International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0254

e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1571 - 1576

Adaptive Neuro-Symbolic Systems for Real Time Ethical Decision-Making in Autonomous Agents

Shadrach C Matthew[1], *Sanjay Siddharthan R*[2], *Elavarasan R*[3]

[1,2]*UG- Information Technology, St. Joseph's College of Engineering, Chennai, TamilNadu, India.*

[3]*Assistant Professor, Information Technology, St. Joseph's College of Engineering, Chennai, TamilNadu, India.*

Email ID: *sshadrach2003@gmail.com*[1], *rsanjaysiddharthan@gmail.com*[2], *elavarasan1304@gmail.com*[3]

Abstract

With the rapid emergence of autonomous systems, appropriate robust frameworks that could make ethical decisions in real time are needed. The adaptive neuro-symbolic approach to decision making by autonomous agents is thus presented here, integrating the advantages of symbolic ability like conventional AI with the adaptability imparted through neural networks. This proposed system enables symbolic reasoning by the AI along with learning from data, thus ensuring transparency and adaptability in decisions. This system, with deep learning models integrated with symbolic representations, would have an ability to make decisions within complex ethical dilemmas that bring adaptations to dynamic environments with decisions in accordance with ethical principles. Simulations across diverse, real-world scenarios demonstrate the potential of the system in autonomous vehicles, robotics, and other critical decision-making applications. This paper investigates the synergy between symbolic reasoning and neural network learning, aiming to bridge the gap between these paradigms. The hybrid approach combines interpretability and generalization strengths of symbolic AI with the learning strengths of neural networks, thereby overcoming the limitation imposed by the exclusive usage of either approach.

Keywords: *Autonomous agent, neural networks, symbolic AI, symbolic reasoning.*

1. Introduction

Decision making systems continue to evolve as it is capable of taking decision on its own and lessens the dependency on humans. Both for crucial and mundane tasks these systems are deployed. Since there is less human intervention, it is greatly helping in autonomous agents or self-aware agents which are deployed across various fields and one such is self-driving cars. The decision that is taken must be appropriate, ethical and compliant with the standards and this is the primary goal. The decision making module in such agents brings a significant impact to the entire system. Over the time various algorithms and techniques have been developed to improve the decision taken by these autonomous agents, and therefore it highlights the need for a ideal decision making module Neuro-Symbolic system harnesses the power of both Symbolic AI and Neural AI Systems. Symbolic AI excels when it comes to explicit reasoning and symbolic manipulation and thus providing transparency and interpretability. On the other hand, Neural AI systems learns from data, adapt based on behavior and capable of recognizing patterns. For an autonomous agent the decision taken so far are either by rule-based systems or by neural systems. When using the either one of the above methods, the autonomous agent will have its limitation in the decision that it takes hence integrating the two systems is much needed. Any autonomous agent is designed in such a way that is able to perceive the environment and process what decision must likely be taken based on the input. This decision which the agent takes it must be logically reasonable as well as ethically correct and to achieve this we need both of Neural and Symbolic system which this paper emphasizes. In traditional system which uses Symbolic reasoning = for decision making, it is capable of performing good when it comes to logical structure, rule-based inference and explicit knowledge representation. However, due to its rigidity it cannot decide during unpredictable real-world scenarios as it relies on pre-defined rules, logic and explicit representation of knowledge. And these systems also struggle to learn from new data or be able to handle ambiguous and noisy input. Autonomous agents with neural AI technology face the problem of "black box problem", which is lack of interpretability. The input is passed on to the complex layers of the neural networks and produces an output. The produced output is difficult to understand and explain why the agent made such a particular decision. Thus, this lack of transparency problem poses a major problem in autonomous agents which takes ethically sensitive decisions. Without the integration of neural and symbolic learning, autonomous agent can exist and it is existing, but such ones are designed only for a particular use case hence to be a generalized and an ideal one it is essential for this integration mainly in agents which takes real-time ethical decisions. However, this integration has procedures, the integration methods and challenges associated with it. The use case of this system focuses on autonomous vehicles and robotics, where real-time responses are crucial.

