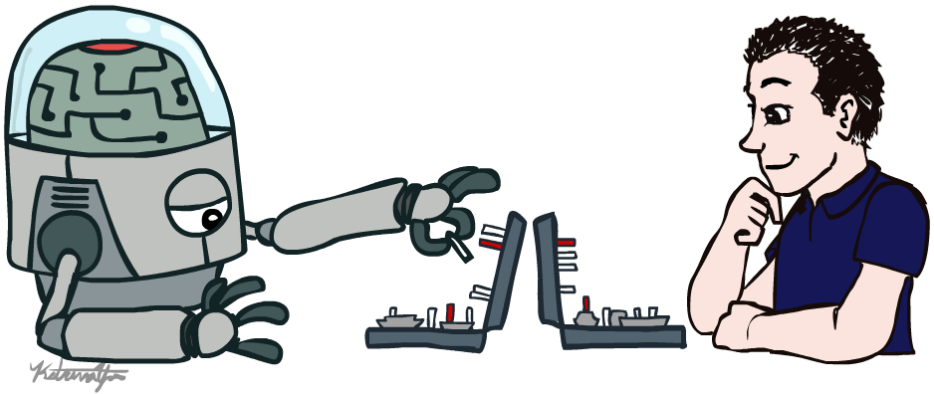
# CS 5955/6955 Advanced Artificial Intelligence

## Introduction



Instructor: Daniel Brown

University of Utah

[some slides and images adapted from those created by Dan Klein and Pieter Abbeel: http://ai.berkeley.edu.]

# Course Staff

## Professor

Daniel Brown

## TA

Connor Mattson

# Course Information

- Communication:
  - Announcements on Canvas (usually also posted on Piazza)
  - Questions and Discussion on Piazza
- Course format:
  - Reading assignments and programming assignments turned in via Canvas.
  - No midterm or final
- Class Website:
  - https://dsbrown1331.github.io/advanced-ai
  - Schedule
  - Assignment instructions
  - Readings
  - Etc.

# Grading

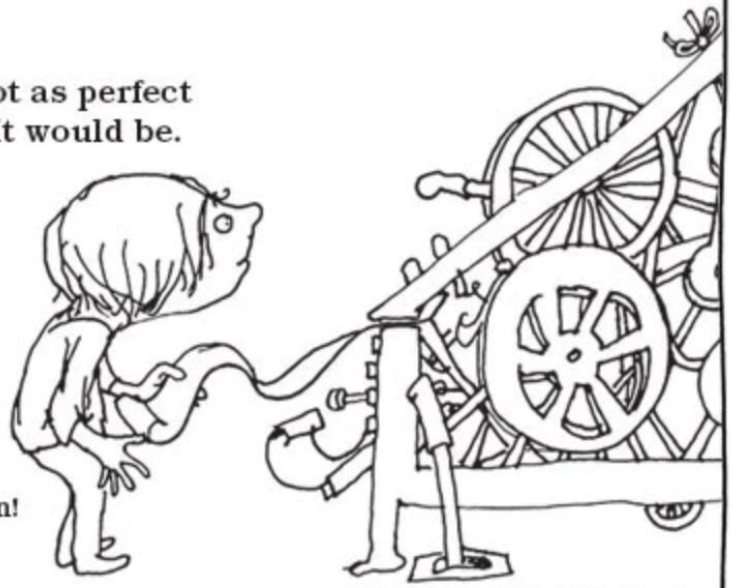
- Programming Assignments: 45%
- Quizzes/Attendance/Reading Reports: 20%
- Final project proposal: 5%
- Final project presentation: 10%
- Final project written report: 20%
- All submissions due electronically **by midnight** on due date.
- There is a moratorium on complaints about grading, etc., of **one week after grades are released.**

# Use of AI in class

- Have fun and use whatever tools you find beneficial for helping with coding, debugging, experimenting, brainstorming, etc.
  - Beware blindly copying and pasting. You won't learn anything that way.
- You need to actually write up the results and discussion and summary yourself.
  - Writing builds critical thinking skills!
  - Forces you to realize what you do and don't understand. If you can't explain it simply in your own words, then you don't really understand something.

## HOMEWORK MACHINE

The Homework Machine, oh the Homework Machine,
Most perfect contraption that's ever been seen.
Just put in your homework, then drop in a dime,
Snap on the switch, and in ten seconds' time,
Your homework comes out, quick and clean as can be.
Here it is—"nine plus four?" and the answer is "three."
Three?
Oh me . . .
I guess it's not as perfect
As I thought it would be.

