

# ChatGPT 能代替医生看病吗？

## 考卷设置十大科目 涵盖基础题和专业实践题

由于 ChatGPT 是依据 2021 年以前的数据训练的，所以题目选自 2020 年临床执业医师资格考试真题库，全部为单选题。

吕海龙介绍，临床执业医师考试内容笔试考试主要包含基础综合、专业综合和实践综合，我们的题目简化为两类，一类是基础题，大都是可背诵的医学知识；另一类是专业实践题，需要进行逻辑推理。

考卷涵盖了十大科目，包括生理学、医学微生物学、病理学、药理学、卫生法规、内科、外科、妇产科、儿科和心理学-精神科，每个单元 10 题，合计 100 题，选择方法尽可能随机。

## 掌握基本诊断推理 但不擅长修正和推测

在我们的猜想中，ChatGPT 应该更擅长可背诵的医学知识，逻辑推理能力则较弱，但实践结果并非如此。

吕海龙发现，对于那种一看就是考题的题干，ChatGPT 有时候会直接给出一个没有解释的错误选项，不排除它从中文网站复制了错误的答案；而那种条件充足、逻辑清晰的推理题，它通常情况都能答对。

不过，当信息不全或者症状不典型时，它的诊断会出现错误，不善于根据实际情况修正诊断，比如内科病例题，它不善于在信息不全的情况下做推理，例如由几个模糊的主诉/症状来寻找乏力的可能原因，所以推测 ChatGPT 在面对患者时如果得不到足够多的信息它可能会不知所措，完全无法给出下一步的诊疗建议，它不会“猜”。

ChatGPT 对精神病学的知识比较了解，能根据症状诊断抑郁症、焦虑症等，但是对于精神药物知识掌握不佳，例如会推荐精神分裂症的患者服用抗抑郁药，来治疗主要症状。

“如果有两种同类型的药可以选择时（例如该用利培酮还是氯丙嗪来治疗精神分裂症），它不会帮你做选择，而是严谨地提议根据医生的建议选择，可能自愧不如人类。而关于这两种药的选择，现在已有超多的研究证据证明利培酮的安全性优于氯丙嗪。ChatGPT 给其他疾病选择药物时也常常选错，它的药理学考试也没有及格。”吕海龙说。

ChatGPT 喜欢做妇产科医生，它会给出很多理由来支持它的选择，虽然有时候选择是错的。另外，ChatGPT 对于中国的医疗相关法律法规不够了解，这可能会给它在实践中招来一些麻烦。



网络图片

## 考试成绩不及格 儿科和心理科最好，内科最差

记者在提问过程中发现，针对不同题目，ChatGPT 的回答速度不同。有些题是秒回，有些题需要更长时间“思考”，平均每题要花半分钟左右。

ChatGPT 对中文的理解完全没问题，有时会直接给出答案，有时给出一串相关知识，或者推理分析的过程。

提问进行到 1 小时时，系统提示“1 小时内的问题太多了，请稍后再试”。看来，人工智能也是会累的嘛！（当然，也可能是这段时间访问量实在太太大，系统无法承载。）

最后经过统计，ChatGPT 的正确率分别为：生理学 40%、医学微生物学 70%、病理学 70%、药理学 50%、卫生法规 50%、内科 30%、外科 60%、妇产科 60%、儿科 70%、心理学-精神科 70%，总分 57 分，差 3 分达到及格。

## 期待未来辅助医生工作 引导前沿医学研究

对 ChatGPT 交出的答卷，吕海龙表示既有惊喜，也有很大不足。就现阶段技术水平来说，他希望医疗方向的 AI 产品能变成医学知识库，起码在识记类的题目上不能出错，必须 100% 正确，给出答案时最好标记出处，是引自某个教科书，还是引自网络信息。

而对于医疗方向 AI 产品的未来，胡少华教授则有着更为积极的展望：“医生把自己的知识经验传给它，医生不在病房时，病人可以直接去问这个机器，输出声音变成人的声音，并且最好它还能向患者提问。同时，AI 在医学研究领域也可以大有作为，做研究的前沿导向者，随时给出新的医学诊断技术、治疗方法的发现。”

胡少华说，有关人工智能的对话，还在不断地完善当中。如今 AI 在识别、记忆、计算、逻辑思维、推理判断等方面已经可以达到甚至超越人脑，但还没有触及到智力的起点，即认知。只要 AI 还无法模拟人类的认知，它就不会全面超越人类，也就无法取代包括医生在内的诸多岗位，但在针对患者提供个性化的诊疗上，它却能够扮演重要角色——成为医生们最得力的智能助手。

（据《钱江晚报》）

最近，ChatGPT 的横空出世，也引发医疗界人士广泛思考——人工智能将对医疗领域带来多大的冲击，未来医生会失业吗？

浙大儿院 NICU 副主任马晓路发朋友圈称：以前吧，觉得会被人工智能取代的，都是那些单调机械重复的体力活。但看到人工智能写的小作文比你写好，画的画比你画得好，回答问题比“砖家”靠谱得多，确实会令人恐慌。

医生们的危机感不无道理。有消息称，OpenAI 公司的程序员已经开始利用高版本 ChatGPT 对日常疾病进行问诊，并根据 ChatGPT 给出的处方进行康复治疗，并取得了良好的反馈。还有一篇预印本的论文表示，ChatGPT 在美国医师执照考试（USMLE）中达到了平均水平以上，准确率达到了 60% 左右。

记者不禁好奇，如果让 ChatGPT 参加中国的执业医师考试，它能考多少分呢？

浙大一院精神卫生科主任胡少华教授团队正在参与一项人工智能相关的医学研究，团队希望未来借助 AI 监测眼球活动、皮肤电势反应等客观数据，辅助临床医生诊断。

于是，记者和胡少华教授团队成员吕海龙一起，设计了一套百分制的执业医师考试卷子，ChatGPT 的最终成绩是 57 分，其中基础得分 56 分，临床得分 58 分。