A Human-Machine Comparative Study on Employee Motivation for Performance Improvement under Negative Performance Feedback

Authors: Wang Guoxuan, Long Lirong, Li Shaolong, Sun Fang, Wang Jiaqing, Huang Shiyingzi, Li Shaolong

Date: 2024-10-20T16:09:16+00:00

Abstract

Negative performance feedback is crucial for employee learning and performance improvement, yet it is often difficult for employees to accept. As artificial intelligence (AI) technology becomes increasingly prevalent in organizational contexts, investigating the effects of AI-delivered negative performance feedback on employee behavior and attitudes has emerged as an important research topic. This study employs four progressive experiments to explore the differential effects and underlying mechanisms of AI versus human managers providing negative performance feedback on individuals' performance improvement motivation. Experiments 1–3 utilize the classic false feedback paradigm, revealing that compared to human managers, AI-provided negative performance feedback elicits higher levels of performance improvement motivation in individuals (Experiment 1). Moreover, in objective tasks, AI (relative to human managers) delivering negative performance feedback triggers higher levels of performance improvement motivation, whereas the opposite effect is observed in subjective tasks (Experiment 2). Additionally, individuals' internal attribution of negative performance feedback explains the underlying mechanism of these relationships (Experiment 3). Experiment 4 adopts a relatively authentic negative performance feedback scenario, replicating the findings from the previous three experiments. This research provides insights into why and when organizations should employ AI to deliver negative performance feedback.

Full Text

Human-AI Comparison in Employee Performance Improvement Motivation Under Negative Performance Feedback

Guoxuan Wang[1], Lirong Long[1], Shaolong Li[2], Fang Sun[3], Jiaqing Wang[1], Shiyingzi Huang[1]

[1] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China
[2] Economics and Management School, Wuhan University, Wuhan 430072, China
[3] Business School, Hubei University of Economics, Wuhan 430205, China

Abstract

Negative performance feedback is crucial for employee learning and performance improvement, yet it is often poorly received by employees. As artificial intelligence (AI) technology becomes increasingly applied in organizational contexts, exploring the impact of AI-delivered negative performance feedback on employee behavior and attitudes has emerged as an important research topic. This study employed four sequential experiments to investigate the differential effects and underlying mechanisms of AI versus human managers delivering negative performance feedback on individuals' motivation to improve performance. Experiments 1–3 adopted the classic false feedback paradigm and found that, compared to human managers, AI-delivered negative performance feedback elicited higher levels of performance improvement motivation (Experiment 1). Moreover, in objective tasks, AI (relative to human managers) delivering negative performance feedback generated higher performance improvement motivation, whereas the opposite pattern emerged in subjective tasks (Experiment 2). Furthermore, individuals' internal attribution for negative performance feedback explained the underlying mechanism of these relationships (Experiment 3). Experiment 4 utilized a relatively realistic negative performance feedback scenario and replicated the findings from the previous three experiments. This research provides insights into why and when organizations should apply AI to deliver negative performance feedback.

Keywords: negative performance feedback, artificial intelligence, performance improvement motivation, internal attribution, task type

---

Negative performance feedback refers to the criticism and negative evaluation that organizations provide to employees who fail to meet performance expectations (Cianci et al., 2010). Typically, managers provide negative performance feedback with the intention of guiding and motivating employee performance (Podsakoff & Farh, 1989; Lam et al., 2011). However, negative performance feedback often triggers negative emotions such as anxiety and sadness, which subsequently reduce employee performance (Kitz et al., 2023; Audia & Locke, 2003). Additionally, because it involves interpersonal communication, negative

performance feedback can damage the quality of manager-employee relationships (Ni & Zheng, 2024). Particularly in Chinese culture, where communication styles tend to be more indirect, managers' negative performance feedback can cause employees to feel guilty and embarrassed, thereby undermining work motivation (Geng et al., 2020). A 2019 Gallup survey also revealed that after experiencing negative emotions (disappointment, frustration) from managers' negative performance feedback, only 10.4% of employees would continue to exert effort or improve their performance levels. Overall, traditional approaches to delivering negative performance feedback through human managers face significant challenges (Kluger & DeNisi, 1996; Xing et al., 2023).

With the development of digital intelligence technologies, using AI to deliver negative performance feedback presents new opportunities for organizations (Lee, 2018; Luo et al., 2021). For example, an algorithm named Enaible monitors employee work behaviors through remote surveillance, diagnoses the reasons for poor performance, and provides suggestions for performance improvement. Additionally, AI evaluation software such as Butterfly can meticulously collect employee behavioral data to help employees improve their performance in a timely manner (Tong et al., 2021). What potential advantages might AI have over human managers when delivering negative performance feedback? Research has found that AI possesses powerful data integration and analytical capabilities and exhibits fewer subjective intentions (Lee, 2018; Garvey et al., 2023). Therefore, compared to human managers, AI-delivered negative performance feedback is more likely to be "task-focused rather than person-focused" (Yalcin et al., 2022), thereby weakening the attribution bias that employees typically exhibit in traditional interpersonal interactions (e.g., attributing negative performance feedback to managerial prejudice) (Xing et al., 2023) and enabling employees to focus more on their own shortcomings and enhance their motivation to improve performance.

Although existing literature has preliminarily indicated that AI and human managers may exhibit different characteristics when providing performance feedback (Garvey et al., 2023; Yalcin et al., 2022) and that employees produce differential reactions when interacting with AI or humans (Tong et al., 2021), few studies have examined employees' attribution processes and subsequent reactions in the context of human versus AI negative performance feedback. Therefore, the objectives of this study include: First, based on human-AI comparative research, this study aims to explore the differential effects of human-AI negative performance feedback (i.e., negative performance feedback delivered by human managers or AI). Second, current research on algorithm attitudes suggests that human appreciation or aversion toward AI depends on task type (Castelo et al., 2019). For instance, in objective tasks, individuals are more willing to accept AI feedback than human feedback due to its greater accuracy. Therefore, this study aims to explore the boundary effect of task type (subjective vs. objective) in human-AI negative performance feedback. Finally, employees make internal or external attributions for negative performance feedback, which determines whether they will improve their performance (Ilgen et al., 1979; Tolli

& Schmidt, 2008). Based on this, the study adopts an attribution theory perspective to investigate the mechanisms underlying the differential outcomes of human versus AI negative performance feedback.

1.1 The Impact of Human-AI Negative Performance Feedback on Individual Performance Improvement Motivation

In traditional interpersonal contexts, the negative effects of negative performance feedback on employees have been extensively examined, including reduced learning motivation (Xing et al., 2023), diminished self-efficacy (Dimotakis et al., 2017), lower goal setting and performance improvement (Podsakoff & Farh, 1989), and hindered creativity (Kim & Kim, 2020). These negative effects can be explained through three primary pathways: interpersonal destructiveness, negative emotions, and self-defense mechanisms. First, employees may perceive negative performance feedback as managerial hostility, leading to decreased willingness to learn and improve (Ni & Zheng, 2024; Cianci et al., 2010). Second, under the influence of negative emotional states such as frustration and shame, employees become demoralized and pay less attention to performance enhancement (Belschak & Den Hartog, 2009; Kim & Kim, 2020). Finally, driven by self-defense mechanisms, negative performance feedback reduces employees' internal attribution, resulting in lower performance levels (Xing et al., 2023).

The characteristics of feedback provided by humans and AI differ substantially. Because human managers rely on subjective experience and intuition, their feedback tends to incorporate personal opinions or biases (Qin et al., 2023; Jiang et al., 2022), subsequently triggering negative emotions and defensive reactions among employees (Ni & Zheng, 2024). In contrast, as a feedback provider, AI is less prone to cognitive fatigue and emotional 失控, and its powerful data analysis and predictive capabilities render AI feedback more objective and comprehensive, while also being less likely to be perceived as malicious or biased (Qin et al., 2023; Xu et al., 2022). Notably, the characteristic differences between human and AI feedback become more pronounced in negative contexts. For instance, when facing negative information delivered by AI, AI's objectivity reduces individuals' perception of intentional harm and increases acceptance of the information provided (Garvey et al., 2023). Conversely, constrained by subjective biases, humans' negative decisions tend to incorporate personal opinions or subjective judgments, thereby reducing acceptance (Tong et al., 2021). Additionally, in negative events, individuals are more likely to accept decisions from AI than from humans. For example, when confronted with unfair product pricing, consumers believe that AI's decisions are generated based on large amounts of objective data, leading to higher levels of trust, whereas human salespeople's decisions may suffer from subjective limitations, triggering stronger intentional evaluations (Song & He, 2020). Furthermore, when individuals are monitored, algorithmic monitoring is perceived as having lower subjective judgment and will compared to human monitoring, making it more acceptable (Raveendhran & Fast, 2021). Based on the above analysis, we propose:

Hypothesis 1: Compared to human managers, AI-delivered negative performance feedback elicits higher levels of employee motivation to improve performance.

1.2 Task Type as a Boundary Condition

In organizations, tasks related to performance can typically be categorized as subjective or objective (Van Dijk & Kluger, 2011). The former involves open-ended or interpretive tasks based on personal opinions or intuition (such as handling interpersonal relationships and communication), whereas the latter comprises quantifiable, fact-based tasks (such as performance analysis and sales forecasting) (Castelo et al., 2019). Task type is a critical factor influencing individuals' preference for human or AI decision-making. For example, compared to humans, users perceive algorithm-based purchase recommendations derived from objective sales data as more fair and reasonable, making them more willing to accept advice from AI (Helberger et al., 2020). Additionally, subjective tasks rely on interpersonal interaction abilities and require processing through intuition, experience, and tacit knowledge (Castelo et al., 2019), where humans possess greater advantages than AI in terms of social and subjective attributes. Therefore, compared to AI, individuals trust human managers' negative performance feedback more in subjective tasks (Newman et al., 2020), thereby generating higher levels of performance improvement motivation. Conversely, objective tasks are characterized by quantifiability (Castelo et al., 2019), which allows AI's negative performance feedback to fully leverage its objective attributes supported by powerful computational capabilities (Logg et al., 2019; Tong et al., 2021), making employees more receptive to the feedback and enhancing their motivation to improve performance. Based on this analysis, we propose:

Hypothesis 2: Human-AI negative performance feedback and task type interactively influence employees' performance improvement motivation. Specifically, in objective tasks, AI (relative to human managers) delivering negative performance feedback elicits higher levels of employee motivation to improve performance; in subjective tasks, human managers (relative to AI) delivering negative performance feedback elicits higher levels of employee motivation to improve performance.

