

# EXPLAINABLE AI

BY : YASMEEN AL-SOOFI  
2024-01-22

# Table of Contents

01

HOW IS THE RESEARCH  
BEING DONE?

02

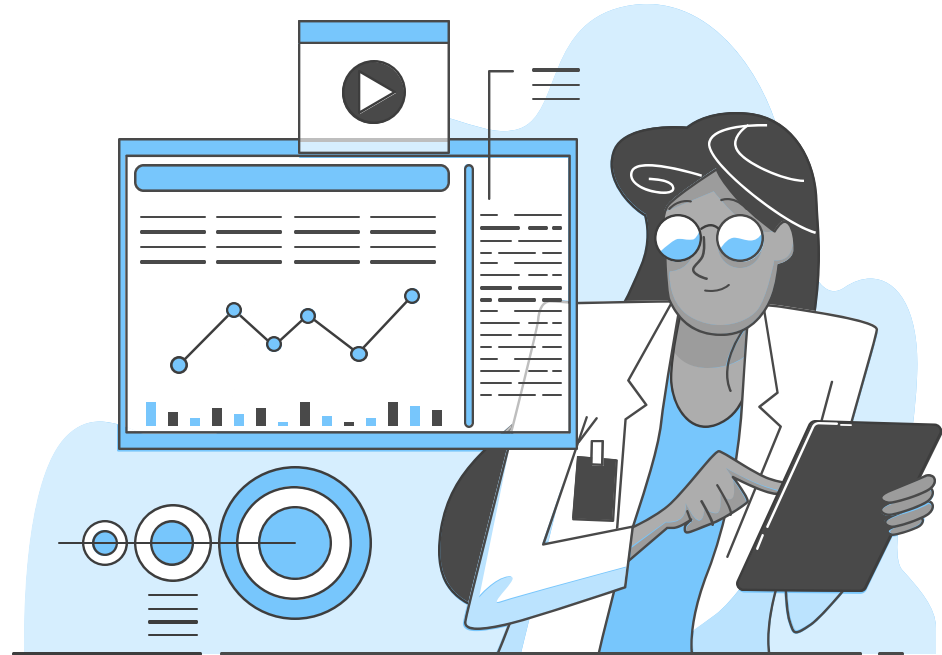
WHERE / WHO IS DOING  
THE RESEARCH

03

WHAT THEY HAVE  
DISCOVERED

04

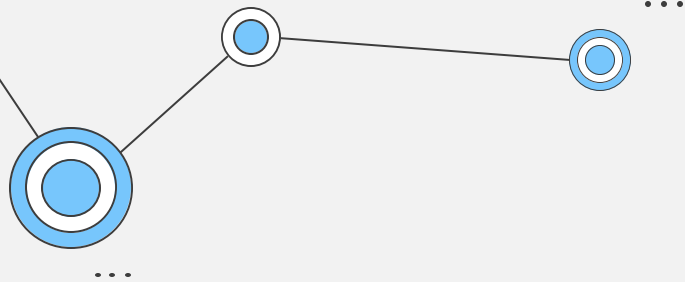
WHY THIS RESEARCH MAY  
BE BENEFICAL TO  
COMPUTER SCIENCE AND  
OTHER FIELDS



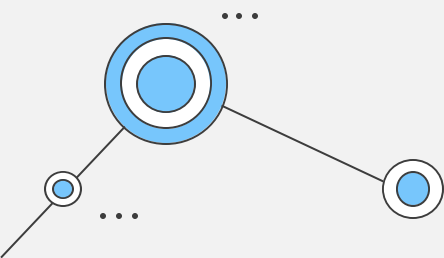
# How Is The Research Being Done?



The field of explainable AI (XAI) aims to shed light on how artificial intelligence (AI) systems reach to their decisions. Research on explainable AI (XAI) focuses on different approaches to improve the comprehension of ... artificial intelligence systems. The creation of approachable tools and methods for interpreting intricate AI models is a crucial strategy. Scientists are attempting to develop algorithms that produce simple justifications for AI judgments. Additionally, they are looking into ways to increase transparency in the AI models' decision-making process so that consumers can identify the important variables affecting results. Researchers at XAI are investigating methods to include human feedback into AI systems so that users may communicate, ask questions, and get clear answers in addition to improving algorithms. Working together, specialists, psychologists, and computer scientists hope to make AI more intuitive and instructive overall in human-computer interaction. Researchers are also using user understanding and alignment with human reasoning tests to assess how well ... different explanation techniques work.



# Where / Who is doing The Research?



Leading companies in the field, including OpenAI, Google, Microsoft, and IBM, are working together to develop the field of explainable AI (XAI) and improve AI interpretability. Through specialized research units, prestigious universities like Carnegie Mellon, Stanford, and MIT have a major influence on the direction of XAI. Cutting-edge insights are provided by research laboratories like Facebook's AI Lab and Google's Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) division. Research integrating computer science, psychology, and human-computer interaction is fostered by collaborative endeavors. Guidelines and standards for responsible AI are offered by groups such as the IEEE Standards Association and the Partnership on AI. Individual academics from all over the world offer their expertise to this dynamic field, which together advances our understanding of Explainable AI.

# What Have They Discovered

Researchers have discovered methods like LIME and SHAP in Explainable AI (XAI) that can make complex AI models easier to grasp. Additionally, they have developed user-friendly interfaces that make interacting with AI systems simple. The integration of human feedback, enhancing transparency, and creating defined measures for efficient explanations are the main areas of ongoing research. The implications of XAI go beyond computer science to include finance, healthcare, and ethics. Prioritizing ethical issues ensures accountability and justice in AI results. The ongoing goal is to improve these findings and preserve ethical standards while increasing the transparency and applicability of AI.



# Why This Research May Be Beneficial To Computer Science And Other Fields ?



Research on explainable AI (XAI) has great potential to advance computer science as well as a few other areas. In computer science, XAI improves the interpretability and transparency of AI models, promoting confidence and comprehension in their decision-making procedures. Beyond computer science, XAI has a lot to offer in a lot of other areas. Clear explanations of AI diagnoses, for example, can boost confidence among medical practitioners and facilitate more assured decision-making. Transparent AI models have the potential to improve risk assessments and market trends in the finance industry. Furthermore, explainable AI is crucial to ensure just and responsible decision-making in domains like ethics and law.

The advantages of XAI extend to fields like geography, where interpretable AI models facilitate the understanding of intricate spatial patterns needed for tasks like environmental monitoring and urban planning. All things considered, XAI's benefits cut beyond academic fields, increasing the reliability, usability, and accessibility of AI systems in a variety of real-world contexts.