

## 人工智能背景下的心理危机行为分析及干预

郜逸博

广西师范大学教育学部心理学系, 桂林

**摘要** | 随着全球心理健康问题的日益严峻, 传统心理危机干预手段面临资源有限、可及性差和实时性不足等挑战。人工智能技术的迅猛发展, 特别是自然语言处理、机器学习和多模态数据分析等领域的突破, 为心理危机的识别、分析与干预带来了革命性的机遇。本综述旨在系统梳理人工智能在心理危机行为分析及干预中的应用现状, 探讨其核心技术、典型应用场景, 并深入分析当前面临的伦理、技术和临床挑战, 最后对未来发展方向进行展望。

**关键词** | 人工智能; 心理危机; 行为分析; 数字干预

Copyright © 2025 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

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



### 1 引言

心理危机, 通常指个体在遭遇重大生活应激事件或内在心理冲突时, 因其惯用的应对机制失效而陷入的一种严重的心理失衡状态 (Caplan, 1964)。这种状态不仅伴随着极度的痛苦、混乱和功能失调, 更是自伤、自杀等高风险行为的关键前兆。传统的心理危机干预模式高度依赖于专业咨询师、精神科医生等人工的即时识别、评估与介入。然而, 这一模式正面临着严峻的全球性挑战。一方面, 心理健康问题的普遍性日益攀升。世界卫生组织 (WHO) 数据显示, 全球有数亿人受到精神障碍的困扰, 其中抑郁症是导致残疾的主要原因之一, 而自杀已成为全球主要死因之一 (World Health Organization, 2021)。另一方面, 精神卫生资源存在着巨大的缺口, 尤其是在中低收入国家, 专业人力严重不足、分布不均, 且心理污名化现象普遍存在, 导致大量有需求的个体无法获得及时、有效的帮助 (Thornicroft et

al., 2022)。

这一供需之间的巨大矛盾, 催生了对创新解决方案的迫切需求, 数字心理健康 (Digital Mental Health) 应运而生。在此背景下, 人工智能技术的迅猛发展, 特别是机器学习、自然语言处理和大数据分析技术的成熟, 为解决上述瓶颈提供了前所未有的机遇 (Torous et al., 2021)。AI不再仅仅是一个辅助工具, 它正从根本上重塑着心理危机的“监测—预警—分析—干预”全链条, 推动该领域从被动响应向主动预防、从普适化向个性化范式转变。

早期研究已经证明了AI在识别心理危机信号方面的可行性。例如, 通过分析社交媒体文本, 自然语言处理模型能够以高于随机水平的准确度识别出包含自杀意念的帖子 (De Choudhury et al., 2016)。这些研究为利用数字足迹进行大规模心理健康筛查奠定了基础。与此同时, 基于人工智能的聊天机器人, 如Woebot (Fitzpatrick et al., 2017) 和Wysa, 已通过随机对照试验证明, 其能

作者简介: 郜逸博, 广西师范大学教育学部心理学研究生, 研究方向: 社会心理学。

文章引用: 郜逸博. (2025). 人工智能背景下的心理危机行为分析及干预. *中国心理学前沿*, 7(11), 1353–1357.

<https://doi.org/10.35534/pc.0711221>

够通过传递认知行为疗法原则,有效减轻用户的抑郁和焦虑症状,展现了AI作为干预媒介的潜力。

因此,本综述旨在系统性地梳理和整合人工智能在心理危机行为分析与干预领域的最新进展。我们将首先阐述其背后的关键技术原理与应用场景,继而分析人工智能驱动的创新干预模式,并深入探讨该领域在伦理、准确性和有效性方面面临的核心挑战与争议,最后对未来发展方向进行展望,以期研究者、临床工作者和政策制定者提供一个全面而深刻的视角,共同推动这一跨学科领域的稳健与负责任发展。

