心理科学进展 2026, Vol. 34, No. 12, 2139–2151 https://doi.org/10.3724/SP.J.1042.2026.2139
Advances in Psychological Science https://cstr.cn/32111.14.2026.2139

·研究构想(Conceptual Framework)·

AI 工作身份威胁的构念、诱因及“双刃剑”效应*

黄 杰[1] 李亚莉[2] 周 空[3] 朱 润[2] 张永军[1]

([1]河南大学商学院, 开封 475004) ([2]南京大学商学院, 南京 210008)
([3]南京理工大学经济管理学院, 南京 210094)

摘 要 AI 的飞速发展和广泛应用, 使员工愈发感受到工作身份威胁, 但 AI 工作身份威胁的内涵、诱因及影响效果尚未得到清晰探讨。本研究构想主要基于身份威胁过程理论, 围绕“新内涵–诱发机制–作用效果”构建 AI 工作身份威胁的理论框架。首先, 本研究基于 AI 威胁源的特点, 尝试理论化提出 AI 工作身份威胁可能存在的新维度及内涵; 其次, 从 AI 拟人性与 AI 可解释性两个核心特征出发, 明确 AI 工作身份威胁的诱发机制及边界条件, 即员工可雇佣性和感知领导 AI 信任的中介作用, 及 AI 应用模式的调节作用; 最后, 揭示 AI 工作身份威胁对员工创新行为的“双刃剑”效应机制, 即工作退却或工作重塑的中介作用, 及特质调节焦点的调节作用。本研究将拓展和深化身份威胁过程理论, 为理解 AI 工作身份威胁的内涵、诱发和后果机制提供理论框架, 并为组织与员工制定更具针对性的 AI 工作身份威胁干预策略提供实践启示。

关键词 AI 工作身份威胁, AI 特征, 工作退却, 工作重塑, 员工创新行为

分类号 B849: C93

1 问题提出

随着人工智能技术的迅速发展, AI 在各种工作场景中的应用变得越来越普及(张志学 等, 2024), 极大地影响了员工的生产和工作方式(赵曙明 等, 2019; Qin et al., 2022), 生成式 AI 的横空出世加剧了这一趋势(彭剑锋, 2023)。国际联合调查结果显示, 世界主要国家的 29%居民认为他们的工作将在 5 年内被 AI 取代(谢小云 等, 2021)。在中国, 19.05%的员工面临被 AI 取代的风险(王林辉 等, 2022)。AI 可能重塑一些行业, 造成社会功能和社会存在的替代(唐小飞 等, 2021), 替代员工的自我价值和功能(王欣 等, 2021), 取代员工工作(Yam et al., 2023), 导致他们产生威胁感(Hunt et al., 2022; Puntoni et al., 2021)。在众多工作领域和岗位中, 员工的工作身份会受到威胁(Gerlich et al., 2007), 产生 AI 工作身份威胁, 需要组织实施有效干预。然而, 现有研究把 AI 工作身份威胁与其他身份威胁混淆在一起, 如污名化相关的身份威胁等(Lyons et al., 2020)。AI 工作身份威胁可能存在独有内涵和维度, 其诱因和对员工的影响效果尚未得到深入探讨。员工 AI 工作身份威胁研究是一个重要的前沿性议题, 理论和实践需求迫切, 本研究将基于身份威胁过程理论展开深入探讨。

尽管学术界关于身份威胁的已有研究为理解 AI 工作身份威胁提供了重要的理论基础和启示, 但是仍存在明显的局限性。第一, 身份威胁的内涵需要拓展, 现有构念难以充分捕捉 AI 情境下工作身份威胁的完整内涵及新维度。原因如下: (1)整体身份威胁、不同身份范畴下的身份威胁, 以及由不同威胁源引发的身份威胁, 在内涵结构上存在显著差异。其中, 整体身份威胁是个体对身份价值(value)、意义(meanings)和呈现(enactment)有

收稿日期: 2025-12-31
* 国家自然科学基金青年项目(72402059); 中国博士后科学基金特别资助(2026T190183); 国家资助博士后研究人员计划 C 档资助(GZC20240389); 中国博士后科学基金面上资助(2024M750786); 国家社会科学基金重点项目(24AGL036); 河南省高等学校哲学社会科学创新团队建设计划资助(2027-CXTD-05)。
通信作者: 李亚莉, E-mail: yali_li@smail.nju.edu.cn

潜在危害的感知(Petriglieri, 2011); 从身份范畴的独特性出发, 学者提出了社会身份威胁构念(Ma et al., 2024); 从威胁源的独特性出发, 有研究开发了污名化相关的身份威胁构念(Lyons et al., 2020)。鲜有研究同时考虑特定身份范畴和特定威胁源的双重独特性, 这限制了现有理论对AI情境下工作身份威胁的解释力; (2)现有身份威胁研究在威胁内容的覆盖上存在结构性不足。受象征性(Symbolic)−工具性(Instrumental)框架启发(Keller, 1993), 工作身份的象征性威胁涉及员工的自我概念威胁, 如自我意义、价值的认同及自我呈现。也就是说, 上述对整体身份威胁、不同范畴身份威胁和不同威胁源身份威胁的研究主要关注象征性方面的威胁, 比如自我意义与价值认同受到挑战、自我呈现空间被束缚等, 而忽略了身份的工具性特征。从工具性角度而言, AI 工作身份威胁具有明显区别于现有身份威胁的工具性特征。如区别于以往技术, AI 技术与人类智能的高度相近性(情感、动机、反思等)会冲击人类的独特性(Gabbiadini et al., 2025; Liu & Sundar, 2018; Stein et al., 2019), 可能产生对人类身份的独特性威胁; 又如, AI 作为高能力、高效率工具, 可能会替代甚至超越个体的知识、技能与能力, 导致身份存续性威胁。因此, 本研究基于身份威胁过程理论, 将象征性−工具性框架作为构念探索的启发性分析框架, 同时考虑工作身份范畴和AI威胁源的双重独特性, 尝试从理论上构建AI工作身份威胁的新内涵和维度。

第二, 尽管AI工作身份威胁的现象逐渐普遍, 但是学术界对于 AI 工作身份威胁诱因机制的研究不足。虽然在传统技术情境下, 有研究探究了技术特点如技术变革的程度和频率对身份威胁的影响(George et al., 2023), 但传统技术特点与 AI 的核心特点具有本质区别(罗映宇 等, 2023)。相较于传统技术, AI 具有两大独特特征：外在形态的拟人性和内在推理的可解释性(Huang et al., 2021; Jin, 2023)。其中, 拟人性指的是AI呈现出高度类似于人类的特征和行为(Kim & Lee, 2026; Li et al., 2026), 可解释性指的是AI给定结果背后的推理被理解的程度(Stohl et al., 2016)。身份威胁过程理论指出, 行业未来特征如技术变化和情境模糊度会使个体产生身份威胁(Petriglieri, 2011)。因此, 本研究聚焦AI拟人性(外在特征)和AI可解释性(内在特征), 明确其对 AI 工作身份威胁的影响机制。具体而言, AI 拟人性并非简单的设计差异, 而是一种“行动者线索”, AI拟人性越强, 越可能导致员工认为 AI 的能力边界更广、可替代性更强, 进而降低对自我可雇佣性的评估(Huang et al., 2021)。而 AI 可解释性高意味着 AI 能够为其输出的信息或工作结果提供清晰、可追溯的推理依据(van den Broek et al., 2021; Rudin, 2019), 使 AI 的决策逻辑更容易被理解和验证, 其输出也更容易被领导视为有理有据、值得信赖, 员工感知到的领导对 AI 的信任度随之提高。另外, 身份威胁过程理论指出, 情境特征如何被建构性解读决定了威胁程度(Petriglieri, 2011)。AI 应用模式划定了 AI 与员工之间的任务边界, 提供了员工解读自我角色边界的框架。所以, AI 应用模式决定了 AI 拟人性和 AI 可解释性这两类 AI 特征信号被员工感知为威胁的程度。因此, 本研究提出 AI 拟人性和 AI 可解释性通过影响员工可雇佣性和感知领导AI信任, 进而引发AI工作身份威胁, 在此过程中, AI 应用模式起边界条件作用。

