

CompSci 1260
Fairness and Privacy: Perspectives of Law and Probability
Fall 2024

Joint class: Mondays, 8:45-10:00 a.m., WCC 2009
Law class: Tuesdays, 9:00-10:15 a.m., WCC 2009
CS class: Wednesdays, 9:00-10:15 a.m., Maxwell-Dworkin G-125

Instructor(s)

Professor Cynthia Dwork

150 Western Ave, Sci&Eng 3.305

dwork@seas.harvard.edu

Office Hours (virtual):

- (1) "I'm Lost" OH: (most) Wednesdays 8-9 PM (and occasionally on Thursdays). Come with questions on the CS lectures and readings.
- (2) Joint Readings Discussion: Mondays 8-9 PM. Shy about speaking out in joint class? Come talk about the joint readings and how they relate to our CS work.

Assisted by Allison Choat

achoat@seas.harvard.edu

Professor Martha Minow

Griswold 407

minow@law.harvard.edu

Office Hours: Email Professor Minow (minow@law.harvard.edu) to make an appointment

Assisted by Caroline Fitzgerald

cfitzgerald@law.harvard.edu

Course description

When -- and how -- can governments and law and private actors use group traits to influence or determine the treatment of individuals? Persistent debates over the use of racial categories in public and private decisions now reach new contexts, such as the use of group traits inferred by large data and deployed by algorithms in criminal legal processes, employment practices, and access to credit, and the distribution of images in generative models. When do these practices run afoul of conceptions of fairness in law or in computer science? When do alternatives even exist? In addition to questions about fairness and equality, privacy faces new jeopardy from contemporary data gathering and deployment in public and private settings. When do these activities—including machine learning, targeted advertising, and interactive Large Language Models -- run afoul of conceptions of privacy in law and ethics? What elements of legal and quantitative reasoning exacerbate or resolve such issues, and how can people with training in one field better collaborate with those from other disciplines? New approaches to data analysis quantify and control individual privacy loss while revealing information about larger groups. What are these techniques and when may they offer alternatives to or modes of compliance for legal restrictions? A persistent theme is: compared to what? Whatever problems are raised by the new tools, how do they compare to the problems with the existing human arrangements?

This intensive course brings together advanced students in computer science, statistics, law, and government to tackle these and related questions while also offering opportunities for hands-on problem-solving and policy development. Offered concurrently by HLS and SEAS, with interwoven tracks emphasizing, respectively, law and computer science, the tracks will meet jointly and separately. A particular learning goal in the course is expanding cross-disciplinary communication and problem-solving. Because the issues are so timely, the readings and topics may change.

CS-Specific Learning Objectives

Students will learn theoretical underpinnings of algorithmic fairness, differentially private data analysis, and cryptography, emphasizing common underlining themes and conceptual breakthroughs. The course will end with three advanced topics lectures.

Through joint readings and weekly class meetings with the HLS track, students will develop disciplinary “bilingualism” and some understanding of the relationship between technology and applicable law.

Additional Instructional Staff

Safwan Hossain, shossain@g.harvard.edu, TF

Course Logistics and Expectations

The Law School and Computer Science classes will meet jointly on Mondays, and the CS track will meet on Wednesdays. The Law track will meet on Tuesdays. CS students are invited to attend the Law classes and vice versa.

Students are expected to attend all classes in their track and all joint classes, and to contribute to in-class and online discussions. Email the TF if you will be absent.

Readings and Access

All CS and joint readings will be available through Canvas or easily obtained online. Most needed links are live in this document. Chance favors the prepared mind! Completing the CS readings before class will substantially help students to get the most out of the lectures, which in turn will help with the problem sets.

A technical article may require mathematical background that you do not have. Don’t panic! Here are some hints on how to proceed. Always read the introduction. If you don’t know the technical background, then read each theorem (or lemma or claim) statement and the language just before and after the theorem -- it is common for that text to contain English descriptions of the point of the theorem. This text can help you understand what the theorems are contributing and why the authors think they are important. Do the best you can; you will get better at it with practice.

Assignments and Grading Procedures

Readings. There will be weekly readings for the individual tracks and for the joint classes. These are mandatory. Students should come to class prepared to discuss the readings. We will have some online form of discussion on the joint readings, and participation is mandatory.

Problem sets will be assigned on a two-week cycle. Students are encouraged to work in groups of 2-3, but solutions must be written up individually. Students will have a fund of 8 late days for problem sets, to be used as needed with a maximum of 2 late days for any given problem set. Assignments should be submitted through Gradescope, typeset in LaTeX.

Cross-Disciplinary Interaction. Students will be placed in small multi-disciplinary groups, mixing HLS and CS students. You are to act as consultants for one another and practice communicating with people from a very different intellectual tradition.

CS Class and Joint Class Participation. Active classroom participation and discussion is an important part of the course! You should be thinking not only about the technical content but also the broader social context for the work and its relation to law. Asking questions is welcome and encouraged. Joint classes will typically be discussions, so come prepared by doing the reading.

We understand that asking questions and sharing ideas in class can be daunting, and if you are having difficulty with this aspect of the course requirements please talk to a member of the instructional staff as soon as possible. Active participation in office hours can contribute to your participation grade, but this is neither necessary nor sufficient for full participation credit.

There are no late days for joint class assignments.

Grading will be roughly: 70% homework, 20% final exam, and 10% class participation, including demonstrating familiarity with joint readings via participation in the joint classes, online discussion, section meetings, and office hours, and consulting with the HLS students on their final projects.

Collaboration, Use of AI, and Academic Integrity

Students may collaborate on the problem sets in groups of 2-3. However, each student must write a complete solution on his/her/their own. Names of your collaborators and all sources, including published papers and web sites, must be cited. Unless otherwise noted in the posted instructions for an assignment, we specifically forbid the use of ChatGPT or any other generative artificial intelligence (AI) tools at all stages of the work process, including preliminary ones. Violations of this policy will be considered academic misconduct. We draw your attention to the fact that different classes at Harvard could implement different AI policies, and it is the student's responsibility to conform to expectations for each course.

Students are expected to maintain the highest standards of academic and social integrity.

Accommodations for Students with Disabilities

Students needing academic adjustments or accommodations because of a documented disability must present their Faculty Letter from the [Accessible Education Office \(AEO\)](#) and speak with the professor by the end of the second week of the term, September 13, 2024. Failure to do so may result in the Course Head's inability to respond in a timely manner. All discussions will remain confidential.

Course Outline

This course studies 3 topics: Algorithmic Fairness, Differential Privacy, and Cryptography, and will also touch on Adaptive/Exploratory data analysis.

The theory of algorithmic fairness. We will introduce 3 flavors of fairness guarantees: group fairness notions, individual (aka metric) fairness, and something in-between, which we will denote multi-X. Algorithms satisfying this in-between notion are required to simultaneously ensure group fairness constraints for a very large set of arbitrarily complex and intersecting groups. Material will mostly be drawn from scholarly articles.

Differential Privacy (DP) is a definition of privacy tailored to the statistical analysis of large datasets. The term now also comprises a collection of algorithmic techniques for ensuring various sorts of data analyses. The course will introduce some basic building blocks and also see some advanced application of these techniques. Material will be drawn mostly from Dwork and Roth, [The Algorithmic Foundations of Differential Privacy](#).

Cryptography. Key cryptographic concepts will include computational indistinguishability, pseudo-randomness, encryption, and zero-knowledge. Material will be drawn from Katz and Lindell, Introduction to Modern Cryptography; and Goldwasser and Bellare, [Lecture Notes on Cryptography](#).

Adaptive Data Analysis. Students will be introduced to the problem of adaptive (aka exploratory) data analysis – that is, data analysis in which the statistics computed or questions asked depend on the data themselves and are not chosen in advance -- and will see one general technique for addressing this challenge. Material will be drawn from adaptivedataanalysis.com.

List of CS Class Lectures

CS Lectures (Wednesdays)

9/4 Intro and Overview; indistinguishability: in DP, in crypto, in fairness
9/11 Fairness 1 Individual and group fairness; impossibility results
9/18 Fairness 2 Individual Fairness: Definitions and basic results
9/25 Fairness 3 Multi-group fairness
10/2 Privacy 1 Reconstruction theorem; intro to DP; properties of DP; Laplace Mechanism
10/9 Privacy 2 DP Primitives: High-Dimensional Laplace; Report Noisy ArgMax; Exponential Mechanism
10/16 Privacy 3 Sparse Vector; Approximate DP; Advanced Composition
10/23 Crypto 1 indistinguishability and hard-core bits
10/30 Crypto 2 pseudo-randomness, PRGs from one-way permutations
11/6 Crypto 3 encryption
11/13 Advanced Topic in DP: Validity in Adaptive Data Analysis
11/20 Advanced Topic in Algorithmic Fairness (causality)
11/27 No CS class (Thanksgiving break)
12/04 Advanced Topic in Cryptography (tentatively: Zero Knowledge)

Joint Class Assignment and Preparation Due Dates

September 9 (class preparation prompt)
September 16 (prompt)
September 30 (joint assignment due at 6pm)
October 7 (prompt)
October 21 (prompt)
November 4 (class preparation prompt AND joint assignment due at 6pm)
November 18 (prompt and 2 videos (20 mins each))

More details about the joint assignments:

There will be two joint assignments involving small teams of both law students and computer science students collaborating and posting results on the class website. We will help constitute the groups during an early joint class session.

