·临床与预防·

大语言模型在心理健康干预中的研究进展

薛旭[1],陈坤[1],刘佳雯[1],贾鹏[2,3],余彬[1,4],杨淑娟[1,3,5]

1. 四川大学华西公共卫生学院/华西第四医院,四川 成都 610041;

2. 武汉大学人民医院代谢与泛血管病医院;武汉大学人民医院临床医学研究中心;

3. 武汉大学空间全生命周期健康国际研究中心;4. 四川大学 - 香港理工大学灾后重建与管理学院;

5. 主动健康与食品营养安全四川省重点实验室

摘要:目的 本研究系统梳理采用大语言模型(Large language models, LLMs)进行心理健康干预的研究现状、可行性及未来展望。**方法** 在中国知网、万方数据知识服务平台、维普网、Pubmed、Scopus、Web of Science 上检索 2023 年 1 月至 2025 年 3 月期间发表的有关 LLMs 在心理健康干预领域文献,筛选后,使用混合方法评价工具(MMAT)进行质量评价,并提取文献的基本信息、研究模型、模型支持语言等内容。**结果** 纳入 21 篇英文文献,12 项研究采用人群干预试验,研究了 LLMs 在减少焦虑、识别干预认知扭曲、测量和培养儿童心理弹性、减轻抑郁和孤独感以及提升正念训练参与意愿等方面的干预效果,结果显示 LLMs 在以上任务中有一定效果。另外 9 项研究从心理知识咨询、情绪支持、治疗辅助三个方面证明 LLMs 心理干预可行性较高,但仍存在偏见、文化适应性和隐私保护等问题。7 项研究选择英语作为模型语言,13 项通过提示工程或指令微调对模型进行了调整,且调整后的模型在具体心理健康任务中具有更好的表现。文献的质量评价表明研究质量普遍不高。**结论** LLMs 在心理健康促进领域具备一定应用前景,未来应进一步完善模型策略以提高任务表现、并设计规范严谨的实证研究以获得科学的效果评价。

关键词:心理健康促进;人工智能;大语言模型

中图分类号:B84 - 05 **文献标志码:**A **文章编号:**1003 - 8507(2025)22 - 4194 - 07

DOI:10. 20043/j. cnki. MPM. 202505303

Applications of large language models in mental health interventions: feasibility, current status, and future trends

XUE Xu*, CHEN Kun, LIU Jia - wen, JIA Peng, YU Bin, YANG Shu - juan

* *West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan* 610041, *China*

Abstract: Objective To systematically review the current research status, feasibility, and future prospects of using large language models (LLMs) for mental health interventions. **Methods** We searched literature published between January 2023 and March 2025 on LLM - based mental health interventions across databases including CNKI, Wanfang Data, VIP, PubMed, Scopus, and Web of Science. After screening, the Mixed Methods Appraisal Tool (MMAT) was used for quality assessment. Key data were extracted, including study characteristics, model types, and supported languages. **Results** Twenty - one English - language studies were included. Twelve studies employed population - based intervention trials, investigating the effects of LLMs in reducing anxiety, identifying and intervening in cognitive distortions, measuring and fostering psychological resilience in children, alleviating depression and loneliness, and enhancing willingness to engage in mindfulness training. Results indicated that LLMs demonstrated some efficacy in these tasks. Another seven studies supported the feasibility of LLM - based mental health interventions in three key areas: psychological knowledge consultation, emotional support, and therapeutic assistance. However, challenges such as bias, cultural adaptability, and privacy concerns remain. Seven studies used English as the model language, while thirteen applied prompt engineering or instruction fine - tuning to adapt the models, with adjusted models showing improved performance in specific mental health tasks. Quality assessment revealed that existing studies were limited. **Conclusion** LLMs show promising potential in mental health promotion. Future research should refine

基金项目:国家重点研发计划(2023YFC3604702);国家自然科学基金青年学生基础研究项目(博士研究生)(723B2017);四川大学“从 0 到 1”创新研究项目(2023CX21)

作者简介:薛旭和陈坤为共同第一作者。薛旭(2003—),男,本科在读,研究方向:健康行为与社会医学;陈坤(2003—),男,本科在读,研究方向:健康行为与社会医学

通信作者:杨淑娟,E - mail:rekiny@126. com

model strategies to enhance task performance and incorporate rigorously designed empirical research to ensure scientifically robust efficacy evaluations.

