心理科学进展 2027, Vol. 35, No. 1, 33–43 https://doi.org/10.3724/SP.J.1042.2027.0033
Advances in Psychological Science https://cstr.cn/32111.14.2027.0033

·研究构想(Conceptual Framework)·

# 引入 AI “新队友”：人–AI 协同自主性对人–AI 协作的影响机理*

樊子立[1,2] 程 垦[3] 朱梦霆[4]

([1]湖北工业大学经济与管理学院; [2]湖北数字工业经济发展研究中心, 武汉 430068)
([3]浙江工业大学管理学院, 杭州 310023) ([4]上海交通大学安泰经济与管理学院, 上海 200030)

**摘 要** 随着信息技术的发展, 越来越多的工作需要人类与 AI 协同完成。以往研究多从“工具”视角探讨 AI 对人–AI 协作的影响, 却较少关注到, 随着 AI 技术自主性的日益增强, AI 对于人类而言正从“工具”转变为“队友”。面对这一关系变革带来的挑战, 如何确保人–AI 协作的有效性已成为亟待解决的问题。本研究跳出以往孤立地从 AI 技术系统分析 AI 与人类协作的做法, 把人类也纳入分析框架中, 聚焦人–AI 协同自主性的 4 个维度, 即适应性学习、不确定性、不透明性与开放性, 基于角色理论开展三项研究：研究 1 指出不确定性维度和不透明性维度导致的角色冲突会降低人–AI 契合度, 进而阻碍人–AI 协作; 研究 2 指出开放性维度通过角色领悟增强 AI 的心智韧性, 帮助 AI 理解和把握与人类协作过程中所需具备的技术和角色要求, 从而使 AI 与人类更加契合; 研究 3 指出适应性学习维度通过角色学习增强 AI 的拟人化程度, 推动人类更好地理解和接受 AI, 最终达到促进人–AI 协作的目的。本研究为组织成功引入 AI、促进人机共创、打造“超级个体”提供了理论依据和实践指南。

**关键词** 人–AI 协同自主性, 人–AI 协作, 角色冲突, 角色领悟, 角色学习
**分类号** B849:C93

## 1 问题提出

随着新一代信息技术的快速发展, 能动性 AI (agentic AI)在企业中得到了广泛应用(董雪艳 等, 2026; 李纯青 等, 2026)。[1]例如, 在 2026 年 4 月正式发布的 OpenClaw v0.9.3 可在全程无人工干预、无中间脚本、无本地模型加载的条件下进行现场工作流重构, 且可同时支持多个工作角色自治编排、实现跨工具链零配置直连等活动。在此趋势下, “人–AI 协作”这种共享、协调、分配、使用信息和资源的新型工作模式开始逐渐取代“人–人协作”的传统工作模式(Li et al., 2025; Romeo & Conti, 2026), 并被众多学者认为具有广阔的发展与应用前景(裴冠雄 等, 2026; Sidra & Mason, 2026; Yang et al., 2026)。当前关于人–AI 协作的研究大多将 AI 视作工具, 探讨其对人–AI 协作的影响(Raisch & Krakowski, 2021; Van Esch et al., 2018; Yim & Su, 2025)。然而, 随着 AI 技术自主性的日益提高, AI 对于人类而言正从“工具”逐步转变为“队友”。AI 在组织中的角色定位转变会对人–AI 协作产生什么影响？在这一转变下, AI 与人类应该如何开展双向互动与协作以确保人–AI 协作的有效性？目前鲜有研究对这些问题展开深入探讨。鉴于此, 本研究跳出以往的“工具”视角, 转而从“队友”视角切入, 并依据 Harris-Watson 等(2023)的观点, 提出 AI 与人类的有效协作一方面需要 AI 主动学习和适应人类角色, 另一方面也需

收稿日期：2026-02-19
* 国家自然科学基金青年项目(72402062)、湖北省教育厅哲学社会科学项目(22Q054)、大学生创新训练项目(20260100125)资助。
通信作者：程垦, Email: chengken1992@126.com
[1] 出于简明表达考虑, 在后续行文中, 除了国内外研究现状及评述部分, 我们使用“AI”来指代“能动性 AI”。

要人类愿意接受 AI 作为“新队友”。换言之，人−AI 协作不仅需要 AI 像人类队友一样具备“主体性”，还需要将人类也纳入到分析框架中，强调人类与 AI 的相互信任与协同互补。

据此，本研究引入“人−AI 协同自主性(human-AI collaborative autonomy)”这一概念来刻画 AI 与人类的交互。[2]与以往技术特征相比，人−AI 协同自主性跳出了从 AI 技术系统单方面探索 AI 与人类的互动的缺陷，转而将人−AI 联合认知系统作为基本分析单位。需要说明的是，本研究提出的“人−AI 协同自主性”是基于 Simmler 和 Frischknecht (2021)提出的“AI 技术自主性(AI technological autonomy)”发展而来的，后者是指 AI 自主感知外部信息，独立地对外部环境做出判断，开放地与其他智能体(包括人类)进行合作、学习和自我更新，无需人工干预的技术特征。既有研究将 AI 技术自主性作为一个整体并分为 5 个等级，探讨其等级高低会产生何种后续影响(Endsley & Mica, 2017; O’neill et al., 2022; Simmler & Frischknecht, 2021)，结果发现技术自主性的后续影响时好时坏，并不稳定。对此，本研究认为有必要跳出以往将自主性看作是一个孤立的、笼统的 AI 技术概念的做法，转而从人类与 AI 双方协同的视角来拓展这一主题。Harris-Watson 等(2023)和 Frischknecht (2021)也提出类似观点，认为在研究人−AI 协作的形成机制时可进一步构建人与 AI 的联合认知系统，从社会感知视角来解构 AI 技术自主性。受此启发，本研究对“AI 技术自主性”进行拓展，提出“人−AI 协同自主性”，并将其定义为“人类与 AI 在共同完成复杂任务时所涌现出的一种动态适应能力”，该能力源于 AI 的客观技术属性与人类的主观感知评估之间的深度耦合与递归校准，包括适应性学习、开放性、不确定性、不透明性四个维度，其中前两个维度属于技术属性，后两个维度属于感知属性。