1.3 The Mediating Role of Internal and External Attribution

Attribution theory (Heider, 1958) distinguishes individuals' attribution styles through the locus of causality perspective, categorizing them as internal or external attribution. Internal attribution emphasizes that individuals tend to seek causes within themselves and believe that current outcomes are related to personal factors (such as ability or personality traits). Conversely, external attribution suggests that individuals' behaviors or outcomes are associated with external factors such as the environment or luck.

When facing negative performance feedback, individuals identify the intentions of the feedback provider to choose between internal or external attribution (Audia & Locke, 2003). Specifically, when perceiving negative performance feedback as stemming from managerial malice (e.g., suppression, harm), individuals tend to make external attributions. Conversely, when negative performance feedback conveys managers' intentions to help employees improve performance, individuals make more internal attributions (Xing et al., 2023; Ni & Zheng, 2024). Because AI's decisions rely on objective data, AI possesses fewer subjective intentions (Garvey et al., 2023). Therefore, compared to humans, AI's decisions in negative contexts exhibit lower intentionality and harmfulness and are more easily accepted. For example, compared to human discrimination, individuals perceive algorithmic discrimination as having lower free will and thus exhibit less desire for moral punishment (Xu et al., 2022). Similarly, when facing higher-than-expected prices, individuals perceive AI (relative to human) pricing as having lower subjective intentions and are more willing to accept it (Garvey et al., 2023).

In summary, compared to human managers, AI's characteristics of being based on objective data and existing facts make negative performance feedback more objective and contain fewer subjective intentions (Tong et al., 2021), thereby enhancing employees' internal attribution (Yalcin et al., 2023). Furthermore, research has found that experiencing negative events such as setbacks can enhance individuals' achievement motivation. Particularly after making internal attributions for negative performance feedback, employees will refine their behavioral performance to achieve higher performance levels and maintain self-esteem (Weiner, 1985). Conversely, when employees attribute performance feedback results to external factors beyond their control, they may experience helplessness and reduced motivation to improve performance (Harvey et al., 2014). Accordingly, this study proposes that when employees attribute negative feedback to internal factors (i.e., internal attribution) rather than external situational factors (i.e., external attribution), their motivation to improve performance increases. Therefore, we propose:

Hypothesis 3: Internal and external attributions mediate the effect of human-AI negative performance feedback on performance improvement motivation. Specifically, compared to human managers, employees exhibit higher levels of internal attribution and lower levels of external attribution for AI-delivered negative performance feedback, which in turn generates higher levels of performance improvement motivation.

Finally, due to the quantifiable nature of objective tasks, AI's stronger data integration and analytical capabilities compared to human managers make it easier to gain individuals' trust, thereby triggering employees' internal attribution (and reducing external attribution) and enhancing performance improvement motivation. Conversely, in subjective tasks, human managers' interpersonal communication experience and interaction abilities enable them to better assess employee performance. Therefore, in subjective tasks, human managers (relative

to AI) delivering negative performance feedback elicits higher levels of internal attribution (and lower external attribution), thereby strengthening individuals' performance improvement motivation (Castelo et al., 2019). In summary, we propose:

Hypothesis 4: Internal and external attributions mediate the interactive effect of human-AI negative performance feedback and task type on performance improvement motivation. Specifically, in objective tasks, compared to human managers, employees exhibit higher levels of internal attribution and lower levels of external attribution for AI-delivered negative performance feedback, which in turn generates higher levels of performance improvement motivation; in subjective tasks, compared to AI, individuals exhibit higher levels of internal attribution and lower levels of external attribution for human manager-delivered negative performance feedback, which in turn generates higher levels of performance improvement motivation.

1.4 Research Overview

This study focuses on the primary question of whether negative performance feedback delivered by AI or human managers influences employees' differential performance improvement motivation. It also examines whether task type serves as a boundary condition in this process and the mediating role of internal and external attributions. This study tested these hypotheses through four sequential experiments. Specifically, to achieve better control over the content of negative performance feedback, Experiments 1–3 employed a false feedback strategy (i.e., participants received identical feedback content). Experiment 1 was conducted on the Credemo platform to test Hypothesis 1, examining whether AI-delivered negative performance feedback leads to higher employee motivation to improve performance compared to human managers. Building on Experiment 1's test of the main effect, Experiment 2 adopted a sampling strategy that recruited employees from various industries and positions through a survey platform (i.e., posting experimental information and requirements on the platform to recruit willing employee participants) to test Hypothesis 2 and further explore the moderating effect of task type. Additionally, Experiment 3 used a more realistic performance feedback format—work emails—to provide feedback information and further test the mediating effects of internal and external attributions (Hypotheses 3 and 4). Finally, to further enhance feedback quality and increase its relative authenticity, Experiment 4 provided participants with more realistic negative performance feedback (based on actual individual task performance with more specific and accurate feedback). To enhance the applicability of the research findings, this study employed different types of AI agents to deliver negative performance feedback. Specifically, Experiment 1 used embedded AI, while Experiments 2–4 used robotic AI.

Experiment 1: The Effect of Human-AI Negative Performance Feedback on Performance Improvement Motivation

2.1.1 Participants

We used G*Power 3.1 (Faul et al., 2007) to calculate the required sample size for this experiment. For the one-way ANOVA applicable to this experiment, we set the effect size at f = 0.25 (medium), significance level at $\alpha = 0.05$, and number of groups at 2. The a priori analysis indicated that at least 128 participants were needed to achieve 80% statistical power. The experiment was posted on the Credamo platform, and 12 samples that failed attention checks, showed patterned responses, or had excessively short completion times were removed in real-time. Additional participants were recruited until we obtained 128 valid responses, including 71 females (55.5%) with a mean age of 33.95 years (SD = 8.24). Participants were randomly assigned to either the human manager or AI negative performance feedback group, with 63 participants in the human manager group and 65 in the AI group. All participants voluntarily joined the experiment and provided informed consent. Participants who passed attention checks and completed the experiment received monetary compensation.

2.1.2 Procedure

Experiment 1 used a single-factor, two-level between-subjects design: human-AI (human manager vs. AI) negative performance feedback. Participants were randomly assigned to one of the two experimental groups. First, considering that self-efficacy levels might influence individuals' attitudes after receiving negative performance feedback (Kluger & Denisi, 1996), and to control for this extraneous variable, participants were asked to complete a 10-item general self-efficacy scale (Scholz et al., 2002) using a 7-point rating scale ranging from "1 = strongly disagree" to "7 = strongly agree" (Cronbach's $\alpha = 0.91$ for this measure in Experiment 1). Participants read a scenario adapted from Tong et al. (2021), which described the daily work context of call center sales representatives in a company and informed participants that they were currently a sales representative in this company. To evaluate the performance of call center sales representatives, the company established a quality control department that recorded and analyzed sales representatives' service calls and provided performance feedback at a fixed time each week through the company's internal performance feedback system.

Previous research has indicated that individuals may reduce their trust in AI due to the opacity of its rules or principles (Glikson et al., 2020). Therefore, in the AI group, participants were provided with appropriate information about AI's attributes and were informed (differences between human and AI feedback are shown in bold): "The quality control department's AI system Xiao Ai (a program based on algorithmic systems, developed by evaluation experts who designed the evaluation criteria, and AI scholars and computer experts to provide performance feedback) provides professional performance feedback to sales

representatives through analysis of call recordings and sales data." In the human manager group, participants were correspondingly informed: "The quality control department's sales manager Xiao Ai (who has undergone systematic performance management training and possesses professional knowledge and work experience) provides professional performance feedback to sales representatives through analysis of call recordings and sales data" (see Appendix 2 for image materials). Subsequently, to test the manipulation of human-AI negative performance feedback, participants completed a manipulation check item ("Who just provided your performance feedback?").

Next, participants in different groups received negative performance feedback from either "Sales Manager Xiao Ai" or "AI System Xiao Ai": "Your work performance is below the departmental average, and you are currently among the lower-performing employees in the department. We hope you will continue to improve" (Belschak & Den Hartog, 2009). After reading the feedback, participants reported their performance improvement motivation using a two-item scale adapted from Wexley et al. (1973) (used in Experiments 1–4): "To what extent do you want to achieve higher performance goals in future work after receiving performance feedback?" and "To what extent do you want to improve your performance in future work after receiving performance feedback?" (rated on a 7-point scale from "1 = not at all" to "7 = very much"). Cronbach's α for this measure in Experiment 1 was 0.84.

Considering that participants in the AI group might have varying levels of familiarity with AI, which could influence their performance improvement motivation after receiving AI feedback, we controlled for this potential effect by asking participants to report their familiarity with AI using a two-item measure adapted from Leo and Huh (2020): "How often do you interact with artificial intelligence in your daily work or life?" and "How familiar are you with the working principles and operating mechanisms of artificial intelligence?" (rated from "1 = not at all familiar" to "7 = very familiar"). Additionally, two attention check items randomly appeared during material reading or questionnaire completion ("For this question, please select 'strongly disagree' " to screen out participants who were not paying attention). Finally, participants reported demographic information including gender and age.

2.2.1 Manipulation Check

To test the effectiveness of the human-AI negative performance feedback manipulation, participants were asked to recall the feedback provider after reading the experimental materials: "Please recall who just provided your performance feedback." All 128 retained participants answered correctly, indicating that the manipulation of human-AI negative performance feedback in Experiment 1 was successful.

