《心理学报》审稿意见与作者回应

题目：负面绩效反馈下员工绩效改进动机的人机比较研究

作者：王国轩, 龙立荣, 李绍龙, 孙芳, 望家晴, 黄世英子

第一轮

审稿人 1 意见：

很高兴有机会审阅《人机负面绩效反馈对个体绩效改进动机的差异化影响机制：基于归因的视角》，本研究主题新颖，具有较好的理论价值和管理实践意义。目前版本中，理论和方法方面仍有一些问题需要澄清和解决，具体如下：

回应：非常感谢您对于本研究主题、理论价值和管理实践意义的认可与肯定！同时也非常感谢您在耐心审阅之后，不仅指出了本文的不足，还为我们提供了明确的修改思路和方向！希望修改后的论文很好地解决了您关注的问题。

意见 1：文聚焦于负面绩效反馈，并将其定义为“负面绩效反馈（negative performance feedback）则是组织对未达到业绩期望的员工所给予的指示、否定和批评（Cianci et al., 2010），旨在引导、激励和强化员工工作行为，最终提升其绩效水平（Podsakoff & Farh, 1989; Lam et al., 2011）。”因此，本文对比的是个体在面对人类管理者或 AI 传递的未达期望绩效目标的负面信息时，后续绩效改进动机的差异。如作者所言，“特别是面对准确和具体负面反馈，较正面反馈，员工的绩效改进动机程度更强（e.g., Podsakoff & Farh, 1989；Borovoi, Schmidtke, & Vlaev, 2022）。”同时，作者在假设 1 的论证中提到，“个体认为来自 AI 的反馈具备更高的准确性、无偏性和一致性（Raveendhran and Fast, 2021），从而增强员工对于提升和改进绩效的动机。”；在假设 2 的论证中提到，“来自 AI 的负面绩效反馈固然更加客观，但由于其缺乏如人类般的沟通与情绪表达能力，致使其较难获得个体的信赖，并较难促使个体产生较高的绩效改进动机水平。”

——以上论证都在说明：负面反馈的质量高低，是否准确和具体，是否得到员工的信任，是后续绩效改进动机的决定因素。那么，为什么没有把“负面反馈的质量、准确性和具体性、公平性和一致性、对负面反馈的信任等”作为中介机制，而是选用了内在归因呢？本研究对于内在归因的论证显得较为薄弱。

回应：非常感谢您的意见。首先，本研究聚焦于组织管理中的负面绩效反馈场景。而当前许多研究显示，归因是负面绩效反馈后员工行动意愿的经典解释机制。比如，Xing 等(2023)发现，相较于外部归因(比如将负面绩效反馈的成因归结于领导者的偏见)，当员工内部归因时(比如关注自身的不足)，会增强学习行为。那么归因的机制是否能够应用于人机反馈的情境中呢？据此，我们考虑使用先前人机研究中较少使用的内部与外部归因作为中介变量，丰富人机中介机制的研究，也据此拓展归因理论在数智化场景中的解释性。其次，正如您所指出的，本研究在前言部分以及假设推理部分使用了过多“AI 准确、具体、无偏”的论断，这使得本研究视角不够聚焦。为此，我们在新版本中聚焦于员工接收负面绩效反馈后的归因心理过程。

比如，我们在引言开头就提出了归因视角下，传统实践中由管理者提供负面绩效反馈的不足。“出于自我服务偏差，员工收到负面绩效反馈后会更多地进行外部归因(例如，将收到负面绩效反馈的原因归结于领导的情绪不好)，而较少进行内部归因(例如，归因于自身的努力程度不够)(e.g., Harvey et al, 2014)。但对负面绩效反馈的外部归因往往不利于员工进行反思及绩效改进(e.g., Harvey & Martinko, 2009)”。此外，在中国的高语境文化下，一方面，出于“人情”与“面子”的顾虑，人们的沟通方式比较含蓄，管理者通常采用较为中性的反馈方式。另一方面，为了避免员工过度愧疚或尴尬，损害其提升绩效的积极性，组织在给予员工负面绩效反馈时也会有所顾虑(马君, 闫嘉妮, 2020; 耿紫珍, 赵佳佳, 丁琳, 2020)。综合看来，负面绩效反馈的实施是组织管理实践中的一大难题(Kluger & DeNisi, 1996; Xing et al., 2021)。”

此外，我们在引言第二段介绍以往 AI 负面绩效反馈的研究时，也增加了对于归因理论视角的论述。“研究发现 AI 是基于数据驱动的程序语言，具备更低的主观意图，并且能够提供更为客观的信息（Lee 2018; Garvey, Kim, & Duhachek, 2023）。因此，相较于人类管理者，AI 提供负面绩效反馈会更加“对事不对人”，这可能会削弱传统人际互动中员工对负面绩效反馈的归因偏差(attribution bias)(例如，将负面绩效反馈归因为领导者的偏见)(Gardner, Karam, Tribble & Cogliser, 2019; Xing et al., 2023)，使得员工更多关注自身存在的不足(即进行内部归因)，并增强绩效改进的意愿。”

最后，在假设提出的 1.3 部分，我们也发现先前版本对于归因理论的回顾不够细致。对此，我们也进行了补充。

“归因理论(attribution theory; Heider, 1958)中的因果控制点(locus of causality)视角将个体的归因划分为内部与外部归因。内部归因强调个体倾向于从自身因素寻找原因，并且相信个人的成功抑或失败与个人因素(比如，个人能力或性格特征等)相关。而外部归因强调个体倾向于认为出现的某种行为或结果与其所处的情境或任务等外部因素有关。

Kelley(1967)指出，信息线索的一致性水平(consistency)会影响个体对于事件的内部与外部归因。具体来说，信息线索的一致性越高(比如经常上班迟到)，说明自身行为所导致的结果并非偶然，个体越会进行内部归因；反之，信息线索的一致性越低(比如偶尔迟到)，个体越倾向于将事件的发生归结于特殊状况，并对其进行外部归因。由于 AI 提供的反馈被认为可以减少人类决策过程的主观性和个体偏见，能够保证负面绩效反馈的结果在不同时间点、对不同的接收者都是标准化的，因此，相较于人类管理者，由 AI 提供的信息线索更具一致性(蒋路远等，2022；Raveendhran & Fast, 2021)，而这会增强员工对 AI 负面绩效反馈的内部归因。

此外，研究发现，经历失败、挫折等负性事件会增强个体的成就动机(achievement motivation)。特别是对负面绩效反馈内部归因后，由于绩效不善源于自身的努力不够或能力较低等内部因素，员工会提高绩效改进动机水平，以期达到更高的绩效来维护自尊水平(Weiner, 1985)。相反，当员工将绩效反馈结果归因于自身无法控制的外部原因时，可能会引发员工的不公平感等负面情绪，导致员工难以增强绩效改进动机(Harvey, Martinko & Douglas, 2006; Zuckerman, 1979)。因此，本研究认为，当员工将负面反馈归因于自身因素（即内部归因）而非外部情境因素（即外部归因）时，能够提高员工随后的绩效改进动机水平。”

主要参考文献（红色标记是指为深化归因理论指导性而新增的参考文献）：

Xing, L., Sun, J. M., Jepsen, D., et al. (2023). Supervisor negative feedback and employee motivation to learn: An attribution perspective. *Human Relations, 3*, 1−31.

Harvey, P., Madison, K., Martinko, M., Crook, T. R., & Crook, T. A. (2014). Attribution theory in the organizational sciences: The road traveled and the path ahead. *Academy of Management Perspectives, 28*(2), 128−146.

Harvey, P., & Martinko, M. J. (2009). An empirical examination of the role of attributions in psychological entitlement and its outcomes. *Journal of Organizational Behavior, 30,* 459−476.

Garvey, A. M., Kim, T. W., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing, 87*(1), 10−25.

Gardner, W. L. , Karam, E. P. , Tribble, L. L. , & Cogliser, C. C. (2019). The missing link? implications of internal, external, and relational attribution combinations for leader–member exchange, relationship work, self-work, and conflict. *Journal of Organizational Behavior, 40*(5), 554−569.

Kelley, H. H. (1967). Attribution theory in social psychology. In Levine, D. (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238).Lincoln, NE: University of Nebraska Press.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548−573.

意见 2：与问题 1 相关，在管理实践中，AI 和人类所给出的反馈（无论是正面还是负面反馈）更有可能存在内容和质量的差异，而不太可能在内容上保持一致（如绩效水平低于部门平均水平），仅仅是在形式上存在差异（如本研究实验中所操控的人类图片和机器人图片）。

现实中 AI 的反馈也可能由人类形象的给出。未来研究可以探索基于同样的反馈内容和形式，操控被试的来源感知（来自人类 vs. AI vs. 人机混合）进行比较。

回应：非常感谢您的这一意见。首先，正如您指出的，在管理实践中，AI 与人类管理者反馈很可能在内容或质量方面存在较大差异，我们非常认同您的这一观点。但作为一项先行研究，本研究试图先探讨在内容和质量保持一致的情况下，人类管理者与 AI 提供的负面反馈是否存在差异化影响。考虑到情境实验中的控制问题，我们尽量突出负面绩效反馈的本质特征，使得人类管理者组与 AI 组反馈的内容标准化，避免无关变量的干扰。当然，我们非常赞同您所说的，认为该问题可以在未来使用 ChatGPT 等更加接近真实反馈场景的研究中进一步完善。为此，我们在文章 5.4 研究不足与展望部分补充道：“由于科技的不断发展，越来越多的 AI 以人类的外在形象（例如虚拟员工）进入工作场所，与人类员工一起工作并提供绩效反馈 (e.g., 许丽颖,喻丰,彭凯平,王学辉, 2022)，未来研究可以在更为现实的人机反馈场景（例如，ChatGPT 反馈），操控被试的绩效反馈来源感知（来自人类 vs. AI vs. 人机混合），从而进行更具深度的人机影响效果比较”。

参考文献：

许丽颖,喻丰,彭凯平,王学辉.(2022).智慧时代的螺丝钉：机器人凸显对职场物化的影响. *心理科学进展, 30*(9),1905−1921.

意见 3：在论证假设 2 时，作者提到绩效任务类型可以分为创造型任务和纠错型任务，主观任务和客观任务。为什么本研究聚焦于主观任务和客观任务，而非创造型任务和纠错型任务？顺带，主观任务和客观任务的定义应该在这里就给出，而不是后面的脚注中。

回应：非常感谢您的意见。首先，Van Dijk 和 Kluger(2011)的研究主要是在传统绩效反馈场景下，关注反馈内容(正面反馈 vs. 负面反馈)与任务类型(创造型 vs. 纠错型)对员工绩效的交互作用。而本研究关注人机背景下的负面绩效反馈。由于既有研究指出，人类在情感、知觉以及经验等方面的能力通常优于 AI，而 AI 在逻辑、运算以及机械性操作方面比人更具优势(e.g., 蒋路远等, 2022)。鉴于人类与 AI 在主观任务与客观任务下的表现可能会有差异(Castelo, Bos & Lehmann, 2019; Newman, Fast & Harmon, 2020)，且相对于创造型/纠错型等任务类型，任务的主观性和客观性是更为典型的任务特征，我们选择了人机反馈场景下更为典型的任务特征(即主观任务与客观任务)作为边界条件研究的切入点。未来的研究，可以进一步深入探究其他任务特征的调节效应。其次，根据您的建议，在提出假设 2 时，我们直接给出了主观任务与客观任务的定义：在组织中，绩效任务通常有主观与客观之分(Van Dijk &

Kluger， 2011)。前者是基于个人观点或直觉的开放式或可解释的任务(比如，处理人际矛盾以及沟通等)；而后者则是可量化、可测量的事实型任务(比如，业绩分析、销量预测等)(e.g., Castelo, Bos & Lehmann, 2019)。

参考文献：

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.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent Algorithm aversion. *Journal of Marketing Research, 56* (5), 809−825.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. O*rganizational Behavior and Human Decision Processes, 160*, 149−167.

Van Dijk, D., & Kluger, A. N. (2011). Task type as a moderator of positive/negative feedback effects on motivation and performance: A regulatory focus perspective. *Journal of Organizational Behavior, 32*(8), 1084−1105.

意见 4：关于假设 2，作者提到，主观任务难以量化，结果是开放性和可解释性的。既然难以量化，又如何理解在一个主观任务中，一个员工的表现低于 82%的参与者这种量化的负面反馈？本研究聚焦于负面反馈，但似乎仅限于量化的负面反馈，未来研究可以探索主观任务的开放性的、可解释性的、针对具体问题的质性的负面反馈。

回应：非常感谢您的意见。首先，“主观任务”类似于主观论述题，的确很难像数学题目等“客观任务”那样有明确的评分标准，但是仍然能够进行评分，只是这种评分主观性较强，容易产生争议或分歧；第二，我们充分参考了以往研究对于负面绩效反馈的经典操纵方式(e.g., Belschak & Den Hartog, 2009; Kim & Kim, 2020)，即在被试完成实验任务(主观任务或客观任务)后，以统计排名的客观方式给予被试评价和反馈。这个排名实际上反映的是被试在同期参加测试群体中的相对表现(e.g., Kim & Kim, 2020)。因此，我们考虑即使主观任务难以量化，相对缺乏明确的标准，但仍可以对参与主观任务被试的表现进行主观评分和排名(类似于做一道论述题，虽然没有标准答案，但被试的回答有优劣之分)。这样操作负面反馈的好处在于，能够使得主观任务和客观任务组的被试接受相同的反馈内容，便于比较不同任务类型(主观 vs. 客观)和人机(人类管理者 vs. AI)负面绩效反馈对个体反应的差异化影响。

此外，您提到了一个非常关键的问题。即偏质性的反馈方式。事实上，绩效反馈有客观型反馈(objective feedback)与评价型反馈(evaluative feedback)之分(Johnson, 2013)。前者指采用定量评价指标进行反馈(类似于本研究采用统计排名的方式)，后者指利用“优良中差”等定性语言进行反馈。考虑到本研究关注的负面反馈本身比较敏感，容易引发被试消极情绪等额外因素，因此我们选择采用客观型反馈这种相对温和的方式。但是，正如您所说，未来研究仍可以采用开放性、质性且针对具体问题的评价式反馈，以探讨不同反馈方式(评价型 vs. 客观型)在人机反馈中对个体反应的影响差异。我们在文章 5.4 研究不足与展望中也对此进

行了补充: “以往研究指出，反馈的特征(例如反馈的频率、即时性、建设性等)是影响绩效反馈效果的重要因素。本研究关注以绩效表现排名为呈现方式的客观型反馈(objective feedback)。由于人类管理者与 AI 在沟通及情感属性方面的差异，未来研究可以探索人机在评价型反馈(evaluative feedback；比如开放性，质性且针对具体问题的反馈)情境中的影响差异(Johnson, 2013)”。

参考文献：

Belschak, F. D., & Den Hartog, D. N. (2009). Consequences of positive and negative feedback: The impact on emotions and extra-role behaviors. *Applied Psychology, 58*(2), 274−303.

