Sociology 190: AI and Society
Spring 2025

Nataliya Nedzhvetskaya
Department of Sociology, UC Berkeley
nataliyan@berkeley.edu

AI is a social technology. It is designed by teams of developers who are employed by corporations, governments, academic institutions, or non-profit organizations. Using massive amounts of (mostly) human generated and annotated data, the technology learns from what it is fed. AI promises tremendous strides for human societies, for instance, improving the output of human workers and speeding the pace of scientific research and development. However, left unexamined, it also reflects and automates the biases that come both from its designers and the data it was trained on. It consumes tremendous amounts of natural resources and leads us to important questions about what sort of energy expenditures we can justify. Without careful consideration of how this technology is designed and integrated into the systems we use, it can reproduce and amplify existing social inequalities– even as it holds immense promise to solve some of our most intractable social problems. This class leads students through an examination of recent academic literature on AI as it relates to society and asks students about what decisions they would make as leaders in this space in order to build the sort of society they wish to see.

This course is part of the Discovery Initiative, a University-wide program designed to give undergraduates the opportunity to get involved with research through coursework. As part of the Discovery Initiative Program, students will have the opportunity to engage in original research with the possibility of contributing to an academic publication. In addition to substantive readings on AI, this class will include methodological training in qualitative research methods

-- PLEASE NOTE: This syllabus is a live document subject to change. Please track course announcements and check the dynamic bCourses syllabus as opposed to this static version weekly.—

Location and Time: Dwinelle 130.Thursdays 4-6pm

Required Texts: Please borrow or purchase in advance: Warren, Caroll and Tracy Karner. (2015). *Discovering Qualitative Methods.*

Reading Strategy: Unless otherwise indicated, all readings on this syllabus are required. However, you will see that some of the readings are marked [skim]. These tend to be pretty technical papers written by academics and policy makers. Students are advised to skim these for high-level comprehension of the data, research methods, and takeaways– as opposed to full technical mastery. (*Rule of thumb: spend no more than 20 minutes! Read the introduction and conclusion first*). No technical background is necessary to take this class. Part of the intention of this class is to acclimate students to conversations that involve technical jargon– and to enforce unfamiliarity as an invitation for learning, not a principle for exclusion.

Academic Integrity: Please review the University policy on academic integrity in the Code of Student Conduct: https://conduct.berkeley.edu/code-of-conduct/. Cheating, plagiarism, and other academic misconduct will result in a failing grade on the assignment and will be reported to The Center for Student Conduct. Students are allowed to use generative AI to aid in completing assignments, however, I *strongly, strongly* caution against *exclusively* relying on LLM output. Assignments that give the appearance of being largely LLM output without the student's own input will receive failing grades.

DSP Accommodations: Students with DSP accommodations should have the DSP office email me with a digital copy of their accommodation letter. Accommodations cannot be applied retroactively.

Office Hours: My office hours will be Monday from 2-3:45pm either in-person (Social Sciences Building 378) or on Zoom starting January 27th. You *must* sign up for office hours in advance–otherwise there is no guarantee I will be in the room or on the Zoom link. All students are required to attend at least one office hours session over the course of the semester.

Sign-up link: https://www.wejoinin.com/nataliyan@berkeley.edu
Zoom link: https://berkeley.zoom.us/j/2440254856

Late Policy: Please reach out in advance if an assignment will be late. I will grant extensions on a case-by-case basis for assignments other than the final paper. For the final paper, students will lose half a letter grade (A to A-) for every 24 hours past the original due date (Monday, May 12th at 11:59pm). *The last date papers will be accepted is Friday, May 16th.*

Week 1. Brave New World: Component Parts of AI (Jan. 23rd)

"But what is a GPT? Visual intro to transformers | Chapter 5, Deep Learning". *3Blue1Brown.* https://www.youtube.com/watch?v=wjZofJX0v4M [27 minute video; focus on intuitive understanding rather than technical comprehension]

Crawford, Kate, and Vladan Joler. (2018). Anatomy of an AI system. https://anatomyof.ai/ [infographic]

[skim] Vaswani, A. et al. (2017)."Attention is All You Need." *Advances in Neural Information Processing Systems.* https://user.phil.hhu.de/~cwurm/wp-content/uploads/2020/01/7181-attention-is-all-you-need.pdf

Week 2. Theorizing Machines (Jan. 30th)

Burrell, Jenna. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*. 1-12.