Joint Discussion Post #1: Due no later than **September 30th by 6pm ET**

During class, we will discuss the privacy issues related to large language models as used in health care. In small groups combining law students and computer science students, each group will look at two or three of the recommendations of Meskó and Eric. J. Topol (first article in required readings for Sept. 30), offer criticisms and alternative recommendations in no more than 4 paragraphs. Each group should post to the class website a summary of their recommendations.

Joint Discussion Post #2: Due no later than **November 4th by 6pm ET**

Working in the same groups as joint discussion post #1, students should find a way to meet outside of class time to discuss this assignment.

The assignment for each group is to address the key elements of an algorithmic hiring tool if turned on the interviewers to determine potential bias along any demographic dimension as perceived by the camera during the first minute of the interview. What are the technical and legal considerations in using such a system to evaluate discrimination on the basis of protected traits in hiring? Summarize your analysis in a few paragraphs.

In addition, as we work on communicating across disciplinary differences, class **preparation for discussion for the November 4th joint class** includes the following assignments:

- Law students should be prepared to explain disparate impact analysis
- Computer science students should be ready to discuss thoughtfully whether programmers may use racial and gender classifiers to test whether their algorithms that do not use those classifications to optimize employment decisions have disparate racial or gender effects.

CS Problem Set Due Dates

(roughly every two weeks; details to be filled in shortly)

Pset 1 Sunday, September 15

Pset 2

Pset 3

Pset 4

Pset 5

Pset 6

CS Weekly Readings

September 4: no assignment

September 11

- Dwork, Hardt, Pitassi, Reingold, and Zemel, “[Fairness through Awareness](#),” ITCS 12 Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, pp. 214-226 (Jan. 8-10, 2012): Sections 1 through 3.1 and Appendix A

September 18

- Dwork, C. and Ilvento, C., 2019. [Fairness Under Composition](#). *10th Innovations in Theoretical Computer Science*, Sections 1 through 4

September 25

- Philip Dawid, “[On Individual Risk](#),” *Synthese* 194 (9):3445-3474 (2017), Sections 1 through 3.3, and 4 through 6.2.
- Hébert-Johnson, Ú., Kim, M., Reingold, O. and Rothblum, G., 2018, July. [Multicalibration: Calibration for the \(computationally-identifiable\) masses](#). In *International Conference on Machine Learning* (pp. 1944-1953), Sections 1, 2, and 6
- Dwork, C., Kim, M.P., Reingold, O., Rothblum, G.N. and Yona, G., 2021, June. Outcome indistinguishability. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (pp. 1095-1108). [ArXiv Version](#). Sections 1, 2, and 4.

October 2

- Dwork, C. and Roth, A., 2014. [The algorithmic foundations of differential privacy](#). *Foundations and Trends® in Theoretical Computer Science*, 9(3–4), Sections 8.1 and Sections 1, 2, and 3 through Example 3.3 on page 34, Section 3.5 up to but not including 3.5.1.

October 9

- Dwork, C. and Roth, A., 2014. [The algorithmic foundations of differential privacy](#). *Foundations and Trends® in Theoretical Computer Science*, 9(3–4), Remainder of Section 3.3 (right after example 3.3 on page 34) and Section 3.4

October 16

- Dwork, C. and Roth, A., 2014. [The algorithmic foundations of differential privacy](#). *Foundations and Trends® in Theoretical Computer Science*, 9(3–4), Section 3.5 except Lemma 3.17.

October 23

- Goldwasser, S. and Bellare, M., 2008. [Lecture Notes on Cryptography](#), Appendix B (complexity theory background)
- Katz, J. and Lindell, Y., 2015. *Introduction to modern cryptography (Second Edition)*. Chapman and Hall/CRC, Sections 1.4 (general principles), Section 7.8 through Definition 7.30 (computational indistinguishability), Sections 7.1 - 7.2 (one-way functions)
- Goldwasser, S. and Bellare, M., 2008. [Lecture Notes on Cryptography](#), Section 1.2 (general), Sections 2.1-2.2 (one-way functions)

October 30

- Goldwasser, S. and Bellare, M., 2008. [Lecture Notes on Cryptography](#), Section 3 (pseudo-random generators)
- Katz, J. and Lindell, Y., 2015. *Introduction to modern cryptography (Second Edition)*. Chapman and Hall/CRC, Section 3.1.2 (the asymptotic approach), Section 3.3.2 (proofs by reduction), 7.4.1-7.4.2 (constructing pseudo-random generators)

November 6

- Goldwasser, S. and Bellare, M., 2008. [Lecture Notes on Cryptography](#), Section 7 (Public-Key Encryption): through Section 7.1, Section 7.3-Section 7.4.4
- Katz, J. and Lindell, Y., 2015. *Introduction to modern cryptography (Second Edition)*. Chapman and Hall/CRC, Section 3.2 (defining computationally secure encryption)

November 13

- Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O. and Roth, A., 2015. [The reusable holdout: Preserving validity in adaptive data analysis](#). *Science*, 349(6248), pp.636-638.
- Dwork, Cynthia, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth. "[Preserving statistical validity in adaptive data analysis](#)." In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pp. 117-126. 2015.

November 20

- Lily Hu, Issa Kohler-Hausman, "[What's Sex Got To DO With Fair Machine Learning?](#)" (June 4, 2020
- <https://simons.berkeley.edu/talks/morning-session-causality>; please watch Moritz Hardt at minute 22

December 4

- TBD

Readings for Joint and Law Classes, & Joint Prompts and Assignments

Week of September 2 (no joint class):

Topic 1 - Introduction: Technological Possibilities, Fairness, and Privacy

September 3: Law Class: Introduction and Emerging Issues in Law

Developments in Artificial Intelligence and Machine Learning generate urgent questions about human values, including fairness and privacy, which in turn call for attention to how these values are and should be defined, assessed, and advanced. What is “determined” by technological and technical developments, what are dangers, and what are promising possibilities? How do existing

legal frameworks address increasing use of computational and algorithmic tools in public and private settings and how should those frameworks change to better realize legal and ethical goals? How might new tools of “artificial intelligence” advance conceptions of fairness and privacy?

Required readings:

- James M. Manyika, “Getting AI Right: Introduction Notes on AI & Society,” *Daedalus* (Spring 2022), <https://www.amacad.org/publication/daedalus/getting-ai-right-introductory-notes-ai-society>
- John Tasioulas, “Artificial Intelligence, Humanistic Ethics,” *Daedalus* (Spring 2022), <https://www.amacad.org/publication/artificial-intelligence-humanistic-ethics>.
- State of AI: <https://www.stateof.ai/>

Optional reading:

- Owen Hughes, Generative AI Defined: How it Works, Benefits, and Dangers TechRepublic (April 25, 2024), <https://www.techrepublic.com/article/what-is-generative-ai/>
- Richard Ngo, Visualizing the Deep Learning Revolution, Medium, <https://medium.com/@richardcngo/visualizing-the-deep-learning-revolution-722098eb9c5>
- Jai Vipra & Sarah Myers West, “Computational Power and AI,” AI Now Institute, (Sept. 27, 2023), <https://ainowinstitute.org/publication/policy/compute-and-ai#h-where-to-go-from-here-points-for-future-policy-intervention>.
- Abigail Sellen and Eric Horvitz, The Rise of the AI Co-Pilot: Lessons for Design from Aviation and Beyond, <https://arxiv.org/abs/2311.14713>

September 4: CS Class: Intro and Overview; indistinguishability: in DP, in crypto, in fairness

Week of September 9:

Joint Topic 2 – Fairness and Algorithms

September 9: Joint Class: What concepts and tools do computational resources offer for realizing legal values and policies? What is “fairness” and how can it be defined in practical ways to design and assess algorithmic and computational tools increasingly used in public and private settings? When does fairness refer to groups and when to individuals—with what consequences?

