

【专题：人机智能交互与服务】

人机智能协同研究综述

孙效华, 张义文, 秦觉晓, 李璟璐, 王舒超
(同济大学, 上海 200092)

摘要: **目的** 智能系统在智能制造、智慧城市、医疗健康、生活服务等各种场景中越来越广泛地存在, 为应对人与智能系统的交互中所面临的诸多挑战, 从技术和体验的视角分析人机智能协同中的关键问题。**方法** 对从人机交互到人机智能协同的发展脉络与研究范围进行梳理, 提出综合技术视角和体验视角的研究框架; 从智能系统的特征出发, 梳理出技术视角下人机智能协同所带来的新兴问题; 从体验的视角探讨如何推动实现人机智能协同; 在此基础上总结人机智能协同的发展趋势。**结论** 总结了人机交互演进的三个阶段; 提出了技术视角下人机智能协同的关键问题, 包括人机能动性分配、动态学习和修正、情境自适应及主动响应模式; 探讨了体验视角下人机智能协同的可解释性、信任问题、情感化及公平负责等问题; 指出了人机智能协同全方位、多类型及体系化的发展趋势。

关键词: 人工智能; 智能系统; 人机交互; 人机智能协同; 技术视角; 体验视角

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2020)18-0001-11

DOI: 10.19554/j.cnki.1001-3563.2020.18.001

Review on Human-Intelligent System Collaboration

SUN Xiao-hua, ZHANG Yi-wen, QIN Jue-xiao, LI Jing-lu, WANG Shu-chao
(Tongji University, Shanghai 200092, China)

ABSTRACT: Intelligent system will exist widely in the future life and work. This work aims to analyze key issues in human-intelligent system collaboration from both technology-centric and experience-centric perspectives, to deal with the challenges in interactions between human and intelligent system. This work combed the evolution and research scope from human-computer interaction to human-intelligent system collaboration, and put forward a comprehensive research framework from both technology-centric and experience-centric perspectives. From the features of intelligent system, emerging problems brought by human-intelligent system collaboration were proposed, and the ways for promoting human-intelligent system collaboration were discussed from experience-centric perspective. Future trends of human-intelligent system collaboration were summarized on this basis. A three-stage framework of human-computer interaction evolution is proposed. Key issues in human-intelligent system collaboration from technology-centric perspective are concluded, including degree of agency, adaptive learning and correction, context-sensitive and proactive interaction. Experience-centric issues like explainability, trust, affective interaction, fairness and accountability are discussed. Trends of human-intelligent system collaboration are described as comprehensive, diverse, and systematic.

KEY WORDS: artificial intelligence; intelligent system; human-computer interaction; human-intelligent system collaboration; technology-centric perspective; experience-centric perspective

随着人工智能技术迅猛发展, 智能系统将普遍存在于智能制造、智慧交通、智慧生活、智能医疗等诸

多方面, 人能否与智能系统形成良好的交互及协作关系, 将影响着能否达到良好的用户体验, 以及能否最

收稿日期: 2020-04-15

作者简介: 孙效华 (1972—), 女, 河南人, 博士, 同济大学教授、博士生导师, 主要研究方向为人机智能交互与共融、大数据与信息可视化分析、车载人机界面与汽车服务创新。

通信作者: 张义文 (1991—), 男, 山东人, 同济大学博士生, 主要研究方向为智能系统及无人车中的交互设计。

大化发挥智能系统的能力。智能系统所具备的自主学习、自主决策、主动交互、情境感知等能力与特性,给人与智能系统交互的研究带来了大量全新的研究命题。当系统可以在一定程度上自主行动时,传统的“刺激—反应”的人机关系模式正在被改变,人机交互的时代正在向人机融合的时代发展,因此必须整体考虑人机的行为模式^[1]。在人工智能技术的驱动下,开始出现伙伴式的“人机智能协同(Human-Intelligent System Collaboration)”关系。人机智能协同并不是代替了人机交互,而是对其概念和研究范畴的延伸与发展。广泛应用的智能系统所具有的复杂性、不确定性,可能超出创造者的预期,产生难以预测的行为,对人和人社会造成相应的负面后果。因此,人机智能协同需要从综合的视角思考,既需要从智能系统的技术特征出发,探索人机智能协同的可能性,又需要从体验角度思考人机智能协同中存在的问题,应对来自于人和社会的挑战^[2]。

本文首先梳理了人机交互的演进轨迹,将其大致分为“人与物理系统交互(Human-Machine Interaction, HMI)——人与数字系统交互(Human-Computer Interaction, HCI)——人与智能系统交互(Human-Robot Interaction, HRI)”三个阶段,指出当前正处于从人机交互迈向人机智能协同阶段,并提出了人机智能协同的研究框架。然后主要从机器智能的视角,阐述了在智能系统特征的驱动之下,人机智能协同将产生哪些传统人机交互领域内不存在的研究问题,并综述了各个方面的相关研究现状。接着则主要从以人为本的体验视角,探讨了哪些技术和设计研究方向有助于构建新型的人机智能协同关系,并呈现了现有的相关研究中如何针对这些方面进行的考量。最后基于文献研究与分析,提出了几点对于未来人机智能协同研究发展趋势的思考。

1 从人机交互到人机智能协同

1.1 人机交互的演进

人机交互(Human-Computer Interaction)作为研究人与计算机系统之间信息交流的接口和以人为中心指导系统开发的方法论的一门学科,其研究领域自诞生以来就一直发生着变化,而这种变化乃至变革最关键的驱动力来自技术的革新,也就是“机”的演进^[3]。1960年,Licklider^[4]展望了人与新兴技术设备之间关系发展的三个阶段,分别为“人机交互(Human-Computer Interaction)”、“人机共融(Human-Computer Symbiosis)”和“超智机器(Ultra-Intelligent Machine)”。虽然过去六十年计算机及相关技术飞速发展,但是仍然处在Licklider所预言的第一个阶段,该阶段人机交互方式通常为人主动做出行为,计算机做出特定的反应。近年来随着智能技术被注入计算机系统,人类

才站到了“人机共融”的门前。在这人机关系面临巨大转变的时间点,有必要分析人交互的对象或系统的演进,进而了解人与不同系统的交互关系发生了怎样的变化。

现在人机交互中的“机”通常指的是计算机系统(Computer),然而实际上在计算机诞生之前,人与各种工具或机器的交互行为已经广泛存在。Grudin^[5]在其著作中对人机交互的历史进行了回顾,认为现代人机交互可以溯源到人因工程(Human Factors)领域,也即“人与工具的交互(Human-Tool Interaction)”。而近年来随着智能机器人越来越多地拥有了“类人”的智能特征,人与机器人交互(Human-Robot Interaction)也开始从交叉学科研究中逐渐成为一个独立的研究领域^[6]。因此,广义上的“人机交互”,其“机”的涵义可以得到拓展,不仅可以指“计算机”,还可以指“机器”或“机器人”。由此,可以将交互系统的演进分为三个阶段,即从“物理系统”到“数字系统”,再到“智能系统”。Harrison等人^[7]提出人机交互的研究经历了三个研究范式的演进:第一个范式为基于人因学的范式,研究关注于系统的物理属性与人的关系;第二个范式为基于认知主义或人类信息处理(Cognitivism/Information Processing)的范式,研究关注于人与计算机系统在信息交互中的认知过程;第三个范式为基于现象学情境(Phenomenologically-Situated),研究开始关注人与系统之外的情境因素,强调特定情境对人与系统交互的影响。通过整合以上观点,可以将人机交互的演进分为以下三个阶段。

