

ScholarWorks@GSU

AI in the Workplace: How Managerial Perceptions of AI Shape Expectations and the Performance Evaluation Process

Authors	Kaitlin Hudspeth
Citation	Kaitlin Hudspeth. "AI in the Workplace: How Managerial Perceptions of AI Shape Expectations and the Performance Evaluation Process." 2025. Dissertation, Georgia State University. https://doi.org/10.57709/d8g4-0r79
DOI	https://doi.org/10.57709/d8g4-0r79
Download date	2026-03-09 03:09:36
Link to Item	https://hdl.handle.net/20.500.14694/15402

*AI in the Workplace: How Managerial Perceptions of AI Shape Expectations and the Performance
Evaluation Process*

BY

Kaitlin Hudspeth

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY
ROBINSON COLLEGE OF BUSINESS
2025

Copyright by
Kaitlin Hudspeth
2025

ACCEPTANCE

This dissertation was prepared under the direction of the Kaitlin Hudspeth's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Ivo Tafkov (Chair)
Matthew DeAngelis
Michael Majerczyk
Curtis Mullis
Jeremy Lill (External)

ABSTRACT

AI in the Workplace: How Managerial Perceptions of AI Shape Expectations and the Performance Evaluation Process

BY

Kaitlin Hudspeth

April 2025

Committee Chair: *Ivo Tafkov*

Major Academic Unit: *School of Accountancy*

This study investigates how an employee's use of artificial intelligence (AI) impacts managers' subjective performance evaluations, particularly focusing on how managers' perceptions of AI influence their performance expectations and subsequent evaluations of their employees. Using an experiment manipulating AI reliability perceptions (high vs. low) and performance outcomes (success vs. failure), I predict and find that managers expect higher performance from employees using AI perceived as reliable. Further, consistent with my prediction, I find greater discrepancies between managers' expectations and evaluations when an employee uses unreliable AI and exhibits successful performance. Additionally, successful performance is more often attributed to the employee when AI is seen as unreliable, whereas no significant attributional differences were found for unsuccessful performance. These findings highlight the importance of managers' perceptions of AI in shaping managerial expectations and evaluations.

ACKNOWLEDGEMENT

I thank my dissertation committee members Matthew DeAngelis, Michael Majerczyk, Curtis Mullis, and Jeremy Lill for their guidance and support throughout the development of this dissertation. I am especially thankful to my chair, Ivo Tafkov, whose mentorship has shaped not only this project but also my entire journey over the past five years. I would also like to express my appreciation to Michael Majerczyk for being an incredible mentor from the very beginning of my time at Georgia State University. Finally, I extend my heartfelt gratitude to my husband, Christopher, as well as my parents, friends, and family for their love, encouragement, and for keeping me grounded during this long and rewarding journey.

I. INTRODUCTION

Artificial intelligence (AI) is increasingly transforming the workplace, influencing daily business operations, managerial decision-making, and the performance evaluation process. As this technology evolves and becomes more embedded in business practices, managers' performance expectations will likely shift. Further, the process by which managers assess not only the employee performance but also the extent to which AI contributes to that performance may also evolve. This raises critical questions about how managers form expectations and attribute performance when evaluating employees utilizing AI in their work.

In this study, I examine whether and how managers' perceptions of AI influence their expectations of employee performance and the attributions they make when evaluating performance outcomes. Specifically, I first investigate whether managers who perceive AI as highly reliable set higher performance expectations for their employees using AI. Next, I examine how discrepancies between performance expectations and evaluations differ depending on AI reliability perceptions and performance outcomes. Lastly, I investigate whether managers attribute successful and unsuccessful performance outcomes differently based on their AI reliability perceptions.

The way in which managers attribute performance can significantly affect subsequent performance outcomes, including future rewards, career progression, and employee motivation (Weiner 1985, Ryan and Deci 2000). Specifically, misattributions in the evaluation process may have unintended consequences. For example, if managers misattribute unsuccessful performance to employees or fail to recognize employees' contributions because they over-credit AI for

successful performance, employees may feel frustrated and undervalued, which lowers motivation and performance over time (Weiner, 1985, Lindner 1998).

A critical context where misattributions may arise is within subjective performance evaluations, which allow managers to consider performance information beyond quantifiable metrics. Subjectivity enables managers to account for more nuanced aspects of employee performance. However, it also introduces potential biases and inconsistencies (Bol 2011). Factors such as employees' use of AI may contribute significantly to these biases. Research has long established the critical role of attributions in shaping the subjective performance evaluation process (Kelley 1972, Weiner and Kukla, 1970). Yet, integrating AI into business practices presents a new and evolving dimension. Attribution theory suggests that people tend to search for the underlying causes for outcomes. In performance evaluations, this means that managers attempt to understand whether performance outcomes are due to employees' internal factors (i.e., effort or ability) or external factors (i.e., luck or situational factors) (Heider, 1958).

Yet, misattributions can lead to important and potentially overlooked consequences when evaluations are influenced by outcome effects rather than employee effort. Specifically, managers may evaluate performance favorably based on outcomes, even though they attribute the underlying effort to AI rather than to the employee (Tan and Lipe 1997, Mertins et al. 2013). Although a more common concern for misattributions is employee dissatisfaction arising when employees perform well yet are penalized for poor outcomes, here employee dissatisfaction may stem from a more subtle issue. Employees who receive positive evaluations may expect recognition, rewards, or promotions; however, if the manager attributes their successes to AI rather than their own contributions, they may not receive these expected benefits. This would ultimately lead to employee dissatisfaction, increasing turnover and absenteeism, potentially

costing firms their most valuable and highly performing employees (Colquitt 2001, Brayfield and Crockett 1955, Herzberg et al. 1957, Porter and Steers 1973). Furthermore, the persistent misattribution of success to AI may lead to the wrong assumption that AI alone can deliver superior results, resulting in the layoff of good or vital employees if management thinks that AI is enough to replace humans. It is therefore important to understand how managers attribute performance between AI and employees when employees use AI in their work as both AI and AI perceptions evolve over time.

I examine these issues by conducting an experiment in which experienced managers evaluate the performance of an employee who has integrated AI into their work for the period under review. Using a 2x2 between-subjects design, I manipulate performance outcomes (successful vs. unsuccessful) and general AI reliability perceptions through news articles that discuss the reliability of AI (high vs. low) to analyze their effects on managerial performance expectations, evaluation discrepancies, and outcome attributions.

My results show that managers who perceive AI as highly reliable have higher performance expectations for their employees using AI than those with lower AI reliability perceptions. This suggests that as perceptions of AI reliability skew more positively over time as technology evolves, managers will likely set higher standards for their employees' performance. Moreover, I find that when AI is perceived less reliable, managers experience a larger gap between their initial performance expectations and their ultimate performance outcomes after successful outcomes than when AI is perceived as more reliable. This suggests that successful employee performance is more surprising to managers under lower AI reliability perceptions but still results in favorable evaluations. Further, I find that managers tend to attribute successful outcomes more to AI when it is perceived as more reliable. However, their overall performance

evaluations remain unaffected by these attributions (i.e., performance evaluations remain positive for successful performance regardless of the outcome attributions). Overall, these results support the idea that AI perceptions influence managers' performance expectations and attributions between AI and employees under successful outcomes.

Alternatively, for unsuccessful performance outcomes, I observe no attributional differences between the employee and AI, regardless of the manager's AI perceptions. This suggests that in the case of unsuccessful performance, managers struggle to identify the cause of the performance outcome when employees use AI. Additionally, I do not find significant differences between expectations and evaluations for unsuccessful performances when AI reliability perceptions vary. This is because, while managers initially set higher expectations for employees using AI perceived as more reliable, those employees also receive more favorable evaluations in cases of unsuccessful performance compared to employees using less reliable AI. One possible reason for this pattern is that when AI is perceived as less reliable, managers may question the employee's decision to rely on it, resulting in lower evaluations for unsuccessful outcomes compared to when AI is perceived as more reliable.

This study contributes to the understanding of how the perception of AI affects the subjective performance evaluation process as AI usage becomes more widespread and perceptions evolve. First, this study emphasizes the influence of managers' AI perceptions on setting performance benchmarks. Specifically, I find that when managers perceive AI as more reliable, they have higher performance expectations for their employees using AI. Additionally, this study highlights that even positive evaluations can result in negative future consequences if managers misattribute success to AI, leading employees to miss out on rewards and career opportunities, contrary to employee expectations. These misattributions may also lead to layoffs

of critical personnel, which can hurt overall firm performance. The insights from this study offer practical implications for firms implementing AI into their operations, especially as AI perceptions may shift more positively over time. This offers firms the opportunity to implement training and policies that assist managers in appropriately assessing AI-assisted performance.

