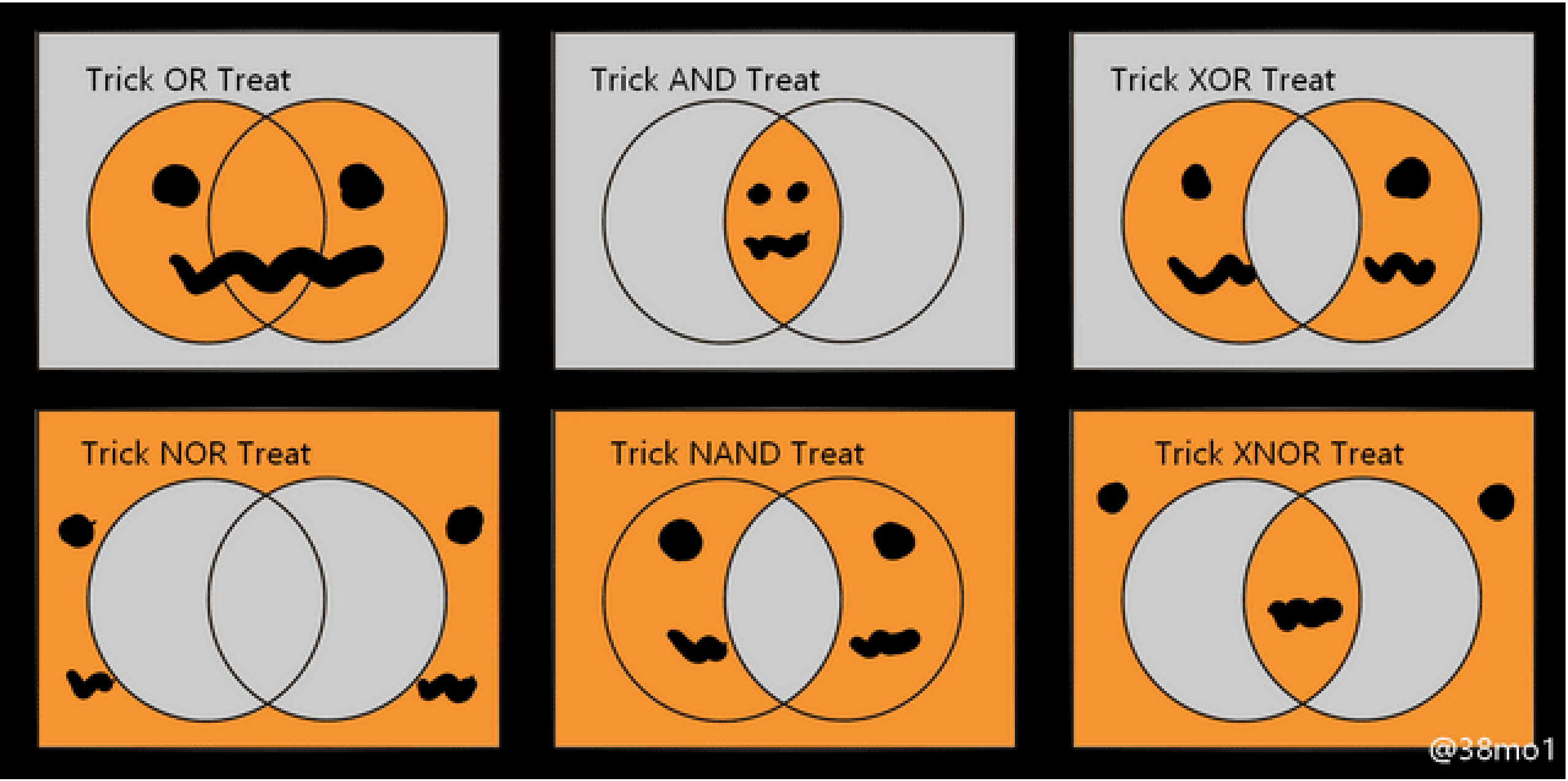
Responsible AI

05-499/899 Fall 2024

Celebrating Accessibility

https://cmu-05-499.github.io

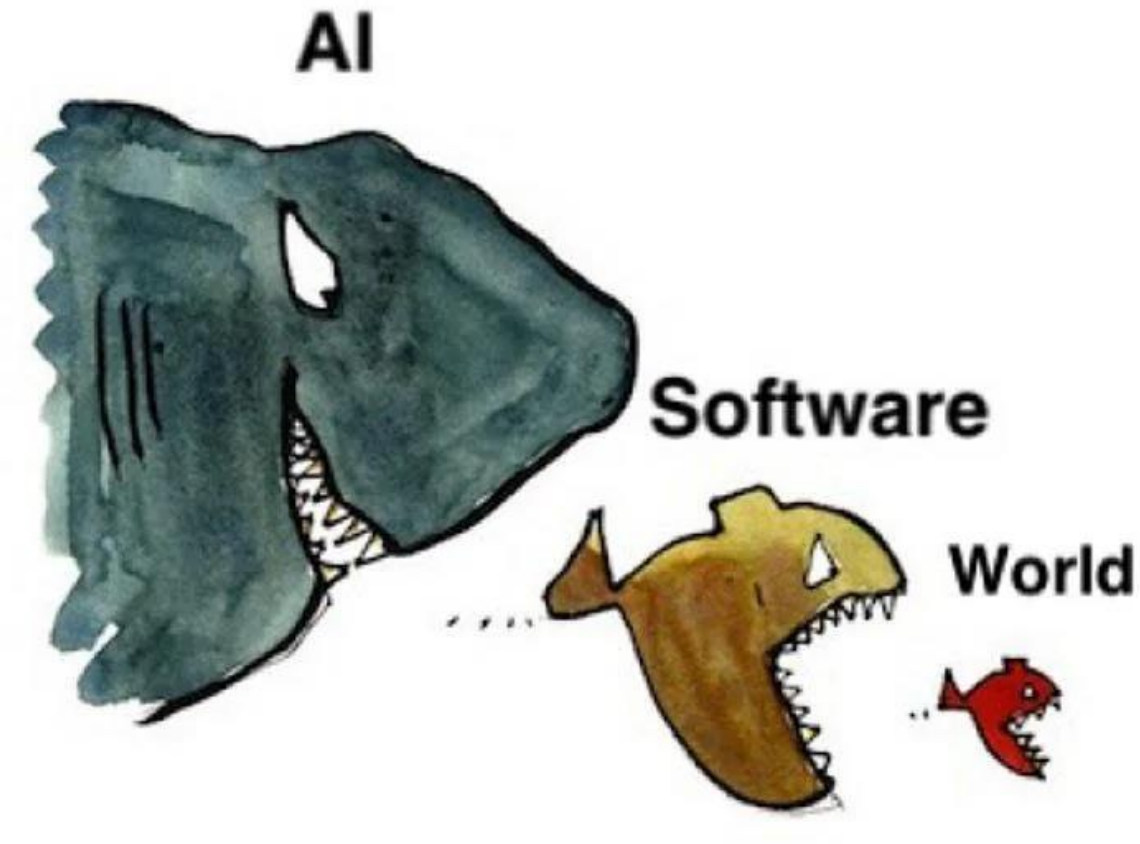
Andrew Begel and Patrick Carrington


Trick OR Treat
Trick AND Treat
Trick XOR Treat
Trick NOR Treat
Trick NAND Treat
Trick XNOR Treat
@38mo1

