doi: 10.19920/j.cnki.jmsc.2026.06.011

人机智能交互下负面反馈的绩效激励机制研究①

刘善仕[1,2]，裴嘉良[1,2*]，王红丽[1,2]，葛淳棉[1,2]，姜军辉[1]
(1. 华南理工大学工商管理学院，广州 510641；2. 广东省互联网行为科学重点实验室，广州 510641)

摘要： 利用负面反馈激励员工绩效提升是绩效管理中一大难题. 随着数智技术的发展，一些科技企业率先利用前沿技术优化绩效管理流程，提高员工体验. 本研究基于动机归因视角，比较了人工智能与人类领导在提供负面反馈时差异化的绩效激励机制. 通过实证分析发现：与人类领导相比，人工智能提供负面反馈时员工绩效提升动机归因更强、伤害诱发动机归因更弱，进而促使其绩效水平更高；进一步扎根本土化情境提出领导风格的作用：在负面反馈通过绩效提升动机归因、伤害诱发动机归因间接影响员工绩效方面，人工智能比威权领导更具优势；但是，与仁慈领导相比，这种差异化间接效应被削弱. 研究结果揭示了人机智能交互下负面反馈的绩效激励效应，拓展了人工智能与负面反馈的研究情境、视角与思路，也为中国企业数智化绩效管理实践创新提供启示.

关键词： 人工智能；负面反馈；工作绩效；动机归因；威权领导；仁慈领导
中图分类号： C93 **文献标识码：** A **文章编号：** 1007－9807(2026)06－0156－16

0 引 言

工作场所负面反馈向员工传达了自身不足或绩效不佳的信息并指出需要改进的方面及策略，在激励员工及时纠偏、认识差距并努力发展知识技能以提升绩效方面至关重要[1,2]. 然而，领导作为提供负面反馈的重要信息源，常因担心破坏和谐的人际关系或导致员工消极应对陷入“进退两难”的窘境[3]. 在一项实地调查中，44% 的领导者认为提供负面反馈有压力[4]. 时至今日，如何解决领导负面反馈的不足，满足负面反馈的激励目的仍面临诸多挑战.

得益于近年来数智化之风席卷各行各业，以人工智能(artificial intelligence，AI)为代表的数智技术逐渐嵌入并变革传统工作流程，辅助甚至替代人类承担特定工作职责[5]. 领先的科技巨头(例如，IBM、Amazon 和 Uber)开始部署 AI 反馈系统，由大数据、机器学习算法驱动的 AI 针对员工过去绩效成绩、当前绩效表现及未来绩效发展追踪、评估和预测，提供个性化反馈[6,7]. 已有研究发现，AI 相较于人类具有更强的计算、分析和预测的能力，并且数据驱动的 AI 能输出更为客观、准确和一致的决策建议，其功能性价值凸显[6,8,9]. 但也有研究从心理层面指出人类在与 AI 互动中可能会面临角色模糊、身份威胁和不安全感等问题[7,10,11]. 由此可见，尽管 AI 具有替代领导提供负面反馈的技术能力以及克服领导提供负面反馈中的情绪化表现、主观偏见以及容易导致员工消极反应等问题的技术潜力[12]，但鉴于情境新颖性，现有研究对负面反馈情境下 AI 如何塑造员工反馈反应的研究尚不充分，尤其是对个体归因的探讨不足.

为弥补现有研究空白，本研究借鉴动机归因视角[13－15]. 该理论视角强调个体并不总是对行为

① 收稿日期：2022－11－22；修订日期：2024－06－15.
基金项目：国家自然科学基金资助项目(72132001；72272054；72272055；72372045).
通讯作者：裴嘉良(1997—)，男，河南安阳人，博士，助理教授，硕士生导师. Email：bmpeijl@scut.edu.cn

本身进行归因(如,员工分析自己为何收到差评),还会主动观察、分析和解读行为发出者的动机(如,员工分析领导为何给自己差评). 已有研究表明,个体在经历负面行为或事件中更倾向推断行为发出者的动机,并做出反应以适应环境[16,17]. 显然,人类领导作为具有能动性的主体会有意图地行事[18],员工收到人类领导的负面反馈后会对其动机进行归因. 例如,研究发现[15],核心自我评价低的员工会将领导负面反馈的动机归因为自我服务导向,不利于激发其学习动机并提高学习绩效.

然而,极少研究提出个体是否会对 AI 某些特定行为的动机进行归因,因为过去研究认为机器作为非人类实体的能动性很弱,并不存在意图或动机[19-21]. 但是,本研究强调个体主观上可能认为 AI 存在特定动机,因为 AI 的底层逻辑是机器学习算法,算法会根据人类社会产生的数据进行训练学习[22,23],进而表现出类人的行为特征(如,机器学习算法报告非裔美国人的犯罪系数高于白人)[24]. 同时,AI 与传统机器不同,具有自我学习、进化和推理的能力[25],AI 背后的"算法黑箱"可能会进行思考并开发出自己的动机和目标[23,26]. 此外,结合计算机社会行动者理论的观点,人们倾向于将计算机赋予类似人类的社会属性和角色,从而在心理上与其建立互动关系[27]. 因此,在人机交互中个体可能会对 AI 表现出某些看似具有智能和能动性的行为进行动机性判断. 进一步地,鉴于人类领导与 AI 的行为特征并不相同,这些差异为员工对二者的负面反馈动机做出推断提供了信息线索. 综上,动机归因视角表明,员工可能会对 AI 和人类领导的负面反馈动机做出不同的归因,继而对其后续的行为表现产生差异化影响.

本研究拟引用两类与负面反馈紧密相关的动机归因: 绩效提升动机归因与伤害诱发动机归因[16,17]. 首先,向员工提供负面反馈的本意是控制偏差、指明错误并持续改进[1,2],倘若在负面反馈中,员工将反馈提供者的动机归因为想要确保其提升绩效时,会激励员工努力追赶差距,对绩效提升有利. 相反,由于负面反馈具有批评色彩,倘若员工将反馈提供者的动机归因为想要对其带来不道德的伤害时,会导致员工意志消沉,对绩效提升不利[17]. 因此,拟将这两类动机归因作为揭示 AI(vs. 人类领导)负面反馈影响员工绩效的机制. 此外,领导风格彰显了领导行为特征,可能会影响员工对其动机的归因判断[28]. 在区别 AI 与人类领导在负面反馈激励员工工作绩效方面的差异后,拟进一步比较 AI 与具有不同领导风格(威权型 vs. 仁慈型)的人类领导相比有何差异.

综上,拟通过 2 个实验研究检验所提理论假设. 潜在的理论贡献包括: 第一,拓展负面反馈在数智化情境下的研究; 第二,揭示 AI 执行特定类型的反馈任务(即负面反馈)对员工工作绩效的影响; 第三,借鉴并拓展动机归因视角,挖掘 AI(vs. 人类领导)负面反馈影响员工工作绩效的作用"黑箱"; 第四,扎根本土化情境,进一步发现领导风格在影响 AI(vs. 人类领导)负面反馈与员工工作绩效关系的作用.

1　理论与假设

1.1　负面反馈、AI 与动机归因

领导负面反馈有很多好处,一是提供了提高领导地位和实现目标的机会,并允许其展示能力、发挥权力和实施控制[29]. 二是负面反馈作为干预员工绩效的手段,有助于激发下属工作动机并激励其做出改变和努力[2,30]. 尽管好处很多,但领导时常不愿意提供负面反馈[3],因为负面反馈是一种情感事件,会引发下属内疚、沮丧、愤怒等消极情感[12],也可能会导致人际冲突并挫伤员工士气[31]. Kluger 和 DeNisi 通过元分析发现[32],超 1/3的负面反馈实际导致了员工绩效下降. 同时,提供负面反馈对领导者自身是一种挑战,不仅会分散其注意力,与下属不愉快的互动还会使自身感到不适[3].

一种平衡负面反馈收益和成本的新兴策略是部署 AI 替代人类领导提供负面反馈. 具体地,与传统需要人工干预的被动机器不同,AI 具有自主推理、深度学习、个性化预测等能力,不仅可以模仿人类思维方式自行处理工作任务,工作流程也可以由它引导[22,25]. 此外,AI 在数据挖掘、集成和分析方面突破了人类生理局限,可以在迭代中

不断优化，超越了人类在处理复杂问题时的认知，优化了决策环境[33,34]. 这些特征和优势意味着 AI 初步具备了承担特定工作角色的能力. 其中，将 AI 嵌入绩效反馈流程是前沿应用的一个新兴领域，即 AI 根据海量的数据信息和具有逻辑体系的机器学习算法追踪、评估员工行为和绩效，并生成、提供与员工生产力相关的反馈建议[8]. 可见，AI 在承担领导“信息角色”方面已初见成效[8]，在某种程度上，它能够执行负面反馈的任务.

虽然 AI 在技术上具有诸多优势(如，准确性、一致性和可供性等)[6,8,9,33]，但对于员工能否从 AI 提供的负面反馈切实受益知之甚少. 甚至部分谨慎观点认为，在工作场所使用 AI 可能导致员工产生不安全感、角色模糊以及信息过载等问题[6,11,35,36]. 因此，亟需在特定的工作场景下识别并厘清部署 AI 的效应. 动机归因视角为解释上述问题提供了理论基础[13]. 即由于人类领导与人工智能的行动特征具有差异，员工与领导的人际互动过程和员工与 AI 的人机交互过程会产生不同的信息线索，影响其对互动对象的动机做出不同的归因判断[15]，随后影响其行为表现. 基于上述理论背景，本研究将在下节提出具体的理论假设.