2. Literature Survey

Neural-symbolic computing has been an active area of research for many years [7] which seeks to bring together robust learning in neural networks with reasoning and explainability via symbolic representations for network models. Examining the literature is crucial to understand how far this technology has come and the possibilities to implement for this use case. [1] Outlines a framework that enhances deep learning models by incorporating symbolic AI for structural constraints and domain knowledge, improving reasoning capabilities. Through experiments, the system demonstrates enhanced performance in tasks requiring complex reasoning and generalization, while reducing data dependency and increasing interpretability. This approach holds significant potential for developing AI systems with more advanced cognitive abilities, resembling human-like understanding and reasoning. [2] discusses the integration of explicit knowledge with data-driven deep learning to improve AI systems. While deep learning excels in learning data-dependent relationships, combining it with knowledge—represented by facts that are universally true—enhances performance, reduces training time, and increases user-level explainability. This approach, termed knowledge-infused learning (KiL), introduces symbolic AI into deep learning, leading to neuro-symbolic AI methods. The paper highlights the development of context-adaptive algorithms for knowledge infusion, categorized into shallow, semi-deep, and deep infusion [3] Emphasizes the growing influence of Artificial Intelligence (AI) and deep learning, while also addressing concerns about the interpretability and accountability of AI systems. It highlights the need for integrating principled knowledge representation and reasoning mechanisms with deep learning to create more explainable and accountable models. Neural-symbolic computing is presented as a solution, combining the learning capabilities of neural networks with the reasoning and interpretability of symbolic AI. The paper surveys recent progress in this area, illustrating how neural-symbolic integration can enhance the creation of explainable AI systems, addressing the demand for more transparent AI technologies. In [4], Pearl proposes a hierarchy consisting of three levels: Association, Intervention and Counterfactual reasoning, and claims that ML is only capable of achieving association (association purely involves statistical relationships. A neuro-symbolic or purely symbolic ML system should be capable of satisfying the requirements of all three of Pearl's levels, e.g. by mapping the neural networks onto symbolic descriptions. Once a symbolic description of the form if A then B has been associated with a neural network, then it is sure that idea of intervention and counterfactual reasoning become possible. [5] Henry Kautz provided a taxonomy for neuro-symbolic AI. Six types of integrating neural network and symbolic systems were proposed. The TYPE 6 system is the highly relevant to the neuro-symbolic computing, in this type a symbolic reasoning is inside of a neural engine unlike the other type which uses symbolic knowledge or neural systems as a part of the process that is taking place inside. It is capable of combinatorial reasoning, possibly by using an attention schema to achieve. [6] A learning architecture that extrapolates to harder symbolic math reasoning problem is introduced. A memory augmented recursive neural networks to address the generalization performance loss on deeper data points is presented. Tree LSTMs is employed to incorporate the structure of symbolic expression. Symbolic expression here used to define relationships between the black-box functions.

3. Current System

The current systems used for decision-making in autonomous agents typically employ a combination of traditional AI techniques, such as rule-based systems, machine learning models (particularly deep learning), and hybrid approaches. These systems are designed to help agents perform tasks like perception, reasoning, planning, and control. Here's an overview of the current state of decision-making frameworks in autonomous agents:

3.1. Rule-Based Systems

- **Overview:** Rule-based systems rely on pre-defined rules and logical structures to make decisions. These rules are often encoded manually by human experts and are based on domain knowledge, ethical guidelines, or operational procedures.
- **Application:** These systems are commonly used in safety-critical application fields, such as autonomous vehicles' emergency braking systems or robotics with specific tasks that require strict adherence to safety or ethical protocols.
- **Limitations:** While highly interpretable, rule-based systems are inflexible and unable to adapt to new, unforeseen situations. They also require exhaustive rules for complex environments of the deployment, which can be impractical to define comprehensively.

3.2. Machine and Deep Learning Models

- **Overview:** Machine learning models, especially deep learning, are widely used in autonomous agents for tasks like object detection, classification, and control. These models learn from large datasets and are able to make decisions based on patterns observed in the data.
- **Application:** In autonomous vehicles, deep neural networks are used for tasks like detecting pedestrians, vehicles, and traffic signs in real-time, enabling the car to make driving decisions. Similarly, reinforcement learning (RL) is used for decision-making in dynamic environments, where agents learn optimal strategies based on trial and error.
- **Limitations:** While these systems can handle complex, real-world data and adapt to changing environments, they face issues such as lack of interpretability, susceptibility to adversarial attacks, and biases in the training data. They also require large volumes of data and can struggle to handle rare or unseen scenarios.