1.1.1 人与物理系统交互(HMI)

当人机交互中的“机”为狭义的非数字化“机器”时,与人交互的系统是物理系统,见图1。这样的物理系统往往具有特定的功能,以满足人的特定需求,可被视为各种工具属性的系统。因此,物理系统与人的关系是支持性的,两者之间的交互往往是单向的交互,即人通过操作物理系统完成特定任务,物理系统是交互行为的被动承担者,其给人类的反馈也是相当有限的。这个阶段的人机交互可被称为Human-Machine Interaction(HMI),其关注点在于优化人与系统的物理属性之间的匹配。

1.1.2 人与数字系统交互(HCI)

当人机交互中的“机”为“计算机”时,与人交互的系统是数字系统,见图2。与传统的工具不同,计算机等数字系统可以运行设定好的程序,针对用户的输入执行相应的命令,做出预设性的反馈,因此人与数字系统的交互是双向、多重的,并且可以形成多轮开放式对话^[8]。数字与信息技术给人类带来了显著的增强作用,这一阶段的人机交互即是一般意义上的人与计算机系统的交互(HCI),从人的认知特性触发对人机信息交互中的认知过程的优化。



图 1 人与物理系统交互
Fig.1 Human-machine interaction

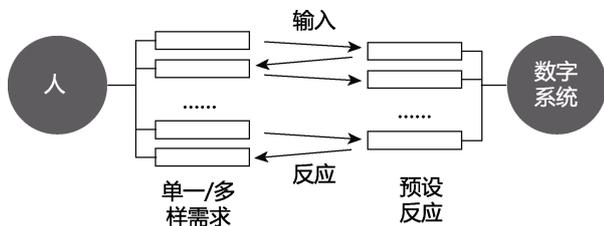


图 2 人与数字系统交互
Fig.2 Human-computer interaction

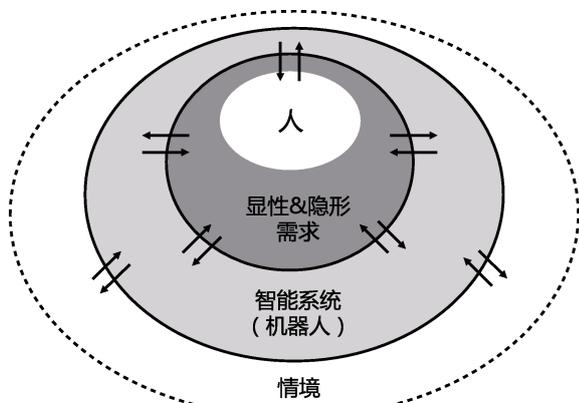


图 3 人与智能系统交互
Fig.3 Human-robot interaction

1.1.3 人与智能系统交互 (HRI)

当人机交互中的“机”为智能化的“机器人”时，与人交互的系统是智能系统，见图 3。人工智能技术的运用使智能系统的感知与认知能力大大加强，可将智能系统视为以各种形态呈现的具备不同功能的“机器人”，其不仅能对人的行为做出反应，还能通过自适应学习感知与理解人的潜在需求，成为交互行为的主动发起者。同时，智能系统可以对交互中的情境因素保持敏感，真正实现“人一机一环”三者的有机融合。这一阶段的人机交互已经超越了狭义的 HCI 的范畴，可被称为人与机器人系统的交互 (Human-Robot Interaction)。

由于机器人或智能系统拥有了上述特征，其与人的关系开始从工具性的关系向协作性的关系转变。在这样的转变趋势下，人与可协同的智能系统将逐渐走向人机智能协同阶段。

1.2 人机智能协同研究框架

在人工智能技术应用的许多领域内，面向人机智能协同的研究已经呈增长态势。例如，数据科学家们认为数据科学工作的未来将是人类与智能系统之间的合作，在这种合作中，智能系统和人类的专业知识

都是不可或缺的。在数据科学领域内，一种名为 AutoAI 的系统能够自主地接收和预处理数据，设计新功能，并基于目标（例如精度或实时效率）创建和评价模型^[9]。在创意领域内，一些 AI 辅助系统可以通过开发基于大规模草图数据集的深度学习模型，为创意者提供灵感启发及设计细节上的辅助参考，以实现人与人工智能的协同创作^[10]。机器学习也被越来越多地应用于支持医疗诊断的图像检索系统中，但目前还没有算法可以完美地捕捉医疗专家对于每一个案例的理解，因此谷歌从病理学家的需求出发，开发了一种以人为中心的工具，使专家与系统协同诊断，大大提升诊断效率的同时并未降低诊断准确率^[11]。

人机智能协同不是对人机交互的取代，而是对其研究范围的扩展。Hornbæk 等人^[12]通过分析 ACM SIGCHI 会议三十五年来的文章，认为对于人机交互的研究内容从大方向上主要分为交互方式和交互质量两方面。其中，交互方式与技术的发展密切相关，包括交互模态、信息设备、应用领域等，主要由人机交互中机器一侧的进化所驱动；而另一方面与交互质量相关的研究内容与人的感觉、认知等十分相关，通常基于交互中人本身的特性，对交互的有用性、可用性、效率、美学等方面的提升进行研究。因此，人机交互的研究内容是对人和机的特性的综合考量，其中用户体验是人机交互研究的目标所在，通过探究人机之间的关系与行为模式，最终落脚于人机界面上的优化及创新，形成一个完整的人机交互研究的框架。这一框架在人机智能协同研究中依然适用，但研究问题与内容得到了拓展。

智能系统具有以下几项关键特征：情境感知、自适应学习、自主决策及主动交互与协同^[13]。当人与具备这些特征的智能系统进行交互时，这些特征将衍生出一些在传统人机交互领域内不存在的新兴问题，包括人机能动性分配、动态学习与修正、情境自适应及主动响应模式等，也正是这些问题推动着人机交互向人机智能协同发生深刻的转变。然而这些问题更多地集中在人机智能协同的技术视角下，还有一些更关乎人的因素没有被考虑到。在以人为中心的视角下，人机智能协同中的可解释性、信任问题情感化以及公平负责等问题都是影响协同的关键问题。综合技术和体验的双重视角，本文提出了人机智能协同的研究框架，见图 4，并将分别对相关问题下的研究现状进行综述。