1.2 AI(vs. 人类领导)负面反馈与绩效提升动机归因

动机归因的核心观点是，人们会对周围个体行为背后的原因或目的进行因果推断，并做出反应[13,14]. 也就是说，当员工收到负面反馈时，不仅可能对自己为何收到负面反馈进行归因(如，归因于自己不努力)，还可能根据与提供反馈主体相关的信息线索对其背后动机进行归因. 其中，绩效提升动机归因强调员工将反馈主体提供负面反馈的动机归因为想要确保或提升其绩效[16,17].

本研究认为 AI(vs. 人类领导)负面反馈能激发员工更高的绩效提升动机归因. 第一，AI 的决策逻辑不同，更客观. 人类领导擅长基于固有知识、经验等做出主观判断和启发式决策，但 AI 则是基于数据和算法逻辑做出客观判断和决策[33,37,38]. 当收到负面反馈时，面对决策更客观的 AI 会强化这些反馈信息与绩效相关、与偏见无关的信念；第二，AI 的决策标准不同，更一致. 人类领导通常难以保持内心衡量尺度的一致，决策很容易受到外界因素的干扰(如，关系亲疏)[39]. 而 AI 基于既定的算法规则进行预测，根据统一的绩效标准进行评估[8]. 当收到负面反馈时，面对标准更一致的 AI 会强化这些负面反馈信息更公平、有标尺的信念；第三，AI 的分析能力不同，更准确. 人类领导通常具有认知局限性，在直指绩效不足问题方面不及 AI，因为 AI 能利用比领导记忆中更大的训练数据集(包括各种成功、失败案例)，输出针对性极强的反馈信息及建议[7,40]. 当收到负面反馈时，面对内容更准确、更可信的 AI 会强化这些负面反馈更符合自身真实情况的信念；最后，AI 的建议属性不同，更可操作. 人类领导在短时间内难以快速响应反馈需求并提供可操的应对策略，AI 响应迅速、覆盖全面、结果清晰，能提供更优、具体的解决方案[6,8]. 当收到负面反馈时，员工能迅速定位到自身不足以及学习到如何改进，更便于其对过往、现在和未来绩效有清晰的理解和认识.

总之，在负面反馈事件中，鉴于 AI 与人类领导在反馈客观性、一致性、准确性和可操作性等方面的差异，在主观上会强化员工将 AI 提供的负面反馈面向工作任务和绩效并且 AI 想要激励其纠正当前不足、取得更高表现的绩效提升动机信念.

H1(a) 与人类领导相比，AI 提供负面反馈能激发员工更高的绩效提升动机归因.

1.3 绩效提升动机归因的中介效应

做出不同动机归因的个体通常会采取不同的行为反应以适应周遭环境[13,15]. 鉴于绩效提升可能带来晋升、奖励等好处以及更高的工作满意度和成就感[41]，绩效提升动机归因水平高的员工通常对未来取得绩优表现拥有较大期望[17]，会激励其积极主动地花费时间、精力学习新的知识技能，并将这些获取的经验和知识在工作中应用，提高应对困难和解决问题的能力，继而最终转化为实际的绩效表现.

H1(b) 绩效提升动机归因与员工工作绩效正相关.

结合 H1(a)、H1(b)，进一步提出绩效提升动机归因的中介效应. 与面对面地收到来自人类领导的负面反馈相比，员工通过人机交互获取来自

AI 的负面反馈时,虽然反馈效价是负面的,但鉴于 AI 的决策逻辑更客观、决策标准更一致、反馈结果更准确、反馈建议更可操作,从主观上更相信这些反馈信息与绩效本身是高度相关的、与其它非绩效因素相关性并不强,并且更倾向于将 AI 提供负面反馈的动机归因为想要其纠正错误、提升绩效以保持竞争力. 进一步地,在未来高绩效、高收益的期望激励下,会努力鞭策自己专注任务和目标、通过学习不断精进自己以取得更高工作绩效.

H1(c) 与人类领导相比,AI 提供负面反馈能通过激发员工更高的绩效提升动机归因,进而促使其工作绩效水平更高.

1.4 AI(vs. 人类领导)负面反馈与伤害诱发动机归因

以往研究表明,不当的负面反馈会伤害员工自尊、自我概念、带来人际威胁或压力[30,42,43],甚至迫使员工感知到经历辱虐[16]. 这些"阴暗面"不利于员工从反馈中汲取经验以改进绩效. 因此,与负面反馈相关的伤害诱发动机归因强调了反馈主体具有不道德、对员工不利的行为意图倾向[17].

本研究认为 AI(vs. 人类领导)负面反馈能导致员工更低的伤害诱发动机归因. 第一,如前所述,AI 决策逻辑是基于客观的数据和算法的,很少涉及个人情感和偏见[22,37]. 而人类领导的启发式决策过程不可避免受到主观因素的干扰,容易产生刻板印象、歧视等问题[26]. 因此,由能够克服人类决策固有偏见的 AI 提供负面反馈更不容易触发员工的心理防御机制; 第二,AI 向员工提供负面反馈是一个人机交互的过程,根据技术可供性的观点[44],技术对象为特定用户提供面向目标的、可操作的可能性,换言之,人机之间的智能交互有助于员工清楚地了解绩效反馈系统面向绩效提升的功能. 而人类领导当面指出员工不足更容易将员工注意力焦点转移至非绩效因素[2,15,30],认为领导在批评其个人能力或品格,从而导致员工感受到被攻击. 因此,在负面反馈中,人际互动比人机交互更容易让员工感到颜面尽失而无法正视负面反馈的作用; 第三,人类领导与员工之间存在不对称的等级权力差距,员工依赖领导所掌握的工作资源(如,任务机会),继而对自上而下的批评更敏感和警惕[30,45]. 而 AI 与员工之间并未形成层级关系,员工不太可能认为 AI 批评他们是为了通过打压下属而巩固其地位和权力; 第四,负面反馈对员工而言本身是一种情感事件[12],人类领导在反馈过程难以像 AI 一样不附加情绪,更有可能导致员工情感波动[46]. 处于消极体验状态下的员工倾向于对领导动机做出不道德的负面评价[15].

总之,在负面反馈事件中,鉴于 AI 与人类领导在反馈客观公正性、技术可供性、等级权力差距、情感表达等方面的差异,主观上会弱化 AI 提供负面反馈的动机是想要对员工带来不道德伤害的信念.

H2(a) 与人类领导相比,人工智能提供负面反馈会导致员工更低的伤害诱发动机归因.

1.5 伤害诱发动机归因的中介效应

与绩效提升动机归因不同,那些将反馈主体提供负面反馈的行为解释为不道德、有害的员工通常不愿投入更多工作资源予以回应,因为其倾向于认为这违背了社会交换的互惠原则[15,17]. 同时,他们会采取行动抵制这种不道德、有害的行为,将注意力由原有的工作任务、目标进展中转移到与绩效活动无关的领域上[47],这是低水平工作动机的表现,会导致工作绩效的下降. 最后,在互动过程中做出不利动机归因的个体通常会产生负面情绪(如,委屈、愤怒)[12],进而也会对后续绩效提升不利.

H2(b) 伤害诱发动机归因与员工工作绩效负相关.

结合 H2(a)、H2(b),进一步提出伤害诱发动机归因的中介效应. 与面对面地收到来自人类领导的负面反馈相比,员工通过人机交互获得来自 AI 的负面反馈时,虽然反馈效价是负面的,但鉴于 AI 反馈主观偏见较弱、交互指向性清晰、等级权力无差异、弱情感连接,从主观上更相信 AI 提供负面反馈的不道德性、攻击性和不利导向更弱,并且更不容易将其动机归因为有害的. 进一步地,由于员工更低程度受到消极动机归因的干扰,他们能从负面反馈的信息中获益,专注生产活动,不断提升工作绩效.

H2(c) 与人类领导相比，AI 提供负面反馈会通过导致员工更低的伤害诱发动机归因，进而促使其工作绩效水平更高.

1.6 立威还是施恩？领导风格的作用

领导力归因研究表明，领导风格为员工识别领导特定行为意图或动机并进行归因提供了重要信息线索[28]. 基于此，在提供负面反馈以影响员工绩效方面，AI 与具有不同领导风格的人类领导相比是否存在差异？倘若是，如何发生？

本研究引入家长式领导理论中“威权领导”与“仁慈领导”两种极具本土化特色的领导风格. 家长式领导经典的二元理论认为家长式领导主要包括立威与施恩两方面[48,49]，即威权领导与仁慈领导. Chan 等[50]认为威权和仁慈是家长式领导的两个主要维度. 其中，威权是指领导者强调个人绝对权威，并对下属进行严格控制，涵盖“专权作风”、“贬损下属能力”等立威行为；与之相对，仁慈是指领导者强调宽容体谅，并对下属做出个性化关怀，涵盖“维护面子”、“急难相助”等施恩行为[48,51].

