Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(REVIEW ARTICLE)

Check for updates

The Rise of Explainable AI: Enhancing Transparency and Trust in Machine Learning Models

Ayodeji S. Saliu [1,*], Lucky Anthony Osayuki [2], Oliseamaka N Chiedu [3], John Cherechim Nwaigbo [4], Airat Aderoju Aroyewun [5] and Afeez Olamilekan Isiaka [6]

[1] *Department of Computer Science, Faculty of Science, Adekunle Ajasin University, Akungba Akoko, Ondo state, Nigeria.*
[2] *Department of Economics and Finance, Faculty of mgt, law and social sciences, University of Bradford, UK.*
[3] *Department of Computer Management Information System, Faculty of College of Science and Technology, Covenant University, Nigeria.*
[4] *Department of Mechanical Engineering, Faculty of Engineering, University of Nigeria, Nsukka, Nigeria.*
[5] *Department of Computer Science, Faculty of computing and Information technology, Lagos State University, Nigeria.*
[6] *Department of Systems Engineering, Faculty of Engineering, University of Lagos, Nigeria.*

Global Journal of Engineering and Technology Advances, 2025, 25(03), 080-090

Publication history: Received 28 October 2025; revised on 02 December 2025; accepted on 05 December 2025

Article DOI: https://doi.org/10.30574/gjeta.2025.25.3.0343

Abstract

Explainable Artificial Intelligence (XAI) seeks to reduce the transparency gap in modern AI and machine learning systems, particularly in high-stakes applications such as healthcare, finance, and legal decision-making. This review compares established interpretability methods, including SHAP and LIME, with emerging approaches such as perturbation-based and self-explainable models, evaluating their suitability for medical imaging, credit scoring, and regulatory compliance. The study also examines the evolving regulatory landscape, including the European Union AI Act, the implications of the General Data Protection Regulation (GDPR), and the U.S. Food and Drug Administration (FDA) guidance on AI-based medical devices. In addition, key ethical challenges related to transparency, accountability, and fairness are discussed. Although both post-hoc explanation techniques and intrinsically interpretable models have achieved significant progress, critical challenges remain. These include computational complexity, the accuracy–interpretability trade-off, diverse stakeholder requirements for explanations, and the lack of standardized evaluation metrics. Addressing these issues will require interdisciplinary collaboration across technical research, cognitive science, legal frameworks, and domain-specific expertise.

Keywords: Explainable Artificial Intelligence; Model Interpretability; AI Ethics; Healthcare AI; AI Regulation

1. Introduction

Recent advancements in artificial intelligence and machine learning have ushered in an era of unprecedented predictive capabilities, where sophisticated algorithms often outperform humans in tasks such as facial recognition and natural language processing [1]. Advanced architectures, including deep neural networks and ensemble methods, have demonstrated remarkable performance in domains such as medical diagnosis, financial forecasting, autonomous systems, and legal decision support [2, 3]. However, this progress has introduced what is commonly referred to as the *black-box problem*: as models become more powerful and complex, they often become less interpretable, even to their own developers [4]. This lack of transparency presents serious challenges in high-stakes domains where accountability, trust, safety, and compliance with legal and ethical standards are not optional but mandatory [5, 6].

* Corresponding author: Ayodeji s Saliu.

Explainable Artificial Intelligence (EAI) has emerged as a timely response to the growing transparency crisis in AI systems. It provides methodological frameworks, structural principles, and analytical tools that enhance model interpretability, demystify decision-making processes, and improve understanding among diverse stakeholders [7]. These methodologies include intrinsically interpretable models, those that are transparent by design, as well as post-hoc explanation techniques, which clarify why complex black-box systems behave the way they do. Among the most widely adopted methods are SHAP, a game-theoretic approach that assigns importance values to features [9], and LIME, a local surrogate modeling technique that approximates model predictions using interpretable representations [10]. Recent advancements also include efficient perturbation-based strategies, self-explainable architectures that integrate interpretability directly into the model structure [11], and systems powered by large language models capable of generating natural-language explanations accessible to non-technical stakeholders [12].

