6.S099: Artificial General Intelligence 2018

https://agi.mit.edu

Lex Fridman

Lecture:

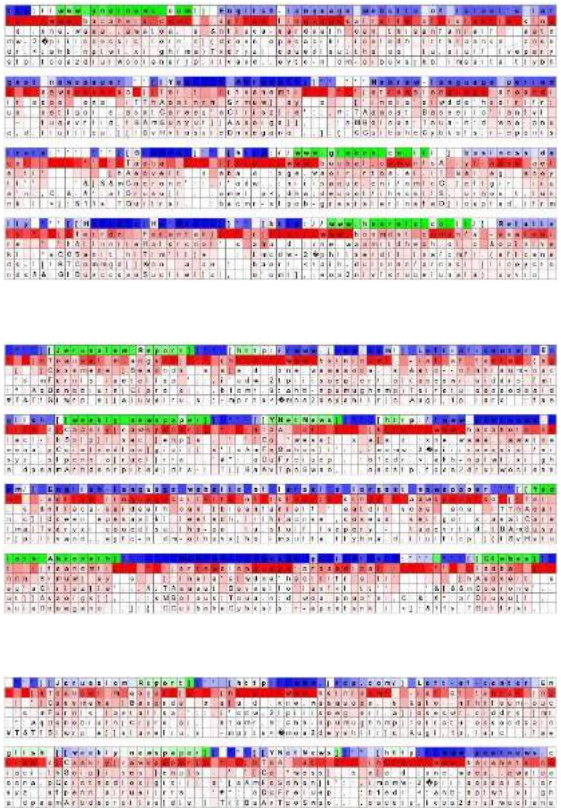
Artificial General Intelligence

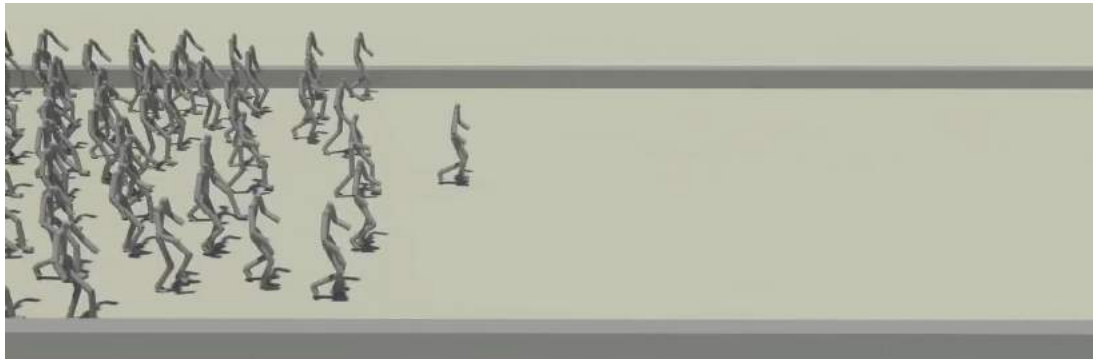
最专业报告分享群：

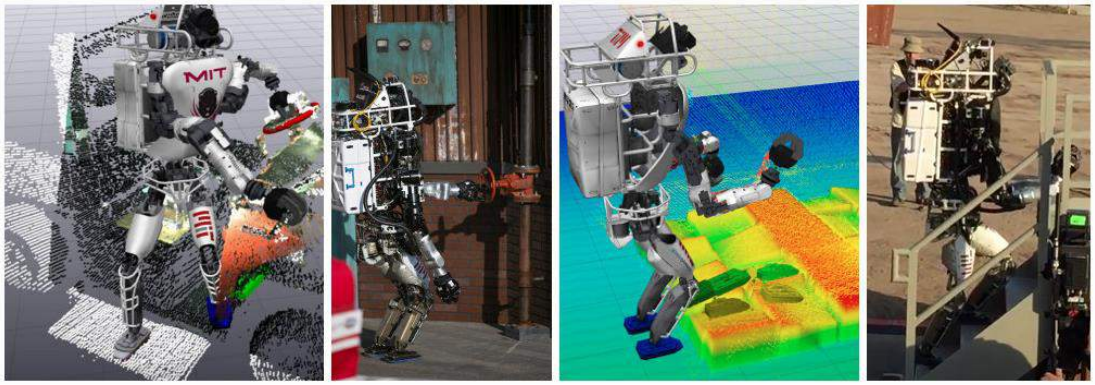
- 每日分享5+科技行业报告
- 同行业匹配，覆盖人工智能、大数据、机器人、智慧医疗、智能家居、物联网等行业。
- 高质量用户，同频的人说同样的话

扫描右侧二维码，
或直接搜索关注公众号：智东西（zhidxcom）
回复“报告群”加入

MIT AGI Mission: Engineer Intelligence







MIT motto: ***Mind and Hand***

MIT AGI Mission: Engineer Intelligence

Provide a balance to “black box” reasoning

Impact on

Future Society

(2028+)

Today’s Society

(2018)

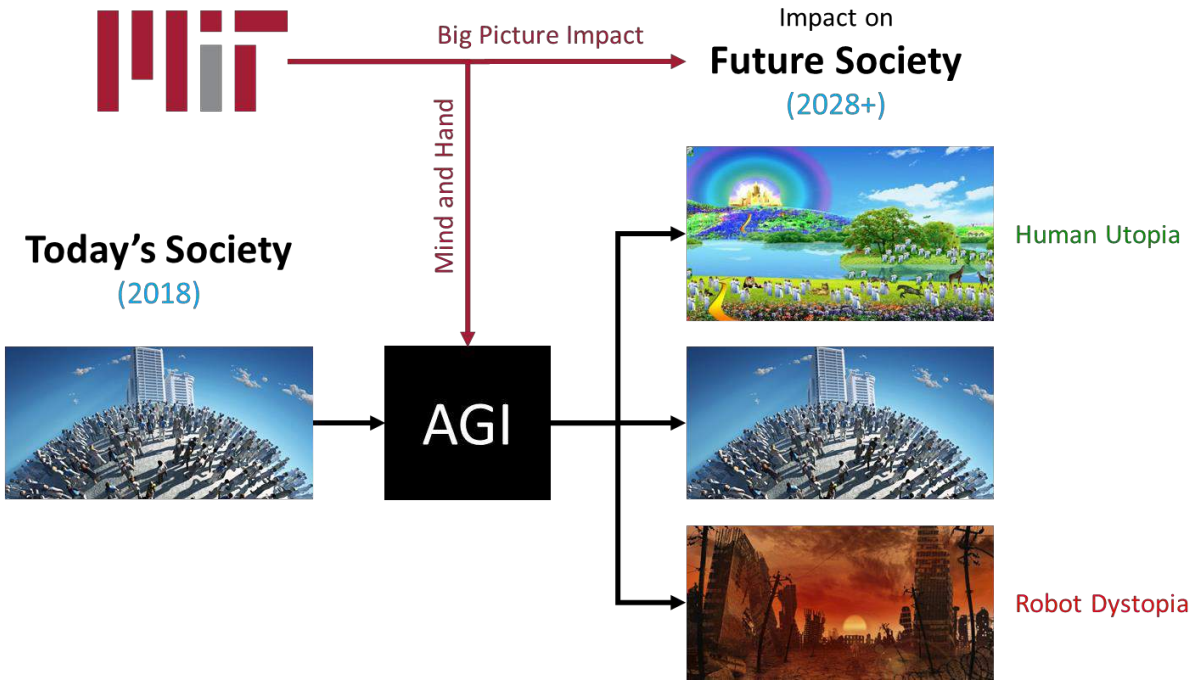
AGI

Human Utopia

Unchanged

Robot Dystopia

Balance Between Paralyzing Technophobia and Blindness to Big Picture Consequences



- Defining metric of disagreement is an **engineering question**: How hard is to create human-level artificial intelligence?
  - Can we build intuition about this without knowing how to build it?

MIT AGI Mission: Engineer Intelligence

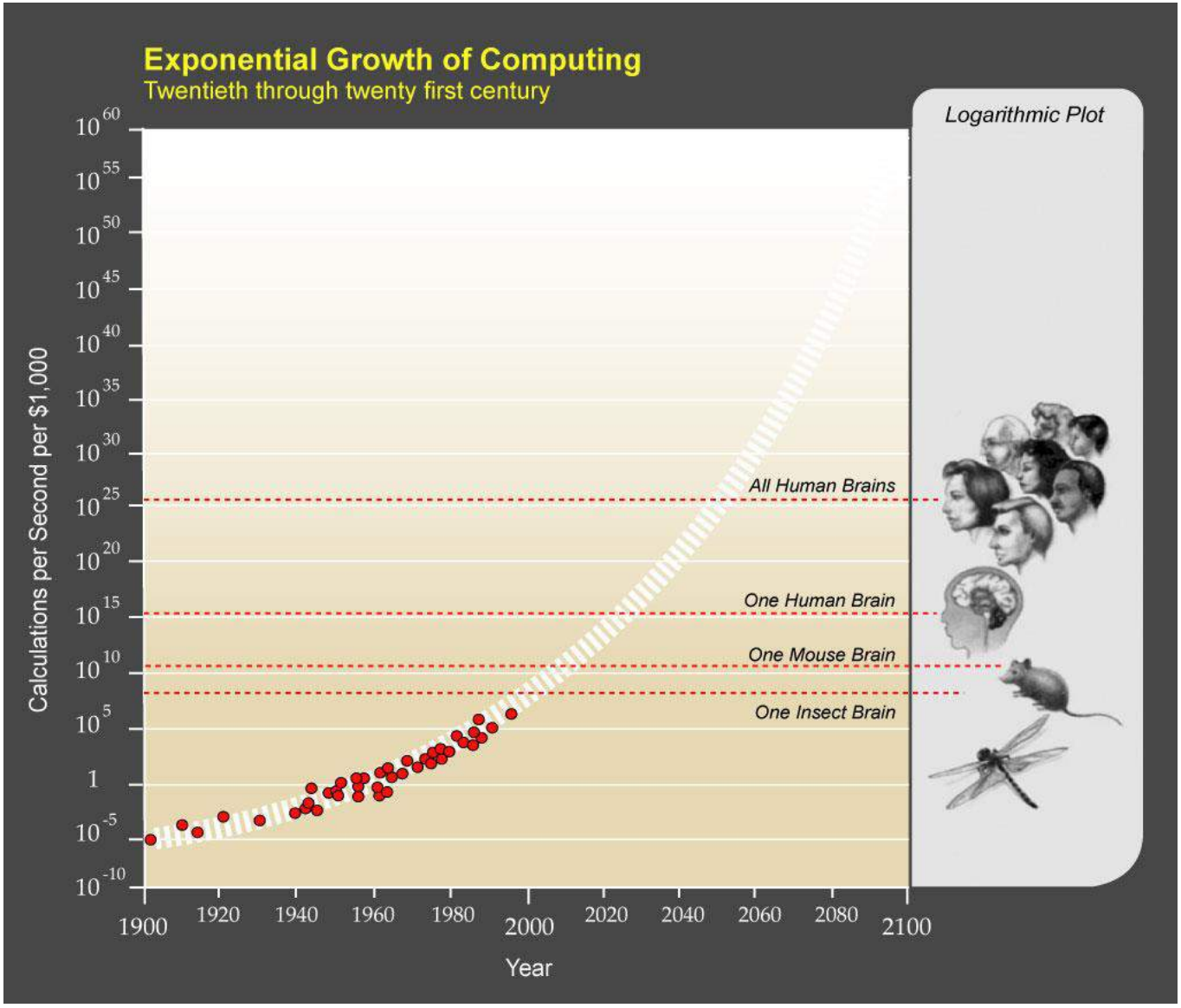
- **Goal 1:** Avoid the pitfalls of "black box" futurism thinking that results in hype that is detached from scientific understanding
- **Goal 2:** Avoid the pitfalls of "I'm just a scientist" that results in ignorance to near-term negative consequences that are preventable with good engineering

Google Intern :

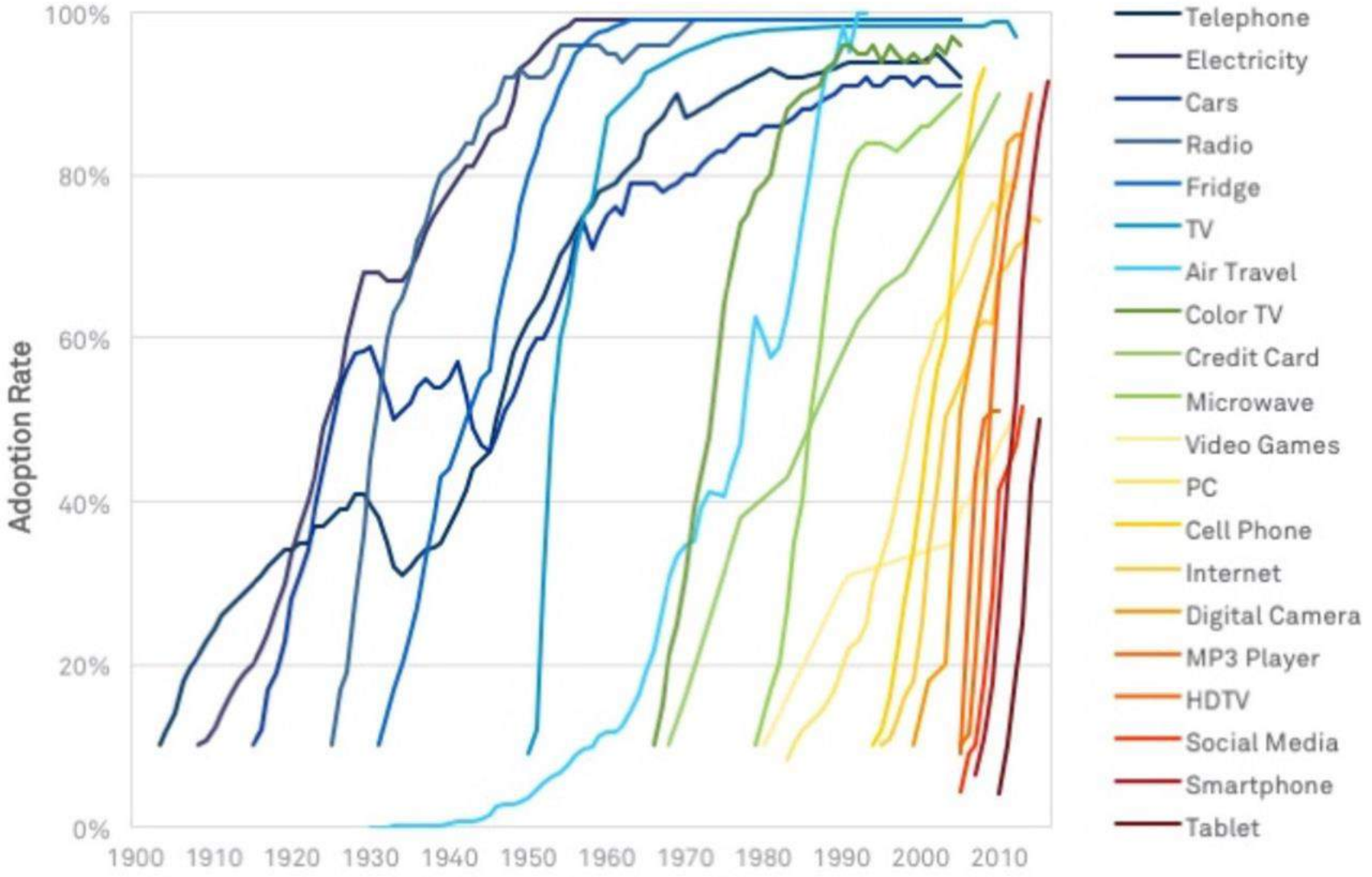
```python
from keras.layers import *
from keras.models import *
from .data import load_data

x, y, x_test, y_test = load_data()

def get_model(num_layers):
    model = Sequential()
    for _ in range(num_layers):
        model.add(Dense(100, activation='sigmoid'))
    model.compile(loss='mse', optimizer='sgd')
    return model

best_model = None
best_loss = None

for i in range(1, 10):
    model = get_model(i)
    model.fit(x, y)
    loss = model.evaluate(x_test, y_test)
    if best_loss is None or loss < best_loss:
        best_loss = loss
        best_model = model
```

Media :

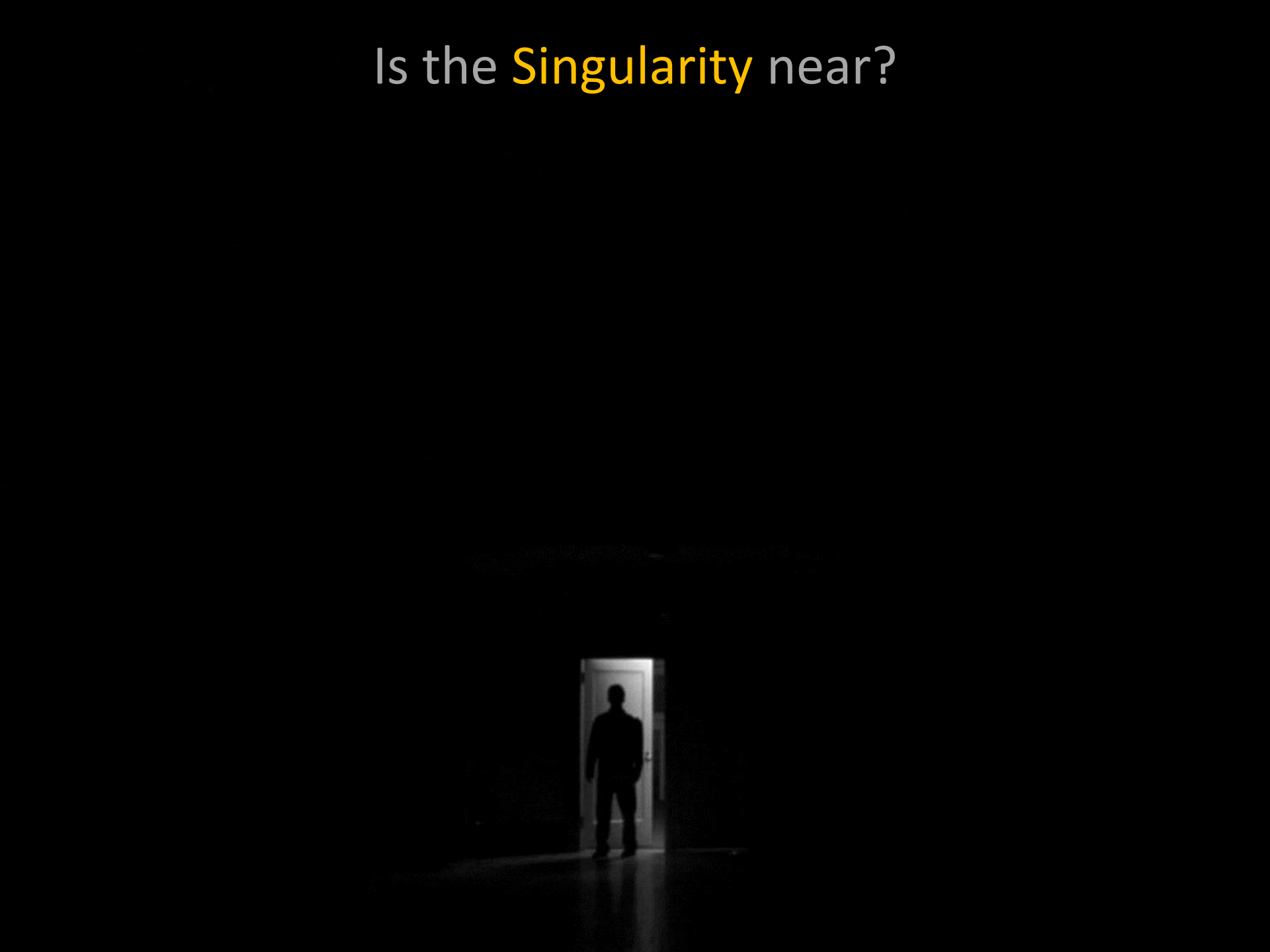
Kurzweil's Law of Accelerating Returns



Increasingly Faster Adoption of New Technology



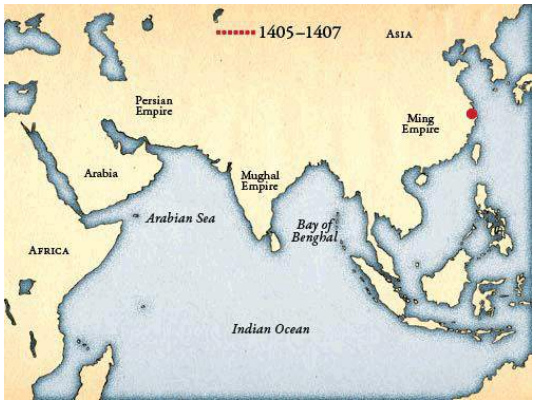
Is the Singularity near?



