Shaping Human-AI Collaboration in Education: Effects of AI-Assisted Decision-Making Paradigms and Human-AI Decision Consistency on Pre-Service Teachers' Psychological States and Performance

Yingying Wang[1,2*], Qin Ni[3†], Haoxin Xu[4], Jiaqi Yin[5], Tingjiang Wei[1,2]

[1]School of Computer Science and Technology, East China Normal University, 3663 Zhongshan North Road, Shanghai 200062, China
[2]Shanghai Institute of Artificial Intelligence for Education, East China Normal University, 3663 Zhongshan North Road, Shanghai 200062, China
[3]Key Laboratory of Multilingual Education with AI, Shanghai International Studies University, 1550 Wengxiang Road, Shanghai 201620, China
[4]School of Education, Shanghai International Studies University, 1550 Wengxiang Road, Shanghai 201620, China
[5]School of Education, Hainan Normal University, 99 Longkun South Road, Haikou 571158, China
52215901020@stu.ecnu.edu.cn, niqin@shisu.edu.cn

Abstract

Artificial intelligence is playing an increasingly important role in supporting decision-making, particularly in educational contexts, where it serves as a critical tool to assist teacher judgment and optimize instructional decisions. However, limited research has examined how different AI-assisted decision-making paradigms influence the Performance of human-AI collaboration, as well as the underlying psychological mechanisms and causal pathways. Therefore, this study investigated 59 pre-service teachers to examine how AI-assisted decision-making paradigms and human-AI consistency influenced their psychological states and task performance. Specifically, this study employed a two-factor mixed experimental design, with the AI-assisted decision-making paradigms as the between-subjects factor and human-AI consistency as the within-subjects factor. Data were analyzed using the Bayesian cumulative link mixed model and structural equation modeling. The results reveal that AI-assisted decision-making paradigms do not have a significant direct effect on task performance. However, when the moderating role of human-AI decision consistency is taken into account, the effect of AI-assisted decision-making paradigms on task performance can exert its influence indirectly through a sequential psychological pathway involving users' confidence and their trust in the AI. Consistency between human and AI decisions not only significantly enhances users' trust in AI, confidence, and task performance, but the proportion of consistent decisions also significantly moderates the impact of AI-assisted decision-making paradigms on users' confidence levels. Notably, our findings indicate that users maintain a moderately level of trust in AI even when their decisions diverge from those of AI. In summary, this study highlights the mediating mechanism by which AI-assisted decision-making paradigms influence task performance through psychological states and identifies the moderating role of human-AI consistency in this pathway. These findings advance the theoretical understanding of human-AI interaction models in educational contexts and offer mechanistic insights to guide the optimization of instructional AI systems.

*Corresponding author. Email: 52215901020@stu.ecnu.edu.cn
†Corresponding author. Email: niqin@shisu.edu.cn

Introduction

Currently, a critical challenge in education is that teachers in large classrooms often struggle to provide timely, personalized assessments and feedback for each student (Schaffer et al. 2017). This issue often hinders educators from accurately tracking students' learning progress and delivering personalized instruction. The application of artificial intelligence (AI) in education effectively addresses this gap and is increasingly becoming a vital tool for educators to evaluate student performance and generate instructional recommendations. Given the unique characteristics of education, the role of AI should be to augment teachers' judgment rather than replace them. Consequently, current AI systems in educational contexts are primarily positioned as "assistive" technologies that emphasize human-AI collaboration instead of full substitution. AI-assisted decision-making in education primarily adopts two typical paradigms: the concurrent paradigm and the sequential paradigm (Tejeda et al. 2022). In the concurrent paradigm, human decision-makers can access AI-generated predictions or recommendations in real time and incorporate these insights as references within their own decision process to make final decisions (Sayres et al. 2019). In contrast, the sequential paradigm requires human decision-makers to first make independent initial decision, and subsequently receive AI suggestions, based on which they revise or confirm their original decisions (Yin, Wortman Vaughan, and Wallach 2019).

In recent years, a growing body of research has investigated the impact of AI-assisted decision-making. Numerous studies across various domains have demonstrated that AI support can significantly enhance both the efficiency and accuracy of decision-making (Gadde 2020; Li et al. 2021; Lysaght et al. 2019). These studies offer valuable initial insights into the mechanisms of human-AI collaboration. Nevertheless, several limitations remain in the existing literature. First, although prior studies have extensively explored how different decision-making paradigms affect collaboration quality and have identified certain psychological effects (Steyvers and Kumar 2024), there is still a lack of systematic theoretical models explaining the underlying psychological mechanisms and causal pathways. Second, regarding psychological variables, prior research has predominantly centered on users' trust in AI, commonly viewed as a critical psychological mechanism underpinning the effectiveness of AI-assisted decision-making (Vereschak et al. 2024). While trust indeed plays a critical role in human-AI collaboration, this perspective tends to overlook another key construct, namely confidence. Prior studies have shown that users' confidence in their own decision, rather than trust in AI, largely determines whether they follow AI-generated recommendations (Chong et al. 2022). This suggests that confidence may exert a profound influence on final decision performance. Therefore, it is imperative to incorporate confidence into current theoretical frameworks and further investigate its interplay with trust. Finally, the majority of existing empirical studies have concentrated on applications in domains such as healthcare (Goh et al. 2025) and law (Fine, Berthelot, and Marsh 2025), where the impact of AI-assisted decision paradigms on human behavior and cognition has been extensively examined. In contrast, despite the growing attention to AI-assisted technologies in education, empirical research on how such systems influence teachers' psychological states and collaborative performance in instructional decision-making remains relatively scarce. In particular, the perspectives and firsthand feedback from teachers have yet to be thoroughly explored.

