



多模可信交互：从多模态信息融合到人–机器人–数字人三位一体式交互模型

王国庆^{1*}, 裴云强¹, 杨阳¹, 徐行¹, 汪政¹, 申恒涛^{1,2}

1. 电子科技大学计算机科学与工程学院, 成都 611731

2. 鹏程实验室, 深圳 518055

* 通信作者. E-mail: gqwang0420@uestc.edu.cn

收稿日期: 2023-05-11; 修回日期: 2023-08-03; 接受日期: 2023-10-20; 网络出版日期: 2024-04-08

国家自然科学基金 (批准号: U23B2011, 62102069, U20B2063, 62220106008)、四川省科技计划 (批准号: 2022YFG0032) 和新基石科学基金 (批准号: XPLOER PRIZE) 资助项目

摘要 信任在人–机器人–数字人协作中扮演着关键角色, 因为它不仅影响了人–机器人–数字人的效率, 同时也存在风险和益处. 然而, 当前的人–机器人–数字人信任研究存在“失衡”现象, 即大部分研究集中于“以人为中心”的信任关系而忽略了智能体 (机器人和数字人) 对人的信任. 为了填补人–机器人–数字人信任研究领域的这一空白, 在使用多模态信号来建立人–机器人–数字人交互中的适度单向信任基础上, 构建跨虚实世界的“三位一体”经历交融共享生态. 本文旨在探讨多模态线索和增强现实在建立人–机器人–数字人之间可信关系方面的现有研究、可行性和未来发展方向. 最后, 本文展望了该模式在社会组织形态及社会事件态势感知和管控中的应用前景, 并指出了未来需要解决的问题. 本文的研究有助于理解多模态线索和增强现实在人–机器人–数字人交互中的作用, 并为“三位一体”趋势的实现提供思路和解决方案.

关键词 增强现实, 多模态交互, 人–机器人–数字人交互, 信任

1 引言

人机交互作为一个跨学科的研究领域^[1,2], 涉及多个学科的交融与发展. 在人工智能和计算机科学的发展推动下, 人机交互领域正逐渐从计算机向智能体演变^[3], 从现实向虚实融合转变^[4~6], 以及从心理学向社会学跨越^[7]. 在这个过程中, “人”, “机器人”和“数字人”作为交互的主体, 开始在人机交互领域中崭露头角. “人”是交互的发起者和接收者, 他们通过自己的感知和行动能力, 与机器人和数字人进行交互. “机器人”是一种具有物理形态的智能体, 它们可以执行各种任务, 如搬运、清洁、探索

引用格式: 王国庆, 裴云强, 杨阳, 等. 多模可信交互: 从多模态信息融合到人–机器人–数字人三位一体式交互模型. 中国科学: 信息科学, 2024, 54: 872–892, doi: 10.1360/SSI-2023-0133
Wang G Q, Pei Y Q, Yang Y, et al. Multimodal trustworthy interaction: from multimodal information fusion to a trinitarian human-robot-digital human interaction model (in Chinese). Sci Sin Inform, 2024, 54: 872–892, doi: 10.1360/SSI-2023-0133

等。“数字人”是一种虚拟实体,它们超越了单纯的机器执行任务,拥有了思考、情感表达和智能决策的能力,通过与人进行自然语言对话和行为交互,在各种领域和场景中发挥重要的作用^[8]。在人-机器人-数字人交互的过程中,每个主体都有其独特的作用。人作为交互的发起者,提出需求和指令;机器人作为执行者,完成各种物理任务;数字人作为智能交互者,提供思考、决策和情感表达。根据 Fan 等^[1]的观点,未来的人机交互将主要表现为“交互人”与“智能体”在多个维度上的感知性交互,而非仅限于计算性交互。当前,这种感知性交互在诸如工业、医疗、教育和娱乐等应用场景中的人与智能机器人的交流互动方面得到了显著体现^[9,10]。本文将重点探讨如何通过多模态线索和增强现实技术^[11,12]来建立和维护人类、机器人和数字人之间的可信关系,以促进更自然、更有效、更安全、更有趣、更有价值的人-机器人-数字人协作。这种协作模式不仅将推动人机交互领域的发展,也将为我们的生活带来前所未有的便利和乐趣。

“人-计算机交互”(human-computer interaction, HCI)和“人-机器人交互”(human-robot interaction, HRI)是两个相关但不同的概念^[13,14]。前者主要研究人类与计算机之间信息传递和处理过程中涉及的技术、方法和理论问题;后者则主要研究人类与具有自主移动能力和执行任务能力的智能机器人之间的交流和互动。两者之间的区别在于, HCI 主要关注人类与计算机之间的信息交互,而 HRI 则更关注人类与机器人之间的协作和互动。在这个背景下,我们引入了“人-机器人-数字人交互”的概念。这是一种更为复杂的交互模式,它不仅包括人与机器人的信息交互、协作和互动,还包括人与数字人的交互和数字人与机器人的交互。

1.1 人-机器人-数字人之间交互的发展

人-机器人交互(human-robot interaction, HRI)这一概念最早可追溯至小说家艾萨克·阿西莫夫(Isaac Asimov)于1941年在其小说《I, Robot》中所提出的,如图1所示。值得注意的是,这一概念出现在实际机器人问世之前^[15]。人-机器人交互主要研究人类与机器人在不同环境中的交流和互动^[15]。图1展示了人-机器人交互历史发展进程。早期研究表明,人们倾向于将计算机“人格化”,从而促进人机之间的有效协作。自20世纪中期以来,美国和英国先后研制出了工业机器人 Unimate 和 RAMP。然而,在1970年之前,两国在机器人研发方面屡遭失败,一度陷入低谷。自1980年起,各国开始重新重视机器人领域。1999年,美国总统顾问委员会的报告将“人机交互和信息处理”确定为21世纪信息技术的主要研究方向之一。2007年,“以人为本的计算”被美国国家科学基金会列为核心技术领域,其中包括多媒体智能、数据可视化和高效人机交互模型等研究方向。2007年,欧盟第七框架计划也将“自然人机交互”技术纳入其研究范畴^[1]。自21世纪初以来, HRI 领域逐渐受广泛关注,越来越多的研究者开始探讨人们如何看待不同类型的机器人^[15]。2011年,美国政府启动了国家机器人计划,将研究重心逐步从人类单向控制转向人-机器人交互。该计划提供的研究证据表明,在人-机器人交互中充分利用人类和机器人的各自优势可以扩展或增强人类能力,从而优化人-机器人协作效率。自2012年起,“人-机器人交互”被美国计算机协会视为计算机科学的重要分支,这标志着人-机器人交互在该领域的日益重要地位。2016年,国家自然科学基金委将人-机器人交互列为《国家自然科学基金“十三五”发展规划》中的重点研究课题之一^[1]。2017年,国务院印发的《新一代人工智能发展规划》将混合增强智能和人机协同纳入规划部署的重点领域之一^[16]。2018年,《科学》杂志旗下的《机器人》杂志将“社交和情感互动”列为未来十年机器人技术面临的十大挑战之一。2021年,《“十四五”机器人产业发展规划》将“人机自然交互技术、情感识别技术”列为重点攻关的机器人前沿技术。同年,国家将虚拟数字技术的发展纳入“十四五”规划,为数字人的发展提供了良好的发展环境和未来发展空间。在2023年1月,《“机器人+”应用行动实施方案》中,中华人民共和国工业和信息化部明确

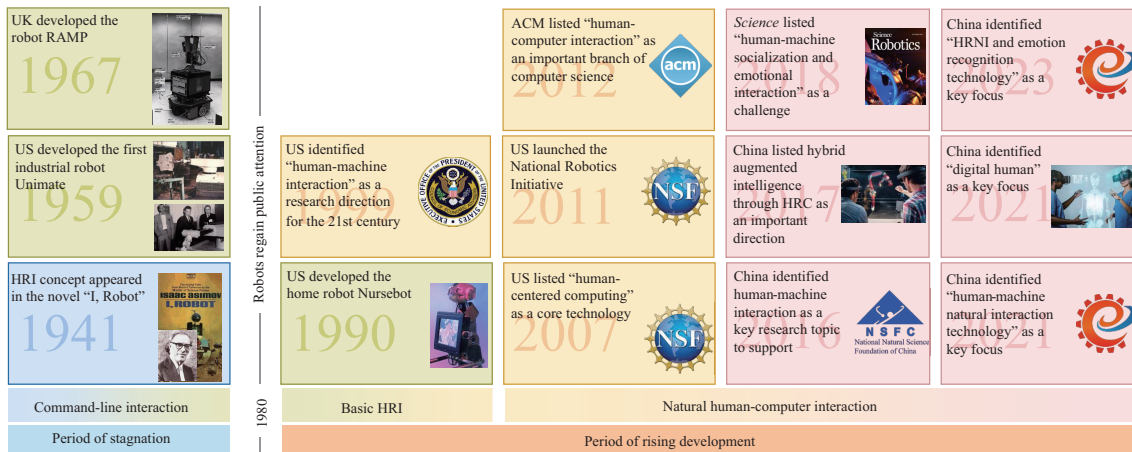


图 1 (网络版彩图) 人-机器人-数字人之间交互的历史发展进程

Figure 1 (Color online) Historical development process of interaction between humans, robots, and digital humans

提出将“多模态量化评估、多信息融合情感识别”作为深化重点领域“机器人+”应用的核心技术。这表明, 机器人产业的发展迫切需要自然交互技术。清华大学徐华教授的研究团队在近年出版的专著中指出, 以“人-机器人交互意图理解”和“多模式情感分析”为代表的人机自然交互已经引起了广泛关注, 并成为学术界深入研究的课题。人-机器人(自然)交互的目标在于定义与机器人交互的人类期望模型, 以指导机器人设计, 从而开发自然有效的人-机器人交互。另一方面, 清华大学史元春教授坚持“人机共生”的理念, 强调机器应适应人的本性和能力。该理念区分了人-机器人交互和人工智能的研究重点和方法, 但认为两者有共同的目标和趋势, 即增强人和机器人之间的交互和感知。2022年, 北京市经济和信息化局发布了国内首个数字人产业专项支持政策——《北京市促进数字人产业创新发展行动计划(2022-2025年)》, 并提出目标: 到2025年, 北京市数字人产业规模将突破500亿元。这些政策和规划的出台, 无疑为人-机器人-数字人交互的发展提供了强大的政策支持和广阔的应用前景。在人-机器人-数字人交互的发展过程中, 从侧重于交互设计和实现的人-机器人交互, 到以人为中心的计算, 再到去中心化的人-机器人-数字人共生系统^[1], 信任都发挥着关键作用。信任是一种心理状态, 包括对另一实体的可靠性、完整性和仁慈的信念^[17]。在人-机器人-数字人交互中, 三者相互之间的信任是一个复杂且多维的结构, 受多种因素影响^[18]。高信任度有助于提高任务绩效和协作效率; 低信任度可能导致不必要的监督或检查。将信任融入机器人或数字人决策模型可以提升人-机器人-数字人团队绩效和交互体验。因此, 建立和维护相互信任对于有效的人-机器人-数字人交互是至关重要的。面向未来“人-机器人-数字人共生”的协作型机器人和数字人的应用, 所具备的以“相互理解、相互感知、相互帮助”为核心的人-机器人-数字人互信交互需要在意图层面和情感层面实现双向对齐^[19]。然而, 谷歌学术的检索结果显示, 截至目前, 关于人-机器人-数字人信任的研究论文中, 有99.8%的文献集中于研究人类对机器人或数字人的信任, 而只有0.2%的文献涉及机器人或数字人对人类的信任。这表明, 目前关于人-机器人-数字人信任的主流研究偏向“以人类为中心”, 而忽视了协作伙伴之间的相互信任。鲜有团队从机器人或数字人的角度出发, 研究机器人或数字人对人类的信任程度^[20,21], 这就导致人-机器人-数字人互信研究失衡, 难以实现人-机器人-数字人协作过程中的相互信任, 并将最终损害人类使用智能系统的预期社会和经济效益^[22]。因此, 人-机器人-数字人都必须关注建立和维护信任, 以保障有效稳定的协作关系^[23]。