第三, AI 工作身份威胁对员工认知及行为的影响机制尚不清晰。过往身份威胁的研究主要基于定性分析, 指出身份威胁既存在消极影响(如增加员工压力、降低工作投入等), 也存在积极影响(田喜洲 等, 2016; Caza et al., 2018; Rothausen et al., 2017; Wang et al., 2024)。然而, AI 工作身份威胁如何影响员工认知及行为, 尚缺乏探究。本研究将关注员工创新行为这一组织关注的重要行为, 原因在于：(1)过往大量研究表明员工创新行为对组织绩效具有重要影响, 在当下的 VUCA 时代, 企业越来越需要员工创新行为来构建组织竞争优势(Anderson et al., 2014; Liehr & Hauff, 2025); (2)创新行为是一种典型的角色外行为, 指员工在不确定与风险情境中主动提出并实践新构想、新方案以提升组织效能(Ng & Lucianetti, 2016)。相较于工作投入和工作绩效等, 创新行为更直接体现员工如何在 AI 情境下重新确认并展示自身工作身份的独特价值。因此, 本研究致力于探究 AI 工作身份威胁如何及何时影响员工创新行为。基于身份威胁过程理论(Petriglieri, 2011), 本研究提出工作退却(身份保护策略)与工作重塑(身份扩展策略)的“双刃剑”机制。本研究进一步关注不同特质个体是否可能通过不同的应对策略影响员工创

新行为。过往研究表明特质调节焦点在决定个体是否主动应对身份威胁方面起重要作用(Shepherd & Williams, 2018)。因此, 本研究认为, 特质防御型焦点员工可能通过工作退却实现身份保护, 消极影响员工创新行为; 特质促进型焦点员工可能通过工作重塑实现身份扩展, 积极影响员工创新行为。换言之, AI 工作身份威胁会对员工创新行为产生“双刃剑”效应。

综上所述, 本研究尝试探索 AI 工作身份威胁的新内涵, 构建其前因、后果的整合理论框架, 对于身份威胁领域具有以下理论贡献：(1)尝试探索 AI 工作身份威胁的内涵, 理论化构建其可能存在的身份独特性威胁与身份存续性威胁新维度, 扩展身份威胁构念; (2)提出 AI 工作身份威胁如何受到 AI 拟人性与 AI 可解释性这两个 AI 核心特征的影响, 扩展身份威胁的诱因研究; (3)构建 AI 工作身份威胁对员工创新行为的“双刃剑”影响机制与边界条件, 为深入理解 AI 工作身份威胁的影响效果提供较系统、全面的思路。此外, 本研究也将有助于管理者和员工更加精准地理解 AI 工作身份威胁的内涵; 有助于管理者识别诱发员工 AI 工作身份威胁的关键因素和机制, 从而实施针对性的干预管理; 同时也有助于管理者和员工调整对 AI 工作身份威胁的消极认知, 并采取相应的应对措施。

2　国内外研究现状

2.1　身份威胁概念及测量

工作身份是指员工基于独特的个人特征、集体成员身份和社会角色而形成的意义集合(Miscenko & Day, 2016), 如工作个人身份、工作组织身份和工作职业身份。Petriglieri (2011)的身份威胁过程理论认为, 身份威胁是对身份价值、意义和呈现有潜在危害的个体感知, 包括身份价值威胁、身份意义威胁和身份呈现威胁三个维度。该构念旨在涵盖所有身份范畴与威胁源, 因而可能难以捕捉特定身份范畴(如工作身份)在特定威胁源(如 AI)情境下所呈现的独特威胁表征。为弥补整体身份威胁构念的局限, 既有研究尝试从身份范畴独特性或威胁源独特性两个方向对身份威胁进行细化。例如, 基于身份范畴差异发展出社会身份威胁构念(Ma et al., 2024); 基于威胁源差异发展出污名化相关身份威胁构念, 如异性恋身份威胁(Lyons et al., 2020)。然而, 这两方面研究往往各自分开进行, 前者强调身份属于何种范畴, 但较少考虑威胁源何以独特; 后者强调威胁源是什么, 却较少区分威胁发生于何种身份范畴。换言之, 现有研究鲜少同时将身份范畴的独特性与威胁源的独特性纳入同一理论框架, 从而限制了我们对 AI 情境下工作身份威胁这一“范畴×威胁源”双重独特现象的解释力与构念精度。而且, 根据象征性−工具性框架, 该构念的三个维度主要覆盖对自我概念的象征性威胁, 却未能充分反映工具性威胁。当个体在组织任务系统中的能力独特性与功能地位被重新评估甚至被技术替代时, 身份威胁不仅是意义与价值受损, 也可能是工具性身份独特性和“身份在未来工作结构中是否仍有位置”的存续性受损。

从测量层面来看, 首先, 整体身份威胁量表虽可用于捕捉一般性威胁体验, 但由于其面向高度抽象的跨范畴、跨威胁源现象, 难以确保 AI 情境下工作身份威胁的内容覆盖与情境贴合度, 进而可能限制测量的精准性与外部效度(George et al., 2023)。其次, 在更精细化的研究中, 有学者通过改编既有单维量表来测量职业身份威胁(Mitchell et al., 2011), 缺乏系统的量表开发与验证流程, 因而在结构效度等方面存在潜在风险。再者, 针对污名化等特定威胁源所开发的量表虽然更具情境指向性(Lyons et al., 2020), 但多为一维结构, 难以区分不同威胁维度在形成机制与行为后果上的关键差异。而且, 这类量表往往未显式区分身份范畴, 难以直接迁移解释 AI 情境下的工作身份威胁。

目前, 关于 AI 威胁源所引发的工作身份威胁, 学界在概念界定、维度结构及针对性测量工具方面仍缺乏一致且清晰的共识。现有研究大致采用了两类测量路径：其一, 直接借用或改编既有身份威胁量表。例如, 有研究将种族身份威胁的一维量表加以修改, 用于测量机器人/AI 身份威胁(Huang et al., 2021; Jin, 2023)。然而, 这类做法通常缺乏系统的量表开发与验证流程, 因此在内容覆盖、结构效度与外部效度上存在明显局限, 难以确保其能够捕捉 AI 情境下身份威胁的独特内涵与潜在新维度。其二, 从 IT 情境出发发展技术相关身份威胁量表, 并迁移至 AI 研究。Craig 等(2019)基于自尊理论和文献开发了 IT 身份威胁量

表, 分为 IT 社会身份威胁、IT 角色身份威胁及 IT 个人身份威胁三个维度。然而, 从构念基础看, George 等(2023)明确指出, 自尊和身份威胁具有明显区别; 从情境适配看, AI 与传统 IT 技术(如办公软件等)在能力边界、替代性、类行动者属性等方面存在本质差异(罗映宇 等, 2023)。综上所述, 未来研究可能需要综合考虑身份的象征性威胁和工具性威胁, 以及身份威胁的工作范畴和 AI 威胁源, 更加清晰和精准地构建 AI 工作身份威胁的独特内涵、维度和量表。

