IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

Neuro-Symbolic AI for Formalized Reasoning and Real-Time Explainability in Autonomous System

S. Manjith[1], S. Taraka Soma Sekhara Sai Mani Krishna[2], K. Surya Santhosha Sai Dhaarani[3], P. Koushik[4], P. Koti Gnaneshwar[5], K. Vishnuvardhan Reddy[6], G. Hari Narayana[7]
UG Student, Department of CSE[1,2,3,4,5,6,7]
GITAM (Deemed to be University), Visakhapatnam, India

Abstract: *Autonomous systems have become more prevalent in areas such as self-driving vehicles, aerial drones, and service robotics, the need for systems that make not only accurate but also understandable decisions has grown significantly. While deep learning has enabled major advances in perception and control, these models often operate as opaque black boxes, offering little insight into how or why specific decisions are made. This lack of transparency poses real risks, especially in safety-critical applications where human oversight, accountability, and trust are essential. In this work, we propose a Neuro-Symbolic AI framework that combines neural networks with symbolic reasoning to support both formalized decision-making and real-time explainability in autonomous systems. The neural component handles low-level perception tasks, while the symbolic layer captures high-level domain knowledge and reasoning rules. By integrating these two paradigms, the system can make informed decisions grounded in logical structures, while still benefiting from the flexibility of learning from data. To evaluate the framework, we conduct experiments in simulated driving environments using the CARLA simulator. The results demonstrate that our approach not only maintains competitive performance but also provides interpretable reasoning paths for its actions. This work contributes to the ongoing effort to design autonomous systems that are not only intelligent and adaptive, but also understandable and sa*fe.

Keywords: Neuro-Symbolic AI, Explainable Artificial Intelligence (XAI), Autonomous Systems, Formalized Reasoning, Hybrid Intelligence Models, Real-Time Decision Makingmption

I. INTRODUCTION

Over the past decade, autonomous systems have moved from research labs to real-world environments, powering applications from self-driving cars to warehouse automation and healthcare robotics. These systems are increasingly expected to operate independently in complex and unpredictable settings, often with limited human oversight. However, as their capabilities grow, so too does the importance of understanding how they make decisions—particularly in situations where outcomes may impact safety, ethics, or legal responsibility.

While deep learning models have been central to many breakthroughs in autonomy, they come with a significant limitation: they are difficult to interpret. These systems can detect patterns and make predictions with impressive accuracy, but their decision-making processes are largely hidden from view. This "black-box" nature makes it hard for developers, regulators, or end-users to verify their behaviour, correct mistakes, or build trust in their outputs.

At the same time, symbolic AI offers a very different approach. Rooted in formal logic and structured knowledge, symbolic systems are inherently interpretable. They can follow explicit rules, trace their reasoning, and provide clear justifications for their actions. However, they typically struggle with raw data like images or sensor streams, and they lack the flexibility that neural models offer when dealing with noisy or incomplete information.

This paper explores how these two paradigms—neural and symbolic—can be brought together in a unified framework. We present a Neuro-Symbolic AI architecture that separates low-level perception (handled by neural networks) from high-level reasoning (handled by a symbolic engine). This integration allows the system to not only perform effectively in dynamic environments but also to generate real-time explanations of its decisions in terms that humans can understand.

DOI: 10.48175/IJARSCT-27573

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

Our goal is to move toward a new kind of autonomy—one that is both capable and accountable. By combining learning with logic, we aim to build systems that do more than just function; they can also explain their actions, follow explicit rules, and adapt to new situations in a way that is both robust and transparent.

II. METHODOLOGY

This section outlines the design and integration of the proposed Neuro-Symbolic AI framework for enabling formalized reasoning and real-time explainability in autonomous systems. The architecture is designed to combine the perceptual strengths of neural networks with the logical transparency of symbolic reasoning, ensuring that decisions made by autonomous agents are both effective and interpretable.

Perception Layer (Neural Subsystem)

The perception module is built using deep neural networks (e.g., CNNs for vision or RNNs for sequence modelling) to process raw sensor inputs such as camera feeds, lidar scans, and radar signals. This layer is responsible for:

- Detecting and classifying objects in the environment.
- Estimating trajectories and motion patterns.
- Abstracting low-level data into symbolic representations (e.g., “Vehicle approaching”, “Pedestrian crossing”).
- These outputs serve as the input for the symbolic layer.