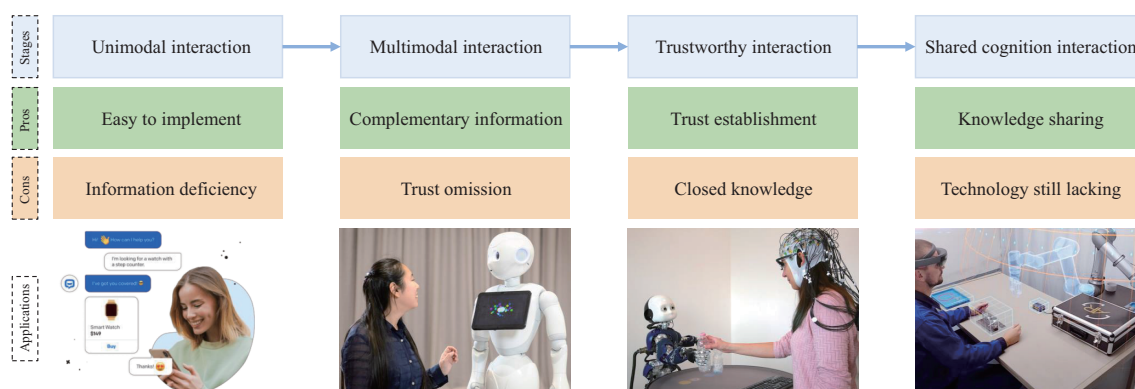


图2 (网络版彩图) 从“单模态交互”到“共享认知交互”: 人-机器人-数字人交互的4大发展阶段

Figure 2 (Color online) From “unimodal interaction” to “shared cognition interaction”: four stages of human-robot-digital human interaction development

1.2 从单模态线索到共享认知线索

单模态线索可能导致信息的不完整性或误解, 原因在于它忽略了其他感官输入所提供的有益信息. 研究表明, 相较于单模态线索, 多模态线索更为有效^[24~26], 因为它们能更有效地激活大脑中负责处理不同输入的区域. 多模态线索是指通过多种感官渠道(如面部表情、身体运动、语言、触摸和皮肤传导)传递信息, 从而促进人-机器人-数字人交互并建立及维持人-机器人-数字人之间的信任. 多模态线索既可以源自人类, 也可以源自机器人和数字人. 人类产生的多模态线索为机器人数字人提供了关于用户心理、生理和社会学状态的宝贵信息^[27~30], 有助于机器人理解并满足用户需求. 而机器人产生的多模态线索则通过语言、动作等方式向人类传达其状态、意图和计划^[31]. 然而, 多模态人-机器人-数字人交互过分关注表象, 却忽略了可信度. 图2展示了人-机器人-数字人交互的4个发展阶段: 单模态交互、多模态交互、可信交互和共享认知交互. 单模态交互只使用一种感官渠道, 易于实现, 但信息量有限, 容易导致误解. 这种交互方式适用于简单的命令行或基本的人-机器人-数字人交互^[24]. 为了提高信息的完整性和准确性, 多模态交互使用多种感官渠道, 如视觉、听觉、触觉等, 提供互补的信息^[32]. 这种交互方式适用于自然的人-机器人-数字人交互或复杂的任务协作. 然而, 多模态交互并不一定能够建立和维护信任关系, 这对于高风险或高收益的领域是至关重要的. 因此, 可信交互通过多模态线索和增强现实技术来提高交互效率和协作效果, 降低风险和成本^[33]. 这种交互方式适用于医疗、军事、教育等领域. 最后, 共享认知交互通过多模态线索、增强现实技术和共享认知机制实现知识和价值的双向对齐, 实现更深层次的理解和协同, 提升创新能力和社会效益^[9]. 这种交互方式适用于未来的社会组织和事件态势感知等领域. 图2从易于实现、信息完整性、信任建立和知识共享4个方面对比了各个阶段的优缺点, 并列举了各个阶段的典型应用领域.

本文旨在探讨信任在人-机器人-数字人交互中的重要性, 以及多模态线索和增强现实在建立人-机器人-数字人之间互信关系方面的现有研究、可行性和未来发展方向. 首先, 本文阐述了人-机器人-数字人交互中信任的理论基础和作用, 探讨了多模态交互信号在缓解过度信任和修复信任受损方面的贡献, 并介绍了信任评估方法. 接着, 文章分析了多模态人-机器人-数字人交互和增强现实在实现互信交互中的重要性, 并探讨了它们之间的关联. 最后, 本文提出了一种“三位一体”式跨媒体虚实混融交互构想, 并探讨了该构想在社会组织和事件态势感知中的应用潜力和挑战.

2 共享式多模可信人-机器人-数字人交互

共享式多模可信人-机器人-数字人交互不仅涉及多模态信息的获取、融合和呈现,还涉及多模态认知线索的共享、协调和对齐。多模态认知线索是指人类和机器人在交互过程中产生和感知的与其内在状态和行为意图相关的多种感官信号,如面部表情、肢体动作、语音语调、触觉反馈等。通过增强现实技术,可以将这些认知线索可视化地呈现给交互对象,从而实现人-机器人-数字人之间的认知线索共享。认知线索共享有助于提高人-机器人-数字人之间的相互信任,从而促进更有效和自然的人-机器人-数字人协作。

2.1 多模态人-机器人-数字人交互

多模态线索是指通过不同的感官渠道(如视觉、听觉和触觉)传递信息,例如面部表情、身体运动、语言、触摸和皮肤传导^[34]。这些线索可以促进人-机器人-数字人交互并建立及维持人-机器人-数字人之间的信任。多模态线索既可以源自人类,也可以源自机器人和数字人。

2.1.1 多模态人-机器人-数字人交互的类型

麦拉宾法则(Mehrabian's rule)明确指出,人们在理解他人意图和情感等信息时,不仅依赖于语言交流的内容,还需考虑非言语信息,其中视觉信息、听觉信息和语言内容分别占据55%、38%和7%^[35]。该法则被广泛应用于人类交流和人-机器人-数字人交互等领域的研究。为实现人-机器人-数字人交互的自然模拟,需充分考虑人类交互过程中视听觉和语义信息的重要性。由于单模态线索无法充分反映人类交互的复杂性,因此需借助多模态线索,以更好地在人-机器人-数字人之间模拟人类之间的交互。视听觉和语义信息的结合能够提供更全面、准确的交互线索,从而使人-机器人-数字人交互更加自然、有效^[36]。多模态线索涵盖听觉、视觉和触觉3种类型,有助于机器人或数字人理解并回应人类的心理、生理和社会学状态,进而建立机器人或数字人对人类的“信任”^[37]。多模态线索能够提供更丰富、更准确、更自然的交互信息,从而增强人类对机器人或数字人的信任感和依赖度,同时也能够提高机器人或数字人对人类的信任感和适应度。

语言作为人-机器人-数字人沟通的主要手段^[38],能够直观地传达人类的意图和情绪。自然语言界面一直是人-机器人-数字人交互的重要研究课题。此外,机器人或数字人还可以通过协作性的语言基础来建立对语言的共同理解,这对于成功的人-机器人-数字人交互至关重要。人-机器人-数字人通过语言沟通来更好地分析人类用户的意图、行为和计划,并作出合理的响应^[39]。值得注意的是,语言不仅可以传达人类意图和行为,还可以通过语音语调传达人类情绪^[40]。例如,当机器人或数字人检测到人类通过言语表达沮丧时,它可以将此解释为不满意的迹象,并及时向人类提供帮助。

肢体动作,如身体姿势、手势和视线,能够反映人类的心理(情绪和偏好)、生理(人机距离)和社会学(社会关系)状态信息^[41,42]。在人际交互中,人类通过肢体动作向交互对象传达自身意图、兴趣、感受和想法,因此肢体动作在交互过程中具有重要意义。同样地,机器人或数字人也能通过观察人类的肢体动作来理解人类的需求和偏好,并做出适应性响应。常见的人体姿势检测设备包括可穿戴式传感器和能检测骨架数据的Kinect传感器^[42]。这种非语言沟通使得机器人或数字人能够主动感知人类行为,通过观察理解人类意图,从而更好地提供服务或协作。李德毅院士在2022年的中国工程科技论坛上的演讲《机器具身交互智能》中指出,多通道跨模态交互(包括肢体动作)对具有具身智能的机器来说至关重要,因为它可以实现行为交互和机器学习。基于具身指称理解(embodied reference understanding, ERU)的人-机器人-数字人交互中,典型的YouReflit多模态融合网络架构在视觉语言双通道基础

上,加入人类手势模态^[43,44],增强机器人或数字人对人类意图的理解.这对构建信任至关重要,因为机器人或数字人能够利用这些信息更好地与人类协作.北京大学朱松纯教授和朱毅鑫教授的团队在人机对话中结合心智模型的研究颇为丰富.他们认为,这样可以比单纯用视觉和语言更好地理解对方.他们建立的模型也能更好地泛化,适应不同的场景.然而,与人类相比,这些模型仍难以从已知信息推理出未知信息^[45].因此,未来仍有很大的提升空间.若机器人或数字人能够运用手势和视线来增强对人类意图的理解,便能更有效准确地响应需求,建立互信.除了提供有关人类心理和生理状态的信息外,肢体线索还可以提供关于其社会因素的信息,如社会角色或关系.机器人或数字人通过理解社会线索,如眼神和手势,以确定该人在群体中的权力或地位,或确定他们与其他群体成员的关系.这些信息可以被机器人和数字人用来调整自己的行为 and 反应,例如对权力或地位较高的人更加恭敬,或者对被群体孤立的人更加支持.同时,人-机器人-数字人的社交距离也是肢体动作的重要体现^[46].进化者公司的智能机器人能够通过距离传感器感知人和机器人之间的社交距离等社交因素,并将社交距离和手势等线索相结合,提升机器人的人性化程度^[47].

