A Mythic Belief Regarding Trust in Artificial Intelligence: Uncovering the Role of Responsibility Perception for AI Use in Decision-Making

Kyootai Lee
Graduate School of Management of Technology
Sogang University
kyootai@sogang.ac.kr

Wooje Cho
School of Business
Seoul National University
woojecho@snu.ac.kr

Han-gyun Woo
Graduate School of Management of Technology
Sogang University
hwoo@unist.ac.kr

Abstract

This study aims to analyze a mechanism of AI responsibility based on attribution theory. It also identifies a new concept, AI locus of control (AI-LOC), reflecting an individual's belief about the degree to which AI determines decision performance. To this end, we built a website with embedded AI systems where participants longitudinally made corporate credit rating decisions. We created a dynamic panel dataset that includes participants' decisions per task and decision performance and attitudes per session. The results revealed that AI-LOC and trust in AI were developed in parallel yet differed over time. AI-LOC positively influenced AI use, but trust in AI did not. We reasoned that individuals would likely exhibit self-serving biases and take an egocentric and disengagement coping strategy regarding their decision-making with AI. This study can contribute to understanding the psychological and behavioral aspects of AI use.

Keywords: Artificial Intelligence, Attribution Theory, Trust, Locus of Control, Decision Making

1. Introduction

Trust in technology can be defined as the belief that a technological artifact possesses certain desirable attributes, making it capable of fulfilling one's expectations [1]. Researchers across academic disciplines have highlighted the importance of trust in artificial intelligence (AI), contending that it is imperative for successful implementation of AI [2, 3]. Accordingly, while focusing mainly on technological characteristics, researchers have also paid attention to identifying the ways that can enhance trust in AI (e.g., [4], [5], [6]). However, some researchers and limited anecdotal evidence have shown that individuals often stop using technologies despite initially forming a positive attitude toward them [7].

Trust in a technology is likely sustained as an outcome of positive experiences with that technology [8, 11] and its perceived reliable performance over time [1]. In the context of decision-making with AI, employees and AI are both prone to making mistakes [7]. Additionally, decision-makers are likely more sensitive to algorithm errors than human mistakes [10] which can increase attributional errors. In this regard, initial trust in AI may not always increase subsequent use because trust is an outcome of attribution [6, 9]. The consequences of individuals' evolving trust in AI [12] should be a critical issue, but it has not yet been investigated longitudinally. Hence, this study aims to answer the question: (RQ1) *How does trust in AI influence AI use over time?*

Users may deliberatively consider not only AI's capabilities but also their own abilities when receiving performance feedback about decision-making [6, 13, 14]. That is, when individuals evaluate the performance of new technologies, an attribution process tends to arise [13, 15], which can influence their continued AI use. Accordingly, attributional thinking is critical in understanding trust [6, 8] and post-adoption IT use behaviors [10]. Although many researchers (e.g., [16], [17], [18], [19]) have begun focusing on responsible AI, they have paid less attention to responsibility development based on performance feedback over time. Though not focusing on attributional theory, several recent researchers have begun emphasizing that decision-makers' sense of AI responsibility can be determined by the perception of decision control [6, 13] in the attribution processes [13, 20]. Hence, the research also aims to answer the following question: (RQ2) *How do individuals perceive the causes of their decision performance?* To this end, this study has a purpose to introduce AI locus of control (AI-LOC)—an individual's beliefs regarding AI as the cause of decision-making, and identify an answer for the following question: (RQ3) *How does AI-LOC affect AI use over time?*

In addressing the above limitations of the extant research, we built a research model to explain the relationships among AI-LOC, trust in AI, and continued AI use based on attribution theory. As such,

URI: https://hdl.handle.net/10125/106538
978-0-9981331-7-1

this study developed AIs capable of providing corporate credit rating recommendations. Next, we built an online website embedding the AIs where participants made corporate credit-rating decisions longitudinally (three sessions over two months). We analyzed data in two steps: auto-regressive (AR) hierarchical linear modeling (HLM) and cross-lagged structural equation modeling (CL-SEM).

The current study makes the following major contributions. First, this research provides a theoretical mechanism to uncover how individuals determine AI responsibility [19, 21] that is not free from self-serving biases in attribution processes [17]. Second, by taking a dynamic and longitudinal perspective as well as delving into the link between trust in AI and its actual use (rather than the self-reported intention to use it), this study reveals the marginal role of trust in facilitating future AI use. Hence, this can extend our understanding of trust in AI (e.g., [4], [6], [13]). Third, while revealing the important role of AI-LOC in increasing AI use, this study demonstrates that individuals are likely to disengage from decision-making processes and increase their dependence on AI advice based on their expected tradeoffs between losses and gains from following AI advice.

2. Theories & Related Work

2.1 Trust in AI

Trust is "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor" ([23], p. 712). Trust was initially defined in the interpersonal domain, but its conceptualization has been increasingly applied to human-technology interactions (e.g., [1], [2], [19], [20]). Researchers have argued that factors enhancing trust are dependent on contexts, for instance, competence, benevolence, and integrity rooted in interpersonal contexts [20] and functionality, helpfulness, and reliability in technological contexts [1, 24]. Particularly, a few researchers (e.g., [14]) have explained that trust development in the context of AI involves calculation, prediction, and capability processes. More specifically, individuals are likely to calculate the costs and benefits of AI use that can occur due to AI agents' (un)trustworthy behaviors. Additionally, people tend to develop confidence in AI's ability to perform as predicted, and recognize AI's capabilities in fulfilling commitments. Thus, the AI features of transparency, reliability, and flexibility can enhance individuals' cognitive trust [14].

