

BIG DATA & ARTIFICIAL INTELLIGENCE

Sandeep Giri

2014

CloudxLab

An environment for learning and researching on Big Data

2014

Amazon

Built High Throughput Systems for Amazon.com site like storm.

2012

InMobi

Built Recommender after churning **200 TB**

2011

tBits Global

Founded tBits Global
Built an enterprise grade Document Management System

2006

D.E.Shaw

Built the big data systems before the
term was coined

2002

IIT Roorkee

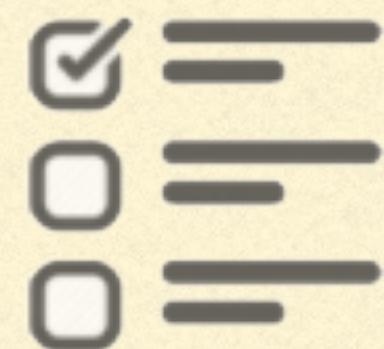
Finished B.Tech



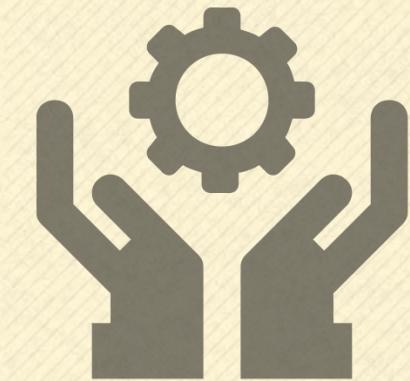
ABOUT



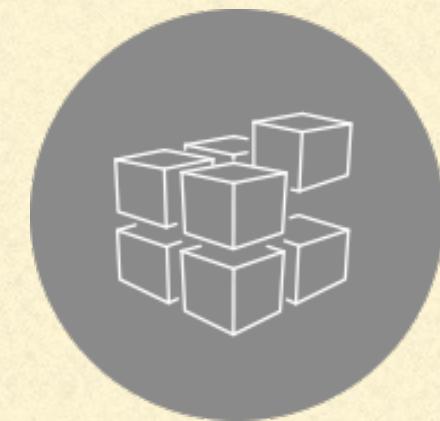
Videos



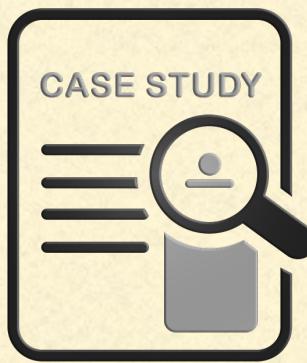
Quizzes



Hands-On



Projects



Case Studies

Gamified Learning in Big Data & AI



HAVE YOU PLAYED MARIO?



HAVE YOU PLAYED MARIO?

How much time did it take you to learn & win the princess?

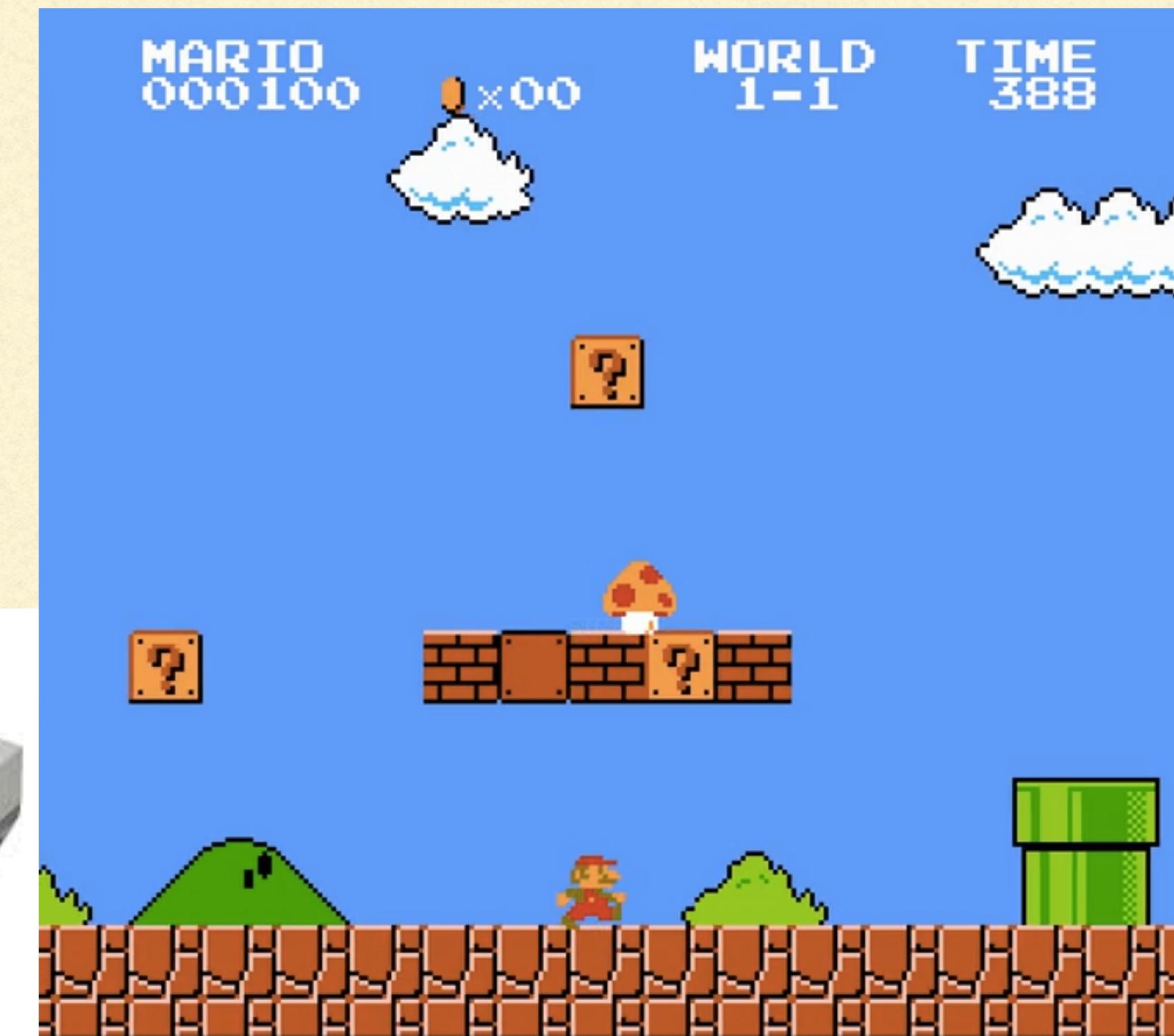


HAVE YOU PLAYED MARIO?

How much time did it take you to learn & win the princess?

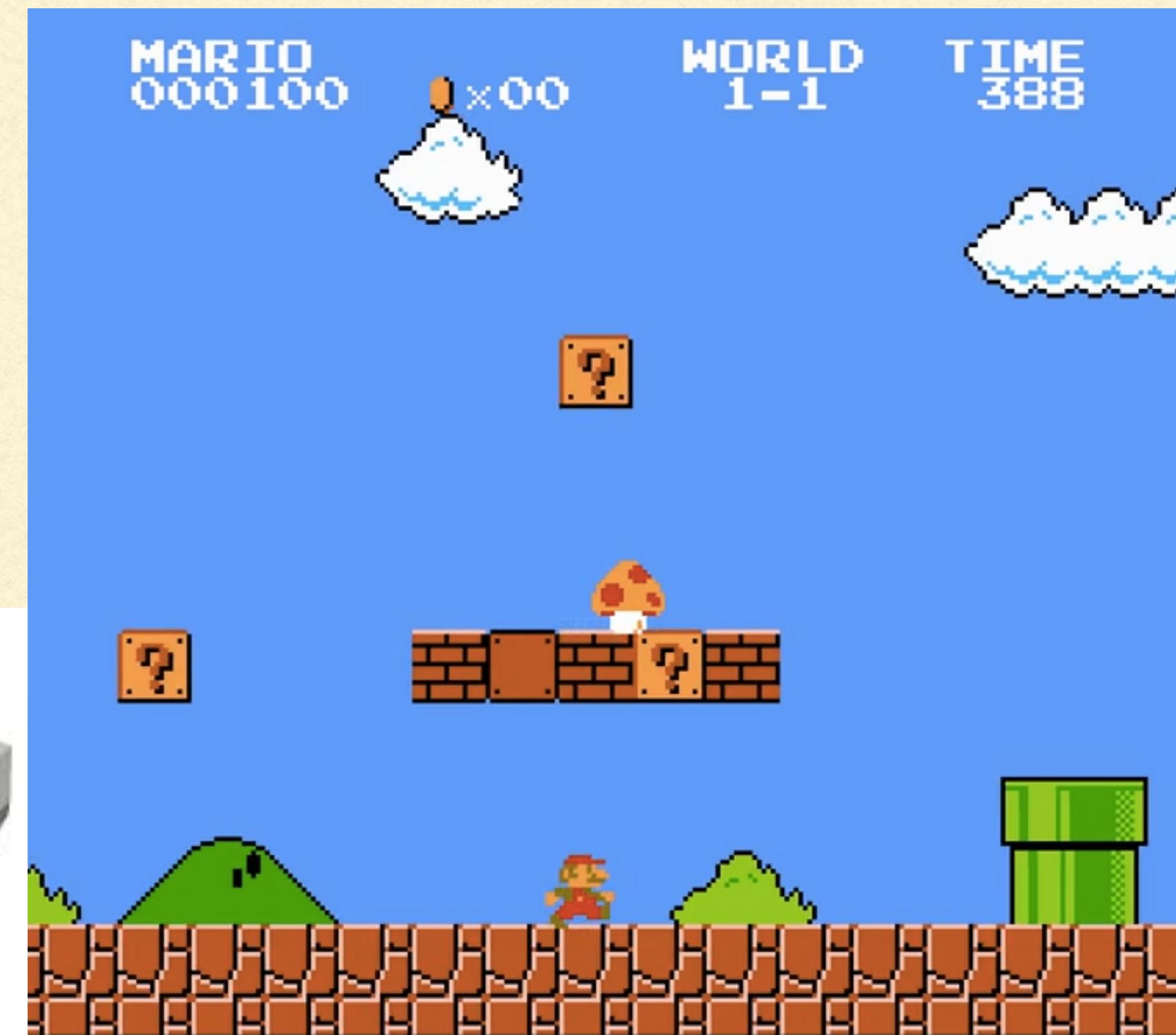
Did any one teach you?

How about automating it?



How about automating it?

- Program Learns to Play Mario



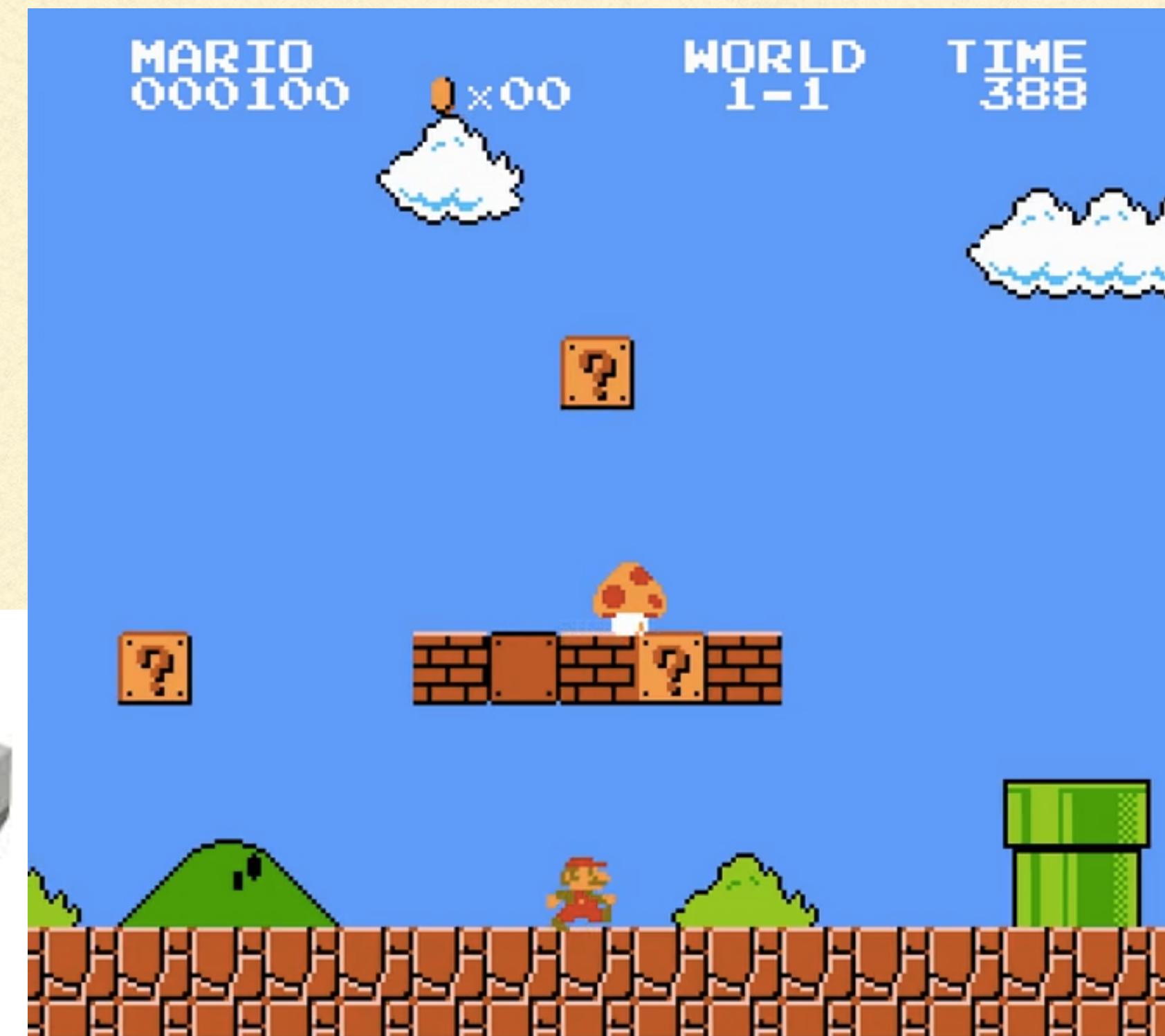
How about automating it?

