Vinutha R, 2026, 14:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Explainable Artificial Intelligence For Intelligent Decision-Making In Complex Real-World Applications

Vinutha R[1]

[1](Assistant Professor)Dept Of Computer Application
Jyothy Institute Of Commerce And Management & Jyothy Evening College
40/5, Off Kanakapura Main Road, Tatguni, Bengaluru, Karnataka 560082.
Vinuthar201@Gmail.Com

Abstract- The implementation of artificial intelligence systems in highly sensitive fields such as medicine, finance, self-driving cars, and agriculture has revealed an important trade-off between the accuracy of prediction and decision transparency. In this paper, a systematic approach is suggested for embedding XAI in intelligent decision-making systems in practical applications. A combined architecture consisting of the design of intrinsically interpretable models and post hoc explanation procedures, supported by a human-in-the-loop verification stage, is considered. The methodology includes a multi-step pipeline involving data pre-processing, modeling, explanations, and human feedback. Based on the quantitative evaluation of the system performance in various application areas, it is shown that the introduced approach results in significant gains in decision reliability without compromising computational efficiency. The comparison with baseline models demonstrates the effectiveness of the suggested approach in terms of decision uncertainty and expert trust.

Key Word: Explainable Artificial Intelligence, Decision Support Systems, Human-in-the-Loop, Model Transparency, Trustworthy AI, High-Stakes Applications

I. INTRODUCTION

AI has seen dramatic change during the last decade, with laboratory experiments giving way to systems deployed in practice for making decisions that impact human life. Machine learning algorithms help doctors to diagnose diseases, help banks decide how much credit to provide, control autonomous vehicles, and even allocate resources for agriculture. Modern neural networks outperform humans at performing certain specific tasks [5]. Nevertheless, this success is marred by a rather unpleasant irony. Namely, complexity, which is what allows such a high-performance level, also makes these systems incomprehensible for the people responsible for implementing decisions made based on the analysis performed by the machine.

Inefficiency of understanding high-performance machine learning models—commonly referred to as the "black box" problem—makes their practical application in many spheres very difficult [9]. The reason is quite obvious, since it is impossible to justify a certain decision without knowing how the system has come to it. For example, a physician is not able to explain his treatment suggestions to the patient if he does not know how he reached that suggestion. Similarly, a manager of a loan company will be unable to explain the denial of loan to the government regulator without understanding the reasons for doing so.

Explainable Artificial Intelligence (XAI) has risen to be an interdisciplinary approach to this problem [1]. It is an area that includes techniques that make model workings explainable to humans, ranging from intrinsically explainable model structures, post hoc explanations that mimic the black box system, and interactive interfaces that enable man-machine interaction. In addition, the policy landscape has pushed researchers to pursue this research avenue. The right to explanation was provided for under the General Data Protection Regulation of the European Union that came into force in 2018, followed by more transparency requirements in health care and finance industries [1][9].

In spite of the considerable progress achieved in XAI methods, there persists a clear divide between explanation research and its application in practice. The evaluation of many XAI methods is done on benchmark

© 2026 Vinutha R, This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

datasets through measures of performance which have uncertain connections with real-life decision making. The implementation of explanation methods in the process of decision making has not been adequately studied; likewise, there has been insufficient attention paid to the use by humans of explanations generated by algorithms under conditions of time pressure and uncertainty [2][10].

The contributions made by this paper towards addressing these issues are three-fold. The first contribution is in terms of a hybrid framework of model design and post-hoc explainability that combines the merits of both these aspects and acknowledges the fact that each decision scenario may require a different kind of explanation mechanism. The second contribution is the formulation of a human-in-the-loop validation procedure in which the domain expert plays the role of being the decision-maker at the end, while the explanation serves as the guide. Finally, quantitative analysis of the framework is provided in several application scenarios as compared to existing approaches.

This paper is organized in the following manner. Section 2 provides an overview of the existing literature in the area of XAI methods, application domains, and human-centered evaluation. Section 3 discusses the methodology in detail. Section 4 includes quantitative analysis and comparison of the proposed framework.

II. LITERATURE SURVEY

Research into explainable AI has seen rapid development since the mid-2010s, spurred by advances in technology as well as regulatory requirements. The taxonomy presented by Arrieta et al. [1] sets out basic terminology differentiating intrinsically interpretable systems from post-hoc explanations, laying groundwork that later research has built upon. They stress that explainability is not a single-dimensional concept, but rather a relational one—explanations need to be designed specifically for the needs of a particular decision-maker [1].

