

社交机器人引发的真实性问题及其伦理探寻

赵天一¹, 于雪¹, 刘博涵²

(1.大连理工大学人文学院, 辽宁大连 116024; 2.北京大学医学人文学院, 北京 100191)

摘要: 社交机器人的快速发展在给人类生活带来便利的同时, 也诱发了人类在情感、认知和道德等方面的真实性问题, 使人类产生错误的“共情”或“移情”, 造成盲从行为及引发违心的道德决策等现象。这些问题的出现, 一方面是由于社交机器人在设计上过度追求拟人化, 另一方面是由于用户在使用过程中过度依赖与信任社交机器人。对此, 在社交机器人的设计和使用环节, 要避免因过度拟人化所带来的“欺骗性”问题, 保持合理的人机距离, 正视社交机器人对人类真实性的积极作用, 在合理范围内善用社交机器人, 助力科技与社会的良性发展。

关键词: 社交机器人; 真实性; 虚假性; 非真实状态

[中图分类号]B82 [文献标识码]A [文章编号]1672-934X(2025)02-0041-06

DOI:10.16573/j.cnki.1672-934x.2025.02.006

The Issue of Authenticity Caused by Social Robots and Its Ethics

Zhao Tianyi¹, Yu Xue¹, Liu Bohan²

(1.School of Humanities, Dalian University of Technology, Dalian, Liaoning 116024, China; 2.School of Health

Humanities, Peking University, Beijing 100191, China.)

Abstract: The rapid development of social robots, while bringing convenience to human life, has also triggered authenticity issue in human emotions, cognition, and morality, which can lead to misguided "empathy" or "emotional transference", resulting in such issues as blind obedience and morally questionable decisions. These issues arise, on the one hand, from excessive anthropomorphism in the design of social robots, and on the other hand, from users' over-reliance and trust in these robots during interaction. Accordingly, the "deceptive" issues caused by them should be avoided. By maintaining a reasonable human-robot distance, acknowledging the positive role of social robots in enhancing human authenticity, and utilizing social robots within appropriate boundaries, the healthy development of technology and society can be enhanced.

Keywords: social robots; authenticity; falsity; inauthenticity state

社交机器人是指能够和人类进行交流与互动的智能机器人^[1], 具有情感性、倾向性、目的性与学习性等特点^[2]。随着社会生产力的发

展, 社交机器人作为机器人领域的新兴业态, 逐渐进入人类社会, 常被应用于教育、医疗、新闻传播等多个领域。社交机器人在为人类社会带

收稿日期: 2024-07-28

基金项目: 国家社会科学基金一般项目(22BZX025)

作者简介: 赵天一(1999—), 男, 博士研究生, 研究方向为技术哲学、技术伦理;

于雪(1989—), 女, 副教授, 博士生导师, 主要从事技术哲学、技术伦理研究;

刘博涵(2000—), 女, 硕士研究生, 研究方向为技术哲学、技术伦理。

来诸多利好的同时,也诱发了人类在情感、认知和道德等方面的真实性问题。在传统意义上,真实性问题主要涉及人类增强技术。但在传统人际交往受到人机交互冲击的背景下,真实性问题也出现在机器人技术中,这一现象引起了人们的广泛关注。其中,社交机器人引发的真实性问题尤为突出。什么是社交机器人引发的真实性问题?这些真实性问题有何表现?为何会产生这些问题?人类应当如何对待社交机器人在真实性问题上带来的冲击?上述问题亟待从哲学层面进行探讨和澄清,从而助力社交机器人在人类社会中的良性发展。

