Instructor

Andre Uhl, PhD
uhl@uchicago.edu

Teaching Assistants

Divyansha Sehgal
divyansha@uchicago.edu

Kayecee Palisoc
kpalisoc@uchicago.edu

Office Hours

TBA

Course Title: The Ethics and Governance of Artificial Intelligence
Course Code: PPHA / KNOW 38850
Quarter: Spring 2024
Meeting Time: Thursdays, 5:00 PM–7:50 PM

Course Overview

This course observes the emergence of AI ethics in law and public policy, examining the norms, values, and political strategies involved in the consensus-building processes that shape the development and governance of AI systems. Students will engage in critical analysis of AI policy documents and delve into core principles such as fairness, accountability, and transparency, exploring their origins and practical applications. The curriculum centers on a series of design thinking workshops that will challenge students to debate over the responsible use of AI in real-world case studies, ranging from issues of human rights to sustainable development and geopolitics. By exercising analytical and rhetorical excellence while engaging with the regulatory complexities inherent in the development of advanced technologies, this course is considered a bootcamp for transformative leaders capable of governing AI for the public good.

Learning Objectives

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

1. Identify and remember core principles, challenges, and developments in AI governance, including essential policy milestones and landmark case studies.
2. Understand the complex interplay of AI development and regulatory practices, comprehending how emerging technologies influence and are influenced by public norms and values.
3. Apply interdisciplinary approaches integrating knowledge from the humanities, social sciences, and public policy to evaluate the social impact of emerging AI applications in real-world scenarios.
4. Analyze emerging AI governance strategies, critically assessing their origins, applications, and effectiveness for responsible AI management and regulation.
5. Exercise leadership and communication skills necessary for advocating ethical decision-making and public interest in AI governance, considering both global impact and specific community needs.
6. Create an original project around a specific challenge that demonstrates critical thinking and innovation in promoting the strategic use of AI for the public good.

Prerequisites

While this course does not have specific prerequisites, it is designed for students who are comfortable engaging in an active learning environment that will include interactive workshops, student debates, and collaborative group projects. Students will be expected to speak frequently in class and navigate the social dynamics of multi-stakeholder negotiations.

Course Schedule and Readings

This is a list of both required and recommended readings to stimulate our discussions. Students are expected to complete required readings ahead of each course meeting, and be prepared to share their thoughts, insights, and questions with the group. All required readings will be provided as digital copies on Canvas—no book purchase is necessary to successfully participate in class.

Week 1: INTEGRITY

Readings
- Joler, Vladan, and Matteo Pasquinelli. "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism." *AI & Society* 36, no. 4 (2021): 1263–1280.

Key Topics
The challenges of evaluating AI-augmented knowledge; the value of intellectual property in the age of knowledge extractivism; and the relevance of academic traditions in emerging knowledge landscapes.

Week 2: TRANSPARENCY

Readings
Required:

- O'Neil, Cathy. "Bomb Parts: What is a Model?" and "Shell Shocked: My Journey of Disillusionment." In *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* 15-50. New York, NY: Crown, 2016.

Recommended:

- Seaver, Nick. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems." *Big Data & Society* 4, no. 2 (2017). https://doi.org/10.1177/2053951717738104
- Eubanks, Virginia. "Automating Eligibility in the Heartland." In *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor,* 39-84. New York, NY: St. Martin's Press, 2018.

Key Topics
The use of technology for professional judgment in critical decision-making processes; the potential for deceptive outputs of machine learning algorithms; and the essential role of human oversight in mitigating the limitations of AI.

Week 3: ACCOUNTABILITY

Readings
Required:

- Elish, Madeleine Clare. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction." *Engaging Science, Technology, and Society* 5 (2019): 40-60.
- Awad, E., Dsouza, S., Kim, R. *et al.* "The Moral Machine Experiment." *Nature* 563 (2018): 59–64.

Recommended:

- Mosqueira-Rey, E., E. Hernández-Pereira, D. Alonso-Ríos, et al. "Human-in-the-Loop Machine Learning: A State of the Art." *Artificial Intelligence Review* 56 (2023): 3005–3054.

