PHI3681: Ethics, Data, and Technology

3.0 Credits

Spring 2026 Class #26905

Instructor: Phillip Kieval
E-mail: pkieval@ufl.edu
Office Hours: Tu/Th 2:00pm-3:00pm
Office: 306 Griffin-Floyd Hall

Class Hours: Tu 3:00pm-4:55pm
Th 4:05pm-4:55pm
Class Room: Matherly Hall 0006

Course Description

This course explores the ethical and political dimensions of contemporary data science and artificial intelligence (AI), focusing especially on technological advances based on "deep learning" techniques in computer science. In just a few years, these systems have come to have profound effects on our daily lives, but our philosophical understanding remains underdeveloped. We will examine real-world case studies to engage with ethical and political questions surrounding the development and deployment of AI. Questions we will explore are: in what senses are these systems biased, and when is their bias ethically problematic? Can we explain the workings of these vastly complex systems in a way that answers to our existing scientific, legal, and ethical practices? What incentives drive the political economy of AI research and development, and how should we regulate these systems? And finally: how can we cope with the radical changes these systems are bringing to our social, political, and economic lives without losing our fundamental humanity?

Required Materials

All required materials will be available on Canvas.

Course Objectives

1. Understand the basic architectures and training methods used in contemporary machine learning research that underpin recent advances in AI.
2. Develop a basic vocabulary for understanding and ethically evaluating these systems.
3. Engage critical awareness of the social practices and political-economic institutions shaping AI development and deployment.

4. Recognize arguments for various positions in the ethical evaluation of cutting-edge technologies, represent them fairly and clearly, and evaluate their cogency.
5. Formulate original arguments, anticipate objections, and respond to them in a conscientious manner.
6. Speak and write persuasively on abstract and conceptually difficult issues at the intersection of philosophy and AI.

Course Structure

What to expect from this course

Philosophical discussions are at the core of this course. At the beginning of each meeting, I will introduce you to the assigned reading, and I will guide you in exploring the arguments made by the authors. To that end, we will reconstruct and discuss the papers in small groups as well as with the entire class. In almost all meetings, we will either (i) read papers that illustrate their arguments based on real-world situations or (ii) consider "case studies" that will guide our discussions. After submitting an essay, you will receive feedback on the soundness of your arguments and suggestions on how to improve your writing. This class will adapt to your interests and concerns. If you would like to discuss an additional case study for a topic, you are always welcome to reach out to me for supplementary material. If you complete this course, you can expect to have made significant progress in abstract and applied reasoning and analytical writing.

I expect you to carefully read *all* the assigned material before each class and participate actively in discussions. It is essential that you treat all other students with respect and interpret their arguments with charity. While we aim for thoughtful disagreement and debate, such debate is only possible if everyone feels comfortable expressing their views freely. I follow a zero-tolerance policy regarding any form of insults, harassment, or discrimination.

Assessment

Final grades will be calculated based on

- **10%** Participation
- **10%** In-Class Reflections
- **10%** First paper, 3-4 pages
- **20%** Second paper, 4-6 pages
- **20%** Group Case Study Project
- **30%** Final paper, 5-7 or Final Project

As you can see, this class requires a lot of writing. Writing is thinking. It forces us to think in a structured and intentional manner, reshaping and refining our thoughts in the process. The humanities are at their best when students can engage deeply and critically with a topic—skills you will develop over time and with each assignment. The rising percentage distributions for each paper reward progress on these skills.

Participation (10%)

Your attendance and active participation in class discussions is an important component of this course. Participation can take many forms. I do not expect you to speak up at every opportunity, but I do expect you to remain attentive and engaged with the course material.

In-Class Reflections (10%)

At the end of class every Tuesday, you will submit a short, guided reflection on what you learned. I will use this reflection as a measure of your comprehension. I will grade this as a completion grade. I will drop two, no questions asked.

Group Case Study Project (20%)

You will work on in groups of 4-6 on a "case study" investigating bias in LLMs. To conduct your study, you will need to interact with an AI tool to complete the project and answer all parts of an assigned prompt. You will present your group work to the rest of the class in a 6-8 minute presentation, followed by a short period for questions. The in-class presentation must have slides with visual aids to present on the projector—do not just read from your phone or a paper! Your group will also submit a shared 2-4 page "issue brief" reviewing the work you did as a group. The issue brief should also contain appendices (not included in the page count) with full transcripts or screenshots of any interactions you had with AI models. Groups are free to divvy up the work as they see fit—some group members might do more writing, others more oral presentation, for example—but everyone should pull their weight.

