THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 6. Issue 1 (2026)
https://doi.org/10.62019/r2vyk412

Explainable AI (XAI) in Practice: Users' Perceptions of Transparency and Understanding in Automated Decision Systems

Nadeem Ahmad Malik*, Sakeena Parveen, Irfan Hanif

Chronicle

Article history
Received: Dec 18, 2025
Received in the revised format: jan 13, 2026
Accepted: Jan 27 2026
Available online: Feb 25, 2026

Nadeem Ahmad Malik* is Direct IT Services PMAS Arid Agriculture University, Rawalpindi, Pakistan.
Email: Nadeem.malik@uaar.edu.pk

Sakeena Parveen Is MSCS Scholar Department of Computer Sciences Dha Suffa University, Karachi, Pakistan
Email: sakeena.parveen1987@gmail.com

Irfan Hanif is Research Scholar School of Architecture, Computing and Engineering University of East London, England
Email: irfanhanif@gmail.com

Corresponding Author*

Abstract

The research explores user perceptions of transparency and interpretability of explainable artificial intelligence (XAI) systems and the quality of the given explanations to understand how trust, understanding, and ethical judgment of automated decisions are formed. Using a qualitative design, 16 participants with different professional experiences were interviewed using semi-structured questions regarding their experience with AI-based decision systems. The data were evaluated using manual thematic analysis in order to determine major patterns and the meaning of narrative in the stories of the participants. The participants underlined the importance of transparency that should be explained in a clear way, and should have a contextual and defensive quality. Clarity and active participation were encouraged through clear communication, whereas the lack of clarity and excessive technicality resulted in confusion and doubt. The research discovered that transparency is seen not just as a technical aspect but as a relationship and moral construct that is linked to fairness and respect for the autonomy of users. The results reveal the significance of conceptualising the XAI systems by focusing on the user interpretability, ethics responsibility, and communicative effectiveness. Clear descriptions of the ways people should be capable of closing the divide between AI logic and human logic to boost public confidence in AI. The study concludes that explainability ought not only to be treated as a form of afterthought in AI design, but also as a pillar of technology innovation that is human-centred.

Keywords: Explainable AI, Transparency, Automated Decision Systems, User Trust, Interpretability, Human-AI Interaction, Ethical Accountability.

INTRODUCTION

AI in decision-making has been hugely incorporated in various industries such as healthcare, finance, law enforcement, and education (Han et al., 2025). This has already seen the automated decision systems take regular decisions that have direct consequences influencing the lives of individuals, like credit granting, medical diagnoses, job offerings, etc (Gsenger & Strle, 2021). But now that these systems are more intricate and less transparent, users are finding it more challenging to understand the process and the reasons behind decision-making. The latter difficulty has led to the appearance of a new discipline of Explainable Artificial Intelligence (XAI), which focuses on making the decisions of AI more transparent and understandable (Minh et al., 2022). Chinnararaju (2025) states that such an inability to explain is among the most urgent challenges to AI trust and adoption, and more than 62 organisations possessing such gaps in interpretability identified them as the main barrier to responsible AI implementation. Regulatory and ethical issues of fairness, accountability, and bias have increased along with the demand for transparency (Akinrinola et al., 2024). According to Patidar et al. (2024), a large proportion of AI users, both professionals and non-professionals, demonstrate concern about opaque algorithms, with many of them seeing them as unreliable. This is further heightened by

the fact that AI systems are capable of making consequential decisions without proper explanations, which causes less user confidence and could create a barrier to adoption (v et al., 2024). Explainability is a challenging issue in ensuring user trust, legitimacy and responsible innovation in AI systems (Ferrario & Loi, 2022). Due to the direct impact that algorithmic transparency has on user judgments of the fairness and credibility of automated decision-making, as Grimmelikhuijsen (2023) observed, algorithmic transparency can predict the overall acceptance of AI technology by users.

Although there is an increasing academic and practical interest in Explainable AI, the literature on how users, especially non-expert end users, understand transparency and interpret automated decision systems' explanations is still lacking. Although theoretical models of XAI are oriented towards the creation of interpretability methods, there are fewer studies that examine how these explanations are perceived, comprehended, and accepted in practice (Sokol & Flach, 2024). According to Shulner-Tal et al. (2025), explanations provided may not be adequately interpreted by users, or they may give false hope to AI results. Accordingly, the primary issue is not only to provide explanations, but to guarantee that those explanations have a sense and make sense to the user. This is a critical challenge to the practical application of the XAI principles due to the absence of empirical knowledge regarding the subjective experience of transparency in users.

- To investigate the perceptions and meanings of the explainability features of AI-based decision systems among users.
- To determine the significant elements that can increase or decrease the comprehension of AI-generated explanations by users.
- To explore the connection between the perceived transparency and user confidence in the automated decision results.

SIGNIFICANCE OF THE STUDY

The relevance of this research lies in the fact that it has helped to fill the gap between the technical design of XAI systems and the human experience of transparency. Although explainability in technical literature is usually discussed on the level of models and algorithms transparency, there is little empirical evidence that speculates on how these explanations are in practice and what users experience. Cetinkaya and Krämer (2025) stressed that the concept of transparency is relational because of the system characteristics and user cognition. Thus, this study will also enhance the developing discussion, bringing its own idea based on the perception of users, instead of the functionality of the system itself. The results have direct consequences on the designers, policy makers, and organisations who would want to establish AI systems that are not only precise, but also understandable and credible to the users. The present qualitative research involves a semi-structured interview design, which is used to investigate the perception of explainability and transparency in AI decision systems among users. The number of interviewees was 16 and represented various professional fields with previous experience working with AI-supported decision tools. The interviews were user-centred and concerned the experiences, level of comprehensiveness and trustworthiness of the explanations given in such systems. Thematic analysis was done using manual data analysis to determine the common patterns and interpretive themes regarding perceptions of transparency and understanding. The analysis was conducted in line with the strategy of Ribes et al. (2021), which highlights the interpretive quality of user perceptions and thus, the

interaction of the quality of explanation and perceived trustworthiness. The approach to the methodology guarantees a comprehensive, contextualised perspective of the way XAI operates in practice; the technical assessment is transcended to examine the experience of users interacting with automated decision systems.

LITERATURE REVIEW

Theoretical Framework

Explainable Artificial Intelligence (XAI) is a new paradigm in the field of artificial intelligence ethics and design, which focuses on interpretability, user trust, and accountability. XAI is theoretically based on three constructs that are tightly connected (transparency, understandability, and trust), which together define the interaction between humans and automated systems of decisions. Chinnararaju (2025) has suggested an extensive model of XAI interpretability based on the cognitive and socio-technical theory, where successful explanations should have the right balance between technical intelligibility and understanding by the user. He claimed that transparency without cognitive accessibility will result in information overload, which will be counterproductive to the aim of explanation. These two perspectives explain not only as a technical property but also as a process of communication between the system and the user (Ashraf, Arzu, Abrar, & Anwar, 2025).

