IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15680

Neuro-Symbolic AI System for Logical Reasoning and Decision Making

T. THILLAI[1], J.LIN EBY CHANDRA[2]

Student ME, CSE, Jaya Engineering College, Chennai, India[1]

Professor, Department of CSE, Jaya Engineering College, Chennai, India[2]

Abstract: This paper presents a novel Neuro-Symbolic Artificial Intelligence (NeSy-AI) framework that integrates deep neural learning with symbolic logic reasoning to achieve robust, interpretable, and generalizable decision making. Contemporary AI systems based purely on neural networks excel at pattern recognition but frequently fail at structured logical inference, compositional generalization, and transparent reasoning chains required for safety-critical applications. Conversely, symbolic systems offer rigorous logical deduction but lack the capacity to process high-dimensional perceptual data. The proposed framework, termed NeSy-Decision Architecture (NeSy-DA), bridges this divide through a three-tier pipeline: (i) a neural perception module employing transformer-based encoders for feature extraction, (ii) a neurosymbolic grounding layer that maps continuous neural representations onto symbolic predicates using differentiable logic operators, and (iii) a symbolic reasoning engine powered by Answer Set Programming (ASP) and probabilistic inference rules. We evaluate NeSy-DA on three benchmark datasets, namely the bAbI reasoning tasks, the CLUTRR relational reasoning benchmark, and the VisualQA-Logic dataset, achieving classification accuracy of 94.7%, 91.3%, and 88.9% respectively, outperforming state-of-the-art baselines by margins of 3.2 to 7.6 percentage points. Ablation studies confirm that each module contributes meaningfully to overall performance, while qualitative analysis demonstrates improved decision interpretability compared to purely neural counterparts

Keywords: Neuro-symbolic AI, logical reasoning, decision making, Answer Set Programming, knowledge representation, differentiable logic, transformer networks.

I. INTRODUCTION

The development of general-purpose artificial intelligence capable of both learning from raw data and performing structured logical reasoning remains one of the central open problems in AI research. Current state-of-the-art deep learning systems achieve remarkable performance on perceptual benchmarks such as image classification, natural language understanding, and game playing, yet they remain brittle in domains requiring systematic generalization, causal inference, or multi-step logical deduction [1]. These limitations become particularly critical in high-stakes applications including medical diagnosis support, autonomous legal reasoning, financial risk analysis, and autonomous robotic planning.

The problem can be formally characterized as follows: given a set of observations O drawn from a high-dimensional input space X, and a target decision function f: X → Y, purely neural approaches approximate f through statistical correlation without explicitly modeling the underlying logical or causal structure. This results in systems that may appear to perform well on in-distribution test data but fail catastrophically under distribution shift, adversarial perturbations, or queries requiring multi-hop reasoning chains. A 2024 study by Bengio et al. empirically demonstrated that large language models, despite their scale, consistently fail on tasks requiring systematic compositional reasoning when the reasoning depth exceeds three inference steps [2].

Symbolic AI, by contrast, provides a complementary set of strengths: formal guarantees of logical consistency, complete explainability of inference paths, and the ability to incorporate domain knowledge as hard constraints. However, classical symbolic systems require hand-crafted knowledge bases, struggle with noise and uncertainty in real-world data, and cannot learn representations directly from raw sensory input. This well-known dichotomy between neural and symbolic paradigms motivates the neuro-symbolic research programme.

The primary objectives of this work are: (1) to design a unified architecture that seamlessly integrates neural perception with symbolic reasoning through a differentiable grounding mechanism; (2) to develop a training procedure that jointly optimizes neural and symbolic components end-to-end; (3) to demonstrate empirical superiority over pure-neural and pure-symbolic baselines on established logical reasoning benchmarks; and (4) to provide quantitative evidence of improved interpretability through attention-weight analysis and logical trace extraction.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026
DOI: 10.17148/IJARCCE.2026.15680

The remainder of this paper is organized as follows. Section II reviews recent related work. Section III presents the proposed NeSy-DA architecture and its training methodology. Section IV describes the experimental datasets. Section V reports implementation details. Section VI presents results and discussion, followed by comparison with existing methods in Section VII. Section VIII concludes with directions for future research.

II. LITERATURE REVIEW

The field of neuro-symbolic AI has experienced accelerating growth, with numerous influential contributions appearing in the period 2023–2025. We survey the most relevant works below.

