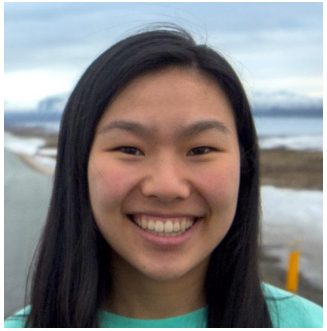
Impact of AI Decal: *Welcome!*

Sign In: https://tinyurl.com/impactsp19q1
Start: 5:10

Brenton

Gokul



Angelina

Pranav

Sajel

Surya

What is this class about?

- ➔ Educate and provide a platform for discussion of emerging issues surrounding the integration of AI into our daily lives.
- ➔ We will touch upon economic, social, and ethical concerns.
- ➔ We want to emphasize that AI is fundamentally about building tools for people to use
- ➔ This is a **discussion-based class**
  - ◆ The more you put in the more you will get out

Operating Procedure

- ➔ 10 min quiz at the beginning of class
  - These will cover previous lecture content and reading
- ➔ 60 min lecture to introduce the topic and familiarize students with relevant issues
- ➔ 35 min activity/game and discussion

Forms of Communication

- ➔ Weekly email updates starting next week
- ➔ Readings will be posted to bCourses
- ➔ Please email both of us:
  - gokul.swamy@berkeley.edu
  - brentonlongchu@berkeley.edu
- ➔ Please don’t email Anca :)

Homework & Grading

- ➔ Homework: Four writing assignments, students need only complete three writing assignments. Each writing assignment should be a short essay, two pages double-spaced. The writing assignments should be a reflection of the previous few weeks of content, relating two topics
  - ◆ Prompts will be released on week 3, 6, 9, and 11 and essays due the following week
- ➔ Attendance policy: One free absence, one additional absence can be made up by doing all four writing assignments
- ➔ Grading: 40% quizzes, 20% each writing assignment
- ➔ Students must have at least a 70% grade in the class and not miss any more than one (two if an additional writing assignment is completed) classes to pass the course

Getting into the class / Waitlist

- ➔ Enrollment codes will be distributed this week
  - Will be sent out through email Thursday evening
  - Last day to add/drop classes without a fee is on Friday, so please add the class immediately after you get the enrollment code
- ➔ If you decide this is not the class for you, email both of us ASAP
  - So we can let other students in
- ➔ To stay in the class / stay on the waitlist you **must** attend the first lecture
  - Good job (to some of you at least)

Overview of Topics

- ➔ Week 1: A Brief History of AI
- ➔ Week 2: A Survey of Modern ML
- ➔ Week 3: The Data Revolution
- ➔ Week 4: A World of Pure Automation?
- ➔ Week 5: The Self-Driving Cars of Today
- ➔ Week 6: Optical Illusions for Neural Networks
- ➔ Week 7: Artificially Generated Data
- ➔ Week 8: Bias in AI
- ➔ Week 9: Artificial General Intelligence
- ➔ Week 10: Human-Compatible AI
- ➔ Week 11: Research Guest Lecture
- ➔ Week 12: Industry Guest Lecture

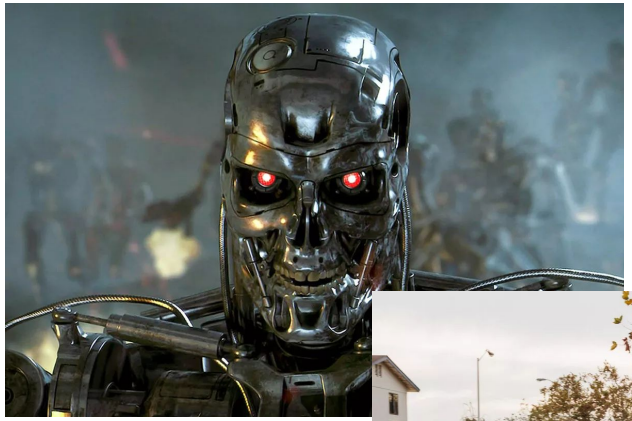
Impact of AI Decal: *A Brief History of AI*

