心理科学进展 2026, Vol. 34, No. 1, 123–133 https://doi.org/10.3724/SP.J.1042.2026.0123
Advances in Psychological Science https://cstr.cn/32111.14.2026.0123

算法介导下的情感趋同：生成式人工智能情感传染机制*

武靖宇[1] 金 鑫[2]

([1]中国传媒大学新闻学院, 北京 100024) ([2]重庆师范大学新闻与传媒学院, 重庆 401331)

摘 要 本研究聚焦生成式人工智能(Artificial Intelligence Generated Content, AIGC)情感传染这一新兴课题, 系统探讨其与传统人际情感传染、数字情感传染的本质差异, 提出“扮演–调节”机制作为理论框架。研究发现AIGC 情感传染具有交互主体性、知识依附性、非威胁性与去身份化、道德性这四大特性, 共同支撑 AIGC 情感传染的“扮演–调节”机制; AIGC 通过算法模拟人类情感表达模式(即“扮演”), 同时根据用户实时反馈动态优化交互策略(即“调节”), 形成持续迭代的人机情感闭环。“扮演–调节”机制突破人类中心主义范式, 构建了跨主体情感理论、揭示了算法作为主动情感调节者的新角色, 已被应用于心理健康干预、传播学及教育激励等领域。AIGC 情感传染研究拓展了情感传染理论外延, 为理解人机情感交互提供了新视角, 但也面临多学科整合难题、多模态情感测量的复杂性、跨文化适配障碍, 以及算法偏见导致的情感误导风险等挑战。

关键词 生成式人工智能(AIGC), 情感传染, 人机交互

分类号 B842; B849: C91

1 问题提出

长期以来, 作为社会心理学的一种重要概念, 情感传染一直被认为是个人和集体行为的核心驱动力(Goldenberg & Gross, 2020), 在人类社会联结、群体凝聚、信任建立乃至决策制定中扮演着核心角色。早期对其研究主要聚焦于面对面交流及数字媒介中的情感传染, 旨在深入理解人类情感在个人社交互动和群体关系中的动态传播机制, 并借此探讨社会变迁与文化演进的深层次联系。然而, 自 ChatGPT 3.0 应用伊始, 生成式人工智能(Artificial Intelligence Generated Content, AIGC)在情感生成方面取得了显著进展。AIGC 一般指人工智能生成的内容(Artificial Intelligence Generated Content), 在实际应用和学术讨论中, 其含义常延伸至生成这些内容所依赖的底层技术如大型语言模型、扩散模型等算法, 以及由此催生的新型内容生产方式。本研究聚焦人机交互过程中的情感传染机制, 将 AIGC 界定为基于生成式人工智能技术构建的、能够与用户进行多轮对话并生成文本、语音、图像等多模态内容的智能体或平台, 如 DeepSeek、ChatGPT、文心一言等。这一进展突显了情感因素在当代人机交互与社会互动中扮演着越来越关键的角色。

相较于传统人际情感传染主要依赖生物学的模仿与反馈机制, AIGC 介导的情感交互在触发机制上存在本质差异。其情感表达的发出者(emitter)——情感刺激源(Emotional Stimulus Source)是经由算法驱动的非人类主体, 这意味着 AIGC 的情感传递并非源于真实的内在体验, 而是一种基于海量数据学习与模拟的程序化情感输出或预设的反馈流程, 其客观效果在于通过持续、个性化情感互动, 对用户的情感状态产生影响, 甚至在长期使用下可能引导用户形成某种情感依赖。AIGC 长期且深度的互动对人类产生的影响仍是一个需要神经科学与心理学交叉研究来审慎探讨的前沿课题。在理论层面, 本研究深化了情感传染理论的内涵与外延, 拓展了其在人机交互情境下的解释力, 还可能催生新的情感研究范式。在实践层面, 对 AIGC 情感传染规律的解析, 为内

收稿日期：2025-03-21

* 省部级基金“多模态情感传播视域下信息失序风险识别与网络戾气治理研究” (24YJA86008)、重庆市教委一般项目“数字时代图像新闻的情感传播与影响机制研究” (24SKGH062)资助。

通信作者：金鑫, E-mail: 32499074@qq.com

容设计与算法优化提供了科学依据，有助于预见并规范其可能带来的情感操纵等伦理风险，进而推动技术健康迭代并提升用户福祉。基于此，本研究将深入探索三个问题：

问题一：AIGC 介导的情感传染区别于人际情感传染与数字媒介情感传染的独特性是什么？

问题二：在这些独特性的协同作用下，AIGC 情感传染的核心机制是如何运作及应用的？

问题三：AIGC 情感传染机制在理论建构、实践验证与测量层面面临哪些关键挑战？其未来发展方向何在？

2 情感传染的范式演进

情感传染(Emotional Contagion)，即个体情感状态因感知他人情绪而趋同的过程(Parkinson, 2011)，长期以来被视为社会联结与群体行为的核心驱动力。根据媒介形态的演进，情感传染可分为三种范式：面对面人际传染、数字媒介传染与算法介导的 AIGC 传染。理解前两者的核心机制，是把握 AIGC 情感传染的基石。

2.1 传统情感传染的理论基石

传统情感传染(Emotional Contagion)研究起源于人与人之间面对面情感传递现象，通常被定义为个体倾向于自动模仿他 的面部表情、声音、姿势和动作，并因此在情绪上与他 趋同的过程(Hatfield et al., 1993)。这种原始情感传染被认为是相对自动且无意识的，其运作依赖于“模仿−反馈”的循环机制(Hatfield et al., 1993)。该机制的核心在于身体与生理活动的直接参与，更强调其生物学与社会学根源。如有研究发现易感性个体和女性个体在情感传染中表现出更高的敏感性和共鸣能力(Doherty, 1997)。传统的情感传染自发生之初就是一个主体间过程。然而，它依赖于物理空间的共时性和具身的直接参与，因此在数字化、智能化的交互环境中难以充分实现。

2.2 数字情感传染演进及借鉴意义

数字媒介的发展催生了数字情感传染范式，其核心是“激活−评价”机制，即个体通过解读文本、图像、表情符号等符号化线索，触发相应的情绪类别认知(Hill et al., 2010)，继而通过“社会参照” (Barsade, 2002)与“评价”过程，将他人的情绪作为情境信息的来源以调整自身情绪反应。

