

Neurosymbolic AI: From research to industry and the RAIL principles

Luis C. Lamb

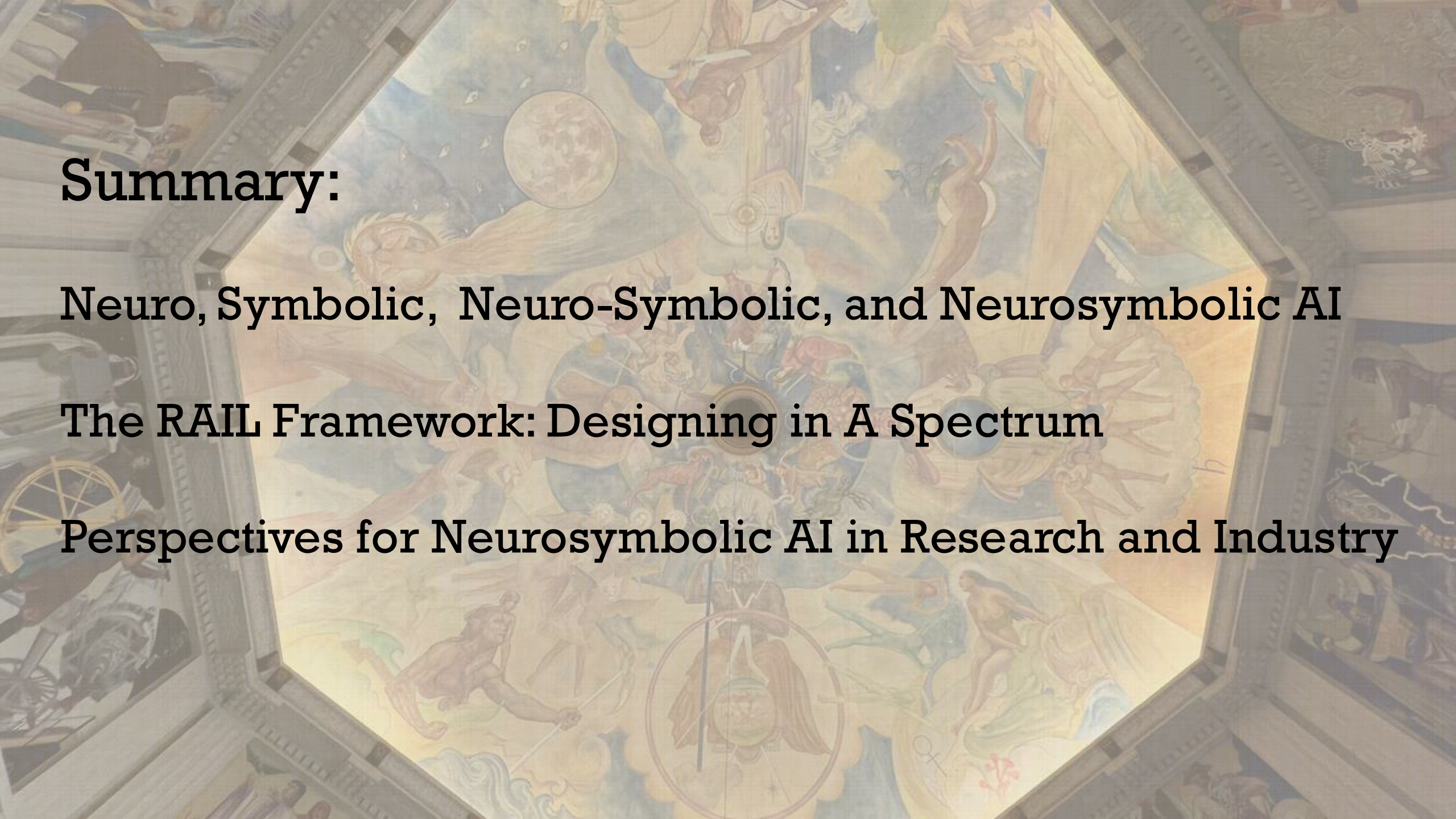
AI Innovation Institute

Stony Brook University, NY

Chiatti, Cochez, Cornelio, Dumancic, Garcez, Lamb, Morra, Niepert, Peharz, Speranzon, Stol, Ten Teije, Thanapalasingam, Van Harmelen, Van Krieken, Vergari, Wang.

Dagstuhl Seminar 25452

Communications of the ACM · to appear September 2026



Summary:

Neuro, Symbolic, Neuro-Symbolic, and Neurosymbolic AI

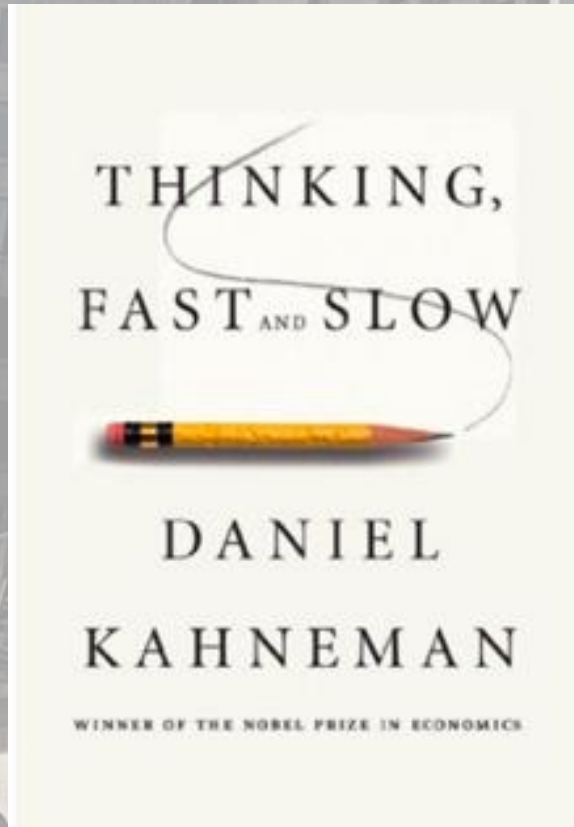
The RAIL Framework: Designing in A Spectrum

Perspectives for Neurosymbolic AI in Research and Industry

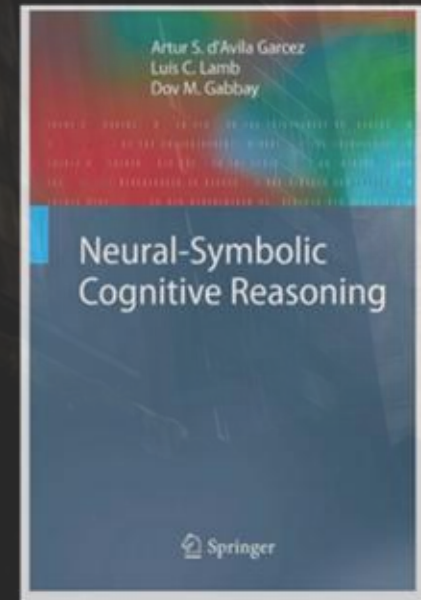
INTRODUCTION

- Neurosymbolic AI brings together the **statistical nature of machine learning and the rigorous logical nature of reasoning**. (GL 2023)
- **Industrial AI systems** that ingest sensor, text, and video streams are typically assembled from **ad-hoc, task-specific modules** that are designed and optimized independently.
- This divide-and-conquer approach accelerates development but **results in non-compositional pipelines with no formal guarantees** on overall behavior
- Error propagation, discontinuities at module interfaces, and **difficulties in interpreting the reasoning of generative models** persist in AI deployment.
- The **persistent problem of LLM hallucinations**, four years after the release of ChatGPT, underscores the need for a more principled solution.
- We will frame neurosymbolic AI around four core principles: **Reasoning, Assurances, Interfacing, and Learning**, providing a blueprint for **building reliable, explainable, and compositional industrial AI** pipelines.

