文章编号:1003 - 2053(2026)04 - 0681 - 09

认知—行为互动视角下人工智能工具合法性建设

张伟齐[1,3],李思涵[2],吴晓波[3]

(1. 温州商学院管理学院,浙江温州 325035; 2. 欧洲工商管理学院新加坡校区,新加坡 138676;
3. 浙江大学管理学院,浙江杭州 310058)

摘 要:研究基于员工认知—行为互动的视角,深入探讨了组织引入人工智能(Artificial Intelligence,AI)工具过程中员工感知的专业身份认同威胁、印象管理行为以及 AI 合法性建设之间的互动关系与机制。研究首先明确了技术属性、任务特征和组织环境作为引发员工身份威胁感知的重要情境因素,并提出了专业地位威胁、专业能力威胁和自主性威胁三种具体的身份威胁感知类型,详细阐明了不同威胁感知如何驱动员工实施特定的印象管理行为,以及该行为如何产生认知强化作用。同时,研究进一步揭示了员工实施的印象管理行为不仅影响 AI 工具在组织内部认知、实践和规范三个层面的合法性建设,而且 AI 工具合法性建设程度也会反过来影响员工的身份认同威胁感知与印象管理行为,从而形成互动反馈机制。研究在理论上整合并深化了身份认同理论、印象管理理论与组织合法性理论的理解,也为管理实践提供了明确的策略指导,帮助组织有效应对 AI 工具应用过程中复杂的员工认知、行为与合法性管理挑战。

关键词:人工智能;组织合法性;身份认同威胁;印象管理行为

中图分类号:C936 **文献标识码**:A

近年来,以 DeepSeek 和 ChatGPT 为代表的生成式人工智能(Generative AI)工具迅速在组织中得到广泛应用,引发了管理实践与学术界的关注[1]。这些 AI 工具凭借在创新创意生成、决策辅助及信息分析等领域的显著优势,逐渐被视为组织实现竞争优势的重要战略工具[2]。但员工通常仍会策略性地修改或修饰工具的输出,以确保最终成果符合组织对原创性和质量的期望[3],深层次反映了他们担忧自身的专业能力、职业地位以及决策自主性被削弱甚至取代[4]。这种专业身份认同的威胁感促使员工主动采取策略性的印象管理行为,以避免在组织内部产生“过度依赖技术”或“专业能力不足”的负面社会评价[5]。这种策略性行为不仅反映了组织内权力关系的微妙变化,更显著阻碍了 AI 技术在组织中的制度化与深入融合[6]。

在理论层面上,尽管已有研究广泛探讨了技术接受、任务 - 技术匹配、技术赋能以及组织制度化过程等多个理论视角,但对 AI 技术合法性建设如何在组织内部受到微观个体认知和行为互动实现的机制仍未深入分析[7]。具体而言,传统技术接受理论主要关注用户对技术的初步接受意愿,忽视了长期使用过程中个体的身份认同与职业形象焦虑[8]。任务—技术匹配理论虽强调技术应用的绩效导向,却未能解释即使技术匹配良好,员工为何仍然存在掩饰技术贡献的动机和行为[9]。而制度理论尽管指出了组织环境对技术使用合法化过程的重要性,却忽视了组织内部个体认知和行为如何具体地作用于技术合法性建设,更未涉及合法性建设不足如何进一步强化员工的身份威胁与印象管理行为,形成持续互动的负反馈循环。

为弥补上述研究缺口,本研究聚焦以下核心问题:(1)员工在何种具体情境下会感受到 AI 工具威胁到自身的专业身份认同?(2)员工不同类型的身份认同威胁感知与印象管理行为是如何互动的?(3)员工的印象管理行为如何影响组织内部 AI 工具合法性在认知、实践和规范层面的构建?(4)AI

收稿日期:2025 - 04 - 12;**修回日期**:2025 - 06 - 17

基金项目:浙江省哲学社会科学重点研究基地浙江大学区域经济开放与发展研究中心课题(2022JDKTYB20);国家社会科学基金重大项目(21&ZD131);教育部人文社会科学研究青年基金项目(22YJCZH257)

作者简介:张伟齐(1998 -),男,讲师,博士。李思涵(1997 -),男,博士后研究员,博士,通讯作者,E - mail:sihan. li@ insead. edu。吴晓波(1960 -),男,教授,博士。

工具的合法性建设如何反过来进一步影响员工的身份认同威胁感知和印象管理行为?