Kim, Y. J., & Kim, J. (2020). Does negative feedback benefit (or harm) recipient creativity? The role of the direction of feedback flow. *Academy of Management Journal, 63*(2), 584−612.

Johnson D. A. (2013). A Component Analysis of the Impact of Evaluative and Objective Feedback on Performance. *Journal of Organizational Behavior Management, 33* (2), 89−103.

意见 5：在汇报实验 3 的结果时，“独立样本 t 检验发现，人类或 AI 提供负面绩效反馈在准确性水平上并无差异，$t(148)= 0.07$，$p = 0.94$。”——两个条件下负面反馈准确性水平的均值分别是多少？参与者认为反馈是大致准确的吗？

回应：非常感谢您的问题。通过独立样本 *t* 检验发现，人类管理者组负面反馈($M = 3.78$, $SD = 1.63$)与 AI($M = 3.80$, $SD = 1.60$)在准确性方面差异并不显著，$t(148) = 0.07$，$p = 0.94$。考虑到我们采用 7 点李克特量表测量构念，两个条件下的反馈准确性均值水平相对不是很高。由于本研究聚焦负面绩效反馈，而既有研究指出，反馈接收者普遍具有评价有利性的偏好(Favorability of Others' Ratings)(e.g., Ilgen & Hamstra, 1972; Podsakoff & Farh, 1989)。即出于自我增强(self-enhancement)和防御的目的，人们更加青睐他人对自身的高评价(Schrauger, 1975)。因此高评价会被个体认为更具准确性，反之低评价更容易被知觉为低准确性。据此，我们认为本研究发现的两种条件下的负面反馈准确性均值是较为合理的。此外，我们与传统视角下的绩效反馈准确性的研究进行了对标。比如，Brett 和 Atwater(2001)发现反馈接收者会知觉负面反馈具有较低的准确性，且来自管理者($r = -0.38$, $p < 0.001$)的绩效评分与员工的消极反应呈显著负相关，与反馈准确性则呈显著正相关($r = 0.31$, $p < 0.001$)。这侧面反映了负面反馈可能会导致较低水平的反馈准确性感知，本研究的反馈准确性均值是可以接受的。

参考文献：

Ilgen, D. R., & Hamstra, B. W. (1972). Performance satisfaction as a function of the difference between expected and reported performance at five levels of reported performance. *Organizational Behavior and Human Performance, 7*, 359−370.

Podsakoff, P. M., & Farh, J. L. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes, 44*(1), 45−67.

Schrauger, J. S. (1975). Responses to evaluation as a function of initial self-perceptions. *Psychological Bulletin, 82*, 581−596.

Brett, J. F., & Atwater, L. E. (2001). 360 °feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86*(5), 930−942.

意见 6：文章中提到“已有研究发现绩效反馈后，个体对人类或 AI 绩效反馈的准确性（e.g., Tong, Jia, Luo, & Fang, 2021）和公平感知（e.g., Newman, Fast, & Harmon, 2020）可能存在差异。”——如何解释本研究中并没有发现反馈准确性的差异？

回应：非常感谢您的这一问题。我们反思了实验三对于人类管理者或 AI 负面反馈的操纵。为保障人机负面绩效反馈的平衡性，防止带来无关变量干扰实验结果。我们对于人类管理者与 AI 反馈的可靠性做了平衡化处理(Tong, Jia, Luo & Fang, 2021)。比如，实验三对于将人类反馈者描述为“经过系统的职业能力测评培训，具有专业知识且经验丰富的测评专家”，与之对应的是，描述 AI 反馈者是“一种算法系统(该算法系统是基于测评专家设计的评价标准，由人工智能学者和计算机专家开发的用于完成能力测评任务的程序)”。我们推测，由于人类管理者与 AI 反馈均具备相对较高的专业性，而既有研究表明，来自专家的建议或推荐更容易得到个体的信赖和采纳(e.g., Bonaccio & Dalal, 2007; Tost, Gino & Larrick, 2012)。据此，在本研究中，个体对于人类管理者与 AI 负面绩效反馈在准确性方面的感知可能并无显著差异。

参考文献：

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.

Bonaccio, S. , & Dalal, R. S. (2007). Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences. *Organizational Behavior & Human Decision Processes, 101*(2), 127−151.

Tost, L. P. , Gino, F. , & Larrick, R. P. (2012). Power, competitiveness, and advice taking: why the powerful don't listen. *Organizational Behavior & Human Decision Processes, 117*(1), 53−65.

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.

Brett, J. F., & Atwater, L. E. (2001). 360 °feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86*(5), 930−942.

希望以上意见对研究提升有所帮助。

回应：非常感谢您给出的建议，这对我们提升研究水平与文章的质量非常重要。

……

审稿人 2 意见：

本文聚焦组织行为与人力资源管理领域的最新研究议题——“人机反馈”。选题方面文章具有较强的现实意义和理论前瞻性。文章采取 3 个递进式实验对理论模型进行检验，研究方法较为规范具有一定新意。研究发现有启发，但仍需进一步地进行理论升华，凸显理论贡献。总之，本文紧跟数智化时代背景，聚焦人工智能在绩效反馈、特别是负面反馈领域的应

用价值，探讨了员工的个体反应，既有一定现实意义，又有一定理论创新，但文章存在一些重要问题亟待完善（详见下文）。

回应：非常感谢您对于本研究选题、理论价值、现实意义和研究方法等方面的认可与肯定！同时也非常感谢您在耐心审阅之后，指出了本文的不足，还为我们提供了明确的修改思路和方向！希望修改后的论文能够很好地解决了您关注的问题。

意见 1：虽然题目可以直观反映研究内容，但题目过长且略显普通，未能吸引读者兴趣。建议修改为“当负面反馈来自 AI：归因视角下的员工个体反应”、“人机对比：负面反馈对员工绩效改进动机的差异化影响机制”等。

回应：非常感谢您的意见。我们借鉴您的建议，已将题目改为 “当负面绩效反馈来自 AI：归因视角下员工的绩效改进动机研究”。

意见 2：核心概念的界定问题。文章在问题提出部分使用的文献主要包含两个部分：一是与 AI 相关的研究；二是与算法(algorithm)相关的研究。那么作者需要注意 AI 与算法是否可以划等号？部分算法研究的文献用于支撑 AI 负面反馈是否合适？此外，AI 还包括 robots、生成式 AI 等，在使用一些关于 AI 的文献进行理论支撑时需要谨慎筛选。建议作者厘清 AI 和算法的关系，并向读者说明使用与算法研究相关的文献为什么合理。

回应：非常感谢您的意见，您的这一意见非常重要。首先，我们查阅了算法的相关文献。算法被定义为基于一种数据结构的计算程序(Moschovakis, 2012)。从广义上来说，算法可以分为脚本型(scripted-based)与机器学习型(machine-learning-based)(e.g., Lyytinen, Nickerson & King, 2020)。前者通常是单次运行的算法，比如常见的代数算法，图论算法。而后者基于 AI 原理，可以自动捕获人类行为的数字痕迹，并生成相关的参数(Tarafdar, Page & Marabelli, 2023)。

其次，我们回顾了 AI 的相关文献。AI 可以分为机器人式 AI(robotic AI；具有类人类外形，可以远程操控，并执行机械或社会属性任务)、虚拟式 AI(virtual AI；无物理外形，但具有虚拟身份。例如聊天机器人和虚拟人)以及嵌入式 AI(embedded AI；既不可见也无独特身份，通常嵌入计算机程序中。例如搜索引擎，GPS 地图等算法)(Glikson & Woolley, 2020)。综合算法与 AI 的文献来看，我们认为算法与 AI 是有交集的构念。而这个交点是机器学习型算法(如图 1)。

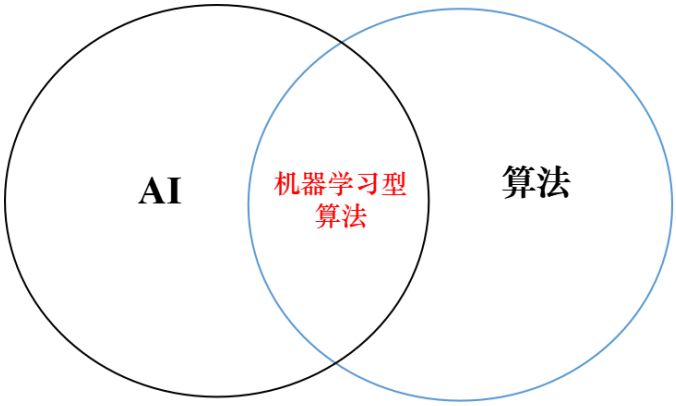

图 1 AI 与算法关系示意图

此外，我们也详细思考了算法与 AI 的关系。文献表明算法是 AI 技术的基础与驱动力。因为只有借助算法分析大量数据，AI 才能得以进化以及适应新环境，并实现自主决策的能力。比如，Tong 等将“AI feedback”详细描述为“一种基于深度学习与神经网络，拥有大量存档数据集，并接受过模型训练的算法程序”。借鉴 Tong 等(2021)的研究，本研究也将实验材料中的人工智能(AI)反馈具体描述为“一种算法系统(该算法系统是基于测评专家设计的评价标准，由人工智能学者和计算机专家开发的用于完成能力测评任务的程序)”。基于上述内容，我们在正文的第 2 页添加了脚注“广义上来说，算法可被分为脚本型(scripted-based)与机器学习型(machine-learning-based)(e.g., Lyytinen, Nickerson & King, 2021)。前者多属于数学概念，后者则兼备人工智能的原理。由于算法是 AI 技术的驱动力，因此本研究也关注机器学习类算法的文献”。

正如您指出的，虽然 AI 与算法联系紧密，但二者不能够等同。我们也进一步检查了文章引用的文献，保留了基于 AI 原理的算法类研究，从而确保了引证的严谨性。表 1 呈现的是我们对本研究主要引用的 6 篇基于 AI 原理的算法类文献的详细回顾。

表 1 本研究引用的 6 篇基于 AI 原理的算法类研究

作者及年份	发表期刊	对于算法的定义	主要研究结论
Castelo 等 (2019)	*Journal of Marketing Research*	基于 AI 原理的自然语言，可以从经验中学习，并模仿人类的行为。	在主观任务中，相对于算法，个体更偏好人类专家的建议。在客观任务中，上述效应发生反转。
Dietvorst 等 (2015)	*Journal of Experimental Psychology: General*	基于 AI 的原理，能够自动抓取数据，并为人类提供分析与预测。	相比于人类犯错，个体更不能容忍算法犯错，并因此产生算法厌恶

Newman 等(2020)	*Organizational Behavior and Human Decision Processes*	基于 AI 的原理，以大数据分析作为工作基础，并为组织提供自动化的决策。	相比于人类管理者，算法会简化决策过程中的一些关键信息，导致人们认为算法的决策更不公平。
Raveendhran 等(2021)	*Organizational Behavior and Human Decision Processes*	基于 AI 的原理，可以自动捕捉员工的行为数据，帮助管理者实时追踪员工的工作表现。	相比于管理者，AI 的绩效监控具有更低的主观意图，可以让员工感受到更多的工作自主性。
许丽颖等(2022)	心理学报	基于 AI 的工作原理，能够在医疗，金融，司法领域为人类提供关键的决策。	相比于人类歧视，算法具有更低的主观意志，因此人们对算法歧视具有更低的道德惩欲。
Yalcin 等(2022)	*Journal of Marketing Research*	基于 AI 的工作原理，能自动分析与识别顾客的喜好，并为他们提供购买推荐。	在有利性决策条件下，相比于算法，顾客更加偏好算法的反馈。而在不利性决策的条件下，上述效应发生了反转。

参考文献：

Moschovakis, Y. (2012). What is an algorithm?. *Mathematics Unlimited & Beyond, 55*(3), 919−936.

Lyytinen, K., Nickerson, J. V., & King, J. L. (2021). Metahuman systems = humans + machines that learn. Journal of Information Technology, 36(4), 427−445.

Tarafdar, M., Page, X., & Marabelli, M. (2023). Algorithms as co-workers: Human algorithm role interactions in algorithmic work. *Information Systems Journal, 33*(2), 232−267.

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

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.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent Algorithm aversion. *Journal of Marketing Research, 56*(5), 809−825.

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.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160,* 149−167.

Raveendhran, R., & Fast, N. J. (2021). Humans judge, algorithms nudge: The psychology of behavior tracking acceptance. *Organizational Behavior and Human Decision Processes, 164*, 11−26.

许丽颖,喻丰, 彭凯平.(2022).算法歧视比人类歧视引起更少道德惩罚欲. *心理学报, 9*,1076−1092.

