

大模型在智慧医院的应用研究综述

白培发¹ 黄宗浩^{1,2,3} 王奕^{2,3*}

¹(复旦大学附属肿瘤医院厦门医院信息部 福建 厦门 361027)

²(复旦大学附属肿瘤医院信息中心-复旦大学上海医学院肿瘤学系 上海 200032)

³(上海肿瘤疾病人工智能工程技术研究中心 上海 200032)

摘要 随着人工智能的快速发展,大模型正成为引领产业变革的关键力量。大模型在智慧医院中具有非常广阔的应用前景。对大模型技术进行简单介绍;通过分析现有文献和案例研究,梳理归纳目前大模型在智慧医院中的潜在应用研究,并总结大模型在医院应用将面临的主要问题;从技术和应用角度对大模型在智慧医院的发展趋势进行展望,以期对未来智慧医院的创新建设提供参考。

关键词 大模型 大语言模型 生成式人工智能 通用人工智能 智慧医院

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.07.001

A REVIEW ON APPLICATION AND RESEARCH OF LARGE LANGUAGE MODELS IN THE SMART HOSPITALS

Bai Peifa¹ Huang Zonghao^{1,2,3} Wang Yi^{2,3*}

¹(Department of Information, Fudan University Shanghai Cancer Center Xiamen Hospital, Xiamen 361027, Fujian, China)

²(Department of Information Center, Fudan University Shanghai Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China)

³(Shanghai Engineering Research Center of Artificial Intelligence Technology for Tumor Diseases, Shanghai 200032, China)

Abstract With the rapid development of artificial intelligence, large Language models (LLMs) become a key power of driving industrial revolution. LLMs have extremely broad application prospects in the smart hospitals. This paper made a brief introduction of the technology about LLMs, and systematically summarized the potential applications of LLMs in the smart hospitals based on the analysis of existing literature and case studies. The main obstacles and challenges of deploying the LLMs applications in the smart hospitals were outlined. An outlook from the technical and application perspectives on the future development of LLMs' application in the smart hospitals was proposed. This paper can offer valuable information and references for the future innovation and construction of the smart hospitals.

Keywords Large language models LLMs Generative artificial intelligence AG Smart hospitals

0 引言

采用随着 ChatGPT^[1] 的横空出世,大模型受到了广泛的关注,并已成为目前人工智能领域最前沿的技术。大模型又称大语言模型(Large Language Models),是使用大量文本数据训练的深度学习模型,可以较好

地理解自然语言和生成自然、连贯、准确的文本。随着大模型的快速发展,各种大模型层出不穷,例如国外的 PaLM^[2]、LLaMA^[3]、Gemini^[4]等,国内的文心一言^[5]、通义千问^[6]、书生大模型^[7]等。大模型的出现为医疗行业带来了全新的思路,带动了大模型在医疗领域的各种研究^[8-16],并出现了专门用于医疗领域的大模型,例如 Med-PaLM 系列模型^[17-19]、BenTsao^[20]等。

智慧医院建设包括面向患者的“智慧服务”、面向临床的“智慧医疗”和面向管理的“智慧管理”三个层面。虽然近年来人工智能已应用到了智慧医院中,然而这些应用仅限于临床工作流程的一小部分,并且智慧化程度有限,而大模型的出现将有望改变这一状况。

1 大模型技术介绍

1.1 大模型原理

作为一种新型 AI 算法,大模型的目标是基于先前的上下文来预测下一个词序列。为了实现这个目标,大模型通常需要含有超大规模的参数(例如 GPT-4 的参数多达 1.8 万亿),同时使用超大规模的语料通过大模型架构进行预训练,再通过有监督微调(Supervised Fine-tuning, SFT)和强化学习(Reinforcement Learning Human Feedback, RLHF)微调,以满足不同行业 and 不同使用场景的需要。现有大模型的训练架构主要使用 Transformer^[21] 架构,它采用了自注意力机制,有效克服了包括卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)在内的传统深度学习算法的一些缺陷。相比 RNN 无法进行并行计算,Transformer 实现了并行计算,显著提高了计算速度;Transformer 的多头注意力机制能够有效实现数据内在之间的关联,解决了 CNN 中数据缺少关联性的问题^[22]。Transformer 整体架构如图 1 所示。