面部表情作为跨文化跨语言的通用表达,能够传达各种情绪.机器人或数字人可借助面部表情检测设备来识别人类的情绪状态,并对面部表情的变化作出动态和实时的响应^[48].面部表情检测有助于确保人-机器人-数字人协作的安全性,例如通过使用红外摄像机或多通道电极系统检测疲劳驾驶,并警示司机或控制车辆驾驶权以避免事故发生^[49].此外,在情感识别中,通过识别人体的情感信号,创建一种能够感知、识别和理解人的情感,并且能够做出智能、灵敏和友好反应的机器人系统.在人类面部表情识别方面,机器人或数字人可以通过摄像头或传感器捕捉人类的面部表情,并根据预先定义的类别(如快乐、悲伤、惊讶等)进行识别和分类,从而理解人类的情绪和意图.这可以帮助机器人提供更合适的反馈和服务,增强人-机器人-数字人之间的信任.在机器人或数字人面部表情生成方面,机器人或数字人可以通过自身的面部结构或显示屏模拟出不同的面部表情,并根据自身的状态或任务进行调整,从而向人类传达自己的情绪和意图.这可以帮助机器人或数字人增加自己的可信度和吸引力,促进人-机器人-数字人之间的沟通和协作.

触觉作为机器人或数字人更好地理解 and 响应人类意图的重要信息来源,对人-机器人-数字人交互具有积极影响.触摸既为机器人或数字人提供来自人类的触觉信息,又为人类提供触觉反馈,以模拟人类之间的肢体接触.这有助于机器人或数字人更好地理解 and 回应人类的情感.此外,为机器人或数字人提供触觉反馈可以帮助它们更好地识别物体并与环境互动.自动驾驶汽车中的多模态人-机器人-数字人交互使用多种交互模态线索,如结合视觉和触摸线索以提高驾驶安全性.通过视觉线索传达有关汽车状态和能力的信息,并将振动和语音提示等视觉和触摸反馈相结合,可及时提供驾驶过程中的危险预警^[50].此外,在操作任务中,使用触觉信息作为机器人的控制参数,例如接触点估计信息、表面法向和曲率等;在抓握任务中,通过测量法向静态力来检测物体滑动情况,例如将接触力的测量值用于辅助抓握力控制;在机器人的灵巧操作任务中,判断施加操作用力的方向也至关重要,例如通过调节法向力和切向力之间的平衡,能够保证抓握的稳定性.

多模态线索(亦称多模态信号),包括面部表情、身体动作、语音和触摸等,能够为机器人和数字人提供丰富的信息,以了解其人类用户的心理、生理和社会状态^[34].其中,社会状态指的是一个人在社会中的地位、角色和关系.例如,身体动作和面部表情可以传达一个人在群体中的权力或地位,或确定他们与其他群体成员的关系.因此,探讨多模态信号不同组合方式对人-机器人-数字人交互体验的影响具有重要意义.为了探讨多模态信号不同组合方式对人-机器人-数字人交互体验的影响,可构建如图3所示的多模态线索检测框架.除了人类主动交互提供的多模态线索外,为实现更有效安全的人-机器人-数字人交互,还应将包括摔倒、抽搐等异常身体反应在内的非主动交互纳入需融合

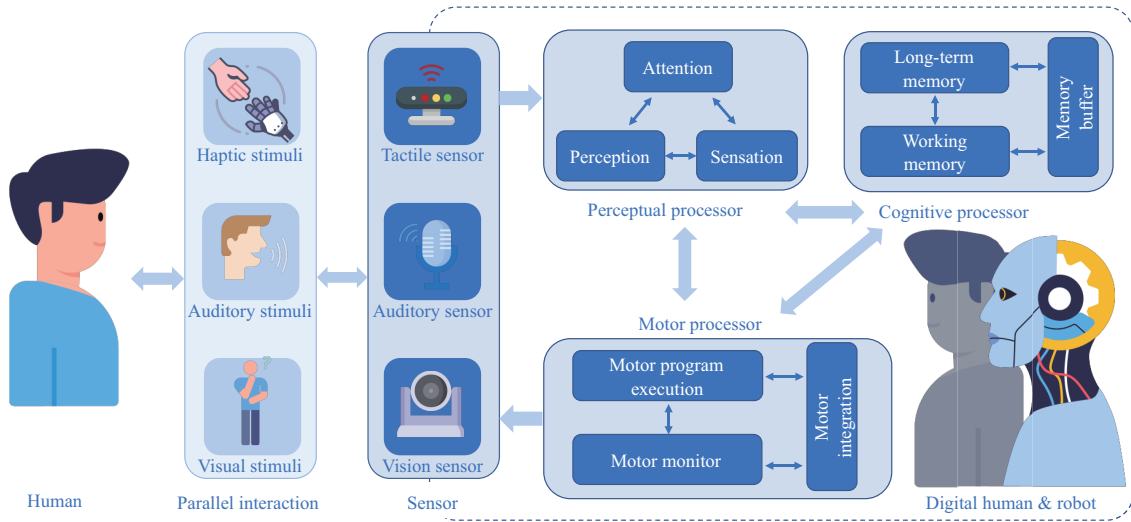


图 3 (网络版彩图) 心智理论模型下多模态人-机器人-数字人协作计算模型框架

Figure 3 (Color online) Multimodal human-robot-digital human collaboration computing model framework based on the mind theory model

的多模态线索中^[47]。机器人通过利用多模态线索,如视听触觉和生理信号等心理、生理和社会线索,能更好地理解用户需求和偏好,并作出适应性响应。多模态交互有助于建立信任^[51],强化沟通和合作,提高生产力。Torta 等^[24]的研究发现,与单模态交互(平均用时 4.8 s)相比,多模态交互(平均用时 1.8 s)能更快地吸引人们的注意力,从而提升交互的安全性。Salem 等^[31]的研究表明,当机器人在说话的同时做手势(即多模态交互)时,所有的品质评分(包括活跃度、生动性、同理心和能力)都有所提高。Maniscalco 等^[51]的研究也进一步证明了在共享环境中使用多模态线索识别并实时响应用户反馈的机器人或数字人可以提高自身的可信度。随着社会对机器人或数字人兴趣的日益浓厚,研究机器人或数字人作为协作者的方向变得越来越有价值。具体而言,研究应关注如何通过有效融合和适当响应这些线索,创造双向价值对齐和互信的人-机器人-数字人交互环境。

2.1.2 多模态融合计算框架

在 20 世纪 80 年代,心理学家开始采用连接主义神经网络来模拟人类的认知处理,其中典型的模型便是贝叶斯网络模型。随着研究的深入,学者们发现人类以多种方式表征外部世界,而这些认知方式深深植根于身体之中,基于具身认知(embodied cognition)和扎根认知理论(grounded cognition)。扎根认知理论主张,大多数储存在大脑中的表征知识源于最初呈现在感觉器官上的外部信息的直接表征。基于这一理论,中国科学院心理研究所的刘焯教授团队^[52]提出了人-机器人-数字人协作心理模型,如图 3 所示。图 3 展示了基于心智理论模型的多模态人-机器人-数字人协作计算模型框架。心智理论模型是指人类能够推测他人的心理状态,如信念、意图、情绪等,并根据这些心理状态来预测和解释他人的行为^[52,53]。该模型综合考虑了人类和计算机的感知、认知和行动模块,以及多模态融合和并行交互计算。感知模块负责接收和分析来自人-机器人-数字人的多模态线索,如视觉、听觉、触觉等,并将这些线索转化为感觉信息。认知模块负责对感觉信息进行加工和存储,并利用长期记忆和工作记忆来形成心智模型。行动模块负责根据心智模型生成和执行运动计划,并通过运动监控和运动整合来调整运动输出。该模型可以实现人-机器人-数字人之间的并行交互,即人-机器人-数字人可以同时进行感知、认知和行动,从而提高交互效率和协作效果。该模型也可以实现多模态融合,

即将不同类型的多模态线索进行有效的组合和表征,从而提高信息的完整性和准确性.该模型旨在实现更贴近人际交互的人-机器人-数字人交互模式,用视觉、听觉、触觉和手势等更自然的多模态交互模式取代传统的 WIMP (windows, menu, icons, pointing devices) 范式中不符合人类用户自然交互习惯的交互媒介与界面元素.

此外,不同类型的多模态信息融合可以组合成复杂的结构,用于多任务学习和跨模态学习^[54].多模态信息融合主要包括基于时序、基于层次和基于处理方式的融合策略.基于时序的多模态信息融合可根据发生的时间分为早期、中期和后期融合^[54].早期融合在机器学习模型处理数据之前将多种模态的数据合并,并在特征层面进行融合表征;中间融合将在不同模态上训练的多个模型的输出相融合,在模型层面对融合结果进行表征;后期融合则将多个模型在不同模态上训练后的输出结合在决策层面进行融合表征^[54].例如,文献^[55]对不同的多模态信息融合策略在动作识别任务上进行了实验比较.实验结果表明,在使用 RGB、骨架和深度 3 种模态的数据时,早期融合策略达到了 86.7% 的准确率,中期融合策略达到了 95.9% 的准确率,而后期融合策略达到了 94.3% 的准确率.这说明,在动作识别任务上,中期和后期融合策略能够更好地利用多模态信息,并提高识别性能.多模态深度学习 (multimodal deep learning) 是数据融合的一个重要方法,它利用卷积神经网络 (convolutional neural network, CNN)、递归神经网络 (recurrent neural network, RNN) 和自注意机制等深度学习技术来融合多种模态的数据^[56],并在融合阶段能够无差别地处理多通道信息^[54].相较于传统的信息融合策略,这些技术可以实现更好的预测和更准确的分类^[56].除了基于时序的融合策略外,还有基于层次和基于处理方式的融合策略.基于层次的融合策略涉及在数据 (特征) 层面、模型层面和决策层面发生的融合,与基于时序的融合策略相似.而基于处理的多模态融合策略则包括基于规则的融合或统计 (基于机器学习) 的融合^[54].贝叶斯决策是基于统计的多模态融合策略,它可以使用基于不完整信息的主观概率估计来修正发生的概率,并根据预期值和修正后的概率做出最佳决策^[54].在多模态联合分布概率部分已知的情况下,贝叶斯决策可以根据历史经验推出某些缺失的模态信号,从而最优化整体多模态信号融合效果.

