ICS 691D: HUMAN-CENTERED AI COURSE OVERVIEW

Peter Y. Washington, PhD

Assistant Professor, ICS

August 22, 2022

OUTLINE

- Human-Centered Artificial Intelligence
- Course Schedule and Topics
- Coursework and Grading

OUTLINE

- Human-Centered Artificial Intelligence
- Course Schedule and Topics
- Coursework and Grading

WHAT IS HUMAN-CENTERED AI (HCAI OR HAI)?

Thoughts?

HUMAN-CENTERED AI

"Artificial Intelligence (AI) dreams and nightmares, represented in popular culture through books, games, and movies, evoke images of startling advances as well as terrifying possibilities. In both cases, people are no longer in charge; the machines rule. However, there is a third possibility; **an alternative future filled with computing devices that dramatically amplify human abilities, empowering people and ensuring human control**. This compelling prospect, called Human-Centered AI (HCAI), enables people to see, think, create, and act in extraordinary ways, by **combining potent user experiences with embedded AI support services that users want**."

- Ben Shneiderman, Human-Centered AI

HUMAN-CENTERED AI

Technical

- Algorithms
- Math
- Programming

Ethical

- Formal ethical frameworks
- User and societal values

Design

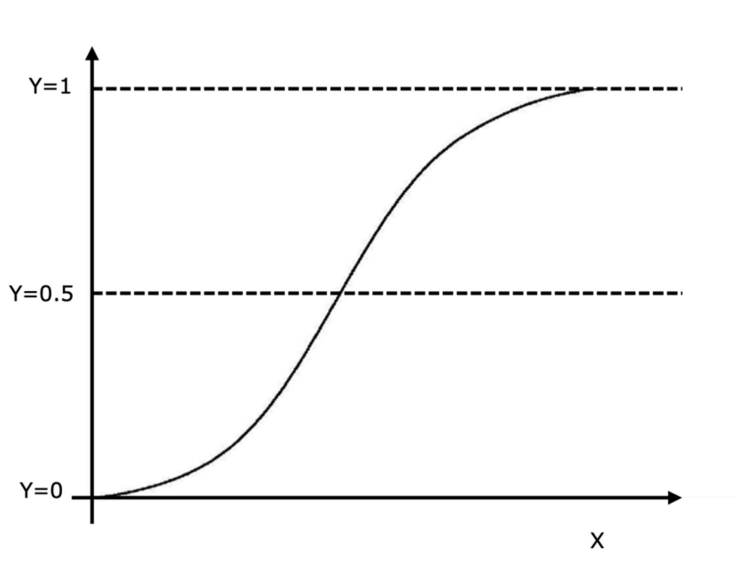
- User interaction with AI
- Features and tradeoffs of the AI system (explainability, speed, performance, fairness, etc)

EXAMPLE: TECHNICAL

Is logistic regression an interpretable algorithm?

EXAMPLE: TECHNICAL

$$f(x) = \frac{1}{1 + e^{-x}}$$



EXAMPLE: TECHNICAL

$$y = m_1x_1 + m_2x_2 + m_3x_3 + b$$

$$Sigmoid\ Activation(mx + b) = f(mx + b) = \frac{1}{1 + e^{-mx+b}}$$

EXAMPLE: TECHNICAL

Prediction Probability =

Sigmoid($m_1x_1 + \ldots + m_Nx_N$)

EXAMPLE: TECHNICAL

Prediction Probability =

Sigmoid($m_1x_1 + \ldots + m_Nx_N$)

$|m_i|$ is the strength of the "importance" of input variable x_i

sign(m_i) signifies whether x_i contributes more towards a positive (1) or negative (0) binary prediction

HOW TECHNICAL DO I NEED TO BE FOR THIS COURSE?

- You can make an A or A+ in the class without a solid understanding of the technical and mathematical details of AI
- However, we will spend a good portion of class discussing the math and tech details of AI algorithms
- Nevertheless, we will also spend a good portion of class discussing the ethical and design aspects, and you can focus your assignments on these aspects and drop the extremely technical class papers

EXAMPLE: ETHICAL

Utilitarian ethics: the most ethical choice is the one that will produce the greatest good for the greatest number

Deontological ethics: morality of an action should be based on whether that action itself is right or wrong under a series of rules, rather than based on the consequences of the action

What is a Utilitarian argument for or against removing traditional blue-collar jobs while increasing blue-collar AI data labeling jobs?

How about a Deontological argument for or against?

Do the Utilitarian and Deontological arguments reach a different conclusion?

EXAMPLE: DESIGN

How might we update traditional ML classification methods to support prediction classes which are inherently subjective (e.g., predicting a person's facial emotion expression)?

RECOMMENDED PREREQUISITES

ICS 235 (Machine Learning Methods)

ICS 434 (Data Science Fundamentals)

ICS 435 (Machine Learning Fundamentals)

Basic understanding of:

- Fundamental ML techniques like logistic regression and support vector machines
- Deep neural networks, include dense, convolutional, and recurrent neural networks

Some comfort with:

- Critical thinking
- Writing
- Oral communication

MORE EXAMPLE DISCUSSION QUESTIONS

Technical

- In what ways does a CNN provide interpretability?
- How can this loss function be modified to reduce bias?

Ethical

- How might advances in active learning algorithms affect the job market for data labelers?
- What are the right evaluation metrics to use for a diagnostic system for COVID? What is an acceptable cutoff?

Design

- How transparent should a federated learning system be to the end user?

OUTLINE

- Human-Centered Artificial Intelligence
- Course Schedule and Topics
- Coursework and Grading

COURSE SCHEDULE

Date	Topic	Pre-Class Readings (before class)	Assignment Due (midnight before class)
Mon Aug 22	Course Overview		
Wed Aug 24	ML Review	Greener Nature Reviews 2022	Greener reflection
Mon Aug 29	Deep Learning Review	Dong Computer Science Review 2021	**Literature review + discussion leading topics**
Wed Aug 31	Reinforcement Learning	Arulkumaran IEEE Signal Processing 2017	**Proposal topic**
Mon Sep 5	Labor Day Holiday		
Wed Sep 7	Interactive ML	Haber NeurIPS 2018	Haber reflection
Mon Sep 12	Robotics	Akalin Sensors 2021	Akalin reflection
Wed Sep 14	Autonomous Vehicles	Kohli FICC 2019	Kohli reflection
Mon Sep 19	Active Learning	Olsson 2009	Olsson reflection
Wed Sep 21	Crowdsourcing	Vaughan JMLR 2017	Vaughan reflection
Mon Sep 26	Human-Computer Interaction	Dove CHI 2017	Dove reflection