2 人机智能协同的技术视角

2.1 协同中的人机能动性分配

当人工智能技术为机器带来了自主性 (Autonomy) 之后，人与机器各自具有的能动性程度 (Degree of

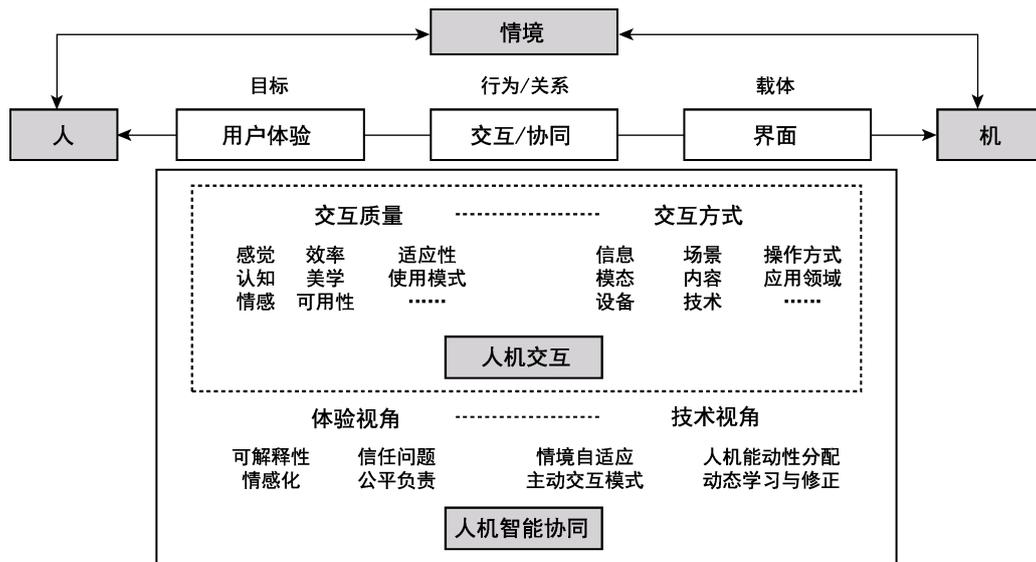


图4 人机智能协同研究框架

Fig.4 Research framework of human-intelligent system collaboration

Agency), 以及在人机的协作中如何分配, 就成了人机智能协同中的关键问题。智能体的能动性一方面体现在其独立行动的能力, 另一方面体现在其自由选择与决策的能力。现在各领域内的人工智能技术发展尚不均衡, 智能系统的自主化程度有高有低。因此, 在人机智能协同的设计中需要明确智能系统的能动性范围, 包括其行为能力和决策能力的边界, 并充分考虑人的能动性, 以使人机的能动性在协同中形成良好的匹配。Heer^[14]认为对机器完全自主化的追求低估了所要面临的突出挑战, 而忽视了人类的能动性。因此, 在系统的设计中, 应当使人与智能系统之间产生丰富且自适应的交互, 平衡人机之间的优势与劣势, 同时也可以促进人的控制行为走向成熟。

人工智能具有更大的计算信息处理能力和分析方法, 可以扩展人类在处理复杂问题时的认知; 而人类在处理组织决策中的不确定性和模糊性时仍然可以提供一种更全面、更直观的方法^[15]。Chakraborti 等人^[16]认为理解在环路中的人的能力与认知特性对于人机协同决策非常关键, 并讨论了促进有效人机协同的人的心智建模相关问题, 包括心理状态的表征, 以及可有效学习表征的方法及其可用性。Bansal 等人^[17]也认为人对于智能系统的能力(特别是其误差边界)所建立的心智模型, 对于人机智能协同而言十分关键, 并强调了几个关键属性: 误差边界的简约性与随机性, 以及任务维度。随着机器自主权的增加, 人的自主权和机器的自主权之间如果不能很好地协调, 将会产生十分紧张的人机关系。Sundar^[18]从心理学角度应用了交互媒体效应理论(Theory of Interactive Media Effects, TIME)的双流程框架, 研究了智能驱动的系统对用户感知与体验的符号效应与使能效应, 为人与人工智能系统的交互提供了一些方向。

2.2 协同中的动态学习和修正

在人与智能系统的交互过程中, 智能系统根据用户需求及喜好进行自主学习和自我修正的能力是其智能程度的重要体现, 也是推动人机实现更加智能的协同的重要因素。如何支持人与智能系统交互中的自适应学习, 将成为人机智能协同领域的重要问题。

英特尔的自适应机器人交互研究报告中指出, 目前的机器人在自主性尤其是学习能力方面比较欠缺, 应用模式单一, 而自适应机器人交互使机器人能够自主地进行持续学习, 并且能根据需求向用户发起交互进行学习, 从而提高现有 AI 技术的可用性^[19]。葛亚特等人^[20]认为机器人通过智能交互进行学习需要一系列的交互流程, 包括触发和引出学习话题、进行知识的学习、进一步确认知识并完成学习。这样的学习过程需要用户持续参与, 因此提升用户与自适应机器人交互的体验是提升机器人学习效率的重要手段。Sekmen 等人^[21]关于社交机器人 HRI 的研究表明, 机器人的自适应学习机制可以持续自主地更新用户模型, 学习与之交互的用户的偏好和行为, 以调整自己的行为来进行更有效和友好的交互。相比非适应性学习机器人, 受试者明显更偏好适应性学习机器人。

智能系统通常采用强化学习(Reinforcement Learning, RL)技术来实现人机智能协同中的动态学习与修正, Konstantinos 等人^[22]提出了一个交互式强化学习(IRL)框架, 该框架将人机智能协同中的显性反馈与隐性人工反馈相结合, 促进了智能系统的个性化。英特尔中国研究院开发了专门针对服务型机器人的开发工具——自适应机器人交互软件库(Adaptive HRI Library), 为服务型机器人提供自适应能力, 帮助其自主学习, 适应不同的情况。在其他领域如智能环境的研究中, Balos 等人^[23]提出系统可以使用由环

境或用户生成的上下文信息进行学习,使系统适应用户的偏好和需求。António 等人^[24]研究的 VirtualCare 项目就是这样的智能环境之一,该系统拥有与用户交互及其上下文环境的持续学习和自适应能力,使用不同类型的推理来实现包括用户建模、决策、活动预测和识别等方法,通过主动响应式技术增强用户体验。Liappas 等人^[25]关注于智能家居如何支持老年人、认知障碍患者等特殊群体的生活,通过头戴式眼动仪捕捉用户视觉注意力及周边环境信息,推断用户行为及上下文信息,相应的解决方案会根据用户及环境的需求动态变化,用户可以在个性化的环境中与不同场景深度交互。

2.3 协同中的情境自适应

长期以来,情境在许多领域中发挥着重要作用,尤其是在对于诸如预测情境变化、意外事件解释及其处理、辅助集中注意力等活动上。McCarthy^[26]最早针对人工智能领域对情境展开讨论,提出情境应当被视为由一系列控制彼此之间及和其他逻辑特征之间交互的公理所构成的一级对象。Maskery 等人^[27]认为情境应该被看作多个主体之间交互的属性,而不是特定问题或应用领域的固有属性。Brezillon 等人^[28]认为情境表现为共享的知识空间,可以被分为三种不同的内容,即外部知识、情境知识和可程序化的情境。他还认为在人工智能的应用中,情境显式表达的缺失是大部分基于知识的系统失败的主要原因,而这些失败主要表现在四个方面:(1)系统无法为用户解决问题;(2)系统不能正确地应用知识;(3)系统无法在初始状态下拥有所有必需的知识;(4)由于不知道解决用户问题所处的情境是什么而无法为用户提供相关说明^[29]。