Key Topics
The ethical challenges of programming moral decisions into machines; the principles that should guide such decisions; and the legal and ethical frameworks for accountability when those decisions lead to harm.

Week 4: FAIRNESS

Readings   Required:

- Kantayya, Shalini, and Sabine Hoffman. *Coded Bias*. United States: 7th Empire Media, 2021.

Recommended:

- Buolamwini, Joy and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018): 77–91.
- Uliasz, Rebecca. "Seeing Like an Algorithm: Operative Images and Emergent Subjects," *AI & Society* 36 (2021): 1233–1241.

Key Topics   The trade-offs between technological efficiency and fairness; the ethical obligations to prevent bias in technological deployments; and the broader societal impacts of relying on flawed systems.

Week 5: CONTROL

Readings   Required:

- Nissenbaum. Helen Fay. "Information Technology's Power and Threat." In *Privacy in Context: Technology, Policy, and the Integrity of Social Life,* 19-64. Stanford, CA: Stanford Law Books, 2010.

Recommended:

- Pasquale, Frank. "Digital Reputation in an Era of Runaway Data." In *The Black Box Society: The Secret Algorithms That Control Money and Information,* 19-59. Cambridge, MA: Harvard University Press, 2015.
- Zuboff, Shoshana. "The Foundations of Surveillance Capitalism, Chapter Two." In *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power,* 25-62. New York, NY: Public Affairs, 2019.

Key Topics   The balance between using technology for collective governance and the potential risks to personal liberty; the impact of algorithmic bias on such systems; and the meaning of user-centric data agency.

Week 6: SUSTAINABILITY

Readings   Required:

- Comber, Rob, and Elina Eriksson. "Computing as Ecocide." *Proceedings of the Ninth Computing within Limits Conference*, Association for Computing Machinery (2023). https://doi.org/10.21428/bf6fb269.9fcdd0c0
- Crawford, Kate, and Vladan Joler. "Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data, and Planetary Resources." AI Now Institute and Share Lab (2018). https://anatomyof.ai/

Recommended:

- Solomon, L., and C. Baio. "An Argument for an Ecosystemic AI: Articulating Connections across Prehuman and Posthuman Intelligences," *International Journal of Communication, Wellbeing and Behaviour* 3 (2020): 559–584.

Key Topics   The environmental costs and benefits of deploying AI systems; and the ethical implications in aligning them with sustainable development goals.

[Week 7: Policy Memo Ideation and Feedback Session]

Week 8: DEMOCRACY

Readings   Required:

- Orlowski, Jeff. *The Social Dilemma.* United States: Exposure Labs, 2020.

Recommended:

- Costanza-Chock, Sasha. "Design Sites: Hackerspaces, Fablabs, Hackathons, and DiscoTechs." *Design Justice: Community-led Practices to Build the Worlds We Need,* 135-172. Cambridge, MA: The MIT Press, 2020.
- Friedman, Batya, and David G. Henry. "Applications." In *Value Sensitive Design: Shaping Technology with Moral Imagination,* 105-167. Cambridge, MA: MIT Press, 2019.

Key Topics   The emergence of AI-based communication systems; the responsibilities of tech companies vs. governmental bodies in regulating content; and the potential impacts on public opinion and values.

Week 9: INTEROPERABILITY

Readings   Required:

- Lee, Kai-Fu. "The Four Waves of AI." In *AI Superpowers: China, Silicon Valley, and the New World Order,* 104-140. Boston: Houghton Mifflin Harcourt, 2018.

Recommended:

- Gal, Danit. "Perspectives and Approaches in AI Ethics: East Asia." In Markus D. Dubber, Frank Pasquale, and Sunit Das (eds). *The Oxford Handbook of Ethics of AI*, 607-625. New York: Oxford University Press, 2020.
- Birhane, Abeba. "Algorithmic Colonization of Africa." In Stephen Cave, and Kanta Dihal (eds). *Imagining AI: How the World Sees Intelligent Machines,* 247-260. New York: Oxford University Press, 2023.

