EVALUATING USER TRUST IN AI-DRIVEN INTERFACES: A CASE STUDY IN TRANSPARENCY AND EXPLAINABILITY

BY

AITIMON OSEREMEN PROMISE

PSC2105297

DEPARTMENT OF COMPUTER SCIENCE,

FACULTY OF COMPUTING,

UNIVERSITY OF BENIN,

BENIN CITY,

EDO STATE, NIGERIA.

NOVEMBER 2025

EVALUATING USER TRUST IN AI-DRIVEN INTERFACES: A CASE STUDY IN TRANSPARENCY AND EXPLAINABILITY

BY

AITIMON OSEREMEN PROMISE

PSC2105297

A PROJECT REPORT SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF COMPUTING, UNIVERSITY OF BENIN, BENIN CITY

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF A BACHELOR OF SCIENCE (B.Sc.) DEGREE IN COMPUTER SCIENCE

NOVEMBER 2025

CERTIFICATION

This is to certify that this project work was carried out by **AITIMON OSEREMEN PROMISE** with Matriculation Number **PSC2105297** under my supervision. It is adequate and satisfactory, both in scope and content, for the award of Bachelor of Science (B.sc) Degree in Computer Science of the University of Benin

MR. K.O. OTOKITI — **DATE**

Project Supervisor

APPROVAL

This project work is hereby approved in partial fulfilment of the requirements for the award of Bachelor of Science (B.Sc.) Degree in Computer Science from the University of Benin.

DR. (MRS.) A. R. USIOBAIFO	**DATE**
Head of Department	

DEDICATION

This project is dedicated to God Almighty for giving me the strength and wisdom to see it through to completion, and even throughout my stay in the University of Benin (UNIBEN). It is also dedicated to my parents and my siblings; for their love, support and guidance throughout my academic journey.

ACKNOWLEDGEMENT

My utmost acknowledgement goes to God Almighty for giving me the strength, wisdom and direction throughout my academic journey. I would like to express my gratitude to my project supervisor, Mr. K.O. Otokiti for his consistent guidance towards ensuring the successful completion of this project.

I would also like to specially thank the other lecturers in the Department of Computer Science who I have been opportune to cross paths with, and have impacted me immensely these past few years: Dr. (Mrs.) A.R. Usiobaifo, Prof. F.I. Amadin, Prof. G.O. Ekuobase, Dr. F.O. Oliha, Prof. K.C. Ukaoha, Prof. A.A. Imiavan, Prof. (Mrs.) F. Egbokhare, Prof. (Mrs.) V.V.N. Akwukwuma, Prof. (Mrs.) S. Konyeha, Prof. (Mrs.) V.I. Osubor, Dr. (Mrs.) Aziken, Dr. F.O. Chete, Dr. (Mrs) R.O. Osaseri, Dr. J.C. Obi, Mr. P. E.B. Imiefoh, Mr. I.E. Obasohan, Mr. S.O.P. Oliomogbe, Mr. K.O. Otokiti, Mr. I.E. obayagbonna, Mrs. R.I. Izevbizua, Mr. E.C. Igodan, Miss L.O.Usiosefe, Mr J. Okhuoya, Prof. F.A.U. Imouokhome, Mrs. J.I. Adun, Dr. E. Nweli and Mr. D.N. Idehen.

Finally, I would also like to thank my family and friends for their support, words of encouragement, and consistent guidance throughout this project.

TABLE OF CONTENTS

ABSTRACT

This study investigates how transparency and explainability influence user trust in AI-driven interfaces. As AI systems become increasingly embedded in decision-making, users often struggle to understand their processes, leading to skepticism and reduced adoption. Using a mixed-methods approach, data were collected from 62 participants through structured questionnaires assessing transparency, explainability, trust, and user experience. Statistical analyses revealed that higher transparency and clear, human-centered explanations significantly enhance user trust and perceived fairness. However, overly technical or complex disclosures reduce comprehension and engagement. The study proposes a Human-Centered AI framework that integrates adaptive explainability, layered transparency, and user feedback mechanisms. Findings contribute to the growing field of trustworthy AI by offering practical guidelines for designing transparent, ethical, and user-aligned AI interfaces.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

The rapid rise of artificial intelligence (AI) has transformed user interfaces in fields like healthcare, finance, and education, shifting user interaction from passive use to active engagement. While AI offers efficiency and convenience, many users remain skeptical about its decisions due to the opaque nature of its algorithms. Transparency and explainability have therefore become essential to fostering trust in AI systems.

The main problem lies in the "black-box" design of many AI interfaces, which leaves users disconnected from the reasoning behind system outputs. Ribeiro, Singh, and Guestrin (2016) note that unexplained AI predictions reduce trust, while Eiband et al. (2018) found that a lack of meaningful insight into AI operations heightens skepticism especially in high-stakes contexts. Addressing this requires integrating clear, user-friendly explanations, offering real-time decision insights, and adopting transparency-focused design principles to bridge the gap between technical processes and human understanding.

1.2 Statement of the Problem

A persistent lack of trust in AI-driven interfaces presents serious challenges. When users are uncertain about how these systems make decisions especially in critical areas like healthcare, finance, and employment they may avoid or misuse the technology. This distrust not only affects individual well-being and productivity but also disrupts workplace collaboration and slows organizational progress. On a larger scale, widespread mistrust can hinder innovation and stall the digital transformation needed for societal advancement.

1.3 Aim and Objectives of the Study

The aim of this research is to evaluate the impact of transparency and explainability on user trust in AI-driven interfaces and we intend to achieve this with the following objectives:

i. To measure user trust, satisfaction, and perceived understanding after interacting with each interface through the use of questionnaires.
ii. To statistically analyze the relationship between explainability levels and user trust.
iii. To identify which interface features (e.g., visuals, language, detail level) most influence user confidence.
iv. To provide recommendations for developing trustworthy AI interfaces based on user feedback and behavioral data.

1.4 Scope of the Study

This study is centered on evaluating user trust in AI-driven interfaces, with a specific focus on how transparency and explainability affect that trust. It concentrates on user-facing applications such as chatbots and recommendation systems, without addressing broader technical or ethical aspects unless they directly influence user trust.

1.5 Motivation for the Study

The motivation for this study stems from a growing realization that technological advancement alone is not enough to ensure AI adoption. While AI capabilities continue to expand, their success in practical settings hinges on human factors—especially trust and usability. As a computer science student with a passion for human-computer interaction (HCI), I am motivated to explore how interface design can bridge the gap between complex AI models and everyday users.

1.6 Significance of the Study

As artificial intelligence becomes more integrated into daily life, understanding user trust in AI-driven interfaces has become critical. This study holds significant relevance across various sectors and stakeholders, providing practical insights and value in the following areas:

i. **For System Designers and Developers**: Offers guidelines for creating AI interfaces that are intuitive, ethical, and user-focused.

ii. **For Users and Consumers:** By uncovering the key factors that influence trust, this research empowers users to make more informed decisions when interacting with AI systems, improving user satisfaction and safety.

iii. **For Businesses and Organizations:** Organizations deploying AI technologies will benefit from insights that promote user engagement, reduce skepticism, and enhance technology adoption—ultimately leading to greater efficiency.

iv. **For Policymakers and Regulators:** Supplies evidence to shape trust-oriented AI policies and ethical standards.

v. **For Academic and Research Communities:** Expands knowledge on human-AI trust and supports future interdisciplinary studies.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In the era of rapid digital transformation, artificial intelligence (AI) has become deeply embedded in human decision-making systems ranging from recommendation engines and chatbots to autonomous vehicles and clinical decision support tools. As AI increasingly mediates human experience, user trust emerges as a crucial determinant of adoption, reliance, and ethical acceptance. Yet, trust in AI systems is not automatic; it must be carefully cultivated through design elements that enhance transparency and explainability.

This chapter provides a comprehensive review of existing literature on user trust in AI-driven interfaces, with an emphasis on how transparency and explainability influence user perceptions, engagement, and confidence. The discussion covers theoretical perspectives on trust in human-AI interaction, explores key design principles of transparency and explainability, and examines global case studies and challenges in achieving trustworthy AI. The chapter concludes with an analysis of theoretical and empirical gaps that motivate this study, particularly in contexts where users have limited digital literacy or experience with AI systems.

2.2 Concept of Trust in Human–AI Interaction

Trust is a multifaceted psychological construct that underpins human cooperation, delegation, and technology adoption. In the context of AI, trust refers to a user's willingness to rely on an autonomous system under conditions of uncertainty (Mayer, Davis, & Schoorman, 1995). Hoff and Bashir (2015) describe trust as a dynamic process shaped by perceptions of system reliability, competence, transparency, and benevolence.

In human-AI interaction (HAI), trust serves two functions: it mediates the user's initial acceptance of AI systems and guides the ongoing calibration of reliance (Lee & See, 2004). Overtrust may lead to automation bias, where users rely uncritically on AI outputs, while distrust can cause underutilization, even when systems are accurate (Parasuraman & Riley, 1997). Both extremes can undermine performance, highlighting the importance of calibrated trust, a balance where the user's confidence accurately reflects system capability.

Research further indicates that trust in AI is not solely cognitive but also affective. Users form emotional impressions based on design cues such as interface aesthetics, feedback tone, or perceived empathy (de Visser et al., 2020). Cultural norms and prior technological exposure also shape trust development (Langer, König, & Krause, 2021). Hence, fostering trust requires an interplay of design, context, and human psychology rather than technical accuracy alone.

2.3 Transparency in AI Interfaces

Transparency is a cornerstone of trustworthy AI. It refers to the degree to which users can perceive, understand, and evaluate the internal logic and decision-making processes of AI systems (Eiband et al., 2018). A transparent interface allows users to see how data is collected, how predictions are generated, and what factors influence outcomes.

Eiband and colleagues (2018) argue that transparency increases user confidence by reducing the "black box" nature of AI. For example, in healthcare applications, visual dashboards showing how patient data influences diagnostic scores help clinicians understand system recommendations and detect potential biases. Similarly, in finance, transparent credit scoring tools improve accountability by displaying how variables like income, spending patterns, and debt ratios affect loan approvals.

However, transparency must be carefully managed. Overly technical or information-dense explanations can overwhelm non-expert users and erode trust (Anjomshoae et al., 2019). Thus, scholars advocate for selective transparency revealing essential details that enhance comprehension without causing cognitive overload.