Several researchers have recognized trust as an outcome of responsibility or attribution [6, 26], though their contexts did not involve AI advice. For instance, Molm et al.'s [26] definition of trust explicitly stated that individuals' attribution of positive intentions to another party is the basis of their trust formation under uncertainty and risk. In line with these studies, the literature has begun arguing that AI accountability and responsibility are important for its users' beliefs and behaviors [11, 22]. However, empirical findings remain limited in longitudinally uncovering the attribution mechanisms underpinning the formation of trust in the organizational decision-making context in which irreducible uncertainties are embedded [28]. Therefore, it is essential to explore how individuals determine responsibility for the (un)intentional consequences of decision performance [14, 29]. Furthermore, we must also investigate how performance feedback affects trust in AI and subsequently influences usage patterns. This exploration can help researchers understand the dynamic nature of trust in AI [8].

2.2 Attribution Theory

Individuals' attribution is at the core of trust [11]. The tenet of attribution theory is to determine how individuals explain events by asking for the reasons [30] behind their perceptions, judgments, and evaluations of behaviors [31]. According to this theory, individuals use performance as a primary informational cue to assess the causal attributions of their actions [31]. The attributional outcomes can then determine their subsequent expectations and responses [15] to ensure similar outcomes [32]. Particularly in AI contexts, Ha et al. [13] argued that when individuals obtain appropriate explanations for the outcomes, they are more likely to attribute the outcomes to the AI and discount other causes.

Causal attribution has three primary dimensions: locus of control (i.e., internal [vs. external] attribution to the decision maker [vs. AI] in this context), controllability (i.e., the level of voluntary control by the decision maker), and stability (i.e., the extent to which the decision maker perceives the cause as dynamic or constant) [15, 20]. When individuals consider the accountability of decision performance, locus of control (LOC) may determine who is responsible for a given outcome [33, 34]. In the context of AI-assisted decision-making, people consider either themselves or AI in the attribution process [13]. Accordingly, we define AI-LOC as the extent to which individuals tend to perceive that AI is accountable for their decisions. Additionally, individuals may feel lower controllability over their decision performance, they can perceive helplessness and vulnerability [33]. Finally, stability can influence individuals' expectations of the future accuracy of AI advice [33], which may, in turn, affect their subsequent usage. As individuals perceive AI accuracy over time, they learn what to expect from AI usage and how to leverage its advice more effectively. As such, trust dynamics may emerge over time [34]. In sum, individuals' perceptions of AI responsibility and their resulting trust in AI can be subject to a longitudinal attributional process.

3. Hypotheses

3.1. Attributions for Decision Performance

When individuals evaluate AI accuracy, they use the outcome information—part of performance feedback—to process the causal attributions of their past actions [6, 14, 15]. When individuals achieve performance beyond their expectations in their AI-assisted decision-making, they are more likely to realize AI's abilities [12]. Accordingly, they are less likely to perceive vulnerability regarding expected outcomes and are instead more likely to build trust in AI [3, 37]. That is, when individuals make decisions with AI and achieve their expected goals over time, they are more willing to develop and sustain their trust in AI [3, 37]. Thus, the so-called positive spirals of trust can be maintained [11]. Thus, we expect that decision performance increases trust in AI (*H1-1*).

Furthermore, it is natural to consider AI capabilities and personal abilities while evaluating decision performance. Attribution processes are not spared from 'errors of judgment' [11]. More specifically, self-serving biases are prevalent in the process [6, 22], highlighting individuals' tendencies to attribute positive (negative) outcomes to internal (external) factors, i.e., the so-called discounting principle [13]. In doing so, individuals are motivated to enhance themselves and seek knowledge about their specific contexts [39]. Similarly, in the context of information systems, Jörling et al. [21] found that when users face disruptive events, they tend to blame software in 73.85% of the cases but blame themselves in only 14.29% of cases. Thus, individuals may be more likely to attribute high performance to themselves but low performance to AI in this context. Overall, when individuals achieve a high level of decision performance, they build trust in AI but attribute the performance to themselves. Thus, we expect that decision performance will increase trust in AI rather than AI-LOC. Formally, we expect that the influence of decision performance on trust in AI is higher than on AI-LOC over time (*H1-2*).

3.2. AI Use Behaviors as an Outcome

How individuals behave tends to rely on expectations of reciprocity [40]–what they expect to gain from their current inputs. Applying this to AI-assisted decision-making contexts, people may make educated guesses about AI's performance and respond accordingly (cf., [23]). Even if they trust AI advice after evaluating the positive outcomes, they may speculate about not only the outcomes relevant to themselves (e.g., potential gains from following the advice [14]) but also the party responsible for the outcomes. Particularly, as AI can recommend only stochastically accurate answers than humans in the context of many organizational decisions [29], individuals should consider who is responsible for the tradeoffs between losses and gains from following AI advice [14]. In short, taking AI advice should be strategic [40].

