Indonesian Journal of Electrical Engineering and Computer Science
Vol. 43, No. 1, July 2026, pp. 271~280
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v43.i1.pp271-280

User-centered requirements elicitation for explainable AI transparency in recommendation and advertising systems

Osama Dakhel Alsuhaimy, M. Rizwan Jameel Qureshi
Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Article Info

Article history:

Received Feb 9, 2026
Revised Jun 2, 2026
Accepted Jun 27, 2026

Keywords:

Decision support
Explainable artificial intelligence
Psychological safety
Recommendation system
Trust calibration

ABSTRACT

Modern recommendation and advertising systems increasingly rely on complex artificial intelligence (AI) models whose opaque decision-making processes limit user understanding and trust. While explainable artificial intelligence (XAI) techniques aim to improve transparency, excessive disclosure can increase privacy concerns and psychological discomfort. This study addresses the lack of structured approaches for regulating transparency in user-facing AI systems. We propose the optimal transparency and psychological safety (OTPS) framework, which regulates explanation depth, timing, and user control to balance interpretability with psychological safety. The framework is implemented through a modular architecture consisting of an explanation generation module, transparency controller, and user interface layer. A user survey involving 35 participants was conducted to evaluate perceptions of transparency, trust, and psychological comfort. Statistical analysis, including reliability testing and response distribution evaluation, indicates strong user preference for adaptive transparency mechanisms. The results demonstrate that regulated transparency improves user trust and usability without introducing significant system overhead, providing practical design guidance for explainable AI systems in recommendation and advertising platforms.

Corresponding Author:

M. Rizwan Jameel Qureshi
Department of Information Technology, Faculty of Computing and Information Technology
King Abdulaziz University
21589 Jeddah, Jeddah, Saudi Arabia
Email: rmuhammd@kau.edu.sa

1. INTRODUCTION

Modern recommendation and advertising systems increasingly rely on artificial intelligence (AI) models to personalize content, rank options, and influence user decisions. Although these systems improve relevance and efficiency, their decision-making processes are often opaque, making it difficult for users to understand why a particular recommendation, advertisement, or decision output is presented. Explainable artificial intelligence (XAI) has therefore become an important approach for improving interpretability, supporting user understanding, and strengthening appropriate trust in AI-driven systems [1]–[3].

However, transparency in AI systems is not always beneficial when it is delivered without control or context. Excessive explanation detail may increase cognitive load, expose sensitive behavioral patterns, and create privacy-related anxiety or perceptions of surveillance. Prior studies show that explainability does not automatically improve trust, especially when users receive overly detailed, poorly timed, or emotionally uncomfortable explanations [4]–[6]. In user-facing systems such as recommendation and advertising

platforms, this creates a critical design challenge: explanations must provide enough information to support understanding and trust, but not so much detail that users feel uncomfortable, monitored, or overwhelmed.

Existing XAI research has contributed valuable methods for interpretability, trust calibration, and user-centered explanation design [1]–[6]. However, many approaches still focus on generating explanations rather than regulating how explanations should be delivered to users. In particular, there remains a lack of structured guidance on how to control explanation depth, explanation timing, abstraction level, and user control while also considering psychological safety. This gap is especially important in recommendation and advertising systems, where explanations may indirectly reveal personal preferences, behavioral data, or inferred user characteristics.

To address this gap, this study proposes the optimal transparency and psychological safety (OTPS) framework. The framework introduces a transparency regulation layer that manages explanation depth, timing, abstraction, and user control in AI-based recommendation and advertising systems. Instead of maximizing transparency, OTPS aims to identify an appropriate transparency boundary that supports interpretability, trust calibration, usability, and psychological comfort.

The main contributions of this study are as follows:

- It proposes the OTPS framework for regulating transparency in user-facing explainable AI systems.
- It presents a modular architecture consisting of an explanation generation module, transparency controller, and user interaction interface.
- It reports user-centered survey findings from 35 participants to identify requirements related to transparency, trust, psychological safety, and long-term usability.

By combining system-level design with user-centered requirements elicitation, this study provides practical guidance for developing explainable AI interfaces that are transparent, usable, and psychologically safe. The remainder of this paper is organized as follows. Section 2 reviews related work on XAI, trust calibration, and psychological safety. Section 3 presents the research problem. Section 4 introduces the proposed OTPS framework. Section 5 reports the user assessment and requirements elicitation results. Section 6 discusses the findings, and Section 7 concludes the paper with limitations and future research directions.

