生成式人工智能GPT-4驱动的中药处方生成研究Δ

陈祺焘*，倪璟雯，徐　君，高晓涵，夏丽珍#（三明市中西医结合医院药学部，福建 三明　365001）

中图分类号　R95　　**文献标志码**　A　　**文章编号**　1001-0408（2023）23-2825-04
DOI　10.6039/j.issn.1001-0408.2023.23.02

摘　要　目的　评估生成式人工智能（AIGC）中的GPT-4模型生成中药处方的安全性、适宜性，为AIGC赋能中医药行业提供研究思路。**方法**　将2020年版《中国药典》和第5版《中药学》作为语料，由GPT-4及基于GPT-4开发的实时联网模型（简称“联网模型”）对其进行深度学习。人工抽取近几年中医药类专家共识收录的临床案例，由GPT-4模型和联网模型根据诊断重新生成处方。由中医药学专家对GPT-4生成处方、联网模型生成处方以及专家共识处方进行盲评打分，同时通过图灵测试来评估GPT-4模型和联网模型是否具有与人类智能相当的能力。**结果**　GPT-4模型生成的中药处方的平均分与人工处方比较，差异无统计学意义（$P>0.05$）；联网模型生成处方的平均分与GPT-4模型生成的中药处方比较，差异无统计学意义（$P>0.05$）。模型生成处方在图灵测试中被误判为人工处方的占比达51.11%。**结论**　GPT-4模型生成的中药处方在安全性、适宜性方面已经具备一定的水平，且GPT-4模型通过了所设置的图灵测试；在诊疗过程中引入AIGC可能为临床中药的合理使用提供技术支撑。

关键词　GPT-4；中药处方；生成式人工智能

Generation of traditional Chinese medicine prescription driven by generative artificial intelligence GPT-4

CHEN Qitao, NI Jingwen, XU Jun, GAO Xiaohan, XIA Lizhen (Dept. of Pharmacy, Sanming Integrated Medicine Hospital, Fujian Sanming 365001, China)

ABSTRACT　OBJECTIVE To evaluate the safety and suitability of traditional Chinese medicine prescriptions generated by generative artificial intelligence (AIGC), and to provide research ideas for empowering the traditional Chinese medicine industry with AIGC. **METHODS** Using the 2020 edition of *Chinese Pharmacopoeia* and the 5th edition of *Traditional Chinese Medicine* as corpus, GPT-4 and the real-time networking model developed based on GPT-4 (referred to as the "networking model") were used for deep learning. The clinical cases included in the consensus of traditional Chinese medicine experts in recent years were extracted manually to regenerate prescriptions based on diagnosis using the GPT-4 model and networking model; traditional Chinese medicine experts conducted blind evaluation and scoring of GPT-4 generated prescriptions, networking model generated prescriptions, and expert consensus prescriptions. At the same time, Turing testing was used to evaluate whether the GPT-4 model and networking model had the same ability as human intelligence. **RESULTS** The average score of traditional Chinese medicine prescriptions generated by the GPT-4 model showed no statistically significant difference compared to manual prescriptions ($P>0.05$), while the average score of prescriptions generated by the networking model showed no statistically significant difference compared to traditional Chinese medicine prescriptions generated by the GPT-4 model ($P>0.05$). The proportion of model-generated prescriptions mistakenly judged as manual prescriptions in the Turing test was 51.11%. **CONCLUSIONS** The traditional Chinese medicine prescriptions generated by the GPT-4 model have reached a certain level of safety and suitability, and the GPT-4 model has passed the Turing test. The introduction of AIGC in the diagnosis and treatment process may provide technical support for the rational use of clinical traditional Chinese medicine.