In this situation, despite AI's higher accuracy rate, decision-makers may not always follow AI advice, because individuals tend to exhibit algorithm aversion and thus prefer riskier (and often suboptimal) human advice [9, 41]. Additionally, people need to learn how to utilize AI advice and recognize the need to put personal effort into making decisions. Similarly, Gefen et al. [10] found that, in online shopping contexts, new users' behaviors tend to be determined by trust, whereas experienced users rely more on recommendation agents' accuracy. While highlighting performance feedback, researchers [14, 35] have shown that though individuals initially build trust in AI, erroneous AI advice can decrease their trust and recovery takes time. Accordingly, although individuals trust AI, they may not always follow its advice. Instead, they may also feel responsible for decision performance and only selectively accept AI advice. This argument can be understood as a dynamic coping process in individuals' adaptations to a new technology, demonstrating that they may change from one coping strategy to another strategy based on an evaluation of their previous efforts [41]. Hence, without individuals' recognizing AI responsibility, trust in AI may not always lead to following AI advice.

On the one hand, recognizing AI abilities and ascribing responsibility for the performance may lead individuals to take AI advice selectively. On the other hand, they may feel less control over their decision contexts and less skill in overcoming situations. For example, because individuals perceive that their past efforts did not contribute enough to their decision performance compared to AI advice, they may feel that fewer cognitive resources are available and may be more motivated to adopt a disengagement coping strategy [42]. Such low-controllability situations tend to increase withdrawal behaviors [30]. This coping strategy may lead individuals to avoid solving their own problems and instead rely on AI advice, increasing AI usage. Likewise, Ha et al. [13] found that individuals with high power status are more likely to perceive greater control over AI outputs than those with low power status. Interestingly, low controllability of AI outputs is positively associated with confidence in AI's capabilities. de Guinea [43] found that individuals tend to switch to a disengagement coping strategy when they feel a lack of control over an IT event. Notably, some researchers found that individuals' decreasing perceptions of control may reduce their AI use when they cannot change AI advice and replace AI roles (e.g., [16], [44]). However, when people can input their expertise in the final decisions but simultaneously perceive less control than AI, they may realize that following AI advice may be optimal. Thus, we expect that AI-LOC increases continued AI use (*H2-1*).

Overall, individuals may change their decision-making strategies as the situations unfold over time, demonstrating dynamic behavioral patterns. Hence, we expect that while the effect of trust on future AI use may not be sustained over time, that of AI-LOC on future AI use can remain intact; Formally, the influence of AI-LOC on continued AI use is stronger than that of trust in AI (*H2-2*).

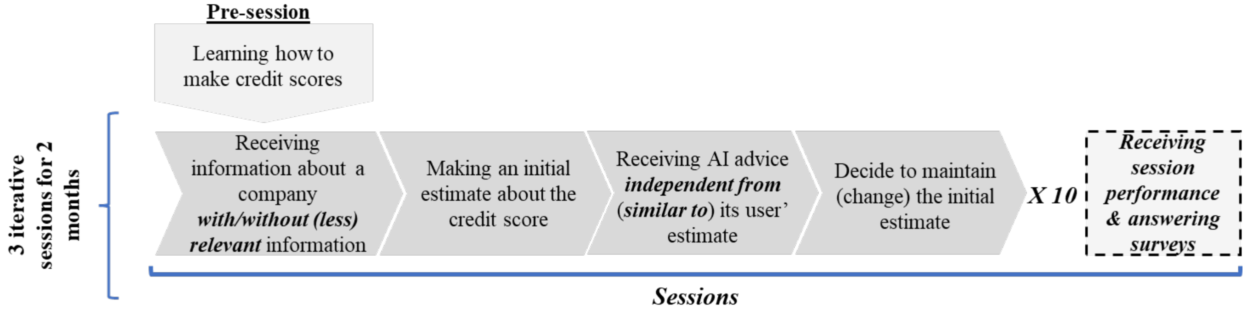

Pre-session
Learning how to make credit scores
3 iterative sessions for 2 months
Receiving information about a company with/without (less) relevant information
Making an initial estimate about the credit score
Receiving AI advice independent from (similar to) its user' estimate
Decide to maintain (change) the initial estimate
X 10
Receiving session performance & answering surveys
Sessions

Figure 1. Credit-rating Decision Procedure

4. Method

4.1. Overview of Longitudinal Studies

This research examines a corporate credit-rating context that naturally includes irreducible uncertainty [28]. Thus, individuals and AI can both only estimate stochastically correct answers. For this study, we developed websites with embedded AI recommendation agents for aiding participants in solving credit rating problems. The websites were designed to provide financial (e.g., total assets, paid-in capital, annual sales) and non-financial (e.g., number of IPs and locations) information as well as AI advice for making credit decisions. We chose the information by following the common industry and academia practices [45]. Furthermore, we differentiated the conditions regarding the AI type (*neutral* or *user-dependent* AI) and the type of extra information (relevant or less relevant information). Thus, we enhanced the generalizability of the results and represented better real-world decision contexts.