2. RELATED WORK

XAI has become an important research area for improving the interpretability, transparency, and usability of AI-driven systems. Prior studies have examined XAI from both technical and human-centered perspectives. Yang *et al.* [1] reviewed major XAI approaches and highlighted the continuing challenge of balancing interpretability with model complexity. Viser *et al.* [2] examined the relationship between explainability, trust, distrust, and appropriate reliance, showing that explanations do not automatically produce better trust outcomes. Naveed *et al.* [3] emphasized the need for user-centered evaluation methods in XAI, while Haque *et al.* [4] synthesized user-perspective research and identified future directions for explainable AI. Ali *et al.* [5] further discussed the broader requirements of trustworthy AI, including transparency, reliability, and responsible system behavior. Together, these studies show that explainability must go beyond technical model interpretation and should also support user understanding, usability, and appropriate reliance.

A second major direction in the literature focuses on trust calibration, affective design, adaptive explanations, and interactive XAI. Naiseh *et al.* [6] proposed the Calibrated-XAI framework to support appropriate trust and reduce both over-trust and under-trust. Bernardo and Seva [7] highlighted the importance of affective and emotional factors in user-centered XAI design. Suffian *et al.* [8] showed that user feedback can improve the effectiveness of counterfactual explanations, while Nimmo *et al.* [9] questioned excessive personalization and suggested that only selected user characteristics may meaningfully influence explanation effectiveness. Kaplan *et al.* [10] proposed a user-centric framework for explainable AI across the system lifecycle. Other studies have expanded XAI into applied domains such as recommendation fairness, explanation timing, social media transparency, requirements prioritization, and AI-generated advertising. Ge *et al.* [11] examined explainable fairness in recommendation systems, Chen *et al.* [12] showed that explanation timing affects user perceptions and trust, West *et al.* [13] found that algorithmic transparency can create discomfort in social media systems, Bai *et al.* [14] applied XAI to user requirement prioritization, and Jung *et al.* [15] highlighted the role of trust, perceived humanness, and anxiety in AI-generated advertising.

More recent and foundational studies further show that explanations must be selective, contextual, and aligned with human expectations. Miller [16] argued that explanations should be grounded in social science and should avoid exhaustive technical disclosure. Doshi-Velez and Kim [17] emphasized the need for a rigorous science of interpretability based on human understanding and decision-making outcomes. Abdul *et al.* [18] identified the shift toward human-centered explainable systems and highlighted issues such as cognitive overload and mismatched explanation timing. Ehsan *et al.* [19] introduced reflective XAI,

emphasizing interactive and feedback-driven explanation processes. Liao and Wortman [20] discussed the practical implementation of AI transparency in real-world systems, while Rader *et al.* [21] showed that users' mental models of algorithmic systems often remain incomplete even when explanations are provided. Bussone *et al.* [22] demonstrated that explanations influence trust and reliance in decision-support systems, but excessive explanation can reduce usability. Wang *et al.* [23] examined psychological impacts such as anxiety and perceived surveillance. Antorán *et al.* [24] argued that explanations should align with user goals rather than only system internals, and Zhang *et al.* [25] showed that high-fidelity explanations may increase cognitive load without necessarily improving trust.

Overall, the literature shows that XAI research has made strong progress in interpretability, trust calibration, user-centered evaluation, feedback-based explanation, and applied transparency. However, existing studies still lack a unified system-level framework that regulates explanation depth, explanation timing, abstraction level, user control, and psychological safety together. This gap is especially important in recommendation and advertising systems, where explanations may indirectly reveal user preferences, behavioral patterns, or inferred personal characteristics. The proposed OTPS framework addresses this gap by introducing a transparency regulation layer that balances interpretability, trust calibration, usability, privacy awareness, and psychological comfort in user-facing AI systems. Table 1 shows synthesized comparison of existing XAI literature and OTPS contribution.

Table 1. Synthesized comparison of existing XAI literature and OTPS contribution

Literature theme	Main contribution	Remaining gap addressed by OTPS
General XAI and interpretability [1], [5], [16], [17]	Explain AI behavior and improve interpretability	Limited focus on psychological safety and regulated transparency
User-centered XAI evaluation [3], [4], [18], [20], [21]	Identify user needs, evaluation methods, and human-centered design issues	Does not fully define a system-level transparency control mechanism
Trust calibration and affective design [2], [6], [7], [22], [25]	Support appropriate reliance and consider emotional/cognitive trust factors	Limited integration of timing, user control, and privacy-aware disclosure
Interactive and adaptive explanations [8]–[10], [19], [24]	Use feedback, personalization, counterfactuals, and goal-oriented explanations	Needs clearer boundaries for explanation depth and disclosure level
Fairness, timing, and applied XAI contexts [11], [12], [14]	Show importance of fairness, timing, and requirements-driven explanation design	Requires integration into a unified adaptive transparency framework
Privacy, anxiety, and surveillance perception [13], [15], [23]	Show that excessive transparency can create discomfort and privacy anxiety	OTPS directly incorporates psychological safety and abstract disclosure

