

4.1 - Computation

Artificial Intelligence Policy

Prof. Jack Reilly

S2026

Think:

- *How do computers work? Can they "simulate" minds? Or can they be minds? Practically, what goes into making a computer, and how do technological constraints structure the nature of AI itself?*

Watch:

- Veritasium, "[The World's Most Important Machine](#)"
- CGP Grey, "[How Machines Learn](#)" (*Note: this video has been retitled "How AI, like ChatGPT, learns"*)
 - and "[How Machines Really Learn](#)" *New Title: How AI, like ChatGPT, really learns*
- 3Blue1Brown, "[Large Language Models explained briefly](#)"
 - *recommended, but not required:* "[Transformers explained visually](#)"

Listen:

- *Complexity Podcast* Episode 3: What Kind of Intelligence is an LLM?

Browse:

- Hicks et al, "ChatGPT is Bullshit" *Ethics and Information Technology*
- Wolfram, S. "[What is ChatGPT doing?](#)"
- A few items on evaluating AI Intelligence:
 - Humanity's Last Exam: [Website](#) and article: Phan et al, *Humanity's Last Exam*
 - Ullman, "[Large Language Models Fail on Trivial Alterations to Theory-of-Mind Tasks](#)"
 - Mitchell, "[How do we know how smart AI systems are?](#)"

Additional Resources:

- Only a few videos from 3Blue1Brown's series on Neural Networks are assigned above, but the entire sequence is worth watching for a non-technical overview: <https://www.3blue1brown.com/topics/neural-networks>
- Witt, Stephen. *The Thinking Machine: Jensen Huang, Nvidia, and the World's Most Coveted Microchip* is worth reading in its entirety, but in particular, chapter 6, "Jellyfish" is an excellent non-technical overview of neural nets
- Similarly, the *Crash Course: Artificial Intelligence* series provides a high level overview of many current AI principles
- There is also a *Crash Course: Computer Science* for more about computers themselves
- Finally, if you are already statistically inclined, the final episodes of *Crash Course: Statistics* also cover elements of machine learning and big data analysis building off of core statistical concepts
- Rumelhart, "The Architecture of the Mind: A Connectionist Approach" *Mind Design III*
- Churchland and Sejnowski, "The Computational Brain" *Mind Design III*
- Cowie and Woodard, "The Mind is Not (Just) a System of Modules Shaped (Just) by Natural Selection" *Mind Design III*
- Vaswani et al, "[Attention Is All You Need](#)" (Google paper)

Tip

- "Read", "Listen", and/or "Watch" items are required content for the day, and should be read/heard/watched before class on that day.
- "Browse" items should be briefly looked at but do not need to be read deeply unless you want to
- "Additional Resources" do not need to be looked at; they are there to serve, if useful, as further references for your debates, final projects, and general edification later.