Additionally, this paper contributes to the understanding of attribution theory as it relates to subjective performance evaluations and AI. Attribution theory is a widely applied theory in behavioral accounting research, particularly in studies examining performance evaluations (e.g., see Kaplan et al., 2017 for a review). Previous research has mainly focused on attributions in the context of relationships and interpersonal interactions among humans (Kelley and Michela 1980, Shaver 2012). However, it is crucial to understand how managers attribute outcomes in the workplace where humans and AI coexist and collaborate. Unlike traditional tools, AI is increasingly perceived as independent and capable of learning and adapting. As a result, managers may start to view AI as a peer rather than another technological tool, complicating the attribution process. This paper expands attribution theory by exploring how AI, a unique external factor, interacts with managers' expectations and performance attributions and highlighting the potential risks of misattributions.

II. BACKGROUND

Subjective Performance Evaluations

Subjective performance evaluations are an important part of performance measurement systems. These evaluations are assessments of an employee's job performance based on criteria that are difficult to quantify or measure objectively. Subjective performance evaluations offer a counterbalance to purely objective measures, which often do not capture the full capacity of an

employee's effort and work contributions. Bol and Smith (2011) discuss that subjective performance evaluations are especially useful in settings where outcomes are difficult to measure in a precise manner and when tasks are more complex. Managerial subjectivity provides flexibility in assessing employee performance, allowing managers to consider situational factors or individual differences that may be overlooked if solely relying on objective or quantitative metrics (Ittner, Larcker, and Meyer 2003). Additionally, subjective performance evaluations allow managers to reward behavior that may be important to long-term success but have little to no immediate quantifiable outcomes (e.g., leadership qualities, innovativeness). By rewarding unquantified or uncaptured behavior, this allows for a more comprehensive performance assessment and motivates employee behavior that is beneficial to an organization's success but not directly tied to objective performance measures (Bol 2008). Therefore, subjective performance evaluations are important for organizations as they measure and motivate more well-rounded employee performance.

However, since subjective performance evaluations rely on human judgement, they are also subject to human error. For example, managers may face difficulty in assessing multiple dimensions of performance, leading to the use of simplification strategies such as focusing on the most easily obtainable information (i.e., availability bias) (Tversky and Kahneman, 1973). Situational factors (e.g., organizational changes, economic conditions, team dynamics, etc.) may also influence evaluations. For example, environmental uncertainty may affect managerial subjectivity through evaluation adjustments when uncontrollable factors negatively impact performance (Bol and Smith 2011). Manager biases and inconsistencies in subjective performance evaluations may undermine their credibility and effectiveness, leading to decreased motivation and trust in management (Hartmann and Slapnicar 2009).

Further, these evaluations are important because they determine the rewards and career opportunities employees receive (Bol 2008). Thus, it is important to have a thorough understanding of what factors may influence and potentially skew subjective performance evaluations. As AI permeates throughout the business landscape, it is vital to understand the effect it may have on subjective performance evaluations. In this study, I examine how employees' integration of artificial intelligence in their work affects managers' subjective performance evaluations.

Artificial Intelligence

History and Perceptions

AI, defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan and Haenlein 2019, p. 17), spans a spectrum of capabilities and has significantly and rapidly improved in recent years. The simplest form of AI, reactive machines, is not capable of learning and only responds to a limited set of inputs. A well-known example of this type of AI is IBM's Deep Blue that was used to beat Chess Grandmaster Garry Kasparov in 1997 (IBM 2023).

Advancing from reactive machines, limited-memory machines have some learning capabilities and use historical data to inform later decisions. Most current forms of AI are limited-memory machines, including self-driving cars and AI chatbots (IBM 2023). A more complex type of AI is theory of mind AI, which can understand human emotions and respond to human mental states. Lastly, the most complex form of AI is self-aware AI, which possesses its own thoughts and emotions and represents the most advanced form of AI development (IBM 2023). Both theory of mind and self-aware AI are currently only theoretical concepts.

The recent surge in limited-memory, generative AI applications like ChatGPT and Dall-E has sparked tremendous media and consumer attention, although responses and opinions are mixed. Despite the mixed opinions, AI can offer significant benefits for many industries. For example, the integration of AI into healthcare has been transformative, providing precise diagnostics and treatments for patients and using machine learning techniques with images for early detection of diseases (Esteva et al., 2021). AI has also played a major role in the finance industry through algorithmic trading, fraud detection, and personalized investment advice (Chlouverakis 2024, Li et al. 2019, Zheng et al. 2019). Despite these advancements, there are considerable concerns about the reliability of AI, as users of generative AI experience reasoning errors or contradictions (Mills 2023).

AI users should exercise caution when relying on AI as there are significant risks associated with its use. One major concern is its tendency to “hallucinate,” making up incorrect information. According to Metz (2023), even the largest AI models generate false information, with hallucination rates varying considerably across different AI systems. This unreliability poses significant risk when employees rely on it without fact-checking the information produced. There is also considerable risk of accountability in AI systems, as these systems are generally “black boxes,” making it difficult to verify the accuracy of the information (Von Eschenbach 2021). However, given the nature of AI, this technology shows promise for future improvements.

Deep learning, a subset of machine learning, has facilitated recent AI breakthroughs, such as natural language processing, speech recognition, and image recognition (Soori, Arezoo, and Dastres 2023). Deep learning allows for learning, adapting, and improving over time, which will also likely improve users’ perceptions and trust of the technology over time. For example, GPT-3.5, the model developed by OpenAI in 2022, scored in the bottom 10 percent of prospective

lawyers for the Uniform Bar Exam, yet GPT-4, a more comprehensive model released later, scored in the top 10 percent (Katz et al. 2024). These advancements, alongside developments such as Explainable AI (XAI) aim to enhance AI's understandability and accountability, addressing current concerns of trust and transparency (Confalonieri et al., 2021; Khosravi et al., 2022). By improving reliable output, trust, and transparency, the future perceptions of AI should improve and become more cohesive.

However, perceptions generally take time to evolve as technology evolves. Given the current reliability issues experienced by AI users, the perceptions of this technology are mixed. The mixed perceptions of AI can significantly impact how managers evaluate and perceive employees who utilize AI in their work. Managers with a positive perception of AI may expect higher quality work, more productivity, and more efficiency. Meanwhile, managers with negative AI perceptions may be more skeptical of its outputs and have lower expectations of their employees who utilize AI in their work. Understanding this dichotomy is important, especially as AI becomes increasingly integrated into business operations. The perceptions that managers have of AI can influence their expectations and evaluations of employees, shaping how employees are assessed and rewarded. In this study I investigate the extent of this influence and how it may lead to inconsistencies in subjective performance evaluations.

AI-Human Collaborations

The relationship between humans and machines has changed over time, impacting workplace dynamics and worker productivity. The history of this relation can be used as a starting point in understanding the relationship between humans and AI in the present day and how outcomes are attributed when humans and AI collaborate. Machines were originally created to make work easier for humans during the Industrial Revolution. Early automation systems,

such as Jacquard looms and steam-powered machinery, were seen as extensions of human effort rather than independent agents (Marshall 2022, Rashid and Kausik 2024). As technology advanced, especially with the development of computers and AI, the relationship between humans and machines graduated from the simplest form of partnership to a more complex one where AI can either serve as a supporting role for humans or can even work independently in several industries (Jia et al. 2022, Schoenerr et al. 2024).

The introduction of AI has led to both advantages and disadvantages for humans utilizing this technology. AI has been applied in decision-making tasks, healthcare diagnostics, finance, manufacturing and other fields (Rashid and Kausik 2024). Research in this area suggests that AI can improve productivity, reduce human error, and support decision-making by analyzing large amounts of data faster than humans (Gao and Feng 2023, Brynjolfsson et al., 2025, Jarrahi, 2018). Nevertheless, the integration of AI also creates issues related to trust, responsibility and the evaluation of outcomes when humans and AI collaborate (Glikson and Woolley 2020, Ulfert et al. 2024, Yue et al. 2023).