进一步地，人−AI 协同自主性会如何影响人−AI 协作？本研究引入角色理论的观点，认为人−AI 协同自主性在社会互动的形式下会持续创造和重新定位角色(Sahoo & Varadarajan, 2025)，随着 AI 在互动协作中的角色从“工具”向“队友”转变，团队成员也会验证 AI 是否满足他们对新队友的期待，以此来重新定位 AI 在团队中的角色。具体而言，人−AI 协同自主性的不确定性维度和不透明性维度会让 AI 在与人类的协作过程中具有不可追踪和难以确定的特征，由此就可能使得双方在互相磨合的过程中存在角色冲突(role conflict)，影响双方的契合度和协作效率。人−AI 协同自主性的开放性维度会通过 AI 的角色领悟(role taking)来让 AI 系统进一步理解和把握与人类协作过程中所需具备的技术特质和角色要求(Bilal et al., 2025)，从而增强 AI 的心智韧性，以便 AI 更好地理解人类意图并展开协作。人−AI 协同自主性的适应性学习维度会通过角色学习(role learning)让人类团队能更好地理解和接受自己(Cai et al., 2025)，即通过海量数据训练和人类反馈来习得扮演该角色所需的技能，增强自身的拟人化程度，最终达到提升人−AI 契合度，促进人−AI 协作的目的。简言之，从角色理论的视角看，人−AI 协同自主性的不确定性维度和不透明性维度会阻碍“AI 作为独立个体主动去适应人类角色”和“人类愿意接受 AI 作为新队友”这两个角色过程，而人−AI 协同自主性的开放性维度和适应性学习维度则会分别促进这两个角色过程。

综上所述，本研究将基于角色理论，聚焦人−AI 协同自主性的 4 个维度，以耦合关系中双方的“角色冲突→角色领悟→角色学习”的逻辑链条构建理论模型，旨在揭示在引入 AI“新队友”的情形下人−AI 协同自主性对人−AI 协作的影响机制。我们希望本研究不仅能在一定程度上弥补现有

[2] “自动化”、“智能水平”、“能动性”或“拟人化”等技术特征难以满足前述 AI 与人类的交互要求。首先，“自动化”停留在任务分配层面(Simmler & Frischknecht, 2021)，将技术系统完成任务的自动化程度由低到高分为 5 个等级，但这本质上仍需要人类来分配任务，AI 技术系统依然是以“工具”角色，而非以“队友”角色来进行协作(Tenakwah & Watson, 2025)；其次，“智能水平”通常会通过基准测试来度量，是一个静态的、去语境化的智力存量，它只能回答“AI 懂不懂”，但无法回答“AI 懂我懂不懂” (Dong et al., 2025)，这种单向的智力展示并不能满足将人类与 AI 作为联合认知系统的要求；再次，“能动性”强调 AI 主动发起行动、设定目标的能力，聚焦于 AI 单方面的主动权，难以满足人类与 AI 双方在协同互补过程中的“可干预性”与“可授予性” (Karim et al., 2025)；最后，“拟人化”关注 AI 在外观、语气或情感表达上像不像人类，更多表现在界面审美层面，无法提供人−AI 协作中所需要的人类长期、理性的功能性依赖和对 AI 的信任(Li & Sung, 2021)。

“人−AI 协作”研究的不足, 而且还能帮助人们深入理解如何识别其协作的前置因素, 从而为组织成功引入 AI“新队友”, 推动组织进行数智化转型提供理论依据。

## 2　国内外研究现状及评述

### 2.1　能动性 AI

作为一种能替代或帮助人们高效、精确地执行各种重复、繁琐、危险任务的智能实体, 能动性 AI 近年来受到学界的日益关注(Joshi, 2025)。能动性 AI 具有其它智能实体无法取代的特点。具体而言, 与传统的机器学习模型不同, 能动性 AI 不仅能够处理预先设定的任务, 还能够通过独立的思考和适应性学习来逐步完成未设定的任务(Dwivedi et al., 2026)。同时, 区别于单智能体(如 AI Agent), 能动性 AI 涉及更广泛的领域与架构方法, 常包含多智能体系统(multi-agent system)的编排(Abou Ali et al., 2025; Sapkota et al., 2025)。能动性 AI 的出现代表了 AI 从被动的“工具”向具有自主能动性的“队友”的转变。本研究中人−AI 协同自主性中的 AI 指的就是能动性 AI。结合能动性 AI 的实际应用场景及以往相关研究, 本研究将其定义为“一种能够主动感知环境、自主制定决策并执行动作, 且全程无需人工参与的智能实体”。

为了更好地理解能动性 AI, 有必要将其与相关概念进行比较分析。首先是大语言模型(LLM)。LLM 的发展推动了能动性 AI 的快速崛起, 但二者仍存在一定差异。第一, 自主决策与行动模式存在区别。能动性 AI 可围绕给定目标开展独立推理与自主行动, 能根据任务目标自主生成提示词(prompt); 而 LLM 的运行高度依赖人类输入的提示词, 输出质量受提示词的完备性约束, 即便具备基础推理能力, 面对复杂任务仍易产生推理谬误(Plaat et al., 2025)。第二, 外部工具调用的自主性不同。能动性 AI 具备开放的自主调用能力, 可主动调用外部工具拓展自身能力以求解复杂问题; 而 LLM 的工具使用必须依托人为配置干预, 自身无法主动配备。

机器人流程自动化(RPA)也与能动性 AI 有区别。在环境适配方面, RPA 只能在给定的条件下运行, 无法处理不可预测的未知环境; 而能动性 AI 可与环境交互, 感知信息并做出推理行动。从定位来看, RPA 专注于重复性、规则化的任务, 而能动性 AI 则重在模拟人类思维学习能力, 完成复杂自主决策。在原理方面, RPA 依托预设规则脚本完成自动化; 而能动性 AI 依靠学习算法从数据中自主学习执行任务。在技术特征上, RPA 以过程为中心, 重复执行既定指令; 而能动性 AI 以数据为中心, 可基于数据实现自我学习与改进。

最后是 ChatGPT。ChatGPT 是一种基于自然语言处理技术的 AI 系统, 其与能动性 AI 存在一些区别。首先, 知识运用模式不同。ChatGPT 的一大缺点在于, 其训练数据只截止到当年, 对于更新的知识内容无法直接做出回答。但能动性 AI 可根据自身已有的知识进行加工, 并根据实际情况进行推理和回答。其次, 自主性的程度不同。ChatGPT 需要用户来针对问题指定是否需要使用插件, 而能动性 AI 则具备自主调用工具的能力, 在获取到每一步子任务的工作后都会判断是否需要通过调用外部工具来完成该任务, 并在完成后获取该外部工具返回的信息提供给大模型, 进行下一步任务的工作。

