

特色专题

迈向人机共生时代：人机协同决策的心理学视角

朱奥¹, 刘华硕¹, 袁佩君^{2*}, 张丹^{1,3*}

摘要 在人工智能技术的驱动下,人机协同决策效能逐步提升,展现出巨大的潜力。然而,当前人机协同决策领域面临情感问题、信任问题、决策权力分配问题和道德对齐问题等挑战。其中,人机协同情感问题涉及机器如何克服人机异质性、准确识别人类的情感状态并做出适当响应;信任问题关系到人类对机器决策的依赖程度和信任水平,需要在过度信任和信任不足之间找到平衡;决策权力分配问题则聚焦于在人机协作中如何合理分配决策权,以防止机器权力过大或不足;道德对齐问题探讨如何确保机器决策符合人类的道德标准和价值观,保障决策的伦理性和社会接受度。提出了人机协同决策的未来发展方向,包括提升机器对人类情感的理解和响应能力、建立稳固的信任机制、优化决策权力的分配方式,以及确保机器决策符合伦理和道德标准。在迈向人机共生的道路上,不仅需要技术创新,还需要结合心理学视角来应对这些挑战,以实现更加高效、安全和可持续的人机协同决策。

关键词 人机共生;人机协作;协同决策;人机信任;心理学视角

人工智能的发展正在改变着人与机器的关系,人类正在从对机器的使用和控制转向与机器的深度协作^[1]。20世纪60年代,著名的心理学家、人工智能先驱 Licklider 提出“人机共生”(human-machine symbiosis)的概念,来描述人类与机器紧密协作、共同达成目标的关系。人机共生的核心思想在于发挥人类与机器各自的优势,将人类的创造力、直觉和判断力等与机器的计算能力、数据处理和信息整合能力相结合,通过人类与机器的优势互补来实现优于单一智能体的表现^[2]。为了实现这一目标,研究人员开始关注如何设计有效的人机协同系统,发挥人类智能与机器智能各自的优势,以提高人机协同效率,并最

终达成人类与机器智能体相互理解、共同决策的人机共生关系。

近年来,机器在认知智能方面取得了长足的进步,尤其在博弈学习、自然语言处理、图像识别等领域已经展现出超过人类表现的潜力。例如,AlphaGo 在2016年战胜了韩国围棋冠军李世石,展示了人工智能在复杂策略游戏中的优势^[3];ChatGPT在生成内容的准确性、效率等方面已经达到甚至超越了人类专家的水平^[4];在医疗领域,人工智能辅助诊断系统已经在某些情况下超越人类医生的平均诊断准确性^[5]。尽管机器尚未具备通用智能,其在情境适应能力、知识泛化能力、创造力、发散性思维和道德判断等方面仍然难以取代人类,但机器智能已能够在诸多领域为人类提供辅助决策信息,具备了与人类共同完成复杂认知决策的潜力^[1]。随着技术进步,智能时代的人机协同决策系统中,机器智能体可能不再仅起辅助决策的

1. 清华大学心理与认知科学系,北京 100084

2. 启元实验室,北京 100095

3. 清华大学脑与智能实验室,北京 100084

收稿日期:2023-09-05;修回日期:2024-04-10

基金项目:教育部高等学校心理学类专业教指委教改项目(20222008);清华大学教学改革项目(DX05-02)

作者简介:朱奥,博士研究生,研究方向为人机协同决策,电子信箱:za20@mails.tsinghua.edu.cn;袁佩君(通信作者),助理研究员,研究方向为人机交互,电子信箱:yuanpeijun@qiyuanlab.com;张丹(共同通信作者),副教授,研究方向为情感计算、脑机接口,电子信箱:dzhang@tsinghua.edu.cn

引用格式:朱奥,刘华硕,袁佩君,等. 迈向人机共生时代:人机协同决策的心理学视角[J]. 科技导报, 2025, 43(3): 37-46; doi:10.3981/j.issn.1000-7857.2023.09.01354

作用,而是在决策过程中逐渐占据主导地位。

然而,人类与机器2种异质智能体之间交互并进行决策的过程面临诸多新挑战。在传统人人协同中面临的一些问题在人机协同中可能更为突出,这可总结为人机内部兼容性(human-agent inner compatibility)问题^[6]。人机内部兼容性指人类智能与机器智能在认知、情感、价值观等方面的内在兼容性,这是影响人机协同决策的重要因素。因此,如何从以人为中心的出发点去研究并设计人机协同决策系统,使机器智能适配人类的目标与价值取向,从而更加安全、可靠和可控地与人合作,成为当前人工智能领域激烈讨论的焦点问题。具体而言,研究人员对多种可能存在的问题进行了详细的归纳和分析。Sheridan^[7]指出,由于人类和机器在心理模型上的不一致,可能导致双方的相互理解不足,进而影响协作效率。此外,人类对机器的信任程度也会显著影响协同任务的质量:信任不足会导致怀疑和不配合,而过度信任则可能导致依赖和失误。为解决这些问题,Sheridan强调动态调整人机决策权力分配的重要性,这样可以实现人类和机器的优势互补,从而提高协作效果。此外,在不同场景中,尤其是高风险领域,如自动驾驶和医疗领域,机器决策的伦理和道德问题同样至关重要。例如,在自动驾驶中,机器需要在紧急情况下做出道德判断,而在医疗中,机器的诊断和治疗决策直接关系到患者的生命安全。Huang等^[8]进一步支持了这些观点,提出机器需要发展出能够理解人类心理状态的共情智能(empathetic intelligence),认为情感计算是人机互理解的关键,准确识别和响应人类情感在提高机器服务质量和人类满意度方面具有关键作用。此外,Huang等还强调,信任是确保高效能人机协同的关键因素,而合理的权力分配不仅能提升机器的服务效率,还能确保协作过程中的公平和透明。上述研究主要涉及人机情感问题、人机信任问题、人机决策权力分配问题及人机道德对齐问题。为应对这些问题,不仅需要从机器性能的角度寻求突破,更需要从人的角度思考解决方案^[7-9]。