通过系统整合身份认同理论、印象管理理论与组织合法性理论,本研究提出了一个互动反馈的理论框架,深入阐述了员工专业身份认同威胁感知、印象管理行为以及 AI 工具组织合法性之间的内在动态关系。这一理论框架不仅明确揭示了技术使用中的微观个体心理与行为如何具体作用于组织的宏观制度过程,也为深入理解技术与个体、组织间复杂互动关系提供了清晰的理论指引。同时,本研究也为管理者提供切实可行的策略,以有效应对 AI 技术引入过程中个体认知与行为的复杂挑战,促进技术制度化应用的顺利推进。

1 理论背景与文献回顾

1.1 AI 工具合法性建设:技术制度化与认知过程

组织合法性是指某种实践或技术在组织内部被广泛感知为适当、恰当并符合社会规范的状态[10]。新技术在组织中的合法性并非自动产生,而是通过组织成员在持续的社会互动过程中,基于技术的实际效用、组织价值观和社会文化等多重因素逐渐建构与赋予的社会认知[11]。具体而言,组织合法性通常可从认知合法性、实践合法性和规范合法性三个维度进行区分,即认知合法性(cognitive legitimacy)、实践合法性(pragmatic legitimacy)和规范合法性(normative legitimacy)[10]。

认知合法性强调技术在组织成员心理上的接受程度,即组织成员是否普遍认为该技术是必要且有价值的;实践合法性则聚焦于技术在日常组织实践中被广泛实际应用的程度,反映技术是否真正融入组织的具体运作流程;规范合法性则涉及技术与组织文化和社会规范之间的契合度。技术在组织内部获得规范合法性的过程,实际上是组织成员逐渐接受并内化该技术的社会规范适用性的过程。这一过程往往受到组织文化、价值观念和社会道德规范的显著影响,体现为成员在道德与规范层面对技术使用的积极态度与行为支持[10]。

尽管已有研究探讨了新技术的制度化和合法化过程,但现有理论大多侧重于宏观制度环境或整体组织情境,鲜有研究深入探讨个体员工的微观认知、行为与技术合法性建设之间的互动关系。

1.2 身份认同理论与身份威胁机制

身份认同理论主张个体在组织中的行为与表现受到其对自身职业角色和身份的主观认知强烈影响[12]。专业身份认同(professional identity)特指个体基于专业知识、技能、经验与自主性所形成的自我定位和角色认知[13],其不仅直接影响员工对工作的投入程度、满意感与职业成就感,也深刻影响员工在组织中的创造性与整体绩效表现。

然而,员工专业身份认同在引入新技术尤其是 AI 技术时常常遭遇显著的挑战与威胁。具体而言,身份威胁(identity threat)指的是员工感知到自身的专业技能、决策权力或职业地位可能被技术所替代或削弱时所产生的心理状态与情绪反应。Elsbach 和 Kramer 提出,员工之所以感知身份威胁,主要是因为他们察觉到新技术可能取代或削弱自身专业技能的价值,从而引发焦虑与防御行为[14]。

理论和实践表明,员工对这些身份威胁的感知并非随机产生,而是受到特定情境因素的显著影响。本文从任务特征、组织环境和技术属性三个维度明确了这些情境因素的作用机制。任务特征方面,员工在从事高度专门化且存在明确绩效评价标准的任务时,更可能因 AI 技术的介入而感受到自身专业能力和地位的威胁[15]。组织环境方面,管理者若过于突出 AI 工具的优势或替代潜力,或组织缺乏有效的沟通反馈机制,都容易加剧员工对职业发展和自主性的焦虑与威胁感知[11]。技术属性方面,AI 工具本身的透明性、可解释性及自主决策能力显著影响员工对自身专业地位与工作自主性的认知。尤其是当 AI 工具决策透明度低且自主性强时,员工对自身的控制感与专业价值感知将明显下降,进一步加剧专业身份认同威胁感[16]。

1.3 印象管理理论与员工技术使用行为

印象管理理论强调个体通过有意识的策略性行为和语言表现,主动影响他人对自身的评价与认知,以达到保护或提升自我形象与社会地位的目的[15]。在组织技术使用情境中,员工印象管理行为体现为个体如何在引入和应用新技术过程中,策略性地调整和控制技术作用的可见性和贡献程度,以维护自身在组织内部的专业形象与职业地位[17]。

在 AI 工具广泛应用的情境下,AI 工具卓越的效率与自主决策能力显著提升了员工感知到自身专业能力和决策自主性被取代或削弱的风险,从而激发了他们更强烈的身份威胁感知[3],从而激发员工的印象管理行为。本文基于以往研究和实践总结出

员工在组织部署 AI 技术环境中通常会实施的三种印象管理策略:淡化 AI 贡献、成果修饰和信息掩盖。淡化 AI 贡献行为是指员工在任务报告或公开表达中有意减少或隐瞒 AI 工具所发挥的作用,以便突出个人的主导地位与专业贡献。成果修饰行为表现为员工对 AI 生成的成果进行个性化的调整和修改,以增强自身专业能力的可见性,强调个人创造力和专业技能的价值。信息掩盖行为则体现为员工故意不公开或模糊技术使用过程中的细节信息,以避免他人对其个人专业能力和工作自主性提出质疑。

上述文献回顾表明,尽管组织合法性理论、身份认同理论和印象管理理论分别被应用于技术和组织管理研究领域,但这些理论尚未得到有效整合。AI 工具的强大功能与社会影响使得其在组织合法性建设过程得到关注,但现有研究未能充分揭示员工认知—行为—合法性之间互动的微观机制。因此,本研究通过融合多个理论视角,构建一个更加完整和精细的理论框架。

2　理论框架与研究命题

2.1　理论模型构建

本研究基于员工微观认知—行为互动的视角提出一个理论模型,旨在系统揭示组织内部引入人工智能(AI)工具后,员工的专业身份认同威胁、印象管理行为与 AI 工具合法性建设之间的动态互动反馈机制。

