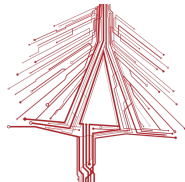
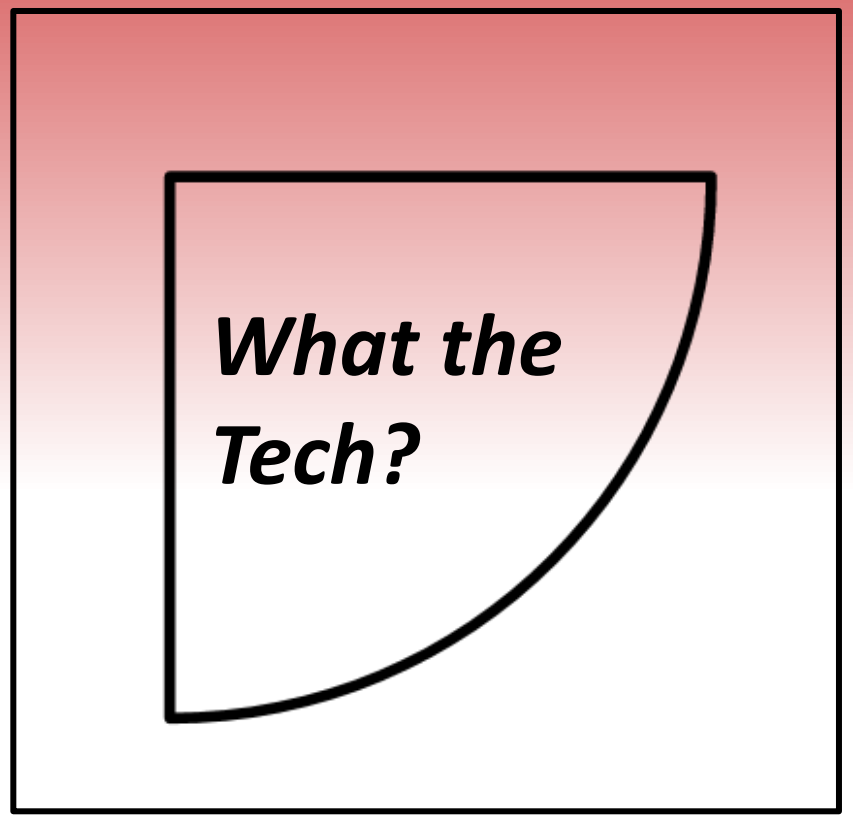
What the Tech?

PART 3: ETHICS IN AI



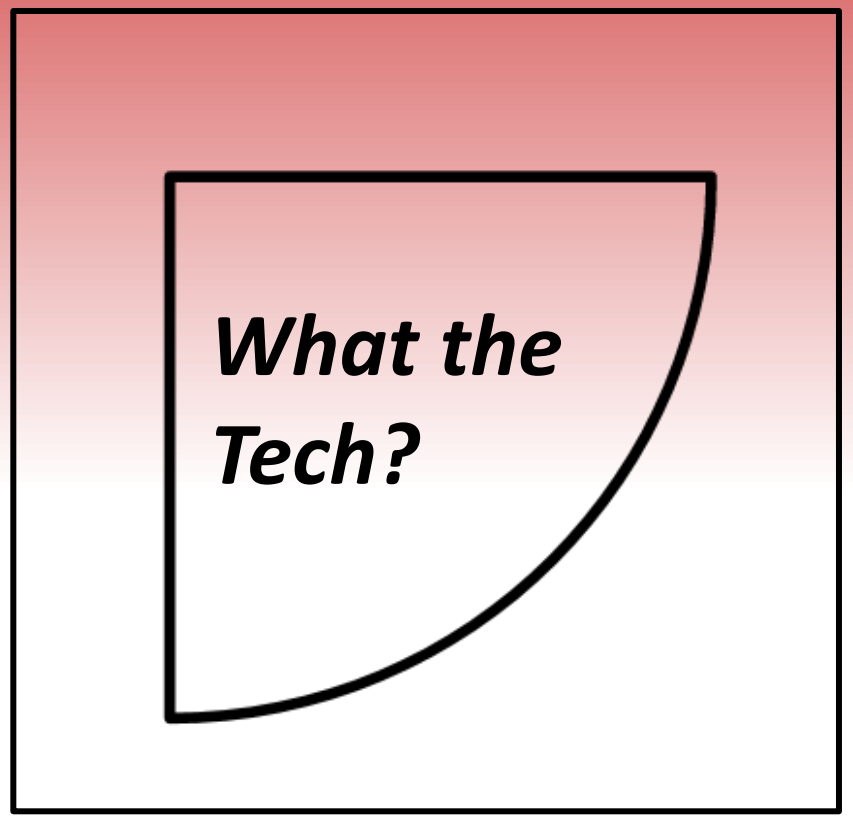
ETHICS IN AI



Last time, we discussed…

- How AI functions
- Key terms in AI
- Tech and AI on campus
- Shared our first weekly reflections!

ETHICS IN AI



This section is about Ethics in AI

Today we will…

- Share second round of weekly reflections
- Talk about Ethics
  - Overview
  - In AI (different settings)
- Small group activity

Questions so far?

Next: Our Weekly Reflections

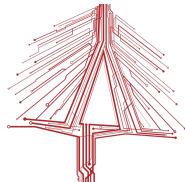
Weekly Reflections

How it works

- Goal: everyone goes once!
- Three volunteers per week
- Three new volunteers each week

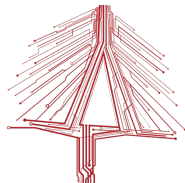
Every Monday

- Week 2:
- Week 3: (Today)
- Week 4:
- Week 5:
- Week 7:

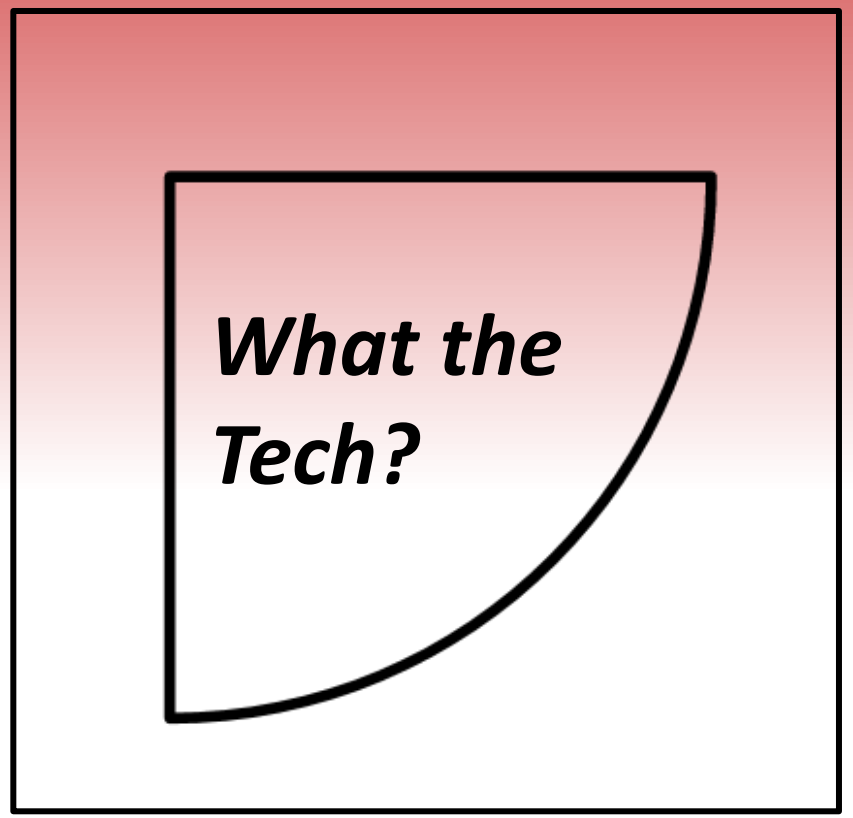


Questions so far?

Next: Ethics in AI (Overview)



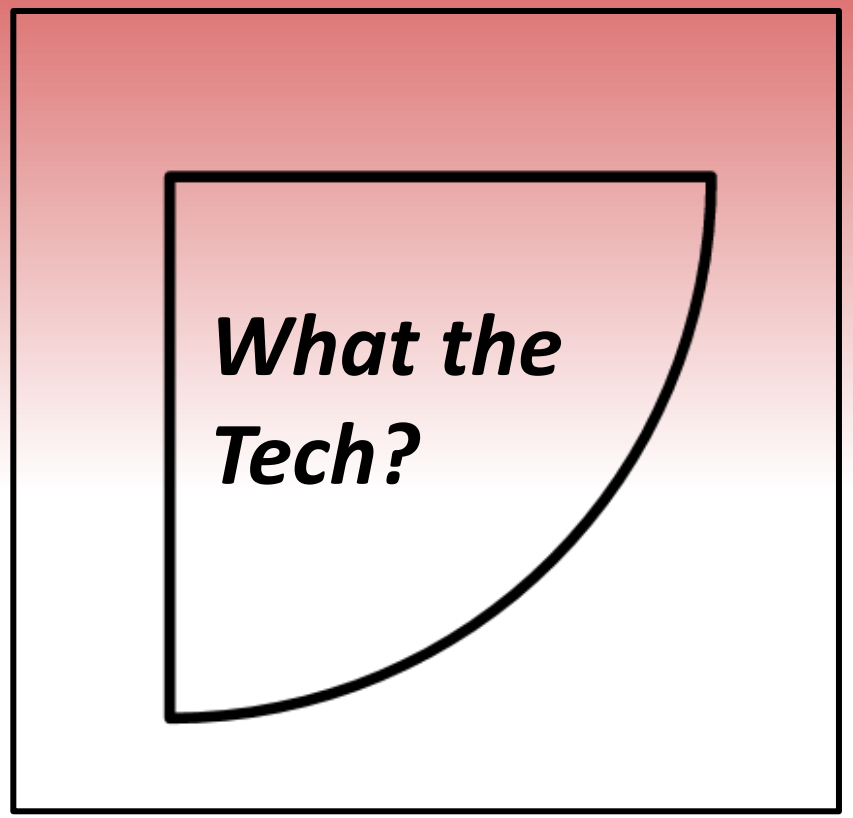
ETHICS IN AI: OVERVIEW



“What *are* ethics?”

