

人工智能聊天机器人助力眼科和科学论文写作

惠延年

引用:惠延年. 人工智能聊天机器人助力眼科和科学论文写作. 国际眼科杂志, 2024,24(1):1-4.

作者单位:(710032)中国陕西省西安市,空军军医大学西京医院眼科 全军眼科研究所

作者简介:惠延年,毕业于第四军医大学,主任医师,教授,博士研究生导师,国家人事部“有突出贡献中青年专家”,全国高等医药院校统编教材《眼科学》第五、六版主编,《国际眼科杂志》(IES)和 International Journal of Ophthalmology (IJO)主编,研究方向:眼底病的基础与临床研究、眼外伤.

通讯作者:惠延年. ynlhui@163.com

收稿日期:2023-10-14 修回日期:2023-11-17

摘要

人工智能聊天机器人 ChatGPT 是使用深度学习技术、能对自然语言输入产生类似人类反应的一种大型语言模型 (LLMs)。它由 OpenAI 公司在 2022.11 开发,属于“生成预训练转换器 (GPT)”模型家族的一种,目前可为公众所用。ChatGPT 能够捕捉人类语言的细微差别和复杂性,生成适当的、与上下文相关的响应。它可以帮助医务人员完成各种任务,如研究、诊断、患者监护和医学教育,从确定研究课题到协助临床和实验室诊断,了解各自领域的新进展和科学写作。ChatGPT 在眼科已吸引了越来越多的关注和广泛应用。然而,目前在这些任务中使用 ChatGPT 和其他人工智能工具仍存在一定的局限性、伦理和法律问题,如可信度、剽窃、版权侵犯和偏见。未来的研究将集中在开发新的方法来减轻这些局限性,同时发挥 ChatGPT 在医疗等相关方面的积极作用。

关键词:人工智能(AI);聊天机器人;大型语言模型;生成预训练转换器(GPT);眼科学;科学写作

DOI:10.3980/j.issn.1672-5123.2024.1.01

Artificial intelligence ChatGPT in ophthalmology and scientific writing

Hui Yannian

Department of Ophthalmology, Xijing Hospital, Air Force Medical University; Eye Institute of PLA, Xi'an 710032, Shaanxi Province, China

Correspondence to: Hui Yannian. Department of Ophthalmology, Xijing Hospital, Air Force Medical University; Eye Institute of PLA, Xi'an 710032, Shaanxi Province, China. ynlhui@163.com

Received:2023-10-14 Accepted:2023-11-17

Abstract

• ChatGPT is a large language models (LLMs) that uses deep learning techniques to produce human-like responses to natural language inputs. It belongs to the family of generative pre-training transformer (GPT) models currently publicly available developed by OpenAI in November 2022. ChatGPT is capable of capturing the nuances and intricacies of human language, generating appropriate and contextually relevant responses. It can assist medical professionals in various tasks, such as research, diagnosis, patient monitoring, and medical education, from identifying research programs to assisting in clinical and laboratory diagnosis, to know new developments in their fields and scientific writing. ChatGPT has also attracted increasing attention and widely used in ophthalmology. However, the use of ChatGPT and other artificial intelligence tools in such tasks comes now with several limitations, ethical and legal concerns, such as credibility, plagiarism, copyright infringement, and biases. Future research can focus on developing new methods to mitigate these limitations while harnessing the benefits of ChatGPT in medicine and related aspects.

• KEYWORDS: artificial intelligence (AI); chatbots; large language models; generative pre-training transformer (GPT); ophthalmology; scientific writing

Citation: Hui YN. Artificial intelligence ChatGPT in ophthalmology and scientific writing. Guoji Yanke Zazhi (Int Eye Sci), 2024, 24(1):1-4.