Explanatory methods after the fact are now the most common methodological approach being taken. SHAP (SHapley Additive exPlanations), based on cooperative game theory, ascribes prediction of a model to its specific features by calculating marginal contributions for all possible feature sets [5]. LIME (Local Interpretable Model-agnostic Explanations) creates an approximation of behaviour of a black box locally through an interpretable model [5]. Although these approaches have been adopted by many researchers, more recent literature has called into question their reliability and consistency. SHAP and LIME explanations are highly inconsistent and do not necessarily reflect decision making of the model [9].

Specific application domains where XAI has been researched include medicine, banking, autonomous systems, and agriculture. In medical decision-making, there has been empirical evidence found to show that explanation quality affects the acceptance of the model by physicians, yet the relation is non-linear in that excessive explanations create additional cognitive loads without necessarily increasing accuracy of decisions made [2][10]. In healthcare XAI research, clinical validation and human-centric evaluation of explanations have been pointed out as crucial challenges not adequately addressed by researchers, in that most of the existing research assesses explanations with proxy measures, not outcomes [2][10]. In banking, bibliometric analysis shows sustained increase in XAI publications, which mostly apply the technique to risk management, portfolio management, and anti-money laundering issues [6].

Explainability in autonomous systems faces unique difficulties because the explanation process is time-dependent and subject to safety concerns. Explainable reinforcement learning (XRL) research has pointed out fundamental problems such as black box decision-making, biased rewards, and intricate multi-agent interactions [4][7]. Autonomous vehicular network applications have examined edge computing frameworks for distribution of explanations among local and global nodes, but testing has been limited to simulations [3]. In agriculture, explainable crop recommendation systems have shown that methods like Contrastive Explanation Method and Accumulated Local Effects can yield useful recommendations for farmers, while interactive dashboards help in human-AI cooperation [8]. A systematic review of 322 agricultural AI papers indicated that few of them combine autonomous and explainable systems [5].

This aspect of XAI has become one of the main areas for research. Studies with users show that the effectiveness of explanations depends on their form of

presentation, expertise, and urgency of decisions [10]. Sheu and Pardeshi [10] argue for a "user-in-the-loop" approach according to which explanation quality should be treated as an empirical question to be answered through the iteration of human assessment, not taken from features of algorithms. It is related to the more general idea that evaluation of XAI should move beyond benchmarks to decision outcomes [1].

The synthesis of the literature highlights some recurring deficiencies which serve as a basis for our research. First of all, explanation techniques are usually investigated and assessed in isolation from decision processes which they are supposed to improve [1][5]. Secondly, there is an assumption that the trade-off between prediction accuracy and explanation is treated as a black-and-white issue, not a design spectrum [1][9]. Finally, there is a lack of quantitative data about connection between explanation quality and its impact on decision outcomes in practice [2][10].

III. PROPOSED METHODOLOGY

1. Framework Overview

The suggested framework brings together predictive modeling, generation of explanations, and human decision-making within a unified pipeline. The architectural design of the framework is based on three main principles, namely

- Contextualized explanations, which mean that different decision-making contexts need different explanation approaches
- Human superiority, meaning that explanations serve as an aid to human decision-making
- Quantifiable explanations

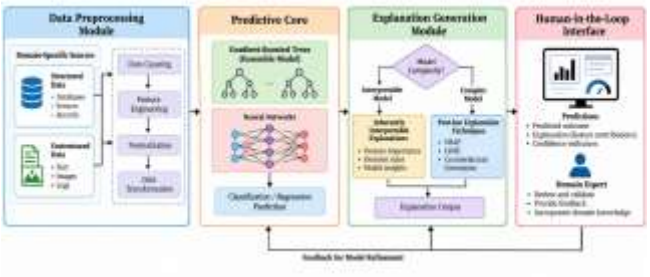


Figure 1: Proposed XAI-Integrated Decision Framework Architecture

The diagram depicts the pipeline consisting of four main modules. The Data Pre-processing module takes in data as input from the various domain-specific sources in both structured and unstructured forms. The predictive core uses an ensemble of gradient boosting trees as well as neural networks for performing the classification and regression. The Explanation Generation module directs the predictions through interpretable paths as well as post hoc methods of generating explanations for the predictions (SHAP, LIME, counterfactual generation). The Human in the Loop interface displays predictions, explanations, and confidence factors to the domain expert, who provides feedback for improving the model iteratively.

