

• 研究构想(Conceptual Framework) •

人工智能反馈寻求行为的驱动机制及其影响效应*

孙 芳¹ 李绍龙² 龙立荣³ 雷 宣² 曾祥麟² 黄夏虹²

(¹湖北经济学院工商管理学院, 武汉 430205)

(²武汉大学经济与管理学院, 武汉 430072) (³华中科技大学管理学院, 武汉 430074)

摘 要 面对高度易变、不确定、复杂和模糊的乌卡时代, 员工需要主动寻求反馈来实现个人发展并提升职场竞争力。人工智能技术为员工主动寻求反馈提供了新契机, 甲骨文公司的调研报告显示超过 50% 的员工倾向于主动向人工智能寻求反馈。然而传统反馈寻求行为的研究未将人工智能纳入反馈寻求目标, 员工向人工智能寻求反馈的驱动机制和影响效果尚不明晰。同时, 人工智能反馈的研究刚兴起, 且主要把员工作为反馈的客体, 缺乏对员工主动寻求反馈的关注。因此, 本文欲融合传统反馈寻求行为和新兴人工智能反馈的研究, 拓展反馈寻求行为的概念, 将人工智能纳入人类员工寻求反馈的对象之中。围绕人工智能反馈寻求行为, 本文欲探究人工智能系统的透明性和拟人性特征对其的驱动机制, 并探讨其对员工绩效改善的影响效果。在此基础上, 本文能为“新兴技术与员工的心理与行为”等前沿方向提供一些研究证据, 并为相关管理实践提供启示。

关键词 人工智能反馈寻求行为, 透明性, 拟人性, 绩效改善

分类号 B849: C93

1 问题提出

根据世界经济论坛(World Economic Forum) *The Future of Jobs Report 2023*¹ 的预测, 由于新技术的引进和人机之间劳动分工的变化, 导致当前在职劳动者所掌握的核心技能在未来 5 年内将有 44% 的比例需要进行更替和变化。由得到和中国人民大学劳动人事学院联合推出的《2022 年中国职场人群发展建议白皮书》² 中也指出, 超过 85% 的职场人士正面临职场困境与挑战, 职业发展瓶颈和职场能力不足成为困扰员工的核心问题。为了应对上述挑战, 由于在职培训和成人教育等渠道存在供给不足或成本高昂等局限, 越来越多的员工采用了在履职过程中主动寻求反馈的策略来

明晰角色定位和工作目标、提升职业技能和工作业绩, 从而提高自身的职场竞争力。例如, 领英(LinkedIn)在其发布的《千禧一代人才管理手册》³ 中就曾提出, 作为当今职场中的重要核心和中坚力量, 千禧一代员工倾向于主动寻求反馈。

员工在主动寻求反馈的过程中存在重要困境, 即反馈寻求目标主要局限于直接上级或同事(张颖, 杨付, 2018; Ashford et al., 2016)。而受限于知识技能水平, 人类管理者可能无法为主动寻求反馈的员工提供准确、客观的反馈内容(Tong et al., 2021)。同时, 由于管理者时间精力有限且存在“重考核, 轻反馈”的传统绩效管理理念, 主动寻求反馈的员工可能无法获得及时的回应(Pulakos et al., 2019)。随着人工智能(Artificial Intelligence, AI)技术的不断发展, 特别是生成式人工智能模型的技术跃迁, 为员工主动寻求反馈提供了新的契机。

收稿日期: 2024-10-30

* 国家自然科学基金项目(72302082, 72272114, 72132001)资助。

通信作者: 李绍龙, E-mail: tli@whu.edu.cn

¹ 来源见 https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf

² 来源见 https://mp.weixin.qq.com/s/yWf7_ZHYKbyQnqZqIcC44vQ

³ 千禧一代(Millennials), 主要指 1983 年到 1995 年间出生的一代人。报告来源见 https://business.linkedin.com/content/dam/me/business/zh-cn/talent-solutions/Event/2019/jul/millennium/millennial-talent_zh_CN.pdf

一方面,人工智能具备高效处理海量数据的能力,可以进行较为快速、准确、客观的评估(Luo et al., 2021)。另一方面,大型语言生成模型等技术的进步,使得人工智能可以提供优质的对话交互体验。目前,已经有一些“先锋”企业逐步在工作中采用人工智能反馈,例如保险业巨头 MetLife 和科技企业 Zoom (Luo et al., 2021)。鉴于此,除了向人类管理者寻求反馈之外,员工是否会主动向人工智能寻求反馈?又能取得怎样的工作结果?

根据甲骨文公司(Oracle)基于 8000 名参与者的 大样本问卷调查得出的“AI@Work Global Study”⁴, 50%以上的员工倾向于主动向人工智能(而不是人类管理者)寻求反馈。遗憾的是,现有文献中仅有少量研究探讨了人工智能等算法反馈对反馈接收者的影响(Luo et al., 2021; Tong et al., 2021; Yam et al., 2022)。且现有研究均主要将员工置于被动接受反馈的地位,尚无理论或实证研究探索员工主动寻求人工智能反馈的驱动机制和影响。因此,有必要将人工智能纳入反馈寻求目标中,以此适应组织现实场景和满足管理实践的需要。

在上述管理现实和研究局限的驱动下,基于反馈寻求行为和人工智能的文献,本文欲拓展反馈寻求行为的概念,将人工智能纳入人类员工寻求反馈的对象之中。本文将人工智能反馈寻求行为(Feedback-seeking from AI)定义为“组织中的个体有意识地向人工智能寻求有价值的信息以适应组织和个人发展需要的一种主动性行为”。上述人工智能包括但不限于机器人人工智能(Robotic AI)、虚拟人工智能(Virtual AI)和嵌入式人工智能(Embedded AI) (Glikson & Woolley, 2020)。在此基础上,本文拟从反馈寻求目标特征(即人工智能透明性和拟人性系统特征)和工作任务特征(即问题解决要求)切入,探索人工智能反馈寻求行为的驱动机制,并探讨人工智能反馈寻求行为对绩效改善的影响。

本文构建了一个系统的人工智能反馈寻求行为理论框架,具有理论意义和实践意义。理论意义方面,首先,本文拓展了反馈寻求行为这一概念,有利于深化人工智能背景下反馈寻求行为的概念研究。其次,本文探讨人工智能反馈寻求行

为的驱动机制和影响,能够丰富数字化转型对组织行为影响的相关研究。实践意义方面,本文为管理者如何有效引导员工从事人工智能反馈寻求行为,并获取相关裨益、规避可能风险,提供一些有针对性的实践指导。

2 研究现状分析

2.1 传统反馈寻求行为研究

2.1.1 反馈寻求行为的内涵

反馈寻求行为(Feedback-seeking behavior)由 Ashford 和 Cummings (1983)首次提出。具体而言,反馈寻求行为指的是组织中的个体为了确定自身行为的正确性和充分性以达到一种有价值的最终状态而主动进行的自我调节行为(Ashford, 1986)。在寻求反馈的过程中,个体采用的手段主要有两种——监控式反馈寻求和询问式反馈寻求(王雁飞, 2005; Ashford & Cummings, 1983)。现有研究主要关注询问式反馈寻求,即个体通过直接或间接询问来获得信息;少数研究关注监控式反馈寻求,即个体通过积极关注他人或环境线索从而推断出相关信息的一种间接手段(Ashford et al., 2016)。同时,反馈寻求的目标主要是直接领导和同事(张颖, 杨付, 2018; Ashford et al., 2016)。

2.1.2 反馈寻求行为的影响因素

(1)个体因素

第一,人口统计学特征。已有研究探讨了年龄、工作年限对员工反馈寻求行为的影响。例如, Gupta 等(1999)研究发现年龄对反馈寻求行为有显著影响,相比于年龄大的员工,年轻员工更愿意主动地寻求反馈。与之相类似,较多研究发现工作年限与员工反馈寻求行为负相关(例如 Anseel et al., 2015; Dahling & Whitaker, 2016)。

第二,特质倾向。在个体特质倾向方面,学者们主要探索了人格和目标取向对反馈寻求行为的影响。例如, Krasman (2010)发现,外倾性与员工的直接/间接询问式反馈寻求正相关。Gong 等(2017)指出,绩效目标导向会促进自我积极类和其他人消极类反馈寻求行为,学习目标导向会促进自我积极类、自我消极类和其他人积极类反馈寻求行为。

第三,动机因素。现有研究主要探讨了员工的工具性动机、自我保护动机以及印象管理动机等对其反馈寻求行为的影响。例如, Sherf 和 Morrison

⁴ 来源见 <https://www.oracle.com/webfolder/s/assets/ebook/ai-work/index.html>

(2020)研究发现, 员工对反馈寻求的价值感知与其反馈寻求行为正相关。林新奇和徐洋洋(2021)基于个体的职业发展动机, 发现员工的未来工作自我清晰度与其反馈寻求行为正相关。

