Scholarly Research Journal for Interdisciplinary Studies
Online ISSN 2278-8808, SJIF 2024 = 8.153
https://www.srjis.com/issues_data/239
Peer Reviewed & Refereed Journal, SEPT-OCT 2024, Vol- 13/85

BUILDING TRUST IN HUMAN-AI COLLABORATION IN THE AGE OF DIGITALIZATION: A COMPREHENSIVE STUDY

Neha Jain

Ph.D Research Scholar, Pacific Academy of Higher Education and Research University, Udaipur. Email- neha.bordia@gmail.com

Dr. Neetu Agarwal

Associate Professor, Pacific Academy of Higher Education and Research University, Udaipur. Email- neetu.agarwal1508@gmail.com

Paper Received On: 20 SEPT 2024

Peer Reviewed On: 24 OCT 2024

Published On: 01 NOV 2024

Abstract

Trust is a cornerstone of effective human-AI collaboration, particularly in an era of rapid digitalization where AI systems are increasingly integrated into decision-making processes across various sectors. This study investigates the critical factors influencing trust namely Transparency, Interpretability, and Satisfaction, and their sector-specific dynamics in healthcare, finance, and customer service. Utilizing a cross-sectional survey of 500 participants and stratified sampling, the research highlights pivotal role of transparency and interpretability in fostering trust, particularly in high-stakes sectors such as healthcare and finance. Transparency (β = 0.512, $p < 0.001$) and interpretability (β = 0.602, $p < 0.001$) significantly enhance trust, with stronger effects observed in healthcare ($R2$ = 0.494) and finance ($R2$ = 0.511) compared to customer service ($R2$= 0.374). Satisfaction had been emerged as a crucial mediating variable that amplifies the relationship between transparency and trust. The indirect effect of transparency on trust through satisfaction (β = 0.223, $p < 0.001$) underscores the importance of user-centric design in building trust. Furthermore, satisfaction demonstrates a stronger influence on trust in customer service (β=0.653), emphasizing its importance in customer-facing applications. This study provides theoretical contributions by extending trust frameworks to sector-specific contexts and offers actionable insights for AI system developers and policymakers. The findings advocate for tailored trust-building strategies, prioritizing transparency and interpretability in healthcare and finance, while emphasizing user satisfaction in customer service. The research will advances the understanding of trust dynamics in human-AI collaboration, addressing the ethical, operational, and design challenges of AI systems in a digitalized world.

***Keywords**: Human-AI Collaboration, Trust Dynamics, Transparency, Interpretability, Satisfaction, Sector-specific analysis, AI System*

INTRODUCTION

Digitalization has transformed industries globally, ushering in an era where artificial intelligence (AI) not only complements but often guides decision-making processes across various domains. This shift brings forth profound implications for how human-AI interactions are structured, understood, and optimized, especially in fields like healthcare, finance, and customer service, where trust is paramount (Ferrario et al., 2019; Rai, 2020). Digitalization's impact on these sectors is substantial, enabling the widespread use of AI for tasks requiring precision, speed, and, increasingly, ethical consideration. As AI continues to advance, human reliance on these systems raises critical questions about trust, accountability, and transparency, which are essential to ensuring the efficacy and adoption of AI-driven decision-making systems (Awad et al., 2018).

A. The Role of Digitalization in Human-AI Collaboration: The digital transformation has altered the dynamics of traditional industries, establishing AI as a powerful partner in data processing, predictive modeling, and decision support. Healthcare is a prime example: AI applications range from diagnostic assistance to patient monitoring and personalized treatment planning, where timely, data-driven decisions are essential (Benjamins & Florez, 2020). In finance, AI is employed to detect fraudulent activities, automate trading processes, and manage risks, benefiting from AI's ability to analyze vast datasets in real time (Theodorou & Dignum, 2020). Similarly, in customer service, AI chatbots and recommendation systems enable more efficient user engagement and personalized experiences, effectively complementing human service agents (Lee et al., 2021).

However, this growing reliance on AI raises the stakes for trust, especially as digitalization integrates these systems deeper into decision-making structures. Unlike traditional software, AI systems in digitalized environments often operate autonomously or semi-autonomously, making decisions based on algorithms that may not always be fully transparent to human users. This opacity can create an inherent barrier to trust, as users may struggle to comprehend or predict AI behavior (Ferrario et al., 2019). Thus, establishing trust has become essential to human-AI collaboration, as it influences not only the acceptance and use of AI systems but also their long-term integration and efficacy across sectors.

B. The Importance of Trust in Human-AI Collaboration: Trust in human-AI interactions is a multi-dimensional concept that encompasses users' beliefs in the reliability, competence, and ethical alignment of AI systems (Lakkaraju et al., 2017). As AI systems are increasingly used for high-stakes decision-making, users' willingness to trust these systems becomes crucial for seamless collaboration. Trust serves as a foundation for user engagement and adoption, especially in areas where the potential consequences of AI decisions are significant, such as medical diagnosis or financial investments (Awad et al., 2018). Without trust, users may resist or underutilize AI systems, thereby limiting their effectiveness and the value they can bring to decision-making processes.

Scholars have identified several factors that contribute to trust in AI, including transparency, interpretability, fairness, and ethical alignment (Rai, 2020; Ribeiro et al., 2016). Transparency allows users to understand how AI systems arrive at specific conclusions, making it easier to trust their output. Interpretability complements transparency by enabling users to grasp not only the outcomes but also the processes and reasoning behind AI decisions (Siau & Wang, 2018). When AI systems are transparent and interpretable, users are more likely to trust and rely on them for critical tasks. Furthermore, fairness and ethical alignment are increasingly seen as non-negotiable attributes, particularly as AI is deployed in sectors where biases or errors could have severe implications for individuals' well-being and rights (Benjamins & Florez, 2020).

