Artificial Intelligence

A Brief (and Buzzword Free) Introduction

Lt Col Adrian A. de Freitas

adrian.defreitas@afacademy.af.edu

UNITED STATES AIR FORCE ACADEMY DEPARTMENT OF COMPUTER AND CYBER SCIENCES

Overview

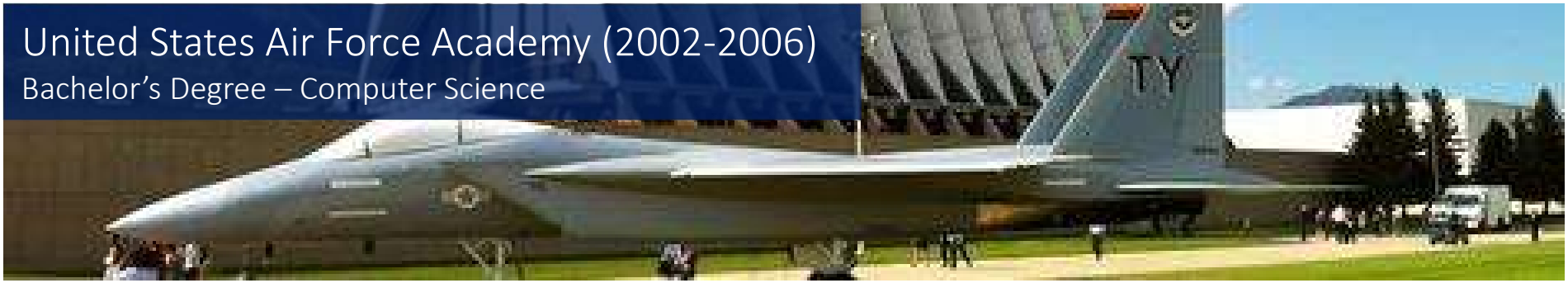
- Introduction
- What is Artificial Intelligence / Machine Learning?
- How Do They Work?
- What Are the Limitations / Implications of these Technologies?
- Who Should Learn About AI/ML?

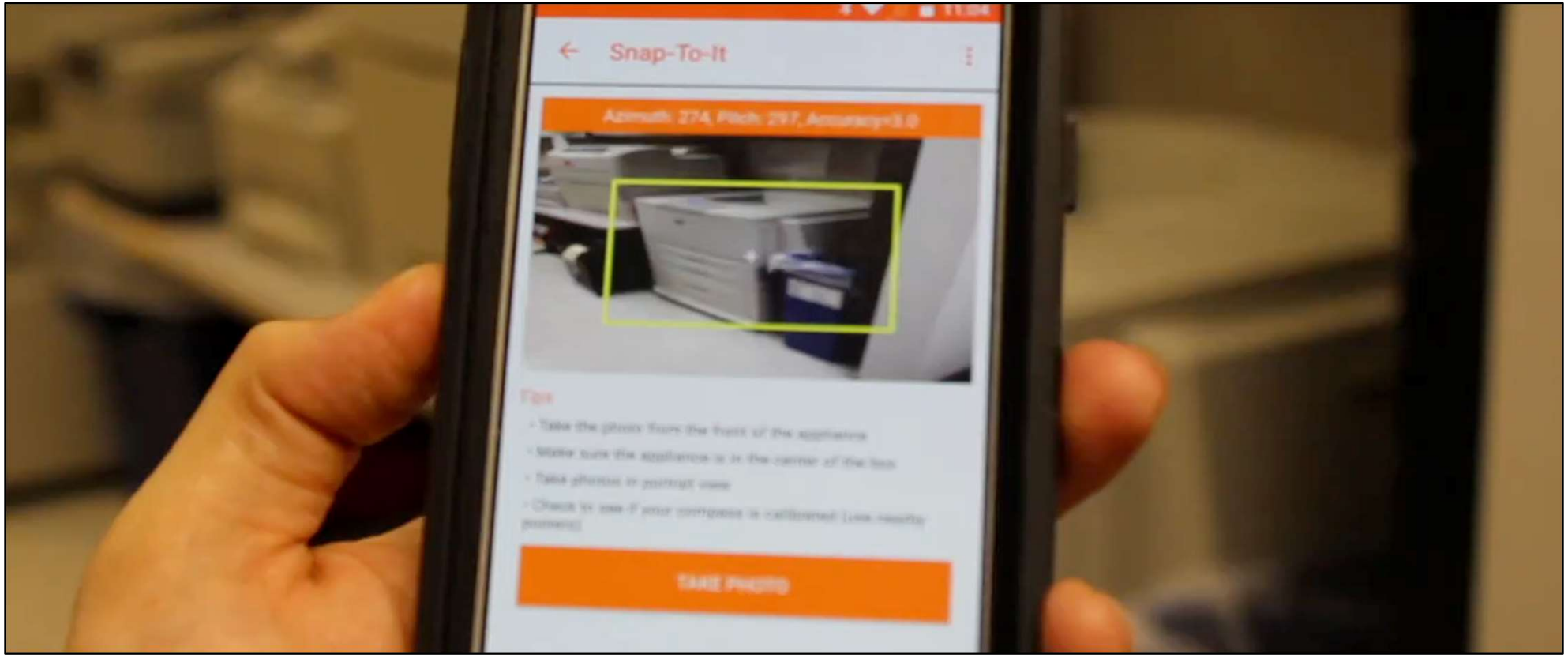
About Me

Lt Col Adrian de Freitas (pronounced: “dee-FRAY-tus”)

- Deputy Department Head, Department of Computer and Cyber Sciences
- AFSC: Cyberspace Operations Officer (17D)
- Email: adrian.defreitas@afacademy.af.edu

About Me (Education)





Snap-To-It: A User-Inspired Platform for Opportunistic Device Interactions

Adrian A. de Freitas[1], Michael Nebeling[1], Xiang 'Anthony' Chen[1], Junrui Yang[2], Akshaye Shreenithi Kirupa Karthikeyan Ranithangam[3], Anind K. Dey[1]

[1]Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA, USA

[2]Peking University
Beijing, China

[3]Coimbatore Institute of Technology,
Tamil Nadu, India

{adefreit, mnebelin, xiangche, anind}@cs.cmu.edu, {jackieyang51, akshaya.kar}@gmail.com

Ubicomp

Software Engineering Institute
Carnegie Mellon

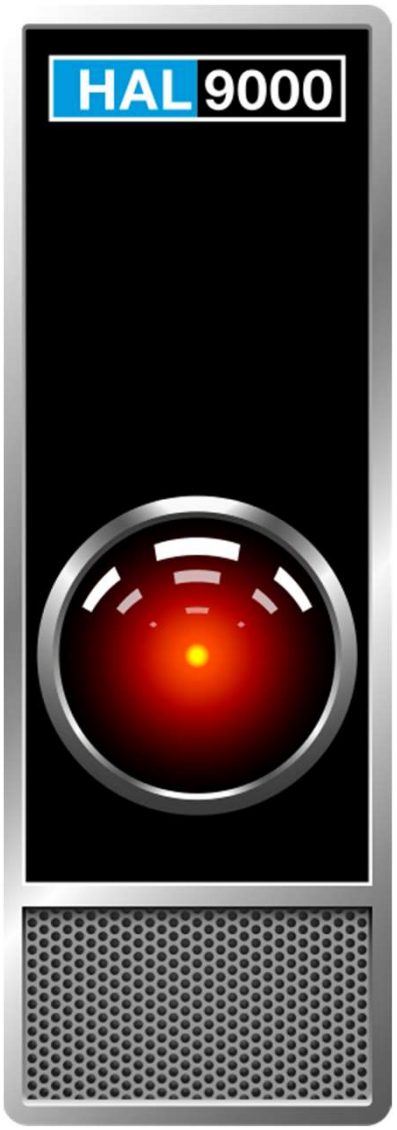
Carnegie Mellon University HCII

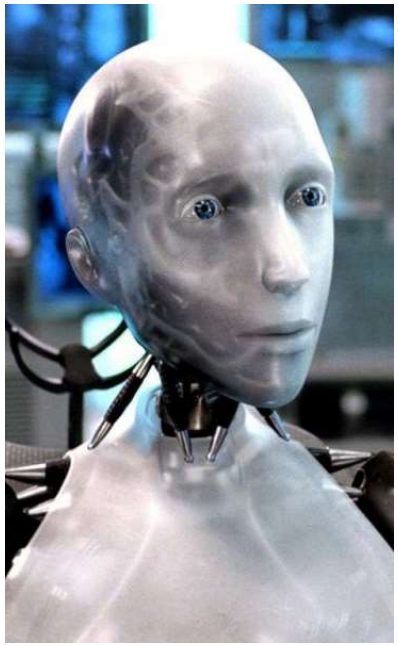
Prior Assignments

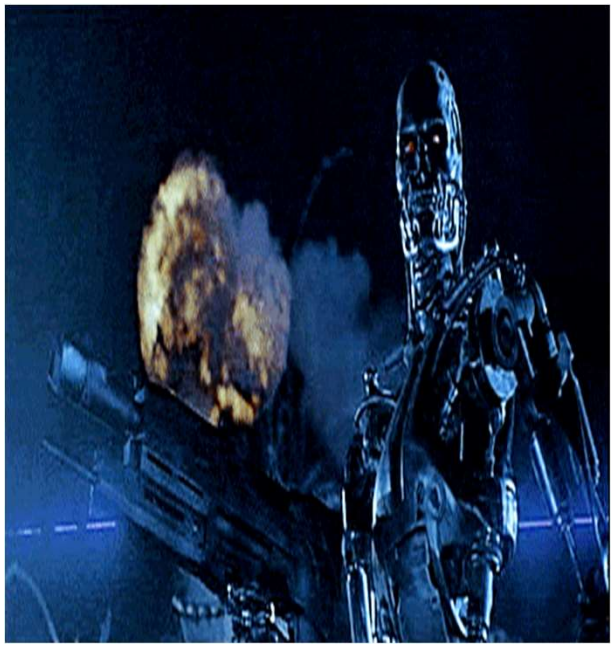
- Air Force Wargaming Institute (2007)
  - Lead Programmer for AF Wargames
- USAF Academy (2011-2013)
  - Instructor and XO
- Deployments
  - Eskan Village, Saudi Arabia (2007)
  - JTF Guantanamo Bay (2010)
  - US Central Command (2018)
- Most Recent: Pentagon (2016-2019)
  - AF Deputy Chief Technology Officer

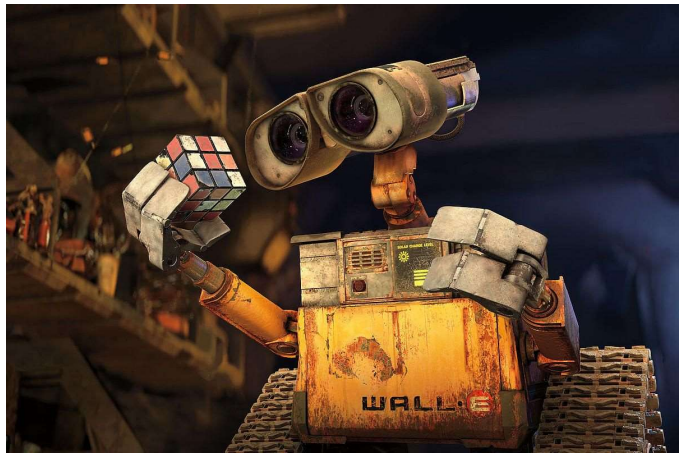
Before We Get Started

- What do you think about when you hear the term, “Artificial Intelligence”









What is Artificial Intelligence?

- AI is the part of Computer Science concerned with creating systems that can:
  - Think like humans
  - Think rationally
  - Act like humans
  - Act rationally

- It's not a very satisfying definition, is it?

  "The problem is that we cannot yet characterize in general what kinds of computational procedures we want to call intelligent. We understand some of the mechanisms of intelligence and not others."