## 2 AI在心理危机行为分析中的关键技术与应用

### 2.1 基于自然语言处理(NLP)的分析

语言作为洞察个体内心状态最直接且丰富的窗口,为心理危机分析提供了关键路径。自然语言处理技术通过理解、解释和量化人类语言中的语义与情感特征,使得从文本和语音中自动识别危机信号成为可能。在文本分析领域,早期研究主要通过词典方法和基础自然语言处理特征来识别心理风险,例如分析社交媒体帖子中第一人称单数代词、负面情感词汇以及绝对化词汇的密度,这些被证明与抑郁和焦虑倾向显著相关(Tausczik & Pennebaker, 2010)。随着深度学习的发展,诸如BERT等预训练语言模型能够更深入地捕捉文本的深层语境与语义,不仅能识别明确的风险词汇,还能有效探测隐喻、告别意图等隐晦表达,从而以高准确度区分含有自杀意念的社交媒体内容(Roy et al., 2020)。此外,NLP的动态监测能力至关重要,通过分析用户语言风格从情绪多变转向持续低迷、或从社交活跃陷入沉默的演变轨迹,可以捕捉到强烈的危机前兆(De Choudhury et al., 2016)。

与此同时,语音分析作为另一关键渠道,通过解析副语言信息来评估个体的情感与生理状态。通过分析通话或录音中的声学特征,如语速减慢、音调范围变窄、音量单调、声音颤抖,以及停顿频率异常等,人工智能能够提取出与心理状态密切相关的客观声学标记。大量临床研究证实,抑郁症患者的语音往往表现出与“精神运动性迟缓”相对应的特征,例如语速显著减慢、反应延迟增加和基频方差减小(Cummins et al., 2015)。这些声学特征因此可作为非侵入性的生物标志物,为远程情绪状态评估和危机风险预警提供有力的辅助手段。

### 2.2 基于计算机视觉的分析

基于计算机视觉的心理危机分析主要通过解析面部表情、眼动模式、肢体姿态等视觉线索来评估个体的心理状态,这种方法为实现非接触式、被动式的心理健康监测提供了可能。在面部表情分析方面,研究表明面部肌肉活动与内心情绪状态之间存在密切联系,抑郁等心

理状态会导致特定的面部特征变化,如嘴角下垂、眉头上扬、笑容减少且持续时间缩短等。为了克服抑郁症缺乏客观生物标志物的临床挑战,一项研究构建了包含患者与正常人群在多情绪状态下的面部数据集,并开发了融合双尺度卷积与自适应通道注意力机制的深度学习模型,该模型不仅能以更高准确度区分抑郁状态,还通过梯度类激活映射技术,可视化出了抑郁症患者区别于常人的特异性面部视觉模式,为诊断提供了可解释的视觉证据(Li Wang et al., 2024a)。例如,一项开创性工作开发了一套集成设备,通过采集情感面部表情视频和情感瞳孔波信号,并利用3DCNN+LSTM及多尺度1DCNN模型分别提取其特征,实现了对汉密尔顿抑郁量表(HAMD)和患者健康问卷(PHQ-9)的自动评估。该研究不仅显著提升了评估精度,还通过对比发现:基于面部表情视频的评估效果优于瞳孔波信号,且对医学量表HAMD的预测精度显著高于自评量表PHQ-9(Li et al., 2024b)。通过深度学习模型对视频中的面部动作单元进行精确编码和分析,可以量化这些微妙的表情变化,Girard等(2013)的研究发现,抑郁症患者在社交互动中特定面部动作单元的强度和使用频率与健康对照组存在显著差异,这些特征可作为辅助诊断的客观指标。此外,肢体姿态和行为分析也提供了重要信息,心理困扰常表现为身体活动的减少、姿势变化的频率降低、手势活动减少以及整体运动的缓慢化,这些运动特征可通过基于视频的姿态估计算法进行提取和分析。值得注意的是,尽管计算机视觉技术前景广阔,但其应用仍面临光照条件、头部姿态、个体差异以及隐私保护等挑战,需要在技术开发和实际应用中加以充分考虑。

### 2.3 基于多模态数据融合与机器学习/深度学习的预测模型

心理危机的表现形式具有高度的复杂性和个体差异性,单一模态的数据往往只能揭示其某个侧面,存在信息盲区与误判风险。为了构建更为稳健、精准的预警系统,研究前沿已从单一模态分析转向多模态数据融合,即协同整合来自行为、语言、视觉、生理及临床等多个来源的异构数据,并利用机器学习与深度学习模型来捕捉其中蕴含的复杂模式与非线性的相互作用,从而实现心理危机的早期、全面洞察。

