3.1 - Embodied Intelligence, Agency, and Responsibility

Artificial Intelligence Policy

Prof. Jack Reilly

S2026

Think:

- *How important is it for intelligence to be "embodied" in a physical agent that interacts with the "real" world, as opposed to digital representations of the world through text, images, video, and the internet? Does embodiment (or lack thereof) change how we think of learning?*
- *What about responsibility? Are AI agents responsible for their behavior? Should they be, legally? Should the companies that make them? Does whether agents are embodied change how we think about this responsibility?*

Listen:

- *Complexity* podcast, Season 2, Episode 4: **"Babies vs. Machines"**

Read:

- Mitchell, et al. "Fully Autonomous AI Agents Should Not be Developed"

Browse:

- Liu and Wu, **"A Brief History of Embodied Artificial Intelligence, and its Outlook"**
  https://cacm.acm.org/blogcacm/a-brief-history-of-embodied-artificial-intelligence-and-its-future-outlook/
- **"NHTSA Finds Teslas Deactivated Autopilot Seconds Before Crashes"**
  https://www.motortrend.com/news/nhtsa-tesla-autopilot-investigation-shutoff-crash/

- Ziegler, Bart. 2023. **"When Will Cars Be Fully Self Driving?"** https://www.wsj.com/articles/cars-self-driving-when-c6ae4fdc
- **"The evolving safety and policy challenges of self-driving cars"**, Brookings https://www.brookings.edu/articles/the-evolving-safety-and-policy-challenges-of-self-driving-cars/

Additional Resources:

- Ganesh, **"The ironies of autonomy"** https://www.nature.com/articles/s41599-020-00646-0
- Chan et al., **"Infrastructure for AI Agents"** https://arxiv.org/abs/2501.10114
- Geistfeld, **"A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation"**
- NHTSA, **"Automated Vehicles for Safety"** https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety
- Google, **AI Responsibility Report** https://ai.google/static/documents/ai-responsibility-update-published-february-2025.pdf
- Stanford Encyclopedia of Philosophy, **"Doing vs Allowing Harm"** (especially the Trolley Problem section)
- Himmelreich, **"Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations"** https://link.springer.com/article/10.1007/s10677-018-9896-4

Submit:

- Discussion question to course chat

Tip

- **Read**, **Listen**, and/or **Watch** items are required content for the day and should be completed before class.
- **Browse** items should be skimmed but do not need deep reading unless you want to.
- **Additional Resources** are optional references for debates, final projects, and future use.