C. Sector-Specific Trust Dynamics: The need for trust in human-AI interactions varies significantly across sectors due to differences in decision-criticality, regulatory standards, and user expectations. In healthcare, trust is paramount because AI-driven diagnostic tools and treatment recommendations directly affect patient outcomes. Healthcare professionals must trust that AI systems are accurate, unbiased, and aligned with ethical standards to adopt them fully into their practices (Ferrario et al., 2019). Studies show that when AI systems are not transparent or fail to provide interpretable insights, healthcare providers may be hesitant to rely on them, preferring traditional decision-making approaches over opaque or uncertain AI recommendations (Chen et al., 2021).

In finance, trust dynamics are shaped by a strong emphasis on reliability and risk management. AI applications in this sector must demonstrate not only accuracy but also resilience in real-time, high-stakes environments where financial stability and client confidence are at risk. As AI becomes central to fraud detection and risk assessment, financial institutions and their clients rely on these systems' ability to process information quickly and accurately without compromising ethical standards (Lee et al., 2021). Customer service, while generally lower-stakes compared to healthcare or finance, also requires trust, especially as AI interfaces increasingly manage interactions that once required human empathy and adaptability. Users need confidence that these systems will respond effectively and fairly to their needs, adapting to unique requests in a way that aligns with the service's quality standards (Theodorou & Dignum, 2020).

D. Challenges to Building Trust in Digitalized Human-AI Systems: Despite the acknowledged importance of trust, several challenges remain in achieving trustworthiness within digitalized AI systems. One of the primary challenges is opacity, as many AI systems operate as "black boxes" with complex algorithms that are not easily understandable by end-users (Ribeiro et al., 2016). Black-box models, particularly deep learning-based AI, often produce high-accuracy predictions but lack interpretability, making it difficult for users to understand the rationale behind specific outcomes. For critical applications, the inability to interpret AI decisions can erode user trust and hinder adoption (Siau & Wang, 2018).

Another challenge is the potential for bias and unfairness in AI systems. When AI algorithms are trained on historical data, they may inadvertently learn and perpetuate biases present in the data, leading to unfair or discriminatory outcomes. This is particularly concerning in sectors like healthcare and finance, where biased decisions can have significant consequences for individuals and communities. Researchers emphasize the importance of fair AI models, particularly in digitalized environments where decisions are automated and reach a large number of users (Lakkaraju et al., 2017). The need for ethical AI is critical, as it directly influences the perception of trustworthiness in AI systems.

Moreover, cultural differences can impact how users perceive and trust AI. A user's cultural background often shapes their expectations and comfort levels with automation and technology (Chen et al., 2021). Thus, trust in AI may not be uniform across different demographic and geographic contexts, underscoring the need for culturally aware AI systems that accommodate diverse user perspectives. This variation poses a challenge for developers and organizations aiming to design universally trusted AI systems, particularly as digitalization connects users across borders.

REVIEW OF LITERATURE

The rise of digitalization has brought artificial intelligence (AI) into the mainstream of critical industries, transforming workflows and decision-making processes in healthcare, finance, and customer service. Human-AI collaboration now requires an understanding of how trust is built and maintained across diverse applications. Trust is recognized as essential for effective human-AI collaboration, particularly as AI systems are increasingly involved in high-stakes and complex decision-making tasks (Benjamins & Florez, 2020; Siau & Wang, 2018). This review synthesizes research on trust mechanisms in AI, focusing on transparency, interpretability, and sector-specific dynamics that influence user acceptance and trust in AI systems.

Transparency and Interpretability in AI: Transparency and interpretability are two critical factors consistently identified in the literature as central to fostering user trust in AI systems (Rai, 2020). Transparency, broadly understood as the degree to which AI systems disclose their decision-making processes and limitations, is essential for user trust, particularly when these systems operate autonomously (Awad et al., 2018). Transparent systems offer users insight into the logic behind AI decisions, helping to alleviate concerns about “black box” operations where the inner workings of AI algorithms are opaque (Ribeiro et al., 2016). Transparency is especially significant in sectors like healthcare and finance, where decisions can have direct consequences for individuals' health and financial well-being (Chen et al., 2021).

Interpretability complements transparency by providing users with explanations of how AI systems arrive at specific decisions. Interpretability frameworks, such as interpretable decision sets or visual explanations, allow users to trace an AI’s reasoning, making the decision-making process more accessible and understandable (Lakkaraju et al., 2017). Studies suggest that interpretability is critical for user trust, as it enables users to assess the validity of AI recommendations, particularly in complex or unfamiliar contexts. For instance, Ribeiro et al. (2016) demonstrated that interpretable models were more likely to be trusted by users in healthcare settings, where understanding the basis for medical decisions is crucial for both professionals and patients. Overall, transparency and interpretability are seen as foundational to establishing initial trust, ensuring that users feel informed and in control when interacting with AI systems (Siau & Wang, 2018).

Fairness, Bias, and Ethical Considerations: Ethical AI is a growing field addressing concerns about bias, fairness, and accountability in AI systems. Research highlights the importance of fair and unbiased AI models, particularly in high-stakes sectors where biased outcomes could have serious consequences (Theodorou & Dignum, 2020). Bias in AI often originates from training datasets that reflect historical inequalities, leading to unfair or discriminatory outcomes in practice. For example, Obermeyer et al. (2019) found that certain healthcare algorithms disproportionately disadvantaged minority groups, raising ethical concerns and highlighting the need for more equitable AI systems.

To address such issues, researchers advocate for fairness-aware algorithms that explicitly seek to reduce bias during the model training process. Ferrario et al. (2019) emphasize that users are more likely to trust AI systems that demonstrate ethical responsibility, as fair treatment aligns with societal values of justice and equality. This perspective underscores the idea that user trust is not solely based on AI's technical accuracy but also on its ethical alignment with human values.

Sector-Specific Dynamics of Trust: The degree to which users trust AI can vary significantly depending on the sector and the nature of the tasks involved. In healthcare, AI systems often assist with diagnostic and treatment decisions, requiring an exceptionally high level of trust due to the potentially life-or-death implications (Rai, 2020). Research shows that healthcare professionals are more likely to trust AI systems when they are transparent and interpretable, as this allows them to validate AI recommendations and integrate them responsibly into patient care (Chen et al., 2021). The sector's regulatory environment also demands rigorous standards for safety and transparency, which may enhance trust among users if met consistently.

