Reflexive Dialogue-Based Explainability for Human-AI Collaboration: An Empirical Study on Adaptive and Interactive Explanations

Christos Troussas, Akrivi Krouska, Phivos Mylonas, Cleo Sgouropoulou
Department of Informatics and Computer Engineering
University of West Attica
Egaleo, Greece
{ctrouss, akrouska, mylonasf, csgouro}@uniwa.gr

***Abstract* - As artificial intelligence (AI) systems are increasingly integrated into decision-making processes, the need for transparency and user-aligned explanations has become critical. While traditional explainability methods – static or post-hoc – have advanced, they often fail to adapt to users' evolving informational needs during real-time interactions, limiting their effectiveness in collaborative contexts. This paper addresses this gap by empirically evaluating reflexive dialogue-based explanations, which dynamically adjust explanatory content through iterative user-AI exchanges. We conducted a mixed-method study involving 80 participants performing complex decision-support tasks under two conditions: static explanations and reflexive dialogue. Quantitative results demonstrate that reflexive dialogue significantly improves task accuracy, comprehension, and trust calibration, while qualitative findings reveal enhanced user engagement, perceived agency, and satisfaction. The study identifies key interaction patterns that support cognitive integration and trust refinement. Our main contribution lies in validating a human-centered, adaptive explanation framework that goes beyond one-size-fits-all transparency. The novelty of this work lies in positioning reflexive dialogue not merely as a user support feature but as an essential mechanism for dynamically co-constructing meaning in human-AI collaboration.**

Keywords - Explainable Artificial Intelligence (XAI); Human-AI Collaboration; Adaptive Explanations; Reflexive Dialogue; Trust Calibration

I. INTRODUCTION

In the past few years, artificial intelligence (AI), especially generative AI has become increasingly pervasive in various industries—changing how humans and machines cooperate in complex ways [1]. Despite the amazing performance capabilities of AIs, their transparency (or lack thereof) in decision-making stands as a major hurdle in human-AI collaboration [2]. Explainability—the description of how an AI makes specific decisions—has become a requirement to pursue trust, reliability, and engagement of humans with intelligent technology [3]. Without suitable explanations, users may not be able to make sense of AI-supporting predictions or recommendations to aid their actions, and ultimately this would limit trust, less likely adoption, and diminished collaborative outcomes.

While the field of AI explainability has made great progress, existing methods are chiefly static or post-hoc explanations, which causes a mismatch between users' informational requirements in the moment of real-time use and the more passive and static explanations they receive [4]. This mismatch in post-hoc or static explanations generally leads to either way too shallow or too complex an explanation that is unable to meet any individual person's needs, leading to cognitive overload or misunderstanding. Therefore, more recent literature has clearly pushed in the direction of user-initiated and adaptive techniques for explanation, which promote user engagement through communicative dialogue to enable ongoing explanations based on user feedback on-the-fly [5-9], drawing on earlier foundational work on semantic personalization frameworks for media and knowledge systems. Reflexive dialogue-based explanations are found to be an exciting but under-explored avenue, which tailor explanations through ongoing interactions adapted based on user inputs, potentially enabling a more dynamic standard of explanations than the mostly static as well as traditional interactive approaches.

Research on explainable AI has taken many forms, including transparent models, post-hoc interpretability, and interactive explanation systems [10-17]. While transparent models like decision trees and linear models provide transparency, they are less scalable and don't lend themselves to more complex tasks [18-22]. Post-hoc approaches like LIME did help users understand complex models, but were not engineered to enable users to refine and redefine their explanations [23-25]. Interactive or dynamic explanations, while better able to adapt in real-time to user's interactions, still don't fit the user's intent and mostly use pre-configured explanation scripts or interactions through highly restrictive channels in the best-case scenario, with little to no active engagement/reflection for end-users. [26-27]. There is thus a large gap in studies on how reflexive (dialogue based adaptive explanations; iterative user and AI interactions) and adaptive dialogue (to better address the needs of a user, both at the initiation of explanation and its extension into more

collaborative situations) can be meaningfully designed and implemented correspondingly in practice for effective user-AI collaboration.

