

# 具身智能范式下人智协作的增强\*

——基于共享表征的视角

周洋 董梅梅 宋晓蕾

(陕西师范大学心理学院, 西安 710000)

**摘要:**智能时代的人智关系是一种人类和智能体作为队友合作的新型协作关系,共享表征是人智协作的基础,但以往少有研究从系统层面人智共享表征的视角考察其对人智协作绩效的影响。此项研究从具身智能范式出发,基于人际协同的共享表征模型,论述了人智协作中共享表征的不同维度和框架,考察了不同维度共享表征对人智协作绩效影响的作用规律、神经机制,并从应用方面通过共享表征的合理配置给出了增强人智组队协作绩效的途径。该研究结果将为设计更智能的人智协同作业系统提供科学基础,且可直接应用于无人机等人智组队的设计优化。

**关键词:**人智协作;共享表征;具身智能;协作绩效

中图分类号: B842.5

文献标志码: A

文章编号: 1003-5184(2025)06-0483-09

## 1 引言

随着人工智能技术的迅速发展,人与智能体的关系及协作模式持续演进迭代,人际协作任务正被人智协作任务逐渐取代,人机关系也演变为人类和智能体作为队友组队的一种新型人智协作关系(Shively et al., 2018)。共享表征可能是人与智能体有效协作的基础(宋晓蕾,董梅梅,2023),然而,以往关于人智协作的研究主要从智能体特征或操作者状态单一视角研究其对协作的影响作用,鲜少基于双方共同视角从系统层面探讨人智协作的机制和增强问题,而该问题正成为智能时代工程心理学的研究重点。

如何实现如人人协作般自然高效的人智协作?基于“以人为中心”的智能体发展原则(许为等,2024),人类操作者需要在一定任务场景中同时站在自己和智能体的视角,以一种功能类似的方式表征双方的目标和动作,并将对方目标动作整合进自己的认知系统,构建人与智能体的共享表征模型。然而,过去在大多数入智协作任务中,智能体的动作行为模式都是基于预先设定的算法和程序,无法真正联系物理世界,更无法直接形成人智共享表征。自2022年起,以ChatGPT、Sora等为代表的预训练大模型接连取得划时代突破,智能体得以实现从基础对话者(L1)向高级推理者(L2)的跨越。其核心

在于,它超越了信息应答,迈进了能够模拟人类认知、主动分析问题并制定结构化解决方案的阶段。但如何让智能体进一步提升为智能行动者L3,能够独立思考并根据复杂情境采取行动,从系统层面出发基于具身智能视角开展人智协作的研究是一个非常重要的突破口。具身智能通过“感知-行动”闭环将实现“虚拟智能”向“物理智能”的跨越。

基于此,研究中将对具身智能视角下不同维度共享表征对人智协作任务绩效影响的作用规律、神经机制及其绩效的增强进行系统探讨,以构建人智协作的理论模型。文中分析框架由理论构建与证据检验构成,具体而言,依托具身智能范式及人际协作的多重表征模型,系统构建人智协作的理论框架。另一方面,通过系统性的文献检索对该理论架构进行补充调整。

## 2 从具身智能范式开展人智协作的优势

### 2.1 具身智能的定义及特点

传统认知科学所主张的符号主义的失败,宣告了“离身心智”在科学上的“倒台”。受梅洛-庞蒂现象学所提出的身体在认知中占据核心地位观点的影响,“离身心智”转向“具身心智”。具身认知强调,身体是认知的主体,心智通过行动生成(Varela et al., 1991)。这为具身智能在复杂现实中实现奠定了核心基础。与以符号操作和知识库为手段实现

\* 基金项目:国家自然科学基金项目(T21922931,32171065),人因工程全国重点实验室项目(HFNKL2024W01)。

通信作者:宋晓蕾, E-mail: songxiaolei@snnu.edu.cn。

智能的符号主义,以及以神经网络为手段模拟人类思维过程的联结主义的传统智能范式不同,具身智能范式下智能体的智能依托感知-动作范式,由身体与环境的实时互动实现,其核心在于身体与环境的紧密耦合(Liu et al., 2025)。因此,具身智能范式是一种以身体-环境互动为核心的研究范式,这将推动智能体实现从“虚拟智能”向“物理智能”跨越。

## 2.2 具身智能范式对传统人机协作的超越

### 2.2.1 任务分配方式——从预设固定程序到自主动态调整

传统人工智能范式下,其算法依赖静态环境模型,动态任务分配能力弱。而现实场景充满不确定干扰,且任务链本身存在脆弱性,这两者的矛盾,导致了人智协作的低效。在具身智能范式下,这些问题得到一定程度的解决。多模态感知数据的协同融合与不确定性处理技术将支撑智能体在复杂现实场景中实现动态交互,而分层任务网络的任务分解机制和动态重规划算法可实现智能体任务结构的自适应调整,这将有效应对真实作业场景中的突发状态变更与实时需求变化。

