PHIL ####—AI ETHICS

AI ETHICS

PROFESSOR: Dr. Blake Hereth ("Dr. H")
PRONOUNS: they/them
EMAIL: Blake_Hereth@uml.edu
OFFICE: Dugan Hall 200L

CLASS MEETINGS: MWF 12:00-1:00pm in Dugan Hall 101
E-OFFICE: Collaborate Ultra (Blackboard)
E-STUDENT HOURS: Tuesday/Thursday 12:00-1:00pm and by appointment

COURSE DESCRIPTION:

Artificial intelligence (AI) is all around us, embedded in every aspect of our lives. We rely on it for decision-making: for driving directions, to remind us of meetings and birthdays, for smart shopping, and to predict wars and criminal behavior. However, the risks of uncritical reliance on AI are numerous. AIs, like their human creators, are shot through with racial, gender, and other biases. Their presence in the economy risks mass unemployment for human laborers. They may have personhood, in which case our use of autonomous weaponry and sex robots becomes deeply suspect. It is therefore of immense moral importance that we anticipate the big moral questions surrounding AI – and that we arrive at correct answers to those questions.

COURSE GOALS:

By the end of the course, students should be able to:

- understand and execute basic logical operations;
- identify and speak intelligently about major issues in AI ethics;
- demonstrate a strong empirical understanding of existing forms of AI; and
- write a philosophically rigorous, empirically informed, and original AI ethics paper that contributes to an existing literature.

ESSENTIAL LEARNING OUTCOMES:

- Social Responsibility and Ethics (SRE): This course meets the Core Curriculum Social Responsibility and Ethics Essential Learning Outcome; it provides students the opportunity to reason about right and wrong conduct, to assess moral beliefs and practices, and to apply that knowledge to make a positive difference in the community and the world.

COURSE REQUIREMENTS:

- Participation (20%): Class will be held face-to-face in Dugan Hall 101. Watch/read/listen to the required content carefully and come prepared to discuss it. Then, when in class, discuss it. You won't receive credit just for being present.

- Group Presentations (20%): In groups of 3-5, students will give a class presentation on an assigned reading. Depending on the size of the class, you will be presenting either alone or in a group. Each presentation should provide a summary of the paper, a reconstruction of the paper's central argument, and questions for class discussion.
- Draft Paper (20%): Each student should write a 2,000-word draft paper (exclusive of notes and bibliography) where they make an original, creative argument on a relevant course topic. (See the Blackboard rubric for further details.) The Draft Paper is due October 31, 2022, at 5pm on Blackboard.
- Final Paper (40%): Each student should write a 4,000-word paper (exclusive of notes and bibliography) in which they revise their Draft Paper in light of my feedback. The idea is to polish, and add to, the Draft Paper. The Final Paper is due December 16, 2022, at 5pm on Blackboard.

GRADING SCALE:

I use a standard grading scale for this course:

A = 90-100%
B = 80-89.99%
C = 70-79.99%
D = 60-69.99%
F = 59.99 or below

TEXTBOOK(S):

- *The Ethics of Artificial Intelligence*, edited by S. Matthew Liao. NY: Oxford University Press, 2020. ISBN: 0190905042.
- All other readings will be made available via Blackboard.

SCHEDULE:

(NOTE: THIS IS A *TENTATIVE* SCHEDULE. I MAY ADJUST IT DEPENDING ON NEED OR CLASS PROGRESS.)

WEEK 1 (SEP 2): COURSE INTRODUCTION

Readings:

- Syllabus

WEEK 2 (SEP 5, 7, 9): WHAT IS AI ETHICS?

Readings:

- Future of Life Institute, "Benefits & Risks of Artificial Intelligence"
- Nick Bostrom and Eliezer Yudkowsky, "The Ethics of Artificial Intelligence"

WEEK 3 (SEP 12, 14, 16): THE VALUE ALIGNMENT PROBLEM

Readings:

- **Watch:** Nick Bostrom, "What Happens When Our Computers Get Smarter Than We Are?" (TED talk)
- Iason Gabriel, "Artificial Intelligence, Values, and Alignment"
- Shakir Mohamed, Marie-Therese Png, and William Isaac, "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence"

WEEK 4 (SEP 19, 21, 23): BIASED AIS

Readings:

- **Watch:** Cathy O'Neill, "The Era of Blind Faith in Big Data Must End" (TED talk)
- Anupam Chander, "The Racist Algorithm?"
- Ayanna Howard and Jason Borenstein, "The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity"

WEEK 5 (SEP 26, 28, 30): BIASED AIS

Readings:

- Linus Ta-Lun Huang, Hsiang-Yun Chen, Ying-Tung Lin, *et al*, "Ameliorating Algorithmic Bias, or Why Explainable AI Needs Feminist Philosophy"
- Nicholas Tilmes, "Disability, Fairness, and Algorithmic Bias in AI Recruitment"
- Ying-Tung Lin, Tzu-Wei Hung, and Linus Ta-Hun Huang, "Engineering Equity: How AI Can Help Reduce the Harm of Implicit Bias"

WEEK 6 (OCT 3, 5, 7): AUTONOMOUS WEAPONS

Readings:

- Robert Sparrow, "Killer Robots"
- Garry Young, "Should Autonomous Weapons Need a Reason to Kill?"
- Ryan Tonkens, "Should Autonomous Robots Be Pacifists?"

WEEK 7 (OCT 10, 11, 12, 14): SEX ROBOTS

Readings:

- Kate Devlin, "The Ethics of the Artificial Lover"
- Nancy S. Jecker, "Nothing to Be Ashamed Of: Sex Robots for Older Adults with Disabilities"
- John-Stewart Gordon and Sven Nyholm, "Kantianism and the Problem of Child Sex Robots"

WEEK 8 (OCT 17, 19, 21): BRAIN DRAIN

Readings:

- Marie Oldfield, "Will AI Take Away Your Job?"
- Aaron James, "Planning for Mass Unemployment: Precautionary Basic Income"
- Tom Parr, "Automation, Unemployment, and Insurance"

WEEK 9 (OCT 24, 26, 28): MORAL ROBOTS

Readings:

- Andrea Loreggia, Nicholas Mattei, Francesca Rossi, and K. Brent Venable, "Modeling and Reasoning with Preferences and Ethical Priorities in AI Systems"
- Ugo Pagallo, "When Morals Ain't Enough: Robots, Ethics, and the Rules of the Law"
- Raul Hakli and Pekka Mäkelä, "Moral Responsibilioty of Robots and Hybrid Agents"

WEEK 10 (OCT 31, NOV 2, 4): ROBOT RIGHTS

Readings:

- **Draft Paper due Monday, October 31, at 5pm**

- S. Matthew Liao, "The Moral Status and Rights of Artificial Intelligence"
- John Danaher, "Welcoming Robots into the Moral Circle: A Defense of Ethical Behaviorism"
- Bartek Chomanski, "If Robots Are People, Can They Be Made for Profit? Commercial Implications of Robot Personhood"

WEEK 11 (NOV 7, 9, 11): SELF-DRIVING CARS

Readings:

- Frances M. Kamm, "The Use and Abuse of the Trolley Problem: Self-Driving Cars, Medical Treatments, and the Distribution of Harm"
- Nassim Jafari Naimi, "Our Bodies in the Trolley's Path, or Why Self-Driving Cars Must *Not* Be Programmed to Kill"
- Johannes Himmelreich, "No Wheel but a Dial: Why and How Passengers in Self-Driving Cars Should Decide How Their Car Drives"

WEEK 12 (NOV 14, 16, 18): CONSERVATION

Readings:

- Peter Singer, "AI Ethics: The Case for Including Animals"
- Irene Nandutu, Marcellin Atemkeng, and Patrice Okouma, "Integrating AI Ethics in Wildlife Conservation AI Systems in South Africa: A Review, Challenges, and Future Research Agenda"

WEEK 13 (NOV 21, 23, 25): CONSERVATION + THANKSGIVING!

Readings:

- TBD
- **No class Wednesday, November 23rd, or Friday, November 25th (Thanksgiving)**

WEEK 14 (NOV 28, 30, DEC 2): AI IN THE CLINIC

Readings:

- Joanna Demaree-Cotton, Brian D. Earp, and Julian Savulescu, "How to Use AI Ethically for Ethical Decision-Making"
- Alice Parfett, Stuart Townley, and Kristofer Allerfeldt, "AI-Based Healthcare: A New Dawn or Apartheid Revisited?"
- Sally Dalton-Brown, "The Ethics of Medical AI and the Physician-Patient Relationship"

WEEK 15 (DEC 5, 7, 9): PREDICTIVE AI

Readings:

- Catherine Greene, "AI and the Social Sciences: Why All Variables Are Not Created Equal"
- Duncan Purves and Jeremy Davis, "Public Trust, Institutional Legitimacy, and the Use of Algorithms in Criminal Justice"
- Tzu-Wei Hung and Chun-Ping Yen, "On the Person-Based Predictive Policing of AI"

WEEK 16 (DEC 12, 14, 16): FINALS WEEK

- **Final Paper due Friday, December 16, at 5pm**