Yalcin, G., Lim, S., Puntoni, S., et al. (2022). Thumbs up or down: Consumer reactions to decisions by algorithms versus humans. *Journal of Marketing Research, 59*(4), 696−717.

意见 3：前言第一段主要对负面绩效反馈的积极、消极效应进行了回顾，但没有凸显引入 AI 向员工提供负面反馈的必要性和重要性。建议作者围绕在工作场所中，领导者由上而下、面对面地向员工提供负面反馈存在哪些挑战（例如，在亚太地区的文化影响下，领导向下属提供负面反馈可能会导致下属觉得丢脸），这些传统的人际互动由新兴的人机互动进行替代是否带来了新的机遇。

回应：非常感谢您的意见。您的这一意见对于我们完善前言部分的写作非常具有帮助，我们结合既有传统文化方面的绩效反馈研究，进一步充实和改善了前言的写作。具体来说，我们在前言第一段中补充道“在中国的高语境文化下，一方面，出于“人情”与“面子”的顾虑，人们的沟通方式比较含蓄，管理者通常采用比较中性的反馈方式。另一方面，为了避免员工过度愧疚或尴尬，损害其提升绩效的积极性，组织在给予员工负面绩效反馈时也会有所顾虑(马君, 闫嘉妮, 2020; 耿紫珍, 赵佳佳, 丁琳, 2020)。综合看来，负面绩效反馈的实施是组织管理实践中的一大难题(Kluger & DeNisi, 1996；Xing et al., 2021)”此外，我们在前言第二段继续写道“随着智能化技术的持续发展，采用 AI 提供负面绩效反馈为组织带来了新的机遇……”。

主要参考文献：

耿紫珍, 赵佳佳, 丁琳. (2020). 中庸的智慧：上级发展性反馈影响员工创造力的机理研究. *南开管理评论 23*(1) ,75−86.

马君，闫嘉妮.(2020).正面反馈的盛名综合症效应：正向激励何以加剧绩效报酬对创造力的抑制？*管理世界,36*(1),105−121+237.

意见 4：虽然文章使用了机器启发视角，但不论在前言部分、还是在假设提出部分，作者都未对该视角进行详细地回顾和阐述，未能说明本文使用该视角的具体原因，建议作者对此问题进行完善和修改。此外，本文的标题强调了归因视角，但在问题提出部分又强调了机器启发视角，这不免使读者困惑。同时，既然本文使用了两个理论视角（机器启发和归因），那么作者还需要向读者说明这两个理论视角进行整合的合理性和必要性。

回应：非常感谢您的意见。我们起初对于研究的理论基础的思考的确不够深入。首先，我们详细回顾了机器启发理论(machine heuristic model; Sundar, 2008)的观点：它假定 AI 或机器比人类更安全、更值得信赖，即当个体认为与其交互的是一台机器而不是人类时，个体会自动

地启动关于机器的刻板印象，即个体会认为它是客观的、意识形态上无偏见的等，从而引发个体的反应(蒋路远, 曹李梅, 秦昕, 谭玲, 陈晨, 彭小斐, 2022)。此外，我们本轮修改更仔细地回顾了机器启发理论的实证型研究，比如，Araujo 等(2020)基于机器启发视角发现，当面对风险决策(例如，医疗、司法等)，人们受到 AI 客观性、无偏性的启发，会认为 AI 的决策质量比人类专家更公平，且更值得信任。此外，Helberger, Araujo & de Vreese (2020) 对 958 名参与者的问卷调查同样发现，受机器启发的影响，人们普遍认为 AI 的决策更公平。

总体来看，机器启发理论的观点更多地指向 AI 提供的信息和决策比人类更公平，或值得信任，这是 AI 研究的重要理论视角。但是这一理论并不是本研究的主要关注点，在前一版本中，本研究原来是想借鉴其某些论点进行理论推导。但是，经过您的点拨，我们反复思考后，认为这确实不是本研究关注的理论机制，且易导致读者对于本研究理论机制选择的困惑。因此，我们在修改版本中明确了归因理论对于本研究的指导性。从而避免理论视角的冲突和模糊。比如，我们在引言开头就提出了归因视角下，传统由管理者提供负面绩效反馈的不足：“出于自我服务偏差，人员收到负面绩效反馈后会更多地进行外部归因(例如，将收到负面绩效反馈的原因归结为领导的情绪不好)，而较少内部归因(例如，归因于自身的努力程度不够)(e.g., Harvey, Madison, Martinko, Crook & Crook, 2014)。员工对负面绩效反馈采取外部归因通常意味着其不会对低绩效的原因进行深入反思(e.g., Harvey & Martinko, 2009)”。此外，我们在引言第二段介绍以往 AI 负面绩效反馈的研究时，也增加了对于归因理论视角的论述：“那么，AI 代替人类管理者为员工提供负面绩效反馈有哪些潜在的优势？研究发现 AI 是基于数据驱动的程序语言，具备更低的主观意图，并且能够提供更为客观的信息(Lee 2018; Garvey, Kim & Duhachek, 2023)。因此，相较于人类管理者，AI 提供负面绩效反馈会更加“对事不对人”，这可能会削弱传统人际互动中员工对负面绩效反馈的归因偏差(attribution bias)(例如，将负面绩效反馈归因于领导者的偏见)(Gardner, Karam, Tribble & Cogliser, 2019; Xing et al., 2023)，使得员工更多关注自身存在的不足(即进行内部归因)，并增强绩效改进的意愿”。

主要参考文献：

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society, 35*(3), 611−623.

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review, 39*, 105456.

Gardner, W. L. , Karam, E. P. , Tribble, L. L. , & Cogliser, C. C. (2019). The missing link? implications of internal, external, and relational attribution combinations for leader–member exchange, relationship work, self-work, and conflict. *Journal of Organizational Behavior, 40*(5), 554−569.

Garvey, A. M., Kim, T. W., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing, 87*(1), 10−25.

Harvey, P., Madison, K., Martinko, M., Crook, T. R., & Crook, T. A. (2014). Attribution theory in the organizational sciences: The road traveled and the path ahead. *Academy of Management Perspectives, 28*(2), 128−146.

Sundar, S. S. (2008). *The MAIN model: A heuristic approach to understanding technology effects on credibility*. In M. J. Metzger & A. J. Flanagin (Eds.), Digital media, youth, and credibility (pp. 72-100). The MIT Press.

Thuillard, M., 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.

Xing, L., Sun, J. M., Jepsen, D., et al. (2023). Supervisor negative feedback and employee motivation to learn: An attribution perspective. *Human Relations, 3*, 1−31.

意见 5：在文章中，一些表述强调人类与 AI 的对比，但一些文献又强调领导与 AI 的对比，并且实验材料是强调领导负面反馈与 AI 负面反馈的对比，那么作者需要统一表述，因为人类互动还包括同事-员工、下属-上级、客户-员工等情况，这些情况与领导向员工提供负面反馈的情况并不相同。

回应：非常感谢您的意见。我们当初设计实验材料时也详细思考过这个问题，即如何介绍人类反馈者的身份？考虑到领导者(leader)或主管(supervisor)对下属而言权力过大，并且相比于 AI 可能更能影响员工薪酬与晋升，造成人机反馈的操纵不对等。因此，我们选择了相对中性的称谓：人类管理者(human manager)。比如，实验一中，人类管理者组的反馈者是“质量控制部门经理”，实验二为“人力资源部门测评专员”，实验三为“测评中心负责人”。这些称谓的共同之处在于，对个体不存在直接上级-下属关系，只是在特定情境下(比如本研究设置的测试，竞赛情境)为个体提供绩效反馈。为进一步保证措辞的严谨性，我们也回顾了 Tong 等(2021)的研究，发现他们也将人类反馈者描述为 human manager(即人类管理者)。据此，我们在全文中明确和统一了表述。

参考文献：

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.

意见 6：尽管绩效改进动机十分重要，但审稿人疑惑，作者为何不将结果变量设置为绩效（job performance）或学习行为（learning behavior）等代表个体行为的变量？这样更能体现负面反馈中人机对比的差异化效果。

回应：非常感谢您的意见。我们当时主要考虑了两点。首先，以往研究表明，负面反馈与绩效改进动机之间存在高相关性，比如，虽然负面绩效反馈可能挫伤个体的自我效能，导致员工降低绩效改进动机(Bandura, 1986; Belschak & Den Hartog, 2009)，但另有研究发现，负面

反馈也传达了实际表现与绩效期望的差距，出于维护自身能力信念和成就期望的目的，员工也会努力改进负面反馈后的绩效表现(Wigfield & Eccles, 2000)。

其次，正如您所说，绩效和学习行为都是绩效反馈非常重要的结果变量。但囿于条件所限，本研究开展的是线上行为实验。我们当时考虑，采用线上模拟或多轮任务的形式很可能难以精确评估绩效，学习表现这类比较客观的变量。其中的干扰因素也比较多(比如任务持续时间过长引发的疲劳效应；专家评定个人绩效的主观偏差因素)。此外，根据 Ilgen et al (1979) 的反馈过程模型。负面绩效反馈后，个体出于维持高水平效能感的目的，会提升绩效改进动机。据此，我们认为绩效改进动机可能是负面反馈后更加直接的结果变量。但是，您的意见也为我们指出了本研究的不足。我们在文章 5.4 研究不足与展望部分也补充道“最后，研究关注人机负面绩效反馈对员工绩效改进动机等近端结果的差异化影响。未来研究可以关注人机提供负面绩效反馈对员工实际行为表现（例如绩效水平，学习行为等）的影响，从而进一步拓展人机负面绩效反馈的影响效果研究”。

参考文献：

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive.* Englewood Cliffs, New Jersey: Prentice Hall.

Belschak, F. D., & Den Hartog, D. N. (2009). Consequences of positive and negative feedback: The impact on emotions and extra-role behaviors. *Applied Psychology, 58*(2), 274−303.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349−374.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68−81.

意见 7：在假设提出部分，文章的理论支撑较为薄弱，这源于作者未能思考清楚本文的理论框架究竟是什么。如前所述，作者使用机器启发模型和归因视角，但未将两个理论视角进行整合，所以导致作者提出的理论假设缺乏核心理论观点的支撑。例如，为什么 AI 提供的负面反馈一致、准确、无偏个体的绩效改进动机就越强？是否个体可能认为 AI 准确，实际反映了“我确实很差”，我可能就此放弃不愿改进？这是否取决于一定的边界条件。倘若无法解释，建议作者调整假设提出的逻辑和顺序，首先提出有调节的模型，即假设 1 涉及任务类型的调节作用。

回应：非常感谢您的意见。首先，针对您所指出的，负面绩效反馈后的内部归因是否也会使得员工“觉得自己很差，进而放弃改进”。我们认为很有启发性，并回顾了以往归因视角下的绩效反馈研究，发现内部归因有利于个体针对负面绩效反馈的绩效改进。比如，Harvey 和 Martinko’ s (2009) 的研究发现，人们出于自我防御的目的通常会对不良结果（例如负面绩效反馈）外部归因，但外部归因会使得个体忽视自身绩效表现的不足，并不利于绩效改进。相

反，对负面绩效反馈的内部归因则有利于员工认识自身不足。且对员工来说，将低绩效表现归结于自身，比归因于外部环境更加可控，因此对负面事件的内部归因更有利于绩效的改进。另外，Weiner(1986)将成就动机理论(achievement motivation theory)与归因理论相结合。并指出，由于人们都渴望取得成就，因此负面绩效反馈会激发员工内驱力以提升绩效表现。特别是对于内部归因的个体，由于低绩效表现与自身相关(例如能力差，不够努力等)，为满足获得成就的需求，员工会努力改进绩效，从而维护自我效能水平。据此，我们在假设 1.3 部分补充了上述内容，从而使得变量之间的关系更为明确与流畅。

其次，您的意见也使我们意识到了本研究的不足。比如，归因理论的视角庞杂，个体归因的类型多样。比如，本研究聚焦于归因理论中的因果控制点视角(locus of causality)，即将员工的归因方式区分为内部和外部(internal and external attribution)。事实上，按照归因的稳定性水平，个体的归因还可被划分为能力归因(ability attribution；即把事件的结果归因于自身的能力)与努力归因(effort attribution; 即归因于自身的努力或投入程度)。以及按照归因的可控性，个体可能会将事件结果归结为可控因素(能力，努力等)，抑或不可控因素(运气，任务难度等)(Weiner, 1986; Harvey et al., 2014)。因此，我们在文章的 5.4 研究展望部分补充道“本研究聚焦于归因理论中的因果控制点视角，即将员工的归因方式区分为内部归因和外部归因。事实上，归因理论的内涵十分丰富。比如，按照归因的稳定性水平，个体的归因可被划分为能力归因(ability attribution；即把事件的结果归因于自身的能力)与努力归因(effort attribution；即归因于自身的努力或投入程度)。按照归因的可控性，个体可能会将事件结果归结为可控因素(能力、努力等)，抑或不可控因素(运气、任务难度等)(Weiner, 1986; Harvey et al., 2014)。由于 AI 能够利用大数据对个体的态度，行为意愿等进行“画像性”分析，因此 AI 对人们性格或能力等个人稳定性因素的分析和预测也更为精准(e.g., Fan 等，2023)。未来研究可以基于更多归因理论的视角，或将多个视角进行结合(例如，AI 提供的负面绩效反馈能否提升员工对于能力的内部归因，并影响绩效改进动机)，从而进一步丰富归因理论对人机绩效反馈产生差异化效果的解释性。

主要参考文献：

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.

Harvey, P., & Martinko, M. J. (2009). An empirical examination of the role of attributions in psychological entitlement and its outcomes. *Journal of Organizational Behavior, 30*, 459−476.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *The Journal of Applied Psychology, 108*(8), 1277−1299.