研发、医学教育、临床医疗、医学研究等。本文仅讨论用于智慧医院建设的临床医疗大模型。

医疗大模型的训练方法如下:首先,商用或者开源大模型等通用大模型为医疗大模型提供最基础的通用能力。其次,通过医学领域专业知识(例如医学文献、医学知识库、药学知识、医学教科书库等)以及医院特有的大数据(例如患者电子病历、医生诊疗规范、临床路径的经验数据等)训练一个具有医学知识的医院预训练大模型,并结合医疗行业中高质量的精标数据对医疗预训练大模型进行有监督微调。最后,采用强化学习技术,通过业务回流数据进一步对模型进行业务对齐精调,得到最终的医疗大模型。基于医疗大模型的智慧医院应用架构如图 2 所示。

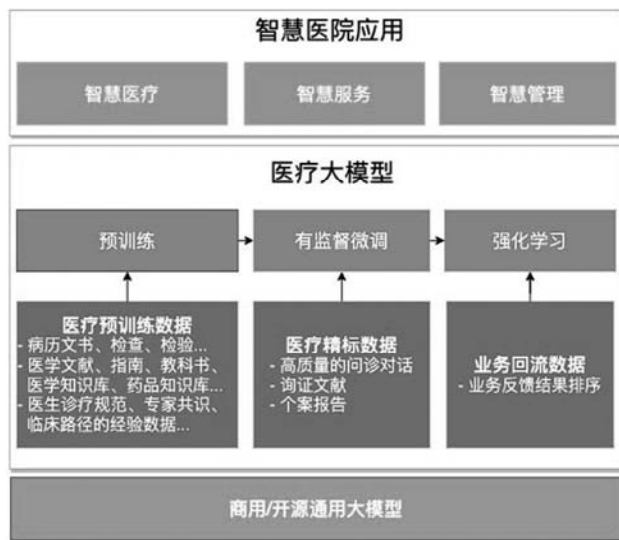


图 2 基于大模型的智慧医院应用架构

2 大模型在智慧医院的潜在应用

大模型在智慧服务、智慧医疗、智慧管理“三位一体”的智慧医院建设中存在着巨大的潜力和广阔的应用前景,主要包括智能医学问答、自动生成病历文书和报告、辅助临床诊疗、辅助临床护理等。

2.1 智能医学问答

基于智能医学问答系统可以为患者提供智能问诊、就医咨询、用药指导等服务,为医生提供医学检索、辅助诊疗决策等服务。迄今为止,大模型在智能医学问答方面已经有不少研究。Samaan 等^[23]通过 ChatGPT 对肥胖外科有关的问题进行问答,结果显示 ChatGPT 对其中 131 个问题(占比 86.8%)提供了“准确且全面”的回答,尤其是对于与“疗效、适应症和手术选项”和“术前准备”相关的问题的回答中高达 93% 以上是“准确且全面”。Li 等^[24]研发的医疗大模型 ChatDoctor 在

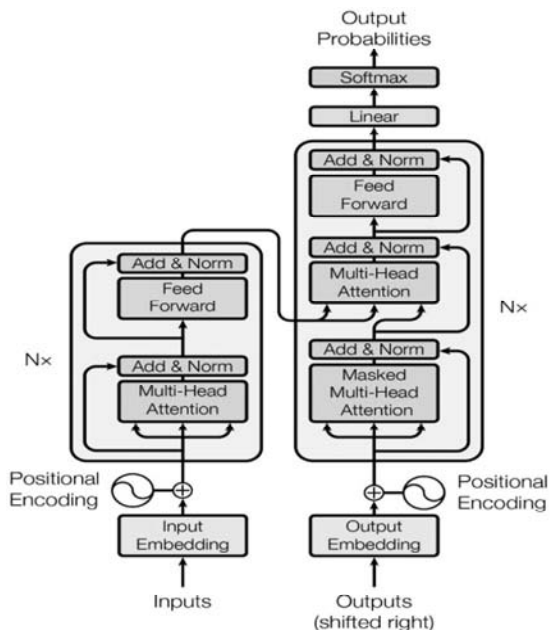


图 1 Transformer 整体架构

1.2 医疗大模型

医疗大模型是通过医疗行业知识进行训练得到的大模型,并用于医疗行业,具体应用场景包括生物药物

处理包括猴痘在内的多种当代医学问答方面表现出色,它不仅能够理解复杂的医学问题,还能即时访问和整合最新的医学信息,提供详尽而准确的回答。谷歌研究院开发的医疗大模型 Med-PaLM^[17]在智能医学问答方面的准确率更是高达 92.6%,这一成绩与人类临床医生的水平(92.9%)已相差无几。

