Chapter 5
Ethical Frameworks for Use in Artificial Intelligence Systems

Veena Christy
https://orcid.org/0000-0001-9987-6253
SRM Institute of Science and Technology, India

Vijaya Kittu Manda
https://orcid.org/0000-0002-1680-8210
PBMEIT, India

Gnanadasan M. L.
SRM Institute of Science and Technology, India

ABSTRACT

The chapter examines the ethical landscape of AI, focusing on the development and deployment of Responsible AI systems. The increased deployment of systems powered by artificial intelligence technologies necessitates making work more predictable, trustworthy, and ethical. The development and deployment of such systems require the use of Responsible AI. A healthcare system, for example, is a classic system where AI is used in healthcare collaborations and decision-making. Studying the ethical landscape of AI allows us to understand the historical context and evolution of AI Ethics and the challenges and risks associated with such systems. Ethical frameworks examines the six AI assurance goals, the existing ethical frameworks and guidelines, and the core principles behind them. Future challenges and considerations for further research that emerge because of evolving concerns and potential misuse of AI systems are discussed.

DOI: 10.4018/979-8-3693-8557-9.ch005

INTRODUCTION

Importance of Ethical Considerations in AI

The rise and increased use of Artificial Intelligence (AI) technologies makes it increasingly important to consider the Ethical aspects. These aspects are best addressed during the system design, development, and implementation stages. If AI is the technology for tomorrow, AI Ethics should be considered the need of the hour. AI technologies will undoubtedly impact people's lives, both at local and global levels. AI technology is becoming increasingly used regularly both in homes and workplaces. AI is now integrated into the software of televisions, cars, and mobile phones (Sreedhar Reddy, 2022). The omnipresence of technology will mean many new issues, such as secrecy or discrimination, will evolve. At the workplace, Generative AI is seen automating tasks. Data analysis and report writing is a significant application that frees managers to focus on more strategic initiatives.

Developers and technology implementers will have to take responsibility for their deployments. For example, concerns over personal data breaches would be raised if facial recognition software shows racial bias (Buolamwini & Gebru, 2018). Moreover, some say this could further widen financial inequality through automated decision-making systems within banking. In contrast, others argue that it likewise applies within employment settings, too, since both areas already struggle greatly with diversity (Mehrabi et al., 2019). According to the IEEE (2019), the organization provides new ethical frameworks and guidelines for future technologies. It also offers principles for the responsible development and application of AI. Other examples of these measures include the European Union's General Data Protection Regulation (GDPR). It protects individuals' privacy rights (European Union, 2016).

Disregarding ethical concerns might slow technological progress. Getting permits and permissions takes much time. However, if products and services are built without permission, public trust in the product/company is reduced. Furthermore, it may even put human lives at risk. Therefore, ethical principles must be incorporated throughout all phases of artificial intelligence (AI) development to develop technology that supports fair results and complies with societal standards. Model evaluations can assess risk, and reinforcement learning from human feedback (RLHF) should be used to fine-tune the models to mitigate the risks (Schuett et al., 2024).

In reinforcement learning, AI models are trained using feedback from human experts to fine-tune AI results. An alternative constitutional AI approach exists. In this, the AI system is trained using rules or guiding principles that act as a "constitution" for the AI. Social Choice Theory (SCT) can be used to align AI to establish, through a collective decision-making process, the goals pursued by AI systems, the