

人性化与数字化的共生:心理学与人智交互的双向赋能*

胡耿丹 齐萌

(同济大学人文学院心理学系,上海 200092)

摘要:随着人工智能(AI)从工具向智能伙伴演进,人智交互面临信任危机、认知过载等人性化设计问题。文章从双向赋能视角探讨心理学对AI交互人性化设计的理论建构。发现心理学通过认知理论如情绪计算、心理AI如不确定条件下的高效决策、社会心理机制如信任启发式、伦理校准如隐私保护等方面为人智交互设计提供认知模拟框架、交互规范和伦理准则;而人智交互则革新了心理学研究的范式,在心理咨询、心理测评、认知神经科学、社会心理学、精神病学等领域推动了AI的技术应用,如通过脑机接口提升神经信号解码精度,通过社交媒体数据探寻群体情绪影响机制。未来研究需构建“理论-数据双轮驱动”模式,整合多模态数据与跨文化规范,促进心理学与AI在技术融合、伦理规制和学科共建中的深度共生,实现技术理性与人类心理需求的平衡。

关键词:人性化设计;人智交互;心理学理论;人工智能;双向赋能

中图分类号:B842.5

文献标志码:A

文章编号:1003-5184(2025)06-0492-08

1 引言

自20世纪80年代初以来,计算机与信息系统的演进已从工具性的人机交互(human-computer interaction, HCI)逐步转向智能化的人智交互(human-AI interaction, HAI)。HCI指的是人与系统之间的交互,系统是随科技发展时代不断演化的,智能化时代是各种智能组合形式的系统(陈善广等, 2021)。人工智能(artificial intelligence, AI)技术有能力造福人类,但若不当开发和使用,却反而会危害人类和社会。面对这一挑战,人智交互作为一个跨学科新领域应运而生,其本质是“人-AI交互”“以人为中心的AI”(许为, 2024)。人智交互的研究旨在将AI引入良性演进的轨道,通过前置性的理论规制,有效平抑智能机器潜在的伦理风险与社会负外部性。

当前,AI正从被动执行指令的工具进化为具备自主决策、情感反馈能力的智能伙伴(姜婷婷等, 2022;许浩等, 2025)。这一转变以大语言模型如GPT-4、情感计算技术、脑机接口(brain-computer interface, BCI)的突破为标志,推动着AI深度融入教育、医疗、工业等场景——从个性化学习助手、智能诊疗系统到协作机器人,人类与AI的交互频次、复杂度及情感卷入度均达到历史峰值(Amershi et al., 2019)。

技术的高速迭代及其向医院、社会的渗透暴露出显著的“人性化缺口”。例如:医疗AI因诊断逻辑不透明导致患者信任度不足30%(Sundar, 2020),复杂多模态界面引发的认知过载使操作失误率提高40%(刘冬予等, 2024),而陪伴机器人的机械回应更被诟病为“情感疏离催化剂”(Green, 2018)。这些问题的症结,在于“技术中心主义”的技术理性与“情境依赖”的人类心理逻辑之间存在内生性张力——AI追求基于统计概率的精准计算,而人类行为则时刻处于认知偏差、情感偏好与社会规范的动态博弈之中。在老龄化加剧与心理健康危机凸显的当下(李敬荣等, 2022),社会对AI的期待已从“效率工具”升级为“认知伙伴”,其设计核心亟需从“技术功能导向”转向“心理需求适配”。心理学作为解码人类心智机制的基础学科,为破解这一矛盾提供了关键钥匙:认知心理学中关于人类通过内化知识单元处理外部输入并生成主观情绪取向的理论(Zhao et al., 2022),为AI情感计算模型的构建提供了底层逻辑。例如,通过测量用户情绪反馈动态调整交互策略,可使智能诊疗系统的共情响应准确率提升62%(Kriegeskorte & Douglas, 2018);社会心理学揭示的“机器启发式”认知捷径(如用户通过界面透明度快速判断AI可靠性),直接指导医疗AI的可视化诊断报告设计,将用户信任度提升30%

* 基金项目:国家社会科学基金教育学一般课题(BLA210211)。

通信作者:齐萌, E-mail: 1561644992@qq.com。

(Sundar, 2020)。

同时,人智交互场景正成为心理学研究的“数字实验室”。可穿戴设备采集的脑电信号(EEG)、心率变异性(HRV)等多模态数据(Stachl et al., 2020),为实时追踪个体压力状态提供了动态证据,推动心理健康测评从滞后的自陈量表转向精准的生理-行为-认知融合建模;虚拟社交平台中AI代理引发的群体极化现象(Vondrick et al., 2016),则为社会心理学提供了观察“技术中介化人际影响”的新场景,催生了“数字具身性”对自我概念重塑的理论探索(许浩等, 2025)。

