Neuro-Symbolic Reasoning: A Roadmap of Unsolved Core Questions

Sahil Rajesh Dhayalkar
Arizona State University
sdhayalk@asu.edu

Abstract

Neuro-symbolic reasoning aims to combine the pattern-recognition power of deep learning with the structure and guarantees of symbolic logic, but progress is fragmented and largely architecture-driven. This paper takes a question-first perspective. We first review subsymbolic and symbolic approaches and propose a unifying taxonomy of neuro-symbolic paradigms: logic-guided learning, differentiable logic, program synthesis and neural program induction, and constraint-based training. Building on this taxonomy, we identify four cross-cutting challenge axes—optimization and stability under symbolic constraints, expressivity and compositional generalization, semantics and explainability, and data and symbol grounding. Along these axes, we formulate eight concrete open problems that cut across existing models and benchmarks. We conclude with a roadmap highlighting short-, medium-, and long-term directions. The result is a structured agenda for turning neuro-symbolic reasoning from a collection of techniques into a principled, scalable paradigm.

1. Introduction

Modern machine learning is split between subsymbolic and symbolic paradigms. Subsymbolic methods (Deep Neural Networks) achieve state-of-the-art performance in vision, language, and multimodal tasks via gradient-based optimization on large datasets [9, 25, 27, 28, 32, 36, 47], excelling at pattern recognition and function approximation. Conversely, classical AI's Symbolic frameworks (logics, rule-based systems, theorem proving) [42, 43, 46, 52] provide explicit semantics, compositional structure, and strong guarantees. Neuro-symbolic reasoning seeks to integrate neural representation learning with symbolic structure and inference [6, 7, 11, 13, 14].

Despite many hybrid architectures, the field lacks a shared view of the *core unresolved questions* for a principled neuro-symbolic paradigm. Existing work is mainly architecture-first/benchmark-first: introducing differentiable logic layers [38, 49, 59], logic-guided training [21, 29, 41], neural-guided program synthesis [15, 35, 45, 54], and constraint-based inference [2, 55, 60], demonstrating task-specific empirical gains. Less attention is given to foundational issues like gradient stability under symbolic constraints, conditions for provable compositional generalization, or the trade-offs between explanation faithfulness and performance [18, 34, 40, 48]. We argue that progress requires clarifying this conceptual and theoretical landscape rather than just bespoke architectures. We focus on four challenge axes: (i) optimization and stability with symbolic constraints, (ii) expressivity and compositional generalization, (iii) semantics, explainability, and performance trade-offs, and (iv) data, supervision, and symbol grounding, using them to motivate eight open problems. Our contributions are: (i) a taxonomy of neuro-symbolic learning; (ii) an analysis of current practices' shortcomings; and (iii) a roadmap of open problems for theoretically grounded neuro-symbolic reasoning.

2. Background and Taxonomy

We review subsymbolic learning and symbolic reasoning, then categorize existing neuro-symbolic approaches.

Subsymbolic learning: Deep neural networks, the dominant subsymbolic approach, represent knowledge in high-dimensional vectors, trained end-to-end via stochastic gradient descent (SGD) [25, 36, 51]. Architectures like CNNs, transformers, and GNNs [8, 27, 31, 32, 58] approximate functions from data, offering scalability, robustness, and universality. However, they lack explicit symbolic structure, show limited compositional generalization [3, 22, 34], and provide few formal guarantees for constraint satisfaction [40].

Symbolic reasoning: Symbolic AI uses discrete, structured formalisms (e.g., first-order logic, production rules, constraint systems) [14, 42, 43, 46]. These support compositional representation, transparent inference, and offer guarantees like soundness and decidability. They underpin automated reasoning [4, 52] and logic programming [23, 37]. Limitations include brittleness to noise, difficulty learning from raw data, and high engineering cost at scale.