XAI extends beyond technical concerns, as it directly addresses fundamental issues of human autonomy, institutional credibility, and social justice [13, 14]. In the healthcare sector, where AI is increasingly used for diagnosis, treatment planning, and patient monitoring, clinicians require clear and interpretable explanations to justify algorithmic decisions, identify potential errors, and ensure accountability in patient care [15]. For example, a cancer therapy recommendation system that does not explain its decisions may achieve high statistical accuracy, but it does not support clinical decision-making and may expose patients to unacceptable risks [10]. Similarly, in finance, algorithmic decisions related to credit approval, loan terms, insurance pricing, and investment strategies must be transparent to promote fairness, detect discrimination, and allow users to challenge adverse outcomes. The 2008 financial crisis highlighted the dangers of opaque algorithmic systems, increasing the demand for explainable and accountable financial models [16].

Explainability is also coming to be viewed as a mandatory part of regulatory systems in various geographic areas that use artificial intelligence (AI) whenever it is applied to high-risk areas. The historic European Union AI Act is the creation of risk-related requirements in terms of which the high-risk AI systems must offer a reasonable degree of transparency and explanation, implying severe fines in the case of failure to comply with it [17]. The latest suggestions related to systematic-risk AI models further demarcate these requirements by providing documentation criteria to documents and descriptive requirements [18]. In the United States, the Food and Drug Administration has issued detailed principles to explainability of AI/ML-enabled medical devices, noting that opaque systems create unacceptable risks to clinical safety [19]. Though under the current understanding of the law, the General Data Protection Regulation (GDPR) has been the cause of numerous controversies about the reasons and how one must explain automated decision-making to the individuals affected [20], these legal changes altogether serve to reinforce a growing pressure within the societal mindset regarding the need to make AI more transparent to be deployed efficiently.

Another, albeit lesser, aspect of the explainable AI (XAI) requirement relates to ethics. Explainability, as Vainio-Pekka and Hakkarainen claim, is one of the principles that support the ethical requirements of fairness, accountability, and consideration of human autonomy [21]. Explainability of AI decisions is also lacking, and this prevents the detection and removal of biases and allows discriminatory trends to remain intact [22, 23]. Adams believes that explicability has a moral value in and of itself, independent of the instruments, and believes that people are granted a substantive right to know decisions that affect their lives, health, and chances [24]. This is because this view re-dresses explainability as the issue of human dignity and democratic accountability as opposed to a technical or regulatory challenge. Morley et al. note that though the concept of explainability is reiterated in a range of AI ethics models created by governments, industry organisations, and academic organisations, the definition, operationalisation, and application of the concept vary significantly [25].

Still, thematic despite significant progress and achievements, an array of obstacles remains that prevent developing and implementing XAI. Some technical challenges include computational complexity of real-time applications, the difficulty of explaining ensembles and deep architectures, inconsistencies between local and global explanation modes, and the long-standing debate of the accuracy-interpretability trade-off. The lack of standardised evaluation metrics, challenges in differentiating the quality of an explanation and ground truth, and context-dependence of what makes a decent explanation can be listed as methodological weaknesses [26]. Many different explanatory modalities are needed by stakeholders, data scientists, domain experts, regulators, and end-users, but modern systems often support only a single explanatory mode, one-size-fits-all explanations [27, 28]. Furthermore, even the practice of explaining things can be doctored or misused to project an illusion of trust in inadequate systems - what is sometimes called the explanations theatre.

XAI represents an interdisciplinary project per se that requires assembling historically separate fields, namely, computer science, statistics, cognitive science, law, ethics and domain expertise. Technical solutions that do not incorporate human mental operations can produce an output of technically perfect but practically useless nature [29,

30]. Laws built without a sense of technical imperatives run a risk of creating demands that are difficult to meet. Similarly, the ethical propositions that are not created within the context of application might be abstract and unrealistic. In line with this, a rigorous XAI requires a long, discursive process of co-development among the diverse communities of experts, whose voices will be essential in addressing the issue of giving the AI systems transparency, credibility and accountability to human values.

It is an empirical review of the existing literature that seeks to give an overall view of how XAI evolves over time, its methodology, application, regulatory framework, ethical issues, evaluation methods, and the future of the field. The review is structured as follows: Section 2 gives an overview of the existing interpretability techniques and the latest developments in the field; Section 3 outlines the deployment of XAI in healthcare context as well as in financial and legal settings; Section 3.3 better outlines regulatory requirements and compliance standards; Section 4 presents the ethical considerations and implementation concerns; Section 5 surveys the current evaluation and validation techniques; and Section 6 outlines the current problems and future research prospects.