第四, 认知因素。已有研究主要基于成本-价值框架、社会认同理论等, 探讨了个体的认知因素对其反馈寻求行为的影响。例如, 基于成本-价值框架, De Stobbeleir 等(2020)研究发现, 团队成员的心理安全感会促进团队成员之间的反馈寻求行为。基于社会认同理论, Nurmohamed 和 Schwingel-Sauer (2024)发现员工意识到自己是某个位置的替代选择会通过减少社会融合而负向影响员工的反馈寻求行为。

(2)情境因素

第一, 领导行为与风格。领导作为组织里最重要的反馈寻求目标, 其行为与风格受到了许多反馈寻求行为研究者的关注。例如, 研究发现真实型领导(张燕红, 廖建桥, 2015)、谦卑型领导(倪清, 杜鹏程, 2017)、伦理型领导(Moss et al., 2020)、授权型领导(王怀勇等, 2022)对下属的反馈寻求行为具有正向影响, 领导的可信度(Anseel et al., 2015)、领导-下属交换关系(Anseel et al., 2015)、领导幽默(任巍, 简浩, 2020)与下属的反馈寻求行为正相关, 领导奖励缺失与下属的反馈寻求行为负相关(张正堂等, 2018)。

第二, 工作环境。部分学者从组织、团队等工作环境切入, 探讨了这些因素对员工反馈寻求行为的影响。例如, 较多学者探究不确定性环境与反馈寻求行为的关系, 且结论并不一致。有的学者发现两者之间存在正相关关系(Ashford, 1986), 有的学者发现两者之间呈 U 型关系(Anseel & Lievens, 2007), 也有学者发现两者呈负相关(Anseel et al., 2015)。此外, Whitaker 和 Levy (2012)从反馈环境出发, 发现工作场所反馈质量将通过员工感知到的反馈效用来影响其反馈寻求行为。Zhu 等(2024)基于反馈干预理论, 发现职场中的负面领导八卦会通过任务反思或负面情绪两种路径, 对员工的反馈寻求行为产生积极和消极的影响。

2.1.3 反馈寻求行为的影响后果

第一, 对绩效的影响。许多学者探讨了反馈寻求行为与绩效之间的关系, 然而并未取得一致的结论。例如, 基于目标导向理论, Gong 等(2017)发现, 寻求关于自我的积极反馈与个体工作绩效

负相关, 寻求关于自我的消极反馈、关于他人的积极/消极反馈均与个体工作绩效正相关。张建平等(2020)的元分析发现, 员工的反馈寻求行为与其绩效正相关。然而, Anseel 等(2015)的元分析发现, 反馈寻求行为和绩效之间只有微弱的关系或者没有关系。

第二, 对积极工作行为的影响。现有研究主要从成本-价值框架、自我决定理论、解释水平理论等出发, 探讨了反馈寻求行为对员工的管理实践能力、创新、建言、领导行为等积极工作行为的影响。例如, Sherf 等(2021)的研究表明, 管理者的反馈寻求行为作为一种关键管理实践, 通过增强其对公平目标的关注度及对员工需求的洞察, 显著提升其管理决策的公正性水平。王宁等(2021)以及 Wang 等(2022)均发现, 向同事主动寻求反馈会正向影响该员工的创新行为。Coutifaris 和 Grant (2022)基于动态视角发现, 领导者的反馈寻求行为有助于增强团队的心理安全从而促进员工的建言。贾建锋等(2024)研究发现, 员工反馈寻求行为和领导的绩效改进动机因将交互影响领导的情绪, 进而影响员工获得上级发展性反馈。

第三, 对职业发展与社会化的影响。由于反馈寻求行为有助于实现目标并减少不确定性, 诸多研究均认为其是员工职业发展以及社会化过程中最重要的主动行为之一。例如, 王振源等(2018)在对师徒关系的研究中, 发现徒弟的反馈寻求行为对师徒关系质量具有正向影响。Vandenberghe 等(2021)通过对反馈寻求行为的动态研究发现, 随着新员工的反馈寻求行为减少, 其组织承诺也会相应减少, 从而导致其更高的离职意愿。

本文对反馈寻求行为的影响因素和影响后果进行了归纳整理和分析, 绘制了研究文献总结图, 如图 1 所示。

2.2 人工智能反馈研究

随着深度学习等技术的不断进步以及数据处理能力的不断增强, 人工智能技术的发展进入了全新阶段, 并在当今日益数字化的组织中发挥着越来越重要的作用(罗文豪等, 2022)。其中, 一个新实践领域逐渐兴起, 人工智能开始扮演反馈者的角色, 通过跟踪员工的工作与学习活动, 利用大数据和机器学习, 快速为员工提供及时、准确、个性化的建议, 从而提高员工的工作绩效(Tong

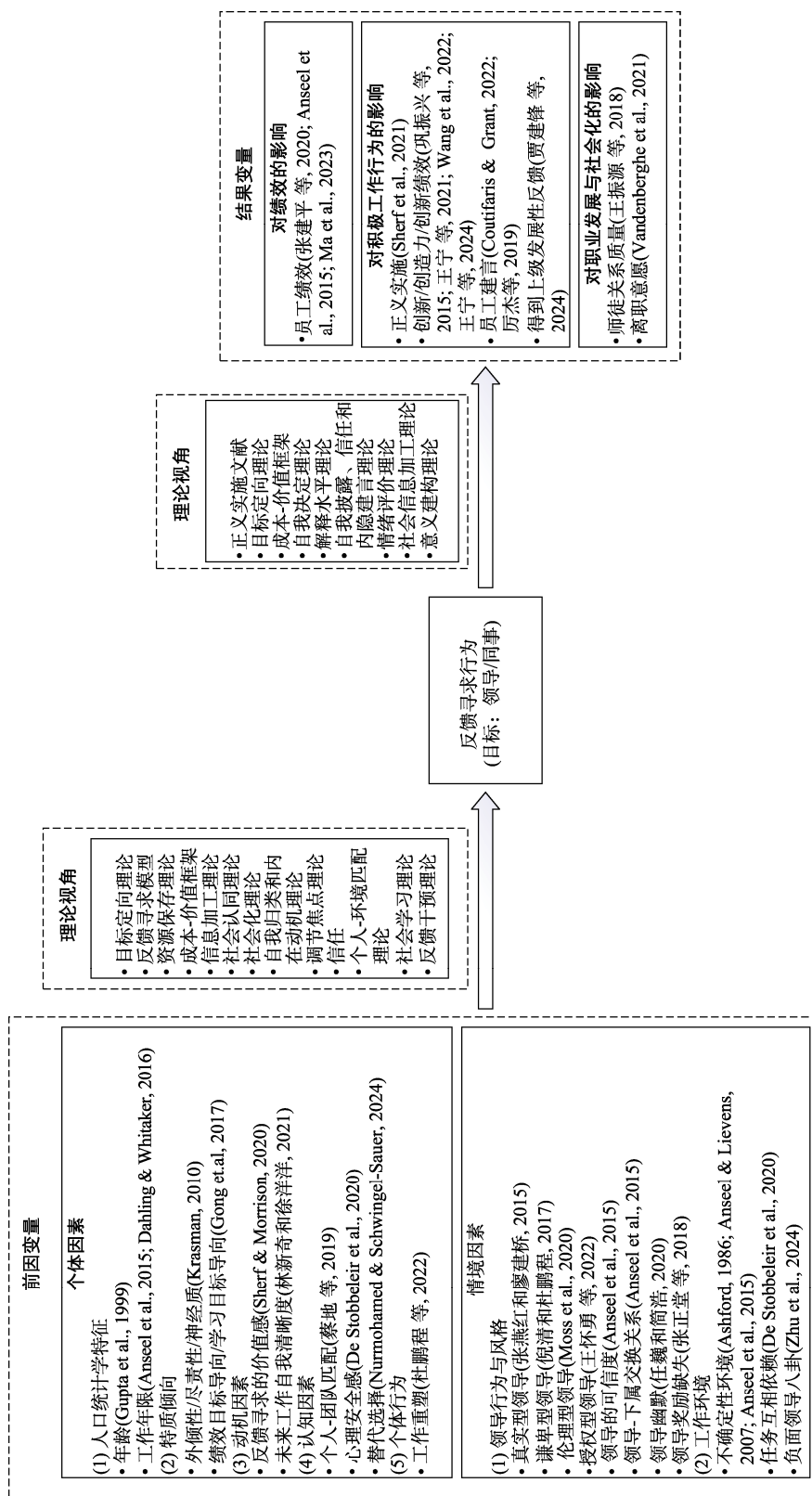


图1 文献总结图

et al., 2021)。与实践浪潮相适应,已有学者开始对人工智能反馈的影响效果进行研究,但是相关研究仍处于起步阶段。总结如下:

第一,人工智能反馈赋能观点。部分研究认为,人工智能所具备的先进的数据分析能力和学习能力,使得它能全面跟踪员工在工作中的行为,准确评估员工的表现,并给予员工个性化的、一致的和准确的反馈意见。因此,相比于人类管理者,由人工智能提供反馈更有助于员工提升绩效。例如,Tong等(2021)和Luo等(2021)发现,同人类管理者相比,人工智能产生的反馈建议对员工的工作绩效提高帮助更大。Hall等(2022)基于反馈过程理论和计划行为理论,发现一个人工智能反馈丰富的环境(即高反馈频率和具体性,及正面反馈效价)将正向影响其销售业绩与组织承诺。Förster等(2025)也发现,伴随着人工智能反馈出现的解释会正向影响人们感知到的人工智能反馈的信息量,从而对学习结果产生积极影响。Pei等(2024)基于调节焦点理论发现,对于担忧丢面子的员工,由人工智能来提供负面反馈能通过增加其学习动机、减少人际反刍从而增加工作绩效。

第二,人工智能反馈的负面影响。部分学者从不同视角出发,探索了人工智能反馈带来的消极影响。例如,Tong等(2021)发现,得知反馈是由人工智能提供的员工往往对反馈质量有较低的信任度,并且对工作取代风险表现出更多的担忧,从而阻碍了员工的学习提升。Thuillard等(2022)的实验结果显示,虽然由计算机和人类提供负面反馈都会给参与人员带来压力,但是由计算机提供负面反馈被认为比由人类提供负面反馈更不公平。Yam等(2022)的研究揭示,相较于非拟人化机器人主管,拟人化机器人主管在传递负面反馈时会显著增强员工的受虐感知(Perceived abusive supervision),进而引发更高频次的针对性报复行为。冯向楠和詹婧(2019)在对外卖骑手的调研中发现,以“智能语音助手”为代表的人工智能技术对骑手的实时指导反馈会导致骑手“去技术化”,降低劳动过程中的技术性。

2.3 人工智能背景下反馈寻求行为研究的挑战

本文在对反馈寻求行为和人工智能反馈领域的研究进行系统梳理之后,结合企业管理实践的趋势,发现以往研究存在以下两大方面的挑战:

第一,反馈寻求行为的研究与企业引入人工

智能技术进行数字化变革的现实背景相脱节,表现在反馈寻求行为研究仍主要聚焦于直接上级或同事等反馈寻求目标。经过近40年的发展,反馈寻求行为的内涵拓展有限,特别是反馈寻求目标仍局限于直接上级或同事(张颖,杨付,2018)。只有少数研究作了一定的拓展,例如将文件纳入反馈寻求目标中(Krasman, 2013)。这一局限可能与过往人工智能技术,特别是生成式人工智能模型的发展缓慢有关。但是近几年来,人工智能不仅可以利用大数据和深度学习技术进行及时、准确的评估(Tong et al., 2021),还可以提供较为优质的对话功能(包括图文和语音对话)。反馈寻求行为的主要策略即为询问式反馈寻求,因此向人工智能寻求反馈可能成为一种趋势。综上所述,反馈寻求行为的研究亟待将人工智能纳入反馈寻求目标中,从而拓展该研究领域。

第二,人工智能反馈的研究处于非常初期的起步阶段,且将员工置于被动接受反馈的位置,缺乏对员工在反馈过程中发挥主观能动性的考量。如果仅仅将人工智能反馈作为管理者的替代(智能自动化),或者作为管理者的辅助手段(智能增强),则可能诱发员工“全景监狱”或“数字泰勒主义”的担忧(刘善仕等, 2022; Kellogg et al., 2020)。实质上,不仅仅是在人工智能反馈领域,在数智化技术引入企业管理的历史过程中一直伴随着人类主体和技术主体优先级与从属关系的讨论(谢小云等, 2021)。过分强调人工智能等技术的主体性,而将人置于从属地位,可能无法促进人工智能与人之间的互动与增益,甚至激起人的“反抗”,最终不利于管理的有效性。学界已经开始呼吁,在利用人工智能等技术带来的准确性和及时性等管理优势的同时,更多地吸纳人本主义的思想(孙伟平, 2020)。相关研究应该促进人类智能与人工智能的有机融合,关注人和人工智能之间相互影响和塑造的关系。鉴于此,在进行以人类员工为客体的人工智能反馈研究的同时,应该开展人类员工主动向人工智能寻求反馈的研究,探讨人工智能反馈寻求行为的驱动机制以及影响后果,从而丰富该研究领域。

3 研究构思

3.1 研究目标

基于上述分析,本文欲将反馈寻求行为的目

标拓展至包含人工智能。围绕人工智能反馈寻求行为,本文希望实现以下具体研究目标:

(1)从人机交互的视角切入,明确驱动员工从事人工智能反馈寻求行为的系统特征因素及机制。具体而言,结合心智感知理论(Mind perception theory; Gray et al., 2007)和信任的相关文献(McAllister, 1995),探究透明性(Transparency)和拟人性(Anthropomorphism)这两种系统特征因素会怎样影响员工的人工智能反馈寻求行为。特别是挖掘员工对人工智能的能动性感知(Perceived agency)和感受性感知(Perceived experience)这两种心智感知,及认知信任(Cognition-based trust)和情感信任(Affect-based trust)这两种信任类型在上述影响过程中的链式中介作用,并探讨问题解决要求(Problem solving)这一工作任务特征的调节作用。

(2)落脚于反馈研究领域关注的核心结果,明晰员工的人工智能反馈寻求行为对其绩效改善(Performance improvement)的影响及机制。具体而言,结合反馈过程理论(Feedback process theory; Ilgen et al., 1979)和人工智能的相关文献,探讨员工的人工智能反馈寻求行为是否会通过影响人工智能信息的准确性(Accuracy)和具体性(Specificity),进而影响其绩效改善。此外,还欲挖掘客观任务/主观任务这一工作任务特性在上述影响过程中的调节效应。

3.2 研究 1: 系统特征对员工从事人工智能反馈寻求行为的影响及机制研究

在 Ashford 等(2016)整理的反馈寻求行为的研究框架中,反馈寻求目标的特征被认为是影响反馈寻求者与反馈寻求目标之间互动过程的关键因素。据此,本研究认为需要从人工智能系统的特征切入,探究员工从事人工智能反馈寻求行为的驱动机制。《2024 年人工智能现状报告》指出,当前的人工智能需要应对的复杂挑战之一即“黑箱模型决策过程神秘性挑战”。与此同时,在越来越多的应用场景中,人工智能系统被设计成具有拟人化特征以增强用户与之互动的意愿。回顾人机交互的研究发现,信任是用户与人工智能交互的关键期望属性。如果人工智能系统无法获得用户的信任,即使是最先进的技术也无法奏效,反而可能引发算法厌恶(Glikson & Woolley, 2020)。Park 等人(2008)深入探讨了影响人机信任的因素,

并提出了人机信任的三要素模型,即机器人系统自身的内部因素、个体自身的特征,以及与环境相关的因素。Hancock 等人(2011)进一步将机器人系统自身的内部因素细分为两个维度:一是与机器人性能相关的因素,具体包括透明性、故障率和自动化水平等;二是与机器人属性相关因素,比如拟人性、机器人个性和适应性等。Ochmann 等人(2024)指出,目前主要通过两种策略来塑造人们对算法的认知:一是通过提供算法底层过程的信息来建立透明度;二是通过拟人化设计来增加个体对算法决策者的熟悉度。因此,在员工向人工智能系统寻求反馈的过程中,系统的透明性和拟人性有可能影响员工的人工智能反馈寻求行为。

综上所述,本研究的第一个研究问题即探索人工智能系统的透明性和拟人性特征是否会影响员工的人工智能反馈寻求行为。同时,整合心智感知理论(Gray et al., 2007)和信任的相关文献(McAllister, 1995),本研究认为员工对人工智能的能动性和感受性心智感知,以及认知和情感信任会分别在透明性和拟人性这两种系统特征对员工的人工智能反馈寻求行为的影响中起到链式中介作用。此外,反馈寻求者的工作任务特征也是传统反馈寻求行为研究中考量的重要因素(Ashford et al., 2016),故本研究还推断问题解决要求这一工作任务特征会调节上述影响。据此,建构如图 2 所示的研究模型。

3.2.1 人工智能透明性和拟人性系统特征对员工的人工智能反馈寻求行为的影响

人工智能透明性即个体对人工智能决策过程和结果等相关信息的可获得程度(蒋路远等, 2022)。透明性有助于个体增强对人工智能决策过程和结果的理解,克服人机交互过程中的“黑箱”障碍,使他们能够有效地对评分结果进行质疑,减轻使用算法系统相关的负面后果如算法偏见等,从而提高个体使用人工智能时的知觉行为控制和使用意愿(Bauer & Gill, 2024)。例如,研究发现智能聊天机器人的透明性会影响用户对其所提供建议的依赖(Wilkinson et al., 2021)。因此,本研究推断:

命题 1a: 人工智能的透明性系统特征会正向影响员工的人工智能反馈寻求行为。

人们经常将非人类主体进行拟人化以理解其复杂行为,尤其是在与计算机、机器人和智能代理进行交互时(Saffarizadeh et al., 2024)。人工智能

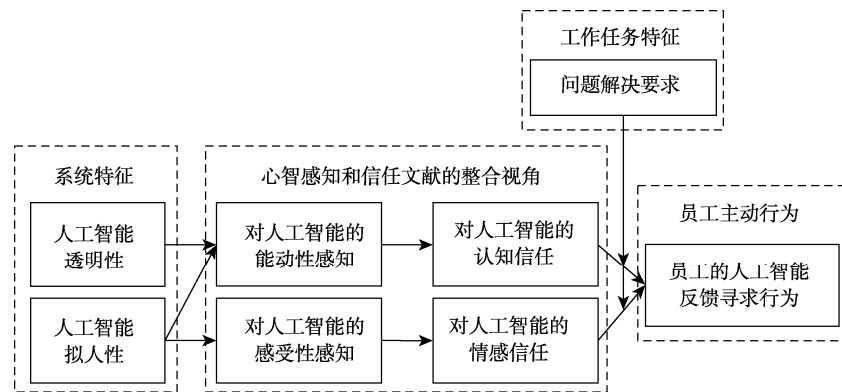


图2 研究1模型

的拟人性即人工智能被赋予意图、情感或动机等人类特征的程度(Yam et al., 2022), 是影响人机互动的关键因素(Epley et al., 2007)。社会反应理论指出, 个体在进行人机互动时, 倾向于将计算机视为具有社会性的行为角色, 而不仅仅是一个工具或机器, 并且采用他们在人际交往中所使用的社会规则和行为准则, 对计算机产生与对人类个体相同的期望(Nass & Moon, 2000)。当人工智能机器人展现出与人类相似的特征或行为时, 人们更有可能将其视为“真实的人”, 并与之进行更深入的互动(Epley et al., 2007)。相关研究发现, 智能聊天机器人的拟人性会增加消费者的技术采用意图(Shreehan et al., 2020)和购买意愿(Han, 2021)。人工智能产品拟人化有助于个体改善对人工智能决策具有低互动性、缺乏人际接触和尊重的感知(Acikgoz et al., 2020)。因此, 在反馈寻求过程中, 拟人化的人工智能通过模仿人类的语言表达、行为或表情等, 提供了更自然的交互体验, 很可能使员工更容易接受和理解人工智能的反馈并将其视为“具备提供反馈能力的人”, 从而主动向其寻求反馈。综上, 本研究推断:

命题 1b: 人工智能的拟人性系统特征会正向影响员工的人工智能反馈寻求行为。

3.2.2 员工对人工智能的能动性感知及认知和情感信任的链式中介作用

根据心智感知理论, 人类会通过能动性感知和感受性感知形成对各类实体的心智状态的感觉和理解(Gray et al., 2007)。能动性感知反映了人类对实体所具备的思考、计划和自主行动等能力的程度的感觉, 而感受性感知反映了人类对实体所具备的感知情绪、痛苦、快乐等能力的程度的感

觉(Gray et al., 2007)。具备较高透明性的人工智能系统设计, 例如对相关代码模块的意图和目的的解释, 有助于员工更好地理解系统的运作机制(Shin & Park, 2019)。当前人工智能技术已经具备了很强的自主规划和运行的能力(Raisch & Krakowski, 2021), 通过高透明性的系统设计, 员工更易于感知到人工智能的自主能力, 进而对人工智能产生较高的能动性感知。据此, 提出以下命题:

命题 2a: 人工智能的透明性系统特征会正向影响员工对其能动性感知。

在社会互动中, 个体出于理解、预测他人行为、建立社会联系及认知世界的需求, 会对他人进行心智感知的推断, 目标个体的特征会对感知者的心智感知过程产生影响, 进而影响其态度和行为(Waytz & Norton, 2014)。当人工智能被拟人化, 即被赋予更多的人类特征时, 个体可能会认为其具有更高的自主性、决策能力和行动意图(Yam et al., 2021)。已有研究表明拟人化增加了人们对动物(Wegner & Gray, 2017)和机器人(van der Woerd & Haselager, 2019)等的能动性感知。与此同时, 拟人化也增加了人们对动物、环境和其他实体在受伤时感到痛苦、在得到帮助时感到快乐的感知(Butterfield et al., 2012)。例如, 当狗被拟人化时, 人们对它被困在汽车中的痛苦更为敏感(Butterfield et al., 2012)。同样, 当机器人展现出类似人类的特征时, 人们倾向于认为它们具有丰富的情感体验能力, 包括感受快乐和痛苦(Gray & Wegner, 2012)。具备较高拟人性的人工智能系统设计, 例如拥有人类头像的交互界面、使用了类似人类的声音、冠以人类常用名字, 由于被赋予

了更多类似人类的特征,会使员工觉得其与人类非常相似,也具备动机、意图和情感(Epley et al., 2007; Yam et al., 2022)。有关动机和意图等能力的判断会使员工倾向于感知到高拟人性的人工智能系统具备较强的能动性,而有关情感能力的判断会使员工同时感知到高拟人性的人工智能系统具备较强的感受性。据此,本研究提出:

命题 2b: 人工智能的拟人性系统特征会正向影响员工对其能动性和感受性的感知。

在此基础上,本研究认为员工对人工智能的心智感知会影响其对人工智能的信任。其中,McAllister (1995)认为信任可以进一步分为认知信任和情感信任。所谓认知信任,主要考量的是对方的能力和可靠性;而情感信任,主要考量的是对方是否存在人际关怀和情感投入(McAllister, 1995)。能动性感知涉及个体对人工智能的思考、规划和自我控制能力的评价(Gray et al., 2007)。如果员工感知到人工智能系统具备较强的能动性,则其可能会认为人工智能系统具有解决问题、执行任务和适应环境变化的能力,这些能力是认知信任的关键因素(McAllister, 1995)。与此同时,感受性感知是人性化的体现(Heflick et al., 2011)。如果员工感知到人工智能系统具备较强的感受性,则其可能会将人工智能系统视为社会行为者。在与人工智能交互的过程中,员工对人工智能的感受性感知越高,越会认为它们能够体验和理解人类的情感需求,并在人际互动中展现出关怀和同理心(Klein et al., 2002)。而对情感的感知和理解能力,是情感信任建立的基础(McAllister, 1995)。据此,本研究提出:

命题 2c: 员工对人工智能能动性的感知会正向影响其对人工智能的认知信任。

命题 2d: 员工对人工智能感受性的感知会正向影响其对人工智能的情感信任。

信任是人机协作中非常重要的影响因素(王振源, 姚明辉, 2022)。对机器人的信任程度对后续人们是否接收和利用机器人提供的信息有着直接影响(Glikson & Woolley, 2020)。由于人工智能系统具备高度的复杂性和不确定性,人类和人工智能的关系中蕴含着较强的风险感,这些因素导致信任会显著影响人类和人工智能的关系(Glikson & Woolley, 2020)。研究表明,对人工智能的认知信任和情感信任会增加其对人工智能系统的使用

(例如 Seitz et al., 2022)。综上所述,提出以下命题:

命题 2e: 员工对人工智能的认知和情感信任会正向影响其人工智能反馈寻求行为。

结合命题 2a 到 2e 的理论推导,本研究提出以下三个链式中介效应假设:

命题 2f: 人工智能的透明性系统特征通过员工对其能动性的感知与认知信任的链式中介作用对员工的人工智能反馈寻求行为产生正向影响。

命题 2g: 人工智能的拟人性系统特征通过员工对其能动性的感知与认知信任的链式中介作用对员工的人工智能反馈寻求行为产生正向影响。

命题 2h: 人工智能的拟人性系统特征通过员工对其感受性的感知与情感信任的链式中介作用对员工的人工智能反馈寻求行为产生正向影响。

3.2.3 问题解决要求的调节作用

在 Ashford 等(2016)总结的反馈寻求行为研究框架中,工作任务特征被认为是影响反馈寻求行为的重要调节因素。随着科技的进步与组织流程的不断优化,现代工作环境的一个显著特征是对知识相关需求的不断增长,其中一个关键的知识工作特征即问题解决要求(Gajendran et al., 2022)。问题解决要求指的是工作在多大程度上需要独特的创意或解决方案,反映了对工作更高的认知处理要求(Morgeson & Humphrey, 2006)。在高问题解决要求的工作中,员工工作涉及产生独特且创新的想法或解决方案,诊断和解决非例行问题,以及预防工作错误并及时纠正错误(Gajendran et al., 2022)。根据资源保存理论,个体倾向于保护、获取和维持对其有价值的资源,如时间、认知能力、社交支持等(Hobfoll, 1989)。当工作面临较高的问题解决要求时,员工需要运用丰富的知识、技能和创造力来应对非例行的问题,这往往导致大量的资源消耗和压力增加(Schmitt et al., 2012)。在这种情况下,员工需要通过获取外部资源来补充自身资源的不足。反馈作为一种重要资源,能够帮助员工评估自己的表现并做出调整(Ashford & Cummings, 1983)。特别是人工智能系统,它能够迅速处理大量信息,提供即时且个性化的反馈和建议,帮助员工应对复杂任务(Tong et al., 2021)。当工作中的问题解决要求较高时,如果员工对人工智能具有较高的认知信任和情感信任,他们更可能将人工智能视为一个有价值的资源来源而主动向其寻求反馈。相反,当工作的问

