CS182: Ethics, Public Policy, and Technological Change

Cross-listings: COMM 180, ETHICSOC 182, PHIL 82, POLISCI 182, PUBLPOL 182

Winter 2025-26

Lectures: Mon/Wed 3-4:20PM PT in CoDa B80

Sections: Fri 3-4:20PM PT (location varies by section assignment)

Course Website: cs182.stanford.edu

Professor Rob Reich
reich@stanford.edu

Professor Mehran Sahami
sahami@cs.stanford.edu

Teaching Team and Course Staff

We have assembled an interdisciplinary teaching team from across the university, including course (teaching) assistants with backgrounds in Computer Science, Philosophy, Communication, and Political Science. Collectively, they bring an array of relevant academic and professional experiences, including serving in government, working at technology companies, and leading initiatives to address many of the issues we'll be discussing. Throughout the course, you will have the opportunity to interact with many different teaching assistants.

Roberta Fischli (Head Course Assistant), fischli@stanford.edu)
Samantha Bennett, sab333@stanford.edu
Mary Fetter, mfetter@stanford.edu
Thay Graciano, thayg@stanford.edu
Aditesh Kumar, aditesh@stanford.edu
Neha Srivathsa, nehasriv@stanford.edu
Dominic Zappia, zappia@stanford.edu

You will find a calendar of Office Hours for the course staff on the course website. The calendar will be populated by the end of the first week of the quarter.

If you have questions at any point during the quarter, please start an inquiry by contacting Roberta, the head Course Assistant. You are also always welcome to contact Rob and Mehran.

Course Description

Our goal is to explore the ethical and social dimensions of technological innovation. Stanford has a special responsibility to address these topics because of its role as a seedbed of Silicon Valley. By integrating perspectives from computer science, philosophy, law, and social science, the course will provide learning experiences that robustly and holistically examine the impact of technology on humans and societies.

The course will challenge students, whatever their choice of major and whatever their career

pathway, to think about their role as enablers and shapers of technological change in society. Instead of accepting a common view that what others do with new technologies is their responsibility, students will explore their responsibilities as innovators, designers, coders, engineers, corporate leaders, policymakers, citizens, and consumers. With every new innovation, students will ask: What am I enabling others to do? What responsibilities does this imply for me as an innovator, a citizen, and a human being?

The content of our course will reflect the latest in technological innovation. We are building, however, on the long history of ethics and CS at Stanford. The CS department began offering courses in this area in the 1980s, taught by Terry Winograd (Computer Science) and Helen Nissenbaum (Philosophy). Eric Roberts (Computer Science) taught classes on Ethics and Computer Science in the 1990s and 2000s.

Targeting students from across multiple disciplines, the course will offer learning opportunities to distinct student populations: giving CS students a greater appreciation of the ethical and policy questions that arise in real-world technical contexts, and students from the humanities, social sciences, and law a deeper understanding of the technical topics underlying many of today's policy and ethical debates.

Course Topics

The course is structured around four core units, which have been selected primarily for two reasons: (a) to preview critical issues you are likely to play a role in shaping over the next decade and (b) to emphasize topics around which technologists could benefit from greater engagement with domains of knowledge found in other disciplines, including philosophy, social science, law, and public policy.

Unit	Sub-topics of interest
Algorithmic Decision-Making	● Use of predictive algorithms in public vs. private settings, with emphasis on the criminal justice system ● Comparative approaches to algorithmic transparency and accountability (e.g., auditing, technological due process) ● Trade-offs between predictive accuracy and competing values (e.g., fairness, transparency, explainability) ● Mapping automated systems onto a society characterized by human judgments, informal norms, and formal rules
Data Collection, Privacy, and Civil Liberties	● Data aggregation, matching, and de-anonymization strategies ● Facial recognition technology (by public and private actors) ● Differential privacy and machine unlearning ● Regulatory, policy, and legal interventions to protect privacy rights, as well as their limitations

Power of Private Platforms	● Transition from an analog to a digital public sphere, with speech and associational rights regulated by companies; virality over veracity in online discourse; tensions between quantity and quality of information; implications for democracy ● Business model concerns, including new conceptions of monopoly and market power of digital platforms, as well as government efforts to promote market competition (e.g., antitrust regulation) ● Technology behind efforts to govern speech in online and AI settings, including content moderation practices (e.g., banning/deleting speech, upranking/downranking content), frontiers/innovations in speech regulation
Frontier AI	● The evolution of AI systems, scaling laws, and functionality of generative AI systems, and potential risks ● Governance of AI systems, including geopolitics and national security concerns ● Labor and economic impact of generative AI systems (e.g., deployment in software engineering) ● Openness of innovation and concerns around open models sparking regulatory proposals at state and national level

Pedagogical Structure

For each of the four modules, we will have a sequence of lectures, discussions, and assignments. Each faculty member will present material relevant to their subject expertise, and course assistants will lead weekly small-group sections.

This approach illustrates our commitment to a genuine integration of multiple disciplinary approaches. The assignments, course materials, and classroom discussions seek to combine the technical and non-technical elements of each topic. Throughout the quarter, we will stimulate discussions and provide assignments that require different disciplinary lenses, including technical exercises, policy memos, and philosophical analyses.

