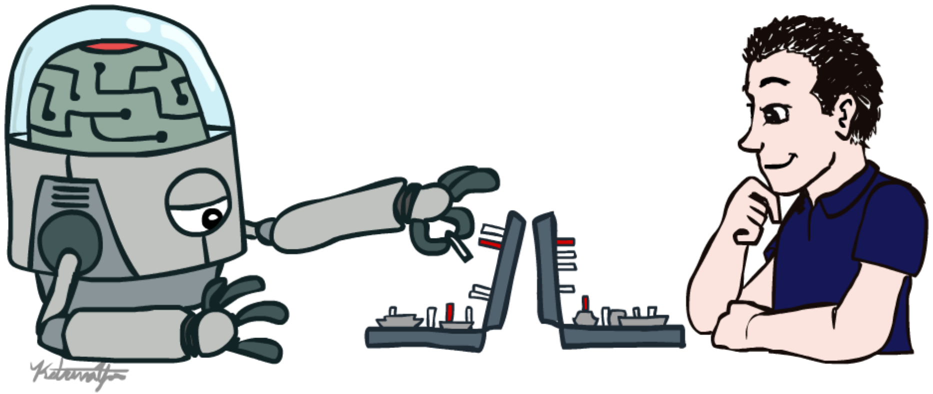
# CSCE 580: Artificial Intelligence

## Introduction



Spring 2025

Prof. Pooyan Jamshidi

University of South Carolina

[Some of these slides were originated by Dan Klein and Pieter Abbeel at UC Berkeley; thanks!]

# First Half of Today: Intros and Logistics

- Staff introductions:
  - Pooyan Jamshidi (me!), Abir Hossen (TA)
- Course Logistics:
  - Lectures, discussions, office hours, and exams
  - Resources and communication platforms
  - Collaboration and academic honesty
  - Disability and extenuating circumstances
  - Stress management and mental health

# Staff Introductions

## Instructor

Pooyan Jamshidi
Assistant Professor
Computer Science and Engineering
University of South Carolina
CV

pjamshid@cse.sc.edu
Storey Innovation Center
Office: 2207 (Lab: 2212)
Follow @PooyanJamshidi

AISys

Home Research AISys Lab Publications Teaching Talks Funding Diversity Bio Service Misc. Consultancy

## TA

**Md. Abir Hossen**
Ph.D. Candidate
Computer Science
University of South Carolina
E-mail: mhossen@email.sc.edu
CV

Area of Interest:

- Autonomous and adaptive systems
- Machine learning systems

**Biography:**

I am a Ph.D. candidate in the Computer Science and Engineering department at University of South Carolina. I am currently working at Artificial Intelligence and Systems Laboratory (AISys) as a Ph.D. research assistant under the supervision of Dr. Pooyan Jamshidi. Prior to starting the Ph.D., I worked as a graduate research assistant at the Unmanned and Swarm System (USS) lab. I completed my M.Sc. in Electrical Engineering from South Dakota School of Mines & Technology and B.Sc. in Electrical & Electronics Engineering from American International University-Bangladesh (AIUB).

# Prof. Pooyan Jamshidi

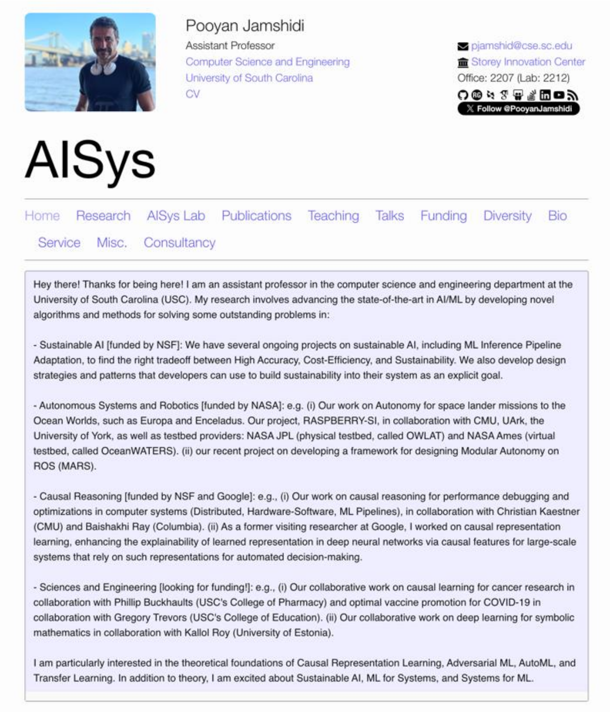
Currently:
- Associate Professor at USC (2025 -- *)
- Director of AISys lab

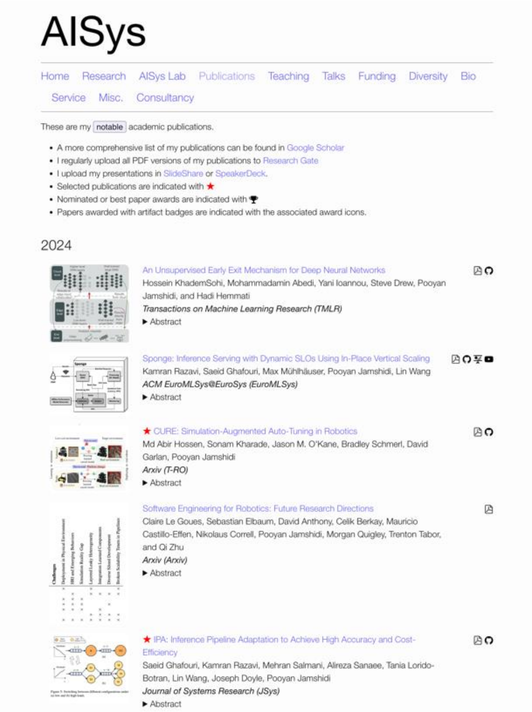
Previously:
- Assistant Professor at USC (2018 -- 2024)
- Researcher at Google (2021-22)
- Postdocs at CMU, Imperial College London

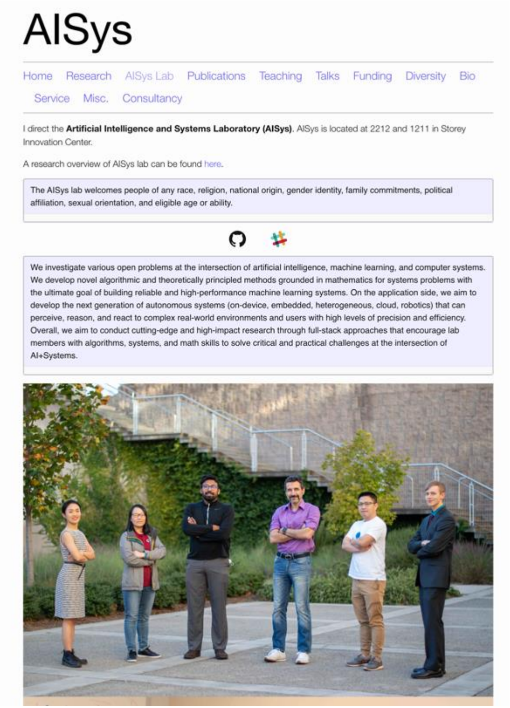
Teaching:
- CSCE 580: Artificial Intelligence
- CSCE 585: ML Systems
- CSCE 212: Intro to Computer Architecture

Research:
- ML Systems
- Causal AI
- Autonomous Systems & Robotics

# Artificial Intelligence and Systems Laboratory



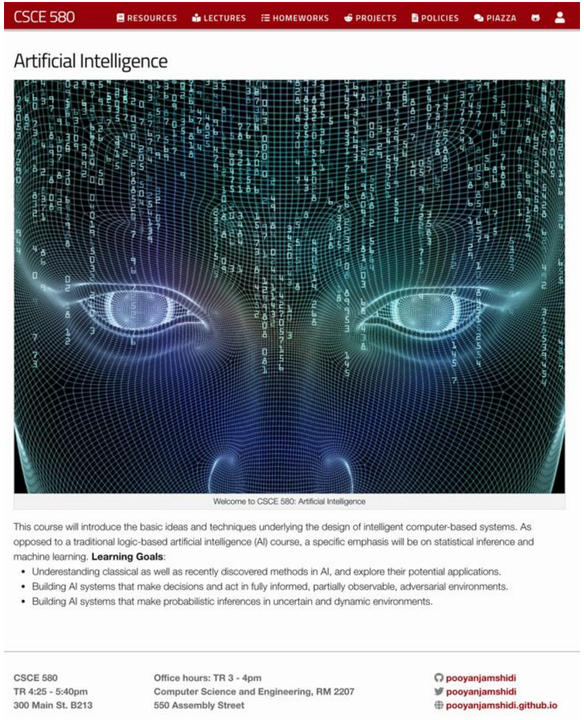




https://pooyanjamshidi.github.io/AISys/

# Course Information

https://pooyanjamshidi.github.io/csce580/

I plan to make minor updates to the lectures, but the syllabus and overall content will remain largely similar to what is currently available.




- Communication:
  - Announcements on webpage/emails/dropbox
  - Questions? Discussion on piazza
  - If not suitable for Piazza?
    - individual staff as needed
- Course technology:
  - Website
  - Piazza
  - Gradescope
  - CSE Dropbox
  - Autograded projects
  - Regular homework
  - Help us make it awesome!

# Course Information

- Course Website: https://pooyanjamshidi.github.io/csce580/
- Piazza: http://piazza.com/sc/spring2025/csce580
  - Discussion boards for each assignment and the course overall
    - PLEASE post questions on course material (don’t be shy)
    - Answer others’ questions - if you know the answer ;-)
    - Learn from others’ questions and answers
    - Check it Often

# Course Information

- Prerequisites:
  - Required: CSCE 350: Data Structures and Algorithms
  - There will be math (and programming)
  - Prior computer programming experience is required. Additional background in data structures and algorithms, linear algebra, and probability will all be helpful.
  - You should be prepared to review basic probability on your own if it is not fresh in your head.

# An awesome colleague asked this earlier last week!

We are trying to put together a MS in Biostatistics and Data Science and were interested in adding your CSCE 580 class as one of the core courses. We were able to look through your github site – which was extremely helpful – and think the class would be perfect.

On the CS website it lists CSCE 350 as a prerequisite for CSCE 580, and we were curious what the main concepts from CSCE 350 that are needed for CSCE 580. I'm copying the topics from 350 below:

1. Structured programming, stacks, queues, lists (3 hours)
2. Determining the Running Time of Programs, Order of Magnitude Analysis (6 hours)
3. Brute force (3 hours)
4. Divide-and-Conquer (4 hours)
5. Dynamic Programming (6 hours)
6. Transform-and-Conquer (4 hours)
7. The Greedy Technique (3 hours)
8. Decrease-and-Conquer (3 hours)
9. Graphs (3 hours)
10. Reviews and exams (4 hours)

# My Answer!

Regarding your question about the CSCE 350 prerequisite, the primary concepts from CSCE 350 that are leveraged in CSCE 580 include:

1. **Order of Magnitude Analysis (Big-O Notation):** Students need a basic understanding of how to evaluate algorithm efficiency and scalability, especially when comparing different approaches in AI and data science.
2. **Graph Algorithms:** Foundational concepts like graph traversal (e.g., BFS and DFS) are used when exploring AI algorithms that interact with structured data or state spaces.
3. **Dynamic Programming and Greedy Algorithms:** These techniques are helpful for understanding optimization problems in AI, though I typically provide a high-level review of these concepts in the course.
4. **Basic Data Structures (Stacks, Queues, Lists):** Familiarity with these is important for understanding how algorithms operate under the hood, but I often provide implementation support or refreshers as needed.

For students without CSCE 350, I think a brief primer on Big-O notation, basic graph concepts, and algorithmic problem-solving would suffice. I’d be happy to work with you to identify resources or strategies to help bridge the gap for those students if needed.

# Course Information

- 5 programming projects: Python, groups of 1 or 2
  - 7 late days for the entire semester (maximum 2 for a given project)
- ~5-10 homework assignments:
  - Written, solve together, write up alone, electronic submission through dropbox
- One midterms, one final
- Participation can help on margins
- Fixed scale
- Academic integrity policy

# Exam Dates

- **Midterm**: March 20, Thursday, 4:25 – 5:40 p.m. (In Class)
- **Final**: Tuesday, May 6 - 4:00 – 6:00 p.m. (In Class)

- There will be no alternative exams, put them on your calendar.
- More logistics closer to the exam

## Final Exam Schedule Spring 2025

### April 30 – May 7, 2025

- The final exam schedule is for standard meeting time classes.
- 2.5 hours is the maximum amount of time that can be allotted for a final exam. Instructors may choose to allot a shorter time.

### Information for Students

Is the class you're taking not listed? Contact your instructor.

### Information for Instructors

Is the class you're teaching not listed? Email Classroom Scheduling.

Deviations from the published schedule are not permitted without prior approval. Please refer to the Faculty Manual for further guidance.

In any course or laboratory that meets weekly, no quiz, test, or examination shall be given during the last class session before the regular examination period. In any course or laboratory that meets two or three times a week, no quiz, test, or examination shall be given during the last two class sessions before the regular examination period. In any course or laboratory that meets more than three times a week, no quiz, test, or examination shall be given during the last three sessions before the regular examination period (per the Faculty Manual, USC - Columbia, page 69).

Expand all

Monday/Wednesday/Friday Meeting Times

Monday/Wednesday Meeting Times

Tuesday/Thursday Meeting Times

| Class Meeting Day and Time | Exam Date, Day and Time |
|---|---|
| TR - 8:30 a.m. | Thursday, May 1 - 9:00 a.m. |
| TR - 10:05 a.m. | Tuesday, May 6 - 9:00 a.m. |
| TR - 11:40 a.m. | Thursday, May 1 - 12:30 p.m. |
| TR - 1:15 p.m. | Tuesday, May 6 - 12:30 p.m. |
| TR - 2:50 p.m. | Thursday, May 1 - 4:00 p.m. |
| TR - 4:25 p.m. | Tuesday, May 6 - 4:00 p.m. |
| TR - 6:00 p.m. | Thursday, May 1 - 7:30 p.m. |
| TR - 7:35 p.m. | Saturday, May 3 - 12:30 p.m. |

# Office Hours

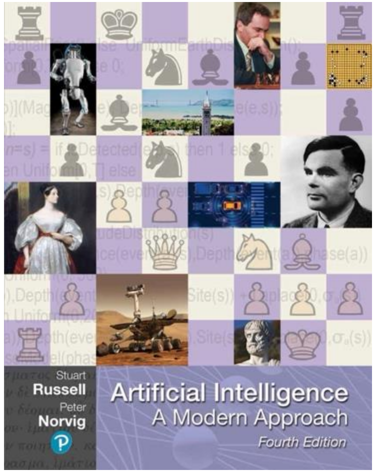
- Join in-person or remotely to talk to staff about content, ask questions on assignments, or raise any concerns you have
- **Office hours: TR 5:40 - 6 pm**
- A queue will become available on website soon
- Office hours start next week (Jan 21)

# Discussion Section

- There will be a few discussion sections
- Topic: review / warm-up exercises
- Will be announced via Piazza

# Textbook

- Not required, but for students who want to read more we recommend
  - Russell & Norvig, AI: A Modern Approach, 4th Ed.
    https://aima.cs.berkeley.edu/




  - Warning: Not a course textbook, so our presentation does not necessarily follow the presentation in the book.