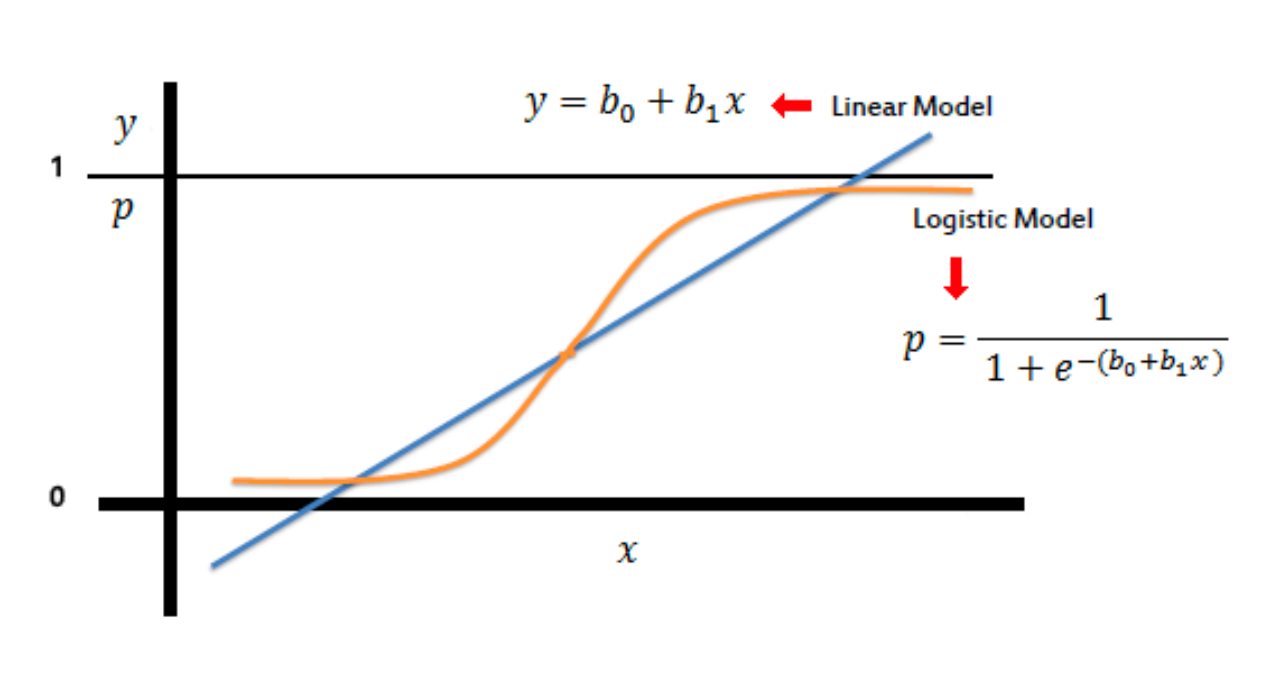
COURSE SCHEDULE

Wed Sep 28	Interpretable ML	Došilović MIPRO 2018	Došilović reflection
Mon Oct 3	Literature Review Presentations		**Literature review writeup (3 pages)**
Wed Oct 5	Literature Review Presentations		
Mon Oct 10	Explainable ML	Alqaraawi IUI 2020	Alqaraawi reflection
Wed Oct 12	Communicating ML results	Varoquaux Neuroimage 2018	Varoquaux reflection
Mon Oct 17	Differential Privacy	Chen Biocomputing 2020	Chen reflection
Wed Oct 19	Federated Learning	Rieke NPJ Digital Medicine 2018	Rieke reflection
Mon Oct 24	Bias	Mehrabi ACM Computing Surveys 2021	Mehrabi relfection
Wed Oct 26	Fairness	Holstein CHI 2019	Holstein reflection
Mon Oct 31	Proposal Checkin Presentations		**Proposal outline**
Wed Nov 2	Proposal Checkin Presentations		
Mon Nov 7	Ethics	Char NEJM 2018	Char reflection

COURSE SCHEDULE

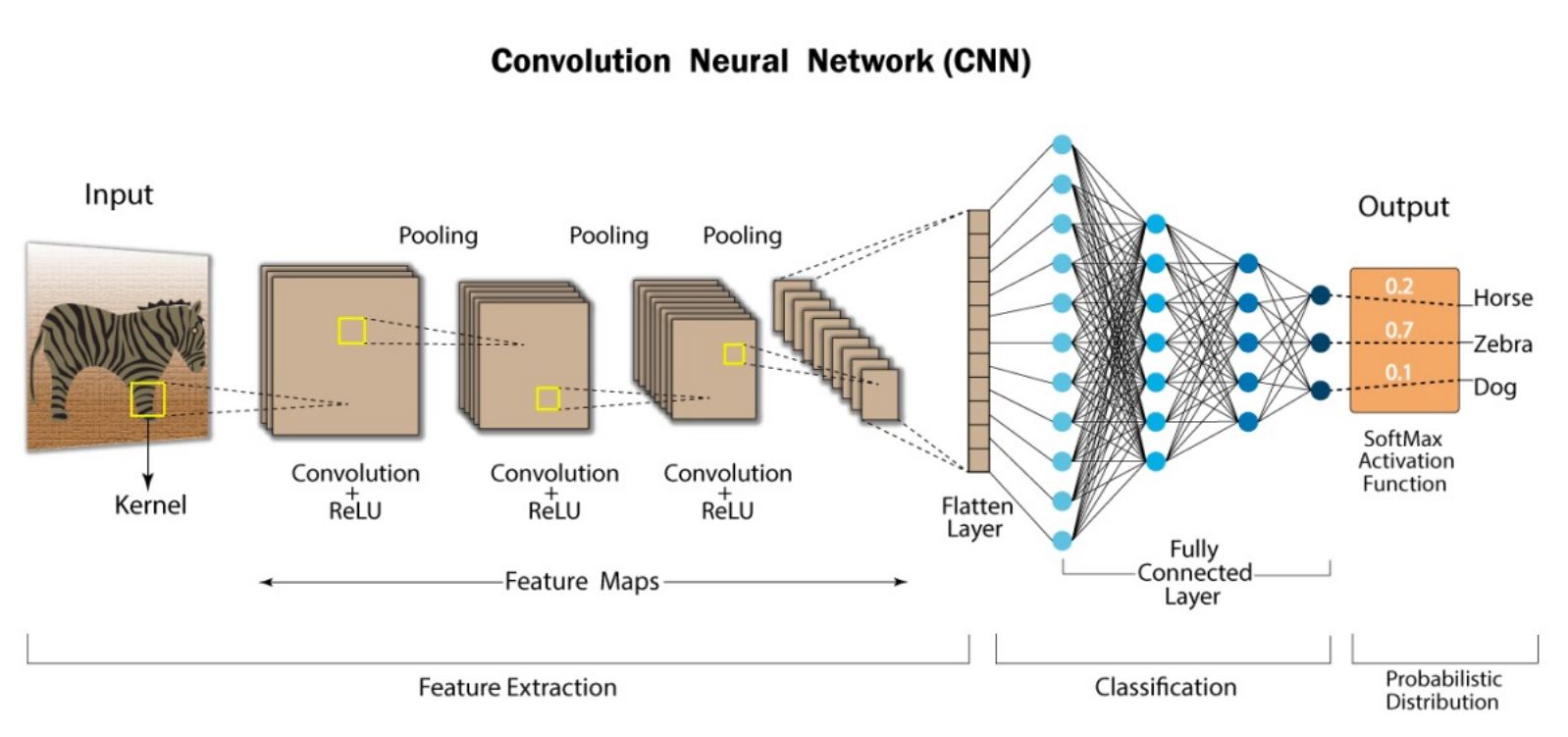
Wed Nov 9	Generative Models	Lyu ICMEW 2020	Lyu reflection
Mon Nov 14	Natural Language Processing	Mishev IEEE Access 2020	Mishev reflection
Wed Nov 16	Biology	Zou Nature Genetics 2019	Zou reflection
Mon Nov 21	Medicine and Health	Hannun Nature Medicine 2019	Hannun reflection
Wed Nov 23	Digital Phenotyping	Omberg Nature Biotechnology 2022	Omberg reflection
Mon Nov 28	Social Network Analysis	Smith PNAS 2021	Smith reflection
Wed Nov 30	Course Summary		
Mon Dec 5	Final Proposal Presentations		**Final project proposal (5 pages)**
Wed Dec 7	Final Proposal Presentations		