2.2.2 Hypothesis Testing

Independent samples t-test results revealed that participants in the AI negative performance feedback group (M = 5.49, SD = 1.18) reported stronger performance improvement motivation than those in the human manager group (M = 4.94, SD = 1.38), t(126) = 2.38, p = 0.019, Cohen's d = 0.43. To verify the robustness of this result, we conducted a one-way ANOVA controlling for participants' self-efficacy. The results indicated that the AI group still showed higher performance improvement motivation than the human manager group, F(1, 127) = 5.97, p = 0.016, ^{2} = 0.046. Next, to further rule out potential effects of participants' gender (male = 1; female = 2) and age on the results, we conducted correlation analysis and independent samples t-tests respectively. The results showed that participants' age was not correlated with performance improvement motivation (r = 0.07, p = 0.447), and there was no significant difference in performance improvement motivation between males (M = 5.18, SD = 1.34) and females (M = 5.25, SD = 1.29), t(126) = 0.34, p = 0.738. Finally, to exclude the potential influence of participants' familiarity with AI on the results, correlation analysis revealed that AI group participants' familiarity with AI was not significantly correlated with performance improvement motivation (r = 0.09, p = 0.47). Therefore, Hypothesis 1 was supported.

Experiment 2: The Interactive Effect of Human-AI Negative Performance Feedback and Task Type

Experiment 1 provided initial support for Hypothesis 1 in a call center sales performance feedback context, demonstrating that AI-delivered negative performance feedback leads to higher individual performance improvement motivation compared to human managers. To verify the robustness of Experiment 1's findings and further test Hypothesis 2, Experiment 2 recruited active employees from various enterprises and changed the performance feedback context (new employee training scenario) to examine the interactive effect of human-AI negative performance feedback and task type on employee performance improvement motivation. Additionally, while Experiment 1 used embedded algorithms as the AI experimental material, Experiments 2–4 adopted robotic AI as the experimental material for the AI group, given that robotic AI may work alongside employees and enter organizations in the future (Yam et al., 2023) and typically elicits higher levels of trust and experiential quality in human-AI interactions (Glikson & Woolley, 2020).

3.1.1 Pretest of Experimental Materials

Before formally conducting Experiment 2, to test the reliability of the self-developed task type (subjective vs. objective) and human-AI (human manager vs. AI) feedback stimulus images, we recruited 60 participants on the Credamo platform (29 males, 31 females, mean age = 29.07 years, SD = 7.97) to conduct a pretest of the experimental materials. Participants were randomly assigned

to either the subjective or objective task condition and completed three task questions (with at least 100 characters per response). The three questions in the subjective task involved conflict resolution, emergency response, and interpersonal communication—common issues faced in organizations. The three questions in the objective task involved personnel ranking, plan calculation, and sales forecasting (see Appendix 1 for details). Additionally, participants evaluated the task objectivity of both subjective and objective tasks ("To what extent do you consider the above tasks to be objective tasks?" 1 = not at all, 5 = to a great extent) and task difficulty ("Please evaluate the difficulty level of the above tasks" 1 = not at all difficult, 5 = very difficult). Independent samples t-test results showed that the objective task (M = 4.10, SD = 0.80) was rated significantly higher in task objectivity than the subjective task (M = 1.53, SD = 0.82), $t(58) = 12.25$, $p < 0.001$, Cohen's d = 0.80. Moreover, t-test results indicated no significant difference in difficulty between the two task types, $t(58) = 0.26$, $p = 0.25$. To test whether the two task types approximated real work scenarios, we used a 5-item scale (Fields et al., 2023) and asked participants to rate face validity, with a sample item being: "To what extent do you believe the actual content of this test is clearly relevant to daily work?" (1 = not at all, 5 = completely). The results showed high face validity for both task types, with a mean of 4.33 for the subjective task and 3.92 for the objective task. Therefore, the experimental task design was reasonable and aligned with real work scenarios.

Additionally, referencing experimental materials from Garvey et al. (2023), we tested whether the images of AI and human manager faces differed in facial attractiveness and eeriness (see Appendix 2 for image materials). Independent samples t-test results showed no significant difference in facial attractiveness between the two images, $t(59) = 0.24$, $p = 0.59$. The AI image exhibited only slight eeriness (M = 1.47, SD = 0.57). Therefore, the selection of stimulus images for Experiment 2 was reasonable.

3.1.2 Participants

We used G*Power 3.1 software (Faul et al., 2007) to calculate the required sample size for this experiment. For the two-way ANOVA applicable to this experiment, we set the effect size at f = 0.25 (medium), significance level at $\alpha = 0.05$, and number of groups at 4. The a priori analysis indicated that at least 146 participants were needed to achieve 85% statistical power. Experiment 2 commissioned Wenjuanwang to post experimental information and recruit active employee participants, importing experimental materials onto the platform to conduct online behavioral experiments. Considering potential incomplete or invalid responses, Experiment 2 recruited 168 active enterprise employees. After excluding 8 participants who failed attention checks, showed patterned responses, or had abnormal completion times, we obtained 160 valid responses, including 61 females (38.1%) with a mean age of 33.29 years (SD = 4.98). Participants primarily came from manufacturing, software, finance, education, and

fast-moving consumer goods industries and worked in management (55 participants, 34.4%), production operations (37 participants, 23.1%), R&D (10 participants, 6.3%), marketing (25 participants, 15.6%), and product design (33 participants, 20.6%). All participants voluntarily joined the experiment and provided informed consent. Participants who passed attention checks and completed the experiment received monetary compensation.

3.1.3 Procedure

Experiment 2 used a two-factor between-subjects design: 2 (human-AI negative performance feedback: human manager vs. AI) × 2 (task type: subjective vs. objective). Participants were randomly assigned to one of the four experimental groups. Before the formal experiment, participants completed the general self-efficacy scale (Scholz et al., 2002) as a control variable (Cronbach's $\alpha = 0.89$ for this measure in Experiment 2). Participants then read a scenario informing them that they were a new employee at a daily necessities company. After a monthly departmental work summary, the department arranged a vocational ability test for five new employees (including the participant) to better develop personalized training plans and provide a basis for subsequent job placement. Participants were told that the following question was a representative example from the test. In the subjective task group, participants completed a question about resolving interpersonal conflicts in the workplace; in the objective task group, participants completed a question about forecasting future product sales (see Appendix 1). To ensure participants engaged seriously with the scenario, they were required to provide at least 100 characters per response.

After completing the test question, participants were informed that they would receive feedback on their performance in 2 minutes. In the human manager group, participants were told (differences between human and AI feedback are shown in bold): "The company invited Wang Liang, an evaluation specialist from the human resources department, to assess and provide feedback on your performance. Evaluation specialist Wang Liang (a trained, knowledgeable, and experienced evaluation expert) will read your responses, assess their quality, conduct statistical ranking, and provide evaluative feedback on your test performance," accompanied by an image of Wang Liang (see Appendix 2). In the AI group, participants were correspondingly informed: "The company uses the AI evaluation assistant Xiao Ai, developed by the human resources evaluation center, to assess and provide feedback on your performance. The AI evaluation assistant Xiao Ai, based on an algorithmic system (developed by AI scholars and computer experts based on evaluation criteria designed by assessment experts), will automatically recognize and analyze your responses, assess their quality, conduct statistical ranking, and provide evaluative feedback on your test performance" (accompanied by an image of Xiao Ai, see Appendix 2). Subsequently, participants completed manipulation check items (e.g., "Who provided your test feedback?" 1 = evaluation specialist Wang Liang, 7 = AI evaluation assistant Xiao Ai; and "Please rate the objectivity of the test question you just completed"

1 = very subjective, 7 = very objective).

After approximately 2 minutes, participants received negative performance feedback from either "evaluation specialist Wang Liang" or "AI evaluation assistant Xiao Ai": "In this test, your performance was below 80% of your colleagues, placing you in the bottom 20%. Your performance needs improvement" (Kim & Kim, 2020). Finally, participants completed the performance improvement motivation scale (Cronbach's $\alpha = 0.84$ in Experiment 2) and reported demographic information including gender, age, industry, and position. Participants in the AI group also reported their familiarity with AI. Two attention check items randomly appeared during questionnaire completion ("For this question, please select 'strongly disagree' ") to screen out participants who were not paying attention.

3.2.1 Manipulation Check

First, to test the effectiveness of the human-AI negative performance feedback manipulation, participants were asked to recall the feedback provider after reading the experimental materials: "Please recall who provided your performance feedback just now" (1 = evaluation specialist Wang Liang, 7 = AI evaluation assistant Xiao Ai). The results showed that the AI negative performance feedback group (M = 6.24, SD = 1.05) rated significantly higher than the human manager group (M = 1.95, SD = 1.11), $t(158) = 25.11$, $p < 0.001$, Cohen's $d = 0.86$, indicating that the manipulation of human-AI negative performance feedback in Experiment 2 was successful.

Second, to test the effectiveness of the task type manipulation, participants were asked to rate the objectivity of the test question: "How objective do you think the test example you just completed was?" (1 = very subjective, 7 = very objective). The results showed that the objective task group (M = 5.96, SD = 0.79) rated significantly higher in task objectivity than the subjective task group (M = 2.33, SD = 0.87), $t(158) = 27.77$, $p < 0.001$, Cohen's $d = 0.83$, indicating that the task type manipulation was successful.

3.2.2 Hypothesis Testing

Independent samples t-test results revealed that participants in the AI negative performance feedback group (M = 5.67, SD = 0.79) reported stronger performance improvement motivation than those in the human manager group (M = 5.36, SD = 1.13), $t(158) = 2.00$, $p = 0.048$, Cohen's $d = 0.46$. A one-way ANOVA controlling for self-efficacy showed that the AI group still demonstrated higher performance improvement motivation than the human manager group, $F(1, 157) = 4.64$, $p = 0.033$, $^2 = 0.029$. To rule out potential effects of AI familiarity on the results, correlation analysis showed that AI group participants' familiarity with AI was not significantly correlated with performance improvement motivation ($r = 0.10$, $p = 0.40$). In summary, Hypothesis 1 was again supported.