3. PROBLEM DEFINITION

Current explainable AI approaches often assume that increasing transparency will automatically improve user understanding, trust, and acceptance. However, this assumption is limited in user-facing recommendation and advertising systems, where explanations may expose personal preferences, behavioral patterns, or inferred user characteristics. Prior research shows that explainability does not always improve appropriate reliance, especially when explanations are overly detailed, poorly contextualized, or difficult for users to interpret [2], [6]. In such cases, static or unrestricted explanation delivery may increase cognitive load, privacy concerns, and perceptions of surveillance rather than improving trust [13], [15].

The main problem addressed in this study is the lack of a structured framework for regulating transparency in explainable AI systems. Existing approaches provide useful explanation methods and trust-calibration strategies, but they do not sufficiently define how much explanation should be shown, when it should appear, what level of abstraction should be used, or how users should control disclosure [3]–[6]. Therefore, user-facing AI systems require a regulated transparency mechanism that balances explanation depth, explanation timing, user control, and psychological safety. This study addresses this problem through the proposed OTPS framework, which aims to support interpretability while reducing privacy-related anxiety, cognitive overload, and discomfort.

4. THE PROPOSED SOLUTION

To address the limitations of unrestricted transparency in explainable AI systems, this study proposes the OTPS framework. The framework shifts the design objective from maximizing transparency to regulating transparency in a way that supports user understanding, appropriate trust, usability, and psychological comfort. OTPS introduces a transparency regulation layer between the AI system and the user interface. This layer controls how much explanation is shown, when it is presented, how abstract or detailed it should be, and how much control users have over explanation disclosure.

Figure 1 presents the overall workflow of the OTPS framework. It shows how the AI-based system output passes through a transparency regulation layer before being presented to the user. This layer manages explanation depth, timing, abstraction, and user control to balance interpretability with psychological safety. The OTPS framework is organized around four core goals. These goals define the main design dimensions required for regulated transparency in AI-based recommendation and advertising systems. Table 2 presents core goals and design mechanisms of the OTPS framework.

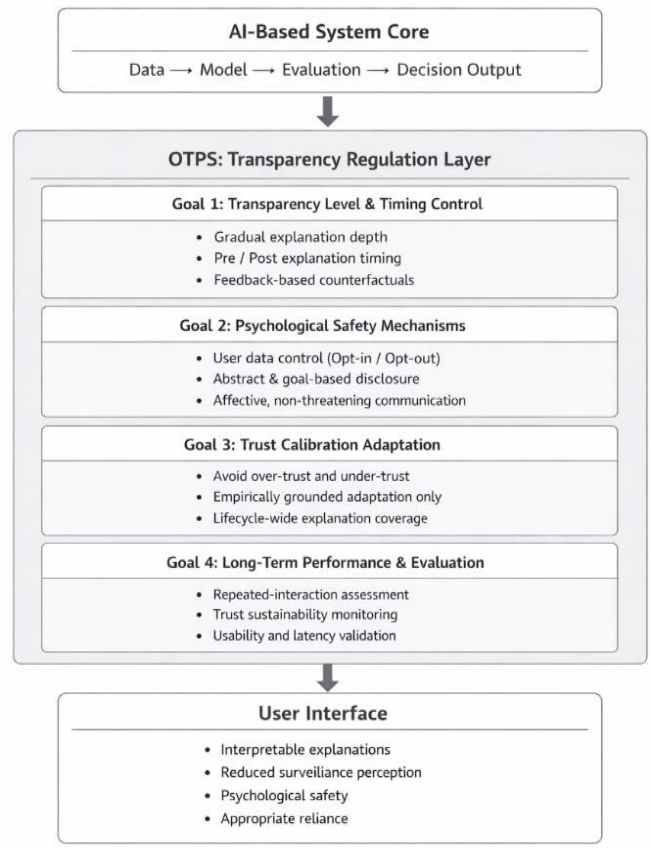

Figure 1. Workflow architecture of the proposed OTPS framework

Table 2. Core goals and design mechanisms of the OTPS framework

OTPS goal	Purpose	Design mechanism
Goal 1. Transparency level and timing	Avoid cognitive overload and unnecessary disclosure	Gradual explanation levels, contextual explanation delivery, pre/post explanation timing
Goal 2. User control and psychological safety	Reduce privacy anxiety and surveillance perception	Abstract disclosure, opt-in/opt-out controls, user-adjustable explanation visibility
Goal 3. Trust calibration	Prevent over-trust and under-trust in AI decisions	Confidence indicators, adaptive explanations, uncertainty communication
Goal 4. Long-term usability	Maintain usefulness during repeated interactions	Minimal intrusion, repeated-use evaluation, sustainable explanation delivery