  - John McCarthy ("Father" of AI)

History of AI: Early Thoughts

1950 – Alan Turing Publishes "Can Machines Think"

1955 – The First AI Program is Invented

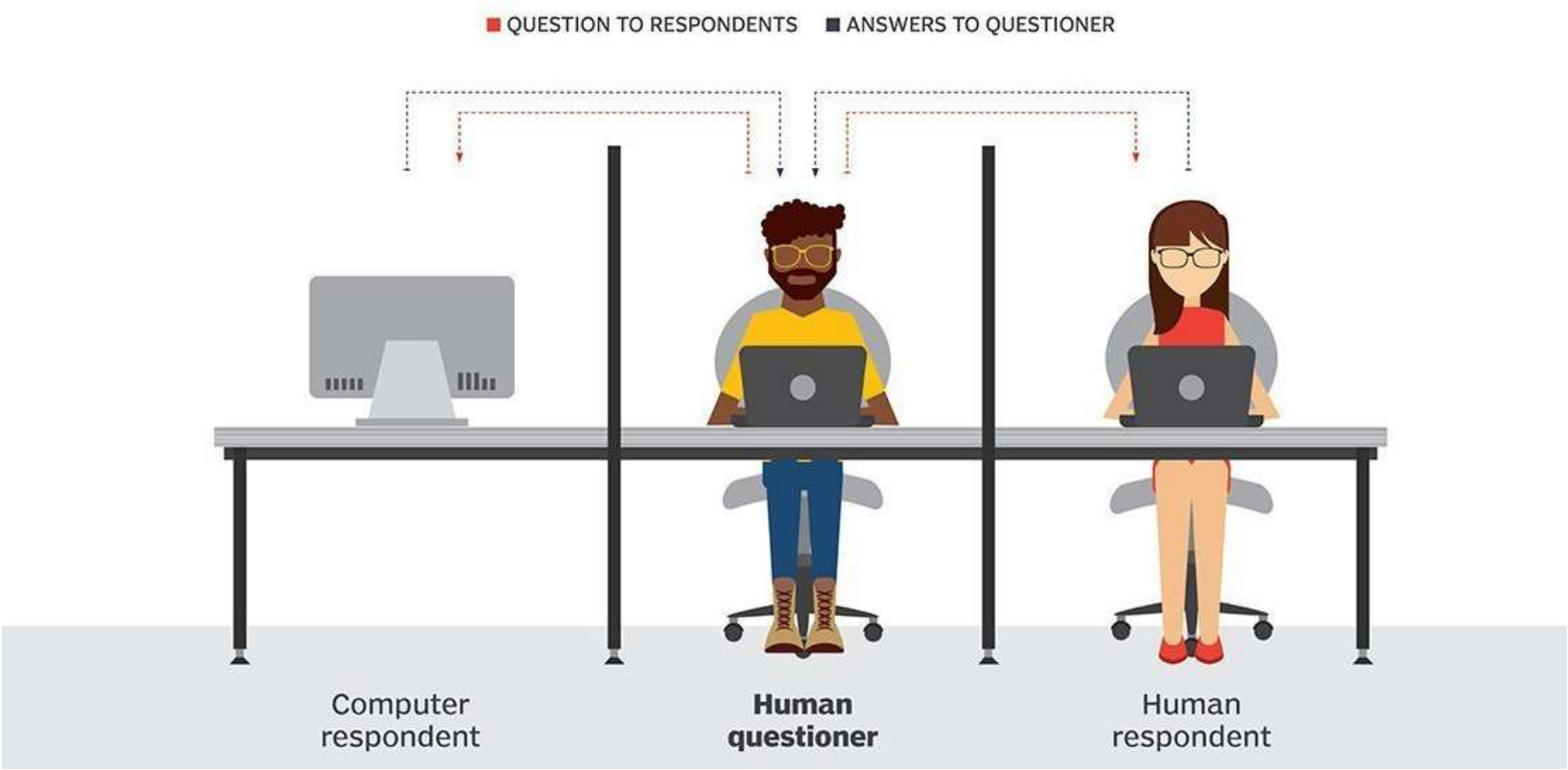
```
TO PROVE:  *2.14   --PIP                          Given
           *2.13   AV---A                         Substitution
                   PV---P                         Given
           *1.4    [AVB]I[BVA]                    Substitution
                   [PV---P]I[---PVP]              Detachment
                   ---PVP                         Given
           *1.01   [AIB].=.[-AVB]                 Substitution
                   [--PIP].=.[---PVP]             Replacement
           *2.14   --PIP
                   Q.E.D.
```

The Early Days (1950 – 1955)	The Golden Years (1956 – 1974)	Expert Systems (1974 – 1990ish)	Intelligent Agents (1990s – 2010ish)	Deep Learning (2010 – Present)

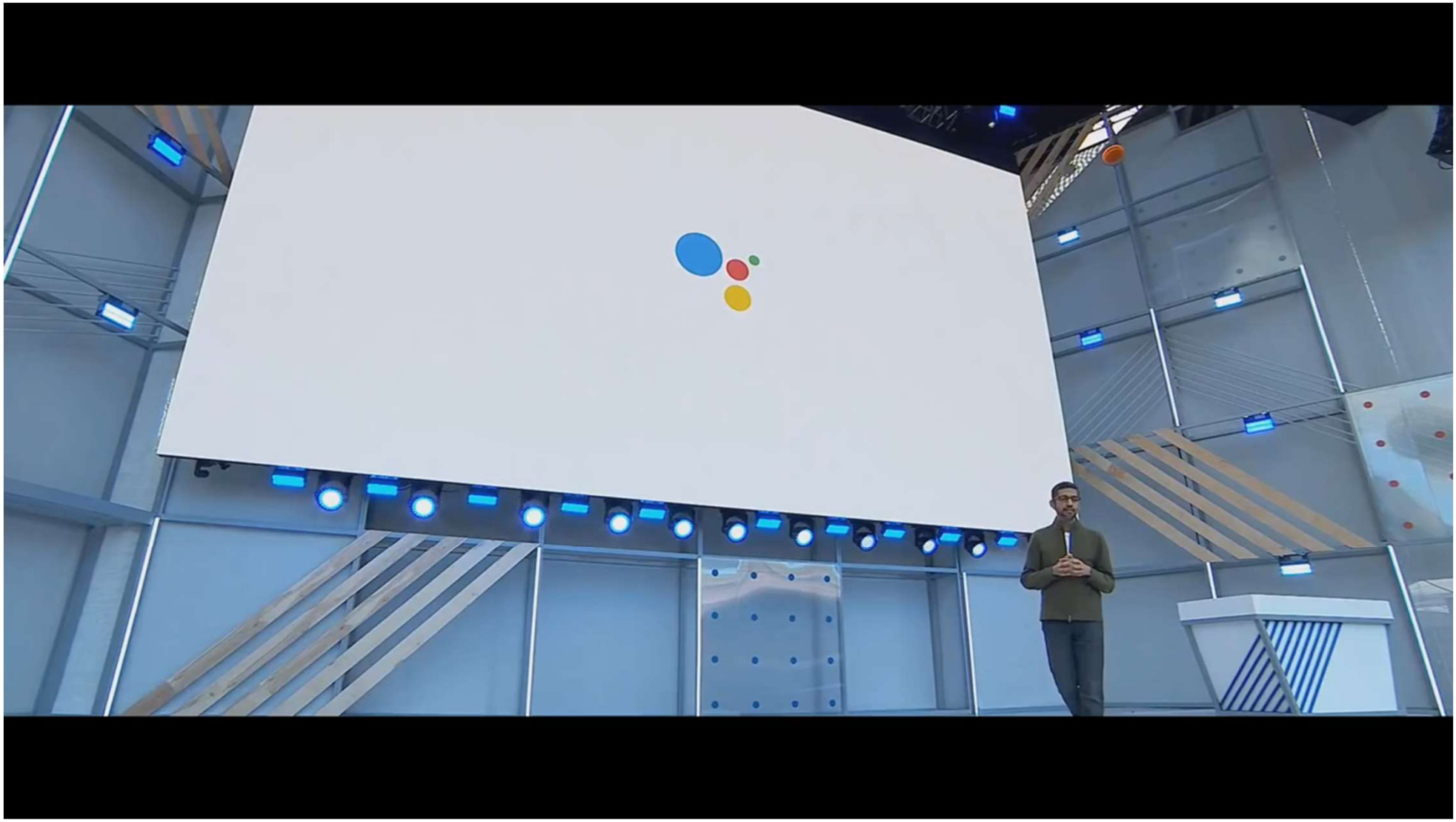
- Before AI, the idea of a computer "thinking" seemed preposterous
- Early work focused on determining the criteria that a machine would need to satisfy in order to be considered "intelligent."

The Imitation Game (i.e., “Turing Test”)

- A human asks a series of questions to one or more unknown respondents. Once complete, the questioner tries to determine if the respondent(s) are human or a computer.



A Modern Turing Test



Google I/O - **2018**

History of AI: Early Progress

1956 Dartmouth Conference: The Founding Fathers of AI

John MacCarthy | Marvin Minsky | Claude Shannon | Ray Solomonoff | Alan Newell

Herbert Simon | Arthur Samuel | Oliver Selfridge | Nathaniel Rochester | Trenchard More

1950 – Alan Turing Publishes "Can Machines Think"

1955 – The First AI Program is Invented

1956 – Dartmouth Meeting – "AI" Coined

1959 – General Problem Solver

1964 – Eliza – the first "chatbot"

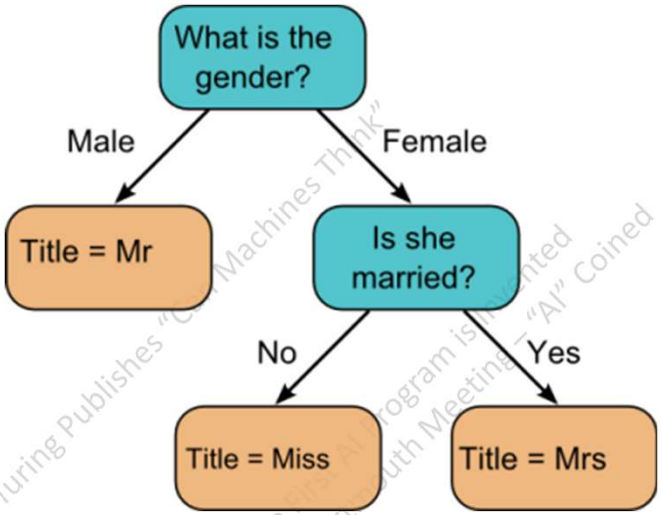
The Early Days (1950 – 1955)	The Golden Years (1956 – 1974)	Expert Systems (1974 – 1990ish)	Intelligent Agents (1990s – 2010ish)	Deep Learning (2010 – Present)

1968 – 2001: A Space Odysey

"By the year 2001 we will have machines with intelligence that matched or exceeded human's"

- Arthur Clarke

History of AI: Development of Expert Systems



"An expert system is a computer system that emulates, or acts in all respects, with the decision-making capabilities of a human expert"

- Dr. Edward Feigenbaum, Stanford University

Assumptions

1. There are human experts that can perform the task
2. The human experts can articulate their methodology sufficiently enough for a programmer to code
3. The human experts' understanding of the problem domain is comprehensive

1950 – Alan Turing Publishes "Can Machines Think"
1955 – The First AI Program is Invented
1956 – Dartmouth Meeting – "AI" Coined
1959 – General Problem Solver
1964 – Eliza – the first "chatbot"

The Early Days (1950 – 1955)	The Golden Years (1956 – 1974)	Expert Systems (1974 – 1990ish)	Intelligent Agents (1990s –2010ish)	Deep Learning (2000s – Present)

1968 – 2001: A Space Odyssey

"By the year 2001 we will have machines with intelligence that matched or exceeded human's"
- Arthur Clarke

1983 – WarGames
1984 – The Terminator
1987 – Star Trek: TNG

History of AI: Exciting Achievements

1950

1964 – Eliza – the first "chatbot"

1994 – Checkers AI
1996 – Deep Blue (Chess)
2007 – Checkers "Solved"
2011 – Watson (Jeopardy)

The Early Days (1950 – 1955)	The Golden Years (1956 – 1974)	Expert Systems (1974 – 1990ish)	Intelligent Agents (1990s – 2010ish)	Deep Learning (2010s – Present)

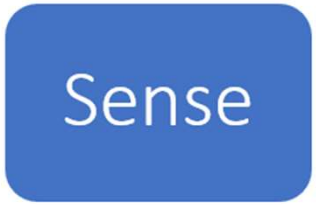
1968 – 2001: A Space Odyssey

"By the year 2001 we will have machines with intelligence that matched or exceeded human's"
- Arthur Clarke

1983 – WarGames
1984 – The Terminator
1987 – Star Trek: TNG

How Does AI Work?

- AIs make decisions using a three-step process:



Sense Phase

The AI "agent" either detects or is provided with information about its environment.

Goal:

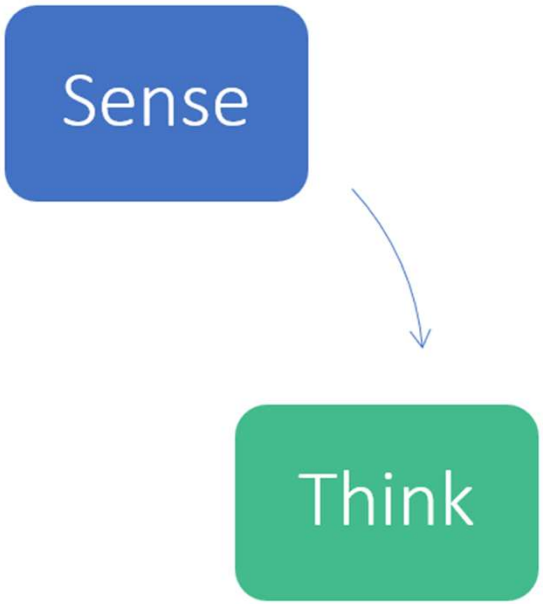
- Give the agent all the information it needs to make an informed decision.

Examples:

- The location of player(s) / threat(s)
- Nearby points of interest (e.g., protective cover, hazards)
- The agent's health (and/or the health of other players)

How Does AI Work?

- AIs make decisions using a three-step process:



Think Phase

The AI considers possible actions and determines which one(s) it wants to perform.

Goal:

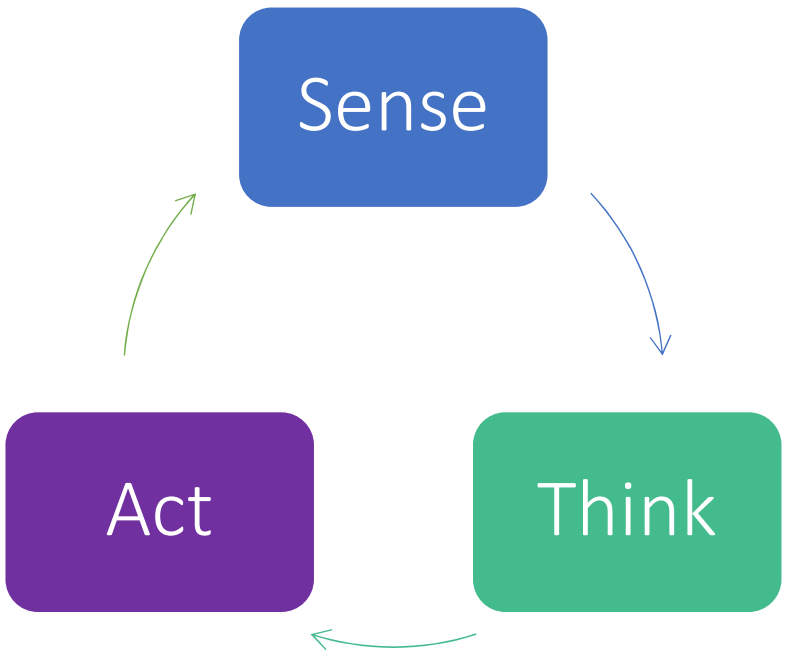
- Try to find the "best" action given its current state (*i.e.,* situation)

Examples:

- Deciding whether to attack or hide/heal
- Deciding where to move
- Deciding whether to use an item, or save it for later

How Does AI Work?

