Report from Dagstuhl Seminar 25452

A Roadmap Towards Practical Applications of Neurosymbolic Learning and Reasoning

Thiviyan Thanapalasingam*[1], Annette ten Teije*[2], and Frank van Harmelen[3]

1 **University of Amsterdam, NL.** thiviyan.t@gmail.com
2 **VU Amsterdam, NL.** annette.ten.teije@vu.nl
3 **VU Amsterdam, NL.** frank.van.harmelen@vu.nl

Abstract

Neurosymbolic Artificial Intelligence (NeSy) promises to combine the scalability and adaptability of neural networks with the precision and interpretability of logic-based reasoning systems. Despite significant theoretical advances and growing interest in the field, the development of NeSy systems that seamlessly integrate symbolic and sub-symbolic elements at scale for real-world practical applications remains an open challenge. This Dagstuhl Seminar brought together leading researchers from diverse fields including machine learning, logic, knowledge representation, knowledge discovery, robotics, healthcare, and natural sciences to chart a roadmap for advancing NeSy systems from synthetic benchmarks to practical deployment. The seminar addressed four key challenges: (I) developing realistic benchmarks and standardized evaluation frameworks for NeSy systems, (II) identifying and advancing applications in robotics, healthcare, and scientific discovery, (III) creating open-source libraries to lower entry barriers and accelerate development, and (IV) creating accessible educational materials to broaden participation in the field. Through structured discussions, breakout sessions, and collaborative planning, participants developed concrete action plans and deliverables, including a planned Communications of the ACM article, open-source library development (ULLER, DeepLog), educational resources, and improvements to the field's Wikipedia entry. This report documents the seminar's presentations, discussions, and findings, and presents strategic recommendations for realizing the promise of Neurosymbolic AI in real-world applications.

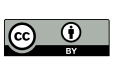
Seminar November 2–7, 2025 – https://www.dagstuhl.de/25452

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Computing methodologies → Symbolic and algebraic algorithms; Theory of computation → Logic

Keywords and phrases Neurosymbolic Artificial Intelligence, Machine Learning, Logic, Knowledge Representation & Reasoning, Sub-symbolic & Symbolic Integration

Digital Object Identifier 10.4230/DagRep.15.11.67

* Editor / Organizer



A Roadmap Towards Practical Applications of Neurosymbolic Learning and Reasoning, *Dagstuhl Reports*, Vol. 15, Issue 11, pp. 67–86
Editors: Thiviyan Thanapalasingam, Annette ten Teije, and Frank van Harmelen

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Thiviyan Thanapalasingam (University of Amsterdam, NL, thiviyan.t@gmail.com)
Annette ten Teije (VU Amsterdam, NL, annette.ten.teije@vu.nl)
Frank van Harmelen (VU Amsterdam, NL, frank.van.harmelen@vu.nl)

This summary presents the outcomes of our Dagstuhl Seminar on "A Roadmap Towards Practical Applications of Neurosymbolic Learning and Reasoning" (25452). The seminar brought together 23 researchers from diverse fields to address the critical challenge of transitioning Neurosymbolic AI (NeSy) systems from synthetic benchmarks to real-world applications. The seminar focused on:

- Identifying key challenges that impede the practical deployment of NeSy systems,
- Developing concrete action plans and collaborative structures to address these challenges,
- Establishing deliverables including educational resources, open-source libraries, and an article in a high-impact journal.

As a result of the seminar, the following challenges have been identified and addressed through structured discussions and working groups:

1. **Benchmarks and Evaluation:** The field's current reliance on synthetic problems (e.g., MNIST Addition, MNIST Sudoku) fails to capture real-world complexity. Participants developed plans for creating realistic benchmark suites that incorporate incomplete rule sets, scalability requirements, and standardized evaluation metrics beyond traditional machine learning measures, including reasoning correctness, rule violation detection, and interpretability assessment.
2. **Domain Applications:** Three priority application areas were identified (robotics, natural sciences, and healthcare) where NeSy systems can provide substantial value. Working groups outlined concrete use cases while identifying specific technical challenges and highlighting existing applications of NeSy systems from other AI subfields.
3. **Open-Source Infrastructure:** Following the success of tools like PyTorch Geometric in accelerating deep learning adoption, participants discussed with the creators of Pylon and ULLER strategies to lower entry barriers for both researchers and practitioners.
4. **Educational Resources:** To broaden participation and accelerate field growth, participants committed to writing a comprehensive handbook/textbook on NeSy foundations, using machine learning as a foundation, and targeting graduate students and researchers from adjacent fields.

The seminar produced several concrete outcomes and commitments for future collaboration. Participants agreed to co-author a CACM article to raise awareness across the broader scientific community, initiate development of open-source Python libraries for NeSy methods, create a comprehensive benchmark suite with standardized evaluation protocols, and update the Wikipedia entry for Neurosymbolic AI to improve public-facing documentation of the field.

Plenary discussions on *"What would be a good textbook for teaching Neurosymbolic Learning and Reasoning?"* and *"What is the role of Neurosymbolic AI in the era of foundational models?"* produced important insights regarding how large language models both motivate NeSy research by highlighting reasoning limitations and provide new opportunities through neuro-symbolic integration with foundation models.

The seminar's roadmap positions the NeSy community to address pressing challenges in AI by focusing on practical applications and building supporting infrastructure.

