



# Explainable AI (XAI): Making AI Decisions more Interpretable

Nicolas Roll\*

Department of Clinical Pharmacy, King Saud University, Saudi Arabia

## INTRODUCTION

Artificial Intelligence (AI) has become a cornerstone of modern technology, driving advancements in fields ranging from healthcare to finance and beyond. However, as AI systems become increasingly complex and autonomous, understanding how these systems arrive at their decisions has become a significant challenge. This has led to the emergence of Explainable AI (XAI), a field dedicated to making AI systems' decisions more transparent, interpretable, and understandable. XAI aims to address the "black box" problem of AI, where the internal workings of models are not easily understood, thereby enhancing trust, accountability, and usability.

## DESCRIPTION

As AI systems are integrated into critical applications such as medical diagnosis, autonomous vehicles, and financial trading the consequences of their decisions can be profound. Users may be hesitant to rely on AI systems if they do not understand how decisions are made. In cases of erroneous or harmful outcomes, it is crucial to identify the cause and hold responsible parties accountable. Increasingly, regulations require that AI systems provide explanations for their decisions, especially in sensitive areas like credit scoring and healthcare. Several approaches are employed to make AI decisions more interpretable. These methods can be broadly categorized into model-specific and model-agnostic techniques. Model-specific techniques are designed for particular types of AI models, often focusing on enhancing the interpretability of complex models. Some AI models are inherently more interpretable than others. For example, linear regression and decision trees offer straightforward explanations of how input features contribute to predictions. These models are often used in scenarios where interpretability is paramount, even if they are less powerful than complex models like deep neural networks. Techniques such as feature importance scoring identify which features have the most influence on the model's predictions. Methods like permutation importance or SHAP

(SHapley Additive exPlanations) values provide insights into how individual features contribute to the final decision. For models like neural networks, visualization techniques can help interpret the internal workings. For instance, activation maps in convolutional neural networks (CNNs) highlight which parts of an image contribute to specific features detected by the model. Model-agnostic techniques apply to any AI model, providing ways to interpret predictions regardless of the underlying algorithm. There is often a trade-off between model complexity and interpretability. Highly complex models like deep neural networks provide greater predictive power but are harder to explain. Balancing accuracy and interpretability remains a challenge. Ensuring that explanations are both accurate and consistent with the model's behavior is crucial. Inaccurate explanations can undermine trust and lead to misguided conclusions. Explanations need to be tailored to different user needs and levels of expertise. For instance, domain experts might require more detailed technical explanations, while non-experts might benefit from simpler, intuitive summaries.

## CONCLUSION

Explainable AI (XAI) represents a critical advancement in the development and deployment of AI systems. By making AI decisions more interpretable, XAI enhances user trust, accountability, and compliance with regulatory requirements. The field encompasses a range of methods, from inherently interpretable models to sophisticated model-agnostic techniques, each contributing to a clearer understanding of AI decision-making processes. As AI continues to evolve and integrate into various aspects of daily life, the need for transparency and interpretability will only grow. Future advancements in XAI will likely focus on improving the trade-off between complexity and interpretability, enhancing the accuracy and consistency of explanations, and addressing regulatory and ethical challenges. By advancing these areas, XAI will play a pivotal role in ensuring that AI systems are not only powerful but also understandable and trustworthy.

<b>Received:</b>	01-July-2024	<b>Manuscript No:</b>	ipbjr-24-21183
<b>Editor assigned:</b>	03-July-2024	<b>PreQC No:</b>	ipbjr-24-21183 (PQ)
<b>Reviewed:</b>	17-July-2024	<b>QC No:</b>	ipbjr-24-21183
<b>Revised:</b>	22-July-2024	<b>Manuscript No:</b>	ipbjr-24-21183 (R)
<b>Published:</b>	29-July-2024	<b>DOI:</b>	10.35841/2394-3718-11.7.65

**Corresponding author** Nicolas Roll, Department of Clinical Pharmacy, King Saud University, Saudi Arabia, E-mail: n\_77@hotmail.com

**Citation** Roll N (2024) Explainable AI (XAI): Making AI Decisions more Interpretable. Br J Res. 11:65.

**Copyright** © 2024 Roll N. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.