Since childhood, I have loved learning about life in its myriad manifestations. The COVID-19 pandemic, however, turned this interest into a call to action. While reading about viruses and infectious diseases during the pandemic, I came across a paper discussing the influence of viruses on the disease-causing ability ("pathogenicity") of the bacteria *Acinetobacter baumannii*. Inspired by this approach, I started investigating the influence of viruses in a related species called *Acinetobacter pittii*, analyzing virus-derived genes and proteins that might influence the bacteria's pathogenicity. Through computational techniques like genome annotation, synteny analysis, and phylogenetic tree construction, I identified 94 virus-derived DNA sequences embedded in the genomes of *A. pittii*, which contained 37 antibiotic resistance genes and 47 virulence factors that could influence bacterial pathogenicity. **I wrote a first-author paper on this work that was published in the peer-reviewed journal *PHAGE* and presented at the Harvard Microbial Science Initiative 20th Anniversary Symposium**.

Through this project, I uncovered new evidence for how bacteria evolve to cause disease and spread among patients worldwide. It made me realize that, with computational tools, I could tackle global biological challenges like the COVID-19 pandemic. I have been developing and applying computational tools to analyze biological systems ever since.

Intellectual Merit

I am interested in developing "AI scientist" approaches to investigating biological systems. These AI scientists integrate information from the molecular to the clinical level in generating novel scientific hypotheses about molecules, cells, and organisms.

The AI Scientist: Agentic Approaches for Scientific Discovery

An AI scientist is an AI agent system built from large language models (LLMs) that uses tools and decision-making to mine scientific data for novel hypotheses. In the Zitnik Lab at Harvard, I helped build an AI scientist, TxAgent, that uses 200+ tools and a multi-agent system to answer biomedical questions. When asked a question like "What are the proteins involved in the disease mechanism of leukemia?", TxAgent conducts a multi-step reasoning process where—at each step—it finds a specific tool (*e.g.*, to retrieve protein sequences), calls the tool, analyzes the output, and proceeds to the next step.

My first contribution was expanding TxAgent's toolbox. I reflected on the biology I learned while competing in the USA Biology Olympiad National Finals and my molecular biology research on Alzheimer's in the Young-Pearse Lab at Brigham & Women's Hospital. These experiences helped me incorporate tools into TxAgent's toolbox that biologists, chemists, and other domain experts use in their research. For example, I added tools to access disease-associated symptoms from the Human Phenotype Ontology database. Since TxAgent's toolbox contains hundreds of tools, I also created training data for a retrieval-augmented generation (RAG) model that helps TxAgent quickly identify the best tool for a particular hypothesis or scientific analysis without requiring that TxAgent memorize all 200+ tools.

Moreover, I built a training and benchmark dataset called Description Prescribing Card that tests whether AI scientists like TxAgent can reason about a class of therapeutic molecules (*e.g.*, statins) independent of the name of any specific molecule. Using this training data and benchmark, we showed that TxAgent outperforms leading LLMs like GPT-4o on therapeutic reasoning.

Finally, we partnered with 29 rare disease foundations to evaluate TxAgent on biomedical questions related to each disease. I built an evaluation platform, which rare disease scientists from each foundation used to evaluate TxAgent. Their evaluations showed that TxAgent can answer rare disease questions more accurately and with better explanations than its base model. This project showed me that the "AI scientist" was not just a concept; by training TxAgent to use gold-standard research tools and show its evidence-grounded reasoning, we had built a tool that is useful across many disease applications.

However, TxAgent is just a prototype. I noticed it sometimes makes biological inaccuracies, as its tools only retrieve textual data. A true AI scientist must reason across disciplines and data modalities,

thinking simultaneously about chemical structures, genomic sequences, and cell biology. To build such a model, I needed to first understand this interdisciplinary reasoning process myself.

As a result, I enrolled in a graduate drug development class at MIT and worked on a project with chemical engineers and a patent lawyer. The chemical engineers optimized the binding affinity of molecules to our protein target, while I predicted the pharmaceutical properties of our top candidates with a graph neural network. We then selected a final molecule that the lawyer helped us protect with patent claims. Together, we designed a novel drug for FSHD, a rare muscular dystrophy. This was my first encounter with AI for modeling molecules, and I quickly saw how these AI tools could help us answer questions like "Will this molecule bind to these proteins?" that text-based models could not answer.

AI for Molecular Modeling

To learn how we can build AI scientists that understand multiple scientific data modalities, and not just text, I dove into the field of molecular modeling: building AI models for capturing the complex 3D geometric structures of proteins and small molecules. In my junior fall, I worked with computer scientists and protein biologists to design MUTAGENIC, a method for redesigning a protein sequence so that the final protein molecule performs a different biological function. To do this, I investigated embeddings of protein function from ESM3 (a language model trained on protein sequences, structures, and functions). I showed that these embeddings reflect differences in protein function and thus can be used as references to figure out which amino acids in a protein sequence to redesign (mask and re-decode) with a protein language model when trying to attain a particular biological function. **This work was presented as a co-first-author poster at the 2025 Learning Meaningful Representations of Life workshop at the International Conference on Learning Representations.** This project showed me the value in combining structural biology and computer science in building AI for biomolecular modeling.