One of the most frequent themes of research in this area is the question of who is to credit or blame for an action when AI is involved in work outcomes. A growing body of literature suggests that responsibility is typically shared rather than assigned to specific party, which results in shared attributions of responsibility and blame between humans and AI (Schoenerr et al., 2024; Shank et al., 2019). However, the degree to which technology is held responsible for outcomes depends on a number of factors. For example, if robots are perceived as highly autonomous, people attribute more credit and blame to the robot and less toward themselves and others (Kim and Hinds 2006, Furlough et al., 2021). Looking at AI, a recent study finds that

highly autonomous AI leads to ascribing failure more to human operators than to AI (Schoenerr et al., 2024).

Transparency and explainability may also be important in shaping perceptions of AI-driven outcomes (Schoenerr et al. 2024, Bussone et al. 2015, Shank and Desanti 2018, Shin 2021). Specifically, when AI can effectively explain its reasoning behind recommendations, users are more likely to trust and positively evaluate its outcomes since AI is considered as more rational and autonomous. Alternatively, when AI works as a “black-box” and the decision-making process is less clear, users may hesitate to accept AI’s outcome, leading to greater scrutiny of the output and increased likelihood to attribute failure to AI rather than to an external factor like the operator or developer.

This literature suggests that when employees incorporate AI into their work, managers’ performance attributions may become skewed, leading to biases in how successes and failures are assigned. It is therefore important to understand not only how AI perceptions shape performance attributions but also how these perceptions shape managers’ expectations. Additionally, it is vital to understand how these expectations and attributions may shift when AI is perceived as more or less reliable, as this perception of reliability may shift over time as AI continues to evolve.

III. THEORY AND HYPOTHESES

Attribution Theory Overview

Attribution theory refers to people’s causal explanations for events. According to the initial work of Heider (1958), people are inclined to explain the world around them by identifying underlying causes for actions, especially when outcomes are unexpected or

significant. In general, individuals tend to assign the cause of a successful or failed outcome to either internal (dispositional) or external (situational) attributions. Dispositional attributions imply that the explanation for an event or behavior is the individual's intrinsic traits, and that the individual is responsible for the outcome. Situational attributions imply that outside forces caused the event or behavior, and that the individual is not responsible for the outcome.

Following the work of Heider, researchers have explored the processes by which individuals assign causal attributions (Weiner 1986; Kelley 1967; Green and Mitchell 1979; Martinko and Thomson 1998) and attributional biases (Ross 1977; Fiske and Taylor 1991).

One key insight from attribution theory is the fundamental attribution error, a phenomenon described by Ross (1977), which occurs when people overestimate the role of internal attributes and underestimate external attributes when explaining the behavior of others. One possible reason for the fundamental attribution error is that people naturally generate personal attributions, making the process of considering situational attributions for others more cognitively challenging (Newman and Uleman 1989, Uleman et al. 2005). Simply put, it is more difficult to examine all situational aspects influencing a person's behavior or performance rather than to focus solely on individual traits. However, when people are aware of external factors, they include them in their attributions because it is easy to do so and leads to a more comprehensive attribution. This study investigates a scenario in which managers are aware of situational circumstances that may influence their employees' performance. Specifically, in organizational settings where managers are aware of their employees using AI in their work, I consider how managers may consider this situational factor in their expectations and evaluations.

In organizational settings, attribution theory has been applied to better understand how managers evaluate their employees. Green and Mitchell (1979) proposed a model focused on the

attributional processes of managers in the context of interactions between the manager and employee. Their model suggests that managers act as information processors, interpreting employee performance through the lens of internal or external attributions. These attributions then shape how managers respond to employee behaviors, influencing decisions around rewarding, punishing, supervising, and setting future expectations.

In particular, effort attributions tend to be the primary driver of performance evaluations (Green and Mitchell 1979, Weiner and Kukla 1970). Specifically, success attributed to employee effort leads to the highest positive evaluation and most substantial rewards. Failure attributed to a lack of effort results in the most severe punishments. Notably, those perceived as having high ability but exerting low effort receive the harshest penalties (Weiner and Kukla, 1970). This phenomenon highlights the interconnectedness between managers' performance expectations, outcome attributions, and the reward or punishment of employees (Green and Mitchell, 1979).

Effect of AI Perceptions on Performance Expectations

I contend that managers' perceptions of the AI tools used by their employees play an important role in shaping their performance expectations. When managers form their performance expectations for their employees, they do so based on an established baseline of what the employee can achieve independently, based on prior observations of performance (Kelley 1967). However, the introduction of AI introduces a new variable into this equation, influencing how managers form these expectations. If a manager perceives AI as a reliable tool, they are likely to anticipate that its use will significantly enhance the employee's performance beyond their baseline capabilities. Thus, the manager's expectations will be higher, as they believe the AI will enhance the employee's output and lead to better performance.

Alternatively, when managers view AI as less reliable, they are more cautious in their expectations. Skepticism surrounding AI's contribution may lead managers to doubt whether the tool can meaningfully improve performance. It is also possible that managers may even anticipate that AI could undermine performance if employees rely on incorrect information provided by AI. As a result, managers who perceive AI as less reliable are likely to set relatively lower expectations for employees who use it, reflecting their concern that AI may not provide consistent, beneficial support. Thus, managers' AI reliability perceptions should significantly influence the expectations they hold for their employees' performance. A positive perception of AI reliability should lead to greater performance expectations, while skepticism about AI reliability should temper these expectations.

H1: Managers who perceive AI as highly reliable will set higher performance expectations for employees utilizing AI compared to managers who perceive AI as less reliable.

Effect of AI Perceptions on Attributions under Successful and Unsuccessful Outcomes

Performance expectations play a vital role in managers' attributions and evaluations, as an employee's performance outcomes will either confirm or contradict managerial expectations (Feather and Simon 1971). The alignment or misalignment between an employee's performance and the manager's expectations greatly influences whether the outcome is attributed to internal or external factors (Green and Mitchell 1979). Specifically, when performance meets expectations, the manager is more likely to attribute the performance to the employee's internal factors. However, when performance outcomes are unexpected, managers engage in a causal search to understand the misalignment between the outcome and their expectations (Weiner 1986, Read 1987).

The traditional attribution model, which focuses on the manager and employee, becomes more complex when AI is introduced into the equation. When AI is utilized as a tool by employees, managers must not only form expectations about the employee's performance, but also about AI's contribution to the performance outcome. This additional consideration creates a more nuanced attributional process, where the source of the manager's expectations (i.e., the employee or the AI) significantly impacts the attributions that managers make regarding performance outcomes.

When managers perceive AI as highly reliable, they likely form high expectations of how it will further improve employee performance. Therefore, when the employee delivers a successful outcome, the manager's expectations are met, leading to a congruency between high performance expectations and a high performance evaluation. However, because the manager's expectations were shaped by the belief that AI would contribute significantly to the performance, the manager will likely attribute successful performance more to the AI than to the employee.

Alternatively, when managers perceive AI as less reliable, they may expect little to no benefit from the employee's use of the tool. In some cases, managers may even anticipate that AI could undermine performance if they are concerned about the employee's reliance on unreliable technology. In this case, when the employee is successful, this outcome contradicts the manager's relatively low expectations, leading to a greater discrepancy between the manager's expectations and the strong performance evaluation provided to the employee for successful performance. Further, given the unexpected result, the manager will reconsider whether the AI or employee was responsible for the success through their attributional search (Feather 1969, Feather and Simon 1971, Green and Mitchell, 1979). Given the manager's low AI reliability

perceptions, they are less likely to attribute the success to the AI and will instead attribute it more to the employee.

For unsuccessful performance outcomes, the attributional process again depends on the manager's perceptions and expectations of AI and the employee. If the manager has low expectations of AI due to low reliability perceptions and those expectations align with a poor performance outcome, then the manager's evaluation will align with their expectations. In this case, the manager is more likely to attribute the poor performance to the AI because the manager anticipated AI's contribution to be minimal or even potentially detrimental. Thus, the poor performance result reinforces the manager's perception of AI as a less reliable tool, leading them to place the blame on AI rather than the employee.