Taxonomy of neuro-symbolic paradigms: Neuro-symbolic learning seeks to combine these strengths [6, 7, 11, 13]. We define four paradigms:

Logic-guided learning. Symbolic knowledge (rules, structure) acts as a teacher or regularizer, shaping training signals or architectures without making logic differentiable [8, 10, 14, 21, 29, 41]. Optimization remains unconstrained.

Differentiable logic. Logic connectives and quantifiers are relaxed into continuous operators, letting gradients flow through ”soft” logical expressions [20, 38, 49, 57]. Logical constraints are optimized alongside standard losses using truth values in $[0, 1]$, risking approximation error or instability from saturation.

Program synthesis and neural program induction. Symbolic programs or logical forms are the learning targets. Neural models induce programs from input/supervision, which symbolic executors run [15–17, 19, 35, 45, 54]. The symbolic part provides semantics/compositionality; the neural part provides priors/heuristics. Challenges include search complexity and sparse feedback.

Constraint-based training and inference. Logical or structural constraints are embedded into optimization/inference. Examples include differentiable optimization layers [1, 2], differentiable SAT solvers [21, 60], structured prediction with constraints [5, 55, 56], or projecting neural outputs onto constraint sets. These import tools from constrained optimization.

Limitations of current practice: Work is highly empirical and model-specific. Despite many techniques for coupling logic and networks, there are few general principles on when integration helps, how to ensure stable optimization with discrete structure, or how to quantify the semantic faithfulness of symbolic explanations/constraints [18, 34, 40, 48]. This taxonomy will identify cross-cutting challenge axes and core open questions.

3. Core Challenge Axes in Neuro-Symbolic Reasoning

The taxonomy in Section 2 highlights the diverse ways symbolic structure integrates with gradient-based learning. Across all paradigms, four challenge axes emerge—optimization and stability, compositional expressivity, semantics and explainability, data and symbol grounding—which motivate the open-problem list in Section 4.

3.1. Optimization and Stability with Symbolic Constraints

Gradient-based learning assumes smooth, unconstrained objectives. Neuro-symbolic methods introduce non-smooth, non-convex, or discrete logical/structural constraints. This occurs through logic-guided loss terms [29, 41], differentiable logic using fuzzy operators for losses [21, 38, 49, 57], or constraint-based layers/projection methods [1, 2, 60]. The resulting loss landscape structure can cause training instability, poor convergence, or degenerate solutions.

Soft relaxations, used in differentiable logics, map truth values in $[0, 1]$ via t-norms, s-norms, and aggregation operators [20, 49, 57]. While enabling end-to-end optimization, the losses can be non-linear and suffer *saturation*: gradients vanish once a formula is nearly satisfied [57]. Different fuzzy logic choices exhibit varied optimization/approximation properties, with naive choices potentially biasing learning [57]. Moreover, conflicting constraints can cause standard SGD to oscillate or converge to solutions with persistent violations [21, 24].

Fundamental questions arise: When do symbolic constraints improve generalization, and when do they destabilize training? How can relaxations and penalty schemes be designed for both Boolean faithfulness and optimization amenability? What principled alternatives exist beyond heuristic combinations of task and constraint loss (e.g., projected gradient methods)? Answering these requires a joint view of learning theory, optimization, and logical semantics, currently lacking in practice.

3.2. Expressivity and Compositional Generalization

A main goal of neuro-symbolic reasoning is to unlock compositional generalization, difficult for purely subsymbolic models [3, 16, 17, 30, 33, 34, 50]. Classical symbols are inherently compositional, aligning with human systematic generalization. Standard neural architectures often fail on systematicity benchmarks like SCAN [30, 34, 50] without engineered inductive biases [3, 33, 44].

Hybrid models (with symbolic layers, program-like representations, or logical constraints) are claimed to improve compositional generalization [19, 39, 61]. However, the specific causal aspect is unclear: is it discrete variables, program schemas, or explicit variable binding/quantification? Conversely, overly rigid symbolic vocabularies can underfit or force brittle generalization if misaligned with concepts.