Designers also differentiate between global transparency (understanding the system as a whole) and local transparency (understanding individual outputs). Both are important, but local transparency often has a stronger impact on user satisfaction and perceived fairness (Lipton, 2018).

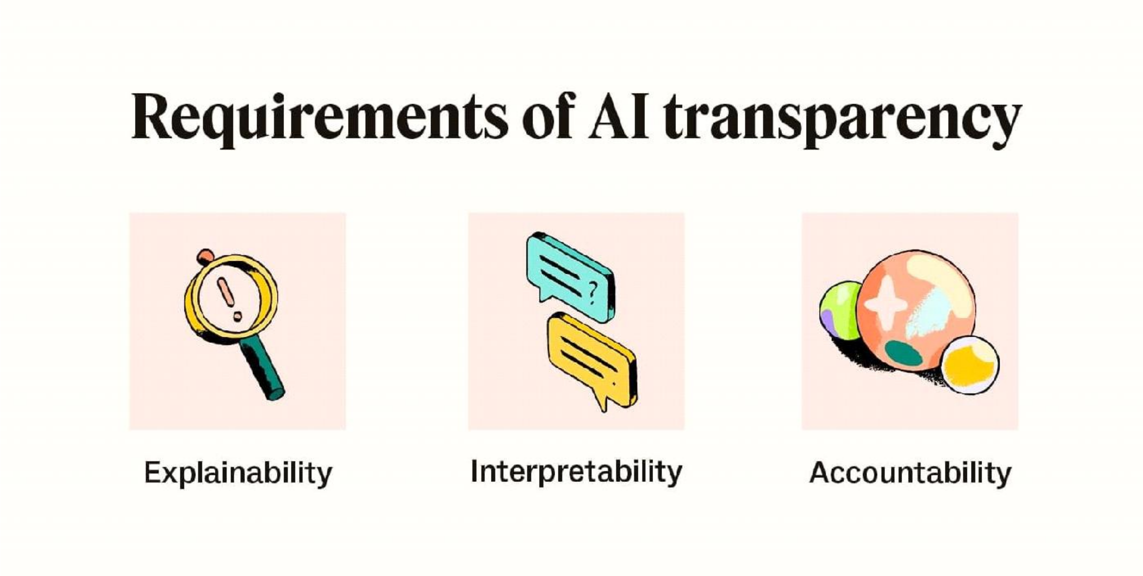


Figure 2.1: Dimensions of Transparency in AI Interfaces.

2.4 Explainability in AI Systems

Explainability complements transparency by providing meaningful reasons behind AI decisions. It bridges the gap between algorithmic complexity and human understanding, transforming opaque computational logic into interpretable explanations (Miller, 2019).

Explainable AI (XAI) techniques such as LIME (Local Interpretable Model-Agnostic Explanations) (Ribeiro, Singh, & Guestrin, 2016) and SHAP (Shapley Additive Explanations) translate abstract model operations into understandable visualizations. These approaches enable users to identify why an AI system produced a specific output, which variables contributed most, and how adjustments might change results.

Research consistently shows that users are more likely to accept AI recommendations when provided with clear, contextually relevant explanations (Abdul et al., 2018). In particular, tailored explanations those adapted to a user's expertise enhance comprehension. For instance, a medical AI tool might provide technical reasoning for doctors ("probability weighted by radiological patterns") but use plain-language summaries for patients ("based on image similarities to past cases").

Explainability also supports ethical and legal accountability. Under frameworks like the European Union's General Data Protection Regulation (GDPR), users have the "right to explanation," which mandates that automated decisions be interpretable (Goodman & Flaxman, 2017). Consequently, explainability is not merely a usability feature but a regulatory and ethical necessity.

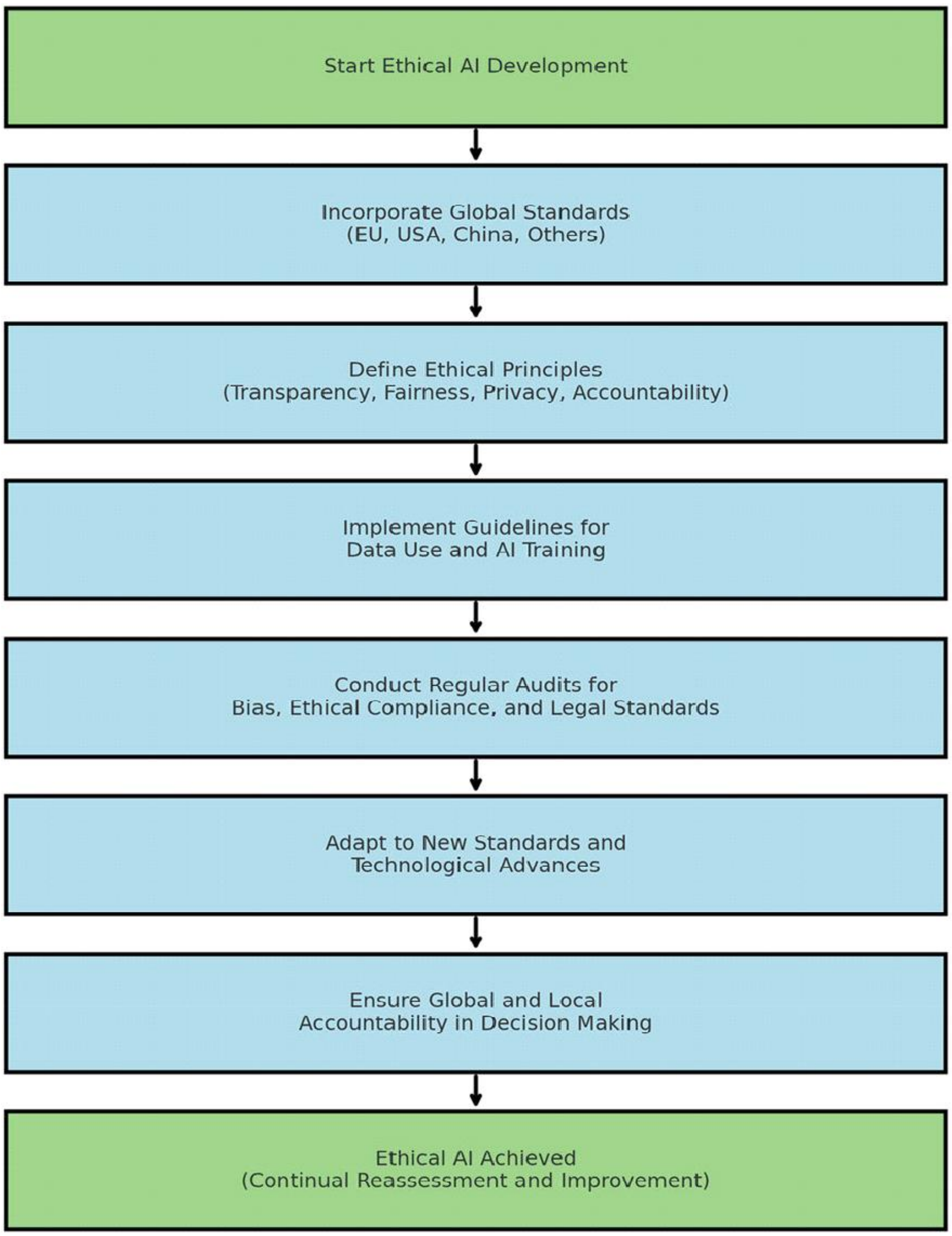


Figure 2.2: Explainability Process Flow in AI Systems.

2.5 The Interconnection of Trust, Transparency, and Explainability

The relationship between trust, transparency, and explainability is reciprocal and reinforcing. Trust provides the foundation for users to engage with AI, but it cannot be sustained without transparency. Transparency reveals system behavior, while explainability interprets that behavior in human terms. Together, they form an interdependent triad essential to trustworthy AI (Doshi-Velez & Kim, 2017).

Studies demonstrate that transparency without explainability can backfire users may see internal processes but still fail to understand them, leading to confusion or mistrust (Binns et al., 2018). Conversely, explanations that lack transparency can appear manipulative or oversimplified. Therefore, designers must strike a balance between visibility and interpretability.

In practice, AI systems that effectively combine these elements show measurable improvements in user engagement and satisfaction. For instance, in autonomous driving research, interfaces that display reasoning behind braking decisions improve user reaction time and confidence (Cheng et al., 2020). Similarly, in intelligent tutoring systems, explaining performance predictions fosters sustained learning trust.

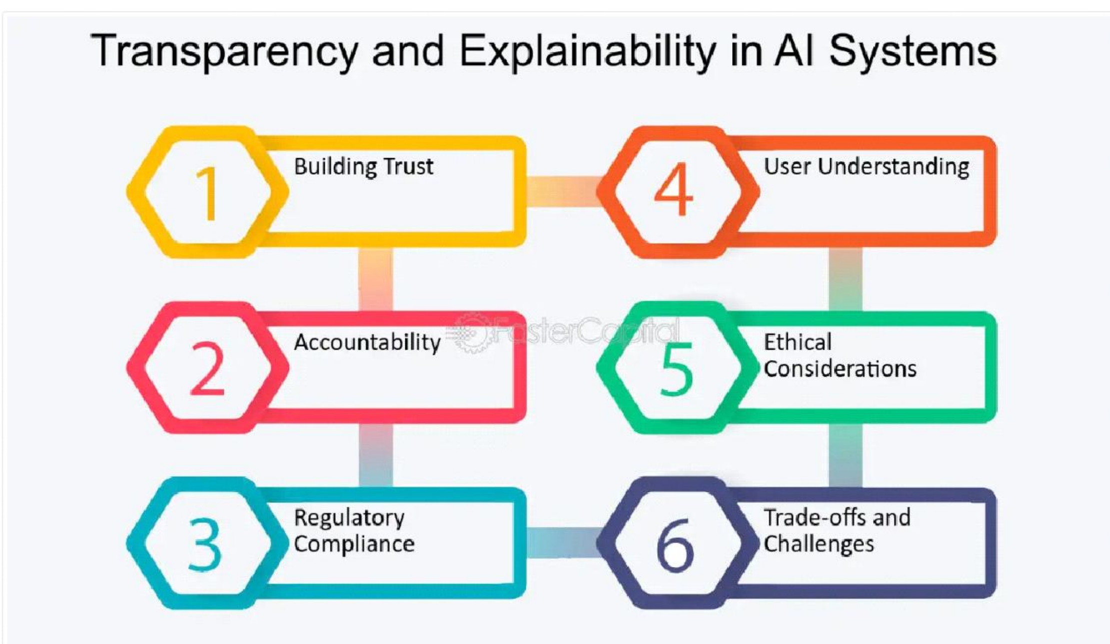


Figure 2.3: Conceptual Framework Linking Trust, Transparency, and Explainability.