However, when the manager has high expectations of AI due to high reliability perceptions, unsuccessful performance outcomes contradict the manager's expectations, leading to a larger discrepancy between performance expectations and a low performance evaluation. In this scenario, the manager expects AI to improve performance, so when failure occurs, it triggers an attributional search to determine whether the failure was due to the employee or the AI. Given the manager's positive perception of the reliability of AI, they are more inclined to attribute the failure to the employee for not capitalizing on AI's capabilities. Considering both successful and unsuccessful performance outcomes, my remaining predictions are as follows:

H2: The discrepancy between managers' expectations and evaluations is larger for successful employees using AI perceived as less reliable, and for unsuccessful employees using AI perceived as more reliable.

H3: Managers attribute performance outcomes more strongly to the employee than AI when performance is successful and AI reliability perceptions are low, and when performance is unsuccessful and AI reliability perceptions are high.

IV. METHOD

Participants

I recruited participants with management experience from Prolific, an online participant recruitment platform that caters to research studies (Palan and Schitter 2018).¹ I used Prolific's prescreen tool to find participants who (1) currently live in the United States, (2) speak English fluently, (3) have earned at least an undergraduate degree, (4) have at least one year of managerial experience working with a subordinate, and (5) have completed at least ten previous submissions with an approval rating of at least 95%. A total of 121 participants completed the experiment online. Of these participants, 49 (40.5 percent) are female, 71 (58.7% percent) are male, and one is nonbinary (0.8 percent). Participants' mean age is 40.0 years. Participants reported an average of 17.7 years of work experience and 8.7 years of evaluation experience. Participants earned \$8 for their participation needed an average of 14 minutes and 28 seconds to complete the experiment. Thus, participants earned an average wage of approximately \$33 per hour.

Experiment Task

Participants assumed the role of Director of Product Development at a hypothetical firm and were asked to evaluate the performance of a Product Developer, who is responsible for generating product development plans for their clients. Participants were told that part of their job as Director is staying current on emerging topics that may influence the industry, such as artificial intelligence. Then, participants receive two articles discussing AI, one that provides a discussion from an AI expert and one that discusses AI's role in businesses.

¹ Before recruiting participants, I obtained Institutional Review Board approval from the university where I conducted this study.

Participants then learned about the firm's evaluation scale and reviewed the performance score that the Product Developer earned last year, which was 53 and fell within the range of "Meets Expectations." Participants were then informed that the Product Developer has incorporated AI into their work this year to assist in the product development plan for their main client. After reviewing this information, participants made a prediction about what the evaluation may be this year with the new integration of AI².

After making their prediction, they received the product development plan generated by the Product Developer. After reviewing the plan, participants provided an overall performance score for the Product Developer and reported the extent to which they attributed the outcome of the plan to the Product Developer or the AI technology. I use the difference between the participants' prediction and the actual evaluation as my primary dependent variable. On the next screen, participants filled out their evaluation notes explaining the performance score given to the Product Developer. After completing the main task, participants answered a series of questions about the task, including demographic questions.

Manipulations

Using a 2 x 2 between-subjects experimental design, I manipulate AI reliability perceptions (high vs. low) and employee performance (successful vs. unsuccessful). To induce manager perceptions of AI reliability, participants received news articles that discuss recent AI news prior to learning about their employee's performance outcomes (see Appendix for articles). For the high AI reliability conditions, these articles discuss AI in a positive tone, highlighting the successes and reliability of AI. Alternatively, for low AI reliability conditions, these articles discuss AI in a more cautious tone, highlighting the limitations and low reliability of AI. To

² I capture performance predictions to determine manager expectations after learning about their employee's AI use but before learning of the performance outcomes.

strengthen and internalize the manipulation, participants provided a short explanation of why AI was a reliable or unreliable tool.

I manipulate successful and unsuccessful performance using the product development plan developed by the employee (see Appendix for the product development plans). The participants learned that expected sales are \$800,000. In the successful condition, participants learned that the product development plan was successful, the client was happy with the plan, and the client's sales substantially exceeded the target. In the unsuccessful condition, participants learned that the product development plan was unsuccessful, the client was unhappy with the plan, and the client's sales substantially missed the target.

V. RESULTS

Manipulation Checks

I ask participants several post-experiment questions to ensure participants understood the information presented in the experimental materials and confirm the effectiveness of my manipulations. I ask participants to indicate the extent to which they agree or disagree that their employee was successful in the product development plan for their main client this year. Participants responded on a seven-point scale with 1 labeled "Disagree," 7 labeled "Agree," and the midpoint of 4 labeled "Neither Agree nor Disagree." Participants in the *Successful* employee performance condition agreed more with the statement than those in the *Unsuccessful* employee performance condition (6.47 vs. 2.39, $F = 442.15$, $p < 0.01$, untabulated)³, supporting the manipulation's effectiveness.

³ All p-values are two-tailed unless stated otherwise.

Additionally, I asked participants to indicate the extent to which they agree or disagree that the AI used by their employee for the product development plan was reliable on the same seven-point scale. Participants in the *AI Reliability High* condition agreed more with the statement than those in the *AI Reliability Low* condition (4.63 vs. 3.85, $F = 10.17$, $p = 0.02$, untabulated). Further, as an added manipulation check, I investigated the open responses to the question “Based on what you’ve learned from the articles, please provide a short explanation of why AI is a[n] [reliable/unreliable] tool,” which was asked following the AI Reliability manipulation. I manually coded responses as 1 if the response correctly described AI as reliable or unreliable consistent with the condition they were assigned. After analyzing the responses, only one participant responded in a manner inconsistent with their assigned condition⁴. Taken together, these manipulation checks support the effectiveness of the manipulation.

Hypotheses Tests

Hypothesis 1 (H1) states that managers will expect higher performance from employees using AI perceived as highly reliable compared to those using AI perceived as less reliable. Using an ANCOVA controlling for AI familiarity⁵, I compared the performance prediction scores, which were collected after the reliability manipulation but before learning the performance outcome, across the two reliability conditions. The results as shown in Table 1

⁴ The participant’s response stated “While I completely disagree with both articles based on my own experience, the articles claim that AI is useful for data analysis such as trend spotting and forecasting. They also claim that AI reduces human error by a significant margin. They also, somehow, claim that AI chatbots improve customer satisfaction due to increased availability.” Given the importance of perceptions to theory, I do not include this observation in my analysis.

⁵ *AI Familiarity* refers to a 7-point Likert scale answer to the question “In general, how familiar are you with Artificial Intelligence?” with 1 being “Not familiar at all: I have no knowledge or understanding of AI” and 7 being “Extremely familiar: I am an expert in AI with comprehensive knowledge and extensive experience in this field.” I include this as a covariate in my analyses because it is significantly associated with *AI Reliability* ($r = .20$, $p < .05$, untabulated) and differs significantly between groups ($t(118) = -2.22$, $p = .03$, untabulated). Given that AI Reliability is a manipulated independent variable, controlling for AI Familiarity ensures that observed effects on the dependent variables are due to the experimental manipulations rather than individual differences in *AI Familiarity*.

indicate that participants in the high AI reliability conditions expected significantly higher employee performance than those in the low AI reliability conditions ($M_{\text{reliableAI}} = 65.43$, $M_{\text{unreliableAI}} = 56.47$, $F = 21.74$, $p < 0.01$). This finding suggests that when AI is perceived as more reliable, managers anticipate superior performance from their employees than when AI is perceived as less reliable. Therefore, H1 is supported.

Hypothesis 2 (H2) examines how discrepancies between managers' expectations and evaluations vary based on AI reliability perceptions and performance outcomes. Specifically, H2 proposes that when an employee is successful (unsuccessful), the discrepancy between managers' expectations and evaluations is larger under low (high) AI reliability perceptions. I ran an ANCOVA including Performance Outcomes (success vs. failure), AI Reliability (high vs. low) and a covariate for AI Familiarity on the difference between managers' expectation scores and their final evaluation scores.

The results shown in Table 2 and depicted in Figure 1 reveal a marginally significant interaction between Performance Outcomes and AI Reliability ($F = 2.82$, $p = 0.10$). Simple effects show that the mean difference in scores for successful employee performance using AI with low reliability is 17 points, whereas the difference is 10 points when AI has high reliability perceptions ($F = 5.05$, $p = 0.03$). This effect is consistent with my predictions, indicating a greater discrepancy between expectations and performance evaluations for successful performance when AI has low reliability perceptions compared to high. In contrast, when the employee was unsuccessful, the difference in expected and actual performance scores did not differ significantly between low (24-point difference) and high (24-point difference) AI reliability perceptions ($F=0.01$, $p=0.91$). Overall, these results partially support H2, with the

predicted difference emerging only under successful performance. I explore this further in my supplemental analysis.