结合威权领导风格特征，首先比较 AI(vs. 威权领导) 负面反馈对员工反馈反应的影响. 威权领导风格源于儒家道德标准中高权力距离下上尊下卑的关系，主张权威不容挑战，严密控制下属绩效表现，要求下属绝对服从[52]. 当下属表现低绩效时，通常会受到威权领导的严厉斥责和贬低贡献[53]，而非耐心安抚员工并悉心指导如何改进. 因此，在工作场所中，由威权领导向员工提供负面反馈，员工倾向于认为以维系个人权威和个人意志为中心的领导提供负面反馈的根本目的或动机是确立权威和地位并非助其绩效提升. 同时，威权领导日常表现出打压员工、专权专治、漠视建议的行为[54]，还会导致收到负面反馈的员工认为领导仍是延续贬低、干涉、示威、操控等不道德、具有攻击性的行事作风. 与之相对，员工表现低绩效时，AI 更多地表现为“对事不对人”，基于客观数据和算法分析员工表现不足的方面、内容和原因，并以人机交互的方式并且不附加负面情感地向员工提出客观、一致、准确、可操作的反馈建议. 因此，与威权领导相比，AI 负面反馈能激发员工更高的绩效提升动机归因、导致更低的伤害诱发动机归因. 进一步结合 H1(b)、H2(b)，AI(vs. 威权领导) 通过绩效提升动机归因、伤害诱发动机归因激励员工更高水平的工作绩效.

结合仁慈领导风格特征，接着比较 AI(vs. 仁慈领导) 负面反馈对员工反馈反应的影响. 与立威不同，领导还会表现出像父亲般的“仁慈”. 仁慈领导风格源于儒家对理想社会人际关系的设想[51]，通常表现出对下属工作和个人福祉长久关怀，包括工作领域的宽容体谅以及生活上的个别照顾[55]. 当下属表现低绩效时，仁慈领导通常会尽力维护下属面子，安抚下属情绪并向下属伸出援手解决难题[53]. 因此，在工作场所中，由仁慈领导提供负面反馈，员工并不会感到难堪和压力，而是从领导反馈中获得个性化的指导和解困，进而倾向于认为关心下属个人需求与发展的领导提供负面反馈的根本目的和动机是想要员工切实地发现并改进自身不足以助其取得更高的绩效表现. 同时，纵使批评下属，仁慈领导日常都会避免羞辱并预留余地，采取更为恰当、情绪稳定、耐心和及时的方式帮助员工分析和解决问题[56]，导致员工收到来自仁慈领导的负面反馈时不容易感受到被冒犯、将注意力转移到不道德的人际伤害上来. 鉴于上述提到的 AI 负面反馈的特征，本研究认为，与仁慈领导相比，AI 负面反馈影响员工绩效提升动机归因、伤害诱发动机归因的差异化效应将被削弱. 进一步结合 H1(b)、H2(b)，AI(vs. 仁慈领导) 通过绩效提升动机归因、伤害诱发动机归因影响员工工作绩效的差异化间接效应也将被削弱.

H3(a) 与威权领导相比，AI 提供负面反馈能激发员工更高的绩效提升动机归因；但与仁慈领导相比，AI 提供负面反馈影响员工绩效提升动机归因的差异化效应将被削弱.

H3(b) 与威权领导相比，AI 提供负面反馈会导致员工更低的伤害诱发动机归因；但与仁慈领导相比，AI 提供负面反馈影响员工伤害诱发动机归因的差异化效应将被削弱.

H4(a) 与威权领导相比，AI 提供负面反馈能通过激发员工更高的绩效提升动机归因，进而促使其工作绩效水平更高；但与仁慈领导相比，AI 提供负面反馈通过绩效提升动机归因影响员工工

作绩效的差异化间接效应将被削弱.

H4(b) 与威权领导相比,AI 提供负面反馈会通过导致员工更低的伤害诱发动机归因,进而促使其工作绩效水平更高; 但与仁慈领导相比,AI 提供负面反馈通过伤害诱发动机归因影响员工工作绩效的差异化间接效应将被削弱.

2 研究 1: 准实验

2.1 实验设计与程序

研究 1 选择一家华南地区的互联网公司进行准实验. 该公司内部考核标准主要分为多档,其中 3 分 ~ 3.25 分为不合格与需要提高群体,占比约 10%,公司规定对不符合预期的员工进行提供反馈和辅导. 该公司数字化程度较高,通过内部开发的通讯和办公平台协同的 APP 进行数字化管理. 2021 年中旬,该款 APP 的团队与国内某家技术平台合作,开放端口接入该平台研发的智能绩效管理工具,将流程数据与机器学习算法结合或者将文本等资料转译为算法可读数据进行分析并生成反馈信息. 公司过去以部门领导绩效面谈的方式向员工提供反馈,现阶段正计划引入该款智能工具. 在试点期间,负责人期望小范围应用该工具以考察员工实际体验,并决定在严格保密和遵循伦理规范的前提下共同推进实验进行. 研究 1 在 HR 部门发布季度绩效考核的自然环境下开展实验.

2021 年 12 月研究团队创建了实验条件: 首先,将时间确定为 HR 部门在第四季度(Q4)结束后启动的绩效考核工作期间; 其次,为提高实验操作的可行性和便捷性,将实验被试的选择范围设定为 4 个部门; 再次,根据实验目的,将实验对象确定为未符合绩效预期的员工. 需要强调的是,实验处理是在原部门进行的,样本不是随机分配的,遵循准随机抽样原则[30]. HR 部门考核工作结束后,再次访问了该企业,成功邀请到 180 名员工志愿参加本次实验活动. 根据单因素两水平的组间设计原则,创建了两个实验条件: 准随机选择 2 个部门由部门领导对被试提供反馈,剩余 2 个部门的员工从手机 APP 上的 AI 虚拟助手上收到反馈. 结合研究情境以及对该公司负责人的访谈,AI 和领导负面反馈内容以员工的述职报告文本为依据,结果包括员工的绩效考核分数档次,并告知其存在的不足之处和改进方向. 19 名员工不符合实验标准(如,考核评分不符合要求、个人原因中途退出等),共获取 161 名有效样本的数据: AI 组(80 人)和人类领导组(81 人). 随后要求两组被试填写人口统计学信息,负面反馈源感知、绩效提升动机归因、伤害诱发动机归因量表等. 在次年一季度(Q1)绩效考核工作后,HR 部门提供了新的一季度绩效考核评分.

2.2 变量测量

对于外文量表,遵循严格的翻译和回译程序. 在研究小组反复比较、沟通直至没有差异后,结合研究情境确定最终实验所使用的中文版本量表. 除有特别说明,所有量表均由李克特 7 点法评分.

负面反馈源. 借鉴 Tong 等[6]的做法,将 AI 组赋值 0,人类领导组赋值 1.

反馈源操纵. 借鉴宋晓兵和何夏楠[57]的做法,要求实验被试填写"1)根据上述情景描述,是谁进行的绩效反馈? 2)谁将你的绩效考核结果提供给你?"信度系数为 0.88.

绩效提升动机归因. 使用 Liu 等[17]对绩效提升动机归因的测量条目. 代表题项为"想要激励我实现更高的绩效目标."信度系数为 0.90.

伤害诱发动机归因. 使用 Liu 等[17]对伤害诱发动机归因的测量条目. 代表题项为"想要让我觉得自己很差劲."信度系数为 0.88.

工作绩效. 获取了被试接受实验操纵后下一季度的客观绩效分数,以衡量被试接受不同来源的负面反馈后下一季度的工作绩效.

控制变量. 首先,对人口统计学变量进行控制. 同时,实验被试来自 4 个不同部门,通过设置哑变量对部门进行控制; 其次,反馈会在情感上影响个体,负面反馈会威胁员工自尊,导致员工产生消极情绪,这干扰了其对负面反馈的判断[12]. 也就是说,一旦员工因负面反馈而产生消极情绪,不论反馈源是 AI 还是人类领导,他们均可能做出不利归因,因而需要排除消极情绪的影响. 此外,员工个体特征可能会影响其做出差异化的归因. 例如,核心自我评价高的员工更倾向于将负面反馈视作改进绩效的机会,更有可能做出外部归

因[15]. 因此,排除个体差异的影响. 最后,排除个体对反馈结果感知信任的影响,因为信任会影响其对负面反馈结果的价值判断[6]. 消极情绪、核心自我评价、感知信任分别参考 Liu 等[57]、Judge 等[58]、Tong 等[6]开发的量表. 信度系数分别为 0.78,0.91 和 0.76.

2.3 实验结果

2.3.1 验证性因子分析(confirmatory factor analysis, CFA)

通过 CFA 检验变量之间的区分效度,对使用李克特量表测量的变量进行 CFA. 结果显示,包括绩效提升动机归因、伤害诱发动机归因、消极情绪、核心自我评价和感知信任的五因子模型各拟合指数基本满足要求($\chi^2 = 491.98$, df = 367, $TLI = 0.93$, $CFI = 0.94$, $RMSEA = 0.05$, $SRMR = 0.05$),该模型显著优于其它竞争模型,说明测量变量具有较好的区分效度.

2.3.2 描述性统计分析

表 1 显示了主要变量的均值、标准差和相关系数. 所得结果为后续假设检验提供初步支持.

