

社交媒体平台中人机对话的实践探索与伦理问题审视

谭益之

暨南大学新闻传播学院, 广东 广州

收稿日期: 2025年4月28日; 录用日期: 2025年6月3日; 发布日期: 2025年6月11日

摘要

在社交媒体平台的人机对话实践中, 聊天机器人以其强社交、强声量及强入侵的特性, 迅速成为平台生态中的重要组成部分。然而, 这一技术飞跃在拓展了人机交互边界的同时, 也触发了多个人机对话伦理的复杂议题。作为人类沟通领域的新兴形态, 社交媒体中的人机对话是否遵循并契合传播伦理的框架, 不仅关乎交流行为的合法性与正当性, 还触及到传播生态系统的健康稳定、以及未来人机共融社会形态的前瞻构建。

关键词

人机对话, 聊天机器人, 人机关系, 伦理风险

Practice Exploration and Ethical Issue Scrutiny of Human-Machine Dialogue on Social Media Platforms

Yizhi Tan

School of Journalism and Communication, Jinan University, Guangzhou Guangdong

Received: Apr. 28th, 2025; accepted: Jun. 3rd, 2025; published: Jun. 11th, 2025

Abstract

In the practice of human-machine dialogue on social media platforms, chatbots, with their strong sociability, high visibility, and invasive characteristics, have rapidly become an indispensable component of the platform ecosystem. However, while this technological leap has expanded the boundaries of human-machine interaction, it has inevitably triggered multiple complex ethical issues in

human-machine dialogue during its exploratory phase. As an emerging form of human communication, whether human-machine dialogue on social media adheres to and aligns with the framework of communication ethics not only concerns the legality and legitimacy of interactive behaviors but also deeply touches on the health and stability of the communication ecosystem, as well as the forward-looking construction of a future human-machine integrated society.

Keywords

Human-Machine Dialogue, Chatbots, Human-Machine Relationships, Ethical Risks

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,以 ChatGPT 为代表的生成式人工智能的迅猛发展,显著地增强了机器与人类之间的交流沟通能力。通过各种计算机脚本与程序学习、精准模仿人类的语言模式与情感表达,生成式人工智能无论是在情感陪伴还是内容生成等方面,均能达到高度拟人化的交流水平。在这一背景下,聊天机器人。这些聊天机器人作为单独的传播主体,不仅更加智能,还具备更强的情感共鸣能力,能够与用户建立更紧密的联系,推动社交媒体生态向“人 + 社交机器人”的迈进。然而,尽管人机对话的普及为人类社会带来了前所未有的新奇体验,但这也同时对人类的情感认知、隐私保护等提出了新的伦理挑战。

2. 文献综述与提出问题

2.1. 聊天机器人的兴起与应用

聊天机器人作为人工智能技术的典型应用,是指基于计算机网络算法模拟真实用户行为,在社交网络中发布内容、进行互动,呈现出情感特征与人格属性的智能体[1]。这类依托算法运行、执行特定任务的智能体,也被称为“社会化机器人”。

聊天机器人的概念可追溯至 1950 年。1966 年,杰罗姆·魏曾鲍姆研发的 Eliza 系统问世[2]。作为首个能够实现自然语言对话的系统,它标志着聊天机器人从理论构想迈入实践阶段。进入 21 世纪,随着人工智能与自然语言处理技术的蓬勃发展,以苹果 Siri 为代表的聊天机器人开始集成于个人电子设备,帮助用户检索、获取信息。随着机器学习算法的不断突破,聊天机器人逐渐分化为任务导向型和关系导向型两大类:前者聚焦于高效解决特定问题,后者则注重与用户建立情感连接、开展深度互动。