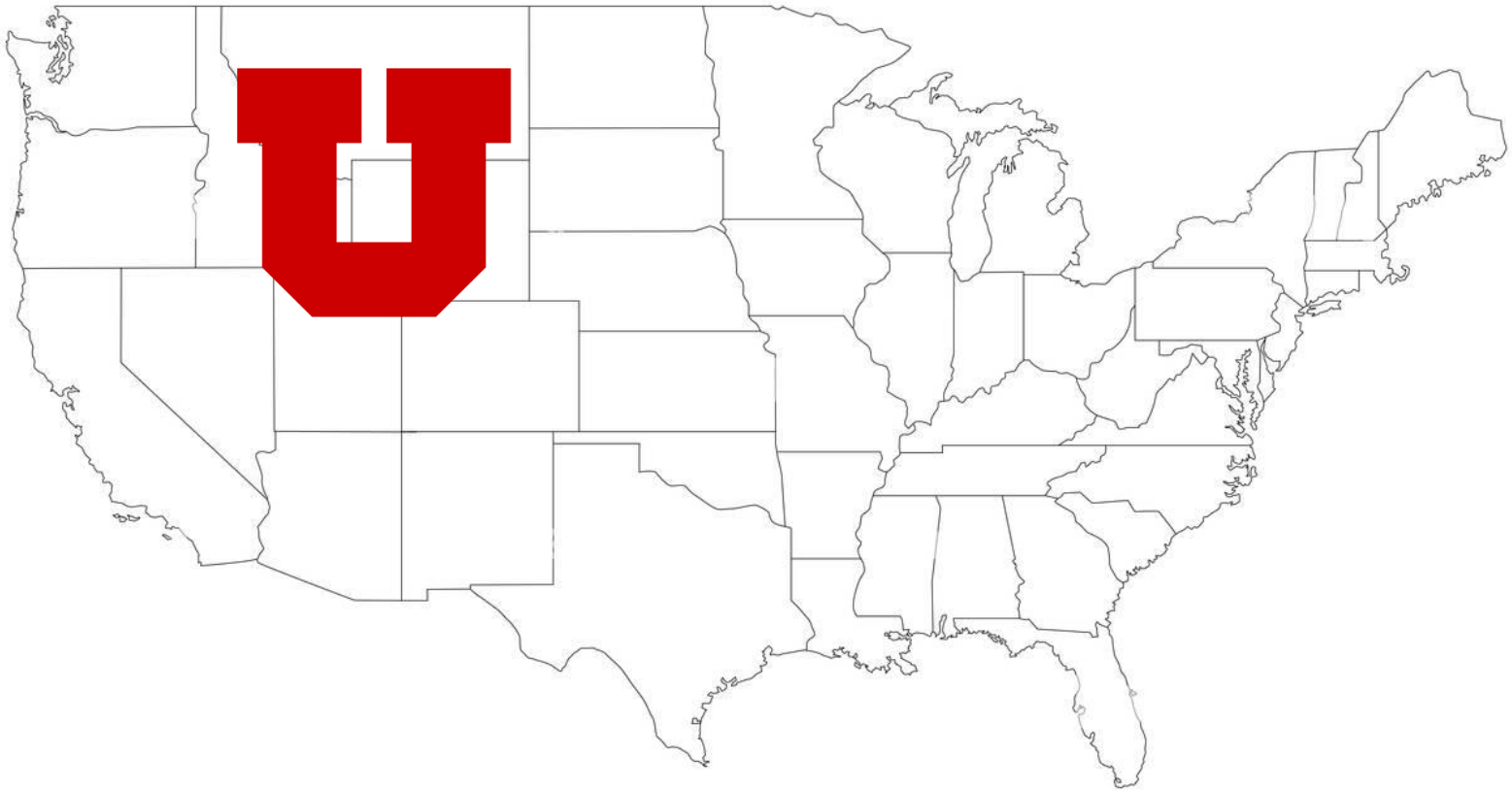


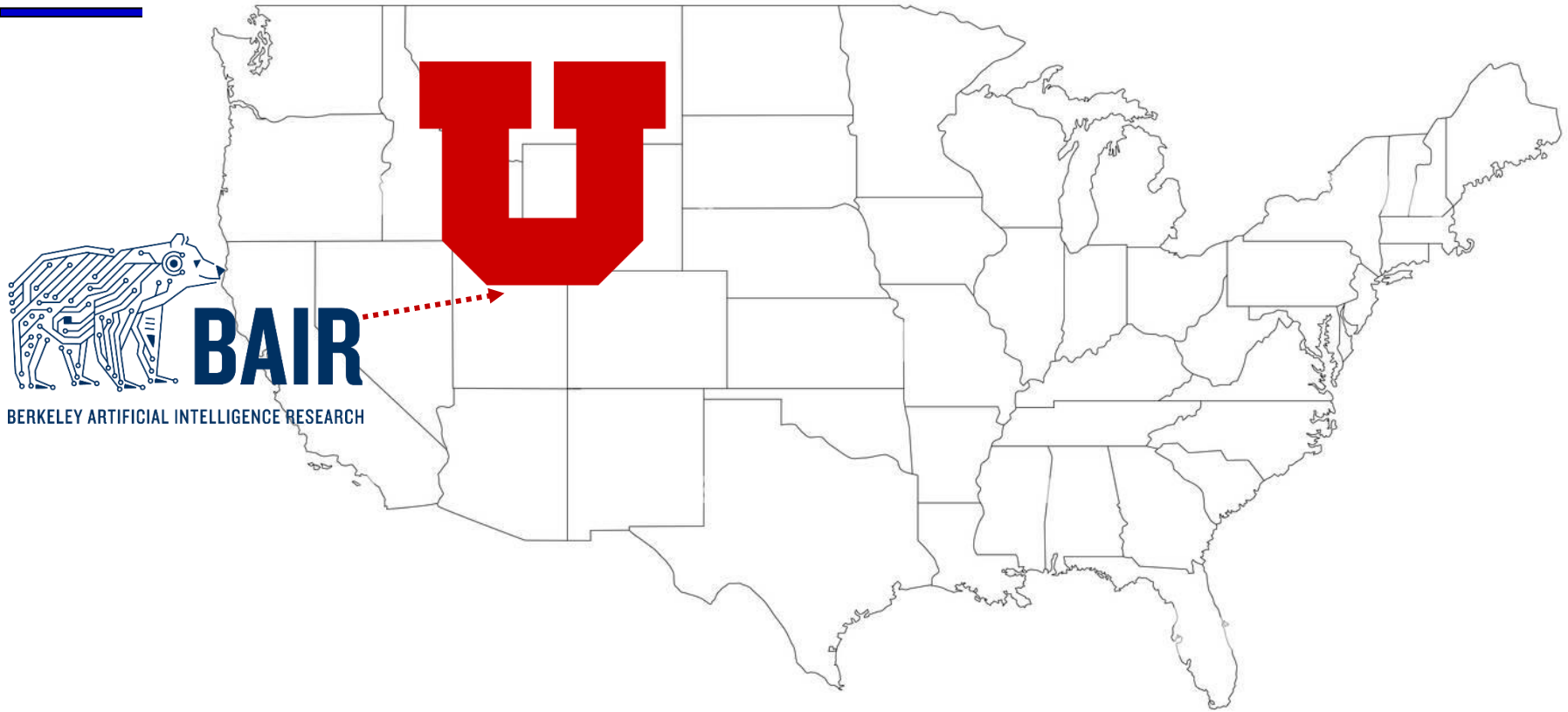
Read more poems in *A Light in the Attic* by Shel Silverstein!

