The Philosophy of Artificial Intelligence

Hauptseminar

Instructor: Dr. Michael T. Stuart

Email: mike.stuart.post@gmail.com

https://uni-tuebingen.de/studium/studienorganisation/semester-und-studienplanung/semestertermine/

This course will explore the social, political, moral, metaphysical and epistemological consequences of new technologies, especially artificial intelligence. We will ask questions like: Can machines create art? How have new technologies affected the roles of women? Can technology be racist? How does technology affect our social interactions with each other? Will machines be able to think and have subjective experiences? Could machines deserve moral treatment? Is there value in attempting to mimic the character and structure of the human mind in a machine?

Evaluation: ECTS will be given for assignments, 3 points for a presentation, 6 for a paper (10-15 pages), 5 points for an oral exam.

Teaching methods: The course will be conducted in English, and online via Zoom.

Course Schedule

Introduction (April 19th)

> Norvig, P. and Russell, S. 2009. *Artificial Intelligence: A Modern Approach*, Prentice. Pp. 1-33.

Foundations and Definitions of AI (April 26th)

> Boden, Margaret. 2016. *AI: Its Nature and future*. Oxford University Press. Chapter 1.

> Optional:

>> Short video on Turing machines

>> Play with a Turing machine

>> On "virtual machines," consider the Redstone computers people make in Minecraft (e.g., here)

>> One rarely told part of the history of AI is about Walter Pitts. You can read about him here

>> Simulations of different neural nets

The State of Modern AI (May 3rd)

Overall

State of AI Report 2020: https://www.stateof.ai/

Stanford's State of AI report (summary): https://hai.stanford.edu/news/state-ai-10-charts

Neural networks

YouTube: "But what is a neural network? | Chapter 1, Deep learning." 3Blue1Brown. https://www.youtube.com/watch?v=aircAruvnKk

YouTube: "Gradient descent, how neural networks learn | Chapter 2, Deep learning." 3Blue1Brown. https://www.youtube.com/watch?v=IHZwWFHWa-w

YouTube: "What is backpropagation really doing? | Chapter 3, Deep learning." 3Blue1Brown. https://www.youtube.com/watch?v=Ilg3gGewQ5U

"A Beginner's Guide to Neural Networks and Deep Learning" *A.I. Wiki*. https://wiki.pathmind.com/neural-network

Generated Adversarial Networks

YouTube: "A Friendly Introduction to Generative Adversarial Networks (GANs)." Luis Serrano. https://www.youtube.com/watch?v=8L11aMN5KY8

Genetic Algorithms

YouTube: "Genetic Algorithms Explained By Example." Kie Codes. https://www.youtube.com/watch?v=uQj5UNhCPuo

More philosophical

"Philosophers on GPT-3." *Daily Nous*. https://dailynous.com/2020/07/30/philosophers-gpt-3/

Problems facing AI (May 10th)

The Frame Problem:

YouTube. "The Frame Problem in AI." https://www.youtube.com/watch?v=EVAKG6Y2uIg

YouTube. "The biggest problem in AI? Machines have no common sense." *Big Think*. https://www.youtube.com/watch?v=IHE63fxpHCg

Dennett, D. 1984. "Cognitive wheels: The frame problem of AI" In Christopher Hookway (ed.), *Minds, Machines and Evolution*. Cambridge University Press.

Shanahan, Murray, "The Frame Problem." *The Stanford Online Encyclopedia of Philosophy*.

Solution?

Bermudez, Luis. 2021. "Overview of Embodied Artificial Intelligence." *Medium*.

The Symbol Grounding Problem:

Harnad, S. 1990. "The Symbol Grounding Problem." *Physica* D 42: 335-346.

Solution(s)?

Taddeo. Mariarosaria and Floridi, Luciano. 2005. "Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research." Journal of Experimental & Theoretical Artificial Intelligence 17(4).

Machine Consciousness (May 17th)

Dehaene, S., Lau, H. and Kouider, S. 2017. "What is consciousness, and could machines have it?" *Science* 358: 486-492.