We aim to develop three core skills. First, students will develop an understanding of how to **identify and frame ethical issues** with technological design, development, and deployment choices. For instance, should designers of criminal justice algorithms have obligations to ensure that such tools do not exacerbate demographic disparities? We will ground these in philosophical frameworks to understand how (a) “consequentialist” or utilitarian approaches – which focus on the consequences of actions (in the utilitarian case, maximizing the greatest good for the greatest number) – or (b) “deontological” approaches – which focus on moral rules, rights, and duties – would guide technological choices. For instance, a utilitarian approach might justify the use of criminal justice algorithms on the safety benefits for the community at large, but a deontological approach might object based on whether such algorithms engage in fair treatment of individuals.

Second, we will explore **technical approaches** that can exacerbate or mitigate such ethical problems. In many instances, spotting ethical issues can inspire technological solutions. The field of algorithmic fairness, for instance, has developed technical approaches to mitigating demographic disparities. In other instances, the rise of technology can make such ethical problems worse. The widespread deployment of facial recognition technology, for instance, may exacerbate concerns about privacy and surveillance. In many instances, technology is not a silver bullet to resolving such ethical issues, requiring other social interventions.

Third, we will explore how to **analyze proposed policy interventions** to address the governance of technologies. When technical solutions fall short, there is often a call for legal and policy interventions. Students will learn fundamentals of policy and social science analysis, such as the evaluation of policy alternatives, identification of stakeholders, decision process, and choice of regulatory institution. For instance, how should we assess proposals to mandate red teaming of Generative AI models or to impose liability for harms to Generative AI developers? How do policy choices navigate the tensions and tradeoffs in light of stakeholder dynamics? Understanding what is institutionally and technically feasible becomes central to effective policy.

Throughout the course, we will surface the following themes:

- **Promise and Perils**: introduction to topic and competing values at stake
- **Technical Deep Dive**: overview of relevant computer science topics
- **Rights and Responsibilities**: governance implications
- **Tensions and Trade-offs**: understanding the tradeoffs of design decisions on competing values (e.g., privacy vs. accuracy)
- **Distributive Effects**: the relevance of social categories such as race and gender to the design, development, funding, and deployment of new technologies
- **Making Choices**: designing a product/system/policy in light of competing values, distributive effects, and trade-offs

Course Requirements, Assignments, and Grading Breakdown

We have assigned a carefully curated set of readings for each class session. We expect you to complete this reading in advance of each lecture, and to come to section prepared to engage the materials in a facilitated discussion.

Most of the readings for the course are easily accessible through the reading list on the course website. Some readings, however, come from books not available through our university library. ***Students must purchase a digital coursepack from the Stanford Bookstore in order to access this copyrighted material.*** The bookstore will provide a digital code that allows you to access the coursepack, which contains excerpts from books we'll be discussing during the quarter. We will provide additional information on how to access the coursepack during the first week of the class.

Assignments and Grading Breakdown

The course includes five assignments. You will receive more information about each of the assignments well in advance of their due dates. The assignments are as follows:

- Technical assignment 1 (algorithmic decision-making) – **due January 27 at 11:59pm PST**
- Philosophy paper – **due February 12 at 11:59pm PST** [NOTE: WIM students will have an additional revision due **March 1 at 12noon PST**]
- Technical assignment 2 (social network) – **due Feb 26 at 11:59pm PST**
- Group policy assignment – **due March 12 at 11:59pm PST**
- Final reflection paper – **due on March 17 at 11:59pm PST**

Grades will be calculated as follows:

Non-WIM Students	**WIM Students**
• Participation – 20% • Technical Assignment 1 – 14% • Philosophy Paper – 20% • Policy Assignment – 20% • Technical Assignment 2 – 6% • Final Reflection – 20%	• Participation – 15% • Technical Assignment 1 – 14% • Philosophy Paper (original) – 9% • Philosophy Paper (revision) – 21% • Policy Assignment – 20% • Technical Assignment 2 – 6% • Final Reflection – 15%

Stanford University and its faculty are committed to ensuring that all courses are financially accessible to all students. If you are an undergraduate who needs assistance with the cost of course readings, supplies, materials and/or fees, you are welcome to approach us directly. If you would prefer not to approach us directly, please note that you can reach out to the First Generation and/or Low Income Student Success Center (https://fli.stanford.edu/about-flissc).

Sections

In addition to lectures on Mondays and Wednesdays, sections will take place every Friday from 3:00 – 4:00pm PT, starting the first week of class. You will receive notification of the location and teaching assistant for your section by email. **You do not need to sign up for section on Axess, as that will not reflect your actual section assignment.**

Class Participation and Lecture Attendance

Class participation can take a variety of forms, ranging from the obvious (e.g., talking intelligently in class/section) to the less obvious (e.g., sharing articles/podcasts that are relevant to course discussions). *At a minimum, it is crucial that you join class on time*, *having done the reading*, *and prepared to talk and engage your fellow classmates*. Because the class will use small group discussions, adequate preparation, willingness to contribute, and capacity for empathetic listening are all required. You are also required to attend a section every week. A portion of your grade will be based on your participation.

Lecture and Section Participation: This course encourages vigorous intellectual exchange, the expression of various viewpoints, and the ability to speak effectively and cogently. Participation includes but is not limited to in-class discussion. As part of the participation grade, the section leaders may assign activities and written assignments.