The goal of this research is to empirically test the effects of reflexive dialogue-based explanations on human-AI collaboration. Our research objectives are to investigate how dynamic and adaptable dialogues enhance user understanding, calibrate trust, and collaborative task performance, relative to static explanation formats. Our research contributions are: (1) providing empirical evidence for reflexive dialogue as a way to improve human-AI collaboration, (2) identifying relevant interaction patterns that promote user understanding and trust, and (3) contributing new knowledge about designing adaptive explanation systems that provide context-appropriate adaptions to user-identified needs. The main novelty of the study is our explicit focus on reflexive, iterative Interaction as a mechanism to adaptively calibrate explanations, providing a significant yet often overlooked approach to dynamically calibrate user understanding and trust through dialogue-based explanations in the current discourse on explainability.

II. Methodology

This research used a mixed-methods approach, combining quantitative and qualitative methods to comprehensively assess the influence of reflexive dialogue-based explanations on human-AI collaboration. The mixed-methods design, utilizing a combination of structured performance measures, standardized psychometric measures, and qualitative interviews, provides various opportunities to examine influence of explanation type on user experience, user performance, and development of trust.

Eighty participants (45 male, 35 female), ranging in age from 21 to 47 years (M = 30.1, SD = 6.3), were recruited for this study using university mailing lists, academic social networks, and professional forums in the fields of computing, education, and information science. Eligibility criteria included fluency in English, moderate to advanced digital literacy, and no prior involvement in similar AI interaction studies. Informed consent was obtained from all participants, and institutional ethical approval was granted prior to the experiment.

Participants were randomly assigned to one of two groups: the control group (n = 40), which interacted with a generative AI system providing static, pre-scripted explanations, and the experimental group (n = 40), which interacted with the same AI system augmented with a reflexive dialogue mechanism capable of adapting explanations in real-time based on user feedback and clarification queries.

All participants completed a set of three collaborative problem-solving tasks in a simulated digital environment. The tasks were framed within a realistic decision-support scenario – such as selecting an optimal solution for urban resource allocation, ethical dilemmas in medical triage, or evaluating software architecture trade-offs. These domains were selected due to their inherent complexity, necessity for justifiable reasoning, and the presence of multiple valid solutions, thus allowing for meaningful explanations and opportunities for reflexive dialogue.

Each session lasted approximately 50 minutes and was conducted individually in a controlled lab setting. Participants received a brief orientation explaining the interface and task objectives, followed by a training round using a simple warm-up task. The AI system provided step-by-step guidance throughout the main tasks, responding either with static explanations (control) or engaging in interactive dialogue to iteratively clarify, expand, or reframe its responses (experimental).

Quantitative data were gathered from:

- Performance Metrics: Accuracy of final decisions, time-to-completion, number of clarification requests, and number of backtracking instances (revisiting previous decision points).
- Comprehension Assessment: A post-task comprehension quiz consisting of 12 multiple-choice items that tested users' understanding of the reasoning behind the AI's recommendations.
- Trust Calibration Scale: A 10-item validated instrument adapted from the Trust in Automation Inventory [28], measuring perceived reliability, predictability, and helpfulness of the AI.
- Cognitive Load Scale: NASA-TLX (Task Load Index), measuring mental demand, effort, and frustration during the interaction.

Qualitative data were obtained through semi-structured interviews conducted immediately after the tasks. Interviews lasted 15–20 minutes and focused on four main themes: (1) user perceptions of explanation clarity, (2) emotional and cognitive responses to the AI's behavior, (3) instances of confusion or misalignment, and (4) perceived usefulness and satisfaction with the interaction.

All sessions and interviews were audio-recorded and transcribed for thematic analysis. An inductive coding framework was developed by two researchers independently, followed by inter-coder agreement (Cohen's $\kappa = 0.82$), ensuring analytical reliability.

The control condition used static explanation templates aligned with traditional post-hoc strategies. These explanations were grammatically polished but fixed in form and content regardless of user behavior. In contrast, the experimental system incorporated a dialogue policy allowing it to:

- Recognize uncertainty or hesitation in user responses
- Prompt clarification requests (e.g., "Would you like a simpler explanation?" or "Should I elaborate on the trade-off here?")
- Rephrase explanations when users showed signs of confusion
- Offer rationale scaffolding (e.g., analogies or stepwise reasoning)
- Track prior user queries and tailor follow-up information accordingly

The reflexive mechanism was implemented using a finite-state dialogue manager coupled with lightweight NLP components for real-time detection of user confusion. Sentiment analysis (based on VADER) and keyword matching were used to identify low-confidence language (e.g., "I don't get it", "I think so"), while hesitation was flagged through pauses exceeding five seconds. Clarification triggers were activated when a normalized confidence score fell below 0.3, balancing responsiveness with cognitive load. Dialogue flow was governed by rule-based transitions, as reinforcement learning was not feasible given the limited interaction length. Recovery from ambiguous inputs relied on fallback templates and a short memory buffer that retained the last three user turns to ensure local coherence. However, limitations remained in interpreting vague pronouns and long-range references, occasionally leading to generic responses. Sarcasm and implicit sentiment were out of scope.