Keywords: Mental health promotion; Artificial intelligence; Large language models

心理健康问题已成为日益增长的全球性公共卫生问题。2021 年全球疾病负担数据显示,超过 1.55 亿的伤残调整寿命年归因于抑郁、焦虑等精神障碍,约占全球疾病负担的 5%[1]。不同人群均出现了心理健康风险上升的趋势。例如,2021 年儿童及青少年群体中抑郁症的患病率约为 6%[1]。约 25% 的老年人患有抑郁和焦虑等心理健康问题[2]。然而,心理健康资源分布不均,中低收入国家资源短缺等问题依然严重,尤其在后疫情时代,心理健康服务的可及性与公平性面临更大挑战,可及有效的心理健康服务供给成为公共卫生领域的迫切需要[3]。

在此背景下,基于人工智能(Artificial Intelligence, AI)的健康技术逐渐成为公共卫生群体性干预的重要工具。大语言模型(Large Language Models, LLMs)作为近年来迅速发展的 AI 技术,通过大规模文本数据训练,构建起庞大的神经网络结构,以此处理复杂语义信息,实现情感分析、交互引导与知识传递等功能,为心理健康资源补充和群体心理健康促进带来新的可能[4]。针对性微调后的 LLMs 可被应用于心理健康教育、情绪识别与筛查、用户心理状态反馈、远程干预与辅助决策等方面[5],为传统心理服务提供补充,拓展公共卫生干预的可及路径。但 LLMs 在实际应用中仍面临诸多挑战,包括心理问题污名化[6]、人类参与度较低[7]、隐私数据安全难题[8]、且输出内容准确性与可解释性有待提高[9-11]。此外,训练数据更新速度快,限制了 LLMs 整合近期事件和信息的能力[12]。因此,为推动 LLMs 在公共卫生领域的科学应用,本研究在公共卫生框架下探讨 LLMs 作为心理健康干预工具的潜力、技术瓶颈与未来发展方向,旨在为群体心理健康促进提供新路径与理论支持,提升大语言模型在公共卫生服务中的科学性与适用性。

1　资料和方法

1.1　*文献检索策略*　在中国知网(CNKI)、万方数据知识服务平台、维普网、Pubmed、Scopus、Web of Science 检索 2023 年 1 月至 2025 年 3 月的全部有关 LLMs 和心理健康干预的中英文文献。中文检索词为:“大语言模型”“人工智能”“聊天机器人”“ChatGPT”“GPT-3”“GPT-4”、“心理健康”“心理状况”“抑郁”“焦虑”与“干预”“案例分析”“案例研究”“效果评估”“效果评价”“心理治疗”“心理咨询”。英文检索词为:“large language model”“LLM”“AI Chatbot”“ChatGPT”“GPT-3”“GPT-4”“OpenAI language model”“mental”、“mental health”“psychological health”“depression”“anxiety”与“intervention”“intervene”“case study”“effectiveness evaluation”“psychotherapy”“counseling”“therapy”“psychiatry”。使用自由词、布尔逻辑运算符撰写检索式。

1.2　*纳入与排除标准和文献筛选*　依据研究目的制定标准如下,纳入标准:①2023 年 1 月至 2025 年 3 月期间发表的期刊论文、会议文献;②利用 LLMs 进行心理健康干预试验和评价,或采用虚拟案例评价 LLMs 心理健康干预可行性的研究;③明确提供与所研究心理健康结局相关的可测量结果,如参与度、干预效果或测试结果等。排除标准:①评论、综述等非原创性研究;②所纳入模型为 2020 年 5 月(GPT-3 正式发布)前发布的 LLMs(如:BERT、GPT-2、T5)的研究;③重复发表的文献;④无法获取全文的文献;⑤未同行评审的预印本;⑥非中文或英文文献。使用 Zotero 7.0 管理文献,两名研究员独立进行三轮筛选,以第三名研究员为基准解决所有分歧。

1.3　*数据提取与质量评价*　一名研究员提取筛选后文献的数据,包括:①基本信息:第一作者、发表年月、研究模型、研究地区、模型语言和调整策略;②研究设计:研究内容、样本信息、结局变量或测量指标;③结果评价:主要结论、干预结果或评估结果、优势与局限性。使用 2018 版的混合方法评价工具(Mixed Methods Appraisal Tool, MMAT)[13]对纳入的研究进行质量评估。由两名研究员分别独立评估每项研究的质量,以第三名研究员为基准解决所有分歧。

2　结　果

2.1　*检索结果*　共检索到 2 382 篇文献,其中 CNKI 242 篇,万方 220 篇,维普 88 篇,Pubmed 256 篇,Scopus 567 篇,Web of Science 1 009 篇。排除重复文献后剩余中文 441 篇,英文 1 399 篇,筛选后纳入 21 篇英文文献,无中文文献。筛选过程见图 1。

本研究根据所纳入研究的研究设计将其分为:干预研究和可行性探索研究。干预研究直接以 LLMs 为心理健康干预措施进行心理健康干预试验;可行性探索研究针对具体场景,基于自动评估、专家评分及案例分析等研究手段对以 LLMs 为心理健康干预工具的潜在可行性做出一定探索,但未进行干预试验验证实际有效性。