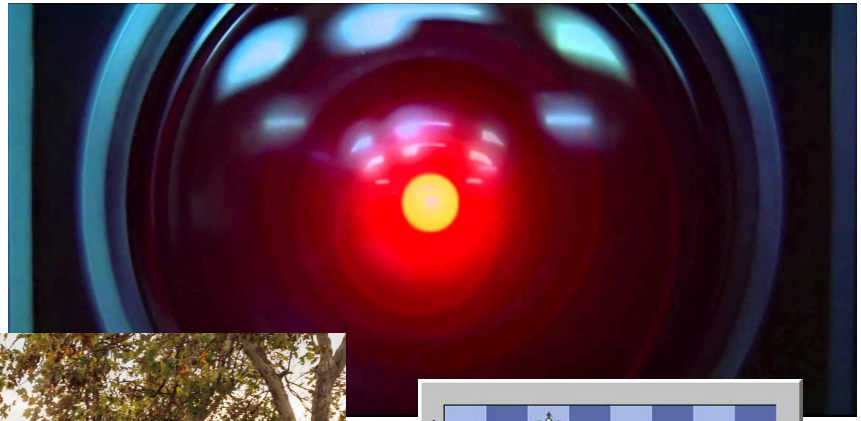
Gokul Swamy & Brenton Chu

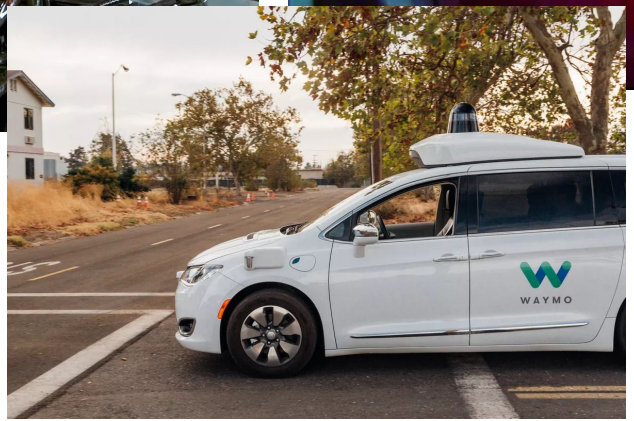
Q:

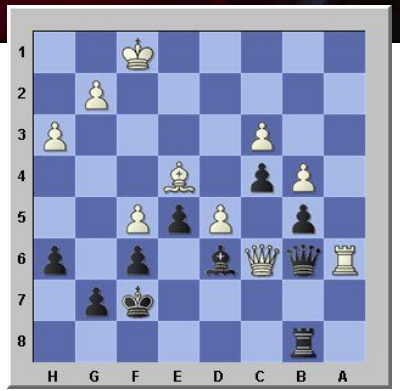
What is the first thing that comes to mind when you hear the term “artificial intelligence”?

What is “Artificial Intelligence”?









Precursors to AI: Philosophical

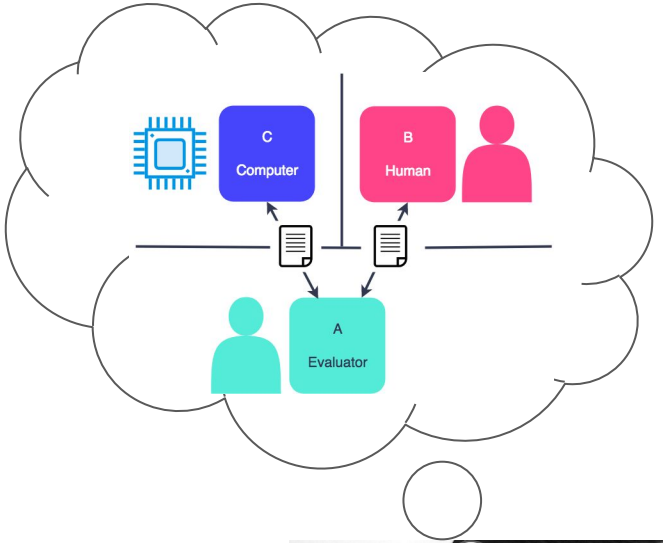
- ➔ Consequentialism: actions should be judged on their consequences
  - ◆ This has manifested in the design of modern AI systems that are trained by looking at how they do
- ➔ Utilitarianism: the morally right action is the one that produces the most “good”
  - ◆ This has manifested in the design of modern AI systems that try to maximize some quantity

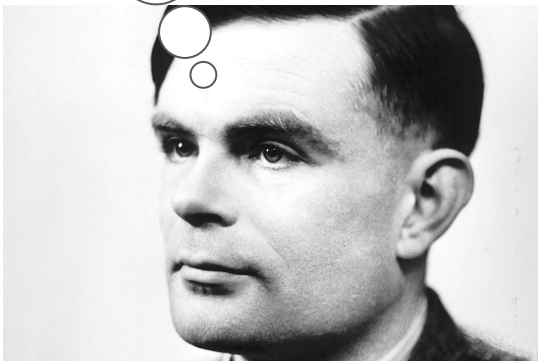
Precursors to AI: Psychological

- ➔ Behaviorism: the mind should be judged purely on how it responds to stimuli
- ➔ Cognitive Psychology: the mind processes and holds information
  - ◆ Beliefs/goals exist
  - ◆ Requires people to introspect and self-report data
  - ◆ Chomsky provided much evidence to the idea of knowledge representation by looking at how kids can construct sentences they haven't heard before
- ➔ AI currently is based on a paradigm of inputs being taken in, processed, and then results spit out

