

## **The Ethics & Politics of Artificial Intelligence: An Introduction**

**Lauren M.E. Goodlad** and **Alex Guerrero**

Rutgers University – New Brunswick

Spring 2022

Course No: 350:641 (English) and 730:680 (Philosophy)

Thursday - 10:20 a.m. – 1:20 p.m.

Murray Hall 207

### **Course Description**

“Artificial intelligence” (AI) is ubiquitous in both the imagined and actual worlds of our day. This introductory graduate seminar provides frameworks for critical thinking on the topic: from the conceptual, technical, and historical, to the ethical, social, and political. As “AI” conjures fictional formulations, real-world transformations, and a culture of hype, students in the humanities must learn to distinguish between what is and what might be while considering the ethical and political questions of the present and future.

Our course begins by familiarizing students with “artificial intelligence.” What is AI today, what has it been in the past, and what might it become in the near or not-so-near future? How and why does the AI of science fiction differ from the commercial products and research horizons of industry and academia? Why is the hype so pervasive and who benefits from the attendant confusion and anxiety?

We proceed by engaging some of the pressing ethical and political questions that AI raises including automated decision-making and attendant problems of bias; privacy, data mining, and surveillance; large language models and their “stochastic” approach and social and environmental effects; robots, automation, and the future of work; the impact of algorithms on democracy and political life; and the uses and abuses of AI in love, art, and sex. As we consider the longer-term horizons of AI, we will address possibly far-fetched concerns from the possibility of conscious robots (with rights!?), to superintelligence, “the singularity,” human replacement, and brain emulations. Thinking about these possibilities is philosophically revealing regarding humanity, even if the problems these technological possibilities pose are unlikely to arise in our lifetimes, in several lifetimes, or ever.

This is an interdisciplinary seminar designed for humanists: knowledge of philosophical ethics and technical expertise is welcomed but not at all required. (Background concepts from ethics or technology will be fully explained.)

## Introduction

1. 1/20	Introduction to AI	<p>Daron Acemoglu, “AI’s Future Doesn’t Have to Be Dystopian” with commentary</p> <p>Rodrigo Ochigame, “The Invention of ‘Ethical AI’”</p> <p><a href="#">Meredith Broussard, <i>Artificial Unintelligence: How Computers Misunderstand the World</i> (Chapter 6, 7, 8, 11)</a></p> <p><a href="#">Lauren M.E. Goodlad, “Excerpts from <i>Critical AI Proposal</i>”</a></p> <p>Background Reading (optional):</p> <p>Meredith Broussard, <i>Artificial Unintelligence</i>, Chapter 3</p>
2. 1/27	Technology and Humanity: What’s New?	<p>Albert Borgmann, <a href="#">Technology and the Character of Contemporary Life</a> (Chapters 1, 2, 7, 8-11, 13-16, 18, 22-23)</p> <p>Meredith Whittaker, "<a href="#">The Steep Cost of Capture</a></p> <p><u>Suggested Further Reading:</u></p> <p>Sara T. Roberts, "Your AI is Human." Ed. Thomas S. Mullaney et al. <a href="#">Your Computer is on Fire (2021)</a>: Chapter 2</p>

## The Ethical and Political Issues of Now and the Near Term

3. 2/3	Automated Decisions and Algorithmic Bias	<p>Cathy O’Neil, <a href="#">Weapons of Math Destruction</a>: Introduction, Chapter 1, Chapter 10, Conclusion</p> <p>Safiya Umoja Noble, <a href="#">Algorithms of Oppression: How Search Engines Reinforce Racism</a> Introduction, Chapter 2, Conclusion</p> <p>Ruha Benjamin, <i>Race After Technology</i>, Introduction</p> <p><u>AI in the Popular Press: (optional)</u></p> <p>"The Reddit Revolt that Led to CEO Ellen Pao's Resignation, Explained <i>Vox</i> July 2015.</p> <p>"Machine Bias," <i>Pro Publica</i> May 23, 2016.</p> <p>"A Face Scanning Algorithm Increasingly Decides If You Get a Job." <i>Washington Post</i>. November 6, 2019.</p> <p>ShotSpotter: <a href="https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220">https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220</a></p> <p>"The Movement to Hold AI Accountable Gains More Steam." <i>Wired</i> Dec. 2 2021.</p>
4. 2/10	Automated Decisions and Algorithmic Bias	<p>Gabrielle Johnson, "<a href="#">Algorithmic bias: on the implicit biases of social technology</a>," Synthese (2020)</p> <p>Gabrielle Johnson, "<a href="#">Are Algorithms Value-Free? Feminist Theoretical Virtues in Machine Learning</a>," <i>Journal of Moral Philosophy</i> (2022)</p> <p>Cathy O’Neil and Hanna Gunn, "<a href="#">Near-Term Artificial Intelligence and the Ethical Matrix</a>" in <i>Ethics of Artificial Intelligence</i> (Ed. by Liao, 2020)</p> <p><u>Suggested Further Reading:</u></p> <p>Lily Hu, "Race, Policing, and the Limits of Social Science." <i>Boston Review</i> (2021)</p> <p>Lily Hu, "<a href="#">What is 'Race' in Algorithmic Discrimination on the Basis of Race?</a>" (manuscript).</p> <p>-Maximilian Kasy and Rediet Abebe, "Fairness, Equality, and Power in Algorithmic Decision-Making"</p> <p>-Mar Hicks, "Sexism is a Feature, Not a Bug." Ed. Thomas Mullaney. <i>Your Computer is on Fire</i> (2021): Chapter 6</p>