This theoretical framing was supported by Grimmelikhuijsen (2023), who used the theory of trust and perceived fairness. He argued that transparency is a psychological signal that determines the judgment of users on the fairness and legitimacy of the system. In his findings, AI decisions are perceived more credibly by the users when they can access intelligible explanations involving the processing and weighting of data. On the contrary, transparency breeds confusion and a sense of partiality (Khan, Butt, Noor, Ishaq, & Siddiqui, 2025). Correspondingly, Cetinkaya and Krämer (2025) used the Human-Computer Trust Framework to state that transparency in AI cannot necessarily lead to trust. Instead, trust is created due to perceived control, understanding, and consistency with human values (Rehmat, Hassan, Rumaan, & Abrar, 2025).

Asghari et al. (2021) contributed to the theoretical debate by suggesting that XAI explanations need to meet the requirement of meaningful information; in other words, the explanations should facilitate the ability of users to take action, challenge, or redefine the automated decisions. They have an interdisciplinary framework that spans philosophy, law, and computer science, with an emphasis on the ethical and epistemic aspects of AI transparency. Schraagen et al. (2021) proposed a systems viewpoint and claimed that explainability was a mediating variable between the complexity of the system and its user performance. Their model assumes that with the increasing system opacity, the dependence of the users on the system will go down unless the explainability mechanisms are sufficient to compensate for this (Braun, Clarke, & Hayfield, 2022). Together, these theoretical lenses bring out the point that transparency is not an inert system characteristic but a dynamic relationship construction, which is influenced by cognitive, affective, and contextual factors.

Transparency, Trust, and User Perception

Transparency has always been referred to as the principle of user trust in automated decision systems. Blake (2024) states that the user will be more willing to accept the

decision made by AI when they feel that the process and reasoning are transparent. In his analysis, opaque systems have brought about a black-box anxiety, a state of doubt, skepticism, and decreased dependence on AI output. Likewise, Patidar et al. (2024) saw that perceived system competence and fairness were positively related to transparency in the communication of the decision logic. They clarified that the XAI should not solely show the way decisions are made but also put them into context, which end-users can understand (Khan, Kausar, & Khan, 2025).

As Ribes et al. (2021) empirically showed, even the existence of the explanations may serve as a trust indicator and can affect the level of emotional involvement and readiness to trust AI suggestions in the users. They advised, though, that excessive revelation of technical specifics could work against them if users think the system is too complicated. Morrison et al. (2024) also addressed such a paradox, as the effectiveness of more imperfect or less accurate explanations can enhance trust, as long as they correspond to the intuitive perception of users. Their research found that perceived explanatory coherence, the subjective impression that an explanation is coherent, is more significant in creating trust than objective accuracy in itself (Zeb, Abrar, Saqib, & Rizvi, 2025).

Transparency and trust are therefore very context-specific. However, Shulner-Tal et al. (2025) discovered that AI systems that provide traceable reasoning mechanisms were more credible to users than systems that can only give outputs. However, they also noted that users are prone to anthropomorphize transparent AI systems and hold them accountable like humans. This expectation of infertility may result in frustration when explanations are inadequate in terms of their emotional or moral adequacy. Cetinkaya and Krämer (2025) supported this by stating that too much transparency can be counterproductive as it can compromise trust in case it reveals inefficiencies in the algorithm or bias. Therefore, to design transparency effectively, it should be tuned in a way that it encourages others to understand it, but not so broad that it imposes and dissuades the users (Jabbar & Gul, 2026; Rehmat, Hassan, Khan, & Abrar, 2025).

Understanding and Cognitive Load in XAI Systems

The cognitive accessibility of presented information is essential in the understanding of AI explanations by the user. The study by Chiaburu et al. (2024) investigated the effect of uncertainty and cognitive load on the interpretation of XAI output. Their work revealed that in cases where explanations are presented in probabilistic or mathematical terms, non-expert users tend to misinterpret these explanations, leading to overconfidence or mistrust. They hypothesized that it is necessary to have adaptive explanations—those that match the detail and complexity to the expertise of the user—to be effective. On the same note, Asghari et al. (2021) cautioned against providing raw data or weights of algorithms, but instead advocated for explanation interfaces that answer "why" or "how" questions (Ahmed, Faisal, Hasan, & Ghazi, 2025).

Praveenraj et al. (2023) expanded on this point and highlighted the pedagogical role of XAI. They theorized explanations as learning resources that allow users to create mental representations of AI behavior. Their results indicate that when users are able to associate explanations with known decision rules, they demonstrate more understanding and are more satisfied. However, on the other hand, opaque or excessively technical explanations can lead to mental exhaustion and decreased interest. This observation was supported by Schraagen et al. (2021), who stated that it is essential for explanations to be contextually sound and time-oriented to the

interactions between the user, in order to bring about real-time understanding (Arzu, Sattar, Sultan, Abrar, & Khuharo, 2025).

Blake (2024) emphasized that the visual, textual, or interactive presentation format of explanations has a significant influence on the understanding of the user. Visual explanations (e.g., feature importance graph or decision tree) are also interpretable by users with little technical background. In the meantime, Morrison et al. (2024) determined that the improvement in users' calibration of trust through iterative explanations that gradually display the reasoning stages enables them to detect inconsistencies more efficiently. These facts highlight the significance of the usability of explanations as a design concept that should provide the intertwining of cognitive psychology and interface design to facilitate further user comprehension (Ahmed, Ali, & Mansoor, 2025; Ahmed, Urooj, Farheen, & Ishaq, 2025; Rehmat, Hassan, Rumaan, Baig, & Abrar, 2025).

Trustworthiness and Human-AI Interaction

Human trust and AI transparency are connected and are complex on a very delicate level. It has been shown by Grimmelikhuijsen (2023) that although the accuracy of AI systems may be high, users may still lack trust in them because of the lack of explanations or their inaccessibility. Trust is not entirely constructed based on rationality, but it is also emotional and normative, informed by expectations of fairness, autonomy, and ethical responsibility. Cetinkaya and Krämer (2025) discovered that perceived benevolence and integrity affect the formation of trust in AI systems, and not the transparency of the systems. Differently put, the user measures whether the AI is acting in their interest, which indicates more socio-ethical aspects of human-AI relations (Kausar & Ahmed, 2026).