Finally, we tested whether task type could serve as a boundary condition. Two-way ANOVA results indicated a significant interactive effect of human-AI negative performance feedback and task type on individuals' performance improvement motivation, $F(1, 156) = 39.65$, $p < 0.001$, $^2 = 0.203$. Simple effects analysis revealed (see Table 1 and Figure 1 [Figure 1: see original paper]) that in the objective task group, the AI negative performance feedback group ($M = 5.71$, $SD = 0.66$) showed significantly higher performance improvement motivation than the human manager group ($M = 4.60$, $SD = 1.01$), $F(1, 156) = 37.75$, $p < 0.001$, $^2 = 0.195$. In the subjective task group, the human manager negative feedback group ($M = 6.13$, $SD = 0.59$) showed significantly higher performance improvement motivation than the AI group ($M = 5.63$, $SD = 0.90$), $F(1, 156) = 7.63$, $p = 0.006$, $^2 = 0.047$. To test the robustness of these results, we controlled for participants' self-efficacy as a covariate and found that the interactive effect between human-AI negative performance feedback and task type remained significant, $F(1, 155) = 40.58$, $p < 0.001$, $^2 = 0.207$. Hypothesis 2 was supported.

Table 1 Means (Standard Deviations and Sample Sizes) of Performance Improvement Motivation for Human Manager and AI Negative Performance Feedback in Subjective and Objective Tasks

Task Type	Human Manager Negative Performance Feedback	AI Negative Performance Feedback
Subjective	6.13 (0.59; n = 40)	5.63 (0.90; n = 40)
Objective	4.60 (1.01; n = 40)	5.71 (0.66; n = 40)

Figure 1 [Figure 1: see original paper] Interactive Effect of Human-AI Negative Performance Feedback and Task Type on Performance Improvement Motivation

Experiment 3: The Mediating Role of Internal and External Attributions

Experiment 3 aimed to use a more realistic email-based performance feedback format to further test the robustness of the findings from Experiments 1 and 2 and to examine the mediating role of internal and external attributions.

4.1.1 Participants

We used G*Power 3.1 software (Faul et al., 2007) to calculate the required sample size for this experiment. For the two-way ANOVA applicable to this experiment, we set the effect size at $f = 0.25$ (medium), significance level at $\alpha = 0.05$, and number of groups at 4. The a priori analysis indicated that at least 146 participants were needed to achieve 85% statistical power. Similar to Experiment 2, Experiment 3 commissioned Wenjuanwang to post experimental

information and recruit active employee participants. Considering potential incomplete or invalid responses, Experiment 3 recruited 160 active employees. All participants voluntarily joined the experiment and provided informed consent. Participants who passed attention checks and completed the experimental tasks received monetary compensation. After excluding 10 participants who failed attention tests, did not complete responses, or provided invalid responses, the final valid sample for Experiment 3 was 150 participants, including 86 females (57.3%) with a mean age of 29.70 years (SD = 4.97) and average work experience of 6.17 years (SD = 4.23). Participants came from nine industries including internet, construction, manufacturing, information and communication, commodity sales, education, healthcare, finance, and services. In terms of job positions, participants included 52 management personnel (34.87%), 29 operations personnel (19.40%), 19 technical personnel (12.50%), 26 marketing personnel (17.43%), and 24 creative design personnel (15.79%).

4.1.2 Procedure

Experiment 3 used a two-factor between-subjects design: 2 (human-AI negative performance feedback: human manager vs. AI) × 2 (task type: subjective vs. objective). Participants were randomly assigned to one of the four experimental groups. Participants were informed that they would participate in a vocational ability competition. Their work email addresses were collected in advance to send corresponding performance feedback later. The competition consisted of two stages (competition and performance feedback). In the competition stage, participants needed to complete competition questions according to requirements. First, participants completed the self-efficacy scale as a control variable (Cronbach's $\alpha = 0.91$ for this measure in Experiment 3). Next, using materials from Experiment 2's pretest, participants completed three subjective or objective competition questions according to their group assignment (see Appendix for specific questions). To ensure participants engaged seriously with the scenario, they were required to provide at least 100 characters per response. After completing the competition tasks, participants were informed that a vocational assessment center at a university would be responsible for evaluating and providing feedback on their competition performance.

In the human manager group, participants were told (differences between human and AI feedback manipulations are shown in bold): "To evaluate your performance in this vocational ability competition, we invited Wang Liang, the head of the assessment center at a university, to assess and provide feedback on your performance. Assessment center head Wang Liang (a trained, knowledgeable, and experienced assessment expert) will read your responses, assess their quality, conduct statistical ranking, and provide evaluative feedback on your competition results," accompanied by an image of Wang Liang (see Appendix 2). In the AI group, participants were correspondingly informed: "To evaluate your performance in this vocational ability competition, we will use the AI assessment assistant Xiao Ai, newly introduced by the assessment center

at a university. The AI assessment assistant Xiao Ai, based on an algorithmic system (developed by AI scholars and computer experts based on evaluation criteria designed by assessment experts), will automatically recognize and analyze your responses, assess their quality, conduct statistical ranking, and provide evaluative feedback on your competition results" (accompanied by an image of Xiao Ai, see Appendix 2). Subsequently, participants completed manipulation check items ("Who provided your competition feedback?" 1 = assessment center head Wang Liang, 7 = assessment center AI assistant Xiao Ai; and "Please rate the objectivity of the competition questions you just completed" 1 = very subjective, 7 = very objective). Participants were then informed: "Since we need to wait for and evaluate other participants' performance and conduct final ranking, performance feedback will take approximately 20 minutes. The final competition results and questionnaire link will be sent to your email address." To reduce interference from potential attention loss during the waiting period, all participants were asked to watch a 20-minute introductory video about the university's assessment center.

In the performance feedback stage, experimenters sent competition feedback results to participants' work email addresses using pre-prepared email accounts (e.g., assessment center head Wang Liang or AI assessment assistant Xiao Ai). To strengthen the manipulation of human-AI negative performance feedback, participants in the human manager group received: "Hello! In this vocational ability competition, your performance was below 82% of participants, placing you in the bottom 18%" (Kim & Kim, 2020). In addition to receiving the same negative performance feedback content as the human manager group, participants in the AI group also saw a note at the end of the email: "This email was automatically sent by an AI assistant. Please do not reply." Participants were then prompted to complete the second-stage questionnaire attached to the email, which included recalling and briefly describing the performance feedback content (to ensure subsequent responses were based on the feedback), reporting AI familiarity, performance improvement motivation (Cronbach's $\alpha = 0.84$ for this measure in Experiment 3), and internal and external attribution measured using Russell's (1982) 6-item scale. Three items measured internal attribution (sample item: "To what extent do you believe the assessment feedback provided by the evaluator was based on your personal effort?"; Cronbach's $\alpha = 0.78$). External attribution was measured with items such as "To what extent do you believe the assessment feedback provided by the evaluator was based on environmental factors (e.g., difficult questions)?" (Cronbach's $\alpha = 0.74$). All items used a 7-point rating scale from "1 = not at all" to "7 = to a great extent."

Previous research has found that individuals may perceive differences in the accuracy (Tong et al., 2021) and fairness (Newman et al., 2020) of performance feedback from humans or AI. For example, because AI is essentially a data-driven program model with higher objectivity and impartiality, AI (relative to human managers) delivering negative performance feedback may elicit higher perceptions of accuracy or fairness (Jiang et al., 2022), thereby differentially influencing performance improvement motivation. To rule out these two

alternative explanatory mechanisms, participants completed scales measuring feedback accuracy (Brett & Atwater, 2001) and fairness (Chory & Westerman, 2009). Feedback accuracy was measured with two items ("To what extent do you believe the feedback you received was an accurate assessment of your performance?" and "To what extent do you believe the feedback you received was correct?" rated on a 7-point scale from "1 = not at all" to "7 = very much"; Cronbach's $\alpha = 0.89$). Fairness perception was measured with six items (sample items: "I believe the feedback provided by the evaluator was: 1 = unfair; 7 = fair" and "I believe the feedback provided by the evaluator was: 1 = biased; 7 = impartial"; Cronbach's $\alpha = 0.97$). Finally, participants reported demographic variables including gender, age, work experience, industry, and position. Two attention check items randomly appeared during questionnaire completion ("For this question, please select 'strongly disagree' ") to screen out participants who were not paying attention.

4.2.1 Manipulation Check

First, to test the effectiveness of the human-AI negative performance feedback manipulation, participants were asked to recall the feedback provider after reading the experimental materials: "Please recall who provided your performance feedback just now" (1 = assessment center head Wang Liang, 7 = assessment center AI assistant Xiao Ai). The results showed that the AI negative performance feedback group (M = 5.26, SD = 1.47) rated significantly higher than the human manager group (M = 2.49, SD = 0.86), $t(148) = 13.10$, $p < 0.001$, Cohen's $d = 0.71$, indicating that the manipulation of human-AI negative performance feedback in Experiment 3 was successful.

To test the effectiveness of the task type manipulation, participants rated the objectivity of the experimental task: "How objective do you think the competition questions you just completed were?" The results showed that the objective task group (M = 5.54, SD = 0.98) rated significantly higher in task objectivity than the subjective task group (M = 3.41, SD = 1.95), $t(148) = 8.23$, $p < 0.001$, Cohen's $d = 0.63$, indicating that the task type manipulation was successful.

4.2.2 Hypothesis Testing

Independent samples t-test results revealed that participants in the AI negative performance feedback group (M = 5.06, SD = 1.21) reported stronger performance improvement motivation than those in the human manager group (M = 4.66, SD = 1.10), $t(148) = 2.10$, $p = 0.037$, Cohen's $d = 0.44$. Controlling for self-efficacy, we still found that AI-delivered negative performance feedback led to higher performance improvement motivation compared to human managers, $F(1, 147) = 4.05$, $p = 0.046$, $^{2} = 0.027$. Hypothesis 1 was supported. Next, to rule out potential effects of AI familiarity on the results, correlation analysis showed that AI group participants' familiarity with AI was not significantly correlated with performance improvement motivation ($r = 0.06$, $p = 0.61$).