4.1.1. AI-based Systems Development

We created four websites representing unique decision conditions, allowing us to identify their impact on individuals' attitudes (see Figure 1). First, we developed two AIs to identify the effect of AI advice on user attitudes [47]. We assumed that (1) a system recommending credit scores based only on historical data for credit events (i.e., neutral AI) may be less preferred than (2) a system recommending credit scores based on both historical data and users' initial judgments (i.e., user-resembling AI). It is because individuals tend to prefer the recommendations including their opinions [3]. Then, we nested the AIs in the websites.

Second, we offered either the focal company's previous-year credit score (relevant information) or the average previous-year credit score of companies in the same region (less relevant information) along with the AI advice. This condition represents the practices that employees are exposed to multiple information cues in their decision-making processes. Several recent studies on non-AI contexts have identified that even one additional source of information can exert a sizable effect on a final decision [47]. Therefore, depending on its relevance, they tend to leverage the extra information differently [48].

We employed historical credit data from Korea Enterprise Data, one of South Korea's largest credit-rating service companies. Additionally, we developed AI algorithms based on linear regression, random forest, and gradient-boosted decision tree techniques offered by the caret and xgboost packages in R statistics. Then, the neutral AI was trained with 113,000 credit scores from 2002 to 2014 in the database, yielding a root mean square error (RMSE) of 2.513. Further, we compared the difference between the advice and credit scores in the database ($dif_{AI\text{-}DB}$) and the difference between participants' initial estimates and credit scores in the database ($dif_{IE\text{-}DB}$). Our finding indicated that AI was significantly more accurate than individuals ($t = 28.502$, $df = 7,709$, $p < .001$). From 2015 to 2017, 842 credit events in the database were used to administer the tasks with which individuals made decisions using AI advice.

We developed four types of online sites. Site (A) provided neutral AI advice, and Site (B) offered user-resembling AI advice. Neither contained any (less) relevant additional information. That is, the participants who used websites (A) and (B) had a different AI-type condition but an identical condition in terms of extra information cues. Site (C) offered user-resembling AI advice and relevant extra information. As for site (D), it gave user-resembling AI advice and less relevant information. Compared to the participants who used site (B), those who employed sites (C) and (D) were exposed to unique conditions regarding the existence of (less) relevant information. Additionally, sites (C) and (D) could create different conditions regarding the informational relevance. Through these sites, we can efficiently create diverse conditions and effectively isolate the possible impact of the unique attributes.

In total, 226 participants who had completed the three sessions were included in the analysis. On sites (A), (B), (C), and (D), 57, 65, 63, and 41 participants completed the three sessions, respectively. Out of 226 participants, 135 were men, 107 were younger than 25, 71 were between 26 and 35, and 48 were over 35. Among the participants, 156 were in business-related disciplines. The remaining were in engineering and the humanities but had business and economics minors.

4.1.2. Research Procedure and Participants

We recruited individuals from four major South Korean universities, compensating them with 15,000 Korean won (KRW, 1 USD = 1,100 KRW as of January 2021) for their participation. They were all senior-level undergraduate or master-level students. We motivated them to enhance their learning and decision accuracy efforts by offering 50,000 to 100,000 KRW based on performance outcomes. The different website links were distributed via email, thereby making participation voluntary. We randomly assigned websites to the participants. Individuals did not know the type of websites, which remained the same for all three sessions. The firms' names were anonymized so that participants could not rely on external sources of information.

Once participants accessed the website, they created IDs and passwords. As shown in Figure 1, before starting the credit rating tasks, participants learned how to rate credit scores based on numerical data about organizations. Participants needed to rate ten firms per session. They could then advance to the next session five days after finishing the previous session and were required to finish all three sessions. All participants were given the same questions, but the task order was randomized. If all ten tasks were not completed before leaving the site, they could not participate in the experiment again.

Table 1. Summary of Measures

Constructs	Mean (SD)		
Variant to time and experiments			
	Session 1	Session 2	Session 3
Decision Performance	16.141 (1.759)	17.202 (1.538)	17.788 (1.074)
Continued AI Use	.317 (.248)	.308 (.246)	.307 (.252)
Trust in AI	3.400 (.790)	3.662 (.819)	3.706 (.912)
AI-LOC	2.673 (.858)	2.761 (.912)	2.814 (.953)
Invariant to time			
$PreFoc_{ip}$	.279 (.449)		
$PreLoc_{ip}$	.181 (.386)		
AI Type (*$AIType_{ip}$*)	.748 (.435)		
Invariant to time and experiments			
Major	2.991 (2.288)		
Job	6.345 (2.518)		
Gender	1.403 (1.456)		
Age	3.164 (1.527)		

PreFoc (Previous Year Credit Score of Focal Company), PreLoc (Previous Year Average Credit Score of Local Companies)

Participants initially estimated each company's credit score before receiving AI advice and then had a chance to adjust the initial score after receiving it. Individuals using sites (A) and (B) made preliminary estimates based only on companies' numerical information. Those using sites (C) and (D) could use the focal firm's information combined with the (less) relevant additional information. Next, individuals received AI advice. We informed participants that they could leverage the advice strategically rather than merely following it. Based on their valuation of AI advice, they could discretionally adjust their initial estimates only if they answered "Yes" to the question "Will you change the initial estimate?" After completing the ten tasks in a session, they could review their session (not each task) performance. Based on the feedback, they could update their attitudes toward AI advice across sessions.

