Complex Reasoning over Vision and Language –Leveraging Neurosymbolic AI

Parisa Kordjamshidi

Michigan State University, East Lansing, USA
Computer Science and Engineering Department
kordjams@msu.edu

Abstract

Recent research highlights the lack of reliability of large language models (LLMs) in tasks requiring complex reasoning. While they can produce impressively fluent text in response to prompts, they can fail on basic reasoning skills, such as recognizing that left is the opposite of right. They struggle even more with grounding such concepts in real-world contexts involving perception and action. Addressing real-world problems, however, typically requires models composed of multiple interdependent learners, with strong capabilities for composition and reasoning. In this talk, I will discuss the reasoning challenges of LLMs and discuss how symbolic representations can enhance neural models by enabling Spatial and Compositional Reasoning over complex linguistic structures, grounding language in visual perception, integrating multiple modalities, and dealing with uncertainty. I will overview recent research in Neurosymbolic (NeSy) modeling and emphasize the need for community-driven libraries to advance this direction. As part of this effort, I will introduce the DomiKnowS framework developed by my team, which combines symbolic and sub-symbolic representations to tackle complex, AI-complete problems, integrating symbolic and logical knowledge seamlessly into deep models and LLMs through a range of underlying algorithms.

Code: Heterogeneous Learning and Reasoning Lab: —
https://github.com/HLR/

DomiKnowS framework —
https://github.com/HLR/DomiKnowS

Introduction

Recent research demonstrates that large language models (LLMs) lack consistent reliability when tasked with complex reasoning. Although they can produce fluent, well-structured natural language in response to prompts, they often fail on basic reasoning tasks—for example, understanding that left is the opposite of right—and struggle even more to ground such concepts in real-world contexts involving perception and action (Zhang et al. 2025a,b). Their performance is highly sensitive to prompt structure, and even slight paraphrasing of semantically equivalent questions can lead to inconsistent or unstable answers. This fragility arises in part from their learning from next-token prediction as the core mechanism for generating outputs. While state-of-the-art LLMs are steadily improving on existing benchmarks, a fundamental question remains: how can these models *generalize* to novel situations and address problems that require more than pattern matching or memorization on one hand, while avoiding *overgeneralization* and related issues such as hallucination on the other hand?

In this talk, I argue for the necessity of integrative paradigms that combine symbolic and subsymbolic representations, thereby enabling both pattern recognition and explicit reasoning. I will first discuss current evaluation efforts and the limitations they reveal in LLM reasoning. I will then highlight key dimensions of reasoning including spatial reasoning, compositional reasoning, and reasoning under uncertainty where existing models fall short. Following this, I will discuss how neural and symbolic processing modules can be jointly designed and effectively integrated to tackle these challenges.

Finally, I will advocate for a community-wide effort to develop general frameworks and libraries that support neurosymbolic (NeSy) modeling. I will briefly introduce one such framework developed by my team, outlining its key components and the requirements for advancing this line of research.

Empirical Evaluation of Complex Reasoning

I will review recent evaluation efforts on the reasoning capabilities of large language models, focusing on compositional reasoning (Kamali, Barezi, and Kordjamshidi 2025), spatial reasoning (Premsri and Kordjamshidi 2025; Mirzaee et al. 2021), and reasoning under uncertainty (Nafar, Venable, and Kordjamshidi 2025). These studies examine large models across both language and vision modalities. A central challenge revealed by these evaluations is the limited ability of current models to generalize to novel situations, particularly in three cases: (a) when they encounter novel combinations of concepts in the visual modality that are not present in training data (Kamali, Barezi, and Kordjamshidi 2025; Kamali and Kordjamshidi 2023, 2025); (b) when processing nested structures and complex linguistic constructs; and (c) when reasoning about entirely new basic concepts, such as spatial frames of reference, that are absent from large-scale internet training corpora (**?**Tanawan Premsri 2025) and require interactions with real world.