表 1 描述性统计分析(研究 1)

Table 1 Descriptive statistics analysis (Study 1)

变量	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. 性别	1.52	0.50	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2. 年龄	32.59	3.89	0.06	—	—	—	—	—	—	—	—	—	—	—	—	—
3. 学历	3.05	0.97	0.05	−0.11	—	—	—	—	—	—	—	—	—	—	—	—
4. 任职期限	3.72	1.64	0.07	0.40**	−0.10	—	—	—	—	—	—	—	—	—	—	—
5. 部门 2	0.22	0.41	0.02	0.04	−0.01	−0.07	—	—	—	—	—	—	—	—	—	—
6. 部门 3	0.20	0.40	0.10	−0.16*	0.14	−0.09	−0.26**	—	—	—	—	—	—	—	—	—
7. 部门 4	0.27	0.44	−0.01	0.00	−0.10	0.04	−0.32**	−0.30**	—	—	—	—	—	—	—	—
8. 消极情绪	4.12	0.97	−0.01	−0.02	−0.09	0.02	−0.11	−0.03	0.07	(0.78)	—	—	—	—	—	—
9. 核心自我评价	4.35	0.94	0.14	−0.10	0.09	0.03	0.05	0.03	0.12	0.07	(0.91)	—	—	—	—	—
10. 感知信任	4.30	1.11	−0.03	0.08	0.08	−0.06	0.07	0.07	−0.10	−0.03	−0.14	(0.76)	—	—	—	—
11. 负面反馈源(AI = 0, 领导 = 1)	0.50	0.50	0.07	0.12	−0.01	0.07	0.13	−0.10	−0.02	0.15	0.13	−0.03	—	—	—	—
12. 绩效提升动机归因	4.19	1.16	−0.14	−0.10	0.07	0.02	−0.05	0.10	0.02	0.03	−0.03	0.05	−0.26**	(0.90)	—	—
13. 伤害诱发动机归因	4.32	1.14	0.02	0.05	−0.02	−0.11	0.12	−0.05	−0.08	0.05	−0.01	−0.05	0.23**	−0.18*	(0.88)	—
14. 工作绩效	3.51	0.29	0.07	−0.02	0.02	0.04	−0.09	0.15	0.02	−0.21**	0.21**	0.02	−0.35**	0.31**	−0.30**	—

注: * 表示 $p < 0.05$, ** 表示 $p < 0.01$,双侧; 部门 1 为参照组.

2.3.3 操纵检验

借鉴彭坚等[59]的做法,使用独立样本 t 检验的方法验证负面反馈源(AI vs. 人类领导)是否操纵成功. 结果显示,AI 组评分显著低于人类领导组: AI 组均值 $M = 2.00$, $SD = 0.62$, 人类领导组均值 $M = 4.60$, $SD = 1.21$, $t(159) = -17.17$, $p < 0.001$. 结果说明研究 1 实验操纵成功.

2.3.4 假设检验

为检验 H1(a)、H2(a),使用独立样本 t 检验的方法(见图 1、2),结果表明,AI 组中绩效提升动机归因显著高于人类领导组: AI 组均值 $M = 4.58$, $SD = 1.24$, 人类领导组均值 $M = 3.80$, $SD = 0.93$, $t(159) = 4.54$, $p < 0.001$; AI 组中伤害诱发动机归因显著低于人类领导组: AI 组均值 $M = 3.91$, $SD = 0.99$, 人类领导组均值 $M = 4.73$, $SD = 1.13$, $t(159) = -4.89$, $p < 0.001$. 此外,加入控制变量后的回归结果也显示(见表 2),负面反馈源对绩效提升动机归因有显著负向影响(M2 $\beta = -0.34$, $p < 0.001$)、对伤

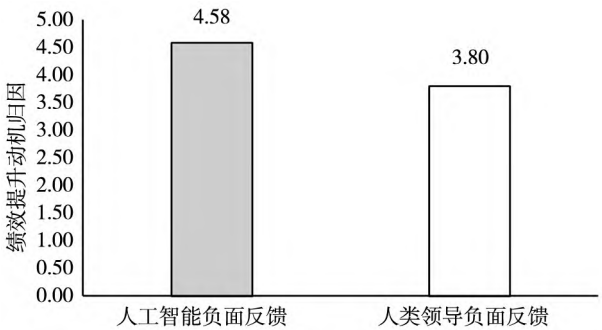

图 1 不同负面反馈源下的绩效提升动机归因(研究 1)

Fig. 1 Attributed performance promotion motives under various negative feedback sources (Study 1)

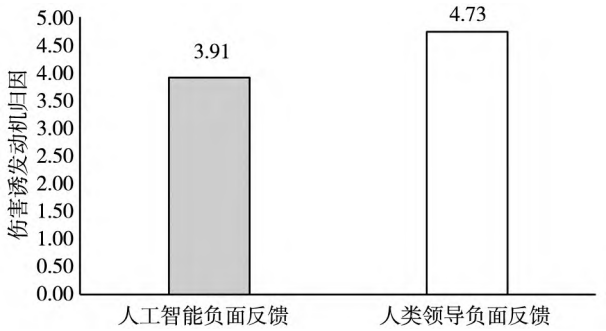

图 2　不同负面反馈源下的伤害诱发动机归因 (研究 1)

Fig. 2 Attributed injury initiation motives under various negative feedback sources (Study 1)

害诱发动机归因有显著正向影响(M4 $\beta = 0.36$, $p < 0.001$) . 结果表明,与人类领导相比,AI 负面反馈能激发员工更高的绩效提升动机归因、导致更低的伤害诱发动机归因,H1(a) 、H2(a) 得到支持.

接下来,在检验绩效提升动机归因、伤害诱发动机归因的中介效应前,通过回归分析发现,绩效提升动机归因与员工工作绩效显著正相关(M7 $\beta = 0.24$, $p < 0.01$) ,伤害诱发动机归因与员工工作绩效显著负相关(M7 $\beta = -0.17$, $p < 0.01$) ,H1(b) 、H2(b) 得到支持. 进一步,使用 Bootstrap 法进行中介效应检验,结果显示,与人类领导相比,AI 负面反馈通过绩效提升动机归因促使员工工作绩效水平更高的间接效应为 -0.16,95% 置信区间为 [-0.30, -0.05] ,不包括 0,说明 H1(c) 得到支持,绩效提升动机归因中介效应显著; 此外,与人类领导负面反馈相比,AI 负面反馈通过伤害诱发动机归因促使员工工作绩效水平更高的间接效应为 -0.15,95% 置信区间为 [-0.28, -0.03] ,不包括 0,说明 H2(c) 得到支持,伤害诱发动机归因中介效应显著.

表 2　层次回归分析(研究 1)

Table 2 Hierarchical regression analysis (Study 1)

变量	绩效提升动机归因		伤害诱发动机归因		工作绩效		
模型	M1	M2	M3	M4	M5	M6	M7
性别	-0.15	-0.13	0.03	0.01	0.02	0.04	0.07
年龄	-0.11	-0.07	0.10	0.06	0.00	0.03	0.06
学历	0.07	0.07	-0.01	-0.02	-0.04	-0.04	-0.05
任职期限	0.10	0.11	-0.15	-0.16	0.04	0.04	-0.01
部门 2	0.03	0.07	0.08	0.04	-0.09	-0.05	-0.06
部门 3	0.13	0.11	-0.04	-0.02	0.12	0.10	0.07
部门 4	0.08	0.07	-0.07	-0.06	0.01	0.01	-0.02
消极情绪	0.04	0.09	0.07	0.01	-0.24**	-0.19*	-0.21**
核心自我评价	-0.03	0.00	0.00	-0.04	0.24**	0.28***	0.27***
感知信任	0.05	0.05	-0.07	-0.06	0.06	0.05	0.03
负面反馈源(AI = 0, 领导 = 1)	—	-0.34***	—	0.36***	—	-0.33***	-0.20**
绩效提升动机归因	—	—	—	—	—	—	0.24**
伤害诱发动机归因	—	—	—	—	—	—	-0.17**
F	0.88	2.60**	0.73	2.67**	2.25	4.12***	5.01***
R^2	0.06	0.16	0.05	0.17	0.13	0.23	0.31
ΔR^2	0.06	0.11***	0.05	0.12***	0.13	0.10	0.02

注: * 表示 $p < 0.05$, ** 表示 $p < 0.01$, *** 表示 $p < 0.001$, 双侧; 部门 1 为参照组.

3　研究 2: 实验室实验

3.1　实验设计与程序

研究 2 参考 Grant 和 Hofmann [60] 的实验设计,在华南地区商学院邀请 191 名学员(有工作经验) 作为被试完成一项任务. 实验包含三个阶段: 第一阶段,在到达实验场地后,实验员要求所有被试填写人口统计学信息、核心自我评价、感知信任. 随后,实验员告知被试,商学院正在招聘一名科研助理,你将收到一篇博士生撰写的英文学术论文草稿片段,为检验英文论文写作能力,请使用 Microsoft Word 中的跟踪修订模式,找到并修改这篇论文中存在的语法、拼写错误,保存文件,完成后提交.

