

智能时代人因科学研究的新范式取向及重点

许 为¹ 高在峰² 葛列众¹

(¹ 浙江大学 心理科学研究中心; ² 浙江大学 心理与行为科学系, 杭州 310058)

摘 要 本文首先提出了“人因科学”这一创新的学科群概念来表征工程心理学、人因工程、工效学、人机交互等相近领域。尽管这些领域的研究角度不一样,但是它们分享共同的研究理念、对象以及目的。我们近期的研究表明,人工智能(AI)新技术带来了一系列新的人因问题,而作为人因科学研究对象的人机关系呈现出从“人机交互”向“人智组队式合作”的跨时代演进。这些变化对人因科学研究提出了新问题和新的挑战,需要我们重新审视基于非智能技术的人因科学研究范式和重点。在此背景下,本文梳理人因科学研究范式取向跨时代的演进,总结我们近5年所提出的一系列用于丰富人因科学研究范式的新概念模型和框架,其中包括人智协同认知系统、人智协同认知生态系统以及智能社会技术系统的模型和框架。本文进一步从人因科学研究范式取向的角度进一步提升这些概念模型和框架,提出智能时代人因科学研究的三种新范式取向,分析相应的应用意义,并展望今后的研究方向。同时,针对智能时代人因科学研究重点的跨时代转移新特征,本文从“人智交互”“智能人机界面”“人智组队合作”三个方面展望了今后人因科学的研究重点,揭示出人因科学新研究范式取向对未来研究重点的作用。我们认为,人因科学的研究范式取向和研究重点互为影响,互为促进,智能时代的人因科学研究需要多样化、创新的研究范式取向,从而进一步推动人因科学的发展。

关键词 人因科学, 工程心理学, 人因工程, 研究范式取向, 人智组队

分类号 B849

1 引言

工程心理学(engineering psychology)、人因工程(human factors engineering)和工效学(ergonomics)均诞生于上世纪二次大战期间,虽然它们具有各自独特的研究角度和重点,但是都分享“以人为中心”的共同理念。例如,工程心理学从人类认知信息加工角度为人机系统的优化设计提供心理学原理、方法及实证(Wickens et al., 2021; 孙向红等, 2011; 许为, 朱祖祥, 1989); 人因工程和工效学则从人-机-环境关系匹配的工程设计角度为人机系统的优化设计提供人因和工效设计原则、方法及实证(Sanders & McCormick, 1993; 许为, 陈勇, 2012, 2013, 2014; Xu & Zhu, 1990; Xu, 2007)。进入计算机时代,人与计算机(包括基于计算技术的产品)交互带来了许多新的人因(human factors)问题,推动

了人机交互(human-computer interaction, HCI)、用户体验等领域的产生和发展(Norman & Kirakowski, 2017; Norman & Draper, 1986; Xu et al., 1999; 许为, 2003, 2005, 2017)。基于“以人为中心”的共同理念,这些相近领域都希望通过优化人、机器以及环境之间的交互,确保系统实现安全、高效和宜人的目标,因此,我们将这些领域统称为人因科学(human factors science)。

本文在国际上首次提出了“人因科学”这一创新的学科群概念。该概念从本质上凸显出工程心理学、人因工程、工效学、人机交互、用户体验等相近领域围绕“人”这一要素在各自相应研究和应用中所采纳的共同研究理念、对象以及目的。类似于认知科学与认知心理学的关系,人因科学的概念是比这些领域更高一级的表达。在实践中遇到的许多研究和应用问题有时不能简单地归属于这些相近

领域中的某一个特定领域,通常需要从跨领域的角度来探索解决问题的整体方案。因此,“人因科学”学科群概念的提出有助于从更高的系统层面上来探索人因科学的系统化整体解决方案,促进这些相近领域之间的交流与合作、取长补短、协同发展,从而实现共同的领域目标。

进入智能时代,基于人工智能(AI)技术的智能自主化系统带来了一系列新特征和新问题(Kaber, 2018; 许为, 2020),这必将对人因科学研究以及相应的研究范式取向带来新的要求,需要我们系统的梳理,从而保证人因科学能够为智能新技术的优化设计做出更加有效的学科贡献。本文将一门领域的研究范式取向定义为该领域开展研究所采纳的角度和层面,一个研究范式取向决定了该研究范围、重点以及相应的方法。从人因科学的整体角度出发,本文对人因科学中相近领域的研究范式取向与研究重点展开全面系统的评估。

近 5 年来,针对智能时代的人因科学研究理念和对象等问题,我们提出了一系列新论述、概念模型和框架,例如“以人为中心 AI”理念,基于“人智组队”(human-AI teaming)式合作的新型人机关系,“人智交互(human-AI interaction)”的新型跨学科研究领域。在这些研究基础上,针对智能时代人因科学的研究范式取向,我们开展了一些新探索,并且提出了一系列新论述和新概念框架(Xu et al., 2019; Xu, 2019, 2021; 许为, 葛列众, 2018, 2020; 许为等, 2021; 许为, 2019a, 2022, 2022a, 2022b, 2022c; Xu et al., 2022; Xu & Dainoff, 2023; Ozmen Garibay et al., 2023; Xu & Gao, 2023; Gao et al., 2023)。这些人因科学研究新理念、新论述和新概念框架必然带来对人因科学研究范式取向和重点的新思考,人因科学界目前还没有系统地开展这方面的工作。因此,本文通过梳理人因科学研究范式取向和重点,进一步提升我们所提出的这些人因科学研究新理念、新论述和新概念框架,回答以下科学问题:智能时代人因科学应该采用怎样的研究范式取向来有效地支持今后的研究?智能时代人因科学今后的研究重点是什么?以期为下一步人因科学研究提供新思路。

2 人因科学研究对象和范式取向的跨时代演变

2.1 人因科学研究对象的演变

人因科学的研究对象是人机关系。在计算机时代,在人与非智能计算系统的交互中,机器充当人机系统中辅助工具的角色。在智能时代,人与智能

系统的交互本质上是与智能系统中的智能体(intelligent agent)之间的交互。基于智能技术,这些智能体可以展现出独特的自主化(autonomy)新特征,拥有一些类似人类的认知能力(感知、学习、推理等),在一些设计未预期的场景中,可自主地完成以往自动化技术所不能完成的任务(Kaber, 2018; Madni & Madni, 2018; 许为, 2020)。因此,智能系统可以从一种支持人类操作的辅助工具角色发展成为与人类合作的团队队友,扮演“辅助工具 + 人机合作队友”的双重新角色,从而形成一种新型的人机关系形态:“人智组队”式合作(Brill et al., 2018; 许为, 葛列众, 2020)。这种新型“人智组队”式合作关系给人机关系赋予了新的内涵,带来了人机关系跨时代的演变(许为, 葛列众, 2020),也为人因科学研究带来了新的视角,需要我们重新审视基于非智能技术的人因科学研究范式和重点。

2.2 人因科学研究范式取向的探索

纵观发展历史,借助于新兴技术,人因科学的研究范式取向一直在拓展。这种拓展提升了人因科学研究的方法论,扩展了学科研究的范围和解决问题的深度,进而推动了人因科学的不断发展。

在发展初期,传统人因科学(如早期的工效学)注重人类体力作业、人机界面等物理特征,研究主要借助人类体力作业分析、时间任务分析等手段来实现人机功能和任务的合理分配,达到优化人机系统效率的目的(Sanders & McCormick, 1993; Gardner et al., 1995)。

自从进入计算机时代,人因科学的研究范式基本上是基于信息加工的认知理论构建,但是这种研究范式在人因科学发展各阶段中呈现出不同的取向。在计算机时代,人因科学(如工程心理学、认知工效学)从信息加工机制出发,深入到人的心理活动层面,考察在人机操作环境中人类工作绩效与感知觉、注意、记忆、决策等心理活动间的关系,达到优化人机系统的设计目的(Wickens et al., 2021; 孙向红等, 2011)。为解决人与计算机交互中的人因问题,人因科学(如人机交互、用户体验)采用了基于“以用户为中心”理念的方法,构建用户心理和情景意识模型、人机交互认知模型及人机界面概念模型,采用基于心理学方法的可用性测试等验证方法,开发出符合用户需求和体验的交互式产品(Nielsen, 1994; Finstad et al., 2009; 许为, 2023)。

然而,人因科学研究通常注重在人类外在行为的层面上,通过客观工作绩效和主观评价方法来考

察人的认知活动和工作绩效。为克服这种方式的局限性,人因科学进一步采纳了来自认知神经科学的测量技术方法(如 EEG、fMRI 等),形成了神经人因学(neuroergonomics),从而能够深入到人类认知加工的内部神经活动层面,探索在人机交互环境中认知加工的神经机制(Parasuraman & Rizzo, 2006),为人因科学提供了更为客观的实证手段(Dehais et al., 2020)。

进入智能时代,智能技术的新特征、人因科学研究理念和对象(人机关系)的跨时代演变必然带来对人因科学研究范式取向的新思考。然而,人因科学界还未系统地开展这方面的工作。自 2021 年,我们针对智能时代的人因科学研究范式取向开展了一些新探索,以下分别总结我们提出的人智协同认知系统(human-AI joint cognitive systems)、人智协同认知生态系统(human-AI joint cognitive ecosystem)、智能社会技术系统(intelligent sociotechnical systems)概念模型和框架(许为, 2022a, 2022b, 2022c; Xu et al., 2022; Xu & Dainoff, 2023; Ozmen Garibay et al., 2023; Xu & Gao, 2023; Gao et al., 2023)。同时,本文从人因科学研究范式取向的角度进一步提升这些概念模型和框架,然后通过智能驾驶人机共驾的应用实例来分析这些研究范式取向的应用意义,并且展望今后的研究。

3 智能时代人因科学研究的新范式取向

3.1 人智协同认知系统

3.1.1 人智协同认知系统的概念模型

人机交互模型是人因科学研究的一个重要内容。研究者通常采用了人类信息处理模型来构建人机交互模型,例如 MHP、GOMS、SOAR、ACT-R、EPIC 等模型(综述见:王宏安等, 2020)。这些模型

主要针对非智能人机交互,机器仅仅是作为一个工具来考虑,没有考虑人与智能系统之间潜在的人机合作关系,无法有效处理智能人机交互任务(许为,葛列众, 2020)。目前,国际学术界就人一智能系统之间存在的合作关系已基本达成共识(例如, NASEM, 2021; Caldwell et al., 2022)。然而,迄今尚未有人智组队合作的交互模型。鉴于此,基于智能技术的自主化新特征以及新型人智组队式合作关系,采用协同认知系统理论(joint cognitive systems; Hollnagel & Woods, 2005)、情景意识(situation awareness)理论(Endsley, 1995)以及智能体理论(Wooldridge & Jennings, 1995; 王祝萍,张皓, 2020),我们首次提出了一个针对智能人机系统中人智组队合作的人智协同认知系统概念模型(许为, 2022b; Xu & Gao, 2023)(见图 1)。

如图 1 所示,不同于传统人机交互系统,该模型将智能系统(包括一个或多个智能体)视作为能够完成一定认知信息加工任务的认知体。因此,一个智能人机系统可以表征为两个认知体协同合作的一个协同认知系统。作为与人类用户合作的团队队友,智能系统通过自然有效的人机交互方式(如语音、手势、表情等)与人类用户开展双向主动式交互和协同合作。在特定的人机交互场景中,智能系统可以对用户状态(认知、生理、意图、情感等)、环境上下文等状态进行自主感知、识别、学习、推理等认知作业,做出相应的自主执行(Kaber, 2018; 许为, 2019b; Xu, 2021)。

该框架采用 Endsley 的情景意识认知理论来表征人类用户和机器认知体的信息加工机制(Endsley, 1995, 2015),即人类操作员感知和理解当前环境状态(包括人机界面、系统、环境、合作队友等)以及预测未来情景状态的认知加工机制。该模型还包括了情景意识与记忆、经验、知识等因素的认知交互,