Key Topics   The global context of AI governance; the feasibility of harmonizing policies between nations with fundamentally different norms and values; and the potential for creating standards that respect both technological progress and cultural diplomacy.

Assignments

Attendance, Participation, and Self-Reflection (20%)

Description

Students are expected to demonstrate active listening, constructive peer feedback, and collaborative problem-solving in navigating the course's interactive group exercises, thereby contributing to an active learning environment that encourages constructive discourse in the face of controversy. Through instructor observation, regular self- and peer-evaluations, and occasional homework assignments, students will be assessed for their adaptability, empathy, and self-awareness necessary to effectively engage in debates where varied views, values, and beliefs can be openly shared and critically examined.

Deliverables

Attend each session and complete a self- and peer-evaluation provided via Canvas after each class, sharing your reflections on course experiences, behaviors, and developments.

AI Policy Document Analysis (Group Project, 40%)

Description

Students will work in small groups and select an AI policy document from a provided list including governmental regulations, industry guidelines, and ethics codes from professional organizations. The task is to conduct a comprehensive analysis of the chosen document evaluating its approach to AI governance, including its stated goals, the challenges it addresses, and its potential impact on AI development and deployment. The analysis should also include a critical assessment of the policy's strengths and weaknesses, considering both theoretical perspectives and practical implications. The culmination of this assignment is a group presentation, where students will share their findings and engage in a Q&A session with the class.

Deliverables

(1) deliver a group presentation in class (20 minutes + Q&A);
(2) submit a co-authored white paper (6-8 pages + appendix, double-spaced) via Canvas by the end of the week following the presentation.

AI Policy Memo (Individual Project, 40%)

Description

Identify a public policy challenge related to the social impact of AI and develop a policy memo that outlines an original and feasible recommendation for addressing the challenge, drawing on both ethical and technical considerations. The memo should articulate the problem clearly, review existing policies or measures (if any), and detail the proposed recommendation, including its goals, anticipated impact, and implementation considerations. The memo is expected to be concise, persuasive, and well-researched, demonstrating the student's ability to think critically and innovatively about AI governance.

Deliverables

(1) complete a draft of the policy memo and participate in the ideation and feedback session on May 2;
(2) submit the final version (2-3 pages + appendix, double-spaced) via Canvas by May 19, 11:59 pm.

Course Policies

Attendance and Coursework

Active and consistent participation in person is crucial to the success of your learning experience in this course. Given the course's interactive exercises, we are unable to accommodate remote participation. It is also imperative to submit assignments promptly to ensure timely feedback. Late submissions will incur a grade reduction of one-third of a letter grade (e.g., from A to A-) for each day they are overdue. Similarly, unexcused absences will negatively affect your grade in the "Attendance, Participation, and Self-Reflection" assignment in an equivalent manner. Should you encounter an emergency or circumstance that hinders your ability to attend a session or meet a deadline, please contact the course's TA team in advance.

Technology

Technology, including generative AI, is an essential aid to our classroom learning environment. You are encouraged to use a laptop, smart phone, or other device to explore concepts related to course discussions and in-class activities. For content co-created with AI tools, you are required to follow the new Chicago Manual of Style citation guideline for AI tools to support the University of Chicago's Academic Honesty and Plagiarism policy. Keep in mind, however, that the success of this class is also dependent on your presence and active contributions. This means that all students are expected to actively listen to one another and participate in classroom activities. Please refrain from using technology for any other purposes than the course's learning objectives during our meetings.

Tolerance and Respect

This course is committed to honoring diversity in all its forms and across all dimensions of identity, social location, and experience. Please be prepared to listen with interest, patience, and resilience to views, values, beliefs, and styles of expression that may be different from your own, and support others to share their insights and understandings of reality freely and generously in the specific terms, idioms, and modalities familiar to them in the environments in which they feel at home.

Accessibility

Students with disabilities who have been approved for the use of academic accommodations by Student Disability Services (SDS) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by SDS for using accommodations. Timely notifications are required to ensure that your accommodations can be implemented. Please meet with me to discuss your access needs in this class after you have completed the SDS procedures for requesting accommodations. For more information, visit disabilities.uchicago.edu.