EJSIT

European Journal of Science, Innovation and Technology

ISSN: 2786-4936

www.ejsit-journal.com

Volume 5 | Number 1 | 2025

The Impact of AI Explainability on Cognitive Dissonance and Trust in Human-AI Recruitment Teams

Tetiana Sydorenko
Master of Arts in User Experience Design
Falmouth University
Penryn Campus, Penryn TR10 9FE, United Kingdom

ABSTRACT

This study examined how varying levels of AI explainability affect trust and cognitive dissonance within human-AI teams, focusing on collaborative recruitment scenarios. The results indicate a complex interplay: while no significant differences were observed between the high-explainability and no-explainability conditions, low explainability led to a statistically significant increase in cognitive dissonance. This finding suggests that partial explanations may exacerbate rather than mitigate uncertainty, possibly due to confirmation bias, where users selectively interpret incomplete information to align with pre-existing beliefs. In such cases, explanations may provide enough detail to provoke skepticism but insufficient justification to resolve doubts, leaving users conflicted between their intuition and the AI's suggestions. These results highlight the need to prioritize explanation quality and completeness over sheer detail when designing XAI systems for collaboration.

Keywords: Explainable AI (XAI), human-AI collaboration, trust in AI, cognitive dissonance, AI-assisted recruitment, algorithmic transparency, organizational behavior, human-computer interaction (HCI)

INTRODUCTION

The increasing integration of artificial intelligence into workplace teams is transforming the nature of collaboration. As AI takes on more complex roles, effective human-AI teamwork becomes critical for organizational success. However, a key challenge in these collaborations is establishing trust in AI-driven decision-making, particularly when users have a limited understanding of the system's inner workings (Glikson & Woolley, 2020), as a deficiency of trust can hinder AI adoption and effective utilization, leading to suboptimal outcome (Gillespie et al., 2023).

Research indicates that Explainable AI (XAI) enhances trust in human-AI collaboration by improving transparency and interpretability (Sharma et al., 2023; Giovine & Roberts, 2024). By providing insights into the AI's reasoning process, XAI seeks to foster a well-calibrated trust relationship between humans and AI (Arrieta et al., 2020). However, the relationship between explainability and trust in human-AI teams remains underexplored, especially in complex, high-stakes decision-making scenarios like recruitment. While increased transparency is often assumed to enhance trust, the effectiveness of explanations is complex and can be influenced by factors such as user biases (Zhang et al., 2020; Bashkirova & Krpan, 2024). Moreover, when AI recommendations contradict human judgment, individuals may experience cognitive dissonance, leading them to align with or reject the AI’s guidance (Festinger, 1962; Sivaraman et al., 2023).

This study investigates the complex interplay between AI explainability, trust, and cognitive dissonance in the context of recruitment. Specifically, we examine how varying levels of AI explainability influence trust and cognitive dissonance among HR professionals and hiring managers interacting with an AI recruitment assistant. Using a simulated recruitment scenario, we address the following research questions:

RQ1: How do different levels of AI explainability impact cognitive dissonance and trust in human-AI recruitment teams?

RQ2: To what extent does AI explainability predict trust in human-AI recruitment teams?

MATERIALS AND METHODS

Research Design

The study employed a quantitative, experimental design to investigate the impact of AI explainability on cognitive dissonance and trust in human-AI recruitment teams. Experimental designs are well-suited for examining cause-and-effect relationships by manipulating independent variables (in this case, levels of explainability) and measuring their effects on dependent variables (cognitive dissonance and trust) (Langer et al., 2023). A between-subjects design was used, with participants randomly assigned to one of three explainability conditions: no explainability (E0), low explainability (E1), or high explainability (E2). This approach enables direct group comparisons, allowing for an assessment of how different levels of explainability influence the targeted psychological constructs.

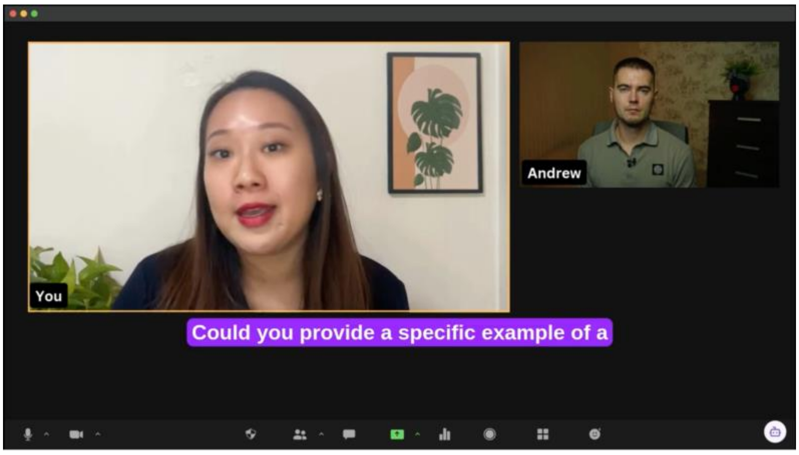
Respondents of the Study

Seventy-two participants (Age: $\bar{x}$ = 34.89; Sex: 67.4% Female, 32.6% Male) were recruited via LinkedIn and completed the online study through Google Forms after providing informed consent. Eligibility criteria required participants to be at least 18 years old and working in Human Resources with recruitment responsibilities. A targeted LinkedIn recruitment strategy ensured compliance with pre-defined selection criteria and access to professionals with relevant experience. No financial compensation was offered. Participants who did not complete the full study were excluded from the analysis. The sample size of 72 was primarily determined by feasibility constraints.