3.3. Reinforcement Learning (RL)

- **Overview:** RL is a class of machine learning algorithms where agents learn by interacting with their environment and receiving feedback in the form of rewards or penalties. Actions are based on the past feedback received. This method is highly suited for decision-making in dynamic and uncertain environments.
- **Application:** RL is particularly useful for tasks like robotic navigation, game playing, decision-making in environments where actions affect future states and where less human intervention would be needed. For example, autonomous robots use RL to learn how to navigate obstacles or execute tasks through trial and error.
- **Limitations:** RL requires a substantial computational resources and time to train the agents, as the learning process involves exploring a vast range of possible actions. It also struggles with safety in decision-making, especially in real-world applications where errors can lead to catastrophic outcomes.

4. Proposed System

The proposed system integrates symbolic AI with neural networks to develop an adaptive, transparent framework for real-time ethical decision-making in autonomous agents. This is a hybrid approach that brings in interpretability and logical structure from symbolic reasoning with the versatility of neural networks. The symbolic component underpins the encoding of the ethics principles, rules, and domain-specific knowledge such that decisions taken by the agent are in tandem with human values and comply with regulatory standards. It uses a neural network component that learns about dynamic inputs or adapts to situations not envisioned by the system by having processes process unstructured data from sensors and their environment.The integration allows an agent to settle decisions both in accordance with predefined ethical rules and through real-time environmental learning, thus providing a strong solution for practical challenges. The idea presented aims at overcoming the limitations that come forth from traditional frameworks in making decisions. While rule-based systems are too rigid, and neural networks too opaque, this hybrid model provides flexibility and transparency. Autonomous agents may be operated effectively within complex, changing environments, such that decisions will be both ethical and context-sensitive. It is especially useful for applications such as autonomous vehicles, healthcare robotics, and industrial automation, for which high stakes permeate every decision made, as ethics and safety must be paramount. It combines symbolic reasoning and neural learning to create a richer, more reliable, and more ethically better-founded framework for decision-making in autonomous agents.

5. Implementation

5.1. Symbol Manipulation

Symbolic AI is based on representing knowledge and reasoning through symbols, logic, and well-structured rules. The representations of concepts are explicit; every symbol is a real-world entity or concept and decisions are drawn by logical manipulation of those symbols. They are very transparent because their decision-making processes can clearly be interpreted. They are well suited to well-defined knowledge-intensive tasks such as solving rule-based problems and translating languages. They are weak, however, in dynamic environments where situations cannot be comprehensively captured by rules-only techniques and cannot be easily adapted to new or unforeseen situations. The system applies logical rules to manipulate these symbols, drawing inferences or making decisions based on their relationships. This process allows the system to simulate reasoning and problem-solving by following structured steps. Through symbol manipulation, AI can derive new knowledge or conclusions from a given set of rules and facts, providing transparency and clarity in its decision-making.

5.2. Neural Networks

Neural networks provide adaptability and learning ability for the proposed hybrid system needed to operate agents autonomously in non-deterministic dynamic environments. Though symbolic AI systems possess a more structured rule-based reasoning, neural networks let the system process huge unstructured data amounts, like sensor inputs, images, and environmental signals vital to be dealt with in decision-making real-time tasks. In this project, neural networks enable the autonomous agent to view patterns, learn new situations, and experience that allows it to make decisions depending on how it has interacted with the environment. This capability is quite crucial in applications like detecting an object, motion planning, or navigation, in which the agent continuously learns about the surroundings. (Figure 1)

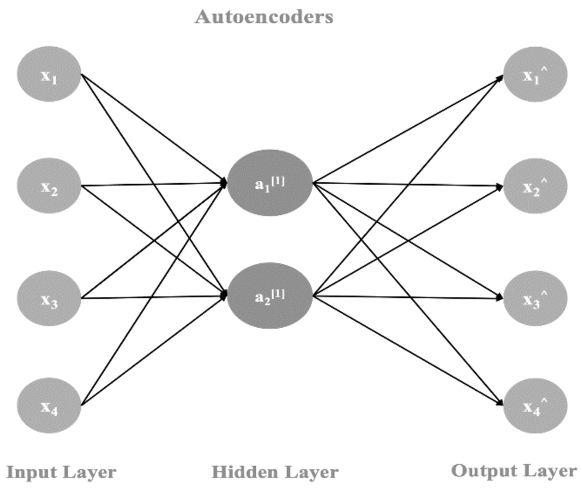

Figure 1 **Neural Network Representation**

This also improves the flexibility of systems since agents can now work with complex, noisy, or even ambiguous inputs that cannot be processed by symbolic systems alone. Such networks also allow generalization from learned experiences and make it possible to handle new or unseen scenarios like rare events or edge cases. The integration of neural networks and symbolic AI makes it able to not only respond to the environment present today, but learn and improve in its decision-making process over time. The integration between learning and reasoning means the autonomous agent will be better-positioned to make context-aware, ethically grounded decisions while it improves its performance and conditions keep changing; accordingly, such applications fit into areas like autonomous driving and robotics.