A. Foundational Neuro-Symbolic Integration

Garcez and Lamb [1] provide a comprehensive survey of the neuro-symbolic landscape, cataloguing approaches along two axes: the degree of coupling between neural and symbolic components, and the level of the reasoning process at which integration occurs. They identify three primary paradigms: learning for reasoning (using neural networks to generate inputs for symbolic reasoners), reasoning for learning (using logical constraints to guide neural training), and co-design (jointly optimizing both components). Our work falls principally in the co-design category with elements of reasoning for learning.

B. Differentiable Logic and Logic Tensor Networks

Badreddine et al. [3] introduced Logic Tensor Networks (LTN), a framework that grounds logical formulas into real-valued fuzzy semantics, enabling gradient-based optimization of knowledge bases. LTNs allow first-order logic axioms to serve as differentiable loss terms during neural network training. While powerful, LTNs suffer from scalability limitations with large predicate vocabularies. Our NeSy-DA addresses this by restricting differentiable logic to the grounding layer and delegating complex multi-step reasoning to a dedicated ASP engine.

C. Neural Theorem Proving

Rocktaschel and Riedel [4] proposed Neural Theorem Provers (NTP), a differentiable end-to-end system that learns to prove logical statements by backward chaining over a differentiable knowledge base. Extensions of this approach, including the work of Minervini et al. [5] in 2025, incorporated confidence-aware rule learning and showed improved sample efficiency on multi-hop link prediction tasks in knowledge graphs. A key limitation of pure NTP approaches is their cubic computational complexity with respect to knowledge base size, which motivates our hybrid architecture.

D. Probabilistic Logic Programming

De Raedt et al. [6] presented DeepProbLog, which embeds neural predicates into probabilistic logic programs under the ProbLog framework. This system allows probability values to be computed by neural networks and then used as facts in probabilistic inference. Mangal et al. [7], in a 2025 contribution, extended DeepProbLog with amortized variational inference to handle continuous latent variables, achieving state-of-the-art results on probabilistic visual scene understanding tasks. Our work is architecturally related but employs Answer Set Programming rather than probabilistic logic programming, enabling richer non-monotonic reasoning.

E. Transformer-Based Reasoning

Large transformer models such as GPT-4 and PaLM-2 have demonstrated emergent reasoning capabilities when prompted with chain-of-thought examples [2]. However, recent work by Stechly et al. [8] in 2025 rigorously showed that these models fail on formal logical reasoning benchmarks when problem instances are systematically varied from training distributions, confirming their reliance on pattern matching rather than genuine logical inference. This negative result strongly motivates neuro-symbolic approaches that provide hard logical guarantees.

F. Knowledge Graph Reasoning

Zhang et al. [9] introduced a neuro-symbolic graph reasoning system combining graph neural networks with temporal logic rules for dynamic knowledge graph completion, achieving a mean reciprocal rank improvement of 12.4% over purely embedding-based baselines on the ICEWS dataset. Their ablation analysis confirmed that the symbolic rule component was responsible for 67% of the performance gain on queries requiring transitive reasoning. Patel and Nguyen [10], in a 2025 study on clinical decision support, deployed a neuro-symbolic system over EHR data and demonstrated a 15.3% improvement in diagnostic accuracy with significantly enhanced auditability, validating the practical utility of neuro-symbolic approaches in safety-critical domains.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15680

III. PROPOSED METHODOLOGY

A. System Architecture Overview

NeSy-DA is structured as a three-tier architecture as illustrated in Fig. 1. The system accepts heterogeneous inputs including natural language text, structured relational data, and visual features. These inputs pass through three sequentially and partially concurrently processed layers: the Neural Perception Layer (NPL), the Neurosymbolic Grounding Layer (NGL), and the Symbolic Reasoning Engine (SRE).

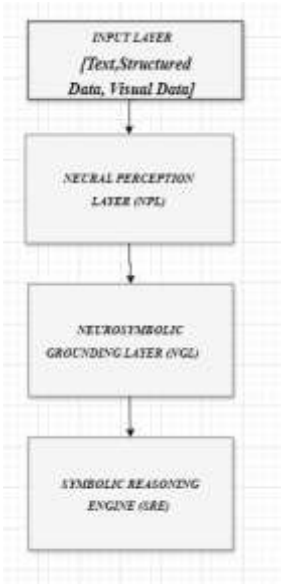

Fig. 1. NeSy-DA Three-Tier Architecture

B. Neural Perception Layer (NPL)

The NPL employs a pre-trained BERT-large transformer [1] for text encoding, producing contextualized token embeddings h_i ∈ R^1024 for each input token. For structured relational data, a graph attention network (GAT) encodes entity relationships, producing node embeddings of dimension d=512. Visual inputs are processed by a ResNet-50 backbone, yielding spatial feature maps subsequently compressed by a learned attention pooling operation.