Yiu, Eunice, Eliza Kosoy, and Alison Gopnik. (2023). "Transmission Versus Truth, Imitation Versus Innovation: What Children Can Do That Large Language and Language-and-Vision Models Cannot (Yet)." *Perspectives on Psychological Science*. 19 (5):874-883.

[skim] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., . . . Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency. Proceedings from FAT* '19.*

Week 3. Theorizing the AI Economy (Feb. 6th)

Mallaby, Sebastian. (2022). *The Power Law: Venture Capital and the Making of the New Future.* Introduction. 1-16.

Zuboff, Shoshana. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* [Introduction: Section III. What Is Surveillance Capitalism?]. 14-18.

Warren, Caroll and Tracy Karner. (2015). *Discovering Qualitative Methods.* Chapter 6: The Interview: From Research Questions to Interview Questions. 126-150.

Week 4. Theorizing the Labor Behind AI (Feb. 13th)

Smith, Adam. (1776). *The Wealth of Nations.* "Education of Youth". 445-447.

Marx, Karl. (1844). *The Marx-Engels Reader.* "Estranged Labor" in "The Economic and Philosophic Manuscripts of 1844". Pages 70-81.

Gray, Mary L. and Siddharth Suri. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* Ch. 3. Algorithmic Cruelty and the Hidden Costs of Ghost Work. 67-93.

Warren, Caroll and Tracy Karner. (2015). *Discovering Qualitative Methods.* Chapter 7: The Interview: Interaction, Talk, and Text. 151-168 (skip "Transcription Technologies")

CITI Certification DUE 11:59pm

Interview Plan DUE 11:59pm

Week 5. Meaning of Work Under AI (Feb. 20th)

Acemoglu, Daron, Autor, David, and Simon Johnson. (2023). "Can We Have Pro-Worker AI? Choosing a path of machines in the service of minds". *MIT Shaping the Future of Work Initiative.* 1-13.

SAG-AFTRA. (2023). "SAG-AFTRA Statement on the Use of Artificial Intelligence and Digital Doubles in Media and Entertainment". https://www.sagaftra.org/sag-aftra-statement-use-artificial-intelligence-and-digital-doubles-media-and-entertainment

[skim] Dell'Acqua, Fabrizio, et al. (2023). "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." *Harvard Business School Technology & Operations Mgt. Unit Working Paper.* (24-013). 1-39.

Warren, Caroll and Tracy Karner. (2015). *Discovering Qualitative Methods.* Chapter 9: Analyzing Qualitative Data: Fieldnotes, Transcripts, Documents, and Images. 209-239.

Week 6. AI Assistants and Bosses (Feb. 27th)

[skim] Keynes, J.M. (1930). "Economic Possibilities for Our Grandchildren". In: *Essays in Persuasion.* 1-7.

Levy, Karen. (2015). "The contexts of control: Information, power, and truck-driving work." *The Information Society* 31(2): 160-174.

Griesbach, K., Reich, A., Elliott-Negri, L., & Milkman, R. (2019). Algorithmic Control in Platform Food Delivery Work. *Socius.* 1-15.

Week 7. Ghost in the Machine: AI Safety and Bias (March 6th)

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies.* Introduction. 1-14.

Denton, E., Hanna, A., Amironesei, R., Smart, A., & Nicole, H. (2021). On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society,* 8(2). 1-14.

Week 8. Environmental Impact of AI (March 13th)

Crawford, Kate. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Chapter 1: Earth. Pages 23-52.

Hao, Karen. (2020). "We read the paper that forced Timnit Gebru out of Google. Here's what it says." *MIT Technology Review.* https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

[skim] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?🦜. *Proceedings from Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.*

Week 9. AI, Bias, and Discrimination (March 20th)

Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* Chapter 1: A Society, Searching. 15-63.

