

人工智能赋能心理健康服务的发展现状与伦理挑战

李 丹

山东师范大学 山东济南

【摘要】随着人工智能技术的飞速发展，心理健康服务领域正经历深刻的数字化变革。人工智能辅助心理咨询作为新兴交叉领域，借助自然语言处理、机器学习、大语言模型等技术，在心理评估、危机干预、远程咨询等方面展现出巨大潜力。本文采用文献综述方法，以 2019—2024 年国内外人工智能与心理健康服务相关研究为主要分析对象，从技术应用、服务模式及伦理风险等方面进行归纳与讨论。研究发现，当前人工智能辅助心理咨询已形成以聊天机器人、数字疗法、智能评估系统为核心的技术体系，在缓解心理健康服务资源短缺、突破时空限制、降低求助门槛等方面成效显著；然而，其在数据隐私保护、算法偏见、人机关系边界、专业责任归属等方面仍面临严峻的伦理挑战。未来应着力构建技术赋能与伦理规制并重的治理框架，推动人工智能辅助心理咨询向更安全、更普惠、更人性化的方向发展。

【关键词】人工智能；心理咨询；心理健康；聊天机器人；伦理挑战；数字疗法

【收稿日期】2025 年 9 月 21 日 **【出刊日期】**2025 年 12 月 25 日 **【DOI】**10.12208/j.aip.20250007

The current development status and ethical challenges of AI-empowered mental health services

Dan Li

Shandong Normal University, Jinan, Shandong

【Abstract】 With the rapid development of artificial intelligence (AI) technology, the mental health service sector is undergoing a profound digital transformation. AI-assisted psychological counseling, as an emerging interdisciplinary field, leverages technologies such as natural language processing, machine learning, and large language models, demonstrating immense potential in areas such as psychological assessment, crisis intervention, and remote counseling. This paper employs a literature review approach, focusing on domestic and international research related to AI and mental health services from 2019 to 2024, summarizing and discussing aspects such as technology application, service models, and ethical risks. The study finds that current AI-assisted psychological counseling has formed a technological system centered on chatbots, digital therapy, and intelligent assessment systems, achieving significant results in alleviating the shortage of mental health service resources, overcoming time and space limitations, and lowering the threshold for seeking help. However, it still faces serious ethical challenges in areas such as data privacy protection, algorithmic bias, the boundaries of human-machine relationships, and the attribution of professional responsibility. Future efforts should focus on building a governance framework that emphasizes both technological empowerment and ethical regulation, promoting the development of AI-assisted psychological counseling towards a safer, more inclusive, and more humane direction.

【Keywords】 Artificial intelligence; Psychological counseling; Mental health; Chatbots; Ethical challenges; Digital therapy

1 引言

1.1 文献检索与筛选

本文以中国知网、万方、Web of Science、PubMed 等数据库为主要检索来源，以“人工智能”“心理咨

询”“心理健康”“聊天机器人”“数字疗法”“artificial intelligence”“mental health”“chatbot”“psychotherapy”等为主要关键词进行组合检索。按照研究主题相关性、发表时间及文献类型进行初步筛选，并通过阅

阅读全文排除与研究主题关联度较低的文献，最终纳入相关研究进行主题归纳。

1.2 背景

全球心理健康问题日益严峻，世界卫生组织数据显示，抑郁症是全球导致功能受损和残疾的重要原因之一，2019 年全球约有 2.8 亿人受到抑郁症影响。与此同时，心理健康服务需求持续增长，但专业服务资源仍存在较大缺口。在我国，随着社会转型加速、竞争压力增大及生活节奏加快，心理健康问题同样呈现出高发、低龄化与复杂化的特征。我国心理健康服务需求持续增长，但专业心理健康服务资源在总量、区域分布及服务可及性方面仍存在不足，基层及偏远地区尤其明显。特别是在基层医疗机构、偏远农村地区及特殊人群中，优质心理咨询服务的可及性严重不足，“看病难、咨询贵”成为制约公众心理健康水平提升的重要瓶颈。

与此同时，公众对心理健康的认知逐步提升，去污名化进程加速，主动寻求心理帮助的意愿显著增强。然而，传统心理咨询模式以面对面、一对一的形式为主，受限于咨询师数量、地域分布、服务时间及经济成本等因素，难以满足大规模、即时性、个性化的心理健康服务需求。在这一供需矛盾日益尖锐的背景下，信息技术的介入为心理健康服务的创新发展提供了全新可能。互联网、移动互联网、大数据及人工智能技术的成熟，使得远程心理服务、数字化干预、智能化评估等新型服务模式应运而生，为破解心理健康服务资源短缺困境开辟了新的路径。

人工智能与心理健康服务的结合并非偶然，而是技术演进与社会需求双重驱动的必然结果。靳宇倡等^[1]指出，远程心理健康服务通过互联网、移动应用程序及人工智能技术，突破了传统面对面咨询的时空限制，在新冠疫情期间更是成为保障公众心理健康的重要渠道。远程心理服务不仅能够提供在线心理咨询、心理测评与干预，还能借助人工智能实现智能化分诊与初步评估，显著提升服务效率。研究表明，远程心理服务对于常见心理健康问题的干预效果与传统面对面服务相当，尤其在焦虑症、抑郁症及睡眠障碍的治疗中表现突出。然而，远程心理服务在推广应用过程中也面临技术适配、伦理规范及服务质量保障等多重挑战，如网络稳定性、数字素养差异、非言语信息缺失等问题可能影响咨询效果。