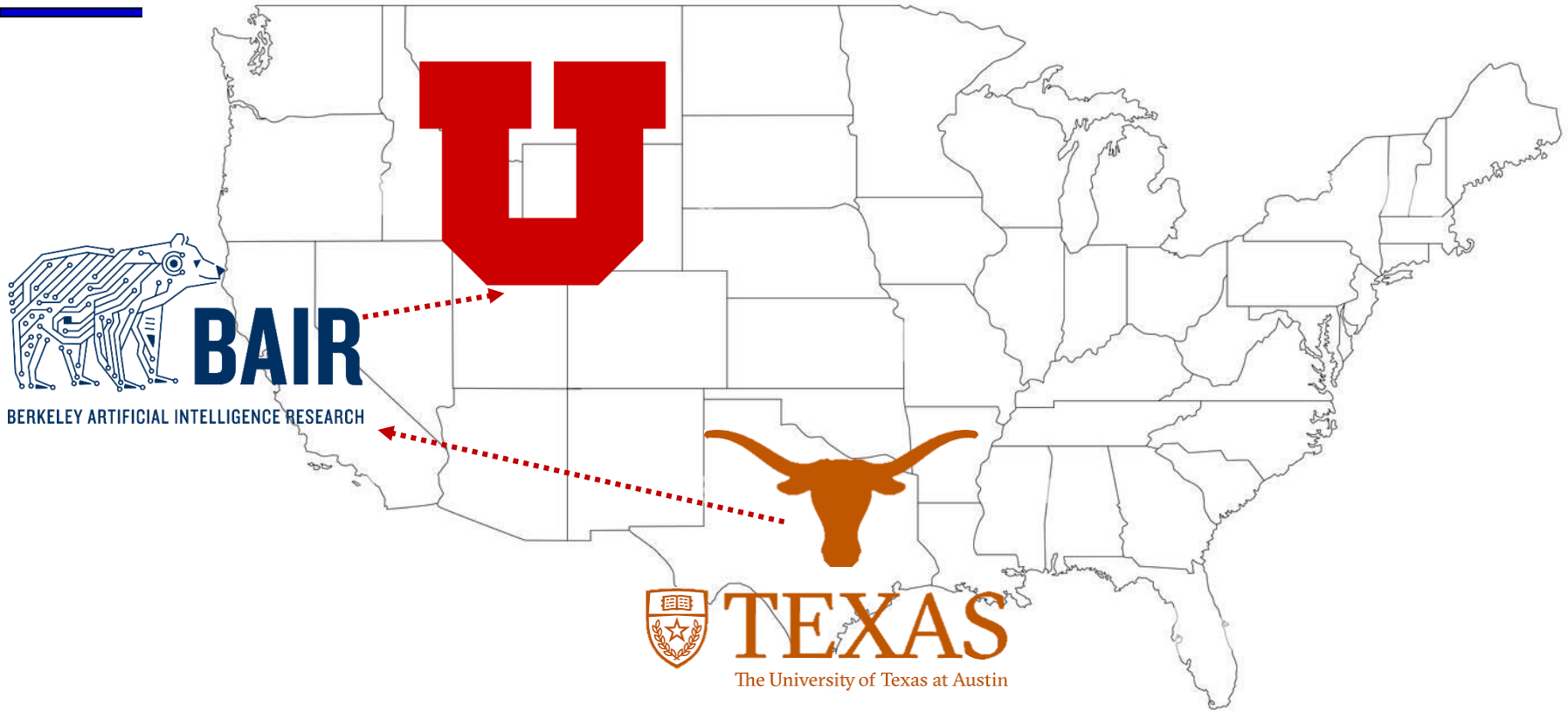
# Important for this week

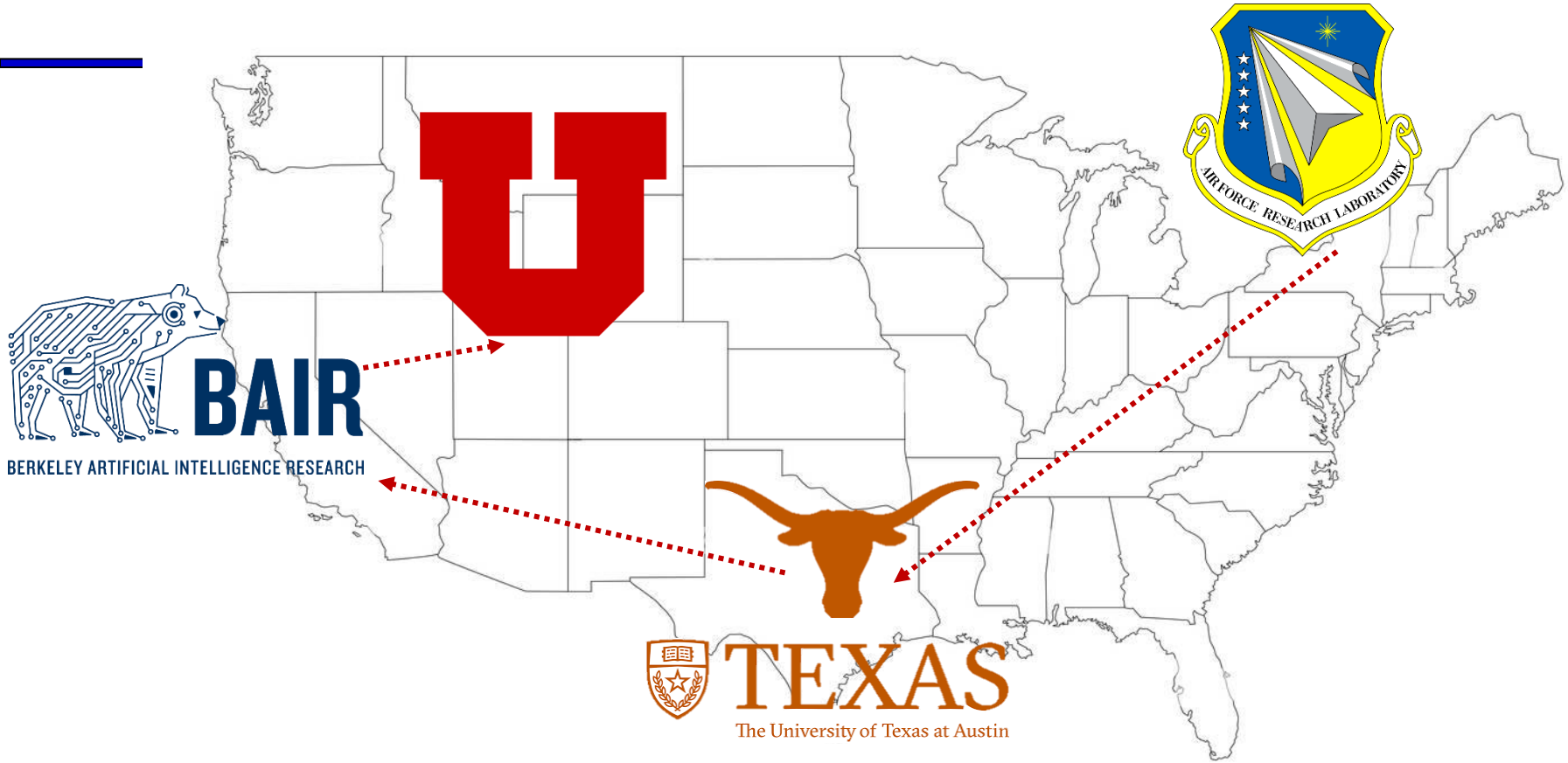
- Register for class on Piazza
- Brush up on Python if you're rusty (see links on class website)