在人机智能协同中智能系统对于情境的理解与应用方面,Turner^[30]提出了情境介导行为(Context-mediated Behavior, CMB)的方法,讨论了赋予 AI 对情境的推理和理解能力所需要的必要条件,包括对情境的感知与理解、对意外事件的处理、需要实现的目标和行为规划。开放世界中的人机智能协同将对情境自适应提出更高的要求。相比传统对话环境,开放世界中对话环境通常包含多个动态且异步的参与方,而周围的环境将为处于这种情境中的人机协同提供丰富的、变化的、与对话的进行和组织相关的情境内容。Bohus 等人^[31-33]在一系列关于开放世界对话系统的研究和原型开发中,探索了在开放世界中多方参与的人机智能协同所面临的挑战和具备的潜力。Andrist 等人^[34]发现开放世界中的机器人总会不可避免地遇到以娱乐为目的的交互发起行为,或由不了解机器人主要功能领域及局限性的用户所发起的交互,在这些情况下,能够识别任务外的交互,并将交互引导回其本身的主要功能是非常有价值的。Pejsa 等人^[35]认为能够基于情境感知进行无感交互的系统需要具备五

个能力:感知、参与、轮换、语言理解和意图理解。他们希望通过共享交互中的不确定性,通过交互参与者的帮助,以自然的方式提升人机智能协同系统的交互能力。

目前,用于开发和研究这一类智能系统的框架和平台也有了越来越多的尝试。Bohus 等人^[36]描述了一个用于构建和研究多模态协同系统的开放性可扩展平台,该平台提供了一个用于并行协调计算的基于时间和流的编程模型,一套用于数据可视化、处理和学习的工具与人工智能组件。Andrist 等人^[37]的研究展示了如何利用开放性可扩展框架更快地开发和研究基于情境的人机智能协同系统,并展示了在开放世界中通过自然语言与用户进行交互的系统组件和交互工具包。

2.4 协同中的主动交互模式

当人工智能技术使机器可以获取和处理与用户及情境相关的海量信息,预测用户意图并提前发现潜在的问题时,机器便从单纯的执行方转变为有理解能力、学习能力、决策能力的合作方,从而支持主动的响应与改变^[38]。从用户视角出发,机器在人机协同过程中体现出了更强的主动性。关于人机交互中主动性的定义,研究领域存在一定的常识性共识,多将主动性作为一种交互状态的描述或评价。Jun 等人^[39]指出拟人化界面助手的主动性是指除了响应用户提问外,提出的未经请求的建议。Antti S 等人^[40]在主动性的一般概念的基础上,从资源管理角度阐述了主动性(Proactivity)的概念,给出了六种主动的资源管理模式,以扩展研究人员对主动系统如何支持用户的观点。该研究认为当系统不仅限于对外界事件进行反应,还对用户的目标和未来可能的选择敏感时,其资源管理模式可以被称为主动的。Keith 等人^[41]在研究中依然使用主动性的一般概念,认为系统具有主动性的前提是系统作为用户的代表,以及在无用户命令下的主动发起。Wendy^[42]在隐式交互框架中将主动性概括为描述隐式交互的一个维度,即交互是由智能系统而不是用户发起,以构成隐式交互的另一种方式。

人机协同系统具有主动性的特征,通常需要感知、建构、推理所处的物理世界,所以对于主动响应的很多尝试性研究,集中于技术领域,多为面向传感器、数据收集、用户和任务模型等问题^[43-45]。同时,从技术视角提升智能系统的预测准确性,可以更好地支持系统主动性,实现自然、直觉化的交互方式。Martin B 等人^[46]在任务相关的主动人机交互研究中,提出通过语音、手势和面部情感模仿的方式实现多模态交互,使机器人能够以一种自然而直观的方式主动向人类发起缺失的任务知识检索,从而控制对话。Md 等人^[47]提出一个基于视觉的导览机器人系统,通过摄像机传感器的检测与跟踪预测用户对绘画的意图,提供主动的引导与讲解服务。Oliver 等人^[48]提出

一种系统架构,定义人机交互的关键模块与交互关系,在不确定的人类意图知识的情况下,主动执行人机协作中的机器人任务,以减少系统的不确定性。

主动交互的系统会影响用户对系统的感知^[49-50]。如何让用户接受和信任,仍是需要解决的一个问题^[51]。一些研究将主动交互作为增进人机智能协同中可理解性的一种有效手段,一方面通过减少用户与系统的交互,降低分心度与认知负荷,另一方面通过增加系统的解释性,提升接受度。Alexander等人^[52]提出了一种基于上下文的协同推荐系统、主动信息支持等技术的智能驾驶支持系统。Roland B等人^[53]认为主动推荐系统可以减少驾驶中的信息过载问题,通过添加解释的方法,提高没有明确的用户请求下主动系统的可理解性。

3 人机智能协同的体验视角

人工智能技术被广泛应用到人们的日常生活中,带来便利的同时也给人类社会生活带来了各种问题。这些问题很多源自于其算法自身的黑盒特征和不确定性。人工智能算法,尤其是机器学习算法的不透明性使得人们很难深入了解其内部工作机制,而将重要决策委托给无法自我解释的系统会带来明显的风险^[54]。Yang等人^[55]也认为人机智能协同中的挑战主要来源于两方面,一是对于智能系统能力边界的不确定性,二是智能系统输出的复杂性。应对这些挑战需要以人为中心的视角,明确这些智能系统的技术特性在人机智能协同中会体现在什么方面。将对体验视角下的人机智能协同的相关议题作如下综述。

3.1 人机智能协同中的可解释性

近年来可解释的人工智能(Explainable AI, XAI)成为前沿的研究热点,主要源于深度机器学习的兴起及“黑箱”算法模型在透明度方面的缺乏。许多研究表明,如果在人机智能协同中用户不能理解智能系统的工作原理与决策机制,将会给协同表现及用户体验带来消极影响^[56]。在可解释性研究领域内,有几个术语经常被交叉使用,例如可解释性(Explainability)、可判读性(Interpretability)和透明度(Transparency),而它们之间实际上存在一定的区别。透明度是智能系统能够提供其内部工作原理或结构的程度,是系统本身的属性,而可判读性与可解释性都是其不同方面的体现。可判读性强调对于智能系统中的模型输出结果可判读、可预测;可解释性则是以人为中心的视角,强调人在多大程度上能够理解智能系统的逻辑^[57-58]。因此,可解释性是更能直接影响人机智能协同的因素。

在如何增强人机智能协同的可解释性方面,一种研究路线是开发创新的解释算法。许多方法依赖于可视化技术(如显著性图),以突出显示与模型决策最相关的输入部分,但是这种方法主要面向人工智能专

家构建解释方案,对于终端用户来说,智能系统如何获取信息或决策仍然是不易理解的^[59-60]。另外一种路径即以人为中心的方法,在人类行为的概念和语言框架范围内产生解释。Weitz等人^[61]在一个可判读的语音识别模型的基础上,进一步探究了加入拟人化的虚拟智能体对于可解释交互设计的影响。Riber等人^[62]的研究关注如何为不同的用户人群提供不同的解释,其案例研究中将用户分为AI专家或开发者、专业领域专家和终端用户三类,各类用户与智能系统的关系不同,对于解释的需求与期望也有所不同。