基于上述观点,本文从机器智能如何能够以人为中心地完成人机协同决策的角度,重点讨论了情感理解、人机信任、决策权分配、道德对齐4个方面的内容。

1 人机协同决策:发展与基础

1.1 人机协同决策的发展

20世纪50年代,早期的决策支持系统主要依赖人力计算和协调的单一数学模型。然而随着问题的复杂性增加,决策过程所依赖的模型数量逐渐增多,超出了传统的人力处理能力范畴。因此,科研人员开始寻求大型计算机的辅助,由此开启了人机协同决策的篇章^[10-12]。

20世纪60年代,交互式计算机技术取得突破性进展,以IBM System 360为代表的大型计算机得到广泛应用,推动了人机协同决策的快速发展。此时期典型的决策支持应用为管理信息系统(management information system, MIS),这类初级的系统主要为管理人员提供结构化的定期报告,但并未提供交互式支持以协助决策^[1]。1971年,Scott Morton首次提出了决策支持系统(decision support system, DSS)的概念,实现数据处理和模型计算的有效集成,标志着机器智能正式参与到决策过程中。

20世纪七八十年代,决策支持系统开始蓬勃发展,不仅被用来帮助人类决策者进行数据分析与计算,也开始通过交互界面向人类提供决策建议^[13-14]。这些系统通常基于大量历史数据和特定规则构建,主要用于专业性较强的领域(如医学诊断、法律或者财务等)^[15-16]。然而,Sprague^[16]提出决策支持系统只能为决策者提供有用的辅助信息,并不能替代决策者进行决策制定,决策仍然需要人类来完成。在这个观念的影响下,同时也因为计算机和相关信息技术发展得相对不成熟,人的作用在决策支持过程中占据主导地位。

随着技术的发展,尤其是进入21世纪后,数据挖掘、机器学习等技术的进步提升了决策支持系统的效能:机器在处理和析庞大数据集方面的能力得到提升,使得决策支持系统可以更有效地挖掘数据中的模式和关系^[17];机器学习技术使决策支持系统拥有对复杂问题的解决和预测能力,并且可以在与人类决策者的互动中不断学习和适应^[18];自然语言处理技术的发展使得决策支持系统可以更好地理解人类的需求和意图,通过更自然的方式与决策者进行交流,有助于提高人机协同决策的效率和效果^[19]。由于

决策支持系统的快速发展,其应用扩展至农业发展、交通管理、灾害应对等更广泛的领域^[20]。在应用过程中,这些决策系统在某些方面甚至展示出超越人类的决策能力,对人机交互的形态提出新的挑战,促使研究人员开始关注和讨论人机协同决策的未来。

1.2 人机协同决策的基础

人机协同决策的重要基础是对人类决策过程和机制的理解。人类的决策过程是一个由信息感知、信息处理和解读、可选项评估、结果预期等多重复杂心理机制驱动的过程。人类在处理复杂问题和未知情况上表现出显著优势,能够快速理解并适应多样化的情境,利用跨领域的知识和经验进行决策,甚至运用创新和发散性思维探索全新的解决途径^[21-22];机器智能在灵活性^[23]、知识整合能力^[24-25]、应对新颖和突发问题能力等方面则相对较弱,这可能是由于现有的机器通常专注于优化特定任务的性能,而非寻求全新的解决途径^[26];此外,得益于人类的情感和社会智能,人类能够根据其立场、目标和价值观,权衡利弊,进而做出复杂道德决策,而机器则通常依赖于人类预设的规则和程序。

人类决策能力也存在局限。第一,人类容易受到各种认知偏见的影响,如过度自信、损失厌恶、后悔厌恶和赌徒谬误等^[27],而机器依赖算法和数据,在处理信息时更加客观^[28]。第二,人类在处理大量数据和复杂信息时可能会受到认知能力的限制,而机器智能体在处理大量数据和复杂数学计算方面具有显著优势^[26,29-30]。第三,人类记忆容量有限,在面对复杂问题时,可能难以有效提取过去的有关经验和知识,而以计算机为硬件基础的决策支持系统可以快速准确地存储和检索相关信息^[31]。第四,人类的决策可能受到情绪因素的影响,导致非理性的行为,而机器可以保持稳定地输出^[32]。第五,人类在面临问题过于复杂、信息过载、长时间工作的情况下,可能会导致认知负荷过高和决策疲劳^[33],而机器能够遵循相同的决策过程,保证性能的持续性和稳定性^[3]。