To address these research gaps, this study designed an instructional task embedded with an AI decision-support system and recruited 59 pre-service teachers as participants. A two-factor mixed experimental design was employed to systematically examine the effects of AI-assisted decision-making paradigms and human-AI decision consistency on users' psychological states and task performance. The study not only investigated statistical associations among these factors but also focused on uncovering the underlying causal pathways. By integrating Bayesian cumulative link mixed model (CLMM) and the structural equation modeling (SEM), this study not only examines how AI-assisted decision-making paradigms and human-AI decision consistency shape users' psychological states and task performance, but also develops and validates a moderated sequential mediation model. This model specifies that human-AI decision consistency moderates the effect of AI-assisted decision-making paradigms on task performance and highlights how users' decision confidence and their trust in AI jointly serve as the psychological pathway through which these effects unfold. Taken together, this work extends theoretical understandings of human-AI collaboration in educational contexts and offers psychologically grounded insights for the design and optimization of trustworthy instructional AI systems.

Related Work

With the rapid advancement of AI, its empowering potential has been increasingly recognized across a wide range of domains. In the field of education, AI-assisted decision-making has been widely adopted in various subdomains such as intelligent assessment, instructional feedback, and student behavior analysis (Zawacki-Richter et al. 2019). Prior studies have explored its application in collaborative decision-making from diverse perspectives. For example, Dai, Thomas, and Rawolle (2025) points out that AI holds significant potential in educational management decision-making, capable of enhancing decision accuracy, optimizing resource allocation, and improving overall efficiency. Rastogi et al. (2022) examined low-quality decision outcomes under a concurrent AI-assisted paradigm in student performance prediction tasks, attributing the issue primarily to anchoring bias during the decision process. While these studies have yielded valuable insights for the field of AI-assisted decision-making, they have predominantly focused on the direct effects of interaction modes and strategies on user behavior. In contrast, the underlying psychological mechanisms that support these behavioral changes remain underexplored. In particular, there is a lack of systematic investigation into how such psychological responses further influence the overall quality of human-AI collaboration.

As AI technology is increasingly integrated into human decision-making processes, researchers have begun to pay growing attention to the psychological states underlying such interactions. This shift reflects a broader effort to understand human-AI collaboration from a human-centered perspective. For instance, Vereschak et al. (2024) explored how AI practitioners and decision-makers perceive trust in human-AI collaboration. Zhang et al. (2025) investigated students' satisfaction with AI-assisted learning in programming courses and found that students generally exhibited a high level of acceptance toward AI-assisted learning. Abdullah et al. (2025) explored the relationships among cognitive load, decision-making styles, and university instructors' reliance on AI, revealing a significant positive correlation between cognitive load and AI reliance, a positive association between intuitive decision styles and AI reliance, and a negative correlation between rational decision styles and AI reliance. Collectively, existing literature has begun to establish a psychological variable framework for AI-assisted decision-making; However, the causal pathways by which different AI decision-making paradigms affect user behavior through these psychological mechanisms remain insufficiently explored, especially within educational contexts. Moreover, as a critical contextual factor, the role of human-AI consistency has not been systematically studied.

This study focuses on instructional support contexts and systematically compares the effects of two AI-assisted decision-making paradigms, concurrent and sequential, on the collaborative task performance of pre-service teachers. We develop a sequential mediation model with trust and confidence as key mediators and examine the moderating role of human-AI decision consistency within this mechanism. By grounding the investigation in educational tasks, this study aims to elucidate the influence pathways of AI-assisted decision-making paradigms on human-AI collaboration performance, providing psychologically aligned and adaptable theoretical support for the design of educational AI systems.