The First Great AI debate 2019: integrating learning & reasoning via neurosymbolic AI



- Until roughly 2018, mainstream ML largely ignored *The Algebraic Mind*
- But *The Algebraic Mind* inspired the seminal book on neurosymbolic approaches
- And, as we will see, *Algebraic* also anticipated much of Yoshua's current argument



2008

Neural-Symbolic Cognitive Reasoning

Gary Marcus in the Great AI Debate, Montreal
23 december 2019.

Understanding the Science of AI

“When no one knew **why things worked**, even radical technological shifts **failed to generate the applications that made them useful in production processes**. In **such a world**, economic resources were not allocated towards improving technologies, as potential investors were just as likely to **waste valuable resources** on fruitless inventions that would never work, such as perpetual motion machines or artificially created gold (Mokyr, 2002, p. 31–32).”
Nobel Prize announcement, 2025.

And quoted by Professor Joshua Gans, October 13, 2025.

Scientific Background to the Sveriges Riksbank Prize in
Economic Sciences in Memory of Alfred Nobel 2025.

The Committee for the Prize in Economic Sciences in Memory
of Alfred Nobel.

Understanding the Science & Technology of AI

Applying the RAIL framework offers a **unified view of seemingly disparate AI systems**, ranging from physics-aware machine learning to Google DeepMind's Alpha-* suite, causal learning and tool-augmented Large Language Models.

Importantly, using the neurosymbolic RAIL principles as a much-needed new evaluation standard for AI will enable engineers to make **better-informed** and more **principled decisions** about the design and deployment of **production-level AI** systems.

Suggested References and Background

The RAIL Principles for Neurosymbolic AI: Reasoning, Assurances, Interfacing and Learning

AGNESE CHIATTI, Politecnico di Milano, Italy

MICHAEL COCHEZ, ELLIS Institute Finland & Åbo Akademi University, Finland

CRISTINA CORNELIO, Samsung AI, UK

SEBASTIJAN DUMANCIC, Delft University of Technology, The Netherlands

ARTUR D'AVILA GARCEZ, City St. George's, University of London, UK

LUIS C. LAMB, Stony Brook University, NY, USA

LIA MORRA, Politecnico di Torino, Italy

MATHIAS NIEPERT, University of Stuttgart, Germany

ROBERT PEHARZ, Graz University of Technology, Austria

ALBERTO SPERANZON, Lockheed Martin, Advanced Technology Labs, MN, USA

MAARTEN STOL, BrainCreators & Vrije Universiteit Amsterdam, The Netherlands

ANNETTE TEN TEIJE, Vrije Universiteit Amsterdam, The Netherlands

THIVIYAN THANAPALASINGAM, University of Amsterdam, The Netherlands

FRANK VAN HARMELEN, Vrije Universiteit Amsterdam, The Netherlands

EMILE VAN KRIEKEN, Vrije Universiteit Amsterdam, The Netherlands

ANTONIO VERGARI, University of Edinburgh, UK

BENJIE WANG, UCLA, CA, USA

<https://arxiv.org/abs/2608.04285>

Artificial Intelligence Review (2023) 56:12387–12406
<https://doi.org/10.1007/s10462-023-10448-w>

Neurosymbolic AI: the 3rd wave

Artur d'Avila Garcez¹ · Luís C. Lamb² 

<https://link.springer.com/article/10.1007/S10462-023-10448-W>

Why Neurosymbolic AI in the Current State of Affairs

Generative AI and Large Language Models (LLMs):

- powerful at learning statistical patterns in data/language, but
 - make **logical errors** or contradict themselves (hallucinations).
 - expensive to train, maintain, update, and deploy.
 - integrating **logic**, as constraints, post-processing, or embedded logical / symbolic reasoning helps to improve **consistency, semantics, and explainability**.

