IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

The Future Of Explainable AI And The Challenges Associated With It

Homeshwar Gupta[1] , Aarti Pandey[2] , Navita Srivastava[3],

[1,2,3,]Department of Computer Science, A.P.S. university, Rewa (M.P.)

Abstract: Explainable Artificial Intelligence (XAI) is rapidly increasing in importance as AI systems are deployed in multiple sectors such as healthcare, finance, autonomous systems, educational research, and policymaking. Previous AI models, even though they are accurate, often work as "black boxes." They offer minimal insight into how decisions are made. XAI aims to address this opacity by ensuring AI models are interpretable, fully transparent, and accountable. This increases trust, ensures ethical compliance, and improves collaboration between human and AI systems.

This paper has brief discussion of the future of explainable AI (XAI), its uses, technical and ethical challenges, and create ways to design strong and reliable AI frameworks that meet human and legal expectations.

Keywords: Explainable AI, XAI, Machine Learning, Transparency in AI, Trust in AI, Challenges

1. INTRODUCTION

Artificial Intelligence (AI) has developed the generation in various fields, enabling complex problem-solving techniques with efficiency. Some AI models, especially deep learning networks work in hidden manner. Their internal logic is not readily understandable, which poses significant difficulties in critical applications.

For instance, in field of healthcare, AI systems help in diagnosis and treatment planning, but due to lack of an explained decision can adversely impact patient health. In finance, black-box algorithms do not work well, leading to wrong credit decisions and flawed risk assessments. Making these systems understandable, reliable and transparent is important.

Moreover, transparent decision-making allows experts to review recommendations, catch tiny errors, and re-correct processes — ultimately leading to better results for society's future.

2. AI EVOLUTION & BLACK-BOX PROBLEM

In beginning, AI with rule-based systems followed explicit instructions to perform any task. But with time, computational power and data availability increased, AI evolved into machine learning models capable of learning patterns from data and taking decisions. Deep learning and reinforcement learning model have grown in ai capabilities but reduced interpretability.

Black-box models are very good and high performing models but they have almost no explanation for their decisions. Transparent models like explainable AI increase accuracy with clarity and making humans believe on the results provided by them.

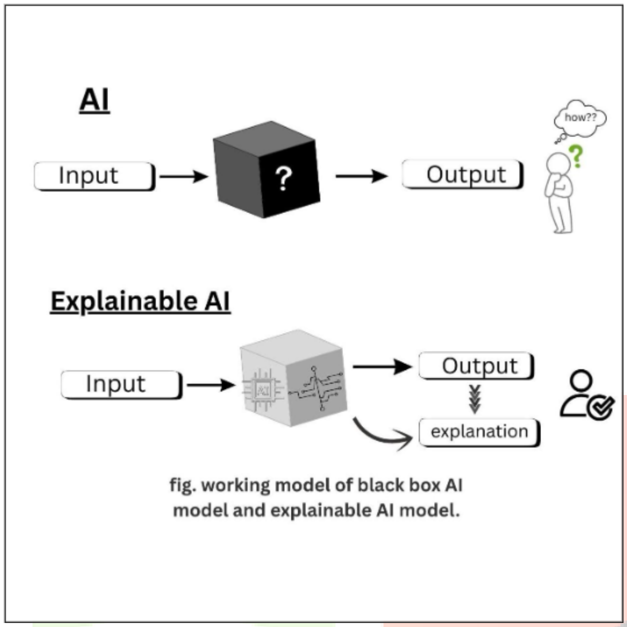

Fig. 1. Working model of black-box AI and Explainable AI (XAI)

3. LITERATURE REVIEW XAI TECHNIQUE

A lot of methods have been proposed to enhance AI interpretability:

3.1. Model-agnostic approaches:

Model-agnostic approaches, such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive Explanations), have become popular methods for interpreting advanced complex models. LIME approximates local behavior with a simpler, easy-to-understand model. SHAP assigns importance values to individual features based on game-theoretic principles.

3.2. Counterfactual explanations:

How a small change in input variables could totally change the predicted outcome, thereby offering valuable insight into causal relationships. This perspective is provided by counterfactual explanations.

3.3. Inherently interpretable models:

Inherently interpretable models are much easier to understand and explain but for the complex tasks, these models are not usually as accurate as deep learning model. Linear regression, decision trees and rule-based systems are three examples of inherently interpretable models.

Table I: Comparative Analysis of XAI Techniques

Approach	**Eg. /Methods**	**Strengths**	**Limitations**	**Suitable Domains**
Model-agnostic	LIME, SHAP	Works with any ML/DL model Provides local feature importance Easy to implement	High computational cost Explanations may vary with sampling	Healthcare, Finance, NLP
Counterfactual Explanations	Perturbation-based input changes	Shows 'what-if' scenarios Useful for fairness & causal reasoning	Difficult to generate realistic counterfactuals May not generalize to all models	Finance, Policy-making
Inherently Interpretable Models	Decision Trees, Rule-based models, Linear Regression	Simple and transparent Fast training & explanation	Not suitable for complex tasks Lower accuracy than deep models	Education, Risk scoring
Visualization-based	Heatmaps, Saliency maps, Grad-CAM	Visual insights into model's focus Good for image data	Often noisy and unstable Limited for tabular/text data	Computer Vision, Healthcare Imaging
Hybrid/Ensemble XAI	Combining interpretable + black-box models	Balances accuracy and interpretability Layered explanations	Complexity increases Harder to standardize	Autonomous Systems, Security

4. APPLICATIONS OF XAI IN DIFFERENT DOMAINS

4.1. Healthcare:

AI-assisted diagnostics, such as cancer detection, blood pressure detection, diabetes detection require explanations for clinical decisions to ensure safety and compliance and these explanations increase the patient trust.