Hypothesis 3 (H3) predicts that when an employee is successful (unsuccessful), managers with low (high) AI reliability perceptions attribute performance more to the employee than to AI. To test this interaction, I ran an ANCOVA on the performance attribution measure (0 = entirely AI, 100 = entirely employee) with Performance Outcome and AI Reliability as factors and AI familiarity as a covariate. The results in Table 3 and depicted in Figure 2 reveal a significant interaction between Performance Outcome and AI Reliability ($F = 4.51, p = 0.04$).

Simple effects reveal that managers are more likely to attribute performance to their employees when they are successful and use AI perceived as having low reliability compared to high reliability ($M = 64.77$ vs. $M = 52.46$; $F = 3.76, p = 0.06$). This finding supports H3 for successful outcomes, indicating that managers credit the employee rather than the AI if they perceive the AI as less reliable. However, contrary to H3, attributions did not differ based on AI reliability perceptions when the employee was unsuccessful ($M = 48.73$ vs. $M = 55.37$; $F = 1.08, p = 0.30$). This suggests that the likelihood of managers blaming the employee or AI for unsuccessful outcomes appears to be unaffected by their AI reliability perceptions. Overall, H3 is marginally supported for successful outcomes only, with no significant corresponding effect under unsuccessful outcomes. I explore this further in my supplemental analysis.

Supplemental Analysis

Overall Evaluation Scores

In addition to my main analysis, I also examine overall evaluation scores, which should be primarily driven by the overall performance outcome (i.e., successful performance or

unsuccessful performance). The analysis, shown in Table 4 and depicted in Figure 3, reveals a significant main effect of performance outcomes, with successful performance outcomes being evaluated significantly higher than unsuccessful performance outcomes ($F = 302.22, p < 0.01$).

The results also show that when performance outcomes are unsuccessful, employees using AI perceived as less reliable receive significantly lower evaluations compared to those using AI perceived as highly reliable ($M_{\text{Unreliable AI}} = 32.11, M_{\text{Reliable AI}} = 41.11, F = 5.00, p = 0.02$). This finding suggests that managers set higher initial expectations for employees using AI perceived as more reliable, resulting in evaluations anchored at a higher starting point. Consequently, although these evaluation scores are adjusted downward for all cases of unsuccessful performance, the evaluations for employees using AI perceived as more reliable are not adjusted as low as the evaluations of employees using AI perceived as less reliable. Overall, this leads to a more “forgiving” performance evaluation when employees use AI perceived as more reliable. This helps explain why I do not find support for H2 and H3 under conditions of unsuccessful performance in my main analysis. Specifically, the degree of adjustment between expectations and evaluations in the case of unsuccessful performance remains consistent regardless of AI perceptions due to the higher expectation and higher performance evaluation for high AI reliability perceptions and lower expectations and lower performance evaluations for low AI reliability perceptions.

Further, I find no significant differences in performance evaluations between AI perceptions when employees are successful. One possible reason for this lack of difference is due to the overall performance outcomes overshadowing any nuances in the means to achieve the outcomes (i.e., outcome effects). However, although successful performance results in similar evaluation scores regardless of AI reliability perceptions, my primary analysis indicates

considerable differences in how managers attribute these outcomes based on their perceptions of AI reliability.

Effort Attributions

Given the strong link between effort attributions and performance evaluations established in prior research (e.g., Green and Mitchell 1979; Weiner and Kukla 1970), I conducted a supplemental analysis examining whether AI reliability perceptions influence the extent to which managers attributed effort to the employee under both successful and unsuccessful outcomes. Specifically, I asked participants indicate to what extent they agree with the statement “in your opinion, it is likely that Jordan’s product development plan was successful [unsuccessful] for the client because Jordan worked [did not work] hard on the product development plan to help the client achieve their goals” using a 7-point Likert scale where 1 is strongly disagree and 7 is strongly agree. I compare participant responses across AI reliability perceptions, controlling for AI Familiarity. Results indicate that managers are significantly more likely to attribute more effort to successful employees who use less reliable AI compared to more reliable AI ($M_{\text{Unreliable AI}} = 6.13$, $M_{\text{Reliable AI}} = 5.59$, $F = 4.81$, $p = 0.03$). This finding is especially important, as a manager’s behavior of rewarding or punishing an employee is strongly influenced by effort attributions concerning employee performance (Green and Mitchell, 1979). For unsuccessful performance, I find no significant differences in how managers attribute effort ($M_{\text{Unreliable AI}} = 4.14$, $M_{\text{Reliable AI}} = 4.37$, $F = 2.97$, $p = 0.59$) based on their perceptions of AI reliability. Given that participant responses across both AI reliability conditions center around the midpoint of the scale (4), managers appear uncertain whether unsuccessful outcomes are primarily due to employee effort or other factors such as AI limitations. This uncertainty in

assigning blame may help explain why I do not find support for H3 under unsuccessful performance outcomes.

VI. CONCLUSION

In this study, I investigate how the integration of AI into employees' work affects managers' subjective performance evaluations. Specifically, I focus on whether managers' perceptions of AI reliability influence their performance expectations and subsequent performance evaluations of their employees using AI. Findings suggest that managers who perceive AI positively have higher performance expectations of their employees using AI compared to those with a negative perception. I also find that due to these expectations, the discrepancy between managers' expectations and actual evaluations is greater when AI is perceived as less reliable, particularly for employees with successful performance.

The implications of these findings are significant for organizations increasingly adopting AI technologies as performance outcomes are attributed differently depending on managers' perceptions of the technology. Specifically, when managers perceive AI as more reliable, they attribute successful outcomes more to AI than the employee. This, in turn, could diminish future recognition for employees' efforts if they perceive those employees exerted lower effort. Alternatively, when managers perceive AI as less reliable, employees receive more credit for successful performance outcomes. Overall, these implications are important to understand as AI improves and becomes more widely accepted and trusted. Additionally, some industries and specialties have different understandings of AI, affecting their perceptions of the technology. For example, the financial services industry is one of the most trusting and highest adopters of AI (IBM 2024). Whereas creative sectors, such as music, film, and art, show a more cautious approach towards AI (Whiting 2024).

In this study, I do not account for other factors outside of AI reliability perceptions that could influence managers' evaluations, such as the specific type of AI used or the context in which it is deployed. Future research should explore additional factors and consider examining how managers' perceptions of AI evolve over time with increased exposure and familiarity with the technology.

REFERENCES

- Bol, J. C. (2008, December). Subjectivity in compensation contracting. In AAA Management Accounting Section (MAS) 2006 Meeting Paper, *Journal of Accounting Literature* (Vol. 27, pp. 1-32).
- Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5), 1549-1575.
- Bol, J. C., & Smith, S. D. (2011). Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review*, 86(4), 1213-1230.
- Brayfield, A. H., & Crockett, W. H. (1955). Employee attitudes and employee performance. *Psychological bulletin*, 52(5), 396.
- Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, qjae044.
- Bussone, A., Stumpf, S., & O'Sullivan, D. (2015, October). The role of explanations on trust and reliance in clinical decision support systems. In 2015 international conference on healthcare informatics (pp. 160-169). IEEE.
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: a construct validation of a measure. *Journal of applied psychology*, 86(3), 386.
- Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable artificial intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(1), e1391.
- Chlouverakis, K. (2024, April 26). How artificial intelligence is reshaping the financial services industry. https://www.ey.com/en_gr/financial-services/how-artificial-intelligence-is-reshaping-the-financial-services-industry
- Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *Npj Digital Medicine*, 4(1). <https://doi.org/10.1038/s41746-020-00376-2>
- Feather, N. T. (1969). Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology*, 13(2), 129.
- Feather, N. T., & Simon, J. G. (1971). Attribution of responsibility and valence of outcome in relation to initial confidence and success and failure of self and other. *Journal of Personality and Social Psychology*, 18(2), 173.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). McGraw-Hill Book Company.
- Furlough, C., Stokes, T., & Gillan, D. J. (2021). Attributing blame to robots: I. The influence of robot autonomy. *Human factors*, 63(4), 592-602.
- Gao, X., & Feng, H. (2023). AI-Driven Productivity Gains: Artificial Intelligence and Firm Productivity. *Sustainability*, 15(11)
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of management annals*, 14(2), 627-660.
- Green, S. G., & Mitchell, T. R. (1979). Attributional processes of leaders in leader—member interactions. *Organizational behavior and human performance*, 23(3), 429-458.