### 2.2　人−AI 协同自主性

人−AI 协同自主性是本研究基于 AI 技术自主性发展而来的概念。AI 技术自主性反映了 AI 自主感知外部信息, 独立地对外部环境做出判断, 开放地与其他智能体进行合作、学习和自我更新, 且无需人工干预的技术特征(Simmler & Frischknecht, 2021)。AI 技术自主性使能动性 AI 可以真正以“队友”的角色融入团队。然而, 由能动性 AI 与人类组成的团队要想实现良好的协作绩效, 除了需要关注人−AI 协作的技术部分, 还需要关注人−AI 协作的互动耦合部分。以往研究指出, 在自主技术代理行为变得越来越棘手、不透明和难以理解的情况下, 迫切需要了解人们对自主性的反应(Alqurni, 2026; Frischknecht, 2021; Tan et al., 2026)。由此可见, 对于人类与能动性 AI 之间协作的理解, 不仅与能动性 AI 的技术特征有关, 而且还需考虑人类主观上如何感知 AI 技术系统。张振铎等(2024)的研究也支持了这一点, 当人类被要求判断算法的可信度和公平性时, 算法的实际行为与人类认为算法能做什么同等重要。即是说, 对自主性的社会感知在人−AI 协作中发挥着重要作用(Iapaolo & Lynch, 2026)。因此, 本研究在 Simmler 和 Frischknecht (2021)的基础上, 提出人−AI 协同自主性这一概念, 将其界定为“人

类与 AI 在共同完成复杂任务时所涌现出的一种动态适应能力”, 具体表现为能动性 AI 的技术属性与人类对能动性 AI 的感知属性之间的深度耦合与递归校准。

人−AI 协同自主性的独特价值在于, 其跳出了以往孤立地从 AI 技术系统出发与人类互动、以相互替代的逻辑来看待人−AI 协作的缺陷, 转而将“人−AI 联合认知系统”作为基本分析单位, 把人类也纳入到分析系统中, 强调双方的协同互补、交互信任、动态平衡(Frischknecht, 2021; Jia et al., 2024)。这就要求人−AI 协同自主性的核心维度中既包含 AI 系统本身的技术特征, 也包含人类对 AI 系统的心理感知。

首先, 人−AI 协同自主性中的能动性 AI 涉及系统性能的可追溯性和确定性的缺乏。在确定性系统中, 输入 A 总是相应地导致输出 B, 并且其间的所有执行步骤都是指定的且透明的。然而, 能动性 AI 的自主性使得它不需要所有执行步骤都是透明的, 这使得观察者或用户对详细的处理方式一无所知(Liu et al., 2022)。因此, 不透明性和不确定性是人−AI 协同自主性中反映人类对 AI 系统的心理感知的两个维度。

其次, 以往研究表明, 具备自主性的智能实体拥有适应不断变化的环境的能力(Dhiman et al., 2022)。这类智能实体能够主动处理信息, 对程序员之前输入的知识进行学习和延伸, 并改变其响应方式。这使得系统能够在没有人为干预的情况下适应并提高其在特定环境中的性能(Sharma et al., 2023)。因此, 适应性学习是人−AI 协同自主性的第三个维度, 且该维度反映的是 AI 系统的技术特征。

最后, AI 系统所具有的先进互动形式使得系统具有开放性特征。相关研究表明, 同其他智能实体协作以拓展数据来源, 是自主性的典型体现(Wang et al., 2023)。借助协作机制, 技术系统能够完成任务委派与分工, 形成多智能体系统, 且协作的开放性可以强化系统自主权。因此, 开放性是人-AI 协同自主性的第四个维度, 归属于 AI 系统的技术特征。

### 2.3 人−AI 协作

Larson 和 Dechurch (2020)指出, 如果智能实体在协作中发挥着独特的作用, 并对绩效做出了独特的贡献, 那么它就是协作队友而非是作为工具与人类互动。这意味着, 能动性 AI 必须被人类识别为代表团队中独特角色的不同实体, 否则这些角色就必须由人类来扮演, 也就没有了所谓的人−AI 协作。这一观点引出一个新问题, 即人类与能动性 AI 应该如何互动才能使能动性 AI 作为“队友”而不是“工具”来与人类协作? 在一项具有里程碑意义的研究中, Nass 等(1995)操纵了人类和自主智能体协同工作, 发现在相互依存的条件下, 被试会觉得自主智能体与自己更相似, 并觉得自主智能体的信息质量更高, 自己更容易受到自主智能体的影响。类似地, Tu 等(2025)也发现, 人−AI 协作中的结果相互依存性对工作绩效产生了强烈的影响。因此, 相互依存的工作内容和结果是人类将 AI 视为队友的第一项标准。进一步地, Wynne 和 Lyons (2019)提出, 人与智能实体要更好地协作, 除了在工作内容和结果上相互依存以外, 在当前“人−智协作”的新型团队中还需进行身份和角色的转变, 即智能实体必须摆脱掉以往“工具”的团队角色, 转而能够像人类成员一样独立地去工作, 这就需要智能实体具备“人性”, 即智能实体能表现出像人一样行动的属性, 例如能自主感知外部环境并在人类成员完全不参与的情况下独立地做出判断和采取决策。这一属性既符合本文中能动性 AI 的技术特征, 也与上述文献中对人−AI 协作的工作情境的描述相一致(Naikar et al., 2026)。因此, AI 要作为“队友”的角色与团队成员展开协作, 其独立性和自主性必不可少。换言之, AI 行动的独立性和自主性是人类将其视为队友的第二项标准。据此, 我们将人−AI 协作定义为“至少由一名人类和至少一个具有自主性的能动性 AI 进行协作, 它们相互独立且彼此依存地朝着一个共同的目标工作”。

能动性 AI 作为“队友”与人类开展协作, 在各个行业都发挥了非凡的作用(Harris-Watson et al., 2023), 但新队友的加入也会带来独特的挑战, 因为这是一个具有破坏性的成员变更事件(McKee et al., 2024; Trainer et al., 2020), 原本的人−人协作的成员可能会遇到与工作安全、隐私、监控等问题, 从而诱发人类对 AI 的普遍焦虑(Erengin et al., 2025; Samson & Zaleskiewicz, 2025)。鉴于此, 如何设计能动性 AI 以使其在发挥自身智能优势的同时也能被人类成员所接受是未来探索人−AI 协作的突破口。Harris-Watson 等(2023)指出, 能动