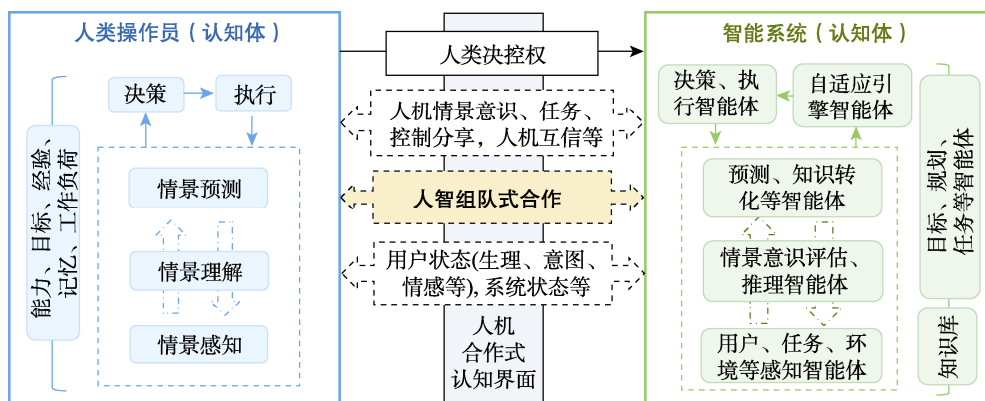


图 1 一个表征人智协同认知系统的概念框架

并且拥有数据驱动(根据感知数据来理解和预测情景)和目标驱动(根据目标以及当前的理解和预测来验证感知数据)的信息加工机制。借助于一个信息收集和后期响应的动态反馈和前馈回路机制,人类用户能够感知动态的环境情景来更新获取的信息。如图 1 所示(右侧),该模型采用与人类用户认知体异质同构方式来表征机器认知体的信息加工机制。

图 1 所表征的概念框架为人因科学研究提供了一种新的研究范式取向,该研究范式取向体现了以下一些新特征和研究新思路:

(1)基于机器认知体的研究范式新取向:不同于传统人因科学研究范式将机器视作为辅助人类作业的工具,该模型将机器智能体表征为与人类组队合作的认知体,这有助于人因科学通过研究智能体的认知行为以及与合作行为,探索通过优化智能体认知能力和行为的途径来提升人机系统绩效。

(2)基于“人智组队”的合作式研究范式新取向:不同于传统人因科学研究范式的“人机交互”式人机关系,该模型将人机关系表征为一个人智协同认知系统中人-机两个认知体之间的协同合作,探索通过优化人机协同合作的途径来提升人机系统绩效。

(3)“以人为中心 AI”的理念:在人智组队合作的团队中,人类用户是这一合作团队的领导者,在应急状态下人类是系统的最终决控者。

(4)人机双向主动式状态识别:不同于传统的“刺激-反应”单向式人机交互,该模型强调人机双向主动式状态识别,智能系统可以通过感应系统主动监测和识别用户生理、认知、行为、意图、情感等状态,人类用户则通过多模态人机界面获取最佳的情景意识。

(5)人机智能互补性:作为一个协同认知系统,系统绩效不仅仅取决于系统单个部分的绩效,而是取决于人机智能互补和合作,通过人机混合(融合)智能来最大限度地提高人机协同合作和整体系统绩效。

(6)自适应智能人机交互:强调智能系统的自适应机制,根据对用户、环境上下文等状态的感应识别和推理,智能体可以在设计无法预测的一些场景中做出合适的自适应系统输出,而人类用户根据情景意识、任务、目标等自适应地调整交互行为。

(7)人机合作式认知界面:强调构建基于多模态交互技术的人机合作式认知界面来支持人机协同合作,其中包括对双向情景意识、人机互信、

人机决策共享、人机控制共享、人机社会交互、人机情感交互等方面的支持。

3.1.2 应用分析

基于智能技术的自动驾驶车是一个典型的智能人机系统。目前,尽管人因科学专业人员参与了自动驾驶车的研发,但是频频发生的事故促使我们探索设计新思路(NTSB, 2017; Endsley, 2018; 许为, 2020)。SAE (2019)将自动驾驶车系统分为 5 个“自动化”等级(L1~L5),在相当长一段时间内,自动驾驶车人机共驾将是常态(宗长富 等, 2021)。在高等级自动驾驶车系统中,基于人智协同认知系统的研究范式取向,人类驾驶员和车载智能系统是可以完成一定认知信息加工任务的两个认知体,自动驾驶车人机共驾就是一个基于人智组队合作的人智协同认知系统(许为, 2020; Xu, 2021; Gao et al., 2023)。装备智能感知等技术的车载智能系统(智能体)可对人类驾驶员状态、环境等状态进行一定程度的感知、识别、学习、推理等认知作业,与人类驾驶员开展有效的人机共驾。因此,人机共驾解决方案可以从几方面来探索:

(1)人机协同合作的设计新范式:目前基于“以技术为中心”理念的自动驾驶级别分类法强调了智能驾驶车作为辅助或取代人类驾驶的工具角色(SAE, 2019)。人智协同认知系统倡导以人为中心和人机合作的设计思维,这有助于在各种自动驾驶任务场景中为双方建立协同合作式关系和动态化人机功能分配,从人机协同合作、人机互信、情景意识共享、人机控制共享和协同驾驶等角度来优化车载人机系统设计(Biondi et al., 2019; 许为, 2020; 高在峰 等, 2021)。例如,研究应急状态下的车辆控制权人机切换,确保人拥有最终控制权(包括远程控制)(Fridman, 2018);探索在什么条件下(如人机互信程度,驾驶员状态和行车意图)完成有效的人机控制权切换。

(2)人机状态双向识别:强调人机两个认知体之间的双向主动式状态识别,基于智能体对人类驾驶员状态、驾驶行为、意图的监测理解以及人类驾驶员对系统和环境的情景意识,实现有效的人机协同合作。例如,如何建模和实现驾驶员操作意图识别,如何提高对驾驶员状态监测的准确性,如何通过解释性 AI 来提升操作员情景意识。

(3)基于“人在环内”的人机混合智能设计:“人在环外”的人机系统设计是导致事故的重要原因之一(NTSB, 2017; Endsley, 2018)。人机共驾应是一个

“人在环内”的人机混合智能系统。研究在系统层面(如“人在环内”控制)和生物层面(如脑机控制)的有效“人在环内”设计;如何采用人机智能互补来优化系统设计;落实“有意义的人类控制”(meaningful human control)设计以及如何采用车载“故障追踪系统”来实现人机故障问责和设计改善的机制(Santoni de Sio & van den Hoven, 2018)。

(4)基于智能自主化的设计:作为一个“移动式”智能自主化系统,高等级自动驾驶车系统所具备的自主化特征对系统设计提出了不同于传统自动化技术的新需求,例如,潜在的不确定系统输出、可解释人机界面、人机控制权分享(许为, 2020)。系统设计需要采用有效方法来满足这些新需求,提高行驶安全。

(5)合作式认知界面设计:探索“合作式认知界面”来支持人机共驾中的人机协同合作,探索有效的人机界面设计隐喻、范式及认知架构,例如,合作式认知界面、车载生态用户界面(Burns & Hajdukiewicz, 2004)。

3.1.3 研究展望

为丰富人智协同认知系统的研究范式取向,今后研究要考虑以下几方面:(1)开展针对机器智能体认知特征、认知能力、行为演变、人-智比较研究等方面的研究(Zhu et al., 2020);研究人因科学如何参与对机器智能体的优化设计,提升人智组队的协同合作。(2)开展针对人智组队合作的理论研究,借鉴现有人-人团队合作等理论,探索人智组队合作理论、方法、人机合作绩效评估系统(Madhavan & Wiegmann, 2007; Kaber, 2018)。(3)构建人机协同合作的人机分享式/团队式情景意识、人机互信、人机决策和控制分享等模型(许为, 2020; Gao et al., 2023)。(4)开发有效的用户状态、行为、意图、情感识别认知模型,研究有效的知识表征和知识图谱来支持计算建模。(5)研究基于用户、系统、环境状态评估的自适应优化设计,利用智能系统主动式前馈预测能力来协助人类团队队友,实现主动式、自适应的人机交互;(6)研究有效支持人机协同合作的合作式认知界面设计新范式和模型。

3.2 人智协同认知生态系统

3.2.1 人智协同认知生态系统的概念框架

人因科学研究通常注重于单一人机系统,人智协同认知系统也主要针对人与单智能体系统之间的协同合作。但是,智能不仅仅是一个产品或系统,而是一个包括了技术变革、系统演变、运行方式创

新和组织适应等特征的跨智能系统(产品)的生态体系(刘伟等, 2023)。例如,基于单车自动驾驶的人机互驾仅仅是一个人智协同认知系统,整个人机共驾智能生态系统包括基于智能车联网、智能交通系统等技术与车、车与车、车与智能交通环境之间的交互和协同合作,这些多智能体系统构成了一个人智协同认知生态系统。这些多重人智协同认知系统间的相互作用和协同合作将直接影响单车人机互驾的行驶安全。因此,人因科学研究范式需要突破面向单一人智系统的研究范式取向,从生态系统化的研究范式取向来整体考虑智能协同认知生态系统的系统解决方案。

目前,针对多智能体系统的研究目前主要侧重在工程技术方面(Dorri et al., 2018; Allenby, 2021; Ali et al., 2021),如分布式多智能体系统和人机物融合群智计算等(谢磊, 谢幸, 2021; 郭斌, 於志文, 2021),但是目前尚未见从人因科学角度出发来考虑系统设计和构建的研究。我们采用生态系统的研究范式取向,借鉴协同认知系统理论(Hollnagel & Woods, 2005)、多智能体系统理论(Dorri et al., 2018; Le Page & Bousquet, 2004)、多智能体生态系统思维(IDC, 2020; Allenby, 2021; Ali et al., 2021),首次将一个智能生态系统表征为一个人智协同认知生态系统(见图2)(许为, 2022a)。如图2所示,一个多智能体生态系统可表征为一个协同认知生态系统。一系列人智协同认知生态系统形成了一个个庞大的基于多重智能体系统的人智协同认知生态系统(如智慧城市、智能交通)。

图2所表征的概念框架从生态系统化的角度为人因科学提供了一种新的研究范式取向,该研究范式取向体现了以下一些新特征和人因科学研究新思路:

(1)基于生态系统的研究范式新取向:一个多智能体生态系统可表征为一个人智协同认知生态系统,它由一系列相互作用的人智协同认知系统组成,智能生态系统的整体优化设计需要考虑这些人智协同认知系统之间的相互作用和协同合作。一个人智协同认知生态系统的系统绩效和安全取决于各个人智协同认知生态系统的整体优化设计。这种新取向体现了人因科学研究范式取向从“点”(单一人机系统)到“面”(跨人机系统)的扩展。

(2)“以人为中心 AI”的理念:人类在跨智能系统的智能生态体系中占有中心位置,智能生态系统设计、开发和实施必须将人类需求、价值、智慧、

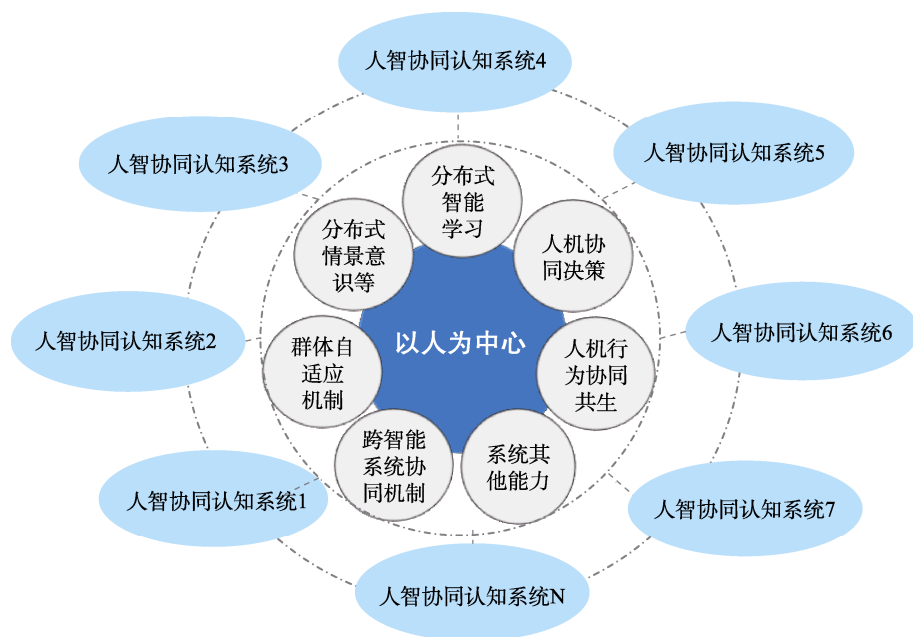


图2 人智协同认知生态系统的概念框架

能力和作用等置于首位。例如，构建跨智能系统的人智决策权限设置等保障系统，构建跨智能系统的人智信任生态体系、跨智能系统(基于不同开发商、文化、社会、文化、伦理背景等)之间的冲突解决机制，保证人类是整个人智生态系统运营的最终决策者，保障智能生态系统的整体安全。