Administrivia

- Tuesday, Nov 5 is Election Day. No class. Vote if you are eligible!
- P4 – Project Milestone 1 due November 14, 2024.
- P5 – Project Milestone 2 due November 26, 2024.
- P4 and P5 are posted on the course home page.

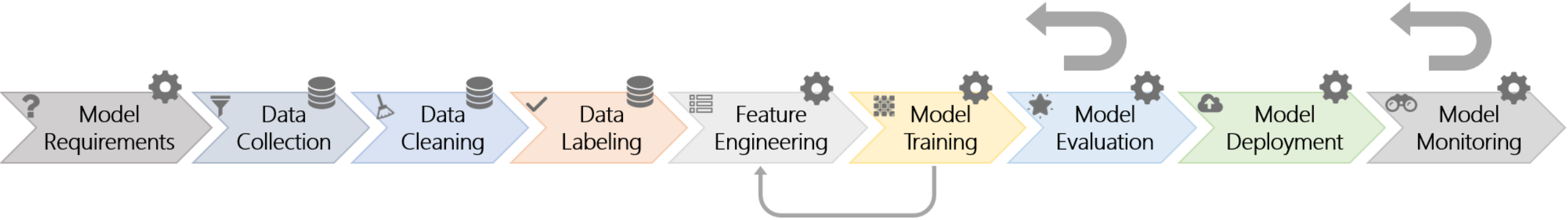
AI is eating the world



What common products use AI?