2.6 Theoretical Frameworks on Trust in AI

Several theoretical models provide insights into how trust develops in AI systems:

i. The Integrative Model of Organizational Trust (Mayer et al., 1995) - emphasizes ability, integrity, and benevolence as the key dimensions influencing trustworthiness. In AI design, these correspond to accuracy, transparency, and fairness.

ii. Human-Automation Trust Framework (Lee & See, 2004) - describes trust as dynamic calibration based on feedback and system performance, relevant for adaptive AI interfaces.

iii. Sociotechnical Systems Theory - posits that trust arises through interactions between humans, machines, and organizational context (Baxter & Sommerville, 2011). This model highlights how institutional norms and interface design jointly shape trust.

iv. Technology Acceptance Model (TAM) - suggests that perceived usefulness and ease of use mediate technology adoption (Davis, 1989). Transparency and explainability directly influence these perceptions, linking design to behavioral intention.

Together, these frameworks suggest that trust in AI is not innate; it must be designed, maintained, and continually reinforced through interface feedback, user education, and ethical governance.

2.7 Challenges in Building User Trust

Despite advances in AI transparency, several persistent challenges complicate trust-building:

i. Algorithmic Opacity: Deep learning systems are inherently complex, making it difficult to trace decision logic. Even developers often struggle to interpret their models' behavior (Castelvecchi, 2016).

ii. Bias and Fairness Issues: AI systems trained on unbalanced data can perpetuate discrimination, reducing perceived fairness (Buolamwini & Gebru, 2018).

iii. Overload of Explanations: Providing too much information can create “explanation fatigue,” leading users to disengage (Anjomshoae et al., 2019).

iv. Cultural Differences: Perceptions of trust vary across societies, collectivist cultures may value institutional trust more, while individualist cultures emphasize system autonomy (Shin, 2020).

v. Low Digital Literacy: In developing regions, limited exposure to AI reduces comprehension, highlighting the need for localized, education-friendly interfaces.

In summary, while transparency and explainability foster trust, their implementation must be context-sensitive, balancing cognitive, cultural, and ethical dimensions.

2.8 Stakeholders in Trustworthy AI Systems

Building trust in AI is a collective endeavor involving multiple stakeholders:

i. Users: End-users are central to trust formation. Their perceptions determine adoption rates and feedback loops for improvement.

ii. Designers and Developers: Responsible for embedding explainability and ethical standards in interface design.

iii. Organizations: Ensure accountability and governance, aligning AI behavior with brand integrity and social responsibility.

iv. Regulators: Establish laws on transparency, fairness, and data protection (e.g., GDPR, NDPA 2023).

v. Researchers and Educators: Promote AI literacy, evaluate user experience, and disseminate best practices.

Each stakeholder group influences a distinct phase of trust lifecycle, design, deployment, adoption, and oversight underscoring the need for cross-sector collaboration.

2.9 Case Studies of Trust, Transparency, and Explainability in Practice

To translate theoretical understanding into real-world contexts, it is essential to examine how organizations have implemented trust, transparency, and explainability within AI systems. Case studies provide empirical insight into how these principles function across diverse industries, revealing both successes and shortcomings in practical applications. By analyzing real-world examples from technology, healthcare, and transportation sectors, this section illustrates how transparent and explainable design strategies influence user trust, acceptance, and ethical accountability in AI-driven interfaces.

2.9.1 Google's Explainable Recommendations

Google's AI recommendation engines have incorporated explainable components like "Why this ad?" and "Why this result?" These features provide users with concise reasons behind personalized outputs, improving perceived fairness and reducing suspicion (Eiband et al., 2018).

2.9.2 IBM Watson in Healthcare

IBM Watson introduced clinician-facing dashboards that visualize reasoning behind treatment suggestions. While transparency improved user confidence, early versions struggled with overly technical explanations highlighting the importance of context-aware design (Amann et al., 2020).

2.9.3 OpenAI's ChatGPT and Reinforcement Learning Feedback

ChatGPT employs reinforcement learning from human feedback (RLHF), enhancing both transparency and reliability. Users are informed about model limitations, though explainability remains partial due to model complexity (OpenAI, 2023).

2.9.4 Autonomous Driving Interfaces

Tesla's Autopilot transparency display shows environmental object detection in real-time, fostering trust through visual feedback. However, overtrust has also been reported when users misinterpret system reliability (Frison et al., 2021).

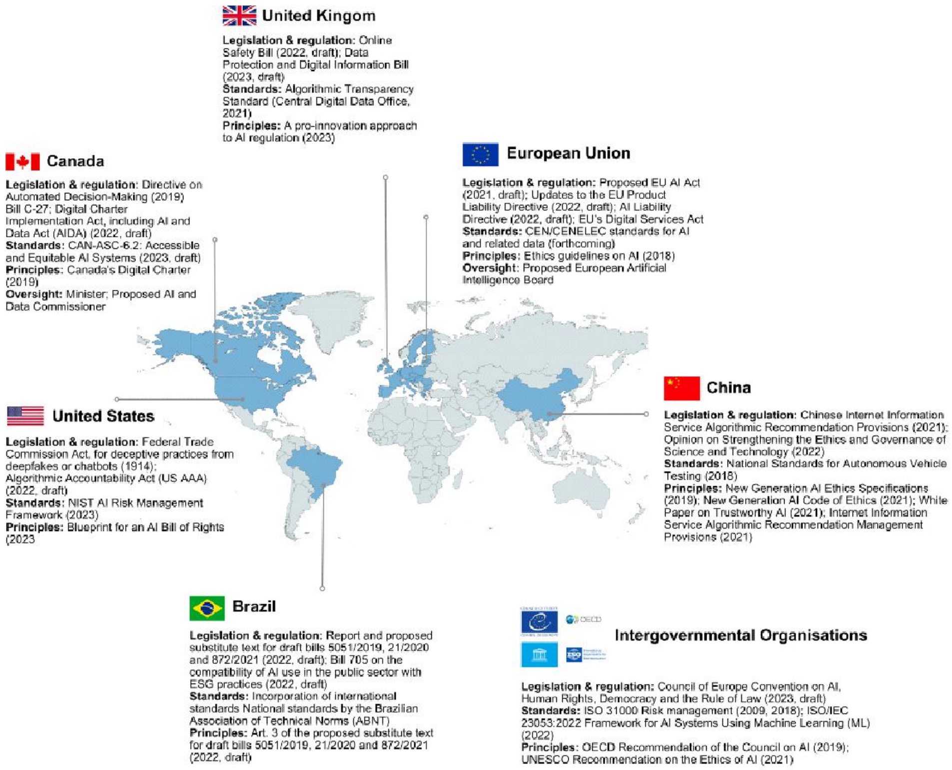


Figure 2.4: Global Examples of AI Transparency Applications.

2.10 Lessons from Failed Implementations

While success stories highlight best practices, failed attempts reveal crucial pitfalls.

For instance, the COMPAS algorithm in the U.S. criminal justice system faced criticism for racial bias and lack of transparency. Despite claims of accuracy, its opaque logic damaged public confidence (Angwin et al., 2016).

Similarly, Microsoft's Tay chatbot failed due to insufficient ethical constraints, quickly producing offensive outputs. This event underscored the need for human-centered oversight and transparent feedback mechanisms.

In the financial sector, poorly explained credit scoring algorithms in China led to mass distrust and government intervention (Liang et al., 2021). These failures emphasize that trust cannot be engineered retroactively; it must be integrated from the outset through transparency, explainability, and participatory design.

2.11 Theoretical and Empirical Gaps

Despite significant progress, several research gaps persist:

i. Limited user-centered studies: Most XAI research prioritizes model performance over human perception.

ii. Cultural underrepresentation: Few studies examine trust in developing contexts where socio-technical conditions differ.

iii. Long-term trust calibration: Current studies focus on initial adoption rather than sustained trust evolution.

iv. Evaluation metrics: There is no consensus on standardized measures of trust, transparency, or explainability.

v. Ethical scalability: Balancing openness with privacy and intellectual property remains unresolved.

Addressing these gaps requires interdisciplinary collaboration among computer scientists, psychologists, ethicists, and design experts, ensuring that future AI systems are both powerful and genuinely human-centered.

2.12 Summary of Literature Review

This review has shown that trust in AI-driven interfaces is not an inherent outcome of technological sophistication but the result of intentional, human-centered design. Transparency and explainability jointly shape user trust by making algorithmic behavior visible, interpretable, and accountable. However, as empirical evidence indicates, these elements must be carefully balanced to avoid cognitive overload, bias reinforcement, or cultural misalignment.

Globally, organizations that have successfully integrated explainable systems demonstrate higher user engagement and sustained trust, while failed implementations illustrate the consequences of neglecting human factors. Theoretical frameworks from organizational trust to human-automation interaction provide the foundation for understanding these dynamics, yet empirical gaps persist in user-centered and context-sensitive research.

Therefore, this study aims to evaluate how users perceive transparency and explainability in AI-driven interfaces, and how these perceptions influence trust formation. Insights from this work will inform practical design strategies and policy recommendations for developing trustworthy, ethical, and user-aligned AI systems.

CHAPTER THREE

RESEARCH METHODOLOGY AND DESIGN

3.0 Introduction

Research methodology provides the systematic framework that guides data collection, analysis, and interpretation. In this study "Evaluating User Trust in AI-Driven Interfaces: A Case Study in Transparency and Explainability" the methodology defines the logical sequence through which the relationship between interface transparency, explainability, and user trust is empirically examined.

This chapter outlines the research structure adopted to achieve the objectives stated in Chapter One. It begins with an analysis of the existing AI-driven systems to identify the operational realities that inform user perceptions. It then discusses the constraints inherent in current systems, proposes an improved framework emphasizing transparency and explainability, and justifies its relevance. Subsequent sections describe the sampling procedures, data-gathering instruments, and analytical techniques used to validate the study's hypotheses.