Estimated training cost of select AI models, 2016–24

Source: Epoch AI, 2024 | Chart: 2025 AI Index report

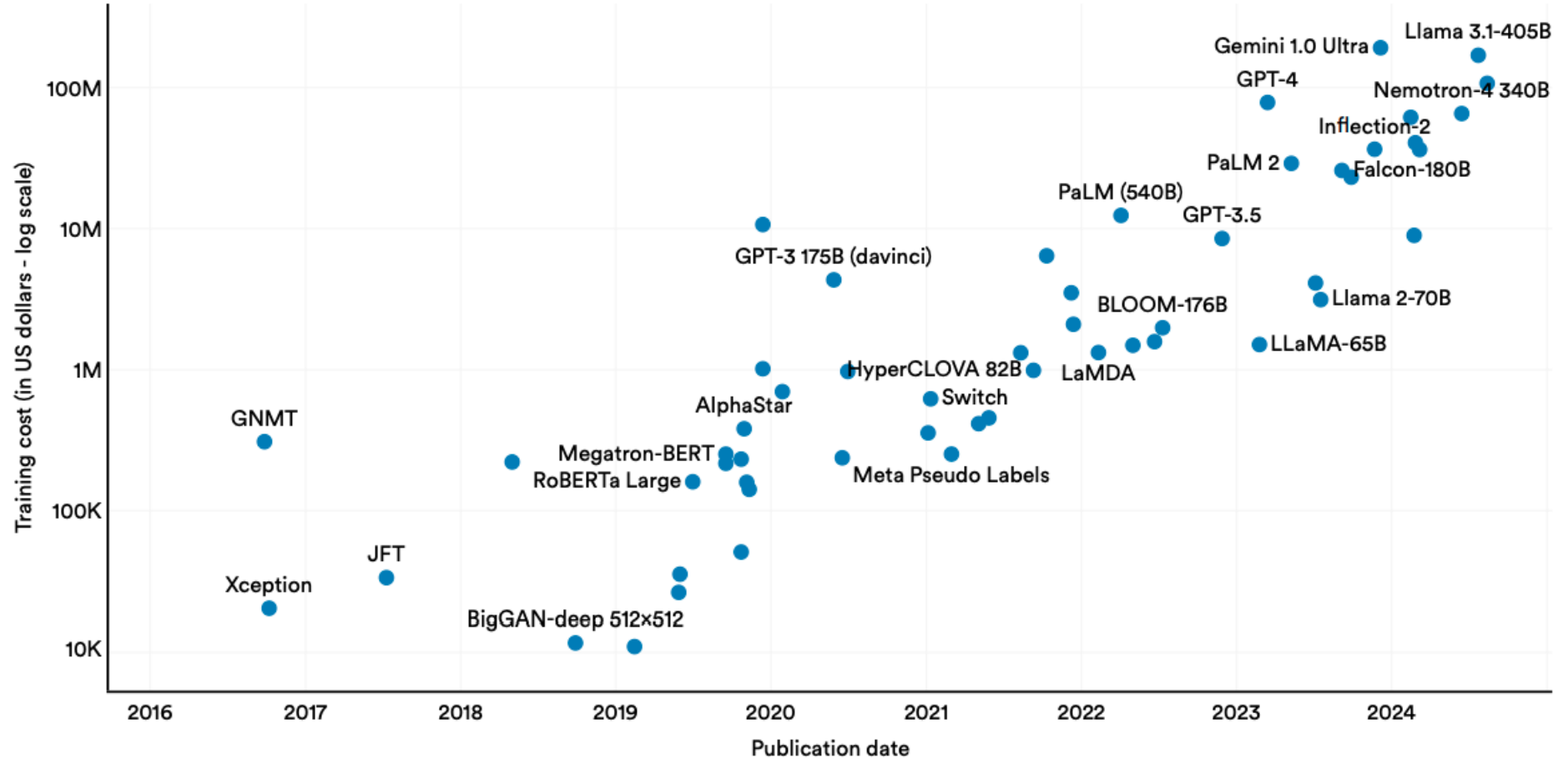


Figure 1.3.25

AI Index Report, Stanford HAI 2025

https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf

Early Neurosymbolic AI?

McCulloch & Pitts, 1943. A Logical Calculus of the Ideas Immanent in Nervous Activity.

S. C. Kleene, 1951.
Representation of Events in
Nerve Nets and Finite Automata

Summary: To what kinds of events can a McCulloch-Pitts nerve net respond by firing a certain neuron? More generally, to what kinds of events can any finite automaton respond by assuming one of certain states? This memorandum is devoted to an elementary exposition of the problems and of results obtained on it during investigations in August 1951.

REPRESENTATION OF EVENTS
IN NERVE NETS AND FINITE AUTOMATA

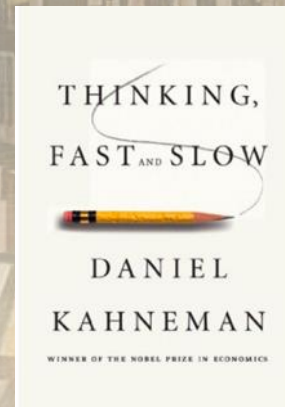
S. C. Kleene

Dual Process Theory

Kahneman: System 1 & System 2

System 1 - intuitive, fast parallel system - is capable of understanding language
Deep Learning system 1

System 2, deliberative, reflexive, most probably performing symbol manipulation.



Kahneman:

“...so far as I'm concerned, System 1 certainly knows language... System 2... does involve certain manipulation of symbols.” AAI2020

Why Neurosymbolic AI

Neurosymbolic AI:

Learning from experience and reasoning about what has been learned from an uncertain environment in a computationally efficient way.

Motivation: there is a need for systems that:

- Learn from changes in the environment (think of temporal logics).
- Reason about commonsense knowledge (space/time/uncertainty...).
- Integrate reasoning (computation) and learning: we combine the logical nature of reasoning and the statistical nature of learning, see Valiant.
- But... Reasoning can be hard.

Why Neurosymbolic AI

But, in AI, agents “want” to reason about, time, knowledge and interact with other agents and the environment.

How to learn temporal, epistemic, and probabilistic logic/abstractions?

Classic Literature:

Ronald Fagin, Joseph Y. Halpern, Yoram Moses, Moshe Y. Vardi:
Reasoning About Knowledge. MIT Press 1995.

Joseph Y. Halpern: **Reasoning about uncertainty.** MIT Press 2005

How to do it: Connectionist Modal Logics

Modal logic goes beyond propositional reasoning: Vardi, 1997.

A proposition is necessary (*box*) in a possible world (state of affairs) if it is true in all worlds which are possible in **relation** to that world.

A proposition is possible (*diamond*) in a possible world (state of affairs) if it is true in at least one world which is possible in relation to that same world (reference state).

Modalities also used for reasoning about uncertainty (following J. Halpern).

Relational learning/reasoning is notoriously hard.

Conectionist Modal Logics

- Insight: *assume that neurons are possible worlds.*
- Propositional Modal Logic = decidable fragment of FOL with two variables.
- Full solution of Muddy Children puzzle and other testbeds.
- Ensembles of NNs that are seen as possible world.

Connectionist Modal Logics

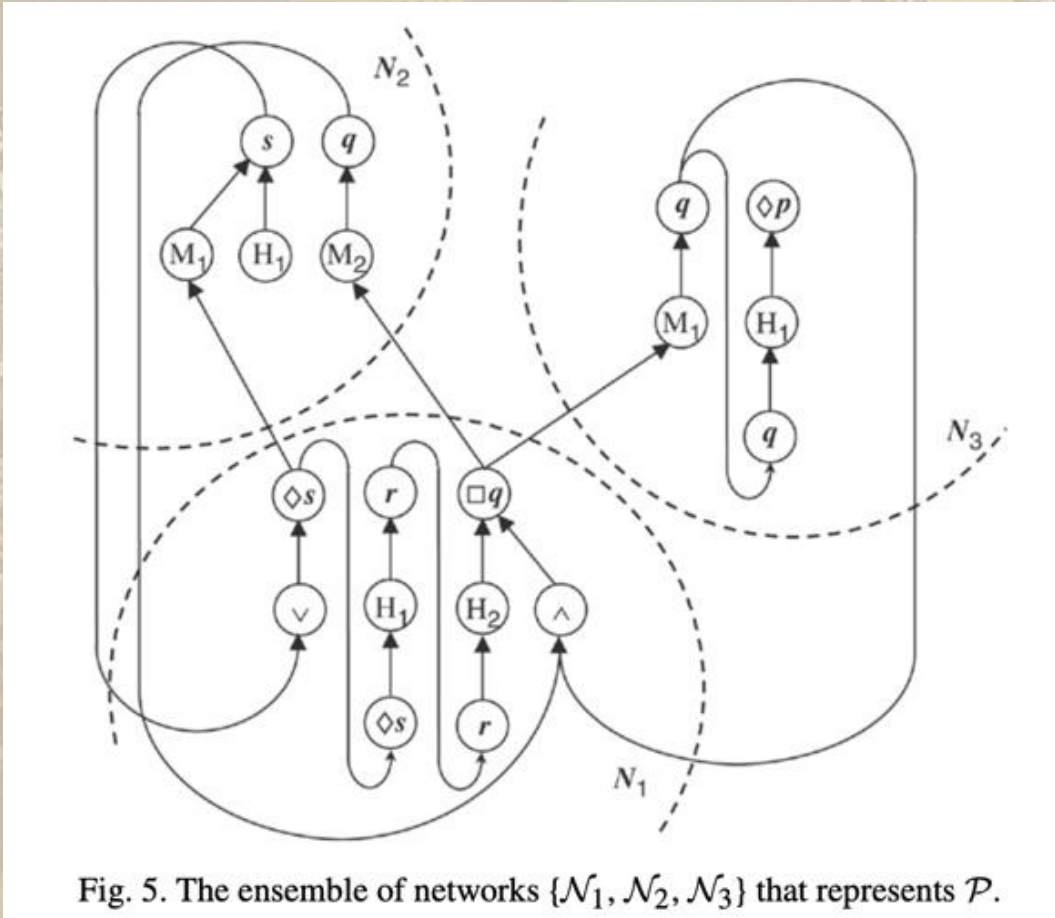


Fig. 5. The ensemble of networks $\{N_1, N_2, N_3\}$ that represents P .

$$P = \{\omega_1 : r \rightarrow \Box q; \omega_1 : \Diamond s \rightarrow r; \omega_2 : s; \omega_3 : q \rightarrow \Diamond p; R(\omega_1, \omega_2), R(\omega_1, \omega_3)\}$$

THEOREM : For any modal/temporal logic program P there exists an ensemble of neural networks N such that N computes P .
 Garcez, Lamb, 2006, *Neural Computation*, vol 18(7).