具体而言,研究首先明确了技术属性、任务特征以及组织环境是引发员工身份认同威胁感知的核心情境因素。这些情境因素引发的专业身份认同威胁具体表现为专业地位威胁、专业能力威胁与自主性威胁三种类型。

身份威胁感知驱动员工采取相应的印象管理策略,与印象管理行为形成持续的互动反馈关系[17]。一方面,身份认同威胁会促使员工实施特定的印象管理行为(如淡化 AI 贡献、成果修饰、信息掩盖),以维护自身的专业形象和社会认可;另一方面,这些印象管理行为本身又会反过来强化员工的身份认同威胁感知,形成身份认同威胁与印象管理行为之间的自我强化互动循环。

同时,员工实施的印象管理行为还会进一步作用于 AI 工具合法性建设的具体维度,包括认知合法性、实践合法性和规范合法性[10]。这些行为通过模糊工具实际贡献的信息传播,削弱组织成员对工具真实价值与实践效果的认知,从而降低 AI 工具在组织中的合法性建设水平。此外,AI 工具合法性建设的状况也会动态地反作用于员工的身份认同威胁感知与印象管理行为。当工具合法性水平较低时,员工可能进一步感知到职业地位与能力的风险,从而更频繁地实施印象管理行为,形成更广泛的负面互动反馈机制。

综上所述,本研究提出以下理论模型(图 1)。

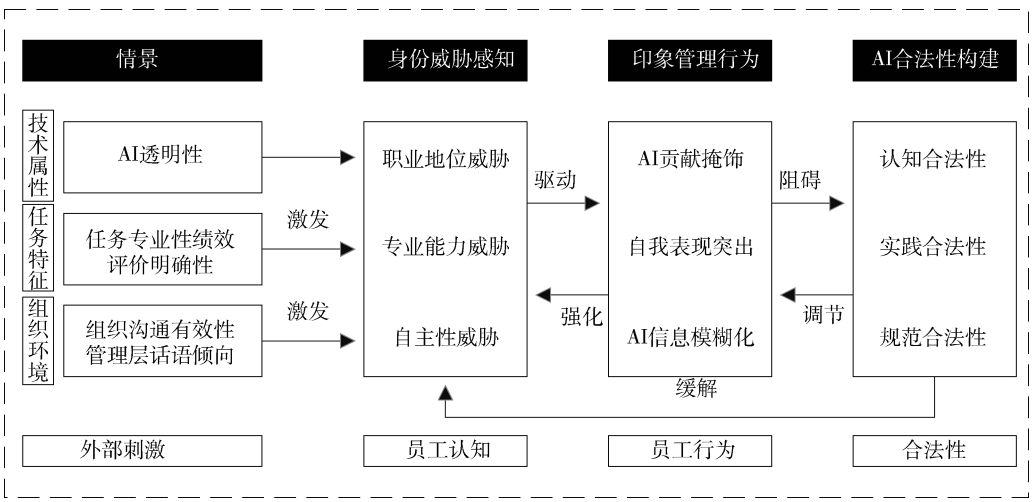

图 1　研究理论模型

Figure 1　Theoretical model

基于上述理论模型提出的动态互动反馈机制,本研究将在下一部分具体提出和阐述一系列明确的

研究命题。

2.2 研究命题提出

(1)身份认同威胁产生的不同情景

①技术属性维度

员工对自身专业身份的威胁感知往往源于具体情境因素的显著影响,其中技术属性是关键的情境维度之一。首先,从 AI 工具的决策透明性与解释性来看,当 AI 工具缺乏足够的决策透明性和解释性时,员工难以明确 AI 工具决策背后的逻辑、规则与具体界限,因而无法精准地界定自身与 AI 工具各自的贡献边界。这种模糊感直接导致员工感知到自身专业贡献的独特性和不可替代性被弱化,从而强化了员工对自身专业地位和专业能力的威胁感。同时,AI 决策的模糊性和不确定性也降低了员工对自身工作流程和决策自主性的掌控感,使员工感到其决策自主性显著被工具侵占[18]。因此,研究提出以下命题:

命题 1a:AI 工具的决策透明性与解释性越低,员工身份认同威胁感知越强烈。

其次,从 AI 工具自主决策能力的视角分析,当 AI 工具展现出较高的自主决策能力,即工具能够相对独立地完成核心任务和重要决策时,员工将直观地感知到自身专业角色的重要性、专业技能的独特性以及决策权威性面临直接且明确的替代风险。AI 工具自主决策能力越高,员工越可能感知到其在组织内部的专业地位被明显边缘化,专业技能和知识的不可替代性显著下降,甚至可能被完全替代。同时,高自主决策能力的 AI 工具也更可能直接替代员工原本所拥有的决策权力和任务控制权,员工因此可能明显感受到自主决策空间被大幅挤压或替代,从而全面强化员工在专业地位、专业能力和决策自主性上的威胁感知。因此,研究提出以下命题:

命题 1b:AI 工具自主决策能力越强,员工身份认同威胁感知越强烈。

②任务特征维度

任务特征也显著影响员工身份认同威胁感知。具体而言,当任务涉及高度专门化和稀缺性专业知识或技能时,这些能力和知识不仅构成了员工在组织中的核心竞争优势,也直接关系到员工的职业身份认同和社会地位。员工通常将高度专业化的知识和技能视为自身独特的职业优势与社会资本,这种优势和资本能够确保其在组织内的独特价值[19]。当 AI 工具在此类任务中表现出明显优势或高效时,员工将直观感知到自身原本具备的专业优势与竞争力被工具显著削弱甚至取代,从而直接引发员工对自身专业能力与社会地位的威胁感知[20]。这种对员工核心技能价值的威胁感知,会进一步放大员工对自身专业地位和职业前景的焦虑,从而产生强烈的身份认同威胁感[21]。因此,研究提出以下命题:

命题 1c:任务的专业化程度与稀缺性越高,员工身份认同威胁感知越强烈。

此外,当任务绩效评价标准明确且存在明显的绩效比较情境时,员工的专业表现与 AI 工具的表现被同时呈现和比较,员工更易感知到自身在专业任务中的相对劣势和可能被工具所取代的风险[5]。明确的绩效标准与直接的绩效对比不仅使员工的专业贡献更加透明化,也使员工专业能力的不足或被替代的可能性变得更加突出,直接影响了员工专业地位的稳定性。此外,绩效差距的清晰可见性也会降低员工对自身任务执行自主性和决策权威的认知,使员工感到其决策空间和任务掌控权被 AI 工具的高效表现明显侵占,进而强化其专业身份认同威胁的感知。因此,研究提出以下命题:

命题 1d:任务绩效评价标准越明确,员工身份认同威胁感知越强烈。

③组织环境维度

组织环境维度强调了组织的沟通方式与管理策略如何影响员工对 AI 工具所带来的身份威胁感知。当组织管理者在推广 AI 工具时过度强调 AI 工具的绩效优势或潜在替代性时,员工将直接感受到来自组织的明确或隐性信号,使其不仅担忧自身专业技能价值下降和地位受到侵蚀,也深刻地感知到自身原有决策权威和自主性可能受到技术显著挤压[22],激发员工对于职业发展的负面情绪与焦虑,进而强化了员工对专业身份认同威胁的感知与防御反应。因此,研究提出以下命题:

命题 1e:当管理者越强调 AI 工具的高效表现与替代潜力,员工身份认同威胁感知越强烈。

同时,当组织内部沟通反馈机制缺乏或不畅通时,员工无法有效地获得关于自身与 AI 工具角色定位的明确反馈信息,进而产生强烈的不确定感与焦虑情绪。这种信息不确定性不仅使员工难以确定自身的专业地位和专业技能是否仍然被组织认可和需要,同时也使员工无法明确理解自身在任务决策和工作自主性方面的具体边界,导致其感受到决策空间与自主控制权的显著弱化[23],进一步强化其对自

身专业身份认同的威胁感知。因此,研究提出以下命题:

命题 1f:组织内部沟通与反馈机制越缺乏或不畅通,员工身份认同威胁感知越强烈。

(2)身份认同威胁与印象管理行为

专业地位是员工在组织中获得认可、尊重和社会评价的重要基础。当 AI 工具以高度效率与显著优势完成原本依靠员工专业知识和技能才能完成的任务时,使得员工在组织内部的专业地位可能被动摇或削弱[12]。如果员工察觉到 AI 工具所产出的成果可能导致其在组织中的专业地位降低,即认为组织中的同事或管理者会认为他们不再具备足够的专业价值或创新贡献时,他们便可能主动采取某种策略以保护自身的社会评价和专业地位。为了防止这种社会评价的负面变化,员工往往采取淡化 AI 工具贡献的印象管理行为,如刻意弱化甚至隐瞒 AI 工具在任务中的作用,以强调个人主导的贡献程度,试图保护自身在组织内部的核心地位与社会认同感[24]。因此,研究提出以下命题:

命题 2a:员工感知到的专业地位威胁越高,其实施淡化 AI 贡献行为和成果修饰行为的倾向越强烈。

员工的专业能力体现了其核心的组织价值和竞争优势。当 AI 工具的强大能力与高效表现明确表现出替代或超过员工专业技能的趋势时,员工清晰意识到自身技能和知识的独特性面临显著威胁。这种情况下,员工倾向于通过成果修饰和信息掩盖两种印象管理行为,以强调自身专业能力的持续有效性与必要性。一方面,员工通过在 AI 工具输出的成果基础上添加更多个人意见与修改,使组织成员难以清晰区分 AI 工具与员工个人的贡献,突显自身专业价值;另一方面,员工通过刻意隐瞒工具使用细节与决策过程信息,避免组织成员直接认知到 AI 工具对专业能力的替代程度。这种行为并非仅仅是被动反应,而是一种员工主动管理专业能力认同与评价的策略性行为。因此,研究提出以下命题:

命题 2b:员工感知到的专业能力威胁越高,其实施成果修饰行为和信息掩盖行为的倾向越强烈。

自主性体现为员工在工作任务中拥有的决策权与控制空间。当 AI 工具自主完成决策与任务执行能力增强,员工直接感知到自身工作自主性和决策空间受到明显挤压甚至替代时,员工倾向于主动采取信息掩盖和淡化工具贡献的印象管理行为,以重新明确自身在组织决策与任务执行中的主导地位和自主空间[5]。具体来说,员工可能在任务汇报、成果分享或团队沟通中刻意降低 AI 工具决策贡献的透明度与可见性,以凸显个人在决策过程中的主导地位,从而在组织内部维持自身的自主决策者形象。因此,研究提出以下命题:

命题 2c:员工感知到的自主性威胁越高,其实施 AI 信息掩盖行为和淡化 AI 贡献行为的倾向越强烈。

尽管印象管理行为短期内可以保护员工在组织中的社会形象与专业评价,但频繁实施这些行为本身也会逐渐强化员工的内在威胁感知和焦虑情绪[25]。当员工不断重复地淡化 AI 工具贡献、修饰成果或隐瞒信息时,其行为本身实际上持续确认并强化了 AI 工具对自身专业地位、能力和自主性的威胁存在。随着这种行为频率和强度的增加,员工内心的身份认同不确定感和焦虑感会持续加剧,从而形成持续自我强化的负面互动循环。这种负反馈循环不仅对员工个体有负面影响,还会严重阻碍 AI 工具在组织内部的有效应用与长期制度化。因此,研究提出以下命题:

命题 3:员工实施印象管理行为的频率与强度越高,其感知到的专业身份威胁越强烈,从而形成持续自我强化的负面反馈机制。

(3)印象管理行为与 AI 合法性

认知合法性指的是组织成员明确理解和认可一项新技术的实际功能、优势和适用情境[10]。在组织中引入新技术(如 AI 工具)时,组织成员需要通过清晰、真实的体验来判断该技术的潜在价值与实际效能,从而建立初步的认知合法性。然而,当员工出于维护自身专业身份和社会评价的动机,频繁且强烈地淡化 AI 工具在完成任务中的贡献时,这一行为直接影响了组织成员对 AI 工具真实效用的准确认知,使得组织成员误以为 AI 工具在任务中的实际作用有限,甚至可能将高质量成果的实现完全归功于员工的个人能力或经验。这种人为导致的认知偏差直接阻碍了组织成员对 AI 工具潜在优势及具体应用价值的准确理解[26]。这种信息扭曲和缺乏清晰展示 AI 工具作用的行为会使组织成员形成对 AI 工具较低的认知合法性认知,即组织成员无法准确意识到 AI 工具的真实效益与优势所在。因此,研究提出以下命题:

命题 4a:员工淡化 AI 贡献的印象管理行为越

频繁,组织成员对 AI 工具认知合法性的感知程度越低。

实践合法性根植于组织成员通过直接的任务执行与体验而明确感知到的技术应用效益[10]。对于 AI 工具而言,实践合法性的核心表现为组织成员在具体任务中的应用感受以及实际产出效果的直观体验。然而,当员工频繁对 AI 工具生成的成果进行大量修改和个性化修饰,以凸显自身的个人风格和专业贡献时,这种行为显著降低了组织成员直接体验和观察 AI 工具实际产出效益的可能性,形成实践层面技术效能信号的模糊化和不确定性。长期来看,这种持续的"信号模糊化"直接阻碍 AI 工具实践合法性的明确构建,使得组织成员难以产生持续且明确的应用共识和制度化使用。因此,研究提出以下命题:

命题 4b:员工的成果修饰印象管理行为越频繁,组织成员对 AI 工具实践合法性的感知程度越低。

规范合法性强调组织成员认为技术的使用是否符合组织的社会规范、文化价值和伦理准则[27]。频繁的信息掩盖行为,即员工故意隐瞒 AI 工具使用过程与决策细节,会直接破坏组织内透明与开放的信息共享文化,使组织成员产生对技术使用本身的伦理性与正当性严重怀疑[28]。具体而言,这种行为不仅破坏了组织成员对技术应用规范性与透明性的认知,也导致组织内对技术道德与伦理边界的不信任不断扩散和深化[16]。长期来看,这种负面的规范信号直接削弱了 AI 工具使用的伦理与道德正当性认同,进而严重阻碍技术在组织内的长期制度化与规范认可。因此,研究提出以下命题:

命题 4c:员工的信息掩盖印象管理行为越频繁,组织成员对 AI 工具规范合法性的感知程度越低。

当 AI 工具在组织内部合法性不足(认知、实践、规范合法性水平较低)时,该情境下,组织成员对技术应用效能和规范的质疑会不断加剧。在这种情境下,员工若持续使用 AI 工具,便可能面临组织成员对其专业判断、能力和贡献的质疑或负面评价。这种对自身职业形象的潜在负面影响促使员工实施更加频繁的印象管理行为,以积极维护自身在组织中的专业形象与社会评价。然而,这种进一步强化的印象管理行为却反向再次削弱组织成员对 AI 工具的认知合法性、实践合法性和规范合法性,从而形成一种持续的、自我强化的负反馈循环机制。因此,研究提出以下命题:

命题 5:AI 工具合法性建设水平越低,员工实施印象管理行为的频率越高,进一步降低 AI 工具合法性的建设水平,形成持续的负面互动反馈机制。

(4)AI 合法性与员工身份认同威胁

根据身份认同理论,员工的专业身份认同在很大程度上取决于组织成员对他们专业地位和贡献的明确认可与尊重[12]。在 AI 工具合法性建设不足的情境下,组织成员可能频繁表达对工具应用的不确定性、怀疑甚至否定意见。这种组织内部广泛存在的负面态度和质疑声音可能使员工产生更强烈的专业身份威胁感[29]。他们担忧自己与 AI 工具的关系会被组织内部错误解读,可能被视为专业能力不足或过度依赖技术的表现,进而影响其在组织内部的专业地位与职业前景。此外,当 AI 工具合法性建设不足时,员工在使用工具过程中可能更加谨慎甚至抗拒,担忧自身的能力评价因此被负面影响[30]。这种负面的外部环境氛围明显强化了员工对专业地位不稳固、专业能力不足以及自主决策空间受限的焦虑和威胁感知。因此,研究提出以下命题:

命题 6:AI 工具合法性建设越不充分,员工感受到的专业身份认同威胁感越强烈。

理论模型进一步指出,因 AI 工具合法性建设不足而强化的员工身份认同威胁感知并非停留在单一的心理感受层面,而是显著影响员工后续的行为选择。员工为应对这种持续强化的威胁感,倾向于频繁且明显地实施印象管理行为(如淡化工具贡献、成果修饰和信息掩盖),以保护自身在组织中的专业形象与社会认可。然而,这些行为本身却进一步扭曲了组织成员对 AI 工具真实贡献和应用效果的准确认知,降低了 AI 工具在组织内部的合法性认可水平,形成了一种负面且持续自我强化的互动反馈循环机制。因此,研究提出以下命题:

命题 7:AI 工具合法性建设水平越低,员工感知到的身份认同威胁越强烈,进而促使员工更频繁实施印象管理行为,这些行为又进一步降低了 AI 工具的合法性水平,形成持续的负面互动反馈机制。

3 讨论

3.1 理论贡献

本研究的理论贡献主要体现于对身份认同理

论、印象管理理论与组织合法性理论的深入整合。

首先,传统身份认同理论虽然关注员工在面临威胁时的认知和行为反应,但缺乏对专业身份认同威胁的精细化分类与具体机制阐释。本研究首次明确地将专业身份威胁区分为专业地位威胁、专业能力威胁和自主性威胁三种具体维度,并揭示了不同类型身份威胁如何系统且差异化地驱动员工实施特定印象管理行为,从而显著提升了身份认同理论的精细化程度与理论解释力。

其次,研究深化了印象管理理论在技术应用情境下的适用边界与理论机制。以往印象管理理论多关注一般社会互动情境,未能深入解释员工在面对技术冲击(尤其是 AI 工具)时如何策略性地管理自身专业形象及其长期后果。本研究明确揭示了员工在 AI 工具情境下实施印象管理行为的内在动机及其反馈机制,指出这些行为不仅源于身份威胁,更反向加剧员工自身身份威胁感知,从而提出了一种动态、自我强化的理论互动机制,扩展了印象管理理论的应用。

最后,研究系统地提出了个体层面的印象管理行为如何具体影响组织内技术合法性的认知、实践与规范建构过程。传统合法性理论多强调制度化的宏观视角,忽视了微观个体行为在合法性建设中的显著作用。本研究则基于微观个体认知—行为互动视角,揭示了员工身份认同威胁感知和印象管理行为如何作用于组织成员对技术的合法性判断与制度化进程,拓展了组织合法性理论的微观分析视野。

3.2　管理建议

基于本研究理论模型提出的复杂互动机制,组织管理者需要采取积极、有效的干预策略,以应对和缓解这些挑战。

首先,为缓解员工因 AI 工具引入而产生的专业身份威胁,组织应在实际工作流程中明确具体的任务分配与评价标准。例如,管理者可以将任务清晰划分为“AI 工具辅助任务”和“员工主导任务”,确保员工清晰了解自身贡献与 AI 工具的差异,帮助员工在 AI 工具环境中确认自身持续的专业价值和独特性[7]。

其次,为减少员工实施印象管理行为的实际动机,组织应建立一种真实而非惩罚性的绩效沟通机制。例如,管理者可以建立公开透明的成果汇报与分享平台,鼓励员工明确展示和客观评价 AI 工具在任务中的实际作用及贡献。同时,组织可推行明确的激励机制,以降低员工进行信息掩盖和成果修饰行为的动机,进而营造更加健康的绩效文化。

最后,为有效推进 AI 工具的合法性建设,组织应设置具体且实操的技术应用反馈闭环机制。例如,可在项目实施过程中定期组织技术复盘会,由专门团队定期对反馈进行汇总并迅速做出技术优化与调整。同时,管理者应在实际工作场景中设立 AI 技术应用的明确伦理与规范界限,确保 AI 工具使用与组织伦理规范高度一致,推动技术长期应用的有效落地。