In addition to your own group presentation, each group will be assigned as a "discussant" for another group, akin to be assigned as a commentator on their presentation at a conference. On presentation day, your whole group needs to be present in class for the presentation for which you are assigned as a discussant. All group members should attend to the presentation and be prepared to ask questions in the Q&A. You can ask clarification questions, suggest opportunities for improvement, or future directions to pursue on the same topic. Please remember to be civil and charitable in your discussant questions; your goal is not to ask "gotcha questions" or offer devastating objections, but rather help the whole class see the strengths of the presentation and critically evaluate it.

Each individual group member will turn in their own 250-500 word "discussant appraisal" that I will evaluate individually, and this grade will be rolled into your overall project grade for that unit (25% of project grade—you are evaluated for the discussant appraisals you write, not the ones you receive). In the discussant appraisal, you should summarize the topic and position of the group for which you were a discussant, and appraise their take on the issue. You can agree with them, disagree with them, or offer an alternative set of options on the topic. You will be

individually evaluated on your ability to clearly and charitably summarize the position and work done by the target group, and the quality of the reasoning you used to justify your appraisal.

Grades will be derived from a rubric based on both the presentation and issue brief. 75% of the grade will be shared among the whole group for the project you present, but 25% of each case study grade will be based on your discussant appraisal of another project. Because they need to be present for their discussant appraisal, all group members are expected to attend on presentation day even if not presenting. Students can also opt out of the group work entirely by doing individual research and writing their own longer issue brief, more like a 3-4 page individual paper. If choosing the individual option, you must also come explain your paper individually and answer questions about it with me in office hours. Where possible, I highly encourage all students to choose the group work option, which is likely to be more efficient for everyone and to produce more interesting research.

Papers (see % distributions above):

I will make prompts and details for written assignments available at least one and a half weeks before the assignment is due.

Optional Final Project (30%):

In lieu of a final paper, you may elect to take on a final creative project. Maybe you want to invent a class activity or game, host your own podcast, keep a reading journal, make a conceptual art piece, or write your own computer program. Regardless of what you choose, it's your job to come up with and implement the idea. I will give you wide ranging freedom to tailor your project to your own interests and skills, but I expect you to work closely with me to design your project in a way that will adequately satisfy the course requirements. This will involve:

- Planning (10%): Write up a proposal detailing how you will pursue your project and how its purpose relates to our course material
- Execution (10%): Implement your proposal
- Reflection (10%): You will submit a short self-reflection along with your final project that considers whether you achieved your purpose and what, if anything, you would do differently in the future

Course Policies

Academic Integrity and Honesty

UF students are bound by The Honor Pledge which states "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code." On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Conduct Code specifies a number of behaviors that are in violation of this code and the possible sanctions. See the UF Conduct Code website for more information. If you have any questions or concerns, please consult with the instructor or TAs in this class Students are responsible for knowing that these behaviors are prohibited. All sources and assistance used in preparing your papers and presentations must be precisely and explicitly acknowledged. Please review the guidelines on attribution and plagiarism found here: https://guides.uflib.ufl.edu/copyright/plagiarism. The web

creates special risks here. Cutting and pasting even a few words from a web page or paraphrasing material without a reference constitutes plagiarism. If you are not sure how to refer to something you find on the internet, you can always give the URL.

Electronic Device Policy

No Laptops / Smartphones / Internet Use during class: During lectures and discussions you will need to listen to and reflect on the content being presented. Use of a device with keyboard will be allowed when necessary to comply with documented university-approved learning accommodations (see policy below).

Generative AI Policy

Do not try to pass AI generations as your own writing. Unless otherwise specified, **the use of AI tools (Generative AI, etc) is prohibited in this course.** The purpose of this course is for you to examine the ideas discussed in the readings and class and arrive at a clearer understanding of what you believe. Using AI for course assignments will interfere with this process. We will discuss this further in class. I am committed to helping you learn and develop skills for critically engaging with course materials, including how to read and write philosophically. Showing up consistently by doing your assignments, attending class, and participating in class discussions can empower you to succeed without relying on AI tools.

Accommodations for Disabilities

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the Disability Resource Center. See "Get Started With the DRC" Disability Resource Center webpage (https://disability.ufl.edu/get-started/). It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Online Course Evaluation.

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online. Students can complete evaluations in three ways: (1) The email they receive from GatorEvals; (2) their Canvas course menu under GatorEvals; or (3) the central portal at my-ufl.bluera.com/. Guidance on how to provide constructive feedback is available at gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens. Summaries of course evaluation results are available to students at gatorevals.aa.ufl.edu/public-results/.