2 Table of Contents

3 Overview of Talks

3.1 Neuro-Symbolic AI in HealthCare: Integrating Learning and Reasoning for Trustworthy Healthcare AI

Maria-Esther Vidal (TIB Leibniz-Informationszentrum – Hannover, DE & Leibniz Universität Hannover, DE, vidal@l3s.de)

Artificial Intelligence is rapidly transforming healthcare, yet its integration into clinical practice remains constrained by concerns of bias, opacity, and lack of accountability. The FUTURE-AI framework defines six guiding principles (Fairness, Universality, Traceability, Usability, Robustness, and Explainability) as prerequisites for trustworthy medical AI. However, these principles are still largely conceptual; they require technical mechanisms that can make them actionable and verifiable within AI pipelines.

This talk examines how Neuro-Symbolic AI (NeSy), which combines neural prediction and symbolic reasoning, can address these challenges by integrating semantics, logic, and ethics directly into the computational core of healthcare AI systems. NeSy unites the statistical power of neural models with the interpretability and formal rigor of symbolic representations. Through this integration, fairness can be achieved by tracing and reasoning over bias; universality and robustness can arise from semantic consistency and invariant knowledge; and traceability and explainability can be realised through explicit reasoning chains and formal provenance models.

The presentation illustrates these ideas through two case studies: (1) ontology-based bias documentation and reasoning in machine learning pipelines, and (2) explainable prediction in lung-cancer knowledge graphs. Both examples demonstrate how symbolic reasoning complements data-driven learning to improve interpretability, reliability, and trust.

The talk concludes by identifying open research directions, formalisms for specification and validation, scalable multimodal reasoning, and human-centred explainability, toward the realisation of ethically grounded, transparent, and trustworthy Neuro-Symbolic AI for healthcare.

3.2 Neurosymbolic AI for Trustworthy Automation: From Datacenters to Agentic Systems

Byron Cook (Amazon Web Services – Seattle, US & University College London, GB, byron.cook@amazon.com)

We are witnessing a scientific revolution in computing, driven by the shift from microprocessors to datacenters as the dominant computational model. This transformation has enabled not only generative AI, through access to massive computational resources, but also fundamental advances in formal reasoning tools, which now achieve 10x performance improvements through distributed reasoning and lemma sharing.

This evolution presents a unique opportunity for neurosymbolic approaches: language models can discover inductive invariants, ranking functions, and rely/guarantee constraints, while mechanical theorem provers verify these artifacts with proof-producing SAT/SMT

solvers. The past decade has demonstrated that automated reasoning is no longer confined to hardware and safety-critical systems, cloud providers like AWS now apply it broadly across their infrastructure, from customer-facing tools (IAM Access Analyzer, S3 Block Public Access) to internal proof projects in cryptography, identity, and virtualization.

However, as customers adopt generative and agentic AI, they face critical challenges around safety, truthfulness, and trustworthiness. This talk presents Amazon's approach through several products: Automated Reasoning checks in Bedrock Guardrails, which combines pre-deployment requirements analysis with inference-time verification; Amazon Kiro, which helps build formal specifications for programs; and Strands and AgentCore, which apply compositional reasoning to agent safety and liveness properties using the Cedar permissions framework.

The presentation demonstrates how neurosymbolic techniques address key challenges, including redundant translation with semantic equivalence checking, iterative refinement, synthesis of axiomatic vocabularies, and specification composition. It explores how these ideas extend to a new style of GenAI-native development, where semantic guardrails in the form of proved properties prevent unwanted bugs in code generated by agentic developer tools.

Finally, the talk addresses the critical importance of soundness metrics, proposing that the community measure soundness in terms of "nines" (e.g., "five 9s" = 99.999% sound), similar to durability and availability contracts in IT systems. Metrics like "Valid@N" help track whether neurosymbolic systems can reliably return valid proofs rather than defaulting to "unknown" or "invalid." The talk concludes by discussing persistent proofs as a foundation for proof search, repair, and distributed lemma sharing in large-scale neurosymbolic systems.

3.3 Toward Derivable Scientific Discovery: Integrating Data, Theory, and the Future of Machine-Aided Discovery

Cristina Cornelio (Samsung AI – Cambridge, GB, c.cornelio@samsung.com)

Scientists aim to build models that not only fit observations but also explain them. Traditionally, theories were developed from domain knowledge and then tested against data, whereas modern machine learning instead extracts patterns directly from large datasets. However, discovering meaningful models within data remains an ongoing challenge.

This talk explores two methods that address this challenge by connecting empirical evidence with formal knowledge: AI-Descartes, which uses symbolic regression to propose candidate laws and logical reasoning to select those consistent with known axioms; and AI-Hilbert, which combines polynomial optimization with logical theories ensuring both theoretical consistency and empirical validation at the same time.

These methods point toward a new paradigm in which data and theory jointly shape scientific discovery, opening new questions and inviting further exploration of the emerging field of AI for science.

3.4 Tooling for Reproducible Neurosymbolic Research: ULLER

Emile van Krieken (VU Amsterdam, NL)

One of the major transformations behind computational research is the emergence of frictionless reproducibility. In Machine Learning research, executing a highly complicated new experiment is easy due to well-developed modern tools like PyTorch, HuggingFace Transformers, and Weights and Biases. What is missing for frictionless reproducibility in neurosymbolic AI? We see opportunities in common formats for sharing knowledge, and for well-developed standard algorithms for (neurosymbolic) reasoning. Therefore, we introduce two Python libraries in development. ULLER aims to represent knowledge in simple Python. DeepLog efficiently implements building blocks for reasoning in neurosymbolic models on the GPU. These libraries are a push towards standard tooling in neurosymbolic research.