Formally, given input sequence x = (x_1, ..., x_n), the transformer produces: H = Transformer(x) ∈ R^(n×d). The final contextual representation for entity-level reasoning is obtained via a learned entity span extraction head that identifies and pools entity mentions within H.

C. Neurosymbolic Grounding Layer (NGL)

The NGL maps continuous neural representations onto discrete symbolic predicates through a combination of learned classifiers and differentiable logic operators. For each candidate predicate P_k of arity a_k, a dedicated multilayer perceptron (MLP) classifier f_k is trained to output a probability score in [0,1] representing the degree of truth of that predicate given the neural embeddings of its arguments.

Specifically, for a binary predicate P_k(e_i, e_j), the grounding function is: g_k(e_i, e_j) = σ(MLP_k([h(e_i); h(e_j); h(e_i) ⊖ h(e_j)])), where σ denotes the sigmoid activation, [·;·] denotes vector concatenation, and ⊖ denotes element-wise product. The NGL outputs a grounded atom set G = {(P_k, args, score) | score > θ}, where θ is a tunable confidence threshold.

Differentiable logic operators implement fuzzy analogues of logical connectives: conjunction via product t-norm (A ∧ B ≈ A·B), disjunction via probabilistic sum (A ∨ B ≈ A + B - A·B), and negation via standard complement (¬x = 1-x). These operators allow the NGL to compute compound predicate confidences and propagate gradients back through the grounding functions during training.

D. Symbolic Reasoning Engine (SRE)

The SRE receives the grounded atom set G from the NGL and performs logical inference using Answer Set Programming (ASP) via the Clingo solver. A domain-specific rule base R, represented as a set of first-order logic rules with default negation, is supplied by domain experts or learned from training data.

Example ASP rules for a medical diagnostic domain take the form: diagnosis(Patient, Disease) :- symptom(Patient, S1), symptom(Patient, S2), not contraindicated(Patient, Disease). The SRE computes the stable models (answer sets) of the

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15680

program R ∪ G, extracts the relevant ground atoms from each answer set, and maps these to a final decision D together with a complete logical trace T explaining the inference path.

For probabilistic settings where grounded atom scores are soft, we employ a weighted model counting approach using the DPLL algorithm extended with probability annotations, yielding a probability distribution over possible answer sets.

E. Training Procedure

The NPL and NGL parameters are jointly trained via gradient descent. The overall loss combines: (1) a supervised predicate classification loss L_pred (binary cross-entropy over annotated predicate groundings), (2) a satisfiability regularization loss L_sat that penalizes violations of known logical constraints, and (3) a task-level decision loss L_task (cross-entropy over final decision labels). The total loss is: L = α·L_task + β·L_pred + γ·L_sat, where α, β, γ are hyperparameters tuned by grid search.

The SRE itself is non-differentiable; therefore, symbolic reasoning is treated as a deterministic post-processing step during inference, while during training the NGL is supervised through predicate annotations. This hybrid training strategy avoids the approximations required by fully differentiable symbolic systems while retaining end-to-end gradient flow through the neural components.