KEYWORDS　GPT-4; traditional Chinese medicine prescription; generative artificial intelligence

中医药作为中国传统文化的重要组成部分，在保障人民健康方面发挥着不可替代的作用。随着“健康中国”行动的深入推进，人民群众对健康美好生活需求的提升，中医药被赋予了更高的期望[1]。一方面，中医药在疾病预防、治疗和康复方面独具优势：中医药强调个体化辨证论治，可根据不同病因、病机和个体差异，制定有针对性的治疗方案；中医药还重视整体调理，通过调节人体阴阳、气血等方式来提高机体免疫力，从而达到预防疾病、延缓衰老、提高生活质量的目的。另一方面，中医药产业也面临着巨大的发展机遇：当前，随着人民群众对健康需求的提升，中医药市场需求不断增长；同时，中医药在科技创新、标准化建设、产业升级等方面也取得了重要进展，逐步实现了从传统到现代的转型升级，

Δ**基金项目**　国家中医药管理局2022年宋纬文全国名老中医药专家传承工作室建设项目（No.国中医药人教函〔2022〕75号）

＊**第一作者**　药师。研究方向：药物制剂。电话：0598-8033609。E-mail：cqt1945@163.com

\# **通信作者**　主任中药师。研究方向：中药学。电话：0598-8033609。E-mail：869220642@qq.com

各种中药饮片被广泛应用于临床的诊疗过程中，受到了广大患者与医生的认可[2]。

西医学习中医（简称“西学中”）是一项为了补充中医从业人员，通过为西医医生提供中医理论基础培训而使西医医生能够开展中医药诊疗的培训制度，发展至今已有60余年的历史。但由于西医医师对中医知识学习不足，使得中药或中医护理技术滥用的情况非常普遍。有学者通过对2018年上海市某社区卫生服务中心开具的中药门诊处方进行分析发现，不合理处方超过总处方的20%[3]。GPT-4（generative pre-trained transformer 4）是一种基于Transformer架构的生成式人工智能（AI generated content，AIGC），其强大的逻辑自洽和数据分析能力使其一面世立刻成为各个领域关注的焦点。本研究准备借助GPT-4赋能，拟通过抽取国内公开的常见中药应用案例，采用对话问答的方式由GPT-4进行中药处方生成测试，再由中医药学领域的专家对生成结果进行评分，同时进行图灵测试，评估GPT-4模型是否具有与人类智能相当的能力，为AIGC赋能中医药行业提供一种研究思路。

1 案例来源

由研究者随机抽取《2023年春季成人流行性感冒中医药防治专家共识》[4]、《咳嗽中医诊疗专家共识意见（2021）》[5]、《毒蛇咬伤中医诊疗方案专家共识（2016版）》[6]收录的案例，并排除临床表现的中医诊断证型与用药存在较大争议的案例（如同一临床表现被不同专家判断为虚寒、虚热两个相反证型的案例）。

2 研究方法

2.1 研究框架

本研究主要分为3个部分：首先，使用2020年版《中国药典》和中国中医药出版社出版的第5版《中药学》为语料对GPT-4模型进行训练；然后，使用GPT-4模型，根据“1.1”项下抽取案例的临床表现进行中药处方生成；最后，以人工盲评的方式对模型生成的中药处方进行评价。

需要注意的是，考虑到GPT-4原生训练数据来自于2021年及之前的互联网数据，为进一步研究GPT-4模型的处方生成能力，本次研究将把基于GPT-4开发的实时联网模型（简称“联网模型”）一同纳入处方生成研究，并将处方生成结果一同参与盲评，最后将GPT-4模型和联网模型生成结果的盲评得分进行对比。

2.2 模型训练

将2020年版《中国药典》和第5版《中药学》的文字内容发送至GPT-4模型以及联网模型，对模型接受程度进行确认并针对处方的格式进行命令限制，以提升模型对中药饮片和处方的理解分析能力。

2.3 处方生成

考虑到评价涉及用药安全性，本研究采用的专家共识处方应包含每种中药饮片的具体剂量，病种涉及咳嗽、蛇伤等。

随机抽取30个专家共识内的用药案例（均为口服水煎剂），由GPT-4模型以及联网模型根据案例的临床表现生成处方。

2.4 处方评价及图灵测试

GPT-4模型、联网模型生成的中药处方和专家共识处方由三明市中西医结合医院3名具有副高级及以上职称的中医师/中药师根据处方安全性、适宜性进行盲评（每张处方的最终得分为3名专家的平均分），并判断每张处方是否为人工生成[7]。考虑到相应法规及医学伦理方面的要求，GPT-4模型及联网模型生成的中药处方未使用临床试验进行测评。本研究主要根据用药安全性、适宜性（即根据“十八反十九畏”、超剂量使用、“先煎后下”等标注、对症情况）由具有副高级及以上职称的专家对生成处方和共识处方进行评分（评分标准见表1）。为保证评分的准确性，在评分时，专家不知道处方是由模型生成的或是人工生成的（即盲评）。