Integration of Symbolic and Subsymbolic AI

The interplay of neurosymbolic can be explained by the concept of System 1 and System 2 thinking described in (Kahneman 2011). Research in this field aims to create an ideal integration that seamlessly supports ”thinking fast and slow” (Booch et al. 2021; Fabiano et al. 2023). Here, neural processing can be seen as a manifestation of System 1, while System 2 corresponds to the slower, more deliberate symbolic reasoning. Different methods for the integration of symbolic reasoning and neural programming have been explored such as employing logical constraint satisfaction, integer linear programming, differentiable reasoning, probabilistic logic programming. I will discuss how integrating symbolic processing with large language models can enhance their reasoning capabilities. I will describe various ways in which symbolic and subsymbolic components can interact and complement each other. This includes efforts to incorporate logical knowledge and constraints during training, differentiable program execution, as well as approaches that introduce logical or probabilistic inference.

Integration of Symbolic Knowledge for Training

One approach that has been explored in various ways is using logical knowledge as a source of supervision during training (Xu et al. 2018; Nandwani et al. 2019; **?**). In this paradigm, the model operates as a neural network, but supervision derived from logical reasoning encourages the model to form abstractions over the data that support reasoning over logical reasoning patterns. Although this hypothesis has not yet been fully evaluated mechanistically, empirical studies demonstrate that it can improve generalizability in complex, previously unobserved reasoning tasks.

The primary technical approach involves adding a term to the loss function that measures violations of logical constraints, which the training process then seeks to minimize. Some techniques to implement this approach include using a differentiable loss that surrogates the logical violations or using sampling and reinforcement. In the first case, these methods often employ soft logical surrogates of logical operations which has been shown to significantly affect training effectiveness in some case studies. I will discuss these techniques and present results in the contexts of spatial reasoning (Premsri and Kordjamshidi 2025; Nafar, Venable, and Kordjamshidi 2024) and reasoning over uncertain text. Although this type of technique follows a neural architecture training, the results indicate that the way supervision from the data and logic is provided to the models can impact their internal abstractions and generalization capabilities. Another approach for NeSy training is to map the raw data including vision and language to an executable and differentiable program where the program execution will provide the solution/outputs (Hsu et al. 2023). The errors can be propagated back to the neural models after the differentiable execution. This approach can be seen as a two phases—neural and symbolic processing—one after another, and it has been employed in a variety of problem settings, particularly in language grounding and concept learning (Kamali, Barezi, and Kordjamshidi 2025).

Symbolic Reasoning for Inference

Using logic as a source of supervision during training can help models capture underlying logical patterns as discussed above, but it does not guarantee that outputs strictly adhere to hard logical constraints. This is problematic for sensitive and safety-critical application domains. Consequently, depending on the sensitivity of the application, explicit formal reasoning can be applied at inference time using algorithms that follow explicit reasoning and imposing the domain constraints. A widely used method over the years has been integer linear programming (Roth and Srikumar 2017), while alternative approaches include probabilistic logical reasoning or constraint satisfaction techniques. With the advent of large language models and their ability to translate language into executable code, new avenues for the integration of symbolic reasoning in neural modeling have opened up. Large language models now offer the potential to generate concrete, verifiable programs (Faghihi et al. 2024). The domain knowledge can be formalized, domain predicates can be invented as required and formal tools can be employed to verify the outputs of the models using self-verification and correction. These new capabilities alleviate many historical limitations of classical symbolic reasoning models and provide the capacity to exploit their strengths. I will provide an overview of these approaches and discuss their application to compositional reasoning tasks, including visual question answering and grounding referring expressions in the visual modality and self-correction for text to image generation (Tanawan Premsri 2025; Kamali, Barezi, and Kordjamshidi 2025; Kamali and Kordjamshidi 2025).

