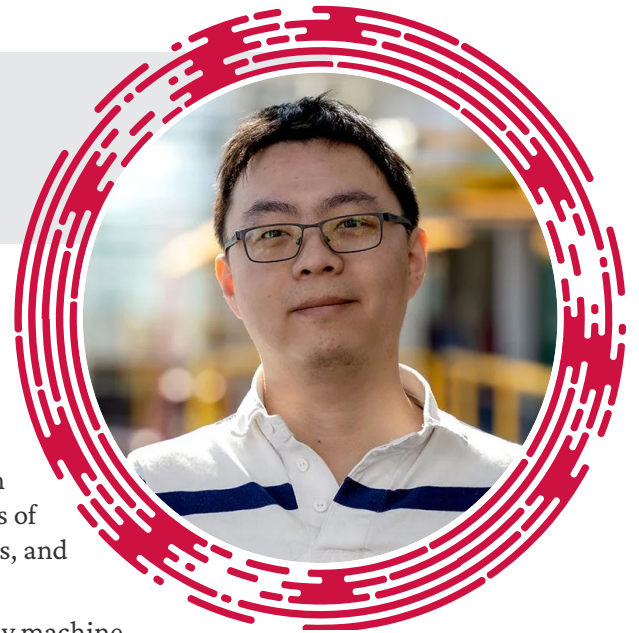# Advancing Agentic, Data-Driven, and Trustworthy Hypothesis Generation in Biomedical Research

Yang Lu, PhD
Assistant Professor
Cheriton School of Computer Science
University of Waterloo

Rapid developments in high-throughput sequencing have enabled biologists to collect large volumes of multi-omics data with unprecedented resolution. However, interpretation of such an increasing amount of heterogeneous biological data becomes highly nontrivial. In my talk, I will present a data-driven research paradigm to discover testable hypotheses directly from biological data in an interpretable and trustworthy fashion. In particular, the talk will focus on three recent works that address key aspects of biomedical research: analyzing data, generating and prioritizing hypotheses, and engaging with users:

(1) An interpretation method that detects non-additive interactions from any machine learning (ML) models. The detected interactions, treated as hypotheses, are rigorously controlled for statistical errors without relying on p-values. This method was the first to demonstrate to the community that higher-order interpretations of ML models can be achieved with confidence guarantees.

(2) An AI-driven agent that automatically translates biologists' needs into actionable insights. The agent we developed enables the automatic execution of off-the-shelf Python-based bioinformatics tools, allowing researchers to generate analysis results with minimal tool-specific knowledge and coding expertise. This method was the first initiative to streamline the automatic and codeless execution of general-purpose bioinformatics tasks via conversation.

(3) A critical reevaluation of problematic statistical estimation of the Basic Alignment Search Tool (BLAST), a cornerstone tool used in daily biomedical analysis over the past 30 years. We have introduced an alternative method to address this issue, ensuring that it does not yield inflated estimates of significance. Our study has the potential to influence and reshape numerous conclusions drawn by researchers.

ABOUT the SPEAKER

Yang Lu is an assistant professor at Cheriton School of Computer Science, University of Waterloo. Prior to that, he was a postdoctoral researcher in Prof. William Noble's group at the University of Washington. He obtained his PhD in Computational Biology and Bioinformatics under the supervision of Prof. Fengzhu Sun from University of Southern California. Before moving to the United States, he received MS and BS degrees in Computer Science and Engineering from Shanghai Jiao Tong University. Yang Lu's research focuses on developing machine learning and statistical methods for genomics and proteomics data analysis. He is particularly interested in developing interpretation methods to find scientifically interesting and statistically confident hypotheses from complex biological data.

Thursday, December 5 at Noon
2188 Mechanical Engineering Building