Human Drive to Explore and Uncover the Mysteries of the Universe

- What drives humans to explore the unknown?
  - "For all the different forms it takes in different historical periods, for all the worthy and unworthy motives that lie behind it, exploration—travel for the sake of discovery and adventure—is a human compulsion, a human obsession even; it is a defining element of a distinctly human identity, and it will never rest at any frontier, whether terrestrial or extra-terrestrial."
    – Stewart Weaver, Exploration: A Very Short Introduction

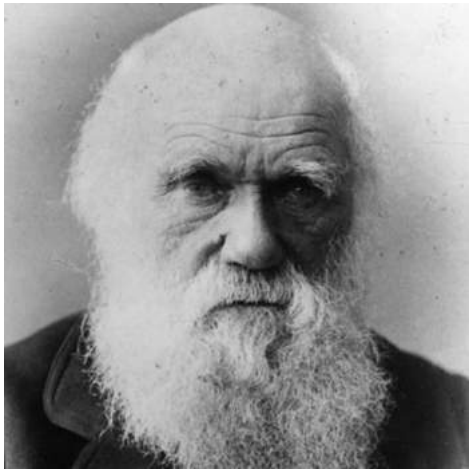
Pytheas of Massalia, 325 B.C.E.
7,500 miles of ocean travel for first known reporter of the Arctic

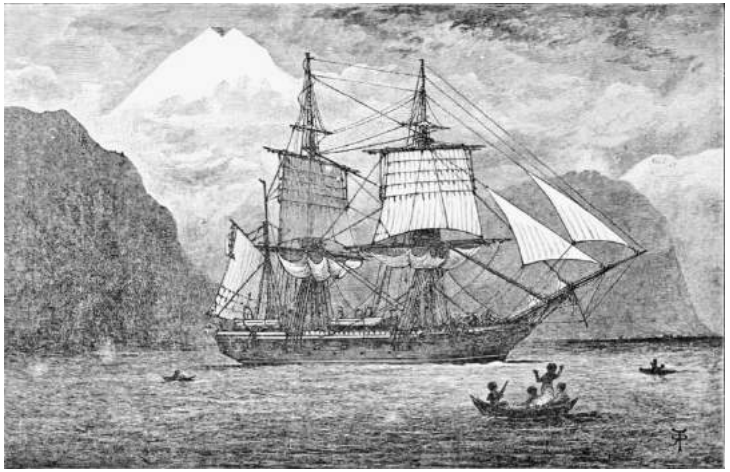


Zheng He, 1405-1433
China's imperial expeditions. Treasure voyages. 7 expeditions. First had 287 vessels and 27,780 men.

Christopher Columbus, 1492
Paved the way for European colonization of Americas. PS1: Didn't "discover" Americas. PS2: His approach receives harsh criticism modern scholarship.

Human Drive to Explore and Uncover the Mysteries of the Universe





"Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. …
Whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved."

Charles Darwin's Voyage of the Beagle (sailed in 1831)

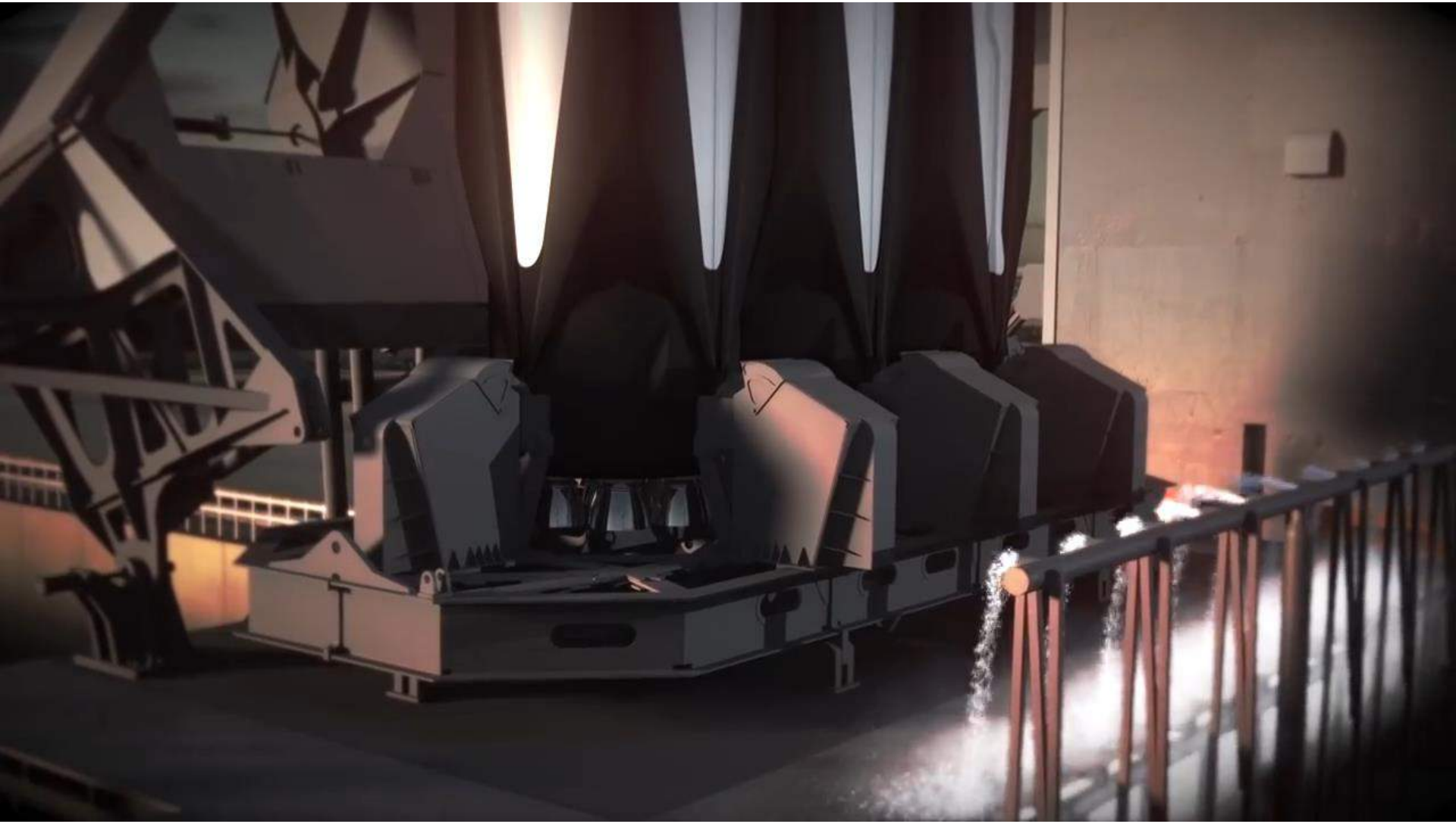
Human Drive to Explore and Uncover the Mysteries of the Universe

"The Earth is blue ... it is amazing,"

Yuri Gagarin, first human in space (April 12, 1961).

Human Drive to Explore and Uncover the Mysteries of the Universe

Human Drive to Explore and Uncover the Mysteries of the Universe



Artificial General Intelligence

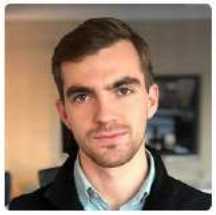
6.S099: Artificial General Intelligence

Lex Fridman

Instructor

Michael Glazer

TA



Jack Terwilliger

TA

Li Ding

TA

Dan Brown

TA

Julia Kindelsberger

TA

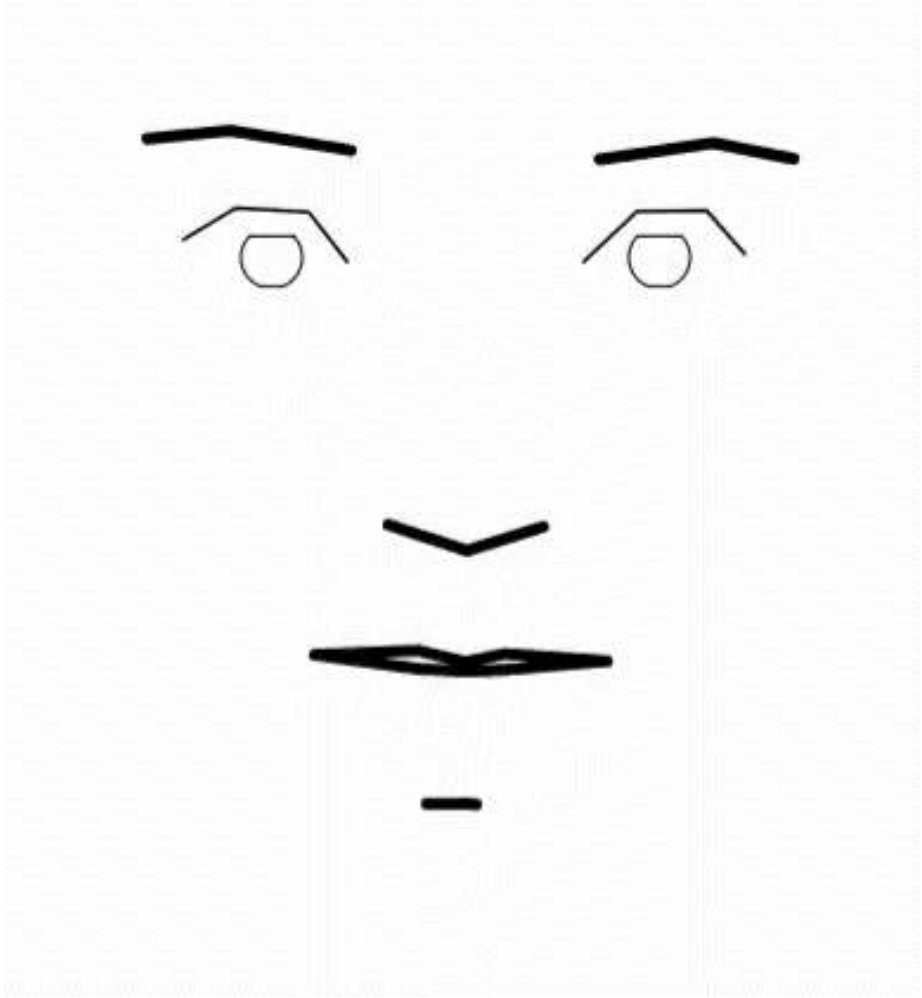
- **Website:** agi.mit.edu
- **Email:** agi@mit.edu
- **Slack:** deep-mit.slack.com
- **For registered MIT students:**
  - Create an account on the website.
  - Submit 5 new links to VoteAI and vote on 10.
  - Submit entry to one of the competition
- **Projects**
  - DreamVision
  - ANGEL
  - EthicalCar
  - VoteAI
- **Guest Speakers** (see schedule)
- **Shirts** free in-person, available online: https://teespring.com/agi-2018

DreamVision

https://agi.mit.edu/dreamvision

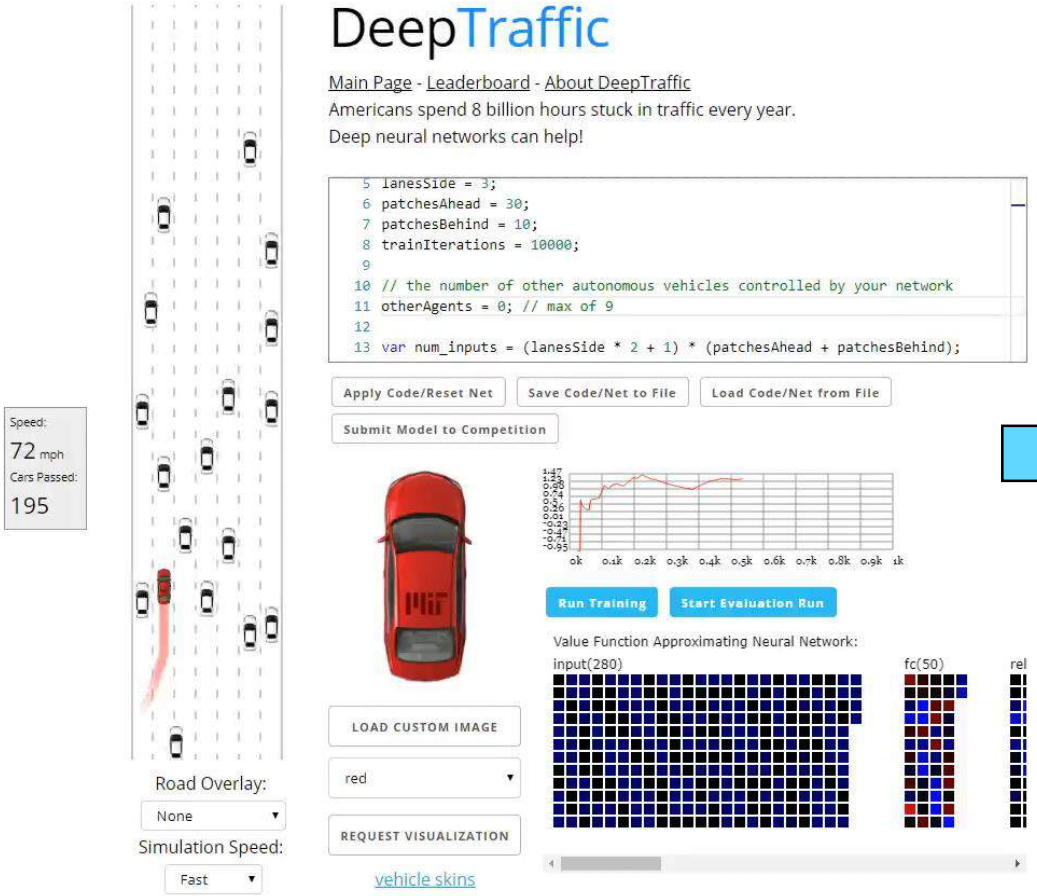
ANGEL: Artificial Neural Generator of Emotion and Language

https://agi.mit.edu/angel



EthicalCar: Machine Learning Approach

https://agi.mit.edu/ethicalcar



Add human life to the loss function.

VoteAI

https://agi.mit.edu/vote-ai

42 Vote on how informative the article is (whether you agree with it or not).

P 23 Click if you believe the article is more **Postive** toward the future impact of AI.

N 16 Click if you believe the article is more **Negative** toward the future impact of AI.

Title:

Link:

Submit

Lectures and Guest Talks

Lecture Mon, Jan 22, 7pm Room 54-100
Artificial General Intelligence
[Slides] - [Lecture Video] *(Available Soon)*

Guest Talk Tue, Jan 23, 7pm Room 54-100
Josh Tenenbaum: Computational Cognitive Science
Professor, MIT

Guest Talk Wed, Jan 24, 1pm Room 10-250
Ray Kurzweil: How to Create a Mind
Google

Guest Talk Thu, Jan 25, 7pm Room 54-100
Lisa Feldman Barrett: Emotion Creation
Northeastern University

Guest Talk Fri, Jan 26, 7pm Room 54-100
Nate Derbinsky: Cognitive Modeling
Northeastern University

Guest Talk Mon, Jan 29, 1:30pm Room 26-100
Andrej Karpathy: Deep Learning
Director of AI, Tesla
Previously: OpenAI, Stanford University

Guest Talk Mon, Jan 29, 7pm Room 54-100
Stephen Wolfram: Knowledge-Based Programming
Wolfram Research

Guest Talk Tue, Jan 30, 7pm Room 54-100
Richard Moyes: AI Safety and Autonomous Weapon Systems
Co-Founder and Managing Director, Article36

Guest Talk Wed, Jan 31, 7pm Room 54-100
Marc Raibert: Robotics
CEO, Boston Dynamics
Previously: MIT

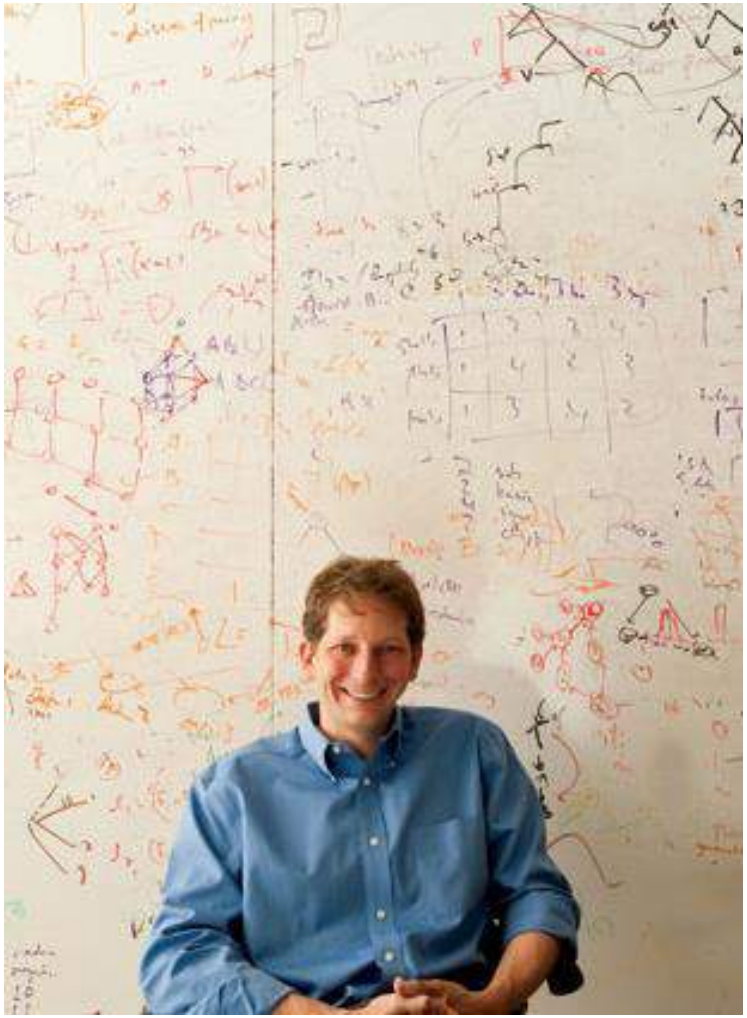
Guest Talk Thu, Feb 1, 7pm Room 54-100
Ilya Sutskever: Deep Reinforcement Learning
Co-founder, OpenAI
Previously: Google Brain, Stanford, U of Toronto

Lecture Fri, Feb 2, 7pm Room 54-100
Human-Centered Artificial Intelligence
[Slides] - [Lecture Video] *(Available Soon)*

Josh Tenenbaum, MIT

Computational Cognitive Science

Tue, Jan 23, 7pm (Room 54-100)

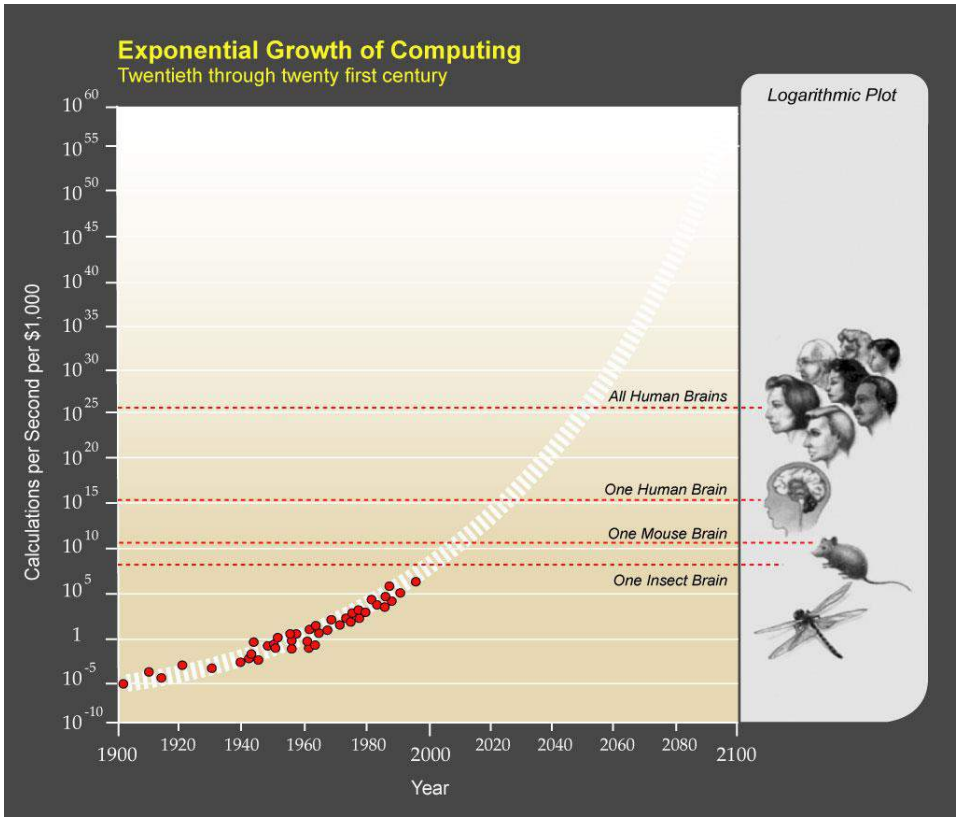


- **Common sense understanding:** How can we see a world of physical objects, their interactions and our own possibilities to act and interact with others ("intuitive physics") – not simply classify patterns in pixels?
- **Rapid model-based learning:** How can we learning new concepts from so little experience – often just a single example?
- Integrating best ideas of how to think about intelligence computationally:
  - Probabilistic generative models
  - Symbol-processing architectures
  - Neural networks

