Ethics of Emerging Technologies
PHI 350/CHV 356 | Spring 2021 | Fridays 1:30-4:20pm ET

Johann Frick
jdfrick@princeton.edu

Michal Masny
mmasny@princeton.edu

Brief description
This course examines key technological developments and challenges of the 21st century from an ethical perspective. We will discuss some of the following topics: Self-driving cars and autonomous weapons systems; surveillance and the value of privacy; the use of predictive algorithms in the criminal justice system and the question of algorithmic unfairness; the impact of technology on employment and the promise of unconditional basic income schemes; human enhancement and genetic testing; and the risk of human extinction.

Office hours
We will be hosting joint office hours on **Tuesdays, 5:30-6:50pm**. Please sign-up for a 20-minute slot using our WASE calendar.

Course requirements and assessment
This course will be held in a seminar format. You will read and watch the assigned materials in advance of the meeting, and then we will discuss them together during the seminar.

Assessment for this course has a number of different components:

(i)	Once during the semester	Oral presentation	10%
(ii)	01 March at 11:59pm ET	First paper (1000 words)	15%
(iii)	22 March at 11:59pm ET	Second paper (1000-1500 words)	15%
(iv)	05 May at 11:59pm ET	Third paper (2500 words)	30%
(v)	Whole semester	Class participation	30%

Class participation
As you can see, class participation is very important for this course. We are asking you do two things.
   a) **Post a reaction to at least one of the assigned materials on the Canvas discussion board by 11:59pm ET on the day before the seminar.** You can clarify a tricky aspect of the paper, raise an objection, ask a pertinent question, bring attention to an interesting case study, or simply identify an issue that you would like us to talk about during the seminar. We will read your comments and do our best to take some of them up during the seminar.
   b) **Actively participate in the discussions during the seminar.** The topics covered during this course can be controversial, so be respectful of one another. At the same time, do not hesitate to present arguments in support of unpopular positions or objections to those widely held.

Oral presentation
You are required to do one oral presentation on a topic of your choosing (pending availability). Presentation slots will be assigned during our first seminar. A presentation should be no more than 8-10 minutes. It can consist of a philosophical commentary on one of the assigned readings, an explanation of the technical aspects of one of the technologies under discussion, or a case study applying that week's philosophical concepts to a specific practical issue. You must meet with us by Tuesday evening of the week in which you will be giving your presentation to go over an outline for your ideas.

Schedule of topics and assigned materials

You should review all assigned materials before the corresponding seminar. All readings and videos will be available through the course website on Canvas. See the course website for further, optional readings.

(1) Introduction to ethical theory (05 Feb)

Shafer-Landau explains the ambitions, the limits and the tools of ethics, and then introduces consequentialism, highlighting its virtues as an ethical theory.

- Russ Shafer-Landau. (2012). Introduction. In his *Fundamentals of Ethics*. (17 pages)
- Russ Shafer-Landau. (2012). Consequentialism: Its Nature and Attractions. In his *Fundamentals of Ethics*. (16 pages)

(2) Self-driving cars (12 Feb)

Thomson's article is a classic statement of the famous 'Trolley Problem'. Frick presents an opinionated overview of various philosophical responses to this problem. Bard offers a brief commentary on what Tesla's Autopilot can see on the road. Nyholm and Smids examine the relevance of the Trolley Problem for the design of autonomous vehicles.

- Judith Thomson. (1986). Killing, Letting Die, and the Trolley Problem. In her *Rights, Restitution, and Risk: Essays in Moral Theory*. (16 pages)
- Johann Frick's lecture on the Trolley Problem (recorded this fall for Intro to Moral Philosophy) (45 minutes)
- Wes Bard. (2020). Tesla Autopilot is better than you think! Here's why. (part 1). (7 minutes)
- Sven Nyholm. & Jilles Smids. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem? *Ethical Theory and Moral Practice*. (15 pages)

(3) Technology and war (19 Feb)

McMahan's piece is an accessible introduction to contemporary Just War Theory by today's leading philosopher of war. Shane makes the case for the use of drones and unmanned aerial vehicles as more precise and humane alternatives to traditional military alternatives. Kaag and Kreps raise the worry that unmanned drones may lower the threshold for resorting to force. Sparrow considers who should be held responsible when an autonomous weapon system is involved in an atrocity that would normally be described as a war crime.