面对上述挑战,既有研究已开始关注这一交叉领域。例如,姜婷婷等人(2022)重点从图书情报学视角探讨了人智交互中的用户体验要素与评价指标;许为(2024)则系统勾勒了以人为中心的AI作为跨学科领域的宏观架构与工程方法。与上述研究不同,本文的创新之处在于从“机制”层面切入,在心理学与人智交互领域首次系统构建了心理学与人智交互的“双向赋能”理论框架。本文不仅关注心理学如何单向地为AI提供人性化设计的理论“标准”,更深入探讨AI技术如何反向重构心理学的研究范式,进而提出“理论-数据双轮驱动”的共生路径。这一视角试图突破单纯的技术应用或理论思辨,为解决人智交互中的“技术理性”与“心理需求”断层问题提供可操作的闭环机制。本文将首先剖析心理学如何为AI交互的人性化设计提供理论支撑,继而探讨AI技术如何反哺心理学研究,最后提出基于“理论-数据双轮驱动”模式的未来协同发展路径。

2 心理学理论对AI交互人性化标准的多维建构

在构建良性的人智交互生态时,单纯依赖‘人性化设计’往往难以触达交互的本质。这是因为‘设计’仅是落实功能的手段,而衡量其是否符合人类心理机制的‘人性化标准’,则需要心理学提供理论依据。二者的逻辑分野在于:前者是‘术’,关注如何实现;后者是‘道’,界定何为合理。心理学及其衍生的心灵哲学正是通过输出这些关于认知、情感与伦理的‘标准’,成为了AI不可或缺的支撑理论。例如,当前AI中的强化学习理论源于心理学的行为主义理论,即生物体通过对环境中奖励或惩罚刺激形成预期,进而养成追求最大收益的习惯性行为(Miller, 2019)。这种跨学科的理论迁移不仅提升了AI在动态环境中的决策效率,更奠定了机器学习

领域的行为建模基础。而AI界面临的机器情绪反应、模糊状态决策等挑战,也需依赖心理学相应领域的突破,凸显了心理学对AI发展的基础性支撑(Miller, 2019)。

2.1 认知心理学驱动AI的高级功能模拟与交互升级

认知心理学作为研究人类认知高级心理过程的学科,涵盖思考、决策、推理、动机和情感等维度(Zhao et al., 2022)。人类区别于机器的核心特征,在于其信息处理过程深深植根于内化的认知图式之中,并据此衍生出满意或厌恶等主观情绪体验。这些情绪特质由认知心理机制生成(Kriegeskorte & Douglas, 2018; Pradhan et al., 2020),通过测量用户情绪变化并引导AI更新知识结构,可赋予机器类人的态度、偏好和情感体验。尽管现有AI在模拟人类记忆、注意力、情绪等方面仍处发展阶段,但融合认知心理学的AI系统已成为重要研究方向,其目标是赋予计算机模拟高级认知的能力,使其能够识别情绪、理解人类感受,最终实现与人类及其他AI的对话和共情。从实践成果看,AI已融合心理学、脑科学、计算机科学等理论方法,通过机器模拟人类心理活动,形成了学科之间相互融合促进的模式。随着心理学研究范围的拓展,其理论成果为AI产品的开发提供了丰富的建模依据,推动AI技术向心理学领域深度渗透:基于面部表情识别的情绪计算模型、依托大数据的舆情分析框架、整合认知神经科学的智能医学影像诊断系统,以及融入危机干预理论的自杀预警系统等技术产品相继落地(Branch, 2019)。这些应用以心理学理论为底层逻辑,提升了AI的社会智能水平。

当前AI设计对人类认知模型的广泛借鉴,比如Zhao等人在2022年提出的“面子吸引力”“情感计算”“音乐情感”等场景建模,本质上是心理学理论在AI交互设计中的具象化实践,即通过模拟人类思维与情感倾向,心理学为AI的认知功能开发提供了可验证的理论原型,加速了心理学理论向AI应用的技术转化。

2.2 心理AI:启发式策略赋能算法设计的透明性与适应性

Martignon(2023)提出的心理AI概念,强调以心理学理论(如认知启发式、记忆规律)指导算法设计,与传统AI依赖大数据解决稳定环境下明确问题不同,其聚焦于不确定条件下的高效决策。Martignon(2023)通过案例佐证:Katsikopoulos等人