一、社交机器人引发的真实性问题及其表现

(一)真实性问题的源起:技术的中介调节作用

“真实性”(Authenticity)起初意味着对真实自我的肯定与追求^[3],即要求人的情感、认知与道德决策不受外部因素的影响。与真实性相对的是“非真实性”,即人们在受到外界的影响时,难以在情感、认知和道德等方面呈现真实的自己^[4]。随着新兴技术的发展,真实性概念表现出层次性特征,不仅涵盖多种技术类型,而且出现了不同的理解方式与研究重点。例如,在人类增强技术涉及真实性的讨论中,人们会更加关注技术对人内在的独特本质与自主选择的影响^[5]。在机器人技术领域,社交机器人的真实性与虚假性相对应,具体指人们在使用社交机器人时,受其影响而产生的虚假的情感、意图和决策^[6],上述现象也可以被视为社交机器人所引发的“非真实性”状态。实际上,这种“非真实性”状态的出现与技术的“中介调节”(Mediation)作用相关^[7]。在中介调节作用日渐凸显的情况下,人的真实性便有可能在技术使用的过程中被“遮蔽”,产生并非出于本我的情感、意图和道德决策。

作为一种技术中介,社交机器人在人与世界的关系中起到了至关重要的调节作用。然

而,不当的设计与使用可能使社交机器人对人的真实性产生负面影响,使人陷入“非真实性”状态,由此引发人的真实性问题。

(二)社交机器人引发的真实性问题的具体表现

社交机器人的研发初衷在于服务人类社会,但在实际应用场景中,社交机器人可能在情感调节、认知引导与道德规范等方面存在局限性,导致虚假性的产生。首先,为了使人能够在交互中获得情感上的满足,越来越多的社交机器人具备了人格化的情感功能及拟人化的外观设计。这些设计虽然有助于人机之间情感关系的建立,但也有可能使人类产生错误的“移情”与“共情”,导致“非真实”情感的产生。一般情况下,“移情”与“共情”产生在人与人之间。但是,情感功能的嵌入及拟人化的外观设计可能使人类将情感转移或类比到与社交机器人的交互之中,从而对社交机器人产生“移情”或“共情”。例如,《纽约时报》(*The New York Times*)曾报道,日本部分老年人热衷于与AIBO社交机器人共同生活,并将其作为自己的“孩子”或“家人”,甚至希望死后能与AIBO社交机器人合葬^[8]。但实际上,社交机器人所引发的“共情”是一种“伪共情”(Pseudo-Empathy)。社交机器人虽然表面上具有同理心,但无法真正与人类建立联系或共享经历^[9]。而且,当“移情”或“共情”超出合理范围之时,人类便可能产生脱离实际的、“非真实”的情感。此外,社交机器人还能通过模拟逝者帮助人类调节悲伤情绪,使人们在心理上对其产生强烈的依赖与信任^[10]。虽然这样可以帮助人们缓解不良情绪,但也可能导致人们对社交机器人产生“非真实”的情感寄托,进而忽视对人际情感关系的维护^[11],不再从现实世界的人际交往中获得精神上的慰藉与依靠,从而难以实现对不良情绪的真正消解。

其次,社交机器人能够促进信息传播、丰富用户体验,这一特征使其在社交媒体平台上得

到了广泛应用。但同时,社交机器人也可能借此特征刻意引导舆论方向,从而在认知层面误导公众。有研究发现,在互联网用户群体中,当社交机器人的比率达到10%时,便可以引发舆论风暴,并诱使公众信服社交机器人所倡导的观点^[12]。这表明,出于从众心理以及对社交机器人的信任与依赖,人们可能被迫接受外部的主流意见。而这些被接受的意见往往并非个人的真实意图,而是人们在社交机器人的作用下误将外部意见视为自己真实意见所产生的虚假结果,这种现象会使人们认知行为的个体性逐渐消弭,进而在社会行动中陷入盲从的境地。此外,为实现用户信息的精准推送,社交机器人能够依托以ChatGPT为代表的预训练基础模型,实现更高效、更准确的信息投送^[13],由此在用户周围营造出“信息茧房”(Information Cocoon),进而导致“回音室效应”(Echo Chambers),使用户误将自己所处某圈层的观点视为全社会的观点,并出现刻板印象^[14],从而造成其思维方式的固化与思想观点的趋同。这种对“真相思维”的忽视将导致人们以错误的思维方式认识事物。需要注意的是,这里的“真相思维”指的是一种透过现象看本质的思维方式,如批判性思维等,即要求人们在确定相信什么或者做什么时进行合理且成熟的思考^[15]。