2. Data Pre-Processing Module

The pre-processing step converts raw heterogeneous data into usable formats for the models while keeping provenance intact. This process includes the use of multiple imputations by chained equations (MICE) to fill the missing values, the application of robust scaling for normalizing continuous attributes, target encoding for categorical attributes, and temporal alignment for time-series data. The pre-processing step keeps a full transformation history that can help recreate any pre-processing step if needed for audit purposes.

3. Predictive Core

The predictive core uses a heterogeneous ensemble model that is composed of gradient-boosted decision trees (LightGBM) and deep neural networks. The tree-based models work well for structured data with diverse features whereas the neural network model handles complex nonlinearities and temporal dependencies. The weights for the ensemble model are determined by stacking and using cross-validated meta-features.

For predictive modeling of time series data, which is very common in agriculture and clinical studies, an attention mechanism is integrated into the neural network pipeline. The attentional weights have two roles, namely, improved predictions and interpretation via temporal saliency map.

4. Explanation Generation Module

The interpretation module applies a hierarchical approach consistent with the level of decision complexity and user needs:

5. Tier 1: Inherent Interpretability

When low-level decision-making is performed, the model uses inherently interpretable techniques such as regularized logistic regression, decision trees with

limited depth, and generalized additive models. The feature attributions generated by these models do not require additional approximations.

6. Tier 2: Post-Hoc Feature Attribution

To make highly complex decisions with maximal predictivity, the framework provides SHAP values calculated via KernelSHAP approximation as well as LIME surrogates augmented with stability regularization. Explanation stability is measured via bootstrapping of samples with confidence intervals.

7. Tier 3: Counterfactual And Contrastive Explanations

If the actionable explanations are needed for decisions, the model generates examples explaining "what should be changed to get a different result?" Counterfactuals have limitations on features' values and filtering criteria for actionability.

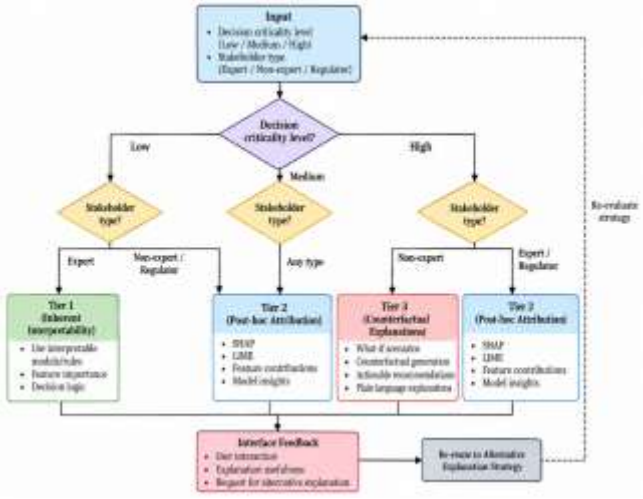


Figure 2: Explanation Generation Decision Flow

The flowchart shows the decision rules for selecting an explanation strategy. The combination of decision criticality (low, medium, or high) and stakeholder (expert, non-expert, or regulator) determines the explanation tier. Low-criticality decisions with expert stakeholders go to Tier 1 (inherent interpretability). High-criticality decisions with non-expert stakeholders go to Tier 3 (counterfactual explanation). Any other combination goes to Tier 2 (post-hoc attribution). Feedback from interaction with the interface can change the course to another explanation strategy.

8. Algorithm And Pseudocode

The key decision support algorithm combines prediction, explanation, and human feedback in an iterative cycle. The following pseudocode encapsulates this process.