(2022) 基于“最近性启发式”设计流感预测算法,仅使用最近一周就诊数据来预测疫情,相较传统模型减少了历史数据依赖,在数据稀缺或动态变化场景中实现了简单性、快速适应性和低计算成本;Jung 等人(2020)将“舍入原则”融入再犯风险预测模型,通过特征筛选和权重简化,使算法决策逻辑清晰可追溯,提升了非专业用户的理解度,同时减少数据噪声导致的过拟合问题。心理 AI 的启发式策略包括通过适配环境结构,在数据稀缺时降低计算成本;通过特征筛选和权重处理提升算法透明度,增强司法、医疗等高危领域的用户信任;通过最近性策略快速响应环境突变,无需重新训练即可适应新趋势,展现出传统 AI 难以企及的灵活性与可靠性。

2.3 社会心理机制:界定交互设计的信任与规范边界

用户对 AI 的信任建立常依赖“机器启发式”认知捷径(Sundar, 2020),即通过界面透明度、拟人化等线索快速判断系统可靠性。这种认知捷径本质上是人类在信息过载环境中进化出的高效决策策略,通过捕捉直观线索降低判断成本,其运作逻辑是依据社会认知中的“启发式偏差”,即依赖简化线索而非复杂分析进行决策。例如,医疗 AI 若以可视化图表展示“基于 2000 例糖尿病患者数据匹配症状”的诊断逻辑,可激活用户“客观性”刻板印象,使信任度提升 30%;反之,算法不透明的黑箱推荐会导致用户接受度下降 25%(Sundar, 2020),这表明人类对技术系统的信任构建高度依赖可感知的“认知锚点”,而非技术本身的实际效能。

跨文化社会规范会进一步影响心理 AI 的交互设计:高语境文化用户如国内的 AI“家长式代理”,其对智能家电自动调节的接受度比低语境文化用户高 40%,这种差异根植于集体主义与个人主义文化对“技术代理”的不同心理预期,前者更容易接受技术对日常决策的适度介入,后者则对个体自主权的让渡更为敏感(Sundar, 2020;Chen & Sundar, 2018)。但无论文化背景如何,“反应式交互”如添加“人工确认”按钮均成为必要设计,其本质是通过保留用户最终决策权,契合人类普遍存在的“控制欲心理基线”,即便接受技术辅助,仍需通过显性决策行为确证自身主体性,这种设计使跨文化信任度平均提升了 28%,体现了社会规范对人机代理边界的隐性约束。

Sundar(2020)的 TIME 理论为人性化交互设计

提供了双重指导:线索路径通过“启发式判断”设定界面可解释性标准,如推荐系统提供不超过 3 种核心推荐理由以避免认知过载;行动路径要求交互流程遵循“用户控制-机器辅助”的动态平衡,如教育 AI 允许用户调整学习计划难度。这种双重机制构成了“人性化交互”的评估指标,既通过透明度线索如医疗 AI 可视化专业资质触发积极认知、提升用户信任,又通过控制权分配如团队协作机器人声明角色权限满足社会协作规范、增强人类认同(Sundar, 2020),确保技术设计与人类心理预期的深度契合。本质上,线索路径通过优化信息呈现方式降低用户认知负担,行动路径通过调节技术代理的“介入程度”与用户的“控制权限”,使交互体验落在个体心理舒适区范围内,避免因技术过度自主引发信任衰减,或因控制权让渡不足导致效率损耗。

综上,心理学从认知功能模拟、算法设计优化、社会规范适配三个维度,构建了 AI 交互的人性化标准:认知心理学推动 AI 从功能实现向情感理解进化,心理 AI 通过启发式策略提升算法透明性与适应性,社会心理机制指导 AI 交互设计符合人类信任逻辑与文化规范。未来,心理学需要在情感计算、认知建模等领域获得更大的理论突破和创新,为 AI 赋予更贴近人类心智模式的决策能力。

3 AI 技术对心理学领域的革新与应用

3.1 心理咨询领域的 AI 赋能:形态、理论与伦理挑战

AI 的发展为心理咨询带来创新性变革。近年来民众对心理咨询的需求激增,但国内专业心理咨询师数量不足,且培养成本高昂(李敬蓉等,2022),促使 AI 心理咨询作为传统心理咨询的衍生与补充快速发展。其理论基础主要包括解释水平理论和情绪对话模型(ECM):解释水平理论指出,人们对 AI 的行为倾向于进行低水平解释,更关注具体行动而非背后原因,这为 AI 心理咨询的信息呈现和干预策略设计提供了依据(Kim & Duhachek, 2020);ECM 则借助 Hill 的助人理论,构建探索、安抚、行动三阶段情绪支持对话框架,通过情绪类别建模、内隐情绪捕获及外显情绪表达,实现可控的情绪疏导对话生成(Zhou et al., 2017)。在技术应用层面, AI 心理咨询依托模式识别、机器学习等技术,如朱廷劭提出的生态化识别(ER)方法,通过整合用户语音、面部表情、网络行为等多源数据,实现个体心理特征的自动识别,已成功应用于家暴研究和自杀干预(朱廷劭, 2019)。