- Jeff McMahan. (2012) Rethinking the 'Just War,' Part 1. and Part 2. *The New York Times*. (~8 pages)
- Scott Shane. (2012). The Moral Case for Drones. *The New York Times*.
- John Kaag and Sarah Kreps. (2012). The Moral Hazard of Drones. *The New York Times*.
- Robert Sparrow. (2007). Killer Robots. *Journal of Applied Philosophy*. (16 pages)

Film screening (attendance optional): In the evening, starting at 6pm, we will host a joint viewing and discussion of Gavin Hood's *Eye in the Sky* (2016), starring Helen Mirren and Alan Rickman.

(4)   Privacy and surveillance I – The right to privacy (26 Feb)
Thomson argues that putative violations of the right to privacy are really just violations of other rights, such as the right to control how other people use our body and property. Marmor argues that there *is* a distinctive right to privacy which is grounded in people's interest in having control over the ways in which they can present themselves to others.

- Judith Thomson. (1975). The Right to Privacy. *Philosophy & Public Affairs.* (20 pages)
- Andrei Marmor. (2015). What is the Right to Privacy? *Philosophy & Public Affairs.* (24 pages)

First paper due on 01 March (Monday) at 11:59pm ET.

(5)   Privacy and surveillance II – Big data and state surveillance (05 Mar)
Veliz describes the extent to which corporations and governments collect data on us. The Guardian provides a brief overview of government surveillance programmes revealed by Edward Snowden in 2013. Edward Snowden explains how easily and extensively our actions are being tracked through our phones. Kosinski and colleagues demonstrate how reliably sensitive information about people can be inferred from what might seem like innocuous online activity.

- Carissa Veliz. (2020). Introduction *and* Chapter 1: Data Vultures *and* Chapter 2: How Did We Get here?. In her *Privacy is Power: Why and How You Should Take Back Control of Your Data.* (24 pages)
- The Guardian. (2013). [The NSA and surveillance ... made simple - animation](#). Video. (3 minutes)
- Edward Snowden. [How Your Cell Phone Spies on You](#). JRE Podcast. (24 minutes)
- Michal Kosinski, et al. (2013). [Private traits and attributes are predictable from digital records of human behavior](#). *Proceedings of the National Academy of Sciences*. (4 pages)

(6)   Algorithmic discrimination I – Race, criminal justice, and search engines (12 Mar)
Angwin and colleagues argue that hat an algorithm used widely in the criminal justice system is biased against Black people. Noble describes how search engines can perpetuate harmful racist stereotypes. Zimmerman and colleagues argue that problems arising from the use of algorithms cannot be solved simply by designing better algorithms. Alexander identifies and explains huge racial disparities at every stage of the criminal justice process, from the initial stop, search, and arrest to the plea bargaining and sentencing phases.

- Julia Angwin, et al. (2016). [Machine Bias](#). ProPublica. (~6 pages)
- Safiya Noble. (2018). [Google's Algorithm: History of Racism Against Black Women](#). TIME Magazine. (~4 pages)
- Annette Zimmerman, Elena di Rosa, and Hochan Kim. (2020). [Technology Can't Fix Algorithmic Injustice](#). Boston Review. (~10 pages)
- Michelle Alexander. (2010). Chapter 3: The Color of Justice. In her The New Jim Crow: Mass Incarceration in the Age of Colorblindness. (42 pages)

(7)   Algorithmic discrimination II – How to measure algorithmic unfairness (19 Mar)
Hellman argues that two otherwise plausible ways of measuring algorithmic fairness sometimes conflict with one another. (Feel free to skip all legal footnotes.) Long responds to the ProPublica investigation and argues against one measure of algorithmic fairness. (Math-phobes beware: This article is slightly technical, but rewarding. Be sure to set aside extra time for this reading.)