Procedure

Participants were randomly assigned to one of three AI explainability conditions: no explainability (E0), low explainability (E1), or high explainability (E2). In the no-explainability condition (E0), the AI system provided only numerical ratings. The low-explainability condition (E1) included a brief summary of key factors influencing the AI’s rating. The high-explainability condition (E2) offered a detailed justification with specific evidence and reasoning.

Participants viewed pre-recorded video interviews in which an applicant responded to three behavioral questions on problem-solving, communication, and adaptability. For each question, they observed a recruiter asking a question (~2 minutes) and the applicant replying (~3 minutes), with the full video lasting approximately 15 minutes. These interviews, featuring professional actors who consented to participate, were presented with subtitles and embedded within a simulated meeting software to create a realistic and accessible virtual interview environment (Figure 1).



Could you provide a specific example of a

Figure 1: Screenshot of the Interview Video Used in the Experiment

This experimental setup was designed to elicit cognitive dissonance by contrasting the applicant's ambiguous non-verbal behaviors with the AI assistant’s objective interpretation. Unlike human recruiters, who may rely on intuition and subjective impressions, the AI assistant evaluates candidates using a fixed set of criteria. Human recruiters can be influenced by cognitive biases, such as the halo effect or confirmation bias, whereas the AI applies the same criteria uniformly, ensuring a more consistent evaluation based on predefined factors.

To enhance ambiguity, we intentionally incorporated non-verbal cues in the interview videos—such as long pauses, firm body posture, and limited eye contact—which are often misinterpreted as signs of dishonesty or uncertainty, leading individuals to make biased judgments based on these behaviors, as suggested by existing research (Denault, 2020). For example, decreased eye contact is commonly perceived as a sign of deception or lack of confidence, even when it may stem from cultural differences or personality traits (Denault, 2020). Similarly, rigid posture—potentially a sign of nervousness or discomfort—might instead be misread as defensiveness or hostility (Zhang, 2021).

By contrast, the AI system, unaffected by cognitive biases, focuses solely on the content of the candidate’s responses rather than non-verbal cues. This fundamental difference in decision-making could lead the AI to assign high ratings in cases where a human recruiter remains uncertain, potentially triggering cognitive dissonance.

After each applicant answered the three questions, participants viewed the AI assistant’s numerical performance rating before completing an online survey. Cognitive dissonance was measured using a modified three-item version of the Post-Purchase Cognitive Dissonance Scale (Sweeney et al., 2000), while trust in AI was assessed with a multidimensional trust scale adapted from Jian et al. (2000).

Following the experiment, participants received a debriefing document addressing potential concerns about cognitive dissonance and ensuring access to clarification or support from the researcher if needed (Figure 2).

The AI system used in this study was designed to simulate a real-world AI recruitment assistant that evaluates candidates based on structured, predefined criteria (Smelyakov et al., 2023). Although it did not algorithmically generate responses, the system was designed to mimic a decision tree model that applies weighted rules to candidate answers. Using a Wizard of Oz approach, the evaluations and explanations were manually predetermined to appear as AI-generated outputs. Similar to commercial AI-driven hiring tools, this rule-based system assessed candidates across multiple factors—particularly the relevance and completeness of their answers, as well as clarity of communication. Each criterion contributed to the AI’s overall rating, with content quality receiving the highest weight. This structured, interpretable model was chosen to simulate how an AI recruitment assistant might systematically assess responses while ensuring experimental control and explanation consistency.

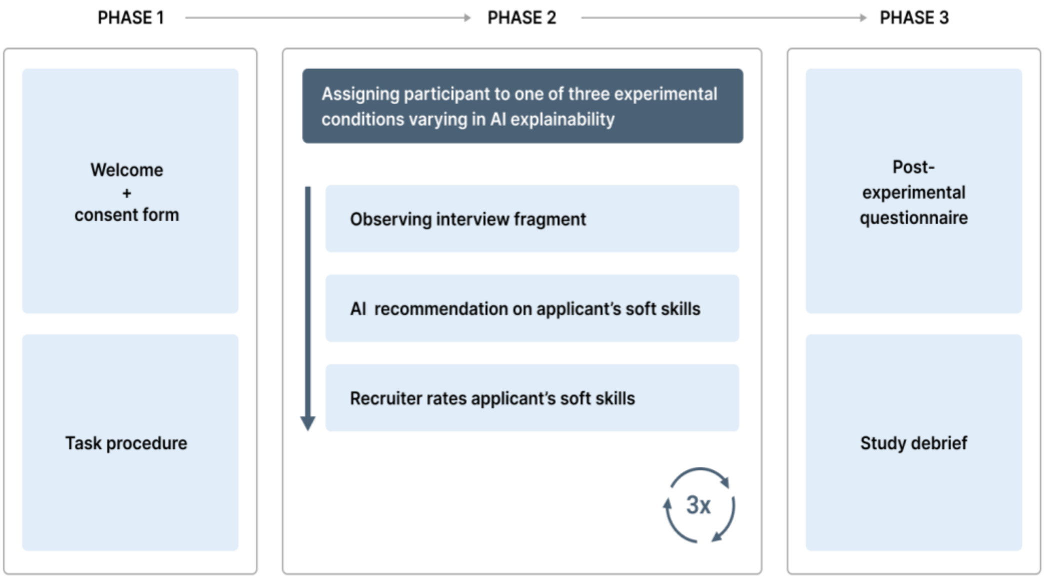

PHASE 1
PHASE 2
PHASE 3
Welcome
+
consent form
Task procedure
Assigning participant to one of three experimental conditions varying in AI explainability
Observing interview fragment
AI recommendation on applicant's soft skills
Recruiter rates applicant's soft skills
3x
Post-experimental questionnaire
Study debrief