4.2. Measures

The unit of analysis was session–individual observations. Accordingly, we created continued AI use and decision performance for the i^{th} individual for the q^{th} tasks at the p^{th} session and aggregated the two variables at the session level. We also created AI-LOC and trust in AI for the i^{th} individual at the p^{th} session by asking a series of survey questions after each completed session. Table 1 summarizes the measures.

Continued AI Use. We measured continued AI use in terms of individuals' switching decision after they had received AI advice per task with their answer to the question, "would you change the initial estimate?" (AIU_{ipq} = 1 for yes, 0 for no). Then we averaged the measures per session to indicate session-level AI use ($AIU_{ip} = \Sigma AIU_{ipq}/10$)

Decision Performance. We measured individuals' decision accuracies (DP_{ipq}) by computing the difference between their final estimates (FE) and the actual credit scores in the database (*D_Val*) as 20 - $|FE_{ipq} - D\_Val_{ipq}|$ per tasks. Decision accuracies per task were then aggregated by averaging them ($DP_{ip}= \Sigma DP_{ipq}/10$).

Trust in AI. We employed Venkatesh et al.'s [49] scale to measure trust in AI (7-point Likert scale). Participants were asked to answer the questions after completing each session (Cronbach's Alpha = .820).

AI-LOC. Ifinedo's [50] single item was used to measure AI-LOC (7-point Likert scale). Participants replied to a question after finishing each session.

Control variables. We included *AI type* that participants used and *information type* that represented whether received (less) relevant information. We also included educational background and job to control for domain knowledge's effect on using AI advice [28]. Age, gender, and university were added to control for the unobserved effects of individuals [51]. Finally, we included session dummies to control for individuals' experience (Exp_{ip}) by counting the sequential session number of (1-3).

4.3. Analysis

We built dynamic panel datasets, which include 678 (= 226 individuals X 3 sessions) longitudinal observations. Our datasets contain the same respondents' responses at multiple times. The session-level data are nested in individuals. Thus, within- and between-session variances may not be independent of individuals' unobserved characteristics. In addition, this study aims to identify the time-lagged effects of AI-LOC and trust in AI. Accordingly, we leveraged two analytical methods–autoregressive hierarchical linear modeling (AR-HLM; [52]) and cross-lagged structural equation modeling (CL-SEM; [53]).

AR-HLM was used to identify the relationships of decision performance with AI-LOC and trust in AI over sessions while controlling for session invariant factors (i.e., individuals' demographics). Fit indices consistently indicate that a model with autocorrelation fits better than that without it, supporting the first-order autocorrelation (AR[1]). We controlled the serial correlations among variables measured at different time points from the same individuals. Intercepts were random at the individual level (Level 2), and the covariates were assessed at both the individual level (Level 2) and the session level (Level 1). CL-SEM seeks to ascertain the causality of trust in AI and AI-LOC to continued AI use by temporal separations.

5. Results

Before testing the hypotheses, we examined within- (between-) individual variations of AI-LOC and trust in AI. One-way analysis of variance (ANOVA) with random effects revealed that 21.041% and 18.631% of the variance in AI-LOC and trust in AI were explained by individuals (computed from the intercept variance; cf. null models [Models 1 and 5] in Table 2). The results consistently indicated that experience positively influences trust in AI (minimum coefficient = .105, p = .007) and AI-LOC (minimum coefficient = .071, p = .043). Regarding the variance of trust in AI and AI-LOC, 21.997% and 21.311%, respectively, are explained by experience and individual demographics (Models 2 through 4 for trust in AI and Models 6 through 8 for AI-LOC in Table 2). The results highlight the heterogeneity across individuals and the dynamics of AI-LOC and trust in AI over time, i.e., the primary assumptions of our research. The results (Model 4) indicate that decision performance significantly relates to trust in AI (coefficient = .094, p = .028), supporting H1-1. However, it did not influence AI-LOC (coefficient = -.060, p = .191). The results indicate that when individuals achieve higher performance, they increase their trust in AI but do not attribute the performance to AI. These results support H1-2.

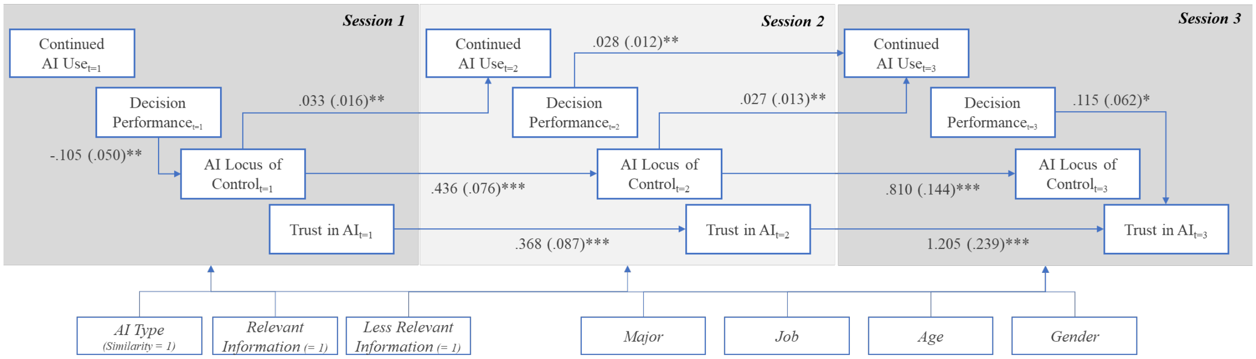
Figure 2 displays the results of a cross-lagged model that specifies the relationships among focal variables over time. The fit statistics for the model, χ^2 (37) = 72.467, CFI = 0.935, RMSEA = .064, and SMSR = 0.050, confirmed a good fit with the data. Hence, we proceeded to test the structural model. After accounting for the covariates, AI-LOC in session 1 was positively related to continued AI use in session