Garcez, Lamb, Gabbay. *Neural-Symbolic Cognitive Reasoning*, Springer, 2009.

Garcez, Lamb. Connectionist Model for Epistemic and Temporal Reasoning. *Neural Computation*, 18:1711-1738, July 2006

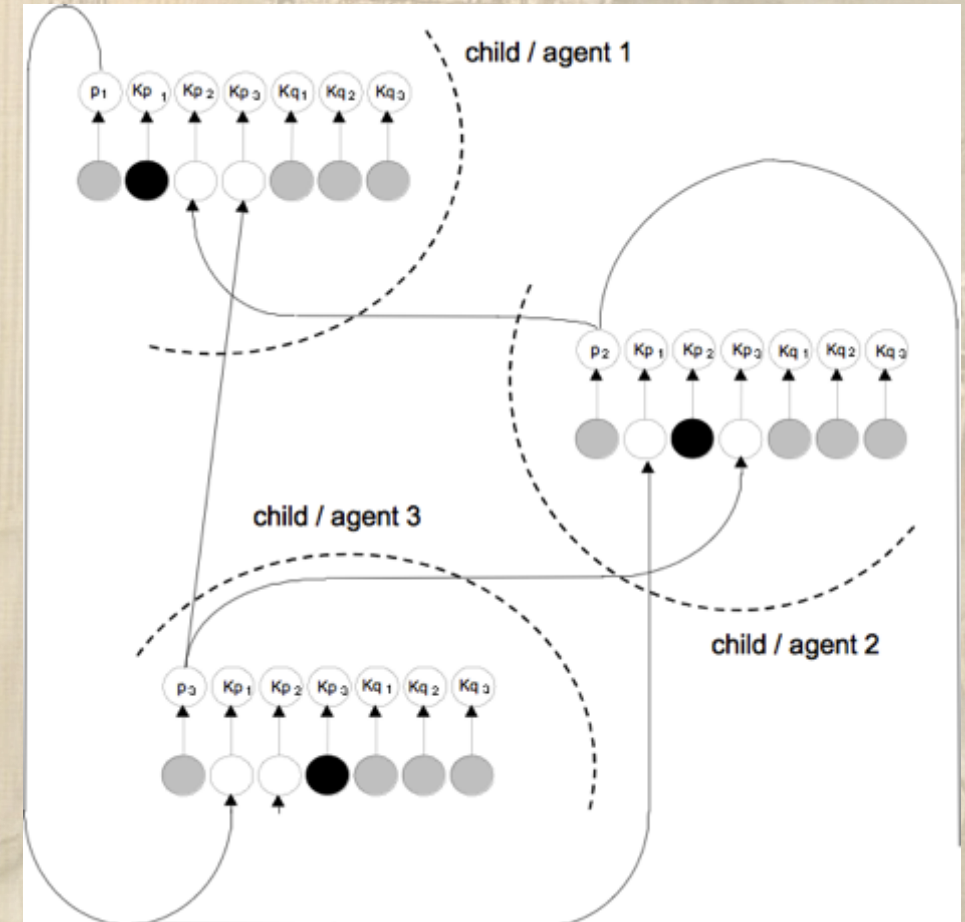
Conectionist Modal Logics

Table 1
Rules for modality operators

$\frac{\begin{array}{c} [R(\omega, g_\varphi(\omega))] \\ \vdots \\ g_\varphi(\omega) : \varphi \\ \hline \omega : \Box \varphi \end{array} \Box I}{\omega : \Diamond \varphi} \Diamond E$	$\frac{\omega_1 : \Box \varphi, R(\omega_1, \omega_2)}{\omega_2 : \varphi} \Box E$
$\frac{\omega : \Diamond \varphi}{f_\varphi(\omega) : \varphi, R(\omega, f_\varphi(\omega))} \Diamond I$	$\frac{\omega_2 : \varphi, R(\omega_1, \omega_2)}{\omega_1 : \Diamond \varphi} \Diamond I$

"Learning to reason" here literally means:
compile the logic into an initial architecture, then gradient-descent on examples to refine it, while the $\vee/\wedge/M$ inter-network wiring keeps enforcing the Kripke-style dependencies throughout training.

Learning = ordinary backpropagation, locally, per world



An Early Application of Neurosymbolic AI, now a standard?

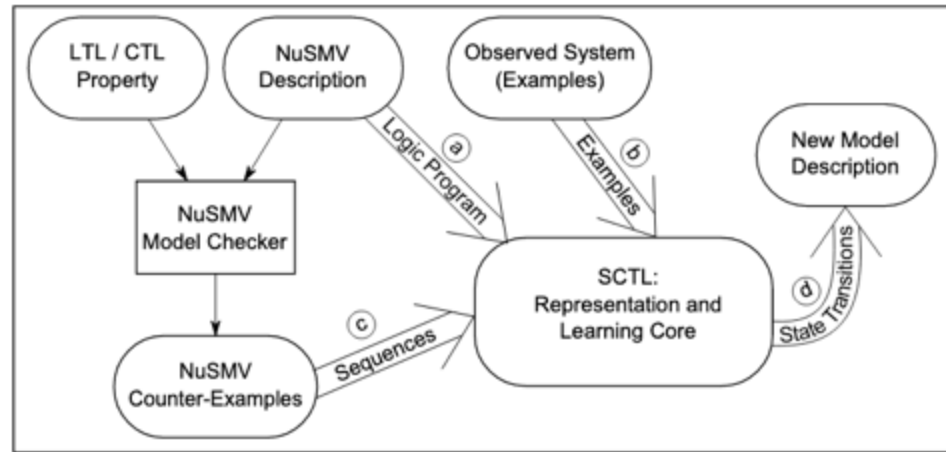


Figure 1: General diagram of the proposed framework

This model can then be subject to verification by NuSMV.

If the model does not satisfy the property, a set of counter-examples is returned.

Counter-examples are used as input of the adaptation engine in order to obtain a new, improved model.

This new model can be subject to the same process until the properties are satisfied.

An initial description of a model can be expressed in NuSMV or as a temporal logic program.

Source: R. Borges, A. Garcez, L.C. Lamb, IEEE T. on Neural Networks 2011, ICSE 2011, ASE2011 and more.

Similar approaches:

Labor Division with Movable Walls: Composing Executable Specifications with Machine Learning and Search.