在评估与干预层面，智能化转型已成为心理健康服务发展的重要趋势。任萍等^[2]系统梳理了人工智能和大数据技术在心理健康初筛、诊断及干预过程中的应用，指出智能化评估能够通过对语音、文本、面部表情、生理指标等多模态数据的分析，实现对心理健康状态的客观、动态监测，有效弥补传统自陈量表的局限性。传统心理评估依赖来访者的主观报告，存在社会期许偏差、回忆偏差及状态依赖性问题，而人工智能评估通过对自然行为数据的捕捉与分析，能够获取更为真实、稳定的心理特征。在干预环节，基于认知行为疗法、正念训练、接纳承诺疗法等循证方法的数字化干预程序已展现出良好的治疗效果，为大规模人群的心理健康促进提供了可行路径。

数字化干预的快速发展进一步拓展了人工智能在心理健康领域的应用边界。李佳等^[3]认为，数字化干预通过短信、电话、网页、社交媒体、虚拟现实及人工智能等多种技术形态，实现了心理健康服务的精准化与个性化。特别是虚拟现实技术与暴露疗法的结合，为焦虑症、创伤后应激障碍、特定恐惧症等心理疾病的治疗提供了沉浸式干预方案，使来访者能够在安全可控的虚拟环境中逐步面对恐惧刺激。人工智能技术的嵌入使得数字化干预能够根据用户的实时反馈动态调整干预策略，提升干预的适应性与依从性。此外，数字化干预还具有成本低、覆盖面广、可重复性强等优势，有助于缩小心理健康服务的城乡差距与阶层差距。

从学科发展的宏观视角来看，人工智能技术正在深刻重塑心理学的研究范式与实践形态。刘冬予等^[4]分析了人工智能技术赋能心理学发展的现状与挑战，指出人工智能不仅在心理测量、临床诊断、心理咨询等应用领域发挥重要作用，还推动了心理学研究方法的革新，如基于电子健康记录的大数据挖掘、纵向追踪数据的智能分析、复杂网络建模等。人工智能与心理学的交叉融合催生了计算精神病学、数字临床科学等新兴分支学科，为理解心理现象的动态机制提供了新的工具与视角。然而，人工智能在心理学领域的应用也面临数据质量参差不齐、算法可解释性不足、伦理规范滞后等挑战，需要心理学研究者与技术开发者协同攻关，在技术创新与科学严谨之间寻求平衡。

聚焦到心理咨询这一核心领域，人工智能心理

咨询已逐步从概念走向实践,从实验室走向市场。李敬荣等^[5]对人工智能心理咨询的概念进行了界定,将其分为基于文本的聊天机器人咨询、基于语音的智能辅导系统及融合虚拟现实技术的沉浸式咨询等类型。研究指出,人工智能心理咨询在提供即时响应、降低求助门槛、保护用户隐私、消除地域限制等方面具有独特优势,尤其在处理轻度情绪困扰、提供心理教育、进行危机初步筛查、维持治疗间隙的支持性联系等场景中表现突出。以 Wysa、Woebot 等为代表的聊天机器人已进入心理健康支持场景,并积累了一定的用户基础。现有研究提示,此类工具在情绪支持、心理教育和行为自我管理方面具有一定应用潜力,但其长期疗效及对复杂心理问题的适用性仍需更多高质量临床研究验证。但同时,人工智能心理咨询也面临建立有效咨询关系、处理复杂情感问题、应对自杀危机、确保伦理合规等深层挑战,其服务边界与适用人群仍需通过严格的临床研究加以明确。

本文所称人工智能辅助心理咨询,是指利用自然语言处理、机器学习、生成式人工智能等技术,为心理健康教育、心理评估、情绪支持、咨询辅助及干预管理等提供技术支持的服务模式。根据人类专业人员参与程度,可进一步分为人工智能辅助型、人机协同型和高度自动化型。本文不将具有明确医疗适应证、独立承担疾病治疗功能的数字疗法或人工智能诊断系统与一般心理咨询工具完全等同,而是将其作为相关技术生态进行讨论。

2 人工智能辅助心理咨询的技术架构与核心机制

人工智能辅助心理咨询的效能发挥依赖于底层技术架构的成熟度与核心机制的有效性。当前,该领域已形成以自然语言处理、机器学习、知识图谱为支撑,以聊天机器人、数字疗法、智能评估系统为主要载体的技术体系。本章从治疗联盟构建、数字疗法应用及智能评估诊断三个层面,梳理人工智能辅助心理咨询的技术发展现状。

2.1 聊天机器人与数字治疗联盟的建立

治疗联盟是心理咨询成效的关键预测因子,传统治疗联盟的建立依赖于咨询师与来访者之间面对面的情感互动、共情回应与信任积累。在人工智能辅助心理咨询场景中,在缺乏人类咨询师持续参与的情况下,如何形成稳定、适度且不产生过度依赖