2 (coefficient = .033, p = .016), and a significant relationship was also found between sessions 2 and 3 (coefficient = .027, p = .013). However, trust in AI in session 1 (2) was not related to continued AI use in session 2 (3). These results provide evidence to support H2-1 and H2-2.

Table 2. HLM Results

Dependent Variable:	Model 1			Model 2			Model 3			Model 4		
Trust in AI	Coef.	SE	p-value	Coef.	SE	p-value	Coef.	SE	p-value	Coef.	SE	p-value
(Intercept)	3.582	.044	.000	3.527	.257	.000	3.410	.290	.000	3.463	.292	.000
Experience				**.153**	**.032**	**.000**	**.153**	**.032**	**.000**	**.105**	**.038**	**.007**
Major				-.011	.020	.573	-.010	.020	.634	-.009	.020	.657
Job				-.024	.019	.211	-.019	.019	.318	-.019	.019	.323
Age				.026	.033	.421	.030	.035	.394	.031	.035	.377
Gender				-.106	.094	.261	-.095	.095	.315	-.091	.095	.338
PreFoc							-.233	.120	.053	-.301	.124	.016
PreLoc							-.055	.135	.685	.037	.142	.797
AI type							.174	.128	.175	.218	.130	.094
Decision Performance										**.094**	**.043**	**.028**
AIC	1606.025			1613.667			1622.306			1623.943		
BIC	1624.095			1654.259			1676.375			1682.499		
logLik	-799.012			-797.834			-799.153			-798.971		
Dependent Variable:	Model 5			Model 6			Model 7			Model 8		
AI LOC	Coef.	SE	p-value	Coef.	SE	p-value	Coef.	SE	p-value	Coef.	SE	p-value
(Intercept)	2.748	.048	.000	2.608	.277	.000	2.789	.314	.000	2.755	.314	.000
Experience				**.071**	**.035**	**.043**	**.071**	**.035**	**.043**	**.101**	**.042**	**.016**
Major				.036	.021	.091	.034	.022	.116	.034	.021	.119
Job				-.024	.020	.245	-.018	.021	.375	-.018	.021	.370
Age				-.035	.035	.314	-.059	.038	.123	-.060	.038	.117
Gender							.107	.102	.297	.105	.102	.306
PreFoc							-.170	.129	.191	-.126	.133	.343
PreLoc							-.052	.146	.722	-.109	.152	.472
AI type							-.100	.138	.469	-.129	.139	.356
Decision Performance										-.060	.046	.191
AIC	1678.253			1703.357			1712.589			1717.214		
BIC	1696.323			1743.949			1766.658			1775.770		
logLik	-835.126			-842.678			-844.294			-845.607		

PreFoc (Previous Year Credit Score of Focal Company), PreLoc (Previous Year Average Credit Score of Local Companies)



Chi Squared = 72.467 (df = 37); CFI = .935; RMSEA = .064; SRMR = .050

Figure 2. Cross lagged Model Results

6. Discussion and Conclusion

6.1. Discussion

This study identifies an egocentric and dynamic coping strategy when individuals leverage their decision-making with AI advice. Hence, trust in AI may not play an important role in enhancing subsequent AI use, while AI-LOC both positively and consistently affects AI use over time. The results can help answer the following three questions (detailed below), which contribute to constructing a deeper understanding of AI use behaviors based on attributional mechanisms.

How do individuals ascribe AI responsibility? Researchers have highlighted the concept of responsible AI [17, 19, 45], which plays a fundamental role for building trust [6, 13]. In this regard, our longitudinal investigation provides preliminary evidence on the mechanism of trust formation and responsibility in the AI context based on attribution theory. It also identifies a new concept of AI-LOC. By doing so, this study illuminates how individuals attribute decision performance when employing AI advice. AR-HLM results show that people build AI-LOC and trust in AI over time. Interestingly, after recognizing outcomes and evaluating the causes [37], people do not attribute their outcomes to AI despite their increasing trust in AI. These results demonstrate that assigning responsibility may not be error-free [11]. Such biased perceptions may distort the perception of performance benefits from AI.