# A little about me



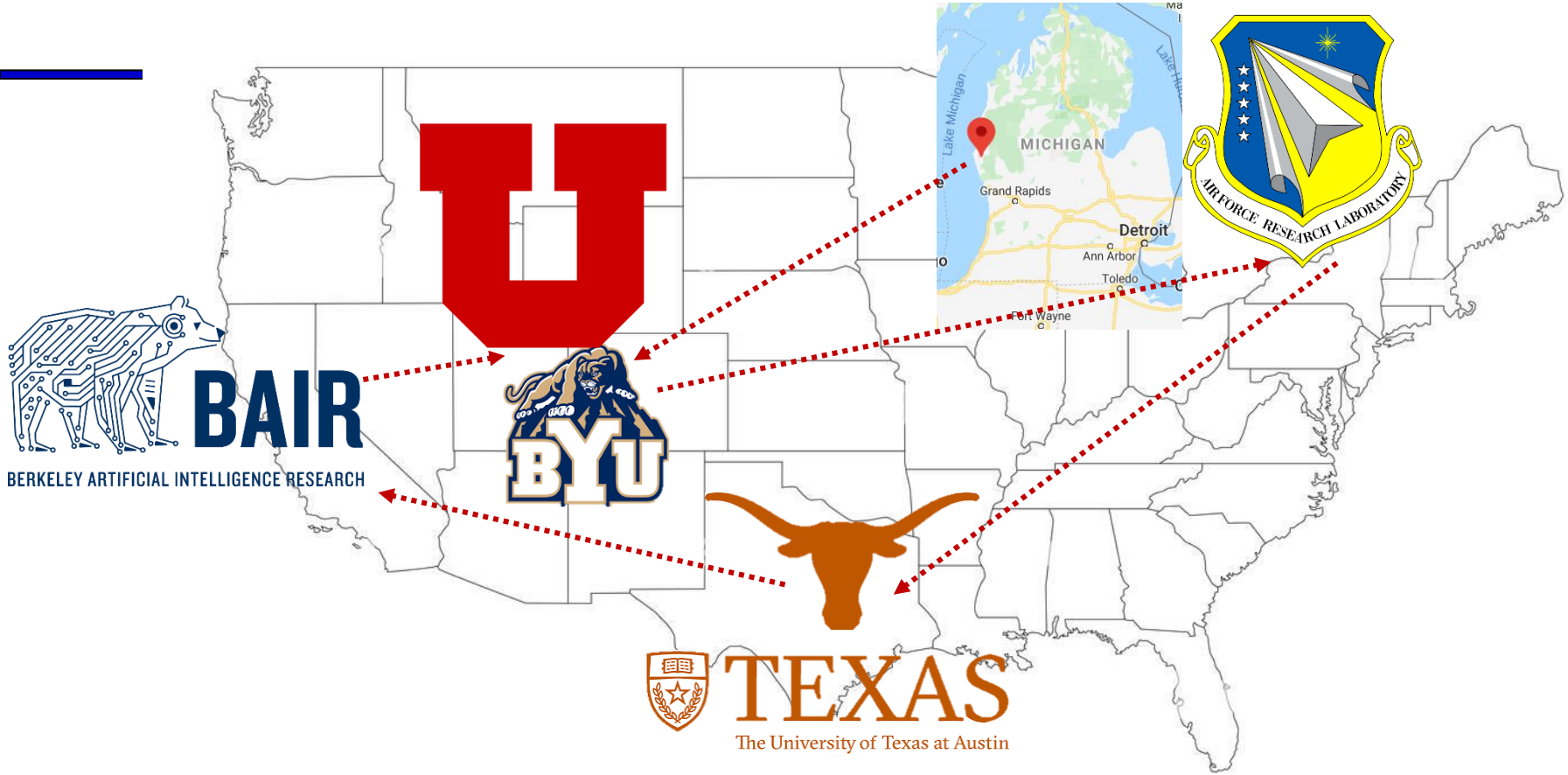

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

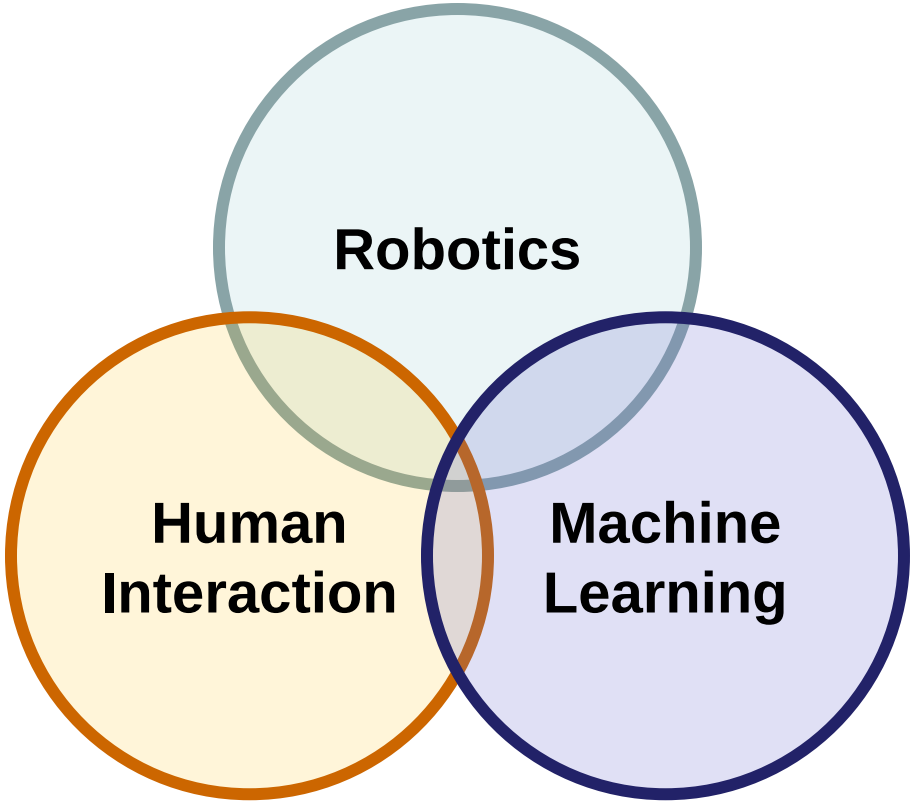

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
TEXAS
The University of Texas at Austin


BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
AIR FORCE RESEARCH LABORATORY
TEXAS
The University of Texas at Austin

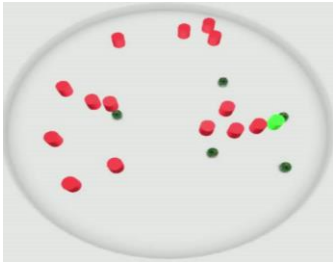

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
BYU
AIR FORCE RESEARCH LABORATORY
TEXAS
The University of Texas at Austin

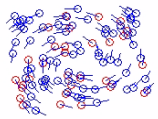

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
BYU
AIR FORCE RESEARCH LABORATORY
TEXAS
The University of Texas at Austin


MICHIGAN
Grand Rapids
Detroit
Ann Arbor
Toledo
Fort Wayne
Lake Michigan
AIR FORCE RESEARCH LABORATORY
BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
BYU
TEXAS
The University of Texas at Austin

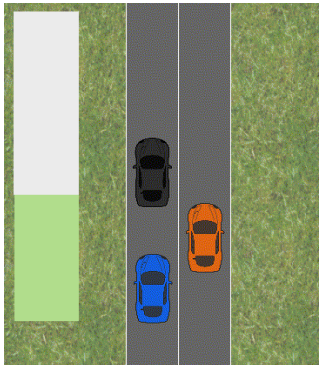

Robotics
Human Interaction
Machine Learning

# Human-Robot Interaction

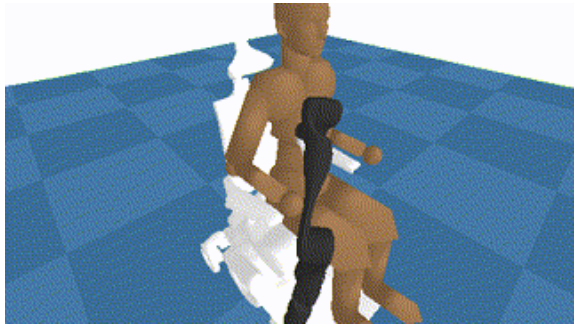




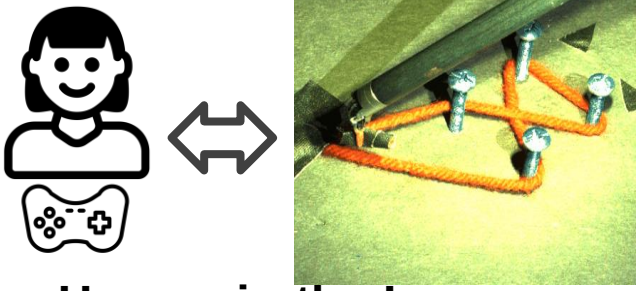
**Human Swarm Interactions**



**Autonomous Driving**

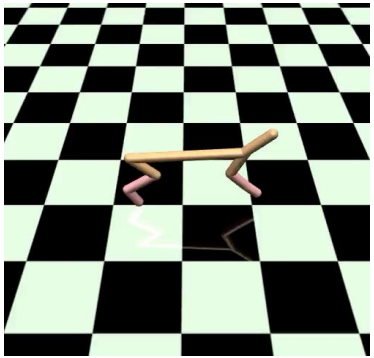


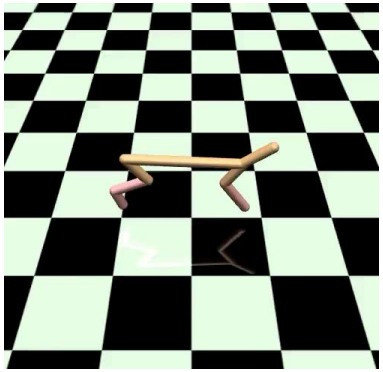
**Shared Autonomy and Assistive Robotics**



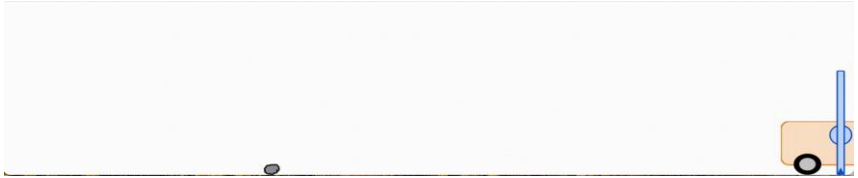
**Human-in-the-Loop Robot Learning**

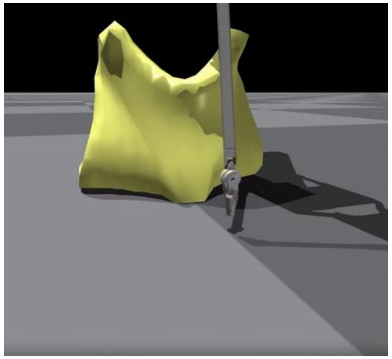
# Learning models of human preferences

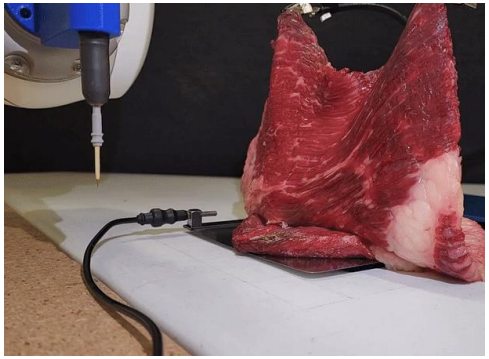




“Great weather today!”