- Program Learns to Play Mario
- Observes the game & presses keys



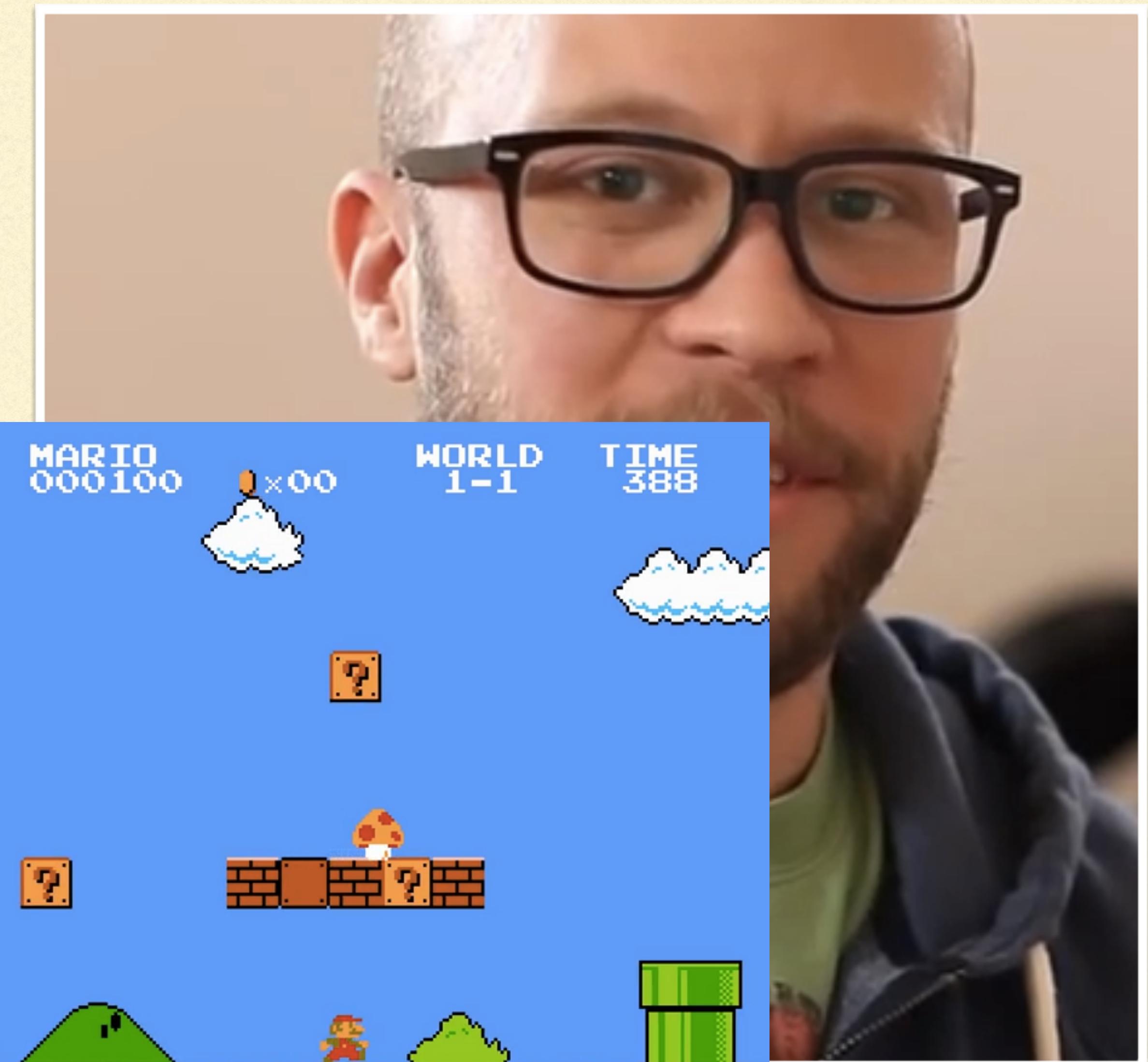
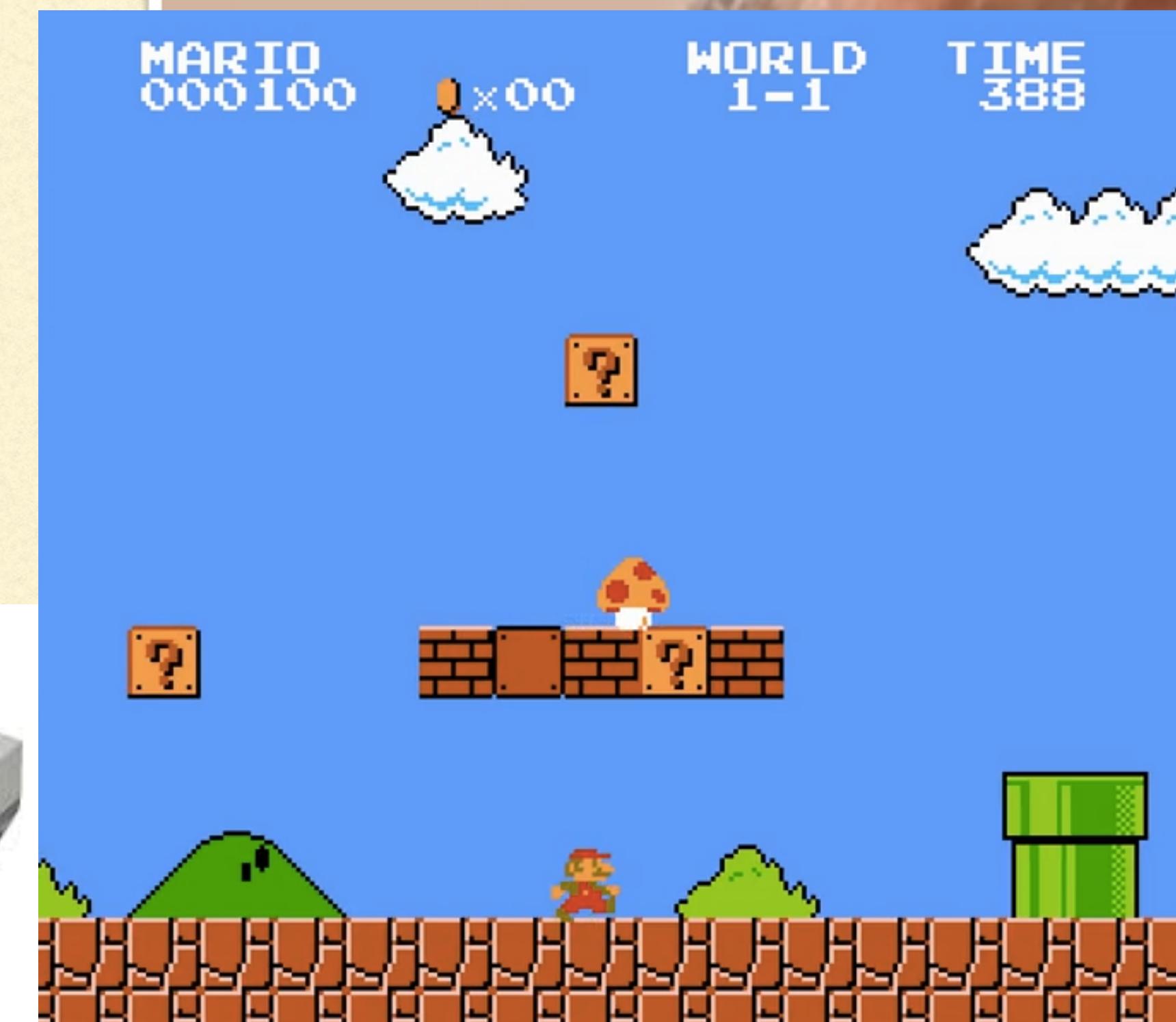
How about automating it?

- Program Learns to Play Mario
- Observes the game & presses keys
- Maximises Score



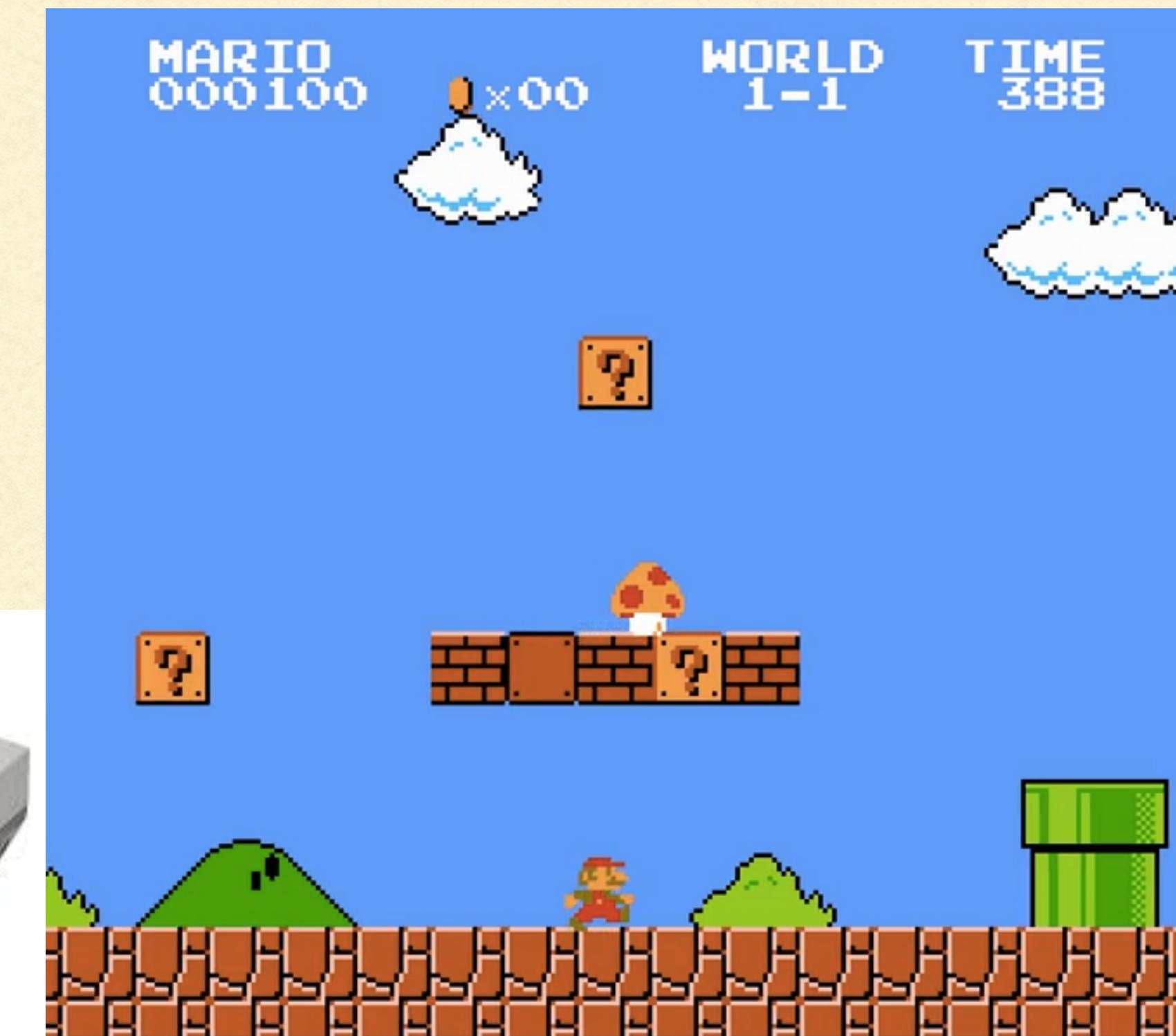
How about automating it?

- Program Learns to Play Mario
- Observes the game & presses keys
- Maximises Score





So?



So?

- Program Learnt to play Mario
 - and other games
- Without any need of programming

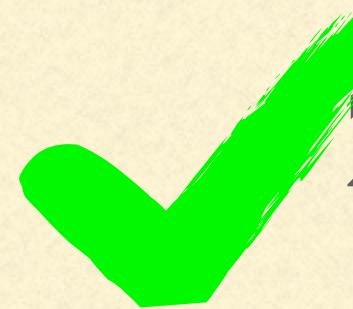


Question:

To make this program learn any other games such as PacMan we will have to ...

1. write new rules as per the game
2. just hook it to new game and let it play for a while

Q: To play make this program learn any other games such as snakes and ladders, we will have to ...

- I. write new rules as per the game
 - 2. just hook it to new game and let it play for a while**
- 

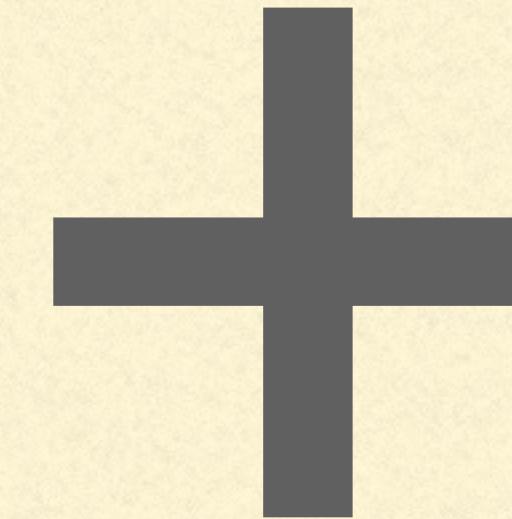


NOW IMAGINE DOING THE SAME FOR LIFE

Gather data and automatically solve problems

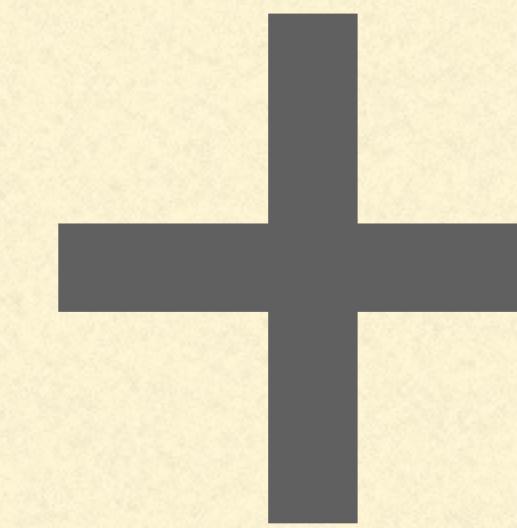
**To Gather data
&
Automatically solve problems
What do we need?**

Data



Intelligence

Big
Data



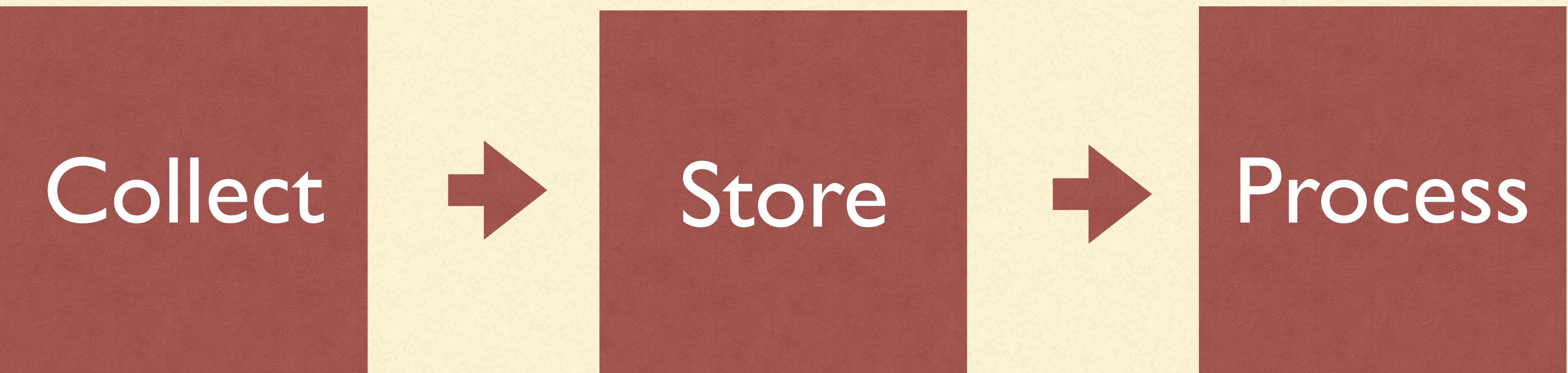
Artificial
Intelligence

WHAT IS BIG DATA?