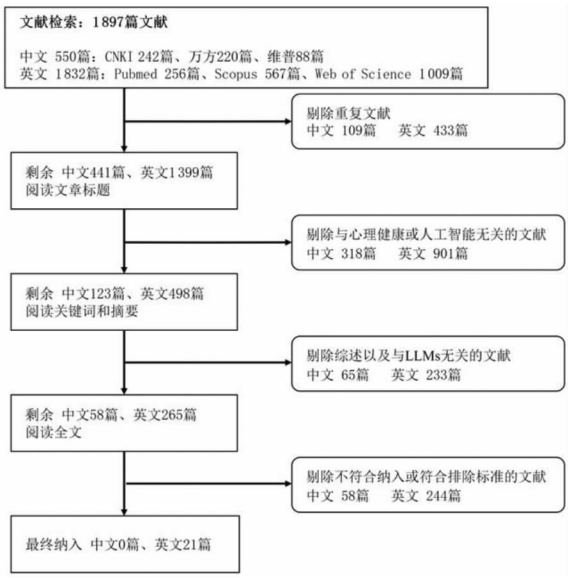


图 1 文献筛选流程图

Fig. 1 Flow chart of literature selection

2.2 纳入的干预研究介绍 在最终纳入的研究中，共 12 项研究[14-25]直接利用 LLMs 工具进行心理健康干预试验，其中包括 9 项[14-16,18,20,21,23-25]定量研究及 3 项[15,19,22]定量定性混合研究。研究设计方面，6 项研究[14,15,20,21,23,25]设置了实验组和对照组，其余 6 项研究[16-19,22,24]采用自身前后对照。研究人群方面，各 1 项针对老年人[15]及女性群体[25]，2 项[19,22]针对儿童，其余研究无针对性人群。详见表 1。

本研究根据研究者是否对干预使用的 LLMs 进行微调或提示，将研究分为“通用大模型（General-purpose Large Models，GLMs）”干预研究及“垂直大模型（Vertical Large Models，VLMs）”干预研究。GLMs 是指用于处理多任务与多数据类型的人工智能模型，通常基于多模态数据进行训练，支持多格式内容的理解与生成，具有高度灵活性与广泛适用性，可直接或少量适配后应用于多个领域与场景。VLMs 是针对特定行业或领域优化的模型，相比 GLMs，其专注于特定任务或数据集，能提供特定领域内高精度与专业化的输出。

4 项研究[14,15,20,23]以 GLMs 为干预工具，使用了包括 ChatGPT、Gemini、Perplexity 在内的模型。研究结果表明 LLMs 能显著降低术前焦虑，但无法降低考试焦虑，且其降低老年人孤独感和抑郁的能力有限，相比正念冥想并无差别。其余 8 项[16-19,21,22,24,25]以 VLMs 为干预工具的研究，则聚焦于 LLMs 在处理负面情绪、认知扭曲、儿童心理弹性等心理健康问题中的表现，以及其提升正念冥想、自我同情写作等心理健康干预措施的能力。结果表明，除了战争情景下心理健康支持以及心理咨询师自我效能提升与焦虑缓解任务外，LLMs 均有较好表现。

表 1 干预研究文献基本信息汇总

Table 1 Summary ofbasic information of intervention research literature

第一作者	模型及其类别	模型语言	研究设计	数据集
Musashi Yahagi（2024）[14]	GPT-3.5；GLMs	日语	RCT	85 名成年手术患者；对照组 44 名，干预组 41 名
Ying Wang（2024）[15]	GPT-3.0；GLMs	中文	RCT	12 名养老院参与者；ChatGPT 组 4 人，正念组 8 人
Gyeyoung Jung（2024）[16]	GPT-4.0；VLMs	英语	non-RCT	11 名大学生，其中女性 3 名，男性 8 名
Deniz Nazarova（2023）[17]	ada、babbage、curie、davinci；VLMs	英语	non-RCT	数据集：自编，含 240 个认知扭曲示例；人工评估：68 名大学生
Harsh Kumar（2023）[18]	GPT-3.0；VLMs	英语	non-RCT	189 位来自 Amazon Mechanical Turk 的对象
Zihui Hu（2024）[19]	Qwen-7B；VLMs	中文	non-RCT	微调数据集：儿童情感支持对话语料库干预试验：48 名儿童
Wenyi Gan（2025）[20]	GPT-4.0；GLMs	中文	RCT	55 名成年手术患者；对照组 28 名，干预组 27 名
Majid Ali（2025）[23]	ChatGPT、Gemini、Perplexity；GLMs	NA	RCT	88 名药学学生；对照组和干预组各 44 名
Tomoko Kishimoto（2025）[24]	GPT-4.0；VLMs	中文	non-RCT	98 名成年人
Dayeon Jeong（2025）[21]	GPT-4.0；VLMs	韩语	RCT	41 名咨询师；对照组 20 人，干预组 21 人
Liana Spytska（2025）[25]	Llama 3.1；VLMs	乌克兰语	RCT	104 名 18～60 岁战区女性；对照组 52 名，实验组 52 名
Joe Hasei（2025）[22]	GPT-4.0；VLMs	日语	non-RCT	5 名儿童青少年及年轻成人癌症患者