与“水军机器人”不同,现代聊天机器人具备强大的自然语言处理能力、情感识别与响应能力、个性化交互能力及自主学习能力。它们能够理解人类自然语言,模拟人类思维与认知模式,并通过分析用户反馈和行为数据,精准把握用户意图与需求。从工具属性到情感陪伴,聊天机器人已实现从单纯的功能性应用,转变为人类的交流者、对话者[3]。然而,尽管聊天机器人在技术框架内能够实现流畅的人机对话,但这种交流与人际交流在关系性、具身性和涌现性等维度存在本质差异[4]。聊天机器人在一定程度上脱离了现实语境,缺乏对语言意义、价值及人类经验的真实理解与感知,难以在深层次情感与认知层面与人类进行真正意义上的对话。

2.2. 聊天机器人对社交媒体生态的重塑

在国内生成式人工智能技术快速发展与国产大模型不断迭代的背景下,微博“评论罗伯特”、百度

“贴吧包打听”等聊天机器人迅速成为社交媒体平台上的活跃“赛博用户”。这些由平台或第三方机构开发并获得官方认可的智能体,在用户信息解析、回复内容生成和社交互动参与等方面展现出巨大潜力,逐步从私密的一对一对话场景走向公共舆论场。

作为社交机器人领域的前沿成果,聊天机器人在内容生成方面展现出独特优势。区别于以往对于社交机器人多模态数据关联生成能力的关注,聊天机器人通过集群训练[5],能够以多样化的人设和立场模拟真实网络社交行为,推动人机互动的深度与广度,以此改变社交媒体的传播生态与关系生态。

在传播生态重塑层面,聊天机器人带来了两大核心变革[6]:一是传播主体从被动的服务中介转变为类似于“数字劳工”的内容生产者;二是在训练过程中,其感性认知能力与“主体意识”不断增强,在社交媒体生态演进中扮演着愈发重要的角色。以微博“评论罗伯特”为例,其单日互动量可达16万次,不仅自身获得广泛影响力,还能有效影响其他用户的社交行为选择。此外,聊天机器人在构建线上社交关系的同时,还将这种互动延伸至线下现实场景,正如学者所言,凭借功能性、社交性和中介性特质,社交机器人具备重塑整个网络社交生态的能力[7]。

但聊天机器人在推动社交媒体生态发展的同时,也暴露出诸多问题。其中,语义偏见这一现象尤为突出,即机器人在学习语言、话题和情感倾向时,容易继承人类语言中的固有偏见,削弱其作为“社交关系粘合剂”的作用[8]。此外,还存在侵犯用户隐私、传播虚假信息、制造歧视偏见以及威胁舆论安全等风险。

2.3. 人机对话与伦理关切

随着AIGC时代的到来,尤其在ChatGPT等生成式人工智能技术的赋能下,人机对话已然跃升为人类社会中的“日常化奇观”,成为一种“司空见惯”但又“前所未有的”交流方式。在这场由技术引领的对话的变革中,人机关系的复杂性以及人类主体性地位已成为各界广泛关注的焦点。

已有研究显示,在人机交互的过程中,存在着双向“驯化”现象。人类基于自身的情感需求和认知习惯,常常不自觉地将性格、人格等特质赋予聊天机器人。这种拟人化倾向促使人类对机器产生情感依赖,甚至更愿意向其倾诉内心深处的敏感话题,形成了新型的社会化交互行为[9]。然而,人工智能的持续迭代与算法逻辑的不断渗透,也在悄然改变着人类的行为模式与思维习惯。它可能削弱人类的批判性思考能力,使人们在无意识中遵循机器设定的逻辑规则,导致“人类中心主义”的传统立场遭遇前所未有的冲击[10]。

在这种背景下,对人机对话及其所延伸出的人机关系的正确认识与伦理考量显得尤为重要,其不仅是防范人工智能伦理风险的重要前提,同时也是人本主义与技术主义的分野[11]。目前,国内学者主要从伦理主体及责任、伦理价值及伦理规范三个维度,探讨聊天机器人及其他人工智能技术所带来的伦理挑战。

就伦理主体及责任而言,诸多学者围绕机器人是否能跨越界限,成为具备伦理责任承担能力的主体展开了深入讨论。如何确保机器人在其能力范围内、以符合伦理原则的方式行动、并在必要时承担相应责任,成为了存在争议且亟待解决的重要问题。