- Simply: Data of Very Big Size
- Can't process with usual tools
- **Distributed Architecture Needed**
- Structured / Unstructured

FOR DATA, WE NEED ...



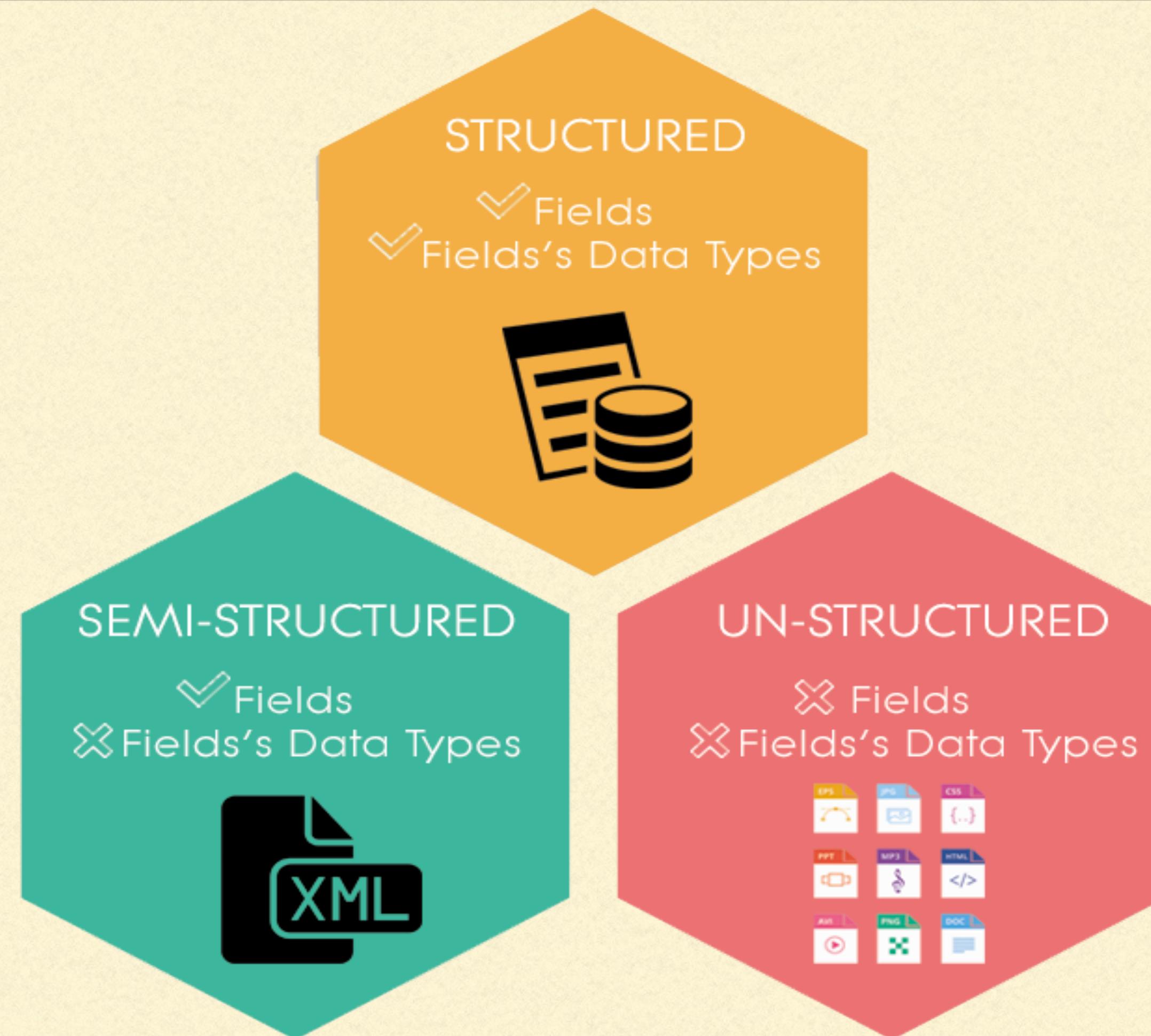
COLLECT DATA - IOT



Phone & Devices
Cheaper, faster and smaller.

Connectivity
Wifi, 4G, NFC, GPS

STORE DATA - TYPES



STORE DATA - CHARACTERISTICS

VOLUME

Data At Rest



Problems related to storage
of huge data reliably.

e.g. Storage of Logs of a
website, Storage of data by
gmail.

FB: 300 PB. 600TB/ day

VELOCITY

Data In Motion



Problems Involving the handling of
data coming at fast rate.

e.g. Number of requests being
received by Facebook, Youtube
streaming, Google Analytics

VARIETY

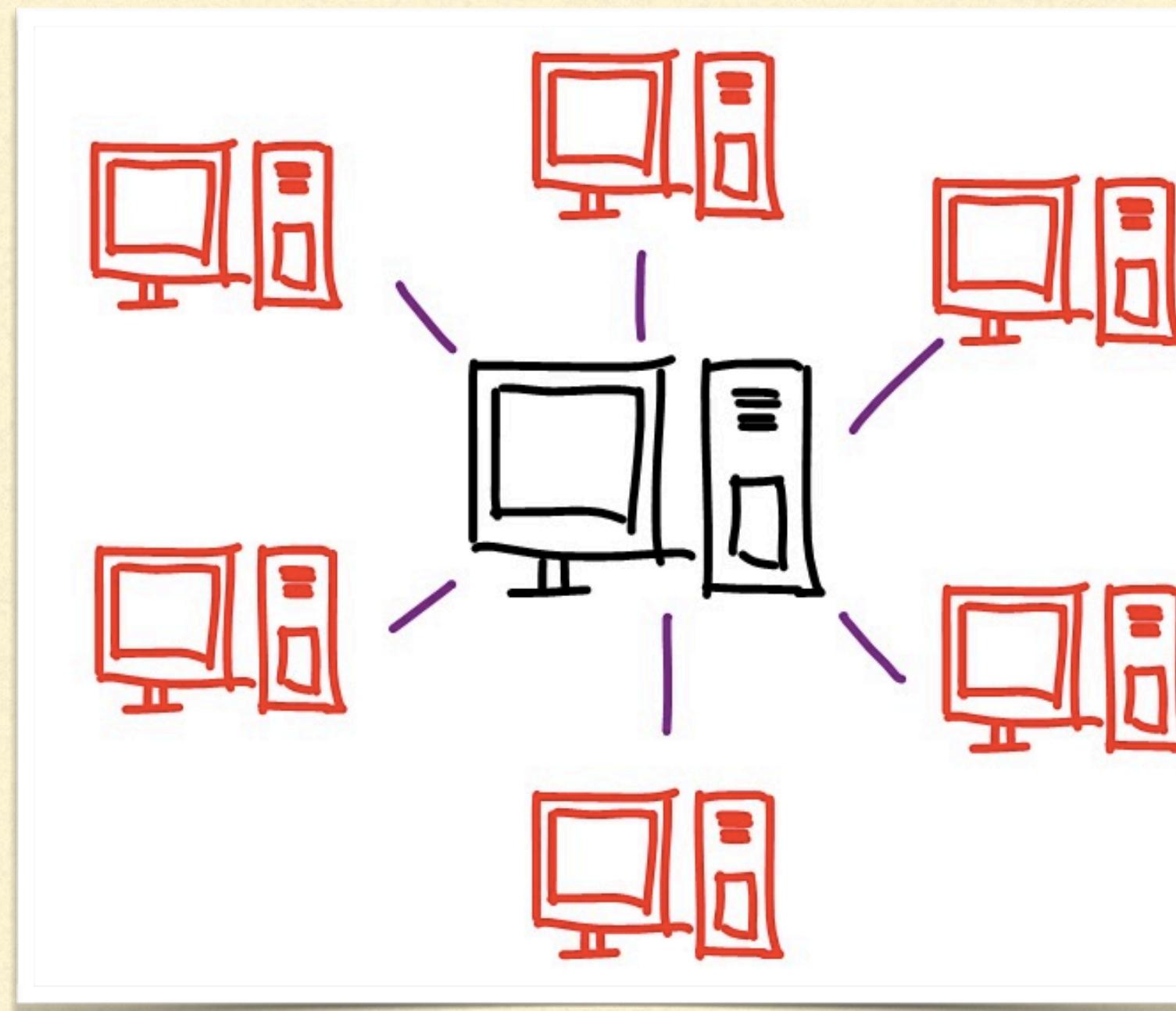
Data in Many Forms



Problems involving complex data
structures

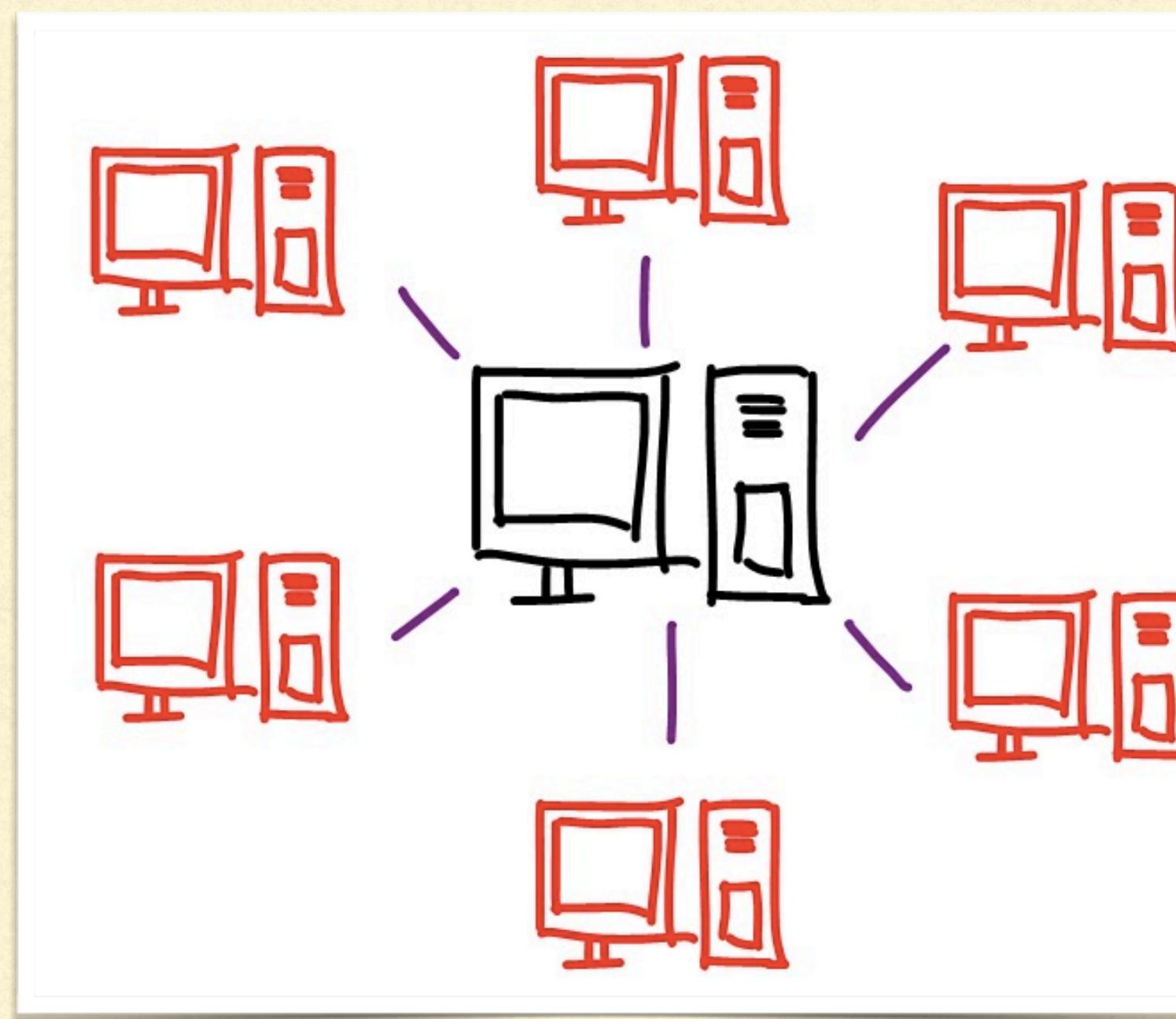
e.g. Maps, Social Graphs,
Recommendations