Figure 2: Experimental Study Procedure (with The Main Study Task (Phase 2) Being Repeated 3 Times)

RESEARCH INSTRUMENT

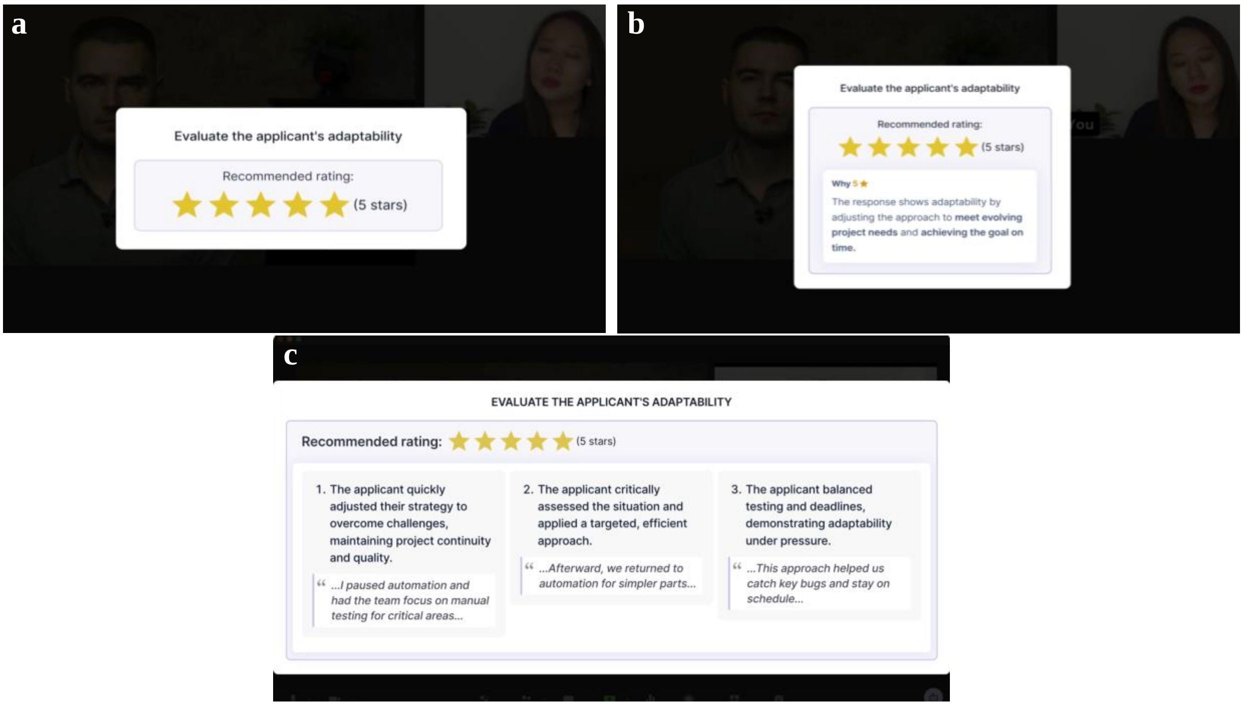
AI Explainability Manipulation

This study manipulated AI explainability by varying the level of detail in explanations about the AI’s decision-making process during a simulated recruitment task. Participants were randomly assigned to one of three conditions: no explainability, low explainability, or high explainability.

No Explainability (E0): Participants received only the AI's candidate ranking without an explanation of how it was generated. This condition represents a "black box" scenario in which the AI's internal workings remain opaque to the user (Figure 3, a).

Low Explainability (E1): Participants received the AI's candidate ranking along with a brief, one-sentence justification for each ranking. These justifications summarized the applicant’s response to the question, providing a basic understanding of the AI’s reasoning (Figure 3, b).

High Explainability (E2): Participants received the AI's candidate ranking alongside a detailed explanation for each ranking. The system provided a natural language justification, incorporating a relevant quote from the interview transcript and explaining how it supported the assigned rating. This level of explainability aimed to maximize transparency in the AI's decision-making process (Figure 3, c).



Evaluate the applicant's adaptability

Recommended rating:

★★★★★ (5 stars)

Evaluate the applicant's adaptability

Recommended rating:

★★★★★ (5 stars)

Why 5 ★

The response shows adaptability by adjusting the approach to **meet evolving project needs** and **achieving the goal on time.**

EVALUATE THE APPLICANT'S ADAPTABILITY

Recommended rating: ★★★★★ (5 stars)

1. The applicant quickly adjusted their strategy to overcome challenges, maintaining project continuity and quality.

   “ *...I paused automation and had the team focus on manual testing for critical areas...*

2. The applicant critically assessed the situation and applied a targeted, efficient approach.

   “ *...Afterward, we returned to automation for simpler parts...*

3. The applicant balanced testing and deadlines, demonstrating adaptability under pressure.

   “ *...This approach helped us catch key bugs and stay on schedule...*

Figure 3: Screenshot from the Interview Video with AI Recommendation under different explainability conditions: No Explainability (E=0) (a), Low Explainability (E=1) (b), and High Explainability (E=2) (c)

Cognitive Dissonance Measure

Cognitive dissonance was measured using a modified three-item version of the Post-Purchase Cognitive Dissonance Scale (Sweeney et al., 2000), originally a seven-item measure assessing discomfort after difficult decisions. For this study, three items were selected to capture emotional tension, cognitive discrepancy, and uncertainty, as these dimensions were most relevant to evaluating AI-driven recruitment recommendations. This shortened version improves conciseness while retaining the scale’s core components of post-decision dissonance.

