



Politecnico
di Torino

ScuDo

Scuola di Dottorato - Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Artificial Intelligence for Industry 4.0 (37th cycle)

Neuro-Symbolic Integration in Artificial Intelligence and its Applications

By

Alice Bizzarri

Supervisor(s):

Prof. Evelina Lamma, Supervisor

Prof. Fabrizio Riguzzi, Co-Supervisor

Doctoral Examination Committee:

Prof. Luca Oneto, Referee, University of Genova

Prof. Felix Weitkaemper, Referee, LMU Munich

Politecnico di Torino

2025

Abstract

In recent years, artificial intelligence (AI) has gained an increasingly central role in many fields, including medicine, engineering, and finance. Today, AI is deeply embedded in our daily lives, influencing areas such as recommendation systems, smart agriculture, Industry 4.0, healthcare, and autonomous driving. With growing interest and investments in this field, the performance and effectiveness of AI systems have rapidly evolved over the past few decades.

Currently, AI is primarily structured into two main approaches: symbolic AI and neural AI. Symbolic AI is based on explicit rules and logic for problem-solving, while neural AI, which is built on deep neural networks, learns from data and generalizes complex behaviors. Both paradigms have led to significant advancements, but they also exhibit inherent limitations. Neural systems, while highly effective in terms of performance, are often difficult to interpret and require significant computational power. In contrast, symbolic models are more transparent and interpretable but tend to be less flexible when dealing with new or unforeseen situations.

A potential solution to overcome these limitations lies in the integration of symbolic AI and neural AI, a synergy that can combine the accuracy and learning capabilities of neural models with the transparency and robustness of symbolic systems. This hybrid approach, known as *neurosymbolic AI*, represents a promising avenue for enhancing the explainability, computational efficiency, and overall robustness of AI systems, making them more reliable and versatile across a wide range of applications.

This thesis explores the potential of neurosymbolic integration, focusing on the importance of combining the learning capabilities of deep neural networks with the explicit and interpretable reasoning characteristic of symbolic AI. The primary objective is to develop hybrid models that ensure not only high performance but also interpretability, efficiency, and robustness, with particular attention to critical

domains such as healthcare and cybersecurity, where decision transparency is of paramount importance.

Throughout the research, various types of integration were explored using neurosymbolic frameworks, and tailored solutions were developed. Specifically, neurosymbolic ensemble learning systems were created, capable of combining a symbolic subsystem with a neural subsystem through the use of explainable mechanisms. Furthermore, the integration of constraints during the training process was analyzed, with the aim of creating more precise and robust anomaly detection systems. A hybrid system for the "open set recognition" scenario was also developed, combining neural and symbolic components. In addition to technical aspects, the thesis proposes new benchmarks for the evaluation of neurosymbolic systems, filling a gap due to the lack of shared standards in the field.

Practical applications of the developed systems were discussed, with a detailed presentation of the obtained results. Despite the promising results, the thesis highlights some ongoing challenges, such as the computational complexity of hybrid models and the difficulties in integrating deep learning with symbolic reasoning. Another critical aspect concerns the applicability of these models to unexplored domains, such as robotics or autonomous systems, which require further study.

In conclusion, neurosymbolic integration represents an important step towards the development of AI systems that are more transparent, reliable, and capable of tackling complex challenges in critical contexts. The research presented in this thesis lays the foundation for future developments in AI, aiming to make these systems increasingly efficient and interpretable, while simultaneously improving their adaptability across different application domains.