STORE DATA - SYSTEMS



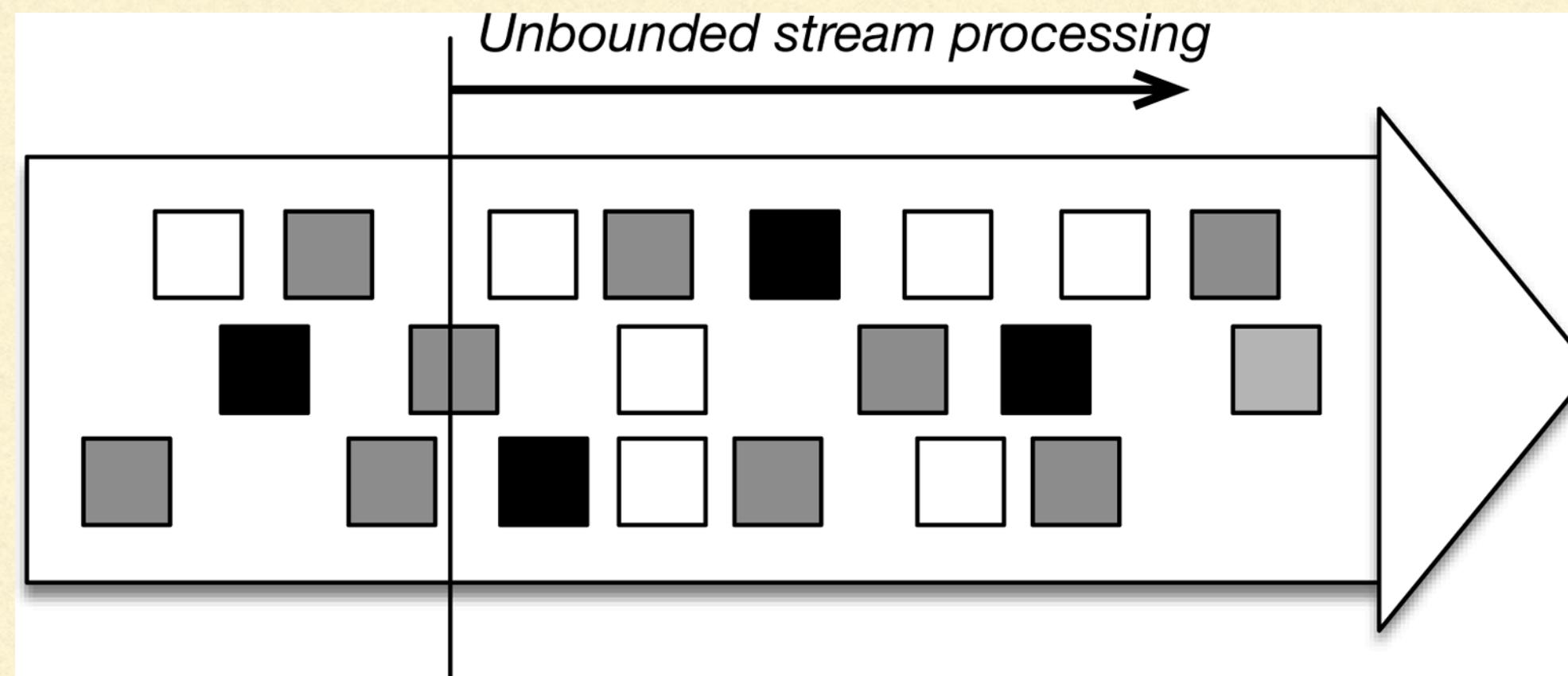
- Big Data Storage Systems
 - Unstructured (Files)
 - Structured (Record wise)

STORE DATA - SYSTEMS



- Big Data Storage Systems
 - Unstructured (Files)
 - HDFS - Hadoop Distributed File Systems
 - Amazon S3
 - Structured (Record wise)
 - HBASE
 - Cassandra

STORE DATA - FAST DATA



- To Store fast data
 - Kafka
 - Amazon SQS

PROCESS DATA - WHAT IS PROCESSING



- Any instruction - Requiring CPU
- Measured By Big-O
- **Examples**
 - Code - Read/Write, Conditions, Jumps
 - Queries (SQL)
 - Complex Algorithms - Machine Learning

PROCESS DATA - RESOURCES

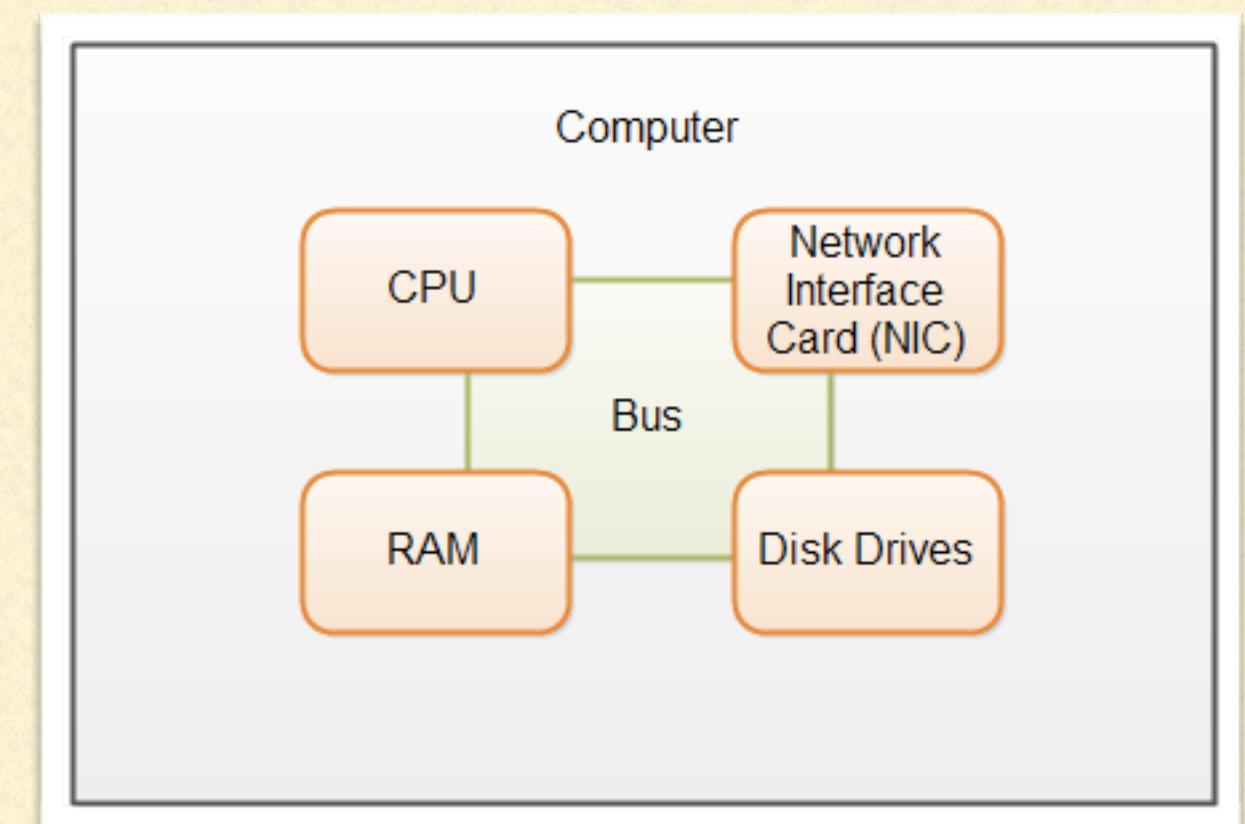


1. CPU Speed



2. RAM - Speed & Size

To process data
We need...



4. Network



3. HDD or SSD
Disk Size + Speed

PROCESS DATA

Which components impact the speed computing?

- A. CPU
- B. Memory
- C. Memory Read Speed
- D. Disk Speed
- E. Disk Size
- F. Network Speed
- G. All of Above

PROCESS DATA

Which components impact the speed computing?

- A. CPU
- B. Memory
- C. Memory Read Speed
- D. Disk Speed
- E. Disk Size
- F. Network Speed
- G. All of Above**

PROCESS DATA - PARALLEL COMPUTING

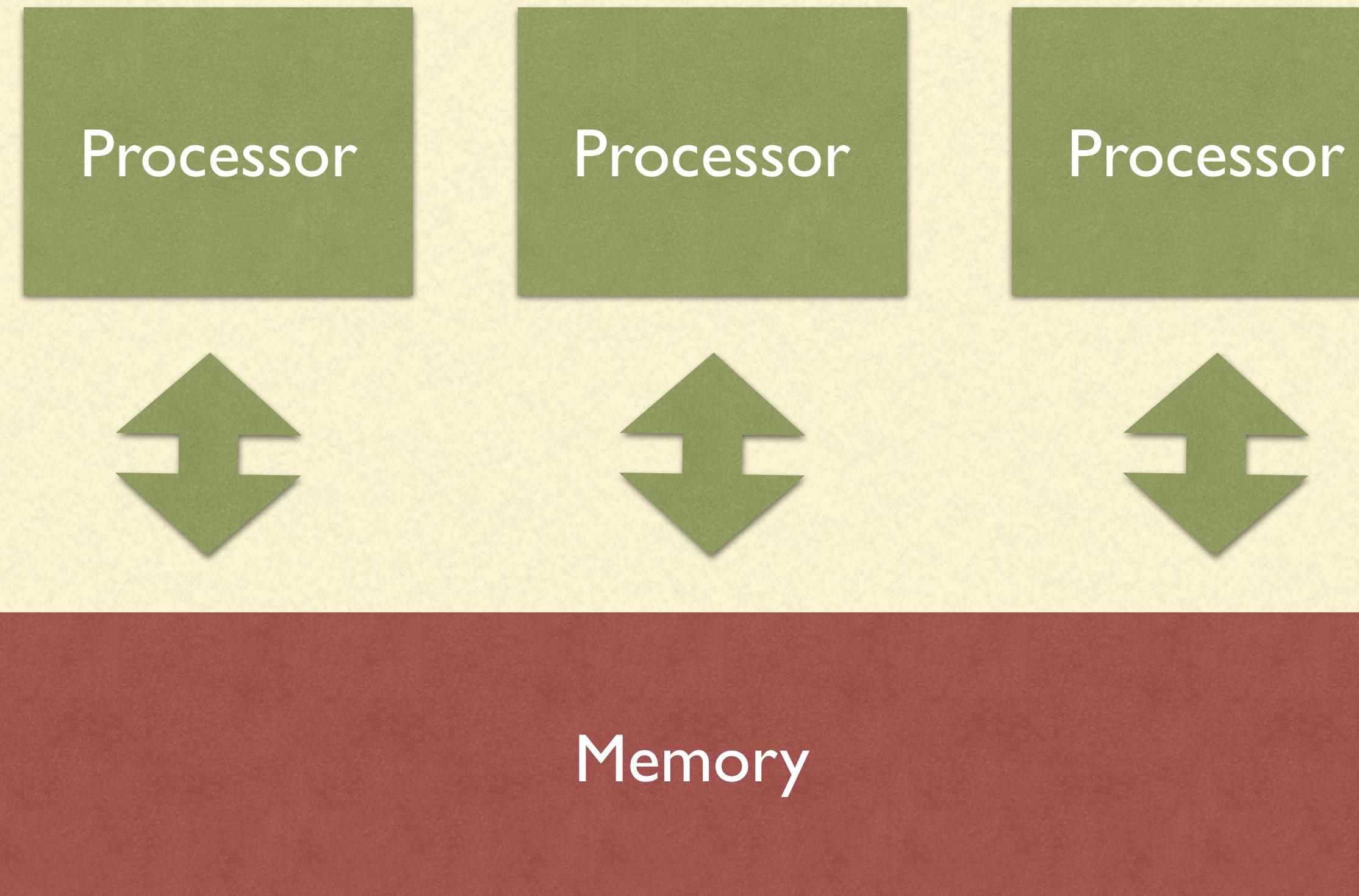
Distributed



- Groups of networked computers
- Interact with each other
- To achieve a common goal.

PROCESS DATA - PARALLEL COMPUTING

Multi Core + GPGPU (General Purpose Graphics Processing Units)