Prompt: **Law Students:** be ready to discuss with Computer Science students your thoughts about what challenges arise with the use of criteria that are not specified in advance, as in clinical decision-making?

Prompt: **CS Students:** be ready to discuss with Law Students (1) from a technical perspective, why might it be problematic not to specify criteria in advance? (2) What legal issues do you expect might be raised when law uses prediction rather than retrospective evidence?

Required joint readings:

- Barbara D. Underwood, "Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment," 88 YALE L.J. 1409 (1979); <https://digitalcommons.law.yale.edu/ylij/vol88/iss7/2> (be sure to read pp. 1409-top of 1420, rest is optional)
- Virginia D. Eubanks, "The Allegheny Algorithm," Ch. 4 in *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2015), <https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:5895035d-26ee-4d04-99c8-ed5a11b94dad>
- Mark L. Huffman, "When the Blue Bus Crashes the Gate: The Problem with *People v. Collins* in the Probabilistic Evidence Debate," 46 U. MIAMI L. REV. 975 (1992), https://holis.harvard.edu/permalink/f/1mdq5o5/TN_cdi_proquest_journals_1291722854_read_sections_I,_II,_and_V
- Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, Rich Zemel, "Fairness Through Awareness" (2011), Sections 1 through 3.1, Section 6, and Appendix A. <https://acrobat.adobe.com/link/review?uri=urn%3Aaaid%3Acds%3AUS%3Ae529553d-d5d0-4a4c-ab34-1856ced03ce0>

Optional joint reading:

- Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings" (July 2016), <https://arxiv.org/abs/1607.06520>

September 10: Law Class: Fairness in Law

What are the contrasting conceptions of fairness at work in uses of probability and statistics? How do these contrasting views appear in discussions about the use of risk scores to inform decisions about access to credit, eligibility for pre-trial release without bail, and other legal matters (more detailed analyses of these problems to come in future classes)? How should calibration vs. balance be treated; what is fair and unfair about randomization; how do ex ante and ex post notions of fairness relate and differ? What does or could it mean for there be to probabilities for an individual or risk scores for an individual? Randomness: is fairness ex ante really fair?

Prompt: What do philosophical and legal conceptions of fairness offer to current debates over machine learning and use of probabilities in decision-making?

Required readings:

- Jon Elster, "Fairness and Norms," *Social Research*, Vol. 73, No. 2 (Summer 2006), pp. 365, <https://www.istor.org/stable/40971828>

- Cass R. Sunstein, “Two Conceptions of Procedural Fairness,” *Social Research*, Vol. 73, No. 2 (Summer 2006), pp. 619-646, <https://www.jstor.org/stable/40971840>

Optional readings:

- Charles Nesson, “The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts,” 98 HARV. L. REV. 1357 (1985) (excerpt), <https://www.jstor.org/stable/pdf/1340951.pdf>

September 11: CS Class: Fairness 1: Individual and group fairness; impossibility results

Week of September 16:

Joint Topic 3 - Privacy: What does and should it mean?

What does “privacy” mean with regard to data of individuals, and what technical and/or legal solutions could increase protections for data privacy? What tradeoffs might be involved? What is differential privacy? Is privacy in the context of consumer exchanges and social media a luxury people should individually pay for, or should government curb techniques of surveillance capitalism to protect individual privacy?

September 16: Joint Class: Prompt: Assuming privacy protections can be built, should the burden be on the individual to assert it (as in the right to be forgotten) or on the platforms/digital companies to develop and comply with pro-active protection requirements?

Required joint readings:

- Congressional Research Service, “U.S.-EU Trans-Atlantic Data Privacy Framework” (June 2022), <https://crsreports.congress.gov/product/pdf/IF/IF11613>
- Helen Nissenbaum, “Privacy as contextual integrity,” 79 Washington Law Review 119- (2004), <https://core.ac.uk/download/pdf/267979739.pdf>, p. 123 starting with “Before Proceeding” to p 138 up to section A, p. 155 Conclusion to the end.
- Shoshana Zuboff, “The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power,” pp. 55-61 (*PublicAffairs*, 2019) (right to be forgotten), https://hollis.harvard.edu/permalink/f/1lqd3jo/01HVD_ALMA512441641590003941
- Comparing U.S. State Data Privacy Laws vs. the EU’s GDPR, July 11, 2023, <https://pro.bloomberglaw.com/insights/privacy/privacy-laws-us-vs-eu-gdpr/>

Optional joint readings:

- Julie E. Cohen, “What Privacy Is For,” 126 HARV. L. REV. 1904 (2013); <https://harvardlawreview.org/print/vol-126/what-privacy-is-for/>

- Danielle Keats Citron, “Sexual privacy,” 128 *The Yale law journal* 1870–1960 (2019), <https://www.yalelawjournal.org/article/sexual-privacy> (browse introduction, table of contents and conclusion).
- European Commission, “Questions & Answers: EU-US Data Privacy Framework” (July 10, 2023). https://ec.europa.eu/commission/presscorner/detail/en/qanda_23_3752
- Cynthia Dwork and Deirdre Mulligan, It’s Not Privacy and It’s Not Fair, 66. *Stan. L. Rev.* (2013), <https://www.stanfordlawreview.org/online/privacy-and-big-data-its-not-privacy-and-its-not-fair/>

September 17: No Law Class

September 18: CS Class: Fairness 2: Individual Fairness: Definitions and Basic Results

Week of September 23: Joint Topic 4 – Privacy Tools

Legal jurisdictions are adopting laws and regulations specifying privacy protection for personal data; how do and how should differences in the emerging requirements be addressed, and how should corporations and other entities approach compliance?

September 23: Joint Class: Prompt: Might we interpret differential privacy as providing the right to be forgotten?

Required joint readings:

- Cynthia Dwork, “A Firm Foundation For Private Data Analysis,” *Communications of the ACM* 86-95”(Jan. 2011), pages 86-91. <https://cacm.acm.org/magazines/2011/1/103226-a-firm-foundation-for-private-data-analysis/fulltext> (download the pdf for proper display of equations)
- C. Dwork, C., Smith, A., Steinke, T. and Ullman, J., 2017. “Exposed! a survey of attacks on private data.” *AnnuStat. Appl*, 4(1), pp.61-84 (Dec. 2016), https://projects.iq.harvard.edu/files/privacytools/files/pdf_02.pdf, (Read Section 1.)
- Miriam Jordan, “New Findings Detail Trump Plan to Use Census for Partisan Gain”, *New York Times*, (July, 2022), <https://www.nytimes.com/2022/07/20/us/census-citizenship-question-oversight.html>

September 24: Law Class: Prompt: What are key elements of the European and US (national/states’) approaches to privacy, and, if you were advising the Massachusetts government, what should it adopt?

Required readings:

- Osman Husain, Global Privacy Laws in 2024 (Updated!), Enzuzo (April 11, 2024), <https://www.enzuzo.com/blog/data-privacy-laws>

- Matt Burgess, “How GDPR Is Failing,” *Wired* (May 23, 2022), <https://www.wired.com/story/gdpr-2022/>
- Alexandra Wood, et al. “Differential Privacy: A Primer for a Non-Technical Audience,” 21 *VAND. J. ENT. & TECH.* 209 (2018), https://dash.harvard.edu/bitstream/handle/1/38323292/4_Wood_Final.pdf?sequence=1&isAllowed=y

Optional reading:

- EU, General Data Protection Regulation, GDPR, Article 4, Definitions, and Chapter 3, articles 12-23, <https://gdpr-info.eu/art-4-gdpr/>, <https://gdpr-info.eu/art-12-gdpr/>, <https://gdpr-info.eu/art-13-gdpr/>, <https://gdpr-info.eu/art-14-gdpr/>, <https://gdpr-info.eu/art-15-gdpr/>, <https://gdpr-info.eu/art-18-gdpr/>, <https://gdpr-info.eu/art-19-gdpr/>, <https://gdpr-info.eu/art-20-gdpr/>, <https://gdpr-info.eu/art-21-gdpr/>, <https://gdpr-info.eu/art-22-gdpr/>, <https://gdpr-info.eu/art-23-gdpr/>
- Margot E. Kaminski, “Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability,” *S. CAL. L. REV.* 92:6 (2019); U. of Colorado Law Legal Studies Research Paper No. 19-9, pp. 31-79, <https://scholar.law.colorado.edu/faculty-articles/1265/>
- Byron Tau, “Selling Your Cellphone Location Data Might Soon Be Banned in U.S. for First Time,” *WSJ* (July 10, 2023). <https://www.wsj.com/articles/first-u-s-ban-on-sale-of-cellphone-location-data-might-be-coming-fbe47e53>
- Sihao Huang and Justin Curl, “Decoding China’s Ambitious Generative AI Regulations” (April 16, 2023). <https://freedom-to-tinker.com/2023/04/16/decoding-chinas-ambitious-generative-ai-regulations/>
- *Commonwealth v. Averyk Carrasquillo*, 489 Mass. 107, (September 8, 2021 - February 7, 2022), <http://masscases.com/cases/sjc/489/489mass107.html>

September 25: CS Class: Fairness 3: Multi-group fairness

Week of September 30:

Joint Topic 5 – Privacy: Theories and Contemporary Issues for Patients, Individual Residents, and Consumers and First Joint Project

September 30: Joint Class: We will spend some time in class in small group discussions tackling proposed regulation of LLMs in the health-care context. We will introduce tools for protecting privacy in various applications, including secure communication, safe storage of data, and data analysis.

