USC Annenberg

COMM 499 Human-AI Interaction: Ethics, Impacts, and Values
4 Units

Fall 2025
Section #20773
Monday/Wednesday 2:00pm
Location: ANN L101

Instructor: Angel Hsing-Chi Hwang
Office: ASC 307
Office Hours: TBD
Contact info: angel.hwang@usc.edu

Course Description
This course provides an exploration of artificial intelligence (AI) and its impact on human society, focusing on the complexities of human-AI interaction, ethical considerations, and technical challenges in designing and regulating AI systems. Specifically, the course consists of three key building blocks:

1. Foundation of human-AI interaction. We begin with an introduction to generative AI and the socio-technical dynamics of human-AI interaction, setting the foundation for understanding how AI systems are designed and perceived. Students will comprehend human-AI interaction through both technical and social aspects. On the technical end, we delve into algorithmic thinking, learning how AI processes information and makes decisions. On the social side, we explore how humans interact with AI agents, examining the anthropomorphism of AI, emotional responses to robots, and social dynamics in human-AI relationships.

2. Ethics, value alignment, and challenges in building human-centered AI. Throughout the course, we examine various ethical implications of AI, with a particular focus on issues of fairness and bias in machine learning systems. Students engage in critical discussions about how AI can reinforce societal inequalities and explore ways to mitigate these biases. Specifically, we delve into a common set of "AI alignment problems," exploring how certain features of AI make it particularly difficult to align with human values in practice.

3. Applications of AI and their societal impact. The course also examines how AI systems are built and used in applied domains. These include (1) how AI applications mediate social interactions as they are increasingly built into communication technologies and social media platforms; (2) how AI reinnovate the future of work and labor markets as professionals of all kinds rapidly adopt AI tools to accomplish their day-to-day jobs; and (3) how AI supports individuals' well-being as healthcare services are increasingly carried out with AI assistance.

Student Learning Outcomes
By the end of this course, students will be able to:

- Identify and compare different human-AI interaction paradigms and their implications
- Evaluate claims of novelty and impact related to emerging AI capabilities
- Articulate their own perspectives about the social, political, and economic implications of AI applications

Prerequisite(s): NA

Course Notes

Reading & Course Materials

All reading and course materials will be distributed through Brightspace. No textbook is required, and you don't need to purchase any course materials.

Attendance

I expect consistent attendance in class. Though I will not formally take attendance, many of your assignment questions will likely be related to content that we discuss in class. If you miss two or more classes in a row, there's a good chance that you will find it challenging to complete an assignment. In that case, please reach out to me in advance.

Communication

I hope the classroom will be a safe space for every student to express their ideas. To make this happen, please be thoughtful and respectful when you speak and interact with others in class and in any other course-related interactions that take place outside of our regular class time (e.g., meetings for group projects). Please be mindful that your classmates might come from very different backgrounds and have quite different life experiences.

Outside of the classroom, Brightspace announcements and emails will be the primary communication channels we use for this class. Likewise, please be mindful and respectful while communicating through these channels as well.

If you encounter any uncomfortable or discriminatory situations during any of the above-mentioned interactions, please contact me directly.

Technological Proficiency and Hardware/Software Required

Although this course is about new technologies, the only technology required is Brightspace (Brightspace.usc.edu).

Description and Assessment of Assignments

Grades will comprise two types of assessments. The first category includes participation, engagement, and reflections on class readings, both of which will be assessed weekly throughout the semester. The second category is a semester-long project in which you will analyze a specific AI-powered application. You will choose an application at the beginning of the semester and then revisit it in subsequent assignments. Please avoid changing your project topic (i.e., your choice of AI application) in the middle of the semester.

Weekly Readings & Reflections (2 points x 10 reflections, assessed weekly)

You are expected to read all the required readings before coming to classes. Across the entire semester, please write at least 10 short reading reflections (up to 500 words) and submit them through Brightspace. You can choose which of the 10 weeks to write reading reflections on. Submissions for each week's reading reflection will close at 5 pm on each Friday.

In-class Midterm Exam (15 points)

The midterm exam will take place in class during Week 9. The exam will consist of short essay questions. These questions will be identical to discussion questions that have been covered during previous classes.

Course Project

Students will pick an application and produce a plan to study its social and/or economic impact. Students are expected to work on the course project throughout the semester instead of cramming all the work toward the end of the semester. As such, the project will be graded across multiple milestones.

