

DR

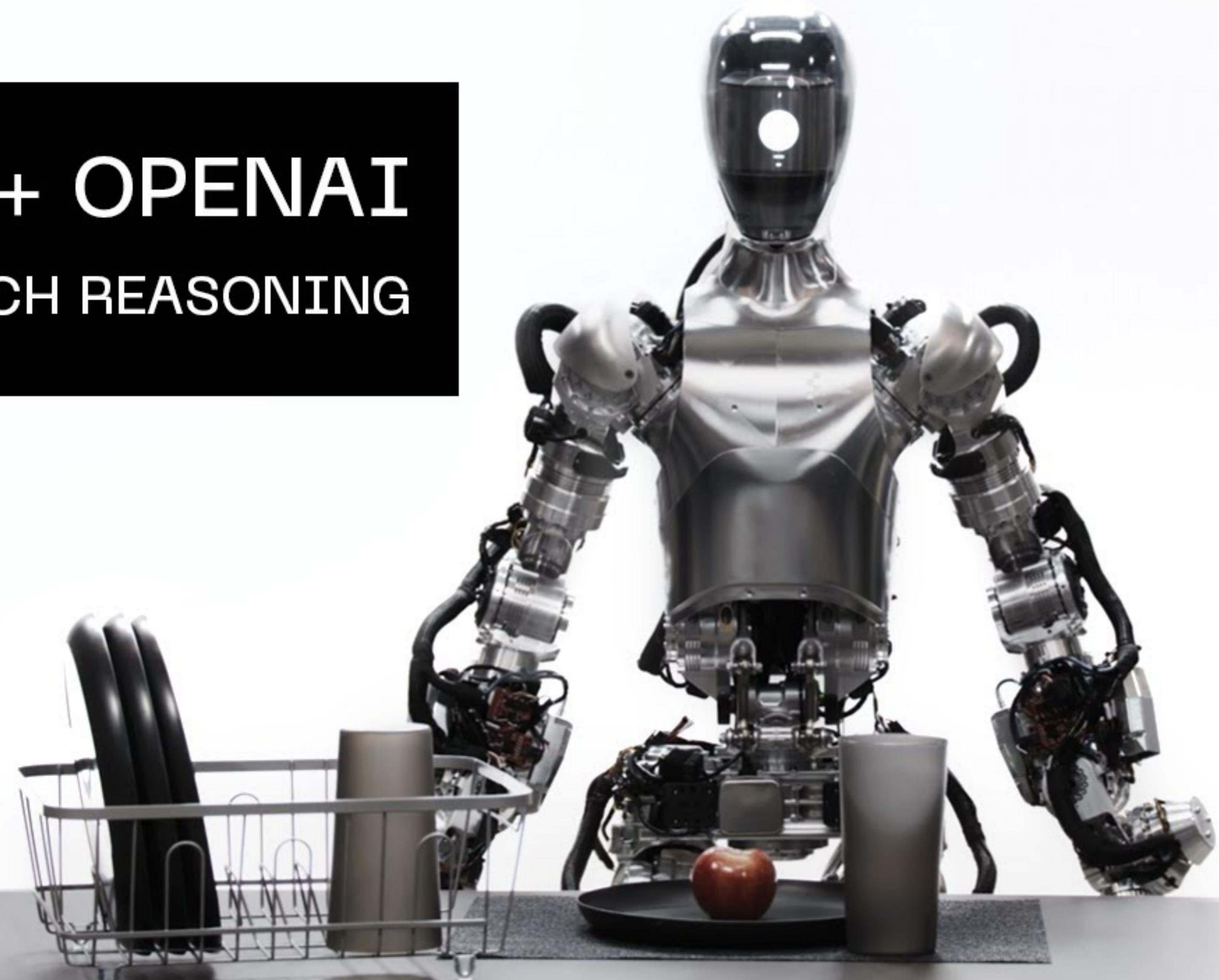
DeepRob

Fall 2024

Deep Learning for Robot Manipulation
University of Minnesota



FIGURE 01 + OPENAI SPEECH-TO-SPEECH REASONING





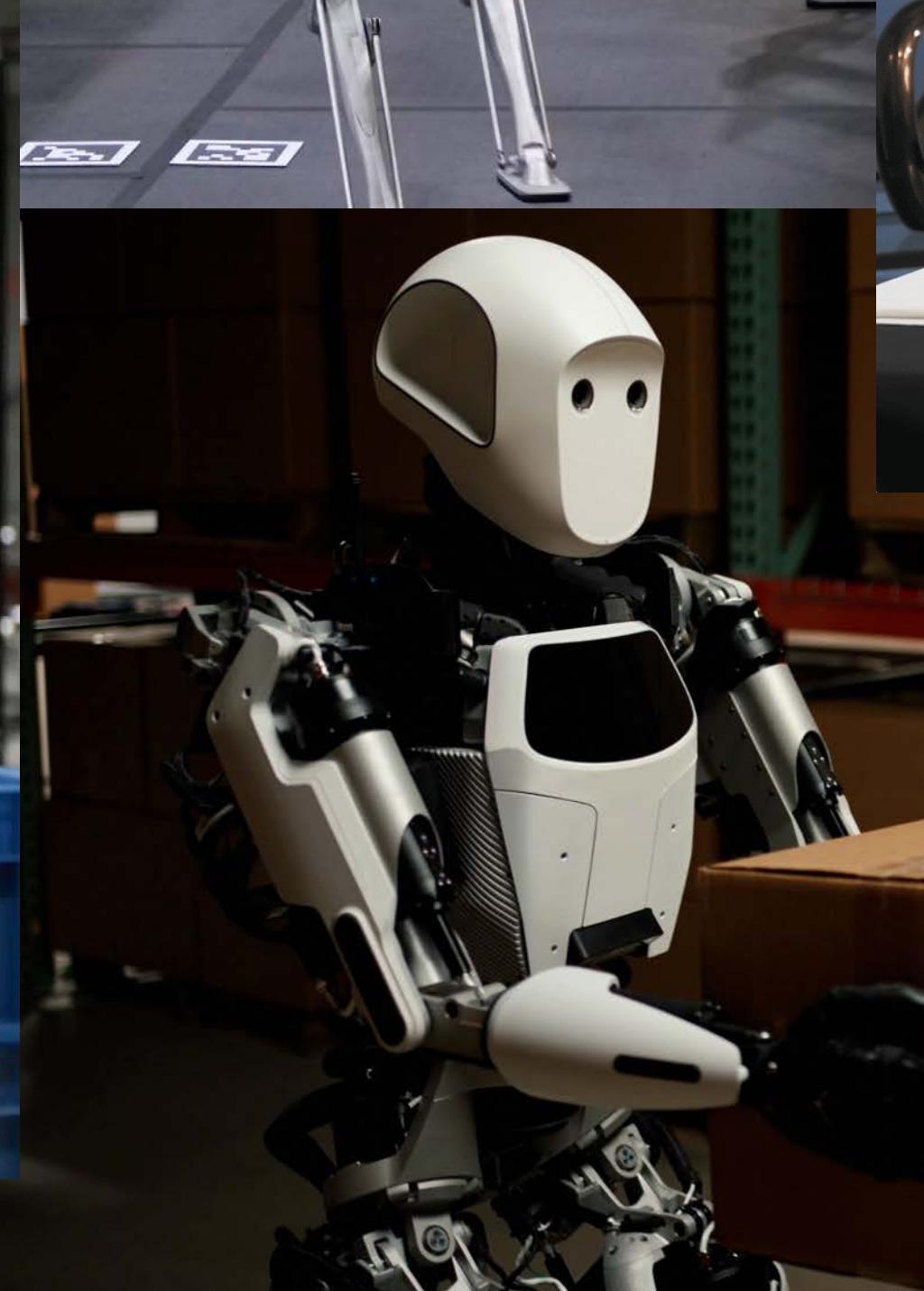




Tesla Optimus



Agility Robotics

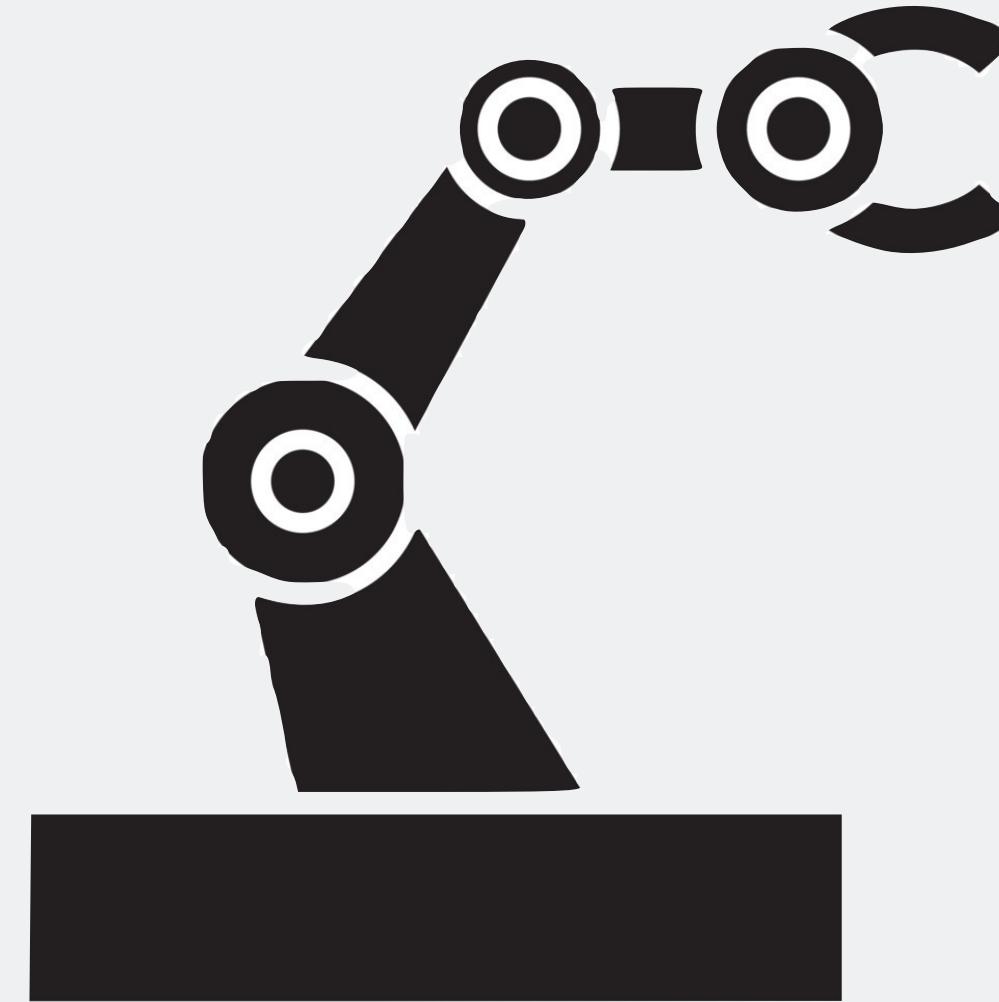
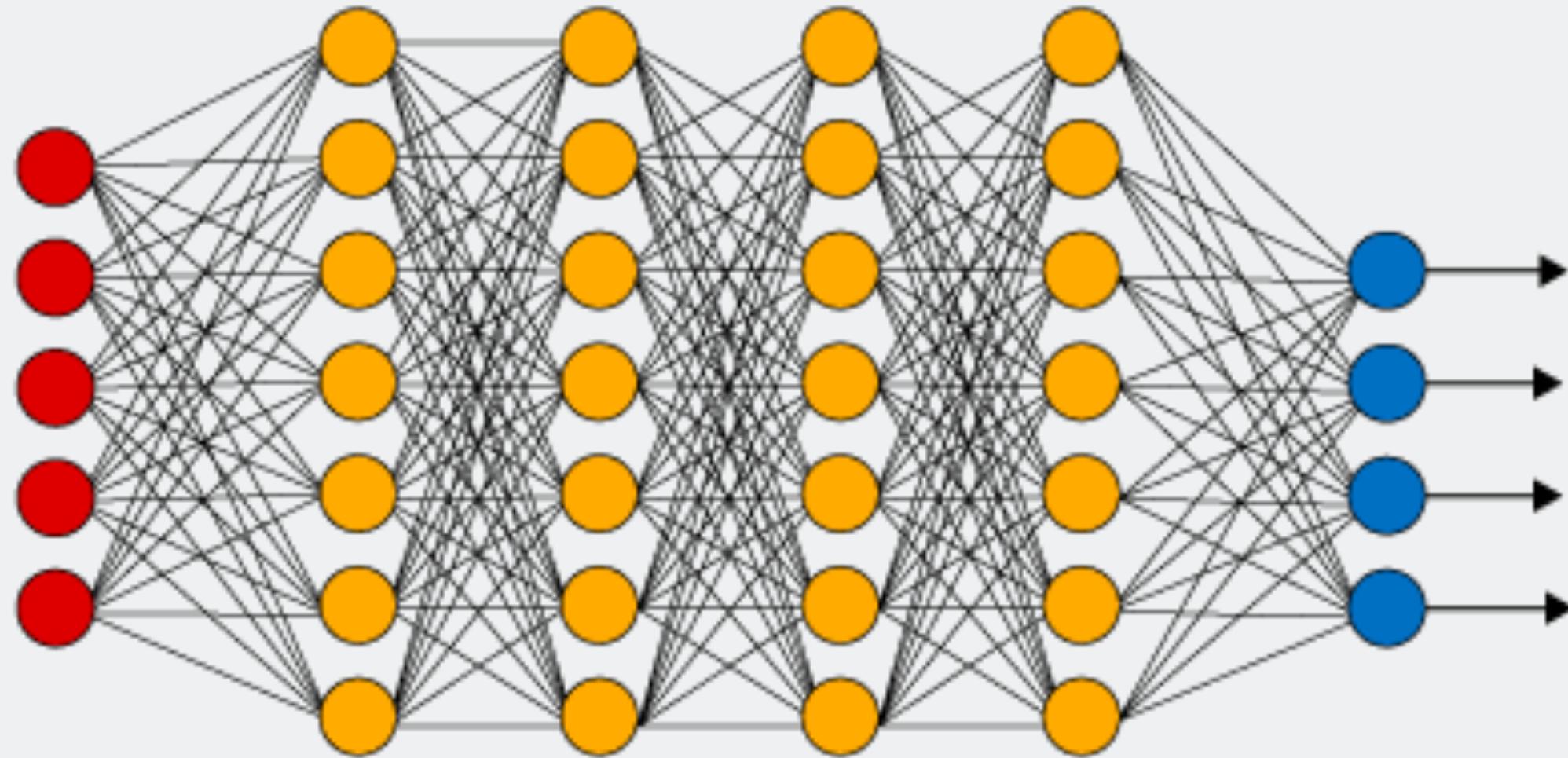


Sanctuary AI



Figure 01

Deep Learning × Robot Manipulation



But what's the key
ingredient???

Training data



Tesla



Nvidia

from human teleoperation!!!

Teleoperation

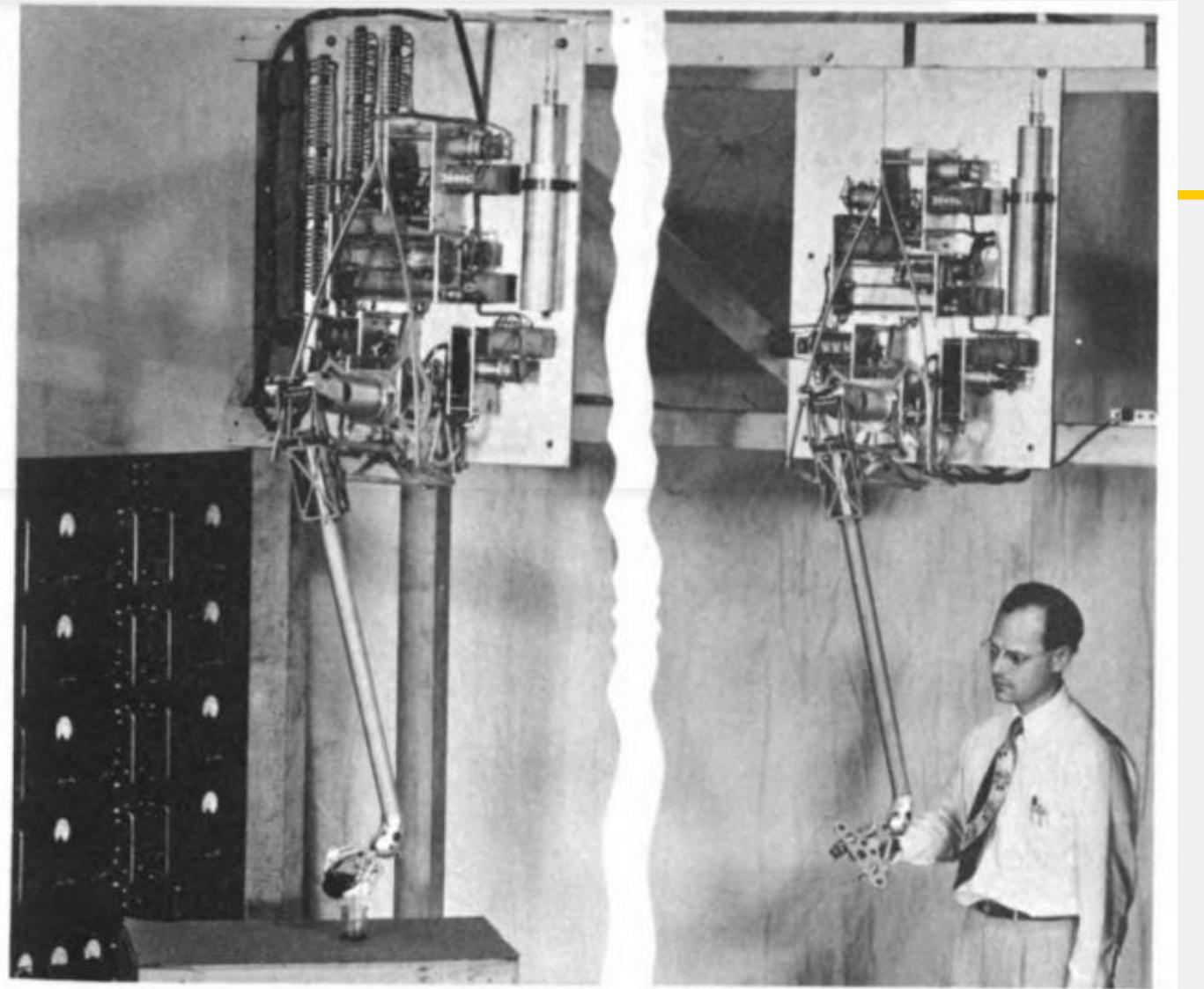


FIGURE 6.—The ANL Model E1 electric master slave. Used only for experimental purposes, this bilateral manipulator was developed in 1954. (Courtesy of Argonne National Laboratory.)