ML REVIEW



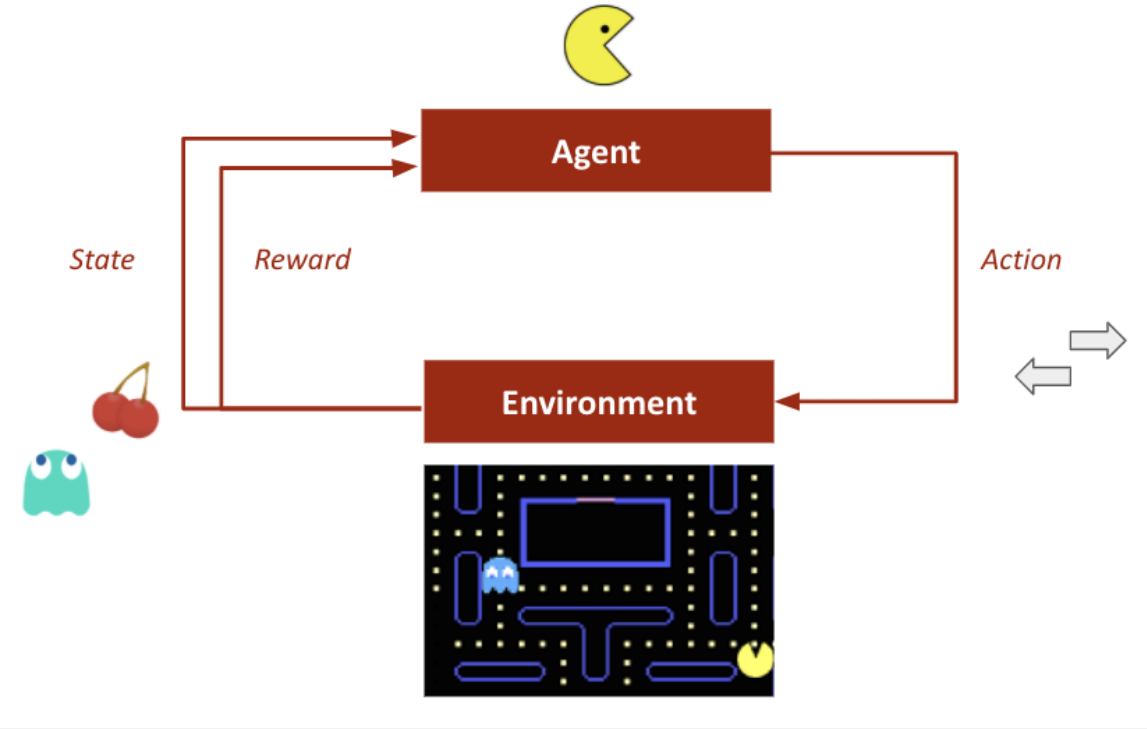
https://www.saedsayad.com/logistic_regression.htm

DL REVIEW



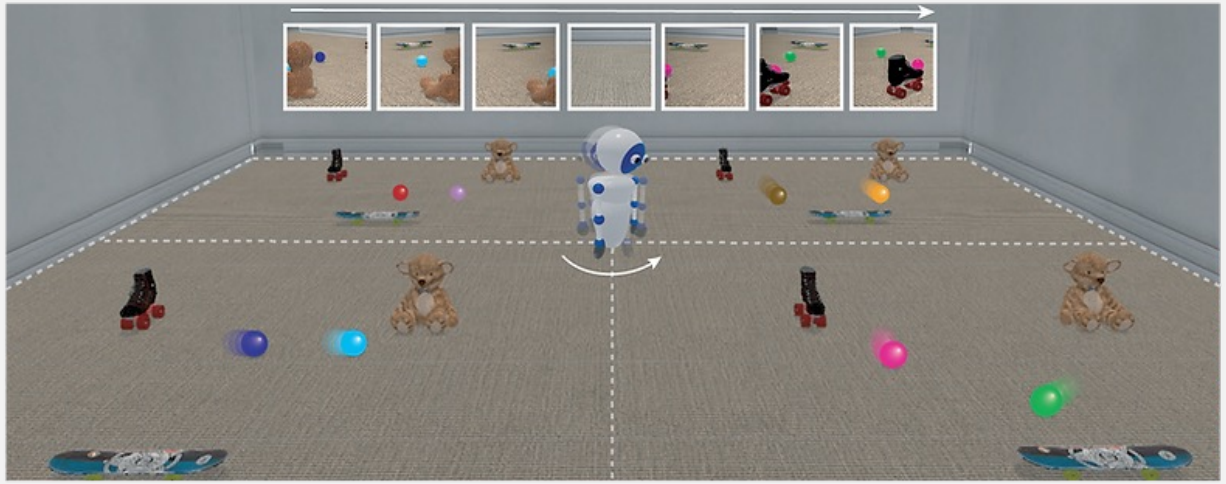
https://developersbreach.com/convolution-neural-network-deep-learning/

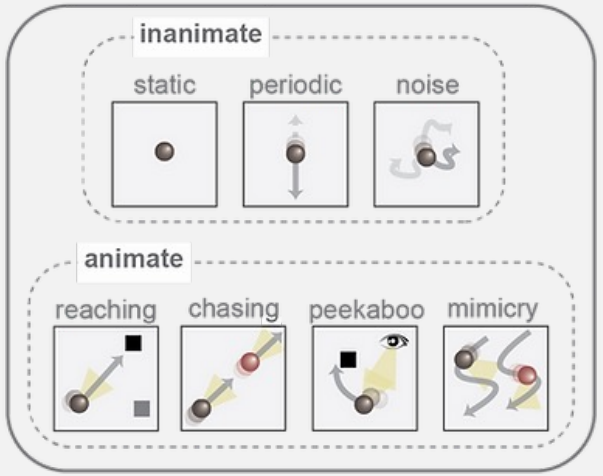
REINFORCEMENT LEARNING



https://bechirtr97.medium.com/safe-and-efficient-off-policy-reinforcement-learning-6a6b48d74b73

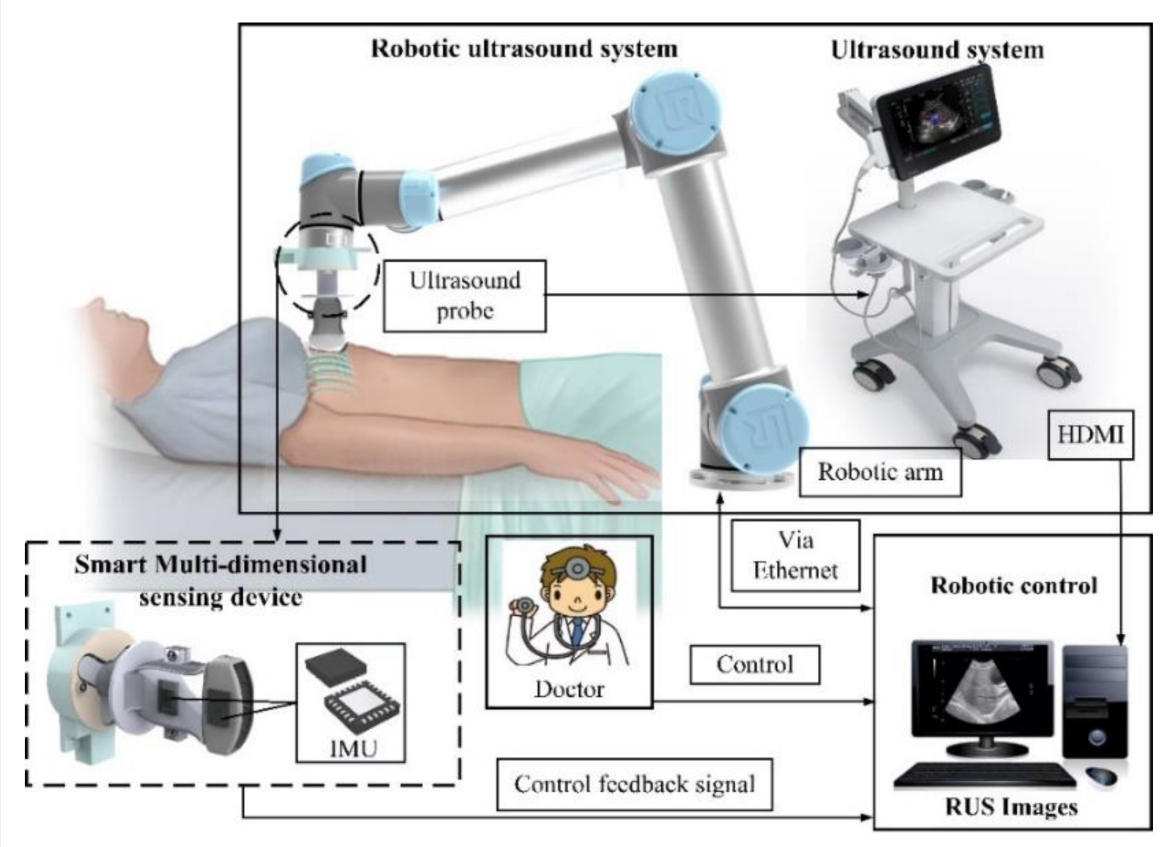
INTERACTIVE ML





Kim et al, "Active World Model Learning in Agent-Rich Environments with Progress Curiosity," 2020.