- **Project Milestone #1 (5 points):** Decide on the AI application that you would like to study for your course project. In a short essay (500-1000 words), elaborate on what drives your interest and motivation to study this application and provide some preliminary ideas for what you might want to address through this study plan.
- **Project Milestone #2 (15 points):** Conduct secondary research to understand what we know and don't know about this AI application. Finalize the key questions you want to address through this study plan and write up your ideas in a short essay (1000-1500 words).
- **Project Milestone #3 (15 points):** Propose a preliminary plan for how you want to study the AI application of your choice and draft a study plan (1000-1500 words). You will also be asked to present your study plan during the last week of classes to collect feedback from me and your classmates.
- **Final Project Deliverable (25 points):** Finalize your study plan based on the feedback you collected from your presentation and write up your final study plan (2000-3000 words).

Note: On Week 2, we will vote in class to decide whether you want to carry out the course project as an individual project or group project.

Grading Breakdown

Description of assessments and corresponding points and percentage of grade.

Assessment	Due	Points	% of Grade
10 Weekly Reflections	Fridays @ 5:00 PM	2 * 10 = 20	20%
Midterm Exam	Week 9 (Date TBD)	100	25%
Project Milestone #1	Week 4, September 19 @ 5PM	100	5%
Project Milestone #2	Week 7, October 10 @ 5PM	100	15%
Project Milestone #3	Week 12, November 14 @ 5PM	100	15%
Final Project Deliverable	December 12 @ 5PM	100	20%

Course Grading Scale

Letter grade and corresponding numerical point range		
94% to 100%: A	80% to 83%: B- (B minus)	67% to 69%: D+ (D plus)
90% to 93%: A- (A minus)	77% to 79%: C+ (C plus)	64% to 66%: D

Letter grade and corresponding numerical point range		
87% to 89%: B+ (B plus)	74% to 76%: C	60% to 63%: D- (D minus)
84% to 86%: B	70% to 73%: C- (C minus)	0% to 59%: F

Grading Standards
What each letter grade demonstrates.

Letter Grade	Description
A	**Excellent; demonstrates extraordinarily high achievement; comprehensive knowledge and understanding of subject matter; all expectations met and exceeded.**
B	**Good; moderately broad knowledge and understanding of subject matter; explicitly or implicitly demonstrates good, if not thorough understanding; only minor substantive shortcomings.**
C	**Satisfactory/Fair; reasonable knowledge and understanding of subject matter; most expectations are met; despite any shortcomings, demonstrates basic level of understanding.**
D	**Marginal; minimal knowledge and understanding of subject matter; more than one significant shortcoming; deficiencies indicate only the most rudimentary level of understanding.**
F	**Failing; unacceptably low level of knowledge and understanding of subject matter; deficiencies indicate lack of understanding.**

Course Policies
Assignment Submission
All assignments should be submitted by their corresponding due dates (see the specific time and date in the course schedule). I will deduct 5% of the total points of each assignment for every 24 hours late. I understand (unexpected) things happen in life, and so everyone has the opportunity to submit one of your assignments late without penalty, but no later than 3 days after the due dates. Afterward, I will again deduct 5% of the total grades of the assignment. If you are more than 3 days past due, please email me and discuss your situation before the assignment's deadline.

Use of Generative AI
This class does not ban the use of AI for course purposes. As a researcher who studies AI, I personally believe AI will become more commonplace in your life. Therefore, if you see a way to apply AI to facilitate your coursework, that could be a helpful skill to pick up. However, as the bulk part of the coursework requires you to *apply* knowledge, if you simply use AI to automatically complete some/all parts of your assignments or projects, it is unlikely that you will get the most out of the course. This might reflect on your grades as well.

If you use AI for your coursework:

1. Be thoughtful when you formulate your prompts. Low-quality prompts result in low-quality output.
2. Think critically about the output you received. AI-generated output can have flaws and involve hallucinations.
3. Cross-check the output with trustworthy sources. As mentioned above, AI-generated output can be wrong. If you adopt false content in your assignments or projects, you are responsible for it and might risk point deductions.
4. Highlight where you adopt AI-generated text in your written assignments and projects, if any.
5. If you use AI in other ways (see below), please declare them at the end of your submission, copy the prompt(s) you use, and the raw, unedited output from AI. This part will not count toward the word/page limit of your submission.
    a. Brainstorming and idea generation
    b. Background and secondary research
    c. Source valuation and validation
    d. Creating an outline for your responses
    e. Drafting
    f. Paraphrasing and finding synonyms
    g. Revising and polishing
    h. Transforming styles
    i. Other usage (please specify)