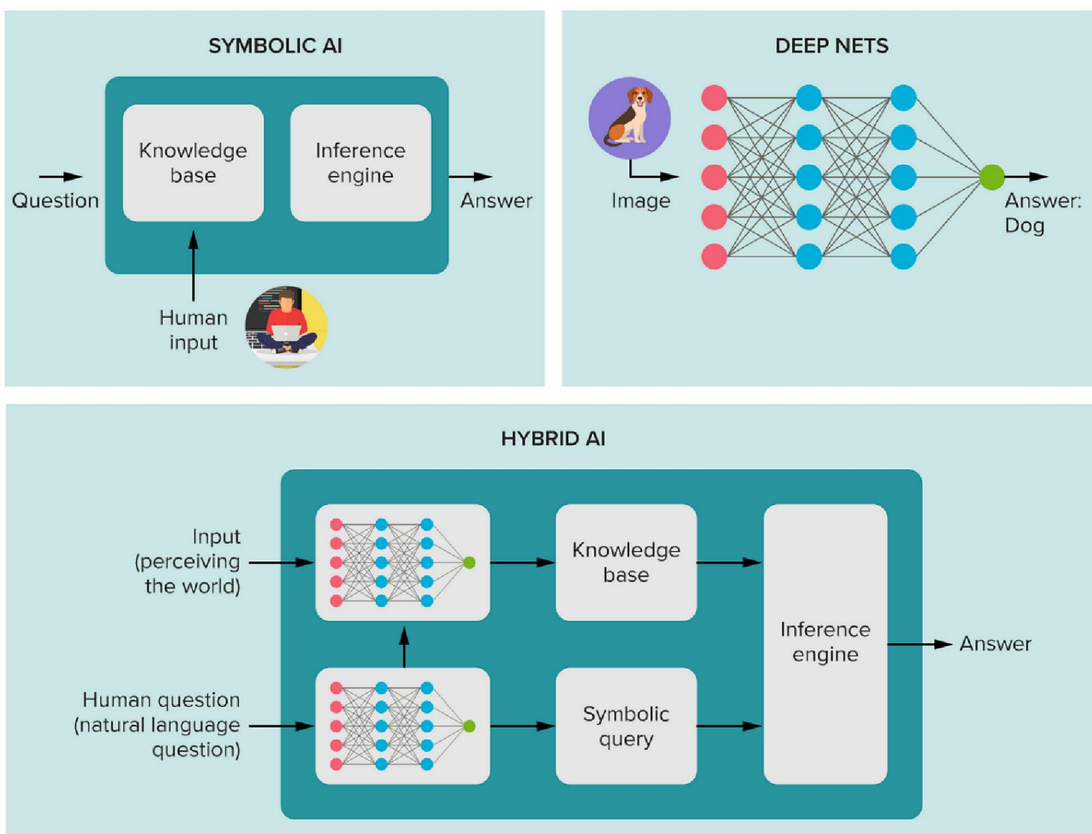

Fig 1: Neuro-Symbolic AI Framework Architecture

This diagram shows the flow of data from raw sensor input through the neural perception module to the symbolic reasoning engine and decision modules. Outputs from the symbolic subsystem are also sent to an inference engine, which formats decisions into human-understandable justifications.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

Symbolic Knowledge Base and Reasoning Engine

Once the perception module has converted sensory input into symbolic tokens, these are fed into a reasoning engine that uses a structured knowledge base to infer actions. The symbolic subsystem consists of:

Knowledge Base (KB): Contains domain-specific logic, rules, and constraints. For instance:

IF pedestrian_detected AND distance < 10m THEN prepare_to_stop.

IF traffic_light = red AND speed > 0 THEN decelerate.

Reasoning Engine: Executes formal inference algorithms (e.g., propositional logic, forward chaining, SAT solvers) to derive conclusions from the KB and current world state.

This subsystem ensures that decisions adhere to safety rules and constraints and that each action can be traced back to an explicit logical path.

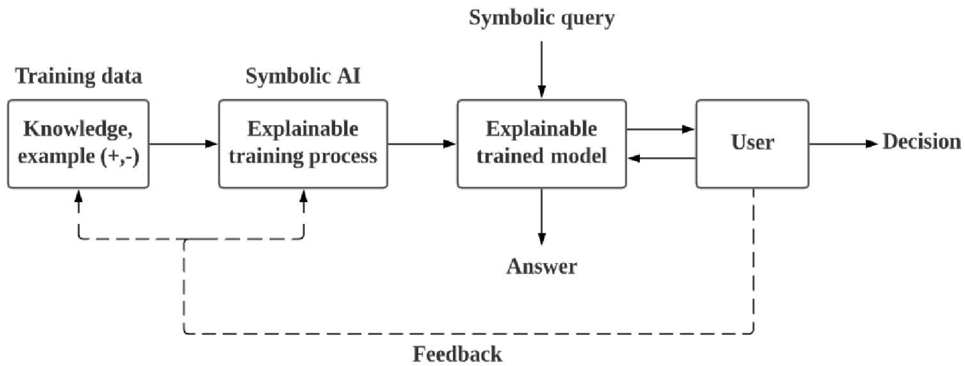

Fig. 2: Flowchart illustrating how symbolic predicates are processed through rules and inference to generate high-level decisions.