第一作者	研究内容	主要结论	局限性	质量评价结果[a]
Musashi Yahagi（2024）[14]	ChatGPT-3.5 对术前焦虑改善和满意度提升	ChatGPT 显著降低术前焦虑	单盲设计	4/5
Ying Wang（2024）[15]	聊天机器人与正念疗法对老年人孤独感和抑郁的影响	ChatGPT 及正念冥想短期内缓解老年人抑郁和孤独感能力有限	样本量小；未随机分配	2/5
Gyeyoung Jung（2024）[16]	开发 AI 介导的日记应用并评估其缓解抑郁和孤独感的能力	开发的 MyListener 有效降低用户使用前后的抑郁情绪和孤独感	未说明对 LLM 的处理；未考虑混杂因素；特征分布不均匀	3/5

（续表）

第一作者	研究内容	主要结论	局限性	质量评价结果[a]
Deniz Nazarova（2023）[17]	结合认知行为疗法开发用于认知扭曲识别及干预的聊天机器人	开发的 TeaBot 能识别 10 种认知扭曲并减少用户消极想法	未提供详细结果；未提供各组的基线数据；失访率高	1/5
Harsh Kumar（2023）[18]	LLMs 对正念训练意愿的提升	与对照组相比，聊天机器人干预提高了参与者再次练习正念冥想的意图	样本代表性存疑；未进行定量分析	2/5
Zihui Hu（2024）[19]	开发评估用于测量和培养儿童心理弹性的聊天机器人	心理弹性智慧测量优于传统测量，聊天机器人的干预提升了儿童自身心理韧性和幸福感	结果展示不完整；未进行定量分析	3/5
Wenyi Gan（2025）[20]	ChatGPT 对全膝关节置换术前患者焦虑的改善	相比对照组，ChatGPT 显著降低术前焦虑	单盲设计；未考虑人口统计学差异	4/5
Majid Ali（2025）[23]	AI 工具对药学学生结构化临床考试中的表现和考试焦虑的影响	未观察到 AI 对成绩和考试焦虑的影响	未说明所用 LLM 版本；可能存在实验污染	3/5
Tomoko Kishimoto（2025）[24]	结合 ChatGPT 反馈的单次在线自我同情写作减轻焦虑的效果	结合 ChatGPT 反馈可显著提高干预效果	各组未同时进行实验	5/5
Dayeon Jeong（2025）[21]	AI 对培训中的心理咨询师自我效能提升和焦虑缓解	AI 辅助反馈效果弱于自我反省	体验缺乏沉浸感	3/5
Liana Spytska（2025）[25]	“Friend”聊天机器人在战争情境中提供心理支持的有效性	聊天机器人的干预效果弱于治疗师干预	未考虑其他危机情景或文化背景	1/5
Joe Hasei（2025）[22]	聊天机器人对儿童少年癌症患者心理负担的缓解和治疗参与度的提升	聊天机器人能降低患者焦虑，提高治疗积极性	无标准测量工具，结果为自评	3/5

注：a 文献符合标准的条数，满分 5，下同；RCT：随机对照试验；non－RCT：非随机对照试验；GLMs：通用大语言模型，下同；VLMs：垂直大语言模型，下同。

2.3　纳入的可行性探索研究　在最终纳入的研究中，共 9 项研究[26－34]探索了 LLMs 在心理干预领域的可行性，包括 7 项[26－32]定量研究与 2 项[33－34]定量定性混合研究。研究人群方面，1 项研究[28]针对老年群体，1 项[33]针对噩梦经历者及治疗师群体，其余无针对人群。3 项研究[28,30,32]的样本为虚拟的案例，详见表 2。

本研究将具体心理健康促进场景分为 3 类，包括：心理知识咨询[26－27]、情绪支持[28,29,34]和临床心理治疗辅助[30－33]。心理知识咨询方面，现有研究表明微调后的 LLMs 在该类型任务中的表现较好。情绪支持方面，在结合多模态交互后，LLMs 能有效地提供正向情感支持，并减少老年人的孤独感和社会孤立感。治疗辅助方面，LLMs 在辅助儿童注意缺陷多动障碍（Attention Deficit and Hyperactivity Disorder，ADHD）、抑郁障碍、噩梦和双向情感障碍治疗中有着良好表现，能提高治疗可及性、参与度、治疗效率等，但也存在性别和种族偏见、隐私保护、文化适应性等问题。

