

Sociology SOCIOL 190
Artificial Intelligence and Society: The Promises and Limits of Technological Futures

Instructor: Skyler Wang

Semester: Fall 2023

Seminar meetings: Thursday 4-6 PM @ Dwinelle 187

Email: skyler.wang@berkeley.edu

Office hours: <https://calendly.com/skylrwang>

Little of our lives today remains untouched by Artificial Intelligence (AI), which makes understanding its reach and influence on society increasingly pertinent. This course uses an interdisciplinary approach to critically dissect AI's origins, proliferation, and ubiquity from social, political, and philosophical angles. We explore questions such as: what makes intelligence of this kind 'artificial,' and how does it differ from other types of intelligence, such as those embodied by humans or animals? What is the relationship between AI, natural language processing, machine learning, big data, and algorithms? Why is it so difficult to create AI systems that align with human values? How can we critically examine the production processes of generative or large language models such as ChatGPT to understand who they help and leave behind? By incorporating academic research, sci-fi literature, films, and a variety of guest lectures by AI practitioners, this course offers a dynamic look at the promises and limits of AI in delivering a utopic technological future.

Course structure

The seminar will meet once a week for two hours. Each class will consist of a wide array of activities, including but not limited to weekly student presentations, in-depth discussions, debates, and guest lectures. Every week, a group of students will lead the facilitation of a portion of the class by first opening up with a presentation of the assigned readings, followed by an organized class discussion (total time should not exceed 40 minutes). The students in-charge of facilitating that week's class will have the opportunity to plan the time allotted in a way they believe will lead to the most constructive discussion of the materials as long as they adhere to the theme of the topic. All students are expected to come to class having done all the required readings assigned for the week.

Required texts

All required texts will be accessible via bCourses, except for these two books (please borrow or buy in advance):

- 1) Brian Christian—*The alignment problem: Machine learning and human values*
- 2) Meghan O'Gieblyn—*God, human, animal, machine: Technology, metaphor, and the search for meaning.*

Grading and evaluation

The evaluation of the course is broken down as such:

1) Weekly group presentation & discussion facilitation (20%)

- The group presentations will draw focus from the assigned readings. The presenters' main objective is to add value to the readings by providing additional insight into the theoretical and empirical contributions of the papers. The presenters are free to choose their style of presentation, incorporating media content as they please. At the end of the day, the goal of the presentation is to generate discussion and invoke new paradigms of thinking. The presenters can critique, revise or extend the scope of knowledge as they deem fit. The presentations should last approximately 30

minutes and be followed by a 10-minute guided discussion. The entire process should not take longer than 40 minutes.

2) First writing assignment: Critical examination of an AI platform (25%)

- Students will explore the features of any AI-driven product they choose for at least three days and write a recommendation brief for the respective platform. The goal of the assignment is to get students to think critically about the structural design of AI and how they contribute to challenges for different user populations, leading to inequitable user experiences. By incorporating knowledge from course readings, students will identify areas of improvement and make evidence-based design or interface suggestions. Students are required to demonstrate their ability to use academic knowledge in a real-world setting by making their briefs accessible to the companies that they hope to preach to. The briefs should not be longer than 5 pages (double-spaced).
- **Due date: Oct 20 (Fri) before 11:59 PM.**

3) Second writing assignment: Research paper (40%)

- Students can use this opportunity to explore an area of AI that interests them the most. Although you have the freedom to choose a substantive topic of your interest, the paper should be *social* in nature; that is, it needs to involve empirical data collection to solicit opinions, attitudes, or lived experiences from people who are implicated by AI.
- Students are required to collect primary data for the final paper, either in the form of content analysis, surveys, interviews, or ethnographic (either online or offline) research. This means you should start thinking about your projects early in the semester. I recommend attending office hours to discuss your topic with me if you are not confident about the direction of your paper.
- A two-pager (double-spaced) outlining your research questions, a preliminary literature review, and methods is due on **Nov 10 (Fri) at 11:59 PM.**
- The research paper should be between 15-20 pages long (double-spaced) and include a bibliography.
- **Due date for final paper: Dec 10 (Sun) at 11:59 PM.**

4) Attendance & participation (15%)

- Students will prepare at least one question for each assigned reading and bring them to class to discuss. Participation will be graded based on the thoughtfulness of the student's questions, in-class responses, general participation, and attendance.

Academic Integrity

Any test, paper, report or homework submitted under your name is presumed to be your own original work that has not previously been submitted for credit in another course. All words and ideas written by other people must be properly attributed: fully identified as to source and the extent of your use of their work. Cheating, plagiarism, and other academic misconduct will result in a failing grade on the assignment, paper, quiz, or exam in question and will be reported to The Center for Student Conduct. Please be sure to review UC Berkeley's rules on Academic Integrity and email me if you have any questions: <https://teaching.berkeley.edu/resources/design/academic-integrity>

Accommodations

For students who require accommodations for accessibility reasons, please send me an email in the first week of class and request that the DSP office provide me with a digital copy of your accommodation letter.