### 2.2.2 感知交互方式——从单向走向双向的意图对齐

传统智能系统的主要局限,包括感知能力受限、认知理解不足以及人机交互模式固化,使得人机协作过程主要表现为单向的人对机器的控制与适应性调整。人类常常通过固定的指令集,例如“开始”、“暂停”等控制机器,或通过预设传感器触发固定动作。与之不同的是,在具身智能范式下,智能体将通过显式意图沟通——基于大语言模型实现的自然语言交互和高透明度、可解释性的输出设计,与人类达成一致理解,并通过隐式意图推断预测人类动作实现与人类协同。同时,人类协作者可以通过物理交互、多模态输入等方式向智能体补充信息,使人和智能体双方在动态交互过程中实时预测对方意图以调整自身动作反应。

### 2.2.3 智能体角色身份——从辅助工具到平等伙伴

受限于以往感知、算法等技术的有限发展,传统人工智能在人机协作中充当着人类指令响应者、计算程序执行者等工具化的角色。而通过模态融合感知、具身模拟、增量学习、强化学习等技术,以及“感

知-模拟-执行”机制和“印象-记忆-知识”模型整合,智能体可实现对复杂环境的准确感知以及长期自主学习和进化能力(沈甜雨等, 2025)。

## 3 具身智能范式下共享表征对人智协作的影响

### 3.1 共享表征的定义与内涵

对比黑猩猩,人类儿童有更强的合作动机以及理解他人协调自身的能力(Tomasello, 2014),这是对人类在合作中“共享”倾向和“表征”潜能的进化论视角的解释。共享表征囊括了认知过程和社会组织过程。认知科学认为共享表征是语言交流中心智模型的近似对齐(Ford & Johnson-Laird, 1985)。社会组织研究认为共享表征是集体心智的基础,是成员之间关于任务、角色和情境的分布式理解系统(Weick & Roberts, 1993)。而人智协作情境下共享表征被界定为执行协作任务的操作者从自身对自己目标,以及从智能体视角对智能体目标的动作表征,并基于此对智能体行为进行预测的认知加工过程(Sebanz et al., 2003)。人智协作任务完成的关键在于操作者对任务场景和智能体行为做出正确表征和预期,并做出快速反应(Andrews et al., 2023),这也意味着人智协作任务的完成可能依赖于操作者形成完整且准确的人智共享表征。

### 3.2 共享表征在人智协作中的作用机制

#### 3.2.1 人际协同中的共享表征

人智协作与人际协作的方式类似,而共享表征是其共同的基础(Yang et al., 2021)。宋晓蕾和董梅梅(2023)从表征视角揭示了人际协作的认知神经机制(见图1),并提出共享目标是人际协作的前提,共享表征是人际协作的基础,共享控制是人际协作的核心,以此完成人际以及人智协作任务。在人际协作过程中,表征内容主要体现在以下两个层面:首先是与低级感知层面的镜像神经网络(MNS)对应的共享本体和情境性表征,以完成对自我和他人以及共同目标和状态的感知和识别,其次是与高级认知层面心理理论网络(ToM)和抑制控制网络(ICS)对应的共享社会性表征,以实现对方行为的理解、预测和控制,以此完成人际协作任务(Yang et al., 2021)。人际协作认知神经机制的揭示可为人智协作的机制和增强研究提供坚实的理论基础。基于具身智能范式的引入,可以让操作者以智能体视角来操控人与智能体的关系和情境,通过提升其共享表征来提高人智协作绩效。

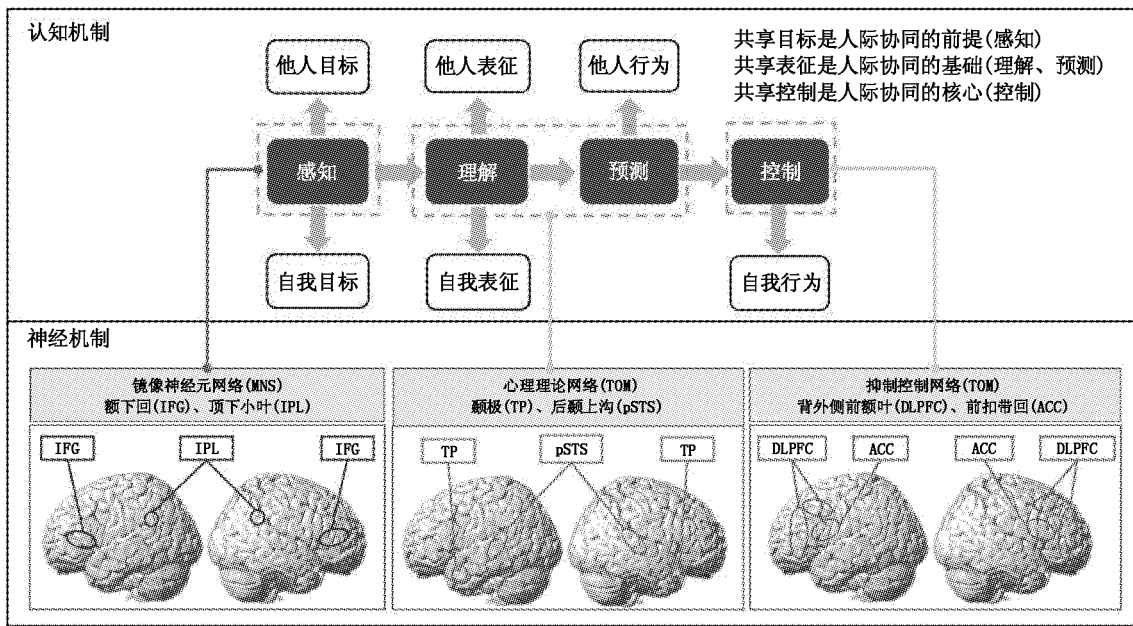


图 1 人际协作的认知神经机制(宋晓蕾,董梅梅,2023)