To test Hypothesis 2, two-way ANOVA results (see Table 2 and Figure 2) revealed a significant interactive effect of human-AI negative performance feedback and task type on individuals' performance improvement motivation, $F(1, 146) = 20.00$, $p < 0.001$, $^2 = 0.120$. Simple effects analysis showed that in the objective task group, the AI negative feedback group ($M = 5.60$, $SD = 0.88$) exhibited significantly higher performance improvement motivation than the human manager group ($M = 4.35$, $SD = 0.92$), $F(1, 146) = 24.47$, $p < 0.001$, $^2 = 0.144$. In the subjective task group, the human manager negative feedback group ($M = 4.92$, $SD = 1.17$) showed marginally significantly higher performance improvement motivation than the AI group ($M = 4.57$, $SD = 1.26$), $F(1, 146) = 3.00$, $p = 0.085 < 0.10$, $^2 = 0.02$. To test the robustness of these results, we controlled for participants' self-efficacy as a covariate and found that the interactive effect between human-AI negative performance feedback and task type remained significant, $F(1, 145) = 22.79$, $p < 0.001$, $^2 = 0.136$. Hypothesis 2 was again supported.

Table 2 Means (Standard Deviations and Sample Sizes) of Performance Improvement Motivation for Human Manager and AI Negative Performance Feedback in Subjective and Objective Tasks

Task Type	Human Manager Negative Performance Feedback	AI Negative Performance Feedback
Subjective	4.92 (1.17; n = 41)	4.57 (1.26; n = 40)
Objective	4.35 (0.92; n = 33)	5.60 (0.88; n = 36)

Figure 2 [Figure 2: see original paper] Interactive Effect of Human-AI Negative Performance Feedback and Task Type on Performance Improvement Motivation

To test the mediating effects of internal and external attributions, we used PROCESS macro Model 4 with 2,000 bootstrap samples. The results indicated that the indirect effect of internal attribution in the relationship between human-AI negative performance feedback and performance improvement motivation was significant, with an effect size of 0.17 and a 95% confidence interval of [0.015, 0.350] that did not include zero. Additionally, the indirect effect of external attribution in the relationship between human-AI negative performance feedback and performance improvement motivation was 0.08, with a 95% confidence interval of [-0.04, 0.21] that included zero, indicating that the indirect effect of external attribution was not significant. Hypothesis 3 was partially supported.

Furthermore, we tested the indirect effects of internal and external attributions in the interactive effect of human-AI negative performance feedback and task type on performance improvement motivation using PROCESS macro Model 8 with 2,000 bootstrap samples. The results showed that in subjective tasks, the indirect effect of internal attribution in the relationship between human-AI negative performance feedback and performance improvement motivation was significant, with a 95% confidence interval of [-0.40, -0.03]. In objective

tasks, the indirect effect of internal attribution was also significant, with a 95% confidence interval of [0.20, 0.84]. The difference in the moderated indirect effects between the two task types was significant, with an effect size of 0.67 and a 95% confidence interval of [0.277, 1.178], indicating a significant moderated mediation effect. However, the moderated indirect effect of external attribution was 0.12, with a 95% confidence interval of [-0.065, 0.375] that included zero. Therefore, Hypothesis 4 was also partially supported.

To test whether feedback accuracy and fairness could serve as alternative explanatory mechanisms, independent samples t-tests revealed no difference in accuracy between human manager (M = 3.78, SD = 1.63) and AI-delivered negative performance feedback (M = 3.80, SD = 1.60), t(148) = 0.07, p = 0.94. However, AI-delivered negative performance feedback (M = 5.31, SD = 1.35) was perceived as fairer than human manager feedback (M = 4.84, SD = 1.50), t(148) = 2.00, p = 0.047, Cohen's d = 0.32. Additionally, the interactive effects of human-AI negative performance feedback and task type on feedback accuracy and fairness were not significant, F(1, 146) = 0.12, p = 0.73 and F(1, 146) = 0.58, p = 0.45, respectively. Finally, the indirect effects of feedback accuracy (95% CI [-0.082, 0.065]) and fairness (95% CI [-0.06, 0.24]) in the relationship between human-AI negative performance feedback and performance improvement motivation were both non-significant. In summary, Experiment 3 ruled out feedback accuracy and fairness as alternative explanatory mechanisms.

Experiment 4: Relatively Real Negative Performance Feedback

Experiment 3 used a more organizationally realistic feedback method to test the robustness of Experiments 1 and 2 and further identified the mediating role of internal attribution. This provided a good explanatory mechanism for the differential effects of AI and human managers delivering negative performance feedback on employee performance improvement motivation across different task contexts. Moreover, Experiment 3 ruled out alternative explanatory mechanisms of feedback accuracy and fairness.

Experiment 3 failed to find a mediating effect for external attribution, possibly because individuals, for self-esteem maintenance and self-defense purposes, tend to exhibit some level of external attribution regardless of whether negative performance feedback comes from human managers or AI (Hareli & Hess, 2008). For example, in Experiment 3, both the human manager group (M = 3.60, SD = 1.27) and the AI negative performance feedback group (M = 3.92, SD = 1.33) showed some external attribution, but the difference was not significant, t(148) = 1.52, p = 0.13. Consistent with Experiment 3's results, Yalcin et al. (2022) also found no significant difference in external attribution between feedback from human or AI customer service in unfavorable decision contexts (e.g., company rejection).

Furthermore, Experiments 1–3 used the false feedback paradigm commonly em-

ployed in performance feedback research (Cianci et al., 2010; Kim & Kim, 2020), which has the advantage of controlling feedback content consistency across participants. However, because participants' actual task performance was not assessed, individuals might perceive the negative performance feedback as less accurate and realistic. To address this issue and further enhance feedback quality, Experiment 4 aimed to provide more specific and personalized performance feedback based on participants' actual task performance, thereby re-testing the overall model in a relatively realistic feedback context.

5.1.1 Participants

We used G*Power 3.1 software (Faul et al., 2007) to calculate the required sample size for this experiment. For the two-way ANOVA applicable to this experiment, we set the effect size at f = 0.25 (medium), significance level at $\alpha = 0.05$, and number of groups at 4. The a priori analysis indicated that at least 146 participants were needed to achieve 85% statistical power. Similar to Experiments 2 and 3, Experiment 4 commissioned Wenjuanwang to post experimental information and recruit active employee participants. Considering potential incomplete or invalid responses, Experiment 4 recruited 166 active employees. All participants voluntarily joined the experiment and provided informed consent. Participants who passed attention checks and completed the experimental tasks received monetary compensation. After excluding 6 participants who failed attention tests, did not complete responses, or provided invalid responses, the final valid sample for Experiment 4 was 160 participants, including 65 females (40.60%) with a mean age of 29.54 years (SD = 6.07). Approximately 91.3% of participants had bachelor's degree or higher education. Participants primarily came from five industries: manufacturing, software, business services, finance, and scientific research and education. In terms of job positions, participants included 50 R&D personnel (31.30%), 34 management personnel (21.30%), 31 production and operations personnel (19.40%), and 26 marketing personnel (16.30%).

5.1.2 Procedure

Experiment 4 used a two-factor between-subjects design: 2 (human-AI negative performance feedback: human manager vs. AI) × 2 (task type: subjective vs. objective). Participants were randomly assigned to one of the four experimental groups. Two preparatory steps were taken before the formal experiment. First, five experimenters were trained to clearly and proficiently master the key points or evaluation criteria for each question. Second, negative performance feedback templates were prepared in advance and personalized based on participants' performance during formal responding to control feedback content and reduce feedback delivery time.

In the formal experiment stage, participants were informed that they were a manager at a paint company about to participate in a management capacity assessment for middle managers, which would provide reference for subsequent

training and learning. The assessment was divided into two stages (task and performance feedback). In the assessment task stage, participants needed to complete the following steps: First, participants completed the self-efficacy scale as a control variable (Cronbach's $\alpha = 0.96$ for this measure in Experiment 4). Second, similar to Experiments 2–3, participants in different task groups completed two subjective or objective tasks (see Appendix 1), with at least 100 characters required per response. Third, after completing the assessment tasks, participants were informed that the company would conduct professional evaluation and feedback on their assessment performance.

In the human manager group, participants were told (differences between human and AI feedback manipulations are shown in bold): "To evaluate your performance in this management capacity assessment, we invited Wang Liang, a management capacity assessment specialist from the company, to assess and provide feedback on your performance. Assessment specialist Wang Liang (a trained, knowledgeable, and experienced assessment expert) will read your responses, assess their quality, conduct statistical ranking, and provide evaluative feedback on your management capacity assessment results," accompanied by an image of Wang Liang (see Appendix 2). In the AI group, participants were correspondingly informed: "To evaluate your performance in this vocational ability competition, we will use Xiao Ai, the latest AI assessment assistant introduced by the company's HR department, to assess and provide feedback on your performance. AI assessment assistant Xiao Ai, based on an algorithmic system (developed by AI scholars and computer experts based on evaluation criteria designed by assessment experts), will automatically recognize and analyze your responses, assess their quality, conduct statistical ranking, and provide evaluative feedback on your management capacity assessment results" (accompanied by an image of Xiao Ai, see Appendix 2). Subsequently, participants completed a task type manipulation check item: "To what extent do you think the assessment questions you completed belong to objective tasks?" (1 = very subjective to 7 = very objective).