综上,人类可以利用自身的情境适应能力、知识和经验泛化能力、创新思维,以及道德判断能力,而机器则可以发挥其高速计算能力、大数据处理能力、稳定决策等优势。在未来人机协同决策中,利用机器的强大能力和人类的独特优势,可以有效提高决

策质量和效率,人机协同决策将成为决策科学的重要发展方向。

2 人机协同决策的心理学

基于智能技术的机器正在从人类的辅助工具逐步发展成为具备一定程度人类认知能力的自主化智能体(intelligent agent),但是由于人机在决策能力上存在异质性,这给人机协同决策带来了新挑战。有研究将技术进步带来的挑战总结为情感和社交交互、信任和依赖性、任务分配和决策权力分配等问题^[7];也有研究人员认为当前问题主要包括如何在决策过程中平衡人类与机器的权力、如何在人机协同中确保有效地沟通与协作、如何处理人机之间的信任问题及人工智能的道德与伦理问题^[8]。本研究从心理学视角出发,基于人机决策能力的差异,将目前人机协同决策面临的主要挑战总结为人机情感问题、人机信任问题、人机决策权力分配问题及人机道德对齐问题。

2.1 人机情感问题

人工智能先驱 Minsky^[34]在 1985 年提出,机器智能的重要标志之一是拥有情感能力。因为情感是影响人类决策最重要的因素之一^[35],在人机协同决策中让机器理解人类情感并进行适当响应,能够促进人机信任,优化决策效率,进而增进协作效果^[36-38]。具体而言,机器需要能够识别人类的情感状态,在决策互动中动态调整对人类情感的判读,并进行适当响应和适配。有研究证明,工业生产中当机器人能够理解并响应人类操作员的情绪状态时,操作员的工作体验将得到改善,协同效率也会提高^[39-40]。

针对如何让机器拥有情感能力的问题,Picard 在 1997 年提出了“情感计算”(affective computing)的概念,即一种与情感相关的、能够影响情感的计算,旨在赋予机器识别、理解、适应并表达人的情感的能力。当前,情感计算主要基于外部和内部表征来识别人类的基本情绪。外部表征包括语音、文本、面部表情等,内部表征主要有脑电、心电等神经生理信号^[41-45]。然而,机器对人类情感的识别存在诸多不足,如识别准确性较低,对不同文化背景的理解欠佳等,尤其在复杂和微妙的情感识别方面仍存在很大的挑战。情感识别的问题有可能在人机协同决策中

导致机器对人类情感的误解,从而影响协同效果^[46]。同时,由于人类情感有随情境与事件变化的灵活性^[47],机器应该具备灵活调整的能力。当检测到人类表现出情绪状态的变化时,机器需要能够调整其策略或反馈方式,以适应人类的情感需求。这要求机器具有高度的情境理解能力和社交智能,但目前的机器在这方面仍然有待进一步优化^[48]。此外,个体差异也是情感识别的挑战,不同的人可能有不同的情绪反应和表达方式,而这些差异可能会对情绪识别和人机交互产生深远的影响^[49]。而情感表达是指机器以某种形式回应人类情感,如改变反馈语音的音调或使用表情符号等。研究显示,机器可以通过情感交互提高用户的满意度和参与度^[50];在医疗领域,具备情感识别和社交功能的系统可以更准确地理解患者的需求和情感状态,从而帮助医生做出更有效的治疗决策^[51]。然而,如何量化地设计和实现这种情感交互仍是一个开放的问题。

综上所述,在未来的人机协同决策过程中,机器不仅需要具备高效的任务执行能力,更需要具备情感智能,以便更好地理解和适应人类的情感和行为。情感计算在人机协同决策中举足轻重,但也面临着许多待解决的问题。如何准确地分析和理解用户的内在情感体验,已经成为这个领域关键的研究方向。为了在人机协同决策中赋予机器情感智能,可以优化深度学习模型以精准识别情绪,并提高情感智能的时效性。同时,还需要深入理解人类情感机制,从心理学的角度支持情感计算研究。

2.2 人机信任问题

人机信任直接影响了人类接受机器并遵循其决策或建议的意愿程度,因此,人机信任决定了人类能够在多大程度上受益于机器智能的独特优势^[52]。过高或过低的人机信任程度都会降低人机协作任务的完成质量^[53]。当前,机器智能为人类解决复杂问题提供了辅助,但有学者认为用户在人机协作时有时会遵从“不信任、不使用”的原则^[54],而没有充分地信任机器,这导致机器可能会被弃用或误用^[55]。另外,对协作机器的过度信任可能会导致用户对机器的过度依赖,在关键情况下可能会导致决策失败^[56]。

过度信任是指人类决策者在决策过程中过于相信机器,当人类过度信任机器时可能会忽略自己的

直觉和经验,完全依赖机器的建议和决策,可能导致人类决策者对机器的错误不加质疑,从而使问题变得更加严重^[57]。同时,过度信任可能使人类对机器的依赖性增强,削弱对自己能力的信心,降低警觉性和应对突发情况的能力,导致在关键时刻无法独立做出正确的决策^[58]。以医疗为例,人工智能辅助诊断系统的应用,可能导致医生过度信任机器的建议,忽视自己的专业判断、出现误诊^[59]。在人机协作的情境中,人类可能对机器的决策能力过于乐观,从而导致其过分依赖机器,更倾向于接受机器的建议,这可能导致过度信任现象^[60]。