在多模态信息融合的基础上,信任成为人-机器人-数字人协同共生发展的关键因素^[57],尤其在自动驾驶领域.在自动驾驶领域,执行阶段需将任务决策转化为控制指令,以实现任务目标并解决实时响应问题.为达成此目标,人-机器人-数字人之间须优化协作、合理分配权限、互补能力,以增强人与系统能力的融合.因此,混合智能技术的两个关键领域需要取得突破:人-机器人-数字人交互技术和智能可信技术.多模态融合计算框架能够实现更高效、更准确、更灵活的信息处理和决策,从而提高人类对机器人或数字人的信任感和满意度,同时也能够提高机器人和数字人对人类的信任感和适应度.人-机器人-数字人交互技术关注人与智能体之间的无缝信息交流,通过适应驾驶员的生理、情感、动机和认知特点,在感知通道 (如视觉、听觉和触觉) 和行为通道 (如生理、肢体和语音) 层面建立自然交互方式.这有助于优化自动驾驶汽车对驾驶员建立的心智模型,提高个性化响应能力,改善操作环境,并不断提高交互水平:从操作交互到知识交互,再到意识交互.智能可信技术的目标则是通过提升智能系统的可解释性,实现人-机器人-数字人知识深度融合和人-机器人-数字人价值三向对齐,从而建立人-机器人-数字人之间的相互信任.结合这两个突破性技术,可以进一步提升交互水平:从意识交互到互信交互.提高人工智能的可解释性是建立人类与智能体 (机器人和数字人) 之间信任的关键因素.相反,无法解释的人工智能决策将阻碍人-机器人-数字人互信的建立.尽管人工智能设计者能够解释模型设计原则和细节,但他们无法解释人工智能决策的缘由.由于缺乏有效的解释机制,当人工智能出现错误时,难以让人类理解其原因,这将进一步阻碍人-机器人-数字人互信的建立.人-机器人-数字人信任的建立受阻于隔绝人机深度交互的“黑箱”,“黑箱”的形成源

于元数据的缺乏、数据预处理的不当、智能算法的不透明、数据授权机制的模糊. 突破“黑箱”障碍是实现人-机器人-数字人互信的必要条件, 当前可行的解决思路是通过沉浸式的自然人-机器人-数字人交互手段增强对复杂数据的解释. 此外, 多模态线索可以帮助人和机器人更全面地认识彼此的优劣势. Wang 等^[58]提出了“人机互信的知识自动化与混合增强智能”的概念, 即人类智能和机器智能在复杂系统认知管控机制中的协作. 该模型确保人-智能体双方清楚彼此的能力上下限, 从而建立互信, 并为共同的目标而协作. 人-机器人-数字人互信对于智能化战争至关重要, 因为它需要人与装备的紧密联系与深度协同. 人-机器人-数字人互信意味着人工智能辅助决策不仅要提供结论, 还要提供先决条件和适用范围等附加信息, 以保证人-机器人-数字人间的高效协同互补. 可信人-机器人-数字人交互模式能够实现更高层次、更深度、更广泛的信息共享和知识融合, 从而提高三方对彼此的信任感和依赖度, 并实现价值双向对齐和目标共同追求. 人-机器人-数字人一体化通过识别手势、语义、行为等多模态线索来实现高效交互, 建立人-机器人-数字人间的互信交流机制, 实现无人系统与有人系统的快速联动. 为了进一步构建人-机器人-数字人互信交互机制, 在以上人-机器人-数字人协作心理模型“感知, 认知, 行动”优势的基础上, 增强现实能够通过提供“共享, 历史和计划”来强化该优势. 通过增强现实技术, 可以实现更高效、更自然、更可靠的人-机器人-数字人交互方式.

2.2 可信人-机器人-数字人交互模式

在人-机器人-数字人交互中信任的重要性及其社会心理学理论基础. 随着机器人和数字人技术的发展, 人-机器人-数字人交互越来越受到关注. 在这一背景下, 信任成为了一个关键因素, 因为它影响着人-机器人-数字人之间的关系和协作效果. 社会心理学中的信任理论为研究人-机器人-数字人交互中信任提供了理论基础, 主要包括基于认知的理论、基于情感的理论、心理契约理论和社会交换理论^[17, 59, 60]. 这些理论虽然最初是用于解释人际信任的形成机制, 但随着机器人和数字人的“人性化”趋势, 它们对人-机器人-数字人交互中信任的形成也具有一定的指导意义^[61]. 然而, 人-机器人-数字人交互中的信任形成机制与人际交互中的信任机制并不完全相同, 仍需进一步研究和理解.

信任在人-机器人-数字人交互中的多重作用及其影响. 信任在人-机器人-数字人交互中具有多重作用, 包括提高协作效率、促进双方专注于各自擅长的任务, 以及提高决策的准确性和及时性^[62~65]. 然而, 人-机器人交互中也存在过度信任和信任受损的风险. 过度信任可能导致人类过度依赖机器人或数字人, 从而引发事故^[66]; 而信任受损则可能导致人类彻底失去信任, 进而破坏人-机器人-数字人协作^[67]. 因此, 在人-机器人-数字人之间建立适当的信任水平对于安全和有效地使用机器人和数字人至关重要^[68].

人-机器人-数字人信任的影响因素. 为了识别信任异常并确定主要影响因素, 需要检测人-机器人-数字人交互中信任的影响因素. 根据 Kim 等^[18]的研究, 这些因素可分为三大类: 与人相关的因素、与智能体(机器人或数字人)相关的因素和与环境相关的因素. 与人相关的因素包括个体能力、特性、经验、文化、价值观和性格等^[2, 69]; 与智能体相关的因素包括性能和属性等^[69]; 环境因素包括交互环境和任务特征等^[69]. 智能体性能和属性对信任发展具有最重要的影响, 而环境因素起次要作用^[69]. 因此, 在信任的发展和维护方面, 不仅需要考虑人和环境的因素, 还需重视与智能体相关的因素, 尤其是智能体的性能和属性. 有效地协调“人-智-环”三者之间的相互作用, 有助于建立人-机器人-数字人之间的相互理解和信任, 并不断提高个人和团队的能力^[22]. 总之, 影响人-机器人-数字人交互的因素复杂且相互关联, 相关领域的研究者、设计师和开发者需要仔细考虑智能体的设计、

人和交互环境,以确保有效可靠的人-机器人-数字人交互。同时,利用增强现实技术结合多模态线索,可以提供更全面、更直观的信息,有望有效解决异常信任水平的问题。

利用多模态信号减少对机器人的过度信任。过度信任是指人类对机器人或数字人的能力或意图有过分乐观或不切实际的想法^[68],导致人-机器人-数字人协作中的懈怠或缺乏行事谨慎^[70]。这种信任偏差会增加误差或事故等风险^[71]。因此,智能体(机器人和数字人)设计者应提高对过度信任的认识,并采取相应措施来减轻其影响。例如,向用户提供关于智能体能力和限制的明确信息^[71],以及设计透明和值得信赖的系统。单模态线索蕴含的外界信息过于片面,会加剧过度信任形成不良后果。在自动驾驶领域,人类对自动驾驶汽车技术能力的过度信任导致了特斯拉和优步的事故^[72]。这些事故中,人类驾驶员过分依赖自动驾驶系统,没有对其进行足够的监督。然而,如果采用多模态线索,即使使用多种线索来强化信息传达的可靠性,通常比单模态线索更有效,更可信^[73]。Cao等^[74]研究了多模态社交线索在人-机器人-数字人交互中的作用,他们发现手势、光线和/或声音模态的组合可以提高人类对智能体社交意图和情感状态的理解和感知。他们的研究还表明,智能体的多模态交流可以通过模仿人类的交互模式来建立社会关系,从而增强智能体的社会影响力,提高智能体的帮助性,并减少任何误判或过度依赖的风险。这与人际交流中的多模态线索类似,一个有效的社交交流通常需要同时使用语言和非语言线索,如说话、手势、面部表情和目光接触等。在目标定位场景中,触觉线索^[75]与视听觉线索的组合优于单独使用的任何一种模式^[73]。这是因为多种模式比单一模式提供更多的信息,使决策的准确性更高^[76]。这些研究都基于一个观念,即用户期望从智能体那里获得与人类相似的交流线索。因此,通过利用多模态线索,研究人员模仿人际交互,改善交互体验,平衡用户信任。此外,多模态学习比只使用其子集的模态实现了更小的群体风险^[76]。这表明,在准确性和可信性方面,多模式学习比单模式学习更有效。多模态线索唯一的缺点就是信息呈现方式受限于2D平面。而增强现实技术能够以更直观、自然的方式将多模态线索可视化至3D空间,并向人类传达“过度信任”的警告以及呈现机器人的状态、意图和行为^[77],从而减少过度信任。

通过多模态信号修复对机器人的信任受损。机器人信任违背(trust violations)在人-机器人-数字人交互中是不可避免的现象,这可能导致人类同伴的信任受损^[78]。信任损害是指人类由于负面经历或事件而对智能体信任的降低或破坏^[79]。在人-机器人-数字人交互过程中,信任损害可能对人类与智能体之间的沟通和协作产生不利影响,降低任务执行效率^[67],甚至导致不信任。相较于信任行为,不信任行为会使参与者在一段时间内对可信度、信任意向和信任行为评价较低。信任修复是在一方破坏另一方信任后重建信任的过程,涉及人类与智能体之间的互动。视觉导航领域(vision-and-language navigation, VLN)和具身指称理解(ERU)领域的研究表明,信任损害在人-机器人-数字人交互中可能导致负面结果^[43,44]。单模态信任修复策略,如仅使用文本或语音来道歉或解释^[80],可能缺乏真诚和温暖的感觉,从而削弱了信任修复的效果^[79]。此外,现有的单模态信任修复研究多数采用视频观察法,而非真实场景下的人-机器人-数字人交互,因此其真实性和生态效度有待提高^[81,82]。相比之下,多模态信任修复策略,如同时使用文本、面部表情和手势来道歉或解释,可以克服单模态信任修复策略在真诚度和认知负荷方面的不足。多模态信任修复策略可以利用不同的交流形式来传达智能体(机器人和数字人)的信任修复意图和情感态度,从而增强智能体的温暖或诚挚感^[83]。同时,多模态信任修复策略可以减轻用户的认知负荷,因为用户可以根据自己的偏好选择最合适的信息来源^[84]。这与人际关系中的信任修复类似,一个真诚和温暖的信任修复通常需要同时使用语言和非语言线索,如说“对不起”并配以微笑或拥抱。因此,考虑采用基于多模态线索的信任修复方法来整合各种线索。为解决部分智能体不具备表情和肢体动作等意图表达能力的问题,可利用增强现实等沉浸式可视化技术,以更直观、自然地传达智能体意图和行为,帮助重建人类对智能体的信任^[85]。

人-机器人-数字人交互中的信任评估方法. 信任评估在人-机器人-数字人交互中至关重要,因为它有助于研究人员和设计师了解信任的形成和维护机制,并找到改善信任建立干预的方法. 评估人-机器人-数字人交互中的信任可采用自我报告测量、行为测量和生理测量等工具和方法^[86]. 自我报告测量是一种主观测量方法,用于评估人类对机器人或数字人的信任水平,以及了解人类对与机器人和数字人交互的看法. 自我报告测量是评估人与机器人或数字人交互中信任度的常用方法,包括调查、问卷和访谈等方式,通过收集人类被试关于他们对机器人或数字人信任程度的相关数据. 自我报告测量方法对于评估人类对机器人或数字人的总体信任水平具有重要价值,并可帮助确定影响人-机器人-数字人交互中信任的因素. 客观信任测量方法分为基于行为的和基于生理状态的两类^[72]. 前者观察或测量参与者与机器人或数字人互动时的行为反应,后者监测参与者的生理指标.