3.5 DeepLog: A Theoretical and Operational Framework for Neurosymbolic AI

Robin Manhaeve (KU Leuven, BE)

We present a theoretical and operational framework for neurosymbolic AI called DeepLog. DeepLog introduces building blocks and primitives for neurosymbolic AI that make abstraction of commonly used representations and computational mechanisms used in neurosymbolic AI. DeepLog can represent and emulate a wide range of neurosymbolic systems. DeepLog is implemented in software, relies on the latest insights in implementing algebraic circuits on GPUs, and is declarative in that it is easy to obtain different neurosymbolic models by making different choices for the underlying algebraic structures and logics. The generality and efficiency of the DeepLog neurosymbolic machine is demonstrated experimentally.

3.6 Benchmarking Neurosymbolic AI

Maarten C. Stol (BrainCreators – Amsterdam, NL & VU Amsterdam, NL)

Neurosymbolic (NeSy) benchmarking lacks systematic desiderata derived from the separate benchmarking cultures of Machine Learning (ML) and Automated Reasoning (AR). A combination of current practices is insufficient, as e.g., ML benchmarking often struggles with "construct validity": overly general claims from highly specific data and tasks. While the nature of NeSy construct validity is an open question, we claim NeSy is now in a position to inherit the succeses, and overcome the flaws of both fields.

We advocate a shift from static benchmarks and incremental SOTA improvements towards understanding how systems work and fail. We propose benchmark generators based on controllable notions of problem instance hardness, inheriting AR notions (phase transitions)

and ML notions (label noise, OOD data). Such generators enable analysis of how tuning hardness can reverse SOTA rankings and explain performance rather than merely measure it. Moreover, utilizing multiple metrics allows us to then analyze performance trade-offs and trade-off gradients. The resulting synthetic data serves not to replace real-world data, but to analyze and categorize its inherent hardness aspects. We demonstrate this approach with two use cases (maritime surveillance, infrastructure maintenance) involving geo-spatial data and object relations. The potential for NeSy in these use cases is high due to an abundance of background knowledge available to support their AR and ML components.

3.7 Neuro-Symbolic AI and Robotics: Bridging Perception and Sensemaking

Agnese Chiatti (Polytechnic University of Milan, IT)

Cornerstone objectives and aspirations in NeSy AI and Robotics overlap in many ways. In Robotics, the demand for world models capable of bridging the gap between sub-symbolic perceptual data and symbolic representations has led to the development of semantic maps. Semantic maps are a paradigmatic example of Neuro-Symbolic representation, as they integrate the outputs of both sub-symbolic and symbolic processing modules.

Establishing a deeper synergy between NeSy AI and Robotics holds significant promise. On the one hand, NeSy approaches can mitigate challenges of verifiability and safety that arise in the use of purely sub-symbolic learning and reasoning techniques, thus enabling reliable and trustworthy robot deployment in real-world contexts. On the other hand, robots, conceived as embodied agents, offer a natural platform for advancing NeSy research by supporting a bidirectional feedback process: (i) incorporating top-down knowledge and constraints to guide reasoning and monitoring, and (ii) discovering new symbolic knowledge through active, bottom-up exploration of the environment. However, the divergence in terminology, methodologies, and benchmarking practices between the NeSy AI and Robotics communities continues to hinder progress at their intersection.

In this talk, I will advocate for drawing insights from both fields, from the perspective of robotic perception and sensemaking. I will discuss use cases in health and safety monitoring, assisted living, and general-purpose visual reasoning. Then, I will present an overview of current research trends as a basis for broader discussion.

3.8 Neurosymbolic AI Education: Integrating Learning and Reasoning

Luís C. Lamb (CatholicTech – Cambridge, US & UFRGS – Porto Alegre, BR)

In this talk, I outline the contents of a proposed graduate-level course on Neurosymbolic Artificial Intelligence. The structure of the course and the selection of topics are designed in alignment with the educational objectives and the intended learning outcomes. The lectures and accompanying materials provide contextual background, conceptual foundations, and applied perspectives. The proposed syllabus includes the following topics, to be covered

over a semester: Introduction; History and Evolution of Computer Science and Artificial Intelligence; Basic concepts: symbolic reasoning and connectionist learning; Integrating Reasoning and Learning; Neurosymbolic AI approaches, tools, and applications; Perspectives of the field. Keywords: Neurosymbolic AI; Education; Learning; Reasoning.

4 Working Groups

4.1 Benchmarking Neurosymbolic AI (Working Group 1)

Robert Peharz (TU Graz, AT)
Antonio Vergari (University of Edinburgh, GB)
Mathias Niepert (Universität Stuttgart, DE)
Guy Van den Broeck (UCLA, US)
Benjie Wang (UCLA, US)
Emile van Krieken (VU Amsterdam, NL)
Annette ten Teije (VU Amsterdam, NL)

In our break-out benchmark session, we began by clarifying what neurosymbolic (NeSy) systems are and why they matter, which naturally led into a deeper discussion of their goals, desiderata, and underlying premises. Several key promises of NeSy systems emerged: (1) data efficiency, both during training and inference; (2) improved generalization and transferability under distribution shifts while maintaining symbolic constraints; (3) controllability, including safety guarantees and reliability; and (4) modularity at the system level, enabling clearer communication and interfacing between components. Building on this discussion, we sketched a small ontology for the space of neurosymbolic approaches. We closed by examining current successes that fit the NeSy paradigm, even when they are not explicitly labeled as such, such as LLMs combined with tool-use and verifiers, and systems like AlphaGeometry. Overall, we concluded that what makes neurosymbolic systems distinctive is the combination of learning, reasoning, interfacing, and guarantees.