Now… what do all these answers have in common?

ETHICS IN AI: OVERVIEW



“Is It Right or Wrong?”

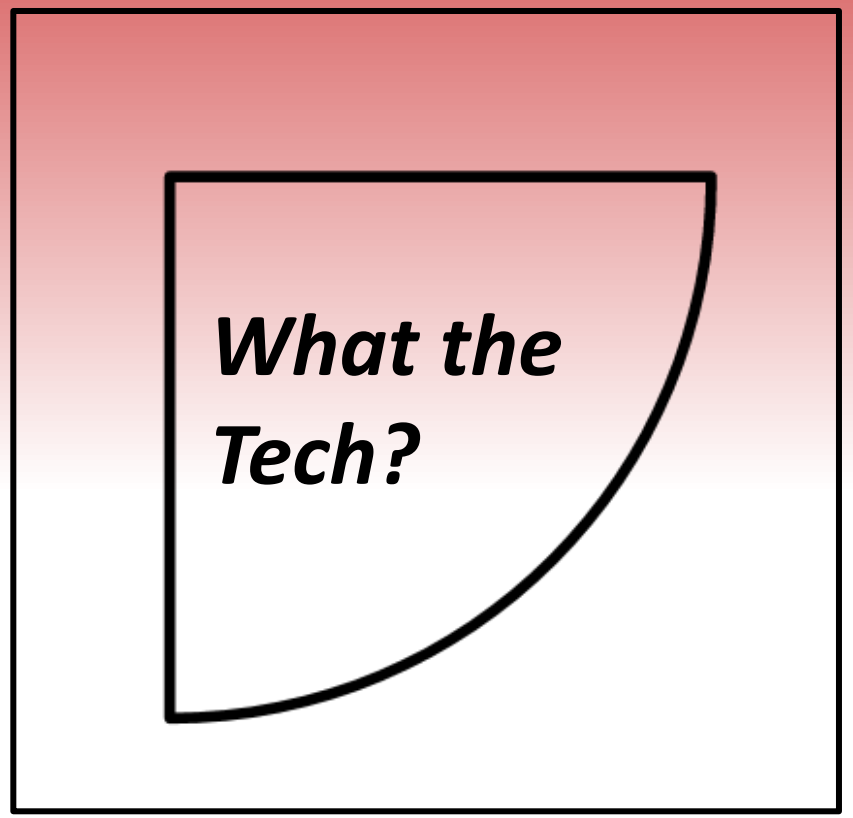
- The number one question in life
  - Directly or indirectly
- We learn this stuff as kids
- Most big and small questions are influenced by ethics

- It’s a uniquely human trait

ETHICS IN AI: OVERVIEW

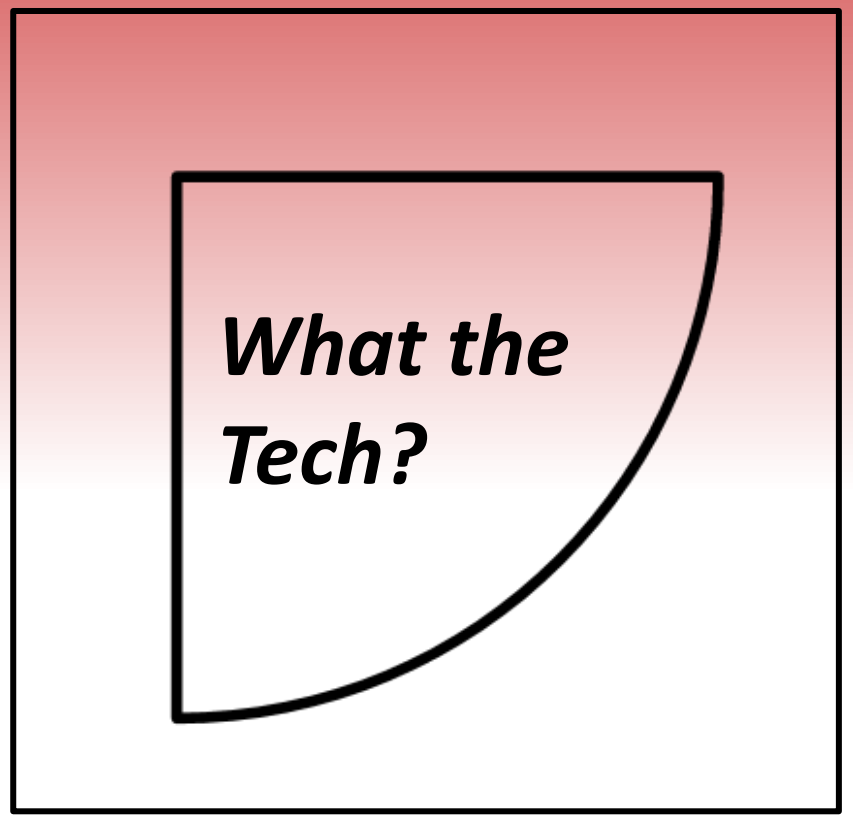
What the Tech?

ETHICS IN AI: OVERVIEW



“Nuance”

ETHICS IN AI: OVERVIEW

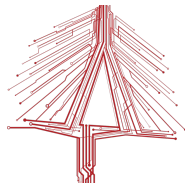


Nuance is…

- Something we learn to identify through “lived experience”
  - Same as right, wrong, and ethics
- Basis for decision-making
  - But what about in technology?
  - What about *AI?*
    - AI is very literal

Questions so far?

Next: Risks



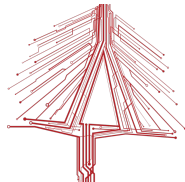
NOTHING IS RISK-FREE

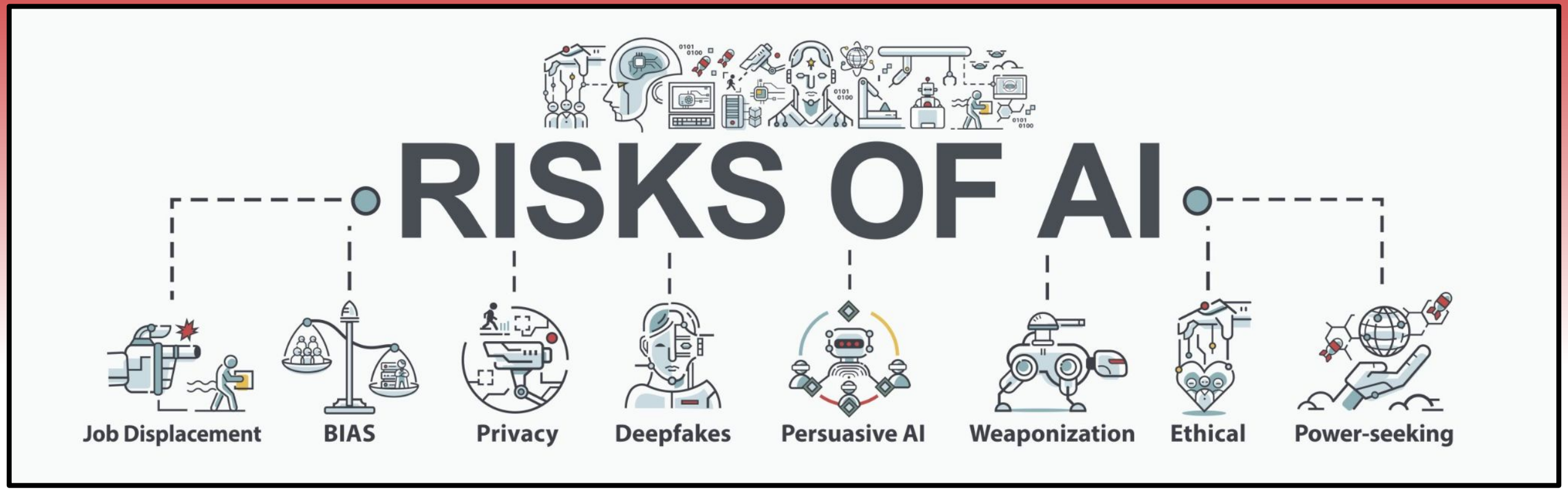
- All technology comes with the potential for harm and for good.
  - Earliest example: discovery of fire and inventing the wheel.
- But with AI, the stakes are higher.
  - We live in the age of deep fakes, fake news, and rising *automation*.
- This means we have to be able to tell the difference between…

Being smart enough to use AI

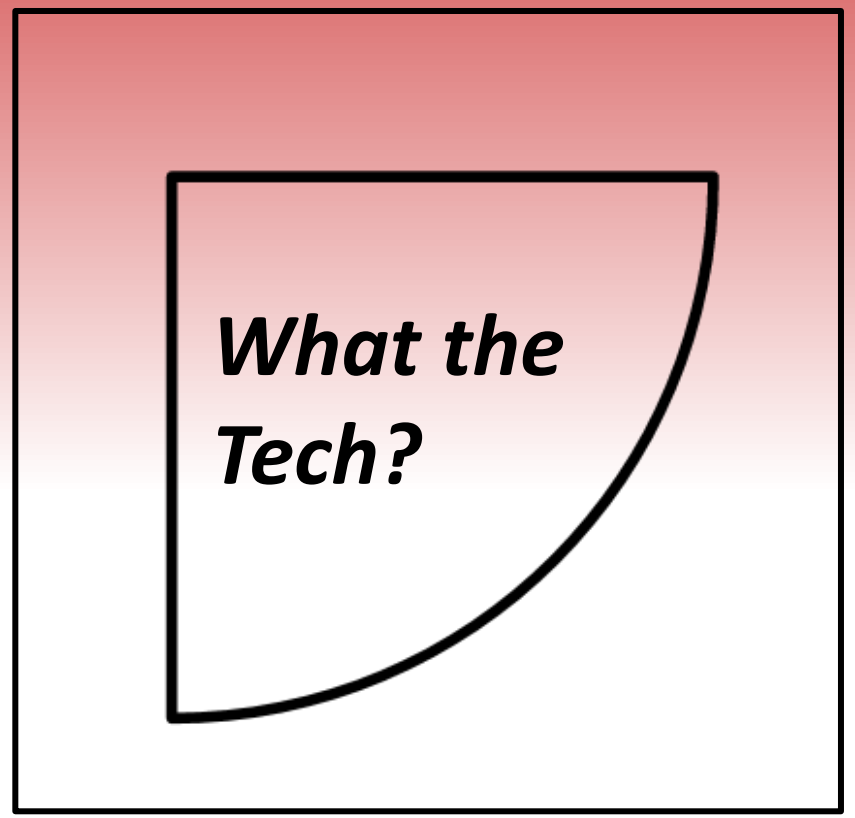
vs.