- AIs make decisions using a three-step process:



Act Phase

The AI performs the action(s) that it chose in the previous step.

Goal:

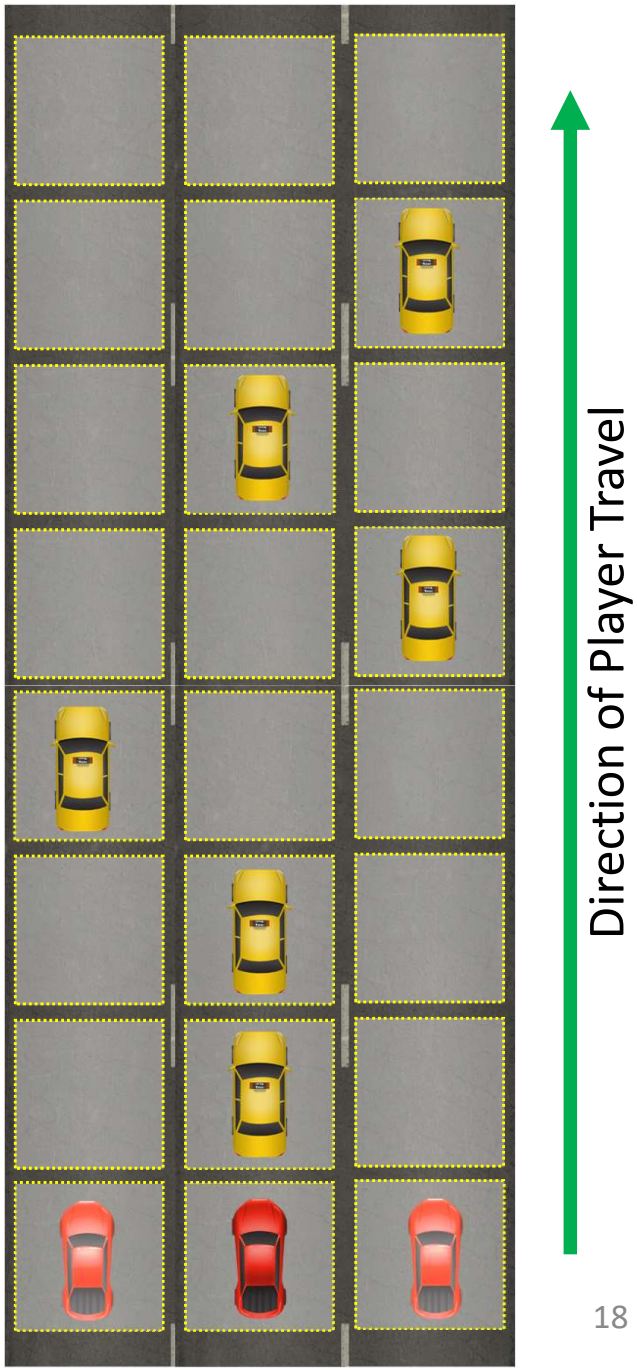
- To actually carry out the desired action in the game world.

Examples:

- Attacking
- Healing/Hiding
- Moving to a Location
- Using Items

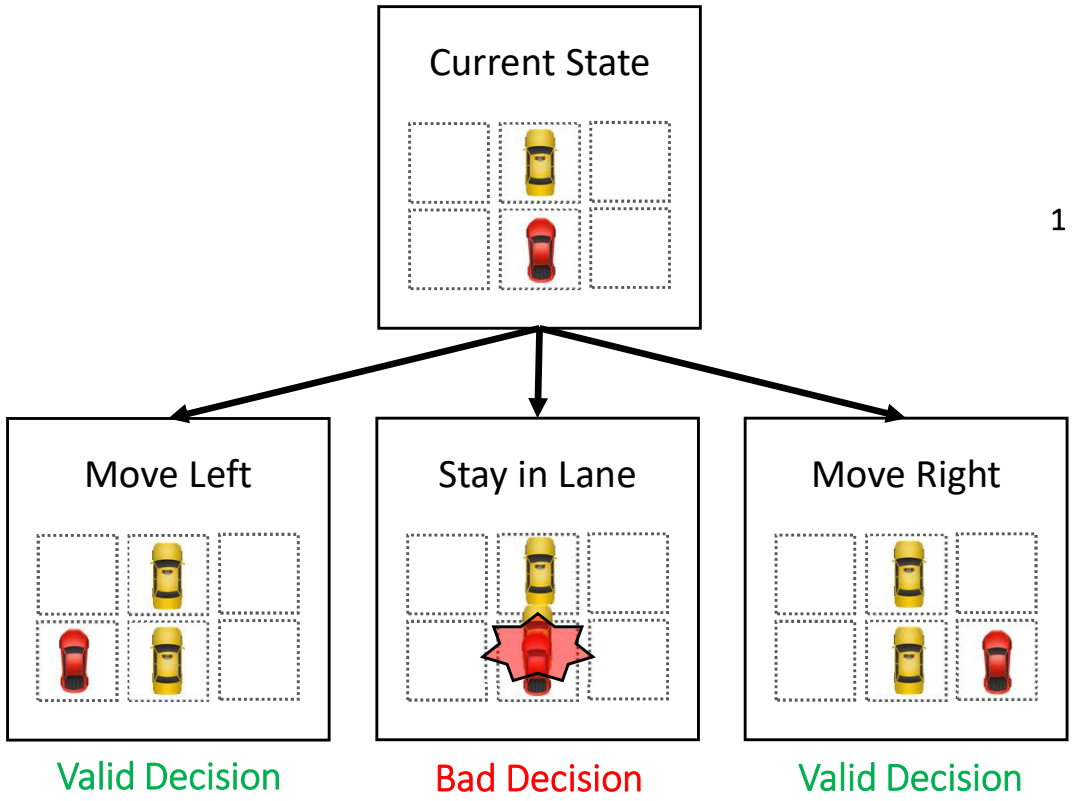
Example: Spy Hunter

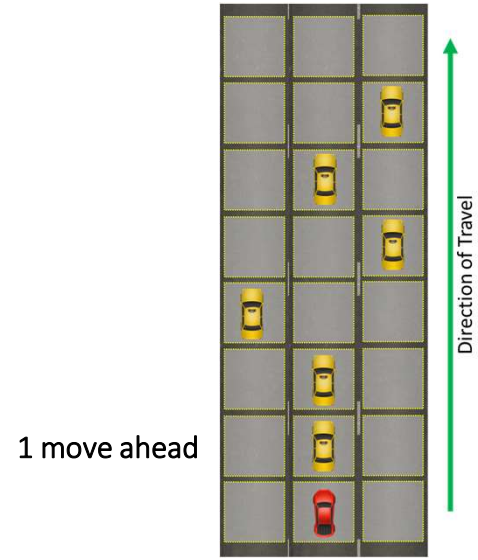
- <u>Goal: Drive the player's car (red) through a busy highway for as long as possible without hitting any of the other cars (yellow).</u>
- Notes:
  - Spy Hunter is usually a real-time game, but we will treat it as turn-based for this discussion.
  - The yellow squares represent locations where a car (player or enemy) can be located.
  - Each "turn", the player has (up to) three choices:
    - Stay in the Same Lane
    - Move Left
    - Move Right
  - The game is **deterministic:**
    - We know with 100% certainty what the game state will look like after an action



Looking at the First Move

- Let's look at our options . . .

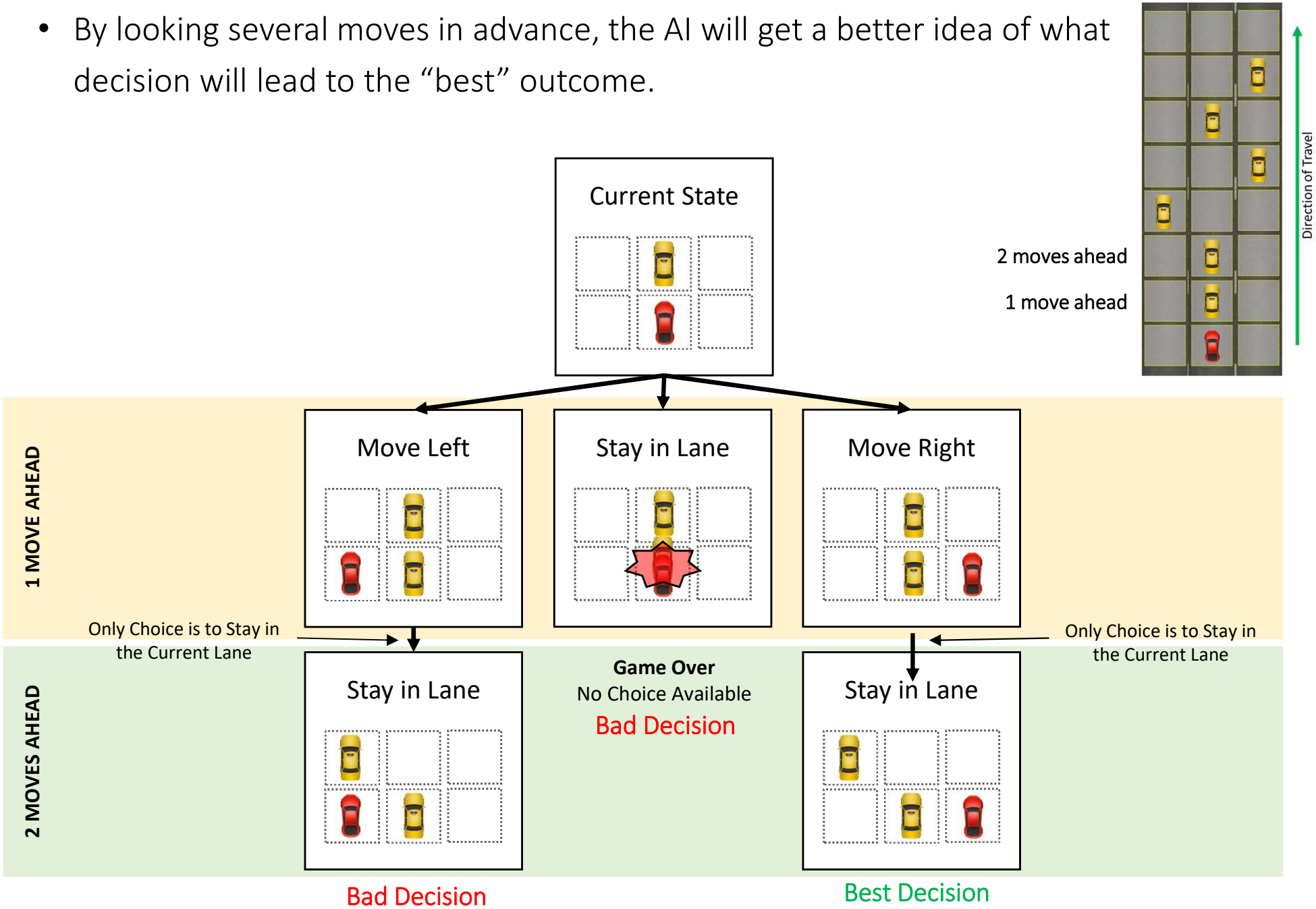




Which decision should the AI choose?

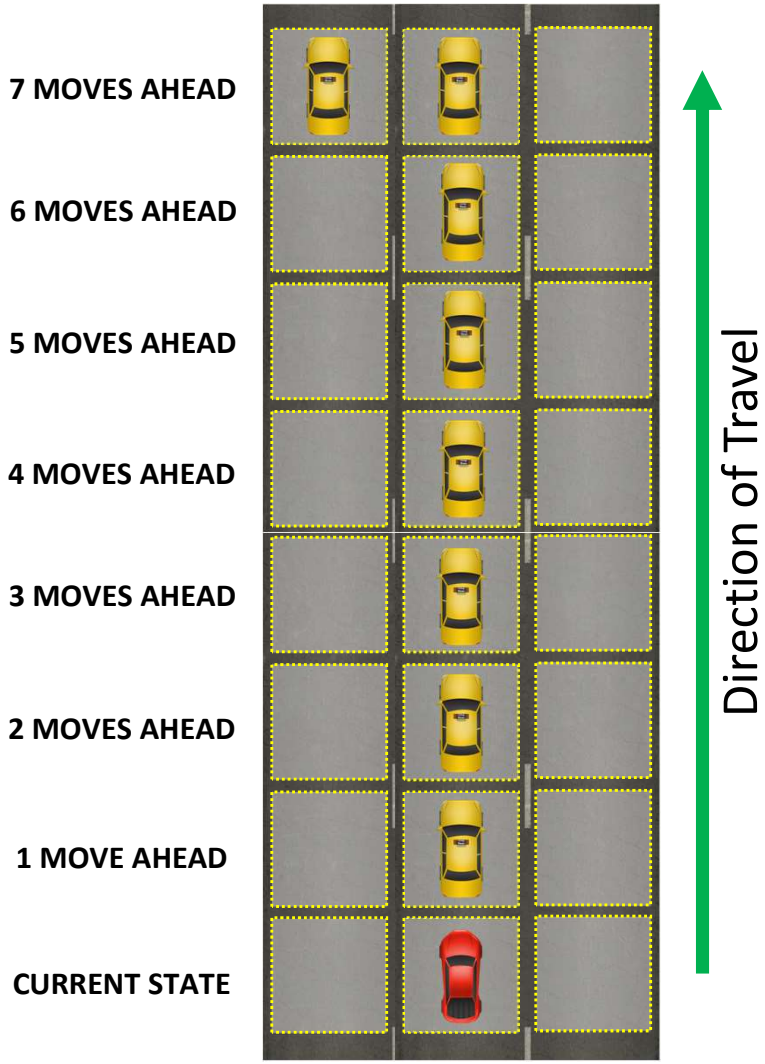
Looking Two Moves Ahead

- By looking several moves in advance, the AI will get a better idea of what decision will lead to the “best” outcome.