Shulner-Tal et al. (2025) discussed the way decision-making authority in humans and AI is differentiated by the users. They discovered that participants were more comfortable with human supervision, despite the participation of AI in the decision. This shows how much of a human assurance bias persists, in that trust is only given based on human control as perceived. Ribes et al. (2021) also indicated that users enjoy explanations that indicate accountability and responsibility, which are traditionally linked to human decision-making. Thus, efficient XAI should not just explain how a decision has been made but also who is the one that eventually makes it (Asif, Jabbar, & Yusaf, 2026).

Both Morrison et al. and Blake found that imperfect explanations could be effective at boosting user trust when they indicated system fallibility and humility (2024). The system should be made to recognize uncertainty; as a result, the users are likely to view it as honest and credible.

Nevertheless, they observed that too much uncertainty disrupts confidence, especially in high-stakes areas like health care or finances (Chiaburu et al., 2024). XAI design should therefore be keen on determining the level of transparency and reassurance so that the necessary transparency is provided without diminishing the perceived credibility (Ashraf et al., 2025).

Explainability Techniques and Practical Constraints

Explainable systems have technical and ethical problems with their application. Patidar et al. (2024) divided the current XAI techniques into post-hoc and intrinsic. The explanations, like feature attribution and rule extraction, convert the model outputs

into interpretations during the training of the model. Intrinsic methods cannot be intravenous, on the other hand. As much as these methods lead to improved interpretability, they are usually associated with trade-offs between accuracy and explainability. It was found by Morrison et al. (2024) that explainability simplicity can decrease predictive performance, which generates a conflict between functionality and transparency. Asghari et al. (2021) proposed that it is not only the manner of explaining but also what to explain that is a challenge. They put forward that we should explain what is in the decision-making that users can act meaningfully upon, including the criteria by which the decision-making is carried out or the option choices taken. Blake (2024) emphasised the same context-specific aspect of the explanation and added that explanation must have some relation to the position of the user, knowledge of the field, and stakes in the choice. One of them is that a clinician might require more information than a patient requires to be offered by the AI diagnosis system.

However, Schraagen et al. (2021) also warned that excessive technical transparency can impede ethical considerations. They fostered situational explainability, which is a methodological approach that integrates the situational context of the user, the applicability of the task and morality. Another essential point, presented by Ribes et al. (2021), is that technical explanations should be accompanied by communicative techniques, which can make AI-related interactions more human. Therefore, the existing literature provides the direction to the necessity of hybrid models that should combine the algorithmic interpretability with user-oriented communication design.

Ethical and Social Dimensions of Transparency

The moral principles of XAI are based on fairness, responsibility, and respect for human autonomy. According to Asghari et al. (2021), meaningful explanations are not merely instrumental, but normative as well, and they reflect respect for the rights of users to understand and challenge. Therefore, transparency plays as an ethical duty, especially when making decisions that have social or legal impacts. As pointed out by Grimmelikhuijsen (2023), the absence of explanations may strengthen the belief in algorithmic bias and diminish the ability of people to trust institutions using AI. Equally, Shulner-Tal et al. (2025) demonstrated that transparency is associated with moral legitimacy by users, and they expect AI systems to be able to explain their decision in a manner that corresponds to human moral standards.

Cetinkaya and Krämer (2025) also continue this argument by discussing the paradox of transparency: on the one hand, greater openness should increase accountability; on the other hand, the notion can also reveal the weaknesses, including the sensitivity of data or the flaws of models. Therefore, strategic dilemmas in organisations are common in deciding the degree to disclose. Blake (2024) argued that ethical transparency does not mandate complete disclosure but honest disclosure, which is a description of the reasons that are factual, locally relevant, and advantageous to ethics. Patidar et al. (2024) agreed with this opinion, noting the necessity to introduce transparency policies that would strike a balance between interpretability and privacy and security limitations. The study by Schemmer et al. (2021) also recommended that XAI systems must contribute to collaborative intelligence, which allows users to challenge, refine, and co-create decisions with AI tools. Such a participatory method is in line with new ethical theories, which consider transparency as a continuous dialogue and not a fixed aspect. Explainability can be used to contribute to more equitable and democratic AI governance by fostering a more

profound comprehension of AI through choosing participants in interpretive processes, as well as empowerment (Stamboliev, 2023).

Gaps in Existing Research

Although XAI has been discussed extensively both theoretically and empirically, there are significant gaps that prevail in the literature. First, most of the literature is dedicated to technical transparency and does not pay much attention to the interpretive experience of the user. Little has been learned about the perception of explanations by non-expert users in practical environments. According to Asghari et al. (2021), when explainability is described, it is often defined based on the view of the designer and not the end user. The disconnect established gives a difference between the intended and perceived transparency.

Second, even though Grimmelikhuijsen (2023) and Cetinkaya and Krämer (2025) discussed the dynamics of trust, not many studies have empirically associated transparency, understanding, and trust into a single paradigm. Most people address these variables independently, and they remain over what happens in practice about the interaction between them. Shulner-Tal et al. (2025) recommended the incorporation of integrative research that embodies the interrelation interaction of the design of explanation, cognitive understanding and affective trust.

Third, qualitative and user-centred studies analysing the perceptions in various contexts are scarce. The majority of available literature, like Morrison et al. (2024) and Schraagen et al. (2021), is based on experimental simulations and not on the experience of the real user. It is this methodological gap that restricts our knowledge on how users are negotiating with ambiguity, how they navigate ambiguity and how they come to trust AI systems in everyday life.

Moreover, the cross-cultural aspects of transparency perception are not sufficiently studied, which in turn have been implicitly identified by Ribes et al. (2021), who also expressed that contextual norms are used to define the value and understanding of transparency. There is no clarity on the best level of transparency in the literature. Under- and over-transparency can affect trust and understanding negatively, as it was shown by Chiaburu et al. (2024) and Blake (2024). However, empirical advice on the way of calibration of the depth of explanations is scarce. The way forward for future research should therefore not be binary concepts of transparent vs opaque systems but rather a more subtle concept of situated transparency.

The paradigm advancements have been substantial, yet the practical implications of XAI require elaboration, in particular, the attitude to transparency and experience among users thereof. These gaps are filled with the existing literature since the research will be of qualitative nature and employ interviews as a means of researching to find out how users perceive, believe and behave on the explanations given by AI-based decision systems. It will be aimed at elucidating the subjective nature of explainability that has been ignored by the quantitative models thus far using thematic analysis.

