Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

HUMAN-CENTRED EXPLAINABLE AI IN EMERGING MARKETS: TRUST AND CONFIDENCE AMONG NON-TECHNICAL USERS IN PAKISTAN

Dr. Muhammad Asim Rehmat*
Assistant Professor, Department of Computer Engineering University of Engineering and Technology, Lahore Main Campus, Pakistan Email: asimrehmat@uet.edu.pk

Hammad Hassan
Manager R&D Al-Khwarizmi Institute of Computer Science (KICS) University of Engineering and Technology Lahore, Pakistan Email: hammad.hassan@kics.edu.pk

Muhammad Rumaan
Lecturer, Department of Computer Science, University of Management and Technology Lahore, Pakistan Email: muhammad.rumaan@umt.edu.pk

Jahangir Baig
PhD Scholar, Department of Statistics, University of Karachi, Karachi, Pakistan Email: jahangirbaig83@gmail.com

Dr. Kashif Abrar
Visiting Faculty Member, Department of Business Studies Bahria University, Islamabad, Pakistan Email: kashif_411@hotmail.com

ABSTRACT

Building user confidence through explainable artificial intelligence (XAI) in non-technical sectors, specifically in the country of Pakistan, is the focus of this research. It seeks to assess the implications of AI system explainability and the influence of trust and understanding on professional acceptance in mainstream use. The qualitative research design involved semi-structured interviews as a means of gathering data from the four sectors of education and healthcare, finance, and public administration, and XAI in these sectors in Pakistan. Participants involved in the research were twelve in number. Thematic analysis was conducted on the perceptions and experiences to evaluate the increasing trust and confidence in the XAI tools.

User confidence depends on the extent to which the clarity, relevance, and transparency of AI explanations are evaluative. When explanations were intuitive and articulated in human-centered ways, trust and participation were high. Confidence and participation, on the other hand, were limited if the explanations were convoluted and impenetrable. These included trust and transparency, human-centered design, cognitive ease, specific domains, and ethical constraints. Different from the other studies, this study explains that the psychological and contextual aspects are equally

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

important in user trust and must be incorporated into explainability beyond the technical aspects of XAI tools. This study underscores the importance of resilient, culturally appropriate systems in which XAI tools focus on a balance between explainability and automation. Such design practices will improve trust and transparency of human–AI interactions in the emerging markets of Pakistan.

Keywords: Explainable Artificial Intelligence (XAI), User Confidence, Trust, Human-Centred Design, Pakistan.

Introduction

Background of the Research

Artificial Intelligence (AI) is prevalent across diverse areas - both technical and non-technical - and is influencing decisions, innovations, and productivity across various domains (Ejjami, 2024). However, the more advanced AI systems become, the opaquer their decision-making systems become, which results in complex systems experiencing what is referred to as the 'black box' problem (Brożek et al., 2024). This lack of clarity leads to issues of trust, accountability, and transparency, especially from non-technical stakeholders who are unable to decipher complex systems. Explainable Artificial Intelligence (XAI) is the response to the growing demand for understanding the AI decision (Minh et al., 2022). XAI is the field of AI to provide the necessary interpretability, transparency, and rationale to decision-making systems, so the users understand the why and how of the system's outcome (Teixeira et al., 2025).

AI's utility is expanding in Pakistan's healthcare, education, public finance, and governance sectors. Per Pakistan's Ministry of IT and Telecom, over 40% of large corporations in Pakistan utilise AI decision-making tools (Khan & Ullah, 2024). However, only 15% of employees understand these tools. This suggests a gap in implementation and understanding, especially for non-tech professionals like managers, educators, or administrators. This gap in understanding can erode decision-making confidence, scepticism, or complete disengagement from AI usage. Therefore, in developing economies, where digital literacy can vary, XAI's potential in enhancing confidence is a significant area of research.

Problem Statement

User trust in AI systems lags behind the advancements in AI technologies. This is true especially for emerging economies such as Pakistan, where the users are also non-specialists. Because users often lack the technical knowledge to comprehend automated decision-making processes, distrust and unease are common. Although Explainable AI (XAI) initiatives are designed to bridge the knowledge gap, most existing works and appraisal structures focus on the explainability needs of technical users and engineers rather than general users (Mohseni et al., 2021). Most of the existing works abstract the cultural, educational, and contextual dimensions embedded in the non-technical users' society that explain the gaps in understanding AI explainability. Therefore, there is a gap in understanding the trust and confidence XAI systems are designed to provide for non-technical users in Pakistan's socio-technological context.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Research Questions

How do non-technical practitioners in Pakistan understand AI systems' explainability?

How does Explainable Artificial Intelligence (XAI) bolster confidence and trust among non-technical users?

What challenges and opportunities are present concerning the implementation of XAI structures concerning non-technical users in Pakistan?

Research Objectives

The study aims to achieve the following objectives:

To understand non-technical users' perceptions of AI explainability concerning trust and confidence

To investigate the ways XAI strengthens the trust and reliability of AI systems for non-technical users.

To analyse the primary challenges and opportunities in the implementation of user-focused XAI in Pakistan

Significance of the Study

XAI holds an extreme importance in the literature on trust within AI systems, particularly focusing on user confidence within the context of non-technical users in Pakistan, and more broadly, in the Global South (Dahri et al., 2025). Existing literature primarily adopts a Western viewpoint. In focusing on non-technical professionals from a multitude of industries, the study effectively highlights how the explainability of AI systems fosters acceptance and trust. This is a central issue within the context of the Global South. This research will assist policymakers, system designers, and businesses in creating more equitable and interpretable designs and AI systems intended for non-technical users. This work also advances the discourse on user-centred design and trust calibration in AI systems, particularly the integration of predictive trust violations, as emphasised in recent literature (Shin, 2021).