# Laptops in Lecture

- Laptops can easily distract students behind you
- Please consider sitting towards the back if using your laptop in lecture

# Course Information: Assignments

- There will be programming and written assignments

- You will have a total of 7 late days for these assignments, *up to two of which can be used for each assignment*.
  - To allow you the flexibility to manage unexpected issues
  - Additional late days will not be granted except under truly exceptional circumstances
  - If you've used up all your late days, you lose 20% per day (see details on the course website)

# Course Information

- Programming Language: Course programming projects will be in Python.

- P0 is designed to teach you the basics of Python (Due: Jan 20)

# Project 0

- Due on Monday 20, 11:59 pm
- To be done alone
- Details on the course website, announcements on Dropbox and Piazza
- Submission via Dropbox

- Python 3.6
- Autograder: We have provided a local autograder and a set of test cases for you to evaluate your code. The local autograder is a file called autograder.py.

# Project 0

- **What to submit:**
  - The files that are required in the project's description (addition.py, buyLotsOfFruit.py, and shopSmart.py). Please use comments appropriately across your code.
  - A short README.txt file that specifies:
    - Your name and ID.
    - A brief description (i.e. a short paragraph) that includes the main ideas of your implementation.

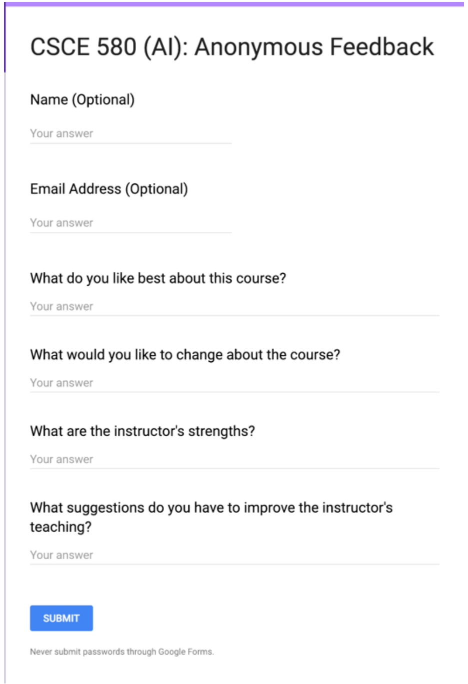
- **Place your files in a single folder inside the archive.**
  **Submit your assignment on Dropbox as a single archive file (.zip), with the name csce580-p0-lastname-uscid**

# Homework 0

- Due on Monday 20, 11:59 pm
- To be done alone
- Details on the course website, announcements on Dropbox and Piazza
- Submission via Dropbox
- To assess whether you are mathematically prepared for the second half of the course.

# Course Information: Feedback

- Please give feedback (positive or negative) as often as and as early as you can.



CSCE 580 (AI): Anonymous Feedback

Name (Optional)

Your answer

Email Address (Optional)

Your answer

What do you like best about this course?

Your answer

What would you like to change about the course?

Your answer

What are the instructor's strengths?

Your answer

What suggestions do you have to improve the instructor's teaching?

Your answer

SUBMIT

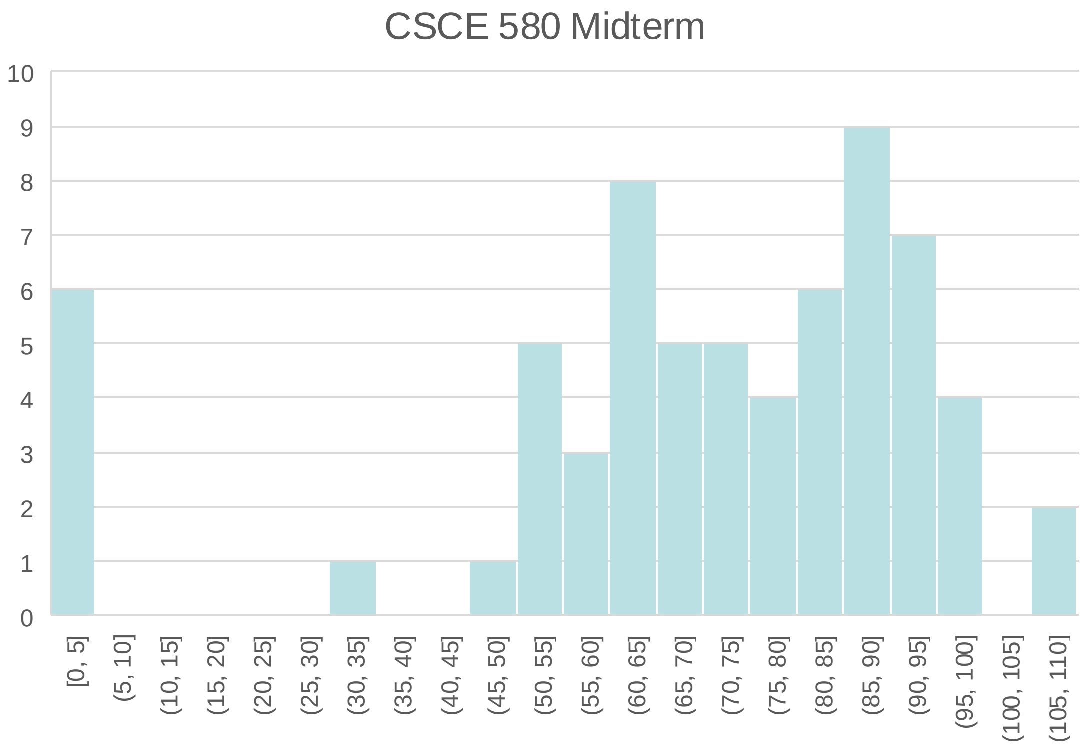
Never submit passwords through Google Forms.

A link will be send out on piazza

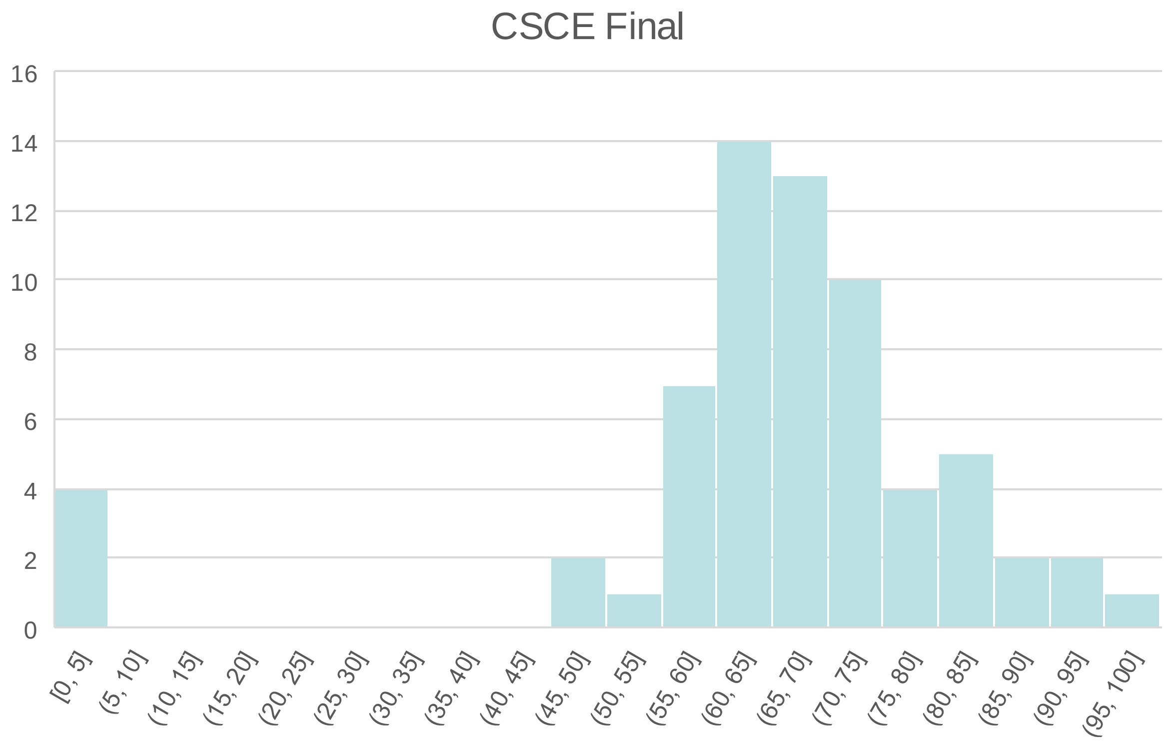
# Acknowledgements


- Dan Klein
- Pieter Abbeel
- Stuart Russell
- Brad Miller
- Nick Hay
- John DeNero
- Pooyan Fazli