Theoretically, we lack an expressivity characterization for neuro-symbolic models that captures both components. Existing theorems for neural networks and logical systems apply in isolation, not to their combination’s effect on hypothesis class, sample complexity, or inductive biases. We also lack agreed-upon compositionality tests tailored to hybrid architectures. Developing these characterizations and benchmarks is essential for principled design of “just expressive enough” symbolic interfaces.

3.3. Semantics, Explainability, and Performance Trade-offs

Symbolic structure is also invoked for transparent, controllable models. Logical rules, programs, and proofs have explicit semantics, serving as human-readable explanations [13, 18]. Neuro-symbolic architectures generate symbolic traces (e.g., proof trees) accompanying neural outputs [16, 17, 19, 20, 35, 38] and enforce consistency with interpretable background knowledge [21, 24, 29]. This raises questions in semantics, faithfulness, and performance.

First, symbolic artifacts are not automatically *faithful* accounts of computation. A neural module may emit correlating logical forms without mediating the decision, resulting in post-hoc or adversarially optimized explanations [18, 48]. Second, enforcing strict symbolic constraints can reduce empirical performance if constraints are misspecified or misaligned with the task loss [21, 24]. This creates an accuracy–consistency trade-off. Third, differentiable logic semantics are often approximate: soft truth values break classical equivalences [57], complicating interpretation of "satisfied" constraints.

Addressing these requires formal notions of *semantic faithfulness* for explanations, metrics quantifying how symbolic traces track causal mechanisms, and theoretical understanding of how enforcement strategies (hard vs soft) affect the accuracy–consistency trade-off. Empirical protocols must jointly evaluate predictive performance and explanation quality.

3.4. Data, Supervision, and Symbol Grounding

Neuro-symbolic reasoning relies on symbolic knowledge (theories, rules, schemas). In practice, this knowledge is often incomplete, noisy, or absent. *Symbol grounding*—aligning abstract symbols with data patterns—is a major challenge [6, 14, 26]. Many methods assume hand-crafted knowledge or dense symbolic supervision (e.g., gold logical forms), limiting scalability.

When supervision is scarce, hybrid systems must infer subsymbolic parameters and symbolic structure from weak signals: language, approximate patterns, partial constraints, or human feedback [14, 19, 41]. This raises questions of identifiability (which theories fit?), robustness (how to revise rules?), and knowledge interaction. Naive integration of symbolic knowledge bases risks amplifying encoded biases and errors.

A principled treatment would connect inductive logic programming, program synthesis, and representation learning, providing guarantees about coherence and convergence of the symbolic component as data accumulates. We have many techniques for injecting/extracting symbols, but few systematic methods for *maintaining* and *evolving* symbolic knowledge that respects both logic and empirical evidence. The open problems in Section 4 frame these challenges for cross-disciplinary progress.

4. A Structured Open-Problem List

Building on Section 3, this section outlines eight open problems that together form a roadmap spanning all architectures and paradigms discussed in Section 2.

Open Problem 1: When Do Symbolic Constraints Help or Hurt Optimization? All major paradigms incorporate symbolic constraints into the loss or inference objective [1, 21, 29, 38, 49, 60]. Empirically, constraints act as inductive biases, improving generalization, but can destabilize training when conflicting with data or capacity [21, 24]. The open problem is characterizing the regimes (data size, noise, constraint structure) where symbolic constraints are beneficial or harmful. We seek analogues of classical regularization theory: conditions for provably reduced generalization error and lower bounds on when constraints cannot help. Progress would make constraint design a principled component of model selection.