- Deborah Hellman. (2020). Measuring algorithmic fairness. *Virginia Law Review.* (56 pages)
- Robert Long. (manuscript). Fairness in machine learning. Against false positive rate equality as a measure of fairness. (21 pages)

Second paper due on 22 March (Monday) at 11:59pm ET.

(8)   The future of work I – The goods and bads of work (26 Mar)
Vox considers whether technological developments will lead to mass unemployment. Gheaus and Herzog argue that work is important because it gives us an opportunity to attain excellence at something, make a social contribution, experience community, and gain social recognition. Danaher argues that work is bad and keeps getting worse, and that its technological elimination would be desirable.

- Vox. (2017). [The big debate about the future of work, explained](#). Video. (9 minutes).
- Anca Gheaus & Lisa Herzog. (2016). The Goods of Work (Other Than Money!). Journal of Social Philosophy. (20 pages)
- John Danaher. (2019). Chapter 3: Why you should hate your job. In his *Automation and Utopia: Human Flourishing in a World without Work.* (34 pages)

(9)   The future of work II – Universal Basic Income (02 Apr)
Bregman outlines the basic case for Universal Basic Income. Van Parijs and Vanderborght consider the objection that Universal Basic Income would be unfair. Gourevitch and Stanczyk critique Universal Basic Income from the Left.

- Rutger Bregman. (2014). [Why we should give everyone a basic income](#). TedX Talk. (17 minutes)
- Philippe van Parijs and Yannick Vanderborght. (2017). Chapter 5: Ethically Justifiable? Free Riding Versus Fair Shares. In their *Basic Income: A Radical Proposal for a Free Society and a Sane Economy.* (34 pages)
- Alex Gourevitch and Lucas Stanczyk. (2018). The Basic Income Illusion. Catalyst Journal. (22 pages)

(10)   Enhancement and genetics I – Procreation (09 Apr)
Buchanan and colleagues argue that, with some exceptions, parents should be free to use genetic intervention techniques to produce the best offspring that they can. Sandel argues that the pursuit of perfection will undermine our appreciation for the gifted character of human life and parental relations.

- Allen Buchanan et al. (2001). Chapter 5: Why not the best? In their *From Chance to Choice: Genetics and Justice.* (47 pages)
- Michael Sandel. (2004). [The Case Against Perfection](#). *The Atlantic Monthly.* (~8 pages)

(11) Enhancement and genetics II – Status quo bias and genetic privacy (16 Apr)

Bostrom and Ord argue that status quo bias may be responsible for opposition to cognitive enhancement and propose a heuristic for reducing this bias. Kilbride considers whether patients have a moral duty to share genetic findings with family members if the information could help them avoid disease. Ethics Unwrapped describe some recent controversies concerning home DNA tests.

- Nick Bostrom and Toby Ord. (2006). The Reversal Test: Eliminating Status Quo Bias in Applied Ethics. *Ethics*. (24 pages).
- Madison Kilbride. (2018). Genetic Privacy, Disease Prevention, and the Principle of Rescue. *The Hastings Center Report*. (8 pages)
- Ethics Unwrapped. (2019). [Ethical Use of Home DNA Testing]. (2 pages)

(12) Existential risk (23 Apr)

Ord explains why safeguarding humanity's future is the defining challenge of our time. Beckstead argues that even relatively small reductions in existential risk are much more important than significantly improving the lives of currently existing people. Scheffler suggests that many activities and projects that we care deeply about would lose their value if human extinction was imminent

- Toby Ord. (2020). [The Precipice: existential risk and the future of humanity]. Narrated presentation summarising his new book by the same title. (20 min)
- Nick Beckstead. (2019). A brief argument for the overwhelming importance of shaping the far future. In *Effective Altruism: Philosophical Issues*. Edited by Hilary Greaves and Theron Pummer. Oxford University Press. (19 pages).
- Sam Scheffler. (2018). Chapter 2: Reasons to Worry – Interests and Love. In his *Why Worry about Future Generations?* (28 pages).

Third paper due on 05 May (Dean's Date) at 11:59pm ET.