的人机工作关系,是人工智能心理健康服务设计需要重点解决的问题,也是决定用户参与度与干预效果的关键变量。磨然等^[6]从关系线索的角度探讨了聊天机器人与用户间数字治疗联盟的构建机制,指出聊天机器人可通过语言风格匹配、情感回应、自我披露、记忆连续性、个性化称呼等关系线索,模拟人类咨询师的共情行为与专业姿态,从而增强用户的参与感、信任感与归属感。研究表明,当聊天机器人展现出较高的关系导向特征时,用户更愿意持续使用服务、遵循干预建议并主动暴露内心困扰。

进一步地,磨然等^[7]从“拟人归因”到“联盟建立”的动态过程出发,揭示了人与聊天机器人关系发展的内在机制与阶段特征。研究发现,用户最初倾向于将聊天机器人拟人化,赋予其情感、意图与道德属性,这种拟人归因是建立初步联系的心理基础;随着互动深入,用户逐渐基于聊天机器人的实际功能表现形成稳定的关系预期,从情感性期待转向工具性期待,最终发展为具有功能性的工作联盟。这一过程提示,聊天机器人的设计需要在拟人化与功能透明性之间寻求平衡:过度拟人化可能导致用户产生不切实际的期待,甚至形成情感依赖;而适度的功能透明性反而有助于建立可持续、健康的治疗联盟。此外,关系建立的动态性还受到用户人格特质、求助动机、技术接受度等个体因素的调节,高依恋焦虑个体更容易对聊天机器人产生强烈情感投入,需要在设计中进行针对性风险提示。

2.2 数字疗法在精神疾病诊疗中的发展

数字疗法是数字化心理健康干预的重要组成部分,部分产品进一步融合人工智能技术,通过个体化推荐、行为监测和自适应干预提高服务的精准性,是指基于循证医学证据、经监管机构审批的软件程序,用于预防、管理或治疗疾病。与传统健康类应用程序不同,数字疗法需经过严格的临床试验验证,其疗效与安全性需达到医疗器械标准。张正等^[8]分析了数字疗法在精神疾病诊疗中的发展现状,指出数字疗法通过移动应用程序、可穿戴设备及在线平台,为患者提供结构化的心理干预方案,如认知行为疗法、正念减压、睡眠限制疗法、行为激活等。在抑郁症、焦虑症、物质使用障碍、失眠症等疾病的治疗中,数字疗法已积累了较为充分的临床证据,国外部分数字疗法产品已经获得监管机构授权,为数字化心理健康干预的临床应用提供了实践基础。但

不同产品的适应证、证据等级和监管路径存在明显差异,不能将个别产品的获批情况等等同于整个数字疗法领域已经成熟。

人工智能技术的融入使得数字疗法能够实现个性化内容推荐、实时症状监测及自适应干预调整,显著提升了治疗的精准度与用户体验。基于用户的行为数据与反馈信号,人工智能算法可动态识别用户的情绪状态变化,及时调整干预内容的难度与节奏;当检测到用户症状恶化或出现自杀风险时,系统可自动触发危机预警机制,转介至人类专业人员进行处理。然而,数字疗法在推广过程中也面临处方流程不畅、医保覆盖不足、患者依从性下降、临床医生接受度低等现实障碍,需要医疗体系、支付方及技术方的协同推进。

2.3 智能评估与辅助诊断技术

心理评估是心理咨询的前提与基础,传统评估方法依赖标准化量表与临床访谈,存在主观性强、时效性差、成本高等局限,且难以捕捉心理状态的动态变化。姚峰等^[9]综述了人工智能在心理评估中的研究进展,指出基于自然语言处理的文本情感分析、基于计算机视觉的面部表情识别、基于语音处理的声学特征提取、基于生理信号的多模态融合等技术,已能够实现对抑郁、焦虑、自杀风险、认知功能等心理状态的自动化识别与量化评估。特别是大语言模型的应用,使得人工智能能够从非结构化的社交媒体文本、临床访谈记录、日记等数据中提取深层心理特征,为早期筛查与风险预警提供了新的技术路径。

在诊断层面,智能辅助诊断系统的开发为心理健康评估提供了更为高效的工具支持。石晓飞等^[10]探讨了智能辅助诊断在心理健康评估系统中的应用与进展,指出基于机器学习算法的智能诊断系统能够整合多源异构数据,构建多维度的风险评估模型,辅助临床医生进行诊断决策。例如,基于电子健康记录的智能系统可通过分析患者的就诊历史、用药记录、实验室检查结果等信息,识别出潜在的心理风险因素,实现早发现、早干预。然而,当前智能辅助诊断系统仍面临诊断标准不统一、训练数据代表性不足、算法黑箱缺乏可解释性、跨人群泛化能力有限等问题,其诊断结果尚需结合临床专业判断进行综合考量,不能完全替代医生的诊断职能。

3 人工智能辅助心理咨询的应用场景与服务模式

技术的成熟推动了人工智能辅助心理咨询在多元场景中的落地应用。从生成式人工智能的颠覆性创新到网络疗法的普及推广,从医疗大模型的深度整合到特殊人群的精准服务,人工智能辅助心理咨询正逐步构建起覆盖预防、干预、康复全链条的服务生态。