Open Problem 2: How Can We Design Differentiable Logic with Provable Approximation and Gradient Properties? Differentiable logic relies on continuous relaxations of Boolean connectives via fuzzy or probabilistic semantics [20, 38, 49, 57]. These operators vary widely in semantic approximation, compositional behavior, and gradient properties. The key open problem is to design relaxation families with *provable* guarantees: bounded approximation error to Boolean truth tables, controlled gradient magnitudes or Lipschitz constants, and convergence results linking relaxed satisfaction to discrete satisfaction. Such results would enable selecting operators with predictable optimization behavior rather than relying on heuristic choices that bias learning or cause saturation.

Open Problem 3: What Minimal Symbolic Interfaces Enable Compositional Generalization? Neuro-symbolic models aim for systematic, compositional generalization [3, 30, 33, 34, 50]. Hybrid architectures (program-like representations) show promise [19, 39, 61], but the essential symbolic aspects remain unclear. The open problem is identifying *minimal* symbolic scaffolds (types, predicates, grammars) sufficient to guarantee specific forms of compositional generalization for a task family. Ideally, we would relate symbolic interface properties (e.g., variable binding) to formal generalization guarantees, guiding the design of "just expressive enough" layers.

Open Problem 4: How Do We Measure Semantic Faithfulness of Neuro-Symbolic Explanations? Neuro-symbolic systems use symbolic artifacts (rules, proofs) as model explanations [13, 19, 35, 38]. Existing evaluation relies on human judgment or proxy metrics, not directly testing if the symbolic explanation reflects the model's internal computation [18, 48]. The open problem is developing formal *semantic faithfulness* notions: metrics quantifying how much an explanation constrains model behavior under counterfactuals or distribution shifts. This requires experiments to distinguish plausible explanations from mechanistic ones, characterizing the faithfulness/simplicity trade-off. Without such measures, claims of "interpretable" models are hard to substantiate.

Open Problem 5: Can We Learn Symbolic Knowledge from Data While Preserving Logical Coherence? Most work assumes symbolic knowledge (theories, rules) is external [6, 14]. In practice, this knowledge is often incomplete/noisy; hybrid systems must *learn*, refine, and debug symbolic components from data [19, 41]. The open problem is devising learning dynamics that jointly update subsymbolic parameters and symbolic structures while preserving logical properties. This connects inductive logic programming and representation learning, raising questions about identifiability and safe rule revision. A solution would allow symbolic knowledge bases to evolve with data.

Open Problem 6: How Should We Evaluate Neuro-Symbolic Systems Beyond Task Accuracy? ML benchmarks emphasize task accuracy/reward. For neuro-symbolic systems, this is insufficient. Constraint satisfaction, sample efficiency, compositional generalization, explanation quality, and semantic consistency are needed [18, 34, 40]. The open problem is designing evaluation protocols and benchmark suites that explicitly measure these dimensions, stress-testing hybrid models beyond raw performance. This requires multi-objective metrics and community-agreed "gold standards" for consistency, as current benchmarks make comparison difficult.

Open Problem 7: What Are the Right Theoretical Abstractions for Neuro-Symbolic Models? Hybrid systems span a heterogeneous landscape (e.g., differentiable logic circuits, constrained deep networks) [2, 38, 45, 49, 60]. We lack unifying abstractions analogous to graphical models [12]. The open problem is developing mathematical frameworks (e.g., differentiable constraint systems) to capture a broad class of architectures under a common formalism and proving general theorems about expressivity, learnability, and robustness. Such abstractions would facilitate insight transfer and generalized reasoning.

Open Problem 8: How Do We Scale Neuro-Symbolic Methods Without Losing Their Benefits? Many neuro-symbolic prototypes work on small-scale tasks/restricted logical theories [19, 38, 39, 61]. Scaling to large models/knowledge bases introduces computational bottlenecks (e.g., search) and compromises (e.g., relaxation) that risk eroding key benefits: interpretability and controllability. The open problem is identifying architectural/algorithmic patterns that retain symbolic benefits at scale: modular knowledge bases, amortized inference, and certified reasoning. Addressing this is crucial for neuro-symbolic reasoning to impact mainstream large-scale ML.