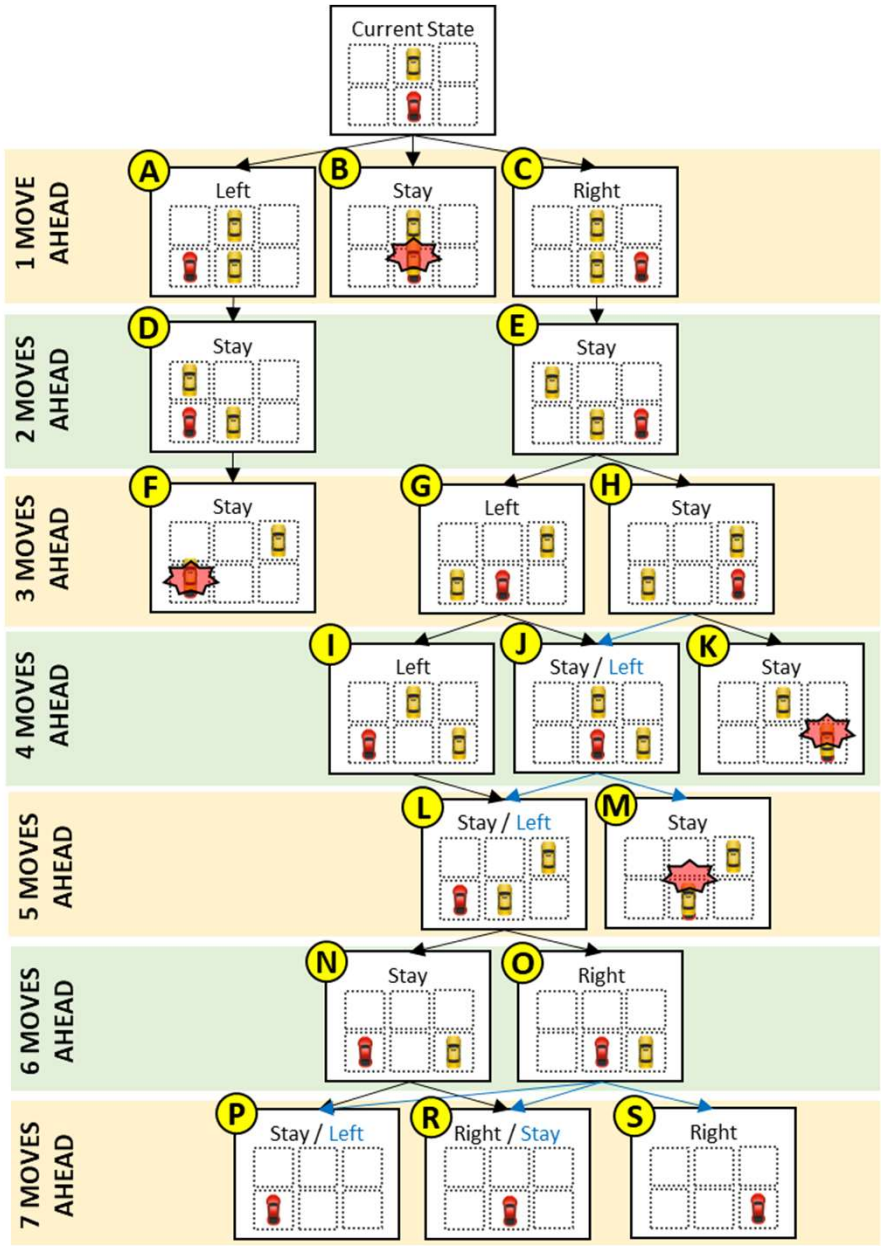
How Far Ahead

- Looking two moves ahead got us past a trap this time, but can you think of any situations where we need to look further into the future?



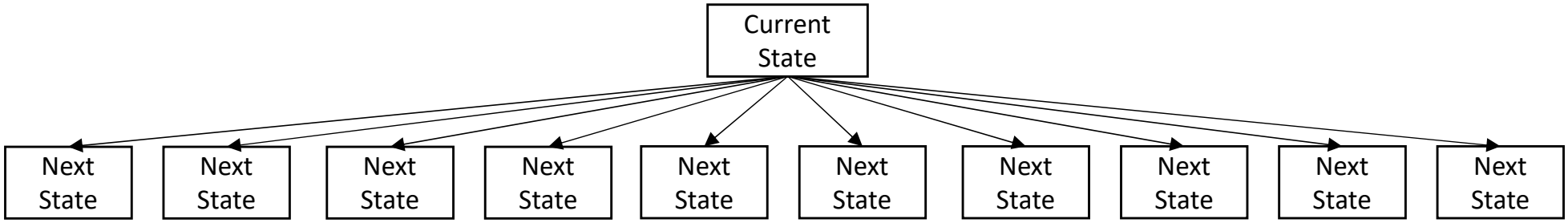
Exploring All Possible Decisions

- To play effectively, we need to expand the tree to depict:
  - All possible decisions
  - All possible game states



Handling Complexity

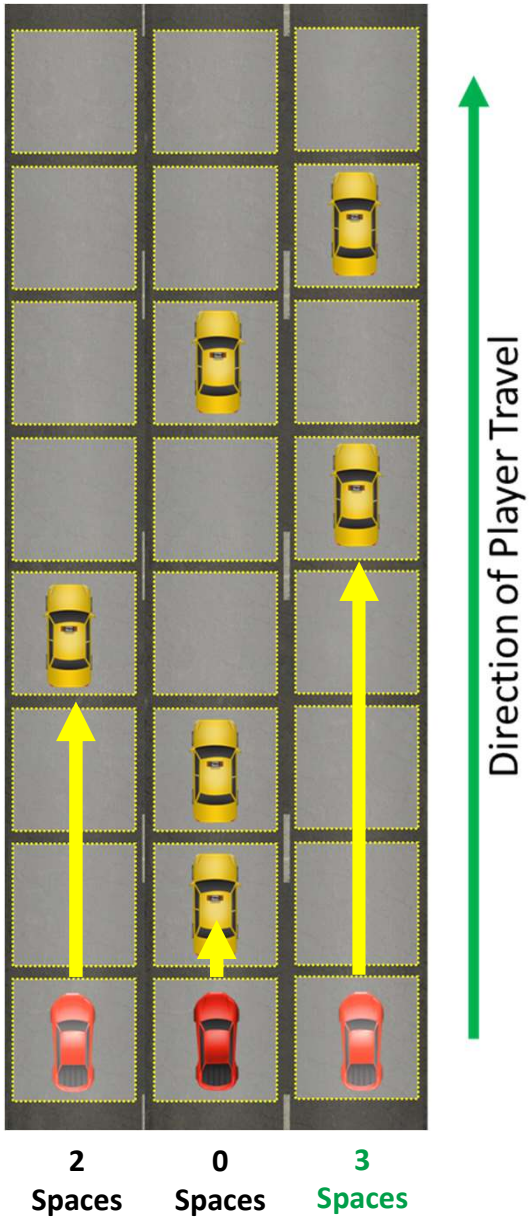
- Even though Spy Hunter only has <u>up to</u> 3 decisions per turn, there ended up being over 20 states in our search tree.
- Most games have a significantly larger search space:
  - On average, a Chess game has 35-50 possible actions per turn (10^{120} states)
  - Real-Games like Call of Duty have a near-infinite number of actions that an AI could take at any given instant.
    - Move (Here? Or there? Or Maybe Here?)
    - Attack (Who? Where should I aim? Should I intentionally miss?)
- How can we examine ALL of these states?
  - **Method 1**: Only look at some of them



  - **Method 2**: Develop a way to "estimate" the goodness of a state without having to look at all the lower levels.

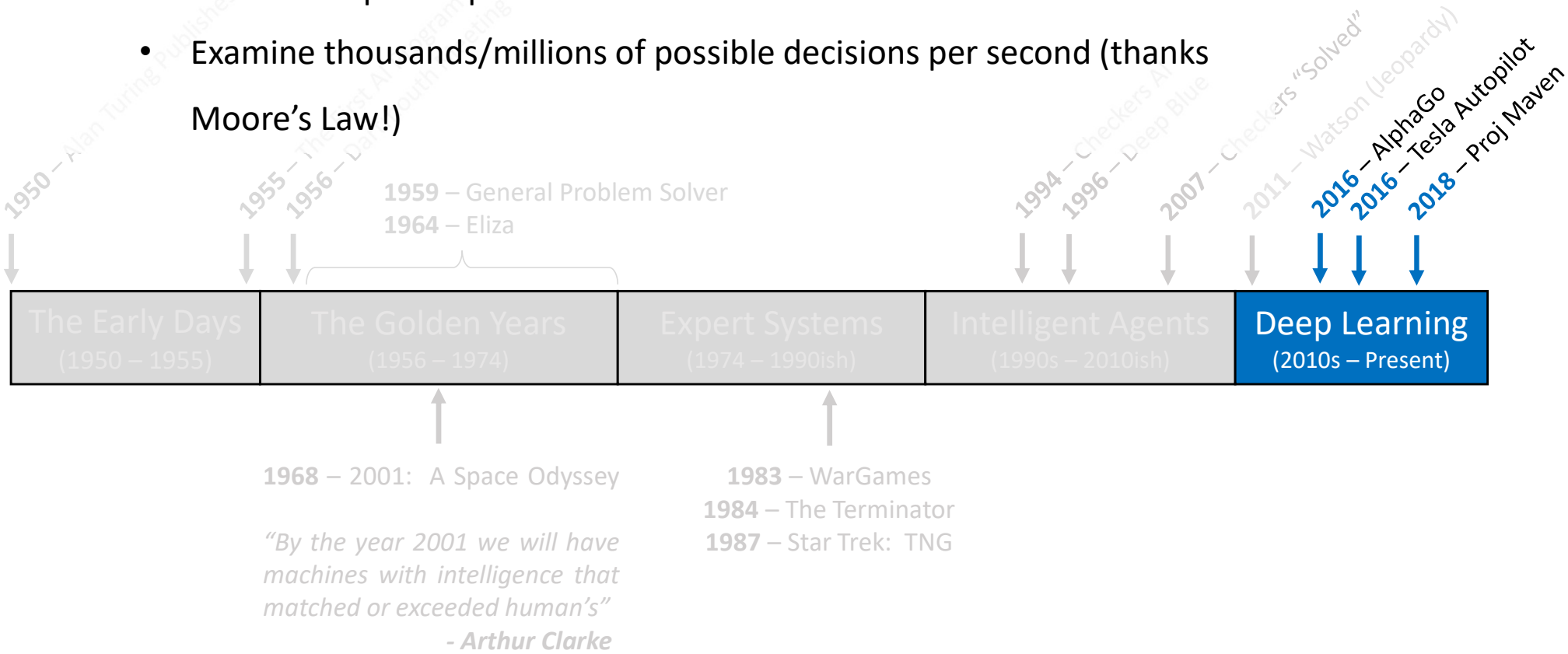
Heuristics to the Rescue!

- A heuristic is a methodology/formula for quantifying and approximating the "goodness" of a state.
- Pros
  - It gives us a way to compare game states to each other.
  - It gives us a way to estimate the outcome of a decision without having to look 2, 3, 4 etc. moves ahead.
- Cons
  - Coming up with a good heuristic requires domain-level knowledge.
  - It's not always easy to express what makes a game state "good"
  - <u>Heuristics are not guaranteed to lead to the optimal outcome.</u>



History of AI: Rise of Machine Learning

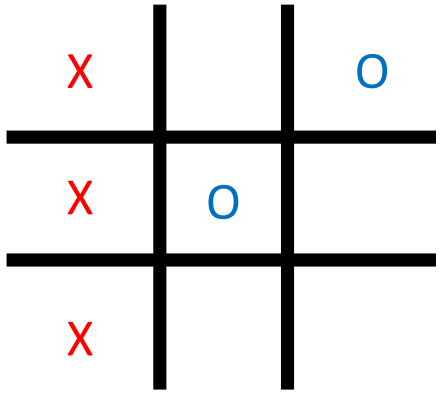
- Can we get computers to learn the rules on their own?
- Faster processors and larger storage allowed researchers to create Artificial Intelligences that can:
  - **Extract patterns** from vast quantities of training data
  - **Learn** from past experiences
  - Examine thousands/millions of possible decisions per second (thanks Moore's Law!)



