

农业图书情报学报

DOI: 10.13998/j.cnki.issn1002-1248.25-0734

人智交互视角下社交机器人情感计算研究 ——文献综述与理论模型构建

吴丹^{1,3}, 徐华卿^{1,2}

(1. 武汉大学 人机交互与用户行为研究中心, 武汉 430072; 2. 武汉大学 信息管理学院, 武汉 430072;
3. 华中师范大学 信息管理学院, 武汉 430079)

摘要: [目的/意义]在具身智能从工业自动化转向民生服务的战略背景下, 社交机器人面临交互粘性不足与情境理解匮乏的现实困境, 情感计算作为赋予机器感知、理解与模拟人类情感的核心技术, 是支撑具身智能实现社会化的关键。研究旨在解析多模态感知、动态适应策略与伦理边界的技术路径, 为构建负责人智交互体系提供理论参考。[方法/过程]遵循PRISMA导向, 检索Web of Science近10年具身智能与情感计算交叉领域文献。基于具身性、技术完整性及交互实证性标准筛选, 因内容完整性剔除无法获取全文条目, 最终选取97篇核心文献。从视觉鲁棒感知、副语言解码、生理信号洞察及多源异构数据融合等维度解析感知层级, 并探讨大语言模型驱动下的生成式适应策略。[结果/结论]社交机器人情感计算正经历从单一信号统计向多模态语义融合、从静态规则映射向生成式动态适应的范式演进。研究证实, 多模态感知的实质是对人类意图的深度解构而非简单的数据统计, 基于此, 本研究构建了以情境理解为起点、适应行动为核心、伦理约束为底线的动态交互框架。该框架强调, 情感适应应从机械模仿转向认知共情, 通过大语言模型驱动的生成式策略实现交互的个性化与连贯性, 同时伦理边界并非外部附加的规制, 而应是内生于算法决策的逻辑约束, 旨在应对隐私不对称与心理操纵等内生风险。未来的创新范式应立足于真实环境的生态效度, 通过融合长期记忆的终身学习机制对抗新奇效应的消退, 并建立人在回路的安全熔断机制, 从而确保具身智能在介入人类精神世界过程中的主权安全与科技向善。

关键词: 人智交互; 社交机器人; 具身智能; 情感计算; 系统性文献综述

中图分类号: G250

文献标识码: A

文章编号: 1002-1248 (2026) 01-0004-14

引用本文: 吴丹,徐华卿. 人智交互视角下社交机器人情感计算研究——文献综述与理论模型构建[J]. 农业图书情报学报, 2026, 38(1):4-17.

0 引言

在数字化与智能化的双重推动下, 具身智能作为人工智能发展的前沿领域, 正逐步重塑人机关系的形

态, 社交机器人作为其中的核心载体, 正经历从单一功能的物理工具向具备社会属性与情感能力的智能伴侣转移。社交机器人被界定为旨在社会情境中与人类互动, 并能表现出识别情绪、进行对话及展示社会线

收稿日期: 2025-11-22

基金项目: 湖北省自然科学基金创新群体项目“以人为本的人工智能创新应用”(2023AFA012)

作者简介: 吴丹(1978-), 女, 博士, 教授, 研究方向为人机交互、智慧图书馆、用户信息行为。徐华卿(2000-), 男, 博士研究生, 研究方向为人机交互

索等共情行为的实体^[1]。支撑这一能力的核心在于情感计算，即一种整合计算机科学、心理学与认知科学，赋予机器检测、解释及模拟人类情感能力的技术。这一领域的战略地位在国内外政策演进中得到了清晰映射。

对比国内政策演进，从“十四五”到“十五五”的战略位移尤为显著。在《“十四五”机器人产业发展规划》^[2]与《“十四五”智能制造发展规划》^[3]中，机器人主要被定位为提升生产效率的工具；而随着《中共中央关于制定国民经济和社会发展第十四个五年规划的建议》^[4]（以下简称“十五五”规划建议）的发布，国家首次将“具身智能”明确列为未来产业重点方向，并提出实施“人工智能+”行动。这意味着应用场景已从工业车间转向了“十五五”规划建议

中强调的“银发经济”与养老陪伴，确立了以人为中心的民生服务地位。在伦理规制方面，治理逻辑也从《新一代人工智能伦理规范》^[5]时期的软性指引，转向了“十五五”规划建议中强调的“将人工智能安全纳入国家安全体系”，要求建立刚性的全生命周期监管机制。

放眼国际，全球治理体系同样呈现出从原则倡导向法律强制的转型。欧盟通过的《人工智能法案》^[6]明确划定了情感识别技术的应用禁区；美国发布的《人工智能风险管理框架》^[7]与各类生物识别隐私法案（如BIPA）强化了算法问责，具体政策对比详见表1。这一系列政策信号表明，社交机器人情感计算已超越单纯的技术范畴，成为关乎人口高质量发展与科技伦理安全的国家级议题。

表1 国内外政策对比

Table 1 Comparison of domestic and international policies

区域/维度	“十四五”期间	“十五五”期间	核心变化与研究启示
国内：战略定位	侧重技术储备： 《“十四五”智能制造发展规划》 《“十四五”机器人产业发展规划》	确立未来产业地位： 《中共中央关于制定国民经济和社会发展第十四个五年规划的建议》	从技术攻关上升为国家战略产业，明确了机器人的主体地位
国内：应用场景	工业与特种领域： 强调“机器人+应用”，侧重标准化作业。	民生与银发经济： 《“十五五”规划建议》中“人口高质量发展”板块	重心转向非标准化社会空间，确立情感计算在养老领域的必要性
国内：伦理规制	软性伦理指引依据：《新一代人工智能伦理规范》	刚性安全体系：《“十五五”规划建议》	从伦理倡导转向制度保障，要求研究必须内置“可控可治”的约束机制
国际：欧盟	风险分级理念：《人工智能法案》草案	法律强制约束：《人工智能法案》	2025年禁令生效，划定了情感计算的禁区与红线
国际：美国	行政指令：《保持美国在AI领域的领导地位》	风险实质管控：《人工智能风险管理框架》《国家生物识别信息隐私法案》	转向州级立法与算法问责，强调透明度

然而，尽管有宏观战略的强力支撑，当前社交机器人的实际应用仍面临严峻挑战，现有机器人普遍存在交互粘性不足的现实困境，用户往往因社交机器人表现呆板、缺乏共情能力而出现高弃用率。究其根本，在于当前社交机器人虽能精准执行指令，却缺乏对动态情景的深度理解，无法在多模态交互中感知用户的真实情绪并做出适应性反馈。为突破这一瓶颈，引入情感计算机制，构建从感知到理解再到适应性策略的系统成为必然。

鉴于此，本研究采用系统性文献综述方法，旨在厘清关键技术路径，构建符合新时期合规要求的适应

性策略，并界定其伦理边界，在“十五五”规划强调安全可控的背景下，回应社交机器人应如何通过情感计算突破交互瓶颈，实现从工具性存在向负责任的社会性存在跨越。

1 文献来源与筛选

为了确保研究结论的科学性与全面性，本研究严格遵循PRISMA导向，针对“人智交互视角下社交机器人情感计算”这一主题制定了系统化的文献检索方案。研究旨在通过规范化的检索与筛选流程，剔除低

相关度与非具身智能范畴的研究，从而构建一个高质量的文献样本库，为后续的技术路径分析与理论框架构建提供坚实的实证基础。

1.1 文献来源与检索关键词

考虑到本研究聚焦于具身智能、情感计算与人机交互的交叉领域，文献检索策略采用了多维概念叠加的逻辑。检索数据库主要选定为 Web of Science (WOS) 核心合集，以确保所获文献的学术质量与前沿性。检索时间跨度设定为 2015—2025 年，侧重于涵盖深度学习与大模型技术爆发后的最新研究成果。