Except in cases of an OAE accommodation or another valid reason that is brought to the course managers and a student's course assistant's attention *before* lecture, **attendance at lectures and sections is mandatory.** If a student has a prolonged illness or a personal situation that might lead to more than one section absence, the student should contact his or her Course Assistant before missing section. Under certain conditions, a student may be provided an opportunity to make up work missed in section. In other words, make-up work is at the discretion of the instructor. Note: insufficient section attendance can result in failure of the course.

In order to be prepared for discussion, it is essential that you come to each lecture having read and understood the materials assigned and having given some thought as to how the readings

relate to the course in general. This will allow you to benefit from the class presentations and discussions and in turn prepare yourself to discuss the issues in depth in section. You should come to section with considered views about:

- *what the main claims offered in the texts or case studies are;*
- *the arguments offered in favor of these claims;*
- *whether these are good or plausible arguments;*
- *whether the claim is, all things considered, strong or plausible;*
- *what alternatives to the claims and arguments exist; and*
- *whether some alternative is superior to the claim under discussion.*

Objections are important. But keep in mind that raising puzzles and problems (even interesting puzzles and problems) for a view is easy: we can be certain in advance that every view will face some problems. Still, we are trying to decide what to think about important issues of enormous consequence, not playing a game or showing off debater's skills. The really hard part is to figure out what to think – what we should think – once we understand the range of theoretical options and competing arguments.

Participation in section will be evaluated on the following guidelines, which stress the quality rather than the quantity of contributions.

- **A range:** The student is fully engaged and highly motivated. This student is well prepared, having studied the assigned material, and having thought carefully about the materials' relation to issues raised in lecture and section. The student's ideas and questions are substantive (either constructive or critical); they stimulate class discussions. This student listens and responds respectfully to the contributions of other students.
- **B range:** The student participates consistently in discussion. This student comes to section well-prepared and contributes regularly by sharing thoughts and questions that show insight and a familiarity with the material. This student refers to the materials discussed in lecture and shows interest in other students' contributions.
- **C range:** The student meets the basic requirements of section participation. This student is usually prepared and participates once in a while but not regularly. The student's contributions relate to the texts and the lectures and offer a few insightful ideas but do not help to build a coherent and productive discussion. (Failure to fulfill satisfactorily any of these criteria will result in a grade of "D" or below.)

<u>Statement on Open Discourse</u>

The course is a space for students committed to a rigorous examination of ethics, technology, public policy, and related topics. The course is also a space for respectful, critical inquiry through the free exchange of ideas. Our goal is to come to a greater understanding of – not a consensus on – the issues the course addresses. To that end, this space is defined by mutual respect that allows us, together, to grapple with a range of ideas, evidence, values, and conclusions. The following principles guide our interaction in this space:

- All viewpoints are welcome.
- Treat every member of the course with respect, even if they disagree with another student's view.
- Treat every claim as open to examination, even if it comes from someone with more experience or expertise than you.
- Reasonable minds can differ on any number of perspectives, opinions, and conclusions.
- Our passions and social and political commitments are welcomed in this space. They are also subject to respectful challenge.
- Some perspectives, opinions, and conclusions are unreasonable or based on falsehoods

and should be identified as such.
- No ideas are immune from scrutiny and debate.
- Evidence and reasoning guide our conclusions.

For more information, see:
- Free Expression, University of Chicago
- Class Community Commitments: A Guide for Instructors, CTL
- Transformative Class Conversations, CTL

The Honor Code

Violating the Honor Code is a serious offense, even when the violation is unintentional. The Honor Code is available at: https://communitystandards.stanford.edu/policies-guidance/honor-code. You are responsible for understanding the University rules regarding academic integrity; you should familiarize yourself with the code if you have not already done so. In brief, conduct prohibited by the Honor Code includes all forms of academic dishonesty, among them copying from another student's work, unpermitted collaboration and representing as one's own work the work of another (including generative AI, such as ChatGPT). If you have any questions about these matters, see your teaching assistant during office hours.

Statement on the Use of Automated Writing/Coding Tools

The development of generative AI (a topic in the class) provides opportunities to support learning, but also opportunities to cheat. We will provide additional guidance about the use of automated writing or coding tools (e.g., ChatGPT or CoPilot) in submitting assignments.

FERPA

Student Record Privacy Policy
https://privacy.stanford.edu/policies/ferpa-overview

Access and Accommodations

Stanford is committed to providing equal educational opportunities for disabled students. Disabled students are a valued and essential part of the Stanford community. We welcome you to our class.

If you experience disability, please register with the Office of Accessible Education (OAE). Professional staff will evaluate your needs, support appropriate and reasonable accommodations, and prepare an Academic Accommodation Letter for faculty. To get started, or to re-initiate services, please visit https://oae.stanford.edu.

If you already have an Academic Accommodation Letter, we invite you to share your letter with us. Academic Accommodation Letters should be shared at the earliest possible opportunity so we may partner with you and OAE to identify any barriers to access and inclusion that might be encountered in your experience of this course.

Additional Resources for Learning

Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Student Disability Resource Center (SDRC) located within the Office of Accessible Education (OAE). SDRC staff will evaluate the request with required documentation, recommend reasonable accommodations and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is being made. Students

should contact the SDRC as soon as possible since timely notice is needed to coordinate accommodations.

