

XAI in Healthcare

Sanjog Ray

Associate Professor, Information Systems
Indian Institute of Management Indore
India
sanjogr@iimidr.ac.in

Abstract

Artificial Intelligence systems are experiencing widespread deployment across numerous domains. These systems are frequently characterized as "black box" models, employing complex algorithms that render their decision-making processes challenging for human comprehension. This opacity has led to a significant deficit of trust in AI systems, despite their widespread deployment. Consequently, there is a growing imperative for AI systems to elucidate their reasoning and decision outcomes, giving rise to the field of eXplainable Artificial Intelligence (XAI). XAI systems are specifically designed to furnish explanations for their decisions or predictions, thereby offering valuable insights into their operational behaviours. In healthcare sector, because of the importance the of trust in the systems, the explanatory capabilities of AI systems are paramount for their successful adoption. This study undertakes an analysis of XAI system applications within healthcare, exploring the existing research on their deployment in this critical domain. Our investigation reveals that extant literature has focused on AI features such as recommendations, interpretation, analysis, prediction, process evaluation, and content generation within the medical domain. While Large Language Models have been proposed for various clinical applications, including text classification, patient support, and diagnostic assistance, the emphasis on explainability within these AI applications has been notably limited, despite its potential for substantial domain benefits. Furthermore, we identify unresolved research problems within the current literature and propose novel approaches for their investigation.

Keywords

XAI, Explanations, Healthcare, LLMs and Recommendations.