1 大型语言模型聊天机器人的开发、能力与医学应用^[1-4]

人工智能 (artificial intelligence, AI) 的开发应用正在迅猛发展,日新月异,并广泛渗透到人类社会各个领域。聊天机器人 (chatbots, 一种能够理解和使用基于文本的界面生成响应的程序) 是近年受到高度关注的 AI 技术之一。人工智能聊天机器人 ChatGPT 在眼科也吸引了越来越多的评论、验证和广泛应用。它是由 OpenAI 公司在 2022.11 开发的一种聊天机器人,属于一种使用生成预训练转换器 (generative pre-training transformer, GPT) 的大型语言模型 (large language models, LLMs)。实质上它就是一种 AI 软件,使用深度学习技术对自然语言输入产生类似人类的反应,是目前最大的、公开可用的语言模型之一。以“ChatGPT”为关键词检索 PubMed,2022 年仅有 4 篇文献,而 2023 年迄今 (2023.10.14) 检出了 1463 篇。对其关注度之高、应用、检验和评论之广泛,值得引起眼科同道的注意。

生成预训练的 GPT3 有一个转换器架构,通过来自互联网的书籍和文章获得庞大的文本数据集(>4000 亿词)进行训练,生成类似人类的文本。一旦模型得到训练,它就可以通过提示、允许预测下一个单词来生成新文本。然后,模型使用这个预测的单词作为下一个预测的上下文,重复这个过程,直到生成一个完整的句子或段落。ChatGPT 不仅仅是预测下一个单词,因为它针对对话进行了优化,并使用人类反馈进行了训练,这使得它能够在回答问题时理解并回应人类的期望。

基于大量的文本数据库,ChatGPT 通过算法编程来理解自然语言输入,能够捕捉人类语言的细微差别和复杂性,生成适当的、与上下文相关的响应。这些回应是预先编写的,或者是 AI 新生成的。通过识别的关键词或短语来模拟与人类用户的对话,可以集成到各种平台中,例如网站、移动应用程序和消息传递平台。

ChatGPT 通过强化技术不断改进、自然语言处理和机器学习,提高了理解和响应用户需求的能力。具体来说,用户可以在对话中问任何问题,并收到一个快速而充分的、类似于人类的书面答复,如就给定的主题写一篇文章,获取有关感兴趣主题的信息,撰写具有特定语气、特定内容并针对特定人员的电子邮件或信息,还能纠正文本的形狀或改变其措辞,以及解决问题。

据一篇对纳入 60 篇在卫生保健中应用 ChatGPT 报告的系统性综述,85.0% 的报告记录了 ChatGPT 的积极作用,包括在医学研究中的实用性,如有效分析数据集,生成代码,文献综述,节省时间专注于实验设计和药物的发现和开发;在实践中简化工作流程、节省成本、个性化医疗和提高卫生素养;在医疗保健教育上改进个性化学习,注重批判性思维和基于问题的学习;改善科学写作,增强研究的公平性和通用性等。这些积极作用可能促使某些医疗模式的转变。但该文同时指出,96.7% 的报告对 ChatGPT 的使用有担忧,包括伦理、版权、透明度和法律问题,偏见风险、剽窃、缺乏原创性、不准确的内容和幻觉风险;有限的知识风险、不正确的引用、网络安全问题和信息流行病风险等。考虑到这些局限性,目前应该在一定的道德准则下非常谨慎和负责任地使用 ChatGPT^[5]。

2 ChatGPT 在眼科的应用

LLMs 的快速发展为包括眼科教育、医疗临床决策和研究等在内的各领域开辟了应用的可能性,并初步显示了在疾病预防、诊断、治疗、监测和患者支持等方面的巨大潜力。已有 40 多篇报告涉及此方面的应用、检验和评论。

为了检验 LLMs 在眼科知识问答领域的表现,Antaki 等^[1]从两个题库(美国眼科学会基础和临床科学课程 BCSC 自我评估计划和 OphthoQuestions 在线题库)生成了 260 个问题的模拟考试,测试了 ChatGPT 两个版本(“legacy”和 ChatGPT Plus)在眼科问答场所的准确性,正确率以百分比表示。结果,使用 ChatGPT Plus 的准确率分别提高到 59.4% 和 49.2%;在控制考试部分和认知水平的情况下,对更容易的问题准确性提高。另一项研究评估了 GPT-4.0 与 GPT-3.5 在眼科知识评估项目(OKAP,美国眼科住院医师的年度多项选择考试)中的表现,使用

BCSC 题库的 12 个亚专科 180 道题。结果 GPT-4.0 的表现明显优于 GPT-3.5(81% vs 57%) 回答的 167 个 OKAP 问题,提示 ChatGPT 可以用来补充传统的学习方法。未来的研究应集中在改善其模型,成为用于医学教育和实践的最优化方法^[6]。