在人-机器人-数字人交互场景下,人-机器人-数字人信任是自主化技术发展中的一个重要且复杂问题. 为提高人-机器人-数字人信任,需构建有效且适应性强的机器人或数字人认知模型.

2.3 虚实混融式认知线索共享

传统的人-机器人-数字人交互与人际交互的核心差异在于交流方式. 人类依赖于语言和手势进行沟通,而机器人或数字人则采用文本命令等数字形式,这可能导致误差和误解. 增强现实技术能够将机器人或数字人的意图、情感和行为等信息以更直观的虚拟知识形态呈现在现实空间中,实现虚实深度融合,消除人机沟通障碍,并提供高带宽和低模糊度的替代交互机制. 增强现实是一种将计算机生成的内容与现实世界相结合的沉浸式交互技术^[87]. 通过视觉、听觉或其他感官信息,它能够增强人们对周围环境的感知. 增强现实在物理世界中创建虚拟模型,实时地为用户提供各类信息. 相较于虚拟现实,增强现实在“半虚半实”的环境中创造了低模糊度和高交互自然性. 在机器人领域,增强现实为人类与机器人或数字人等自主系统之间的交互和信息交流提供了新的媒介^[5,6,88]. 通过添加计算机生成的视觉元素、声音和其他刺激,增强现实提升了人类对机器人或数字人以及现实世界环境的感知. 此外,在高危作业场景(如高放废物深地质处置)中,增强现实被用于提升远程人机操作技术;同时,它还能提高风险环境下劳动者的生产力和安全性.

在非共享增强现实空间中,机器人和数字人的多模态线索存在局限性,如可展示文本内容受限、语音线索不适用于紧急场景、部分机器人和数字人不具备肢体动作和表情. 文本交流方面,由于机器人和数字人自带的显示屏尺寸有限,呈现的文本或可视化信息受限. 语音交流方面,人类需要在接收到完整的机器人和数字人表达的情况下才能理解其意,并存在遗漏重要信息的风险,因此语音交流不适用于紧急场景. 肢体动作和面部表情方面,非人形机器人和数字人难以通过肢体动作或表情传达意图或行动. 增强现实为人-机器人-数字人交互创造了更自然、更沉浸式的交互空间,并广泛应用于人-机器人-数字人协作,丰富了人类与自主系统之间的信息交流^[89]. 增强现实技术能突破机器人和数字人多模态线索的局限,结合多模态信息实现人-机器人-数字人互信. 沉浸式人-机器人-数字人交互中的共享认知线索是指:利用增强现实技术对各类多模态线索进行增强;通过沉浸式技术来增强协作方(人-机器人-数字人)的状态、意图和计划的视觉表达方式;这些表达共享在虚实融合的交互空间,可改善非共享增强现实空间中的协作效率^[88],如图4所示. 图4展示了基于“能力互补+知识共享+意图对齐”的人-机器人-数字人互信框架. 该框架旨在通过评估人-机器人-数字人之间的信任水平,动态调整机器人和数字人的决策权限,以实现人-机器人-数字人之间的有效协作. 该框架包括3个主要部分:人-机器人-数字人意图对齐^[19]、人-机器人-数字人交互可靠性保障和信任评估. 人-机器人-数字人意图对齐部分负责确保人-机器人-数字人对整体任务目标和当前状态的理解一致. 这可以通过增强现实技术来实现,即通过可视化机器人和数字人的内在状态和意

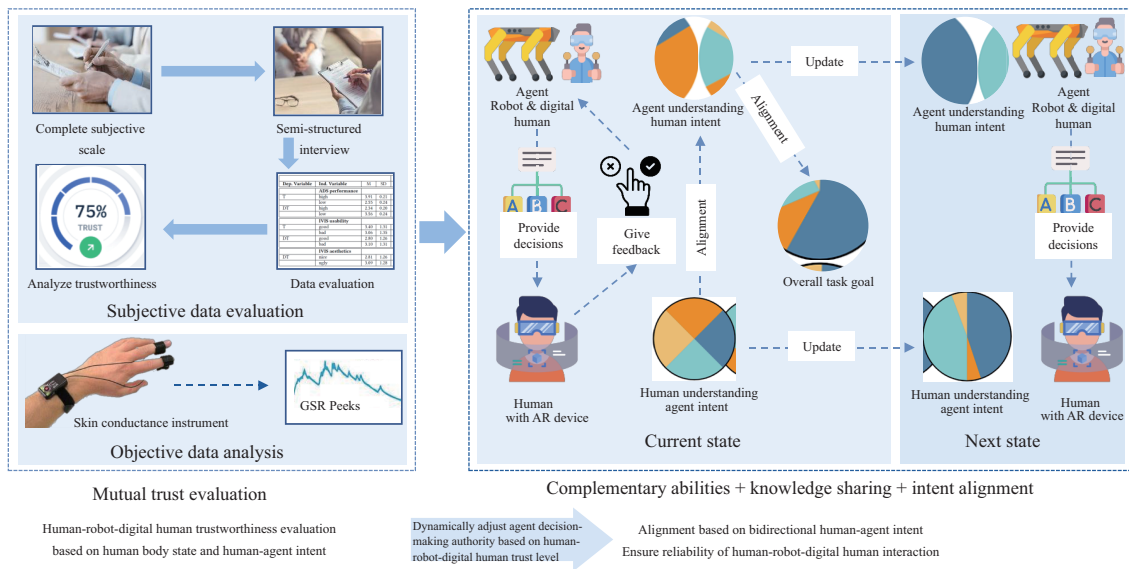


图 4 (网络版彩图) 基于“能力互补 + 知识共享 + 意图对齐”的人 - 机器人 - 数字人互信框架

Figure 4 (Color online) Human-robot-digital human trust framework based on “complementary abilities + knowledge sharing + intent alignment”

图, 帮助人类更好地理解机器人和数字人的行为. 同时, 机器人和数字人也可以通过感知和分析人类的多模态线索来理解人类的意图和需求, 并作出适应性响应. 人 - 机器人 - 数字人交互可靠性保障部分负责确保人 - 机器人 - 数字人交互过程中信息传递的准确性和及时性. 这可以通过主观数据评估和客观数据分析来实现, 即通过主观量表、半结构化访谈等方式收集人类对交互过程的主观评价, 并通过皮肤电反应峰值、皮肤电导仪等仪器测量人类的生理反应, 从而分析交互过程中的信任水平. 人 - 机器人 - 数字人信任评估部分负责根据人体状态和人 - 机器人 - 数字人意图对齐情况来评估人 - 机器人 - 数字人之间的信任水平. 根据评估结果, 可以动态调整机器人的决策权限, 以实现更有效的协作. 例如, 当评估结果显示人类对机器人和数字人的信任水平较高时, 可以提高机器人的决策权限; 反之, 则降低机器人和数字人的决策权限. 总之, 该框架通过能力互补、知识共享和意图对齐 3 个方面来建立和维护人 - 机器人 - 数字人之间的互信关系, 从而实现更有效、更安全、更可靠的协作.

通过提供更沉浸的交互体验, 使机器人和数字人能够更清晰地传达其状态、行为和意图, 从而提高人类对机器人和数字人的信任. 增强现实设备能够在用户的真实世界中叠加虚拟信息, 如 3D 图像、动画、符号或图标, 这些设备包括抬头显示器、头戴式混合现实设备、全息显示器、智能眼镜和手持式移动设备等^[88]. 为了实现更有效的双向交互, 增强现实设备可用于显示机器人和数字人的情绪、意图、能力或限制, 以及状态信息和交互界面, 使得人 - 机器人 - 数字人类能够更好地理解和响应彼此的交流^[19]. 为了实现这一目标, 相关研究者应设计清晰直观的增强现实界面元素, 并定制化信息呈现方式. 同时, 还需控制增强现实可视化信息的呈现量, 以避免过度干扰用户的注意力. Chandan 等^[87]利用模仿学习方法实现了增强现实的智能化, 并开发出“模仿学习式 AR 可视化”的技术框架. 如图 4 所示, 增强现实与多模态线索的结合可以互相补足, 有助于人 - 机器人 - 数字人互信交互的发展. 一方面, 多模态线索为增强现实提供了更多的信息通道, 提高了可视化信息的可靠性. 相较于单模态交互, 多模态融合交互允许用户使用多种输入模式与机器人和数字人交互, 提供更多元、自然和直观的交互体验. 然而, 单模态界面在解释用户的复杂命令或意图方面能力有限, 因为它依赖于单个输入模

态. 这使机器人和数字人难以准确有效地响应人类的需求. 融合多种输入模式可以促使每种输入模式的相互补偿, 这使得人与机器人和数字人更容易理解彼此通过各种输入通道提供的完整信息, 从而构建更自然的人-机器人-数字人交互生态. 另一方面, 增强现实将多源多模态线索信息共享于人-机器人-数字人之间, 有利于人-智能体之间意图、情感和行为的对齐^[19]. Qiu 等^[90] 设计和研发了一种增强现实共享空间, 在此空间中, 机器人和数字人不仅能感知自己视野中的虚拟信息, 还能通过感知人类的凝视和姿势来推断人类的意图, 并建立双向信息交互. 通过增强现实技术, 人-机器人-数字人交互可以实现虚实深度融合, 提供高带宽和低模糊度的共享认知线索. 然而, 在人-机器人-数字人交互中, 并非所有信息都能够直接呈现在真实空间中. 有些信息需要跨越多种媒体形式, 在虚拟和现实之间进行转换和融合. 为了解决这一问题, 本文提出了“三位一体”式跨媒体虚实混融交互构想.

3 “三位一体”式跨媒体虚实混融交互构想

前文介绍了多模可信交互的概念、类型和框架, 以及多模态线索和增强现实技术在建立信任方面的作用. 然而, 随着人-机器人-数字人交互技术的不断发展, 人类不仅需要与机器人和数字人进行自然、有效、可信的交互, 还需要将自己的意图、情感、知识等投射到虚拟空间中, 与其他智能体进行混融式交互. 这就要求人-机器人-数字人交互系统能够实现跨媒体虚实混融智能交互, 即在真实世界和虚拟世界之间建立无缝连接, 让人类可以自由地在两个世界中切换和协作. 为了进一步发展跨媒体虚实混融智能交互, 我们提出了一种“三位一体”式复杂交互与协同系统, 即将自然人、数字人和机器人融合为一个整体^[9,91], 通过数字分身作为人类的交互载体, 在真实世界和虚拟世界之间进行高效、自然、可信的沟通和协作. 本节将介绍“三位一体”式复杂交互与协同系统的新框架、关键技术和应用前景. 专注于混合现实研究的 Meta 公司^[91] 提出了一种基于物理引擎和神经网络的人形模拟系统, 该系统能够将机器人和由人操作的数字人之间的社会化协作通过高保真的 3D 渲染呈现在室内家庭环境中, 并通过鼠标/键盘或 VR 等设备进行交互. 该系统在社会导航 (机器人非接触式跟随数字人) 和社会重排协作 (数字人和机器人协作移动物品) 任务上进行了评估, 结果显示该系统能够显著提高协作效率和用户满意度. 该系统是最接近“三位一体”式跨媒体虚实融合交互构想的案例, 但它仍缺少虚实融合、认知线索共享、多模可信交互等核心思想.