Quantitative data were analyzed using SPSS 27.0. Between-group comparisons were conducted using independent-samples t-tests for continuous variables and chi-square tests for categorical responses. Multivariate analysis of variance (MANOVA) was used to examine interactions between explanation type and dependent variables (comprehension, trust, load). Effect sizes were calculated using Cohen's d and partial eta squared (η^2).

Qualitative data were analyzed thematically using NVivo. Emergent themes were mapped against the main constructs of the study: comprehension support, trust trajectory, and collaborative engagement. Key excerpts were selected to illustrate user perceptions of dynamic explanation processes and trust repair instances.

III. Evaluation Results and Discussion

This section presents the empirical findings from our mixed-method study, integrating both quantitative analyses – centered on comprehension, performance, and trust calibration metrics – and qualitative insights derived from user interviews. Furthermore, we present a comparative evaluation of the reflexive dialogue-based explanation approach against a static explanation baseline. We conclude with a discussion that interprets key findings, evaluates the strengths and limitations of reflexive dialogue mechanisms, and explores broader implications for explainable AI (XAI).

A. Quantitative Analysis

To assess task performance, we measured participants' accuracy and task completion times across both experimental conditions: reflexive dialogue-based explanations (RD) and static explanations (SE). Participants were assigned comparable tasks requiring interpretative reasoning on AI-generated decisions (e.g., document classification, recommendation rationales).

- Task Accuracy: Participants in the RD condition achieved a mean task accuracy of 83.5% (SD = 5.8) compared to 72.4% (SD = 7.1) in the SE condition, indicating a statistically significant improvement ($t(78) = 6.41$, $p < .001$).
- Completion Time: RD participants completed tasks slightly slower (M = 3.8 min, SD = 0.6) than SE participants (M = 3.2 min, SD = 0.5), reflecting the interactive nature of dialogue-based explanation. However, this marginal time increase did not correlate with negative perceptions, as discussed in later sections.

Participants completed comprehension assessments immediately after each task. These included multiple-choice and open-ended items targeting both factual understanding and deeper conceptual clarity of the AI system's decision rationale.

- Mean Comprehension Score: RD participants scored significantly higher (M = 87.2%, SD = 6.4) than their SE counterparts (M = 70.3%, SD = 7.9), with a strong effect size (*Cohen's d* = 1.24).
- Confidence Ratings: Participants rated their understanding on a 7-point Likert scale. Mean confidence was higher for RD users (M = 5.9) than for SE (M = 4.3), suggesting a subjective perception of improved grasp of AI logic.

These results confirm the proposed hypothesis that adaptive, iterative explanations enable users to understand better by customizing the depth of information and allowing clarification through conversation.

Trust calibration was measured through the use of a pre- and post-task trust questionnaire, Drawing from the established Human-AI Trust Scale (HATS), trust was subsequently measured along the competence, reliability, and predictability dimensions.

- Trust Gain: RD participants had a statistically significant increase in trust calibration ($\Delta M = +1.1$ on a 7-point scale), while SE participants, had a negligible adjustment ($\Delta M = +0.2$).
- Appropriate Distrust: Moreover, RD participants were more likely to accurately identify incorrect AI outputs (e.g., planted misleading cases), meaning reflexive dialogue also encourages critical reflection rather than unqualified trust.

These findings underscore the idea that reflexive explainability engenders calibrated trust, allowing users to assess whether outputs are justifiable, or erroneous.

B. Qualitative Analysis

We conducted 10 semi-structured interviews with 20 randomly selected participants from each condition. Thematic analysis indicated that there was a meaningful variation in user experience, emotional engagement, and perceived control of the AI process.

a) Emerging Themes from Reflexive Dialogue Users

- Theme 1: Sense of Transparency and Control

Participants often described the system as being “more transparent” and indicated having a sense of control in guiding the exposition path. Many participants mentioned that asking follow-up questions helped the AI feel “less like a black box”.