Course Schedule

All course materials, including required readings, will be uploaded to Bright Space before class.
Important note to students: *Be advised that this syllabus is subject to change.*

	TOPICS	REQUIRED READING	DELIVERABLE
WEEK 1 DATES: 8/25-8/29	**Introduction**	• Narayanan, A. & Kapoor, S. (2024). The Long Road to Generative AI. In *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton: Princeton University Press. • Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, 25(1), 74-88. • Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301 • [Recommended] Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1	
WEEK 2 DATES: 9/1-9/5	**Algorithmic Thinking (a.k.a. how to think like AI)**	Kartik Hosanagar. (2019). *A Human's Guide to Machine Intelligence : How Algorithms Are Shaping Our Lives and How We Can Stay in Control*. Penguin Books. • Chapter 3: Omelet Recipes for Computers • Chapter 4: Algorithms Become Intelligent	Submission for weekly reading reflection #1

(9/1 LABOR DAY)		• Chapter 6: The Psychology of Algorithms • [Recommended] Chapter 5: Machine Learning and the Predictability-Resilience Paradox	closes at 5pm on Friday
WEEK 3 DATES: 9/8-9/12	**Interaction with agent**	• Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1-2), 119-155. • Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864 • Krämer, N.C., von der Pütten, A., Eimler, S. (2012). Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction. In: Zacarias, M., de Oliveira, J.V. (eds) Human-Computer Interaction: The Agency Perspective. Studies in Computational Intelligence, vol 396. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-25691-2_9 • Fox, J., & Gambino, A. (2021). Relationship development with humanoid social robots: Applying interpersonal theories to human–robot interaction. *Cyberpsychology, Behavior, and Social Networking*, *24*(5), 294-299. [Optional readings] • Waytz, A., Cacioppo, J., & Epley, N. (2010). Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism. Perspectives on Psychological Science, 5(3), 219-232. https://doi.org/10.1177/1745691610369336 • Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social Cognition Unbound: Insights Into Anthropomorphism and Dehumanization. Current Directions in Psychological Science, 19(1), 58-62. https://doi.org/10.1177/0963721409359302 • Troshani, I., Rao Hill, S., Sherman, C., & Arthur, D. (2020). Do We Trust in AI? Role of Anthropomorphism and Intelligence. *Journal of Computer Information Systems*, *61*(5), 481–491. https://doi.org/10.1080/08874417.2020.1788473	Submission for weekly reading reflection #2 closes at 5pm on Friday
WEEK 4 DATES: 9/15-9/19	**AI fairness in an unfair world**	• Buolamwini, J. & Gebru, T.. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* in *Proceedings of Machine Learning Research* 81:77-91 Available from https://proceedings.mlr.press/v81/buolamwini18a.html. • Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press. • Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press. • Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, *8*(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902 • Humans are biased too, so why does machine learning bias matter? (8-min video) Thomas, Video series on Ethics and Machine Learning	Submission for weekly reading reflection #3 closes at 5pm on Friday Project Milestone #1 due at 5pm on 9/19
WEEK 5 DATES: 9/22-9/26	**An AI-infused social world**	• Narayanan, A. & Kapoor, S. (2024). Why Can't AI Fix Social Media? (pp. 179-226). Princeton: Princeton University Press. • Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. Journal of Computer-Mediated Communication, 25(1), 89-100. • Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about Invisible Algorithms in News Feeds. *Proceedings of the 33rd Annual ACM Conference on Human*	Submission for weekly reading reflection #4 closes at 5pm on Friday