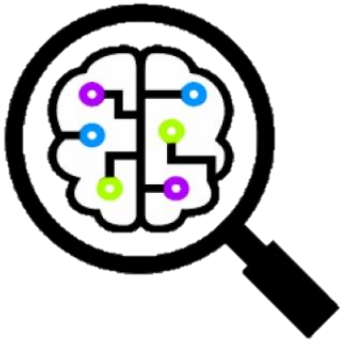






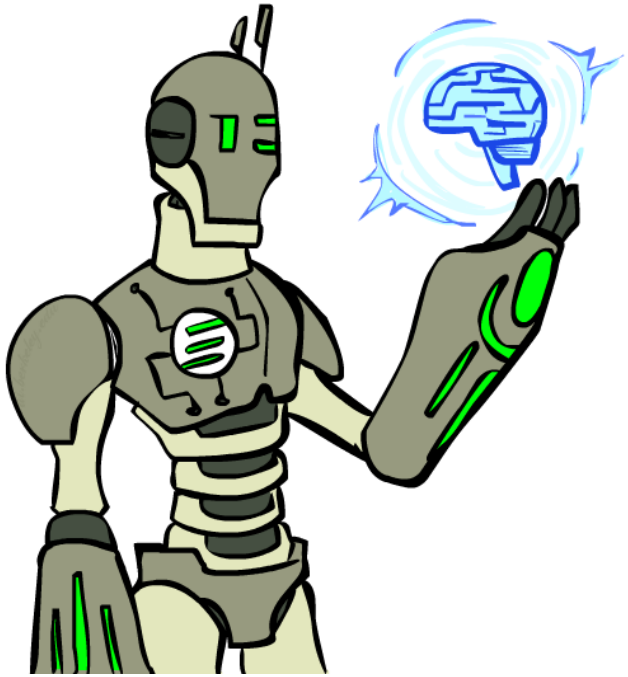
“Didn’t get the job offer…”

# AI Safety and Robustness



# Today

- What is artificial intelligence?
- How is it different from machine learning?
- What will we cover in this class?




ambi
ROBOTICS


foods, better choices


WAYMO

# Smart Chatbots

ChatGPT

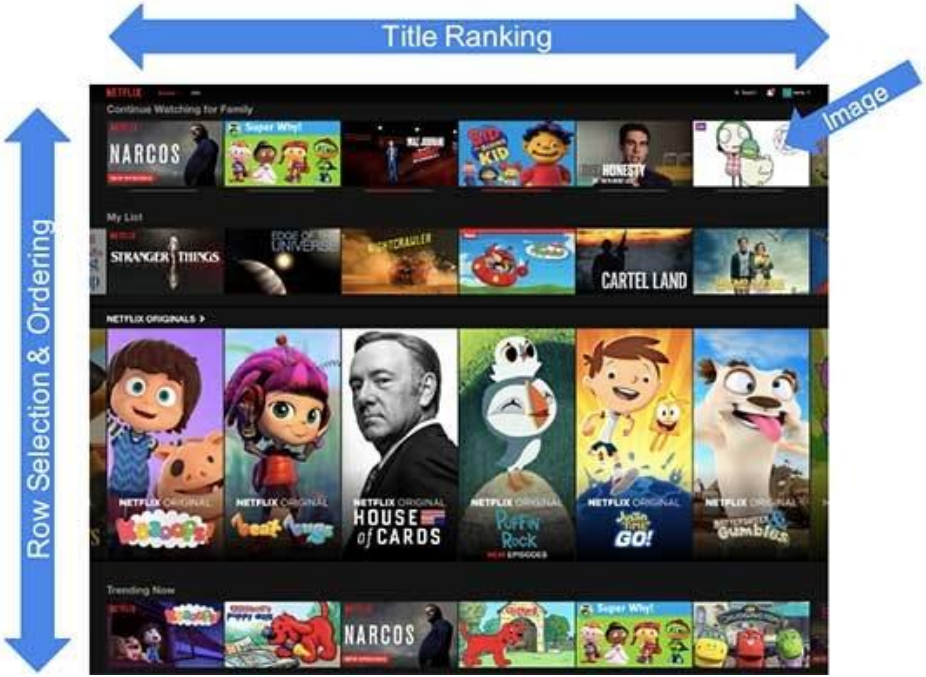
Gemini

Claude

BY ANTHROP\C

# Entertainment

## Everything is a Recommendation

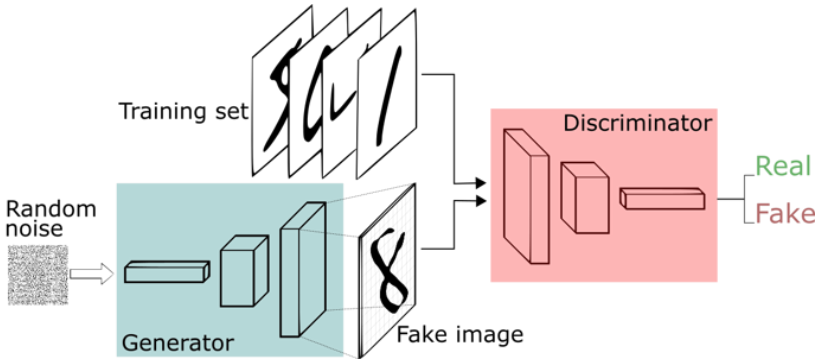


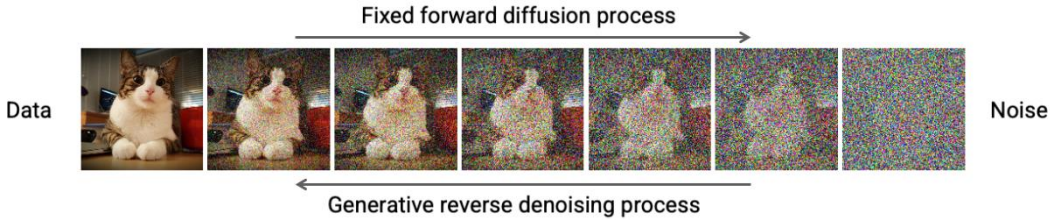

Recommendations are driven by machine learning algorithms

**Over 80%** of what members watch comes from our recommendations

# Education


# Generating Images



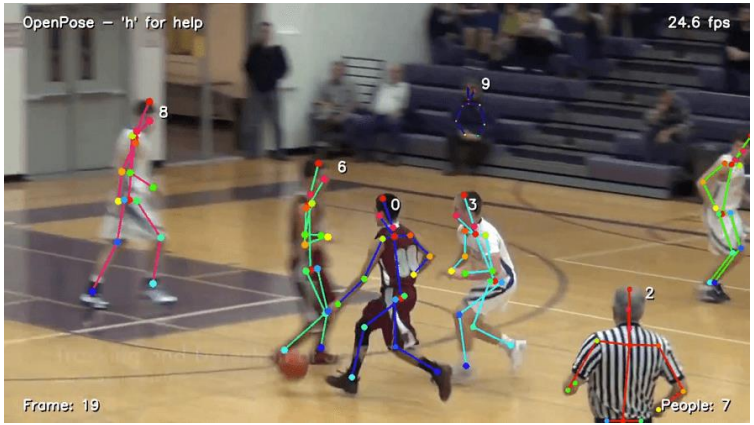

# Text to Images (DALL-E)

- “An astronaut riding a horse in a photo-realistic way”

- “An armchair in the shape of an avocado.”