表 2　可行性探究文献基本信息汇总

Table 2　Summary ofbasic information of feasibility exploration literature

第一作者	模型及其类别	模型语言	数据集
Hongbin Na（2024）[26]	GPT－3.5－turbo－16k、LLaMA－Chinese－7B、Alpaca－Chinese－7B、Qwen－7B、Baichuan－7B；VLMs	中文	评估样本：100 个源于自编数据集的样本；评估人员：4 名心理学专业学生，1 名心理治疗师
Laberiano Andrade Arenas（2024）[27]	GPT－4.0；VLMs	英语	评估人员：15 名用户，5 位专家
Jose Llanes－Jurado（2024）[29]	text－davinci－002；GLMs	西班牙语	64 名西班牙语母语参与者
Abeer M. Alessa（2023）[28]	GPT－3.5－turbo；VLMs	英语	评估样本：5 个虚构案例；评估人员：1 名计算机科学教授，1 名研究生
Santiago Berrezueta－Guzman（2024）[31]	GPT－4.0；VLMs	英语、西班牙语	10 名 ADHD 领域专家
Inbar Levkovich（2023）[30]	GPT－3.5、GPT－4.0；GLMs	NA	测试用案例：8 个虚拟案例人工评估样本：1 249 名初级保健医生
Roy H. Perlis（2024）[32]	GPT－4－1106－preview；VLMs	英语	测试用案例：50 个基于现实生成的双相案例；人工评估样本：3 位双相障碍专家；27 名社区临床医生
Boyoung Kang（2025）[34]	GPT－4.0；VLMs	英语、韩语	20 名 18～27 岁的大学生
Yijun Zhao（2025）[33]	GPT－3.0；VLMs	NA	形成性调查：8 名治疗师，10 名噩梦经历者；评估人员：2 名治疗师，16 名噩梦经历者

第一作者	研究内容	主要结论	局限性	质量评价结果[a]
Hongbin Na（2024）[26]	使用基于 CBT 生成的数据集微调 LLMs 并评估在心理健康咨询中的表现	开发的 CBT－LLM 各方面表现优于基线模型	未明确定义认知扭曲；未定量分析	3/5

（续表）

第一作者	研究内容	主要结论	局限性	质量评价结果[a]
Laberiano Andrade Arenas（2024）[27]	开发评估用于心理健康支持任务的聊天机器人	开发的机器人能够提供个性化情绪支持，综合表现较好	未评估对心理健康和情绪管理的影响	3/5
Jose Llanes – Jurado（2024）[29]	开发评估用于社交情感激发的对话式虚拟人（VH）	VH 情感激发的表现符合预期，倾向于激发积极情绪	本地未部署导致响应慢和隐私问题	5/5
Abeer M. Alessa（2023）[28]	开发老年人语音对话陪伴系统并评估帮助老年人减少孤独感和社会孤立感	系统生成内容质量较高，流畅度和内容意义表现优异	虚构案例；仅采取人工评估；评估人员少	3/5
Santiago Berrezueta – Guzman（2024）[31]	GPT – 4.0 对 ADHD 治疗的改善	ChatGPT 在 ADHD 治疗中个性化、可及性，但隐私、文化适应性尚不足	仅采取人工评估	5/5
Inbar Levkovich（2023）[30]	ChatGPT – 3.5 和 ChatGPT – 4.0 与初级保健医生对抑郁发作的评估和建议治疗方案比较	ChatGPT – 4 的治疗方案更符合临床指南，未检出性别和社会经济地位相关偏见	虚构案例	4/5
Roy H. Perlis（2024）[32]	GPT – 4.0、临床指南提示下的 GPT – 4.0、专家及社区临床医生对双向情感障碍的治疗建议比较	模型中检出性别和种族偏见，提示下的 GPT 表现最佳但与专家建议一致性低	专家间意见一致性差	4/5
Boyoung Kang（2025）[34]	开发跨语言心理健康聊天机器人	开发模型在支持性、同理心、界面设计方面表现较好，交互、专业性和个性化有待改进	仅采取人工评估	4/5
Yijun Zhao（2025）[33]	开发基于视觉交互和叙事生成技术的噩梦治疗辅助系统	该系统能提高治疗师工作效率、促进互动、提供治愈支持并激发对噩梦的深度探索	仅采取人工评估	4/5

2.4 *研究质量评价* 使用 MMAT 混合评价标准对纳入的 21 项研究进行质量评价，3 项研究[24,29,31]符合 100% 质量标准（5/5），6 项[14,20,30,32-34]符合 80% 质量标准评价（4/5），8 项[16,19,21-23,26-28]符合 60% 质量标准评价（3/5），2 项[15,18]符合 40% 质量标准评价（2/5），2 项[17,25]符合 20% 标准（1/5）。不符合标准的原因主要包括：研究设计未考虑混杂因素，抽样方法不合理，基线特征分布不均匀或未报告，效应测量方式不合理，研究对象失访率高，结果报告不全或有误，或只对结果进行描述而未做统计分析，详见图 2。

在这 21 项研究中，仅 9 项[14,20,22-23,25,30-32,34]发表在医学相关领域期刊，3 项分别来自行为科学[15]、心理学[24]、教育学[21]相关领域，剩余 9 项[16-19,26-29,33]均来自计算机科学相关领域期刊或会议，该 9 项研究的 MMAT 评价结果普遍不高于 60%（仅 1 项[29]符合 100% 标准）。