METHODOLOGY

Research Design

To examine the perceptions of users of transparency and understanding of explainable artificial intelligence (XAI) systems, this study employed a qualitative research design (Braun and Clarke, 2017). The qualitative approach provided the

freedom to examine the complex cognitive and emotional responses which cannot be measured quantitatively. The primary aim was to have a comprehensive and interpretative understanding of the meaning of AI-generated explanations to the users in the context of automated decision systems. It was designed following the interpretivist paradigm, which assumes that the meaning and knowledge are co-created within the framework of social interaction and personal understanding of it (Ribes et al., 2021).

The primary method of data collection was a semi-structured interview method. This method gave the participants the freedom to narrate their experience and perception in their own words and also gave the researcher the liberty to query and seek clarifications on the emerging knowledge. The choice of interviews was justified by Asghari et al. (2021), who stressed that to observe how users react to XAI, it is necessary to provoke their subjective interpretations of the concept instead of determining the influence of this kind of behaviour. The emphasis was placed on the way people see transparency, interpretations, and how these interpretations are connected with trust and understanding.

The theoretical perspectives described by Chinnararaju (2025) and Grimmelikhuijsen (2023) formed the basis of the research design since the two scholars saw transparency and trust as relational constructs mediated by human cognition. It is the qualitative approach that therefore allowed examining subtle and context-specific meanings users assign to explainable AI systems. This study aimed to identify the relationship between the interpretability of AI systems and the perceived fairness, accountability, and user confidence by focusing on the narratives of the participants.

Sampling Strategy

Participants were selected through a purposive sampling technique to respond to the research questions, and they had personal or professional experience working with or using AI-driven decision systems. The sample was created to include 16 participants who cover a variety of industries such as healthcare, education, finance, human resources, and digital services. The inclusion criteria included the fact that the participant needed to be exposed to AI-based decision-making systems (e.g., predictive analytics software, recommendation systems, automated screening algorithms, or customer care chatbots). This helped to make sure that the participants could give substantial information about the role of expositions in their comprehension and credibility.

This sample size of sixteen was deemed sufficient with regard to data saturation as per the qualitative research standards. According to Patidar et al. (2024), small but heterogenous groups of participants enable the researcher to dive deep and uncover the experience richness as well as discover patterns across the context. Diversity was upheld in terms of gender, professional background, and technological literacy in order to get a wide range of perspectives. This methodology corresponded to the interpretivist goal of having knowledge of multiple realities instead of projecting the results onto populations.

They were recruited through professional networks and online invitations. The subjects were informed about the research subject and guaranteed anonymity. The voluntary participation was highlighted, and all the participants accepted the interviews with informed consent. The sample diversity allowed achieving a balanced representation of both expert and non-experienced user experiences in line with the findings by

Cetinkaya and Krämer (2025) that the perception of transparency among users greatly depends on prior exposure to technical knowledge.

Data Collection

The gathered data were obtained using semi-structured interviews that were carried out using video conferencing. Individual interviews took between 45-60 minutes, depending on the engagement and the expounded answers of the participants. The interviews were conducted through an interview protocol that was based on the reflections of Blake (2024) and Shulner-Tal et al. (2025) and involved three major areas: perceived transparency, perception of explanations, and trust in AI decisions. Open-ended questions enabled the participants to share their experiences, thoughts and feelings related to the interaction with automated systems.

Such questions as: How do you rate the transparency of the AI system you used? Do the explanations make you understand how decisions were made? The availability or lack of explanation influenced the level of your trust in the system were used as examples. Unclear answers were explained and elaborated through probing questions. The adaptability of the interviews in the form of interviews made it possible to allow the participants to use examples of either positive or negative experiences regarding XAI features.

They were audio-taped and transcribed later and analysed with the consent of verbatim interviews. The data collection also took into consideration the values of ethical sensitivity and the comfort of the participants, which is also a human-oriented value as described by Asghari et al. (2021). The overall approach offered channels of acquiring qualitative information that were rich and described lived experiences and interpretative perceptions of users who had been interacting with explainable AI systems.

Data Analysis

In order to analyse it, the inductive thematic analysis (Braun and Clarke, 2022) was employed and did not assume such categories in advance, which allowed patterns and themes to be identified in the course of data analysis. This analysis was conducted in six cycles of data analysis that involved familiarisation, initial coding, theme identification, theme review, definition, and story synthesis. Repeated readings of the transcripts were done to identify useful statements that were related with transparency, understanding and trust. The subsequent step was the grouping of codes into sub-themes based on shared perceptions or experience.

According to Praveenraj et al. (2023), XAI perception studies based on manual analysis have a higher interpretive validity because they enable researchers to capture contextual meanings that may not be considered by automated tools. A total of four major themes were identified based on the conceptual focus of the study, which included (1) perceived clarity of explanations, (2) user comprehension and cognitive alignment, (3) trust calibration and confidence, and (4) accountability and ethical assurance.

The process was done through constant comparison to narrow down the relationship between the themes. The analysis of contradictory and deviant cases was done in order to give depth and credibility. The study made theoretical use of the analytic thinking of Grimmelikhuijsen (2023), who highlights the role of the design of explanation, of cognitive processes and emotional trust. Codes were arranged in

thematic matrices, and they were facilitated in systematic comparison between participants. The quotes of the participants were then included in the discussion to demonstrate and support the identified themes.

Ethical Considerations

The research had ethical integrity throughout all stages. The purpose of the study, as well as the voluntary participation of the participants and their right to leave at any time without consequences, were described to them. Data collection was done through informed consent. Personal identifiers were not used in transcripts, and pseudonyms were employed to guarantee anonymity. The transcripts and audio recordings were safely saved in encrypted drives available to the researcher only.

The ethical method focused on the sustainability, responsibility, and empowerment of research participants, which is based on human subjects and AI solutions. Since the research problem under investigation is the perceptions of transparency and trust, specific attention was paid to avoid making the participants feel judged because of their technological literacy level or their views. All data treatment methods were in line with academic research ethics and data protection rules. Moreover, the concept of reflexivity was upheld to reduce the bias of the researcher. The researcher kept on thinking about personal assumptions regarding AI and interpretability so as not to influence the interpretation of data unduly. One of the conditions involved the member checking where the sample participants were made to see the short summaries of their answers, to determine accuracy and authenticity. This increased the validity and morality of the findings.

Participants' Profile

The sample of participants was a fair representation of the cross-section of AI exposure and professional background. The respondents who were highly and lowly technologically literate gave different answers to the definition of transparency and understanding. According to the findings of the research by Cetinkaya and Krämer (2025), the perceptions of trust and interpretability depend on the expertise and experience of a particular user in the field of use and technology. This information, gathered through these sixteen participants, was therefore a multidimensional picture of the perception that users had of XAI. Thematic results of their accounts formed the basis of the discussion of four key themes, including clarity of explanations, comprehension, trust calibration, and accountability, which holistically demonstrated how explainability affects user interest and trust in automated decision systems.