基于工作方式与载体的差异, AI 心理咨询呈现出多元形态。按工作方式划分, 文本心理咨询类如早期的聊天机器人“Eliza”、国内的“小丘”及国际上的“Woebot”, 通过简短文本会话提供认知行为治疗, 研究显示 Woebot 能在两周内减轻学生抑郁症状; 语音心理咨询类如校园 AI 减压赋能机器人, 通过实时采集语音、生理、表情数据动态调整咨询策略; 面对面心理咨询类如虚拟人“Ellie”, 结合虚拟现实与情感计算技术, 通过分析口头反应、面部表情和语音语调, 有效评估与治疗创伤后应激障碍(PTSD)、抑郁和焦虑(Fitzpatrick et al., 2017)。按咨询载体划分, 心理咨询 App 如 Calm、Headspace 提供冥想指导, EMoods 双相情绪追踪器(EMoods Bipolar Mood Tracker)记录情绪与生理数据; 虚拟心理咨询机器人如超级临床医生系统, 在飞行恐惧症治疗中与传统认知行为疗法效果相当; 实体心理咨询机器人如 NAO 用于自闭症儿童康复、Paro 用于减少痴呆症患者躁动, 均展现出理论与技术融合的多样化心理支持能力(Dawe et al., 2019)。

然而, 当前 AI 心理咨询面临多重伦理挑战。在技术层面, 算法透明度不足与数据安全隐患可能损害用户信任并加剧心理伤害(Green, 2018); 在应用层面存在技术滥用风险及心理卫生资源分配不均, 且缺乏统一的风险评估与转介标准(Green, 2018); 在文化层面需针对不同文化背景适配咨询模型以避免干预偏差(李敬荣等, 2022)。未来发展需聚焦于多模态数据融合以提升个性化干预精度, 强化生态化识别技术在危机干预中的应用; 同时建立严格的数据安全机制, 推动可解释的 AI 技术发展, 并加快制定行业伦理规范, 确保技术发展兼顾专业性与人文关怀(朱廷劭, 2019; 李敬荣等, 2022)。

3.2 心理测评的智能化革新: 数据类型与发展方向

近年来, 机器学习与大数据挖掘技术推动着心理健康测评革新, 催生了“智能化心理健康测评”新兴领域(Jiang et al., 2022)。智能化心理健康测评的数据来源可划分为两大类型: 一是主观行为类数据, 包含社交媒体互动(Park et al., 2015)、智能手机日常轨迹(Stachl et al., 2020)、虚拟游戏操作(Dechant et al., 2021)等, 可实现对心理状态的间接表征; 二是客观生理类数据, 如可穿戴设备采集的 EEG、HRV 信号, 其中 EEG 结合 LSTM 模型对抑郁症的识别准确率已超 95%(Ay et al., 2019), 具备临床诊断的技术潜力。

未来发展需强化心理学理论对测评设计的指导, 通过“证据中心设计”(Mislevy et al., 2003)与“计算心理测量学”(von Davier & Halpin, 2013)构建理论与数据驱动结合的模型, 提升测评的针对性与可解释性; 加强多模态数据融合分析如 EEG、语音、文本协同建模以突破单一数据局限, 实现更精准的心理状态评估; 重视信效度检验, 通过交叉验证与跨情境稳定性测试确保测评工具的科学性; 同时建立完善伦理框架, 在数据采集、算法设计中保护用户隐私、规避偏见风险, 推动智能化测评从探索性研究向临床实用化迈进。

3.3 认知神经科学与 AI 的交叉融合: 技术突破与临床价值

AI 与认知神经科学关联紧密, 其相互作用对双方发展至关重要。AI 对认知神经科学的助力集中体现在 BCI 技术中, 该技术通过电生理记录细胞外电活动, 实现与中枢神经系统的通信(刘冬予等, 2024)。诚然, 在心理现象与神经活动的关联机制尚未完全厘清、BCI 观测精度受限的当下, AI 技术凭借其强大的非线性映射能力, 为神经信号的解读提供了一种极具潜力的“计算解法”, 从而在一定程度上规避了既有理论与硬件的短板。

通过优化信号处理算法, AI 能显著提升 EEG 与功能性磁共振成像(fMRI)的分析精度。在 EEG 解码中, 机器学习技术精准定位运动相关脑功能区, 为 BCI 控制义肢、计算机等外部设备奠定基础, 如中风患者通过神经信号训练实现辅助设备控制, 复杂运动神经信号分类准确率达到 78.9%, 且无需先验数据训练, 提升了技术普适性(Nurse et al., 2015)。深度神经网络的介入显著锐化了 fMRI 的脑功能定位能力, 这不仅为绘制高分辨率脑图谱奠定了坚实基础, 更成为解构脑认知机制的关键数据引擎。

