Jiani Huang: Research Statement

Modern machine learning has revolutionized problem-solving across diverse fields, including software engineering, scientific discovery, and medicine. With advances in foundation models for language, image, and multi-modal data, it is becoming feasible for end users to accomplish complex tasks that would otherwise require significant expertise and resources.

Yet, despite these remarkable strides, deep learning faces many limitations. Importantly, it struggles with problems requiring structure, logic, and planning—areas where traditional symbolic reasoning excels. In his 2011 classic *Thinking Fast and Slow*, Kahneman describes human cognition as an interplay between an intuitive, associative "System 1," akin to neural networks, and a deliberative, logical "System 2," resembling symbolic reasoners. Combining the complementary strengths of these two paradigms into a unified system is a fundamental challenge in artificial intelligence.

Neurosymbolic programming is a promising emerging paradigm that aims to address this challenge. My research focuses on the *foundations of neurosymbolic programming*—spanning formal semantics, language design, and learning algorithms—and its *applications to real-world problems* involving natural language reasoning, computer vision, and multi-modal integration. To this end, I have pursued two complementary directions of research: Scallop, a framework for general-purpose neurosymbolic programming, published in (NeurIPS 2021), (PLDI 2023), (AAAI 2024), and an invited monograph in the Foundations and Trends in Programming Languages series (FnTPL 2024); and a series of progressively advanced applications that enhance the complexity of reasoning and integrate increasingly diverse modalities, published in (ICML 2020), (ACL 2023), and (TR 2024).

Research on Foundations of Neurosymbolic Programming

My research on Scallop was driven by the observation that while previous approaches had shown the immense potential of neurosymbolic programming in specific applications, these benefits did not readily transfer across applications. Developing each neurosymbolic application required duplicating effort to tackle recurring challenges. What was missing was a general programming framework to make solutions to these challenges broadly applicable and accessible.

The first prototype of a general-purpose neurosymbolic framework. In my (NeurIPS 2021) work, I set out to develop an early prototype of Scallop as a framework for scalable differentiable reasoning in neurosymbolic programs by leveraging ideas from probabilistic deductive databases. My key insight was that Datalog, a declarative logic programming language, is expressive enough to specify rich forms of symbolic reasoning (such as recursion and aggregation), yet efficient enough to scale to large datasets in real-world applications. Furthermore, I extended the theory of provenance semirings developed by the database community to make Datalog programs end-to-end differentiable. To enable efficient learning in the presence of complex loss surfaces induced by arbitrary Datalog computations, I introduced the *top-k-proofs* semiring which is parameterized by a hyperparameter k to control the tradeoff between efficiency and accuracy. In doing so, my work not only generalized the exact semantics of differentiable reasoning in previous work on DeepProbLog, while significantly improving its applicability to complex neurosymbolic tasks. I demonstrated the effectiveness of Scallop on Visual Question Answering with Reasoning (VQAR), a challenging task I introduced by extending the widely-studied problem of VQA with common-sense reasoning using an external knowledge base. The scene graph of images in this task encompasses hundreds of object names, attributes, and relationships, whereas the knowledge base comprises thousands of facts, providing a comprehensive testbed for complex reasoning. Scallop was able to achieve an accuracy of 84%, significantly outperforming VQA-specific models based on Neural Module Networks and Transformers by 12% to 22%.

Scallop: From prototype to a neurosymbolic programming language. Encouraged by this result, together with my collaborator Ziyang Li, I redesigned Scallop and implemented a full-fledged compiler to support a wide range of neurosymbolic applications (PLDI 2023). To ensure Scallop's applicability across diverse domains, I studied the data representations of various modalities, the necessary reasoning constructs, ease of use, and learning performance across a suite of tasks spanning image, natural language text, tabular data, and their combinations. These studies informed the language design, leading to the incorporation of a comprehensive set of reasoning constructs, such as negation and aggregation; a rich library supporting discrete, probabilistic, and differentiable reasoning with flexible trade-offs between efficiency and accuracy; and user-friendly PyTorch APIs for seamless integration into machine learning pipelines. My hypothesis that the neurosymbolic paradigm can universally benefit the AI tasks, was further confirmed by the experiments over the benchmark. For instance, in the PacMan-Maze task, Scallop achieved a 1,000× speed-up in training episodes and a near-perfect success rate of 99.4%, outperforming DQN by 14.5%. Besides accuracy, Scallop outperformed existing models in various aspects, such as runtime and data efficiency, interpretability, and generalizability.

