

## **The Ethics and Governance of Artificial Intelligence**

Professor Joi Ito, Director of the MIT Media Lab, co-taught with Professor Jonathan Zittrain  
MIT  
Spring 2018

### Syllabus

#### I. Course Description

This course will pursue a cross-disciplinary investigation of the implications of emerging technologies, with an emphasis on the development and deployment of Artificial Intelligence. We will cover a variety of issues, including the complex interaction between governance organizations and sovereign states, the proliferation of algorithmic decision making, autonomous systems, machine learning and explanation, the search for balance between regulation and innovation, and the effects of AI on the dissemination of information, along with questions related to individual rights, discrimination, and architectures of control. The course will entail an intense array of learning and teaching methods. Students will be expected to participate in a variety of activities. The class may include Media Lab and Berkman Klein Center fellows and affiliates.

#### II. Class Schedule

The class begins with a conference that starts on February 2nd and ends on February 4th. The class will meet six additional times on Tuesdays from 5-7pm at the MIT Media Lab during the remainder of the semester. The specific dates and times for all class meetings are listed below.

**Conference** (all class meetings will be held on the 6th floor of the MIT Media Lab):

Friday, February 2nd: dinner from 6-8pm (students and Assemblers only)

Saturday, February 3rd: 10am - 4pm (lunch will be provided)

Sunday, February 4th: noon - 4pm (lunch will be served at 11:30am)

**Tuesday class dates** (all classes will meet on the 6th floor of the MIT Media Lab from 5-7pm):

February 6th, February 13th, February 27th, March 20th, April 3rd, and April 10th

#### III. Class Requirements

This class is taking up issues that have only recently come into focus and are being approached by a variety of disciplines at different paces. The students in this class also hail from methodologically diverse backgrounds. Our goal in soliciting work from you is that it be meaningful, rigorous and original and that we allow you to figure out what you can bring to the table.

The final project for this class will require that you apply what you have learned from class discussions to propose a solution to a problem or a way to move the ball forward in this field. The proposal may take any form (project proposal, policy recommendation, code, etc.) that you feel adequately explains and/or demonstrates your idea. You may work on your own or in small groups and we encourage you to meet with the professors to discuss your proposals. There will also be opportunities to discuss the final project requirements during class time. Your final paper or project proposal will be due on Friday, April 20th, and may be submitted by email to Natalie Saltiel ([nsaltiel@media.mit.edu](mailto:nsaltiel@media.mit.edu)) and Samantha Bates ([sbates@law.harvard.edu](mailto:sbates@law.harvard.edu)).

#### IV. Class Materials

There is no required textbook for this class. All readings will be available online through the links in this syllabus. They will also be available on the class h2o playlist:  
<https://h2o.law.harvard.edu/playlists/53282>

Students should bring their laptops to class to use for daily activities, but all electrical devices must remain off during lectures.

Please note that elements of the class may be recorded and/or live streamed on the Internet. Students will receive a waiver via email before the beginning of the semester. Please make sure you fill out the waiver before our first class meeting.

#### V. General Information

Professor Ito and Professor Zittrain will be available for office hours by appointment. Please email Yuko Barnaby to schedule a meeting with Professor Ito and Annabel Kupke to schedule a meeting with Professor Zittrain (see contact information below).

For general scheduling/appointment questions, please contact:

##### **Yuko Barnaby**

*Assistant to Media Lab Director Ito*

Office Phone: [617-999-5579](tel:617-999-5579)

Email: [yuko@media.mit.edu](mailto:yuko@media.mit.edu)

Office location: MIT Media Lab, E-14, Director's Office 245

##### **Annabel Kupke**

*Assistant to Professor Zittrain*

Office Phone: 617-495-8351

Email: [a2jz@law.harvard.edu](mailto:a2jz@law.harvard.edu)

Office location: Harvard Law School, Griswold Hall, 5th floor, Room #505

For course related (readings/paper) questions, please contact:

**Samantha Bates**

*Research Associate*

Harvard Law School

Email: [sbates@law.harvard.edu](mailto:sbates@law.harvard.edu)

**Natalie Saltiel**

MIT Media Lab

Email: [nsaltiel@media.mit.edu](mailto:nsaltiel@media.mit.edu)

**Jordi Weinstock**

*Lecturer on Law*

Email: [jordi.weinstock@law.harvard.edu](mailto:jordi.weinstock@law.harvard.edu)

## Reading Assignments

### Opening Event (February 3rd and 4th)

- [“Minds, Brains, and Programs”](#) by John Searle (Behavioral and Brain Sciences, 1980)  
(Up to “Open Peer Commentary”)
- [“Physiognomy’s New Clothes”](#) by Agüera y Arcas, Mitchell, and Todorov (Medium, 2017)
- [“A Few Useful Things to Know about Machine Learning”](#) by Pedro Domingos (University of Washington)
- [“Machine Learning that Matters”](#) by Kiri L. Wagstaff (JPL)
- [“Thinking Machines: The Search for Artificial Intelligence”](#) by Jacob Roberts (Chemical Heritage Foundation, 2016)
- [“AI Now 2017 Report”](#) by Crawford et al. (2017)
- [“The Ethics of Artificial Intelligence: Mapping the Debate”](#) by Brent Daniel Mittelstadt et al. (Big Data & Society, 2016).
- [Toward an ethics of algorithms: Convening, observation, probability, and timeliness](#) by Ananny, M. (Science, Technology, and Human Values, 2016).
- [“The Doomsday Invention”](#) by Raffi Khatchadourian (New Yorker, 2015).

### Optional:

- [“How the machine “thinks”: Understanding Opacity in Machine Learning Algorithms”](#) by Jenna Burrell (*Big Data and Society*, 2016).
- [“Robots as Legal Metaphors”](#) by Ryan Calo (Harvard Journal of Law and Technology, 2016).

- [“Why Zuckerberg and Musk are Fighting About the Robot Future”](#) by Ian Bogost (The Atlantic, July 27, 2017).
- [“How I learned to Stop Worrying and Love A.I.”](#) by Robert Burton (The New York Times, September 21, 2015).

## Classes

### Day 1 (2/6) - Autonomy, System Design, Agency, and Liability

- [“Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction”](#) by M.C. Elish (We Robot, 2016).
- [“Machines without Principals \(sic\): Liability Rules and Artificial Intelligence”](#) by David C. Vladeck (Washington Law Review, 2014).
- [“Whose Life Should Your Car Save?”](#) by Azim Shariff et al. (New York Times, 2016).
- [“Machine Learning and the Law: Five Theses”](#) by Thomas Burri (Machine Learning and the Law Conference, 2017).
- [A.I.Versus M.D.](#) By Siddhartha Mukherjee (The New Yorker, 2017)

### Optional:

- [“When your self-driving car crashes, you could still be the one who gets sued”](#) by Madeleine Claire Elish and Tim Hwang (Quartz, 2015)
- [“Executive Summary: The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems”](#) (IEEE).

### Day 2 (2/13) - Algorithmic Bias

- The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford #NIPS2017  
[https://www.youtube.com/watch?v=fMym\\_BKWQzk](https://www.youtube.com/watch?v=fMym_BKWQzk)
- Cathy O’Neil, *Weapons Of Math Destruction : How Big Data Increases Inequality And Threatens Democracy* [e-book]. New York: Crown; 2016.
  - Read the Introduction and Chapter 1 (will print for the students)
- [“Principles for Accountable Algorithms and a Social Impact Statement for Algorithms”](#) by Nicholas Diakopoulos et al. (Fairness, Accountability, and Transparency in Machine Learning)
- [“Technology is Biased too. How do we Fix it?”](#) By Laura Hudson (Five Thirty Eight, July 20, 2017).

### Risk Assessment:

- [“Should Prison Sentences Be Based On Crimes That Haven’t Been Committed Yet?”](#) by Anna Maria Barry-Jester, Ben Casselman and Dana Goldstein (FiveThirtyEight, 2015)
- [State of Wisconsin vs. Eric L. Loomis](#), Supreme Court of Wisconsin (2016)
- [“How we Analyzed the COMPAS Recidivism Algorithm”](#) by Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin (ProPublica, 2016)
- ["A computer program used for bail and sentencing decisions was labeled biased against blacks. It’s actually not that clear"](#) by Avi Feller, Emma Pierson, Sam Corbett-Davies and Sharad Goel (Washington Post, 2016)

### **Predictive Policing:**

- [Databite No. 90 – Kristian Lum: Predictive Policing](#) (Data & Society). Reference [“To Predict and Serve?”](#) by Kristian Lum and William Isaac.
  - **Optional:**
    - [“Stuck in a Pattern: Early evidence on ‘predictive policing’ and civil rights”](#) by David Robinson & Logan Koepke (Upturn, 2017)
    - [“Predicting Financial Crime: Augmenting the Predictive Policing Arsenal”](#) by Brian Clifton, Sam Lavigne, & Francis Tseng (The New Inquiry). See also <https://whitecollar.thenewinquiry.com>

### **Credit Scoring:**

- ["Knowing the Score: New Data, Underwriting, and Marketing in the Consumer Credit Marketplace"](#) (Robinson and Yu, 2014)
- [“Credit Scoring in the Era of Big Data”](#) by Hurley and Adebayo

### **Media (image recognition)**

- [“Machines Taught by Photos to Learn a Sexist View of Women”](#) by Tom Simonite (Wired Magazine, August 21st, 2017)
- [“Why Stanford Researchers Tried to Create a ‘Gaydar’ Machine”](#) by Heather Murphy (New York Times, 2017)

### **Day 3 (2/27) - Ownership, Control, and Access**

- [“The Threat of Algocracy: Reality, Resistance and Accommodation”](#) by John Danaher (Philosophy and Technology, 2016) ([Preprint 1](#): [Preprint 2](#))
- [“Information Operations and Facebook”](#) by Facebook Newsroom
- [“Million-dollar babies”](#) (The Economist, 2016)

- [“The Race For AI: Google, Baidu, Intel, Apple In A Rush To Grab Artificial Intelligence Startups”](#) (CBInsights, 2017)
- [“The Current State of Machine Intelligence 3.0”](#) by Shivon Zilis and James Cham (O’Reilly, 2017)
- [“Artificial Intelligence Pushes the Anti-Trust Envelope”](#) by Michaela Ross (Bloomberg BNA, 2017)
- [“How Facebook’s Algorithm Suppresses Content Diversity \(Modestly\) and How the Newsfeed Rules Your Clicks”](#) by Zeynep Tufekci (Medium, 2015)
- [“Facebook Figured Out My Family Secrets. And Won’t Tell Me How”](#) by Kashmir Hill (Gizmodo, 2017).

### Optional

- [“Big Data: Bringing Competition Policy to the Digital Era - Executive Summary”](#) (OECD, 2017)
- [“Why big tech companies are open-sourcing their AI systems”](#) by Patrick Shafto (The Conversation, 2016)
- [“Data Monopolists Like Google Are Threatening the Economy”](#) by Kira Radinsky (HBR, 2015)
- [“How Baidu Will Win China’s AI Race - and, Maybe, the World’s”](#) by Jessi Hempel (Wired, 2017)
- [“How AI Startups Must Compete with Google”](#) (Startup Grind)

### Day 4 (3/20) - Governance, Explainability, and Accountability

- [“Computer says no: why making AIs fair, accountable and transparent is crucial”](#) by Ian Sample (Guardian, 2017).
- [“Transparent, explainable, and accountable AI for robotics”](#) by Sandra Wachter, Brent Mittelstadt, and Luciano Floridi (Science, 2017).
- [“Our Machines Now Have Knowledge We’ll Never Understand”](#) by David Weinberger (Backchannel, 2017).
- [“Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability”](#) by Mike Ananny and Kate Crawford (New Media and Society, 2016).
- [“Is Effective Regulation of AI Possible? Eight Potential Regulatory Problems”](#) by John Danaher (Philosophical Disquisitions 2015).

- [“Algorithmic Transparency for the Smart City”](#) by Robert Brauneis and Ellen P. Goodman (Yale Journal of Law and Technology, 2017)

**Optional:**

- [“The Doctor Just Won’t Accept That!”](#) by Zachary C. Lipton (NIPS 2017 Interpretable ML Symposium).
- [“Is Artificial Intelligence Permanently Inscrutable?”](#) by Aaron M. Bornstein (Nautilus, 2016).

**Day 5 (4/3) - Labor, Automation, and Regulation**

- [“Is This Time Different? The Opportunities and Challenges of Artificial Intelligence,”](#) by Jason Furman, (expanded remarks from the AI Now expert workshop, 2016)
- [“Regulating the Loop: Ironies of Automation Law”](#) by Meg Leta Ambrose (We Robot, 2014)
- [“Society-in-the-Loop: Programming the Algorithmic Social Contract”](#) by Iyad Rahwan
- [“10 Breakthrough Technologies: Self-Driving Trucks”](#) by David H. Freedman (MIT Technology Review, 2017)
- [“Beijing Wants A.I. to Be Made in China by 2030”](#) by Paul Mozur (New York Times, 2017)

**Optional:**

- [“Where machines could replace humans, and where they can’t \(yet\)”](#) by Michael Chui, James Manyika, and Mehdi Miremadi (McKinsey Quarterly, 2016)
- [“The Relentless Pace of Automation”](#) by David Rotman (MIT Technology Review, 2017)
- [“Basic Income: A Sellout of the American Dream”](#) by David H. Freedman (2016)
- [“End of the Road: Will automation put an end to the American trucker?”](#) by Dominic Rushe (The Guardian, 2017)
- [“China’s Plan to ‘Lead’ in AI: Purpose, Prospects, and Problems”](#) by Graham Webster, Rogier Creemers, Paul Triolo and Elsa Kania (August 1, 2017)

**Day 6 (4/10) - Conclusion: Ethics, Morals, and Frontiers**

- [Moral Machine](#) from MIT Media Lab
- [“Full Tilt: When 100% of Cars are Autonomous”](#) (New York Times, 2017). Read “Picturing the Self-Driving City.”
- [“Resisting Reduction: A Manifesto”](#) by Joi Ito
- [“Everything You Need To Know About Sophia, The World’s First Robot Citizen”](#) by Zara Stone (Forbes, 2017)