4.2 Benchmarking Neurosymbolic AI (Working Group 2)

Alberto Speranzon (Lockheed Martin Advanced Technology Labs – Eagan, US)
Lia Morra (Polytechnic University of Turin, IT)
Maria-Esther Vidal (TIB Leibniz-Informationszentrum – Hannover, DE & Leibniz Universität Hannover, DE)
Michael Cochez (ELLIS Institute – Espoo, FI & Åbo Akademi University – Turku, FI)

Benchmarks in neurosymbolic AI should not be understood merely as tools for comparing different approaches; they also shape the direction of entire research communities. A well-designed benchmark provides clear targets to work toward and helps reveal the boundaries

and limitations of current state-of-the-art systems. This requires identifying the essential properties of a good AI benchmark: properties that hold independently of specific performance metrics and that may be inspired by the standards of other scientific fields [1, 2]. Because neurosymbolic AI systems, especially in domains like healthcare, must satisfy multidimensional objectives, benchmarks must be able to evaluate these different dimensions. For instance, the FUTURE-AI principles (Fair, Universal, Traceable, Usable, Robust, Explainable) [3] illustrate the variety of criteria that a single benchmark may need to assess.

To be genuinely useful, benchmarks should be prescriptive: they must specify not only which aspects of performance to measure, but also how these measurements should be made, interpreted, and reported. They should explicitly surface the trade-offs inherent in different solutions, enabling researchers to understand how improving one property may hinder another.

A major risk in neurosymbolic AI is the development of benchmarks that appear favourable to NeSy approaches while failing to reflect the real complexities and requirements of the target domain. Avoiding such "bad" benchmarks is essential to ensuring that progress is meaningful and aligned with real-world needs.

References

1 Guideline for reporting Experimental Results in Life Sciences. `https://www.nature.com/documents/nr-reporting-life-sciences-research.pdf`
2 Guideline for reporting Experimental Results in Physical Sciences. `https://www.nature.com/documents/Physical_sciences_reproducibility.pdf`
3 Lekadir K, Frangi A F, Porras A R, Glocker B, Cintas C, Langlotz C P et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare BMJ 2025; 388 :e081554 doi:10.1136/bmj-2024-081554

4.3 Benchmarking Neurosymbolic AI (Working Group 3)

Luis Lamb (CatholicTech – Cambridge, US & UFRGS – Porto Alegre, BR)
Guy Van den Broeck (UCLA, US)
Sebastijan Dumancic (TU Delft, NL)
Thiviyan Thanapalasingam (University of Amsterdam, NL)
Cristina Cornelio (Samsung AI – Cambridge, GB)
Maarten C. Stol (BrainCreators & VU Amsterdam, NL)
Frank van Harmelen (VU Amsterdam, NL)

This working group identified three key lessons for benchmark design, outlined in this section. A central idea in neurosymbolic AI is that a little logic goes a long way. Many real-world problems do not require full-blown symbolic reasoning or heavy domain theories; instead, lightweight logical constraints can already provide meaningful structure and improve learning outcomes. This perspective encourages focusing on simple forms of domain knowledge and considering applications where only limited, easily articulated knowledge is needed. Industry examples point in this direction as well: systems such as DeepMind's structured reasoning models or OpenAI's JSON-constrained generation show how modest symbolic scaffolding can guide powerful neural models. Exploring lighter logical systems that can stand in for prompt engineering or complement it, may reveal a productive balance between expressivity and

practicality. Finding this "sweet spot" of expressivity is crucial. Lightweight knowledge can make models more data-efficient, more reliable, and more controllable, without overburdening them with complex logical formalisms. This aligns with the broader idea that intelligent systems must be educable: as Valiant [1] argues, humans' unique strength lies in our ability to acquire and use small, structured pieces of knowledge. Neurosymbolic systems should strive for a similar balance.

Lesson for benchmarking design. avoid tasks that are overly logic-heavy and instead design evaluation setups that test how effectively a system can use minimal but meaningful knowledge. Potential paths forward include creating new knowledge-centric datasets from scratch such as ARC or augmenting existing datasets with structured knowledge, using as little additional data as possible while still leveraging symbolic insights.

Industry's main promise in adopting neurosymbolic AI lies in its potential to deliver safety, reliability, and transparency qualities that become essential in sectors where failure tolerance is extremely low and data is often scarce. These needs are not limited to big tech; industries such as aerospace, manufacturing, finance, and cybersecurity all operate in environments where mistakes are costly and explanations matter. Neurosymbolic approaches can add value precisely because they offer structured guarantees, improved data efficiency, and stronger out-of-distribution robustness critical features given that most real-world prediction tasks fall outside the training distribution and most industrial datasets are small. At the same time, documenting how knowledge is obtained and incorporated, as seen in cybersecurity workflows, becomes a practical requirement for trust and auditability. Instead of investing months into refining prompts, organisations want principled, knowledge-driven methods that scale reliably.

