Building the Foundation for Automated Scientific Discovery

Scientific breakthroughs, from developing life-saving drugs to understanding the world, are urgently needed. Accelerating and scaling scientific discovery is now critical. My research addresses this by developing **AI agents** capable of automating scientific discovery – the process of understanding and reasoning about scientific phenomena, controlling scientific tools, and interpreting outcomes. However, current AI agents face key challenges: (1) limited understanding of multimodal scientific inputs; (2) insufficient planning abilities to solve complex problems; and (3) safety concerns requiring verifiable, reliable predictions. To address these challenges, I have led the following research efforts:

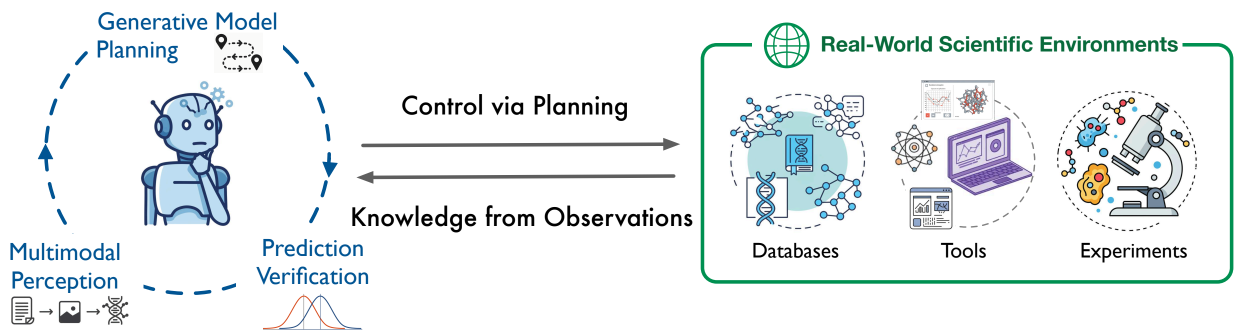

Figure 1: Research Roadmap

Multimodal Understanding of Scientific Data Scientific data include many modalities beyond natural text and data, e.g. molecule graphs, protein sequences. Understanding these modalities is the basis of using AI for scientific discovery. In TorchDrug [Preprint 22] and PEER [NeurIPS 22], we made the earliest effort to comprehensively benchmark different architectures for understanding proteins, molecules, and their interactions. Further in RSA [EMNLP 24], I identify the importance of retrieval-based features for understanding proteins. RSA replaces the dominant MSA feature with dense retrieval, achieving 300x faster efficiency and higher accuracy. Beyond understanding biosequences, I explored integrating textual reasoning with scientific input. [EMNLP 24] combinates the use of RSA with Large Language Models (LLMs) to reason about protein sequences; GIMLET [NeurIPS 23] pre-train a text-graph transformer to reason about molecule properties; and UTGDiff [Preprint 24] explores text-guided molecule design. These multimodal methods have key advantages (e.g. reasoning intensive, multi-task pre-training), which can enable the unification and scaling of scientific foundation model.

Enhancing Generative Model Planning I initially explored LLM agents for Automated Research in 2022 and found that LLMs (e.g., ChatGPT) at the time could barely handle basic scientific tasks like literature review. Therefore, I started with the basics and investigated the deficiencies of LLMs as agents. I developed a virtual environment platform, AgentBoard [NeurIPS 24 Oral], where agents iteratively generate API-based actions and process textual observations to complete tasks. Using AgentBoard, I identified long-horizon planning as a major bottleneck and aim to enhance this capability. In Predictive-Decoding [ICLR 25], I address suboptimal LLM planning caused by greedy decoding using a global-aware sampling method, significantly improving planning without additional training. Follow-up work [submitted to ACL 25] optimizes its efficiency, while another [submitted to ACL 25] adapts it into a scalable RL framework. We also explored using cross-domain middle-training to enhance planning [submitted to COLM 25] and learning an expandable action space to reduce planning complexity [COLM 24]. These advancements empower LLMs to handle complex planning tasks.

Verification Frameworks for Trustworthy Predictions Reliability is vital for AI in real-world science, and verifying model predictions enables trustworthy outcomes. In CISS [ICML 22], Zhao and I developed a scalable verification method for pretrained language models to ensure robustness against text perturbations. We also verified model consistency in data-scarce [ACL 24] and parameter-efficient tuning scenarios [NAACL 25 Oral]. I plan to extend this work to the verification of LLM agents in the future.

Building Agents for Automatic Scientific Discovery AI Agents perform scientific research by interacting with real-world scientific environments. One example is agents automatically performing research by browsing scientific databases. In [EMNLP 24], I build a protein agent, ProteinChat, that actively queries databases (e.g. protein sequence database, gene ontology, and literature search databases) to accomplish novel tasks related to protein sequences, such as design a protocol for measuring the function of this given protein. In our recent work, BioMaze [submitted to ACL 25], we build a LLM agent that draws hypothetical conclusions about interventions simply by exploring a biological pathway database, e.g. *the effect of blocking muscarinic receptors M3 on taste receptor cells*. These "virtual" agents have the potential to autonomously obtain knowledge and perform reasoning, but lack grounding on real-world experiments. We are now building another platform to enable agents interaction with scientific softwares, extending our work on computer-using agent [submitted to COLM 25]. Interaction with softwares that simulate, process and visualize experimental results (e.g. Molecular Dynamics software for protein-ligand docking) could stimulate the agents to perform experimental research.

Future Work In summary, my research lays the groundwork for developing AI agents capable of conducting scientific research tasks, with the ultimate aim of accelerating and facilitating scientific breakthroughs. For future work, I intend to explore the following three directions:

- ***Agents as Co-Scientists:*** I aim to expand upon my work on scientific agents to create more seamless and efficient workflows for human researchers. These agents would autonomously handle routine tasks such as information browsing, and simplify the utilization of scientific software. This would enable scientists to focus on higher-level problem-solving.
- ***Agents for Autonomous Laboratory Operations:*** Developing agents capable of conducting experiment-driven tasks is a crucial step toward fully autonomous scientific discovery. This area is particularly labor-intensive and in great demand of automation. I plan to design integrated virtual-physical environments to streamline interactions. For example, considering a drug-assay agent, I aim to train reasoning-focused agent models that can propose and optimize drug candidates through iterative interactions with an experimental platform. This platform would combine prediction models trained on wet-lab data with embodied devices capable of performing physical experiments. The "virtual lab" would enable rapid iterations and model training, while the "physical lab" would validate results in real-world conditions.
- ***Learning Scientific Laws from Agentic Experiences:*** While scientific agents primarily excel at **inductive** research, **deductive** reasoning is indispensable for research rooted in first principles. Agents could advance deductive research by gathering experiential data through interactions with their environment. Leveraging the aforementioned platforms, I plan to collect large-scale trajectory data and employ models to extract fundamental rules and scientific laws from these observations. Just as the law of gravity was deduced by observing objects falling, similarly, generalizable scientific principles could be derived from extensive agentic experiences.