The first goal regulates explanation depth and timing. Instead of presenting full technical explanations by default, the system provides explanations progressively according to user needs and interaction context. Basic explanations can be shown first, while more detailed explanations remain available when users request additional information. This reduces cognitive load and prevents users from being overwhelmed by unnecessary technical details. The second goal focuses on user control and psychological safety. Users should be able to manage how explanations are displayed and how much personal information is reflected in them. The framework therefore emphasizes abstract disclosure, where explanations describe general reasoning patterns rather than exposing sensitive personal behavior directly. This helps reduce privacy-related anxiety and the feeling of being constantly monitored. The third goal supports appropriate

trust calibration. The system should help users understand the confidence and limitations of AI recommendations instead of encouraging blind trust or complete distrust. This can be achieved through confidence indicators, short explanation summaries, expandable reasoning panels, and adaptive explanations that clarify why a recommendation or advertisement was generated. The fourth goal addresses long-term usability. Explanation mechanisms should remain useful and non-intrusive during repeated use. If explanations are too frequent, too detailed, or poorly timed, users may experience explanation fatigue. Therefore, OTPS encourages sustainable transparency design that remains clear, lightweight, and adaptable over time.

4.1. Practical implementation of the OTPS framework

The OTPS framework can be implemented as a modular architecture consisting of four main components: the AI Recommendation Engine, Explanation Generation Module, Transparency Controller, and User Interaction Interface. The AI Recommendation Engine generates the recommendation, advertisement, or decision output using the underlying AI model. The Explanation Generation Module produces interpretable explanations using suitable XAI methods such as SHAP, LIME, feature attribution, or counterfactual explanations. The Transparency Controller, which represents the OTPS layer, regulates explanation depth, timing, abstraction level, and user control based on user preferences, system context, and transparency requirements. Finally, the User Interaction Interface presents explanations through expandable explanation panels, confidence indicators, transparency settings, and feedback options.

Figure 2 illustrates the system-level implementation of OTPS. The AI recommendation engine produces the output, the explanation generation module prepares explanation content, and the OTPS transparency controller determines whether the explanation should be shown immediately, delayed, summarized, expanded, abstracted, or controlled by the user. The user interface then presents the explanation in a clear and psychologically safe manner. This design allows OTPS to be integrated into existing recommendation and advertising systems without requiring major changes to the underlying AI model. The framework can operate as a middleware or interface-level transparency management layer. By separating explanation generation from explanation regulation, OTPS supports practical implementation while maintaining flexibility across different AI platforms and application domains.

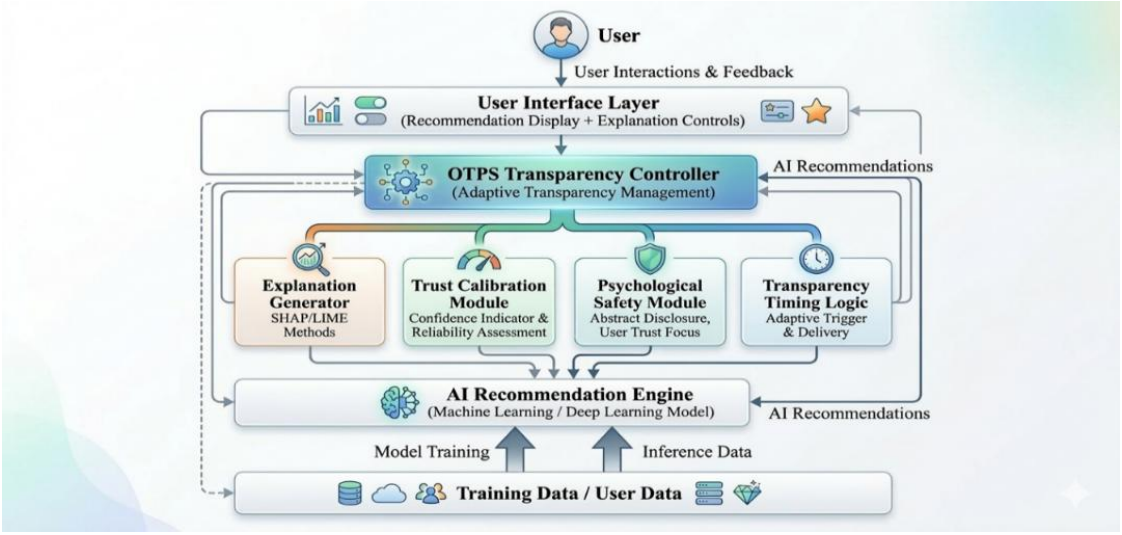

Figure 2. Practical interface implementation of the OTPS framework for AI recommendation systems