		Factors in Computing Systems, 153–162. https://doi.org/10.1145/2702123.2702556 • Hannah Mieczkowski, Jeffrey T. Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 17 (April 2021), 14 pages. https://doi.org/10.1145/3449091 • [Recommended] Christin, A., Bernstein, M. S., Hancock, J. T., Jia, C., Mado, M. N., Tsai, J. L., & Xu, C. (2024). Internal Fractures: The Competing Logics of Social Media Platforms. *Social Media + Society*, *10*(3). https://doi.org/10.1177/20563051241274668	
WEEK 6 **DATES:** **9/29-10/3**	**The AI alignment problem**	• Christian, B. (2021) *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton & Company, New York, NY. • Leike, J. & Sutskever, I. Introducing Superalignment. https://openai.com/index/introducing-superalignment/ • Rakowski, R., Kowaliková, P. The political and social contradictions of the human and online environment in the context of artificial intelligence applications. *Humanities and Social Sciences Communications* 11, 289 (2024). https://doi.org/10.1057/s41599-024-02725-y • Khamassi, M., Nahon, M. & Chatila, R. Strong and weak alignment of large language models with human values. *Scientific Reports* 14, 19399 (2024). https://doi.org/10.1038/s41598-024-70031-3 • Garcia, P. Aversion to external feedback suffices to ensure agent alignment. *Sci Rep* 14, 21147 (2024). https://doi.org/10.1038/s41598-024-72072-0 • Collins, K.M., Sucholutsky, I., Bhatt, U. *et al.* Building machines that learn and think with people. *Nature Human Behaviour* 8, 1851–1863 (2024). https://doi.org/10.1038/s41562-024-01991-9 • [Recommended] Sorensen, T., Moore, J., Fisher, J., Gordon, M., Mireshghallah, N., Rytting, C. M., Ye, A., Jiang, L., Lu, X., Dziri, N., Althoff, T., & Choi, Y. (2024). A roadmap to pluralistic alignment. *Proceedings of the 41st International Conference on Machine Learning*, *235*, 46280–46302.	Submission for weekly reading reflection #5 closes at 5pm on Friday
WEEK 7 **DATES:** **10/6-10/10**	**Emerging methods to addressing the alignment problem**	**Participatory and community-centered approach** • Bergman, S., Marchal, N., Mellor, J. *et al.* STELA: a community-centred approach to norm elicitation for AI alignment. *Scientific Reports* 14, 6616 (2024). https://doi.org/10.1038/s41598-024-56648-4 **Seeking public input for value alignment** • Huang, S. *et al.* Collective Constitutional AI: Aligning a Language Model with Public Input. in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* 1395–1417 (Association for Computing Machinery, New York, NY, USA, 2024). doi:10.1145/3630106.3658979. **Understanding AI alignment through experts' lenses** • Li, J. *et al.* User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* 1–18 (Association for Computing Machinery, New York, NY, USA, 2024). doi:10.1145/3613904.3642114. **Algorithmic regulation** • Henderson, P., Mitchell, E., Manning, C., Jurafsky, D. & Finn, C. Self-Destructing Models: Increasing the Costs of Harmful Dual Uses of Foundation Models. in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* 287–296 (Association for Computing Machinery, New York, NY, USA, 2023). doi:10.1145/3600211.3604690. **Large-scale, interactive survey** • Kirk, H. R. et al. (2024) The PRISM Alignment Project: What Participatory, Representative and Individualised Human Feedback Reveals About the	Submission for weekly reading reflection #6 closes at 5pm on Friday Project Milestone #2 due at 5pm on 10/10