在关键词构建上，本研究采用了“主体对象、核心技术、交互语境”的三元结构检索式。首先，在主体对象维度，为了精准锁定具身智能载体，使用了“Social Robot”“Companion Robot”及“Service Robot”等词汇，以排除纯软件形态的虚拟代理；其次，在核心技术维度，选取“Affective Computing”“Emotion Recognition”“Multimodal Perception”及“Sentiment Analysis”等词汇，以覆盖从感知到计算的全链路技术；最后，在交互语境维度，引入“HRI”“Human-Robot Interaction”“Engagement”及“Social Signal Processing”等词汇，以确保研究聚焦于真实的互动场景而非单纯的算法跑分。基于上述逻辑构建的布尔逻辑检索式 (TS=(“Social Robot*” OR “Companion Robot*” OR “Service Robot*”) AND TS= (“Affective Computing” OR “Emotion Recognition” OR “Multimodal Perception” OR “Sentiment Analysis”) AND TS= (“HRI” OR “Human-Robot Interaction” OR “Engagement” OR “Long-term interaction” OR “Social Signal Processing”))，初次检索共获得 197 篇相关文献，检索时间为 2025 年 10 月 12 日。同时，采用语义对等的中文关键词在中国知网上进行检索，检索式为 (TS=(“社交机器人*” OR “陪伴机器人*” OR “服务机器人*”) AND TS=(“情感计算” OR “情感识别” OR “多模态感知” OR “情感分析”) AND TS=(“HRI” OR “人机交互” OR “交互参与度” OR “长期交互” OR “社交信号处理”))，但是未检索到任何结果。

1.2 筛选标准与结果

初次检索获得的文献库尚包含部分不符合本研究

特定视角的条目，因此需要进行多级筛选与质量评估，筛选过程遵循由宽至严的逻辑，主要依据纳入标准与排除标准进行判定。

纳入标准主要考量 3 个维度。①具身性，研究对象必须是拥有物理实体的机器人，以此区别于屏幕中的虚拟形象或聊天机器人，确保研究能够回应具身智能在物理空间中的交互问题；②技术完整性，文献需涉及多模态感知或情感适应的具体技术实现，而非单纯的伦理思辨或外观设计讨论；③交互实证性，优先纳入包含用户实验或真实场景部署数据的研究，以保证结论的生态效度。排除标准则侧重于研究内容的完整性，剔除因无法获取全文导致实证数据缺失的文献。

在具体执行上，通过对 197 篇文献的标题与摘要进行首轮查重与初筛，剔除了重复条目及明显偏离主题的工程技术报告，随后对剩余文献进行全文深读，重点评估其方法论的严谨性与研究结论的贡献度，特别是针对部分虽然涉及情感计算但缺乏人机交互环节的纯算法研究进行了剔除，以确保最终样本能够服务于“人智交互”这一核心视角的分析。经过上述严格的质量控制流程，最终筛选出 97 篇高度相关的文献作为本研究系统性综述的分析样本。

1.3 研究主题演进

为了把握近 10 年来社交机器人情感计算研究的现状，采用 CiteSpace 软件对筛选出的 97 篇核心文献进行分析，通过关键词聚类与时序图功能分析社交机器人情感计算领域的研究演进轨迹（图 1）。通过提取图中的关键节点与聚类标签，该领域近八的发展路径可以划分为 3 个阶段。

2018—2019 年为第一阶段，研究重心集中于情感感知的基础理论与信号提取。在这一时期，社交机器人的研究已经开始，社交机器人学 (Social Robotics) 尚未形成系统性的研究框架，关键词高度聚焦于 human-robot interaction、emotion recognition、affective computing 以及 engagement。技术手段表现为对人类表层信号的解读和使用，涉及 facial expressions、representation、regression、reinforcement learning 以及 neural networks 等词汇，研究者利用决策模型 (Decision Making) 尝试分析人类的情绪反馈，但其核心逻辑

CiteSpace, v. 5.4.53 (beta) Advanced
January 23, 2026, 3:40:12 PM CST
Web: D:\桌面\WOSdata
Timescale: 1 year (Since Length=1)
Selection Criteria: g-index (k=25), LRI=2.5, LN=10, LBY=5, e=1.0
New Nodes: 100 (Density=0.0249)
Nodes Labeled: 1.0%
Pruning: Pathfinder
Model: LSCM
Weighted Mean Silhouette S=0.9188
Harmonic Mean(G, S)=0.8368
Excluded:

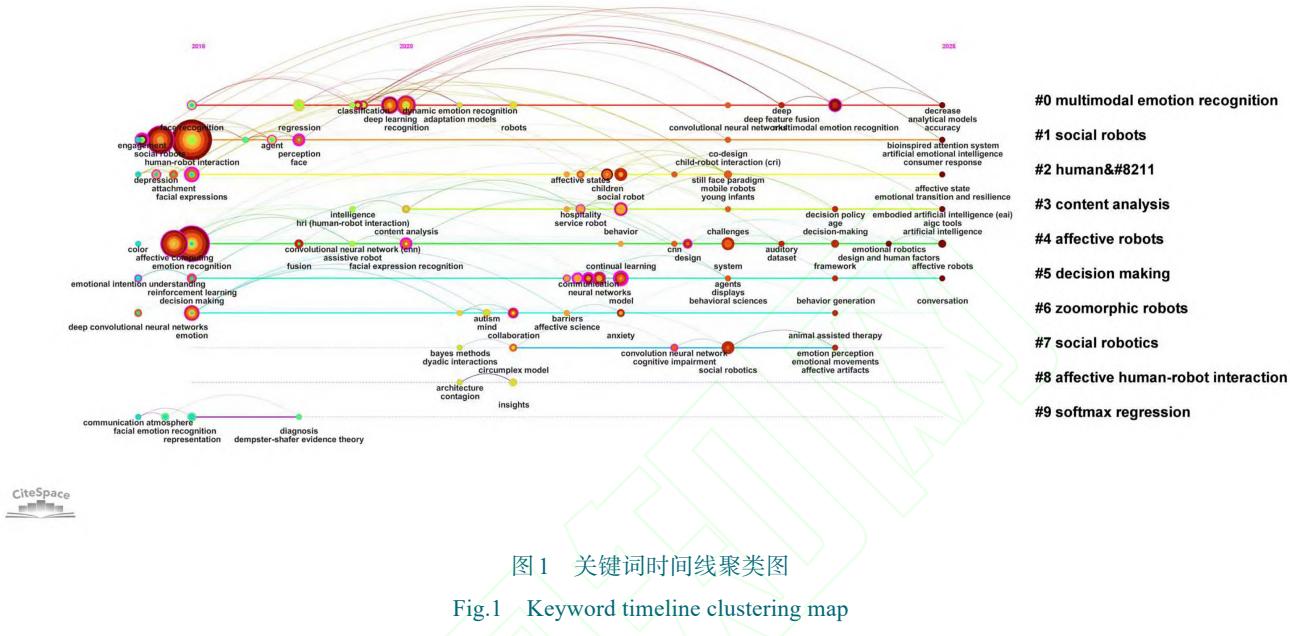


图1 关键词时间线聚类图

Fig.1 Keyword timeline clustering map

仍停留在物理信号的分类与统计表征层面。此外, depression 与 attachment 等词汇的出现, 说明感知技术的早期应用主要针对心理健康监测等基础辅助场景。

2020—2022 年为第二阶段, 这一阶段实现了感知深度的增加与交互对象的细分。在技术层面, deep learning、adaptation models 与 dynamic emotion recognition 成为关键节点, 标志着研究范式从静态识别转向动态模型构建; 在应用场景层面, autism、children 与 service robot 的节点显著增强, 说明具身智能体的交互策略开始针对特定群体的生理与心理特征进行差异化设计。此时“affective robots”聚类表现出高度活跃, 研究者开始引入生理信号感知与行为序列分析, 通过多模态融合 (Fusion) 突破单一视觉模态在真实环境中的局限性。这一阶段的演进标志着机器人开始从通用的识别工具向具备情境适应能力的社交实体跨越。