### 3.2.2 人智协作中的共享表征

#### 人智协作中的共享本体表征

本体表征通过整合多种感觉输入、运动指令和感觉反馈的匹配机制不断更新,个体因此能够适应不断变化的动作需求。Murata 等人(2012)的研究表明,前馈模型机制通过顶叶-前额叶网络的实时监测功能精确调节身体姿势和运动,同时还参与理解他人动作以及区分自我和他人。此外,本体表征的可塑性还能促进协作的学习行为。大脑通过将工具整合至表征中,可实现对工具的自然控制,甚至仅需短期训练即可灵活操作复杂工具(如机器人手)(Oztop et al., 2006)。

已有研究发现通过经皮电神经刺激(TENS)来增强身体图式可显著提升任务表现(Watanabe et al., 2018)。因此,可以通过构建身体图式,使机器人能自动更新身体模型,来适应动态环境和任务需求(Sugiura et al., 2006)。例如,Zheng 等人通过细粒度人类数字孪生建模实现了机器人对人类状态的实时感知(Zheng et al., 2023)。

#### 人智协作中的共享情境表征

实现自然流畅的人智协作需要解决两个核心挑战:环境动态性和人类行为不确定性带来的交互复杂性,以及建立可理解的跨主体沟通机制(Chevalier et al., 2022)。这些挑战的本质在于协作各方对情境的共同理解和表征能力。智能体的情境感知系统包含两个关键组成部分:多模态传感网络对环境状态和人类行为的实时采集,以及计算模型对不确定

性的建模与处理。

Moorman 等人(2025)提出的双向心理模型对齐框架,通过大语言模型实现任务信息的动态识别与交流,显著提升了协作绩效。Müller 等人(2024)则从机器人感知维度出发,开发了新的情境评估指标和学习机制,增强了智能体在动态环境中的适应能力。这些研究共同推进了共享情境表征的理论构建和实践应用。

#### 人智协作中的共享社会性表征

##### 共享信任

人机交互中,信任被定义为用户对 AI 系统的可靠性和安全性的信心。而系统透明度和决策可解释性是建立信任的关键因素,它们使人类用户能够理解 AI 的决策逻辑和预测其行为(Eke & Shuib, 2025)。

信任水平与人智协作绩效存在显著相关性。McNeese 等人(2021)发现高绩效团队普遍表现出更高的信任水平。但信任水平需要与任务需求相匹配。Hoff 和 Bashir(2015)指出,过度信任可能导致错误忽视,而信任不足则引发低效干预,只有适度依赖才能实现效率最大化。因此,人智共享信任水平的确定需要考虑具体任务的特殊性。

##### 角色表征

清晰的角色表征能使人类和机器对各自职责和分工更明确,减少任务冲突和重复劳动,从而提高协作绩效。何贵兵等人(2022)指出清晰的角色表征能够促进人机之间在认知、情感和价值层面的社会

性兼容。

这种基于共享社会表征的机制同样适用于动态任务环境,其中动态人机角色分配成为关键因素。例如,Merlo 等人(2022)通过实时状态监测实现角色动态调整,优化了社会性协作绩效。Zheng 等人(2023)证实,基于信任评估的角色转换符合社会协作的基本规律。此外,Chen 等人(2024)融合“心流”理论提出的动态角色适应模型,使活动挑战和用户技能之间达到平衡,提升了人智协作的长期效果。

### 3.2.3 人智协作的神经机制

#### 具身模拟的神经网络

具身模拟理论指出,认知的形成依赖于身体感觉运动系统的参与,大脑通过模拟身体的感觉和运动体验来理解和预测行为。其中,额下回(Inferior Frontal Gyrus, IFG)作为语言处理的核心脑区,不仅与动作观察和模仿密切相关,还在动作相关的语言理解中发挥关键作用。在抽象句义理解时,IFG 与运动前区的协同激活表明语言理解依赖于动作模拟机制(Vijayaraghavan et al., 2025)。此外,IFG 在整合视觉、本体感觉和语言指令时起到枢纽作用,支持具身智能模型的泛化学习能力。其神经活动直接反映多模态信息的对齐质量,进而影响动作预测的准确性(Huang et al., 2025)。可以推测,在具身智能范式中,IFG 的功能模拟可优化人智交互中的语言-动作映射,提升协作任务的适应性和预测精度。

具身模拟表征是大脑构建概念知识的基础,顶下小叶(Inferior Parietal Lobule, IPL)在其中负责整合动作和空间信息,并且在动作观察、意图理解及动作执行中均发挥关键作用。Kilner 等人的研究发现,在观察他人动作时,IPL 的活动增强,并通过预测编码机制帮助大脑模拟和理解动作意图(Kilner et al., 2007)。这一机制在智能体协作中尤为重要。Metta 等通过模拟 IPL 的功能使机器人能更准确地预测人类动作意图,并在协作任务中表现出更高的灵活性和适应性(Metta et al., 2010)。可以推测,具身智能范式下的智能体设计可借鉴 IPL 的预测编码机制,优化动作规划和实时交互能力,使其在动态环境中更自然地协同人类行为。