多模态融合的技术路径主要分为三种层级。在特征级融合中,模型首先从各模态数据(如从文本中提取的情感向量、从面部视频中提取的动作单元强度、从可穿戴设备中提取的心率变异性特征)中分别提取高维特征,随后在输入预测模型前将这些特征向量进行拼接或对齐,形成一个统一的表征。这种方法能够保留原始信息的丰富性,但面临不同模态特征在时间尺度与量纲上不一致的对齐挑战。在决策级融合中,每个模态会先由一个独立的子模型(例如,一个模型处理语音,另一个处理文本)进行处理并生成初步的风险概率或分类结

果,最后通过加权平均、投票或元学习器等策略整合各子模型的输出。这种策略灵活性高且易于实现,但可能损失模态间的底层交互信息。目前,最先进的研究倾向于采用模型级融合,尤其是基于注意力机制的深度学习架构,如多模态变换器。这类模型能够动态地学习不同模态在特定上下文中的重要程度(注意力权重),并模拟它们在时序上的相互关联,例如,模型可以学习到当一个人在语音上表现出颤抖(声学模态)的同时,其文本中出现了“绝望”一词(文本模态),此时的危机风险权重应被显著提高。

实证研究已证明了多模态融合的显著优势。例如,一项针对抑郁症评估的研究表明,融合了面部表情视频与语音声学特征的端到端模型,其评估结果与临床医生基于HAMD量表的判断具有更高的一致性(Yang et al., 2021)。这些发现强有力地说明,多模态数据融合能够通过信息互补,更全面地勾勒出个体的心理状态图谱,是推动人工智能心理危机分析从实验室走向实际临床应用的关键技术路径。

### 3 人工智能在心理危机干预中的创新模式

#### 3.1 人工智能驱动的聊天机器人与对话系统

人工智能驱动的聊天机器人与对话系统代表了当前数字心理健康领域最显著的应用之一,它们通过模拟人类对话,为面临心理危机的个体提供可及、即时且低门槛的初步支持。这些系统通常基于自然语言处理技术,并整合了循证心理干预原则,旨在缓解即时情绪困扰、提供心理教育,并引导用户发展健康的应对技能。

其核心运作模式是构建一个安全、非评判性的对话空间,使用户能够随时表达自身的痛苦与压力。早期的代表性系统,如Woebot,其交互内容根植于认知行为疗法原则。它通过结构化的对话引导用户识别并挑战扭曲的认知,例如“全有或全无”思维,并鼓励其参与愉悦活动以改善情绪。一项随机对照试验表明,与仅使用信息手册的对照组相比,与Woebot交互的年轻成年人在抑郁症状方面表现出显著的减轻(Fitzpatrick et al., 2017)。类似地,Wysa和Tess等聊天机器人则整合了正念练习、情绪记录和问题解决等多种技术,以适应不同用户的需求偏好(Vaidyam et al., 2019)。一项范围综述指出,使用者普遍认为心理健康聊天机器人具有便利性和可及性的优点,许多人对其非评判性和随时可用的特性表示赞赏(Abd-Alrazaq et al., 2021),这为那些因污名化、经济限制或地理不便而无法获得服务的个体提供了关键的支持入口。

然而,现有系统仍面临一定的局限性。当前的对话机器人大多依赖于预编写的脚本和有限范围内的响应生成,在理解高度复杂、新颖或充满隐喻的用户表达时

可能出现偏差。其共情回应本质上是程序化的模拟,尚无法替代真实人类治疗师所提供的情感联结与深度共情(Vaidyam et al., 2019)。因此,未来的发展方向将集中于提升对话系统的上下文理解与生成能力,并探索其与专业人工督导相结合的人机协同干预模式,以确保在提供普惠服务的同时,能有效识别并转介高风险个案。