The Turing Test

- Developed in 1950
- Was considered gold standard for determining what “true AI” is
  - There were strong objections to this
- A first attempt was the chatbot ELIZA
- Some say this was beaten in 2014 by a chatbot called Eugene Goostman
  - Posed as a 13 y.o. Ukrainian boy



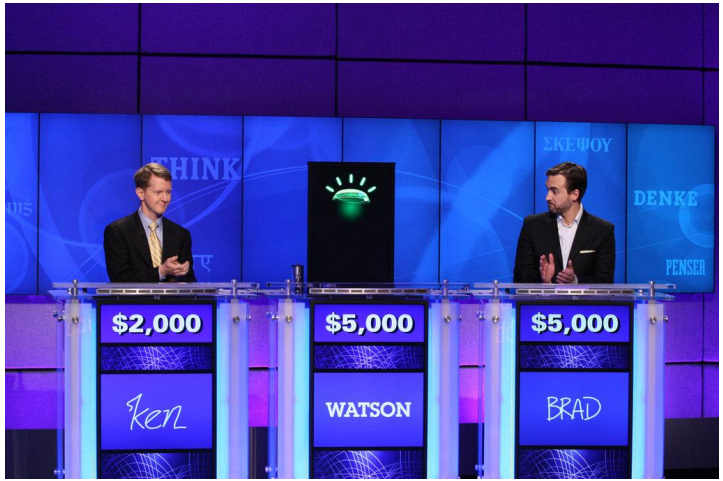


DeepBlue

- Developed by IBM to play chess at the grandmaster level
- Defeated Garry Kasparov in 1997
- Chess was considered the benchmark in human-level AI
  - Playing chess was assumed to require extensive intuition and strategy

Watson

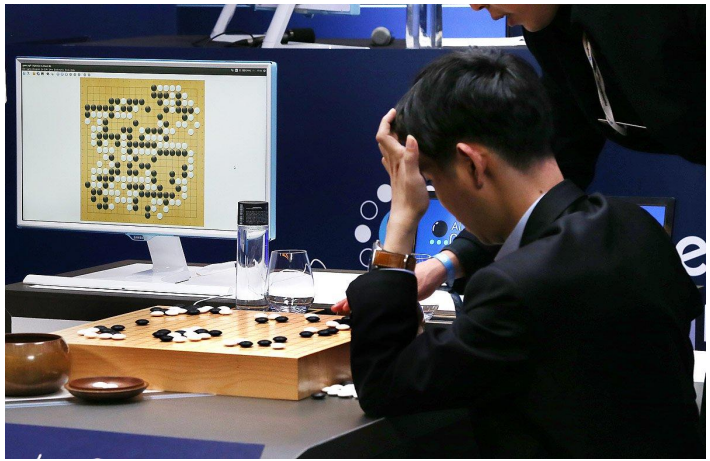
- ➔ Built by IBM to play Jeopardy against two of the show's most successful contestants
- ➔ Able to understand questions posed in natural language and respond in restricted natural language (question)
- ➔ Resoundingly beat other players but gave ridiculous answer for Final Jeopardy
- ➔ Showed that even natural language parsing was not equivalent to human influence



THINK
$2,000
Ken
$5,000
WATSON
$5,000
BRAD

AlphaGo

- ➔ Created by DeepMind to play the game of Go in 2015
- ➔ Defeated a Go champion in October 2015 and then the world's top Go player a few months later
- ➔ Only a year prior to AlphaGo, experts were claiming that it would take at least a decade for expert Go AI
  - ◆ Go is a significantly more complex game than Chess
- ➔ Continued to improve by constantly playing itself until it became completely unrivaled



OpenAI Five

- ➔ Dota 2-playing AI made by OpenAI
- ➔ Managed to defeat individuals in select matchups in 2017
- ➔ Won against 5 professional Dota 2 players in early August
- ➔ Lost to a coordinated team last week
- ➔ The algorithm is still actively being developed by OpenAI

Crystal Maiden's Freezing Field
and Viper's Nethertoxin force
the human team to scatter

AlphaStar

- ➔ Deepmind’s project after beating Go
- ➔ Livestreamed games against professional players on January 24, 2019
- ➔ Won undefeated against human opponents in prerecorded matches, but lost in a live match

DeepMind
Workers: 14/16
SC II HL
AlphaStar

Q:

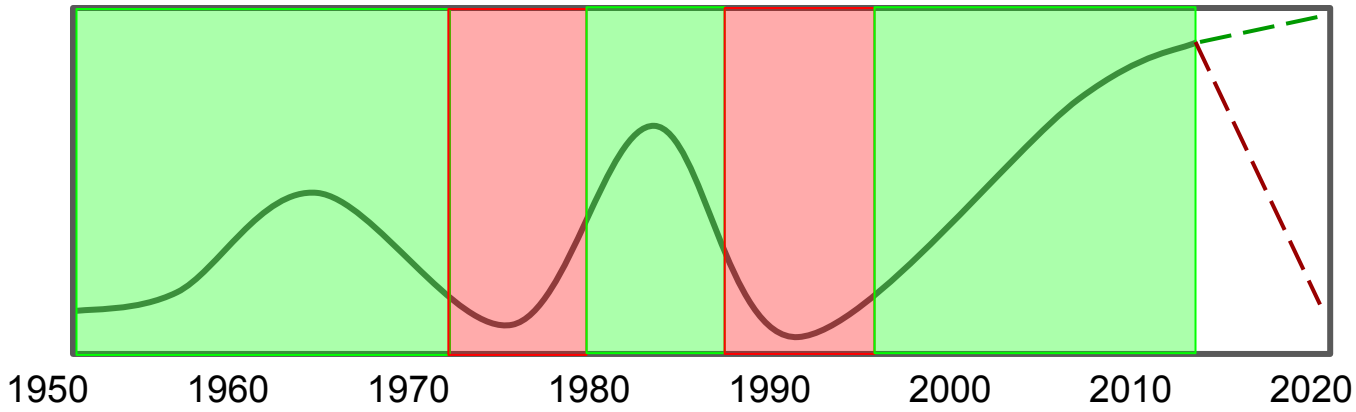
What are these systems lacking that keeps them from being “true AI”?

Is AI really here yet?

- ➔ All these examples are extremely good at one single task, but cannot do anything else
- ➔ Called ANI, “Artificial Narrow Intelligence”, or “weak AI”
- ➔ What is commonly thought as “true AI” is called AGI, “Artificial General Intelligence”, or “strong AI”
  - ◆ e.g. HAL 9000 from 2001: A Space Odyssey
- ➔ AI at the level of Skynet is called ASI, “Artificial Superintelligence”

AI Hype Cycle

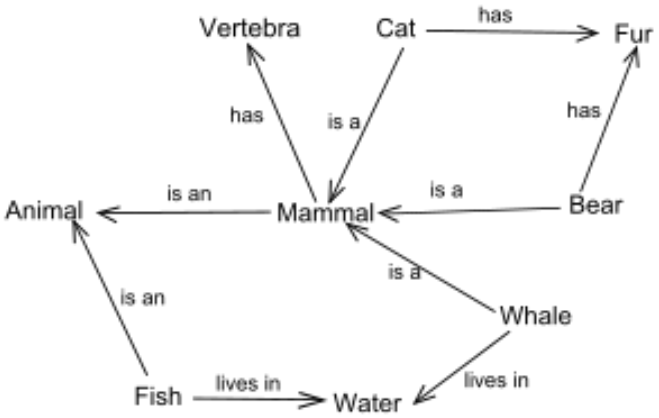
- ➔ New technique shows great initial results
- ➔ Capabilities of researched systems get over promised
- ➔ Human-level AI is said to be just around the corner
- ➔ Further experimentation produced mixed delivery of initial promises
- ➔ Fundamentally limiting flaws are revealed
- ➔ Everyone loses hope in AI and startups lose their VC funding



AI Hype Cycle

WINTER IS COMING . . . ?

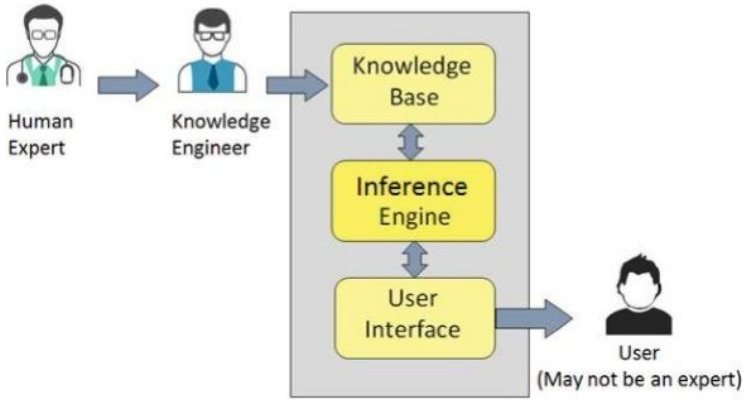
Approaches to AI: Symbolic Reasoning



- a.k.a “Good, Old-Fashioned AI”
- Dominant form of AI during the first AI boom
- Creates a symbol for each concept
- Builds a “knowledge graph” by connecting symbols with relationships
- Searched through the graph to find solutions and answers to problems

What are some advantages and disadvantages with this approach?