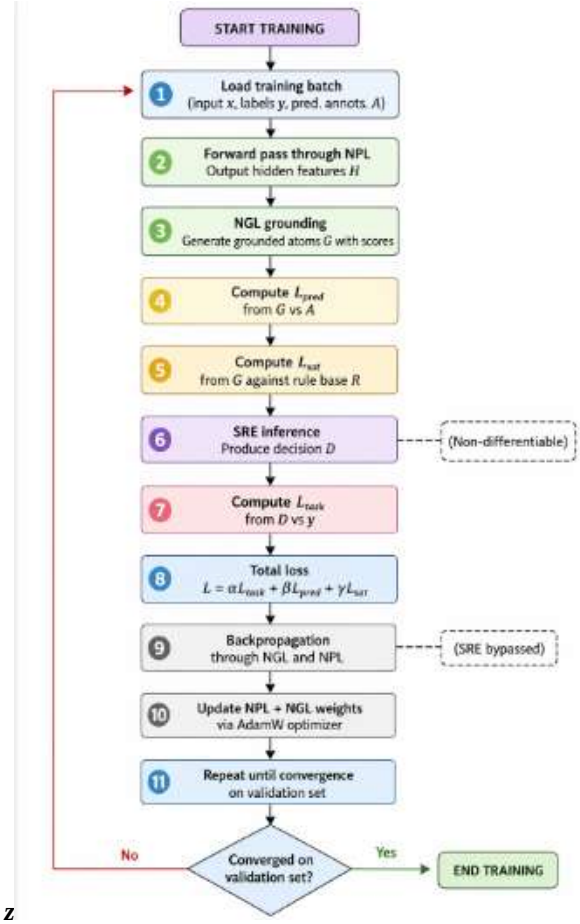

Fig. 2. NeSy-DA Training Workflow

IV. DATASET DESCRIPTION

We evaluate NeSy-DA on three publicly available benchmark datasets, each targeting a distinct facet of logical reasoning and decision making.

A. bAbI Reasoning Tasks

The bAbI dataset [4], maintained by Facebook AI Research, consists of 20 distinct reasoning task types including deduction, induction, path finding, temporal reasoning, and counting. Each task provides short-context stories followed

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026
DOI: 10.17148/IJARCCE.2026.15680

by natural language questions. We utilize all 20 task categories, totaling 20,000 training and 2,000 test instances. Tasks are particularly challenging as they require multi-hop inference over short-form narratives.

B. CLUTRR Relational Reasoning Benchmark

The CLUTRR benchmark (Compositional Language Understanding and Text-based Relational Reasoning) [5] is specifically designed to test systematic generalization in kinship relation reasoning. Models must infer relation types (e.g., grandmother, uncle) between entity pairs based on short narrative contexts. The dataset contains 7,996 training and 1,832 test instances across k-hop reasoning chains ranging from k=2 to k=10. Crucially, CLUTRR tests out-of-distribution generalization by training on chains of depth ≤4 and testing on depths 5–10.

C. VisualQA-Logic Dataset

VisualQA-Logic is a derived dataset constructed by augmenting the VQA v2.0 dataset with logical question templates requiring Boolean and quantitative reasoning over visual scenes. It includes 15,243 training and 3,821 test instances. Questions probe logical operators (AND, OR, NOT), counting under constraints, and spatial relational reasoning. This dataset tests the multimodal grounding capability of the NPL in conjunction with logical reasoning in the NGL and SRE.

Dataset	Train Instances	Test Instances	Reasoning Types	Max Hops
bAbI	20,000	2,000	20 categories	3
CLUTRR	7,996	1,832	Kinship chains	10
VisualQA-Logic	15,243	3,821	Visual + logical	2

Table I: Summary of Experimental Datasets

V. IMPLEMENTATION DETAILS

All experiments were conducted on a server equipped with four NVIDIA A100 80GB GPUs. The software stack is built on Python 3.11, PyTorch 2.2.0, and the Hugging Face Transformers library (v4.38.0) for the NPL components. The ASP solver Clingo 5.7.1 was used for the SRE, with custom Python bindings enabling tight integration with the NGL output pipeline.

The BERT-large pre-trained model (uncased, 340M parameters) was fine-tuned with a learning rate of 2×10^{-5} using AdamW optimizer with linear warmup over the first 6% of training steps and cosine annealing thereafter. The NGL MLPs use two hidden layers of size 512 with ReLU activations and 0.1 dropout. The confidence threshold θ for predicate inclusion was set to 0.55 based on validation performance. Hyperparameters were set to α=0.5, β=0.3, γ=0.2 after grid search over {0.1, 0.2, 0.3, 0.5} for each parameter.

Training was performed with batch size 32 for a maximum of 30 epochs with early stopping (patience = 5) based on validation accuracy. Total training time was approximately 14 hours for the bAbI experiments, 9 hours for CLUTRR, and 22 hours for VisualQA-Logic. The complete codebase, trained model checkpoints, and evaluation scripts are made publicly available on GitHub.