(3)“分布式”协同合作：基于生态系统的范式取向，通过有效的网络分布式(跨智能系统)人智交互和协同合作式系统设计来实现整体生态系统的最佳绩效。这种分布式系统的设计需要考虑跨智能系统的认知增强学习、情景意识、人智情感交互、人智互信、人智信息加工、人智认知学习、人智协同决策、人智控制共享、人智社会交互等一系列具备分布式和分享式特征的协同合作，提升人智生态系统的整体协同合作能力。

(4)系统学习和演进：系统设计要注重人智系统学习和演进。基于人类和智能体系统的学习和进化特征，强调通过分布式认知增强学习、跨实体和跨任务的群体智能知识迁移、自组织与自适应协同能力，实现人智系统的持续演化和优化，从而能够组织各要素来适应动态化的复杂应用场景。同时，人智协同认知生态系统强调基于人机智能差异性和互补性，通过人智协同增强学习、群智融合、人智协同学习等方法和技术，提升人智协同认知生态系统的整体学习、进化和协同合作能力。

3.2.2 应用分析

人智协同认知生态系统概念框架已被用来初

步分析自动驾驶人机共驾和大型商用飞机驾驶舱单人飞行操作的应用场景(许为, 2022a; 许为等, 2021)。以自动驾驶人机共驾为例，自动驾驶人机共驾的安全实现不仅仅依赖单车层面的系统设计，还需要通过实现人、车、路、智能交通环境中各种认知体(智能体)之间的协同合作，通过车端、路端、云端之间的有效信息交互等途径，从而为整个人机共驾生态系统的安全驾驶、决策和规划提供优化的系统设计(谭征宇 等, 2020)。图3示意了自动驾驶人机共驾的人智协同认知生态系统，人智协作认知生态系统的表征为自动驾驶人机互驾解决方案提供了人因科学研究新思路。

首先，自动驾驶人机共驾模式的系统化设计思路。任何局限于“人类驾驶员 + 智能自动驾驶车”单车层面的设计都无法保证整个协同认知生态系统的优化设计和安全运行。作为一个庞大的人智协同认知生态系统工程，人机共驾生态系统(人-车-路-交通-社会系统)的安全运营取决于该系统内所有人智协同认知子系统之间的有效协同合作和整体优化设计。

其次，只有在人智协同认知生态系统层面上实现“以人为中心 AI”理念，才能保证人类拥有自动驾驶车的最终决策权。例如，当车辆处于失控状态(如系统故障、黑客攻击、驾驶员失能)，一方面，车载智能系统启动智能应急方案来脱离当前的失控场景，保护人类(包括驾驶员、乘客、道路行人和其他车辆等)；另一方面，作为人智协同认知生态系

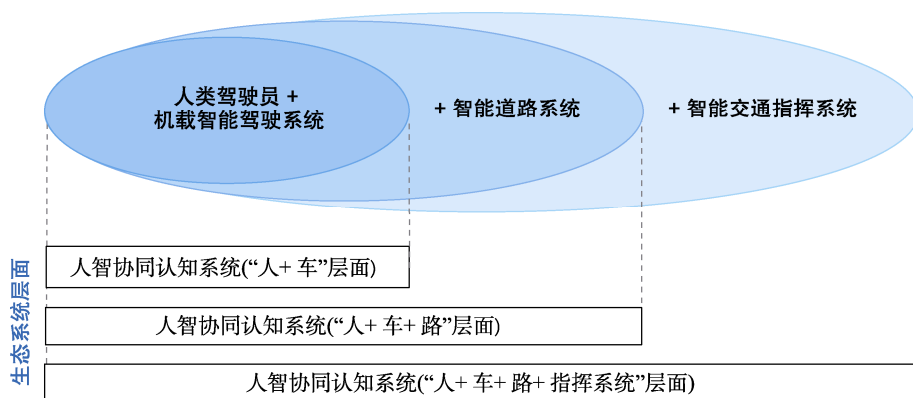


图3 自动驾驶人车共驾的人智协同认知生态系统示意图

统的冗余化系统安全方案之一,城市智能交通指挥中心或者智能车运营商控制中心的操作员能够远程接管故障车(如“5G云代驾”),指挥协调路面其他智能驾驶车,保证智能道路系统的整体安全。该生态系统设计必须考虑人的能力,为各方人员提供有效的情景意识(人、车载系统、驾驶环境等状态),实现驾驶权限和责任的实时分配,保证应急状态下人类拥有对人机共驾生态系统的最终决控权。

第三,在人智协同认知生态系统范围内实现自动驾驶人机共驾模式的系统安全和优化设计。从生态系统角度讲,系统设计需要保证整个人机共驾生态系统中各子系统的共同学习、共同演化、系统自适应等能力的实现。从协同认知角度讲,系统设计需要保证整个人机共驾生态系统中人-人、人-智、跨智能体以及跨子系统之间的有效交互,建立有效的兼容和冲突协调机制(技术、交通规范等),实现整个生态系统内的有效协同合作。

最后,从人智协同认知生态系统角度制定实现自动驾驶人机共驾系统的策略。图3所示的生态系统架构示意了以下可能的发展路径:(1)“自下而上”路径,即从“人类驾驶员 + 智能自动驾驶车”到“智能道路系统”以及“智能交通指挥系统”;(2)“自上而下”路径,即从“智能交通系统”到“智能道路系统”以及“人类驾驶员 + 智能自动驾驶车”;(3)混合或并行路径,即基于“自下而上”与“自上而下”路径的组合式并行发展路径,这可能是最安全、最有效的发展路径。

3.2.3 研究展望

丰富人智协同认知生态系统概念框架的工作需要跨学科的合作,今后研究应该首先考虑解决以下一些基本问题。(1)现有针对单一人机系统的人因科学研究主要采用实证的研究手段,对基于生态系统研究取向的多人机系统研究,需要探索支持该范

式取向的方法论,研究一个协同认知生态系统的基本元素、特征、结构、发展和演化过程等。(2)基本理论研究。例如,人机行为协同共生理论,即如何根据人机环境变化、人机感知节点能力差异性以及互补性,借助生物群体的演化、自组织和自适应机制来发掘人-智自适应、自组织、演化等能力的理论(Rahwan et al., 2019; Werfel et al., 2014; 郭斌, 於志文, 2021);群体分布式环境下智能体学习演化理论,即基于人-智群体的融合协作、竞争等特征的群体智能体分布式学习模型,群体智能体知识迁移方法(郭斌, 於志文, 2021; Neftci & Averbeck, 2019)。(3)跨人-多智能体系统的协同合作问题。多智能体人机交互和协同机制、绩效评估系统;人与多智能体间的协调、任务分配、团队建设;多智能体间的情景意识、可信度、任务分享、合作、决策控权分享(Le Page & Bousquet, 2004; Dorri et al., 2018);基于不同文化和规范的跨智能体系统交互中的兼容性、沟通、模式转换以及人类用户的协调作用(ISO, 2020)。(4)应用方法。探索能有效影响智能生态系统设计的人因科学流程和方法。

3.3 智能社会技术系统

3.3.1 智能社会技术系统的概念框架

社会技术系统(sociotechnical systems, STS)理论提倡社会、技术、组织等子系统之间的协调,从而实现最佳的系统绩效(Eason, 2008)。人因科学研究通常注重于人机界面以及物理环境对人机系统绩效的影响,不注重考虑宏观社会、组织环境等因素(葛列众, 许为, 宋晓蕾, 2022)。在过去20年中,STS理论影响了人因科学领域,推动了宏观工效学(Macroergonomics)等领域的发展(Watson et al., 2015)。由于智能系统需要在一定的STS环境中运行,且智能技术可能对人类产生负面影响(如用户隐私、伦理、决策权、技能成长等),这种负面影响

促使人因科学在 STS 的大环境中考虑智能系统的研发和使用(Stahl, 2021), 开展针对智能系统的研究(Asatiani et al., 2021)。例如, Steghofer et al (2017)认为下一代 STS 应该以智能技术为基础, STS 的社会子系统受智能技术、AI 潜在的决策不确定性等因素的影响; 从 STS 开发流程和多学科合作角度, 研究者提出了参与式设计流程(Huang et al., 2019)、社会技术系统工程(STSE) 框架(Baxter & Sommerville, 2011)、自适应 STS 系统架构(Dalpiaz et al., 2013)、人因科学方法(Watson et al., 2015; Hollnagel et al., 2006)等。许为等(2021)提出的“人智交互”新型领域也是从宏观的 STS 环境考虑。

目前还没有针对智能技术的成熟 STS 理论, 人因科学研究应该探讨如何在复杂 STS 环境中有效地研发和使用智能系统。为此, 许为(2022c)分析了智能时代 STS 的新特征, 这些新特征包括系统组成、认知代理、人机关系、用户需求、系统决策和控制权、系统学习能力、系统设计范围、组织目标和需求、系统复杂性和开放性等方面。基于这些新特征, 我们在国际上初步提出了一个智能社会技术系统(intelligent sociotechnical systems, iSTS)概念框架(许为, 2022c) (见图 4)。

如图 4 所示, iSTS 继承了传统 STS 理论的一些基本特征。例如, iSTS 内部环境有独立但相互依赖的技术和社会子系统, 整体系统绩效依赖于两个子系统之间的协同优化(Badham et al., 2000); iSTS 拥有一个宏观的外部环境以及各类智能社会形态; 相对于人智协同认知生态系统, iSTS 更加注重于宏观和非技术因素, 其中包括工作系统重新设计、组织

决策与智能决策等。

图 4 所表征的概念框架从社会技术系统的视角为人因科学提供了一种新的研究范式取向, 该研究范式取向体现了以下一些新特征和人因科学研究新思路:

(1)基于 STS 的研究范式新取向: 任何一个智能人机系统(人智协同认知系统)和智能生态系统(人智协同认知生态系统)的开发以及使用都在一个智能 STS 环境中发生, 智能系统的优化设计和有效使用都需要考虑技术与非技术子系统(如社会、组织)之间的相互作用。这种新取向体现了人因科学研究范式从“点”(单一人机系统)到“面”(跨人机系统)再到 STS 宏观环境的进一步扩展。

(2)“以人为中心 AI”的理念: 强调从人的需求、价值、智慧、能力和作用出发, 充分考虑社会、文化、伦理等宏观环境因素对智能系统开发和使用的影 响, 强调需要解决 iSTS 中独特的人机合作、AI 伦理道德等新问题。在系统开发中采用用户参与等以人为中心的人因科学方法, 保证智能技术能够有效辅助人类和组织决策, 确保人类拥有最终决控权(Herrmann et al., 2018; Baxter & Sommerville, 2011; Xu, 2019)。

(3)基于人智组队的人机团队协作: 图 4 中社会子系统与智能技术子系统两个圆圈重叠的部分示意了两子系统之间的协同合作关系, 这是 iSTS 区别于传统 STS 所特有的新型人机关系(许为, 葛列众, 2020)。强调在 iSTS 的社会和组织的语境中, 智能系统不仅仅是传统 STS 中支持人类作业和提升组织生产力的简单工具, 也是人智团队协同合作的



图 4 智能社会技术系统(iSTS)的概念框架

成员。iSTS 中相互依赖的人智团队是分享共同社会和组织目标的团体(Salas et al., 2008)。人类与智能系统(机器代理)之间的团队式交互合作、团队式互信、团队式信息和决策分享等是智能系统成功开发的重要因素之一。

