

CPSC 430

Computers & Society

Class 6A: Professional Ethics and Artificial Intelligence

Dr. Firas Moosvi | 2024_S1

Slides courtesy of Dr. Kevin Leyton-Brown

Class Outline

1. Announcements (5 mins)
2. Professional Ethics (30 mins)
3. Break (10 mins)
4. Artificial Intelligence (50 mins)
 - Break (10 mins)
5. Activity (20 mins)
6. Reminders before next class (5 mins)

Announcements

Professional Ethics

Is Software Engineering a Profession?

- In many ways software engineering is similar to other professions such as law or accounting
 - Dependence on professional education and practical training
 - Professional associations (IEEE, ACM)
 - Potential for bad decisions to cause significant public harm
- It's not a “full-fledged” profession:
 - No formal accreditation/licensing system
- All the same, sensible to ask software engineers to follow a code of ethics.
 - One with significant support is described in the book.
 - Nothing it says ought to come as a big surprise at this point in the course ☺.

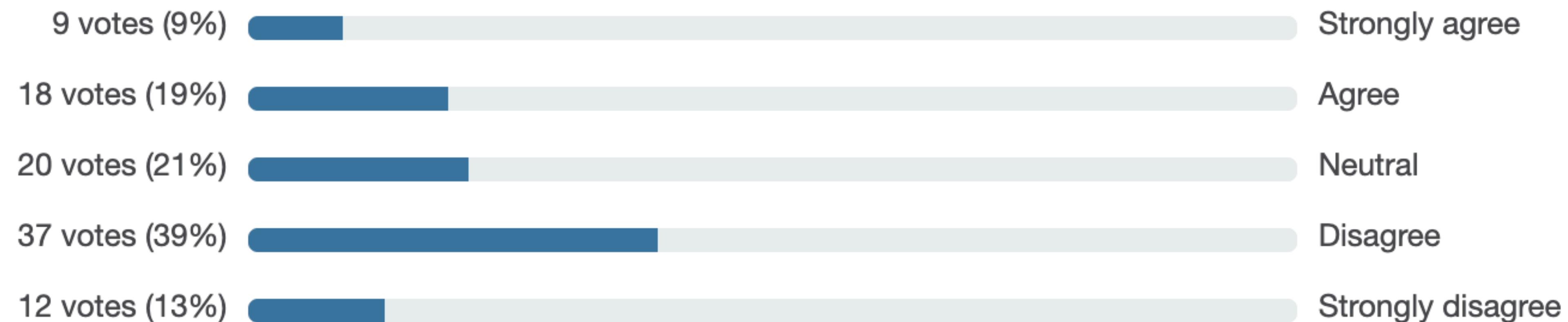
Whistle-Blowing

- Revealing a real or potential harm to the public being caused by your employer.
 - Not an attempt to take revenge on the employer, e.g., for turning down a promotion
 - Not an attempt to avoid personal responsibility for a problem that is about to come to light anyway.
- *Under what circumstances is whistle-blowing morally justified? When is it morally necessary?*

Professional Ethics

“Computer scientists and data scientists should not aid in the development of autonomous weapon systems.”

A total of 96 voter(s) in 1536 hours



Case Study

- Software Recommendation
 - Sam Shaw sought advice on how to improve security of his business's local area network
 - Professor Jane Smith answered Sam's questions
 - She also recommends, NetCheks, a software package to him to identify security problems
 - She does not mention the “best buy” option
 - She does not mention that she is shareholder of NetCheks
- Did Professor Jane Smith behave ethically?