意见 8：如上所述，如果作者同意调整假设提出的顺序和逻辑，在实验部分建议重新调整或修改部分实验内容。

回应：非常感谢您的这一意见。我们在文章前言部分，以及假设提出部分进一步明确了归因理论的观点。进而增强了变量之间的逻辑关系。

意见 9：实验分析建议增加操纵检验。

回应：非常感谢您的这一意见。我们在三个实验的研究结果部分均增加了操纵检验。

意见 10：实验 2 将场景切换为大学生熟知的学业成绩反馈是否合适？因为作者提到实验 2 是在实验 1 的基础上检验人机负面绩效反馈×任务类型是否会交互影响员工的绩效改进动机。建议作者使用其它实验方法或将实验样本切换为 MBA 学员。

回应：非常感谢您的意见。本研究采用大学生被试主要有两点考虑：首先，以往绩效反馈的研究也多采用学生被试作为研究对象(e.g., Kluger, Lewinsohn & Aiello, 1994; Lam et al., 2011)，因为对于学生群体而言，成绩就是代表其绩效的重要指标之一。采用学习成绩反馈的情境可以一定程度上模拟真实的绩效反馈过程。

但正如您所说，由于实验 1 与 2 的递进关系，实验 2 直接采用大学生被试验证假设的考虑的确欠妥。为此，我们补充了一个实验，并替换了原来的大学生实验作为本研究的实验 2。实验 2 聚焦于组织的绩效反馈情境。采用 2(人类管理者负面绩效反馈 vs. AI 负面绩效反馈)×2(主观任务 vs. 客观任务)的实验设计。并参考实验 4 的做法，发布实验信息，以定向招募不同行业(比如，制造、软件、金融、教育及快消品等行业)与岗位(比如，管理生产运营、技术研发、市场营销与产品设计)的在职员工被试参与研究。最终招募有效的被试 160 名。实验结果也支持了本研究的假设。

参考文献：

Kluger, A. N., Lewinsohn, S., & Aiello, J. R. (1994). The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal. *Organizational Behavior and Human Decision Processes, 60*(2), 276−299.

Lam, C. F., DeRue, D. S., Karam, E. P., et al. (2011). The impact of feedback frequency on learning and task performance: Challenging the "More is better" assumption. O*rganizational Behavior and Human Decision Processes, 116*(2), 217−228.

Yalcin, G., Lim, S., Puntoni, S., et al. (2022). Thumbs up or down: Consumer reactions to decisions by algorithms versus humans. *Journal of Marketing Research, 59*(4), 696−717.

意见 11：理论贡献部分建议作者增加讨论本研究发现对人机反馈领域的贡献和突破，核心文献包括 Tong 等、Luo 等、Yam 等人的研究发现。此外，建议作者在厘清算法和 AI 核心概念区别的基础上，对理论贡献的第二点进行修改，重点关注对 AI 领域文献的对话。最后，建议作者思考机器启发模型、归因理论两个理论视角的关系，并结合本研究发现对两个理论视角的贡献进行阐述。

回应：非常感谢您的这一意见。我们重新梳理了本研究的理论贡献。在明确归因理论作为本研究的理论视角的基础上，侧重于与传统负面绩效反馈研究，人机反馈研究，数智化场景下归因理论的应用三个研究领域进行了理论贡献的总结。并重点与您提到的 AI 领域文献进行了对话。以下内容是我们对于理论贡献部分的修改。

首先，本研究丰富了负面绩效反馈的研究。传统基于人际互动(比如，领导等)的负面绩效反馈不利于员工的情绪，甚至负面影响其绩效表现(Belschak & Den Hartog, 2009)。既有研究也从不同角度探索了提升负面绩效反馈实施效果的途径。例如，反馈特征层面(负面绩效反馈的即时性以及建设性等)(e.g., Kuvaas, Buch, & Dysvik, 2017)，员工个体层面(对负面绩效反馈的积极归因，核心自我评价等)(马璐,谢鹏,韦依依, 乔小涛，2021；Xing et al., 2023)。而本研究则从人机负面绩效反馈这个新兴的视角出发，发现 AI 替代人类管理者提供的负面绩效反馈有利于员工后续的绩效改进动机，并挖掘出主观和客观任务这一重要的边界条件，为负面绩效反馈研究领域的人机差异化效果提供了证据。

其次，本研究丰富了人机反馈的研究。当前，为数不多的人机反馈研究基于不同的反馈特征以及理论视角发现了不尽相同的研究结论。一方面，基于算法欣赏的视角，研究者发现 AI 能够提升绩效反馈的精确性，从而提升员工的绩效水平(Tong et al., 2021)。但另一方面，也有研究从算法厌恶角度出发，发现 AI 缺乏真诚性与独特性，因此当组织披露绩效反馈(尤其是带有鼓励，赞扬性质的正面反馈)来源于 AI 时(Yalcin et al., 2022)，会降低个体的积极表现(Tong et al., 2021; Luo et al., 2019)。而本研究聚焦于管理实践中开展难度较大的负面绩效反馈，并验证了 AI 提供负面绩效反馈对员工绩效改进动机的积极影响。此外，以往研究显示，个体对于 AI 绩效反馈的欣赏抑或厌恶并不是绝对的，因此探索其中的边界条件十分重要。比如，Tong 等(2021)发现，对于任期较长的员工而言，由于他们与组织建立了更强的情感纽带，对于组织采用 AI 提供绩效反馈的变革也更为支持。因此员工的任期会缓解 AI 提供绩效反馈的负面效果。此外，Luo 等(2019)发现，顾客对于 AI 的了解程度会降低个体对于 AI 的刻板印象(例如缺乏知识和同理心)，从而缓解由 AI 提供反馈造成的产品销量下降。本研究则发现了人机负面绩效反馈与任务类型对员工绩效改进动机的交互作用(具体来说，在主观任务中，相较于 AI，人类管理者提供的负面绩效反馈导致更高的绩效改进动机。在客观任务中，上述结果发生反转)，从而丰富了人机反馈边界条件的研究。

最后，本研究基于归因理论为解释 AI 和人类管理者在提供负面绩效反馈过程中产生差

异化的结果提供了一个有效的解释机制，丰富了内部归因理论的相关研究。归因理论常应用于社会心理学或组织管理的传统场景中(e.g., Tolli & Schmidt, 2008)，用于解释个体对事件发生后的成因归结(e.g., Xing et al., 2023)。本研究将传统归因理论的观点(比如，高一致性的信息线索有利于个体进行内部归因)(e.g., Kelly, 1967)，与当前人机研究的发现(比如，相较于人类，AI 能够提供一致性更高的信息)(e.g., 蒋路远等，2022)进行结合，从而解释了人机负面绩效反馈的差异化影响，拓展了归因理论在人机反馈情境中对于员工动机或行为的解释。此外，AI 对人类行为影响的机制研究一直以来备受关注，以往研究更多从信任(Glikson & Woolley, 2020)、公平感(Newman et al., 2020)等角度解释人机对比效应的机理。而本研究则是从内部归因视角为人机提供负面绩效反馈的差异化效果提供了一个较新的解释。

意见 12：文章对局限和展望讨论的不够深刻，特别是负面反馈中人机对比的理论视角、内在机制、边界条件、文化情境等因素未进行讨论和展望。建议进行拓展和完善。

回应：非常感谢您的这一意见。我们重新反思了本研究存在的不足。并结合两位审稿专家之前为我们提出的宝贵意见，对研究局限与展望部分进行了归纳。具体不足和展望如下：

本研究也存在一些局限。首先，未来 AI 可能以人类的外在形象（例如虚拟员工）进入工作场所，与人类员工共事并提供绩效反馈 (e.g., 许丽颖,喻丰,彭凯平,王学辉, 2022)。未来研究可以在更为现实的人机反馈场景（例如，虚拟同事的反馈），操控被试的绩效反馈来源感知（来自人类 vs. AI vs. 人机混合），从而进行更具深度的人机影响效果比较。其次，以往研究指出，反馈的特征（例如反馈的频率，即时性，建设性等）是影响绩效反馈效果的重要因素。本研究关注以绩效表现排名为呈现方式的客观型反馈（objective feedback）。由于人类管理者与 AI 在沟通及情感属性方面的差异，未来研究可以探索人机在评价型反馈（evaluative feedback；比如开放性，质性且针对具体问题的反馈）情境中的影响差异（Johnson, 2013）。

另外，未来研究可以更多关注人机绩效反馈中的文化因素。比如，在中国传统的中庸与谦和文化的影响下，为提升负面绩效反馈的实施效果，管理者往往采用“三明治”式的反馈形式，即在负面的反馈中夹杂鼓励与赞扬性质的正面反馈。由于 AI 常常因缺乏“人情味”或同理心而遭到个体的厌恶(e.g., Dietvorst, Simmons & Massey, 2015; Luo, Tong, Fang & Qu, 2019)，具备类人性化特征的 AI 往往更能够被人们接受(许丽颖, 喻丰, 彭凯平, 2022; Yalcin et al., 2022)。未来研究可以探索 AI 采用“三明治”式的绩效反馈形式是否会对员工产生更为积极的影响，从而探索传统文化因素在 AI 提供绩效反馈过程中的作用。

再者，本研究聚焦于归因理论中的因果控制点视角，即将员工的归因方式区分为内部归因和外部归因。事实上，归因理论的内涵十分丰富。比如，按照归因的稳定性水平，个体的归因可被划分为能力归因(ability attribution；即把事件的结果归因于自身的能力)与努力归因(effort attribution；即归因于自身的努力或投入程度)。按照归因的可控性，个体可能会将事件结果归结为可控因素(能力、努力等)，抑或不可控因素(运气、任务难度等)(Weiner, 1986;

Harvey et al., 2014)。由于 AI 能够利用大数据对个体的态度，行为意愿等进行“画像性”分析，因此 AI 对人们性格等个人稳定性因素的分析和预测也更为精准(e.g., Fan 等，2023)。未来研究可以基于更多归因理论的视角，或将多个视角进行结合(例如，AI 提供的负面绩效反馈能否提升员工对于能力的内部归因，并影响绩效改进动机)，从而进一步丰富归因理论对人机绩效反馈产生差异化效果的解释性。

最后，研究关注人机负面绩效反馈对员工绩效改进动机等近端结果的差异化影响。未来研究可以关注人机提供负面绩效反馈对员工实际行为表现（例如绩效水平，学习行为等）的影响，从而进一步拓展人机负面绩效反馈的影响效果研究。

主要参考文献：

Luo, Tong, Fang, Qu, & Zheng. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases.*Marketing Science, 38*(6), 937−947.

Granulo, Fuchs, Puntoni, & Stefano.(2021).Preference for Human (vs. Robotic) Labor is Stronger in Symbolic Consumption Contexts. *Journal of Consumer Psychology, 31*(1), 72−80.

Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology, 108*(8), 1277−1299.

意见 13：文章仍存在一些错字、漏字等细节问题，请作者注意修改，在此不一一列举。

回应：非常感谢您的这一意见。我们仔细通读了全文，对文章中表述的疏漏，不清晰之处，以及错别字，漏字问题进行了修改与完善。比如，文章中“调节作用与交互作用”、“人机负面绩效反馈与人-AI 作为反馈源”的混用，我们统一将本研究的自变量描述为“人机负面绩效反馈”，并对其进行了界定。此外，鉴于本研究属于实验类研究，我们统一表述为“人机负面绩效反馈与任务类型对员工绩效改进动机的交互作用”。

第二轮

审稿人 1 意见：

本研究关注来自 AI 和人类管理者的负面反馈对员工绩效改进动机的影响，以及任务类型的调节作用和内部归因的中介作用。论文在理论和方法上都有了进一步的改善，仍有以下问题值得思考：

回应：衷心感谢您给出上轮修改的肯定！

意见 1: 在研究中需要关注反馈质量因素（如准确性、具体性、针对性等因素。）

对于负面反馈是否可以增进绩效改进动机，很重要的一点，或者可以说最重要的一点，在于负面反馈本身的质量，如是否准确、是否有针对性、是否详细具体等。有质量的负面反馈才有可能“有意义地”增进绩效改进动机。假如负面反馈本身是不准确的、低质量的或者模糊的，员工将在错误的方向上继续努力，或者无所适从。

因此不能仅仅对负面反馈的来源——人 vs. 机进行对比来研究对绩效改进动机的影响，而是需要考虑负面反馈质量的因素。

尤其是，在本研究中，在主观或者客观任务的表现上，被试是有绩效表现高低的区分的。在研究中一律给予“低于 80%（或者 82%）同事”的负面反馈，被试的感受会有所不同。那些任务绩效表现高的被试，可能怀疑负面反馈的准确性，因此无法将负面反馈转化为绩效改进动机。值得注意的是，研究 3 中也汇报，被试认为绩效反馈的准确性平均分在 3.8 左右（7 点评分），SD = 1.6 左右。因此总的来说，被试认为反馈的准确性一般，而且存在较大被试间差异。——一个被认为不准确的反馈，如何能够引发内归因呢？

回应：非常感谢您的这一意见。我们十分赞同您所指出的，即负面绩效反馈的质量、准确性以及具体性会影响反馈接收者对反馈的归因和绩效改进动机。首先，想跟您解释的是，既有关于绩效反馈的操纵方式大体上分为两种：虚假反馈(e.g., Kim & Kim, 2020)(详见表 1)与真实反馈(e.g., Goodman & Wood, 2004)(详见表 2)。虚假反馈的特点是给予被试相同内容的绩效反馈(本研究实验 1~3 参考了经典的虚假反馈操纵，这种操作方法已比较成熟且被广泛采用)。其优点在于，在探索研究变量间的关系(如本研究的人机负面绩效反馈对绩效改进动机的影响)时，能够控制反馈内容或信息的一致，从而便于关注研究变量之间的关系。因此，该操纵方法比较适用于实验室或情景实验研究。此外，在实验 1~3 中，负面绩效反馈展现的是个体的相对绩效水平(relative performance)(比如实验 1 中被试会收到“你的工作表现低于部门平均水平，你现在是部门绩效表现较低的员工之一，希望你能持续改进”；实验 2 与 3 中被试收到“在本次测试中，你的表现低于 80%的同事，位于后 20%，表现有待提升”)。这样做的好处在于能够减少被试对于负面绩效反馈的怀疑或不信任感(e.g., Kim & Kim, 2020)。因为不同于绝对绩效水平(absolute performance)，相对绩效水平反映了个体在团队或集体中绩效的相对水平。简单来说，即使在实验任务中绝对绩效表现较好的被试，也可能因为其他被试的绝对绩效水平更好而收到相对的负面绩效反馈(通俗来讲，人外有人)。考虑到相对于绝对绩效水平，个体自身更难评估相对绩效水平(Cianci, Klein, & Seijts, 2010)，并且既有研究显示，相对绩效评估也能够减少个体的情绪反应并提升绩效(e.g.,Gjedrem,William,Gilje,Kvaloy,& Ola,2020)。因此，本文的实验 1~3，采用这种方式尽可能提高被试对反馈信息质量的感知。