Student Disability Resource Center Office of Accessible Education https://oae.stanford.edu/

The Hume Writing Center works with Stanford students taking WIM classes and any course that includes writing assignments. In free one-to-one sessions, trained writing consultants help students brainstorm and get started on assignments; learn strategies for revising, editing, and proofreading; and improve organization, flow, and argumentation. We also have digital media consultants who work with students to develop strategies to improve visual and multimodal communication in media such as research posters and PowerPoint and oral communication tutors to help students prepare or refine a presentation. Students can make an appointment with a lecturer or advanced graduate student consultant or drop in to meet with an undergraduate peer tutor. For further information, to see hours and locations, or to schedule an appointment, visit the Hume website at: http://hume.stanford.edu.

The Technical Communication Program (TCP) is a writing and public speaking resource for Stanford students of all levels. TCP is a resource specifically tailored for WIM courses offered in the School of Engineering. TCP instructors will also be working with WIM students to provide support on writing assignments. To request a consultation, please visit: https://engineering.stanford.edu/tcp.

Reading List and Course Outline

Below is a schedule of class topics and readings. They are subject to change by the instructors, and all changes will be communicated.

**Required readings that are excerpted from books and are NOT hyperlinked below will appear in the course reader, which you can obtain through the Stanford Bookstore.* Please reach out to your course assistant if you are having trouble finding the readings.*

Monday, January 5: Opening session (no pre-reading)

Wednesday, January 7: Computer Science, Optimization, and Automation

- ACM Code of Ethics and Professional Conduct
- Reich, Sahami, and Weinstein, "Why Silicon Valley's Optimization Mindset Sets Us Up For Failure," excerpt from *System Error: Where Big Tech Went Wrong and How We Can Reboot*, in Time Magazine
- Wang, Kapoor, Barocas, and Narayanan, "Against Predictive Optimization" (2023)
- Bardach's Eightfold Path to More Effective Problem Solving – Atlas of Public Management (2017)
- Carl V. Patton, David S. Sawicki, and Jennifer J. Clark, Basic Methods of Policy Analysis and Planning (Routledge 2016), pp. 30-33
- Sven Nyholm, Chapter 2, "What is Ethics?" in *This Is Technology Ethics: An Introduction* (2023).

Supplementary:

- The Limits of "Thinking like an Economist" - LPE Project by Elizabeth Popp Berman (2022)
- "Feynman's Error: On Ethical Thinking and Drifting" by Dan Munro (Dan's blog, November 2018)
- "Code is Law." Lessig, Lawrence. January 2000. Harvard Magazine.
- "Of Course Congress Is Clueless About Tech—It Killed Its Tutor" (WIRED, 2016)

Friday, January 9 (SECTION): Introductory Section

Monday, January 12: Algorithmic Decision-Making | Promise & Perils

- Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Crown Publishing Group, 2016), pp. 28-31 [available on EBSCOhost and as a scanned pdf]
- Constantaras, Geiger, et al., "Inside the Suspicion Machine", WIRED and Lighthouse Reports (2023).
- John Rawls, *Stanford Encyclopedia of Philosophy*, Sections 4.3 "The Two Principles of Justice as Fairness," 4.6 "The Original Position," and 4.7 "The Argument from the Original Position: The Selection of Principles"

Supplementary:

- John Rawls, A Theory of Justice, pp. 10-24, Section 3 "The Main Idea of the Theory of Justice," Section 4 "The Original Position and Justification," and Section 5 "Classical Utilitarianism," (Harvard University Press, 1971; revised 1999)
- The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition by OS Keyes (Proceedings of the ACM on Human-Computer Interaction, Vol. 2, No. CSCW, Article 88. Publication date: November 2018.)
- "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," pp. 1-18, 22-24 (White House, May 2016)

Wednesday, January 14: Algorithmic Decision-Making | Technical Deep Dive

- Handout on "Introduction to Probability and Machine Learning" for Background
- "Machine Bias" by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner (ProPublica, 2016)
- "How we Analyzed the COMPAS Recidivism Algorithm" by Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin (ProPublica, 2016)
- "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning" by Sam Corbett-Davies, Sharad Goel (ArXiv, 2018), Section 2.2
- "Can you make AI fairer than a judge? Play our courtroom algorithm game" by Karen Hao and Jonathan Stray (MIT Technology Review, 2019) – *the reading is optional, but it's helpful to play the game to visualize different fairness metrics*

Supplementary:

- "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning" by Sam Corbett-Davies, Sharad Goel (ArXiv, 2018), full paper
- "21 Fairness Definitions and Their Politics" by Arvind Narayanan (Tutorial at Conference on Fairness, Accountability, and Transparency, 2018)
- "Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges" by Songul Tolan (ArXiv, 2019)
- "Algorithmic decision making and the cost of fairness" by Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, Aziz Huq (Proceedings of KDD'17, 2017)
- "Inherent Trade-Offs in the Fair Determination of Risk Scores" by Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, sections 1 and 5 (Proceedings of Innovations in Theoretical Computer Science, 2017)

Friday, January 16 (SECTION)