为了实现跨虚实世界的沉浸式多模态互信人-机器人-数字人交互, 我们提出了一种“三位一体”式跨媒体虚实混融交互构想. 该构想基于以下两个假设: (1) 自然人、数字人和机器人之间存在着共享的认知和情感基础, 可以通过多模态信号进行有效的信息交流和情感表达. 在这种情况下, 建立可信的交互关系至关重要. 通过可信的交互关系, 自然人、数字人和机器人可以更好地理解彼此的意图和需求, 从而实现更有效、更安全、更可靠的协作; (2) 自然人、数字人和机器人之间存在着共享的价值和目标, 可以通过协作和竞争来实现共同的利益和成就. 在这种情况下, 建立可信的协作关系至关重要. 通过可信的协作关系, 自然人、数字人和机器人可以更好地分享知识和资源, 从而实现更高效、更创新、更有价值的成果. 基于这两个假设, 我们构建了一个由自然人、数字人和机器人组成的三元交互模型, 旨在通过增强现实技术将真实空间和虚拟空间进行无缝融合, 从而实现更加自然、更加有效、更加可靠的沉浸式多模态互信人-机器人-数字人交互. 该模型包括 3 个主要部分: 真实空间、虚拟空间和内部关系. 在内部关系部分, 我们设计了一套人-机器人-数字人信任评估机制, 用于动态评估人-机器人-数字人之间的信任水平, 并根据评估结果调整机器人的决策权限. 该机制包括主观评估和客观评估两个部分. 主观评估部分通过问卷调查、访谈等方式收集用户对交互过程中信任感受的

主观反馈; 客观评估部分通过生理信号检测、行为分析等方式收集用户在交互过程中的客观反映. 通过综合分析主观评估和客观评估的结果, 我们可以动态调整机器人和数字人的决策权限, 以实现更有效、更安全、更可靠的协作.

3.1 三位一体智能交互新框架

沉浸式数字世界建模与交互. 当前, 沉浸式自然人-机器人-数字人交互尚未能有效地融合跨模态信号, 导致人与虚拟环境之间的边界难以模糊化. 为解决该问题, 研究者基于跨模态关联协同表征, 探讨虚拟世界三维模型的拟人化构建、虚实融合信息传递技术以及虚拟环境增强模型, 旨在构建流畅自然的沉浸式人-机器人-数字人交互环境. 此外, 研究者还需探索数字世界中虚拟对象智能感知与行为的数学表达与建模方法, 以解决虚拟对象与虚拟环境和用户之间进行自然、流畅、沉浸交互的问题. 因此, 亟需发展数字世界智能对象建模的技术与方法体系, 以实现具备沉浸式、智能化、跨媒体等特点的人-机器人-数字人交互^[9].

跨媒体人机认知交互是实现自然交互和自主协同的关键, 然而, 其目前仍存在局限性, 如仅限于感知驱动的真实交互, 无法满足认知意图引导的交互以及多任务多场景的需求. 为克服这些局限, 研究者可在现有跨媒体感知式人-机器人-数字人交互框架的基础上, 探索多模态交互测量理论的双向价值建模与对齐机制, 从而建立起认知驱动的可信、可靠型人-机器人-数字人交互. 同时, 研究者还可以研究跨场景跨任务条件下的人机智能共生的行为增强与脑机协同机制, 从而形成“人与智能体在回路”的混合增强智能交互与协同范式. 最后, 针对动态真实世界环境下的多维度融合情境理解及人-智能体群组协同, 可研究面向人-智能体协同的共享记忆与知识演化方法, 实现更加高效的人-智能体群组协同^[57].

共享式跨媒体虚实混融交互正逐渐成为新型人-机器人-数字人交互的主要趋势. 然而, 这一趋势也带来了更加复杂的交互关系和更多的交互对象, 因此亟需发展进入跨媒体虚实混融智能交互的高级阶段. 为了进一步发展跨媒体虚实混融智能交互, 探索不同实体之间的信任机制, 我们提出了一种“三位一体”式复杂交互与协同系统, 即将自然人、数字人和机器人融合为一个整体, 通过数字分身作为人类的交互载体, 在真实世界和虚拟世界之间进行高效、自然、可信的沟通和协作. 这种新型人-机器人-数字人交互模式不仅包括自然人和机器人之间的多模可信交互, 还包括自然人和数字人之间、数字人和机器人之间以及三者之间的混融式交互^[9].

图5展示了一种“三位一体”式跨媒体虚实混融交互构想. 该构想旨在通过构建一个由自然人、数字人和机器人组成的三元交互模型, 实现跨虚实世界的沉浸式多模态互信人-机器人-数字人交互. “三位一体”中存在“互信问题”, 因为建立和维护信任并非易事, 人对智能体(机器人和数字人)的信任^[18]和智能体对人^[20,21]的信任是一个复杂且多维的结构, 受多种因素影响. 该模型包括3个主要部分: 真实空间、虚拟空间和内部关系. 真实空间部分包括自然人和高仿真人形机器人. 自然人是指具有真实身份的人类, 他们可以通过身体动作、语言和面部表情等多种方式与机器人和数字人进行交互. 高仿真人形机器人是指外观和行为类似于人类的机器人, 它们可以通过模拟人类的面部表情和身体动作来与自然人进行交互. 虚拟空间部分包括虚拟数字人. 虚拟数字人是指在虚拟世界中存在的数字化角色, 它们可以通过计算机生成的图像和声音来与自然人或高仿真人形机器人进行交互. 不同实体之间的信任机制可能会有所不同, 具体差异取决于这些实体在认知能力、交互方式和决策过程等方面的具体表现. 内部关系部分负责描述自然人、数字人和机器人之间的相互关系. 这些关系包括IP合作、会议、赞赏与合作、虚实营销、劳动分配与协作、教学、工作、数据交换/行为交互和归属感等^[9]. 这些关系可以通过沉浸式多模态互信人-机器人-数字人交互来实现, 即通过提供沉浸式的视

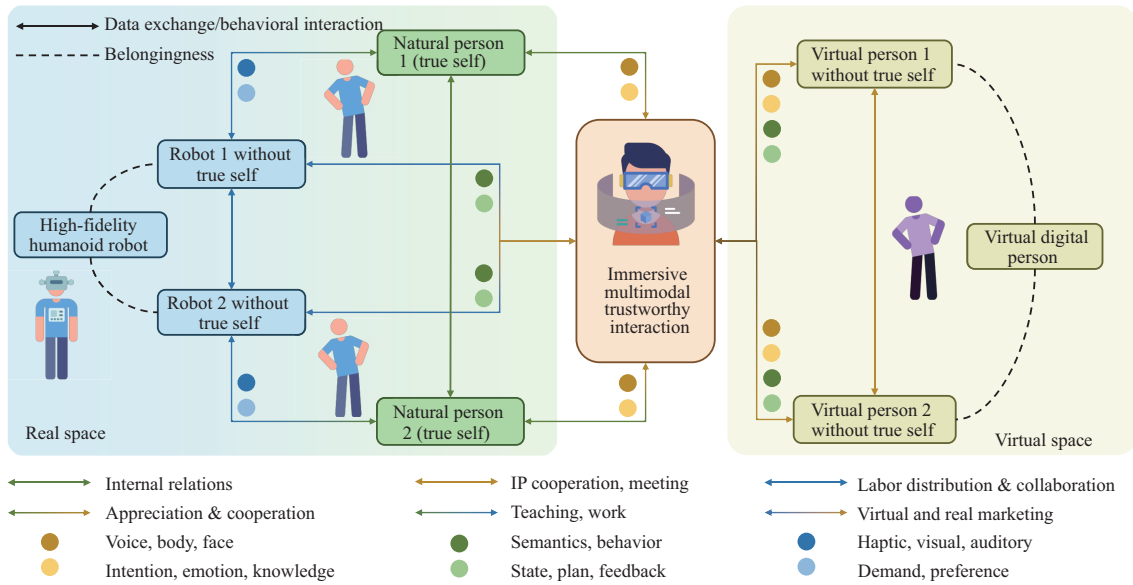


图 5 “三位一体”式跨媒体虚实混融交互构想

Figure 5 Immersive cross-modal “human-robot-digital human” triune interaction model

觉、听觉和触觉反馈来增强人-机器人-数字人之间的信任感。在“三位一体”中多模态技术可以作用于感知、认知和行动 3 个部分, 帮助人-机器人-数字人更好地相互理解和互相适应, 提高机器人和数字人的智能水平和决策能力。总之, 该构想提供了一种新颖的跨媒体虚实混融交互模型, 旨在通过构建一个由自然人、数字人和机器人组成的三元交互系统, 实现更加自然、更加有效、更加可靠的沉浸式多模态可信人-机器人-数字人交互。

3.2 三位一体智能社会组织模式

三位一体智能社会组织模式涉及自然人、虚拟数字人和高仿人机器人 3 种主体之间的交互模式。基于真身-化身-假身“三位一体”与自然人-高仿人机器人-虚拟数字人“三人行”的模式, 将被用于探索虚实相融的新型人-机器人-数字人交互方式。为实现这种模式, 需依赖先进技术手段, 如人工智能、区块链和 Web3.0 等。该模式有助于构建新的社会组织模式及社会事件态势感知框架。通过真身-化身-假身“三位一体”, 用户可在不同场景和角色中自由切换, 体验多样时空模态和交互模式, 增强沉浸感和代入感。通过自然人-高仿人机器人-虚拟数字人“三位一体”组织架构, 用户可与不同类型智能体建立新的社会关系网络, 分享虚拟数字人作为真实分身的人的认知框架和惯习, 跨越文化背景和沟通障碍。结合扩展现实 (VR, AR, MR)^[4] 技术和 5G/6G 低延速率, 用户可在虚实场景之间自然地跳转和过渡, 构建跨虚实世界的“三位一体”经历交融共享生态。借助高仿人机器人和虚拟数字人的外形、语音、表情、动作等特征, 用户可在现实世界中与机器人和数字人进行更真实和复杂的互动, 改变现有的连接方式和社会组织形态。利用虚拟数字人作为信息传递和事件感知的载体, 用户可获取更全面和深入的社会事件态势感知框架, 提升社会参与度和责任感。

在应用方面, 这种模式可广泛应用于各个领域, 如教育、医疗、金融等。在教育领域, 高仿人机器人可作为辅助教学工具, 与自然人进行交互, 提升教学效果; 在医疗领域, 虚拟数字人可作为医生的助手, 帮助医生进行诊断和治疗; 在金融领域, 高仿人机器人可作为客户服务代表, 与客户进行交互并提供服务。总之, “三位一体”智能社会组织模式具有广泛的应用前景, 并将对未来社会发展产生深远

影响.

