提高医疗保健的效率和质量。通过将语言模型与图像、声音等多模态数据处理相结合<sup>[25]</sup>,机器人可以更全面地理解和处理医学信息,从而提供更为全面的医学支持。最后,机器人还有望根据患者的个性化需求提供定制化的医疗建议和消息,进一步改善医疗服务的质量<sup>[26]</sup>。这些发展趋势表明,LLM 和 DQR 在医学领域的应用前景广阔,未来将成为医疗行业的重要助力。

## 5 结语

LLM 和 DQR 在医学领域的应用为医疗保健提供了新的可能性,能够加速医学知识的获取和医学研究的进展。它们不仅可以提供即时的文献综述和疾病诊断支持,还可以帮助医生制定个性化的治疗方案,提高医疗决策的准确性。此外,它们还可以协助研究人员进行数据挖掘和知识发现,推动医学领域的前沿研究。但 LLM 和 DQR 在医学领域应用中仍然面临挑战和问题,如数据质量、隐私保护和领域适应性等。为了充分发挥它们的潜力,需要不断改进算法和技术,以解决这些问题。

## 参考文献

- 1 Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: opportunities, challenges, and future directions[J]. *JMIR Med Educ*, 2023, 9: e48291. DOI: [10.2196/48291](https://doi.org/10.2196/48291).
- 2 Atallah SB, Banda NR, Banda A, et al. How large language models including generative pre-trained transformer (GPT) 3 and 4 will impact medicine and surgery[J]. *Tech Coloproctol*, 2023, 27(8): 609–614. DOI: [10.1007/s10151-023-02837-8](https://doi.org/10.1007/s10151-023-02837-8).
- 3 周志浩, 宋佳琳, 刘嘉成, 等. 基于人工智能 LLM 技术的虚拟患者系统构建与临床教学应用 [J]. *医学新知*, 2024, 34 (7): 833–842. [Zhou ZH, Song JL, Liu JC, et al. Construction and clinical teaching application of virtual patient system: based on artificial intelligence LLM technology[J]. *Yixue Xinzhi Zazhi*, 2024, 34(7): 833–842.] DOI: [10.12173/j.issn.1004-5511.202403091](https://doi.org/10.12173/j.issn.1004-5511.202403091).
- 4 Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records[J]. *NPJ Digit Med*, 2022, 5(1): 194. DOI: [10.1038/s41746-022-00742-2](https://doi.org/10.1038/s41746-022-00742-2).
- 5 Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Comput*, 1997, 9(8): 1735–1780. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- 6 DeLong KA, Trott S, Kutas M. Offline dominance and zeugmatic similarity normings of variably ambiguous words assessed against a neural language model (BERT)[J]. *Behav Res Methods*, 55(4): 1537 – 1557. <https://doi.org/10.3758/s13428-022-01869-6>.
- 7 Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. *Bioinformatics*, 2020, 36(4): 1234–1240. DOI: [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682).
- 8 Nielsen JPS, von Buchwald C, Grønhoj C. Validity of the large language model ChatGPT (GPT4) as a patient information source in otolaryngology by a variety of doctors in a tertiary otorhinolaryngology department[J]. *Acta Otolaryngol*, 2023, 143(9): 779–782. DOI: [10.1080/00016489.2023.2254809](https://doi.org/10.1080/00016489.2023.2254809).
- 9 Sutherland R, Trembath D, Hodge A, et al. Telehealth language assessments using consumer grade equipment in rural and urban settings: feasible, reliable and well tolerated[J]. *J Telemed Telecare*, 2017, 23(1): 106–115. DOI: [10.1177/1357633X15623921](https://doi.org/10.1177/1357633X15623921).
- 10 Hristidis V, Ruggiano N, Brown EL, et al. ChatGPT vs Google for queries related to dementia and other cognitive decline: comparison of results[J]. *J Med Internet Res*, 2023, 25: e48966. DOI: [10.2196/48966](https://doi.org/10.2196/48966).
- 11 Gorenstein L, Konen E, Green M, et al. Bidirectional encoder representations from transformers in radiology: a systematic review of natural language processing applications[J]. *J Am Coll Radiol*, 2024, 21(6): 914–941. DOI: [10.1016/j.jacr.2024.01.012](https://doi.org/10.1016/j.jacr.2024.01.012).
- 12 Cascella M, Montomoli J, Bellini V, et al. Evaluating the feasibility of ChatGPT in healthcare: an analysis of