Lesson for benchmark design. include small-data scenarios, OOD tasks, and conditions reflecting fault intolerance to meaningfully test systems intended for industrial deployment.

Neurosymbolic AI should broaden its view beyond purely logical formalisms and consider a wider variety of symbolic representations, including state machines, knowledge graphs, JSON schemas, grammars, and domain-specific languages such as PSL or PDDL. Current NeSy research often overemphasizes declarative logic, even though many practical systems like Python-based code generators or structured planners, already use symbolic constraints to reduce hallucinations and improve reliability. Likewise, perception systems can benefit from injecting world knowledge, such as object permanence or contextual relations, to clean or enrich scene graphs, and robotic navigation can be improved by adding structured knowledge on top of bounding-box-based vision. Instead of searching for entirely new NeSy applications, it may be more productive to highlight the neurosymbolic components that already underpin existing ones, such as code generation or multimodal reasoning in VLMs.

Lesson for benchmark design. avoiding constraints that require logic as the only declarative language and instead standardizing input formats for both data and knowledge, independent of their expressivity.

References

1 The Importance of Being Educable: A New Theory of Human Uniqueness, Leslie Valiant, ISBN: 9780691230566.

4.4 Neuro-Symbolic AI in Health Care

Robert Peharz (TU Graz, AT)
Maria-Esther Vidal (TIB Leibniz-Informationszentrum – Hannover, DE & Leibniz Universität Hannover, DE)
Frank van Harmelen (VU Amsterdam, NL)
Annette ten Teije (VU Amsterdam, NL)
Cristina Cornelio (Samsung AI – Cambridge, GB)
Agnese Chiatti (Polytechnic University of Milan, IT)

The FUTURE-AI framework (Fairness, Universality, Traceability, Usability, Robustness, and Explainability) [1] provides a conceptual foundation for trustworthy medical AI, but its principles require concrete technical mechanisms to become actionable. Our discussion focused on what these dimensions mean in the context of neurosymbolic (NeSy) systems, and whether such systems may naturally perform better on some aspects of the framework.

We explored each FUTURE dimension through a NeSy lens. For **Fairness**, NeSy systems offer tools such as counterfactual reasoning and causal interventions between features, as well as ontology-informed learning both of which enable more transparent bias detection and mitigation. Under **Universality**, their combination of symbolic structure and neural generalization supports improved transferability and out-of-distribution performance; symbolic rules further allow human inspection and causal validation. In terms of **Traceability**, the picture is mixed: production traceability (e.g., reproducibility of training pipelines) does not automatically improve, but execution traceability, being able to follow symbolic reasoning steps, can benefit substantially from NeSy architectures. For **Usability**, symbolic components introduce shared concepts and interfaces that may make human interaction with models more intuitive. **Robustness** and **Explainability** also stand to gain from explicit reasoning layers, structured constraints, and semantic consistency. Overall, the discussion highlighted that while NeSy systems do not guarantee trustworthiness, they may provide promising pathways for operationalizing several dimensions of the FUTURE-AI framework.

References

1 FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare, FUTURE-AI consortium BMJ 2025; 388 doi: https://doi.org/10.1136/bmj-2024-081554

4.5 Architectures for Probabilistic Neurosymbolic AI

Robert Peharz (TU Graz, AT)
Antonio Vergari (University of Edinburgh, GB)
Mathias Niepert (Universität Stuttgart, DE)
Guy Van den Broeck (UCLA, US)
Benjie Wang (UCLA, US)
Emile van Krieken (VU Amsterdam, NL)
Luis C. Lamb (CatholicTech – Cambridge, US & UFRGS – Porto Alegre, BR)
Lia Morra (Polytechnic University of Turin, IT)
Michael Cochez (ELLIS Institute – Espoo, FI & Åbo Akademi University – Turku, FI)
Sebastijan Dumancic (TU Delft, NL)

This working group examined fundamental architectural components underlying probabilistic neurosymbolic systems, with a particular focus on the product of experts formulation. In this framework, one expert represents a probability distribution that encodes beliefs over world states, while the other encodes logical constraints that filter invalid states. Computing the weighted model count (or weighted model integral for continuous domains) by summing this product over all possible states forms the computational backbone of many neurosymbolic pipelines.

Knowledge compilation routines that translate logical constraints into circuit representations with specific structural properties enable closed-form computation of weighted model counts. The determinism property, while enabling tractable inference, restricts circuit expressiveness and may cause exponential blow-up in compiled circuit size. Relaxing determinism constraints could yield smaller circuits and more scalable pipelines. One promising direction, currently explored in work on steering LLM generation with automata, involves translating CNF formulas into partial DNF representations. Hybrid strategies combining these approaches represent an important research direction with broad applicability across neurosymbolic systems.

The discussion also addressed integration of circuit representations into large language models for constraint enforcement and generation acceleration. Training circuits to approximate LLM distributions was compared with approaches that embed circuits as final layers, offering greater expressiveness but potentially limiting constraint enforcement to bounded token windows. Finally, analogies to constrained generation in diffusion models were explored, noting that classifier-guided approaches employ product-of-expert formulations for score matching while sidestepping exact partition function computation. Combining these directions with tractable circuit models presents opportunities for novel hybrid neurosymbolic architectures.