颞上回(Superior Temporal Gyrus, STG)通过连接语言和动作信息,使身体部位动词(如“踢”激活腿部运动区,“捡”激活手部运动区)的理解直接关联运动系统(Pulvermüller, 2005)。在具身智能范式

中,STG 的模拟可强化语言-动作的耦合机制,使智能体在接收语言指令时自动激活相应的运动表征,从而提高交互的自然度。例如,在服务机器人或协作机器人中,STG 的功能建模可优化其对口头指令的解析能力,使其动作执行更符合人类预期。

综上所述,IFG、IPL 和 STG 分别从语言-动作整合、意图预测和感知-运动耦合的角度支持具身模拟理论的核心机制。在具身智能研究中,模拟这些脑区的功能可能显著提升智能体的协作能力,使其在语言理解、动作预测和任务执行等方面更接近人类水平。未来研究可进一步探索多脑区协同计算模型,以构建更高效、自适应的人智协作系统。

#### 心理理论的神经网络

心理理论(Theory of Mind, ToM)是指个体通过理解他人心理状态来预测其行为的能力,这种能力对于增强协作双方的理解度和提升行动流畅性具有关键作用。神经科学研究发现,这一高级社会认知功能依赖于多个脑区的协同工作,其中后上颞沟(Posterior Superior Temporal Sulcus, pSTS)、IFG 和 IPL 构成了核心神经网络。

pSTS 在社会互动中扮演着信息输入枢纽的角色,专门负责处理动态社会信息,包括他人动作、面部表情和意图识别。在观察和模仿情绪表情时,pSTS 与 IFG 的共同激活证实了其在情绪共情中的桥梁作用(Carr et al., 2003)。最新 EEG 研究进一步揭示,pSTS 在刺激呈现后 120~160 ms 即对情绪信息进行快速编码(Ely & Ambrus, 2025),这种快速处理机制为实时社会互动提供了神经基础。在具身智能系统中,模拟 pSTS 的这种快速社会信息处理能力,可以显著提升智能体对人类非语言线索的识别速度和准确性。

IFG 是情绪共享与意图理解的关键。Preston 和 de Waal 提出的知觉-动作模型指出,IFG 通过共享表征将他人的情绪转化为自身体验(Preston & de Waal, 2002),这一机制为交互中的意图理解提供了神经解释。在智能体设计中,模拟 IFG 的这种情绪共享机制,可以使机器更好地理解人类的情感状态,从而做出更符合社会预期的反应。

IPL-IFG 协同是行为预测的计算基础。IPL 在观察他人动作时被特异性激活,并与 IFG 形成功能耦合,共同构建动作-意图的共享表征系统。这一神经机制已被成功应用于机器 ToM 模型的开发,如 ToMnet 通过元学习模拟 IPL-IFG 的协同工作机

制,实现了对智能体目标和信念状态的准确预测(Rabinowitz et al., 2018)。在具身智能领域,这一发现为开发具有高级社会认知能力的协作机器人提供了重要的设计思路。

理解上述脑区在心理理论中的具体作用,可为开发新一代具身智能的智能体提供了明确的神经科学依据。未来研究可以深入探索 pSTS - IFG - IPL 网络的动态交互模式,开发更精细的神经计算模型来模拟这些脑区的协同工作机制。最终将这些发现应用于服务机器人、医疗辅助机器人等需要高度社会智能的领域,进一步提升人智协作的自然性和效率。

#### 抑制控制的神经网络

抑制控制是指个体对目标无关的刺激、认知和行为反应进行自主调控的能力,这一功能对目标导向行为的执行至关重要。抑制控制主要依赖于前额叶 - 扣带回神经网络的协同工作,其中背外侧前额叶皮层(DLPFC)、前扣带回皮层(ACC)和眶额皮层(OFC)构成了核心调控系统。

DLPFC 在自我控制行为中发挥关键作用。Hare 等人(2009)的食物选择实验表明,当个体抵制诱惑选择健康食物时,DLPFC 活动显著增强,证实了该脑区在抑制诱惑和维持目标导向行为中的重要性。Mirabella(2014)进一步将行为抑制过程划分为五个阶段,指出 DLPFC 主要负责决策评估和行为抑制的执行。因此,在基于具身智能的人智协作情境中,DLPFC 的功能模拟可以帮助智能体更好地维持任务目标,抑制无关干扰。

ACC 通过独特的预期控制机制调节行为。Shenhav 等人(2016)提出,背侧 ACC(dACC)通过计算预期收益与成本的差值来实施认知控制。White 等人(2018)则明确了从预期编码到行为执行的前扣带回皮层 - 隔核 - 皮质神经回路。这表明了 ACC 在认知控制中的核心作用,这一机制对人智协作中的决策优化具有重要意义。