In the performance feedback stage, referencing Goodman and Wood (2004), feedback quality was enhanced by explaining task purposes and providing specific encouragement or improvement suggestions. First, participants were asked to log in to SalesSmartly (a professional real-time chat interaction website for enterprises and personnel) to receive near real-time performance feedback from experimenters acting as either human managers or AI. Participants could receive assessment feedback by entering their corresponding ID number. Second, regarding feedback content, experimenters first provided participants with an overall evaluation based on the detail, logic, and clarity of their responses, such as: "Dear Participant XX, thank you for completing the management capacity assessment questions. Overall, your responses were somewhat vague (or clear)." Next, participants were explained the purpose of the assessment and the specific abilities being tested. For the subjective task group: "The first official document in this assessment aims to evaluate your conflict resolution ability in team building, while the second aims to test your problem-solving

ability in handling unexpected team events." For the objective task group: "The first official document in this assessment aims to evaluate your computational analysis ability in raw material procurement, while the second aims to test your logical reasoning ability in sales forecasting." Subsequently, based on pre-organized response key points, experimenters scored participants' responses to the two questions and provided specific suggestions. For example: "Compared to other participants, you demonstrated weaker conflict resolution ability [54.15/100] (with specific response deficiencies analyzed and listed). However, you demonstrated better unexpected event resolution ability [69.75/100] (with specific response strengths analyzed and listed)." Additionally, to provide clear negative performance feedback and better control content, all participants uniformly received: "Overall, your total score in this assessment was [61.95/100], which is below 82% of participants and places you in the bottom 18%. Your performance needs improvement." To demonstrate differences between human and AI feedback, assessment specialist Wang Liang thanked participants for their participation and cooperation in the human manager group, while Xiao Ai thanked participants for their usage in the AI group. Finally, to prevent extraneous variables, all negative performance feedback followed the same format and was controlled to approximately 200 characters.

After reading the feedback, participants completed the second-stage questionnaire. To ensure they read the feedback carefully and based their subsequent responses on it, participants were asked to recall and briefly describe the feedback content they received. Participants then completed questionnaire items including manipulation checks for feedback provider (1 = assessment specialist Wang Liang; 7 = assessment assistant Xiao Ai) and feedback content (1 = very negative; 5 = very positive), AI familiarity, performance improvement motivation (Cronbach's $\alpha = 0.89$ for this measure in Experiment 4), internal attribution (Cronbach's $\alpha = 0.79$), and external attribution (Cronbach's $\alpha = 0.83$; same items as Experiment 3). Finally, similar to Experiment 3, considering that feedback fairness (Cronbach's $\alpha = 0.95$) and accuracy (Cronbach's $\alpha = 0.87$) might serve as alternative mediators, participants completed scales measuring these two variables. Two attention check items randomly appeared during questionnaire completion ("For this question, please select 'strongly disagree' ") to screen out participants who were not paying attention.

5.2.1 Manipulation Check

First, to test the effectiveness of the negative performance feedback manipulation, participants were asked to recall the feedback content they received: "How would you rate the feedback you received regarding your assessment performance?" (1 = very negative, 5 = very positive). The results showed that participants' mean rating of feedback content was 2.01 (SD = 0.97), indicating that the negative performance feedback manipulation in Experiment 4 was successful.

Second, to test the effectiveness of the human-AI negative performance feedback

manipulation, participants were asked to recall the feedback provider after reading the experimental materials: "Please recall who provided your performance feedback just now" (1 = assessment center head Wang Liang, 7 = assessment center AI assistant Xiao Ai). The results showed that the AI negative performance feedback group (M = 5.71, SD = 1.72) rated significantly higher than the human manager group (M = 1.60, SD = 1.33), t(158) = 16.92, p < 0.001, Cohen's d = 0.80, indicating that the manipulation of human-AI negative performance feedback in Experiment 4 was successful.

Finally, to test the effectiveness of the task type manipulation, participants rated the objectivity of the experimental task. The results showed that the objective task group (M = 5.49, SD = 1.23) rated significantly higher in task objectivity than the subjective task group (M = 3.41, SD = 1.51), t(158) = 9.53, p < 0.001, Cohen's d = 0.60, indicating that the task type manipulation was successful.

5.2.2 Hypothesis Testing

Independent samples t-test results revealed that participants in the AI negative performance feedback group (M = 6.07, SD = 0.78) reported stronger performance improvement motivation than those in the human manager group (M = 5.76, SD = 0.97), t(158) = 2.24, p = 0.027, Cohen's d = 0.35. Controlling for self-efficacy, we still found that AI-delivered negative performance feedback led to higher performance improvement motivation compared to human managers, F(1, 157) = 4.98, p = 0.027, ^{2} = 0.031. Hypothesis 1 was supported. Next, to rule out potential effects of AI familiarity on the results, correlation analysis showed that AI group participants' familiarity with AI was not significantly correlated with performance improvement motivation (r = 0.058, p = 0.61).

To test Hypothesis 2, two-way ANOVA results (see Table 3 and Figure 3) revealed a significant interactive effect of human-AI negative performance feedback and task type on individuals' performance improvement motivation, F(1, 156) = 44.76, p < 0.001, ^{2} = 0.223. Simple effects analysis showed that in the objective task group, the AI negative feedback group (M = 6.19, SD = 0.72) exhibited significantly higher performance improvement motivation than the human manager group (M = 5.09, SD = 0.81), F(1, 156) = 43.66, p < 0.001, ^{2} = 0.219. In the subjective task group, the human manager negative feedback group (M = 6.43, SD = 0.61) showed significantly higher performance improvement motivation than the AI group (M = 5.95, SD = 0.82), F(1, 156) = 8.14, p = 0.005, ^{2} = 0.05. To test the robustness of these results, we controlled for participants' self-efficacy as a covariate and found that the interactive effect between human-AI negative performance feedback and task type remained significant, F(1, 155) = 44.66, p < 0.001, ^{2} = 0.224. Hypothesis 2 was again supported.

Table 3 Means (Standard Deviations and Sample Sizes) of Performance Improvement Motivation for Human Manager and AI Negative Performance Feed-

back in Subjective and Objective Tasks

Task Type	Human Manager Negative Performance Feedback	AI Negative Performance Feedback
Subjective	6.43 (0.61; n = 40)	5.95 (0.82; n = 40)
Objective	5.09 (0.81; n = 40)	6.19 (0.72; n = 40)

Figure 3 [Figure 3: see original paper] Interactive Effect of Human-AI Negative Performance Feedback and Task Type on Performance Improvement Motivation

To test the mediating effects of internal and external attributions, we used PROCESS macro Model 4 with 2,000 bootstrap samples. The results indicated that the indirect effect of internal attribution in the relationship between human-AI negative performance feedback and performance improvement motivation was significant, with an effect size of 0.17 and a 95% confidence interval of [0.017, 0.359] that did not include zero. Additionally, the indirect effect of external attribution in the relationship between human-AI negative performance feedback and performance improvement motivation was -0.10, with a 95% confidence interval of [-0.086, 0.011] that included zero, indicating that the indirect effect of external attribution was not significant. Hypothesis 3 was partially supported.

Furthermore, we tested the indirect effects of internal and external attributions in the interactive effect of human-AI negative performance feedback and task type on performance improvement motivation using PROCESS macro Model 8 with 2,000 bootstrap samples. The results showed that in subjective tasks, the indirect effect of internal attribution in the relationship between human-AI negative performance feedback and performance improvement motivation was significant, with a 95% confidence interval of [-0.373, -0.035]. In objective tasks, the indirect effect of internal attribution was also significant, with a 95% confidence interval of [0.241, 0.669]. The difference in the moderated indirect effects between the two task types was significant, with an effect size of 0.63 and a 95% confidence interval of [0.349, 0.989], indicating a significant moderated mediation effect. However, the moderated indirect effect of external attribution was 0.013, with a 95% confidence interval of [-0.017, 0.110] that included zero. Hypothesis 4 was partially supported.

To test whether feedback accuracy and fairness could serve as alternative explanatory mechanisms, independent samples t-tests revealed no difference in accuracy between human manager ($M = 5.33$, $SD = 0.91$) and AI-delivered negative performance feedback ($M = 5.48$, $SD = 0.99$), $t(158) = 0.99$, $p = 0.32$. However, AI-delivered negative performance feedback ($M = 6.14$, $SD = 1.21$) was perceived as fairer than human manager feedback ($M = 5.74$, $SD = 1.28$), $t(158) = 2.01$, $p = 0.046$, Cohen's $d = 0.33$. Additionally, the interactive effects of human-AI negative performance feedback and task type on feedback accuracy and fairness were not significant, $F(1, 156) = 0.45$, $p = 0.51$ and $F(1, 156) = 0.20$, $p = 0.65$, respectively. Finally, the indirect effects of feedback accuracy

(effect size = 0.02; 95% CI [-0.012, 0.105]) and fairness (effect size = 0.03; 95% CI [-0.075, 0.051]) in the relationship between human-AI negative performance feedback and performance improvement motivation were both non-significant. In summary, similar to Experiment 3, Experiment 4 ruled out feedback accuracy and fairness as alternative explanatory mechanisms.

6 General Discussion

Grounded in attribution theory, this study employed four sequential experiments to uncover the differential effects and mechanisms of human-AI negative performance feedback on performance improvement motivation. Specifically, the study found that compared to human managers, AI-delivered negative performance feedback led to higher levels of employee motivation to improve performance. Second, human-AI negative performance feedback and task type interactively influenced individuals' performance improvement motivation. In subjective tasks, individuals exhibited stronger motivation to improve performance in response to negative feedback from human managers compared to AI, whereas the opposite pattern emerged in objective tasks. Additionally, based on attribution theory, this study further revealed that internal attribution mediated the interactive effect of human-AI negative performance feedback and task type on performance improvement motivation. The study also employed different types of AI agents (embedded AI in Experiment 1 and robotic AI in Experiments 2–4), varied performance feedback contexts (call center sales representatives in Experiment 1, new employee training in Experiment 2, employee vocational ability competitions in Experiment 3, and middle managers' management capacity assessments in Experiment 4), different performance feedback strategies (fake feedback in Experiments 1–3 and relatively real feedback in Experiment 4), and diverse feedback delivery channels (online experimental platform display in Experiments 1–2, real email delivery in Experiment 3, and real-time conversational interaction website in Experiment 4). Overall, the results across the four experiments demonstrated strong consistency and robustness.

