Ethical Issues in Big Data

Instructor: Alice Huang
Email: alicealice.huang@mail.utoronto.ca

Please do NOT email about substantive questions on the course content. Instead, raise these questions or thoughts either in class or during office hours.

Course Meeting Information

- Tuesdays 15:00-18:00
- Office hours: TBA
- Location: BA2135

Course Description

In this course, we tackle three different topics in the ethics of big data: fairness, privacy and explanability. We look at questions such as: What it means for an algorithm to be fair? What are the risks involved in relying on machine predictions without being able to explain how these predictions came about? What are the ethical concerns unique to using data about people, and what can we do to ensure that their privacy is protected while promoting transparency?

Learning Outcome and Format

Some key skills we will learn throughout the course include: how to write a philosophy paper that involves interdisciplinary engagement, how to clearly explain and reconstruct arguments, how to analyze and evaluate ethical issues related to technology, how to articulate one's own views and engage in philosophical discussions.

Each student has the option to work on a case study project instead of a traditional final paper. In the project, you will study a particular ethical problem that a practitioner might face when making decisions. For example, you might look at the prediction of an algorithm and analyze whether it is fair by engaging with the theories of fairness we look at in the course. AI ethics is a topic that is very pertinent to our society today. The purpose of the project is to give you a taste of the kinds of practical problems we might face when making decisions in machine learning beyond pure theory.

Basic knowledge in mathematics, especially probability, is assumed. Those with some background in statistics and programming are encouraged to use their skills in the case study project.

Evaluation

- Participation (10%)
- Weekly short responses or worksheet (20%)

- Each week, you will be asked to hand in a short response to one of two assigned questions about the readings. These are meant to be basic questions to make sure you stay on track and act as short writing practices.
- Or, you will be asked to do some simple practice exercises to help you understand some of the formal aspects of machine learning, such as applying fairness criteria to prediction outcomes.
- There is no extension on these short assignments, but only on the 8 best short weekly assignments will count towards your grade. This means that you can miss 2 assignments.

- Midterm exam (25%)
- Final project (45%)

Week-by-week Breakdown of In-class Activities

All readings will be available on Quercus

Session/Date	Topic	Readings
Jan 9	Introduction	- The syllabus! - Introduction to Statistical Learning, 2.1-2.2, Gareth and Witten
Jan 16	Moral encroachment	- On the Epistemic Cost of Implicit Bias (2011) Gendler - Belief, Credence and Norms (2014) Buchak
Jan 23	Stereotypes	- Algorithmic bias: on the implicit biases of social technology (2020) Johnson - Normativity, Epistemic Rationality and Noisy Statistical Evidence (2021) Babic
Jan 30	Fairness	- The Badness of Discrimination (2006) Lippert-Rasmussen - The Measure and Mismeasure of Fairness (2018) Goel and Corbett-Davies
Feb 6	Fairness	Guest Lecture
Feb 13	Fairness case study	- Machine Bias, Pro Publica (2017) Angwin - Pro Publica: How We Analyzed the COMPAS Recidivism Algorithm - Is it impossible to be fair? (https://jainfamilyinstitute.github.io/algorithmic-fairness/)
	Reading Week	
Feb 27	Fairness	- Measuring Algorithmic Fairness (2020) Hellman - On Statistical Criteria of Algorithmic Fairness (2021) Hedden

Mar 5	Explanability	- Transparency in Complex Computational Systems (2020) Creel - Understanding from Machine Learning Models (2022) Sullivan
Mar 12	Explanability	- The Ethical Algorithm, chapter 5 (2019) Kearns and Roth - The Algorithmic Explainability Bait & Switch (2024) Babic
Mar 19	Explanability	- Conceptual Challenges for Interpretable Machine Learning (2022) Watson - Why Should I Trust You? Explaining the Predictions of Any Classifier (2016) Ribeiro
Mar 26	Privacy	- The Ethical Algorithm, chapter 1 (2019) Kearns and Roth - "I've Got Nothing to Hide" and Other Misunderstanding of Privacy (2019) San Diego Law Review
Apr 2	Privacy	- Privacy in the Age of Medical Big Data (2019) Cohen and Price - Big Data's Epistemology (2018) Skopek

