PPHA 38830: The Minds Behind the Machine: Artificial Intelligence In Modern Policy

Anita Nikolich

1. **Course Number and Title**
   PPH38830 "The Minds Behind the Machine: Artificial Intelligence in Modern Policy"

2. **Class Meeting Time, Location**
   Thursday 5 PM – 7:50 PM In person Keller 2112

3. **Instructor Information**
   Office Hours are either in person immediately after class, or virtually by Zoom. Students can request office hour appointment times via email or Canvas. Standard office hour times will be posted by the start of class.
   Contact information is anikolich@uchicago.edu

4. **Teaching Assistant Information**
   TA is still TBD

5. **Course Description**
   This non-technical course introduces important topics in Artificial Intelligence that policy makers must consider. We start with foundational AI concepts such as the role of data, the meaning of algorithms, and ways in which AI is used in decision making. By using case studies on issues such as predictive policing, disinformation propagation, deep fake videos, AI in hiring, student proctoring, and criminal facial recognition, students will explore AI concepts of fairness, ethics, privacy, transparency and governance and tie them to important policy recommendations.

6. **Course Goals or Objectives**
   This course will arm students with basic knowledge about fundamental aspects of AI and major areas in which AI is used today. The students will be able to pose critical questions about how to shape future policies around its use in society, the appropriate creation of AI databases, the ownership of AI, the deployment of AI locally and globally, and the use of AI for social good. Students will be able to articulate the critical social, legal, political and ethical issues around AI, be equipped to write position papers on major topics in AI and be able to describe how AI is used in society today.

   Given the nature of the issues, and emerging policy in these areas, there is an expectation of thoughtful, energetic weekly discussions and debates. Active listening and participation is crucial. Many uses of AI are controversial, and regulation is minimal in the US. Students should be prepared to listen to a variety of opinions about the use and potential abuse of AI.

7. List of Course Texts, Required or Recommended Materials/Equipment

The readings will primarily consist of open source online material, with the addition of some academic papers and books available through the campus library system. The material is a combination of academic papers, popular press articles, critical commentary, policy analysis, government documents and short videos. Required materials versus optional readings will be noted. Materials will be uploaded into Canvas.

8. Reading/Topic List
Background Reading [Optional but helpful]

"Preparing for the Future of AI", National Science and Technology Council (2016)

2019 National AI R&D Strategic Plan, National Science and Technology Council (2019)

"Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights", White House Office of Science and Technology Policy, (2016)

The National AI Research Resource

Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence"

"AI Now Report", 2018

Week 1: An Introduction to Modern AI

Basic Definitions. What is machine learning? What is Big Data? We're going to start with the basics!

What's an algorithm and how does it work: Supervised Learning, Unsupervised Learning, Deep Learning, Language Modeling

The Importance of Data

Initial Framing of AI Issues for Policymakers

Case Study: Microsoft Tay Chatbot

Readings:

"Why Microsoft's 'Tay' AI bot Went Wrong" by Hope Reese, March 24, 2016

"A robot wrote this entire article. Are you scared yet, human?", The Guardian, Sept 8, 2020

Week 2: Can AI be Fair?

Definitions of Fairness, Accountability, Transparency, Ethics (FATE)

How do AI experts construct data sets? What's a Large Language Model?

Training Data

Data Governance

The Important Role of AI Ethics

Algorithmic Accountability

Case Study: AI for Hiring – HireVue

Case Study: AI for Test Proctoring

Week 3: AI for Social Good

How is AI Used for Drug Discovery

AI for email filtering, malware detection and cybersecurity

AI for Agriculture

AI for Wildfires and Crisis Response

AI and Social Credit in China

Guest Lecture: Brian Martin, AbbVie

Readings:

"Face Mask Recognition Has Arrived", National Geographic 2020

"How China uses AI to influence its Citizens", Engadget Sept 2020

Week 4: AI for Social Bad

AI Algorithms for Criminal Sentencing

AI for Predictive Policing

AI for Credit Scoring

AI and Smart Cities

Case Study: COMPAS Recidivism Tool

Readings:

"Machine Bias" by Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin, May 23, 2016

"AI Judges and Juries", Communications of the ACM, December 2018

"Algorithms and Sentencing, What Does Due Process Require", Brookings, March 2019

"Wrongfully Accused by an Algorithm", NY Times June 24, 2020

"Sentenced by Algorithm" by jed S. Rakoff, June 10, 2021

"QLD police will use AI to 'predict' domestic violence before it happens. Beware the unintended consequences", The Conversation, Sept 16, 2021

Week 5: AI for Facial and Speech Recognition

The mystery of chatbots

Facial Recognition

Who owns the data the underlies facial and speech recognition?