2.2 自动生成病历文书和报告

在临床诊疗过程中,书写病历文书和报告对于医生而言,不仅仅繁琐、耗时,还容易出现填写不规范或缺漏的情况。大模型通过强大的文本理解、分析整合能力可以帮助医生自动生成病历文书和报告。Veen 等^[25]对包含 GPT-4 在内的八种主流大模型进行适配,适配后的大模型总结概括临床文档的质量在完整性、正确性和简洁性方面都优先于人类专家。姜胜耀等^[26]基于 ChatGLM2-6B 大模型设计了多阶段质控模板生成方案,该方案能够有效提取电子病历的关键信息,自动生成规范、完整、满足医疗文书质控要求的出院小结模板内容。谷歌研究团队利用医学大模型 Med-PaLM M^[19]对 246 份胸部 X 光片影像生成放射学报告,研究表明临床医生在高达 40.50% 的情况下更倾向于选择大模型生成的报告。

2.3 辅助临床诊疗

医疗的复杂性和每位患者身体和疾病的个性化差异导致了临床诊疗具有高度的复杂性和个性化。大模型强大的知识体系和分析能力,可以帮助医生从各种复杂的数据中快速提取有价值的信息并进行分析,并提供高质量的临床辅助诊断与决策支持。Liu 等^[27]利用 ChatGPT 进行临床辅助诊断的实验结果表明 ChatGPT 表现不俗,并具有独特的观点,可以帮助专家改进他们自己的临床决策建议。Rao 等^[28]研究发现 ChatGPT 在对乳腺癌筛查方面准确率高达 88.9%。彼特等^[29]利用 GPT-4 成功为一名婴儿诊断出一种非常罕见的疾病。Wang 等^[30]研发的 ChatCAD 将大模型整合到医学图像的计算机辅助诊断,诊断性能提高了 16.42%,还为患者和医生提供了友好的交互式解释和医疗建议。

2.4 辅助临床护理

目前大模型在辅助临床护理的研究还相对欠缺。尽管如此,大模型在这方面的潜力和价值不容小觑。首先,凭借强大的文本处理和整合能力,大模型可以快速准确地提取现病史、既往史病史、诊断报告、生命体征等信息,识别患者健康风险、预测潜在的并发症或治疗反应,帮助护理人员制定积极的护理计划。此外,大模型可以促进患者主动沟通和宣教。大模型可以为患

者及其家属生成量身定制的健康宣教内容,解释复杂医学概念,生成清晰易懂的文字描述,促进护理人员与患者及其家属之间的有效沟通,帮助患者提高患者对治疗和护理计划的理解,从而提高患者对医疗护理的依从性,改善治疗结果。最后,通过大模型,护理人员可以为患者提供以前无法实现的个性化护理。

3 大模型在智慧医院的应用问题

3.1 成本问题

大模型的训练需要依赖强大的算力,而强大的算力需要巨大的成本投入来支撑。例如,训练 ChatGPT-3 的算力成本就已超过数亿或数十亿美元。大模型的算力成本大大限制了其在医院落地的可行性。

3.2 信任问题

由于直接关乎医疗质量安全,临床诊疗过程要求每一步都做到准确、符合相关诊疗规范和指南。而大模型目前在准确性、可解释性、可验证性等方面还存在一定的问题。这些问题加剧了对大模型的不信任感。首先,大模型的训练数据质量、规模、实时性以及算法自身携带的“幻觉”问题^[31]都会影响到结果的准确性。其次,大模型算法的不透明性会带来结果缺乏可解释性,这将令医护人员无法知道大模型回答的依据是什么。最后,由于大模型前所未有的多功能性,大模型的验证也变得困难。