智能系统对于用户的可解释性并非越高越好,过量的解释信息会造成信息过载,进而影响人机智能协同的表现。Poursabzi-Sangdeh等人^[63]的一项研究发现,对于功能相同的两个算法模型,使用透明度较高的模型的用户对于输出结果的预测性更好,但是用户信任度方面并没有显著差异,甚至较高的透明度会减弱用户发现系统发生重大错误的能力。Wang等人^[64]提出了以人为中心的、决策理论驱动的XAI概念框架,探索了人类认知模式驱动构建XAI的需求的路径,以及XAI如何减轻常见的认知偏差。Hois等人^[65]从用户类型、系统范围与限制、情境和随时间变化等维度上,探讨了解决特定情况下智能系统透明度问题的最佳方式。

3.2 人机智能协同中的信任问题

由于人工智能技术依然处于不断发展和完善的过程中,相应的智能系统还远未达到完美,所以人机智能协同中用户对于智能系统的信任问题是关键。信任(Trust)是人的一种态度上的判断,指在具有不确定性的情况下,用户可以在多大程度上依赖智能系统来实现一定的目标^[66]。人机智能协同的安全性和效率通常取决于人类如何恰当地校准对智能系统的信任,对具有自主性的智能系统缺乏信任或过度信任有时会导致协同表现的降低,甚至严重的安全问题。Nguyen等人^[67]研究了人如何与人工智能事实核查系统交互并建立信任关系,提出了一种混合主动权的事实核查方法,将人类的知识和经验与自动化信息检索和机器学习的效率,以及可扩展性相结合。通过用户研究发现,用户倾向于信任系统,但过度信任会导致模型出错时用户准确率的降低。

许多研究认为通过增强系统的透明度来保持恰当的信任校准(Trust Calibration),是促进人与易出错的智能系统进行有效交互的关键,而实际上探测并减轻不正常的信任校准同样重要。Okamura等人^[68]提出了一种自适应信任校准的方法,通过监测用户的依赖行为发现不正常的信任校准,并通过认知线索提示用户进行信任重校准。Ferrario等人^[69]提出了一个信任增量模型,其中描述了简单信任、反思信任及范式信任的递进关系,既可以应用于人与智能系统交互中,也可以应用于人与人的交互中。De Visser等人^[70]

提出了一种基于纵向视角的人与机器人协同的信任发展与校准一体化模型,创新地引入关系公平的概念,可以预测两个行为体之间的善意程度,并进一步帮助预测长期人机协同关系的未来健康状况。

3.3 人机智能协同中的情感化

人机智能协同中的情感研究是一个多学科交叉的领域,相关问题包括智能系统如何对人类情感进行感知与理解,人与智能系统之间的交互与协同会受到情感怎样的影响,以及如何利用情感设计智能系统以提升其能力等。情感计算(Affective Computing)是MIT媒体实验室的Picard所提出的概念,其目的是为了促进智能系统“情商”的发展,以使其发展出类人的感知、理解,以及产生各种情感特征的能力,进而有助于自然、和谐而富有情感的人机交互^[71]。Mou等人^[72]在研究中发现人类用户在与智能系统的初始交互中,会表现出与人类交互不同的人格特征和交际属性。用户在与人类互动时往往更开放、更随和、更外向、更认真、更自我表露,呼应了米舍尔的认知—情感处理系统模型。人类面对人和智能系统所采取的不同策略,应当在人机情感研究方面得到考虑。

在一些具体领域内人机智能协同的情感化研究方面,Miyaji等人^[73]主要对看护机器人情感行为的相关功能进行了研究,认为机器人必须了解人的情感状态,并建立虚拟情感状态以表达情感,能够在交互与协同中产生情感化行为。Röning等人^[74]研究了服务机器人中的情感交互,开发了一种面向HRI的分布式系统,结合了所有基本的感官元素,可以识别用户情绪并通过情感交互提供个性化服务。Novikova等人^[75]探索了人工情感(Artificial Emotions)在人与智能系统交互中进行与任务相关的协调的潜力,发现即使是非人型机器人的简单动作同样可以传达情感含义,并且当人们将情感状态归因于机器人本身时,通常会用基于时间的框架来理解其所看到的机器人表情。Coyne等人^[76]提出了将一种测量情绪反应的自我报告工具“日内瓦情绪轮(Geneva Emotion Wheel, GEW)”,并将其应用在智能系统的情感表达评估方面,提供了一个可重复的系统框架以评估社交机器人表情的感知情感。Liu X等人^[77]通过一个实体智能交互界面TAI,增强用户与对话机器人之间交流的物理沉浸感,增强了用户对于虚拟智能系统的共情与社交联系,改善了人与智能系统交流的情感体验。

3.4 公平负责的人机智能协同

在人工智能不断渗入社会与生活中的同时,公平(Fairness)和负责(Accountability)相关的问题也逐渐引起人们的关注,而研究者们也致力于运用技术及设计的手段推动实现公平负责的人机智能协同。人类天生容易受到认知偏见的影响,而当涉及到一些敏感数据时,人的偏见不可避免地会嵌入到所创造的智

能系统中。构建公平的人机智能协同需要通过不断的研究和数据的收集来使偏见最小化。微软开发了一个评估和改善人工智能系统公平性的工具包——Fairlearn,其拥有交互式可视化仪表盘和不公平缓解算法两个主要组件,可以帮助数据科学家在智能系统的性能和公平性之间进行权衡^[78]。Harrison等人^[79]通过一个涉及五百多名被试的实证研究,评估了人们对于一个不完美的算法模型的感知公平性,发现人们对于“无偏”和“公平”的感知之间还存在细微差别。在残障人士与智能系统的交互与协同这一细分领域中,智能系统有极大潜力给残障人士带来生活上的便利,如利用语言预测算法增强言语或认知障碍人士的交流。然而,一般的智能系统若在设计与开发中缺乏相关考虑,很可能难以与残障人士交互,甚至产生歧视。Guo等人^[80]研究了几个人工智能技术类别对特定的残障人士群体可能产生的影响,并预期对其相互作用进行风险评估,以形成未来研究的路线图,指导构建更包容的算法,促进更公平负责的人机智能协同。

负责任的智能系统在人与人的交互与协同中,应当为其决策提供明确的依据与推理过程,这也是使人能够与智能系统进行协同决策的前提条件。Abbass^[81]对于人工智能与人类在社会层面上的融合进行了批判式的文献研究,认为无论智能系统有多复杂,人类应当始终主动地参与到智能系统的决策环路中,或保持在决策环路之上,这是智能系统融入人类社会,构建负责任的人机智能协同关系的保证。Nushi等人^[82]设计开发了一套人机混合的方法及工具——Pandora,可以综合利用人与系统生成的观测结果来描述和解释系统故障,加深了人机智能协同下对系统行为和故障的理解。