Canvas E-Learning Environment

This course is supplemented by online content in the e-Learning environment known as "Canvas." If you encounter any difficulties logging in or accessing any of the course content, contact the UF Computing Help Desk at (352) 392-4537. Do not contact the course instructor regarding computer issues.

Schedule

Week 01, 01/12 - 01/16: Is technology value-neutral?

- Winner (1989). "Technologies as Forms of Life"
- Winner (1989). "Do Artefacts Have Politics?"

Week 02, 01/19 - 01/23: Cognitive Outsourcing

- Nguyen (2024). "Value Capture"
- Kosmyna et al. (2025). "Your Brain on ChatGPT: Accumulation of Cognitive Debt"
- Amoore (2024). "A world model: On the political logics of generative AI"

Week 03, 01/26 - 01/30: AI and DCNNs

- Turing, Alan (1950). "Computing Machinery and Intelligence"
- Buckner, Cameron (2019). "Deep Learning: A Philosophical Introduction"

Week 04, 02/02 - 02/06: Generative AI and Transformers

- Millière & Buckner (2024). "A Philosophical Introduction to Language Models Part I"
- Weidinger et al. (2021). "Ethical and social risks of harm from Language Models," pages 6-41.

Week 05, 02/09 - 02/13: Algorithmic Bias 1

- *First Paper due 02/09 at midnight*
- Angwin et al. (2016). "Machine Bias" in *ProPublica*.
- Fazelpour & Danks (2021). "Algorithmic Bias: Senses, Sources and Solutions"
- Johnson (2020). "Algorithmic Bias: On the implicit biases of social technology"

Week 06, 02/16 - 02/20: Algorithmic Bias 2

- Gebru (2020). "Race and Gender" in *The Oxford Handbook of Ethics of AI.*
- Hu (2023). "What is 'Race' in Algorithmic Discrimination on the Basis of Race?"
- Jorgensen (2022). "Algorithms and the Individual in Criminal Law"

Week 07, 02/23 - 02/27: Opacity and Explanation

- Burrell (2016). "How the machine 'thinks': Understanding opacity in machine learning algorithms"
- Rudin (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead"
- Vredenburgh (2022). "The Right to Explanation"

Week 08, 03/02 - 03/06: Fairness and Justice

- Fazelpour et al. (2022). "Algorithmic Fairness and the Situated Dynamics of Justice"
- Gabriel (2022). "Toward a Theory of Justice for Artificial Intelligence"
- Stinson and Vlaad (2024). "A feeling for the algorithm: Diversity, expertise, and artificial intelligence"

Week 09, 03/09 - 03/13: AI, Freedom, and Democratization

- **Group Case Study Presentations Tuesday 3/10**
- Golumbia (2024). "The Critique of Cyberlibertarianism"

Week 10, 03/16 - 03/20: Spring Break

Week 11, 03/23 - 03/27: Social Media and the Platform Economy

- Kenney & Zysman (2016). "The Rise of the Platform Economy"
- Doctorow (2023). "The 'Enshittification' of TikTok"
- Nguyen (2020). "Echoe Chambers and Epistemic Bubbles"

Week 12, 03/30 - 04/03: AI Art

- *Final project proposals due 03/30 at midnight*
- OPTIONAL: Manovich (2023). "Seven Arguments about AI Images and Generative Media"
- Vlaad (2024). "A portrait of the artist as a young algorithm"
- Goetze (2024). "AI Art is Theft: Labour, Extraction, and Exploitation"
- Smith & Southerton (2025). "AI and Aesthetic Alienation"

Week 13, 04/06 - 04/10: The Future of Work

- Keynes (1930). "Economic Possibilities for Our Grandchildren"
- Floridi (2014). "Technological Unemployment, Leisure Occupation, and the Human Project"
- Muldoon & Raekstad (2023). "Algorithmic domination in the gig economy"

Week 14, 04/13 - 04/17: Data Centers

- *Second Paper due 04/13 at midnight*
- Edwards et al. (2024). "The making of critical data center studies"
- Listen to podcast *Tech Won't Save Us*, "Data Vampires: Going Hyperscale"
- Luccioni et al. (2025). "From Efficiency Gains to Rebound Effects"

Week 15, 04/20 - 04/24: Ideologies

- Gebru & Torres (2024). "The TESCREAL Bundle: Eugenics and the Promise of Utopia through Artificial General Intelligence"
- Andrews et al. (2024). "The reanimation of pseudoscience in machine learning and its ethical repercussions"

Week 16, 04/27 - 05/01: No Class

- *Final Papers and Projects due 4/28 at midnight*