

当 ChatGPT 遇见医疗……

丁思月 / 整理

最近,美国开放人工智能研究中心(OpenAI)研发的语言模型 ChatGPT 火爆全网。据悉,ChatGPT 仅上线 5 天注册用户就突破了 100 万,两个月后的月活用户突破 1 亿,成为史上用户增长速度最快的消费级应用程序。

迭代多次,并非突然出现

ChatGPT 是一个以对话方式进行交互的模型,其中 Generative Pre-trained Transformer(GPT)是一种基于互联网可用数据训练的文本生成的深度学习模型。它用于问答、文本摘要生成、机器翻译、分类、代码生成和对话 AI。

截至目前,GPT 的整个发展史历经 3 次迭代。

2018 年,GPT-1 诞生,这一年也是 NLP(自然语言处理)的预训练模型元年。性能方面,GPT-1 有一定的泛化能力,能够用于和监督任务无关的 NLP 任务。其常用任务包括自然语言推理,即判断两个句子的关系(包含、矛盾、中立);问答与常识推理,即输入文章及若干答案,输出答案的准确率;语义相似度识别,即判断两个句子语义是否相关;分类,即判断输入文本属于指定的哪个类别。

即便 GPT-1 在未经调试的任务上有一些效果,但其泛化能力远低于经过微调的有监督任务,因此 GPT-1 只能算得上一个不错的语言理解工具而非对话式 AI。

2019 年,GPT-2 问世。不过,GPT-2 并没有对原有的网络进行过多的结构创新与设计,只使用了更多的网络参数与更大的数据集,最大模型共计 48 层,参数量达 15 亿,学习目标则使用无监督预训练模型做有监督任务。在性能上,除了理解能力,GPT-2 在生成方面第一次表现出了强大的天赋——阅读摘要、聊天、续写、编故事,甚至生成假新闻、钓鱼邮件或在网上进行角色扮演。

2020 年,GPT-3 出现了,作为一个无监督模型(现在经常被称为自监督模型),几乎可以完成自然语言处理的绝大部分任务,例如面向问题的搜索、阅读理解、语义推断、机器翻译、文章

生成和自动问答等。

该模型在诸多任务上表现卓越,例如在法语-英语和德语-英语机器翻译任务上达到当前最佳水平,自动产生的文章几乎让人无法辨别出自人还是机器(仅 52% 的正确率,与随机猜测相当)。更令人惊讶的是,在两位数的加减运算任务上达到几乎 100% 的正确率,甚至还可以依据任务描述自动生成代码。

2022 年底,衍生自 GPT3.5 的 ChatGPT 问世,并于 2023 年初火爆互联网。

可提高效率,但无法做出决策

从目前看,ChatGPT 带来的影响是巨大而深远的,它不只是一个现象级的产品,其背后的 GPT3.5 更是一种颠覆性的技术,在各行各业都有落地应用的巨大前景。

而医学被认为是人工智能(AI)应用中有可能率先实现商业化的细分领域。在“政产学研用”的多方努力下,全球智能时代加速到来,而医疗行业也正加速进入数智化的爆发期。数据显示,全球医疗保健 AI 市场规模预计从 2023 年的 146 亿美元,增长到 2028 年的 1027 亿美元,年复合增长率达 47.6%。

那么,ChatGPT 的出现将给医疗行业带来哪些影响,它能代替医生执业吗? AI 与医疗相结合将面临何种挑战?

上海市第十人民医院老年医学科主任彭沪认为,医疗 AI 的发展目标应该是辅助医生,而非代替。AI 可以帮助医生快速完成一些琐碎的重复性工作,提高效率、改善工作质量、提升治疗水平、减轻医务人员工作负担。但最终的判断和决策仍需在医生监督下进行。“尤其在强医疗场景中,AI 的应用还涉及伦理、政策、论证等多方面因素。”

ChatGPT“自己”也这样认为。当它被询问是否可以取代医生时,它回答说,“我不能取代医生,我只是一个辅助工具。我可以帮助医生收集患者的健康信息,提供临床诊断支持,协助医生进行病例管理,提供专业的医疗咨询服务,以及协助医生进行病例分析



图片来源:
摄图网

和研究。”

中国科学院院士葛均波也曾发文表示,AI 在医学应用方面存在诸多重大挑战,想要完全取代医生,目前技术不可能实现。他指出,ChatGPT 是 AI 发展到一定阶段的代表性产品,影响 AI 发展的三个要素是算法、算力及数据库。

在算力方面,他相信随着芯片等硬件技术提高,问题不大。“但在算法及数据库方面,医学是有特殊性的。”首先,医学的参数变量主观性较强,有时候不能精确定量评价,如疼痛、疲劳、瘙痒、困倦等,个体差异很大。其次,医学参数变异度及相互影响很大,比如医生经常使用的一个参数——6 分钟步行试验,即让患者在 6 分钟内拼命走路,然后测量其最大走动距离,患者在不同身体状态、不同医生陪同监督下都会出现差别较大的结果。

葛均波认为,医学是一门模糊科学、概率科学,很难精确预判,只能大概率的判断。在实践中,再简单的手术,都不可能百分百没有并发症;再高明的专家,在临床实践中,也不可能百分百准确诊断。医学变量的模糊性及变异性,以及参数之间可能会发生交互作用,对 AI 在医学中的应用提出了更高要求。

医学数据库的建立,同样存在巨大难度。医学变量和常规变量不一样,有时候难以准确收集,况且一个人所涵盖的生物信息参数非常大,再加上受制于个人隐私、医学伦理,医院之间信息壁垒,人力投入不够等因素,要建立一个完整的海量生物信息数据库是一个非常大的挑战。

“一本正经胡说八道”现隐忧

值得注意的是,ChatGPT 很多时候固然能给出智能且知识性超强的回答,但也不断有人找出它犯的一些低级错

误,认为它在一些问题是“一本正经地胡说八道”。

有媒体要求 ChatGPT 对当前在中国上市的丁型肝炎 RNA 试剂盒做调研,并直接以表格形式列出试剂盒的产品名称及相关信息。ChatGPT 给出了一份看似满意的答卷,然而经过人工搜索,发现所列出的产品信息均属杜撰,继续对其提问也无法获取公司相关信息。同时,在多次“审问”后,ChatGPT 给出了生成此表格的模板,例如公司名称通过填写 × × 生物生成。

从本次测试情况看,ChatGPT 有很强的总结能力,但是,引用 ChatGPT 做案头研究需要审慎判断答案的真实性和时效性。

此外,ChatGPT 会给出一些看似有意义实则空洞无物的回答。

当 ChatGPT 被问及能否充当一个经验丰富的医药代表,作为前辈指导新人工作。它给出的回答(部分)是“了解医生的治疗需求并说明产品如何满足他们的需求;保持与医生的联系,加强与他们的关系,以便以后与他们进行更多的合作”。

医药公司的人表示,该回答具有相对完备的产品介绍框架,对医药代表新人可能会起到一些指导作用,但是内容稍显空泛,而新人代表可能会期待其提供更有价值的销售经验。

中国工程院院士、国家基层糖尿病防治管理办公室主任贾伟平则指出,医生和患者面对面不止于信息交流,其间还有心理层面的沟通,因此可以肯定的是,“AI 不能够代替人类,不能够代替医生”。

山西省人民医院妇产科主任医师王素琴直言,“ChatGPT 不可能取代医生。医学是人文科学,医生除了需要很强的技术背景还需要人文关怀,医生对患者的关爱是有温度的,AI 做不到有温度的关怀。”