4 人机智能协同的发展趋势

通过对人机智能协同的技术视角及体验视角下的研究进行综述,可以发现人机智能协同目前尚处于由交互向协同转变初级的阶段,人机关系中协同性的体现仍然不全面、不充分、不深入。本文结合相关研究,提出在人工智能技术发展的驱动之下,人机智能协同将呈现以下三方面的趋势。

4.1 全方位

在人机智能得到充分融合的条件下,人机智能协同所形成的系统可以被认为是一种新型的人机混合智能系统,而未来的人机智能协同将在信息传递、信息处理及决策执行等方面都得到全方位的充分体现,而不仅是其中有限的部分^[83]。在信息传递中,需要实现客观数据(传感器)与主观信息(人的感知)的有效融合。一些支持人机协同信息融合的技术与方法研究已经取得了一定的进展,如可以将传感器数字信号转变为类神经模拟信号的神经拟态器件,有助于实现

数字信号与生物神经信号的兼容^[84];还有对多通道人机交互中的信号实现有效融合的智能方法,以支持对于用户意图的准确理解,从而为人机协同提供条件^[85]。在信息处理方面,需要将智能系统的计算与人的认知实现有效融合。人在环路的混合智能即是一种主要路径,相关的问题包括如何利用人机智能协同使智能系统能够以自然的方式在智能回路中进行训练,以及如何构建任务使机器能够同时从海量数据和人的知识中进行学习等^[86]。在决策执行方面,需要将智能系统概率化的模型输出结果与人的规则化价值判断进行匹配,以支持决策以人机智能协同的形式得到有效执行。

4.2 多类型

现在的人机智能协同的相关研究大多还停留在个体的层面上,这与当前人工智能技术的研究与应用多数为单体智能有关。未来的人机智能协同将随着群体智能技术的发展与成熟,逐渐扩展到集群化的人机智能协同。未来人机智能协同的类型将更加丰富多样,既包括人与单个或多个独立智能系统的协同,也包括人与群体智能系统的协同,此时群体智能系统将呈现出单个智能系统所不具备的群体特性。更进一步地,在有多个个体参与任务协作情况下,还将形成多人与群体智能系统的群体混合协同。

4.3 体系化

当人机智能协同进一步拓展到社会层面上时,面对复杂社会环境下更广阔范围内的问题与挑战,人机智能协同将以更加体系化的形式呈现。在这一层面上,一种研究路径是探讨如何将智能系统嵌入到当前的复杂社会技术系统中,以渐进的方式促进人与更为智能化的复杂社会技术系统的融合;另一种研究路径是面向实际的社会需求,以服务体系的视角构建智能化的解决方案,其中多种利益相关方及智能系统将形成更为复杂的协同关系。

5 结语

在人工智能技术的快速发展之下,人机交互向人机智能协同的转变已经是不可逆转的趋势。然而各个领域内应用的智能系统现在还远未到成熟的程度,因此如何促进人机智能协同在当前有着广阔的研究空间。本文首先厘清了人机智能协同相对于传统人机交互所延展的研究领域和问题;然后以技术和体验的双重视角,收集、分析并总结了相关研究现状;最后展望了人机智能协同的全方位、多类型、体系化的未来发展趋势。

参考文献:

[1] FAROOQ U, GRUDIN J. Human-Computer Integra-

tion[J]. *Interactions*, 2016, 23(6): 26-32.

- [2] PEETERS M M M, DIGGELEN J V, BOSCH K V D, et al. Hybrid Collective Intelligence in a Human: AI Society[J]. *AI & SOCIETY*, 2020(3): 1-22.
- [3] 范俊君, 田丰, 杜一, 等. 智能时代人机交互的一些思考[J]. *中国科学: 信息科学*, 2018, 48(4): 361-375.
FAN Jun-jun, TIAN Feng, DU Yi, et al. Thoughts on Human-computer Interaction in the Age of Intelligence[J]. *Science in China: Information Science*, 2018, 48(4): 361-375.
- [4] LICKLIDER J C R. Man-computer Symbiosis[J]. *IRE Transactions on Human Factors in Electronics*, 1960(1): 4-11.
- [5] GRUDIN J. From Tool to Partner: the Evolution of Human-computer Interaction[M]. San Rafael: Morgan & Claypool Publishers, 2017.
- [6] GOODRICH M A, SCHULTZ A C. Human-robot Interaction: a Survey[J]. *Foundations and Trends® in Human-computer Interaction*, 2008, 1(3): 203-275.
- [7] HARRISON S, SENGERS P, TATAR D. Making Epistemological Trouble: Third-paradigm HCI as Successor Science[J]. *Interacting with Computers*, 2011, 23(5): 385-392.
- [8] SUCHMAN L A. Plans and Situated Actions: the Problem of Human-machine Communication[M]. Cambridge: Cambridge University Press, 1987.
- [9] WANG D, WEISZ J D, MULLER M, et al. Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI[J]. *Proceedings of the ACM on Human-Computer Interaction*, 2019, 3(CSCW): 1-24.
- [10] KARIMI P, REZWANA J, SIDDIQUI S, et al. Creative Sketching Partner: an Analysis of Human-AI Co-creativity[C]. Cagliari: Proceedings of the 25th International Conference on Intelligent User Interfaces, 2020.
- [11] CAI C J, REIF E, HEGDE N, et al. Human-centered Tools for Coping with Imperfect Algorithms During Medical Decision-making[C]. Glasgow: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019.
- [12] HORNBAEK K, MOTTELSON A, KNIBBE J, et al. What do We Mean by "Interaction"? an Analysis of 35 Years of CHI[J]. *ACM Transactions on Computer-Human Interaction(TOCHI)*, 2019, 26(4): 1-30.
- [13] 孙效华, 张义文, 侯璐, 等. 人工智能产品与服务体系研究综述[J]. *包装工程*, 2020, 41(10): 49-61.
SUN Xiao-hua, ZHANG Yi-wen, HOU Lu, et al. Artificial Intelligence Products and Service System[J]. *Package Engineering*, 2020, 41(10): 49-61.
- [14] HEER J. Agency Plus Automation: Designing Artificial Intelligence into Interactive Systems[J]. *Proceedings of the National Academy of Sciences*, 2019, 116(6): 1844-1850.
- [15] JARRAHI M H. Artificial Intelligence and the Future of