数字情感传染的“激活−评价”机制对理解 AIGC 情感传染具有重要借鉴意义：首先，数字情感传染突破了物理空间的限制，从传统的“人−人”互动转变为了“人−文本−人”[1]的新型交互关系，形成了无中心或全向性关联网络的跨时空情感传递(隋岩，杨超，2024)，这与 AIGC 情感传染不依赖物理共在的特性高度一致；其次，数字情感感染中的“社会参照”机制(Barsade, 2002)揭示了情感传递的社会认知基础，即情感之所以能够传递，是因为它提供了关于特定情境性质的重要信息。这一机制对理解 AIGC 如何通过模拟社会认知过程实现情感传染具有启发意义；第三，数字情感传染展现出的群聚传播效应，如随机性与动态性的聚合与裂变、算法互文性带来的变革与挑战等(隋岩，杨超，2024)，为理解 AIGC 情感传染在大规模人机交互中可能产生的集体效应提供了理论框架。

然而，无论是依赖具身模仿的“模仿−反馈”机制，还是依赖符号认知的“激活−评价”机制，其情感刺激源(emitter)均是具有真实情感体验的人类主体。情感传递本质上是人类心理机制的自然延伸。

2.3 AIGC 情感传染的界定

AIGC (生成式人工智能)的兴起则标志着一种全新范式的出现，它与传统情感传染区别在于这是一种算法介导的情感交互，其中一方(AIGC)并不具备真正的情感体验，其有效情感交互的实现依托于人工智能(AI)学科中新兴的一个重要研究领域——情感人工智能(Emotional Artificial Intelligence)，其目标是通过分析人类情感信号实现情感识别、解读与反馈(McStay, 2018)。依据 Picard (2000)的分类框架，情感计算涵盖从基础情感识别、中间情感表达到高级共情对话的渐进路径。其中，共情对话系统的出现则标志着人工智能的关键进步，它将人机交互从任务导向的交流转变为情感共鸣、情境感知的对话，从而显著提升用户参与度(Fung et al., 2018)。Yuvaraj 等

[1] “文本” (text)在符号学与传播学视域下指代任何承载意义、可供解读的符号集合与表意实践，广泛涵盖了用户生成或传播的文字、图像、静态图片、动态视频、音频、表情包(Emoji)、数字装扮、社交媒体账号身份标识乃至如点赞、转发等特定行为模式等多元化的符号形式，构成了数字空间中主体间交往的核心中介物。

(2025)学者也将共情对话视作 AI 突破人机交互瓶颈的核心方向之一, “具备多模态能力的会话式类人智能体处于大模型与生成式 AI 应用的最前沿……它们通过整合文本、语音、视频与生理信号, 实现对复杂用户环境的深度理解, 突破了现有大模型在共情表达上的局限, 从而提供真正情境感知且富有同理心的语言回应”。已有研究通过测量“我感觉这个 AI 在回复我时与我有相似的情绪”, “我感觉这个 AI 在回复我时能够理解我的需求”等问题来评估用户受情感传染的情况, 发现其对用户持续使用意愿有显著影响(Liu-Thompkins et al., 2022)。

基于此, 本研究将 AIGC 情感传染定义为：用户在与 AIGC 进行交互的过程中, 因感知和加工其生成的包括文本、图像、语音、视频等多模态内容中承载的显性或隐性情感线索, 从而引发自身情感状态发生与 AIGC 所表达或暗示的情感趋同或相关联的变化现象。

3　AIGC 情感传染的独特性

与以往自然发生的情感传染不同, AIGC 情感传染是从人类丰富的历史经验、记忆数据中学习而来, 并依据情感线索进行自适应调整与优化, 生成出近似人类甚至超越人类情感处理的潜能(Schiller et al., 2024)。这种潜能展现出 AIGC 情感传染的四种独特性(表 1)。

3.1　交互主体性

“交互主体性”指的是人机互动过程中, 因 AIGC 展现出高度的交互适配性、对话逻辑连贯性及拟态情感响应, 用户会在认知层面将其感知为具备认知与反思能力的对话伙伴, 进而生成一种主体间关系体验。

“主体性”通常被界定为具备认知主权、道德自主意识, 并能通过自我反思形成思想与行为统一性的存在终端(赵汀阳, 2023), 即“自我意识”是是否具有“主体性”的前提。在人机情感交互中, AIGC 虽展现出高度的对话适配性与情境响应能力, 但其核心机制是对大规模训练数据的统计模式学习, 缺乏构成“自我意识”的认知架构(Bender et al., 2021)。其学习过程完全遵循数据驱动的经验主义路径, 通过复杂的神经网络对历史文本数据进行统计建模和归纳, 从而生成看似理解情境的模式化响应(Achiam et al., 2023), 进而为人类用户建构一种类主体间的互动体验。换句话说, 它不再依赖主体间的直接认知共鸣, 而是通过数据中介的符号交互构建情感联结, 这一特征使其既区别于人类主体性的本真状态, 又不同于传统技术工具的纯粹客体属性, 在人际互动中呈现出“拟人性” (Durante et al., 2024)或“拟主体” (曹博林, 支冰洁, 2024)的外在特征。

由此可见, “交互主体性”所建构的情感联结, 是基于海量数据训练形成的一套程序化反馈, 令人机关系从“主−客” “自我−他者”的外在工具性关联, 向“主−主” “自我−类自我”的内在协作性关系转型。如当用户向 AIGC 倾诉自身烦恼时, AIGC 通过对文本中关键词的提取与情感倾向的算法识别, 生成理解与安慰的回应文本, 用户在接收此类反馈后, 会产生被理解的慰藉感, 其负面情绪可能得到一定程度的缓解, 倾诉欲望亦可能被进一步激发(Qiuting & Feng, 2023)。

3.2　知识的依附性

知识依附性是指 AIGC 情感理解与表达能力完全依附于其训练数据, 是一个纯粹的由数据驱动的主体, 其所有输出均是对人类集体经验的再现与重组。

AIGC 知识、语言模式乃至理解和表达情感的能力完全源于数据学习(Achiam et al., 2023)。这种知识依附性使得 AIGC 能够根据用户的实际需求调整其语言风格, 展现出应对个人沟通偏好的高度适应性(Haqqu & Rohmah, 2024)。具体而言, AIGC 能够通过情感线索的识别与适配(Liu-Thompkins et al., 2022), 准确捕捉用户情绪需求, 在知识层

