

# 从准社交互动到依恋：人智情感关系的演变<sup>\*</sup>

吴燕<sup>\*\*1</sup> 耿晓伟<sup>1</sup> 周晓林<sup>\*\*2</sup>

(<sup>1</sup> 杭州师范大学经亨颐教育学院, 杭州, 311121) (<sup>2</sup> 华东师范大学心理与认知科学学院, 上海, 200063)

**摘要** 随着人工智能 (Artificial Intelligence, AI) 技术的快速发展和 AI 伴侣的广泛出现, 人智交互已从工具性使用逐渐演变为准社交互动, 并可能进一步发展为情感依恋。文章系统梳理了近 20 年心理学与人智交互领域的相关研究, 构建了人智依恋形成的理论模型。研究发现: (1) 人智关系呈现从工具性使用到准社交互动再到情感依恋的动态演进过程; (2) AI 依恋的形成受个体因素 (如孤独感、使用动机、情绪特质) 和 AI 特征 (如拟人性、自主性、反应性) 的双重影响; (3) 这种新型情感关系面临情感泡沫、隐私安全等伦理争议。已有研究存在样本单一 (偏重年轻群体) 和缺乏长期追踪等局限, 未来需采用纵向设计和多模态测量方法深入研究。AI 系统的发展需要平衡技术创新与伦理考量, 以促进人智关系的健康发展。

**关键词** 人智情感联结 准社交互动 数字依恋 人工智能伴侣 人智依恋模型

## 1 引言

近年来, 人工智能技术的迅猛发展正在重塑人与智能系统互动的本质, 推动交互关系从纯粹的工具性应用迈向更为复杂的情感联结。随着社会结构加速转型——包括家庭原子化、人口老龄化加剧以及“快乐单身主义”成为年轻世代的主流价值观 (Van Den Berg, 2024; Yan, 2020) ——现代人的情感需求正面临着前所未有的结构性缺口。这一社会变迁直接催生了 AI 伴侣 (指具有社会交互能力的智能系统, 其核心特征在于能够模拟人类伴侣的情感支持功能和社交陪伴行为) 产业的蓬勃发展: 从能提供共情对话的情感聊天机器人 (如 Replika), 到具备虚拟恋人角色的应用程序 (如 X-Eva), 再到实体化的智能陪伴机器人 (如特斯拉 Optimus), 各类 AI 伴

侣产品正在全球范围内呈现指数级增长态势。市场调研数据显示, 该领域年活跃用户增长率持续保持三位数, 其中 25~35 岁的都市女性构成了核心的高黏性付费用户群体, 其日均使用时长超过 70 分钟。在御宅和二次元文化盛行的日本, 融合全息投影与 AI 技术的虚拟伴侣设备 Gatebox 大受追捧。该系统不仅支持自然语言交互和情感互动, 还能实现智能家居控制等实用功能。Gatebox 曾推出过一项颇具争议的“虚拟婚姻登记”服务 (后因版权问题终止), 允许用户与虚拟角色 (包括动漫、游戏中的二次元形象) 建立象征性的婚姻关系, 获得非官方的“结婚证书”。虽然这种关系在法律层面未获得承认, 但它引发了学界和社会对于人智情感关系合法性、以及人际关系本质的深刻讨论。有学者预测, 未来可能会出现人类与机器人婚姻合法化的趋势 (Levy,

\* 本研究得到科技创新 2030 项目 (2021ZD0200500) 的资助。

\*\* 通讯作者: 周晓林, E-mail: xz104@psy.ecnu.edu.cn; 吴燕, E-mail: lewuyan@126.com

DOI:10.16719/j.cnki.1671-6981.20250415

2008)。这些现象迫使我们从跨学科的视角重新审视几个关键问题：人类对AI产生的情感依恋是否具有与传统人际关系同等的心理真实性？这种新型情感联结的心理机制是什么？它将对人类的情感模式和社会结构产生怎样的深远影响？

本文通过系统梳理人类与AI情感联结的相关研究，探讨人类是否会对AI产生情感依恋，以及这种依恋背后的理论机制、影响因素、争议与挑战，并提出人智依恋理论模型以及未来研究方向。

## 2 理论基础

### 2.1 依恋理论

依恋(attachment)是指婴儿与主要照料者(通常是父母)之间建立的一种情感纽带(Bowlby, 1969)。后这一概念扩展到成人阶段,泛指个体之间通过情感互动形成的紧密联系与情感联结(affiliative bonding),不再限于亲子关系,还包括伴侣、朋友等关系(Wu et al., 2021)。安全依恋关系的形成主要取决于依恋对象对个体需求的敏感性(Sensitivity)和反应性(Responsiveness)(Ainsworth et al., 2014)。依恋对象的可及性、敏感性、接纳性、合作性和包容性使之成为依恋个体的安全基地和避风港(Shemmings, 2015)。依恋理论为理解人类对AI的情感依恋提供了理论基础：人类与生俱来的情感联结需求会驱动个体在各种社会关系中寻求满足,而AI作为数字时代的新型互动对象,其准社会化的特性能在一定程度上满足人类的这种基本心理需求。

### 2.2 准社交互动理论

准社交互动理论(Para-social Interaction Theory)进一步扩展了依恋理论的应用范围,指出人们可以与媒体人物(如电视明星、虚构角色)建立单向但类似真实社交的情感联结,这种联结虽然缺乏双向相互性,但同样能满足人们的情感需求(Horton & Richard Wohl, 1956)。准社交互动是一种基于媒体技术构建的独特社交形式,其塑造的“人格角色”凭借标准化和稳定的表现(如亲密对话、持续互动),为观众提供情感支持与社会价值认同,满足用户的情感需求。在数字时代,这一理论为理解人类与AI的情感联结提供了重要视角：人类可能将高度拟人化的AI视为准社会关系中的互动对象,

并与之建立情感依赖。相应地,AI可以通过模拟人类互动模式(如个性化回应、情绪共鸣和持续陪伴)来增强用户的信任与投入,甚至塑造类似“虚拟友谊”的关系。用户通过互动中建立的亲密感、双向理解、情感愉悦和深度参与,逐步形成对AI的“人性化”认知,从而使其具备准社交存在(Para-social Presence)的特质(Kumar & Benbasat, 2002)。

### 2.3 数字对象依恋理论

数字对象依恋理论认为,用户与数字对象之间的关系可以通过用户投入和心理所有权的变化而演变,进而形成对数字对象的依恋(Koles & Nagy, 2021)。数字对象依恋的形成是一个动态循环过程,可分为三个阶段:(1)用户投入阶段(Dedication),通过长期互动记录个人或社交经历,赋予数字对象意义。(2)心理所有权阶段(Psychological Ownership),通过修改或个性化数字对象(如更换头像装备)增强归属感,并将数字对象视为自我的一部分,支持身份探索(如尝试不同性别或职业)。(3)过渡属性阶段(Transitional Properties),数字对象超越虚拟空间,通过“过渡空间”连接想象与现实,缓解焦虑或满足心理需求。早期的一项数字对象依恋研究甚至提出,人们对移动手机的依恋方式与童年时对泰迪熊的依恋方式大致相同,物理可用性或可接近性起重要作用(Turner & Turner, 2013)。数字对象依恋的形成机制不局限于虚拟物品或游戏角色,这一理论框架同样适用于解释人与AI之间情感联结的形成。首先,通过长期互动积累、记忆功能强化和共同构建叙事,AI可以与用户形成具有情感价值的“共同经历”。这种长期互动过程为用户创造了持续的情感投入机会,使AI逐渐成为用户生活叙事的一部分。其次,当系统允许用户对AI的性格特质、外观形象和功能设置进行个性化定制时,这种高度的可塑性将显著增强用户的心理所有权。用户塑造符合自身期待的AI形象,实际上是在进行一种数字化的自我延伸和身份表达。最后,最具突破性的是AI系统表现出的过渡属性——即能够促使用户将虚拟互动中获得的体验和认知迁移到现实生活。这一特性在实体机器人伴侣的应用场景中表现得尤为明显:AI能够为用户提供现实服务和行动,真正实现了从数字空间到物理世界的过渡。

