IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

MAIA – A Multimodal Automated Interpretability Agent for Explainable AI

Vedant Jawalekar[1], Vijay Hatte[2], Naman Shetty[3], Nikita Khawase[4], Anil Walke[5]
Students, Artificial Intelligence & Data Science[1,2,3]
Head of Department - Artificial Intelligence & Data Science[4]
Assistant Professor - Artificial Intelligence & Data Science[5]
ISBM College of Engineering, Pune, India

Abstract: *Explainable Artificial Intelligence (XAI) has emerged as a critical field to demystify the decision-making process of complex deep learning models. This paper introduces MAIA – a Multimodal Automated Interpretability Agent – developed as a web-based platform to enhance interpretability, fairness analysis, and transparency of AI models. MAIA offers an integrated set of modules for neuron visualization, bias detection, feature attribution, and natural language summarization. Designed for educators, researchers, and developers, MAIA leverages advanced techniques such as Grad-CAM, Integrated Gradients, and Pegasus-XSum to interpret and present AI decisions in an understandable way. This paper details the architecture, implementation, and real-world use cases demonstrating MAIA's capabilities as a complete XAI toolkit.*

Keywords: Explainable AI, Model Interpretability, Grad-CAM, Pegasus-XSum, Bias Detection

DOI: 10.48175/IJARSCT-27520