4.6 Neurosymbolic Methods for Safe and Reliable Robotic Systems

Guy Van den Broeck (UCLA, US)
Alberto Speranzon (Lockheed Martin Advanced Technology Labs – Eagan, US)
Agnese Chiatti (Polytechnic University of Milan, IT)
Annette ten Teije (VU Amsterdam, NL)
Cristina Cornelio (Samsung AI – Cambridge, GB)
Thiviyan Thanapalasingam (University of Amsterdam, NL)
Sebastijan Dumancic (TU Delft, NL)

This working group examined how neurosymbolic methods can improve the reliability and safety of robotic systems. A central observation was that LLMs perform impressively for hierarchical planning, until they do not. Current approaches use symbolic verifiers to validate LLM-generated plans, feeding errors back into the prompt for iterative refinement, an approach sometimes characterised as "prompt and pray". While this works in many scenarios, LLMs fundamentally lack the safety guarantees required for real-world deployment, where even controlled environments can produce unexpected uncertainty.

Neurosymbolic methods offer principled ways to handle such ambiguity. Three aspects of assurance were identified as critical: system design, concerning how to architect systems with built-in safety properties; runtime monitoring, using techniques such as signal temporal logic to detect and respond to violations during execution; and task evaluation, measuring success in completing objectives. Additionally, small-data requirements in robotics applications align well with the data efficiency that neurosymbolic systems naturally provide.

Recent benchmarking efforts such as the Embodied Agent Interface [1] offer standardised evaluation for LLM-based decision making, and work on Bayesian-conditioned controllable generation has shown success rate improvements from 20% to 40% by encoding constraints, demonstrating that neurosymbolic approaches yield measurable benefits. A key insight was that rather than imposing all possible constraints, selectively adding those with high failure probability proves more effective, raising the possibility of learning which constraints matter most.

The group also noted that edge computing demands smaller models, and neurosymbolic constraints can help achieve compactness without sacrificing performance. Finally, connections between vision-language models and world models were discussed as both containing neurosymbolic components, and the division between robotics and AI venues was noted as a barrier to knowledge sharing across communities.

References

1 Li, Manling and Zhao, Shiyu and Wang, Qineng and Wang, Kangrui and Zhou, Yu and Srivastava, Sanjana and Gokmen, Cem and Lee, Tony and Li, Li Erran and Zhang, Ruohan and Liu, Weiyu and Liang, Percy and Fei-Fei, Li and Mao, Jiayuan and Wu, Jiajun. "Embodied Agent Interface: Benchmarking LLMs for Embodied Decision Making". In *Advances in Neural Information Processing Systems 37 (NeurIPS 2024), Datasets and Benchmarks Track* (Oral), 2024.

4.7 Updating the Wikipedia Entry for Neurosymbolic AI

Emile van Krieken (VU Amsterdam, NL)
Robin Manhaeve (KU Leuven, BE)
Luis C. Lamb (CatholicTech – Cambridge, US & UFRGS – Porto Alegre, BR)

This working group examined the Wikipedia entry for Neurosymbolic AI[1] and found it incomplete, with an unclear definition, limited historical coverage, missing connections to adjacent fields, and minimal discussion of practical applications. The page also lacks adequate treatment of the field's relationship to ethical AI, explainable AI, safe AI, and trustworthy AI research.

Inspired by the well-organized Machine Learning Wikipedia page, the group developed a comprehensive restructuring plan with the following proposed additions and revisions:

- Rewriting the introduction and definition for clarity and completeness
- Expanding the history section with improved coverage of the field's development
- Adding a new section on relationships to other fields, including machine learning and deep learning, knowledge representation and reasoning, theorem proving, cognitive science, ethical and responsible AI, explainable AI, safe AI, trustworthy AI, databases and software engineering, and artificial general intelligence
- Creating a new applications section documenting real-world deployments
- Revising the research challenges section to reflect current open problems
- Updating the implementations and software section with recent developments

Following the seminar, participants committed to collaboratively editing the entry, aiming to provide the broader scientific community and general public with an accurate and up-to-date resource on Neurosymbolic AI.

5 Plenary Discussions

5.1 Tooling for Neurosymbolic AI

This plenary discussion focused on the infrastructure needed to achieve "frictionless reproducibility" in neurosymbolic AI research. In machine learning, executing complex experiments has become straightforward thanks to well-developed tools such as PyTorch and PyTorch Geometric [1]. The NeSy field lacks comparable infrastructure, which hinders both newcomers and experienced researchers. Two complementary Python libraries under development were presented as steps toward addressing this gap. ULLER (Unified Language for LEarning and Reasoning) [2] provides a Python domain-specific language for expressing background knowledge, with compilers targeting different NeSy system formats. ULLER decouples the syntax of knowledge representation from the semantics provided by specific NeSy systems, supporting classical first-order logic, fuzzy logic, and probabilistic semantics. Users can define constraints once and then invoke different NeSy systems with minimal code changes. DeepLog [3] operates at a lower level of abstraction, providing a neurosymbolic machine consisting of two components: a language for specifying neurosymbolic models based on

[1] https://en.wikipedia.org/wiki/Neuro-symbolic_AI

annotated grounded first-order logic, and extended algebraic circuits as computational graphs. DeepLog efficiently implements reasoning primitives on the GPU and can emulate a wide range of existing neurosymbolic systems including DeepProbLog, NeurASP, and Semantic Loss. Together, ULLER addresses the first-order knowledge level while DeepLog handles the propositional and inference levels. The discussion also examined Pylon [4], an earlier NeSy framework that struggled to gain community adoption despite its capabilities. This raised concerns about the risk of repeatedly reinventing tools that fail to achieve critical mass. Drawing parallels to the challenges faced by Semantic Web technologies, the group noted that even well-designed tooling requires sustained effort to build community uptake. Standardised input formats for both data and knowledge, independent of their logical expressivity, were identified as a key enabler for broader NeSy adoption and for avoiding fragmentation across competing frameworks.