- Phone-based customer support
- Recognizing money in vending machines
- Credit card fraud detection
- Resume filtering
- Social media content moderation
- Cruise control / self-driving cars

AI Practitioners Use a Machine Learning Workflow



Workflow used by many teams at Microsoft

AI does not always work well for people with disabilities

- Lack of diversity in training sets and designers
- Inadvertent propagation of systemic societal bias
- Outlier elimination erases disabled experiences
- Unexpected input confuses the AI

Physical disabilities are an afterthought

- Speech recognition systems have trouble with the ‘accent’ of a deaf or hard of hearing speaker or a person whose cerebral palsy affects their vocal cords. Makes it quite difficult to call your bank.
- reCAPTCHA systems expect people to be able to see or hear, but there are many people who cannot do either, making it impossible for them to prove their humanity.
  - Example from Karen Nakamura. 2019. My Algorithms Have Determined You're Not Human: AI-ML, Reverse Turing-Tests, and the Disability Experience. In ASSETS '19. ACM, New York, NY, USA, 1–2. DOI:https://doi.org/10.1145/3308561.3353812

- Training sets and application designs often lack diversity in representation and imagination.

Past bias predicts future bias

- Resume filtering systems support recruiters' ability to scale.
- Unusual resumes are weeded out of job application systems because they do not represent the norm.
- People with disabilities are subject to systemic discrimination in society and often have patchy job histories with large gaps in employment.
- These are helpfully weeded out by the machine learning classifier.
  - Example from Moss, H. (2020). Screened out onscreen: Disability discrimination, hiring bias, and artificial intelligence. *Denv. L. Rev.*, *98*, 775.
- Machine learning systems often perpetuate societal biases.

Overzealous outlier elimination

- Boosting increases signal by emphasizing the dominant pattern and improve training. But outliers are real people's experiences, and they are not irrelevant.
- If you move differently, e.g. you don't walk or bike, but instead use a wheelchair, self-driving cars may not realize that you're a person to be avoided.
  - Example from Shari Trewin, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. 2019. Considerations for AI fairness for people with disabilities. AI Matters 5, 3 (September 2019), 40–63. https://doi.org/10.1145/3362077.3362086
- Even worse, once alerted to the problem, algorithm designers created a fix, that when tested in simulation, ran over people in wheelchairs with much higher precision and recall.
  - Some people in wheelchairs find it faster to push themselves *backwards* through intersections. Cars swerving to avoid them *mispredicted* the direction wheelchairs would move.

Unexpected input confuses AI

- Many voice assistants have difficulty processing conversational repair.
- Cognitive impaired people can't figure out what's wrong and persist in speaking to the assistant as though it was a real person who can understand them.
  - Example from Clayton Lewis. 2020. Implications of developments in machine learning for people with cognitive disabilities. SIGACCESS Access. Comput., 124, Article 1 (June 2019). https://doi.org/10.1145/3386308.3386309.

- Object recognition algorithms are trained on clear, focused, properly framed photos.
- Blind and visually impaired people's camera-phone photos are often off center and blurry, inhibiting successful recognition.
  - Example from Erin Brady, Meredith R. Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. **Visual Challenges in the Everyday Lives of Blind People.** ACM Conference on Human Factors in Computing Systems (CHI), 2013

Case management systems' failure of imagination

- In 2008, Medicaid asked Indiana resident Omega Young to recertify her eligibility in an in-person meeting. Suffering from ovarian cancer, she called to say she could not make the meeting as she was in the hospital. Her benefits were cut off anyway. The ML system automating welfare eligibility said she exhibited a "failure to cooperate."
  - Example from Virginia Eubanks: https://www.npr.org/sections/alltechconsidered/2018/02/19/586387119/automating-inequality-algorithms-in-public-services-often-fail-the-most-vulnerab