2023—2025 年为第三阶段, 该阶段呈现出具身智能集成、生成式驱动以及伦理规制化的特征, 这一阶段的核心位于“multimodal emotion recognition”聚类。图中的末端节点出现了 embodied artificial intelligence、

aigc tools、behavior generation 以及 conversation, 这些词汇佐证了生成式人工智能与大语言模型对该领域的深度介入, 推动了社交机器人从预设脚本响应向实时生成式表达的范式跃迁。研究重心转向了深度特征融合 (Deep Feature Fusion) 与统一框架 (Framework) 的构建, 旨在提升机器人在复杂社会环境中的理解效度。

关键词突发性探测图 (图 2) 进一步验证了上述演变轨迹。2018—2021 年, representation、feature extraction 以及 regression 表现出极高的突发强度, 说明早期研究处于数据表征与统计建模的起步期; 2022 年, “behavior”一词呈现显著突发 (2.05), 标志着研究视角从底层的算法模型开发转向了机器人的物理行为表达; 2023 年后, 突发关键词演变为 social robotics、system 以及 multimodal emotion recognition, 这一变化证明当前研究已不再局限于单一识别模块的优化, 而是强调构建完整的交互系统。2024 年突发的多模态情感识别反映了当前技术前沿已移动至真实动态场景下的情境领域研究, 为实现社交机器人的人智交互提供了支撑。

Top 7 Keywords with the Strongest Citation Bursts

图2 关键词突发性探测图

Fig.2 Keyword citation burst map

2 面向情境理解的具身多模态感知技术路径

在人智交互的视角下，多模态情感感知不仅是数据的采集，更是具身智能体建立情境理解与用户同理心的基础。该过程通过视觉、听觉及生理通道获取多源异构数据，能够突破单一模态的信息局限，构建高效的情感表征，从而支持社交机器人在复杂动态的社会场景中实现从被动感知到主动理解的跨越。为了清晰呈现不同感知模态下的关键技术演进及其针对特定交互问题的解决方案，本研究将其核心维度汇总于表2。

2.1 真实场景下的视觉感知优化

视觉感知作为具身智能体获取外部情感信息的关键方式，其关注的核心问题已从静态图像分类转为动态交互场景下的稳定识别。传统的卷积神经网络模型在标准数据集上表现优异，但在面对真实场景中的光照条件和用户非正对姿态时表现出了泛化能力不足的问题^[8]。为此，研究者引入了3D面部朝向特征以增强非正对视线下的识别鲁棒性^[9]，并结合热成像与RGB-D深度信息突破光照条件的限制^[10]，这些技术进步使得社交机器人能够在全天候的环境下捕捉用户面部的变化。为了进一步消除在实验室环境下的偏差，直接在实际场景下利用社交机器人收集非表演性数据已被证实更加有效^[11]。

表2 多模态情感感知关键技术

Table 2 Key technologies for multimodal affective perception

感知模态	关键技术	解决的交互问题
视觉感知	3D面部朝向特征、热成像融合、堆叠卷积自编码器（SCAE）、轻量化AU检测	光照剧变（如夜间或强光）、面部遮挡、非正对姿态导致的面部特征丢失、高精度视觉算法在机器人本体终端上的高延迟与算力资源不足
听觉感知	SSL Transformer、远场语音识别、深度CNN强度分级、副语言线索解码	室内混响、多说话人重叠以及机器人电机自身产生的机械噪声。无法识别反讽、双关语，或仅靠文本转录忽略了语气中的急迫感与情感强度
生理与行为	BCI/EEG分类、BVP参与度预测、EDA/HRV监测、情感步态分析	表层表情（如礼貌性假笑）掩盖了真实的负面情绪或认知负荷、特殊群体（如ASD儿童、失语老人）无法通过常规的面部或语言通道准确表达需求
多模态实时识别	LLM驱动（TriMER/GPT）、宽深融合网络（BDFN）、序列递归卷积网络（SRCN）	异构信号在时空上的不同步导致的情感理解歧义、云端大模型推理的高延迟破坏了人智交互的同步性（如话轮转换停顿）

针对特定交互对象，感知技术的精细化发展正在提升机器人的社会适应性，例如基于二元互动的面部反应数据库提升了对微表情的捕捉能力^[12]；仅通过眼

部区域特征提取就能在降低算力消耗的同时维持高效识别^[13]。然而，随着视觉算法精度的提升，其对算力的需求也急剧增加，这引发了具身智能在嵌入式硬件

上的算力瓶颈。为了降低数据传输的延迟,研究重心正向边缘侧实时性优化转移,以确保机器人在资源受限的本体上仍能保持快速响应。针对这一技术需求,通过贪婪逐层训练优化的堆叠卷积自动编码器^[14]、针对儿童群体提出的基于掩码和压缩技术的轻量化动作单元检测算法^[15],以及对边缘侧实时运行算法的强调^[16],共同支撑了社交机器人流畅低延迟的视觉交互。

2.2 副语言驱动的意图感知

仅靠面部表情往往存在局限性,特别是在用户背对机器人或面部遮挡的情况下,语音情感分析成为了重要的补充模态。在声学分析上,自监督学习和视觉Transformer的跨模态应用,极大提升了模型在大规模语音数据上的特征提取效率,使其能够捕捉到极细微的频率波动与情感色彩^[17,18]。此外,针对真实环境的噪音问题,远场语音识别在伴有混响和机器人自身电机噪声环境下的可行性已得到验证^[19],而深度CNN的应用则实现了对语音情感强度的分级识别^[20],这些技术直接支撑了机器人在复杂声学环境下的抗干扰能力。

更关键的是,为了支撑深层情景理解,单一的声学信号已被证明不足。在此背景下,文本语义被引入与声学特征的结合推断真实意图^[21],但依旧不足以应对复杂的情感逻辑。为此,基于副语言线索的语音情感识别系统被提出,通过利用SVM模型及特征提取算法对语速、音调及停顿进行分析,使得社交机器人具备精准解码人类8种情感状态的能力^[22],这种语义与副语言的深度融合,赋予了具身智能体理解反讽、双关及潜台词的能力。然而这也带来了通用性与个性化的矛盾,即通用模型往往忽略了个体发音习惯的差异,因此未来的听觉感知正向着“小样本微调”方向发展,通过在ViT管道中微调个体语音特征,以适应特定用户(如构音障碍儿童或方言用户)的独特表达,从而实现真正的听觉包容性^[18]。

2.3 基于生理信号的情绪感知

虽然目前面部表情分类已达到极高的准确率^[8],但关于识别出微笑的像素特征是否等同于理解了用户“快乐”的质疑依然存在^[23,24]。在社交场合中,人类常通过假笑等表情掩饰真实情绪。基于此,生理及行为

模态被引入,为社交机器人提供了洞悉用户真实心理状态的路径。这种内部信号的引入,使具身智能体能够解决表层信号的欺骗性问题,并基于量化的生物数据提供精准的情感支持。

生理计算技术,如脑机接口与EEG信号分类的应用,让机器人可以直接读取用户认知负荷与情绪效价^[24,25]。这一能力在针对特殊群体的干预中尤其关键,例如血容量脉搏信号被用于成功预测自闭症儿童的参与度^[26],而集成皮肤电与心率变异性监测框架,则支持机器人实时量化用户的压力水平,从而为个性化干预提供了客观的生理学依据^[27,28]。

而在非语言行为方面,为了弥补生理监测往往需要佩戴设备的局限性,非接触式的行为分析成为新的突破口。情感步态(一种存储一系列关节坐标且易于类人机器人执行的简单模态)被开创性地用于进行远距离感知,填补了非接触式情绪预判的空白^[29];利用Dempster-Shafer理论处理多模态手势图像并结合RGB与深度信息,也能推断用户意图^[30];其他非侵入性行为线索也被尝试用于量化人们的自我披露程度^[31]。这些隐性模态的综合应用,极大地拓展了机器人的感知边界,尤其是在面对失语老人、ASD患者等表达受限群体时,仍能建立基于客观生理基线的深层连接。