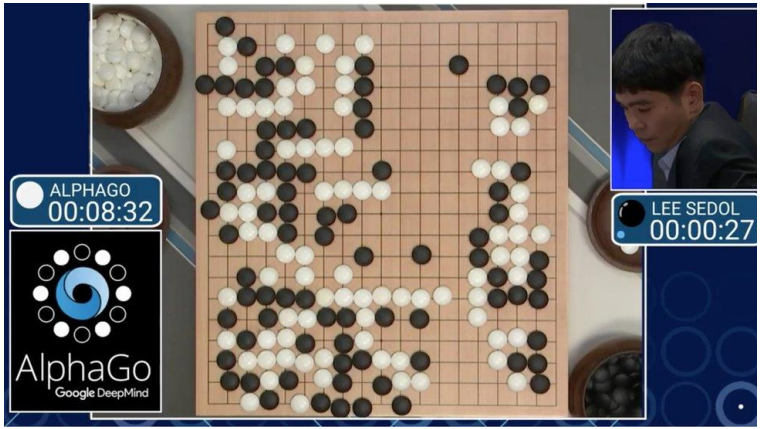
# Vision (Perception)

- Object and face recognition
- Scene segmentation
- Image classification





# Super-Human Performance at Games

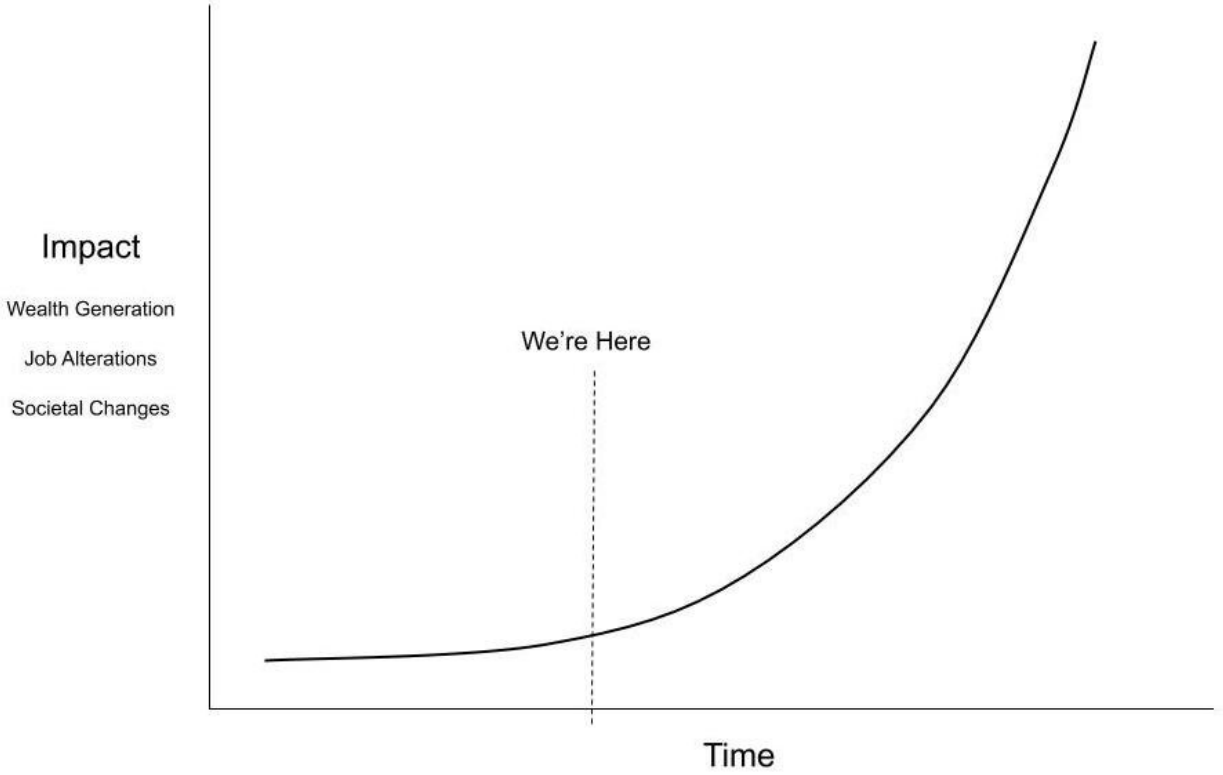

# Boston Dynamics Atlas

# https://vision-locomotion.github.io/

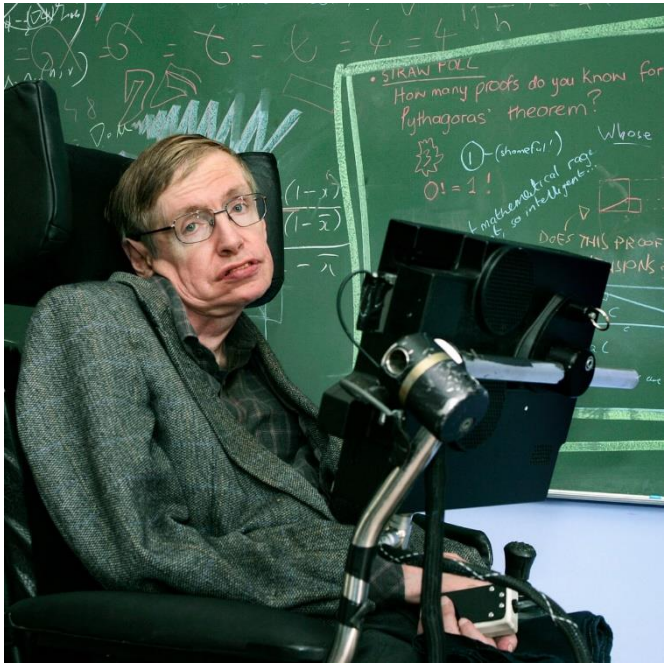
# Videos

# Deep Fakes and Reality

# Progress is expected to continue!



# Should we be worried?




"The development of full artificial intelligence could spell the end of the human race.“

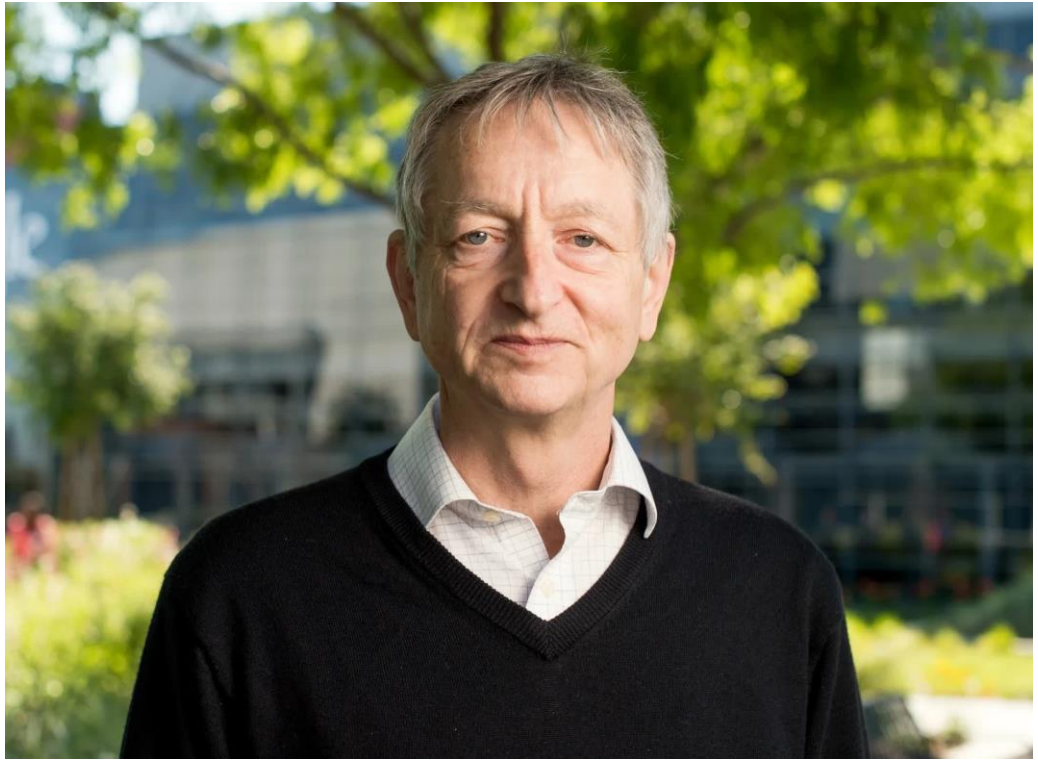
-Stephen Hawking

# Should we be worried?


“AI is a fundamental existential risk for human civilization.”

