

理论知识不输专家，AI 为何仍不能替代医生

□新华社记者 褚怡

头痛是不是脑梗的前兆？咳嗽要不要拍个片子？体检报告上的指标异常意味着什么？在去医院之前，越来越多人愿意先把健康问题抛给人工智能（AI）。输入症状、上传报告，几秒钟后，一份看似专业、条理清晰的分析便出现在屏幕上。对不少人来说，AI正在成为“24小时在线”的医学咨询窗口。但这真的意味着AI会看病吗？

标准化测试的“高分选手”

德国马尔堡大学等机构参与的团队近日发布的一项研究显示，在针对急性肾损伤的标准化知识测试中，多款AI大语言模型平均得分高于接受测试的医学专业人员。

研究选取了13个公众可使用的大语言模型，并将其与123名志愿者的表现进行比较。志愿者是2025年德国内科学会年会参会人员，其中包括内科执业医师。

测试采用同一套急性肾损伤知识问卷，包含两个模拟病例和15道选择题。结果显示，接受测试的大语言模型平均答对约90%的题目，多个模型达到满分；志愿者答题正确率约48.7%，且人类答题时间明显长于大语言模型。

研究人员认为，这表明在标准化测试情境中，大语言模型已经能够较可靠地调取并应用符合指南的相关医学知识，具有为临床工作快速提供事实性信息的潜力。

年初发表于“施普林格-自然出版集团”旗下《Cureus》医学科学杂志的一项研究也显示，一些大语言模型在标准化医师资格考试中的表现可比肩专业人员。研究人员选取美国全国医学考试委员会题库中的105道选择题，对GPT-4 Turbo模型进行测试，其正确率高达90.99%。

临床过程的“推理短板”

标准化测试中的高分，并不意味着AI具备真实临床诊疗所需的判断力。美国麻省总医院布里格姆医疗中心机构研究人员近日在《美国医学会杂志·网络开放》上发表研究说，大语言模型在临床推理方面的能力仍然不足，在相关数据收集齐全情况下，这些模型通常能给出较准确的最终诊断，但在病例早期、信息仍然匮乏时，它们往往不具备鉴别诊断的能力。

为还原真实临床过程，研究人员采取分步输入方式，评估了21个大语言模型对29个标准化临床案例的诊断情况。研究人员先输入患者年龄、性别和症状等基础信息，再补充体格检查和实验室结果。模型每个阶段表现由医学专业学生进行评估，并据此计算得分。

结果显示，所有受测试模型在超过80%的情景下都未能在病情尚未明确、信息仍不完整时给出恰当的鉴别诊断，即未能准确判断最可能的病因或排除严重疾病，并据此为下一步检查和排查提供可靠方向。

“鉴别诊断是临床推理的核心，也是目前AI尚无法复制的‘医学艺术’的基础。”研究论文通讯作者马克·苏奇说，现阶段AI在临床医学中的潜力，在于其能够辅助而非取代医生的推理过程。

哈佛大学医学院和斯坦福大学等机构研究人员年初在《自然-医学》杂志发表的一项研究也显示，大语言模型在标准化医学考试中表现优异，但在基于医患对话记录进行诊断时明显吃力。

研究论文通讯作者、哈佛大学医学院副教授普拉纳夫·拉杰普尔卡尔说，医疗对话具有动态性，需要在恰当时机提出恰当问题，将零散信息整合起来，并根据症状推理，这种独特挑战远非答题可比。“当场景从标准化测试转向自然对话时，即使是最先进的AI模型，诊断准确性也会显著下降。”

医生主导下的人机协作

既然AI还无法独立诊疗，它应当以何种身份进入医疗实践？在18日开幕的2026年德国内科学会年会上，德国杜伊斯堡-埃森大学人工智能医学研究所所长延斯·克莱西克说，随着AI的发展，医生与计算机的协作正在加强。数字系统不再只是提供支持，而是通过病例记录、协调流程等方式主动地介入医疗过程，“这将从根本上改变医疗服务”。他认为，要让AI真正发挥潜力，前提是高质量、结构化且可互操作的数据，以及足够可靠的技术基础设施。

但医生的主体责任并未因此削弱。克莱西克强调，人的因素仍至关重要，仍需要由具备医学专业能力、能够理解并合理使用AI技术的医生来推动和把关。

在医生主导下由人机协作开展医疗服务的效果已得到研究支持。斯坦福大学等机构研究人员近期在《自然合作期刊-数字医学》杂志上发表的一项随机对照试验显示，在经过设计的人机协作流程中，医生诊断准确性可由传统资源条件下的75%提高至80%以上。

专家强调，推动AI技术融入临床诊疗的同时须警惕伴随风险。美国密苏里大学医学院副教授法里斯·阿拉赫达卜认为，经验丰富的临床医生通常能够识别AI提供的错误建议，而医学学生和年轻医生往往缺乏相应的判断力，难以识别那些细微却可能致命的错误。

阿拉赫达卜指出，更隐蔽的风险在于，过度使用AI可能削弱医生的批判性思维。医生可能会在不知不觉中把推理过程“外包”给AI。模型给出的答案越流畅、越完整、越像是正确的，使用者就越可能放弃独立检索信息、批判性思考及知识整合。久而久之，那些本应持续训练的能力将逐渐退化。

（新华社柏林4月20日电）



►4月19日，天津医科大学肿瘤医院医护人员使用达芬奇手术机器人给病人做手术。

新华社记者李然 摄



►2025年9月20日，在第22届中国—东盟博览会现场，广西医科大学越南留学生范忠孝（右）向媒体介绍“泌语医谈”AI数字医生。

新华社发

▼4月16日，在哈尔滨思哲睿智能医疗设备股份有限公司生产车间，工作人员在对设备进行调试。

新华社记者张涛 摄