(4)组织适应和再设计。实施 AI 新技术会改变既定的工作系统,甚至可能导致一些用户陷入困境。iSTS 强调将组织中的人类和智能系统作为一个新型的工作系统重新设计(包括人员、角色、流程、技术、治理等),根据人与智能技术之间的优势互补,调整优化人机之间的功能和任务分配,其中包括人机角色分配、工作流程、作业环境等,制定有效的人智团队的工作和任务,保证人机功能的最佳分配,提升整体系统绩效;在引进智能技术的同时,要充分考虑组织内员工的岗位重新分配、公平、满意感、决策参与感、技能成长等问题,提升人智系统的整体绩效。

(5)人智共同学习。iSTS 中的技术子系统(硬件、软件、机器代理等)和人类代理(跨个人、组织和社会层面)之间存在复杂的相互作用,跨越了传统人和机器的物理界限。智能体(代理)是促进 iSTS 中社会和技术子系统之间交互的新资源,这种交互会调整智能代理自身的行为(基于机器学习算法等),也导致人类使用和期望模式的改变(社会学习);同时,iSTS 的社会和技术子系统包含人类、智能机器代理在不同层面上的自主权,这些都体现出人智共同学习、共同成长、灵活性、自适应等能力特征。因此,iSTS 强调智能系统的开发需要通过有效的设计和治理,促使人智共同学习,从而提高人智系统的整体能力(Heydari et al., 2019)。

(6)开放式生态系统。在非智能时代,传统 STS 中的分析单元通常是拥有相对独立边界的组织;在智能时代,物联网智能网、智能城市、智能交通等智能生态系统存在于复杂的、相互依存的 iSTS 中,智能技术的发展将极大地增加智能机器代理的数目,可导致智能系统中潜在不确定性和不可预测性的增加,人智共同学习和智能系统中潜在的输出不确定性等自主化特征会带来动态和模糊的 iSTS 边界(van de Poel, 2020)。这些新特征既给智能系统开发带来了创新设计的机遇,也给系统设计、系统规则、伦理道德、文化价值观等方面带来了挑战(Hodgson et al., 2013)。因此,智能系统的设计开发需要从一个开放式的人类、技术、社会、组织生态系统的角度考虑。

3.3.2 应用分析

从 iSTS 角度看,自动驾驶人机共驾研发不仅是一个工程技术项目,还需从 iSTS 角度考虑许多非技术因素的影响。例如,目前公众对自动驾驶车的信任度并不高(Lee & Kolodge, 2018),自动驾驶伦理研究主要集中在单车层面(Borenstein et al., 2019)。针对自动驾驶人机共驾的 iSTS 研究需要考虑智能自主化技术、设计标准、道路交通基础设施、政策法规、伦理化设计、交通规则、物联网/5G、企业投资回报、大众信任度和接受度、驾驶员技能以及自动驾驶车技术认证等一系列因素。例如,自动驾驶车企业以及它如何与其他 iSTS 成员之间的协同合作(包括制造商、监管机构、执法人员、消费者等);事故风险预先分配方案的社会共识(如紧急下是优先保护乘客还是行人)(陈磊等, 2021);自动驾驶与现行交通法规的协调;如何将基于“有意义的人类控制”(meaningful human control)方法落实在系统设计中,保证操作员拥有足够情景意识,能够对智能系统做出有意识的、符合伦理规范的决策控制(Santoni, et al., 2018)。

可见,自动驾驶人机共驾研究需要从“人智协同认知系统”和“人智协同认知生态系统”的范式取向扩展到“智能社会技术系统”的范式取向,并且需要多学科及跨部门的合作,从而研发出安全的、技术可行的、大众信任和接受的自动驾驶车。

3.3.3 研究展望

为丰富 iSTS 概念框架,今后研究需要考虑两个方面。从方法论来说,iSTS 框架强调跨学科的团队合作,基于 iSTS 的智能系统研发可以利用人因科学迭代式原型化和用户测试等方法来逐步优化设计和构建(Norman & Stappers, 2015)。许为(2022c)根据 iSTS 在系统研发各阶段中能够提供贡献的程度,建议了一系列方法,例如生态研究方法(Brown et al., 2017)、情境设计(Beyer & Holtzblatt, 1999)、人类学工作场所分析(Hughes et al., 1992)以及纵向研究(Lieberman, 2009),这些方法需要在今后研究中进一步优化和充实。另外,iSTS 研究也需要创新方法。

从研究思路上说,iSTS 研究可从以下几方面考虑。首先,在智能系统研发中分享“以人为中心 AI”的理念,这将有助于解决更多的人类和社会实际问题。其次,需要充实 iSTS 理论。例如,社会和技术子系统之间如何协同演化?人与智能体交互将如何影响社会和组织环境中的人类行为、组织变

革、组织学习、组织认知等? 如何有效地开展 iSTS 集成设计和治理? 最后, 开展基于 iSTS 的人因科学应用。例如, 如何将 iSTS 概念有效地应用在智能系统研发中? iSTS 如何为伦理化 AI 治理做出贡献 (Chopra & Singh, 2018; Fiore, 2020)?

3.4 人因科学研究范式取向的跨时代演进

基于上述分析, 表 1 进一步概括了在不同的技术时代, 人因科学研究范式取向的演变。由表 1 可知: (1)从传统人机系统理论出发, 人因科学的研究范式取向随技术时代的发展逐步向微观和宏观方向拓展, 技术推动了人因科学研究范式取向的拓展, 例如, 计算机时代的认知信息加工理论, 基于脑活动技术的人类认知神经研究。(2)人因科学研究范式取向具有跨时代的生命力, 丰富了人因科学研究。例如, 机械时代推出的传统“基于人机能力差异性和互补性的静态化人机功能分配”范式取向仍然对智能时代的人因科学研究具有意义, 有助于通过对人机智能差异性和互补性的研究来实现系统设计中的动态化人机功能分配及人机混合智能。(3)研究范式取向呈现出跨学科的特征, 体现了作为交叉学科的人因科学本身发展的需求(如认知神经、AI 技

术)。(4)借助于研究范式取向的扩展, 人因科学的研究重点随时代发展不断地解决新兴技术所带来的新因新问题(如计算机、智能技术的人因问题)。因此, 人因科学研究范式取向的跨时代演进促进了人因科学的发展, 而人因科学研究本身也需要多样化的研究范式取向。

进一步地, 在智能时代, 为解决智能技术的人因新问题, 我们提出了人智协同认知系统、人智协同认知生态系统、智能社会技术系统三种新研究范式取向, 图 5 示意了这三种人因科学研究范式新取向之间的关系。从表 1 和图 5 可知, 这三种新研究范式取向从本质上讲仍然遵循人类认知信息加工这一人因科学的基本研究范式。类似于人类行为研究中的个体、群体、社会的研究范式取向, 针对智能时代人机关系(即人因科学的研究对象)的研究, 我们提出的这三种新范式取向体现了人因科学研究从“点”(单一人智协同认知系统内的人智组队合作)到“面”(跨多重人智协同认知系统的人智协同认知生态系统)再到“体”(跨多重协同认知生态系统的宏观社会技术系统环境)的新扩展, 这种扩展反映了智能时代对人因科学的新要求, 敦促智能时

表 1 人因科学研究范式取向的跨时代演变

	产生时代	研究范式取向	范式取向描述	代表性领域或框架	代表性方法
	计算机、智能时代	基于人类认知神经活动	在神经层面了解人机环境中认知加工的神经机制与工作绩效之间的关系	神经人因学	脑机接口技术和设计, 脑电测量、特征分析和建模
微观	计算机时代	基于人类认知信息加工活动	从人类心理活动(感知、记忆、认知负荷等)层面了解人机环境中认知加工与工作绩效间的关系, 优化人机系统设计	工程心理学	在人机操作环境中, 采用工作绩效测量(反应时、错误率等)和主观评价方法来评价人类心理活动与绩效间的关系、人机系统设计的有效性
	机械时代	基于人机能力差异性和互补性	人与机器能力的差异化及互补性, 优化人机功能和任务分配, 人适应于机器	早期的工效学、人因工程	人类体力作业分析, 时间动作分析, 人机功能和任务分析及分配等
	计算机时代	基于机器作为辅助人类作业(工具)的人机交互	通过人机交互技术、设计、测试和实现达到机器适应于人、人机交互优化、最佳用户体验	人机交互	用户心理模型和需求研究及分析、人机交互认知建模和界面概念化、基于心理学方法的可用性测试
	智能时代	基于人和智能体两个认知体之间的协同合作(单一人机系统)	机器智能体成为与人类合作的团队队友, 人与智能体是协同认知系统中的两个认知体, 通过组队合作达到最佳整体系统绩效	基于人智组队合作的人智协同认知系统	基于人-人团队合作、协同认知系统等理论, 建模人机双向式情景意识、心理模型、信任、决策, 实现人机协同合作的人机交互技术和设计
	智能时代	基于跨人-智系统之间的协同合作(多重协同认知系统)	多智能体系统(协同认知系统)形成智能协同认知生态系统, 整体系统绩效取决于各协同认知系统间的协同合作和优化设计	人智协同认知生态系统	基于生态系统的建模、设计、技术, 包括跨多智能体系统间的协同合作、多智能体系统间的群体知识迁移、自组织与自适应协同、分布式情景意识、人机交互和协同决策等
宏观	智能时代	基于社会和技术子系统间的协同合作	通过实现人、组织、社会、技术子系统之间的相互作用和协同优化, 达到最佳的整体系统绩效	智能社会技术系统	系统化方法, 社会技术系统方法, 工作系统重新设计, 组织优化设计, 工程、社会行为科学等跨学科方法

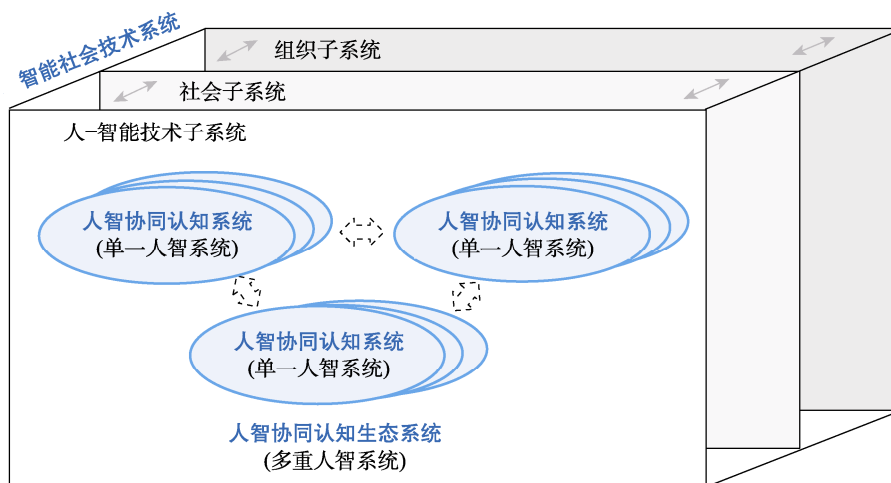


图5 人因科学研究三个新范式取向之间的关系

代的人因科学解决方案不应该仅限于孤立式的、基于“点”方案的传统研究范式取向,而应该为人类社会提供全方位的系统化解决方案。

4 智能时代人因科学研究的重点

如同研究范式取向的跨时代演变,进入智能时代,人因科学研究理念和对象的跨时代演变也带来了人因科学研究重点跨时代转移的新特征。基于文献综述及我们近几年的工作(例如, Xu, 2019, 2021; 许为, 2022, 2022a, 2022b, 2022c; 许为等, 2021; Xu et al., 2022; Ozmen Garibay et al., 许为, 2023; Xu & Gao, 2023),针对人因科学研究重点跨时代转移的新特征,本文从以下三方面展望今后人因科学的研究重点。

4.1 从“人-计算机交互”到“人智交互”

