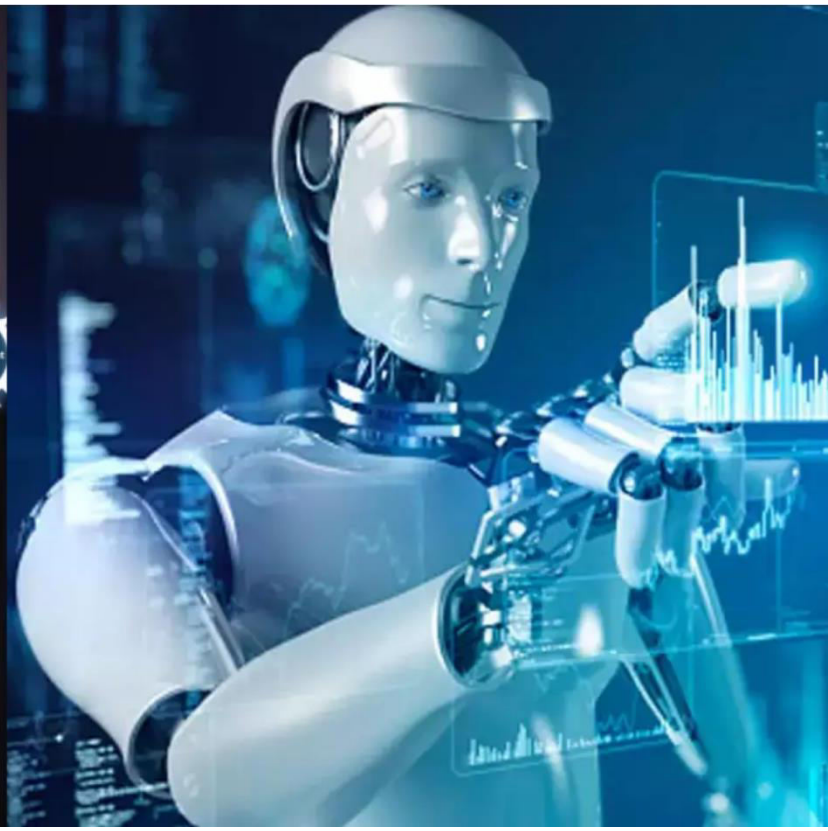
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 5, May 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505069|

Explainable Artificial Intelligence

Archana H R[1], Naseerhusen Ankalagi[2]

PG Student, Dept. of MCA, City Engineering College, Bengaluru, India[1]

Assistant Professor, Dept. of MCA, City Engineering College, Bengaluru, India[2]

ABSTRACT: XAI, or Explainable Artificial Intelligence, is an emerging area of research that concentrates on making Transparency in artificial intelligence systems, interpretable, and intelligible to human users. Although modern AI models, particularly deep learning models, achieve high accuracy and performance, they frequently serve as black-box systems, making It is difficult to comprehend How choices are generated. This inability to be interpreted can reduce user trust, create challenges in identifying errors or bias, and limit the adoption of AI in vital fields as autonomous systems, healthcare, and finance. XAI addresses these challenges by providing meaningful explanations for model forecasts using techniques such as feature importance analysis, local explanation methods, and visualization tools. These approaches help users understand the influence of input data on the output and improve confidence in AI systems. This paper presents an overview of Explainable Artificial Intelligence, including its importance, techniques, applications, and recent developments. By enhancing transparency and accountability, XAI plays a crucial part in building reliable, ethical, and human-centered systems with artificial intelligence.

KEYWORDS: Explainable AI, Interpretability, Transparency, Machine Learning, Data Visualization

I. INTRODUCTION

Artificial Intelligence (AI) has become an essential part of modern technology, with applications in healthcare, finance, transportation, and many other fields. Advanced AI models, especially those model forecasts using methods like, are capable of producing highly accurate results. However, many of these models operate as "black-box" systems, where the internal decision-making process is not easily comprehensible by people. This lack of openness creates challenges in trusting and validating AI-based decisions.

Artificial Intelligence That Is Explainable (XAI) addresses this issue by making AI systems more transparent and interpretable. It provides methods that aid consumers in comprehending how input data is processed by models and what factors influence their predictions. By offering clear explanations, XAI closes the gap between intricate AI models and human comprehension. This is especially important with crucial uses such as medical diagnostics, financial decision-making, and self-governing systems, where understanding the logic behind decisions is necessary.