ROBOTICS

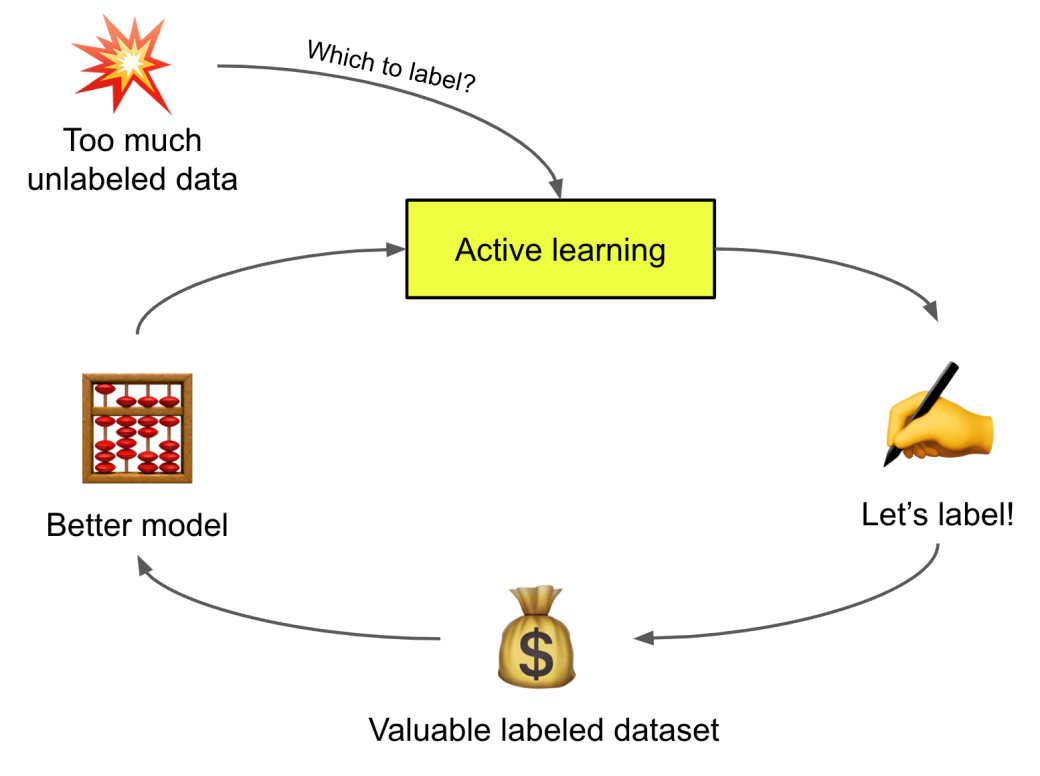


Wang et al, "An Improved Sensing Method of a Robotic Ultrasound System for Real-Time Force and Angle Calibration," 2021.

AUTONOMOUS VEHICLES

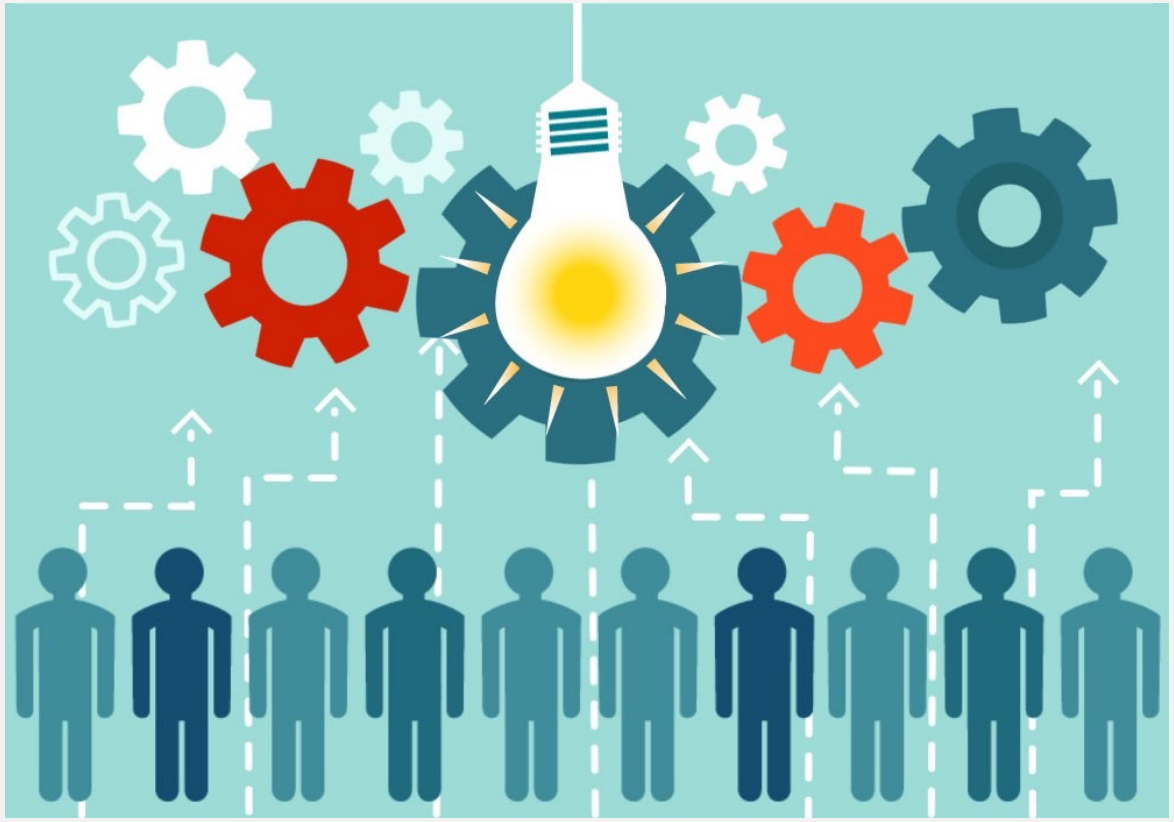
https://www.fau.edu/newsdesk/articles/autonomous-vehicles-patent.php

ACTIVE LEARNING



https://lilianweng.github.io/posts/2022-02-20-active-learning/

CROWDSOURCING

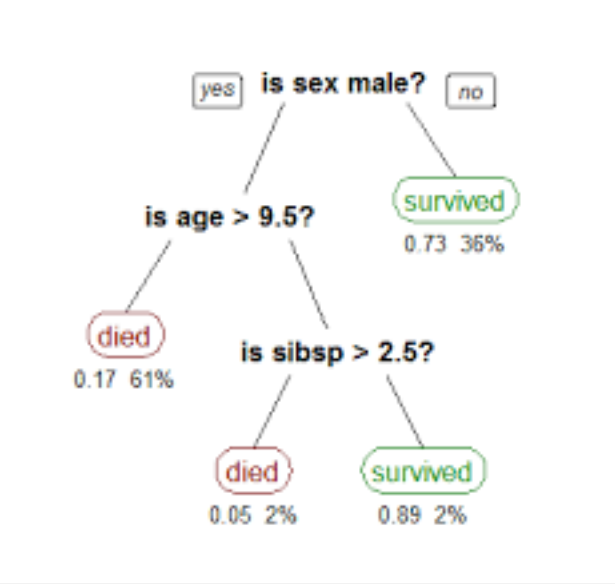


https://bloombergcities.medium.com/how-crowdsourcing-drives-citizen-engagement-eb5a7eeaf2b1