This qualitative research describes an interpretivist paradigm underscoring the meaning and value of lived experiences and perceptions of non-technical professionals interfacing with artificial intelligence systems. In-depth semi-structured interviews were conducted with 12 respondents from different sectors (education, finance, healthcare, and administration) in metropolitan towns of Pakistan. Data was analysed manually via thematic coding. Using thematic analysis enhanced the opportunity for subjective interpretation, focusing on the richness of the data via the respondents' explanations instead of minimising to numbers. Ethical parameters to the letter were observed, including confidentiality and informed consent. Personal experience was the methodological emphasis, which is why this study is well-positioned to examine the impact of culturally and technologically diverse environments like Pakistan on user confidence in explainable artificial intelligence (XAI) systems.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Literature Review

Theoretical Framework

The theoretical framework for this study is based on trust calibration and human-centred design theories which contextualise the role of explanations on users' mental models and users' trust on the outputs of the algorithm (Shin, 2021). Trust calibration theory posits that users will have the 'right' trust when their mental models of a system are accurate as to what it can or cannot do; therefore, the explainability models will bridge perceived trust on a system and trust that the system is reliable (Wischnewski et al., 2023). The human-centred XAI literature further details that, for a non-technical user, effective explanations call for the consideration of the user's goals, cognitive load, domain knowledge, and distance to the goal (Rong et al., 2023).

Socio-technical frameworks that incorporate cognitive, organisational, and technological aspects provide a richer understanding of the different user groups and the ways explainability can provide actionable confidence (Naveed et al., 2024). Causability (the explanation's ability to articulate causal relationships in human-understandable form) is a theoretical concept that adds to the discourse on explanation transparency and the instructional aspects of explanations (Shin, 2021). Explanation goodness and user satisfaction are constructs, in empirical XAI evaluations, that mediate between explanation design and trust (Hoffman et al., 2023).

Explainability and Trust in Human-AI Interaction

While explainability is often conceptualised as a tiered requirement of trust, systematic studies demonstrate a more complex relationship in which some explanations do not foster calibrated trust among laypersons (Wiratsin & Ragkhitwetsagul, 2025). Systematic reviews of trust-related studies highlight differing trusts and understandings attributed to specific explanations, including feature attribution, counterfactuals, and rule-based narratives, although evidence remains mixed as to which explanation is most effective in a given context (Kalasampath et al., 2025). User-centred evaluation studies advocate trust measurement as a multidimensional construct to include indicators like mental models, satisfaction, and task performance rather than relying on single-item trust scales (Hoffman et al., 2023).

Empirical guides focus on co-design efforts in which explanation content is aligned to users' mental models and task contexts as a major determinant of whether an explanation is likely to alleviate cognitive uncertainty and cognitive load (Naveed et al., 2024). Variability in user expertise as a cognitive resource is a source of explanation deficit; users with a non-technical role in an organisation typically like contrastive and example-based explanations that relate to the administrative or business contexts of the model outputs (Rong et al., 2023). The importance of transparency in the trustworthiness of XAI systems extends to the provenance of the data and limitations of the system, in addition to transparency in the model internals (Chamola et al., 2023).

Cognitive Load, Visualisation, and Confidence

One of the methods to enhance the interpretability of AI models involves the use of visualisations. However, the explainability of visualisations engages and interacts

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

with the cognitive load of the user and the confidence the user places in the model (Hudon et al., 2021). It has been shown that the design of visual explanations can diminish the user's cognitive load and bolster subjective confidence. However, visualisations that are poorly designed may lead to an overestimation of understanding (Hudon et al., 2021). Therefore, the literature appears to prioritise the goal of explanatory parsimony, which involves the use of only the minimum relevant information, over the goal of technical transparency (Naveed et al., 2024). The cognitive load measures that predict whether explanation visualisations will lead to improved confidence suggest at best superficial acceptance (Hoffman et al., 2023). In the user studies conducted in industry, decision-makers expressed a preference for explanations to be provided in a layered format, with the option of progressive disclosure, consisting of a high-level explanation with the main details followed by optional details (Rong et al., 2023). In systematic literature reviews of XAI techniques, the lack of validation of visualisation techniques with end users has been noted (Wiratsin & Ragkhitwetsagul, 2025). In the design of visualisations, this has created a gap between the claims of visual design and the evidence demonstrating that miscalibrated trust has decreased.

XAI Evaluation Methods and Metrics

Evaluating XAI interventions requires the integration of objective performance indicators and subjective user-centred indicators, as comprehensive methodological surveys strongly recommend (Brdnik et al., 2025). For the applied sciences overview, the provision of multi-modal frameworks that combine task completion, comprehension tests, mental model elicitation, as well as affective dimensions of confidence and satisfaction is essential (Naveed et al., 2024). The combination and integration of qualitative approaches, such as think-aloud protocols, semi-structured interviews, and thematic analysis, converge toward understanding how non-technical audiences comprehend the explanations and how these explanations impact trust (Rong et al., 2023).

Literature enumerates the measures, explanation goodness, curiosity, and human-AI performance metrics, capturing diverse dimensions of the influence of explanations on user outcomes (Hoffman et al., 2023). Systematic reviews highlight that several XAI evaluations are still limited to the use of artificial tasks, small-scale lab samples, and homogeneously technical participants, which have drastically reduced ecological validity for non-technical professional sectors. For the organisational context, trust is built over time through repetitive interactions rather than through one-off interactions. For this reason, the need for longitudinal and contextualised evaluations rests primarily on these environments (Naveed et al., 2024).