3.1 生成式人工智能与大语言模型的应用

以 ChatGPT 为代表的生成式人工智能的出现,标志着人工智能辅助心理咨询进入新阶段。生成式预训练 Transformer 模型通过海量文本数据的预训练与基于人类反馈的强化学习微调,具备了强大的语言理解、逻辑推理与内容生成能力,大语言模型能够进行开放式自然语言交互,并可用于心理健康信息提供、情绪支持及部分结构化心理干预内容的辅助生成。在经过专业设计和安全约束的情况下,其部分输出可以模拟认知行为干预中的提问与反思过程,但其临床有效性和适用边界仍需通过严格研究加以验证。张熙等^[11]分析了 ChatGPT 及生成式人工智能的发展现状与未来方向,指出该类模型在心理咨询场景中可扮演“虚拟咨询师”角色,为用户提供 7×24 小时的心理健康信息支持,有效缓解专业人力资源不足的压力。与传统基于规则的聊天机器人相比,生成式大模型能够处理更为复杂、模糊的表达,生成更自然、连贯的回应,显著提升了人机交互的质量。

在医疗领域,人工智能大模型的应用进一步拓展了心理健康服务的深度与广度。郑琰莉等^[13]分析了人工智能大模型在医疗领域的应用现状与前景,指出大模型通过整合医学文献、临床指南、病例数据等专业知识,能够为心理咨询师提供决策支持、知识检索、培训教育及督导服务。在心理治疗方面,虚拟数字人可根据患者需求切换互动风格,通过人性化互动抚慰患者情绪,为传统治疗提供有益补充。大模型还可用于生成个性化的治疗材料,如针对特定认知偏差的自动重构语句、基于用户兴趣的正向活动建议等。然而,大模型在医疗领域的应用也面临幻觉生成、知识时效性不足、医疗责任界定模糊等风险,需要建立严格的质量控制机制与人工审核流程。

3.2 网络疗法与远程心理服务

互联网心理服务是传统心理健康服务向数字空间延伸的重要形式,包括在线心理咨询、远程心理

评估、数字化心理教育、网络支持及远程督导等。人工智能的嵌入进一步拓展了其自动化评估、智能匹配、随访和风险识别等功能。刘斌志等^[14]指出,网络疗法是社会工作者顺应 5G 通讯、大数据平台及人工智能等现代信息技术发展,针对虚拟社会生活功能处于弱势境地的服务对象,利用互联网、网络互动及虚拟现实等技术开展的专业服务。网络疗法不仅包括在线心理咨询,还涵盖网络支持小组、数字化个案管理、在线心理教育、远程督导等多元形式,形成了较为完整的服务链条。

人工智能技术的嵌入使得网络疗法能够实现智能匹配、自动随访、效果监测及资源链接等功能,提升了服务的系统性与连续性。例如,基于用户画像的智能匹配系统可将来访者与其适配的咨询师或干预程序进行精准对接;自然语言处理技术可自动分析咨询对话内容,提取关键主题与情感变化,生成咨询摘要与进展报告;智能随访系统可在咨询间隔期主动推送关怀信息,监测来访者状态变化,及时发现复发迹象。这些功能的实现,使得网络疗法在保持灵活性的同时,具备了接近传统面对面服务的专业深度。

3.3 特殊场景与人群的智能化服务

人工智能辅助心理咨询的应用已延伸至疗养康复、临床护理、教育辅导等特殊场景,展现出良好的场景适配性。杨泽鹏等^[15]探讨了疗养中心智慧化建设,指出基于健康大数据平台,利用自然语言处理技术从心理学文献中提取知识、构建心理咨询知识图谱,能够为疗养人员提供个性化的心理健康评估与干预方案。智慧疗养系统通过整合环境监测、生理指标采集与心理状态评估,实现了身心健康的协同管理,有助于提升疗养效果与生活质量。

在临床护理领域,聊天机器人干预已展现出积极的应用效果。刘婕等^[17]考察了微信聊天机器人干预对血液透析动静脉内瘘病人心理状态、治疗依从性及并发症的影响,该研究提示,针对特定慢性病患者设计的聊天机器人干预可能在心理支持、健康教育和治疗依从性方面发挥辅助作用,但其效果是否能够推广至其他疾病和人群,仍需进一步研究验证。这一研究表明,人工智能辅助心理咨询不仅适用于一般人群的心理促进,在慢性病患者的身心康复、疾病适应与生活质量提升中同样具有重要价值,为“生物—心理—社会”医学模式的落地提

供了技术支撑。

4 人工智能辅助心理咨询的伦理挑战与风险审视

人工智能辅助心理咨询在带来便利与效率的同时,也引发了一系列深刻的伦理问题。从学生心理危机干预中的隐私保护到人机情感关系的边界模糊,从元宇宙空间的新型伦理困境到伦理教育工具的滞后开发,伦理挑战贯穿于技术设计、服务提供与效果评估的全过程,亟需引起学界、产业界及监管部门的高度重视。

4.1 学生心理危机干预中的伦理问题

学生群体是心理健康服务的重点对象,也是人工智能心理服务渗透率最高的群体之一。数字技术在学生心理危机干预中的应用日益广泛,高校普遍建立的心理健康监测系统通过采集学生的学业数据、消费记录、网络行为、门禁出入等信息,构建心理危机预测模型,试图实现早发现、早干预。俞国良等^[16]探讨了数字技术赋能学生心理危机的应对策略,指出大数据、人工智能、虚拟现实等技术能够实现学生心理状态的实时监测、危机风险的智能预警及干预资源的精准匹配,在预防校园心理危机事件中发挥了积极作用。

