

---

# AI目前或不能协助公众做出更好日常健康决策

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/38242.html>

**本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！**

中新网北京2月10日电 (记者 孙自法)施普林格·自然旗下专业学术期刊《自然-医学》最新发表一篇医学研究论文指出，基于人工智能(AI)技术的大语言模型(LLM)，目前或许还不能协助公众做出更好的日常健康决策。研究人员认为，这类AI工具的未来设计需要更好地支持真实用户，才能安全用于向公众提供医学建议。

该论文介绍，全球医疗机构提议将大语言模型作为提升公众获取医疗信息的潜在工具，让个人在向医生求助前进行初步健康评估和疾病管理。但之前的研究显示，控制场景下在医师资格考试中得分很高的大语言模型，并不保证能有效完成真实世界的交互。

在本项研究中，英国牛津大学牛津互联网研究所研究团队与合作者一起，测试了大语言模型是否能协助公众精准辨别医疗病症，如普通感冒、贫血或胆结石，并选择一种行动方案，如呼叫救护车或联系全科医生。研究团队给1298名英国受试者每人指派了10种不同的医疗情景，并让他们随机使用三个大语言模型(GPT-4o、Llama3或Command R+)中的一个，或使用他们的常用资源(对照组)，如互联网搜索引擎。

研究结果显示，不用人类受试者进行测试时，大语言模型能准确完成这些情景，平均能在94.9%的情况下正确辨别疾病，在56.3%的情况下选择正确的行动方案。不过，当受试者使用相同的大语言模型时，相关病症的识别率低于34.5%，选择正确行动方案的情况低于44.2%，这一结果未超过对照组。研究团队人工检查了其中30种情况的人类-大语言模型交互并发现，受试者常向模型提供不完整或不准确的信息，并且大语言模型有时也会生成误导性或错误的信息。

论文作者总结认为，当前的大语言模型未准备好部署用于直接的患者医疗，因为将大语言模型与人类用户配对，会产生现有基准测试和模拟交互无法预测到的问题。(完)

( 原题：《国际研究：AI目前或不能协助公众做出更好日常健康决策》 )

作者：孙自法 来源：中新网

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

---

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://iikx.com)转发