Harel et al. **AAAI 2019**

Robust Deep Reinforcement Learning Using Formal Verification

Raviv et. al. **TASE 2025**

Transitioning into Industry

THE WALL STREET JOURNAL.

World Business U.S. Politics Economy Tech Markets & Finance Opinion Arts Lifestyle

STEVEN ROSENBUSH

Meet Neurosymbolic AI, Amazon's Method for Enhancing Neural Networks

A hybrid approach to AI is powering Amazon's Rufus shopping assistant and cutting-edge warehouse robots



By [Steven Rosenbush](#) [Follow](#)

Aug. 12, 2025 11:00 am ET

*“At our inference time or after inference, you can **move the statements coming out of a language model, put them into logic and then prove or disprove** the correctness of them.*

Automated reasoning is the symbolic manipulation of, like you move symbols around and you deduce things that are true about the semantics that those formally represent.

*You can do all kinds of things **in combination with machine learning.**”*

“The rise of interest in neurosymbolic AI caused people, while they were using the tool, to realize how important this work was.”

Byron Cook, Distinguished Scientist and Vice President at AWS

VentureBeat, Aug. 2025

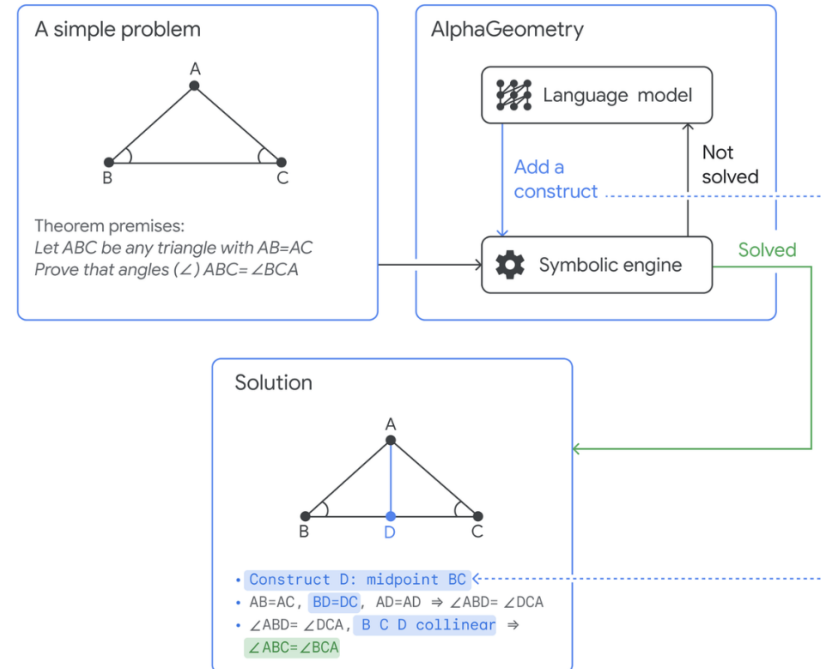
Transitioning into Industry



AlphaGeometry is a system made up of **2** parts:

- A neural language model, which can predict useful geometry constructions to solve problems
- A symbolic deduction engine, which uses logical rules to deduce conclusions

Both work together to find proofs for complex geometry theorems.



AlphaGeometry is a **neuro-symbolic** system made up of a neural language model and a symbolic deduction engine, which work together to find proofs for complex geometry theorems.

Akin to the idea of “thinking, fast and slow”, one system provides fast, “intuitive” ideas, and the other, more deliberate, rational decision-making.

“AlphaGeometry's output is impressive because it's both verifiable and clean.”

Solving olympiad geometry without human demonstrations

[Trieu H. Trinh](#) , [Yuhuai Wu](#), [Quoc V. Le](#), [He He](#) & [Thang Luong](#) 

[Nature](#) **625**, 476–482 (2024) | [Cite this article](#)

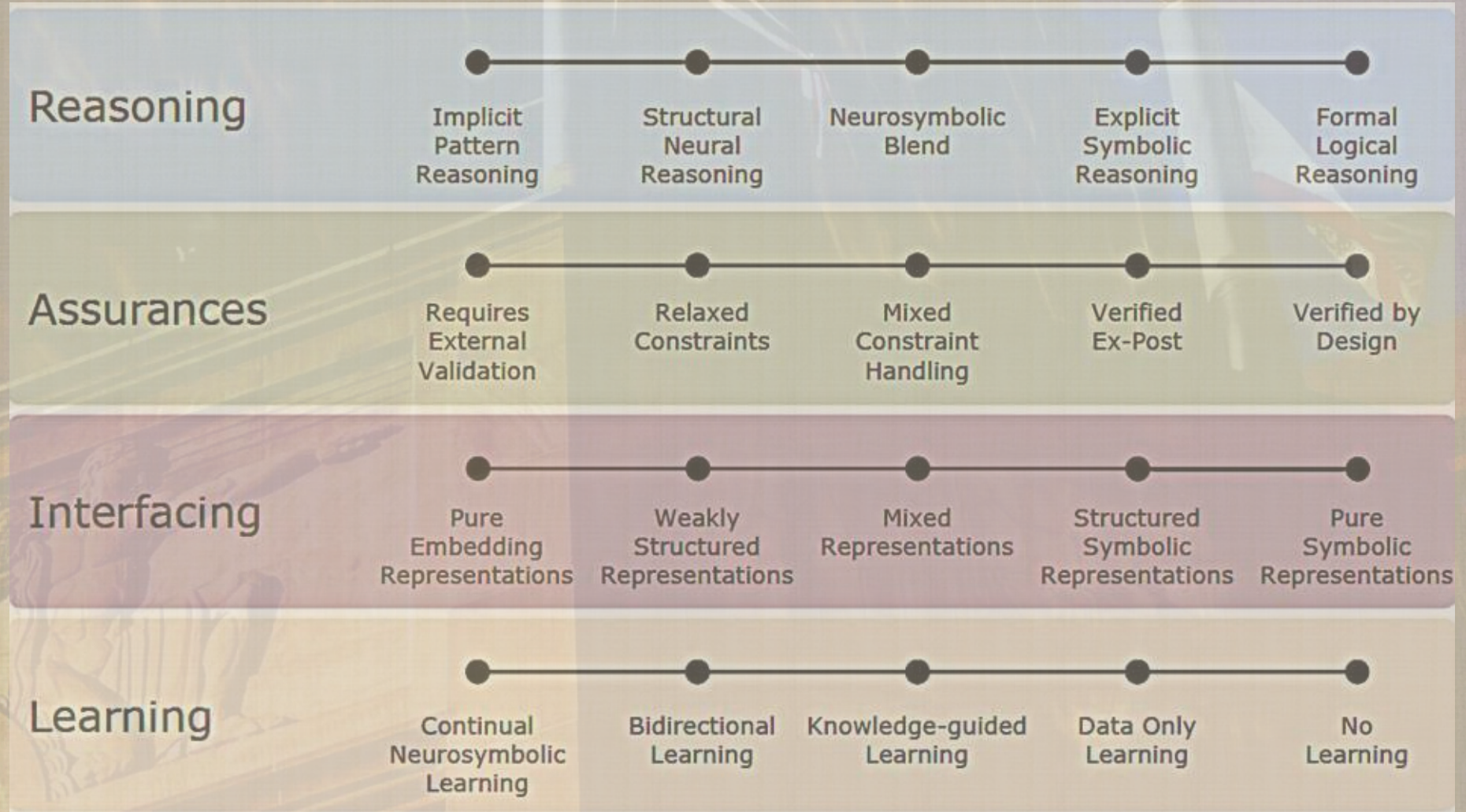
THE FOUR RAIL PRINCIPLES

RAIL Principles

Neurosymbolic AI systems can be read along four design dimensions that trade off against one another:

- 1 Reasoning**
How inference moves from implicit pattern completion to explicit, formal symbolic reasoning.
- 2 Assurances**
What guarantees the system can provide about safety, performance, and reliability requirements.
- 3 Interfacing**
How knowledge and explanations flow between humans and the system, in both directions.
- 4 Learning**
How much of the system's behaviour comes from data versus from injected structured knowledge.