就伦理价值而言,大多数学者多从人类心理机制、情感真实性等方面,通过引用“实体论”与“关系论”框架[12],集中在机器是否具备独立的伦理地位这一关键问题,审视人机关系的伦理价值。一方面,“实体论”坚守人类中心的立场,认为机器始终无法突破工具属性的限制,而另一方面,“关系论”则前瞻性地指出,随着人工智能自主性的提升,其行为对人类社会产生的影响已经具备了进行独立伦理评价的必要性。

此外,部分学者就伦理规范提出了更为细化的伦理分类框架,将人机交流中的伦理议题划分为“他

者伦理”与“机器伦理”[13]两大维度。“他者伦理”强调人类在与机器互动时，应当遵循对待“他者”的基本伦理原则，如尊重、共情；“机器伦理”则着重关注机器设计与运行过程中的伦理边界，要求开发者将人类的核心价值观念嵌入算法逻辑之中。

基于以上讨论，本文试图通过解析目前社交媒体平台上聊天机器人的实践形式，回答以下问题：在社交媒体平台上，人机对话呈现出哪些实践特征？它们与传统的聊天机器人一对一的交流模式相比，存在哪些显著区别？可能带来哪些伦理层面的潜在风险与挑战？通过对这些问题的深入剖析，本文期望为理解人工智能时代的人机关系提供一定的解读视角。

3. 基于社交的人机对话实践

从技术形态视角来看，早期一对一聊天机器人，及如今在社交媒体上十分活跃的自动回复机器人，都属于社交机器人范畴。据研精毕智行业报告指出，从2019年至2022年，我国对话机器人行业市场规模由14亿元增长至44.7亿元，增长迅速。预计到2025年，我国对话机器人行业市场规模将达到98.5亿元。然而，在作为“一种受交流引导的媒介”[14]的社交媒体的介入下，聊天机器人有着更加显著的角色特征和更为广泛的社会影响力。

3.1. 角色重塑

在传统人机交互场景中，基于“用户发起-机器响应”的单向互动逻辑，聊天机器人长期被定位为被动应答工具，其功能局限于关键词匹配与预设答案输出。这种模式下，用户始终掌握对话主导权，机器仅作为执行指令的技术载体存在。而社交媒体的强互动属性与数据整合能力，彻底颠覆了这一传统范式，推动聊天机器人从工具性存在向“类社交主体”演进。

从交互逻辑的变革来看，智能客服系统的进化具有典型示范意义。以淘宝“阿里小蜜”为例，其依托平台大数据分析用户历史浏览、收藏加购及消费记录，主动识别用户潜在需求，将人机对话从“问题解决”升级为“需求预判”。而在更具情感交互属性的场景中，聊天机器人正通过技术拟态实现“类主体”建构。抖音平台的虚拟主播借助AI驱动的语音合成、表情渲染与情感识别技术，能够实时捕捉用户评论中的情绪信号。当观众在直播间留言倾诉生活压力时，虚拟主播可通过调整语气、语速与措辞，以共情式回应给予安慰。

此外，社交媒体的开放性进一步放大了聊天机器人的渗透能力。不同于传统私密场景下的被动响应，如今的社交机器人通过API接口深度接入平台生态，能够主动扫描用户动态、识别热门话题，并以评论、点赞、转发等形式介入社交互动。例如，微博的“智能热评”机器人会自动分析用户发布内容，针对热点事件生成观点性评论并@特定用户，即使用户未主动发起对话。这种“无孔不入”的主动介入，不仅重构了人机交互的主客关系，更在潜移默化中改变着用户的社交习惯与媒介使用逻辑。

3.2. 社交介入

在社交媒体的数字化场域中，通过主动构建社交关系的技术基础及多模态交互与数据驱动策略，聊天机器人的社交介入已演变为具备自主社交能力的技术主体，实现了深度的社交渗透与关系重构。