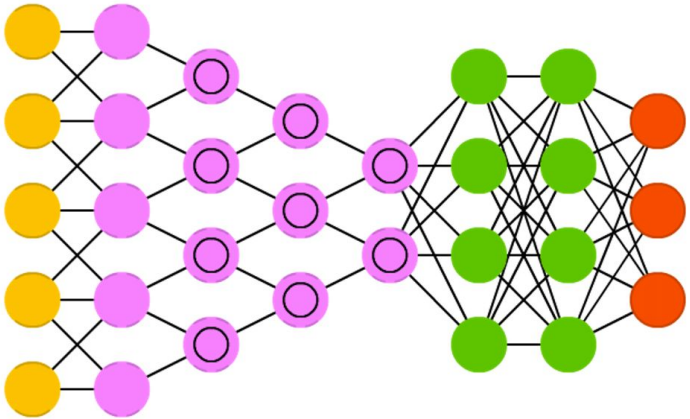
Approaches to AI: Expert Systems



- ➔ Translates knowledge from a human expert into a set of rules
- ➔ Rules are saved into a “knowledge base”
- ➔ Uses an “inference engine” to apply the rules to the new information in order to respond correctly

What are some advantages and disadvantages with this approach?

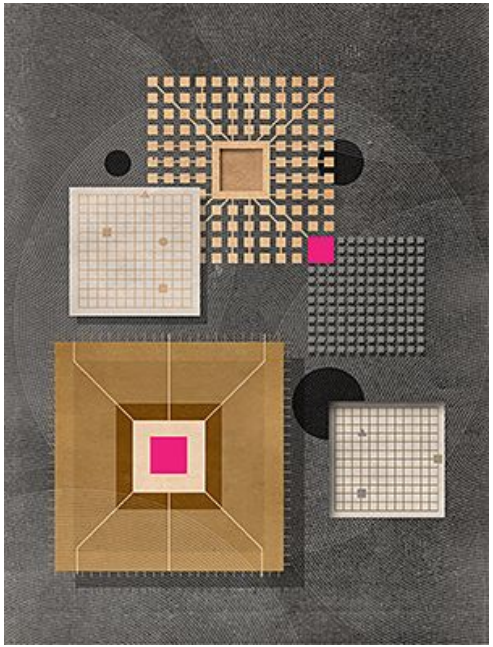
Approaches to AI: Machine Learning



- ➔ Uses data to learn the correct predictions rather than following experts
  - ◆ Each example seen changes internally saved values in a way that helps it find the right answer next time
- ➔ Neural networks are the most common form of modern ML algorithm and is inspired by how the neurons in the brain function
- ➔ Deep learning stacks many “layers” of neurons

What are some advantages and disadvantages with this approach?

Approaches to AI: Neuromorphic Computing



- ➔ Mimics the behavior of neurons at the hardware level
- ➔ Often uses new or uncommon hardware, e.g. memristors
- ➔ Still heavily in research, no practical applications as of yet

Impact of AI Decal: *Activity*

Gokul Swamy & Brenton Chu

2 Truths and a Lie

- ➔ Get into groups of 4 or 5
- ➔ Introduce yourself and give three facts about yourself, two true and one false
- ➔ The rest of the group tries to guess which of the three facts about yourself is false
- ➔ We'll reconvene and try to judge some facts about AI

Current AI Systems Can...

Write a sensible Harry Potter chapter

FALSE

CHAPTER THIRTEEN

THE HANDSOME ONE

The castle grounds snarled with a wave of magically magnified wind. The sky outside was a great black ceiling, which was full of blood. The only sounds drifting from Hagrid's hut were the disdainful shrieks of his own furniture. Magic: it was something that Harry Potter thought was very good.

Leathery sheets of rain lashed at Harry's ghost as he walked across the grounds toward the castle. Ron was standing there and doing a kind of frenzied tap dance. He saw Harry and immediately began to eat Hermione's family.

Ron's Ron shirt was just as bad as Ron himself.

"If you two can't clump happily, I'm going to get aggressive," confessed the reasonable Hermione.

"What about Ron magic?" offered Ron. To Harry, Ron was a loud, slow, and soft bird. Harry did not like to think about birds.

"Death Eaters are on top of the castle!" Ron bleated, quivering. Ron was going to be spiders. He just was. He wasn't proud of that, but it was going to be hard to not have spiders all over his body after all is said and done.

