Applied Data Science and Analysis
Vol.2025, **pp**. 1–2
DOI: https://doi.org/10.58496/ADSA/2025/001; ISSN: 3005-317X
https://mesopotamian.press/journals/index.php/ADSA

Editorial Article

Explainable AI: Methods, Challenges, and Future Directions

Ahmed Hussein Ali[1,*], Marwan Ali Shnan[2]

[1] *Department of Computer, College of Education, AL-Iraqia University, Baghdad, Iraq.*

[2] *Faculty of Computer Systems and Software Engineering, University Malaysia Pahang, Kuantan, Pahang, Malaysia.*

As artificial intelligence (AI)[1] systems become increasingly complex and pervasive, the need for transparency and interpretability has become a critical concern. Explainable AI (XAI)[2, 3] seeks to bridge the gap between opaque machine learning models and human users by providing insights into the decision-making processes of AI systems. This editorial explores the various methods employed in XAI, the challenges faced in achieving interpretability, and potential future directions for the field.

The rapid adoption of AI in critical domains such as healthcare, finance, and criminal justice has raised concerns about the "black-box" nature of many AI models[4]. While these models often achieve high accuracy, their decision-making processes remain obscure, making it difficult to diagnose errors, ensure fairness, and build user trust. Explainable AI aims to address these concerns by developing techniques that offer transparency and interpretability without compromising performance.

1. METHODS OF EXPLAINABLE AI

Several techniques have been developed to improve the explainability of AI models. These methods can be broadly categorized into intrinsic and post-hoc approaches:

1. Intrinsic Explainability: Models designed with interpretability in mind, such as decision trees, linear regression, and rule-based systems, inherently provide transparent reasoning. While these models are easier to understand, they often lack the predictive power of more complex architectures[5].
2. Post-hoc Explainability: Applied to black-box models like deep neural networks and ensemble methods, post-hoc techniques include[6, 7]:
    - Feature Attribution: Methods like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) help identify the contribution of individual features to a model's predictions.
    - Visualization Techniques: Tools like saliency maps and Grad-CAM (Gradient-weighted Class Activation Mapping) provide graphical representations of how deep learning models make decisions.
    - Surrogate Models: Simplified, interpretable models (e.g., decision trees) are trained to approximate the behavior of complex AI systems.

2. CHALLENGES IN EXPLAINABILITY

Despite the progress in XAI, several challenges remain:

1. Trade-off Between Accuracy and Interpretability: Many inherently interpretable models lack the predictive power of deep learning, creating a tension between explainability and performance.
2. Scalability Issues: As AI models grow in complexity, existing XAI methods struggle to provide meaningful explanations without overwhelming users.
3. Bias and Fairness: Ensuring that explanations do not inadvertently reinforce biases or provide misleading justifications remains an open problem.
4. Human-Centered Evaluation: Effective explanations should be comprehensible to non-experts, yet there is no universal standard for evaluating the quality of explanations.

*Corresponding author. Email: msc.ahmed.h.ali@gmail.com

3. FUTURE DIRECTIONS

To advance the field of XAI, future research should focus on:

1. Hybrid Approaches: Combining intrinsic and post-hoc methods to balance interpretability and accuracy.
2. Context-Aware Explanations: Tailoring explanations to different stakeholders (e.g., doctors, policymakers, end-users) to enhance usability.
3. Regulatory and Ethical Considerations: Developing guidelines and frameworks to ensure that XAI aligns with ethical principles and regulatory standards.
4. Human-AI Collaboration: Designing interactive tools that allow users to query AI systems and refine explanations dynamically.

Explainable AI is essential for fostering trust, accountability, and fairness in AI-driven decision-making. While significant progress has been made, achieving robust and scalable explainability remains an ongoing challenge. Future efforts should focus on refining existing techniques, addressing ethical concerns, and enhancing the usability of explanations to make AI more transparent and trustworthy.

Conflicts of Interest

The author's paper explicitly states that no funding was received from any institution or sponsor.

Funding

None.

Acknowledgment

None

References

[1] R. Dwivedi *et al.*, "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1-33, 2023.

[2] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: an analytical review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 5, p. e1424, 2021.

[3] D. Saraswat *et al.*, "Explainable AI for healthcare 5.0: opportunities and challenges," *IEEe Access*, vol. 10, pp. 84486-84517, 2022.

[4] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of explainable AI techniques in healthcare," *Sensors*, vol. 23, no. 2, p. 634, 2023.

[5] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, "Explainable artificial intelligence: a comprehensive review," *Artificial Intelligence Review*, pp. 1-66, 2022.

[6] P. Gohel, P. Singh, and M. Mohanty, "Explainable AI: current status and future directions," *arXiv preprint arXiv:2107.07045*, 2021.

[7] R. Confalonieri, L. Coba, B. Wagner, and T. R. Besold, "A historical perspective of explainable artificial intelligence," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 1, p. e1391, 2021.