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

Instructor: Skyler Wang
Semester: Spring 2023
Seminar meetings: Thursday 2-4PM @ Dwinelle 250
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? What is the relationship between AI, machine learning, big data, and algorithms? Should we really worry about AI becoming 'sentient' or 'superintelligent'? How can we critically examine the production processes and organizational makeups of AI to understand who it helps and leaves behind? And lastly, how should we think about imminence of the Metaverse, a digital society where decades of AI research culminate? 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. The only material you need to purchase is Brian Christian's *The alignment problem: Machine learning and human values*.

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 + presentation (25%)

- Students will explore the features of any AI-driven product of their choice for at least three days and write a recommendation brief to 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).
- Time will be dedicated to the presentation of these briefs. Each student will be given 3 minutes to present to a panel (which is formed by class participants who would role-play as company representatives). The idea of this component is to have students present or sell their ideas in a professional and creative manner. The panel will be given the opportunity to ask questions and challenge the presenter's arguments. Evaluations will be based on both content and delivery. The presentation will be worth 10% of this component of the final grade.
- **Due date: <u>Mar 6 (Mon)</u> before 11:59PM.**

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 on 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 **<u>Apr 10 (Mon)</u> at 11:59PM**.
- The research paper should be between 15-20 pages long (double-spaced).
- The last session will be dedicated to individual presentations of the final paper (the presentation constitutes 10% of this component of the final grade).
- **Due date for final paper: <u>May 4 (Thu)</u> at 11:59PM.**

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.

<u>**Academic Integrity**</u>

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: Jan 19 – 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: Jan 26 – Human vs artificial intelligence
- French, Robert M. "The Turing Test: the first 50 years." *Trends in cognitive sciences* 4, no. 3 (2000): 115-122.
- 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).

Week 3: Feb 2 – 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: Feb 9 – The 'thinking' machine
- Waytz, Adam, Joy Heafner, and Nicholas Epley. "The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle." *Journal of experimental social psychology* 52 (2014): 113-117.
- Burrell, Jenna. "How the machine 'thinks': Understanding opacity in machine learning algorithms." *Big data & society* 3, no. 1 (2016): 2053951715622512.

Week 5: Feb 16 – 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: Feb 23 – 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 7: Mar 2 – Playing games | Guest speaker: Emily Dinan, Research Engineer (Meta AI)
- 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=WXuK6gekU1Y

==**First writing assignment due on <u>Mar 6 (Mon)</u> before 11:59PM.**==

Week 8: Mar 9 – Recommendation brief presentations

Week 9: Mar 16 – AI as social practice
- 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.
- Alkhatib, Ali. "To live in their utopia: Why algorithmic systems create absurd outcomes." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-9. 2021.
- *Optional*: 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 10: Mar 23 – Whose job
- Lee, Kai-fu. "10 Jobs That Are Safe in an AI World." (2018):
  https://kaifulee.medium.com/10-jobs-that-are-safe-in-an-ai-world-ec4c45523f4f
- Wilson, H. James, Paul Daugherty, and Nicola Bianzino. "The jobs that artificial intelligence will create." *MIT Sloan Management Review* 58, no. 4 (2017): 14.
- Huang, Ming-Hui, Roland Rust, and Vojislav Maksimovic. "The feeling economy: Managing in the next generation of artificial intelligence (AI)." *California Management Review* 61, no. 4 (2019): 43-65.

Week 11: Mar 30 – Spring break

Week 12: Apr 6 – 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).

==**Two-page outline due <u>Apr 10 (Monday)</u> at 11:59PM.**==

Week 13: Apr 13 – The Metaverse | Guest speaker: Krista Schnell, Research Program Manager (Accenture)
- Accenture's Technology Vision, 2022. Pages: 21-38
  https://www.accenture.com/content/dam/accenture/final/a-com-migration/custom/_acnmedia/thought-leadership-assets/pdf-5/Accenture-Meet-Me-in-the-Metaverse-Full-Report.pdf

- Duan, Haihan, Jiaye Li, Sizheng Fan, Zhonghao Lin, Xiao Wu, and Wei Cai. "Metaverse for social good: A university campus prototype." In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 153-161. 2021.

Week 14: Apr 20 – Toward general artificial intelligence?
- Mitchell, Melanie. "Why AI is harder than we think." arXiv preprint arXiv:2104.12871 (2021).
- Fjelland, Ragnar. "Why general artificial intelligence will not be realized." *Humanities and Social Sciences Communications* 7, no. 1 (2020): 1-9.

Week 15: Apr 27 – Wrap up & final presentations

Second writing assignment due May 4 (Thu) at 11:59PM.