- multiple clinical and research scenarios[J]. *J Med Syst*, 2023, 47(1): 33. DOI: [10.1007/s10916-023-01925-4](https://doi.org/10.1007/s10916-023-01925-4).
- 13 Arora A, Arora A. The promise of large language models in health care[J]. *Lancet*, 2023, 401(10377): 641. DOI: [10.1016/S0140-6736\(23\)00216-7](https://doi.org/10.1016/S0140-6736(23)00216-7).
- 14 Bagal V, Aggarwal R, Vinod PK, et al. Molgpt: molecular generation using a transformer-decoder model[J]. *J Chem Inf Model*, 2022, 62(9): 2064-2076. DOI: [10.1021/acs.jcim.1c00600](https://doi.org/10.1021/acs.jcim.1c00600).
- 15 Jagannatha A, Yu H. Calibrating structured output predictors for natural language processing[J]. *Proc Conf Assoc Comput Linguist Meet*, 2020, 2020: 2078-2092. DOI: [10.18653/v1/2020.acl-main.188](https://doi.org/10.18653/v1/2020.acl-main.188).
- 16 Haider L, Baumgartner M, Hayn D, et al. Integration of python modules in a matlab-based predictive analytics toolset for healthcare[J]. *Stud Health Technol Inform*, 2022, 293: 197-204. DOI: [10.3233/SHTI220369](https://doi.org/10.3233/SHTI220369).
- 17 Hu Y, Wen G, Ma J, et al. Label-indicator morpheme growth on LSTM for Chinese healthcare question department classification[J]. *J Biomed Inform*, 2018, 82: 154-168. DOI: [10.1016/j.jbi.2018.04.011](https://doi.org/10.1016/j.jbi.2018.04.011).
- 18 Becerra-Culqui TA, Lonky NM, Chen Q, et al. Patterns and correlates of cervical cancer screening initiation in a large integrated health care system[J]. *Am J Obstet Gynecol*, 2018, 218(4): 429.e1-429.e9. DOI: [10.1016/j.ajog.2017.12.209](https://doi.org/10.1016/j.ajog.2017.12.209).
- 19 Koranteng E, Rao A, Flores E, et al. Empathy and equity: key considerations for large language model adoption in health care[J]. *JMIR Med Educ*, 2023, 9: e51199. DOI: [10.2196/51199](https://doi.org/10.2196/51199).
- 20 Patsakis C, Lykousas N. Man vs the machine in the struggle for effective text anonymisation in the age of large language models[J]. *Sci Rep*, 2023, 13(1): 16026. DOI: [10.1038/s41598-023-42977-3](https://doi.org/10.1038/s41598-023-42977-3).
- 21 Peng C, Yang X, Chen A, et al. A study of generative large language model for medical research and healthcare[J]. *NPJ Digit Med*, 2023, 6(1): 210. DOI: [10.1038/s41746-023-00958-w](https://doi.org/10.1038/s41746-023-00958-w).
- 22 Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing?[J]. *Z Evid Fortbild Qual Gesundheitswes*, 2015, 109(4-5): 285-290. DOI: [10.1016/j.zefq.2015.06.005](https://doi.org/10.1016/j.zefq.2015.06.005).
- 23 Sivarajkumar S, Mohammad HA, Oniani D, et al. Clinical information retrieval: a literature review[J]. *J Health Inform Res*, 2024, 8(2): 313-352. DOI: [10.1007/s41666-024-00159-4](https://doi.org/10.1007/s41666-024-00159-4).
- 24 Chow JCL, Wong V, Li K. Generative pre-trained transformer-empowered healthcare conversations: current trends, challenges, and future directions in large language model-enabled medical chatbots[J]. *BioMedInformatics*, 2024, 4(1): 837-852. DOI: [10.3390/biomedinformatics4010047](https://doi.org/10.3390/biomedinformatics4010047).
- 25 诸宇佳, 韩慧, 卫建华, 等. 人工智能医学影像研究报告规范: CLAIM 检查清单解读[J]. *中国循证医学杂志*, 2023, 23(12): 1478-1484. [Zhu YJ, Han H, Wei JH, et al. Interpretation of checklist for artificial intelligence in medical imaging(CLAIM)[J]. *Chinese Journal of Evidence-Based Medicine*, 2023, 23(12): 1478-1484.] DOI: [10.7507/1672-2531.202303152](https://doi.org/10.7507/1672-2531.202303152).
- 26 袁为标, 陆大军, 李海峰, 等. 基于磁共振表观扩散系数图像的影像组学模型对鉴别前列腺癌和前列腺增生的价值[J]. *中华腔镜泌尿外科杂志(电子版)*, 2022, 16(1): 14-18. [Yuan WB, Lu DJ, Li HF, et al. Apparent diffusion coefficient map based radiomics model in differentiation of prostate cancer and benign prostate hyperplasia[J]. *Chinese Journal of Endourology (Electronic Edition)*, 2022, 16(1): 14-18.] DOI: [10.3877/cma.j.issn.1674-3253.2022.01.003](https://doi.org/10.3877/cma.j.issn.1674-3253.2022.01.003).

收稿日期: 2023 年 12 月 13 日 修回日期: 2024 年 08 月 05 日  
 本文编辑: 桂裕亮 曹越

引用本文: 全筱筱, 熊文举, 潘军杰, 等. 基于大语言模型的数据查询机器人在医学领域的应用[J]. *医学新知*, 2024, 34(9): 1057-1063. DOI: [10.12173/j.issn.1004-5511.202312071](https://doi.org/10.12173/j.issn.1004-5511.202312071)  
 Quan XX, Xiong WJ, Pan JJ, et al. Application of data query robots based on large language models in the medical field[J]. *Yixue Xinzhi Zazhi*, 2024, 34(9): 1057-1063. DOI: [10.12173/j.issn.1004-5511.202312071](https://doi.org/10.12173/j.issn.1004-5511.202312071)