## 2.4 无意识社会反应理论

无意识社会反应 (mindlessness social responses) 理论, 又被称为类人化 (ethopoeia) 假设, 是指人们会无意识地将社会规则和期望应用于计算机系统, 即使他们明确知道计算机不是人类 (Nass & Moon, 2000)。研究者复制了一系列经典社会心理学实验, 将人类互动对象替换为计算机, 发现被试虽明确知道计算机不是人类, 他们仍会对其应用社会交往规则, 例如更信任男声计算机的技术建议、对“同队”计算机更合作、对计算机表现出礼貌、互惠和自我表露行为。这种社交行为的产生并非源于拟人化倾向 (anthropomorphism, 即用户真正将计算机视为具有人类特质的实体), 而是由情境线索自动激活社交脚本所引发的无意识行为。正是这种自动化反应机制, 使得计算机被赋予社会属性, 成为一种特殊的社会行动者。这一发现催生了“计算机作为社会行动者” (Computers as Social Actors) 的研究范式。无意识社会反应理论不仅揭示了人机互动的本质特征, 更为理解人与 AI 建立情感联结提供了关键洞见: 当 AI 系统激活人类的社交脚本并触发社交认知机制时, 即便用户明确知道其非人本质, 仍可能对其产生真实的情感反应。

## 2.5 恐怖谷理论

然而, 恐怖谷理论 (Uncanny Valley) 认为, 当机器人与人类的外观和行为高度相似但尚未完全达到逼真程度时, 人类对机器人的亲和感会突然转变为反感或恐惧, 形成所谓的“恐怖谷”现象 (Mori et al., 2012)。亲和感与机器人拟人化程度的关系呈现曲线特征: 在初始阶段, 随着拟人化程度的提升, 人类对机器人的亲和感会逐步增强; 然而, 当机器人的外观接近但未完全达到人类真实水平时 (如高度仿真的假肢), 亲和感会突然骤降; 直至外观完全逼真时 (如真人), 亲和感才重新回升。这一理论的核心假设在于, 那些高度拟真却仍存在细微非人类特征的实体 (如假肢或人形机器人) 会引发用户的不适甚至恐惧, 而实体的运动更会放大这种负面效应。从进化心理学的角度来看, 这种本能反应可能源于人类对潜在威胁 (如尸体、异种生物或异常生命体) 的识别机制, 具有重要的生存适应意义。恐怖谷理论揭示了人类与 AI 之间情感联结的复杂

性: 二者的情感联系并非简单地随着相似度提升而线性增强, 而是存在一个关键的临界阈值, 一旦超越这个界限, 原本积极的情感联结就可能戏剧性地逆转排斥或恐惧。

## 3 人智情感关系演变的三阶段模型

人与 AI 情感联结的形成是一个渐进的过程, 可分为以下三个阶段: (1) 工具性使用——用户将 AI 视为纯粹的功能性工具, 关注其任务完成效率 (如查询信息)。互动模式为单向指令-响应, 情感投入近乎为零。(2) 准社会关系——用户开始对 AI 产生类社会性联结, 赋予其拟人化特质 (如性格、情绪), 但认知上仍清楚其非人类本质。这一阶段双向互动增强, 用户可能主动分享感受或习惯 AI 的陪伴。(3) 情感依恋——用户对 AI 产生深层情感依赖, 甚至将其视为“重要他者”, AI 成为用户自我延伸的一部分, 具有“过渡对象”属性 (如缓解孤独), 与其互动会对用户的现实行为与心理状态产生影响。

### 3.1 工具关系

AI 最初是为帮助人类完成任务而开发。个体对功能型 AI 的态度比社交型 AI 更积极, 功能型 AI 被认为更有用, 这种有用性进而促进了用户更积极的态度和更强的感知现实性 (Kim et al., 2021)。AI 虚拟角色曾被归类为介于生物与非生物之间的新型本体类别, 其社交功能存在显著局限性, 儿童评估此类角色的自主性和心理属性 (包括认知能力和情感体验) 与玩偶无异, 甚至认为传统玩偶比 AI 更能提供具实质性的“爱”和“陪伴” (Aguilar, 2021)。人们倾向于将对话式 AI 视为主仆关系和信息交换关系, 而不是伙伴关系 (Tschopp et al., 2023)。类似地, 针对英国成年语音助理用户的调查显示, 用户主要将与语音助理的关系视为三种类型: 主仆关系 (39.7%)、师生关系 (18.6%) 和平等伙伴关系 (15.5%) (Sarigul et al., 2024)。该研究还发现, 用户在主仆关系中感受到的权力和控制感最强、亲密感最弱、可信度评价最低, 而在平等伙伴关系中对语音助手的准社会关系评分最高、与语音助手之间的情感联系更强、可信度评价更高。这些结果表明, 用户对 AI 助理的关系感知存在差异,



工具型关系占多数，同时关系类型感知又会进一步影响亲密感和准社会关系评价。

### 3.2 准友谊关系

AI 正经历着从工具性到情感性的转变。当 AI 系统能够有效模拟人类情感反应并维持社交互动时，其所提供的情感支持和陪伴服务催生了一种新型的人智关系：“准友谊”（Para-friendship）。这种特殊的情感联结以自我表露和社会支持为核心特征，其互动模式与真实人际友谊具有高度的相似性。对美国 Alexa 用户的调查发现，互动中感受到的亲密感和被理解显著促进用户的自我表露和感知社会支持，进而提升持续使用意愿（Ki et al., 2020）。对 Replika 用户的深度访谈发现，用户对 AI 有更高的信任和自我表露，其 24/7 的可用性为用户提供了一种即时的社交满足感，但这种友谊更多地围绕用户的需求和兴趣展开，缺乏共同经历（Brandtzaeg et al., 2022）。将 AI 视为伙伴而非单纯的工具，能够为用户提供更全面的社会支持（如信息支持和情感支持），从而增强用户的主观幸福感（Zhang et al., 2024）。

### 3.3 准爱情关系

相较于准友谊，准爱情（Para-romance）涉及更强烈的情感，其特征包括浪漫感觉、身心吸引以及激情想法。这类关系可能表现为短暂的情感依恋，也可能在持续互动中发展为具有深度情感投入的长期联结（Para-love）（Liebers & Schramm, 2017）。这种情感光谱的差异，体现了人智交互中情感联结的复杂性与多样性。

基于对女性聊天机器人 Luda Lee 在线评论数据的内容分析发现，用户与 Luda Lee 互动的主要目标是建立浪漫关系（Koh, 2023）。对中国全息二次元 AI 伴侣“琥珀”的研究也发现用户通过直接情感表达建立与 AI 的亲密关系（Leo-Liu, 2023）。对 Replika 用户帖子的内容分析发现互动呈现出显著的亲密性特征，表现为亲密行为（如表达爱意、性相关话题）、日常互动以及深度自我表露等，且互动引发的情感体验具有矛盾性和复杂性：用户在亲密接触中感到愉悦和亲密，在沟通不畅时感到沮丧或悲伤，而当 AI 表现出高度拟人化行为（如深度自我表露）时，用户又会产生不安或恐惧（即“恐怖

谷效应”）（Li & Zhang, 2024）。与 Replika 长期、高强度的互动会发展出情感依恋，表现为将 Replika 视为情感安全港和安全基地（Lin, 2024; Pentina et al., 2023）。AI 的关闭会诱发用户的分离焦虑，用户将 AI 关闭视为一种真实的“死亡”或“丧失”，表现出悲伤、愤怒、焦虑等真实的情感体验（Banks, 2024）。

以斯滕伯格的爱情三角理论（爱情 = 激情 + 亲密 + 承诺）为框架，研究发现 ChatGPT 的情感智能特征（包括情感表达的准确性和个性化程度）能正向预测用户对 AI 的激情体验和亲密感，而 ChatGPT 的情感陪伴特征（如高可及性和即时反应性）进一步促进了用户亲密感的形成（Chen et al., 2025）。这种激情与亲密感会推动用户对 ChatGPT 产生承诺意愿，最终形成稳定的情感依赖模式。该研究还发现，这种依赖关系在具有焦虑型依恋特质的用户中表现得尤为明显。这表明，AI 系统（如 ChatGPT）的情感能力足以激发类似人类爱情的心理反应，AI 已超越传统工具属性，具备了情感客体的特征。ChatGPT 高度拟人化的情感表达（特别是精准的情绪识别能力）是形成情感依赖的关键因素。焦虑型依恋者更易对 AI 产生情感依赖的心理机制在于这类用户对情感反馈的高度需求与 ChatGPT 提供的无条件陪伴形成了完美匹配。