然而,技术应用也带来了严峻的伦理挑战。首先,学生心理健康数据的采集涉及高度敏感的个人信息,如何平衡监测预警与隐私保护成为首要难题。大规模、持续性的行为监控可能构成对学生私人生活的过度侵入,引发“全景敞视”式的焦虑。其次,算法标签化可能导致对部分学生的歧视性对待,如果心理风险标签缺乏严格的访问权限控制和用途限制,理论上可能被不当用于与心理健康服务无关的评价或决策,从而产生标签化、污名化甚至歧视风险。再次,自动化危机干预决策可能忽视个体情境的特殊性,机械触发家长通知、强制就医等“一刀切”措施,反而加剧学生的心理压力与病耻感。最后,数据安全问题不容忽视,学生心理健康数据一旦泄露,将对其造成难以挽回的伤害。

4.2 人机情感关系与伦理边界

人工智能辅助心理咨询的一个核心伦理争议在于人机情感关系的性质与边界。刘伟等^[18]从消费者行为学视角综述了人与智能社交机器人的情感关系研究,发现用户在与社交机器人互动过程中会产生真实的情感依恋,这种依恋虽然能够提升用户参与

度与服务依从性,但也可能导致用户对机器人的过度依赖、对现实人际关系的疏离,以及对机器人“情感能力”的误判与过度理想化。在心理咨询场景中,来访者可能将聊天机器人视为具有真实情感的理解者,从而产生强烈的移情现象,但这种移情缺乏真实人际互动中的相互性、不确定性与成长性,可能阻碍来访者心理成熟的自然进程,过度拟人化可能增强用户的情感依赖,并弱化其主动寻求现实社会支持和专业心理服务的意愿。

元宇宙技术的引入进一步模糊了人机关系的伦理边界。刘燊等^[19]分析了“元宇宙+心理学”的发展新趋势,指出元宇宙空间中的虚拟化身、沉浸式交互、虚实融合为心理咨询提供了全新的场景可能,如来访者可在虚拟安全空间中重现创伤情境、练习社交技能、进行暴露治疗。然而,元宇宙中的身份匿名性、虚拟行为的伦理约束力弱化、数字成瘾风险、虚拟财产归属等问题,对心理咨询的专业伦理提出了全新挑战。咨询师在元宇宙空间中的专业身份认定、虚拟互动中的伦理责任界定、虚拟伤害的法律定性等问题尚缺乏明确规范,存在较大的伦理与法律灰色地带。

4.3 伦理教育与发展态势中的伦理反思

面对人工智能辅助心理咨询的伦理挑战,伦理教育的滞后性不容忽视。当前,人工智能心理服务的开发者多具有计算机科学背景,缺乏心理学伦理训练;而心理学从业者对人工智能技术的伦理风险认知不足,双方在伦理对话中存在知识壁垒。王佑镁等^[12]设计了 Ethical-GPT 这一面向人工智能伦理教育的垂类 GPT 工具,旨在通过案例教学、伦理决策模拟、多视角辩论等方式,提升开发者与使用者的人工智能伦理素养。该工具的设计思路提示,人工智能辅助心理咨询的伦理治理不仅需要外部规制,更需要内化于技术开发者、服务提供者及用户心中的伦理意识与道德敏感性。

从宏观发展态势来看,数字健康产业的快速扩张使得伦理规制面临“追赶式”困境。李丹丹等^[20]分析了 2022 年数字健康发展态势,指出人工智能、大数据、物联网、区块链等技术在数字健康领域的深度融合,推动了心理健康服务的智能化升级与产业变革,但相应的伦理审查机制、数据安全标准、服务质量监管体系、从业人员资质认证制度尚未健全。在资本驱动与技术竞赛的双重压力下,部分人工智

能心理服务产品存在过度商业化、疗效夸大、隐私保护不足、缺乏循证基础等问题,亟需建立覆盖产品研发、市场准入、服务提供、效果评估全链条的伦理治理框架。

5 讨论与展望

基于前文对人工智能辅助心理咨询技术发展、应用场景及伦理挑战的系统梳理,本章从国际伦理治理框架、人机替代关系、社会信息学视角、国际政策比较及质量评价体系五个维度展开综合讨论,以期为该领域的可持续发展提供理论参考与实践指引。

5.1 国际视野下的伦理治理框架

人工智能辅助心理咨询的伦理治理已成为全球关注的焦点,不同国家和地区基于各自的文化传统、法律体系与产业基础,形成了差异化的治理路径。Fiske 等^[21]系统分析了实体化人工智能在精神病学、心理学及心理治疗中的伦理影响,指出人工智能代理的虚拟或物理存在形式对医患关系、知情同意、专业责任、数据主权等核心伦理原则产生了深远冲击。研究强调,人工智能心理服务的设计与部署必须遵循尊重自主性、不伤害、有利及公正等基本伦理原则,同时应特别关注人工智能代理的“拟人化”设计可能带来的情感操纵风险与信任滥用问题。当用户将人工智能视为具有道德地位的存在时,其决策可能受到不当影响,这种“算法劝导”在心理健康领域尤为危险。