Additionally, the original seven-point Likert scale was adjusted to a six-point scale to eliminate the neutral midpoint, a modification shown to improve response distributions in online surveys (Nuño & John, 2015). Higher scores on the modified scale indicate greater cognitive dissonance.

Trust Measure

Trust in the AI recruitment assistant was measured using a three-item scale adapted from Jian et al.'s (2000) multidimensional trust scale. This study selected three items assessing comfort with the AI's recommendations, belief in the accuracy of its judgments, and confidence in relying on its decisions during the experiment. These dimensions were most relevant to user acceptance and reliance on AI in recruitment.

As with the Cognitive Dissonance scale, the original seven-point Likert scale (1 = Strongly Disagree to 7 = Strongly Agree) was modified to a six-point scale (1 = Strongly Disagree to 6 = Strongly Agree) to eliminate the neutral midpoint and encourage more definitive responses. This concise version reduces participant burden while still capturing essential aspects of trust in human-AI collaboration. Higher scores on the adapted scale indicate greater trust in the AI recruitment assistant.

DATA ANALYSIS

This study employed a mixed-methods approach to data analysis, incorporating both descriptive and inferential statistics. Descriptive statistics, including means and standard deviations, were used to summarize participant demographics (age, sex, experience) and assess levels of cognitive dissonance and trust in AI. Inferential analyses were conducted to examine the relationship between AI explainability and these variables.

A one-way ANOVA with post-hoc Tukey's HSD tests was used to assess the impact of explainability on cognitive dissonance. Due to the non-normal distribution of trust scores, non-parametric tests, including the Kruskal-Wallis test and pairwise Mann-Whitney U tests with Bonferroni correction, were employed to analyze the effect of explainability on trust. Spearman's rank-order correlations were performed to explore the relationship between cognitive dissonance and trust across different explainability conditions. Finally, linear regression analysis was conducted to assess the predictive power of AI explainability on trust while controlling for age, sex, and work experience.

RESULTS AND DISCUSSION

RQ1: How do different levels of AI explainability impact cognitive dissonance and the development of trust in human-AI recruitment teams?

To examine the effect of AI explainability (E0 = No Explainability, E1 = Low Explainability, E2 = High Explainability) on cognitive dissonance and trust in AI, a series of statistical analyses were conducted, including group comparisons, correlation tests, and predictive modeling.

Effect of Explainability on Cognitive Dissonance

A one-way ANOVA examining the effect of explainability (No Explainability [E0], Low Explainability [E1], High Explainability [E2]) on cognitive dissonance revealed a non-significant overall effect, F(2, 69) = 2.45, p = .093. Pairwise comparisons were conducted to further examine differences in cognitive dissonance between conditions. The analysis revealed a significant increase in dissonance between E0 and E1 (t = -2.16, p = .036). No statistically significant differences were found between E1 and E2 (t = 0.90, p = .371) or between E0 and E2 (t = -1.33, p = .188) (Figure 4).

These findings suggest that introducing a low level of explainability may paradoxically increase cognitive dissonance. In the absence of any explanation (E0), users may rely on existing beliefs and mental models to interpret the AI's recommendations, as explored in research on user perceptions of algorithmic decision-making (Bashkirova & Krpan, 2024). This aligns with broader sensemaking research, which suggests that individuals fill gaps in understanding with pre-existing knowledge (Alvarado et al., 2020). Providing minimal explanations (E1) might disrupt this process by raising awareness of the AI's decision-making without sufficiently clarifying it, leading to greater uncertainty and dissonance. This effect may occur because partial explanations highlight discrepancies between users' intuitive understanding and the AI's rationale, creating ambiguity.

Confirmation bias may have played a role: when the AI's assessment matched expectations (e.g., confirming that a candidate was 'strong' or 'weak'), it was perceived as more credible (Bashkirova & Krpan, 2024), resulting in lower dissonance. Conversely, when the AI's evaluation contradicted initial beliefs, participants may have rationalized or dismissed the AI's reasoning to maintain their original judgment. This rationalization serves as a coping mechanism in situations where algorithmic predictions clash with human intuition, manifesting in selective attention—where individuals focus on details that support their viewpoint while disregarding conflicting information. Nickerson (1998) describes this as "case-building," in which individuals selectively gather or interpret evidence to reinforce their beliefs, often without conscious intent. In the low-explainability condition (E1), partial explanations may have provided just enough detail to trigger disagreement but not enough to fully convince participants, leaving them torn between their instincts and the AI's suggestion.

The dissonance scale's low reliability (α = .47) in the E1 condition further supports this interpretation, suggesting that ambiguity increased response variability. In contrast, the high-explainability condition (E2) may have provided sufficient detail to alleviate this tension or make participants' reasoning for agreement or disagreement more explicit, potentially reducing the uncertainty-driven dissonance observed with partial explanations.

However, the marginal significance of the overall ANOVA (p = .093) warrants caution in interpreting these findings. Future research with larger samples and more robust dissonance measures is needed to explore this complex relationship. Further investigation into how cognitive biases influence XAI-assisted decision-making could provide additional insights. Different theoretical perspectives on cognitive dissonance may offer alternative interpretive lenses. Specifically, exploring how forcing functions reduce overreliance on AI (Buçinca et al., 2021) may help mitigate the increased dissonance observed in the low-explainability condition. Additionally, research on how users' familiarity with recommender systems influences their behavior (Ghori et al., 2021) could further illuminate the impact of explanation depth on cognitive dissonance.