5. Roadmap and Outlook

The open problems in Section 4 define a research agenda complementing architecture-driven progress with question-driven inquiry. In the short term, systematic studies of differentiable logic operators and constraint-optimization schemes are valuable, including comparisons of their approximation and gradient properties [20, 49, 57]. Developing small, controlled benchmarks that explicitly test compositional generalization, semantic faithfulness, and constraint satisfaction—beyond just task accuracy—would provide a shared yardstick for evaluation [18, 30, 34, 50].

In the medium term, we expect convergence between learning theory, optimization, and logic, leading to guarantees on when symbolic constraints help/hurt, how scaffolds affect generalization, and how to safely learn/revise symbolic knowledge from data [14, 40]. Progress will likely depend on unifying theoretical abstractions for neuro-symbolic models, analogous to graphical models and PAC theory in earlier AI waves [12, 53].

In the longer term, addressing these questions could make neuro-symbolic reasoning a mainstream paradigm: enabling toolchains for specifying knowledge and constraints, debuggers spanning neural parameters and symbolic rules, and curricula teaching hybrid modeling. Reaching this point requires shifting from ad hoc experimentation to a principled, question-first perspective. The eight open problems provide a scaffold for organizing future work and building a theoretically grounded, empirically careful, and reliable neuro-symbolic reasoning community.

6. Conclusion

Neuro-symbolic reasoning intersects powerful subsymbolic models and symbolic formalisms. This paper adopts a question-first stance, moving beyond new architectures. We introduced a taxonomy, identified four challenge axes, and distilled eight open problems spanning optimization, compositional generalization, semantics/explainability, and symbol grounding. This roadmap outlines a research agenda to transform neuro-symbolic reasoning from isolated techniques into a principled, scalable paradigm. We hope this encourages work coupling empirical innovation with theoretical clarity, developing benchmarks for hybrid capabilities, and creating tools that treat symbolic knowledge as a first-class citizen in mainstream machine learning.

References

[1] Akshay Agrawal, Brandon Amos, Shane Barratt, Stephen Boyd, Steven Diamond, and Zico Kolter. Differentiable convex optimization layers, 2019. 2, 3

[2] Brandon Amos and J. Zico Kolter. Optnet: differentiable optimization as a layer in neural networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, page 136–145. JMLR.org, 2017. 1, 2, 4

[3] Dzmitry Bahdanau, Shikhar Murty, Michael Noukhovitch, Thien Huu Nguyen, Harm de Vries, and Aaron Courville. Systematic generalization: What is required and can it be learned?, 2019. 1, 2, 3

[4] Clark Barrett and Cesare Tinelli. Satisfiability modulo theories. In *Handbook of Model Checking*, pages 305–343. Springer International Publishing, 2018. 1

[5] David Belanger and Andrew McCallum. Structured prediction energy networks. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, page 983–992. JMLR.org, 2016. 2

[6] Tarek R. Besold, Artur S. d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro M. Domingos, Pascal Hitzler, Kai-Uwe Kühnberger, L. Lamb, Daniel Lowd, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, Hoifung Poon, and Gerson Zaverucha. Neural-symbolic learning and reasoning: A survey and interpretation. *ArXiv*, abs/1711.03902, 2017. 1, 2, 3, 4

[7] Bikram Pratim Bhuyan, Amar Ramdane-Cherif, Thipendra P. Singh, and Ravi Tomar. The emergence of neuro-symbolic artificial intelligence. In *Neuro-Symbolic Artificial Intelligence*, pages 1–25. Springer Nature Singapore, 2024. 1, 2

[8] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021. 1, 2

[9] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2020. Curran Associates Inc. 1

[10] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2990–2999, New York, New York, USA, 2016. PMLR. 2

[11] Brandon C. Colelough and William Regli. Neuro-symbolic ai in 2024: A systematic review, 2025. 1, 2