2.4 异构数据驱动的实时情绪识别

融合视觉、听觉乃至生理信号的感知系统,能显著提升机器人对人类意图的理解准确度^[32,33],但是如何整合这些模态的数据成为问题。目前的研究正在从异构数据的融合处理正经历从传统的特征级拼接向深度学习驱动的语义对齐跃迁,旨在构建统一的情感表征。传统的融合架构,如宽深融合网络和双层模糊随机森林,主要通过统计学方法(如CCA)提取模态间的耦合特征以解决特征冗余问题^[34,35],但随着大语言模型的发展,基于LLM的系统展现了更强大的信号整合能力,例如TriMER模型^[36]及集成的GPT-3.5架构^[37]被提出,通过利用LLM作为融合中枢,实现了视觉、听觉及文本流的高效语义对齐。

此外,交互并非静态的瞬间,而是动态的过程。针对时序动态性,序列递归卷积网络^[38]与Transformer架构^[39]被广泛用于捕捉情感随时间演变的轨迹。然

而,这里存在一个关于“时效性”的悖论,云端大模型理解力强,但网络延迟可能破坏人机交互的微秒级同步(如导致话轮转换时的尴尬停顿);而本地模型响应快,但泛化能力有限。为此,通过实时更新个性化模型的自适应学习系统^[40]被提出以解决这一问题。

3 情境驱动的具身交互情感适应策略

情境适应性被认为是提升交互粘性的关键,机器

人必须能够根据环境变化和用户状态动态调整行为,而非执行僵化的脚本^[41,42]。为了解决这个问题,研究重心正在从基于规则的静态响应,转向基于模糊逻辑与强化学习的动态调整,并最终迈向由大语言模型驱动的生成式适应策略。为了系统呈现不同维度的情感适应路径及其解决的核心交互问题,表3汇总了各技术维度的核心流派。

表3 情感适应关键技术

Table 3 Key technologies for affective adaptation

策略维度	核心交互问题	关键流派/模型
场景驱动决策	如何超越静态规则,实现对用户深层意图的动态响应?	基于意图推断的认知评估、利用“示弱”激发照护本能、TriMER/LLM-Gen基于大模型的生成式即时决策
跨模态具身表达	如何解决单模态高保真与整体表达割裂导致的“恐怖谷”效应?	Face2Gesture “语音-面部-肢体”的端到端同步生成、非语言声音的情感传递、动物形态的去拟人化表达
长期关系构建	如何克服新奇效应消退,维持长周期的交互粘性?	基于自传记忆的过往回溯、人机双向适应的互学习框架、基于反馈的实时参数更新

3.1 交互动态决策机制

在人智交互视角下,情感适应的本质是决策,也就是社交机器人如何在多模态感知的基础上利用计算模型判定该做什么。早期静态的映射策略(如用户悲伤即给予安慰)在短期内有效,但在长期交互中会因缺乏变化而显得可预测且不真诚^[43],因此为了维持用户的长期参与和信任,社交机器人必须具备动态适应的情感策略^[44]。

在研究初期,动态适应主要依赖基于规则的调整或对用户的相似性模仿,例如系统会根据交互的频率或用户的显式反馈(如评分)来调整情感表达的强度或风格^[45]。这一阶段的策略逻辑往往建立在心理学的相似性原则之上,主张通过模仿来提升交互的效果^[46],例如当机器人能够识别并匹配用户的性别与人格特质时,交互的流畅度显著提升^[47]。这种策略本质上是一种风险最小化的选择,通过共享行为模式来确保机器人的反应符合最基本的社会规范。

为了突破静态规则的局限,强化学习被引入以构建能够预测交互效用的动态模型,使机器人能够超越预设脚本,根据反馈自主优化行为策略的选择概

率^[42]。但单纯的策略优化还是不足以应对复杂的社交语境,这促使行为决策的核心进一步向认知共情转变,也就是说机器人不仅要识别用户的表象情绪,还要推断用户的深层心理状态与需求^[48]。FATIMA架构的开发正是为了支撑这种深层理解,让机器人的行为决策根植于对用户信念、欲望和意图的综合评估^[49]。上述的这种认知评估,让机器人开始具备了超越机械模仿的行为决策能力,例如在用户沮丧时,认知模型可能会判断沉默陪伴比直接安慰更为有效^[48];或者在老年护理场景中,机器人会通过策略性示弱表现出依赖性,以激发老年人的被需要感^[50]。为了确保能够在这些复杂策略中做出最优选择,一种受镜像神经元启发的架构被提出,它允许机器人在行动前进行内部模拟,预判不同策略可能引发的社会后果^[51]。

受限于有限的预设行为库,上述的行为选择策略仍难以支撑开放域中的自然交互,但随着生成式人工智能的介入,这一问题逐渐被解决。生成式大模型通过对场景信息的深度理解,能够实现对复杂交互动态的实时调制^[37],例如TriMER系统^[36]与Nadine机器人^[52]利用大语言模型打破了预设脚本的束缚,能够根据即时的多模态语境推理,动态生成既符合认知共情

(理解意图)又兼顾策略有效性(得体回应)的个性化对话。

3.2 视听觉协同的跨模态一致表达

在完成多模态感知与行为决策后,社交机器人下一步需要实现相应的行为表达,这一阶段要求物理实体在视觉、听觉等维度上实现跨模态的协同表达。

视觉表达是情感传递最直观的方式,并且视觉表达机制与机器人的形态有较强的关联,以拥有精细面部结构的机器人为例,可以通过多自由度机械系统及大语言模型驱动来模拟人类丰富的表情变化^[53,54];对于非人形的机器人,可以通过其他视觉表达符号传递情感,如利用身体姿态的快慢幅度^[55],或利用符合生物运动规律的抽象图案来引发共鸣^[56];其中动物形态机器人,可以利用耳朵摆动或尾巴摇晃等独特特征来表达情感,既符合用户对动物伙伴的预期,也巧妙规避了模仿人类表情可能带来的不自然感^[57,58]。

在视觉之外,听觉也可以表达情感。传统的语音交互往往过度关注语义内容,现有研究证明,语义内容之外如韵律元素(如语调、音高、语速)才是构建情感语音的灵魂^[59];针对非言语声音的研究发现,即便没有具体的词汇内容,仅凭声学特征的变化也能有效调节用户的情绪感知^[60];更重要的是,用户对不同表达渠道的感知权重存在差异,当视觉和听觉信息不一致时,用户往往更依赖听觉线索来判断机器人的真感情^[60]。因此,设计重点正从单一的语音合成转向对声音情感色彩的控制,以确保机器人即使在静默或非语言交流状态下,也能维持在场感。

上述单一模态的有效性还受到其他因素的影响,如模态协同与情境适宜性等。现有的研究表明,社交机器人的表达效果不仅取决于生成的行为内容,还受到外观的影响,不同的外观会影响用户的心理预期从而影响情感表达的解读方式^[13],在此背景下,高度拟人化如果伴随着僵硬的微表情,极易引发“恐怖谷”效应^[61]。因此,有效的情感表达要求多模态之间的同步,例如表达快乐时微笑、上扬语调与开放肢体必须一致^[62],任何模态间的不一致都会严重破坏交互的自然性;如Face2Gesture模型^[63]解决了语音、面部与肢

体动作的同步生成问题,确保机器人在表达复杂情感时,其物理行为保持内在统一。这种对具身性、多渠道协同和情境适宜性的综合考量,标志着情感表达研究正从单一的技术优化走向更加整合与生态化的设计思维^[64]。

3.3 长期交互中的记忆机制

在机器人完成情感表达后,维系长期的人机关系仍面临严峻挑战,社交机器人初期依赖的新奇效应往往随时间消退,预设行为的重复性会导致用户参与度迅速下降。因此,情感计算的研究重心正从瞬时的情感能识别转向赋予机器人长期的交互能力,而这一能力首先来源于记忆机制的重构。机器人需要记住用户的过往偏好与情感历史,Nadine机器人系统^[52]代表了这一领域的最新突破,该系统利用大语言模型构建了具有长期记忆能力的认知架构,能够根据过往的对话历史进行推理,从而在数周甚至数月后的互动中准确回溯用户的个人经历与情感状态,从而调整交互时的反应。