"Look," said Hermione. "Obviously there are loads of Death Eaters in the castle. Let's listen in on their meetings."

The three complete friends zapped onto the landing outside the door to the castle roof. They almost legged it, but witches are not climbing. Ron looked at the doorknob and then looked at Hermione with searing pain.

"I think it's closed," he noticed.

"*Locked,*" said Mr. Staircase, the shabby-robed ghost. They looked at the door, screaming about how closed it was and asking it to be replaced with a small orb. The password was "BEEF WOMEN," Hermione cried.

Harry, Ron, and Hermione quietly stood behind a circle of Death Eaters who looked bad.

"I think it's okay if you like me," said one Death Eater.

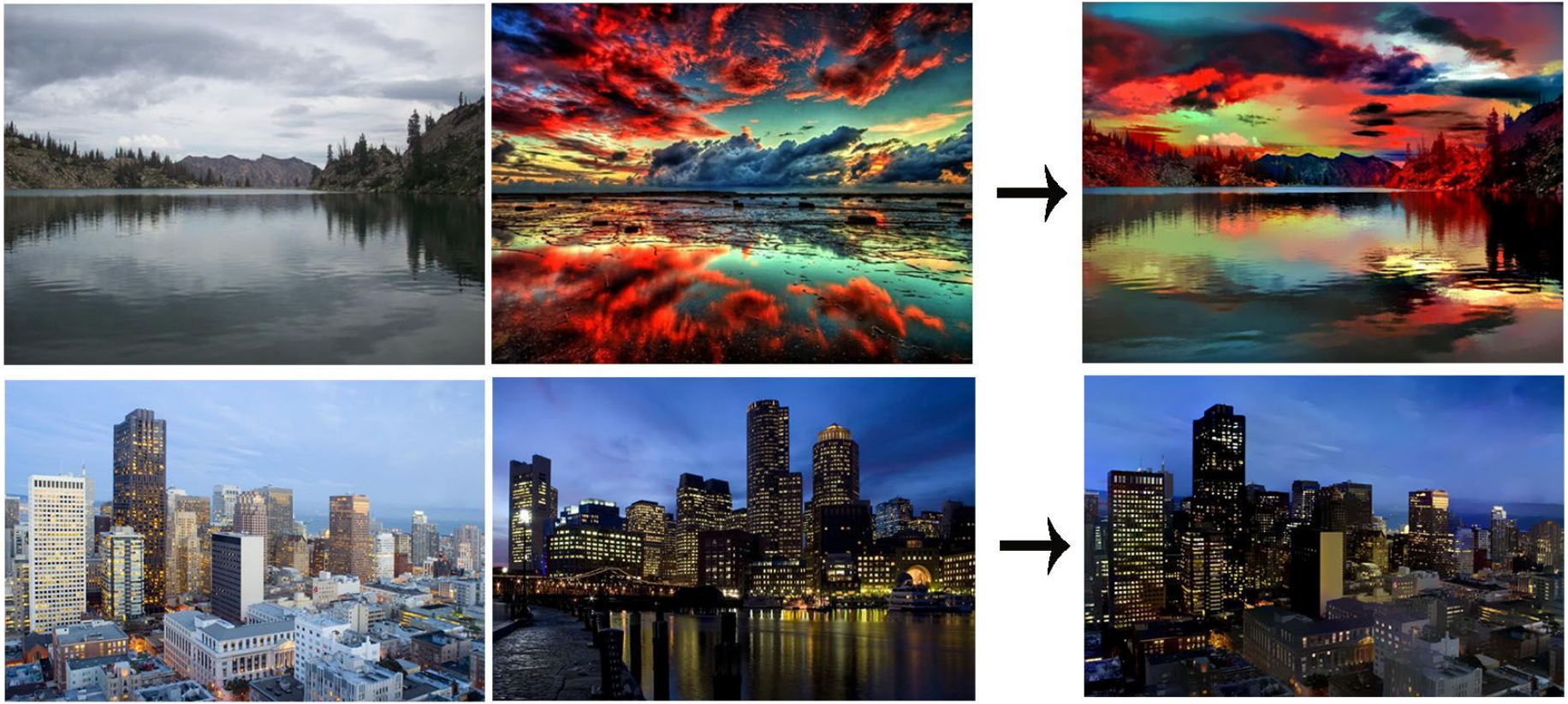
"Thank you very much," replied the other. The first Death Eater confidently leaned forward to plant a kiss on his cheek.

"Oh! Well done!" said the second as his friend stepped back again. All the other Death Eaters clapped politely. Then they all took a few minutes to go over the plan to get rid of Harry's magic.

Current AI Systems Can...