3.3 安全问题

大模型将带来全新的安全问题。研究发现,大模型有潜力被用于尝试识别个人,从而泄露患者的隐私^[32]。大模型还可能带来政治敏感、犯罪违法、偏见、歧视等伦理安全问题。例如,大模型可能会提供获取非法药品的方法。这些新的安全问题都可能给患者和医院造成严重危害。

3.4 责任问题

尽管业界对包括大模型在内的人工智能主要作为主要辅助工具,而不能作为绝对标准,更不能代替医生有相同的共识,当模型产生错误建议或决策引发事故时,如何追溯伤害、划分责任、分配赔偿仍尚未清晰。这些问题在大模型应用上将更加突出,亟待合理的责任管理办法和规章制度来指导规范医疗行为。

3.5 监管问题

尽管大模型在医疗实践中非常具有潜力,但也带来各种风险,因此大模型技术的使用须得到有效的监管,以满足医疗质量和患者安全等需要。但对大模型

的监管本身也是一个巨大的挑战,这涉及医学、计算机、伦理学、法学等不同学科。一方面需要制定相应的规章制度、办法指南进行指引和规范。尽管国家相关部门已颁布了《生成式人工智能服务管理暂行办法》^[33],但仍需要进一步建立医疗机构的大模型监管办法,以及更详细的监管指南和制度。另一方面也需要建立评估标准、建设评估系统持续、系统和透明地对大模型各方面进行评估和系统监管,而这方面目前尚未成熟。

4 大模型在智慧医院的发展趋势

虽然目前大模型技术大多处于学术研究阶段或者应用尝试阶段,但大模型正在快速发展。未来大模型将重塑整个智慧医院应用,贯穿患者服务、临床医疗、医院管理各个环节。下面从技术和应用角度展望大模型在智慧医院的发展趋势。

4.1 医疗大模型在技术上的发展趋势

4.1.1 数据模态走向多模态

由于医院诊疗过程中存在着病历记录、医学影像、基因组学数据等多模态数据,例如,医学影像富含有关患者身体状况的关键信息。同时,医院中大量的多模态数据可以显著增加大模型训练数据的规模,有助于进一步提升大模型的性能。因此,具备多模态能力的医疗大模型在整个诊疗过程中具有更大的潜力和价值。随着 Med-PaLM M^[19]、Visual Med-Alpaca^[34]、XrayGPT^[35] 等优秀多模态大模型的发布,医疗大模型进入了多模态研究阶段,多种数据模态无缝融合也是未来医疗大模型在数据模态上的方向。

4.1.2 规模量级走向轻量化

由于现实中大部分医院受算力和成本的限制,未来医疗大模型在规模量级上将走向轻量化。虽然大模型训练数据量越大、参数规模越大,越有助于大模型性能的提升。但是规模越大意味着训练的成本越高。通过模型剪枝、量化、多轮训练^[36] 等技术对大模型进行压缩后得到的轻量级大模型可实现在保障大模型质量的情况下,还能大幅降低训练成本,且更易于本地化部署,更适合在医院部署和使用。例如,构建于 LLaMa-7B 之上的 Visual Med-Alpaca^[34]、BenTsao^[20] 等医疗大模型由于参数规模相对较小,可实现在单个消费级 GPU 上的部署和运行。这种轻量化方案显著增强了医疗大模型的实用性和可行性。

4.1.3 风险质量走向安全可靠

由于医疗行业具有数据敏感、专业性强、高风险、

低容错等特点,因此信任问题和安全问题成为大模型在医院应用的核心挑战。如何解决这两个问题,确保大模型回答的安全性和可靠性,已成为大模型技术的主要研究方向之一,多种优化方法和技术被不断提出,包括提示词工程 (Prompt Engineering)^[37-39]、微调 (Fine-Tuning)^[40-41] 以及检索增强生成 (Retrieval-Augmented Generation, RAG)^[42-43] 等。其中,RAG 技术通过引入外部专业知识库与大模型的知识相结合,不仅成本低,还能有效地改善了大模型在幻觉、实时性、透明性和安全性方面的问题,但目前 RAG 仍存在部分关键问题有待深入探究,包括长上下文的处理、RAG 与微调的协同作用、RAG 的工程应用等。随着大模型优化技术的不断发展,医疗大模型将变得更加安全可靠。