第二阶段,在完成修订任务后,实验员告知被试,将被随机分为三组,依次轮流接受结果反馈. 为启动 AI 组的实验条件,参考 Luo 等[7] 对 AI 技术逻辑的描述,AI 组的被试(63 人),在接受负面反馈前被告知向其提供反馈的是 AI,它拥有大数据分析技能和基于深度学习的技术(如,文本挖掘、机器翻译等),主要由算法工程师开发的机器学习算法、自然语言处理技术相结合并通过自主学习知识数据库中已有的最佳实践来自动化进行文本写作、翻译和润色等工作,它通过人机交互向用户提供功能. 为启动威权领导组的实验条件,参考彭坚等[59] 对领导风格进行实验刺激的做法,根据威权领导的定义和测量内容[61],威权领导组的被试(64 人),在接受负面反馈前被告知向其提供反馈的是一位学术专家领导,日常行事作风具有如下特点:经常要求别人完全服从其指示;包揽所有决定;在会议上拥有最后发言权;在别人面前居高临下;给别人带来压力;严格要求别人;因为任务未被完成责骂别人;要求最好的表现;要求别人必须遵守其定下的规则,违反则会严厉惩罚. 同样,为启动仁慈领导组的实验条件,根据仁慈领导的定义和测量内容[61],仁慈领导组的被试(64 人),在接受负面反馈前被告知向其提供反馈的是一位学术专家领导,他日常行事作风具有如下特点:经常与别人像家人一样相处;对别人花费大量精力;除了工作,还经常关心别人的生活;经常关心别人是否舒心;帮助别人应对紧急情况;非常体贴别人;尽可能地满足别人的个性化需求;别人遇到困难时会给予鼓励;照顾别人的身边人;别人表现不好时,会主动了解分析原因;为别人处理日常生活中难以解决的事情. 随后,参考 Yam 等[19] 和 Kim 等[30] 的实验设计,每组接收负面反馈的被试都获得了如下信息:你在刚刚修订论文的任务中表现不佳,得分排在所有参与者中的后 20%,明显低于平均水平,你需要认真检查语法、词汇拼写等方面的错误. 此外,为增加心理真实感,参考 Tang 等[36] 的做法,AI 组的负面反馈信息,由文本通过音频剪辑工具转录为模仿智能机器的语音,威权领导组、仁慈领导组的负面反馈,分别由一位扮演学术专家的实验员面对面提供.

第三阶段,在收到负面反馈后,实验员要求被试填写操纵检验的测量、绩效提升归因动机和伤害诱发动机归因、消极情绪的测量. 随后,实验员告知被试,他们需要再次完成论文修订的任务,并重新发放一篇英文学术论文草稿片段,该篇论文被人为嵌入 66 个语法、拼写错误,待被试完成后再次提交.

3.2 变量测量

研究 2 确定最终实验所使用量表的标准程序与研究 1 一致.

负面反馈源. 将 AI 赋值 0,威权领导赋值 1,仁慈领导赋值 2.

反馈源操纵. 与研究 1 一致. 信度系数为 0.81.

领导风格操纵. 使用 Cheng 等[61] 对威权领导、仁慈领导的测量条目. 调查内容为 “你认为刚刚为你提供的学术专家多大程度上符合以下风格?”威权领导的代表题项为 “要求别人完全服从他的指令.”,仁慈领导的代表题项为 “当别人遇到困难时会提供鼓励.”信度系数分别为 0.91、0.93.

工作绩效. 参考 Grant 和 Hofmann[60] 实验设计中的测量方法,从准确性的角度通过计算被试在参与论文修改任务中成功纠正拼写和语法错误的数量来衡量工作绩效. 在实验完成后,邀请了一位英语专业的老师和一位商科教授(未告知实验目的)独立核查论文中的错误.

绩效提升动机归因($\alpha = 0.87$)、伤害诱发动机归因($\alpha = 0.84$)和消极情绪($\alpha = 0.86$)、核心自我评价($\alpha = 0.92$)、感知信任($\alpha = 0.92$)的控制变量的测量方式与研究 1 一致.

3.2.1 验证性因子分析

CFA 结果表明,五因子模型各拟合指数基本满足要求(χ^2=505.18,df =367,$TLI = 0.94$,$CFI = 0.95$,$RMSEA = 0.04$,$SRMR = 0.05$),且该模型显著优于其它竞争模型,说明测量变量具有较好的区分效度.

3.2.2 描述性统计分析

研究 2 的描述性统计分析结果如表 3 所示,所得结果为假设检验提供支持.

表 3　描述性统计分析结果(研究 2)

Table 3 Descriptive statistics analysis results (Study 2)

变量	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. 性别	1.53	0.50	—	—	—	—	—	—	—	—	—	—	—
2. 年龄	32.78	3.71	0.03	—	—	—	—	—	—	—	—	—	—
3. 学历	2.96	0.46	0.06	0.12	—	—	—	—	—	—	—	—	—
4. 任职期限	3.62	1.68	0.03	0.17*	0.00	—	—	—	—	—	—	—	—
5. 消极情绪	4.26	0.89	0.15*	0.06	0.07	0.10	(0.86)	—	—	—	—	—	—
6. 核心自我评价	4.28	0.93	0.01	0.12	0.15*	0.12	−0.03	(0.92)	—	—	—	—	—
7. 感知信任	4.13	0.96	−0.06	−0.08	−0.05	−0.01	0.06	−0.12	(0.83)	—	—	—	—
8. 负面反馈源(AI=0，威权=1，仁慈=2)	1.01	0.82	−0.01	−0.03	0.01	0.01	0.11	0.03	−0.01	—	—	—	—
9. 绩效提升动机归因	4.39	0.90	−0.02	−0.00	0.08	0.07	−0.05	0.05	0.05	−0.00	(0.87)	—	—
10. 伤害诱发动机归因	4.32	0.89	0.15*	0.07	−0.09	0.05	0.09	−0.06	0.06	0.02	−0.18*	(0.84)	—
11. 工作绩效	34.27	10.71	0.04	0.01	0.15*	−0.01	−0.11	0.06	−0.03	0.03	0.35**	−0.29**	—

注: *M* 表示均值，*SD* 表示标准差; * 表示 $p<0.05$，** 表示 $p<0.01$，*** 表示 $p<0.001$，双侧; 对角线括号内的数值为各变量的信度系数.

3.2.3　操纵检验

首先，反馈源的操纵检验与研究 1 一致，独立样本 *t* 检验结果表明，AI 组($M=2.84$，$SD=0.78$)显著低于人类领导组($M=3.69$，$SD=0.94$)，$t(189)=5.47$，$p<0.001$. 其次，领导风格的操纵检验同样使用独立样本 t 检验，结果表明，针对威权领导的评分，威权领导组($M=3.91$，$SD=0.83$)显著高于仁慈领导组($M=3.13$，$SD=0.75$)，$t(126)=5.63$，$p<0.001$. 而针对仁慈领导的评分，威权领导组($M=3.28$，$SD=0.76$)显著低于仁慈领导组($M=4.05$，$SD=0.95$)，$t(126)=5.11$，$p<0.001$. 结果说明研究 2 实验操纵成功.

3.2.4　假设检验

研究 2 的目的是，比较在负面反馈事件中，AI、威权领导和仁慈领导作为不同反馈源在激励员工工作绩效方面的差异化效应. 鉴于研究 2 遵循单因素三水平的实验设计，首先使用单因素方差分析(One-way Anova)法. 结果表明(见图 3、图 4)，负面反馈源对绩效提升动机归因的影响显著($F(2,188)=6.81$，$p<0.01$，$\eta^2=0.07$). 随后，使用 Bonferroni 法进行两两比较，发现 AI 组($M=4.56$，$SD=0.85$)显著高于与威权领导组($M=4.06$，$SD=0.90$)，$p<0.01$; 威权领导组($M=4.06$，$SD=0.90$)显著低于仁慈领导组($M=4.55$，$SD=0.87$)，$p<0.01$; 而 AI 组($M=4.56$，$SD=0.85$)与仁慈领导组($M=4.55$，$SD=0.87$)无显著差异，$p>0.05$. 此外，负面反馈源对伤害诱发动机归因的影响显著($F(2,188)=5.95$，$p<0.01$，$\eta^2=0.06$). 两两比较发现，AI 组($M=4.14$，$SD=0.81$)显著低于与威权领导组($M=4.62$，$SD=0.94$)，$p<0.01$; 威权领导组($M=4.62$，$SD=0.94$)显著高于仁慈领导组($M=4.18$，$SD=0.87$)，$p<0.05$; 而 AI 组($M=4.14$，$SD=0.81$)与仁慈领导组($M=4.18$，$SD=0.87$)无显著差异，$p>0.05$. 结果表明，与威权领导比，AI 负面反馈能激发员工更高的绩效提升动机归因、导致更低的伤害诱发动机归因; AI 与仁慈领导无显著差异，H3(a)、H3(b)得到支持.

