AI and algorithms

Objectives

- Define an algorithm
- Identify deanonymization methods used to breach data privacy
- Introduce differential privacy as a method for protecting personally identifiable information
- Define algorithmic fairness
- Assess methods for making algorithms more "fair" and discuss their benefits and drawbacks
- Define false discovery, explain its potential harm to scientific discovery, and examine methods to reduce or prevent it
- Examine doomsday scenarios of algorithms gone wild

Assigned readings

- Kearns, M., & Roth, A. (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.
- Angwin, J., Larson, J., Kirchner, L., & Mattu, S. (2016, May 23). Machine Bias. *ProPublica*.

Optional readings

- Problems implementing differential privacy for the 2020 U.S. Census
  - Can a set of equations keep U.S. census data private?
  - https://www.nytimes.com/interactive/2020/02/06/opinion/census-algorithm-privacy.html
- Molnar, C. (2019). *Interpretable machine learning*. Lulu.com. - technical introduction to methods for interpreting black box algorithms

Response paper prompt

Kearns and Roth (2019) discuss the inherent difficulties in defining algorithmic fairness and what constitutes a fair outcome. Suppose you are tasked with constructing a model to predict the likelihood that a convicted felon will commit a new felony (e.g. re-offend) at some point in the next three years. Judges will use your recommendation to help determine the final sentence imposed on the defendant. Given your understanding of algorithmic fairness, how would you design the "best" model?

Don't worry about the mathematical elements of the model (e.g. neural network, tree-based model, penalized regression). Instead, think about the basic inputs and outputs for such a model. What are the core goals for the model? How would you prioritize the trade-off between fairness and accuracy? What types of inputs (variables) would you need to collect? Are there additional concerns of fairness or equity your model would need to address?