Ray Kurzweil, Google

Future of Intelligence, Artificial and Natural

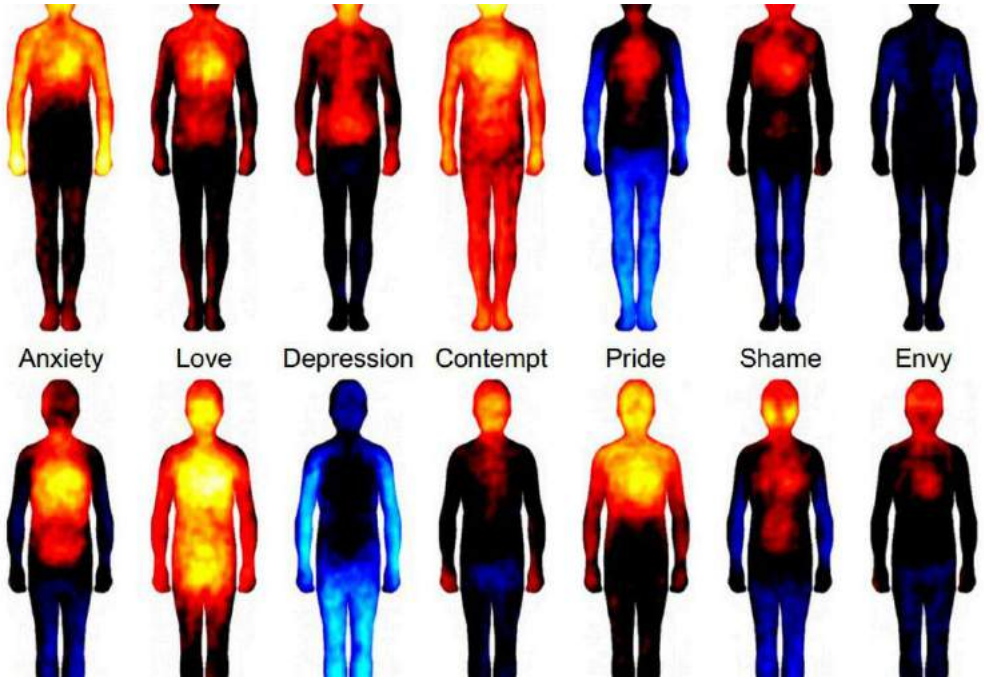
Wed, Jan 24, 1pm (Room 10-250)



Lisa Feldman Barrett, NEU

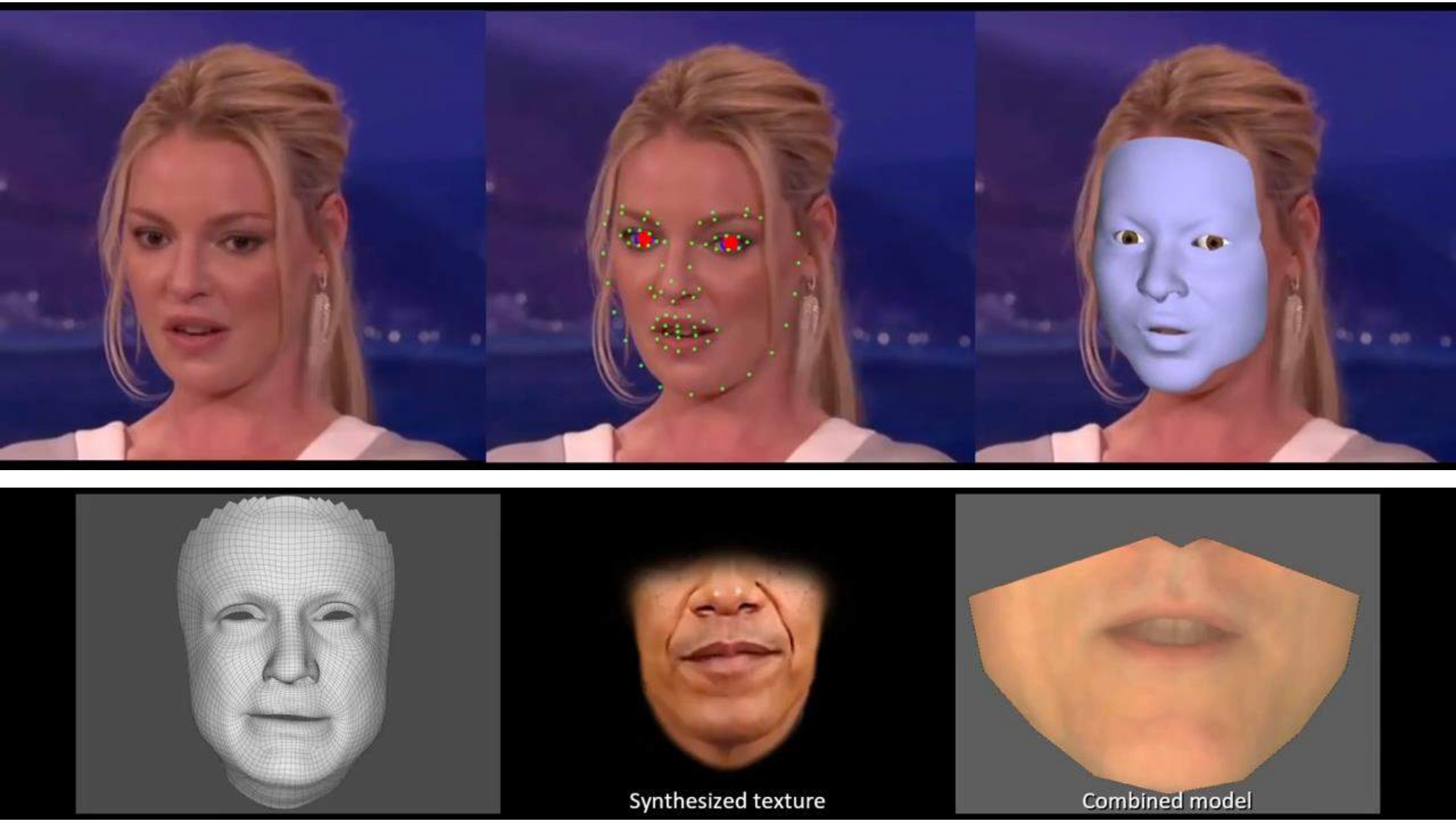
Emotion Creation

Thu, Jan 25, 7pm (Room 54-100)



Re-Enacting Intelligence

- Start with human data and manipulate its visual and auditory contents

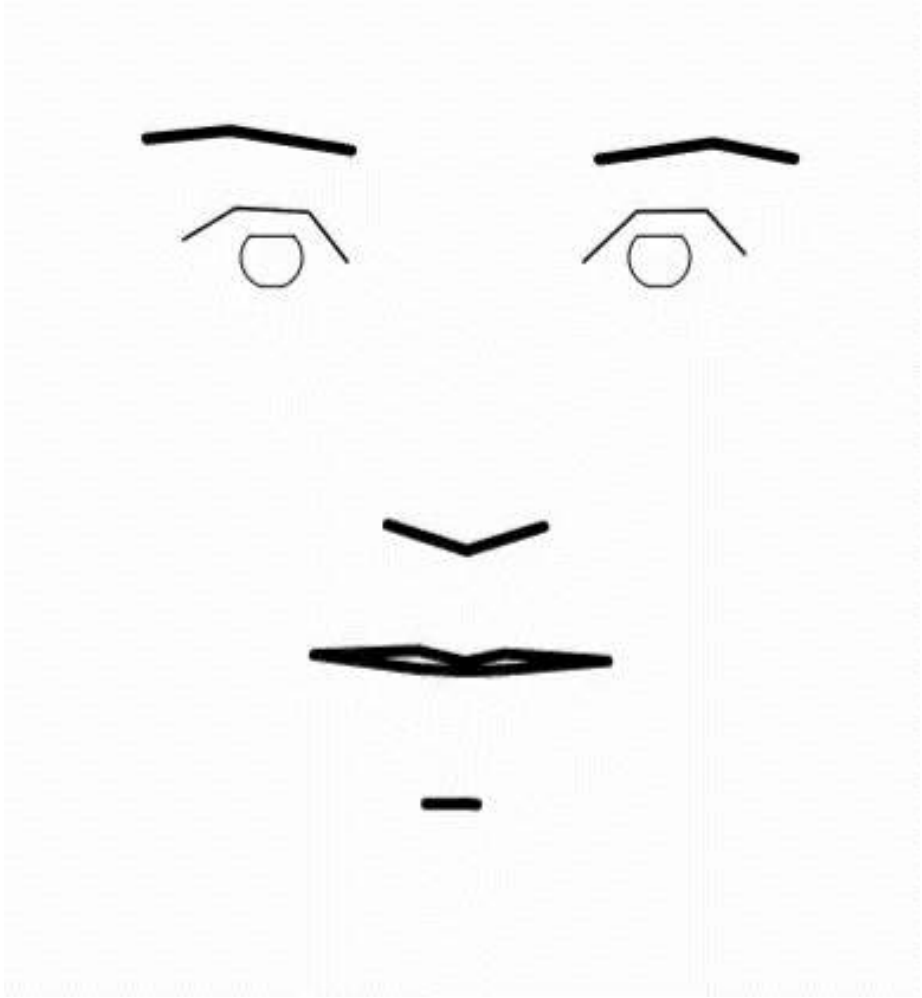


Sophia: Embodied Re-Enactment

(PS: Sophia is **not** a strong NLP system)

ANGEL: Artificial Neural Generator of Emotion and Language

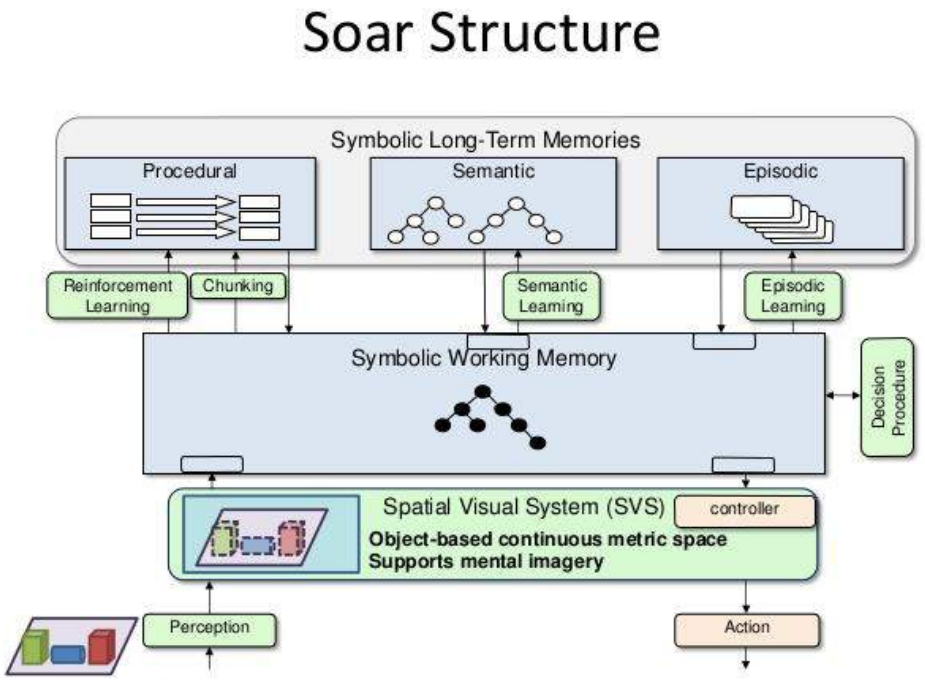
https://agi.mit.edu/angel



Nate Derbinsky, NEU

Cognitive Modeling

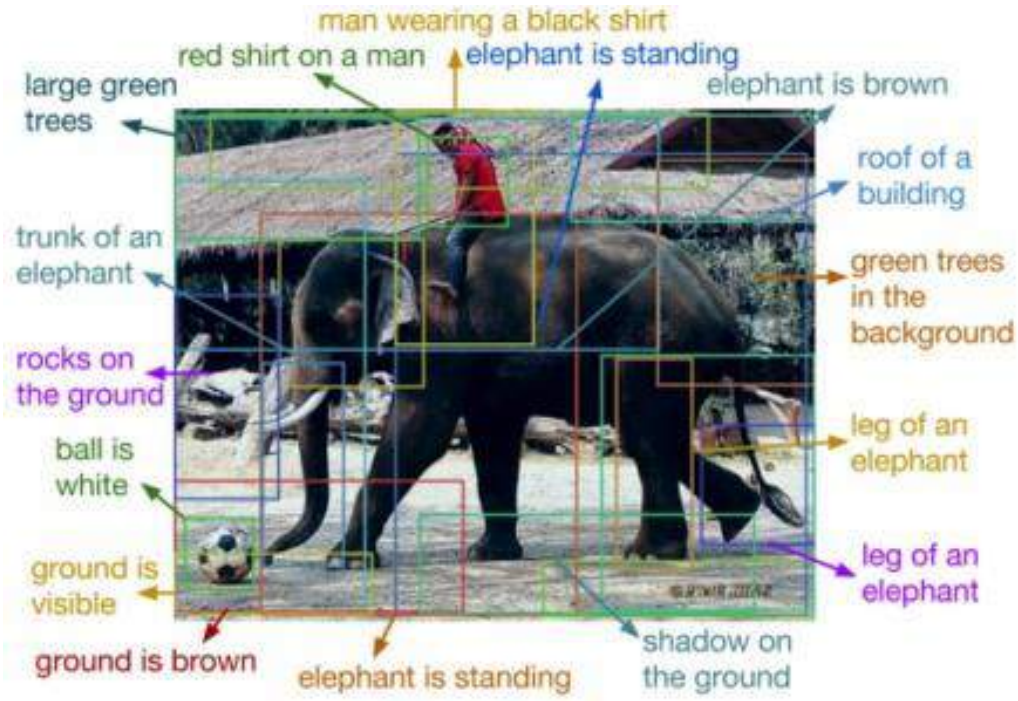
Fri, Jan 26, 7pm (Room 54-100)



Andrej Karpathy

Deep Learning

Mon, Jan 29, 1:30pm (Room 26-100)



Deep Learning:

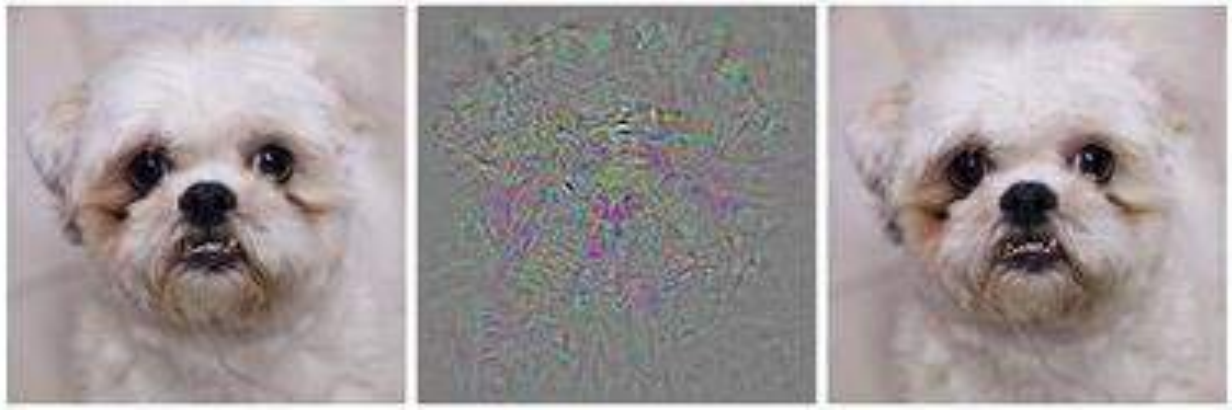
Our intuition about what's "hard" is flawed (in complicated ways)

"Encoded in the large, highly evolve sensory and motor portions of the human brain is a **billion years of experience** about the nature of the world and how to survive in it.... Abstract thought, though, is a new trick, perhaps less than **100 thousand years** old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it."
- *Hans Moravec, Mind Children (1988)*

Visual perception: 540,000,000 years of data

Bipedal movement: 230,000,000 years of data

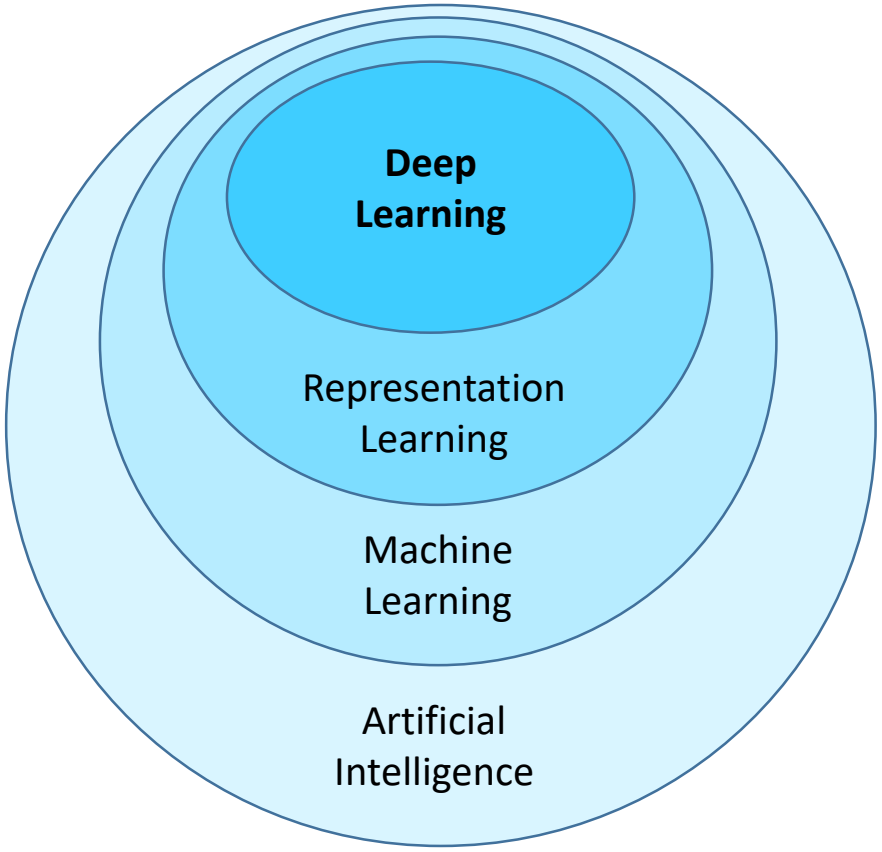
Abstract thought: 100,000 years of data

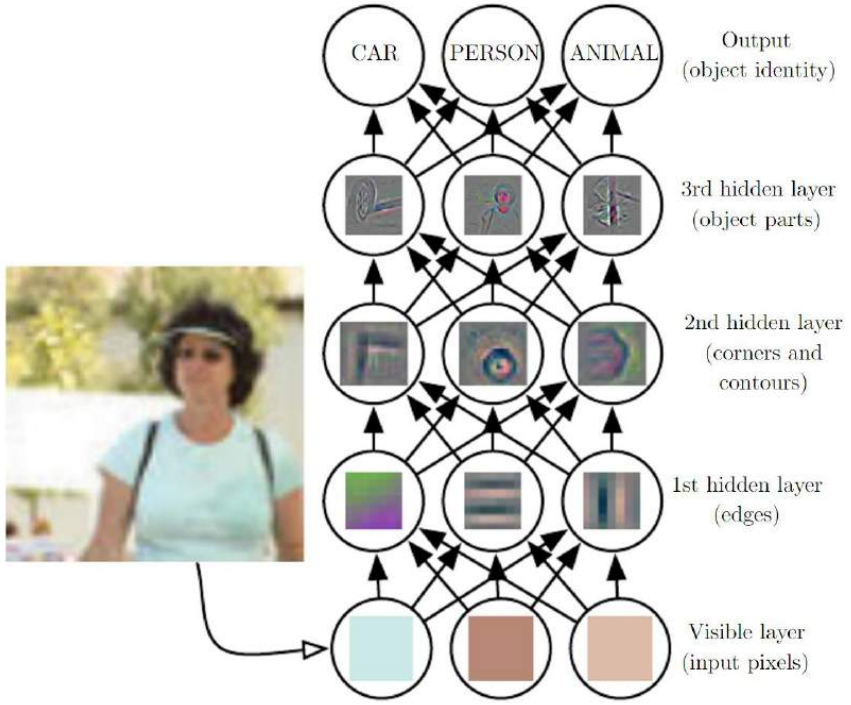


Prediction: **Dog** + Distortion Prediction: **Ostrich**

Deep Learning is **Representation Learning**

(aka Feature Learning)