人机协同决策也同样受到信任不足的威胁。信任不足可能导致人类决策者过多地干预系统,从而降低整体决策效率^[61]。人机信任不足的典型表现是“算法厌恶”,即尽管算法通常能做出优于人类的决策,但人类对机器算法充满不信任从而排斥其意见^[28]。人对机器信任不足可能有2方面原因,一是因为目前很多机器学习的模型(尤其是深度学习模型)的决策过程过于复杂,导致人类决策者难以理解,从而降低对机器的信任。二是人类对机器错误的过度敏感,即人类认为机器不应该犯错,所以对机器错误的容忍度较低^[53]。

为了实现有效的人机协同决策,需要帮助用户在理解机器的能力和局限的基础上,建立适度的信任。为此,可以从以下几方面考虑解决方案。首先,提高算法透明度,设计更加透明且能让人易懂、速懂的人工智能算法,让用户更容易理解算法的决策过程,从而帮助建立信任关系。研究者正在开发可解释的人工智能技术(explainable AI),如使用可视化技术对决策依据进行展示,或在人类行为概念和语言框架内提供解释。值得注意的是,过多解释信息可能引发信息过载,反而不利于提升协同效能^[62]。系统透明度并不总是与用户信任度正相关,过高的透明度甚至可能影响用户察觉重大错误的能力^[63]。其次,通过教育和培训,让人类理解机器算法性能和限制,有助于消除不必要的恐惧和怀疑。另外,有效的人机界面设计也可以促进沟通、提高用户对当前情况的理解,从而建立信任^[57]。最后,信任的建立也受社会因素的影响,研究表明,当机器表现出类似人类的特征时,如情感能力,人类更容易建立信任^[64]。通

过将这些特征纳入未来的决策支持系统,可以更高效地将人机信任调整到适度范围。

2.3 人机决策权力分配问题

如何在人类和机器之间平衡决策权力,从而充分利用各自的优势实现高效的协同决策是人机协同决策领域的重要议题^[65-66]。当机器在决策过程中占据主导地位时,可能会导致人类过分依赖技术,引发系统安全风险等问题^[67];但如果人类占据绝对领导地位,则无法发挥机器的性能优势^[66]。

目前,大部分人机协作情境中,决策权仍然由人类决策者主导^[66,68]。在一项经济风险决策实验中,研究人员把人机团队分为“人类主导组”和“人机平权组”,结果显示,人类主导组中人类决策相对人机平权组更激进,并且在遭遇失败后也会延续冒险风格,这可能是由于人类主导组中决策者更有控制力^[66];而人机平权组中人类激进决策显著更少,但是部分协同决策的选择又过于保守。该研究结果表明,人机决策权力的分配会显著影响决策的策略,从心理学的角度,除信任因素外,这可能是由于人类存在对环境或自身行为的控制需求。这种需求源于人类对稳定性和可预测性的追求,当人类认为自己对某个过程或结果具有控制力时,往往感到更加安全和自信,从而更容易冒险^[69]。

合理的人机权力分配应具备动态性和灵活性。为了实现这一目标,研究人员提出从机器角度提升的方法,如自适应自动化系统,它可以根据实时的工作环境、任务需求调整决策权力,从而避免争夺决策权带来的风险^[70-71]。从人的角度出发,可以采取一系列措施,例如,增加人类在决策过程中的参与感受,满足人类的控制需求^[72];提供有关机器智能及其安全性的教育和培训等;根据人类状态和外部条件来调整决策权力分配,从而实现更高效和安全的协同决策。在人工智能更加发达和值得信赖的未来,可能会看到机器获得更多决策权,但是需要明确人机协同的重要目标是无论任务由谁主导,都应该注重人机之间的互动和合作,而不是完全由一方主导。未来的人机决策权力分配会朝着更加平衡、互补的方向发展,以让人类和机器各自发挥自己的优势。

2.4 人机道德对齐问题

随着机器被赋予越来越多的决策权力,人机道

德的差异问题日益凸显。当前人类对机器道德判断机制理解不全面,人类会质疑机器决策的道德性^[73];同时,人类在与机器共同承担责任时,人类可能出现道德逃避行为,试图将责任推给机器^[60]。

