PHI 3625, Honors AI Ethics

Dr. Mark Tunick
3 credits
No prerequisites
Fall 2020

Meets TR 1230-1350 in HC 115
Prof. Mark Tunick, HC 104
Office Hours: TW 10-12, 2-3
Contact: tunick@fau.edu; (561) 799-8670

Description: This course offers an interdisciplinary approach to ethical issues raised by artificial intelligence (AI) including liability issues of self-driving cars/drones, economic impacts of AI, ethical concerns with predictive analytics using big data, and the ethical status of robots.
Rapid developments in artificial intelligence (AI) are raising numerous ethical issues. For example, how should one go about programming a machine to be ethical? What are the economic consequences of AI: will robots replace humans, leading to massive unemployment? Will they become so intelligent they will threaten humanity? Who should be morally and legally responsible if a self-driving car or military drone causes an accidental death? Is it moral to sell or use sex robots? Should robots be caretakers of the young or old? Because AI is used to analyze and mine 'big data' we will also address ethical issues associated with big data, such as ethical concerns about the use of predictive analytics in health care, and the creation of filter bubbles that shape what news we receive on the internet. Finally, we will address a broad set of issues concerning whether robots could ever be persons with rights. Doing so requires us to consider what it is to be a person, or a human being, or to have dignity, or intelligence, or free will. This course will take an interdisciplinary approach to these issues, drawing on science fiction, films, philosophy, political theory, economics, and works by scholars in the health care and computer science fields who explore the implications of AI technology. There are no prerequisites.

Course Objectives:

- To develop an understanding of AI capacities, the ethical issues it raises, and critical approaches to thinking about these issues.
- To understand the distinction between humans, persons, and various forms of non-humans including robots and the ethical implications of these distinctions.
- To develop the ability to read complex ethical arguments independently.
- To think critically and write clearly and with precision.

Note on Honors Distinction: This Honors course differs substantially from a non-Honors course in the amount of work expected from students: it is discussion based, and draws on some challenging theoretical and philosophical primary works from a variety of disciplines. In addition, standards and expectations for written expression, including editing and revision, will be demanding.

Requirements: Class will be discussion-based. It is important for students to come to class prepared to discuss the reading scheduled for that meeting. Grading is based on quizzes/discussion boards (20%), one short (1-2 page) paper (5%), two papers of 6-8 pages (40% total), class participation (20%), and a group project presentation (15%). Because this is a discussion-based course, attendance is important and so the participation grade will be reduced by 2 points (out of a possible 100 points) for each unexcused absence beyond 2. The default grading scale is 94-100 (A), 90<94 (A-), 87<90 (B+), 84<87 (B), 80<84 (B-), 77<80 (C+), 74<77 (C), 70<74 (C-), 67<70 (D+), 64<67 (D), 61<64 (D-), <61 (F). I may adjust the grading scale for each assignment. A would be 93-100 instead of 94-100, and so forth. For the final grade—so that every student was treated the same—based on whether it seems in my professional judgment that the quizzes/paper topics were too hard, compared to grade distributions in my other courses in the past. In no case would the adjustment yield a lower grade for any student.

Reading: Cathy O'Neil, Weapons of Math Destruction (2017); Isaac Asimov, I Robot and The Bicentennial Man; all other material is available in Canvas. Reading listed under each class is to be done prior to that class meeting.
Films: *2001*; *Her*; *War for the Planet of the Apes*.
Canvas: This course makes use of Canvas: each topic has its own Module with background information and links to the reading, quizzes to help you understand the material, and in some cases discussion boards. Be sure to check Canvas regularly.
Schedule/Topics
I. Weeks 1-2: Introduction to AI and its capabilities.
This course presupposes no knowledge of computer programming, but we begin with some introductory material on A.I.
Week 1: A.I. and its uses and potential
Rdg: Russell and Norvig, Artificial Intelligence, pp. 26-7 and Ex. 1.5; ch. 2 and Ex. 2.4; ch. 3; Brendan Lake, Joshua Tenenbaum et.al., "Building Machines that Learn and Think Like People," *Behavioral and Brain Sciences* 40 (2017); Amanda Schaffer, "Boosting," *MIT News* July/August (2018), 12-16; Fei Fei Li, "Teaching Computers to Understand Pictures" (Ted Talk, 2015, 17:59): online; "Using AI to manage employees and make them more productive or efficient," *New York Times* 6/24/19, online
For those interested: Neudert, "Teaching propaganda how to talk," *MIT Technology Review* 121(5):72-3 (2018).
Week 2: Will AI surpass humans and pose a threat?
Rdg: Nick Bostrom, "Are we Living in a Computer Simulation?", *Philosophical Quarterly* 53(211):243-55 (2003); Khatchadourian, "The Doomsday Machine," *The New Yorker*, Nov. 23, 2015; David Chalmers, "The Singularity: A Philosophical Analysis," *Journal of Consciousness Studies* 17(9-10):7-65 (2010); and Ralph Nader's blog post.
Film: *2001*
For those interested: Campbell, "The Last Evolution" (1932 sci fi short story).

II. Weeks 3-5. Can robots be 'persons'?
Week 3: Can robots have consciousness and think?
Rdg: A.M. Turing, "Computing Machinery and Intelligence," *Mind* 59:433-60 (1950)-omit section 5 (439-442); John Searle, "Chinese Room Argument," online; Paul Churchland, Matter and Consciousness (1988), Introduction; Hubert Dreyfus, "A History of First Step Fallacies," *Minds and Machines* 22:87-99 (2012).
For those interested: Tomasello, Natural History of Human Thinking (2014); Diane Proudfoot, "Rethinking Turing's Test," *Journal of Philosophy* 110(7):391-441 (2013).
Short paper due

Week 4: What is it to be a person? Non-human animals vs Robots vs Humans vs Plants vs Clones
Rdg: Mary Ann Warren, Moral Status: Obligations to Persons and other living things (1997), excerpts; Robert Sparrow, "The Turing Triage Test," Ethics and Information Technology 6:203-13 (2004) (2005).
Film: *War for the Planet of the Apes*
For those interested: TV Series *Extant*; Christopher Stone, Earth and other Ethics: The Case for Moral Pluralism (1987).

Week 5: Can robots have dignity and free will?
Rdg: Isaac Asimov, The Bicentennial Man (1976), online; Angela Martin, "On respecting animals, or can animals be wronged without being harmed," *Res Publica* 25:83-99 (2019); Stephen Petersen, "The Ethics of Robot Servitude," *Journal of Experimental and theoretical Artificial intelligence* 19(1):43-54 (2007).
For those interested: David Gunkel, The Machine Question (2012).

III. Weeks 6-7: Can morality be programmed into a machine?
Week 6: Ethical frameworks: utilitarian, deontological, virtue theory
Rdg: Wendell Wallach et.al., Moral Machines, ch. 5; Etzioni and Etzioni, "Incorporating Ethics into A.I.," *Journal of Ethics* 21:403-18 (2017).
For those interested: Wallach et.al., "A Conceptual and Computational Model of Moral Decision Making in Human and Artificial Agents," *TopiCS* 2(2010):454-85.
Week 7: How to program robots
Rdg: Isaac Asimov, I, Robot; Michael and Susan Leigh Anderson, "Machine Ethics: Creating an Ethical Intelligent Agent," *AI Magazine* Winter 2017:15-26.
For those interested: Alan Winfield et.al., "Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection," ResearchGate/net/publication/268741533 (2014).

IV. Weeks 8-15: Topics and Applications
Week 8: AI and Business/Economics
Rdg: Kiron and Schrage, "Strategy for and with AI," *MIT Sloan Management Review* (Summer 2019), 30-35; "The Glass-Half-Full Argument about AI and Jobs," *MIT Sloan Management Review* (Summer 2019), 17-19; Derek Thompson, "A World without Work," *The Atlantic Monthly* July/Aug 2015, online; Carlson, "The Robotic Reporter," *Digital Journalism* 3(3):416-31 (2015); and Ben Casselman, "Amazon's Latest Experiment: Retraining its Work Force," *New York Times*, July 11, 2019.
For those interested: David Autor, "Why are there still so many jobs? The history and future of workplace automation," *Journal of Economic Perspectives* 29(3):3-30 (2015); Christer Clerwall, "Enter the Robot Journalist," *Journalism Practice* 8(5):519-31 (2014).
Paper 1 Due

Weeks 9-10: Self-Driving Cars and Military Drones
Week 9 Rdg: David C. Vladeck, "Machines without Principals: Liability Rules and Artificial Intelligence," *Washington Law Review* 89(1):117-50 (2014); Jeffrey Gurney, "Imputing Driverhood," ch. 4 in Lin, et.al. ed. Robot Ethics 2.0 (2017); Edmond Awad, et.al., "The Moral Machine Experiment," *Nature* 563:59-65 (Nov. 2018); *Brouse v. U.S.,* 83 F. Supp. 373 (1949)
Week 10 Rdg: Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24(1)62-77 (2007).

For those interested: Coca-Vila, "Self-driving cars in dilemmatic situations," *Criminal Law and Philosophy* 12:59-82 (2018).

Weeks 11-12: Robot Sex and Love
Week 11: Sex
Rdg: Neil McArthur, "The Case for sexbots," in Danaher and McArthur, eds. Robot Sex (2018); Romy Eskens, "Is Sex with Robots Rape?," *Journal of Practical Ethics*; J. Danaher, "Robotic Rape and Robotic

Child Abuse: Should they be Criminalized?," *Criminal Law and Philosophy* 11(1)71-95 (2017).
Week 12: Love
Rdg: Michael Hauskeller, "Automatic Sweethearts for Transhumanists," in Danaher and McArthur, eds. Robot Sex (2018); Savulescu and Sandberg, "Neuroenhancement of Love and Marriage," *Neuroethics* 1:31-44 (2008);
Film: *Her*
For those interested: Marina Adshade, "Sexbot-induced social change: an economic perspective," in Danaher and McArthur, eds. Robot Sex (2018); Daniel Engber, "The Strange Case of Anna Stubblefield," *NYT Magazine*, Oct. 20, 2015, online; follow up article, *NYT Magazine*, April 5, 2018; Danaher, "Robots, law and the retribution gap," *Ethics Info Technol* 18:299-309 (2016); R. Sparrow, "Robots, Rape and Representation," *International Journal of Robotics* (2017); Ian Yoeman, Michelle Mars, "Robots, men and sex tourism," *Futures* 44:365-71 (2012).

Weeks 13-14: AI, big data, and politics
Week 13: The uses and abuses of big data
Rdg: O'Neil, Weapons of Math Destruction (2017)
Week 14: Politics and Big Data
Rdg: Pariser, The Filter Bubble, Intro; Will Knight, "Inside the race to catch the worryingly real fakes that can be made using A.I.," *MIT Technology Review* 121(5):37-41 (2018); Tiffany Hsu, "1.6 Million Followed her on Instagram. She doesn't exist," *New York Times*, June 18, 2019; Zynep Tufekci, "The Road from Tahrir to Trump," *MIT Technology Review* 121(5):11-17 (2018).
For those interested: Mayer-Schoenberger and Cukier, Big Data: A Revolution that will transform how we work and think (2013); Pariser, The Filter Bubble (2011), chs. 1-4.

Week 15: Big Data and Health Care
Rdg: Eric Topol, Deep Medicine: How AI can Make Healthcare Human Again (2019), ch. 1, 11-13; Daisuke Wakabayashi, "Google and University of Chicago Are Sued Over Health-Data Sharing," *New York Times* June 26, 2019.
Paper 2 due
Group Project presentations

Additional notes:
Policy on Accommodations: In compliance with the Americans with Disabilities Act Amendments Act (ADAAA), students who require reasonable accommodations due to a disability to properly execute coursework must register with Student Accessibility Services (SAS) and follow all SAS procedures. SAS has offices across three of FAU's campuses – Boca Raton, Davie and Jupiter—however disability services are available for students on all campuses. For more information, please visit the SAS website at www.fau.edu/sas/.

Counseling and Psychological Services (CAPS) Center: Life as a university student can be challenging physically, mentally and emotionally. Students who find stress negatively affecting their ability to achieve academic or personal goals may wish to consider utilizing FAU's Counseling and Psychological Services (CAPS) Center. CAPS provides FAU students a range of services – individual counseling, support

meetings, and psychiatric services, to name a few – offered to help improve and maintain emotional well-being. For more information, go to http://www.fau,edu/counseling/

Code of Academic Integrity Policy Statement: Students at Florida Atlantic University are expected to maintain the highest ethical standards. Academic dishonesty is considered a serious breach of these ethical standards, because it interferes with the university mission to provide a high quality education in which no student enjoys an unfair advantage over any other. Academic dishonesty is also destructive of the university community, which is grounded in a system of mutual trust and places high value on personal integrity and individual responsibility. Harsh penalties are associated with academic dishonesty. For more information, see University Regulation 4.001 and the WHC Academic Honor Code at http://www.fau.edu/honors/academics/honor-code.php.

Classroom Etiquette Policy: In order to enhance and maintain a productive atmosphere for education, personal communication devices such as smartphones are to be disabled during class.
Attendance Policy: Students are expected to attend all of their scheduled University classes and to satisfy all academic objectives as outlined by the instructor. The effect of absences upon grades is determined by the instructor, and the University reserves the right to deal at any time with individual cases of non-attendance. Students are responsible for arranging to make up work missed because of legitimate class absence, such as illness, family emergencies, military obligation, court-imposed legal obligations or participation in University-approved activities. Examples of University-approved reasons for absences include participating on an athletic or scholastic team, musical and theatrical performances and debate activities. It is the student's responsibility to give the instructor notice prior to any anticipated absences and within a reasonable amount of time after an unanticipated absence, ordinarily by the next scheduled class meeting. Instructors must allow each student who is absent for a University-approved reason the opportunity to make up work missed without any reduction in the student's final course grade as a direct result of such absence.

Florida Atlantic University policies regarding incomplete grades can be found in the University Catalog. Late work is subject to a grade reduction.