In class, we will discuss the privacy issues related to large language models as used in health care and in small groups combining law students and computer science students, each group will look at two or three of the recommendations of Meskó and Eric. J. Topol (first article below), offer criticisms and alternative recommendations in no more than 4 paragraphs. **By 6 p.m. on Sept. 30**, each group should post to the class website a summary of their recommendations.

Required readings:

- Bertalan Meskó and Eric. J. Topol, The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digit. Med.* 6, 120 (2023). <https://www.nature.com/articles/s41746-023-00873-0> **This will be the prompt for joint exercise #1 so be sure to read in advance of class. See assignment in the paragraph above.**
- Barbara J. Evans, "Big Data and Individual Autonomy in a Crowd," (2018); <https://www.cambridge.org/core/books/big-data-health-law-and-bioethics/big-data-and-individual-autonomy-in-a-crowd/D1C166C853F32315CEF7D9621095A15B>
- National Conference of State Legislatures, "Differential Privacy for the Census Explained," (2021), <https://www.ncsl.org/technology-and-communication/differential-privacy-for-census-data-explained#:~:text=Differential%20privacy%20will%20mean%20that,used%20to%20protect%20small%20populations.>
- Lindell, Yehuda. "Secure multiparty computation." *Communications of the ACM* 64, no. 1 (2020): 86-96, (read first 3 pages). [Secure multiparty computation \(acm.org\)](https://dl.acm.org/doi/10.1145/3365553)
- Review: Cynthia Dwork, "A Firm Foundation For Private Data Analysis," *Communications of the ACM* 86-95" (Jan. 2011), pages 86-91. <https://cacm.acm.org/magazines/2011/1/103226-a-firm-foundation-for-private-data-analysis/fulltext> (download the pdf for proper display of equations)
- Ryan Yackel, "What is Homomorphic Encryption?," (July 6, 2021), <https://www.keyfactor.com/blog/what-is-homomorphic-encryption/>
- [Browse the Wikipedia entry on homomorphic encryption](https://en.wikipedia.org/wiki/Homomorphic_encryption)

October 1: Law Class: Developments in privacy laws

Prompt: Identify what is helpful and what is inadequate in these assessments of privacy risks and health apps:

Required readings:

- Robbert Goddard, Privacy at Risk with Health and Wellness Apps, <https://www.ispartnersllc.com/blog/data-privacy-health-apps/>
- Aditya Singhal, Nikita Neveditsin, Hasnaat Tanveer, and Vijay Mago, Toward Fairness, Accountability, Transparency, and Ethics in AI for Social Media and Health Care: Scoping Review. *JMIR Med Infom.* 2024 Apr. 3; <https://medinform.jmir.org/2024/1/e50048/>
- California Privacy Act: Updated, <https://oag.ca.gov/privacy/ccpa>
- Alan Friel, Julia Jacobson, Kyle Fath, Kyle Dull, Gicel Tomimbang, & Sasha Kiosse, Health (and Health-ish) Data and Advertising Under Scrutiny, *Privacy World* (June 20, 2023), <https://www.privacyworld.blog/2023/06/health-and-health-ish-data-and-advertising-under-scrutiny/>
- Which States Have Consumer Data Privacy Laws?, Bloomberg (March 18, 2024), <https://pro.bloomberglaw.com/insights/privacy/state-privacy-legislation-tracker/>
- Ariel Latzer, Complying with New and Existing Biometric Data Privacy Laws, 16 *J. Bus.*

Entrepreneurship & L. 201 (2023),

<https://digitalcommons.pepperdine.edu/jbel/vol16/iss1/7/>

- Ian Cohen, The Rising Stakes of Healthcare Data Privacy in 2024, MedCity News (May 2024), <https://medcitynews.com/2024/05/the-rising-stakes-of-healthcare-data-privacy-in-2024-the-need-for-practical-guidance/>

Optional readings:

- Willaim Betten, Medical Data Privacy in a Connected World, MDDI (Jan. 17, 2024), <https://www.mddionline.com/software/medical-data-privacy-in-a-connected-world>
- Marta Ziosi, David Watson, and Luciano Floridi, A Genealogical Approach to Algorithmic Bias, https://www.researchgate.net/publication/380292889_A_Genealogical_Approach_to_Algorithmic_Bias
- FTC Business Blog, Updated FTC Health Breach Notification Rule (April 2024), <https://www.ftc.gov/business-guidance/blog/2024/04/updated-ftc-health-breach-notification-rule-puts-new-provisions-place-protect-users-health-apps>
- European Union Agency for Fundamental Rights, GDPR in practice – Experiences of data protection authorities (June 11, 2024), <https://fra.europa.eu/en/publication/2024/gdpr-experiences-data-protection-authorities>

October 2: CS Class: Privacy 1: Reconstruction theorem; intro to DP; properties; Laplace Mechanism

Heads up: Make plans for Joint Project #2:

Joint project #2, due 6 pm on November 4:

Meet in cross-disciplinary groups to address the key elements of an algorithmic hiring tool if turned on the interviewers to determine potential bias along any demographic dimension as perceived by the camera during the first minute of the interview. What are the technical and legal considerations in using such a system to evaluate discrimination on the basis of protected traits in hiring? Summarize your analysis in a few paragraphs.

Week of October 7:

Joint Topic 6 - Fairness as Transparency: Accountability, Auditing

October 7: Joint Class: Does a requirement of transparency or explanation in the use of

algorithms in decision-making promote fairness? How would it work and what would be limitations? What are complex/simple model practices? Come to class prepared to thoughtfully discuss the following questions:

1. For what properties is the system to be audited? The denominator problem. Test for differentially expressive data (as discussed in Eubanks)?
2. What is an explanation in administrative procedures? What is an explanation in computer science (work in progress; may not be ready to discuss)?
3. Does “useful and complete” auditing imply that the auditor knows a (proto-)metric?
4. When people call for an explanation, what are they really seeking?
5. What does it mean to understanding something, and how can we check that understanding has been achieved?

Required joint readings:

- Zachary C. Lipton, “The Mythos of Model Interpretability,” 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016, New York, NY) (Mar. 6, 2017), <https://arxiv.org/abs/1606.03490>
- Andrew D. Selbts and Solon Barocas, “The Intuitive Appeal of Explainable Machines,” 87 *Fordham L. Rev.* 1085 (2018), <https://ir.lawnet.fordham.edu/flr/vol87/iss3/11>
- Ben Miller, “Making AI Work for Government: It All Comes Down to Trust,” GovTech, <https://www.govtech.com/opinion/making-ai-work-for-government-it-all-comes-down-to-trust>.
- Wolfgang Blau, et al., Protecting Scientific Integrity in an Age of Generative AI, PNAS (May 21, 2024), <https://www.pnas.org/doi/10.1073/pnas.2407886121>

Optional joint readings:

- Alfred Ng, “Can Auditing Eliminate Bias from Algorithms?” *The Mark-up* (Feb.23, 2021), <https://themarkup.org/the-breakdown/2021/02/23/can-auditing-eliminate-bias-from-algorithms>
- Cliff Kuang, “Can A.I. Be Taught to Explain Itself?,” *N.Y. Times Magazine* (Nov. 21, 2017), <https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html>
- Airlie Hilliard, “Bias Audit Laws in the US: The State of Play for Automated Employment Decision Tools,” *Holistic AI*, (Jan. 15, 2024), <https://www.holisticai.com/blog/automated-employment-decision-tool-bias-audit-laws>.

- Roland Neil, and Christopher Winship, Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing, 2 Ann. Rev. Criminology (Jan. 2019), <https://doi.org/10.1146/annurev-criminol-011518-024731>
- Alex Kalai and Santosh S. Vempala, Calibrated language models must hallucinate. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (pp. 160-171(June 2024)), [2311.14648 \(arxiv.org\)](https://arxiv.org/abs/2311.14648)

October 8: Law Class: Prompt: Compare Facebook/Meta meeting calls for public accountability—with the White House Blueprint for an AI Bill of Rights: What are strengths and weaknesses of each?