机器对人类价值观和道德的忽视主要是由机器过度理性、机器决策的偏见导致的。机器过度理性是由于其通常遵循预先设定的算法,其决策往往依赖于数学模型,这可能导致它们在处理复杂、模糊或多面性问题时过于理性。人类在类似情况下往往会运用直觉、经验和道德观念来平衡多个目标,然而机器可能缺乏这种平衡能力,这种过度的理性可能导致忽视人类的情感、道德观念、文化习惯等非理性因素,忽略人类价值观和道德需求^[28]。一个典型的例子是在自动驾驶汽车的道德困境中,机器可能会在减少伤害的同时违背道德原则,例如,牺牲一个无辜者的生命来保护更多人^[73]。同时,机器偏见问题一直以来受到社会关注,因为机器学习算法通常依赖于大量历史数据进行训练,然而这些数据可能包含了人类社会普遍存在的偏见和歧视,这可能导致机器在决策过程中无意中放大这些偏见,导致不公平和歧视性的结果。有研究人员发现,在面部识别系统中,一些商业面部识别算法在识别非洲裔和女性面孔时的准确率较低^[74],这是因为训练数据集包含了更多的白人和男性面孔,这种偏见导致了潜在的歧视问题。为了解决机器过于理性的问题,研究人员尝试将情感、道德和其他非理性因素纳入机器学习模型中,这包括在自动驾驶汽车领域引入道德规则^[75],以及开发具有情感智能的机器人与人进行互动。机器偏见的解决方案则包括使用无偏见的训练数据、更新算法以消除偏见,并确保算法的公平性^[76]。

在人机协同中道德责任归属是复杂的问题,特别是在出现错误或损害时,决策不仅会影响到参与者本人收益,也会影响其他参与者的收益。有研究人员通过联合决策实验来模拟人类与机器协同决策的场景,决策不仅会影响参与者本人收益也会影响其他参与者的收益,结果显示,当人类与机器共享责任时,其更有可能做出更自私的决策而忽略其他人的利益。这表明与机器共享责任会导致人类产生道德逃避行为,即倾向于将责任归咎于机器^[60]。针对此问题,不少学者呼吁出台相关新法律、明确法律框

架,以确定人类和机器在不同情况下的责任^[77]。

为了实现更高质量的人机协同决策,需要关注机器决策中的过理性和偏见问题,明确人机责任界

定,从法律制定的层面上填补相关空缺,尽可能统一道德标准,减轻该问题的影响。

综上,人机协同中的挑战现状如表1所示。

表1 人机协同决策挑战现状

挑战类别	理想人机协同决策	人机协同决策现状
情感问题	机器能够准确识别、理解并响应人类的情感状态,动态调整以适应人类情感需求	机器情感识别准确性和情境适应性有限,复杂情感识别存在挑战
信任问题	人类对机器的信任度适中,理解机器的能力和局限,既不完全依赖机器,也不完全弃用机器	人类对机器过度信任或信任不足
权力分配问题	人机决策权力分配具备动态性和灵活性,能够根据实时环境和任务需求进行调整,人类和机器发挥各自优势	人类主导决策,动态自适应调整能力欠缺
道德对齐问题	机器能够在决策过程中平衡理性与人类的情感、道德观念,避免偏见和歧视,明确人机协同决策责任归属	机器决策主要依赖偏向理性的预设算法,人类的情感和道德观念关注不足;没有建立明确的责任归属

3 展望

随着人工智能技术的持续演进,人机协同决策的模式正经历深刻变革。2017年7月8日,国务院印发《新一代人工智能发展规划》,明确了人工智能作为新一代科技革命和产业变革核心动力,在国家战略中的关键地位。规划针对2030年中国新一代人工智能发展提出了战略目标,将人机协同的混合增强智能列为人工智能发展的核心方向之一^[78]。在可预见的未来,人机协同将引领生产和服务领域的创新模式,这对人机协同决策系统的要求也相应提升。

在此背景下,人机协同决策系统不仅需要具备高效的执行能力,更需要考虑减少协作时的摩擦,这涉及到人机情感问题、人机信任问题、决策权力分配问题,以及责任划分等道德问题。从机器的角度,需要具备情感智能,以便深入理解和适应人类的情感。机器还需要提高可解释性,使用户能理解其决策过程及依据,以形成可信可靠的人机协同决策模式^[79]。未来的人工智能系统还应具备泛化能力,适应不同的领域和场景,自适应地根据人类和机器的实时性能、信任度和需求进行权力调整,实现更高效和安全的协同决策。此外,需明确人类与机器共同决策时的相关法律和制度以预防道德分歧问题。在人机协同决策的未来,不仅需要提升协同决策的能力,还需要赋予机器更高的社会智能和情感理解能力,综合考量多方面因素,以使协同决策更加优化,例如,使

机器能够更深入地理解社交规则、礼仪和人际关系,从而实现和谐的人机互动^[80]。

从人类的角度,政府和企业需要采取相应措施,如推动教育和培训,使人类决策者适应新的技术和工作岗位,确保人类在人工智能时代的竞争力和价值^[81]。为实现人机协同决策,人类需要了解机器在不同领域的能力边界,以充分利用其优势并避免过度依赖^[6]。同时,还需考虑人的适应问题,如适应不断变革的智能化工作方式,处理可能出现的心理问题等^[82]。面对通往人机共生的过程中不断涌现的新问题,需要从人的角度应对这些挑战,以实现更高效、安全和可持续的人机协同决策。

展望未来,新技术有望显著提升人机协同效能,尤其是脑机接口(brain-computer interface,BCI)技术将成为关键推动力。BCI技术通过直接读取和解释大脑信号,实时捕捉用户的意图和情感状态,并将其反馈给机器,使其能够更准确地理解人类的情感、需求和意图,从而实现更加自然和高效的人机交互^[83]。这种实时交互不仅可以显著提高决策的速度和准确性,还能显著增强人机互动的自然性和流畅性,这些技术进步将为未来人机共生时代奠定坚实基础,并推动人机协同效能的全面提升^[84]。

