Abstract:

The abstract emphasizes the need for Explainable AI (XAI) due to the widespread influence of AI systems in critical domains, particularly decision-making and individual rights. It calls for a multidisciplinary approach to address these challenges, integrating technical, philosophical, social, and psychological dimensions. The complexity of articulating AI reasoning to non-experts is highlighted, necessitating foundational models and a comprehensive theoretical framework. Translating complex AI-generated explanations into accessible language is crucial for bridging the gap between AI reasoning and human comprehension. Surrogate models, like expert systems, could enhance understanding of complex AI reasoning among broader audiences. The abstract calls for a robust theoretical foundation encompassing diverse disciplines to effectively tackle Explainable AI's complexities in contemporary socio-technical landscapes.

Introduction :

The introduction discusses the role of Artificial Intelligence (AI) in decision-making processes, focusing on its impact on individual rights and societal well-being. It distinguishes between automated decisions and genuine AI systems, highlighting the misuse of the term "AI" and its misconceptions. The complexity of Explainable AI (XAI) is also discussed, highlighting the challenges of making complex AI reasoning accessible to those unfamiliar with technicalities. The introduction categorizes complex AI mechanisms into different levels of explainability, highlighting the difficulties in articulating AI decision-making in societal domains and public opinion. Transparency is emphasized, as it helps maintain public confidence in AI systems' performance and validity. The introduction calls for a balanced approach that integrates technical understanding with human comprehension in complex socio-technical landscapes.