Required readings:

- Caroline Mimbs Nyce, “The New ‘Transparent’ Twitter Isn’t Very Transparent” The Atlantic, (April 5, 2023) <https://www.theatlantic.com/technology/archive/2023/04/elon-musk-twitter-algorithm-code-public/673642/>
- Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, Moritz Hardt, Algorithmic amplification of politics on Twitter, (Jan. 4, 2022), <https://pubmed.ncbi.nlm.nih.gov/34934011/>
- [should-do-more-on-voter](#)
- “FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence,” (Oct. 2023), <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>.
- Alistair Knott, Dino Pedreschi, “Human, or human-like? Transparency for AI-generated content,” OECD.AI, <https://oecd.ai/en/wonk/human-or-human-like-transparency-for-ai-generated-content>.

Optional readings:

- Christiano Lima, “Biden’s Former Tech Advisor on What Washington Is Missing About AI,” Washington Post (May 30, 2023). <https://www.washingtonpost.com/politics/2023/05/30/biden-former-tech-adviser-what-washington-is-missing-about-ai/>
- Review: Andrew D. Selbst and Solon Barocas, “The Intuitive Appeal of Explainable Machines,” 87 *Fordham L. Rev.* 1985 (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3126971
- Richard H. Fallon, Jr., “A Theory of Judicial Candor,” 117 COLUM. L. REV. 2265 (2017), <https://columbialawreview.org/content/a-theory-of-judicial-candor/>
- Jiaming Ji, et al., AI Alignment: A Comprehensive Survey, <https://arxiv.org/abs/2310.19852#:~:text=AI%20alignment%20aims%20to%20make>,

[methodology%2C%20and%20practice%20of%20alignment.](#)

- Yuntao Bai et al., Anthropic’s Constitutional AI: Constitutional AI: Harmlessness from AI Feedback, (2022) <https://arxiv.org/abs/2212.08073>
- Tyna Eloundou and Teddy Lee, OpenAI’s democratic opinion polling: “Democratic inputs to AI grant program: lessons learned and implementation plans,” OpenAI, 2024, <https://openai.com/blog/democratic-inputs-to-ai-grant-program-update>

October 9: CS Class: Privacy 2: DP Primitives: High-Dimensional Laplace, Report Noisy ArgMax, Exponential Mechanism

Week of October 14:

Oct. 14: No Joint Class

Oct. 15: No Law Class

October 16: CS Class: Sparse Vector, Approximate Differential Privacy, and Advanced Composition

Week of October 21:

Joint Topic 7 - Differential Privacy and the Census: Technical, Legal, and Political questions

October 21: Joint Class: Differential Privacy and US Census Bureau Litigation

Prompt: What are the promises and limits (technically, legally, functionally) of using differential privacy in the U.S. Census?

1. What does the law require in terms of privacy protections in the Census?
Semantic privacy vs “tracing back”.
2. How does differential privacy interact with the constitutional uses of census data, specifically, electoral representation? How about redistricting? What role does citizenship information play in redistricting, given that it is computed until now based on sampling?
3. Census practice under President Trump allowed citizenship to be inferred based on administrative data paired with census data. How the bureau will handle this, and the impact on the privacy budget?
4. How should the Census address issues posed by very small populations (such as Indian tribes) and differential privacy?
5. How might different techniques for protecting privacy differently affect electoral districting?

Required joint readings:

- Priyanka Nanayakkara and Jessica Hullman, “States Are Suing the Census Bureau Over Its

Attempts to Make Data More Private,” *Slate* (Aug. 12, 2021),

<https://slate.com/technology/2021/08/census-bureau-differential-privacy-lawsuit.html>

- Danah Boyd and Jayshree Sarathy, “Differential Perspectives, Epistemic Disconnects Surrounding the US Census Bureau’s use of Differential Privacy,” *2 Harvard Data Science Review*, (June 24, 2022), <https://hdsr.mitpress.mit.edu/pub/3vj5j6i0>
- National Strategy to Advance Privacy-Preserving Data Sharing and Analytics (March 2023). <https://acrobat.adobe.com/link/review?uri=urn%3Aaaid%3Asc%3AUS%3A7dd56ab2-6afc-41da-9608-e61db4bd1003>
- Mark Bun, et al., Statistical Inference is Not a Privacy Violation, <https://differentialprivacy.org/inference-is-not-a-privacy-violation/> (June 3, 2021).

Optional joint reading:

- *State of Alabama v. U.S. Census Department*, M.D.Ala. (June 29, 2021), <https://www.brennancenter.org/our-work/court-cases/alabama-v-us-dept-commerce>
- V. J. Hotz et al., Balancing data privacy and usability in the federal statistical system. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2104906119 (2022), <https://www.pnas.org/doi/full/10.1073/pnas.2104906119>
- Jarmin, R.S., Abowd, J.M., Ashmead, R., Cumings-Menon, R., Goldschlag, N., Hawes, M.B., Keller, S.A., Kifer, D., Leclerc, P., Reiter, J.P. and Rodríguez, R.A., 2023. An in-depth examination of requirements for disclosure risk assessment. *Proceedings of the National Academy of Sciences*, 120(43), p.e2220558120, <https://pubmed.ncbi.nlm.nih.gov/37831744/>

October 22: Law class: Prompt: Should the legal parameters around the census change in light of new technologies and new needs? What actions should the Census bureau, Congress, or federal or state consumer law take in light of recent assessments?

Required readings:

- U.S. Census Privacy Policy, <https://www.census.gov/about/policies/privacy/privacy-policy.html>
- U.S. Census: Differential Privacy, <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html>
- Kenneth A. Bamberger, et. al., “Can you Pay for Privacy? Consumer Expectations and the Behavior of Free and Paid Apps” (2020). <https://acrobat.adobe.com/link/review?uri=urn%3Aaaid%3Asc%3AUS%3A17884bbc-5624-4d5a-9066-cd1b912bf8da>

Optional reading:

- A History of U.S. Census Privacy Protections, <https://www.census.gov/library/visualizations/2019/comm/history-privacy-protection.html>
- JASON/Mitre, “Report on Consistency of Data Products and Privacy Methods for the 2020

Census,

https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/plan/memo-series/2020-memo-2022_03.html (full report)

October 23: CS Class: Cryptography 1: one-way functions and hardcore bits

Week of October 28:

Joint Topic 8 - Fairness Again: Prediction

Guest Speaker: Charlotte Burrows: Chair of the U.S. Equal Employment Opportunity Commission

Required joint readings:

- Jon Kleinberg et al., “Human Decisions and Machine Predictions,” *Quarterly Journal of Economics* (2018), pp. 237-293, <https://academic.oup.com/qje/article/133/1/237/4095198>
- Philip Dawid, “On Individual Risk,” *Synthese* 194 (9):3445-3474 (2017), <https://arxiv.org/abs/1406.5540>
- Christopher Thomas, Antonio Pontón-Núñez, “Automating Judicial Discretion: How Algorithmic Risk Assessments in Pretrial Adjudications Violate Equal Protection Rights on the Basis of Race,” 40(2) *Law & Ineq.* 371 (2022), DOI: <https://doi.org/10.24926/25730037.649>.

Optional joint readings:

- Julia Angwin, et al., “Machine Bias,” *Pro Publica* (May 22, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Alexandra Chouldechova, [Fair prediction with disparate impact: A study of bias in recidivism prediction instruments](#). *Big data*, 5(2), pp.153-163, 92017), [[1703.00056](#)] [Fair prediction with disparate impact: A study of bias in recidivism prediction instruments \(arxiv.org\)](#) (Sections 1 and 2. Be sure to read and understand Equation 2.6 on Page 7.)
- *State of Wisconsin v. Loomis*, 881 NW 2d 749 (2016) <https://harvardlawreview.org/2017/03/state-v-loomis/>
- Jon Kleinberg et al., “Discrimination in the Age of Algorithms,” *Journal of Legal Analysis* (Apr. 22, 2019), <https://academic.oup.com/jla/article/doi/10.1093/jla/laz001/5476086>
- Benjamin van Heffen, Dennis Herhausen, and Tobias Fause, “The pitfalls and perils of

algorithms: A classification of machine learning biases and mitigation methods,” 144 J. of Bus. Res. 93 (May 2022),

<https://www.sciencedirect.com/science/article/pii/S0148296322000881>

- Cass Sunstein. Governing by Algorithm? No noise and less bias. (2022). <https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=4114&context=dlj>
- Sandra G. Mayson, “Bias in, bias out,” 128 *The Yale law Journal* 2218–2300 (2019), <https://www.yalelawjournal.org/article/bias-in-bias-out>, p. 2218-2226; 2253-2281.