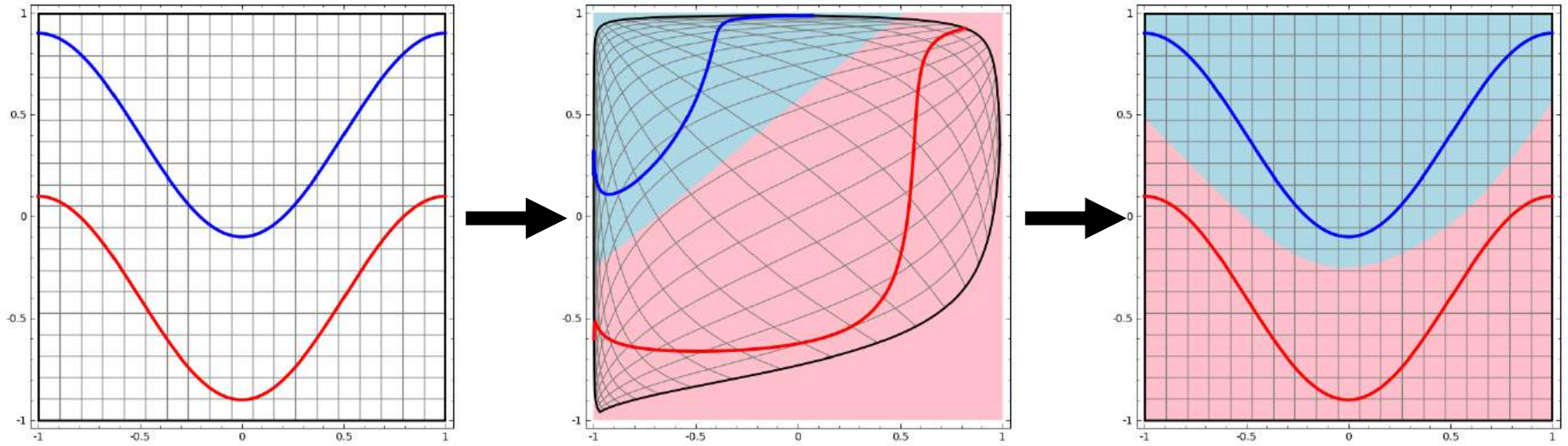


Intelligence: Ability to accomplish complex goals.

Understanding: Ability to turn complex information to into simple, useful information.

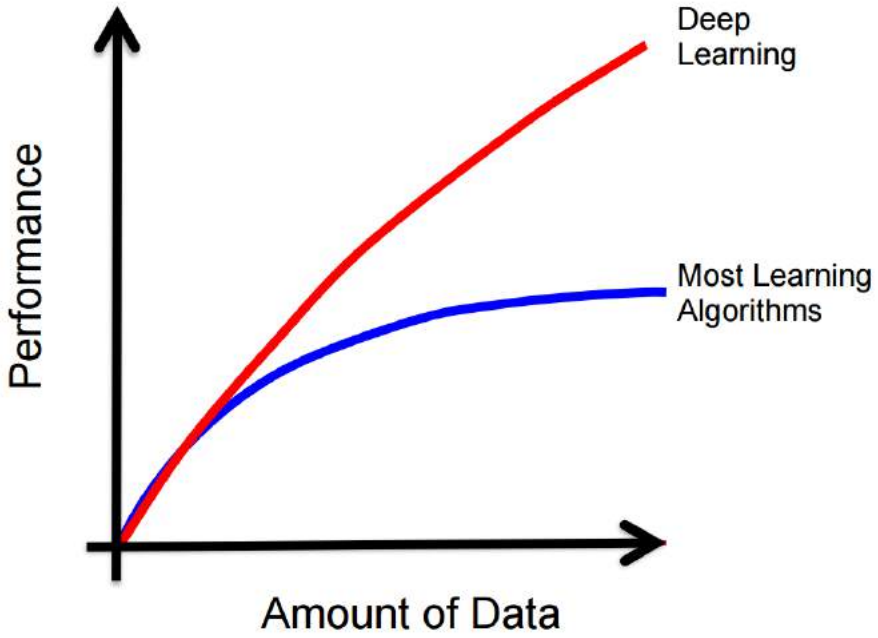
Deep Learning is **Representation Learning**

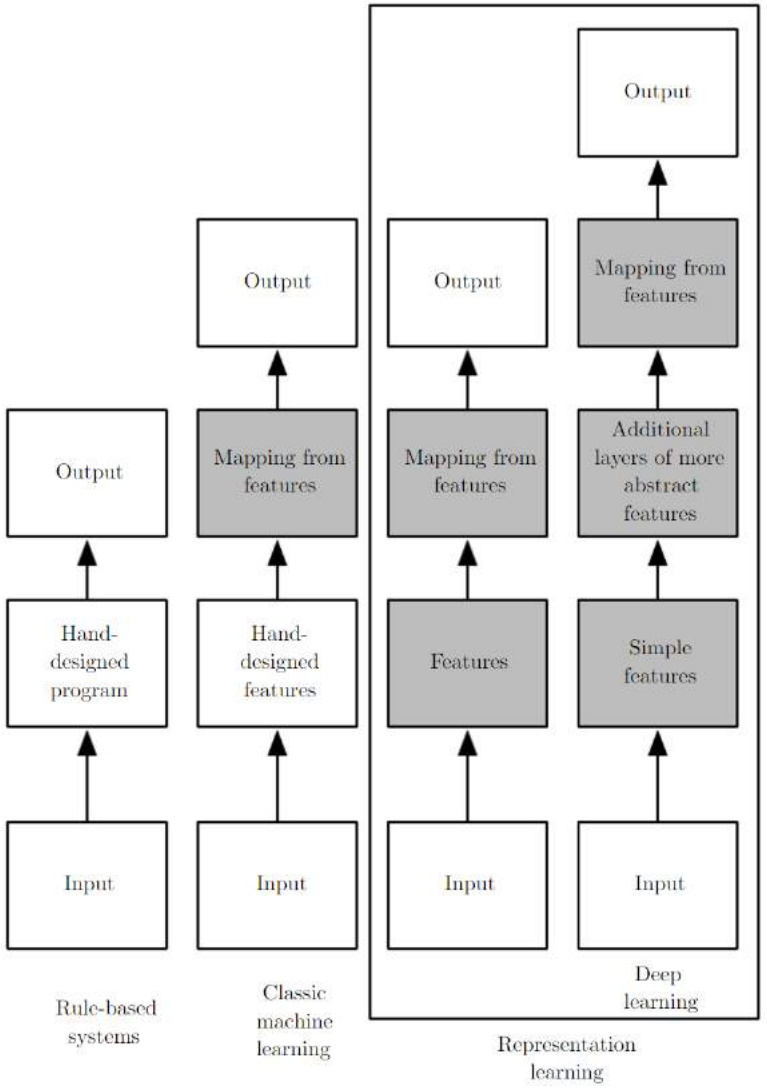
(aka Feature Learning)



Task: Draw a line to separate the **blue curve** and **red curve**

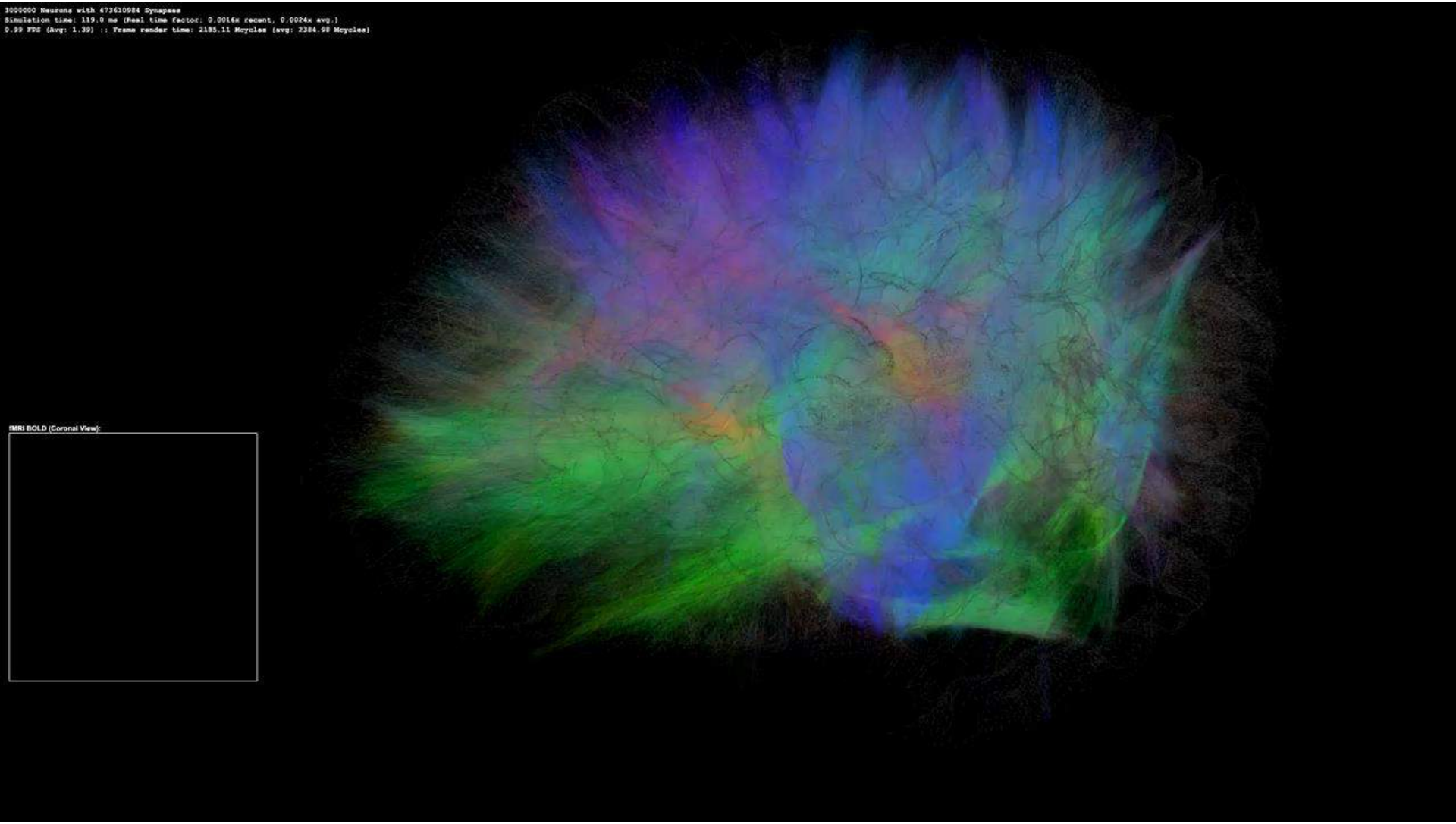
Deep Learning: **Scalable** Machine Learning





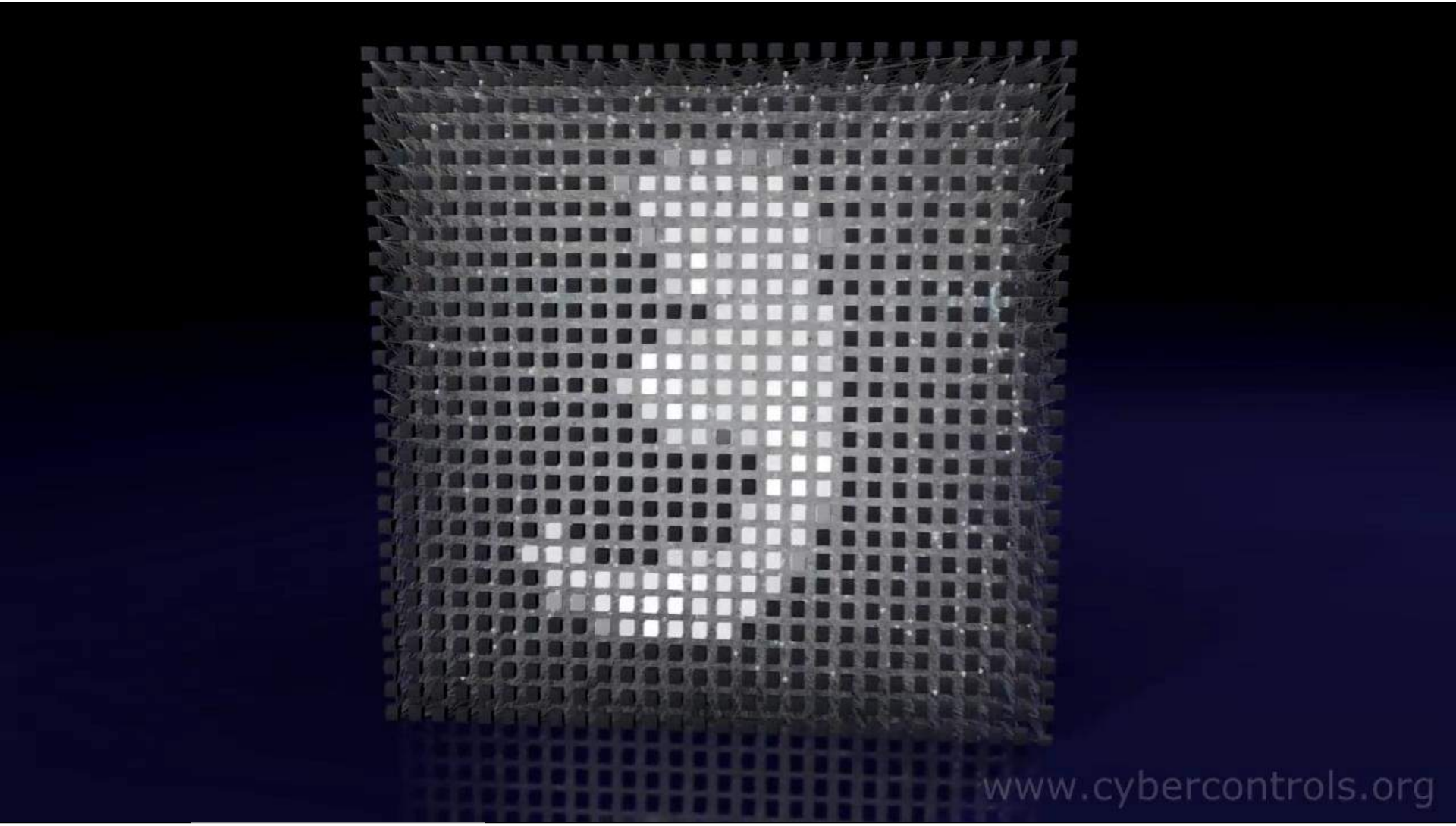
Biological Neural Network

- Thalamocortical brain network (simulation video shown below)
  - 3 million neurons, 476 million synapses
- Full human brain:
  - 100 billion neurons, 1,000 trillion synapses



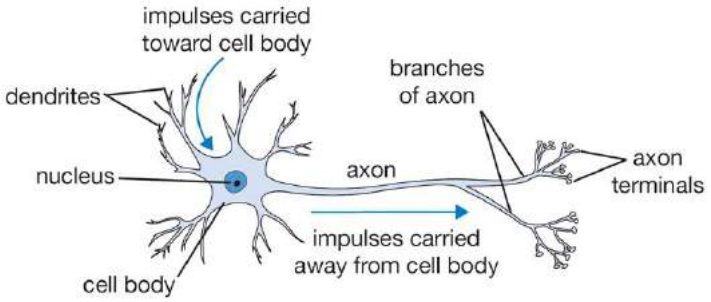
Artificial Neural Network

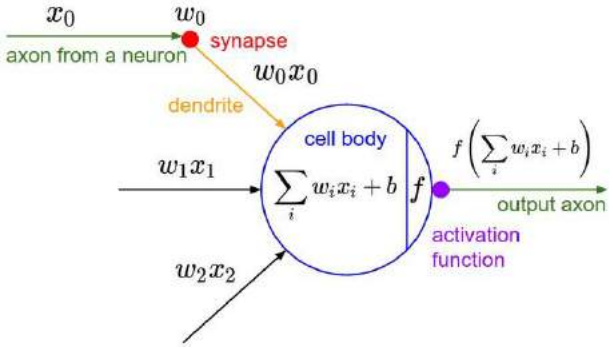
- **Human neural network:** 100 billion neurons, 1,000 trillion synapses
- **ResNet-152 neural network:** 60 million synapses



Neuron: Biological Inspiration for Computation

- **Neuron:** computational building block for the brain





- **(Artificial) Neuron:** computational building block for the "neural network"

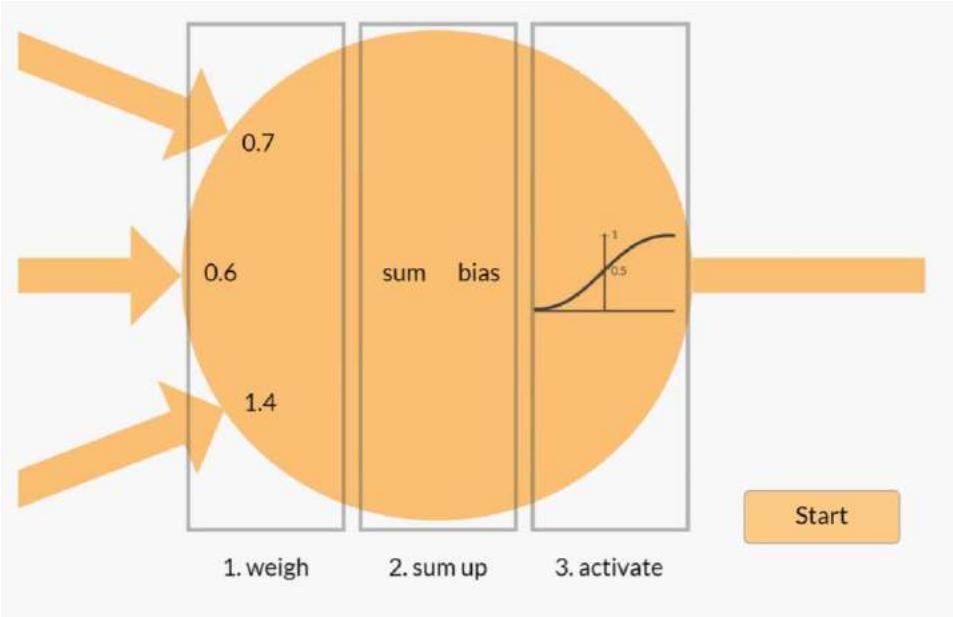
Differences (among others):

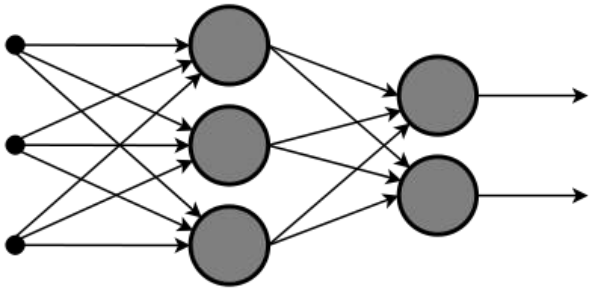
- **Parameters:** Human brains have ~10,000,000 times synapses than artificial neural networks.
- **Topology:** Human brains have no "layers". Topology is complicated.
- **Async:** The human brain works asynchronously, ANNs work synchronously.
- **Learning algorithm**: ANNs use gradient descent for learning. Human brains use … (we don't know)
- **Processing speed**: Single biological neurons are slow, while standard neurons in ANNs are fast.
- **Power consumption:** Biological neural networks use very little power compared to artificial networks
- **Stages:** Biological networks usually don't stop / start learning. ANNs have different fitting (train) and prediction (evaluate) phases.