Being wise enough to use AI




RISKS OF AI
Job Displacement
BIAS
Privacy
Deepfakes
Persuasive AI
Weaponization
Ethical
Power-seeking

ETHICS IN AI: RISKS

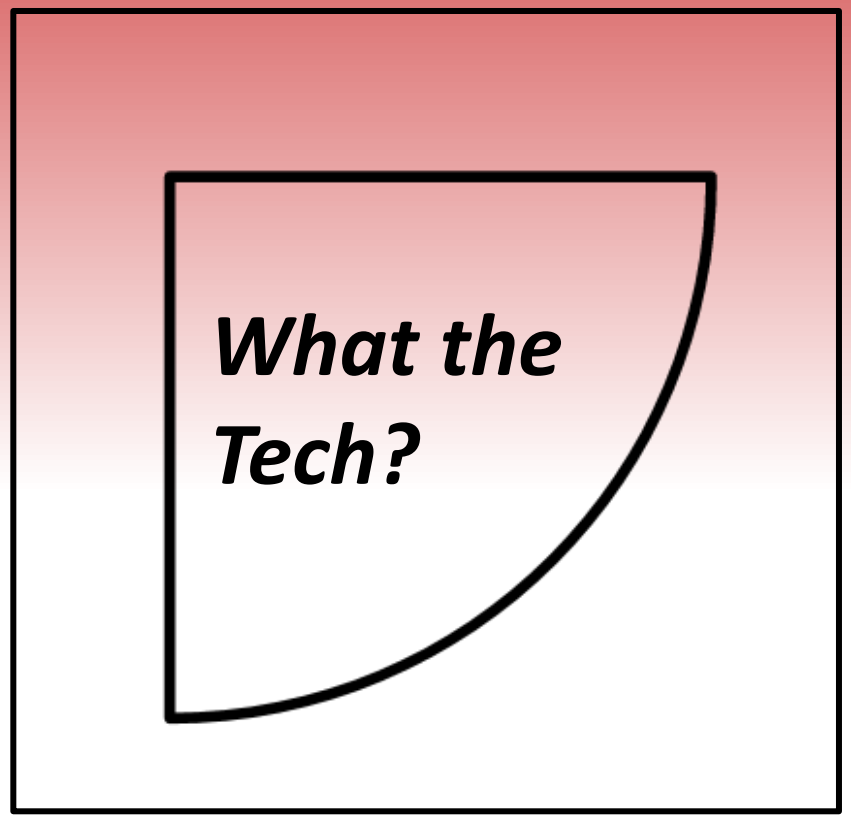


Some examples…

- **Unequal access**
- **Environmental impacts**
- **Unanticipated impacts**
- **Algorithmic bias**

- Over Reliance on tech
- Hacking and privacy
- Job loss

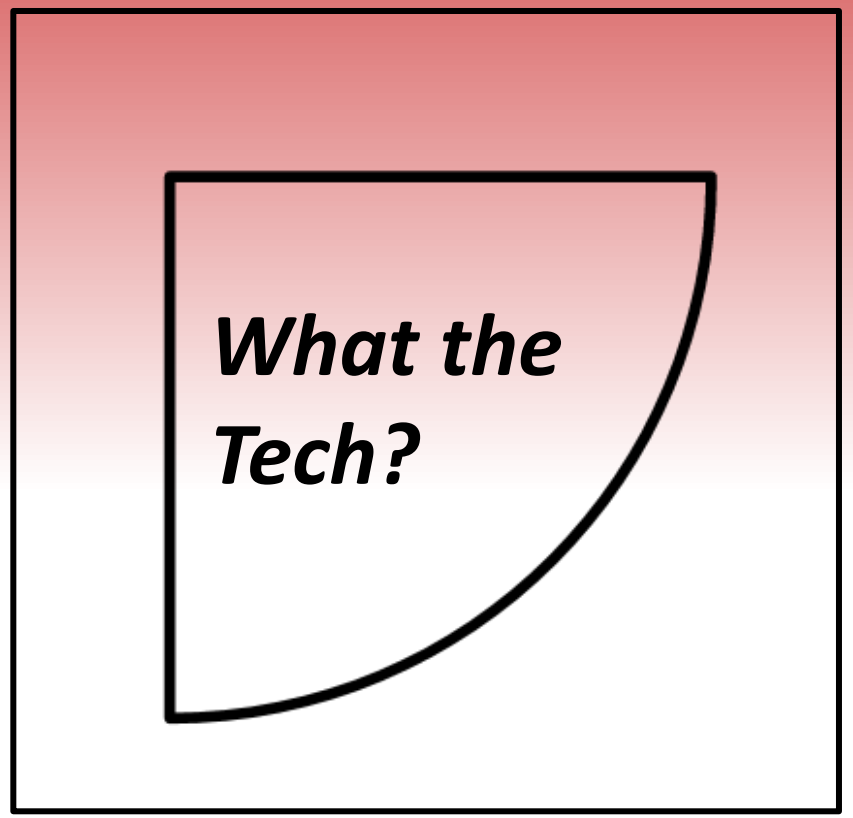
ETHICS IN AI: RISKS



Unequal Access

- Initial access to new technology is often limited.
- Several reasons for this:
  - It’s expensive
  - Not many people know how to use it
  - Limited applicability/practicality
- There are plenty of examples of this:
  - Television sets
  - Mobile phones
  - Cars
  - Other cases?
- AI is no different.

ETHICS IN AI: RISKS

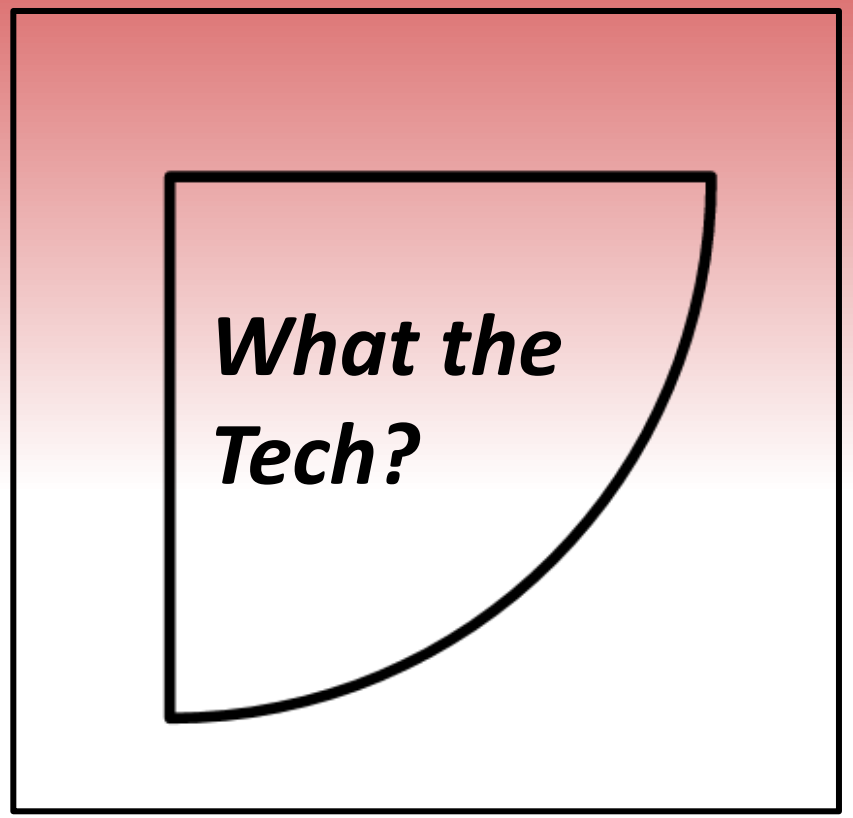


(Hypothetical) Examples of Unequal Access in AI

- Communities: Do some cities have the financial ability to build superior predictive AI models for future planning?
  - Our focus next week

- Education: Who knows how to use ChatGPT (and related tools) to enhance their work? What if it was trained on language different from your own?
  - Our focus the week after next