4.2 基于大模型的智慧医院应用发展趋势

4.2.1 应用交互走向自然对话式交互

大模型带来了全新的自然对话式人机交互方式体验,多轮对话与多模态的能力进一步提升了这种交互体验。未来基于大模型的智慧医院应用将在传统的 UI 界面上融入自然对话式的人机交互。通过友好的对话式人机交互界面,患者、医护人员、医院管理者都可以采用自然对话方式与系统进行互动,提出自定义、个性化、开放性、跨模态的问题,获取准确可靠的信息。

4.2.2 智慧服务走向多功能智慧就医助手

目前医院智慧服务是基于多个应用,各应用提供单一功能的方式向患者提供服务(例如,智能导诊应用提供智能导诊服务,用药指导应用提供用药指导服务)。未来基于大模型技术,医院可建设新型的多功能智慧就医助手,统一提供智能问诊、智能分诊、就医咨询、用药指导等多功能、全方位的患者服务。同时,智慧就医助手还能够根据患者的具体情况和历史数据提供个性化的就医建议和健康方案。

4.2.3 智慧医疗走向全面性人机共诊

目前医院的诊疗大多是依靠人类医生完成。虽然医院存在着部分 AI 应用(例如,肺结节 AI 辅助诊断系统),但目前的 AI 应用都是为了实现特定任务而开发的,只能用于特定功能,而且无法实现自然和深度的人机交互。未来的智慧医疗将依靠大模型实现全面性的人机共诊。未来基于多模态医疗大模型的新型电子病历系统不仅能读懂、听懂并理解医生的指令意图,还能够实现跨模态的数据分析,通过分析患者的各模态数据帮助医生及时发现疾病的早期征兆和异常信号,为医生提供准确的诊断分析、决策建议、治疗方案,推算患者预期生存曲线和可能预后结果,实现医生和大模

型的全面性人机协同会诊。

4.2.4 智慧管理走向顾问式主动管理

大模型对医院的管理系统将产生巨大的转变。未来基于大模型的医院智慧管理系统,不仅是医院管理者的得力“助手”,还将成为医院管理者的高级“顾问”,实现从传统的反应式被动管理到顾问式主动管理的转变。新型的智慧管理系统不仅能理解和回答医院管理者的提问,还能提供基于深度融合医院各种数据分析的专业建议和决策支持,甚至还能够基于最新的医疗技术和市场趋势,提出创新的管理策略和建议,促进医院管理的持续改进和创新。

5 结语

大模型正处于医疗人工智能的创新前沿。这项新技术在智慧医疗、智慧服务、智慧管理“三位一体”的智慧医院建设中存在着巨大的潜力。但同时也面临诸多问题和挑战,主要涉及有关成本、信任、安全、责任、监管等问题。面对这些问题与挑战,不仅要依靠大模型本身技术的发展,还需要跨学科合作、相关政策和法规的制定,共同推动大模型在智慧医院的发展和应用。尽管如此,大模型的发展势不可挡。在技术层面,医疗大模型将走向多模态、轻量化、安全可靠,以更好满足医院落地应用的要求;在应用层面,大模型将重塑智慧医院的所有应用,助力医院的高质量发展,造福广大患者和医护人员。

