Ethics and Algorithms in Public Administration and Policy

Professor Jason Anastasopoulos
Department of Public Administration and Policy
University of Georgia
Email: ljanastas@uga.edu

Office Hours: Tuesday, Thursday, 3:30-4:30 PM
Course Time and Day: Wednesdays, 6:50-9:50 PM

Course Objectives

The big data revolution is transforming public policy and governance. The goal of this course is to provide a non-technical overview of some of the methods driving big data methodologies and to explore how these technologies are shaping and will shape the future of public policy and government with an eye towards the ethical dilemmas that these technologies raise. We begin the course with a discussion of some of the fundamental theories and applications of machine learning methods, which form the basis of big data and artificial intelligence technologies. We then move on to an in-depth discussion of the ethical and societal promises and perils that these technologies pose for decision making in government more broadly, focusing on the potential of these techniques to shape government and policy.

Attendance and Participation

The most important content from this class will come from the lectures and group assignments during lecture time. Because of this and the rather technical nature of this class, attendance and participation in class is extremely important. If you cannot attend a lecture you must present me with a valid excuse at least 24 hours prior to the start of class unless the situation you encountered was an emergency. Either way, absence requires explanation and documentation if you do not want want points taken off your final participation grade.

Grading and Requirements

- *Participation/discussion leader: 40%.*
- *Three (3) problem sets: 20% (lowest of 3 is dropped)*
- *Final group project proposal: 5%*
- *Final group report and presentation: 35%.*

Key Assignment Dates (to be submitted via the ELC)

- **Problem Set 1**: Distributed 01/22, Due 01/29 @ 6:49PM
- **Problem Set 2**: Distributed 01/29, Due 02/05 @ 6:49PM
- **Problem Set 3**: Distributed 02/05, Due 02/12 @ 6:49PM
- **Discussion Leader Presentations**
- **Final Group Project Proposal:** Due 03/18 @ 6:49PM
- **FInal project report and presentations:** Due 04/22 @ 6:49PM.

Key Discussion Leader Dates

- **Class 4 02/12:** *Algorithms and government - overview.*
- **Class 5 02/19:** *Algorithms, ethics and public policy I: Introduction and approaches.*
- **Class 6 02/26:** *Algorithms, ethics and public policy II: Normative approaches to fair & ethical machine learning/AI.*
- **Class 7 03/04:** *Algorithms, ethics and public policy III: Causal inference approaches to fair & ethical machine learning/AI.*
- **Class 8 03/18:** *Algorithms and government - decision making in theory.*
- **Class 9 03/25:** *Algorithms and government - decision making in practice - police and judges*
- **Class 10 04/01:** *Strengths and Weaknesses of Machine Learning Systems for Public Policy*
- **Class 11 04/08:** *Machine Learning and Bias I: Overview*
- **Class 12 04/15:** *Machine Learning and Bias II: Sources and Pathways in Practice*
- **Class 13 04/22:** *Machine learning, big data and ethics: the modern panopticon?*

Discussion Leaders

Groups of students will be assigned the role of “discussion leaders.” Discussion leaders will lead the class discussion for that week by reading the assigned content, preparing a 15-30 minute presentation summarizing the readings and will propose a series of 3-5 and discussion points to start off our discussion about the content. Every student MUST participate in a group as a discussion leader at least once. Discussion leader groups can have a maximum of 2 people and if you do not sign up for one week as a discussion leader you will be assigned to a week.

Problem Sets

During the first three weeks of the course, there will be three problem sets which you can work on in groups and which are designed to give you some hands on experience with machine learning algorithms in a policy context. These problem sets will involve some rudimentary

programming in the statistical language **R** and will teach you about some of the basic machine learning algorithms used in practice today.

Group Project

A major portion of your grade will involve writing a policy memorandum in response to a question that I will assign two weeks prior to the memorandum due date. For guidelines on how to write a policy memorandum, please see this excellent guide by Iris Malone: http://web.stanford.edu/~imalone/Teaching/ps1winter17/PS1-PolicyMemo.pdf. Policy memoranda will be graded on the basis of the quality and clarity of your writing and the quality and clarity of the ideas that you present.

Required and Recommended Texts*

Kearns, Michael, and Aaron Roth. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design.* Oxford University Press, 2019.

Machine Learning: The New AI Ethem Alpaydin (2016) MIT Press. Referred to in the schedule as **EA**.

James, Witten, Hastie and Tibshirani. 2015. *An Introduction to Statistical Learning with Applications in R*. Springer Science. Available for free here: http://www-bcf.usc.edu/~gareth/ISL/. Referred to in the schedule as **JWHT**.