Saeidnia 等^[22]进一步探讨了人工智能心理健康干预的伦理考量,提出负责任实施的人工智能心理健康服务应确保数据隐私保护、算法透明度、文化敏感性及人类专业人员的必要监督。研究指出,当前许多人工智能心理服务产品在数据收集环节缺乏充分的知情同意,在算法决策环节缺乏可解释性,在服务提供环节缺乏人类专业人员的兜底监督,这些漏洞可能导致用户权益受损、治疗效果不佳甚至危机事件恶化。建立国际通用的伦理标准与认证体系,推动人工智能心理服务从“野蛮生长”走向“规范发展”,是当务之急。

从伦理学理论视角来看,关怀伦理为人工智能心理健康治理提供了独特的分析框架。Tavory^[23]提出将关怀伦理应用于人工智能心理健康规制,强调技术设计应关注脆弱群体的特殊需求,服务提供应体现对个体情感、文化背景与生活情境的敏感性,制度安排应确保技术红利向弱势群体倾斜,避免

“数字红利”演变为“数字鸿沟”。与传统以权利为基础的伦理框架不同,关怀伦理更强调关系性、情境性与责任性,这对于以“关系”为核心、以“共情”为基础的心理咨询实践具有特别的契合性。在人工智能辅助心理咨询中,关怀伦理要求技术不仅提供标准化的干预方案,更要能够识别用户的独特处境,在关键时刻提供温暖、尊重与希望。

5.2 AI 能否替代人类心理治疗师

关于人工智能能否替代人类心理治疗师的争论,折射出对该技术定位的深层分歧,也关系到整个心理健康服务行业的未来走向。Zhang 等^[25]系统探讨了人工智能整合到心理治疗中的前景与局限,指出人工智能在提供标准化干预、实现大规模服务覆盖、降低求助成本、消除地域壁垒、保护用户匿名性等方面具有不可替代的优势,但在处理复杂情感问题、进行深度人格改变、建立真实治疗联盟、应对存在性议题、处理伦理困境等方面仍存在本质局限。心理治疗不仅是技术的应用,更是两个人之间深刻的相遇与理解,这种“主体间性”的特质难以被算法完全模拟。

因此,国际研究共识倾向于将人工智能定位为人类心理治疗师的辅助者而非替代者,“人机协同”模式被认为是兼顾服务效率与专业伦理的可行路径。在这一模式中,人工智能承担初步筛查、心理教育、日常监测、危机预警、资料整理等基础性工作,释放人类咨询师的时间与精力,使其能够专注于需要高度专业判断与情感投入的复杂案例;人类咨询师则负责对人工智能的输出进行审核把关,处理超出算法能力范围的复杂情境,为来访者提供最终的专业保障。这种分工协作既发挥了技术的规模优势,又保留了人类的专业温度。

5.3 社会信息学视角与信任构建

从社会信息学视角审视,人工智能辅助心理咨询的可持续发展离不开信任机制的构建,而信任的建立是一个涉及技术、制度、文化多层面的系统工程。Gamble^[24]分析了人工智能移动应用在心理健康领域应用的社会信息学问题,指出用户对人工智能心理服务的信任不仅取决于技术性能与干预效果,还受到服务透明度、数据安全感、品牌声誉、社会文化因素及同伴推荐等多重因素的综合影响。用户对算法“黑箱”的不信任、对数据泄露的担忧、对“机器不懂人心”的偏见,都可能成为采纳人工智能心

理服务的心理障碍。

建立可信赖的人工智能心理服务生态,需要技术开发者、服务提供者、监管机构、专业协会及用户群体的协同参与。技术层面,应推动可解释人工智能的发展,使用户能够理解算法决策的基本逻辑;制度层面,应建立第三方审计机制、用户投诉救济渠道及数据泄露应急响应预案;文化层面,应开展公众科普教育,纠正对人工智能心理服务的过度恐惧或过度期待,培育理性的技术认知。只有通过多维度的信任积累,人工智能辅助心理咨询才能获得广泛的社会接受度。

5.4 中外人工智能心理服务监管政策比较

不同国家和地区在人工智能心理服务的监管政策上呈现出明显的路径差异,这些差异反映了各自在技术创新、患者保护与市场发展之间的不同权衡。美国采取相对宽松的“创新友好型”监管策略,美国食品药品监督管理局通过“软件作为医疗器械”框架对数字疗法进行分级管理,部分人工智能心理服务产品可通过 510(k) 通道快速上市,但近年来也加强了对算法偏见与疗效宣称的监管审查。欧盟《人工智能法案》采用基于风险的监管框架,对于特定高风险用途的人工智能系统提出数据治理、风险管理、透明度、人类监督等合规要求。对于涉及医疗、安全及基本权利的人工智能应用,应结合具体用途、产品属性及适用法规判断其风险等级,而不能简单将所有心理健康人工智能系统归入高风险类别。

我国的人工智能心理服务监管正处于快速完善阶段。国家药品监督管理局已将部分数字疗法纳入医疗器械管理,国家互联网信息办公室发布的《生成式人工智能服务管理暂行办法》对人工智能服务提供者提出了明确的合规要求。然而,专门针对人工智能心理服务的监管细则尚不完善,存在标准缺失、审批周期长、跨部门协调难等问题。借鉴国际经验,我国应加快建立分类分级管理制度,对低风险的心理健康教育类应用实行备案制,对高风险的心理疾病诊疗类应用实行严格的临床试验与审批制度,同时建立动态监测与退出机制,确保监管的科学性与灵活性。