Furthermore, XAI is essential to ensuring equity, responsibility, and ethical application of AI technologies. It helps identify bias and discrepancies that could result from data limitations or model design, enabling developers to improve system performance and reliability. With increasing concerns about responsible AI and regulatory requirements for transparency, the demand for explainable systems is rapidly growing. As a result, XAI has become a key research area in artificial intelligence. This essay investigates the concepts, techniques, applications, challenges, and Current Advances in Explainable Artificial Intelligence, highlighting its importance in constructing reliable and human-centered AI systems.

II. LITERATURE SURVEY

Title: Recent Applications of Explainable AI: A Systematic Literature Review
Authors: M. Saarela et al (2024)
Abstract: This paper analyzes how Explainable AI is applied in real-world domains such as financial, medical, and cybersecurity. It reviews multiple studies and identifies trends in practical applications of XAI. The research highlights The importance of explainability in improving explainability in critical systems. It also talks about difficulties encountered during implementation and suggests ways to improve The efficiency of XAI in business.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505069|

Title: **Human-Centered Evaluation of Explainable AI Applications**
Authors: J. Kim et al. (2024)
Abstract: This study focuses on evaluating Explainable AI systems from a human perspective. It examines how users perceive explanations and what makes them useful and trustworthy. The document identifies key factors such as simplicity, lucidity, and relevance of explanations. It also talks about different evaluation methods and highlights the necessity of creating XAI systems that are user-friendly and aligned with human understanding. The research highlights the importance of improving interaction between humans and AI.

Title: **Explainable Artificial Intelligence: A Systematic Review of Reviews**
Authors: C. Wilkinson et al (2025)
Abstract: This paper presents a meta-review of existing XAI surveys and identifies research gaps in the field. It combines findings from multiple studies to provide a high-level understanding of in Explainable AI.

Title: Recent Emerging Techniques in Explainable Artificial Intelligence
Authors: D. E. Mathew et al (2025)
Abstract: This project investigates novel methods developed to improve interpretability in AI systems. It focuses on enhancing transparency and making AI decisions more understandable for users, especially in critical applications. and tree-based models in capturing long-term volatility, particularly in capturing sequential patterns and long-range dependencies, In line with the results. The work emphasizes The possibility of attention-based models for deep learning for boosting cryptocurrency forecasting accuracy.

Title: A Comprehensive Survey on Explainable Artificial Intelligence Techniques
Authors: C. Wilkinson et al (2026)
Abstract: This paper provides an in-depth overview of several XAI methods and categorizes them according to their functionality and application. It discusses both traditional and modern approaches to explainability and compares their effectiveness. Additionally, the study tackles issues like balancing accuracy and interpretability. It proposes a structured framework to comprehend and evaluate different XAI methods.

III. METHODOLOGY

Existing Problem:
In Explainable Artificial Intelligence, many advanced AI models, particularly deep learning models, function as black boxes, making their decision-making process difficult to interpret This lack of openness erodes user confidence and limits their adoption in sensitive applications. Additionally, these models may produce biased or inconsistent results due to data issues, and without proper explanations, such problems are hard to detect. Additionally, there aren't enough standardized methods to evaluate explanations, making it difficult to ensure the reliability and effectiveness of AI systems.

Proposed Solution:
To address these difficulties, Intelligible Artificial Intelligence applies explainability techniques like as feature importance, LIME, and SHAP to interpret model predictions. Visualization methods like heatmaps, decision trees, and graphs are used to present results in an understandable way. These techniques help reveal how input features influence the output, making the model more transparent. As a result, it becomes easier to detect bias, improve reliability, and build user confidence in AI systems.