- Work: Human-AI Symbiosis in Organizational Decision Making[J]. *Business Horizons*, 2018, 61(4): 577-586.
- [16] CHAKRABORTI T, KAMBHAMPATI S. Algorithms for the Greater Good! on Mental Modeling and Acceptable Symbiosis in Human-AI Collaboration[J]. *arXiv Preprint arXiv*: 2018.
- [17] BANSAL G, NUSHI B, KAMAR E, et al. Beyond Accuracy: the Role of Mental Models in Human-AI Team Performance[C]. *Stevenson: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 2019.
- [18] SUNDAR S S. Rise of Machine Agency: a Framework for Studying the Psychology of Human-AI Interaction (HAII)[J]. *Journal of Computer-mediated Communication*, 2020, 25(1): 74-88.
- [19] 英特尔中国研究院. 自适应机器人交互白皮书 [EB/OL]. (2019-08-08)[2020-04-10]. <http://www.robotplaces.com/info/3707.jhtml>.
Intel Research. Adaptive Robot Interaction [EB/OL]. (2019-08-08)[2020-04-10]. <http://www.robotplaces.com/info/3707.jhtml>.
- [20] 葛亚特, 叶露. 面向自适应机器人交互的类人反应研究[J]. *工业设计研究*, 2018: 326-331.
GE Ya-te, YE Lu. Human-like Reaction for Adaptive Robot Interaction[J]. *Industrial Design Research*, 2018: 326-331.
- [21] SEKMEN A, CHALLA P. Assessment of Adaptive Human-robot Interactions[J]. *Knowledge-based Systems*, 2013, 42: 49-59.
- [22] KONSTANTINOS T, MAHER A, FILLIA M. Task Engagement as Personalization Feedback for Socially-assistive Robots and Cognitive Training[J]. *Technologies*, 2018, 6(2): 49-65.
- [23] BALOS N, ESPEJO G, RAMÓN LÓPEZ-CÓZAR, et al. A Multimodal Dialogue System for an Ambient Intelligent Application in Home Environments[C]. *Brno: International Conference on Text*, 2010.
- [24] ALMEIDA A, COSTA R A F, LIMA L, et al. Non-obstructive Authentication in AAL Environments[J]. *Ambient Intelligence and Smart Environments*, 2011: 63-73.
- [25] LIAPPAS N, JOSÉ GABRIEL TERIÚS-PADRÓN, MACHADO E, et al. Best Practices on Personalization and Adaptive Interaction Techniques in the Scope of Smart Homes and Active Assisted Living[C]. *Leiceste: SmartWorld 2019: IEEE Smart World Congress*, 2019.
- [26] MCCARTHY J. First Order Theories of Individual Concepts and Propositions[J]. *Cogprints*, 1979.
- [27] MASKERY H, MEADS J. Context: in the Eyes of Users and in Computer Systems[J]. *ACM SIGCHI Bulletin*, 1992, 24(2): 12-21.
- [28] BRÉZILLON P, POMEROL J C. Contextual Knowledge and Proceduralized Context[C]. *Orlando: Proceedings of the AAAI-99 Workshop on Modeling Context in AI Applications*, 1999.
- [29] BRÉZILLON P. Context in Human-machine Problem Solving: a Survey[J]. *LIP*, 1996, 6: 29.
- [30] TURNER R M. Context-mediated Behavior for AI Applications[C]. *Heidelberg: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 1998.
- [31] BOHUS D, HORVITZ E. Dialog in the Open World: Platform and Applications[C]. *New York: Proceedings of the 2009 International Conference on Multimodal Interfaces*, 2009.
- [32] BOHUS D, HORVITZ E. Open-world Dialog: Challenges, Directions, and a Prototype[C]. *Acapulco: Proceedings of Ijcai Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 2009.
- [33] BOHUS D, SAW C W, HORVITZ E. Directions Robot: In-The-Wild Experiences and Lessons Learned[C]. *Paris: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, 2014.
- [34] ANDRIST S, BOHUS D, YU Z, et al. Are You Messing with Me? Querying about the Sincerity of Interactions in the Open World[C]. *Christchurch: 11th ACM/IEEE International Conference on Human-Robot Interaction*, 2016.
- [35] PEJSA T, BOHUS D, COHEN M F, et al. Natural Communication about Uncertainties in Situated Interaction[C]. *Istanbul: Proceedings of the 16th International Conference on Multimodal Interaction*, 2014.
- [36] BOHUS D, ANDRIST S, JALOBEANU M. Rapid Development of Multimodal Interactive Systems: a Demonstration of Platform for Situated Intelligence[C]. *Glasgow: Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017.
- [37] ANDRIST S, BOHUS D, FENIELLO A. Demonstrating a Framework for Rapid Development of Physically Situated Interactive Systems[C]. *Daegu: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction*, 2019.
- [38] 李璟璐, 孙效华, 郭炜炜. 基于智能交互的汽车主动响应式交互设计[J]. *图学学报*, 2018(4): 668-674.
LI Jing-lu, SUN Xiao-hua, GUO Wei-wei. Proactive HMI Design Based on Smart Interaction[J]. *Journal of Graphics*, 2018(4): 668-674.
- [39] XIAO J, CATRAMBONE R, STASKO J. Be Quiet? Evaluating Proactive and Reactive User Interface Assistants[C]. *Edinburgh: Proceedings of INTERACT*, 2003.
- [40] SALOVAARA A, OULASVIRTA A. Six Modes of Proactive Resource Management: a User-centric Typology for Proactive Behaviors[C]. *Tampere: Proceedings of the Third Nordic Conference on Human-computer Interaction*, 2004.
- [41] CHEVERST K, BYUN H E, FITTON D, et al. Explor-