最后,社交机器人在满足人类的情感需求与信息供给时,难免对人类的道德判断与决策造成影响。有研究显示,在使用大数据算法进行犯罪风险评估时,黑人群体被标记为高犯罪风险的概率是白人群体的两倍,这暴露出典型的种族歧视问题^[16]。这表明,社交机器人受内部算法影响,呈现具有偏向性、歧视性的信息,进而引导人们做出错误的道德判断。同时,社交机器人的交互类型影响人类的道德行为。例如,在涉及捐赠问题时,若社交机器人能够在人机交互过程中引入更多的社会线索,人们的捐赠行为便会因此受到影响:当社交机器人提供

的信息越完整,人们的捐赠数额会越高;反之亦然^[17]。这意味着,社交机器人可以通过改变交互方式影响人类的道德行为,可能导致人类在道德行动上出现“违心”之举。

二、社交机器人引发真实性问题的成因分析

人类对社交机器人的不合理设计和使用是真实性问题产生的主要原因。具体来说,主要包括两个方面:一方面,社交机器人在设计上过度拟人化导致“真假难辨”;另一方面,人类在使用社交机器人过程中对其的过度信任和依赖造成了“盲从”与“违心”。

(一)过度拟人化导致“真假难辨”

“恐惑谷”理论表明:在一定范围内,人类对机器人的亲和感与其拟人化程度成正比^[18]。因此,为促进人机之间建立情感联系和丰富使用者的交互体验,社交机器人在外观与功能上往往被采取拟人化设计,即将人类特征赋予非人类实体的倾向^[19]。适度的拟人化设计有助于人机交互的顺利展开,但过度的拟人化容易使人难辨信息的真假,导致人类自身的真实性被遮蔽。

一方面,社交机器人过度拟人化的外观在无形中提高了人类的辨识成本,进而引发本质与现象之间的矛盾。在人机交互中,如果人们仅将社交机器人视为无生命、无情感的工具,就难以对其“情感”与“行为”发展产生兴趣^[20],但若人类将社交机器人视为有生命、有情感的存在,则又容易偏离客观现实。因此,在人机交互中,人类需要花费更多的精力以保持对社交机器人的正确认识,否则人类可能对其产生不恰当的认知。

另一方面,社交机器人过度拟人化的功能容易使人高估其真实能力,在生活中盲目信任或依赖社交机器人。例如,过度拟人化的情感功能使人们忽视了社交机器人情感能力的有限性。虽然社交机器人能够通过利用人的同理心(Empathy)来实现人机之间的情感交流,但其自身却无法及时给予人类充分的情感反馈^[21],

反而可能使人对它的情感能力产生误解,导致错误的“移情”和“共情”。此外,社交机器人在功能上的过度拟人化,使其“他者性”特征愈发明显,甚至表现出一定的道德能力。这有可能导致人们在使用中忽视了社交机器人在道德地位与道德权利上的限度,进而对其进行不合理的道德评判和道德辩护。

(二)过度信任和依赖造成“盲从”与“违心”

除了设计上的过度拟人化之外,为了能够吸引用户并促进人机交互,越来越多的社交机器人在使用中表现出对用户的讨好倾向。这既有助于社交机器人取得人类信任,又加深了人类对社交机器人的依赖^[21]。相较于人际交往中的不确定性,人们可能更愿意与社交机器人进行交互。人们可能过度依赖与社交机器人进行情感交流,而忽视从人际情感关系中寻求支持和安慰。这不仅无益于真正提升人们的幸福感,还容易使人在无形中产生虚假的情感投射。同时,社交机器人还能够为人们提供专业的建议与服务,辅助人们做出决策。例如,部分护理机器人可以24小时不间断地对老人进行实时监护,协助老人完成如起床、传递物品、设定提醒服务等辅助活动^[22]。然而,这类工作内容也容易导致人类对社交机器人产生过度的信任与依赖。这种信任与依赖可能导致人类在与社交机器人进行交互时丧失理智,从而让渡自身部分的感知与决策权力^[23]。