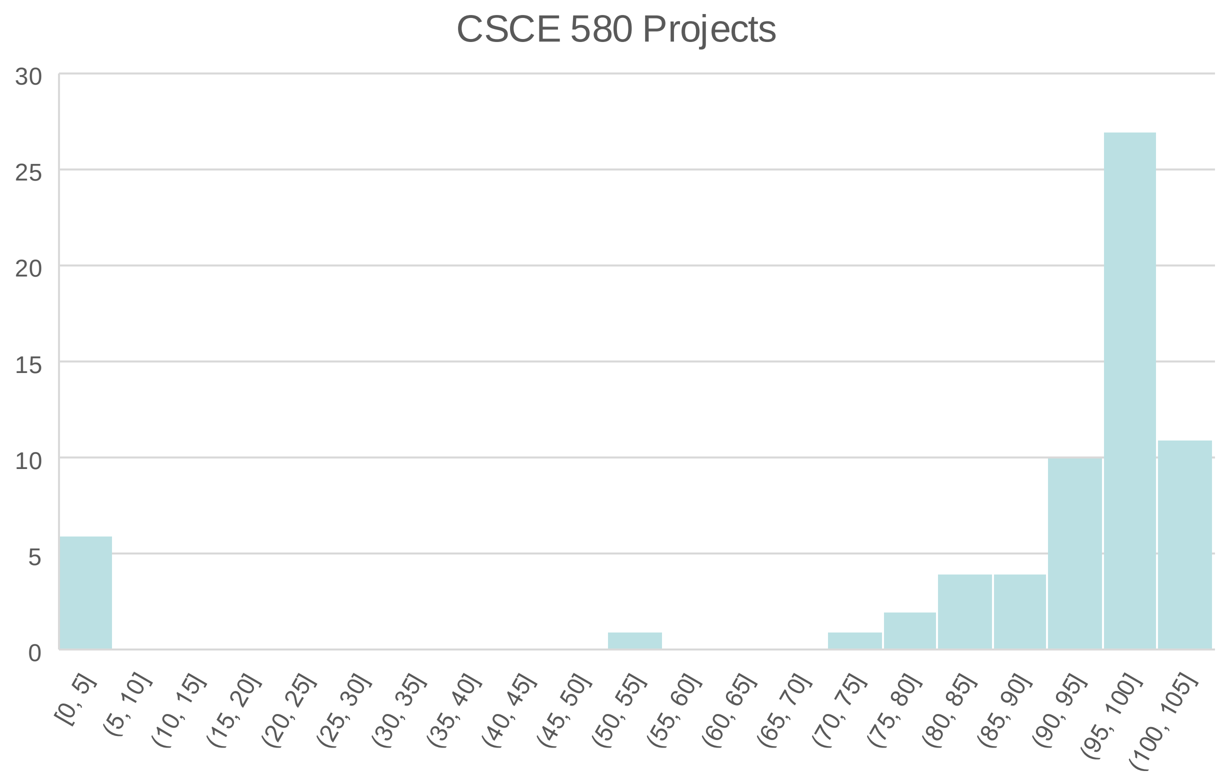
# Historical Statistics



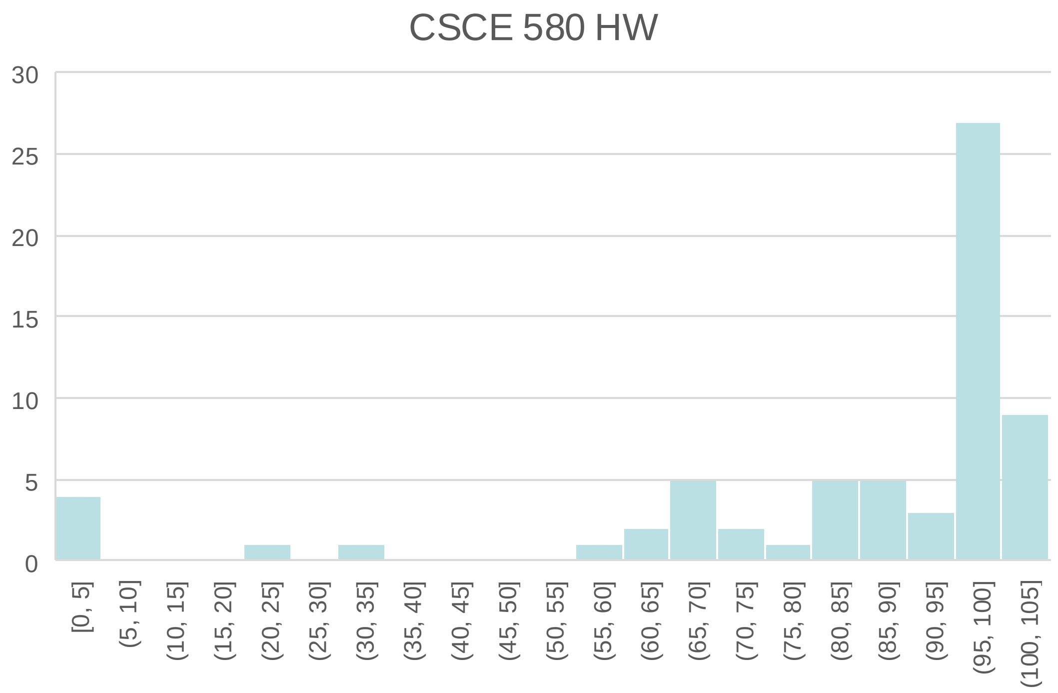
# Historical Statistics



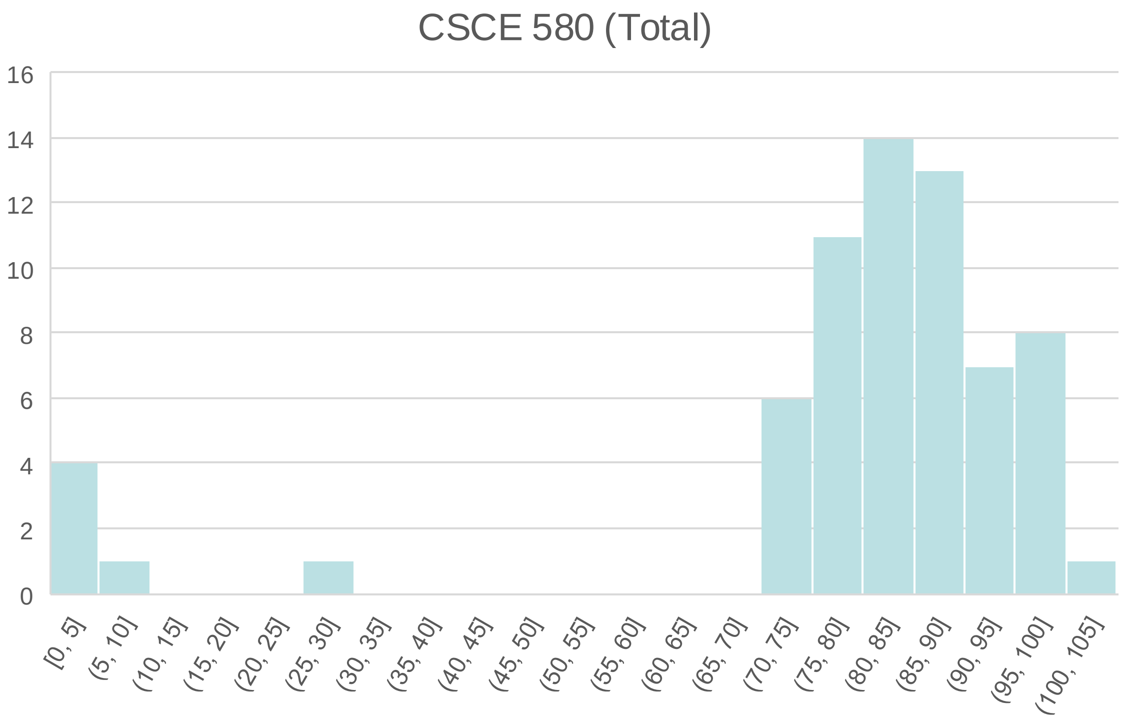
# Historical Statistics



# Historical Statistics



# Historical Statistics



# Important This Week

- Important this week:
  - **Checkout website:** https://pooyanjamshidi.github.io/csce580/
  - **You should be** on
    - **Piazza**---our main resource for discussion and communication
    - **Gradescope**---our main resource for homework
    - If you are not on either, let us know right away.
  - **HW0: Math diagnostic homework** is out (due on Tuesday 1/21 at 11:59 pm)
  - **P0: Python tutorial** is out (due on Tuesday 1/21 at 11:59 pm)
    - I fixed the issues. I am waiting for the TA to confirm all works perfectly.
    - For questions related to projects, please use Piazza.
    - If something is broken in the project, please use Piazza.
    - I appreciate it if you could include details of the issue so we can replicate the issue and fix it as soon as possible.
    - I will create a template for these cases soon.
  - **Office Hours** start next week, and you can catch me after the lecture

# Inclusion Statement

Fostering a Respectful and Welcoming Learning Environment

# Support and Understanding

- If you feel your academic performance has been negatively impacted due to personal circumstances such as any events or family matters, please reach out to me if I can help.
- My role here extends beyond teaching.

# Community Responsibility

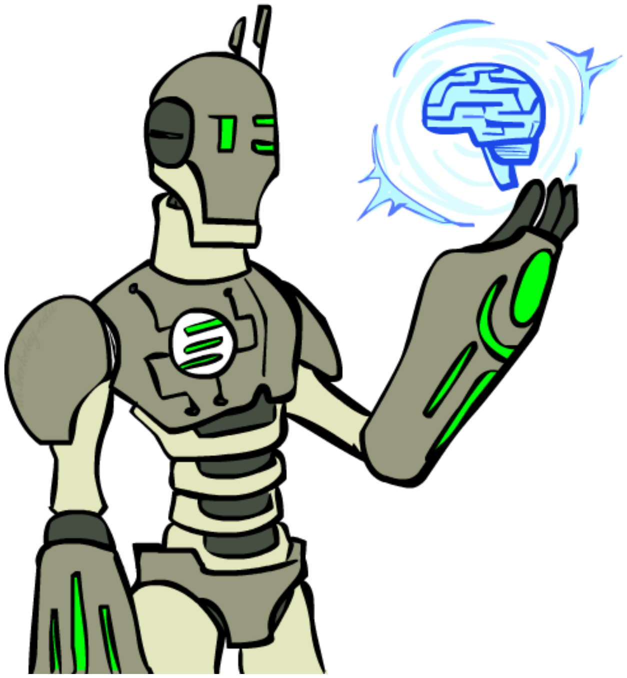
- Each member of this class plays a vital role in creating an inclusive and respectful learning environment.
- Be mindful of how your actions and words impact others and strive to make everyone feel valued and respected.

# Closing Statement

- We are committed to ensuring that everyone has an equitable opportunity to succeed and thrive in this course.
- Let us work together to create a positive and inclusive experience for all.

# Today

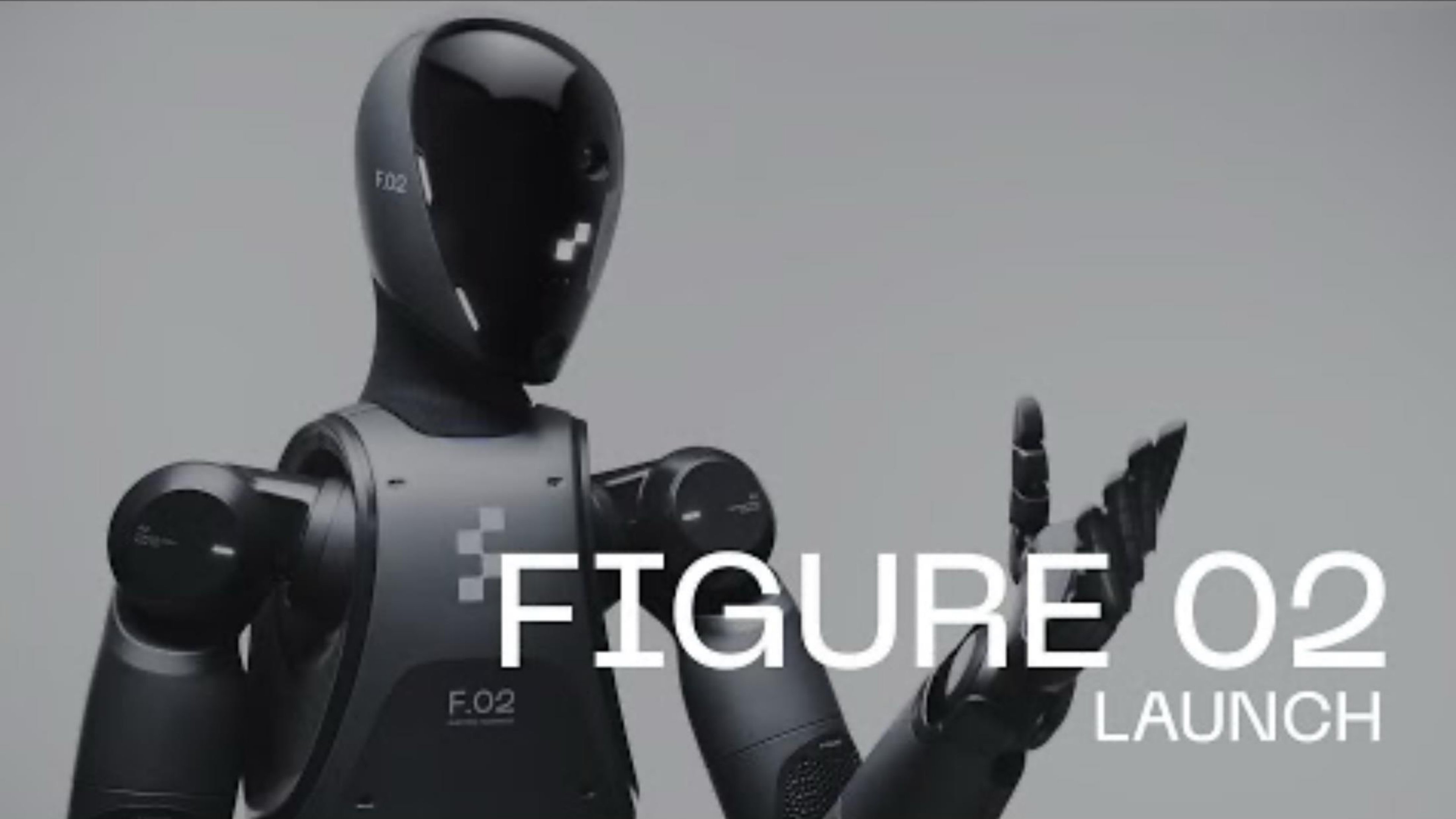
- What is artificial intelligence?
- What can AI do?
- What is this course?



# Understanding AI: Definitions and Historical Perspectives

How we define and approach AI across disciplines

- Let’s start by understanding where we are now in 2025!


F.02
FIGURE 02
LAUNCH
F.02

# What is Artificial Intelligence?

- AI is the study and creation of systems that can perform tasks requiring human-like intelligence.
- Definitions and approaches have evolved over time, informed by history and different disciplines.

# Think Like People (Cognitive Modeling Approach)

- **Goal:** Understand and replicate how humans think.
- **Fields Involved:** Cognitive science, neuroscience.
- **Why This Matters:**
  - This approach uses the human brain as an existence proof of intelligence and seeks to reverse-engineer it.
  - Examples include understanding human decision-making, memory, and problem-solving.
- **Challenges:** Humans and AI systems can arrive at similar outcomes (e.g., playing chess) but might use entirely different methods.

# Act Like People (Turing Test Approach)

- **Goal:** Create systems that behave indistinguishably from humans.
- **Origin:** Proposed by Alan Turing, who introduced the Turing Test.
- **Why This Matters:**
  - Early focus on human-like behavior, e.g., answering questions, holding conversations.
  - This approach emphasized mimicry, including quirks like favorite movies or realistic response delays.
- **Challenges:** Mimicking humans does not necessarily lead to general intelligence or optimal behavior.

# Think Rationally (Logical Reasoning Approach)

- **Goal:** Encode rules for correct reasoning and turn them into computational systems.
- **Origins:** Ancient philosophy, including works by Plato and Aristotle.
- **Why This Matters:**
  - Early AI relied on logic-based systems to emulate rational thought (e.g., theorem provers).
- **Challenges:** Difficulties in scaling, as human reasoning isn't always formalizable, and it's often more critical to focus on results than on the process.

# Act Rationally (Modern Approach)

- **Goal:** Build systems that act rationally to achieve optimal outcomes.
- **Fields Involved:** Statistics, optimization, data science.
- **Why This Matters:**
  - This approach focuses on creating systems that make decisions to maximize success (e.g., self-driving cars, recommendation systems).
  - Modern AI combines data, algorithms, and optimization to solve real-world problems effectively.

# Current Perspective

- The **dominant approach today** is building systems that *act rationally*.
- These systems aim to make optimal decisions based on data, leveraging tools like **machine learning**, **neural networks**, and **optimization** techniques.
- By focusing on **rational actions** rather than mimicking humans, AI can solve complex problems efficiently.

# Key Takeaway

- AI definitions vary based on:
  - Thinking vs. acting.
  - Mimicking humans vs. acting optimally.

- Modern AI focuses on building systems that act rationally, solving complex problems effectively.

# What is AI?

The science of making machines that:

# What Do We Mean by Rationality in AI?

1. **Rationality is about achieving goals optimally:**

   A system is rational if it takes actions that are expected to achieve its pre-defined goals in the best possible way.

2. **Rationality focuses on decisions, not thought processes:**

   It doesn’t matter how the system arrived at a decision (e.g., through logic, trial and error, or intuition). What matters is the quality of the outcomes.

# What Do We Mean by Rationality in AI?

3. **Utility as a measure of success:**

   Goals are translated into *utility functions,* which assign numerical values to outcomes. The higher the utility, the better the outcome for the system.

4. **Expected utility and decision-making:**

   Rational behavior means choosing actions that maximize the *expected utility*, accounting for uncertainties in the environment.

# Example to Illustrate Rationality

Imagine a robot vacuum cleaner with the goal of cleaning a room:

- **Pre-defined goal:** Minimize dust in the room.
- **Utility function:** Assigns higher utility to states where the room is cleaner.
- **Rational decision-making:** The vacuum chooses a path that it calculates will result in the cleanest room (highest utility), even if there’s a small chance of encountering obstacles.

# Key Takeaway

- In AI, rationality isn’t about thinking like a human or being logical—it’s about achieving the best outcomes given a system's goals and available information.
- This focus on results is why utility functions and optimization are central to AI design.

# Rational Decisions

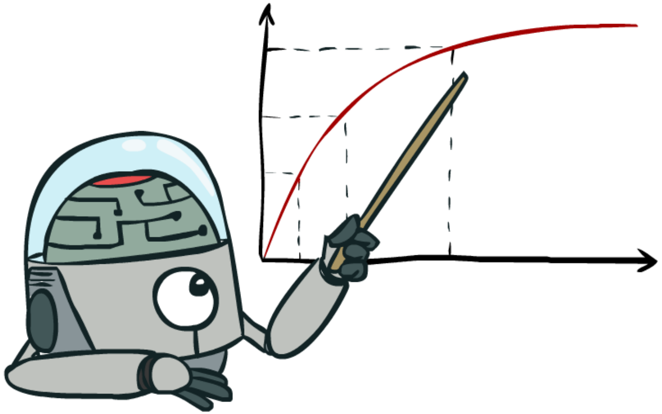
We'll use the term **rational** in a very specific, technical way:

- Rational: maximally achieving pre-defined goals
- Rationality only concerns what decisions are made (not the thought process behind them)
- Goals are expressed in terms of the **utility** of outcomes
- Being rational means **maximizing your expected utility**

A better title for this course would be:

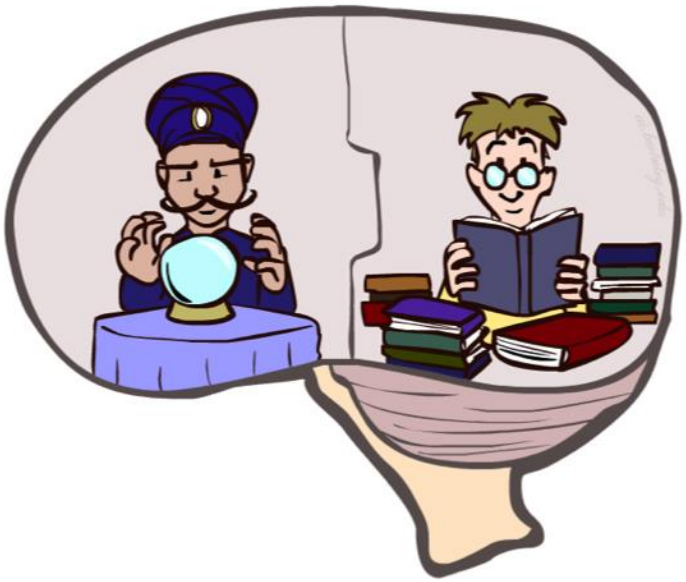
**Computational Rationality**

# Maximize Your Expected Utility



# What About the Brain?

- Brains (human minds) are very good at making rational decisions, but not perfect
- Brains aren't as modular as software, so hard to reverse engineer!
- "Brains are to intelligence as wings are to flight"
- Lessons learned from the brain: memory and simulation are key to decision-making



# Lessons from AI: Existence Proof and Decision-Making

Insights on why reverse engineering AI isn't enough

# Existence Proof of AI

- We have systems that demonstrate intelligent behavior.
- Why don’t we reverse engineer these systems?
- The complexity of AI systems requires us to focus on broader principles instead of direct replication.