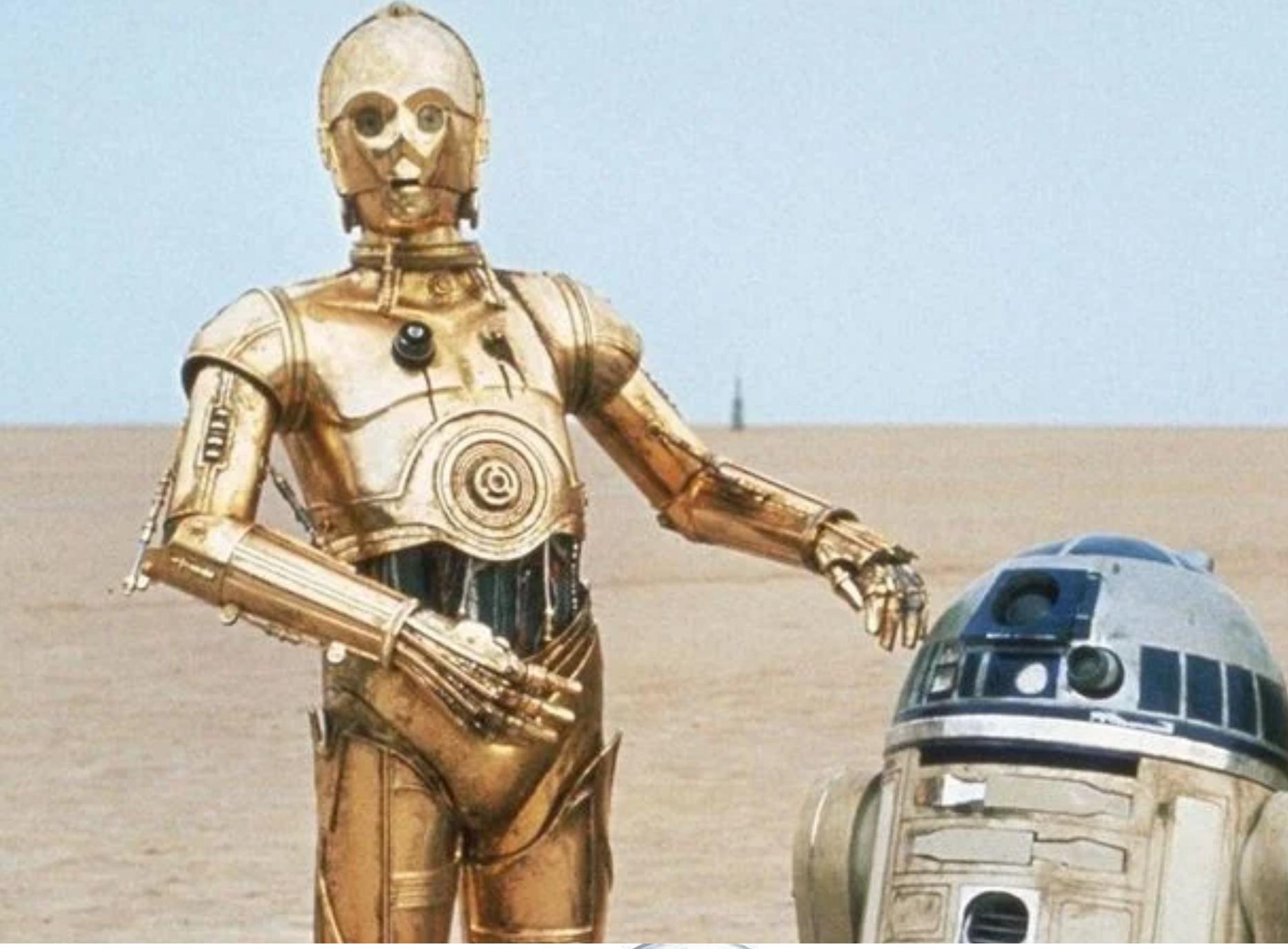
Case Study

- **Antworm**
 - The internet is plagued by a new worm that infected PCs by exploiting a security hole
 - Tim Smart created an antworm that exploited the same security hole, but fixed it
 - Tim released the antworm, taking precautions to ensure that it cannot be traced back to him