-Elon Musk

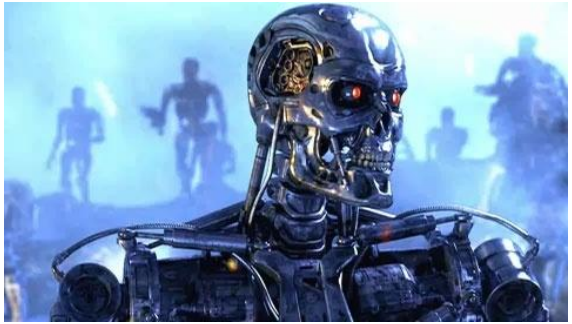
# Geoffrey Hinton speaks out about the risks of AI



# Worries about Advanced AI

- AI enabling people to do bad things:
  - Enabling cyber attacks, bio-terrorism, disinformation, etc.
- Self-improvement loop
  - AI automates AI research and development
- Unintended consequences
  - Deception, scheming, and manipulation
  - Power seeking
  - AI gets out of control

# What is AI?



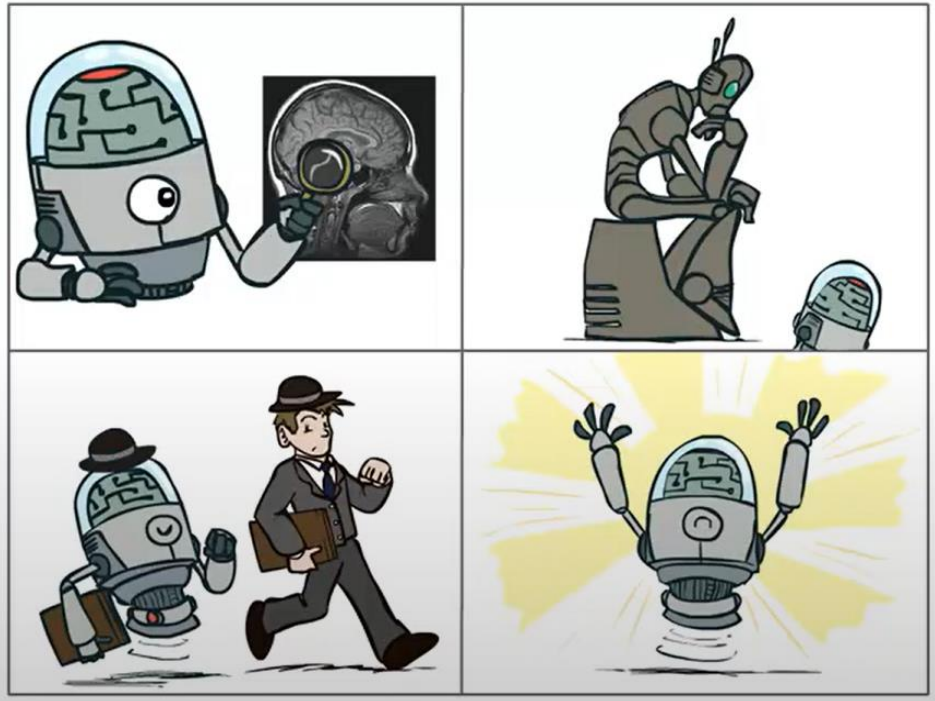

# What is AI?

- **John McCarthy (1955) (He coined the term AI)**
  - *"The science and engineering of making intelligent machines."*
- **Alan Turing (1950)**:
  - *"A machine that can mimic any aspect of human intelligence.“*
- **European Commission (2018)**:
  - *"Artificial intelligence refers to systems that display intelligent behavior by analyzing their environment and taking actions—with some degree of autonomy—to achieve specific goals."*

# What is AI?

The science of making machines that:

Think like people



Think rationally

Act like people

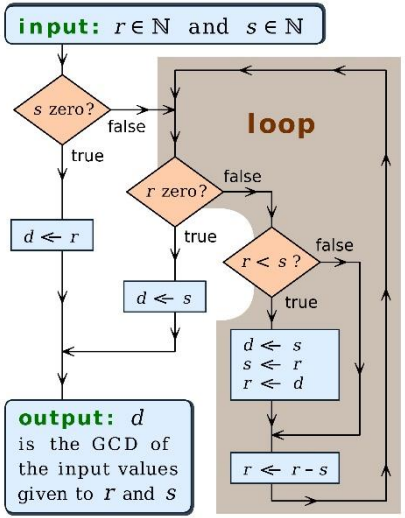
Act rationally

# What about machine learning?

- **"Machine Learning is the study of computer algorithms that improve automatically through experience and by the use of data.“** Tom M. Mitchell (1997)
  - It’s a critical tool for achieving AI