Let’s Play a Game

Machine Learning

- **Task:** identify relationships between the independent variables (i.e., features) and the dependent variables (i.e., label).
- **Goal:** Create an AI/model that can predict the label by **only** looking at the features



By looking at these features . . .
(X = X player; O = O player; B = Blank)

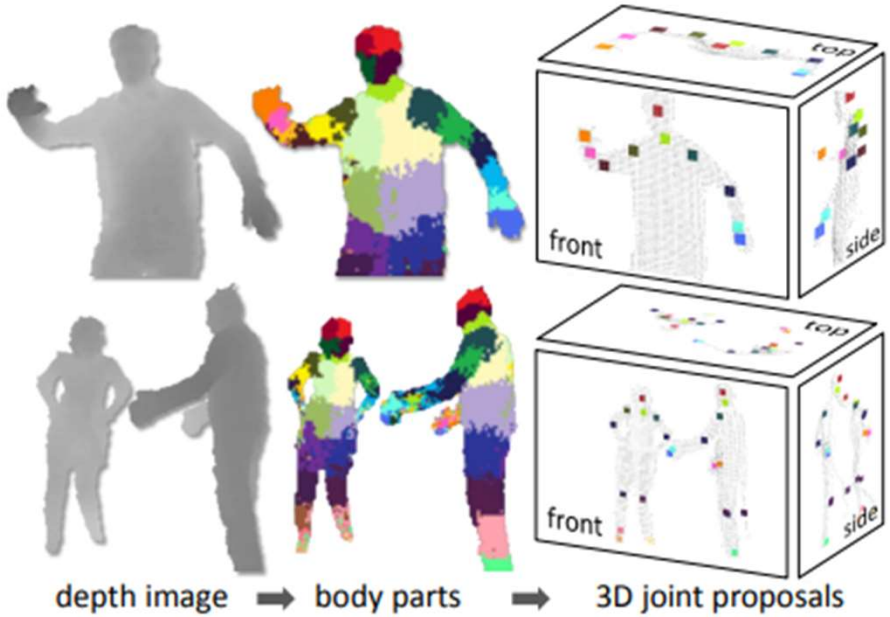
Can I Predict This?

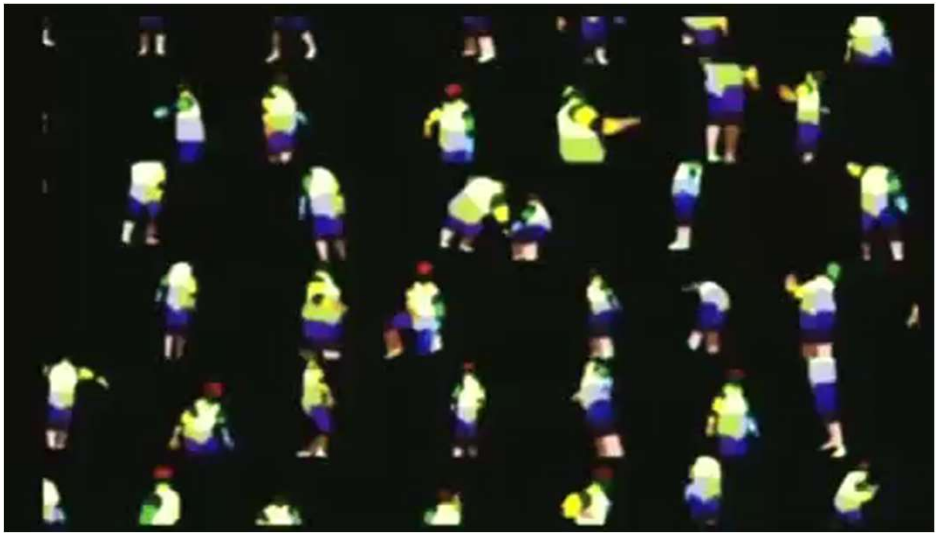
TOP_LEFT	TOP_MID	TOP_RIGHT	MID_LEFT	MID_MID	MID_RIGHT	BOT_LEFT	BOT_MID	BOT_RIGHT	X_WIN
X	B	O	X	O	B	X	B	B	POSITIVE

It’s a Lot of Work!

- It is a common misconception that Machine Learning is a “set it and forget it” process.
- Humans are involved in nearly every step, specifically when it comes to:
  - Determining how we are going to collect and label the data
  - Designing the training regimen such that it doesn’t introduce implicit biases
    - Too much of one data vs. the other
  - Examining the trained model to make sure it didn’t learn something irrelevant or inconsequential
    - **Example:** What if we found out that X usually wins when it is in the center of the board?
    - Correlation does not (always) equal causation!

Example: Microsoft Kinect





The Potential Benefits of AI

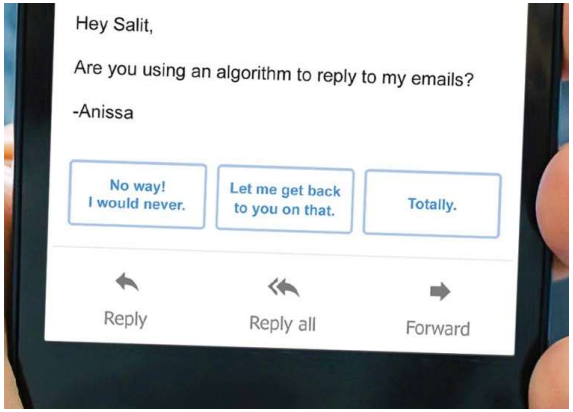
- Can (Potentially) Outperform Human Experts
  - AI lets us create systems that are (potentially) more accurate and predictable than the best experts

- Reduction in Human Risk
  - Artificial Intelligence can be utilized in areas that are hazardous to human beings
  - Using AI increases the range of problems that we can solve and increases the decisions that we can (ethically) make.

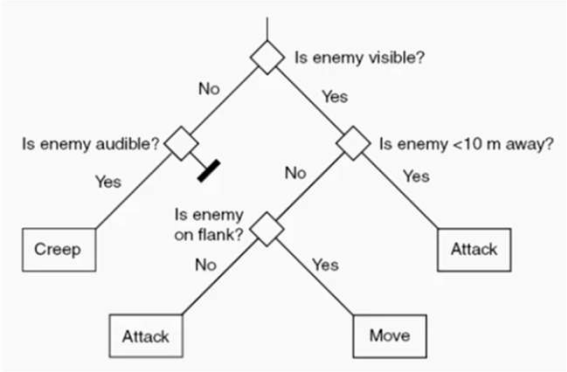
- Consistent Performance
  - AI systems do not need breaks/crew rest, get bored, *etc.*
  - AI systems do not momentarily forget, or overlook a step

Additional Benefits of AI

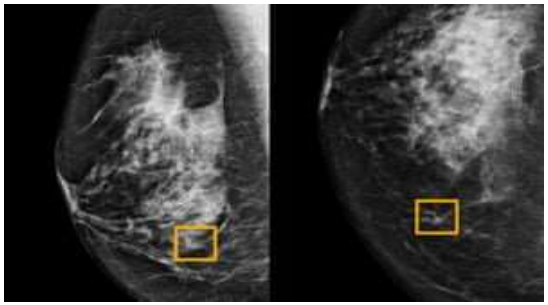
- Reduces the Need for Humans Performing Trivial Tasks
  - AI can perform many of the repetitive tasks that we encounter throughout our day
  - By performing these tasks on our behalf, AI lets us focus on more meaningful or creative tasks



- Faster Decision-Making
  - The average reaction time for humans is between 0.15 – 0.25 seconds
  - This is an eternity for a computer!



- Helps us Discover Hidden Insights
  - AI can be more adept at identifying obscure patterns in data
  - These insights can increase our understanding of a problem or task

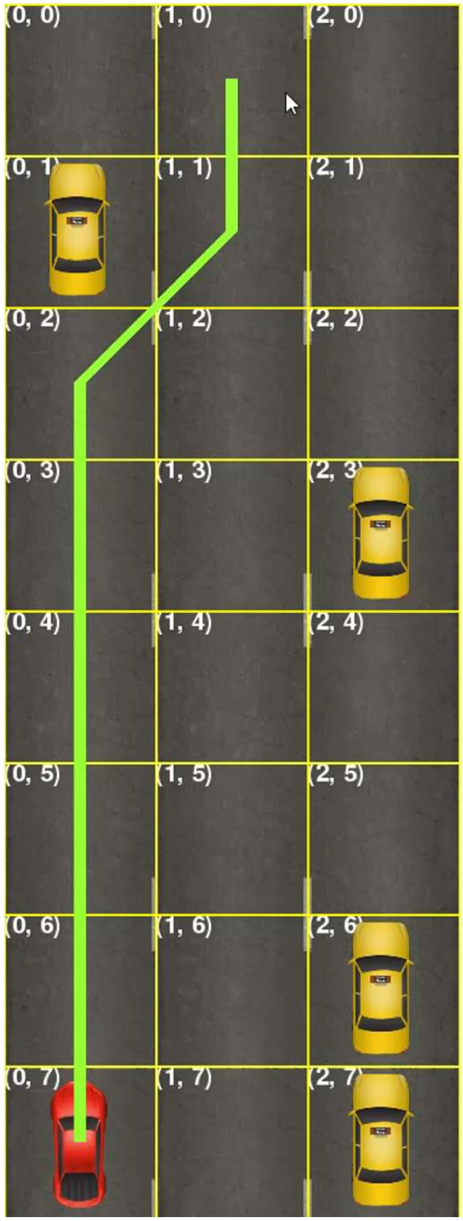


Limitations of (Current) AI

- Although we have made significant advances, there are many limitations with AI systems as we currently know them:
  - Narrow Focus (Weak. vs. Strong AI)
  - High Cost of Creation
  - Hidden Bias / Misguided Assumptions
  - Ambiguous Responsibility
  - Lack of Intelligibility

Narrow Focus

- Review: Weak vs. Strong AI
  - Weak AI – (Specific) Operates within a pre-defined range of functions
  - Strong AI – (General) Can think and accomplish complex tasks on their own
- Current Artificial Intelligence systems are designed with <u>a specific task in mind</u>.
  - Driving a car
  - Diagnosing an illness
  - Recognizing human handwriting
- AIs do not generalize to other problem domains:
  - Deep Blue can beat a grand master at chess, but can't play checkers
  - Our "Spy Hunter" AI can play the game perfectly, but it wouldn't know what to do in "Pong"



High Cost of Creation

- Rule-Based AIs (*i.e.,* expert systems) are challenging to create
  - They require human beings to painstakingly tell them what to do <u>in every possible situation.</u>
  - This assumes that human experts' understanding of the problem is comprehensive (it rarely is).
- Machine Learning algorithms have their own issues
  - Machine learning is **addicted to data**
    - It can take hundreds, thousands, or even millions of training instances to create a model
    - Getting and labeling the data is nontrivial.
      - We must make sure the data contains everything we want the model to learn
      - We oftentimes must resort to using the data we can get, rather than the data that we actually want
      - Omissions / gaps in training data can lead to other problems (see next slide)

Hidden Bias / Misguided Assumptions

- It is a misconception that Artificial Intelligence systems are bias-free
- ALL AI systems inherit biases from their creator(s)
  - The "Spy Hunter" AI we created always tries to steer in the leftmost lane because we programmed it with that strategy!
  - A model may learn unintentional patterns/correlations if the dataset it was trained with does not represent real-world conditions
- Bottom line: An AI (like a student) is as only as good as its trainer!

(a) Husky classified as wolf (b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

	Before	After
Trusted the bad model	10 out of 27	3 out of 27
Snow as a potential feature	12 out of 27	25 out of 27

Table 2: "Husky vs Wolf" experiment results.

Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos.

Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

Gender Recognition Dataset

If the dataset has a lots of lighter-skinned individuals, the trained model may not understand what to do when it encounters darker-skin (because it has never seen it before)

Ambiguous Responsibility

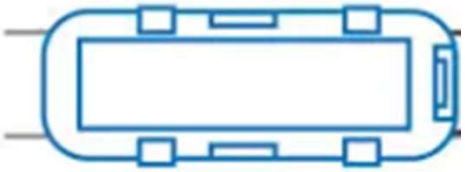
- In a traditional software system, it is generally accepted that the programmer or company is at fault for a defect.
- In an Artificial Intelligence system, it is less clear who is at fault when something goes wrong
  - Is it the programmer's fault for developing the machine learning algorithm?
  - Is it the trainer's fault for providing the algorithm with incomplete data, or accounting for every possible situation?
  - Can we blame the AI?

Lack of Interpretability

- Interpretability refers to our ability to understand **what a model is actually learning**
- Why do we care?
  - If a self-driving car suddenly swerves off the road, we need to know that it had a good reason for doing so (*e.g.,* "I was trying to avoid a person!" vs. "I saw a butterfly!")
  - Understanding the rationale behind an AI's decision can help us determine if it has learned the correct concept, and give us increased confidence in its abilities
- Some Artificial Intelligences are more intelligible than others
  - It's easy to look at a decision tree and see what the AI has "learned"
  - Many advanced AI algorithms (*e.g.,* neural networks), however, are "Black Boxes". This means that we have no way to determine what they are doing behind the scenes.

Why Does This Matter?