Guilbeault, D., Delecourt, S., Hull, T., Desikan, B. S., Chu, M., & Nadler, E. (2024). Online images amplify gender bias. *Nature*, 626(8001), 1049-1055.

[skim] Joy, B., & Gebru, Timnit. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification *Proceedings of the 1st Conference on Fairness, Accountability and Transparency.*

Interview Transcript DUE 11:59pm

--- NO CLASS (Spring Break. March 27th)---

Week 10. Simulacra and Synthetic Data (April 3rd)

New York Times The Daily. (2024). A.I.'s Original Sin. [Listen]
https://www.nytimes.com/2024/04/16/podcasts/the-daily/ai-data.html

Please watch the following video before reading: https://www.youtube.com/watch?v=1Yxg2_6_YLs
[skim] Baudrillard, Jean. (1994). "The Precession of Simulacra." *Simulacra and Simulation.* Pages 1-42.

Jacobsen, B. N. (2023). Machine learning and the politics of synthetic data. *Big Data & Society,* 10(1). 1-12

[skim] Chen, Richard J., Ming Y. Lu, Tiffany Y. Chen, Drew FK Williamson, and Faisal Mahmood. "Synthetic data in machine learning for medicine and healthcare." *Nature Biomedical Engineering* 5, no. 6 (2021): Pages 493-497.

Peer Feedback DUE in-class assignment

Week 11. Fake News, Misinformation, and Public Opinion (April 10th)

M., R. Shoaib, Wang Z., T. Ahvanooey M., and Zhao J. 2023. Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. *International Conference on Computer and Applications (ICCA)*.

Zhou, Di and Yinxian Zhang. 2024. "Political biases and inconsistencies in bilingual GPT models—the cases of the US and China." *Nature Scientific Reports*. 14 (1):25048.

[skim with a focus on p. 12: The Deepfakes/Cheap Fakes Spectrum] Paris, Britt and Joan Donovan. Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. 2019. *Data & Society*. https://datasociety.net/library/deepfakes-and-cheap-fakes/

[skim] Artificial Intelligence Index Report. (2024). "Chapter 9: Public Opinion". *Stanford University Human-Centered AI*. 435-457. https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf

Week 12: Regulating AI (April 17th)

Artificial Intelligence Index Report 2024. "Chapter 7: Policy and Governance". *Stanford University Human-Centered AI*. Pages 366-411. https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf

Anthropic. 2023. "Collective Constitutional AI: Aligning a Language Model with Public Input". https://www.anthropic.com/news/collective-constitutional-ai-aligning-a-language-model-with-public-input

World Economic Forum. 2024. "Davos 2024: Sam Altman on the future of AI". https://www.weforum.org/agenda/2024/01/davos-2024-sam-altman-on-the-future-of-ai/#:~:text=AI%20will%20be%20able%20to%20explain%20its%20reasoning%20to%20us&text=%22I%20can't%20look%20in,to%20do%20the%20same%20thing.

[spend 15-20 minutes exploring] 2024. EU Artificial Intelligence Act. *Future of Life Institute*. https://artificialintelligenceact.eu/

[spend 15-20 minutes exploring] Office of Science and Technology Policy. 2024. "Blueprint for an AI Bill of Rights." https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Week 13: AI and Global Inequality/ Voices Outside the Valley (April 24th)

Amrute, Sareeta. (2019). "Of Techno-Ethics and Techno-Affects." *Feminist Review,* 56-73.

Arun, Chinmayi. (2020). "AI and the Global South: Designing for Other Worlds." In *The Oxford Handbook of Ethics of AI*. 589–606.

Posada, Julian. (2021). Unbiased: Why AI Needs Ethics from Below. In *A New AI Lexicon*. 1-4.

[skim] UN General Assembly. (2024). Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development. https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=mIp0qL6b7e405QgeP8&fe=true

Paper Proposal DUE 11:59pm

Week 14. Ethics and Existential Risk (May 1ˢᵗ)

Christian, Brian. (2020). *The Alignment Problem: Machine Learning and Human Values*. "Introduction".