Monogan III, James E. 2015. Political Analysis Using R , Springer. http://link.springer.com/book/10.1007%2F978-3-319-23446-5 . Referred to in the schedule as **M3** .

***** Most texts will be available online on the course site.

Course Outline

1. Introduction to machine learning and artificial intelligence.
2. Introductions to algorithms in government.
3. Algorithms, ethics and public policy.
    - Introduction.
    - Normative approaches.
    - Causal inference approaches.
4. Algorithms, behavior and decision-making in government.
    - Decision making by humans and machines.
    - Strength and weaknesses of machine learning and AI.
    - Machine learning and fairness.
5. Algorithms and political institutions.
    - War and international affairs
    - Policymaking.
    - Accountability.

COURSE SCHEDULE

Class 1 01/15 : *Introduction to Machine Learning and Artificial Intelligence: Fundamentals*

Programming fundamentals.

- Introduction to programming in **R**.

Overview of machine learning

- Machine learning in public organizations.
- What is machine learning?
- Supervised & unsupervised learning.
- Inference versus prediction.

Readings:

- Larson et al "How We Analyzed the COMPAS Recidivism Algorithm" https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- **EA** Chapter 1.
- Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z., 2015. Prediction policy problems. American Economic Review, 105(5), pp.491-95. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4869349/pdf/nihms776714.pdf .
- **JWHT** – Introduction, pp 1-15.

Class 2 01/22 : *Introduction to statistical learning theory*

- Training, testing and cross–validation.
- Assessing model accuracy.
- Overfitting.
- Regression vs. classification problems.
- The Bias–Variance tradeoff.
- **Application**: H1–B Visa Certification. H1–B Application Data.

Required Readings

- **JWHT –** Statistical Learning, pp 15-37, 176–184.
- **EA** - Chapter 2.

NOTE: I AM GIVING A TALK IN CAMBRIDGE ON 01/29 AND THERE WILL BE NO CLASS.

Class 3 02/05: *Understanding supervised machine learning through examples: decision trees and regression.*

- "Pure ML": Decision tree algorithms and CART.
- "Statistical ML": Linear regression as a machine learning algorithm.
- Application 1: Preventative policing: pre–crime targeting and detection. NYC Stop and Frisk Data: 2003–2016

Required Readings

- **JWHT –** Chapter 3.
- **EA -** Chapter 3.

Class 4 02/12: *Algorithms and government - overview.*

- Overview of administrative decision making and machine learning.
- Bureaucracy and technology.

Required Readings

- Coglianese and Lehr. 2017. "Regulating by Robot Administrative Decision Making in the Machine-Learning Era". Georgetown Law Journal.
- A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. *Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, Rhema Vaithianathan* ; PMLR 81:134-148
- Dick, Philip K. *The minority report. https://cwanderson.org/wp-content/uploads/2011/11/Philip-K-Dick-The-Minority-Report.pdf* (short story)

Optional Readings

- Lee, Ronald M. "Bureaucracies, bureaucrats and information technology." *European Journal of Operational Research* 18, no. 3 (1984): 293-303.

Class 5 02/19: *Algorithms, ethics and public policy I: Introduction and approaches.*

- Overview of algorithmic fairness.

Required Readings

- Kearns, Michael, and Aaron Roth. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design.* Oxford University Press, 2019. **pp 1-94**

Optional Readings

- ❖ Lepri, Bruno, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck. "Fair, transparent, and accountable algorithmic decision-making processes." *Philosophy & Technology* 31, no. 4 (2018): 611-627. https://link.springer.com/content/pdf/10.1007/s13347-017-0279-x.pdf

- ❖ Glymour, Bruce, and Jonathan Herington. "Measuring the Biases that Matter: The Ethical and Casual Foundations for Measures of Fairness in Algorithms." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 269-278. ACM, 2019.

Class 6 2/26: *Algorithms, ethics and public policy II: Normative approaches.*

- Normative approaches to creating fair algorithms.

Required Readings

- ❖ Binns, Reuben. "What Can Political Philosophy Teach Us about Algorithmic Fairness?" *IEEE Security & Privacy* 16, no. 3 (2018): 73-80.