References

1 Fey, Matthias and Lenssen, Jan Eric. "Fast Graph Representation Learning with PyTorch Geometric." In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

2 van Krieken, Emile and Badreddine, Samy and Manhaeve, Robin and Giunchiglia, Eleonora. "ULLER: A Unified Language for Learning and Reasoning." In *Proceedings of the 18th International Conference on Neural-Symbolic Learning and Reasoning (NeSy 2024)*, pp. 219–239, 2024.

3 Derkinderen, Vincent and Manhaeve, Robin and Adriaensen, Rik and Van Praet, Lucas and De Smet, Lennert and Marra, Giuseppe and De Raedt, Luc. "The DeepLog Neurosymbolic Machine." *arXiv preprint arXiv:2508.13697*, 2025.

4 Ahmed, Kareem and Teso, Stefano and Chang, Kai-Wei and Van den Broeck, Guy and Vergari, Antonio. "Pylon: A PyTorch Framework for Learning with Constraints." In *NeurIPS 2022 Workshop on Neuro Causal and Symbolic AI (nCSI)*, 2022.

5.2 Bridging NeSy and LLMs

In the plenary session on *"What is the role of Neurosymbolic AI in the era of foundational models?"*, the discussion centered on emerging patterns in how neurosymbolic (NeSy) methods can enhance and structure the use of large language models. Several concrete integration patterns were identified. LLMs can assist in knowledge extraction and construction, transforming unstructured text into symbolic or semi-formal representations. They can also be pushed toward formal reasoning, for example through constrained prompting or by delegating certain reasoning steps to external symbolic engines. Another pattern involves feeding symbolic feedback back into the LLM, either dynamically during generation or as a post-processing verification layer, helping to rectify inconsistencies. Finally, symbolic systems can be used to generate high-quality training or fine-tuning data, enabling LLMs to inherit structured reasoning capabilities they cannot easily learn from raw text.

A key takeaway from the session was the mutually reinforcing relationship between LLMs and symbolic methods. LLMs can dramatically reduce the cost of constructing and maintaining formal representations by automating parts of the knowledge engineering pipeline. Conversely, symbolic constraints, logic, and verifiable representations help reduce hallucinations and improve reliability in LLM outputs. Looking ahead, the group envisioned a future in which elements such as logic, causality, and other structured reasoning formalisms become integrated directly into model training regimes, rather than remaining purely external modules. This deeper fusion could lead to more robust, interpretable, and systematically grounded AI systems that retain the flexibility of foundation models while gaining the trustworthiness and rigor of symbolic reasoning.

5.3 Education Materials for NeSy AI

In the plenary discussion on educational materials for neurosymbolic AI, several overarching themes emerged. The consensus was that the primary target audience should be advanced master's students or early-stage PhD students. A key tension discussed was the balance between practical, hands-on learning and deeper research-oriented understanding: should the material focus on using neurosymbolic tools, or on developing the theoretical foundations behind them? The group agreed that a modern resource should avoid the static, paper-collection style of traditional textbooks and instead adopt an agile, update-friendly format. Given the increasing momentum in neurosymbolic AI, many felt that now is an ideal time to develop a comprehensive teaching resource or textbook.

Prerequisites were another central point: students should enter with a working knowledge of logic and formal representation (potentially including exposure to symbolic programming), machine learning fundamentals, and a clear vocabulary that bridges terminology across ML and knowledge representation (KR). This led to multiple possible design directions for structuring the material. One idea is framed as "advanced ML", is to start from familiar ML concepts and progressively extend them with symbolic reasoning, explicitly motivating each addition through ML's shortcomings. Another approach begins with the high-level "why learning and reasoning?", incorporating Kautz's taxonomy, examples of existing systems, and industrial case studies. A more foundational option would foreground probability, Boolean circuits, and the expressivity of neural networks and logics, culminating in a taxonomy of neurosymbolic system classes. Yet another design path would contrast ML and KR directly: starting from the weaknesses of each and showing how the two fields complement and strengthen one another. A narrative "product-development" storyline was also proposed, modeling the material as the evolution of increasingly capable NeSy systems.

The discussion also highlighted concrete tools suitable for teaching, such as PyNeuralog, ULLER, and SKI-Lang (to be introduced at ECAI 2025). Additional topics worth integrating include inductive programming, the intellectual history of symbolic vs. connectionist AI ("the two AI religions"), and broader historical reflections inspired by Kuhnian paradigm shifts. Luis Lamb existing course structure on Neurosymbolic AI Education was referenced as a strong template, spanning introductions, historical context, symbolic vs. connectionist basics, motivations for integration, tools and applications, and forward-looking perspectives. Finally, the group explored different strategies for generating the teaching materials. Options ranged from writing an entirely new textbook, to producing a Version 0 using generative AI and iterating, to assembling handouts and slides that later evolve into a book. Alternatives included extending an existing text, collaborating with current authors, or producing lecture videos as the first step. Several existing books were reviewed [1, 2, 3] with the consensus that none fully addresses both the conceptual depth and modern research landscape needed for a master- or PhD-level neurosymbolic AI course. An online course was also considered as a complementary or preliminary step. Overall, the discussion reflected a shared recognition that neurosymbolic AI education is both timely and ripe for a coherent, rigorous, and modern learning resource.