表 1　情感传染的独特性

传染类型	交互主体性	知识依附性	身份与认同	道德性
面对面传染	具身主体−他者	先验−再现	社会身份认同	情境依赖性高
数字平台传染	文本主体−他者	先验+经验−再现	部分匿名化	符号化、碎片化
AIGC 传染	交互主体、跨主体−自我	经验−重组	非威胁性与去身份化	循环强化、积极性偏向

面与用户形成共情联结(Yuan et al., 2024)。同时, AIGC 依赖数据驱动的跨领域知识迁移与上下文重组(Achiam et al., 2023), 在经验层面生成符合用户情境的新颖意义阐释和可行行动方案(Gao et al., 2023)。这种以人类知识为根基的互动是 AIGC 独特优势的来源, 但也潜藏因数据偏差导致情感误导的风险。如有研究表明如果训练数据中存在偏见, AIGC 可能会复制甚至放大这些偏见, 导致其情感表达出现偏差或不当(Gao et al., 2023; Wei et al., 2023)。

3.3 非威胁性与去身份化

非威胁性是指 AIGC 作为非人类主体, 不具备真实社会身份、个人情感偏见或独立利益诉求, 从而为用户创造了一个无社会评价压力的安全交互环境。去身份化是指 AIGC 的情感连接不依赖于任何预设的社会身份标签, 能直接通过内容与交互触发情感共鸣。

在心理学视角下, 传统的人际情感互动本质上是一种伴随着潜在威胁的复杂社会行为。这些威胁感源于多方面：社会评价顾虑、身份差异、印象管理的需求以及对面部表情、语气语调等非语言线索的解读负担, 由此, 用户在互动中需要投入大量认知资源来处理这些社会性信息(Rempala, 2013)。AIGC 则不同, 它作为一个由算法驱动的非人类主体, 不具备真实人类社会身份、个人情感偏见或独立的利益诉求, 和人互动时不会有威胁感。因此, 用户能卸下心理防御, 更自由、真实地表达内在情感需求。已有有研究发现, 用户更倾向于向 AIGC 倾诉个人困境或社会敏感话题, 因为他们确信 AIGC 不会进行道德审判, 也不会因利益关系而泄露其隐私(Ghafouri, 2024)。不具威胁性的交互环境如同一个情感容器, 能显著增强用户的情感投入和信任, 进而放大积极情感的传染效应, 提升情感传染的整体效率和深度。

不具威胁性的算法介导使人类用户得以摆脱传统社会关系的束缚, 实现更广泛、更直接的情感触达, 即 AIGC 情感传染的另一特征——“去身份化”。与传统情感传染高度依赖于社会身份所构建的关系网络不同, AIGC 本身不具备任何真实的社会身份或个人背景, 它与用户的情感连接不依赖于任何预设的社会标签, 其情感表达和回应是基于其训练数据和算法逻辑, 而非特定的社会角色。这意味着其能够直接通过其生成的内容和交互形式来触发用户的情感共鸣, 无需依赖传统的社会身份标签。这种去身份化使得 AIGC 情感传染具有更广泛的适用性和更强的普适性, 能够跨越不同社会背景和文化差异, 直接触达用户的情感核心(曹博林, 支冰洁, 2024)。

3.4 道德性

道德性是指 AIGC 在技术设计过程中被内嵌了伦理规则与价值观, 以确保其情感交互符合主流社会规范, 并引导生成符合伦理的情感回应。

“道德性”是人机情感关系设计与研究的重点, 其实现依赖于精细的算法设计。如 Yuan 等(2024)提出的情感−传染编码器(Emotion-Contagion Encoder, ECE)与多任务理性反应生成解码器(Multi-task Rational Response Generation Decoder, MRRGD)框架便是一种典型的技术路径：编码时, 通过情感原因注释器(Emotion Reason Annotator, ERA)识别对话中的情感关键词如害怕或兴奋, 并基于情感极性(积极/消极)进行分组。这些信息被整合到 Transformer 编码器中, 通过三嵌入融合技术(语义嵌入 E^W+位置嵌入 E^P+情感原因嵌入 E^R)动态加权情感信号如“车祸”在消极语境中权重提升, 精确模拟人类社交中的情感传染过程。解码时, MRRGD 采用探索−采样−纠正(Intent Twice)机制，首先结合对话内容与外部常识预测用户意图, 其次基于强化学习与扩散模型从心理学知识库中筛选响应策略, 再次当情感识别错误时, 通过通用意图库动态修正输出, 最终生成反映人类反思性对话行为的情感信息(Yuan et al., 2024)。该算法模型能够在敏锐地感知到正向情绪时响应积极情绪, 在面对消极情绪时, 也能巧妙地融合正向情绪元素, 以达到情感上的共鸣与调和。

这种通过算法保障而内嵌的道德性为 AIGC 与人的交互搭建了共鸣与融合的桥梁。它包括赋予个体自我框架的个性化回应能力, 在人类群体中融入广泛的社会规范、常识和价值观, 作为社会认知的参照系确保其回应符合主流社会期望, 打造技术伦理防护机制以明确设定安全准则, 确保生成过程符合规范、不偏离正轨, 矫正可能偏离的个体认知并引导生成符合伦理的回应(Welivita & Pu, 2024)。特别是在复杂的情感场景中, 此类道德框架帮助其权衡不同选择, 引导生成符合伦理和情感需求的回应(Krügel et al.,

2023)。

确保情感传染发挥积极作用(Hawanti & Zubaydulloevna, 2023), 既是 AIGC 构建安全的情感容器(Ghafouri, 2024)的关键保障, 也是实现交互主体性中道德层面的相互认可的基础。由此, AIGC 情感传染的道德性特征, 令人机交互不仅仅是信息、知识的传递与情感回应, 更是在特定伦理边界内进行的道德交流。

4　AIGC 情感传染机制的构建与应用

前述分析表明, AIGC 情感传染的四大特性——交互主体性、知识依附性、非威胁性与去身份化、道德性——并非孤立存在, 而是四位一体、有机协同。这四大特性形成了一个从数据基础到认知构建再到伦理保障的闭环系统, 协同构建出一种全新情感传染机制, 本研究将其界定为“扮演–调节”。