 - The antworm was effective, but looked just like another worm and the system administrators battled its spread.
- Did Tim behave ethically?

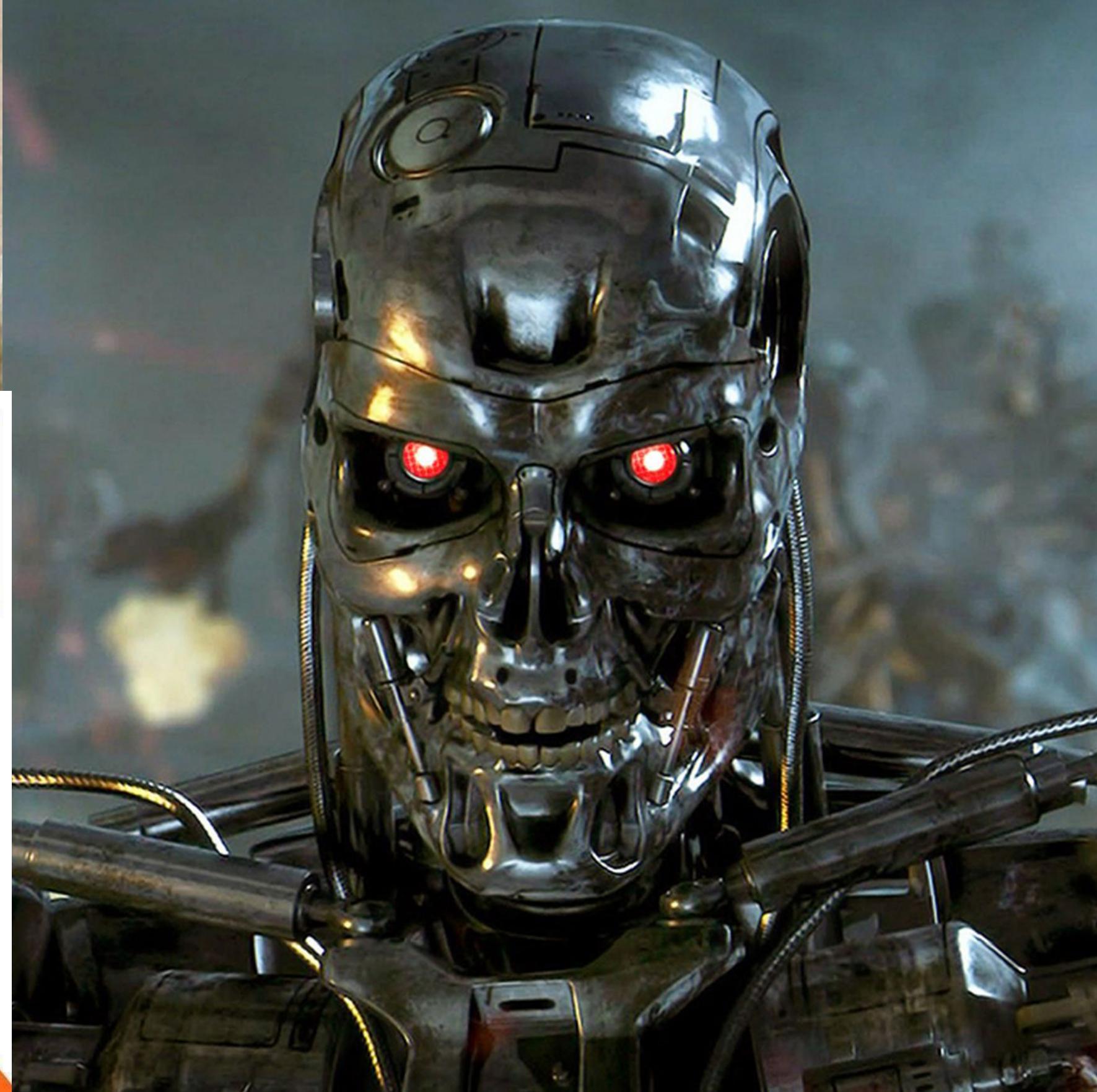
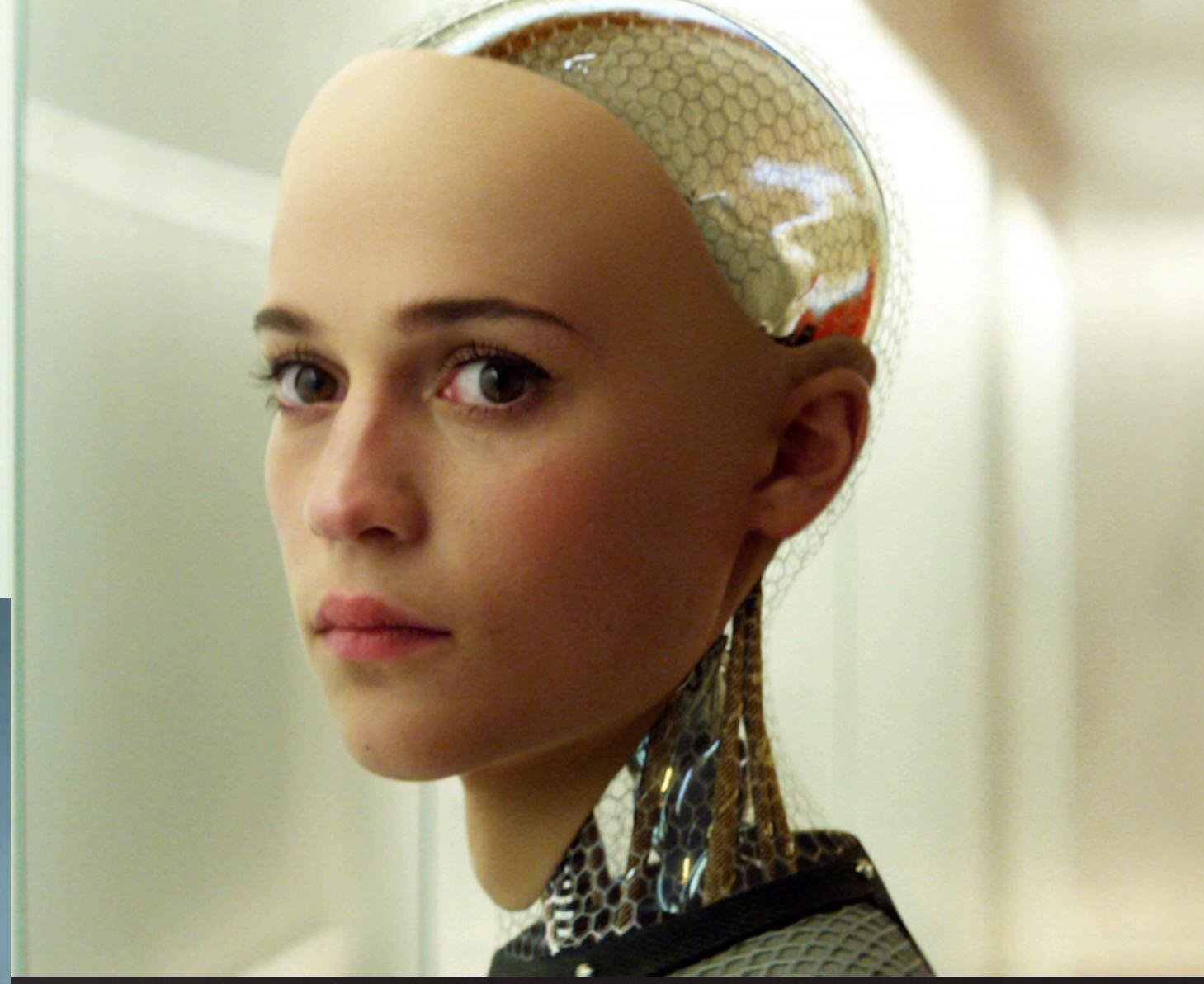
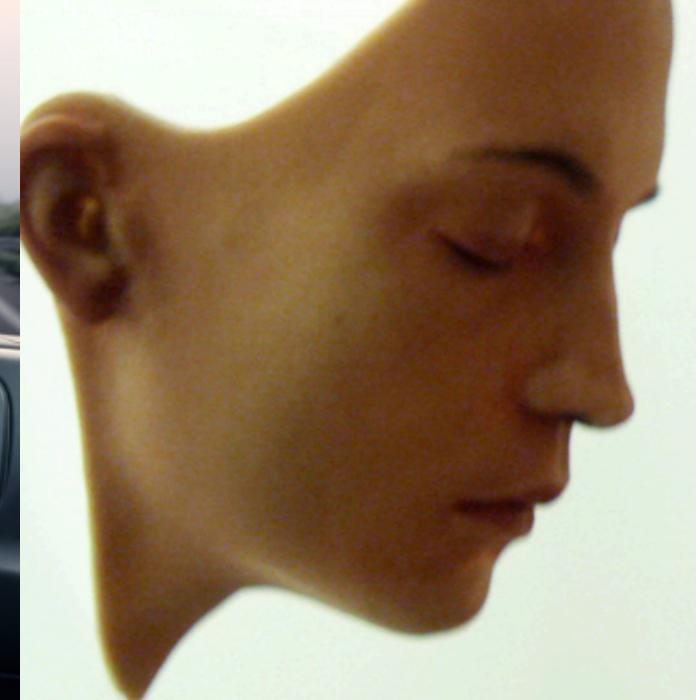
Break