October 29: Law Class: Prompt: What legal rules can and should the jurisdiction of your choice adopt to govern predictive policing, risk scores, and other AI/ML tools when used by governments and other powerful actors?

Required readings:

- Adam Williams, “What Could Horizontal AI Legislation Look Like In the US? Exploring the US Algorithmic Accountability Act,” *Holistic AI* (Jan. 9, 2023), <https://www.holisticai.com/blog/us-algorithmic-accountability-act>.
- Renata M. O’Donnell, “Challenging Racist Predictive Policing Algorithms Under the Equal Protection Clause,” 94 N.Y.U. L. REV. 544 (2019), <https://www.nyulawreview.org/issues/volume-94-number-3/challenging-racist-predictive-policing-algorithms-under-the-equal-protection-clause/>
- Dhruv Mehrota, Surya Mattu, Annie Gilbertson, and Aaron Sankin, “How We Determined Predictive Policing Software Disproportionately Targeted Low-Income, Black, and Latino Neighborhoods,” *Gizmodo* (Dec. 2, 2021), <https://gizmodo.com/how-we-determined-predictive-policing-software-dispropo-1848139456>
- Katie Johnston, “Can AI Help Employers Screen for Honesty?”, *Boston Globe* (May 21, 2023). <https://www.bostonglobe.com/2023/05/21/business/can-ai-help-employers-screen-honesty/>

Optional readings:

- Andrew Guthrie Ferguson, “Big Data and Predictive Reasonable Suspicion,” 163 U. PENN. L. REV. 327 (2015), https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?httpsredir=1&article=9464&context=penn_law_review
- Simone Wallk, “Shotspotter and the Expanding Scope of Reasonable Suspicion Under the Fourth Amendment” HLS Fairness and Privacy paper, (2023).
- Betsy Anne Williams et al., “How Algorithms Discriminate Based on Data They Lack,” *Journal of Information Policy* 8, pp. 78-115 (2018) <https://pdfs.semanticscholar.org/8dbc/2a67f0142104c82c32e1ebced118c8d69937.pdf>
- Letter and Statement re: Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns (July 8, 2019) <https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns>
- Anthony W. Flores et al., “False Positives, False Negatives, and False Analyses: A Rejoinder to ‘Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks,’” 80:2 *Federal Probation* (Sept. 2016)
- https://hollis.harvard.edu/permalink/f/1mdq5o5/TN_cdi_proquest_reports_1918592458

- Ifeoma Ajunwa, *The Quantified Worker*, pp. 75-101 (2023).
<https://acrobat.adobe.com/link/review?uri=urn%3Aaaid%3Aacds%3AUS%3Ad117dc4c-c501-45d5-a64f-80891c28db22>
- Victoria Angelou, Will Dobbie, and Crystal S. Yang, “Algorithmic Recommendations and Human Discretion” (draft shared with permission of authors)
<https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:3d3dba42-1955-4aac-a65d-6e74134fd6c2>

October 30: CS Class: Cryptography 2: Pseudo-randomness from one-way permutations

Week of November 4:

Joint Topic 9 - Fairness and Equal Protection, Disparate Impact, and Other Legal Frameworks; Intention vs. Effects

Prompt: May programmers use racial and gender classifications to test whether their algorithms that do not use those classifications to optimize employment decisions have disparate racial or gender effects?

HEADS UP: Law students should be prepared to explain disparate impact analysis and **computer science students** should be ready to discuss thoughtfully whether programmers may use racial and gender classifiers to test whether their algorithms that do not use those classifications to optimize employment decisions have disparate racial or gender effects.

Reminder: Joint project #2: due 6 pm on Nov. 4:

Meet in cross-disciplinary groups to address the key elements of an algorithmic hiring tool if turned on the interviewers to determine potential bias along any demographic dimension as perceived by the camera during the first minute of the interview. What are the technical and legal considerations in using such a system to evaluate discrimination on the basis of protected traits in hiring? Summarize your analysis in a few paragraphs.

Required joint readings:

- Kevin Tobia, “Note: Disparate Statistics,” *126 Yale L.J.* 2382 (2017),
<https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=5815&context=yli>
- Emily Black, John Logan Koepke, Paulin Kim, Solon Barocas, and Mingwei Hsu, *Less Discriminatory Algorithms* (October 2, 2023). 113 *Georgetown Law Journal*, No. 1, 2024, Washington University in St. Louis Legal Studies Research Paper Forthcoming, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4590481

or https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4590481

Optional joint readings:

- Mark MacCarthy, “Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms,” 48 CUMBER. L. REV. 1 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3154788 [GET WHOLE PAPER]
- Paige Smith, “Machine Learning Deployed to Help EEOC Predict Discrimination,” *Bloomberg Law* (Dec. 26, 2018), <https://news.bloomberglaw.com/daily-labor-report/machine-learning-deployed-to-help-eeoc-predict-discrimination>
- Martha Minow, “Equality, Equity, And Algorithms: Learning From Justice Rosalie,” 73 U. Toronto L.J. 163 (2023), <https://utpjournals.press/doi/10.3138/utlj-2023-0064>
- Ian Ayres and Jack M. Balkin, *The Law of AI is the Law of Risky Agents without Intentions* (forthcoming 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4862025#:~:text=In%20this%20sense%20AI%20programs,different%20parts%20of%20the%20law

November 5: No Law Class: Election Day

November 6: CS Class: Cryptography 3: Encryption

Week of November 11:

Joint Topic 10 – Correctness and Accountability

November 11: Joint Class: Prompt: How do scientific and legal experts assess the current uses of facial recognition? What do you recommend as policy, especially in the light of historical issues with “scientific evidence” relied upon in courts and public policies?

Required joint readings:

- Written Testimony of Professor Andrew Guthrie Ferguson Before the House of Representatives Committee on Oversight and Reform Hearing On: Facial Recognition Technology: (Part 1) Its Impact on our Civil Rights and Liberties May 22, 2019, <https://docs.house.gov/meetings/GO/GO00/20190522/109521/HHRG-116-GO00-Wstate-FergusonA-20190522.pdf>
- Andrew Gelman and Eric Loken, “The garden of forking paths: Why multiple comparisons can be a problem, even when there is no ‘fishing expedition’ or ‘p-hacking’ and the research hypothesis was posited ahead of time,” (2013), http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf
- Ella Jakubowska, “The EU wants to make facial recognition history — but it must be done for the right reasons,” *EuroNews*, (Nov. 16, 2023), <https://www.euronews.com/2023/11/16/the-eu-wants-to-make-facial-recognition->

[history-but-it-must-be-done-for-the-right-reasons#:~:text=In%20June%202023%2C%20the%20European,us%20all%20as%20walking%20barcodes](#)

Optional joint readings:

- Paresh Dave, “US Cities are Backing Off of Banning Facial Recognition as Crime Rises,” *Reuters* (May 12, 2022), <https://www.reuters.com/world/us/us-cities-are-backing-off-banning-facial-recognition-crime-rises-2022-05-12/>
- Clare Garvie, “Garbage In. Garbage Out. Face Recognition on Flawed Data,” *Georgetown Law Center of Privacy and Technology* (May, 2019) <https://www.flawedfacedata.com>.
- Jason R. Bent, “Is Algorithmic Affirmative Action Legal?” *GEO. L.J.* 108, (Apr. 16, 2019), <https://ssrn.com/abstract=3372690>
- Tate Ryan-Mosley, “The Movement to Limit Face Recognition Tech Might Finally Get A Win,” *MIT Technology Review* (July 20, 2023), <https://www.technologyreview.com/2023/07/20/1076539/face-recognition-massachusetts-test-police/>
- Urs Gasser and Carolyn Schmitt, “The Role of Professional Norms in the Governance of Artificial Intelligence”, in Markus D. Dubber, et. al. (eds), *The Oxford Handbook of Ethics of AI*, Oxford University Press pp. 141-156 (2020). <https://doi.org/10.1093/oxfordhb/9780190067397.013.8>
- Review previously assigned paper: [Roland Neil](#), and [Christopher Winship](#), Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing, 2 *Ann. Rev. Criminology* (Jan. 2019), <https://doi.org/10.1146/annurev-criminol-011518-024731>

November 12: Law Class: Prompt: Make the case for specific uses of machine learning, AI, or other digital tools used by legal actors in fields such as human rights, health care, education, or contracts—with modifications/regulatory restrictions as you think best.