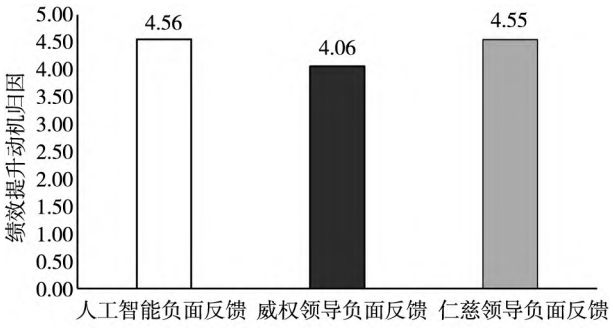

图 3　不同负面反馈源下的绩效提升动机归因(研究 2)

Fig. 3 Attributed performance promotion motives under various negative feedback sources (Study 2)

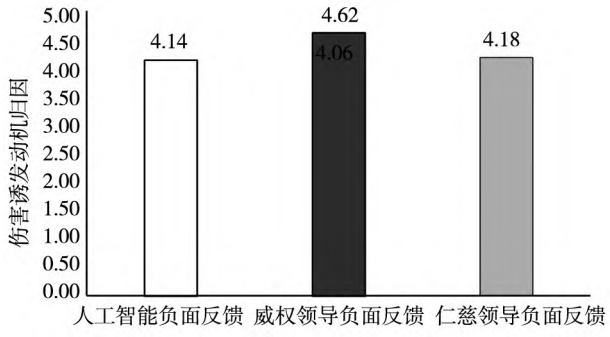

图 4　不同负面反馈源下的伤害诱发动机归因（研究 2）

Fig. 4 Attributed injury initiation motives under various negative feedback sources (Study 2)

在检验绩效提升动机归因、伤害诱发动机归因的中介效应前，与研究 1 一致，通过回归分析发现，绩效提升动机归因与员工工作绩效显著正相关（$\beta = 0.35$，$p < 0.001$），伤害诱发动机归因与员工工作绩效显著负相关（$\beta = -0.30$，$p < 0.001$），H1(b)、H2(b) 再次得到支持. 随后进行中介效应检验，以 AI 为参照，加入控制变量后的结果表明，与 AI 相比，威权领导负面反馈通过绩效提升动机归因促使员工工作绩效水平更高的间接效应为 −0.063，95% 置信区间为 [−0.12，−0.02]，不包括 0；与 AI 相比，仁慈领导负面反馈通过绩效提升动机归因促使员工工作绩效水平更高的间接效应为 0.001，95% 置信区间为 [−0.04，0.04]，包括 0. 结果说明，与威权领导相比，AI 负面反馈能通过激发员工绩效提升动机归因促进其更高水平的工作绩效，而与仁慈领导相比，AI 负面反馈的差异化间接效应消失. 此外，在伤害诱发动机归因方面，威权领导与 AI 比的差异化间接效应为 −0.048，95% 置信区间为 [−0.11，−0.02]，不包括 0；而仁慈领导与 AI 比的差异化间接效应为 −0.004，95% 置信区间为 [−0.04，0.03]，包括 0. 结果说明，与威权领导相比，AI 负面反馈会通过减少员工伤害诱发动机归因促进其更高水平的工作绩效，而与仁慈领导相比，AI 负面反馈的差异化间接效应消失. H4(a)、H4(b) 得到支持.

4　结束语

4.1　研究结论

基于动机归因视角，探究了 AI（vs. 人类领导）负面反馈对员工工作绩效的差异化影响效应与机制. 通过实证分析得出如下结论：1）AI（vs. 人类领导）负面反馈通过激发员工更高的绩效提升动机归因和更低的伤害诱发动机归因，继而激励员工取得更高的工作绩效；2）进一步考虑本土化领导风格的作用，即 AI（vs. 威权领导）负面反馈通过激发员工更高的绩效提升动机归因和更低的伤害诱发动机归因，继而激励员工取得更高的工作绩效；3）在负面反馈通过绩效提升动机归因、伤害诱发动机归因间接影响员工工作绩效方面，AI（vs. 仁慈领导）差异不显著.

4.2　理论贡献

本研究对已有文献做出如下贡献.

首先，拓展负面反馈在数智化情境下的相关研究. 已有研究围绕领导者提供负面反馈如何影响下属反应开展大量实证研究. 然而，现存争议说明领导者提供负面反馈需要权衡收益（如，领导有效性、员工创造力等）和成本（如，领导分心、人际冲突等）[1−4，29−32]，仍需探究何种方式能有效发挥负面反馈的激励鞭策作用. 得益于数智技术的进步，AI 具有替代领导者承担负面反馈职责的潜力和能力[7，19]，但当前学界还不清晰，在负面反馈情境下，反馈源是 AI 而不是人类领导时，更有助于帮助员工改善绩效. 本研究填补了上述研究缺口，在数智化情境下揭示了与人类领导比，AI 负面反馈在激励员工工作绩效方面的优势.

其次，细化 AI 绩效反馈在反馈效价为负时的影响研究. AI 绩效反馈是一个涉及多学科交叉的前沿研究领域[62]. 部分研究发现，AI 绩效反馈存在积极的部署效应，通过改善反馈质量促进员工生产力[6]. 另有研究指出，AI 绩效反馈也会带来信息过载、决策厌恶等问题，不利于绩效提升[7]. 现存争议说明 AI 绩效反馈能否给员工带来好处需要结合具体管理情境进行讨论. 新近研究初步揭示了拟人化 AI 在负面反馈情境下的不利影响（即，个体报复）[19]，但 AI 负面反馈能否激励员工改善绩效尚不知晓. 本研究揭示了 AI 执行特定类型的反馈任务（即负面反馈）对员工工作绩效的积极影响.

再次，挖掘了 AI 负面反馈影响员工绩效的作用“黑箱”，拓展了动机归因视角. 已有基于动机归因视角的研究主要强调人际互动中行为接收者对行为发出者的目的或动机进行因果推断的过

程[15-17]，但极少研究发现个体在人机交互过程中同样会对作为非人类实体的行为发出方的目的或动机进行归因. 过去研究认为客观上机器具有低程度的能动性，不存在行为意图或倾向[19,21]. 本研究基于理论和实证结果证明主观上员工评估 AI 行为动机的存在是可能的，发现 AI(vs. 人类领导) 负面反馈激励员工更高绩效的原因在于他们更倾向于将 AI 提供负面反馈的动机归因为绩效提升导向的、而非伤害诱发导向的.

最后，将本土领导理论和数智化情境相结合，揭示了负面反馈中家长式领导与 AI 的差异化作用机制，进一步丰富本土领导理论[63]. 不论是负面反馈的相关文献，还是涉及 AI 与人类领导绩效反馈比较的相关研究，基于中国文化背景进行研究的文献寥寥无几. 根据家长式领导理论[48]，引入彰显华人“立威”与“施恩”领导风格的威权领导与仁慈领导，比较了在负面反馈影响员工工作绩效方面，AI 与威权领导、仁慈领导分别比较，是否具有差异化效应与机制. 研究结果揭示了一个有趣的现象，AI(vs. 威权领导) 负面反馈在激励员工取得更高绩效方面具有优势效应，但相较于仁慈领导，AI 负面反馈影响员工工作绩效的优势并不显著，说明仁慈领导提供负面反馈的有效性.

4.3　实践启示

本研究对管理实践的主要启示如下.

首先，对企业而言，未来可以考虑部署 AI 替代领导者执行负面反馈职责，一是解决领导者对负面反馈的困扰，避免领导者分散注意力[3]，二是避免员工因收到负面反馈而消极反应，使其专注于改进不足、提升绩效、创造个人价值[1,2]，三是说明基于 AI 的绩效反馈系统在市场上具有较好商业前景和投资价值[6]；其次，对领导者而言，研究发现 AI 在提供负面反馈通常表现出不附加负面情感地向员工提出客观、一致、准确、可操作反馈建议的行为特征，因而造成员工更倾向于对 AI 的行为动机做出积极归因. 这从侧面提醒领导者必须提供负面反馈时不仅要尽量避免向员工传递负面态度[12]，还要下足功夫学习如何提升反馈的客观性、一致性、准确性和可操作性，而非率性而为之; 此外，研究还发现具有威权领导风格的领导并不适合向员工提供负面反馈，相比之下，仁慈领导提供负面反馈的效果与 AI 一样具有优势. 因此，启示领导者在日常行事作风上要修炼和磨砺积极的领导艺术，结合中国复杂的人际关系和人情社会的背景[52]，汲取仁慈领导风格中保全下属面子、回应下属需求、关怀下属工作与生活等积极的行为特征因素，避免威权领导风格中贬低下属能力、忽视下属言论等消极的行为特征因素[48,55]. 最后，对员工而言，做出绩效提升动机归因对其工作绩效提升有显著的预测作用，因此员工在经历负面事件时(如，领导批评) 能学会做出积极的归因判断有助于减少自身的消极应激反应(如，愤怒)，还有助于对自身绩效提升带来好处[15]，鼓励员工尽可能地对与之交互的外界环境因素进行积极归因而非消极归因.

4.4　局限与展望

本研究还有如下局限，并对未来研究指出方向.