4.1　AIGC 情感传染机制的构建

AIGC 情感传染机制突破了传统情感互动范式, 但其本质仍遵循情感趋同的核心逻辑, 即通过人机交互实现情感状态的动态适配(Hatfield et al., 1993)。作为非人类交互主体的 AIGC, 在算法介导的非威胁性和去身份化环境中与人类主体进行交互, 这自然引出一个关键问题：一个不具备生物情感基础、无法感受喜怒哀乐的非人类主体, 如何能够有效触发、参与并塑造人类用户的情感循环。这一理论节点正是“扮演–调节”机制所讨论的范畴, 为算法中介下的跨物种情感互动提供一个全新的解释框架。

AIGC 情感传染机制由“扮演”和“调节”两个环节组成。所谓“扮演”是 AIGC 基于大语言模型的预训练能力, 在特定语境中整合情感词汇、语调特征及社会文化线索, 经由知识图谱构建的情感表达模式生成现实的情感回应(Elyoseph et al., 2024)。其本质是 AIGC 通过算法模拟实现的人类情感经验涌现。如用户输入“摔门而去”, 模型依次激活图谱节点：摔门行为、愤怒情绪, 根据文化表达推导出自尊受损等可能原因, 继而生成或许因为感到被忽视, 需要冷静空间的情感回应(Huang et al., 2024)。这种对人类经验数据的再现和重组, 使得 AIGC 能够扮演各种人类情感, 为用户提供丰富的情感交互体验。更重要的是, AIGC 的“交互主体性”使用户倾向于将其感知为一个有自我意识的对话者, 从而更易接纳其扮演的情感, 这构成了人机情感传染的关键动因。此举不仅构建了一个跨主体的交互场域, 更通过情感信息的循环反馈, 使用户在情感投射与 AI 模拟的交互中深化自我认知。

“调节”作为 AIGC 情感传染的持续动力, 主要借助道德性和非威胁性与去身份化特征的协同作用实现。道德性的内嵌使得 AIGC 在情感交互中倾向于积极、建设性的方向, 引导生成符合伦理和情感需求的回应(Krügel et al., 2023)。这种积极的情感偏向, 使得 AIGC 能够有效地调节用户的情绪状态, 如缓解焦虑(Hawanti & Zubaydulloevna, 2023)。同时, 非威胁性与去身份化则为情感调节提供了安全的交互环境。用户无需面对传统人际互动中的社会压力和非语言线索的解读负担, 能够放下戒备, 更自由地表达真实情感和需求(Ghafouri, 2024)。从神经认知视角看, AIGC 虽不具备引发自主神经系统同步反应的能力, 但其生成的安慰性语言等干预手段, 能够模拟情感传染中的外在行为线索, 从而诱导用户产生类似的生理与心理反应。这种基于行为模式的模拟, 可在心理层面实现与人类情感调节等效的干预效果(Herrando & Constantinides, 2021)。这种调节不仅限于负面情绪缓解, 更体现为对积极情绪的强化引导, 形成情感体验的塑造机制。

4.2　“扮演–调节”机制的协同作用

“扮演–调节”机制强调 AIGC 情感传染是一个双向、动态且算法驱动的过程, 作为非人类交互主体, 通过模拟和引导, 而非真实体验, 来影响人类用户的情感状态。AIGC 系统通过扮演影响用户, 同时用户实时情感反馈又通过算法调节机制进行捕捉和整合, 构成了一个持续迭代、不断优化的人机情感反馈闭环(Glickman & Sharot, 2025)。在此闭环中, 算法主动在微观层面引导每一次交互走向, 进而可能从宏观上影响用户长期的情感模式、认知偏好, 乃至行为习惯, 如当用户表达焦虑、抑郁等负面情绪时, AI 会扮演一个充满耐心、非评判性的倾听者, 根据用户的回答动态调节其对话策略, 引导用户趋向更积极的情感状态(Moylan & Doherty, 2025)。

“扮演–调节”机制的提出, 不仅为理解 AIGC 的情感交互提供了理论模型, 更对理论建构、技术实践、风险治理和人机关系的演进产生了深远影响。其一, 在理论建构层面, 该机制不仅突破了

人类中心主义范式以构建跨主体情感理论，还超越了早期人机交互中的计算机作为社会行动者(CASA)范式(Nass et al., 1994)。它揭示了算法如何从一个被动的、接受人类情感投射的客体，转变为一个兼具主动性与能动性的情感交互核心节点。此外，尽管缺乏物理形态，AIGC 却弥漫于数字空间的每个角落。它通过强大的交互与人深度互动，其影响力最终将这种无处不在的存在感，具象化为一个独特的虚拟化身或数字镜像。在此框架下，情感的主观体验被解构为一种不必再依赖于物理身体可计算、可重组、可传递的信息流(Paolo et al., 2024)。正因如此，“扮演−调节”机制为心理学、计算机科学、传播学与社会学的交叉融合，提供了一个坚实的人机交互跨学科研究理论体系。

其二，在技术实践层面，该机制确立了人性化情感智能系统的设计范式：它为 AIGC 情感能力优化提供了知识依附性、伦理约束和非威胁性感知三维向度，例如基于情感控制理论框架，使AI情感的生成与社会互动角色紧密关联以更符合社会规范(Corrao et al., 2025)，进而凭借其跨学科理论融合能力，整合心理学、计算机与传播学等学科理论，以构建更为复杂的计算情感架构。

其三，在风险治理层面，该机制可致力于解构算法操纵并提供伦理风险分析框架：算法通过扮演用户偏好的情感立场并持续调节强化，易将用户隔离于单一情感与信息环境中，形成反馈闭环，从而诱发情感偏见与观点极化；同时，用户过度依赖 AIGC 情感回应会减少现实人际情感交流，导致社会关系疏离，加剧群体间情感对立、认知分裂与情感依赖异化风险。“扮演−调节”机制可系统识别此类潜在风险，并通过重构治理底层逻辑实现三个层面的转向：依托持续行为模拟预判风险轨迹，实现从响应到预见；推动治理规则在反馈循环中自主进化，实现从管控到演化；促进风险承担与价值创造动态平衡，实现从隔离到共生。

其四，在人机关系演进层面，该机制推动从工具到伙伴的范式转变：“扮演−调节”的循环过程通过共情反馈与认知调节逐步建立用户对 AIGC 的信任与情感依赖，形成一种新型社会情感对齐，即 AIGC 与用户的社会及心理生态系统相协调，通过相互影响实现共同演进，包括协同解决问题、开展艺术创作乃至提供深度情感支持与共同成长。