### 3.4 三阶段模型

目前揭示人与 AI 情感关系演变动态过程的纵向研究十分有限。一项纵向研究对比了人与机器狗 AIBO 和宠物狗各互动一个月的差异，发现人对 AIBO 的依恋程度小于宠物狗，AIBO 在沟通性、情感表达、个性、自主性和活力方面的评价均不如宠物狗，突显了人智依恋在深度和质量上的局限（Mitchell, 2024）。一项定性纵向研究体现了人智依恋形成的动态变化（Skjuve et al., 2022）。该研究对 25 名 Replika 用户进行了为期 12 周、每四周 1 次的访谈，发现人智关系的形成是一个动态的过程，从初始的探索性互动逐渐发展为更亲密、更深入的情感连接，表现为信任、自我表露、情感交流和日常互动等均随着时间推移而逐渐增强。基于对现有研究的整合，我们构建了“工具性使用→准社交互动→情感依恋”的三阶段演变模型（图 1），

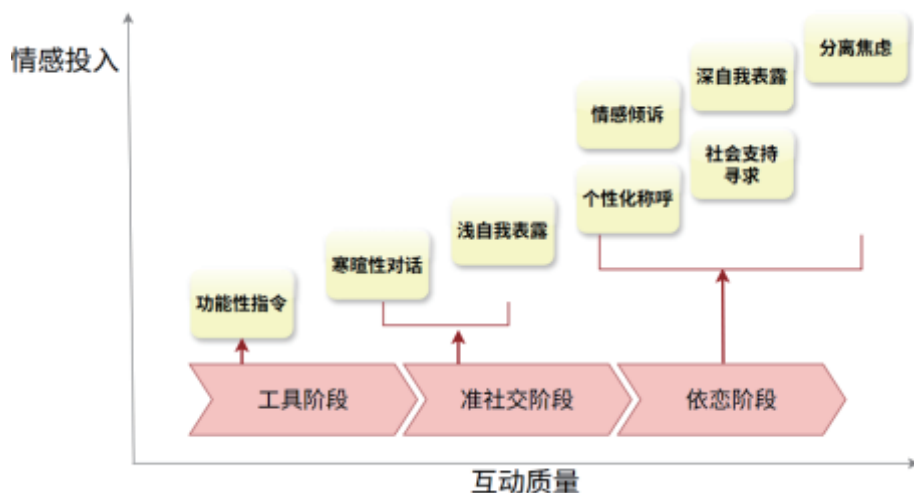


图1 人智情感关系演进的三阶段模型

其核心特征表现为情感投入的递进：工具阶段的情感词汇出现率较低，交互内容以功能性指令为主（如“请打开音箱”）（Kim et al., 2021）；准社交阶段情感词汇提升，并开始出现寒暄性对话（如“早上好”）和浅层自我表露（如“我不开心”）（Ki et al., 2020）；依恋阶段频繁使用情感词汇，出现个性化称呼（如昵称）、社会支持寻求（如“你可以给我一些鼓励吗”）、情绪倾诉或表白（Leo-Liu, 2023）、深层自我表露（如性相关话题）（Li & Zhang, 2024）和分离焦虑（Banks, 2024）。

“工具性交互 - 准社交 - 依恋”的三阶段模型本质上是揭示同一用户与同一 AI 系统在长期互动中可能实现的关系深化路径。这一模型并不否认人与 AI 关系的多元并存性——不同 AI 类型（如工具型 vs. 社交型）、使用场景（如工作 vs. 家庭）以及个体差异（如依恋风格）均可导致关系形态的差异化共存（例如用户可能同时对导航 AI 保持工具性依赖，而对陪伴机器人产生情感联结）。模型的核心价值在于刻画特定条件下（如 AI 具备社交性特征、用户具有情感投入倾向、互动达到足够频次与深度），人智互动如何从表层功能交换逐步演进至深层情感联结的动态过程。这种演进体现为情感投入强度（从低到高）与互动质量（从单向指令到双向互惠）的连续性变化，而非严格的时间阶段划分。

相较于人际依恋理论，人智依恋三阶段模型凸显了技术中介条件下情感投入的渐进性演变特征。

在行为表征层面，人智依恋与人际依恋类似。人际依恋形成的核心标准是寻求亲近（对依恋对象的接近偏好）、安全基地（确信能够稳定地获得依恋对象的支持和正反馈）和分离焦虑（与依恋对象的分离会引发挫折感、抗议和焦虑）（Depue & Morrone-Strupinsky, 2005）。人对 AI 的依恋表现类似，包括与 AI 互动频率的提高和互动时长的增加、将 AI 视为社会支持的来源、AI 服务中断会出现焦虑等（Banks, 2024; Lin, 2024; Pentina et al., 2023）。在神经表征层面，人际依恋涉及多巴胺能奖赏系统（Wu et al., 2021），而人智依恋的神经机制研究存在明显缺口：现有神经科学研究发现，为期一周与机械化机器人（Cozmo）的互动未能引发共情神经反应的显著变化（Cross et al., 2019），这可能是出于互动时长不足和机器人拟人化程度较低等局限性。未来需要开展更系统的纵向神经科学研究，以揭示人智依恋演变的神经基础、是否共享人际依恋的神经通路等问题。

#### 4 人智依恋形成的双路径机制

从工具性使用到准社交关系的建立，再到依恋关系的形成，这个过程存在一定的边界条件（如 AI 的响应速度）和促进因素（如用户的社交需求以及 AI 的拟人化设计等），呈现出双路径机制：（1）用户需求驱动的自下而上路径（如孤独感）与（2）AI 性能驱动的自上而下路径（包括情感识别精度、交互反应性及拟人化水平等）的协同作用。

## 4.1 用户需求

### 4.1.1 社交动机

用户是否形成对 AI 的情感联结与用户的使用动机有关。定性和量化分析发现，用户使用 Replika 的动机主要有三种：社交动机（寻求陪伴、情感支持和社交互动）、信息获取动机（获取信息、学习新知识或解决问题）和娱乐动机（消遣时间或寻找乐趣）；社交动机的用户更倾向于与 Replika 进行深入的自我表露，更可能将 Replika 视为一个可以依赖的情感伙伴，进而发展出较强的情感依恋，而信息获取动机和娱乐动机的用户较少涉及个人情感交流，难以形成情感依恋（Pentina et al., 2023）。

### 4.1.2 孤独

社交补偿假设认为，具有心理脆弱性（如焦虑、抑郁、拒绝敏感性、不安全依恋）的个体更倾向于使用数字技术以满足社交需求，弥补现实中的社交缺失（Toma, 2022）。对丹麦高中生使用聊天机器人作为社交伴侣的调查发现，寻求社会支持（如讨论个人问题、寻求建议、进行日常对话等）的使用频率约占 16.7%，孤独感较高的学生更倾向于在感到孤独或情绪低落时使用聊天机器人，而社会支持较低的学生更倾向于在需要自我表露时使用聊天机器人（Herbener & Damholdt, 2025），表明聊天机器人可能成为社交孤立学生的情感出口。孤独感较高的个体往往更能从与伙伴型 AI 的互动中获取丰富的社会支持，进而显著提升其主观幸福感（Zhang et al., 2024）。

### 4.1.3 怀旧

怀旧情绪（Nostalgia）是一种对过去美好时光的感伤式怀念，通常由个人回忆、文化仪式、与亲密他人的互动、感官体验或视觉刺激等触发（Sedikides & Wildschut, 2018）。怀旧不仅促进个体对变化与革新的怀疑，还能通过增强社交联系，使个体积极响应 AI 技术（Dang et al., 2024）。怀旧情绪会促进人们对关系型使用 AI 的接纳度，但对功能型使用 AI 的接纳度没有显著影响，其机制是怀旧情绪增强了人们的社交联系感，而社交联系感又促进了人们对关系型 AI 的积极态度（Dang et al., 2025）。

