Virtual Guest Lecture
Multiple Models, Explainability, and Bias
Prepared by Michael S. Branicky, Spring 2024

Overview
Students explore the space of good models and reflect on those that are more explainable and open to examination for interpretation by subject matter experts and for bias by policy makers, lawyers, and the general public.

Pre-Lecture Preparation (15 minutes)
Students perform the following readings from the standard text: Stuart Russell & Peter Norvig. *Artificial Intelligence: A Modern Approach.* 4th US ed., 2021. ISBN: 978-0134610993

> Section 19.9.4; Decision Lists, including Figs. 19.10-11
> Section 19.5.1: Definitions of Explainability

Lecture Time (75 minutes)
Instructor reviews the required reading and then plays clips from the following video, hosted by the National Science Foundation:

> Prof. Cynthia Rudin, Duke University:
> "Do Simpler Machine Learning Models Exist and How Can We Find Them?
>
> Specifically, the following selected clips should be played (MM:SS, H:MM:SS in video):
>
> | | |
> |---|---|
> | 1:52–44:21 | (talk itself) |
> | | |
> | Q&A session | |
> | 49:55–53:30 | (dimension reduction, model comparison) |
> | 56:45–58:15 | (how simple are trees?) |
> | 58:56–1:01:44 | (computer vision, name giver) |
> | 1:08–1:10:31 | (repeated like Reinforcement Learning) |
> | 1:10:33–1:12:15 | (anomalies, wicked cool) |
> | 1:19:20–1:20:53 | (software) |

Reflection Questions for Homework (30 minutes)
1. What examples of potential bias and ethical implications for ML models were explored in the lecture?
2. Given that multiple good models (with the same performance) exist, how important is model explainability and transparency with respect to model usabilty and bias?

Supplementary Material
$1M Squirrel Award