2.2 身份威胁的诱发机制研究

以往研究主要涉及个体、组织、技术等方面的经历、线索或变化对身份威胁的影响(Petriglieri, 2011)。关于身份威胁最常见、研究最多的个体诱发因素是关于自己的明确负面信息和评价。例如, 失业会导致身份威胁(Shepherd & Williams, 2018); 组织诱发因素方面, 组织变革(Kyratsis et al., 2017)、组织危机和丑闻(Piening et al., 2020)等也会对身份产生威胁。Piening 等(2020)指出, 来自媒体的组织负面评价会威胁员工的组织身份, 他们会参与归因过程来理解这种威胁; 技术诱发因素方面, 技术变革的程度和频率对教师身份威胁具有促进作用(George et al., 2023)。AI 相对于传统技术, 带来很多新优势和变化。如研究表明, AI 机器人在狭窄任务领域内的明显优势, 可能会导致人们将 AI 机器人视为对他们就业的威胁(Yam et al., 2023)。Mirbabaie 等(2021)指出, AI 导致的工作变化和地位丧失可以预测员工的社会身份威胁。研究还表明, AI 会使员工产生职业身份威胁(Shonhe & Min, 2025)。然而, 以往研究对 AI 特征(AI 拟人性和 AI 可解释性)如何引发身份威胁的机制尚不清楚。

对于身份威胁诱发机制的边界条件, 以往研究主要探讨个体地位、价值观和个体评价对身份威胁诱发机制的调节作用。个体地位方面, 有研究发现, 地位的丧失对自我完整性构成严重威胁, 地位高时, 地位丧失对身份威胁的消极影响更大(Marr & Thau, 2014); 个体价值观方面, 例如面对歧视, 一些人可能会感到身份威胁, 而另一些人可能不会。当个体具有自给自足的自我价值时, 他们感知到的身份威胁减弱, 他们会更加开放, 削弱消极反馈导致的身份威胁(Kang & Kim, 2022); 来自其他个体的评价对身份威胁感知也起调节作用, 例如来自竞争对手的挑战可能加剧向上社会比较产生的身份威胁(Kilduff et al., 2010)。Kang 和 Kim (2022)指出, 来自其他专家基于实际结果的不确定反馈会加剧消极绩效反馈对专家身份构成的威胁。本研究则聚焦 AI 应用模式在 AI 工作身份威胁诱发机制中的调节作用。

2.3 身份威胁对个体结果的影响研究

以往研究主要探讨身份威胁对个体认知、情感和行为的消极影响。关于个体认知结果, 身份威胁会引发个体的自我调节过程, 进而损耗其自我调节资源和认知资源(Schmader et al., 2008); 关于个体情感、情绪结果, 当个体感知到身份威胁时, 会产生心理契约违背, 降低幸福感(Vrontis et al., 2022), 及其他消极情绪(Eury et al., 2018), 如情绪疲惫(George et al., 2023)和压力(Rothausen et al., 2017), 导致不真实、痛苦、恐惧、愤怒、内疚和羞耻感等(Greenbaum et al., 2022; Vescio et al., 2021); 关于个体行为结果, 身份威胁会导致工作投入、工作效率、工作满意度和工作绩效的下降(王艳霞, 周珊, 2024; Greenbaum et al., 2022; Shewach et al., 2019; Wang et al., 2024), 引发个体离职(George et al., 2023)、退休(Vough et al., 2015)和不道德行为(Zhou & Dou, 2023)。其他一些研究表明身份威胁还会增加群体冲突, 使组织变革受阻(Fiol et al., 2009)。对于身份威胁的积极一面, 以往研究更多地进行理论和定性分析。如有学者探讨个人如何以认知或推论方式积极管理他们受到威胁的职业身份(Wang et al., 2024), 提出个体可以通过积极的身份重构(Petriglieri, 2011)和身份工作(Caza et al., 2018)来应对身份威胁, 如个体在面临消极绩效反馈导致的身份威胁时, 会识别问题、进行改进并采取纠正措施。Tajfel 和 Turner (2004)认为, 个体在受到威胁的情况下发现自己处于低地位群体时, 会寻找方法走出该群体, 进入更高地位的群体, 以应对威胁。Kreiner 等(2009)提出的身份威胁柔术, 也定性描述了一些旨在将负面威胁转化为积极行动的行为：这些行为既保留受威胁的身份, 又通过将威胁重新定义为机会(而非贬低构成威胁的因素), 寻求改善自身与情境之间的关系。另外, 只有极少数研究涉及身份威胁结果效应的个体价值观方面边界条件, 如 Kang 和 Kim (2022)指出, 自我肯定可以增强自我价值, 帮助专家认识到特定的身份威胁不

会破坏他们的整体自我价值, 削弱威胁的显著性。另外, 自尊也能削弱职业身份威胁导致的消极情绪(Yang & Zhang, 2023)。因此, 探讨 AI 工作身份威胁对员工的“双刃剑”效应及其边界条件具有必要性。

3　研究构想

本研究首先基于 AI 的特点与身份威胁过程理论, 构想 AI 工作身份威胁区别于已有身份威胁的可能内涵和维度, 并在此基础上, 明晰员工 AI 工作身份威胁的诱发机制, 以及其对员工创新行为的“双刃剑”效应框架。

3.1　研究 1：AI 工作身份威胁的内涵

身份威胁过程理论提出整体身份威胁概念, 指个体感知其工作身份的价值、意义与呈现可能受到损害(Petriglieri, 2011), 即包含身份价值威胁、身份意义威胁与身份呈现威胁三个维度。George 等(2023)据此开发了相应量表, 但该构念与测量较为宽泛, 难以捕捉不同身份范畴与不同威胁源下身份威胁的独特内涵。为深化研究, 部分学者从身份范畴出发提出社会身份威胁(Ma et al., 2024), 但未区分威胁源; 另一些学者从威胁源出发提出污名化相关身份威胁(如异性恋身份威胁; Lyons et al., 2020), 但未区分身份范畴。总体而言, 鲜有研究同时在“特定身份范畴×特定威胁源”框架下刻画身份威胁。

目前对 AI 工作身份威胁最相近的测量是 Craig 等(2019)所开发的 IT 身份威胁量表, 该研究基于自尊理论和文献, 把 IT 身份威胁构念分为 IT 社会身份威胁、IT 角色身份威胁及 IT 个人身份威胁三个维度, 该量表被 Mirbabaie 等(2021)用于测量 AI 身份威胁。然而, 工作身份只是个体身份的一种。而且, AI 与传统 IT 技术(如办公软件)具有明显区别(罗映宇 等, 2023)。更重要的是, 自尊与身份威胁也具有明显区别：自尊是一种相对稳定的特质, 而身份威胁则是对身份潜在危害所做出的短暂性评估。此外, 身份威胁是对特定身份的感知评估, 而自尊是对自我的整体感知评估。最后, 身份威胁不一定伴随着自尊的削弱(George et al., 2023)。

另外, 过往身份威胁的内涵研究主要关注身份的象征性属性受到威胁, 忽视了其工具性属性。象征性−工具性框架指出, 产品/个体形象由象征性意义(关涉自我概念, 如意义、价值认同与自我呈现)与工具性属性(关涉客观功能与有形特征)构成(Keller, 1993)。工作身份是个体自我概念的重要组成与意义来源(Petrou et al., 2017), AI 既可能威胁员工的自我意义与价值, 也可能威胁其功能、能力(王欣 等, 2021; Stein & Ohler, 2017)。然而, 既有关于整体、不同范畴与不同威胁源的身份威胁研究多聚焦象征性威胁(意义、价值与呈现), 对工具性威胁关注不足。在 AI 情境下, AI 工作身份威胁可能存在新的内涵：如 AI 与人类智能的高度相近性(如情感、动机与反思等类人特征)可能会冲击人的独特性(Gabbiadini et al., 2025), 从而可能引发员工在工作身份功能与能力独特性上的威胁感(Liu & Sundar, 2018; Stein et al., 2019)。身份独特性威胁强调员工的工作能力和角色贡献是否仍具有差异性与不可替代性。再如, AI 作为高能力、高效率工具, 可能会替代人力资本和人类工作身份(唐小飞 等, 2021), 产生身份存续性威胁。身份存续性威胁强调员工的工作身份是否能够在未来组织任务结构中持续存在。根据研究 1 的初步构想, 独特性威胁与存续性威胁属于并列维度, 虽可能相关, 但并非同义概念。