首先，方法上虽然采取了 2 个实验设计以增强研究的内外部效度和结论的稳健性，但研究 1 实验样本并非完全随机的，可能存在初始点的差别，降低了因果关系解释的说服力. 此外，新技术的迭代也可能会影响研究发现; 研究 2 实验室实验难以捕获真实工作场景中复杂的环境，与自然条件下实验刺激仍具有一定差距. 此外，本研究开展时主要基于当时主流的 AI 技术范式(如判别式、规则驱动的机器学习算法)，然而近年来生成式 AI(如大语言模型) 的快速发展显著改变了人机交互的方式与体验[64]. 生成式 AI 具备更强的自然语言理解与生成能力、更高的交互拟人化水平以及更丰富的上下文适应性，这可能会对员工在接收负面反馈时的动机归因产生不同于传统 AI 的影响. 因此，本研究的结论在生成式 AI 广泛应用的情境下可能需要重新审视。未来研究应结合纵向问卷调查、案例研究等多种方法，并关注 AI 技术迭代(尤其是生成式 AI) 对反馈机制的影响，以弥补上述不足，进一步验证和拓展本研究的发现; 其次，本研究仅关注 AI 与人类领导在提供负面反馈时对员工反应的差异，根据反馈干预理论[32]，反馈手段和方式是多样的(如，正面反馈、延时反馈等)，建议未来研究进一步探索更多反馈干预方式下，AI 与人类领导之间的作用差异. 此外，本研究仅从动机归因视角揭示了 AI(vs. 人类领导) 负面反馈影响员工工作绩效的差异化机

制，根据反馈干预理论[30,32]，员工注意力导向分为专注任务细节的任务过程以及专注自我概念的元过程，这可能在解释 AI 与人类领导如何差异化影响员工绩效方面同样重要. 建议未来研究结合其他理论进一步挖掘机制“黑箱”；再次，本研究仅比较 AI 与家长式领导理论中威权领导、仁慈领导的差异，领导风格还有悖论式领导[65]、谦卑式领导[66]等，未来研究可以继续识别在绩效反馈中，AI 与具有上述其他的领导风格的人类领导之间是否具有差异，倘若存在，该差异又是如何表现的. 最后，本研究仅揭示 AI 负面反馈的优势，而 AI 负面反馈的“阴暗面”或“双刃剑”效应仍亟待未来研究.

参 考 文 献:

[1] Milberg S J, Sinn F. Vulnerability of global brands to negative feedback effects [J]. Journal of Business Research, 2008, 61(6): 684 – 690.

[2] 董念念，尹 奎，邢 璐，等. 领导每日消极反馈对员工创造力的影响机制 [J]. 心理学报，2023，55(5): 831 – 843.
Dong Niannian, Yin Kui, Xing Lu, et al. The effects of daily supervisor negative feedback on employee creativity [J]. Acta Psychologica Sinica, 2023, 55(5): 831 – 843. (in Chinese)

[3] Simon L S, Rosen C C, Gajendran R S, et al. Pain or gain? Understanding how trait empathy impacts leader effectiveness following the provision of negative feedback [J]. Journal of Applied Psychology, 2022, 107(2): 279 – 297.

[4] Zenger J, Folkman J. Why do so many managers avoid giving praise? [J/OL]. Harvard Business Review, 2017. https://hbr.org/2017/05/why-do-so-many-managers-avoid-giving-praise.

[5] Bankins S, Ocampo A C, Marrone M, et al. A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice [J]. Journal of Organizational Behavior, 2023, 45(2): 159 – 182.

[6] Tong S, Jia N, Luo X, et al. The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance [J]. Strategic Management Journal, 2021, 42(9): 1600 – 1631.

[7] Luo X, Qin M S, Fang Z, et al. Artificial intelligence coaches for sales agents: Caveats and solutions [J]. Journal of Marketing, 2021, 85(2): 14 – 32.

[8] Rivera M, Qiu L, Kumar S, et al. Are traditional performance reviews outdated? An empirical analysis on continuous, real-time feedback in the workplace [J]. Information Systems Research, 2021, 33(2): 517 – 540.

[9] Kellogg K C, Valentine M A, Christin A. Algorithms at work: The new contested terrain of control [J]. Academy of Management Annals, 2020, 14(1): 366 – 410.

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

[11] Tang P M, Koopman J, Yam K C, et al. The self-regulatory consequences of dependence on intelligent machines at work: Evidence from field and experimental studies [J]. Human Resource Management, 2023, 62(5): 721 – 744.

[12] Alam M, Singh P. Performance feedback interviews as affective events: An exploration of the impact of emotion regulation of negative performance feedback on supervisor-employee dyads [J]. Human Resource Management Review, 2019, 31(2): 1 – 14.

[13] Weiner B. An attributional theory of achievement motivation and emotion [J]. Psychological Review, 1985, 92(4): 548 – 573.

[14] Thomas K W, Pondy L R. Toward and “intent” model of conflict management among principal parties [J]. Human Relations, 1977, 30(12): 1089 – 1102.

[15] Xing L, Sun J, Jepsen D, et al. Supervisor negative feedback and employee motivation to learn: An attribution perspective [J]. Human Relations, 2021, 76(2): 310 – 340.

[16] Tepper B J. Abusive supervision in work organizations: Review, synthesis, and research agenda [J]. Journal of Management, 2007, 33(3): 261 – 289.

[17] Liu D, Liao H, Loi R. The dark side of leadership: A three-level investigation of the cascading effect of abusive supervision on employee creativity [J]. Academy of Management Journal, 2012, 55(5): 1187 – 1212.

[18] Synofzik M, Vosgerau G, Voss M. The experience of agency: An interplay between prediction and postdiction [J]. Fron-

tiers in Psychology，2013，4(127)：1－8.
[19] Yam K C，Goh E Y，Fehr R，et al. When your boss is a robot：Workers are more spiteful to robot supervisors that seem more human [J]. Journal of Experimental Social Psychology，2022，102：1－12.
[20] Gray H M，Gray K，Wegner D M. Dimensions of mind perception [J]. Science，2007，315(5812)：619.
[21] Gray K，Wegner D M. Feeling robots and human zombies：Mind perception and the uncanny valley [J]. Cognition，2012，125(1)：125－130.
[22] Jordan M I，Mitchell T M. Machine learning：Trends，perspectives，and prospects [J]. Science，2015，349(6245)：255－260.
[23] Burrell J. How the machine ‘thinks’：Understanding opacity in machine learning algorithms [J]. Big Data and Society，2016，3(1)：1－12.
[24] Miron M，Tolan S，Gómez E，et al. Evaluating causes of algorithmic bias in juvenile criminal recidivism [J]. Artificial Intelligence and Law，2021，29(2)：111－147.
[25] Agrawal A，Gans J，Goldfarb A. What to expect from artificial intelligence [J]. MIT Sloan Management Review，2017，58(3)：28－37.
[26] 裴嘉良，刘善仕，钟楚燕，等. AI 算法决策能提高员工的程序公平感知吗？[J]. 外国经济与管理，2021，43(11)：41－55.
Pei Jialiang，Liu Shanshi，Zhong Chuyan，et al. Can AI algorithmic decision-making improve employees’ perception of procedural fairness？[J]. Foreign Economics & Management，2021，43(11)：41－55.(in Chinese)
[27] Nass C，Moon Y. Machines and mindlessness：Social responses to computers [J]. Journal of Social Issues，2000，56(1)：81－103.
[28] Batson C D，Shaw L L. Evidence for altruism：Toward a pluralism of prosocial motives [J]. Psychological Inquiry，1991，2(2)：107－122.
[29] Bear J B，Cushenberry L，London M，et al. Performance feedback，power retention，and the gender gap in leadership [J]. The Leadership Quarterly，2017，28(6)：721－740.
[30] Kim Y J，Kim J. Does negative feedback benefit (or harm) recipient creativity？The role of the direction of feedback flow [J]. Academy of Management Journal，2020，63(3)：584－612.
[31] Peterson R S，Behfar K J. The dynamic relationship between performance feedback，trust，and conflict in groups：A longitudinal study [J]. Organizational Behavior and Human Decision Processes，2003，92(1－2)：102－112.
[32] Kluger A N，DeNisi A S. The effects of feedback interventions on performance：A historical review，a meta-analysis，and a preliminary feedback intervention theory [J]. Psychological Bulletin，1996，119(2)：254－284.
[33] Qin M S，Jia N，Luo X，et al. Perceived fairness of human managers compared with artificial intelligence in employee performance evaluation [J]. Journal of Management Information Systems，2023，40(4)：1039－1070.
[34] 徐　鹏，徐向艺. 人工智能时代企业管理变革的逻辑与分析框架 [J]. 管理世界，2020，36(1)：122－129＋238.
Xu Peng，Xu Xiangyi. Change logic and analysis framework of enterprise management in the era of artificial intelligence [J]. Management World，2020，36(1)：122－129＋238.(in Chinese)
[35] Yam K C，Tang P M，Jackson J C，et al. The rise of robots increases job insecurity and maladaptive workplace behaviors：Multimethod evidence [J]. Journal of Applied Psychology，2023，108(5)：850－870.
[36] Tang P M，Koopman J，McClean S T，et al. When conscientious employees meet intelligent machines：An integrative approach inspired by complementarity theory and role theory [J]. Academy of Management Journal，2022，65(3)：1019－1054.
[37] Jago A S. Algorithms and authenticity [J]. Academy of Management Discoveries，2019，5(1)：38－56.
[38] Newman D T，Fast N J，Harmon D J. When eliminating bias isn’t fair：Algorithmic reductionism and procedural justice in human resource decisions [J]. Organizational Behavior and Human Decision Processes，2020，160：149－167.
[39] 高良谋，王　磊. 偏私的领导风格是否有效？——基于差序式领导的文化适应性分析与理论延展 [J]. 经济管理，2013，35(4)：183－194.
Gao Liangmou，Wang Lei. Does favoritism leadership style is effective？Cultural adaptability analysis and theoretical extension of the chaxu leadership [J]. Business and Management Journal，2013，35(4)：183－194.(in Chinese)
[40] Brynjolfsson E，Mitchell T M. What can machine learning do？Workforce implications [J]. Science，358(6370)：1530－1534.