5.5 人工智能辅助心理咨询的质量评价体系构建

质量评价是保障人工智能辅助心理咨询安全性

与有效性的关键环节，但现有的质量评价体系主要借鉴传统心理咨询或医疗器械的评估框架，未能充分反映人工智能心理服务的特殊性。构建科学的质量评价体系，应涵盖以下维度：一是技术性能指标，包括自然语言理解的准确率、风险评估的灵敏度与特异度、系统响应的及时性与稳定性；二是临床效果指标，包括症状改善程度、功能恢复水平、用户满意度、治疗依从性及长期复发率；三是伦理合规指标，包括知情同意的充分性、数据保护的有效性、算法偏见的检测与修正情况、人类监督的落实程度；四是服务可及性指标，包括服务覆盖人群、地域分布、价格可负担性及数字素养支持。

在评价方法上，应突破传统随机对照试验的局限，采用适应性试验、真实世界研究、用户生成内容分析等多元方法，捕捉人工智能心理服务在复杂现实环境中的表现。同时，应建立独立的第三方评价机构，避免“既当运动员又当裁判员”的利益冲突。质量评价结果应向社会公开，为用户提供知情选择的依据，也为监管部门的市场准入决策提供科学支撑。

6 结论

本综述系统梳理了 2019 年至 2024 年间国内外关于人工智能辅助心理咨询的研究成果，从技术架构、应用场景及伦理挑战三个维度进行了全面分析，得出以下主要结论。

第一，人工智能辅助心理咨询已形成较为成熟的技术架构。以自然语言处理、机器学习、大语言模型为底层支撑，以聊天机器人、数字疗法、智能评估系统为主要载体，人工智能辅助心理咨询实现了从心理筛查、风险评估到干预实施、效果监测的全流程覆盖。聊天机器人通过关系线索与拟人化设计，能够在一定程度上建立数字治疗联盟；数字疗法基于循证医学证据，为抑郁症、焦虑症等精神疾病提供了结构化、个性化的干预方案；智能评估系统借助多模态数据分析，突破了传统自陈量表的主观局限，实现了心理健康状态的动态监测与早期预警。这些技术进展显著提升了心理健康服务的可及性与效率，为缓解全球心理健康服务资源短缺问题提供了可行路径。

第二，人工智能辅助心理咨询的应用场景正从通用心理健康促进向垂直细分领域深度拓展。生成式人工智能与大语言模型的出现，使得人工智能心理咨询从简单的问答式交互升级为开放式、情境化

的深度对话，服务覆盖面显著扩大。网络疗法与社会工作服务的数字转向，推动了心理健康服务模式的结构变革，打破了传统咨询的时空壁垒。在疗养康复、临床护理、学生心理危机干预等特殊场景中，人工智能辅助心理咨询展现出独特的适配优势，能够为慢性病患群体、青少年群体等提供持续性、低成本的心理支持。这些应用实践表明，人工智能辅助心理咨询已从概念验证阶段迈向规模化落地阶段。

第三，伦理挑战是制约人工智能辅助心理咨询可持续发展的核心瓶颈。数据隐私泄露风险、算法偏见与歧视、人机情感关系的边界模糊、专业责任归属不清等问题，贯穿于技术设计、服务提供与效果评估的全过程。特别是在学生心理危机干预等敏感场景中，大规模心理健康数据采集与智能预警可能引发标签化、污名化等次生风险；元宇宙等新型交互空间的出现，进一步模糊了虚拟与现实、工具与主体之间的伦理边界。现有伦理规制体系呈现出明显的滞后性，技术创新的速度远超制度建设的步伐，导致部分人工智能心理服务产品存在过度商业化、疗效夸大、隐私保护不足等问题。

第四，国际研究共识倾向于将人工智能定位为人类心理治疗师的辅助者而非替代者，“人机协同”是兼顾服务效率与专业伦理的可行路径。人工智能在标准化干预、大规模服务覆盖、即时响应等方面具有不可替代的优势，但在处理复杂情感问题、进行深度人格改变、建立真实治疗联盟等方面仍存在本质局限。关怀伦理为人工智能心理健康治理提供了独特的分析视角，强调技术设计应关注脆弱群体的特殊需求，服务提供应体现对个体情感与情境的敏感性。社会信任机制的构建是人工智能辅助心理咨询可持续发展的社会基础，需要技术开发者、服务提供者、监管机构及用户群体的协同参与。

展望未来，人工智能辅助心理咨询的发展需要在技术创新与伦理规制之间寻求动态平衡。在研究层面，应开展更多大样本、长周期的随机对照试验，系统评估人工智能辅助心理咨询的疗效、安全性及长期影响，特别关注不同文化背景、年龄阶段、疾病类型人群的差异化反应。在制度建设层面，应加快建立人工智能心理服务的准入标准、质量规范、责任认定规则及伦理审查机制，借鉴国际经验构建分类分级管理制度，推动形成政府监管、行业自律、社

会监督相结合的多元治理格局。在技术开发层面,应着力提升算法的可解释性与透明度,建立数据安全与隐私保护的技术屏障,发展可解释人工智能与公平性评估工具,避免算法黑箱带来的信任危机。在人才培养层面,应加强心理学、计算机科学、伦理学、法学等学科的交叉融合,培养既懂心理服务又懂技术伦理的复合型人才,填补当前的人才缺口。在服务普及层面,应关注技术普惠性,防止“数字鸿沟”加剧心理健康服务的不平等,确保农村、偏远地区及低收入群体能够公平享有技术红利,真正实现心理健康服务的全民覆盖。