5. 2/17	Large Language Models, “Stochastic Parrots,” and the ‘Dangers’ Thereof	<p>Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” (2021)</p> <p>Goodlad, "Data-Centrism and its Discontents." <i>Critical AI</i>. October 2020          Marcus &amp; Davis, "GPT-3 Bloviator." <i>MIT Review</i>. August 22, 2020</p> <p><u>Suggested Further Reading:</u></p> <p>Thompson, Greenewald, Lee, Manso: “Deep Learning’s Diminishing Returns” (2021)</p> <p>Abid, Farooqi, and Zou: “Persistent Anti-Muslim Bias in Large Language Models” (2021)</p> <p>Myers: “Rooting Out Anti-Muslim Bias in Popular Language Model GTP-3” (2021)</p> <p>Welble, Glaese, and Uesato: “Challenges in Detoxifying Language Models” (2021)</p>
6. 2/24	AI and Sex, Art, and Love	<p>Kostas Terzidis et al., "<a href="#">Unintentional intentionality: art and design in the age of artificial intelligence</a>" <i>AI &amp; Society</i> (2021)</p> <p>Robert Nozick, <a href="#">The Experience Machine</a></p> <p>-film <i>Her</i></p> <p>-Kate Devlin, "<a href="#">The Ethics of the Artificial Lover</a>" in <i>Ethics of Artificial Intelligence</i> (Liao, ed., 2020)</p> <p><u>Additional Readings (optional):</u></p> <p>-Folstad et al., "<a href="#">Future directions for chatbot research: an interdisciplinary research agenda</a>" <i>Computing</i>, 2021</p> <p>Arielle Pardes, "The Emotional Chatbots are Here to Probe Our Feelings" <i>Wired</i>, 2018</p> <p>Alexandra S. Levine, "Suicide Hotline Shares Data with For-Profit Spinoff, Raising Ethical Questions." <i>Politico</i>, January 2022.</p> <p>Joanne McNeil, "Crisis Text Line and the Silicon Valleyfication of Everything " <i>Vice</i>, Feb 10 2022</p>

7. 3/3	Surveillance, Privacy, and "Datafication"	<p>-Shoshana Zuboff, <a href="#"><i>The Age of Surveillance Capitalism</i></a>, Chapter 1, 9.</p> <p>-Jathan Sadowski, <a href="#"><i>Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives, and Taking Over the World</i></a> Chapter 1, 2, 7, 8</p> <p>-Christopher Fascella, "<a href="#">Amazon Ring Master of the Surveillance Circus</a>," <i>Federal Communications Law Journal</i> (3: 2021)</p> <p><u>Suggested Further Reading:</u></p> <p>-Todd Feathers, "<a href="#">This Private Equity Firm is Amassing Companies that Collect Data on America's ChildrenLinks to an external site.</a>," <i>The Markup</i>. Jan. 11, 2022</p>
8. 3/10	Automation and Future of Work	<p>Tom Parr, "<a href="#">Automation, Unemployment, and Taxation</a>" Social Theory and Practice (forthcoming 2022)</p> <p>Aaron James, "<a href="#">Planning for Mass Unemployment: Precautionary Basic Income</a>" in Ethics of Artificial Intelligence (Liao, ed., 2020)</p> <p>Kai Erikson, "<a href="#">On Work and Alienation</a>" <i>American Sociological Review</i> 1986, pp. 1-8</p> <p><u>Suggested Further Reading:</u></p> <p>Kurt Vonnegut, <a href="#">Player Piano</a></p> <p>Aaron Benanav, "<a href="#">Automation and the Future of Work</a>" Part 2. <i>New Left Review</i> Nov/Dec 2019.</p> <p>Mary Gray and Siddharth Sureth, "<a href="#">Introduction</a>," and "<a href="#">Chapter 1</a>," <i>Ghost Work: How to Stop Silicon Valley from Creating a New Global Underclass</i></p>