总之, 本研究基于身份威胁过程理论, 融合工作身份范畴和 AI 威胁源的双重独特性, 将象征性−工具性框架作为启发性分析框架, 认为 AI 工作身份威胁构念可能不仅包含 AI 对工作身份象征性的威胁, 如价值、意义和呈现的威胁感知, 也可能包含 AI 对工作身份工具性的威胁, 如独特性和存续性(存在和继续)的威胁感知, 还可能形成不同于当前设想的五维度结构。

3.2　研究 2：员工 AI 工作身份威胁的诱发机制

本研究聚焦 AI 特征, 探讨 AI 的两种独特特征(即 AI 拟人性和 AI 可解释性)通过员工可雇佣性和感知领导 AI 信任影响 AI 工作身份威胁的诱发机制, 并明晰 AI 应用模式所起的调节效应。理论模型如图 1 所示。

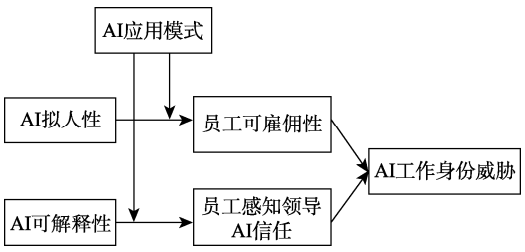


图 1　AI 工作身份威胁诱发机制模型

3.2.1 AI 拟人性、员工可雇佣性与 AI 工作身份威胁

AI 拟人性体现于那些栩栩如生的虚拟 AI 人物和 AI 机器人(Mishra et al., 2022)。其中, 虚拟 AI 人物指的是由计算机生成的、完全在线存在的图像或动画数字角色, 是以数字化身呈现的代理人, 设计得很像人类员工(Arsenyan & Mirowska, 2021)。而 AI 机器人, 如 AI 服务机器人, 是能够交互、通信和向客户提供服务的自主机器人(Wirtz et al., 2018)。AI 拟人性是指将人类的特征和行为(如智能化、自主性和学习能力)赋予非人类 AI 的现象(MacInnis & Folkes, 2017)。AI 拟人性包括外表拟人性和行为拟人性, 其中外表拟人性指 AI 机器人在形象上与人的相似性。人们对 AI 机器人的体验主要是由机器人的外观驱动的。相关研究表明, 机器人外表是否拟人化会影响人们对其心智与人类相似程度的推断。AI 外表拟人性程度高会使员工感到 AI 与人没有差异, AI 可作为实体替代他们, 从而使他们感到自身存在受到威胁(Ho & MacDorman, 2010), 对人类独特性的自我概念构成威胁(Stein et al., 2019)。行为拟人性是指 AI 机器人在行为上与人的相似性(Glikson & Woolley, 2020)。研究发现, 行为拟人性高的 AI 会剥夺员工的工作机会和工作资源, 使人类日益过时, 进而取代人类, 造成失业(Yogeeswaran et al., 2016)。由此可见, AI 拟人性与员工的 AI 工作身份威胁之间存在正相关关系, 即 AI 拟人性越高, 员工的 AI 工作身份威胁越强; AI 拟人性越低, 员工的 AI 工作身份威胁越弱。

可雇佣性是指员工对保持现有工作和获得理想工作的能力感知, 包括内部可雇佣性和外部可雇佣性两个维度(Rothwell & Arnold, 2007)。其中, 内部可雇佣性是指员工感知的在当前工作单位的可雇佣性, 表明了员工保持现有工作的能力感知; 外部可雇佣性是指员工感知的在当前工作单位之外的劳动力市场上的可雇佣性, 表明了员工获得理想工作的能力感知。Hillage 和 Pollard (1998)也认为, 可雇佣性是员工在保持就业的状态下, 能够在劳动力市场充分流动, 以实现自我潜能的能力感知。基于 AI 拟人性会加剧 AI 对人类员工的取代并导致失业这一趋势(Yogeeswaran et al., 2016), 以及 AI 拟人性越高表明 AI 功能越强大、甚至可能超越人类员工能力这一观点(Huang et al., 2021), 本研究认为, 相较于 AI 拟人性低的情况, AI 拟人性高意味着 AI 与人类员工的相似性更高, 对员工工作机会和工作资源的取代与剥夺程度也更高, 因而更容易导致员工对自身可雇佣性的感知降低, 进而引发更高水平的 AI 工作身份威胁感。身份威胁过程理论对身份威胁情境类诱发因素的论述表明, 行业发展现状、未来特征(如技术变化)等方面的线索会影响个体的自我认知和自我意识, 使个体感知到未来身份的贬值, 及身份在未来的不可持续, 阻止未来身份的设定(Petriglieri, 2011)。本研究基于该理论, 认为 AI 技术的高拟人性会使员工对自己可雇佣性的认知和评估降低, 进而增加 AI 工作身份威胁感。因此, 本研究提出如下命题:

命题 1: AI 拟人性通过负向影响员工可雇佣性感知, 进而增加 AI 工作身份威胁。

3.2.2 AI 可解释性、员工感知领导 AI 信任与 AI 工作身份威胁

AI 可解释性是指人们能够理解 AI 输出结果背后推理过程与依据的程度(Stohl et al., 2016)。由于许多 AI (尤其是深度学习模型)具有高度复杂性, 其决策逻辑往往呈现黑箱化特征, 可解释性不足由此成为制约其应用的瓶颈。可解释性的重要作用在于为 AI 输出提供可被审视与验证的逻辑, 进而提升使用者对信息的认知确定性与信任感(Lebovitz et al., 2022)。例如, 在医疗场景中的研究发现, 当 AI 能够解释其诊断依据时, 使用者更容易将其输出理解为有理由、可追溯、可辩护的判断, 从而提升对结果的确定性感与可信度(van den Broek et al., 2021)。类似地, 在组织管理情境下, 当 AI 可解释性高时, 员工可能会预期领导认为自己的专业能力可被 AI 很好地替代, 即员工感知到的领导 AI 信任随之增强。员工会形成对自身能力与地位被重新评估、专业边界被压缩、乃至岗位可持续性下降的预期, 进而感受到更高的 AI 工作身份威胁, 尤其表现为独特性威胁和存续性威胁等工具性层面威胁的增强。AI 可解释性低意味着 AI 难以为其输出的信息或工作结果提供清晰的推理依据, 甚至完全无法呈现其推理过程(即呈现“黑箱”特征) (Rudin, 2019), 员工可能会预计到领导因难以理解或追溯 AI 的决策逻辑, 而对 AI 生成的结果不放心(Venkatesh, 2021), 即感知领导 AI 信任减弱, 从而使得员工不容易被 AI 所

取代, 员工 AI 工作身份威胁较低。身份威胁过程理论也指出, 情境的模糊度高会使个体产生身份威胁(Petriglieri, 2011)。AI 可解释性程度高意味着 AI 的透明度高, 模糊度低, 员工感知领导 AI 信任增强, AI 工作身份威胁增强。也就是说, AI 可解释性可以通过员工感知领导 AI 信任正向影响 AI 工作身份威胁。因此, 本研究提出如下命题:

命题 2: AI 可解释性通过正向影响感知的领导 AI 信任, 进而增加 AI 工作身份威胁。

3.2.3　AI 应用模式的调节作用

Raisch 和 Krakowski (2021)总结了 AI 的两种应用模式, 即自动化 (automation) 和增强 (augmentation), AI 自动化与 AI 增强主要反映特定工作任务中 AI 与人的功能配置方式。在 AI 自动化模式下, AI 可以接管任务、包揽工作, 一般应用于基于规则的、结构良好的工作任务, 如工业制造、零售、酒店和餐饮等行业。因此, AI 自动化可能会强化 AI 拟人性与员工可雇佣性之间的负向关系, 即同等水平的 AI 拟人性, 当员工处在 AI 自动化模式下, 所感知到的可雇佣性更低。AI 自动化也可能会强化 AI 可解释性与员工感知领导 AI 信任之间的正向关系, 即同等水平的 AI 可解释性, 当处在 AI 自动化模式下, 员工所感知到的领导 AI 信任更强。在 AI 增强模式下, AI 可以拓展人类能力、辅助人类进阶, 一般应用于复杂的、规则和模型不是完全已知的工作任务, 如负责招聘的 HR 利用 AI 招聘程序进行简历初筛和首轮面试, 以辅助 HR 做出录用决策(张建民 等, 2022)。Mirbabaie 等(2021)认为, AI 增强模式下的员工(与 AI 协作互动是其工作中不可或缺的组成部分)不容易感知到工作被取代。首先, 与 AI 协作的 AI 增强模式能够在解决与工作相关的任务时为员工提供支持(Brachten et al., 2020), 将他们从繁杂、冗余的工作任务中解放出来(罗映宇 等, 2023), 减少他们在工作场所的感知工作量, 使他们执行任务更有效率; 其次, AI 增强模式下 AI 能够扩展员工的自我, 员工在与 AI 协作时可能把 AI 伙伴当作所有物或自我延伸的一部分, 纳入自我意识(Mirbabaie et al., 2021)。由此可见, AI 增强可能会缓解 AI 拟人性与员工可雇佣性之间的负向关系, 即同等水平的 AI 拟人性, 当员工处在 AI 增强模式下, 所感知到的可雇佣性更高。AI 增强也可能会减缓 AI 可解释性与员工感知领导 AI 信任之间的正向关系, 即同等水平的 AI 可解释性, 当处于 AI 增强模式下, 员工所感知到的领导 AI 信任更低。身份威胁过程理论强调, 情境特征之所以引发身份威胁, 关键在于个体如何解读这些特征对自我角色边界的含义(Petriglieri, 2011)。AI 应用模式划定了 AI 与员工之间的任务边界, 提供了员工解读自我角色边界的框架, 从而决定了 AI 拟人性和 AI 可解释性这两类 AI 特征信号被员工感知为威胁的程度。因此, 本研究提出如下命题:

命题 3: AI 应用模式在 AI 拟人性与员工可雇佣性的负向关系间起调节作用。当员工处于 AI 自动化模式时, AI 拟人性与可雇佣性的负向关系增强, 当员工处于 AI 增强模式时, AI 拟人性与可雇佣性的负向关系减弱。

命题 4: AI 应用模式在 AI 可解释性与员工感知领导 AI 信任的正向关系间起调节作用。当处于 AI 自动化模式时, AI 可解释性与员工感知领导 AI 信任的正向关系增强, 当处于 AI 增强模式时, AI 可解释性与员工感知领导 AI 信任的正向关系减弱。

命题 5: AI 应用模式调节员工可雇佣性在 AI 拟人性与 AI 工作身份威胁之间的中介效应。AI 自动化模式下, 员工可雇佣性的中介效应增强; AI 增强模式下, 员工可雇佣性的中介效应减弱。

命题 6: AI 应用模式调节员工感知领导 AI 信任在 AI 可解释性与员工 AI 工作身份威胁之间的中介效应。AI 自动化模式下, 员工感知领导 AI 信任的中介效应增强; AI 增强模式下, 员工感知领导 AI 信任的中介效应减弱。

3.3　研究 3: AI 工作身份威胁对员工创新行为的“双刃剑”效应机制

员工创新行为代表员工在不确定与风险情境中主动提出并实践的新构想、新方案, 以提升组织效能, 是组织持续改进与创新的动力来源(Ng & Lucianetti, 2016)。它直接体现了员工如何在 AI 情境下重新确认并展示自身的独特价值。AI 工作身份威胁如何通过不同的应对策略传导给员工创新行为? 何时对员工创新行为产生消极影响(身份保护策略)? 何时对员工创新行为产生积极影响(身份扩展策略)? 以往研究在这方面的实证检验不足(田喜洲 等, 2016)。本研究基于身份威胁过程理论, 引入特质调节焦点作为边界条件, 明晰不同类型特质调节焦点员工在感知 AI 工作身

份威胁后通过工作退却减少创新行为，或通过工作重塑增加创新行为的“双刃剑”效应机制。理论模型如图 2 所示。

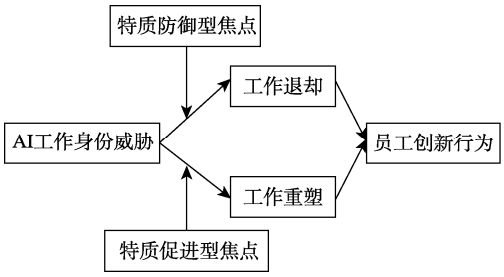


图 2 AI 工作身份威胁对员工创新行为的“双刃剑”效应模型

根据身份威胁过程理论，身份威胁涉及员工的不愉快感知(如不确定感)，由于员工不喜欢并希望规避，他们会想办法消除这些感觉(Lind & van den Bos, 2002)。研究表明，身份威胁可以触发员工的认知工作来分析这种感知和可能的应对资源，并促进他们对威胁的消极和积极反应，以处理威胁(Taylor, 1991)。不同特质个体可能通过相反的身份策略，如退却行为或改变行为来应对(Nach, 2015; Petriglieri, 2011)。本研究认为，特质调节焦点在决定个体是否主动应对身份威胁方面起重要作用(Shepherd & Williams, 2018)。特质调节焦点是个体在成长过程中形成的较为稳定的调节焦点特质倾向(Brockner & Higgins, 2001)，包括特质防御型焦点和特质促进型焦点，解释了员工在避免痛苦、追求快乐时的相反状态(Bipp & Demerouti, 2015)。其中，特质防御型焦点是指员工促使自己与应该自我(ought self，即自己所认为的应该状态)保持一致。特质防御型焦点员工更有可能看到 AI 工作身份威胁所带来的伤害和损失，对 AI 工作身份威胁的反应可能是一种基于回避的方式(Wang et al., 2024)，将会采取工作退却的消极应对策略(Chong et al., 2020)。工作退却是一种消极工作行为，表现为员工试图通过逃避工作来缓解压力，从而减少投入工作的时间和精力(Spector et al., 2006)，例如做白日梦、分心、迟到、早退、缺勤以及不合理的休息等。特质防御型焦点员工在感知到 AI 工作身份威胁后，可能通过这些退却行为和策略消极影响员工创新行为。因此，本研究提出如下命题：

命题 7：特质防御型焦点水平越高的员工，AI 工作身份威胁对工作退却的促进作用越强。

命题 8：特质防御型焦点水平越高的员工，工作退却在 AI 工作身份威胁与员工创新行为之间的中介效应越强。

特质促进型焦点是指员工出于成长和发展的需要，以问题为中心，以改进为目标，激励自己与理想自我(ideal self，即自己所希望的理想状态)保持一致(Higgins, 1987)。特质促进型焦点员工更有可能看到威胁中固有的潜在可能性，愿意进行身份尝试(Shepherd & Williams, 2018)，更有可能以主动的方式应对 AI 工作身份威胁，将威胁转化为利益(Wang et al., 2024)。这是因为，AI 相对于人类，创新思维仍显不足(韩明燕，赵静幽，2026)。员工创新行为的提升可能意味着原有的工作身份在创意领域得到了扩展，或者一种更有价值和意义的工作身份开始形成，从而使员工在 AI 工作身份威胁中实现收益。特质促进型焦点员工可能通过工作重塑的积极应对策略，正向影响员工创新行为，实现在创意领域的身份扩展，应对 AI 工作身份威胁。工作重塑是指员工为了满足个人或群体需求，获得工作意义，自下而上积极改变工作内容、方式与边界的主动行为(Akkermans & Tims, 2017)。它是一种自我导向行为，允许员工对自己的工作界限进行重新定义(Berg et al., 2013)，包括认知重塑、关系重塑和任务重塑。认知重塑是指员工对所从事工作看法的改变，关系重塑是指员工对工作互动对象和频率的改变，任务重塑是指员工对工作流程或范围的改变。特质促进型焦点员工在感知 AI 工作身份威胁后，可以采取这三种类型的工作重塑，以认知重塑改变自己对 AI 工作身份威胁的看法；以关系重塑改变与 AI 的互动方式，充分利用 AI 增强创新行为；以任务重塑改变当前工作流程，与 AI 协作提升创新行为。因此，本研究提出如下命题：