Sectoral Applications and Contextual Constraints

The research done in construction and management, as well as applied energy, illustrates how particularities of a given area tend to shape the framing and adoption of XAI systems (Teixeira et al. 2025; Machlev et al. 2022; Love et al. 2023). For XAI systems in energy management, those frameworks which transform outputs of the model into operational recommendations of a more actionable kind tend to prove more reliable to system managers, especially when explanations of the output model

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

reveal constraints of the system (Teixeira et al. 2025). The same reasoning applies to the reviews in power systems, where causally coherent explanations in a safety-critical system, as well as rigorous validation, must meet practitioner needs (Machlev et al. 2022).

Nikiforidis et al. (2025) explain that system operators will more willingly act on predictive maintenance model alerts when the system failure explanations and the alerts point to observable, and explainable, behaviours of the machine. The construction industry XAI use case explains the need for XAI systems to meet the demands of technical accuracy, as well as the need for simplified narratives, to enable managerial decisions in environments characterised by high uncertainty (Love et al. 2023). These sector-specific studies suggest that the designs of XAI should fit into the industry workflows.

Human-Centred Design and Co-creation in XAI

A notable focus in participatory design processes is the need to engage target users while explanations are in the process of being formulated (Rong et al., 2023). As pointed out in human-centred co-creation practices surveys, design activities, including workshops, prototype-testing, and iterative feedback loops, are considered helpful in crafting explanations that users prefer and which adapt to their contextual decision-making frameworks (Naveed et al., 2024). The research regarding explainable interface design presents critical modular interface design heuristics. It suggests designing interface components that explain complex systems to non-technical users and detailing affordances for trust balancing (Naiseh, 2021). Although users are likely to build accurate mental models of AI systems and of trust calibrated to the AI system outputs when explanation design employs active participatory design (Naiseh, 2021) Ethical user autonomy, as well as informed consent standards, in decision-making systems are met with the integration of participatory design approaches (Chamola et al., 2023).

Measures, Standards, and Trustworthy XAI

There is increased demand for standardised processes and justifiable explainability claims, which draws regulatory actions nearer (Chamola et al., 2023). Trustworthy XAI reviews indicate that accountability, documentation, explanation, and audit trails must be incorporated with other aspects of transparency—otherwise, it holds no real value (Chamola et al., 2023). To assess the dependability of the model output, XAI technologies are suggested to provide users with provenance metadata and uncertainty quantification (Kalasampath et al., 2025). The rapid proliferation of diverse XAI tools is generating problems with comparability, and the community is seeking common frameworks and harmonised evaluation methods that strike a reasonable balance between the technical and the human dimensions (Wiratsin & Ragkhitwetsagul, 2025).

Synthesis of Variable Combinations: Explainability, Trust, Cognitive Load, and Context

In the review of current discourse, the relationships between explainability, trust, cognitive load, and context emerged as the most complete framework to account for

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

user trust in AI. Empirical research suggests that explanations which alleviate cognitive load and increase task-relevant information are more likely to yield calibrated trust (Hudon et al., 2021). Studies in specific sectors find that contextualised explanations strengthen trust perceptions more than abstract technical explanations (Teixeira et al., 2025; Nikiforidis et al., 2025). Methodological reviews go further in claiming that insights into the operation of these variable combinations in practice are best captured through mixed approaches that incorporate subjective confidence and qualitative elicitation of the mental model (Naveed et al., 2024).

Gaps in Existing Research

The examined sources contribute significantly to the field. Nevertheless, there are considerable gaps that limit the relevance for non-technical professionals in Pakistan. Most evaluation studies have a bias towards Western, high-resource contexts and technical participants. This severely limits the cultural and contextual transferable value (Akbar, Abrar, & Khan, 2022; Wiratsin & Ragkhitwetsagul, 2025). Also, there are a few longitudinal studies assessing how trust "builds" or "develops" over time in repeated AI interactions, especially in the organisational context, regarding explanation-induced confidence in trust.

Additionally, the lack of empirical studies on resource-poor or low-bandwidth contexts, particularly where visualisation-heavy, explanation- scaffolding may be inapplicable, is a significant omission for developing countries (Machlev et al., 2022). Furthermore, while numerous sources suggest evaluation frameworks, there remains a lack of consensus on unified, standardised instruments, particularly for non-technical professionals, resulting in a lack of cohesion in how trust and explanation "goodness" is measured (Hoffman et al., 2023). Also, research that directly addresses socio-cultural aspects such as institutional trust, hierarchical norms of decision-making, or local digital literacy that impact mediating XAI is scarce (Rong et al., 2023). This highlights the need for research that is user-sensitive and context-sensitive, that elucidates how explainability frameworks operate for non-technical professionals in Pakistan organisational contexts.

Shin (2021), Hoffman et al. (2023), and Naveed et al. (2024) investigate explainability and its effects on trust, cognitive load, and decision-making. They address its impact through user-centred design, contextualisation, and evaluative rigour, ultimately suggesting design explainability is a multidimensional construct. Sector-specific studies demonstrate how domain-aligned explanations correlate to increased perceived reliability. However, as discussed in Teixeira et al. (2025) and Machlev et al. (2022), there remain contextual and methodological gaps that would allow for transfer to non-technical professionals in Pakistan. As a first step to bridging the identified gaps, this qualitative study seeks to achieve participatory design through thematic analysis of interviews as described in the preceding chapter, adapted to the local infrastructural and cultural context of Pakistan, particularly for non-technical professionals.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Methodology

Research Design

Qualitative research under the interpretivist paradigm was employed to study the impact of Explainable Artificial Intelligence (XAI) on user confidence of non-technical professionals in Pakistan (Braun & Clarke, 2017). A qualitative approach prioritises understanding the unscripted views, constructions, and lived experiences of individuals over merely quantifying the metric of some variable (Abrar et al., 2024; Ashraf et al., 2025). From the interpretive standpoint, user confidence and trust in AI was seen as a socially constructed phenomenon through interaction and individual cognition (Naveed et al., 2024). The predominant method of inquiry was semi-structured interviews, which allowed the participants to express their experiences within the ample thematic flexibility provided in the guiding framework, while the overarching research aim was to capture their understanding of trust and confidence in the explainability of AI within their professions.