In finance, trust in AI systems is essential but shaped by slightly different factors, such as risk management and reliability. Financial institutions leverage AI for real-time data analysis, fraud detection, and predictive modeling, and the need for trust is closely tied to these systems' accuracy and reliability under dynamic market conditions (Theodorou & Dignum, 2020). Studies indicate that trust is likely to erode if AI systems produce unreliable or inconsistent results, especially in high-risk scenarios like trading or credit evaluation (Lee et al., 2021). Financial users also benefit from transparency and interpretability, which enable them to assess risk in AI-driven predictions and make informed decisions.

Customer service, while generally less high-stakes than healthcare or finance, still requires a level of user trust for AI systems to be effective. AI tools in this sector, such as chatbots and recommendation engines, facilitate efficient customer interaction, but trust is essential for users to feel confident in the responses and solutions provided by these systems (Benjamins & Florez, 2020). The literature indicates that customer satisfaction with AI tools in service contexts is often contingent on responsiveness, adaptiveness, and the ability to handle complex, individualized requests. Trust in customer service AI is influenced by user perceptions of empathy and relevance, which are less technical but critical for sustained engagement.

Gaps in Current Research

Despite extensive research, several gaps remain in the understanding of trust in human-AI collaboration. First, there is limited empirical research examining trust dynamics over time, particularly as users interact with AI systems across various stages of familiarity and experience (Ferrario et al., 2019). Longitudinal studies could provide insights into how trust evolves or degrades based on user experiences with AI, revealing whether initial trust factors continue to hold significance or if new factors emerge. Second, cultural considerations in AI trust are underexplored, even though culture can shape expectations and comfort levels with technology (Chen et al., 2021). Cross-cultural studies could help to identify diverse trust requirements and create AI systems that accommodate a broader range of user perspectives.

Finally, the impact of AI transparency and interpretability on trust remains context-dependent, and future research is needed to develop sector-specific guidelines that address unique trust requirements across industries (Awad et al., 2018). Addressing these gaps will be critical to advancing the design of AI systems that are trustworthy, fair, and adaptable, thereby enhancing human-AI collaboration in digitalized environments.

RESEARCH OBJECTIVES AND HYPOTHESES

RO1: To evaluate the impact of Transparency and Interpretability on Trust in human-AI collaboration across healthcare, finance, and customer service sectors.

RO2: To analyse the mediating role of Satisfaction in the relationship between Transparency and Trust in human-AI interactions.

RO3: To examine sector-specific differences in how Trust is influenced by Transparency, Interpretability, and Satisfaction.

The aforementioned objectives are tightly aligned with the following research hypotheses and ensures a focused exploration of trust dynamics within human-AI collaboration.

H_1: Transparency positively influences Trust in human-AI collaboration.

H_2: Interpretability positively influences Trust in human-AI collaboration.

H_3: Transparency positively influences Satisfaction with AI systems.

H_4: Satisfaction mediates the relationship between Transparency and Trust in human-AI collaboration.

H_5: Satisfaction positively influences Trust in human-AI collaboration.

H_6: The impact of Transparency and Interpretability on Trust is stronger in healthcare and finance than in customer service.

H_7: Satisfaction has a stronger influence on Trust in customer service compared to healthcare and finance.

RESEARCH METHODOLOGY

This study employs a mixed-methods approach to explore the dynamics of trust in human-AI collaboration across healthcare, finance, and customer service sectors. By focusing on key constructs such as Transparency, Interpretability, Satisfaction, and Trust, the methodology ensures a comprehensive examination of trust-building factors and their sector-specific impacts. The primary data collection method involves a cross-sectional quantitative survey, enabling the study to capture a broad range of user perceptions within a defined time frame. This approach aligns with established practices for investigating user attitudes in digitalized environments (Groves et al., 2009).

A. Research Design: A cross-sectional survey design was chosen for its ability to provide a snapshot of user attitudes and perceptions. This design facilitates the measurement of trust and its influencing factors across different sectors simultaneously, ensuring consistency in data collection. The quantitative survey includes constructs such as perceived transparency (Rai, 2020), interpretability (Lakkaraju et al., 2017), and overall trust in AI systems (Siau & Wang, 2018). Each construct was carefully adapted from validated scales to align with the research objectives and hypotheses. The survey's Likert-scale format, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), allows respondents to express varying degrees of agreement with statements related to trust, ethical considerations, and satisfaction with AI systems.

B. Sampling and Population: The target population consists of AI users from healthcare, finance, and customer service sectors, chosen for their distinct trust requirements and user expectations (Benjamins & Florez, 2020). A stratified sampling approach was employed to ensure sectoral representation, capturing variations in trust dynamics across these domains. An initial sample size of 500 participants was determined using power analysis to ensure statistically meaningful results (Cohen, 1992). Each sector was allocated approximately 165 respondents to maintain balanced representation. The demographic diversity of participants, including factors such as age, education, and professional experience, was considered to enhance the study's generalizability.

C. Data Collection: The primary data collection tool was a structured survey instrument, developed using validated scales tailored to the study's focus. Key constructs measured include:

- **Transparency:** Adapted from Rai (2020), focusing on the degree to which users perceive openness in AI decision-making processes.
- **Interpretability:** Based on Lakkaraju et al. (2017), assessing the clarity of AI system explanations.
- **Satisfaction:** Custom items evaluating user satisfaction with AI system performance and decision-making.
- **Trust:** Derived from Siau and Wang (2018), capturing overall confidence in AI systems.

The survey instrument underwent pilot testing with 30 participants to refine question clarity and relevance, resulting in minor adjustments. The final survey was distributed online to ensure wide reach and convenience for respondents.