#### 3.2 个性化干预方案推荐

人工智能通过整合多源数据与算法模型,正在推动心理危机干预从标准化向个性化范式转变。这类系统通过分析用户的数字足迹(如社交媒体文本、可穿戴设备数据)、临床记录及实时情绪报告,构建动态的个人心理画像——这一过程被称为个人感知,它为理解心理健康提供了新的维度(Mohr et al., 2017)。基于此画像,机器学习算法可实现多层次精准匹配:根据问题类型推荐特定的认知行为疗法模块,依据严重程度调整干预强度,并参照用户偏好(例如对视频、音频或文字形式的倾向)匹配合适的媒介形式。研究显示,医疗提供者对在实践中有针对性地使用此类工具表现出浓厚兴趣(Schueller et al., 2016)。

这种个性化推荐显著提升了干预的精准性与参与度。其中,适时自适应干预(JITIs)作为前沿方向,代表了这一领域的核心进展。适时自适应干预能够根据实时情境数据(如地理位置、生理指标或自我报告的情绪状态)在个体最需要支持且最容易接受的时刻,推送高度个性化的干预策略(Nahum-Shani et al., 2018)。例如,系统在检测到用户情绪剧烈波动且处于社交孤立情境时,可能会主动启动安全计划或引导进行紧急的减压练习。

然而,该领域仍面临数据隐私、算法透明度及数字鸿沟等挑战。未来需开发更可靠的联邦学习框架,并在个性化推荐中加强临床专业知识与伦理监督,确保技术应用既高效又负责任。

### 4 挑战与局限性

尽管人工智能在心理危机分析与干预领域展现出巨大潜力,其发展仍面临技术有效性、伦理道德及临床实践等多层面的严峻挑战。在数据隐私与算法公平性方面,心理健康数据作为最敏感的个人信息,在其收集、存储与分析过程中存在显著的泄露风险;同时,训练数据的偏差可能导致算法歧视,例如对少数族裔或特定性别群体的风险评估不准确,若这种系统性偏差未被纠正,可能加剧现有的医疗不平等。在技术可靠性方面,现有模型的“黑箱”特性使得临床工作者难以理解其决策逻辑,显著降低了系统的可信度;此外,模型在真实场景中的泛化能力存疑,当应用于与训练数据差异较大的新群体时,其性能往往显著下降。在临床整合层面,缺乏明确的监管框架和责任界定使得AI工具在出现误判

时的责任归属成为难题；更重要的是，当前系统在危机情境下缺乏真正的情感共鸣，过度依赖技术手段可能削弱重要的人际支持。这些相互交织的挑战共同构成了AI在心理健康领域深入应用必须克服的障碍，未来的发展需要在技术进步、伦理规范和临床需求之间寻求更加谨慎的平衡。

## 5 未来展望

人工智能在心理危机干预领域的未来发展将呈现出更加集成化、精准化和人性化的趋势。在技术层面，基于多模态数据融合的动态风险评估系统将实现从静态评估向动态轨迹预测的重要转变，通过整合文本、语音、视觉、生理信号等多维数据，构建个体的动态心理状态画像，为早期预警和预防性干预提供可能。在应用模式上，深度融合的人机协同干预系统将成为主流发展方向，形成人工智能监测预警—医生决策干预的闭环 workflow，既发挥人工智能在持续监测和数据处理方面的优势，又保留专业医生在复杂情况判断和情感支持中的核心作用。同时，可解释人工智能（XAI）与算法公平性的提升将成为推动临床采纳的关键，通过开发能够提供透明决策依据的模型，并采用去偏差算法和多样化训练数据等策略，建立临床信任并确保系统在不同群体中的公平性。在应用范围上，人工智能系统将从危机干预向前延伸至全域心理健康促进，通过日常心理状态监测和个性化干预，帮助用户建立积极的心理资本，实现从“治已病”到“治未病”的范式转变。此外，健全的伦理与法律框架将是可持续发展的基础，需要建立涵盖数据隐私、算法问责、临床责任等领域的完善法规体系。总之，未来AI在心理危机干预领域的发展将更加注重技术与人文的平衡，通过构建安全、有效、可信的系统，为促进全球心理健康提供有力支持。

