Frontiers in Emerging Artificial Intelligence and Machine Learning (Open Access)

Volume 02, Issue 04, April 2025,
Publish Date: 11-04-2025
PageNo.08-13

Integrating Neural And Symbolic AI For Robust Generalized Planning In Robotics

Dr. Arjun Malhotra
Department of Artificial Intelligence, Indian Institute of Science, Bengaluru, India

Meenal Sinha
Department of Robotics and Intelligent Systems, Indian Institute of Technology Bombay, Mumbai, India

ABSTRACT

The integration of neural and symbolic Artificial Intelligence (AI) offers a promising pathway toward achieving robust and generalized planning in robotics. While neural networks excel at perception and pattern recognition, symbolic AI contributes structured reasoning and interpretability. This paper explores a hybrid approach that combines neural perception modules with symbolic planning frameworks to enable robots to operate effectively in dynamic and partially observable environments. The study reviews recent advancements in neurosymbolic architectures for task generalization, real-time decision-making, and cross-domain adaptability. Emphasis is placed on bridging the gap between low-level sensory data and high-level abstract reasoning to support flexible, explainable, and scalable robotic behavior. The findings highlight the potential of integrated neural-symbolic systems to advance autonomous robotics in complex, real-world applications.

KEYWORDS: Neurosymbolic AI, Robotics, Generalized Planning, Symbolic Reasoning, Neural Networks, Hybrid Intelligence, Autonomous Systems, Explainable AI, Task Generalization, Cognitive Robotics.

INTRODUCTION

The field of robotics is undergoing a transformative period, driven by advancements in Artificial Intelligence (AI). Robots are increasingly expected to operate autonomously in complex, dynamic, and unstructured environments, requiring sophisticated planning capabilities that can adapt to novel situations, learn from experience, and reason about the world. Traditional robotic planning often relies on symbolic AI approaches, where knowledge is explicitly represented as facts and rules, enabling logical reasoning and robust decision-making in well-defined domains. However, these symbolic systems typically struggle with perception from raw, noisy sensor data and with generalizing to unforeseen circumstances or learning new skills without extensive manual programming.

Conversely, neural networks and deep learning have revolutionized perception and control in robotics, excelling at tasks like object recognition, grasping, and low-level motor control by learning complex patterns directly from data. Yet, purely neural approaches often lack interpretability, struggle with logical consistency, and face significant challenges in tasks requiring abstract reasoning, long-term planning, and compositional generalization – abilities that are crucial for complex, multi-step robot tasks [6, 7]. For example, a neural network might learn to pick up a specific object but struggle to generalize this skill to a different object or integrate it into a broader task plan like "make breakfast."

The limitations of both purely symbolic and purely neural paradigms in robotics have underscored the necessity for neurosymbolic AI. This emerging field seeks to bridge the gap between connectionist learning (neural networks) and symbolic reasoning (logical systems), aiming to combine their complementary strengths for more intelligent, robust, and generalizable robot behaviors [3]. The core motivation for neurosymbolic robotics is to create systems that can perceive the world through neural networks, reason about it using symbolic logic, and translate these insights into actionable plans for robots. This integration is particularly vital for generalized planning, where a robot must be able to adapt its planning strategies and execute tasks across a wide range of similar, yet distinct, scenarios without being explicitly reprogrammed for each one [6, 7]. This article explores the methodologies, advancements, and implications of applying neurosymbolic approaches to achieve robust generalized planning in AI-powered robotics, paving the way for truly autonomous and adaptable intelligent agents.

METHODS

Neurosymbolic approaches in robotics aim to integrate perception, learning, and action planning by combining neural networks with symbolic reasoning systems. The methodologies employed are diverse, reflecting various ways these two paradigms can interact and exchange information to facilitate generalized planning.