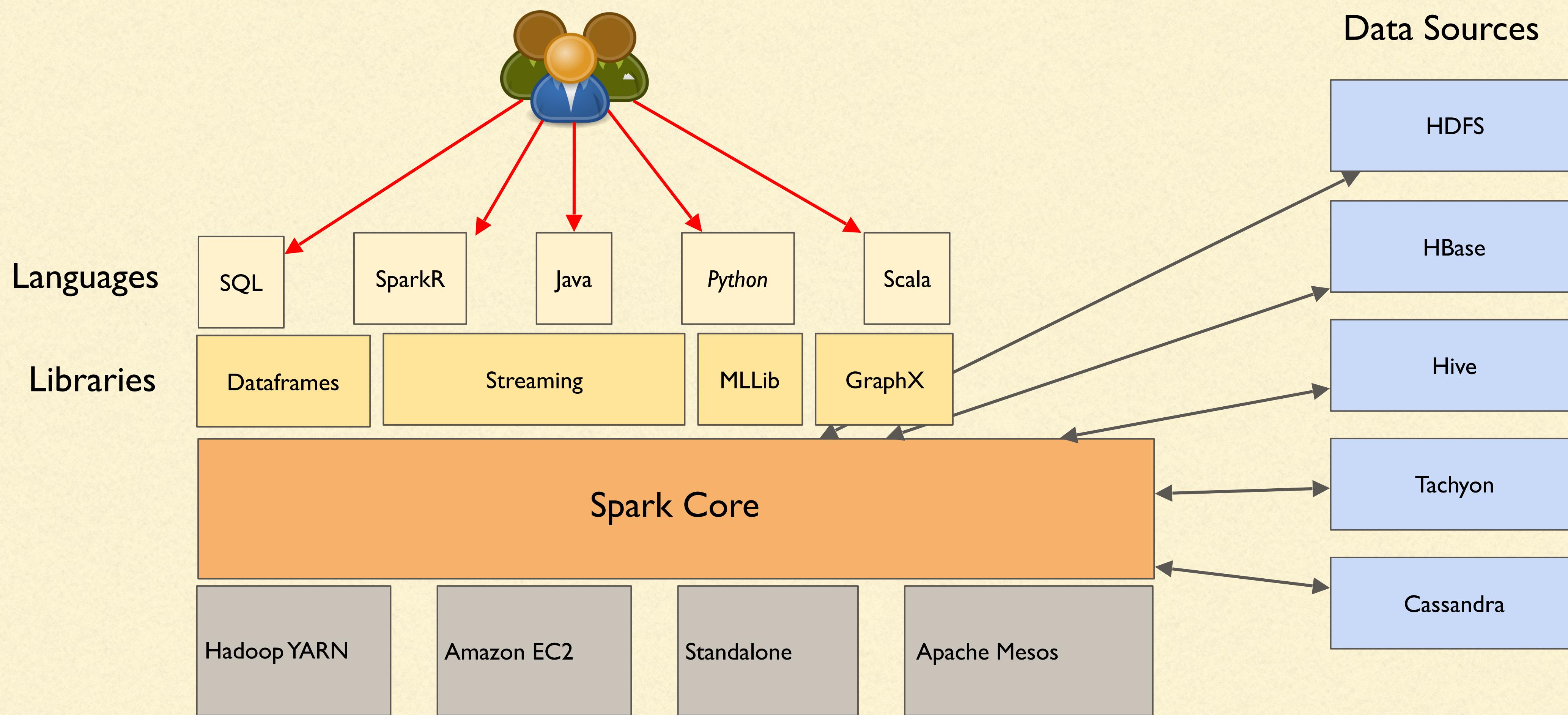


- Many processors or Cores
- Perform tasks and interact using
- Memory or bus

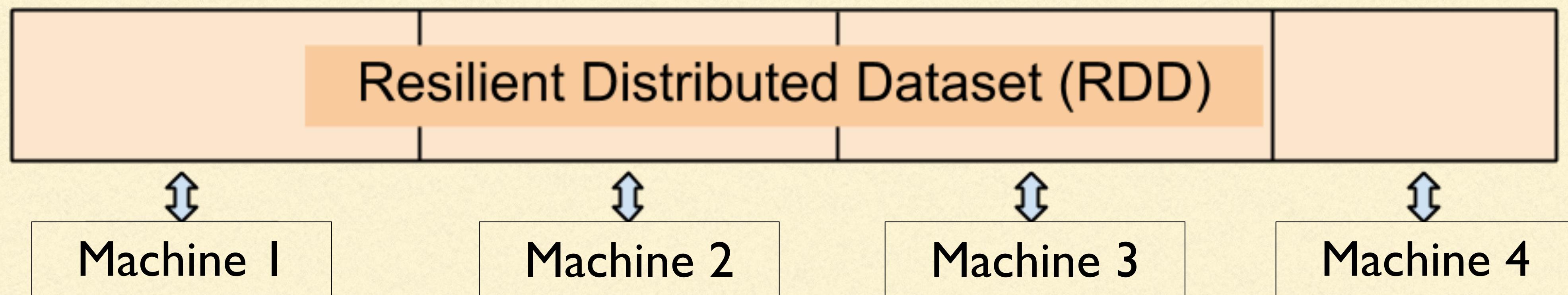
PROCESS DATA - PARALLEL COMPUTING

	MULTI CORE	GPGPU	DISTRIBUTED
CAN HANDLE HUGE DATA? (DISK READ INTENSIVE)			
REALLY FAST COMMUNICATION BETWEEN CPUS			
GREAT FOR MATHS/GRAFICS?			
TOOLS	Hadoop MR, Apache Spark	Keras, Tensorflow, Caffe, Spark (Exp)	Hadoop MR, Apache Spark

PROCESS DATA - APACHE SPARK



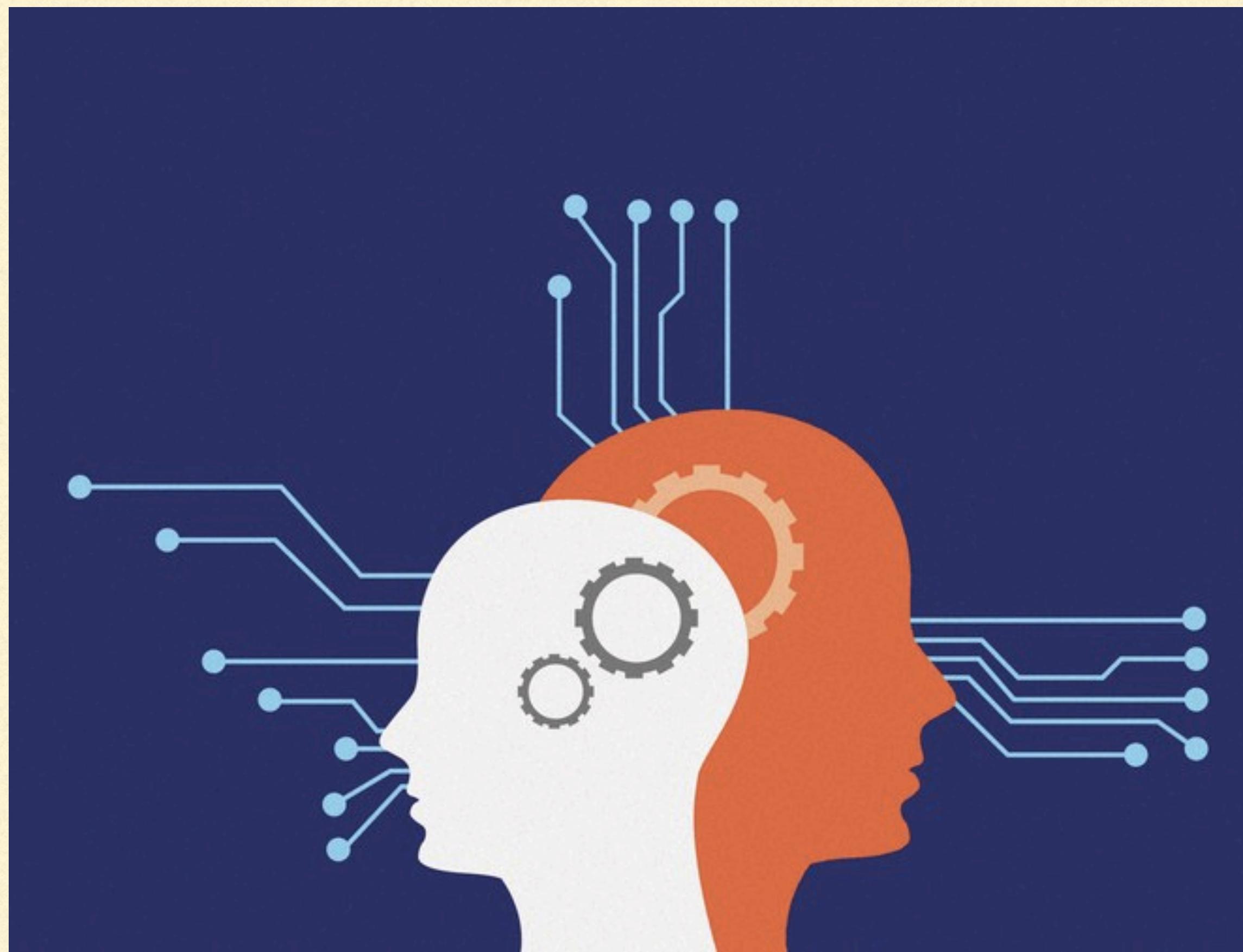
PROCESS DATA - APACHE SPARK



WHAT IS AI?

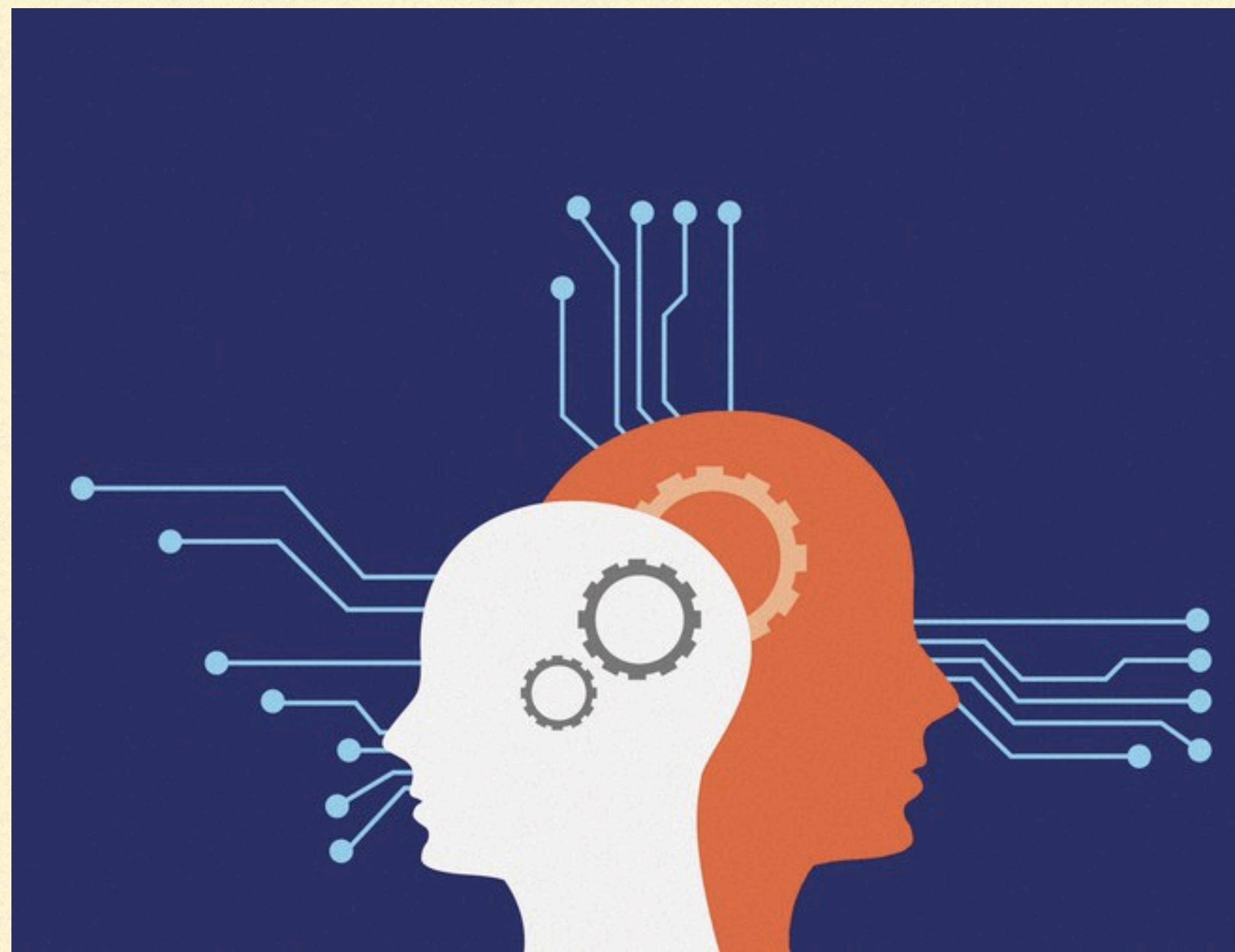
Artificial intelligence (AI) is intelligence exhibited by machines.

WHAT IS AI?



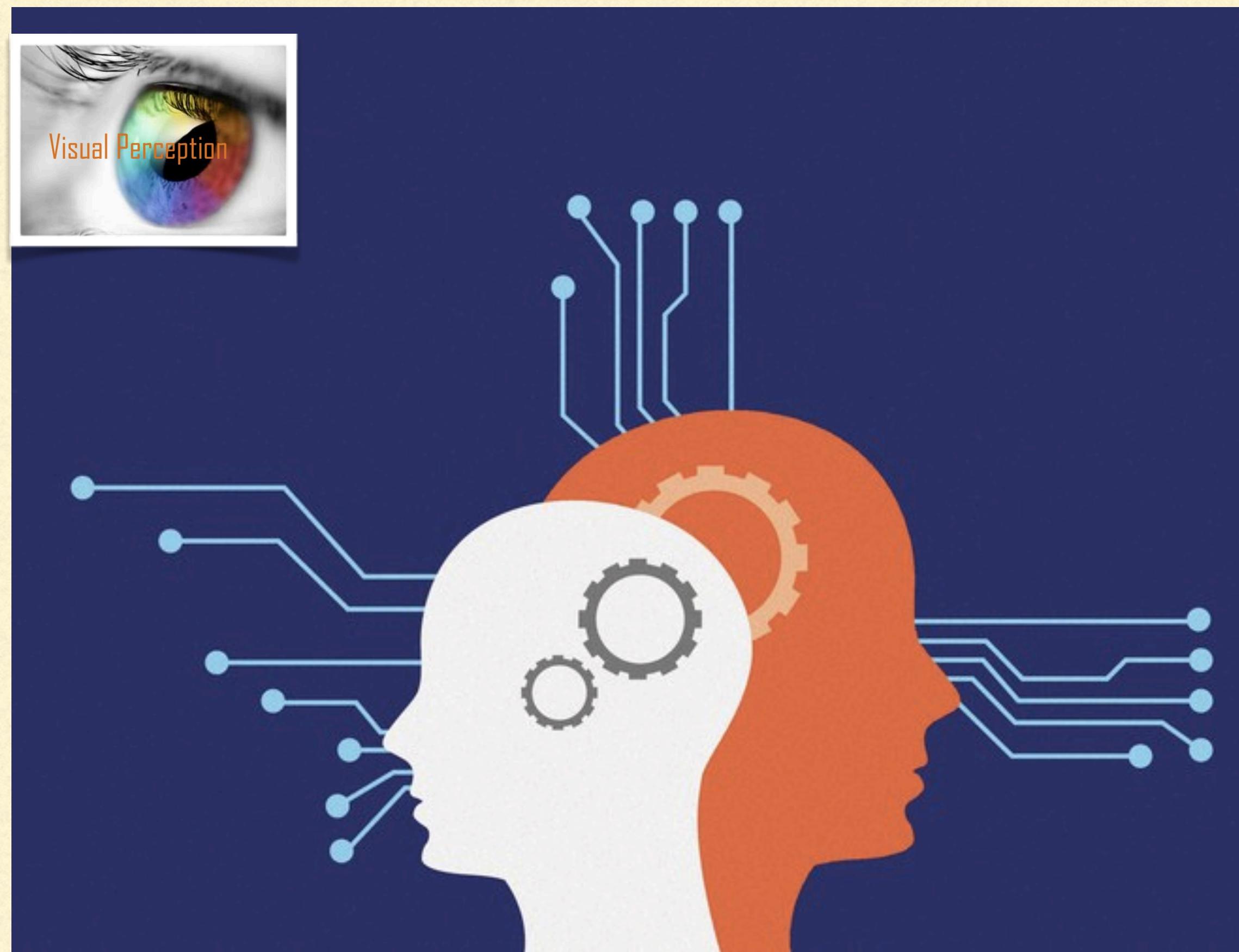
- The theory and development of computer systems

WHAT IS AI?