参 考 文 献

- [1] OpenAI. ChatGPT[EB/OL]. [2024-02-26]. <https://chat.openai.com>.
- [2] Chowdhery A, Narang S, Devlin J, et al. Palm: Scaling language modeling with pathways[J]. *Journal of Machine Learning Research*,2023,24(240):11324-11436.
- [3] Touvron H, Lavril T, Izacard G, et al. Llama: Open and efficient foundation language models[EB]. arXiv:2302.13971,2023.
- [4] Team G, Anil R, Borgeaud S, et al. Gemini: A family of highly capable multimodal models[EB]. arXiv:2312.11805,2023.
- [5] 百度. 文心一言[EB/OL]. [2024-02-26]. <https://yiyan.baidu.com>.
- [6] 阿里云. 通义千问[EB/OL]. [2024-02-26]. <https://tongyi.aliyun.com>.
- [7] 上海人工智能实验室. 书生通用大模型[EB/OL]. [2024-02-26]. <https://intern-ai.org.cn>.
- [8] Moor M, Banerjee O, Abad Z S H, et al. Foundation models for generalist medical artificial intelligence[J]. *Nature*,2023,616,259-265.
- [9] Yeo Y H, Samaan J S, Ng W H, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma[J]. *Clinical and Molecular Hepatology*,2023,29(3):721-732.
- [10] Chintagunta B, Katariya N, Amatriain X, et al. Medically aware GPT-3 as a data generator for medical dialogue summarization[C]//2nd Workshop on Natural Language Processing for Medical Conversations. Online; Association for Computational Linguistics,2021:66-76.
- [11] 王树义,张庆薇,张晋. AIGC时代的科研工作流:协同与AI赋能视角下的数字学术工具应用及其未来[J]. *图书情报知识*,2023,40(5):28-38.
- [12] Ali S R, Dobbs T D, Hutchings H A, et al. Using ChatGPT to write patient clinic letters[J]. *The Lancet Digital Health*,2023,5(4):179-181.
- [13] Biswas S. ChatGPT and the future of medical writing[J]. *Radiology*,2023,307(2):e223312.
- [14] Hisan U K, Amri M. ChatGPT and medical education: A double-edged sword[J]. *Journal of Pedagogy and Education Science*,2023,2(1):71-89.
- [15] Li H Z, Moon J T, Purkayastha S, et al. Ethics of large language models in medicine and medical research[J]. *The Lancet Digital Health*,2023,5(6):333-335.
- [16] Meskó B, Topol E J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare[J]. *NPJ Digital Medicine*,2023,6(1):120.
- [17] Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge[EB]. arXiv:2212.13138,2022.
- [18] Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models[EB]. arXiv:2305.09617,2023.
- [19] Tu T, Azizi S, Driess D, et al. Towards generalist biomedical AI[EB]. arXiv:2307.14334,2023.
- [20] Zhang H B, Chen J Y, Jiang F, et al. HuatuoGPT, towards taming language model to be a doctor[EB]. arXiv:2305.15075,2023.
- [21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//31st International Conference on Neural Information Processing Systems,2017:6000-6010.
- [22] 任欢,王旭光. 注意力机制综述[J]. *计算机应用*,2021,41(S1):1-6.
- [23] Samaan J S, Yeo Y H, Rajeev N, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery[J]. *Obesity surgery*,2023,33:1790-1796.