表 1 本研究参考虚假反馈操纵的文献来源

作者及年份	发表期刊	对负面绩效反馈的操纵	主要研究结论
Cianci, Klein, & Seijts (2010)	*Journal of Applied Psychology*	“参与者在此项任务中的平均得分为 95 分，而你的得分四舍五入后为 60 分”	尽责性与任务目标交互影响负面绩效反馈后个体的紧张感与绩效水平，相对于学习性目标，当高尽责性个体怀揣绩效目标时，其负面绩效反馈后的绩效水平较低。
Belschak & Den Hartog (2009)	*Applied Psychology*	“目前你的工作表现低于部门平均水平，你现在是部门绩效表现较低的员工之一，公司希望你能持续改进”（注：本研究实验 1 参考）	个体接收负面绩效反馈后会产生负面的情绪，从而导致更高水平的反生产行为，离职意愿，以及更低水平的组织公民行为和组织承诺。
Kim & Kim(2020)	*Academy of Management Journal*	“在所有参与任务的小组中，你和你搭档所组成的小组在创造力方面的评分位于后 20%”（注：本研究实验 2~3 参考）	负面绩效反馈的指向性影响其效果。当由下属指向领导时，人员更加关注提升绩效的策略；而当领导指向下属，或发生在同事之间时，人员更加关注负面绩效反馈的内容。

与虚假反馈这种反馈的操作化方式不同，真实反馈基于被试客观的任务表现，其特点是反馈内容更加具体、准确以及个性化(e.g.,Goodman & Wood, 2004)，不仅包含结果性的反馈(例如，告知被试任务表现水平及排名)，还包括过程性的反馈(例如，指出参与者任务表现中的优缺点)。

表 2 操纵真实反馈的代表性研究

作者及年份	发表期刊	对负面绩效反馈的操纵	主要研究结论
Goodman & Wood, 2004	*Journal of Applied Psychology*	（举例）“你在任务过程中给予了杰克错误的任务分配，但好在你给予了尼尔正确的任务奖励”（本研究补充的实验 4 参考）	高具体的反馈有利于个体从高绩效表现中学习，但并不利于个体从低绩效表现中汲取经验。

总结而言，采用虚假反馈的策略操纵负面绩效反馈是目前比较常见且成熟的方法。但我们认可并非常重视您为本研究提出的有关反馈质量、准确性或具体性不足的意见。特别是，实验 3 被试知觉到的准确性水平的确一般。因此我们参考 Goodman & Wood (2004)的做法，补充了实验 4，用改进的虚假反馈（以看起来更加真实的反馈形式改善了负面反馈的操作），以重复先前实验的结果。并进一步提升研究整体的效度(详细内容请见正文的实验 4 部分 P54-59)。修改如下：

简单来说，实验 4 为被试提供了相对真实的负面绩效反馈(依据被试真实的任务表现给予其具体和准确的反馈)。实验 4 采用了线上实验室实验的方法，在正式实验开始前，首先对 5 名主试人员进行培训，使他们充分了解，并熟练掌握任务内容以及评价标准。实验 4 分为任务阶段与反馈阶段。任务阶段要求被试完成两个主观或客观任务(与实验 3 类似)。反馈阶段由主试扮演反馈者(人类管理者或 AI)，通过 SalesSmartly(一个专业的企业-员工实时聊天交互网站，https://app.salesmartly.com/)为参与者提供几乎实时的绩效反馈。反馈的内容方面，参考 Goodman & Wood(2004)的做法，为个体提供具体的、过程性质的反馈。比如，向被试解释测试的目的，为其分析任务表现中的优缺点，并且所有负面绩效反馈统一包含 4 个方面的内容：1. 依据被试的作答整体情况给予其总评(见表 3 的绿色字体)。由于我们要求被试对每个任务的作答字数不得少于 100 字，因此总评是基于个体回答的详细性，清晰性，逻辑性三个方面给出；2. 解释任务题目的目的以及所考察的能力(见表 3 的红色字体)。相同任务组别的个体会收到一致的任务目的解释，这样做的目的在于让被试了解任务的内涵从而提升反馈的质量；3. 分析参与者在任务表现中的优缺点并给出相应评分(见表 3 的蓝色字体部分)。我们在实验开始前，对 5 名主试进行了培训，重点在于使各位主试熟练地掌握每个题目的作答要点。考虑到在作答过程中，被试的作答普遍比较认真，并且他们会不同程度地回答到作答要点。因此我们选择被试作答质量相对较好的题目(即回答了更多的作答要点)进行相对正面的反馈，而作答相对较差(回答较少或没有回答作答要点)的题目进行负面反馈。选择负面夹杂正面反馈策略的好处是，使得反馈内容更加真实，客观并增加可信度(为确保严谨性，我们也对反馈信息的操纵进行了检验)；4. 在上述内容的铺垫下给出被试获得的总分并提供负面绩效反馈(见表 3 的橙色字体部分)。不同于前三点均基于被试真实表现作出的

评价，第四点给予被试的总分以及负面反馈是虚假且在被试间保持一致的。这样做可以加强对负面绩效反馈的控制，使反馈内容保持更好的组间一致性。同时，为控制反馈内容，所有负面反馈内容的字数统一控制在 200 字左右。

总的来说，借鉴 Goodman & Wood (2004)的方法，实验 4 提升了负面绩效反馈操作的质量，一方面给个体更加具体、个性化的反馈(比如，基于真实作答为个体清晰地指出任务表现的优缺点)，另一方面参考虚假反馈的经典做法，一定程度上加强了对反馈内容的控制(比如，统一给予被试假的总分以及作答排名的负面绩效反馈，并保证每条反馈的总字数在 200 左右)。提升反馈质量的同时也考虑到对反馈信息的控制。以下我们随机抽取了 4 个组别中(人机反馈×任务类型)为被试提供的反馈信息作为例子（黑色斜体字部分展示了人类与 AI 反馈操作的不同点，除最后感谢被试的话术略有不同外，两组参考的反馈形式是相同的）。

表 3 实验 4 操纵真实反馈的例子

组别	负面绩效反馈内容
主观任务	【人类管理者组】亲爱的 A035 参与者，感谢您完成本次管理能力测评的试题。整体上，您回答的内容比较详实，条理清晰。本次测试的第一封公文旨在考察您在建设团队中的矛盾化解能力。第二封意在测试您应对团队突发事件的问题解决能力。相较于同期的参与者，您表现出了较好的矛盾化解能力【69.75/100】（您分点作答，且方案有效可行性高）。但您表现出的突发事件解决能力较弱，得分为【54.15/100】（您未谈及寻求多部门协作这个关键点）。总的来说，您在本次测试中的总分为【61.95/100】，低于 82%的同事，位于后 18%，表现有待提升。 ***请您阅读反馈后进行下一轮测试，感谢您的配合！*** 【AI 组】亲爱的 A037 参与者，感谢您完成本次管理能力测评的试题。整体上，您回答的内容比较简略。本次测试的第一封公文旨在考察您在建设团队中的矛盾化解能力。第二封意在测试您应对团队突发事件的问题解决能力。相较于同期的参与者，您表现出了较弱的矛盾化解能力【54.15/100】（您仅谈及了促进团队成员沟通，未提到重新制定工作计划或目标）。但您表现出的突发事件解决能力较好，得分为【69.75/100】（您分点作答，且给出的方案可行性较好）。总的来说，您在本次测试中的总分为【61.95/100】，低于 82%的同事，位于后 18%，表现有待提升。 ***请您阅读反馈后进行下一轮测试，感谢使用小 ai！***
客观任务	【人类管理者组】亲爱的 A066 参与者，感谢您完成本次管理能力测评的试题。整体上，您回答的内容比较详实，条理清晰。本次测试的第一封公文旨在考察您在原料采购中的运算分析能力。第二封意在测试您预测销售量的逻辑推理能力。相较于同期的参与者，您表现出了较好的运算分析能力【69.75/100】(利用加权法推算，排序应为 C>A>D>B，您的回答基本正确)。但您表现出的逻辑推理能力比较欠缺，得分为【54.15/100】（第 11 个月的状态应为畅销，销量大于 100）。总的来说，您在本次测试中的总分为【61.95/100】，低于 82%的同事，位于后 18%，表现有待提升。 ***请您阅读反馈后进行下一轮测试，感谢您的配合！***

	【AI 组】亲爱的 A074 参与者，感谢您完成本次管理能力测评的试题。整体上，您回答的思路明确，表达流畅。本次测试的第一封公文旨在考察您在原料采购中的运算分析能力。第二封意在测试您预测销售量的逻辑推理能力。相较于同期的参与者，您表现出了较低的运算分析能力【54.15/100】（利用加权法推算，正确排序应为 C>A>D>B）。但您表现出的逻辑推理能力比较好，得分为【69.75/100】（您对第 11 个月的销量状态推理基本正确）。总的来说，您在本次测试中的总分为【61.95/100】，低于 82%的同事，位于后 18%，表现有待提升。 ***请您阅读反馈后进行下一轮测试，感谢使用小 ai！***

相较于实验 1~3，我们认为实验 4 的反馈内容在质量，准确性以及具体性方面有了明显的进步。主要表现在(1) 实验 4 充分考虑了被试真实的作答情况。现实作答过程中，参与者的回答均存在优缺点(虽然客观任务存在明确的标准答案，但被试展示解题思路的严谨性与详实性是不同的，这也为提供负面反馈增加了合理性)，我们详细为其一一指出，这样能够提升反馈的总体质量，并且实验 4 的反馈信息更加具体和真实。值得注意的是，实验 4 同样测量了反馈准确性(7 点评分)，相较于实验 3 中被试知觉到的反馈准确性(M = 3.79, SD = 1.61)，实验 4 中被试感知的反馈准确性得到明显的提升(M = 5.41, SD = 0.95)。（2）实验 4 也考虑到对于反馈内容的控制，因为依据真实表现的反馈可能带来个性化的问题，导致反馈内容在被试间不完全一致。为了缓解这个问题，我们首先在被试间采取了相同的反馈形式(即向被试解释任务目的，指出作答的优点与缺点，最后提供负面绩效反馈的总评)。其次，我们严格控制了反馈内容的字数(200 字左右)。在为参与者提供更具体和准确的反馈的同时，尽可能保证反馈内容的标准化。（3）实验 4 也考量到操纵负面绩效反馈的严谨性，如前文所述，考虑到现实中被试的作答同时存在优缺点，为提高反馈质量和准确性，我们采取正面反馈结合负面反馈以及最终给予被试负面反馈总评的策略。为检验实验 4 对负面绩效反馈的操纵效果，我们增加了对反馈内容的操作检验“请问您收到的反馈对您在测评中的表现的评价是 (1 = 很负面，5 = 很正面)。结果显示，个体对反馈内容感知的均值为 2.01(SD = 0.97)，说明实验 4 对于负面绩效反馈的操纵是成功的。

以上是补充实验 4 的基本内容，希望能对您指出的问题有所回应。由于增加了一个实验，我们也对文章的整体逻辑进行了调整和完善。详细内容请见正文中的蓝色字体部分。

参考文献：

Gjedrem, William, Gilje,Kvaloy, &, Ola. (2020). Relative performance feedback to teams. *Labour economics,66*, 166−184.

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology, 95*(4), 618−630.

Goodman, J.S., & Wood, R.E..(2004).Feedback specificity, learning opportunities, and learning.*Journal of Applied Psychology,89*(5),809−821.

Kim, Y. J., & Kim, J. (2020). Does negative feedback benefit (or harm) recipient creativity? The role of the direction of feedback flow. *Academy of Management Journal, 63*(2), 584−612.