综上所述,未来的人机协同决策研究应该注重多学科融合,以应对复杂多变的决策环境,以更好地实现人机互理解和互适应,从而迈向人机共生的新时代。

参考文献(References)

- [1] 黄海丰, 刘培森, 李擎, 等. 协作机器人智能控制与人机交互研究综述[J]. 工程科学学报, 2022, 44(4): 780-791.
- [2] Licklider J C R. Man-computer symbiosis[J]. IRE Transactions on Human Factors in Electronics, 1960(1): 4-11.
- [3] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
- [4] Openai. Gpt-4 technical report[EB/OL]. [2023-08-05]. <https://cdn.openai.com/papers/gpt-4.pdf>.
- [5] Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639): 115-118.
- [6] 何贵兵, 陈诚, 何泽桐, 等. 智能组织中的人机协同决策: 基于人机内部兼容性的研究探索[J]. 心理科学进展, 2022, 30(12): 2619-2627.
- [7] Sheridan T B. Human-robot interaction: Status and challenges[J]. Human Factors, 2016, 58(4): 525-532.
- [8] Huang M H, Rust R T. Artificial intelligence in service[J]. Journal of Service Research, 2018, 21(2): 155-172.
- [9] 孙效华, 张义文, 秦觉晓, 等. 人机智能协同研究综述[J]. 包装工程, 2020, 41(18): 1-11.
- [10] Raymond R C. Use of the time-sharing computer in business planning and budgeting[J]. Management Science, 1966, 12(8): B-363-B-381.
- [11] Turban E. The use of mathematical models in plant maintenance decision making[J]. Management Science, 1967, 13(6): B-342-B-358.
- [12] Power D J. Decision support systems: A historical overview [M]//Handbook on Decision Support Systems 1. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008: 121-140.
- [13] Davis G B. Management information systems: Conceptual foundations, structure, and development[M]. New York: McGraw-Hill, Inc., 1984.
- [14] Power D J. Decision support systems: Concepts and resources for managers[M]. Westport: Quorum Books, 2002.
- [15] Kunz J C, Shortliffe E H, Buchanan B G, et al. Computer-assisted decision making in medicine[J]. The Journal of Medicine and Philosophy, 1984, 9(2): 135-160.
- [16] Sprague R H. A framework for the development of decision support systems[J]. MIS Quarterly, 1980, 4(4): 1-26.
- [17] Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making[J]. Big Data, 2013, 1(1): 51-59.
- [18] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. Cambridge: MIT Press, 2018.
- [19] Demner-Fushman D, Chapman W W, McDonald C J. What can natural language processing do for clinical decision support? [J]. Journal of Biomedical Informatics, 2009, 42(5): 760-772.
- [20] Basso B, Antle J. Digital agriculture to design sustainable agricultural systems[J]. Nature Sustainability, 2020, 3: 254-256.
- [21] Cropley D H, Kaufman J C, Cropley A J. Measuring creativity for innovation management[J]. Journal of Technology Management & Innovation, 2011, 6(3): 13-30.
- [22] Guilford J P. Creativity: Yesterday, today and tomorrow[J]. The Journal of Creative Behavior, 1967, 1(1): 3-14.
- [23] Lake B M, Ullman T D, Tenenbaum J B, et al. Building machines that learn and think like people[J]. The Behavioral and Brain Sciences, 2017, 40: e253.
- [24] Gigerenzer G, Gaissmaier W. Heuristic decision making[J]. Annual Review of Psychology, 2011, 62: 451-482.
- [25] Woolley A W, Chabris C F, Pentland A, et al. Evidence for a collective intelligence factor in the performance of human groups[J]. Science, 2010, 330(6004): 686-688.
- [26] Brynjolfsson E, McAfee A. The second machine age: Work, progress, and prosperity in a time of brilliant technologies [M]. New York: W.W. Norton & Company, 2014.
- [27] Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases[J]. Science, 1974, 185(4157): 1124-1131.
- [28] Dietvorst B J, Simmons J P, Massey C. Algorithm aversion: People erroneously avoid algorithms after seeing them err [J]. Journal of Experimental Psychology General, 2015, 144(1): 114-126.
- [29] Davenport T H, Kirby J. Only humans need apply: Winners and losers in the age of smart machines[M]. New York: Harper Business, 2016.
- [30] Simon H A. A behavioral model of rational choice[J]. The Quarterly Journal of Economics, 1955, 69(1): 99.
- [31] Squire L R. Memory systems of the brain: A brief history and current perspective[J]. Neurobiology of Learning and Memory, 2004, 82(3): 171-177.
- [32] Lerner J S, Keltner D. Beyond valence: Toward a model of emotion-specific influences on judgement and choice[J]. Cognition & Emotion, 2000, 14(4): 473-493.
- [33] Lim J, Wu W C, Wang J J, et al. Imaging brain fatigue