对眼科考题的测试和验证还有类似的报告。如,Cai 等^[7]利用 BCSC 中的 250 个问题,对 3 个 LLMs(Bing chat-Microsoft 和 ChatGPT 3.5 和 4.0)进行了评估。结果,人类应答者的平均准确率为 72.2%。GPT-3.5 得分最低(58.8%),而其它两种表现相当。与诊断性问题相比,GPT-4.0 在检查型问题上表现出色,但在图像解释方面表现不佳。此研究提示 LLMs 在回答 BCSC 评估项目的问题中与人类受访者表现相似;而错觉和非逻辑推理的频率表明,会话代理的表现还有改进的空间。Mihalache 等^[8]报告使用 OphthoQuestions 实践题库 166 个选择题,ChatGPT 答对了大约一半的问题。其在全科医学类别中表现最好(79%),在视网膜和玻璃体最差(0)。该调查中使用的 ChatGPT 没有正确回答足够的多项选择题,无法为认证提供实质性帮助。

在眼病的辅助诊断应用方面,已有多篇关于神经眼科、角膜病、青光眼和葡萄膜炎等的一些报告。Madadi 等^[9]从一个公开的在线数据库中选择了 22 个神经眼科病例,包括专科医生常见的各种慢性和急性疾病。作者将每个病例中的文本作为新的提示符插入 ChatGPT 3.5 和 ChatGPT Plus 4.0,要求进行最可能的诊断。同时向两名神经眼科医生提供了准确的信息,并记录了他们的诊断,随后与 GPT 的反应进行比较。结果,在 22 例中,GPT 两种版本与两位医生分别正确 13 例(59%)、18 例(82%)、19 例(86%) 和 19 例(86%);最终诊断基本一致,但医生列出的诊断更系统具体。由此提示,在缺乏亚专业训练的神经眼科医生的临床环境中,使用 LLMs 的适用性值得进一步研究。

在角膜病诊断方面,Delsoz 等^[10]进行了类似的比较。作者从爱荷华大学可公开访问的在线数据库中随机选择 20 例角膜疾病,包括角膜感染、营养不良、变性和损伤。然后将每个病例描述的文本输入到 GPT 中,要求进行诊断;并与 3 位角膜病专家的诊断进行比较。结果,GPT-4.0 的诊断准确率为 85%(17 例);而 1 位专家全部准确,另 2 位 90%(18 例)准确。提示 GPT-4.0 在诊断角膜疾病方面的准确性明显提高,具有潜在的临床应用前景。同一研究组也对 11 例原发性青光眼(开角型、幼年型、正常眼压型和闭角型)和 4 例继发性青光眼进行了 GPT 与专家的诊断比较。结果,基于 GPT 的临时诊断正确率为 72.7%(8/11);3 位眼科住院医师正确率分别为 54.5%(6/11)、72.7%(8/11)、72.7%(8/11)。作者认为结合具体病例,GPT 对青光眼的诊断准确率与资深眼科住院医师相当或更好^[11]。

Rojas-Carabali 等^[12]使用 25 个符合葡萄膜炎命名指南的标准病例,对 GPT 与 5 名接受葡萄膜炎培训的眼科医生的诊断准确性和管理建议进行了评估。结果,医生的诊断率(60%-92%)高于 AI(60%)。考虑到完全和部分

准确的诊断,医生的成功率为76%–100%;AI达到72%。两者在48%的病例中诊断一致,在91.6%病例的治疗方案上表现一致。AI的整体表现显得落后。作者认为,GPT在葡萄膜炎诊断和管理方面具有潜力,但需要进一步提高其在诊断和推荐方面的精确度。