意见 2: 在文章 1.3 部分对内部和外部归因的中介作用进行论述时，作者采用了归因理论的“线索一致性水平”进行论述。这个论述并不适用于本研究。线索一致性是需要出现多次线索进行比较，才有一致性的高低。本研究中的负面反馈均为一次性的，即一次性给出负面评价（“低于平均表现水平”），这里的线索一致性体现在哪里呢？被试如何感知到线索的一致性？

作者还提到：在主观任务中，来自 AI 的负面绩效反馈信息一致性水平更低。然而在本研究的主观任务中，AI 也跟人类管理者同样给出了“低于 80%（或者 82%）同事”的负面反馈，为什么 AI 的负面绩效反馈信息线索一致性更低？

只有在多次给出负面反馈的情境中，才有可能讨论线索的一致性水平。例如 AI 每次给的反馈结果一致，或者，如同作者假设，在主观任务中，AI 每次给的反馈结果不一致，（虽然未必）。因此，本研究不适合“线索一致性水平”的内外归因论述。

回应：非常感谢您的这一问题。经过反思，由于本研究的实验中并不包含对负面绩效反馈的重复操纵，因此，我们认同您指出的“线索一致性不适用于本研究的假设推导”。经过对归因理论的反复推敲，我们采取归因理论结合人类与 AI 不同特征的策略展开中介机制的推理。我们修改了部分内容(见正文 P42-43)：

面对负面绩效反馈，个体会识别反馈提供者的行为意图，从而选择内部或外部归因(e.g., Audia & Locke, 2003)。具体来说，当感知负面绩效反馈出于管理者的恶意时(例如，打压，伤害)，个体会对负面反馈进行外部归因，从而降低学习动机与绩效改进。相反，当负面绩效反馈传递出管理者帮助员工改善绩效的动机时，个体会更多地内部归因，进而提升绩效水平(Xing et al., 2023; Ni & Zheng, 2024)。由于 AI 依赖客观数据进行信息输出，导致 AI 具备更少的主观意图(e.g., Garvey, Kim & Duhachek, 2023)。因此，相比于人类，AI 或算法在负面情境中输出的决策具有更低的蓄意性与伤害性，也更容易被接受。比如，相比于人类歧视，个体认为算法歧视具有更低的自由意志，因此对其道德惩罚欲更少(许丽颖, 喻丰, 彭凯平, 2022)。再如，面对高于预期的价格，个体认为 AI(较人类)的出价具备较低的主观意图并更愿意接受(Garvey, Kim & Duhachek, 2023)。总结来看，相比于人类管理者，AI 基于数据与事实的信息输出特点使负面绩效反馈更加客观，且具有更少的伤害意图(e.g.,Tong, Jia, Luo, & Fang, 2021)，因此，来自 AI 的负面绩效反馈会使个体更多关注绩效改进事件，并进行内部归因(e.g.,Yalcin et al., 2023)。

参考文献：

Audia, P. G., & Locke, E. A. (2003). Benefiting from negative feedback. *Human Resource Management Review, 13*(4), 631−646.

Garvey, A. M., Kim, T. W., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing, 87*(1), 10−25.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160,* 149−167.

Raveendhran, R., & Fast, N. J. (2021). Humans judge, algorithms nudge: The psychology of behavior tracking acceptance. *Organizational Behavior and Human Decision Processes, 164,* 11−26.

宋晓兵, 何夏楠.(2020).人工智能定价对消费者价格公平感知的影响. *管理科学, 33*(5), 3−16.

意见 3: 实验 2 委托了专业调查机构，这里需要给出机构的具体背景信息，解释为什么该机构是可信的。

回应：感谢您提出的这一问题。在上一版本中，我们在正文中没有明确专业调查机构的信息。我们在这一版本中进行了补充。实际上，实验 2~4 均委托了问卷网(http://www.wenjuan.com)向企业定向招募愿意参与本研究的员工被试。问卷网是当前市面上比较主流的调查机构，据其官方网站介绍，“问卷网当前已为 2000 万用户收集超过 18 亿份数据，并服务中小企业上万家”。此外，根据我们的了解，调查机构的样本服务一般分为公共被试库与定向招募两种。前者是长期在平台参与调查研究的被试人群，而后者则是平台利用企业资源为研究实时招募参与人员。实验 2~4 选择定向招募企业员工被试的取样策略。因为相比于调查机构的公共样本库，定向招募的员工被试会看到我们事先发给调查平台的招募信息，初步了解研究的目的以及指导语，从而决定是否参与研究。据此，我们认为定向招募的方式更具生态效度。另外，我们正文方法部分也将 “委托专业调查机构”的描述明确为 “委托问卷网”。

意见 4: 表 1 除了平均数和标准差外，需要给出每组具体的样本数。另外，表 1 中值得注意的是：4 个组的标准差差异不小，最小的.59，最大的 1.01。——这里有什么特别的原因吗？

同样，表 2 需要给出每组具体的样本数。这里 4 个组的标准差差异同样不小，最小的.88，最大的 1.17。——是否有合理的解释？

回应：非常感谢您的这个意见。首先，我们分别在正文表 1、2 和 3 中给出了各组具体的样本数。为更清晰地展示，我们将实验 2~3 中各组均值、标准差以及样本数汇总在了表 4 中。

表 4 实验 2~4 主观和客观任务下人类管理者与 AI 负面绩效反馈的绩效改进动机的平均数(标准差及每组样本量)

		主观任务组	客观任务组
实验 2	人类组	6.13 (0.59; *n* = 40)	4.60 (1.01; *n* = 40)
	AI 组	5.63 (0.90; *n* = 40)	5.71 (0.66; *n* = 40)
实验 3	人类组	4.92 (1.17; *n* = 41)	4.35 (0.92; *n* = 33)
	AI 组	4.57 (1.26; *n* = 40)	5.60 (0.88; *n* = 36)
实验 4(本轮补充)	人类组	6.43 (0.61; *n* = 40)	5.09 (0.81; *n* = 40)
	AI 组	5.95 (0.82; *n* = 40)	6.19 (0.72; *n* = 40)

需要跟您说明的是，在实验 3 中，涉及发送真实邮件向参与者提供绩效反馈的环节(时长约 20 分钟)，为保证被试在等待过程中专注于实验，我们选择一段介绍某高校测评中心的

视频供其观看(时长约 20 分钟)。因此，相较于实验 2 和 4，实验 3 的总耗时较长，我们招募到了 160 名员工被试参与实验。完成全部实验后，我们进行了作答质量检查。剔除没有通过注意力测试，未完成作答或作答无效的被试 10 名，实验 3 最终有效样本为 150。具体而言，其中人类-主观任务组(实际收取 42 份数据，41 份有效)有 1 名被试作答时间过长(发送绩效反馈 40 分钟后仍未提交问卷)；人类-客观任务组(实际收取 38 份数据，其中 33 份有效)有 3 名被试作答时间过短(填写问卷时间未满 5 分钟)，而 2 名被试存在规律性作答的问题；AI-主观任务组(实际收取 42 名，其中 40 名有效)有 2 名被试未按时提交答卷；AI-客观任务组(实际收取 38 名，其中 36 名有效)有 1 名被试未按时提交答卷，另 1 名被试存在规律性作答问题。

其次，确实如您所说，实验 2 和 3 的结果存在标准差差异的问题。我们仔细分析了数据，发现报告的标准差并没有错误。我们仔细思考了几天，并咨询了相关专家，也没有想到什么特别的原因。不过，经过对相关文献的回顾，我们认为这可能与我们研究的议题 “负面绩效反馈” 有关。具体来说，目前一个比较稳健的研究发现是，面对负面反馈，个体会表现出 “评价有利性偏好” (Favorability of Others' Ratings)(e.g., Ilgen & Hamstra, 1972; Podsakoff & Farh, 1989)，即相对于负面评价，人们更倾向于获得正面的反馈。这导致被试间对于负面绩效反馈的接受度可能是不同的，因此在数据层面可能出现一定程度波动。

例如下表 6 所列的两篇论文中，不同实验组的标准差也存在一定差异。这两项研究与本研究存在较高相关性。

表 6 两篇关不利情景下人机研究各组的均值与标准差

作者及年份	发表期刊	各组均值与标准差	主要研究假设
Garvey, Kim & Duhachek, 2023	*Journal of Marketing*	在商家价格高于预期组：Study2:人类组的积极反应(M = 2.53, SD = 1.40);而 AI 组为(M = 3.17, SD = 0.94) (两组标准差差异大约是 0.40)。	当面对出价高于自身预期这种不利性状况时，相对于人类的决策，个体对 AI 决策的积极反应更强。
Yalcin et al., 2022	*Journal of Marketing Research*	当面对公司的拒绝服务 Study1a: 人类组对公司的口碑效应(M = 3.02, SD = 1.82); 而 AI 组为(M = 3.12, SD = 2.11) (两组标准差差异约为 0.30)。	在有利性条件下，相对于 AI，个体对人类的正面反馈接受度更高；在不利性条件下，相比于人类，个体对 AI 负面反馈的接受度更强。

参考文献：

Garvey, A. M., Kim, T. W., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing, 87*(1), 10−25.

Yalcin, G., Lim, S., Puntoni, S., et al. (2022). Thumbs up or down: Consumer reactions to decisions by algorithms versus humans. *Journal of Marketing Research, 59*(4), 696−717.

. . . . . .

审稿人 2 意见：

经过一轮评审，作者已按照审稿人的意见进行了认真修改和回复，修改稿在理论依据、实验设计和发现讨论等方面已有明显进步。通读下来，仍有一些问题想要与作者讨论并且希望作者做出进一步完善。

回应：衷心感谢您对上一轮修改所给出的肯定！

意见 1：虽然作者按照审稿人对文章题目做出了修改，但此版本的题目仍不够令人印象深刻和简洁，并且文章的核心强调了人机比较。建议作者将题目修改为“负面绩效反馈下员工绩效改进动机的人机比较研究”。

回应：非常感谢您的这一意见。我们已将题目修改为 “负面绩效反馈下员工绩效改进动机的人机比较研究”。

意见 2：前言第一段在引出问题时仍需改进。作者需要思考一个问题，即本研究的出发点是解决人类管理者在提供负面绩效反馈方面的压力、不准确等方面的问题，还是想要解决员工作为反馈接收者的绩效改进问题。据我了解，本文的重心应放在后者，因此在问题提出部分重点要放在传统人际场景下负面绩效反馈对员工消极影响方面的论述。

回应：非常感谢您的这一意见。在上一版本中，我们在前言论述中的确没有紧紧围绕文章的核心议题。因此，在这一版中，遵照您的意见，我们将前言写作重点放在了传统人际反馈情景下负面绩效反馈对员工消极影响的论述(见正文 P39)：

负面绩效反馈(negative performance feedback)是组织对未达到业绩期望的员工所给予的指示、否定和批评(Cianci et al., 2010)。通常来说，领导者向员工传达负面绩效反馈的目的在于引导和激励员工工作行为，最终提升员工的绩效水平(Podsakoff & Farh, 1989; Lam et al., 2011)。然而遗憾的是，负面绩效反馈往往会引发员工焦虑、悲伤等负面情绪，并且降低员工的自我效能感，使其感受到不被组织认可或需要，从而不利于员工的绩效提升(e.g., Kitz et al., 2023; Kim & Kim, 2020)。此外，由于涉及人际间的沟通环节，负面绩效反馈还会降低领导与下属间的关系质量。特别是在中国文化背景下，考虑到人们的沟通方式较为含蓄或“面子”因素，领导者的“低评价”会使得员工产生愧疚和尴尬情绪，进而损害员工提升绩效的积极性(e.g., 耿紫珍, 赵佳佳, 丁琳, 2020)。另外，盖勒普(Gallup)的调查显示，在对领导者

负面绩效反馈产生负面情绪(感受到被批评，丧失动力，失望以及沮丧)后，仅有 10.4%的员工会继续投入工作或改善绩效水平。综合看来，传统由人类管理者提供负面绩效反馈的方式面临着较大的挑战(Kluger & DeNisi, 1996; Xing et al., 2023)。

参考文献：

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology, 95*(4), 618−630.

耿紫珍, 赵佳佳, 丁琳. (2020). 中庸的智慧：上级发展性反馈影响员工创造力的机理研究. *南开管理评论 23*(1) ,75−86.

Kitz, C. C., Barclay, L. J., & Breitsohl, H. (2023). The delivery of bad news: An integrative review and path forward. *Human Management Review,33*(3), 1−23.

Kluger, A. N., & Denisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254−284.

Lam, C. F., DeRue, D. S., Karam, E. P., et al. (2011). The impact of feedback frequency on learning and task performance: Challenging the "More is better" assumption. O*rganizational Behavior and Human Decision Processes, 116*(2), 217−228.

Xing, L., Sun, J. M., Jepsen, D., et al. (2023). Supervisor negative feedback and employee motivation to learn: An attribution perspective. *Human Relations, 3*, 1−31.

Kim, Y. J., & Kim, J. (2020). Does negative feedback benefit (or harm) recipient creativity? The role of the direction of feedback flow. *Academy of Management Journal, 63*(2), 584−612.

Podsakoff, P. M., & Farh, J. L. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes, 44*(1), 45−67.