4.2. Finance:

decisions like giving loans or assessing risk can have a big impact on people's lives. Transparent credit scoring can help to increase customer trust in banks.

4.3. Autonomous Vehicles:

for real-time monitoring of vehicle, XAI is needed. It helps to detect errors or risky patterns before accident occur, this can save human lives

4.4. Education:

Explainable models can understand which concepts a student struggles with most, helping educators focus their teaching on areas where students need attention. Which can help in improving of literacy level of country also.

4.5. Security and Defense:

XAI helps in threat detection systems by explaining anomaly detections, ensuring actionable decisions. Transparency on each decision of detection is really needed.

5. FUTURE TRENDS IN XAI

Future XAI will integrate with:

• **Generative AI:**
Understands outputs of large language and image generation models.

• **Reinforcement Learning & Multi-Agent Systems:**
Explains decisions in dynamic, adaptive environments.

• **Human AI Collaboration Tools:**
Interfaces designed to improve decision-making and trust.

• **Regulatory Compliance:**
making clear standards for increasing transparency and more clarity in the areas of banking and finance, healthcare, education and autonomous system as well. The main focus will be on real-time explanations with maintaining accuracy.

6. CHALLENGES ASSOCIATED WITH XAI

6.1. Technical Challenges:

providing real-time explanations without affecting performance is becoming difficult just because deep learning and ensemble models are highly complex but ongoing research and new technologies are making this possible day by day.

6.2. Operational Challenges:

In the existing environment, it is difficult to integrate explainable system into it because it needs significant resources, well-trained staff and adjustments to establish it.

6.3. Ethical & Legal Challenges:

Ensuring clarity, transparency as well as fairness, accountability, and compliance with regulations is essential for legal rights and human rights.

6.4. User Acceptance:

if user do not understand or trust the explanations then fully interpretable technical model will become useless for users. User Acceptance of explanation is mandatory.

7. ETHICAL, LEGAL AND SOCIAL IMPLICETION

XAI helps reduce unfairness by offering insight into how decisions are made by AI, thereby promoting fairness.

In terms of legal perspective and human rights: explanation must be provided to human to meet transparency and accountability.

8. RECOMMENDATIONS AND FUTURE

Directions:

● Develop human-centered and scalable frameworks for explainable systems and processes.

●Standardize evaluation metrics to consistently measure interpretability across different industries across the globe.

●Continuous learning is mandatory so explanations can adapt with changing environments and take right decisions.

● Mutual collaboration among researchers, ethicists, domain experts, and policymakers is important.

●Design human–system collaboration interfaces that provide actionable insights.

9. CASE STUDIES AND REAL-WORLD EXAMPLE

9.1. Healthcare:
AI systems help in cancer detection by providing visual explanations for clinicians.

9.2. Finance:
Transparent credit algorithms decisions and improving fairness is maintaining customer trust also.

9.3. Autonomous Systems:
Drones and self-driving vehicles can generate instant (real-time) solution of with explanation which making it easier to monitor and understand their decisions.

9.4. Education:
Intelligent tutoring systems can offer detailed explanations to students which can help them build a bright future.

10. CONCLUSION

By showing accountability, transparency as well as interpretability to humans, it helps for building trust, supports ethical practices, and it enables meaningful collaboration between technology and people.

Explainable AI is essential for putting AI systems into action in a way that people can trust, that are ethical, and aligned with human needs. Future research should focus on developing robust frameworks, establishing standards, and encouraging interdisciplinary approaches to ensure AI systems are socially responsible and reflect human values.

It is essential to make real-world predictions by expanding explainable AI to make a positive impact on human lives.

REFERENCES

[1] Ali, S. (2023). Explainable Artificial Intelligence (XAI): What we know and what we don't. Journal of AI Research, 56(4), 234-256.

[2] Saranya, A. (2023). A systematic review of Explainable Artificial Intelligence approaches. Computer Science Review, 47, 100529.

[3] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 22(1), 1–34.

[4] Cheng, Z., Wang, Y., & Liu, J. (2025). A Comprehensive Review of Explainable Artificial Intelligence in Computer Vision. Artificial Intelligence Review, 58, 1107–

[5] Alkhanbouli, R., & Al-Mutairi, H. (2025). The role of explainable artificial intelligence in disease prediction. Journal of Biomedical Informatics, 136, 104204.

[6] Yang, W., & Li, X. (2023). Survey on Explainable AI: From Approaches, Limitations to Applications. Journal of Big Data Analytics, 10(2), 45–72.

[7] Longo, L., & Mariani, G. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of the next generation. Information Fusion, 95, 101987.

[8] Rosenbacke, R., Schulte, C., & Greifeneder, R. (2024). How Explainable Artificial Intelligence Can Increase Clinicians' Trust in AI-Driven Clinical Decision Support Systems. JMIR AI, 5(1), e53207.

[9] Bhati, D., & Kumar, A. (2024). A Survey on Explainable Artificial Intelligence (XAI) in Medical Imaging. Journal of Imaging, 10(10), 239.