新加坡学者Lim等^[13]最近比较分析了3种LLMs(ChatGPT-3.5、ChatGPT-4.0和Google Bard)在近视护理中提供准确的响应方面的性能和表现。近视是一个常见的话题,患者和家长通常在网上寻找信息。作者整理出31个与近视护理相关的常见问题,将其分为发病机制、危险因素、临床表现、诊断、治疗和预防以及预后6个专题。每个问题都提交给LLMs,其回答由3位顾问级别的小儿眼科医生独立地按照3分的准确性量表(差、边缘、良好)进行评分。采用多数一致的方法来确定每个回答的最终评级。获得“良好”评级的回答在5分制上进一步评估其综合性。相反,评分“差”的回答会进一步提示自我纠正,然后重新评估准确性。结果,ChatGPT-4.0显示出更高的准确性,80.6%的回答被评为“良好”,而ChatGPT-3.5和Google Bard的回答分别为61.3%和54.8%。所有3个LLMs的平均综合得分都很高,也表现出自我纠正能力,纠正后的回答得到了改善。除了“治疗和预防”之外,LLMs在各个专题表现都是一致的,为近视相关的查询提供了准确和全面的回应。此研究提示,利用这些平台加强公众对预防近视的了解,将有助于缓解日益严重的近视大流行。然而,LLMs还处于起步阶段,为其提供量身定制的、特定领域的训练是当务之急,以确保准确的信息传播,避免提供错误的信息。

基于ChatGPT具备高水平的自然语言处理能力、对非结构化数据的理解能力、新内容生成能力和高通用性,有助于实现以患者为中心,其范围从简单的问答扩展到协助诊断护理。公众可以在类似聊天的界面上轻松地访问并与之互动。包括眼科知识的搜索和查询;对医疗文本数据、医疗记录和眼科图像的分析,以及分析结果的通知。患者还可以预测疾病的未来病程。这比简单的语言传递更能有效地向患者传达疾病的特征和预后。例如,患者可以了解视网膜手术后黄斑形态的变化并决定手术。通过新型的交互工具,使远程医疗中的沟通更加顺畅。在医生稀缺的地区,患者可以通过虚拟人轻松了解各种医疗信息,且不受语言的限制。其缺点包括提供最新的医学知识和研究成果困难,当前的模型在生成图像方面的性能仍然不足,提供不准确的信息可能导致严重后果。需要通过持续的数据采集和先进的算法开发来弥补这些局限性^[4]。

3 ChatGPT 辅助于科学论文写作^[1,3,14-15]

已有数十篇报告讨论了AI聊天机器人和ChatGPT用于科学论文写作的问题。有的文章还附有由其写成的短文例子来讨论在写作中的应用^[16]。总体上看,ChatGPT正在成为科学写作的有用工具,但目前对其能否成为“作者”均持否定态度^[2]。

ChatGPT可以用于组织材料、总结数据或信息、自动生成论文的完整草稿、撰写摘要和总结、为参考文献和标题提供建议、语言翻译、校对等任务。可以使写作工作来

得更快、更容易;在语言评论中使文本更具可读性。AI虽然不能产生新的想法,但它可以组织和发展研究人员的想法,创造一个初步的草案。

对于文献检索,ChatGPT和其它AI工具可以通过查找学术论文,总结其结论,快速了解某一特定主题的知识现状,并确定需要解决的潜在差距。但所提供的摘要可能相当笼统,并且不能批判性地分析研究之间的差异^[1]。

Buholayka等^[17]检验了ChatGPT撰写病例报告的准确性、完整性和可读性。作者要求ChatGPT仅根据5个病例介绍和最终诊断(不提供该疾病的其他信息)生成病例报告,并提供生成信息的参考文献。结果发现生成的病例报告在科学准确性方面存在严重缺陷,出现了错误的诊断和捏造的参考文献。作者认为AI目前尚不具备可靠的或独立的撰写科学病例报告的能力,它可能是一个有意义的补充,如对检查语法、提供同义词和替代短语很有用。由于ChatGPT并不是为科学出版而专门设计的,为了获得适合科学出版的LLMs,该算法应专门针对该任务进行训练,访问可靠的、最新的和多样化的来源数据库,以提供准确的引用。多位作者提出,自动生成的文本远不能取代人类专家的知识、创造力和批判性思维。对于生成的科学信息,应该加以规范,需要经过专家的仔细审查和修改。