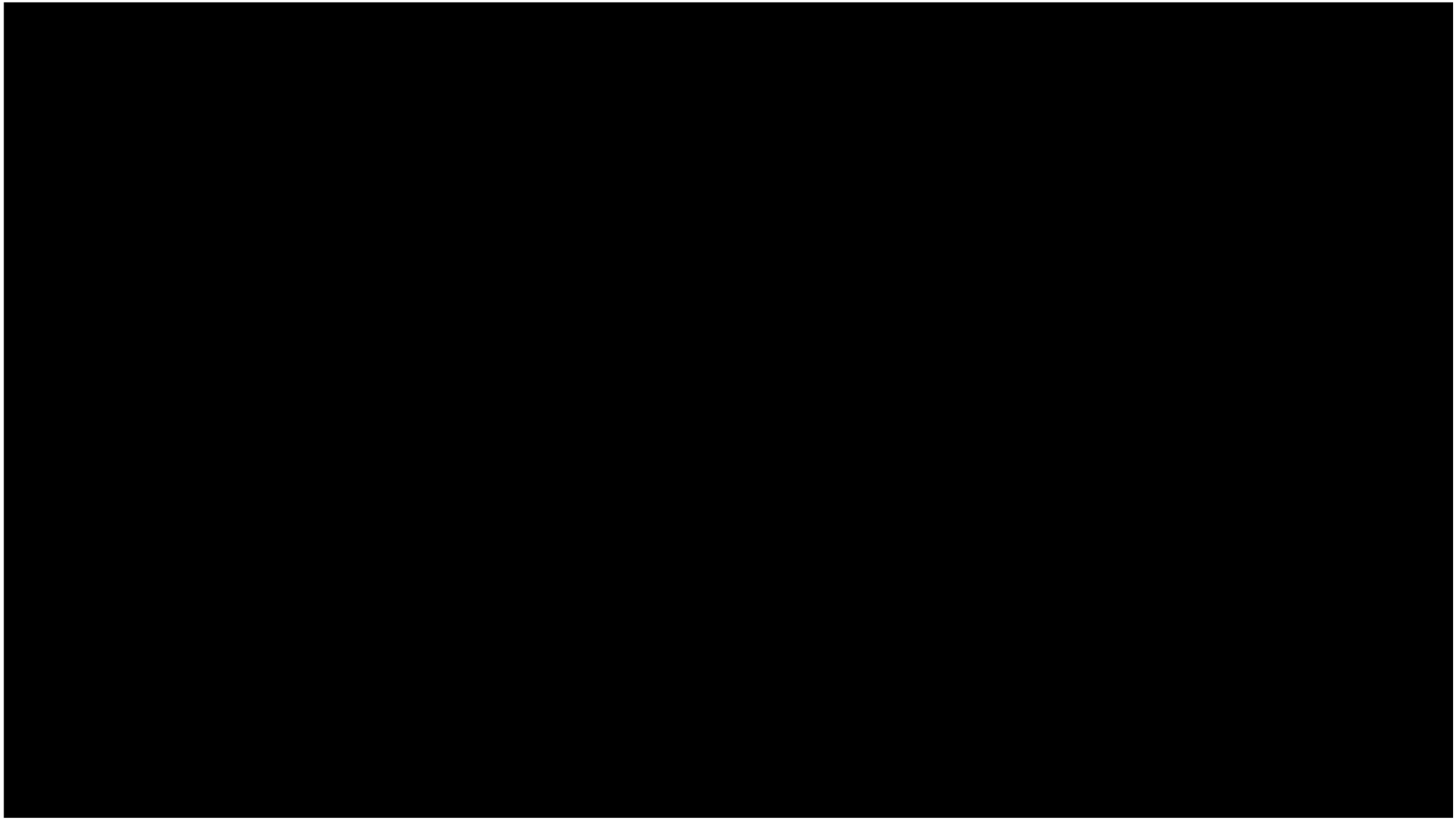
- AI is starting to be used in systems that can have a measurable impact on our lives
- The Trolley Problem



- What is the correct answer?
- Are you comfortable with an AI making this decision?
  - What model is the AI using to make this decision?
  - How was this model trained? <u>Who</u> trained it?
  - How would you feel if you were the elderly person?

Tesla Autopilot

- Are we ready for this? Do we have a choice?



This is Not Science Fiction

NEWS · 24 OCTOBER 2019 · UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

REPORT

Assessing employer intent when AI hiring tools are biased

Caitlin Chin · Friday, December 13, 2019

Who Should Learn About AI?

- As AI becomes more pervasive, there will be a growing need for the general populace to understand how it impacts their lives.

- Problem(s)
  - What do they Need to Know?: AI has traditionally been an upper-level course. What topics should an AI curriculum cover?
  - Complexity: AI can quickly become math heavy, making it inaccessible for many students.

- Our Challenge: "How can we teach Artificial Intelligence and Machine Learning (AI/ML) to a wider student audience?"
  - Freshmen (*i.e.,* First-Year) Students
  - Non-Computer Science Majors

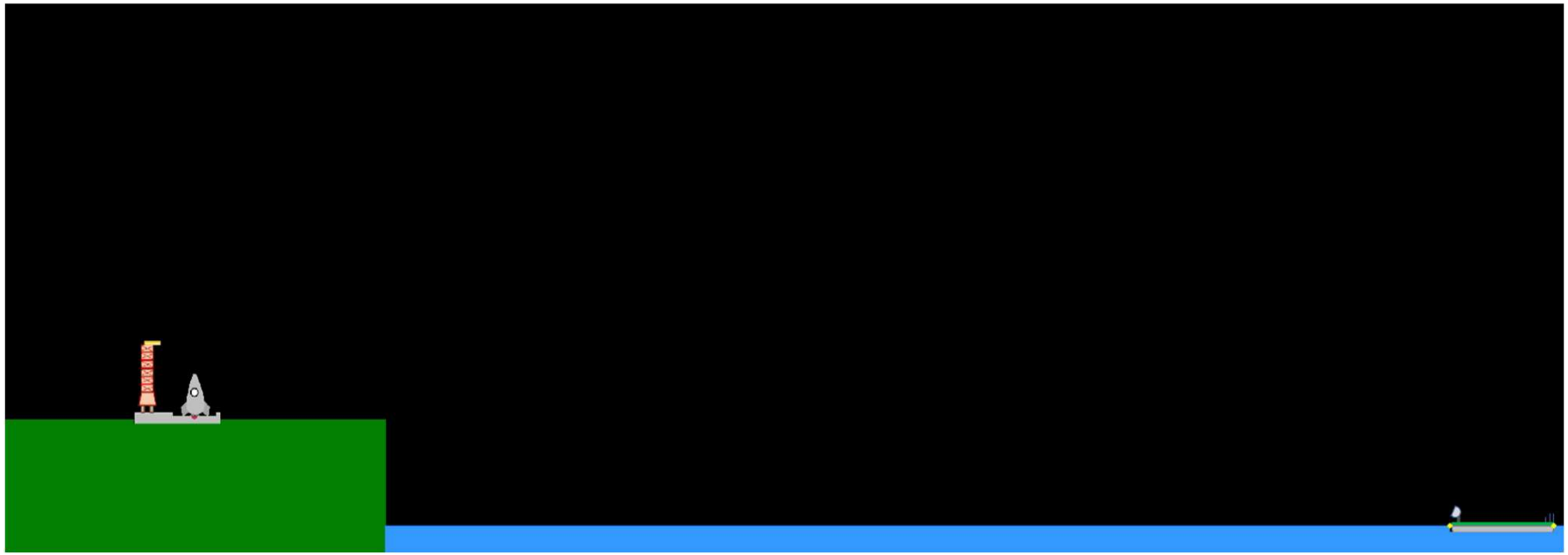
Creating an Introductory AI Curriculum

Comp Systems	Algorithmic Reasoning (i.e., Programming)	**AI / ML**	Cybersecurity	Admin / Exams
4 Lessons	23 Lessons	**4 Lessons**	5 Lessons	6 Lessons

- Lesson 1: What is AI?
  - Introduces Cadets to AI Terminology and History
  - Students apply the Turing Test to AI applications to see if they are "intelligent"
- Lesson 2: Sense-Think-Act
  - Shows students how to build rule-based agents to play games
- Lesson 3: Intro to Machine Learning
  - Walks students through each step of the machine learning process
  - Students label a dataset, train it, and evaluate the model
- Lesson 4: AI Limitations / Ethics
  - Discusses AI's (current) limitations
  - Demonstrates how ML models can be unintentionally biased
  - Has students compare/contrast ethical frameworks of academia, industry, and the Department of Defense

Big Takeaway:
AI / ML is NOT Magic! It is a useful technology with strengths and limitations.

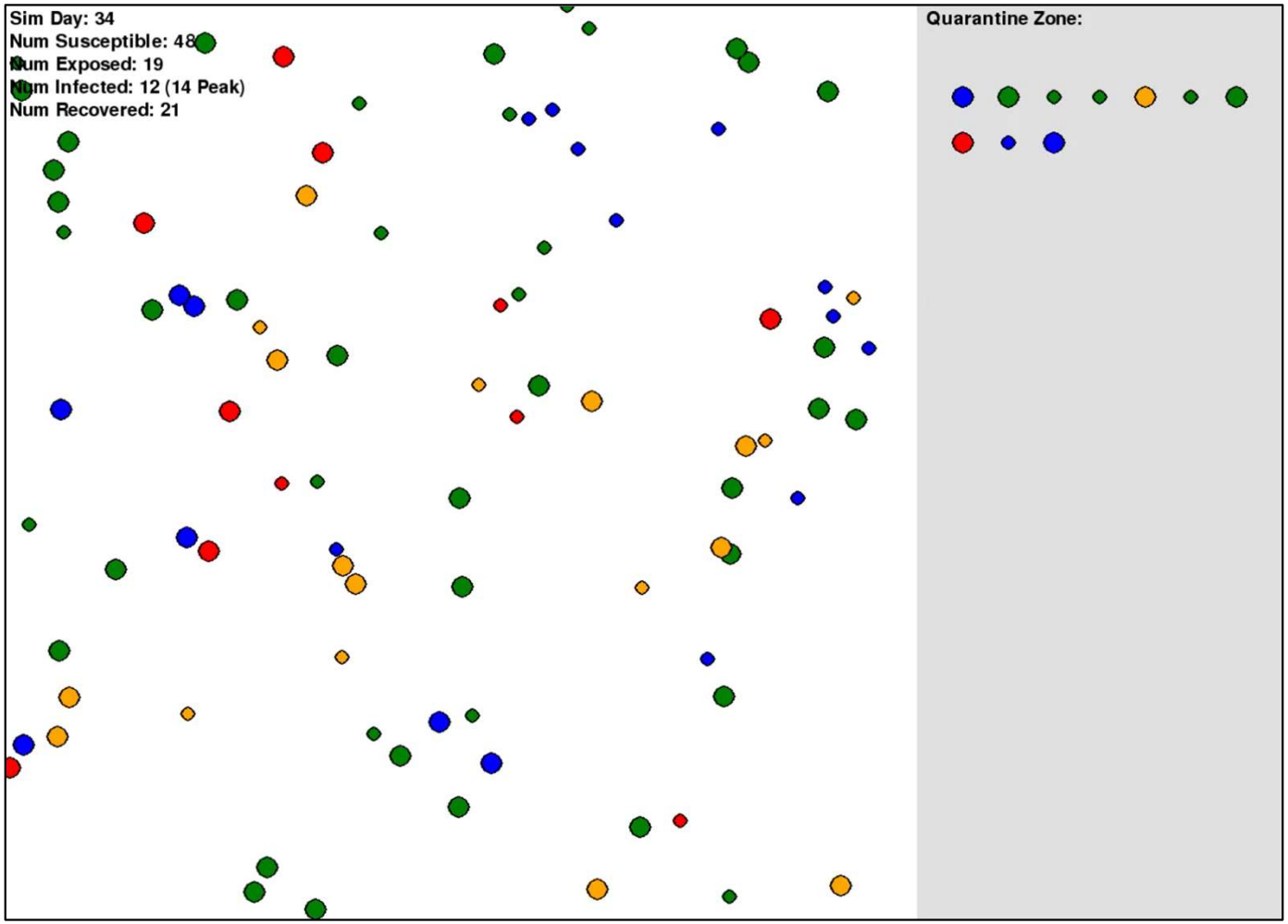
AI Block Final Project



- Spring 2020
  - Students program a “SpaceX” style rocket-landing simulation in Python
  - Students write an AI that controls the rocket
    - **Part 1:** Students write a rules-based agent to land the rocket on the boat
    - **Part 2:** Students tune the performance of their agent using a genetic algorithm library
      - Students define a fitness function, and “breed” an AI that maximizes performance
  - AIs compete against each other in an online competition

AI Block Final Project

- Fall 2020
  - Created a new “disease simulator” final project
    - Goal: Train an AI policy to minimize infections while maximizing individual freedom



Susceptible Exposed Infected Recovered

Creating an Advanced AI Curriculum

Overview

- Introduction
- What is Artificial Intelligence / Machine Learning?
- How Do They Work?
- What Are the Limitations / Implications?
- Who Should Learn About AI/ML?

Acknowledgments

- This work was sponsored by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-20-S-0003
- Thank you to the entire Department of Computer and Cyber Sciences for supporting and developing this curriculum!