[12] Darwiche, Adnan. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009. 4

[13] Artur S. d'Avila Garcez, Marco Gori, L. Lamb, Luciano Serafini, Michael Spranger, and S. Tran. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *FLAP*, 6:611–632, 2019. 1, 2, 3

[14] De Raedt, Luc and Kersting, Kristian and Natarajan, Sriraam and Poole, David. Statistical relational artificial intelligence: Logic, probability, and computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 10(2):1–189, 2016. 1, 2, 3, 4

[15] Jacob Devlin, Jonathan Uesato, Surya Bhupatiraju, Rishabh Singh, Abdel-rahman Mohamed, and Pushmeet Kohli. Robustfill: neural program learning under noisy i/o. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, page 990–998. JMLR.org, 2017. 1, 2

[16] Sahil Rajesh Dhayalkar. A constructive framework for nondeterministic automata via time-shared, depth-unrolled feedforward networks, 2025. 2, 3

[17] Sahil Rajesh Dhayalkar. Symbolic feedforward networks for probabilistic finite automata: Exact simulation and learnability, 2025. 2, 3

[18] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv: Machine Learning*, 2017. 1, 2, 3, 4

[19] Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sable-Meyer, Luc Cary, Lucas Morales, Luke Hewitt, Armando Solar-Lezama, and Joshua B. Tenenbaum. Dreamcoder: Growing generalizable, interpretable knowledge with wake-sleep bayesian program learning, 2020. 2, 3, 4

[20] Richard Evans and Edward Grefenstette. Learning explanatory rules from noisy data. *J. Artif. Int. Res.*, 61(1):1–64, 2018. 2, 3, 4

[21] Marc Fischer, Mislav Balunovic, Dana Drachsler-Cohen, Timon Gehr, Ce Zhang, and Martin Vechev. DL2: Training and querying neural networks with logic. In *Proceedings of the 36th International Conference on Machine Learning*, pages 1931–1941. PMLR, 2019. 1, 2, 3

[22] Daniel Furrer, Marc van Zee, Nathan Scales, and Nathanael Schärli. Compositional generalization in semantic parsing: Pre-training vs. specialized architectures, 2021. 1

[23] Gebser, Martin and Kaufmann, Benjamin and Krennwallner, Thomas and Schaub, Torsten. *Answer Set Solving in Practice*. Springer Science & Business Media, 2012. 1

[24] Eleonora Giunchiglia, Mihaela Catalina Stoian, and Thomas Lukasiewicz. Deep learning with logical constraints. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 5478–5485. International Joint Conferences on Artificial Intelligence Organization, 2022. Survey Track. 2, 3

[25] Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron. *Deep Learning*. MIT Press, 2016. 1

[26] Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1):335–346, 1990. 3

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[28] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504–507, 2006. 1

[29] Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. Harnessing deep neural networks with logic rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2410–2420, Berlin, Germany, 2016. Association for Computational Linguistics. 1, 2, 3

[30] Daniel Keysers, Nathanael Schärli, Nathan Scales, Hylke Buisman, Daniel Furrer, Sergii Kashubin, Nikola Momchev, Danila Sinopalnikov, Lukasz Stafiniak, Tibor Tihon, Dmitry Tsarkov, Xiao Wang, Marc van Zee, and Olivier Bousquet. Measuring compositional generalization: A comprehensive method on realistic data, 2020. 2, 3, 4

[31] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. 1

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. 1

[33] Brenden M. Lake. *Compositional generalization through meta sequence-to-sequence learning*. Curran Associates Inc., Red Hook, NY, USA, 2019. 2, 3

[34] Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40:e253, 2017. 1, 2, 3, 4

[35] Guillaume Lample and François Charton. Deep learning for symbolic mathematics, 2019. 1, 2, 3

[36] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015. 1

[37] Lloyd, John W. *Foundations of Logic Programming*. Springer Science & Business Media, 1987. 1