Promises and perils of Facial and speech recognition

Readings:

 "Concerns as face recognition tech used to Identify Criminals", New Scientist, December 2016

 "The "Criminality From Face" Illusion", Walter Scheirer (June 2020) – *skim this and/or watch the accompanying video:*
https://www.wjscheirer.com/misc/videos/Scheirer_Face_attributes_Talk.mp4

"Automated Inference on Criminality Using Face Images", Xiaolin Wu and Xi Zhang (2016) – *skim this*

"Facial Recognition Software Not Being Published", Harrisburg University press release

"Large Image Datasets: A pyrrhic win for Computer Vision", Anonymous submission, *briefly skim this*

Accompanying press release for the paper "Large Image Datasets" from MIT CSAIL

"MIT Apologizes", July 1, 2020

Week 6 : AI, Big Data and Corporations

Exploration of "Surveillance Capitalism"

Discuss the AI data ecosystem – data brokers, personalization, fraud detection, AdTech and how corporations use AI

Readings:

"Discrimination in Online Ad Delivery", Latanya Sweeney (2013)

Excerpt from "Surveillance Capitalism", Shoshanna Zuboff

Week 7: AI's Role in Online Content: Fake News, Disinformation, Misinformation, Deep Fakes

Brief History of Information Operations

Explanation of Deep Fakes and Cheap Fakes

Deep Fakes as a Threat

How Does AI perpetuate Disinformation?

Case Study: 2016 Election and Russian Trolls

Case Study: Cambridge Analytica and 2016 elections

Readings:

"Get Ready for Deepfakes to be Used in Financial Scams by Jon Bateman"

"Weaponized Tweets: AI Could Help Defend Against Adversary Attacks in Social Media", War on the Rocks, Oct 16, 2020

Week 8: Assessing AI Safety, Risk and Regulation

How is AI developed and tested in Autonomous Vehicles, Autonomous Weapons, Robots

Who Regulates AI?

Discussion of the roles of NIST, the FTC, FAA, FDA and Other Agencies

Cultural and Ethical Differences in Creating and Using AI Systems

International Implications in AI Safety, Risk and Regulation

"[Using AI and Algorithms](#)" FTC Bureau of Consumer Protection

"[Why We Need to Audit Algorithms](#)", HBR November 2018

"[New York City's Algorithm Task Force is Fracturing](#)", The Verge 2019

"Auditing Algorithms: Research Methods for Detecting Discrimination", Sandvig (2014)

Week 9: Attacking AI

What is Red/Blue/Purple Teaming?

What is Adversarial AI

How do we know an AI System was Attacked?

Readings TBD

Week 10: Project Presentations

9. **Major Assignments:**
   The course includes 3 written/presentation assignments:
   i. Two (2) short (2 pages max) policy focused "Position Papers" on a topic of interest to the student with regard to an AI Policy issue. Examples: The use of AI and Facial Recognition should be adopted by social service agencies to assist citizens; Facebook AI-driven content should be regulated; the use of AI in credit scores must be further regulated. White papers must explicitly describe why this is an AI-related policy issue and the outcome the paper is seeking to influence. Examples will be provided in Canvas. One is due by the end of Week 4 and one is due by the end of Week 8. These must be submitted online via Canvas. The brief position papers are an exercise for the students in having to summarize complex topics in succinct prose and convey policy recommendations in a simple, easy to read format.

   ii. A final group project presentation (3-5 students per group) which will be presented during the final class in Week 10. The material for the group presentation (slides, a 3-4 page paper) must be submitted by at least one of the group via Canvas, with the group members noted in the submission. The group

projects are meant to explore more complex topics such as Regulating Ethics in AI or Legislating Transparency and present a 10-15 min talk conveying background and recommendations for policy. Students will pick a topic of interest to the group and submit this to the TA or Instructor via Canvas no later than Week 7. Example presentation topics might include: creating a plan for a disinformation campaign using AI, outlining a framework for Ethical AI, creating a simple deep fake video on a timely topic, describing how you'd re-create an unbiased Automated Decision System (ADS), or anything creative that displays an understanding of the nuances of AI Policy.