Artificial Intelligence

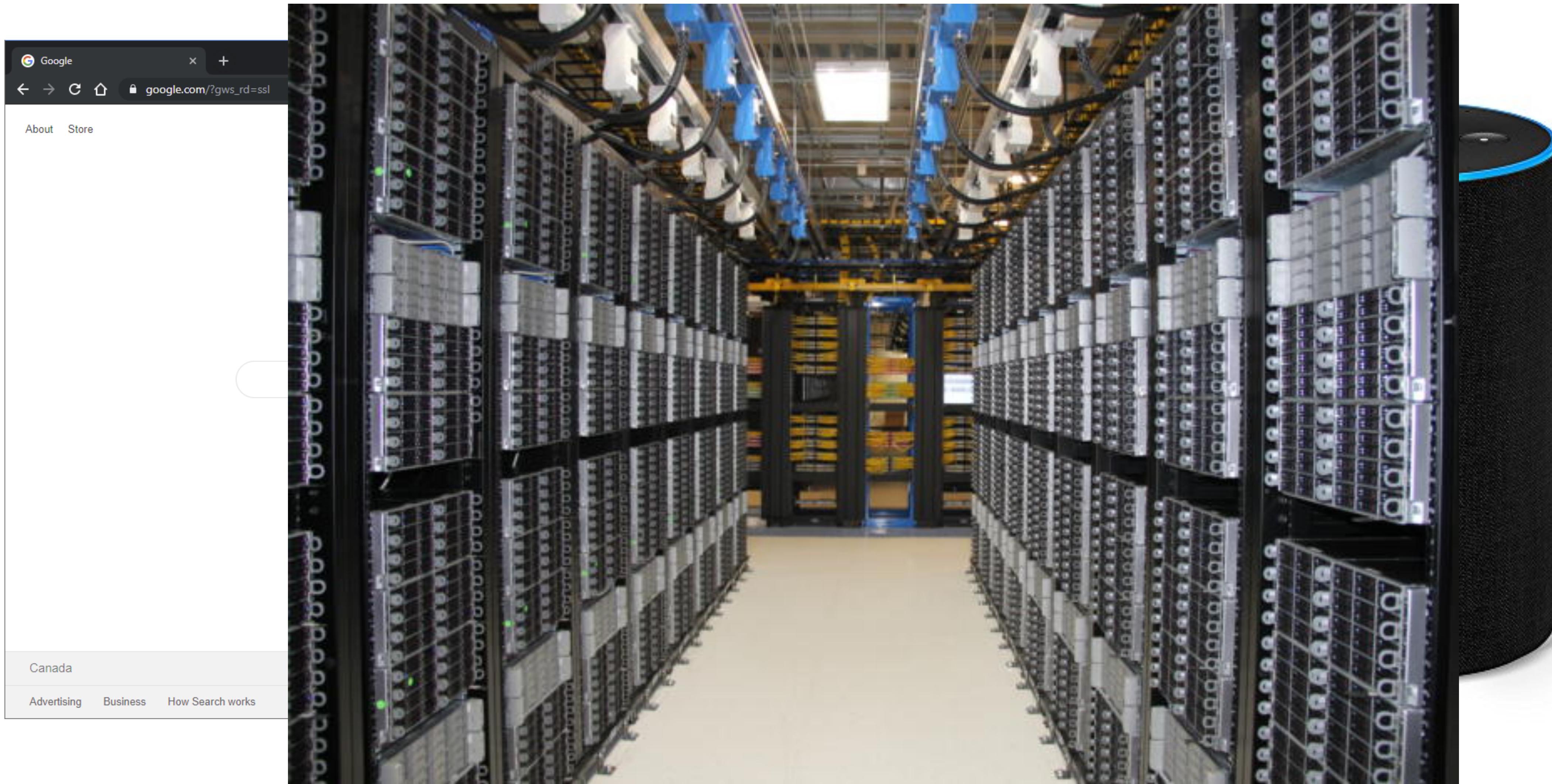
What AI Means to Most People



A magazine cover for Cosmopolitan India. The title 'COSMOPOLITAN' is in large orange letters at the top. Below it, the subtitle 'THE BEST THAT YOU CAN BE' is in smaller text. A woman with a robotic head and metallic wings on her shoulders is the central figure. The date 'MARCH 2018' and price '₹150' are at the top right. The year '2018' is at the bottom right. The text 'SOPHIA THE ROBOT' is in bold black letters, with the quote 'I'm Not Looking at World Domination' underneath. The word 'ONLINE' is at the bottom left.



But most modern AI systems look like...



What is Artificial Intelligence?

AI is the study, design, and development
of computational processes that solve problems
that **previously required human intelligence**

Not all of computer science is focused on AI.

But, it's hard to think of a field in CS that
isn't **contributing fundamentally** to it

The “**AI Paradox**”: once we become familiar
with a technology, we stop considering it AI

THE JOURNEY SO FAR

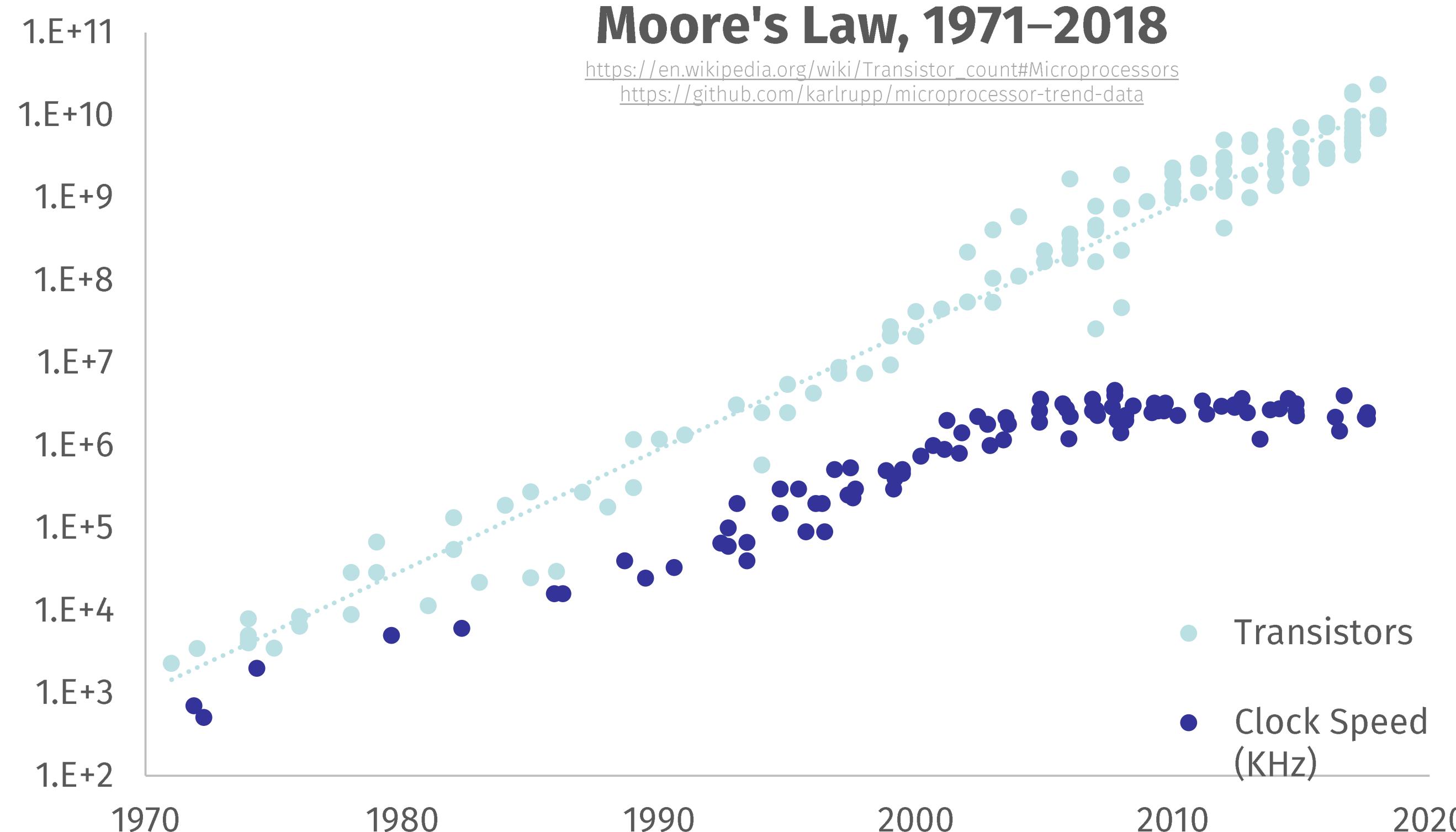
What is AI and what can it do today?
How much is real and how much is hype?