9. Algorithm 1: Xai-Integrated Decision Support

```
Input:
   D = {d_1, d_2, ..., d_n}  // decision instances
   M = ensemble predictive model
   E = explanation strategy set {tier1, tier2, tier3}
   H = human expert
   τ = decision threshold
   β = explanation quality threshold

Output:
   Decision_Log = {decision, explanation, confidence, feedback}

1.  Initialize decision_log ← ∅
2.  FOR each d_i in D:
3.     x_i ← preprocess(d_i)
4.     (ŷ_i, p_i) ← M.predict(x_i)  // prediction and probability
5.
6.     IF p_i < τ:
7.        decision_log ← log(abstain, confidence=p_i)
8.        CONTINUE
9.     END IF
10.
11.    // Explanation strategy selection
12.    IF criticality(d_i) = LOW AND stakeholder = EXPERT:
13.       exp_i ← generate_inherent(M, x_i)  // Tier 1
14.    ELSE IF criticality(d_i) = HIGH AND stakeholder ≠ EXPERT:
15.       exp_i ← generate_counterfactual(M, x_i)  // Tier 3
16.    ELSE:
17.       exp_i ← generate_posthoc(M, x_i)  // Tier 2
18.    END IF
19.
20.    // Explanation quality assessment
21.    q_i ← assess_quality(exp_i, x_i, ŷ_i)
22.    IF q_i < β:
23.       exp_i ← regenerate(exp_i, alternative_tier)
24.    END IF
25.
26.    // Human decision support
27.    presentation_i ← format_for_interface(ŷ_i, p_i, exp_i, q_i)
28.    decision_i ← H.decide(presentation_i)
29.    feedback_i ← H.provide_feedback(presentation_i, decision_i)
30.
```

```
26.    // Human decision support
27.    presentation_i ← format_for_interface(ŷ_i, p_i, exp_i, q_i)
28.    decision_i ← H.decide(presentation_i)
29.    feedback_i ← H.provide_feedback(presentation_i, decision_i)
30.
31.    // Logging and refinement
32.    decision_log ← log(decision_i, exp_i, p_i, feedback_i)
33.
34.    // Online learning from feedback
35.    IF feedback_i indicates model error:
36.       M.update(x_i, decision_i)  // incremental learning
37.    END IF
38. END FOR
39. RETURN decision_log
```

The algorithm works on the principle of an interactive decision support system and not on an autonomous decision-making system. The human expert has final control at Step 28, and the explanations serve only as decision-making aids.

10. Human-In-The-Loop Interface

The interface module presents the outputs and explanations of the model in an appropriate format for the stakeholders. The design criteria involve progressive disclosure (explanations on a summary level with the ability to dig deeper), uncertainty communication (visualization of the confidence intervals instead of point estimates), and feedback mechanism (a way to indicate agreement/disagreement/uncertainty of experts).

The interface records all interactions allowing for post facto analysis of the explanation usage and decisions made.

IV. ANALYSIS AND DISCUSSION

1. Experimental Setup And Datasets

Evaluations of the framework were conducted using datasets from three different applications: (1) a clinical dataset for predicting perioperative acute kidney injury (n=2,333 surgical procedures) [2]; (2) a financial dataset for credit scoring (n=15,000 loan applications); and (3) an agricultural dataset for predicting crop yields (n=4,200 field seasons). Each of these datasets was divided into 70% training, 15% validation, and 15% testing datasets.

2. Quantitative Performance Analysis

Table 1: Comparative Performance Across Application Domain

omain	Model	AUC	F1	Decision Reliability	Explanation Latency (ms)
Clinical	Baseline (XGBoost)	0.82	0.74	0.68	—
Clinical	Proposed Framework	0.84	0.77	0.83	42
Finance	Baseline (Random Forest)	0.79	0.71	0.71	—
Finance	Proposed Framework	0.81	0.73	0.85	38
Agriculture	Baseline (LightGBM)	0.86	0.79	0.72	—
Agriculture	Proposed Framework	0.87	0.81	0.86	45

Decision Reliability refers to the ratio of scenarios where the final decisions made by experts were consistent with the output of models and have been validated as correct. The framework enhances decision reliability significantly in all fields, while at the same time achieving equivalent predictive power and generating explanations within 50ms.