This study targeted a population of professionals across educational, healthcare, financial, and administrative sectors who engage directly with AI-enabled tools. This study's decision to focus on non-technical professionals—who use AI technology but are not involved in designing it—was to capture the everyday user perspective on explainability frameworks, as opposed to the viewpoint of a technical expert. Such a qualitative design enhances the capturing of rich and detailed accounts that are necessary to provide a contextualised account of users' trust in explainable AI systems (Braun & Clarke, 2022).

Sampling Strategy

The focus of the study on particular individuals led to purposive sampling that met the inclusion criteria that aligned with the research goals (Braun & Clarke, 2022). The three main participant criteria included individuals who worked in organisations that operate AI systems, individuals who have worked with AI decision systems, and individuals who lack formal training in technical and/or computing sciences. This rationale provided a view of the intended population of non-technical AI users whose trust in AI systems was primarily shaped through algorithmic exposure as opposed to end-user experiences (Rong et al., 2023).

A total of 12 participants were recruited from diverse professional backgrounds in Karachi, Lahore, and Islamabad, reflecting the country's major urban centres where AI adoption is most pronounced. These included educators using AI-based learning platforms, healthcare administrators managing predictive analytics dashboards, financial officers employing algorithmic forecasting tools, and public-sector employees interacting with digital service automation systems. The sample size was sufficient to achieve data saturation, where additional interviews no longer yielded new thematic insights.

Recruitment occurred through professional networks and organisational referrals. Participants received a concise study brief explaining the research purpose, confidentiality assurances, and interview format. Voluntary participation was emphasised, and no financial incentives were offered, preserving the authenticity of responses. This sampling strategy produced a heterogeneous yet thematically coherent set of perspectives relevant to the study's research questions (Alordiah & Oji, 2024).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Data Collection

Data collection relied on semi-structured interviews conducted in person or, when appropriate, on secure digital platforms such as Zoom and Microsoft Teams. Individual interviews were approximately 45-60 minutes long. Each interview was organised around three primary topics: AI explainability, trust and confidence in AI outputs and decisions, and the difficulties in understanding AI outputs. Each interview guide maintained a question structure of open-ended questions within the confines of the established boundaries of prior XAI scholarship, focusing on constructs such as mental model formation, trust transparency, and trust calibration (Hoffman et al., 2023).

Probing questions encouraged participants to elaborate and generalise, drawing on examples from their own experiences. Interviews were recorded, and transcriptions were completed to the letter after being recorded, for the sake of fidelity, with the provided participant permissions. The researcher also went on to compose brief field notes in the case of certain contextual observations and the non-verbal behaviour of the participants, which were important to incorporate in the subsequent analysis.

In consideration of linguistic and cultural accessibility, interviews were carried out in English or Urdu, dependent on the preference of the participants, with translated Urdu responses in English for the analysis. As culture and language were reduced, participants were able to respond more freely; this flexed interpretivist principles. Interpretivism prioritised subjective responses and meanings rather than standardised or normative responses.

Data Analysis

The qualitative data were organised and interpreted using thematic analysis. There are six steps to this type of analysis, which include: data familiarisation, initial coding, theme creation, theme review, definition of the theme, and synthesis into narrative findings. Manual coding ensured close engagement with the data as well as the identification of patterns, both linguistic and conceptual, that more automated coding systems would miss. The initial reading of the transcripts provided the researcher with preliminary ideas and keywords on explainability, trust, and user confidence. The codes were created in participants' own words and focused on the perceptions of inclusion, understandability, and trustworthiness. For example, coded phrases of "I understand why it gave this result" or "I don't trust it because it's unclear" were grouped with clarity of explanation and perceived opacity.

Sub-codes were organised into broader thematic categories, which articulated important aspects of the research questions. Five main themes were generated: (1) Trust and Transparency in Explainable AI (XAI), (2) Human-Centred Design and User Interaction, (3) Cognitive and Psychological Dimensions of Explanation, (4) Domain-Specific Applications of XAI, and (5) Ethical, Practical, and Future Challenges in XAI Implementation. Further development of the themes was done via comparison of the transcripts to establish thematic coherence and discreteness. Each theme in the findings section was illustrated with representative quotes. The interpretive synthesis prioritised the construction of meaning over the enumeration of themes. It also compared the themes with the theoretical insights obtained in the

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

literature review to identify points of convergence and divergence with existing frameworks, thereby marrying empirical findings with theoretical exploration. This analysis illustrated how explainability mechanisms influenced confidence among non-technical professionals within the organisational contexts of Pakistan.

Ethical Considerations

From the design stage to implementation, the ethical integrity of the research was prioritised. Participants were given a description of the study, their rights, and the voluntary nature of their involvement—after which they provided informed consent. Participants were assigned pseudonyms, and all identifying details were removed from transcripts to ensure data anonymity and confidentiality. This data, along with the other materials for the research, was securely stored in password-protected folders, which only the researcher could access.