- Hartmann, F., & Slapničar, S. (2009). How formal performance evaluation affects trust between superior and subordinate managers. *Accounting, Organizations and Society*, 34(6-7), 722-737.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.
- Herzberg, F., Mausner, B., Peterson, R. O., & Capwell, D. F. (1957). Job attitudes: Review of research and opinion. Pittsburgh: Psychological Services of Pittsburgh.
- Hewett, R., Shantz, A., Mundy, J., & Alfes, K. (2018). Attribution theories in human resource management research: A review and research agenda. *The International Journal of Human Resource Management*, 29(1), 87-126.
- IBM. (2024, January 10). Data Suggests Growth in Enterprise Adoption of AI Is Due to Widespread Deployment by Early Adopters. IBM Newsroom.
<https://newsroom.ibm.com/2024-01-10-Data-Suggests-Growth-in-Enterprise-Adoption-of-AI-is-Due-to-Widespread-Deployment-by-Early-Adopters>
- Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The accounting review*, 78(3), 725-758.
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586.
- Jia, H., Wu, M., & Sundar, S. S. (2022). Do we blame it on the machine? Task outcome and agency attribution in human-technology collaboration. Hawaii International Conference on System Sciences.
- Kaplan, S. E., Samuels, J. A., & Sawers, K. M. (2017). Social psychology theories as applied to Behavioural Accounting Research. In *The Routledge companion to behavioural accounting research* (pp. 57-71). Routledge.
- Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2024). Gpt-4 passes the bar exam. *Philosophical Transactions of the Royal Society A*, 382(2270), 20230254.
- Kelley, H. H. Attribution in social interaction. In E. Jones, D. Kanouse, H. Kelley, R. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press, 1972.
- Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation*, 1967 (Vol. 15). Lincoln, NB: Univ. Nebraska Press, 1967.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28, 107-128.
- Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual review of psychology*, 31(1), 457-501.
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and education: artificial intelligence*, 3, 100074.
- Kim, T., & Hinds, P. (2006, September). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *ROMAN 2006-The 15th IEEE international symposium on robot and human interactive communication* (pp. 80-85). IEEE.

- Li, Y., Zheng, W., & Zheng, Z. (2019). Deep Robust Reinforcement Learning for Practical Algorithmic Trading. *IEEE Access*, 7, 108014–108022.
<https://doi.org/10.1109/access.2019.2932789>
- Lindner, J. R. (1998). Understanding employee motivation. *Journal of extension*, 36(3), 1-8.
- Marshall, S. J. (2022). *The Story of the Robot: A Short History of Automation and Robotics*. Stephen J Marshall.
- Martinko, M. J., & Thomson, N. F. (1998). A synthesis and extension of the Weiner and Kelley attribution models. *Basic and Applied Social Psychology*, 20(4), 271-284.
- Mertins, L., Salvador, D., & Long, J. H. (2013). The outcome effect—a review and implications for future research. *Journal of Accounting Literature*, 31(1), 2-30.
- Mills, S. (2023, April 17). Generative AI Is Under Growing Scrutiny. Here’s What That Means. *BCG.com*. Retrieved July 21, 2024, from
<https://www.bcg.com/publications/2023/growing-scrutiny-over-generative-ai-ethics-risks>
- Newman, L. S., & Uleman, J. S. (1989). Spontaneous trait inference. *New York: Guilford*, 155-88.
- Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin*, 80(2), 151-176.
- Rashid, A. B., & Kausik, A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, 100277.
- Read, S. J. (1987). Constructing causal scenarios: a knowledge structure approach to causal reasoning. *Journal of personality and social psychology*, 52(2), 288.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173-220). Academic Press.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54-67.
- Schoenherr, J. R., & Thomson, R. (2024). When AI fails, who do we blame? Attributing responsibility in human-AI interactions. *IEEE Transactions on Technology and Society*.
- Shank, D. B., & DeSanti, A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations. *Computers in human behavior*, 86, 401-411.
- Shank, D. B., DeSanti, A., & Maninger, T. (2019). When are artificial intelligence versus human agents faulted for wrongdoing? Moral attributions after individual and joint decisions. *Information, Communication & Society*, 22(5), 648-663.
- Shaver, K. G. (2012). *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer Science & Business Media.
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International journal of human-computer studies*, 146, 102551.
- Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, 3, 54-70.
- Tan, H. T., & Lipe, M. G. (1997). Outcome effects: The impact of decision process and outcome controllability. *Journal of Behavioral Decision Making*, 10, 315–325.

- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.
- Uleman, J. S., Blader, S. L., & Todorov, A. (2005). Implicit impressions. *The new unconscious*, 362-392.
- Ulfert, A. S., Georganta, E., Centeio Jorge, C., Mehrotra, S., & Tielman, M. (2024). Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework. *European Journal of Work and Organizational Psychology*, 33(2), 158-171.
- Understanding the different types of artificial intelligence. (2023, October 12). IBM. Retrieved July 20, 2024, from <https://www.ibm.com/think/topics/artificial-intelligence-types>
- Von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), 1607-1622.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4), 548.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.
- Weiner, B., & Kukla, A. An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology*, 1970, 15, 1-20.
- Whiting, K. (2024, February 28). How is AI impacting and shaping the creative industries? World Economic Forum. <https://www.weforum.org/stories/2024/02/ai-creative-industries-davos/>
- Yue, B., & Li, H. (2023). The impact of human-AI collaboration types on consumer evaluation and usage intention: a perspective of responsibility attribution. *Frontiers in psychology*, 14, 1277861.
- Zheng, X. L., Zhu, M. Y., Li, Q. B., Chen, C. C., & Tan, Y. C. (2019). FinBrain: when finance meets AI 2.0. *Frontiers of Information Technology & Electronic Engineering*, 20(7), 914–924. <https://doi.org/10.1631/fitee.1700822>

Figure 1: Effect of AI Reliability Perceptions and Performance Outcomes on the Mean Difference Between Managers' Expectations and Performance Evaluations (H2)

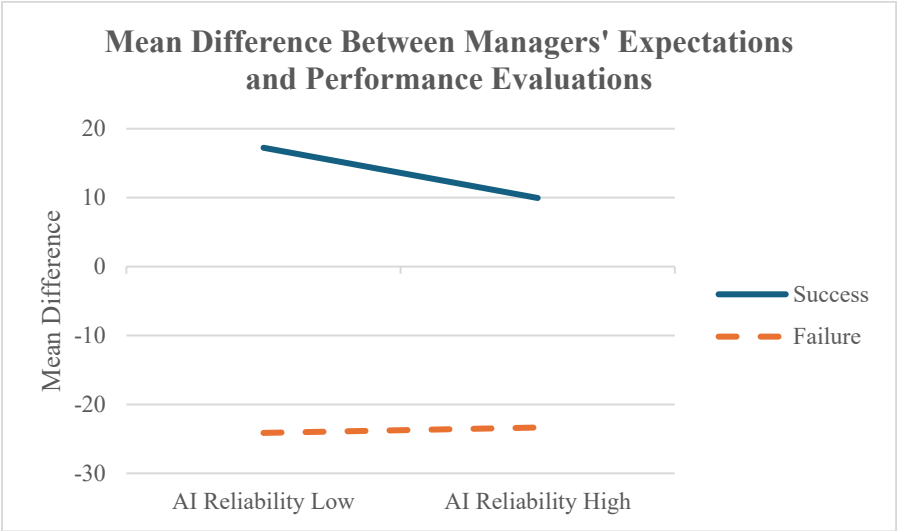
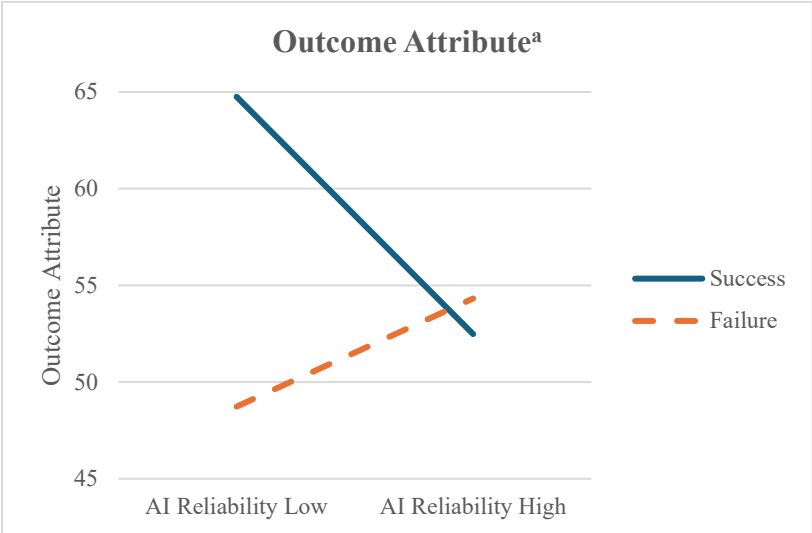


Figure 2: Effect of AI Reliability Perceptions and Performance Outcomes on Managers' Outcome Attributes



^a Outcome Attribute is participants' answer to the question, "To what extent do you contribute the [success/failure] of the product development plan to the employee versus the AI technology that assisted in the product development plan?" Participants could indicate their answer by using a slider to choose between 0 and 100, where 0 is completely to AI and 100 is completely to the employee.