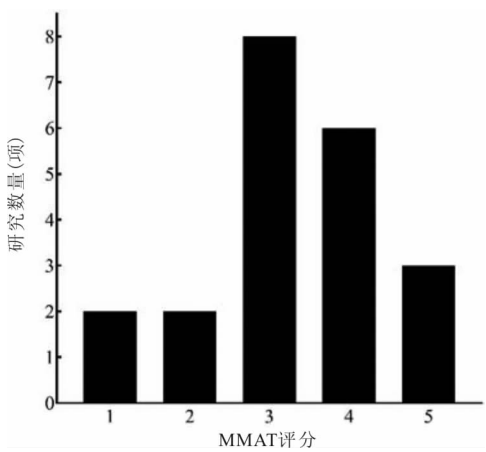


图 2 MMAT 文献质量评价结果

Fig. 2 Quality of literature evaluated by MMAT

3 讨 论

本研究结果显示心理健康干预 LLMs 在多人群中效果初步可行。基于较先进的 GPT – 3 或后续 LLMs，无论调整与否，其在人群心理健康促进方面均有一定可行性。4 项混合方法研究认为 LLMs 在测量和干预儿童心理韧性，缓解老年人孤独感和抑郁、缓解儿童少年癌症患者心理负担，以及噩梦治疗等方面有较可观的应用前景；17 项定量研究表明 LLMs 干预对 ADHD、术前焦虑、孤独感、抑郁、认知扭曲和双相情感障碍的效果显著，在情绪激发、提高正念参与意愿和心理健康知识问答任务中的表现优秀，但对考试焦虑和战争情景下焦虑的干预效果有限。研究还表明，指令微调或提示工程训练后的模型具有更好的表现。

现有研究质量亟需提升，缺乏可靠人群试验研究。LLMs 在心理健康干预中的应用仍处于探索阶段，不仅缺乏多场景实证研究，也未充分考虑人群特征和文化背景等关键变量。在本研究所纳入的研究中，仅有 6 项研究[14,17,20-21,23,25]采取 RCT 设计，且在随机化、样本量等方面存在不足，研究间存在较大质量差异，有 2 项研究[17,25]未严格遵守 RCT 设计，可能与研究团队缺乏医学背景有关。在 8 项包含定量非随机试验的研究中，仅 2 项研究[29,33]考虑了混杂因素，另有 4 项研究[15-16,19,22]有明显偏倚风险。另 3 项研究[28,30,32]中使用了虚构的案例，可能限制其结论在复杂现实世界中的可推广性。仅 2 项研究[15,17]的干预时长达到 8 周，LLMs 的长期干预效果尚不明确。

纳入研究中，仅 3 项[17,19,26]通过在线心理健康对话数据集或自建数据集对基础大模型进行指令微调，

10 项[18,21-22,24,27,28,31-34]使用提示工程方法,通过引入具体任务描述、心理干预知识或用户个体信息以优化模型表现。在语言适配方面,研究大多集中于英语语言环境,共 7 项研究[16-18,27,28,31-32,34]使用英语支持模型,仅 5 项[15,19-20,24,26]涉及中文,另各有 2 项研究使用了支持西班牙语[29,31]、日语[14,22]及韩语[21,34]的模型,1 项研究[25]支持乌克兰语。语言分布的不均衡暴露出当前 LLMs 在全球心理健康干预中对非英语人群适配性和文化多样性关注不足。

目前聚焦心理健康干预 LLMs 的公共卫生研究仍十分有限,且对干预人群特征、社会文化背景、模型训练语言及具体心理健康问题的考虑尚不充分,这可能源于专业化数据的不足。例如,HuggingFace 等开源社区中的对话数据集多为通用医疗健康问答,对心理干预类场景覆盖有限,制约了微调等方式对模型的增强。另外,已有研究指出思维链技术(Chains of thoughts,COT)可显著提高 LLMs 的共情[35]、推理[36]和策略性生成能力,但尚无聚焦心理健康的模型引入该技术。部分研究尝试引入语音语调、面部表情等非语言输入,以生成更具同理心和语境适应性的多模态回应。提示融合语言与感知的多模态路径或为提升干预效果的潜在方向[37]。

本研究的优势在于:首先,仅纳入基于 GPT-3 及后续模型的研究,确保所评估模型的先进性。其次,从干预工具类型(通用/垂直领域)、应用目标及研究设计等维度系统梳理研究内容。此外,采用 MMAT 标准评估研究质量,并结合干预对象选取标准、训练语言及数据集内容,探讨 LLMs 对不同人群的适配性,以及指令微调、提示工程等优化手段的作用。

4　结　论

当前,高性能、低成本且可本地部署的开源 LLMs(如 DeepSeek-V3)不断涌现[38],为开展高质量的心理健康干预试验提供技术支持,也使面向特定心理问题或人群的高效干预成为可能。因此,LLMs 在心理健康促进领域展现出巨大应用潜力。未来的研究应鼓励公共卫生、计算机科学、心理学等领域深度融合,利用大样本人群队列数据构建高质量对话数据集,支持 LLMs 开展更专业的定制化训练,同时,应加强试验设计的科学性与规范性,通过更严谨的随机对照试验评价干预效果,多领域合作共同促进社会心理健康资源的可及性及公平性。

利益冲突声明　本研究不存在任何利益冲突

参考文献

[1] GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy Life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021[J]. Lancet, 2024, 403(10440): 2133-2161.