[38] Robin Manhaeve, Sebastijan Dumančić, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Deepproblog: Neural probabilistic logic programming, 2018. 1, 2, 3, 4

[39] Jiayuan Mao, Chuang Gan, Pushmeet Kohli, Joshua B. Tenenbaum, and Jiajun Wu. The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision, 2019. 2, 3, 4

[40] Gary Marcus. Deep learning: A critical appraisal, 2018. 1, 2, 4

[41] Giuseppe Marra, Francesco Giannini, Michelangelo Diligenti, and Marco Gori. Integrating learning and reasoning with deep logic models, 2019. 1, 2, 3, 4

[42] John McCarthy. Mathematical logic in artificial intelligence. 1989. 1

[43] Allen Newell. Physical symbol systems. *Cognitive Science*, 4(2):135–183, 1980. 1

[44] Santiago Ontañón, Joshua Ainslie, Zachary Fisher, and Vaclav Cvicek. Making transformers solve compositional tasks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3591–3607, Dublin, Ireland, 2022. Association for Computational Linguistics. 2

[45] Emilio Parisotto, Abdel-rahman Mohamed, Rishabh Singh, Lihong Li, Dengyong Zhou, and Pushmeet Kohli. Neuro-symbolic program synthesis. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. 1, 2, 4

[46] David L. Poole and Alan K. Mackworth. *Artificial Intelligence: Foundations of Computational Agents*. Cambridge University Press, New York, 2010. 1

[47] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019. 1

[48] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. 1, 2, 3, 4

[49] Tim Rocktäschel and Sebastian Riedel. End-to-end differentiable proving. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. 1, 2, 3, 4

[50] Laura Ruis, Jacob Andreas, Marco Baroni, Diane Bouchacourt, and Brenden M. Lake. A benchmark for systematic generalization in grounded language understanding. 2, 3, 4

[51] Rumelhart, David E and Hinton, Geoffrey E and Williams, Ronald J. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. 1

[52] Russell, Stuart J and Norvig, Peter. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3rd edition, 2010. 1

[53] Shalev-Shwartz, Shai and Ben-David, Shai. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014. 4

[54] Armando Solar-Lezama. The sketching approach to program synthesis. In *Proceedings of the 7th Asian Symposium on Programming Languages and Systems*, page 4–13, Berlin, Heidelberg, 2009. Springer-Verlag. 1, 2

[55] Aviv Tamar, Yi Wu, Garrett Thomas, Sergey Levine, and Pieter Abbeel. Value iteration networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, page 2154–2162, Red Hook, NY, USA, 2016. Curran Associates Inc. 1, 2

[56] Ben Taskar, Carlos Guestrin, and Daphne Koller. Max-margin markov networks. In *Advances in Neural Information Processing Systems*. MIT Press, 2003. 2

[57] Emile van Krieken, Erman Acar, and Frank van Harmelen. Analyzing differentiable fuzzy logic operators. *Artificial Intelligence*, 302:103602, 2022. 2, 3, 4

[58] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. 1

[59] Gustav Šourek, Vojtěch Aschenbrenner, Filip Železný, Steven Schockaert, and Ondřej Kuželka. Lifted relational neural networks: efficient learning of latent relational structures. *J. Artif. Int. Res.*, 62(1):69–100, 2018. 1

[60] Po-Wei Wang, Priya L. Donti, Bryan Wilder, and Zico Kolter. Satnet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver. In *International Conference on Machine Learning*, 2019. 1, 2, 3, 4

[61] Kexin Yi, Jiajun Wu, Chuang Gan, Antonio Torralba, Pushmeet Kohli, and Joshua B. Tenenbaum. Neural-symbolic vqa: disentangling reasoning from vision and language understanding. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 1039–1050, Red Hook, NY, USA, 2018. Curran Associates Inc. 2, 3, 4