Method

Participants

The participants in this study were students majoring in education at a normal university. As pre-service teachers, they not only possess solid disciplinary knowledge but also demonstrate strong competence in applying new technologies. Representing the future direction of the teaching profession in the digital age, this group holds particular significance for the present research. A total of 60 participants were recruited and randomly assigned to one of two experimental conditions: concurrent AI-assisted decision-making paradigm (n = 30) or sequential AI-assisted decision-making paradigm (n = 30). Among the 60 participants who formally took part in the experiment, 59 completed all procedures and questionnaires. One participant was excluded due to incomplete responses. As a result, the final valid sample consisted of 29 participants in the concurrent condition and 30 in the sequential condition. Both groups interacted with AI systems that had identical decision-making accuracy; the only manipulated variable was the type of AI-assisted decision-making paradigm, concurrent vs. sequential. This study received ethical approval from the university's institutional review board for research involving human subjects. All participants received compensation upon completion of the experiment.

Procedure

This study selected the student academic performance prediction task as the experimental decision-making scenario, which represents a common task type in research on AI-assisted decision-making (Rastogi et al. 2022; Alamri et al. 2020). In this task, participants were required to predict students' final academic performance based on a set of features, including demographic information, learning behaviors and learning performance information. The dataset used for the task was derived from the Student Performance Dataset available in the University of California, Irvine (UCI) Machine Learning Repository (Cortez and Silva 2008). This dataset contains samples from students, each comprising 33 attributes (e.g., age, study time), along with actual final grades, which served as the benchmark for evaluating participants' prediction accuracy. To ensure that the task maintained an appropriate level of complexity and effectively supported pre-service teachers in constructing meaningful decision models, we conducted feature selection based on the correlation between individual attributes and prediction performance. From the original 33 features, eight key variables were selected for their strong relevance to students' final academic performance. These features encompass demographic factors (e.g., parental education levels: Fedu, Medu), learning behaviors (e.g., aspiration for higher education: higher, weekly study time: studytime, number of absences: absences), and prior academic performance (e.g., number of past class failures: failures, first-term grade: G1, second-term grade: G2). These selected features are suitable for constructing human prediction models, and similar approaches have been validated in previous studies as effective for modeling human decision-making (Bansal et al. 2019; Ma et al. 2023).

In the experiment, participants were divided into two groups based on the AI-assisted decision-making paradigm. In the prediction task, participants in the concurrent paradigm group were presented with the AI-generated outcome before making their own decisions, whereas those in the sequential paradigm group first made an initial decision and were then shown the AI's prediction, allowing them to review and revise their original decision accordingly. Participants experienced one of the AI-assisted decision-making paradigms under randomly assigned conditions. We implemented a random forest model to support AI-generated predictions (Huynh-Cam, Chen, and Le 2021), and built a customized intelligent decision-support platform, embedding the instructional task to enable controlled human-AI interaction. Participants proceeded through three phases during the study: (1) Introduction to the experimental procedure: We provided the participants with an overview of the task process and conducted a survey on demographic information. (2) Task instruction: A concise tutorial was given to explain the attributes and meaning of the student information provided during the prediction task, as well as the overall prediction procedure. To ensure attentiveness, two attention check questions were included at the end of the tutorial; data from participants who failed these checks were excluded from further analysis. Two sample trials were then presented to help participants become familiar with the task. (3) Main task: Participants completed 25 prediction trials with AI assistance. After each trial, they rated their level of trust in the AI prediction and their confidence in their final decision. During this phase, no feedback or ground truth information was provided.

Measures

Trust refers to the extent to which participants rely on the AI's suggestions or decision during interaction, while confidence denotes participants' belief in the accuracy of their own decisions(Mayer, Davis, and Schoorman 1995; Chong et al. 2022). The measurements of trust and confidence were adapted from the single-item momentary assessment approach proposed by (Castro et al. 2023). Following each task round, participants evaluated their trust in the AI's recommendation and their confidence in the accuracy of their final decision, based on their direct interaction experience with the AI system. All items were rated on a 7-point Likert scale (1 = "Strongly Disagree", 7 = "Strongly Agree").

This study adopted decision error as the primary measure of task performance, operationalized as the absolute error between a participant's prediction and the true value. As a non-directional measure of bias, absolute error effectively captures the degree of deviation between predicted values and actual outcomes. This metric provides a critical basis for assessing the impact of AI-assisted decision-making under different experimental conditions on participants' behavioral performance. The use of absolute error has been widely adopted in prior research (Ahn et al. 2024), where it demonstrated strong reliability and validity, thus offering theoretical support for its selection in the present study.

Data Analysis

This study employed a two-step analytical strategy to enhance the robustness and interpretability of the results. First, CLMM (DeYoreo and Kottas 2018) were used in R (version 4.5.1) to examine the main and interaction effects of collaboration mode, human-AI decision consistency, and their effects on task performance, trust, and confidence. Random intercepts for participants and task rounds were included to account for repeated measurements and the hierarchical structure of the data. Trust and confidence were measured using 7-point Likert scales and treated as ordinal variables. Due to the ordinal nature of Likert scales, CLMMs are appropriate for modeling the relationships between these ratings. Task performance was quantified as the absolute error, which, after excluding outliers, ranged from 0 to 4 as discrete integers. Given the small, bounded, and ordered distribution of these data, CLMM was also deemed appropriate for modeling task performance. Second, to examine the relationships among AI-assisted decision-making paradigms, psychological variables, and user performance, SEM (Hair 2014) was constructed using SmartPLS 4.0. Confidence and trust were modeled as mediators linking collaboration mode to task performance, while human-AI decision consistency was specified as a moderator on the path from collaboration mode to confidence. Bootstrapping with 5000 resamples was employed to assess the significance of direct, indirect, and moderated effects.