Monday, January 19: MLK Day, No class

Wednesday, January 21: Algorithmic Decision-Making | Rights & Responsibilities

- Sven Nyholm, Chapter 6, "Responsibility and Technology: Mind the Gap(s)" in *This Is Technology Ethics: An Introduction* (2023).
- Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*, Chapter 5.
- "A Guide to Solving Social Problems with Machine Learning" by Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan (Harvard Business Review, 2016)
- "If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions" by Alex Albright, pp. 2-6

Supplementary:

- "If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions" by Alex Albright (full paper)
- "How Corporations Turned Prison Tablets into a Predatory Scheme" by Thomasso Bardelli. *Dissent Magazine.* 2021.
- Alan Gerber and Donald Green, *Field Experiments*, Chapter 1
- "Human Decisions and Machine Predictions" by Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan (The Quarterly Journal of Economics, 2018)
- "Improving Refugee Integration through Data-Driven Algorithmic Assignment" by Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, Jeremy Weinstein (Science, 2018)

- "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation" by Scott E. Carrell, Bruce I. Sacerdote, and James E. West (Econometrica, 2013)
- "Randomized Controlled Field Trials of Predictive Policing" by Mohler et al, (Journal of the American Statistical Association, 2015)
- "Predictive Policing Software Terrible At Predicting Crimes" (The Markup, 2023)

Friday, January 23 (SECTION): Algorithmic Decision-Making | Tensions & Tradeoffs I

- Case Study: Algorithmic Decision-Making and Accountability

Supplementary:

- "When an Algorithm Helps Send You to Prison" by Ellora Israni (New York Times, 2017)

Monday, January 26: Algorithmic Decision-Making | Making Choices

Note: Technical Assignment #1 is due tomorrow at 11:59pm PST!

- Frank Pasquale, *The Black Box Society*, pp. 140-143, 147-153 [jump to "The Lawful Use of Data" on p. 147] (Harvard University Press, 2016)
- Solon Barocas, Moritz Hardt, and Arvind Narayanan, Fairness and Machine Learning. Limitations and Opportunities (2023), pp. 160-169 "Regulating Machine Learning"
- "Discrimination in the Age of Algorithms" by Kleinberg et al, pp. 1-6 (NBER, 2019)
- Wright et al., Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability (FAccT 2024)

Supplementary:

- "Algorithmic Impact Assessments: Toward Accountable Automation in Public Agencies" by Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, pp. 7-20 (AI Now Institute, 2018)
- "Discriminating Systems: Gender, Race, and Power in AI" by Sarah West, Meredith Whittaker, and Kate Crawford (AI Now Institute, 2019)
- "The Scored Society: Due Process for Automated Predictions" by Danielle Citron and Frank Pasquale, pp. 18-33 (Washington Law Review, 2014)
- Frank Pasquale, *The Black Box Society*, all of ch. 5-6
- "Data and Society Report Algorithmic Impact Assessment" (Policy Brief), by Emmanuel Moss, Elizabeth Anne Watkind, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf (2021)
- "Solving the Problem of Discriminatory Ads on Facebook" by Jinjang Zang. Brookings Institute. 2021.
- "Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies," by Brianna Vecchione, Solon Barocas, Karen Levy (EAAMO 2021)
- Birhane et al., AI auditing: The Broken Bus on the Road to AI Accountability (ArXiv, 2024)

Wednesday, January 28: Data Collection, Privacy, Civil Liberties | Promise & Perils

- "Why Democracy Needs Privacy," by Carissa Veliz (Boston Review, 2021)
- Daniel J. Solove, The Limitations of Privacy Rights, 98 Notre Dame Law Review 975 (2023), Introduction
- "The Digital Poorhouse" by Virginia Eubanks (Harper's Magazine, 2018)
- "The Secretive Company That Might End Privacy as We Know It" by Kashmir Hill (New York Times, 2020)
- Stanford Administrative Guide, 6.1.1 Privacy and Access to Electronic Information

Supplementary:

- Nothing to Hide, pp. 21-32, by Daniel Solove (Yale University Press, 2013)

- "A Contextual Approach to Privacy Online" by Helen Nissenbaum (Daedalus 140 (4), Fall 2011)
- "Limitless Worker Surveillance" by Ifeoma Ajunwa, Kate Crawford, and Jason Schultz (California Law Review, 2016)
- "Privacy Harms" by Danielle Keats Citron & Daniel J. Solove (Boston University Law Review, 2022)
- "Facial Recognition: What Happens When We're Tracked Everywhere We Go?" by Kashmir Hill (New York Times, 2021)
- "The Constant Boss: Work Under Digital Surveillance" by Aiha Nguyen (Data and Society, 2021)

Friday, January 30 (SECTION): Algorithmic Decision-Making | Tensions & Tradeoffs II

- Nigel Duara, What the failure of Prop. 25 means for racial justice in California (CalMatters, 2020)
- Stephanie Wylie, How Profit Shapes the Bail Bond Industry, pp. 1-2 (Brennan Center, 2024)
- Logan Koepke and David Robinson, Civil Rights and Pretrial Risk Assessment Instruments (Upturn, 2019), pp. 4-12
- Ben Green, Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness (SSRN, 2021), Section 5

Monday, February 2: Data Collection, Privacy, Civil Liberties | Technical Deep Dive I