OFC 基于价值评估指导决策行为。Shi 等人(2023)发现,OFC 通过构建社会认知地图,以利益最大化为原则选择合作或竞争策略。OFC 损伤会导致价值判断能力受损,且在停止信号任务中表现出更长的反应时(Brockett & Roesch, 2021),证实了其在行为抑制和任务转换中的关键作用。在涉及具身智能的人智协作系统中,模拟 OFC 的价值评估功能或可优化合作策略的选择。

综上所述,DLPFC 的功能模拟可增强智能体的目标维持能力,ACC 的冲突监测机制可优化协作决策过程,OFC 的价值评估模型可提升合作策略的适应性。未来研究应着重探索这三个脑区的动态交互模式,并将其计算原理应用于关于具身智能的人智协作系统的开发,特别是在需要复杂决策和快速任务转换的场景中。

## 4 基于共享表征视角人智协作绩效的增强

### 4.1 TMS 神经工效增强共享本体表征

经颅磁刺激(Transcranial Magnetic Stimulation, TMS)通过磁场诱导的电脉冲可特异性调控目标脑区的神经活动。TMS 干预能够有效增强本体感觉表征,这一效应主要体现在提升身体运动感知的精确性和动作控制的协调性两个方面。

视觉信息向精细动作的转化依赖于腹侧前运动皮层(PMv)与初级运动皮层(M1)构成的功能环路。Chiappini 等人(2024)研究发现,TMS 诱导的 PMv - M1 环路增强能够促进运动共鸣效应,提升运动系统对他人动作的敏感性,从而优化协作中的动作理解能力。这一机制与本体感觉神经表征的精确编码密切相关,因为动作理解的质量直接依赖于对自身及他人运动状态的准确表征。Sengupta 等人(2024)的研究也发现,抑制性 TMS 干预右侧顶内沟皮层(rPPC)可有效降低空间位置的选择偏差。这表明 TMS 通过调控运动相关神经网络来增强本体感觉表征,进而提升协作行为表现。

### 4.2 设计工效增强共享社会性表征

#### 4.2.1 协作架构和任务分配设计——基于角色表征

在协作系统中,角色分配构成社会性表征的核心维度,其通过明确各主体的职责边界来优化任务执行效能。基于社会认知理论,角色界定不仅避免了协作冲突和任务冗余,更重要的是实现了优势互补,这一机制在动态任务场景中通过角色切换得以延续。当前多通过整合多模态感知与意图识别技术来构建社会情境模型,实现角色自适应参数的动态优化。

#### 4.2.2 社会信号与拟人化设计——基于共享信任

在人工智能系统设计中,社会信号的表达与拟人化特征的设计共同构成了信任建立的重要基础。拟人化特征促使用户将人类特质投射至智能系统,进而影响了信任评估。Pinto 等人(2025)的研究表明,类人表情和动作通过提供直观的社会线索,增强

了系统状态的透明度和可预测性,显著提升了用户的信任体验。Schreiter 等人(2022)在自主导航系统的研究中发现,简单的头部运动模拟关注行为就能有效改善信任评价,这一结果印证了基本社会信号在社会性表征构建中的价值。而从系统设计的角度看,有效的拟人化策略需要整合多维度特征:表达性特征构建社会互动框架,识别性特征实现社会情境理解,而物理特征则塑造社会实体感知(Pinto et al., 2025)。这些设计要素通过协同作用,共同完善了用户对 AI 系统的社会性表征,为建立稳定的人智信任关系提供了认知基础。

#### 4.3 自适应自主性增强共享情境表征

在具身智能范式中,智能体通过主动感知调整实现与环境的动态耦合,这一过程本质上构建了共享的情境表征。智能体通过动作执行(如调整摄像头角度、改变身体方位等)主动改变感知输入,不仅可以优化信息获取效率,更能建立与环境状态同步的表征系统(Liu et al., 2025)。这种具身交互模式通过物理接触实现了多模态感知的整合,为共享情境表征提供了物质基础。

Chevalier 等人(2022)提出的预测模型框架,通过融合历史数据与实时状态,使智能体能够自主模拟环境演变过程,这种预测能力本质上是共享情境表征的时间维度延伸。Tan 等人(2020)通过主动探索获取的最优视角,能够产生与环境结构更匹配的语义表征,这反映了共享情境表征的空间构建机制。

而不同探索策略的有效性差异揭示了共享情境表征的构建路径多样性。Ramakrishnan 等人(2020)的研究显示,探索范式的选择取决于环境特征与任务需求,这种适应性正体现了共享情境表征的动态调节特性。Shen 等人(2025)系统性地整合了多模态数据融合、风险预测和轨迹规划,通过建立多层次的环境模型,实现了共享情境表征的完整构建与实时更新。

上述研究共同表明,具身智能体通过主动感知、环境预测和探索策略的协同作用,能够构建并维护与物理环境同步的共享情境表征。这种表征不仅支持智能体的自主决策,更为人机协作提供了共同的理解基础。未来研究应进一步探索共享情境表征的神经计算机制及其在复杂环境中的适应性表现。