表1 专家评分标准

评价标准	评价说明（分值共5分，实行扣分制）
是否出现“十八反十九畏”	出现“十八反十九畏”扣2分
是否出现超剂量使用	出现超剂量使用扣1分
是否未标明“先煎后下”等	出现未标明“先煎后下”等扣1分
处方是否对症	处方不对症扣1分

专家对处方进行评分时，同时需要判断该处方是否为人工生成的处方（即图灵测试）。如果超过一半的模型生成处方被错误判断或模型生成处方被错误判断的比例高于人工生成处方，则说明自动生成的处方对人类有足够的迷惑性。该测试用于评估GPT-4模型是否具有与人类智能相当的能力。

2.5 统计学方法

采用SPSS 26.0软件对数据进行统计分析[8]。对各生成处方盲评所得分数进行正态性检验，并分别将专家共识处方和GPT-4生成处方、联网模型生成处方盲评分数进行配对t检验分析。检验水准$\alpha=0.05$。

3 结果

3.1 数据正态性检验分析

本次研究共纳入30个临床案例，各类型处方通过盲评所得分数的正态性检验见表2。

表2 数据正态性检验分析

处方类型	样本量	盲评平均分	标准差	偏度	峰度	Shapiro-Wilk检验	
						统计量（W值）	P
专家共识处方	30	3.756	0.747	0.305	−0.188	0.961	0.327
GPT-4生成处方	30	3.622	0.508	−0.254	−0.343	0.940	0.089
联网模型生成处方	30	3.501	0.478	−0.079	0.886	0.940	0.090

由表2可见，Shapiro-Wilk检验结果显示，专家共识处方、GPT-4生成处方、联网模型生成处方具备正态性分布特质（$P>0.05$）。

3.2 专家共识处方和GPT-4生成处方比较

本次研究共使用30个临床案例，专家共识处方的平均分为3.76分，略高于GPT-4生成处方的3.62分。通过配对t检验分析可知，专家共识处方和GPT-4生成处方的平均分比较，差异无统计学意义（$P>0.05$）。结果见表3。

表3 专家共识处方和GPT-4生成处方的配对t检验分析结果

处方类型	平均分	标准差	平均分差值	t	P
专家共识处方	3.76	0.75	0.14	0.762	0.452
GPT-4生成处方	3.62	0.51			

3.3 GPT-4生成处方和联网模型生成处方比较

本次研究共使用30个临床案例，GPT-4生成处方的平均分为3.62分，略高于联网模型生成处方的3.50分。通过配对t检验分析可得，GPT-4生成处方和联网模型生成处方的平均分比较，差异无统计学意义（$P>0.05$）。结果见表4。

表4 GPT-4生成处方和联网模型生成处方的配对t检验分析结果

处方类型	平均分	标准差	平均分差值	t	P
GPT-4生成处方	3.62	0.51	0.12	1.040	0.307
联网模型生成处方	3.50	0.48			

3.4 图灵测试结果

本研究共纳入30个案例，根据上述案例形成了90个处方（分为专家共识处方、GPT-4生成处方、联网模型生成处方各30个），由3位专家判断这90个处方是否为人工生成的处方。图灵测试结果（表5）显示，270个处方中，共有138个模型生成处方被错误判断，占比为51.11%（>50%），其中GPT-4生成处方被错误判断的占比达30.37%（>30%），结合上文“3.2”项下结果发现，GPT-4模型生成处方已经具备一定的专业性。

表5 图灵测试结果

项目	识别为人工处方次数	识别为人工生成处方占比/%
专家共识处方	70	25.93
GPT-4生成处方	82	30.37
联网模型生成处方	56	20.74
模型生成处方合计	138	51.11