"Consciousness." *Stanford Online Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/consciousness/

YouTube. "Can You Upload Your Mind & Live Forever?" (https://www.youtube.com/watch?v=4b33NTAuF5E)

YouTube. "Scientists Put the Brain of a Worm into a Robot…and It MOVED" (https://www.youtube.com/watch?v=eYS7UIUM_SQ)

YouTube. "These Self-Aware Robots Are Redefining Consciousness." (https://www.youtube.com/watch?v=chukkEeGrLM&t=1s)

Bostrom, N. 2012. "The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents." *Minds and Machines* 22: 71–85.

AI and Art (May 31st)

Coeckelbergh, M. 2017. "Can Machines Create Art?" *Philos. Technol.* **30,** 285–303. https://doi.org/10.1007/s13347-016-0231-5

YouTube. "How Does A.I. Art Stack Up Against Human Art?" *Vice News*. https://www.youtube.com/watch?v=hws1ZTlkz_I

McFadden, Christopher. 2019. "7 of the Most Important AI Artists That Are Defining the Genre." *Interesting Engineering*. https://interestingengineering.com/7-of-the-most-important-ai-artists-that-are-defining-the-genre

Kelly, Sean Dorrance. 2019. "A philosopher argues that an AI can't be an artist." *MIT Technology Review*. https://www.technologyreview.com/2019/02/21/239489/a-philosopher-argues-that-an-ai-can-never-be-an-artist/

Adajian, Thomas. 2018. "The Definition of Art." *Stanford Online Encyclopedia of Philosophy* https://plato.stanford.edu/entries/art-definition/

Mikalonytė, Elzė Sigutė & Kneer, Markus. Forthcoming. "Can Artificial Intelligence Make Art?" https://philpapers.org/archive/MIKCAI-2.pdf

AI and Medicine (June 7th)

Amann, Julia, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I. Madai, and the Precise4Q consortium. 2020. "Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective." BMC Medical Informatics and Decision Making 20 (1): 310. https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6

Kiener, Maximilian. 2020. "Artificial Intelligence in Medicine and the Disclosure of Risks." AI & SOCIETY, October. https://link.springer.com/article/10.1007/s00146-020-01085-w

Podcast. Onora O'Neill on Medical Consent. Philosophy Bites. https://philosophybites.com/2007/10/onora-oneill-on.html

Love and AI (June 14th)

YouTube: Devlin, K. 2019. "Sex robots: The truth behind the headlines." https://www.youtube.com/watch?v=B20E31ea9RQ&t=572s

Helm, Bennett. 2017. "Love." The Stanford Encyclopedia of Philosophy https://plato.stanford.edu/archives/fall2017/entries/love/

Sparrow, R. 2017. "Robots, Rape, and Representation." *Int J of Soc Robotics*. https://link.springer.com/article/10.1007/s12369-017-0413-z

Eskens, R. 2017. "Is sex with robots rape?" *Journal of Practical Ethics* 5: 62-76. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3087200

Turkle, S. 2011. *Alone Together*. Basic Books. Chapter 3.

AI, War and Weapons (June 21st)

Sparrow, R. 2007. "Killer Robots." *Journal of Applied Philosophy*, Vol. 24, No. 1. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-5930.2007.00346.x

Musgrave, Z. and Roberts, B. 2015. "Humans, Not Robots, Are the Real Reason Artificial Intelligence Is Scary" *The Atlantic*. https://www.theatlantic.com/technology/archive/2015/08/humans-not-robots-are-the-real-reason-artificial-intelligence-is-scary/400994/

Horowitz, M. 2016. "The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons." *Daedalus* 145 (4): 25–36. https://doi.org/10.1162/DAED_a_00409

YouTube. "A.I. Is Making it Easier to Kill (You). Here's How." | NYT, YouTube: https://www.youtube.com/watch?v=GFD_Cgr2zho

YouTube. "The future of modern warfare: How technology is transforming conflict | DW Analysis." https://www.youtube.com/watch?v=TmlBkW6ANsQ

AI and Politics (June 28th)

Edenberg, Elizabeth. 2021. "The Problem with Disagreement on Social Media: Moral not Epistemic." In *Political Epistemology*. OUP.