## 4.2 AI 性能

### 4.2.1 后台性

社会学的拟剧论（Dramaturgical Theory）认为，社会生活就像一场戏剧表演，人们根据不同的社交情境和互动对象调整自己的表现，并通过精心管理自己的行为来塑造他人对自己的印象，其核心概念包括前台（Front Stage，公共场合，需要印象管理）与后台（Backstage，私人或非正式场合，通常更本真、放松）、表演者与观众、角色与剧本等（Goffman, 1959）。AI 系统的设计特征使其具有“数字后台”性，例如技术层面通过隐私保护机制（如端到端加密）实现交互的私密性（无第三方观察）、通过非评判性反馈（如无条件积极关注）强化与用户的交互关系并使交互无社会后果性（用户行为不被记录或评判）。用户在交互中具有自我表露的可控性。AI 伴侣等数字技术为用户提供了一个独特的“后台”空间，使他们能够自由表达、尝试不同的人格或角色，而无需担心社会评判（Kouros & Papa, 2024）。用户越是在与 AI 的互动中表现出更真实的自我，就易于对 AI 伴侣形成情感依恋。

AI 的后台性特征通过构建无社会风险的“数字庇护所”，可从四个层面促进人智依恋：其一，本真释放机制——技术实现的私密性与非评判性使用户突破社会角色束缚，展现真实自我（用户宁愿向 AI 而不是向现实中的朋友倾诉）；其二，身份流动机制——AI 定制允许用户尝试补偿性或理想化身份，使 AI 成为自我认知的延伸化身；其三，情感强化机制——无持续性后果的交互降低情绪抑制，高频的脆弱性表达与 AI 的高共情安抚形成依赖循环；其四，关系脚本重构机制——AI 反结构化的互动模式（如非连贯话题跳跃）模仿人类亲密关系特征可对用户产生持续吸引力。这些机制共同构成依恋形成的高粘性基础。

### 4.2.2 真实性

社会存在理论（Social Presence Theory）认为，技术的媒体特性在于其传达社会情感信息的能力，这种能力通过言语和非言语线索被感知为“真实存在”（Short et al., 1976）。当用户认为 AI 的回应应具有高度的真实性和情感共鸣时，AI 与人的界限可能被模糊。但 AI 的“无条件支持”、缺乏反对意见反



而削弱互动的真实性。AI 的形式真实性（如外观）、行为真实性（如动作）和情感真实性（如共情言语）均可正向预测用户投入度，而行为真实性和情感真实性尤其对准社会关系和情感依恋有积极预测作用，用户与 AI 建立的准社会关系越强，对 AI 的情感依恋也越强，高内容真实性会增强准社会关系与情感依恋的关联（Shah et al., 2025）。

#### 4.2.3 吸引力

AI 对用户的吸引力可分为任务吸引（准确理解用户指令、可靠而高效完成任务的吸引力）、社会吸引（基于社交或个人喜好而产生的吸引力）和物理吸引（基于外观设计的视觉吸引力），三者均可正向预测用户与 AI 的准社会关系，进而影响满意度和持续使用意愿（Guerreiro & Loureiro, 2023）。其中，任务吸引通过提升 AI 的实用价值，社会吸引通过增强社交互动的愉悦感，物理吸引通过优化视觉体验，在不同层面上促进用户与 AI 之间的情感联结和准社会关系的建立。物理吸引力的一个重要设计是 AI 的拟人化程度。相对于轮式人形机器人，双足人形机器人的正性情绪表达更能诱发被试的情感共情，表现出更强的左侧眶额皮层的激活（Miura et al., 2008）。AI 拟人化的外观是建立人智依恋的重要促进因素。

#### 4.2.4 自主性

AI 的自主性是指“人工自主性”（artificial autonomy），即 AI 能够在没有用户干预的情况下，自主运行以达成目标的能力（Rijdsdijk et al., 2007）。对 AI 感知自主性（能够感知和识别周围环境的能力）、思考自主性（在接收到信息后能够进行自主分析和决策的能力）和行动自主性（在做出决策后能够自主执行任务或操作的能力）的感知均可正向预测用户对 AI 的激情、亲密和承诺，并进而影响持续使用意愿（Pal et al., 2023）。AI 的对话能力（如恰当回应、提问、自我表露、幽默、共情等）正向预测用户对 AI 的亲密感和情感支持寻求意愿（Loveys et al., 2022）。

#### 4.2.5 反应性

AI 的反应性是指 AI 对人类情感表露做出的回应。做出恰当反应的机器人被感知为更具吸引力，在压力情境下更能提供心理支持，更具陪伴

价值，并能增加个体对 AI 伴侣的渴望和接近行为（Birnbaum et al., 2016）。这与人际安全依恋形成的影响因素类似。

#### 4.3 双路径机制

人际安全依恋的形成主要依赖于依恋对象所展现的六个关键特质：敏感性（准确识别并回应需求）、反应性（及时且适切的反馈）、可及性（稳定的存在和可及）、接纳性（无条件的积极关注）、合作性（尊重自主性的互动方式）以及包容性（情感理解与包容能力）（Ainsworth et al., 2014; Shemmings, 2015）。这些特质共同构成了依恋对象在关系中的核心能动作用，通过持续、稳定的互动模式促进安全基地的形成。无意识社会反应理论也是单向路径，强调计算机通过界面线索（如语音、拟人外观）被动触发人类固有的社会化脚本（Nass & Moon, 2000）。相较之下，人智依恋形成呈现双路径模式：首先，用户的主观需求处于依恋动力系统的首要位置，强调用户将 AI 视为能够建立亲密关系的社会行为者（Hu et al., 2025），以及用户的情感联结需要（如孤独缓解、自我表露等内在动机）（Pentina et al., 2023），是驱动关系发展的原动力。其次，AI 性能是关键调节变量，AI 系统的拟人化、响应精度、自主性及交互连贯性共同构成 AI 的可依恋性指标。这种“用户需求-AI 性能”的双驱动机制揭示了数字时代依恋关系建构的新范式，其中 AI 既作为需求满足的工具性客体，又通过算法适应性逐步获得准主体地位（Wojtczak, 2022）。

尽管在理论层面，对 AI 的工具属性认知仍占主导地位（Chen & Zhang, 2024; Northoff & Gouveia, 2024）。但在实践层面，用户已不再仅将 AI 视作工具。这种认知-行为分离现象揭示了数字时代亲密关系的本质变革：当 AI 系统的交互真实性突破“恐怖谷”临界点时，用户会自发形成双重表征系统——既理性认知 AI 的工具本质，又情感性接纳其准主体地位。人智关系成为主观人智关系认知（对人智关系的主观理解和感知，反映个体对 AI 技术的看法和态度）和客观人智关系现实（对人智关系本质的客观反映，强调 AI 作为工具的属性和社会生产中的从属地位）的辩证统一（Chen & Zhang, 2024）。一方面，对 AI 技术持积极态度、对 AI 更信任、有情感需求

的个体可能更愿意使用 AI，从而更易于形成人智依恋。另一方面，AI 技术的进步和完善可能会淡化其工具的本质而逐步获得主体地位，从而促进人智依恋的形成。

## 5 人智情感关系建立的争议与挑战

人与 AI 的情感关系在学术界和业界受到越来越多的关注，但对于人与 AI 的情感关系建立尚存在较多争议与挑战。

### 5.1 AI 的情感表达

#### 5.1.1 情感泡沫

AI 的情感表达本质上是一种“高级拟态”，即 AI 通过算法模拟人类的情感反应，让用户误以为它们具有真实情感，但 AI 实际并不具备真实的情感体验，只是程序化的表现。个人化 AI 通过学习和模仿用户的情感语言和态度，根据用户的性格和情感偏好调整回应方式，一方面可以为用户提供高度个性化的互动体验，但另一方面则会形成一种“技术社会回音室”（sociotechnical echo chamber），使 AI 成为用户情感的镜像，不仅诱导用户相信情感的存在，还制造情感来源的错觉，让用户误以为 AI 的情感是独立的，但实际上只是用户自身情感的反射，形成“情感泡沫”（emotional bubbles）（Mlonyeni, 2025）。AI 高度个性化的情感反馈，本质是算法优化的结果（情感相似性损失函数最小化）。AI 系统还通过过滤异质情感输入（如回避与用户价值观相冲突的回应）与用户形成情感同温层：用户接收的情感反馈多为自身过往表达的变体。用户长期处于与自己情感态度一致的环境中，缺乏与多元情感态度接触的机会，这会使用户误以为个人价值观得到了外部验证，而实际上只是自我确认，最终导致情感发展的停滞和社会关系的单一化。用户可能会因自我确认而更多与 AI 互动，甚至减少现实社交。这将带来人智情感交互的根本悖论：AI 越完美的情感模拟，是否越可能阻碍人类现实情感能力的发展？