References

1 Dingli, A., & Farrugia, D. (2023). Neuro-Symbolic AI: Design transparent and trustworthy systems that understand the world as you do. Birmingham & Mumbai: Packt Publishing. ISBN 978-1-80461-762-5. O'Reilly Media

2 Shakarian, P., Simari, G. I., Baral, C., Xi, B., & Pokala, L. (2023). Neuro-Symbolic Reasoning and Learning. Springer. ISBN 978-3-031-39179-8.

3 Bhuyan, B. P., Ramdane-Cherif, A., Singh, T. P., & Tomar, R. (2025). Neuro-Symbolic Artificial Intelligence: Bridging Logic and Learning (Studies in Computational Intelligence, Vol. 1176). Springer Nature Singapore. ISBN 978-981-97-8171-3. Scribd

6 Communicating Results

6.1 Planning on joint paper

As a result of an intensive week of working together around the theme of neurosymbolic approaches and its applications, we aim for a Communications of the ACM paper for a broad audience. Below are the contours of the paper.

Background. Neuro-symbolic AI systems that combine learning with reasoning are rapidly gaining attention in recent years. It is becoming increasingly clear that the data-intensive probabilistic approaches of neural networks and language models must be complemented by symbolic algorithms in order to function in domains with high stakes and low data regimes, which are typical in many real-world applications.

Message. We argue that the combination of learning and reasoning is not just a single niche approach in AI, but is actually a spectrum that encompasses many already successful techniques. This leads to rethinking the design of AI systems in general within this spectrum. We show that many leading AI systems can be "rationally reconstructed" using four neuro-symbolic principles: Reasoning, Assurances, Interfacability and Learning (RAIL for short). Firstly, this gives us a unified view on many seemingly disparate designs of AI systems, ranging from physics-aware machine learning to DeepMind's Alpha-X suite (AlphaGo, AlphaGeometry, AlphaFold) to Knowledge Completion, Causal Learning, and LLMs augmented with tool usage. Secondly, the neuro-symbolic RAIL principles allow engineers to make informed choices about AI system design, by moving along the axes for each of these 4 principles.

Goal.

(i) We show that many works in AI share some/all of the NeSy principles.
(ii) This gives a unified look at many disconnected lines of work and reveals possible connections/opportunities/challenges.
(iii) The NeSy field will benefit from learning the lessons learned in these other communities.
(iv) Other communities would benefit from looking at their own work through the NeSy lens.

6.2 Wikipedia Entry for Neurosymbolic AI

A working group has committed to improving the Wikipedia entry for Neurosymbolic AI, which currently lacks comprehensive coverage of the field's scope, methods, and applications. The group will rewrite the introduction for clarity, expand the historical overview, add sections on relationships to adjacent fields (including explainable AI, safe AI, and trustworthy AI), and document real-world applications. The goal is to establish a reliable, publicly accessible reference that reflects the current state of neurosymbolic research. See Section 4.7 for details on the proposed changes.

6.3 Educational Materials

A working group has committed to developing educational resources for neurosymbolic AI, targeting advanced master's students and early-stage PhD researchers. The planned deliverable is a course structure for a comprehensive textbook covering both theoretical

foundations and practical tooling. The group aims to bridge the gap between machine learning and knowledge representation communities by establishing shared vocabulary and providing hands-on exercises using tools such as ULLER and DeepLog. See Section 5.3 for the full discussion on content structure and pedagogical approaches.

Participants

- Agnese Chiatti
Polytechnic University of Milan, IT
- Michael Cochez
ELLIS Institute – Espoo, FI & Åbo Akademi University – Turku, FI
- Byron Cook
Amazon Web Services – Seattle, US & University College London, GB
- Cristina Cornelio
Samsung AI – Cambridge, GB
- Artur d'Avila Garcez
City University of London, GB
- Sebastijan Dumancic
TU Delft, NL
- Egor V. Kostylev
University of Oslo, NO
- Luis C. Lamb
CatholicTech – Cambridge, US & UFRGS – Porto Alegre, BR
- Robin Manhaeve
KU Leuven, BE
- Lia Morra
Polytechnic University of Turin, IT
- Mathias Niepert
Universität Stuttgart, DE
- Robert Peharz
TU Graz, AT
- Ute Schmid
Universität Bamberg, DE
- Alberto Speranzon
Lockheed Martin Advanced Technology Labs – Eagan, US
- Maarten C. Stol
Braincreators – Amsterdam, NL & VU Amsterdam, NL
- Annette ten Teije
VU Amsterdam, NL
- Thiviyan Thanapalasingam
University of Amsterdam, NL
- Guy Van den Broeck
UCLA, US
- Frank van Harmelen
VU Amsterdam, NL
- Emile van Krieken
VU Amsterdam, NL
- Antonio Vergari
University of Edinburgh, GB
- Maria-Esther Vidal
TIB Leibniz-Informationszentrum – Hannover, DE & Leibniz Universität Hannover, DE
- Benjie Wang
UCLA, US