- ❖ *Rawls, John. A theory of justice. Harvard university press, 2009.* "The Main Idea of a Theory of Justice (15 pages): https://www.csus.edu/indiv/c/chalmersk/econ184sp09/johnrawls.pdf

- ❖ Leben, Derek. "A Rawlsian algorithm for autonomous vehicles." *Ethics and Information Technology* 19, no. 2 (2017): 107-115. https://link.springer.com/content/pdf/10.1007/s10676-017-9419-3.pdf

Optional Readings

- Keeling, Geoff. "Against Leben's Rawlsian collision algorithm for autonomous vehicles." In 3rd Conference on" Philosophy and Theory of Artificial Intelligence, pp. 259-272. Springer, Cham, 2017. https://www.researchgate.net/profile/Geoff_Keeling/publication/327274610_Against_Leben%27s_Rawlsian_Collision_Algorithm_for_Autonomous_Vehicles/links/5c9fb00e45851506d736183b/Against-Lebens-Rawlsian-Collision-Algorithm-for-Autonomous-Vehicles.pdf

Class 7 03/04: *Algorithms, ethics and public policy III: Causal inference applications*

- Causal inference approaches for creating fair algorithms.

Required Readings

- Holland, Paul W. "Statistics and causal inference." *Journal of the American statistical Association* 81, no. 396 (1986): 945-960. https://www.jstor.org/stable/pdf/2289064.pdf

- Loftus, Joshua R., Chris Russell, Matt J. Kusner, and Ricardo Silva. "Causal reasoning for algorithmic fairness." *arXiv preprint arXiv:1805.05859* (2018). https://arxiv.org/pdf/1805.05859.pdf

Optional Readings

- Kilbertus, Niki, Mateo Rojas Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. "Avoiding discrimination through causal reasoning." In *Advances in Neural Information Processing Systems*, pp. 656-666. 2017. http://papers.nips.cc/paper/6668-avoiding-discrimination-through-causal-reasoning.pdf

SPRING BREAK: March 9-March 13

Class 8 03/18: *Algorithms and government - decision making in theory.*

- Decision making by humans v. machines.

Required Readings

- *Keiser, Lael. 2010. "Understanding Street Level Bureaucrats Decision Making," Public Administration Review. 70 (02) pp.247-57.* ***JSTOR***
- *Dawes et al. Clinical versus actuarial judgement*
- *Dietvorst. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*

Class 9 03/25: *Algorithms and government - decision making in practice - police and judges*

- Technology and street level bureaucracy.
- Applied examples: judges and police.

Required Readings

- Harcourt, Bernard E. "Against prediction: Sentencing, policing, and punishing in an actuarial age." (2005).
- *Green, Ben, and Yiling Chen. "Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments." In Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 90-99. ACM, 2019.* https://dl.acm.org/doi/pdf/10.1145/3287560.3287563?download=true

Optional Readings

- Kleinberg, Jon, et al. "Human decisions and machine predictions." *The quarterly journal of economics* 133.1 (2017): 237-293.

Class 10 04/01: *Strengths and Weaknesses of Machine Learning Systems for Public Policy*

- Strengths and weaknesses of the machine learning approach and how it might apply to public policy.

Required Readings

- *Breiman. Statistical Modeling: The Two Cultures*
- *Lazer et al. The parable of Google Flu.*
- *Olteanu et al. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries*

Optional Readings

- *Norvig. On Chomsky and the Two Cultures of Statistical Learning*

Class 11 04/08: *Machine Learning and Bias I: Overview*

- Defining bias in the machine learning context.

Required Readings

- *Angwin et al. Machine Bias*
- *Angwin & Larson. Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say*
- *Chouldechova. Fair Prediction with Disparate Impact: A study of bias in recidivism prediction instruments.*
- Kleinberg et al. Inherent Trade-Offs in the Fair Determination of Risk Scores

Optional Readings

- *Corbett-Davies et al. Algorithmic Decision Making and the Cost of Fairness.*

Class 12 04/15: *Machine Learning and Bias II: Sources and Pathways in Practice*

- Machine learning bias in practice.

Required Readings

- *Pierson et al. A large-scale analysis of racial disparities in police stops across the United States*
- *Caliskan et al. Semantics Derived Automatically from Language Corpora Contain Human-like Biases*
- *Torralba & Efros. Unbiased Look at Dataset Bias*

Class 13 04/22: *Machine learning, big data and ethics: the modern panopticon?*

- Ethical considerations of big data and machine learning.

Required Readings

- *Foucault. "Panopticism" in Discipline and Punishment. Pp 195-228*
- *Ohm & Peppet. What if Everything Reveals Everything?*
- *Kosinski, Stillwell, and Graepel. Private Traits and Attributes Are Predictable From Digital Records of Human Behavior*
- *Wu & Zhang. Automated Inference on Criminality using Face Images*
- *Wang & Kosinsky. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images*