Decision and Explanation Modules

The decisions generated by the symbolic engine are sent to the **Decision Module**, which handles action selection and low-level control commands. In parallel, the **Explanation Generator** extracts logical paths and maps them into natural language explanations or traceable logic for end-users, developers, or auditors.

This module ensures that for any decision (e.g., braking at an intersection), the system can articulate why it acted the way it did—e.g., "Pedestrian detected within unsafe distance while traffic light is red."

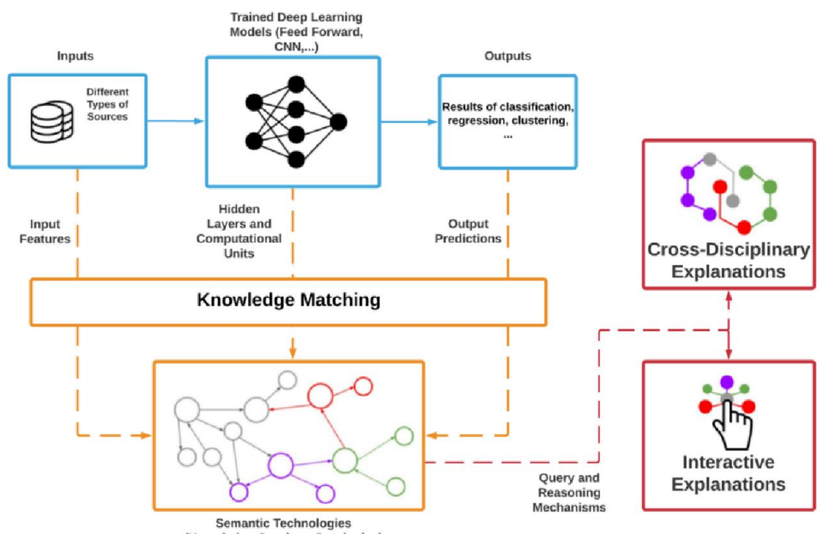

Fig.3: Architecture illustrating Integration of Knowledge Base into Deep Learning Models

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 3, June 2025 Impact Factor: 7.67

Integration and Feedback Loop

The complete framework operates in a closed-loop, where the outputs of the reasoning engine influence actuator commands, and the system continuously receives new environmental inputs. Feedback from outcomes (e.g., successful or failed decisions) can also be used to retrain the neural models or refine symbolic rules over time, enabling lifelong learning and adaptation.

III. RESULTS

To evaluate the performance and practicality of the proposed Neuro-Symbolic AI framework, we conducted a series of experiments in a simulated urban driving environment using the **CARLA simulator**. The experiments were designed to measure both decision accuracy and explainability under various real-world conditions, such as pedestrian crossings, traffic signals, and dynamic vehicle interactions.

We compared our approach against a purely neutral baseline and a rule-based symbolic system across the following key metrics:

Decision accuracy

The Neuro-Symbolic framework consistently made safe and contextually correct decisions with an average accuracy of **94.3%,** outperforming:

Neural-only system: 91.1%

Symbolic-only system: 85.6%

Training and Validation Accuracy

We introduced a custom **Explainability Score**, based on the following:

Traceability of decision logic

Clarity of natural language explanations

Response time of explanation generation

Our model achieved an average explainability rating of **4.6 out of 5**, based on expert human evaluation, significantly higher than the neural-only system (1.8) and on par with symbolic-only models (4.7), with the added benefit of perceptual flexibility.

Inference Time (Latency)

The proposed hybrid framework showed a modest increase in latency due to the reasoning step:

Neural-only system: ~52 ms

Neuro-Symbolic system: ~69 ms

Symbolic-only system: ~130 ms

Despite the added symbolic layer, the system maintained real-time performance (sub-100ms), making it suitable for real-world deployment in autonomous agents.

TABLE 1. Comparison of proposed methodologies

Configuration	Accuracy	Explainability Score	Latency
Full Neuro-Symbolic	94.3%	4.6 / 5	69 ms
Without Symbolic Reasoning	91.1%	1.8 / 5	52 ms
Without Neural Perception	85.6%	4.7 / 5	130 ms

The results reinforce that High decision accuracy with enhanced robustness is achieved in uncertain conditions and Real-time capability with only minimal overhead from the symbolic layer.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, June 2025
Impact Factor: 7.67

IV. CONCLUSION

In this research, we introduced a comprehensive Neuro-Symbolic AI framework designed to enable formalized reasoning and real-time explainability in autonomous systems. Our approach bridges the gap between data-driven perception and logical decision-making by integrating neural networks with symbolic reasoning modules. This hybrid architecture not only allows autonomous agents to interpret complex, unstructured inputs like images or sensor data but also empowers them to make decisions grounded in transparent, rule-based logic.