D. Data Analysis: Quantitative data were analyzed using SPSS software to perform descriptive and inferential statistical tests. Reliability analysis confirmed internal consistency, with Cronbach's alpha values exceeding 0.7 for all constructs (Tavakol & Dennick, 2011). Factor analysis verified the validity of the survey items, ensuring accurate measurement of constructs.

A series of regression analyses were conducted to test the hypotheses, focusing on the influence of Transparency and Interpretability on Trust, as well as the mediating role of Satisfaction. ANOVA was used to compare trust levels across sectors, identifying significant differences in user perceptions between healthcare, finance, and customer service. In addition, mediation analysis was performed to assess whether Satisfaction acts as a bridge between Transparency and Trust, providing empirical evidence for the hypothesized relationships.

E. Validity and Reliability: To ensure the robustness of the findings, multiple strategies were employed to validate the methodology. Construct validity was established through factor analysis, confirming that survey items accurately represented the intended dimensions. The survey instrument's reliability was demonstrated through high Cronbach's alpha values. Additionally, a pilot study provided preliminary insights and facilitated adjustments to improve clarity.

F. Ethical Considerations: Ethical guidelines were strictly followed to protect participants' rights and confidentiality. Respondents were informed about the study's purpose, assured of voluntary participation, and provided the option to withdraw at any time. Informed consent was obtained prior to data collection, and data anonymization ensured participant privacy. The research protocol received approval from an institutional review board, confirming compliance with ethical standards (Creswell, 2014).

G. Limitations: While the methodology is designed to provide comprehensive insights, certain limitations must be acknowledged. The cross-sectional design offers a snapshot of trust dynamics but does not account for changes over time. Additionally, the sample, while representative of key sectors, may not fully capture geographic and cultural variations in trust perceptions. Future research could address these limitations by adopting a longitudinal approach and including more diverse samples to explore trust dynamics across different contexts (Ferrario et al., 2019).

DATA ANALYSIS AND INTERPRETATION

Reliability Analysis: The reliability of the constructs used in the study, including Transparency, Interpretability, Satisfaction, and Trust, was assessed using Cronbach's alpha. A Cronbach's alpha value of 0.70 or higher indicates good internal consistency of the items within each construct (Tavakol & Dennick, 2011).

Table 1: Cronbach's Alpha Reliability Test Results

Construct	Number of Items	Cronbach's Alpha	Interpretation
Transparency	5	0.832	High reliability
Interpretability	5	0.814	High reliability
Satisfaction	4	0.780	Good reliability
Trust	6	0.857	High reliability

Source: Primary Data

The Cronbach's alpha values for all constructs exceed the recommended threshold of 0.70, indicating that the survey items measuring these constructs are internally consistent and reliable. With a Cronbach's alpha of 0.832, the Transparency demonstrated high reliability, suggesting that the items effectively measure users' perceptions of AI system transparency. Interpretability construct with Cronbach's alpha of 0.814, reflected high internal consistency in assessing how well users understand AI system decisions and processes. The alpha value of 0.780 for satisfaction had indicated good reliability, and had showed that the items reliably capture the user satisfaction levels regarding AI system interactions. Trust construct have recorded highest Cronbach's alpha value (0.857) and have highlighted very high reliability in measuring users' overall trust in AI systems.

Frequency Distribution Analysis of Demographics: Demographic profile of the respondents was analysed to provide insights into the sample characteristics. Below is the table for frequency distribution.

Table 2: Frequency Distribution of Demographics

Demographic Variable	Category	Frequency (n = 500)	Percentage
Gender	Male	260	52.0%
	Female	240	48.0%
Age Group	18–25	120	24.0%
	26–35	180	36.0%
	36–50	150	30.0%
	Above 50	50	10.0%
Sector	Healthcare	165	33.0%
	Finance	170	34.0%
	Customer Service	165	33.0%
Education Level	Undergraduate	150	30.0%
	Graduate	250	50.0%
	Postgraduate and above	100	20.0%

Source: Primary Data

With reference to the demographic distribution of the sample (n = 500), balanced representation across key demographic variables, ensuring diverse insights into trust dynamics in human-AI collaboration have been noticed. For Gender aspect, sample included 52% male respondents (n = 260) and 48% female respondents (n = 240), providing a nearly even gender distribution. This balance ensures the study captures potential gender-based variations in trust, transparency, and interpretability in AI systems. Further, for age as a demographic variables it was noticed that respondents were distributed across various age groups, with the majority (36%, n = 180) falling in the 26–35 age range. The 18–25 group constitutes 24% (n = 120), followed by 30% (n = 150) in the 36–50 range, and 10% (n = 50) aged above 50. This spread reflects a concentration of young to middle-aged individuals who are likely more familiar with digitalized environments and AI systems, while also including older age groups for comprehensive insights.

Additionally, it was noticed that the sample was evenly distributed across three key study sectors: healthcare (33%, n = 165), finance (34%, n = 170), and customer service (33%, n = 165). This stratification ensures the study captures sector-specific trust dynamics, recognizing that each industry involves unique user expectations and interaction scenarios with AI systems. Educational attainment among respondents indicated that 50% (n = 250) respondents were graduate, 30% (n = 150) respondents were undergraduates, and remaining 20% (n = 100) respondents have possessed postgraduate or higher qualifications. This distribution emphasizes the inclusion of a relatively educated sample, critical for understanding perceptions of complex AI concepts like transparency, interpretability, and trust. So, overall the demographic breakdown demonstrated a well-rounded sample representing diverse age groups, genders, education levels, and industry sectors. This diversity strengthens the study's generalizability and relevance, enabling robust analysis of trust dynamics in human-AI collaboration across different contexts.