Vieira: Extending relational programming with foundational models. While Scallop significantly enhanced the usability and efficiency of neurosymbolic methods, it faced a fundamental limitation: the neural components assumed a closed-world vocabulary, restricting its applicability to real-world tasks such as image generation. The advent of foundation models offered a promising solution. These models are powerful enough to unify tasks within a specific domain but remain narrow in their applicability to generalize beyond the modalities they are trained on. My key insight was that foundation models can be treated as stateless functions with relational inputs and outputs. A logical reasoning language can then act as the glue to compose diverse sub-models into a cohesive framework for complex tasks. I realized this insight by building Vieira (AAAI 2024), an extension to Scallop that supports a customizable foreign interface to incorporate foundation models as plugins. I mentored a cohort of four undergrads to implement support for 12 foundation models, including GPT, CLIP, and SAM. Vieira not only expanded the scope of downstream tasks that our neurosymbolic system could address but also achieved performance on par with or exceeding competitive solutions. Leveraging a zero-shot combination of foundation models, we demonstrated Vieira's effectiveness across challenging tasks spanning language, vision, vector databases, and even image editing and generation. For instance, Vieira achieved 100% accuracy on the NLP task of Tracking Shuffled Objects (TSO) from BIG-Bench, significantly enhancing reasoning performance and reducing hallucinations compared to 84% by GPT-4 using Chain of Thought.

Broader Impact. A significant goal of my research is to facilitate the widespread adoption of the neurosymbolic paradigm alongside advancing its foundations. To that end, I created and presented invited tutorials on Scallop at the Eleventh Summer School on Formal Techniques in 2022 and the Summer School on Neurosymbolic Programming in 2024, as well as at conferences such as Learning on Graphs (LoG 2022) and Programming Language Design and Implementation (PLDI 2023). I also wrote a comprehensive treatise titled "Neurosymbolic Programming in Scallop: Principles and Practice" and published by NOW Publishers.

All of the artifacts I have developed in my research are publicly available. Other researchers are already independently building upon these artifacts; a notable example is work by a group of vision researchers appearing at the 2024 International Conference on Pattern Recognition and Artificial Intelligence. Titled "Open-World Visual Reasoning by a Neuro-Symbolic Program of Zero-Shot Symbols," it is an example of Scallop's adoption by external researchers who are beginning to use neurosymbolic approaches in their respective domains.

Research on Language and Vision Reasoning Applications

Language and vision are two fundamental modalities through which humans interact with the world. They provide a wide array of applications that naturally require both neural perception and logical reasoning. A key challenge lies in defining the correct data representation that bridges the symbolic nature of reasoning with the unstructured domain of perception. By exploring progressively sophisticated applications that advance reasoning complexity and integrate diverse modalities, I have worked toward developing generic semantic representations for visual data, natural language, and their combinations, which in turn has helped to lay the foundation for my vision of a general and practical neurosymbolic paradigm.