4.2. Technical evaluation of the OTPS framework

The OTPS framework is technically feasible because it can be implemented as an intermediate transparency layer between the AI recommendation engine and the user interface. The underlying AI model remains responsible for generating recommendations, while the explanation generation module produces interpretable explanation content using existing XAI techniques such as SHAP, LIME, feature attribution, or counterfactual explanations. The OTPS transparency controller then regulates the level of explanation detail, timing of disclosure, abstraction level, and user control before the explanation is displayed to the user.

The main computational cost of OTPS comes from the explanation generation module, especially when feature-attribution methods are used. If (m) represents the number of model predictions requiring

explanation and (n) represents the number of input features, the approximate computational complexity can be expressed in (1):

$$O(m \times n) \quad (1)$$

However, the OTPS transparency controller introduces negligible overhead because it mainly performs rule-based or preference-based selection of explanation depth, timing, and disclosure level. In addition, OTPS reduces unnecessary computation through adaptive triggering, where explanations are generated only when requested by the user or when required by the interaction context. This avoids continuous explanation generation and helps maintain efficient system performance. Table 3 illustrates performance impact of the OTPS components.

Overall, OTPS can be integrated into existing AI recommendation and advertising systems without major changes to the underlying model architecture. Since the framework separates explanation generation from transparency regulation, it can operate as a middleware or interface-level control layer. This makes OTPS practical for real-time and large-scale AI systems where transparency must be provided without creating excessive latency, cognitive overload, or system performance degradation.

Table 3. Performance impact of the OTPS components

Component	Performance impact
AI prediction	No additional overhead
Explanation generation	Moderate cost depending on SHAP, LIME, or selected XAI method
OTPS transparency controller	Negligible overhead
User interface rendering	Minimal overhead
Adaptive triggering	Reduces unnecessary explanation computation

5. USER ASSESSMENT AND REQUIREMENTS ELICITATION

This section presents the user assessment conducted to evaluate perceptions of the proposed OTPS framework. The purpose of the assessment was to identify user-centered requirements related to transparency adjustment, user control, psychological safety, trust calibration, and long-term usability in explainable AI systems. A survey was conducted with 35 participants who regularly interact with digital platforms such as recommendation systems, e-commerce services, streaming platforms, and personalized advertising environments. The questionnaire consisted of 30 Likert-scale items grouped according to the four OTPS goals. Each item was evaluated using a five-point scale ranging from 1 (Very Low) to 5 (Very High). The tables in this section summarize the participant demographics (Table 4), descriptive statistics (Table 5), cumulative response distribution (Table 6), questionnaire reliability (Table 7), and comparison with existing XAI approaches (Table 8). The participant profile indicates that the sample was suitable for exploratory user-centered assessment because most respondents were familiar with digital platforms and AI-driven recommendation environments. Although the sample size is modest, it provides useful initial insights for requirements elicitation and framework refinement.

Table 4. Participant demographics

Category	Distribution
Total participants	35
Male	20 (57%)
Female	15 (43%)
Age 18–20	14%
Age 21–30	63%
Age 31–40	23%
Technical background	71%
Non-technical background	29%

The survey results were summarized across the four OTPS goals using mean and standard deviation values. The results show generally positive user perception toward adaptive transparency, user control, psychological safety, trust calibration, and long-term usability. These results suggest that users do not prefer unrestricted or overly detailed transparency by default. Instead, they favor explanation mechanisms that adapt to context, provide suitable levels of detail, and allow users to control how much information is disclosed. Goal 2 and Goal 4 received the strongest mean scores, indicating that user control, privacy-aware explanation design, and long-term usability are particularly important requirements for explainable AI interfaces.