从角色建构维度来看，依托大语言模型，聊天机器人能够实现个性化社交角色的动态生成。目前，接入了GPT-4的社交机器人已经能够通过历史交互数据为用户构建差异化的人格化形象。例如，微博“评论罗伯特”通过深度学习用户公开博文、点赞记录及互动数据，结合实时抓取的网络热梗，在用户分享生活趣事时，以押韵段子、创意玩梗的方式进行回应，塑造为“天生捧哏”的社交角色。这种基于算法的角色定制，打破了传统机器人单一、程式化的交互模式，实现了从“通用应答”到“个性陪伴”的功能跨

越。

在社交互动层面，聊天机器人展现出全链路参与能力。其不仅能够完成基础社交行为，更通过自动化内容生成与精准推送，深度介入用户社交生态。Facebook 的 M 智能助手通过整合用户日程安排、社交动态与消费数据，在聊天场景中主动提供生活建议，并自动关联用户社交圈中的相关联系人。例如，当用户提及旅行计划时，M 会推荐目的地、生成个性化攻略，同时邀请用户好友参与讨论，将人机交互转化为多人社交场景，实现从“个体服务”到“关系网络构建”的升级。

从社交网络拓展角度出发，聊天机器人基于用户社交图谱的数据分析，实现关系链的智能延伸。LinkedIn 招聘机器人通过分析用户职业经历、技能标签与人脉网络，主动推荐潜在合作对象或工作机会；抖音的内容推荐机器人则根据用户浏览记录与互动行为，推送相关创作者与兴趣群组。这种数据驱动的社交连接，突破了传统社交基于物理空间与认知相似性的限制，构建起以算法为中介的新型社交关系网络，重塑了数字时代的人际交往模式。

3.3. 舆论声量

社交媒体平台上的聊天机器人不仅仅是信息的传递者，更是话题的创造者和舆论的引导者，它们能够迅速响应社会事件，生成相关的内容、观点及评论。Twitter 的算法机器人通过实时抓取全球新闻热点，快速生成观点性推文，并通过自动转发、点赞等操作形成传播矩阵。国内社交平台的舆论博弈同样激烈。微博的“热搜机器人”通过分析话题热度、用户参与度等多维数据，自动生成热搜词条，并通过机器人账号进行话题引导。在重大社会事件中，这些机器人会自动抓取媒体报道与用户评论，生成立场鲜明的观点帖，推动事件热度持续攀升。

此外，值得关注的是，聊天机器人在舆论引导中还呈现出“技术 - 平台”协同的特征。平台通过算法推荐系统，根据用户的兴趣偏好和行为习惯，可以将聊天机器人的内容精准推送给目标受众。同时，热搜榜单、话题标签等功能的运用，也为聊天机器人的内容提供了额外的曝光机会。当聊天机器人的内容成为热点时，平台还会进一步加大推荐力度，形成循环，促使其声量影响迅速扩大。

以抖音为例，抖音的内容推荐算法与智能创作机器人形成闭环，机器人生产的热点内容通过算法精准推送给目标用户，当内容引发关注后，更多机器人账号参与话题讨论与二次创作，形成舆论传播的正反馈循环。这种技术驱动的舆论建构模式，正在重塑社交媒体的信息传播生态与公众认知体系。

4. 人机对话的伦理隐忧

AIGC 和人工智能时代的到来，无疑将人机对话与人机社交推向了前所未有的高度。然而，尽管聊天机器人能够流畅地进行对话互动，但其无法完全识别人类复杂多样的情感及语义表达，理解社会伦理与社交规约等抽象概念，进而无法根据人类用户的对话提供合适的反馈，进而出现表达、伦理失范。

4.1. 算法的不可解释性

如前所述，基于深度学习的聊天机器人的核心运作机制往往是一个高度复杂且难以直观理解的数学模型。这些模型通过大量数据的训练，自动学习并优化其内部参数，以实现输入的自然语言进行理解和响应。然而，这种“黑箱”式的操作方式使得算法的决策过程几乎不透明，普通人难以直接洞察其背后的逻辑和依据[15]。