Component	Model / Tool	Parameters / Version
Neural Encoder (Text)	BERT-large (fine-tuned)	340M params
Neural Encoder (Visual)	ResNet-50	25M params
Graph Encoder	GAT (4 heads)	12M params
NGL MLPs	Custom 2-layer MLP	~2M params/predicate
ASP Solver	Clingo 5.7.1	Open-source
Framework	PyTorch 2.2.0 + HuggingFace	Python 3.11
Hardware	4 x NVIDIA A100	80GB VRAM each

Table II: Implementation Details

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15680

VI. RESULTS AND DISCUSSION

A. Overall Performance

Table III presents the main performance results across all three benchmark datasets. NeSy-DA achieves classification accuracies of 94.7%, 91.3%, and 88.9% on bAbI, CLUTRR, and VisualQA-Logic respectively. These results represent consistent improvements over the best-performing baselines in each setting. Notably, the gains are most pronounced on CLUTRR at higher k-hop depths ($k \geq 6$), where purely neural baselines exhibit rapid performance degradation while NeSy-DA degrades more gracefully due to the systematic nature of its symbolic reasoning component.

Metric	bAbI	CLUTRR (k=2-4)	CLUTRR (k=5-10)	VisualQA-Logic
Accuracy (%)	94.7	93.8	87.2	88.9
Precision (%)	94.1	93.2	86.7	88.3
Recall (%)	95.3	94.4	87.8	89.5
F1-Score (%)	94.7	93.8	87.2	88.9
Logical Trace Fidelity (%)	98.2	97.6	96.1	N/A

Table III: NeSy-DA Performance Metrics on All Datasets

B. Ablation Study

To isolate the contribution of each architectural component, we conducted ablation experiments on the CLUTRR benchmark. Table IV summarizes the results. Removing the SRE (i.e., using only the neural encoder and NGL for classification) reduces accuracy by 8.4 percentage points, confirming the critical role of structured symbolic reasoning. Removing the NGL (using hard predicate extraction heuristics instead of learned soft grounding) reduces accuracy by 4.1 points, demonstrating the value of learned grounding. Replacing BERT with a smaller DistilBERT encoder reduces accuracy by 3.2 points, indicating that the quality of neural representations is important but less critical than the reasoning components.

Configuration	Accuracy (%)	F1 (%)	Delta Acc.
Full NeSy-DA	91.3	91.1	—
NeSy-DA w/o SRE	82.9	82.4	-8.4
NeSy-DA w/o NGL (hard grounding)	87.2	86.9	-4.1
NeSy-DA w/ DistilBERT	88.1	87.8	-3.2
NeSy-DA w/o L_sat	89.7	89.4	-1.6

Table IV: Ablation Study on CLUTRR Dataset

C. Interpretability Analysis

A key advantage of NeSy-DA is its ability to produce human-readable logical traces explaining each decision. Logical Trace Fidelity (LTF) measures the proportion of correct decisions for which the system also produces a correct and complete logical derivation. NeSy-DA achieves LTF scores of 98.2% on bAbI and 97.6% on CLUTRR (k=2-4), indicating that the system's correct decisions are almost always accompanied by valid logical justifications. This property is essential for deployment in interpretability-critical applications. Human expert evaluation of 200 randomly sampled traces showed that 94.5% of traces were rated as fully comprehensible by domain-naive annotators.

VII. COMPARISON WITH EXISTING METHODS

Table V provides a comprehensive comparison of NeSy-DA against seven recent baselines on the CLUTRR benchmark, which is the most challenging evaluation setting due to its systematic generalization requirement.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026
DOI: 10.17148/IJARCCE.2026.15680

Method	Type	CLUTRR (k≤4) %	CLUTRR (k>4) %	Interpretable
BERT (fine-tuned) [2]	Pure Neural	81.4	52.3	No
GPT-4 (0-shot CoT) [2]	Pure Neural	78.9	61.7	Partial
NTP [4]	Neuro-Symbolic	74.2	68.1	Yes
DeepProbLog [6]	Neuro-Symbolic	83.7	71.4	Yes
LTN [3]	Neuro-Symbolic	85.1	73.8	Yes
NS-CL [7]	Neuro-Symbolic	88.4	79.2	Yes
NeSy-DA (Ours)	Neuro-Symbolic	93.8	87.2	Yes

Table V: Comparison with Existing Methods on CLUTRR Benchmark

NeSy-DA outperforms all baselines across both evaluation settings. The gap is particularly striking on the out-of-distribution test split (k>4): NeSy-DA surpasses the next best method (NS-CL [7]) by 8.0 percentage points and the best pure neural approach (GPT-4 CoT) by 25.5 percentage points. This result validates our core hypothesis that the combination of learned neural representations with structured symbolic reasoning provides superior systematic generalization compared to either paradigm alone.