The RAIL Principles at a Glance



The RAIL Principles for Neurosymbolic AI: Each principle exists on a qualitative spectrum, with intermediate positions showing different combinations of neural and symbolic approaches.

Source: Chiati et al, **Comm of the ACM to Appear 2006.** <https://arxiv.org/abs/2608.04285>

From Implicit Patterns to Formal Logic

REASONING

Reasoning chains inference steps to reach a conclusion: deductive, abductive, inductive, or analogical.

This axis spans from implicit pattern completion to fully formal symbolic proof.

1	2	3	4	5
Implicit Pattern	Structural Neural	Neurosymbolic Blend	Explicit/ Symbolic	Formal / Logical
Emergent, latent reasoning by similarity: e.g. an LLM's chain-of-thought.	Reasoning is encoded in a neural network's own structure and weights.	Systems such as DeepProbLog and Logic Tensor Networks mix implicit and explicit inference.	A neural model calls symbolic computation, e.g. generating code and calling an interpreter.	Sound inference in classical or non-classical logic, with verifiable steps.

Reliable systems span the full spectrum of reasoning.

From External Validation to Verified by Design

ASSURANCES

Assurances are the guarantees a system provides against safety, performance, and reliability requirements: from unconstrained models to guarantees baked into the architecture.

1	2	3	4	5
External Validation	Relaxed Constraints	Mixed Constraints	Verified Ex-Post	Verified by Design
Neural components trained with no requirements at all; correctness must be checked outside the system.	Requirements are relaxed into differentiable form, e.g. fuzzy logic, for training and inference-time reasoning.	Shield layers assign zero probability to outputs that would violate a constraint, e.g. an output grammar.	Properties or outcomes are checked for correctness only after the system has been trained.	Requirements are baked directly into the network architecture itself.

Balancing assurances against their scalability limits remains a core open challenge.

From Pure Embeddings to Pure Symbols

INTERFACING

Interfacing is the shared representation that lets humans and systems exchange domain knowledge and explanations, in both directions.

1	2	3	4	5
Pure Embedding	Weakly Structured	Mixed Repres.	Structured Symbolic	Pure Symbolic
High-dimensional latent representations capture statistical structure but resist direct inspection.	Partial organization: e.g. attention patterns or neuron groupings — without a defined semantics.	Graphs and causal models support intervention while staying compatible with subsymbolic systems.	Logic programs, annotated code, or structured data formats prioritize explicit compositional structure.	Formal knowledge representation with unambiguous semantics: expressive, but usually hand-specified.

Interfacing is often asymmetric: guiding a system is easy, but getting an explanation back out remains hard.

From Continual to No Learning

LEARNING

Learning captures how much of a system's behaviour comes from data versus from injected, structured knowledge.

1

Continual NS

Learning continues over time as knowledge evolves through feedback loops such as reinforcement learning.

2

Bidirectional

The system learns to evolve its own symbolic knowledge, rather than treating it as fixed background.

3

Knowledge-Guided

Structured knowledge shapes learned representations, improving generalisation and data efficiency.

4

Data-Only Learning

Behaviour is acquired purely by fitting parameters to data, with limited control over what is learned.

5

No Learning

Pre-defined symbolic knowledge is used as-is, relying on knowledge engineering rather than statistics.

Most industrial systems combine data-only components with pockets of knowledge-guided learning.

Where RAIL Shows Up Today

SIX DOMAINS

RAIL is not just a design ideal: major areas of AI already sit somewhere on all four RAIL dimensions.

Knowledge Discovery: Embedding-based link prediction versus rule-learning over knowledge graphs.

Tool-Augmented LLMs: Language models that write code, call APIs, and reason step-by-step with external tools (e.g. ReAct).

Physics-Aware ML: Physical laws, symmetries, and conservation principles act as symbolic knowledge for scientific ML.

Neuro-Guided Search: DeepMind's Alpha family (AlphaGo, AlphaFold, AlphaGeometry) pairs neural search with symbolic game and proof trees.

Causal Learning: Causal graphs mediate between statistical estimation and formal do-calculus reasoning.

Industrial Architectures: Modular production pipelines already combine learned components with symbolic glue code.

Knowledge Discovery on RAIL

Embedding-based methods map graph entities into vector space and predict new links by geometric similarity.

Rule-based methods such as AnyBURL learn explicit logical rules from the graph and apply them to infer new facts.

	Embedding-Based	Rule-Based
Reasoning	Implicit: reasons over vector similarity and distances in the embedding space.	Explicit: applies symbolic rules discovered from the graph.
Assurances	Relaxed constraints only; loss minimisation raises likelihood but gives no guarantee.	Can move toward verified constraints via ex-post ontology checks (types, cardinality).
Interfacing	Weakly structured: embeddings are hard for humans to inspect directly.	Fully symbolic: graph, rules, and new relations stay human-readable.
Learning	Data-only, or knowledge-guided when informed sampling strategies are used.	Bidirectional: rules are learned from the graph, then applied back to it.

The two categories are almost opposite on RAIL: embeddings favour learning, rule-learning favours assurances and interfacing.

Tool Use by Language Models on RAIL

Coupling LLMs with programs re-introduces neurosymbolic principles in three increasingly explicit forms.

Program Synthesis

The LLM generates executable code from a natural-language spec; reasoning stays implicit, but the output is explicit and testable.

Tool Invocation

The LLM calls existing APIs or database queries, gaining assurances from execution and unit tests it did not need to author.

Agentic Use (ReAct)

Reasoning and action interleave: the model deliberates, acts, observes the result, and refines its next step.

Interleaving reasoning and action curbs hallucination, but early tool-selection errors can still cascade through longer sequences.

MORE LESSONS

What Four More Systems Teach Us About RAIL

Neuro-Guided Search

Alpha family

Stronger formal assurances need more explicit reasoning: AlphaGo's implicit game rules vs. AlphaProof's LEAN-verified proofs.

Causal Learning

Causal graphs

A causal graph mediates between data-driven estimation and formal do-calculus reasoning about interventions.

Physics-Aware ML

Physical laws

Symmetries and conservation laws act as symbolic knowledge, giving assurance-by-design through equivariant and Hamiltonian architectures.