- "Private traits and attributes are predictable from digital records of human behavior" by Michal Kosinski, David Stillwell, and Thore Graepel (PNAS, 2013)
- "Harvard Researchers Identify Accuracy Concerns in Census Bureau's New Privacy System" by Kate N. Guerin (The Harvard Crimson, 2021)
- "Why 'Anonymous' Data Sometimes Isn't" by Bruce Schneier (WIRED, December 2007)
- Williams v. City of Detroit | American Civil Liberties Union

Wednesday, February 4: Data Collection, Privacy, Civil Liberties | Technical Deep Dive II

- Michael Kearns and Aaron Roth, *The Ethical Algorithm,* pp. 31-34 (Oxford University Press, 2020)
- "Differential Privacy: A Primer for a Non-technical Audience" by Alexandra Wood et al. (Vanderbilt Journal of Entertainment & Technology Law, 2018), pp. 211-214 (Executive Summary) and pp. 225-246 (Sections III and IV)
- Ken Ziyu Liu, Machine Unlearning in 2024 - Ken Ziyu Liu - Stanford Computer Science, Sections 1 and 2

Supplementary:

- "What is Differential Privacy?" by Matthew Green (A Few Thoughts on Cryptographic Engineering, 2016)
- "The Promise of Differential Privacy: A Tutorial on Algorithmic Techniques" by Cynthia Dwork (Microsoft Research, 2011)

Friday, February 6 (SECTION): Data Collection, Privacy, Civil Liberties | Tensions & Tradeoffs I

- Case Study: Facial Recognition
- "The two-year fight to stop Amazon from selling face recognition to the police" by Karen Hao (MIT Technology Review, 2020)

Supplementary:

- "The End of Trust" from McSweeney's and Electronic Frontier Foundation

- "The Perpetual Line-Up: Unregulated Police Face Recognition In America" from Center on Privacy and Technology at Georgetown Law
- Podesta report "Big Data: Seizing Opportunities, Preserving Values" (especially pp. 58-68)
- "Report on the Telephone Records Program Conducted under Section 215" (Privacy And Civil Liberties Oversight Board, 2014)
- "Facial recognition technology: The need for public regulation and corporate responsibility" by Brad Smith (Microsoft, 2018)
- "Most US government agencies are using facial recognition" by Russell Brandom (The Verge, 2021)

Monday, February 9: Data Collection, Privacy, Civil Liberties | Making Choices

- GDPR, Art. 5 "Principles relating to processing of personal data"
- "Comparing Privacy Laws: GDPR v. CCPA" by DataGuidance and Future of Privacy Forum (2018), pp. 26-35
- Mary D. Fan, The Hidden Harms of Privacy Penalties, 56 U.C. Davis L. Rev. 71 (2022), https://digitalcommons.law.uw.edu/faculty-articles/934, Introduction and pp. 89-93
- Kashmir Hill, "Clearview AI Used Your Face. Now You May Get a Stake in the Company." (New York Times, 2024)
- Rationales, Mechanisms, and Challenges to Regulating AI: A Concise Guide and Explanation, Section II: Types of Regulatory Interventions

Supplementary:

- "A Design for Public Trustee and Privacy Protection Regulation" by Priscilla M. Regan (Seton Hall Legislative Journal, 2020)
- Eric Glen Weyl and Eric Posner, *Radical Markets*, ch. 5 "Data as Labor" (Princeton University Press, 2018)
- "Why Don't We Just Ban Targeted Advertising? "by Gilad Edelman. (WIRED, 2020).
- "Privacy and Information Sharing" by Lee Rainie and Maeve Duggan, pp. 1-8 (skim the rest), (Pew Research Center, 2016)
- "Privacy and human behavior in the age of information" by Alessandro Acquisti, Laura Brandimarte, and George Loewenstein (Science, 2015)
- "Americans feel the tensions between privacy and security concerns" by Shiva Maniam (Pew Research Center, 2016)
- "Jaron Lanier Fixes the Internet" by Jaron Lanier and Adam Westbrook (video series) (New York Times, 2019)
- "Information Fiduciaries and the First Amendment" by Jack Balkin (UC Davis Law Review, 2016)
- "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI" by Sandra Wachter and Brent Mittelstadt, pp. 1-18, 78-85 (Columbia Business Law Review)
- "We May Own Our Data, but Facebook Has a Duty to Protect It" by Nathan Heller (New Yorker, 2018)
- "Privacy, Autonomy, and the Dissolution of Markets" by Kiel Brennan-Marquez and Daniel Susser (Knight Institute, 2022)
- "Privacy and Data Protection in an International Perspective" by Lee A. Bygrave, sections 3-5 (Scandinavian Studies in Law, 2010)
- "Privacy Self-Management and the Consent Dilemma" by Daniel Solove (Harvard Law Review, 2012)
- "Nudging Privacy: The Behavioral Economics of Personal Information" by Alessandro Acquisti (IEEE Security & Privacy, 2009)

Wednesday, February 11: Power of Private Platforms | Promise & Perils

Note: Your philosophy paper is due tomorrow at 11:59pm PST!