Required readings:

- I. Glenn Cohen and Harry S. Graver, “Cops, Docs, and Code: A Dialogue Between Big Data in Health Care and Predictive Policing,” 51 *U. CAL. (DAVIS) L. Rev.* 437 (2017) https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Cohen_Graver.pdf
- David B. Black, “Blockchain Smart Contracts Aren’t Smart and Aren’t Contracts,” *Forbes* (Feb. 4, 2019), <https://www.forbes.com/sites/davidblack/2019/02/04/blockchain-smart-contracts-arent-smart-and-arent-contracts/>
- United Nations High Commissioner for Human Rights, *Humanity Should Get the Best, Not the Worst –UN Disability Rights Experts* (May 2022), <https://www.ohchr.org/en/stories/2022/05/humanity-should-get-best-ai-not-worst-un-disability-rights-expert>
- Khari Johnson, “AI Could Change How Blind People See the World,” *Wired* (July 5, 2023), <https://www.wired.com/story/ai-gpt4-could-change-how-blind-people-see-the-world/>
- Orly Lobel, “The Law of AI for Good” (Jan. 2023, draft), available following the abstract

at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4338862

Optional readings:

- Lorna McGregor, Daragh Murray, and Vivian Ng, “International Human Rights Law as a Framework for Algorithmic Accountability,” 68:2 INT’L AND COMP. L.Q. 309 (April 2019), <https://www.cambridge.org/core/journals/international-and-comparative-law-quarterly/article/international-human-rights-law-as-a-framework-for-algorithmic-accountability/1D6D0A456B36BA7512A6AFF17F16E9B6>
- Christine Chambers Goodman, “Just Aled: An Essay on Just Applications of Artificial Intelligence in Education”, West Virginia L. Rev., Vol. 123 (2021). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3882873
- Giovana Carneiro, Regulatory Sandboxes for Artificial Intelligence Regulation (draft 2024), <https://acrobat.adobe.com/link/track?uri=urn:aaid:scds:US:cd290b40-48c2-4759-9ab3-0c92f7bed26f>

November 13: CS Class: Advanced Topic in Differential Privacy: Validity of Adaptive Data MAnalysis (law students might want to hear this lecture which will include some responses to problems raised by Gelman and Loken)

HEADS UP: Before class on **November 18**, students should watch:

- Video talk by Dorothy Roberts (starting approximately at minute 1:31:54): [Morning Session: Genomics, Medicine, Epidemiology and Data Interpretation | Simons Institute for the Theory of Computing \(berkeley.edu\)](https://simons.berkeley.edu/talks/morning-session-genomics-medicine-epidemiology-data-interpretation)
- Video talk by Jay Kaufman (starting approximately at minute 49:15): <https://simons.berkeley.edu/talks/morning-session-genomics-medicine-epidemiology-data-interpretation>

Week of November 18:

Joint Topic 11 - Arguments Over the Proper Uses and Limits on Uses of Algorithms in Medicine

November 18: Joint class: Prompt: What does the historical treatment of race in medicine and in employment offer as guides and cautions for current policies and policy discussion?

Possible Guest speaker: Dorothy Roberts, [George A. Weiss University Professor of Law and Sociology and the Raymond Pace and Sadie Tanner Mossell Alexander Professor of Civil Rights](#)

Dorothy Roberts is the George A. Weiss University Professor of Law & Sociology at the University of Pennsylvania, with joint appointments in the Departments of Africana Studies and Sociology and the Law School, where she is the inaugural Raymond Pace and Sadie Tanner Mossell Alexander Professor of Civil Rights. She is also founding director of the Penn Program on Race, Science & Society.

Her path breaking work in law and public policy focuses on urgent social justice issues in reproductive justice, policing, family regulation, science, medicine, and bioethics. Her major books include *Killing the Black Body: Race, Reproduction, and the Meaning of Liberty* (Pantheon, 1997); *Shattered Bonds: The Color of Child Welfare* (Basic Books, 2001); *Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-First Century* (The New Press, 2011); and *Torn Apart: How the Child Welfare System Destroys Black Families—and How Abolition Can Build a Safer World* (Basic Books, 2022), as well as more than 100 articles and essays in books and journals.

Required viewing:

- Video talk by Dorothy Roberts (starting approximately at minute 1:31:54): [Morning Session: Genomics, Medicine, Epidemiology and Data Interpretation | Simons Institute for the Theory of Computing \(berkeley.edu\)](https://simons.berkeley.edu/talks/morning-session-genomics-medicine-epidemiology-data-interpretation)
- Video talk by Jay Kaufman (starting approximately at minute 49:15): <https://simons.berkeley.edu/talks/morning-session-genomics-medicine-epidemiology-data-interpretation>

Required joint readings:

- Lundy Braun and Evelyn Hammonds, “The Dilemma of Classification: The Past in the Present,” in Wailoo, Nelson, et al., editors, *Genetics and the Unsettled Past* (Rutgers University Press, 2011),
https://hollis.harvard.edu/permalink/f/1mdq5o5/TN_cdi_jstor_books_j_ctt5hj79f_9
- Duana Fullwiley, 2008. The Biological Construction of Race: Admixture Technology and the New Genetic Medicine, 38 *Social studies of science*, 38(5), pp.695-735 (2008),
<https://www.jstor.org/stable/25474605>

Optional joint readings:

- Duana Fullwiley, “The ‘Contemporary Synthesis’: When Politically Inclusive Genomic Science Relies on Biological Notions of Race,” *Isis*, 105:4 (Dec. 2014), pp. 803-814,
<https://www.jstor.org/stable/10.1086/679427>
- Dorothy Roberts, “The Problem With Race-Based Medicine: American journal of Physical Medicine & Rehabilitation,” *American Journal of Physical Medicine & Rehabilitation* (July 2023),
https://journals.lww.com/ajpmr/Fulltext/2023/07000/The_Problem_With_Race_Based_Medicine.1.aspx
- Lily Hu, Issa Kohler-Hausman, “What’s Sex Got To DO With Fair Machine Learning?” (June 4, 2020), <https://arxiv.org/abs/2006.01770>
- Braun, Fausto-Sterling, Fullwiley, Hammonds, Nelson, Quivers, Reverby, and Shields, “Racial Categories in Medical Practice: How Useful Are They?” in *PLOS Medicine* 4, no. 9 (2007),
<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0040271>
- L. . Ebony Boulware, et al., “Race-Free Estimation of Kidney Function: Clearing the Path Toward Kidney Health Equity,” *JAMA Network* (June 6, 2022).
https://jamanetwork.com/journals/jama/fullarticle/2793291?guestAccessKey=7c9bf9ae-98cc-404b-9365-c52a8596f6e6&utm_source=silverchair&utm_medium=email&utm_campaign=article_alert-jama&utm_content=olf&utm_term=060622
- Yuval Barack-Corren et al, Validation of an Electronic Health Record–Based Suicide Risk Prediction Modeling Approach Across Multiple Health Care Systems, (2020).
<https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2763237>

November 19: Law Class: Mitigating Risks

Prompt: How should public and private organizations—health care and others-- identify and manage risks from use of AI tools, including Generative AI? What actors should be subject to regulation and liability?

Required readings:

- Paul Ohm, “How to Regulate Harmful Inferences,” *Jotwell* (Dec. 22, 2021) <https://cyber.jotwell.com/how-to-regulate-harmful-inferences/>

November 20: CS Class: Advanced Topic in Algorithmic Fairness (causality)
(Law students especially invited)

Week of November 25: Joint Topic 12 - Algorithms for Democracy, Governing, and Access to Law

November 25: Joint Class: Prompt: How can algorithms and other tools damage or improve democracy, governing, justice, and access to law?

Required joint readings:

- Colleen V. Chien, Miriam Kim, Akhil Raj, and Rohit Rathish, How Generative AI Can Help Address the Access to Justice Gap Through the Courts (January 4, 2024). *Loyola of Los Angeles Law Review*, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=4683309>
- Stephanie Kelley, Anton Ovchinnikov, Adrienne Heinrich, and David R. Hardoon, “Removing Demographic Data Can Make AI Discrimination Worse,” (2023), <https://hbr.org/2023/03/removing-demographic-data-can-make-ai-discrimination-worse>

Optional readings:

- Carrie Arnold, “The Mathematicians Who Want to Save Democracy,” *Scientific American*, <https://www.nature.com/articles/546200a>
- Laura Barrón-López, Shrai Popat, “How AI-generated misinformation threatens election integrity,” *PBS NewsHour*, (Feb. 13, 2024), <https://www.pbs.org/newshour/show/how-ai-generated-misinformation-threatens-election-integrity>
- Ngozi Okidegbe, “The Democratizing Potential Of Algorithms?,” (2020), (53 *Connecticut Law Review* 739 (focus on pp. 35-47). <https://papers.ssrn.com/abstract=3835370> or https://scholarship.law.bu.edu/faculty_scholarship/3138/
- Mikaela Meyer et al., *Flipping the Script on Criminal Justice Risk Assessment: An actuarial model for assessing the risk the federal sentencing system poses to defendants*, (2022), <http://arxiv.org/abs/2205.13505> (read abstract and introduction).