Regression Analysis: Transparency Significantly Influences User Trust in AI Systems: In order to analyse the impact of transparency over the trust of user for human-AI collaboration linear regression analysis was performed, results are hereunder:

Table 3: Regression Analysis Statistics of Transparency Impact on Users' Trust in Human-AI Collaboration Systems

Model Summary						
Model	**R**	**R^2**	**Adj. R^2**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Transparency	0.521	0.271	0.269	0.484	113.85	< 0.001
Regression Coefficients						
Variable		**β Coefficient**		**Std. Error**	**t-Value**	**Sig.**
Transparency		0.512		0.048	10.67	< 0.001
Constant		1.245		0.153	8.14	< 0.001
ANOVA for Model Fit						
Source	**Sum of Squares**		**df**	**Mean Square**	**F**	**p-Value**
Regression	26.562		1	26.562	113.85	< 0.001
Residual	71.265		498	0.143		
Total	97.827		499			

Source: Primary Data

The R^2 value of 0.271 indicated that 27.1% of the variance in user trust is explained by transparency. The model is statistically significant ($F = 113.85$, $p < 0.001$), and suggested that the predictor variable, transparency, meaningfully contributes to explaining trust in AI systems. Transparency is a significant predictor of trust in AI systems ($\beta = 0.512$, $p < 0.001$). The positive β value indicates that higher levels of perceived transparency are associated with greater trust. For every unit increase in transparency, trust increases by 0.512 units. The ANOVA statistics had confirmed the model's statistical significance ($F = 113.85$, $p < 0.001$), indicating that the regression model fits the data well. The statistics of the regression analysis lead into acceptance of Hypothesis 1 i.e. "Transparency positively influences Trust in human-AI collaboration", showing that transparency significantly and positively influences user trust in human-AI collaboration. Transparency explains a substantial portion of the variance in trust, emphasizing its critical role in enhancing users' trust for AI systems human-AI collaboration.

Regression Analysis: Interpretability Significantly Influences User Trust in human-AI collaboration: In order to analyse the impact of interpretability over the trust of user for human-AI collaboration linear regression analysis was performed, results are hereunder:

Table 4: Regression Analysis Statistics of Interpretability Impact on Users' Trust in Human-AI Collaboration

Model Summary						
Model	**R**	**R²**	**Adj. R²**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Interpretability	0.563	0.317	0.316	0.465	230.85	< 0.001
Regression Coefficients						
Variable		**β Coefficient**		**Std. Error**	**t-Value**	**Sig.**
Interpretability		0.602		0.040	15.21	< 0.001
Constant		0.957		0.129	7.42	< 0.001
ANOVA for Model Fit						
Source	**Sum of Squares**		**Df**	**Mean Square**	**F**	**p-Value**
Regression	33.285		1	33.825	230.85	< 0.001
Residual	72.502		498	0.146		
Total	106.327		499			

Source: Primary Data

The R^2 value of 0.317 indicated that 31.7% of the variance in trust is explained by interpretability. The high F-value (230.85) and its significance ($p < 0.001$) had confirmed that interpretability has a meaningful influence on users' trust for human-AI collaboration. Interpretability is a significant predictor of trust in human-AI collaboration ($\beta = 0.602$, $p < 0.001$). A positive β coefficient suggested that an increase in interpretability is associated with a proportional increase in trust. For every unit increase in interpretability, trust increases by 0.602 units. The ANOVA statistics had confirmed the model's statistical significance ($F = 230.85$, $p < 0.001$), indicating that the regression model fits the data well. The statistics of the regression analysis lead into acceptance of Hypothesis 2 i.e. "Interpretability positively influences Trust in human-AI collaboration", showing that interpretability significantly and positively influences user trust in human-AI collaboration. Interpretability explains a substantial portion of the variance in trust, emphasizing its critical role in enhancing users' trust for AI systems human-AI collaboration.

Regression Analysis: Transparency Positively Influences Satisfaction with AI Systems: In order to analyse the impact of transparency over the users' satisfaction for AI systems linear regression analysis was performed, results are hereunder:

Table 5: Regression Analysis Statistics of Transparency Impact on Users' Satisfaction with AI Systems

Model Summary						
Model	**R**	**R²**	**Adj. R²**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Transparency	0.591	0.349	0.336	0.452	267.94	< 0.001
Regression Coefficients						
Variable		**β Coefficient**		**Std. Error**	**t-Value**	**Sig.**
Transparency		0.592		0.036	16.37	< 0.001
Constant		1.115		0.122	9.14	< 0.001
ANOVA for Model Fit						
Source	**Sum of Squares**		**Df**	**Mean Square**	**F**	**p-Value**
Regression	34.792		1	34.792	267.94	< 0.001
Residual	62.237		498	0.131		
Total	100.029		499			

Source: Primary Data

The R^2 value of 0.349 indicated that 34.9% of the variance in users' satisfaction with AI systems is explained by transparency. The F-value of 267.94 ($p < 0.001$) showed that the model is highly statistically significant. Transparency is a significant predictor of satisfaction with AI systems ($\beta = 0.592$, $p < 0.001$). A positive β coefficient suggested that higher transparency leads to greater satisfaction. Specifically, for every unit increase in perceived transparency, satisfaction increases by 0.592 units. The ANOVA statistics had confirmed the model's statistical significance ($F = 267.94$, $p < 0.001$), indicating that the regression model fits the data well. The statistics of the regression analysis lead into acceptance of Hypothesis 3 i.e. "Transparency positively influences Satisfaction with AI systems", showing that transparency significantly and positively influences users' satisfaction in AI systems. Transparency had explained a substantial portion of the variance in trust, emphasizing its critical role in enhancing users' satisfaction for AI systems.

Regression Analysis: Satisfaction Mediates the Relationship between Transparency and Trust in Human-AI Collaboration: To assess mediation, a stepwise regression analysis following Baron and Kenny's (1986) framework was performed, which involves three steps:

- Regress Transparency on Trust (Direct Effect).
- Regress Transparency on Satisfaction.
- Regress both Transparency and Satisfaction on Trust (Testing Mediation).

Further, the Sobel test was also conducted to confirm the mediation effect statistically.