- Chapter 2 of *On Liberty* by John Stuart Mill (republished by Heterodox Academy, 2019), pp. 6-9, 20-23, 32-35
- "Democracy and the Digital Public Sphere" by Josh Cohen and Archon Fung from Digital Technology and Democratic Theory, Bernholz, Landemore, Reich, eds (University of Chicago Press, 2021), pp. 26-38 "Democracy and the Public Sphere"
- "Will Free Speech Survive the Internet," Chapter 7 from System Error: Where Big Tech Went Wrong and How We Can Reboot, Reich, Sahami, and Weinstein (2021)

Supplementary:

- "A Declaration of Independence of Cyberspace" by John Perry Barlow (1996)
- "Democracy's Dilemma" by Henry Farrell and Bruce Schneier (Boston Review, 2019)
- "It's the (Democracy-Poisoning) Golden Age of Free Speech" by Zeynep Tufekci (WIRED, 2018)
- "Is the First Amendment Obsolete?" by Tim Wu, pp. 2-17 (Columbia Public Law Research Paper, 2018)
- Tim Wu, The Curse of Bigness, Intro + ch. 7
- Stanford Encyclopedia of Philosophy, Freedom of Speech

Friday, February 13 (SECTION): Data Collection, Privacy, Civil Liberties | Tensions & Tradeoffs II

- "The movement to limit face recognition tech might finally get a win" by Tate Ryan-Mosley (MIT Technology Review, 2023)
- Face Recognition Technology Evaluation (FRTE) 1:1 Verification
- "The US wants to use facial recognition to identify migrant children as they age" by Eileen Guo (MIT Technology Review, 2024)

Supplementary:

- GAO-24-107372, Facial Recognition Technology: Federal Law Enforcement Agency Efforts Related to Civil Rights and Training (U.S. Government Accountability Office, 2024)

Wednesday, February 18: Power of Private Platforms | Technical Deep Dive I

- CHI Lites 2019: Joseph Konstan - What makes a good recommendation? (Video, 17:32)
- "Exposure to ideologically diverse news and opinion on Facebook" by Eytan Bakshy, Solomon Messing, Lada A. Adamic (2015)
- Paresh Dave, "The Scramble to Save Twitter's Research From Elon Musk" (WIRED, 2023)
- Fukuyama et al., Report of the Working Group on Platform Scale | FSI (2020), Sections IV.C. and IV.E.

Supplementary:

- "Personalized News Recommendation Based on Click Behavior" by Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen (Google, 2009)
- "Filter Bubbles, Echo Chamber, and Online News Consumption" by Seth Flaxman, Sharad Goel, and Justin M. Rao (Public Opinion Quarterly, 2016)
- "I Was the Head of Trust and Safety at Twitter. This Is What Could Become of It." by Yoel Roth (New York Times, 2022)
- "The Algorithmic Rise of the "Alt-Right"" by Jessie Daniels Contexts, 17(1), 60–65. https://doi.org/10.1177/1536504218766547

Friday, February 20 (SECTION): Power of Private Platforms | Tensions & Tradeoffs I

- Case Study: Platforms

Supplementary:

- "The Social Responsibility of Business is to Increase its Profits" by Milton Friedman (The New York Times Magazine, 1970)

Monday, February 23: Power of Private Platforms | Technical Deep Dive II

- Halevy et al., "Preserving Integrity in Online Social Networks" (Communications of the ACM, 2022)
- Arvind Narayanan and Sayash Kapoor, *AI Snake Oil* (2024), Ch. 6 "Why Can't AI Fix Social Media?", pp. 179-205 (available online via DeGruyter)

Supplementary:

- Andrew Deck, "AI moderation is no match for hate speech in Ethiopian languages" (Rest of World, 2023)
- Casey Newton, "The secret lives of Facebook moderators in America" (The Verge, 2019)
- "Trends in the Diffusion of Misinformation on Social Media" by Hunt Allcott, Matthew Gentzkow, Chuan Zu (2018)
- "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature" by Josh Tucker (Hewlett Foundation, 2018)
- "The Disinformation Report: The Tactics & Tropes of the Internet Research Agency" (New Knowledge, 2018)
- "Free Speech is a Triangle" by Jack Balkin (Columbia Law Review, 2018)
- "Testing the Marketplace of Ideas" by Daniel E. Ho and Frederick Schauer (NYU Law Review, 2015)
- "Tech Platforms and the Knowledge Problem" by Frank Pasquale (American Affairs, 2018)

Wednesday, February 25: Power of Private Platforms | Making Choices

Note: Technical Assignment #2 is due tomorrow at 11:59pm PST!

- "From Private Bads to Public Goods: Adapting Public Utility Regulation for Informational Infrastructure" by K. Sabeel Rahman and Zephyr Teachout (Knight First Amendment Institute, 2020)
- "The New Governors: The People, Rules, and Processes Governing Online Speech" by Kate Klonick, pp. 1599-1603 (Harvard Law Review, 2018)
- "Free Speech Is Not the Same As Free Reach" by Renee DiResta (WIRED, 2018)
- Dealing with Disinformation: Evaluating the Case for Amendment of Section 230 of the Communications Decency Act" by Tim Hwang, pp. 269-280 in Social Media and Democracy Edited by Nathaniel Persily and Joshua A. Tucker (Cambridge University Press, 2020)

Supplementary:

- "Dealing with Disinformation: Evaluating the Case for Amendment of Section 230 of the Communications Decency Act" by Tim Hwang, pp. 252-285 in Social Media and Democracy Edited by Nathaniel Persily and Joshua A. Tucker (Cambridge University Press, 2020), full paper
- "The New Governors: The People, Rules, and Processes Governing Online Speech" by Kate Klonick, pp. 1599-1603, 1616-31, 1662-70 (Harvard Law Review, 2018), full paper
- EU Digital Policy and International Trade (2021) prepared by Congressional Research Service for Members of Congress
- Chapter 4: "Private Platforms and Public Infrastructure" from *Platform Socialism* by James Muldoon. Pluto Press. 2022.
- Daphne Keller, Platform Transparency and the First Amendment (*Journal of Free Speech Law*, 2023)

Friday, February 27 (SECTION): Power of Private Platforms | Tensions & Tradeoffs II

Note [WIM students only]: Your revised philosophy paper is due on March 1 at 12noon PST!