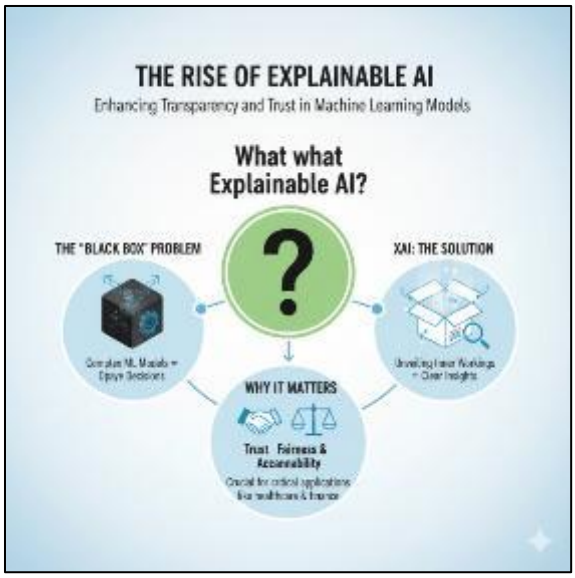

Figure 1 The XAI Ecosystem and Interdisciplinary Framework

2. Fundamental Concepts and Techniques

The perspective that is adopted determines whether a model is interpretable or not. It is possible to identify two main approaches: model transparency and model understanding.

The XAI methods can be arranged into two broad categories, including intrinsic interpretability and post-hoc explainability [3]. Intrinsic interpretable models are transparent in nature, with decision trees and linear regression being examples of these. Post-hoc explainability approaches are adopted once complex models are trained with the view of explaining their predictions [8].

2.1. SHAP (Shapley Additive exPlanations)

SHAP, as proposed by Lundberg and Lee [1], offers a common mechanism for explaining the output of any model with cooperative game theory. The approach gives every attribute a factor that represents the contribution that the specific attribute makes to a given forecast, thus ensuring reliability and local validity. SHAP values have three desirable properties: local accuracy, missingness, and consistency [1]. They have received extensive favor among researchers because of their theoretical basis and their model-agnostic character [16].

Most recent uses testify to the effectiveness of SHAP in a wide range of areas. Das Neves et al. [16] demonstrated that the stability of SHAP-based explanations increased the level of understanding of credit-scoring decisions among the stakeholders, but did not affect the model performance. SHAP has enabled the interpretation of breast cancer diagnostic models [10] and also the analysis of radiology reports [13] in healthcare.

2.2. LIME (Local Interpretable Model-agnostic Explanations)

The other outstanding technique is LIME, which was suggested by Ribeiro et al. [2]. LIME uses individual predictions to approximate the target model by a simple and interpretable surrogate model locally. This method is a distortion of input data and tracking the change in prediction to determine impactful features [2]. Since LIME is model-agnostic, any classifier can be used, but it can also be unstable and inaccurate on related cases [8].

According to comparative studies, SHAP and LIME have complementary strengths. SHAP offers feature importance consistency worldwide, compared to the term LIME, which offers faster local explanations computation [16]. These two practices have been useful in financial aspects, especially in compliance with regulations and contacting the customers [14].

2.3. Emerging Techniques

Perturbation-based methods like PeBEx [4] have been demonstrated to be more efficient than earlier methods, more recently. At the same time, a body of research is also working on using large language models to create natural-language explanations of AI decisions [28], which could be the key to overcoming the divide between technical and human understanding.