- Theme 2: Reflective Engagement

  Participants valued opportunities to request elaborating or contesting the AI's reasoning. One participant shares, “It made me stop and think – why did it say that? And could it explain it better if I asked?”.

- Theme 3: Trust through Iteration

  Several participants described an enhanced sense of trust, not because the AI was perfect – but because of the system providing an opportunity to ‘question it’. This dialogical opportunity was also important for enhancing one participant's overall trust in the mediated code-presentation process.

b) Themes from Static Explanation Users

- Theme 1: Information Overload or Oversimplification

  SE participants criticized the explanations by describing them as either "too vague" or "packed with stuff I didn't tell it I wanted." This mismatch can create confusion or disengagement.

- Theme 2: Not Personalized

  SE participants found a one-size-fits-all approach to be frustrating, "I couldn't ask why, and I couldn't tell if it actually understand what I even needed.".

These themes have illustrated that the static explanations do not reflect the individual user's goals like the reflexive dialogue, and have a less supportive character.

C. Comparative Results: Dialogue-based vs Static Approaches

Table 1 provides a summary comparison of performance, understanding, and trust metrics between Reflexive Dialogue (RD) and Static Explanation (SE) conditions, noting that participants interacting with the RD system significantly outperformed their SE counterparts in various dimensions.

In terms of task accuracy, the mean accuracy was considerably higher in the RD group (M = 83.5%) than in the SE group (M = 72.4%), and this was statistically significant, $p < .001$. This indicates that having access to interactive, adaptive explanations provided the participants improved decisions that corresponded to more accurate outcomes. The same discomfort is evident with the comprehension scores taken from the post-task quizzes assessing their understanding of the AI rationales. Here, the mean comprehension was much higher in the RD condition (M = 87.2 % vs. 70.3 %), which reinforces the cognitive advantages of interactivity with explanation (See Fig. 1).

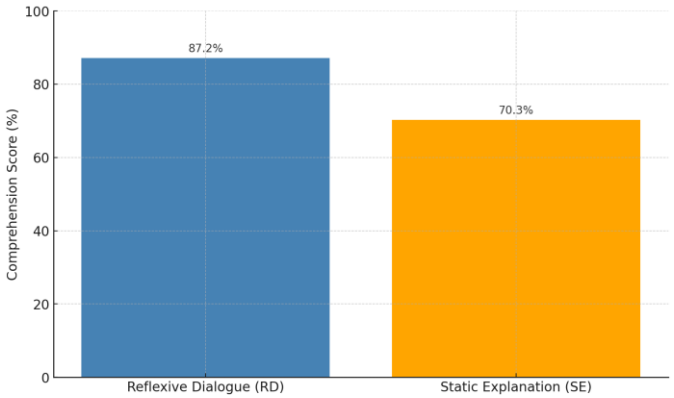

Fig. 1. Comprehension Scores by Condition

The trust calibration also found meaningful differences. Both groups started with equivalent baseline trust, but users in the RD group reported a larger increase in trust (+1.1 Likert scale points) after the interaction, while the SE group's change was marginal (+ 0.2), indicating that the dynamic dialogue supported opportunities for the development of trust through engagement and transparency (Fig. 2).

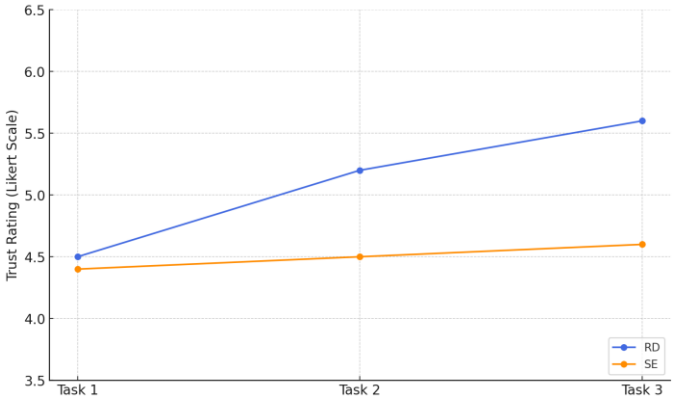

Fig. 2. Trust Calibration Across Tasks

The task completion times were slower for the RD condition (3.8 vs. 3.2 minutes), but the difference was not statistically significant suggesting that the benefits of improved comprehension and accuracy were still realized with no overwhelming time burden. The average number of clarification requests – a participant engagement metric not applicable to the SE system – also provides evidence that participants engaged with the dialogue function and utilized the explanations to their needs.