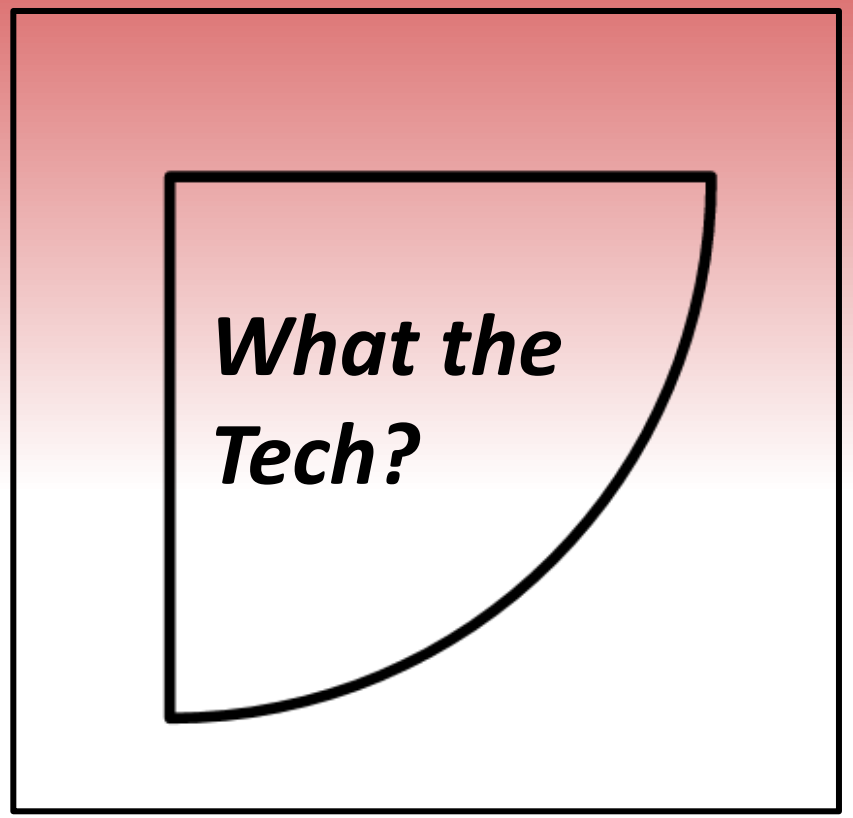
ETHICS IN AI: RISKS



Environmental Impacts

- AI has a very large impact on our natural resources. Can you guess a few?
  - CO2 emissions
    - GPT-3 produced the equivalent of around 500 tons of carbon dioxide.
  - Fresh water usage
    - Servers need water to stay cool
    - Google’s 2023 water usage rose by 20%, thanks in part to its AI work

ETHICS IN AI: RISKS

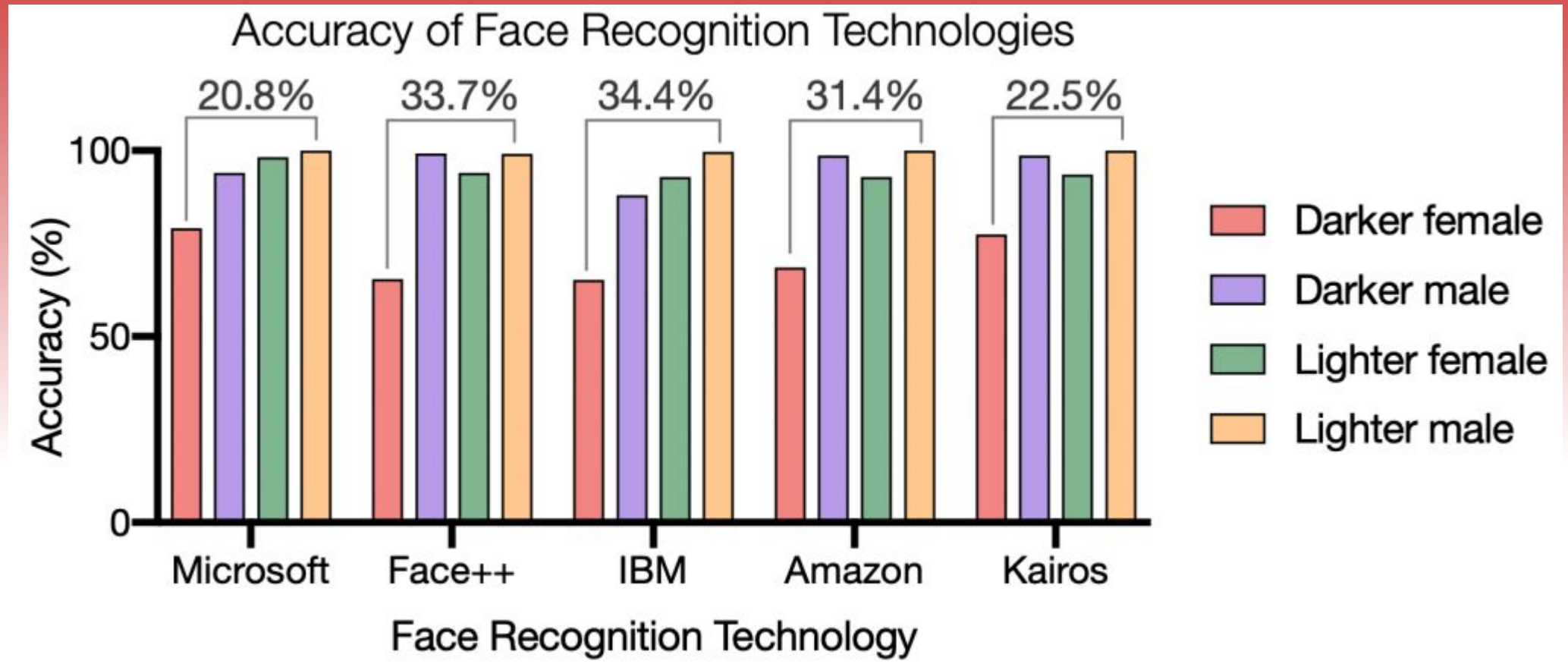


Algorithmic bias

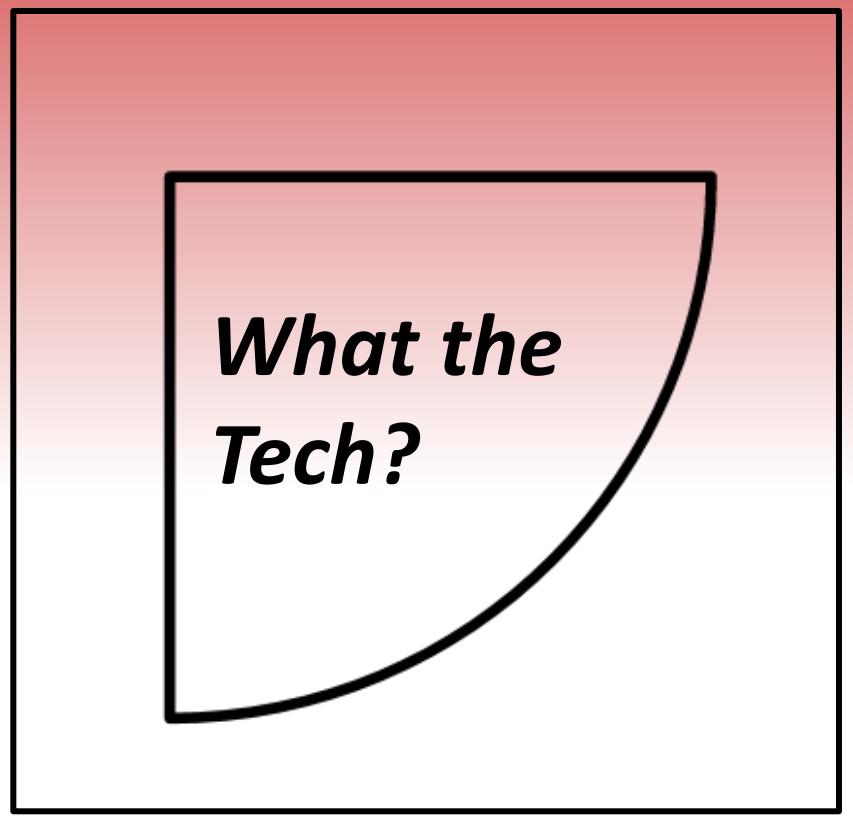
- Algorithmic bias occurs when algorithms make decisions that systematically disadvantage certain groups of people.
  - VIDEO: Gender Shades by Joy Boulamwini

