PNP Reading Class

Deep Learning: Philosophy and Cognitive Science

L64 PNP 390 02, Fall 2023

Michael Barkasi
barkasi@wustl.edu

Schedule

- **T, 8/29:** **Icebreaker:** Biever, Celeste. 2023 "ChatGPT broke the Turing test – The race is on for new ways to assess AI" // Céliz, Carissa. 2023. "Chatbots shouldn't use emojis"
- **T, 9/5:** **Essay writing and ChatGPT:** Kasneci, Enkelejda et al. 2023. "ChatGPT for good? On opportunities and challenges of large language models for education" // Herbold, Steffen et al. 2023. "AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT-generated essays"
- **T, 9/12:** **Do DNNs understand?** Tamir, Michael and Shech, Elay. 2023. "Machine understanding and deep learning representation" // Nefdt, Ryan M. 2022. "Are machines radically contextualist?"
- **T, 9/19:** **Deepfakes:** Atencia-Linares, Paloma and Artiga, Marc. 2022. "Deepfakes, shallow epistemic graves: On the epistemic robustness of photography and videos in the era of deepfakes" // Rini, Regina and Cohen, Leah. 2022. "Deepfakes, deep harms"
- **T, 9/26:** **Moral responsibility:** Constantinescu, Mihaela et al. 2022. "Blame it on the AI? On the moral responsibility of artificial moral advisors" // Véliz, Carissa. 2021. "Moral zombies: Why algorithms are not moral agents"
- **T, 10/3:** **Medical diagnoses and legal advice:** Alvarado, Ramón. 2021. "Should we replace radiologists with deep learning? Pigeons, error and trust in medical AI" // Tan, Jinzhe et al. 2023. "ChatGPT as an artificial lawyer?"
- **T, 10/10:** No Class (Fall Break)
- **T, 10/17:** **Social trust of deep learning:** von Eschenbach, Warren J. 2021. "Transparency and the black box problem: Why we do not trust AI" // Ross, Amber. 2022. "AI and the expert; a blueprint for the ethical use of opaque AI"
- **T, 10/24:** **Epistemic challenges of DNNs in science:** Sullivan, Emily. 2022. "Understanding from Machine Learning Models" // Räz, Tim and Beisbart, Claus. 2022. "The importance of understanding deep learning"
- **T, 10/31:** **Empiricism? DNNs as models of human concept learning:** Buckner, Cameron. 2018 "Empiricism without magic: Transformational abstraction in deep convolutional neural networks" // Childers, Timothy et al. 2023. "Empiricism in the foundations of cognition"
- **T, 11/7:** **Insights into human cognition?** Jha, Aditi et al. 2023. "Extracting low-dimensional psychological representations from convolutional neural networks" // Quilty-Dunn, Jake et al. 2022. "The best game in town: The re-emergence of the language of thought hypothesis across the cognitive sciences"
- **T, 11/14:** **Brain models:** Doerig, Adrien et al. 2023. "The neuroconnectionist research programme" // Lillicrap, Timothy P. et al. 2020. "Backpropagation and the brain"
- **T, 11/21:** **Anyone can code a DNN (we'll do it in class!)** // Eklöf, Jon et al. 2023. "Abstraction, mimesis and the evolution of deep learning"
- **T, 11/28:** **Models of vision:** Lindsay, Grace W. 2021. "Convolutional neural networks as a model of the visual system: Past, present, and future" // Pogodin, R. et al. 2021. "Towards biologically plausible convolutional networks"
- **T, 12/5:** TBD
- No Final