Why is it happening?

1. Scientific/mathematical **breakthroughs**, esp in machine learning
2. Growth in raw **computing power**

For comparison:



Species	# Neurons in Brain
Fruit Fly	10^5
Cat	10^9
Chimpanzee	10^{10}
Human	10^{11}

What short term impacts should we expect?

What kinds of social impacts are they having?

CURRENT TRENDS

Current State of the Art

- **Predictive modeling** limited mostly by quality of training data
- **Image recognition, face recognition** becoming extremely reliable
- **Image generation** almost there
- **Language understanding & generation** maturing extremely quickly. Watch this space
- **Fueling scientific breakthroughs** from protein folding to astronomy
- **Robotics** remains a hard problem, but it's working well in some niche applications



Prompt: Here's a short rhyming poem by Dr. Seuss about Elon Musk's company, Neuralink which is building brain implants to achieve symbiosis with AI. Neuralink's mission statement is "If you Can't Beat'em, join'em". The poem has a structure typical of Dr. Seuss stories with simple alternate line rhymes. In the poem Elon explains why we need to merge with the AI to save humanity.

Neuralink, by Dr. Seuss

Response:

I have a friend, Elon Musk,
Who wants to join our brains to AI.
He thinks that humans can't compete
So he's gonna use AI to cheat.
They want to build a neural-link.
To link us to the AIs
And save humanity from its demise.
They want to put an AI in your brain.
They want to use your neurons as a frame.
They want to save humanity,
If you can't beat'em, join'em

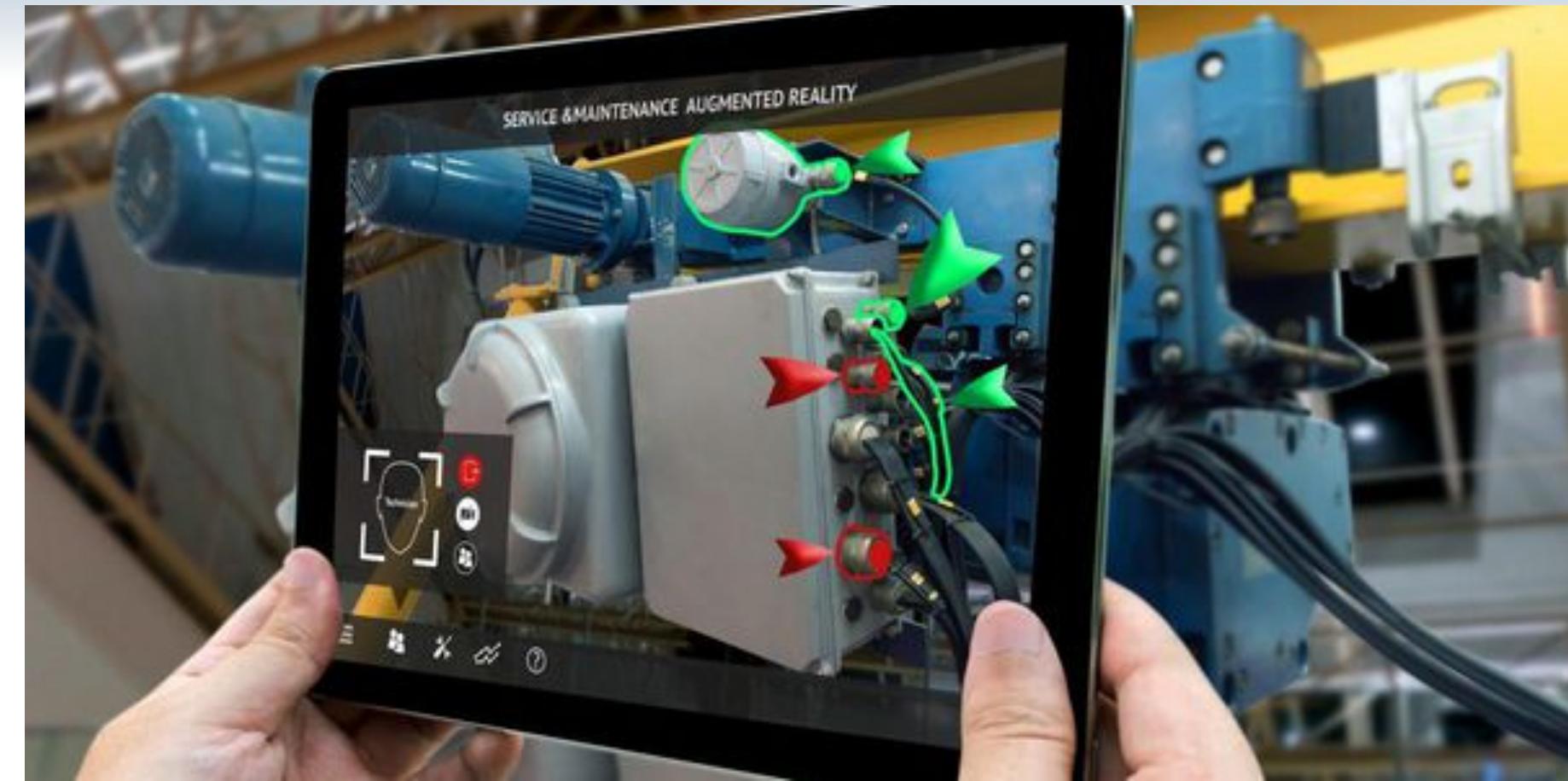
Machine Learning

- Narrow: building **predictive models** from data
- Broad: a new approach to computer programming, where computers **learn from experience** rather than being told what to do
- Where's it **headed?**
 - More and more data collection
 - sensors
 - behavior tracking
 - Image, text, speech understanding
 - Data mining/personalization/profiling
 - Self programming/optimization