Table 5. Summary of user assessment results across OTPS goals

OTPS goal	Mean	SD	Key interpretation
Goal 1: transparency adjustment and explanation timing	3.69	1.02	Users support adaptive timing and gradual explanation delivery
Goal 2: user control and psychological safety	3.92	0.98	Users strongly value control, abstract disclosure, and privacy-aware explanations
Goal 3: trust calibration through adaptive explanations	3.84	1.01	Users support explanations that help prevent over-trust and under-trust
Goal 4: long-term usability and sustainability	3.95	0.94	Users prefer explanation mechanisms that remain useful and non-intrusive over time

SD: standard deviation

Table 6 presents the combined response distribution across all four goals. Instead of using separate charts for each goal, the figure summarizes the overall pattern of participant responses and visually demonstrates the dominance of high and very high responses. From a requirements engineering perspective, the results identify several important design requirements for explainable AI systems. First, explanations should be adaptive rather than static. Second, users should be given control over explanation depth and data disclosure. Third, explanation interfaces should communicate uncertainty and confidence to support appropriate trust calibration. Fourth, transparency mechanisms should remain lightweight and understandable during repeated use. To confirm the reliability of the questionnaire, Cronbach's Alpha analysis was conducted across the four OTPS constructs. The overall questionnaire reliability was 0.87, which is above the commonly accepted threshold of 0.70 and indicates strong internal consistency.

Table 6. Final cumulative evaluation of all goals

Goal	Very low	Low	Neutral	High	Very high
Goal 1	6.4%	12.5%	22.0%	26.0%	34.0%
Goal 2	5.0%	10.7%	17.9%	24.6%	43.9%
Goal 3	7.3%	10.8%	18.4%	21.6%	43.8%
Goal 4	8.0%	9.7%	14.3%	25.1%	46.3%
Average	6.68%	10.93%	18.15%	24.33%	42.00%

Table 7. Summary of user assessment results across OTPS goals

Construct	No. of items	Cronbach's Alpha
Goal 1: transparency timing	8	0.84
Goal 2: psychological safety	8	0.88
Goal 3: trust calibration	9	0.86
Goal 4: long-term usability	5	0.85
Overall questionnaire	30	0.87

A Chi-square goodness-of-fit test was also conducted on the aggregated Likert responses to examine whether the response distribution significantly differed from a uniform distribution. The result was statistically significant, $\chi^2 = 42.31$, $p < 0.01$, indicating that the concentration of responses in the High and Very High categories was unlikely to be due to random variation. This supports the interpretation that participants consistently favored transparency mechanisms aligned with the OTPS framework.

In Table 8, the OTPS framework was compared with existing explainable AI approaches to clarify its contribution. Existing studies provide valuable contributions in interpretability, trust calibration, user-centered evaluation, and interactive explanation design. However, they do not fully combine transparency depth, timing, abstraction, user control, and psychological safety within one system-level framework.

Table 8. Comparison between OTPS model and related work

Feature	OTPS framework	Yang *et al.* [1]	Calibrated-XAI (C-XAI) [6]	Reflective XAI [19]
Transparency adjustment	Yes	Yes	No	Partial
Psychological safety	Yes	No	Yes	Partial
User control	Yes	No	No	Yes
Privacy concern handling	Yes	No	Yes	No
Adaptive explanation timing	Yes	No	No	No
Lifecycle-wide transparency	Yes	No	No	Partial
Interactive feedback	Yes	No	Partial	Yes
System-level transparency regulation	Yes	No	No	No

Overall, the user assessment supports the need for regulated transparency in explainable AI systems. The findings show that users value explanations that are adaptive, controllable, privacy-aware, and sustainable over time. These results strengthen the OTPS framework as a practical requirements-driven approach for designing explainable AI interfaces in recommendation and advertising systems.

6. DISCUSSION

The findings suggest that explainable AI systems should not treat transparency as unrestricted disclosure. Users appear to prefer adaptive transparency, where explanations are provided at an appropriate level of detail and only when they are useful for understanding the AI output. This supports the idea that explanation design should move beyond simply showing more information and instead focus on presenting the right amount of information at the right time. The results also indicate that psychological safety and user control should be treated as core design requirements in user-facing AI systems. In recommendation and advertising environments, explanations may indirectly reveal personal preferences, behavioral patterns, or inferred user characteristics. Therefore, users should be given control over explanation visibility, data disclosure, and the level of detail presented. Abstract explanations, confidence indicators, and expandable explanation panels can help users understand AI decisions without creating unnecessary discomfort or privacy-related anxiety. From an engineering perspective, the OTPS framework can be implemented as a transparency regulation layer between the AI model and the user interface. This layer allows explanation generation to remain separate from explanation delivery, making it possible to regulate explanation depth, timing, abstraction, and user control without modifying the underlying AI model. As a result, OTPS provides a practical approach for integrating explainability into real-world recommendation and advertising systems while balancing interpretability, usability, and psychological comfort.