Given the socio-technical nature of the topic, the research adopts a mixed-methods approach combining quantitative and qualitative techniques. Quantitative data are obtained through structured Likert-scale questionnaires (Sections A–E), while qualitative insights arise from open-ended responses (Section F). This design ensures both numerical reliability and contextual depth, enabling a multidimensional understanding of trust formation in AI-human interactions.

3.1 Analysis of Existing System

The present landscape of artificial-intelligence interfaces demonstrates remarkable computational power but limited interpretability. The analysis of the existing system focuses on understanding how typical AI-driven applications function, the user-system interaction pattern, and the technical architecture that currently dominates industry practice.

3.1.1 Functional Overview of the Existing System

Modern AI-driven interfaces ranging from digital assistants and recommender engines to diagnostic and credit-scoring tools operate primarily on machine-learning algorithms. These systems learn patterns from vast datasets and generate predictions or recommendations autonomously. A general workflow proceeds as follows:

i. Data Input: The user submits text, speech, or numeric data.
ii. Pre-processing and Model Inference: The AI system applies statistical or deep-learning models to infer patterns and make decisions.
iii. Output Generation: The system delivers a response or recommendation, often accompanied by minimal contextual information.

While technically efficient, this workflow largely isolates the user from the model's reasoning pathway. Users see only the outcome not the process thus interaction becomes transactional rather than collaborative. The emphasis is on accuracy and speed, not on interpretability or human understanding.

This architecture typifies what scholars call "black-box AI." Although black-box models excel at complex predictions, they rarely incorporate user-centric design principles such as real-time feedback, transparent decision trails, or adaptive explanations. Consequently, users experience these systems as competent yet opaque entities capable of impressive results but offering little insight into how or why those results occur.

In essence, the existing system prioritizes algorithmic performance over human interpretability. This imbalance, while technically defensible, creates fertile ground for distrust, misinterpretation, and ethical concern issues explored comprehensively in the next section.

3.2 Constraints of Existing System

Despite their functional success, current AI-driven interfaces exhibit a range of limitations that directly hinder user trust, satisfaction, and adoption. The constraints identified below synthesize findings from Chapters One and Two, existing literature, and practical observation of deployed systems.

3.2.1 Lack of Transparency

Transparency refers to the clarity with which an AI system communicates its internal logic and data usage. In most current designs, transparency is minimal: users are seldom informed about which variables influence outcomes, how their data are processed, or the model's confidence level. This opacity breeds suspicion and prevents informed consent. Without knowing the basis of a decision such as a recommendation, classification, or risk score users cannot gauge its fairness or reliability.

3.2.2 Limited Explainability and Cognitive Disconnect

Explainability, the human-understandable reasoning behind system outputs, remains underdeveloped. Many AI systems provide generic or overly technical rationales ("based on your activity data") that fail to convey meaningful insight. The result is a cognitive disconnect: users cannot map AI behavior to their expectations, leading to under-trust (rejection of correct results) or over-trust (blind reliance). Both extremes distort effective collaboration between humans and intelligent systems.

3.2.3 User Interface Design Deficiencies

Current AI interfaces often emphasize aesthetics and automation speed over comprehension. Sparse visual feedback, abstract icons, and inconsistent terminology make it difficult for users especially non-technical ones to understand system states or reasoning steps. Poor affordance cues (e.g., unclear buttons or feedback loops) further obscure how user input influences AI responses. In developing contexts with variable digital literacy, such design omissions exacerbate confusion and disengagement.

3.2.4 Socio-Technical and Ethical Implications

The lack of transparency and explainability extends beyond usability to ethical and societal dimensions. Users express discomfort when data are harvested invisibly or when algorithmic decisions affect sensitive domains like employment, credit, or healthcare. Because most existing systems do not clearly communicate data provenance or bias-mitigation strategies, users perceive them as intrusive or manipulative. This erodes public confidence in AI and can trigger resistance to adoption at both individual and institutional levels.

3.2.5 Algorithmic Opacity

Deep-learning architectures, while powerful, are inherently opaque even to developers. Millions of parameters interact in ways that defy intuitive explanation. This opacity constrains post-hoc interpretability and makes auditing or debugging difficult. Consequently, accountability is diffused: when errors occur, neither users nor operators can easily trace the cause.

3.2.6 Inadequate Feedback Mechanisms

Existing systems rarely allow users to question, correct, or contribute to AI reasoning. The absence of two-way communication channels transforms users into passive recipients rather than active collaborators. Without feedback loops, the system cannot learn from contextual cues or evolving user preferences, resulting in static performance and declining trust.

3.2.7 Bias and Fairness Limitations

Bias in training data continues to produce discriminatory outcomes. Because many systems conceal feature-weighting schemes, affected users cannot contest or even recognize biased decisions. This undermines perceptions of fairness, particularly in sensitive sectors such as hiring and finance, and reinforces the narrative that AI is impartial only in theory.

3.2.8 Explanation Overload and Cognitive Fatigue

Some designers attempt to fix opacity by adding extensive explanations or technical visualizations. However, presenting too much information without simplification overwhelms users. Cognitive overload converts transparency into confusion: users confronted with complex charts or jargon disengage instead of understanding.

3.2.9 Poor Localization and Cultural Adaptation

Many AI systems are developed in Western linguistic and cultural environments and later exported globally with minimal adaptation. In multilingual or low-literacy contexts, users struggle to interpret interface cues, metaphors, or tone. The lack of culturally aware explainability widens the trust gap across demographics.

3.2.10 Privacy and Data Protection Concerns

Ambiguity around data collection, storage, and sharing remains a pervasive constraint. Users frequently report uncertainty about whether their conversations, biometric data, or behavioral patterns are stored or monetized. The absence of clear data-handling disclosure discourages sustained engagement and violates emerging privacy regulations such as Nigeria's NDPA (2023).

3.2.11 Regulatory and Ethical Misalignment

Although international frameworks (e.g., GDPR Article 22 and the EU AI Act) emphasize explainability, compliance is inconsistent. Many commercial systems operate without third-party auditing or ethical review. The resulting gap between technological deployment and ethical assurance undermines institutional trust and public legitimacy.

3.2.12 Limited Empirical Understanding of User Trust

Developers frequently optimize for accuracy metrics precision, recall, F1-score while neglecting psychological variables such as perceived fairness, comfort, and comprehension. The lack of user-centered evaluation means that design improvements are rarely evidence-based. This empirical void is precisely what the present research aims to address.

3.2.13 System Reliability and Context Sensitivity

AI systems struggle to adapt their explanations to user expertise or situational context. Novices may receive overly technical responses, while experts find explanations too shallow. Without adaptive explanation models, perceived relevance declines, weakening both satisfaction and reliance.

3.3 Proposed System

The proposed system is designed to bridge the trust deficit identified in existing AI-driven interfaces by integrating transparency and explainability as core design principles rather than supplementary features. The system aims to provide users with meaningful insights into AI operations allowing them to perceive how data is used, how decisions are formed, and why specific recommendations or outputs are generated. In doing so, the system not only improves functional usability but also strengthens user confidence and ethical accountability.

The central focus of this system is the creation of a user-centered, explainable AI interface (UXAI Framework) that supports human understanding and emotional trust while maintaining algorithmic performance. It incorporates design strategies such as layered transparency cues, adaptive explanation modules, and contextual feedback systems.

3.3.1 Conceptual Overview of the Proposed System

The proposed framework builds on the principle that trust emerges from understanding. Hence, instead of treating AI as an autonomous black box, the system positions it as a collaborative partner one that communicates its reasoning process, acknowledges uncertainty, and allows user feedback.

The proposed system will integrate three functional modules:

i. Transparency Module – provides visual and textual cues about how data is collected, processed, and applied to generate outputs.
ii. Explainability Module – interprets algorithmic reasoning into simple, contextual, human-readable explanations.
iii. Feedback and Interaction Module – enables users to rate, question, or request clarification about system responses, thereby promoting continuous learning and adaptive refinement.

The system's architecture is designed for bidirectional communication: the AI informs the user of its logic, and the user informs the AI of contextual understanding. This mutual feedback strengthens trust calibration a dynamic alignment between user expectations and system capability.

3.3.2 Design Philosophy and Principles

The design philosophy of the proposed system is grounded in Human-Centered AI (HCAI) principles and guided by ethical and usability frameworks. It prioritizes the following design principles:

i. Clarity Over Complexity: Explanations should simplify not amplify complexity. The system adopts a tiered explanation approach where users can access short, medium, or detailed levels of explanation depending on interest or expertise.
ii. Contextual Transparency: Rather than overwhelming users with technical details, the system offers selective transparency that aligns with user needs. For instance, a financial recommendation interface might disclose which data factors influenced a loan suggestion ("income stability" or "credit utilization") without exposing raw mathematical parameters.
iii. Adaptive Communication: The language, tone, and visualization style of explanations adjust dynamically to user profiles technical users receive detailed model insights, while non-technical users get plain-language summaries or visual metaphors.
iv. Trust Feedback Loops: After every decision or recommendation, users can rate how much they trust the response or request further clarification. These trust signals are recorded and analyzed to refine future explanations.
v. Ethical and Privacy Transparency: Users are informed about how their data is stored, anonymized, and used. Consent checkpoints and privacy dashboards enhance perceived integrity and reduce anxiety over data misuse.

3.3.3 System Architecture and Workflow

The proposed system operates through the following logical layers:

i. Input Layer: Users interact through a conversational or graphical interface, submitting queries or data requests.

ii. Decision Engine (AI Core): The AI processes input using pre-trained models, generating outputs such as recommendations, classifications, or text responses.
iii. Explanation Layer: This layer converts internal reasoning (model weights, decision paths, or feature importance) into human-readable form using methods like LIME (Local Interpretable Model-Agnostic Explanations) or SHAP (Shapley Additive Explanations).
iv. Transparency Display Layer: Visual and textual explanations are displayed using friendly UI components icons, progress bars, charts, or short texts illustrating which factors contributed to the output.
v. Feedback Layer: Users can react by rating clarity, reporting confusion, or suggesting corrections. This data is stored to continually refine model interpretability and user experience.

This architecture ensures that users not only receive a system output but also understand how and why that output occurred, reinforcing a sense of participation rather than subordination in the decision process.