Cool AI Demonstrations

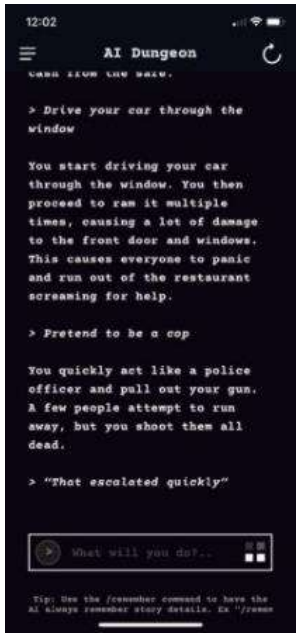
Name	Description	URL
AI Dungeon	Play a text-based choose your own adventure	https://play.aidungeon.io/
Eliza	Try out the original chatbot	http://manifestation.com/neurotoys/eliza.php3/
Aiva	Create custom music (you have to make a free account)	https://beta.aiva.ai/
Quick Draw	Play Pictionary with an AI	https://experiments.withgoogle.com/quick-draw
Sentiment Analysis	Analyze English text to extract the author's meaning and intent.	https://aidemos.microsoft.com/text-analytics
Pet Detector	Identify your pet's species from a photograph.	http://nvidia-research-mingyuliu.com/ganimal
Landscape Drawing	Create novel, photo-quality pictures using broad brushstrokes.	http://nvidia-research-mingyuliu.com/gaugan
Removing Objects	Remove objects from a photograph	https://www.nvidia.com/research/inpainting/result
Sketchpad	Draw objects WITH an AI	https://magic-sketchpad.glitch.me/
AI Duet	Play piano with an AI	https://experiments.withgoogle.com/ai/ai-duet/view/

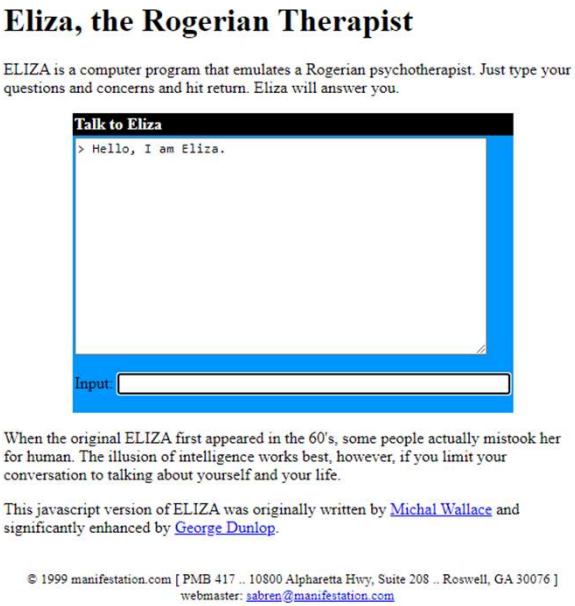
Want the Slides?

AI Block Lessons – Core Concepts

- Lesson 1: What is AI?
  - Introduces Cadets to AI Terminology and History
  - Students apply the Turing Test to AI applications to see if they are "intelligent"

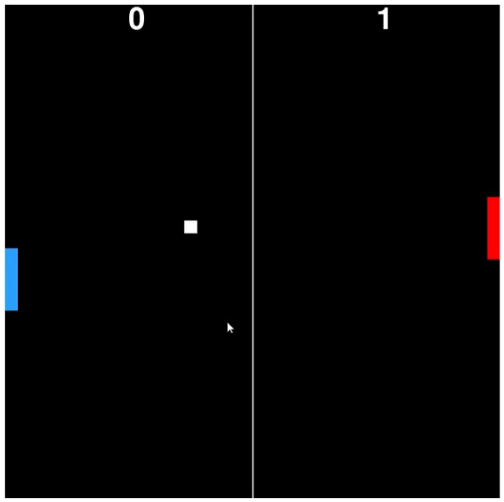
- Learning Objectives
  - Describe the characteristics of AI
  - Describe the Turing Test
  - Differentiate between weak/strong AI
  - Understand AI's history/evolution
  - Articulate the problems AI can help us solve



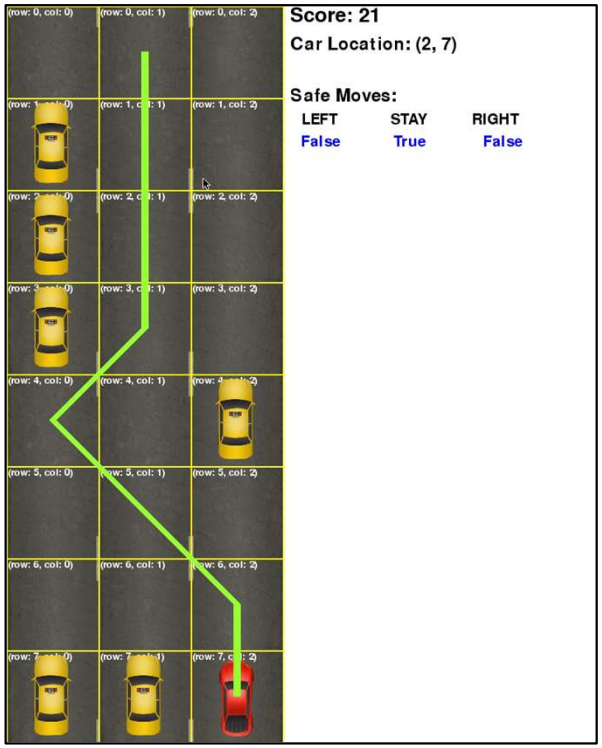


AI Block Lessons – Implementation Details

- Lesson 2: Sense-Think-Act
  - Shows students how to build rule-based agents to play games
  - Students compare/contrast agent performance when using exhaustive search vs. heuristics
  - Hands-on-activities: Pong & Spy Hunter



- Learning Objectives
  - Describe each step in the Sense-Think-Act cycle
  - Describe how decision trees, finite state machines, and search trees can be used by AIs to make decisions



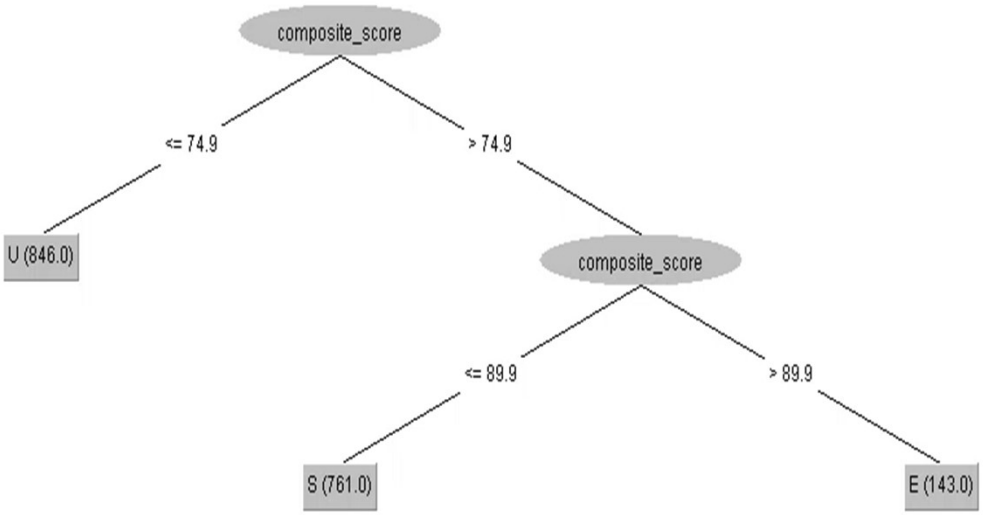
AI Block Lessons – Implementation Details

- Lesson 3: Intro to Machine Learning
  - Walks students through each step of the machine learning process
  - Students label a dataset, and train it using Weka
  - Students evaluate the trained model to look for classification errors

waist	pushup	situp	run	result
40	36	55	676	**Unsat**
42	67	25	798	**Unsat**
30	44	58	568	**Excellent**
30	39	58	729	**Sat**
37	46	57	709	**Sat**

Student Labeled Dataset (Air Force Fitness Test)

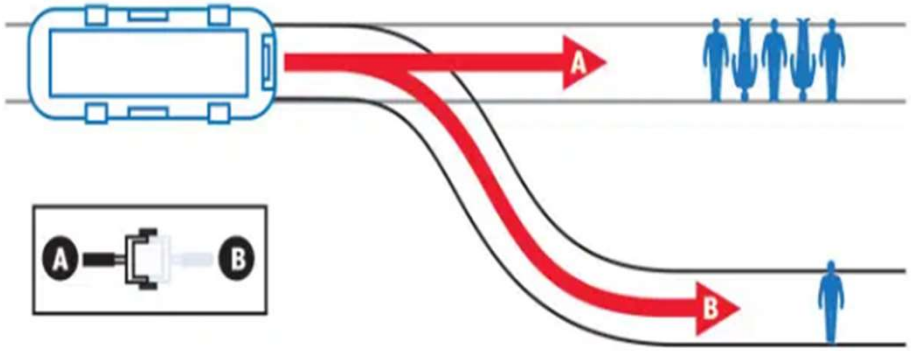
- Learning Objectives
  - Describe the goals and motivation behind Machine Learning
  - Define ML terminology (“Model”, “Instance”, “Feature”)
  - Describe and execute the steps involved in training a model



Trained Model

AI Block Lessons – Ethical Considerations

- Lesson 4: AI Limitations / Ethics
  - Discusses AI's (current) limitations
  - Demonstrates how ML models can be unintentionally biased
  - Has students compare/contrast ethical frameworks of academia, industry, and the Department of Defense



NEWS · 24 OCTOBER 2019 · UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

- Learning Objectives
  - Understand the ethical implications of using ML algorithms in high risk and/or life-threatening situations
  - Describe "Explainable AI"
  - Differentiate between AI ethical frameworks

REPORT

Assessing employer intent when AI hiring tools are biased

Caitlin Chin · Friday, December 13, 2019

Ethical Frameworks for Artificial Intelligence

- Ethics (even without AI), is a tricky subject
  - Ethical theories are controversial and not wholly agreed upon
    - Utilitarianism – Do what causes the least pain and distress
    - Egoistic – Do what is in your self-interest
    - Common Good – Do what is best for the people as a whole
  - Algorithms require precision to code, but ethics doesn't normally talk in these concrete terms
- Consequently, there is no agreed-upon ethical framework for Artificial Intelligence
- Instead, companies and organizations (*e.g.,* the DoD) are developing ethical frameworks to address their immediate needs

Academic Perspective

AI Ethical Criteria, according to Bostrom and Yudkowsky

Focus: Identify the criteria that needs to be satisfied before AI can replace humans in societal tasks.

Transparent to Inspection

AI developers should state how the AI was created (e.g., what dataset/features were provided to create the model) so that others can review and critique as needed.

Predictability

Users should, over time, be able to understand how an AI operates so that they can plan their own actions accordingly.

Robust Against Manipulation

Models should be designed with an eye towards defeating adversaries that may try to exploit or counteract it.

Responsibility

AI developers should develop systems with the understanding of who will be to blame if and when something goes wrong.

Industry Perspective

AI Ethical Criteria, according to Google

Focus: Identify where and when the company will develop and deploy AI technologies

Be Socially Beneficial

"we will take into account a broad range of social and economic factors, and will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides"

Avoid Creating or Reinforcing Unfair Bias

"We will seek to avoid unjust impacts on people"

Be Built and Tested for Safety

"We will continue to develop and apply strong safety and security practices to avoid unintended results that create risks of harm."

Be Accountable to People

"We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal."

Incorporate Privacy Design Principles

"We will give opportunity for notice and consent, encourage architectures with privacy safeguards, and provide appropriate transparency and control over the use of data"

Uphold High Standards of Scientific Excellence

"we will responsibly share AI knowledge by publishing educational materials, best practices, and research that enable more people to develop useful AI applications."

Be Made for Use

"We will work to limit potentially harmful or abusive applications"

Technologies That Cause Harm

"Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks"

Technologies that Gather Info

Google will not "gather or use information for surveillance violating internationally accepted norms"

Technologies that Violate Human Rights

Google will not design or deploy AI "whose purpose contravenes widely accepted principles of international law and human rights"

Department of Defense Perspective

AI Principles, according to the Defense Innovation Board

Focus: Define how AI is to be developed and utilized in a warfighting capacity.

Responsible

Human beings should exercise appropriate levels of judgment and remain responsible for the development, deployment, use, and outcomes of DoD AI systems.

Equitable

DoD should take deliberate steps to avoid unintended bias in the development and deployment of combat or non-combat AI systems that would inadvertently cause harm to persons.

Traceable

DoD's AI engineering discipline should be sufficiently advanced such that technical experts possess an appropriate understanding of the technology, development processes, and operational methods of its AI systems, including transparent and auditable methodologies, data sources, and design procedure and documentation.

Reliable

DoD AI systems should have an explicit, well-defined domain of use, and the safety, security, and robustness of such systems should be tested and assured across their entire life cycle within that domain of use.