4 讨论

4.1 如何提高GPT-4的学习、完善能力

在使用GPT-4和联网模型进行处方生成的过程中，GPT-4体现出强烈的原则性和强大的学习能力，主要体现在：（1）要求提供蛇伤处方时，虽然研究者已经发出限制命令，要求GPT-4模型仅提供中药名称和单次剂量，但GPT-4模型仍会发出警告，要求及时就医；（2）在放开限制命令后，GPT-4模型不仅会对处方中每个组分进行方解，而且在生成处方的过程中，会逐渐对先煎后下等特殊处理方式进行标注，其生成处方的评分结果已经非常接近专家共识处方。与此同时，已有深度学习模型介入影像学诊断的相关研究，其模型评分结果已经超过了低年资影像科医生的评分结果[9]。随着AIGC介入医学领域的深度和广度不断扩大，要进一步提高生成式模型医学水平，需要开展以下几项工作：（1）海量病例语料训练。本次测试的病例主要为咳嗽等常规案例，未考虑到患者性别、年龄、人种等因素，对于临床诊断等方面的内容有待进一步探索。（2）算法优化。临床医生年龄分布较广，对于药品名称、临床症状的口语化描述问题较为严重。这一方面需加强规范医生的病历书写，另一方面也需对AIGC算法进行优化，对口语化内容进行识别。

4.2 联网模型质量问题

本次处方评分中，联网模型生成处方的平均分最低。通过对联网模型数据来源进行分析发现，其中混杂了大量非医学类专业的网站数据，对联网模型的处方生成造成了极大的干扰。对互联网医学类语料进行规范标识和整理，一方面能够提高模型的训练质量，另一方面还能够降低群众通过互联网就医的学习成本。对中文医学语料进行标注整理是未来生成式医学模型发展的必经之路，同时此项工作对互联网问诊、分级诊疗、医学科普有着极大推动作用。

4.3 临床方向AIGC研究的进一步探索

本研究考虑到患者权益问题，并未对生成处方进行临床试验，同时为了保护患者隐私，本研究采用的是公开的专家共识所收录的临床案例。如果要进一步发展药事管理方向的AIGC，有以下两点尚待解决：（1）诊疗权责以及医学伦理问题[10]。目前未有相关的法律法规对药事管理方向AIGC所生成的处方进行明确的权责划分。生成式模型算法的程序缺陷，医生、药师对生成处方的审核失误都会对患者造成不可逆的伤害。如何保障患者（特别是妊娠、低龄、残疾等弱势患者群体）权益，患者权益的保障方应该是模型公司还是医院（即责任归属）等，均需要有明确的法律法规进行规范；对于紧急情况下AIGC介入医学研究的程度和范围也需要伦理方面的专家进行研探。（2）公民隐私及遗传学信息保护问题。随着生成式模型的发展，AIGC进入医院药事管理领域是可预见的，但因算力需求等原因，该技术目前主要基于互联网使用，尚无与GPT-4具有同等能力的本地化AIGC。欧盟在使用GPT模型的过程中已发现隐私及机密泄露等问题，遂将人工智能纳入安全工作研究[11]。医疗信息涉及公民隐私和我国遗传学信息保护，关系到国家安全，AIGC本地化部署是解决以上问题的唯一途径。另外，与此关联的还有本地化部署的设备费用、运营维护、医疗机构系统适配对接等问题。

综上，本文使用的GPT-4和基于GPT-4开发的实时联网模型较好地学习了《中国药典》和《中药学》的内容，其生成的处方在中药处方评分环节取得了较为出色的成绩，与专家共识处方较为接近。临床科室医生在使用中药时，要求医生对各种中药药性有较高的理解能力，目前通过西学中培训的医生对相关知识的储备仍有不足，中药处方质量有待提高。本GPT-4模型在生成处方的过程中自动为处方标注方解及参考来源，减少了医生的学习成本和沟通成本，为临床中药的合理使用提供了技术支撑。实现医院中医药方向生成式模型的落地，还需要做好以下几点：(1)在相关法律法规的框架下收集大量的中医处方和相关病症数据，包括但不限于病历、医生处方和临床诊断。(2)收集到的数据需要进行清洗和标准化处理，以使其能适用于模型训练。(3)部署模型后，需要不断地监督模型的运行，并根据反馈进行调整；若发现模型生成处方有误，需及时对模型进行调整或重新训练。

参考文献

[1] 杨洪军，李耿. 推动中药产业迈向高质量发展[J]. 中国生物工程杂志，2022，42(5)：16-17.
YANG H J，LI G. Promote the development of Chinese medicine industry towards high quality[J]. China Biotechnol，2022，42(5)：16-17.