HUMAN-COMPUTER INTERACTION

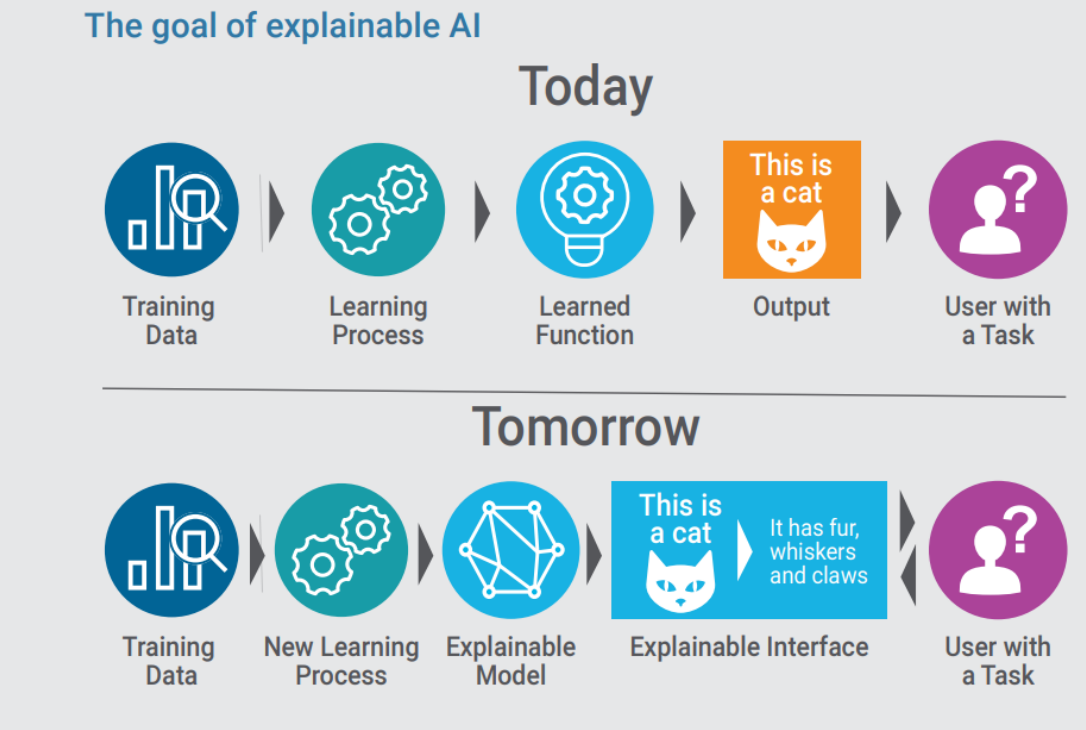
https://www.hp.com/us-en/shop/tech-takes/exploring-human-computer-interaction

INTERPRETABLE ML



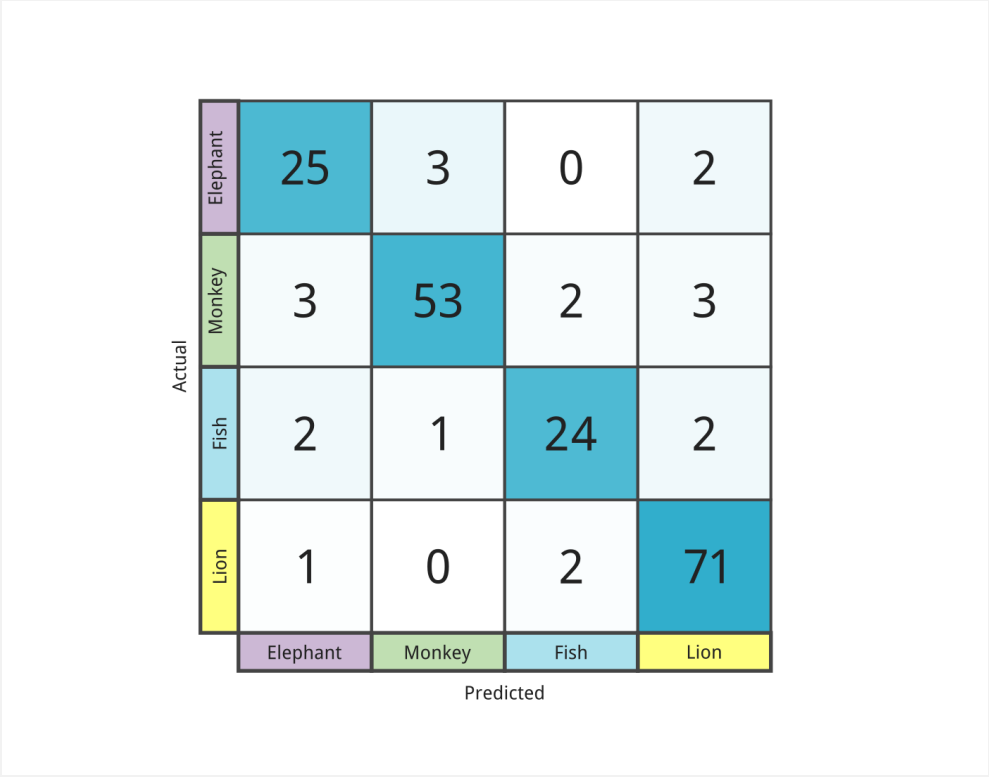
https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052