- from sustained mental workload: An ASL perfusion study of the time-on-task effect[J]. *NeuroImage*, 2010, 49(4): 3426–3435.
- [34] Minsky M. *Society of mind*[M]. New York: Simon and Schuster, 1988.
- [35] Lerner J S, Li Y, Valdesolo P, et al. Emotion and decision making[J]. *Annual Review of Psychology*, 2015, 66: 799–823.
- [36] Corti K, Gillespie A. A truly human interface: Interacting face-to-face with someone whose words are determined by a computer program[J]. *Frontiers in Psychology*, 2015, 6: 634.
- [37] McDuff D, El Kaliouby R, Senechal T, et al. Automatic measurement of ad preferences from facial responses gathered over the Internet[J]. *Image and Vision Computing*, 2014, 32(10): 630–640.
- [38] Breazeal C. Toward sociable robots[J]. *Robotics and Autonomous Systems*, 2003, 42(3/4): 167–175.
- [39] Franti E, Ispas I, Dragomir V, et al. Voice based emotion recognition with convolutional neural networks for companion robots[J]. *Science and Technology*, 2018, 20(3): 222–240.
- [40] Castellano G, Villalba S D, Camurri A. *Recognising human emotions from body movement and gesture dynamics*[M]// *Affective Computing and Intelligent Interaction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 71–82.
- [41] Eyben F, Weninger F, Gross F, et al. Recent developments in openSMILE, the Munich open-source multimedia feature extractor[C]//*Proceedings of the 21st ACM International Conference on Multimedia*. New York: ACM, 2013: 835–838.
- [42] Koelstra S, Muhl C, Soleymani M, et al. DEAP: A database for emotion Analysis; Using physiological signals[J]. *IEEE Transactions on Affective Computing*, 2012, 3(1): 18–31.
- [43] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality [EB/OL]. [2023–05–06]. <https://arxiv.org/abs/1310.4546v1>.
- [44] Zhang Z P, Luo P, Loy C C, et al. Learning deep representation for face alignment with auxiliary attributes[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(5): 918–930.
- [45] Hu X, Chen J J, Wang F, et al. Ten challenges for EEG-based affective computing[J]. *Brain Science Advances*, 2019, 5(1): 1–20.
- [46] Calvo R A, D’Mello S. Affect detection: An interdisciplinary review of models, methods, and their applications[J]. *IEEE Transactions on Affective Computing*, 2010, 1(1): 18–37.
- [47] Hu X, Wang F, Zhang D. Similar brains blend emotion in similar ways: Neural representations of individual difference in emotion profiles[J]. *NeuroImage*, 2022, 247: 118819.
- [48] Riek L D, Rabinowitch T C, Chakrabarti B, et al. How anthropomorphism affects empathy toward robots[C]//*Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*. New York: ACM, 2009: 245–246.
- [49] Rafaeli A, Vilnai-Yavetz I. Emotion as a connection of physical artifacts and organizations[J]. *Organization Science*, 2004, 15(6): 671–686.
- [50] Bickmore T W, Picard R W. Establishing and maintaining long-term human-computer relationships[J]. *ACM Transactions on Computer-Human Interaction*, 2005, 12(2): 293–327.
- [51] Riek L D. Healthcare robotics[J]. *Communications of the ACM*, 2017, 60(11): 68–78.
- [52] Freedy A, Weltman G, Freedy E, et al. Adaptive delegation interfaces (ADI) for improved situation awareness and reduction of workload in controlling multiple unmanned vehicles (UV) [EB/OL]. [2023–03–01]. https://www.researchgate.net/publication/235131830_Adaptive_Delegation_Interfaces_ADI_for_Improved_Situation_Awareness_and_Reduction_of_Workload_in_Controlling_Multiple_Unmanned_Vehicles_UV.
- [53] Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse[J]. *Human Factors*, 1997, 39(2): 230–253.
- [54] Schaefer K E, Chen J Y C, Szalma J L, et al. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems[J]. *Human Factors*, 2016, 58(3): 377–400.
- [55] Kok B C, Soh H. Trust in robots: Challenges and opportunities[J]. *Current Robotics Reports*, 2020, 1(4): 297–309.
- [56] Chen M, Nikolaidis S, Soh H, et al. Trust-aware decision making for human-robot collaboration: Model learning and planning[J]. *ACM Transactions on Human-Robot Interaction*, 2020, 9(2): 1–23.
- [57] Lee J D, See K A. Trust in automation: Designing for appropriate reliance[J]. *Human Factors*, 2004, 46(1): 50–80.