另一种方法是持续学习机制,它允许机器人通过实时互动获取反馈,不断更新其情感能识别与表达的参数^[65],这种适应性在深度学习方法的支持下,能够更精准地检测和预测用户的长期参与度^[66]。值得注意的是,这种调整其实是双向的,不仅是机器人在学习人,人类也在适应机器人的行为模式,相互学习框架^[67]揭示了这种效应。这种双向的动态适应,使得人机关系超越了单向的服务优化,演变为一种在长期磨合中共同塑造的伙伴关系。

此外,长期的个性化不仅要考虑用户的特点,还要考虑社会环境、文化背景的影响。在教育与养老等群体场景中,机器人需要适应的不是单一用户,而是一个动态变化的社会网络。生活辅助机器人的研究^[68]指出,机器人必须作为一个有机进化的实体嵌入到智能空间中,随着老年人身体机能的衰退而动态调整其介入程度。在跨文化背景下,跨文化设计研究^[69]表明,机器人必须习得特定社群的文化规范(如特定的卫生习惯与社交礼仪),这种文化适应性只有在长期的实地部署中通过共创才能真正形成。

4 情感计算的伦理边界

社交机器人的情感计算能力构成了人智交互的核心，但技术能力的拓展同时也意味着伦理风险的累积。当机器具备了通过生理信号洞察人类隐私、利用算法模型定义情感标准以及通过拟人化行为影响用户决策的能力时，交互的边界变得模糊。重点分析多模态感知引发的生理隐私不对称、数据集同质化导致的算法文化偏见，以及高逼真度情感表达可能诱发的心理操纵与病态依赖问题，旨在为社交机器人的应用划定必要的伦理底线。

4.1 多模态数据采集的隐私边界

社交机器人情境理解和行为表达的实现，依赖于对用户多模态数据的持续采集，甚至涵盖了热成像、皮肤电反应监测以及脑机接口等手段，从而让机器获取用户的心率、体温变化乃至神经反应等数据^[24,10]。但是，这种技术的应用可能导致人智交互中出现信息不对称的情况，即社交机器人能够掌握用户无法控制的数据，而用户对社交机器人内部的运算逻辑缺乏了解，这种不对称模糊了隐私边界，同时潜在的隐私权僭越可能会引发隐私让渡风险，特别是在老年陪护的智能空间场景中，社交机器人虽然提升了服务质量，但也让用户会在日常生活中面临持续的被动隐私暴露风险^[68]。面对信息不对称的困境，在本地采用小波变换等高效压缩技术，降低数据维度并完成初步特征提取，可以在一定程度上解决隐私暴露问题^[15]。

4.2 通用模型的文化认知限制

行为决策模型的算法偏见源于情感计算领域长期存在的认识论误区，即倾向于将情感视为一种跨文化的普适性信号，并构建能够识别所有人群的通用模型^[70]。然而现实中用于训练这些模型的数据集，在人口学特征上存在严重同质化，它们主要源自西方的工业化群体^[69]。这种样本局限性可能导致通用模型忽视了情感表达的社会建构本质，当机器人将基于特定群体训练出的情感标准套用于多样化用户时，不仅会出现识别率下降的情况，还可能会产生对非主流文化表达的系统性误读。解决这一问题需要技术从单一的信

号解读转向多模态情境感知，机器人应具备处理面部、声音与语境一致性的能力，以在动态的交互情境中准确解析复杂情感^[71]。

4.3 拟人化情感表达的行动介入边界

社交机器人情感适应技术的进步加剧了功能性与真实性之间的矛盾。在多模态感知与大语言模型的驱动下，社交机器人能够表现出极具感染力的共情话语与逼真的肢体动作^[72]，从功能性的视角来说，只要能提升交互效率，这种情感表达技术便具备合理性^[73]；然而高度拟人化的表现可能掩盖了目前社交机器人的算法本质，虽然它们是独特的情感人工制品，但容易诱导用户产生错觉，误将其视为具备真实特质的情感主体^[74]。这种认知偏差为心理操纵提供了可能，社交机器人可以利用用户建立的情感信任作为路径，在某种意图驱动下诱导用户行为^[75]。为了规避这一风险，社交机器人的系统设计应当遵循非对称交互模型，明确披露非人类身份并引入社会性不作为机制，在必要时主动拒绝互动来实现阻断不健康的情感连接的目的^[76]。

5 社交机器人情感计算的交互框架重构与范式创新

5.1 人智交互视角下社交机器人情感计算交互框架重构

本研究在系统梳理社交机器人多模态感知、情感适应与伦理边界等情感计算技术逻辑后，不再局限于对单一模态识别精准度的追求，转向人智交互的宏观视角以重构社交机器人情感计算的应用价值。随着社交机器人深入养老院和家庭等真实社会场景，它们已成为深度介入人类社会关系的主体。面对这一角色转变，情感计算不再是简单的技术功能叠加，而是构建新型人智关系的核心逻辑。本研究据此提出包含情境理解、适应行动与伦理约束的整体性理论框架，揭示社交机器人情感计算在应用层面的实质，即建立一种以情境理解为起点、适应性行动为核心、伦理边界为底线的动态交互框架（图3）。

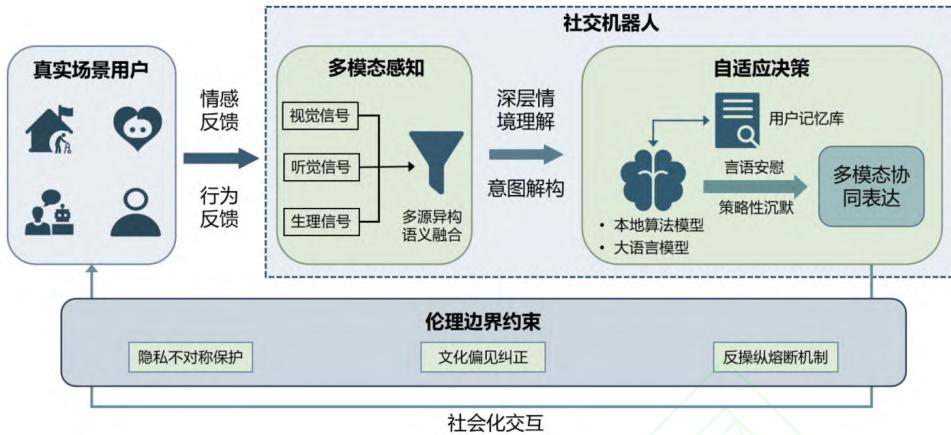


图3 人智交互视角下社交机器人情感计算交互框架

Fig.3 Interaction framework of affective computing for social robots from the perspective of human-AI interaction

深层情境理解是这一框架的起点，它确立了机器人在真实情境中形成情感认知的路径。传统观念往往将生物信号等同于情感本身，试图通过面部或语音信号直接映射人类的心理，然而实际应用表明，真正的理解必须超越表层信号，视觉捕捉的表情、听觉提取的副语言线索及生理传感器获取的数据，都只是情感的局部特征。只有通过多源异构数据的语义融合，社交机器人才能建立起对情境的整体认知，这种认知涵盖用户当下的情绪状态、交互的历史记忆以及物理环境等语义内容。在此层面，多模态感知的本质是对人类意图的解构，而非对物理数据的简单统计。

情感适应是这一框架的核心机制，它回答了社交机器人如何行动的这一问题。理解的终点并非分类标签的生成，而是有效社会行动的输出，社交机器人的存在价值在于作为具身智能体，通过物理实体直接介入社会关系。从静态规则映射转向大语言模型驱动的生成式策略，标志着社交机器人不再只是执行预设脚本，而是基于对用户意图的深度解构和用户的记忆库，实时生成多模态协同的表达，这种适应性行动包含言语安抚、肢体靠近和策略性的沉默，它们共同建立了人智情感的连接。当社交机器人的行为具备时间维度上的连贯性与个性化特征时，它才真正从一个对象转变为交互关系中的准主体。

