

# AI-Driven Scientific Visualization

## 1. Introduction:

Artificial Intelligence (AI) has demonstrated transformative potential across multiple scientific disciplines, including biology, medicine, climate science, physics, and materials science. The integration of domain-specific AI into scientific data analysis and discovery automation remains in its nascent stages. Significant new challenges persist in bridging the gaps between data, users, and AI models, which currently limit the full potential of AI-driven analytical systems for scientific innovations and advances. The overarching goal of my research is to establish the **core foundations of next-generation AI-driven scientific data analysis and visualization systems** capable of processing large-scale, high-dimensional, and in situ scientific data with enhanced *expressivity*, *explainability*, and *scalability*. My research vision, as shown in Fig. 1, encompasses three key pillars: 1) **AI Generative Model**: novel neural rendering architectures for interactive analytical systems with high visualization quality and low rendering latency. 2) **Human-AI Interaction**: human exploratory behavior learning, and Large Multimodal Models-based agentic assistant for effective user interaction. 3) **AI-Friendly Data Representation**: Implicit Neural Representation (INR) with efficient training and inference for complex data on modern neural processing platforms.

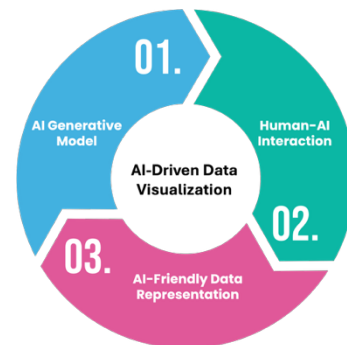


Figure 1: My research is on the foundations of modern AI-driven data analysis and visualization.

## 2. Research Directions

**Transformative Integration of Data Enabling Science:** This future direction tackles the critical challenge of preparing scientific data for artificial intelligence by ensuring it is semantically rich, interoperable, and universally discoverable. This project aims to provide solutions to three core thrusts: 1) Enabling efficient feature learning from sophisticated data through indexed semantic embeddings. 2) Establishing robust linkages between these embeddings and diverse data artifacts (e.g., simulation and observational data, intermediate attention results for inferences, and shared reasoning contexts). 3) Designing federated data management frameworks that automate multimodal indexing, composition, query processing, and reuse of embeddings and linked artifacts. By providing a unified framework for the discovery, navigation, update, and reuse of scientific data and knowledge, this research work bridges a significant gap in the AI-assisted research ecosystem. Furthermore, the direction also addresses the optimization issues arising in the meantime to meet the extreme performance and scalability demands of modern scientific computing on heterogeneous hardware.

**Intelligent Data Exploration Assistant for Science:** My solution to the problem of exascale scientific data understanding is twofold: 1) AI-enhanced visualization and analysis framework designed to support intelligent, scalable, and interactive exploration of complex scientific data. This direction will continue the exploration of constructing AI-friendly data representations from scientific data with diverse types and scales through efficient learning and modeling of a compact INR, therefore enhancing the downstream tasks of analysis and visualization. 2) AI assistant that interacts with domain experts with insights and suggestions by understanding and reasoning through Multimodal Large Models (MLMMs). This direction will empower the agentic AI assistant with capabilities of processing, understanding, and reasoning from the large-scale data and parameter space that is not tractable to humans. Critical uncertainty, topology, and time-varying features will be learned through LMMs, such as the Vision Language Model (VLM), to interpret user intent, adapt to dynamic workflows, and orchestrate contextual data analysis and visualization, enabling scalable and proactive exploration of complex scientific data.

**Automatic Data Analysis for Accelerating Science:** Future AI-driven data analytical workflows for accelerating laboratory experiments and discovery in diverse critical scientific domains, including biotechnology, material science, chemistry, physical simulation, etc, rely on large-scale simulation, big data management, data analysis, and visualization with advanced AI-assisted learning, reasoning, and generative capability. I would like to investigate: 1) Improve the accessibility for data processing, training, and visualization on state-of-the-art cloud-based high-performance computing clusters and AI infrastructures, providing a level of automation, throughput, and ease of use. 2) Bridge post-experiment stages of data transformation, analysis, and visualization, to accelerate science and discovery by orders of magnitude. 3) Facilitate a more effective, more efficient, more inclusive way of doing science through establishing standards for experiment protocols, metadata, parameter space exploration, and the agentic AI. The objective is to accelerate scientific innovation by leveraging a network of AI agents to automate discovery processes, devise novel scientific methods, and uncover unifying principles across disciplines at a global scale.