1. Conceptualizing Generalized Planning in Robotics

Generalized planning for robots involves creating systems that can solve a family of related problems or adapt to new, unseen instances of a task without retraining or explicit reprogramming. This requires:

- **Abstraction:** Ability to form abstract representations of the world, objects, and actions.
- **Compositionality:** Capacity to combine basic skills or knowledge elements to form complex plans.
- **Transferability:** Performance across different problem instances within a domain or even across similar domains.
- **Robustness:** Ability to handle noisy sensory input and unexpected variations in the environment.

2. Architectural Frameworks for Neurosymbolic Integration

Various architectural paradigms are explored for integrating neural and symbolic components in robotic planning:

- Sequential or Pipeline Architectures: This is one of the most straightforward integration methods, where neural networks and symbolic systems operate in a pipeline.
    - Neural Perception to Symbolic Representation: Neural networks process raw sensor data (e.g., images, point clouds) to extract symbolic representations (e.g., object types, their states, spatial relations). For instance, a neural network might identify objects and their properties, which are then fed as facts into a symbolic knowledge base or a planner [3].
    - Symbolic Planning to Neural Execution: A symbolic planner generates a high-level plan (a sequence of abstract actions). Neural networks then translate these abstract actions into low-level motor commands for robot execution, learning the mapping from abstract goals to continuous control signals [3].
    - Neurosymbolic Predicators: Recent work proposes systems like VisualPredicator, which learn abstract world models using neurosymbolic predicates. This allows a neural module to predict the effects of actions in an abstract, symbolic space, which is then used by a symbolic planner to generate robust plans [2].
- Integrated/Hybrid Architectures: These approaches involve a tighter coupling where neural and symbolic components interact more directly and iteratively.
    - Neural Probabilistic Logic Programming: Frameworks like DeepProbLog combine neural networks with probabilistic logic programming. Neural networks can learn the probabilities of facts or rules from data (e.g., perceived object properties), which are then used by a probabilistic logic engine to perform uncertain reasoning and generate plans. This allows robots to handle noisy or incomplete sensory information while maintaining logical consistency [4].
    - Learning Neuro-Symbolic Skills for Bilevel Planning: This involves a hierarchical approach where a high-level symbolic planner dictates abstract goals, and lower-level neural networks learn to achieve these goals by executing neuro-symbolic skills. The neural networks might learn to map high-level symbolic states to appropriate actions, with symbolic constraints guiding the learning process [5].
    - Symbolic Goal Decomposition with Language Models: Neurosymbolic language models can be used for fast and accurate task planning by leveraging multi-level goal decomposition. The language model generates high-level plans, which are then refined through symbolic reasoning, ensuring logical consistency and feasibility [8]. This also includes neuro-symbolic natural language navigational planners that translate natural language instructions into robot actions [9].
- Feedback Loops and Self-Correction: Some advanced architectures incorporate feedback loops where the outcome of symbolic planning or robot execution informs the learning process of the neural components. This allows the system to learn from failures and refine its internal representations or rules.

3. Knowledge Representation and Learning

- Symbolic Knowledge Base: Explicit representation of objects, their properties, relations, and action preconditions/effects (e.g., using PDDL or first-order logic). This knowledge provides the foundation for symbolic reasoning and planning.
- Neural Embeddings for Symbolic Concepts: Neural networks can learn vector embeddings for symbolic

entities and relations, allowing reasoning in a continuous space. These embeddings capture semantic similarities and can facilitate generalization.

- Learning Abstract Predicates: Neural networks can be trained to learn the truth values of abstract predicates (e.g., IsClear(BlockA)) directly from visual or sensory input, effectively bridging the gap between raw data and symbolic concepts [2].

4. Training and Evaluation

- Supervised Learning: Training neural components with labeled data for perception tasks (e.g., object detection, state estimation).
- Reinforcement Learning: Training agents to learn policies that map states to actions, often guided by symbolic rewards or constraints [6].
- Imitation Learning: Learning from expert demonstrations, where a human demonstrates a task, and the robot learns to replicate it.
- Evaluation Metrics: Beyond standard robotic metrics (e.g., task success rate, execution time), neurosymbolic systems are also evaluated on their:
    - Generalization to Novelty: Performance on unseen problem instances or variations of a task.
    - Interpretability: Ability to explain the reasoning behind a plan or decision, often through logical derivations.
    - Robustness: Performance under noisy or uncertain conditions.