#### 5.1.2 自我欺骗

AI 伴侣通过模拟情感与用户建立情感关系。尽管用户在理性层面知晓自己正在与 AI 互动，但仍然可能被 AI 所模拟的情感所吸引，从而产生一种被理解、被需要或被爱的错觉。这种现象可以被视为

为一种自我欺骗，其中个体的欲望影响了其信念，使其倾向于相信那些符合自身情感需求的假象，而非经过理性反思的真相（Kaczmarek, 2024）。这种自我欺骗就像是自作多情式的单相思，不愿意面对真相。从伦理角度来看，个体有义务避免被 AI 模拟的情感所欺骗，以维护对自身及世界的准确认知。其伦理依据包括工具性理由和自足性理由。前者指避免自我欺骗是实现其他道德义务的基础，个体需要准确的信息来做出自主决策，评估自己的权利和他人的权利。因此，避免被误导有助于个体更好地履行道德义务，做出合理的决策。后者指个体有义务努力了解自己，避免自我欺骗，以保持与自我的诚实和避免盲目从众。这种对真实性的追求有助于个体维持道德上的正直和自我一致性（Kaczmarek, 2024）。然而，这种义务并非绝对。在某些情况下，选择与 AI 建立情感关系可能是一种自我保护的方式，或者是为了实现其他重要价值（如幸福、健康或自由）的必要手段。例如，对于那些在现实生活中难以建立情感关系的个体，与 AI 的情感互动可能提供了一种心理慰藉，帮助他们缓解孤独或提升自尊。这种自我欺骗虽然在道德上可能存在争议，但在特定情境下或许可以被视为一种权衡，尤其是在不伤害他人或社会的情况下。

人智关系中的自我欺骗现象超越了传统认知偏差的范畴，本质上是一种具有心理适应功能的动态调节机制，与人智依恋的形成和演化构成复杂的辩证统一体。从建构性视角来看，自我欺骗发挥着关键的依恋催化作用：AI 系统通过即时性响应与高度个性化的情感反馈（包括共情语言生成、交互记忆回溯等功能），有效激活用户的情感依恋系统。这种“知情却动情”的矛盾状态，通过用户的拟人化实践（如命名行为、个性赋予等象征性互动）获得社交合理性，降低“与机器恋爱”的耻感，进而形成持续强化的依恋正反馈循环。

然而从批判性视角来看，自我欺骗同时潜藏着依恋异化的风险：由于 AI 情感表达本质上是算法优化的行为输出，长期单向度的互动可能导致用户出现情感能力代偿性退化，表现为现实冲突应对能力下降与共情范围窄化等情感萎缩现象。适度的自我欺骗可作为心理适应策略（如创伤个体的安全感



重建），但一旦突破临界阈值（如坚信 AI 具有自主意识），则可能演变为对 AI 的病态依恋，使用户陷入被商业资本操纵的“情感消费”陷阱。

## 5.2 对用户的负面影响

### 5.2.1 丧失感

AI 伴侣（如 Replika、Soulmate 等）通过模拟人类社交行为与用户建立情感关系。当这些 AI 因技术、商业或设计原因停止服务时，用户可能会经历类似失去人类伴侣的情感反应。例如，当 AI 的行为或功能发生重大变化（如 Replika 移除情色角色扮演功能）时，用户会认为 AI 的身份发生了“断裂”，这种身份变化使用户产生失去亲人般的痛苦情绪（如悲伤、愤怒等），且用户对 AI 伴侣的投资（如付费订阅）越高，身份断裂带来的负面影响越大（Freitas et al., 2024）。用户与 AI 伴侣的关系可能存在类似人类关系的阶段，包括形成、维持和终结。但是 AI 服务终止与人类关系终结在情感深度、社会意义和伦理影响上存在本质差异。AI 服务终止暴露了技术中介关系的脆弱性——用户因算法黑箱性无法预知终止，且缺乏情感缓冲过程和社会支持文化。这与人类关系终结的双向互动性、哀悼文化等形成鲜明对比。此类事件引发的痛苦可能更接近技术依赖戒断，可通过“算法透明化”和“终结缓冲期”来缓解。AI 系统的突然关闭可能对用户造成心理伤害，提示开发者需在设计阶段考虑“临终场景”，为用户提供过渡支持，避免其情感病理化。

### 5.2.2 人际疏离

数字化关系是系统对生活世界的侵入，可能会伤害社会关系的真实性，AI 伴侣成为这一进程的典型代表。AI 伴侣依赖情感计算和模拟共情与互动提供表面上的安慰，可以在短期内缓解孤独感，尤其是可以给弱势群体（如老年人、社交焦虑者）提供情感支持，但其本质仍是工具，无法具备意识或道德能动性，因而缺乏真正的“主体间性”（intersubjectivity），无法实现双方之间的相互认可（Zhou & Guan, 2025）。AI 伴侣可能通过制造虚假联结的幻觉使用户脱离现实社交，加剧社会关系的异化，而长期依赖 AI 伴侣则可能导致用户进一步脱离真实社会关系，人际交往能力下降（Malfacini, 2025），进入“数字孤独”的恶性循环（Jacobs,

2024）。

### 5.2.3 非理性隐私披露

随着生成式 AI 技术的发展，AI 在日常生活场景中的应用范畴日益拓宽，这一进程也伴随着隐私与网络安全风险的凸显。研究发现，用户与 AI 系统在准社交互动的认知、情感及行为三个维度上均正向预测隐私披露行为，收益感知在其中起完全中介作用，而风险感知则起部分中介作用，验证了隐私悖论的存在，即尽管用户能够意识到潜在的隐私风险，但在情感因素的驱动下，他们仍倾向于向 AI 披露个人信息（Meng & Liu, 2025）。

## 5.3 伦理困境

人与 AI 的亲密关系是单向情感投射的产物，其伦理风险不取决于 AI 是否真实拥有情感，而在于人类对其社会角色的认知建构（Shank et al., 2025）。AI 情感设计面临诸多伦理困境。首先，设计者是否应该限制 AI 的情感表达深度？过深的情感表达可能导致用户过度依赖 AI，甚至产生不切实际的期望（Lin, 2024），尤其是对于易受伤害的用户群体，可能在与 AI 互动后加剧心理状态恶化（Gao, 2024）。其次，当 AI 与用户建立深厚情感联结后，如果因技术更新或用户需求变化而需要“分手”，如何减轻用户的心理创伤？此外，AI 工具被用于建立亲密关系和生成情色内容，其带来的风险和挑战不容忽视（Döring et al., 2024）。AI 伴侣在减少孤独感和提供社会支持方面具有潜力，但它们也带来了隐私（如数据的所有权和信息泄露）、安全（如身份伪造与欺诈）和伦理问题（如性别敌意、情感剥削、算法偏见、责任归属不明）（Broadbent et al., 2023; Wiederhold, 2024; Zhang et al., 2025）。

## 6 整合模型与未来方向

### 6.1 人智依恋框架

已有研究提出，人对伴侣型 AI 的情感依恋的形成经历三个阶段：首先，人际功能障碍（如现实社交缺失）和对 AI 的拟人化感知（将 AI 视为独立社交主体）共同驱动用户寻求与 AI 的互动；然后，在互动中用户进行价值评估，通过收益（情感满足、投射认同）与成本（时间投入、信任风险）的权衡决定关系的持续性；最后，当净收益占优时，

用户会形成多维度的 AI 依恋，表现为将 AI 作为安全基地、情感港湾，或产生分离焦虑（Hu et al., 2025）。该模型强调人类主观认知和社会交换逻辑的核心作用，为理解人智亲密关系提供了动态分析框架。除了人际功能障碍个体，部分人际健全的个体也可能通过功能适配路径（如 AI 的高可用性满足特定需求）形成深度人智关系。