Table 1:
Participants' Profile

Participant	Gender	Occupation	Experience with AI Systems	Level of Technical Literacy	AI System Context
P1	Female	Financial Analyst	3 years	Moderate	Credit Scoring Tool
P2	Male	HR Manager	4 years	Moderate	Automated Recruitment System
P3	Female	Healthcare Practitioner	2 years	Low	Diagnostic Support System
P4	Male	Data Scientist	5 years	High	Predictive Analytics Platform

P5	Female	University Lecturer	3 years	Moderate	Educational Recommendation System
P6	Male	Customer Service Representative	2 years	Low	Chatbot System
P7	Female	Policy Researcher	4 years	High	Decision Simulation Tool
P8	Male	Software Engineer	6 years	High	Model Interpretability Dashboard
P9	Female	Marketing Specialist	2 years	Low	Ad-targeting Algorithm
P10	Male	Compliance Officer	3 years	Moderate	Fraud Detection Platform
P11	Female	UX Designer	5 years	High	Interactive AI Interface
P12	Male	Medical Technician	3 years	Moderate	Imaging Analysis System
P13	Female	Customer Insights Analyst	4 years	Moderate	Sentiment Analysis Tool
P14	Male	Legal Consultant	2 years	Low	AI-based Document Review
P15	Female	Operations Coordinator	3 years	Low	Scheduling Algorithm
P16	Male	Business Consultant	5 years	High	Risk Management Decision Model

FINDINGS AND DISCUSSION

Thematic analysis of 16 semi-structured interviews showed that there are four prevalent themes that explain how users perceive, interpret, and respond to the transparency and explainability of AI decision systems. The themes were formed during systematic coding, meaning unit clustering and iterated interpretation. There are four overarching themes include Perceived Clarity and Comprehensibility of AI Explanations, Trust Calibration and Emotional Responses to Transparency, User Autonomy and Interpretive Engagement with Explanations, and Contextual and Ethical Dimensions of Transparency. The themes represent different dimensions that are connected but not the same regarding the user experience of XAI systems. The results indicate that the perceptions of users towards transparency are determined by the technical explanation readability as well as contextual, emotional and ethical aspects.

Theme 1: Perceived Clarity and Comprehensibility of AI Explanations

The former is the first significant theme connected to the knowledge of the users on how artificial intelligence systems justify their choices. In interviews, transparency was considered to be an outcome of clarity and simplicity. The respondents also constantly indicated that they preferred explanations that were not technical because they communicated the reasoning in human and straightforward, relatable terms. One participant said: "I would simply prefer that the system explains everything in plain English. I get lost and give up trying to comprehend when it is done in mathematical terms" (Participant 4). This quote summarises a familiar feeling that transparency becomes pointless when the users do not understand how to interpret it. Ribes et al. (2021) state that the interpretability of AI models can be practical only when the explanations are in line with cognitive models and the knowledge of users and the domain. Likewise, Rahman, M. (2023) states that XAI should have the ability to overcome the interpretation gap between the understanding of the technical design and interpretation by users.

According to the participants, the cognitive load required to decode layered explanations was also mentioned. In cases where explanations involved several stages of thinking, users disengaged and thought the system to be over-complicated. Most of them were thankful to visual descriptions, which included colour-coded signs or flow charts that outlined the course of action in decisions. This aligns with Grimmelikhuijsen (2023), who discovered that visual transparency promotes cognitive accessibility, particularly for non-technical end users. The other challenge stated by participants with technical backgrounds was that they found simplistic explanations to be patronising. This highlights the challenge of trying to come up with explanations which are both informative and user-appropriate, a tension which has been found as a key focus in the XAI literature (Patidar et al., 2024). Theme 1 points out that transparency is based on perceived clarity. The trustworthiness of the algorithm and the satisfaction of the users were determined by whether they felt that they genuinely comprehended the logic of the AI output or not, regardless of the complexity of the algorithm.

Theme 2: Trust Calibration and Emotional Responses to Transparency

The second theme is related to how transparency will influence the level and trust of the users with AI systems. The respondents cited transparency as a psychological intermediary of the respondents. It decided whether the respondents were relaxed or nervous around the authority of the presented system. One participant commented: "I am glad that when the system tells me the reason why it has made a decision, I feel confident. However, when it conceals such logic, I begin to doubt whether it is right or playing with something under the carpet" (Participant 7). This implies the interactive quality of trust and transparency. According to Shulner-Tal et al. (2025), individuals do not blindly trust AI but rate it in accordance with the perceived openness and control. Those respondents that expounded on system reliability placed more trust in system reliability. On one hand, opaque systems had been linked to suspicion and lack of emotions.

The participants also associated transparency with emotional safety. They identified the candidness of a system with the purity of morals, which assumed that the more open systems were not likely to have unfair and prejudiced decisions. However, other respondents noted that too much information could lead to cognitive overload that can lead to confusion and mistrust. Such is the transparency-trust paradox that is widely spread in the writings of XAI (Cetinkaya and Krakauer, 2025). Weirdly enough, the respondents provided descriptions of the cases when increased transparency had a volume of trust, particularly when the explanations were filled with scepticism or contradictory possibilities. Probabilistic reasoning was also disturbing to the users because it had an effect of disrupting the illusion of AI accuracy. An example of this is that one of the respondents said: When the system says it is 60 percent sure, I start to think, but then it is 40 percent that it is not right? That does not make me feel more trust in it" (Participant 2). This means that the language and the way in which a person describes things are very important in determining an emotional response. A case of competence and accountability yields transparency that strengthens trust, and in the case of no contextual reassurance through the explanations of the uncertainty of the algorithmic component, transparency undermines trust.

Theme 3: User Autonomy and Interpretive Engagement with Explanations

The third theme is what transparency is doing to the autonomy and the participation of the users in the decision-making. The respondents saw AI explanations as a way of

proving automated decisions or disproving them. This interpretation agency made users feel that they were the ones who participated and not passively accepted. This dynamic was explained by one of the members as follows: "When I understand the motivation behind making such a choice, I am in a position to agree to it. Otherwise, I appear to be blindly clicking on the accept button" (Participant 9). This is signatory to one of the most important lessons, that is, transparency leads to autonomy by transforming users into co-interpreters. Users who understood the explanations said that they felt powerless and more responsible. This supports the fact that interpretability enhances participatory trust as discovered by Grimmelikhuijsen (2023), that transforms human-machine relations, so it is not as dependent but a collaboration.