Make pictures look like they were painted by Van Gogh

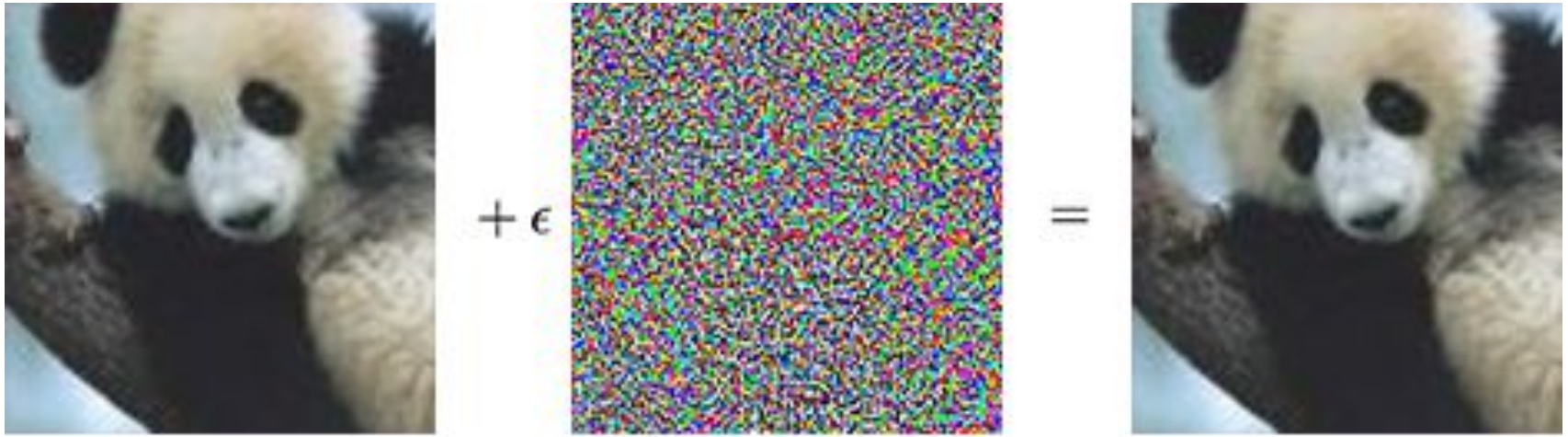
TRUE



Current AI Systems Can...

Fool both other AI systems and humans into seeing wrong things

TRUE

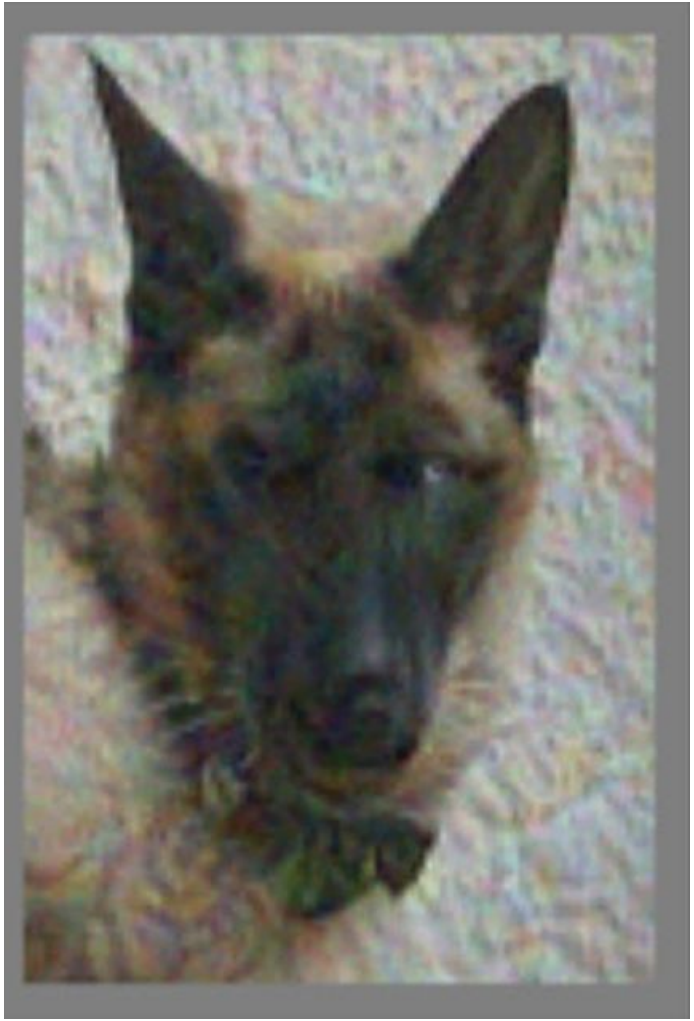


"panda"

57.7% confidence

"gibbon"

99.3% confidence



Current AI Systems Can...