To explore molecular modeling further, I collaborated with scientists in the Rosetta Commons (an international protein modeling community) this summer to build an AI tool for modeling protein-protein complexes—such as that formed from a virus protein bound to a human cell surface protein. My goal was to improve DFMDock, a diffusion model that could propose potential structures of a protein-protein complex but wasn't good at ranking these structures. DFMDock learned to predict the energy landscape for protein-protein binding, but the learned landscape was imperfect and didn't assign lower energies to more biophysically plausible binding interactions. I tried multiple techniques to improve DFMDock, including steering the model with physics-based energy functions and Feynman-Kacs steering, as well as fine-tuning with direct preference optimization (DPO) on biophysically "good" and "bad" poses. When these techniques both failed, I looked into computer science research from the Learning-to-Rank subfield, as I suspected that these techniques could help DFMDock rank protein-protein complexes better. This is how I came across the LambdaLoss for training ranking models, which I used to improve my model beyond state-of-the-art performance on ranking protein-protein complexes. Drawing connections across computer science and biophysics was deeply fulfilling and showed me the importance of interdisciplinary collaboration. I have continued these conversations with experts across many fields while **presenting this first-author work at the 2025 Summer RosettaCon, 2025 Molecular Machine Learning conference, and 2025 Machine Learning for Structural Biology workshop (co-localized with NeurIPS).**

Unifying Agentic AI and Molecular Modeling: Multi-Disciplinary AI Scientists for Discovery

This fall, drawing upon my experiences in both agentic AI and molecular modeling, I started expanding TxAgent's toolbox into a "ToolUniverse" for building AI scientists that contained more tools from molecular modeling, such as graph neural networks for predicting chemical properties and diffusion models for modeling intermolecular interactions. Using ToolUniverse, I built another AI scientist prototype that answers questions in biology and chemistry like "How can I design a molecule that binds to HMG-CoA reductase and does not cross the blood-brain barrier?" Watching the AI scientist manipulate

tools from a half dozen subdisciplines to answer queries, I started grinning—by equipping the model to conduct interdisciplinary reasoning, we were one step closer to an autonomous AI scientist. **I am the second author on the ToolUniverse paper currently under review at Nature Biotechnology.**

Ultimately, to build AI scientists, I aim to answer the following question: **How can we teach AI to reason like scientists, integrating evidence across chemistry, biology, and clinical data to generate testable hypotheses?** By developing AI systems that think across scales, from molecular interactions to human physiology, I hope to develop foundational AI methods for diverse scientific reasoning tasks.

Broader Impacts

Through the Zitnik Lab, I have started building a community for research on AI scientists. Our ToolUniverse project is open-source and allows anyone to build an AI scientist from standard LLMs. As we collaborate with biomedical experts across the U.S. and world, I have also built platforms to crowdsource evaluations of our TxAgent AI scientist to understand the strengths and weaknesses of current models. Finally, to stimulate interest in this field, **we are hosting CURE-Bench, a competition to benchmark AI for therapeutic reasoning, at the 2025 Neural Information Processing Systems (NeurIPS) conference**. I built an agentic AI system for evaluating submissions to CURE-Bench and am helping identify the leading AI scientist models. To date, the competition has attracted 300+ teams, and I am incredibly excited to co-organize this effort to advance the frontier of AI scientists.

Outside the Zitnik Lab, I have also led conversations and collaborations across fields in my other extracurriculars. I served as co-president of Harvard GAMI, a global health technology innovation community, where I oversaw projects ranging from supporting Kenyan blood donors with AI chatbots to diagnosing neurodegenerative disease in India using statistical models. Working with both physicians and engineers, I led nine student teams in building solutions to unmet global health needs.

While leading GAMI, I often encountered "language barriers" between members in technical and non-technical disciplines. For example, [redacted], our team lead for the blood donation project, had little experience in chatbot development. I worked with [redacted] to lay out a technical framework for building the chatbot, as well as plans for obtaining high-quality blood donation reference material and conducting safety testing to ensure the chatbot wouldn't fabricate medical information. Through my mentorship, I helped [redacted] develop a plan to lead both technical members building the chatbot and non-technical members focused on data and safety. Through teamwork, his group built a chatbot aligned with the needs of our Kenyan collaborators: when asked about blood donation, it would answer with evidence, but it would refuse other questions.

By the end of my tenure, GAMI had over seventy members spanning multiple universities and academic fields. Reflecting on these experiences, I realized GAMI's biggest gift was teaching me how to bring people together across disciplines. Whether working with clinicians, software engineers, or biology students, I had learned to rally everyone around a unified goal.

Beyond GAMI, I have also taught in a wide variety of disciplines. In Summer 2024, I taught a week-long crash course on fish biology to Chinese high schoolers at the Harvard Summit for Youth Leaders in China conference. Subsequently, I served as a teaching fellow for CS1200 (Introduction to Algorithms) and as a peer tutor for courses in systems programming, probability, and efficient ML. Next semester, I will serve as a teaching fellow for Stat171 (Introduction to Stochastic Processes). With the support of the GRFP, I will continue this community-building in AI for Science during my PhD and future career.

Future Goals

I hope to eventually lead an AI research lab collaborating closely with experts from different disciplines—chemists, biologists, physicists—to develop AI scientists that accelerate research across domains, from analyzing thousands of molecules a day to engineering new protein drugs and biosensors.