3.3　未来研究方向

基于本研究的内容,我们提出了以下值得进一步探讨的研究方向与问题。

首先,尽管研究已提出了员工专业身份认同威胁、印象管理行为与 AI 工具合法性建设之间的互动框架,但未来研究仍需深入探讨不同组织环境与任务特征如何影响该机制的适用性与表现形式。例如,未来研究可以比较不同产业(如技术密集型与服务导向型)以及任务复杂性对员工身份认同威胁具体表现与强度的差异。

其次,未来研究应着眼于动态纵向视角,深入分析员工印象管理行为与 AI 工具合法性建设之间互动关系随时间推移的演变趋势与机制。具体而言,可以通过长期追踪或多阶段实地案例研究,深入揭示员工的印象管理行为如何随技术实施阶段的变化而动态演变,并进一步探讨组织文化、领导风格和绩效评价制度对这一互动效应的具体调节作用。

最后,未来研究需明确分析技术本身特征对身份认同威胁与印象管理行为的差异化作用机制,聚焦 AI 工具的透明性、解释性和自主决策程度等技术属性如何不同程度地影响员工的身份认同威胁感知与印象管理行为,并系统地检验不同技术特征下,这些机制的表现形式与效果差异。

参考文献:

[1] Boussioux L, Lane J N, Zhang M, et al. The crowdless future? Generative AI and creative problem - solving [J]. Organization Science, 2024, 35(5): 1589 - 1607.

[2] Kemp A. Competitive advantages through Artificial Intelligence: Toward a theory of situated AI[J]. Academy of Management Review, Academy of Management, 2023: 1 - 1.

[3] Lebovitz S, Lifshitz - Assaf H, Levina N. To engage or

not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis[J]. Organization Science, 2022, 33(1): 126 – 148.

[4] Ashforth B E, Schinoff B S, Rogers K M. "I identify with her," "I identify with him": Unpacking the dynamics of personal identification in organizations[J]. Academy of Management Review, 2016, 41(1): 28 – 60.

[5] 涂艳,蒿坡,龙立荣. 工作替代还是工作转型? 技术型工作不安全感的内涵、影响后果及来源[J]. 心理科学进展, 2023, 31(8): 1359 – 1373. Tu Y, Gao P, Long L R. Job replacement or job transformation? Definition, consequences, and sources of technology – driven job insecurity[J]. Advances in Psychological Science, 2023, 31(8): 1359 – 1373.

[6] Mollick E. AI in organizations: Some tactics[EB/OL]. https://www.oneusefulthing.org/p/ai – in – organizations – some – tactics. 2025 – 01 – 26/2025 – 04 – 09.

[7] Craig K, Thatcher J B, Grover V. The IT identity threat: A conceptual definition and operational measure[J]. Journal of Management Information Systems, 2019, 36(1): 259 – 288.

[8] Venkatesh V, Davis F D. A theoretical extension of the technology acceptance model: Four longitudinal field studies[J]. Management Science, INFORMS, 2000, 46(2): 186 – 204.

[9] Goodhue D L, Thompson R L. Task – technology fit and individual performance[J]. MIS Quarterly, JSTOR, 1995, 19(2): 213.

[10] Suchman M C. Managing legitimacy: Strategic and institutional approaches[J]. The Academy of Management Review, 1995, 20(3): 571.

[11] Suddaby R, Bitektine A, Haack P. Legitimacy[J]. Academy of Management Annals, 2017, 11(1): 451 – 478.

[12] Ashforth B E, Mael F. Social identity theory and the organization[J]. The Academy of Management Review, 1989, 14(1): 20.

[13] Ibarra H. Provisional selves: Experimenting with image and identity in professional adaptation[J]. Administrative Science Quarterly, 1999, 44(4): 764 – 791.

[14] Elsbach K D, Kramer R M. Members' responses to organizational identity threats: Encountering and countering the business week rankings[J]. Administrative Science Quarterly, JSTOR, 1996, 41(3): 442.

[15] Petriglieri J L. Under threat: Responses to and the consequences of threats to individuals' identities[J]. Academy of Management Review, 2011, 36(4): 641 – 662.

[16] Glikson E, Woolley A W. Human trust in Artificial Intelligence: Review of empirical research[J]. Academy of Management Annals, 2020, 14(2): 627 – 660.

[17] Leonardi P M, Treem J W. Behavioral visibility: A new paradigm for organization studies in the age of digitization, digitalization, and datafication[J]. Organization Studies, 2020, 41(12): 1601 – 1625.

[18] Raveendhran R, Fast N J. Humans judge, algorithms nudge: The psychology of behavior tracking acceptance[J]. Organizational Behavior and Human Decision Processes, Elsevier, 2021, 164: 11 – 26.

[19] Ashforth B E, Schinoff B S. Identity under construction: How individuals come to define themselves in organizations[J]. Annual Review of Organizational Psychology and Organizational Behavior, 2016, 3(1): 111 – 137.

[20] Li J, Li X, Zhang C. Understanding physicians' noncompliance use of AI – aided diagnosis—A mixed – methods approach[J]. Decision Support Systems, Elsevier, 2025, 191: 114420.