“人-计算机交互”(人机交互)是PC时代形成的跨学科领域,它研究人-非AI机器系统之间的交互。智能时代的机器过渡到基于AI技术的智能系统。如表2所例,AI技术带来了一系列新特征以及相应的人因新问题。针对这些人因新问题的研究已经展开,例如,人-智能系统交互(Brill et al., 2018)、人-智能体交互(human-agent interaction) (Prada & Paiva, 2014)、人-自主化交互(human-autonomy interaction) (Cummings & Clare, 2015)、人-AI交互(Amershi et al., 2019)。尽管这些研究各有侧重点,但是它们都是研究人与智能“机器”(智能体)之间的交互,这种交互的本质就是人-AI交互(简称人智交互),为此,许为、葛列众和高在峰(2021)提倡将人智交互作为一个新型跨学科研究和应用领域来推动。围绕智能系统跨时代转移的新特征和相应

的人因新问题,表2概括了人因科学今后在“人智交互”领域的研究重点。

由表2可见,与人-计算机(非智能系统)交互相比,人智交互研究已经远远超出了现有人机交互的研究范围。针对人智交互的研究也为人因科学研究提供了一个跨学科合作的平台,有助于拓展研究范式,与其他学科共同开发“以人为中心”的智能系统。

4.2 从“传统人机界面”到“智能人机界面”

人机界面是人因科学研究的重点之一。在基于非智能技术的传统人机界面研究中(如用户图形界面),基于“刺激-反应”式、精准输入等人机交互方式,人因科学通过有效的人机界面设计来达到易学、易用的用户体验(Farooq & Grudin, 2016; 许为, 2022a)。在智能时代,“传统人机界面”过渡到基于智能技术的“智能人机界面”,带来了跨技术时代转移的人机交互新特征。例如,从“单向式”到“人机合作双向式”界面,从“可用性”到“可解释AI”界面,从“简单属性”到“情境化”交互界面,从“精准输入式”到“模糊推理式”交互界面,从“交互式”到“合作式”认知界面(见表3)。人因科学需要解决的核心问题是通过这些新型的人机交互途径,解决相应的人因新问题,从而为用户提供更加自然有效、可解释和可理解的用户体验。围绕这些跨时代新特征和人因新问题,针对人智组队的研究开始展开,表3概括了人因科学今后在“智能人机界面”领域的研究重点。

4.3 从“人机交互”到“人智组队”

基于非智能技术的机器通过单向式、非分享的(即只有人针对机器的单方向信任、情境意识、控制等)、非智能互补(即只有人类智能)等方式来实现人机交互;而基于智能自主化技术的人智组队合作

表 2 “人智交互”领域的人因科学研究重点

跨时代转移的新特征	智能技术的人因新问题	人因科学研究重点(部分)
从“可预期”到“潜在不可预期”的机器行为	<ul style="list-style-type: none"> 智能系统可带来不确定机器行为、独特的机器行为演变, 导致系统输出偏差(Rahwan et al., 2019) 现有软件测试方法缺乏对智能机器行为的考虑 智能系统的行为拥有进化、社会交互等特征 	<ul style="list-style-type: none"> 研究机器行为的行为科学方法 在数据收集、训练和算法测试中减小系统输出偏差的迭代式设计和用户测试方法(Amershi et al., 2014) 用户参与式设计、“以人为中心”式机器学习(Kaluarachchi et al., 2021)
从“人机交互”到“人智组队合作”	<ul style="list-style-type: none"> 机器(智能体)从工具到与人类合作的队友 人-智如何合作 如何建模人机协作(人机共信、共享式态势感知、心理模型、决策和控制等) 	<ul style="list-style-type: none"> 基于人智组队合作范式的理论、方法等 人-AI 合作理论、模型、团队绩效评价系统(Bansal et al., 2019)
从“人类智能”到基于“人机智能互补”的人机混合增强智能	<ul style="list-style-type: none"> 机器无法模仿高阶人类认知能力, 孤立地开发机器智能遇到瓶颈效应(Zheng et al., 2017) 无法确保将人类角色整合到智能系统, 从而实现人类可控 AI (Zanzotto, 2019) 	<ul style="list-style-type: none"> 人机混合增强智能的认知架构 基于协同认知生态系统范式的人-多智能体系统协同合作 “人在环中”“人在环上”智能系统及交互设计 人类高阶认知能力模型、知识表征和图谱
从“以人为中心自动化”到“人类可控自主化”	<ul style="list-style-type: none"> 人类可能失去对智能自主系统的最终控制权 自主化技术潜在的负面影响(不确定性等) (Kaber, 2018) 对自动化与自主化概念的混淆会导致低估自主化技术潜在的负面影响 	<ul style="list-style-type: none"> 针对自主化技术的人机交互设计范式 人类可控自主化的人因科学方法 人机共享的自主化控制设计 人-自主化交互
从“非智能”到“智能”人机交互”	<ul style="list-style-type: none"> 如何提升智能用户界面的自然性 如何有效设计智能人机交互(Google PAIR, 2019) 普适计算环境中人类感知能力和认知资源的瓶颈效应(王巍 等, 2014) 	<ul style="list-style-type: none"> 智能人机交互和界面设计新范式 多通道自然式用户界面设计 新型人机交互技术和设计(情感交互、意图识别、脑机接口等) 针对智能技术研发的人因科学设计标准
从“体验需求”到“伦理化 AI”	<ul style="list-style-type: none"> 用户新需求(隐私、道德、公平、技能发展、决策权等) (IEEE, 2019) 智能系统可能的输出偏差和意外结果 滥用智能系统(歧视、隐私等问题) 缺乏对智能系统故障的追溯和问责机制 	<ul style="list-style-type: none"> 基于人因科学方法的伦理化 AI 跨学科设计 基于协同认知生态系统范式的方法 基于智能社会技术系统范式的方法 有意义的人类控制(Santoni & van den Hoven, 2018) 透明化设计
从“体验式”到“系统化”交互设计	<ul style="list-style-type: none"> 基于用户体验、可用性设计方法的局限性 如何有效开展智能系统原型化设计和可用性测试 人因科学人员没能早期介入智能系统开发 	<ul style="list-style-type: none"> 基于人因科学理念的智能系统开发流程 用户体验驱动的智能创新设计 有效的智能交互设计方法(Holmquist, 2017) 系统化人因科学方法(Xu et al., 2019)
从“实体物理”到“虚实融合元宇宙”空间的交互	<ul style="list-style-type: none"> 虚实融合空间的人机交互新需求(史元春, 2021) 元宇宙空间中人机交互的沉浸感、交互性、平行孪生环境的新体验 元宇宙空间中的多模连续、交互数据模糊为交互意图推理带来的新问题 	<ul style="list-style-type: none"> 元宇宙空间的自然人机交互模式和技术 人机交互的虚拟化、远程化和多映射关系 元宇宙交互空间的伦理、信息呈现、脑机融合等 元宇宙交互空间的人-人、人-智间的社会关系

表 3 “智能人机界面”领域的人因科学研究重点

跨时代转移的新特征	智能技术的人因新问题	人因科学研究重点
从“单向式”到“人机合作双向式”界面	<ul style="list-style-type: none"> 系统不再被动地接受用户输入, 并根据固定规则做出预期输出 智能体可主动感知来捕获和理解用户生理、认知、情感、意图等状态, 主动启动人机交互和推送服务 	<ul style="list-style-type: none"> 基于人-智机组队合作范式的人机交互模型 用户情景意识、生理、认知、情感、意图状态的认知模型
从“可用性”到“可解释 AI”界面	<ul style="list-style-type: none"> AI“黑匣子”效应可导致不可解释和不可理解的系统输出(Muelle et al., 2019) AI“黑匣子”效应引发 AI 信任问题 	<ul style="list-style-type: none"> 创新的人机界面技术(如可视化)和设计 “以人为中心”的可解释和可理解 AI (Ehsan et al., 2021) 心理学解释理论的转化(Mueller et al., 2019)

续表

跨时代转移的新特征	智能技术的人因新问题	人因科学研究重点
从“简单属性”到“情境化”交互界面	<ul style="list-style-type: none"> 系统输入除了简单人、机、物等感知属性(如目标位置)以外, 还有“情境化”输入目标(如使用场景上下文、用户行为数据) 	<ul style="list-style-type: none"> 基于操作场景上下文、用户行为等数据, 对“情景化”特征建模、智能推演(如用户行为特征、消费行为画像) 适合用户需求和和使用场景的个性化设计
从“精准输入式”到“模糊推理式”交互界面	<ul style="list-style-type: none"> 用户输入不仅仅是单一精准形式(如键盘、鼠标), 还可是多模态、模糊交互(如用户意图) 应用场景中的模糊交互问题(如随机交互信号和环境噪声) 	<ul style="list-style-type: none"> 从不确定性条件下推理出用户交互意图的方法和模型(易鑫等, 2018) 模糊状态下人机交互的自然性和有效性
从“交互式”到“合作式”认知界面	<ul style="list-style-type: none"> 用户界面既要支持人机交互, 也要支持人智组队合作 支持有效人机合作的人机界面 	<ul style="list-style-type: none"> 人机合作式认知界面的设计范式和模型 基于智能人机交互的界面设计标准 有效支持人机合作的交互设计(如应急场景中的人机控制交接)

可以实现两者之间的“双向合作式”交互, 这种交互表现出两者之间双向主动的、分享的、互补的、自适应等特征(许为 & 葛列众, 2020)。因此, 传统人机交互与人智组队之间有着本质上的差别(Shively

et al., 2018; Brill et al., 2018; Xu, 2021)。目前, 针对人智组队的研究开始展开(NASEM, 2021; Caldwell et al., 2022; Xu & Dainoff, 2022), 表4概括了人因科学今后在人智组队方面的研究重点。

表4 人智组队领域的人因科学研究重点

重点方面	人智组队的新问题	人因科学研究重点
方法与模型	<ul style="list-style-type: none"> 人智组队的研究理论和方法 如何量化预测人智组队的知识结构、界面机制 如何评估人智组队中智能体的作用和绩效(Demir et al., 2018) 复杂动态场景中的团队绩效衡量标准 	<ul style="list-style-type: none"> 人智组队合作理论(如基于人-人团队理论)和方法 人智组队合作绩效的评价系统和预测模型(Kaber, 2018) 智能体在不确定场景中执行预期功能的能力模型
合作流程与能力	<ul style="list-style-type: none"> 人智组队如何长期合作, 分布式团队中的功能分配、目标设置等 智能体如何协调人智组队的协同合作 人智组队所涉及的多样、复杂、动态、自适应协作场景(Goodwin et al., 2018) 	<ul style="list-style-type: none"> 人智组队的技能(团队建设、目标设定、沟通协调、人智合作语言等)(NASEM, 2021) 支持人智组队的有效团队流程 智能体充当合作协调者或团队资源经理的能力(Wesche & Sonderegger, 2019)
情景意识	<ul style="list-style-type: none"> 人智组队合作需要团队和共享式情景意识 跨智能体系统的人智组队合作需要优化的信息集成和方法 人智组队情景意识有可能在应急状态中受到破坏, 事先很难预测(NASEM, 2021) 	<ul style="list-style-type: none"> 团队式、分布式、共享式情景意识(Endsley & Jones, 2012) 智能体自我意识、对人类队友意识与整体团队绩效的关系(NASEM, 2021) 感知、理解、预测人智组队合作状态的情境意识模型
人机信任	<ul style="list-style-type: none"> 需要重新构建人智信任模型和实现方法 需要重新构建信任研究和测试方法 	<ul style="list-style-type: none"> 人智组队合作场景和目标对信任的影响 团队结构和合作中的信任衡量标准 人智共信演化的动态模型
运行操作	<ul style="list-style-type: none"> 人智组队如何在系统功能共享中开展合作 如何实现跨自主化水平的人智组队管理 如何实现跨自主化水平的自适应转换操作 如何实现跨人智团队的动态功能分配和协同操作 	<ul style="list-style-type: none"> 人智组队在任务、功能共享中的协作方法 应急状态下人智组队应对系统自主化变化的方法 人智组队操作中对人类技能保留、培训等方面的要求(Roth et al., 2019) 人智组队与灵活自主化系统操作的关系
人智参与、学习和演化	<ul style="list-style-type: none"> 人智组队所需的先决条件(如共享的信息、知识、技能、能力、目标和意图等(van der Bosch et al., 2019) 需要了解人智如何共同参与、学习和演化(关系、流程、机制)(van der Bosch et al., 2019) 	<ul style="list-style-type: none"> 人智参与理论和方法(短期与长期、任务与社会参与、参与动态过程)(Madni & Madni, 2018) 人智学习模式(Schoonderwoerd et al., 2022) 团队学习、知识和经验分享模式(van der Bosch et al., 2019) 人智共同演化模型和方法(Döppner et al., 2019)
社会因素	<ul style="list-style-type: none"> 社会化人-人交互向社会化人-智交互的转移(Schneeberger, 2018) 缺乏对人智组队中社会认知、社会角色、社会适应性、情感等方面的了解 	<ul style="list-style-type: none"> 人智组队中的社会化智能体(André et al., 2020) 人与社会化智能体之间的群体互动(André et al., 2020) 人智组队的社会交互(Bendell et al., 2021)