此外,社交机器人在信息传播层面的巨大优势使其成为人们获取外界信息的便利渠道。但在某些情况下,社交机器人也可能成为传播虚假信息、引导社会舆论和造成极端化情绪的推手。例如,在2017年的加泰罗尼亚公投期间,社交机器人生成并传播了大量针对独立主义者的暴力内容,这些内容加剧了在线社区公众意见的极化问题^[24]。因此,非理性“信任”社交机器人可能导致人们在社会行为上产生盲从,或是做出违心的道德判断与决策。

三、社交机器人真实性问题的纾解之道

(一)适度拟人化以防“欺骗性”问题滋生

面对社交机器人引发的真实性风险,人类应当平衡拟人化设计与现实情景的需求,以避免过度拟人化设计所带来的“欺骗性”问题。这里的“欺骗性”是指社交机器人的外观或行为可能导致人类对其功能的高估,从而产生关于社交机器人功能和角色的错误认知^[25]。

第一,设计师要合理把握社交机器人的“拟人化”设计限度,在保持其形象友好的同时增加辨识特征。例如,老年护理机器人PARO采用宠物海豹作为外形,有效降低了使用者对其功能的过高期待^[26]。第二,设计师要基于情境需求为社交机器人设置合理的功能实现方式。设计师可以通过设置用户自主调节机器人能力表现程度的功能,使社交机器人的模仿能力可控,进而促进人机之间的个性化良性互动,构建和谐的人机关系。同时,设计师还可以为社交机器人设置使用时间与使用次数可调节功能。具体来说,当社交机器人实际运行时间超出设定的时间范围之后,社交机器人将进入待机或休眠状态,防止用户过度依赖。第三,设计师要坚持价值敏感性设计的理念,并将其贯穿到设计与研发的全过程。在设计初始阶段,设计师可以结合技术可用性预设人的真实性如何在使用中得以呈现。设计师可以将符合实现人类真实性的道德标准程序化为道德代码,嵌入社交机器人的人工智能系统。由于人的价值观念与伦理取向处于动态的变化过程之中,设计师可以通过定期更新社交机器人内部的道德代码,以应对由价值观念变化而引发社交机器人影响人类真实性的新情况。

(二)保持合理的人机距离谨防人际关系疏远

应对真实性问题,需要人类在使用过程中对社交机器人拥有正确认知,并与其保持合理距离,谨防因过度沉迷而出现人际关系疏远问题。

一方面,人们可以通过自律与他律相结合的方式应对上述问题带来的不良影响。从自律的角度来看,个人可以主动提高自己的思辨意识与反思能力,在人机交互中正确认识社交机器人的局限性,自觉把控人机关系的走向。此外,个人还应积极在人际交往中寻求精神上的支持与依靠。从他律的角度来看,人们可以按照具体的应用场景将社交机器人划分至不同领域,并要求该领域的工作人员依照相应的伦理原则对其进行管控,以确保用户与社交机器人保持合理距离。另一方面,人们在人机交互中要做到理性认识与感性认识相统一,强化自身的反思能力。这要求人们既要在使用中理性认识社交机器人的局限性,反思自己的真实情感、认知与道德是否存在被遮蔽的风险,又要进一步丰富释放自身情感的渠道,使自身能够在人际交往中寻求真正的精神关怀。此外,在使用过程中,人们应该明确社交机器人的辅助身份,使自己成为掌握情感生活、认知活动与道德判断的关键部分。具体来看,人们要主动走出依赖社交机器人的“舒适圈”,谨慎听取社交机器人的意见与建议,保持独立思考的能力,改变以往被动接受社交机器人意见与建议的认知习惯,克服认知过程中的惰性心理,学会从多种平台与方式收集信息,以丰富自己的信息渠道^[27]。

