The Ethics and Governance of Artificial Intelligence (Fall 2017)

Elon Musk, Bill Gates, and Stephen Hawking say that AI poses an existential threat to humanity. Whether or not that's true, the mainstreaming of tightly-coupled complex autonomous systems raises pressing questions now. Who bears responsibility for what an autonomous system does? When and how should governments seek to regulate their uses? This reading group will explore questions like these in two areas: self-driving cars and risk scores in the justice system.

Meeting dates:
Monday, September 11th, 5-7pm.
Monday, October 2nd, 5-7pm.
Monday, October 16th, 5-7pm.
Monday, October 30th, 5-7pm.

All meetings will be held in WCC 4059.

H2O playlist: https://h2o.law.harvard.edu/playlists/52829

Meeting #1: Can Computers Think Like Humans?: The Searle Chinese Room Thought Experiment

Readings

- "Minds, Brains, and Programs" by John Searle (Behavioral and Brain Sciences, 1980) http://cogprints.org/7150/1/10.1.1.83.5248.pdf archived at https://perma.cc/7HED-3SBH.
    - *Searle presents a thought experiment reflecting on fundamental semantic and philosophical differences between human minds and computational processes. Is it theoretically possible for a computer to "think" like a human being?*
- "The Chinese Room Argument" by David Cole (Stanford Encyclopedia of Philosophy, Updated 2014) https://plato.stanford.edu/entries/chinese-room/.
    - Read Sections 1, 3, Introduction to 4, 4.1-4.1.1, 5.1-5.3, Conclusion. Remainder optional.
    - *An entry from the Stanford Encyclopedia of Philosophy outlining the Chinese Room argument and several significant critiques it has faced. What are the key issues at stake in the debate surrounding the Chinese Room argument?*
- "Thinking Machines: The Search for Artificial Intelligence" by Jacob Roberts (Chemical Heritage Foundation, 2016) https://www.chemheritage.org/distillations/magazine/thinking-machines-the-search-for-artificial-intelligence archived at https://perma.cc/C55K-CCGT.
    - *This piece outlines key milestones and ideas in the development of AI technology. How have public perceptions of AI development diverged from its realities? Have such perceptions changed over time?*

Meeting #2: AI and Self-Driving Cars

Assignment: http://moralmachine.mit.edu/ (Try out a few scenarios before class).
Readings

- "Whose Life Should Your Car Save?" by Jean-François Bonnefon, Azim Shariff, Iyad Rahwan (New York Times, 2016) https://www.nytimes.com/2016/11/06/opinion/sunday/whose-life-should-your-car-save.html.
    - *The authors introduce challenging ethical questions related to how autonomous vehicles should handle unavoidable accidents with potentially lethal consequences. Should autonomous vehicles prioritize the safety of their passengers relative to that of pedestrians and passengers in other vehicles?*
    - [FOR REFERENCE] "The Social Dilemma of Autonomous Vehicles" by Jean-François Bonnefon, Azim Shariff, Iyad Rahwan (Science, 2016) http://science.sciencemag.org/content/352/6293/1573.full archived at https://perma.cc/T32T-BVGE.
- "The Numbers Don't Lie: Self-Driving Cars Are Getting Good" by Alex Davies (Wired, 2017) https://www.wired.com/2017/02/california-dmv-autonomous-car-disengagement/ archived at https://perma.cc/C3WE-WLBZ.
    - *Davies investigates the improvement of a number of autonomous vehicles. How should we monitor the performance of autonomous vehicles during development and deployment?*
- "Who's Responsible When a Self-Driving Car Crashes?" by Corinne Iozzio (Scientific American, 2016) https://www.scientificamerican.com/article/who-s-responsible-when-a-self-driving-car-crashes/ archived at https://perma.cc/397C-X6UL.
    - *This article outlines a number of liability-related questions that are relevant to autonomous vehicles that cause an accident. Is an autonomous vehicle manufacturer responsible for damages inflicted by its products throughout their lifespan?*
- "How Drive.ai is Mastering Autonomous Driving With Deep Learning" by Evan Ackerman (IEEE, 2017) http://spectrum.ieee.org/cars-that-think/transportation/self-driving/how-driveai-is-mastering-autonomous-driving-with-deep-learning archived at https://perma.cc/QZG8-XT5N.
    - *Ackerman explores the challenges and promises involved in developing heavily AI-dependent autonomous vehicle platforms. What level of "black box" opacity should we accept in an autonomous vehicle's decisionmaking process?*
- "Tesla's Self-Driving System Cleared in Deadly Crash" by Neal E. Boudette (New York Times, 2017) https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html.
    - *This article describes the outcome of an investigation into Tesla's Autopilot mode following a deadly collision in May 2016. In instances where human drivers have*

contributed to the behavior of autonomous vehicle systems, how should liability be distributed between the human and the system?

- "Securing the Future of Driverless Cars" by Darrell West (Brookings, 2016) https://www.brookings.edu/research/securing-the-future-of-driverless-cars/ archived at https://perma.cc/NAT9-S52E.
    - *This report details some prospective benefits and challenges of autonomous vehicle deployment with a strong focus on regulatory action and the creation of standards. How could the rise of autonomous vehicles be disruptive to established legal, social, and economic institutions?*
- "Autonomous Vehicles | Self-Driving Vehicles Enacted Legislation" (National Conference of State Legislatures, 2017) http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx archived at https://perma.cc/QXA5-4MWU.
    - *This source provides an overview of enacted legislation governing autonomous vehicle systems. How could discrepancies between states' regulations on autonomous vehicles pose a challenge to their deployment? How should autonomous vehicle manufacturers handle different legislation across jurisdictions?*

Meeting #3: AI and Justice

- "Sent to Prison by a Software Program's Secret Algorithms" by Adam Liptak (New York Times, 2017) https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html.
    - *This article provides an overview of the major events and issues surrounding the use of a sentencing algorithm in the case of Wisconsin man Eric Loomis. Do you agree with the article's claim that "There are good reasons to use data to ensure uniformity in sentencing"?*
- *State of Wisconsin vs. Eric L. Loomis*, Supreme Court of Wisconsin (2016) http://www.scotusblog.com/wp-content/uploads/2017/02/16-6387-op-bel-wis.pdf archived at https://perma.cc/639U-ZDSZ.
    - *State of Wisconsin v. Eric L. Loomis is a 2016 case heard by the Supreme Court of Wisconsin. The opinion reflects on the use and limitations of algorithms as applied to sentencing practices. Can algorithms trained on data from a broad swath of the population help inform suitably individualized sentencing?*
- "How We Analyzed the COMPAS Recidivism Algorithm" by Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin (Pro Publica, 2016) https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm archived at https://perma.cc/8BXL-NU4D.
    - *A data-driven investigative analysis of the COMPAS algorithm used in the Loomis case and elsewhere to determine whether a criminal is likely to commit*

another offense. How can technically engaged public activists shed light on the problems and failures of deployed AI technology?

Meeting #4: Missing the forest for the trees: Is alarm about AI justified?

- "Our Fear of Artificial Intelligence" by Paul Ford (MIT Technology Review, February 11, 2015) https://www.technologyreview.com/s/534871/our-fear-of-artificial-intelligence/ archived at https://perma.cc/KQ6K-LT49. (Should we replace "The Doomsday Invention" with this article?)
    - *This piece provides a good overview of the central arguments of Nick Bostrom's book, Superintelligence, and explains why many others share his alarm about AI. The author suggests that we may be able to avoid harmful consequences if we design AI to respect human interests and values.*
- "Can we Build AI without Losing Control Over It?" by Sam Harris (June 2016) https://www.ted.com/talks/sam_harris_can_we_build_ai_without_losing_control_over_it/transcript archived at https://perma.cc/MA4L-NQYE.
    - *Sam Harris echoes Nick Bostrom's concerns about AI and focuses on the lack of concern expressed by AI supporters and the general public about AI and the future of human civilization. According to Harris, the development of superintelligent machines is inevitable. He proposes that we think about how to build superintelligent AI that respects humans and shares our interests. Given what we've read and discussed, do you think superintelligent AI is inevitable? How should we mitigate the risks that Harris identifies?*
- "Why Zuckerberg and Musk are Fighting About the Robot Future" by Ian Bogost (The Atlantic, July 27, 2017) https://www.theatlantic.com/technology/archive/2017/07/musk-vs-zuck/535077/ archived at https://perma.cc/RK6C-S7Z2.
        - *Bogost points out that the recent back and forth between Mark Zuckerberg and Elon Musk about whether we should be worried or excited about AI is mainly fueled by Zuckerberg's and Musk's business interests. Does this perspective change your opinion about ongoing debates regarding the future of AI?*

- "A Blueprint For Coexistence with Artificial Intelligence" by Kai-Fu Lee (Wired, July 12, 2017) https://www.wired.com/story/a-blueprint-for-coexistence-with-artificial-intelligence/ archived at https://perma.cc/548R-C4V4.
        - *This article claims that our quality of life will improve if humans work with machines. The author acknowledges that machines will displace many human workers, but maintains that there will always be a need for humans because of our capacity to love and connect emotionally with one another. How can machines work with humans rather than against them*

and how can we best address concerns about the impact of AI on the job market?

- "How I learned to Stop Worrying and Love A.I." by Robert Burton (*The New York Times*, September 21, 2015) https://opinionator.blogs.nytimes.com/2015/09/21/how-i-learned-to-stop-worrying-and-love-a-i/ archived at https://perma.cc/EU3V-QEV4.
    - *The author points out that machines may be able to outwit humans when it comes to quantifiable data, but they will always lack emotional intelligence. Will machines eventually develop emotional intelligence and how would it impact our society?*