Similarity (among others):

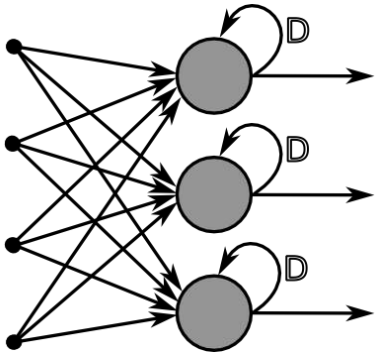
- Distributed computation on a large scale.

Artificial Neurons



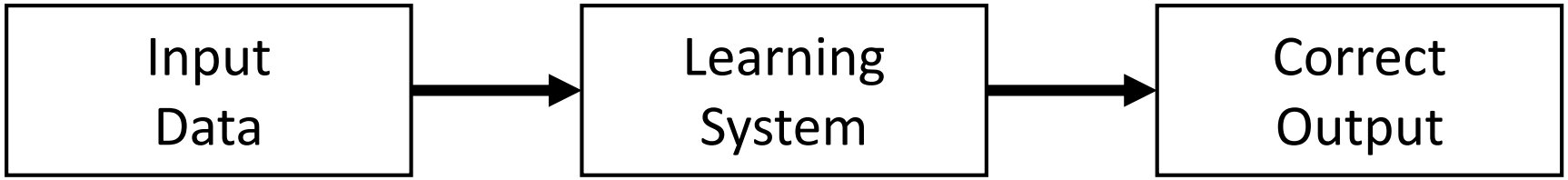


Feed Forward Neural Network



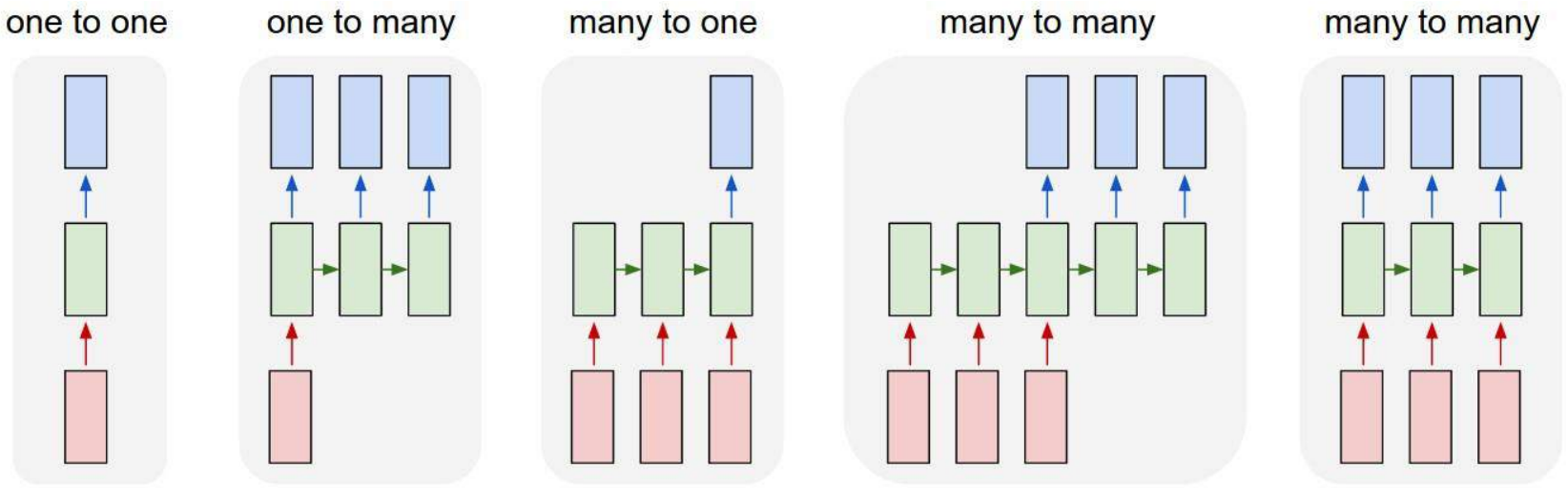
Recurrent Neural Network

What can we do with Deep Learning?

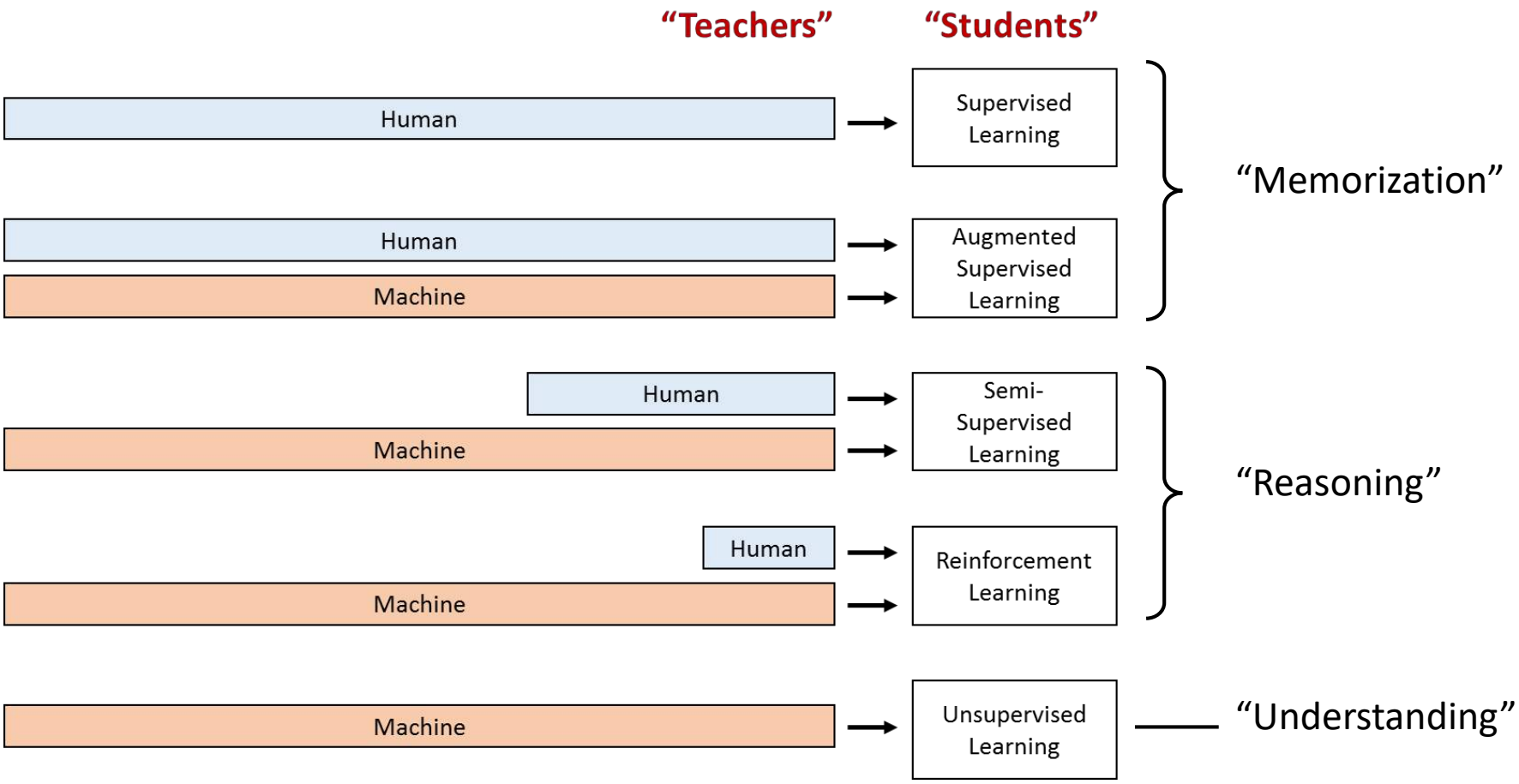


- Number
- Vector of numbers
- Sequence of numbers
- Sequence of vectors of numbers

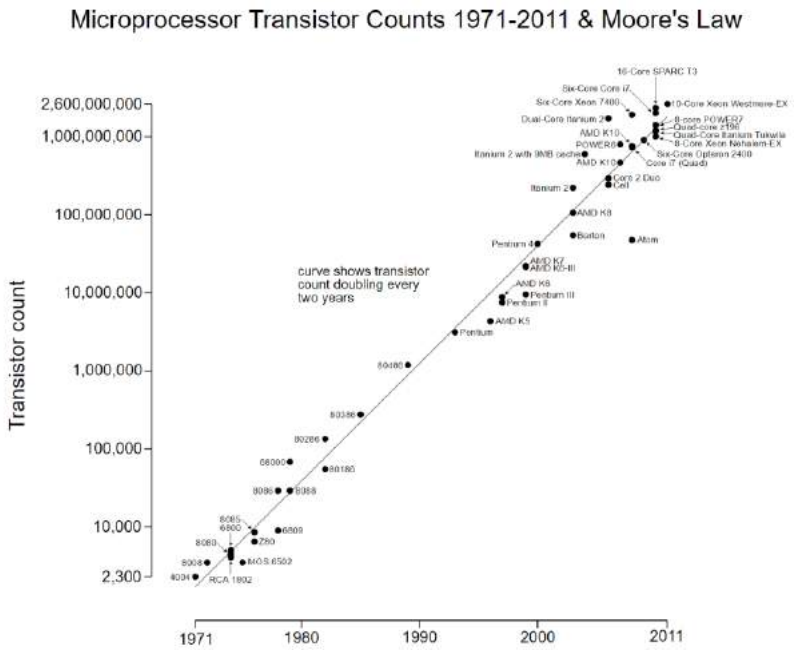
- Number
- Vector of numbers
- Sequence of numbers
- Sequence of vectors of numbers

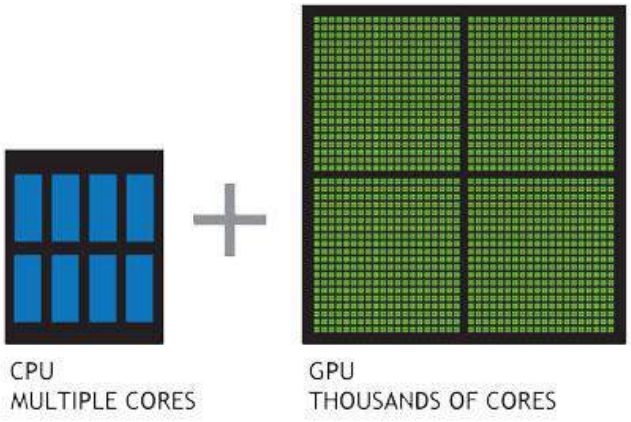


Deep Learning from Human and Machine



Past and Future of Deep Learning Breakthroughs





- **Compute**
  CPUs, GPUs, ASICs
- **Organized large(-ish) datasets**
  Imagenet
- **Algorithms and research:**
  Backprop, CNN, LSTM
- **Software and Infrastructure**
  Git, ROS, PR2, AWS, Amazon Mechanical Turk, TensorFlow, …
- **Financial backing of large companies**
  Google, Facebook, Amazon, …

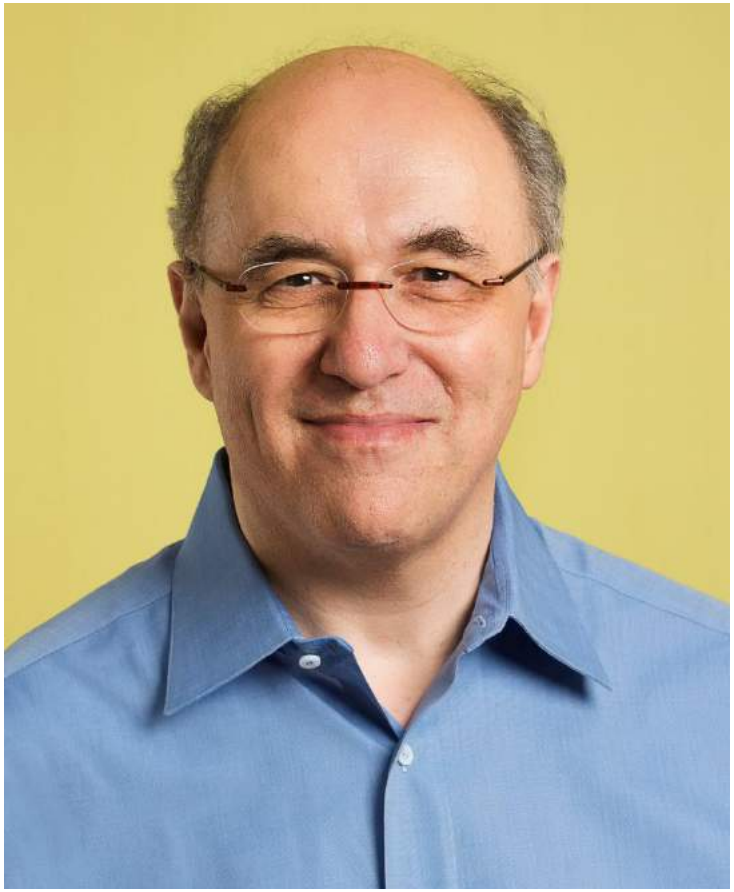
Current Challenges

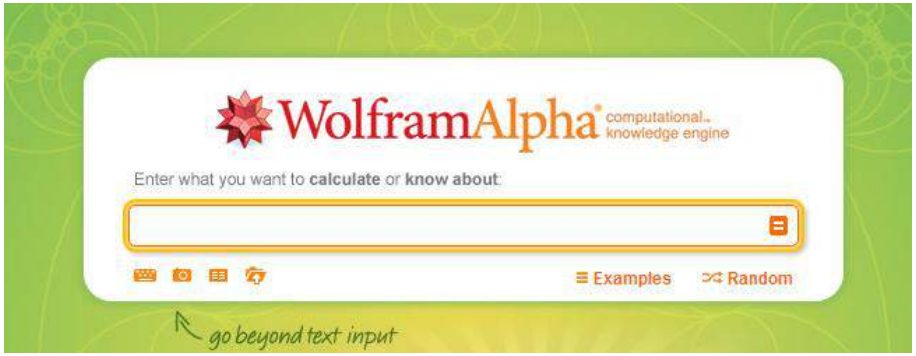
- **Transfer learning:** Unable to transfer representation to most reasonably related domains except in specialized formulations.
  - **Understanding:** Lacks "reasoning" or ability to truly derive "understanding" as previously defined on anything but specialized problem formulations. (Definition used: Ability to turn complex information to into simple, useful information.)
- Requires **big** data: inefficient at learning from data
- Requires **supervised** data: costly to annotate real-world data
- **Not fully automated:** Needs hyperparameter tuning for training: learning rate, loss function, mini-batch size, training iterations, momentum, optimizer selection, etc.
- **Reward:** Defining a good reward function is difficult.
- **Transparency:** Neural networks are for the most part black boxes (for real-world applications) even with tools that visualize various aspects of their operation.
- **Edge cases:** Deep learning is not good at dealing with edge cases.

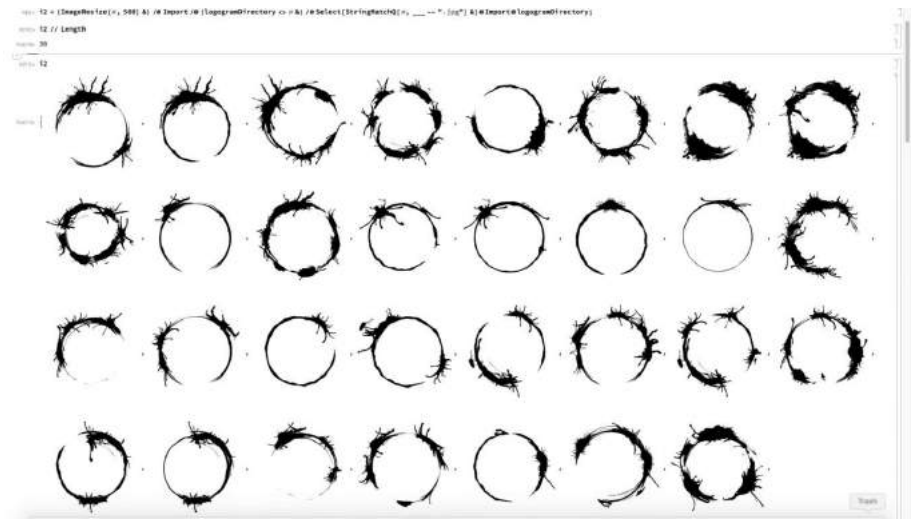
Stephen Wolfram

Knowledge-Based Programming

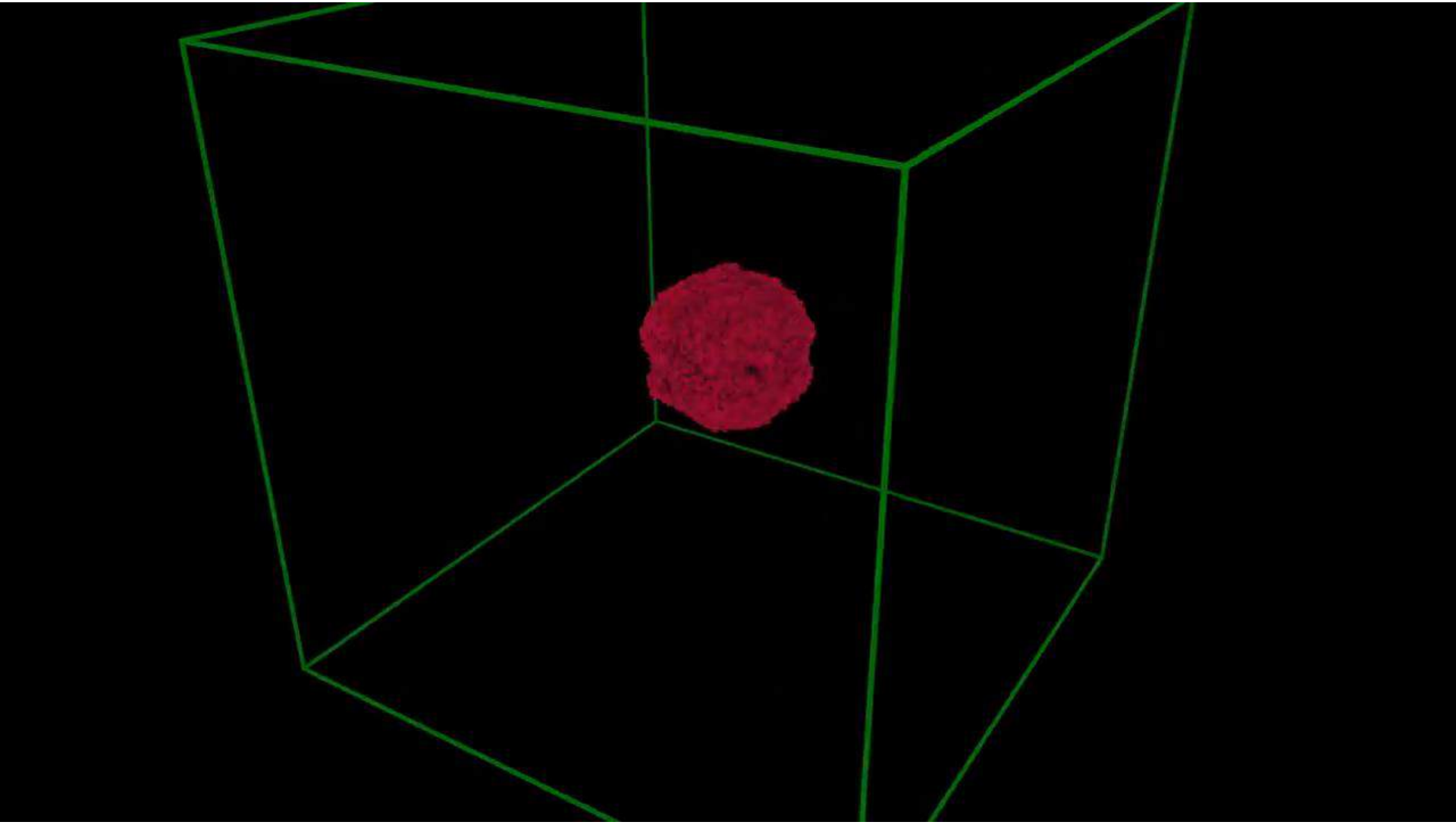
Mon, Jan 29, 7pm (Room 54-100)