命题 9：特质促进型焦点水平越高的员工，AI 工作身份威胁对工作重塑的促进作用越强。

命题 10：特质促进型焦点水平越高的员工，工作重塑在 AI 工作身份威胁与员工创新行为之间的中介效应越强。

4 理论建构

AI 在工作场所的广泛应用使得员工越来越感

受到 AI 工作身份威胁。AI 相对于其他威胁源, 具有显著的独特性, 传统的身份威胁构念和理论框架需要进行扩展, 以便组织更好地实施干预管理。因此, 本研究认为有必要从理论上探索 AI 工作身份威胁的新要素和内涵, 诱发机制及其对员工创新行为的影响。基于此构想, 本研究通过三个紧密相连且层次递进的子研究, 构建以“AI 特征→AI 工作身份威胁→员工创新行为”为主线的理论框架, 如图 3 所示。研究 1 是关于 AI 工作身份威胁本身的研究, 尝试理论探索 AI 情境下工作身份威胁可能包含的维度和内容; 研究 2 从 AI 特征的角度探究 AI 工作身份威胁的诱发机制, 并分析 AI 应用模式的调节作用; 研究 3 构建“AI 工作身份威胁→工作退却或重塑→员工创新行为”的“双刃剑”影响路径, 同时从员工特质调节焦点探究边界条件。研究构想不仅为理解 AI 威胁源带来的工作身份威胁提供了新的理论视角, 也为采用 AI 的企业管理者识别和干预 AI 引发的员工工作身份威胁, 引导员工积极应对 AI 挑战提供理论依据和实践指导。

本研究的理论建构包括三个方面。首先, 聚焦 AI 的独特特征, 同时考虑工作身份范畴和 AI 威胁源的双重独特性, 将象征性−工具性框架作为启发性分析框架, 尝试探索 AI 工作身份威胁可能的新维度, 推动其构念探索。具体而言, 既有身份威胁研究主要刻画身份的象征性威胁(如身份价值、意义与呈现被质疑)。然而在 AI 情境下, 威胁不仅发生于“我是谁/我的价值高低”的象征性层面, 也可能发生于个体能否作为劳动力/人力资本在工作中占有不可替代的位置的工具层面。因此, 本研究尝试将象征性−工具性分析框架引入身份威胁过程理论, 初步将 AI 工作身份威胁内涵划分为两类相互关联但可区分的威胁内容：象征性威胁, 即员工的自我概念受到挑战, 如身份价值、身份意义与身份呈现威胁感知; 工具性威胁, 即作为劳动力的能力/功能受到挑战, 如身份独特性及存续性威胁感知。当然, 未来的构念开发也可能形成不同于当前设想的五维度结构。总之, 本研究能够拓展和深化对身份威胁概念的理解, 扩展身份威胁理论的内容和解释范围, 推动身份威胁领域的理论发展。由此也可见, 未来对身份威胁的研究需要结合具体情境, 更加深入和细致地探讨。未来研究也可以基于研究 1 所探索的构念进行相关理论模型的检验, 以扩展 AI 工作身份威胁的理论体系。

其次, AI 工作身份威胁并非由技术存在本身直接产生, 而是由 AI 特征触发个体对工作身份的认知与评估。本研究聚焦 AI 技术的独特且重要特征(AI 拟人性和 AI 可解释性), 以员工感知自身和感知领导的视角探究 AI 工作身份威胁的两条前因路径。以往研究探讨过传统技术因素如技术变革的程度和频率对身份威胁的影响(George et al., 2023), 但对于解释 AI 工作身份威胁的诱发机制存在不足。本研究从宏观的技术变革特征深化到微观的技术特征层面, 同时考虑员工感知自身和感知领导视角, 开拓性地构建了 AI 特征影响下的 AI 工作身份威胁前因模型。员工感知自身视角下, AI 拟人性强化了 AI 的类人特征, 使员工更可能进行能力比较与可替代性推断, 从而降低对自身可雇佣性与未来就业机会的预期, AI 工作身份威胁感增强; 员工感知领导视角下, AI 可解释性通过增强员工感知到的领导对 AI 的信任, 进而增强

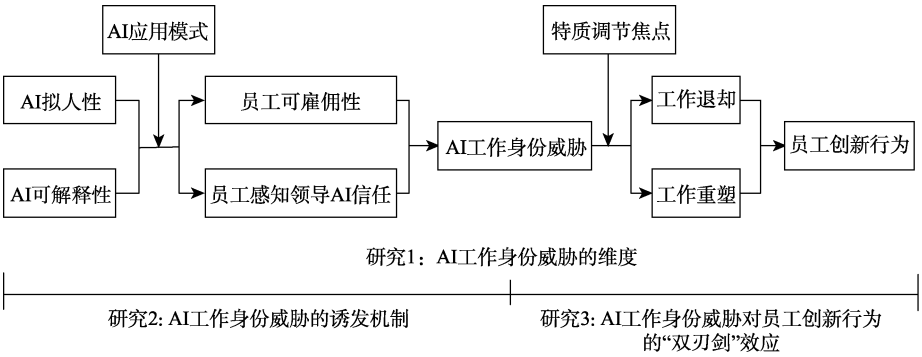


图 3　本研究构想整体理论框架

员工对自身被替代的感知, AI 工作身份威胁感增强。同时, 两种视角下均考虑 AI 应用模式的调节作用。研究结果不仅揭示了 AI 工作身份威胁的诱发过程, 也开辟了身份威胁领域相关议题研究的新思路。未来研究可以进一步探讨其他 AI 特征对 AI 工作身份威胁的诱发机制, 如 AI 关怀和温暖等(Kyung & Kwon, 2022), 也可以探讨不同类型的 AI 所产生的 AI 工作身份威胁差异。

最后, 本研究基于身份威胁过程理论, 揭示 AI 工作身份威胁对员工创新行为的“双刃剑”效应。在身份威胁过程理论中, 威胁激活后并不必然导向单一结果, 而是会通过不同身份策略进行传导。本研究提出 AI 工作身份威胁对员工创新行为呈现“双刃剑”效应, 其消极或者积极效应及内在机制取决于个体的特质调节焦点所驱动的身份策略选择。具体而言, 其一为退却路径：AI 工作身份威胁在防御型调节焦点个体中更易触发风险规避与身份保护, 表现为工作退却, 降低探索与试错, 从而抑制员工创新。其二为重塑路径：AI 工作身份威胁在促进型调节焦点个体中更可能被转化为成长导向的应对, 推动工作重塑与能力边界扩展, 通过在更具创造性与不可替代性的任务上重新确立价值, 从而促进员工创新。本研究为深入理解 AI 工作身份威胁对员工创新行为的影响及其机制提供了系统、全面的研究思路, 也是对 AI 情境下身份威胁个体认知和行为研究的有力补充。未来学者可以对 AI 工作身份威胁所导致的工作重塑策略和过程进行深入细致的剖析和探讨, 也可以进一步探讨 AI 工作身份威胁对其他工作结果的影响, 如任务绩效等。

综上, 身份威胁是学术界近年来关注的热点话题, 相关研究成果丰富(George et al., 2023)。本研究基于前期成果和相关理论基础, 力求在构念扩展和理论构建方面有所突破, 尝试性地对 AI 工作身份威胁的构念结构、前因及对员工创新行为的影响机制进行理论探究, 为深入分析 AI 赋能情境下企业如何干预其带来的工作身份威胁提供新的视角和理论工具。未来研究可以基于本研究理论框架, 进一步丰富 AI 工作身份威胁相关研究。

参考文献

韩明燕, 赵静幽. (2026). 员工−AI 合作对创新行为的非线性影响研究. *科研管理, 47*(7), 181−189.