Gender Shades by Joy Boulamwini

Gender Shades by Joy Boulamwini

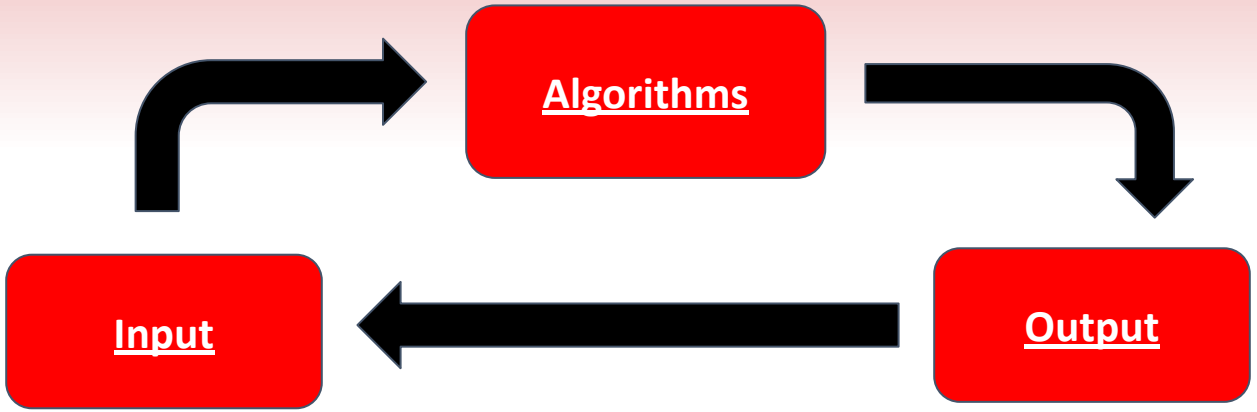


ETHICS IN AI: RISKS

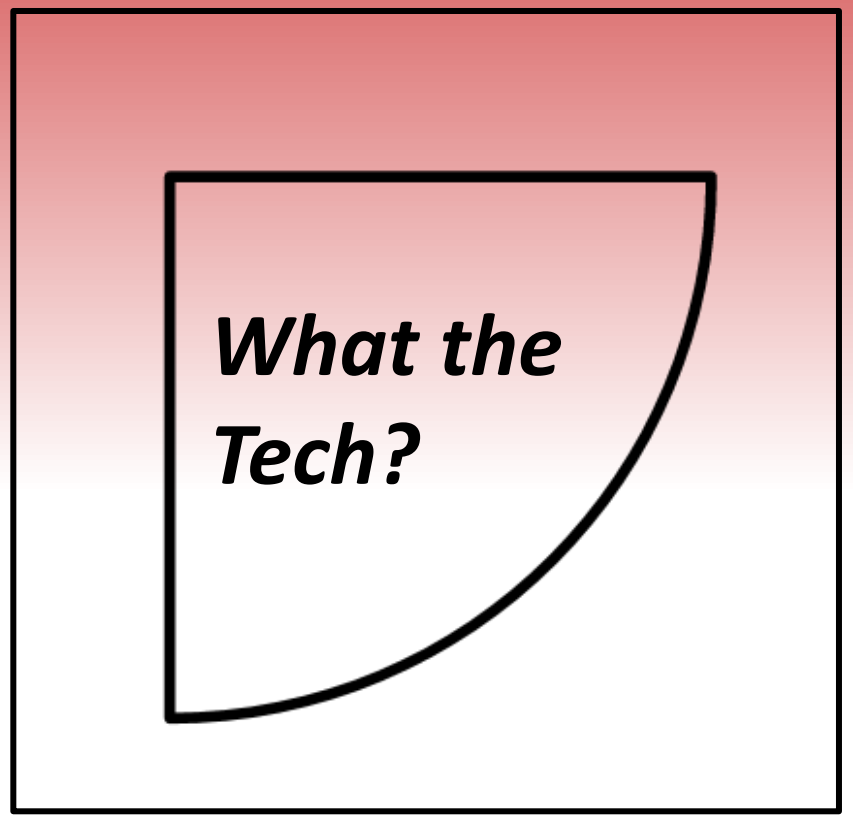


Algorithmic bias

- The risk for bias in AI is rooted in training data
- You only see what you've learned!
- This can lead to…



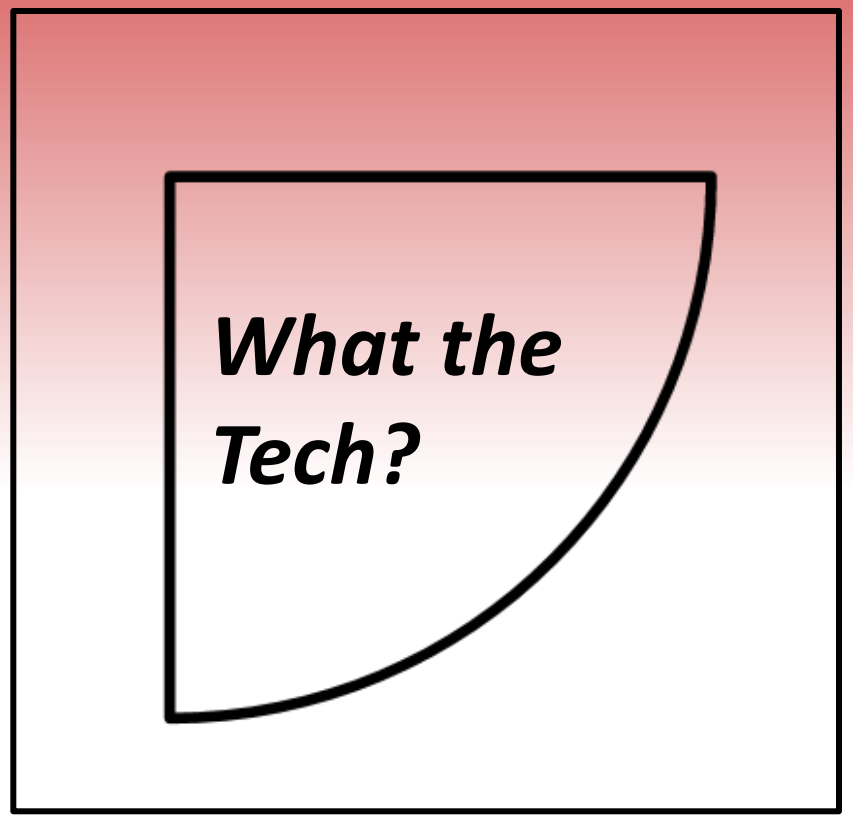
ETHICS IN AI: RISKS



Unanticipated Impacts

- **“Deep Fakes”:** A politician creates an “endorsement” from someone who doesn’t support them.
- **Mistaken predictions**: What happens if a crime is “predicted” incorrectly?

ETHICS IN AI: RISKS



We looked at these…

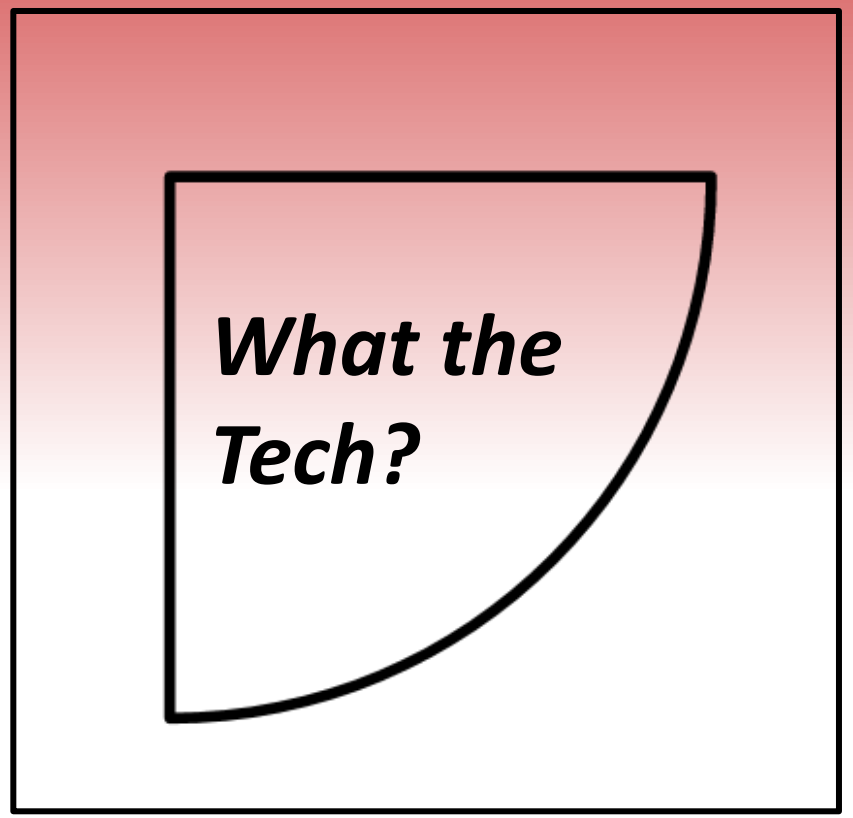
- **Unequal access**
- **Environmental impacts**
- **Unanticipated impacts**
- **Algorithmic bias**

And we will touch on these in the future…

- Over Reliance on tech
- Hacking and privacy
- Job loss

But that's not all… what else is there?

ETHICS IN AI: RISKS



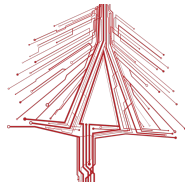
Before the break…

- This is heavy stuff
- But AI isn't all doom and gloom
- Knowing the risks and rewards!
  - It's part of being a good user *and* a good person!

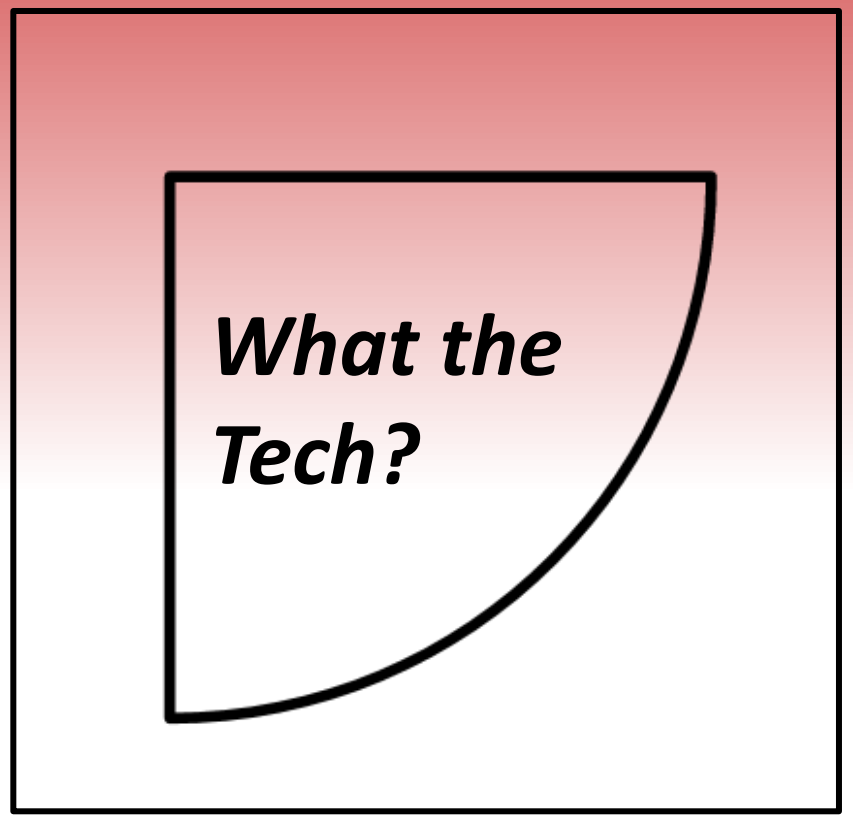
Questions so far?