“Artificial Life Simulation”: Cellular Automata and Emerging Complexity



Richard Moyes, Article36

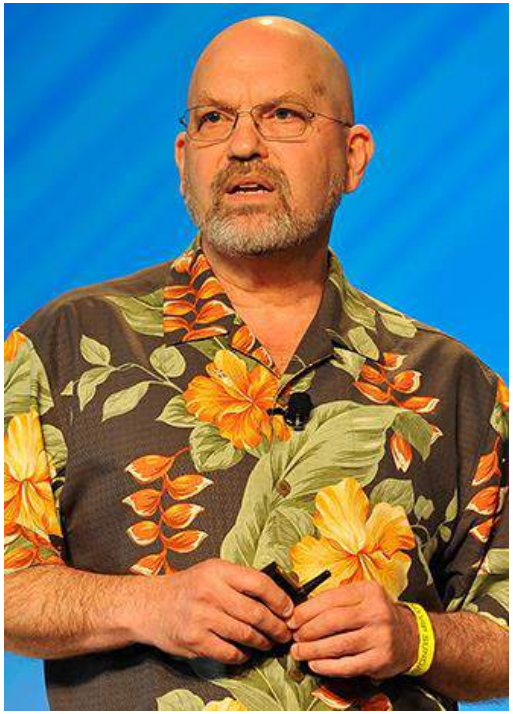
AI Safety and Autonomous Weapon Systems

Tue, Jan 30, 7pm (Room 54-100)

Marc Raibert, CEO, Boston Dynamics

Robots in the Real World

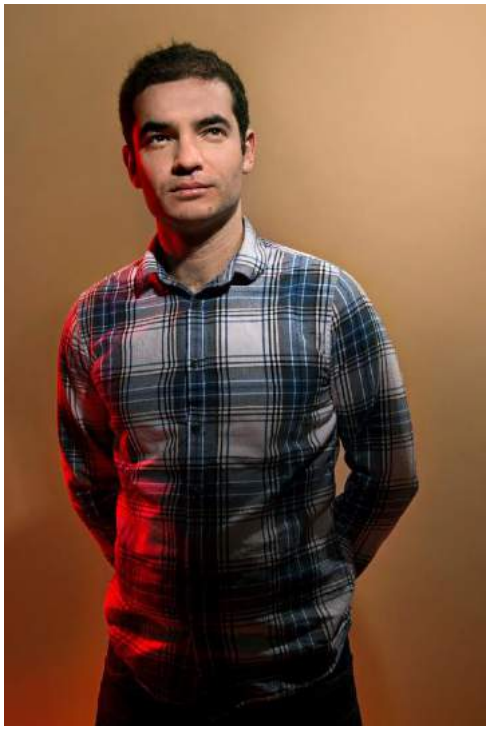
Wed, Jan 31, 7pm (Room 54-100)

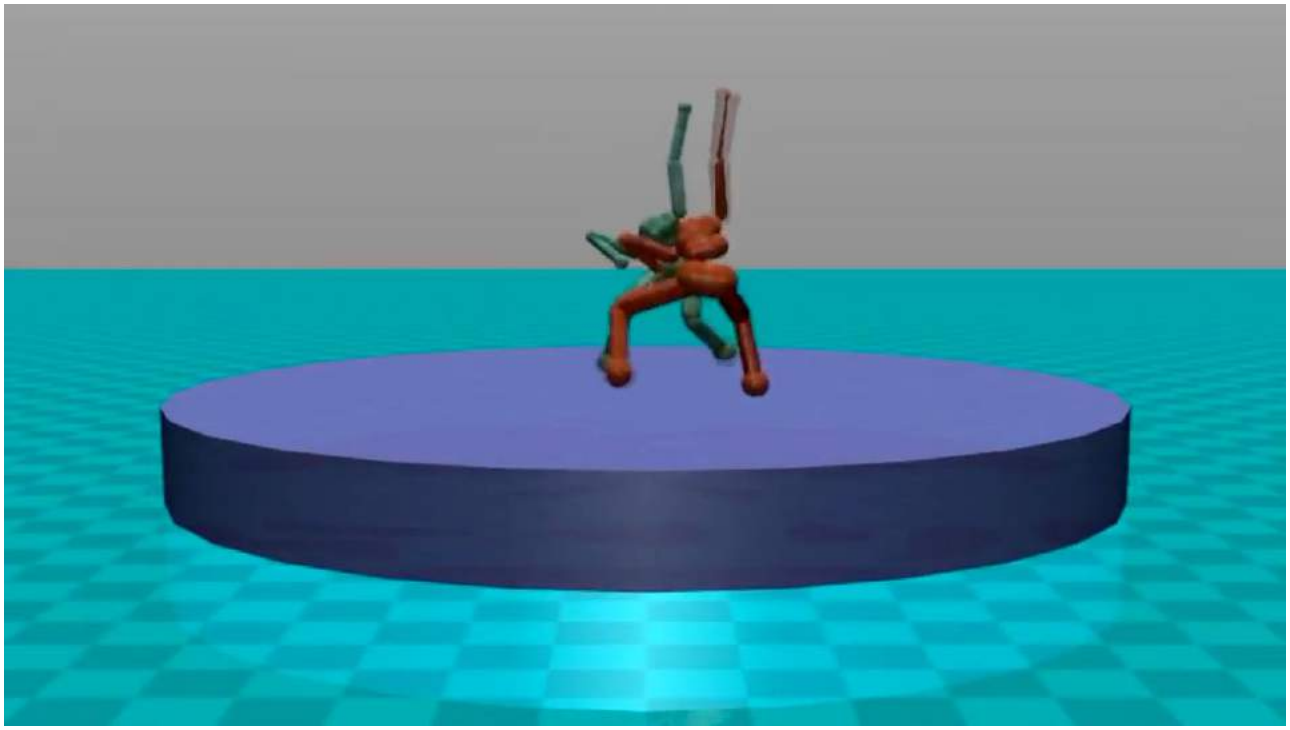


Ilya Sutskever, Co-Founder, OpenAI

Deep Reinforcement Learning

Thu, Feb 1, 7pm (Room 54-100)

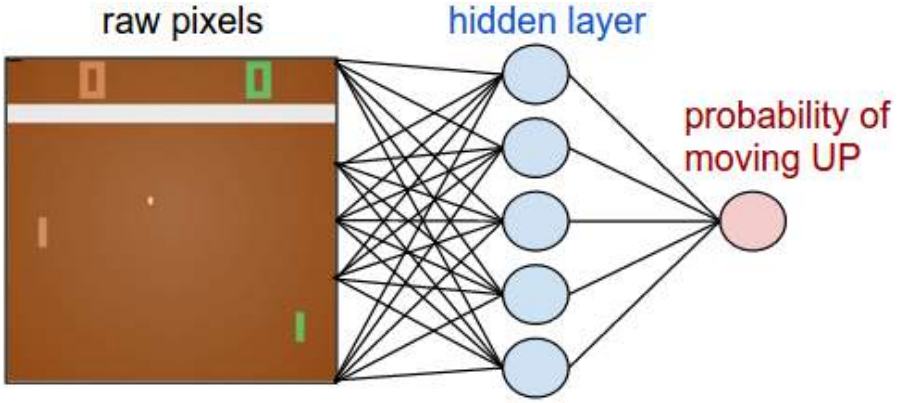




(Toward) General Purpose Intelligence:

Pong to Pixels

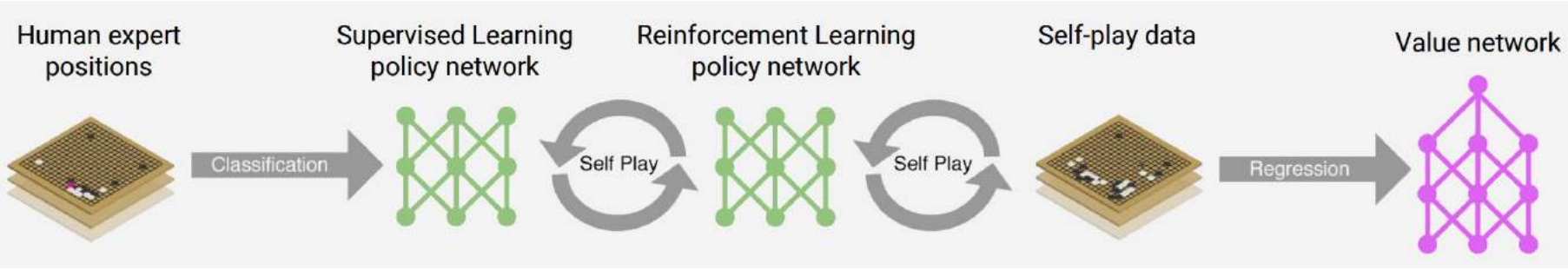
Policy Network:

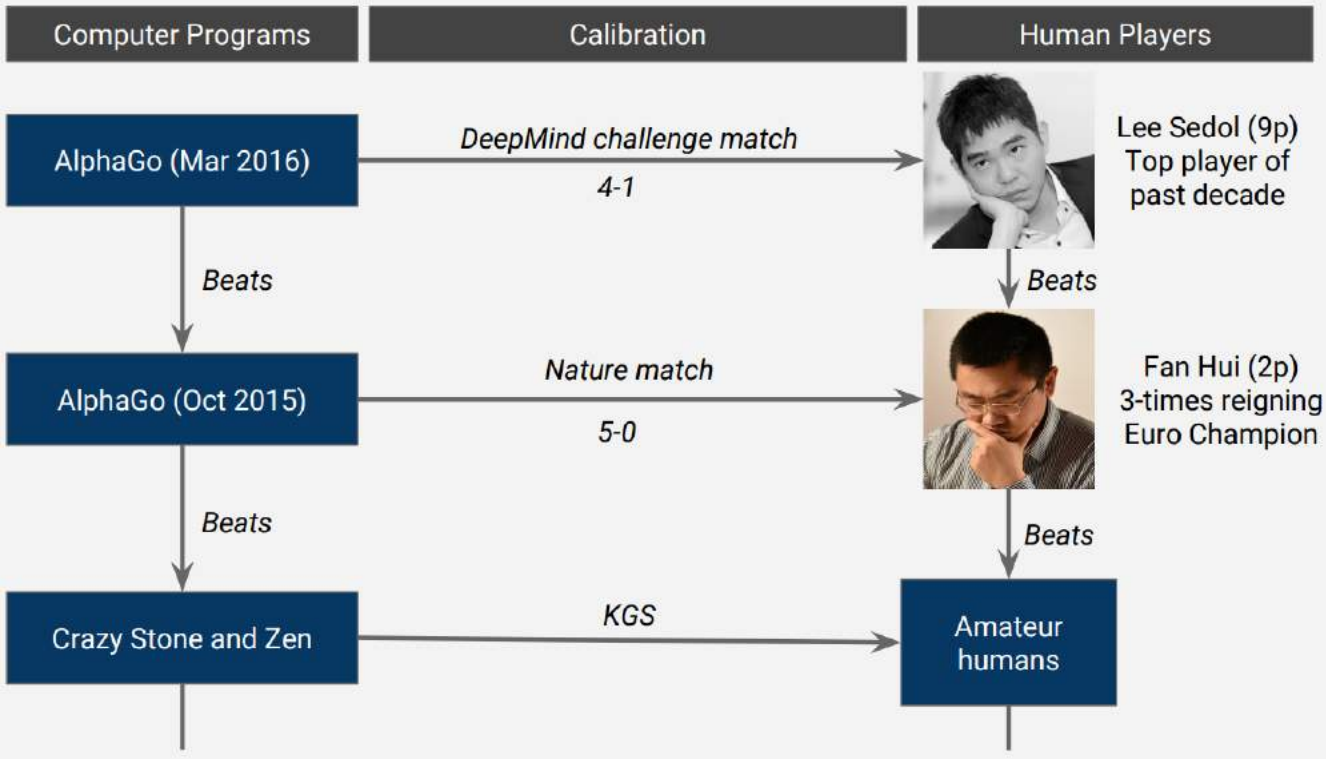


- 80x80 image (difference image)
- 2 actions: up or down
- 200,000 Pong games

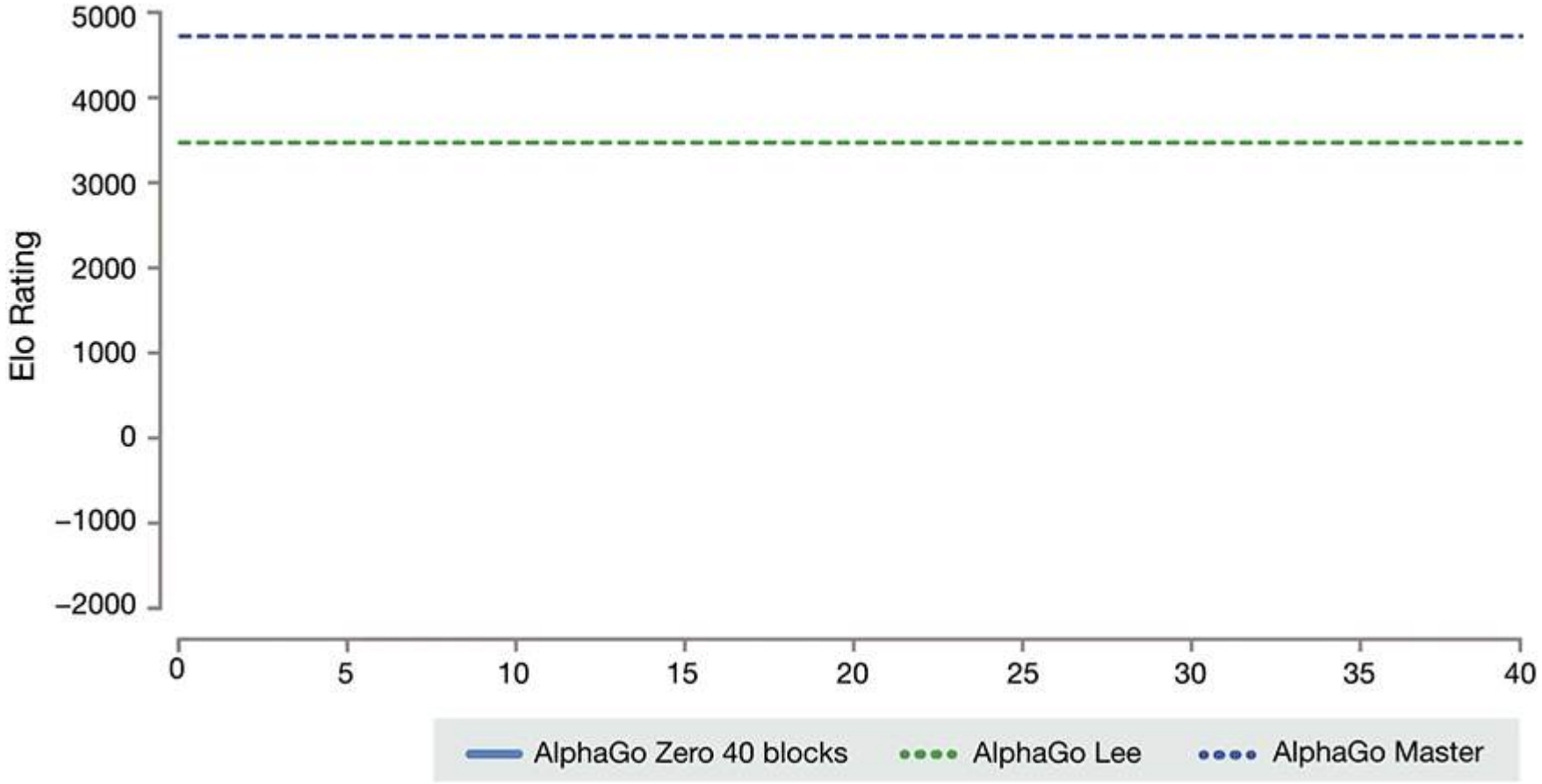
This is a step towards general purpose artificial intelligence!

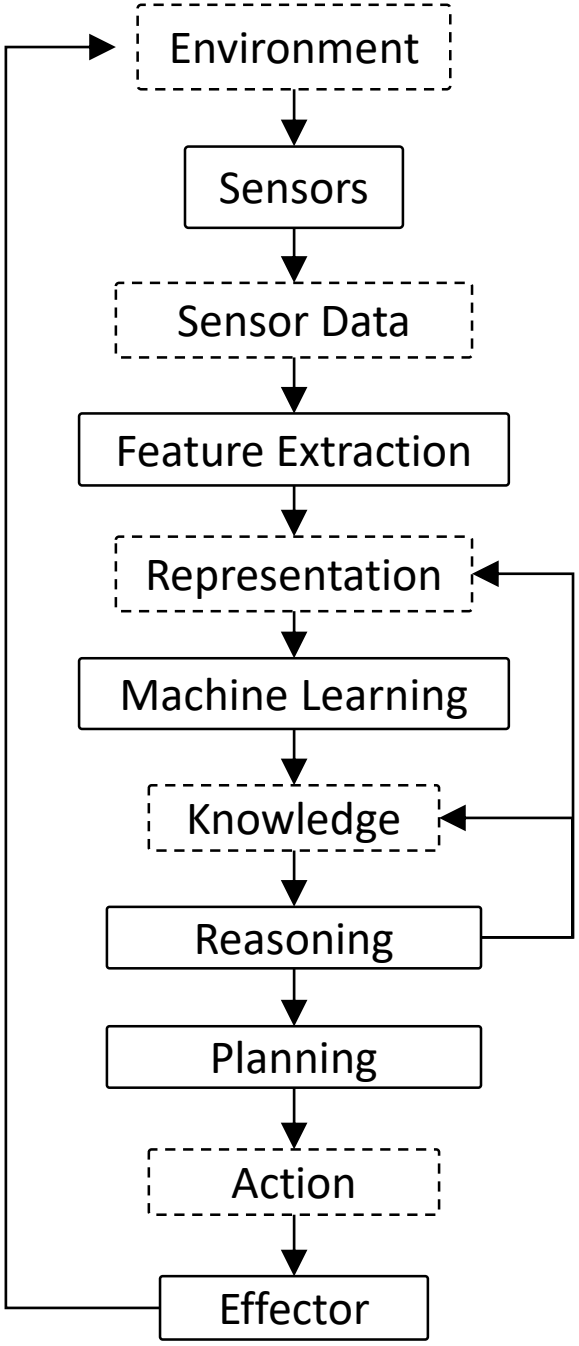
AlphaGo (2016) Beat Top Human at **Go**





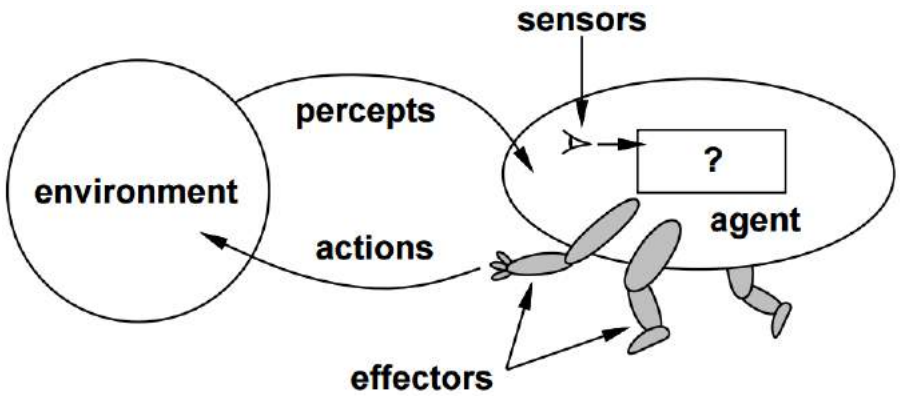
AlphaGo Zero (2017): Beats AlphaGo

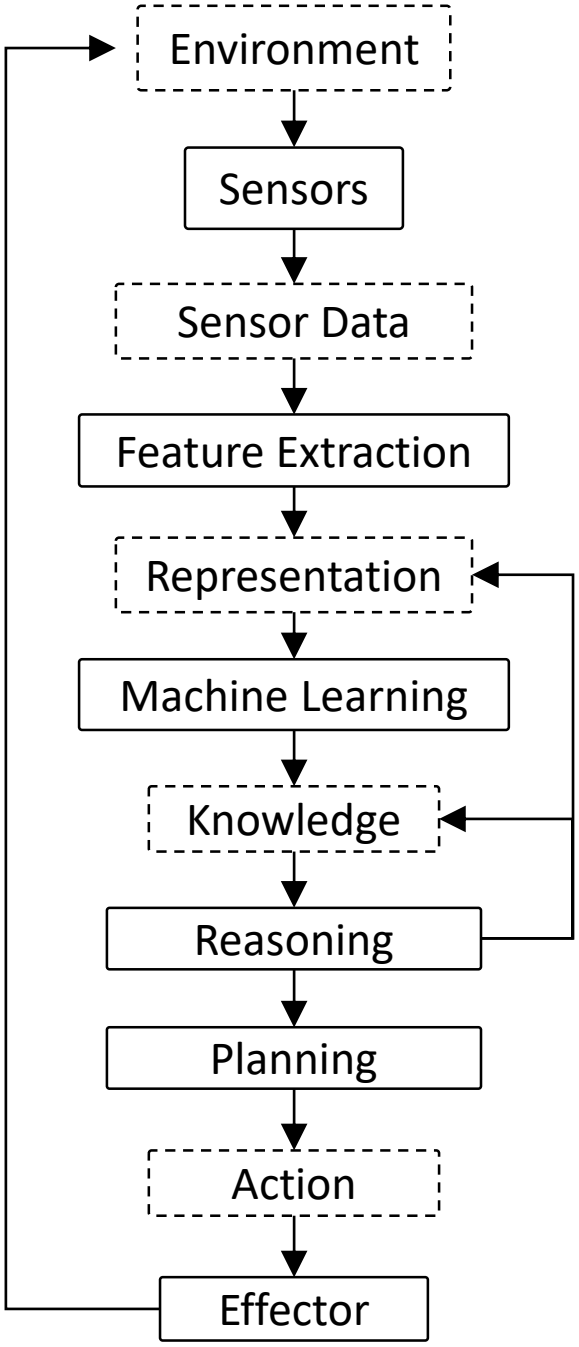




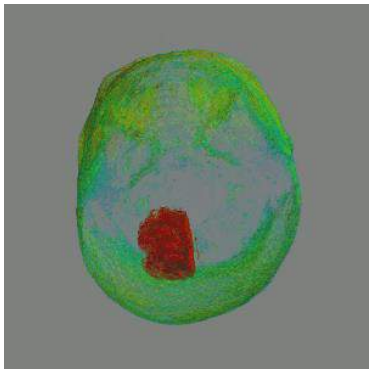
Open Question:

What can we **not** do with Deep Learning?