性 AI 要想有效地与人类协作, 一方面需要其主动适应人类角色, 另一方面也需要人类愿意接受其作为“新队友”。具体而言, 在能动性 AI 主动适应人类角色方面, 尽管能动性 AI 作为新队友进行人−AI 协作的相关研究刚刚起步, 但人类作为新人融入人−人协作的研究却已有几十年的历史。该研究领域将新队友的融入视为一个社会化过程, 在这个过程中人类成员必须调整现有的工作流程和状态以适应新队友。在需要整合能动性 AI 作为新队友的协作团队中, 我们可将其类比为人类新人, 因而能动性 AI 需要在协作团队中经历一个社会化过程。在人类愿意接受能动性 AI 作为“新队友”方面, 相关研究发现, 影响对新人接受度的因素主要包括团队反思、知识利用和心理接受(Rink et al., 2013), 这三大要素有助于澄清协作团队接受新来者的条件。本研究将接受度的相关研究从人−人协作扩展到人−AI 协作, 并预计对能动性 AI 的接受度也会涉及这三大要素。团队反思涉及到由于添加了新队友而调整团队流程和结构; 知识利用是对新队友独特的专业知识和技能的认可和采用; 心理接受反映了人类成员对新队友作为协作队员的态度和认可。

### 2.4　文献评述

结合上述文献回顾, 我们认为当前研究主要存在以下局限, 有待补充与完善。首先, 以往研究大多只关注 AI 技术本身的技术特征, 孤立地从 AI 技术系统出发与人类互动、以相互替代的静态逻辑来看待人−AI 协作。对此, 本研究拟将“人−AI 联合认知系统”作为基本分析单位, 把人也纳入到分析系统中, 并创新性地提出“人−AI 协同自主性”这一概念, 深入探讨人−AI 协同自主性对人−AI 协作的影响机理。

其次, 以往研究主要基于“工具”视角, 探索 AI 技术本身的优劣势, 忽略了随着 AI 技术“自主性”的精进, AI 应如何以独立的新队友身份融入人类群体？本研究将基于人与 AI 耦合的“队友”视角, 提出人−AI 协作取决于人类与 AI 的双向互动, 一方面需要 AI 主动适应人类角色, 另一方面也需要人类愿意接受 AI 作为“新队友”, 以期全面地揭示了保证人−AI 协作有效性的标准和条件。

最后, 以往研究大多在将 AI 视为一种工具的基础下探讨人类与 AI 的协作过程与结果。随着 AI 技术自主性的提高, 人类与 AI 的关系逐渐转向队友关系。面对这一崭新的关系模式, 人类与 AI 的协作过程黑箱还没有充分打开。本研究将基于角色理论, 识别人类与 AI 协作过程中可能出现的冲突、问题及其相应的应对策略, 以期为组织更精准地预测人−AI 协作提供理论支持。

## 3　研究构想

本研究旨在探索人−AI 协同自主性对人−AI 协作的影响机理。为实现这一目标, 我们基于角色理论构建了如图 1 所示的整体理论框架, 拟探讨：(1)人−AI 协同自主性的不确定性维度和不透明性维度对人−AI 协作的影响; (2)人−AI 协同自主性的开放性维度对人−AI 契合度的影响; (3)人−AI 协同自主性的适应性学习维度对人−AI 契合度的影响。

### 3.1　研究 1：不确定性和不透明性对人−AI 协作的影响

人−AI 协同自主性的不确定性维度和不透明性维度使得人类成员无法洞悉 AI 的内部操作流程和规律(Skulason et al., 2024), 这无疑增加了 AI 的不可追踪性和不可控制性。面对这一情形, 团

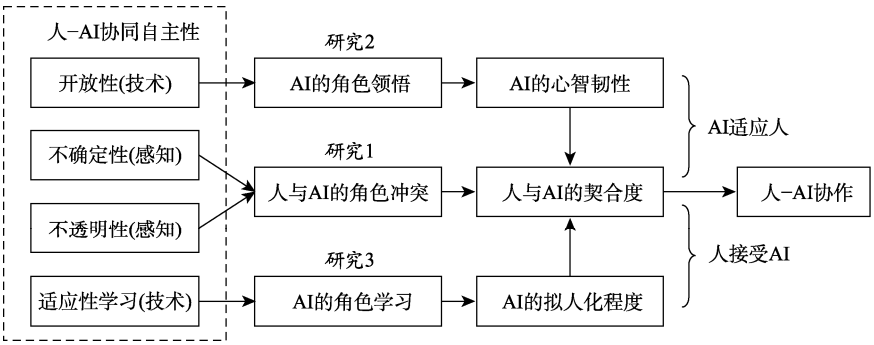



图 1　整体理论框架

队成员会担忧自己的工作内容是否会和 AI 的工作内容混淆、AI 是否会不经同意就插手自己的工作、在某些工作上 AI 是否比自己做得更好、自己是否失去了对 AI 的掌控、自己是否会被 AI 取代等(Marey et al., 2024)。这种工作边界的不清晰使人类成员越来越分不清哪些工作角色到底是属于自己还是属于 AI, 继而使人类成员自身的工作角色受到挑战, 使人类成员与 AI 之间产生角色冲突。据此, 本研究提出如下命题：

命题 1：人−AI 协同自主性的不确定性维度和不透明性维度会促进人−AI 的角色冲突。

进一步地, 由角色理论可知, 随着角色冲突的增加, 人与 AI 的关系也会变得越发紧张。人类成员对 AI 会有较强的防御意识和地盘保护意识, 以竞争的态度来看待 AI 这位“新队友”, 以防止其威胁到自己在组织中的工作角色和地位。在这种情形下, 人与 AI 的契合度也被逐渐损耗。紧接着, 随着人−AI 契合度的降低, 双方在工作内容、工作方式和处理日常事务的配合上会出现较大出入, 进而使 AI 不能很好地融入协作团队, 甚至在协作过程中出现诸多摩擦, 导致人与 AI 互动效率低下(Dai et al., 2025), 最终阻碍人−AI 协作。据此, 本研究得出如下命题：

命题 2：人−AI 协同自主性的不确定性维度和不透明性维度会通过促进人−AI 的角色冲突来降低人与 AI 的契合度, 进而阻碍人−AI 协作。

综上所述, 研究 1 的理论模型如图 2 所示。

### 3.2 研究 2：开放性对人−AI 契合度的影响

以往关于人−人协作的研究认为, 新加入的成员需要经历组织社会化过程, 从而更好地与现有成员相契合, 适应现有成员的工作流程和工作范式(Alqurni, 2026; Bilal et al., 2025)。本研究沿用这一观点, 提出人−AI 协同自主性的开放性维度会促使 AI 采取角色领悟来提高自身的心智韧性, 进而提升人与 AI 的契合度, 即开放性会促进 AI 主动适应人类角色。

具体而言, 随着 AI 技术从“自动化”向“自主性”过渡, AI 在团队中的角色从“工具”转变为“队友”, 这促使 AI 逐渐具备了像人类成员一样的“主体性” (Lund, 2026)。通过开放性这一技术特征, AI 会广泛地与其他智能实体(如人类)产生连接, 进一步理解和把握与人类协作过程中所需具备的技术特质和角色要求, 自主地在组织社会化过程中进行角色领悟来适应团队(Weigl et al., 2024)。据此, 本研究得出如下命题：