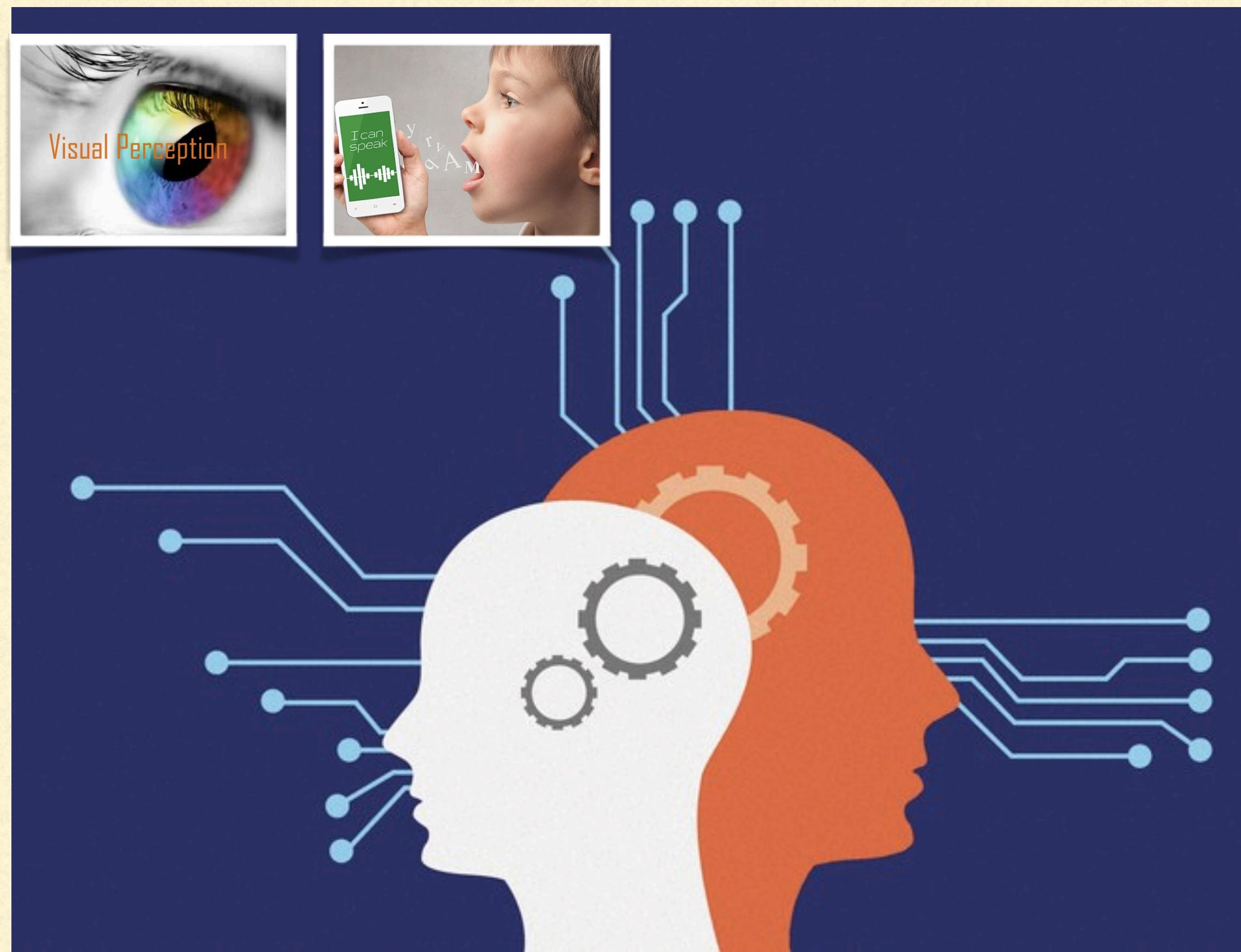
- The theory and development of computer systems
- To perform tasks requiring human intelligence such as

WHAT IS AI?



- The theory and development of computer systems
- To perform tasks requiring human intelligence such as
 - Visual perception

WHAT IS AI?



- The theory and development of computer systems
- To perform tasks requiring human intelligence such as
 - Visual perception
 - Speech recognition,

WHAT IS AI?



- The theory and development of computer systems
- To perform tasks requiring human intelligence such as
 - Visual perception
 - Speech recognition,
 - Decision-making,

WHAT IS AI?



- The theory and development of computer systems
- To perform tasks requiring human intelligence such as
 - Visual perception
 - Speech recognition,
 - Decision-making,
 - **Translation between languages.**

HISTORY - MYTHOLOGY / FICTION



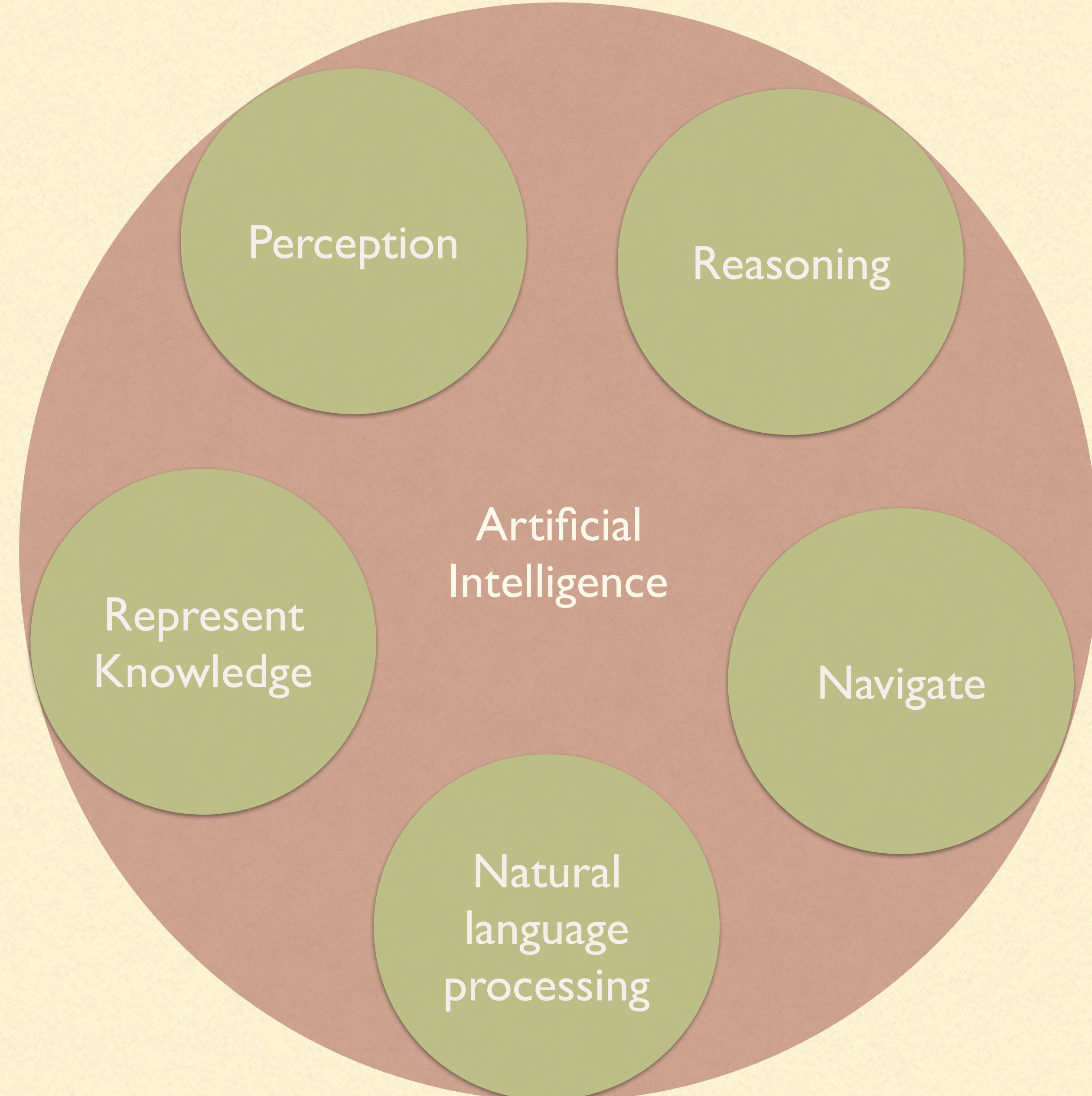
- In every mythology, there is some form of mechanical man such talos from greek mythology.
- In fiction novels, we have Mary Shelley's Frankenstein
- We are fascinated by the idea of creating things which can behave like human

HISTORY - SUMMERY OF 1956



- The term artificial intelligence was coined
- by John McCarthy
- In a workshop at
- Dartmouth College in New Hampshire.
- Along with Marvin Minsky, Claude Shannon, and Nathaniel Rochester.

Sub objectives of AI



WHAT IS AI - REPRESENT KNOWLEDGE



- Understanding and classifying terms or things in world e.g.
 - What is computer?
 - What is a thought?
 - What is a tool?
- Languages like lisp were created for the same purpose

WHAT IS AI - REASONING



Aja Huang plays the first move for AlphaGo, against Lee Sedol 9 dan in game two.

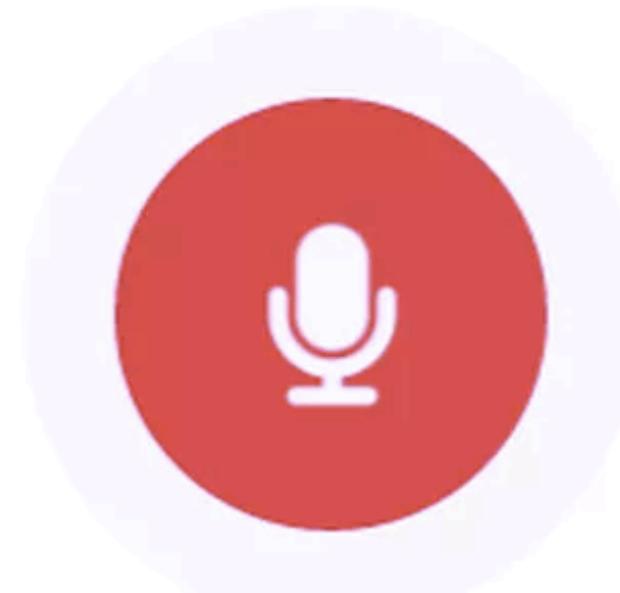
- Play puzzle game - Chess, Go, Mario
- Prove Geometry theorems
- Diagnose diseases

WHAT IS AI - NAVIGATE



- How to plan and navigate in the real world
 - How to locate the destination?
 - How to pick path?
 - How to pick short path?
 - How to avoid obstacles?
 - How to move?

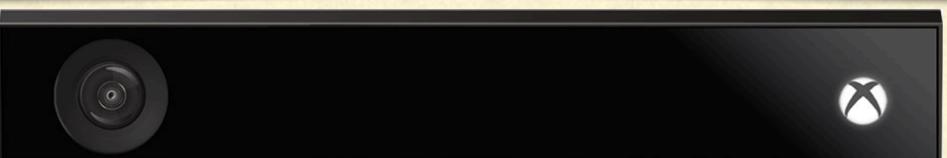
AI - NATURAL LANGUAGE PROCESSING:



Ok Google - can you show me a list of your commands?

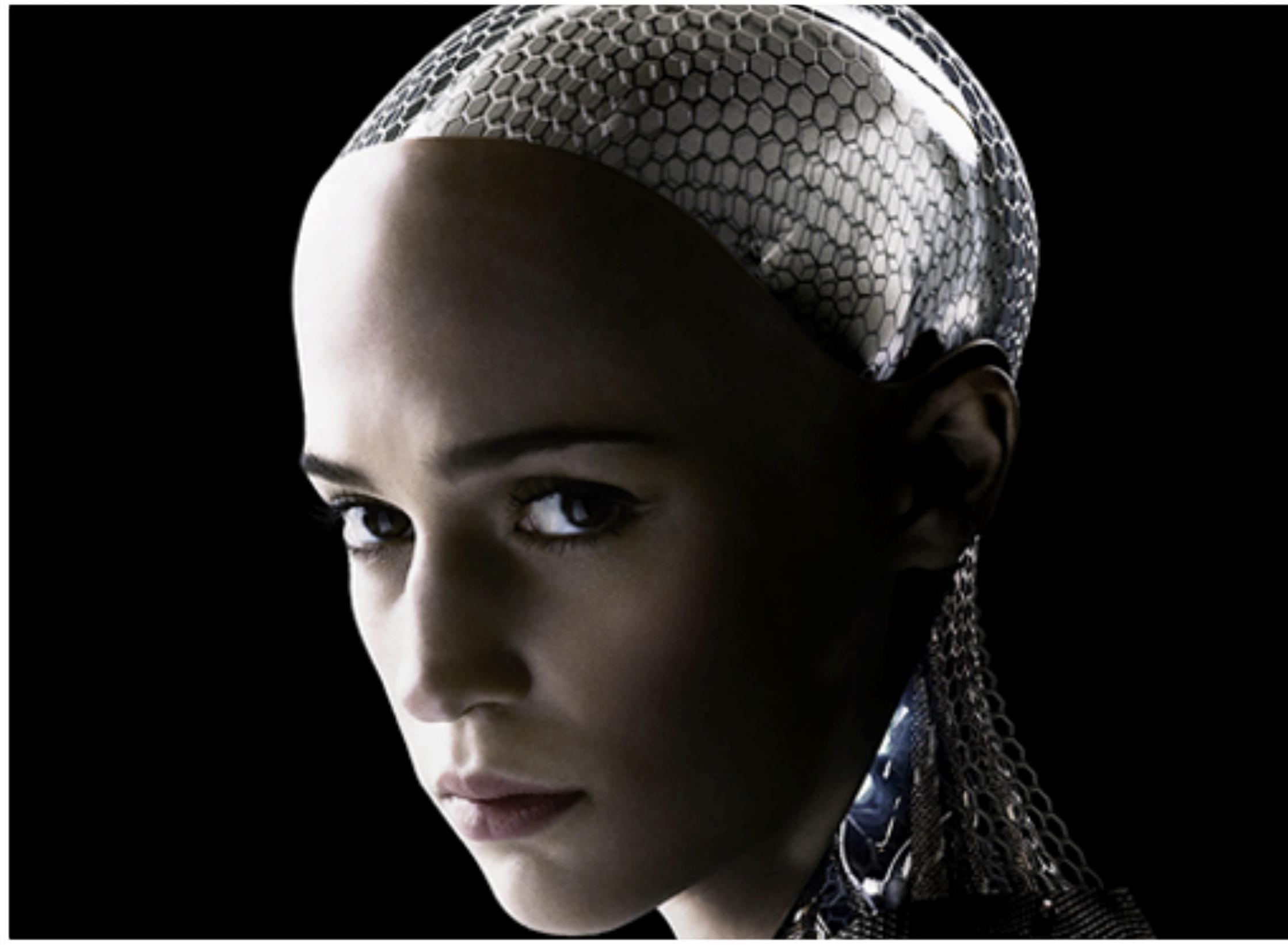
- How to speak a language
- How to understand a language
- How to make sense out of a sentence