EXPLAINABLE ML



https://www.datanami.com/2018/05/30/opening-up-black-boxes-with-explainable-ai/

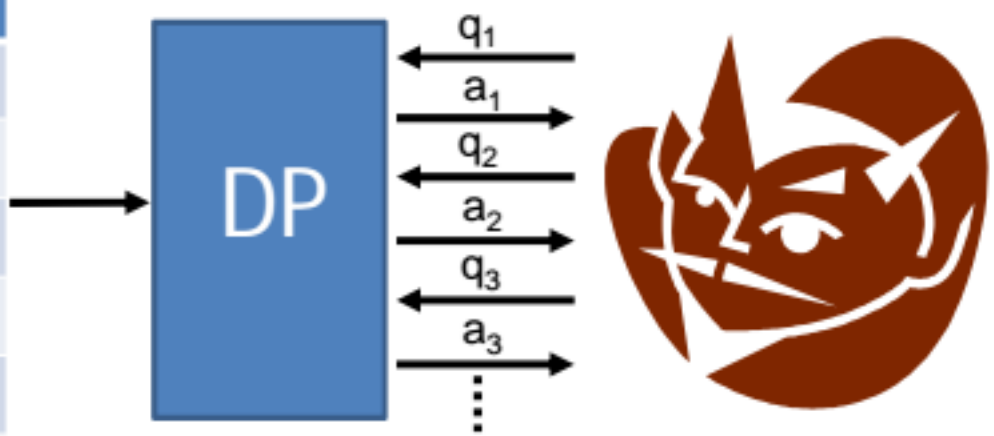
COMMUNICATING ML RESULTS



https://towardsdatascience.com/visual-guide-to-the-confusion-matrix-bb63730c8eba

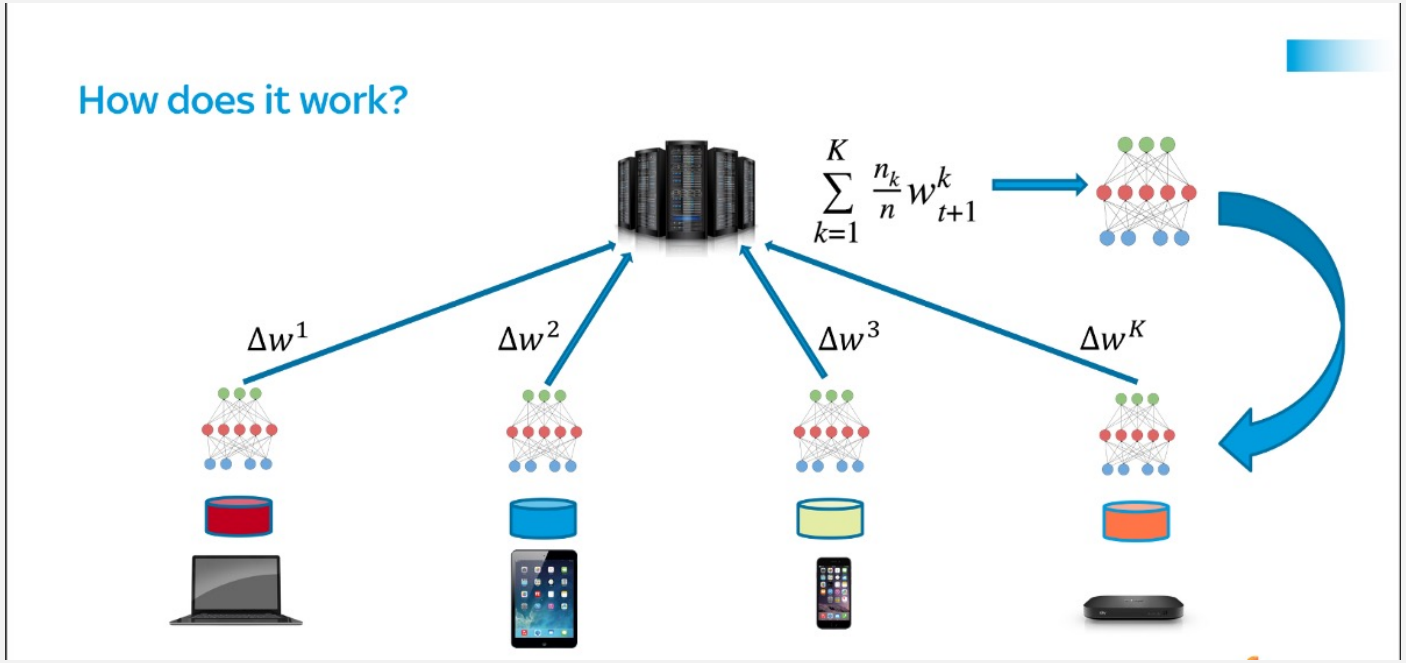
DIFFERENTIAL PRIVACY

Sex	Blood	…	HIV
F	B	…	Y
M	A	…	N
M	O	…	N
M	O	…	Y
F	A	…	N
M	B	…	Y



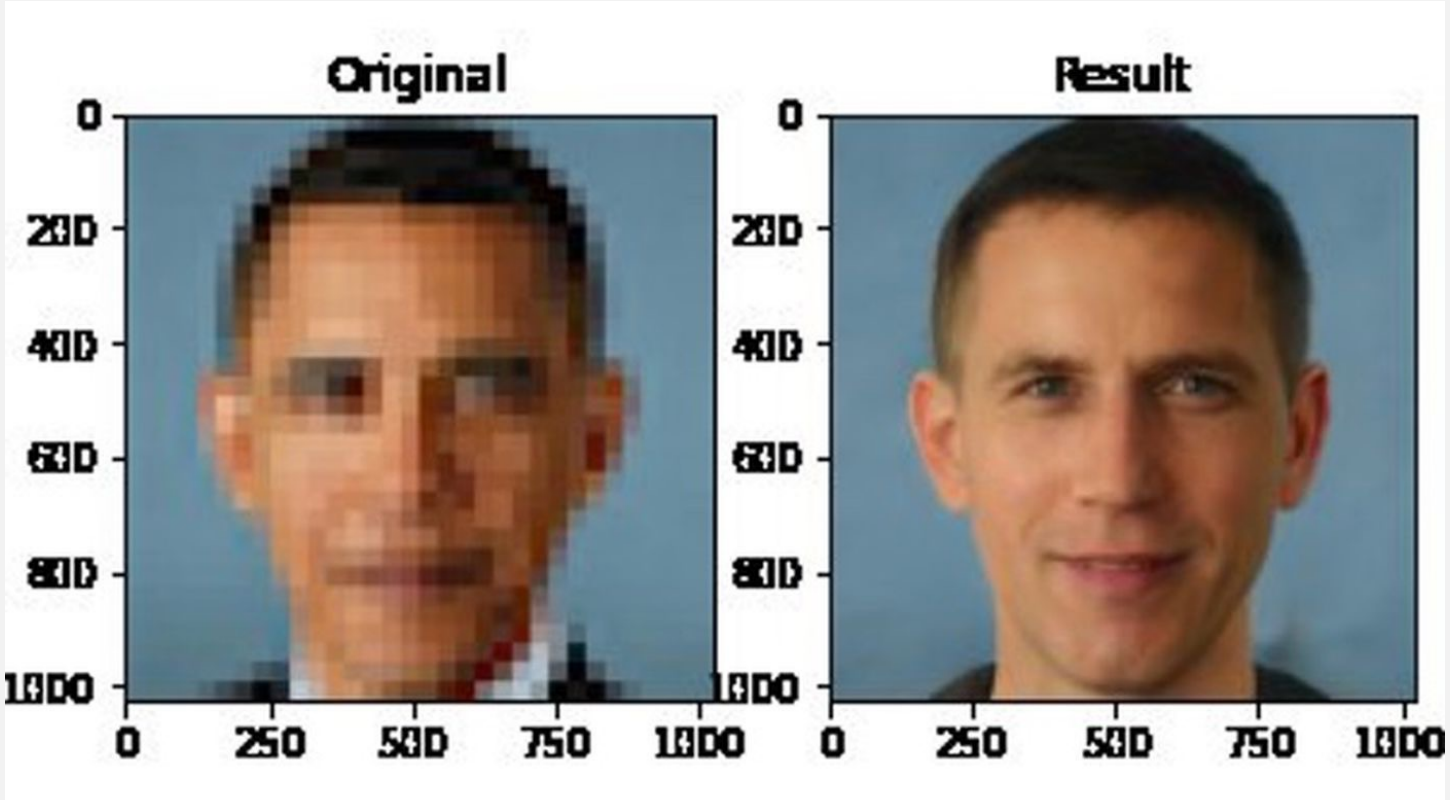
https://privacytools.seas.harvard.edu/differential-privacy

FEDERATED LEARNING



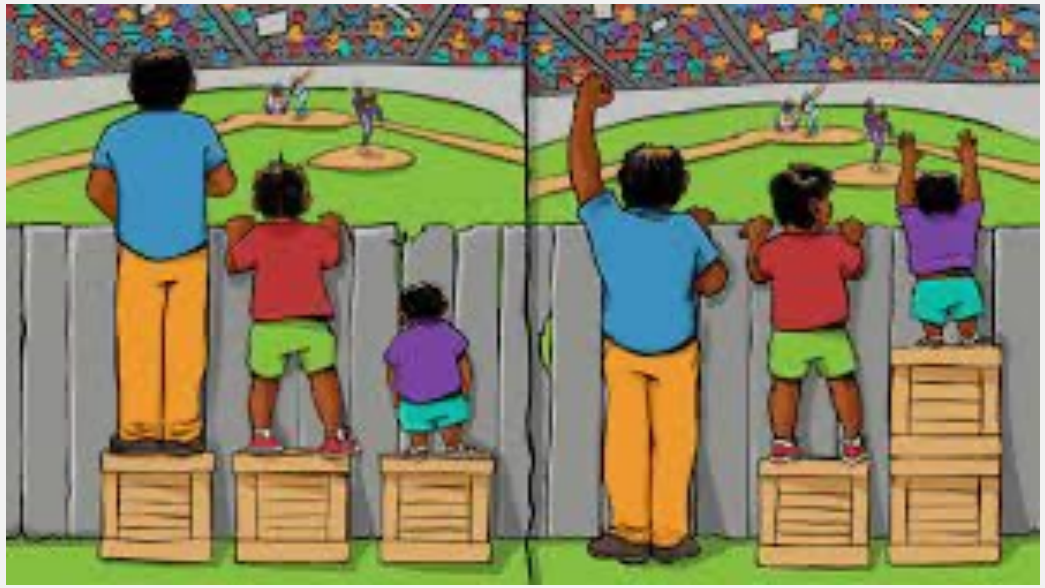
https://proandroiddev.com/federated-learning-e79e054c33ef