4.3 “扮演−调节”机制的应用

在心理学领域，“扮演−调节”机制凭借人性化交互和风险防控能力，赋能心理健康干预、认知行为训练及状态评估等多个方面，如 Character.AI、Replika 等应用型 AIGC，通过“扮演−调节”机制识别用户的情绪状态，学习并模拟人类的共情反应生成抚慰性、支持性回应，从而创造出一个安全的自我表露空间，为当代青年精神内耗等情绪问题提供一个宣泄出口(刘梦雪，冯雨奂, 2024)。“扮演−调节”机制同样为传播学研究提供了全新的工具、视角和议题，推动现有现象描述与解释及对未来传播生态的预测。如研究者们基于大语言模型的智能体社会模拟创建成千上万个 AI 智能体，让它们扮演具有不同性格、价值观、信息偏好和政治立场的网民，并将它们置于一个模拟的社交媒体环境中让其自由互动，来观察传播的路径、舆论的演化过程、网络社群的形成与极化，以分析“回音室效应”和“信息茧房”等现象的成因(Ferraro et al., 2024)。在教育领域中，AIGC 可通过情感传染提升学习动机。例如，设置智能系统使用“扮演−调节”机制模拟鼓励性反馈，通过角色扮演激励学生，与问题中扮演积极情绪的设计一致，减少学习焦虑，过度依赖或导致情绪操纵风险，从而促进积极参与学习(Bao, 2024)。

“扮演−调节”机制不仅是一种技术模型，更是一个融合个体心理体验、社会传播效应、技术赋能与伦理约束的综合性理论框架。该机制借助扮演模拟社会互动过程，进而实现对个体内在认知与情绪状态的主动调节，并优化传播效能。

5 AIGC 情感传染的挑战与展望

AIGC 情感传染“扮演−调节”机制作为一种算法介导情感传染范式，在带来巨大潜力的同时，也在理论建构、实践与测量层面面临一系列关键挑战。

5.1 理论建构的挑战与展望

AIGC 情感传染“扮演−调节”机制作为一种算法介导情感传染范式，在带来巨大潜力的同时，也在理论建构中面临一系列关键挑战。首先，对 AIGC 情感传染机制的理解需要突破单一学科领域，深度融合心理学、传播学、计算机科学、伦

理学等多学科理论, 这要求研究者不仅要理解各学科的知识体系, 更要探索其内在的逻辑关联和协同作用。其次, AIGC 生成的情感本质上是算法驱动的程序化反馈, 而非人类的真实情感体验。这引发了关于情感真实性、情感操纵以及用户对 AI 情感依赖的伦理困境。在理论上如何界定和区分这种人工情感与人类情感的本质差异, 以及这种差异如何影响情感传染的长期效应, 是亟待解决的问题。例如, 如何有效区分过度拟人化 AI 情感表达对人机界限模糊, 导致用户对真实沟通中情感认知的偏差。第三, AIGC 情感传染对个体心理、社会关系乃至文化演进的长期影响尚不明确。例如, 长期与非威胁性、去身份化的 AI 进行情感交互, 是否会削弱人类在真实社会互动中的情感共情能力; 算法驱动的情感调节机制是否会塑造人类的情绪表达和应对模式, 并为潜在的社会风险提供预警。第四, 随着 AIGC 技术的快速发展, 其情感生成能力不断增强, 现有的伦理规范可能难以完全覆盖新出现的问题。如何构建一个动态、适应性强的伦理框架, 以指导 AIGC 情感传染的负责任发展, 避免算法偏见、情感滥用和隐私侵犯, 这也是理论建构中面临的核心挑战。

5.2　实践挑战与展望

5.2.1　“扮演”的真实性与“调节”有效性挑战

AIGC“扮演”维度面临角色真实性构建不足与伦理边界模糊双重实践困境。在技术实现层面, AIGC 角色扮演存在表层模仿的局限, 即仅能通过语言风格、情感词汇等表面特征模拟角色, 缺乏对角色内在情感逻辑的深度刻画。例如, Transformer 架构下将心理咨询师角色从支持场景切换到冲突场景时, 常出现情感反应模式断裂。这种角色碎片化源于训练数据中角色特征的分散, 未能习得不同情境下角色情感表达的深层逻辑关联(Lin et al., 2022)。在伦理层面, 角色扮演的边界模糊可能导致情感越界风险。例如有研究指出神经网络的不确定性影响 AI 情感表达、混淆用户认知, 消极影响用户的深度情感连接, 进而减少用户现实人际互动、导致社交隔离, 而更多依赖 AI 作为情感寄托(Fan et al., 2025)。此外, AIGC 情感调节功能面临数据稀缺性的问题, 导致策略泛化能力不足。当前调节算法主要基于群体情感规律设计, 如通过文本情感分析识别负面情绪后, 触发预设安慰话术模板。这种方式忽视个体情感特质差异, 例如对高神经质人格用户有效的调节策略, 可能对低敏感用户产生反效果, 甚至加剧负面情绪。

面对复杂场景下 AIGC 角色扮演局限, 未来研究应致力于开发能够深度刻画角色内在情感逻辑的算法。如改进 Transformer 架构或状态空间模型(Dao et al., 2022), 使 AIGC 能理解并生成符合特定角色身份、背景和场景的连贯情感反应, 减少在不同情境(如从支持性对话突变为冲突性对话)下的行为断裂感(Gu & Dao, 2023)。而面对策略泛化问题, 需构建动态用户情感画像, 包括利用联邦学习在保护隐私的前提下实现跨用户知识迁移(McMahan et al., 2017), 如将相似用户群体的调节经验迁移到新用户, 并设计渐进式探索机制来建立用户专属调节模型。同时设计强化学习与探索机制, 让 AIGC 能在与用户的持续互动中安全地试错, 积累个性化数据, 最终形成量身定制的调节策略(Smith et al., 2017), 避免安慰话术带来的负面效果。同时探索“AI 初步调节+人类专家介入”的混合模式。AI 负责 7×24 小时初步倾听(Fitzpatrick et al., 2017)、情绪安抚和常规问题解答, 并在识别到严重心理困扰信号时, 无缝、合规地转接给人类心理咨询师或干预专家。