Democratizing Neurosymbolic Modeling

While there are algorithms and models that explore the interplay between neural and symbolic paradigms (Kautz 2022; De Raedt, Kimmig, and Toivonen 2007; Huang et al. 2021), pursuing hybrid neurosymbolic modeling remains relatively less popular in the AI community compared to neural modeling. In this talk, I argue that one key barrier is the lack of libraries that facilitate modeling and programming for neurosymbolic systems. Effective neurosymbolic modeling requires expressing the symbolic structure of data in terms of concepts and relationships (Huang et al. 2021; Kordjamshidi, Roth, and Kersting 2019), yet most current implementations operate at the algorithmic level and are highly task-specific. I will provide a discussion on a) Identifying the main components of existing NeSy frameworks, b) Comparison of frameworks across the identified facets, c) Highlighting the requirements for the next generation of NeSy frameworks, building upon the drawbacks of the current systems and the possible interplays between the neural and symbolic components. I advocate for community efforts to advance this direction through the development of general-purpose libraries. To illustrate this, I will present DomiKnowS (Faghihi et al. 2021), a library developed by my team to support flexible and reusable neurosymbolic modeling.

Future Directions

I advocate for the neurosymbolic modeling to learn and reason in complex situations, particularly in environments that include *perception* from multiple modalities such as natural language and vision in addition to *reasoning* over novel and complex situations. However, one key bottleneck limiting progress in this area is the lack of programming languages and libraries specifically designed to support such modeling. I argue that an effective neurosymbolic paradigm should offer an expressive language for data and knowledge representation, facilitate the transformation of raw data into structured forms, manage the inherent uncertainty of neural components, and enable seamless integration between symbolic and subsymbolic approaches in both learning and decision-making processes by supporting a variety of algorithms. Designing new libraries to support hybrid programming can pave the way for more reliable and trustworthy AI systems, while also making this technology more accessible for both fundamental research and real-world applications. This idea has the potential to advance the new wave of Agentic AI (Qu et al. 2025), paving the way for more robust and trustworthy models.

Acknowledgments

This presented research is partially supported by the Office of Naval Research (ONR) grant N00014-23-1-2417. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Office of Naval Research.

References

Booch, G.; Fabiano, F.; Horesh, L.; Kate, K.; Lenchner, J.; Linck, N.; Loreggia, A.; Murgesan, K.; Mattei, N.; Rossi, F.; and Srivastava, B. 2021. Thinking Fast and Slow in AI. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17): 15042–15046.

De Raedt, L.; Kimmig, A.; and Toivonen, H. 2007. ProbLog: a probabilistic prolog and its application in link discovery. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, 2468–2473. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Fabiano, F.; Pallagani, V.; Ganapini, M. B.; Horesh, L.; Loreggia, A.; Murugesan, K.; Rossi, F.; and Srivastava, B. 2023. Plan-SOFAI: A Neuro-Symbolic Planning Architecture. In *Neuro-Symbolic Learning and Reasoning in the era of Large Language Models*.

Faghihi, H. R.; Guo, Q.; Uszok, A.; Nafar, A.; Raisi, E.; and Kordjamshidi, P. 2021. DomiKnowS: A Library for Integration of Symbolic Domain Knowledge in Deep Learning.

Faghihi, H. R.; Nafar, A.; Uszok, A.; Karimian, H.; and Kordjamshidi, P. 2024. Prompt2DeModel: Declarative Neuro-Symbolic Modeling with Natural Language. arXiv:2407.20513.

Hsu, J.; Mao, J.; Tenenbaum, J. B.; and Wu, J. 2023. What's Left? Concept Grounding with Logic-Enhanced Foundation Models. arXiv:2310.16035.

Huang, J.; Li, Z.; Chen, B.; Samel, K.; Naik, M.; Song, L.; and Si, X. 2021. Scallop: From Probabilistic Deductive Databases to Scalable Differentiable Reasoning. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 25134–25145. Curran Associates, Inc.

Kahneman, D. 2011. *Thinking, fast and slow*. macmillan.

Kamali, D.; Barezi, E. J.; and Kordjamshidi, P. 2025. NeSyCoCo: A Neuro-Symbolic Concept Composer for Compositional Generalization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(4): 4184–4193.