Result

The results of the CLMM and SEM analyses are presented in this section. Descriptive statistics provide an overview of task performance, trust, and confidence across the two AI-assisted decision-making paradigms and the human-AI decision consistency conditions, as summarized in Table 1. Overall, task errors were lower under consistent decision conditions than under inconsistent ones, regardless of the type of AI-assisted decision-making paradigm. Trust ratings were generally higher when participants' decisions were consistent with the AI, with the highest trust observed in the concurrent-consistency condition. Confidence showed a similar pattern, although differences across conditions were less pronounced.

Results of CLMM

The results of the CLMM examined the effects of AI-assisted decision-making paradigms and Human-AI decision consistency on user task performance, trust level, and confidence. The results are presented in Table 2

Regarding task performance, a significant main effect of human-AI decision consistency was observed on decision error (Estimate = 0.26, 95% CI [0.09, 0.42]), with the 95% confidence interval excluding zero, indicating that participants exhibited lower decision error under consistent conditions. The main effect of AI-assisted decision paradigms were not significant (Estimate = -0.04, 95% CI [-0.20, 0.13]), suggesting no overall difference in task performance between the sequential and concurrent paradigms. Moreover, the interaction between AI-assisted decision paradigms and human-AI decision consistency did not reach statistical significance (Estimate = -0.20, 95% CI [-0.43, 0.03]), suggesting that the data provide no evidence for an interaction between the two variables.

With respect to trust, human-AI decision consistency exhibited a significant main effect (Estimate = -2.40, 95% CI [-2.60, -2.19]), indicating that participants' trust substantially decreased when their decisions were inconsistent with those of the AI. In contrast, the main effect of the AI-assisted decision-making paradigm (Estimate = -0.38, 95% CI = [-0.89, 0.11]) and the interaction effect (Estimate = 0.10, 95% CI [-0.14, 0.34]) were not significant, suggesting that neither the human-AI decision-making paradigm nor its interaction with human-AI decision consistency had a significant effect on participants' trust.

Regarding participants' confidence, the CLMM analysis indicated significant main effects of both the AI-assisted decision-making paradigm (Estimate = -0.77, 95% CI [-1.46, -0.12]) and human-AI decision consistency (Estimate = -0.33, 95% CI [-0.51, -0.14]). Specifically, participants in the sequential collaboration paradigm reported lower confidence compared to those in the concurrent paradigm. Similarly, participants' confidence decreased significantly when their decisions were inconsistent with AI recommendations. The interaction between AI-assisted decision paradigms and human-AI decision consistency did not reach statistical significance (Estimate = -0.24, 95% CI [-0.49, 0.01]), suggesting that the two factors did not interact in a statistically meaningful way.

Result of SEM

To further investigate the underlying psychological mechanisms linking AI-assisted decision-making paradigms to task performance, which may not be fully captured by CLMM, we hypothesized that confidence and trust serve as sequential mediators between the decision-making paradigm and task performance, with human-AI decision consistency acting as a moderator. To test this hypothesis, we employed SEM, which allowed us to systematically examine the sequential mediating effects of confidence and trust as well as the moderating role of decision consistency. This approach provided a complementary perspective to the direct effects identified by CLMM.

Figure 1 presents the proposed structural model, while Table 3 provides a detailed summary of the direct effects in the structural equation model. The results indicate that the AI-assisted decision-making paradigm did not have a significant direct effect on task performance, as measured by mean absolute error ($\beta = -0.239$, p = 0.297). However, participants in the sequential mode reported significantly lower confidence than those in the parallel mode ($\beta = -0.613$, p = 0.015). In addition, confidence positively and significantly predicted decision error ($\beta = 0.537$, $p < 0.001$), indicating that higher confidence was associated with greater decision errors and poorer task performance. The sequential

AI-DMP	HAI-DC	Decision Error		Trust		Confidence	
		Mean	SD	Mean	SD	Mean	SD
Concurrent	Consistency	0.80	0.67	6.04	0.814	5.93	0.798
	Inconsistency	0.99	0.864	4.41	1.303	5.94	0.765
Sequential	Consistency	0.80	0.711	5.89	0.910	5.68	0.990
	Inconsistency	0.82	0.818	4.12	0.997	5.26	0.894

Table 1: Descriptive statistics of task performance, trust, and confidence. Note: AI-DMP = AI-assisted decision-making paradigm; HAI-DC = human-AI decision consistency.