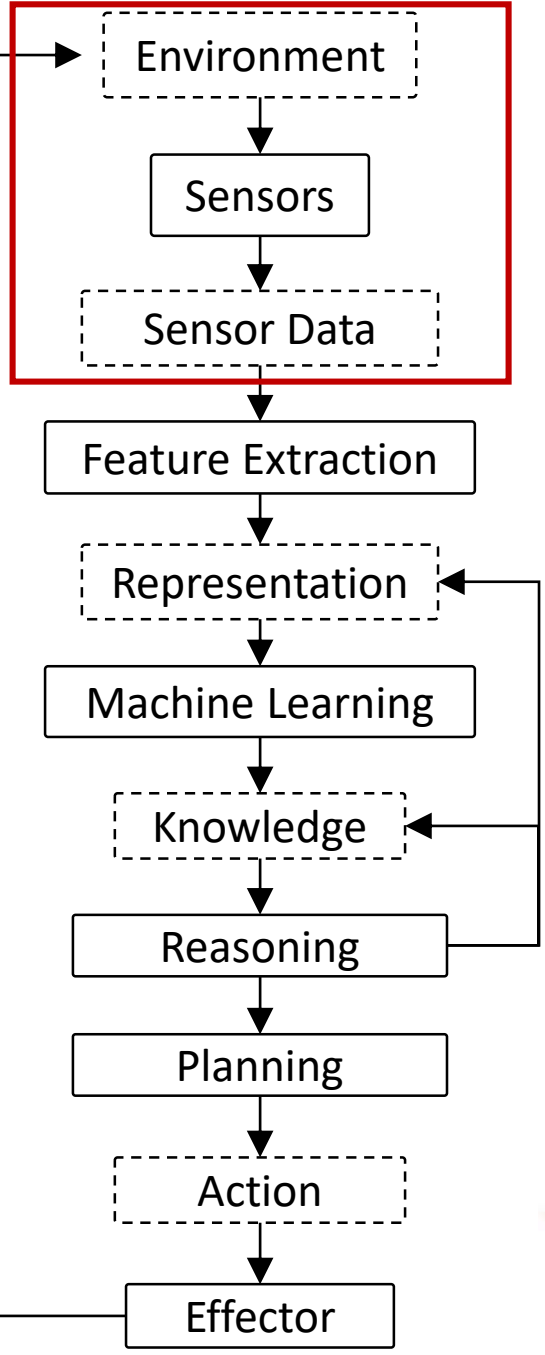
Formal tasks: Playing board games, card games. Solving puzzles, mathematical and logic problems.

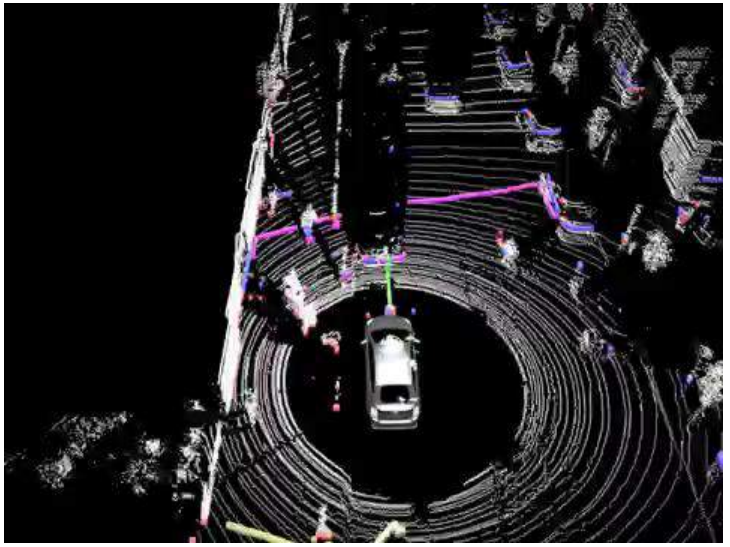


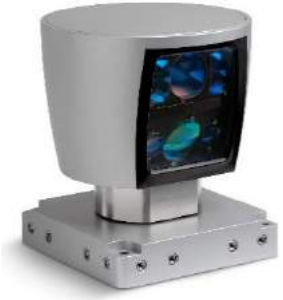
Expert tasks: Medical diagnosis, engineering, scheduling, computer hardware design.

Mundane tasks: Everyday speech, written language, perception, walking, object manipulation.

Human tasks: Awareness of self, emotion, imagination, morality, subjective experience, high-level-reasoning, consciousness.







Lidar

Camera
(Visible, Infrared)

Radar

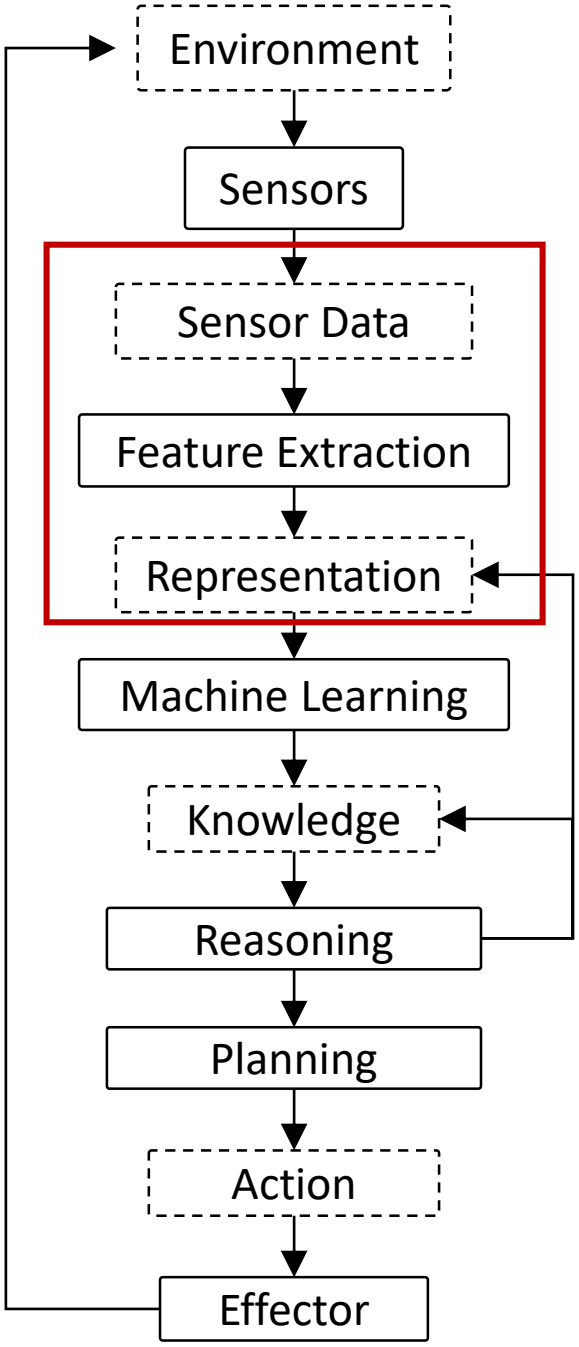
GPS

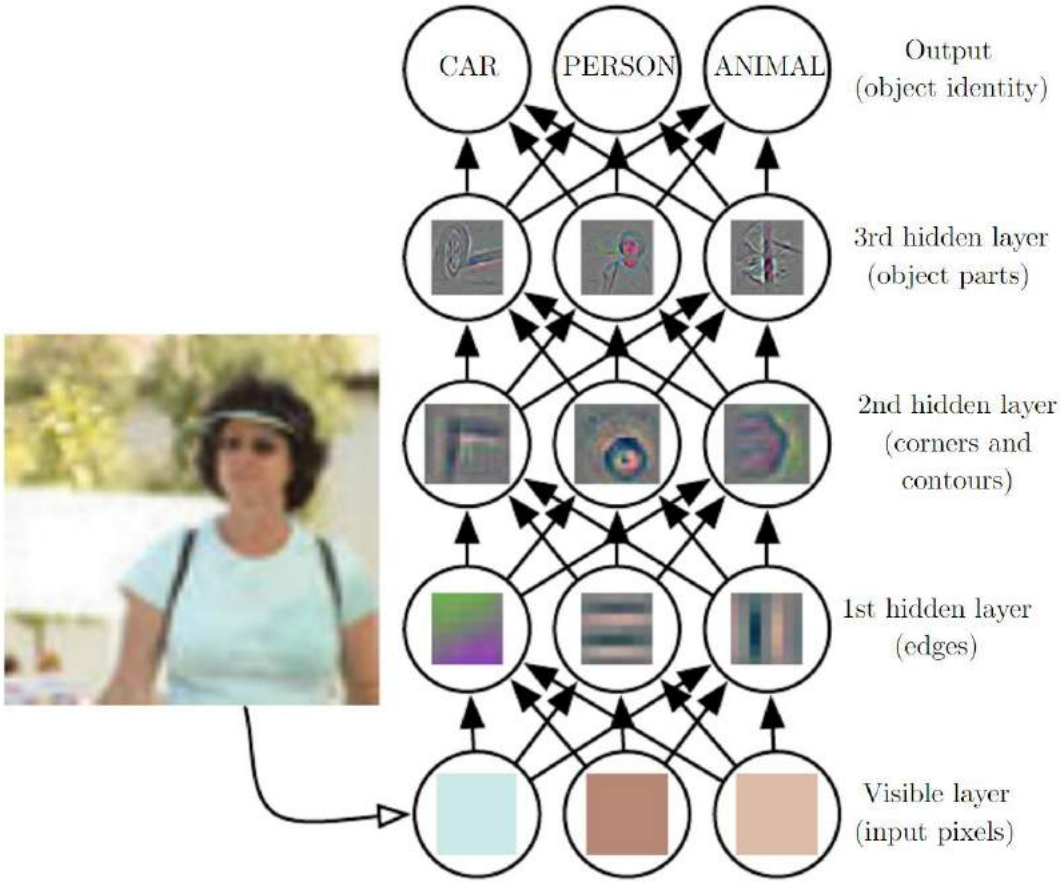
Stereo Camera

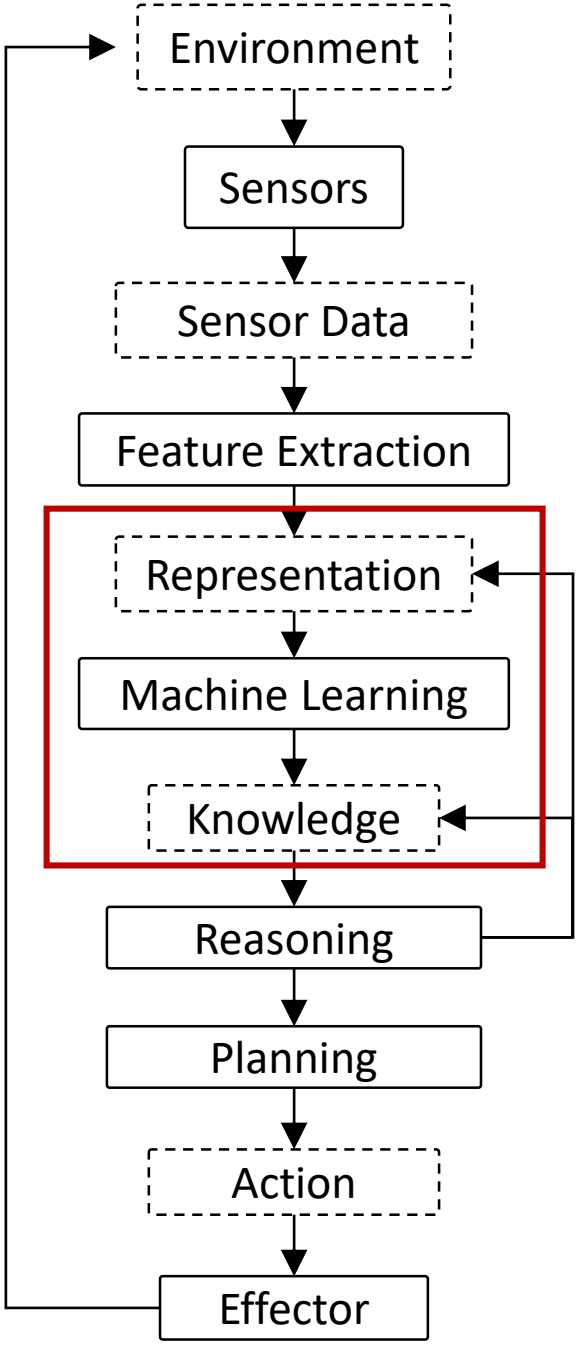
Microphone

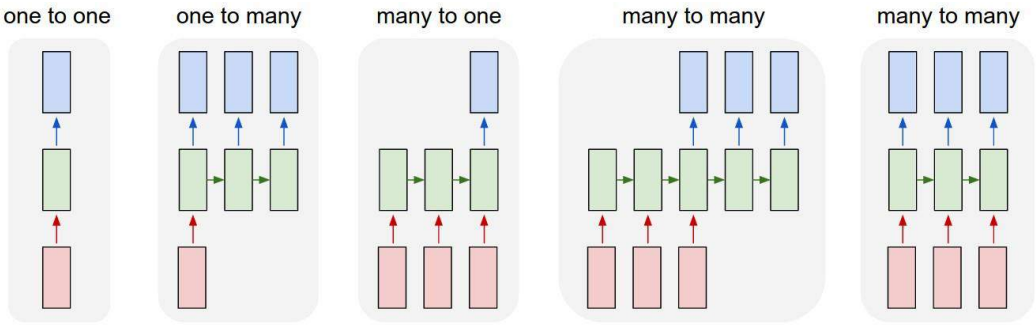
Networking
(Wired, Wireless)

IMU


Environment
Sensors
Sensor Data
Feature Extraction
Representation
Machine Learning
Knowledge
Reasoning
Planning
Action
Effector


CAR
PERSON
ANIMAL
Output
(object identity)
3rd hidden layer
(object parts)
2nd hidden layer
(corners and
contours)
1st hidden layer
(edges)
Visible layer
(input pixels)


Environment
Sensors
Sensor Data
Feature Extraction
Representation
Machine Learning
Knowledge
Reasoning
Planning
Action
Effector


one to one
one to many
many to one
many to many
many to many

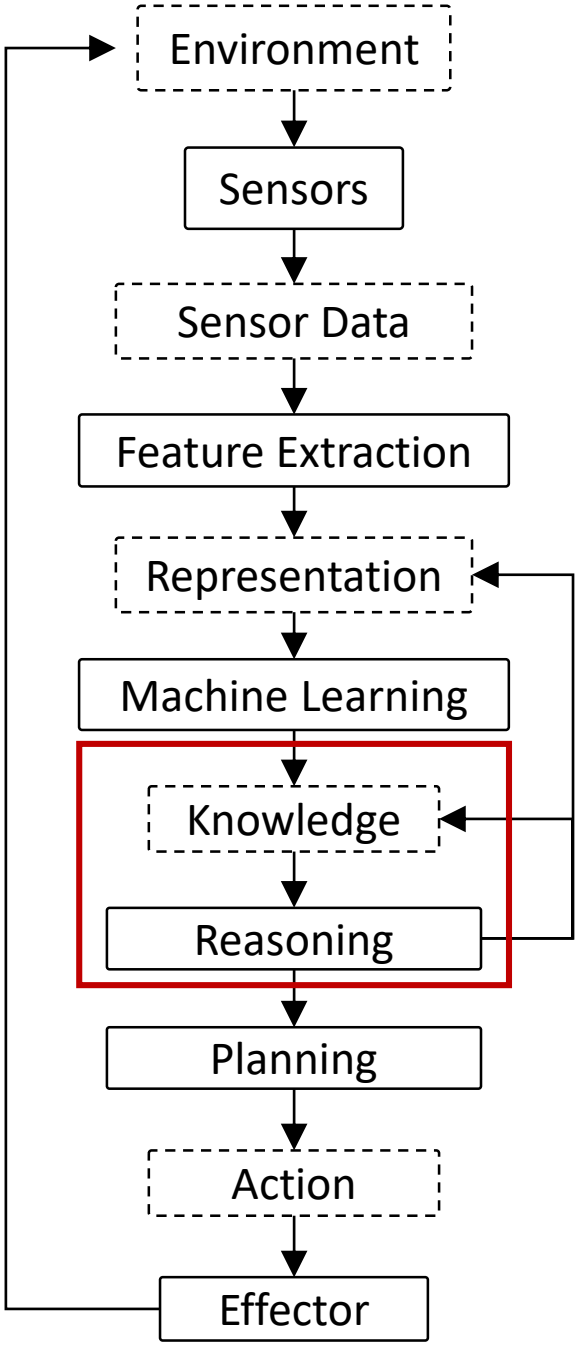
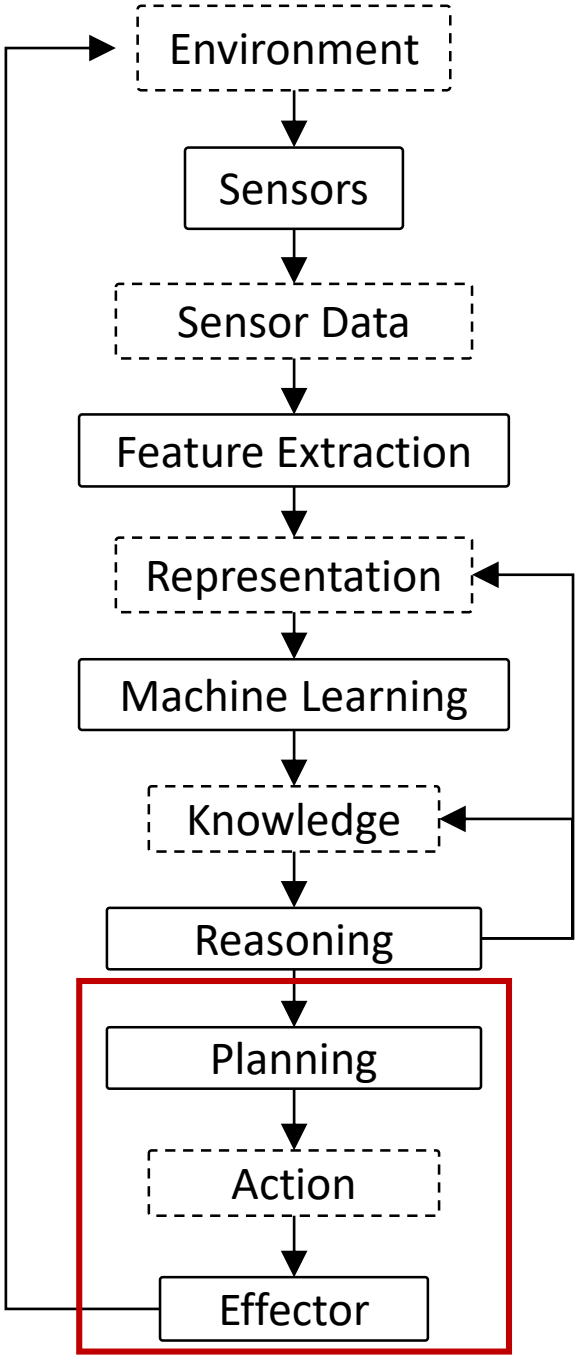
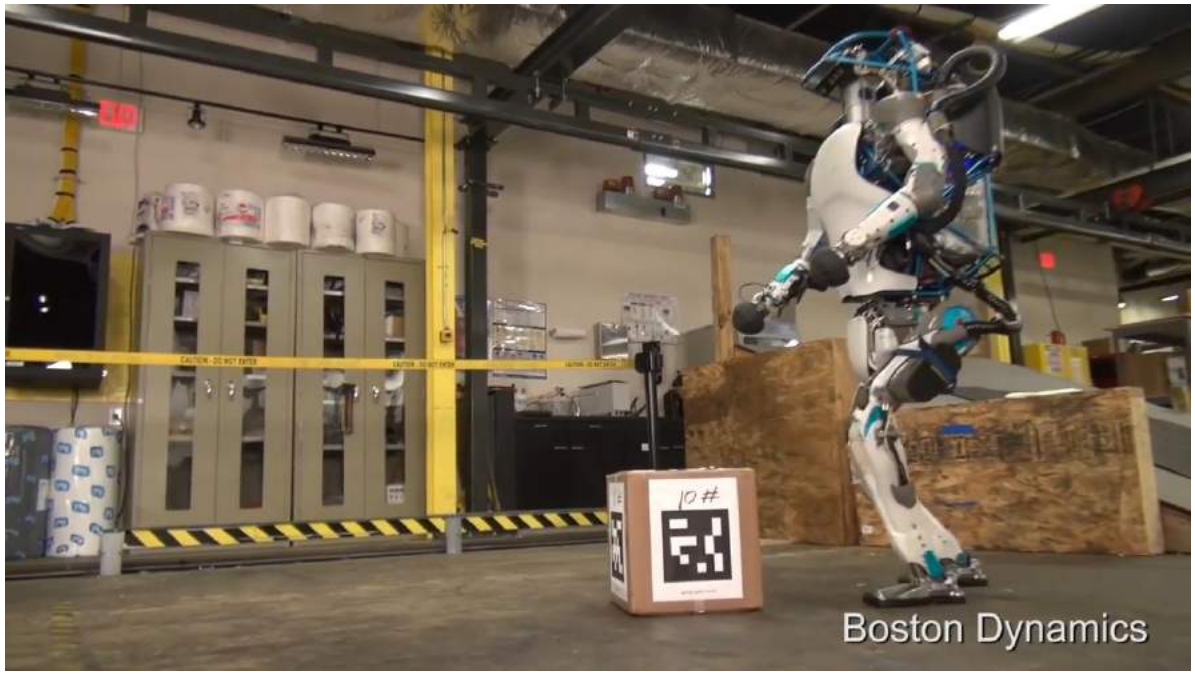