6.1 Theoretical Contributions

First, this study expands the perspective of existing negative performance feedback research. Specifically, whereas traditional negative feedback research focusing on interpersonal interactions has primarily examined feedback from human managers (Kitz et al., 2023), this study uncovers the potential positive effects of AI replacing human managers in delivering negative performance feedback. Previous research has explored various approaches to enhancing the effectiveness of negative performance feedback implementation, such as feedback characteristics (frequency, immediacy, or quality of performance feedback) (Kuvaas et al., 2017; Ni & Zheng, 2024) and employee individual factors (positive attribution of negative performance feedback, employee core self-evaluation, etc.) (Ma et al., 2021; Xing et al., 2023). By integrating the context of the digital

intelligence era and adopting the novel perspective of human-AI negative performance feedback, this study finds that AI (compared to human managers) delivering negative performance feedback enhances employees' subsequent motivation to improve performance, thereby providing empirical evidence for the differential effects of human versus AI negative performance feedback.

Second, this study enriches existing human-AI feedback research. Current debates exist regarding the application of digital intelligence technology in performance feedback (Dong et al., 2022). On one hand, from an algorithm appreciation perspective, researchers have found that AI can enhance the accuracy and reliability of performance feedback, thereby improving employee performance (Tong et al., 2021). On the other hand, from an algorithm aversion perspective, research has found that AI lacks sincerity and uniqueness and threatens human job opportunities, so when organizations disclose that performance feedback (especially positive feedback with encouraging or praising nature) originates from AI (Yalcin et al., 2022), it reduces individuals' positive performance (Tong et al., 2021; Luo et al., 2019). This study focuses on negative performance feedback and finds that AI (relative to human managers) as a feedback provider enhances individuals' performance improvement motivation. Additionally, existing research has examined boundary conditions for differential effects of human-AI feedback. For example, Tong et al. (2021) found that for employees with longer tenure, stronger emotional bonds with the organization lead to greater support for organizational changes involving AI-delivered performance feedback, thus mitigating the negative effects of AI-delivered feedback. Furthermore, Luo et al. (2019) found that customers' familiarity with AI reduces stereotypes about AI (e.g., lacking knowledge and empathy), thereby mitigating the decline in product sales caused by AI-delivered feedback. This study focuses on task type as an external factor of employees' work and finds the interactive effect of human-AI negative performance feedback and task type on employee performance improvement motivation, thereby expanding research on boundary conditions of human-AI feedback.

Moreover, this study contributes to agile performance management research in the digital intelligence era. Specifically, traditional performance management models with annual or quarterly cycles suffer from excessive length, which is not conducive to employees obtaining timely information and improving performance. Scholars have therefore proposed the transformation trend of agile performance management (Pulakos et al., 2019; Schleicher et al., 2018), aiming to enhance the timeliness of performance management and provide employees with accurate, high-quality performance evaluation and feedback. Digital intelligence is the most important factor in enhancing agile performance management, as AI can tirelessly integrate and analyze data to evaluate employee performance objectively and impartially and provide more accurate performance feedback (Qin et al., 2023; Tong et al., 2021). In addition to human-AI performance feedback research, studies have also examined AI coaching. For example, Luo et al. (2021) found that AI coaching relative to human coaching showed an inverted U-shaped distribution of guidance effectiveness across different sales per-

sonnel because lower-performing salespeople faced information overload from AI feedback, while higher-performing salespeople exhibited stronger aversion to AI. This study is consistent with the above literature in exploring the impact and mechanisms of digital intelligence technology on specific aspects of performance management.

Finally, this study deepens the application of attribution theory in organizational contexts. Attribution theory is widely used to explain how individuals understand the causes of their own or others' behaviors in interpersonal interactions (Tolli & Schmidt, 2008). According to classical attribution theory (Heider, 1958), people typically make external attributions for unfavorable outcomes and internal attributions for favorable outcomes for self-defense and self-enhancement purposes. However, these conclusions are moderated by some factors. For example, Xing et al. (2023) found that employees with higher core self-evaluation are more likely to view negative performance feedback as an opportunity to improve performance, thereby increasing internal attribution and learning performance. This study deeply explores the differential effects of human-AI feedback and finds that AI (relative to human managers) delivering negative performance feedback may enhance individuals' internal attribution. This is explained by combining different feedback characteristics of humans and AI (e.g., AI has fewer subjective or harmful intentions compared to humans) (Jiang et al., 2022). The findings suggest that when reducing the negative impact of negative stimuli (e.g., using AI instead of human managers for negative performance feedback), individuals may strengthen internal attribution. This provides new insights for attribution theory in explaining individuals' attribution tendencies or behavioral responses to unfavorable outcomes.

6.2 Management Implications

This study also offers several management implications. First, traditional human manager-dominated negative performance feedback may damage leader-subordinate relationships, trigger negative employee emotions, and reduce performance levels (Ni & Zheng, 2024). This study's findings indicate that AI (relative to human managers) enhances individuals' internal attribution and performance improvement motivation. The results suggest that organizations can apply digital intelligence technology to empower performance management processes and leverage AI's objective and impartial performance feedback advantages. This can not only reduce the pressure on human managers to deliver negative performance feedback but also make negative performance feedback from AI more acceptable to employees, thereby enhancing feedback implementation effectiveness.

Second, although digital intelligence technology offers advantages such as efficiency, objectivity, and standardization, it reduces interpersonal interaction and empathy in the performance feedback process (Dong et al., 2022; Yalcin et al., 2022). Therefore, it is necessary to distinguish between different application scenarios for human and AI feedback. According to this study's results,

organizations should pay attention to task characteristics in human-AI negative performance feedback. For instance, AI's objective and impartial characteristics provide advantages for negative performance feedback in objective tasks (such as performance analysis and sales forecasting). However, compared to human managers, AI lacks social and interaction attributes, resulting in poorer effectiveness for negative performance feedback in subjective work tasks (such as interpersonal communication and conflict management). Accordingly, organizations should identify task types in advance to fully leverage the respective feedback advantages of human managers and AI.

Third, this study provides management insights for organizations to help employees engage in positive psychological construction after receiving human-AI negative performance feedback. Since employees' internal attribution for human-AI negative performance feedback affects their performance improvement motivation—higher internal attribution for negative performance feedback leads to higher motivation to improve performance—organizations need to pay attention to employees' attribution styles after receiving human-AI negative performance feedback and strengthen performance communication to help employees timely identify their shortcomings or improve performance feedback processes, thereby enhancing employee performance.

6.3 Limitations and Future Directions

This study has several limitations. First, future AI may enter the workplace in human-like forms (e.g., virtual employees) to work alongside human employees and provide performance feedback (Yam et al., 2023). Future research could manipulate participants' perception of performance feedback sources (from human vs. AI vs. human-AI hybrid) in more realistic human-AI feedback scenarios (e.g., feedback from virtual colleagues) to deepen human-AI comparisons. Second, previous research has indicated that feedback characteristics are important factors influencing performance feedback effectiveness. This study primarily focused on objective feedback presented through performance ranking. Due to differences between humans and AI in communication and emotional attributes, future research could further explore differential effects of human-AI feedback in evaluative feedback (e.g., open-ended, qualitative, or problem-specific feedback) (Johnson, 2013).

Additionally, future research could examine cultural factors in human-AI performance feedback. For example, under the influence of traditional Chinese culture of moderation and humility, managers often adopt a "sandwich" feedback approach that mixes encouraging and praising positive feedback with negative feedback to avoid conflict. Since AI is often disliked for lacking "human touch" (Dietvorst et al., 2015; Luo et al., 2019), future research could explore the impact of AI adopting a "sandwich" performance feedback strategy on employee performance. Moreover, compared to Western societies, Eastern societies under harmony culture have more indirect interpersonal communication styles (Geng et al., 2020), which may cause individuals in Eastern societies to respond more

negatively to negative performance feedback, thereby affecting the differential effectiveness of human-AI negative performance feedback. Future research could use Western samples to explore the effects of human-AI negative performance feedback under different cultural backgrounds.

Furthermore, this study focused on the causal locus perspective (internal vs. external attribution) in attribution theory. In fact, attribution theory has rich connotations. For example, according to attribution stability, individuals' attributions can be categorized as ability attribution (attributing event outcomes to one's ability) vs. effort attribution (attributing to one's effort or input). According to attribution controllability, individuals may attribute event outcomes to controllable factors (ability, effort, etc.) or uncontrollable factors (luck, task difficulty, etc.) (Weiner, 1985; Russell, 1982). Considering that AI can conduct profiling analysis of individuals based on big data and deeply analyze human characteristics in personality, preferences, and abilities (Fan et al., 2023), future research could adopt more perspectives from attribution theory or combine perspectives (e.g., can AI-delivered negative performance feedback enhance employees' internal attribution for ability and affect performance improvement motivation?) to further enrich attribution theory's explanation of differential effects of human-AI performance feedback.

Finally, this study focused on the proximal outcome of employees' performance improvement motivation after receiving human-AI negative performance feedback. Future research could explore the impact of human-AI negative performance feedback on employees' actual behavioral performance (e.g., performance levels, learning behaviors), thereby expanding research on the consequences of human-AI negative performance feedback.

References

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

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.

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

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

Chory, R. M., & Westerman, C. Y. (2009). Feedback and fairness: The relationship between negative performance feedback and organizational justice. *Western Journal of Communication*, 73(2), 157−181.

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.

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.

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.

Dong, Y., Long, L., Cheng, Z. (2022). Performance management in the era of digital intelligence: Present and future. *Tsinghua Management Review*, 5, 93−100. [董毓格, 龙立荣, 程芷汀. (2022). 数智时代的绩效管理：现实和未来. 清华管理评论, 5,93−100.]

Fan, J., Sun, T., Liu, J., et al. (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.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods*, 39(2), 175−191.