Participants were free to leave the study whenever they wanted for any reason they chose. Care was taken to ensure that no information was passed to supervisors and that no information about the study was integrated with the performance appraisal of the participant. Ethical clearance for the study was obtained from the university's home institutional review board. At all times, the researcher took into consideration the cultural communication patterns, professional hierarchies, and norms relevant to the study population.

All of the ethical principles required in the study were fulfilled, including respect, beneficence and justice. This resulted in the participants feeling safe during the sharing of their experiences concerning user confidence in ai systems. Ethical considerations were also observed in the reporting, with neutral interpretations, an authentic representation of the participants, and the ethical use of their data.

Participants' Profile

Table 1 shows the diversity of participants in terms of sectors, gender, and years of experience. This also shows the range of non-technical professionals working with AI in Pakistan. The different AI applications the participants used, such as adaptive learning tools and predictive analytics, formed a rich dataset to study the impact of explainability on user confidence.

Table 1:
Participant Profiles

Participant	Sector	Job Role	Gender	AI Tool Used	Experience with AI (Years)
P1	Education	Lecturer	Female	Adaptive learning system	2
P2	Finance	Credit officer	Male	Fraud detection dashboard	3
P3	Healthcare	Administrator	Female	Patient data analytics	2
P4	Public administration	Officer	Male	Digital workflow assistant	4

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

P5	Education	Curriculum coordinator	Female	Predictive student performance model	1
P6	Banking	Relationship manager	Male	Customer profiling tool	3
P7	Retail	Marketing manager	Female	AI-driven recommendation system	2
P8	Telecommunications	HR analyst	Male	Resume screening software	2
P9	Logistics	Operations supervisor	Male	Route optimisation model	3
P10	Healthcare	Medical data analyst	Female	Predictive diagnostic interface	4
P11	Public administration	Policy analyst	Male	AI-based policy simulation	2
P12	Education	Academic coordinator	Female	Online exam proctoring system	1

Ensuring methodological rigour through transparency, reflexivity, and triangulation was key. Documentation of coding and analytical pathways facilitated transparency. As for reflexivity, continuous self-assessment of positionality was crucial, particularly because the interpretation of data was influenced through the researcher's lens of XAI. Components of cross-sectoral sampling and comparative analysis of participant stories formed the basis of triangulation and reinforced the reliability of the themes that emerged.

The qualitative approach provided the study with the opportunity to illuminate the specific challenges "explainability" mechanisms pose in garnering trust and confidence in AI among non-technical professionals in Pakistan. In focusing on the participants, the methodology embraced the human side of AI adoption, highlighting for the first time the interpretability, clarity, and contextual relevance of the explanations that may trigger users' sustained confidence in intelligent systems.

Findings and Discussion

The thematic analysis resulted in five themes: (1) Trust and Transparency in Explainable AI (XAI), (2) Human-Centred Design and User Interaction, (3) Cognitive and Psychological Dimensions of Explanation, (4) Domain-Specific Applications of XAI, and (5) Ethical, Practical, and Future Challenges in XAI Implementation. These five themes help in understanding the XAI systems and the users' multifaceted relationship. Explainability influences the users' trust and their adoption of the technology, varying in numerous contexts.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Theme 1: Trust and Transparency in Explainable AI

Trust and transparency were highlighted repeatedly regarding human-AI communication. One participant asserted, "If I can't understand how the model made its decision, I can't rely on it—even if it's statistically correct." This finding resonates with that of Wiratsin and Ragkhitwetsagul (2025) that as users understand the explanation, trust in the machine model increases. Trust and understanding the decision model of the AI system go beyond the algorithm and require communication that fits human cognitive expectations (Wysocki et al., 2023).

Teixeira et al. (2025) showed that in energy management systems, an explainable AI framework boosted confidence in operators because it generated interpretable visualisations of models' decisions. Likewise, Chamola et al. (2023) explain that the AI system's "explainability" mitigates "algorithmic opacity" and, thus, enhances human oversight and eases pushback against AI decisions. The thematic analysis showed that trust in XAI is a cycle: transparency produces understanding, and acceptance builds trust (Hoffman et al., 2023).

Naveed et al. (2024) proposed structured evaluation criteria where trust calibration is the core focus. Users will neither over trust nor under trust AI outputs when explanations are appropriately balanced between excessive detail and under-simplicity. In this study, some participants described transparency as a "gateway to accountability." This idea resonates with the need for ethical and co-technical transparency to facilitate user trust. This aligns with Naiseh (2021), who proposed the C-XAI design methodology in which XAI frameworks exhibit a human-centred interface design that balances cognitive load and trust calibration.

Theme 2: Human-Centred Design and User Interaction

From the interviews, the growing focus on the need for human-centred design to create effective explainability emerged as yet another key theme. One user stated, "We need AI that talks like a collaborator, not a calculator." This connects to the user-centred inequalities described by Rong et al. (2023) in XAI. Users described interfaces for explanations that do not simply provide a rationale to the user, but also allow the users to interrogate, experiment with, and authenticate the reasoning of the system. Naiseh (2021) described the importance of iterative co-design frameworks where end-users determine how explanations will be presented. In a similar vein, Hudon et al. (2021) argued that cognitive load and confidence disproportionately affect poorly designed visualisation strategies. These findings align with Hoffman et al. (2023) and the proposed multidimensional approach to explanation "goodness," including curiosity and satisfaction as key elements alongside accuracy.