Table 6: Regression Analysis Statistics of Measuring the Mediating Role of Satisfaction for Relationship between Transparency and Trust in Human-AI Collaboration

Model Summary (Transparency → Trust)						
Model	**R**	**R^2**	**Adj. R^2**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Transparency	0.512	0.262	0.260	0.471	113.83	< 0.001
Model Summary (Transparency → Satisfaction)						
Model	**R**	**R^2**	**Adj. R^2**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Transparency	0.592	0.350	0.349	0.452	267.94	< 0.001
Model Summary (Transparency and Satisfaction → Trust (Testing Mediation))						
Model	**R**	**R^2**	**Adj. R^2**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Transparency + Satisfaction	0.648	0.420	0.418	0.430	179.88	< 0.001
Regression Coefficients (Transparency → Trust)						
Variable		**β Coefficient**		**Std. Error**	**t-Value**	**Sig.**
Transparency		0.512		0.048	10.67	< 0.001
Constant		1.245		0.153	8.14	< 0.001
Regression Coefficients (Transparency → Satisfaction)						
Variable		**β Coefficient**		**Std. Error**	**t-Value**	**Sig.**
Transparency		0.592		0.036	16.37	< 0.001
Constant		1.115		0.122	9.14	< 0.001
Regression Coefficients (Transparency & Satisfaction → Trust)						
Variable		**β Coefficient**		**Std. Error**	**t-Value**	**Sig.**
Transparency		0.289		0.051	5.67	< 0.001
Satisfaction		0.376		0.044	8.55	< 0.001
Constant		0.872		0.145	6.01	< 0.001

Source: Primary Data

Table 7: Sobel Test Statistics - Summary of Mediation Analysis

Path	**Effect Type**	**β Coefficient**	**p-Value**
Transparency → Trust	Direct Effect	0.512	< 0.001
Transparency → Satisfaction	Direct Effect	0.592	< 0.001
Transparency → Satisfaction → Trust	Indirect Effect	0.223	< 0.001

Source: Primary Data

The Transparency → Trust (Direct Effect) model evaluated the direct influence of transparency on trust in human-AI collaboration. With an R-value of 0.512, the model indicated a moderate positive correlation between transparency and trust. The R^2 value of 0.262 suggested that transparency accounts for 26.2% of the variance in users' trust. The F-statistic of 113.83 ($p < 0.001$) indicated that the model is statistically significant, confirming the predictive relevance of transparency. The regression coefficient ($\beta = 0.512$, $p < 0.001$) demonstrates that transparency has a substantial and positive impact on trust.

Further, the Transparency → Satisfaction model examined the effect of transparency on users' satisfaction. The R-value of 0.592 suggested a stronger positive correlation compared to the transparency-to-trust model. An R^2 of 0.350 indicated that 35% of the variance in satisfaction is explained by transparency, highlighting its importance in user satisfaction with AI systems. The F-value of 267.94 ($p < 0.001$) further supported the model's significance. Transparency's regression coefficient ($\beta = 0.592$, $p < 0.001$) had confirmed a significant positive influence, affirming that transparency directly enhances user satisfaction.

Transparency and Satisfaction → Trust (Testing Mediation) model incorporated satisfaction as a mediating variable between transparency and trust. The combined model achieved a higher R-value of 0.648, indicating a stronger correlation when satisfaction is included. An R^2 of 0.420 revealed that 42% of the variance in trust is explained by both transparency and satisfaction, demonstrating improved explanatory power. The F-value of 179.88 ($p < 0.001$) confirms the statistical significance of the model. Regression coefficients for transparency ($\beta = 0.289$, $p < 0.001$) and satisfaction ($\beta = 0.376$, $p < 0.001$) highlighted substantial contribution of both variables. While transparency continues to have a direct positive effect on trust, the inclusion of satisfaction significantly strengthens the model.

The mediation analysis revealed that transparency significantly impacts trust in AI systems both directly ($\beta = 0.512$, $p < 0.001$) and indirectly through satisfaction ($\beta = 0.223$, $p < 0.001$). Additionally, transparency strongly influences satisfaction ($\beta = 0.592$, $p < 0.001$), highlighting its role in enhancing user satisfaction. The significant indirect effect demonstrates that satisfaction partially mediates the relationship between transparency and trust, amplifying the overall impact of transparency on trust. These findings underscore the dual importance of transparency in directly fostering trust and indirectly strengthening it by ensuring user satisfaction, emphasizing the critical role of user-centric design in building trust in AI systems. Hence, Hypothesis 4 i.e. "Satisfaction mediates the relationship between Transparency and Trust in human-AI collaboration", is accepted.

Regression Analysis: Satisfaction Positively Influences Trust in Human-AI Collaboration: To statistically determine the impact of user satisfaction on users' trust in human-AI collaboration linear regression analysis was performed, results are hereunder:

Table 8: Regression Analysis Statistics of Impact of Users' Satisfaction on Users' Trust in Human-AI Collaboration

Model Summary						
Model	**R**	**R^2**	**Adj. R^2**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Satisfaction	0.642	0.412	0.401	0.434	348.63	< 0.001
Regression Coefficients						
Variable		**β Coefficient**		**Std. Error**	**t-Value**	**Sig.**
Satisfaction		0.642		0.034	18.67	< 0.001
Constant		1.029		0.123	8.37	< 0.001
ANOVA for Model Fit						
Source	**Sum of Squares**	**Df**		**Mean Square**	**F**	**p-Value**
Regression	65.932	1		65.932	348.63	< 0.001
Residual	93.768	498		0.188		
Total	159.700	499				

Source: Primary Data

The regression analysis of Impact of Users' Satisfaction on Users' Trust in Human-AI Collaboration demonstrated that satisfaction has a strong positive influence on trust in human-AI collaboration. The model shows an R-value of 0.642, indicates a strong correlation between satisfaction and trust. The R^2 value of 0.412 suggested that 41.2% of the variance in trust is explained by satisfaction alone, highlighting its significant contribution to trust-building. The F-value of 348.63 ($p < 0.001$) confirms the model's overall significance. The regression coefficient for satisfaction ($\beta = 0.642$, $p < 0.001$) indicates substantial and statistically significant positive impact on trust, with every unit increase in satisfaction leading to a corresponding 0.642 increase in trust. The ANOVA statistics confirmed that the regression model is highly significant, with an F-value of 348.63 and a p-value of < 0.001. This indicates that the variation in trust is significantly explained by satisfaction. The regression sum of squares (65.932) accounts for a substantial portion of the total variance (159.700), further supporting the strong relationship between satisfaction and trust in human-AI collaboration. These results strongly support Hypothesis 5 i.e. "Satisfaction positively influences Trust in human-AI collaboration", and confirming that higher satisfaction levels positively influence trust in AI systems.