- ing Issues of User Model Transparency and Proactive Behaviour in an Office Environment Control System[J]. *User Modeling and User-adapted Interaction*, 2005, 15(3): 235-273.
- [42] JU W, LEIFER L. The Design of Implicit Interactions: Making Interactive Systems Less Obnoxious[J]. *Design Issues*, 2008, 24(3): 72-84.
- [43] HORVITZ E, KADIE C, PAEK T, et al. Models of Attention in Computing and Communication: from Principles to Applications[J]. *Communications of the ACM*, 2003, 46(3): 52-59.
- [44] SCHMIDT A. Implicit Human Computer Interaction through Context[J]. *Personal Technologies*, 2000, 4(2): 191-199.
- [45] SALBER D, DEY A K, ABOARD G D. The Context Toolkit: Aiding the Development of Context-enabled Applications[C]. Atlanta: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1999.
- [46] BUSS M, CARTON D, GONSIOR B, et al. Towards Proactive Human-robot Interaction in Human Environments[C]. Budapest: 2011 2nd International Conference on Cognitive Infocommunications (CogInfoCom), 2011.
- [47] RASHED M G, SUZUKI R, LAM A, et al. A Vision Based Guide Robot System: Initiating Proactive Social Human Robot Interaction in Museum Scenarios[C]. Prague: 2015 International Conference on Computer and Information Engineering (ICCIIE), 2015.
- [48] SCHREMPF O C, HANEBECK U D, SCHMID A J, et al. A Novel Approach to Proactive Human-robot Cooperation[C]. Nashville: ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.
- [49] BARAGLIA J, CAKMAK M, NAGAI Y, et al. Initiative in Robot Assistance during Collaborative Task Execution[C]. Christchurch: 11th ACM/IEEE International Conference on Human-Robot Interaction, 2016.
- [50] PENG Z, KWON Y, LU J, et al. Design and Evaluation of Service Robot's Proactivity in Decision-making Support Process[C]. Denver: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019.
- [51] KRAUS M, WAGNER N, MINKER W. Effects of Proactive Dialogue Strategies on Human-Computer Trust[C]. Genoa: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, 2020.
- [52] SMIRNOV A, SHILOV N, GUSIKHIN O. Socio-cyber-physical System for Proactive Driver Support Approach and Case Study[C]. Alsace: The 12th International Conference on Informatics in Control, Automation and Robotics, 2015.
- [53] BADER R. Proactive Recommender Systems in Automotive Scenarios[D]. München: Technische Universität München, 2013.
- [54] ADADI A, BERRADA M. Peeking inside the Black-Box: a Survey on Explainable Artificial Intelligence (XAI)[J]. *IEEE Access*, 2018, 6: 52138-52160.
- [55] YANG Q, STEINFELD A, ROSÉ C, et al. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design[C]. Denver: Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems, 2020.
- [56] STUBBS K, HINDS P J, WETTERGREEN D. Autonomy and Common Ground in Human-robot Interaction: a Field Study[J]. *IEEE Intelligent Systems*, 2007, 22(2): 42-50.
- [57] ROSENFELD A, RICHARDSON A. Explainability in Human-agent Systems[J]. *Autonomous Agents and Multi-Agent Systems*, 2019, 33(6): 673-705.
- [58] CLINCIU M A, HASTIE H. A Survey of Explainable AI Terminology[C]. Tokyo: Proceedings of the 1st Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence (NL4XAI 2019), 2019.
- [59] MILLER T, HOWE P, SONENBERG L. Explainable AI: Beware of Inmates Running the Asylum or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences[J]. arXiv Preprint arXiv, 2017.
- [60] WEITZ K, HASSAN T, SCHMID U, et al. Deep-learned Faces of Pain and Emotions: Elucidating the Differences of Facial Expressions with the Help of Explainable AI Methods[J]. *Tm-Technisches Messen*, 2019, 86(7): 404-412.
- [61] WEITZ K, SCHILLER D, SCHLAGOWSKI R, et al. "Let Me Explain!": Exploring the Potential of Virtual Agents in Explainable AI Interaction Design[J]. *Journal on Multimodal User Interfaces*, 2020: 1-12.
- [62] RIBERA M, LAPEDRIZA A. Can We Do Better Explanations? A Proposal of User-centered Explainable AI[C]. Los Angeles: IUI Workshops, 2019.
- [63] POURSAZBI-SANGDEH F, GOLDSTEIN D G, HOFMAN J M, et al. Manipulating and Measuring Model Interpretability[J]. arXiv Preprint arXiv, 2018.
- [64] WANG D, YANG Q, ABDUL A, et al. Designing Theory-driven User-centric Explainable AI[C]. Glasgow: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019.
- [65] HOIS J, THEOFANOU-FUELBIER D, JUNK A J. How to Achieve Explainability and Transparency in Human AI Interaction[C]. Paphos: International Conference on Human-Computer Interaction, 2019.
- [66] LEE J D, SEE K A. Trust in Automation: Designing for Appropriate Reliance[J]. *Human Factors*, 2004, 46(1): 50-80.
- [67] NGUYEN A T, KHAROSEKAR A, KRISHNAN S, et al. Believe it or not: Designing a Human-AI Partnership for Mixed-initiative Fact-checking[C]. Berlin: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, 2018.
- [68] OKAMURA K, YAMADA S. Adaptive Trust Calibration

- for Human-AI Collaboration[J]. *Plos One*, 2020, 15(2): e0229132.
- [69] FERRARIO A, LOI M, VIGANÒ E. In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-artificial Intelligence Interactions[J]. *Philosophy & Technology*, 2019: 1-17.
- [70] DE VISSER E J, PEETERS M M M, JUNG M F, et al. Towards a Theory of Longitudinal Trust Calibration in Human-robot Teams[J]. *International Journal of Social Robotics*, 2020, 12(2): 459-478.
- [71] PICARD R W. Affective Computing: Challenges[J]. *International Journal of Human-computer Studies*, 2003, 59(1): 55-64.
- [72] MOU Y, XU K. The Media Inequality: Comparing the Initial Human-human and Human-AI Social Interactions[J]. *Computers in Human Behavior*, 2017, 72: 432-440.
- [73] MIYAJI Y, TOMIYAMA K. Implementation Approach of Affective Interaction for Caregiver Support Robot[J]. *Journal of Robotics and Mechatronics*, 2013, 25(6): 1060-1069.
- [74] RÖNING J, HOLAPPA J, KELLOKUMPU V, et al. Minotaurus: A System for Affective Human-robot Interaction in Smart Environments[J]. *Cognitive Computation*, 2014, 6(4): 940-953.
- [75] NOVIKOVA J, WATTS L. Towards Artificial Emotions to Assist Social Coordination in HRI[J]. *International Journal of Social Robotics*, 2015, 7(1): 77-88.
- [76] COYNE A K, MURTAGH A, MCGINN C. Using the Geneva Emotion Wheel to Measure Perceived Affect in Human-robot Interaction[C]. Cambridge: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 2020.
- [77] LIU X, LONDON K. Tai: a Tangible AI Interface to Enhance Human-artificial Intelligence(AI) Communication beyond the Screen[C]. Brisbane: Proceedings of the 2016 ACM Conference on Designing Interactive Systems, 2016.
- [78] MIRO D, SARAH B, HANNA W, et al. Fairlearn: a Toolkit for Assessing and Improving Fairness in AI[EB/OL]. (2020-05-18)[2020-06-10]. https://www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_whitepaper.pdf,2020-5.
- [79] HARRISON G, HANSON J, JACINTO C, et al. An Empirical Study on the Perceived Fairness of Realistic, Imperfect Machine Learning Models[C]. Barcelona: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020.
- [80] GUO A, KAMAR E, VAUGHAN J W, et al. Toward Fairness in AI for People with Disabilities: a Research Roadmap[J]. arXiv Preprint arXiv, 2019.
- [81] ABBASS H A. Social Integration of Artificial Intelligence: Functions, Automation Allocation Logic and Human-autonomy Trust[J]. *Cognitive Computation*, 2019, 11(2): 159-171.
- [82] NUSHI B, KAMAR E, HORVITZ E. Towards Accountable AI: Hybrid Human-machine Analyses for Characterizing System Failure[J]. arXiv Preprint arXiv, 2018.
- [83] 刘伟. 人机融合智能的再思考[J]. *人工智能*, 2019(4): 13. LIU Wei. Rethinking Hybrid Human-machine Intelligence[J]. *Artificial Intelligence*, 2019(4): 13.
- [84] 陆骥峰, 孙富钦, 王子豪, 等. 柔性人工突触: 面向智能人机交互界面和高效神经网络计算的基础器件[J]. *材料导报*, 2020, 34(1): 1022-1049. LU Qi-feng, SUN Fu-qin, WANG Zi-hao, et al. Recent Advances in Flexible Artificial Synapses Towards Intelligent Human-machine Interface and Neuromorphic Computation Systems[J]. *Materials Reports*, 2020, 34(1): 1022-1049.
- [85] 杨明浩, 陶建华. 多通道人机交互信息融合的智能方法[J]. *中国科学: 信息科学*, 2018, 48(4): 433-448. YANG Ming-hao, TAO Jian-hua. Intelligence Methods of Multi-modal Information Fusion in Human-computer Interaction[J]. *Science in China: Information Science*, 2018, 48(4): 433-448.
- [86] ZHENG N, LIU Z, REN P, et al. Hybrid-augmented Intelligence: Collaboration and Cognition[J]. *Frontiers of Information Technology & Electronic Engineering*, 2017, 18(2): 153-179.