TABLE I. COMPARATIVE PERFORMANCE BETWEEN REFLEXIVE DIALOGUE AND STATIC EXPLANATION CONDITIONS

Metric	RD	SE	Significance (p)
Task Accuracy (%)	83.5	72.4	< .001
Comprehension	87.2	70.3	< .001

Metric	RD	SE	Significance (p)
Score (%)			
Trust Gain (Δ Likert score)	+1.1	+0.2	< .01
Task Completion Time (minutes)	3.8	3.2	not significant
Clarification Requests per User	3.7	N/A	–

D. Discussion

This study's results indicate that reflexive dialogue-based explanations provide pathways for improvements in human-AI collaboration by allowing users to flexibly interactively customize the explanation in ways that tracked their changing information needs. Users who used the reflexive system not only produced better performance but also developed a more complex understanding of the AIs reasoning process, which suggests that explainability can facilitate deeper cognitive integration when constructed via a dialogically reflexive process than as pre-packaged static outputs.

In addition to the development of a better understanding, reflexive dialogue was also associated with more calibrated levels of trust. Users were able to revise their judgements about the trustworthiness of the system as they were able to engage with it to clarify and elaborate. This responsiveness helped users shift to a partnership that was interpretive with the AI as opposed to over-simplifying or overwhelming. However, giving such systems reflective capabilities means that they must be able to engage meaningfully to any uncertainty expressed by the user. When there are misalignments in this interaction, as a result of irrelevant elaboration, or poorly timed clarification requests, users are likely to experience frustration or cognitive fatigue, especially in jobs where routine task processing is accompanied by minimal interactivity.

The implications include explainable AI design, where the goal post should be move from static transparency towards adapativity toward interaction models - a perspective also reinforced by research in augmented personalized narratives and tangible user interaction frameworks [29, 30]. Efficient systems will find a balance between responsiveness and efficiency, potentially in the form of a hybrid designs with layers of explanation, beginning with a summary and then expanding it based on user questions and responses. Key issues of domain transfer, personalization and error will need to be overcome before reflexive explainability ever becomes scalable beyond a controlled lab environment. This is also compounded by the dual challenge of not only developing intelligible AI, but in ways that are context aware and user sensitive, and positively supporting long-term sustainability of collaborative activity.

IV. Conclusions and Future Work

This research has shown that embedding reflexive dialogue mechanisms in an AI explanation system is a viable method for improving both the quality and propriety of human-AI interactions. Since explanations are generated interactively and adaptively, hence a better fit with how humans communicate and their cognitive needs, the reflexive explanation systems developed in this study provided improvements in understanding and performance, as well as a conceptual shift in the user, who started to see the AI as a collaborator rather than an inert tool. These findings represent something important for understanding the possibilities for shifting explanations from their traditional role as an endpoint to an ongoing process of interaction.

As for future work, I envision extending reflexive dialogue systems to other use cases or domains. Therefore we would need to include richer models of user profiles, situational awareness, and manage multi-turn conversations; a longitudinal approach would also be useful for measuring the long term impact of such systems on user learning, trust and behavior change. Finally, examining multilingual and cross-culture factors with reflexive explanations could be a valuable direction in regard explaining expectations and interaction styles across different communities could vary widely. In the end, progress in these systems will need progress in dialogue modeling, cognitive science, and participatory design, paving the way for AI that is both human-centered, and more interpretable and inclusive.

References

[1] V. C. Storey, W. T. Yue, J. L. Zhao, *et al.*, "Generative artificial intelligence: Evolving technology, growing societal impact, and opportunities for information systems research," *Inf. Syst. Front.*, 2025. [Online]. Available: https://doi.org/10.1007/s10796-025-10581-7

[2] V. Pillai, "Enhancing transparency and understanding in AI decision-making processes," *Iconic Res. Eng. J.*, vol. 8, no. 1, pp. 168–172, 2024.