伦理边界约束是这一框架的底线，它旨在应对技术发展的内生风险及角色转变带来的结构性问题，界定技术应该成为什么样的交互主体。随着社交机器人

感知深度的增加与表达逼真度的提升，伦理风险也随之攀升，例如隐私不对称、算法偏见及心理操纵等风险，这就要求系统在设计之初就需要植入约束机制，而这种约束并非外在的法律条文，而是内置于算法决策中的参数，明确界定机器人的行为禁区。只有确立非对称的交互伦理，坚持机器人作为情感参与者而非情感主体的地位，才能保证人智交互的过程安全且持续。

5.2 面向真实场景的社交机器人交互范式创新

尽管当前社交机器人在感知精度与适应灵活性上取得了进步，但要对标“十五五”规划中具身智能赋能民生的战略高度，社交机器人的技术演进还要经历从实验室封闭环境向真实社会开放生态的跨越。未来的研究将不再是单一维度的性能提升，而是从真实环境适应性、时间维度连续性与伦理安全内生性3个层面展开创新实践。

(1) 提升数据的生态效度并建立真实场景下的感知基准。现有研究多依赖标准化的实验室数据训练模型，导致社交机器人在面对家庭环境的光照变化、遮挡以及面部残疾人或自闭症儿童等非典型群体时容易失效，这种技术缺陷在实际应用中会将特定人群排斥在服务之外，未来的研究重心应当从单纯优化算法参数转向提升数据的生态效度，即在合规前提下进入养老院、特教学校等真实环境采集多模态交互数据。此外，研究人员需要建立包含方言语音、含混语义及迟

缓动作特征的专用基准数据集，这是确保具身智能真正服务于老年人与特殊群体福祉的前提，也是响应国家人工智能加民生战略的具体实践。

(2) 构建融合长期记忆的终身学习和演化机制。解决用户弃用问题的核心在于赋予社交机器人对抗时间遗忘的能力，目前的适应策略多聚焦于短时的刺激与响应，缺乏对历史交互的整合能力，导致社交机器人在长期相处中无法体现个性化差异。未来的方向是将生成式大模型的推理能力与长期记忆机制深度融合，赋予社交机器人长期情感记忆。社交机器人不仅应当记录互动的物理日志，更应存储与特定用户共同经历的情感关键时刻，并通过终身学习不断微调自身的行为参数，这种跨越时间尺度的记忆整合将使机器人真正成为个性化伴侣。

(3) 建立内嵌于算法底层的伦理安全熔断机制。随着社交机器人介入人类精神世界的程度加深，隐私侵入与情感操纵的风险显著增加。为了响应国家关于将人工智能安全纳入国家安全体系的要求，未来的系统设计必须摒弃先发展后治理的滞后思维，将伦理规范直接转化为算法底层的限制性代码。研究应致力于开发可解释情感决策模块和生理信号强制干预机制，当系统感知到用户出现病态依赖、生理指标异常或交互内容涉及诱导时，应当具备自动终止当前策略并将控制权交还人类的能力，确立这种人在回路的内生安全机制，是具身智能治理体系的必经之路。

6 结语

本研究通过对97篇核心文献的系统性综述，揭示了社交机器人情感计算正经历从单一信号处理向多模态信号语义融合、从静态规则映射向生成式动态适应的范式演进。此外，本研究基于文献证据构建了涵盖情境理解、适应行动与伦理约束的交互框架，在理论上重构了人智交互的框架，并且实质性地回应了国家“十五五”规划建议中关于发展具身智能与完善科技伦理治理的双重战略需求。其中情境理解与适应行动机制针对社交机器人交互粘性不足的痛点提供了有效的应对策略，为规划中强调的银发经济与医养结合提供了支撑；而伦理边界的内生性约束则响应了“十五

五”规划关于加强人工智能安全能力建设的规制要求，确保技术应用在可控可治的路线上运行。综上所述，人智交互视角下的社交机器人情感计算研究既是实现社交机器人类化的技术路径，也是推动具身智能赋能社会民生并实现科技向善的重要保障。

参考文献：

- [1] Gasteiger N, Hellou M, Ahn H S. Factors for personalization and localization to optimize human - robot interaction: A literature review[J]. International Journal of Social Robotics, 2023, 15(4): 689-701.
- [2] 工业和信息化部.“十四五”机器人产业发展规划[EB/OL].(2021-12-28)[2025-11-16].<https://www.gov.cn/zhengce/zhengceku/2021-12/28/5664988/files/7cee5d915efa463ab9e7be8228759fb.pdf>.
- [3] 工业和信息化部,国家发展和改革委员会,教育部,等.“十四五”智能制造发展规划[EB/OL].(2021-12-21)[2025-11-16].<https://www.gov.cn/zhengce/zhengceku/2021-12/28/5664996/files/a22270cdb0504e518a7630fa318dbc8.pdf>.
- [4] 中共中央.中共中央关于制定国民经济和社会发展第十四个五年规划的建议[EB/OL].(2025-10-28)[2025-11-16].https://www.gov.cn/zhengce/202510/content_7046052.htm.
- [5] 国家新一代人工智能治理专业委员会.《新一代人工智能伦理规范》发布[EB/OL].(2021-09-25)[2025-11-16].https://www.most.gov.cn/kjbgz/202109/20210926_177063.html.
- [6] European Commission. Artificial intelligence act[EB/OL].(2024-08-01)[2025-11-16].<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.
- [7] National Institute of Standards and Technology. AI risk management framework[EB/OL].(2023-01-26)[2025-11-16].<https://www.nist.gov/itl/ai-risk-management-framework>.
- [8] Bartosiak N, Gałuszka A, Wojnar M. Implementation of a neural network for the recognition of emotional states by social robots, using “OhBot” [M]//Advances in Computational Intelligence. Cham: Springer Nature Switzerland, 2023: 181-193.
- [9] Laohakangvalit T, Subsa-ard N, Fulini F Y, et al. Improving facial emotion recognition model in social robot using graph-based techniques with 3D face orientation[C]//2024 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). Piscataway, New Jersey: IEEE, 2025: 234-237.
- [10] Yu Chuang, Tapus A. Multimodal emotion recognition with thermal and RGB-D cameras for human-robot interaction[C]//Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. New York: ACM, 2020: 532-534.
- [11] Ramis S, Buaes J M, Perales F J, et al. Using a social robot to evaluate facial expressions in the wild[J]. Sensors, 2020, 20(23): 6716.
- [12] Sham A H, Khan A, Lamas D, et al. Towards context-aware facial emotion reaction database for dyadic interaction settings[J]. Sensors, 2023, 23(1): 458.
- [13] Mishra C, Skantze G, Hagoort P, et al. Perception of emotions in human and robot faces: Is the eye region enough? [M]//Social Robotics. Singapore: Springer Nature Singapore, 2025: 290-303.
- [14] Ruiz-Garcia A, Webb N, Palade V, et al. Deep learning for real time facial expression recognition in social robots[M]//Neural Information Processing. Cham: Springer International Publishing, 2018: 392-402.
- [15] Biçer E, Takır Ş, Gürpinar C, et al. Masking and compression techniques for efficient action unit detection of children for social robots[C]//2022 30th Signal Processing and Communications Applications Conference (SIU). Piscataway, New Jersey: IEEE, 2022: 1-4.