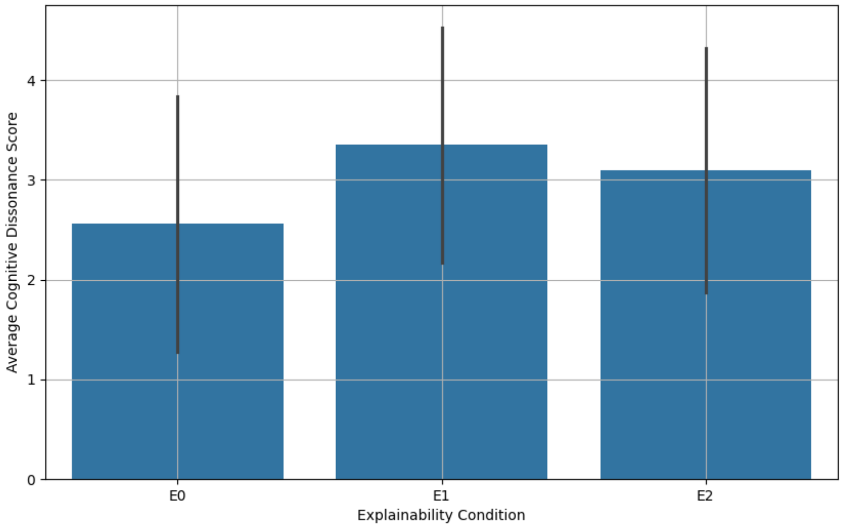


Figure 4: Average Cognitive Dissonance Scores by Explainability Condition

Effect of Explainability on Trust

Given the non-normal distribution of trust scores, a Kruskal-Wallis test was conducted to examine the effect of explainability conditions (no, low, high) on trust. The test revealed no significant effect (H = 0.99, p = .611), indicating that trust scores did not vary significantly across explainability conditions.

To further investigate potential differences between specific conditions, pairwise Mann-Whitney U tests with Bonferroni correction were performed. The results confirmed no significant differences in trust scores between the no-explainability (E0) and low-explainability (E1) conditions (p = .838), the low-explainability (E1) and high-explainability (E2) conditions (p = .665), and the no-explainability (E0) and high-explainability (E2) conditions (p = .866).

These findings consistently demonstrate that AI explainability does not significantly influence trust in AI-driven recruitment decisions. This suggests that factors beyond explainability may play a more substantial role in shaping trust. For instance, research indicates that perceived competence, benevolence, and integrity of the AI system are key determinants of trust (Kaplan et al., 2023; Langer et al., 2023). Additionally, perceived risk, prior experience with AI, and the transparency of the overall recruitment process can also influence trust formation (Schmidt et al., 2020; Ulfert et al., 2024).

In the recruitment context, procedural justice, fairness, and the perceived alignment of AI decisions with organizational values are particularly relevant (Hunkenschroer & Luetge, 2022). These findings align with recent research highlighting the complexity of trust in AI and the limitations of explainability as its sole determinant (de Brito Duarte et al., 2023).

Correlation Between Cognitive Dissonance and Trust

Spearman's rank-order correlation analyses were conducted for each explainability condition (no, low, high) to examine the relationship between cognitive dissonance and trust. It was hypothesized that lower cognitive dissonance would be associated with higher trust. However, contrary to expectations, no significant correlations were found across any of the explainability conditions (Figure 5).

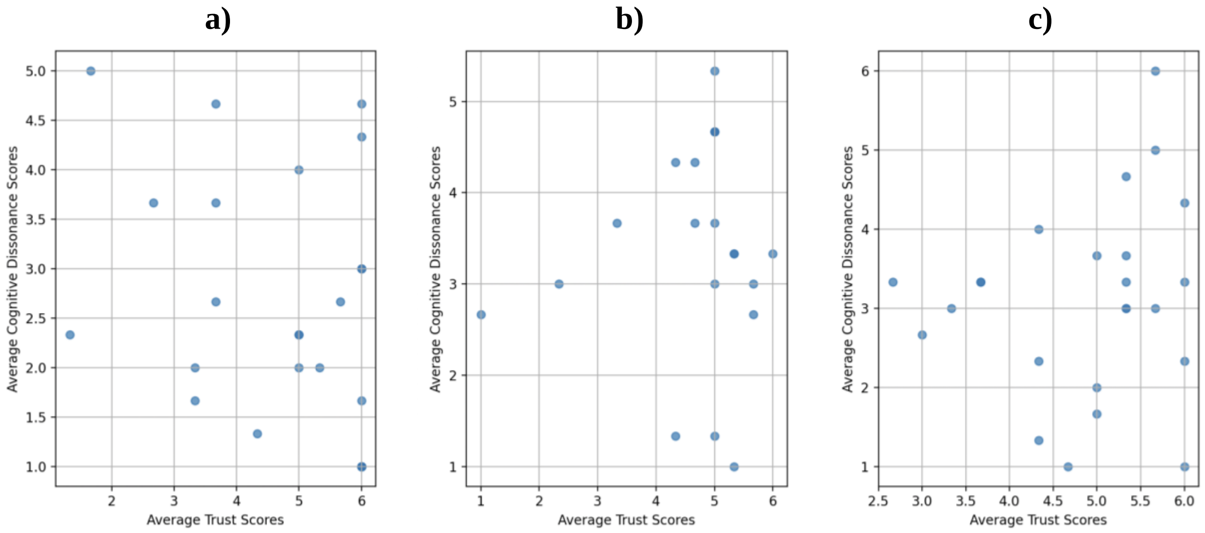


Figure 5: Scatterplot of Average Trust vs. Average Cognitive Dissonance Scores - No Explainability condition (a), Low Explainability condition (b), High Explainability condition (c)

The absence of significant correlations suggests that reducing cognitive dissonance does not necessarily increase trust in AI. This finding indicates that these constructs may operate independently in human-AI interactions, challenging the assumption that greater transparency and reduced dissonance inherently lead to higher trust. Similar null or negative effects of explainability on trust have been reported in prior research (Kästner et al., 2021).