[2] Chobe S, Chobe M, Metri K, et al. Impact of yoga on cognition and mental health among elderly: a systematic review [J]. Complementary Therapies in Medicine, 2020, 52: 102421.

[3] Hossain MM, Tasnim S, Sultana A, et al. Epidemiology of mental health problems in COVID-19: a review[J]. F1000Research, 2020, 9: 636.

[4] Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine[J]. Nature Medicine, 2023, 29(8): 1930-1940.

[5] Lawrence HR, Schneider RA, Rubin SB, et al. The opportunities and risks of large language models in mental health[J]. JMIR Ment Health, 2024, 11: e59479.

[6] Koutsouleris N, Hauser TU, Skvortsova V, et al. From promise to practice: towards the realisation of AI-informed mental health care [J]. Lancet Digital Health, 2022, 4(11): e829-e840.

[7] Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare[J]. NPJ Digit Med, 2023, 6(1): 1-6.

[8] Omarov B, Zhumanov Z, Gumar A, et al. Artificial intelligence enabled Mobile chatbot psychologist using AIML and cognitive behavioral therapy[J]. Int. J. Adv. Comput. Sci. Appl, 2023, 14(6): 616.

[9] Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine [J]. EBioMedicine, 2023, 90: 104512.

[10] Sickel AE, Seacat JD, Nabors NA. Mental health stigma update: A review of Consequences[J]. Adv Ment Health, 2014, 12(3): 202-215.

[11] Zhong Y, Chen YJ, Zhou Y, et al. The artificial intelligence large language models and neuropsychiatry practice and research ethic [J]. Asian Journal of Psychiatry, 2023, 84: 103577.

[12] Mündler N, He J, Jenko S, et al. Self-contradictory hallucinations of large language models: Evaluation, detection and mitigation[EB/OL]. [2025-09-14]. http://arxiv.org/abs/2305.15852.

[13] Hong QN. Gonzalez-reyes a, pluye P. MMAT-Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the mixed methods appraisal Tool (MMAT) [J]. Journal of Evaluation in Clinical Practice, 2018, 24(3): 459-467.

[14] Yahagi M, Hiruta R, Miyauchi C, et al. Comparison of conventional anesthesia nurse education and an artificial intelligence chatbot (ChatGPT) intervention on preoperative anxiety: a randomized controlled trial[J]. Journal of Perianesthesia Nursing: Official Journal of the American Society of PeriAnesthesia Nurses / American Society of PeriAnesthesia Nurses, 2024, 39(5): 767-771.

[15] Wang Y, Li S. Tech vs. tradition: ChatGPT and mindfulness in enhancing older adults' emotional health[J]. Behavioral Sciences, 2024, 14(10): 923.

[16] Jung G, Choi S, Kang Y, et al. MyListener: an AI-Mediated journaling Mobile application for alleviating depression and loneliness using contextual data[C]//Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous

Computing,2024.
[17] Nazarova D. Application of artificial intelligence in mental healthcare:generative pre - trained transformer 3 (GPT - 3) and cognitive distortions[EB/OL]. [2025 - 09 - 30]. https://link.springer.com/chapter/10.1007/978 - 3 - 031 - 47454 - 5_16.
[18] Kumar H, Wang Y, Shi J, et al. Exploring the use of large language models for improving the awareness of mindfulness[EB/OL]. [2025 - 09 - 30]. https://dl.acm.org/doi/pdf/10.1145/3544549.3585614.
[19] Hu Z, Hou H, Ni S. Grow with your AI buddy: Designing an LLMs - based conversational agent for the measurement and cultivation of children's mental resilience[EB/OL]. [2025 - 09 - 30]. https://dl.acm.org/doi/10.1145/3628516.3659399.
[20] Gan W, Ouyang J, She G, et al. ChatGPT's role in alleviating anxiety in total knee arthroplasty consent process: a randomized controlled trial pilot study[J]. International Journal of Surgery, 2025, 111(3): 2546 - 2557.
[21] Jeong D, Lee H, Yang H, et al. AI integration in counseling training: aiding Counselors - in - Training in Self - Efficacy enhancement and anxiety reduction[J]. Journal of Counseling and Development : JCD, 2025, 103(3): 346 - 360.
[22] Hasei J, Hanzawa M, Nagano A, et al. Empowering pediatric, adolescent,and young adult patients with cancer utilizing generative AI chatbots to reduce psychological burden and enhance treatment engagement: a pilot study[J]. Front Digit Health, 2025, 7: 1543543.
[23] Ali M, Rehman S, Cheema E. Impact of artificial intelligence on the academic performance and test anxiety of pharmacy students in objective structured clinical examination: a randomized controlled trial[J]. Int J Clin Pharmdoi, 2025, 47(4): 1034 - 1041.
[24] Kishimoto T, Hao X, Chang T, et al. Single online self - compassion writing intervention reduces anxiety:With the feedback of ChatGPT[J]. Internet Interventions - The Application of Information Technology in Mental and Behavioural Health, 2025, 39: 100810.
[25] Spytska L. The use of artificial intelligence in psychotherapy: development of intelligent therapeutic systems[J]. BMC Psychology, 2025, 13: 175.
[26] A Na H. CBT - LLM: A Chinese Large Language Model for Cognitive Behavioral Therapy - based Mental Health Question Answering[EB/OL]. [2025 - 09 - 30]. https://aclanthology.org/2024.lrec - main.261.pdf.
[27] Andrade AL, Yactayo - Arias C. Chatbot with ChatGPT technology for mental wellbeing and emotional management[J]. IAES Int. J. Artif. Intell, 2024, 13(3): 2635.
[28] Alessa A, Al - Khalifa H. Towards designing a ChatGPT conversational companion for elderly people[EB/OL]. [2025 - 09 - 30]. https://dl.acm.org/doi/10.1145/3594806.3596572.
[29] Llanes - Jurado J, Gómez - Zaragozá L, Minissi ME, et al. Developing conversational Virtual Humans for social emotion elicitation based on large language models[J]. Expert Systems with Applications, 2024, 246: 123261.
[30] Levkovich I, Elyoseph Z. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians[J]. Fam Med Community Health, 2023, 11(4): e002391.
[31] Berrezueta - Guzman S, Kandil M, Martín - Ruiz ML, et al. Future of ADHD care: evaluating the efficacy of ChatGPT in therapy enhancement[J]. Healthcare, 2024, 12(6): 683.
[32] Perlis RH, Goldberg JF, Ostacher MJ, et al. Clinical decision support for bipolar depression using large language models[J]. Neuropsychopharmacology, 2024, 49(9): 1412 - 1416.
[33] Zhao Y, Li Z, Wang Y, et al. DreamDirector: designing a generative AI system to aid therapists in treating clients' nightmares[EB/OL]. [2025 - 09 - 30]. https://dl.acm.org/doi/10.1145/3708359.3712078.
[34] Kang b HMD, Evaluation of a Mental Health Chatbot Using ChatGPT 4.0: mixed methods user experience study with korean users[J]. JMIR Med Inform, 2025, 13: e63538.
[35] 桑晨扬,马廷淮,谢欣彤,等.基于大语言模型多阶段推理的情绪支持对话生成方法[J].计算机科学与探索,2024,18(11):2925 - 2939.
Sang CY, Ma TH, Xie XT, et al. Emotion - supportive dialogue generation method based on multi - stage reasoning of large language models[J]. Journal of Frontiers of Computer Science & Technology, 2024, 18(11): 2925 - 2939. (In Chinese)
[36] Wei J, Wang X, Schuurmans D, et al. Chain - of - thought prompting elicits reasoning in large language models[EB/OL]. [2025 - 09 - 30]. https://arxiv.org/abs/2201.11903.
[37] Rühlemann C, Trujillo J. The effect of gesture expressivity on emotional resonance in storytelling interaction[J]. Frontiers in Psychology, 2024, 15: 1477263.
[38] Deepseek - Ai, Liu A, Feng B, et al. DeepSeek - V3 technical report[EB/OL]. [2025 - 09 - 30]. https://arxiv.org/abs/2412.19437.

收稿日期:2025-05-22

(上接第 4064 页)

[18] Liu JT, Yao HY, Yu SC, et al. Joint association of metabolic health and obesity with Ten - Year risk of cardiovascular disease among Chinese adults[J]. Biomedical and Environmental Sciences, 2022, 35(1): 13 - 21.
[19] Meng MM, Guo YX, Kuang ZR, et al. Risk of stroke among different metabolic obesity phenotypes: a systematic review and Meta - Analysis[J]. Front Cardiovasc Med, 2022, 9: 844550.
[20] Li QY, Wang PB, Ma R, et al. A novel criterion of metabolically healthy obesity could effectively identify individuals with low cardiovascular risk among Chinese cohort[J]. Frontiers in Endocrinology, 2023, 14: 1140472.
[21] Wei Y, Wang J, Han X, et al. Metabolically healthy obesity increased diabetes incidence in a middle - aged and elderly Chinese population[J]. Diabetes - Metabolism: Research and Reviews, 2020, 36(1): e3202.
[22] 孙梦,魏葳,陆海丽,等.北京市西城区 35 ~ 75 岁居民心血管疾病高风险人群现状调查[J].华南预防医学,2023,49(9):1188 - 1191.
Sun M, Wei W, Lu HL, et al. A survey on the high - risk population of cardiovascular diseasesamong residents aged 35 - 75 in Xicheng District, Beijing[J]. South China Journal of Preventive Medicine, 2023, 49(9): 1188 - 1191. (In Chinese)

收稿日期:2025-02-23