4 总结与展望

本研究以人-机器人交互的发展趋势为出发点,深入探讨了人工智能技术在人-机器人-数字人交互领域的应用及挑战.本文首先提出了共享式多模可信人-机器人-数字人交互的概念与框架,随后详细论述了“三位一体”式跨媒体虚实混融交互模式的理论与实践.本文强调,共享式多模可信人-机器人-数字人交互在提高人-机器人-数字人协作效率与质量、增强人-机器人-数字人之间的信任与理解以及促进人-机器人-数字人一体化智能社会组织模式形成方面具有显著优势.然而,本文也指出了当前研究中存在的一些问题与不足之处,例如,增强现实技术在提供虚拟信息时可能会干扰真实环境的感知与认知,“三位一体”式跨媒体虚实混融交互模式需考虑不同媒体间的协调与转换,多模态融合计算框架须更精确地捕捉与分析人类的情感、意图、反馈等信息,以及机器人和数字人的认知模型需更好地适应复杂多变的环境与任务.

针对上述问题,未来研究可从以下几个方面展开:一是基于增强现实的人-机器人-数字人互信交互机制的设计与评估,探讨如何平衡虚拟信息与真实环境之间的关系,以提高用户体验与满意度;二是“三位一体”式跨媒体虚实混融交互模式的理论与实践,研究如何利用不同媒体间的优势与特点,构建更丰富、更灵活、更自然的交互方式;三是基于多模态融合计算框架的人-机器人-数字人协作心理模型的构建与验证,研究如何利用多种传感器数据进行多层次、多维度、多精度的信息处理与表达,以提高信息传递与理解效率;四是基于人工智能技术的机器人和数字人认知模型的优化与改进,研究如何赋予机器人更为强大、灵活、自主、适应性强以及高度可学习的能力.相关工作详见Github¹⁾.

参考文献

- 1 Fan J J, Tian F, Du Y, et al. Thoughts on human-computer interaction in the age of artificial intelligence. *Sci Sin Inform*, 2018, 48: 361-375 [范俊君, 田丰, 杜一, 等. 智能时代人机交互的一些思考. *中国科学: 信息科学*, 2018, 48: 361-375]
- 2 Hopko S, Wang J K, Mehta R. Human factors considerations and metrics in shared space human-robot collaboration: a systematic review. *Front Robot AI*, 2022, 9: 6
- 3 Drury J L, Scholtz J, Yanco H A. Applying CSCW and HCI techniques to human-robot interaction. In: *Proceedings of the CHI Workshop on Shaping Human-Robot Interaction*, Vienna, 2004. 13-16
- 4 Suzuki R, Karim A, Xia T, et al. Augmented reality and robotics: a survey and taxonomy for AR-enhanced human-robot interaction and robotic interfaces. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, New Orleans LA, 2022. 1-33
- 5 Williams T, Hirshfield L, Tran N. et al. Using augmented reality to better study human-robot interaction. In: *Proceedings of the HCI International Conference*, Copenhagen, 2020. 643-654
- 6 Chacko S M, Kapila V. An augmented reality interface for human-robot interaction in unconstrained environments. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Macao, 2019. 3222-3228
- 7 Bera A, Randhavane T, Prinja R, et al. SocioSense: robot navigation amongst pedestrians with social and psychological constraints. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vancouver, 2017. 7018-7025
- 8 Malik A A, Brem A. Digital twins for collaborative robots: a case study in human-robot interaction. *Robot Comput Integr Manuf*, 2021, 68: 102092

1) <https://github.com/MIG-UESTC-HCI/Multimodal-Human-Robot-Digital-Human-Trust-Interaction>.

- 9 Peng Y T, Gao S, You K K, et al. An analysis of human-computer interactive patterns in the metaverse. *J Xi'an Jiaotong Univ Soc Sci*, 2023, 43: 176–184 [彭影彤, 高爽, 尤可可, 等. 元宇宙人机融合形态与交互模型分析. *西安交通大学学报 (社会科学版)*, 2023, 43: 176–184]
- 10 Sharkawy A N. Human-robot interaction: applications. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, Xi'an, 2021. 98–103
- 11 Raybourn E M, Stubblefield W A, Trumbo M, et al. Information design for XR immersive environments: challenges and opportunities. In: *Proceedings of the HCI International Conference*, Orlando, 2019. 153–164
- 12 Hertel J, Karaosmanoglu S, Schmidt S, et al. A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review. In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, Bari, 2021. 431–440
- 13 Scholtz J. Theory and evaluation of human robot interactions. In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, Hawaii, 2003. 10
- 14 Bonarini A. Communication in human-robot interaction. *Curr Robot Rep*, 2020, 1: 279–285
- 15 Bell G. Making life: a brief history of human-robot interaction. *Consumption Markets Culture*, 2018, 21: 22–41
- 16 Wang D X, Zheng Y L, Li T, et al. Multi-modal human-machine interaction for human intelligence augmentation. *Sci Sin Inform*, 2018, 48: 449–465 [王党校, 郑一磊, 李腾, 等. 面向人类智能增强的多模态人机交互. *中国科学: 信息科学*, 2018, 48: 449–465]
- 17 Mayer R C, Davis J H, Schoorman F D. An integrative model of organizational trust. *Acad Manage Rev*, 1995, 20: 709–734
- 18 Kim W, Kim N, Lyons J B, et al. Factors affecting trust in high-vulnerability human-robot interaction contexts: a structural equation modelling approach. *Appl Ergonomics*, 2020, 85: 103056
- 19 Yuan L Y, Gao X F, Zheng Z L, et al. In situ bidirectional human-robot value alignment. *Sci Robot*, 2022, 7: 4183
- 20 Vinanzi S, Patacchiola M, Chella A, et al. Would a robot trust you? Developmental robotics model of trust and theory of mind. *Phil Trans R Soc B*, 2019, 374: 20180032
- 21 Kirtay M, Oztop E, Asada M, et al. Modeling robot trust based on emergent emotion in an interactive task. In: *Proceedings of the IEEE International Conference on Development and Learning*, Beijing, 2021. 1–8
- 22 Zhang F B, Yu J, Lin D F, et al. UnIC: towards unmanned intelligent cluster and its integration into society. *Engineering*, 2022, 12: 24–38 [张福彪, 余婧, 林德福, 等. 无人智群及其社会融合. *工程*, 2022, 12: 24–38]
- 23 Khavas Z R, Ahmadzadeh S R, Robinette P. Modeling trust in human-robot interaction: a survey. In: *Proceedings of the International Conference on Social Robotics*, Golden, 2020. 529–541
- 24 Torta E, van Heumen J V, Piunti F, et al. Evaluation of unimodal and multimodal communication cues for attracting attention in human-robot interaction. *Int J Soc Robotics*, 2015, 7: 89–96
- 25 Lazaro M J, Kim S, Lee J, et al. A review of multimodal interaction in intelligent systems. In: *Proceedings of the HCI International Conference*, Washington DC, 2021. 206–219
- 26 Budkov V Y, Prischepa M, Ronzhin A, et al. Multimodal human-robot interaction. In: *Proceedings of the International Congress on Ultra Modern Telecommunications and Control Systems*, Moscow, 2010. 485–488
- 27 Oliveira R, Arriaga P, Paiva A. Human-robot interaction in groups: methodological and research practices. *Multimodal Technol Interact*, 2021, 5: 59
- 28 Li Y, Chen X X, Zhao H, et al. Understanding embodied reference with touch-line transformer. In: *Proceedings of the International Conference on Learning Representations*, Washington DC, 2023. 1–12
- 29 Islam M M, Mirzaiee R, Gladstone A, et al. CAESAR: an embodied simulator for generating multimodal referring expression datasets. In: *Proceedings of Advances in Neural Information Processing Systems*, 2022. 35: 21001–21015
- 30 Lu Z Y, Pei Y Q, Wang G Q, et al. ScanERU: interactive 3D visual grounding based on embodied reference understanding. 2023. ArXiv:2303.13186
- 31 Salem M, Rohlfing K, Kopp S, et al. A friendly gesture: investigating the effect of multimodal robot behavior in human-robot interaction. In: *Proceedings of the IEEE International Symposium on Robot and Human Interaction Communication, Ethics, and Society*, Atlanta, 2011. 247–252
- 32 Papanastasiou S, Kousi N, Karagiannis P, et al. Towards seamless human robot collaboration: integrating multimodal interaction. *Int J Adv Manuf Technol*, 2019, 105: 3881–3897
- 33 Luebbbers M B, Tabrez A, Hayes B. Augmented reality-based explainable AI strategies for establishing appropriate