题解决要求较低时,员工在工作中的资源损耗较少,即使员工对人工智能具有较高的认知信任和情感信任,也可能不会频繁地向人工智能寻求反馈。据此,提出以下调节效应命题:

命题 3a: 工作的问题解决要求会调节其对人工智能的认知和情感信任与其人工智能反馈寻求行为之间的正向关系。具体而言,相较于问题解决要求较低的工作,在问题解决要求较高的工作中,若员工对人工智能存在较强的认知和情感信任,则其更有可能主动向人工智能寻求反馈。

结合上述链式中介效应和调节效应的推导过程,本研究提出以下三个被调节的链式中介效应:

命题 3b: 工作的问题解决要求会调节人工智能的透明性系统特征通过员工对其能动性的感知与认知信任而正向影响员工的人工智能反馈寻求行为的链式中介效应。具体而言,相较于问题解决要求较低的工作,在问题解决要求较高的工作中,人工智能的透明性系统特征通过员工对其能动性的感知与认知信任的链式中介作用而对员工的人工智能反馈寻求行为产生的正向影响更强。

命题 3c: 工作的问题解决要求会调节人工智能的拟人性系统特征通过员工对其能动性的感知与认知信任而正向影响员工的人工智能反馈寻求行为的链式中介效应。具体而言,相较于问题解决要求较低的工作,在问题解决要求较高的工作中,人工智能的拟人性系统特征通过员工对其能动性的感知与认知信任的链式中介作用而对员工的人工智能反馈寻求行为产生的正向影响更强。

命题 3d: 工作的问题解决要求会调节人工智能的拟人性系统特征通过员工对其感受性的感知与情感信任而正向影响员工的人工智能反馈寻求

行为的链式中介效应。具体而言,相较于问题解决要求较低的工作,在问题解决要求较高的工作中,人工智能的拟人性系统特征通过员工对其感受性的感知与情感信任的链式中介作用而对员工的人工智能反馈寻求行为产生的正向影响更强。

3.3 研究 2: 员工的人工智能反馈寻求行为对其绩效改善的影响及机制研究

在明晰人工智能反馈寻求行为的驱动机制的基础上,本研究欲进一步探讨员工主动向人工智能寻求反馈是否能够带来裨益。绩效一直是传统反馈寻求行为研究关注的重要结果变量(Ashford et al., 2016),然而相关研究并未取得一致的结论(张建平 等, 2020; Anseel et al., 2015; Gong et al., 2017)。因此,本研究欲聚焦于探究员工的人工智能反馈寻求行为对其绩效改善的影响。具体而言,本研究基于反馈过程理论(Ilgen et al., 1979),并结合人工智能的相关文献,探讨人工智能反馈信息的准确性和具体性在上述影响过程中所起的中介作用。在此基础上,进一步明确客观任务/主观任务这一工作任务特征对上述影响过程的调节效应,从而明晰哪些工作任务更适合引导员工从事人工智能反馈寻求行为以获得更多的绩效改善。据此,建构图 3 所示的研究模型。

3.3.1 员工的人工智能反馈寻求行为对其绩效改善的影响

传统反馈寻求行为的研究并未获得有关反馈寻求行为对员工绩效影响的一致结论(Anseel et al., 2015),其中一个重要原因可能是人类管理者存在认知局限、经验不足,导致其提供的反馈可能准确性不佳、具体性不足(Ashford et al., 2016)。因此,业界开始在绩效反馈系统中引入人工智能,

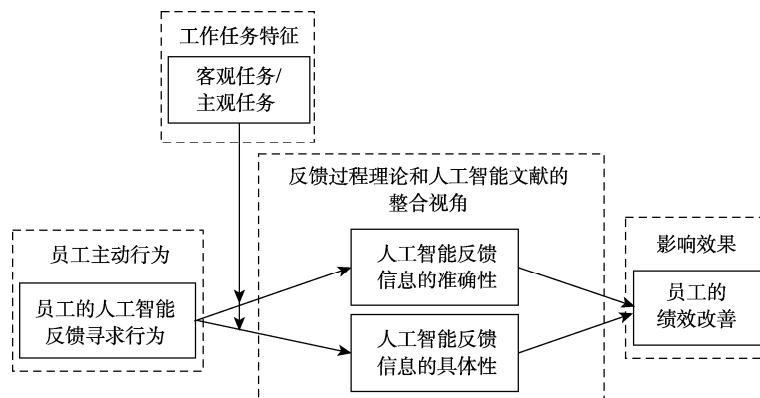


图 3 研究 2 模型

以利用人工智能所具备的先进的数据分析能力,从而提供更为有效的反馈(Luo et al., 2021; Tong et al., 2021)。然而,该领域的研究也并未取得人工智能反馈对绩效存在正面影响的一致结论。例如,Tong 等(2021)发现如果员工得知反馈是由人工智能提供的,则不利于其学习提升。Luo 等(2021)发现对于业绩靠前的销售人员而言,人工智能反馈对绩效的提升有限。上述研究中,接受反馈的员工均为人工智能反馈的客体,存在算法厌恶、低信任等倾向,从而导致人工智能反馈效果不佳。本研究探讨的人工智能反馈寻求行为,反映了员工主动向人工智能寻求反馈,不存在算法厌恶等倾向。因此,员工可以充分利用人工智能反馈优势所带来的好处,并进而提升其绩效。据此,提出以下命题:

命题 4: 员工的人工智能反馈寻求行为对其绩效改善有正向影响。

3.3.2 人工智能反馈信息的准确性和具体性的中介作用

反馈信息准确性反映了反馈内容的正确与否和客观性(Brett & Atwater, 2001),而反馈信息具体性反映了反馈内容的详细程度(Goodman et al., 2011)。基于大数据、物联网等数据收集手段,以及深度学习和神经网络等数据分析和预测技术,人工智能可以全面跟踪并评估员工的工作行为(张敏,赵宜萱,2022; Tong et al., 2021)。与此同时,人工智能具有迭代优化的能力,输入信息越多,越有利于产生具体和准确的信息(Raisch & Krakowski, 2021)。向人工智能寻求反馈已不仅仅是“工具调用”,而是一种实时交互的动态认知协同过程(Turel, 2024)。因此,本研究认为,员工更高水平的人工智能反馈寻求行为会通过提供更丰富多样的数据信息、模型的训练优化和系统迭代学习的循环,使得人工智能系统具备条件提供更加准确和具体的反馈信息。

在向人工智能寻求反馈的过程中,更高水平的反馈寻求行为意味着员工将通过多模态交互界面,更频繁、主动地提供多维度组织情景数据来向人工智能寻求关于自己在组织中以及个人发展的多方面信息。这些数据的广泛性和多样性为人工智能系统提供了更加全面的员工“画像”,从而有利于其更加客观、准确地评估员工的工作表现(Tong et al., 2021),并驱动人工智能系统自动关

联更广泛的资料、调取更多的信息,显著提升跨任务反馈的一致性。例如,Choi 等(2025)通过对 1,241 名专业围棋选手的 24,973 场重大比赛进行分析发现,在采用了人工智能系统进行训练以后,选手的落子与人工智能建议的最佳落子的一致性增加,选手的最佳落子准确性提升幅度达 30.5%。据此,提出以下命题:

命题 5a: 员工的人工智能反馈寻求行为对人工智能反馈信息的准确性存在正向影响。

更高水平的人工智能-员工互动产生的更多上下文信息有助于促进人工智能模型对模糊区域的显式推理,这些数据有助于人工智能系统动态地针对某个个体行为建模,根据员工的工作模式和行为特征,更准确地分析每个员工在工作中的独特情况和挑战,为员工提供个性化和适应性的反馈(Brynjolfsson & Mitchell, 2017; Wambsganss et al., 2025)。并且,员工对人工智能系统提供的反馈信息的响应和后续行为,又会生成用于训练的新数据点,进一步优化模型。这种迭代学习的循环使得人工智能系统能够持续提升其反馈信息的具体性(Bauer et al., 2024; Raisch & Krakowski, 2021)。