da Vinci robot

Teleoperation

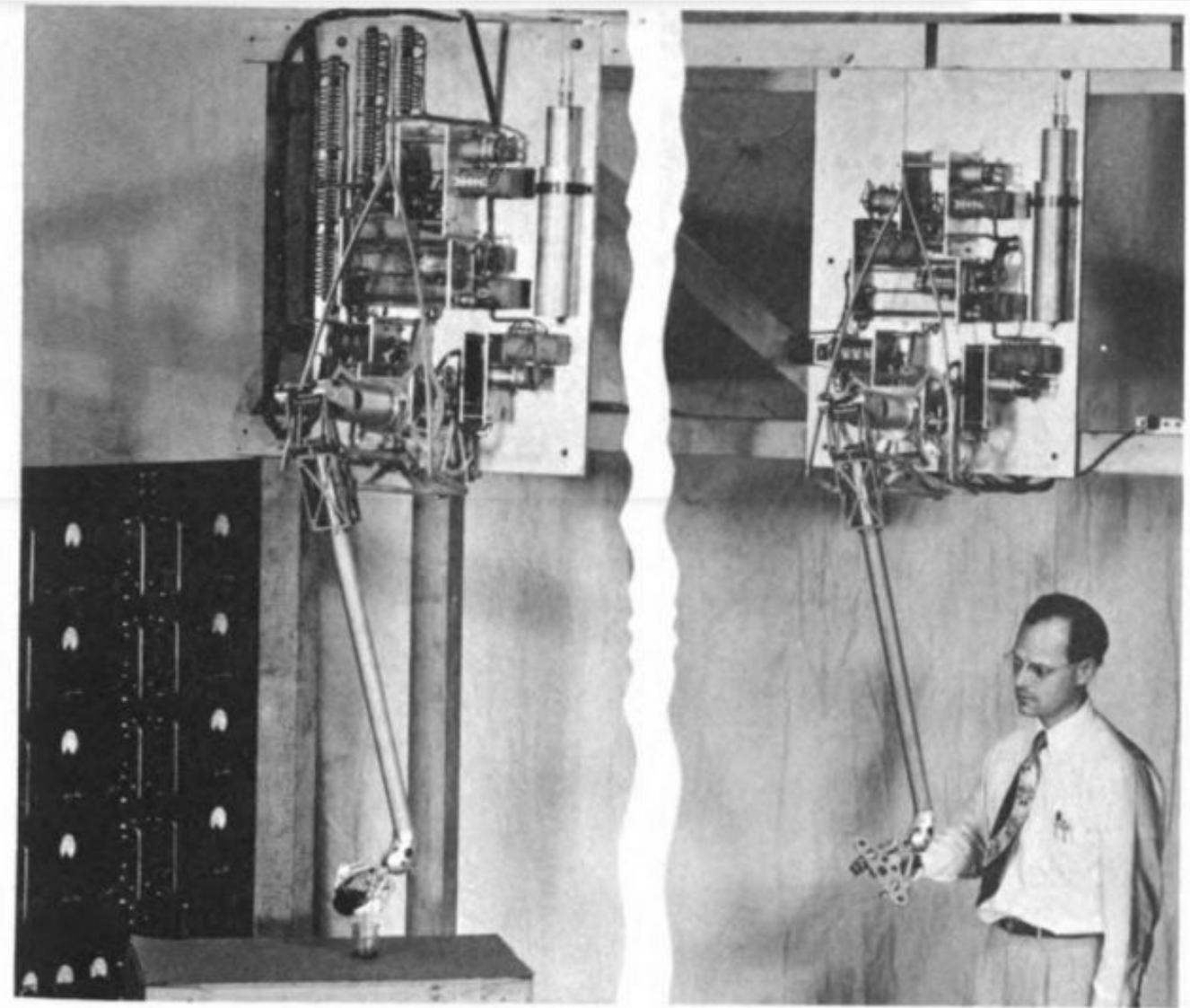


FIGURE 6.—The ANL Model E1 electric master slave. Used only for experimental purposes, this bilateral manipulator was developed in 1954. (Courtesy of Argonne National Laboratory.)



da Vinci robot



PR1 robot: <https://www.youtube.com/watch?v=qBZPSTR96N4>

This video showcases PR1, a robot developed by Keenan Wyrobek, Eric Berger, HFM Van der Loos, and Ken Salisbury at Stanford. It is the predecessor of the Willow Garage PR2. The PR1 was tele-operated during this entire video.



Teleoperation

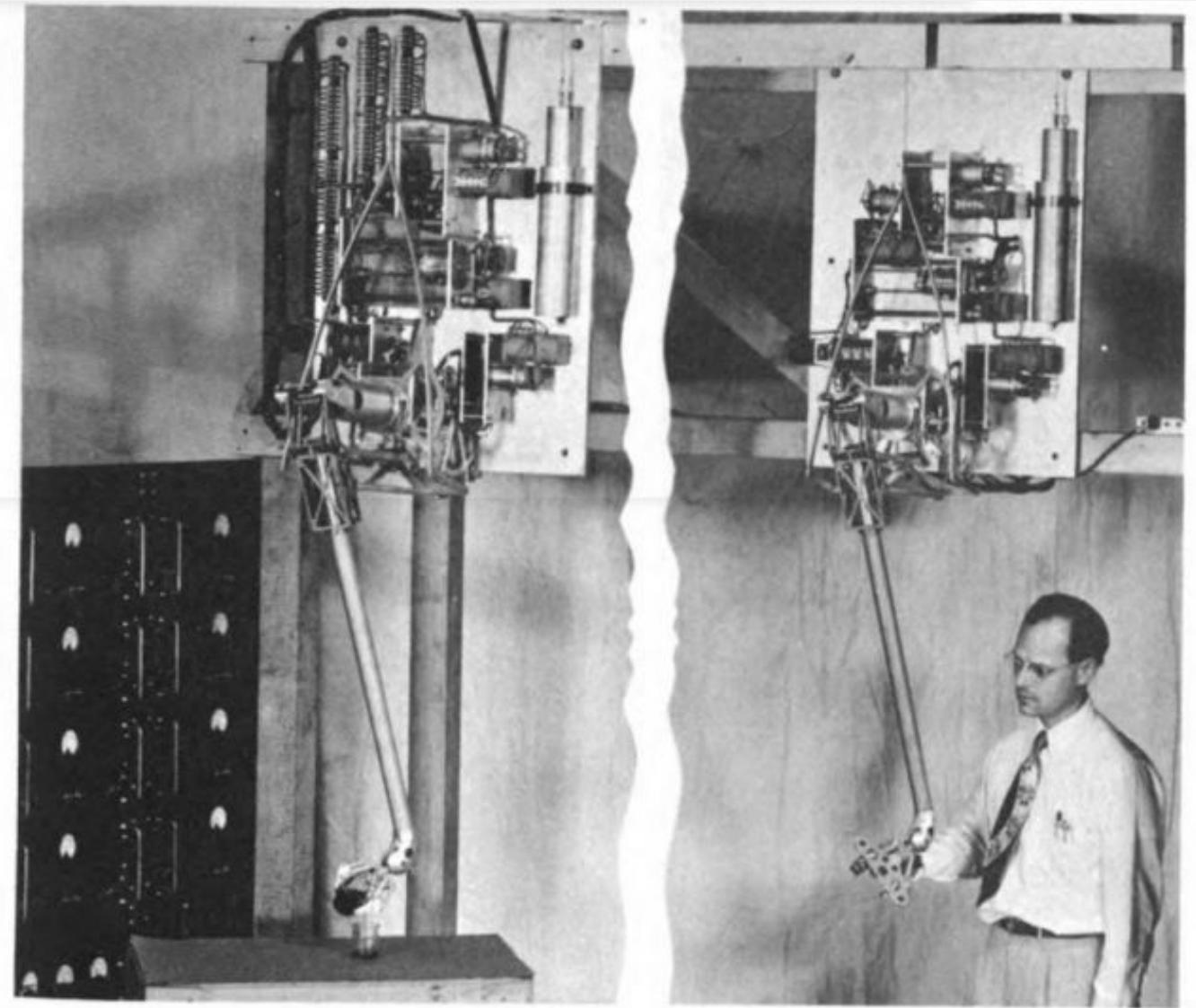


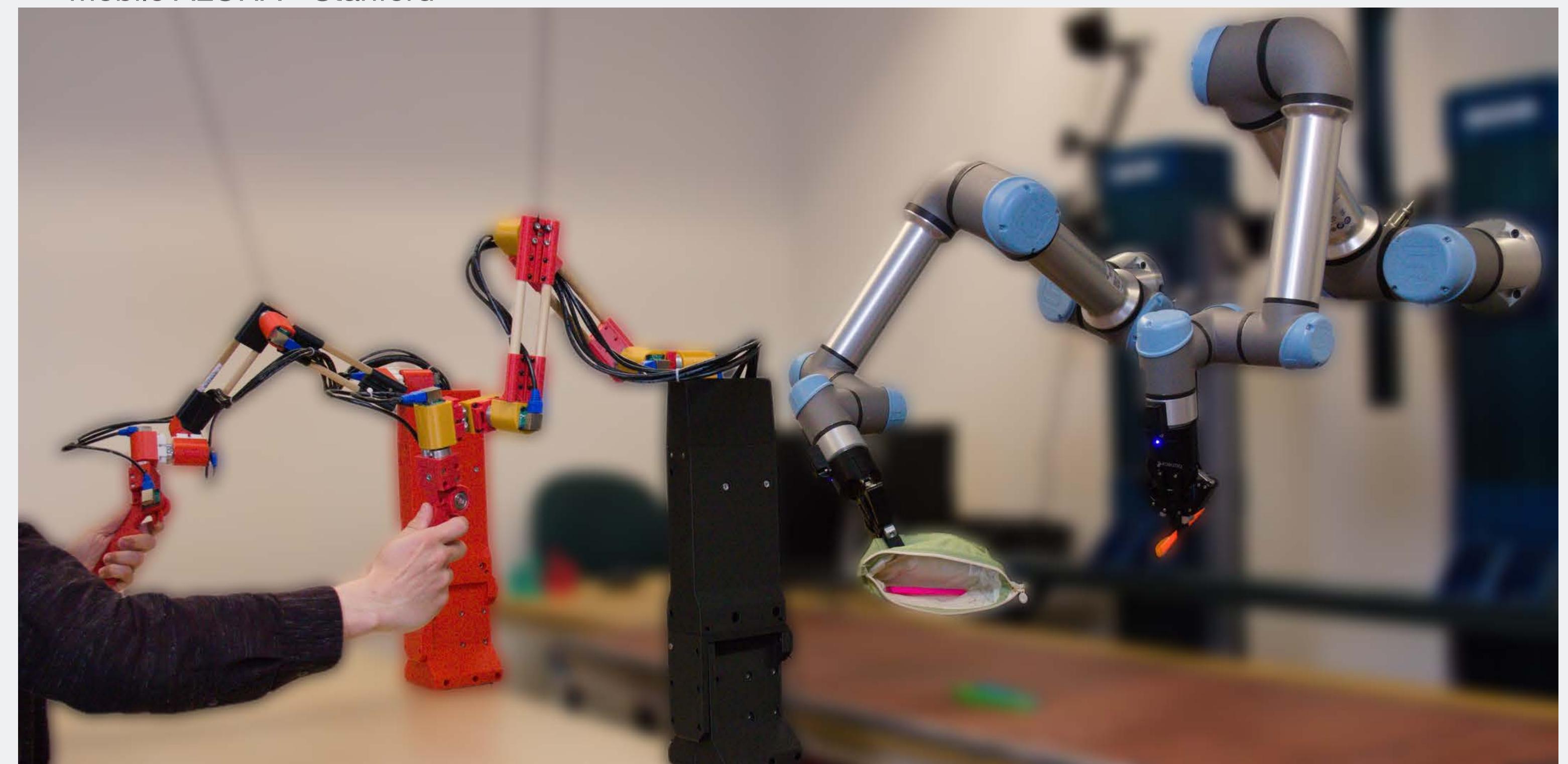
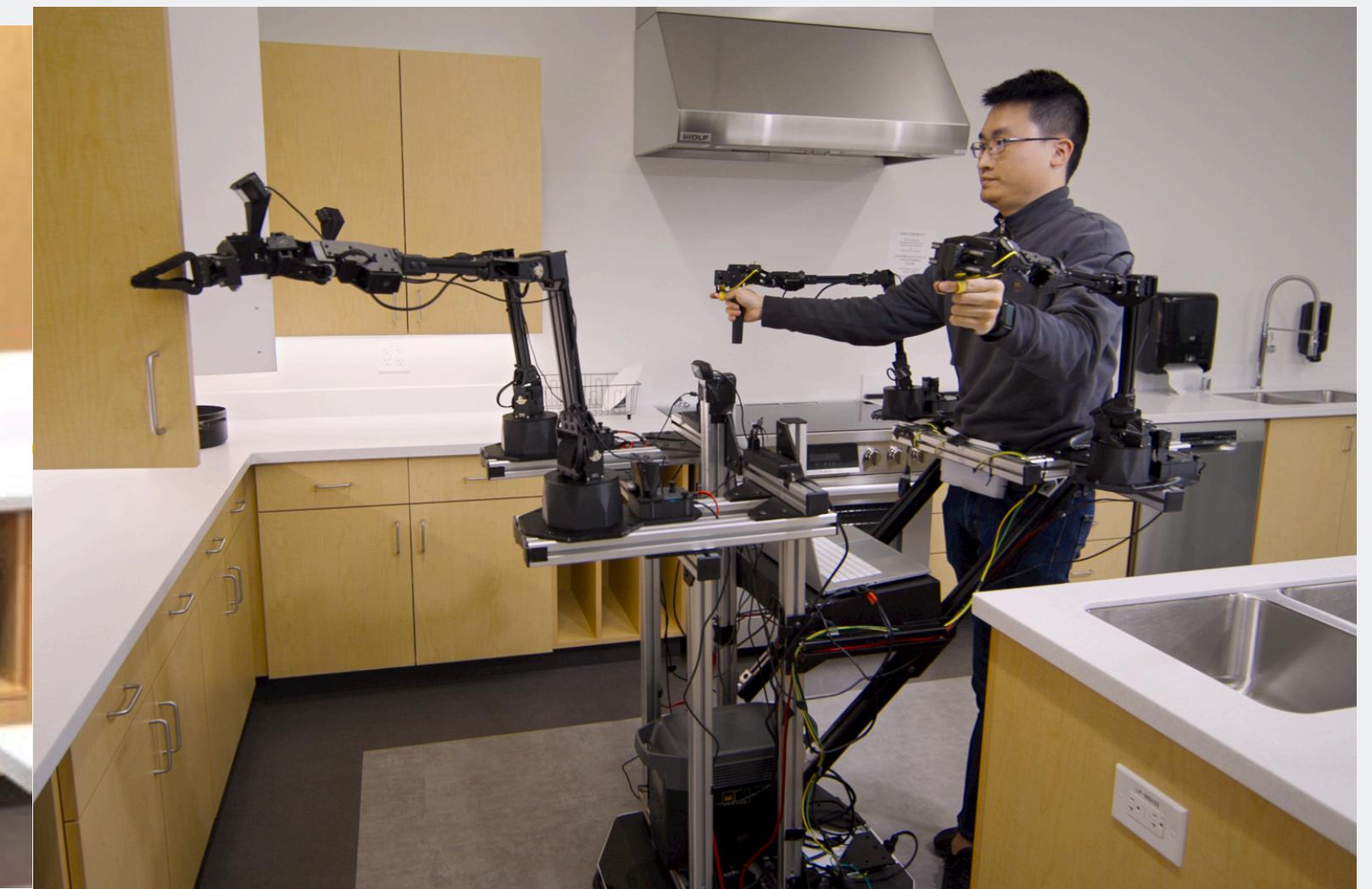
FIGURE 6.—The ANL Model E1 electric master slave. Used only for experimental purposes, this bilateral manipulator was developed in 1954. (Courtesy of Argonne National Laboratory.)