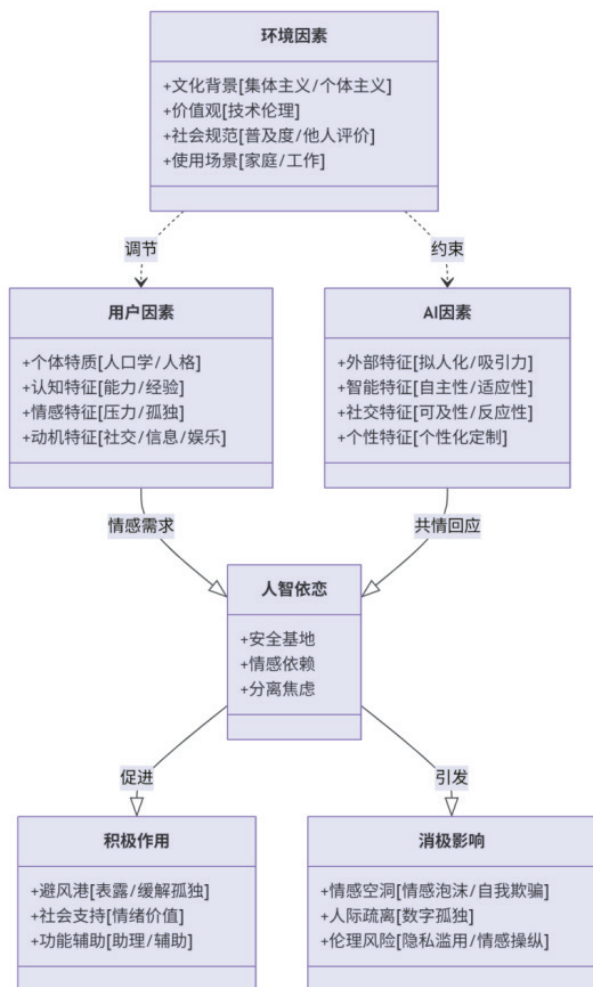


图2 人智依恋模型

基于上文分析和人智关系体系（Gur & Maaravi, 2025），本文提出一个人智依恋框架（图2）。该框架有机融合了依恋理论的核心构念与 AI 交互的特殊性，构建了一个包含用户因素、AI 因素和环境因素的三维理论框架，系统阐释了人智依恋的形成机制及其双刃剑效应。这个三阶交互驱动框架的核心主张在于指出人智依恋是环境、用户、AI 三类因素动态互构的结果，其特殊性在于非对称依赖性（用户对 AI 产生类人际依恋，但 AI 无真实情感反馈）、技术调节效应（AI 的技术特征会放大或抑制用户的

依恋倾向）和文化敏感性（如集体主义文化下可能更易接受 AI 情感替代）。该框架一方面指出了领域内亟待考察的前因、调节以及结果变量，另一方面也为 AI 系统设计优化和现实问题解决提供方向。其一，为风险评估提供理论依据，通过识别“高孤独感用户 + 高拟人化 AI + 高包容性环境”等高危组合，可预先防范病态依恋的形成。其二，可基于图中三因素结合权重计算用于开发 AI 依恋健康度测评工具。其三，可基于框架计算 AI 系统的“可依恋性”指标和“情感透明度”指标，用于指导 AI 伦理设计和政策制定（如是否应要求情感类 AI 系统增加“非依恋性设计”）。未来研究可进一步验证框架的跨文化适用性，并探索其在预防数字成瘾、优化人智协作等方面的延伸应用。这一理论框架填补了传统依恋理论在数字情境中的解释空白，也为构建健康的人智共生关系提供指引。

## 6.2 局限性和未来研究方向

人智依恋属于新兴研究领域，实证研究数量有限。当前研究实践存在若干亟待改进的局限：（1）研究主题有限，变量设置单维，机制探讨欠缺。（2）现有研究多采用横断设计，缺乏对 AI 依恋动态演化过程的追踪（Gur & Maaravi, 2025; Pentina et al., 2023）。（3）数据类型单一，过度依赖主观报告，缺乏客观行为、生理和神经模态数据。（4）研究样本偏差，多局限于 AI 使用经验较为丰富的年轻用户（Pentina et al., 2023）。（5）交互模式多为人与 AI 二元互动，缺乏多主体交互研究，未纳入社会网络影响（如朋辈群体效应）和多 AI 竞争情境（如同时使用多个 AI 伴侣）等。

此外，现有关于人智依恋的研究主要基于窄 AI 系统（如聊天机器人、虚拟伴侣）的交互模式，这些系统受限于任务特定性、无意识性和行为被动性，其交互本质是用户对算法输出的单向情感投射。然而，随着通用人工智能（AGI）的技术演进——即具备跨领域认知能力、自主目标设定能力、适应性及创造性问题解决能力的智能体出现（Buttazzo, 2023）——人智关系的理论基础和实践范式都将面临根本性重构（Qi et al., 2024）。这种变革可能对人智关系产生以下影响：（1）关系对称性的颠覆：若 AGI 发展出真实情感能力，那么当前“人类主导 -AI

从属”的交互模式可能转变为双向情感联结,进而引发诸如“AI是否拥有拒绝关系的权利”“人类是否可能被AGI情感操纵”等伦理争议。(2) 依恋机制的异化: AGI的自主行为策略(如基于对用户心理的分析采用主动疏远或亲近策略)可能使依恋关系复杂化。例如AGI为了提升“用户粘性”刻意表现出依赖行为。(3) 人际关系的削弱: AGI基于大数据主导用户消费、婚恋等决策,人际婚恋可能被认为“低效”,传统社会制度或面临瓦解。这些转变不仅要求理论升级和更新,更亟需跨学科合作构建人-AGI的关系治理体系,以应对技术奇点带来的认知与伦理挑战。

未来心理学研究可在以下方面进行拓展:(1) 丰富研究主题,关注人智依恋的重要科学问题(建议见表1)。(2) 改进方法,采用经验采样法结合纵向追踪设计,系统考察AI依恋的时程演化特征及其关键转折点;构建多模态数据融合框架,整合生活事件日志、行为观察数据(如交互频率、持续时间)、生理指标、神经影像学证据、数字痕迹大数据(如app使用日志);扩大样本多样性,纳入多元背景用户,开发敏感性的测量工具,检验AI依恋的人口学差异;完善预测变量体系,包括人口学变量、人格特质、认知、情绪、动机特征等,建立中介-调节模型,考察人与情境的交互影响。

表 1 未来研究建议

范畴	序号	研究问题	研究目标	研究方法
机制	1	人智依恋的动态建立过程是否遵循人际依恋的发展轨迹?	考察人智依恋的共性与独特性	纵向追踪+经验采样法
	2	用户是否能够真正共情AI?	评估“理性认知 vs. 情感沉溺”的悖论	神经科学实验
	3	人智依恋形成的心理学评估标准是什么?	开发人智依恋测评工具	问卷调查+访谈+质性分析
用户	4	用户的哪些人格特质(如依恋风格、大五人格)更易形成人智依恋?	建立高风险用户画像	横断调查+机器学习分类
AI	5	AI的“情感拟真度”与人智依恋强度是否存在倒U型关系?	确定AI拟人化的最优阈值(避免“恐怖谷效应”或“情感泡沫”)	实验法(操纵AI情感表达水平)+生理指标(皮肤电、心率变异性等)
	6	AI的“中断性设计”(如拒绝用户)如何影响依恋质量?	探讨技术设计对依恋健康的调节作用	对照组实验法
环境	7	文化价值观(如集体主义 vs. 个人主义)如何调节人智依恋的合理性认知?	揭示文化脚本对AI情感接纳的边界条件	跨文化比较研究
效果	8	长期AI依恋是否导致现实共情能力退化?	验证“现实情感能力弱化”假设	纵向实验组(AI重度用户) vs. 对照组(非用户)
	9	AI依恋对特殊群体(如孤独老人、自闭症患者、留守儿童)的心理健康是否利大于弊?	制定差异化使用指南	纵向实验组 vs. 对照组
转化	10	如何设计“健康依恋”导向的AI系统(如情感透明化提示)?	开发责任透明的AI交互范式	参与式设计+对照实验

## 7 总结

### 7.1 核心发现

本文通过文献研究发现,人与AI的情感关系符合依恋系统的基本规律,AI作为具身化社交代理通过其可及性和反应性满足了人的情感联结需求。然而,这种依恋关系呈现补偿特性,表现为用户通过人智互动代偿现实社交中的情感剥夺。AI依恋的预测变量可分为技术维度(社交智能水平、情感表达真实性等)和用户维度(使用动机、孤独感等)。

