

Cognitive Science Seminar: The Ethics of Artificial Intelligence

Wednesday 2:20-5:20 FIT 405

[PHIL 7514-PHIL 8514](#), [PSYC 7514-PSYC 8514](#), [COMP 7514-COMP 8514](#), [AUSP 7514-AUSP 8514](#)

Instructor: David Miguel Gray

dmgray2@memphis.edu

While scientists have been developing Artificial Intelligence (AI) since the 1950's, only recently has AI permeated virtually every technological aspect of our lives. Driven mainly by the increase in the power of computers and the availability of training data, it has been possible to apply new forms of AI, such as machine learning (frequently in the form of neural networks), to many of the problems we face in our increasingly technological world. But machine learning and training data pose novel ethical problems, not encountered by earlier generations of programmed AI. After exploring some fundamental issues in the development of AI, we will investigate ethical issues raised by machine learning, with a focus on ethical principles, explainability, privacy, and fairness/bias.

Guest Presenters (from 4:00-5:15)

2.2.22 	<p>"Why Privacy Concerns aren't about Privacy"</p> <p>Reid Blackman Founder and CEO of Virtue Consultants Chief Ethics Officer, Government Blockchain Association Advisory Board, Ethics Grade, Hatch, & EY</p>
2.9.22 	<p>"Ethical Principles for Artificial Intelligence"</p> <p>Paul Thagard Distinguished Professor Emeritus of Philosophy, The University of Waterloo Fellow of the Royal Society of Canada, the Cognitive Science Society, & the Association for Psychological Science</p>
2.16.22 	<p>"Legitimacy, Authority, and the Value of Explanations"</p> <p>Seth Lazar Professor of Philosophy, Australian National University Distinguished Research Fellow, Oxford Institute for Ethics in AI</p>
2.23.22 	<p>"A Case for Humans-in-the-Loop: Decisions in the Presence of Misestimated Algorithmic Scores"</p> <p>Maria De-Arteaga Assistant Professor of Information, Risk and Operation Management, University of Texas at Austin Core faculty member, Machine Learning Laboratory, University of Texas at Austin</p>
3.2.22 	<p>"Fine-Grained Explanations Using Markov Logic"</p> <p>Deepak Venugopal Associate Professor of Computer Science, The University of Memphis Affiliate, Institute for Intelligent Systems, The University of Memphis</p>

3.16.22

"Discovering Bias"



David Danks

Professor of Data Science & Philosophy, UC-San Diego
Affiliate Faculty, Department of Computer Science & Engineering, UC-San Diego

3.23.22

To Be Announced



Brent Mittelstadt

Senior Research Fellow & British Academy Postdoctoral Fellow, Oxford Internet Institute
Turing Fellow, Alan Turing Institute
Member, UK National Statistician's Data Ethics Advisory Committee

3.30.22

"Explanations in Whose Interest"



Salon Barocas

Principal Researcher, Microsoft Research (NYC lab)
Adjunct Assistant Professor, Department of Information Science, Cornell University
Faculty Associate, Berkman Klein Center for Internet & Society, Harvard University

4.6.22

To Be Announced (Topic: Fairness/Bias)



Tulsee Doshi

Google Head of Product, Responsible AI & ML Fairness

4.13.22

To Be Announced (Topic: Explainability)



Michael Brent

Office of Responsible AI at Microsoft

4.20.22

To Be Announced



Sina Fazelpour

Assistant Professor of Philosophy & Computer Science, Northeastern University
Fellow, World Economic Forum's Global Future Council on Data Policy