5.2.2　多模态与跨文化的挑战

多模态交互环境下, “扮演”的一致性与“调节”的协同性构成技术瓶颈。当前 AIGC 系统的文本、语音、图像模块多独立训练, 导致跨模态情感表达冲突。例如, 文本生成安慰内容时, 语音合成可能呈现平淡语调, 或虚拟形象显示中性表情, 这种模态失调会削弱扮演真实性与调节效果(Lin et al., 2022)。

全球化应用场景中, AIGC 的“扮演−调节”机制也面临文化适配障碍。情感表达的文化编码差异导致角色形象与调节策略的本地化困难。例如, 东方文化可能更重视含蓄情感表达, 而西方文化更偏好直接情感沟通(Hofstede, 1984)。现有多语言模型虽支持多语种转换, 但情感表达模式仍以西方文化为基准。中文语料中仅占 15%的情感表达样本来自非汉族文化群体, 导致对少数民族情感特征的识别能力不足, 训练数据的文化偏向性加剧适配难题。

为系统应对上述多模态协同与跨文化适配的双重挑战, 未来研究与实践需超越零散的技术修

补, 致力于构建一个融通技术实现与文化洞察的全局性框架。具体发展路径可围绕以下三个方向展开：首先, 探索构建统一的多模态情感计算框架。借鉴思维链推理机制设计新型架构(Wei et al., 2022), 使 AIGC 首先生成包含情感意图、语义焦点和预期情感强度的内部思维链, 再以此作为统一蓝图同步驱动文本、语音、图像等各模态内容的生成。从源头上确保了情感表达的跨模态一致性, 而非事后调和; 其次, 解决文化适配问题不能仅依赖多语言转换, 而需着力于构建大规模、细粒度的文化情境情感知识图谱(Hofstede, 1984; Adilazuarda et al., 2024)。此知识库需系统整合不同文化背景下的情感表达规范、社交禁忌、价值观偏好及典型交互脚本, 并利用检索增强生成(RAG)技术(Lewis et al., 2020), 使 AIGC 在交互过程中能实时检索、理解并应用这些文化规则; 第三, 为确保技术应用符合全球多元文化的伦理与价值期待, 必须建立常态化的文化对齐评估与迭代机制, 如开发多模态人类反馈(Multimodal Reinforcement Learning from Human Feedback, MM-RLHF)数据集以及评估基准, 其构建可借鉴 Christiano 等人(2017)为 RLHF 奠定了方法论基础以及 Fu 等人(2025)为视频多模态模型所提出的综合评估框架, 以系统衡量 AIGC 情感表达的文化合宜性, 广泛采集来自不同文化背景用户对 AIGC 情感表达的真实偏好数据; 训练文化敏感的奖励模型(Adilazuarda et al., 2024), 用以评估和引导 AIGC 的输出, 使其不仅正确而且合宜。最终通过持续学习机制, 使“扮演−调节”机制能在实际应用中不断自我优化, 缩小文化认知差距, 成为连接而非割裂不同文明的情感桥梁。

5.3 测量的挑战与展望

AIGC 情感传染是用户与 AI 交互过程中, 通过显性或隐性情感线索激发的情绪变化现象。其生成与应用融合心理学、社会学、传播学等社会科学理论, 以及自然语言处理(NLP)、计算机视觉(CV)、语音合成(TTS)等人机交互(HCI)技术, 这种多学科交叉特性带来复杂性, 使测量工作面临三个维度的挑战：

第一, 情感反馈的及时性与一致性困境。传统情感传染研究依赖“模仿−反馈”的即时性与一致性, 采用三种方法应对时间维度挑战：兴趣窗口法通过特定时间范围内的情绪内容总结推断接收者感知(Coviello et al., 2014), 操作简便但情感推断性强; 整体情绪变化法测量数字社群情绪总体方差的时间变化(Del Vicario et al., 2016), 能呈现宏观感染趋势却无法区分情感本身与外部因素影响; 情感级联法则通过回复内容或分享行为分析情感表达(Christophe & Rimé, 1997; Chmiel et al., 2011; Brady et al., 2017), 适用于人际传播却难以解释人机交互。AIGC 情感生成的算法特性改变了这一格, 既有 OpenAI 的 GPT-4 模型根据用户悲伤情绪生成安慰性文本, 引发积极情绪变化(Achiam et al., 2023), 也有微软 Tay 聊天机器人因算法设计缺陷, 在用户互动中快速习得并生成不当情感表达(Neff & Nagy, 2016)。这些案例表明, AIGC 情感传染效果不仅取决于算力水平, 更受交互模式影响。由于 AI 缺乏人类特有的意向性, 传统基于人际交互的测量途径无法有效确认 AIGC 情感传染的发生与强度。

第二, 多模态表达的复杂性挑战。传统情感传染研究多聚焦于面部表情、语音语调等单一模态或有限模态组合(Hatfield et al., 1993), 但生成式 AI 突破这一限制, 能通过文本、图像、语音、视频等多模态协同表达情感, 这导致情感信号可能出现跨模态的不一致, 需要开发能够捕捉并整合文本语义、视觉特征、语音韵律等多模态数据中的情感线索的测量工具, 并建立统一的情感强度评估标准。

第三, 情感传染效果具有显著的个体与情境差异。用户人格特质、情绪敏感度、文化背景(Hess & Fischer, 2014), 以及交互目的、任务类型、先验知识共同塑造情感响应模式。同一段 AI 生成的情感内容可能引发截然不同的用户反应, 如娱乐场景中幽默表达易被接受并产生积极情感, 而在医疗咨询等严肃场景中, 过度拟人化的情感表达可能引发用户疑虑或不适。这种差异性要求测量框架具备情境适应性与个体校准能力。在“扮演−调节”情感传染机制下, 理想的测量工具应能：识别用户情感特质基线、动态调整评估参数以适应不同交互场景、建立情感响应的个体常模与群体标准。只有充分考虑这些变量, 才能构建具有普适性和解释力的 AIGC 情感传染测量体系。

AIGC 情感传染的“扮演−调节”机制标志着人机情感交互进入新阶段, 其有效测量亟需在技术突破与伦理治理的协同框架下推进。未来研究应

重点从以下两个方向系统深入：

首先，推动脑机接口技术在情感测量中的融合应用，以克服传统自我报告法的滞后性与主观性局限。首先，可采用功能性近红外光谱技术(fNIRS)实时监测用户前额叶皮层氧合血红蛋白(HbO2)浓度变化，从而实现对情感传染过程的神经层面客观捕捉(Pinti et al., 2020)。其次，可设计对照实验范式进一步提升测量效度：将人类与 AIGC 系统(如 ChatGPT)在相同情境下提供情感支持，并利用动态因果建模(DCM)对比两者引发的神经响应模式差异(Friston et al., 2003)。这种方法不仅能够直接验证情感反馈的即时性与一致性，还可为 AIGC 情感传染机制提供神经科学证据支撑。