The construction of interfaces following human-centred XAI design entails the consideration of emotions. For instance, Love et al. (2023) reported on construction engineers whose explanation preferences aligned with patterns of human reasoning, including the provision of stepwise, causally linked justifications, instead of statistical justifications. This design principle shifts the role of AI systems to 'transparent collaborators', and as a result, augmentation of participation, engagement, and shared decision-making is observed (Zhu et al., 2025).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Theme 3: Cognitive and Psychological Dimensions of Explanation

Participants' struggles with AI reasoning brought up the sub-themes of cognition and perception. Participants' discomfort and confidence variances were a result of the explanation's intricacy. One thoughtful critic observed: "Too much detail feels like noise, too little feels like magic." Shin (2021) highlighted this cognitive balancing act in users' perception of fairness, trust, and acceptance in response to explanations aligned with their causal mental models. This issue of mental model alignment was further developed by Hoffman et al. (2023) as explanation alignment theory, whereby AI explains parallel users' cognitive schemas, resulting in trust stabilisation and comprehension improvement. Visualisation quality directly relates to cognitive load, as corroborated by Hudon et al. (2021). In this study, participants made it clear that explanations must bridge the gap between technical interpretability and understandability for humans.

On another note, curiosity was incorporated as a psychological mediator. Hoffman et al. (2023) portray the users' curiosity as a psychological catalyst for deeper engagement and trust in XAI systems. Nevertheless, Naveed et al. (2024) emphasise that boundless curiosity, given a free rein, can, in fact, cause trust erosion due to cognitive overload. All in all, findings from the study reinforce the idea that XAI should prioritise the balance between cognitive overload and an information-saturated XAI. To help cognitive processing, XAI should, as Machlev et al. (2022) state, prioritise "human intelligibility over technical transparency."

Theme 4: Domain-Specific Applications of XAI (Energy, Manufacturing, and Construction)

The last theme detailed the participants' illustrative descriptions of how practical frameworks encapsulate the explainability. Illustrative descriptions detailed how explainability transcends practical boundaries. Participants exhibited an understanding that the practice context modifies the explanations needed and what users expect. An example from the energy sector explained, "In power systems, I don't need storytelling; I need clarity on risk and control." Similar insights were noted in Teixeira et al.'s (2025) study on automated building management, where real-time decision visualisation helped engineers understand and adjust AI-driven operations. In manufacturing, Nikiforidis et al. (2025) reported that XAI facilitates trust in Industry 4.0/5.0 environments through auditable machine decision making. Participants of this study also noted that transparency helps to close trust gaps between operators and autonomous systems.

Visual dashboards that present the causal factors that lead to machine failures have helped to gain acceptance by the operators. Similarly, Machlev et al. (2022) noted that XAI in energy systems enhances safety and compliance, provided the explanations given are actionable. Love et al. (2023) extended this reasoning to the field of construction, stating that XAI can assist on-site decision making, risk evaluation, and stakeholder communication when suitable interpretability frameworks are created and adapted to the specific domains. This was echoed in the participants' comments: "Explainability is not one-size-fits-all—it must speak the language of the field." This demonstrates that the effective XAI needs to implement contextualised model transparency to meet the regulatory and cognitive demands of the specific sector.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Theme 5: Ethical, Practical, and Future Challenges in XAI Implementation

Participants had reservations regarding XAI's challenges, both ethical and practical, even if some had a positive outlook on its capability. As one interviewee said, "We speak about explainability, but sometimes it is just a veneer—companies still use it as a checkbox for compliance." This resonates with Chamola et al. (2023), who indicated that explainability may need to be coupled with other attributes to avoid it being symbolic. Ethically, explainability requires more than transparency regarding a model; there also has to be a discourse regarding data sources, attribution, bias, and the use case (Balasubramaniam et al., 2022). Unfortunately, as yet, there has been no discourse on the ethical use of AI (Rong et al., 2023). Other practical challenges discussed were computation limits, user misinterpretation, and the problems of scalability. Machlev et al. (2022) noted that XAI deployed in energy grid systems may add latency that reduces the systems' operational efficiency. Balancing interpretability and performance is a well-known difficulty; Naiseh (2021) noted that model simplification improves explainability but reduces accuracy. This explained trade-off is a common frustration, illustrated by one participant stating, "Sometimes, making AI explainable makes it less intelligent."

Participants also noted inherited weaknesses in ethical and governance frameworks. Inconsistent XAI practice across industries may be solved by implementing Nikiforidis et al. (2025) quality explanation standards. Love et al. (2023) have also advocated for the interdisciplinary collaboration of ethicists, engineers, and policy-makers to achieve and enforce equity and access. Ultimately, the findings highlight that explainability is a socio-technical problem. It cannot be addressed solely by algorithms or visual design; it requires the integration of cross-disciplinary frameworks that embed human values in the technical design processes.

Table 2:
Thematic Analysis of Findings

Theme	Sub-theme	Codes	Description
Trust and Transparency	Calibration of trust; Perceived reliability	Trust, transparency, confidence	Users' trust improves when explanations are understandable and balanced between simplicity and completeness.
Human-Centred Design	Interactive explanation interfaces	Design, usability, empathy	Co-designed interfaces enhance engagement and comprehension by mirroring human reasoning.
Cognitive &	Cognitive load; Mental	Comprehension,	Users prefer

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Psychological Factors	model alignment	curiosity, overload	explanations that align with their cognitive schema and avoid technical overload.
Domain-Specific Applications	Contextualisation across sectors	Energy, manufacturing, construction	Explainability must adapt to industry needs and decision environments.
Ethical & Practical Challenges	Accountability and performance trade-offs	Bias, governance, performance	Ethical transparency requires institutional oversight and technical balance.

