International Journal of Computer Science and Engineering (IJCSE)
ISSN (P): 2278–9960; ISSN (E): 2278–9979
Vol. 13, Issue 2, Jul–Dec 2024; 73–88
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

NEURO-SYMBOLIC AI FOR AUTONOMOUS ETHICAL DECISION-MAKING IN BRIDGING HUMAN VALUES WITH MACHINE INTELLIGENCE

Balasubramani Murugesan

Director of Engineering, Digit7, United States of America

ABSTRACT

This research explores the potential of neuro-symbolic AI in bridging the gap between human values and the machine intelligence. Neuro-symbolic AI combines the neural networks with symbolic reasoning to deliver an ethical reasoning capability in artificial intelligence systems. Through the integration of the neural networks and symbolic reasoning, neuro-symbolic AI seeks to enhance the learning, reasoning and decision-making capabilities of the AI system. This study examines the theoretical underpinnings, applications, and ethical considerations encompasses by neuro-symbolic AI in various domains such as academia, healthcare and business. Through the utilization of the qualitative research methodology, the study analyzes the existing literature to address the main themes and potential challenges. The findings underline the encouraging the potential of neuro-symbolic AI in the identification of the limitations of traditional AI approaches and promote human-centered AI development. The findings from the research underlines that neuro-symbolic AI can contribute to the explainability, transparency, and fairness of AI. However, there are still open challenges on scalability, interpretability, and cross-cultural adaptability. Future research should be devoted to overcome the limitations above and foster further interdisciplinary collaboration on ethical development and deployment associated to neuro-symbolic AI systems. This paper focuses on the exploration of the innovative methods to integrate fairness into AI models, leveraging Explainable AI (XAI) tools such as SHAP to mitigate biases during training. The framework autonomously maintains fairness through diversified data handling, with applications in income prediction, credit risk, and recidivism. This paper analyzes the neuro-symbolic AI approaches for ethical reasoning, highlighting their integration of deep learning and symbolic logic to enhance transparency, explainability, and decision-making. Applications in healthcare, finance, and education are explored, with a focus on aligning AI with human values, ensuring ethical, trustworthy outcomes across diverse domains.

KEYWORDS: *Neuro-Symbolic AI, Ethical Decision-Making, Human Values Integration, Artificial Neural Networks (ANN), Machine Intelligence, Symbolic Reasoning, Explainability in AI.*

Article History

Received: *22 Nov 2024* **|** ***Revised:*** *26 Nov 2024* **|** ***Accepted:*** *12 Dec 2024*