5.3. Integration Module

This integration module in the project is actually the critical component module that will integrate symbolic AI and neural network systems into a decision-making framework for autonomous agents. Integration is done in such a way it provides smooth communication between the two systems, wherein the structured rule-based reasoning would be from the symbolic part, while the neural network part deals with unstructured data and learning. It manages information flow between these systems in such a way that learned patterns or predictions from the neural network are interpreted in light of the previously encoded ethical rules and domain knowledge in the symbolic system. The integration module, as such, is a kind mediation between the symbolic term in which the neural network is output to be processed logically and vice versa, thus enabling the two components to complement each one and every other in real-time decision. The integration module addresses one of the major problems encountered in getting the two systems to harmoniously work, especially on issues involving contradictory inputs and decision. The module's hybrid design is with respect to choosing ethical reasoning at a spot which is necessary, that ensures in complex novel situations, even if the neural network is returning an unexpected output, the symbolic system may be able to enforce the ethics guidelines and constraints. The hybrid structure of this one allows the system to move flexibly and adaptively like a neural network but keep the transparency and ethical rigor from its symbolic reasoning. Thus, integration modules are one of the important features of making an autonomous agent react smartly as well as perform decisions compatibly with ethical standards. This would provide a secure and reliable framework for critical applications such as autonomous vehicles and robotics. (Figure 2)

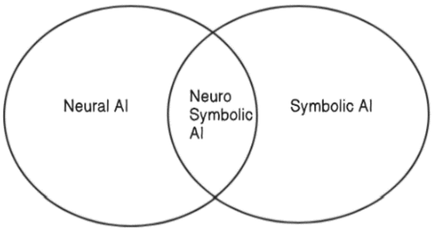

Figure 2 **Integrating Module**

6. Results and Conclusion

In conclusion, this research demonstrates the potential of integrating symbolic AI with neural networks to create a robust, adaptive, and ethically grounded decision-making system for autonomous agents. By combining the interpretability and rule-based reasoning of symbolic AI with the learning capability and flexibility of neural networks, the proposed system overcomes the limitations of traditional approaches. It ensures that autonomous agents can make real-time decisions that are both ethically sound and adaptable to dynamic environments. The integration of these two paradigms offers a promising solution to the challenges of transparency, adaptability, and ethical reasoning in autonomous systems. (Table 1)

Table 1 **AI Models**

AI Models	Accuracy	Reasoning	Transparency
Neuro-Symbolic Ai	90	High	High
Neural Networks	84	High	Low
Symbolic Ai	80	Low	High

This hybrid approach is particularly suited for high-stakes applications, such as autonomous vehicles and robotics, where ethical decision-making and safety are paramount. Future work will focus on further refining the system's scalability and its ability to handle even more complex real-world scenarios.

References

[1]. M.Himabindu, Revathi V, M.Gupta, A.Rana, ``Neuro-Symbolic AI: Integrating Symbolic Reasoning with Deep Learning", IEEE UPCON 2024.

[2]. M.Gaur, K.Gunaratna, S.Bhatt, "Knowledge-Infused Learning: A Sweet Spot in Neuro-Symbolic AI", IEEE 2022.

[3]. Garcez, A., Gori, M., Lamb, L. C., Serafini, L., Spranger, M., & Tran, S. N. . Neural -symbolic computing: An effective methodology for principled integration of Machine Learning and Reasoning. 2020

[4]. J.Pearl, The seven tools of casual inference with reflections on machine learning.ACM.2019

[5]. H. Kautz. The Third AI Summer, Thirty-fourth AAAI Conference on Artificial Intelligence, 2020.

[6]. F.Arabshahi, S. Singh and A AnandKumar. Combining Symbolic expressions and black-box function evaluations in neural programs. ICLR,2018

[7]. R. Evans and E.Grenfenstette. Learning explanatory rules from noisy data. JAIR,2018