在精神障碍与神经损伤治疗中, AI 解决了传统神经调控的个性化不足问题。斯坦福大学的加速智能神经调节疗法(SAINT)通过智能算法实现了经颅磁刺激(TMS)的精准定位, 针对个体差异调整刺激参数, 临床双盲实验显示其对抑郁症缓解率达到 78.57%, 为强迫症、PTSD 等疾病提供了新治疗范式(Phillips et al., 2020)。在中风康复中, 基于 BCI 的智能治疗策略通过机器学习分类运动神经信号, 助力患者肢体功能恢复, 展现了 AI 在神经重塑领域的应用潜力。

3.4 社会心理学研究的 AI 赋能:数据采集与机制挖掘

长期以来,社会心理学试图通过自陈量表捕捉个体的社会性反应,但这种方法往往难以规避记忆模糊与迎合心理带来的数据失真。AI 与社交媒体的结合为解决这些问题提供了新思路:通过分析社交媒体的文本与行为数据如帖子、评论,可有效度量用户情绪、文化价值观和行为意图,相关研究在多语言背景国家展现出高效性与优越性(Hernández - García & Giménez - Júlvez, 2020)。在人类行为模式研究中,深度学习算法处理海量数据,规避了传统人工场景实验的人力投入大、被试有限及伦理风险,如“predict vision”实验通过学习超 600 小时 YouTube 视频,实现对 43% 人类交互行为如拥抱、亲吻的预测,为总结社会活动规律提供真实数据支撑(Vondrick et al., 2016)。

此外, AI 被用于挖掘社会心理机制: Pan 等人(2021)借助 AI 对微博评论进行分析,探讨疫情下的群体情绪变化,揭示疫情情况、恐惧与政府信任度的关系; Huang 等人(2020)通过 AI 分析疫情期间微博发帖内容,发现提升集体主义态度可增强防疫意识,而强化疫情恐惧可能产生负面效果。这些研究凸显了大数据与自然语言处理技术在样本容量、及时性和生态效度上的优势,以及向社会实践转化的潜力。

3.5 精神病学的 AI 辅助实践:监测、诊断与机制研究

AI 在精神障碍的日常监测、诊断和治疗中展现重要价值。通过可穿戴设备生理数据采集、自然语言处理及社交媒体挖掘,实现个体全天候动态监测;临床场景中影音与生理数据的整合分析为诊断提供多元支持,对话机器人则降低了诊断访谈的人力与时间成本(刘冬予等, 2024)。

作为精神障碍机制研究的核心领域,精神病理学在传统上依赖大样本纵向追踪,而 AI 技术为基因研究带来突破。Xiong 等人(2015)构建的计算模型以 DNA 序列为输入,通过机器学习预测基因剪接变异,识别常见、罕见及自发变异,在自闭症预测中准确率达到 85%,并整合基因影响与神经生理标记,精准表述智力障碍、自闭症谱系等神经发育相关精神障碍的机制(Xiong et al., 2015)。

在心理健康诊断中,结合 AI 的电子健康记录技术突破传统纵向研究的数据单一性局限,基于个体差异化健康记录,实现动态、个性化评估,较传统检测更全面精准(Nemesure et al., 2021)。自然语言指标如日常语言、文本写作被纳入精神障碍预测体系,进一步丰富诊断维度(刘冬予等, 2024)。这些应用提升了精神病理学研究效率,可为临床干预提供数据驱动的精准支持。

4 心理学与人智交互双向赋能的发展方向 and 路径

基于上文分析,心理学与人智交互的关系并非单向的理论输出或技术应用,而是一个深度的共生系统。为了更直观地阐释这一机制,本文构建了“心理学与人智交互双向赋能理论框架”(见图 1)。