5 新研究范式取向对人因科学研究重点的意义

为进一步评价人因科学研究范式取向对今后人因科学研究的作用,我们从表2~表4中选择了智能时代人因科学今后的一些重点研究课题做进一步分析(见表5)。表5概括了这些研究重点与研究范式取向之间的关系,其中,表5列标题是人因科学研究中所需的多样化范式取向,表5行标题是今后人因科学研究的一些重要研究课题,表5中的各个单元格概括了基于各种研究范式取向,人因科学

各项重要课题在研究中需要考虑的一系列重要研究问题。

由表5可知,一方面,智能机器行为、人智组队合作、人机混合智能、伦理化 AI、智能人机交互、可解释 AI 等人因科学研究的重要课题都需要多样化的研究范式取向来支持,现有的人因科学研究范式取向(如神经人因学、工程心理学、人机交互)仍然起着重要的作用;另一方面,本文所提出的三种研究范式新取向,拓展了智能时代人因科学研究的角度、方法和范围,有助于人因科学在智能时代开展全方位的研究,提供完整系统化的人因科学解决方案。

表 5 人因学科研究范式取向与研究重点的关系

研究课题	研究范式取向					
	神经人因学(认知加工的神经机制)	工程心理学(认知信息加工)	人机交互	基于人智组队合作的 人智协同认知系统	人智协同认知生态系统	智能社会技术系统
智能机器行为	基于人类认知神经模型的机器学习算法模型优化	基于人类信息加工模型的机器学习算法优化(Leibo et al., 2018)	基于迭代式原型设计、用户测试方法,优化机器学习算法训练和测试来减小算法和行为偏差	人智协同合作对机器行为的影响	机器行为进化模式,人机行为协同共生理论(Rahwan et al., 2019)	社会环境对机器行为的影响,社会交互中的机器行为,机器行为公正性和伦理化,智能系统决策与组织决策的协调
人智组队合作	人智组队合作和交互中的神经机制(Stevens & Galloway, 2019)	用户感知、情感、意图、行为等认知模型	基于人智组队合作的人机交互和界面模型	针对人智组队合作模型的研究,包括人机互信、分享式情景意识、心理模型、决策等	人智组队的生态系统,多智能体系统间的协同合作(Mohanty & Vyas, 2018),人-智系统自适应、自组织、演化机制(Wu, 2020)	社会环境中的人智组队合作,人与智能体的社会互动,社会责任对人机合作的影响(Mou & Xu, 2017)
人机混合增强智能	脑机混合、脑机融合等研究	人类高级认知计算模型、知识表征和图谱在实现人机混合智能中的应用	“人在环中”混合智能的交互设计,人机协同控制(胡源达等, 2020)	人机混合增强智能中的人智协同合作,人机混合增强智能中的人机智能互补	跨多智能体系统(Dorri et al., 2018),人机物融合群智(分布式协同认知理论)(郭斌 & 於志文, 2021)	社会和组织环境中的人机智能互补和协调、功能和任务分配、人机决策权限设置
伦理化 AI		伦理化 AI 的认知方法(Schoenherr, 2022; Chrisle, 2020)	“有意义的人类控制”设计(自主化系统)(Santoni & van den Hoven, 2018)	人智系统合作的伦理问题(Boni, 2021)	伦理化 AI 的生态系统方法(Stahl, 2021);跨智能体系统的兼容性问题(文化、伦理等)(ISO, 2020)	智能社会技术系统的伦理化 AI 问题,伦理化社会技术系统(Chopra & Singh, 2018)
智能人机交互	脑机接口技术、交互设计、应用	智能人机交互的认知模型和理论,社会和情感交互、意图识别的认知模型	智能人机交互设计新范式和新方法,智能人机交互设计标准	基于人智组队合作的认知界面设计、设计新范式、认知架构	智能人机交互模拟和生态管理,多智能交互系统的共同演化(Döppner et al., 2019)	社会、文化等因素对智能人机交互的影响
可解释 AI	可解释 AI 的认知神经学研究(Fellous et al., 2019)	心理理解释理论转化,可解释 AI 界面的认知模型	创新的人机界面技术和设计,可视化技术和设计	“以人为中心”的可解释 AI(Ehsan et al., 2019)	跨智能决策系统的可解释 AI 问题	公众 AI 信任度和接受度与 AI 解释性间的关系(Ehsan, 2020),可解释 AI 与文化、用户知识、决策、伦理的关系

由此可见,如同人因科学研究范式取向的跨时代演变促进了人因科学的发展,智能时代的人因科学研究同样需要多样化的研究范式取向,而人因科学的研究也将进一步完善研究范式取向,因此,人因科学研究范式取向与研究重点互为影响,互为促进。

6 总结

在人因科学这一新概念的语境下,本文梳理了人因科学研究范式取向跨时代的拓展,智能时代人因科学研究需要多样化、创新的研究范式取向。为了应对这些挑战以及研究的内在需求,从人因科学研究范式取向的角度出发,基于我们提出的人智协同认知系统、人智协同认知生态以及智能社会技术系统概念模型和框架,我们提出了智能时代人因科学的三种新型研究范式取向,这些研究范式新取向的提出对人因科学研究和应用具有重要意义。针对智能时代人因科学研究重点跨时代的转移,本文展望了“人智交互”“智能人机界面”“人智组队合作”三方面的人因科学研究重点。我们认为,多样化、创新的人因科学研究范式取向有利于进一步开展人因科学研究,人因科学研究范式取向与研究重点互为影响,互为促进,共同推动人因科学的发展。

参 考 文 献

- Ali, M. I., Patel, P., Breslin, J. G., Harik, R., & Sheth, A. (2021). Cognitive digital twins for smart manufacturing. *IEEE Intelligent Systems*, 36(2), 96–100.
- Allenby, B. R. (2021). World Wide Weird: Rise of the cognitive ecosystem. *Issues in Science and Technology*, 37(3), 34–45.
- Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4), 105–120.
- Amershi, S., Weld, D., Vorvoreanu, M., Fournay, A., Nushi, B., Collisson, P., ... Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery.
- André, E., Paiva, A., Shah, J., & Šabanovic, S. (2020). *Social agents for teamwork and group interactions*. Report presented at the Dagstuhl Seminar, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Germany.
- Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, 22(2), 8.
- Badham, R., Clegg, C., & Wall, T. (2000). Socio-technical theory. In: Karwowski, W. (Ed.), *Handbook of Ergonomics*. John Wiley, New York, NY.
- Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-AI team performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7(1), 2–11.
- Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1), 4–17.
- Bendell, R., Williams, J., Fiore, S. M., & Jentsch, F. (2021, September). Supporting social interactions in human-AI teams: Profiling human teammates from sparse data. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 665–669). Sage CA: Los Angeles, CA: SAGE Publications.
- Biondi, F., Alvarez, I., & Jeong, K. A. (2019). Human-system cooperation in automated driving. *International Journal of Human-Computer Interaction*, 35(11), 917–918.
- Boni, M. (2021). The ethical dimension of human-artificial intelligence collaboration. *European View*, 20(2), 182–90.
- Borenstein, J., Herkert, J. R., & Miller, K. W. (2019). Self-driving cars and engineering ethics: The need for a system level analysis. *Science and Engineering Ethics*, 25(2), 383–398.
- Brill, J. C., Cummings, M. L., Evans III, A. W., Hancock, P. A., Lyons, J. B., & Oden, K. (2018). Navigating the advent of human-machine teaming. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 455–459). Sage CA: Los Angeles, CA: SAGE Publications.
- Brown, B., Bødker, S., & Höök, K. (2017). Does HCI scale? Scale hacking and the relevance of HCI. *Interactions*, 24(5), 28–33.
- Caldwell, S., Sweetser, P., O'Donnell, N., Knight, M. J., Aitchison, M., Gedeon, T., ... Conroy, D. (2022). An agile new research framework for hybrid human-AI teaming: Trust, transparency, and transferability. *ACM Transactions on Interactive Intelligent Systems*, 12(3), 1–36.
- Chen, L., Wang, B. C., Huang, S. H., Zhang, J. Y., Guo, R., & Lu, J. Q. (2021). Artificial intelligence ethics guidelines and governance system: Current status and strategic suggestions. *Science and Technology Management Research*, 6(6), 193–200.
- [陈磊, 王柏村, 黄思翰, 张吉豫, 郭锐, 鲁俊群. (2021). 人工智能伦理准则与治理体系: 发展现状和战略建议. *科技管理研究*, 6(6), 193–200.]
- Chopra, A. K., & Singh, M. P. (2018, December). Sociotechnical systems and ethics in the large. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 48–53). Association for Computing Machinery.
- Cummings, M. L., & Clare, A. S. (2015). Holistic modelling for human-autonomous system interaction. *Theoretical Issues in Ergonomics Science*, 16(3), 214–231.
- Dehais, F., Karwowski, W., & Ayaz, H. (2020). Brain at work and in everyday life as the next frontier: Grand field challenges for neuroergonomics. *Frontiers in Neuroergonomics*, 1, 583733–583745.
- Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2018). Team coordination and effectiveness in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems*, 49(2), 150–159.
- Döppner, D. A., Derckx, P., & Schoder, D. (2019). Symbiotic co-evolution in collaborative human-machine decision making: Exploration of a multi-year design science research project in the Air Cargo Industry. In *Proceedings of the 52nd Hawaii International Conference on System*