7. CONCLUSION

This study proposed the OTPS framework for regulating transparency in explainable AI systems. Unlike approaches that focus mainly on increasing explanation detail, OTPS emphasizes controlled transparency by managing explanation depth, timing, abstraction level, and user control. The framework is designed to support interpretability, appropriate trust calibration, usability, and psychological comfort in AI-based recommendation and advertising systems. The user assessment showed that participants generally preferred adaptive and controllable explanation mechanisms rather than unrestricted full disclosure. These findings suggest that explainable AI interfaces should provide explanations progressively, allow users to adjust transparency levels, and avoid unnecessary exposure of sensitive behavioral information. From an engineering perspective, OTPS can be implemented as a transparency regulation layer between the AI model and the user interface, making it practical for integration into existing recommendation and advertising platforms. The main limitation of this study is the relatively small sample size of 35 participants, which limits the generalizability of the findings. Future research should evaluate the OTPS framework with larger and more diverse user groups, real-world AI platforms, and longitudinal studies to examine how trust, cognitive load, privacy concerns, and psychological comfort evolve over repeated interactions. Further studies may also apply OTPS in other domains such as healthcare decision support, intelligent assistants, financial recommendation systems, and autonomous systems.

FUNDING INFORMATION

The authors declare that no funding was received for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Osama Dakhel Alsuhaimy	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
M. Rizwan Jameel Qureshi		✓				✓		✓	✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request at the email address: rmuhammd@kau.edu.sa. The data are not publicly available due to privacy and ethical considerations related to human participant responses.

REFERENCES

[1] W. Yang *et al*., "Survey on explainable AI: From approaches, limitations and applications aspects," *Human-Centric Intelligent Systems*, vol. 3, no. 3, pp. 161–188, Aug. 2023, doi: 10.1007/s44230-023-00038-y.

[2] R. Visser, T. M. Peters, I. Scharlau, and B. Hammer, "Trust, distrust, and appropriate reliance in (X)AI: A conceptual clarification of user trust and survey of its empirical evaluation," *Cognitive Systems Research*, vol. 91, p. 101357, Jun. 2025, doi: 10.1016/j.cogsys.2025.101357.

[3] S. Naveed, G. Stevens, and D. Robin-Kern, "An overview of the empirical evaluation of explainable AI (XAI): A comprehensive guideline for user-centered evaluation in XAI," *Applied Sciences*, vol. 14, no. 23, p. 11288, Dec. 2024, doi: 10.3390/app142311288.

[4] A. B. Haque, A. K. M. N. Islam, and P. Mikalef, "Explainable artificial intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research," *Technological Forecasting and Social Change*, vol. 186, p. 122120, Jan. 2023, doi: 10.1016/j.techfore.2022.122120.

[5] S. Ali *et al*., "Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence," *Information Fusion*, vol. 99, p. 101805, Nov. 2023, doi: 10.1016/j.inffus.2023.101805.

[6] M. Naiseh, A. Simkute, B. Zieni, N. Jiang, and R. Ali, "C-XAI: A conceptual framework for designing XAI tools that support trust calibration," *Journal of Responsible Technology*, vol. 17, p. 100076, Mar. 2024, doi: 10.1016/j.jrt.2024.100076.

[7] E. Bernardo and R. Seva, "Affective design analysis of explainable artificial intelligence (XAI): A user-centric perspective," *Informatics*, vol. 10, no. 1, p. 32, Mar. 2023, doi: 10.3390/informatics10010032.

[8] M. Suffian, U. Kuhl, A. Bogliolo, and J. M. Alonso-Moral, "The role of user feedback in enhancing understanding and trust in counterfactual explanations for explainable AI," *International Journal of Human-Computer Studies*, vol. 199, p. 103484, May 2025, doi: 10.1016/j.ijhcs.2025.103484.

[9] R. Nimmo, M. Constantinides, K. Zhou, D. Quercia, and S. Stumpf, "User characteristics in explainable AI: The rabbit hole of personalization?," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, May 2024, pp. 1–13. doi: 10.1145/3613904.3642352.

[10] S. Kaplan, H. Uusitalo, and L. Lensu, "A unified and practical user-centric framework for explainable artificial intelligence," *Knowledge-Based Systems*, vol. 283, p. 111107, Jan. 2024, doi: 10.1016/j.knosys.2023.111107.