Kamali, D.; and Kordjamshidi, P. 2023. Syntax-Guided Transformers: Elevating Compositional Generalization and Grounding in Multimodal Environments. In Hupkes, D.; Dankers, V.; Batsuren, K.; Sinha, K.; Kazemnejad, A.; Christodoulopoulos, C.; Cotterell, R.; and Bruni, E., eds., *Proceedings of the 1st GenBench Workshop on (Benchmarking) Generalisation in NLP*, 130–142. Singapore: Association for Computational Linguistics.

Kamali, D.; and Kordjamshidi, P. 2025. NePTune: A Neuro-Pythonic Framework for Tunable Compositional Reasoning on Vision-Language. arXiv:2509.25757.

Kautz, H. 2022. The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture. *AI Magazine*, 43(1): 105–125.

Kordjamshidi, P.; Roth, D.; and Kersting, K. 2019. Declarative Learning-Based Programming as an Interface to AI Systems. *ArXiv*, abs/1906.07809.

Mirzaee, R.; Rajaby Faghihi, H.; Ning, Q.; and Kordjamshidi, P. 2021. SPARTQA: A Textual Question Answering Benchmark for Spatial Reasoning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4582–4598. Online: Association for Computational Linguistics.

Nafar, A.; Venable, K. B.; and Kordjamshidi, P. 2024. Teaching Probabilistic Logical Reasoning to Transformers. In Graham, Y.; and Purver, M., eds., *Findings of the Association for Computational Linguistics: EACL 2024*, 1615–1632. St. Julian's, Malta: Association for Computational Linguistics.

Nafar, A.; Venable, K. B.; and Kordjamshidi, P. 2025. Reasoning over Uncertain Text by Generative Large Language Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(23): 24911–24920.

Nandwani, Y.; Pathak, A.; Mausam; and Singla, P. 2019. A Primal Dual Formulation For Deep Learning With Constraints. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Premsri, T.; and Kordjamshidi, P. 2025. Neuro-symbolic Training for Reasoning over Spatial Language. In Chiruzzo, L.; Ritter, A.; and Wang, L., eds., *Findings of the Association for Computational Linguistics: NAACL 2025*, 2395–2414. Albuquerque, New Mexico: Association for Computational Linguistics. ISBN 979-8-89176-195-7.

Qu, X.; Damoah, A.; Sherwood, J.; Liu, P.; Jin, C. S.; Chen, L.; Shen, M.; Aleisa, N.; Hou, Z.; Zhang, C.; Gao, L.; Li, Y.; Yang, Q.; Wang, Q.; and Souza, C. D. 2025. A Comprehensive Review of AI Agents: Transforming Possibilities in Technology and Beyond. arXiv:2508.11957.

Roth, D.; and Srikumar, V. 2017. Integer Linear Programming formulations in Natural Language Processing. In Klementiev, A.; and Specia, L., eds., *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*. Valencia, Spain: Association for Computational Linguistics.

Tanawan Premsri, P. K. 2025. FoR-SALE: Frame of Reference-guided Spatial Adjustment in LLM-based Diffusion Editing, under review.

Xu, J.; Zhang, Z.; Friedman, T.; Liang, Y.; and Van den Broeck, G. 2018. A Semantic Loss Function for Deep Learning with Symbolic Knowledge. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 5502–5511. PMLR.

Zhang, Y.; Ma, T.; Wang, Z.; Qiao, Y.; and Kordjamshidi, P. 2025a. Vision-and-Language Navigation with Analogical Textual Descriptions in LLMs. In Christodoulopoulos, C.; Chakraborty, T.; Rose, C.; and Peng, V., eds., *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 15028–15036. Suzhou, China: Association for Computational Linguistics. ISBN 979-8-89176-332-6.

Zhang, Y.; Xu, Z.; Shen, Y.; Kordjamshidi, P.; and Huang, L. 2025b. SPARTUN3D: Situated Spatial Understanding of 3D World in Large Language Model. In *The Thirteenth International Conference on Learning Representations*.