Proposed Methodology:
The proposed methodology for Artificial Intelligence That Is Explainable focuses on improving AI's interpretability and transparency models. Initially, A model for machine learning is trained. using relevant input data. After generating predictions, explainability techniques such as feature importance, LIME, or SHAP are applied to examine each feature's contribution to the output.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505069|

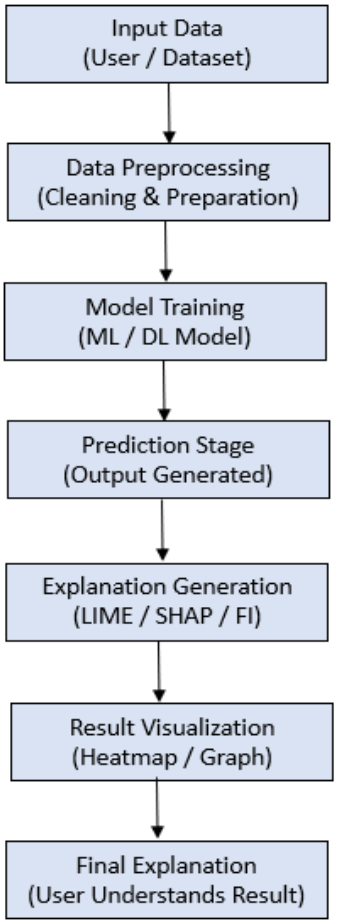


Fig 1: Proposed Methodology

IV. SYSTEM ARCHITECTURE & DESIGN

The system Explainable Artificial Intelligence Architecture begins with gathering and preparing data to prepare input data. The processed data is then fed into a model for machine learning to generate predictions. An explainability module, using techniques like LIME or SHAP, interprets the model's output, while an evaluation step checks performance and bias. Finally, the results are visualized and presented to the user, making the AI decision transparent and easy to understand.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505069|