意见 3：假设 1 的推演逻辑仍需改进。实际上，作者应该在假设 1 之前增加一部分文献述评（关于负面绩效反馈和 AI 相关的研究观点），在此基础上，重点论述人类管理者和 AI 的反馈特征差异，这些差异如何在负面绩效反馈上表征出来，以此来提出假设 1。现有版本在负面绩效反馈方面的讨论较少，对相关概念的介绍，以及背景知识的介绍较多，掩盖了假设推演的逻辑性。

回应：非常感谢您的这一意见。我们在假设 1 部分的确过多地介绍了相关概念与背景知识，对于人机负面绩效反馈对个体绩效改进动机影响的核心论述较少。在现版本中，我们首先评述了传统反馈场景下，负面绩效反馈对员工消极影响的研究，并结合人机的不同特点完善了假设 1 的推导(见正文 P40-41)：

在组织行为研究中，负面绩效反馈对员工动机和绩效的影响已被广泛探讨。传统上，由管理者提供负面绩效反馈的场景中，大部分研究关注其对员工的消极影响。例如，减少员工的学习动机(Xing et al., 2023)，以及自我效能感(Dimotakis, Mitchell, & Maurer, 2017)，降低员工目标设置以及绩效改进(Podsakoff, & Farh, 1989)，或阻碍员工创造力(Kim & Kim, 2020)。以往研究认为，这些消极影响的产生可以通过三条路径解释：人际破坏属性、负面情绪引发和自我防御机制。首先，负面绩效反馈带有人际破坏属性。员工可能将负面绩效反馈知觉为

管理者的敌意，致使学习动机与绩效改进下降(e.g., Ni & Zheng, 2024; Cianci, Klein, & Seijts, 2010)。第二，负面绩效反馈容易引发员工消极的情绪。在悲伤，受挫，羞愧等负面情绪状态的影响下，员工对绩效提升的注意水平下降，转而关注或反思与上级的关系质量(e.g, Kim & Kim, 2020)。最后，在自我防御的驱使下，负面绩效反馈可能减少员工内部归因，导致其绩效水平的降低(e.g., Xing et al., 2023)。

人类与 AI 提供反馈的特征存在较大差异。人类管理者在提供反馈时，其社会属性和主观属性较为突出。他们能够展现高度的情感和人际互动能力，并运用个人经验和知觉来分析信息（蒋路远等, 2022; Newman, Fast & Harmon, 2020; Logg, Minson & Moore, 2019）。然而，这种主观性和经验性可能导致员工将负面反馈感知为带有个人偏见和敌意，从而引发消极情绪和自我防御反应，减少内部归因和绩效改进动机（Ni & Zheng, 2024; Cianci, Klein, & Seijts, 2010）。相对地，AI 在提供反馈时展现出的机械属性和客观属性较为显著。AI 基于算法程序驱动，不具备情感，且处理信息基于数据和事实，这提升了反馈的公平性和无偏性（宋晓兵, 何夏楠, 2020; 许丽颖,喻丰, 彭凯平, 2022）。此外，人类与 AI 提供反馈的特征差异在负面反馈情境中更为明显。面对 AI 提供的负面绩效反馈，AI 的属性减少了员工对其敌意或蓄意意图的感知，个体更愿意相信该反馈针对绩效改进的事件，而非针对员工自身(e.g., Garvey, Kim, & Duhachek, 2023)，从而减少员工对 AI 负面反馈的消极情绪(e.g., Yalcin et al., 2023)。相反，囿于高社会属性与主观属性，人类管理者的反馈容易包含个人看法或主观判断，致使其相对难以提供高客观程度的绩效反馈(e.g., Tong, Jia, Luo, & Fang, 2021)。另外，既有研究也证实，在负面的事件中，相对于人类，个体更容易接受 AI 的决策。比如，当面对商品价格不公平时，相较于人类，消费者会认为 AI 的决策是基于对海量数据的分析，并产生更高水平的信任。而人类销售员的决策可能存在一定主观性与经验性，从而引发消费者较高的蓄意判断(宋晓兵, 何夏楠, 2020)。此外，当个体行为受到侵入式的监控时，相较于人类监控者，AI 算法的监控被认为具有较低的主观判断与意志，使得个体感受到较高的自主性，从而更容易被接受(Raveendhran & Fast, 2021)。

参考文献：

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent Algorithm aversion. *Journal of Marketing Research, 56*(5), 809−825.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629−650.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149−167.

Raveendhran, R., & Fast, N. J. (2021). Humans judge, algorithms nudge: The psychology of behavior tracking acceptance. *Organizational Behavior and Human Decision Processes, 164*, 11−26.

宋晓兵, 何夏楠.(2020).人工智能定价对消费者价格公平感知的影响. *管理科学, 33*(5), 3−16.

Xing, L., Sun, J. M., Jepsen, D., et al. (2023). Supervisor negative feedback and employee motivation to learn: An attribution perspective. *Human Relations, 3*, 1−31.

Dimotakis, N., Mitchell, D., & Maurer, T. (2017). Positive and negative assessment center feedback in relation to development self-efficacy, feedback seeking, and promotion. *Journal of Applied Psychology, 102*(11), 1514−1527

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.

Kim, Y. J., & Kim, J. (2020). Does negative feedback benefit (or harm) recipient creativity? The role of the direction of feedback flow. *Academy of Management Journal, 63*(2), 584−612.

意见 4：在实验部分，实验一和二、三使用的材料为什么不同？一种算法系统小 ai、一种是人工智能评估助手，前一种是嵌入式 AI，后一种是实体 AI，想与作者沟通的是，这样区别的作用是什么？以及如果这种区别本身上就是一种创新，作者可以在讨论部分简要提及。

回应：非常感谢您的这一意见。如您所说，实验 1 与实验 2，3 对于 AI 的操纵的确具有差别。事实上，实验 1 的执行时间较早，起初我们判断现阶段工作场所中嵌入式 AI 的应用更为广泛(例如采用算法进行绩效监控或提供绩效反馈)。因此，实验一我们参考了 Castelo, Bos, & Lehmann(2019)的实验材料，并采用了算法系统的描述。然而近年来实体机器人在工作场所中的发展较为迅猛，前沿的研究也开始关注机器人凸显或涌现对员工工作行为的影响(e.g., 许丽颖,喻丰,彭凯平,王学辉,2022; Yam et al., 2023)，此外，Garvey, Kim, & Duhachek (2023)的实验材料给予我们较大的启发。为减少机器人形象过高的恐怖谷效应进而干扰实验结果，他们对 8 种机器人形象进行了材料预测试，最终发现图 1 所示的机器人形象综合效果最佳。整体考量下，一方面实体机器人已逐渐进入工作场所并与组织成员发生交互，另一方面 Garvey 等(2023)的实验材料较为严谨与可靠。因此实验 2 与实验 3 选择了实体机器人作为刺激材料。

由于以往研究发现，有形性 AI 与嵌入式 AI 可能导致观察者差异化的知觉反应。比如，有形性机器人拥有实体，能够通过对话，表情，肢体等信号与人类产生更高水平的互动，从而产生引发更高水平的人机信任(e.g., Glikson & Woolley, 2020)。考虑到这一因素，相比于算法，我们认为采用 robot 提供反馈能够减少个体对人或机信任方面的差异，从而获得更好的人机组间平衡。总的来说，实验 1 与实验 2~3 在 AI 刺激材料的方面有所不同，但实验间的结果保持了相对较高的一致性。根据您的建议，我们在正文的研究概览以及结果讨论部分对此做出了简要的提及(见正文 P43-44 以及 P59)。

最后，想跟您进一步解释的是，实验 1 与实验 2,3 的材料具有一定的关联性。考虑到 AI(无论嵌入式 AI 抑或实体型 AI)均依赖算法的驱动，我们在实验 1~3 中均向被试解释了 AI 提供负面绩效反馈的机理(比如，实验 2 中展示，人工智能评估助手小 ai，会基于算法系统(该算法系统是基于测评专家设计的评价标准，由计算机专家开发的程序)对你的答案自动进行识别和分析，评估你的作答质量，并进行统计排名，对你本次测试完成情况进行评估反馈)，增加解释的目的在于减少 AI 工作原理的不透明性对个体造成的负面感受，从而平衡人类与 AI 负面绩效反馈在这一因素上的组间差异。

图 1 实验 2 与实验 3 参考 Garvey 等(2023)研究选择的 AI 刺激材料

参考文献：

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent Algorithm aversion. *Journal of Marketing Research*, *56*(5), 809−825.

许丽颖,喻丰,彭凯平,王学辉.(2022).智慧时代的螺丝钉：机器人凸显对职场物化的影响. *心理科学进展, 30*(9),1905−1921.

Yam, K.,Tang, P., Jackson, J., Su, R., Kurt, G.(2023).The rise of robots increases job insecurity and maladaptive workplace behaviors: Multimethod evidence. *Journal of Appied Psychology, 108*(5), 850−870.

Garvey, A. M., Kim, T. W., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing*, *87*(1), 10-25.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627−660.

意见 5：结论部分是否可以删除？实际上在总讨论部分读者已经了解了本文的总体发现。

回应：非常感谢您的这一意见。遵照您的意见，我们已将结论部分删除。

最后，感谢作者对审稿人建议的重视，这篇文章具有很高的实践意义和理论创新性，建议作者在对上述问题修改后，交由团队其他成员交叉阅读，提高文章的可读性和趣味性。

回应：非常感谢您提出的宝贵意见以及对于研究的肯定！我们汲取了团队内其他成员以及外团队成员对于本文写作方面的建议，力求提升文章的趣味性。同时，我们也邀请了团队内成员和外团队成员对论文进行交叉阅读，从而提高论文的逻辑性和可读性。

第三轮

审稿人 2 意见：

修改稿较上一稿已有较大改进，不仅解决了审稿人提出的意见和问题，还通过补充一个实验研究来进一步检验所提假设和模型，修改工作认真且有效。总体上，审稿人认为本篇论文已到达期刊发表的水平。具体而言，在选题方面，本研究结合数智化情境解决了一个传统绩效管理实践中的一大管理难题，即领导向下属提供负面绩效反馈的“两面性”，将组织绩效管理实践引领至人工智能时代;此外，通过比较人工智能和人类管理者在提供负面绩效反

馈中的作用，对过去负面绩效反馈的文献有所推进，一些传统情境下的理论命题和基本假设被人工智能的介入所打破;同时，较以往研究不同的是，本研究在不同任务类型下探讨了负面绩效反馈的人机对比效应，进一步丰富了人工智能的相关研究，在研究方法方面，作者通过 4 个实验对所提假设进行检验，研究设计较为严谨，数据分析过程得当。特别是，作者在实验 4 补充了过程性的反馈，通过结果反馈和过程反馈两种负面绩效反馈方式来验证假设。最后，在生成式 AI 到来的时代，越来越多的组织部署人工智能执行管理流程中的各类职责，因此员工可能在多种环节收到来自人工智能的负面绩效反馈，例如，员工培训、安全监督、临床诊疗等，本文所提发现一方面可以解放传统管理者的“双手”，减少其提供负面绩效反馈中的压力，同时由人工智能所提负面反馈更易被员工接纳、学习和成长，以此实现双赢的目标，实现人机协同，从实践价值上，本文同样具有较好贡献。

诚然，为帮助作者进一步完善研究，审稿人还具有如下建议供作者参考:

回应：衷心感谢您为本研究提出的专业、细致且富有启发性的意见。在三轮的修改过程中，两位审稿专家的意见不仅帮助我们不断完善这个研究，同时也对我们未来的学术思考提供引导与启迪。此外，也非常感谢您对本研究修改工作给出的肯定。看到您对文章的点评后，我们深受启发，并在文章管理启示部分借鉴和引用了您的部分点评“本研究启示组织可以应用数智技术赋能绩效反馈或绩效管理流程，发挥 AI 客观，无偏的绩效反馈优势，这一方面能够减轻人类管理者提供负面绩效反馈的压力，另一方面，来自 AI 的负面绩效反馈更容易被员工接纳，从而提升反馈实施的效果”。以下是我们对您本轮提出意见的逐一回应。

意见 1:在描述人工智能的特征或属性时，建议增加参考这篇:

Qin, M., Jia, N., Luo, X., Liao, C., & Huang, Z. (2023). Perceived fairness of Human Managers Compared with Artificial Intelligence in Employee Performance evaluation. *Journal of Management Information Systems, 40*(4),1039-1070.

回应：非常感谢您提出的这个意见，这篇文献帮助我们更好地学习和描述 AI 反馈的特征。基于这篇文献的介绍，我们完善了描述 AI 特征或属性的论述(详见 P.49~50)“人类与 AI 提供反馈的特征存在较大差异。比如，由于依赖经验与直觉处理信息，人类管理者提供的反馈具备较高水平的主观性(Qin et al., 2023; 蒋路远等, 2022; Newman et al., 2020)，且容易包含个人看法或偏见，继而引发员工的消极情绪和防御反应(Ni & Zheng, 2024)。相反，AI 作为反馈提供者不容易出现认知疲劳和情绪失控，并且由于具备强大的数据分析与预测能力，使 AI 反馈更加客观与全面，也较少被个体知觉为恶意或偏见(Qin et al.,2023; 许丽颖等, 2022)”。

意见 2: 在引言部分，建议作者修改研究动机的表达“由于当前有关 AI 提供负面绩效反馈的研究还处于相对初级的阶段，相关研究还很匮乏(e.g., Tong et al,2021)。建议修改为“尽管已有文献初步表明，AI 和人类管理者在提供反馈时可能呈现出不同的特征，员工在与 AI

和人类管理者互动时也会产生差异化反应，但少有研究在 AI 提供负面绩效反馈情境下探讨员工的归因过程及其后续反应，因此本研究的目标包括:...

回应：非常感谢您提出的这个意见。遵照您的意见，我们完善了对于研究目标的表达(详见 P.49)“尽管已有文献初步表明，AI 和人类管理者在提供绩效反馈时可能呈现出不同的特征(Garvey, Kim & Duhachek, 2023; Yalcin et al., 2022)，员工在与 AI 或人类互动时也会产生差异化的反应(e.g., Tong et al., 2021)，但少有研究在人机提供负面绩效反馈情境下探讨员工的归因过程及后续反应，因此本研究的目标包括：首先，研究拟基于人机比较的相关研究，探索人机负面绩效反馈(即由人类管理者或 AI 提供负面绩效反馈)的差异化影响效应。第二，当前算法态度(algorithm attitude)的研究表明，人类对 AI 表现出的欣赏抑或厌恶可能依据不同类型的任务而定(e.g., Castelo, Bos & Lehmann, 2019)。比如在客观型任务中，相比于人类反馈，个体更信赖客观性更强、精确度更高的 AI 反馈。因此研究拟探索任务类型(主观或客观任务)在人机负面绩效反馈中的边界效应。最后，既有研究发现，员工在接收负面绩效反馈后，会对反馈信息进行归因(将绩效不佳的成因归因于自身或外部环境)，从而决定后续绩效目标的改进(e.g., Ilgen, Fisher & Taylor, 1979; Tolli & Schmidt, 2008)。基于此，研究拟从内部与外部归因的视角切入，进一步解释人机提供负面绩效反馈产生差异化效应的机制”。

意见 3: 图片采用编辑部可编辑的图片而非截图。

回应：非常感谢您提出的这个意见。我们已将文章实验 2~4 三张图片更改为可编辑的格式。

意见 4: 在对归因理论的贡献时，能否提到，过去归因理论认为，个体在收到来自外界的负面刺激后可能会做出外部归因，但本文突破在于，收到来自 AI，作为一种非人类实体的负面刺激后可能会做出内部归因，这对过去的归因研究也是一项重要的突破。

回应：非常感谢您提出这个富有启发的意见。基于您指出的视角，我们进一步探讨了本研究对归因理论的贡献(详见 P.68)“最后，本研究深化了归因理论在组织场景中的研究。归因理论被广泛应用于解释人际互动中个体如何理解自身或他人行为的原因(e.g., Tolli & Schmidt, 2008)。根据经典归因理论的观点(Heider, 1958)，人们通常出于自我防御的目的对不利性结果进行外部归因，或对有利性结果进行内部归因从而获得自我提升。不过上述结论也受到一些因素的调节，比如，Xing 等(2023)发现员工核心自我评价水平越高，越会将负面绩效反馈视作提升与改善绩效的机会，从而提高内部归因与学习绩效。而本研究深入探究人机反馈的差异化影响，发现 AI(较人类管理者)提供负面绩效反馈可能会提升个体的内部归因。并结合人机不同的反馈特征(比如，相较于人类，AI 具备更少的主观或伤害意图)(e.g., 蒋路远等，2022)进行了解释。本研究结果表明，在削弱了外部负面刺激的消极影响时(如采用 AI 替代人类管理者进行负面绩效反馈)，个体可能会增强内部归因，这为归因理论解释不利性结果中个体的归因倾向或行为提供了新的认识”。

参考文献

Tolli, A. P., & Schmidt, A. M. (2008). The role of feedback, causal attributions, and self-efficacy in goal revision. *Journal of Applied Psychology, 93*(3), 692−701.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons Publishing House.