10. **Grading Standards and Criteria**

If students are opting for a Letter grade, the course assignments are weighted as such:

Class Discussion	10%
Position Paper 1	30%
Position Paper 2	30%
Group Project	30%

This class is offered as a Pass/Fail. Students earning a "pass" grade have completed the course in full, including attendance at all lectures (unless extenuating circumstances prevent attendance), participation in class discussions and demonstration of an understanding of the course material.

Students who wish to take the course pass/fail rather than for a letter grade must use the Harris Pass/Fail request form (https://harris.uchicago.edu/form/pass-fail) and must meet the Harris deadline, which is generally 9am on the Monday of the 5th week of courses. Students who take the course pass/fail must attend class meetings and turn in all assignments, achieving marks on assignments that are overall commensurate with at least a C- letter grade.

To earn a passing grade, students taking the course pass/fail must: complete and submit all assignments; attend classes and participate in class discussions.

11. **Policies**

    a. Students should make every effort to attend the class each week except in extenuating circumstances.
    b. Assignments will only be submitted though Canvas.
    c. Late assignments are accepted with prior approval for extenuating circumstances.
    d. Required citation style or other requirements for written assignments
    e. Students are expected to participate in class discussions. Given the often-controversial nature of AI policy topics, class discussions will not be recorded officially or unofficially. It is the expectation that students can respectfully debate controversial topics and respect the opinions of fellow students. Students will not be permitted to record any class discussions. Violation of this policy will be referred to the Area Dean of Students.
    f. Questions should be directed to the TA

g. The class will use Canvas for course materials, Zoom for office hours, and Panopto for pre-recorded videos.

h. By attending course sessions, students acknowledge that:

    i. They will not: (i) record, share, or disseminate University of Chicago course sessions, videos, transcripts, audio, or chats; (ii) retain such materials after the end of the course; or (iii) use such materials for any purpose other than in connection with participation in the course.

    ii. They will not share links to University of Chicago course sessions with any persons not authorized to be in the course session. Sharing course materials with persons authorized to be in the relevant course is permitted. Syllabi, handouts, slides, and other documents may be shared at the discretion of the instructor.

    iii. Course recordings, content, and materials may be covered by copyrights held by the University, the instructor, or third parties. Any unauthorized use of such recordings or course materials may violate such copyrights.

    iv. Any violation of this policy will be referred to the Area Dean of Students.

12. *Resources available to students*

The University has long offered a comprehensive set of student support services (described here), including student health services. And in response to the COVID-19 pandemic and associated disruptions, the University has provided links for students via its "Learning Remotely" website; specific resources are listed here. Specifically, instructors may wish to explicitly remind students about available counseling services in particular: *If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, remote counseling services are available. Student Counseling Service (SCS) urges you to attend to your mental wellbeing and to reach out to them for support during these challenging times. All SCS services are covered by the Student Life Fee, and there is no additional cost for students to access their services. See https://wellness.uchicago.edu/mental-health/student-counseling-service-spring-quarter-faq/. Students seeking new services/resources can call 773.702.9800 during business hours (Monday–Friday 8:30 a.m.–5 p.m.) and ask to speak with a clinician**. Students needing urgent mental health care can speak with clinicians over the phone 24/7 by calling the SCS at 773.702.3625.***

a. Similarly, the Harris School itself provides both academic and non-academic support services for students. These resources are described (and links provided) via the Canvas site *Harris Student Handbook*, which all Harris instructors can access.

13. Students should be aware of both the University and Harris School-specific guidelines and principles in at least three key domains: academic integrity; students with disabilities; and diversity and inclusion.

Academic Integrity

All University of Chicago students are expected to uphold the highest standards of academic Integrity and honesty. Among other things, this means that students shall not represent another's work as their own, use un-allowed materials during exams, or otherwise gain unfair academic advantage.

- The University's policies regarding academic integrity and dishonesty are described here. It is worth explicitly stating the University's approach here: "It is contrary to justice, academic integrity, and to the spirit of intellectual inquiry to submit another's statements or ideas as one's own work. To do so is plagiarism or cheating, offenses punishable under the University's disciplinary system. Because these offenses undercut the distinctive moral and intellectual character of the University, we take them very seriously."
- The Harris School's policies are available in the *Harris Student Handbook* Canvas site.
    - The *Academic Honesty and Plagiarism* section expresses the main principles.
    - Detailed guidelines for more specialized student work (e.g., problem sets including computer code) are offered in the section titled *Harris Integrity Policy for Problem Sets Involving Code.*
    - Harris's specific procedures for handling suspected violations of these policies are available in the section *Harris Procedures for Allegations of Plagiarism, Cheating, and Academic Dishonesty* and are also re-produced as an Appendix to this document.
    - All students suspected of academic dishonesty will be reported to the Harris Dean of Students for investigation and adjudication. The disciplinary process can result in sanctions up to and including suspension or expulsion from the University.
    - At the instructors' discretion, the student may receive a failing grade for the course regardless of their performance on elements of the course. Students found in violation of academic dishonesty will receive a failing grade in this course.