Variable	Effect	Estimate	Est.Error	*Z*	95% CI	
					Lower	Upper
Decision error	AI-DMP (Sequential)	-0.04	0.09	-0.42	-0.20	0.13
	HAI-DC (Inconsistency)	0.26	0.09	3.02	0.09	0.42
	Interaction	-0.20	0.12	-1.73	-0.43	0.03
Trust	AI-DMP (Sequential)	-0.38	0.25	-1.53	-0.89	0.11
	HAI-DC (Inconsistency)	-2.40	0.10	-23.15	-2.6	-2.19
	Interaction	0.10	0.12	0.85	-0.14	0.34
Confidence	AI-DMP (Sequential)	-0.77	0.34	-2.25	-1.46	-0.12
	HAI-DC (Inconsistency)	-0.33	0.10	-3.43	-0.51	-0.14
	Interaction	-0.24	0.13	-1.89	-0.49	0.01

Table 2: Summary of Fixed Effects. Note: AI-DMP = AI-assisted decision-making paradigm; HAI-DC = human-AI decision consistency.

mode significantly increased users' trust in AI ($\beta = 0.399$, $p = 0.049$), and this trust was found to significantly and negatively predict decision errors ($\beta = -0.575$, $p < 0.001$). Additionally, confidence had a significant positive effect on trust ($\beta = 0.682, p < 0.001$), while the consistency between human and AI decisions significantly moderated the effect of the AI-assisted decision-making paradigm on confidence ($\beta = 0.789, p < 0.001$).

Table 4 presents the moderated serial mediation effect. The results indicate that the indirect effect of the AI-assisted decision-making paradigm on task performance through trust is marginally significant ($\beta = -0.233$, $p = 0.086$). although p is slightly above 0.05, it is close to the significance threshold, indicating a potential trend. Similarly, the indirect path through confidence alone is also marginally significant ($\beta = -0.337$, $p = 0.062$), despite p being slightly above 0.05, this effect remains close to significant, highlighting the importance of confidence in the AI-assisted decision-making paradigm. The serial mediation path involving both confidence and trust is also marginally significant ($\beta = 0.249$, $p = 0.081$). This result suggests that confidence and trust may jointly form a meaningful psychological transmission mechanism, linking the AI-assisted decision-making paradigm to task performance. Although p is slightly above 0.05, this still reflects the potential importance of this path. Notably, within the moderated serial mediation model, human-AI decision consistency significantly moderated the first stage of the serial mediation pathway ($\beta = -0.313$, $p = 0.022$). Overall, these findings reveal a chain-based moderated mediation mechanism: under the moderating effect of human-AI decision consistency, the AI-assisted decision-making paradigm influences task performance through the pathways of confidence and trust.

Table 5 presents the conditional direct and indirect effects across levels of human-AI decision consistency. The results indicate that human-AI decision consistency significantly moderated the effect of AI-assisted decision-making paradigm on user confidence. At low consistency, the sequential paradigm substantially reduced confidence compared to the concurrent paradigm ($\beta = -1.405, p < 0.001$). At average levels of decision consistency, this negative effect was weakened, yet it remained statistically significant ($\beta = -0.613$, $p = 0.015$). At high consistency, the effect was non-significant ($\beta = 0.178$, $p = 0.602$), indicating that higher decision alignment buffers the confidence-reducing impact of the sequential paradigm. Moreover, the sequential mediation path from AI-assisted decision-making paradigm to task performance via confidence and trust was also moderated by human-AI decision consistency. At low consistency, the indirect effect was significant and positive ($\beta = 0.577$, $p = 0.021$), compared to the concurrent paradigm, suggesting that in situations with lower consistency, the sequential paradigm more significantly reduces users' confidence, which in turn decreases trust and increases decision errors. At mean consistency, the effect was smaller and marginally significant ($\beta = 0.252$, $p = 0.081$), while at high consistency, the effect was non-significant ($\beta = -0.073$, $p = 0.602$), indicating that the indirect effect diminishes as decision consistency increases.