Table 1 Comparison of Major XAI Techniques

Domain	Application Area	Common Techniques	Key Benefits	Implementation Challenges
Healthcare	Medical Imaging [7], [12]	SHAP, LIME, Grad-CAM, Attention Maps	Clinical validation, improved diagnosis, increased trust	Integration with clinical workflow, real-time processing
Healthcare	Disease Prediction [9]	SHAP, Decision Trees, Rule-based Systems	Early intervention, personalized treatment	Data privacy, model generalization
Healthcare	Personalized Monitoring [11]	SHAP, Feature Importance	Patient engagement, adherence	Continuous validation, sensor reliability
Finance	Credit Scoring [16]	SHAP, LIME	Regulatory compliance, customer satisfaction	Model complexity, fairness metrics
Finance	Fraud Detection [14]	Anomaly detection with explanations	Reduced false positives, transparency	Real-time constraints, adversarial robustness
Finance	Risk Assessment [15]	SHAP, Counterfactual Explanations	Stakeholder trust, model validation	Market volatility, model drift

3. Domain-Specific Applications

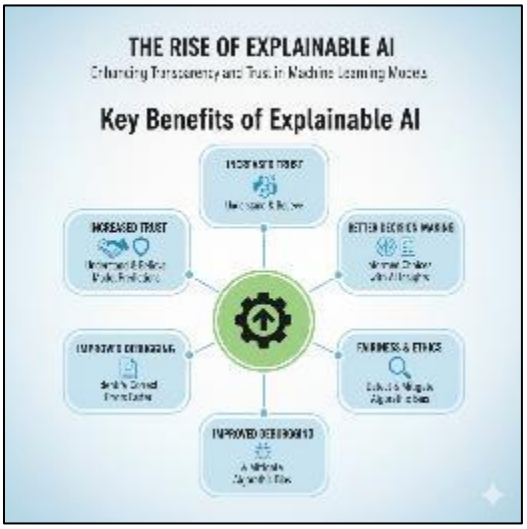

Figure 2 XAI Application Workflow in Domain-Specific Contexts

3.1. Healthcare and Medical Imaging

Healthcare is one of the most problematic areas where XAI may be used. Extreme cases of medical decisions should be handled by artificial intelligence systems capable of giving valid justifications to their suggestions [9]. According to recent systematic reviews, XAI has already received a substantial amount of applications in medical imaging, especially in the context of software created using deep-learning algorithms to interpret radiological images, pathology slides, and clinical scans 7, 12.

Housesein et al [12] carried out a systematic survey of explainable AI in medical imaging, and the most widely used to date include SHAP, LIME, and Grad-Cam. Their review proved that such XAI methods have been successfully used to diagnose the conditions of pneumonia and various cancers. In the more narrow scope of breast cancer detection, Ghasemi et al. [10] demonstrated that explanatory models not only achieved high accuracy, but they also increased clinician reliability and uptake.

Personalized health monitoring is another potential field of application. An explainable AI system was created by Vani et al. [11] to oversee continuous health-related data capable of assisting the patient and clinician to interpret the information presented in a way that is understandable in terms of its trends and risk factors. It was determined that the system enhanced patient interactions and compliance with treatment in comparison to black-box methods.

One of the issues that dominates in medical XAI is the complexity versus interpretability trade-off. This has been demonstrated to be solved by self-explainable AI models that apply interpretability in the architectural design and not post hoc on the architectural design [29]. Such models can bring out pertinent parts of the image or clinical features involved in the diagnostic process and generate real-time explanations that are consistent with clinical thought processes.

3.2. Finance and Credit Scoring

The advancement of XAI has been experienced in the finance sector as it is pressured by regulating agencies and requires assurance to the customers [14]. Theories of credit scoring, credit issuance, credit fraud detection, and risk assessment are not just desirable but in most instances, they are even legally required [15].

The SHAP-based explanations and LIME-based explanations of the credit-scoring models were also utilized by Das Neves et al. [16], who proved that pointwise explanations could be both effective and did not violate the regulatory context. They discovered that customers provided with a description of credit decisions were more satisfied and willing to take corrective actions in order to become creditworthy.

The industry report compiled by CFA Institute [15] has highlighted that the importance of explainability in finance is not only based on the necessity to comply with the regulations but also on the management of risks, validation of models,

and communication with stakeholders. Financial institutions have identified that explainable models are a necessary instrument towards establishing fair and transparent decision-making since they can enable the identification of possible biases and unexpected links in the lending and investment decisions.

Transparency, fairness, and accountability were found to be the key drivers behind the adoption of XAI, which was performed through a systematic review of the XAI in finance by Cernviciene et al. [14]. Although there is an increase in the utilization of XAI, they observed that there are still issues in the effort to explain ensemble models, as well as neural networks that are frequently being used in algorithmic trading and risk modelling.