科技期刊编辑在ChatGPT时代面临着新的挑战和机遇。在Lira等^[18]发表的报告中,首先用斜体字显示了一个作为科学编辑的挑战和优势的文本,除标题外,全部是在ChatGPT的帮助下编写的。通过这个文本,读者知道了ChatGPT是如何描述期刊编辑的。须知并不是每个科研人员都是或都能接触到优秀的医学作家。尤其在高影响力科学期刊上发表论文存在着激烈竞争。如果应用AI克服了语言障碍,有效地展示和讨论结果,撰写摘要,选择关键词,甚至设计吸引人的标题,这些都有助于产出适当的科学论文。因此,如果严格遵守所有的伦理原则和出版政策,ChatGPT可以是一个有用的写作工具,也能缩短数据分析和科学知识发表之间的时间。

由于诸多争议,该编辑团队决定不允许发表ChatGPT撰写的文章^[18]。科学期刊应该是传播创新的先锋,它们的编辑应该确保读者获得准确的信息来源,并保护版权。目前禁止发表源自AI的文章,并不意味着它将被永久禁止。在明确的定义编辑与创作、版权、道德偏差和潜在侵权责任之前,必须采取谨慎的态度。其中的一个担忧是,AI算法可能会创造出与现实不符的叙事和系列。这些生成技术允许用户从很少的输入信息中生成文本、图像和视频,制造欺骗性和虚假的内容。在信息的真实性至关重要的医学领域,这可能是危险的。传播假新闻也可能对公众健康造成严重后果。为此,必须开发和改进反AI的检查工具。就像已有的检查文本抄袭的策略一样,需要一定的机制来验证算法生成内容的真实性。期刊的“作者指南”部分也将随着科学界对这些新兴技术的谨慎使用而逐步更新。

参考文献

[1] Antaki F, Touma S, Milad D, et al. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmol Sci*, 2023, 3(4):100324.

- [2] Lee JY. Can an artificial intelligence chatbot be the author of a scholarly article? *J Educ Eval Health Prof*, 2023, 20:6.
- [3] Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*, 2023, 6:1169595.
- [4] Choi JY, Yoo TK. New era after ChatGPT in ophthalmology: advances from data-based decision support to patient-centered generative artificial intelligence. *Ann Transl Med*, 2023, 11(10):337.
- [5] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*, 2023, 11(6):887.
- [6] Teebagy S, Colwell L, Wood E, et al. Improved performance of ChatGPT-4 on the OKAP examination: a comparative study with ChatGPT-3.5. *J Acad Ophthalmol (2017)*, 2023, 15(2):e184-e187.
- [7] Cai LZ, Shaheen A, Jin A, et al. Performance of generative large language models on ophthalmology board - style questions. *Am J Ophthalmol*, 2023, 254:141-149.
- [8] Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol*, 2023, 141(6):589-597.
- [9] Madadi Y, Delsoz M, Lao PA, et al. ChatGPT assisting diagnosis of neuro-ophthalmology diseases based on case reports. *medRxiv*, 2023:2023.09.13.23295508.
- [10] Delsoz M, Madadi Y, Munir WM, et al. Performance of ChatGPT in diagnosis of corneal eye diseases. *medRxiv*, 2023:2023.08.25.23294635.
- [11] Delsoz M, Raja H, Madadi Y, et al. The use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports. *Ophthalmol Ther*, 2023, 12(6):3121-3132.
- [12] Rojas - Carabali W, Cifuentes - González C, Wei X, et al. Evaluating the diagnostic accuracy and management recommendations of ChatGPT in uveitis. *Ocul Immunol Inflamm*, 2023:1-6.
- [13] Lim ZW, Pushpanathan K, Yew SME, et al. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. *EBioMedicine*, 2023, 95:104770.
- [14] Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care*, 2023, 27(1):75.
- [15] Dashti M, Londono J, Ghasemi S, et al. How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing? *J Prosthet Dent*, 2023:S0022-S3913(23)00371-2.
- [16] Altmäe S, Sola-Leyva A, Salumets A. Artificial intelligence in scientific writing: a friend or a foe? *Reprod Biomed Online*, 2023, 47(1):3-9.
- [17] Buholayka M, Zouabi R, Tadinada A. The readiness of ChatGPT to write scientific case reports independently: a comparative evaluation between human and artificial intelligence. *Cureus*, 2023, 15(5):e39386.
- [18] Lira RPC, Rocha EM, Kara - Junior N, et al. Challenges and advantages of being a scientific journal editor in the era of ChatGPT. *Arq Bras Oftalmol*, 2023,86(3):5-7.