Yudkowsky, Eliezer. (2023). "Pausing AI Developments Isn't Enough. We Need to Shut it All Down". *TIME*. https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/

The Future of Life Institute. (2023). "Pause Giant AI Experiments: An Open Letter". https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Ali, S. J., Christin, A., Smart, A., & Katila, R. (2023). Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs. *Proceedings from Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency.*

Paper Draft DUE 11:59pm

Final Papers DUE May 12ᵗʰ 11:59pm

Assignments

The student's grade will be based on a combination of course participation and engagement with the readings (45% of the grade) and research tasks culminating in a final research paper (55% of the grade). All assignments are due by 11:59pm on the day unless otherwise noted. The categories are broken down as follows:

Final Research Project (55%)		
CITI Certification	5%	Thursday, Feb. 13th
Interview Plan	5%	Thursday, Feb. 13th
Interview Transcript	3%	Thursday, March 20th
Peer Feedback (Coding Exercise)	2%	Thursday, April 3rd (in-class)
Paper Proposal (1 paragraph)	2%	Thursday, April 24th
Paper Draft (5-7 pages)	8%	Thursday, May 1st
Final Paper	30%	Monday, May 12th
Participation/Attendance in Seminar (30%)		
Participation/Attendance	25%	-
Seminar Presentation	4%	-
Discovery Initiative Surveys	1%	In-class
Discussion Questions/Comments (15%)		

Individual Assignment Breakdown

Participation/Attendance in Seminar– 30%

Participation/Attendance– 25%, in-class and office hours

This Sociology 190 seminar serves as the culmination of your undergraduate studies, offering a small class setting focused on exploring crucial sociological topics and enhancing your research abilities. Unlike traditional lecture-style courses, seminars prioritize interactive, discussion-based learning. The success of these sessions hinges on active engagement and contributions from all students. If you're uncertain about how to engage meaningfully, a helpful approach is to come prepared with a question or two, as I consistently encourage inquiries regarding the assigned readings.

Each student is allowed one unexcused absence in the course. Any additional absences must be discussed beforehand or students must provide some documentation if an unexpected emergency prevents them from attending. Class will start on Berkeley time. Regular late arrivals will impact your attendance grade.

In addition to attending seminars, student are *required* to attend at least one office hours over the course of the semester.

Seminar Presentation– 4%, in-class

Every student in the class will be required to sign up to present one of the week's readings over the course of the seminar. For longer texts, students may have the option to pair up. Synopses should be a minimum of 3 minutes and not go over five minutes. Students should focus on (a) summarize the main arguments of the text, (b) explaining what sort of evidence was used to justify the argument, and (c) their own personal take on how this reading relates to the class. I encourage students to come to office hours to discuss the reading if they are unsure of any of those three pieces.

Complete Discovery Initiative Program pre- and post-survey — 1% in-class

As part of the Discovery Initiative Program, students are encouraged to fill out a pre- and post-survey tracking what they've learned in the class. You'll receive an email for a pre-survey on February 6th around 4pm and a post-survey on May 1st around 4pm. You must fill out *both* surveys to receive 1% credit in the class. We will have dedicated class time to filling out the surveys.

Discussion Questions/Comments– 15%, due by noon the day before class

Each week, students will compose a minimum of three comments or questions on the assigned readings. On weeks with multiple readings, students must discuss at least two different texts. Comments and questions should be no longer than a paragraph and demonstrate engagement with the texts, including citations to specific passages or quotations. Students are not required to answer their own questions but should include their own reasoning in the question. Students are allowed to bring in outside texts for additional evidence or context so long as they provide citations.

Wrong way to write a discussion question:
    Why is opacity in technology harmful for society?

Right way to write a discussion question:
    On page 4 of "How the machine 'thinks'", Burrell describes opacity as technical illiteracy as "stem[ming] from an acknowledgement that...writing (and reading) code and the design of algorithms is a specialized skill... [that] remains inaccessible to the majority of the population. Does this hold true now that generative AI has significantly lowered the barrier to writing and reading code? My take is that this is *even* more important given the black box and proprietary nature of this technology (see also "opacity as intentional corporate or state secrecy" on page 3).