Break Time

Next: Ethics in AI (Small group Activity)



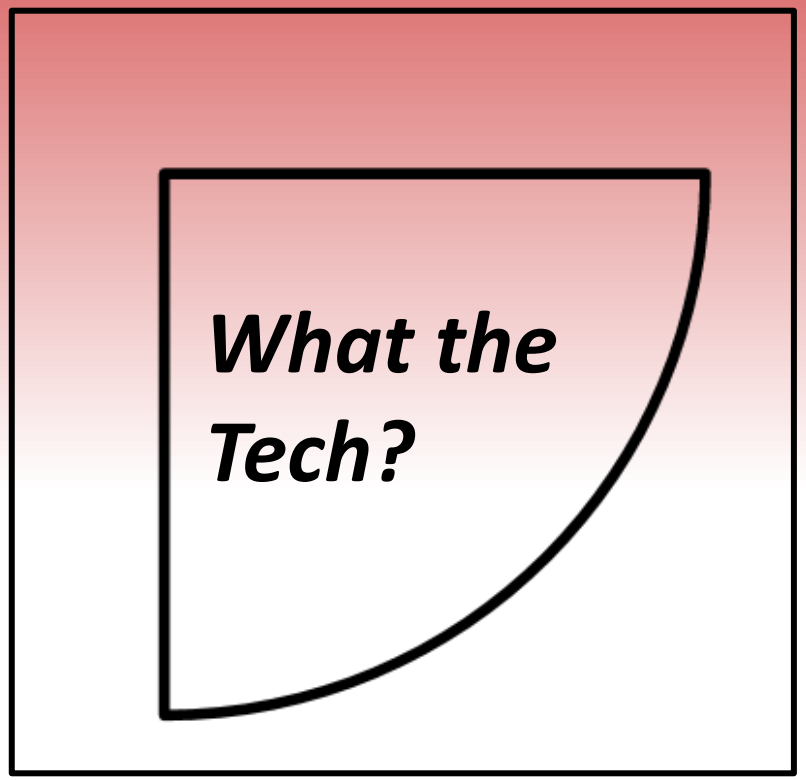
ETHICS IN AI: SMALL GROUP ACTIVITY



GET INTO GROUPS OF TWO/THREE

1. Pick an envelope at random
2. With your teammates, read the ethical dilemma and come up with answers to the questions listed
    a. Feel free to write on the back of your sheet
3. We will regroup in 20 minutes to share our answers

Next time:



- Deep dive
  - Ethics, AI, Technology
- Debrief
- Look ahead at next week
  - Predictive Analytics and AI

ETHICS IN AI: BONUS READING

NEWS > LOCAL NEWS

How AI-powered robots in law enforcement could become a tool for 'supercharging police bias'

LISTEN · 7:50 ALL THINGS CONSIDERED SHARE

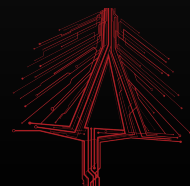
One of Boston Dynamics' Spot robots during a 2021 demonstration at their facilities in Waltham, Mass.

Josh Reynolds / AP

WGBH News: April 30, 2024

https://www.wgbh.org/news/local/2024-04-30/how-ai-powered-robots-in-law-enforcement-could-become-a-tool-for-supercharging-police-bias

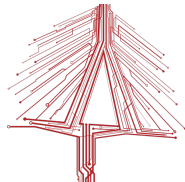
Halfway Point



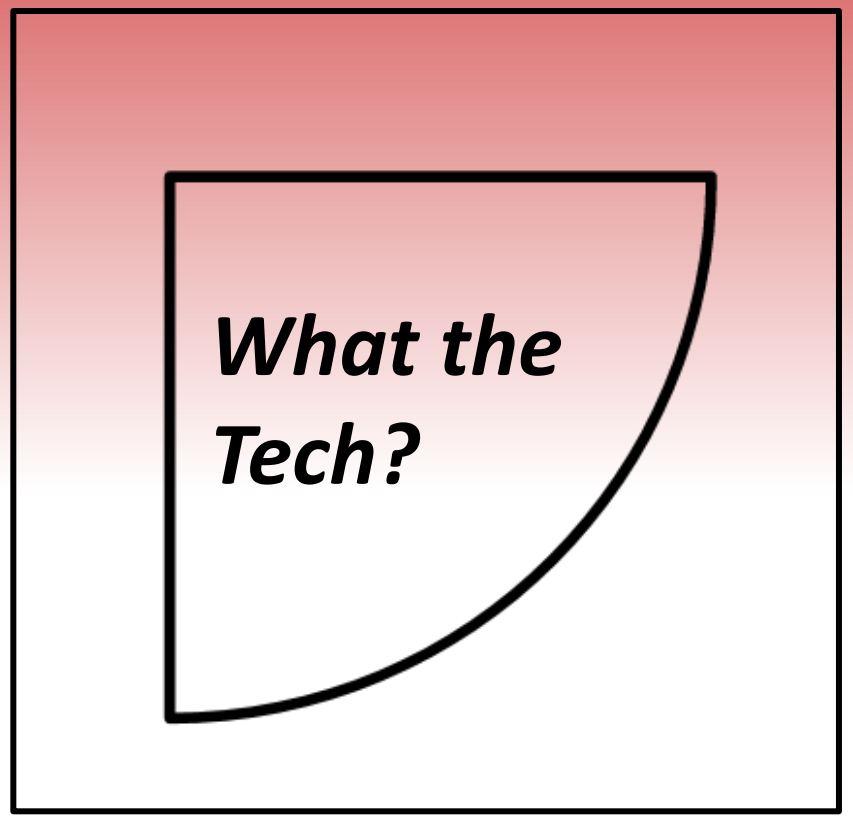
What the Tech?

Week 3, Part II:

DEEP DIVE: ETHICS IN AI



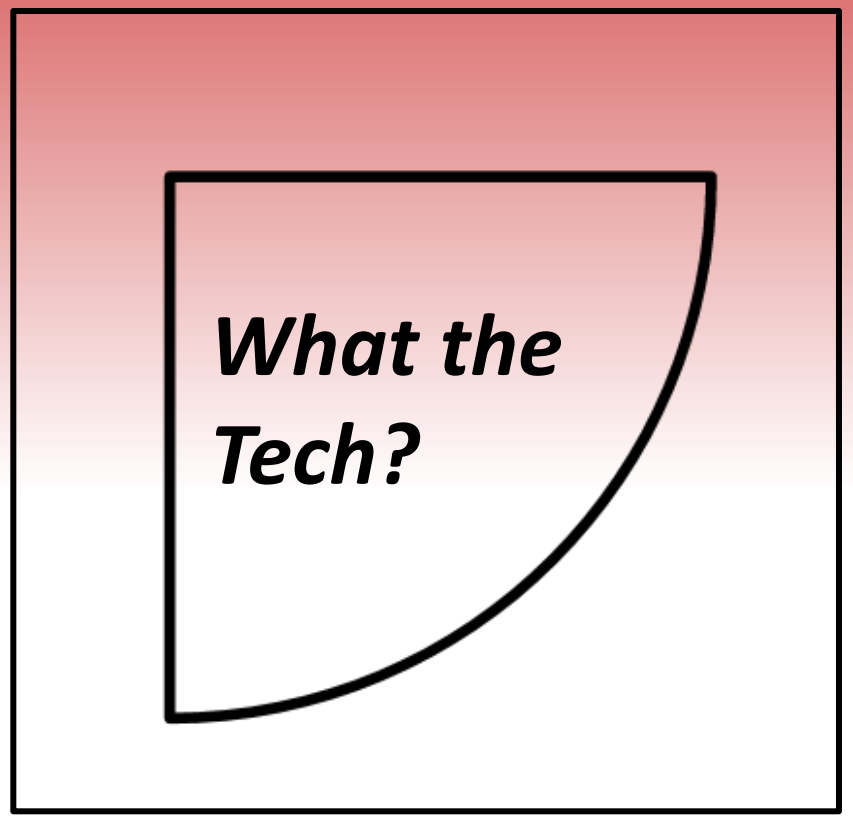
ETHICS IN AI



Last time, we discussed…

- Ethics
  - In general
  - In AI

ETHICS IN AI



Today…

- Deep dive: “Thinking and feeling machines?
  - Originally by Dr. Vance Ricks
- Debrief
- Any questions? Refreshers?

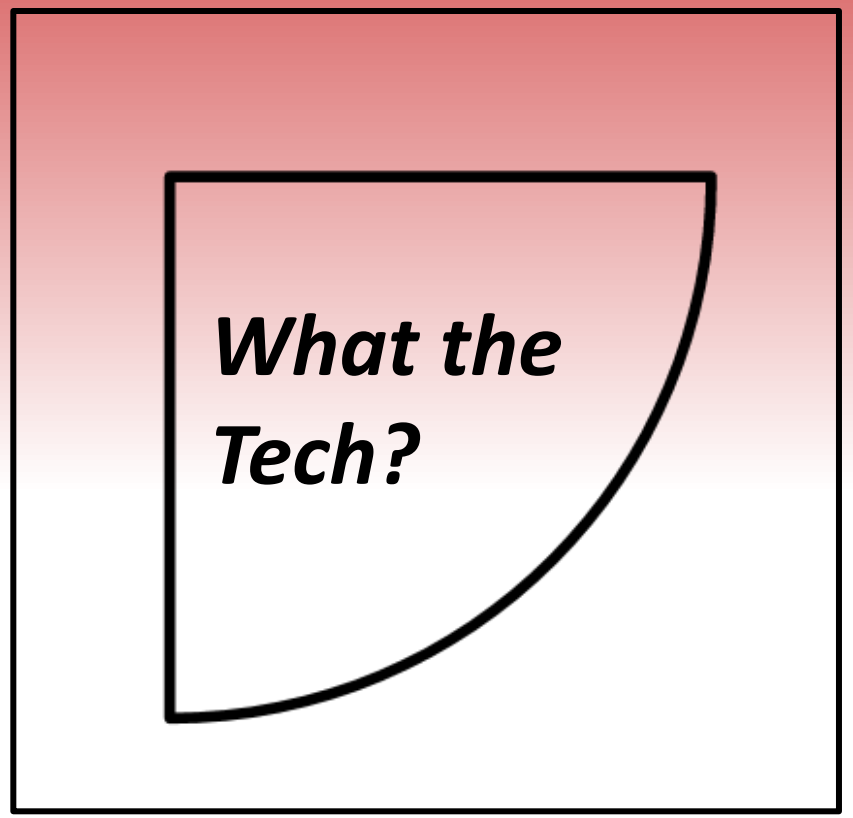
Northeastern University
College of Social Sciences
and Humanities

Thinking and feeling machines?