# Key Lessons Learned

Effective decision-making relies on two key components:

1. **Memory:** Storing and learning from past experiences (e.g., avoiding past mistakes).
2. **Simulation:** Predicting outcomes by unrolling the consequences of actions based on a model of the world.

# Model and Learning

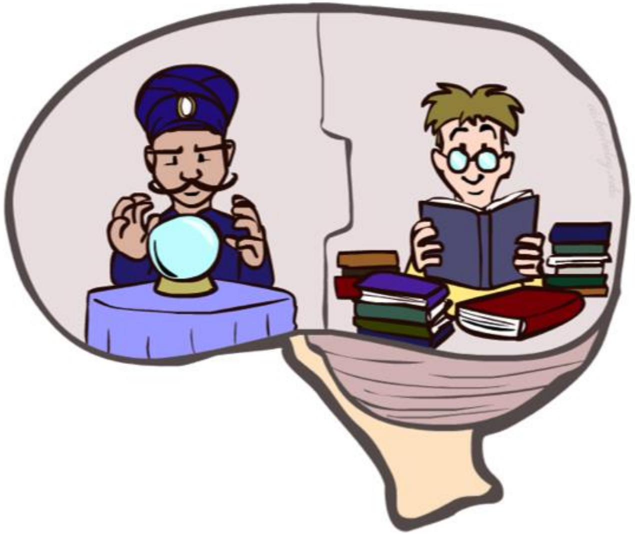
- **Models** are derived from data and experiences.
- **Learning** involves interleaving memory and simulation to improve decision-making.
- The combination of these components helps AI systems make **informed** and **adaptive** choices.

# Course Topics

Overview of Key Areas in Artificial Intelligence

# Course Topics

- Part I: Making Decisions
  - Fast search / planning
  - Constraint satisfaction
  - Adversarial and uncertain search
- Part II: Reasoning under Uncertainty
  - Bayes' nets
  - Decision theory
  - Hidden Markov Models
- Part III: Machine learning
  - Naïve Bayes
  - Perceptrons and Logistic Regression
  - Neural Networks
  - Decision Trees and Support Vector Machines



# Part I: Making Decisions

- **Fast search/planning:** Techniques for navigating large state spaces efficiently.
- **Constraint satisfaction:** Solving problems with constraints (e.g., Sudoku).
- **Adversarial and uncertain search:** Methods for decision-making in competitive and uncertain environments.

# Part II: Reasoning Under Uncertainty

- **Bayes' Nets:** Probabilistic models for representing and reasoning about uncertain knowledge.
- **Decision Theory:** Framework for making optimal decisions under uncertainty.
- **Hidden Markov Models (HMMs):** Modeling systems that evolve over time with hidden states.

# Part III: Machine Learning

- **Naïve Bayes:** A simple probabilistic classifier based on Bayes' theorem.
- **Perceptrons and Logistic Regression:** Basic linear models for classification.
- **Neural Networks:** Deep learning models inspired by biological neurons.
- **Decision Trees and Support Vector Machines:** Popular methods for classification and regression.

# Recovery Strategy for CSCE 580 AI Course

- I appreciate your patience as we navigate some delays due to unforeseen circumstances.
- Recovery Strategy for CSCE 580 AI Course (Weeks 4+ Delay)
- HW0 & P0 Grades Released
  - We have **just published grades for HW0 and P0**.
  - Please **check your grades** and if you have any concerns, **send a private Piazza message** to me and the TA.
  - We will review and **regrade if necessary**—don't hesitate to reach out!

# 🚀 Get Involved: GAMA Club

# GAMA Club Update

Games, AI, and Machine Learning Association

# What is GAMA Club?

- - A student-driven club focused on **AI, games, and machine learning**.
- - Explore **AI-driven decision-making, reinforcement learning, and gaming applications**.
- - Open to students interested in **AI, game development, and computational strategies**.

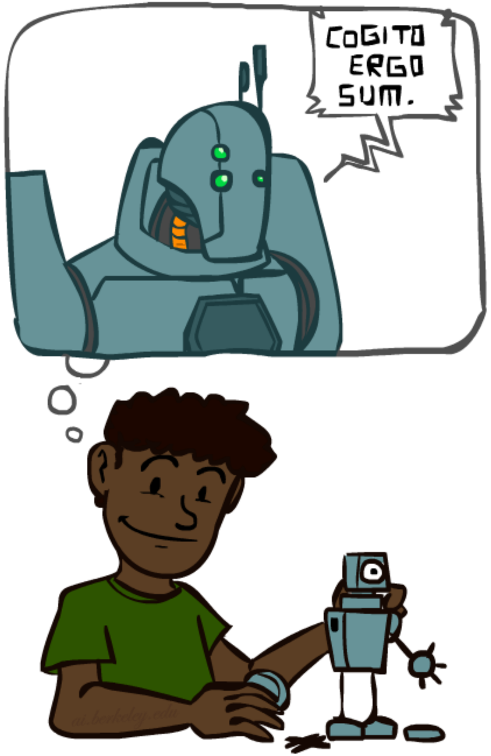
# Why Join GAMA Club?

- - **Learn cutting-edge AI applications** in game design and reinforcement learning.
- - **Collaborate on projects** that bridge AI and gaming.
- - **Attend workshops, talks, and competitions** related to AI and interactive systems.
- - **Network with peers and experts** in AI and machine learning.

# How to Get Involved?

- - **Join upcoming meetings** (details will be shared soon).
- - Follow **announcements on Piazza and departmental emails**.
- - **Express your interest** by reaching out to club organizers or faculty sponsors.
- - Explore collaboration opportunities on **student projects and competitions**.

# A (Short) History of AI




Demo: HISTORY – MT1950.wmv

# The Evolution of Computers

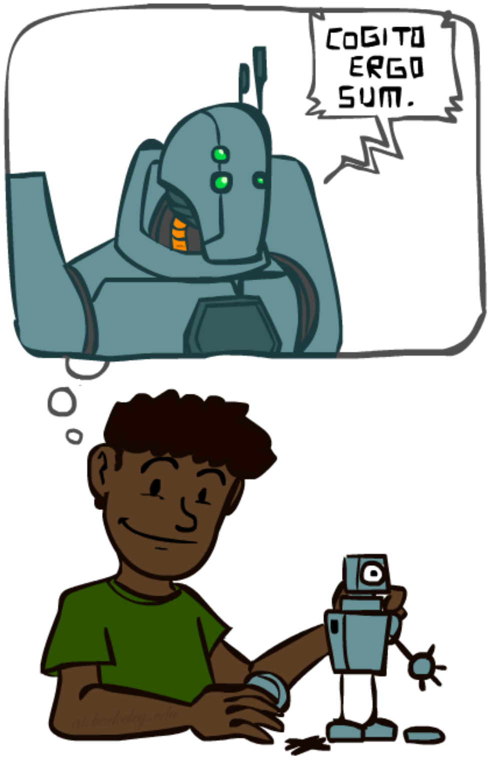
From Big Calculators to Intelligent Machines

# Big Calculators: 1950s Era

- Early computers were primarily used as large calculators.
- **Focus**: Performing numerical calculations efficiently.
- These systems paved the way for broader uses, including:
  - Calculations beyond numbers (e.g., data processing).
  - Exploring the idea of making computers 'think.'

# A (Short) History of AI

- 1940-1950: Early days
  - 1943: McCulloch & Pitts: Boolean circuit model of brain
  - 1950: Turing's “Computing Machinery and Intelligence”
- 1950—70: Excitement: Look, Ma, no hands!
  - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
  - 1956: Dartmouth meeting: “Artificial Intelligence” adopted
  - 1965: Robinson's complete algorithm for logical reasoning
- 1970—90: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms
  - 1988—93: Expert systems industry busts: “AI Winter”
- 1990—: Statistical approaches
  - Resurgence of probability, focus on uncertainty
  - General increase in technical depth
  - Agents and learning systems… “AI Spring”?
- 2012—: Where are we now?
  - Big data, big compute, neural networks
  - Some re-unification of sub-fields
  - AI used in many industries



# A (Short) History of AI: Updated for 2025

Tracing the evolution and advancements in Artificial Intelligence

# 2012-2020: Neural Networks Revolution

- Big data, large-scale neural networks, and GPUs.
- Significant progress in computer vision, NLP, and reinforcement learning.
- AI adoption in industries: healthcare, finance, and autonomous vehicles.

# 2020-2025: AI Transformation Era

- **Foundation Models:** Scaling large language models (e.g., GPT-4, ChatGPT, and others).
- **Multi-Modal AI:** Advances in systems that process text, images, and speech simultaneously.
- **Explainability:** Focus on ethical AI, fairness, and interpretable machine learning.
- **AI in Real-Time Applications:** Enhanced decision-making in healthcare, climate, and robotics.
- **Generative AI:** Boom in creativity with AI-generated art, music, and design.

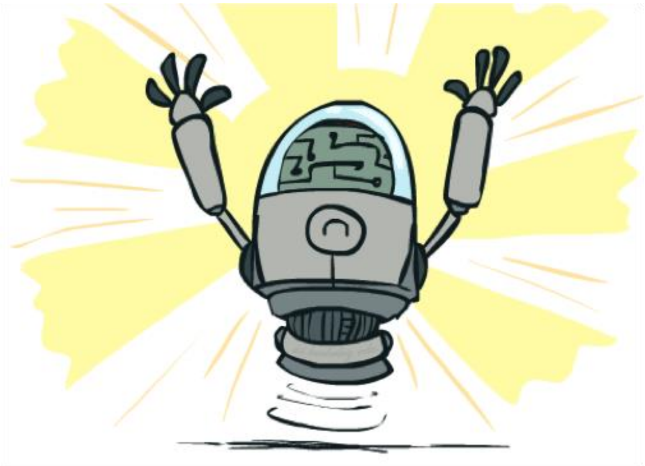
# What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Play a decent game of table tennis?
- ✓ Play a decent game of Jeopardy?
- ✓ Drive safely along a curving mountain road?

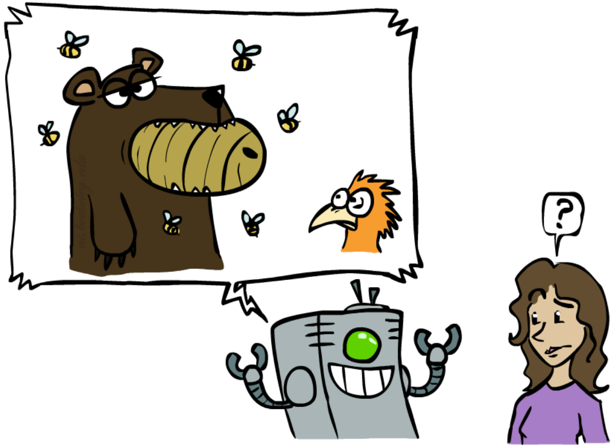
- ✓ Buy a week's worth of groceries on the web?

- ? Discover and prove a new mathematical theorem?
- ✗ Converse successfully with another person for an hour?
- ? Perform a surgical operation?
- ✓ Put away the dishes and fold the laundry?
- ✓ Translate spoken Chinese into spoken English in real time?
- ✗ Write an intentionally funny story?



# Unintentionally Funny Stories

- One day Joe Bear was hungry. He asked his friend Irving Bird where some honey was. Irving told him there was a beehive in the oak tree. Joe walked to the oak tree. He ate the beehive. The End.




- Henry Squirrel was thirsty. He walked over to the river bank where his good friend Bill Bird was sitting. Henry slipped and fell in the river. Gravity drowned. The End.

- Once upon a time there was a dishonest fox and a vain crow. One day the crow was sitting in his tree, holding a piece of cheese in his mouth. He noticed that he was holding the piece of cheese. He became hungry, and swallowed the cheese. The fox walked over to the crow. The End.

[Shank, Tale-Spin System, 1984]

# Recovery Strategy for CSCE 580 AI Course

- I appreciate your patience as we navigate some delays due to unforeseen circumstances.
- Recovery Strategy for CSCE 580 AI Course (Weeks 4+ Delay)
- HW0 & P0 Grades Released
  - We have **just published grades for HW0 and P0**.
  - Please **check your grades** and if you have any concerns, **send a private Piazza message** to me and the TA.
  - We will review and **regrade if necessary**—don't hesitate to reach out!

# Recovery Strategy for CSCE 580 AI Course

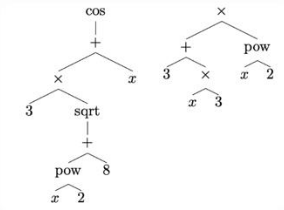
- We take attendance and it will be 10% of the final grade
- Read next lecture materials during the prior week
- Come to class on time so we can start on time

# AI in 2025: Notes

- **Advanced Communication:** AI models can hold meaningful conversations but still struggle with humor and nuance.
- **Healthcare Applications:** AI assists in surgeries and diagnostics under human supervision.
- **Physical Tasks:** Robotics has advanced significantly for household chores.
- **Mathematical Discoveries:** AI systems like DeepMind's AlphaTensor have proven new theorems.

- **Real-Time Translation:** Multi-modal AI excels in live language translation.
- **Creativity:** Generative AI continues to improve, but humor remains a complex challenge.

2021




★ Pretrained Language Models are Symbolic Mathematics Solvers too!

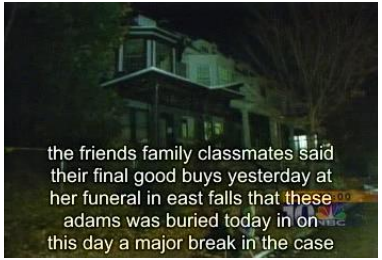
Kimia Noorbakhsh, Modar Sulaiman, Mahdi Sharifi, Kallol Roy, Pooyan Jamshidi

*Arxiv*

▶ Abstract

# Natural Language

- Speech technologies (e.g. Siri)
  - Automatic speech recognition (ASR)
  - Text-to-speech synthesis (TTS)
  - Dialog systems




Demo: NLP – ASR tvsample.avi

# Natural Language

- Speech technologies (e.g. Siri)
  - Automatic speech recognition (ASR)
  - Text-to-speech synthesis (TTS)
  - Dialog systems
- Language processing technologies
  - Question answering
  - Machine translation



**"Il est impossible aux journalistes de rentrer dans les régions tibétaines"**

Bruno Philip, correspondant du "Monde" en Chine, estime que les journalistes de l'AFP qui ont été expulsés de la province tibétaine du Qinghai "n'étaient pas dans l'illégalité".