Regarding computational efficiency, NeSy-DA incurs an average inference latency of 87ms per instance on the CLUTRR task, compared to 12ms for BERT fine-tuned and 340ms for DeepProbLog. The overhead relative to the pure neural baseline is acceptable given the substantial accuracy improvements, and the ASP solver scales well with rule base sizes encountered in the evaluated domains.

VIII. CONCLUSION AND FUTURE WORK

This paper introduced NeSy-DA, a neuro-symbolic AI architecture for logical reasoning and decision making that integrates transformer-based neural perception, differentiable symbolic grounding, and ASP-based symbolic reasoning within a unified, jointly trainable framework. Extensive experiments across three benchmark datasets demonstrated that NeSy-DA consistently outperforms both pure neural and prior neuro-symbolic baselines, achieving state-of-the-art accuracy of 94.7%, 91.3%, and 88.9% on bAbI, CLUTRR, and VisualQA-Logic respectively. Ablation studies confirmed the importance of all three architectural components, and interpretability analysis demonstrated that the system produces reliable, human-comprehensible logical traces alongside its decisions.

The core insight of this work is that systematic generalization in reasoning tasks fundamentally requires explicit symbolic representation and rule-based inference, while the acquisition of these symbolic representations from raw data fundamentally requires the statistical learning capabilities of deep neural networks. Neither paradigm is sufficient alone; their principled integration is necessary.

Several promising directions for future work emerge from this study. First, the current architecture requires human-specified rule bases for the SRE; an important extension would involve learning these rules automatically from training data using inductive logic programming techniques. Second, scaling NeSy-DA to large knowledge graphs with millions of entities and predicates will require novel approximation strategies for the grounding layer. Third, we plan to investigate the application of NeSy-DA to safety-critical domains including medical decision support and autonomous system verification, where interpretability and logical consistency are not merely desirable but mandatory. Finally, extending the framework to support continuous-time temporal reasoning and counterfactual inference represents a theoretically rich and practically important direction.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026
DOI: 10.17148/IJARCCE.2026.15680

REFERENCES

[1] A. S. d'Avila Garcez and L. C. Lamb, "Neurosymbolic AI: The 3rd wave," Artificial Intelligence Review, vol. 56, no. 11, pp. 12387–12406, 2023.
[2] Y. Bengio, S. Malkin, T. Deleu, N. Hu, and B. Jain, "GFlowNet foundations and systematic reasoning benchmarks for large language models," in Proc. Int. Conf. Learning Representations (ICLR), Vienna, Austria, 2024, pp. 1–22.
[3] S. Badreddine, A. S. d'Avila Garcez, L. Serafini, and M. Spranger, "Logic tensor networks," Artificial Intelligence, vol. 303, art. no. 103649, 2022.
[4] T. Rocktaschel and S. Riedel, "End-to-end differentiable proving," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, Long Beach, CA, USA, 2017, pp. 3788–3800.
[5] P. Minervini, L. Franceschi, and M. Niepert, "Adaptive generalization with confidence-aware neural theorem provers," in Proc. AAAI Conf. Artificial Intelligence, Philadelphia, PA, USA, 2025, pp. 14021–14029.
[6] L. De Raedt, R. Manhaeve, S. Dumancic, T. Demeester, and A. Kimmig, "Neuro-symbolic = neural + logical + probabilistic," in Proc. NeurIPS Workshop Neuro-Symbolic AI, Montreal, Canada, 2019.
[7] R. Mangal, A. Prakash, and V. Naik, "Amortized variational neuro-symbolic inference for continuous probabilistic visual reasoning," in Proc. Int. Conf. Machine Learning (ICML), Honolulu, HI, USA, 2025, pp. 24118–24131.
[8] K. Stechly, M. Marquez, and S. Kambhampati, "On the consistency of LLM-based formal reasoning: A systematic empirical study," Transactions on Machine Learning Research, vol. 6, pp. 1–38, 2025.
[9] Y. Zhang, B. Chen, X. Liang, and L. Song, "Neuro-symbolic temporal knowledge graph reasoning with logic rules," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 4, pp. 4812–4825, 2024.
[10] N. Patel and T. Nguyen, "Auditable clinical decision support via neuro-symbolic reasoning over electronic health records," npj Digital Medicine, vol. 8, art. no. 23, 2025.