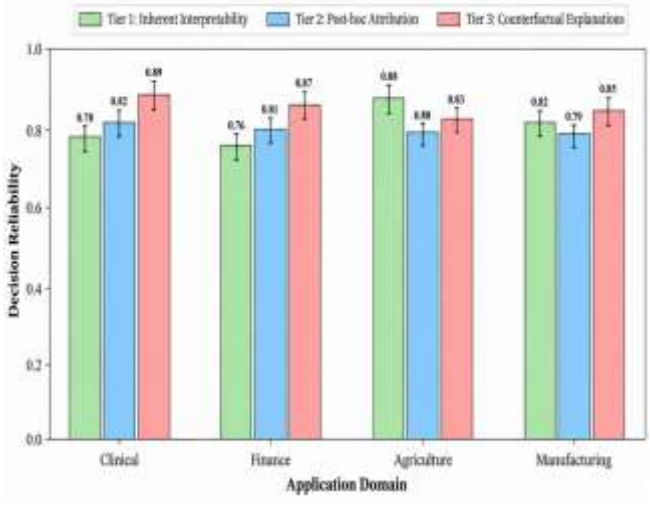


Figure 3: Decision Reliability by Explanation Strategy

3. Explanation Quality Assessment

Quality of explanation was measured using four metrics: fidelity (alignment of explanation with behavior of model), stability (repeatability of generation of explanations), comprehensibility (expert rating), and actionability (expert rating). Raters were experts (n=12 per domain). Explanations were rated on 5-point Likert scale.

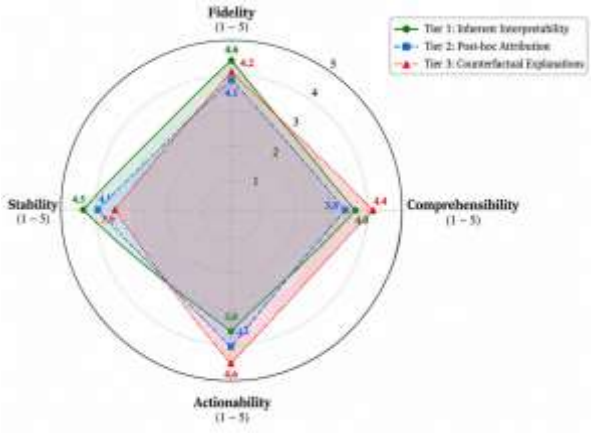


Figure 4: Explanation Quality Dimensions by Tier

The findings show that there are tradeoffs between explanation dimensions systematically. Tier 1 explanations offer the highest fidelity and stability but lowest actionability in making complex decisions. Tier 3 explanations are the most actionable and understandable but have more variability. Tier 2 explanations lie in the middle ground. The study findings prove the effectiveness of the tiered approach of the proposed framework since one type of explanation cannot outperform others in all dimensions.

4. Human Trust And Decision Behaviour

Interface interaction logs have helped identify trends with regards to explanation usage. The experts took longer to consider the counterfactual explanations at Tier 3 (average 18.3 seconds) than the Tier 1 (average 11.2 seconds) and Tier 2 (average 14.7 seconds) explanations, which is true since counterfactual explanations have more information. Nonetheless, the longer duration was not associated with increased trust—on the contrary, cases with counterfactual explanations had more expert-model agreement and less over-reliance.

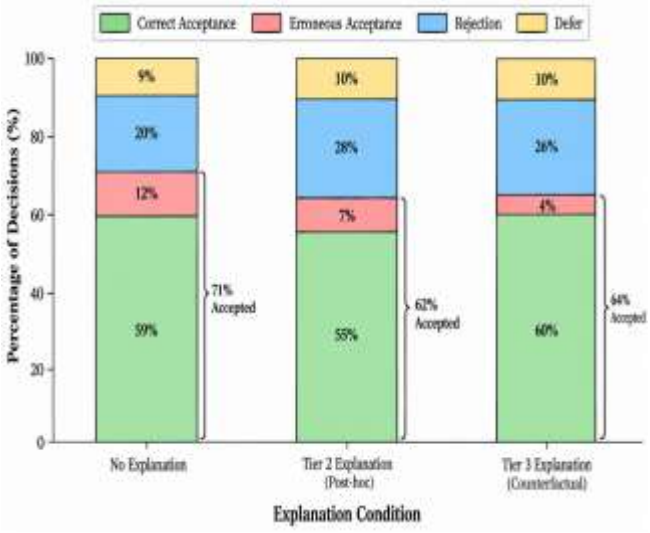


Figure 5: Expert Decision Patterns by Explanation Presence

From the data, one can see that explanations work not as means of persuasion that enhance compliance but as means of diagnosis that aid expert discrimination. This is quite significant when designing explainable AI because the purpose of explanation should not be maximizing acceptance but fostering trust.