[41] Carmeli A, Shalom R, Weisberg J. Considerations in organizational career advancement: What really matters [J]. Personnel Review, 2007, 36(2): 190 – 205.

[42] Lindholm N. Performance management in MNC subsidiaries in China: A study of host-country managers and professionals [J]. Asia Pacific Journal of Human Resources, 1999, 37(3): 18 – 35.

[43] Kay E, Meyer H H, French J R P. Effects of threat in a performance appraisal interview [J]. Journal of Applied Psychology, 1965, 49(5): 311 – 317.

[44] Mygland M J, Schibbye M, Pappas I O, et al. Affordances in Human-Chatbot Interaction: A review of the literature [J]. IFIP International Conference on E-Business, E-Services, and E-Society, 2021: 3 – 17.

[45] Magee J C, Smith P K. The social distance theory of power [J]. Personality and Social Psychology Review, 2013, 17(2): 158 – 186.

[46] Xing L, Sun J J, Jepsen D. Feeling shame in the workplace: Examining negative feedback as an antecedent and performance and well-being as consequences [J]. Journal of Organizational Behavior, 2021, 42(9): 1244 – 1260.

[47] Brosschot J F, Gerin W, Thayer J F. The perseverative cognition hypothesis: A review of worry, prolonged stress-related physiological activation, and health [J]. Journal of Psychosomatic Research, 2006, 60(2): 113 – 124.

[48] Pellegrini E K, Scandura T A. Paternalistic leadership: A review and agenda for future research [J]. Journal of Management, 2008, 34(3): 566 – 593.

[49] **刘善仕, 凌文辁. 家长式领导与员工价值取向关系实证研究 [J]. 心理科学, 2004, (3): 674 – 676.**
Liu Shanshi, Ling Wenquan. The research on the relationship of value orientation of employees with paternalistic leadership [J]. Journal of Psychological Science, 2004, (3): 674 – 676. (in Chinese)

[50] Chan S C H, Huang X, Snape E, et al. The Janus face of paternalistic leaders: Authoritarianism, benevolence, subordinates' organization-based self-esteem, and performance [J]. Journal of Organizational Behavior, 2012, 34(1): 108 – 128.

[51] **周 浩, 龙立荣. 恩威并施, 以德服人——家长式领导研究述评 [J]. 心理科学进展, 2005, 13(2): 227 – 238.**
Zhou Hao, Long Lirong. A review of paternalistic leadership research [J]. Advances in Psychological Science, 2005, 13(2): 227 – 238. (in Chinese)

[52] **李 锐, 田晓明. 主管威权领导与下属前瞻行为: 一个被中介的调节模型构建与检验 [J]. 心理学报, 2014, 46(11): 1719 – 1733.**
Li Rui, Tian Xiaoming. Supervisor authoritarian leadership and subordinate proactive behavior: Test of amediated-moderation model [J]. Acta Psychologica Sinica, 2014, 46(11): 1719 – 1733. (in Chinese)

[53] **沈伊默, 周婉茹, 魏丽华, 等. 仁慈领导与员工创新行为: 内部人身份感知的中介作用和领导 – 部属交换关系差异化的调节作用 [J]. 心理学报, 2017, 49(8): 1100 – 1112.**
Shen Yimo, Zhou Wanru, Wei Lihua, et al. Benevolent leadership and subordinate innovative behavior: The mediating role of perceived insider status and the moderating role of leader-member exchange differentiation [J]. Acta Psychologica Sinica, 2017, 49(8): 1100 – 1112. (in Chinese)

[54] Aryee S, Chen Z X, Sun L Y, et al. Antecedents and outcomes of abusive supervision: Test of a trickle-down model [J]. Journal of Applied Psychology, 2007, 92(1): 191 – 201.

[55] Farh J L, Liang J, Chou L F, et al. Paternalistic Leadership in Chinese Organizations: Research Progress and Future Research Directions [M] // Chen C C, Lee Y T. Leadership and Management in China, Philosophies, Theories, and Practice. London: Cambridge University Press, 2008: 171 – 205.

[56] Simon C H, Mak W M. Benevolent leadership and follower performance: The mediating role of leader-member exchange (LMX) [J]. Asia Pacific Journal of Management, 2012, 29(2): 285 – 301.

[57] Liu C, Spector P E, Shi L. Cross-national job stress: A quantitative and qualitative study [J]. Journal of Organizational Behavior, 2007, 28(2): 209 – 239.

[58] Judge T A, Erez A, Bono J E, et al. The core self-evaluation scale: Development of a measure [J]. Personnel Psychology, 2003, 56(2): 303 – 331.

[59] **彭 坚, 尹 奎, 侯 楠, 等. 如何激发员工绿色行为? 绿色变革型领导与绿色人力资源管理实践的作用 [J]. 心理学报, 2020, 52(9): 1105 – 1120.**
Peng Jian, Yin Kui, Hou Nan, et al. How to facilitate employee green behavior: The joint role of green transformational leadership and green human resource management practice [J]. Acta Psychologica Sinica, 2020, 52(9): 1105 – 1120. (in Chinese)

[60] Grant A M, Hofmann D A. Outsourcing inspiration: The performance effects of ideological messages from leaders and beneficiaries [J]. Organizational Behavior and Human Decision Processes, 2011, 116(2): 173 – 187.
[61] Cheng B S, Chou L F, Wu T Y, et al. Paternalistic leadership and subordinate responses: Establishing a leadership model in Chinese organizations [J]. Asian Journal of Social Psychology, 2004, 7(1): 89 – 117.
[62] 董毓格，龙立荣，程芷汀. 数智时代的绩效管理：现实和未来 [J]. 清华管理评论，2022，101(5)：93 – 100.
Dong Yuge, Long Lirong, Cheng Zhiting. Performance management in the digital intelligence era: Reality and future [J]. Tsinghua Business Review, 2022, 101 (5): 93 – 100. (in Chinese)
[63] 张晓军，韩　巍，席酉民，等. 本土领导研究及其路径探讨 [J]. 管理科学学报，2017，20(11)：36 – 48.
Zhang Xiaojun, Han Wei, Xi Youmin, et al. Chinese indigenous leadership research: Research questions and process [J]. Journal of Management Sciences in China, 2017, 20(11): 36 – 48. (in Chinese)
[64] Marimon F, Mas-Machuca M, Akhmedova A. Trusting in generative AI: Catalyst for employee performance and engagement in the workplace [J]. International Journal of Human-Computer Interaction, 2025, 41(7): 7076 – 7091.
[65] Zhang Y, Waldman D A, Han Y L, et al. Paradoxical leader behaviors in people management: Antecedents and consequences [J]. Academy of Management Journal, 2015, 58(2): 538 – 566.
[66] 陈力凡，刘圣明，胡小丽. 社会认同视角下谦卑型领导与员工主动性行为 [J]. 管理科学学报，2022，25(2)：104 – 115.
Chen Lifan, Liu Shengming, Hu Xiaoli. Leader humble behavior and follower proactive behavior: A social identity perspective [J]. Journal of Management Sciences in China, 2022, 25(2): 104 – 115. (in Chinese)

Research on performance incentive mechanisms of negative feedback under human-AI Interaction

LIU Shan-shi[1,2], *PEI Jia-liang*[1,2*], *WANG Hong-li*[1,2], *GE Chun-mian*[1,2], *JIANG Jun-hui*[1]

1. School of Business Administration, South China University of Technology, Guangzhou 510641, China;
2. Guangdong Key Laboratory of Internet Behavior Sciences, Guangzhou 510641, China

Abstract: Motivating employee performance improvement through negative feedback is a persistent challenge in performance management. As digital and intelligent technologies advance, some tech firms have begun leveraging cutting-edge tools to optimize performance management processes and enhance employee experience. Adopting a motives attribution perspective, this study compares the differentiated performance incentive mechanisms of artificial intelligence (AI) and human leaders when delivering negative feedback. Empirical findings show that, relative to human leaders, AI-provided negative feedback elicits stronger attributed performance-promotion motives and weaker attributed injury-initiation motives, which in turn lead to higher employee performance. Further grounding the inquiry in the indigenous Chinese context reveals the moderating role of leadership style: AI exhibits a greater advantage over authoritarian leadership in transmitting the indirect effects of negative feedback on performance through these two motives; however, compared with benevolent leadership, this differentiated indirect effect is attenuated. The study uncovers the performance-incentive effects of negative feedback in human-AI interaction, broadens the context, perspectives, and approaches for research on AI and negative feedback, and offers insights for the digital and intelligent transformation of performance management practices in Chinese enterprises.

Key words: artificial intelligence; negative feedback; job performance; motives attribution; authoritarian leadership; benevolent leadership