The discussion posts will be graded as complete or incomplete. Students are required to submit their posts by noon the day before class starting the second week of class. *Every student is allowed four passes over the course of the semester.* Additional passes will only be allowed if a student has additional excused absences over the course of the semester. Extra credit will not be given for additional posts but students will be allowed to make up for previous posts marked "incomplete" if they wish to do so.

Final Research Project – 55% total

As part of the [Discovery Initiative Program](#), students will have the opportunity to engage in original research with the possibility of contributing to an academic publication. All of these efforts will culminate in a final research paper based on the student's original research. The final research project, worth 55% of the final grade, involves multiple components, explained below.

CITI Human Subjects Research Certification: Due Feb. 13th (5%)

Before February 13th, all students will be required to complete the online CITI Human Subjects Research Training (https://cphs.berkeley.edu/training.html). I will provide more information about this in class. The CITI credential is an official document that certifies you to conduct human subjects research and is a great credential to have for anyone considering a future in research, either academic or industry. Please upload a PDF of your certificate to bCourses to complete the assignment.

Interview Plan: Due Feb. 13th (5%)

In preparation for the final paper, each student will conduct one interview of a worker who was harmed by AI listed in the AI Incidents Database (https://incidentdatabase.ai/). The interview plan should include three potential interview subjects that have been identified through the database (mentioned by name in a story) along with how you plan to contact them. It should also have a brief synopsis of the incident itself (who was involved, what happened, when and where, and what the outcome was). Include a sentence or two about why this incident is interesting to you and five interview questions you plan to ask them alongside our standard interview guide. This is also a great opportunity to propose research questions you might answer through your final paper but this is not strictly required.

Individual Interview Transcript (Cleaned and Coded): Due March 20th (3%)

Once you have conducted the interview, you will be asked to provide a cleaned and coded transcription of the full interview. We'll discuss transcription options and coding strategies in class and go through a group coding exercise.

Peer Feedback Exercise: Due April 3rd, in-class assignment (2%)

Students will be paired and asked to provide feedback on one another's interviews and coding structures. This can include overall takeaways from the interview, pointers on how the interview could have been improved, and additional codes that might be relevant. This exercise will include a brief 1-2 minute presentation of your interview and your findings to the class. As part of this exercise, you will also get a chance to consider using other students' interviews for your final paper project. This assignment will be graded pass/fail.

Final Paper Proposal (1 paragraph): Due April 24th (2%)

The proposal should be a paragraph (*maximum nine sentence*) description of your plan for the overall paper. It should include a reference to the interview you will use, a description of the interview, and a reference to at least one class reading you will utilize for your analysis. It will be graded pass/fail.

Final Paper Draft Due (minimum five pages; maximum seven pages): May 1st (8%)

The draft should be at least five pages long and should summarize the expected main argument of the paper, along with evidence from the interview. This assignment will be graded pass/fail and is designed to give the student the opportunity to receive feedback before their final submission.

Final Papers Due: May 12th (30%)

The final paper, due May 12th, will focus on a question related to the course themes and utilize the interview materials. Students get to determine the question with feedback from the instructor. I encourage students to regularly revisit their research question over the course of the class and discuss it with me at least once during office hours. The final paper should be between 15 and 20 pages, double-spaced, 12 point Times New Roman font, with 1 inch margins. It should include at least five academic citations, which can be works from the syllabus. You can use any citation style you wish so long as it's consistent.

Grade Points
A 100 to 93.33
A- < 93.33 to 90.0
B+ < 90.0 to 86.66
B < 86.66 to 83.33
B- < 83.33 to 80.0
C+ < 80.0 to 76.66
C < 76.66 to 73.33
C- < 73.33 to 70.0
D+ < 70.0 to 66.66
D < 66.66 to 63.33
D- < 63.33 to 60
F < 60 to 0.0