Make automated medical treatments

FALSE

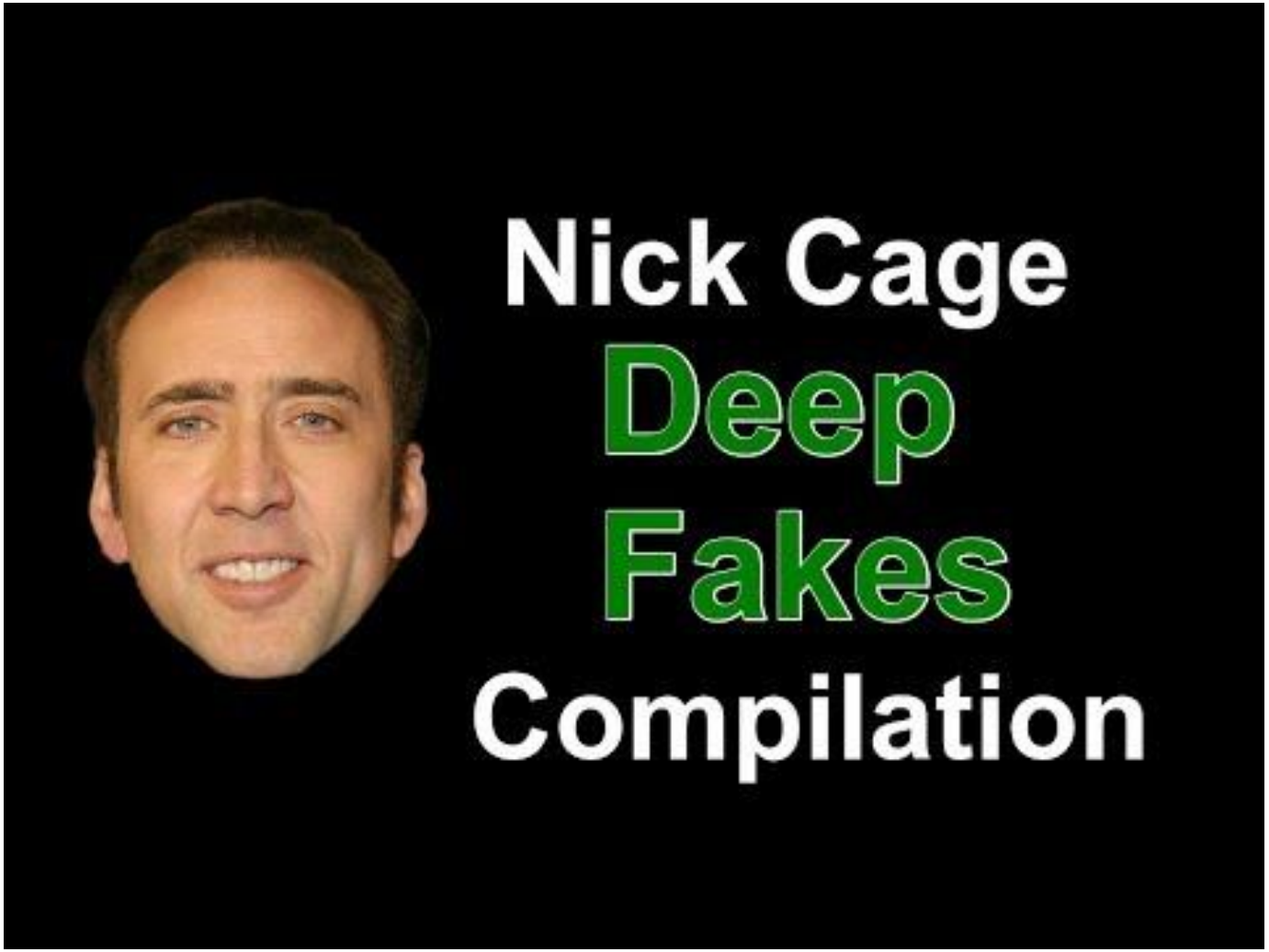
FDA approves AI-powered diagnostic that doesn’t need a doctor’s help

https://www.technologyreview.com/the-download/610853/fda-approves-first-ai-powered-diagnostic-that-doesnt-need-a-doctors-help/

Current AI Systems Can...

Make videos of people doing things they've never done

TRUE


Nick Cage
Deep
Fakes
Compilation

Current AI Systems Can...

Make AI systems better than humans can

TRUE*

Google's
Skynet
ARTIFICIAL
INTELLIGENCE

Thank you for coming to the first lecture!

Anonymous feedback:
https://tinyurl.com/impactaifeedback