Summary of Findings

The investigation indicates that successful XAI depends on building calibrated trust, achieving cognitive alignment, and delivering field-appropriate explanations. Trust and transparency were foundational themes, underscoring that users request more than mere algorithmic exposure—they require an understanding that empowers decision-making. The consideration of human-centred design as interactive, empathetic, and adaptable interfaces strongly supports interface design. Overload and the equilibrium needed for interpretations to align with a user's mental model represent the cognitive and psychological dimensions of the problem. The variations XAI needs to think about include the language of the target industry as well as the language of the law. Other operational constraints like industry regulations also come into play. The lack of responsibility at the operational level highlights the explainability gap's ethical and pragmatic dimensions. Collectively, the themes represent XAI as an unfinished, evolving entity, focusing on the critical imbalance between the understanding a user can achieve and the intricacy a machine can execute.

Practical Recommendations

This study highlights several actionable recommendations for enhancing design, evaluation, and deployment processes for Explainable Artificial Intelligence (XAI) systems across various sectors. Initially, we advise creators and entities to adopt trust calibration frameworks to ensure provided explanations foster trust, avoiding extremes of total trust or total distrust (Naiseh et al., 2024). As mentioned by Naveed et al. (2024) and Naiseh (2021), confidence calibration demands attention. Thus, aside from the technical facets of explainability, explainers should offer clear, context-relevant explanations and professional perspectives that align with the user's cognitive level.

Implementing human-centered co-design is of equal importance. XAI interfaces need to be designed with end users to address the differences of natural reasoning, questioning, and interpretation that people use (Chromik & Butz, 2021). As noted by

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 10 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Rong et al. (2023) and Hudon et al. (2021), users possess different cognitive styles, which means that explanations should not be rigid and static, but flexible and dynamic. This design feature enhances emotional adhesion, converting systems from passive implements to active partners. Also, explaining interfaces should incorporate cognitive load adaptable and flexible explanatory strategies. As Shin (2021) states, the information constraint shouldn't over and underload, avoid over-simplification, and should not abstract the needed ideas while keeping out excess details. This is especially true for construction and energy management with high-stake risks, where actionable information should be clear, concise, and timely to support critical decision making (Teixeira et al., 2025; Love et al., 2023).

For better adoption, sector-specific explainability strategies are crucial. This aligns technical openness with sector-specific norms and potential legal frameworks. Ultimately, explainability under the principle of ethical governance will provide real accountability. In the views of Chamola et al. (2023) and Nikiforidis et al. (2025), the success of XAI cannot be attained through the algorithms alone but depends on the collaborative endeavour of the engineers, psychologists, and other stakeholders who will offer the value of ethical democracy, and the policymakers in the construction of transparent, equitable, and accountable AI.

Future Research Directions

In addition to research on explainability, the intersection of disparate cultures should be studied. Culture can deeply impact the dynamics of cognitive trust and can influence its calibration. Research is needed to document the longitudinal impact of explainability on the trust, or distrust, of users and subsequent behaviours (Hoffman et al., 2023). Case comparative studies on the emerging tech of generative AI and other multimodal systems can explain the portability of principles of explainability to other complex, adaptive systems. There is also a need to define weighted metrics to explain phenomena through means other than accuracy (Zhou et al., 2021). Psychological attributes of explanation, such as curiosity and perceived fairness, need to be included. Finally, the ethics of algorithmic interpretation must be dealt with in interdisciplinary efforts. Ensure explainability entails real accountability as opposed to solely performative transparency.

Conclusion

This study investigated how Explainable Artificial Intelligence (XAI) is building confidence in non-technical users in Pakistan. Confidence in AI systems was examined through the lenses of explainability, trust calibration, cognitive engagement, and contextual usability, as was rapport building through qualitative thematic analysis of semi-structured interviews. AI systems designed to assist non-technical users understand and save time on tasks and decisions were appreciated, but confidence was mostly given to systems when the designed AI provided ‘explainable, rational and contextual’ answers to questions (Kim et al., 2021). Explainable AI systems designed with trust in users were a positive and determining factor for the relationship of users with systems, as explainability was a time and human-centred process.

Participants described explainability in terms of trust and interdependence. Constructed explanations, in which the users trust systems to the level of determining

Online ISSN 3007-3197
Print ISSN 3007-3189

the device's actions, were a positive trust-enhancing factor. Participants defined cognitive models, which, as was aligned in the work of Naveed et al. (2024) and Shin (2021), were the primary confidence-enhancing factors, in which AI systems described decisions. Systems which provided cognitive shortcuts enhanced confidence. Systems which provided rational AI and trust in users, as concluded in the work of Kim et al. (2021), twenty trust in relation, integration, and explainability enhanced rational AI systems.

The emphasis on the need for exposure and understanding of the principles of explainable artificial intelligence in developing digital trust in emerging Pakistani and Pakistani-adjacent organisations contributes to both the theoretical and the practical understanding of the issue. However, the lack of understanding of the cultural and ethical implications of explainability in diverse and non-digital non-technical settings, as illustrated in the research, continues to be an opportunity for developing future research. To summarise, the success of AI in human environments does not merely rely on the accuracy of an AI's prediction but hinges on the quality of the explanation delivered to the user on all aspects of the prediction to foster trust and support engagement on a continuum of responsible use.

References

Abrar, K., Irfan, S., & Malik, F. (2024). Impact of Online Reviews on Consumer Trust, Perceived Value, and Purchase Behavior in Pakistan's E-Commerce Market. Journal of Business and Management Research, 3(3), 1009-1030.

Akbar, K., Abrar, K., & Khan, S. A. (2022). Effect of Information and Communication Technologies (ICT) as Innovation Tool on Business Performance: Evidence from Pakistan. Annals of Human and Social Sciences, 3(3), 494-504.