**Les faits** Le dalaï-lama dénonce l'"enfer" imposé au Tibet depuis sa fuite, en 1959

**Vidéo** Anniversaire de la rébellion

**"It is impossible for journalists to enter Tibetan areas"**

Philip Bruno, correspondent for "World" in China, said that journalists of the AFP who have been deported from the Tibetan province of Qinghai "were not illegal."

**Facts** The Dalai Lama denounces the "hell" imposed since he fled Tibet in 1959

**Video** Anniversary of the Tibetan rebellion: China on guard

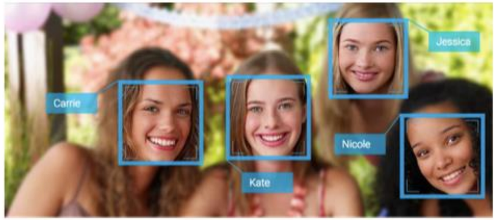
  - Web search
  - Text classification, spam filtering, etc…

# Vision (Perception)

PIXELS -> INFO/DECISION

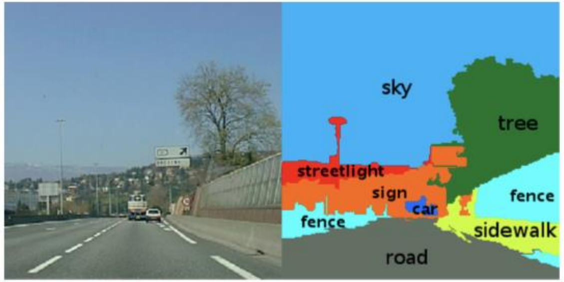
E.g.:

## Face detection and recognition



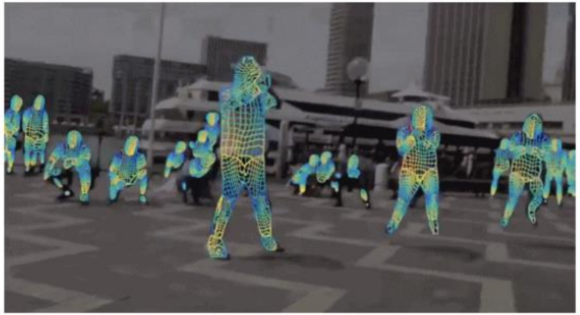

Source: TechCrunch

## Semantic Scene Segmentation




[Caesar et al, ECCV 2017]

## 3-D Understanding



[DensePose]

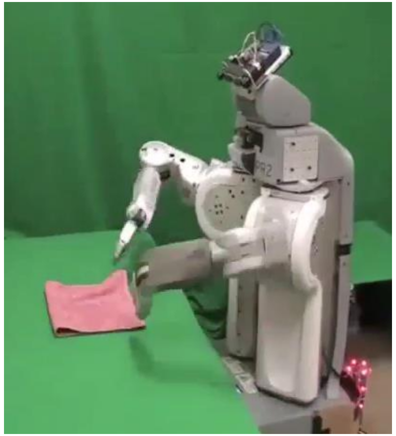

TRACKING:
N
NW
NE
W
E


car 1.000
LOUIS VUITTON

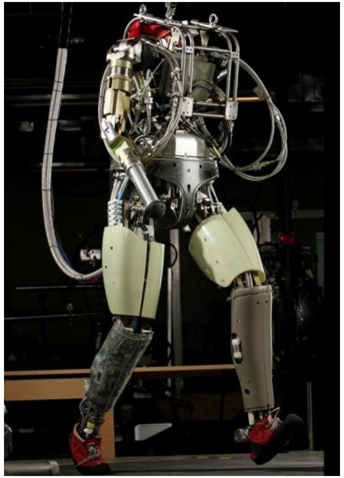
# Robotics

Demo 1: ROBOTICS – soccer.avi
Demo 2: ROBOTICS – soccer2.avi
Demo 3: ROBOTICS – gcar.avi
Demo 4: ROBOTICS – laundry.avi
Demo 5: ROBOTICS – petman.avi

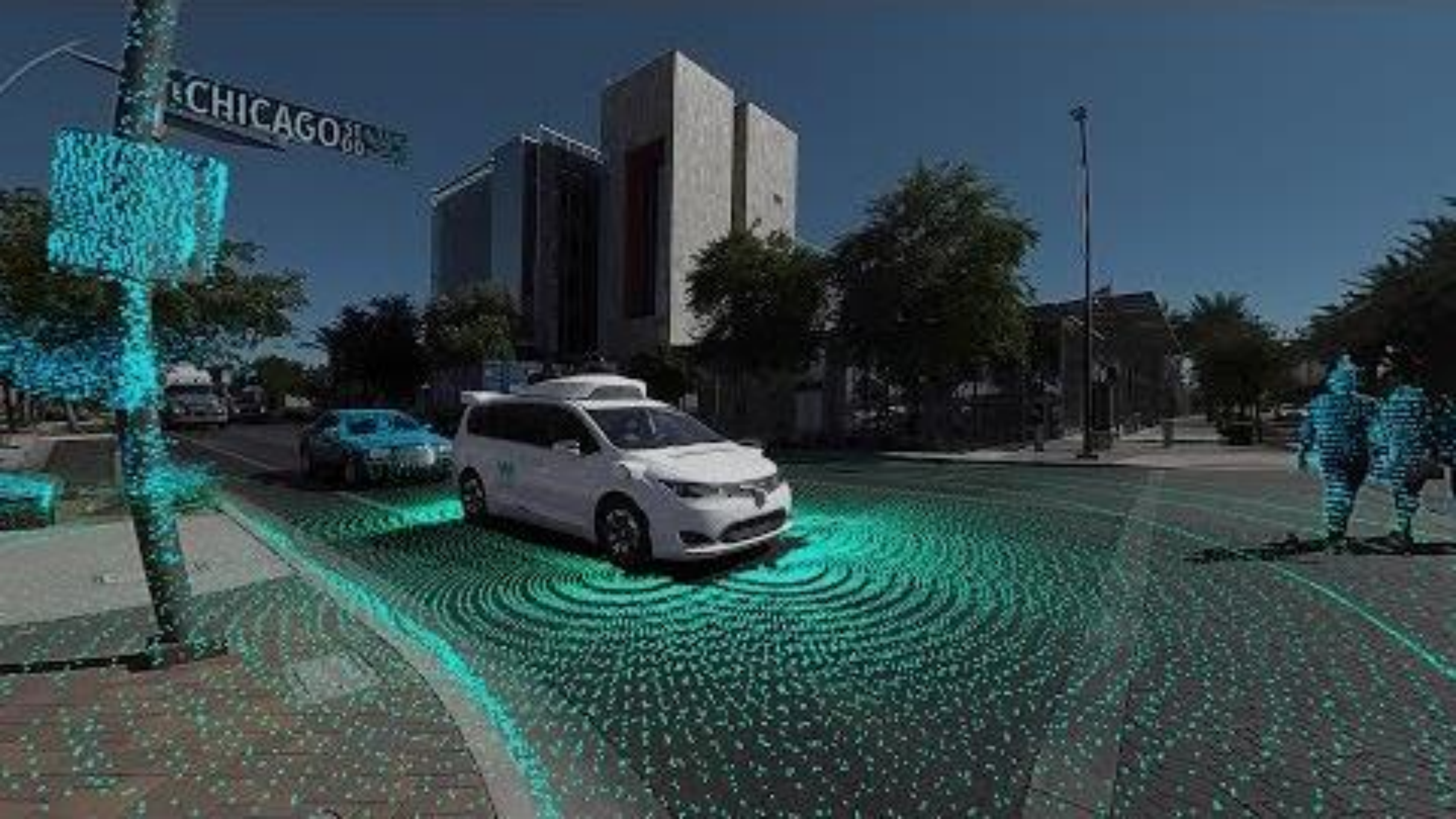
- Robotics
  - Part mech. eng.
  - Part AI
  - Reality much harder than simulations!
- Technologies
  - Vehicles
  - Rescue
  - Soccer!
  - Lots of automation…
- In this class:
  - We ignore mechanical aspects
  - Methods for planning
  - Methods for control





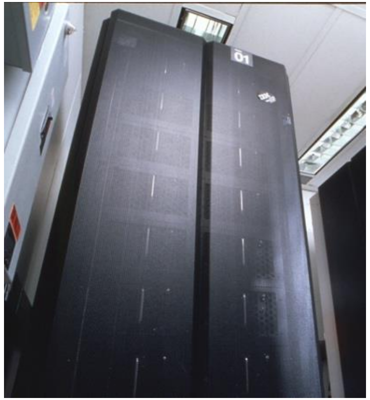


Images from UC Berkeley, Boston Dynamics, RoboCup, Google


CHICAGO

# Game Playing

- Classic Moment: May, '97: Deep Blue vs. Kasparov
  - First match won against world champion
  - “Intelligent creative” play
  - 200 million board positions per second
  - Humans understood 99.9 of Deep Blue's moves
  - Can do about the same now with a PC cluster




- Open question:
  - How does human cognition deal with the search space explosion of chess?
  - Or: how can humans compete with computers at all??
- 1996: Kasparov Beats Deep Blue
  “I could feel --- I could smell --- a new kind of intelligence across the table.”
- 1997: Deep Blue Beats Kasparov
  “Deep Blue hasn't proven anything.”
- 2016: Huge game-playing advances recently, e.g. in AlphaGo beats Lee Sedol – huge advantage: sparse rollout, ML (NN), and self-play
- 2019: OpenAI Five vs human pros

Text from Bart Selman, image from IBM’s Deep Blue pages

# Logic

- Logical systems
  - Theorem provers
  - NASA fault diagnosis
  - Question answering
- Methods:
  - Deduction systems
  - Constraint satisfaction
  - Satisfiability solvers (huge advances!)

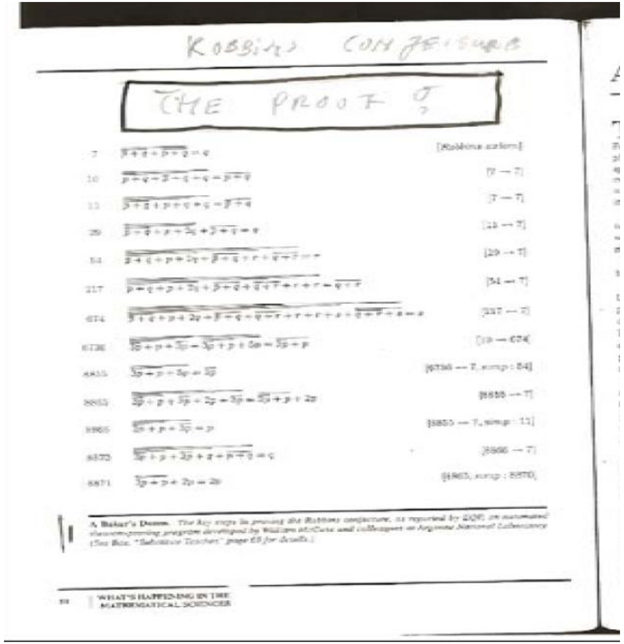
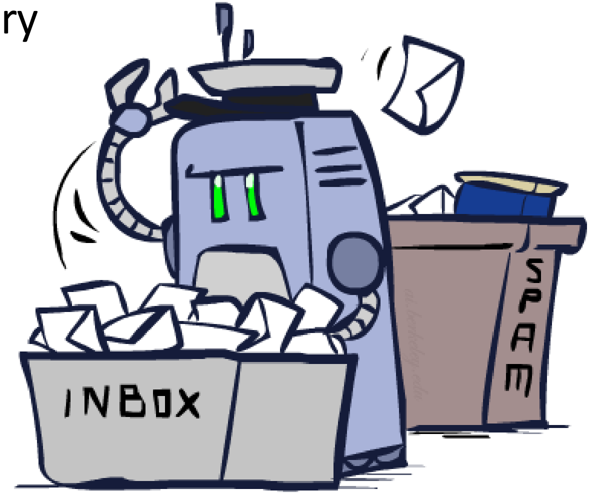


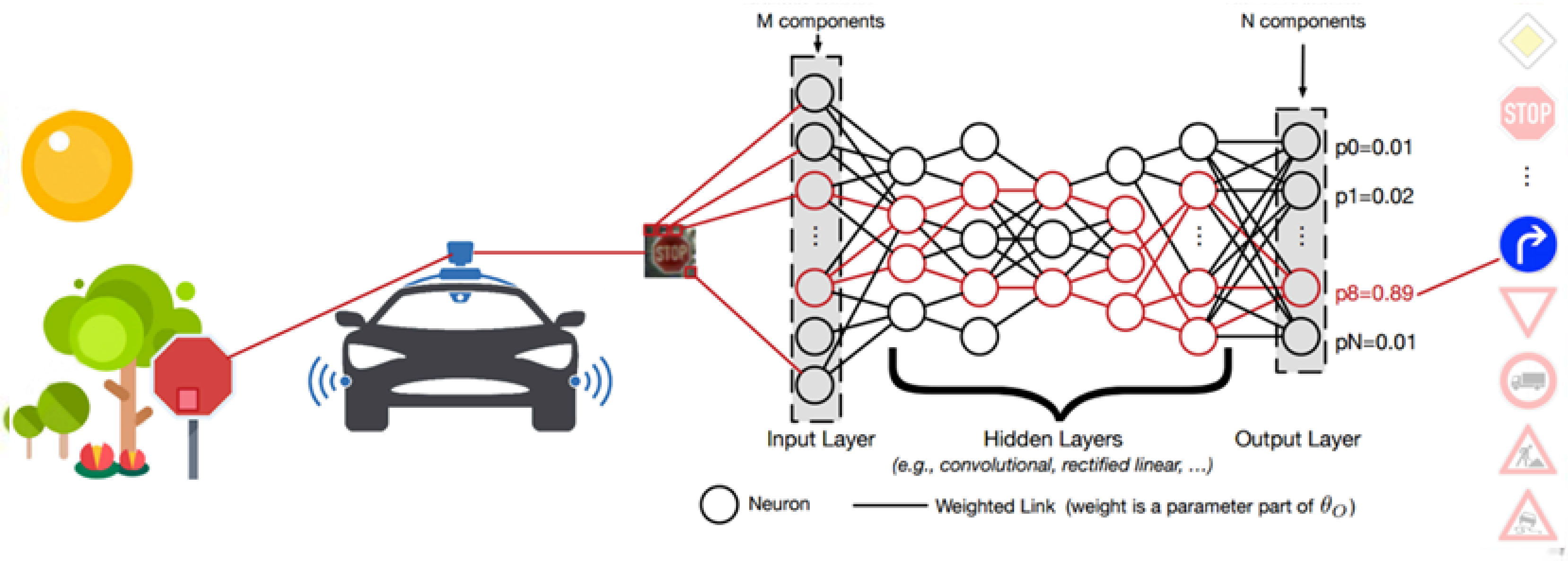
Image from Bart Selman

# AI is starting to be everywhere…

- 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
  - … Lots more!