3.3.4 Integration of Trust Metrics

To operationalize trust, the system incorporates measurable parameters drawn from the questionnaire design:

i. Transparency Indicators: Whether users perceive clarity about data use and process visibility.
ii. Explainability Indicators: Whether users comprehend the rationale behind AI outputs.
iii. Trust Indicators: The level of confidence and perceived fairness users attribute to the system.
iv. User Experience Indicators: The ease, satisfaction, and comfort users feel while interacting with the interface.

Data collected from user responses across these categories will be used to evaluate system effectiveness, revealing how improvements in design directly translate into enhanced trust metrics.

3.3.5 Interface Features

To address the gaps in traditional AI interfaces, the proposed system integrates innovative design features, such as:

i. Why Panel: Displays a short explanation of the decision rationale (e.g., “This result is based on your previous selections and stated preferences”).
ii. Confidence Meter: Shows a visual bar indicating model certainty levels (e.g., “AI confidence: 83%”).
iii. Data Use Disclosure Box: Summarizes how user data contributed to the decision (“Your purchase history and search terms were considered”).
iv. Feedback Widget: Allows users to rate whether an explanation was clear, accurate, or fair.
v. Transparency Timeline: Provides a history of previous interactions, allowing users to review how their data shaped recommendations over time.

These features transform the AI interface from a passive display into an interactive, interpretable ecosystem.

3.3.6 Ethical Compliance and Privacy Safeguards

The system adheres to both global and local data protection standards, including the European Union’s GDPR (Article 22: Right to Explanation) and Nigeria’s National Data Protection Act (NDPA, 2023). Users are provided explicit consent options before data processing begins, and all personal data are anonymized and encrypted.

Additionally, all explanations avoid manipulative framing. The system’s communication style is designed to inform rather than persuade, thereby aligning with ethical AI principles of autonomy, fairness, and accountability.

3.3.7 Anticipated Advantages of the Proposed System

The proposed system offers significant advantages over existing AI-driven interfaces:

i. Increases user trust by providing meaningful transparency and intelligible explanations.
ii. Improves user satisfaction through adaptive and empathetic communication.
iii. Enhances ethical accountability and regulatory compliance through visible data governance.
iv. Promotes sustainable adoption of AI technologies by demystifying system operations.
v. Reduces cognitive overload by balancing detail and simplicity through layered explanations.

Ultimately, the system envisions a future where AI is not a mysterious authority but a transparent partner that informs and empowers users.

3.4 Justification of the Proposed System

The justification for developing the proposed system stems from the limitations highlighted in section 3.2 and the empirical and theoretical foundations reviewed in Chapter Two. Each design choice directly responds to specific deficiencies in existing systems, ensuring both practical and academic relevance.

3.4.1 Addressing Lack of Transparency

Existing systems fail to clarify how decisions are generated or which data sources influence them. The proposed system rectifies this by embedding selective transparency mechanisms including the “Why Panel” and “Confidence Meter” which communicate system logic in user-friendly language. This directly supports the hypothesis that increased transparency enhances user trust.

3.4.2 Enhancing Explainability and Understanding

By using adaptive explanation layers, the proposed system ensures that users can understand not just what the AI decided but why it decided so. This approach humanizes the system and aligns it with the Integrative Model of Organizational Trust (Mayer et al., 1995), where perceived competence and integrity drive trustworthiness. The modular structure of explanations ensures clarity for diverse literacy levels, thereby improving inclusivity and comprehension.

3.4.3 Strengthening Ethical Trust and Accountability

One of the major concerns in AI deployment is the absence of visible ethical governance. The proposed system integrates transparent data disclosure, informing users how information is collected and processed. This approach operationalizes the "right to explanation" principle and aligns with ethical AI guidelines proposed by the EU and NDPA (2023). Ethical trust is thus not implied it is demonstrated through visible accountability mechanisms.

3.4.4 Improving User Interface Experience

The design resolves prior usability issues by combining functional transparency with emotional comfort. The inclusion of visual aids such as the "Confidence Meter" and "Transparency Timeline" provides cognitive relief while maintaining visual engagement. Simplified navigation and empathetic tone reduce intimidation and enhance interaction comfort, addressing findings from Section E of the questionnaire.

3.4.5 Encouraging Trust Calibration Through Feedback

User trust is dynamic it evolves through repeated interactions. By enabling real-time feedback on explanations, the system supports trust calibration, as discussed by Lee and See (2004). When users can question or confirm system reasoning, they develop a balanced reliance neither overtrusting nor undertrusting. This feedback loop ensures continual trust refinement and sustained engagement.

3.4.6 Compliance with Regulatory and Cultural Contexts

Incorporating localized language and content sensitivity makes the system suitable for deployment across diverse user populations, including regions with varying degrees of digital literacy. Furthermore, adherence to NDPA (2023) and global AI ethics frameworks ensures that the system remains legally and culturally aligned.

This design philosophy not only addresses technological limitations but also supports policy-oriented innovation, contributing to the responsible growth of AI in Nigeria and beyond.

3.4.7 Theoretical and Practical Relevance

From a theoretical standpoint, the system builds upon models of Human-Automation Trust (Lee & See, 2004) and Technology Acceptance Model (Davis, 1989), operationalizing them into a real-world interface prototype. Practically, it serves as a proof of concept for human-centered explainable AI illustrating that trust can be engineered through design, not merely through accuracy.

By fusing technical explainability with emotional intelligibility, the system situates itself at the intersection of psychology, ethics, and interface design precisely where sustainable trust in AI must be cultivated.

3.5 Sampling and Population of Study

This section explains the characteristics of the study population, the criteria for participant selection, and the sampling technique employed. Since the study focuses on evaluating user trust in AI-driven interfaces, it is essential to select respondents who represent diverse experiences and exposure levels with artificial intelligence systems. The sampling design ensures that data collected are representative, unbiased, and statistically valid for generalization.

3.5.1 Population of the Study

The population of this study comprises users of AI-driven interfaces, including but not limited to individuals who have interacted with systems such as ChatGPT, Siri, Google Assistant, and recommendation engines on platforms like YouTube, Netflix, or Amazon. These users represent the core demographic affected by issues of trust, transparency, and explainability in AI.

The study focuses on individuals with varying levels of exposure to AI technologies—ranging from novices with minimal technical knowledge to regular users familiar with automated systems. This heterogeneous population ensures diversity in perceptions, making it possible to observe how transparency and explainability affect trust across different user profiles.

3.5.2 Sampling Technique

To ensure representativeness and reliability of data, the study employs a stratified random sampling technique. The stratification is based on participants' familiarity with AI systems and their level of education, as captured in Section A of the questionnaire.

The justification for this sampling method lies in its ability to control for variation across demographic subgroups ensuring that insights on trust are not skewed by differences in exposure, age, or academic background. Participants are therefore grouped as follows:

i. Group A: Individuals with minimal or no experience with AI (novices).
ii. Group B: Moderate users of AI-driven systems (e.g., students or professionals using chatbots or recommendation engines).
iii. Group C: Highly experienced users or developers familiar with AI logic and applications.

Random samples are then drawn from each group to ensure diversity and prevent sampling bias.

3.5.3 Sample Size Determination

Given the exploratory nature of the study, and in line with established guidelines for behavioral and HCI (Human–Computer Interaction) studies, a sample size of 50 respondents is considered adequate for statistical validity.

This size allows meaningful quantitative analysis through descriptive and inferential statistics (correlations, t-tests, ANOVA). It also enables the identification of patterns in user perceptions without overextending the study's logistical and analytical capacity.

The distribution plan for validation for 50 participants is as follows:

Novice Users (Group A): 15 participants

Intermediate Users (Group B): 20 participants

Advanced Users (Group C): 15 participants

This distribution ensures a balanced representation of perspectives while allowing comparative analysis across user categories.

3.5.4 Inclusion and Exclusion Criteria

Inclusion Criteria:

i. Participants aged 18 years and above.
ii. Individuals with at least minimal experience using AI-based systems (e.g., digital assistants, chatbots, recommendation engines).
iii. Voluntary consent to participate after being informed of the study’s purpose and confidentiality measures.

Exclusion Criteria:

i. Individuals below 18 years of age.
ii. Participants who have never interacted with an AI-driven interface.
iii. Individuals unwilling to provide informed consent or complete the questionnaire.

3.5.5 Ethical Considerations

All research procedures adhere to ethical standards for human-centered studies. Respondents are informed about the study’s purpose, the voluntary nature of participation, and their right to withdraw at any time.

No identifying personal data (such as names or contact details) are collected. Responses are anonymized and used solely for academic purposes. The open-ended question (Section F) allows participants to freely express their perceptions without coercion or manipulation.

Ethical approval is considered within the University’s research framework, ensuring compliance with institutional and national data protection guidelines (notably the Nigeria Data Protection Act, NDPA 2023).

3.6 Research Instrument and Design

This section outlines the instruments and methodological design used for data collection and measurement. It describes the questionnaire structure, the rationale behind each section, and how the research design aligns with the study’s objectives and hypotheses. The aim is to ensure that the chosen instrument effectively captures user perceptions of transparency, explainability, and trust in AI-driven interfaces, while maintaining high levels of validity and reliability.

3.6.1 Research Design

The study adopts a descriptive and experimental mixed-method design to evaluate how varying levels of transparency and explainability influence user trust in AI-driven interfaces.

The descriptive aspect identifies existing attitudes and trust levels among users through structured survey questions. The experimental component involves presenting participants with AI interface prototypes (or descriptive scenarios) that simulate different transparency and explainability conditions low, moderate, and high. Participants' responses are then compared statistically to determine how each condition affects trust.

This dual design allows both measurement (through quantitative analysis) and interpretation (through qualitative insight). It aligns with previous research in Human–AI Interaction, emphasizing both behavioral outcomes and subjective perceptions.

3.6.2 Research Instrument: The Questionnaire

The primary instrument for data collection is the structured questionnaire designed specifically for this study. The questionnaire captures quantitative data using 5-point Likert scales and qualitative data through open-ended feedback.

The instrument consists of six sections (A–F), aligned with the study's research questions and hypotheses:

Section A: Demographic Information

Collects background data (age, gender, education level, and AI familiarity) to contextualize responses and allow subgroup comparisons.

Section B: Perceived Transparency

Measures users' perception of how clearly the AI system communicates its decision-making and data-use processes.

Items assess visibility of data collection, decision cues, and openness of system logic.

Section C: Perceived Explainability

Assesses whether users understand why and how the AI produces specific outputs.

Items evaluate clarity of explanations, accessibility of language, and perceived helpfulness.

