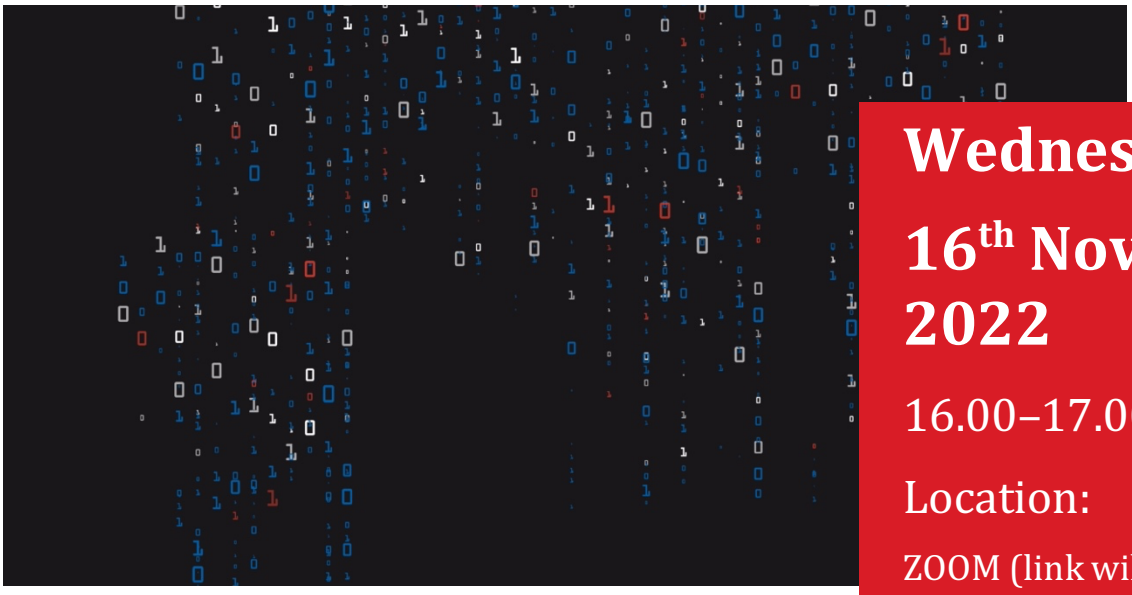
HEIBRIDS
HELMHOLTZ EINSTEIN INTERNATIONAL BERLIN RESEARCH SCHOOL IN DATA SCIENCE

LECTURE SERIES



Wednesday

16th November 2022

16.00–17.00

Location:

ZOOM (link will be shared via email)

Directions in Interpretability

Speaker: Ruth Fong, Princeton University

Abstract

As machine learning algorithms are increasingly applied to high impact yet high risk tasks, such as medical diagnosis or autonomous driving, it is critical that researchers can explain how such algorithms arrived at their predictions. In this talk, I'll highlight our work on explaining the decisions and internal representations of deep neural networks. We'll compare how the past decade of interpretability research has tracked with the broader research communities in machine learning and computer vision and highlight several novel directions that are important for keeping pace with the next decade of research breakthroughs.

Next Lecture: November 30, 2022

EINSTEIN CENTER Digital Future

HELMHOLTZ GEMEINSCHAFT