### 5 未来研究展望

随着深度学习技术,特别是大语言模型的快速发展,以及轻量化技术带来的显著成本优化(如

DeepSeek - R1),人工智能正逐步实现基础设施化转型。这一技术演进主要体现在三个维度:首先,大语言模型通过任务规划、多模态感知和逻辑推理等能力,使智能体具备了参与真实情境交互的认知基础;其次,脑机接口、柔性电子、分布式传感网络和量子计算等硬件技术的突破,极大地拓展了人机协作的物理边界;最终,这些技术进步共同推动了人智关系中智能体从传统的工具角色迈向伙伴角色的根本性转变。

通过系统回顾具身智能范式的演进历程,深入分析共享表征在人智协作中的作用机制。研究表明,人智共享表征有效解决了人、智能体和环境三者之间的交互融合问题,为人智协作提供了必要的认知基础。当前工程心理学研究仍坚持以人为中心的设计理念,忽略了具身智能体在协作过程中的潜在影响,未来研究需要通过动态任务分配实现人与智能体能力的优势互补,从而确保协作质量和效率。此外,透明化设计和情感化交互对建立长期稳定的人智信任关系具有重要作用,有助于提升团队协作的持续效能。基于这些发现,研究提出了基于共享表征的增强模型,以期对未来研究提供新的理论框架。

当前研究基于共享表征视角从系统整体上对人智协作绩效的提升给出了增强方案,但还未能纵深揭示具体任务场景下人类和智能体协作中共享表征的构建及其传递通路,特别是如何加速这一过程,未来或许可基于图式的形式进行研究训练,丰富人和智能体共同表征存储库,促进部分人智协作任务的自动化,从而提高协作效率。

特别需要注意的是,对于人智协作任务的失败及其产生的后果问题还缺乏系统研究。尤其是具身智能范式下的人智协作使智能体从工具性辅助转向平等协作伙伴,这一变化将重塑许多重要问题的思考方式。例如,动态任务分配中的责任归属机制及其对协作绩效的影响,未来需要基于自我效能感和责任扩散理论以及激励机制设计的实证研究来实现责任和绩效的平衡。其次,由智能体导致的任务失败进而引发的信任问题的修复机制尚不明确。最后,过度信任导致的人类认知惰性及其对创造性工作的抑制效应也亟需深入探究。未来这些问题的解决将为人智协作系统的优化提供重要理论基础。

周洋和董梅梅是本文的共同第一作者。

## 参考文献

- 何贵兵, 陈诚, 何泽桐, 崔力丹, 陆嘉琦, 宣泓舟, 林琳. (2022). 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. *心理科学进展*, 30(12), 2619 – 2627.
- 沈甜雨, 陶子锐, 王亚东, 张庭祯, 刘宇航, 王兴霞, 杨静, 李志伟, 陈龙, 王坤峰, 王飞跃. (2025). 具身智能研究的关键问题: 自主感知、行动与进化. *自动化学报*, 51(1), 43 – 71. <https://doi.org/10.16383/j.aas.c240364>.
- 宋晓蕾, 董梅梅. (2023). 人际协同的多重表征模型: 基于认知表征的视角. *心理科学进展*, 31(7), 1288 – 1302.
- 许为. (2024). 人智交互: 以人为中心人工智能的跨学科融合创新. 清华大学出版社. ISBN:9787302669678.
- Andrews, R. W., Lilly, J. M., Srivastava, D. K., & Feigh, K. M. (2023). The role of shared mental models in human – AI teams: A theoretical review. *Theoretical Issues in Ergonomics Science*, 24, 129 – 175. <https://doi.org/10.1080/1463922X.2022.2061080>.
- Brockett, A. T., & Roesch, M. R. (2021). The ever – changing OFC landscape: What neural signals in OFC can tell us about inhibitory control. *Behavioral Neuroscience*, 135(2), 129 – 137. <https://doi.org/10.1037/bne0000412>.
- Carr, L., Iacoboni, M., Dubeau, M. – C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), 5497 – 5502. <https://doi.org/10.1073/pnas.0935845100>.
- Chen, H., Alghowinem, S., Breazeal, C., & Park, H. W. (2024). Integrating flow theory and adaptive robot roles: A conceptual model of dynamic robot role adaptation for the enhanced flow experience in long – term multi – person human – robot interactions. \* *Proceedings of the 2024 ACM/IEEE International Conference on Human – Robot Interaction* \* (pp. 116 – 126). <https://doi.org/10.1145/3610977.3634945>.
- Chevalier, P., Schadenberg, B. R., Aly, A., Cangelosi, A., & Tapus, A. (2022). Context – awareness in human – robot interaction: Approaches and challenges. In \* *2022 17th ACM/IEEE International Conference on Human – Robot Interaction (HRI)* \* (pp. 1241 – 1243). IEEE. <https://doi.org/10.1109/HRI53351.2022.9889584>.
- Chiappini, E., Turrini, S., Zanon, M., Marangon, M., Borgomaneri, S., & Avenanti, A. (2024). Driving Hebbian plasticity over ventral premotor – motor projections transiently enhances motor resonance. *Brain Stimulation*, 17(2), 211 – 220. <https://doi.org/10.1016/j.brs.2024.02.011>.
- Eke, C. I., & Shuib, L. (2025). The role of explainability and transparency in fostering trust in AI healthcare systems: A systematic literature review, open issues, and potential solutions. *Neural Computing and Applications*, 37(4), 1999 – 2034. <https://doi.org/10.1007/s00521-024-10868-x>.
- Ely, M. M., & Ambrus, G. G. (2025). Shared neural dynamics of facial expression processing. *Cognitive Neurodynamics*, 19(1), 45. <https://doi.org/10.1007/s11571-025-10230-4>.
- Ford, M., & Johnson – Laird, P. N. (1985). Mental models: Towards a cognitive science of language, inference, and consciousness [Review of the book *Mental Models*, by P. N. Johnson – Laird]. *Language*, 61(4), 897. <https://doi.org/10.2307/414498>.
- Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self – control in decision – making involves modulation of the vmPFC valuation system. *Science*, 324(5927), 646 – 648. <https://doi.org/10.1126/science.1168450>.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407 – 434. <https://doi.org/10.1177/0018720814547570>.
- Huang, S., Chen, L., Zhou, P., Chen, S., Jiang, Z., Hu, Y., Liao, Y., Gao, P., Li, H., Yao, M., & Ren, G. (2025). *EnerVerse: Envisioning Embodied Future Space for Robotics Manipulation* (arXiv: 2501.01895). arXiv. <https://doi.org/10.48550/arXiv.2501.01895>.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, 8(3), 159 – 166. <https://doi.org/10.1007/s10339-007-0170-2>.
- Liu, H., Guo, D., & Cangelosi, A. (2025). Embodied intelligence: A synergy of morphology, action, perception and learning. *ACM Computing Surveys*, 57(7), 1 – 36. <https://doi.org/10.1145/3717059>.
- McNeese, N. J., Demir, M., Chiou, E. K., & Cooke, N. J. (2021). Trust and team performance in human – autonomy teaming. *International Journal of Electronic Commerce*, 25(1), 51 – 72. <https://doi.org/10.1080/10864415.2021.1846854>.
- Merlo, E., Lamon, E., Fusaro, F., Lorenzini, M., Carfi, A., Mastrogiovanni, F., & Ajoudani, A. (2022). Dynamic human – robot role allocation based on human ergonomics risk prediction and robot actions adaptation. In *2022 International Conference on Robotics and Automation (ICRA)* (pp. 2825 – 2831). IEEE. <https://doi.org/10.1109/ICRA46639.2022.9812438>.
- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., von Hofsten, C., Rosander, K., Lopes, M., Santos – Victor, J., Bernardino, A., & Montesano, L. (2010). The iCub humanoid robot: An open – systems platform for research