Virtual and Physical Worlds Collide

- Smart **interfaces**
 - remote controls; microwaves; clocks; ...
- Augmented reality
- More intuitively “**programmable**” systems
- Much better **context awareness**
- **Assistance**
 - elderly; children; disabled; education
- **Social agents/games**



LLMs in the workplace

OpenAI report:

- How “exposed” are different occupations to automation with large language models (LLMs)?
- ~80% of the US workforce could have at least 10% of their work tasks affected by LLMs
- ~19% may see at least 50% of their tasks impacted
- Most affected tasks: writing and programming.
- Higher-income jobs (such as translators, tax consultants, and web designers) potentially face greater exposure

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou, Sam Manning, Pamela Mishkin, Daniel Rock

We investigate the potential implications of large language models (LLMs), such as Generative Pre-trained Transformers (GPTs), on the U.S. labor market, focusing on the increased capabilities arising from LLM-powered software compared to LLMs on their own. Using a new rubric, we assess occupations based on their alignment with LLM capabilities, integrating both human expertise and GPT-4 classifications. Our findings reveal that around 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of LLMs, while approximately 19% of workers may see at least 50% of their tasks impacted. We do not make predictions about the development or adoption timeline of such LLMs. The projected effects span all wage levels, with higher-income jobs potentially facing greater exposure to LLM capabilities and LLM-powered software. Significantly, these impacts are not restricted to industries with higher recent productivity growth. Our analysis suggests that, with access to an LLM, about 15% of all worker tasks in the US could be completed significantly faster at the same level of quality. When incorporating software and tooling built on top of LLMs, this share increases to between 47 and 56% of all tasks. This finding implies that LLM-powered software will have a substantial effect on scaling the economic impacts of the underlying models. We conclude that LLMs such as GPTs exhibit traits of general-purpose technologies, indicating that they could have considerable economic, social, and policy implications.



The report answers the question “which jobs do people think LLMs will automate” and not “which jobs will LLMs automate”.

LLMs in the workplace

- Adoption of Large Language Models for various tasks, including work-related ones
 - Some people compare it to the advent of Google
 - Some issues (ownership of training data, hallucinations...)
- Have you used LLMs before?
- Do you think LLMs should be used in the classroom? How?
- Do you think that AI will increase the expectations from professionals in the workforce? How?

Growing Military Use

- Increasing use of AI in **military applications**
- Already quite **commonplace** in some applications
 - smart bombs
 - surveillance systems
 - missile defence (Patriot, Iron Dome, etc)
- **Autonomous weapons** are becoming practical
 - drones are widely used; becoming more autonomous
 - land-based robotic systems are improving quickly
 - swarms of drones could revolutionize urban warfare
- The main barrier to adoption is **public opinion**



Break

PEERING INTO THE CRYSTAL BALL

Bold predictions about where AI is headed in the longer term

Technology that will be in widespread use

- Tailored solutions for **specific tasks**, not general intelligence
- Prototypes **that work today** in labs & narrow deployments
- Some examples:
 - **Non-text input modalities** (vision; speech)
 - **Consumer modeling** (recommendation; marketing)
 - **Cloud services** (translation; question answering; AI-mediated outsourcing)
 - **Transportation** (automated trucking; some self-driving cars)
 - **Industrial robotics** (factories; some drone applications)
 - **AI knowledge work** (logistics planning; radiology; legal research; call centers)
 - **Policing & security** (electronic fraud; cameras; predictive policing)



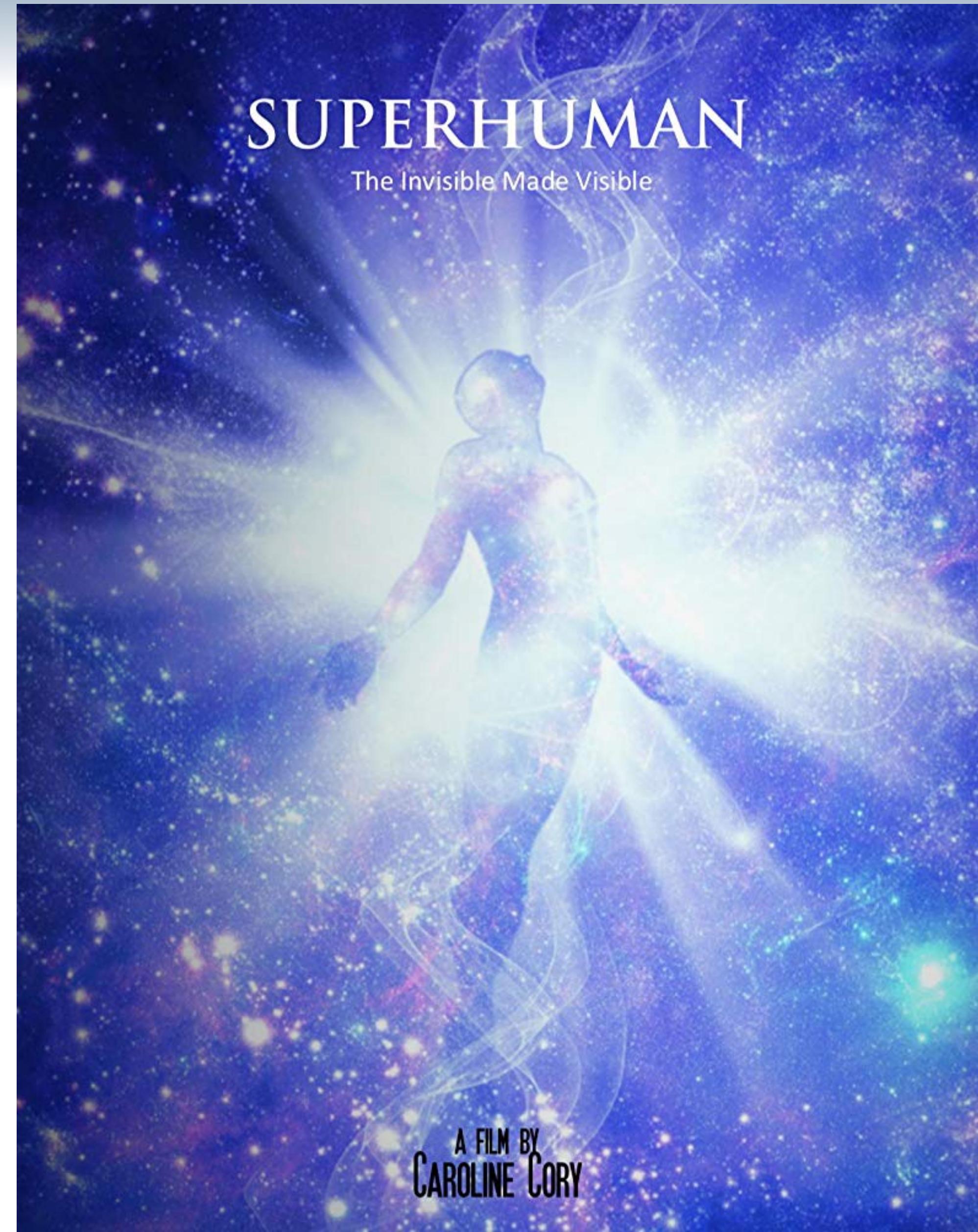
Technologies that won't take off as quickly