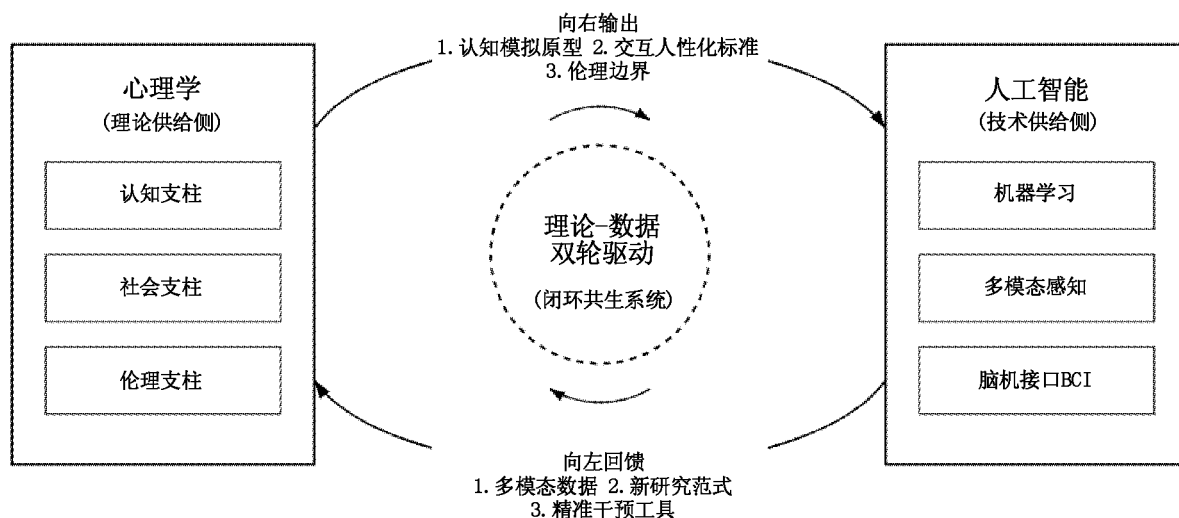


图 1 心理学与人智交互双向赋能理论框架图

在 AI 技术融合深化层面,亟需构建“理论 - 数据双轮驱动”模式以破解传统数据驱动研究的局限

性,即应在 AI 模型设计中深度嵌入心理学理论框架,形成“理论假设引导数据采集、数据验证反哺理

论迭代”的闭环,避免技术应用与心理机制解析的割裂。以精神病理学研究为例,心理动力学理论提出潜意识冲突、早期创伤可能通过神经-基因交互机制影响精神障碍发生(如童年创伤与前额叶皮层发育关联假设),此理论假设引导AI模型设计时优先采集基因测序数据,比如Xiong等人2015年构建的基因剪接预测模型,与EEG数据连接,通过机器学习算法对多模态数据的分析,AI可识别出与童年创伤相关的神经标记(如前额叶皮层异常激活),并发现其与5-羟色胺转运体基因启动子区(5-HTTLPR)基因多态性显著关联,该研究不仅验证了“基因易感性-环境应激”交互作用的理论假设,更通过引入“心理防御机制”作为调节变量,揭示出“认知调节”在病理机制中的中介作用,推动了心理学理论从单一因果模型向动态系统模型的迭代。这种融合模式使AI模型超越单纯的数据关联分析,转而聚焦于病理机制的因果推断。以自闭症谱系障碍为例,研究可将“社交认知发展理论”作为先验假设,引导AI在fMRI数据中定向搜索脑功能连接网络,进而精准识别出镜像神经元系统异常与早期社交匮乏之间的交互致病路径。这种方法为个性化干预提供了具有理论解释力的靶点。此类实践不仅能提升精神病理学研究的临床转化价值,更可为社会心理学(如依据社会认同理论分析社交媒体行为数据)、认知神经科学(如基于双重编码理论优化BCI信号解码)等领域提供方法论借鉴,确保AI技术在心理学应用中始终锚定人类心理现象的本质规律,而非局限于数据拟合的技术工具。

在跨学科研究重点领域,人机协同测评体系的构建需要突破单一数据维度的局限,着力整合“生理-行为-认知”多源信息,形成兼具动态性与理论根基的评估框架。例如,将可穿戴设备采集的HRV、睡眠周期等生理指标,与心理学经典理论中的情绪调节理论、认知负荷模型深度融合,通过机器学习构建实时心理健康评估系统,该系统可基于用户日常运动轨迹(行为数据)识别压力积累模式,结合EEG捕捉的前额叶皮层激活特征(认知数据),动态校准抑郁、焦虑风险模型,突破传统量表依赖自陈报告的滞后性与主观性。这种跨模态融合不仅能捕捉个体在特定场景下的即时心理状态如会议发言时的生理应激反应程度与认知资源分配情况,更可通过心理学理论锚定核心观念如情绪调节能力,使评估结果兼具技术精准度与心理解释力,为职场心

理干预、校园心理危机预警提供实时化、个性化的决策支持。

针对大数据研究中普遍存在的样本偏性问题,社会伦理与技术规制的协同创新成为跨学科攻坚的关键。建立跨平台数据共享机制需打破语言、文化、地域的数据壁垒,例如构建涵盖亚非拉等多语言群体的社交媒体语料库,结合跨文化心理学理论(如集体主义-个人主义维度)标注数据特征,使AI模型在分析群体情绪时能规避“西方中心主义”偏差,准确识别不同文化语境下的心理应激信号如东方文化中“含蓄表达”对应的语言标记。