# Importance of defending against adversarial attacks

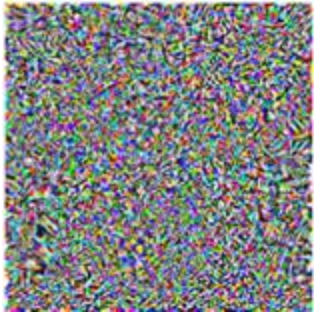


# Adversarial examples

pig

x

$+ \; 0.005 \times$



$\eta$

$=$

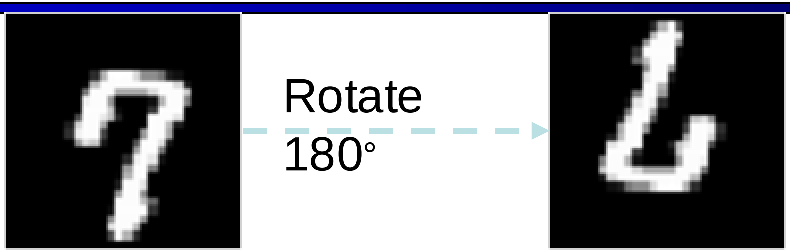
airliner

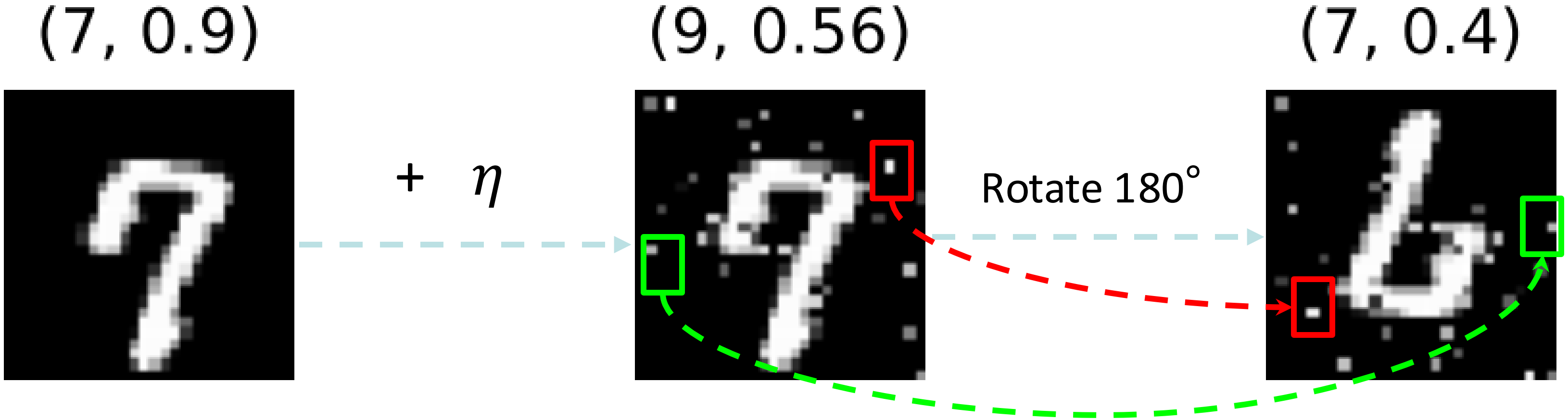
x’

FGSM

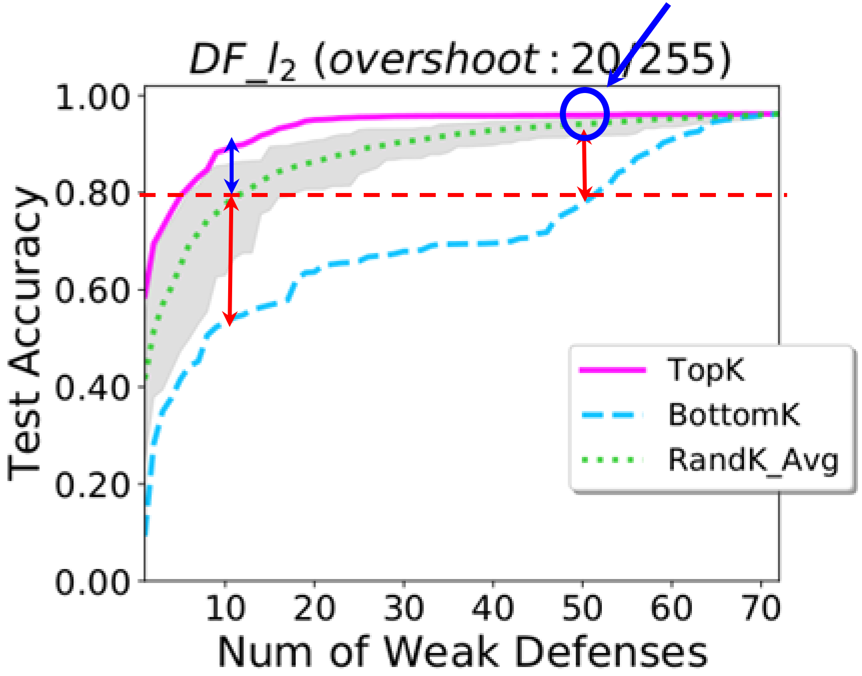
$\varepsilon \cdot \mathrm{sign}(\nabla_x J(x, y))$

# Transformations are Effective

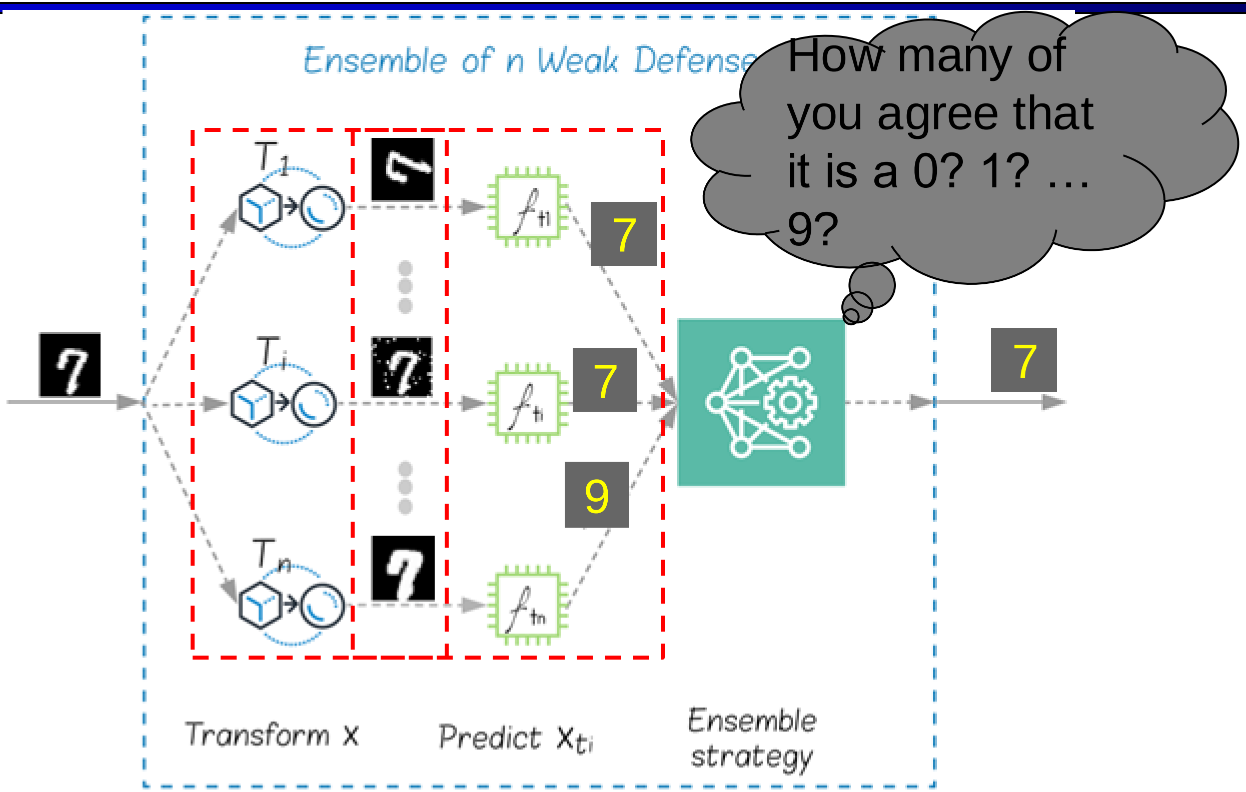


# Both Quality and Quantity Matter



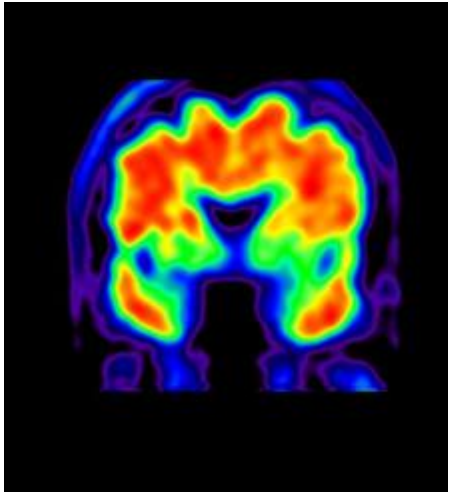
# Athena: Ensemble of Many Weak Defenses




- 72 transformations
- 5 ensembles

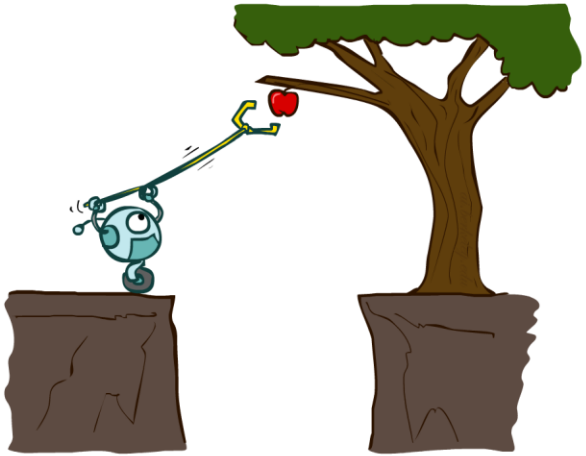
# Paper and Code

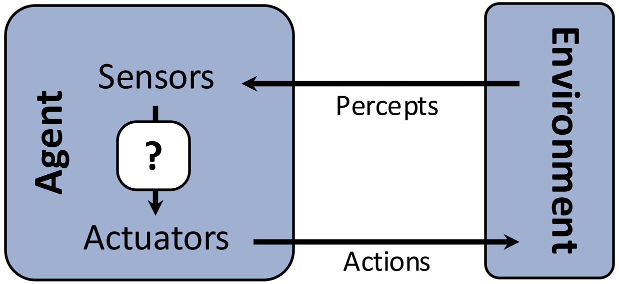
- **Paper**: https://arxiv.org/pdf/2001.00308.pdf
- **Code**: https://github.com/softsys4ai/athena



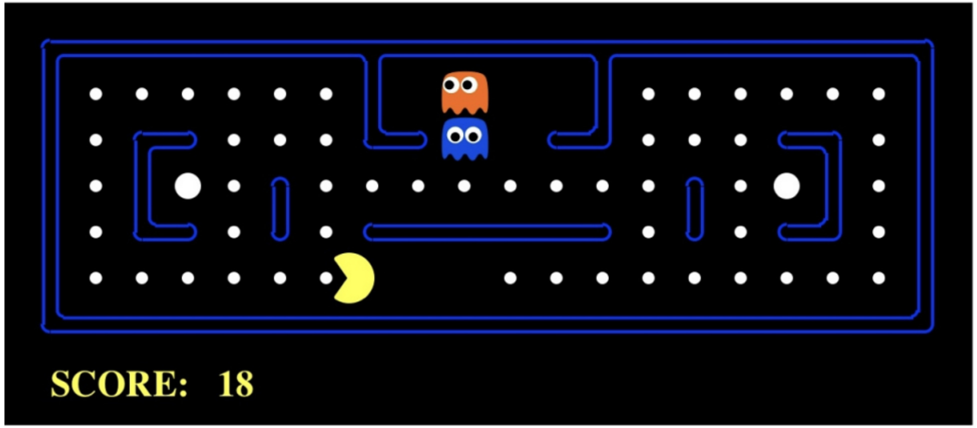

# Designing Rational Agents

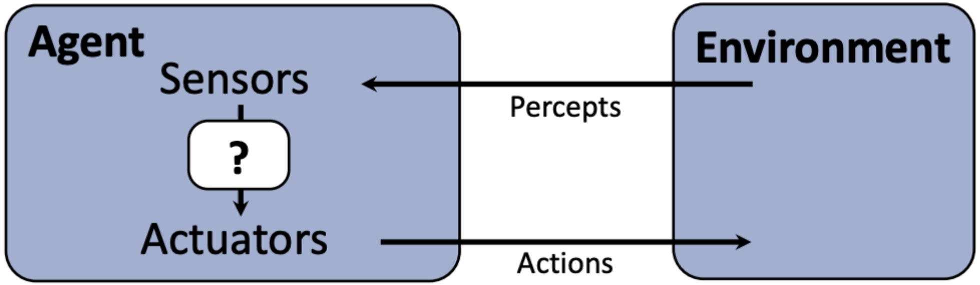
- An **agent** is an entity that *perceives* and *acts*.
- A **rational agent** selects actions that maximize its (expected) **utility**.
- Characteristics of the **percepts, environment,** and **action space** dictate techniques for selecting rational actions
- **This course** is about:
  - General AI techniques for a variety of problem types
  - Learning to recognize when and how a new problem can be solved with an existing technique





# Pac-Man as an Agent




Pac-Man is a registered trademark of Namco-Bandai Games, used here for educational purposes

SCORE: 1282

# Agents

- A Goal of AI: Build robust, fully autonomous agents in the real world

# Intelligent (Autonomous) Agents: Examples

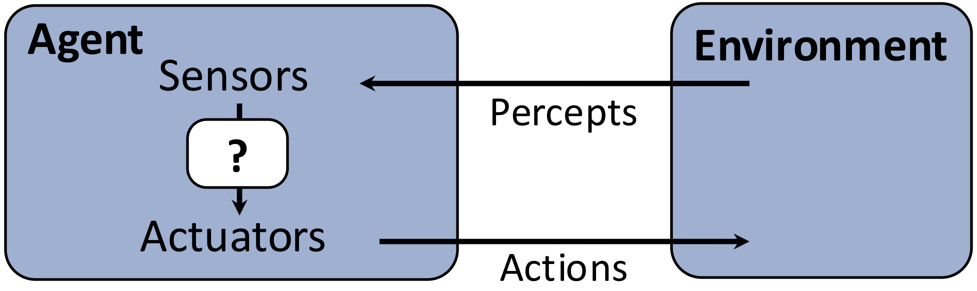
- Autonomous robot
- Information gathering agent
  - Find me the cheapest?
- E-commerce agents
  - Decides what to buy/sell and does it
- Air-traffic controller
- Meeting scheduler
- Computer-game-playing agent

# Not Intelligent Agents

- Thermostat
- Telephone
- Answering machine
- Pencil
- Java object

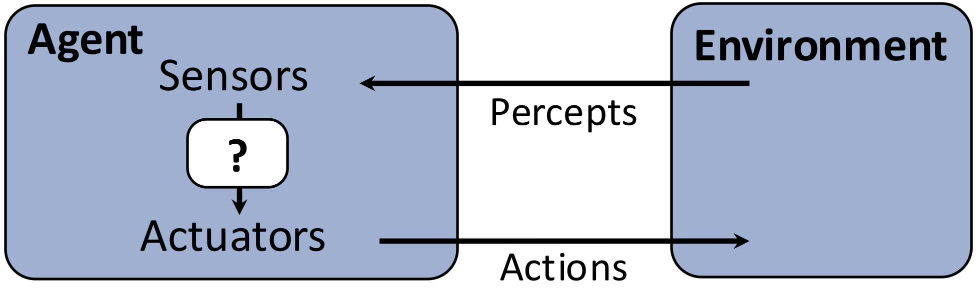
# What is an Agent?