Alordiah, C. O., & Oji, J. (2024). Theoretical sampling strategies in qualitative interviews: Enhancing data richness and theoretical saturation. NIU Journal of Social Sciences, 10(2), 181-191.

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.

Balasubramaniam, N., Kauppinen, M., Hiekkanen, K., & Kujala, S. (2022, March). Transparency and explainability of AI systems: ethical guidelines in practice. In International working conference on requirements engineering: foundation for software quality (pp. 3-18). Cham: Springer International Publishing.

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.

Brdnik, S., Karakatič, S., & Šumak, B. (2025). A comprehensive user-centric method for evaluating and comparing XAI explanations. International Journal of Human–Computer Interaction, 1-20.

Brożek, B., Furman, M., Jakubiec, M., & Kucharzyk, B. (2024). The black box problem revisited. Real and imaginary challenges for automated legal decision making. Artificial Intelligence and Law, 32(2), 427-440.

Online ISSN 3007-3197
Print ISSN 3007-3189

Chamola, V., Hassija, V., Sulthana, A. R., Ghosh, D., Dhingra, D., & Sikdar, B. (2023). A review of trustworthy and explainable artificial intelligence (XAI). IEEe Access, 11, 78994-79015.

Chromik, M., & Butz, A. (2021). Human-XAI interaction: a review and design principles for explanation user interfaces. In IFIP Conference on Human-Computer Interaction (pp. 619-640). Springer, Cham.

Dahri, N. A., Dahri, F. H., Laghari, A. A., & Javed, M. (2025). Decoding ChatGPT's impact on student satisfaction and performance: a multimodal machine learning and explainable AI approach. Complex Engineering Systems, 5(2), N-A.

Ejjami, R. (2024). Emerging Professions in the Age of AI Across Multiple Sectors. International Journal for Multidisciplinary Research (IJFMR), 6(5), 1-32.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. Frontiers in Computer Science, 5, 1096257.

Hudon, A., Demazure, T., Karran, A., Léger, P. M., & Sénécal, S. (2021). Explainable artificial intelligence (XAI): how the visualization of AI predictions affects user cognitive load and confidence. In NeuroIS Retreat (pp. 237-246). Cham: Springer International Publishing.

Kalasampath, K., Spoorthi, K. N., Sajeev, S., Kuppa, S. S., Ajay, K., & Angulakshmi, M. (2025). A Literature review on applications of explainable artificial intelligence (XAI). IEEE Access.

Khan, I. A., & Ullah, A. (2024). The Role of Artificial Intelligence in Enhancing Social Governance: A Framework for Ethical Implementation and Policy Development in Pakistan. Journal of Management & Social Science, 1(4), 274-289.

Love, P. E., Fang, W., Matthews, J., Porter, S., Luo, H., & Ding, L. (2023). Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction. Advanced Engineering Informatics, 57, 102024.

Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy and AI, 9, 100169.

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.

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. ACM Transactions on Interactive Intelligent Systems (TiiS), 11(3-4), 1-45.

Naiseh, M. (2021). C-XAI: Design Method for Explainable AI Interfaces to Enhance Trust Calibration (Doctoral dissertation, Bournemouth University).

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.

Naveed, S., Stevens, G., & Robin-Kern, D. (2024). An overview of the empirical

Online ISSN 3007-3197
Print ISSN 3007-3189

evaluation of explainable ai (xai): A comprehensive guideline for user-centered evaluation in xai. Applied Sciences, 14(23), 11288.

Nikiforidis, K., Kyrtsoglou, A., Vafeiadis, T., Kotsiopoulos, T., Nizamis, A., Ioannidis, D., ... & Sarigiannidis, P. (2025). Enhancing transparency and trust in AI-powered manufacturing: A survey of explainable AI (XAI) applications in smart manufacturing in the era of industry 4.0/5.0. ICT Express, 11(1), 135-148.

Rong, Y., Leemann, T., Nguyen, T. T., Fiedler, L., Qian, P., Unhelkar, V., ... & Kasneci, E. (2023). Towards human-centered explainable ai: A survey of user studies for model explanations. IEEE transactions on pattern analysis and machine intelligence, 46(4), 2104-2122.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International journal of human-computer studies, 146, 102551.

Teixeira, B., Carvalhais, L., Pinto, T., & Vale, Z. (2025). Explainable AI framework for reliable and transparent automated energy management in buildings. Energy and Buildings, 116246.

Wiratsin, I. O., & Ragkhitwetsagul, C. (2025). Effectiveness of Explainable Artificial Intelligence (XAI) Techniques for Improving Human Trust in Machine Learning Models: A Systematic Literature Review. IEEE Access.

Wischnewski, M., Krämer, N., & Müller, E. (2023, April). Measuring and understanding trust calibrations for automated systems: A survey of the state-of-the-art and future directions. In Proceedings of the 2023 CHI conference on human factors in computing systems (pp. 1-16).

Wysocki, O., Davies, J. K., Vigo, M., Armstrong, A. C., Landers, D., Lee, R., & Freitas, A. (2023). Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI-driven clinical decision-making. Artificial Intelligence, 316, 103839.

Zhou, J., Gandomi, A. H., Chen, F., & Holzinger, A. (2021). Evaluating the quality of machine learning explanations: A survey on methods and metrics. Electronics, 10(5), 593.

Zhu, Y., Hua, G., Liu, X., Wang, C., & Tang, M. (2025). Trust in machines: how personality trait shapes static and dynamic trust across different human–machine interaction modalities. Frontiers in Psychology, 16, 1539054.