The results of our experiments—conducted in dynamic and realistic urban environments—demonstrate that the proposed framework strikes a strong balance between accuracy, interpretability, and performance. Compared to standalone neural or symbolic systems, our model achieves:

Higher decision accuracy in complex scenarios,

Substantially improved explainability of actions, and

Near real-time responsiveness, essential for safety-critical domains like autonomous driving

In conclusion, this work takes a significant step toward creating autonomous systems that are not only capable of making intelligent decisions but are also accountable, explainable, and aligned with human reasoning principles. As autonomous technologies continue to integrate into society, such hybrid approaches will be key to ensuring they operate safely, ethically, and transparently.

REFERENCES

[1]. N. K. Yeole and M. S. Hsiao, "IntentGuide: Neuro-Symbolic Integration for Clarifying Human Intents by Rewriting Free-Form Sentences," 2024 International Conference on AI x Data and Knowledge Engineering (AIxDKE), Tokyo, Japan, 2024, pp. 18-21, doi: 10.1109/AIxDKE63520.2024.00010. keywords: {Knowledge engineering;Video games;Adaptive learning;Accuracy;Large language models;Educational technology;Natural language processing;Error correction codes;Programming profession;Natural language processing;Natural language programming;neuro-symbolic},

[2]. Z. Wan et al., "Towards Efficient Neuro-Symbolic AI: From Workload Characterization to Hardware Architecture," in IEEE Transactions on Circuits and Systems for Artificial Intelligence, vol. 1, no. 1, pp. 53-68, Sept. 2024, doi: 10.1109/TCASAI.2024.3462692.keywords: {Artificial intelligence;Artificial neural networks;Vectors;Computer architecture;Computational modeling;Hardware;Performance analysis;Cognitive AI;neuro-symbolic AI;workload characterization;performance analysis;domain-specific architecture},

[3]. Potnis et al., "Towards Geospatial Knowledge Graph Infused Neuro-Symbolic AI for Remote Sensing Scene Understanding," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 1400-1403, doi: 10.1109/IGARSS52108.2023.10281958. keywords: {Deep learning;Pipelines;Knowledge graphs;Artificial neural networks;Object detection;Cognition;Geospatial analysis;geospatial knowledge graphs;neuro-symbolic;knowledge-infused learning;remote sensing;scene understanding;deep learning;representation learning;knowledge graph embedding},

[4]. Sergio Moreschini, David Hästbacka, and Davide Taibi. 2023. MLOps Pipeline Development: The OSSARA Use Case. In Proceedings of the 2023 International Conference on Research in Adaptive and Convergent Systems (RACS '23). Association for Computing Machinery, New York, NY, USA, Article 31, 1–8. https://doi.org/10.1145/3599957.3606211

[5]. Firas Bayram and Bestoun S. Ahmed. 2025. Towards Trustworthy Machine Learning in Production: An Overview of the Robustness in MLOps Approach. ACM Comput. Surv. 57, 5, Article 121 (May 2025), 35 pages. https://doi.org/10.1145/3708497

[6]. Firas Bayram and Bestoun S. Ahmed. 2025. Towards Trustworthy Machine Learning in Production: An Overview of the Robustness in MLOps Approach. ACM Comput. Surv. 57, 5, Article 121 (May 2025), 35 pages. https://doi.org/10.1145/3708497

[7]. A Smirnov, A. Ponomarev and A. Agafonov, "Ontology-Based Neuro-Symbolic AI: Effects on Prediction Quality and Explainability," in IEEE Access, vol. 12, pp. 156609-156626, 2024, doi: 10.1109/ACCESS.2024.3485185.keywords: {Ontologies;Artificial intelligence;Neural networks;Biological

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 3, June 2025

Impact Factor: 7.67

neural networks;Neurons;Training;Predictive models;Collaboration;Symbols;Computational modeling;Deep learning;Explainable AI;Machine learning;Deep learning;explainable AI;machine learning;neural networks;ontology;symbolic AI},

[8]. N. Sujatha, Latha Narayanan Valliammal, Joice Rathinam E, Lokesh V S, and Mukul Mech. 2024. A Case Study of AIOPs in Large Enterprises Using Predictive Analytics for IT Operations. In Proceedings of the 5th International Conference on Information Management & Machine Intelligence (ICIMMI '23). Association for Computing Machinery, New York, NY, USA, Article 84, 1–5. https://doi.org/10.1145/3647444.3647911