Podcast: CBC Radio One: AI, Privacy and Power. 2021. https://www.cbc.ca/listen/live-radio/1-55-spark/clip/15821774-498-the-spark-guide-civilization-part-five-privacy

Savaget, Chiarini, and Evans. 2019. "Empowering political participation through artificial intelligence." *Science and Public Policy* 46: 369–380. https://doi.org/10.1093/scipol/scy064

Acemoglu, D. 2021. "AI's Future Doesn't Have to Be Dystopian." *Boston Review*. https://bostonreview.net/forum/science-nature/daron-acemoglu-redesigning-ai

Oltermann, P. 2021. "'At first I thought, this is crazy': the real-life plan to use novels to predict the next war." *The Guardian*. https://www.theguardian.com/lifeandstyle/2021/jun/26/project-cassandra-plan-to-use-novels-to-predict-next-war

AI Ethics (July 5th)

Etzioni, A. and Etzioni, O. 2017. "Incorporating Ethics into Artificial Intelligence." The Journal of Ethics. https://link.springer.com/article/10.1007/s10892-017-9252-2

Thoma, J. 2021. "How should artificial agents make risky choices on our behalf?" https://www.lse.ac.uk/philosophy/blog/2021/06/08/how-should-artificial-agents-make-risky-choices-on-our-behalf/

van Wynsberghe, A., Robbins, S. "Critiquing the Reasons for Making Artificial Moral Agents." Sci Eng Ethics 25, 719–735 (2019). https://doi.org/10.1007/s11948-018-0030-8

Bostrom and Yudkowsky, 2014. "The Ethics of Artificial Intelligence" In *Cambridge Handbook of Artificial Intelligence*, edited by Keith Frankish and William Ramsey. New York: Cambridge University Press. https://intelligence.org/files/EthicsofAI.pdf

Bryson, J. 2010. "Robots Should Be Slaves." In *Close Engagements with Artificial Companions*, Y. Wilks (ed.), pp 63-74. http://www.cs.bath.ac.uk/~jjb/ftp/Bryson-Slaves-Book09.html

Schwitzgebel and Mara, 2015. "A Defense of the Rights of Artificial Intelligence." *Midwest Studies in Philosophy* XXXIX https://onlinelibrary.wiley.com/doi/pdf/10.1111/misp.12032?saml_referrer

AI, Automation, and the Meaning of Work (July 12th)

YouTube. "The big debate about the future of work, explained." Vox. https://www.youtube.com/watch?v=TUmyygCMMGA

YouTube "The future of work: is your job safe?" The Economist. https://www.youtube.com/watch?v=gUc5oN_ffRo

Russell, B. 1932. "In Praise of Idleness" Harper's Magazine.

Graber, D. 2013. On the Phenomenon of Bullshit Jobs. Strike! Magazine.

Autor, D. H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *The Journal of Economic Perspectives* 29: 3–30.

Arntz, M., Gregory, T. and Zierahn, U. 2016. "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis." *OECD Social, Employment and Migration Working Papers*. No. 189, OECD Publishing, Paris.

AI and Education (July 19th)

Ayoub, D. 2020. "Unleashing the Power of AI for Education." MIT Technology Review. Available here: https://www.technologyreview.com/2020/03/04/905535/unleashing-the-power-of-ai-for-education/

YouTube. "The Science of Thinking" Veritasium. https://www.youtube.com/watch?v=UBVV8pch1dM.

Luckin, R. 2019. "Is education ready for artificial intelligence? Machine learning and EdTech." Cambridge Assessment. https://www.cambridgeassessment.org.uk/insights/is-education-ready-ai-rose-luckin/

Chen, L, Chen, P. and Lin, Z. 2020. "Artificial Intelligence in Education: A Review." IEEE. DOI: 10.1109/ACCESS.2020.2988510. https://ieeexplore.ieee.org/abstract/document/9069875

Final Presentations, General Revision, Final Paper, Oral Exam Advice (July 26th)

No readings assigned, but have a look at some paper writing guides:

- Guidelines on Writing a Philosophy Paper (James Pryor)

- Writing a Philosophy Paper (Peter Horban)

- Guide to the Study of Philosophy (Garth Kemerling)

- Undergraduate Philosophical Writing (Martin Young)

- Tips on Writing a Philosophy Paper (Douglas Portmore)

- 7 Steps to a Better Philosophy Paper (Bryan Roberts)