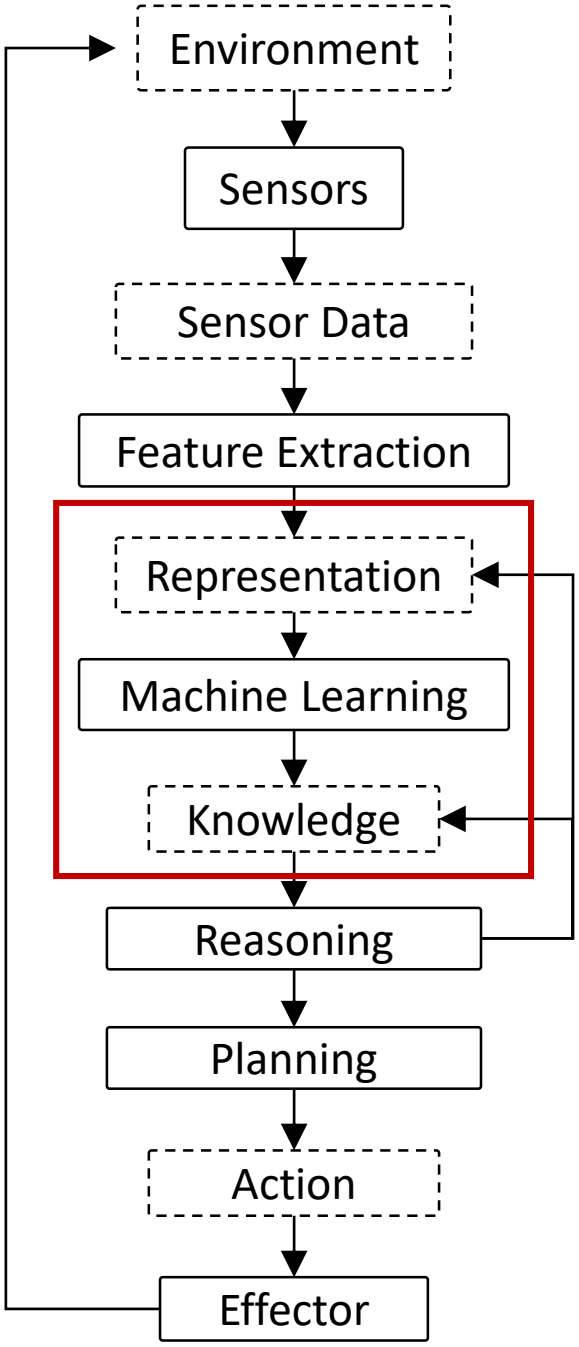
Image Recognition:
If it looks like a duck

Audio Recognition:
Quacks like a duck

Activity Recognition:
Swims like a duck

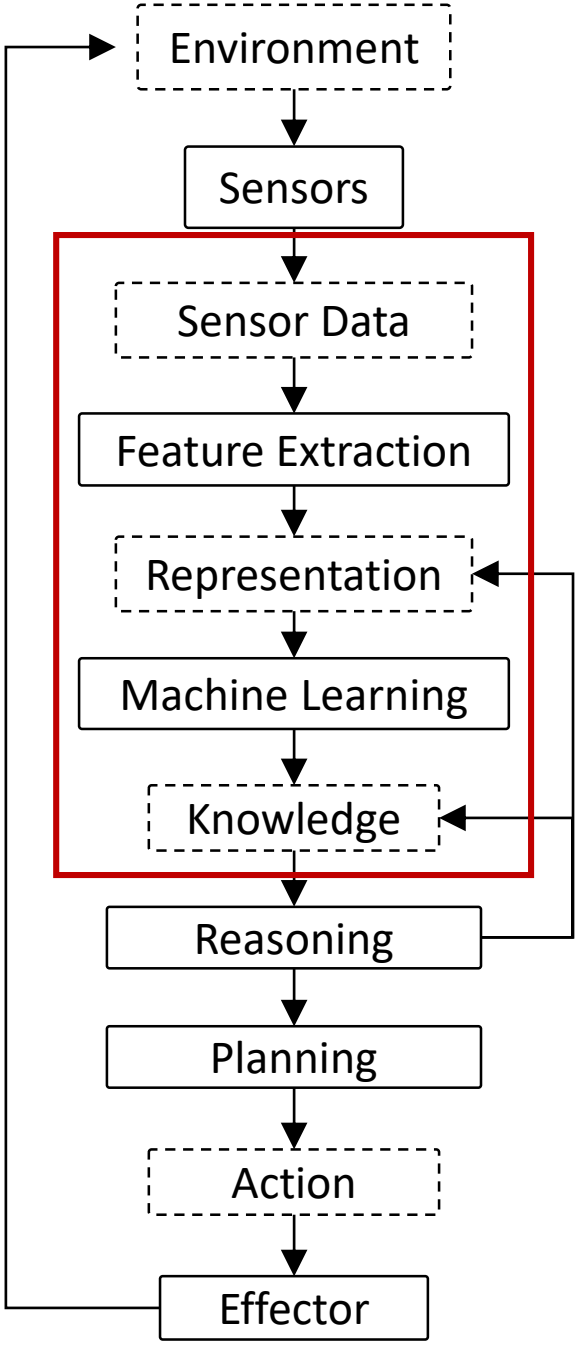

Environment
Sensors
Sensor Data
Feature Extraction
Representation
Machine Learning
Knowledge
Reasoning
Planning
Action
Effector


Boston Dynamics



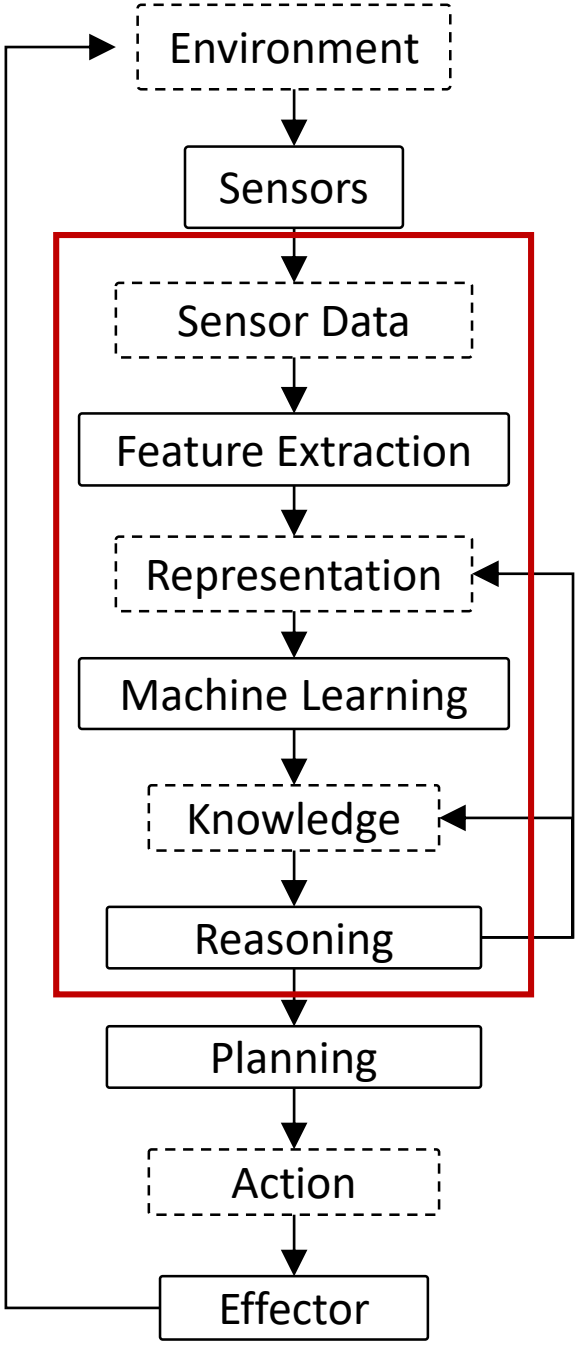
Open Question:

How much of this AI stack can be **learned**?



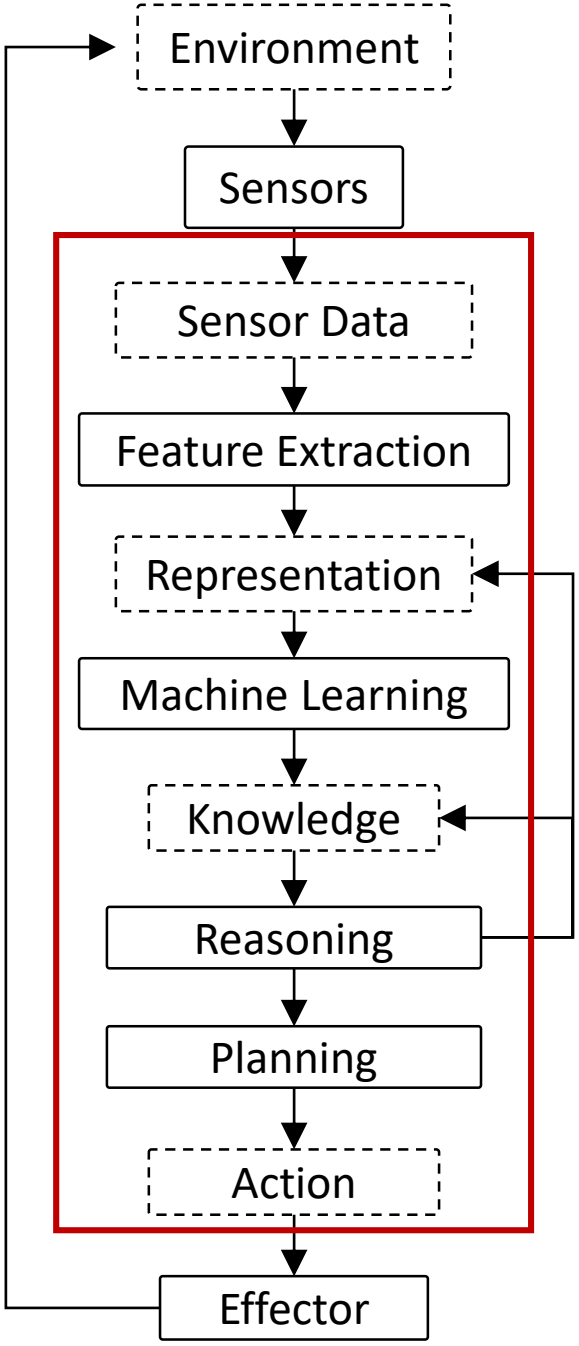
Open Question:

How much of this AI stack can be **learned**?



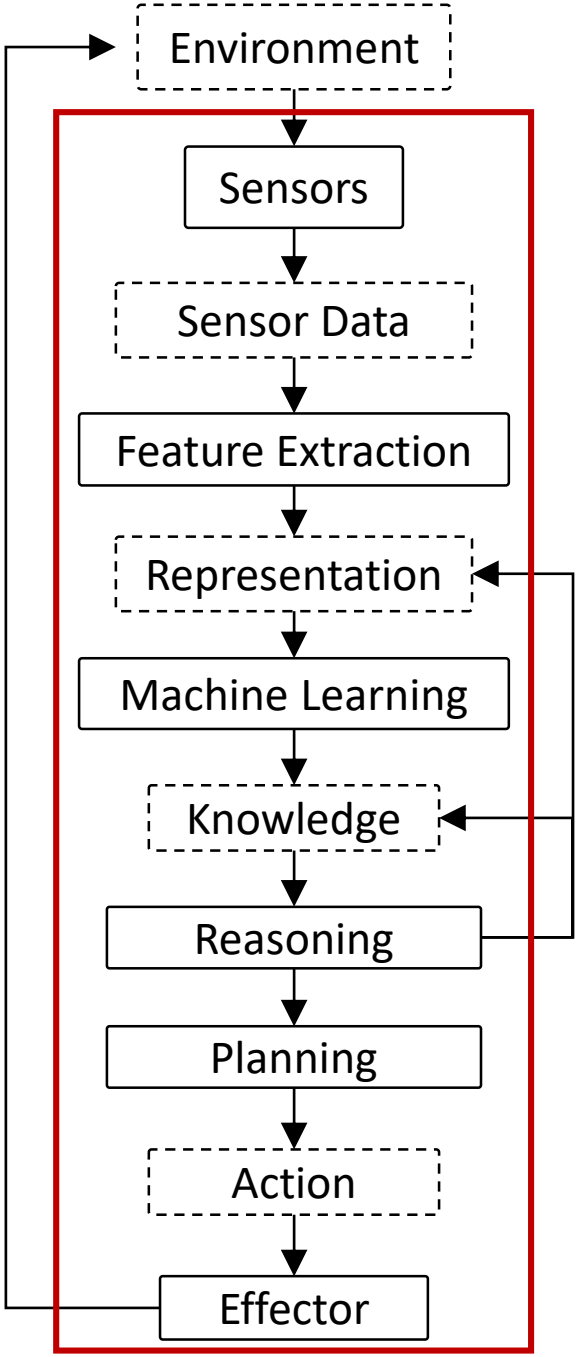
Open Question:

How much of this AI stack can be **learned**?



Open Question:

How much of this AI stack can be **learned**?



Open Question:

How much of this AI stack can be **learned**?

Lectures and Guest Talks

Lecture Mon, Jan 22, 7pm Room 54-100
Artificial General Intelligence
[Slides] - [Lecture Video] *(Available Soon)*

Guest Talk Tue, Jan 23, 7pm Room 54-100
Josh Tenenbaum: Computational Cognitive Science
Professor, MIT

Guest Talk Wed, Jan 24, 1pm Room 10-250
Ray Kurzweil: How to Create a Mind
Google

Guest Talk Thu, Jan 25, 7pm Room 54-100
Lisa Feldman Barrett: Emotion Creation
Northeastern University

Guest Talk Fri, Jan 26, 7pm Room 54-100
Nate Derbinsky: Cognitive Modeling
Northeastern University

Guest Talk Mon, Jan 29, 1:30pm Room 26-100
Andrej Karpathy: Deep Learning
Director of AI, Tesla
Previously: OpenAI, Stanford University

Guest Talk Mon, Jan 29, 7pm Room 54-100
Stephen Wolfram: Knowledge-Based Programming
Wolfram Research

Guest Talk Tue, Jan 30, 7pm Room 54-100
Richard Moyes: AI Safety and Autonomous Weapon Systems
Co-Founder and Managing Director, Article36

Guest Talk Wed, Jan 31, 7pm Room 54-100
Marc Raibert: Robotics
CEO, Boston Dynamics
Previously: MIT

Guest Talk Thu, Feb 1, 7pm Room 54-100
Ilya Sutskever: Deep Reinforcement Learning
Co-founder, OpenAI
Previously: Google Brain, Stanford, U of Toronto

Lecture Fri, Feb 2, 7pm Room 54-100
Human-Centered Artificial Intelligence
[Slides] - [Lecture Video] *(Available Soon)*

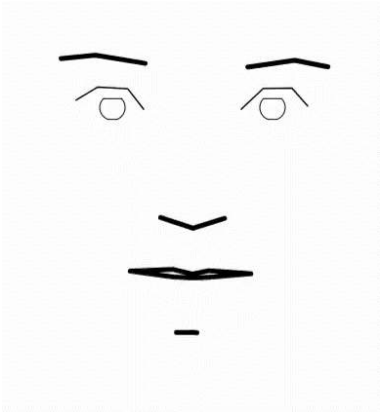
Timeline: AGI Approaches

- MIT 6.S099: Artificial General Intelligence (first 2 weeks)
  - Deep learning
  - Deep reinforcement Learning
  - Cognitive modeling
  - Computational cognitive science
  - Emotion creation
  - Knowledge based programming
  - AI Safety
  - Human-centered artificial intelligence
- MIT 6.S099: Artificial General Intelligence (in 2018)
  - AI ethics and bias
  - Creativity in generating music and art
  - Brain simulation
  - Computational neuroscience
  - Turing test and natural language processing
  - *…and much more…*

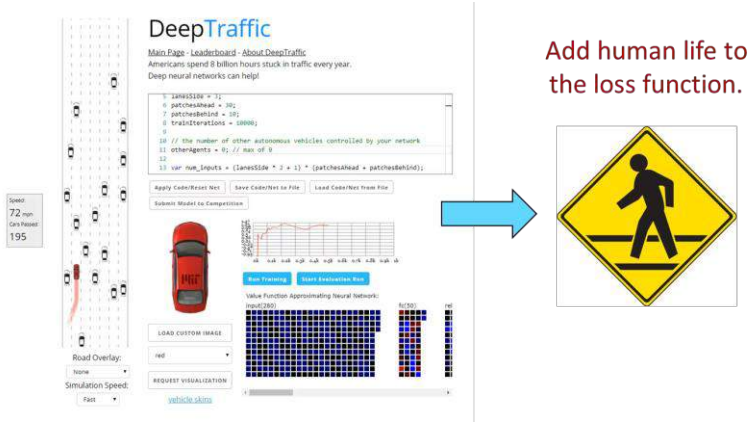
Thank You

DreamVision
https://agi.mit.edu/dreamvision

ANGEL
https://agi.mit.edu/angel



EthicalCar
https://agi.mit.edu/ethicalcar



VoteAI
https://agi.mit.edu/vote-ai