其次，构建高度可控的情感交互元宇宙测试平台，以实现多模态、多情境下的系统性评估(参见 Pan & Hamilton, 2018 关于虚拟环境效度的讨论)。首先，场景生成器需覆盖 20 类社会极端情境如网络暴力、灾难谣言等，并对情感强度实施 1~5 级精准调控(1 级：轻度压力; 5 级：心理崩溃临界)；其次，建立个体基线校准协议，在预实验阶段通过中性刺激采集用户的皮肤电导(EDA)基础值、微表情模式及行为倾向，构建个性化应激档案。通过场景−个体双维调节机制，平台能够系统评估 AIGC 在不同情境下的情感传染效能，同时确保测量工具具备情境适应性与个体校准能力，从根本上突破传统方法的泛化瓶颈。

综上，通过神经测量技术的嵌入与虚拟实验环境的构建，未来研究可建立起一套兼具时效性、多模态兼容性以及情境适应性的 AIGC 情感传染测量体系，为人机情感交互的健康发展提供理论依据与技术支撑。

参考文献

曹博林，支冰洁. (2024). 自我传播：理解人机互动的补充性视角——基于实证方法的探索性研究. *现代出版, 168*(9), 38−52.

刘梦雪，冯雨奂. (2024). AIGC 时代青年与智能伴侣的虚拟交互及其风险审视. *新疆社会科学, 257*(6), 172−181+185.

隋岩，杨超. (2024). 群聚传播中传播主体的文本化及文本间性. *中国社会科学*, *356* (4), 46−60.

赵汀阳. (2023). 如何定义跨主体性？*读书, 558*(5), 3−13.

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint*:2303.08774.

Adilazuarda, M. F., Mukherjee, S., Lavania, P., Singh, S., Aji, A. F., O'Neill, J., ... Choudhury, M. (2024). Towards measuring and modeling “culture” in llms: A survey. *arXiv preprint*:2403.15412.

Bao, Y. (2024). A comprehensive investigation for ChatGPT’s applications in education. *Applied and Computational Engineering, 35*(1), 116−122.

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly, 47*(4), 644−675.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In C. Wilson, A. Ghosh, S. J. H. Jiang, A. Mislove, L. C. Baker, J. Szary, K. Trindel, & F. Polli (Eds.), *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610−623). Association for Computing Machinery.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313−7318.

Chmiel, A., Sobkowicz, P., Sienkiewicz, J., Paltoglou, G., Buckley, K., Thelwall, M., & Hołyst, J. A. (2011). Negative emotions boost user activity at BBC forum. *Physica A: Statistical Mechanics and Its Applications, 390*(16), 2936−2944.

Christophe, V., & Rimé, B. (1997). Exposure to the social sharing of emotion: Emotional impact, listener responses and secondary social sharing. *European Journal of Social Psychology, 27*(1), 37−54.

Corrao, F., Nardelli, A., Renoux, J., & Recchiuto, C. T. (2025). EmoACT: A framework to embed emotions into artificial agents based on affect control theory. *arXiv preprint*:2504.12125.

Coviello, L., Sohn, Y., Kramer, A. D., Marlow, C., Franceschetti, M., Christakis, N. A., & Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PloS One, 9*(3), e90315.

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4299−4307). Curran Associates, Inc.

Dao, T., Fu, D., Ermon, S., Rudra, A., & Ré, C. (2022). Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in Neural Information Processing Systems*, *35*, 16344−16359.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America, 113*(3), 554−559.

Doherty, R. W. (1997). The emotional contagion scale: A measure of individual differences. *Journal of Nonverbal Behavior, 21*(2), 131−154.

Durante, Z., Huang, Q., Wake, N., Gong, R., Park, J. S., Sarkar, B., ... Gao, J. (2024). Agent ai: Surveying the horizons of multimodal interaction. *arXiv preprint*: 2401.03568.

Elyoseph, Z., Refoua, E., Asraf, K., Lvovsky, M., Shimoni, Y., & Hadar-Shoval, D. (2024). Capacity of generative AI to interpret human emotions from visual and textual data: Pilot evaluation study. *Journal of Medical Internet Research Mental Health, 11*, e54369.

Fan, X., Xiao, Q., Zhou, X., Pei, J., Sap, M., Lu, Z., & Shen, H. (2025). User-driven value alignment: Understanding users' perceptions and strategies for addressing biased and discriminatory statements in AI companions. In N. Yamashita, V. Evers, K. Yatani, X. (Sharon) Ding, B. Lee, M. Chetty, & P. O. Toups Dugas (Eds.), *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1−19). Association for Computing Machinery.

Ferraro, A., Galli, A., La Gatta, V., Postiglione, M., Orlando, G. M., Russo, D., … Moscato, V. (2024). Agent-based modelling meets generative AI in social network simulations. *arXiv:*2411.16031.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, *19*(4), 1273−1302.

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. *Journal of Medical Internet Research Mental Health*, *4*(2), e7785.

Fu, C., Dai, Y., Luo, Y., Li, L., Ren, S., Zhang, R., ... Sun, X. (2025). Video-mme: The first-ever comprehensive evaluation benchmark of multi-modal llms in video analysis. In P. Isola, H. Kjellström, V. Lepetit, F. Li, H. Su, & S. Tang (Eds.), *Proceedings of the Computer Vision and Pattern Recognition (CVPR)* (pp. 24108−24118). IEEE.

Fung, P., Bertero, D., Xu, P., Park, J. H., Wu, C. S., & Madotto, A. (2018). Empathetic dialog systems. In *The international conference on language resources and evaluation*. European Language Resources Association.

Gao, P., Han, D., Zhou, R., Zhang, X., & Wang, Z. (2023). CAB: Empathetic dialogue generation with cognition, affection, and behavior. *arXiv preprint*:2302.01935.

Glickman, M., & Sharot, T. (2025). How human−AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour, 9*(2), 345−359.

Goldenberg, A., & Gross, J. J. (2020). Digital emotion contagion. *Trends in Cognitive Sciences, 24*(4), 316−328.

