ICFO

INSIGHT SEMINAR: Interpret or explain? What to keep in mind when learning physics from neural networks

ANNA DAWID-LEKOWSKA

May 08, 2025
12:00 to 13:00
Elements Room

ABSTRACT:
The importance of machine learning (ML) in quantum physics has been rising, especially in studies of quantum phases of matter or finding ground states of interacting Hamiltonians. The more widespread its use, the more urgent the critical questions of how to extract meaningful physical insights from ML. This talk will explore the differences between explaining a trained model's behavior (post-hoc explainability) and designing machine learning models with interpretable parts from the ground up. Using two case studies from our research - neural networks applied to the Su-Schrieffer-Heeger (SSH) model and the TetrisCNN model tailored to detecting phase transitions and their order parameters in spin systems - we will show the limitations of post-hoc explainability and the advantages of

interpretable architectures. We argue that the most insightful interpretable models are largely task-dependent, and we share our recipe for their design.

BIO:

Anna Dawid is an assistant professor at the Leiden Institute of Advanced Computer Scienc (LIACS) and the Leiden Institute of Physics (LION) at Leiden University in the Netherlands. He research interests include interpretable machine learning for science, ultracold quantu simulation platforms, and machine learning theory. Her passion is to transform automate computational methods (especially neural networks) into a new, unique lens, enabling a fres perspective on difficult, established scientific problems. Before taking up the position i Leiden, she was a researcher at the Center for Computational Quantum Physics at th Flatiron Institute in New York. In 2022, she defended her PhD in physics and photonics unde the supervision of Prof. Micha? Tomza (Faculty of Physics, University of Warsaw) and Prof Maciej Lewenstein (ICFO - The Institute of Photonic Sciences, Spain). Previously, sh completed her master's degree in quantum chemistry and bachelor's degree i biotechnology at the University of Warsaw. She is the first author of the book "Machin Learning in Quantum Science", which will be published by Cambridge University Press i 2025. She is also a laureate of the START scholarship of the Foundation for Polish Scienc (2022) and one of the participants of the 74th Nobel Laureate Meeting in Lindau (2024).

Hosted by: Prof. Dr. Maciej Lewenstein