November 26: Law Class : Prompt: What should the Federal Trade Commission, Department of Justice, or Congress consider as priorities for AI regulation, including for transparency, and interpretability? What could serve as a remedy for deception if the government requires auditing and accountability with regard to social media policies and practices?

Required Readings:

- Nicol Turner Lee, Jack Malamud, “How Congress can secure Biden’s AI legacy,” Brookings Institute, (Jan. 25, 2024), <https://www.brookings.edu/articles/how-congress-can-secure-bidens-ai-legacy/>
- Sonia K. Katyal, “Democracy & Distrust in an Era of Artificial Intelligence,” *Daedalus* (Spring 2022), <https://direct.mit.edu/daed/article/151/2/322/110619/Democracy-and-Distrust-in-an-Era-of-Artificial>

Optional Readings:

- Pauline Kim, Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action, (California Law Review, Forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4018414
- Jia Qing Yap and Ernest Lim, A Legal Framework for Artificial Intelligence Fairness Reporting, Cambridge Law Journal (June 2022). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4128641
- Evelyn Douek, “From Posts as Trumps to Proportionality and Probability,” *121 Colum. L.Rev.* (2020), <https://columbialawreview.org/content/governing-online-speech-from-posts-as-trumps-to-proportionality-and-probability/>
- Michelle Miller, Worker Power and Voice in the AI Response, (2024). <https://clje.law.harvard.edu/app/uploads/2024/01/Worker-Power-and-the-Voice-in-the-AI-Response-Report.pdf>
- Mark Weber, Mikhail Yurochkin, Sherif Botros, Vanio Markov, “Black Loans Matter: Fighting Bias for AI Fairness in Lending,” (2020). <https://mitibmwatsonailab.mit.edu/research/blog/black-loans-matter-fighting-bias-for-ai-fairness-in-lending/>
- Matthew Dahl, et al., Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models, *J. Legal Analysis* (2024), <https://academic.oup.com/jla/article/16/1/64/7699227>

Nov. 27: No CS Class, Thanksgiving Break

Week of December 1

Joint Topic 13 – Responsible Computing

December 1: Joint Class: Prompt: How might you define “responsible scaling” for the moderating of political speech on social media? Prediction algorithms are trained on data which may be inadequate for the task or the class of networks on which the training is carried might be insufficiently expressive to capture reality (or both). Discuss when it is and is not appropriate to make do with what is given or to require development of what is not yet available.

Joint required readings:

- Cynthia Dwork, Martha Minow; “Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law,” *Daedalus* 2022; 151 (2): 309–321.
doi: https://doi.org/10.1162/daed_a_01918
- Susan Silbey, Rotten Apples or a Rotting Barrel, MIT,
<https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:52c3c4a4-ac38-41a5-96ce-e3fd23a76318>
- Kate Crawford, Atlas of AI (excerpts), AI (2021), pp 32 – 36, starting from “The Mineralogical Layer,” p. 43-46 “The Myth of Clean Tech,” p.63-69 “Potemkin AI and the Mechanical Turks,” p. 89-95 “Data,” p. 136-149 “Training Sets as Classification Engines: The Case of ImageNt,”
<https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:2f07d2ac-5bc3-45ca-9581-1e7c3d02e92c>

Optional joint readings:

- Dennis D. Hirsch, [Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1021623), 41 Georgia L. Rev. 1 (2006),
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1021623
- Andries Nentjes, Frans P de Vries, and Doede Wiersma, “Technology-forcing through environmental regulation,” 23 *Eur. J. Polit. Econ.* 903 (Dec. 2007),
<https://www.sciencedirect.com/science/article/pii/S0176268007000286>
- Urs Gasser and Carolyn Schmidt, “the Role of Professional Norms in the Governance of Artificial Intelligence,” in Markus D. Dubber, Frank Pasquale, and Sunit Das, eds., *The Oxford Handbook of Ethics in AI* (2020),
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3378267

- Colin Doyle, “How Algorithms Expose the Law,” (2021).
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3894617
- Joseph Gordon Levitt, “Artificial Intelligence Residuals,” Wash. Post. (July 7, 2023),
<https://www.washingtonpost.com/opinions/2023/07/26/joseph-gordon-levitt-artificial-intelligence-residuals/>
- Roger McNamee, *Zucked: Waking Up to the Facebook Catastrophe* (Penguin, 2019), pp. 31- 79, <https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:d255f0bc-3c8c-351c-b119-6d9819a53e07>

December 2: Law Class: Law’s Potential Roles in Encouraging and Requiring Responsible Computing

Prompt: Should law create economic incentives, command-and-control regulations, performance standards, or other techniques to encourage or require responsible computing? And what law— legislative, administrative, or judicially-made? State, federal, international? What approaches are likely to be effective, and what likely to be feasible?

Required readings:

- Guido Calabresi and A. Douglas Melamed, “Property Rules, Liability Rules, and Inalienability: One View of the Cathedral,” 85 *Harv. L. Rev.* 1089 (April 1972), <https://cyber.harvard.edu/openlaw/DVD/articles/calabresi.html>
- Alda Joaquin Acosta, “Three Practical Tools to Help Regulators Develop Better Laws and Policies: Policy Paper on Autonomous Vehicles,” Berkman Klein Center, (July 2018), https://cyber.harvard.edu/sites/default/files/2018-07/2018-07_AVs04_1.pdf
- Joaquin Quiñonero-Candela, Yuwen Wu, Brian Hsu, Sakshi Jain, Jen Ramos, Jon Adams, Robert Hallman, and Kinjal Basu “Disentangling and Operationalizing AI Fairness at LinkedIn,” In 2023 ACM Conference on Fairness, Accountability, and Transparency (FACCT ’23), June 12–15, 2023, Chicago, IL, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3593013.3594075>, or https://dl.acm.org/doi/pdf/10.1145/3593013.3594075?casa_token=D90pWVy7z5MAAAAA:owGjvscq0IzCcTFbIEzrR2lY3XupmjpBQ2HUV10vCeWnyUqyMmih_rr1hUvX2lVpMG6i8VR35d8kh
- Cade Metz, Katie Robertson, “OpenAI Seeks to Dismiss Parts of The New York Times’s Lawsuit,” (Feb. 27, 2024), <https://www.nytimes.com/2024/02/27/technology/openai-new-york-times-lawsuit.html>
- FTC Letter to EU Commission describing recent FTC actions and commitments: https://www.ftc.gov/system/files/ftc_gov/pdf/2023.06.08letter-to-commissioner-reynders.pdf

Optional readings:

- Paola Inveradi, “The European Perspective on Responsible Computing” (Apr. 2019), <https://dl.acm.org/doi/pdf/10.1145/3311783>
- Daniel H. Cole and Peter Z. Grossman, “When Is Command-and-Control Efficient?”

Institutions, Technology, and the Comparative Efficiency of Alternative Regulatory Regimes for Environmental Protection" (1999). Articles by Maurer Faculty. 590, <https://www.repository.law.indiana.edu/facpub/590/>

- Tom Wheeler, "Are the FTC's Tools strong Enough for Digital Challenges?", Brookings, (May 10, 2023). <https://www.brookings.edu/articles/are-the-ftcs-tools-strong-enough-for-digital-challenges/>
- "Two New Books Explore The Upside of Big Data and AI", The Economist (Nov. 30, 2022). <https://www.economist.com/culture/2022/11/30/two-new-books-explore-the-upside-of-big-data-and-ai>
- Salome Viljoen, "A Relational Theory of Data Governance," 131 Yale L.J. 573 (2021), [131.2 Viljoen 1n12myx5.pdf \(yalelawjournal.org\)](https://www.yalelawjournal.org/131.2_Viljoen_1n12myx5.pdf)
- Matthijs Maas and Jose Jaime Villalobos, International AI Institutions: A Literature Review of Models, Examples, and Proposals, https://www.researchgate.net/publication/374148901_International_AI_Institutions_A_Literature_Review_of_Models_Examples_and_Proposals
- Gerrit De Vynck, Microsoft calls for new laws on deepfake fraud, AI sexual abuse images, The Washington Post (July 30, 2024), <https://www.washingtonpost.com/politics/2024/07/30/microsoft-ai-deepfake-law-fraud/>

December 4: CS class: Advanced Topic in Cryptography (possibly Zero Knowledge)

