Research Statement

Houru Jiang
tifeng236@gmail.com juliourias@stu2023.jnu.edu.cn

Overview

My research lies at the intersection of **generative models**, **trustworthy AI**, and **AI for education**. I am particularly interested in diffusion and flow-based generative models—both in extending them to broader downstream applications and in making their training more efficient—as well as in understanding how neural models learn, represent, and apply knowledge. My work has spanned generative modeling for computer vision, mechanistic interpretability, knowledge tracing, and AI evaluation.

Past and Current Research

Knowledge Tracing and Cognitive Diagnosis

Knowledge tracing—modeling a student's evolving mastery of knowledge components from their learning history—is a central problem in AI for education. In **NSR-KT**, we introduced a neuro-symbolic residual learning framework that disentangles sequential regularities from sparse learning interactions, achieving strong generalization with interpretable skill-level estimates. Building on this, **All-as-Vision** reformulates exercise representation as a transferable visual encoding problem, enabling cross-domain knowledge tracing without task-specific feature engineering.

In the cognitive diagnosis setting, I developed **Auditing Graph Utilization in Cognitive Diagnosis**, a model-agnostic framework that instruments graph-based diagnosis models with causal, information-theoretic, and probabilistic probes. This work was among the first to systematically audit how graph structure is actually used—rather than assumed to be used—by cognitive diagnosis models.

Trustworthy AI: Evaluation and Interpretability

A thread running through my research is the question: *how do we know if our benchmarks and models are actually measuring what we think they are?* In my **Information-Theoretic Evaluation Framework for Knowledge Tracing**, we formalize benchmark quality through information-theoretic quantities—redundancy, synergy, and unique information—between benchmark design, model predictions, and true latent states. This framework can diagnose when a benchmark is saturated, when a model is exploiting spurious correlations, and when performance numbers are misleading.

Complementing this evaluation perspective, I investigate the internal mechanisms of large models—how learned representations encode task-relevant features and how these representations can be interpreted through sparse autoencoders and causal intervention methods. I am particularly interested in multimodal settings: understanding how vision and language representations interact within large models, and how interpretability tools developed for language models transfer to vision-language architectures. Together, these two threads—evaluation from the outside and interpretation from the inside—provide a holistic view of model trustworthiness.

Generative Models and Computer Vision

In **FUSE**, we proposed a flow-based unified framework for zero-shot segmentation by expression, demonstrating that flow matching can serve as a powerful backbone for dense visual prediction tasks. This work connects generative modeling with classical computer vision, and opens directions for controlled generation and editing at the pixel level. I am continuing to explore diffusion and flow-based models for both generation and representation learning.

Future Directions

Broader Downstream Applications of Generative Models

I am excited by the potential of diffusion and flow-based generative models to serve as general-purpose backbones for a wide range of downstream tasks beyond image generation. My work on FUSE demonstrated that flow matching can be adapted for dense visual prediction (zero-shot segmentation), and I plan to extend this paradigm to other modalities and tasks: video understanding, 3D content creation, controllable generation, and multimodal synthesis. A key direction is developing unified generative frameworks that can handle heterogeneous inputs and outputs—text, images, video, 3D—within a single modeling paradigm, enabling downstream applications in creative tools, scientific visualization, and simulation.

Efficient Training of Generative Models

Despite their impressive capabilities, diffusion and flow-based models remain computationally expensive to train and sample from. I am interested in improving the efficiency of generative model training through several complementary angles: better noise schedules and sampling strategies that reduce the number of function evaluations; architectural innovations that improve parameter efficiency and convergence speed; and training curricula that adaptively structure the learning process based on the model's current capability—an idea that bridges my knowledge tracing background with generative model training. These efficiency improvements are essential for democratizing access to powerful generative models and enabling their deployment in resource-constrained settings.

Conclusion

My research is unified by a focus on **generative models as a foundation for building, understanding, and evaluating AI systems**. Whether through developing diffusion and flow-based models for broader downstream applications, improving their training efficiency, or building evaluation and interpretability tools that reveal model internals, I aim to advance AI systems that are not only more capable and efficient, but also more transparent and trustworthy.