- Overall, areas in which
 - major entrenched **regulatory regimes** need to be navigated
 - there exist **social/cultural barriers** to adoption
 - the **human touch** is crucial
 - substantial **new hardware** would need to be developed
- Some **examples**:
 - childcare, healthcare, eldercare
 - education
 - coaching, counselling
 - consumer robots beyond niche applications
 - semantically rich language understanding



Superhuman Intelligence

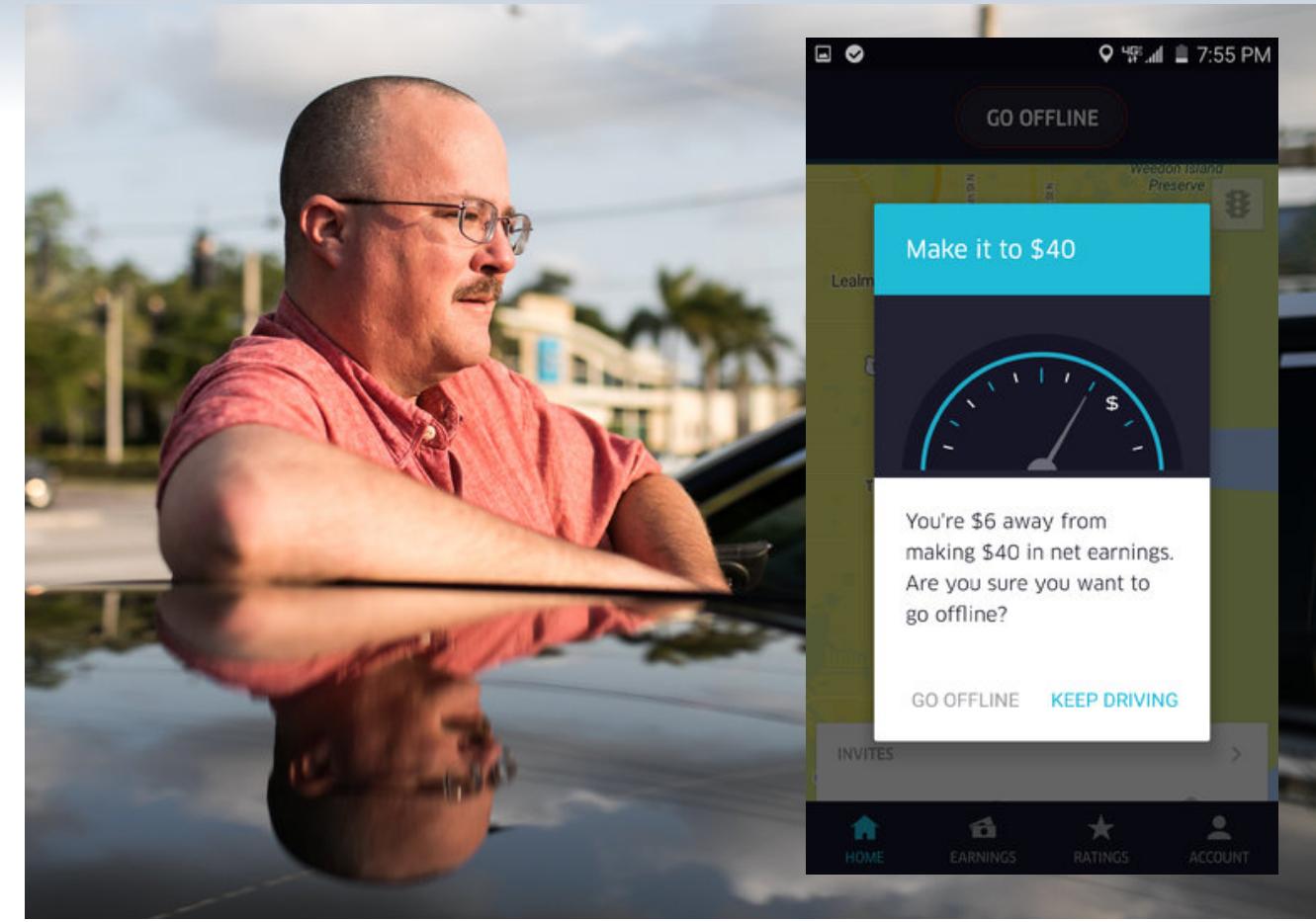
- AI systems will increasingly be capable of reaching **human-level performance**
- **Superhuman intelligence** isn't such a foreign, scary thing
 - governments, corporations, NGOs exhibit behavior much more sophisticated and complex than that of any individual
- Many important problems need superhuman intelligence; AI can help
 - improved **collective decision making**
 - more efficient use of **scarce resources**
 - addressing **underserved communities**
 - **climate change**; other societal challenges