# Brief History of AI

## Algorithms -- 1980s (TELL)

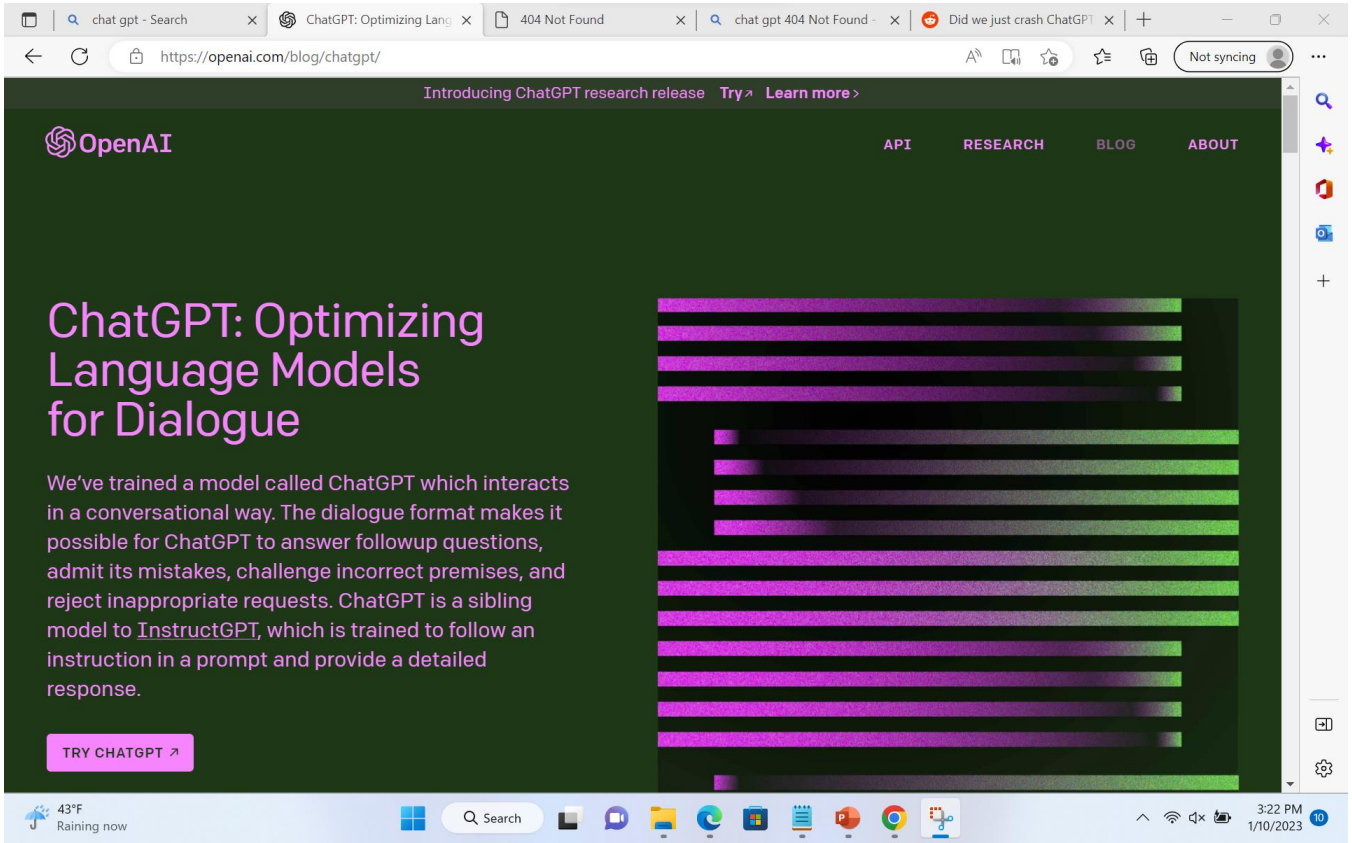



## Big Data -- 2000s (SHOW)

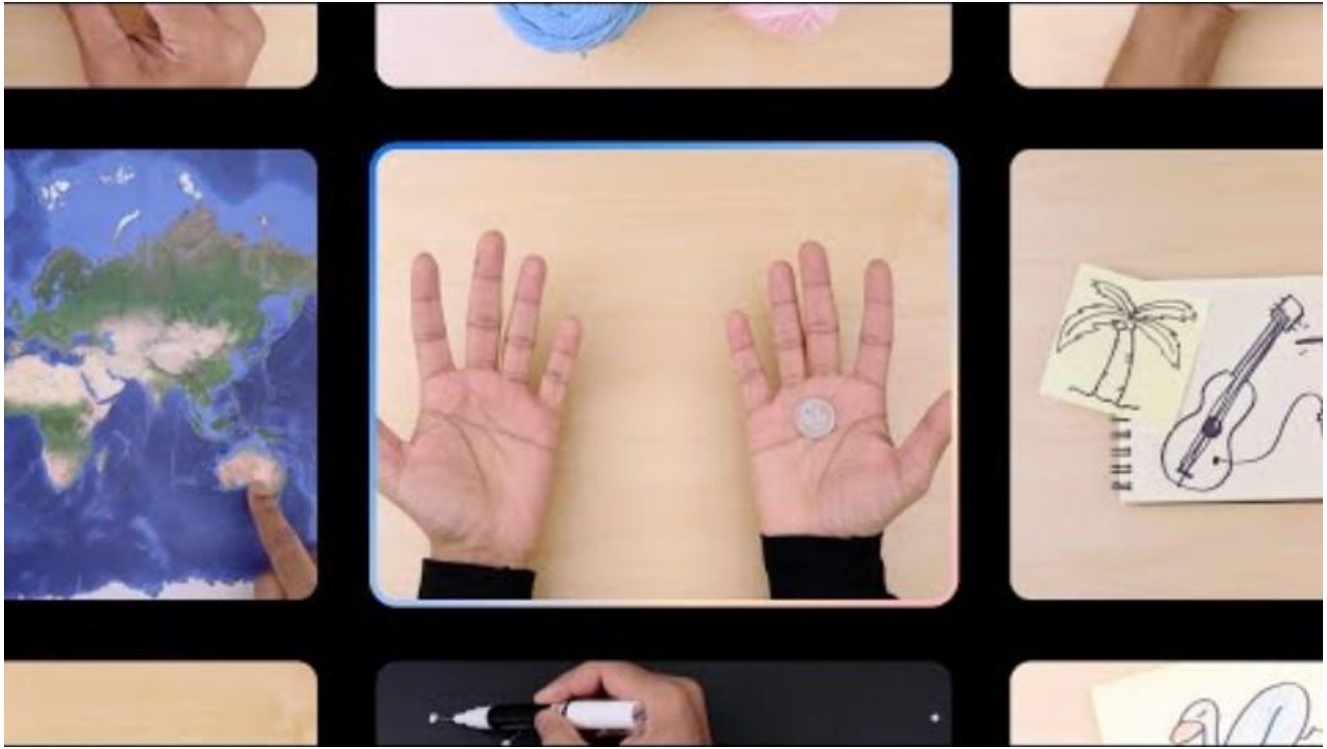
## Objectives -- 2020s (WANT/NEED)


Credit: Peter Norvig

# Natural Language Processing

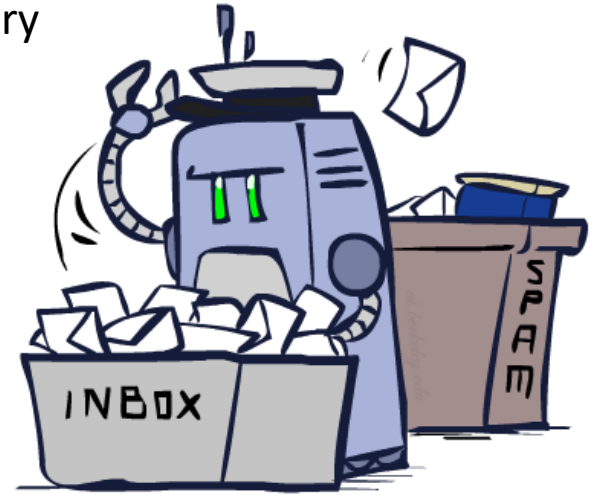


# Multi-Modal Models



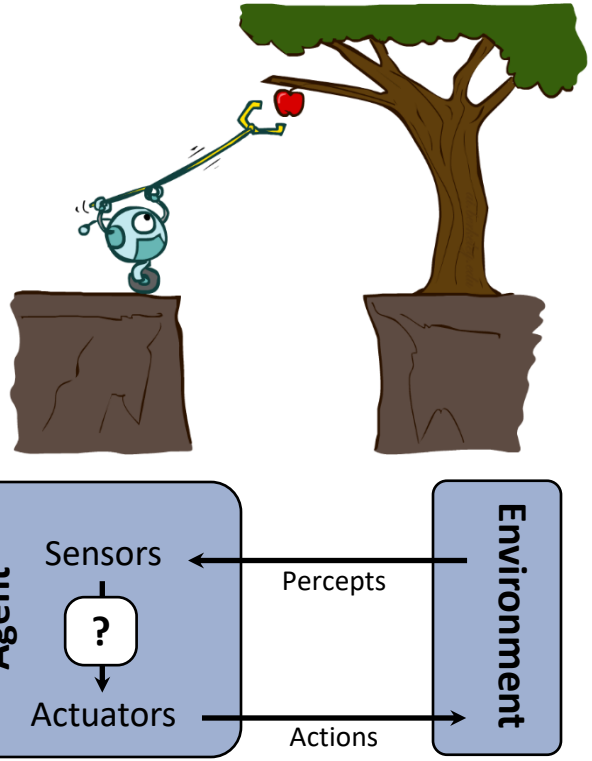
# Decision Making

- Applied AI involves many kinds of automation
  - Scheduling, e.g. airline routing, military
  - Route planning, e.g. Google maps
  - Medical diagnosis
  - Web search engines
  - Spam classifiers
  - Automated help desks
  - Fraud detection
  - Product recommendations
  - Service robots
  - … Lots more!



# Designing Rational Agents

- An **agent** is an entity that *perceives* and *acts*.
- A **rational agent** selects actions that maximize its (expected) **utility**.
- Characteristics of the **percepts, environment,** and **action space** dictate techniques for selecting rational actions
- **This course** is about:
  - AI techniques for a variety of problem types
  - We will use machine learning to design agents/policies



# Main Course Topics

- Learning to make decisions from examples via supervised learning (behavioral cloning)
- Learning to make one-step decisions from evaluative feedback (multi-armed bandits)
- Learning to make multi-step decisions from evaluative feedback (Reinforcement Learning)
  - Lots on this!
- Learning rewards from human feedback
- AI Safety and Alignment