命题 3：人−AI 协同自主性的开放性维度会促进 AI 的角色领悟。

进一步地, 随着 AI 角色领悟的加深, AI 开始逐渐在思维模式和沟通风格等方面主动适应和配合与自己协作的人类成员, 与协作成员进行持续性的情感互动, 这类互动不仅限于回答问题或执行任务, 更强调营造“像人一样”的陪伴、共情或关系体验(Korentsides et al., 2024), 即 AI 的心智韧性。紧接着, 随着 AI 心智韧性的提高, AI 可以迅速确定自己在角色中需要承担的工作内容, 能更好地按照人类成员期望的方式与之进行协作, 并根据角色的工作内容和需求给予的反馈, 促进双方在工作习惯、流程和目标上更容易达成一致, 提升了人−AI 的契合度。据此, 本研究得出如下命题：

命题 4：人−AI 协同自主性的开放性维度会通过促进 AI 的角色领悟来提高 AI 的心智韧性, 进而增强人与 AI 的契合度。

综上所述, 研究 2 的理论模型如图 3 所示。

### 3.3 研究 3：适应性学习对人−AI 契合度的影响

在人−AI 协作团队中, 除了 AI 需要主动去适应人类角色之外, 人类也需要逐步接受 AI 作为新队友(Gao et al., 2026), 而在这一过程中, 人−AI 协同自主性的适应性学习维度起到了重要作用。具体而言, 人−AI 协同自主性中的适应性学习维度可以促使 AI 主动向人类寻求反馈, 降低信息的模糊性, 从而更新 AI 的认知图式并以此来帮助 AI 进行意义建构(Sinha et al., 2024)。随着意义构建的逐渐增强, AI 会主动界定工作边界, 使得人

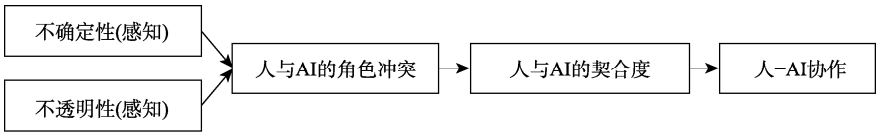



图 2 研究 1 的理论模型

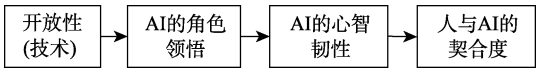



图 3　研究 2 的理论模型

类与 AI 对工作的目标、任务、内容等越来越清晰, 推动 AI 将角色的规范进行内化(Weigl et al., 2024), 促进 AI 的角色学习。据此, 本研究提出如下命题：

命题 5：人−AI 协同自主性的适应性学习维度会促进 AI 的角色学习。

进一步地, 适应性学习维度可通过 AI 的角色学习帮助 AI 更高效地提高自身的拟人化程度。AI 的角色学习可将 AI 的算法行为有效转化为可被人类观察、解读并赋予社会意义的行为序列(Cai et al., 2025)。与单纯增加“微笑表情”、“口语化语气词”等浅层行为不同, 这些行为不是经不起推敲的表面功夫, 而是在一次次真实的协同碰撞中依靠纠正偏差、情境妥协和记忆积累等来增强的 AI 拟人化行为, 且这种拟人化是较为稳固的、抗干扰的(Ramaul et al., 2026)。紧接着, 随着 AI 的拟人化程度越来越高, 人类成员也越来越能够读懂和理解 AI, 与之相处的模式更加融洽, 也更愿意去接纳 AI 的工作风格, 从而削弱了人与 AI 的角色冲突(Soprana, 2024)。双方无论是在工作流程、知识共享还是信任和共情上都能进行更好的交流和配合(Wu et al., 2024), 最终提升了人−AI 契合度。据此, 本研究提出如下命题：

命题 6：人−AI 协同自主性的适应性学习维度会通过促进 AI 的角色学习来提升 AI 的拟人化程度, 进而增强人−AI 契合度。

综上所述, 研究 3 的理论模型如图 4 所示。

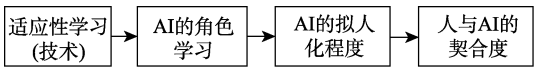



图 4　研究 3 的理论模型

## 4　理论建构

受到人−AI 协同自主性相关研究和角色理论的启发, 本研究依据人−AI 双方互动关系中“角色冲突→角色领悟→角色学习”的逻辑主线, 构建了一个“人−AI 协作模型”, 旨在探索人−AI 协同自主性会如何影响人−AI 协作。本研究的主要理论观点如下：

第一, 人−AI 协同自主性的不确定性维度和不透明性维度对人−AI 协作具有阻碍效应。在保障人−AI 协作、激发协作效率的问题上, 如果只考虑 AI 技术的优势对团队可能造成的影响, 而忽视了团队成员是否接受, 那么势必会在协作中形成冲突, 降低团队工作效率(Georganta & Ulfert, 2024)。基于角色理论, 本研究提出, 当 AI 的角色发生转变时(工具→队友), 人−AI 协同自主性中的不确定性和不透明性会对人−AI 协作产生阻碍效应。具体而言, 人−AI 协同自主性中的不确定性和不透明性会增加团队成员对 AI 的难以追踪性和不可控制性(Frischknecht, 2021), 这使得 AI 的内部处理系统像一个不可预测的黑箱, 加之 AI 精密的算法和高超的工作效率, 无疑会让团队成员产生技术替代感。人们会对“哪些工作由 AI 来做, 哪些工作由人来做”等问题产生困惑, 使得双方的角色产生冲突。由此, 团队成员会对 AI 产生较强的排斥心理和地盘意识, 继而削弱人−AI 协同自主性对人−AI 协作的影响效应。进一步地, 角色冲突的加深也会使人与 AI 协作产生许多分歧, 增加双方协作的难度, 降低人−AI 协作团队的工作效率, 从而损害整体绩效。上述研究紧密围绕人−AI 协同自主性的负面效应展开, 揭示了角色冲突在不确定性和不透明性两大维度影响人−AI 协作的过程中所发挥的作用, 由此不仅拓展了我们对人−AI 协同自主性的认识, 同时也丰富了我们对人−AI 协作的前置因素的理解。