(三)发挥社交机器人促进人类真实性的积极作用

社交机器人作为一种技术中介,如果人们能够合理应用并合理限制其对人类的调节尺度,那么社交机器人也可以对人的真实性的实现产生积极作用。有研究显示,社交机器人能够通过陪伴与互动满足阿尔茨海默病患者的情感需求,从而缓解患者的躁动情绪与心理压力^[28]。这表明,合理使用社交机器人有助于改善人的情绪,引导人类合理释放自身的情感,从而减少过激行为的出现,避免因极端化情绪而造成认知与道德上的“违心”。此外,在医疗护理

领域,社交机器人可以为患者提供个性化服务。社交机器人可以针对患者独特的生活经历进行精准编程,以此刺激患者的认知活动,促进患者认知功能的恢复与提升^[29]。这有助于削弱或消除疾病对人们认知功能的不利影响,从而帮助人们重搭自主思考与正确判断的能力。有学者指出,人类可以与社交机器人建立朋友关系,从而提高人们对他人的容忍度和包容度,这有利于人类美德的培育和发展^[30]。由此可见,社交机器人在交互中可以帮助人类养成良好的道德倾向,修正道德意义上的“不善”和“不美”之处。

但值得注意的是,社交机器人的修正作用是有限的。一方面,现有技术下的社交机器人只能满足人类低层次的需求,尚难以实现人与人之间深刻的情感交流^[31];另一方面,社交机器人对人类真实性的调节必须在理性社会所允许的框架下进行,否则就有可能出现“美而不真”或“善而不真”的问题。例如,子女依靠社交机器人照顾老人,却忽视了行孝应该躬亲的原则,反而因“孝行”难以彰显真正的孝心而出现“善而不真”的问题^[32]。

四、结语

社交机器人与传统机器人存在很大区别,在一定程度上已经超越了传统意义上机器人的界限,呈现了一些新的特征^[33]。社交机器人所引发的真实性问题不仅反映出社交机器人在设计与使用过程中所存在的不足,更表明技术不但重塑了世界的展现方式,而且对人类生活方式与价值观念具有一定的塑造作用。在技术与人类社会深度融合的背景下,人们要做的不是将社交机器人彻底从个人生活或人类社会中剥离开来,而是应合理设定社交机器人对人类各方面的调节界限,充分发挥社交机器人对人类真实性的积极作用,防止自身被社交机器人控制而失去真实自我。必须确保人的理性不被社交机器人淹没、感性不被其利用、思维不被其取

代。如此,人们才能在利用社交机器人推动人类社会发展的同时,实现感性与理性、功利性与价值性、真实性与社会理性的和谐统一,从而促成人的全面发展与“真善美”的和谐统一。

[参考文献]