Regression Analysis: Transparency and Interpretability's Impact on Users' Trust across Sectors: The analysis involves regression analyses for each sector (Healthcare, Finance, and Customer Service) followed with the comparison of the results to identify differences in the strength of the relationships results are hereunder:

Table 9: Regression Analysis Statistics of Transparency and Interpretability's Impact on Users' Trust across Sectors

Model Summary						
Sector	**R**	**R^2**	**Adj. R^2**	**Std. Err. of Estimate**	**F-Value**	**Sig.**
Healthcare	0.703	0.494	0.489	0.411	55.47	< 0.001
Finance	0.715	0.511	0.506	0.392	59.50	< 0.001
Customer Service	0.612	0.374	0.369	0.482	39.83	< 0.001

Regression Coefficients					
Sector	**Variable**	**β Coefficient**	**Std. Error**	**t-Value**	**Sig.**
Healthcare	Transparency	0.371	0.062	5.98	< 0.001
	Interpretability	0.425	0.058	7.33	< 0.001
	Constant	1.028	0.184	5.58	< 0.001
Finance	Transparency	0.396	0.058	6.83	< 0.001
	Interpretability	0.446	0.055	8.11	< 0.001
	Constant	0.947	0.174	5.44	< 0.001
Customer Service	Transparency	0.285	0.071	4.01	< 0.001
	Interpretability	0.328	0.067	4.90	< 0.001
	Constant	1.217	0.199	6.12	< 0.001

ANOVA for Model Fit						
Sector	**Source**	**Sum of Squares**	**df**	**Mean Square**	**F-Value**	**Sig.**
Healthcare	Regression	41.28	2	20.64	55.47	< 0.001
	Residual	60.15	162	0.371		
	Total	101.43	164			
Finance	Regression	43.21	2	21.61	59.50	< 0.001
	Residual	58.76	162	0.363		
	Total	101.97	164			
Customer Service	Regression	33.35	2	16.68	39.83	< 0.001
	Residual	68.10	162	0.420		
	Total	101.45	164			

Source: Primary Data

The regression analysis of Transparency and Interpretability's Impact on Users' Trust across Sectors demonstrated that the impact of transparency and interpretability on users' trust is stronger in healthcare and finance sector compared to customer service sector. The model summary revealed higher R^2 values for healthcare (0.494) and finance (0.511), indicating that 49.4% and 51.1% of the variance in trust, respectively, are explained by transparency and interpretability. In contrast, customer service showed a lower R^2 value of 0.374, explaining only 37.4% of the variance in trust. Furthermore, the adjusted R^2 and significant F-values (p<0.001) across all sectors confirmed the models' robustness, with healthcare and finance displaying better model fit compared to customer service. Regression coefficients indicated that both transparency and interpretability are significant predictors of users' trust in all three sectors. In healthcare, interpretability (β=0.425, p<0.001) has a slightly stronger influence on users' trust than transparency (β=0.371, p<0.001). Similarly, in finance, interpretability

(β=0.446, p<0.001) has stronger effect than transparency (β=0.396, p<0.001). However, in customer service, the effects of transparency (β=0.285, p<0.001) and interpretability (β=0.328, p<0.001) are weaker, indicating reduced influence compared to the other sectors. The ANOVA results supported the acceptance of Hypothesis 6 i.e. "The impact of Transparency and Interpretability on Trust is stronger in healthcare and finance than in customer service", which posits that the impact of transparency and interpretability on users' trust is stronger in healthcare and finance sector than in customer service sector. The F-values for healthcare (F=55.47, p<0.001) and finance (F=59.50, p<0.001) are significantly higher than for users' of customer service (F=39.83, p<0.001) sector, and indicates better model fit and stronger relationships in the healthcare and finance sectors.

Regression Analysis: Satisfaction's Impact on Users' Trust across Sectors: The analysis involves regression analyses for each sector (Healthcare, Finance, and Customer Service) followed with the comparison of the results to identify differences in the strength of impact of satisfaction on user' trust across sectors, results are hereunder:

Table 10: Regression Analysis Statistics of Satisfaction's Impact on Users' Trust across Sectors

Model Summary

Sector	R	R^2	Adj. R^2	Std. Err. of Estimate	F-Value	Sig.
Healthcare	0.567	0.321	0.317	0.488	75.97	< 0.001
Finance	0.584	0.341	0.338	0.472	84.10	< 0.001
Customer Service	0.653	0.426	0.423	0.430	119.80	< 0.001

Regression Coefficients

Sector	Variable	β Coefficient	Std. Error	t-Value	Sig.
Healthcare	Satisfaction	0.567	0.065	8.71	< 0.001
	Constant	1.024	0.151	6.78	< 0.001
Finance	Satisfaction	0.584	0.063	9.17	< 0.001
	Constant	0.988	0.145	6.82	< 0.001
Customer Service	Satisfaction	0.653	0.059	11.34	< 0.001
	Constant	0.905	0.139	6.51	< 0.001

ANOVA for Model Fit

Sector	Source	Sum of Squares	df	Mean Square	F-Value	Sig.
Healthcare	Regression	37.41	1	37.41	75.97	< 0.001
	Residual	78.85	163	0.484		
	Total	116.26	164			
Finance	Regression	39.69	1	39.69	84.10	< 0.001
	Residual	76.58	163	0.470		
	Total	116.27	164			
Customer Service	Regression	49.58	1	49.58	119.80	< 0.001
	Residual	66.74	163	0.410		
	Total	116.32	164			