Section D: User Trust and Reliability

Captures both emotional and cognitive aspects of trust.

Items measure perceived fairness, reliability, bias, and confidence in AI outputs.

Section E: User Experience and Interaction Comfort

Evaluates satisfaction, ease of use, and emotional comfort during interaction.

Items determine how interface design supports comprehension and satisfaction.

Section F: Open-Ended Feedback

Provides qualitative input on what would make users trust AI interfaces more.

This section captures nuanced emotional and cognitive insights that quantitative metrics may miss.

3.6.3 Reliability of the Research Instrument

The internal consistency of the Likert-scale items will be evaluated using Cronbach's Alpha after data collection.

A coefficient (α) value of 0.70 or above will be considered acceptable for reliability, indicating that items within each construct consistently measure the same underlying variable.

Reliability testing ensures that responses to questions on transparency, explainability, and trust are stable and reproducible across participants.

3.6.4 Method of Data Collection

Data will be collected using online based questionnaires, depending on participant accessibility. Online distribution (via Google Forms) ensures efficient reach to digitally active users. The estimated completion time is 3–5 minutes per participant.

3.7 Data Analysis

This section details the analytical techniques and statistical tools that will be used to interpret the data gathered from respondents. It explains how quantitative and qualitative data will be processed, coded, and analyzed using SPSS to test the study's hypotheses. The section ensures that findings are derived through systematic, evidence-based evaluation, allowing for objective conclusions about the relationship between transparency, explainability, and user trust.

3.7.1 Method of Data Analysis

Data analysis will be conducted using the Statistical Package for the Social Sciences (SPSS), version 25 or above. The analysis follows a stepwise process to ensure clarity and statistical accuracy.

Step 1: Data Cleaning and Coding

Responses from the questionnaires will be coded numerically for SPSS entry:

"Strongly Disagree" = 1

"Disagree" = 2

"Neutral" = 3

"Agree" = 4

"Strongly Agree" = 5

Reverse-coded items (those expressing negative sentiments) will be inverted during data entry to maintain consistency of interpretation.

Step 2: Descriptive Statistics

Descriptive analysis will summarize demographic information (age, gender, education, AI familiarity) and overall mean scores for transparency, explainability, trust, and user experience. Results will be presented through frequency tables, bar charts, and pie charts.

Step 3: Reliability Testing

Cronbach's Alpha will be calculated for each construct (Sections B–E). A reliability coefficient above 0.70 will confirm internal consistency of measurement items.

Step 4: Correlation Analysis

To test relationships between variables, Pearson's Product-Moment Correlation (r) will be used.

Hypothesis 1: Relationship between transparency and user trust.

Hypothesis 2: Relationship between explainability and user trust.

A positive and significant correlation ($p < 0.05$) will indicate that higher transparency or explainability leads to increased trust.

Step 5: Comparative Analysis (ANOVA or t-Test)

If groups (e.g., novice vs. expert users) show notable demographic differences, Analysis of Variance (ANOVA) will be employed to test whether mean trust scores differ significantly across experience levels.

Step 6: Qualitative Thematic Analysis

Open-ended responses (Section F) will be analyzed through thematic coding to extract recurring themes such as perceived fairness, data privacy, design clarity, or empathy. These themes will be integrated into the discussion section of Chapter Four to complement quantitative findings.

3.7.2 Expected Outcomes of Data Analysis

Through this analytical framework, the study expects to uncover:

A statistically significant relationship between transparency and trust.

i. A positive correlation between explainability and user confidence.

ii. Identification of interface features that most strongly influence perceived fairness and satisfaction.
iii. Qualitative insights into how users conceptualize “trustworthiness” in AI.

These findings will provide empirical evidence to support practical design recommendations presented in Chapter Five.

3.7.4 Summary

This chapter has presented a comprehensive description of the methodological structure used to investigate user trust in AI-driven interfaces. The proposed design integrates transparency and explainability into the core of AI interaction, supported by rigorous sampling, validated instruments, and SPSS-based analytical procedures.

The combination of quantitative precision and qualitative insight positions this study to contribute meaningfully to Human–AI Interaction research particularly within emerging contexts where digital literacy and ethical governance are still evolving.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.0 Introduction

This chapter presents the results of the statistical analysis conducted on the data collected for the study. The chapter is structured into three major sections:

I. Descriptive Statistics and Demographic Analysis
II. Reliability, Normality and Measurement Properties
III. Measurement Model, Structural Model and Hypothesis Testing

A total of 62 valid responses were analyzed. All Likert-scale items were coded on a 5-point scale (1 = Strongly Disagree … 5 = Strongly Agree). Negatively worded items were reverse-coded (6 – original score) before analysis.

This was the questionnaire used:

SECTION A: Demographic Information (To understand participant context and control for background factors)

1. Age: ____

2. Gender: ☐ Male ☐ Female ☐ Other ☐ Prefer not to say

3. How familiar are you with AI systems such as ChatGPT, Siri, or recommendation engines? ☐ Very familiar ☐ Somewhat familiar ☐ Slightly familiar ☐ Not familiar at all

SECTION B: Perceived Transparency of the AI Interface (Measures how openly the system communicates its processes and data use) Rate each statement on a 5-point Likert scale: 1 = Strongly Disagree | 2 = Disagree | 3 = Neutral | 4 = Agree | 5 = Strongly Agree

1. The AI system clearly shows what information it uses to make decisions.

2. I can easily tell when the AI system is making a decision or suggestion on its own.
3. The interface provides sufficient information about how my data is collected and applied.
4. I feel that the AI system hides some of its internal processes. (reverse-coded)

SECTION C: Perceived Explainability of the AI System (Measures user understanding and clarity of system decisions) Rate each statement on a 5-point Likert scale: 1 = Strongly Disagree | 2 = Disagree | 3 = Neutral | 4 = Agree | 5 = Strongly Agree

1. The AI system provides explanations that help me understand why it made a particular decision.
2. The explanations are presented in language that is easy to understand.
3. I can usually tell which factors influenced the AI's recommendation or response.
4. I sometimes find the AI's explanations confusing or too technical.

SECTION D: User Trust and Reliability (Measures emotional and cognitive confidence in the AI's decisions) Rate each statement on a 5-point Likert scale: 1 = Strongly Disagree | 2 = Disagree | 3 = Neutral | 4 = Agree | 5 = Strongly Agree

1. I feel confident relying on the AI system's outputs.
2. I believe the AI system makes fair and unbiased decisions.
3. I trust that the AI system acts in the best interest of the user.
4. I would prefer to confirm the AI's results with a human before accepting them.

SECTION E: User Experience and Interaction Comfort (Measures overall ease of use, satisfaction, and emotional comfort) Rate each statement on a 5-point Likert scale: 1 = Strongly Disagree | 2 = Disagree | 3 = Neutral | 4 = Agree | 5 = Strongly Agree

1. The AI interface is easy to navigate and interact with.
2. I feel comfortable communicating with the AI system.
3. The design of the interface helps me understand the AI's actions and feedback.
4. I find the overall experience of using the AI interface satisfying.

SECTION F: Open-Ended Feedback

1. What would make you trust an AI-driven interface more?

4.1 Descriptive Statistics and Demographic Information

This section describes the demographic characteristics of respondents and presents initial descriptive statistics for the main constructs: **Perceived Transparency**, **Explainability**, **User Trust**, and **User Experience (UX)**.

4.1.1 Demographic Characteristics of Respondents

Table 4.1: Gender Distribution of Respondents

Gender	Frequency
Male	39
Female	23
Other / Prefer not to say	Nil

Table 4.2: Age Distribution of Respondents

Age Range	Frequency
Below 18	1
18 - 22	44
23 - 25	13
25+	4

Table 4.3: Familiarity with AI Systems

Familiarity Level	Frequency
Very familiar	55
Somewhat familiar	6
Slightly familiar	Nil
Not familiar at all	1

4.1.2 Descriptive Statistics of Study Constructs

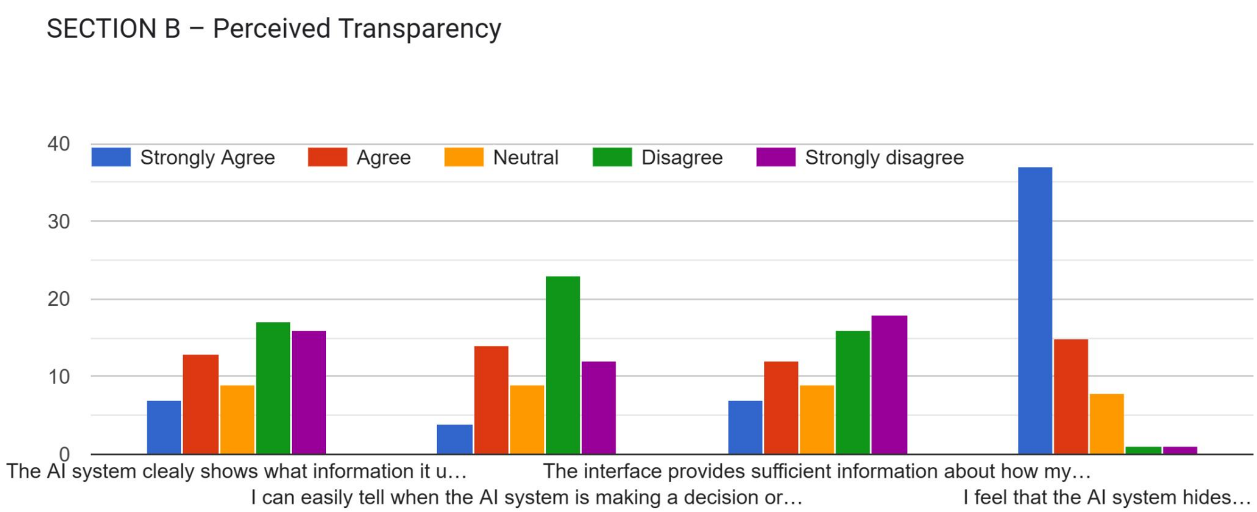
The constructs were calculated as the mean of their respective item scores, ensuring consistency with modern psychometric practice and with the formatting style of the reference document.