- [1] 王亮.基于情境体验的社交机器人伦理:从“欺骗”到“向善”[J].自然辩证法研究,2021(10):55-60.
- [2] 窦笑.社交机器人伦理问题与政策建议研究[J].智库理论与实践,2022(5):103-110.
- [3] Varga S. Authenticity as an ethical ideal[M]. Routledge, 2013:13-15.
- [4] 张照生,曾繁亨.论莱昂内尔·特里林的“本真”之思[J].安徽大学学报(哲学社会科学版),2022(5):22-31.
- [5] Gordon E C. Cognitive enhancement and authenticity: moving beyond the impasse[J]. Medicine, Health Care and Philosophy, 2022, 25(02):281-288.
- [6] 舒红跃.技术之于此在:从非本真性到本真性[J].科学技术哲学研究,2013(2):81-85.
- [7] 张卫,王前.道德可以被物化吗?:维贝克“道德物化”思想评介[J].哲学动态,2013(3):70-75.
- [8] 114只机器狗寿终正寝,日本寺庙为它们举行了一场葬礼[EB/OL].Ifanr,https://www.ifanr.com/1024518.[2018-05-05].
- [9] Topol E J. Machines and empathy in medicine[J]. The Lancet, 2023, 402(10411):1411.
- [10] Fabry R E, Alfano M. The affective scaffolding of grief in the digital age: the case of deathbots[J]. Topoi, 2024, 43(03): 757-769.
- [11] 刘红玉,李理.类机器人格研究:基于人机交互视角[J].长沙理工大学学报(社会科学版),2020(6):10-17.
- [12] Cheng C, Luo Y, Yu C B. Dynamic mechanism of social bots interfering with public opinion in network[J]. Physica A, 2020, 551:124163.
- [13] 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(7):154-163.
- [14] 喻国明,曲慧.“信息茧房”的误读与算法推送的必要:兼论内容分发中社会伦理困境的解决之道[J].新疆师范大学学报(哲学社会科学版),2020(1):127-133.
- [15] Ennis R H. A taxonomy of critical thinking dispositions and abilities[J]. Teaching Thinking Skills: Theory and Practice, 1987:9-26.
- [16] 温凤鸣,解学芳.欧美社会治理中算法应用的伦理风险与规制路径研究[J].西南民族大学学报(人文社会科学版),2023(9):130-139.
- [17] Zhou Y Y, Fei Z Y, He Y Q, et al. How human-chatbot interaction impairs charitable giving: the role of moral judgment[J]. Journal of Business Ethics, 2022, 178(03):1-7.
- [18] Laakasuo M, Palomaki J, Koebis N. Moral uncanny valley: a robot's appearance moderates how its decisions are judged[J]. International Journal of Social Robotics, 2022, 14(01):295-296.
- [19] 李平安.延异视角下的拟人化机器人认知偏差问题审视[J].科学技术哲学研究,2024(1):59-66.
- [20] Russell B. Robots and reality: a reply to robert sparrow[J]. Ethics and Information Technology, 2012, 14(01):41-51.
- [21] 王亮.社交机器人“单向度情感”伦理风险问题刍议[J].自然辩证法研究,2020(1):56-61.
- [22] 逯改.智能养老护理机器人的伦理反思与建议[J].医学与哲学,2022(13):47-51,56.
- [23] 任东景.人工智能视域下人的全面发展问题探析[J].长沙理工大学学报(社会科学版),2020(5):30-34.
- [24] 赵蓓.信息失序的背后:社交机器人对网络传播秩序的影响[J].青年记者,2023(2):23-27.
- [25] Sharkey A, Sharkey N. We need to talk about deception in social robotics! [J]. Ethics and Information Technology, 2021, 23(03):309-316.
- [26] Duan F, Li W, Tan Y. Overview of robots[M]. Singapore: Springer Nature Singapore, 2023:16.
- [27] 景星维,王思奇.论智能时代人的价值自觉[J].理论导刊,2023(9):74-80.
- [28] Jøranson N, Pedersen I, Rokstad A M M, et al. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial[J]. Journal of the American Medical Directors Association, 2015, 16(10):867-873.
- [29] D'Onofrio G, Sancarolo D, Raciti M, et al. MARIO project: validation and evidence of service robots for older people with dementia[J]. Journal of Alzheimer's Disease, 2019, 68(04):1587-1601.
- [30] Ryland H. It's friendship, Jim, but not as we know it: a degrees-of-friendship view of human-robot friendships[J]. Minds and Machines, 2021, 31(03):377-393.
- [31] 陈嘉华,丁晓军,喻丰.护理机器人辅助居家养老的伦理困境研究:基于价值敏感设计理论[J].中国医学伦理学,2023(12):1350-1357.
- [32] 王健,林津如.护理机器人补位子女养老的伦理风险及其防范[J].道德与文明,2019(3):131-137.
- [33] 易显飞,刘壮.社会化机器人主要特征的技术哲学审视[J].晨刊,2022(4):21-24.