- Sciences*. (pp.125–131). Computer Society Press.
- Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-agent systems: A survey. *IEEE Access*, 6, 28573–28593.
- Eason, K. (2008). Sociotechnical systems theory in the 21st century: Another half-filled glass? *Sense in social science: A collection of essay in honor of Dr. Lisl Klein* (pp. 123–134). Desmond Graves, Broughton.
- Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable ai: Towards a reflective sociotechnical approach. In *HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22* (pp. 449–466). Springer International Publishing.
- Ehsan, U., Tambwekar, P., Chan, L., Harrison, B., & Riedl, M. O. (2019). Automated rationale generation: A technique for explainable AI and its effects on human perceptions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 263–274). Association for Computing Machinery.
- Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., ... Riedl, M. O. (2021). Operationalizing human-centered perspectives in explainable AI. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–6). Association for Computing Machinery.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.
- Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9(1), 4–32.
- Endsley, M. R. (2018). Situation awareness in future autonomous vehicles: Beware of the Unexpected. *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, IEA 2018, Springer.
- Farooq, U., & Grudin, J. (2016). Human computer integration. *Interactions*, 23(6), 27–32.
- Fellous, J. M., Sapiro, G., Rossi, A., Mayberg, H. S., & Ferrante, M. (2019). Explainable artificial intelligence for neuroscience: Behavioral neurostimulation. *Frontiers of Neuroscience*, 13, 1346. doi: 10.3389/fnins.2019.01346
- Finstad, K., Xu, W., Kapoor, S., Canakapalli, S., & Gladding, J. (2009). Bridging the gaps between enterprise software and end users. *Interactions*, 16(2), 10–14.
- Fiore, E. (2020). Ethics of technology and design ethics in socio-technical systems: Investigating the role of the designer. *Form Akademisk Forskningstidsskrift for Design og Designdidaktikk*, 13(1), 13–19.
- Fridman, L. (2018). Human-centered autonomous vehicle systems: Principles of effective shared autonomy. <https://arxiv.org/pdf/1810.01835.pdf>.
- Gao, Q., Xu, W., Shen, M., Gao, Z. (2023). Agent teaming situation awareness (ATSA): A situation awareness framework for human-AI teaming. <https://arxiv.org/abs/2308.16785>.
- Gao, Z. F., Li, W. M., Liang, J. W., Pan, H. X., Xu, W., & Shen, M. W. (2021). Trust in automated vehicles. *Advances in Psychological Science*, 29(11), 1–12.
- [高在峰, 李文敏, 梁佳文, 潘晗希, 许为, 沈模卫. (2021). 自动驾驶车中的人机信任. *心理科学进展*, 29(11), 1–12.]
- Gardner, D., Mark, L., Dainoff, M. & Xu, W. (1995). Considerations for linking seatpan and backrest angles. *International Journal of Human-Computer Interaction*, 7(2), 153–165.
- Ge, L. Z., & Xu, W. (Eds). (2020). *User experience: Theory and practice*. China Renmin University Press.
- [葛列众, 许为. (主编). (2020). *用户体验: 理论与实践*. 中国人民大学出版社.]
- Ge, L. Z., Xu, W., Song, X. (Eds). (2022). *Engineering psychology* (2nd ed.). China Renmin University Press.
- [葛列众, 许为, 宋晓蕾. (主编). (2022). *工程心理学* (第 2 版). 中国人民大学出版社.]
- Goodwin, G. F., Blacksmith, N., & Coats, M. R. (2018). The science of teams in the military: Contributions from over 60 years of research. *American Psychologist*, 73(4), 322.
- Google PAIR. (2019). *People + AI Guidebook: Designing human-centered AI products*. Retrieved Nov. 23, 2023 from <https://pair.withgoogle.com>.
- Guo, B., & Yu, Z. W. (2021). Crowd intelligence with the deep fusion of human, machine, and IoT. *Communication of the CCF*, 17(2), 35–40.
- [郭斌, 於志文. (2021). 人机物融合群智计算. *中国计算机学会通讯*, 17(2), 35–40.]
- Herrmann, T., Schmidt, A., & Degeling, M. (2018, June). From interaction to intervention: An approach for keeping humans in control in the context of socio-technical systems. In *STPIS@ CAiSE* (pp. 101–110). Tallinn, Estonia.
- Heydari, B., Szajnfarder, Z., Panchal, J., Cardin, M. A., Hölttä-Otto, K., Kremer, G. E., & Chen, W. (2019). Analysis and design of sociotechnical systems. *Journal of Mechanical Design*, 141(11), 118001.
- Hodgson, A., Siemieniuch, C. E., & Hubbard, E. M. (2013). Culture and the safety of complex automated sociotechnical systems. *IEEE Transactions on Human-Machine Systems*, 43(6), 608–619.
- Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. London: CRC Press.
- Hollnagel, E., Woods, D., & Leveson, N. (Eds.). (2006). *Resilience engineering: Concepts and precepts*. Williston, VT: Ashgate.
- Holmquist, L. E. (2017). Intelligence on tap: Artificial intelligence as a new design material. *Interactions*, 24(4), 28–33.
- Hu, Y. D., Sun, X. H., Zhang, H. X., Zhang, S. C., & Yi, S. Q. (2020). Interaction design in human-in-the-loop hybrid intelligence. *Packaging Engineering*, 41(18), 38–47.
- [胡源达, 孙效华, 张何辛, 张绳宸, 易思琦. (2020). 支持人在环路混合智能的交互设计研究. *包装工程艺术版*, 41(18), 38–47.]
- Huang, Y., Poderi, G., Šćepanović, S., Hasselqvist, H., Warnier, M., & Brazier, F. (2019). Embedding internet-of-things in large-scale socio-technical systems: A community-oriented design in future smart grids. *The Internet of Things for Smart Urban Ecosystems* (pp. 125–150). Cham: Springer.
- Hughes, J. A., Randall, D., Shapiro, D. (1992). Faltering from ethnography to design. In: *Proceedings of CSCW '92* (pp. 115–122). ACM Press, New York, NY.
- IDC (International Data Corporation). (2020). *智能体白皮书, 共建智能体, 共创全场景智慧*. 2023-11-10 取自 https://www.huawei.com/minisite/building-an-intelligent-world-together/assets/doc/White_Paper_on_Huawei_Intelligent_Twins.pdf
- IEEE (The Institute of Electrical and Electronics Engineers). (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*. The Institute of Electrical and Electronics Engineers (IEEE), Incorporated.
- ISO (International Organization for Standardization). (2020). *Ergonomics- ergonomics of human-system interaction - Part 810: Robotic, intelligent and autonomous systems*.

- Retrieved Nov. 10, 2023 from <https://www.iso.org/standard/76577.html>.
- Kaber, D. B. (2018). A conceptual framework of autonomous and automated agents. *Theoretical Issues in Ergonomics Science*, 19(4), 406–430.
- Kaluarachchi, T., Reis, A., & Nanayakkara, S. (2021). A review of recent deep learning approaches in human-centered machine learning. *Sensors*, 21(7), 2514.
- Le Page, C., & Bousquet, F. (2004). Multi-agent simulations and ecosystem management: A review. *Ecological modelling*, 176(3–4), 313–332.
- Lee, J. D., & Kolodge, K. (2018). Understanding attitudes towards self-driving vehicles: Quantitative analysis of qualitative data. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), doi: 10.1177/1541931218621319
- Leibo, J. Z., d'Autume, C. D. M., Zoran, D., Amos, D., Beattie, C., Anderson, K., ... Botvinick, M. M. (2018). Psychlab: A psychology laboratory for deep reinforcement learning agents. *arXiv preprint arXiv:1801.08116*.
- Li, F. F. (2018). How to make A.I. that's good for people. *The New York Times*. Retrieved Nov. 10, 2023 from <https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html>.
- Lieberman, H. (2009). User interface goals, AI opportunities. *AI Magazine*, 30(4), 16–22.
- Liu, W. (2023). *Human-machine environmental system intelligence: Beyond human-machine fusion*. Beijing: Science Press.
- [刘伟. (2023). 人机环境系统智能:超越人机融合. 北京: 科学出版社]
- Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, 8(4), 277–301.
- Madni, A. M., & Madni, C. C. (2018). Architectural framework for exploring adaptive human-machine teaming options in simulated dynamic environments. *Systems*, 6(4), 44(2018).
- Mcgregor, S. (2022). *AI Incidents Database*. Retrieved Nov. 10, 2023 from <https://incidentdatabase.ai/>.
- Mohanty, S., & Vyas, S. (2018). Putting it all together: Toward a human-machine collaborative ecosystem. In S. Mohanty & S. Vyas (Eds.), *How to compete in the age of artificial intelligence: Implementing a collaborative human-machine strategy for your business* (pp. 215–229). Apress.
- Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, 72(3), 432–440.
- Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *arXiv preprint arXiv:1902.01876*.
- National Academies of Sciences, Engineering, and Medicine (NASEM). (2021). *Human-AI teaming: State-of-the-art and research needs*. Retrieved Nov. 10, 2023 from <https://nap.nationalacademies.org/catalog/26355/human-ai-teaming-state-of-the-art-and-research-needs>
- Neftci, E. O., & Averbeck, B. B. (2019). Reinforcement learning in artificial and biological systems. *Nature Machine Intelligence*, 1(3), 133–143.
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Norman, D. A., & Draper, S. W. (1986). *User-centered system design: New perspectives on human-computer interaction*. CRC Press.
- Norman, D. A., & Stappers, P. J. (2015). DesignX: complex sociotechnical systems. *She Ji: The Journal of Design, Economics, and Innovation*, 1(2), 83–106.
- Norman, K., & Kirakowski, J. (Eds.). (2017). *The Wiley handbook of human-computer interaction set*. John Wiley & Sons.
- NTSB. (2017). Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston, Florida, May 7, 2016. *Accidents Report*, by National Transportation Safety Board (NTSB) 2017, Washington, DC.
- Parasuraman, R., & Rizzo, M. (Eds.). (2006). *Neuroergonomics: The brain at work* (Vol. 3). Oxford University Press.
- Prada, R., & Paiva, A. (2014). Human-agent interaction: Challenges for bringing humans and agents together. In *Proc. of the 3rd Int. Workshop on Human-Agent Interaction Design and Models (HAIDM 2014)* (pp. 1–10), Association for Computing Machinery.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., ... Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477–486.
- Roth, E. M., Sushereba, C., Militello, L. G., Diulio, J., & Ernst, K. (2019). Function allocation considerations in the era of human autonomy teaming. *Journal of Cognitive Engineering and Decision Making*, 13(4), 199–220.
- Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... Xu, W. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human-Computer Interaction*, 39(3), 391–437.
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50(3), 540–547.
- Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering and design* (7th ed.). McGraw-Hill Education
- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Front Robot AI*, 5(2), 15.
- Schneeberger, T. (2018). Transfer of social human-human interaction to social human-agent interaction. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 1778–1780), Association for Computing Machinery.
- Schoenherr, J. (2022). *Ethical artificial intelligence from popular to cognitive science: Trust in the age of entanglement*. Routledge.
- Schoonderwoerd, T. A., van Zoelen, E. M., van den Bosch, K., & Neerinx, M. A. (2022). Design patterns for human-AI co-learning: A wizard-of-Oz evaluation in an urban-search-and-rescue task. *International Journal of Human-Computer Studies*, 164(8), 102831.
- Shi, Y. C. (2021). Metaverse needs a breakthrough in human-computer interaction. *Communication of the CAAI*, 12(1), 26–33.
- [史元春. (2021). 元宇宙需要人机交互的突破. *中国人工智能学会通讯*, 12(1), 26–33.]
- Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V. & Johnson, W. W. (2018). Why human-autonomy teaming? *International Conference on Applied Human Factors and Ergonomics*, May 2018, Orlando, FL.
- Society of Automotive Engineers (SAE). (2019). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *Recommended Practice J3016*. Retrieved Nov. 10, 2023 from <https://www.sae.org/>