Course/Departmental/Divisional policies

- You are expected to attend and participate in classes. Please do the readings beforehand.
- Assignments should be submitted through Quercus.
- Late penalty is 1/3 letter grade per day (e.g., A becomes A- after one day.) unless extensions are granted. No late assignment is accepted for weekly short assignments. Please ask for extension before the day of the deadline.
- All regrade requests must be made within 24 hours - 2 weeks of the assignment return date.

Respect for Classmates

The University of Toronto is committed to equity, human rights and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect where all members of our community can express themselves, engage with each other, and respect one another's differences. U of T does not condone discrimination or harassment against any persons or communities.

Academic Integrity

Students must adhere to the Code of Behavior on Academic Matters. You are responsible for ensuring that you do not act in such a way that would constitute cheating, misrepresentation, or unfairness,

including but not limited to, using unauthorized aids and assistance, personating another person, and committing plagiarism. For more information see U of T Academic Integrity website.

Academic integrity includes understanding appropriate research and citation methods. If you are uncertain about this, please seek out additional information from the instructors or from other institutional resources including the following:

- This tip sheet provides clear and helpful information about appropriate academic citation: http://guides.library.utoronto.ca/citing
- This site offers a series of scenarios to help students understand how to prevent themselves from being subject to academic offence allegations https://www.utm.utoronto.ca/academic-integrity/students/scenarios
- Before handing in assignments students can also review this academic integrity checklist provided by the UofT Centre of Teaching Support & Innovation:
    - I have acknowledged the use of another's ideas with accurate citations.
    - If I used the words of another (e.g., author, instructor, information source), I have acknowledged this with quotation marks (or appropriate indentation) and proper citation.
    - When paraphrasing the work of others, I put the idea into my own words and did not just change a few words or rearrange the sentence structure.
    - I have checked my work against my notes to be sure I have correctly referenced all direct quotes or borrowed ideas.
    - My references include only the sources used to complete this assignment.
    - This is the first time I have submitted this assignment (in whole or in part) for credit.
    - Any proofreading by another was limited to indicating areas of concern which I then corrected myself.
    - This is the final version of my assignment and not a draft.
    - I have kept my work to myself and did not share answers/content with others, unless otherwise directed by my instructor.
    - I understand the consequences of violating the University's Academic Integrity policies as outlined in the Code of Behaviour on Academic Matters.
- Please note that the use of generative AI (e.g., ChatGPT) for assignments is considered a violation of academic integrity in this course, unless prior approval from the instructor is granted.

Accessibility

Students with diverse learning styles and needs are welcome in this course. In particular, if you have a disability or health consideration that may require accommodations, please feel free to approach me/us and/or the Accessibility Services Office as soon as possible. The Accessibility Services staff are available by appointment to assess specific needs, provide referrals and arrange appropriate accommodations. The sooner you let them and me know your needs, the quicker we can assist you

in achieving your learning goals in this course. For more information, or to register with Accessibility Services, please visit: http://studentlife.utoronto.ca/as.

Student Mental Health Resources

- U of T's Central Hub for Student Mental Health Resources
- Student Life Health and Wellness
- MySSP: 1-844-451-9700 (or use the app)
- Good2Talk: Call: 1-866-925-5454 or Text: GOODTOTALKON to 686868

UofT Academic Success Centre

- Offers group workshops and individual counselling to develop strategies for a range of learning challenges such as time management, stress and anxiety, memory, exams, note taking, textbook reading, concentration.

Writing Support

- The University of Toronto provides a number of resources to help students improve their writing. Visit the Writing Center: https://writing.utoronto.ca/