

Knowledge Centric AI for Scientific Discovery

Carla P. Gomes

¹ Cornell University

gomes@cs.cornell.edu

Data-centric AI, exemplified by the rapid advancement of deep learning and large language models, has fueled discussions of Artificial General Intelligence. However, for scientific discovery and high-stakes decision-making, purely data-driven methods face significant limitations. These include opaque behavior with limited interpretability, restricted use of prior knowledge, and brittle performance outside the training distribution. Furthermore, these models often struggle with heavy data requirements and complex multi-objective trade-offs. I will discuss a knowledge-centric AI agenda designed to overcome these hurdles. This approach combines first-principles reasoning with data-driven learning, integrating prior scientific knowledge with data to produce interpretable and informed recommendations. I will discuss recent work on solving inverse problems such as inferring crystal structures from X-ray diffraction data. I will also highlight the generality of the approach across other scientific applications and how it facilitates discovery, prediction, and decision.