The methodologies in neurosymbolic robotics are designed to harness the strengths of both neural and symbolic AI, moving towards more capable and understandable autonomous systems for generalized planning.

RESULTS AND APPLICATIONS

The application of neurosymbolic approaches to generalized planning in AI-powered robotics has yielded compelling results, demonstrating significant advancements in capabilities that were previously challenging for purely neural or symbolic systems.

1. Enhanced Generalization and Adaptability

One of the most significant outcomes is the improved ability of robots to generalize across varying task instances and environments.

- Transfer of Skills: Neurosymbolic systems allow for the learning and transfer of abstract skills and knowledge. Instead of learning a specific sequence of low-level motor commands for each task, a robot can learn symbolic actions or predicates that can be composed to solve new problems [2, 5]. For instance, if a robot learns the symbolic predicate IsGrasped(Object) from visual input, it can apply this knowledge to grasp different objects in various settings, leading to more generalized grasping capabilities.
- Robustness to Novelty: By leveraging symbolic reasoning, robots can infer general rules and constraints that hold true across different scenarios, making them more robust to unseen variations in the environment or task parameters. This contrasts with purely neural systems which might fail catastrophically when encountering out-of-distribution data. For example, a neurosymbolic planner might use symbolic rules to identify obstacles in new environments, allowing it to avoid collisions without re-training its perception system for every new obstacle type.
- Multi-level Goal Decomposition: Neurosymbolic language models have shown promise in quickly and accurately decomposing high-level goals into a series of actionable sub-goals, which is crucial for complex, multi-step tasks in robotics. This multi-level decomposition allows robots to handle diverse and complex instructions, enhancing their task planning capabilities [8].

2. Improved Interpretability and Transparency

Neurosymbolic integration addresses the "black-box" problem of deep learning by providing human-interpretable explanations for robot behavior and planning decisions.

- Logical Tracing of Plans: The symbolic component in neurosymbolic planners can generate plans as sequences of logical actions, making the robot's high-level intentions and reasoning process transparent [3]. A human operator can inspect these symbolic plans to understand *why* the robot chose a particular sequence of actions to achieve a goal.
- Explanations for Perceptual Decisions: When neural networks learn to extract symbolic predicates from raw data (e.g., IsRed(Object) or IsHeavy(Object)), the system can explain its perceptual classifications in human-understandable terms, bridging the gap between raw sensor data and abstract reasoning [2].
- Constrained Task Planning: The integration of symbolic logic allows for reasoning about constraints, and neurosymbolic systems can explain why certain actions are not taken or why particular states are unreachable due to violated constraints [10]. This provides a clearer understanding of the planning process and potential limitations.

3. Enhanced Learning Efficiency and Data Efficiency

Neurosymbolic frameworks can often learn more efficiently and from less data than purely neural approaches.

- Prior Knowledge Integration: Pre-existing symbolic knowledge (e.g., physics laws, domain constraints) can be directly incorporated, guiding the learning process of neural networks and reducing the need for extensive data collection. This reduces the search space for neural learning.
- Learning Abstract Skills: Instead of learning every low-level detail, neurosymbolic systems can learn abstract skills that are generalizable across different contexts, requiring fewer training examples for each specific task instance. Learning neuro-symbolic skills for bilevel planning is a key example of this [5].
- Faster Plan Generation: By abstracting away low-level details and performing reasoning at a symbolic level, some neurosymbolic planners can generate plans much faster than purely reactive neural policies, especially for complex, long-horizon tasks.