		Subjective and Multicultural Alignment of Large Language Models. https://doi.org/10.48550/arXiv.2404.16019. ○ Unveiling the PRISM Alignment Project. https://mlcommons.org/2024/05/prism/	
WEEK 8 DATES: 10/13-10/17	**What does AI optimize for?: Goals, rewards, and reinforcement learning**	• Lambert, et al. (2022) Illustrating Reinforcement Learning from Human Feedback (RLHF). *Hugging Face Blog*. https://huggingface.co/blog/rlhf • Casper, S. et al. (2023) Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. https://doi.org/10.48550/ARXIV.2307.15217. • Stiennon, N. et al. (2020) Learning to summarize from human feedback. in Proceedings of the 34th International Conference on Neural Information Processing Systems 3008–3021 (Curran Associates Inc., Red Hook, NY, USA). • Companion blog post: https://openai.com/index/learning-to-summarize-with-human-feedback/	Submission for weekly reading reflection #7 closes at 5pm on Friday
WEEK 9 DATES: 10/20-10/24	**When does AI fail?: Robustness, adversarial attacks, and red-teaming**	• Overview of Adversarial Machine Learning (2023) Software Engineering Institute, Carnegie Mellon University. Overview video: https://www.youtube.com/watch?v=C8jJ4H6BL1c • O'Sullivan, C. (2024) What Is Adversarial Machine Learning? Types of Attacks & Defenses. https://www.datacamp.com/blog/adversarial-machine-learning • Zou, A. et al. (2023) Universal and Transferable Adversarial Attacks on Aligned Language Models. https://doi.org/10.48550/ARXIV.2307.15043. • Companion blog post: https://llm-attacks.org/	Submission for weekly reading reflection #9 closes at 5pm on Friday In-Class Midterm Exam: 10/24
WEEK 10 DATES: 10/27-10/31	**Why did AI just do that?: Interpretability and explainability of AI behaviors**	• Nussberger, AM., Luo, L., Celis, L.E. *et al.* Public attitudes value interpretability but prioritize accuracy in Artificial Intelligence. *Nature Communications* 13, 5821 (2022). https://doi.org/10.1038/s41467-022-33417-3 • Morrison, K. *et al.* The Impact of Imperfect XAI on Human-AI Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 8, 183:1-183:39 (2024). • Steixner-Kumar, S., Rusch, T., Doshi, P. et al. Humans depart from optimal computational models of interactive decision-making during competition under partial information. Sci Rep 12, 289 (2022). https://doi.org/10.1038/s41598-021-04272-x • Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). Association for Computing Machinery, New York, NY, USA, 2119–2128. https://doi.org/10.1145/1518701.1519023 • Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376590	Submission for weekly reading reflection #9 closes at 5pm on Friday
WEEK 11 DATES: 11/3-11/7	**Human control and ownership in human-AI interaction**	• Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513 • Russell, S. J. (2019). *Human Compatible : Artificial Intelligence and the Problem of Control*. • Greenblatt, R., Shlegeris, B., Sachan, K. & Roger, F. (2024) AI Control: Improving Safety Despite Intentional Subversion. Proceedings of the 41st International Conference on Machine Learning. https://doi.org/10.48550/arXiv.2312.06942 ○ Companion blog post: https://www.alignmentforum.org/posts/d9FJHawgkiMSPjagR/ai-control-improving-safety-despite-intentional-subversion	Submission for weekly reading reflection #10 closes at 5pm on Friday

WEEK 12 DATES: 11/10-11/14	**Ethics and responsible AI**	• Boggust, A., Hoover, B., Satyanarayan, A. & Strobelt, H. Shared Interest: Measuring Human-AI Alignment to Identify Recurring Patterns in Model Behavior. in *CHI Conference on Human Factors in Computing Systems* 1–17 (ACM, New Orleans LA USA, 2022). doi:10.1145/3491102.3501965. • Hendrycks, D. et al. (2023) Aligning AI With Shared Human Values. In *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2008.02275 • Pan, A. et al. (2023) Do the rewards justify the means? measuring trade-offs between rewards and ethical behavior in the MACHIAVELLI benchmark. In *Proceedings of the 40th International Conference on Machine Learning*. vol. 202 26837–26867 (JMLR.org, Honolulu, Hawaii, USA, 2023). https://dl.acm.org/doi/10.5555/3618408.3619525 • Aharoni, E., Fernandes, S., Brady, D.J. et al. Attributions toward artificial agents in a modified Moral Turing Test. Scientific Reports 14, 8458 (2024). https://doi.org/10.1038/s41598-024-58087-7 • Jones, A. (2024) Can we scale human feedback for complex AI tasks? An intro to scalable oversight. AI Safety Fundamentals https://aisafetyfundamentals.com/blog/scalable-oversight-intro/	Submission for weekly reading reflection #11 closes at 5pm on Friday Project Milestone #3 due at 5pm on 11/14
WEEK 13 DATES: 11/17-11/21	**Human-AI interaction in specific domains (1): AI and labors**	• Tyna Eloundou *et al.* (2024) GPTs are GPTs: Labor market impact potential of LLMs. *Science*. **384**, 1306-1308. DOI:10.1126/science.adj0998 • M.R. Frank, D. Autor, J.E. Bessen, E. Brynjolfsson, M. Cebrian, D.J. Deming, M. Feldman, M. Groh, J. Lobo, E. Moro, D. Wang, H. Youn, I. Rahwan, Toward understanding the impact of artificial intelligence on labor, *PNAS*. U.S.A. 116 (14) 6531-6539, https://doi.org/10.1073/pnas.1900949116. • Eloundou, T., Manning, S., Mishkin, P. & Rock, D. (2024) GPTs are GPTs: Labor market impact potential of LLMs. *Science* 384, 1306–1308. https://www.science.org/doi/10.1126/science.adj0998 • Baily, M. N., Brynjolfsson, E., & Korinek, A. (2023) Machines of mind: The case for an AI-powered productivity boom. *The Brookings Institution*. https://www.brookings.edu/articles/machines-of-mind-the-case-for-an-ai-powered-productivity-boom/	Submission for weekly reading reflection #12 closes at 5pm on Friday
WEEK 14 DATES: 11/24-11/28 (11/26-30 THANKSGIVING)	**Human-AI interaction in specific domains (2): AI and healthcare**	• Suh, J., Pendse, S.R., Lewis, R., et al. (2024) Rethinking technology innovation for mental health: framework for multi-sectoral collaboration. *Nature Mental Health* 2, 478–488. https://doi.org/10.1038/s44220-024-00232-2 • Moor, M., Banerjee, O., Abad, Z.S.H. et al. Foundation models for generalist medical artificial intelligence. Nature 616, 259–265 (2023). https://doi.org/10.1038/s41586-023-05881-4 • Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello ai": Uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24. https://doi.org/10.1145/3359206	Submission for weekly reading reflection #13 closes at 5pm on Friday
WEEK 15 DATES: 12/1-12/5	**Future of human-AI interaction through your eyes**	• Zaidan, E., Ibrahim, I.A. AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective. *Humanities and Social Sciences Communications* 11, 1121 (2024). https://doi.org/10.1057/s41599-024-03560-x • Lahusen, C., Maggetti, M. & Slavkovik, M. Trust, trustworthiness and AI governance. *Scientific Reports* 14, 20752 (2024). https://doi.org/10.1038/s41598-024-71761-0	Presentations of course projects during the last class
STUDY DAYS (**DATES: 12/6-12/9**)			
FINAL EXAM PERIOD (**DATES: 12/10-12/17**)			Final project deliverable due at 5pm on Dec 12