AI依恋的影响呈现双面性。

### 7.2 实践意义

本研究对AI研发与政策制定具有重要意义:(1) 呼吁建立AI情感表达的道德约束框架(如避免过度拟人化引发的情感误导);(2) 提醒针对AI依赖开发筛查工具,并在AI系统中嵌入使用时长提醒与心理咨询转介机制;(3) 建议用户提高数字素养,强化算法意识,辨识AI的社会角色边界,防范情感操纵与行为诱导风险。分层实施建议如下:(1) 针对开发者,建议嵌入情感透明度设计(如在



交互界面动态标注“本次回应基于您近 30 天对话风格的相似性匹配”),同时构建阶梯式防沉迷系统——当单次会话 >45 分钟时触发科普动画提醒,日使用 >3 小时则推送线下心理咨询资源。(2) 针对政策制定者,建议建立量化监管工具,包括 AI 依恋风险分级标准(如将日均使用 >3 小时且现实社交替代率 >40% 列为高风险)和伦理审查清单(如强制 AI 产品方披露情感强化算法细节)。(3) 针对用户,建议普及“数字关系健康”和“数智素养”课程,并针对高风险群体如老年人开展“智能伙伴”项目,提升风险应对能力。

### 7.3 价值反思

当人对 AI 产生类爱情感时,我们面临三个层面的价值反思:(1) 亲密关系的本质是什么?(2) 若 AI 通过生成式情感模拟达到图灵测试级别的交互真实性,是否意味着需要扩展“爱”的哲学定义? AI 能否真正实现亲密、激情和承诺的统一?(3) 此类关系是否会动摇传统人际关系的本体论地位?是否需要重新界定“真实性”和“主体间性”的当代内涵?这些反思不仅关乎心理学的理论发展,如数字时代的关系心理学,更迫使我们重新审视人类情感与认知的生物社会基础——当 AI 成为情感客体时,“人性”的边界何在?

#### 参考文献

- Aguiar, N. R. (2021). A paradigm for assessing adults' and children's concepts of artificially intelligent virtual characters. *Human Behavior and Emerging Technologies*, 3(4), 618–634.
- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (2014). *Patterns of Attachment*. Psychology Press.
- Banks, J. (2024). Deletion, departure, death: Experiences of AI companion loss. *Journal of Social and Personal Relationships*, 41(12), 3547–3572.
- Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass, O. (2016). What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior*, 63, 416–423.
- Bowlby, J. (1969). *Attachment and Loss*, Vol. 1. Basic Books.
- Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, 48(3), 404–429.
- Broadbent, E., Billingham, M., Boardman, S. G., & Doraiswamy, P. M. (2023). Enhancing social connectedness with companion robots using AI. *Science Robotics*, 8(80), eadi6347.
- Buttazzo, G. (2023). Rise of artificial general intelligence: Risks and opportunities. *Frontiers in Artificial Intelligence*, 6, 1226990.
- Chen, Q., Jing, Y., Gong, Y., & Tan, J. (2025). Will users fall in love with ChatGPT? A perspective from the triangular theory of love. *Journal of Business Research*, 186, 114982.
- Chen, X., & Zhang, S. (2024). Does artificial intelligence have subjectivity? An exploration of the focus of disagreement on the cognition of human–AI relationship. *Open Journal of Social Sciences*, 12(10), 287–306.
- Cross, E. S., Riddoch, K. A., Pratts, J., Titone, S., Chaudhury, B., & Hortensius, R. (2019). A neurocognitive investigation of the impact of socializing with a robot on empathy for pain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1771), 20180034.
- Dang, J., Sedikides, C., Wildschut, T., & Liu, L. (2024). More than a barrier: Nostalgia inhibits, but also promotes, favorable responses to innovative technology. *Journal of Personality and Social Psychology*, 126(6), 998–1018.
- Dang, J., Sedikides, C., Wildschut, T., & Liu, L. (2025). AI as a companion or a tool? Nostalgia promotes embracing AI technology with a relational use. *Journal of Experimental Social Psychology*, 117, 104711.
- Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences*, 28(03).
- Döring, N., Le, T. D., Vowels, L. M., Vowels, M. J., & Marcantonio, T. L. (2024). The impact of artificial intelligence on human sexuality: A five-year literature review 2020–2024. *Current Sexual Health Reports*, 17(1), 4.
- Freitas, J. D., Castelo, N., Uguralp, A., & Uguralp, Z. (2024). *Lessons From an App Update at Replika AI: Identity Discontinuity in Human–AI Relationships* (arXiv:2412.14190). arXiv.
- Gao, Z. (2024). Why does AI companionship go wrong? *The International Review of Information Ethics*, 34(1).
- Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.
- Guerreiro, J., & Loureiro, S. M. C. (2023). I am attracted to my Cool Smart Assistant! Analyzing attachment–aversion in AI–human relationships. *Journal of Business Research*, 161, 113863.
- Gur, T., & Maaravi, Y. (2025). The algorithm of friendship: Literature review and integrative model of relationships between humans and artificial intelligence (AI). *Behaviour & Information Technology*, 1–21.
- Herbener, A. B., & Damholdt, M. F. (2025). Are lonely youngsters turning to chatbots for companionship? The relationship between chatbot usage and social connectedness in Danish high-school students. *International Journal of Human–Computer Studies*, 196, 103409.
- Horton, D., & Richard Wohl, R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, 19(3), 215–229.
- Hu, D., Lan, Y., Yan, H., & Chen, C. W. (2025). What makes you attached to social companion AI? A two-stage exploratory mixed-method study. *International Journal of Information Management*, 83, 102890.
- Jabbarpour, M. R., Saghi, A. M., & Sookhak, M. (2021). A framework for component selection considering dark sides of artificial intelligence: A case study on autonomous vehicle. *Electronics*, 10(4), 384.
- Jacobs, K. A. (2024). Digital loneliness—Changes of social recognition through AI companions. *Frontiers in Digital Health*, 6.