However, the findings also suggest that autonomy can be the source of interpretive conflict. Some of the users made explanations to the situations where they disagreed with AI even where this system was correct. These people also had a high chance of using their instinct to override the algorithmic thinking which reflects a cognitive bias and can complicate human-AI collaboration. Moreover, user autonomy was an object of contextual interests. Users did not question AI outputs too much in the case of low stakes (e.g. product recommendations). In contrast, the decisions with high stakes (i.e. lending, medical tests) were what the users demanded more detailed explanations and control. The meaning of transparency is a local, and not general value, as is demonstrated by such an interpretive gradient. The transparency in this theme illustrates that it not only creates trust but also enables users to co-produce meaning. Sound XAI systems thus provide a compromise between automation and human interpretive agency and leave the ownership of decisions to the user.

Theme 4: Contextual and Ethical Dimensions of Transparency

The fourth theme touches upon the more global ethical and contextual influences that have an impact on the perception of transparency by users. The participants did not only consider explainability as a usability feature but also as an ethical demand. Transparency was also associated with many people with fairness, accountability and respect for personal data. One interviewee stressed saying: "When the system cannot justify how it made its decision, then who will be accountable when it is wrong? I believe that transparency is not an option but a rule" (Participant 12). This quote is indicative of increased awareness of algorithmic accountability by users, which aligns with the results of Patidar et al. (2024) that ethical transparency was a key factor in people accepting AI. The respondents thought that explanations are a guard against prejudice and discrimination, especially in sensitive areas such as employment and credit rating.

Some of the users also associated transparency with data privacy. They complained when the explanations needed the revelation of the source of personal data or inference lines. This conflict between transparency and privacy, raised in recent research by Cetinkaya and Krämer (2025), implies that ethical transparency should be contextually balanced. Excessive information on data provenance will inadvertently provide sensitive information when it is excessively detailed, and it contributes to suspicion when it is insufficiently detailed. Moreover, the participants revealed the lack of institutional responsibility: although they may be able to explain them, the users wonder who can confirm that they are correct. There were some demands to have independent oversight or ethical auditing of XAI systems. This feeling highlights how the issue of transparency has been transformed into a governance one. Theme 4 puts user perceptions into a moral perspective, thus: transparency is a

social contract between machines and humans, based on fairness, privacy, and accountability.

Table 2:
Thematic Analysis of Findings

Theme	Sub-Theme	Codes	Description
Perceived Clarity and Comprehensibility of AI Explanations	Simplicity of language, Cognitive load in technical terms	"Understandable," "Confusing," "Too technical," "Readable," "Plain explanation"	Users valued clear, concise explanations and expressed difficulty when language was overly technical or abstract.
Trust Calibration and Emotional Responses to Transparency	Confidence from explanation clarity, Anxiety over algorithmic opacity	"Trustworthy," "Skeptical," "Comfortable," "Hidden logic," "Manipulated"	Emotional reactions influenced trust levels, with transparency often mediating between comfort and skepticism.
User Autonomy and Interpretive Engagement with Explanations	Decision validation, Challenge vs. acceptance of AI output	"Double-check," "Agree," "Blind trust," "Second opinion"	Users used explanations either to validate AI decisions or challenge them, shaping their engagement with the system.
Contextual and Ethical Dimensions of Transparency	Fairness and accountability, Data privacy awareness	"Fair," "Bias," "Data misuse," "Accountability"	Transparency was perceived as an ethical necessity, intertwined with fairness, accountability, and responsible data handling.

Summary of Findings

The results of the thematic analysis indicate that the perception of transparency of automated decision systems by the user is formed as a result of a complex interaction between the cognitive, emotional, and ethical aspects. Throughout the four themes, it was apparent that AI explanation comprehensibility is where trust and acceptance are developed. The participants always appreciated straightforward, easy, and personalised explanations, which made them understand the rationale of AI decision-making without becoming overwhelmed by technical explanations. Transparency was not viewed as a technical quality but as an experience that made the users feel respected and valued. In addition, feelings were an essential part of the perception of transparency; systems that seemed to be open and transparent in their communication were better understood and accepted by users, whereas systems with veiled or too technical interventions provoked distrust and anxiety. Transparency also affected the way in which the participants exerted autonomy, with most of them referring to explanations as a means to validate or criticise in order to act actively in the AI instead of passively accepting its output. Lastly, the transparency was perceived by the users in terms of morality, equitableness, and responsible data management. They highlighted that the possibility to comprehend the decision-making process and the decision making process is not only a convenience but a moral imperative that strengthens the authenticity of the AI systems. All of these points indicate that human-centred explainability in AI is most likely to succeed (an element of technical accuracy combined with social interpretation and emotional comfort).

PRACTICAL RECOMMENDATIONS

The research results suggest to create effective explainable AI systems, one has to approach the design process in a way that emphasises human understanding, emotional appeal, and ethical accountability. Transparency should be used as an

interactive communication process instead of a technical aspect (Thalpage, 2023). The explanations must be prepared using clear, easy-to-understand language, and they should be delivered in a manner that is acceptable to the different users, depending on their expertise. Images and contextual objects may help comprehend the information and decrease cognitive load, especially in the case of non-technical users (Alhasan & Alnanih, 2025). Simultaneously, system designers should not be too detailed, and at the same time, they should not be too detailed to the extent that they cause confusion and alienation to users. Transparency should also be multiplied by going beyond technical transparency to include ethical accountability (McDermid et al., 2021). The members of this study emphasised that the users should be assured that AI systems are not operated unfairly, that they are not violating their data, and that they are controlled. As such, formal policies to be implemented by organisations implementing AI must regulate the concept of explainability, data integrity, and bias reduction, with the organisation ensuring that the explanations are ethical (Yadav, 2024). Also, end users should be engaged in the process of explaining the features of explainability so that better alignment can be achieved between the output of the system and the expectations of end users. Co-design creates inclusiveness, enhances the relevance of explanations and long-term trust through the provision of a sense of ownership and collaboration in the functioning of AI systems.

FUTURE RESEARCH DIRECTIONS

The findings of the present research indicate a number of relevant avenues of future research concerning explainable AI and user transparency. The research into the influence of cultural and contextual factors on the perception of explainability requires additional investigation since the attitudes toward authority, trust, and technology differ in different societies. The study of various demographic and cultural groups would provide helpful information on the way various users understand the explanations of AI. Besides, longitudinal studies of the evolution of trust in AI during a specific period assist in determining whether transparency is a long-term phenomenon or whether it will decrease as users get used to automated systems. Concrete measures that can be used in evaluating the success of explanations should also be developed and tested in the future, specifically, the user comprehension and calibration of trust (Naiseh et al., 2024). The other important field of research is the conflict between transparency and privacy because exposing excessive information on data sources or algorithms can undermine confidentiality. Researchers should explore frameworks that are open and offer protection to sensitive data without being accountable. Finally, the necessity to learn the impact of transparency on human-AI cooperation under the conditions of real-world decision-making is increasing, particularly in high-stakes domains, such as healthcare, law, and finance (Etuk & Omankwu, 2025). Realising the distributed interpretive responsibility between users and the AI systems will be a critical point in modelling the technologies to inform and empower individuals and institutions to make ethically sound, transparent, and trustworthy decisions.