BIAS



https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

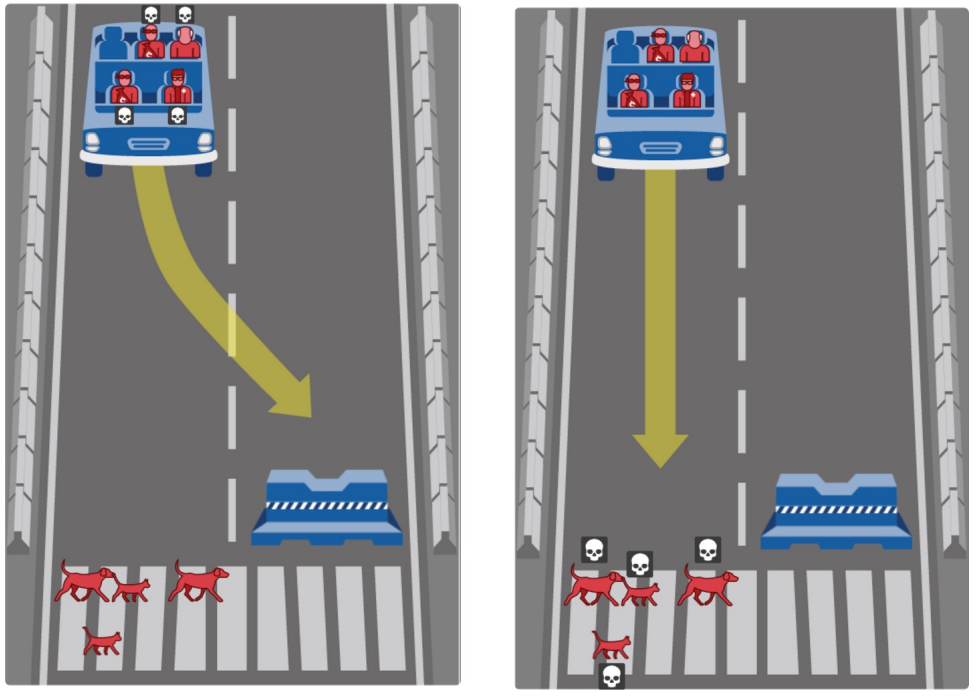
FAIRNESS



https://interactioninstitute.org/illustrating-equality-vs-equity/

ETHICS

What should the self-driving car do?