Xing, L., Sun, J. M., Jepsen, D., & Zhang, Y. J. (2023). Supervisor negative feedback and employee motivation to learn: An attribution perspective. *Human Relations, 3*, 1−31.

Garvey, A. M., Kim, T. W., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing, 87*(1), 10−25.

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.

意见 5: 在局限和展望部分，建议作者增加样本地区的局限性，因为相较于西方社会，中国社会对来自领导负面反馈可能反应更强。未来研究应增加对西方样本的检验。

回应：非常感谢您提出这个富有启发的意见。遵照您的意见，我们在文章(P.69~70)的研究局限处补充道“此外，相比于西方，东方社会在和谐文化的影响下，人际间的沟通方式更为含蓄(e.g.,耿紫珍等, 2020)，这可能导致东方社会中的个体更加消极地应对负面绩效反馈，继而影响人机负面绩效反馈的差异化效果。未来研究可以采用西方样本，探究文化背景差异下人机负面绩效反馈的效果”。

……

审稿人 3 **意见：**作者的论文选题新颖，且具有很强的现实意义。作者设计了 4 个递进式的研究，研究设计合理。在审阅本论文的同时，审稿人还仔细看了作者的论文修改过程及修改说明。审稿人认为，作者在两位审稿人的精心指导与帮助下对论文进行了认真的修改，论文质量有了明显提升，尤其是补充了研究 4。尽管如此，审稿人还有以下 3 点建议，供作者参考:

回应：衷心感谢您为本研究提出的专业、细致且富有启发性的意见。在三轮的修改过程中，两位审稿专家精心的指导不仅帮助我们不断完善这个研究，同时也对我们未来的学术思考提供引导与启迪。同时，也非常感谢您对本研究修改工作给出的肯定。以下是我们对您本轮提出意见的逐一回应。

意见 1：进一步挖掘并提炼研究发现的贡献性或意义，以及对管理实践的启迪意义，以丰富论文的讨论部分。建议作者结合绩效管理 (performance management) 研究的前沿动态，进一步提炼本研究发现的理论价值，而不是仅仅局限于丰富相关的研究文献，同时，还可以拓展研究发现对绩效管理实践的启迪意义。

回应：非常感谢您提出的这个意见。我们在原文基础上，对绩效管理的文献进行了仔细回顾，其中包括 Pulakos 等(2019)以及 Schleicher 等(2018)对于绩效管理的综述型研究。这两篇论文是绩效管理领域近年来非常重要的论文，对于绩效管理研究的前沿动态进行了很好的归纳和

总结。其中，相关文献均探讨了传统绩效管理周期过长的缺陷，并提出敏捷绩效管理的未来趋势与构想。由于绩效反馈是绩效管理的重要一环，我们遵照您的意见，增加了一段对于绩效管理研究的理论贡献，并完善了理论与实践意义的写作(详见本文 P67~69)：

参考文献

Pulakos, E. D., Mueller-Hanson, R. A., & Arad, S. (2019). The Evolution of Performance Management: Searching for Value. *Annual Review of Organizational Psychology and Organizational Behavior, 6*, 249−271.

Schleicher, D. J. , Baumann, H. M. , Sullivan, D. W. , Levy, P. E. , Hargrove, D. C. , & Barros-Rivera, B. A. . (2018). Putting thesysteminto performance management systems: a review and agenda for performance management research. *Journal of Management, 44*(6), 2209−2245.

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.

理论意义方面：

首先，本研究拓展了既有负面绩效反馈研究的视角。具体而言，传统基于人际互动的负面反馈研究大多关注来自人类管理者的反馈(e.g., Kitz et al., 2023)，而本研究则发现 AI 替代人类管理者提供负面绩效反馈潜在的积极效应。既有研究从多种角度探索提升负面绩效反馈实施效果的途径，例如，反馈特征层面(绩效反馈的频率，即时性或质量等)(e.g., Kuvaas et al., 2017; Ni & Zheng, 2024)，员工个体层面(对负面绩效反馈的积极归因，员工核心自我评价等)(马璐等, 2021；Xing et al., 2023)。而本研究结合数智化时代背景，基于人机提供负面绩效反馈的新兴视角，发现 AI(相较人类管理者)提供的负面绩效反馈提升员工后续的绩效改进动机，为负面绩效反馈的人机差异化影响效果提供了研究证据。

其次，本研究丰富了既有人机反馈的研究。当前数智化技术在绩效反馈中的应用引发了一些争论(董毓格等, 2022)。一方面，基于算法欣赏的视角，研究者发现 AI 能够提升绩效反馈的准确性和可靠性，从而提升员工的绩效水平(e.g., Tong et al., 2021)，但另一方面，也有研究从算法厌恶角度出发，发现 AI 缺乏真诚性与独特性，并且会威胁人类的工作机会，因此当组织披露绩效反馈(尤其是带有鼓励、赞扬性质的正面反馈)来源于 AI 时(e.g., Yalcin et al., 2022)，会降低个体的积极表现(Tong et al., 2021; Luo et al., 2019)。本研究聚焦于负面绩效反馈，并发现 AI（较人类管理者）作为反馈提供者提升个体的绩效改进动机。此外，既有研究也关注人机反馈产生差异化效果的边界条件，比如，Tong 等(2021)发现，对于任期较长的员工而言，由于他们与组织建立了更强的情感纽带，对于组织采用 AI 提供绩效反馈的变革也更为支持，因此员工的任期会缓解 AI 提供绩效反馈的负面效果。此外，Luo 等(2019)发现，顾客对于 AI 的熟悉程度会降低个体对于 AI 的刻板印象(例如，缺乏知识和同理心)，从而缓解由 AI 提供反馈造成的产品销量下降。本研究关注员工工作的任务类型这一外部因素，并发现人机负面绩效反馈与任务类型对员工绩效改进动机的交互作用，从而拓展了人机反馈边界条件的研究。

此外，本研究对绩效管理领域的研究具有边际贡献。传统以年，季度为时间单位的绩效管理模式存在周期过长的缺陷，不利于员工即时获取信息并提升绩效。为此，有学者提出敏

捷(agile)绩效管理的变革趋势(Pulakos et al., 2019; Schleicher et al., 2018)，旨在提升绩效管理的时效性，并为员工提供准确，高质量的绩效评估与反馈。数智化是敏捷绩效管理最为重要的驱动力，具体表现在 AI 能够不知疲倦地整合并分析数据，为员工提供更为即时的反馈，同时也能提供客观、无偏地评估员工的绩效表现，并提供个性化的绩效反馈(Qin et al.,2023; Tong et al., 2021)。除人机绩效反馈的研究外，也有研究关注了 AI 绩效指导(AI coach)。比如，Luo 等（2021）发现 AI 教练相对于人类教练的指导效果在不同的销售人员中呈倒 U 形分布。这是因为绩效排名靠后的销售会面临 AI 反馈信息过载的问题，而绩效排名靠前的销售对 AI 的厌恶程度较高。本研究与上述文献一致，均探索了数智化技术对绩效管理中特定环节的影响和机制。

最后，本研究深化了归因理论在组织场景中的研究。归因理论被广泛应用于解释人际互动中个体如何理解自身或他人行为的原因(e.g., Tolli & Schmidt, 2008)。根据经典归因理论的观点(Heider, 1958)，人们通常出于自我防御的目的对不利性结果进行外部归因，或对有利性结果进行内部归因从而获得自我提升。不过上述结论也受到一些因素的调节，比如，Xing 等(2023)发现员工核心自我评价水平越高，越会将负面绩效反馈视作提升与改善绩效的机会，从而提高内部归因与学习绩效。而本研究深入探究人机反馈的差异化影响，发现 AI(较人类管理者)提供负面绩效反馈可能会提升个体的内部归因。并结合人机不同的反馈特征(比如，相较于人类，AI 具备更少的主观或伤害意图)(e.g., 蒋路远等，2022)进行了解释。本研究结果表明，在削弱了外部负面刺激的消极影响时(如采用 AI 替代人类管理者进行负面绩效反馈)，个体可能会加强内部归因。这为归因理论解释不利性结果中个体的归因倾向或行为提供了新的认识。

管理启示方面：

本研究也具有一定的管理启示。首先，传统由人类管理者主导的负面绩效反馈可能破坏领导-下属关系，为员工带来负面情绪并降低绩效水平(e.g., Ni & Zheng, 2024)。而本研究的结果表明，AI(较人类管理者)增强了个体的内部归因与绩效改进动机。本研究启示组织可以应用数智化技术赋能绩效管理流程，发挥 AI 客观、无偏的绩效反馈优势。这一方面能够减轻人类管理者提供负面绩效反馈的压力，另一方面，来自 AI 的负面绩效反馈更容易被员工接纳，从而提升反馈实施的效果。

第二，尽管数智化技术具有高效、客观、标准化等优势，但它减少了绩效反馈过程中的人际互动或同理心(董毓格等, 2022; Yalcin et al., 2022)，因此需要区分人机反馈不同的应用场景。根据本研究的结果，组织应关注人机负面绩效反馈中的任务特征，比如，AI 以其客观和无偏的特征为客观任务(比如业绩分析、销量预测等)中的负面绩效反馈提供优势。但相比于人类管理者，由于 AI 缺乏社会或互动属性，因而在主观工作任务(比如人际沟通、矛盾处理等)中进行负面绩效反馈的效果较差。因此，组织应事先辨别任务的类型，充分发挥人类管理者与 AI 各自的反馈优势。

第三，本研究为人机负面绩效反馈后，组织帮助员工进行积极的心理建设提供了管理启示。由于员工对人机负面绩效反馈的内部归因会影响员工的绩效改进动机，对负面绩效反馈的内部归因越高，绩效改进的动机也越高。因此，组织需要关注人机负面绩效反馈后员工的归因方式，并加强绩效沟通，帮助员工及时地发现自身不足或改善绩效反馈流程，从而提升员工绩效。

意见 2：“绩效改进动机”的测量工具在 4 个研究是否相同?如是，需要作明确的说明。

回应：非常感谢您提出的这个意见。为保证测量的一致性，以及实验间结果的可对比性，本研究 4 个实验的因变量“绩效改进动机”均采取了相同的测量题项(Wexley, Singh, & Yukl, 1973)。遵照您的意见，我们在实验 1 描述因变量测量时补充道：“实验 1~4 均采用该测量的题项”(详见 P. 53)。

参考文献

Wexley, K. N., Singh, U. A., & Yukl, G. A. (1973). Subordinate personality as a moderator of effects of participation in 3 types of appraisal interviews. *Journal of Applied Psychology, 58*(1), 54−59.

意见 3：论文的部分语言表达需再斟酌或作明确的说明，例如，“这能够增强 AI 机械属性与客观属性的优势”(第 42 页)，其中的“机械属性”的内涵是什么?等等。

回应：非常感谢您提出的这个意见。依据蒋路远, 曹李梅, 秦昕, 谭玲, 陈晨, 彭小斐(2022)的描述，机械属性指 AI 依托整合大量数据并分析的固定模式输出结果，而无法输出直觉或经验性的类人化决策。因此 AI 提供的反馈通常更加客观，但缺乏互动和情感体验。我们思考再三删除了关于“机械属性”的描述和内容(详见论文的 P.49~50)。原因是，一方面，机械属性相对晦涩难懂，另一方面，与本研究想表达的 AI 具备的客观属性有所重叠。此外，遵照您的意见，我们将论文在团队中交叉阅读，对部分语言表达进行了完善。

参考文献

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.

第四轮

审稿人 3 意见：作者对审稿人的建议进行了认真细致的修改，审稿人赞同这些修改。但审稿人对讨论部分的修改仍有一些建议，例，“此外，本研究对绩效管理领域的研究具有边际贡献。(第 68 页)“边际贡献”指的是什么?"…同时也能提供客观、无偏地评估员工的绩效表现…"(第 68 页)这句话似乎不通顺，应该调整为"….同时也能客观、无偏地评估员工的绩效表现.…吧?

回应：非常感谢您对论文修改的认可以及指出的问题。首先，在理论贡献部分(P.68)，我们思索下认为“边际贡献”的表述不够规范，已修改为 “本研究丰富了敏捷型(agile)绩效管理的研究”。此外，我们非常认可您指出的表达不通顺的问题，在文章 P.68 部分，我们修改了整句话的表述 “表现在 AI 能够不知疲倦地整合并分析数据，以客观、无偏的方式评估员工的绩效表现，并提供更加准确的绩效反馈”。

编委意见：同意录用。

主编意见：同意发表。