Part 3: Ethics in AI

The goal of these slides is to review the previous week's content and to dive into a deeper analysis of the ethical questions surrounding AI.

This document is designed to walk you through the content (originally split into two 90 minute sessions) and to collect notes that might be helpful as you go.

Slideshow contents:

- Recap/Looking ahead
- Weekly reflections
- Ethics in AI (Overview)
    - "Is It Right or Wrong?"
    - "Nuance"
- Risks
    - Nothing is risk-free
    - Some examples of ethical risks in AI
        i. Unequal access
        ii. Environmental impacts
        iii. Algorithmic bias
        iv. Unanticipated Impacts
        v. Additional examples
- Ethics in AI (Small group activity)
    - Explain
    - Hand out activity guide/envelopes
    - Share answers
- Bonus reading
- [Halfway Point] Deep Dive: Ethics in AI
    - Different types
    - Activity
- Deep dive: "Thinking and feeling machines?
    - Originally by Dr. Vance Ricks
- Debrief
- Week 3 Reflection Activity Prompts

Last time, we discussed:

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

Today will be about…

- Share second round of weekly reflections
- Talk about Ethics
    - Overview
    - In AI (different settings)
    - Small group activity
- [Halfway point] Deep Dive: “Thinking and feeling machines?”
- Debrief and new reflection prompts

Our Weekly Reflections:

- How it works (schedule)

Ethics in AI (Overview)

- "What *are* ethics?"
    - What do the participants' answers all have in common?

- "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

- "But even if we know right from wrong, we also know that life isn't always black and white. One of the tricky things about being a human living in this world of ours is that more often than not, we are weighing right against wrong and seeing that there are shades of gray between the black and the white. But that is another uniquely human trait…"

Nuance

- 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

- "That brings risks"

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

Some examples of risks:

- The ones we'll look at:
    - Unequal access
    - Environmental impacts
    - Unanticipated impacts
    - Algorithmic bias

- Other big examples:
    - Over Reliance on tech
    - Hacking and privacy
    - Job loss

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.

- Hypothetical cases:
    - Communities
    - Education

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

Algorithmic Bias

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

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

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?

Other Examples and a Reminder

- 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?

- Before we 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!

Activity: AI and Tech on the Street

- The activity is based around putting ethical concerns to the test.
- Hand out worksheets. Please note that the prompts included were for the original *WTT* held at Northeastern Univerity's Boston campus. Feel free to make some of your own!

Next time:

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

- ETHICS IN AI: BONUS READING
    - Link on screen

[Halfway point] Last time, we…

- Ethics
    - In general
    - In AI
- Small group activity

Today, we will look at…

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

Deep dive: "Thinking and feeling machines? Originally by Dr. 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.
- [Read opening quote: "Technology is neither good nor bad; nor is it neutral"
- Part 1 - Two classic 20th-c thought experiments about (the possibility of digital) "thinking machines"
    - The Imitation Game(s) – "Can a machine think?"
        - Original version v. 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).
            - 20 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)
  - Link: https://philosophy.hku.hk/joelau/?n=Main.TheChineseRoomArgument
  - ONE formulation of Searle's argument – "Can a software system understand?"
    - According to "Strong AI", any suitably-programmed computer system can genuinely understand and have other mental states (such as beliefs) that humans have.
    - 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.
    - 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?
            - Read text on slide
        - Where is hitchBOT now?
            - Read text on slide

- Questions**?**

Activity: AI and Tech on the Street

The activity is based around examining just how many data-gathering devices are in use publicly and considering how the information it collects could potentially be used in AI.

Hand out worksheets. Please note that the map included is for Northeastern Univerity's Boston campus. You will (naturally) want to create a similar map for your uses.

Debrief and Upcoming reflections

- 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 reflections… 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?