Industrial Architectures

Modular pipelines

Learned components and symbolic glue code already coexist in production, but assurances remain the hardest RAIL dimension to fill.

RAIL Meets Industry

Three deployed systems already sit at different points on the RAIL spectrum — neurosymbolic AI is no longer just a research idea.

1 **Amazon Rufus: Reasoning & Learning**

Amazon's shopping assistant blends an LLM's implicit reasoning with structured product knowledge, and keeps learning from customer interactions. (Wall Street Journal, 2025)

2 **AWS Bedrock Guardrails: Assurances**

Automated, LEAN-based reasoning checks validate LLM outputs against logical rules, reportedly catching hallucinations with up to 99% accuracy.

3 **AlphaGeometry: Interfacing**

A language model proposes constructions in natural language; a symbolic deduction engine turns them into formally verifiable Olympiad-level proofs. (ALPHABET, 2024)

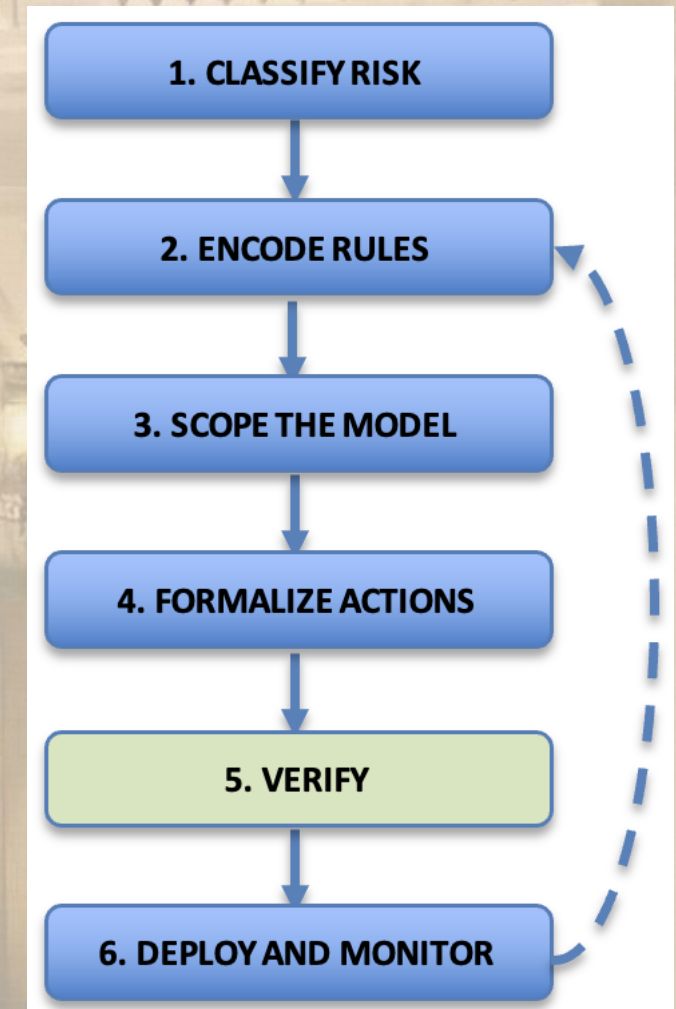
NeSyOps: NeuroSymbolic Operations

NeSyOps: neurosymbolic verification as a pipeline stage.

Not a replacement for AgentOps. One new stage.

It denotes the **addition of a formal (neurosymbolic) verification stage**, operating over an explicit rule base held separately from the agent, to existing AgentOps, LLMOps, and MLOps pipelines.

To fill the methodological gap in which models or **agents' validation is run via testing with no formal guarantees** when developing Agentic AI systems.



NeSyOps: NeuroSymbolic Operations

Verification as a pipeline stage.

PREPARE THE ACTION

CHECK

RUN & LEARN

1

Classify risk

Rank actions by reversibility and severity; decide which ones need formal checking.

NIST AI RMF, OPA

2

Encode rules

Write domain constraints in a formal language, held separately from the agent.

OWL, Datalog, TLA+

3

Scope the model

The agent proposes; for risky actions it holds no decision power.

JSON Schema, NeMo Guardrails

4

Formalize

Translate the agent's natural-language proposal into a formal claim.

SMT-LIB, first-order logic, Lean

5

Verify

A solver checks the claim against the rules and returns a proof or a counterexample.

6

Deploy & monitor

Verified actions run; cases the rules did not cover are flagged.

Bedrock Guardrails, OPA at runtime



Feedback: an uncovered case is added to the rule file, not patched into the prompt

Stage 5 is the neurosymbolic step:

instead of a test score, a guarantee over every case the encoded rules describe.

NeSyOps in Practice

Same six stages, two kinds of failure: a contradiction hidden in a policy, and a deadlock hidden in a protocol.

Airline refund chatbot

Based on AWS Automated Reasoning checks (ARc)

Risk	Wrong eligibility answers create legal exposure at scale.
Rules	Refund policy encoded formally; a second expert reviews it.
Formalize	“Denied boarding $\rightarrow$ full refund” becomes premises $\Rightarrow$ refund-eligible.
Verify	Two rules fire and clash (flight operated, not travelled $\Rightarrow$ not eligible; denied boarding $\Rightarrow$ eligible). The solver flags a gap in the policy, not a chatbot error.
Fix	Policy owner adds an explicit denied-boarding exception.

After expert refinement: soundness 96.8% $\rightarrow$ 100%, recall 25% $\rightarrow$ 45.5%.

C. An et al(2026) A Neurosymbolic Approach to Natural Language Formalization and Verification. In CAV 2026 – AWS Team.

Multi-agent data pipeline

Dijkstra’s dining philosophers, three agents on a ring

Risk	Silent full stall: no crash, no exception, low observability.
Rules	3 agents, 3 partition locks, protocol “left lock, then right”, specified in TLA+.
Formalize	Property $AG \neg \text{Deadlock}$, checked over every interleaving.
Verify	Model checker returns a counterexample: A0 waits on A1, A1 on A2, A2 on A0.
Fix	Resource hierarchy (lower-numbered lock first); re-check returns a proof.

A failure that testing and runtime monitoring would likely miss.

Open problems: the Stage 4 translation must itself be checked; guarantees cover only what Stage 2 encodes; no principled rule yet for how much to formalize given the risk.

See:

Conclusions

RAIL is a design space, not a checklist.

Four axes constrain how to frame a neurosymbolic architecture:

- 1** Reasoning spans implicit pattern-matching to formal logic.
- 2** Assurances span external validation to built-in guarantees.
- 3** Interfacing is two-way: guidance in, explanations out.
- 4** Learning spans no learning to continual learning.
- 5** These four axes trade off against each other.

This is a qualitative map.

Turning RAIL into a practical tool for AI will consider operational criteria and concrete metrics. NeSyOps is a step in that direction (see next)

Conclusions

From a design space to an engineering practice.

RAIL tells us how to position a neurosymbolic system. **NeSyOps** puts that position into the development pipeline.

RAIL AXIS DESIGN SPECTRUM

Reasoning Implicit pattern-matching → formal logic

Assurances External validation → built-in guarantees

Interfacing Two-way: guidance in, explanations out

Learning No learning → continual learning

WHERE **NESSYOPS** ACTS

Stage 4 formalizes the agent's proposal, moving it to the formal end.