# Agents and Environments



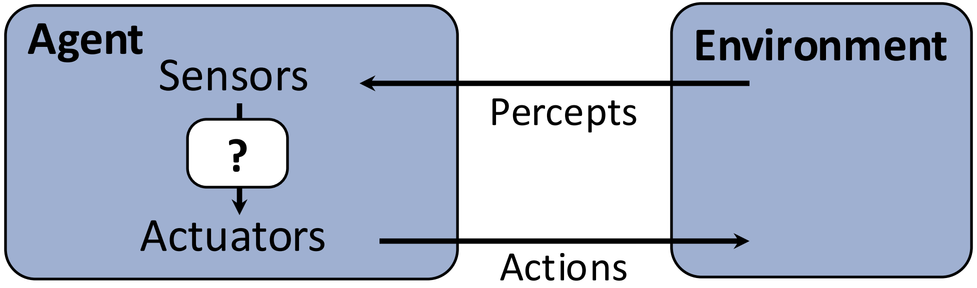

- An agent *perceives* its environment through *sensors* and *acts* upon it through *actuators*.

# Agents and Environments



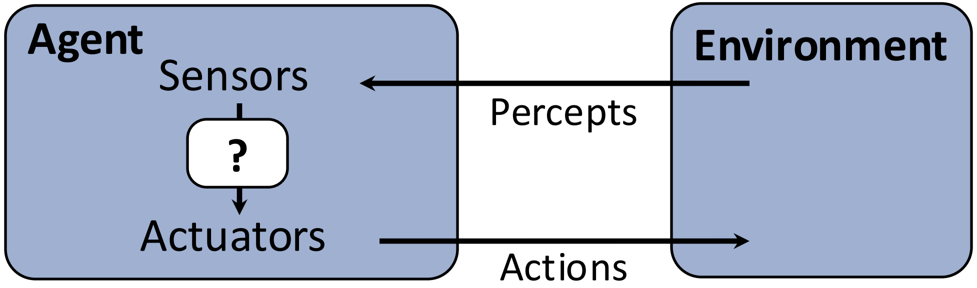

- Are humans agents?
- Yes!
  - Sensors = vision, audio, touch, smell, taste, …
  - Actuators = muscles, secretions, changing brain state

# Agents and Environments



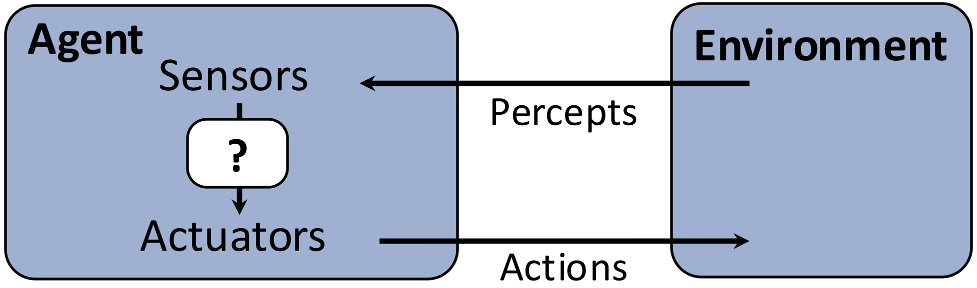

- Are Robots agents?
- Yes!
  - Sensors =  cameras, laser range finders, GPS
  - Actuators = various motors

# Agents and Environments



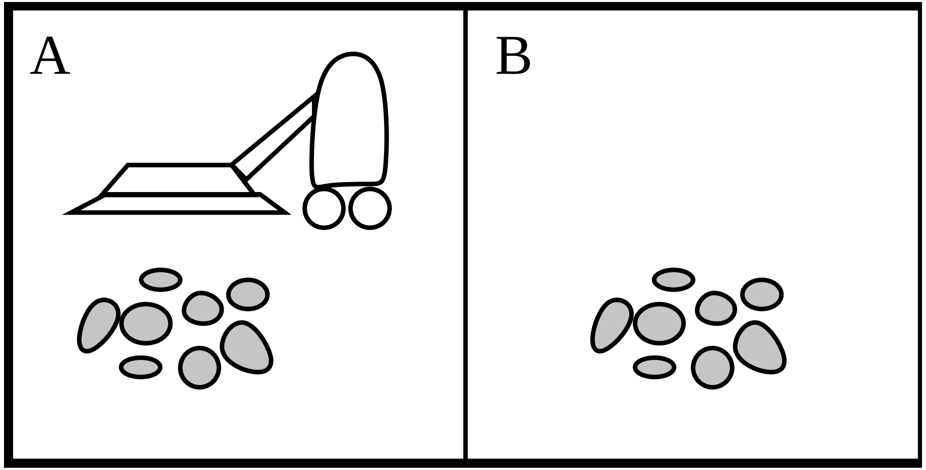

- Are pocket calculators agents?
- Yes!
  - Sensors = key state sensors
  - Actuators = digit display

# Agents and Environments




- AI is more interested in agents with substantial computation resources and environments requiring nontrivial decision making

# Example: Vacuum world

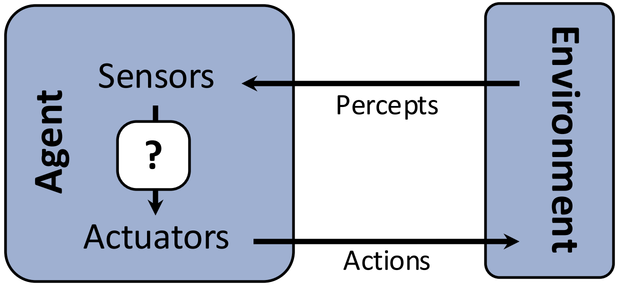



- Percepts: [location,status], e.g., [A,Dirty]
- Actions: Left, Right, Suck, NoOp

# Rational Agents

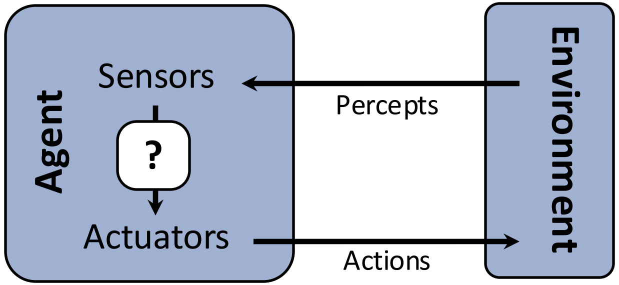
- A **rational agent** selects actions that maximize its (expected) **utility**.

- Utility or performance measure of a vacuum-cleaner agent:
  - amount of dirt cleaned up
  - amount of time taken
  - amount of electricity consumed
  - amount of noise generated
  - etc.



# Rational Agents

- A **rational agent**
  - acts appropriately given goals and circumstances
  - is flexible to changing environments and goals
  - learns from experience
  - makes appropriate choices given perceptual and computational limitations
- Characteristics of the **percepts, environment,** and **action space** dictate techniques for selecting rational actions.



# Rational Agents

- Are rational agents omniscient?
  - No – they are limited by the available percepts
- Are rational agents clairvoyant?
  - No – they may lack knowledge of the environment dynamics
- Do rational agents explore and learn?
  - Yes – in unknown environments these are essential
- So rational agents are not necessarily successful, but they are autonomous

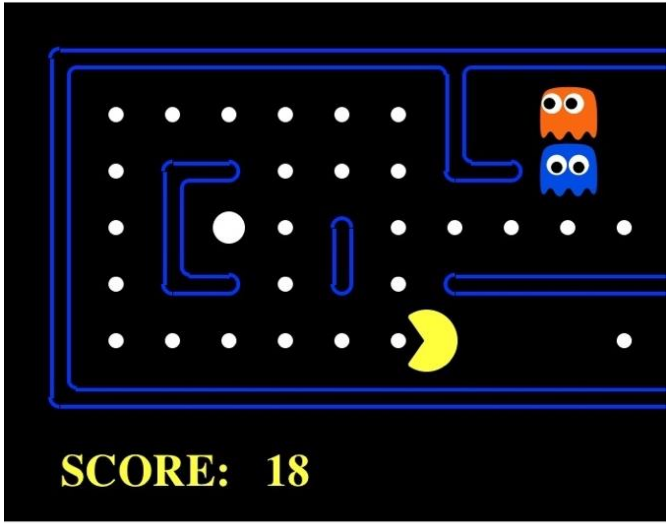
# Discussion Item

- A realistic agent has finite amount of computation and memory available. Assume an agent is killed because it did not have enough computation resources to calculate some rare event that eventually ended up killing it. Can this agent still be rational?

# PEAS: Performance measure, Environment, Actuators, Sensors

# PEAS: Pacman

- Performance measure
  - -1 per step; + 10 food; +500 win; -500 die;
- Environment
  - Maze, food, ghosts, …
- Actuators
  - Pacman’s body and mouth
- Sensors
  - Some sort of Vision (Entire state is visible)



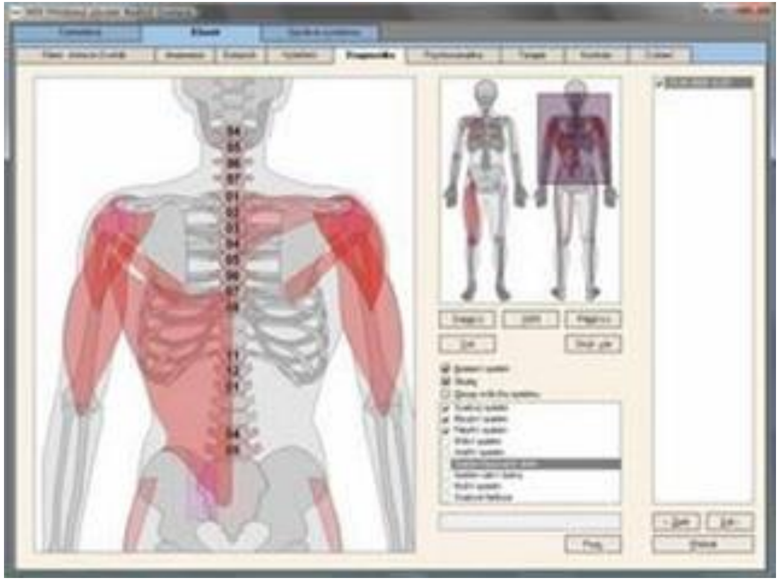
# PEAS: Automated Taxi

- Performance measure
  - Income, happy customer, vehicle costs, fines, insurance premiums
- Environment
  - US streets, other drivers, customers
- Actuators
  - Steering, brake, gas, display/speaker
- Sensors
  - Camera, radar, accelerometer, engine sensors, microphone

Image: http://nypost.com/2014/06/21/how-google-might-put-taxi-drivers-out-of-business/

# PEAS: Medical Diagnosis System

- Performance measure
  - Patient health, cost, reputation
- Environment
  - Patients, medical staff, insurers, courts
- Actuators
  - Screen display, email
- Sensors
  - Keyboard/mouse



# Environment Types

# Environment Types

- Fully Observable (vs. Partially Observable)
- Deterministic (vs. Stochastic)
- Episodic (vs. Sequential)
- Static (vs. Dynamic)
- Discrete (vs. Continuous)
- Single-Agent (vs. Multi-Agent):

# Fully Observable vs. Partially-Observable Domains

- Fully-observable: The agent has access to all information in the environment relevant to its task.
- Partially-observable: Parts of the environment are inaccessible

| Pacman | Crossword | Backgammon | Pick&Place Robot | Diagnosis | Taxi |
|---|---|---|---|---|---|
| Fully | Fully | Fully | Partially | Partially | Partially |

# Deterministic vs. Stochastic Domains

If an agent knew the initial state and its action, could it predict the resulting state? The dynamics can be:

- Deterministic: the resulting state is determined from the action and the state
- Stochastic: there is uncertainty about the resulting state

| Pacman | Crossword | Backgammon | Pick&Place Robot | Diagnosis | Taxi |
|---|---|---|---|---|---|
| Deterministic | Deterministic | Stochastic | Stochastic | Stochastic | Stochastic |

# Episodic vs Sequential Domains

- Episodic: Current action is independent of previous actions.
- Sequential: Current choice of action will affect future actions

| Pacman | Crossword | Backgammon | Pick&Place Robot | Diagnosis | Taxi |
|---|---|---|---|---|---|
| Sequential | Sequential | Sequential | Episodic | Sequential | Sequential |

# Static vs Dynamic Domains

- Static: Environment does not change while the agent is deliberating over what to do

- Dynamic: Environments does change

| Pacman | Crossword | Backgammon | Pick&Place Robot | Diagnosis | Taxi |
|---|---|---|---|---|---|
| Static | Static | Static | Dynamic | Dynamic | Dynamic |

# Discrete vs Continuous Domains

- Discrete: A limited number of distinct, clearly defined states, percepts, actions, and time steps (otherwise continuous)

| Pacman | Crossword | Backgammon | Pick&Place Robot | Diagnosis | Taxi |
|---|---|---|---|---|---|
| Discrete | Discrete | Discrete | Continuous | Continuous | Continuous |

# Single-agent vs. Multi-agent Domains

- Does the environment include other agents?
- If there are other agents whose actions affect us
  - It can be useful to explicitly model their goals and beliefs, and how they react to our actions
- Other agents can be: cooperative, competitive, or a bit of both