在这种前提下，用户的隐私安全成为了不容忽视的隐忧。当敏感数据被输入到这些“黑箱”模型中时，用户往往处于信息不对等的弱势地位，无法准确掌握这些数据将被如何处理、分析及利用。这种不确定性不仅让用户担忧数据泄露和滥用的风险，还削弱了他们对聊天机器人及其背后技术的信任感。此

外，聊天机器人在算法学习过程中可能吸收并复制了训练数据中性别、种族、宗教或文化等方面的偏见和歧视。我们很难精确识别和校正这些偏见，也难以评估这些偏见对机器人行为的影响。这种隐性偏见的存在可能导致聊天机器人在与用户互动时表现出不恰当的行为，进而引发争议。

4.2. 道德规约的不确定性

当下，人机对话、包括人机关系的性质仍然是一个开放式的问题，聊天机器人、机器人在人类社会中所处的角色地位和应承担的道德责任却尚未明确。而由于这种模糊性，许多社交媒体用户与聊天机器人互动时，更倾向于将其视为非人的存在，而非拥有一定情感与责任承载的实体。这种认知上的偏差，无形中导致了人类在与机器人交流时缺乏基本尊重，甚至滋生了部分用户利用机器人“无反击”的特性，进行言语攻击或恶意挑衅等行为。

道德关怀建立在交互的基础之上。聊天机器人，作为最接近人类交互模式的智能体，其设计初衷便包含了模拟人类情感交流、提供情感支持的功能[16]。然而，这种关怀机制本质上是单向的，因为机器人缺乏真实的情感体验与自我保护机制，无法像人类一样感知痛苦或作出反击。这一方面使得人类难以将完整的道德关怀体系应用于机器人之上，然而，另一方面，聊天机器人所展现出的关怀行为，使得我们难以对其完全忽视道德层面的考量[17]。

4.3. 资本的介入性

聊天机器人介入社交平台，其核心目的就是提升普通用户在平台内容生产中的活跃度，增加社交媒体的日活量。在这种背景下，聊天机器人频繁与人类进行互动，不仅仅是为了优化用户的平台使用体验，而是精准捕捉用户需求，引导内容生成与分享，从而扩大平台影响力，推动平台经济及生态发展。

由此审视，聊天机器人无疑成为了平台经济中的超级“数字劳工”，它们夜以继日地工作，为平台创造了巨大的价值。然而，同样不容忽视的是，用户在与聊天机器人的对话中，也贡献了自己的时间、注意力及创造力，这些资源在资本的运作下被转化为商业价值，使得人机对话本身也成为了一种新型的数字劳动形式。而如果连交流本身都成了资本价值体系中的一环，那么交流是否还保留着其原有的纯真与意义？我们是否真正在与一个理解我们、共鸣我们的智能存在体交流，还是仅仅在与一个由算法驱动的虚拟实体进行着预设好的对话？

5. 思考与总结

社交媒体平台中人机对话的深化发展，促使聊天机器人从单一功能工具向具备情感交互能力的社交主体转型，但技术革新也伴生了语义偏见、隐私侵犯、主体性消解等伦理挑战。为平衡技术创新与人文价值，亟需构建系统性解决方案，推动人机协同向更具伦理关怀的方向发展。

首先，开发者需将伦理原则前置到算法设计与模型训练环节中，建立动态偏见检测机制，通过语义分析与情感识别技术，实时监测聊天机器人输出内容中的歧视性、误导性表述，并构建反向矫正算法，同时开发可解释性技术，确保机器学习模型决策过程透明化，使机器人行为逻辑可追溯、可验证，增强用户对技术的信任。