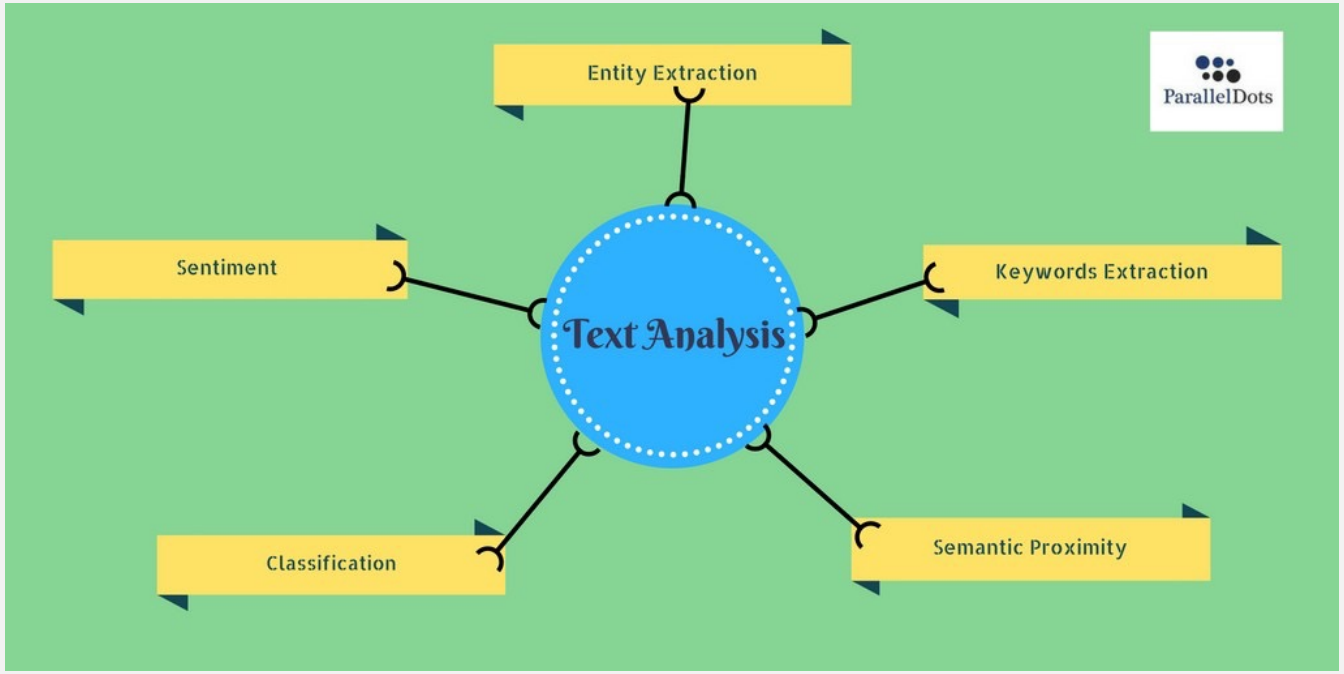


https://techcrunch.com/2016/10/04/did-you-save-the-cat-or-the-kid/

GENERATIVE MODELS

https://www.arabnews.com/node/2044866/media

NATURAL LANGUAGE PROCESSING



https://miro.medium.com/max/1024/1*Am_xQ4sMBIcwTsaPEdqQow.jpeg

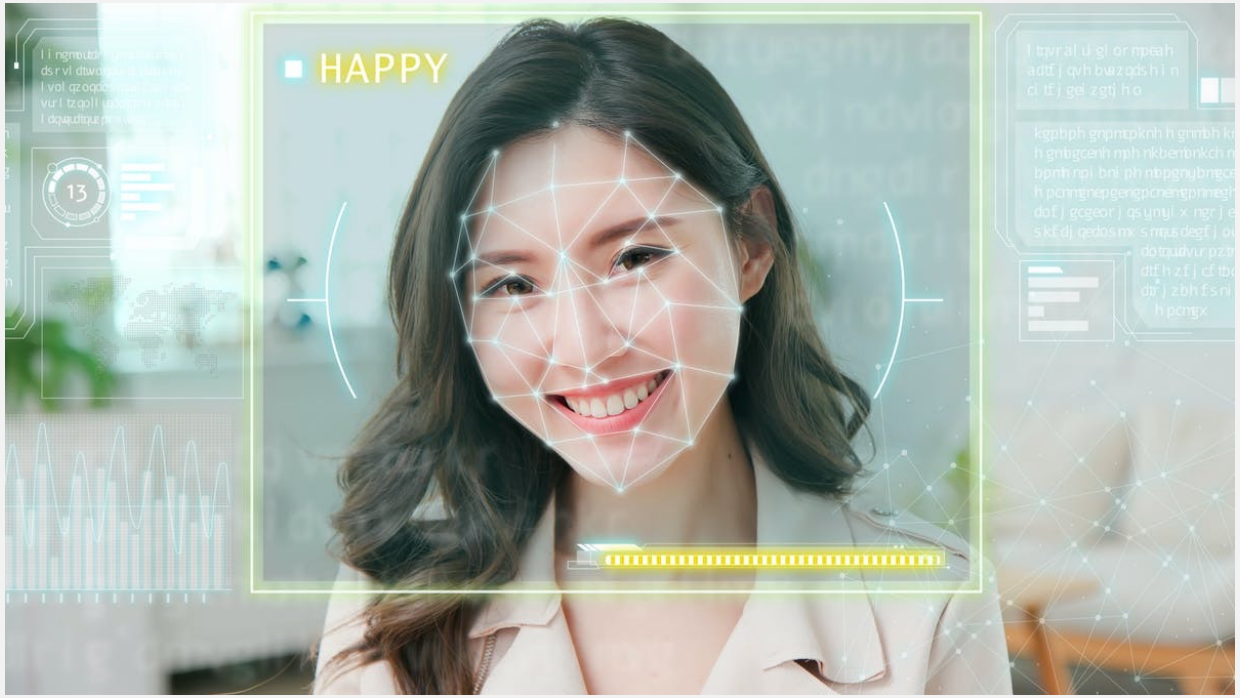
BIOLOGY

https://www.mayo.edu/research/core-resources/bioinformatics-core/overview

MEDICINE AND HEALTH

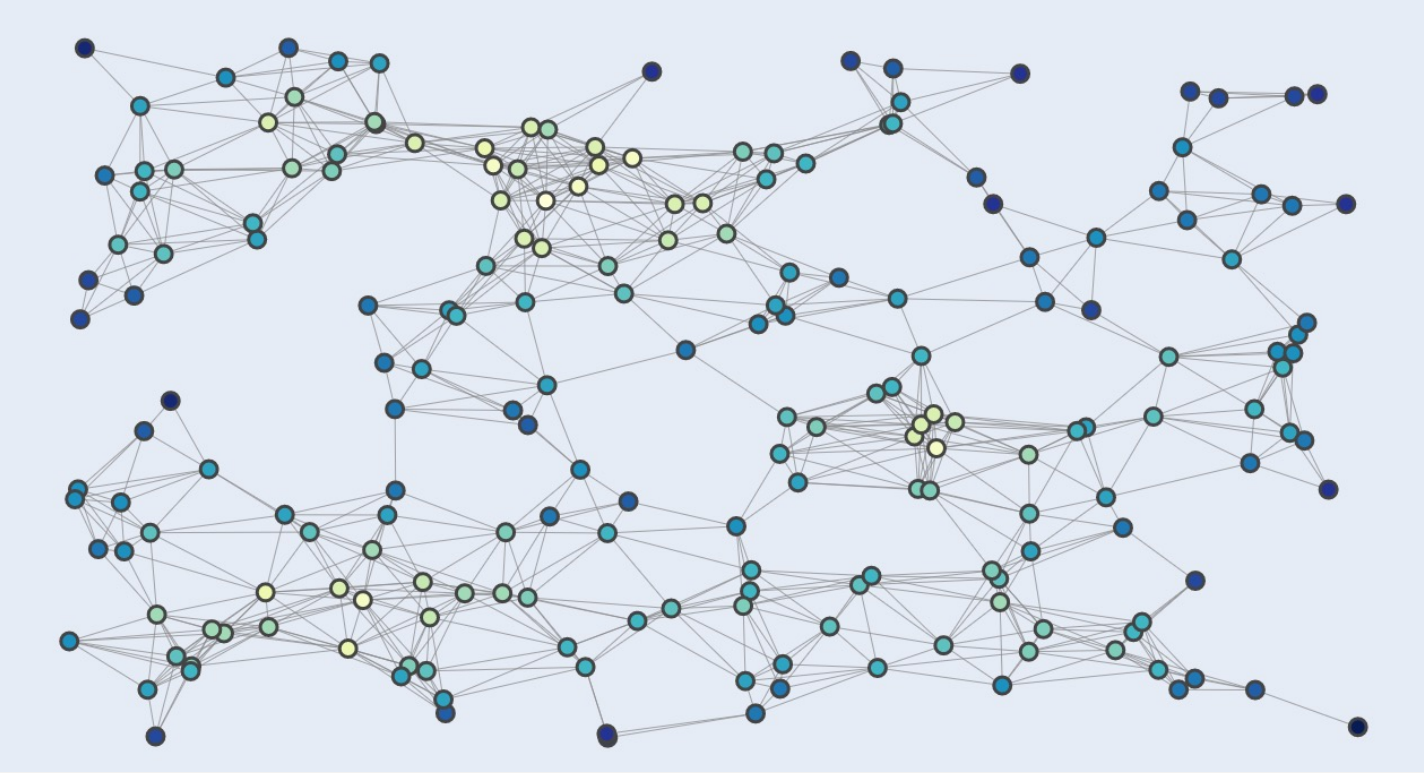
https://humane-ai.nl/students-corner/artificial-intelligence-in-health-care-why-ai-applications-are-not-trustworthy-yet/

DIGITAL PHENOTYPING



https://theconversation.com/ai-is-increasingly-being-used-to-identify-emotions-heres-whats-at-stake-158809

SOCIAL NETWORK ANALYSIS



https://miro.medium.com/max/1838/1*geuswBuLVAbonIviFKzesQ.png

OUTLINE

- Human-Centered Artificial Intelligence
- Course Schedule and Topics
- Coursework and Grading

ASSIGNMENTS AND GRADING BREAKDOWN

- **Paper reflection paragraphs**: 20% (up to 5 can be dropped; extra credit if not dropped)
- **Topic review paper** (3 pages): 20%
- **Project proposal paper** (5 pages): 30%
- **In-class presentations**: 10%
- **Participation and attendance**: 10%
- **Leading discussion on chosen topic**: 10%

All assignments are submitted online via Laulima

PAPER REFLECTION PARAGRAPHS

- Submit a one-paragraph reflection of the paper assigned for each class BEFORE the start of the class
- This is NOT a summary of the paper, but rather a critical reflection
- Do not spend more than 30 minutes reading each paper
  - Unless you really want to
  - The papers can be long and dense: learn to extract the key information
  - This is good practice for reading research papers in general

TOPIC REVIEW PAPER

- 3 pages
- Choose one of the course topics
- Identify a sub-topic
- Review at least 8 papers
- Not a summary of papers, but rather a critical analysis of the papers (for an A- or above)
- We will cover details about review paper format in class
- Option to use this paper as a contribution for co-authoring a review paper on HAI (not required)

PROJECT PROPOSAL PAPER

- 5 pages
- Propose a novel research project in HAI
- If feasible / a good match, opportunity to take project forward in future semesters via ICS 499, ICS 699, ICS 700, etc
- Can be in any topic covered in class or beyond
- We will cover details about the proposal paper format in class

IN-CLASS PRESENTATIONS

- (1) Literature review
- (2) Project proposal midpoint presentation
- (3) Project proposal final presentation
- All presentations are limited to 6 minutes
- We will cover details about presentation format in class

LEADING A DISCUSSION SECTION

- Lead class discussion, for 1 class per semester, about the reading for that day's class
- Come up with discussion questions beforehand and ask the class
- While at least 1 question should be about the reading, some can also be about broader topics related to the reading or the class topic in general
- Call on classmates to answer
- Keep the conversation going and keep it interesting
- Will require much more thorough reading of the paper
- Keep the conversation going

PARTICIPATION AND ATTENDANCE

- Show up to every class
  - Email me if you need to miss class for a valid reason (e.g., academic conference, sick, family emergency; too much homework is not valid)
- Make at least one substantial comment per class

LEARNING OBJECTIVES

- Gain familiarity with the state-of-the-art in HAI
- Learn to analyze HCI and AI research critically
- Practice communicating and discussing state-of-the-art ideas at the intersection of HCI and AI
- Practice ideating, developing, and communicating novel and impactful research ideas in HAI both in written and verbal form

UPCOMING ASSIGNMENTS DUE

- Wednesday August 24: Greener reflection
- Monday August 29: Literature review topic
- Wednesday August 31: Proposal topic

LITERATURE REVIEW TOPIC SUBMISSION

- Choose a topic from the course calendar that interests you based on the overview from last class
- Provide an additional sentence summarizing the lens through which you will analyze the topic
- Submit on Laulima
- Example submissions:
  1. *Topic: Crowdsourcing. Lens: I will review the use (or lack thereof) of fair payment practices for data labeling on crowdsourcing platforms such as Amazon Mechanical Turk.*
  2. *Topic: Interpretable ML. Lens: Lens: I will review the state-of-the-art in methods for interpretability of multimodal prediction models.*

DISCUSSION TOPIC SUBMISSION

- Provide your ranking of the top-10 topics you'd like to lead
- Alternatively, you may say "I am fine with leading any discussion topic"
- Submit on Laulima

PROPOSAL TOPIC SUBMISSION

- Choose a topic from the course calendar that interests you based on the overview from last class
- Provide a paragraph-long summary describing your project proposal
- Submit on Laulima
- Requested format:
  - *Project Category: [choose from course calendar]*
  - *Societal and/or Technical Motivation: [text]*
  - *Dataset: [include size, data description, and whether it is a publicly available dataset or how the data will be collected]*
  - *Methods: [a few sentences about the proposed methodology]*