Prof. Vance Ricks
Departments of Philosophy and Computer Science
Northeastern University

What The Tech?
Wednesday 08 May 2024

ETHICS IN AI: ABOUT VANCE RICKS

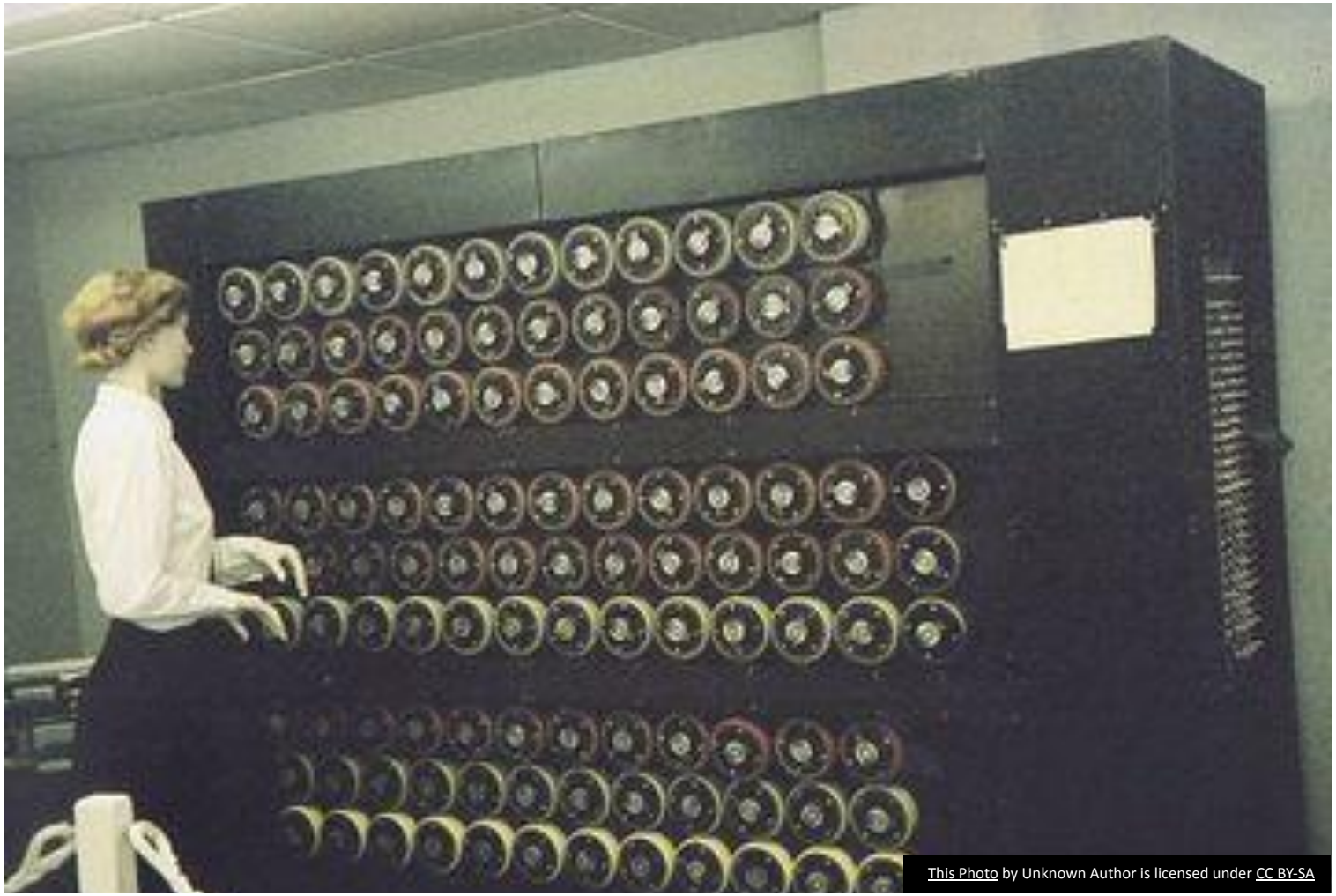


- Associate Teaching Professor of Philosophy and Computer Science
- Work focuses on *moral philosophy*
  - Spans from the works of John Stuart Mill in the 1800s to the ethics of today's digital technologies.

Technology is neither
good nor bad;
nor is it neutral.

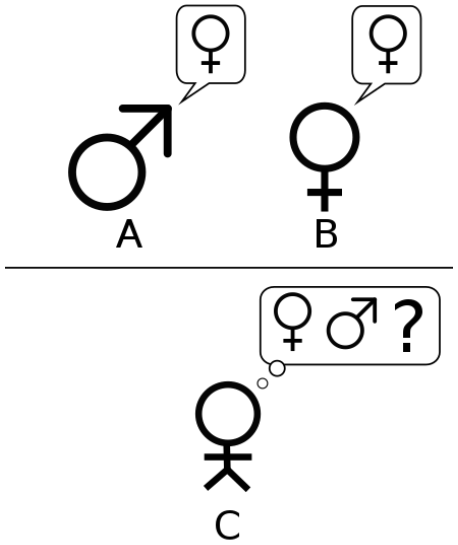
Melvin Kranzberg
First Law of Technology (1986)

Part 1 - Two classic 20th-c thought experiments about (the possibility of digital) “thinking machines”

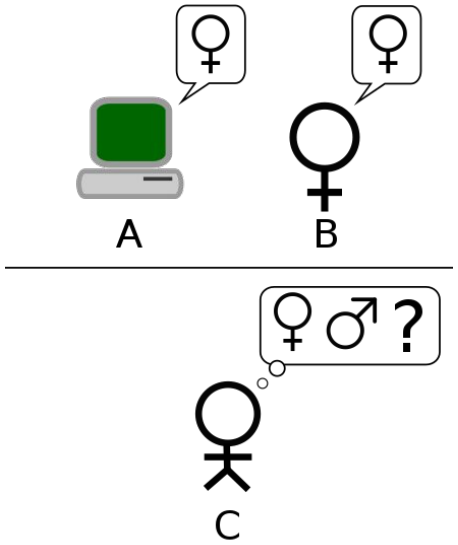


1. The Imitation Game(s) – “Can a machine think?”

Original version



Turing’s (Main) Version



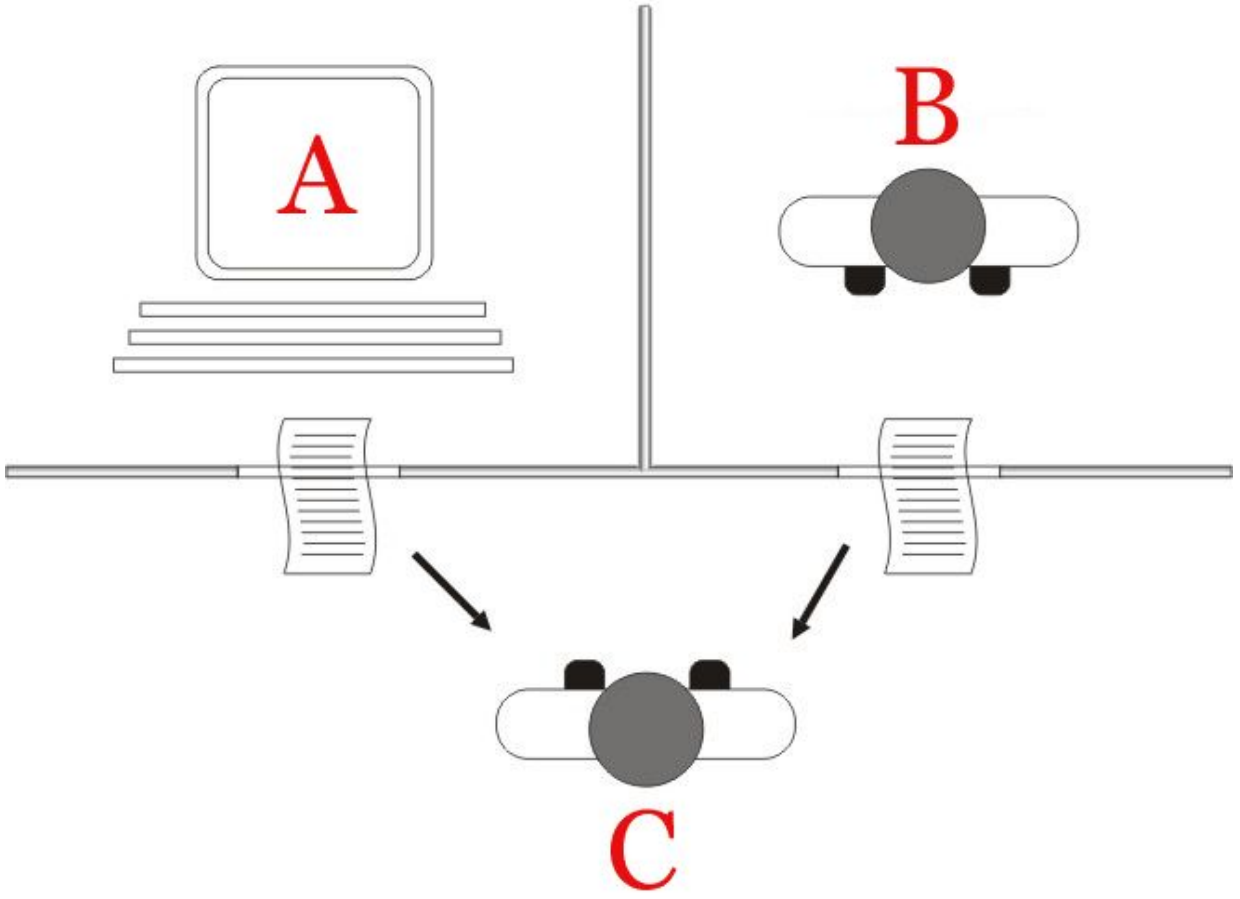
Turing's variation

Imagines a variation on the original party game:

C communicates (via text) with **Players A and B**.