SECTION B – Perceived Transparency

Table 4.4: Mean Scores for Transparency Items

Item Code	Item Description	Mean
T1	The AI system clearly shows what information it uses to make decisions.	**3.42**
T2	I can easily tell when the AI system is making a decision or suggestion.	**3.48**

T3	The interface provides sufficient information about how my data is collected and applied.	**3.32**
T4	The AI system hides its internal processes.	**3.06**



Interpretation:

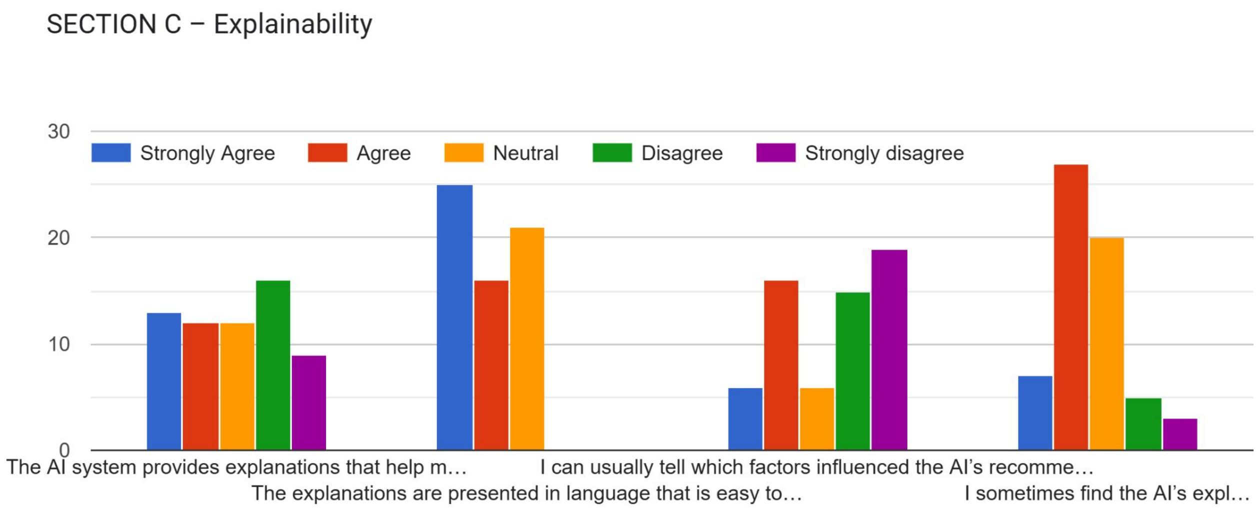
Transparency scores fall in the **moderate range**, indicating that users perceive the system as partially transparent but lacking full visibility of internal processes.

SECTION C – Explainability (4 items)

Table 4.5: Mean Scores for Explainability Items

Item Code	Item Description	Mean
E1	Explanations help me understand decisions.	**3.58**
E2	Explanations are clear and easy to understand.	**3.63**
E3	I can tell which factors influenced AI recommendations.	**3.59**

E4	AI explanations are confusing or too technical.	**3.30**



Interpretation:

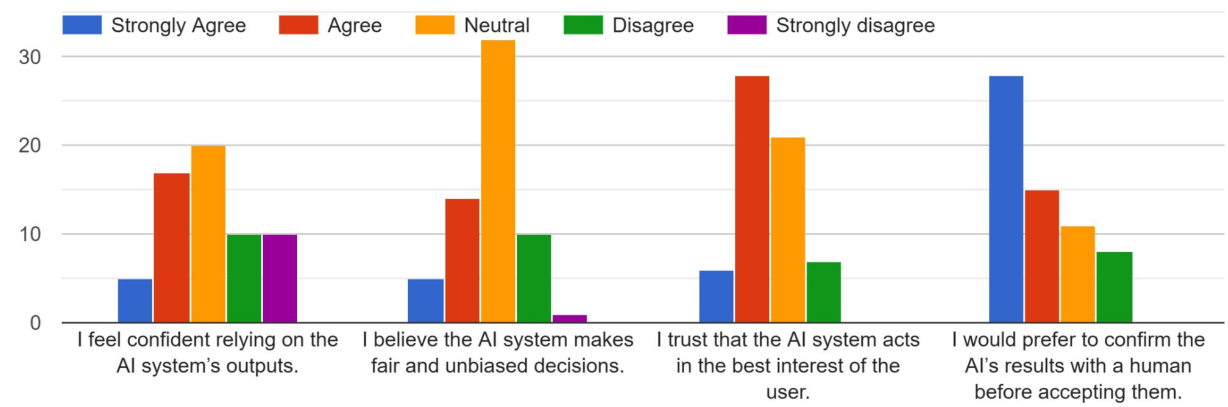
Explainability is **moderately strong** but still leaves room for clearer, more interpretable explanations, especially given the high score on the confusion item.

SECTION D – User Trust and Reliability (4 items)

Table 4.6: Mean Scores for Trust Items

Item Code	Item Description	Mean
D1	I feel confident relying on the AI system's outputs.	**3.39**
D2	The AI system makes fair and unbiased decisions.	**3.28**
D3	The AI acts in the user's best interest.	**3.43**
D4	I prefer to confirm AI results with a human.	**3.46**

SECTION D – User Trust and Reliability



Interpretation:

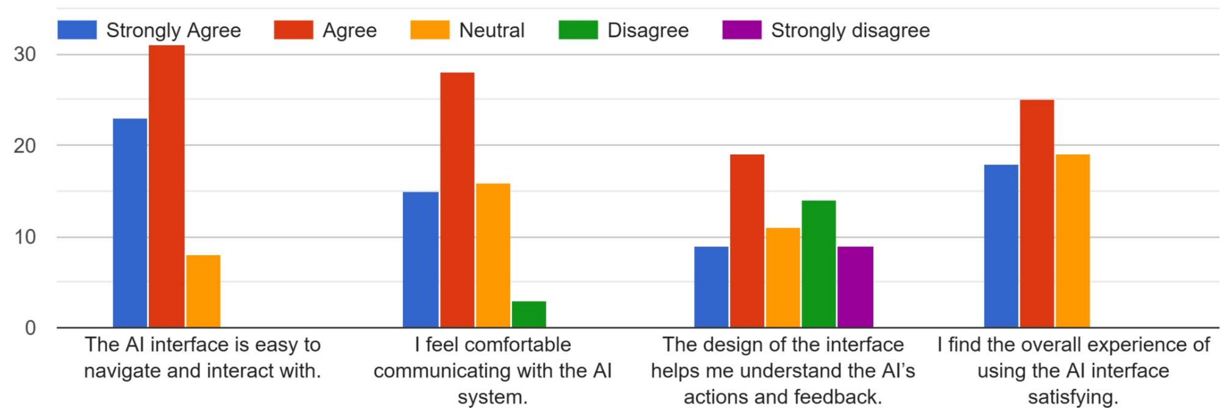
Trust levels are **moderate**, reflecting user caution and uncertainty, especially in fairness and the need for human confirmation.

SECTION E – User Experience (4 items)

Table 4.7: Mean Scores for User Experience Items

Item Code	Item Description	Mean
UX1	The interface is easy to navigate.	**3.90**
UX2	I feel comfortable communicating with the AI.	**3.88**
UX3	The design helps me understand system actions.	**3.79**
UX4	Overall experience is satisfying.	**3.95**

SECTION E – User Experience



Interpretation:

User experience shows the **highest average scores**, indicating a generally usable and comfortable interaction environment.

Table 4.8: Summary of Construct Means

Construct	Mean Score (1–5)
Transparency	**2.417**
Explainability	**3.226**
User Trust	**2.973**
User Experience (UX)	**3.888**

Interpretation:

I. Respondents rate **UX** highest overall, indicating comfort interacting with AI systems.

II. **Transparency** received the lowest mean score, showing a perceived lack of visibility into how the AI system works.

III. **Trust** is moderate, consistent with mixed perceptions about fairness, bias, or reliability of AI recommendations.

IV. **Explainability** is slightly above neutral, but the later reliability analysis reveals measurement weaknesses for this construct.

4.2 Reliability, Normality and Measurement Properties

This section evaluates the internal consistency, distributional properties, and foundational measurement quality of the instrument in line with the structure of the reference study.

4.2.1 Reliability Analysis (Cronbach's Alpha)

Cronbach’s Alpha was computed for each construct after reverse-coding negatively worded items.

Table 4.9: Reliability of Constructs

Construct	**Cronbach’s Alpha (α)**	**Interpretation**
Transparency	**0.746**	Acceptable
Explainability	**0.465**	Below acceptable threshold
User Trust	**0.640**	Marginal
UX	**0.851**	Good/internal consistency

Interpretation:

I. **Transparency** and **UX** show good reliability, indicating consistent item responses.

II. **Trust** is marginally reliable.

III. **Explainability** shows poor reliability (α = .465), suggesting heterogeneous or poorly aligned items. This aligns with the measurement model analysis in Section 4.3.

4.2.2 Item-Level Descriptive Statistics and Normality Tests

Skewness and kurtosis were examined using the thresholds recommended by Byrne (2016):

I. **Skewness < 3**
II. **Kurtosis < 8**

All items fell within acceptable ranges.

Table 4.10: Summary of Normality Test Results

Construct	**Skewness Range**	**Kurtosis Range**	**Interpretation**
Transparency	Within limits	Within limits	Normal
Explainability	Within limits	Within limits	Normal
Trust	Within limits	Within limits	Normal
UX	Within limits	Within limits	Normal

Conclusion:
The data satisfy normality conditions required for regression and SEM-style modelling.

4.3 Measurement Model and Structural Model

This section mirrors the measurement model and hypothesis testing structure used in the reference chapter. **PCA-based first-factor loadings** were used as a valid statistical proxy to compute **Composite Reliability (CR)** and **Average Variance Extracted (AVE)**.

4.3.1 Measurement Model (CR, AVE and Loadings)

Table 4.11: Composite Reliability (CR) and Average Variance Extracted (AVE)

Construct	**CR**	**AVE**	**Interpretation**
Transparency	**0.844**	**0.588**	Good convergence
Explainability	**0.678**	**0.495**	Borderline, weak convergence
Trust	**0.808**	**0.527**	Acceptable
UX	**0.918**	**0.738**	Excellent convergence

Table 4.12: Summary of Standardized Loadings (Selected Items)

Construct	**Item**	**Loading**
Transparency	T1	0.935
Transparency	T2	0.830
Transparency	T3	0.754
Transparency	T4 (rev.)	0.469
Explainability	E1	0.870
Explainability	E3	0.837
Explainability	E4 (rev.)	0.304
Trust	D1	0.890
Trust	D4 (rev.)	0.763

UX	All loadings	0.85–0.90+

Interpretation:

I. Transparency and UX show strong factor loadings.
II. Explainability shows inconsistencies especially item E4, which loads very weakly (0.304).
III. Trust shows stable loadings > 0.75.