- reliance and trust in human-robot teaming. In: Proceedings of the International Workshop on Virtual, Augmented, and Mixed Reality for HRI, Hokkaido, 2022. 1–5
- 34 Chen X T, Li S. Survey on emotion recognition in conversation. *Comput Eng Appl*, 2023, 59: 33–48 [陈晓婷, 李实. 对话情绪识别综述. *计算机工程与应用*, 2023, 59: 33–48]
- 35 Mohammed S N, Hassan A K A. A survey on emotion recognition for human robot interaction. *J Comput Inf Technol*, 2020, 28: 125–146
- 36 Axelsson A, Skantze G. Multimodal user feedback during adaptive robot-human presentations. *Front Comput Sci*, 2022, 3: 135
- 37 Kirtay M, Oztop E, Kuhlen A K, et al. Forming robot trust in heterogeneous agents during a multimodal interactive game. In: Proceedings of the IEEE International Conference on Development and Learning, London, 2022. 307–313
- 38 Wudarczyk O A, Kirtay M, Pischetta D, et al. Robots facilitate human language production. *Sci Rep*, 2021, 11: 16737
- 39 Marge M, Espy-Wilson C, Ward N G, et al. Spoken language interaction with robots: recommendations for future research. *Comput Speech Language*, 2022, 71: 101255
- 40 Rázuri J G, Sundgren D, Rahmani R, et al. Speech emotion recognition in emotional feedback for human-robot interaction. *Int J Adv Res Artif Intell*, 2015, 4: 20–27
- 41 Inamura T, Mizuchi Y. SIGVerse: a cloud-based VR platform for research on multimodal human-robot interaction. *Front Robot AI*, 2021, 8: 549360
- 42 Liu H, Wang L. Gesture recognition for human-robot collaboration: a review. *Int J Industrial Ergonomics*, 2018, 68: 355–367
- 43 Chen Y, Li Q, Kong D, et al. YouReflT: embodied reference understanding with language and gesture. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, 2021. 1385–1395
- 44 Jiang K, Stacy S, Dahmani A L, et al. What is the point? A theory of mind model of relevance. In: Proceedings of the Annual Meeting of the Cognitive Science Society, Seattle, 2022. 3669–3675
- 45 Zheng Z L, Qiu S W, Fan L F, et al. GRICE: a grammar-based dataset for recovering implicature and conversational reasoning. In: Proceedings of the Findings of the Association for Computational Linguistics, New York, 2021. 2074–2085
- 46 Peters C, Li C J, Yang F, et al. Investigating social distances between humans, virtual humans and virtual robots in mixed reality. In: Proceedings of the International Conference on Autonomous Agents and MultiAgent Systems, Stockholm, 2018. 2247–2249
- 47 Zhang X L, Lyu F, Cheng S W. Interaction paradigm in intelligent systems. *Sci Sin Inform*, 2018, 48: 406–418 [张小龙, 吕菲, 程时伟. 智能时代的人机交互范式. *中国科学: 信息科学*, 2018, 48: 406–418]
- 48 Spezialetti M, Placidi G, Rossi S. Emotion recognition for human-robot interaction: recent advances and future perspectives. *Front Robot AI*, 2020, 7: 145
- 49 Galarza E E, Egas F D, Silva F M, et al. Real time driver drowsiness detection based on driver's face image behavior using a system of human computer interaction implemented in a smartphone. In: Proceedings of the International Conference on Information Technology & Systems, Libertad, 2018. 563–572
- 50 Yun H, Yang J H. Multimodal warning design for take-over request in conditionally automated driving. *Eur Transp Res Rev*, 2020, 12: 1–11
- 51 Maniscalco U, Storniolo P, Messina A. Bidirectional multi-modal signs of checking human-robot engagement and interaction. *Int J Soc Robotics*, 2022, 14: 1295–1309
- 52 Liu Y, Wang Y M, Bian Y L, et al. A psychological model of human-computer cooperation for the era of artificial intelligence. *Sci Sin Inform*, 2018, 48: 376–389 [刘焯, 汪亚珉, 卞玉龙, 等. 面向智能时代的人机合作心理模型. *中国科学: 信息科学*, 2018, 48: 376–389]
- 53 Fan L F, Xu M J, Cao Z H, et al. Artificial social intelligence: a comparative and holistic view. *CAAI Artif Intell Res*, 2022, 1: 144–160
- 54 Yang M H, Tao J H. Intelligence methods of multi-modal information fusion in human-computer interaction. *Sci Sin Inform*, 2018, 48: 433–448 [杨明浩, 陶建华. 多通道人机交互信息融合的智能方法. *中国科学: 信息科学*, 2018, 48: 433–448]
- 55 Boulahia S Y, Amamra A, Madi M R, et al. Early, intermediate and late fusion strategies for robust deep learning-

- based multimodal action recognition. *Machine Vision Appl*, 2021, 32: 121
- 56 Stahlschmidt S R, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Briefings Bioinf*, 2022, 23: 569
- 57 Wang H W, Li J, Liu J G, et al. Research on human-machine integration complex social system. *Chin J Manage Sci*, 2023, 31: 1–21 [王红卫, 李珏, 刘建国, 等. 人机融合复杂社会系统研究. *中国管理科学*, 2023, 31: 1–21]
- 58 Wang F Y, Guo J B, Bu G Q, et al. Mutually trustworthy human-machine knowledge automation and hybrid augmented intelligence: mechanisms and applications of cognition, management, and control for complex systems. *Front Inform Technol Electron Eng*, 2022, 23: 1142–1157
- 59 Cropanzano R, Mitchell M S. Social exchange theory: an interdisciplinary review. *J Manage*, 2005, 31: 874–900
- 60 Atkinson C. Trust and the psychological contract. *Employee Relations*, 2007, 29: 227–246
- 61 Bankins S, Formosa P. When AI meets PC: exploring the implications of workplace social robots and a human-robot psychological contract. *Eur J Work Organizational Psychol*, 2020, 29: 215–229
- 62 Lewis M, Sycara K, Walker P. The role of trust in human-robot interaction. In: *Foundations of Trusted Autonomy*. Cham: Springer, 2018. 117: 135–159
- 63 Miller L, Kraus J, Babel F, et al. More than a feeling-interrelation of trust layers in human-robot interaction and the role of user dispositions and state anxiety. *Front Psychol*, 2021, 12: 592711
- 64 Vinanzi S, Cangelosi A, Goerick C. The collaborative mind: intention reading and trust in human-robot interaction. *iScience*, 2021, 24: 102130
- 65 Javaid M, Estivill-Castro V. Explanations from a robotic partner build trust on the robot’s decisions for collaborative human-humanoid interaction. *Robotics*, 2021, 10: 51
- 66 Wagner A R, Borenstein J, Howard A. Overtrust in the robotic age. *Commun ACM*, 2018, 61: 22–24
- 67 de Visser E J, Peeters M M M, Jung M F, et al. Towards a theory of longitudinal trust calibration in human-robot teams. *Int J Soc Robotics*, 2020, 12: 459–478
- 68 Kok B C, Soh H. Trust in robots: challenges and opportunities. *Curr Robot Rep*, 2020, 1: 297–309
- 69 Hancock P A, Billings D R, Schaefer K E, et al. A meta-analysis of factors affecting trust in human-robot interaction. *Hum Factors*, 2011, 53: 517–527
- 70 Sharkey A, Sharkey N. We need to talk about deception in social robotics! *Ethics Inf Technol*, 2021, 23: 309–316
- 71 Aroyo A M, de Bruyne J, Dheu O, et al. Overtrusting robots: setting a research agenda to mitigate overtrust in automation. *Paladyn J Behav Robotics*, 2021, 12: 423–436
- 72 Frison A, Wintersberge P, Riener A, et al. In UX we trust: investigation of aesthetics and usability of driver-vehicle interfaces and their impact on the perception of automated driving. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Glasgow, 2019. 1–13
- 73 White T L, Hancock P A. Specifying advantages of multi-modal cueing: quantifying improvements with augmented tactile information. *Appl Ergonomics*, 2020, 88: 103146
- 74 Cao H L, Scholz C, de Winter J, et al. Investigating the role of multi-modal social cues in human-robot collaboration in industrial settings. *Int J Soc Robotics*, 2023, 15: 1169–1179
- 75 Yang Y, Deng M Y, Zhou Y H, et al. The effects of multimodal takeover request on distracted drivers’ takeover performance and perception. *Hum Factors Transp*, 2022, 60: 25
- 76 Huang Y, Du C Z, Xue Z H, et al. What makes multi-modal learning better than single (provably). In: *Proceedings of Advances in Neural Information Processing Systems*, 2021. 34: 10944–10956
- 77 Delmerico J, Poranne R, Bogo F, et al. Spatial computing and intuitive interaction: bringing mixed reality and robotics together. *IEEE Robot Automat Mag*, 2022, 29: 45–57
- 78 Connor E, Lionel P R. Do you still trust me? Human-robot trust repair strategies. In: *Proceedings of the 30th IEEE International Conference on Robot and Human Interactive Communication*, Vancouver, 2021. 183–188
- 79 Sebo S S, Krishnamurthi P, Scassellati B. “I don’t believe you”: investigating the effects of robot trust violation and repair. In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, Daegu, 2019. 57–65
- 80 Esterwood C, Robert J L P. Three Strikes and you are out!: the impacts of multiple human-robot trust violations and repairs on robot trustworthiness. *Comput Hum Behav*, 2023, 142: 107658
- 81 Shahrardar S, Park C, Nojournian M, et al. Human trust measurement using an immersive virtual reality autonomous

- vehicle simulator. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, 2019. 515–520
- 82 Hald K, Weitz K, André E, et al. “An Error Occurred!” -Trust repair with virtual robot using levels of mistake explanation. In: Proceedings of the 9th International Conference on Human-Agent Interaction, 2021. 218–226
- 83 Lyons J B, Hamdan I A, Vo T Q. Explanations and trust: what happens to trust when a robot partner does something unexpected? *Comput Hum Behav*, 2023, 138: 107473
- 84 Papagni G, de Pagter J, Zafari S, et al. Artificial agents’ explainability to support trust: considerations on timing and context. *AI Soc*, 2023, 38: 947–960
- 85 Kassem K, Ungerböck T, Wintersberger P, et al. What is happening behind the wall? Towards a better understanding of a hidden robot’s intent by multimodal cues. In: Proceedings of the ACM on Human-Computer Interaction, New York, 2022. 1–19
- 86 Kohn S C, de Visser E J, Wiese E, et al. Measurement of trust in automation: a narrative review and reference guide. *Front Psychol*, 2021, 12: 604977
- 87 Chandan K D, Albertson J, Zhang S. Learning visualization policies of augmented reality for human-robot collaboration. In: Proceedings of the 6th Conference on Robot Learning, Copenhagen, 2023. 1233–1243
- 88 Makhataeva Z, Varol H A. Augmented reality for robotics: a review. *Robotics*, 2020, 9: 21–48
- 89 Piumsomboon T, Dey A, Ens B, et al. The effects of sharing awareness cues in collaborative mixed reality. *Front Robot AI*, 2019, 6: 5
- 90 Qiu S, Liu H, Zhang Z, et al. Human-robot interaction in a shared augmented reality workspace. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, 2020. 11413–11418
- 91 Xavier P, Eric U, Andrew S, et al. Habitat 3.0: a co-habitat for humans, avatars and robots. 2023. ArXiv:2310.13724

Multimodal trustworthy interaction: from multimodal information fusion to a trinitarian human-robot-digital human interaction model

Guoqing WANG^{1*}, Yunqiang PEI¹, Yang YANG¹, Xing XU¹, Zheng WANG¹ & Hengtao SHEN^{1,2}

1. *School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China;*

2. *Peng Cheng Laboratory, Shenzhen 518055, China*

* Corresponding author. E-mail: gqwang0420@uestc.edu.cn

Abstract Trust plays a key role in the intelligent collaboration among humans, robots, and digital humans, as it affects not only the efficiency of their interaction, but also the risks and benefits involved. However, the current research on human-agent trust suffers from an “imbalance” phenomenon; that is, most studies focus on the “human-centered” trust relationship and ignore the trust of intelligent agents (robots and digital humans) towards humans. To fill this gap in the research field of human-robot-digital human trust, we construct a “trinitarian” experience blending and sharing ecology across the virtual and real worlds, based on the use of multimodal signals to establish moderate unidirectional trust in human-robot-digital human interaction. This paper aims to explore the existing research, feasibility, and future development directions of multimodal cues and augmented reality in establishing trustworthy relationships among humans, robots, and digital humans. Finally, this paper envisions the application prospects of this model in social organization and social event situation perception and control, and points out the problems that need to be solved in the future. This paper contributes to the understanding of the role of multimodal cues and augmented reality in human-robot-digital human interaction, and provides ideas and solutions for the realization of the “trinitarian” trend.

Keywords augmented reality, multimodal interaction, human-robot-digital human interaction, trust