AI - PERCEPTION



- How to we see things in the real world
- From sound, sight, touch, smell

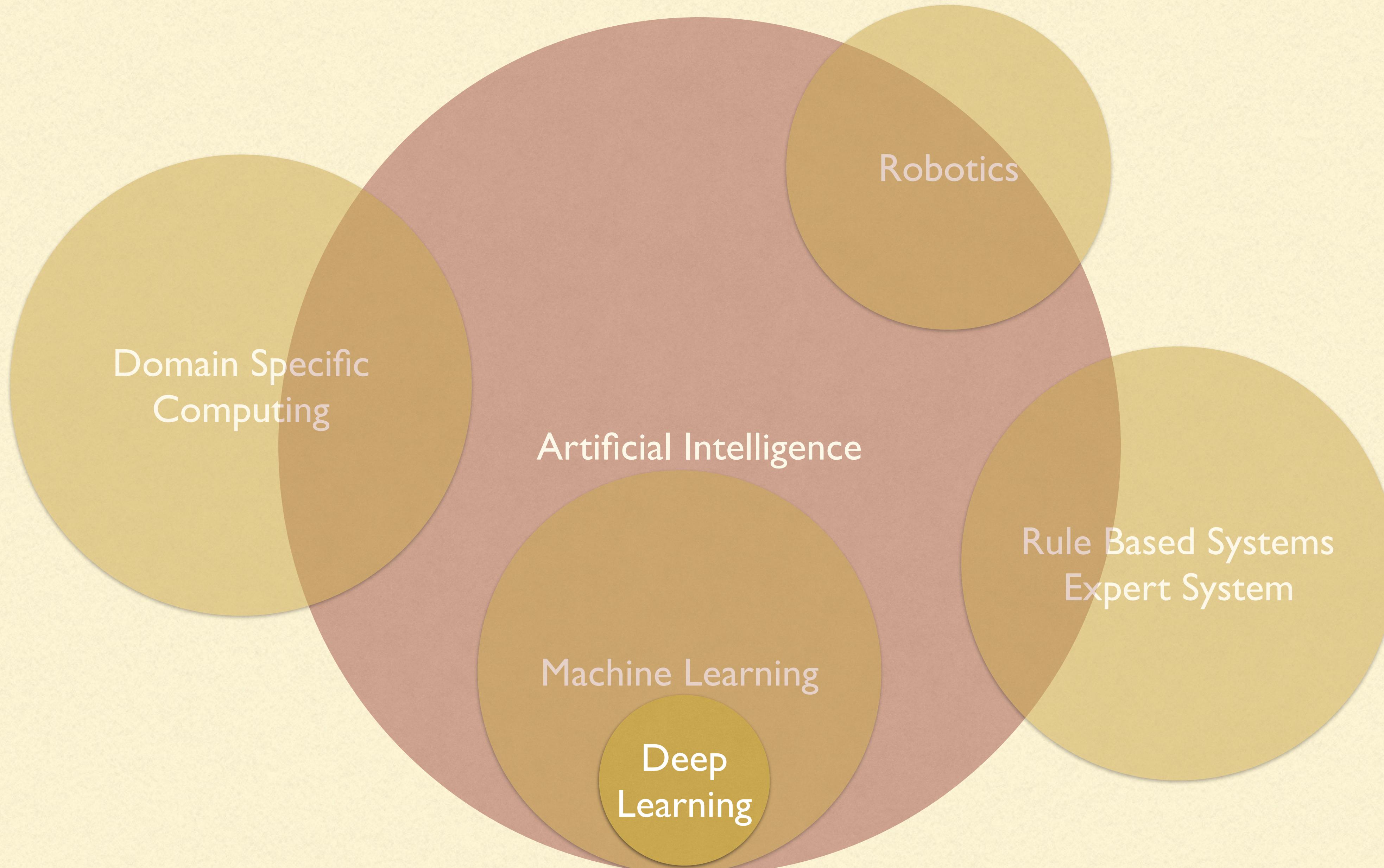
AI - GENERALISED INTELLIGENCE:



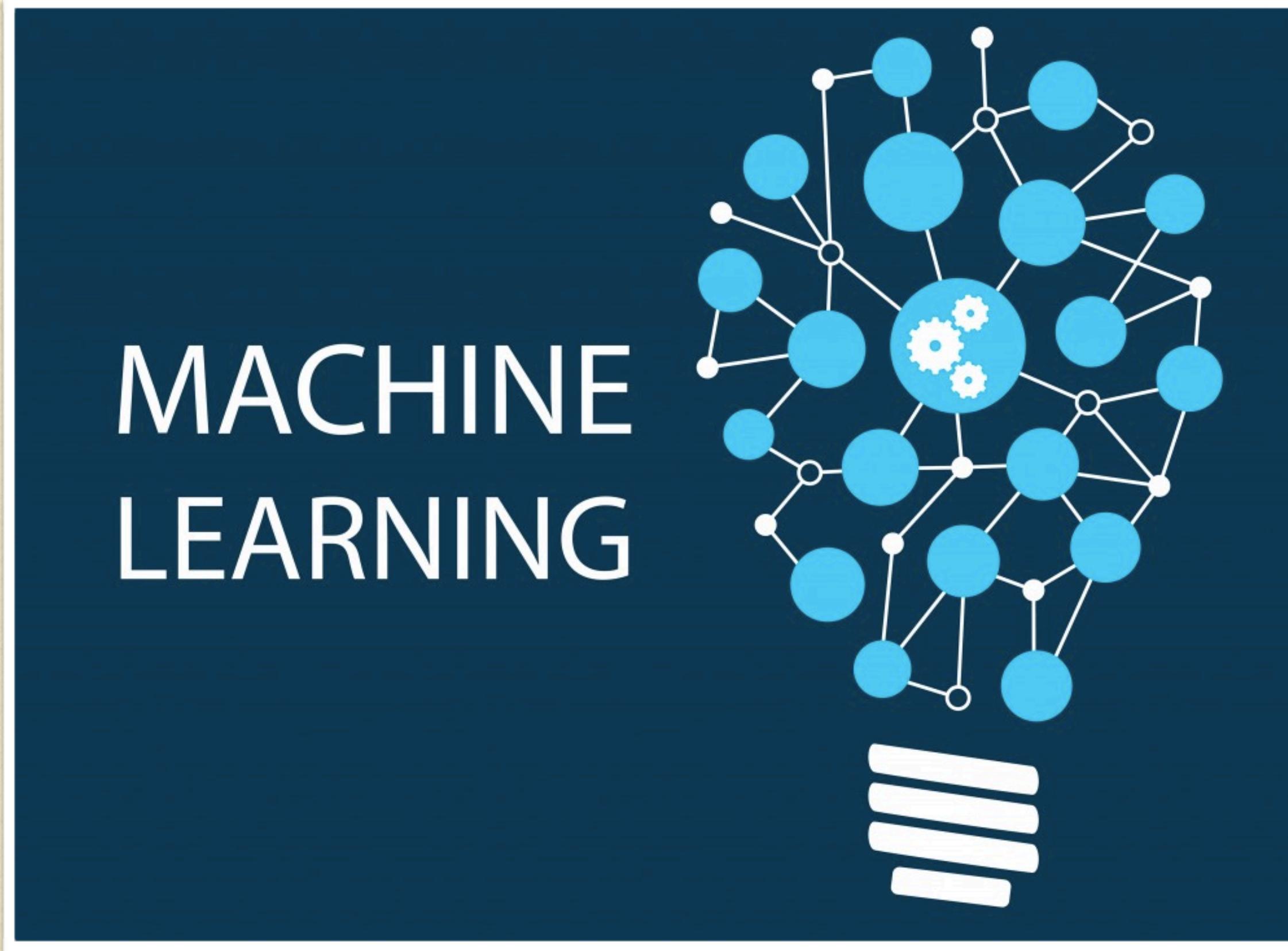
With these previous building blocks, the following should emerge:

- Emotional Intelligence
- Creativity
- Reasoning
- Intuition

How to achieve AI?



MACHINE LEARNING?

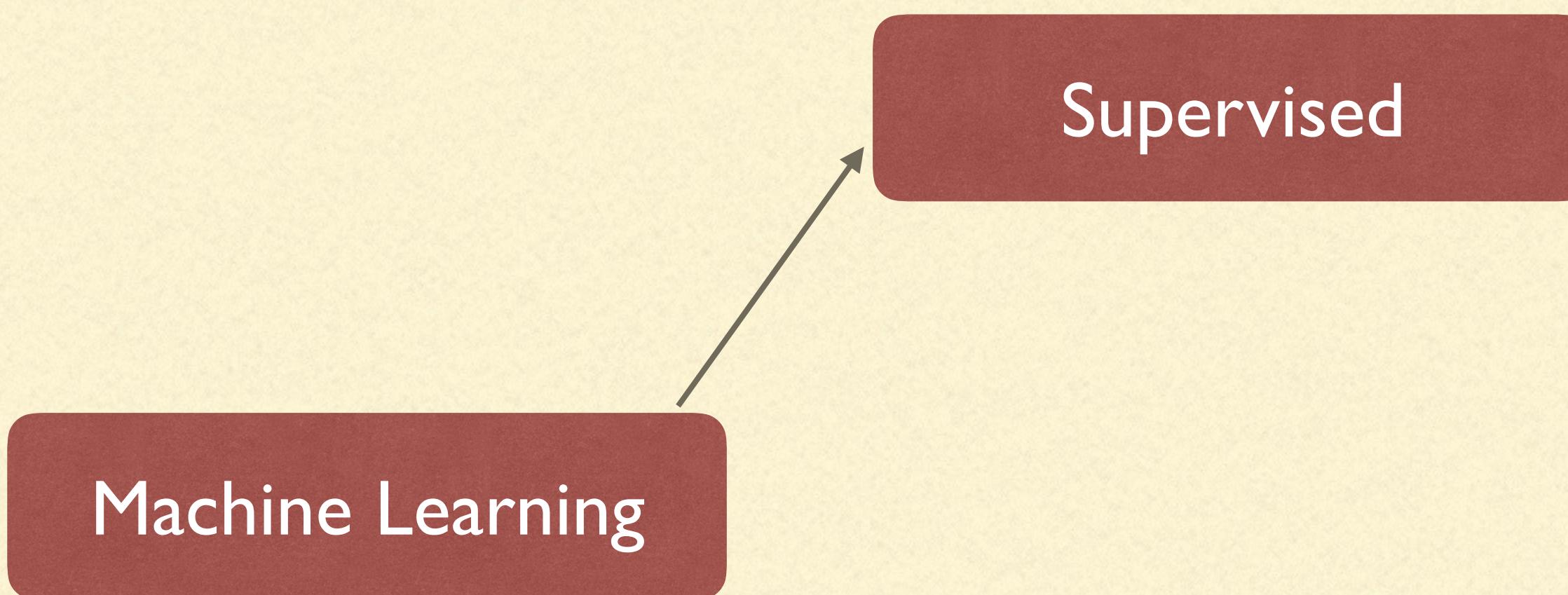


Field of study that gives "computers the ability to learn without being explicitly programmed."

-- Arthur Samuel, 1959

Branch of AI

MACHINE LEARNING - TYPES?



Given example inputs & outputs,
learn to map inputs to outputs

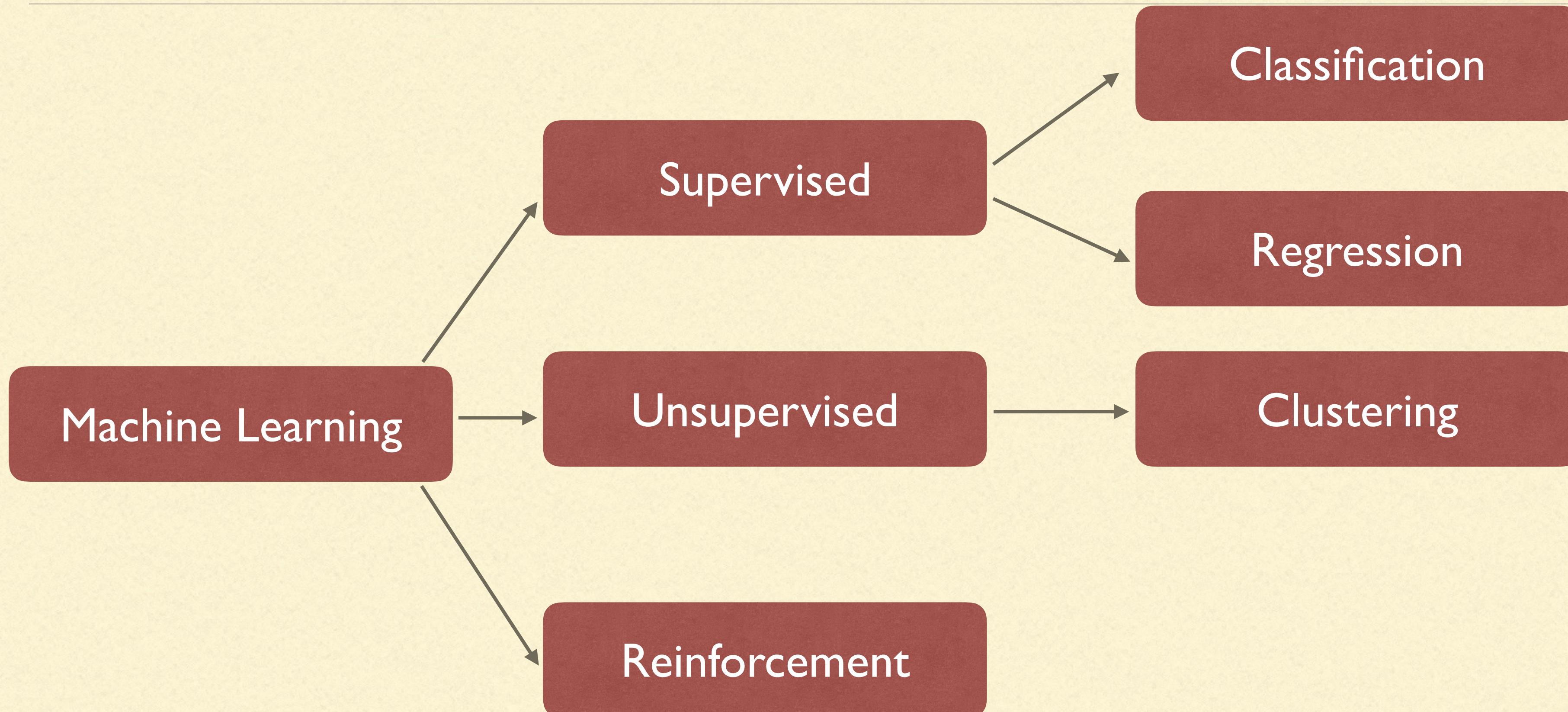
MACHINE LEARNING - TYPES?