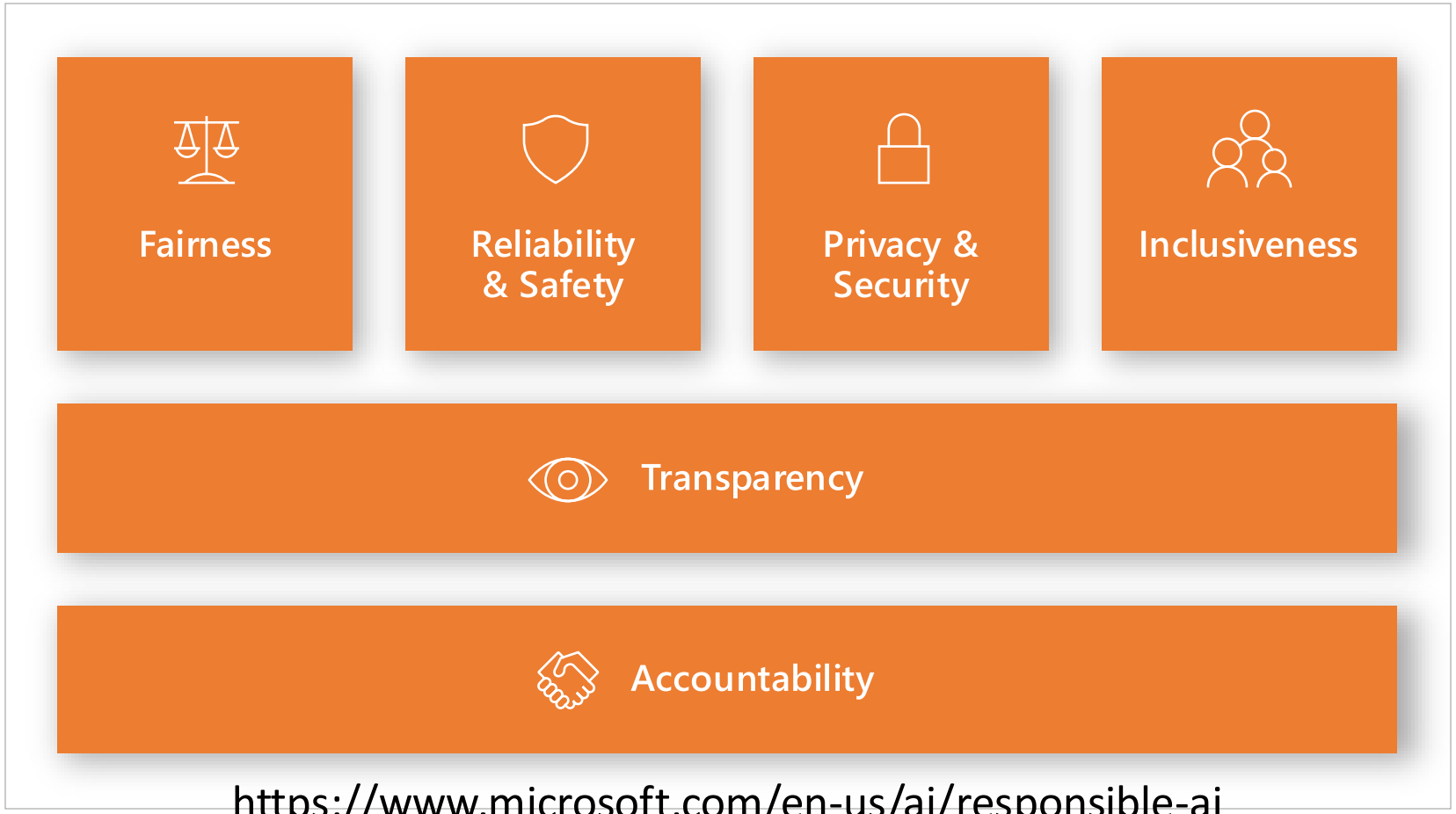
- Algorithmic case handling may not recognize extenuating circumstances that human case workers would never miss.

Human-Centered AI Requires a Balance

1. What do people need?
2. What are the capabilities of AI?
3. How far can these capabilities be pushed?
4. How can we mitigate our limitations?
5. Are we taking the right approach to achieving our goals?