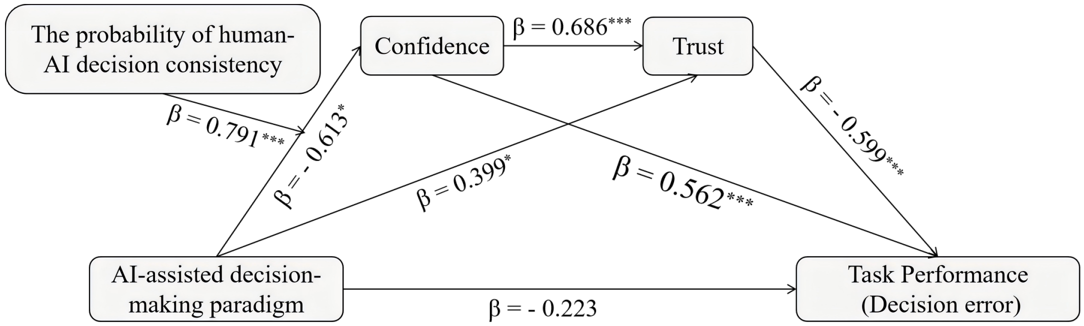

Figure 1: PLS-SEM model diagram

Path	O	M	Std. Deviation	95% CI		p
				Lower	Upper	
AI-DMP → Decision error	-0.223	-0.239	0.213	-0.641	0.185	0.297
AI-DMP → Confidence	-0.613	-0.613	0.253	-1.098	-0.099	0.015
Confidence → Decision error	0.562	0.537	0.158	0.201	0.831	$<$ 0.001
AI-DMP → Trust	0.399	0.399	0.203	-0.014	0.790	0.049
Trust → Decision error	-0.599	-0.575	0.157	-0.845	-0.214	$<$ 0.001
Confidence → Trust	0.686	0.682	0.095	0.475	0.845	$<$ 0.001
HAI-DC → Confidence	-0.440	-0.444	0.143	-0.741	-0.157	0.002
HAI-DC × AI-DMP → Confidence	0.791	0.789	0.225	0.318	1.201	$<$ 0.001

Table 3: Direct effects of the structural equation model. Note: AI-DMP = AI-assisted decision-making paradigms; HAI-DC = human-AI decision consistency

Disscusion

The Effect of AI-Assisted Decision-Making Paradigms on Task Performance: A Chain Psychological Mechanism under the Moderation of Human-AI Consistency

This study found that, in human-AI collaboration, AI-assisted decision-making paradigms did not exert a significant direct effect on task performance, which is consistent with previous research Buçinca, Malaya, and Gajos (2021); Tejeda et al. (2022). However, further analyses revealed that human-AI consistency significantly moderated the effect of AI decision-making paradigms on task performance, and this effect operated through a chain psychological mechanism involving users' confidence and trust in AI. This indicates that the influence of AI decision-making paradigms on task performance is neither entirely absent nor universally present, but rather depends on the level of human-AI consistency and is mediated by psychological factors. Specifically, when human-AI consistency is low, the sequential paradigm is more likely to undermine users' confidence, which in turn reduces trust in AI and ultimately impairs task performance; under high human-AI consistency conditions, this negative chain effect is substantially weakened or may even disappear. Overall, this study highlights the critical role of human-AI consistency and users' psychological states in designing effective human-AI collaboration. In particular, within the context of educational decision-making, the effectiveness of AI-assisted decision-making paradigms may depend on whether users' psychological responses are successfully activated. Therefore, future AI system design should focus on optimizing interaction experiences to effectively guide users' psychological states, thereby ultimately improving decision quality.

The Role of Human-AI Decision Consistency in Shaping Confidence, Trust, and Task Performance

This study demonstrates that the consistency between human and AI decision-making exerts a significant main effect on all three key indicators: task performance, self-confidence, and trust. Compared with high consistency, low consistency leads to increased decision errors, reduced trust, and diminished self-confidence. This indicates that decision consistency not only reflects cognitive alignment but also enhances psychological affirmation during human-AI interactions. Such consistency does not occur by chance; rather, it functions as a critical interactive resource. Specifically, in sequential paradigms, low human-AI consistency is more likely to undermine users' self-confidence, which in turn negatively affects task performance through its impact on trust. Conversely, high consistency mitigates these adverse effects. These findings suggest that, in professional domains such as education, AI systems should be matched to users' task execution capabilities. Excessive discrepancies between human and AI abilities can lead to high inconsistency, which not only erodes users' confidence and trust in AI but also impairs collaboration quality, with these effects being particu-

Path	O	M	Std. Deviation	95% CI		p
				Lower	Upper	
AI-DMP → Trust → DE	-0.239	-0.233	0.139	-0.559	-0.011	0.086
AI-DMP → Confidence → DE	-0.345	-0.337	0.184	-0.822	-0.059	0.062
AI-DMP → Confidence → Trust → DE	0.252	0.249	0.145	0.046	0.653	0.081
HAI-DC × AI-DMP → Confidence → Trust → DE	-0.325	-0.313	0.141	-0.697	-0.107	0.022

Table 4: Moderated serial mediation effects. Note: AI-DMP = AI-assisted decision-making paradigms; HAI-DC = human-AI decision consistency; DE = Decision error

Path	O	M	Std. Deviation	95% CI		p
				Lower	Upper	
AI-DMP → Confident (HAI-DC at +1 SD)	0.178	0.176	0.341	-0.517	0.840	0.602
AI-DMP → Confident (HAI-DC at -1 SD)	-1.405	-1.403	0.336	-2.037	-0.707	< 0.001
AI-DMP → Confident (HAI-DC at Mean)	-0.613	-0.613	0.253	-1.098	-0.099	0.015
AI-DMP → Confident → Trust → DE (HAI-DC at +1 SD)	-0.073	-0.064	0.140	-0.340	0.226	0.602
AI-DMP → Confident → Trust → DE (HAI-DC at -1 SD)	0.577	0.562	0.250	0.143	1.106	0.021
AI-DMP → Confident → Trust → DE (HAI-DC at Mean)	0.252	0.249	0.145	0.018	0.579	0.081