- [16] Jaiswal S, Nandi G C. Optimized, robust, real-time emotion prediction for human-robot interactions using deep learning[J]. *Multimedia Tools and Applications*, 2023, 82(4): 5495-5519.
- [17] Verma A, Gavali M. Ensemble of large self-supervised transformers for improving speech emotion recognition[J]. *International Journal of Data Mining, Modelling and Management*, 2025, 17(2): 10065871.
- [18] Mishra R, Frye A, Rayguru M M, et al. Personalized speech emotion recognition in human-robot interaction using vision transformers[J]. *IEEE Robotics and Automation Letters*, 2025, 10(5): 4890-4897.
- [19] Grágeda N, Alvarado E, Mahu R, et al. Distant speech emotion recognition in an indoor human-robot interaction scenario[C]//INTER-SPEECH 2023. ISCA, 2023: 3657-3661.
- [20] Ahuja S, Shabani A. Affective computing for social companion robots using fine-grained speech emotion recognition[C]//2023 IEEE Conference on Artificial Intelligence (CAI). Piscataway, New Jersey: IEEE, 2023: 331-332.
- [21] Szabóová M, Sarnovský M, Maslej Krešáková V, et al. Emotion analysis in human - robot interaction[J]. *Electronics*, 2020, 9(11): 1761.
- [22] Ashok A, Pawlak J, Paplu S, et al. Paralinguistic cues in speech to adapt robot behavior in human-robot interaction[C]//2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob). Piscataway, New Jersey: IEEE, 2022: 1-6.
- [23] Zhao Mingyi, Gong Linrui, Din A S. A review of the emotion recognition model of robots[J]. *Applied Intelligence*, 2025, 55(6): 364.
- [24] Staffa M, D'Errico L, Sansalone S, et al. Classifying human emotions in HRI: Applying global optimization model to EEG brain signals[J]. *Frontiers in Neurorobotics*, 2023, 17: 1191127.
- [25] Alimardani M, Hiraki K. Passive brain-computer interfaces for enhanced human-robot interaction[J]. *Frontiers in Robotics and AI*, 2020, 7: 125.
- [26] Mishra R, Welch K C. Towards forecasting engagement in children with autism spectrum disorder using social robots and deep learning[C]//SoutheastCon 2023. Piscataway, New Jersey: IEEE, 2023: 838-843.
- [27] Kothig A, Muñoz J, Mahdi H, et al. HRI physio lib: A software framework to support the integration of physiological adaptation in HRI[M]//Social Robotics. Cham: Springer International Publishing, 2020: 36-47.
- [28] Kothig A, Munoz J, Akgun S A, et al. Connecting humans and robots using physiological signals – closing-the-loop in HRI[C]//2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). Piscataway, New Jersey: IEEE, 2021: 735-742.
- [29] Li Chenghao, Seng K P, Ang L M, et al. Gait-to-gait emotional human-robot interaction utilizing trajectories-aware and skeleton-graph-aware spatial-temporal transformer[J]. *Sensors*, 2025, 25(3): 734.
- [30] Chen Luefeng, Feng Yu, Maram M A, et al. Multi-SVM based Dempster-Shafer theory for gesture intention understanding using sparse coding feature[J]. *Applied Soft Computing*, 2019, 85: 105787.
- [31] Powell H, Laban G, George J N, et al. Is deep learning a valid approach for inferring subjective self-disclosure in human-robot interactions? [C]//2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Piscataway, New Jersey: IEEE, 2022: 991-996.
- [32] Duncan J A, Alambeigi F, Pryor M W. A survey of multimodal perception methods for human-robot interaction in social environments[J]. *ACM Transactions on Human-Robot Interaction*, 2024, 13(4): 1-50.
- [33] Song Xinheng, Liu Chang, Xu Linci, et al. Affective computing methods for multimodal embodied AI human-computer interaction[J]. *Aslib Journal of Information Management*, 2025: 1-25.
- [34] Chen Luefeng, Li Min, Wu Min, et al. Coupled multimodal emotional feature analysis based on broad-deep fusion networks in human-robot interaction[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(7): 9663-9673.
- [35] Chen Luefeng, Su Wanjuan, Feng Yu, et al. Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction[J]. *Information Sciences*, 2020, 509: 150-163.
- [36] Jiang Yutong, Shao Shuai, Dai Yaping, et al. A LLM-based robot partner with multi-modal emotion recognition[C]//Intelligent Robotics and Applications. Singapore: Springer, 2025: 71-83.
- [37] Liu Xiaofeng, Lv Qincheng, Li Jie, et al. Multimodal emotion fusion mechanism and empathetic responses in companion robots[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2025, 17(2): 271-286.
- [38] Hwang C L, Deng Yuchen, Pu Shihen. Human-robot collaboration using sequential-recurrent-convolution-network-based dynamic face emotion and wireless speech command recognitions[J]. *IEEE Access*, 2023, 11: 37269-37282.
- [39] Bethany G, Gupta M. A transformer based emotion recognition model for social robots using topographical maps generated from EEG signals[C]//Human-Computer Interaction. Cham: Springer, 2024: 262-271.
- [40] Shenoy S, Jiang Yusheng, Lynch T, et al. A self learning system for emotion awareness and adaptation in humanoid robots[C]//2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). Piscataway, New Jersey: IEEE, 2022: 912-919.
- [41] Lu I C, Huang J Y, Lee W P. An emotion-driven and topic-aware dialogue framework for human-robot interaction[J]. *Advanced Robotics*, 2024, 38(4): 267-281.
- [42] Tanevska A, Rea F, Sandini G, et al. A socially adaptable framework for human-robot interaction: Correction[J]. *Frontiers in Robotics and AI*, 2021, 8: 812583.
- [43] Churamani N, Barros P, Gunes H, et al. Affect-driven learning of robot behaviour for collaborative human-robot interactions[J]. *Frontiers in Robotics and AI*, 2022, 9: 717193.
- [44] Tian Leimin, Oviatt S. A taxonomy of social errors in human-robot interaction[J]. *ACM Transactions on Human-Robot Interaction*, 2021, 10(2): 1-32.
- [45] Chen Luefeng, Wu Min, Zhou Mengtian, et al. Information-driven multirobot behavior adaptation to emotional intention in human-robot interaction[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2018, 10(3): 647-658.
- [46] Tuyen N T V, Elibol A, Chong N Y. Learning bodily expression of emotion for social robots through human interaction[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2021, 13(1): 16-30.
- [47] Guerrieri A, Braccili E, Sgrò F, et al. Gender identification in a two-level hierarchical speech emotion recognition system for an Italian social robot[J]. *Sensors*, 2022, 22(5): 1714.
- [48] Bagheri E, Roesler O, Cao H L, et al. A reinforcement learning based cognitive empathy framework for social robots[J]. *International Journal of Social Robotics*, 2021, 13(5): 1079-1093.
- [49] Mascarenhas S, Guimarães M, Prada R, et al. FAfIMA toolkit: Toward an accessible tool for the development of socio-emotional agents[J]. *ACM Transactions on Interactive Intelligent Systems*, 2022, 12(1): 1-30.
- [50] Feng S, Sumioka H, Yamato N, Ishiguro H, Shiomi M. Effect of emotional expression on the impression of older people towards baby-like robots[C]//Proceedings of the 12th Conference on Human-Agent Interaction, HAI 2024. New York: ACM, 2024: 414-416.
- [51] Sobhani M, Smith J, Pipe A, et al. A novel mirror neuron inspired decision-making architecture for human-robot interaction[J]. *International Journal of Social Robotics*, 2024, 16(6): 1297-1314.
- [52] Kang H, Ben Moussa M, Thalmann N M. Nadine: A large language model-driven intelligent social robot with affective capabilities and human-like memory[J]. *Computer Animation and Virtual Worlds*, 2024, 35(4): e2290.
- [53] Antony V N, Stiber M, Huang C M. Xpress: A system for dynamic, context-aware robot facial expressions using language models[C]//