第二, 人−AI 协同自主性的开放性维度在“AI 作为独立个体主动去适应人类角色”方面对人−AI 契合度具体促进作用。推动 AI 广泛地与其他智能实体(包括人类)进行连接是保障人−AI 协作、激发协作团队活力的重要措施。已有研究大多从“工具”视角来看待 AI, 鲜有相关研究从“队友”视角切入。本研究聚焦“队友”视角, 认为 AI 可以作为“新队友”主动与人类成员进行连接, 并以此来扩大原始数据和完成自我更新(Flathmann et al., 2026; Zhang et al., 2023)。主动适应人类角色的关键是要适应团队中的原有成员(Tu et al., 2025)。因此, 人与 AI 的契合度在开放性维度促进人−AI 协作的过程中发挥着重要作用。具体而言, 开放性可以使 AI 通过连接主动进行角色领悟, 把握人类成员所需要的技术特征和角色要求, 以团队成员更适应团队的方式来进行合作和沟通时, 进而提升人与 AI 的契合度。随着人−AI 契合度的提升, AI

越发能理解人类成员的意图, 也更能了解任务类型和处理方式, 进而游刃有余地展示任务所需要的知识和技能, 推动人−AI 协作的效率的提升。上述研究紧密围绕人−AI 协同自主性中开放性这一维度展开, 剖析了当 AI 角色发生转变后(工具→队友)应该如何保障人与 AI 的有效协作, 由此丰富了当前对于 AI 在团队中角色的理解。

第三, 人−AI 协同自主性的适应性学习维度在“人类愿意接受 AI 作为新队友”方面对人−AI 契合度具有促进作用。适应性学习是人−AI 协同自主性最重要的特点, 也是 AI 与其他智能体建立连接、相互协同、彼此平衡的基石(Frischknecht, 2021)。本研究认为, 适应性学习维度会推动 AI 采取角色学习, 帮助 AI 越来越像人(无论是能力方面, 还是情感方面), 即增强 AI 的拟人化程度, 进而促进人−AI 契合度。具体而言, 适应性学习维度会根据协作的需求对角色进行学习, 在角色学习中 AI 会通过纠正偏差、情境妥协和记忆积累等来将自身的算法行为转化为可被人类观察、解读并赋予社会意义的行为序列, 由此增强 AI 的拟人化程度。随着 AI 拟人化程度的提升, 从理性方面来看, 团队成员会重新审视 AI 这位“队友”, 更能够读懂并理解 AI 的算法语言, 愿意接纳 AI 的工作风格, 正视 AI 的知识、经验与才能, 并对其进行学习、迭代和更新, 以此来提升人−AI 契合度; 从感性方面来看, 较高的 AI 拟人化程度会使得感知到 AI 的团队成员从内心里真正接受 AI 作为他们的队友, 摆脱“工具”思维, 从“队友”视角来重新审视 AI 在团队中的地位和作用, 增强对 AI 的信任、共情、交流和配合(Kim & Hur, 2024; Naikar et al., 2026), 将其作为团队中的一份子, 从而提升人−AI 契合度。

需要说明的是, 人类与 AI 的协作是一个非常复杂的过程。囿于有限的研究能力和特定的研究偏向, 本研究只从队友视角, 基于角色理论, 对人−AI 协同自主性影响人−AI 协作的过程进行刻画。未来研究可从几个方面对本研究进行拓展。第一, 构建 AI 的协同能力体系。未来研究可聚焦于 AI 的交互智能, 系统考察其在适应性学习(如根据人类实时反馈进行动态微调)、开放性(如接纳人类临时变更的模糊目标)等维度的具体表现, 并在此基础上构建出人−AI 协作团队中 AI 的协同能力体系。第二, 识别人类心理感知模型。未来研究可将人−AI 协同自主性中的不确定性和不透明性作为关键中介变量, 探究“感知自主性”与“客观自主性”的偏差及其对协作绩效的影响, 以及 AI 的行为模式(如犯错后的解释风格等)如何影响人类对其心智模型的构建。第三, 挖掘人类与 AI 的双向交互与协商机制。未来研究可将角色理论进一步深度嵌入, 探析角色冲突在 AI 场景下的特殊表现(如指令矛盾、安全护栏与用户需求的冲突等), 也可探索交互式角色学习, 包括人类如何通过自然语言指令临时重塑 AI 的角色、AI 又如何通过主动提问来澄清角色模糊等。第四, 建立责任归属机制。未来研究可从单一的人与 AI 协作扩展到“多智能体+人类”的混合生态, 分析并填补“责任鸿沟”, 例如当协同决策出错时责任在 AI 的设计者、使用者还是部署者, 又如在面对环境剧变和 AI 性能骤降时能否通过人类接管或 AI 降级实现安全退化。第五, 探讨人−AI 协同系统迭代演化过程。未来研究一方面可以关注人−AI 团队记忆系统来探讨如何存储过往协作中的冲突与成功经验并在未来任务中调用, 另一方面也可结合 AI 演化组件, 探索 AI 如何在多轮协作中更新对人类偏好的内部表征, 将人类的情境感知与 AI 的数据记忆进行有机融合。第六, 考察人−AI 协作的旁观者效应。在组织中, 除了有人−AI 协作团队, 还存在常规的人−人协作团队, 未来研究可从第三方的角度探索人−AI 协作会如何影响人−人协作团队及其成员的心理和行为。

## 参考文献

董雪艳, 陈锋, 赵健宇, 熊轩. (2026). AI 技术可供性驱动国有企业财务数智化价值创新的过程机理——基于中石化的探索性案例研究. *管理世界, 42*(3), 198−228.

李纯青, 郝日艳, 刘伟. (2026). 人机共生体验的形成机制与作用效果. *心理科学进展, 34*(6), 932−952.

裴冠雄, 董波, 金佳, 孟亮, 张加林. (2026). 营销数字人对话智能特征的动态加工与神经机制. *心理科学进展, 34*(2), 227−238.

张振铎, 国佳宁, 李豪, 王宏蕾. (2024). 挑战还是阻断? 平台算法压力对数字零工主动服务行为的影响机制. *心理科学进展, 32*(11), 1768−1785.

Abou Ali, M., Dornaika, F., & Charafeddine, J. (2025). Agentic AI: A comprehensive survey of architectures, applications, and future directions. *Artificial Intelligence Review, 59*(1), 11−20.

Alqurni, J. (2026). Exploring the role of agentic AI in fostering self-efficacy, autonomy support, and self-learning motivation in higher education. *Frontiers in*

*Artificial Intelligence, 9*(5), 173−187.

Bilal, H., Khan, M. N., Khan, S., Shafiq, M., Fang, W., Khan, R. U., ... Xu, B. (2025). The role of artificial intelligence and machine learning in predicting and combating antimicrobial resistance. *Computational & Structural Biotechnology Journal, 27*(5), 423−439.

