



# Deconstructing Machine Learning Bias

## 21:219:400 (3 Credits)

### Course Description

This course is designed to use contemporary case studies in algorithmic bias to teach students to identify and deconstruct machine learning bias. Students will learn how to combine critical reasoning, and their understanding of both the modeling process and ML techniques to identify different types of bias, to assess the impact of technical bias on the model (outcome), discuss the social and economic impact of deploying a biased model. In the second half of the term, students will apply and critique a statistical-ML solution to mitigate algorithmic bias in a case study dataset.

### Learning Objectives

This course is structured to provide students with exposure to the most frequently used ML and sampling techniques and how they can when misused introduce technical bias. Students will leverage an analytic framework to understand bias entry points and deconstruct machine learning bias. Upon completion of this class, the student will be able to work backwards from the algorithm and the data used to generate the algorithm to identify types of bias, their technical and social impact and propose or implement bias-remediation techniques in Python. As future practitioners, students will understand their responsibility to adhere to model development and deployment best practices in the face of cultural influences that could detract from ethical model development.

### Pre-Requisites

Students must have completed 21:198/219:329 (Statistics and Machine Learning) OR 21:220:303 Introduction to Econometrics (Formerly 220:322) OR are concurrently enrolled in 21:219:329 (Statistics and Machine Learning) AND 21:219:400 (Deconstructing Machine Learning Bias)

### Course Materials

#### Technical Reference Texts

- [Mastering Machine Learning Algorithms](#): Expert techniques for implementing popular machine learning algorithms, fine-tuning your models, and understanding how they work, 2nd Edition (Giuseppe Bonaccorso - paperback and Kindle version available)
- An Introduction to Statistical Learning with Application in R (James, Witten, Hastie, Tibshirani) <http://faculty.marshall.usc.edu/gareth-james/ISL/>

#### Articles and Case Studies

- Articles and Case Studies to be provided

### Software (Python and AIF360)\

- **Anaconda** (Python-Jupyter Notebook-Spyder): Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. You can install Anaconda by going to <https://www.anaconda.com/products/individual>

**Videos** – Topics include definitions of fairness and AI Bias

### Topics Covered

- Introduction to AI Bias and Fairness
- Analytic framework and Bias Identification: Machine Learning-Data Science Workflow
- Demographics and Select Sampling Techniques
- Select ML techniques and automated model data pipelines
- Contemporary Case Studies in AI and ML Bias
- Impact assessment
- Hands-on Bias Mitigation- Algorithm versus the Data Scientist

**Department of Mathematics and Computer Science, Smith Hall, 216 Warren  
Street, Newark, New Jersey, 07102, Phone: (973) 353-1004**