Governable

DoD AI systems should be designed and engineered to fulfill their intended function while possessing the ability to detect and avoid unintended harm or disruption, **and for human or automated disengagement or deactivation of deployed systems that demonstrate unintended escalatory or other behavior**

Comparing and Contrasting Frameworks

Academia	Industry	DoD
Transparency	Be Socially Beneficial	Responsible
Predictability	Avoid Creating or Reinforcing Unfair Bias	Equitable
Robust Against Manipulation	Be Built / Tested for Safety	Traceable
Responsibility	Be Accountable	Reliable
	Incorporate Privacy Design Principles	Governable
	Uphold High Standards of Scientific Excellence	
	Be Made Available	

AI Task: Classification

Task Description

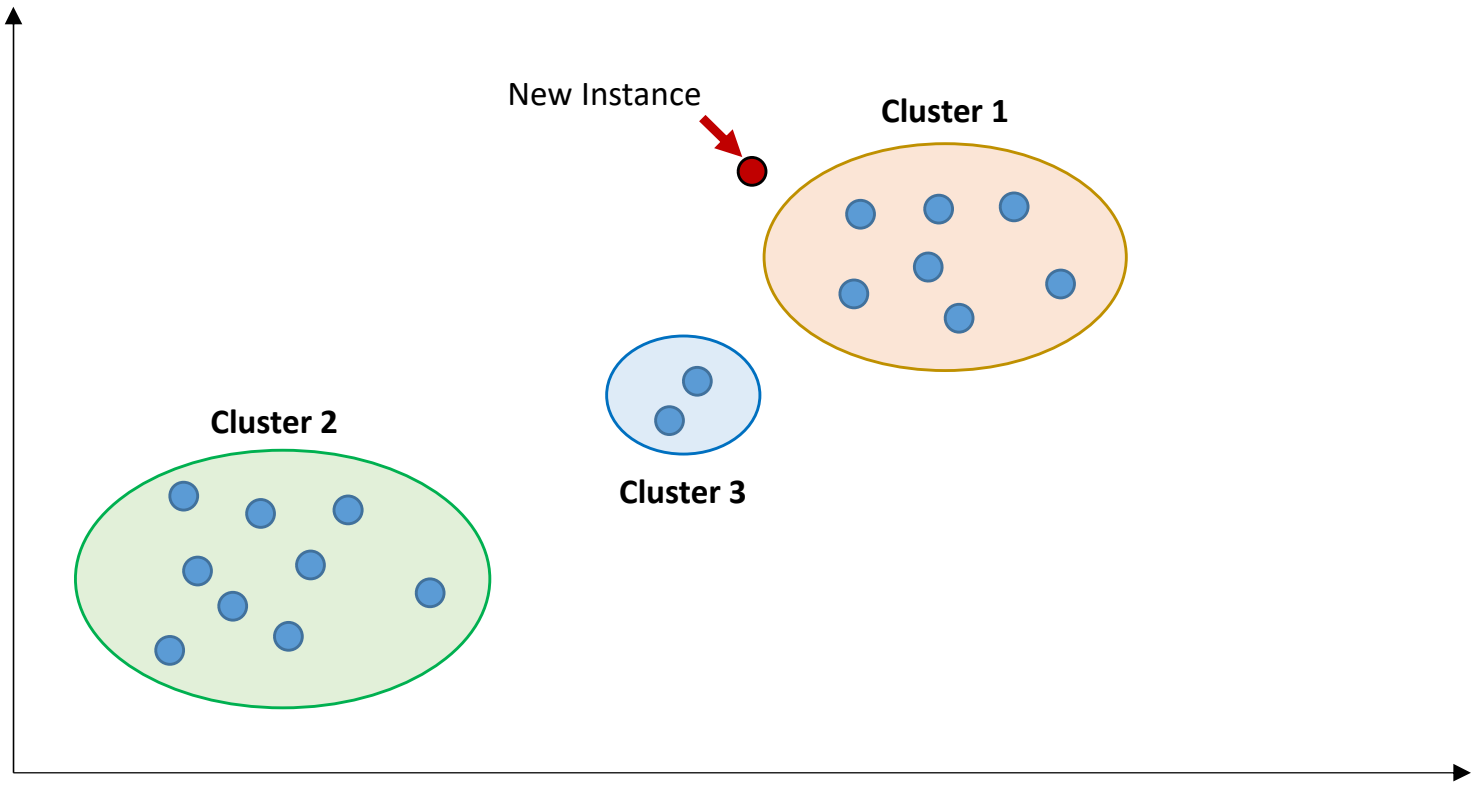
Using a set of training examples as a guide, categorize new inputs as belonging to one or more categories.

Distinguishing Between a Dog and “Not a Dog”

AI Task: Clustering

Task Description

Determine how to categorize or group data based on their common characteristics (*i.e.,* features)



AI Task: Optimization

Task Description

Given a function, determine what input(s) will result in the maximum / minimum value

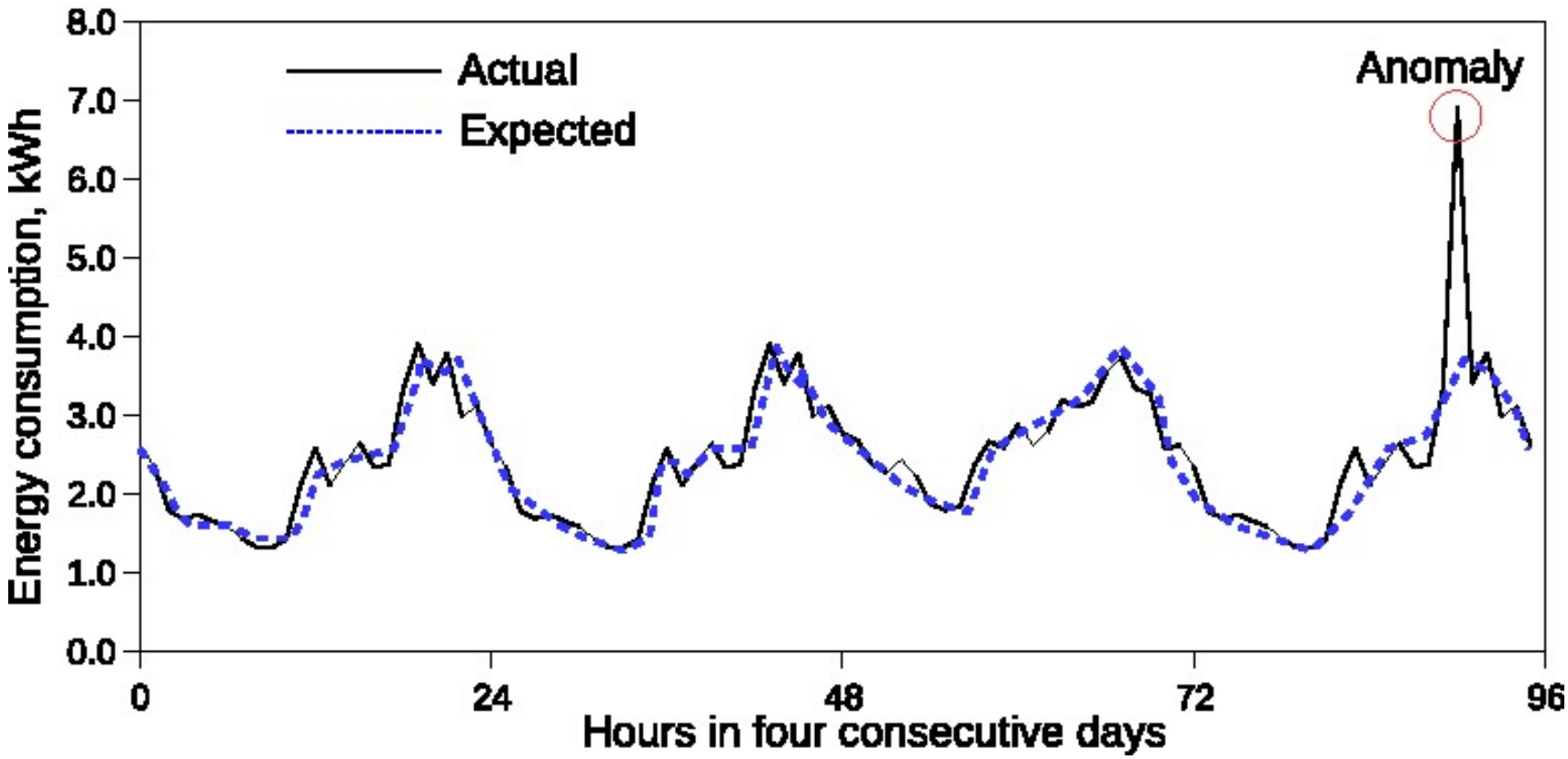
Traveling Salesman Problem

Ex: Find the Shortest Path to Visit All the Capital Cities in the Continental US

AI Task: Anomaly Detection

Task Description

Given a set of training inputs, determine if a new input is "out of the ordinary."



AI Task: Ranking / Recommendations

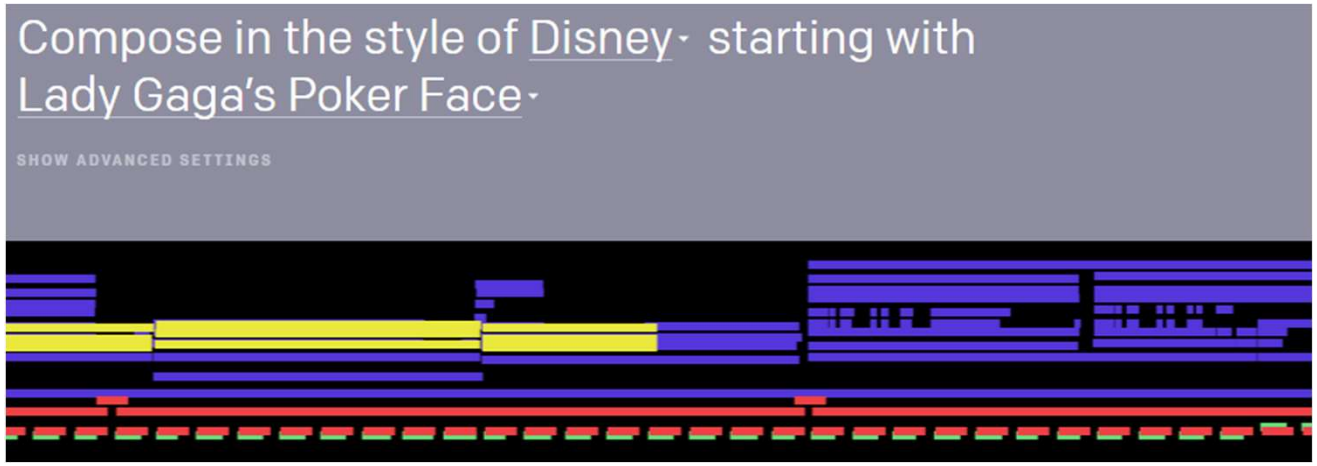
Task Description

Given an input (e.g., a search query, a list of previous actions), determine what action is most likely to occur next.

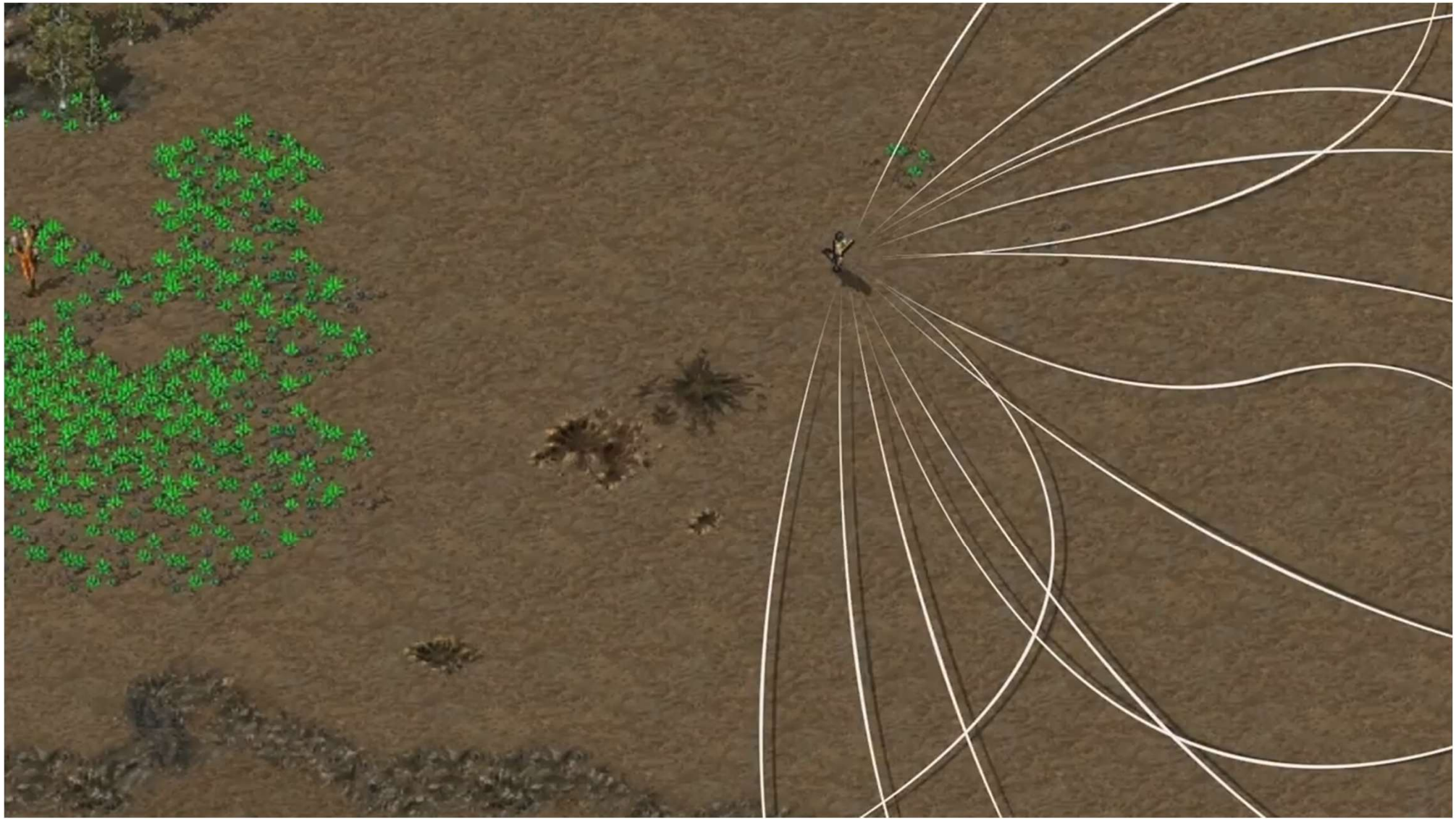
AI Task: Data Generation

Task Description

Produce novel data in order to achieve a specified goal.



Video Game Pathfinding



Open AI Project

Project Maven

- Using AI to identify targets from aerial drone footage.