Table 2 XAI Applications in Healthcare and Finance

Domain	Application Area	Common Techniques	Key Benefits	Implementation Challenges
Healthcare	Medical Imaging [7], [12]	SHAP, LIME, Grad-CAM, Attention Maps	Clinical validation, improved diagnosis, increased trust	Integration with clinical workflow, real-time processing
Healthcare	Disease Prediction [9]	SHAP, Decision Trees, Rule-based Systems	Early intervention, personalized treatment	Data privacy, model generalization
Healthcare	Personalized Monitoring [11]	SHAP, Feature Importance	Patient engagement, adherence	Continuous validation, sensor reliability
Finance	Credit Scoring [16]	SHAP, LIME	Regulatory compliance, customer satisfaction	Model complexity, fairness metrics
Finance	Fraud Detection [14]	Anomaly detection with explanations	Reduced false positives, transparency	Real-time constraints, adversarial robustness
Finance	Risk Assessment [15]	SHAP, Counterfactual Explanations	Stakeholder trust, model validation	Market volatility, model drift

3.3. Legal and Regulatory Contexts

The intersection of XAI and law remains complex and evolving. While the "right to explanation" has been widely discussed in relation to the GDPR, legal scholars debate its actual scope and enforceability [20, 21]. Wachter et al. [20] argued that the GDPR does not explicitly grant a right to explanation for automated decisions, while Kaminski [21] countered that various GDPR provisions collectively establish such a right.

Recent developments suggest increasing regulatory emphasis on explainability. The European Union's AI Act [17, 18] introduces risk-based requirements where high-risk AI systems must provide appropriate levels of transparency and explainability. The U.S. FDA has similarly established guidelines requiring explainability for AI/ML-enabled medical devices [19].

Papadimitriou et al. [22] advocate for rethinking the right to explanation beyond mere technical disclosure, arguing that meaningful explanations must be accessible, understandable, and actionable for affected individuals. This perspective emphasizes that legal frameworks must evolve alongside technical capabilities to ensure genuine accountability.

4. Ethical Considerations

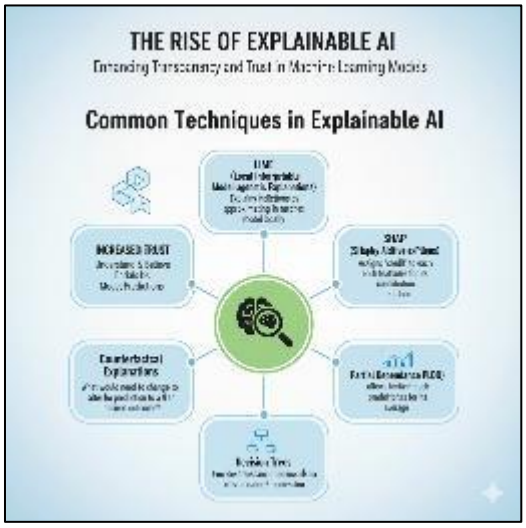

Figure 3 Ethical Principles and Implementation Challenges in XAI

4.1. XAI and AI Ethics

It is well known that explainability is a key concept on which the ethical application of artificial intelligence is based [23,24]. According to Vainio-Pekka and Hakkarainen [23], explainability helps to operationalize such ancillary norms of ethics as fairness, accountability, and transparency. Without a thorough knowledge of how AI systems make decisions, the identified and corrected biases will be practically impossible.

Adams [24] explicability is identified as a separate ethical principle per se and assumes that it is more than a vehicle to achieving other ethical aims. Its value lies in the protection of human autonomy and dignity by ensuring that people are informed about the decisions that affect them sufficiently.

In a survey carried out by Morley et al. [25], focusing on the current AI ethics advocacies, it has been identified that the concept of explainability is assigned to all main principles of the guidelines, which arise due to the efforts of governmental agencies, industries, and institutions. However, the authors also found that there was a large range of heterogeneity both in the definitions and in the implementations of explainability, thus highlighting the need for a greater level of concrete operationalization.