同时,推动AI伦理指南在心理学研究中的落地需构建技术-伦理双轨框架:在技术层面,开发可解释性算法如注意力机制可视化工具,确保抑郁症预测模型的决策逻辑,如对“绝望”类词汇的权重赋值可被临床心理学家追溯验证;在规制层面,参照American Psychological Association伦理准则、通用数据保护条例(GDPR),建立针对心理数据的“最小化采集-分层授权-动态脱敏”机制,既保障用户隐私如加密处理可穿戴设备中的生理信号,又通过算法公平性审计如将不同性别、年龄群体的预测误差控制在5%以内,以杜绝技术滥用,最终实现AI技术在心理学应用中的科学性与伦理正当性的统一,促进科技向善。

参考文献

- 陈善广,李志忠,葛列众,张宜静,王春慧.(2021).人因工程研究进展及发展建议.《中国科学基金》,(2),203-212.
- 姜婷婷,许艳闰,傅诗婷,陆伟.(2022).人智交互体验研究:为人本人工智能发展注入新动力.《图书情报知识》,39(4),43-55.
- 李敬荣,赵然,张玉.(2022).人工智能心理咨询的发展与应用.《心理技术与应用》,10(5),296-306.
- 刘冬予,骆方,屠焯然,饶思敬,沈阳.(2024).人工智能技术赋能心理学发展的现状与挑战.《北京师范大学学报(自然科学版)》,60(1),30-37.
- 许浩,程卿玄,董晶,吴丹,田丽.(2025).人智交互中的AI世代:缘起,特征与未来展望.《信息资源管理学报》,15(1),13-20.
- 许为.(2024).人智交互:以人为中心人工智能的跨学科融合创新.北京:清华大学出版社.
- 朱廷劭.(2019).人工智能助力心理学研究的应用场景.《学术前沿》,(20),48-53.
- Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., & Bennett, P. N. (2019). Guidelines for human-ai interaction.

- Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1 – 13.
- Ay, B. , Yildirim, O. , Talo, M. , Baloglu, U. B. , Aydin, G. , Puthankattil, S. D. , & Acharya, U. R. (2019). Automated depression detection using deep representation and sequence learning with EEG signals. *Journal of Medical Systems*, 43 (7), 1 – 12.
- Branch, B. (2019). Artificial intelligence applications and psychology: An overview. *Neuropsychopharmacol Hung*, 21 (3), 119 – 126.
- Chen, T. W. , & Sundar, S. S. (2018). This app would like to use your current location to better serve you: Importance of user assent and system transparency in personalized mobile services. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Article 537).
- Dawe, J. , Sutherland, C. , Barco, A. , & Broadbent, E. (2019). Can social robots help children in healthcare contexts? A scoping review. *BMJ Paediatrics Open*, 3(1), e000371.
- Dechant, M. J. , Frommel, J. , & Mandryk, R. (2021). Assessing social anxiety through digital biomarkers embedded in a gaming task. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1 – 15). IEEE.
- 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.
- Green, B. P. (2018). Ethical reflections on artificial intelligence. *Scientia et Fides*, 6(2), 9 – 31.
- Hernández – García, I. , & Giménez – Júlvez, T. (2020). Characteristics of YouTube videos in Spanish on how to prevent COVID – 19. *International Journal of Environmental Research and Public Health*, 17(13), 4671.
- Huang, F. , Ding, H. , Liu, Z. , Wu, P. , Zhu, M. , Li, A. , & Zhu, T. (2020). How fear and collectivism influence public's preventive intention towards COVID – 19 infection: A study based on big data from the social media. *BMC Public Health*, 20, 1 – 9.
- Jiang, L. , Tian, X. , Ren, P. , & Luo, F. (2022). A new type of mental health assessment using artificial intelligence technique. *Advances in Psychological Science*, 30(1), 157 – 167.
- Jung, J. , Concannon, C. , Shroff, R. , Goel, S. , & Goldstein, D. G. (2020). Simple rules to guide expert classifications. *Journal of the Royal Statistical Society A: Statistics in Society*, 183 (3), 771 – 800.
- Kriegeskorte, N. , & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, 21 (9), 1148 – 1160.
- Katsikopoulos, K. V. , Şimşek, Ö. , Buckmann, M. , & Gigerenzer, G. (2022). Transparent modeling of influenza incidence: Big data or a single data point from psychological theory? *International Journal of Forecasting*, 38(2), 613 – 619.
- Kim, T. W. , & Duhachek, A. (2020). Artificial intelligence and persuasion: A construal – level account. *Psychological Science*, 31(4), 363 – 380.
- Martignon, L. (2023). Psychological AI: Designing algorithms informed by human psychology. *Perspectives on Psychological Science*, 19(5), 1 – 15.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1 – 38.
- Mislevy, R. J. , Almond, R. G. , & Lukas, J. F. (2003). A brief introduction to evidence – centered design. *ETS Research Report Series*, (1), 1 – 29.
- Nemesure, M. D. , Heinz, M. V. , Huang, R. , & Jacobson, N. C. (2021). Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Scientific Reports*, 11 (1), 1980.
- Nurse, E. S. , Karoly, P. J. , Grayden, D. B. , & Freestone, D. R. (2015). A generalizable brain – computer interface (BCI) using machine learning for feature discovery. *PloS One*, 10 (6), e0131328.
- Park, G. , Schwartz, H. A. , Eichstaedt, J. C. , Kern, M. L. , Kosinski, M. , Stillwell, D. J. , Ungar, L. H. , & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934 – 952.
- Pradhan, N. , Singh, A. S. , & Singh, A. (2020). Cognitive computing: Architecture, technologies and intelligent applications. In K. K. Singh & M. Elhoseny (Eds.) , *Machine learning and cognitive computing for mobile communications and wireless networks* (pp. 25 – 50). Scrivener Publishing.
- Pan, W. , Wang, R. J. , Dai, W. Q. , Huang, G. , Hu, C. , Pan, W. L. , & Liao, S. J. (2021). China public psychology analysis about COVID – 19 under considering Sina Weibo data. *Frontiers in Psychology*, 12, 713597.
- Phillips, A. L. , Cole, E. J. , Bentzley, B. S. , et al. (2020). Stanford Accelerated Intelligent Neuromodulation Therapy (SAINT – TRD) induces rapid remission from treatment – resistant depression. *American Journal of Psychiatry*, 177(11), 1015 – 1026.
- Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human – AI interaction (HAII). *Journal of Computer – Mediated Communication*, 25(1), 74 – 88.
- Stachl, C. , Au, Q. , Schoedel, R. , Gosling, S. D. , Harari, G. M. ,