Thematic outline

Week 1: Aug 24 – Introduction & class overview

- “The History of Artificial Intelligence” by Rockwell Anyoha in *Science in the News* (2017): <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

Week 2: Aug 31 – Human, animal, and artificial intelligence

- Yiu, Eunice, Eliza Kosoy, and Alison Gopnik. "Imitation versus Innovation: What children can do that large language and language-and-vision models cannot (yet)?" arXiv preprint arXiv:2305.07666 (2023).
- Lake, Brenden M., Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. "Building machines that learn and think like people." *Behavioral and brain sciences* 40 (2017).
- “Are We Smart Enough to Know How Smart AIs Are?” by Robert Long in *Asterisk Magazine* (2023): <https://asteriskmag.com/issues/03/are-we-smart-enough-to-know-how-smart-ais-are>

Week 3: Sep 7 – Values

- Lee, Kai-Fu, and Chen Qiufan. *AI 2041: Ten visions for our future*. Currency, 2021. Chapter(s): 1
- Christian, Brian. *The alignment problem: Machine learning and human values*. WW Norton & Company, 2020. Chapter(s): Prologue, Introduction, 1

Week 4: Sep 14 – The ‘thinking’ machine

- O’Gieblyn, Meghan. *God, human, animal, machine: Technology, metaphor, and the search for meaning*. Anchor, 2021. Chapter(s): 1
- Burrell, Jenna. "How the machine ‘thinks’: Understanding opacity in machine learning algorithms." *Big data & society* 3, no. 1 (2016): 2053951715622512.

Week 5: Sep 21 – Training data

- Geiger, R. Stuart, Kevin Yu, Yanlai Yang, Mindy Dai, Jie Qiu, Rebekah Tang, and Jenny Huang. "Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from?" In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 325-336. 2020.
- Denton, Emily, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. "On the genealogy of machine learning datasets: A critical history of ImageNet." *Big Data & Society* 8, no. 2 (2021): 20539517211035955.

Week 6: Sep 28 – Fairness & bias

- Christian, Brian. *The alignment problem: Machine learning and human values*. WW Norton & Company, 2020. Chapter(s): 2
- Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. "Language (technology) is power: A critical survey of" bias" in nlp." *arXiv preprint arXiv:2005.14050* (2020).

Week 7: Oct 5 – How size matters

- Kitchin, Rob. "Big Data, new epistemologies and paradigm shifts." *Big data & society* 1, no. 1 (2014): 2053951714528481.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610-623. 2021.

Week 8: Oct 12 – Playing games

- Bory, Paolo. "Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo." *Convergence* 25, no. 4 (2019): 627-642.
- Meta Fundamental AI Research Diplomacy Team (FAIR), Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried et al. "Human-level play in the game of Diplomacy by combining language models with strategic reasoning." *Science* (2022): eade9097.
- *Optional: AlphaGo - The Movie* by Deep Mind:
<https://www.youtube.com/watch?v=WXuK6gekUIY>

Week 9: Oct 19 – Writing week (no class)

First writing assignment due on Oct 20 (Fri) before 11:59 PM.

Week 10: Oct 26 – ChatGPT & generative dialogue agents

- Rudolph, Jürgen, Samson Tan, and Shannon Tan. "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?" *Journal of Applied Learning and Teaching* 6, no. 1 (2023).
- Wang, Skyler, Ned Cooper, Margaret Eby, Eun Seo Jo. "From Human-Centered to Social-Centered Artificial Intelligence: Assessing ChatGPT's Impact through Disruptive Events." (2023). Manuscript under review.

Week 11: Nov 2 – AI as social practice | Guest speaker: Dr. Paco Guzman (Research Scientist Manager at Meta AI)

- Hoffman, Steve G. "Managing ambiguities at the edge of knowledge: Research strategy and artificial intelligence labs in an era of academic capitalism." *Science, technology, & human values* 42, no. 4 (2017): 703-740.
- Joyce, Kelly, Laurel Smith-Doerr, Sharla Alegria, Susan Bell, Taylor Cruz, Steve G. Hoffman, Safiya Umoja Noble, and Benjamin Shestakofsky. "Toward a sociology of artificial intelligence: A call for research on inequalities and structural change." *Socius* 7 (2021): 2378023121999581.

Week 12: Nov 9 – AI taking jobs?

- Frank, Morgan R., David Autor, James E. Bessen, Erik Brynjolfsson, Manuel Cebrian, David J. Deming, Maryann Feldman et al. "Toward understanding the impact of artificial intelligence on labor." *Proceedings of the National Academy of Sciences* 116, no. 14 (2019): 6531-6539.
- Gonzalez, Manuel F., Weiwei Liu, Lei Shirase, David L. Tomczak, Carmen E. Lobbe, Richard Justenhoven, and Nicholas R. Martin. "Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes." *Computers in Human Behavior* 130 (2022): 107179.

Two-page outline due Nov 10 (Fri) at 11:59 PM.

Week 13: Nov 16 – AI governance | Guest speaker: Dr. Miles Brundage (Head of Policy Research, OpenAI)

- Wirtz, Bernd W., Jan C. Weyerer, and Benjamin J. Sturm. "The dark sides of artificial intelligence: An integrated AI governance framework for public administration." *International Journal of Public Administration* 43, no. 9 (2020): 818-829.
- Taeihagh, Araz. "Governance of artificial intelligence." *Policy and society* 40, no. 2 (2021): 137-157.

Week 14: Nov 23 – Thanksgiving (no class)

Week 15: Nov 30 – Toward artificial general intelligence?

- O'Gieblyn, Meghan. *God, human, animal, machine: Technology, metaphor, and the search for meaning*. Anchor, 2021. Chapter(s): 5
- Mitchell, Melanie. "Why AI is harder than we think." arXiv preprint arXiv:2104.12871 (2021).

Second writing assignment due Dec 10 (Sun) at 11:59 PM.