Disability Accommodations

The University's policies regarding students with disabilities are available here. Students who have disability accommodations awarded by the University Student Disability Services Office should inform the Harris Dean of Students office by the end of the first week of class. The Harris Dean of Students Office will work with the student and instructor to coordinate the students' accommodations implementation.

Harris students are not required to submit their accommodations letter to the instructor. Students from other divisions in the University must submit their accommodations letter to either the instructor or the Harris Dean of Students Office.

Students who do not yet have formal accommodations in place but who feel they need accommodations on a temporary or ongoing basis should contact the Harris Dean of Students Office or Student Disability Services.

Diversity and Inclusion

The Harris School welcomes, values, and respects students, faculty, and staff from a wide range of backgrounds and experiences, and we believe that rigorous inquiry and effective public policy problem-solving requires the expression and understanding of diverse viewpoints, experiences, and traditions. The University and the Harris School have developed distinct but overlapping principles and guidelines to insure that we remain a place where difficult issues are discussed with kindness and respect for all.

- The University's policies are available here. Specifically, the University identifies the freedom of expression as being "vital to our shared goal of the pursuit of knowledge, as is the right of all members of the community to explore new ideas and learn from one another. To preserve an environment of spirited and open debate, we should all have the opportunity to contribute to intellectual exchanges and participate fully in the life of the University."
- The Harris School's commitments to lively, principled, and respectful engagement are available here: "Consistent with the University of Chicago's commitment to open discourse and free expression, Harris encourages members of the leadership, faculty, student body, and administrative staff to respect and engage with others of differing backgrounds or perspectives, even when the ideas or insights shared may be viewed as unpopular or controversial." We foster thought-provoking discourse by encouraging community members not only to speak freely about all issues but also to listen carefully and respectfully to the views of others.

Appendix: Full Harris Academic Integrity Procedures

Harris Procedures for Allegations of Plagiarism, Cheating, and Academic Dishonesty

First Violation

If a student is accused by an instructor or teaching assistant of plagiarism, cheating, or any other form of academic dishonesty, the student will be summoned to meet with the Dean of Students and the instructor. In the meeting, the student and instructor both present information about the situation. If it is determined by the instructor and the Dean of Students that the student has, in fact, plagiarized or cheated, the following sanctions will be imposed for the first violation:

- The student will generally receive a grade of 0 on the assignment or exam in question. Please note that grading decisions are fully at the discretion of the instructor, who may decide to impose harsher grade penalties.
- The student may be asked to re-do the assignment or retake the exam (without credit) to ensure that the student has learned how to properly cite sources or demonstrate that he or she has command of material covered.

- A formal letter of finding is sent to the student stating that the student has been found in violation of the code of academic honesty and what the sanctions were. The letter, along with any evidence presented, is archived in Harris Student Affairs records until the student graduates if the student has no other violations.
- Students found in violation of the academic honesty policy are not permitted to withdraw from the course to avoid grade penalties from the instructor.
- In cases where plagiarism or academic dishonesty is egregious, the case may be referred to the Area Disciplinary Committee even on a first offense. The Dean makes all decisions about which cases will go before the Area Disciplinary Committee.

Second Violation

If a student who has already been found in violation academic dishonesty is again accused of academic dishonesty, the case will be sent to the Harris Area Disciplinary Committee. Details about the Area Disciplinary Committee procedures can be found in the University Student Manual. Information about the first violation, including the formal letter of finding any evidence, will be presented to the Area Disciplinary Committee, along with evidence of the current allegation. If the student is found in violation of academic honesty a second time, the Area Disciplinary Committee can assign sanctions including transcript notes, disciplinary probation, suspension or expulsion from the University.

Academic Dishonesty Appeals

If a student has been found in violation of academic honesty and does not believe that either the finding or the sanction is fair or correct, the student has the right to appeal the finding by requesting a hearing from the Area Disciplinary Committee. More information about the Area Disciplinary Committee is available here.