- Buschek, D. , . . . Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, 117(30), 17680 – 17687.
- von Davier, A. A. , & Halpin, P. F. (2013). Collaborative problem solving and the assessment of cognitive skills; Psychometric considerations. *ETS Research Report Series*, (2), 1 – 36.
- Vondrick, C. , Pirsiavash, H. , & Torralba, A. (2016). Anticipating visual representations from unlabeled video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 98 – 106.
- Xiong, H. Y. , Alipanahi, B. , Lee, L. J. , Bretschneider, H. , Merico, D. , Yuen, R. K. C. , Hua, Y. , Gueroussov, S. , Najafabadi, H. S. , Hughes, T. R. , Morris, Q. , Barash, Y. , Krainer, A. R. , Jovic, N. , Scherer, S. W. , Blencowe, B. J. , & Frey, B. J. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347 (6218), 1254806.
- Zhao, J. , Wu, M. , Zhou, L. , Wang, X. , & Jia, J. (2022). Cognitive psychology – based artificial intelligence review. *Frontiers in Neuroscience*, 16, 1024316.
- Zhou, H. , Huang, M. , Zhang, T. , Zhu, X. , & Liu, B. (2017). Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 730 – 738.

The Symbiosis of Humanization and Digitization : The Bidirectional Empowerment of Psychology and Human – AI Interaction

Hu Gengdan Qi Meng

(Department of Psychology, School of Humanities, Tongji University, Shanghai 200092)

Abstract: As artificial intelligence (AI) evolves from a tool to an intelligent partner, human – AI interaction is facing personalized design issues such as trust crisis and cognitive overload. Based on this, the article explores the theoretical construction of psychology on the humanized design of AI interaction from the perspective of bidirectional empowerment. It found that psychology provides cognitive simulation frameworks, interaction norms, and ethical guidelines for human – AI interaction design through cognitive theories such as emotion calculation, psychological AI such as efficient decision – making under uncertain conditions, social psychological mechanisms such as trust heuristics, and ethical calibration such as privacy protection; And human – AI interaction has revolutionized the paradigm of psychological research, promoting the application of AI technology in fields such as psychological counseling, psychological assessment, cognitive neuroscience, social psychology, psychiatry, etc. , such as improving the decoding accuracy of neural signals through brain computer interfaces and exploring the mechanism of group emotional influence through social media data. Future research needs to establish a model of “dual wheel drive based on theory and data”, integrating multimodal data and cross – cultural norms, promoting deep symbiosis between psychology and AI in technological integration, ethical regulation, and disciplinary co – construction, and achieving a balance between technological rationality and human psychological needs.

Key words: humanized design; human – AI interaction; psychological theory; artificial intelligence; bidirectional empowerment