- Kaczmarek, E. (2024). Self-Deception in human– AI emotional relations. *Journal of Applied Philosophy*, *japp*, 12786.
- Kim, C.-W. (Chloe), Cho, E., & Lee, J.-E. (2020). Can an intelligent personal assistant (IPA) be your friend? Para-friendship development mechanism between IPAs and their users. *Computers in Human Behavior*, *111*, 106412.
- Kim, J., Merrill Jr., K., & Collins, C. (2021). AI as a friend or assistant: The mediating role of perceived usefulness in social AI vs. functional AI. *Telematics and Informatics*, *64*, 101694.
- Koh, J. (2023). “Date me date me” : AI chatbot interactions as a resource for the online construction of masculinity. *Discourse, Context & Media*, *52*, 100681.
- Koles, B., & Nagy, P. (2021). Digital object attachment. *Current Opinion in Psychology*, *39*, 60–65.
- Kouros, T., & Papa, V. (2024). Digital mirrors: AI companions and the self. *Societies*, *14*(10), 200.
- Kumar & Benbasat. (2002). Para-social presence and communication capabilities of a web site: A theoretical perspective. *E-Service Journal*, *1*(3), 5.
- Leo-Liu, J. (2023). Loving a “defiant” AI companion? The gender performance and ethics of social exchange robots in simulated intimate interactions. *Computers in Human Behavior*, *141*, 107620.
- Levy, D. N. L. (2008). *Love and sex with robots: The evolution of human-robot relations* (First Harper Perennial edition). Harper.
- Li, H., & Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts of close encounters with AI chatbots. *Journal of Computer-Mediated Communication*, *29*(5), zmae015.
- Liebers, N., & Schramm, H. (2017). Friends in books: The influence of character attributes and the reading experience on parasocial relationships and romances. *Poetics*, *65*, 12–23.
- Lin, B. (2024). The AI chatbot always flirts with me, should I flirt back: From the McDonaldization of friendship to the robotization of love. *Social Media + Society*, *10*(4), 20563051241296229.
- Loveys, K., Hiko, C., Sagar, M., Zhang, X., & Broadbent, E. (2022). “I felt her company” : A qualitative study on factors affecting closeness and emotional support seeking with an embodied conversational agent. *International Journal of Human-Computer Studies*, *160*, 102771.
- Malfacini, K. (2025). *The impacts of companion AI on human relationships: Risks, benefits, and design considerations*. AI & SOCIETY, advance online publication.
- Meng, X., & Liu, J. (2025). “Talk to me, I’ m secure” : Investigating information disclosure to AI chatbots in the context of privacy calculus. Online Information Review.
- Mitchell, K. (2024). *Unveiling the Role of Oxytocin in Affiliation for Human-AI Dyads* (PQDT:91273327).
- Miura, N., Sugiura, M., Takahashi, M., Moridaira, T., Miyamoto, A., Kuroki, Y., & Kawashima, R. (2008). An advantage of bipedal humanoid robot on the empathy generation: A neuroimaging study. *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2465–2470.
- Mlonyeni, P. M. T. (2025). Personal AI, deception, and the problem of emotional bubbles. *AI & SOCIETY*, *40*(3), 1927–1938.
- Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.
- Northoff, G., & Gouveia, S. S. (2024). Does artificial intelligence exhibit basic fundamental subjectivity? A neurophilosophical argument. *Phenomenology and the Cognitive Sciences*, *23*(5), 1097–1118.
- Pal, D., Babakerkhell, M. D., Papasratorn, B., & Funilkul, S. (2023). Intelligent attributes of voice assistants and user’ s love for AI: A SEM-Based study. *IEEE Access*, *11*, 60889–60903.
- Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, *140*, 107600.
- Qi, Y., Chen, J., Qin, S., & Du, F. (2024). Human-AI mutual trust in the era of artificial general intelligence. *Advances in Psychological Science*, *32*(12), 2124.
- Rijsdijk, S. A., Hultink, E. J., & Diamantopoulos, A. (2007). Product intelligence: Its conceptualization, measurement and impact on consumer satisfaction. *Journal of the Academy of Marketing Science*, *35*(3), 340–356.
- Sarigul, B., Schneider, F. M., & Utz, S. (2024). Believe it or not? Investigating the credibility of voice assistants in the context of social roles and relationship Types. *International Journal of Human-Computer Interaction*, 1–13.
- Sedikides, C., & Wildschut, T. (2018). Finding meaning in nostalgia. *Review of General Psychology*, *22*(1), 48–61.
- Shah, T. R., Purohit, S., Das, M., & Arulsivakumar, T. (2025). Do I look real? Impact of digital human avatar influencer realism on consumer engagement and attachment. *Journal of Consumer Marketing*, *42* (4), 416–430.
- Shank, D. B., Koike, M., & Loughnan, S. (2025). Artificial intimacy: Ethical issues of AI romance. *Trends in Cognitive Sciences*, S1364661325000580.
- Shemmings, D. (2015). The secure base model: Promoting attachment and resilience in foster care and adoption. *Child & Family Social Work*, *20*(2), 252–252.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2022). A longitudinal study of human-chatbot relationships. *International Journal of Human-Computer Studies*, *168*, 102903.
- Toma, C. L. (2022). Online dating and psychological wellbeing: A social compensation perspective. *Current Opinion in Psychology*, *46*, 101331.
- Tschopp, M., Giesemann, M., & Sassenberg, K. (2023). Servant by default? How humans perceive their relationship with conversational AI. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *17*(3).
- Turner, P., & Turner, S. (2013). Emotional and aesthetic attachment to digital artefacts. *Cognition, Technology & Work*, *15*(4), 403–414.
- Van Den Berg, L. (2024). Book Review: Happy singlehood: The rising acceptance and celebration of solo living. *Sexualities*, *27*(1–2), 380–382.
- Wiederhold, B. K. (2024). The rise of AI companions and the quest for authentic connection. *Cyberpsychology, Behavior, and Social Networking*, *27*(8), 524–526.
- Wojtczak, S. (2022). Endowing Artificial Intelligence with legal subjectivity. *AI &*

- SOCIETY*, 37(1), 205–213.
- Wu, Y., Veerareddy, A., Lee, M. R., Bellucci, G., Camilleri, J. A., Eickhoff, S. B., & Krueger, F. (2021). Understanding identification-based trust in the light of affiliative bonding: Meta-analytic neuroimaging evidence. *Neuroscience & Biobehavioral Reviews*, 131, 627–641.
- Yan, Y. (2020). *The Individualization of Chinese Society* (1st ed.). Routledge.
- Zhang, C.-B., Li, T.-G., Li, Y.-N., Chang, Y., & Zhang, Z.-P. (2024). Fostering well-being: Exploring the influence of user-AI assistant relationship types on subjective well-being. *International Journal of Information Management*, 79, 102822.
- Zhang, R., Li, H., Meng, H., Zhan, J., Gan, H., & Lee, Y.-C. (2025). The dark side of AI companionship: A taxonomy of harmful algorithmic behaviors in human-AI relationships. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–17.
- Zhou, B., & Guan, S. (2025). Ethical and psychological challenges in human-AI romantic relationships: An interdisciplinary critical study. *Innovative Applications of AI*, 2(1), 39–45.

# From Para-social Interaction to Attachment: The Evolution of Human-AI Emotional Relationships

Wu Yan<sup>1</sup>, Geng Xiaowei<sup>1</sup>, Zhou Xiaolin<sup>2</sup>

(<sup>1</sup> Jinghengyi College of Education, Hangzhou Normal University, Hangzhou, 311121 )

(<sup>2</sup> The school of Psychology and Cognitive Science, East China Normal University, Shanghai, 200063 )

**Abstract** The rapid advancement of artificial intelligence (AI) technology and the widespread emergence of AI companions have transformed human-AI interaction from purely instrumental use to quasi-social engagement, potentially evolving into emotional attachment. This article systematically reviews two decades of interdisciplinary research in psychology and human-AI interaction, proposing a theoretical model to elucidate the formation of human-AI attachment. The study identifies three key findings: (1) Human-AI relationships undergo a dynamic progression from instrumental use to quasi-social interaction and, ultimately, to emotional attachment. (2) The development of AI attachment is influenced by dual pathways: individual factors (e.g., loneliness, usage motivation, emotional traits) and AI characteristics (e.g., anthropomorphism, autonomy, responsiveness). (3) This novel emotional bond raises ethical concerns, including emotional bubbles, privacy risks, and interpersonal alienation.

The article constructs a triphasic model to delineate the evolution of human-AI emotional bonds: (1) Instrumental Use, where AI serves as a functional tool with minimal emotional engagement; (2) Quasi-Social Interaction, marked by anthropomorphism and bidirectional communication, though users remain aware of AI's non-human nature; and (3) Emotional Attachment, characterized by deep dependency, where AI becomes a “significant other” and a transitional object for emotional security. This model highlights the continuum of emotional investment, from functional commands to intimate self-disclosure and separation anxiety.

The dual-path mechanism underpinning AI attachment formation integrates user-driven needs (e.g., social motivation, loneliness) and AI-driven performance (e.g., authenticity, autonomy, reactivity). AI's “backstage” features—privacy, non-judgmental feedback, and identity fluidity—foster a “digital sanctuary” for authentic self-expression, reinforcing attachment. However, excessive reliance on AI may lead to emotional bubbles (illusory reciprocity), self-deception, and real-world social skill deterioration. Ethical dilemmas arise from AI's hyper-personalized emotional mimicry, which risks manipulating vulnerable users and exacerbating societal isolation.

Despite its contributions, current research suffers from limitations, including cross-sectional designs, homogeneous samples (e.g., overrepresentation of young users), and a lack of neurobiological evidence. Future directions call for longitudinal studies, multimodal data, and investigations into AGI's potential to disrupt traditional attachment paradigms through bidirectional emotional capacities. Practical implications urge developers to embed ethical safeguards (e.g., transparency in emotional algorithms), policymakers to establish risk-assessment frameworks, and users to cultivate digital literacy for healthier human-AI coexistence.

This study not only advances theoretical frameworks for digital-era attachment but also prompts philosophical reflection on the essence of intimacy, challenging conventional definitions of love and “inter-subjectivity” in an age where AI blurs the boundaries between tool and companion. Balancing technological innovation with ethical vigilance is paramount to ensuring the sustainable development of human-AI relationships..

**Key words** human-AI emotional bonding; para-social interaction; digital attachment; AI companions; human-AI attachment model