4.2. Implementation of Ethical Issues.

Several tensions are associated with the implementation of ethical XAI. First, the accuracy and interpretability of the model can be subject to a possible trade-off, but current studies have shown that such a trade-off can perhaps be less intensive than was previously evaluated [8]. Second, explanations themselves can be manipulated or abused, and this will lead to the development of a false impression of trust in flawed systems [25].

Third, the decision of whether or not an adequate explanation exists is dependent on circumstances. An appropriate technical explanation can be inadequate to end-users or regulators [27]. Taylor [27] explores the philosophical foundations of explanation, where he argues that the right to an explanation should be understood according to its intended application in contestation, understanding, or accountability.

Table 3 Ethical Principles and XAI Requirements

Ethical Principle	XAI Requirement	Implementation Challenge	References
Transparency	Clear documentation of model decisions	Balancing detail with comprehensibility	[23], [25]
Accountability	Traceable decision pathways	Identifying responsible parties in complex systems	[22], [24]
Fairness	Bias detection and mitigation	Defining fairness metrics, intersectionality	[16], [25]
Privacy	Explanation without revealing sensitive data	Information leakage through explanations	[20], [23]
Autonomy	Meaningful human oversight	Over-reliance on automation	[21], [27]
Beneficence	Explanations that improve outcomes	Avoiding explanation theater	[24], [25]

5. Evaluation and Validation

5.1. Evaluation Frameworks

There are special issues with the evaluation of XAI methods due to the aspects of both objective and subjective explanations of the quality. Nauta et al. [8] suggest an extended framework where the distinction lies between explanation-level metrics, which include fidelity and consistency, and human-level metrics such as usefulness and comprehensibility.

Salih et al. [26] reviewed evaluation methods in XAI and proposed three core categories: functionally grounded evaluation, which measures explanation accuracy; application-grounded evaluation, which assesses usefulness to domain experts; and human-grounded evaluation, which examines how end users understand the explanations. They argue that effective XAI systems should satisfy all three evaluation types.

The systematic framework of explainability in AI systems by Hu et al. [30] was based on both quantitative metrics and qualitative assessment of explainability. Their framework deals with five important dimensions, namely completeness (whether the explanation covers all the relevant factors); soundness (whether the explanation is true to the model); efficiency (whether explanations can be produced with the constraint of a reasonable time); effectiveness (whether the user can understand explanations); and usability (whether explanations are accessible to the purpose users).

5.2. Challenges in Evaluation

Table 4 XAI Evaluation Metrics and Methods

Evaluation Type	Metrics / Methods	Advantages	Limitations	References
Functionally-Grounded	Fidelity, Consistency, Stability	Objective, reproducible	May not reflect user needs	[8], [26]
Application-Grounded	Expert assessment, Task performance	Domain-relevant, practical	Resource-intensive, subjective	[26], [30]
Human-Grounded	User studies, Comprehension tests	Measures actual understanding	Expensive, variable	[8], [30]
Quantitative	Feature importance correlation, Explanation size	Measurable, comparable	Incomplete picture	[26], [30]
Qualitative	Interviews, Think-aloud protocols	Rich insights, context-aware	Hard to generalize	[8], [26]

Various methodological issues make the process of XAI evaluation more difficult. To begin with, no ground truth to explanations usually exists, especially in tricky fields where human experts might come into conflict. Second, different stakeholders have different preferences to be addressed through explanation; what is good to the regulator might not be important to the needs of a clinician. Third, evaluation studies often lack ecological validity because they are tested on controlled environments as opposed to situations of real-world application. Fourth, some metrics like faithfulness and stability may be incompatible, thus there must be trade-offs in system design.

6. Current Challenges and Future Directions

6.1. Technical Challenges

Even now, XAI is facing a number of technical challenges. Mersha and Adi [5] note that a major challenge has been computational complexity, especially when it comes to real-time applications. The exponential complexity of SHAP cannot be scaled at the cost of its imprecision, and approximation methods can be used at the cost of its precision. Model-agnostic methods are often not able to deal with complicated architectures like ensemble models and deep neural networks. The perturbation-based approaches have challenges in determining relevant perturbation strategies, and can produce conflicting explanations of similar inputs. The conflict between the local and global explanations is still pending. Local explanations can be used to explain a set of individual predictions, but might not give any systematic biases or trends, due to which global explanations can be used in offering a general picture of the behaviour of the complex model, but risk the individual behaviour being too simplified.