RQ2: To what extent does AI explainability predict trust in human-AI recruitment teams?

A multiple linear regression analysis was conducted to examine the relationship between AI explainability and trust in the context of recruitment. This analysis assessed whether the explainability level (no, low, high) predicted trust in the AI recruitment assistant while controlling for potential confounding variables.

The regression model, with trust as the dependent variable and explainability level, age, and work experience as predictors, was not statistically significant, F(3, 68) = 0.056, p = .982. These results indicate that AI explainability did not significantly predict trust in the AI recruitment assistant (β = 0.036, p = .794). Similarly, neither age (β = -0.033, p = .841) nor work experience (β = 0.041, p = .811) was a significant predictor of trust.

Predictors	β	t	p	95% CI
const	3.468	2.599	1	[-0.267, 0.267]
Explainability	0.036	0.262	0.794	[-0.239, 0.311]
Age	-0.033	-0.202	0.841	[-0.364, 0.298]
Experience	0.041	0.240	0.811	[-0.298, 0.379]

Figure 6: Regression Analysis Results for Predictors of Trust

The non-significant relationship between explainability and trust contradicts some prior research suggesting that providing explanations enhances trust in AI systems (Arrieta et al., 2020). However, our findings align with Schmidt et al. (2020), who reported that transparency can sometimes negatively impact trust. In this simulated recruitment task, varying levels of explainability did not significantly influence participants' trust in AI, suggesting that explanations alone may not be sufficient to enhance trust in all scenarios.

This outcome may stem from the specific nature of the explanations provided, the characteristics of the participant sample, or the simplified design of the simulated recruitment task. Additionally, factors beyond explainability—such as perceived fairness of the AI system or prior experience with AI—may play a more significant role in shaping trust. Further research is needed to disentangle these factors.

LIMITATIONS AND FUTURE STUDIES

Reliability of Cognitive Dissonance Scale

The significantly lower internal consistency of the Cognitive Dissonance scale in the Low Explainability condition (α = .52) compared to the No Explainability (α = .74) and High Explainability (α = .79) conditions raises concerns about the reliability of dissonance measurement when explanations are only partially informative. This discrepancy may be driven by confirmation bias, which, as discussed earlier, can shape how individuals process ambiguous or incomplete information. Research suggests that confirmation bias strongly influences how individuals evaluate and reconcile conflicting evidence (Bashkirova & Krpan, 2024), a fundamental aspect of cognitive dissonance.

In the Low Explainability condition, participants may have selectively focused on elements of the AI’s explanation that aligned with their initial impressions of the candidate while disregarding conflicting details. When explanations were vague or incomplete, individuals had greater interpretative flexibility, allowing them to rationalize the AI’s reasoning in a way that reinforced their preexisting beliefs. This selective processing could explain the wide variability in dissonance scores—some participants resolved inconsistencies by interpreting the AI’s output as supporting their perspective, while others experienced heightened discomfort when the explanation partially contradicted their expectations but lacked enough justification to prompt genuine reconsideration.

These findings suggest that confirmation bias, rather than the structural properties of the explanation itself, may have been the primary driver of dissonance responses in the Low Explainability condition. Future research should further investigate how confirmation bias influences trust and cognitive dissonance in AI-assisted decision-making. Exploring bias-awareness interventions, structured decision protocols, or more explicit AI-generated explanations could help mitigate its effects. Additionally, refining dissonance measurement tools to account for bias-driven variability could improve our understanding of how individuals reconcile conflicts between AI-generated recommendations and their own intuitions, particularly in recruitment and other high-stakes decision-making contexts.

Simplified Interaction

To induce cognitive dissonance and isolate the effects of explainability, the experiment employed simplified interactions with somewhat exaggerated applicant behaviors and consistently high recommendation scores. While this design choice effectively heightened dissonance and allowed for a clearer examination of explainability, it also diverges from the complexities of real-world recruitment. In practice, applicant behaviors are more varied, and recommendations rarely exhibit uniformly high scores. This simplification may have amplified the observed effects of explainability on trust and dissonance, potentially overestimating their impact in more realistic settings.

Future research should address these limitations by incorporating more diverse applicant profiles, behaviors, and recommendation patterns to reflect real-world recruitment dynamics better. Enhancing ecological validity in this way would provide a more accurate understanding of how explainability influences trust and dissonance in AI-assisted hiring decisions. Additionally, integrating elements such as applicant feedback or interactive dialogue could further enrich the experimental design, offering a more comprehensive assessment of explainability’s role in shaping trust and decision-making within human-AI recruitment teams.

Geographical Diversity and Generalizability

This study did not restrict participants by country of origin; instead, HR professionals were recruited via LinkedIn, a platform with global reach. In theory, this open recruitment approach could yield an internationally diverse sample, enhancing the broad applicability of the findings. However, because geographic data were not collected, the cultural and regional composition of the sample remains unknown. This limitation prevents an assessment of cross-cultural representativeness and the potential influence of cultural differences on participants' responses. Given that attitudes toward AI and decision-making processes can vary significantly across cultures, the absence of location data constrains the generalizability of these results.

To enhance external validity, future research should explicitly account for geographic and cultural diversity. Expanding recruitment efforts to ensure broader representation and examining how cultural factors shape interactions with AI recommendations would strengthen the robustness of these findings and improve their applicability across different populations and organizational contexts.