Fields, B., Carbery, M., Schulz, R., Rodakowski, J., Terhorst, L., & Still, C. (2023). Evaluation of Face Validity and Acceptability of the Care Partner Hospital Assessment Tool. *Innovation in Aging*, 7(2), 557−568.

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.

Geng, Z. Z., Zhao, J. J., & Ding. L. (2020). The "Wisdom" of Golden Mean Thinking: Research on the Mechanism between Supervisor Developmental Feedback and Employee Creativity. *Nankai Management Review*, 23(1), 75−86. [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.]

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

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

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125−130.

Hareli, S., & Hess, U. (2008). The role of causal attribution in hurt feelings and related social emotions elicited in reaction to others' feedback about failure. *Cognition and Emotion*, 22(5), 862−880.

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.

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

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.

Jiang, L., Cao, L., Qin, X., Tan, L., Chen, C., & Peng, X. (2022). Fairness Perception of Artificial Intelligence Decision-Making. *Advances in Psychological Science*, 30 (5), 1078−1092. [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.]

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.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366−410.

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.

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.

Kuvaas, B., Buch, R., & Dysvik, A. (2017). Constructive supervisor feedback is not sufficient: Immediacy and frequency is essential. *Human Resource Management*, 56(3), 519−531.

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. *Organizational Behavior and Human Decision Processes*, 116(2), 217−228.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society*, 5(1), 1−16.

Leo, X., & Huh, Y. E. (2020). Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms. *Computers in Human Behavior*, 113(4), 106520.

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

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.

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

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

Ma, L., Xie, P., Wei, Y., & Qiao, X., (2021). Is negative feedback from leaders really harm for employee innovative behavior ? the role of positive attribution and job crafting. *Science and Technology Progress and Policy*, 38(12), 144−150. [马璐, 谢鹏, 韦依依, 乔小涛. (2021). 领导者负面反馈真的不利于员工创新吗——积极归因与工作重塑的作用. 科技进步与对策, 38(12), 144−150.]

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.

Ni, D., Zheng, X. M. (2024). Does negative performance feedback always lead to negative responses? The role of trust in the leader. *Journal of Occupational and Organizational Psychology*, 97(2), 623−646.

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.

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.

Qin, S., 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.

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.

Russell, D. (1982). The causal dimension scale: a measure of how individuals perceive causes. *Journal of Personality and Social Psychology*, 42(6), 1137−1145.

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

Scholz, U., Gutiérrez-Doña, B., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment*, 18(3), 242−251.

Song, X., He, X. (2020). The effect of artificial intelligence pricing on consumers' perceived price fairness. *Journal of Management Science*, 33(5), 3−16. [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.]

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.

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.

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.

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

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.

Xing, L., Sun, J. M., Jepsen, D., & Zhang, Y. J. (2023). Supervisor negative feedback and employee motivation to learn: An attribution perspective. *Human Relations*, 76(2), 1−31.

Xu, L., Yu, F., & Peng, K. (2022). Algorithmic discrimination causes less desire for moral punishment than human discrimination. *Acta Psychologica Sinica*, 54(9), 1076−1092. [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, 54(9), 1076−1092.]

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.

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.

Appendix 1 (Self-Developed Subjective and Objective Tasks)

Note: Experiment 2 used Subjective Task 1 and Objective Task 3; Experiment 3 used all tasks; Experiment 4 used Subjective Tasks 1 and 2, and Objective Tasks 2 and 3.

Subjective Tasks

[Subjective Task 1] Assume today is [date], and you are [position] at [company]. You have received three official emails that require your attention.

[Subjective Task 1] You have received an official email from your colleague [name] in the [department], which mentions difficulties encountered by [name] in leading a team to complete a project. Please handle the following official correspondence.

Official Document Type: Email
To: [Name]
From: [Name]
Date: [Date]
Subject: Regarding Project Team Building Issues

Recently, the company assigned me to lead a team responsible for a new project. Our project team successfully completed preliminary work based on past experience and received confirmation and recognition from the client. However, due to tight timelines and heavy workloads, we requested and obtained two new team members from the company. After project implementation began, frequent disputes occurred between original team members and new members, with each side blaming the other for mistakes. Original team members believed new members were inefficient and delayed project progress, while new members believed original team members were difficult to work with and communicate with effectively. I initially considered this a normal team 磨合 process and did not intervene extensively. However, after two months of project implementation, I realized that continuing this way would definitely cause problems. What suggestions do you have? Please advise.

[Department] [Name]
[Date]

Question: Please provide effective suggestions to your colleague [name] in the [department] based on your work experience (suggestions should be specific, feasible, and logically clear, with at least 100 characters).

[Subjective Task 2] You have received an official email from your colleague [name] in the [department], which mentions difficulties encountered by [name] in leading a team to complete a project. Please handle the following official correspondence.

Official Document Type: Email
To: [Name]
From: [Name]
Date: [Date]
Subject: Regarding Handling Unexpected Situations

Currently, I am primarily responsible for a project that has reached its later stages. However, due to a sudden recurrence of the pandemic, Xiao Li, a col-

league who was responsible for an offline work task, has been quarantined outside. Due to cost control, our project team cannot hire new staff this year, and this offline work was previously only mainly handled by Xiao Li, with others not understanding it and having relatively saturated workloads. Due to limitations, this work cannot be conducted online. What suggestions do you have? Please advise.

[Department] [Name]
[Date]

Question: Please provide effective suggestions to your colleague [name] in the [department] based on your work experience (suggestions should be specific, feasible, and logically clear, with at least 100 characters).

[Subjective Task 3] You are a workplace mentor for [name], who sent you an official email mentioning difficulties encountered in recent work. Please handle the following official correspondence.

Official Document Type: Email
To: [Name]
From: [Name]
Date: [Date]
Subject: Request for Help with Work Communication

Today, the general manager assigned me to complete an important work task together with [name]. When I asked the leader when the work needed to be delivered, the leader replied "as soon as possible." Originally, I planned to work overtime today to complete the task, but [name] believed that since the leader didn't specify a delivery time, we could complete it slowly next week. I tried to persuade him for a long time, but he still felt there was no need to work overtime. How should I persuade him to work overtime with me to complete this important task? Can you give me some suggestions? Please advise.

Question: Please provide effective suggestions to your workplace mentee [name] based on your work experience (suggestions should be specific, feasible, and logically clear, with at least 100 characters).

Objective Tasks

[Objective Task 1] Assume today is [date], and you are [position] in [department] at [company]. You have received three official emails that require your attention.

[Objective Task 1] You have received an official email from your colleague [name] in the [department], which mentions issues regarding work arrangements. Please handle the following official correspondence.

Official Document Type: Email
To: [Name]
From: [Name]

Date: [Date]
Subject: Regarding Work Arrangement Issues

Tomorrow, we have invited four candidates (A, B, C, D) to simultaneously participate in a three-stage interview at our company's project department. The company requires that each candidate must first be interviewed by a recruitment specialist, then by a project supervisor, and finally by a project manager. No queue-jumping is allowed (i.e., the order of the four candidates must be the same at each stage). Due to different professional backgrounds and work experiences, each candidate's interview time varies at each stage, as shown in the table below (unit: minutes):

Candidate	Recruitment Specialist Interview	Project Supervisor Interview	Project Manager Interview
A			
B			
C			
D			

Since our company is far from the hotel where candidates are staying, we need to arrange a shuttle bus to take all four candidates to leave the company together. Interviews begin at 8:00 AM tomorrow. How should you arrange their interview order so they can leave the company together as early as possible? What is the earliest time? Please provide specific arrangements.

Question: Please use logical reasoning and computational ability to reply to [name] in the [department] about interview order arrangements. Sort candidates A, B, C, D and calculate the earliest time they can leave the company. (Please provide specific logical analysis and reasoning, with at least 100 characters).

[**Objective Task 2**] You have received an official email from your colleague [name] in the [department], which mentions issues regarding investment and purchase of project production equipment. Please handle the following official correspondence.

Official Document Type: Email
To: [Name]
From: [Name]
Date: [Date]
Subject: Regarding Production Equipment Investment and Purchase

Our project team recently plans to invest in and purchase a project production equipment and conduct comprehensive evaluations of different equipment types based on six decision indicators: maximum operating speed (C1), maximum output (C2), maximum load (C3), cost (C4), reliability (C5), and sensitivity (C6). Four equipment models are available for selection, with specific indicator

values shown in the table. The weight of each attribute is 20%, 10%, 10%, 10%, 20%, and 30% respectively. Please advise us on how to make the purchase.

Equipment Model	Maximum Operating Speed (units/day)	Reliability Score	Sensitivity Score
A			
B			
C			
D			

Question: Please use logical reasoning and computational ability to reply to [name] in the [department] about equipment investment and purchase. Rank equipment A, B, C, D from best to worst. (Please provide specific logical analysis and reasoning, with at least 100 characters).

[Objective Task 3] You are a workplace mentor for [name], who sent you an official email mentioning difficulties encountered in recent project preliminary research. Please handle the following official correspondence.

Official Document Type: Email
To: [Name]
From: [Name]
Date: [Date]
Subject: Request for Help with Project Research

Recently, I have been conducting preliminary research for a project I am responsible for and need to understand the sales situation of a product involved in the project to decide whether to include it in our project plan. The table below shows the product's sales statistics for the past 10 months. According to company sales standards: (1) Sales $<$ 60,000 units: unsalable; (2) 60,000 $\leq$ sales $\leq$ 100,000 units: average; (3) Sales $>$ 100,000 units: best-selling. According to company requirements, if the product is best-selling next month (month 11), it can be included in our project plan. Please advise whether we should include this product in our project plan.

Product Sales Statistics Table
Sales (10,000 units): 40

Question: Please use logical reasoning and computational ability to reply to your mentee [name] about the project research request. Analyze the product's sales status for month 11 (unsalable, average, best-selling) and predict the sales volume for month 11. (Please provide specific logical analysis and reasoning, with at least 100 characters).

Note: Figure translations are in progress. See original paper for figures.