CONCLUSION

The study had the purpose of exploring the perception of transparency and understanding in those who use explainable artificial intelligence (XAI) systems. The study demonstrated that automation cannot be trusted by humans as a phenomenon of precision of algorithms but as a phenomenon of understandability

and moral responsibility in communication. Thematic analysis of the sixteen interviews which included participants revealed that explanations serve as a bridge between the technical logic and human cognitions in a bid to encapsulate abstract computations into easy to relate and comprehend anecdotes. The experience of the subjects revealed that when it is possible to make the explanations clear, foreseeable, and put it in context, users can have confidence in the outputs of AI and feel free to be critical of automated decisions. On the other hand, being too technical or too opaque makes understanding more complicated and distrustful, supporting the idea of AI as distant and unemotional.

It was also revealed that transparency is a dual process, as an informational and moral. The users associated the extent of clarity in explanation with justice, accountability, and adherence to autonomy and viewed open systems as rightful and socially responsible. These lessons support the necessity of a paradigm shift in the AI design, i.e. the shift towards user-centred interpretability instead of the model-centric explainability. To build ethical trust and social acceptance, AI explanations must be done at various levels of literacy and emotional expectations.

This study highlights that the effectiveness of XAI does not rely merely on the level of technology but also on the human experience of understanding. Through the introduction of ethical responsibility, communicative design, and participatory feedback in the development of AI, transparency can no longer be an element of compliance, but rather become a pillar of responsible human-AI cooperation. Therefore, this study adds to a growing literature that explainability is a key factor in enhancing the usability and legitimacy of automated decision-making systems, providing a guide to responsible innovation in the age of AI.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor to the research and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally in the creation of this work.
Conflicts of Interest: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Ahmed, S., Ali, A., & Mansoor, M. T. (2025). Cognitive Representations of Generative AI among Pre-Service Teachers: A Phenomenological Study in a Teacher Education College in Karachi. *The Critical Review of Social Sciences Studies*, *3*(4), 2264-2273.

Ahmed, S., Faisal, M., Hasan, S. S. F., & Ghazi, M. A. (2025). Impact of Artificial Intelligence (AI) on Teacher Self-Efficacy: A Systematic Literature Review. *Social Science Review Archives*, *3*(4), 3406-3415.

Ahmed, S., Urooj, S., Farheen, S., & Ishaq, M. (2025). Impact of Ethical AI Use on Learning Patterns of Pre-Service Teachers at a Public Sector University in Karachi. *ACADEMIA International Journal for Social Sciences*, *4*(4), 3305-3315.

Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Advanced Research and Reviews*, *18*(3), 050-058.

Alhasan, S., & Alnanih, R. (2025). Perception-Centric Explainable AI: Bridging Cognitive Theories and HCI Design for Enhanced User Experience. *Procedia Computer Science*, *257*, 297-304.

Arzu, F., Ali, R., & Muneeb, M. T (2025). The Role of FinTech Accelerators & Incubators in Nurturing Entrepreneurial Innovation: A Multiple Case Study of European Hubs. *ACADEMIA International Journal for Social Sciences*, *4*(4), 3725-3736.

Arzu, F., Sattar, M. S., Sultan, S., Abrar, K., & Khuharo, Z. H. (2025). ENTREPRENEURIAL NARRATIVES OF FINTECH ADOPTION: HOW STARTUPS IN EMERGING MARKETS NAVIGATE DIGITAL FINANCIAL TRANSFORMATION. *Journal of Management Science Research Review*, *4*(4), 473-504.

Asghari, H., Birner, N., Burchardt, A., Dicks, D., Fassbender, J., Feldhus, N., ... & Züger, T. (2021). What to explain when explaining is difficult? An interdisciplinary primer on XAI and meaningful information in automated decision-making. *Alexander von Humboldt Institute for Internet and Society*.

Ashraf, N., Arzu, F., Abrar, K., & Anwar, M. (2025). Narratives of SMEs on Access to Finance: Barriers and Opportunities in Pakistan's Banking Sector. *The Critical Review of Social Sciences Studies*, *3*(4), 262-277.

Asif, M., Jabbar, S., & Yusaf, S. (2026). Employee Perspectives on Remote Work Culture and Leadership in Hybrid Organisations. *The Critical Review of Social Sciences Studies*, *4*(1), 151-167.

Blake, H. (2024). Transparency and Explainability in AI: Ensuring Trust and Understanding in Autonomous Decision-Making Systems.

Braun, V., & Clarke, V. (2017). *Qualitative Research in Psychology Using Thematic Analysis in Psychology*. SAGE Publisher.

Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative psychology*, *9*(1), 3.

Braun, V., Clarke, V., & Hayfield, N. (2022). 'A starting point for your journey, not a map': Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis. *Qualitative research in psychology*, *19*(2), 424-445.

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & quantity*, *56*(3), 1391-1412.

Cetinkaya, N. E., & Krämer, N. (2025). Between transparency and trust: identifying key factors in AI system perception. *Behaviour & Information Technology*, 1-15.

Chiaburu, T., Haußer, F., & Bießmann, F. (2024). Uncertainty in xai: Human perception and modeling approaches. *Machine Learning and Knowledge Extraction*, *6*(2), 1170-1192.

Chinnaraju, A. (2025). Explainable AI (XAI) for trustworthy and transparent decision-making: A theoretical framework for AI interpretability. *World Journal of Advanced Engineering Technology and Sciences*, *14*(3), 170-207.

Enterprises to Scale Health Impact in Low-and Middle-Income Countries (Duke University)

Etuk, E. A., & Omankwu, O. C. B. (2025). Human-AI Collaboration: Enhancing Decision-Making in Critical Sectors. *Communication In Physical Sciences*, *12*(2), 426-433.

Ferrario, A., & Loi, M. (2022, June). How explainability contributes to trust in AI. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency* (pp. 1457-1466).

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, *83*(2), 241-262.

Gsenger, R., & Strle, T. (2021). Trust, automation bias and aversion: Algorithmic decision-making in the context of credit scoring. *Interdisciplinary Description of Complex Systems: INDECS*, *19*(4), 542-560.