Sample Specificity

This study focused specifically on HR professionals, which limits the generalizability of the findings to other domains where human-AI teams operate. While explainability, trust, and cognitive dissonance are relevant across various human-AI collaborations, the recruitment context and HR professionals’ expertise may have influenced the observed patterns. Given their domain knowledge and experience in candidate evaluation, HR professionals might exhibit different trust and dissonance responses compared to professionals in other fields interacting with AI.

Future research should explore these constructs in diverse professional settings, such as healthcare, finance, and engineering, to assess the extent to which these findings generalize across different domains and expertise levels. This would provide a more comprehensive understanding of how explainability, trust, and dissonance interact in a broader range of human-AI team environments.

CONCLUSIONS

This study investigated the impact of explainable AI on trust and cognitive dissonance in human-AI collaborative recruitment scenarios. Our findings revealed a nuanced relationship between explainability and cognitive dissonance. A statistically significant increase in dissonance was observed in the low explainability condition (E=1) compared to no explainability (E=0), suggesting that partial or ambiguous explanations introduce uncertainty, thereby heightening dissonance. However, the low reliability of the dissonance measure in this condition warrants cautious interpretation, as confirmation bias may have confounded the results. In the absence of clear explanations, participants may have selectively interpreted incomplete information in ways that reinforced their initial beliefs, reducing the need to engage with conflicting details and contributing to variability in dissonance scores. This aligns with existing research on AI-assisted decision-making, which suggests that individuals are more likely to trust and accept AI-generated recommendations when they align with their prior judgments (Bashkirova & Krpan, 2023). These findings highlight the need for more robust assessments that account for cognitive biases when evaluating the effects of explainability. They also reinforce the importance of clear, comprehensive explanations in mitigating dissonance, aligning with research emphasizing the role of explanation clarity in trust calibration (Naiseh et al., 2023).

Interestingly, despite the observed effects of explainability on dissonance, we found no significant relationship between explainability and participants' trust in the AI recruitment assistant. This suggests that while ambiguous explanations may heighten cognitive discomfort, they do not necessarily erode trust in AI. Instead, trust formation in AI-assisted decision-making may be influenced by other factors beyond explainability, such as perceived fairness, reliability, and prior user experience. This interpretation aligns with recent research proposing a more complex, non-linear relationship between explainability and trust (e.g., Morandini et al., 2023). Given that confirmation bias appeared to shape dissonance responses in the low explainability condition, future research should further examine how individual cognitive tendencies—such as susceptibility to confirmation bias, cognitive styles, and prior experience with AI—mediate both trust and dissonance in human-AI teams. Investigating these factors will provide deeper insights into how explainability strategies can be optimized to support trust calibration and mitigate bias in AI-assisted decision-making.

REFERENCES

Alvarado, O., Heuer, H., Vanden Abeele, V., Breiter, A., & Verbert, K. (2020). Middle-aged video consumers' beliefs about algorithmic recommendations on YouTube. *Proceedings of the ACM on Human-Computer Interaction, 4*(CSCW2), 121. https://doi.org/10.1145/3415192

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58,* 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Bashkirova, A., & Krpan, D. (2024). Confirmation bias in AI-assisted decision-making: AI triage recommendations congruent with expert judgments increase psychologist trust and recommendation acceptance. *Computers in Human Behavior: Artificial Humans, 2*(1), 100066. https://doi.org/10.1016/j.chbah.2024.100066

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., ... & Eckersley, P. (2020, January). Explainable machine learning in deployment. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 648-657). https://doi.org/10.1145/3351095.3375624

Buçinca, Z., Malaya, M. B., & 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), 188. https://doi.org/10.1145/3449287

Bussone, A., Stumpf, S., & O'Sullivan, D. (2015, October). The role of explanations on trust and reliance in clinical decision support systems. In *2015 International Conference on Healthcare Informatics* (pp. 160-169). IEEE. https://doi.org/10.1109/ichi.2015.26

de Brito Duarte, R., Correia, F., Arriaga, P., & Paiva, A. (2023). AI trust: Can explainable AI enhance warranted trust?. *Human Behavior and Emerging Technologies*, *2023*(1), 4637678. https://doi.org/10.1155/2023/4637678

Denault, V. (2020). Misconceptions about nonverbal cues to deception: A covert threat to the justice system?. *Frontiers in Psychology, 11,* 573460. https://doi.org/10.3389/fpsyg.2020.573460

Došilović, F. K., Brčić, M., & Hlupić, N. (2018, May). Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 0210-0215). IEEE. https://doi.org/10.23919/mipro.2018.8400040

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society, 20*(5), 325–344. https://doi.org/10.1080/01972240490507974

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*(4), 93-106.

Ghori, M. F., Dehpanah, A., Gemmell, J., Qahri-Saremi, H., & Mobasher, B. (2021). *How does the user's knowledge of the recommender influence their behavior?* arXiv. https://doi.org/10.48550/arxiv.2109.00982

Gillespie, N., Lockey, S., Curtis, C., Pool, J. K., & Akbari, A. (2023). *Trust in Artificial Intelligence: A global study*. Brisbane, Australia; New York, United States: The University of Queensland; KPMG Australia. https://doi.org/10.14264/00d3c94

Giovine, C., & Roberts, R. (2024). *Building AI trust: The key role of explainability.* McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/building-ai-trust-the-key-role-of-explainability

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Haufe, S., Wilming, R., Clark, B., Zhumagambetov, R., Panknin, D., & Boubekki, A. (2024). *Explainable AI needs formal notions of explanation correctness*. arXiv. https://doi.org/10.48550/arxiv.2409.14590