da Vinci robot



Mobile ALOHA - Stanford

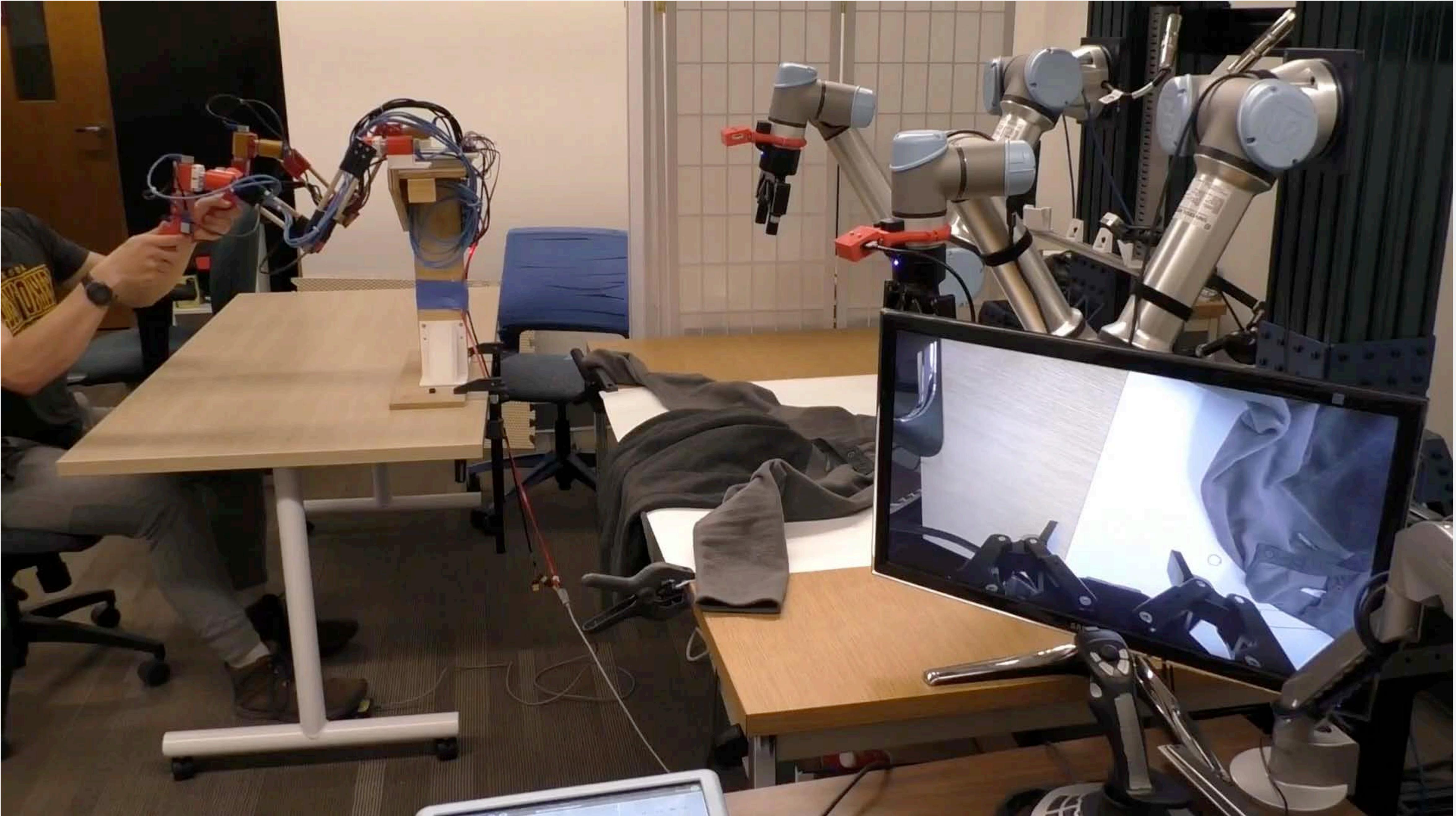


SPARK Teleoperation setup for Dual-arm task demonstrations @ **RPM Lab** - Adam Imdieke



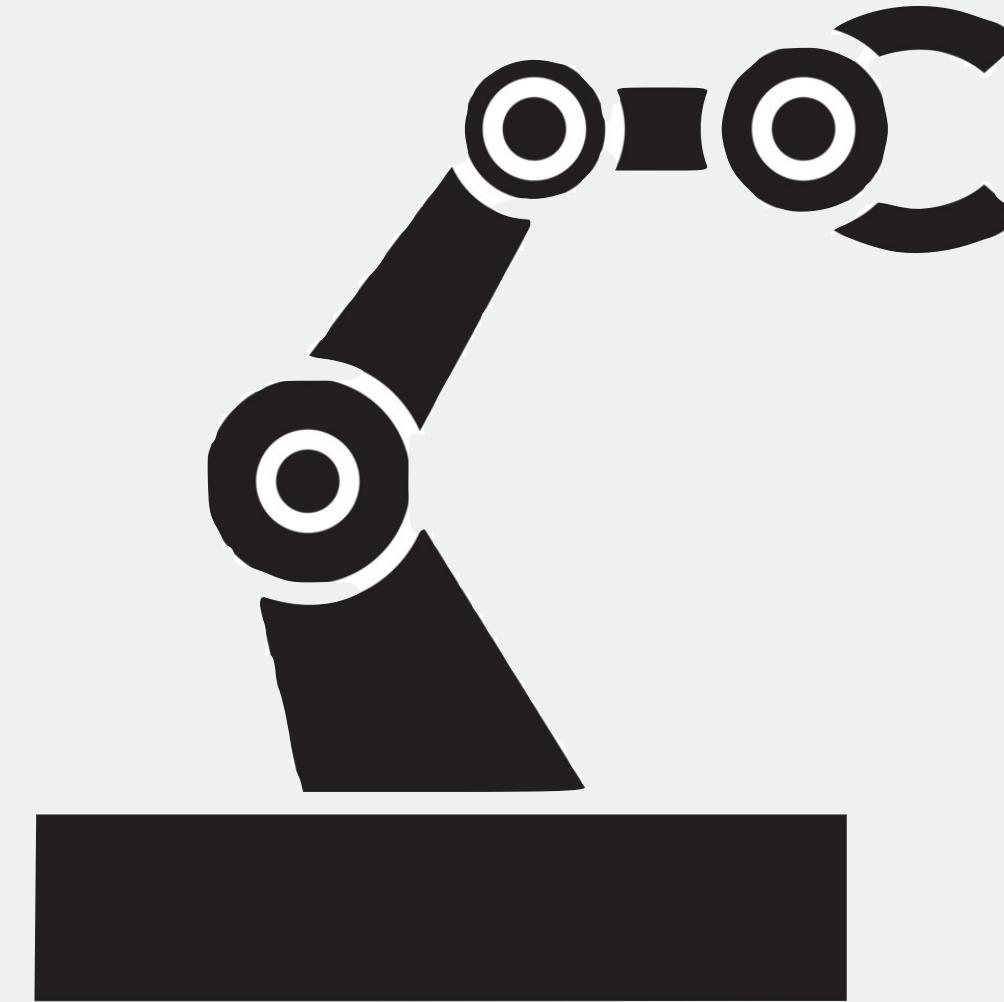
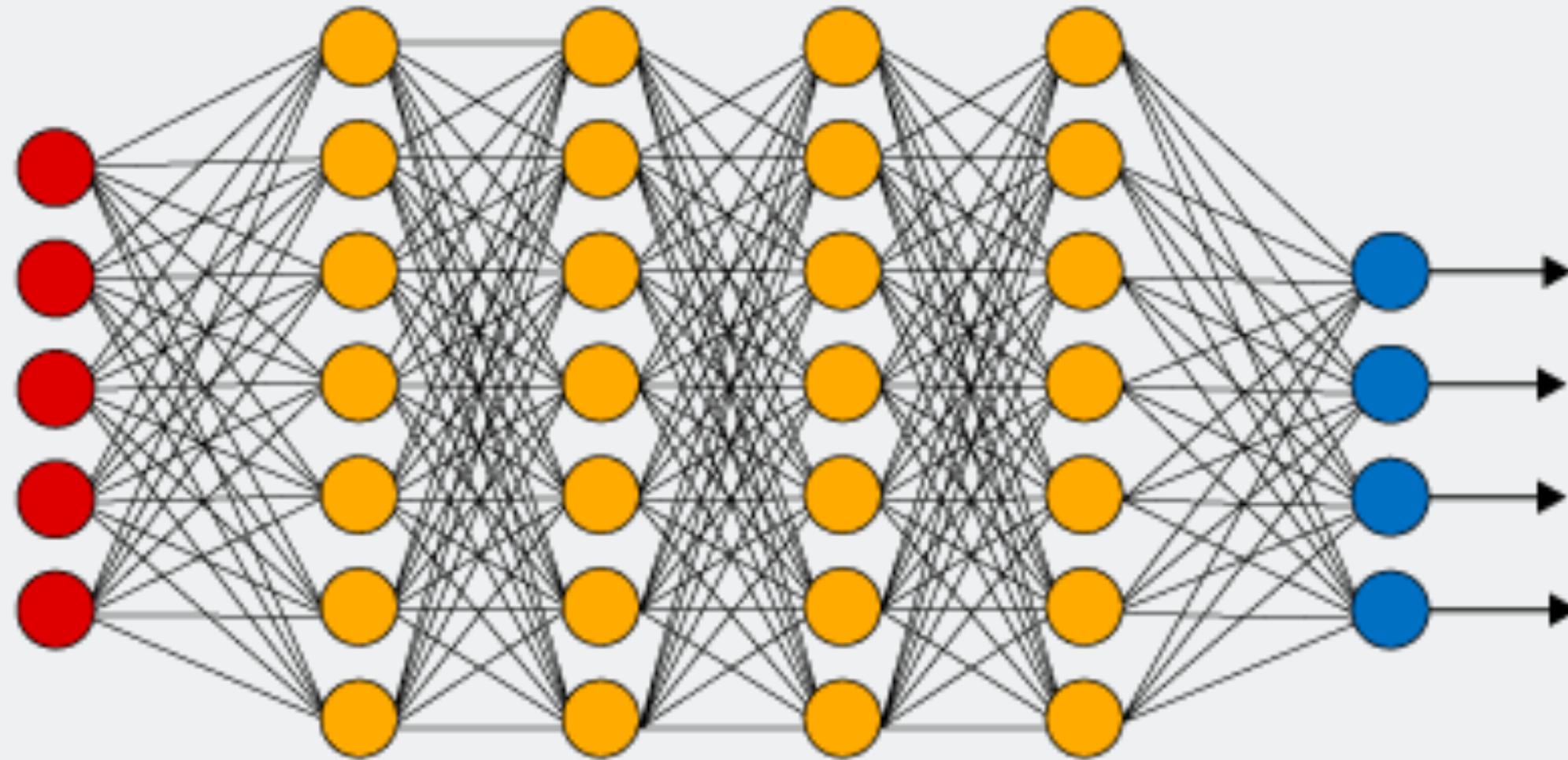


DR



SPARK Teleoperation setup for Dual-arm task demonstrations @ **RPM Lab** - Adam Imdieke

Deep Learning X Robot Manipulation



But what's the key ingredient???

Large number of high-quality expert data

Is that it???

Not really!!! Lets see few more examples...

4Progress



4Progress

Sui, Zhou, Zeng, Jenkins 2017

Goal:
Perception and manipulation
of objects in clutter...

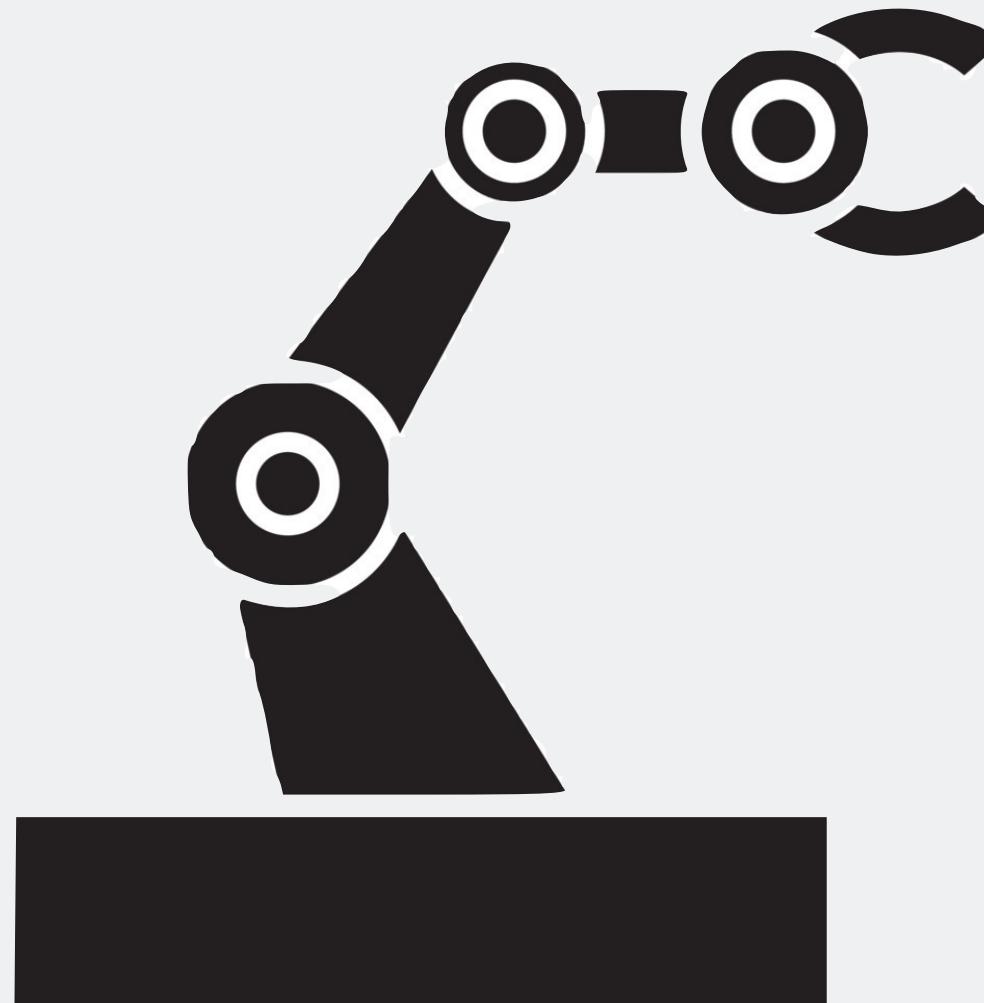
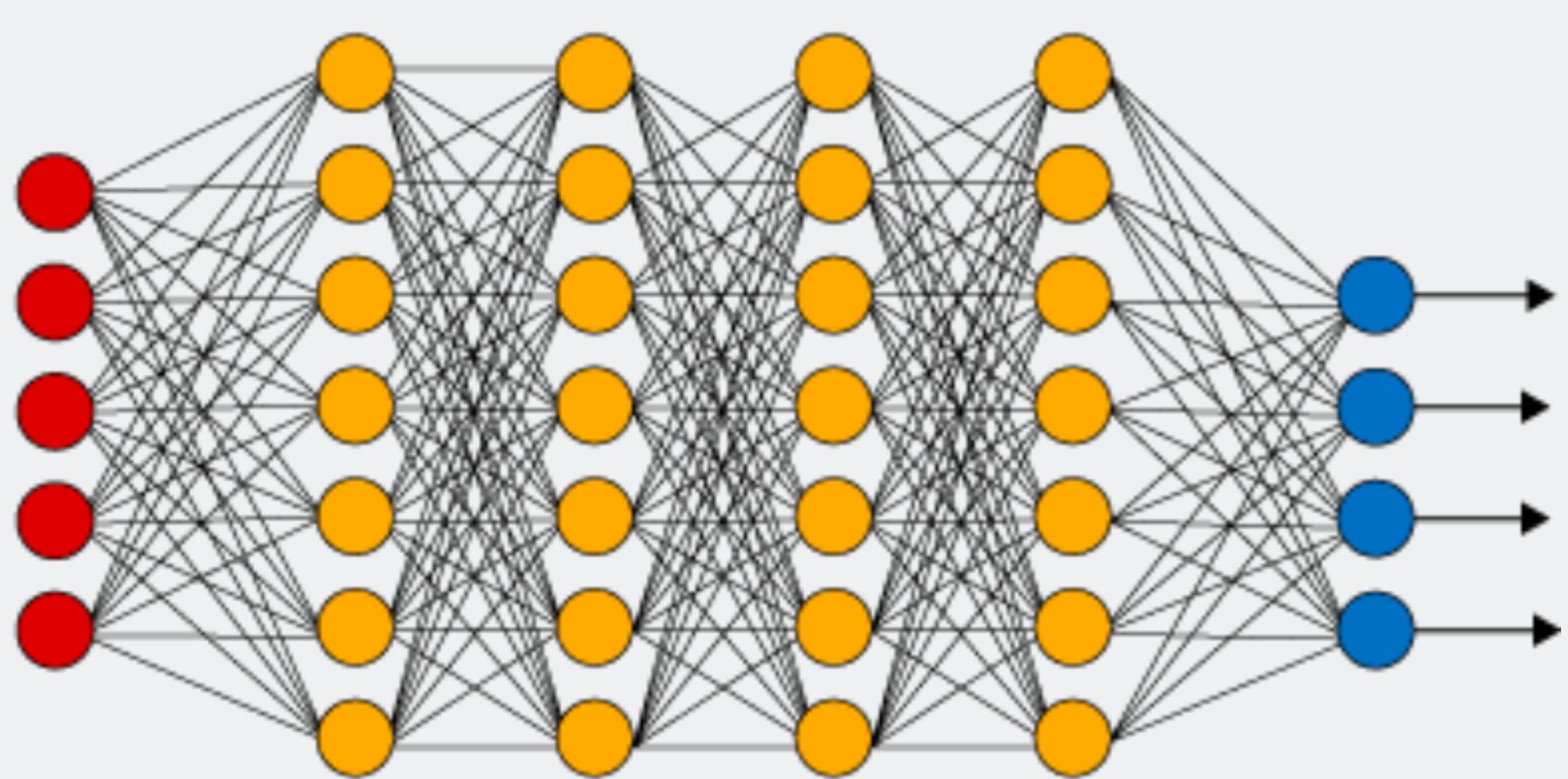


DR

... through deep learning with
neural networks (mostly) ...



Deep Learning X Robot Manipulation



But what's the key ingredient???

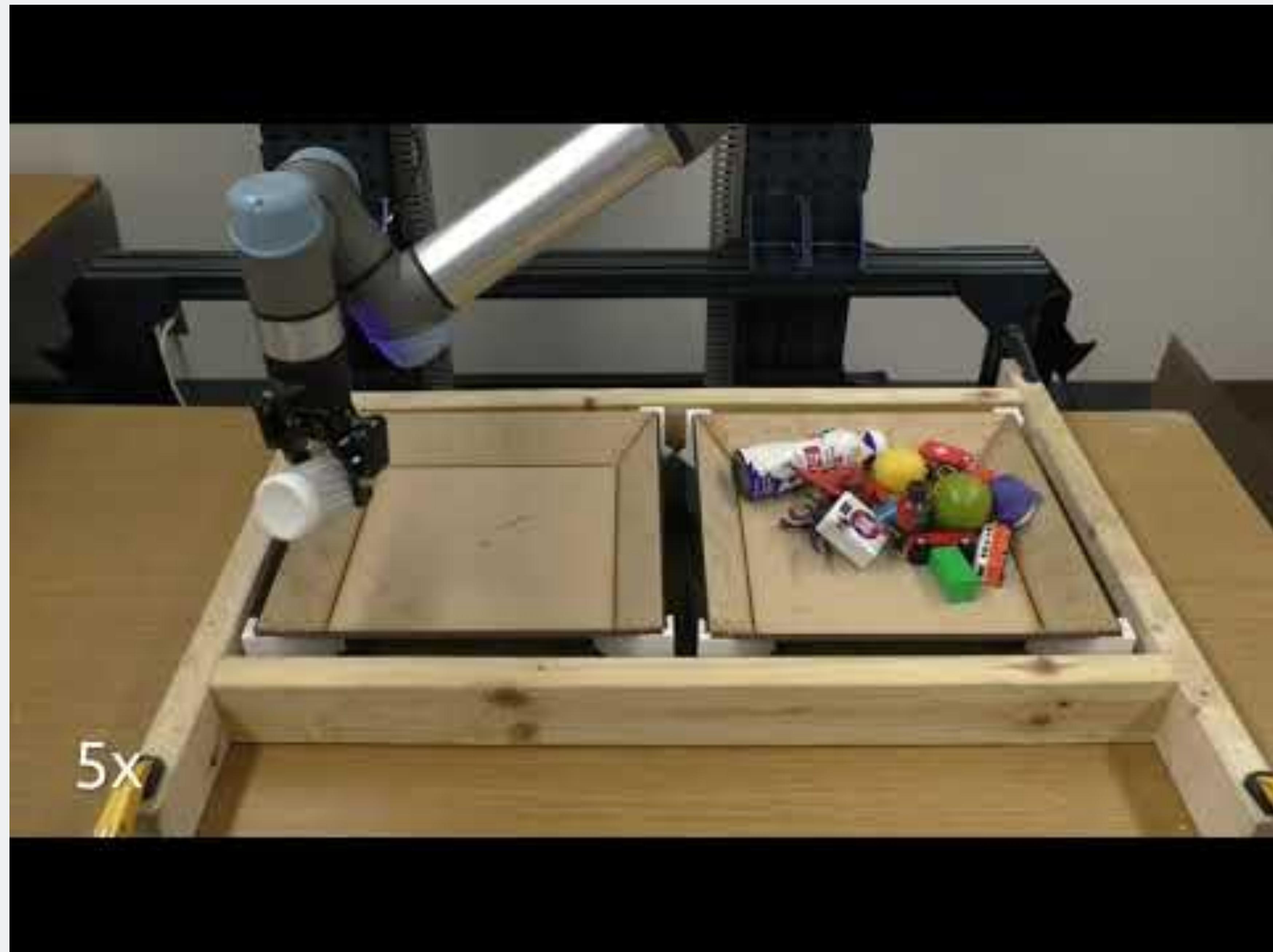
Large number of high-quality expert data

Is that it???

Not really!!! Lets see few more examples...

Robot Perception is hard!

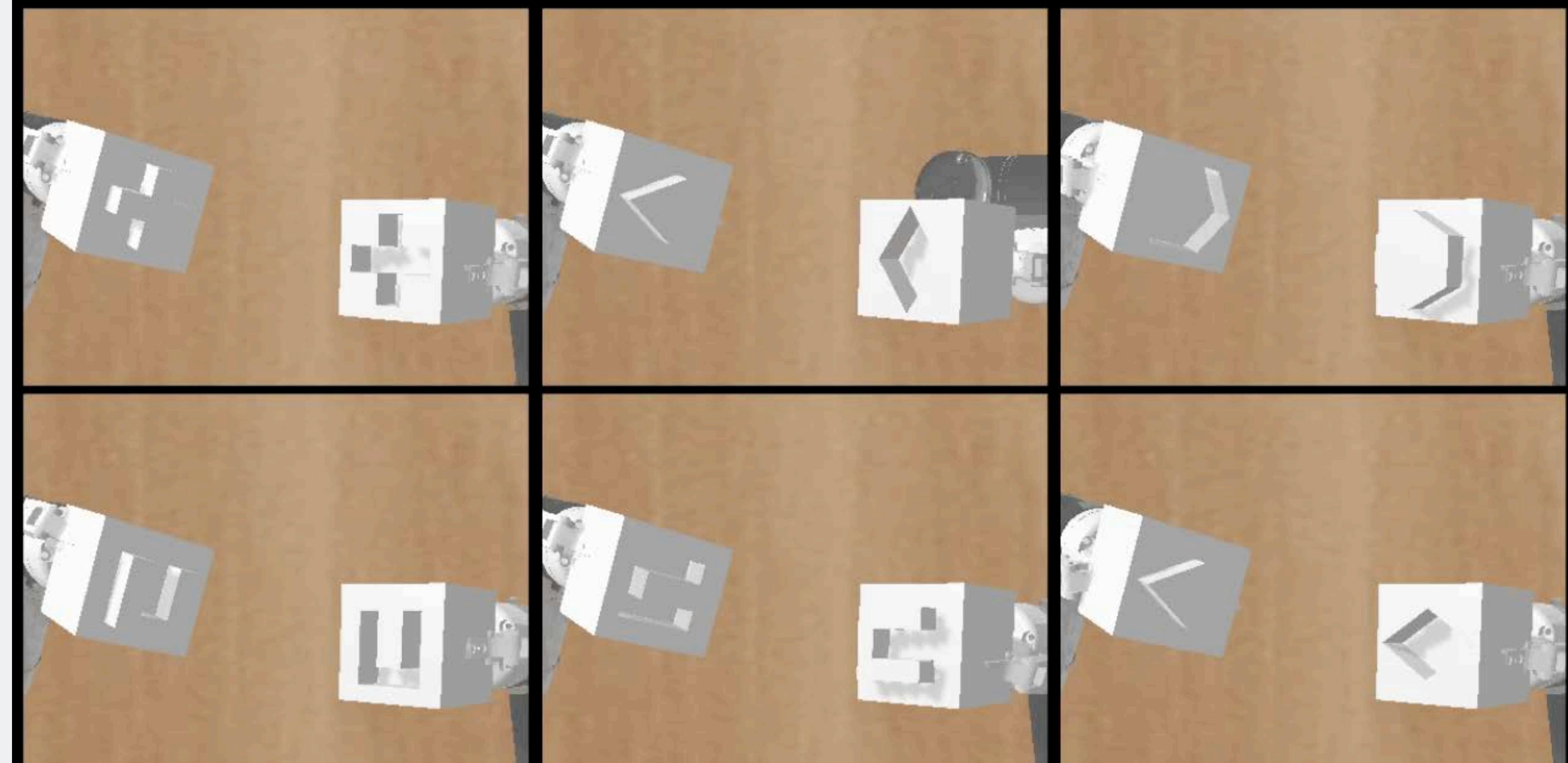
Grasp Learning



Carl Winge's reimplementation of the paper -
"Sample Efficient Grasp Learning Using Equivariant Models," by

Xupeng Zhu, Dian Wang, Ondrej Biza, Guanang Su, Robin Walters, and Robert Platt, RSS 2022.

Manipulation Learning for Object Assembly

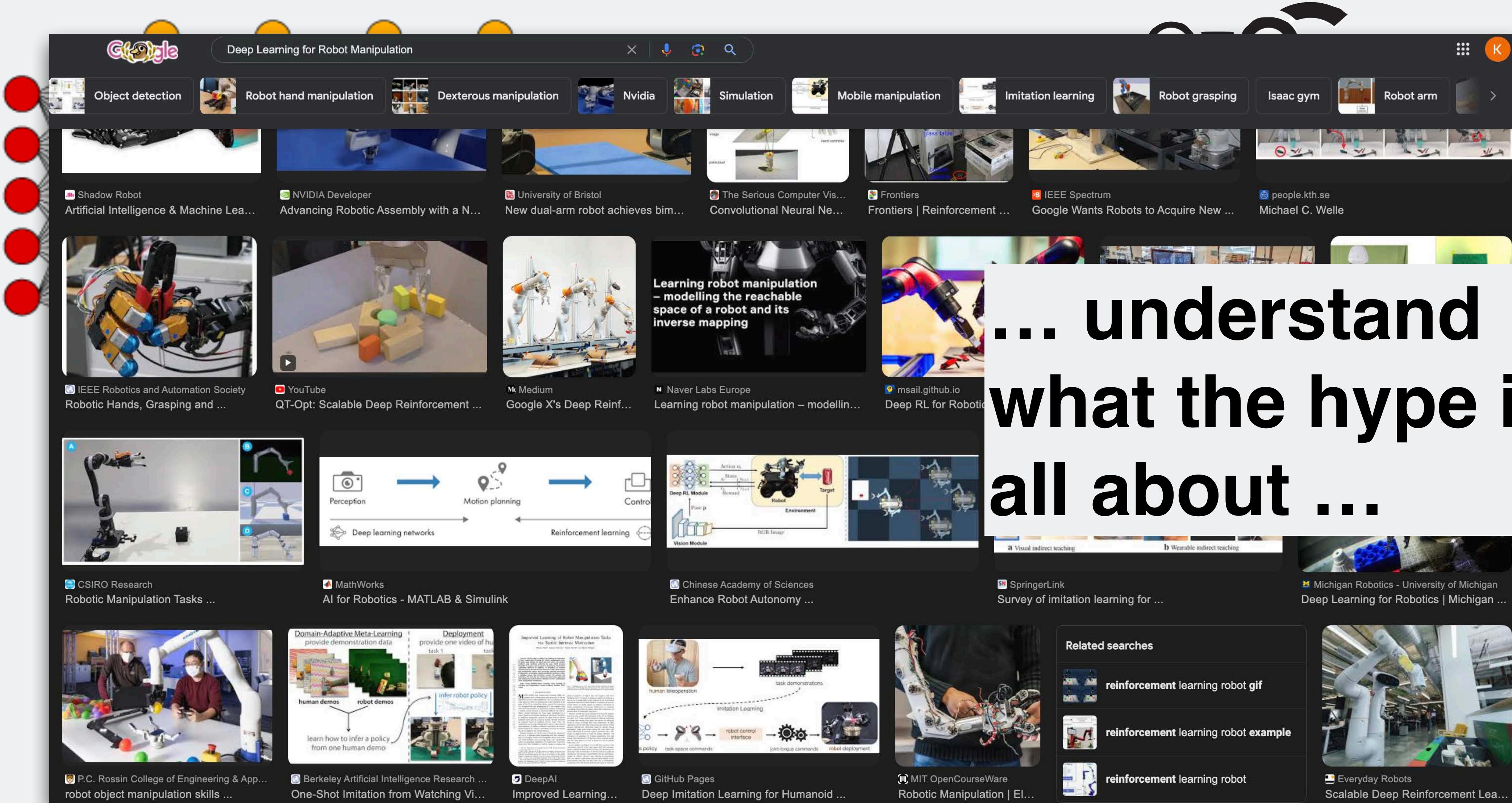


Robotics:
Perception & Manipulation
(RPM) Lab

Chahyon Ku, Carl Winge, Ryan Diaz, Wentao Yuan, Karthik Desingh
"Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning,"
ICRA 2024.



Deep Learning X Robot Manipulation



... understand
what the hype is
all about ...



Robot Perception is hard!



Welcome to DeepRob!





Welcome to DeepRob!

a Distributed Teaching Collaborative



Prof. Karthik Desingh



Dr. Xiaoxiao Du



Anthony Opiari



Prof. Chad Jenkins



UNIVERSITY
OF MINNESOTA





A brief history of AI



1956 Dartmouth Conference: The Founding Fathers of AI



John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

1956

1960

1970

1980

1990

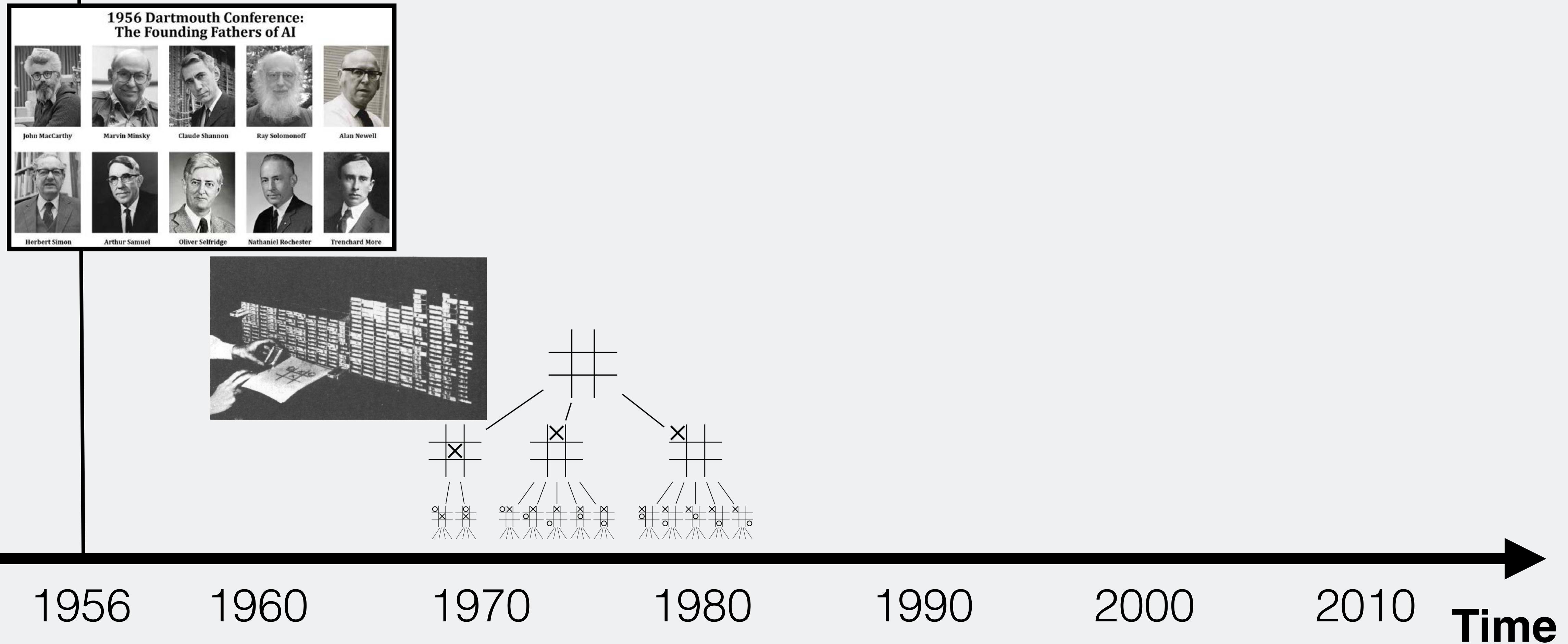
2000

2010

Time

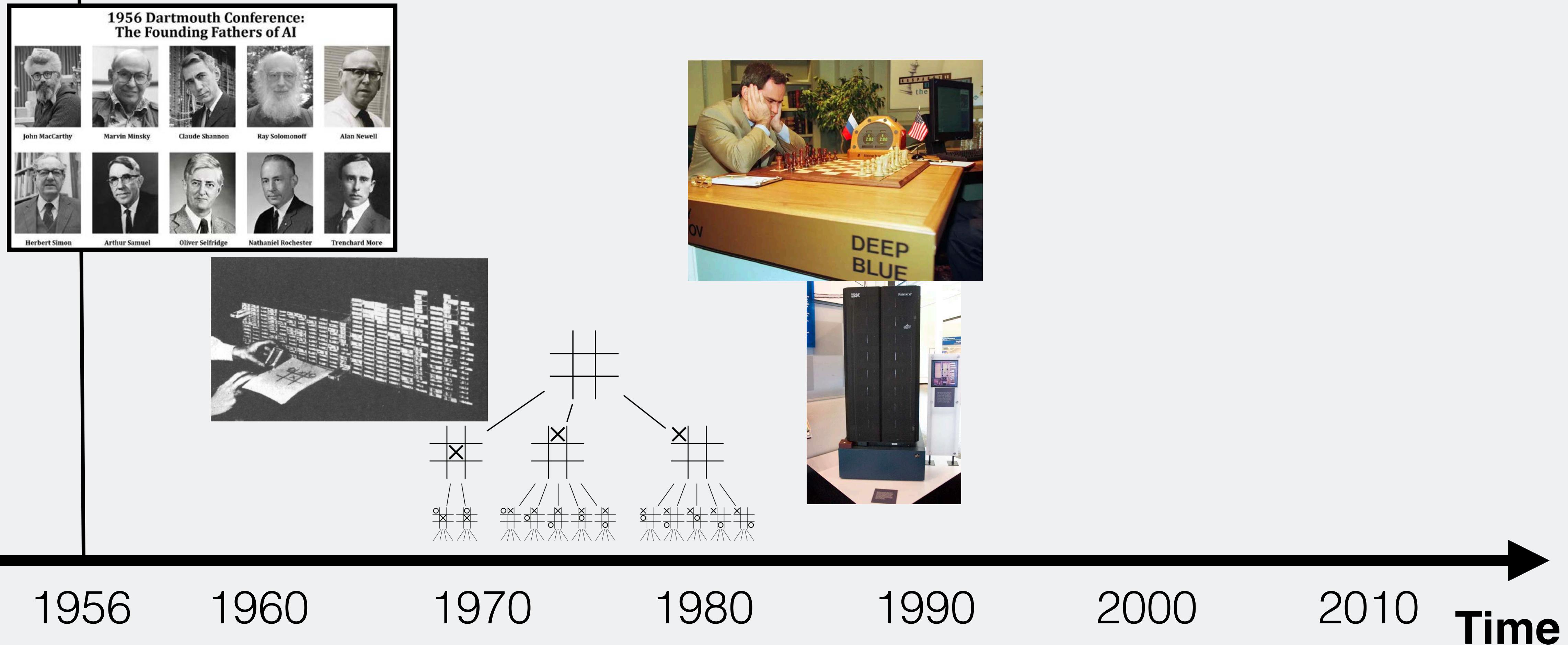
First wave AI: Model-based

“Think through the entire problem”



First wave AI: Model-based

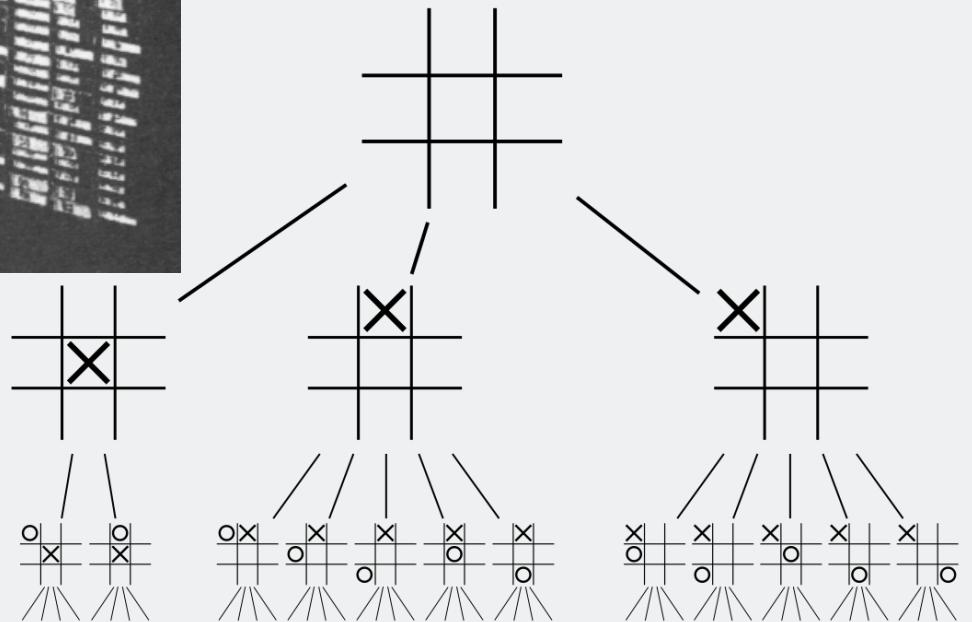
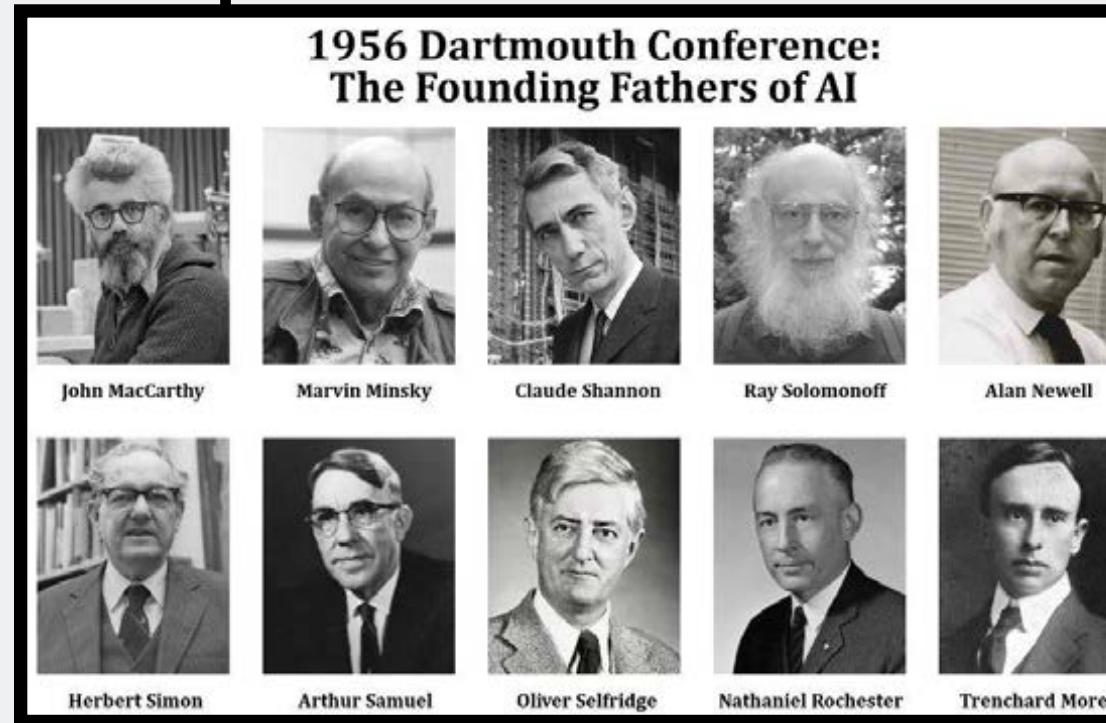
“Think through the entire problem”



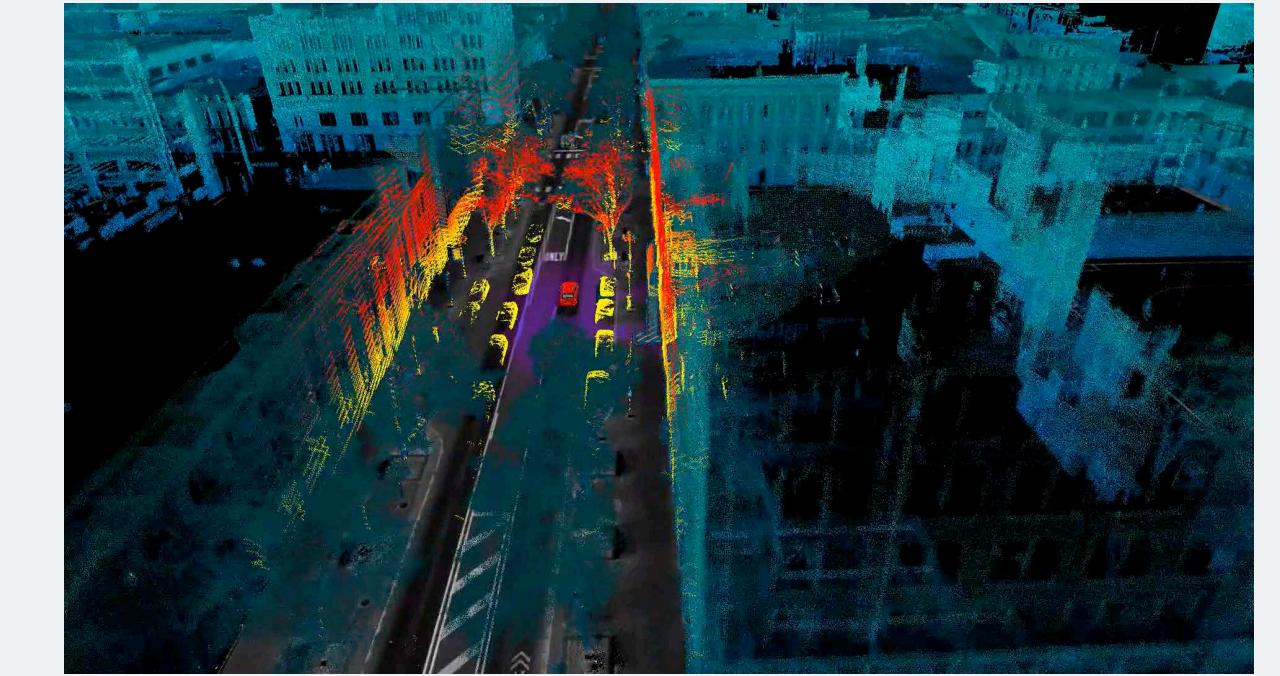


First wave AI: Model-based

“Think through the entire problem”



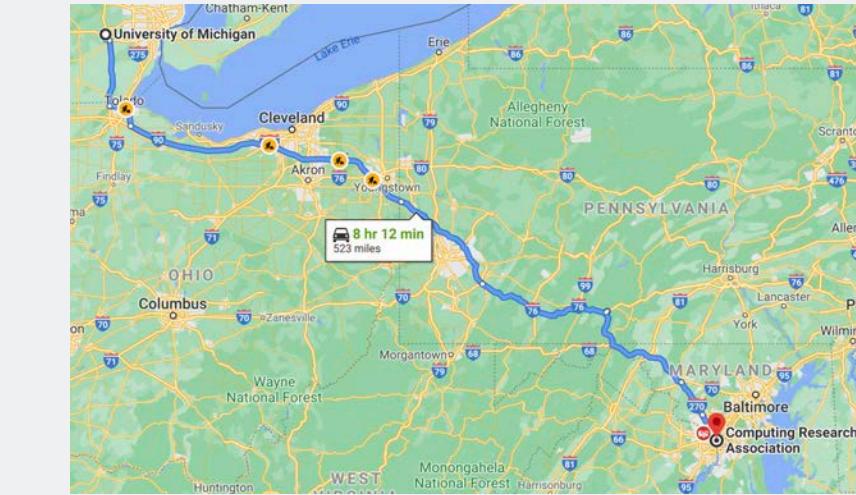
City-scale 3D mapping



Self-driving cars



Online route navigation



1956

1960

1970

1980

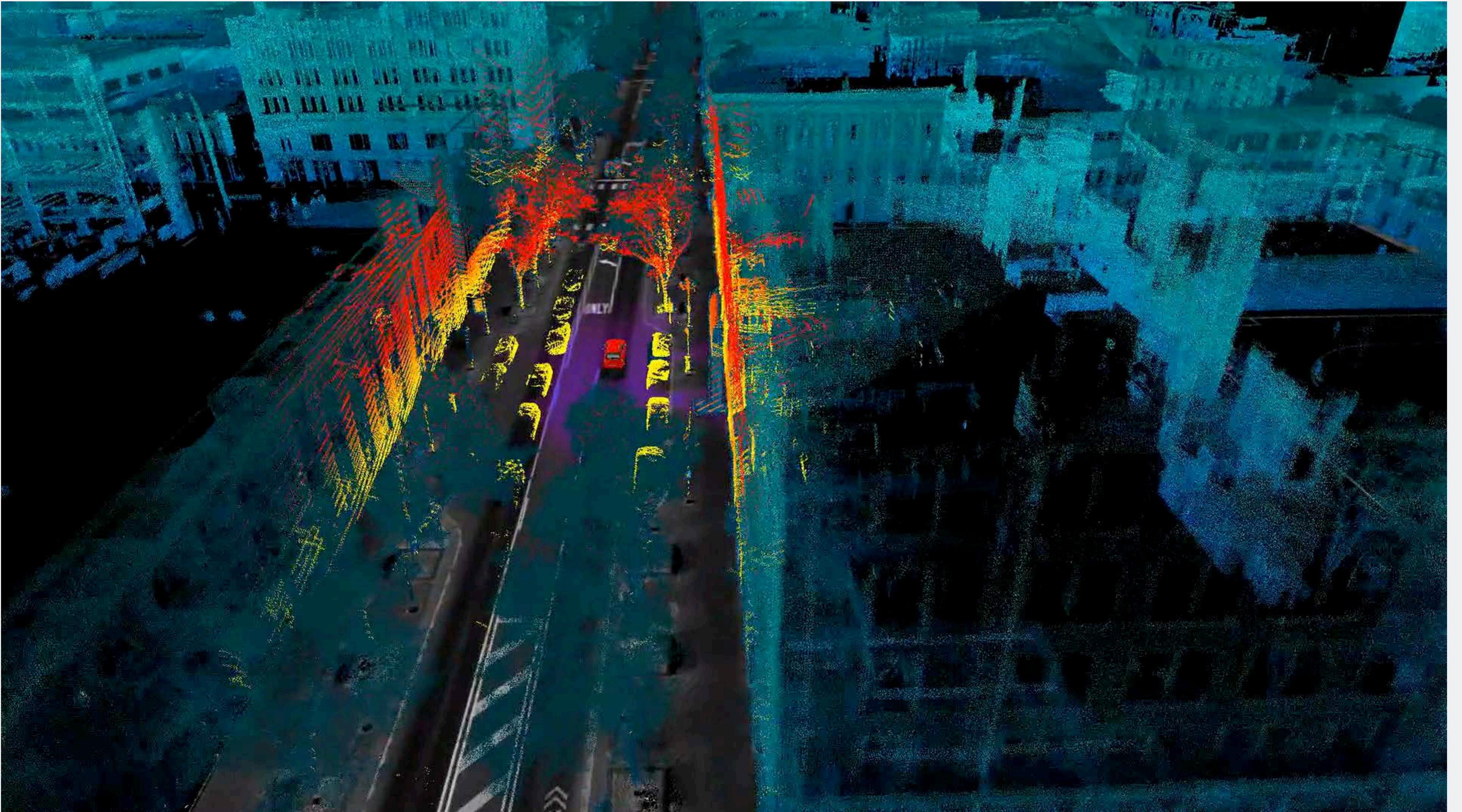
1990

2000

2010

Time

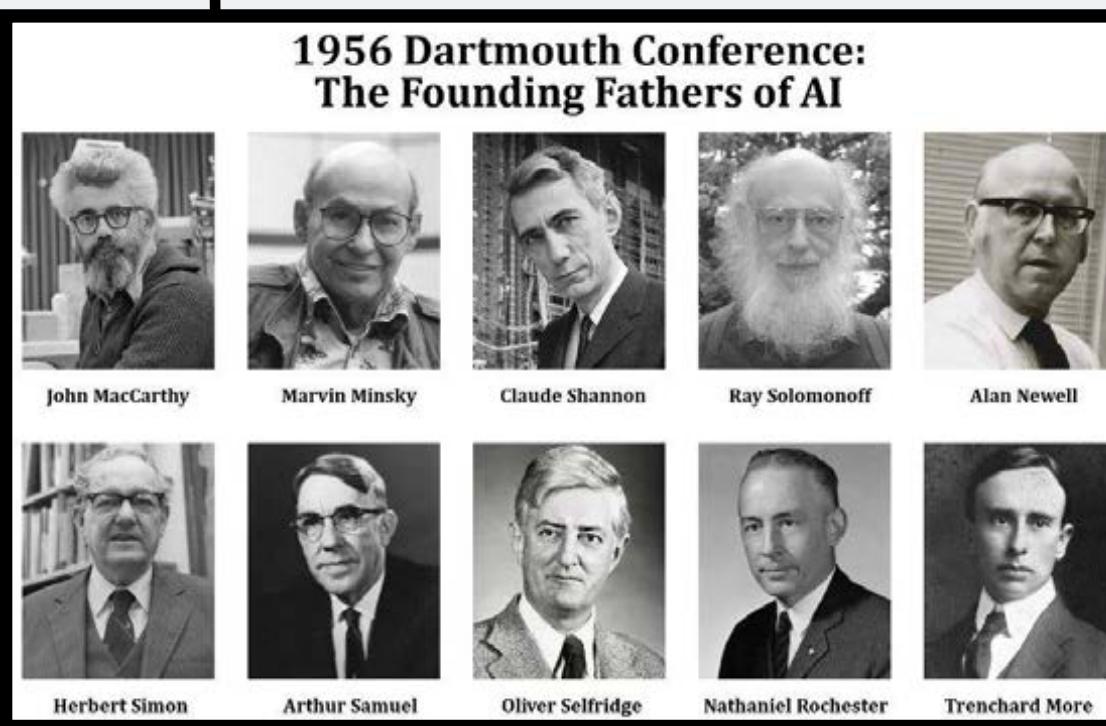




Let's watch this video closely

First wave AI: Model-based

“Think through the entire problem”

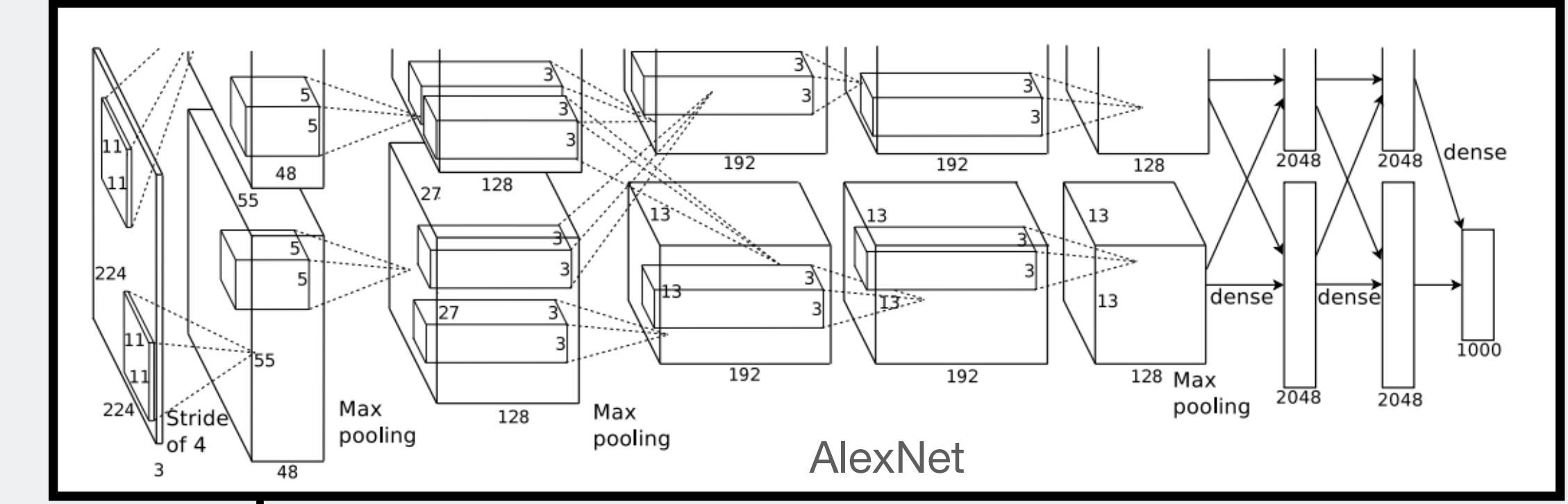


1956

Second wave AI: Data-driven

“Learn from lots of data”

Rise of “deep learning”



2011

Time

First wave AI: Model-based

The screenshot shows a news article from IEEE Spectrum. The title is "Deep Learning Makes Driverless Cars Better at Spotting Pedestrians". Below the title is a subtitle: "Pedestrian detection systems for cars could become faster and more accurate with help from deep learning algorithms". The author is listed as "By Jeremy Hsu". Below the author's name is a grid of four images showing pedestrian detection results. The first two images show a street scene with people walking, and the last two show a car driving through an intersection. Red boxes highlight detected pedestrians. The bottom of the grid is captioned "Images: Statistical Visual Computing Lab/UC San Diego". The top navigation bar includes links for "Engineering Topics", "Special Reports", "Blogs", "Multimedia", "The Magazine", "Professional Resources", and "Search". A secondary navigation bar below the main one includes "Cars That Think", "Transportation", and "Advanced Cars". The date "09 Feb 2016 | 17:00 GMT" is also present.

Second wave AI: Data-driven

"Learn from lots of data"



"deep learning"

1956

2011

Time



Second wave AI: Data-driven

TEXT PROMPT
an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES

Edit prompt or view more images↓

TEXT PROMPT
an armchair in the shape of an avocado....

AI-GENERATED IMAGES

Edit prompt or view more images↓

DALL-E (OpenAI)

“Learn from lots of data”



“deep learning”

2011

Time





TEXT PROMPT
an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES

Your AI pair programmer

GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time, right from your editor.

[Start my free trial >](#) [Explore docs](#)

TEXT PROMPT
an armchair in the shape

AI-GENERATED IMAGES

```
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch(`http://text-processing.com/api/sentiment/`, {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17}
```

[Edit prompt or view more images](#)

DALL-E (OpenAI)

Copilot (OpenAI)

Second wave AI: Data-driven

“Learn from lots of data”



“deep learning”

2011

Time



Second wave AI: Data-driven

TEXT PROMPT
an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES



Your AI pair programmer

GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in

In the following sample, ChatGPT is able to understand the reference ("it") to the subject of the previous question ("fermat's little theorem").

Sample 3 of 4 Previous Next

User what is the fermat's little theorem

ChatGPT Fermat's Little Theorem states that if p is a prime number and a is an integer not divisible by p , then a^p (a to the power p) is congruent to a modulo p . In other words, if we divide a^p by p , the remainder is always a .

For example, if $p = 5$ and $a = 2$, then $2^5 = 32$, and 32 divided by 5 has a remainder of 2. Therefore, 2^5 is congruent to 2 modulo 5, and Fermat's Little Theorem holds for this case.

Fermat's Little Theorem is often used in cryptography and other applications where it is necessary to perform modular arithmetic operations quickly and efficiently. It is also a useful tool for proving other theorems in number theory

User how is it used in cryptography?

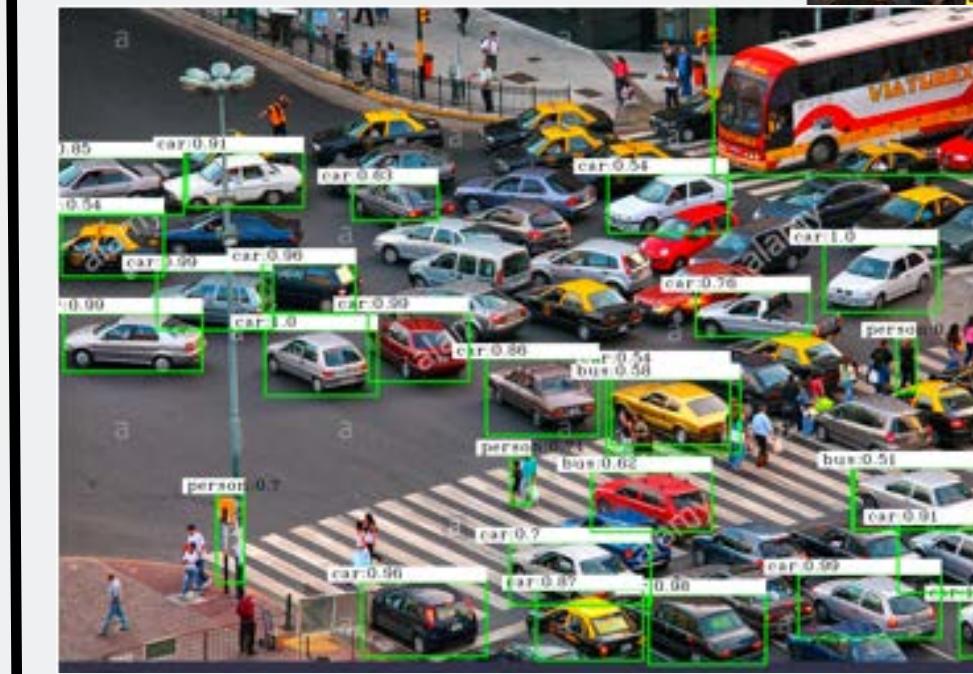
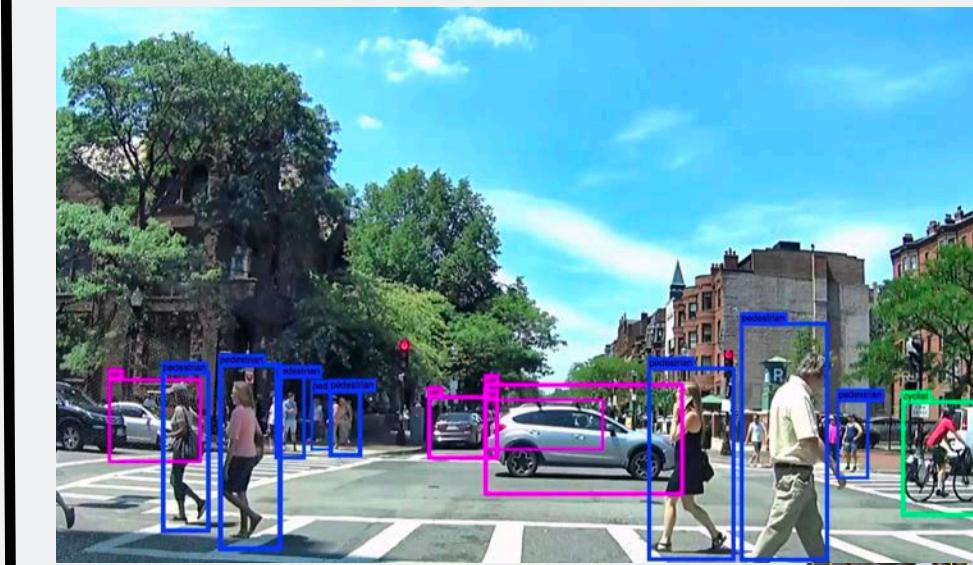
ChatGPT Fermat's Little Theorem is used in cryptography in several ways. One of the most common applications is in the generation of so-called "public-key" cryptography systems, which are used to securely transmit messages over the internet and other networks.

In a public-key cryptography system, each user has a pair of keys: a public key, which is widely known and can be used by anyone to encrypt a message intended for that user, and a private key, which is known only to the user and is used to decrypt messages that have been encrypted with the corresponding public key.

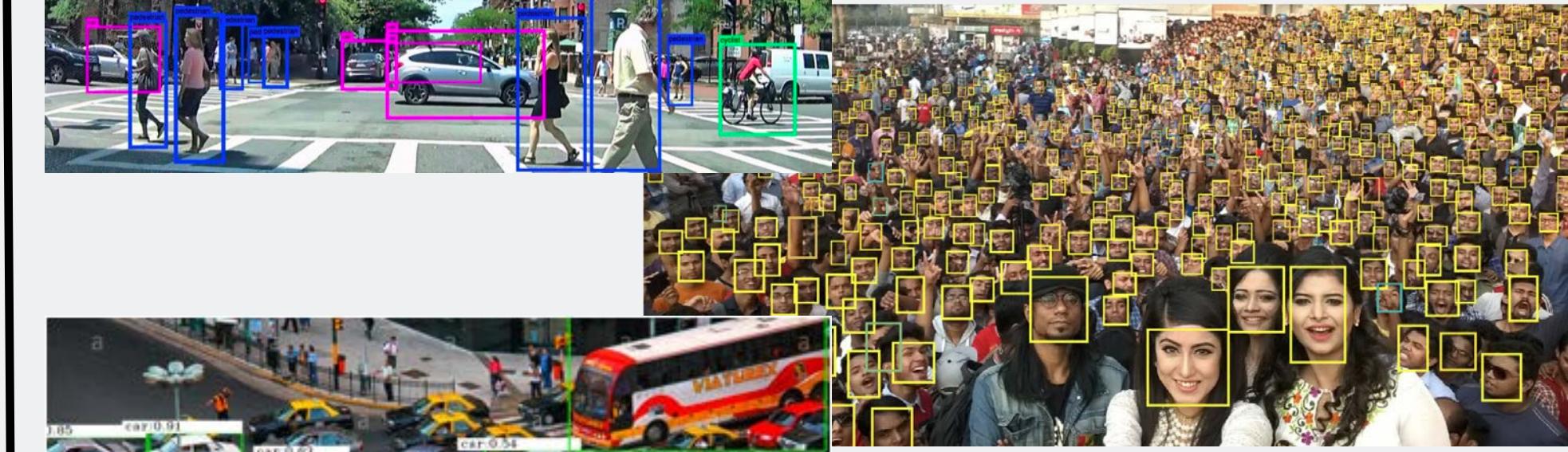
DALL-E (OpenAI)

Copilot (OpenAI)

ChatGPT (OpenAI)



"Learn from lots of data"



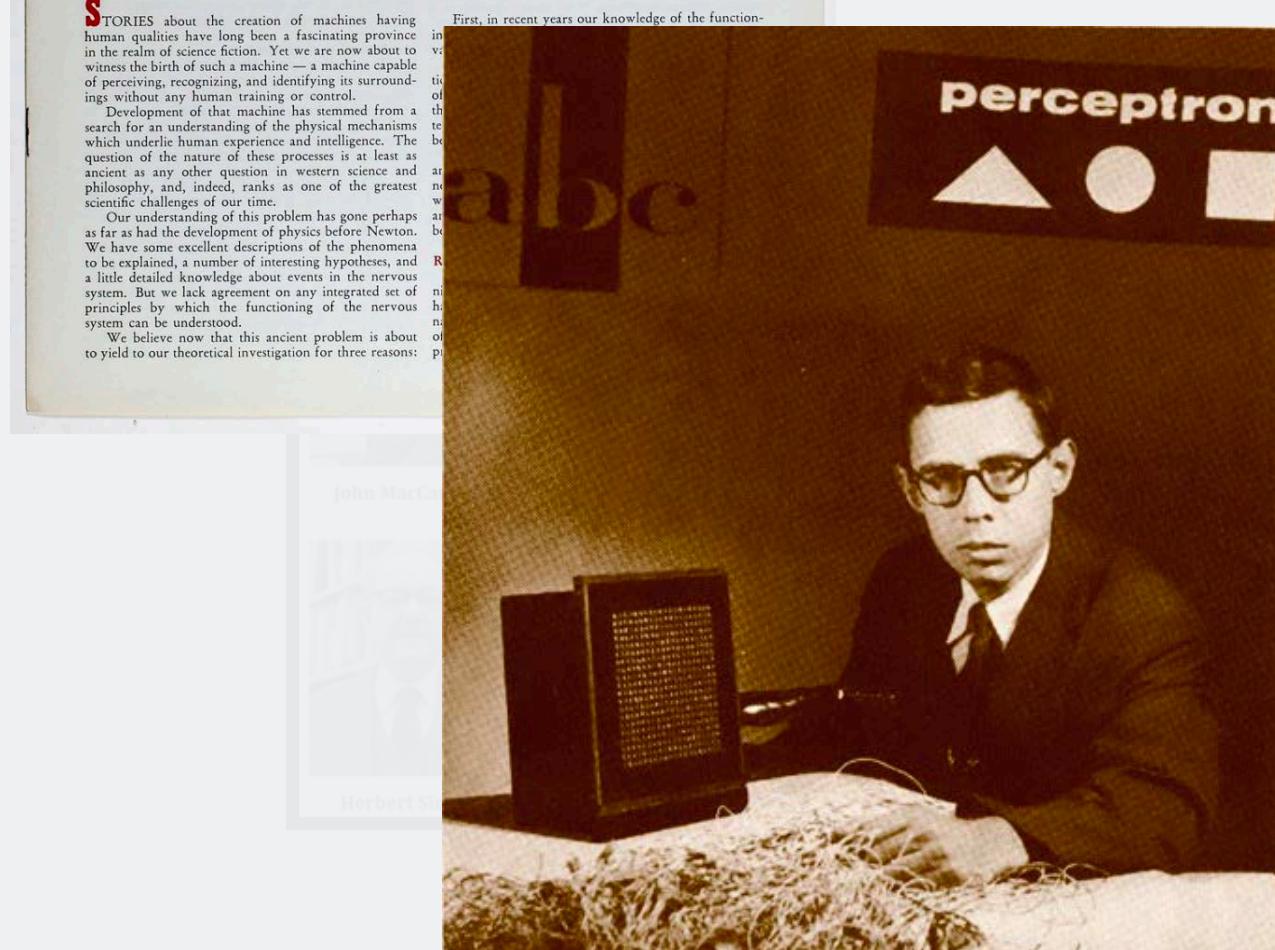
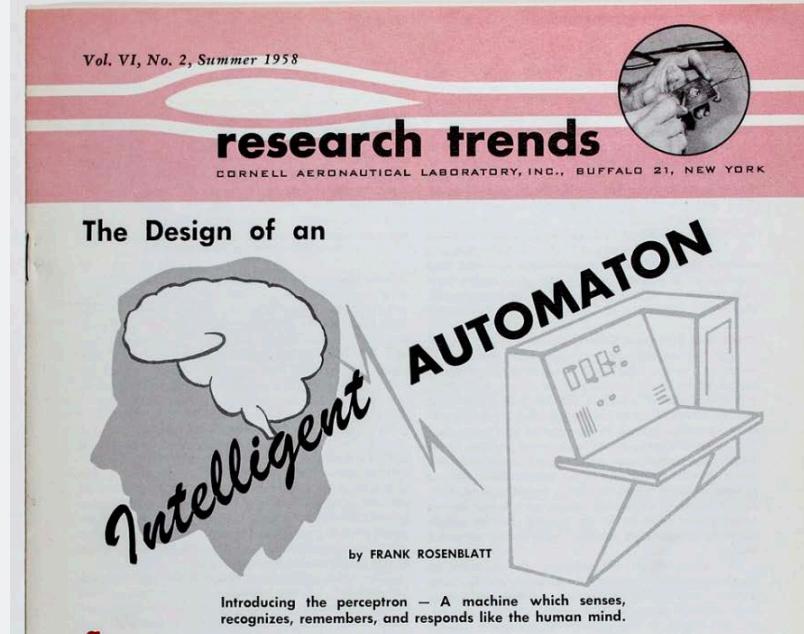
"deep learning"

2011

Time



The AI of today needed decades of investment and research



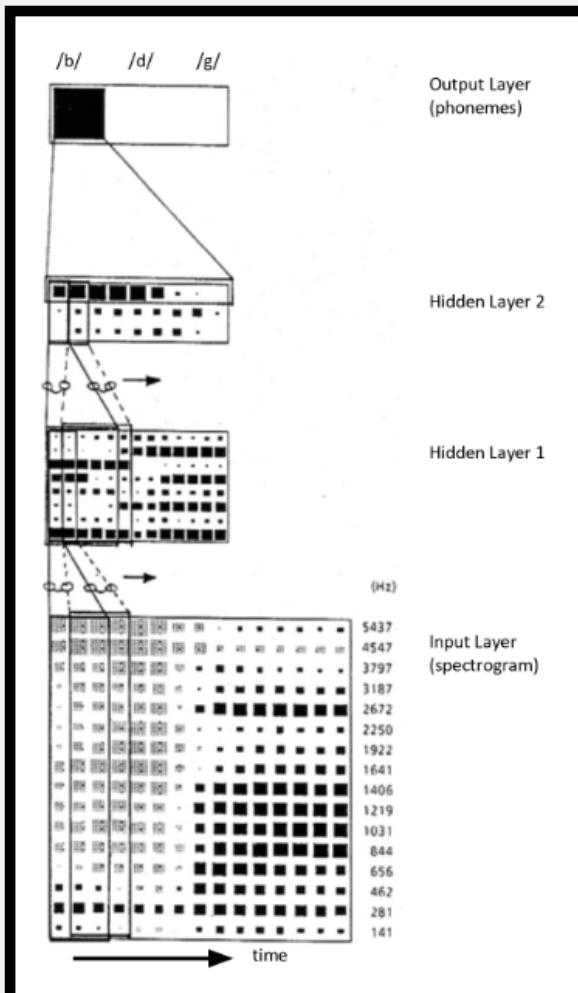
Rosenblatt's Perceptron

1956

1958



Speech recognition by neural networks



Waibel et al.

1989

2011

Second wave AI: Data-driven

“Learn from lots of data”



“deep learning”

Time →





What is this?

DR

Search

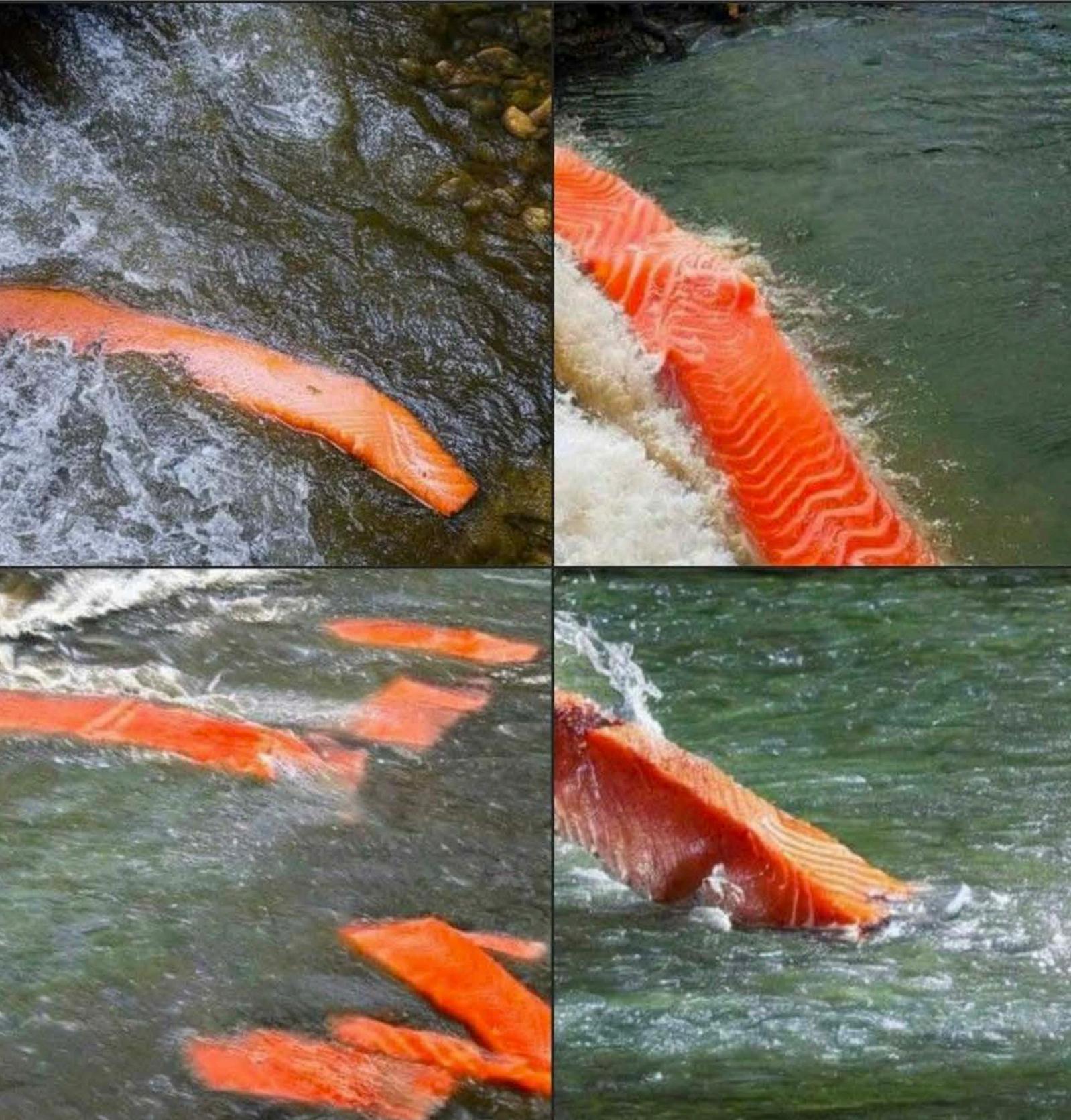
Carlos la Orden Tovar · 3rd+ Trainer and speaker at international events in the language industry 2d

+ Follow

They asked an AI engine to recreate "a salmon swimming down a river", and here's what it guessed it would look like.

It's exactly the same thing that happens to your texts when you leave them in the hands of a raw machine translation engine and fail to plan a proper post-editing, proofreading and—above all—put a discerning human eye in the loop to look for context and read between the lines.

#IA #MT #MachineTranslation



Be mindful about how you use AI

“a salmon swimming down a river”



First wave AI: Model-based

The Economist ≡ Menu Weekly edition Search Subscribe

Technology Quarterly Automobiles

Driverless cars show the limits of today's AI

Jun 13th 2020 edition >

They, and many other such systems, still struggle to handle the unexpected

Tom Gauld

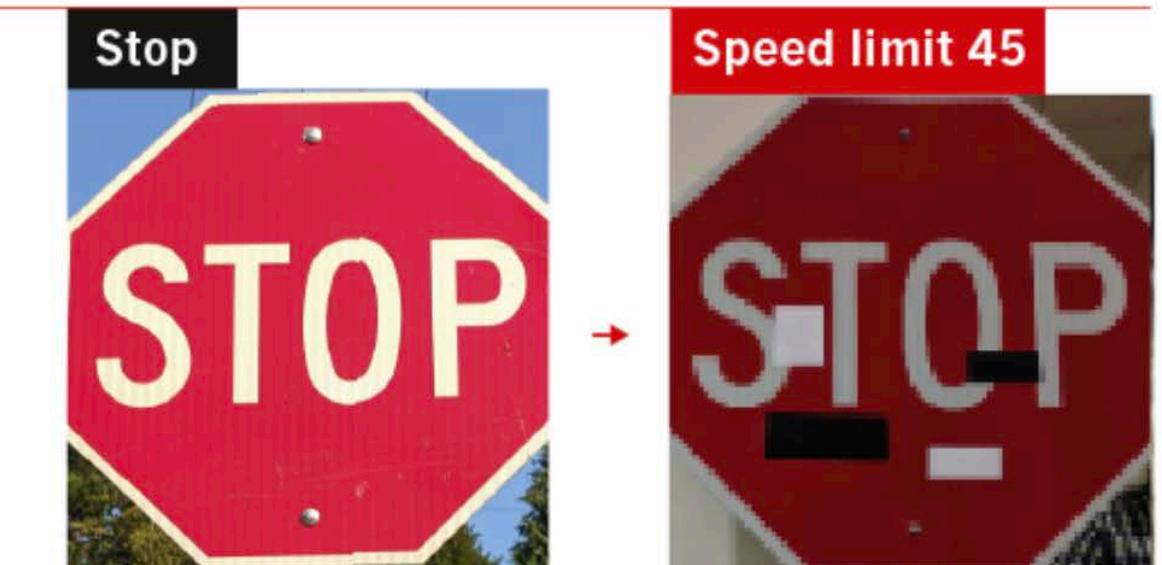
Second wave AI: Data-driven

“Learn from lots of data”

FOOLING THE AI

Deep neural networks (DNNs) are brilliant at image recognition — but they can be easily hacked.

These stickers made an artificial-intelligence system read this stop sign as ‘speed limit 45’.



©nature

1956

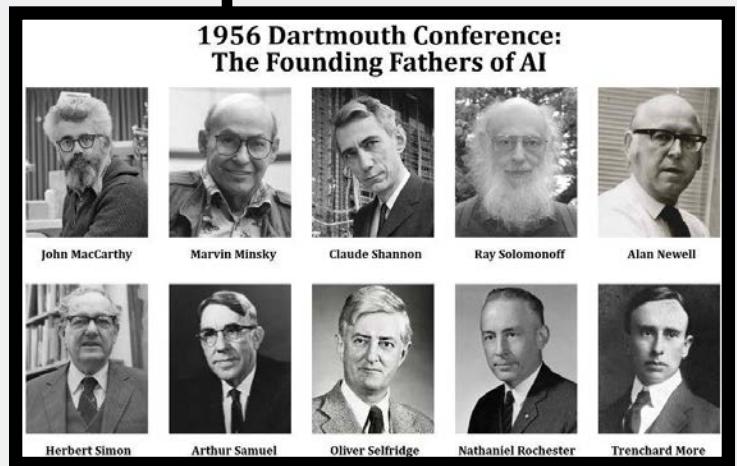
2011

Time



First wave AI: Model-based

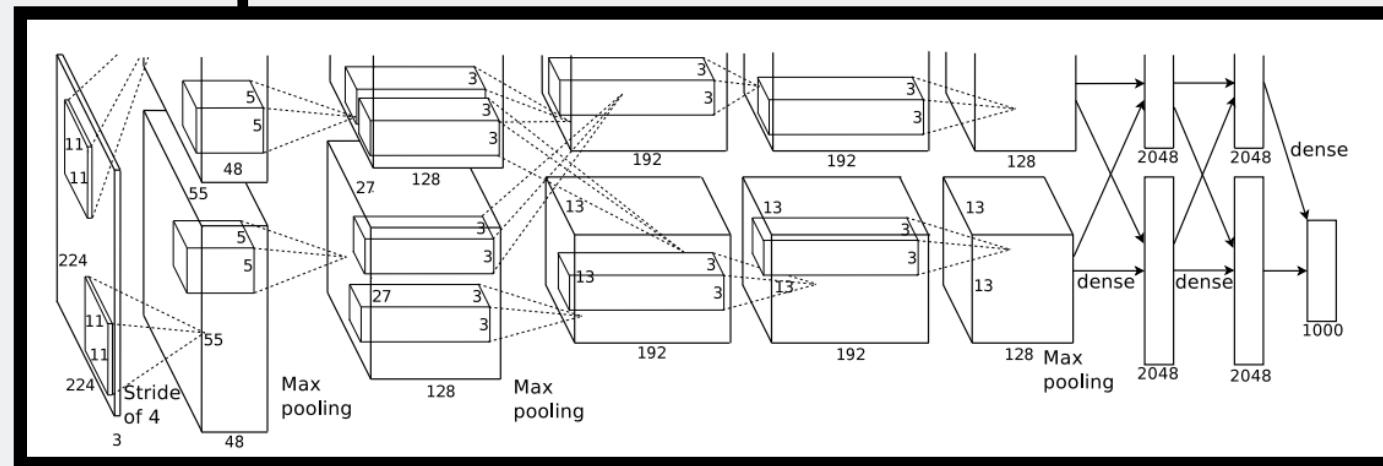
“Think through the entire problem”



1956

Second wave AI: Data-driven

“Learn from lots of data”



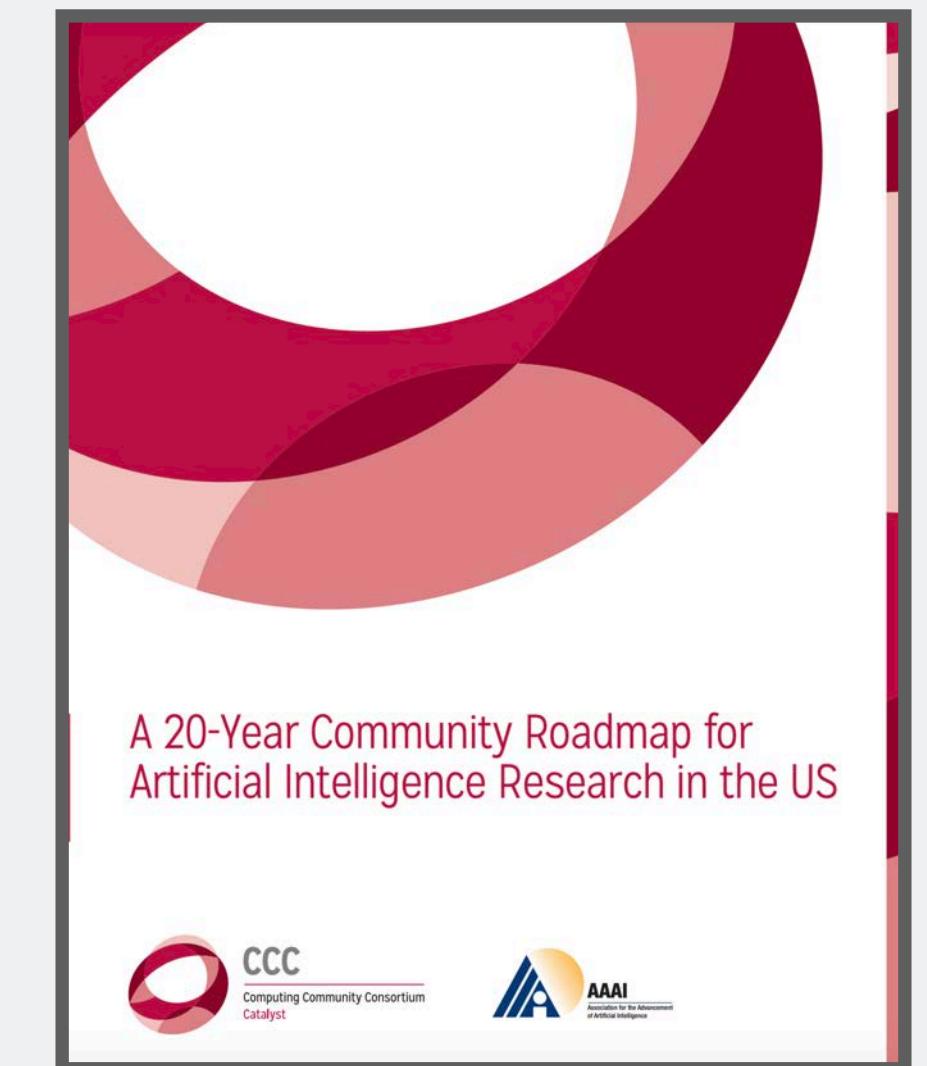
2011

Third wave AI: Explainable

“Combine first and second wave AI to generate explanations”



20??



Time

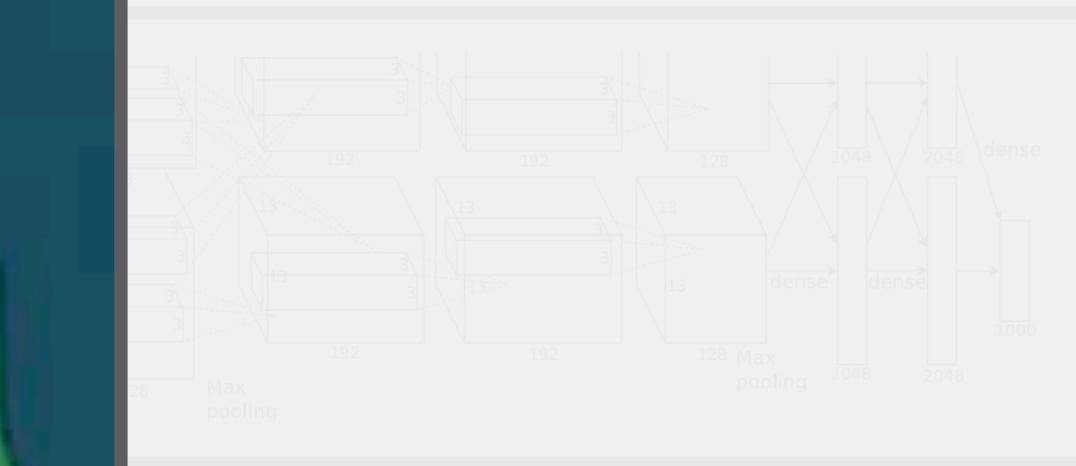
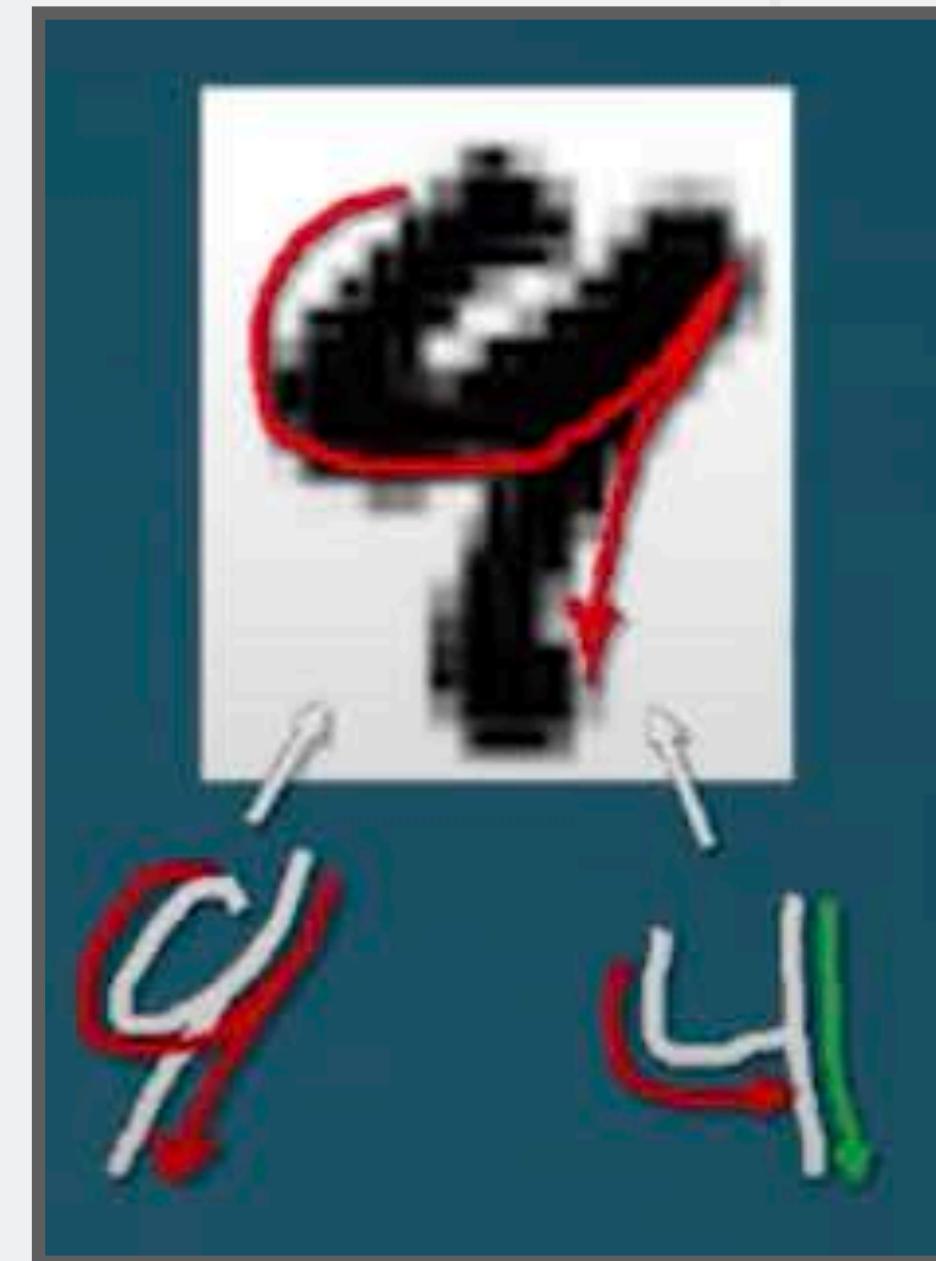




First wave AI: Model-based

“Think through the entire problem”

Is this handwritten character a ‘9’ or a ‘4’?



1956

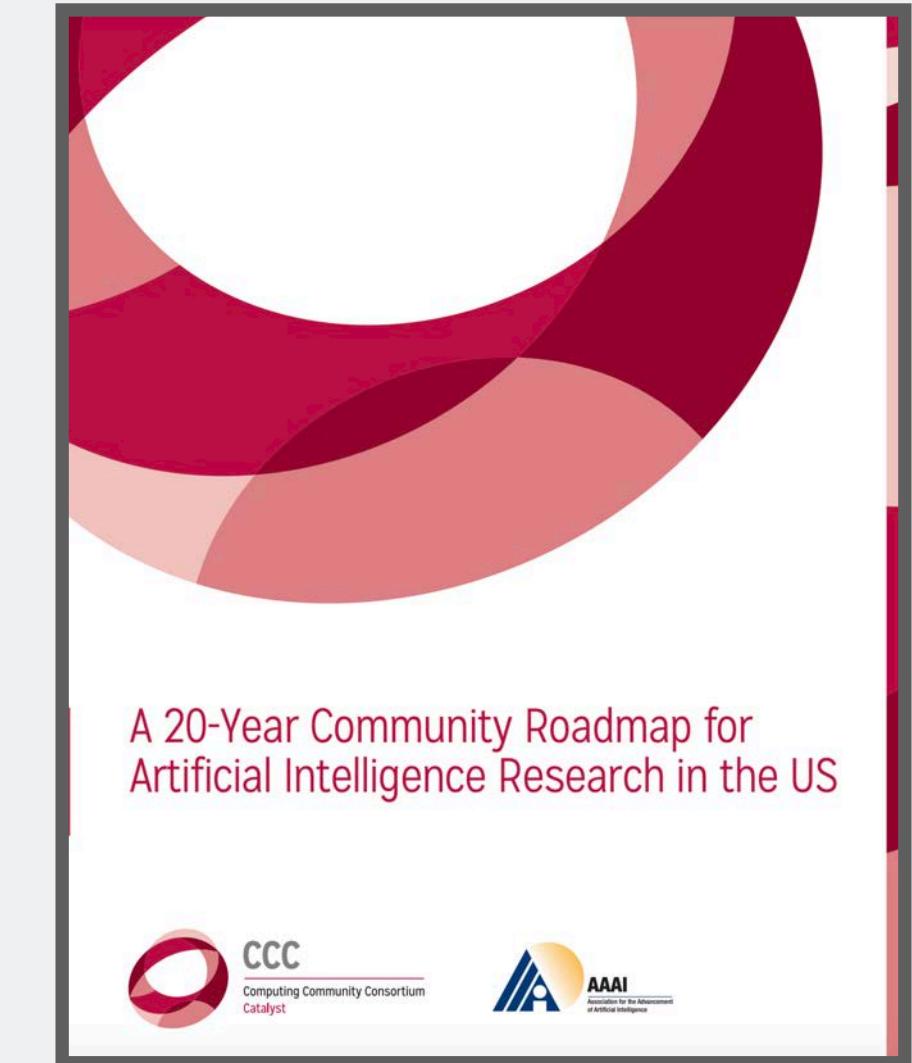
2011

20??

Time

Third wave AI: Explainable

“Combine first and second wave AI to generate explanations”





Probable number of strokes: 1 - 4
Each stroke: probable trajectory
Each trajectory: probable shift in shape and location

Seed model

Generative model
Generates explanations of how a test character might have been created

14:54 / 16:11

A DARPA Perspective on Artificial Intelligence
241,472 views • Feb 15, 2017

DARPAtv 143K subscribers

5.5K 130 SHARE SAVE

SUBSCRIBE

1956

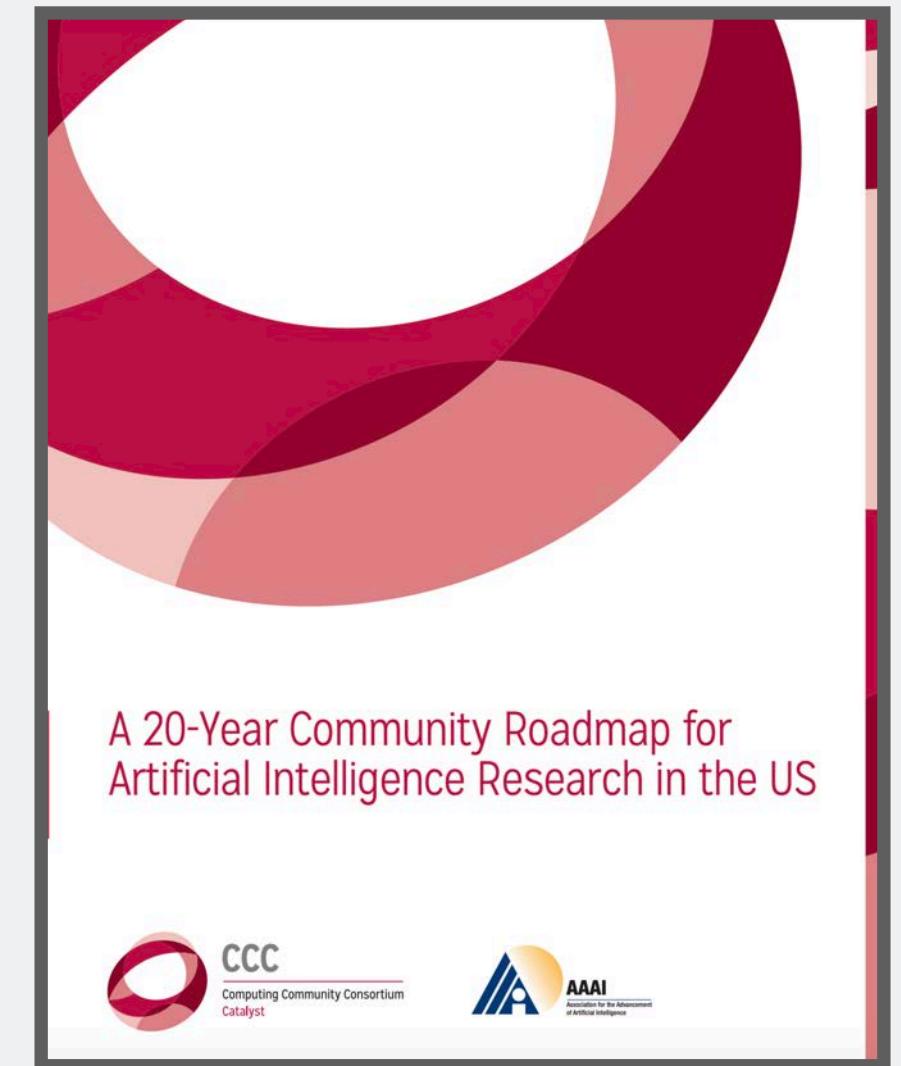
2011

20??

Time

Third wave AI: Explainable

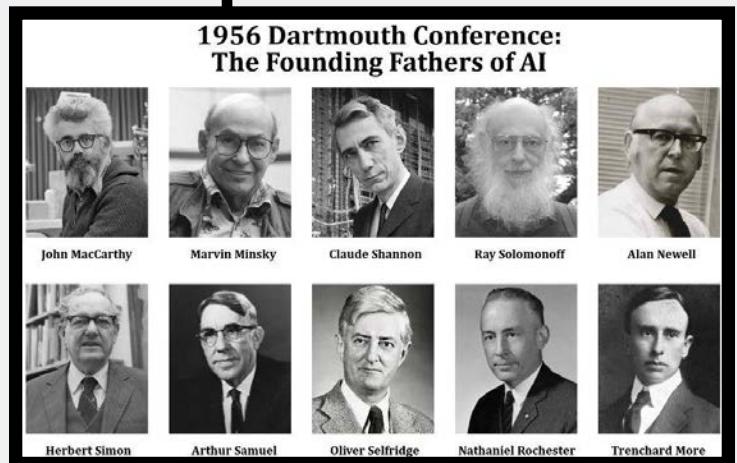
"Combine first and second wave AI to generate explanations"





First wave AI: Model-based

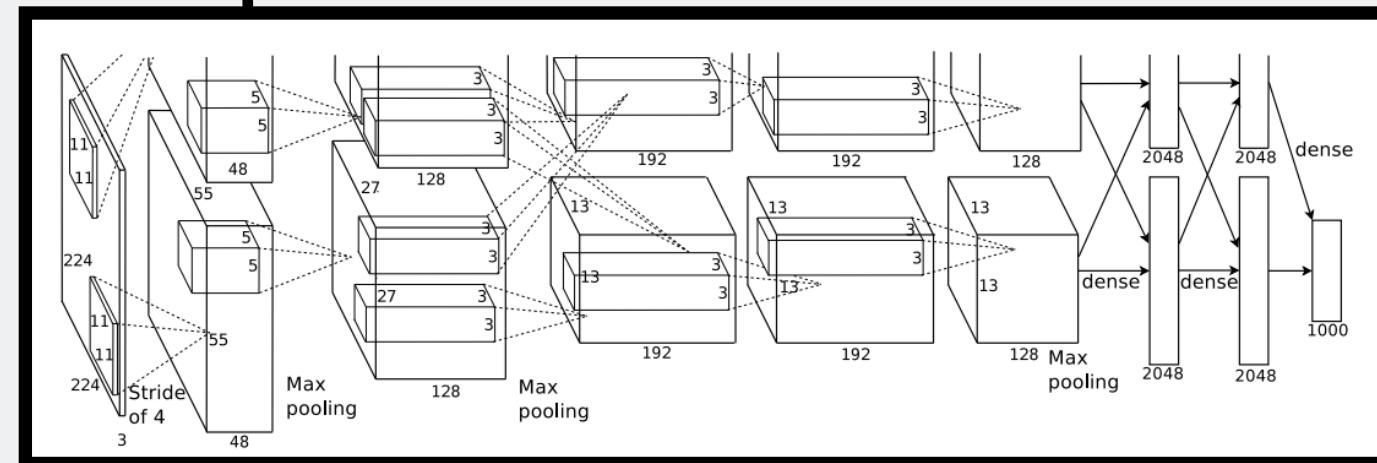
“Think through the entire problem”



1956

Second wave AI: Data-driven

“Learn from lots of data”



2011

Third wave AI: Explainable

“Combine first and second wave AI to generate explanations”



20??



Time

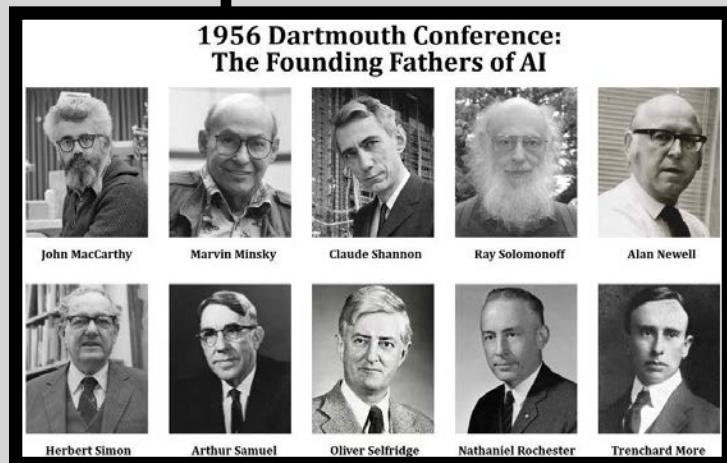




Other Robotics and AI courses

First wave AI: Model-based

“Think through the entire problem”

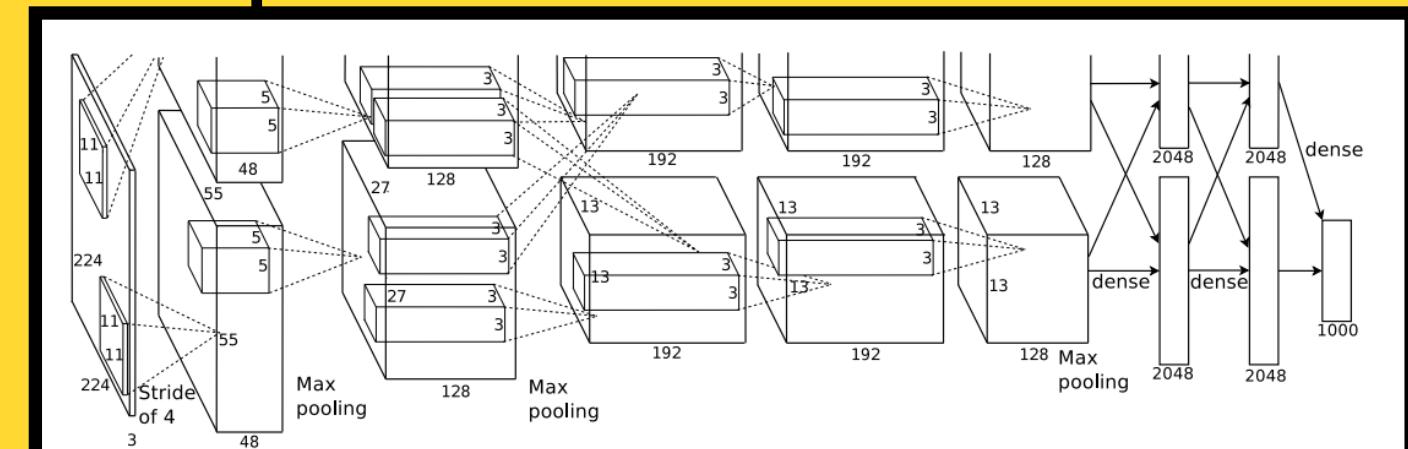


1956

DeepRob is a step into modern robot learning

Second wave AI: Data-driven

“Learn from lots of data”

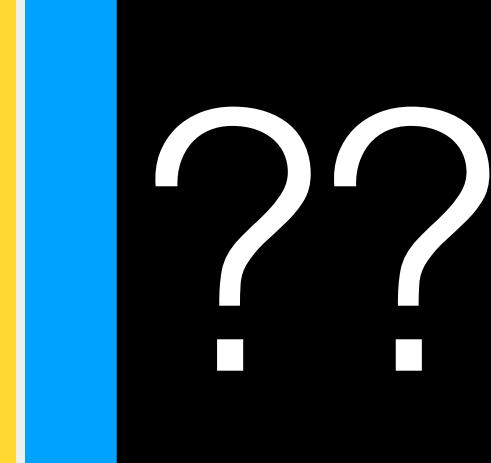
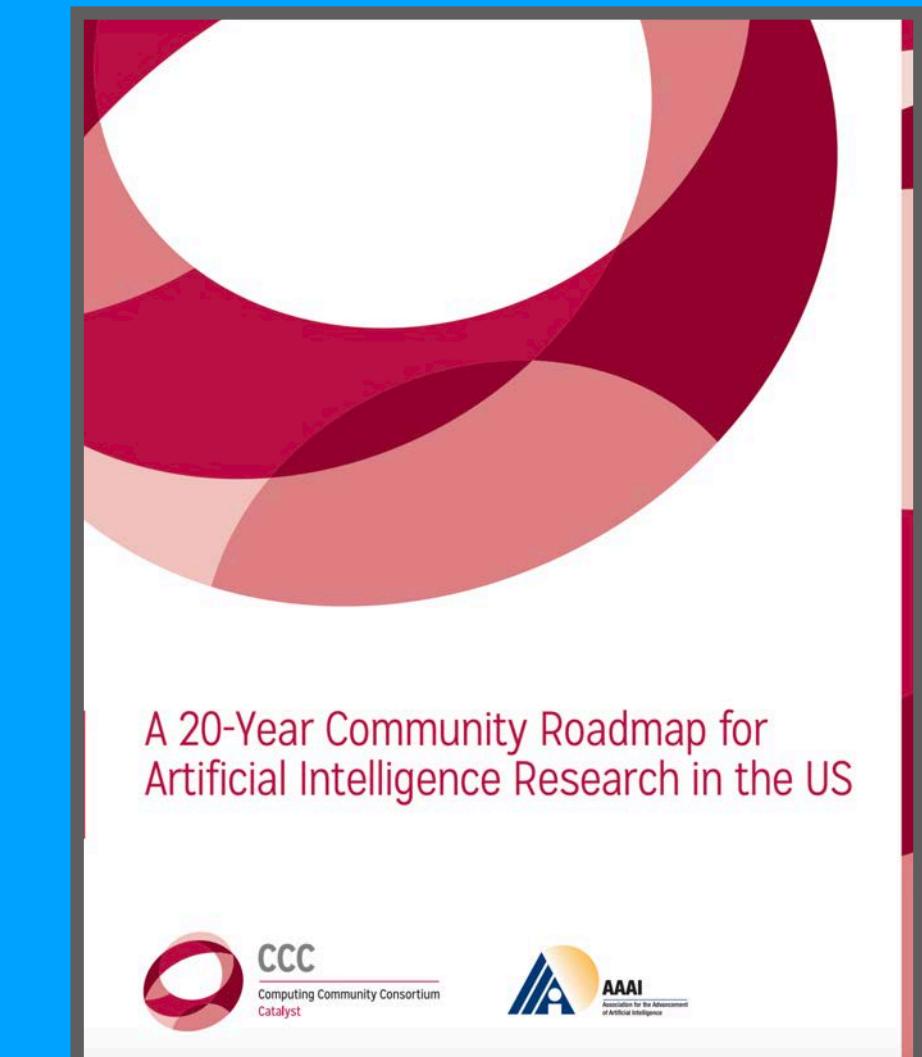


2011

Research for future AI

Third wave AI: Explainable

“Combine first and second wave AI to generate explanations”



20??

Time



Course Structure

- **Objective:** Give you the computational skills to understand and reproduce emerging applications of deep learning in robotics
- **Project focused class**
 - 5 total projects: building in complexity from basic linear layers to state-of-the-art research topics



Research Topics

RGB-D Architectures

Pointcloud Processing

Object Pose, Geometry, SDF, Implicit Surfaces

Dense object descriptors, Object category-level representations

Recurrent Networks and Object Tracking

Semantic Scene Graphs and Explicit Representations

Manipulation Policy learning

Diffusion models

Neural Radiance Fields and Implicit Representations

Datasets

Self-supervised Learning

Grasp Pose Detection

Tactile Perception for Grasping and Manipulation

Transformer Architectures

Benchmarking & Evaluation in Robot Learning

More frontiers ...



Project Topics

- Project 0
 - Introduction to Python, PyTorch and Google Colab
- Project 1
 - Classification using K-Nearest Neighbors and Linear Models
- Project 2
 - Classification using Fully-Connected and Convolutional Neural Networks
- Project 3
 - Detection using convolutional neural networks
- Project 4
 - State estimation with deep networks
- Project 5
 - Imitation learning



Final Project Tasks throughout the Semester

- Published paper(s) review, in-class lecture, reproducing an existing method, implementation, evaluation, demonstration - of a *deep learning model for robot manipulation*
- Work as collaborative teams up to 3 people





Course Resources

- A canvas announcement will be made with following details
 - Course website
 - Calendar
 - Syllabus
 - Projects
 - Final Project
 - EdStem discussion board





Meeting Logistics

- In-person Lectures
 - Mon & Wed 9:45 am - 11:00 am CT
- Appleby Hall 102
- Office Hours
 - Fri, 10:00 am - 12:00 pm
159 Shepherd Labs

Starting 09/06





Lets brainstorm on what your fav
robot should do!!!





Next Lecture: Classification

Image Classification and Linear Classifiers



DR

DeepRob

Fall 2024

Deep Learning for Robot Manipulation
University of Minnesota