Table 5: Conditional indirect and direct effects across levels of HAI-DC. Note: AI-DMP = AI-assisted decision-making paradigms; HAI-DC = human-AI decision consistency; DE = Decision error

larly pronounced in sequential paradigms. Furthermore, the design of human-AI collaboration systems should consider mechanisms to enhance users' perceived consistency. For example, when AI outputs frequently conflict with users' judgments, providing timely and interpretable feedback can buffer the negative psychological effects caused by sequential information presentation, thereby maintaining user confidence and supporting effective collaboration.

Notably, although disagreement between human and AI decisions significantly reduced trust in the AI system, the average trust score under inconsistency conditions remained above the neutral midpoint (greater than 4 on a 7-point scale). This indicates that users did not completely lose trust in the AI system when faced with conflicting advice. Instead, this phenomenon highlights the resilience of trust, a decline in trust does not equate to its collapse. This finding supports the theory of dispositional trust (Marsh and Dibben 2003), which suggests that individuals hold a general tendency to trust automation, independent of specific contexts or systems. Dispositional trust is considered a relatively stable trait over time (Hoff and Bashir 2015). Therefore, while trust is susceptible to situational challenges, it is not inherently fragile or easily destroyed (Stanley and Dorton 2023). In fact, moderate conflict in AI recommendations may serve as an opportunity for users to re-evaluate the AI's expertise, potentially triggering deeper cognitive engagement and renewed reliance (Shin 2021).

Conclusion

This study investigates how AI-assisted decision-making paradigms and human-AI consistency influence the psychological states and task performance of pre-service teachers. It highlights the critical role of human-AI Decision Consistency and psychological factors in the design of educational AI systems, suggesting that future AI optimization should not focus solely on technical improvements, but also on the positive shaping and guidance of users' psychological states. Moreover, the study deepens theoretical understanding of human-AI collaborative decision-making in educational contexts, providing theoretical support and design implications for developing more adaptive and human-centered AI systems in teaching and learning.

Ethical Statement

This study has been approved by the Human Subject Protection Committee of East China Normal University (HR4-0032-2024). All experiments were conducted with voluntary participants. Participants' data were anonymized and handled with strict confidentiality to ensure privacy. Potential risks were minimized, and no harm was anticipated to participants.

Acknowledgments

We would like to thank all participants and colleagues for their helpful discussions and valuable feedback during this study. We also thank the reviewers and program chairs for their constructive suggestions. This work was supported by the National Natural Science Foundation of China under Grant No. 62477011, and the Special Project on Artificial Intelligence-Driven Research Paradigm Reform and Discipline Leapfrog Development Empowerment by Shanghai Municipal Education Commission, Institute of Language Sciences, Shanghai International Studies University.

References

Abdullah, Q.; Rehman, H.; Shehzad, M.; et al. 2025. Relationship between Cognitive Load and AI Dependence among University Teachers: Moderating Role of Decision Making Styles. *Research Journal of Psychology*, 3(3): 26–35.

Ahn, D.; Almaatouq, A.; Gulabani, M.; et al. 2024. Impact of Model Interpretability and Outcome Feedback on Trust in AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–25. ACM.

Alamri, H. L.; Almuslim, S. R.; Alotibi, M. S.; et al. 2020. Predicting Student Academic Performance Using Support Vector Machine and Random Forest. In *Proceedings of the 2020 3rd International Conference on Education Technology Management*, 100–107. IEEE.

Bansal, G.; Nushi, B.; Kamar, E.; Lasecki, W. S.; Weld, D. S.; and Horvitz, E. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, 2–11. AAAI Press.

Buçinca, Z.; Malaya, M. B.; and Gajos, K. Z. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1): 1–21.

Castro, M. S.; Bahli, B.; Ferreira, J. J.; et al. 2023. Comparing Single-Item and Multi-Item Trust Scales: Insights for Assessing Trust in Project Leaders. *Behavioral Sciences*, 13(9): 786–800.

Chong, L.; Zhang, G.; Goucher-Lambert, K.; et al. 2022. Human Confidence in Artificial Intelligence and in Themselves: The Evolution and Impact of Confidence on Adoption of AI Advice. *Computers in Human Behavior*, 127: 107018.

Cortez, P.; and Silva, A. M. G. 2008. Using Data Mining to Predict Secondary School Student Performance. *Journal of Educational Data Mining*, 5(1): 22–48.

Dai, R.; Thomas, M. K. E.; and Rawolle, S. 2025. The Roles of AI and Educational Leaders in AI-Assisted Administrative Decision-Making: A Proposed Framework for Symbiotic Collaboration. *The Australian Educational Researcher*, 52(2): 1471–1487.