罗映宇, 朱国玮, 钱无忌, 吴月燕, 黄静, 杨智. (2023). 人工智能时代的算法厌恶：研究框架与未来展望. *管理世界, 39*(10), 205−233.

彭剑锋. (2023). 新一代人工智能对组织与人力资源管理的影响与挑战. *中国人力资源开发, 40*(7), 8−14.

唐小飞, 孙炳, 张恩忠, 梅发贵. (2021). 类人智能机器人社会价值替代与风险态度研究. *南开管理评论, 24*(6), 4−15.

田喜洲, 董强, 马珂. (2016). 工作身份转变研究述评. *外国经济与管理, 38*(8), 81−94.

王林辉, 胡晟明, 董直庆. (2022). 人工智能技术、任务属性与职业可替代风险：来自微观层面的经验证据. *管理世界, 38*(7), 60−79.

王欣, 朱虹, 姜帝, 夏少昂, 肖春曲. (2021). 人工智能产品“协助者”与“替代者”形象对消费者评价的影响. *南开管理评论, 24*(6), 39−49.

王艳霞, 周珊. (2024). 个人-工作匹配视角下的人机协作身份威胁与工作重塑. *中国人力资源开发, 41*(8), 23−37.

谢小云, 左玉涵, 胡琼晶. (2021). 数字化时代的人力资源管理：基于人与技术交互的视角. *管理世界, 37*(1), 200−216.

张建民, 顾春节, 杨红英. (2022). 人工智能技术与人力资源管理实践：影响逻辑与模式演变. *中国人力资源开发, 39*(1), 17−34.

张志学, 华中生, 谢小云. (2024). 数智时代人机协同的研究现状与未来方向. *管理工程学报, 38*(1), 1−13.

赵曙明, 张敏, 赵宜萱. (2019). 人力资源管理百年：演变与发展. *外国经济与管理, 41*(12), 50−73.

Akkermans, J., & Tims, M. (2017). Crafting your career: How career competencies relate to career success via job crafting. *Applied Psychology, 66*(1), 168−195.

Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of Management, 40*(5), 1297−1333.

Arsenyan, J., & Mirowska, A. (2021). Almost human? A comparative case study on the social media presence of virtual influencers. *International Journal of Human-Computer Studies, 155*(1), 102694.

Berg, J. M., Dutton, J. E., & Wrzesniewski, A. (2013). Job crafting and meaningful work. In B. J. Dik, Z. S. Byrne, & M. F. Steger (Eds.), *Purpose and meaning in the workplace* (pp. 81−104). American Psychological Association.

Bipp, T., & Demerouti, E. (2015). Which employees craft their jobs and how? Basic dimensions of personality and employees' job crafting behaviour. *Journal of Occupational and Organizational Psychology, 88*(4), 631−655.

Brachten, F., Brünker, F., Frick, N. R. J., Ross, B., & Stieglitz, S. (2020). On the ability of virtual agents to decrease cognitive load: An experimental study. *Information Systems and e-Business Management, 18*(2), 187−207.

Brockner, J., & Higgins, E. T. (2001). Regulatory focus theory: Implications for the study of emotions at work. *Organizational Behavior and Human Decision Processes, 86*(1), 35−66.

Caza, B. B., Vough, H., & Puranik, H. (2018). Identity work

in organizations and occupations: Definitions, theories, and pathways forward. *Journal of Organizational Behavior, 39*(7), 889−910.

Chong, S., Huang, Y., & Chang, C. H. D. (2020). Supporting interdependent telework employees: A moderated-mediation model linking daily COVID-19 task setbacks to next-day work withdrawal. *Journal of Applied Psychology, 105*(12), 1408−1422.

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

Eury, J. L., Kreiner, G. E., Treviño, L. K., & Gioia, D. A. (2018). The past is not dead: Legacy identification and alumni ambivalence in the wake of the Sandusky scandal at Penn State. *Academy of Management Journal, 61*(3), 826−856.

Fiol, M. C., Pratt, M. G., & O'Connor, E. J. (2009). Managing intractable identity conflicts. *Academy of Management Review, 34*(1), 32−55.

Gabbiadini, A., Ognibene, D., Baldissarri, C., & Manfredi, A. (2025). The emotional impact of generative AI: Negative emotions and perception of threat. *Behaviour & Information Technology, 44*(4), 676−693.

George, M. M., Strauss, K., Mell, J. N., & Vough, H. C. (2023). When "who I am" is under threat: Measures of threat to identity value, meanings, and enactment. *Journal of Applied Psychology, 108*(12), 1952−1978.

Gerlich, L., Parsons, B. N., White, A. S., Prior, S., & Warner, P. (2007). Gesture recognition for control of rehabilitation robots. *Cognition, Technology & Work, 9*(4), 189−207.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627−660.

Greenbaum, R. L., Deng, Y., Butts, M. M., Wang, C. S., & Smith, A. N. (2022). Managing my shame: Examining the effects of parental identity threat and emotional stability on work productivity and investment in parenting. *Journal of Applied Psychology, 107*(9), 1479−1497.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*(3), 319−340.

Hillage, J., & Pollard, E. (1998). *Employability: Developing a framework for policy analysis* (Research Report No. RR85). Department for Education and Employment.

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior, 26*(6), 1508−1518.

Huang, H. L., Cheng, L. K., Sun, P. C., & Chou, S. J. (2021). The effects of perceived identity threat and realistic threat on the negative attitudes and usage intentions toward hotel service robots: The moderating effect of the robot's anthropomorphism. *International Journal of Social Robotics, 13*(7), 1599−1611.

Hunt, W., Sarkar, S., & Warhurst, C. (2022). Measuring the impact of AI on jobs at the organization level: Lessons from a survey of UK business leaders. *Research Policy, 51*(2), 104425.

Jin, S. V. (2023). To comply or to react, that is the question: The roles of humanness versus eeriness of AI-powered virtual influencers, loneliness, and threats to human identities in AI-driven digital transformation. *Computers in Human Behavior: Artificial Humans, 1*(1), 100011.

Kang, S., & Kim, J. W. (2022). The fragility of experts: A moderated-mediation model of expertise, expert identity threat, and overprecision. *Academy of Management Journal, 65*(2), 577−605.

Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing, 57*(1), 1−22.

Kilduff, G. J., Elfenbein, H. A., & Staw, B. M. (2010). The psychology of rivalry: A relationally dependent analysis of competition. *Academy of Management Journal, 53*(5), 943−969.

Kim, H., & Lee, S. W. (2026). Sorry, it's my fault: Politeness, attribution, and anthropomorphism in managing generative AI hallucinations. *International Journal of Information Management, 86*, 102996.

Kreiner, G. E., Hollensbe, E. C., & Sheep, M. L. (2009). Balancing borders and bridges: Negotiating the work-home interface via boundary work tactics. *Academy of Management Journal, 52*(4), 704−730.

Kyratsis, Y., Atun, R., Phillips, N., Tracey, P., & George, G. (2017). Health systems in transition: Professional identity work in the context of shifting institutional logics. *Academy of Management Journal, 60*(2), 610−641.

Kyung, N., & Kwon, H. E. (2022). Rationally trust, but emotionally? The roles of cognitive and affective trust in laypeople's acceptance of AI for preventive care operations. *Production and Operations Management, 34*(1), 1−20.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science, 33*(1), 126−148.

Li, G., Wang, Y., Zhu, H., & Zhu, M. (2026). The impact of AI service robots' cognitive anthropomorphism on consumers' subjective well-being − a moderated mediation model. *The Service Industries Journal, 46*(1-2), 1−26.

Liehr, J., & Hauff, S. (2025). Promoting employees' innovative work behavior through innovation-specific leader behavior: An AMO-approach. *Journal of Management & Organization, 31*(1), 215−232.