Vision only: Image scene graph and programmatic referring expressions. In my (ICML 2020) work, I focused on a simple vision task with intriguing implications for the neurosymbolic paradigm: synthesizing referring expressions. Given a symbolic representation of an image and a target object within it, the goal was to generate a relational program that uniquely identifies the object. By interfacing the perceptual and reasoning components through a scene graph representation, I designed and implemented an executable program interpreter. This interpreter took a scene graph and a synthesized programmatic referring expression, and provided feedback on whether the expression satisfies the desired properties. Leveraging the feedback, a policy network was trained via reinforcement learning to synthesize referring expressions, resulting in a significant performance improvement, surpassing various state-of-the-art program synthesis approaches by 51% to 91%. While this work achieved impressive accuracy, it also highlighted key limitations in neurosymbolic systems at the time. Without a principled method for integrating differentiable logic inference into learning, this approach suffered from low data efficiency and long training times. Further, the program interpreter was tailored specifically to the CLEVR task, making it a point solution that lacked generalizability to other applications. These limitations highlighted the need for a systematic approach to transform the reinforcement learning framework into an algorithmic supervised learning scheme—a realization that ultimately inspired the creation of Scallop.

Natural language only: long-range reasoning with common sense knowledge. With Scallop, a general-purpose differentiable neurosymbolic language, I endeavored to push the boundaries of its applicability to complex natural language tasks. This led to my work on DSR-LM (ACL 2023), a Differentiable Symbolic Reasoning framework. The core innovation in this work was the integration of a pre-trained large language model, GPT-3, as a source of commonsense knowledge, with Scallop as the reasoning engine to perform robust long-range reasoning. Leveraging foundation models not only significantly reduced the need for manually crafting commonsense knowledge bases—an inherently labor-intensive and infeasible task—but also provided adaptability to accommodate various symbolic data representations. Powered by

GPT-3 and Scallop, DSR-LM achieved 60.98% accuracy on the CLUTRR task, significantly surpassing fine-tuned GPT-3 (34.3%) and the structured RN model (39.9%). The success of DSR-LM inspired the development of Vieira, which extended this concept by integrating multiple foundation models, further advancing the neurosymbolic paradigm.

Language-vision: aligning video and caption for weak supervision. Having established structured representations for both vision and natural language, the natural next question arose: what kind of representation can effectively bridge the gap between these two modalities? This inquiry led me to build LASER (TR 2024), a novel framework that aligns the symbolic representation of a video with its caption. The first challenge was defining a suitable symbolic representation for the two modalities. I extended scene graph representations to videos with the Spatio-Temporal Scene Graph (STSG) and designed the Spatio-Temporal Specification Language (STSL) for video captions, based on Linear Temporal Logic. With this symbolic data representation, I developed a neurosymbolic alignment checker in Scallop, providing differentiable alignment scores for weak supervision. STSGs were generated from video via object trajectory extraction, while captions were processed into STSL programs with GPT-4 for temporal and spatial orderings, validated by a parser. On a complex real-world video benchmark OpenPVSG, LASER achieved substantial improvements, outperforming fully-supervised STSG extraction baselines by 12.65%. LASER opened new avenues for video understanding, addressing the labor-intensive nature of low-level annotations by leveraging symbolic alignment for weak supervision. It also inspired extending Scallop to support Linear Temporal Logic and temporal databases, advancing temporal reasoning in neurosymbolic systems.

My Future Research Directions

New paradigms in neurosymbolic learning. My current encapsulation of neurosymbolic learning focuses on unifying symbolic reasoning and neural perception via probabilistic relational databases, enabling tasks that combine structured logical inference with neural perception. However, there are numerous other potential approaches that could expand the horizons of this field, enhancing scalability and applicability. One promising direction is vector symbolic architecture (VSA), which combines the power of symbolic manipulation with the distributed representation capabilities of neural networks. VSAs enable encoding symbolic relationships as high-dimensional vectors, offering a pathway to perform operations like binding and superposition directly within neural systems. Another area of exploration is statistical relational artificial intelligence (StarAI), which merges statistical models with relational logic to address uncertainty and relational complexity simultaneously. This paradigm could significantly extend neurosymbolic systems by incorporating probabilistic programming techniques that introduce variables and confidence intervals, facilitating more informed and robust decision-making processes in uncertain and dynamic environments.