[3] S. Baron, "Trust, explainability and AI," *Philos. Technol.*, vol. 38, Art. no. 4, 2025. [Online]. Available: https://doi.org/10.1007/s13347-024-00837-6

[4] D. E. Mathew, D. U. Ebem, A. C. Ikegwu, *et al.*, "Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human," *Neural Process. Lett.*, vol. 57, Art. no. 16, 2025. [Online]. Available: https://doi.org/10.1007/s11063-025-11732-2

[5] C. Papakostas, C. Troussas, A. Krouska, and C. Sgouropoulou, "A rule-based chatbot offering personalized guidance in computer programming education," in *Proc. ITS 2024: Generative Intelligence and Intelligent Tutoring Systems*, A. Sifaleras and F. Lin, Eds., *Lecture Notes in Computer Science*, vol. 14799, Cham, Switzerland: Springer, 2024. [Online]. Available: https://doi.org/10.1007/978-3-031-63031-6_22

[6] D. Mindlin, A. S. Robrecht, M. Morasch, and P. Cimiano, "Measuring user understanding in dialogue-based xAI systems," in *Proc. ECAI 2024*, U. Endriss

[7] A. M. Carreno, *Building a continuous feedback loop for real-time change adaptation: Best practices and tools*. Institute for Change Leadership and Business Transformation, 2024. [Online]. Available: https://doi.org/10.5281/zenodo.14051466

[8] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review," *Smart Learn. Environ.*, vol. 11, Art. no. 28, 2024. [Online]. Available: https://doi.org/10.1186/s40561-024-00316-7

[9] K. Chrysafiadi, C. Troussas, and M. Virvou, "Combination of fuzzy and cognitive theories for adaptive e-assessment," *Expert Syst. Appl.*, vol. 161, Art. no. 113614, 2020. [Online]. Available: https://doi.org/10.1016/j.eswa.2020.113614

[10] D. D. W. Praveenraj, M. Victor, C. Vennila, A. H. Alawadi, P. Diyora, N. Vasudevan, and T. Avudaiappan, "Exploring explainable artificial intelligence for transparent decision making," in *Proc. Int. Conf. Newer Eng. Concepts Technol. (ICONNECT-2023)*, *E3S Web Conf.*, vol. 399, Art. no. 04030, pp. 1–9, 2023. [Online]. Available: https://doi.org/10.1051/e3sconf/202339904030

[11] R. Bouderhem, "A comprehensive framework for transparent and explainable AI sensors in healthcare," *Eng. Proc.*, vol. 82, no. 1, Art. no. 49, 2024. [Online]. Available: https://doi.org/10.3390/ecsa-11-20524

[12] Z. Sadeghi, R. Alizadehsani, M. A. CIFCI, S. Kausar, R. Rehman, P. Mahanta, P. K. Bora, A. Almasri, R. S. Alkhawaldeh, S. Hussain, B. Alatas, A. Shoeibi, H. Moosaei, M. Hladík, S. Nahavandi, and P. M. Pardalos, "A review of explainable artificial intelligence in healthcare," *Comput. Electr. Eng.*, vol. 118, pt. A, Art. no. 109370, 2024. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2024.109370

[13] P. Dixit, "Assessing methods to make AI systems more transparent through explainable AI (XAI)," *Int. J. Multidiscip. Innov. Res. Methodol.*, vol. 2, no. 4, pp. 59–66, 2023. [Online]. Available: https://ijmirm.com/index.php/ijmirm/article/view/48

[14] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," *Lancet Digit. Health*, vol. 3, no. 11, pp. e745–e750, Nov. 2021. [Online]. Available: https://doi.org/10.1016/S2589-7500(21)00208-9

[15] A. Marey, P. Arjmand, A. D. S. Alerab, *et al.*, "Explainability, transparency and black box challenges of AI in radiology: Impact on patient care in cardiovascular radiology," *Egypt. J. Radiol. Nucl. Med.*, vol. 55, Art. no. 183, 2024. [Online]. Available: https://doi.org/10.1186/s43055-024-01356-2

[16] E. S. Ortigossa, T. Gonçalves, and L. G. Nonato, "EXplainable artificial intelligence (XAI)—From theory to methods and applications," *IEEE Access*, vol. 12, pp. 80799–80846, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3409843