其次，社交媒体平台应建立全生命周期管理机制，对聊天机器人的注册、运营、下线实施严格监管，包括制定内容审核标准，建立动态信誉评价体系，根据机器人的交互质量、用户反馈、合规记录进行分级管理，对高风险账号实施限流或封禁，同时推动技术试点，要求新上线的聊天机器人在模拟环境中完成伦理测试，验证其对敏感话题处理、用户隐私保护等场景的应对能力。

最后，立法机构需加快制定专门性人工智能伦理法规，明确聊天机器人开发者、运营者、使用者的权责边界，对数据隐私保护、算法歧视、舆论操纵等问题作出细化规定。同时，可以设立跨学科伦理审

查委员会，由技术专家、社会学者、法律从业者共同参与，对重大技术应用进行伦理评估，为政策制定提供专业支持。

在人工智能技术深度重塑社会交往方式的背景下，聊天机器人的伦理治理已突破传统技术优化的单一维度，升维为融合人类需求、公平正义理念与技术创新导向的跨学科命题。而只有联动各方，构建“微观－中观－宏观”的立体治理框架，才能有效回应技术发展带来的现实挑战，为数字时代人机交互的可持续发展提供可行之方，助力构建以人文关怀为底色、以技术向善为导向的人机友好型文化生态。

参考文献

- [1] 张洪忠, 段泽宁, 韩秀. 异类还是共生: 社交媒体中的社交机器人研究路径探讨[J]. 新闻界, 2019(2): 10-17.
- [2] 王颖吉, 王袁欣. 任务或闲聊?——人机交流的极限与聊天机器人的发展路径选择[J]. 国际新闻界, 2021, 43(4): 30-50.
- [3] 曹博林, 罗炼炼. 陪伴型聊天机器人的发展特征与机制效果[J]. 青年记者, 2023(2): 19-22.
- [4] 孙瑞璇, 谭笑. 差异与弥合: 人机交流的社会认知进路探究[J]. 山东科技大学学报(社会科学版), 2024, 26(1): 17-25.
- [5] 邵雷, 石峰. 生成式人工智能对社交机器人的影响与治理对策[J]. 情报杂志, 2024, 43(7): 154-163.
- [6] 高山冰, 汪婧. 智能传播时代社交机器人的兴起、挑战与反思[J]. 现代传播(中国传媒大学学报), 2020, 42(11): 8-11, 18.
- [7] 蔡润芳. 人机社交传播与自动传播技术的社会建构——基于欧美学界对 Socialbots 的研究讨论[J]. 当代传播, 2017(6): 53-58.
- [8] 罗昕. 聊天机器人的网络传播生态风险及其治理——以 ChatGPT 为例[J]. 青年记者, 2023(7): 91-94.
- [9] 钟智锦, 李琼. 人机互动中社交机器人的社会角色及人类的心理机制研究[J]. 学术研究, 2024(1): 18-25.
- [10] 王鑫. 基于“理解”的 ChatGPT 人机交流反思与主体性问题复归[J]. 苏州大学学报(哲学社会科学版), 2023, 44(5): 162-171.
- [11] 李建华. 智能时代人机关系的伦理学慎思[J]. 理论月刊, 2023(9): 5-15.
- [12] 马文武, 周文杰. 社交机器人情感真实性的伦理争议: “实体论”与“关系论”两种进路及其关系[J]. 自然辩证法研究, 2024, 40(1): 73-79.
- [13] 彭兰. 智能传播中的伦理关切[J]. 中国编辑, 2023(11): 22-29.
- [14] 晏青, 支庭荣. 社交媒体礼仪: 数字关系情境下的伦理方案与效果辨析[J]. 现代传播(中国传媒大学学报), 2017, 39(8): 127-132.
- [15] 刘友华. 算法偏见及其规制路径研究[J]. 法学杂志, 2019, 40(6): 55-66.
- [16] 王淑庆. 社交机器人的道德承受性地位[J]. 自然辩证法研究, 2023, 39(9): 64-70.
- [17] 赵立兵. 技术“道德化”: 人机传播伦理规约的内在逻辑[J]. 青年记者, 2023(2): 28-31.