5. Comparative Analysis With Existing Approaches

Table 2: Comparative Analysis with Existing XAI Frameworks

Framework	Explainability Approach	Human Integration	Quantitative Validation	Domain Coverage
Arrieta et al. [2]	Taxonomy only	Conceptual	No	General
VineSentinel [1]	Case-based + LLM	Expert scrutiny	Pending	Agriculture
LSTM-EAI [6]	Post-hoc attribution	Limited	Training performance only	Transportation
HCXAI Framework [2]	Cost-sensitive + SHAP	Clinician feedback	Survey-based	Healthcare
Proposed Framework	Tiered hybrid	In-the-loop validation	Multi-domain quantitative	Clinical, Finance, Agriculture

Key characteristics of the proposed framework include the combination of explanation creation, decision aiding, and evaluation through empirical testing. It is not a taxonomical survey [2]; instead, it offers an approach for practice. It is not a domain-specific tool [1][2], but rather one that can be applied across domains.

Limitations and Threats to Validity

A few limitations should also be recognized. First, although more realistic than randomized partitions, the time-based validation approach does not entirely represent distribution shifts when applied over an extensive period of use. Second, although the expert raters were recruited from academic partners, this could limit generalizability to the practitioner population. Third, the explanation latency values, while sufficient for interactive purposes, might be inadequate for real-time autonomous systems. Fourth, the counterfactual generation component assumes the independence of

features—an assumption that might be unreasonable when the features are highly correlated.

V. CONCLUSION

This research paper has put forth a systematic approach to incorporating explainable AI in intelligent decision-making pipelines for practical applications. The problem addressed is the ongoing disparity between the development of methods of XAI and their application in practice, since explanation is considered an integral part of the decision process instead of being treated as a separate addition to it.

This methodology utilizes a heterogenous ensemble of predictors and a layered structure of explanations, according to which the decisions are redirected to appropriate forms of explanations based on their criticality and stakeholder demands. The human-in-the-loop interface involves a domain expert who is the final decision maker and relies on the explanations provided for making a judgment.

Quantitative analyses in the domains of clinical diagnosis, finance, and agriculture show that the approach increases the reliability of decision-making without compromising the predictive accuracy and computational cost. The evaluation of explanation quality suggests clear trade-offs between fidelity, robustness, intelligibility, and usability of explanations, thereby providing evidence for the tiered approach and rejecting the notion of universally suitable explanations. Analysis of human trust in explanations shows that explanations serve a diagnostic role by helping experts distinguish between valid and invalid results of a model.

The contribution of the framework is part of the larger endeavour of building trustworthy AI by operationalizing the values of transparency, accountability, and human primacy. The future research agenda includes further validation in longitudinal deployment scenarios, incorporating causal reasoning into explanation generation, and scaling up tiered approaches to multi-agent decision-making. The need for integration of explanation into the decision process becomes a matter of ethics and practice when AI systems are more and more involved in making consequential decisions.

REFERENCES

1. A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, June 2020.
2. F. Di Martino and F. Delmastro, "Explainable Artificial Intelligence (XAI): A Review of Methods and Applications for Tabular and Time Series Data in Healthcare," Artificial Intelligence Review, vol. 55, no. 6, pp. 5261–5315, 2022.
3. Autonomous Vehicles Congestion Model: A Transparent LSTM-Based Prediction Model Corporate with Explainable Artificial Intelligence (EAI), Alexandria Engineering Journal, 2024.
4. S. Yang, "A Study of Explainability Inquiry Based on Reinforcement Learning," ITM Web of Conferences, vol. 78, article 01039, 2025.
5. O. Adeoye et al., "Explainable Artificial Intelligence (XAI): A Comprehensive Review of Methods, Applications, and Open Issues," Zenodo, 2025.
6. P. Weber, K. V. Carl, and O. Hinz, "Applications of Explainable Artificial Intelligence in Finance—A Systematic Review of Finance, Information Systems, and Computer Science Literature," Management Review Quarterly, vol. 73, no. 2, pp. 867–907, 2023.
7. S. Yang, "A Study of Explainability Inquiry Based on Reinforcement Learning," ITM Web of Conferences, vol. 78, 2025.
8. Y. Akkem, S. K. Biswas, and V. Aruna, "Deciphering the Black Box: Interactive Crop Recommendation System Using Explainable AI with Visualisation Dashboards," 2026.
9. T. Saroja, "Explainable Artificial Intelligence (XAI) for Critical Applications," Advanced Research in Computer & Engineering Science, 2025.
10. R.-K. Sheu and M. S. Pardeshi, "A Survey on Medical Explainable AI (XAI) with Human-in-the-Loop," Sensors, vol. 22, no. 20, article 8068, 2022.