据此,提出以下命题:

命题 5b: 员工的人工智能反馈寻求行为对人工智能反馈信息的具体性存在正向影响。

基于反馈过程理论,反馈信息的准确性是影响反馈效果的关键因素之一,能够减少员工对信息的怀疑并增强其对信息的接受意愿,并进而促使员工对反馈进行行为响应(Ilgen et al., 1979; Steelman et al., 2004)。已有研究表明,人类对算法决策存在显著的厌恶心理(Dietvorst et al., 2015),而后续研究进一步指出,这种现象的存在使得人工智能系统必须达成更高水平的决策准确性才能获得用户接纳(Mahmud et al., 2022)。一旦人工智能产生了不够准确的决策,人们便会对其失去信心(Bogert et al., 2021; Dietvorst et al., 2015)。因此,若人工智能反馈信息具有较高的准确性,则员工更可能接受反馈,并对反馈进行行为响应,从而有助于绩效改善。Tong 等(2021)研究发现接受来自人工智能更具有准确性的反馈信息的员工比接受来自人类管理者反馈的员工工作绩效高出 12.9%。据此,提出以下命题:

命题 5c: 人工智能反馈信息的准确性会正向

影响员工的绩效改善。

根据反馈过程理论,反馈的具体性同样是影响员工反馈接受度和行为响应的关键因素(Ilgen et al., 1979)。具体的反馈信息能够为员工提供更为明确的修正方向,帮助员工制定具体的行动计划以改善其行为表现(Ilgen et al., 1979)。而高具体性的人工智能反馈可以针对特定员工行为进行细致分析,为其提供个性化和丰富的具体改进建议(关键等, 2025),从而让员工能更有效地采取行动,最终有助于其绩效改善。据此,提出以下命题:

命题 5d: 人工智能反馈信息的具体性会正向影响员工的绩效改善。

结合命题 5a 到 5d 的理论推导,本研究提出以下两个中介效应假设:

命题 5e: 人工智能反馈信息的准确性会中介员工的人工智能反馈寻求行为对其绩效改善的正向影响。

命题 5f: 人工智能反馈信息的具体性会中介员工的人工智能反馈寻求行为对其绩效改善的正向影响。

3.3.3 客观任务/主观任务的调节作用

员工的反馈寻求行为一般是针对工作任务进行的,故有必要探讨工作任务特征的调节作用。本研究欲探讨客观任务/主观任务这一工作任务特征在上述影响过程中的调节效应。客观任务被定义为可量化和可测量的事实任务,而主观任务被定义为基于个人观点或直觉的开放或可解释的任务(Castelo et al., 2019)。以往人工智能的相关研究发现,在执行结构化和定义明确的任务时,人工智能技术所具备的卓越的数据分析和逻辑推演能力更有利于其发挥优势(Castelo et al., 2019; Tong et al., 2021)。由于客观任务依赖基于逻辑和规则的分析(Inbar et al., 2010),故人工智能可以有效地处理客观任务,并输出具体、准确的结果。据此分析,针对客观任务,即使员工的人工智能反馈寻求行为较少,人工智能输出信息的准确性和具体性依然会保持较高的水准。

主观任务更依赖感觉和直觉(Inbar et al., 2010),一直是人工智能的短板,人们甚至在主观任务中存在对人工智能的算法厌恶(Castelo et al., 2019)。劣势中存在机会,更好地完成主观任务正是人工智能未来可以突破的领域。实际上,人工智能在处理主观任务方面的能力已经得到了很大提升,

而这与大规模数据的训练和学习密切相关。ToM (Theory of Mind)能力此前被认为是人类特有的能力(例如同理心等),在一项最新研究中发现 GPT 模型在 2022 年 1 月能够解决 70% 的 ToM 测试,而在 2022 年 11 月的版本中已经可以解决 93% 的 ToM 测试(Kosinski, 2023)。因此,本研究推测,针对主观任务,员工如果能够频繁地向人工智能寻求反馈,人工智能将拥有更多的数据进行训练和学习,则人工智能提供的反馈信息将更为准确和具体。相比于客观任务中,员工的人工智能反馈寻求行为对人工智能输出信息准确性和具体性的提升影响有限,员工的人工智能反馈寻求行为在主观任务中对人工智能输出信息准确性和具体性的提升影响更大。综上所述,提出以下调节效应命题:

命题 6a: 客观任务/主观任务这一任务特征会调节员工的人工智能反馈寻求行为与人工智能信息准确性及具体性之间的正向关系。具体而言,相对于客观任务,在主观任务中,员工的人工智能反馈寻求行为对人工智能信息准确性和具体性的正向影响更强。

结合上述中介效应和调节效应的推导过程,本研究提出以下两个被调节的中介效应:

命题 6b: 客观任务/主观任务这一任务特征会调节员工的人工智能反馈寻求行为通过人工智能反馈信息准确性而正向影响其绩效改善的中介效应。具体而言,相较于客观任务,在主观任务中,员工的人工智能反馈寻求行为通过人工智能反馈信息的准确性的中介作用而对其绩效改善的正向影响更强。

命题 6c: 客观任务/主观任务这一任务特征会调节员工的人工智能反馈寻求行为通过人工智能反馈信息具体性而正向影响其绩效改善的中介效应。具体而言,相较于客观任务,在主观任务中,员工的人工智能反馈寻求行为通过人工智能反馈信息的具体性的中介作用而对其绩效改善的正向影响更强。

4 理论构建

作为“第四次工业革命”的核心,人工智能技术有可能极大地改变整个劳动力结构、组织和工作的设计以及员工工作方式等方面。但是,组织行为的相关研究和理论发展却严重滞后于管理实

践。本文不仅能够通过将人工智能纳入反馈寻求目标从而拓展反馈寻求行为在新兴技术背景下的研究,同时还能够将员工的主动性纳入考量而丰富人工智能反馈在“人本主义”呼声中的研究。

理论上,第一,本文创新性地将人工智能纳入员工反馈寻求的目标,从而深化了人工智能背景下的反馈寻求行为研究。经过近40年的发展,反馈寻求行为的内涵扩展有限,特别是反馈寻求目标仍局限于直接上级或同事(张颖,杨付,2018; Ashford et al., 2016)。只有少数研究作了一定的拓展,例如将文件纳入反馈寻求目标中(Krasman, 2013)。随着生成式人工智能模型的快速发展,人类员工向人工智能寻求反馈的行为正越来越普遍地被应用于工作场所之中。但是,仅有少数研究在人工智能背景下探讨了由人工智能提供反馈带来的影响(冯向楠,詹婧,2019; Luo et al., 2021; Thuillard et al., 2022; Tong et al., 2021; Yam et al., 2022),缺乏对员工主动向人工智能寻求反馈的关注。本文将人工智能纳入反馈寻求目标中,为后续深入探讨人工智能背景下员工向人工智能这一对象寻求反馈现象、揭示其独特来源和作用效果奠定了基础,拓展了反馈寻求行为这一研究领域。

第二,结合人工智能系统特征,本文揭示了员工的人工智能反馈寻求行为的驱动机制。反馈寻求目标的特征是反馈寻求行为的重要前因变量(Ashford et al., 2016)。以往关于反馈寻求目标的研究,主要探讨了领导的行为与风格,例如真实型领导(张燕红,廖建桥,2015)、谦卑型领导(倪清,杜鹏程,2017)、授权型领导(王怀勇等,2022)、领导幽默(任巍,简浩,2020; 张正堂等,2018),对员工的反馈寻求行为的影响。整体而言,对于人工智能这一反馈寻求目标的特征如何影响员工的反馈寻求行为的研究尚处于空白阶段。本文从人工智能系统的透明性和拟人性特征切入,探讨人类员工对人工智能评估中关注的重要系统特征(蒋路远等,2022; Glikson & Woolley, 2020)对员工的人工智能反馈寻求的影响,揭示了人工智能反馈寻求行为的独特影响因素及作用机制,为后续人工智能背景下反馈寻求行为的研究提供了理论指导。具体而言,结合心智感知理论(Gray et al., 2007)和信任的相关文献(McAllister, 1995),本文探究了员工对人工智能的能动性感知和感受性感知这两种心智感知以及认知信任和情感信任这两

种信任类型在透明性和拟人性对人工智能反馈寻求行为影响过程中的链式中介作用。此外,基于资源保存理论,探讨了问题解决要求这一工作任务特征在其中起到的调节作用。通过揭示这些影响因素及其作用机制,本文加深了系统特征与人工智能反馈寻求行为之间关系的理解。