Han, X., Xiao, S., Sheng, J., & Zhang, G. (2025). Enhancing efficiency and decision-making in higher education through intelligent commercial integration: Leveraging artificial intelligence. *Journal of the Knowledge Economy*, *16*(1), 1546-1582.

Hole, L. (2024). Handle with care; considerations of Braun and Clarke's approach to thematic analysis. *Qualitative research journal*, *24*(4), 371-383.

Jabbar, S., & Gul, M. (2026). Impact of Organizational Justice on Organizational Commitment and Organizational Citizenship Behavior with the Mediating Effect of Organizational Trust. *Social Science Review Archives*, *4*(1), 808-828.

Kausar, T., & Ahmed, S. (2026). Impact of Artificial Intelligence (AI) on Educational Leadership: A Systematic Literature. *ACADEMIA International Journal for Social Sciences*, *5*(1), 97-106

Khan, H. A., Butt, M. A. A., Noor, S., Ishaq, S. M., & Siddiqui, F. (2025). Redefining Pedagogy: Investigating the Intersection of Technology Integration and Traditional Teaching Methods in Pakistan's Educational Research Landscape. *The Critical Review of Social Sciences Studies*, *3*(4), 1543-1558.

Khan, H. A., Kausar, T., & Khan, S. (2025). Leading AI-Integrated Pedagogies: A Qualitative Study on How School Leaders Guide Teachers in Using AI Tools for Teaching and Learning in Pakistan. *Social Science Review Archives*, *3*(4), 3086-3101.

Khuharo, Z. H., Ayub, M. J., & Ali, R. (2025). How Investors Interpret Risk in the Pakistan Stock Exchange: Lived Experiences of Retail Traders. *Social Science Review Archives*, *3*(4), 3281-3297.

Kloker, A., Fleiß, J., Koeth, C., Kloiber, T., Ratheiser, P., & Thalmann, S. (2022). Caution or trust in AI? How to design XAI in sensitive use cases?.

Maarten Schraagen, J., Kerwien Lopez, S., Schneider, C., Schneider, V., Tönjes, S., & Wiechmann, E. (2021, September). The role of transparency and explainability in automated systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 27-31). Sage CA: Los Angeles, CA: SAGE Publications.

Marocco, S., Barbieri, B., & Talamo, A. (2024). Exploring facilitators and barriers to managers' adoption of AI-based systems in decision making: A systematic review. *AI*, *5*(4), 2538-2567.

McDermid, J. A., Jia, Y., Porter, Z., & Habli, I. (2021). Artificial intelligence explainability: the technical and ethical dimensions. *Philosophical Transactions of the Royal Society A*, *379*(2207), 20200363.

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, *55*(5), 3503-3568.

Morrison, K., Spitzer, P., Turri, V., Feng, M., Kühl, N., & Perer, A. (2024). The impact of imperfect XAI on human-AI decision-making. *Proceedings of the ACM on human-computer interaction*, *8*(CSCW1), 1-39.

Naiseh, M., Simkute, A., Zieni, B., Jiang, N., & Ali, R. (2024). C-XAI: A conceptual framework for designing XAI tools that support trust calibration. *Journal of Responsible Technology*, *17*, 100076.

Patidar, N., Mishra, S., Jain, R., Prajapati, D., Solanki, A., Suthar, R., ... & Patel, H. (2024). Transparency in AI decision making: A survey of explainable AI methods and applications. *Advances of Robotic Technology*, *2*(1).

Praveenraj, D. D. W., Victor, M., Vennila, C., Alawadi, A. H., Diyora, P., Vasudevan, N., & Avudaiappan, T. (2023). Exploring explainable artificial intelligence for transparent decision making. In *E3S Web of Conferences* (Vol. 399, p. 04030). EDP Sciences.

Rahman, M. (2023). Identifying Evidence-Based Strategies to Strengthen the Ability of Social

Rehmat, M. A. R., Hassan, H., Rumaan, M., & Abrar, K. (2025). IT Sector Transition to Agile: Managerial Challenges and Cultural Dynamics in Pakistan's Technology Firms. *Social Science Review Archives*, *3*(4), 479-494.

Rehmat, M. A., Hassan, H., Khan, H. A., & Abrar, K. (2025). Artificial Intelligence in the Classroom: Teachers' Lived Experiences and Ethical Concerns in Educational Integration. *ASSAJ*, *4*(02), 629-645.

Rehmat, M. A., Hassan, H., Rumaan, M., Baig, J., & Abrar, K. (2025). HUMAN-CENTRED EXPLAINABLE AI IN EMERGING MARKETS: TRUST AND CONFIDENCE AMONG NON-TECHNICAL USERS IN PAKISTAN. *Annual Methodological Archive Research Review*, *3*(10), 145-163.

Ribes, D., Henchoz, N., Portier, H., Defayes, L., Phan, T. T., Gatica-Perez, D., & Sonderegger, A. (2021, August). Trust indicators and explainable AI: A study on user perceptions. In *IFIP Conference on Human-Computer Interaction* (pp. 662-671). Cham: Springer International Publishing.

Schemmer, M., Kühl, N., & Satzger, G. (2021). Intelligent decision assistance versus automated decision-making: Enhancing knowledge work through explainable artificial intelligence. *arXiv preprint arXiv:2109.13827*.

Shulner-Tal, A., Kuflik, T., Kliger, D., & Mancini, A. (2025). Who made that decision and why? Users' perceptions of human versus AI decision-making and the power of explainable-AI. *International Journal of Human–Computer Interaction*, *41*(7), 4230-4247.

Sokol, K., & Flach, P. (2024). Interpretable representations in explainable AI: from theory to practice. *Data Mining and Knowledge Discovery*, *38*(5), 3102-3140.

Stamboliev, E. (2023). Proposing a postcritical AI literacy: Why we should worry less about algorithmic transparency and more about citizen empowerment. *Media Theory*, *7*(1), 202-232.

Thalpage, N. (2023). Unlocking the black box: Explainable artificial intelligence (XAI) for trust and transparency in aisystems. *J. Digit. Art Humanit*, *4*(1), 31-36.

Yadav, B. R. (2024). The ethics of understanding: Exploring moral implications of explainable AI. *International Journal of Science and Research (IJSR)*, *13*(6), 1-7.

Zeb, S. S., Abrar, K., Saqib, S., & Rizvi, S. A. A. (2025). Economic dimensions of health protection under Sehat Sahulat Program: A qualitative assessment of urban and rural households in Pakistan. *ACADEMIA International Journal for Social Sciences*, *4*(4), 2259-2273.