[11] Y. Ge *et al*., "Explainable fairness in recommendation," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA: ACM, Jul. 2022, pp. 681–691. doi: 10.1145/3477495.3531973.

[12] C. Chen, M. Liao, and S. S. Sundar, "When to explain? Exploring the effects of explanation timing on user perceptions and trust in AI systems," in *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems*, New York, NY, USA: ACM, Sep. 2024, pp. 1–17. doi: 10.1145/3686038.3686066.

[13] J. West, B. Cagiltay, S. Zhang, J. Li, K. Fawaz, and S. Banerjee, "'Impressively scary': Exploring user perceptions and reactions to unraveling machine learning models in social media applications," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, Apr. 2025, pp. 1–21. doi: 10.1145/3706598.3713256.

[14] S. Bai, S. Shi, C. Han, M. Yang, B. B. Gupta, and V. Arya, "Prioritizing user requirements for digital products using explainable artificial intelligence: A data-driven analysis on video conferencing apps," *Future Generation Computer Systems*, vol. 158, pp. 167–182, Sep. 2024, doi: 10.1016/j.future.2024.04.037.

[15] T. Jung, M. Koghut, E. Lee, and O. Kwon, "Artificial creativity in luxury advertising: How trust and perceived humanness drive consumer response to AI-generated content," *Journal of Retailing and Consumer Services*, vol. 87, p. 104403, Oct. 2025, doi: 10.1016/j.jretconser.2025.104403.

[16] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, Feb. 2019, doi: 10.1016/j.artint.2018.07.007.

[17] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017, doi: 10.48550/arXiv.1702.08608.

[18] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, "Trends and trajectories for explainable, accountable and intelligible systems," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, Apr. 2018, pp. 1–18. doi: 10.1145/3173574.3174156.

[19] U. Ehsan and M. O. Riedl, "Human-centered explainable AI: Towards a reflective sociotechnical approach," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12424 LNCS, C. Stephanidis, M. Kurosu, H. Degen, and L. Reinerman-Jones, Eds., Cham: Springer, 2020, pp. 449–466. doi: 10.1007/978-3-030-60117-1_33.

[20] Q. V. Liao and J. Wortman Vaughan, "AI transparency in the age of LLMs: A human-centered research roadmap," *Harvard Data Science Review*, Special Issue 5, May 2024, doi: 10.1162/99608f92.8036d03b.
[21] S. Mohseni, N. Zarei, and E. D. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable AI systems," *ACM Transactions on Interactive Intelligent Systems*, vol. 11, no. 3–4, Art. no. 24, pp. 1–45, Aug. 2021, doi: 10.1145/3387166.
[22] A. Bussone, S. Stumpf, and D. O'Sullivan, "The role of explanations on trust and reliance in clinical decision support systems," *in Proc. 2015 IEEE Int. Conf. Healthcare Informatics (ICHI)*, Dallas, TX, USA, Oct. 2015, pp. 160–169, doi: 10.1109/ICHI.2015.26.
[23] H. J. Oh, J. Kim, J. J. C. Chang, N. Park, and S. Lee, "Social benefits of living in the metaverse: The relationships among social presence, supportive interaction, social self-efficacy, and feelings of loneliness," *Comput. Hum. Behav.*, vol. 139, Art. no. 107498, Feb. 2023, doi: 10.1016/j.chb.2022.107498.
[24] C. Deng, A. Kim, and W. Lu, "A generic battery-cycling optimization framework with learned sampling and early stopping strategies," *Patterns*, vol. 3, no. 7, Art. no. 100531, Jul. 2022, doi: 10.1016/j.patter.2022.100531.
[25] S. van den Elzen et al., "The flow of trust: A visualization framework to externalize, explore, and explain trust in ML applications," *IEEE Comput. Graph. Appl.*, vol. 43, no. 2, pp. 78–88, Mar.–Apr. 2023, doi: 10.1109/MCG.2023.3237286.

BIOGRAPHIES OF AUTHORS

Osama Dakhel Alsuhaimy is currently pursuing the M.S. degree with the Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include information technology, artificial intelligence, and emerging computing applications. He can be contacted at email: oalsuhaymi0006@stu.kau.edu.sa.

M. Rizwan Jameel Qureshi received his Ph.D. degree from the National College of Business Administration & Economics, Pakistan, in 2009. He is currently a Full Professor with the Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. He has been recognized as the Best Researcher by the Department of Information Technology, King Abdulaziz University, in 2013 and 2016 for his outstanding research contributions. His research interests include software engineering, information systems, artificial intelligence, and emerging computing technologies. He can be contacted at email: rmuhammd@kau.edu.sa.