第三,结合人类员工同人工智能的交互特点,本文揭示了员工的人工智能反馈寻求行为的影响后果及其作用机制。绩效是反馈寻求行为研究中最关注的影响结果(Ashford et al., 2016)。在传统反馈寻求行为研究中,员工的反馈寻求和其绩效间的关系研究未取得一致的结论。有的研究发现员工的反馈寻求行为与其绩效正相关(张建平等,2020),有的研究发现这两者之间只有微弱的关系或者没有关系(Anseel et al., 2015)。相较于人类管理者,人工智能有其特定的优势,比如其具备高效处理海量数据的能力,可以快速地提供准确和具体的反馈内容(Luo et al., 2021; Tong et al., 2021)。因此,探索员工主动向人工智能寻求反馈对其绩效的影响,有助于进一步明晰反馈寻求与绩效之间的关系。基于反馈过程理论(Ilgen et al., 1979),并结合人工智能的相关文献,本文认为员工的人工智能反馈寻求行为会通过正向影响人工智能信息的准确性和具体性,从而对员工的绩效改善产生正向影响,并且客观任务/主观任务这一工作任务特性在该影响过程中起到调节作用。本文有助于拓展人工智能反馈寻求行为的影响后果研究并厘清其作用机制,以及丰富人工智能反馈寻求行为研究的理论边界。

本文也有一定的实践价值。本文揭示了人工智能反馈寻求行为的影响因素和作用结果,为管理者有效引导员工主动向人工智能寻求反馈,并获得相应的裨益、规避可能的风险,提供了一些有针对性的实践指导。具体而言,通过解析人工智能系统特征对员工人工智能反馈寻求行为的影响机制,为设计激发员工主动反馈寻求的人工智能系统提供理论依据。同时,探究员工的人工智能反馈寻求行为对绩效改善的影响及其机理,可以帮助管理者明晰在哪些工作任务中应该鼓励员工主动向人工智能寻求反馈。在此基础上,本文希望将人工智能反馈寻求行为作为管理者更好地理解新兴技术和员工如何交互影响并共融的窗口,启发实践者的思考。

参考文献

- 冯向楠, 詹婧. (2019). 人工智能时代互联网平台劳动过程研究——以平台外卖骑手为例. *社会发展研究*, 6(3), 61-83+243.
- 关键, 李文朴, 何国华, 张新安. (2025). 人工智能反馈: 文献述评与研究展望. *外国经济与管理*, 47(3), 83-100.
- 贾建锋, 葛羿京, 焦玉鑫. (2024). 有求必应, 还是相“机”而动? 员工反馈寻求行为对上级发展性反馈的影响机制研究. *管理工程学报*, 38(6), 1-12.
- 蒋路远, 曹李梅, 秦昕, 谭玲, 陈晨, 彭小斐. (2022). 人工智能决策的公平感知. *心理科学进展*, 30(5), 1078-1092.
- 林新奇, 徐洋洋. (2021). 未来工作自我清晰度与员工反馈寻求行为的关系: 变革型领导和工作投入的影响. *管理评论*, 33(7), 216-226.
- 刘善仕, 裴嘉良, 葛淳棉, 刘小浪, 谌一璠. (2022). 在线劳动平台算法管理: 理论探索与研究展望. *管理世界*, 38(2), 225-239+14-16.
- 罗文豪, 霍伟伟, 赵宜萱, 王震. (2022). 人工智能驱动的组织与人力资源管理变革: 实践洞察与研究方向. *中国人力资源开发*, 39(1), 4-16.
- 倪清, 杜鹏程. (2017). 基于链式中介效应的谦卑型领导与经理人反馈寻求行为关系研究. *管理学报*, 14(6), 823-831.
- 任巍, 简浩. (2020). 领导幽默对员工反馈寻求行为的影响研究: 促进焦点的中介作用. *广东财经大学学报*, 35(6), 32-41.
- 孙伟平. (2020). 人工智能与人的“新异化”. *中国社会科学*, (12), 119-137+202-203.
- 王怀勇, 岳思怡, 沈晓寻. (2022). 授权型领导对员工反馈寻求行为的影响机制: 领导差异化授权两阶段的调节作用. *心理与行为研究*, 20(1), 130-137.
- 王宁, 杨芮, 周密, 于玲玲. (2021). 差序格局视角下个体反馈寻求行为对创新绩效的影响研究. *软科学*, 35(8), 69-75.
- 王雁飞. (2005). 组织中的个体反馈寻求行为. *心理科学进展*, 13(5), 679-685.
- 王振源, 段永嘉, 王燕榕. (2018). 徒弟寻求反馈促进导师知识共享? 一个有中介的调节作用模型. *管理工程学报*, 32(4), 20-27.
- 王振源, 姚明辉. (2022). 工作场所人机协作对员工影响的研究述评. *外国经济与管理*, 44(9), 86-102.
- 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+13.
- 张建平, 秦传燕, 刘善仕. (2020). 寻求反馈能改善绩效吗? ——反馈寻求行为与个体绩效关系的元分析. *心理科学进展*, 28(4), 549-565.
- 张敏, 赵宜萱. (2022). 机器学习在人力资源管理领域中的应用研究. *中国人力资源开发*, 39(1), 71-83.
- 张燕红, 廖建桥. (2015). 团队真实型领导、新员工反馈寻求行为与社会化结果. *管理科学*, 28(2), 126-136.
- 张颖, 杨付. (2018). 反馈寻求行为的影响因素和本土化发展. *心理科学进展*, 26(6), 1121-1130.
- 张正堂, 丁明智, 刘宁. (2018). 领导奖励缺失对员工反馈寻求行为的影响机制. *经济管理*, 40(9), 107-122.
- Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, 28(4), 399-416.
- Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2015). How are we doing after 30 years? A metanalytic review of the antecedents and outcomes of feedback-seeking behavior. *Journal of Management*, 41(1), 318-348.
- Anseel, F., & Lievens, F. (2007). The relationship between uncertainty and desire for feedback: A test of competing hypotheses. *Journal of Applied Social Psychology*, 37(5), 1007-1040.
- Ashford, S. J. (1986). Feedback-seeking in individual adaptation: A resource perspective. *Academy of Management Journal*, 29(3), 465-487.
- Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance*, 32(3), 370-398.
- Ashford, S. J., De Stobbeleir, K., & Nujella, M. (2016). To seek or not to seek: Is that the only question? Recent developments in feedback-seeking literature. *Annual Review of Organizational Psychology and Organizational Behavior*, 3(1), 213-239.
- Bauer, K., & Gill, A. (2024). Mirror, mirror on the wall: Algorithmic assessments, transparency, and self-fulfilling prophecies. *Information Systems Research*, 35(1), 226-248.
- Bauer, K., Heigl, R., Hinz, O., & Kosfeld, M. (2024). Feedback loops in machine learning: A study on the interplay of continuous updating and human discrimination. *Journal of the Association for Information Systems*, 25(4), 804-866.
- Bogert, E., Schecter, A., & Watson, R. T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports*, 11(1), 8028.
- Brett, J. F., & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, 86(5), 930-942.
- Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370), 1530-1534.
- Butterfield, M. E., Hill, S. E., & Lord, C. G. (2012). Mangy mutt or furry friend? Anthropomorphism promotes animal welfare. *Journal of Experimental Social Psychology*, 48(4), 957-960.
- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, 56(5), 809-825.
- Choi, S., Kang, H., Kim, N., & Kim, J. (2025). How does artificial intelligence improve human decision-making? Evidence from the AI-powered Go program. *Strategic Management Journal*, 46(6), 1523-1554. <https://doi.org/10.1002/smj.3694>