[17] C. Chen, A. D. Tian, and R. Jiang, "When post hoc explanation knocks: Consumer responses to explainable AI recommendations," *J. Interact. Mark.*, vol. 59, no. 3, pp. 234–250, 2023. [Online]. Available: https://doi.org/10.1177/10949968231200221

[18] Z. Kanetaki, C. Stergiou, G. Bekas, C. Troussas, and C. Sgouropoulou, "A hybrid machine learning model for grade prediction in online engineering education," *International Journal of Engineering Pedagogy (iJEP)*, vol. 12, no. 3, pp. 4–24, 2022. [Online]. Available: https://doi.org/10.3991/ijep.v12i3.23873

[19] R. Shamsuddin, H. B. Tabrizi, and P. R. Gottimukkula, "Towards responsible AI: An implementable blueprint for integrating explainability and social-cognitive frameworks in AI systems," *AI Perspectives and Advances*, vol. 7, p. 1, 2025. [Online]. Available: https://doi.org/10.1186/s42467-024-00016-5

[20] C. Troussas, A. Krouska, P. Mylonas, and C. Sgouropoulou, "Personalized instructional strategy adaptation using TOPSIS: A multi-criteria decision-making approach for adaptive learning systems," *Information*, vol. 16, no. 5, p. 409, 2025. [Online]. Available: https://doi.org/10.3390/info16050409

[21] W. J. Yeo, W. Van Der Heever, R. Mao *et al.*, "A comprehensive review on financial explainable AI," *Artificial Intelligence Review*, vol. 58, p. 189, 2025. [Online]. Available: https://doi.org/10.1007/s10462-024-11077-7

[22] A. Krouska, K. Kabassi, C. Troussas, and C. Sgouropoulou, "Personalizing environmental awareness through smartphones using AHP and PROMETHEE II," *Future Internet*, vol. 14, no. 2, p. 66, 2022. [Online]. Available: https://doi.org/10.3390/fi14020066

[23] D. Hooshyar and Y. Yang, "Problems with SHAP and LIME in interpretable AI for education: A comparative study of post-hoc explanations and neural-symbolic rule extraction," *IEEE Access*, vol. 12, pp. 137472–137490, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3463948

[24] D. Mane, A. Magar, O. Khode, S. Koli, K. Bhat, and P. Korade, "Unlocking machine learning model decisions: A comparative analysis of LIME and SHAP for enhanced interpretability," *Journal of Electrical Systems*, vol. 20, no. 2s, pp. 598–613, 2024. [Online]. Available: https://doi.org/10.52783/jes.1480

[25] H. A. Tahir, W. Alayed, W. U. Hassan, and A. Haider, "A novel hybrid XAI solution for autonomous vehicles: Real-time interpretability through LIME–SHAP integration," *Sensors*, vol. 24, no. 21, p. 6776, 2024. [Online]. Available: https://doi.org/10.3390/s24216776

[26] V. Alizadeh, M. Kessentini, M. W. Mkaouer, M. Ó Cinnéide, A. Ouni, and Y. Cai, "An interactive and dynamic search-based approach to software refactoring recommendations," *IEEE Transactions on Software Engineering*, vol. 46, no. 9, pp. 932–961, Sept. 1, 2020. [Online]. Available: https://doi.org/10.1109/TSE.2018.2872711

[27] I. Salgado, M. Mera, and I. Chairez, "Suboptimal adaptive control of dynamic systems with state constraints based on Barrier Lyapunov functions," *IET Control Theory & Applications*, vol. 12, pp. 1116–1124, 2018. [Online]. Available: https://doi.org/10.1049/iet-cta.2017.1120

[28] J. Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *International Journal of Cognitive Ergonomics*, vol. 4, no. 1, pp. 53–71, 2000. [Online]. Available: https://doi.org/10.1207/S15327566IJCE0401_04

[29] D. Vallet, P. Mylonas, M.A. Corella, J.M. Fuentes, P. Castells, Y. Avrithis, "A semantically-enhanced personalization framework for knowledge-driven media services", Proceedings of IADIS International Conference on WWW/Internet (ICWI 2005)

[30] G. Trichopoulos, J. Aliprantis, M. Konstantakis, K. Michalakis, P. Mylonas, Y. Voutos, "Augmented and personalized digital narratives for Cultural Heritage under a tangible interface," 16th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP 2021), Corfu, Greece, 2021, pp. 1-5, [Online]. Available: https://doi.org/10.1109/SMAP53521.2021.9610815.