9. 3/24	AI and Democracy	<p>Zeynep Tufekci, <u>Twitter and Tear Gas: The Power and Fragility of Networked Protest</u> (<a href="#">Chapters 1, 6, and 9</a>)</p> <p>Sasha Costanza-Chock, <u>Design Justice: Community-Led Practices to Build the World We Need</u> (Introduction &amp; Chapter 2)</p> <p>The Arab Spring, Retweeted Axel Bruns, Tim Highfield, and Jean Burgess. “<a href="#">The Arab Spring and social media audiences: English and Arabic Twitter users and their networks</a>” <i>American behavioral scientist</i> 57, no. 7 (2013): 871-898.</p> <p><u>Suggested Further Reading:</u></p> <p>Nicole Aschoff, <i>The Smartphone Society: Technology, Power, and Resistance in the New Gilded Age</i> (Beacon)</p>
10. 3/31	AI and Democracy	<p>Adrienne LaFrance, “The Facebook Papers: ‘History will not judge us kindly’”</p> <p>-Yochai Benkler, Robert Faris, and Hal Roberts, <u>Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics</u> (Oxford University Press, 2018) (<a href="#">chapters 1 and 2</a>)</p> <p>Cat Zakrzewski et al., “How Facebook neglected the rest of the world, fueling hate speech and violence in India”</p> <p>C. Thi Nguyen, “<a href="#">Echo chambers and epistemic bubbles</a>” <i>Episteme</i>, 17(2)(2020), pp. 141-161</p> <p><u>Optional Reading:</u></p> <p>Karen Hao, "How Facebook Got Addicted to Spreading Misinformation." <i>MIT Review</i>. March 2021.</p> <p>Siva Vaidhyanathan, <u>Antisocial Media: How Facebook Disconnects Us and Undermines Democracy</u> (2018)</p> <p>Sheera Frankel &amp; Cecelia Kang. <i>An Ugly Truth: Inside Facebook's Battle for Domination</i> (2021)</p>
<p><b>The Ethical and Political Issues of Later or Never</b></p>		

11. 4/7	AI, consciousness, and morality	<p>Susan Schneider, <a href="#">Artificial You: AI and the Future of Your Mind (Introduction, Chapter 1, 2, 3, and 5)</a></p> <p>Additional Reading/Viewing:</p> <p>Peter Railton, <a href="#">"Ethical Learning, Natural and Artificial"</a></p> <p>film Ex Machina</p>
12. 4/14	Intelligence and Anthropomorphism	<p>Alan Turing, "<a href="#">Computing Machinery and Intelligence</a>"</p> <p>Brian Cantwell Smith, <a href="#">The Promise of Artificial Intelligence</a>: Introduction, Chapter 3, Chapter 5/6, Chapter 10, Chapter 13, Optional: Chapter 9 ("World") - note that some of these chapters are quite short.</p> <p>Judea Pearl, "<a href="#">Big Data, Artificial Intelligence, and the Big Questions</a>" in <i>The Book of Why</i></p> <p><u>AI in the Popular Press:</u></p> <p>Rohit Prasad, "The Turing Test Is Obsolete" <i>Fast Company</i>. (December 28, 2020)</p>
13. 4/21	Control and Superintelligence	<p>Nick Bostrom, <a href="#">Superintelligence: Paths, Dangers, Strategies</a> (Preface and pp. 1-6, 22-43, 54-62, 110-139, 155-176, and 226-255)</p> <p>Ted Chiang, <a href="#">The Lifecycle of Software Objects</a></p> <p>Ted Chiang, "Silicon Valley is Turning Into Its Own Worst Fear"</p> <p><u>Suggested Further Reading:</u></p> <p>Steven Shaviro, <i>Discognition</i> (2016), chapter 3 (on Chiang with some passing discussion of Bostrom)</p>

14. 4/28	Rules and Reforms	<p>Frank Pasquale, <a href="#">New Laws of Robotics, Chapter 1</a></p> <p><u>Background Readings:</u></p> <p>European Union <a href="#">Regulatory Overview</a></p> <p><a href="#">European Union Artificial Intelligence Act</a></p> <p>Michael Veale &amp; Frederik Zuiderveen Borgesius, "<a href="#">Demystifying the Draft EU Artificial Intelligence Act</a>" Computer Law Review International</p> <p>Nathalie Smuha et al., "<a href="#">How the EU Can Achieve Legally Trustworthy AI: A response to the European Commission's Proposal for an Artificial Intelligence Act</a>"</p>
----------	-------------------	--