DeYoreo, M.; and Kottas, A. 2018. Bayesian nonparametric modeling for multivariate ordinal regression. *Journal of Computational and Graphical Statistics*, 27(1): 71–84.

Fine, A.; Berthelot, E. R.; and Marsh, S. 2025. Public Perceptions of Judges' Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice. *Behavioral Sciences*, 15(4): 476–496.

Gadde, H. 2020. AI-Assisted Decision-Making in Database Normalization and Optimization. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, 11(1): 230–259.

Goh, E.; Bunning, B.; Khoong, E. C.; Veettil, S.; Lee, A.; Patel, S. Y.; Agha, Z.; Ong, C.; Lin, S.; and Gomez, A. 2025. Physician Clinical Decision Modification and Bias Assessment in a Randomized Controlled Trial of AI Assistance. *Communications Medicine*, 5(1): 59–63.

Hair, J. F. 2014. *A primer on partial least squares structural equation modeling (PLS-SEM)*. SAGE Publications.

Hoff, K. A.; and Bashir, M. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors*, 57(3): 407–434.

Huynh-Cam, T.-T.; Chen, L.-S.; and Le, H. 2021. Using Decision Trees and Random Forest Algorithms to Predict and Determine Factors Contributing to First-Year University Students' Learning Performance. *Algorithms*, 14(11): 318–334.

Li, J.; Wu, J.; Zhao, Z.; Zhang, L.; Li, W.; Yang, J.; Liu, Y.; Wang, X.; Zhang, X.; and Huang, M. 2021. Artificial Intelligence-Assisted Decision Making for Prognosis and Drug Efficacy Prediction in Lung Cancer Patients: A Narrative Review. *Journal of Thoracic Disease*, 13(12): 7021–7032.

Lysaght, T.; Lim, H.; Xafis, V.; and Kerridge, I. 2019. AI-Assisted Decision-Making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research. *Asian Bioethics Review*, 11(3): 299–314.

Ma, S.; Lei, Y.; Wang, X.; et al. 2023. Who Should I Trust: AI or Myself? Leveraging Human and AI Correctness Likelihood to Promote Appropriate Trust in AI-Assisted Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19. ACM.

Marsh, S.; and Dibben, M. R. 2003. The Role of Trust in Information Science and Technology. *Annual Review of Information Science and Technology*, 37: 465–498.

Mayer, R. C.; Davis, J. H.; and Schoorman, F. D. 1995. An Integrative Model of Organizational Trust. *Academy of Management Review*, 20: 709–734.

Rastogi, C.; Zhang, Y.; Wei, D.; et al. 2022. Deciding Fast and Slow: The Role of Cognitive Biases in AI-Assisted Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1): 1–22.

Sayres, R.; Taly, A.; Rahimy, E.; Blumer, K.; Cozad, R. A.; Hammel, N.; Hassanpour, S.; House, H.; Lim, G.; Mahdavi, S.; Raymond, A.; Shieh, W.; Smith, J. R.; Wu, D. J.; Xu, S.; Butt, M.; Chang, P.; Cuadros, J.; Krause, J.; Peng, L.; and Webster, D. R. 2019. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. *Ophthalmology*, 126(4): 552–564.

Schaffer, H. E.; Young, K. R.; Ligon, E. W.; and Chapman, D. D. 2017. Automating individualized formative feedback in large classes based on a directed concept graph. *Frontiers in Psychology*, 8: 260.

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.

Stanley, J. C.; and Dorton, S. L. 2023. Exploring Trust with the AI Incident Database. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 67, 489–494. Los Angeles, CA: SAGE Publications.

Steyvers, M.; and Kumar, A. 2024. Three challenges for AI-assisted decision-making. *Perspectives on Psychological Science*, 19(5): 722–734.

Tejeda, H.; Kumar, A.; Smyth, P.; and Laird, J. E. 2022. AI-Assisted Decision-Making: A Cognitive Modeling Approach to Infer Latent Reliance Strategies. *Computational Brain & Behavior*, 5(4): 491–508.

Vereschak, O.; Alizadeh, F.; Bailly, G.; Goyal, N.; Feustel, C.; and Buschek, D. 2024. Trust in AI-Assisted Decision Making: Perspectives from Those Behind the System and Those for Whom the Decision is Made. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–14. ACM.

Yin, M.; Wortman Vaughan, J.; and Wallach, H. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–12. New York, NY, USA: ACM.

Zawacki-Richter, O.; Marín, V. I.; Bond, M.; and Gouverneur, F. 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1): 1–27.

Zhang, W.; Xiong, Y.; Zhou, D.; et al. 2025. Balancing Human and AI Instruction: Insights from Secondary Student Satisfaction with AI-Assisted Learning. *Interactive Learning Environments*, 1–16.