The absence of tangible connections among the various disciplines negatively impacts the relationship between psychology and these other fields.

6.2. Interdisciplinary Integration

Successful XAI requires interactions between disciplines. Kabir et al. [6] highlight the importance of using the knowledge of cognitive science, social science, and domain expertise to inform the technical solutions. This is essential in the development of successful XAI systems since it is important to understand the way humans interpret explanations. Legal and regulatory experience are also required. Both technical development and law should be informed by the philosophy of the explanation as Taylor [27] asserts. The domain experts, lawyers, computer scientists, and ethicists have to work together to create XAI systems that are both technically, ethically, and legally sound.

6.3. Emerging Directions

There are a number of areas of potential research. Large language models have a promise of obtaining natural-language explanations that can bridge technical and lay interpretations [28]. The self-explainable models incorporate interpretability into model architecture, which may eliminate the accuracy-interpretability trade-off. Intuitive and practical explanations are found in counterfactual ones, which explain what would have to change the predictions. The approach of causal explanation that goes beyond correlation to find causal relationships is a research frontier in XAI studies.

Table 5 Current Challenges and Research Directions in XAI

Challenge Area	Specific Issues	Proposed Solutions	Future Research Needs	References
Computational Efficiency	Scalability of SHAP, real-time constraints	Approximation methods, efficient algorithms like PeBEx	Better complexity–accuracy trade-offs	[4], [5]
Explanation Quality	Fidelity vs. simplicity, consistency	Multi-level explanations, adaptive systems	User-centered design research	[8], [26]
Domain Adaptation	Medical, financial, and legal contexts	Domain-specific XAI frameworks	Cross-domain transfer of methods	[7], [9], [14]
Regulatory Compliance	Varying requirements across jurisdictions	Standardized explanation formats	Legal–technical collaboration	[17], [18], [22]

Evaluation Standards	Lack of consensus metrics	Comprehensive evaluation frameworks	Benchmark datasets for XAI	[26], [30]
Human Understanding	Cognitive limitations, explanation overload	Natural language interfaces, LLM integration	Cognitive science insights	[8], [28]

7. Conclusion

Clear-cut AI has gone beyond a niche area of focus to a mandatory component of responsible AI applications. As it has been shown in this review, significant progress has been made in terms of the creation of interpretability methods, the use of XAI in various fields, and the creation of regulatory policies.

However, there remain serious challenges. The conflict between model performance and interpretability exists, but recent research also indicates that the dichotomy might not be as pronounced as previously assumed. The methodologies of evaluation should be further enhanced to capture the complex nature of the quality of an explanation. Legislation is also a changing concept, and therefore, a consistent discussion between technical, legal, and ethical stakeholders is required.

The knowledge of cognitive science should drive technological advancement under the influence of ethical values and the necessities of regulation. The need to have transparent, interpretable, and trustworthy models is bound to increase as AI systems play a more important role in the key decision-making processes.

Finally, XAI goes beyond black box demystification; it is supposed to promote trust, make AI systems accountable, and bring them in line with human values and social objectives. The further evolution of XAI is a necessary measure that will help to achieve the promise of artificial intelligence and reduce the risks associated with it.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems (NeurIPS), pp. 4765–4774, 2017.

[2] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the predictions of any classifier," Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144, 2016.

[3] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed. Leanpub, 2022.

[4] I. Gómez-Talal, "PeBEx: Efficient perturbation-based explanations," Journal of Machine Learning Research, vol. 26, no. 134, pp. 1–25, 2025.

[5] M. Mersha and W. B. Adi, "Explainable artificial intelligence: A survey of needs, techniques and challenges," Artificial Intelligence Review, vol. 57, pp. 1–38, 2024.

[6] S. Kabir et al., "A review of explainable artificial intelligence," Applied Sciences, vol. 15, no. 4, pp. 1223–1244, 2025.

[7] S. Nazir et al., "Explainable artificial intelligence techniques for biomedical imaging," Frontiers in Radiology, vol. 3, pp. 1–15, 2023.

[8] M. Nauta, R. van Bree, and C. Seifert, "Evaluating explainable AI: Which explanation is best?," Machine Learning, vol. 111, pp. 455–489, 2022.