Stage 5 returns a proof or counterexample, not a test score.

Stages 2 and 3 keep rules and scope outside the model; experts edit them.

Stage 6 grows the rule base from new cases, without retraining.

Takeaway:

The axes still trade off, but verification becomes a pipeline stage, not an afterthought.

Next Steps: Operational criteria (NeSyOps) and metrics for RAIL; verified NL-to-formal translation; a principled way to weigh formalization cost against risk.

Neurosymbolic AI is moving from research into deployable practice.

Path Forward

Several high-impact systems turn out to already be neurosymbolic, unwittingly following the RAIL principles.

RAIL defines a four-dimensional design space whose inter-dependent axes constrain viable neurosymbolic architectures: Mapping existing systems onto it reveals real trade-offs and synergies.

By visualising these relationships, RAIL opens new design opportunities for both research and industrial deployment.

Path Forward: Foundational

(1) First-order logic and higher-order knowledge extraction from very large networks that is provably sound and efficient, explains the entire model and *local network interactions* and accounts for different levels of abstraction.

(2) Goal-directed commonsense and efficient combinatorial reasoning about what has been learned by a complex deep network trained on large amounts of multimodal data.

e.g. Learning to Solve NP-Complete Problems: A Graph Neural Network for Decision TSP. Prates, Avelar, Lemos, Lamb, Vardi. AAAI-2019.

(3) Human-network communication as part of a multiagent system that promotes communication/argumentation protocols between the user and an agent that can ask questions and check her understanding.

Path Forward: Foundational

Axiomatizing Neurosymbolic AI:

Building a consensus around a shared set of foundational principles for NeSy AI.

A unifying theoretical foundation to:

reduce the overhead of describing new systems, facilitate collaboration across subfields

foster clearer communication [within the] area, Avoids continued fragmentation and redundancy.

See:

C. Pryor, L. Getoor: Neural-Symbolic Architectural Axioms of Integration: A Manifesto. NeSy 2025;

Garcez, Lamb, Gabbay: Neural-symbolic Cognitive Reasoning, Springer, 2009.

Full Development of NeSyOps:

- A sound methodology for the design of agentic neurosymbolic AI Systems.
- NeSyOps is the operationalization of the RAIL principles.
- Forthcoming NeSyOps paper by Lamb (submitted)

References and Background

The RAIL Principles for Neurosymbolic AI: Reasoning, Assurances, Interfacing and Learning

AGNESE CHIATTI, Politecnico di Milano, Italy

MICHAEL COCHEZ, ELLIS Institute Finland & Åbo Akademi University, Finland

CRISTINA CORNELIO, Samsung AI, UK

SEBASTIJAN DUMANCIC, Delft University of Technology, The Netherlands

ARTUR D'AVILA GARCEZ, City St. George's, University of London, UK

LUIS C. LAMB, Stony Brook University, NY, USA

LIA MORRA, Politecnico di Torino, Italy

MATHIAS NIEPERT, University of Stuttgart, Germany

ROBERT PEHARZ, Graz University of Technology, Austria

ALBERTO SPERANZON, Lockheed Martin, Advanced Technology Labs, MN, USA

MAARTEN STOL, BrainCreators & Vrije Universiteit Amsterdam, The Netherlands

ANNETTE TEN TEIJE, Vrije Universiteit Amsterdam, The Netherlands

THIVIYAN THANAPALASINGAM, University of Amsterdam, The Netherlands

FRANK VAN HARMELEN, Vrije Universiteit Amsterdam, The Netherlands

EMILE VAN KRIEKEN, Vrije Universiteit Amsterdam, The Netherlands

ANTONIO VERGARI, University of Edinburgh, UK

BENJIE WANG, UCLA, CA, USA

<https://arxiv.org/abs/2608.04285>

Communications of the ACM, 2026, to appear

Also see the Proceedings of the **NeSy Conference**: <https://nesy-ai.org>

Contact: LuisLamb@alumni.mit.edu

Artificial Intelligence Review (2023) 56:12387–12406
<https://doi.org/10.1007/s10462-023-10448-w>

Neurosymbolic AI: the 3rd wave

Artur d'Avila Garcez¹ · Luís C. Lamb² 

<https://link.springer.com/article/10.1007/S10462-023-10448-W>

References on Neurosymbolic AI Computational Foundations, Main Methodologies and Tools

- [1] A. d'Avila Garcez, L. C. Lamb, D.M. Gabbay. **Neural-Symbolic Cognitive Reasoning**. Springer, 2009.
- [2] T.R. Besold, A. d'Avila Garcez, S. Bader, H. Bowman, P. Domingos, P. Hitzler, K-U. Kühnberger, L. C. Lamb, P. Lima, L. de Penning, G. Pinkas, H. Poon, G. Zaverucha: **Neural-Symbolic Learning and Reasoning: A Survey and Interpretation**. In: Neuro-Symbolic Artificial Intelligence, 2022.
- [3] R.Riegel, A. Gray, F. Luus, N.Khan, N. Makondo, I. Akhalwaya, H.Qian, R.Fagin, F. Barahona, U.Sharma, S.Ikbal, H.Karanam, S.Neelam, A.Likhyani, S.Srivastava: **Logical Neural Networks**. arXiv (2020) & PNAS (2023).
- [4] R. Manhaeve, S. Dumančić, A. Kimmig, T. Demeester, L. De Raedt . **DeepProbLog: Neural Probabilistic Logic Programming**. NeurIPS (2018)
- [5] S. Badreddine, A. Garcez, L, Serafini, M. Spranger: **Logic Tensor Networks**. AI Journal (2022)
- [6] A.Tavares, P. Avelar, J.Flach, M.Nicolau, L.C.Lamb, M. Vardi: **Understanding Boolean Function Learnability on Deep Neural Networks**. NeSy (2025).
- [7] D Cunningham, M Law, J Lobo, A Russo. **The Role of Foundation Models in Neuro-Symbolic Learning and Reasoning** arXiv (2024)
- [8] G. Ciravegna, P. Barbiero, F. Giannini, M. Gori, P. Lió, M. Maggini, S.Melacci: **Logic Explained Networks**. AI Journal (2023)
- [9] L.C. Lamb, A. S. d'Avila Garcez, Marco Gori, M. O. R. Prates, Pedro H. C. Avelar, Moshe Y. Vardi: **Graph Neural Networks Meet Neural-Symbolic Computing: A Survey and Perspective**. IJCAI 2020: 4877-4884
- [10] Z. Xu, Zhang, T Friedman, Y Liang, G Van den Broeck. **A semantic loss function for deep learning with symbolic knowledge**. ICML 2018.
- [11] C. Dickens, C. Pryor, C. Gao, A. Albalak, E. Augustine, W.Wang, S. Wright, L. Getoor. **A Mathematical Framework and a Suite of Learning Techniques for Neural-Symbolic Systems**. arXiv (2025)
- [12] Pascal Hitzler, Md. Kamruzzaman Sarker (Editors). **Neuro-Symbolic Artificial Intelligence** IOS Press, 2022.