## 参考文献

- [1] Abd-Alrazaq A A, Alajlani M, Ali N, Denecke K, Bewick B M & Househ M. (2021). Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. *Journal of Medical Internet Research*, 23(1), e17828.
- [2] Caplan G. (1964). *Principles of preventive psychiatry* (pp. xi, 304). Basic Books.
- [3] Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J & Quatieri T F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, (71), 10–49.
- [4] De Choudhury M, Kiciman E, Dredze M, Coppersmith G & Kumar M. (2016). Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. *CHI Conference*.
- [5] Fitzpatrick K K, Darcy A & Vierhile M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2), e19.
- [6] Girard J M, Cohn J F, Mahoor M H, Mavadati S & Rosenwald D P. (2013). Social Risk and Depression: Evidence from Manual and Automatic Facial Expression Analysis. Proceedings of the ... International Conference on Automatic Face and Gesture Recognition. *IEEE International Conference on Automatic Face & Gesture Recognition*, 1–8.
- [7] Li M, Lu Z, Cao Q, Gao J & Hu B. (2024). Automatic Assessment Method and Device for Depression Symptom Severity Based on Emotional Facial Expression and Pupil-Wave. *IEEE Transactions on Instrumentation and Measurement*, (73), 1–15.
- [8] Li M, Wang Y, Yang C, Lu Z & Chen J. (2024). Automatic Diagnosis of Depression Based on Facial Expression Information and Deep Convolutional Neural Network. *IEEE Transactions on Computational Social Systems*, 11(5), 5728–5739.
- [9] Mohr D C, Zhang M & Schueller S M. (2017). Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology*, (13), 23–47.
- [10] Nahum-Shani I, Smith S N, Spring B J, Collins L M, Witkiewitz K, Tewari A & Murphy S A. (2018). Just-in-Time Adaptive Interventions (JITAI) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, 52(6), 446–462.
- [11] Roy A, Nikolitch K, McGinn R, Jinah S, Klement W & Kaminsky Z A. (2020). A machine learning approach predicts future risk to suicidal ideation from social media data. *Npj Digital Medicine*, 3(1), 78.
- [12] Schueller S M, Washburn J J & Price M. (2016). Exploring mental health providers' interest in using web and mobile-based tools in their practices. *Internet Interventions*, (4), 145–151.
- [13] Tausczik Y R & Pennebaker J W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- [14] Thornicroft G, Sunkel C, Alikhon Aliev A, Baker S, Brohan E, El Chammy R, Davies K... Winkler P. (2022). The Lancet Commission on ending stigma and discrimination in mental

- health. *Lancet* (London, England), 400(10361), 1438–1480.
- [15] Torous J, Bucci S, Bell I H, Kessing L V, Faurholt-Jepsen M, Whelan P, Carvalho A F, Firth J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 20(3), 318–335.
- [16] Vaidyam A N, Wisniewski H, Halamka J D, Kashavan M S & Torous J B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Canadian Journal of Psychiatry Revue Canadienne De Psychiatrie*, 64(7), 456–464.
- [17] World Health Organization. (2021). WHO report highlights global shortfall in investment in mental health. from <https://www.who.int/news/item/08-10-2021-who-report-highlights-global-shortfall-in-investment-in-mental-health>.
- [18] Yang L, Jiang D & Sahli H. (2021). Integrating Deep and Shallow Models for Multi-Modal Depression Analysis—Hybrid Architectures. *IEEE Transactions on Affective Computing*, 12(1), 239–253.

## Analysis and Intervention of Psychological Crisis Behavior in the Context of Artificial Intelligence

Gao Yibo

*Department of Psychology, Faculty of Education, Guangxi Normal University, Guilin*

**Abstract:** As global mental health issues become increasingly severe, traditional methods of psychological crisis intervention face challenges such as limited resources, poor accessibility, and insufficient real-time responsiveness. The rapid development of artificial intelligence (AI), particularly breakthroughs in natural language processing, machine learning, and multimodal data analysis, has brought revolutionary opportunities for the identification, analysis, and intervention of psychological crises. This review aims to systematically outline the current applications of AI in the analysis and intervention of psychological crisis behavior, explore its core technologies and typical application scenarios, and provide an in-depth analysis of the ethical, technical, and clinical challenges currently faced. Finally, future development directions are discussed.

**Key words:** Artificial intelligence; Psychological crisis; Behavioral analysis; Digital intervention