Note the order of these questions!

Microsoft Responsible AI Principles



https://www.microsoft.com/en-us/ai/responsible-ai

Class Discussion

- Rua Mae Williams, Louanne Boyd, and Juan E. Gilbert. 2023. **Counterventions: a reparative reflection on interventionist HCI.** In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 653, 1–11. https://doi.org/10.1145/3544548.3581480

- Os Keyes, Jevan Hutson, and Meredith Durbin. 2019. **A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry**. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19). Association for Computing Machinery, New York, NY, USA, Paper alt06, 1–11. https://doi.org/10.1145/3290607.3310433

Guidelines for Human AI Interaction

Learn more: https://aka.ms/aiguidelines

INITIALLY	1 Make clear what the system can do.	2 Make clear how well the system can do what it can do.					
DURING INTERACTION	3 Time services based on context.	4 Show contextually relevant information.	5 Match relevant social norms.	6 Mitigate social biases.			
WHEN WRONG	7 Support efficient invocation.	8 Support efficient dismissal.	9 Support efficient correction.	10 Scope services when in doubt.	11 Make clear why the system did what it did.		
OVER TIME	12 Remember recent interactions.	13 Learn from user behavior.	14 Update and adapt cautiously.	15 Encourage granular feedback.	16 Convey the consequences of user actions.	17 Provide global controls.	18 Notify users about changes.

Human-AI Guidelines Cards

https://www.microsoft.com/en-us/haxtoolkit/library/

Impact Assessments

1. System Purpose
   - What value does your system deliver?
   - How does your system make use of AI to improve on today's solution?
2. Who are the stakeholders? (direct, indirect, malicious actors, marginalized groups, non-stakeholders)
3. How do you intend your system to be used by each stakeholder?
4. How could your system be misused by each stakeholder?
5. What are the design tradeoffs between the system's benefits and harms?
6. How can you mitigate the risk?

Participation Activity: AI for People with Disabilities

- Divide up into up to 7 groups of 3 - 4 people
- Each group takes a scenario and a target user

1. Come up with an application concept
   - What value does your system deliver?
   - How does your system improve on today's solution using AI?
2. Ideate an AI model concept that can achieve the scenario
   - Describe the kind of model you'd train, the model inputs and outputs, the data you need to collect to train the model, the most important features of the model, the evaluation criteria used to judge the goodness of the model
3. Describe the top 3 user tasks that illustrate how your application's UX works
   - Make use of the Human-AI Guidelines: https://docs.microsoft.com/en-us/ai/guidelines-human-ai-interaction/
4. Conduct an impact assessment of your application

Choose one of these activities below:

- E-Commerce: http://aka.ms/gssi-e-commerce
  - Your user is a 26-year-old autistic woman who needs to buy a new computer.
- Movie Recommendations (mobile): http://aka.ms/gssi-movie-recommendations
  - Your user is a 32-year-old woman suffering locked-in syndrome who has been unable to communicate until 3 months ago when she received a brain implant.
- Activity Tracker (step counter): http://aka.ms/gssi-activity-tracker
  - Your user is a 21-year-old paraplegic man who uses a motorized wheelchair to get around his NYC apartment.
- Autocomplete (mobile): http://aka.ms/gssi-autocomplete
  - Your user is a 23-year-old woman with a PhD in neurobiology who has dysarthric motor abilities that affect her ability to type and clearly speak words out loud.
- Social Networks (feed filtering): http://aka.ms/gssi-social-networks
  - Your user is a 45-year-old man with face blindness and emotional instability and a bent towards authoritarianism.
- Voice Assistants: http://aka.ms/gssi-voice-assistants
  - Your user is a 16-year-old non-binary person with a traumatic brain injury that impacts their short-term memory, speech quality, and trusts everyone they talk to.
- Photo Organizers: http://aka.ms/gssi-photo-organizers
  - Your user is a 51-year-old man who can no longer see clearly due to retinosis pigmentaria, which was diagnosed 15 years ago.

Reflect on Lessons Learned

- Report back on your application's design
- How did your design change as you ran through the Human-AI Guidelines?
- How did your design change in response to the Impact Assessment?

- What was difficult about this process?
- What would help make this process easier for you as a user researcher?