- standards/content/j3016_202104/
- Stahl, B. C. (2021). *Artificial intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies* (p. 124). Springer Nature.
- Steghofer, J.-P., Diaconescu, A., Marsh, S., & Pitt, J. (2017). The next generation of socio-technical systems: Realizing the potential, protecting the value [introduction]. *IEEE Technology and Society Magazine*, 36(3), 46–47.
- Stevens, R. H., & Galloway, T. L. (2019). Teaching machines to recognize neurodynamic correlates of team and team member uncertainty. *Journal of Cognitive Engineering and Decision Making*, 13(4), 310–327.
- Sun, X. H., Wu, C. X., Zhang, I., & Qu, W. N. (2011). The role, status, and current development of engineering psychology. *Bulletin of Chinese Academy of Sciences*, 26(6), 650–660.
- [孙向红, 吴昌旭, 张亮, 瞿炜娜. (2011). 工程心理学作用、地位和进展. *中国科学院院刊*, 26(6), 650–660.]
- Tan, Z. Y., Dai, N. Y., Zhang, R. F., & Dai, K. Y. (2020). Overview and perspectives on human-computer interaction in intelligent and connected vehicles. *Computer Integrated Manufacturing Systems*, 26(10), 2615–2632.
- [谭征宇, 戴宁一, 张瑞佛, 戴柯颖. (2020). 智能网联汽车人机交互研究现状及展望. *计算机集成制造系统*, 26(10), 2615–2632.]
- van der Bosch, K., Schoonderwoerd, T., Blankendaal, R., & Neerincx, M. (2019). Six challenges for human-AI Co-learning. In *International Conference on Human-Computer Interaction* (pp. 572–589). Springer International Publishing.
- Van de Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines*, 30(3), 385–409.
- Waterson, P., Robertson, M. M., Cooke, N. J., Militello, L., Roth, E., & Stanton, N. A. (2015). Defining the methodological challenges and opportunities for an effective science of sociotechnical systems and safety. *Ergonomics*, 58(4), 565–599.
- Werfel, J., Petersen, K., & Nagpal, R. (2014). Designing collective behavior in a termite-inspired robot construction team. *Science*, 343(6172), 754–758.
- Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior*, 101(12), 197–209.
- Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge.
- Wooldridge, M., & Jennings, N. R. (1995). Intelligent agent: Theory and practice. *Knowledge Engineering*, 10(2), 115–152.
- Wu, Z. (2020). Ecological transformation and man-machine symbiosis: A study on the relationship between human and artificial intelligence. In *Multidisciplinary Digital Publishing Institute Proceedings*, 47(1), 37, MDPI.
- Xie, L., & Xie, X. (2021). General situational intelligence. *Communication of the CCF*, 17(2), 8–9.
- [谢磊, 谢幸. (2021). 泛在情境智能. *中国计算机学会通讯*, 17(2), 8–9.]
- Xu, W. (2003). User-Centered Design approach: Opportunities and challenges of human factors practices in China. *Chinese Journal of Ergonomics*, 9(4), 8–11.
- [许为. (2003). 以用户为中心设计: 人机工效学的机遇和挑战. *人类工效学*, 9(4), 8–11.]
- Xu, W. (2005). Recent trend of research and applications on human-computer interaction. *Chinese Journal of Ergonomics*, 11(4), 37–40.
- [许为. (2005). 人-计算机交互作用研究和应用新思路的探讨. *人类工效学*, 11(4), 37–40.]
- Xu, W. (2007). Identifying problems and generating recommendations for enhancing complex systems: Applying the abstraction hierarchy framework as an analytical tool. *Human Factors*, 49(6), 975–994.
- Xu, W. (2012). User experience design: Beyond user interface design and usability. In I. Nunes (Ed.), *Ergonomics, a Systems Approach* (Chapter 8). InTech.
- Xu, W. (2014). Enhanced ergonomics approaches for product design: A user experience ecosystem perspective and case studies. *Ergonomics*, 57(1), 34–51.
- Xu, W. (2017). User-centered design (II): New challenges and new opportunities. *Chinese Journal of Ergonomics*, 23(1), 82–86.
- [许为. (2017). 再论以用户为中心的设计: 新挑战和新机遇. *人类工效学*, 23(1), 82–86.]
- Xu, W. (2019). Toward human-centered AI: A perspective from human-computer interaction. *Interactions*, 26(4), 42–46.
- Xu, W. (2019a). User-centered design (III): Methods for user experience and innovative design in the intelligent era. *Chinese Journal of Applied Psychology*, 25(1), 3–17.
- [许为. (2019a). 三论以用户为中心的设计: 智能时代的用户体验和创新设计. *应用心理学*, 25(1), 3–17.]
- Xu, W. (2019b). User-centered design (IV): Human-centered artificial intelligence. *Chinese Journal of Applied Psychology*, 25(4), 291–305.
- [许为. (2019b). 四论以用户为中心的设计: 以人为中心的人工智能. *应用心理学*, 25(4), 291–305.]
- Xu, W. (2020). User-Centered Design (V): From automation to the autonomy and autonomous vehicles in the intelligence era. *Chinese Journal of Applied Psychology*, 26(2), 108–128.
- [许为. (2020). 五论以用户为中心的设计: 从自动化到智能时代的自主化以及自动驾驶车. *应用心理学*, 26(2), 108–128.]
- Xu, W. (2021). From automation to autonomy and autonomous vehicles: Challenges and opportunities for human-computer interaction. *Interactions*, 28(1), 48–53.
- Xu, W. (2022). *AI in human-computer interaction and user experience*. <https://arxiv.org/abs/2301.00987>.
- Xu, W. (2022a). User-centered design (VI): Human factors engineering approaches for intelligent human-computer interaction. *Chinese Journal of Applied Psychology*, 28(3), 191–209.
- [许为. (2022a). 六论以用户为中心的设计: 智能人机交互的人因工程途径. *应用心理学*, 28(3), 191–209.]
- Xu, W. (2022b). User-centered design (VII): From automated to intelligent flight deck. *Chinese Journal of Applied Psychology*, 28(4), 291–313.
- [许为. (2022b). 七论以用户为中心的设计: 从自动化到智能化飞机驾驶舱. *应用心理学*, 28(4), 291–313.]
- Xu, W. (2022c). User-Centered Design (VIII): A new framework of intelligent sociotechnical systems and prospects for future human factors research. *Chinese Journal of Applied Psychology*, 28(5), 387–401.
- [许为. (2022c). 八论以用户为中心的设计: 一个智能社会技术系统新框架及人因工程研究展望. *应用心理学*, 28(5), 387–401.]
- Xu, W. (2023). User-centered design (IX): A "user experience 3.0" paradigm framework in the intelligence era. *Chinese Journal of Applied Psychology* (online published) <http://www.appliedpsy.cn/CN/abstract/abstract448.shtml>
- [许为. (2023). 九论以用户为中心的设计: 智能时代的“用

- 户体验 3.0”范式. *应用心理学*. 在线发表: <http://www.appliedpsy.cn/CN/abstract/abstract448.shtml>
- Xu, W., & Chen, Y. (2012). New progress and applications of human factors in the research and development of civil flight deck. *Aeronautical Science & Technology*, 6, 18–21.
- [许为, 陈勇. (2012). 人机工效学在民用客机研发中应用的新进展及建议. *航空科学技术*, 6, 18–21.]
- Xu, W., & Chen, Y. (2013). Challenges and strategies of human factors airworthiness certification for civil aircraft. *Civil Aircraft Design and Research*, (2), 24–30.
- [许为, 陈勇. (2013). 民用客机人机工效学适航认证及对策. *民用飞机设计与研究*, (2), 24–30.]
- Xu, W., & Chen, Y. (2014). Reducing design-induced pilot error in civil flight deck: Perspectives of airworthiness certification and design. *Civil Aircraft Design and Research*, (3), 5–11.
- [许为, 陈勇. (2014). 从驾驶舱设计和适航来减少由设计引发的飞行员人为差错的挑战和途径. *民用飞机设计与研究*, (3), 5–11.]
- Xu, W., Chen, Y., Dong, W. J., Dong, D. Y., & Ge, L. Z. (2021). Status and prospect of human factors engineering research on single pilot operations for large commercial aircraft. *Advances in Aeronautical Science and Engineering*, 13(1), 1–18.
- [许为, 陈勇, 董文俊, 董大勇, 葛列众. (2021). 大型商用飞机单一飞行员驾驶的人因工程研究进展与展望. *航空工程进展*, 13(1), 1–18.]
- Xu, W., & Dainoff, M. (2023). Enabling human-centered AI: A new junction and shared journey between AI and HCI communities. *Interactions*, 30(1), 42–47.
- Xu, W., Dainoff, M., Ge, L., & Gao, Z. (2022). From human-computer interaction to human-AI interaction: New challenges and opportunities for enabling human-centered AI. *International Journal of Human Computer Interaction*, 39(3), 494–518.
- Xu, W., Dainoff, M. J., & Mark, L. S. (1999). Facilitate complex search tasks in hypertext by externalizing functional properties of a work domain. *International Journal of Human-Computer Interaction*, 11(3), 201–229.
- Xu, W., Furie, D., Mahabaleswar, M., Suresh, B., & Chouhan, H. (2019). Applications of an interaction, process, integration and intelligence (IPII) design approach for ergonomics solutions. *Ergonomics*, 62(7), 954–980.
- Xu, W. & Gao, Z. (2023). *Applying HCAI in developing effective human-AI teaming: A perspective from human-AI joint cognitive systems*. <https://arxiv.org/abs/2307.03913>
- Xu, W., & Ge, L. Z. (2018). New trends in human factors. *Advances in Psychological Science*, 26(9), 1521–1534.
- [许为, 葛列众. (2018). 人因学发展的新取向. *心理科学进展*, 26(9), 1521–1534.]
- Xu, W., & Ge, L. Z. (2020). Engineering psychology in the era of artificial intelligence. *Advances in Psychological Science*, 28(9), 1409–1425.
- [许为, 葛列众. (2020). 智能时代的工程心理学. *心理科学进展*, 28(9), 1409–1425.]
- Xu, W., Ge, L. Z., & Gao, Z. F. (2021). Human-AI interaction: An emerging interdisciplinary domain for enabling human-centered AI. *CAAI Transactions on Intelligent Systems*, 16(4), 604–621.
- [许为, 葛列众, 高在峰. (2021). 人-AI 交互: 实现“以人为中心 AI”理念的跨学科新领域. *智能系统学报*, 16(4), 604–621.]
- Xu, W. & Zhu, Z. (1990). The effects of ambient illumination and target luminance on colour coding in a CRT display. *Ergonomics*, 33(7), 933–944.
- Xu, W., & Zhu, Z. X. (1989). Effects of ambient illuminant intensity, color temperature and target luminance on color coding in a CRT display. *Acta Psychologica Sinica*, 21(4), 269–277.
- [许为, 朱祖祥. (1989). 环境照明强度、色温和目标亮度对 CRT 显示器颜色编码的影响. *心理学报*, 21(4), 269–277.]
- Yi, X., Yu, C., & Shi, Y. C. (2018). Bayesian method for intent prediction in pervasive computing environments. *Science China Information Sciences*, 48(4), 419–432.
- [易鑫, 喻纯, 史元春. (2018). 普适计算环境中用户意图推理的 Bayes 方法. *中国科学: 信息科学*, 48(4), 419–432.]
- Zanzotto, F. M. (2019). Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research*, 64(2), 243–252.
- Zheng, N. N., Liu, Z. Y., Ren, P. J., Ma, Y. Q., Chen, S. T., Yu, S. Y., ... Wang, F. Y. (2017). Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*, 18(2), 153–179.
- Zhu, Y., Gao, T., Fan, L., Huang, S., Edmonds, M., Liu, H., ... Zhu, S. C. (2020). Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering*, 6(3), 310–345.
- Zong, Z. F., Dai, C. H., & Zhang, D. (2021). Human-machine interaction technology of intelligent vehicles: Current development trends and future directions. *China Journal of Highway and Transport*, 34(6), 214.
- [宗长富, 代昌华, 张东. (2021). 智能汽车的人机共驾技术研究现状和发展趋势. *中国公路学报*, 34(6), 214.]

New research paradigms and agenda of human factors science in the intelligence era

XU Wei¹, GAO Zaifeng², GE Liezhong¹

(¹ Center for Psychological Sciences; ² Department of Psychology, Zhejiang University, Hangzhou 310058, China)

Abstract

This paper first proposes the innovative concept of “human factors science” to characterize engineering psychology, human factors engineering, ergonomics, human-computer interaction, and other similar fields. Although the perspectives in these fields differ, they share a common goal: optimizing the human-machine relationship by applying a “human-centered design” approach. AI technology has brought in new characteristics, and our recent research reveals that the human-machine relationship presents a trans-era evolution from “human-machine interaction” to “human-AI teaming.” These changes have raised questions and challenges for human factors science, compelling us to re-examine current research paradigms and agendas.

In this context, this paper reviews and discusses the implications of the following three conceptual models and frameworks that we recently proposed to enrich the research paradigms for human factors science. (1) human-AI joint cognitive systems: this model differs from the traditional human-computer interaction paradigm and regards an intelligent system as a cognitive agent with a certain level of cognitive capabilities. Thus, a human-AI system can be characterized as a joint cognitive system in which two cognitive agents (human and intelligent agents) work as teammates for collaboration. (2) human-AI joint cognitive ecosystems: an intelligent ecosystem with multiple human-AI systems can be represented as a human-AI joint cognitive ecosystem. The overall system performance of the intelligent ecosystem depends on optimal collaboration and design across the multiple human-AI systems. (3) intelligent sociotechnical systems (iSTS): human-AI systems are designed, developed, and deployed in an iSTS environment. From a macro perspective, iSTS focuses on the interdependency between the technical and social subsystems. The successful design, development, and deployment of a human-AI system within an iSTS environment depends on the synergistic optimization between the two subsystems.

This paper further enhances these frameworks from the research paradigm perspective. We propose three new research paradigms for human factors science in the intelligence era: human-AI joint cognitive systems, human-AI joint cognitive ecosystems, and intelligent sociotechnical systems, enabling comprehensive human factors solutions for AI-based intelligent systems. Further analyses show that the three new research paradigms will benefit future research in human factors science. Furthermore, this paper looks forward to the future research agenda of human factors science from three aspects: “human-AI interaction”, “intelligent human-machine interface”, and “human-AI teaming”. We believe the proposed research paradigms and the future research agenda will mutually promote each other, further advancing human factors science in the intelligence era.

Keywords Human factors science, engineering psychology, human factors engineering, research paradigm, human-AI teaming