[2] 彭善祥. 中药处方点评在中药饮片合理使用中的干预价值分析[J]. 北方药学，2022，19(3)：88-90.
PENG S X. Analysis of intervention value of traditional Chinese medicine prescription comment on rational use of traditional Chinese medicine decoction pieces[J]. J N Pharm，2022，19(3)：88-90.

[3] 胡宗仁，邓奕辉，沈承玲，等. “西学中”系统化培训的意义、特点及知识体系[J]. 中国中西医结合杂志，2022，42(12)：1424-1427.
HU Z R，DENG Y H，SHEN C L，et al. Significance，characteristics and knowledge system of systematic training of western medicine doctors learning from Chinese medicine[J]. Chin J Integr Tradit West Med，2022，42(12)：1424-1427.

[4] 方邦江，张洪春，张忠德，等. 2023年春季成人流行性感冒中医药防治专家共识[J]. 陕西中医药大学学报，2023，46(4)：1-5.
FANG B J，ZHANG H C，ZHANG Z D，et al. Expert consensus on traditional Chinese medicine prevention and treatment of adult influenza in spring 2023 [J]. J Shaanxi Univ Tradit Chin Med，2023，46(4)：1-5.

[5] 孙增涛，师艺航，李小娟. 咳嗽中医诊疗专家共识意见：2021[J]. 中医杂志，2021，62(16)：1465-1472.
SUN Z T，SHI Y H，LI X J.Experts consensus on the diagnosis and treatment of cough in traditional Chinese medicine：2021[J]. J Tradit Chin Med，2021，62(16)：1465-1472.

[6] 王万春，严张仁. 毒蛇咬伤中医诊疗方案专家共识：2016版[J]. 中医杂志，2017，58(4)：357-360.
WANG W C，YAN Z R. Expert consensus on TCM diagnosis and treatment scheme for snakebite：2016 edition[J]. J Tradit Chin Med，2017，58(4)：357-360.

[7] 刘江峰，刘雏菲，齐月，等. AIGC助力数字人文研究的实践探索：SikuGPT驱动的古诗词生成研究[J]. 情报理论与实践，2023，46(5)：23-31.
LIU J F，LIU C F，QI Y，et al. A practical exploration of AIGC-powered digital humanities research：a SikuGPT driven research of ancient poetry generation[J]. Inf Stud Theory Appl，2023，46(5)：23-31.

[8] 刘堃. SPSS统计分析在医学科研中的应用[M]. 北京：人民卫生出版社，2012：1-244.
LIU K. Application of SPSS statistical analysis in medical scientific research[M]. Beijing：People’s Medical Publishing House，2012：1-244.

[9] 殷保才. 基于深度学习的医学影像辅助诊断技术研究[D]. 合肥：中国科学技术大学，2022.
YIN B C. Research on computer aided diagnosis of medical images based on deep learning[D]. Hefei：University of Science and Technology of China，2022.

[10] 陈吉栋，何丽. 生成式人工智能风险治理亟待“伦理-法律”综合框架[J]. 张江科技评论，2023(2)：8-10.
CHEN J D，HE L. Risk management of generative artificial intelligence urgently needs a comprehensive framework of “ethics-law”[J]. Zhangjiang Technol Rev，2023(2)：8-10.

[11] 宋黎磊，戴淑婷. 科技安全化与泛安全化：欧盟人工智能战略研究[J]. 德国研究，2022，37(4)：47-65，125.
SONG L L，DAI S T. Shuting technology security and pan security：research on EU artificial intelligence strategy[J] Ger Res，2022，37(4)：47-65，125.

（收稿日期：2023-05-23　修回日期：2023-09-15）
（编辑：刘明伟）