From neurosymbolic systems to trustworthy AI. Neurosymbolic systems inherently provide explainability through their explicit proofs, making them well-suited to address critical aspects of trustworthy AI. To ensure these systems are not only powerful but also reliable and ethical, the transition from neurosymbolic frameworks to trustworthy AI must focus on:

1. Explainability: Neurosymbolic systems inherently offer greater interpretability by making their reasoning processes explicit, with potential expansions including methods to generate user-friendly explanations that enhance understanding and trust in AI decisions.
2. Robustness and Resilience: To be deployed in real-world scenarios, systems must withstand noisy or adversarial inputs. Incorporating techniques that enhance robustness and

recoverability will be critical for neurosymbolic systems.

3. Fairness and Transparency: Ensuring that decisions are equitable across diverse populations and providing transparency in how conclusions are reached are essential for addressing ethical concerns. Neurosymbolic systems, with their structured reasoning, can encode and enforce fairness constraints more effectively than black-box models.
4. Generalization and Adaptability: Scaling neurosymbolic solutions to new domains requires building systems that generalize well across tasks and adapt dynamically to new environments.
5. Verifiability and Credibility: Trustworthy systems must be verifiable, enabling rigorous testing and validation of their outputs. Neurosymbolic frameworks, with their explicit logic and reasoning layers, are uniquely positioned to support formal verification methods.

Video understanding, grounding, and reasoning. Advancing video understanding and reasoning is a transformative application for neurosymbolic frameworks, particularly for tasks requiring long-term temporal dependencies and multi-modal integration. A comprehensive pipeline based on a neurosymbolic framework could facilitate long video analysis, enabling systems to reason across extended temporal spans and complex visual scenes; support robotics and autonomous driving, incorporating domain-specific hard constraints, such as safety rules, navigation protocols, and real-time decision-making, into the perception system; and enhance planning and situational reasoning, allowing AI systems to generate plans or respond to dynamic environments by combining symbolic constraints with perceptual inputs. To achieve these capabilities, extended support for database representations is necessary, including advanced relational and temporal logic systems. These advancements would enable robust, scalable solutions for applications in robotics, healthcare, and education.

References

(ICML 2020) Jiani Huang, Calvin Smith, Osbert Bastani, Rishabh Singh, Aws Albarghouthi, and Mayur Naik. “Generating Programmatic Referring Expressions via Program Synthesis”. In Proceedings of the International Conference on Machine Learning.

(NeurIPS 2021) Jiani Huang*, Ziyang Li*, Binghong Chen, Karan Samel, Mayur Naik, Le Song, and Xujie Si. “Scallop: From Probabilistic Deductive Databases to Scalable Differentiable Reasoning”. In Proceedings of the Conference on Neural Information Processing Systems.

(ACL 2023) Jiani Huang*, Hanlin Zhang*, Ziyang Li, Mayur Naik, and Eric Xing. “Improved Logical Reasoning of Language Models via Differentiable Symbolic Programming”. In Proceedings of the Findings of the Association for Computational Linguistics.

(PLDI 2023) Ziyang Li*, Jiani Huang*, and Mayur Naik. “Scallop: A Language for Neurosymbolic Programming”. In Proceedings of the ACM Conference on Programming Language Design and Implementation.

(AAAI 2024) Ziyang Li, Jiani Huang, Jason Liu, Felix Zhu, Eric Zhao, William Dodds, Neelay Velingker, Rajeev Alur, and Mayur Naik. “Relational Programming with Foundation Models”. In Proceedings of the AAAI Conference on Artificial Intelligence.

(FnTPL 2024) Ziyang Li*, Jiani Huang*, Jason Liu, Mayur Naik. “Neurosymbolic Programming in Scallop: Principles and Practice.” Invited Monograph to *Foundations and Trends® in Programming Languages*, NOW Publishers.

(TR 2024) Jiani Huang, Ziyang Li, Mayur Naik, and Ser-Nam Lim. “LASER: A Neuro-Symbolic Framework for Learning Spatial-Temporal Scene Graphs with Weak Supervision”. Under review at the International Conference on Learning Representations (ICLR), 2025.

* indicates these authors contributed equally.