Source: Primary Data

The regression analysis of users' satisfaction's impact on their trust across sectors (Healthcare, Finance, and Customer Service) demonstrated that satisfaction has a stronger influence on trust in the customer service sector compared to healthcare and finance. The model summary revealed the highest R^2 value for customer service (0.426), indicating that 42.6% of the variance in trust is explained by satisfaction. In comparison, the R^2 values for healthcare (0.321) and finance (0.341) are lower, explaining only 32.1% and 34.1% of the variance in trust, respectively. These findings suggests that satisfaction plays a more significant role in shaping trust of users working in customer service sector. The adjusted R^2 values and significant F-values (p<0.001) across all sectors confirm the robustness of the models, with customer service showing the best fit, followed by finance and healthcare. The regression coefficients further emphasize the stronger impact of satisfaction on trust in customer service sector, where the β coefficient is 0.653, compared to healthcare (β = 0.567) and finance (β = 0.584) sector. This suggests that for users working in customer service sector, an increase in satisfaction leads to a greater improvement in the trust. The ANOVA results support the acceptance of Hypothesis 7 i.e. "Satisfaction has a stronger influence on Trust in customer service compared to healthcare and finance", which posits that satisfaction has a stronger influence on trust of users of customer service sector than to the users of healthcare and finance sectors. The F-values for customer service (F = 119.80, p<0.001) are significantly higher than those for healthcare (F = 75.97, p<0.001) and finance (F = 84.10, p<0.001), indicating a stronger relationship between satisfaction and trust for the users of customer service sector.

CONCLUSION AND CONTRIBUTION

This research had provided a comprehensive exploration of the dynamics of Transparency, Interpretability, Satisfaction, and Trust in human-AI collaboration, focusing on three key sectors: healthcare, finance, and customer service. The findings are closely aligned with the research objectives, offering a nuanced understanding of trust-building mechanisms and their sector-specific variations. Transparency had been emerged as a critical factor positively influencing trust, with a direct effect size of β = 0.512 ($p < 0.001$). This underscores the importance of clear, ethical, and comprehensible AI processes in fostering user trust. Sector-specific analysis revealed that the influence of transparency on trust is more pronounced in healthcare and finance, evidenced by higher R^2 values (healthcare = 0.494, finance = 0.511) compared to customer service (R^2 = 0.374). Similarly, interpretability significantly enhances trust, with a direct effect size of β = 0.602 ($p < 0.001$). The impact of interpretability is particularly strong in healthcare and finance, where comprehensibility of AI outputs is essential for decision-critical tasks.

Satisfaction plays a pivotal role in mediating the relationship between transparency and trust, amplifying the influence of transparency on trust (direct effect = β = 0.512, indirect effect = β = 0.223). This highlights the importance of satisfaction as a key construct in building trust in AI systems. Furthermore, sector-specific differences in trust dynamics were observed. Transparency and interpretability had a stronger impact on trust in the high-stakes contexts of healthcare and finance, while satisfaction was a more significant predictor of trust in customer-facing applications like customer service (customer service β = 0.653; healthcare β = 0.567; finance β = 0.584). These findings underscore the need for tailored strategies to enhance trust based on sector-specific requirements.

This research had made a significant theoretical and practical contributions. It had enriched existing trust frameworks by incorporating sector-specific insights into roles of transparency, interpretability, and satisfaction in human-AI collaboration. These findings provide actionable recommendations for AI system designers and policymakers, emphasizing the prioritization of transparency and interpretability in healthcare and finance, while focusing on user satisfaction in customer service. Moreover, the study highlighted the need for customized trust-building strategies that align with the operational, ethical, and decision-critical requirements of different sectors. By addressing its objectives and validating the hypotheses, this research had advanced understanding of trust dynamics in human-AI collaboration and provides a robust foundation for future investigations and practical applications in AI system design and deployment.

REFERENCES

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., & Rahwan, I. (2018). The moral machine experiment. Nature, 563(7729), 59-64.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173–1182.

Benjamins, R., & Florez, J. L. (2020). Towards a European strategy on AI transparency and explainability. AI and Ethics, 1(3), 279-285.

Benjamins, R., & Florez, P. (2020). Towards a trustworthy AI framework: Foundations and challenges. AI Ethics and Governance Journal, 4(2), 156-168.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

Chen, C., Zhang, Y., & Liu, Y. (2021). Transparency in AI systems: A review of definitions, frameworks, and future directions. Journal of Artificial Intelligence Research, 71, 1-20.

Chen, Y., Korfiatis, N., & Kou, G. (2021). Trust development in human-AI interaction: A socio-cognitive perspective. Information & Management, 58(1), 103285.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155-159.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2017). Designing and conducting mixed methods research (3rd ed.). Sage.

Ferrario, A., Loi, M., & Viganò, E. (2019). In AI we trust incrementally: a multi-dimensional approach. Philosophy & Technology, 32(3), 457-477.

Ferrario, M. A., Simm, W., & Whittle, J. (2019). Empirical ethics in HCI: The role of empathy in designing for trustworthiness. Human-Computer Interaction, 34(3), 1-23.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Survey methodology (2nd ed.). John Wiley & Sons.

King, N. (2012). Doing template analysis. In G. Symon & C. Cassell (Eds.), Qualitative organizational research: Core methods and current challenges (pp. 426-450). SAGE Publications.

Lakkaraju, H., Bach, S. H., & Leskovec, J. (2017). Interpretable decision sets: A joint framework for description and prediction. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1675-1684.

Lee, J., Shin, D., & Yang, Y. (2021). AI trust in human-AI collaborative interaction: A comparative study of the antecedents of trust in AI and humans. Journal of Business Research, 123, 144-153.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16(1), 1-13.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137-141.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. ACM Transactions on Management Information Systems, 9(3), 1-32.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53-55.

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. Nature Machine Intelligence, 2(1), 10-12.