送^[12],将预警信息推送到安卓通知栏;还将考虑软件更深层次的前后端分离,从而提高软件的鲁棒性。

参 考 文 献

- [1] 张舒,刘道宇. 基于 Android 平台的移动办公软件的设计与实现[J]. 信息技术与信息化,2020(4):55-57.
- [2] 葛佩,李丛宇,张巧珍,等. 一种新型水务移动平台系统[J]. 净水技术,2018,37(11):17-21.
- [3] 肖文娟,王加胜. 基于 Vue 和 Spring Boot 的校园记录管理 Web App 的设计与实现[J]. 计算机应用与软件,2020,37(4):25-30.
- [4] 张志远. 基于 Android 混合模式开发的实习宝软件[D]. 北京:北京邮电大学,2016.
- [5] Oliveira W, Oliveira R, Castor F. A study on the energy consumption of Android app development approaches[C]//2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR),2017.
- [6] An J, Zhang L, Zhang X C, et al. The Effect of HTML5 on information security[R]. Science and Engineering Research Center:Science and Engineering Research Center,2017.
- [7] 陈长庆. 计算机基础应用数据库层次化结构设计[J]. 信息通信,2020(9):93-94.
- [8] Cheng X. Research on file upload based on HTML5[C]//International Conference on Service Systems & Service Management,2014.
- [9] 郭庆燕,张敏,杨贤栋. JQuery Ajax 异步处理 JSON 数据实现气象图片的显示[J]. 计算机应用与软件,2016,33(6):20-22.
- [10] 王汝言,蒋子泉,刘乔寿,等. Android 下 Binder 进程间通信机制的分析与研究[J]. 计算机技术与发展,2012,22(9):107-110.
- [11] 彭飞. 基于 XMPP 协议的 Android 消息推送的分析与设计[J]. 长沙大学学报,2018,32(5):41-43.
- [12] 吴昌政. 基于前后端分离技术的 web 开发框架设计[D]. 南京:南京邮电大学,2020.
- [EB]. MedRxiv:23286254,2023.
- [28] Rao A, Kim J, Kamineneni M, et al. Evaluating ChatGPT as an adjunct for radiologic decision-making[EB]. medRxiv:23285399,2023.
- [29] 彼特·李,凯里·戈德伯格,伊萨克·科恩. 超越想象的 GPT 医疗[M]. 芦义,译. 杭州:浙江科学技术出版社,2023.
- [30] Wang S, Zhao Z H, Ouyang X, et al. Chatcad: Interactive computer-aided diagnosis on medical image using large language models[EB]. arXiv:2302.07257,2023.
- [31] Huang L, Yu W J, Ma W T, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[EB]. arXiv:2311.05232,2023.
- [32] OpenAI. Gpt-4 system card[EB/OL]. [2024-02-26]. <https://cdn.openai.com/papers/gpt-4-system-card.pdf>.
- [33] 中国政府网. 生成式人工智能服务管理暂行办法[EB/OL]. [2024-02-26]. https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm.
- [34] Shu C, Chen B A, Liu F Y, et al. Visual Med-Alpaca: A parameter-efficient biomedical LLM with visual capabilities[EB/OL]. [2024-02-26]. <https://github.com/cambridgeltl/visual-med-alpaca>.
- [35] Thawkar O, Shaker A, Mullappilly S, et al. XrayGPT: Chest radiographs summarization using medical vision-language models[EB]. arXiv:2306.07971,2023.
- [36] Tang Y H, Liu F C, Ni Y S, et al. Rethinking optimization and architecture for tiny language models[EB]. arXiv:2402.02791,2024.
- [37] Wei J, Wang X Z, Schuurmans D, et al. Chain-of-Thought prompting elicits reasoning in large language models[C]//36th International Conference on Neural Information Processing Systems,2022:24824-24837.
- [38] Yao S Y, Yu D, Zhao J, et al. Tree of thoughts: Deliberate problem solving with large language models[C]//37th International Conference on Neural Information Processing Systems,2024:11809-11822.
- [39] Wang J Q, Shi E, Yu S, et al. Prompt engineering for healthcare: Methodologies and applications[EB]. arXiv:2304.14670,2023.
- [40] He J, Li P, Liu G, et al. PeFoMed: Parameter efficient fine-tuning on multimodal large language models for medical visual question answering[EB]. arXiv:2401.02797,2024.
- [41] Xiong H L, Wang S, Zhu Y T, et al. DoctorGLM: Fine-tuning your Chinese doctor is not a herculean task[EB]. arXiv:2304.01097,2023.
- [42] Gao Y F, Xiong Y, Gao X Y, et al. Retrieval-augmented generation for large language models; A survey[EB]. arXiv:2312.10997,2023.
- [43] Chen J W, Lin H Y, Han X P, et al. Benchmarking large language models in retrieval-augmented generation[EB]. arXiv:2309.01431,2023.

(上接第5页)

- [24] Li Y X, Li Z, Zhang K, et al. ChatDoctor: A medical chat model fine-tuned on a large language model meta-AI (LLaMA) using medical domain knowledge[J]. Cureus,2023,15(6):2045-2056.
- [25] Veen D, Uden C, Blankemeier L, et al. Clinical text summarization: Adapting large language models can outperform human experts[EB]. arXiv:2309.07430,2023.
- [26] 姜胜耀,李寅驰,薛万东,等. 基于大模型的多阶段出院小结质控方法研究[J]. 中国卫生信息管理杂志,2023,20(6):888-896.
- [27] Liu S, Wright A P, Patterson B L, et al. Assessing the value of ChatGPT for clinical decision support optimization