

## COURSE SUMMARY



The Interpretability Problem

Right to Explanation

Counterfactual Method

Axiomatic Attribution



Intelligible models

Monotonicity

Rationalization

Feature visualization