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda [Review of Ethics of AI-enabled recruiting and selection: A review and research agenda]. *Journal of Business Ethics*, *178*(4), 977-1007. https://doi.org/10.1007/s10551-022-05049-6

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, *4*(1), 53-71. https://doi.org/10.1207/S15327566IJCE0401_04

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, *65*(2), 337-359. https://doi.org/10.1177/00187208211013988

Kästner, L., Langer, M., Lazar, V., Schomäcker, A., Speith, T., & Sterz, S. (2021). *On the relation of trust and explainability: Why to engineer for trustworthiness*. arXiv. https://doi.org/10.48550/arxiv.2108.05379

Keutzer, C. S. (1968). A measure of cognitive dissonance as a predictor of smoking treatment outcome. *Psychological Reports*, *22*(2), 655-658. https://doi.org/10.2466/pr0.1968.22.2.65

Kupfer, C., Prassl, R., Fleiß, J., Malin, C., Thalmann, S., & Kubicek, B. (2023). Check the box! How to deal with automation bias in AI-based personnel selection. *Frontiers in Psychology*, *14*, 1118723. https://doi.org/10.3389/fpsyg.2023.1118723

Langer, M., König, C. J., Back, C., & Hemsing, V. (2023). Trust in artificial intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. *Journal of Business and Psychology*, *38*(3), 493-508. https://doi.org/10.1007/s10869-022-09829-9

Levin, D. T., Harriott, C., Paul, N. A., Zhang, T., & Adams, J. A. (2013). Cognitive dissonance as a measure of reactions to human-robot interaction. *Journal of Human-Robot Interaction*, *2*(3), 3-17. https://doi.org/10.5898/jhri.2.3.levin

Morandini, S., Fraboni, F., Puzzo, G., Giusino, D., Volpi, L., Brendel, H., Balatti, E., Angelis, M.D., Cesarei, A.D., & Pietrantoni, L. (2023). Examining the nexus between explainability of AI systems and user's trust: A preliminary scoping review. *CEUR Workshop Proceedings*, *3554*, 6.

Naiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2023). How the different explanation classes impact trust calibration: The case of clinical decision support systems. *International Journal of Human-Computer Studies*, *169*, 102941. https://doi.org/10.1016/j.ijhcs.2022.102941

Nuño, A., & John, F. A. S. (2015). How to ask sensitive questions in conservation: A review of specialized questioning techniques. *Biological Conservation*, *189*, 5-15. https://doi.org/10.1016/j.biocon.2014.09.047

Oshikawa, S. (1972). The measurement of cognitive dissonance: Some experimental findings. *Journal of Marketing*, *36*(1), 64-67. https://doi.org/10.1177/002224297203600112

Paas, F. G. W. C., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1), 419-430. https://doi.org/10.2466/pms.1994.79.1.419

Peters, T.M., & Visser, R.W. (2023). The Importance of Distrust in AI. In L. Longo (Ed.), *Explainable Artificial Intelligence. xAI 2023. Communications in Computer and*

Information Science (Vol. 1903). Springer, Cham. https://doi.org/10.1007/978-3-031-44070-0_15

Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (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), 83. https://doi.org/10.1145/3512930

Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems*, *29*(4), 260-278. https://doi.org/10.1080/12460125.2020.1819094

Sharma, S., Kaushik, K., Sharma, R., & Chaturvedi, N. (2023). Explainable Artificial Intelligence (XAI). *IJFANS International Journal of Food and Nutritional Sciences, 12*(1), 2660-2666.

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. https://doi.org/10.1016/j.ijhcs.2020.102551

Sivaraman, V., Bukowski, L. A., Levin, J. R., Kahn, J. M., & Perer, A. (2023). *Ignore, trust, or negotiate: Understanding clinician acceptance of AI-based treatment recommendations in health care*. arXiv. https://doi.org/10.48550/arxiv.2302.00096

Small, E., Xuan, Y., Hettiachchi, D., & Sokol, K. (2023). *Helpful, misleading or confusing: How humans perceive fundamental building blocks of Artificial Intelligence explanations*. arXiv. https://doi.org/10.48550/arXiv.2303

Smelyakov, K., Hurova, Y., & Osiievskyi, S. (2023). Analysis of the effectiveness of using machine learning algorithms to make hiring decisions. *CEUR Workshop Proceedings, 3387*, 7.

Sweeney, J. C., Hausknecht, D., & Soutar, G. N. (2000). Cognitive dissonance after purchase: A multidimensional scale. *Psychology & Marketing*, *17*(5), 369-385. https://doi.org/10.1002/(SICI)1520-6793(200005)17:5%3C369::AID-MAR1%3E3.0.CO;2-G

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Ulfert, A. S., Georganta, E., Centeio Jorge, C., Mehrotra, S., & Tielman, M. (2024). Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework. *European Journal of Work and Organizational Psychology*, *33*(2), 158-171. https://doi.org/10.1080/1359432x.2023.2200172

Wang, X., & Yin, M. (2022). Effects of explanations in AI-assisted decision making: Principles and comparisons. *ACM Transactions on Interactive Intelligent Systems*, *12*(4), 27. https://doi.org/10.1145/3519266

Zhang, M. (2021, August). Research on cross-cultural differences in nonverbal communication between America and China. In *Proceedings of the 2021 5th International Seminar on Education, Management and Social Sciences (ISEMSS 2021): Advances in Social Science, Education and Humanities Research* (pp. 954-957). Atlantis Press. https://doi.org/10.2991/assehr.k.210806.181

Zhang, Y., Liao, Q. V., & Bellamy, R. K. (2020, January). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 295-305). https://doi.org/10.1145/3351095.3372852