[21] Lebovitz S, Levina N, Lifshitz – Assaf H. Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know – what[J]. MIS Quarterly, 2021, 45(3): 1501 – 1526.

[22] 陈文晶,康彩璐,杨玥,等. 人工智能潜在替代风险与员工职业能力发展:基于员工不安全感视角[J]. 中国人力资源开发, 2022, 39(1): 84 – 97. Chen W J, Kang C L, Yang Y. The potential substitution risk of Artificial Intelligence and employees' professional ability development: From the perspective of employees' insecurity[J]. Human Resources Development of China, 2022, 39(1): 84 – 97.

[23] Kaya F, Aydin F, Schepman A, et al. The roles of personality traits, AI anxiety, and demographic factors in attitudes toward Artificial Intelligence[J]. International Journal of Human – Computer Interaction, 2024, 40(2): 497 – 514.

[24] Xie C, Fu T, Yang C, et al. Not a good judge of talent: The influence of subjective socioeconomic status on AI aversion[J]. Marketing Letters, 2024, 35(3): 381 – 393.

[25] Bolino M C, Kacmar K M, Turnley W H, et al. A Multi – level review of impression management motives and behaviors[J]. Journal of Management, 2008, 34

(6): 1080 – 1109.

[26] Agarwal N, Moehring A, Rajpurkar P, et al. Combining human expertise with artificial intelligence: Experimental evidence from radiology[R]. National Bureau of Economic Research, 2023.

[27] Scott W R. Institutions and Organizations: Ideas, Interests and Identities[M]. Los Angeles, CA: Sage, 2013.

[28] Rudko I, Bashirpour B A, Fedele M, et al. New institutional theory and AI: Toward rethinking of artificial intelligence in organizations[J]. Journal of Management History, Emerald Publishing Limited, 2025, 31(2): 261 – 284.

[29] Strich F, Mayer A – S, Fiedler M. What do I do in a world of Artificial Intelligence? Investigating the impact of substitutive decision – making AI systems on employees' professional role identity[J]. Journal of the Association for Information Systems, 2021, 22(2): 304 – 324.

[30] 何勤,刘明泽,李鑫悦. 人工智能会引发员工的知识隐藏行为吗? ——基于相对剥夺感的理论视角[J]. 外国经济与管理, 2024, 46(10): 55 – 70. He Q, Liu M Z, Li X Y. Will AI trigger employee knowledge – hiding behavior? From the perspective of relative deprivation[J]. Foreign Economics & Management, 2024, 46(10): 55 – 70.

Legitimacy construction of AI tools: A cognitive – behavioral interaction perspective

ZHANG Wei – qi[1,3], LI Si – han[2], WU Xiao – bo[3]

(1. School of Management, Wenzhou Business College, Wenzhou 325035, China; 2. INSEAD, Singapore 138676; 3. School of Management, Zhejiang University, Hangzhou 310058, China)

Abstract: This study develops a theoretical framework to explain the dynamic interplay among employees' professional identity threats, impression management behaviors, and the legitimacy construction of artificial intelligence (AI) tools within organizational contexts. As AI tools such as DeepSeek and ChatGPT become widely integrated into organizational practices due to their remarkable capabilities in innovation generation, decision – making support, and information analysis, understanding employees' responses and adaptations to these technological advancements becomes essential. Despite the notable improvements in efficiency and output quality facilitated by AI tools, employees frequently engage in impression management behaviors aimed at strategically modifying or obscuring the contributions of these technologies. Such behaviors reflect deeper anxieties regarding threats to their professional identity, particularly concerning perceived risks of diminished professional status, competence, and decision – making autonomy.

Addressing existing theoretical gaps, this research integrates identity theory, impression management theory, and organizational legitimacy theory into a comprehensive, interactive theoretical framework. Prior studies often examined these theories independently, neglecting the nuanced interplay at the micro – level between individual cognition, behavior, and broader organizational legitimacy processes. This study systematically addresses this gap by posing four critical research questions: (1) Under what specific organizational and task – related circumstances do employees perceive threats to their professional identity upon the introduction of AI tools? (2) How do distinct perceptions of identity threats—status threats, competence threats, and autonomy threats—influence employees' specific impression management behaviors? (3) How do these impression management behaviors, in turn, influence the cognitive, pragmatic, and normative legitimacy construction of AI tools within organizations? (4) How does the legitimacy constructed around AI tools reciprocally influence employees' ongoing identity threat perceptions and subsequent behaviors, thereby creating a continuous feedback loop?

To address these questions, the study first identifies and elaborates on key contextual factors that evoke perceptions of identity threats among employees. These factors are categorized into three main dimensions: technological attributes, task characteristics, and organizational contexts. Technological attributes include AI tools' decision transparency and autonomous decision – making capabilities, directly influencing employees' perceptions of control and their professional roles. Task characteristics, such as task specialization, uniqueness, and clarity of performance evaluation criteria, also significantly impact employees' feelings of professional competence and security. Meanwhile, organizational contexts—particularly managerial communication strategies emphasizing AI's efficiency and potential substitutability, alongside the organization's information feedback mechanisms—play critical roles in shaping employees' identity threat perceptions.

(下转第 756 页)