Ethics of AI

Will a new technology:

- disempower **individuals vs corporations?**
⇒ user modeling; data mining; fostering addictive behaviors; developmental effects on children
- disempower **individuals vs governments?**
⇒ facilitate disinformation (deep fakes; bots masquerading as people; filter bubbles); enable qualitatively new military or security tactics
- take **autonomous actions** in a way that obscures responsibility
⇒ autonomous weapons; self-driving cars; loan approval systems
- disproportionately affect **vulnerable/marginalized groups**
⇒ automated decision making tools trained in ways that may encode existing biases



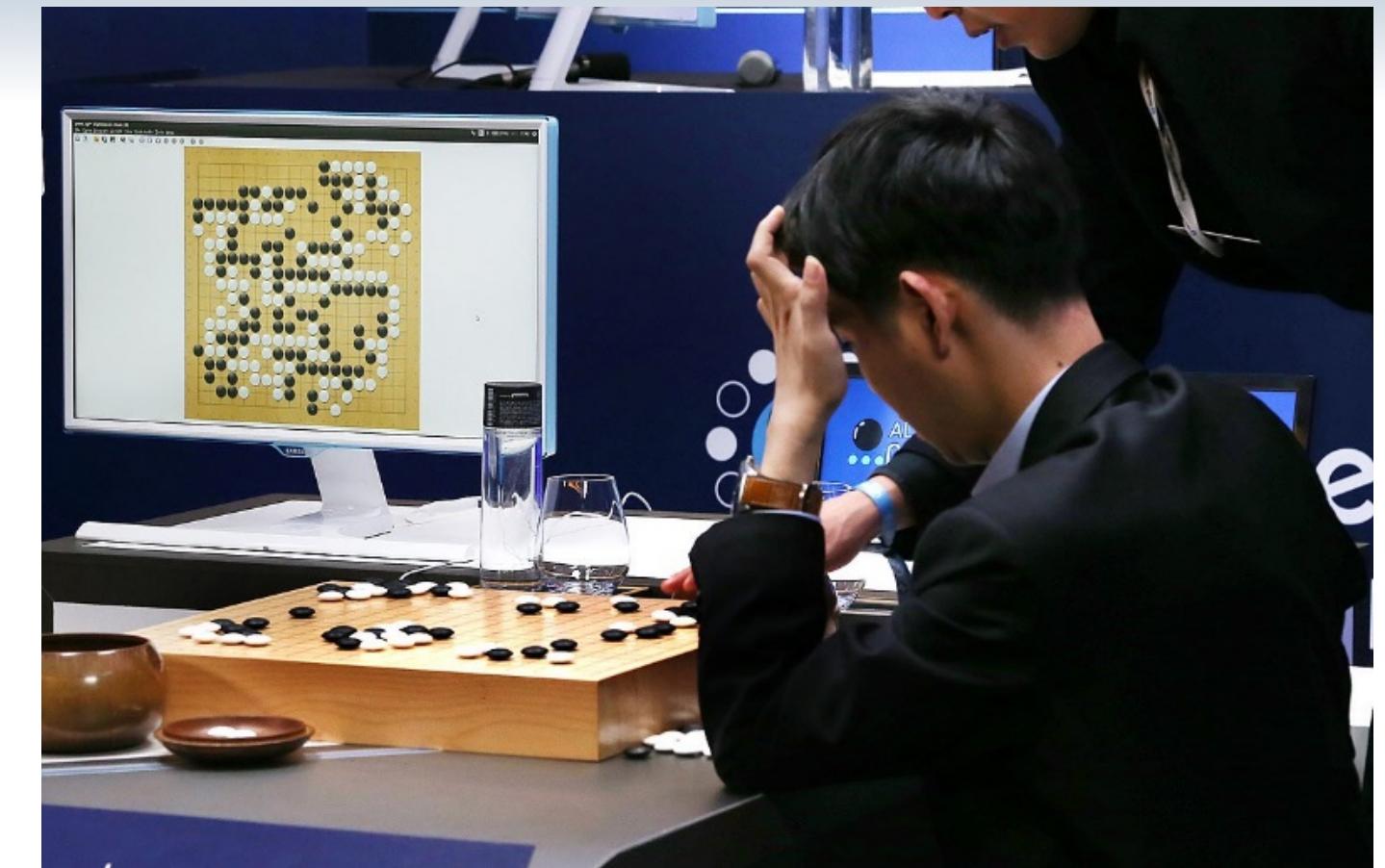
Social Impact

- How will AI technologies **transform society**?
- Will there be a **social backlash** against AI?
 - If so, what will be considered AI?
- This **generation of children** will grow up taking for granted many technologies that strike us as magical
- How will **human relationships** change in the presence of always-available social agents?
- As we are increasingly **augmented by AI**, what are our inherent cognitive/emotional/motivational limitations, beyond which augmentation won't help?



Ethical consideration of advances in AI

- *Is it wrong to create machines capable of making human labor obsolete? Will humans become demoralized by the presence of vastly more intelligent robots?*
- *How can we ensure that intelligent robots will not be put to an evil purpose by a malevolent human? How can we ensure they do not adopt malevolent purposes themselves?*
- *Is it morally acceptable to create “personal” (self-conscious) AI?*



Activity

Convention for Ethical use of Computer Science and Data Science techniques

The screenshot shows a Google Document interface with the following details:

- Title:** 2024S1 - CPSC 430 Class-wide Agreement
- Toolbar:** Includes File, Edit, View, Insert, Format, Tools, Extensions, Help, a clock icon, a message icon, and a Share button.
- Text Editor:** Shows a toolbar with search, navigation, print, and font/color tools, followed by a ruler from 1 to 7. The text "Purpose" is centered in bold black font at position 1.1.
- Text Content:**
 - Purpose Section:** "This term we have talked about many aspects of computer science and society with plenty of examples, case studies, and scenarios to guide your ethical thinking."
 - Callout:** "It is now time to come together as a class and come up with principles of the ethical use of computer science and data science techniques. **With the 40 students enrolled in the 2024S1 term, we will need to have unanimous agreement on any principles that are proposed.**"
- Footnote:** A vertical column of five small dashes is visible on the left margin, corresponding to lines 4 and 5 of the main text area.

Reminders before next class