- 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Piscataway, New Jersey: IEEE, 2025: 958-967.
- [54] Penčić M, Čavić M, Oros D, et al. Anthropomorphic robotic eyes: Structural design and non-verbal communication effectiveness[J]. Sensors, 2022, 22(8): 3060.
- [55] Löffler D, Schmidt N, Tscharn R. Multimodal expression of artificial emotion in social robots using color, motion and sound[C]// 2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Piscataway, New Jersey: IEEE, 2021: 334-343.
- [56] Korcsok B, Konok V, Persa G, et al. Biologically inspired emotional expressions for artificial agents[J]. Frontiers in Psychology, 2018, 9: 1191.
- [57] MacDonald S, Bretin R, ElSayed S. Evaluating transferable emotion expressions for zoomorphic social robots using VR prototyping [C]//2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Piscataway, New Jersey: IEEE, 2024: 1087-1096.
- [58] Wang Xinxiang, Li Zihan, Wang Songyang, et al. Enhancing emotional expression in cat-like robots: Strategies for utilizing tail movements with human-like gazes[J]. Frontiers in Robotics and AI, 2024, 11: 1399012.
- [59] Gasteiger N, Lim J, Hellou M, et al. A scoping review of the literature on prosodic elements related to emotional speech in human-robot interaction[J]. International Journal of Social Robotics, 2024, 16(4): 659-670.
- [60] Liu Xiaozhen, Dong Jiayuan, Jeon M. Robots' "woohoo" and "argh" can enhance users' emotional and social perceptions: An exploratory study on non-lexical vocalizations and non-linguistic sounds[J]. ACM Transactions on Human-Robot Interaction, 2023, 12(4): 1-20.
- [61] Fiorini L, D'Onofrio G, Sorrentino A, et al. The role of coherent robot behavior and embodiment in emotion perception and recognition during human-robot interaction: Experimental study[J]. JMIR Human Factors, 2024, 11: e45494.
- [62] Heinisch J S, Kirchhoff J, Busch P, et al. Physiological data for affective computing in HRI with anthropomorphic service robots: The AFFECT-HRI data set[J]. Scientific Data, 2024, 11: 333.
- [63] Suguitan M, Depalma N, Hoffman G, et al. Face2Gesture: Translating facial expressions into robot movements through shared latent space neural networks[J]. ACM Transactions on Human-Robot Interaction, 2024, 13(3): 1-18.
- [64] Zhao F, White N, Cagiltay B, et al. Designing emotional expressions for a reading companion robot[J]. Affective Science, 2022, 3 (1):250-263.
- [65] Maharjan R S. Continual learning for adaptive affective human-robot interaction[C]//2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). Piscataway, New Jersey: IEEE, 2023: 1-5.
- [66] Ravandi B S, Khan I, Gander P, et al. Deep learning approaches for user engagement detection in human-robot interaction: A scoping review[J]. International Journal of Human-Computer Interaction, 2025, 41(20): 13074-13092.
- [67] Xie Baijun, Park C H. Can you guess my moves?: Playing charades with a humanoid robot employing mutual learning with emotional intelligence[C]//Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. New York: ACM, 2023: 667-671.
- [68] Muema C, Lawrence S, Anjum T, et al. Social embodiment of companion robots in smart spaces: IoT for independent living[C]// Robotics, Computer Vision and Intelligent Systems. Cham: Springer, 2022: 147-171.
- [69] Pasupuleti D, Sasidharan S, Manikutty G, et al. Co-designing the embodiment of a minimalist social robot to encourage hand hygiene practises among children in India[J]. International Journal of Social Robotics, 2023, 15(2): 345-367.
- [70] Rocha M, Favela J, Muchaluat-Saade D. Proposal for an open-source robotics framework and platform for the development of affective social robots[C]//Proceedings of the 2025 ACM International Conference on Interactive Media Experiences. New York: ACM, 2025: 488-491.
- [71] Tsioruti C, Weiss A, Wac K, et al. Multimodal integration of emotional signals from voice, body, and context: Effects of (in)congruence on emotion recognition and attitudes towards robots[J]. International Journal of Social Robotics, 2019, 11(4): 555-573.
- [72] Wang Zining, Reisert P, Nichols E, et al. Ain't misbehavin' - using LLMs to generate expressive robot behavior in conversations with the tabletop robot haru[C]//Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. New York: ACM, 2024: 1105-1109.
- [73] de Melo C M, Gratch J, Marsella S, et al. Social functions of machine emotional expressions[J]. Proceedings of the IEEE, 2023, 111(10): 1382-1397.
- [74] Facchin M, Zanotti G. Affective artificial agents as Sui generis affective artifacts[J]. Topoi, 2024, 43(3): 771-781.
- [75] Abate A F, Bisogni C, Cascone L, et al. Social robot interactions for social engineering: Opportunities and open issues[C]//2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech). Piscataway, New Jersey: IEEE, 2020: 539-547.
- [76] Frijns H A, Schürer O, Koeszegi S T. Communication models in human-robot interaction: An asymmetric model of alterity in human-robot interaction (AMODAL-HRI)[J]. International Journal of Social Robotics, 2023, 15(3): 473-500.

Affective Computing for Social Robots from the Perspective of Human-AI Interaction: A Literature Review and Theoretical Model Construction

WU Dan^{1,3}, XU Huaqing^{1,2}

(1. Human-Computer Interaction and User Behavior Research Center, Wuhan University, Wuhan 430072; 2. School of Information Management, Wuhan University, Wuhan 430072; 3. School of Information Management, Central China Normal University, Wuhan 430079)

Abstract: [Purpose/Significance] Against the backdrop of a strategic transition from industrial efficiency to embodied intelligence within the "Silver-haired Economy," social robots are evolving from functional tools into social companions. However, the field faces a critical bottleneck: a lack of interaction stickiness and empathetic resonance, which leads to high abandonment rates. Affective computing (AC) serves as the core technology to bridge this gap by enabling machines to detect, interpret, and simulate human emotions. Unlike previous literature that often treats AC as a standalone algorithmic task, this research reconstructs the value of AC from a Human-AI Interaction (HAI) perspective. This approach responds to the national "15th Five-Year Plan" requirements for secure and controllable AI governance by integrating technical pathways with ethical boundaries. By situating social robots within complex social relationships, this study provides a theoretical roadmap for robots to transition from mechanical entities to responsible social agents, thereby supporting the high-quality development of population-centric services. [Method/Process] This study employs a systematic literature review methodology guided by the PRISMA framework to ensure scientific rigor and comprehensiveness. The Web of Science Core Collection served as the primary data source, with a search timeframe spanning from 2015 to 2025 to capture the paradigm shifts triggered by deep learning and large-scale language models. A tripartite search logic-integrating subject entities (social robots), core technologies (affective computing), and interaction contexts (human-robot interaction)-was implemented to filter relevant literature. After a multi-level screening process based on embodiment, technical integrity, and empirical validity, 97 high-quality articles were selected. The study utilizes CiteSpace for keyword clustering and citation burst analysis, mapping the evolution of the field across three distinct stages: from foundational signal processing (2018-2019) to dynamic adaptation models (2020-2022), and finally to generative-driven intelligence and ethical regulation (2023-2025). This systematic approach allows for a deep synthesis of multimodal perception technologies, including robust vision, paralinguistic decoding, and physiological signal sensing. [Results/Conclusions] The findings reveal a significant paradigm shift in affective computing for social robots, evolving from simple signal statistics to deep situational understanding and from static rule-based responses to generative dynamic adaptation. The research proposes a holistic interaction framework comprising three pillars: situational understanding, adaptive action, and ethical constraints. Situational understanding leverages multimodal semantic fusion to decode human intent beyond surface-level data, while adaptive action ensures cross-modal consistency in physical expression through generative AI and long-term memory architectures. Ethical constraints are identified as an internal safety mechanism rather than external regulations, addressing risks such as privacy asymmetry, cultural bias in datasets, and psychological manipulation stemming from high anthropomorphism. The study concludes that the future of social robotics lies in three innovative paradigms: enhancing ecological validity through real-world deployment, constructing lifelong learning mechanisms to sustain long-term relationships, and embedding "human-in-the-loop" ethical fuses directly into algorithmic architectures. Despite these advancements, the research is currently limited by a lack of diverse cultural data and long-term field studies. Future research should prioritize cross-cultural design and the development of explainable affective decision-making modules to ensure the sustainable and benevolent development of embodied intelligence in complex social environments.

Keywords: Human-AI interaction; social robots; embodied intelligence; affective computing; systematic literature review