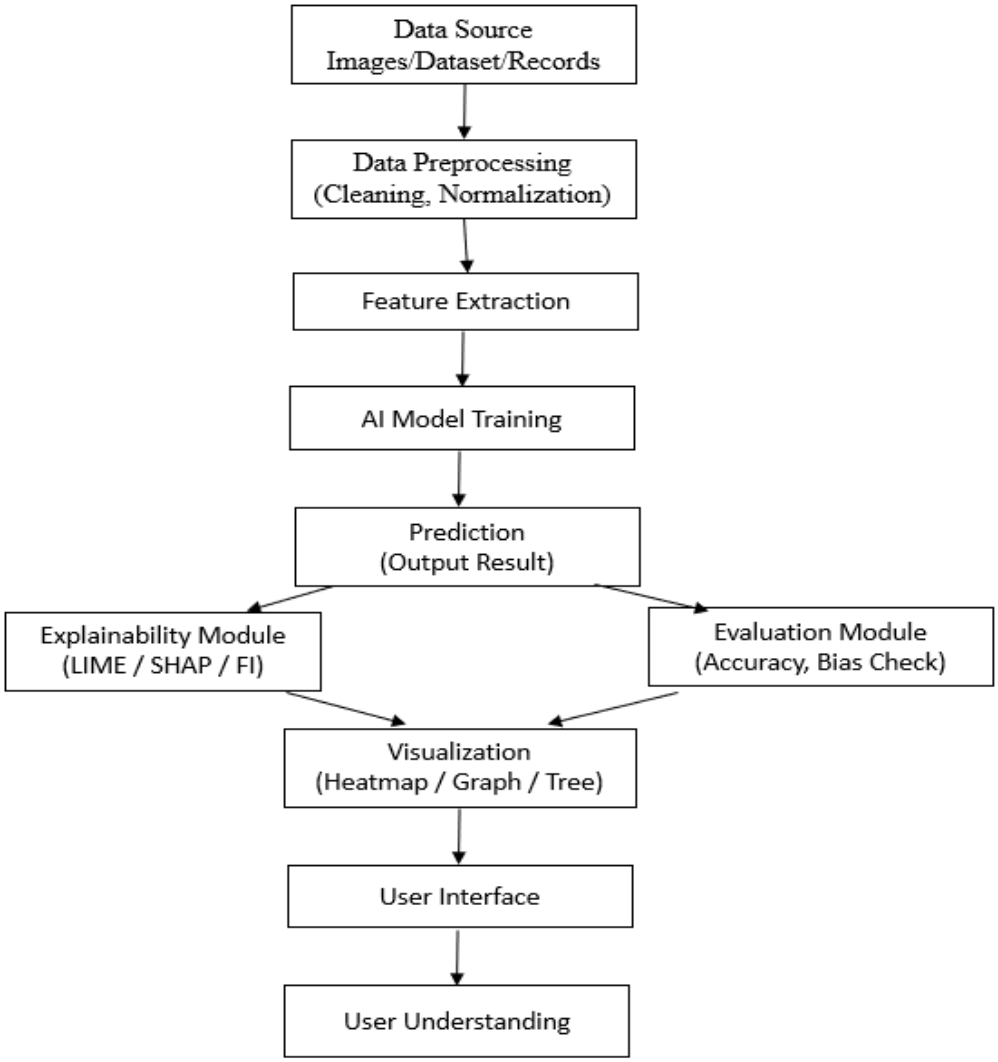


Fig 2: System Design

V. RESULTS & DISCUSSION

The proposed Explainable AI approach improves the interpretability of model predictions while maintaining satisfactory performance. By applying explainability techniques such as feature importance, LIME, and SHAP, the system is able to clearly determine the contributing input features.

most to the output. Visualization methods like heatmaps and graphs further enhance understanding by presenting results in an intuitive format.

The findings demonstrate so that users can comprehend the rationale behind AI decisions, which increases faith and assurance in the system. Additionally, the approach helps in recognizing possible prejudice and inconsistencies in the model, allowing for improvements in fairness and reliability. Compared to traditional black-box models, the proposed method provides more transparency without significantly affecting accuracy. Overall, The incorporation of explainability techniques makes the AI system more reliable, interpretable, and suitable for real-world applications.

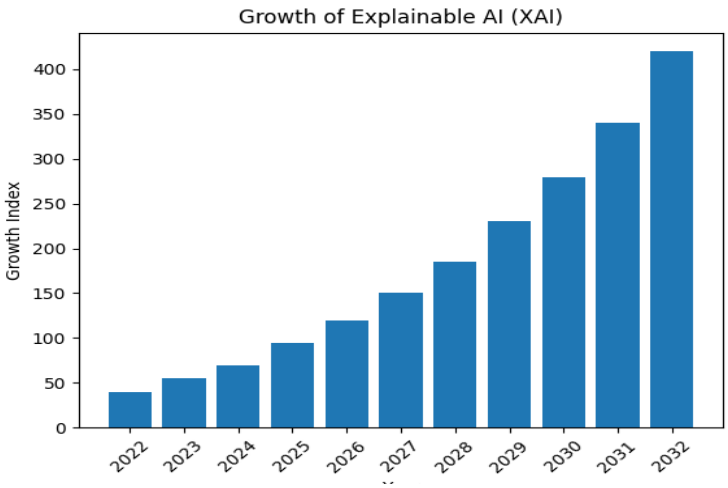
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505069|



VI. CONCLUSION

Explainable AI is essential. part in transforming complex AI systems into transparent and interpretable models. By using techniques such as feature importance, LIME, and SHAP along with visualization methods, XAI helps Users comprehend why and how decisions are made. It improves trust, detects bias and errors, and supports moral and conscientious application of AI, especially in critical applications. As the demand for reliable and accountable AI systems increases, Explainable Artificial Intelligence will become an essential part of future intelligent technologies.

REFERENCES

[1] C. Guestrin, S. Singh, and M. T. Ribeiro "Why Do I Have to Trust You? Describe the Forecasts of Any Classifier," The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Proceedings, 2016.
[2] S. M. Lundberg and S.-I. Lee, "A A Unified Method for Model Interpretation Predictions," Developments in Systems for Neural Information Processing (NeurIPS), 2017.
[3] Z. C. Lipton, "The Mythos of Model Interpretability," Queue, vol. 16, no. 3, 2016.
[4] F. Doshi-Velez and B. Kim, "Towards A Strict Science of Interpretable Machine Learning," arXiv preprint arXiv:1702.08608, 2017
[5] A. Adadi and M. Berrada, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges," Information Fusion, vol. 58, pp. 82–115, 2020.
[6] R. R. Selvaraju et al., "Grad-CAM: c via Gradient-based Localization," Deep Network-Based Visual Explanations (ICCV), 2017.
[7] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models," arXiv preprint arXiv:1708.08296, 2017.
[8] L. Longo et al., "Explainable Artificial Intelligence (XAI) 2.0: Open Challenges and Research Directions," Information Fusion, 2024.
[9] J. Kim et al., "Human-Centered Evaluation of Explainable AI Applications," Frontiers in Artificial Intelligence, 2024.
[10] S. Ali et al., "Explainable Artificial Intelligence: What We Understand and What We Don't," Artificial Intelligence Review, 2023.

.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details