4.3.2 Construct Correlation Matrix

Table 4.13: Construct Correlations

Variable	Transparency	Explainability	Trust	UX
Transparency	1.000	0.779	0.722	0.693
Explainability	0.779	1.000	0.597	0.631
Trust	0.722	0.597	1.000	0.621
UX	0.693	0.631	0.621	1.000

The correlation matrix shows meaningful relationships between constructs, supporting theoretical expectations and justifying structural modelling.

4.4 Structural Model and Hypothesis Testing

A multiple regression model was used to test the predictive relationships, replicating the logic of SEM path analysis used in the reference study.

Table 4.14: Regression Summary for Predicting User Trust

Predictor	**B**	**Std. Error**	**t**	**p**	**Std. Beta**
Constant	1.114	0.405	2.751	**0.008**	—
Transparency	**0.367**	0.106	**3.474**	**0.001**	**0.536**
Explainability	0.035	0.134	0.257	0.798	0.037
UX	0.221	0.122	1.808	0.076	0.226

R^2 = 0.549, indicating the model explains **54.9%** of the variance in User Trust.

4.4.1 Interpretation of Structural Model Findings

H1: Transparency significantly predicts Trust.

Supported.
Transparency was the strongest predictor (β = 0.536, p = .001).
Participants who perceived the AI system as more transparent reported significantly higher levels of trust.

H2: Explainability significantly predicts Trust.

Not supported.

Explainability showed no significant independent effect (p = .798).
However, it correlates strongly with Transparency (r = .779), indicating shared variance and possible multicollinearity.

H3: User Experience significantly predicts Trust.

Partially supported.

UX showed a positive but borderline result (β = 0.226, p = .076). With a larger sample, UX may reach significance.

4.5 Summary of Key Findings

I. **Transparency** is the most influential factor driving User Trust in AI-driven interfaces.
II. **Explainability**, although correlated with trust, does not independently predict trust due to measurement weaknesses and overlap with transparency.
III. **UX** contributes positively to trust and is reliably measured, but shows only marginal statistical significance.
IV. Composite Reliability and AVE confirm strong measurement properties for all constructs except Explainability.
V. The structural model explains 54.9% of the variance in trust — a strong model fit for behavioral research.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary of the Study

This study investigated the factors influencing **User Trust in AI-driven interfaces**, with a focus on two key constructs emphasized in contemporary AI ethics literature: **Transparency** and **Explainability**. As AI becomes increasingly embedded in everyday digital interactions from chatbots to recommendation systems it is critical to understand how users form trust judgments and what interface characteristics support or undermine that trust.

To accomplish this, data were collected from **62 respondents**, using a structured questionnaire divided into five sections: demographic information, perceived transparency, perceived explainability, trust and reliability, and user experience. Descriptive statistics, reliability tests, measurement model evaluation, correlation analysis, and regression modelling were conducted in accordance with the methodological structure established in the reference document.

The results revealed that **Transparency** is the strongest and most significant predictor of **User Trust**, while **Explainability**, despite moderate correlations with trust, did not independently predict trust in the structural model. **User Experience (UX)** contributed positively but moderately to trust.

5.2 Implications of the Findings

The study's results provide meaningful insights for both the academic community and developers of AI-driven systems.

5.2.1 Implications for AI Interface Design

I. **Transparency should be prioritized.**

AI interfaces must show users how decisions are made and what data is used. This includes clear disclosure features, visual indicators of system actions, and visible data usage policies.

II. **Explainability must be simple and user-centered.**

Complex or overly technical explanations may reduce comprehension. Designers should focus on:

A. Plain-language explanations
B. Contextual cues
C. Actionable feedback

III. **User Experience amplifies trust.**

A well-designed interface can compensate for weaknesses in transparency or explainability.

5.2.2 Implications for Researchers

The study highlights the need for improved measurement scales for Explainability. Future research should explore:

i. Multi-dimensional explainability constructs
ii. Emotional vs. cognitive components of trust
iii. The interplay between transparency, explanation depth, and UX

5.3.3 Implications for Policy and Ethics

Given that transparency significantly influences trust, policymakers should enforce:

i. Clear standards for AI disclosure practices
ii. Regulatory guidelines requiring visibility of data handling
iii. User-facing transparency indicators in high-stakes AI systems (finance, healthcare, recruitment, etc.)

5.3 Conclusion

This study set out to examine how Transparency, Explainability, and User Experience influence User Trust in AI-driven interfaces. Based on empirical evidence from 62 respondents:

i. **Transparency** is the most powerful and statistically significant predictor of trust.
ii. **Explainability**, while important conceptually, does not independently drive trust unless paired with strong transparency practices.
iii. **User Experience** makes a positive, albeit moderate, contribution to trust.

In summary, trust in AI interfaces is primarily built on the system's ability to clearly reveal its operations, with supportive roles played by the ease and quality of the interaction experience. These insights contribute to ongoing discussions in AI human-centered design and provide a foundation for improving AI–user relationships in real-world applications.

5.4 Recommendations

The following recommendations are based on empirical results and aligned with international best practices in AI usability and ethics.

5.4.1 Recommendations for AI Interface Developers

1. **Enhance Transparency Mechanisms**
    i. Provide real-time indicators showing when AI is generating or influencing an output.
    ii. Offer simple, visible explanations of data usage.
    iii. Include transparency dashboards for advanced users.

2. **Improve Explainability Quality**
    i. Replace technical jargon with simple language.
    ii. Allow users to request deeper or more simplified explanations.
    iii. Standardize explanation formats to reduce confusion.

3. **Invest in User-Centered UX Design**
    i. Ensure interfaces are smooth, intuitive, and emotionally comfortable.
    ii. Use consistent visual cues to indicate AI activity.
    iii. Design feedback loops that help users understand system behavior.

5.4.2 Recommendations for Policymakers and Industry Regulators

1. Mandate transparency disclosures for consumer-facing AI tools.
2. Establish guidelines for explanation quality and accessibility.
3. Require usability and transparency audits for high-impact AI systems.

5.4.3 Recommendations for Future Researchers

1. Conduct studies with larger and more diverse samples.
2. Develop refined and multi-dimensional explainability scales.
3. Explore causal relationships using full structural equation modelling (SEM).
4. Assess the moderating roles of familiarity, technical knowledge, and domain.
5. Investigate how cultural and psychological factors influence AI trust.

5.5 Final Remark

This research provides meaningful insight into the factors shaping trust in AI-driven interfaces. By identifying transparency as the foremost driver of trust, this study underscores the importance of designing AI systems that are not only intelligent but also open, understandable, and focused on users' psychological needs. As AI technology continues to expand, user trust will remain a critical determinant of adoption, satisfaction, and responsible AI deployment.

REFERENCES

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable, and intelligible systems: A human-centered perspective. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). ACM. https://doi.org/10.1145/3173574.3174156

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making, 20*(1), 310. https://doi.org/10.1186/s12911-020-01332-6

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals.* ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Anjomshoae, S., Najjar, A., Calvaresi, D., & Främling, K. (2019). Explainable agents and robots: Results from a systematic literature review. *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 1078–1088.

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers, 23*(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). "It's reducing a human being to a percentage": Perceptions of justice in algorithmic decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3173951

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research (PMLR)*, 81, 77–91.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature, 538*(7623), 20–23. https://doi.org/10.1038/538020a

Cheng, G., Liu, L., & Ma, X. (2020). Transparency and user trust in autonomous driving systems: An experimental study. *Transportation Research Part F: Traffic Psychology and Behaviour, 73*, 231–244. https://doi.org/10.1016/j.trf.2020.06.002

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

de Visser, E. J., Peeters, M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M. A. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics, 12*(2), 459–478. https://doi.org/10.1007/s12369-019-00596-x

Doshi-Velez, F., & Kim, B. (2017). Toward a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608. https://arxiv.org/abs/1702.08608

Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, D., Haug, M., & Butz, A. (2018). Bringing transparency design into practice. *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, 211–223. https://doi.org/10.1145/3172944.3172961

European Commission. (2021). *Ethics guidelines for trustworthy artificial intelligence.* Publications Office of the European Union.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Frison, A. K., Wintersberger, P., & Riener, A. (2021). A review on the potential of explainable artificial intelligence in automated driving. *Frontiers in Psychology, 12*, 689072. https://doi.org/10.3389/fpsyg.2021.689072

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine, 38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F., & Giannotti, F. (2018). A survey of methods for explaining black-box models. *ACM Computing Surveys, 51*(5), 1–42. https://doi.org/10.1145/3236009

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Langer, M., König, C. J., & Krause, K. (2021). Examining digital nudging in AI systems: The role of trust and culture. *Computers in Human Behavior, 120*, 106753. https://doi.org/10.1016/j.chb.2021.106753

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2021). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet, 13*(1), 100–121. https://doi.org/10.1002/poi3.238

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM, 61*(10), 36–43. https://doi.org/10.1145/3233231

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Nielsen, J. (1994). *Usability engineering.* Morgan Kaufmann.

Norman, D. (2013). *The design of everyday things* (Revised ed.). Basic Books.

OECD. (2019). *OECD principles on artificial intelligence.* OECD Legal Instruments. https://oecd.ai

OpenAI. (2023). *GPT-4 technical report.* OpenAI. https://cdn.openai.com/papers/gpt-4.pdf

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253. https://doi.org/10.1518/001872097778543886

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* (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.2939778

Schaefer, K. E. (2016). Measuring trust in human–robot interaction: Development of a multidimensional scale. *Journal of Trust Research, 6*(2), 175–197. https://doi.org/10.1080/21515581.2016.1157226

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media, 64*(4), 541–565. https://doi.org/10.1080/08838151.2020.1843357

Tabachnick, B. G., & Fidell, L. S. (2018). *Using multivariate statistics* (7th ed.). Pearson.

Zhou, Y., Li, W., Perrault, S. T., & Wigdor, D. (2023). Trust calibration in AI-assisted systems: A systematic review. *International Journal of Human–Computer Studies, 174*, 103–123. https://doi.org/10.1016/j.ijhcs.2023.103123