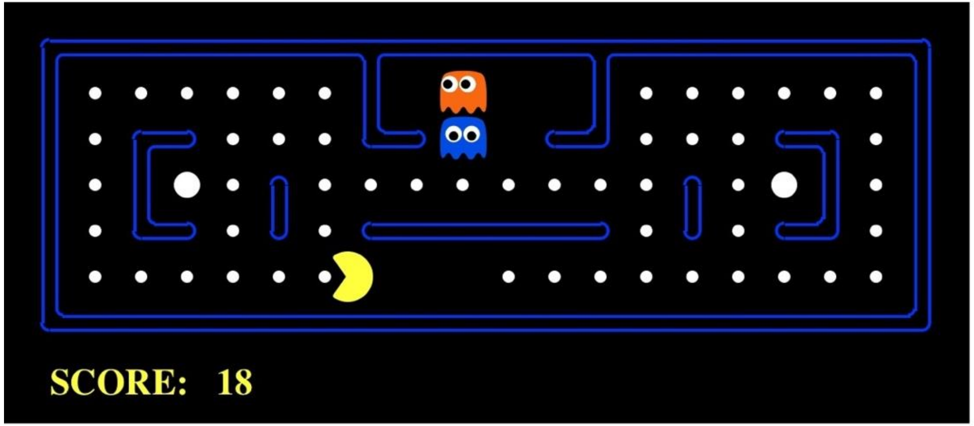
| Pacman | Crossword | Backgammon | Pick&Place Robot | Diagnosis | Taxi |
|---|---|---|---|---|---|
| Multi | Single | Multi | Single | Single | Multi |

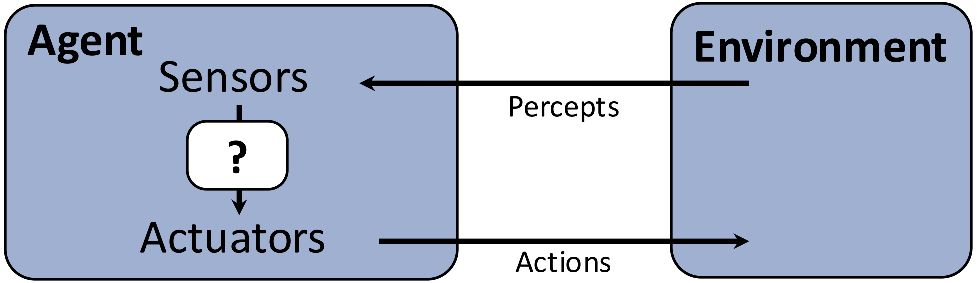
# Environment Types: Summary

| | Pacman | Crossword | Backgammon | Pick&Place Robot | Diagnosis | Taxi |
|---|---|---|---|---|---|---|
| **Fully or Partially Observable** | Fully | Fully | Fully | Partially | Partially | Partially |
| **Deterministic or Stochastic** | Deterministi c | Deterministic | Stochastic | Stochastic | Stochastic | Stochastic |
| **Episodic or Sequential** | Sequential | Sequential | Sequential | Episodic | Sequential | Sequential |
| **Static or Dynamic** | Static | Static | Static | Dynamic | Dynamic | Dynamic |
| **Discrete or Continuous** | Discrete | Discrete | Discrete | Continuous | Continuous | Continuous |
| **Single-agent or Multiagent** | Multi | Single | Multi | Single | Single | Multi |

# Agent Types

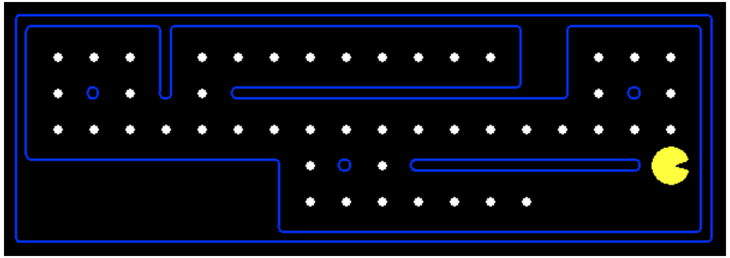
# Pac-Man as an Agent

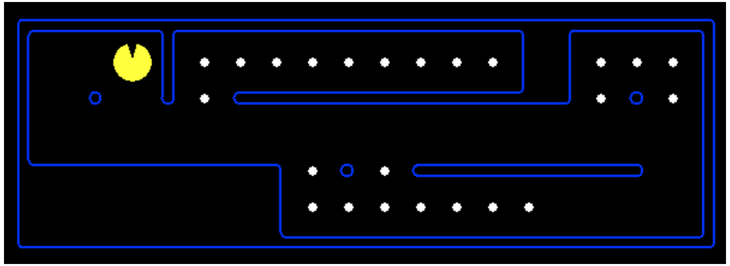


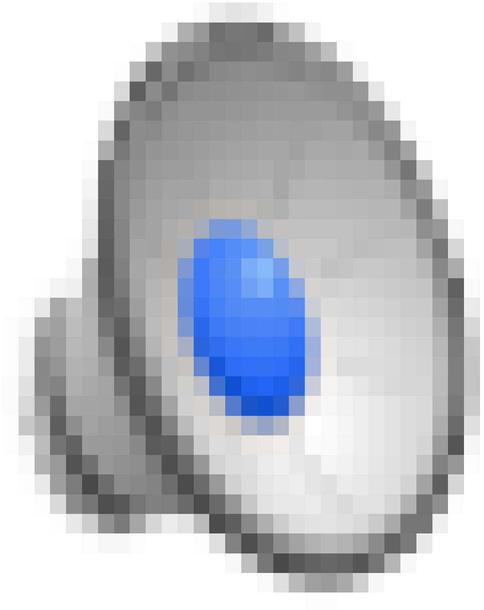
# Reflex Agents

- Reflex agents:
  - Choose action based on current percept (and maybe memory)
  - May have memory or a model of the world's current state
  - Do not consider the future consequences of their actions
  - Consider how the world IS
- Can a reflex agent be rational?

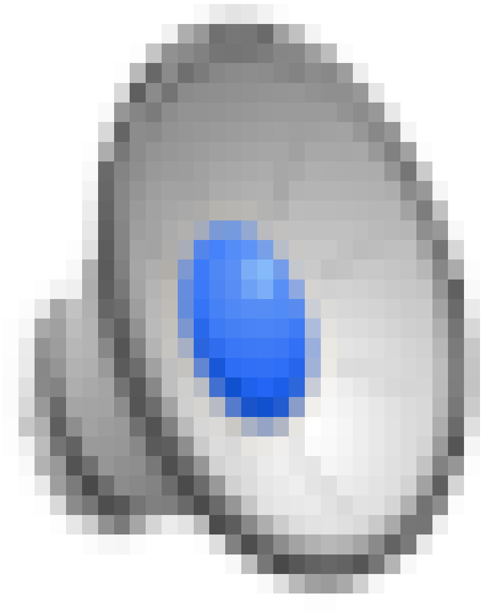




# Video of Demo Reflex Optimal

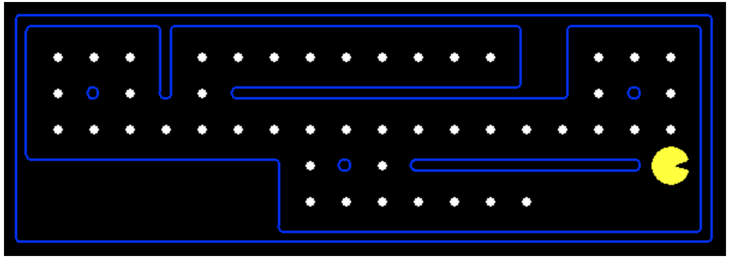


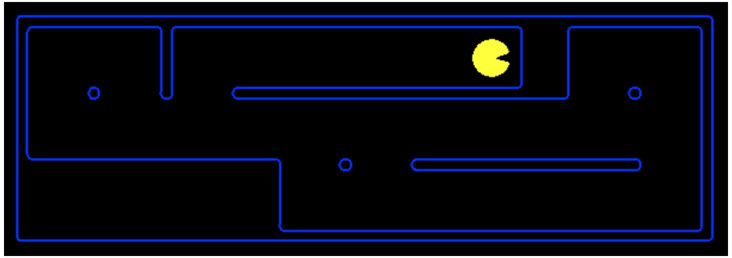
# Video of Demo Reflex Odd



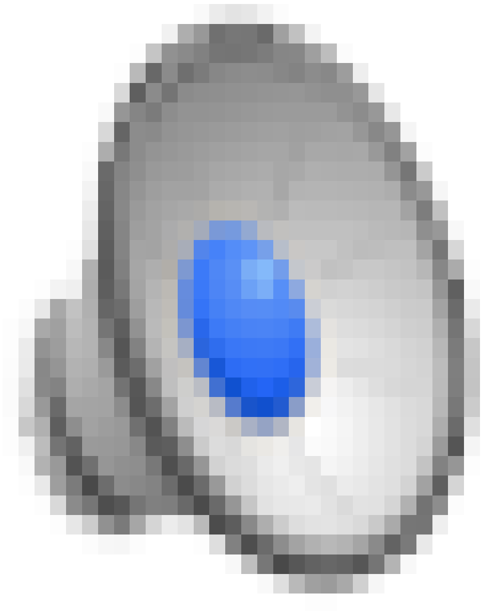
# Planning Agents

- Planning agents:
  - Ask "what if"
  - Decisions based on (hypothesized) consequences of actions
  - Must have a model of how the world evolves in response to actions
  - Must formulate a goal (test)
  - Consider how the world WOULD BE
- Optimal vs. complete planning
- Planning vs. replanning

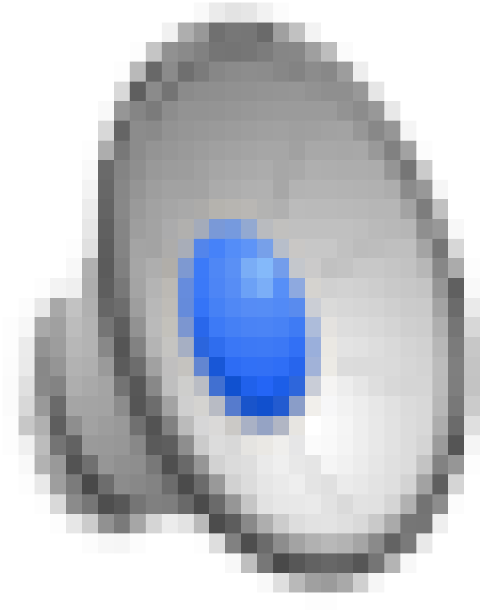


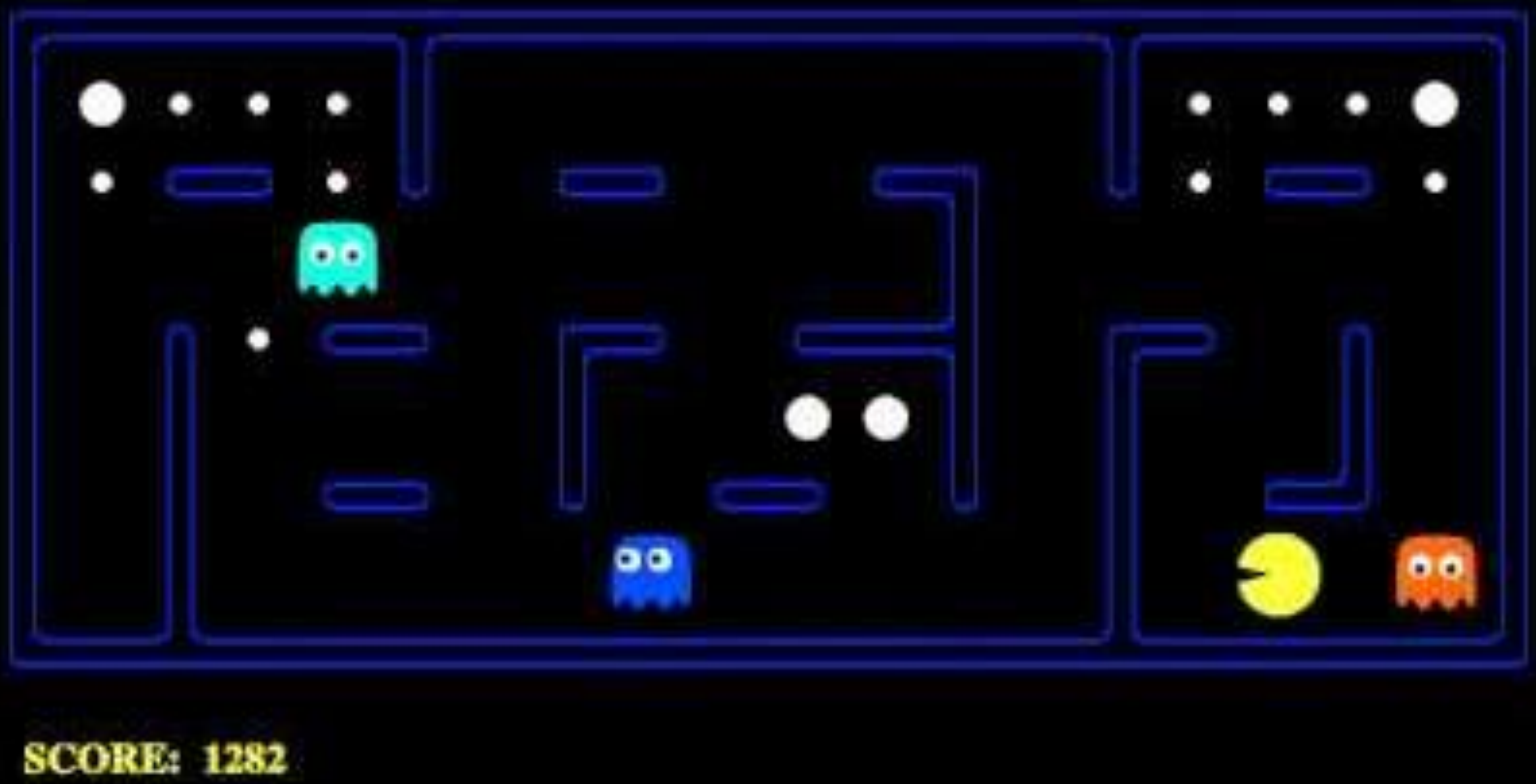


# Video of Demo Re-planning



# Video of Demo Mastermind




SCORE: 1282

# Ethics and Implications

- Robust, fully autonomous agents in the real world

- What happens when we achieve this goal?

# Ethics and Implications

- Who is liable if a robot driver has an accident?
- What will we do with super-intelligent machines?
- Would such machines have conscious existence? Rights?
- Can human minds exist indefinitely within machines (in principle)?

# Announcements

- Project 0: Python Tutorial
  - I encourage team of 2 for doing the projects
  - I encourage pair programming
  - DO NOT SEPARATE THE TASKS BETWEEN EACH OTHER!
  - Due Monday at 11:59pm  (pulse check to see you are in + get to know submission system)
- Homework 0: Math self-diagnostic
  - Important to check your preparedness for second half
  - GradeScope
- Project 1: Search
  - Will go out next week
  - Longer than most, and best way to test your programming preparedness
- Pinned posts on piazza


PAIR
PROGRAMMING