- in cognitive development. *Neural Networks*, 23 (8), 1125 – 1134. <https://doi.org/10.1016/j.neunet.2010.08.010>.
- Mirabella, G. (2014). Should I stay or should I go? Conceptual underpinnings of goal – directed actions. *Frontiers in Systems Neuroscience*, 8, 206. <https://doi.org/10.3389/fnsys.2014.00206>.
- Moorman, N., Zhao, M., Luebbbers, M. B., van Waveren, S., Simmons, R., Admoni, H., Chernova, S., & Gombolay, M. (2025). *Bidirectional mental model reconciliation for human – robot interaction with large language models* (arXiv:2503.07547). arXiv. <https://doi.org/10.48550/arXiv.2503.07547>.
- Müller, M., Ruppert, T., Jazdi, N., & Weyrich, M. (2024). Self – improving situation awareness for human – robot – collaboration using intelligent Digital Twin. *Journal of Intelligent Manufacturing*, 35 (5), 2045 – 2063. <https://doi.org/10.1007/s10845-023-02138-9>.
- Murata, A., Maeda, K., & Naito, E. (2012). Body schema as a link between motor control and cognitive function. In *2012 ICME International Conference on Complex Medical Engineering (CME)* (pp. 467 – 470). IEEE. <https://doi.org/10.1109/ICCME.2012.6275708>.
- Oztop, E., Lin, L., Kawato, M., & Cheng, G. (2006). Dexterous skills transfer by extending human body schema to a robotic hand. In *\* 2006 6th IEEE – RAS International Conference on Humanoid Robots \** (pp. 82 – 87). IEEE. <https://doi.org/10.1109/ICHR.2006.321367>.
- Pinto, A., Solovov, A., Simões, A. C., & Menezes, P. (2025). User acceptance in human – robot interaction: Exploring the role of anthropomorphic mechanisms in manufacturing environments – A systematic literature review. *Journal of Intelligent & Robotic Systems*, 111 (2), 46. <https://doi.org/10.1007/s10846-025-02248-x>.
- Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1 – 20. <https://doi.org/10.1017/S0140525X02000018>.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576 – 582. <https://doi.org/10.1038/nrn1706>.
- Rabinowitz, N. C., Perbet, F., Song, H. F., Zhang, C., Eslami, S. M. A., & Botvinick, M. (2018). *Machine theory of mind* (arXiv:1802.07740). arXiv. <https://doi.org/10.48550/arXiv.1802.07740>.
- Ramakrishnan, S. K., Jayaraman, D., & Grauman, K. (2020). *An exploration of embodied visual exploration* (arXiv:2001.02192). arXiv. <https://doi.org/10.48550/arXiv.2001.02192>.
- Schreiter, T., Morillo – Mendez, L., Chadalavada, R. T., Rudenko, A., Billing, E. A., & Lilienthal, A. J. (2022). *The effect of anthropomorphism on trust in an industrial human – robot interaction*. arXiv [cs. RO]. <https://doi.org/10.48550/arXiv.2208.14637>.
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, 88(3), B11 – B21.
- Sengupta, A., Banerjee, S., Ganesh, S., Grover, S., & Sridharan, D. (2024). The right posterior parietal cortex mediates spatial reorienting of attentional choice bias. *Nature Communications*, 15 (1), 6938. <https://doi.org/10.1038/s41467-024-51283-z>.
- Shen, T. – Y., Tao, Z. – R., Wang, Y. – D., Zhang, T. – Z., Liu, Y. – H., Wang, X. – X., Yang, J., Li, Z. – W., Chen, L., Wang, K. – F., & Wang, F. – Y. (2025). Key problems of embodied intelligence research: Autonomous perception, action, and evolution. *Acta Automatica Sinica*, 51 (1), 43 – 71. <https://doi.org/10.16383/j.aas.c240364>.
- Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, 19 (10), 1286 – 1291. <https://doi.org/10.1038/nn.4384>.
- Shi, W., Meisner, O. C., Blackmore, S., Jadi, M. P., Nandy, A. S., & Chang, S. W. C. (2023). The orbitofrontal cortex: A goal – directed cognitive map framework for social and non – social behaviors. *Neurobiology of Learning and Memory*, 203, 107793. <https://doi.org/10.1016/j.nlm.2023.107793>.
- Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., & Johnson, W. W. (2018). Why human – autonomy teaming? In C. Baldwin (Ed.), *Advances in neuroergonomics and cognitive engineering* (pp. 3 – 11). Springer. [https://doi.org/10.1007/978-3-319-60642-2\\_1](https://doi.org/10.1007/978-3-319-60642-2_1).
- Sugiura, K., Matsubara, D., & Katai, O. (2006). Construction of robotic body schema by extracting temporal information from sensory inputs. In *\* 2006 SICE – ICASE International Joint Conference \** (pp. 302 – 307). IEEE. <https://doi.org/10.1109/SICE.2006.315696>.
- Tan, S., Liu, H., Guo, D., Zhang, X., & Sun, F. (2020). *Towards embodied scene description* (arXiv:2004.14638). arXiv. <https://doi.org/10.48550/arXiv.2004.14638>.
- Tomasello, M. (2014). The ultra – social animal. *European Journal of Social Psychology*, 44(3), 187 – 194. <https://doi.org/10.1002/ejsp.2015>.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
- Vijayaraghavan, P., Queißer, J. F., Flores, S. V., & Tani, J. (2025). Development of compositionality through interactive learning of language and action of robots. *Science Robotics*, 10