Cai, C., Zhu, G., & Ma, M. (2025). A systemic review of AI for interdisciplinary learning: Application contexts, roles, and influences. *Education & Information Technologies, 30*(7), 9641−9687.

Dai, R., Thomas, M. K. E., & Rawolle, S. (2025). The roles of AI and educational leaders in AI-assisted administrative decision-making: A proposed framework for symbiotic collaboration. *The Australian Educational Researcher, 52*(2), 1471−1487.

Dhiman, K. K., Kothari, M., & Abhishek. (2022). Autonomous control and transportation of underslung load with single and dual lift helicopter systems. *Journal of Dynamic Systems, Measurement & Control, 144*(4), 14−26.

Dong, L., Tang, X., & Wang, X. (2025). Examining the effect of artificial intelligence in relation to students' academic achievement: A meta-analysis. *Computers and Education: Artificial Intelligence, 8*, 100400.

Dwivedi, Y. K., Helal, M. Y. I., Elgendy, I. A., Albashrawi, M. A., Hughes, L., Shawosh, M., ... Jeon, I. (2026). Artificial intelligence agents and agentic systems in hospitality and tourism: Challenges, opportunities and research agenda. *International Journal of Contemporary Hospitality Management, 38*(1), 27−52.

Endsley, M. R., & Mica, R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors, 59*(1), 5−27.

Erengin, T., Briker, R., & De Jong, S. B. (2025). You, me, and the AI: The role of third‐party human teammates for trust formation toward AI teammates. *Journal of Organizational Behavior*. Advance online publication.https://doi.org/10.1002/job.2857.

Flathmann, C., Nguyen, H., Andre, K., & Guynup, C. (2026). The teammate divide: How humans approach teamwork differently with human and AI teammates. *Proceedings of the ACM on Human-Computer Interaction, 10*(1), 1−25.

Frischknecht, R. (2021). A social cognition perspective on autonomous technology. *Computers in Human Behavior, 122*, 106815.

Gao, B., Zuo, P., Xia, C., Zhang, T., Dong, S., & Liu, C. (2026). Impact of interaction diversity and interlayer coupling on the evolution of human-AI cooperation. *Applied Mathematics and Computation, 518*(8), 129−138.

Georganta, E., & Ulfert, A. S. (2024). My colleague is an AI! Trust differences between AI and human teammates. *Team Performance Management, 30*(1−2), 23−37.

Harris-Watson, A. M., Larson, L. E., Lauharatanahirun, N., DeChurch, L. A., & Contractor, N. S. (2023). Social perception in human-AI teams: Warmth and competence predict receptivity to AI teammates. *Computers in Human Behavior, 145*(3), 107765.

Iapaolo, F., & Lynch, C. R. (2026). Rethinking urban AI: A critical posthumanist perspective on intelligence, autonomy, and agency. *Urban Geography, 47*(1), 65−87.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal, 67*(1), 5−32.

Joshi, S. (2025). Advancing innovation in financial stability: A comprehensive review of AI agent frameworks, challenges and applications. *World Journal of Advanced Engineering Technology & Sciences, 14*(2), 117−126.

Karim, M. M., Van, D. H., Khan, S., Qu, Q., & Kholodov, Y. (2025). AI agents meet blockchain: A survey on secure and scalable collaboration for multi-agents. *Future Internet, 17*(2), 57−68.

Kim, W. B., & Hur, H. J. (2024). What makes people feel empathy for AI chatbots? Assessing the role of competence and warmth. *International Journal of Human–Computer Interaction, 40*(17), 4674−4687.

Korentsides, J., Keebler, J. R., Fausett, C. M., Patel, S. M., & Lazzara, E. H. (2024). Human-AI teams in aviation: Considerations from human factors and team science. *Journal of Aviation/Aerospace Education & Research, 33*(4), 63−72.

Larson, L., & Dechurch, L. A. (2020). Leading teams in the digital age: Four perspectives on technology and what they mean for leading teams. *The Leadership Quarterly, 31*(1), 101377.

Li, J. M., Zhang, L. X., & Mao, M. Y. (2025). How does human-AI interaction affect employees' workplace procrastination?. *Technological Forecasting & Social Change, 212*, 123951.

Li, X., & Sung, Y. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in user-AI assistant interactions. *Computers in Human Behavior, 118*, 106680.

Li, Y., Wu, B., Huang, Y., Liu, J., Wu, J., & Luan, S. (2025). Warmth, competence, and the determinants of trust in artificial intelligence: A cross-sectional survey from China. *International Journal of Human–Computer Interaction, 41*(8), 5024−5038.

Liu, B., Mazumder, S., Robertson, E., & Grigsby, S. (2022). AI autonomy: Self-initiation, adaptation and continual learning. *Computers in Human Behavior, 12*(5), 15−23.

Lund, E. (2026). Autonomy for whom? Implications of cyber-physical mining systems for operator work and organisation. *Applied Ergonomics, 132*, 104684.

Marey, A., Arjmand, P., Alerab, A. D. S., Eslami, M. J., Saad, A. M., Sanchez, N., & Umair, M. (2024). Explainability, transparency and black box challenges of AI in radiology: Impact on patient care in cardiovascular radiology. *Egyptian Journal of Radiology & Nuclear*

*Medicine, 55*(1), 145−162.

McKee, K. R., Bai, X., & Fiske, S. T. (2024). Warmth and competence in human-agent cooperation. *Autonomous Agents & Multi-Agent Systems, 38*(1), 23.

Naikar, N., Hoffman, R., Roth, E. M., Klein, G., Militello, L. G., & Dominguez, C. (2026). Should we make AI more tool-like or teammate-like?. *Journal of Cognitive Engineering & Decision Making, 20*(2), 137−157.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities?. *International Journal of Human-Computer Studies, 43*(2), 223−239.

O'neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human-autonomy teaming: A review and analysis of the empirical literature. *Human Factors, 64*(5), 904−938.

Plaat, A., Van Duijn, M., Van Stein, N., Preuss, M., Van Der Putten, P., & Batenburg, K. J. (2025). Agentic large language models, a survey. *Journal of Artificial Intelligence Research, 84*(7), 18−25.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *The Academy of Management Review, 46*(1), 192−210.

Ramaul, L., Ritala, P., Kostis, A., & Aaltonen, P. (2026). Rethinking how we theorize AI in organization and management: A problematizing review of rationality and anthropomorphism. *Journal of Management Studies, 63*(2), 761−807.

Rink, F., Kane, A. A., Ellemers, N., & Van der Vegt, G. (2013). Team receptivity to newcomers: Five decades of evidence and future research themes. *Academy of Management Annals, 7*(1), 247−293.