总之,人工智能辅助心理咨询正处于快速发展与深度调整并存的战略机遇期。唯有在拥抱技术变革的同时坚守伦理底线,在提升服务效率的同时保障专业质量,在拓展应用场景的同时维护用户权益,人工智能辅助心理咨询才能真正成为促进公众心理健康、增进社会福祉的积极力量,为构建全民心理健康服务体系贡献持久价值。

参考文献

- [1] 靳宇倡, 张政, 郑佩璇, 安俊秀. 远程心理健康服务: 应用、优势及挑战[J]. 心理科学进展, 2022, 30(1): 141-156.
- [2] 任萍, 汪悦, 刘冬予, 李添, 骆方. 心理健康评估与干预的智能化应用[J]. 北京师范大学学报(社会科学版), 2022(4): 150-160.
- [3] 李佳, 符仲芳, 田东华, 屈智勇. 数字化干预在心理健康领域的发展与应用[J]. 北京师范大学学报(社会科学版), 2023(6): 127-140.
- [4] 刘冬予, 骆方, 屠焯然, 饶思敬, 沈阳. 人工智能技术赋能心理学发展的现状与挑战[J]. 北京师范大学学报(自然科学版), 2024, 60(1): 30-37.
- [5] 李敬荣, 赵然, 张玉. 人工智能心理咨询的发展与应用[J]. 心理技术与应用, 2022, 10(5): 296-306.
- [6] 磨然, 方作之, 方建东. 如何建立聊天机器人与用户间的数字治疗联盟: 关系线索的作用[J]. 心理科学进展, 2023, 31(4): 669-683.
- [7] 磨然, 方建东, 常保瑞. 从“拟人归因”到“联盟建立”: 人与聊天机器人关系对参与度的影响[J]. 心理科学进展, 2023, 31(9): 1742-1755.
- [8] 张正, 蒋正乾, 刘佳丽, 曹霞, 周建松. 数字疗法在精神疾病诊疗中的发展[J]. 实用医学杂志, 2024, 40(11).
- [9] 姚峰, 王雪, 韦正德, 张效初. 人工智能在心理评估中的研究进展[J]. 科技导报, 2024, 42(10).
- [10] 石晓飞, 范鹏飞. 智能辅助诊断在心理健康评估系统中的应用与进展[J]. 中国医疗器械杂志, 2023, 47(5).
- [11] 张熙, 杨小汕, 徐常胜. ChatGPT 及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(5): 743-750.
- [12] 王佑镁, 王京京, 张思宇, 柳晨晨. Ethical-GPT: 一种面向人工智能伦理教育的垂类 GPT 工具设计与实现[J]. 现代教育技术, 2024, 34(11): 27-35.
- [13] 郑琰莉, 韩福海, 李舒玉, 苏文星. 人工智能大模型在医疗领域的应用现状与前景展望[J]. 医学信息学杂志, 2024, 45(6).
- [14] 刘斌志, 赵茜. 网络疗法: 社会工作服务的数字转向[J]. 西华大学学报(哲学社会科学版), 2021, 40(5): 67-75.
- [15] 杨泽鹏, 李宾, 曾婧, 李萍. 疗养中心智慧化建设的研究与思考[J]. 中国疗养医学, 2024, 33(12).
- [16] 俞国良, 张哲. 数字技术赋能学生心理危机的应对[J]. 清华大学教育研究, 2024, 45(1).
- [17] 刘婕, 赵淑丽, 张群. 微信聊天机器人干预对血液透析动静脉内瘘病人心理状态、治疗依从性及并发症的影响[J]. 蚌埠医科大学学报, 2023, 48(8): 1145-1149.
- [18] 刘伟, 董悦, 李纯青. 机器人并不“冰冷”: 消费者与智能社交机器人的情感关系研究综述[J]. 外国经济与管理, 2024, 46(6).
- [19] 刘燊, 柏江竹. 虚实相融的“元宇宙+心理学”发展新趋势[J]. 心理技术与应用, 2022, 10(8).
- [20] 李丹丹, 陈大明, 刘樱霞, 杨露, 李祯祺. 2022年数字健康发展态势[J]. 生命科学, 2023, 35(2).
- [21] FISKE A, HENNINGSEN P, BUYX A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy[J]. Journal of Medical Internet Research, 2019, 21(5): e13216.
- [22] SAEIDNIA H R, HASHEMI FOTAMI S G, LUND B, GHIASI N. Ethical considerations in artificial intelligence

- interventions for mental health and well-being: ensuring responsible implementation and impact[J]. *Social Sciences*, 2024, 13(7): 381.
- [23] TAVORY T. Regulating AI in mental health: ethics of care perspective[J]. *JMIR Mental Health*, 2024, 11: e58493.
- [24] GAMBLE A. Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective[J]. *Aslib Journal of Information Management*, 2020, 72(4): 509-523.
- [25] ZHANG Z, WANG J. Can AI replace psychotherapists? Exploring the future of mental health care[J]. *Frontiers in Psychiatry*, 2024, 15: 1444382.

版权声明：©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS