

# INTRODUCTION to the ETHICS of ARTIFICIAL INTELLIGENCE

WiSe 2019/20

Fri. 12 (c.t.) – 14h

HG / 1.072

Instructor: Dr. Catherine Stinson

## **Outline:**

This course looks at ethical issues surrounding artificial intelligence. Some of the questions examined are: Who is responsible when AI causes suffering? Do we have to give up privacy to get technological innovation? How can bias in algorithms be prevented? Can AI be effectively regulated? What can we do to make AI more ethical?

Students will develop fluency reading technical material drawn from a variety of fields, including computer science, philosophy, legal and media studies. Students will also practice ethical and critical thinking skills, such as weighing the rights of different stakeholders, and thinking through the consequences of technological innovations.

## **Textbooks:**

Broussard, *Artificial Unintelligence*

Zuboff, *The Age of Surveillance Capitalism* (also available in German translation)

plus various online readings.

## **Assessment:**

### *Engagement:*

Students will be expected to come to class having done the week's readings (additional and background readings are not required), and prepared to discuss them in depth. Some discussions will occur in small groups, and there will be additional in-class exercises.

### *Group Presentation:*

In small groups, choose a week of class material to do more in depth research about. The additional readings are good starting points for your research. Find additional sources (in any language) and real-world examples relevant to the week's topic. Develop a 10-15 minute presentation for the class, and lead a follow-up activity or discussion.

### *Final Papers:*

Choose an ethical question related to the course readings, that has a public interest angle. Examine the question, weigh any benefits and harms, consider the perspective of various stakeholders, develop an opinion, and make recommendations for action. Write a paper in the style of a newspaper opinion article or policy paper. Please consult with the instructor about your paper topic during office hours.

## **Accessibility:**

If the course delivery, readings, or assessment methods aren't working for you because of linguistic, disability, health, financial or other barriers, please do consult with the instructor. Reasonable accommodations will be made wherever possible.

## Schedule of Topics & Readings:

October 11	Introduction: board games, chatbots, paperclips & trolleys	Background Reading: Broussard ch. 2–3 Boden, <i>Mind as Machine</i> , ch. 10–12 <a href="https://www.vox.com/future-perfect/2018/11/9/18072678/self-driving-cars-philosophy-safety-trolley-problem-mit">https://www.vox.com/future-perfect/2018/11/9/18072678/self-driving-cars-philosophy-safety-trolley-problem-mit</a>
October 18	Machine learning, big data & bias	Broussard ch. 7 <a href="https://ainowinstitute.org/AI_Now_2017_Report.html">https://ainowinstitute.org/AI_Now_2017_Report.html</a> (“Bias and Inclusion” section)  Additional Reading: <a href="https://ainowinstitute.org/litigatingalgorithms-2019-us.html">https://ainowinstitute.org/litigatingalgorithms-2019-us.html</a> Friedman & Nissenbaum, “Bias in Computer Systems”
October 25	The birth of surveillance capitalism	Zuboff ch. 3,5
November 8	Ubiquitous surveillance	Zuboff ch. 7 <a href="https://www.wired.com/story/google-sidewalk-labs-toronto-quayside/">https://www.wired.com/story/google-sidewalk-labs-toronto-quayside/</a>  Additional Reading: <a href="https://www.blocksidewalk.ca/read_more">https://www.blocksidewalk.ca/read_more</a> (Film) <i>Minority Report</i>
November 15	Facial recognition & policing bodies	Zuboff ch. 8 <a href="http://gendershades.org/overview.html">http://gendershades.org/overview.html</a> <a href="https://twitter.com/katecrawford/status/1173662048208728064">https://twitter.com/katecrawford/status/1173662048208728064</a> <a href="https://www.rt.com/news/368307-facial-recognition-criminal-china/">https://www.rt.com/news/368307-facial-recognition-criminal-china/</a> <a href="https://www.lgbtqnation.com/2019/08/tsa-pessuring-trans-people-showing-genitals-can-board-plane/">https://www.lgbtqnation.com/2019/08/tsa-pessuring-trans-people-showing-genitals-can-board-plane/</a>  Additional Reading: <a href="https://www.excavating.ai/">https://www.excavating.ai/</a> <a href="https://www.buzzfeednews.com/article/meghara/period-tracker-apps-facebook-maya-mia-fem">https://www.buzzfeednews.com/article/meghara/period-tracker-apps-facebook-maya-mia-fem</a> <a href="https://www.bloomberg.com/news/articles/2019-10-03/french-liberte-tested-by-nationwide-facial-recognition-id-plan">https://www.bloomberg.com/news/articles/2019-10-03/french-liberte-tested-by-nationwide-facial-recognition-id-plan</a> <a href="https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset">https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset</a> <a href="https://www.theverge.com/2019/10/2/20896181/google-contractor-reportedly-targeted-homeless-people-for-pixel-4-facial-recognition">https://www.theverge.com/2019/10/2/20896181/google-contractor-reportedly-targeted-homeless-people-for-pixel-4-facial-recognition</a> <a href="https://www.theverge.com/2019/7/25/8929793/emotion-recognition-analysis-ai-machine-learning-facial-expression-review">https://www.theverge.com/2019/7/25/8929793/emotion-recognition-analysis-ai-machine-learning-facial-expression-review</a> Stark, “Facial recognition is the plutonium of AI” <a href="https://betaface.com/demo.html">https://betaface.com/demo.html</a>
November 22	Personal assistants & IoT	Zuboff ch. 9 <a href="https://www.wired.com/story/sex-robot-ethics/">https://www.wired.com/story/sex-robot-ethics/</a>  Additional Reading: <a href="https://www.vox.com/future-perfect/2019/8/2/20746236/ai-robot-empathy-ethics-racism-gender-bias">https://www.vox.com/future-perfect/2019/8/2/20746236/ai-robot-empathy-ethics-racism-gender-bias</a>

November 29	Personality profiling & nudging	Zuboff ch. 10  Additional Reading: Sunstein, “The Ethics of Nudging” <a href="https://www.theatlantic.com/technology/archive/2017/07/the-internet-is-one-big-personality-test/531861/">https://www.theatlantic.com/technology/archive/2017/07/the-internet-is-one-big-personality-test/531861/</a> <a href="https://slate.com/technology/2018/04/how-corporations-convincing-us-that-personality-tests-are-fun.html">https://slate.com/technology/2018/04/how-corporations-convincing-us-that-personality-tests-are-fun.html</a>
December 6	YouTube, 8chan & radicalization	<a href="https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html">https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html</a> <a href="https://www.huffpost.com/entry/what-to-know-about-8chan-the-extremist-forum-where-mass-killers-find-an-audience_n_5d476ff5e4b0acb57fce642b">https://www.huffpost.com/entry/what-to-know-about-8chan-the-extremist-forum-where-mass-killers-find-an-audience_n_5d476ff5e4b0acb57fce642b</a>  Additional Reading: <a href="https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html">https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html</a> <a href="https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2">https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2</a>
December 13	AI in Asia	<a href="https://www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook">https://www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook</a> <a href="https://www.wired.com/story/inside-chinas-massive-surveillance-operation/">https://www.wired.com/story/inside-chinas-massive-surveillance-operation/</a> <a href="https://www.buzzfeednews.com/article/pranavdixit/india-is-creating-a-national-facial-recognition-system-and?utm_source=dlvr.it&amp;utm_medium=twitter">https://www.buzzfeednews.com/article/pranavdixit/india-is-creating-a-national-facial-recognition-system-and?utm_source=dlvr.it&amp;utm_medium=twitter</a>  Additional Reading: <a href="https://www.independent.co.uk/news/world/asia/china-has-made-obedience-to-the-state-a-game-a6783841.html">https://www.independent.co.uk/news/world/asia/china-has-made-obedience-to-the-state-a-game-a6783841.html</a> <a href="https://www.npr.org/2019/10/10/768841864/after-china-objects-apple-removes-app-used-by-hong-kong-protesters">https://www.npr.org/2019/10/10/768841864/after-china-objects-apple-removes-app-used-by-hong-kong-protesters</a> <a href="https://www.wsj.com/video/under-ais-watchful-eye-china-wants-to-raise-smarter-students/C4294BAB-A76B-4569-8D09-32E9F2B62D19.html">https://www.wsj.com/video/under-ais-watchful-eye-china-wants-to-raise-smarter-students/C4294BAB-A76B-4569-8D09-32E9F2B62D19.html</a> <a href="https://www.scmp.com/news/china/politics/article/3027349/artificial-intelligence-watching-chinas-students-how-well-can">https://www.scmp.com/news/china/politics/article/3027349/artificial-intelligence-watching-chinas-students-how-well-can</a>
December 20	Insider fixes	Grosz et al. “Embedded EthiCS: Integrating Ethics Broadly Across Computer Science Education” <a href="https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech">https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech</a> <a href="https://www.digitaltrends.com/web/ways-to-decentralize-the-web/">https://www.digitaltrends.com/web/ways-to-decentralize-the-web/</a>  Additional Reading: <a href="https://knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech">https://knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech</a>
January 10	Outsider fixes	First Nations Information Governance Centre “Ownership, Control, Access, and Possession (OCAP) or Self-Determination Applied to Research” Hamraie, “Sloped Technoscience: Curb Cuts, Critical Frictions, and

		<p>Disability (Maker) Cultures”  <a href="https://cvdazzle.com/">https://cvdazzle.com/</a></p> <p>Additional Reading:  <a href="https://www.techspot.com/news/80729-complete-list-alternatives-all-google-products.html">https://www.techspot.com/news/80729-complete-list-alternatives-all-google-products.html</a>  <a href="https://www.technologyreview.com/f/613409/how-to-hide-from-the-ai-surveillance-state-with-a-color-printout/">https://www.technologyreview.com/f/613409/how-to-hide-from-the-ai-surveillance-state-with-a-color-printout/</a>  <a href="https://www.survivopedia.com/6-ways-to-defeat-facial-recognition/">https://www.survivopedia.com/6-ways-to-defeat-facial-recognition/</a></p>
January 17	Afrofuturism	<p>(Film) <i>Black Panther</i> (screening in class, location TBA)          McGlotten, “Black Data”</p> <p>Additional Reading:          Hasein, “Against Black Inclusion in Facial Recognition”</p>
January 24	The (sick) culture of AI	<p>Broussard ch. 6          Douglas, “The Moral Responsibilities of Scientists”  <a href="https://www.newyorker.com/news/news-desk/how-an-elite-university-research-center-concealed-its-relationship-with-jeffrey-epstein">https://www.newyorker.com/news/news-desk/how-an-elite-university-research-center-concealed-its-relationship-with-jeffrey-epstein</a></p> <p>Additional Reading:          Asimov, <i>I, Robot</i> (“Little Lost Robot”)          The Hacker Manifesto  <a href="https://medium.com/@selamjie/remove-richard-stallman-fec6ec210794">https://medium.com/@selamjie/remove-richard-stallman-fec6ec210794</a>  <a href="https://medium.com/@EthanZ/on-me-and-the-media-lab-715bfc707f6f">https://medium.com/@EthanZ/on-me-and-the-media-lab-715bfc707f6f</a>  <a href="https://onezero.medium.com/facing-the-great-reckoning-head-on-8fe434e10630">https://onezero.medium.com/facing-the-great-reckoning-head-on-8fe434e10630</a>  <a href="https://newrepublic.com/article/154826/jeffrey-epsteins-intellectual-enabler">https://newrepublic.com/article/154826/jeffrey-epsteins-intellectual-enabler</a>  <a href="https://concernedatmit.weebly.com/">https://concernedatmit.weebly.com/</a>  <a href="https://newrepublic.com/article/154826/jeffrey-epsteins-intellectual-enabler">https://newrepublic.com/article/154826/jeffrey-epsteins-intellectual-enabler</a>  <a href="https://www.technologyreview.com/s/614264/mit-media-lab-jeffrey-epstein-joi-ito-nicholas-negro-onte-funding-sex-abuse/">https://www.technologyreview.com/s/614264/mit-media-lab-jeffrey-epstein-joi-ito-nicholas-negro-onte-funding-sex-abuse/</a>  <a href="https://litchin.wordpress.com/2019/10/01/an-oral-history">https://litchin.wordpress.com/2019/10/01/an-oral-history</a></p>
January 31	Diversity and cultural change	<p><a href="https://qz.com/1598345/microsoft-staff-are-openly-questioning-the-value-of-diversity/">https://qz.com/1598345/microsoft-staff-are-openly-questioning-the-value-of-diversity/</a>          Stinson, “What is the Purpose of Diversity in AI?”</p> <p>Additional Reading:  <a href="https://www.dhillonlaw.com/wp-content/uploads/2018/01/Damore-Google-Manifesto.pdf">https://www.dhillonlaw.com/wp-content/uploads/2018/01/Damore-Google-Manifesto.pdf</a>  <a href="https://slate.com/technology/2019/08/google-political-speech-workplace.html">https://slate.com/technology/2019/08/google-political-speech-workplace.html</a>  <a href="https://www.theguardian.com/us-news/2019/mar/28/tech-workers-silicon-valley-activism">https://www.theguardian.com/us-news/2019/mar/28/tech-workers-silicon-valley-activism</a>  <a href="https://www.nytimes.com/2017/10/16/opinion/diversity-tech-women-silicon-valley.html">https://www.nytimes.com/2017/10/16/opinion/diversity-tech-women-silicon-valley.html</a></p>