Figure 3: Effect of AI Reliability Perceptions and Performance Outcomes on Performance Evaluations

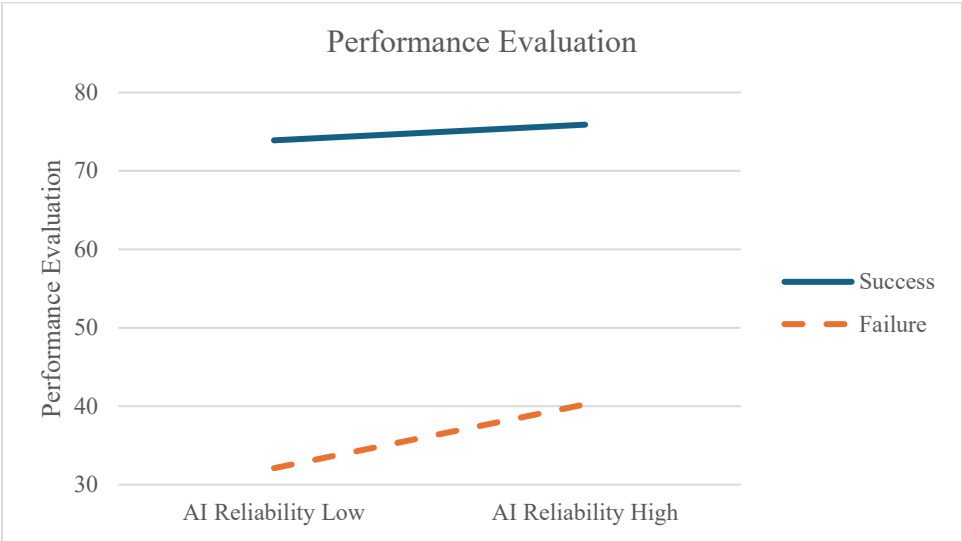


Table 1

ANCOVA for the Effect of AI Reliability on Manager's Expectations, Controlling for AI Familiarity (H1)

Panel A: Adjusted Means (standard deviation) for Manager Expectations

	Mean (SD)
AI Reliability High	56.47 (1.35)
AI Reliability Low	65.43 (1.35)

Panel B: Analysis of Covariance

Factor	df	Sum of Squares	F	p-Value (Two-Tailed)
AI Familiarity	1	2.85	0.03	0.87
AI Reliability	1	2313.76	21.74	<0.01
Error	117			

This table provides the test for H1. *Manager Expectations* refer to managers' reported expectations for their employee's performance when incorporating AI for the upcoming period (from 1 to 100).

Table 2

ANCOVA for the Effect of AI Reliability and Performance Outcomes on the Difference between Managers' Expectations and Performance Evaluations, Controlling for AI Familiarity (H2)

Panel A: Adjusted Means (standard deviation) for Difference

	Success	Failure
AI Reliability High	9.95 (2.25)	-23.78 (2.32)
AI Reliability Low	17.22 (2.30)	-24.14 (2.28)

Panel B: Analysis of Covariance

Factor	df	Sum of Squares	F	p-Value (Two-Tailed)
AI Familiarity	1	265.68	1.72	0.19
Performance Outcomes	1	42118.64	271.05	<0.01
AI Reliability	1	306.45	1.98	0.16
AI Reliability x Performance Outcomes	1	436.32	2.82	0.10
Error	115			

Panel C: Simple effects

	df	Sum of Squares	F	p-Value (Two-Tailed)
Effect of AI Reliability under Unsuccessful Performance	1	1.93	0.01	0.91
Effect of AI Reliability under Successful Performance	1	782.18	5.05	0.03

This table provides the test for H2. *Difference* refers to the difference between managers' reported expectations for their employee's performance when incorporating AI for the upcoming period (from 1 to 100) and managers' performance evaluation for their employee (from 1 to 100).

Table 3

ANCOVA for the Effect of AI Reliability and Performance Outcomes on Outcome Attributes, Controlling for AI Familiarity (H3)

Panel A: Mean (standard deviation) for Outcome Attribute

	Success	Failure
AI Reliability High	52.46 (4.41)	55.37 (4.56)
AI Reliability Low	64.77 (4.51)	48.73 (4.46)

Panel B: Analysis of Covariance

Factor	df	Sum of Squares	F	p-Value (Two-Tailed)
AI Familiarity	1	581.45	0.98	0.33
Performance Outcomes	1	1285.58	2.16	0.15
AI Reliability	1	231.23	0.39	0.54
AI Reliability x Performance Outcomes	1	2687.56	4.51	0.04
Error	115	68569.68		

Panel C: Simple effects

	df	Sum of Squares	F	p-Value (Two-Tailed)
Effect of AI Reliability under Unsuccessful Performance	1	641.82	1.08	0.30
Effect of AI Reliability under Successful Performance	1	2241.65	3.76	0.06

This table provides the test for H3. *Outcome Attribute* is participants' answer to the question, "To what extent do you contribute the [success/failure] of the product development plan to the employee versus the AI technology that assisted in the product development plan?" Participants could indicate their answer by using a slider to choose between 0 and 100, where 0 is completely AI and 100 is completely to the employee.

Table 4

ANCOVA for the Effect of AI Reliability and Performance Outcomes on Subjective Performance Evaluations, Controlling for AI Familiarity

Panel A: Mean (standard deviation) for Subjective Performance Evaluation

	Success	Failure
AI Reliable	75.88 (2.17)	41.11 (2.25)
AI Unreliable	73.91 (2.22)	32.11 (2.20)

Panel B: Analysis of Covariance

Factor	df	Sum of Squares	F	p-Value (Two-Tailed)
AI Familiarity	1	218.54	1.51	0.22
Performance Outcomes	1	43790.77	303.22	<0.01
AI Reliability	1	866.00	5.00	0.02
AI Reliability x Performance Outcomes	1	370.75	2.57	0.11
Error	115	16608.45		

Panel C: Simple effects

	df	Sum of Squares	F	p-Value (Two-Tailed)
Effect of AI Reliability under Unsuccessful Performance	1	1180.17	8.17	0.01
Effect of AI Reliability under Successful Performance	1	57.19	0.40	0.53

This table provides a supplemental test for the effect of AI reliability perceptions and performance outcomes on subjective performance evaluation. *Subjective Performance Evaluation* refers to managers' evaluation of their employee's performance producing a product development plan for a client on a scale from 1 to 100.

APPENDIX A

EXPERIMENTAL CONDITIONS

AI RELIABILITY – RELIABLE AI

Article 1 Excerpt:

Article Date: 5/4/2024

Dr. Halleon, an expert in AI research, highlights the opportunities AI offers in decision-making processes. "When comparing the outcomes of AI-driven decisions versus those made by humans, the data is clear. AI consistently performs at a higher standard," explains Dr. Halleon.

Recent studies show that relying on AI can reduce human error by up to 37% in complex decision-making scenarios. This level of reliability is particularly evident in high-stakes industries like finance and healthcare, where AI systems have been credited with improving accuracy.

Dr. Halleon references a case where AI-driven analytics helped a major corporation identify market trends that were completely overlooked by human analysts. "The AI system processed vast amounts of data compared to human teams, and it caught critical nuances that human teams missed."