Though the HLM results show the relationships over the three sessions, the CL-SEM results highlight the distinct impacts of decision performance on AI-LOC and trust in AI within the sessions. As Figure 2 demonstrates, when we scrutinized the CL-SEM results, decision performance negatively affected AI-LOC only in session 1 and positively influenced trust in AI only in session 3. These findings indicate that individuals may initially attribute their performance to themselves rather than to AI, showing a self-serving bias. However, they may realize they have less control over decision contexts and begin attributing their performance to AI, which helps build trust in AI over time. A correlation between AI-LOC and trust in AI was not significant in sessions 1 and 2 but became significant ($r = .164$, $p < .05$) in session 3. These results further support the notion that these two psychological constructs can be distinct and be developed differently over time. These demonstrate that attribution theory may indeed have great potential in revealing a new mechanism of responsible AI research (e.g., [6], [13]).

How does trust in AI impact AI use? A dominant view in this stream of research on trust in AI has been that if users successfully complete a task with a certain technology, they are likely to build trust and use that same technology again in the future [2, 4, 21]. Our research extends this stream of studies by demonstrating that individuals' trust in AI tends to show a dynamic pattern over time [35], and its outcome behaviors may arise from egocentric reasoning [14, 39]. Thus, even though they trust AI and intend to accept the inherent vulnerability of taking its advice based on their trust, individuals may still evaluate both the potential outcomes of their trust and the responsibility for the outcomes [13]. Therefore, researchers should focus on the dynamic coping strategies employed over time when individuals make decisions using AI advice to better understand the roles of trust in AI and its (continued) use (e.g., [42]). This issue is particularly important in decision contexts where correct answers can only be stochastically expected because users are at risk of incorrect decisions.

How does AI-LOC enhance AI use over time? This study identifies the new concept of AI-LOC. AI-LOC is an external LOC and reflects an individual's belief about the degree to which AI determines decision performance. This, in turn, tends to be developed in parallel with trust in AI in their attribution processes. The results indicate that AI-LOC tends to increase future use behavior. Similar to the findings about the relationship between engagement in the decision process and outcome responsibility reported in attribution studies (e.g., [13], [14], [21]), when individuals perceive that AI is responsible for their performance outcomes, they may become disengaged in the decision processes. This could be because individuals may perceive that their capabilities cannot outweigh AI accuracy. Thus, their dependence on AI advice may increase in discretionary decision-making contexts. That is, if users realize that AI is better at making credit rating decisions, they may perceive their aptitude in performing the tasks to be lacking (e.g., "I'm not good at making credit scores" in this research context). In turn, they are more likely to disengage cognitively from the tasks while increasing their dependency on and continued use of AI. Therefore, our results contribute to contemporary explainable AI research [13, 33] in that when decision-makers recognize that the cause of their experienced outcome is controllable by AI and is stable, their perception of AI-LOC may be an important determinant of AI use behaviors, even without losing trust in AI.

6.2. Limitations & Directions for Future Research

As with other research, this study has several limitations. First, we administered a single item to measure AI-LOC. Hence, this paper is limited in fully identifying its psychometric properties. Future studies must delve into AI-LOC's psychological nature as a part of the external LOC. Second, we argued that individuals may disengage from their decision-making, which can increase their dependence on AI advice. If so, an area of further research is whether AI use can increase automatic cognitive processes. If this is the case, users' learning about leveraging AI advice may facilitate but limit their learning about tasks (credit ratings in this study). Future studies must delve into the cognitive processes that combine learning and attribution, which can offer unique aspects of AI-assisted decision-making.

7. Conclusion

This research demonstrates that employees' daily decision-making events can lead to the emergence of divergent beliefs in and attitudes toward AI over time. In this context, organizations may expect that accurate AI can help employees improve decision performance and increase their initial trust in AI, enhancing their subsequent AI use. The results uncover that the implementation strategy based on such expectations may not always occur. Instead, while attributing the decision accuracy to AI and building trust, employees may disengage in the decision processes. Hence, organizations must highlight who is responsible for decision outcomes so that employees are aware of this when using AI to make decisions.

8. References

[1] McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems* (TMIS), 2(2), 1-25.

[2] Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660.

[3] Haque, A. B., Islam, A. N., & Mikalef, P. (2023). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change*, *186*, 122120.

[4] Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. In *Proceedings of the 54th Hawaii international conference on system sciences* (pp. 5463–5472).

[5] Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change*, *181*, 121763.

[6] Sharan, N. N., & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon*, 6(8), e04572.

[7] Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 351-370.

[8] Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems*, 22(2), 281-303.

[9] Hatzakis, T. (2009). Towards a framework of trust attribution styles. *British Journal of Management*, *20*(4), 448-460.

[10] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170.

[11] Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27, 51-90.

[12] Cabiddu, F., Moi, L., Patriotta, G., & Allen, D. G. (*forthcoming*). Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal*.

[13] Ha, T., Sah, Y. J., Park, Y., & Lee, S. (2022). Examining the effects of power status of an explainable artificial intelligence system on users' perceptions. *Behaviour & Information Technology*, *41*(5), 946-958.

[14] Shamim, S., Yang, Y., Zia, N. U., Khan, Z., & Shariq, S. M. (2023). Mechanisms of cognitive trust development in artificial intelligence among front line employees: An empirical examination from a developing economy. *Journal of Business Research*, *167*, 114168.

[15] Kelley, H., Compeau, D., Higgins, C. A., & Parent, M. (2013). Advancing theory through the conceptualization and development of causal attributions for computer performance histories. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *44*(3), 8-33.