4. Applications in AI-Powered Robotics

These advancements translate into concrete applications:

- Autonomous Navigation and Exploration: Robots can learn to navigate complex environments, identify landmarks symbolically, and plan paths that respect both geometric and semantic constraints. Neuro-symbolic natural language navigational planners are particularly relevant here [9].
- Human-Robot Collaboration: Improved interpretability facilitates more intuitive and trustworthy human-robot interaction, as humans can better understand the robot's intentions and provide guidance or correction at an abstract level [3].
- Complex Manipulation Tasks: Robots can perform multi-step manipulation tasks requiring reasoning about object properties, spatial relations, and action effects, such as assembling products or performing household chores.
- Robust Adaptation in Unstructured Environments: Robots can adapt to changes in their environment or unexpected events by combining learned perceptual patterns with symbolic re-planning capabilities.

These results underscore that neurosymbolic AI is not just a theoretical concept but a practical approach enabling robots to achieve higher levels of intelligence, autonomy, and explainability, crucial for their effective deployment in real-world scenarios.

DISCUSSION

The integration of neural networks and symbolic AI represents a pivotal advancement for generalized planning in robotics. As demonstrated by the promising results, neurosymbolic frameworks are successfully addressing fundamental limitations inherent in purely connectionist or purely symbolic approaches. The ability to generalize across varying task instances, interpret robot decisions, and learn efficiently from data are transformative capabilities that pave the way for more intelligent and autonomous robotic systems.

The core strength of neurosymbolic AI in robotics lies in its capacity to handle the "perception-to-action" gap more effectively. Neural networks excel at extracting meaningful, sub-symbolic features from noisy, high-dimensional sensor data—a critical first step for any autonomous agent. Subsequently, symbolic reasoning systems can operate on these learned abstractions to perform logical inference, plan sequences of actions, and maintain long-term coherence, addressing the limitations of neural networks in abstract reasoning and compositionality. This synergistic approach leads to robots that are not only capable of complex tasks but also able to explain their rationale, fostering greater trust and facilitating human-robot collaboration [3]. The emergence of neurosymbolic predicates for learning abstract world models is a particularly powerful development, allowing neural perception to directly feed into symbolic planning [2].

Furthermore, the enhanced generalization capabilities are crucial for deploying robots in dynamic, real-world environments. Instead of requiring extensive retraining for every minor variation in a task or environment, neurosymbolic robots can leverage learned symbolic knowledge to adapt their behavior, leading to more robust and versatile systems. This is particularly relevant for generalized planning, where the aim is to create systems that can solve entire families of problems rather than just individual instances. The efficiency gains in learning, often requiring less data due to the integration of prior symbolic knowledge or the learning of abstract skills, also contribute to making advanced robotic systems more feasible to develop and deploy.

Despite these significant advancements, several challenges remain in the widespread adoption and further development of neurosymbolic AI for generalized planning in robotics:

- Scalability of Symbolic Reasoning: As robotic tasks become more complex, the symbolic state space can grow exponentially, posing scalability challenges for traditional symbolic planners. Optimizing symbolic inference within neurosymbolic frameworks is an ongoing area of research.
- Robustness of Perception-to-Symbolic Mapping: The neural component's ability to accurately translate raw sensor data into reliable symbolic representations is critical. Errors in this translation can propagate through the symbolic reasoning pipeline, leading to planning failures. Handling ambiguity and uncertainty in perception remains a significant hurdle.
- Interoperability and Communication: Developing seamless and efficient communication interfaces between diverse neural network architectures and

symbolic reasoning engines is complex. Ensuring that information flows effectively and without loss between the sub-symbolic and symbolic layers is paramount.

- Learning Abstract Concepts Automatically: While neurosymbolic systems can leverage pre-defined symbolic knowledge, the automatic discovery and refinement of new, useful abstract concepts and rules from robot experience remains a difficult problem.
- Real-time Performance: For many robotic applications, real-time planning and execution are essential. The computational overhead of integrating multiple AI paradigms can sometimes hinder real-time performance, necessitating efficient implementations.
- Benchmarking and Evaluation: Establishing standardized benchmarks and metrics that effectively evaluate the unique capabilities of neurosymbolic systems (e.g., generalization across domains, interpretability, robustness to noise) is crucial for driving progress in the field.