- Coutifaris, C. G. V., & Grant, A. M. (2022). Taking your team behind the curtain: The effects of leader feedback-sharing and feedback-seeking on team psychological safety. *Organization Science*, 33(4), 1574–1598.
- Dahling, J. J., & Whitaker, B. G. (2016). When can feedback-seeking behavior result in a better performance rating? Investigating the moderating role of political skill. *Human Performance*, 29(2), 73–88.
- De Stobbeleir, K., Ashford, S., & Zhang, C. (2020). Shifting focus: Antecedents and outcomes of proactive feedback seeking from peers. *Human Relations*, 73(3), 303–325.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886.
- Förster, M., Broder, H. R., Fahr, M. C., Klier, M., & Fink, L. (2025). Tell me more, tell me more: The impact of explanations on learning from feedback provided by artificial intelligence. *European Journal of Information Systems*, 34(2), 323–345.
- Gajendran, R. S., Loewenstein, J., Choi, H., & Ozgen, S. (2022). Hidden costs of text-based electronic communication on complex reasoning tasks: Motivation maintenance and impaired downstream performance. *Organizational Behavior and Human Decision Processes*, 169, 104130.
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660.
- Gong, Y., Wang, M., Huang, J.-C., & Cheung, S. Y. (2017). Toward a goal orientation-based feedback-seeking typology: Implications for employee performance outcomes. *Journal of Management*, 43(4), 1234–1260.
- Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes*, 115(2), 253–267.
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619.
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125–130.
- Gupta, A. K., Govindarajan, V., & Malhotra, A. (1999). Feedback-seeking behavior within multinational corporations. *Strategic Management Journal*, 20(3), 205–222.
- Hall, K. R., Harrison, D. E., Ajjan, H., & Marshall, G. W. (2022). Understanding salesperson intention to use AI feedback and its influence on business-to-business sales outcomes. *Journal of Business & Industrial Marketing*, 37(9), 1787–1801.
- Han, M. C. (2021). The impact of anthropomorphism on consumers' purchase decision in chatbot commerce. *Journal of Internet Commerce*, 20(1), 46–65.
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5), 517–527.
- Heflick, N. A., Goldenberg, J. L., Cooper, D. P., & Puvia, E. (2011). From women to objects: Appearance focus, target gender, and perceptions of warmth, morality and competence. *Journal of Experimental Social Psychology*, 47(3), 572–581.
- Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, 44(3), 513–524.
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349–371.
- Inbar, Y., Cone, J., & Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology*, 99(2), 232–247.
- Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410.
- Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14(2), 119–140.
- Kosinski, M. (2023). Evaluating large language models in theory of mind tasks. *arXiv*. <https://arxiv.org/abs/2302.02083>
- Krasman, J. (2010). The feedback-seeking personality: Big five and feedback-seeking behavior. *Journal of Leadership and Organizational Studies*, 17(1), 18–32.
- Krasman, J. (2013). Putting feedback-seeking into “context”: Job characteristics and feedback - seeking behaviour. *Personnel Review*, 42(1), 50–66.
- Luo, X., Qin, M. S., Fang, Z., & Qu, Z. (2021). Artificial intelligence coaches for sales agents: Caveats and solutions. *Journal of Marketing*, 85(2), 14–32.
- Mahmud, H., Islam, A., Ahmed, S., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, 175, 121390.
- McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24–59.
- Morgeson, F. P., & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology*, 91(6), 1321–1339.
- Moss, S. E., Song, M., Hannah, S. T., Wang, Z., & Sumanth, J. J. (2020). The duty to improve oneself: How duty orientation mediates the relationship between ethical leadership and followers' feedback-seeking and feedback-avoiding behavior. *Journal of Business Ethics*, 165(4), 615–631.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*,

- 56(1), 81–103.
- Nurmohamed, S., & Schwingel-Sauer, Z. (2024). Beyond the first choice: The impact of being an alternate choice on social integration and feedback seeking. *Journal of Applied Psychology*, 109(4), 587–598.
- Ochmann, J., Michels, L., Tiefenbeck, V., Maier, C., & Laumer, S. (2024). Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. *Information Systems Journal*, 34(2), 384–414.
- Park, E., Jenkins, Q., & Jiang, X. (2008). Measuring trust of human operators in new generation rescue robots. *Proceedings of the JFPS International Symposium on Fluid Power*, 2008(7-2), 489–492.
- Pei, J., Wang, H., Peng, Q., & Liu, S. (2024). Saving face: Leveraging artificial intelligence-based negative feedback to enhance employee job performance. *Human Resource Management*, 63(5), 775–790.
- Pulakos, E. D., Mueller-Hanson, R., & Arad, S. (2019). The evolution of performance management: Searching for value. *Annual Review of Organizational Psychology and Organizational Behavior*, 6(1), 249–271.
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192–210.
- Saffarizadeh, K., Keil, M., Boodraj, M., & Alashoor, T. (2024). “My name is Alexa. What’s your name?” The impact of reciprocal self-disclosure on post-interaction trust in conversational agents. *Journal of the Association for Information Systems*, 25(3), 528–568.
- Schmitt, A., Zacher, H., & Frese, M. (2012). The buffering effect of selection, optimization, and compensation strategy use on the relationship between problem solving demands and occupational well-being: A daily diary study. *Journal of Occupational Health Psychology*, 17(2), 139–149.
- Seitz, L., Bekmeier-Feuerhahn, S., & Gohil, K. (2022). Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. *International Journal of Human-Computer Studies*, 165, 102848.
- Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, 115, 14–24.
- Sherf, E. N., Gajendran, R. S., & Posner, B. Z. (2021). Seeking and finding justice: Why and when managers’ feedback seeking enhances justice enactment. *Journal of Organizational Behavior*, 42(6), 741–766.
- Sherf, E. N., & Morrison, E. W. (2020). I do not need feedback! Or do I? Self-efficacy, perspective taking, and feedback seeking. *Journal of Applied Psychology*, 105(2), 146–165.
- Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277–284.
- Steelman, L. A., Levy, P. E., & Snell, A. F. (2004). The feedback environment scale: Construct definition, measurement, and validation. *Educational and Psychological Measurement*, 64(1), 165–184.
- Thuillard, S., Adams M., Jelmini, G., Schmutz, S., Sonderegger, A., & Sauer, J. (2022). When humans and computers induce social stress through negative feedback: Effects on performance and subjective state. *Computers in Human Behavior*, 133, 107270.
- Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9), 1600–1631.
- Turel, O. (2024). To learn or not learn from AI? Unpacking the effects of feedback valence on novel insights recall. *European Journal of Information Systems*. <https://doi.org/10.1080/0960085X.2024.2426473>
- van der Woerd, S., & Haselager, P. (2019). When robots appear to have a mind: The human perception of machine agency and responsibility. *New Ideas in Psychology*, 54, 93–100.
- Vandenberghe, C., Landry, G., Bentein, K., Anseel, F., Mignonac, K., & Roussel, P. (2021). A dynamic model of the effects of feedback-seeking behavior and organizational commitment on newcomer turnover. *Journal of Management*, 47(2), 519–544.
- Wambsganss, T., Janson, A., Söllner, M., Koedinger, K., & Leimeister, J. M. (2025). Improving students’ argumentation skills using dynamic machine-learning-based modeling. *Information Systems Research*, 36(1), 474–507.
- Wang, T., Wang, D., & Liu, Z. (2022). Feedback-seeking from team members increases employee creativity: The roles of thriving at work and mindfulness. *Asia Pacific Journal of Management*, 39(4), 1321–1340.
- Waytz, A., & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. *Emotion*, 14(2), 434–444.
- Wegner, D. M., & Gray, K. (2017). *The mind club: Who thinks, what feels, and why it matters*. Penguin.
- Whitaker, B., & Levy, P. (2012). Linking feedback quality and goal orientation to feedback seeking and job performance. *Human Performance*, 25(2), 159–178.
- Wilkinson, D., Alkan, Ö., Liao, Q. V., Mattetti, M., Vejsbjerg, I., Knijnenburg, B. P., & Daly, E. (2021). Why or why not? The effect of justification styles on chatbot recommendations. *ACM Transactions on Information Systems*, 39(4), 1–21.
- Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2021). Robots at work: People prefer—and forgive—service robots with perceived feelings. *Journal of Applied Psychology*, 106(10), 1557–1572.
- Yam, K. C., Goh, E.-Y., Fehr, R., Lee, R., Soh, H., & Gray, K. (2022). When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human.

Journal of Experimental Social Psychology, 102, 104360.
Zhu, Q., Martinescu, E., Beersma, B., & Wei, F. (2024). The
double-edged sword of negative supervisor gossip: When

and why negative supervisor gossip promotes versus
inhibits feedback seeking behavior among gossip targets.
Human Relations, 77(6), 864–886.

Driving mechanisms and impact effects of AI feedback-seeking behavior: A research proposal

SUN Fang¹, LI Shaolong², LONG Lirong³, LEI Xuan², ZENG Xianglin², HUANG Xiahong²

(¹ Business School, Hubei University of Economics, Wuhan 430205, China)

(² Economics and Management School, Wuhan University, Wuhan 430072, China)

(³ School of Management, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: In the current VUCA (volatility, uncertainty, complexity, ambiguity) era, employees must proactively seek feedback to facilitate personal development and enhance their workplace competitiveness. Artificial intelligence (AI) offers new opportunities for proactive feedback-seeking, with a survey by Oracle Corporation indicating that over 50% of employees prefer seeking feedback from AI systems. However, traditional research on feedback-seeking behavior has yet to incorporate AI as a feedback source, leaving the mechanisms and consequences of employee feedback-seeking from AI largely underexplored. Moreover, emerging studies on AI feedback primarily position employees as passive feedback recipients, paying limited attention to their proactive feedback-seeking behaviors. To bridge this gap, this research integrates insights from traditional feedback-seeking literature and emerging studies on AI feedback, expanding the concept of feedback-seeking to include AI as a legitimate source. Specifically, this research explores the driving mechanisms of AI system characteristics, such as transparency and anthropomorphism, on feedback-seeking from AI, and examines the impact of feedback-seeking from AI on employee performance improvement. By doing so, it contributes to the growing literature on emerging technologies and employee psychology and behavior, and offers actionable insights for management practices.

Keywords: feedback-seeking from AI, transparency, anthropomorphism, performance improvement