Chapter 6
Transparency and Accountability

Princy Pappachan
https://orcid.org/0000-0001-6728-0228
Department of Foreign Languages and Literature, Asia University, Taiwan

Massoud Moslehpour
https://orcid.org/0000-0001-8808-2407
Department of Business Administration, Asia University, Taiwan & Department of Management, California State University, San Bernardino, USA

Ritika Bansal
Insights2Techinfo, India

Mosiur Rahaman
International Center for AI and Cyber Security Research and Innovations, Asia University, Taiwan

ABSTRACT

The rapid growth and application of AI has ushered in ground-breaking technologies like LLMs. However, these innovations also bring significant challenges related to transparency and accountability, especially considering the complex neural network architectures and vast training datasets. This chapter thus explores the journey of AI from rule-based systems to the current ML and deep neural network, identifying the black box problem that plagues the decision-making process in LLMs. The chapter introduces strategies for enhancing transparency using explainable AI (XAI) frameworks to address these issues, offering practical solutions to quantify and improve transparency. Accountability is also emphasized through a detailed protocol for assigning responsibility across AI development phases, reinforced by ethical auditing and reporting methodologies. Mathematical equations and frameworks are also

DOI: 10.4018/979-8-3693-3860-5.ch006

presented to compute transparency scores and accountability measures, providing organizations with structured, actionable guidelines for building transparent, fair, and ethical AI systems.

INTRODUCTION

The evolution of Artificial Intelligence (AI) since the 1950s has been marked by various key innovations and shifts in research focus. Early AI pioneers like Alan Turing and McCarthy explored symbolic reasoning and rule-based logic for developing intelligent machines that could simulate every aspect of human intelligence. This laid the groundwork for expert systems, where knowledge was explicitly encoded as "if-then" rules. These systems utilized causal graphs and rule-based reasoning to perform well in well-defined domains like diagnosis and planning (Shao et al., 2022). Following the advent of abundant data and advances in computing technology in the 1990s, a paradigm shift occurred as machine algorithms emerged. These algorithms could generate models and rules from data sets driven by statistical methods and user-defined objectives. This approach, called Machine Learning (ML), allowed systems to automatically extract knowledge from data, removing the painstaking process of manually encoding rules. Accordingly, this AI technique used mathematical algorithms to generate predictive models using patterns derived from known data.

Later, this process was extended to process a wider range of raw data, such as images and text, into machine-understandable features using neural networks with multiple layers, known as deep neural networks (DNNs). These allow systems to recognize visual and textual data patterns through hierarchical feature extraction by training themselves on incoming data (Galván & Mooney, 2021). This approach led to significant breakthroughs in perception tasks, allowing AI systems to identify patterns in complex data like images and natural language (Iba & Nasimul, 2020). Extensive datasets further powered these advances, leading to the development of large language models (LLMs), which could map long-term contextual relationships and generalize across diverse linguistics and natural language processing (NLP) tasks. The capabilities of LLMs bring the AI field closer to achieving general intelligence. Figure 1 illustrates the journey of AI to the present, while Figure 2 depicts the different AI systems over the years.

Related Content