Players A and B are both trying to convince **C** that they're the (woman) human.

If **C** misidentifies the (woman) human "often enough", then…?



Who cares what's happening inside?

1. If machines can regularly win the Imitation Game, then they think (i.e., it's capable of thought).
2. Machines can regularly win the Imitation Game.

Therefore, machines can think.

But doesn't it *matter* what's happening inside?

1. If men can regularly win the original Imitation Game, then they are women.

2. Men can regularly win the original Imitation Game.

Therefore, men are women.

2. John Searle's "Chinese Room" (published 1980)

How the Chinese Room works

From https://philosophy.hku.hk/joelau/?n=Main.TheChineseRoomArgument

ONE formulation of Searle's argument – "Can a software system understand?"

1. According to "Strong AI", any suitably-programmed computer system can genuinely understand and have other mental states (such as beliefs) that humans have.

2. If Strong AI is true, then there is a program for Mandarin Chinese such that if any computing system runs that program, that system thereby comes to understand Chinese.

3. But I (me, you, Searle) could "run a program" for Chinese without thereby coming to understand Chinese.

Therefore Strong AI is false.

https://plato.stanford.edu/entries/chinese-room/

Part 2 – (mis)treating robots?

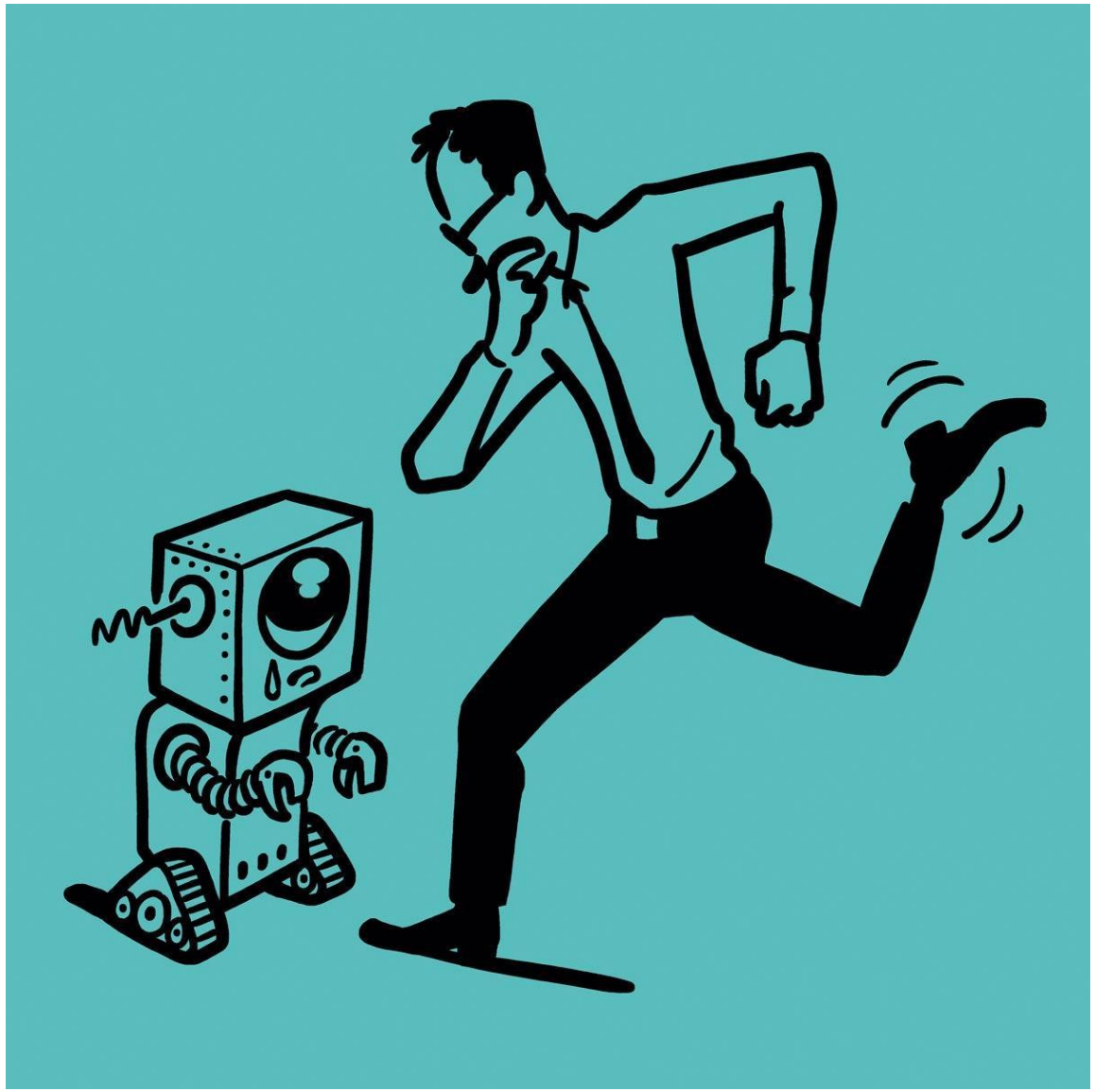
Do robots have rights to not be mistreated?

- No

- Yes – directly; inherent to their moral status
- Yes – indirectly: it'll offend some human

- Yes – indirectly: mistreating them will affect the treatment of non-robots (e.g., humans, nonhuman animals)

- Yes – indirectly: mistreating them deforms our own moral character



Example: hitchBOT

Source: http://www.hitchbot.me/wp-content/media/hB_MediaKit_Summer2014.pdf

What (and why) was hitchBOT?

Professors Frauke Zeller and David Harris Smith designed the robot to learn about how people interact with technology and ask the question, “Can robots trust human beings?”

HitchBot had a GPS tracker, a cellular chip and a camera. It could hold basic conversations, offering charming jests and reciting random facts. The size of a 6-year-old child, it relied on solar charging and cigarette lighters in cars for energy.

Source: Los Angeles Times, 03 August 2015

Where is hitchBOT now?

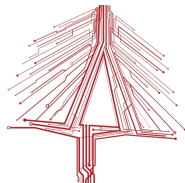
HitchBot, a talking robot built by Canadian researchers, was found maimed and battered beyond repair over the weekend in Philadelphia. It was trying to hitchhike from Massachusetts to California by relying on the kindness of human drivers. Hitchhiking took hitchBot across the length of Canada and most of Germany and the Netherlands.

But it lasted just two weeks in the U.S. and couldn't escape the East Coast.

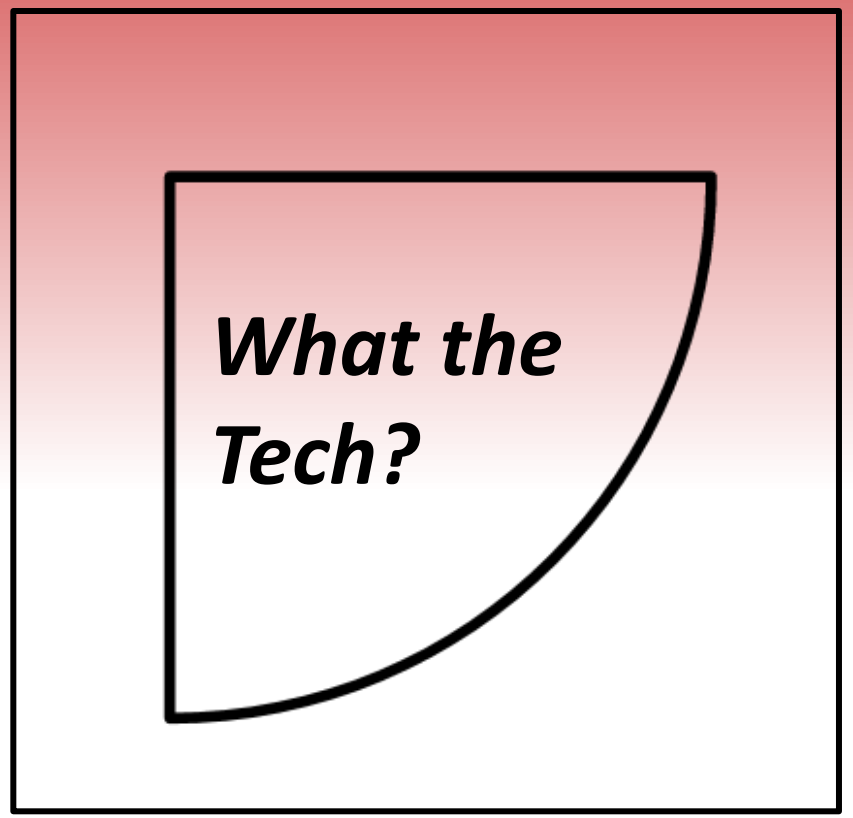
"My trip must come to an end for now, but my love for humans will never fade," the robot said in his last Facebook message.

Source: Los Angeles Times, 03 August 2015

Questions?



ETHICS IN AI



- Debrief!
- A chance for you to tell us how it’s going. For instance:
  - What went well this week?
  - What didn’t?
  - What are you excited for?
  - What are you unsure about?

Weekly Reflection… With a Twist!

- Switching it up this week…
  - Pick a prompt that you DIDN'T have from the small group activity!
- Your mission:
  - Discover which side people you talk with (your friends/family/community members) are taking and what their reasons for doing so are.
  - Based on that information, what would your stance be? *Not just* for yourself but if you are speaking on behalf of the community you spoke with?

END OF CONTENT