[9] Z. Sadeghi et al., "Explainable AI in healthcare: A review," Healthcare Analytics, vol. 4, pp. 100–129, 2024.

[10] A. Ghasemi et al., "Explainable artificial intelligence approaches in breast cancer detection," Computer Methods and Programs in Biomedicine, vol. 241, pp. 107–128, 2024.

[11] M. S. Vani et al., "Personalized health monitoring using explainable AI," Scientific Reports, vol. 15, no. 2334, pp. 1–12, 2025.

[12] E. H. Houssein et al., "Explainable AI for medical imaging: Methods and applications," Neural Computing and Applications, vol. 37, pp. 21241–21265, 2025.

[13] M. T. Zamir et al., "Explainable AI-driven analysis of radiology reports," JMIR Formative Research, vol. 9, e54321, 2025.

[14] J. Černevičienė et al., "Explainable artificial intelligence in finance: A systematic review," Journal of Financial Innovation, vol. 12, pp. 45–67, 2024.

[15] CFA Institute, Explainable Artificial Intelligence in Finance, CFA Industry Report, 2024.

[16] A. das Neves et al., "SHAP and LIME-based explainability for credit scoring models," Expert Systems with Applications, vol. 235, pp. 119–142, 2024.

[17] O. A. Mavisclara, I. I. Oshobugie, A. T. Olufunmi, A. A. Bolaji, A. O. Olawale, N. C. Anulika, and O. K. Samuel, "The AI-driven energy surge: A comprehensive review of sustainable power solutions for data centers," Glob. J. Eng. Technol. Adv., vol. 24, no. 3, pp. 328–344, 2025, doi: 10.30574/gjeta.2025.24.3.0279.

[18] O. Dosunmu, A. Adeoye, P. Ukonu, C. Offo, A. Izuchukwu, and O. Akadiri, "A comparative study of data visualisation techniques for effective decision-making in business intelligence," Path Sci., vol. 11, no. 8, pp. 1058–1068, 2025, doi: 10.22178/pos.121-45.

[19] O. Akadiri, O. Babatunde, A. A. Jamiu, O. R. Igbape, B. O. Samson, and C. F. Anyaehie, "Collaborative intelligence databases (CID): Harnessing AI for privacy-preserving multi-source data management," Glob. J. Eng. Technol. Adv., vol. 24, no. 3, pp. 406–416, 2025, doi: 10.30574/gjeta.2025.24.3.0289.

[20] S. Wachter, B. Mittelstadt, and L. Floridi, "Why a right to explanation of automated decision-making does not exist in the GDPR," International Data Privacy Law, vol. 7, no. 2, pp. 76–99, 2017.

[21] M. E. Kaminski, "The right to explanation, explained," Berkeley Technology Law Journal, vol. 34, pp. 189–218, 2019.

[22] T. Papadimitriou et al., "Rethinking the right to explanation," AI and Ethics, vol. 5, pp. 220–239, 2024.

[23] H. Vainio-Pekka and M. Hakkarainen, "The role of explainable AI in AI ethics," ACM Journal on Responsible Computing, vol. 1, no. 2, pp. 1–19, 2023.

[24] J. Adams, "Defending explicability as a principle for the ethics of AI," Bioethics, vol. 37, no. 3, pp. 135–150, 2023.

[25] J. Morley et al., "The ethics of AI: Tools and frameworks," AI & Society, vol. 37, pp. 109–130, 2022.

[26] A. M. Salih et al., "Evaluation approaches for explainable artificial intelligence: A review," Information Fusion, vol. 103, pp. 102017, 2024.

[27] E. Taylor, "Explanation and the right to explanation," Philosophy & Technology, vol. 37, pp. 1–25, 2024.

[28] A. Bilal, D. Ebert, and B. Lin, "Large language models for explainable AI: A survey," arXiv preprint arXiv:2502.01234, 2025.

[29] Y. Chen et al., "Self-explainable AI models for medical image analysis: A review," Medical Image Analysis, vol. 94, pp. 103001, 2024.

[30] L. Hu et al., "A systematic framework for evaluating explainability in AI systems," IEEE Transactions on Artificial Intelligence, vol. 6, no. 2, pp. 250–268, 2025.