- Lauren Feiner, "Breaking down the DOJ's plan to end Google's search monopoly" (The Verge, 2024)
- Toussaint Nothias, "An Intellectual History of Digital Colonialism" (Journal of Communication, October 2025)

Supplementary:
- K. Sabeel Rahman, The New Octopus (Logic(s) Magazine, 2018)

Monday, March 2: Generative AI | Risks and Opportunities
- Kaplan and McCandlish et al., "Scaling Laws for Neural Language Models" (2020), pp. 1-9
- "Sparks of AGI" pp. 5, 54-60
- Executive summary of the AI index (2025)

Supplementary:
- "Google Eats Rocks, a Win for AI Interpretability, and Safety Vibe Check," Hard Fork, 2024.
- "OpenAI's Reinforcement Finetuning and RL for the masses" by Nathan Lambert (Interconnects, 2024)
- Bender, Gebru, McMillan-Major, and Shmitchell, "On the Dangers of Stochastic Parrots" (FAccT 2021)
- "The real AI fight" by Cory Doctorow (Pluralistic, 2023), pp. 1-3.

Wednesday, March 4: Generative AI | Implications of Using GenAI Systems
- "AI and Just Transitions for American Workers," NAIAC 2024, pp. 3-12
- Perry et al., Do Users Write More Insecure Code with AI Assistants? (2023)
- Weidinger et al., Sociotechnical Safety Evaluation of Generative AI Systems (2023), Sections 1 and 4
- Tamkin and McCain et al., "Clio: Privacy-Preserving Insights into Real-World AI Use" (2024), pp. 1-6

Supplementary:
- Brynjolfsson, Li, Raymond, "Generative AI at Work" (2023), pp. 1-10 Sections 1 and 2
- Vipra and Korinek, "Market Concentration Implications of Foundation Models" (2023), pp. 1-10
- Goldstein et al., "Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations" (2023), Executive Summary

Friday, March 6 (SECTION): Generative AI | Release Policies
- Solaiman et al., "Release Strategies and the Social Impacts of Language Models" (OpenAI, 2019)
- "OpenAI has released the largest version yet of its fake-news-spewing AI," by Karen Hao (MIT Tech Review, 2019)
- Irene Solaiman, "The Gradient of Generative AI Release: Methods and Considerations" (FAccT, 2023)
- "GPT2, Five Years On" Jack Clark, Import AI (2024)
- Verma, Tiku, and Zakrzewski, "OpenAI promised to make its AI safe. Employees say it 'failed' its first test." (Washington Post, 2024)

Monday, March 9: Generative AI | Policy Interventions
- Kapoor and Bommasani et al., On the Societal Impact of Open Foundation Models (2024)
- Reuel and Bucknall et al., Open Problems in Technical AI Governance, Introduction and Policy Brief (2024)

- Henderson et al., Safety Risks from Customizing Foundation Models via Fine-Tuning (2024)

Supplementary:
- California SB1047
- Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Section 4
- National Security Memorandum on Advancing the United States' Leadership in AI, Harnessing AI to Fulfill National Security Objectives, and Fostering the Safety, Security, and Trustworthiness of AI, Framework to Advance AI Governance and Risk Management in National Security
- "AI Red Teaming is not a One Stop Solution for AI Harms", Sorelle Friedler et al. (Data & Society, 2023)

Wednesday, March 11: Wrapping Up
Note: Your group policy memo is due tomorrow at 11:59pm PST!
- "The Ones Who Walk Away from Omelas" by Ursula K. Le Guin (1973)
- "Confucius and the Whistleblower" by Peter Wei, Palladium Magazine (2022)

Supplementary:
- "Uncanny Valley" by Anna Wiener (n+1, 2016)
- "What is Technology?" by Saffron Huang (Letters to a Young Technologist)
- "Value By Instrumentalization" by Jasmine Wang (Letters to a Young Technologist)
- "Study the Past, Create the Future" by Matthew Jordan (Letters to a Young Technologist)
- "To be a Technologist is to be Human" by Saffron Huang, Maran Nelson (Letters to a Young Technologist)
- "It's Time to Govern" by Anna Mitchell (Letters to a Young Technologist)

Friday, March 13 (SECTION): Generative AI | Ghost Work
Note: Your Final Reflection Paper is due on March 17 at 11:59pm PST!
- Data Workers, In Their Own Voices (Tech Policy Press, 2024)
- Training AI takes heavy toll on Kenyans working for $2 an hour (60 Minutes, 2024)
- "U.S tech giants are building dozens of data centers in Chile. Locals are fighting back," (Rest of World, 2024), pp. 1-5