Lind, E. A., & van den Bos, K. (2002). When fairness works: Toward a general theory of uncertainty management. In B. M. Staw & R. M. Kramer (Eds.), *Research in organizational behavior: An annual series of analytical essays and critical reviews* (Vol. 24, pp. 181−223). JAI Press.

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking, 21*(10), 625−636.

Lyons, B. J., Lynch, J. W., & Johnson, T. D. (2020). Gay and lesbian disclosure and heterosexual identity threat: The role of heterosexual identity commitment in shaping de-stigmatization. *Organizational Behavior and Human Decision Processes, 160*, 1−18.

Ma, R., Fink, E. L., & Atwell Seate, A. (2024). When my group is under attack: The development of a social identity threat scale. *Group Processes & Intergroup Relations, 27*(4), 738−758.

MacInnis, D. J., & Folkes, V. S. (2017). Humanizing brands: When brands seem to be like me, part of me, and in a relationship with me. *Journal of Consumer Psychology, 27*(3), 355−374.

Marr, J. C., & Thau, S. (2014). Falling from great (and not-so-great) heights: How initial status position influences performance after status loss. *Academy of Management Journal, 57*(1), 223−248.

Mirbabaie, M., Stieglitz, S., Brünker, F., Hofeditz, L., Ross, B., & Frick, N. R. J. (2021). Understanding collaboration with virtual assistants − The role of social identity and the extended self. *Business & Information Systems Engineering, 63*(1), 21−37.

Miscenko, D., & Day, D. V. (2016). Identity and identification at work. *Organizational Psychology Review, 6*(3), 215−247.

Mishra, N., Ramanathan, M., Tulsulkar, G., & Thalmann, N. M. (2022). Uncanny valley for interactive social agents: An experimental study. *Virtual Reality & Intelligent Hardware, 4*(5), 393−405.

Mitchell, R. J., Parker, V., & Giles, M. (2011). When do interprofessional teams succeed? Investigating the moderating roles of team and professional identity in interprofessional effectiveness. *Human Relations, 64*(10), 1321−1343.

Nach, H. (2015). Identity under challenge: Examining users' responses to computerized information systems. *Management Research Review, 38*(7), 703−725.

Ng, T. W. H., & Lucianetti, L. (2016). Within-individual increases in innovative behavior and creative, persuasion, and change self-efficacy over time: A social-cognitive theory perspective. *Journal of Applied Psychology, 101*(1), 14−34.

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

Petrou, P., Bakker, A. B., & van den Heuvel, M. (2017). Weekly job crafting and leisure crafting: Implications for meaning-making and work engagement. *Journal of Occupational and Organizational Psychology, 90*(2), 129−152.

Piening, E. P., Salge, T. O., Antons, D., & Kreiner, G. E. (2020). Standing together or falling apart? Understanding employees' responses to organizational identity threats. *Academy of Management Review, 45*(2), 325−351.

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing, 85*(1), 131−151.

Qin, X., Chen, C., Yam, K. C., Cao, L., Li, W., Guan, J., Zhao, P., Dong, X., & Lin, Y. (2022). Adults still can't resist: A social robot can induce normative conformity. *Computers in Human Behavior, 127*(2), 107041.

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

Rothausen, T. J., Henderson, K. E., Arnold, J. K., & Malshe, A. (2017). Should I stay or should I go? Identity and well-being in sensemaking about retention and turnover. *Journal of Management, 43*(7), 2357−2385.

Rothwell, A., & Arnold, J. (2007). Self-perceived employability: Development and validation of a scale. *Personnel Review, 36*(1), 23−41.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206−215.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336−356.

Shepherd, D. A., & Williams, T. A. (2018). Hitting rock bottom after job loss: Bouncing back to create a new positive work identity. *Academy of Management Review, 43*(1), 28−49.

Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology, 104*(12), 1514−1534.

Shonhe, L., & Min, Q. (2025). Mitigating AI-induced professional identity threat and fostering adoption in the workplace. *AI & Society, 40*(5), 4079−4092.

Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviors created equal? *Journal of Vocational Behavior, 68*(3), 446−460.

Stein, J. P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior, 95*(3), 73−82.

Stein, J. P., & Ohler, P. (2017). Venturing into the uncanny valley of mind − the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition, 160*, 43−50.

Stohl, C., Stohl, M., & Leonardi, P. M. (2016). Managing opacity: Information visibility and the paradox of transparency in the digital age. *International Journal of Communication, 10*, 123−137.

Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 276−293). Psychology Press.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67−85.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of

developing AI for hiring. *MIS Quarterly, 45*(3), 1557−1580.
Venkatesh, V. (2021). Adoption and use of AI tools: A research agenda grounded in UTAUT. *Annals of Operations Research, 308*(4), 641−652.
Vescio, T. K., Schermerhorn, N. E. C., Gallegos, J. M., & Laubach, M. L. (2021). The affective consequences of threats to masculinity. *Journal of Experimental Social Psychology, 97*(5), 104195.
Vough, H. C., Bataille, C. D., Noh, S. C., & Lee, M. D. (2015). Going off script: How managers make sense of the ending of their careers. *Journal of Management Studies, 52*(3), 414−440.
Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *The International Journal of Human Resource Management, 33*(6), 1237−1266.
Wang, Y., Lau, D. C., & Kim, Y. (2024). Accentuating the positive: How and when occupational identity threat leads to job crafting and positive outcomes. *Journal of Occupational and Organizational Psychology, 97*(1), 131−159.
Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management, 29*(5), 907−931.
Yam, K. C., Tang, P. M., Jackson, J. C., Su, R., & Gray, K. (2023). The rise of robots increases job insecurity and maladaptive workplace behaviors: Multimethod evidence. *Journal of Applied Psychology, 108*(5), 850−870.
Yang, N., & Zhang, R. (2023). Under threat: Emotional and behavioral responses to occupational identity threat. *Journal of Management & Organization, 29*(3), 464−480.
Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction, 5*(2), 29−47.
Zhou, J., & Dou, J. (2023). A double-edged sword: When does identity threat affect unethical behavior? *Journal of Management & Organization, 29*(6), 1120−1138.

AI work identity threat: The construct, antecedents, and double-edged sword effect

HUANG Jie[1], LI Yali[2], ZHOU Kong[3], ZHU Run[2], ZHANG Yongjun[1]

([1] *School of Business, Henan University, Kaifeng* 475004, *China*)

([2] *School of Business, Nanjing University, Nanjing* 210008, *China)*

([3] *School of Economics & Management, Nanjing University of Science and Technology, Nanjing* 210094, *China*)

Abstract: The rapid advancement and widespread adoption of artificial intelligence (AI) have intensified employees' perceptions of threat to their professional identity, yet the connotation, antecedents, and consequences of AI work identity threat remain underexplored. Grounded primarily in identity threat process theory, this study develops a theoretical framework for AI work identity threat organized around three components: novel dimensions, triggering mechanisms, and consequences. First, based on the distinctive characteristics of AI as a source of threat, this study theoretically delineates new dimensions and conceptualizations of AI work identity threat. Second, focusing on two core characteristics of AI—anthropomorphism and explainability—it clarifies the triggering mechanisms and boundary conditions underlying AI work identity threat, namely the mediating roles of employee employability and perceived leader AI trust, and the moderating role of AI application patterns. Finally, the study uncovers a “double-edged sword” mechanism through which AI work identity threat shapes employee innovative behavior, operating through the mediating roles of work withdrawal and job crafting, and moderated by trait regulatory focus. This research is expected to extend and enrich identity threat process theory, offer a theoretical framework for understanding the connotation, antecedents, and consequences of AI work identity threat, and provide practical implications for organizations and employees in designing more targeted intervention strategies against such threats.

Keywords: AI work identity threat, AI characteristics, work withdrawal, job crafting, employee innovative behavior