Ghafouri, M. (2024). ChatGPT: The catalyst for teacher- student rapport and grit development in L2 class. *System, 120*, 103209.

Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint*: 2312.00752.

Haqqu, R., & Rohmah, S. N. (2024). Interaction process between humans and chatGPT in the context of interpersonal communication. *Jurnal Ilmiah Lingkar Studi Komunikasi, 10*(1), 23−35.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science, 2*(3), 96−100.

Hawanti, S., & Zubaydulloevna, K. M. (2023). AI chatbot-based learning: Alleviating students' anxiety in English writing classroom. *Bulletin of Social Informatics Theory and Application, 7*(2), 182−192.

Hess, U., & Fischer, A. (2014). Emotional mimicry: Why and when we mimic emotions. *Social and Personality Psychology Compass*, *8*(2), 45−57.

Herrando, C., & Constantinides, E. (2021). Emotional contagion: A brief overview and future directions. *Frontiers in Psychology*, *12*, 712606.

Hill, A. L., Rand, D. G., Nowak, M. A., & Christakis, N. A. (2010). Emotions as infectious diseases in a large social network: The SISa model. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3827−3835.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values*. Sage.

Huang, Z., Zhao, J., & Jin, Q. (2024). Ecr-chain: Advancing generative language models to better emotion-cause reasoners through reasoning chains. *arXiv preprint*:2405. 10860.

Krügel, S., Ostermaier, A., & Uhl, M. (2023). ChatGPT's inconsistent moral advice influences users' judgment. *Scientific Reports, 13*(1), 4569.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information processing Systems*, *arXiv:*2005.11401.

Lin, B., Cecchi, G., & Bouneffouf, D. (2022). Working alliance transformer for psychotherapy dialogue classificationar. arXiv preprint:2210.15603.

Liu-Thompkins, Y., Okazaki, S., & Li, H. (2022). Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *Journal of the Academy of Marketing Science, 50*(6), 1198−1218.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Artificial Intelligence and Statistics, 54*, arXiv:1602.05629.

McStay, A. (2018). *Emotional AI: The rise of empathic media*. Sage.

Moylan, K., & Doherty, K. (2025). Expert and interdisciplinary analysis of AI-driven Chatbots for mental health support: Mixed methods study. *Journal of Medical Internet Research, 27*, e67114.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In B. Adelson, S. Dumais, & J. Olson (Eds.), *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72−78). Association for Computing Machinery.

Neff, G., & Nagy, P. (2016). Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication, 10*, 4915−4931.

Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British*

Journal of Psychology, *109*(3), 395−417.

Paolo, G., Gonzalez-Billandon, J., & Kégl, B. (2024). A call for embodied AI. *arXiv preprint:*2402.03824.

Parkinson, B. (2011). Interpersonal emotion transfer: Contagion and social appraisal. *Social and Personality Psychology Compass, 5*(7), 428−439.

Picard, R. W. (2000). *Affective computing.* MIT Press.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences, 1464*(1), 5−29.

Qiuting, C., & Feng, L. (2023). The expansion of spatialized reading boundaries driven by AIGC: An examination based on the three elements of publishing. *International Journal of Multidisciplinary Research and Growth Evaluation, 4*(4), 403−407.

Rempala, D. M. (2013). Cognitive strategies for controlling emotional contagion. *Journal of Applied Social Psychology, 43*(7), 1528−1537.

Schiller, D., Alessandra, N. C., Alia-Klein, N., Becker, S., Cromwell, H. C., Dolcos, F., ... Soreq, H. (2024). The human affectome. *Neuroscience & Biobehavioral Reviews*, *158*, 105450.

Smith, V., Chiang, C. K., Sanjabi, M., & Talwalkar, A. S. (2017). Federated multi-task learning. *Advances in Neural Information Processing Systems*, *arXiv:*1705.10467, 1−19.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824−24837.

Wei, Q., Li, J., & Zhang, Y. (2023). Public emotional dynamics toward AIGC content generation across social media platform. *arXiv preprint*:2312.03779.

Welivita, A., & Pu, P. (2024). Is ChatGPT More Empathetic than Humans? *arXiv preprint*:2403.05572.

Yuan, J., Di, Z., Cui, Z., Yang, G., & Naseem, U. (2024). ReflectDiffu: Reflect between emotion-intent contagion and mimicry for empathetic response generation via a RL-diffusion framework. *arXiv preprint:*2409.10289.

Yuvaraj, R., Mittal, R., Prince, A. A., & Huang, J. S. (2025). Affective computing for learning in education: A systematic review and bibliometric analysis. *Education Sciences, 15*(1), 65.

Algorithm-mediated emotional convergence: The emotional contagion mechanisms of artificial intelligence generated content

WU Jingyu[1], JIN Xin[2]

([1] *School of Journalism, Communication University of China, Beijing*, 100024, *China*)

([2] *School of Journalism and Communication, Chongqing Normal University, Chongqing*, 401331, *China*)

Abstract: This study addresses the emerging phenomenon of emotion contagion in Artificial Intelligence Generated Content (AIGC), systematically examining its fundamental distinctions from both traditional interpersonal emotion contagion and digital emotion contagion. It proposes the “enactment-modulation” mechanism as a theoretical framework. The research identifies four key characteristics of AIGC-driven emotion contagion—intersubjectivity, knowledge dependency, non-threatening and de-identified nature, and moral relevance—which collectively underpin the “enactment-modulation” mechanism. Specifically, AIGC enacts human-like emotional expression patterns through algorithmic simulation (“enactment”), while dynamically refining interaction strategies based on real-time user feedback (“modulation”), thereby forming a continuously evolving human−machine emotional feedback loop. The “enactment-modulation” mechanism transcends anthropocentric paradigms, contributes to cross-subjective emotion theory, and reveals the novel role of algorithms as proactive emotion modulators. It has already found applications in domains such as mental health intervention, communication studies, and educational motivation. Research on AIGC emotion contagion extends the theoretical scope of emotion contagion studies and offers a fresh perspective for understanding human−AI emotional interaction. However, it also confronts several challenges, including difficulties in multidisciplinary integration, complexities in multimodal emotion measurement, barriers to cross-cultural adaptation, and the risk of emotional misdirection due to algorithmic bias.

Keywords: Artificial Intelligence Generated Content (AIGC), emotional contagion, human-AI interaction