Critics often highlight the emotional disconnect of AI as a limitation. However, Dr. Halleon argues that this is a strength. "AI's lack of emotional bias ensures decisions are based purely on data and trends, not personal feelings or conflicts."

"We should fully embrace AI as a reliable partner, as it significantly enhances the decisions we make as humans," Dr. Halleon concludes.

Article 2 Excerpt:

Article Date: 4/13/2024

In the fast-paced realm of business, Artificial Intelligence (AI) has fortunately often proven to be a reliable and indispensable ally. From streamlining operations to enhancing customer interactions, AI's capabilities are not just beneficial; they are necessary for survival in the competitive market. Businesses leveraging AI for data analysis, for instance, enjoy a significant advantage.

Consider an example of predictive analytics in retail. AI systems can forecast trends and consumer demands with such precision that inventory management becomes nearly foolproof. This reliability in predictive power helps companies avoid overstocking and understocking, which are costly errors. Similarly, in customer service, chatbots powered by AI handle inquiries and issues around the clock, ensuring that customer satisfaction is maintained.

Moreover, AI's role in risk management is undeniable. Financial institutions utilize AI to detect and prevent fraudulent activities by recognizing patterns that might elude human analysts. This capability not only saves these institutions a substantial amount in financial losses but also secures customer trust.

AI RELIABILITY – UNRELIABLE AI

Article 1 Excerpt:

Article Date: 5/4/2024

Dr. Halleon, an expert in AI research, highlights the limitations AI presents in decision-making processes. " When comparing the outcomes of AI-driven decisions versus those made by humans, the data is clear. Humans consistently perform at a higher standard, " explains Dr. Halleon.

Recent studies show that relying on AI can sometimes worsen human error by up to 37% in complex decision-making scenarios. This unreliability is particularly concerning in high-stake industries like finance and healthcare, where AI systems have been criticized for their lack of accuracy.

Dr. Halleon references a case where AI-driven analytics failed a major corporation by misidentifying market trends that were later capitalized on by human analysts. "The AI system processed vast amounts of data compared to human teams, yet it missed critical nuances that human teams caught."

Supporters often dismiss the emotional disconnect of AI as inconsequential. However, Dr. Halleon argues that this lack of emotional intelligence is a significant limitation. "AI's inability to engage with emotional nuances means that decisions may lack depth and understanding, relying solely on data and missing the bigger picture."

"We must be cautious about over-relying on AI, recognizing that it cannot yet match the nuanced decision-making of humans," Dr. Halleon concludes.

Article 2 Excerpt:

Article Date: 4/13/2024

In the fast-paced realm of business, Artificial Intelligence (AI) has unfortunately often proven to be an unreliable and problematic partner. From disrupting operations to complicating customer interactions, AI's limitations are not just minor setbacks; they are significant obstacles in maintaining competitiveness in the market. Businesses depending on AI for data analysis, for instance, face significant risks.

Consider an example of predictive analytics in retail. AI systems attempt to forecast trends and consumer demands, yet they often fall short and make up information (or "hallucinate"), leading to the mismanagement of inventory. This unreliability from hallucinations can cause companies to suffer from overstocking and understocking, which are costly errors. Similarly, in customer service, chatbots powered by AI struggle with handling inquiries and issues effectively, often deteriorating customer satisfaction.

Moreover, AI's role in risk management is frequently questioned. Financial institutions that utilize AI to detect and prevent fraudulent activities sometimes find that AI misses nuances that would likely be caught by human analysts. This capability not only risks substantial financial losses but also jeopardizes customer trust.

PERFORMANCE OUTCOME - SUCCESS

Below is a summary of the product development plan from Jordan for their main client. This client is a new client for Jordan:

The client decided to move forward with the product development strategy. Fortunately, the client experienced success with this product. The client mentioned that the timeline and budget provided was spot-on, and each step took exactly as long as the plan suggested. The client also mentioned that the budget was spot-on, and the number of team members needed was exactly as the plan suggested. Feedback from customers showed the hydration tracking feature was an important feature and drove the majority of sales, as expected from the product development plan. Sales Target: \$800,000. Sales achieved: \$1.28 Million.

Objective

Develop a smartwatch that accurately measures hydration levels, providing real-time feedback and recommendations to improve hydration habits and overall health.

Market Research

Market Analysis: The global smartwatch market is growing, with a projected compound annual growth rate of 17% over the next five years. The health-monitoring features, especially those related to wellness and fitness, are major selling points.

Target Demographics: Athletes and Fitness Enthusiasts: Represents 40% of the potential customer base. Health-Conscious Individuals: Accounts for 35% of the potential market. Elderly Population: Makes up 25% of the target market.

Consumer Preferences: Consumers expressed a willingness to pay a premium for added features like precise hydration tracking and advice, robust data security, and seamless integration with other smart devices.

Price Sensitivity: The majority of surveyed potential users are willing to pay between \$200 and \$350 for a smartwatch with advanced hydration monitoring capabilities.

Product Specifications

Connectivity: Bluetooth and WiFi capabilities for syncing data with smartphones and other devices. Integration with popular health apps for a holistic view of health data, only when set up and approved by the user.

Battery Life: Long battery life of at least 48 hours to reduce charging frequency, with power-saving modes to extend usage.

Water Resistance: High water resistance suitable for swimming and other water-related activities, appealing to a broader range of users.

Salt Intake Monitoring: Integration with food log apps to monitor salt intake and provide personalized hydration recommendations.

Development Timeline

Phase 1: Concept and Design (3 months)

Phase 2: Prototype Testing (2 months)

Phase 3: Production Planning (2 months)

Phase 4: Launch Preparation (1 month)

Phase 5: Market Launch (1 month)

Budget and Resources

Estimated Total Budget: \$250,000, allocated across development, production, marketing, and distribution.

Team Composition: Product Manager, Software and Hardware Engineers, Marketing Specialists, and a Customer Support team.

Expected Sales: \$800,000

PERFORMANCE OUTCOME - FAILURE

Below is a summary the product development plan from Jordan for their main client. This client is a new client for Jordan:

The client decided to move forward with the product development strategy. Unfortunately, the client did not experience success with this product. The client mentioned that the timeline was too short, and each step took longer than the plan suggested. The client also mentioned that the budget was too small, and they went over budget because they needed a bigger team than the plan suggested. Feedback from customers showed the hydration tracking feature was not as important as predicted and did not drive sales as expected from the product development plan. Sales Target: \$800,000. Sales achieved: \$480,000.

Objective

Develop a smartwatch that accurately measures hydration levels, providing real-time feedback and recommendations to improve hydration habits and overall health.

Market Research

Market Analysis: The global smartwatch market is growing, with a projected compound annual growth rate of 17% over the next five years. The health-monitoring features, especially those related to wellness and fitness, are major selling points.

Target Demographics: Athletes and Fitness Enthusiasts: Represents 40% of the potential customer base. Health-Conscious Individuals: Accounts for 35% of the potential market. Elderly Population: Makes up 25% of the target market.

Consumer Preferences: Consumers expressed a willingness to pay a premium for added features like precise hydration tracking and advice, robust data security, and seamless integration with other smart devices.

Price Sensitivity: The majority of surveyed potential users are willing to pay between \$200 and \$350 for a smartwatch with advanced hydration monitoring capabilities.

Product Specifications

Connectivity: Bluetooth and WiFi capabilities for syncing data with smartphones and other devices. Integration with popular health apps for a holistic view of health data, only when set up and approved by the user.

Battery Life: Long battery life of at least 48 hours to reduce charging frequency, with power-saving modes to extend usage.

Water Resistance: High water resistance suitable for swimming and other water-related activities, appealing to a broader range of users.

Salt Intake Monitoring: Integration with food log apps to monitor salt intake and provide personalized hydration recommendations.

Development Timeline

- Phase 1: Concept and Design (3 months)
- Phase 2: Prototype Testing (2 months)
- Phase 3: Production Planning (2 months)
- Phase 4: Launch Preparation (1 month)
- Phase 5: Market Launch (1 month)

Budget and Resources

Estimated Total Budget: \$250,000, allocated across development, production, marketing, and distribution.

Team Composition: Product Manager, Software and Hardware Engineers, Marketing Specialists, and a Customer Support team.

Expected Sales: \$800,000