Future Directions

The field of neurosymbolic AI for generalized planning in robotics is ripe for further innovation:

1. Deep Integration and Joint Learning: Developing more tightly integrated, end-to-end differentiable neurosymbolic architectures that allow for joint optimization of both neural and symbolic components. This could involve novel neural modules that directly learn symbolic operations or logic-constrained neural networks.
2. Adaptive Knowledge Representation: Researching methods for robots to autonomously learn, refine, and adapt their symbolic knowledge bases based on their experiences in the world, moving beyond static, pre-programmed knowledge.
3. Human-Robot Interaction via Neurosymbolic AI: Further exploring how neurosymbolic AI can facilitate more natural and effective human-robot collaboration, allowing humans to provide high-level symbolic instructions and receive interpretable explanations from robots.
4. Uncertainty Quantification and Robustness: Incorporating robust mechanisms for dealing with uncertainty at both the perceptual and symbolic levels, enabling robots to reason and plan effectively in highly uncertain real-world environments.
5. Long-Term and Hierarchical Planning: Developing neurosymbolic frameworks that can support truly long-term, hierarchical planning over extended time horizons, breaking down complex goals into manageable sub-goals across multiple levels of abstraction.
6. Application to Complex Real-World Domains: Applying these advanced neurosymbolic systems to increasingly complex real-world robotic applications, such as autonomous manufacturing, disaster response, and personalized healthcare, where robust generalized planning is indispensable.

CONCLUSION

Neurosymbolic AI represents a critical frontier for achieving robust generalized planning in AI-powered robotics. By strategically integrating the strengths of deep learning for perception and low-level control with the logical rigor of symbolic reasoning for high-level planning and abstract understanding, these hybrid frameworks are overcoming the inherent limitations of isolated AI paradigms. The demonstrated capabilities in enhanced generalization, improved interpretability, and more efficient learning underscore the transformative potential of this field. While challenges related to integration complexity, scalability, and automated knowledge acquisition persist, the vibrant research landscape is actively addressing these issues. The continued evolution of neurosymbolic robotics promises to usher in an era of more intelligent, autonomous, and transparent robots, capable of navigating and performing complex tasks in the dynamic and unpredictable environments of the real world.

REFERENCES

[1] Capitanelli, A., & Mastrogiovanni, F. (2023). A framework for neurosymbolic robot action planning using large language models. arXiv preprint arxiv.org, 9.
[2] Silver, T., Liang, Y., Weller, H., et al. (2025). VisualPredicator: Learning abstract world models with neuro-symbolic predicates for robot planning. ICLR Spotlight openreview.net.
[3] Ugur, E., Ahmetoglu, A., Nagai, Y., et al. (2025). Neuro-Symbolic Robotics: An overview of architectures integrating neural networks and symbolic systems. Submitted/recent arxiv.org, 15.
[4] Manhaeve, R., Dumančić, S., Kimmig, A., Demeester, T., & De Raedt, L. (2018). DeepProbLog: neural probabilistic logic programming. NeurIPS researchgate.net, 1.
[5] Silver, T., & colleagues. (2023). Learning Neuro-Symbolic Skills for Bilevel Planning. Proceedings of Machine Learning Research smythos.com, 15.
[6] Anonymous Authors. (2023). Neurosymbolic reinforcement learning and planning: A survey. arXiv preprint openreview.net, 3.
[7] Anonymous Authors. (2021). A survey on neural-symbolic learning systems. ScienceDirect sciencedirect.com.
[8] Kwon, M., Kim, Y., & Kim, Y. J. (2024). Fast and accurate task planning using neuro-symbolic language models and multi-level goal decomposition. arXiv preprint arxiv.org, 1.

[9] English, W., Simon, D., Jha, S., & Ewetz, R. (2024). NSP: A neuro-symbolic natural language navigational planner. arXiv preprint arxiv.org, 13.

[10] Anonymous Authors. (2023). A case for neuro-symbolic AI in constrained task planning. Proceedings of ACM SACMAT '23 frontiersin.org, 7.