Statement on Academic Conduct and Support Systems

Academic Integrity

[This first section on academic integrity is required to be included on all USC syllabi.]
The University of Southern California is foremost a learning community committed to fostering successful scholars and researchers dedicated to the pursuit of knowledge and the transmission of ideas. Academic misconduct is in contrast to the university's mission to educate students through a broad array of first-rank academic, professional, and extracurricular programs and includes any act of dishonesty in the submission of academic work (either in draft or final form).

This course will follow the expectations for academic integrity as stated in the USC Student Handbook. All students are expected to submit assignments that are original work and prepared specifically for the course/section in this academic term. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s). Students suspected of engaging in academic misconduct will be reported to the Office of Academic Integrity.

Other violations of academic misconduct include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

Academic dishonesty has a far-reaching impact and is considered a serious offense against the university. Violations will result in a grade penalty, such as a failing grade on the assignment or in the course, and disciplinary action from the university itself, such as suspension or even expulsion.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment or what information requires citation and/or attribution.

Course Content Distribution and Synchronous Session Recordings Policies

USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation, is prohibited. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in

relation to the class, whether obtained in class, via email, on the internet, or via any other media. Distributing course material without the instructor's permission will be presumed to be an intentional act to facilitate or enable academic dishonestly and is strictly prohibited. (Living our Unifying Values: The USC Student Handbook, page 13).

Statement on University Academic and Support Systems

Students and Disability Accommodations:
USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Student Financial Aid and Satisfactory Academic Progress:
To be eligible for certain kinds of financial aid, students are required to maintain Satisfactory Academic Progress (SAP) toward their degree objectives. Visit the Financial Aid Office webpage for undergraduate- and graduate-level SAP eligibility requirements and the appeals process.

Support Systems:
Annenberg Student Success Fund
The Annenberg Student Success Fund is a donor-funded financial aid account available to USC Annenberg undergraduate and graduate students for non-tuition expenses related to extra- and co-curricular programs and opportunities.

Annenberg Student Emergency Aid Fund
Awards are distributed to students experiencing unforeseen circumstances and emergencies impacting their ability to pay tuition or cover everyday living expenses. These awards are not intended to cover full-tuition expenses, but rather serve as bridge funding to guarantee students' continued enrollment at USC until other resources, such as scholarships or loans, become available. Students are encouraged to provide as much information in their application, as well as contact their academic advisor directly with questions about additional resources available to them.

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - 988 for both calls and text messages – 24/7 on call
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline consists of a national network

of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL) – 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-2500
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - (213) 740-0411
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-2850 or otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.