[16] Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433-1450.

[17] Merhi, M. I. (2022). An Assessment of the Barriers Impacting Responsible Artificial Intelligence. *Information Systems Frontiers*, 1-14.

[18] Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, *31*(3), 257-268.

[19] Al-Dhaen, F., Hou, J., Rana, N. P., & Weerakkody, V. (2021). Advancing the Understanding of the Role of Responsible AI in the Continued Use of IoMT in Healthcare. *Information Systems Frontiers*, 1-20.

[20] Bansal, G., & Zahedi, F. M. (2015). Trust violation and repair: The information privacy perspective. *Decision Support Systems*, 71, 62-77.

[21] Jörling, M., Böhm, R., & Paluch, S. (2019). Service robots: Drivers of perceived responsibility for service outcomes. *Journal of Service Research*, *22*(4), 404-420.

[22] Allen, M. S., Robson, D. A., Martin, L. J., & Laborde, S. (2020). Systematic review and meta-analysis of self-serving attribution biases in the competitive context of organized sport. *Personality and Social Psychology Bulletin*, 46(7), 1027-1043.

[23] Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709–734.

[24] Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880-918.

[25] Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for information systems*, 6(3), 72-101.

[26] Molm, L. D., Takahashi, N., & Peterson, G. (2000). Risk and trust in social exchange: An experimental test of a classical proposition. *American Journal of Sociology*, 105(5), 1396-1427.

[27] Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets, 31*(2), 447–464.

[28] Choudhury, P., Starr, E., & Agarwal, R. 2020. Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal,* 41(8)*,* 1381–1411.

[29] Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56-62.

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

[31] Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107-128.

[32] Hufnagel, E. M. (1990, January). User satisfaction-are we really measuring system effectiveness. In *Twenty-Third Annual Hawaii International Conference on System Sciences* (Vol. 4, pp. 437-446). IEEE Computer Society.

[33] Tomlinson, E. C., & Mryer, R. C. (2009). The role of causal attribution dimensions in trust repair. *Academy of Management Review*, *34*(1), 85-104.

[34] Vesa, M., & Tienari, J. (2022). Artificial intelligence and rationalized unaccountability: Ideology of the elites?. *Organization*, *29*(6), 1133-1145.

[35] McKnight, D. H., Liu, P., & Pentland, B. T. (2020). Trust change in information technology products. *Journal of Management Information Systems*, *37*(4), 1015-1046.

[36] Efendić, E., Van de Calseyde, P. P., & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes*, *157*, 103-114.

[37] De Baets, S., & Harvey, N. (2020). Using judgment to select and adjust forecasts from statistical models. *European Journal of Operational Research*, 284(3), 882-895.

[38] Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, 31(1), 457-501.

[39] Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust. *Judgment and Decision Making*, *9*(2), 90-103.

[40] Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*(10), 1302-1314.

[41] Beaudry, A., & Pinsonneault, A. (2005). Understanding user responses to information technology: A coping model of user adaptation. *MIS Quarterly*, 493-524.

[42] Inesi, M. E., Botti, S., Dubois, D., Rucker, D. D., & Galinsky, A. D. (2011). Power and choice: Their dynamic interplay in quenching the thirst for personal control. *Psychological Science*, *22*(8), 1042-1048.

[43] de Guinea, A. O. (2016). A pragmatic multi-method investigation of discrepant technological events: Coping, attributions, and 'accidental' learning. *Information & Management*, *53*(6), 787-802.

[44] Huo, W., Zheng, G., Yan, J., Sun, L., & Han, L. (2022). Interacting with medical artificial intelligence: Integrating self-responsibility attribution, human–computer trust, and personality. *Computers in Human Behavior*, 132, 107253.

[45] Caporale, G. M., Cerrato, M., & Zhang, X. (2017). Analysing the determinants of insolvency risk for general insurance firms in the UK. *Journal of Banking & Finance*, *84*, 107-122.

[46] Himmelstein, M., & Budescu, D. V. (2023). Preference for human or algorithmic forecasting advice does not predict if and how it is used. *Journal of Behavioral Decision Making*, *36*(1), e2285.

[47] Bhatia, N., & Gunia, B. C. (2018). "I was going to offer $10,000 but…": The effects of phantom anchors in negotiation. *Organizational Behavior and Human Decision Processes*, *148*, 70-86.

[48] Switzer III, F. S., & Sniezek, J. A. (1991). Judgment processes in motivation: Anchoring and adjustment effects on judgment and behavior. *Organizational Behavior and Human Decision Processes*, *49*(2), 208-229.

[49] Venkatesh, V., Thong, J. Y., Chan, F. K., Hu, P. J. H., & Brown, S. A. (2011). Extending the two-stage information systems continuance model: Incorporating UTAUT predictors and the role of context. *Information Systems Journal*, *21*(6), 527-555.

[50] Ifinedo, P. (2014). Information systems security policy compliance: An empirical study of the effects of socialisation, influence, and cognition. *Information & Management*, *51*(1), 69-79.

[51] Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103.

[52] Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

[53] Burkholder, G. J., & Harlow, L. L. (2003). An illustration of a longitudinal cross-lagged design for larger structural equation models. *Structural Equation Modelling*, *10*(3), 465-486.