

Explainable AI: Improving Transparency in AI Decision-Making

Aditya Swapan Halder, Shruti Ramesh Sunadalli, Aman Iqbal Shaikh, Reeshi Nitin Kanade
Guru Gobind Singh Polytechnic Nashik, Maharashtra, India

Abstract: *In recent years, the rapid adoption of Artificial Intelligence (AI) through numerous domain names has raised essential issues concerning the transparency, trustworthiness, and interpretability of AI decisionmaking structures. Traditional AI fashions, specifically black container models such as deep neural networks, frequently lack readability in explaining their outputs, which poses challenges in responsibility, equity, and moral compliance. Explainable AI (XAI) targets to bridge this gap through imparting interpretable models and post hoc explanation techniques that make clear AI-pushed selections for both technical experts and cease-users. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive Explanations (SHAP), and characteristic importance analysis have emerged as important equipment to enhance transparency. XAI is especially vital in high-stakes sectors like healthcare, finance, and self-sufficient automobiles, in which understanding AI reasoning can impact lives and regulatory compliance. This paper explores the basics of XAI, its key strategies, actual-world packages, and the ongoing challenges of balancing accuracy and interpretability.*

Keywords: Explainable AI, Transparency, Interpretability, Trustworthiness, Black-Box Models, Ethical AI, Accountability, LIME, SHAP, Fairness