- [58] Merritt S M, Ilgen D R. Not all trust is created equal: Dispositional and history-based trust in human-automation interactions[J]. Human Factors, 2008, 50(2): 194-210.
- [59] Shortliffe E H, Sepúlveda M J. Clinical decision support in the era of artificial intelligence[J]. JAMA, 2018, 320(21): 2199-2200.
- [60] Kirchkamp O, Strobel C. Sharing responsibility with a machine[J]. Journal of Behavioral and Experimental Economics, 2019, 80: 25-33.
- [61] Madhavan P, Wiegmann D A. Similarities and differences between human-human and human-automation trust: An integrative review[J]. Theoretical Issues in Ergonomics Science, 2007, 8(4): 277-301.
- [62] Swar B, Hameed T, Reyachav I. Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search[J]. Computers in Human Behavior, 2017, 70: 416-425.
- [63] Poursabzi-Sangdeh F, Goldstein D G, Hofman J M, et al. Manipulating and measuring model interpretability[C]//Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2021: 1-52.
- [64] Nass C, Moon Y. Machines and mindlessness: Social responses to computers[J]. Journal of Social Issues, 2000, 56(1): 81-103.
- [65] Parasuraman R, Sheridan T B, Wickens C D. A model for types and levels of human interaction with automation[J]. IEEE Transactions on Systems, Man, and Cybernetics Part A, Systems and Humans, 2000, 30(3): 286-297.
- [66] Xiong W, Wang C, Ma L. Partner or subordinate? Sequential risky decision-making behaviors under human-machine collaboration contexts[J]. Computers in Human Behavior, 2023, 139: 107556.
- [67] Parasuraman R, Manzey D H. Complacency and bias in human use of automation: An attentional integration[J]. Human Factors, 2010, 52(3): 381-410.
- [68] Haesevoets T, De Cremer D, Dierckx K, et al. Human-machine collaboration in managerial decision making[J]. Computers in Human Behavior, 2021, 119: 106730.
- [69] Langer E J. The illusion of control[J]. Journal of Personality and Social Psychology, 1975, 32(2): 311-328.
- [70] Byrne E A, Parasuraman R. Psychophysiology and adaptive automation[J]. Biological Psychology, 1996, 42(3): 249-268.
- [71] Scerbo M W. Adaptive automation[M]//Neuroergonomics. Oxford: Oxford University Press, 2006: 239-252.
- [72] Bobadilla-Suarez S, Sunstein C R, Sharot T. The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses[J]. Journal of Risk and Uncertainty, 2017, 54(3): 187-202.
- [73] Bonnefon J F, Shariff A, Rahwan I. The social dilemma of autonomous vehicles[J]. Science, 2016, 352(6293): 1573-1576.
- [74] Buolamwini J, Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification[C]//Conference on fairness, accountability and transparency. New York: PMLR, 2018: 77-91.
- [75] Nyholm S, Smids J. The ethics of accident-algorithms for self-driving cars: An applied trolley problem? [J]. Ethical Theory and Moral Practice, 2016, 19(5): 1275-1289.
- [76] Barocas S, Selbst A D. Big data's disparate impact[J]. California Law Review, 2016, 104: 671.
- [77] Bryson J J, Diamantis M E, Grant T D. Of, for, and by the people: The legal *Lacuna* of synthetic persons[J]. Artificial Intelligence and Law, 2017, 25(3): 273-291.
- [78] 李睿晶, 房超, 陈凯. 新时代我国人工智能发展回顾与展望[J]. 科技智囊, 2023(1): 14-21.
- [79] Miller T. Explanation in artificial intelligence: Insights from the social sciences[J]. Artificial Intelligence, 2019, 267: 1-38.
- [80] Crandall J W, Oudah M, Tennom, et al. Cooperating with machines[J]. Nature Communications, 2018, 9(1): 233.
- [81] Arntz M, Gregory T, Zierahn U. The risk of automation for jobs in OECD countries: A comparative analysis[R]. Paris: OECD Publishing, 2016.
- [82] Sauer J, Hockey G R J, Wastell D G. Effects of training on short- and long-term skill retention in a complex multiple-task environment[J]. Ergonomics, 2000, 43(12): 2043-2064.
- [83] 张丹, 李佳蔚. 探索思维的力量: 脑机接口研究现状与展望[J]. 科技导报, 2017, 35(9): 62-67.
- [84] Gao X R, Wang Y J, Chen X G, et al. Interface, interaction, and intelligence in generalized brain-computer interfaces[J]. Trends in Cognitive Sciences, 2021, 25(8): 671-684.

Towards the era of human-machine symbiosis: A psychological perspective on human-machine collaborative decision-making

ZHU Ao¹, LIU Huashuo¹, YUAN Peijun^{2*}, ZHANG Dan^{1,3*}

1. Department of Psychology and Cognitive Sciences, Tsinghua University, Beijing 100084, China

2. Qiyuan Lab, Beijing 100095, China

3. Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China

Abstract Driven by the advancement in artificial intelligence, the efficiency of human-machine collaborative decision-making has been steadily improved. However, human-machine collaboration faces challenges such as emotional issues, trust issues, authority allocation, and ethical alignment. Emotional issues involve how machines overcome heterogeneity to identify and respond to human emotional states. Trust issues pertain to striking a balance between over-trust and under-trust on machine decisions. Authority allocation focuses on how to reasonably distribute decision-making power in human-machine collaboration so as to prevent excessive or insufficient machine influence. Ethical alignment ensures that machine decisions adhere to human ethical standards and values, safeguarding the ethicality and societal acceptance of decisions. Future directions include enhancing machines' understanding and responsiveness to human emotions, establishing robust trust mechanisms, optimizing authority distribution, and ensuring ethical alignment. On the path to human-machine symbiosis, it requires not only technological innovation but also the integration of psychological perspectives to address these challenges and achieve more efficient human-machine collaboration.

Keywords human-machine symbiosis; human-machine collaboration; collaborative decision-making; human-machine trust; psychological perspective ●



(责任编辑 王微)