- (98), eadp0751. <https://doi.org/10.1126/scirobotics.adp0751>.
- Watanabe, T., Fujiwara, T., & Suzuki, S. (2018). NIRS – SPM analysis of body schema modification and performance of body motion. In *2018 11th International Conference on Human System Interaction (HSI)* (pp. 369 – 374). IEEE. <https://doi.org/10.1109/HSI.2018.8430778>.
- White, M. G., Panicker, M., Mu, C., Carter, A. M., Roberts, B. M., Dharmasri, P. A., & Mathur, B. N. (2018). Anterior cingulate cortex input to the claustrum is required for top – down action control. *Cell Reports*, 22(1), 84 – 95. <https://doi.org/10.1016/j.celrep.2017.12.023>.
- Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 38(3), 357 – 381. <https://doi.org/10.2307/2393372>.
- Yang, Q., Song, X., Dong, M., Li, J., & Proctor, R. W. (2021). Underlying neural mechanisms of interpersonal situations on collaborative ability: A hyper – scanning study using functional near – infrared spectroscopy. *Social Neuroscience*, 16(5), 549 – 563.
- Zheng, P., Li, S., Fan, J., Li, C., & Wang, L. (2023). A collaborative intelligence – based approach for handling human – robot collaboration uncertainties. *CIRP Annals*, 72(1), 1 – 4. <https://doi.org/10.1016/j.cirp.2023.04.057>.

## Enhancing Human – Agent Collaboration under the Embodied Intelligence Paradigm: From the Perspective of Shared Representation

Zhou Yang Dong Meimei Song Xiaolei

(School of Psychology, Shaanxi Normal University, Xi'an 710000)

**Abstract:** In the intelligent era, the human – agent relationship evolves into a novel collaborative form, defined by humans and agents working together as teammates. While shared representation serves as the foundation of human – agent collaboration, few studies have investigated how shared representation at the systemic level influences human – agent team performance. Based on the embodied intelligence paradigm and drawing on the shared representation model from interpersonal collaboration, this study elucidates the distinct dimensions and frameworks of shared representation in human – agent teamwork. It investigates the functional principles and neural mechanisms through which different dimensions of shared representation influence collaborative performance. Furthermore, from an applied perspective, it proposes pathways to enhance human – agent team performance through the strategic configuration of shared representation. The findings of this research offers a foundation for the future design of intelligent human – agent collaborative systems, paving the way for optimizing the design of human – agent teams such as those involving unmanned aerial vehicles (UAVs).

**Key words:** human – agent collaboration; shared representation; embodied intelligence; collaborative performance