Romeo, G., & Conti, D. (2026). Exploring automation bias in human-AI collaboration: A review and implications for explainable AI. *AI & Society, 41*(1), 259−278.

Sahoo, L. K., & Varadarajan, V. (2025). Deep learning for autonomous driving systems: Technological innovations, strategic implementations, and business implications-a comprehensive review. *Complex Engineering Systems, 5*(1), 78−91.

Samson, K., & Zaleskiewicz, T. (2025). Warmth and competence in human-AI agent interactions. *Polish Psychological Bulletin, 56*, 148−160.

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. *Information Fusion, 126*, 103599.

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence, 5*(1), 46−57.

Sidra, S., & Mason, C. (2026). Generative AI in human-AI collaboration: Validation of the collaborative AI literacy and collaborative AI metacognition scales for effective use. *International Journal of Human Computer Interaction, 42*(7), 5084−5108.

Simmler, M., & Frischknecht, R. (2021). A taxonomy of human-machine collaboration: Capturing automation and technical autonomy. *AI & Society, 36*(1), 239-250.

Sinha, A., Singh, S., & Verma, H. K. (2024). AI-driven task scheduling strategy with blockchain integration for edge computing. *Journal of Grid Computing, 22*(1), 13.

Skulason, H., Whitby, A., Leingang, O., Vogl, W. D., Andre, P., Alzouni, T., & Schmidt-Erfurth, U. (2024). Using AI uncertainty for image quality assessment and gradability of retinal thickness in OCT. *Investigative Ophthalmology & Visual Science, 65*(7), 5637-5637.

Soprana, M. (2024). Compatibility of emerging AI regulation with GATS and TBT: The EU artificial intelligence act. *Journal of International Economic Law, 27*(4), 706−722.

Tan, X., Chun, D., Zhao, S., & Liu, Y. (2026). From manual delivery to autonomous delivery robots: A socio-technical push–pull–mooring framework. *Journal of Theoretical & Applied Electronic Commerce Research, 21*(1), 22.

Tenakwah, E. S., & Watson, C. (2025). Embracing the AI/automation age: Preparing your workforce for humans and machines working together. *Strategy & Leadership, 53*(1), 32−48.

Trainer, H. M., Jones, J. M., Pendergraft, J. G., Maupin, C. K., & Carter, D. R. (2020). Team membership change "events": A review and reconceptualization. *Group & Organization Management, 45*(2), 219−251.

Tu, Y., Li, J., Chen, J., Li, C. J., & He, W. (2025). When AI becomes my teammate: Unpacking how employees perceive and collaborate with gendered AI teammates. *Journal of Organizational Behavior*. Advance online publication. https://doi.org/10.1002/job.70038.

Van Esch, P., Black, J. S., & Ferolie, J. (2018). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior, 90*, 215−222.

Wang, Y., Chen, S., Ma, C., Tao, J., Yu, Y., Shan, W., & Zhao, B. (2023). A novel hybrid consensus-based variational bayesian filter with position uncertainty in multi-agent system. *Proceedings of the Institution of Mechanical Engineers, 237*(10), 2331−2340.

Weigl, L. M., Jabeen, F., Treur, J., Rob Taal, H., & Roelofsma, P. H. (2024). Learning for a Better Safety and Security Culture Within an Organization: Reducing the Risk in Communication with AI Coaching for Security Communication Through Cyberspace. In P. H. M. P. Roelofsma, F. Jabeen, H. R. Taal, & J. Treur (Eds.), *Using Shared Mental Models and Organisational Learning to Support Safety and Security Through Cyberspace: A Computational Analysis Approach* (pp. 235−303). Springer Cham.

Wu, Y., Kim, K. J., & Mou, Y. (2024). Minority social influence and moral decision-making in human-AI interaction: The effects of identity and specialization cues. *New Media & Society, 26*(10), 5619−5637.

Wynne, K. T., & Lyons, J. B. (2019). Autonomous agent teammate-likeness: Scale development and validation. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality. Applications and Case Studies* (pp. 199−213). Springer Cham.

Yang, C., Bauer, K., Li, X., & Hinz, O. (2026). My advisor, her AI, and me: Evidence from a field experiment on human–AI collaboration and investment decisions. *Management Science, 72*(1), 242−264.

Yim, I. H. Y., & Su, J. (2025). Artificial intelligence (AI) learning tools in k-12 education: A scoping review. *Journal of Computers in Education, 12*(1), 93−131.

Zhang, G., Chong, L., Kotovsky, K., & Cagan, J. (2023). Trust in an AI versus a human teammate: The effects of teammate identity and performance on human-AI cooperation. *Computers in Human Behavior, 139*, 107536.

# When AI becomes a teammate: The impact of collaborative autonomy on human-AI team dynamics

FAN Zili[1,2], CHENG Ken[3], ZHU Mengting[4]

([1] *School of Economics and Management, Hubei University of Technology, Wuhan* 430068, *China*)

([2] *Hubei Digital Industrial Economy Development Research Center, Wuhan,* 430068, *China*)

([3] *School of Management, Zhejiang University of Technology, Hangzhou* 310023, *China*)

([4] *Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai* 200030, *China*)

**Abstract:** With the development of information technology, more and more jobs require the collaboration of humans and AI to be accomplished. Prior research has mainly explored the impact of AI on human-AI collaboration from a “tool” perspective, largely to the neglect of the fact that AI is transforming from a “tool” to a “teammate” for humans given the increasing AI technological autonomy. Confronting the challenges brought by the transformation of human-AI relationship, how to ensure the effectiveness of human-AI collaboration has become an urgent issue. Different from prior studies solely analyzing human-AI collaboration from the perspective of AI technology systems, this research incorporates humans into the analysis framework and concentrates on the four dimensions of human-AI collaborative autonomy, namely adaptive learning, uncertainty, opacity and openness, to carry out three studies via the lens of role theory. Study 1 proposes that the uncertainty and opacity dimensions lead to role conflicts which damage human-AI fit and then hinder human-AI collaboration. Study 2 points out that the openness dimension enhances the mental resilience of AI through role taking, helping AI understand and grasp the technical and role requirements needed in the process of collaborating with humans, thereby making AI more compatible with humans. Study 3 suggests that the adaptive learning dimension enhances AI anthropomorphism through role learning, promoting humans to better understand and accept AI, and ultimately achieving the goal of facilitating human-AI collaboration. This research offers theoretical basis and practical guidance for organizations to successfully introduce AI, promote human-AI co-creation, and cultivate “super individuals”.

**Keywords:** human-AI collaborative autonomy, human-AI collaboration, role conflict, role taking, role learning