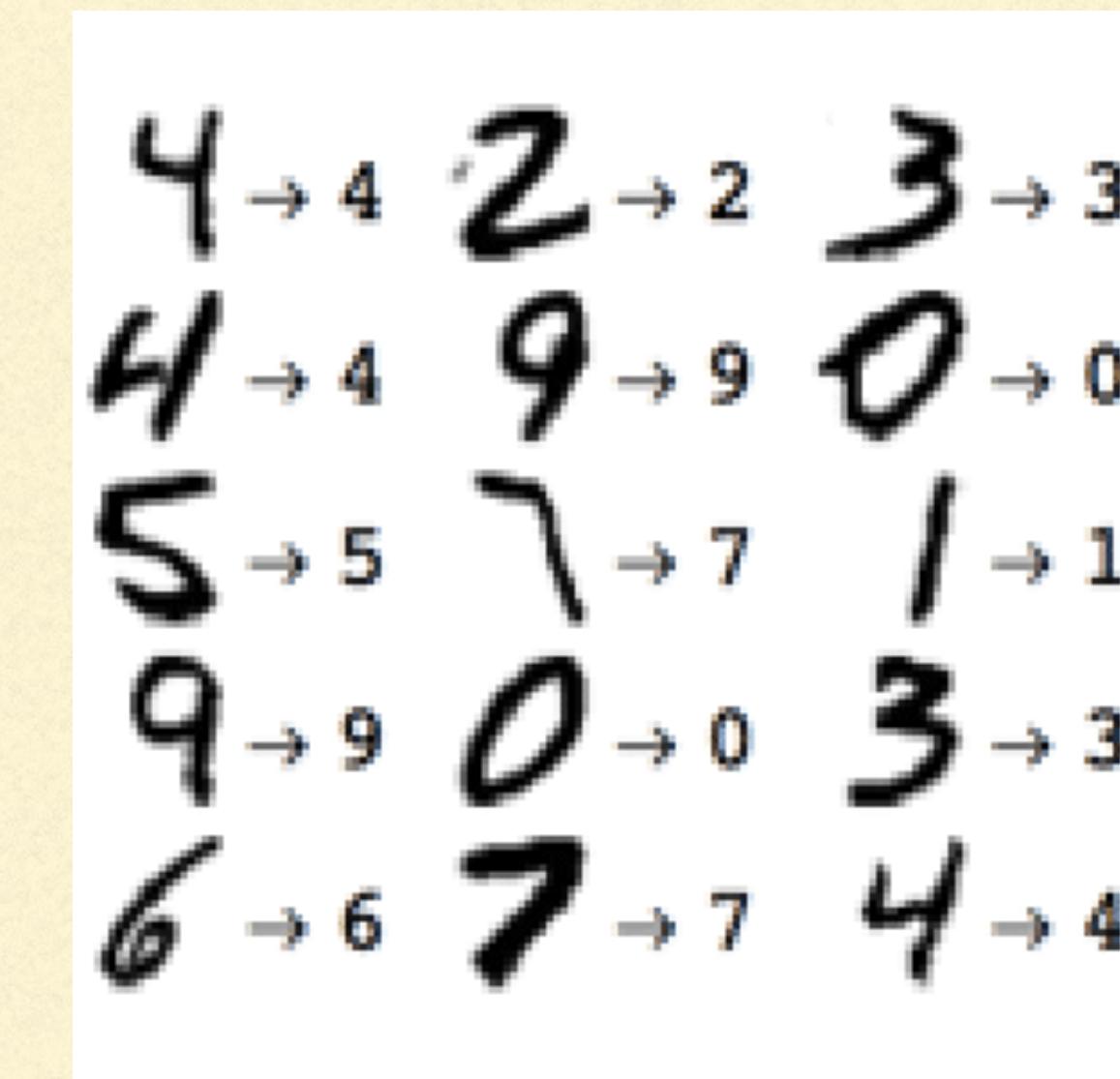
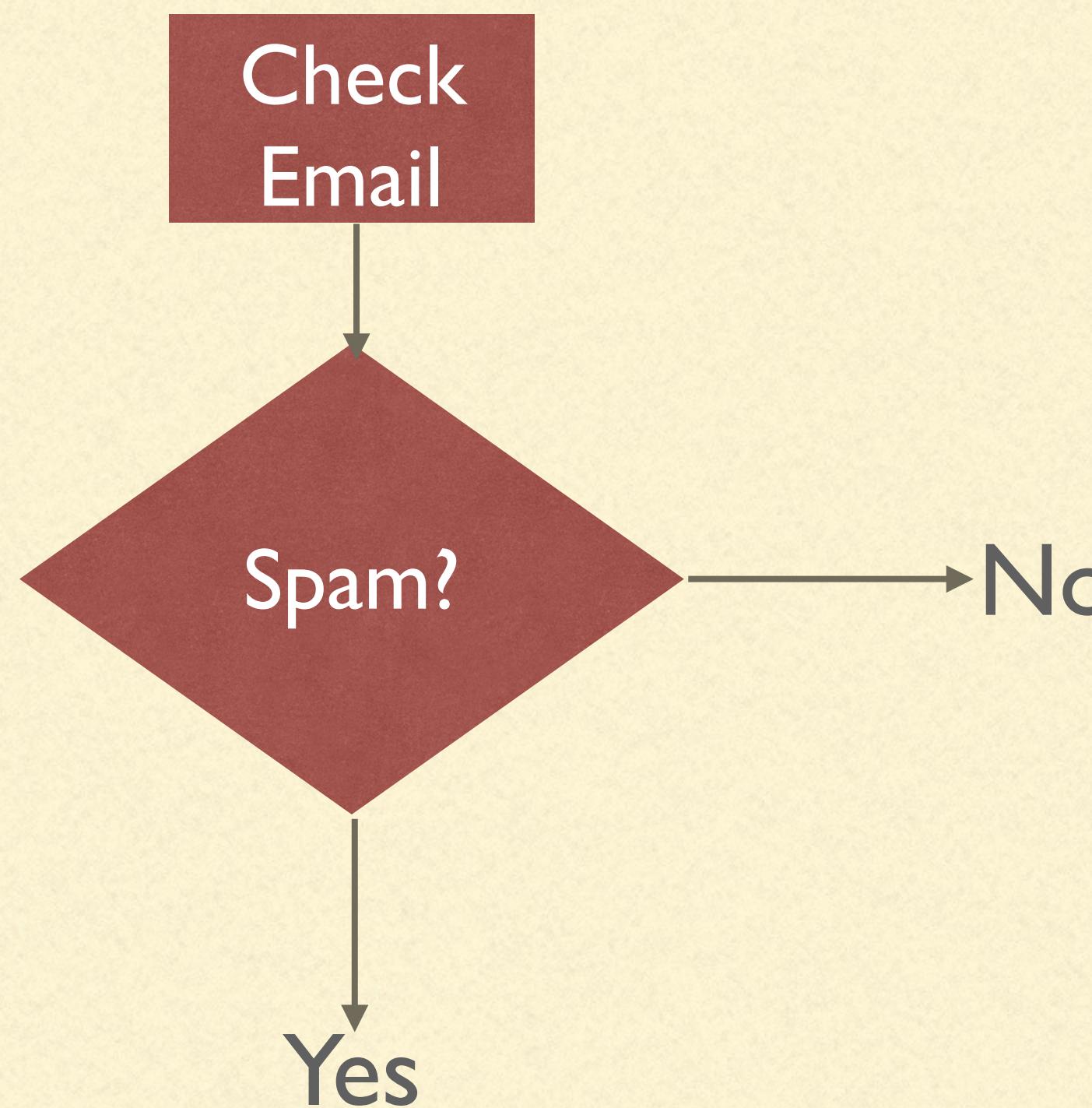
MACHINE LEARNING - TYPES?



MACHINE LEARNING - TYPES?



MACHINE LEARNING - CLASSIFICATION?



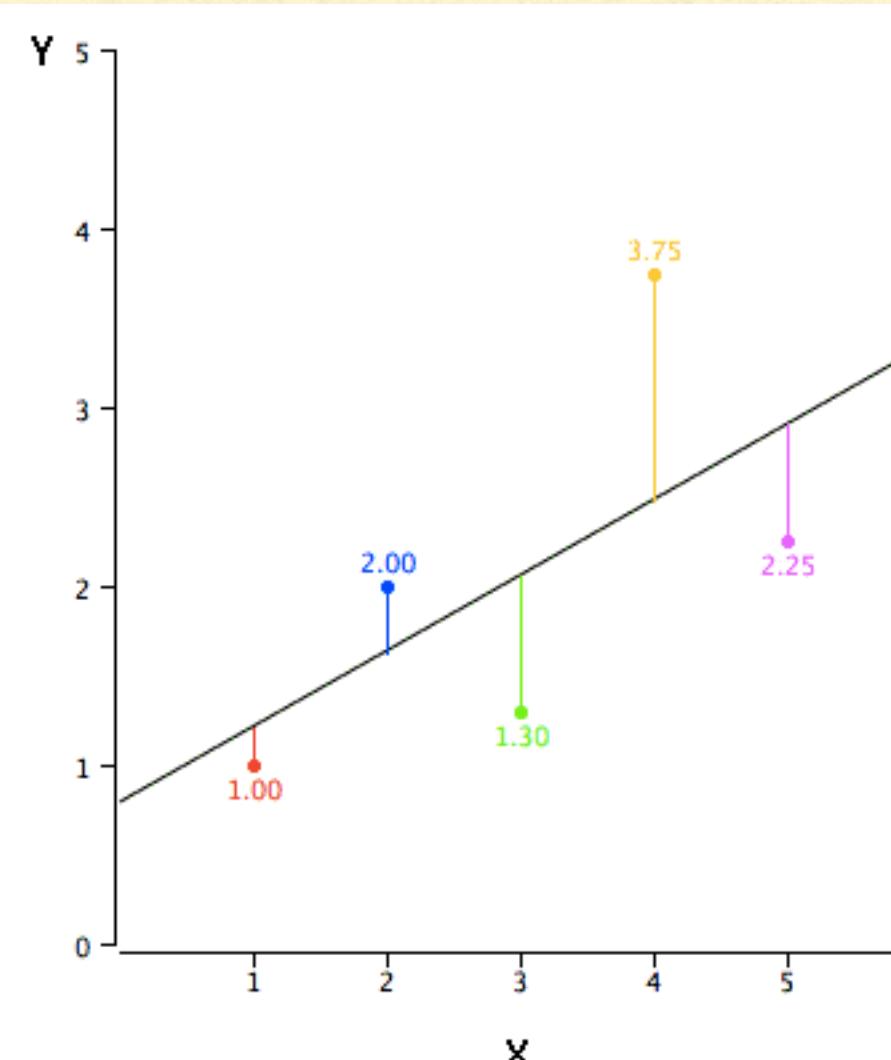
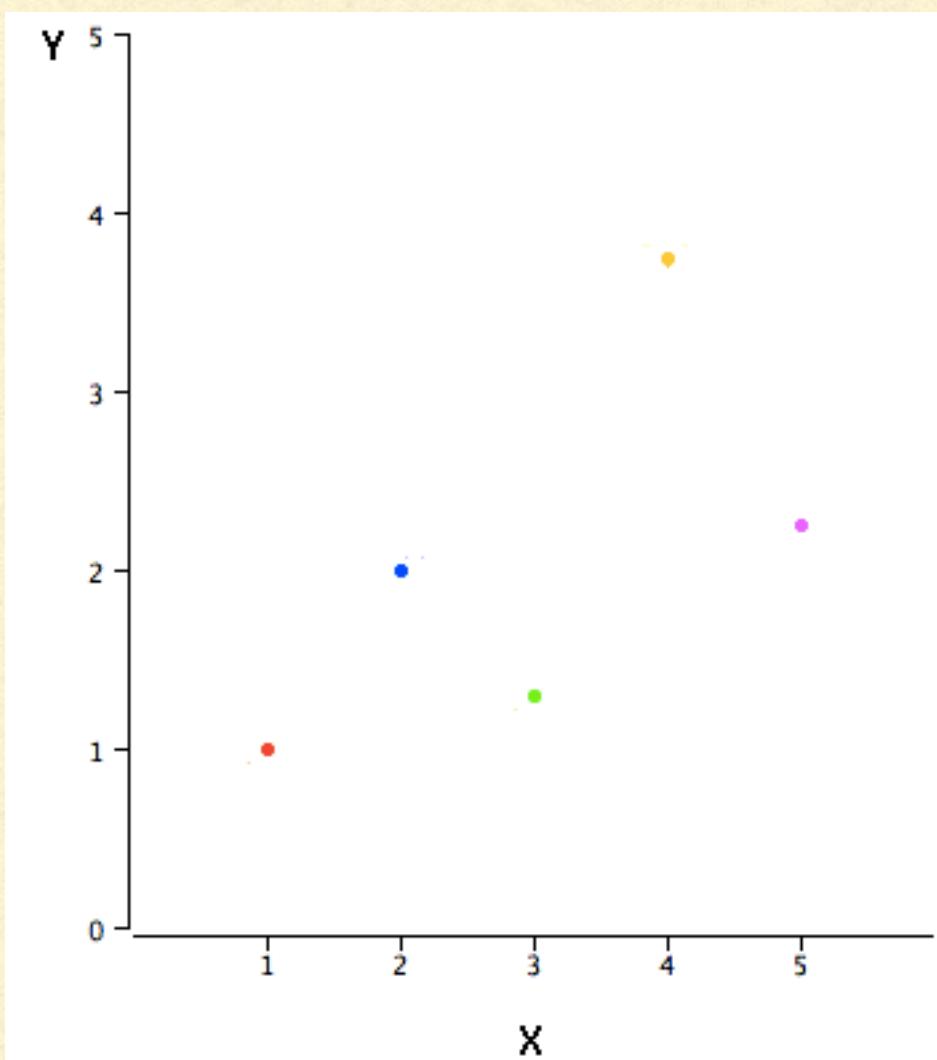
We Use Logistic Regression

MACHINE LEARNING - REGRESSION?

Table 1. Example data.

X	Y
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

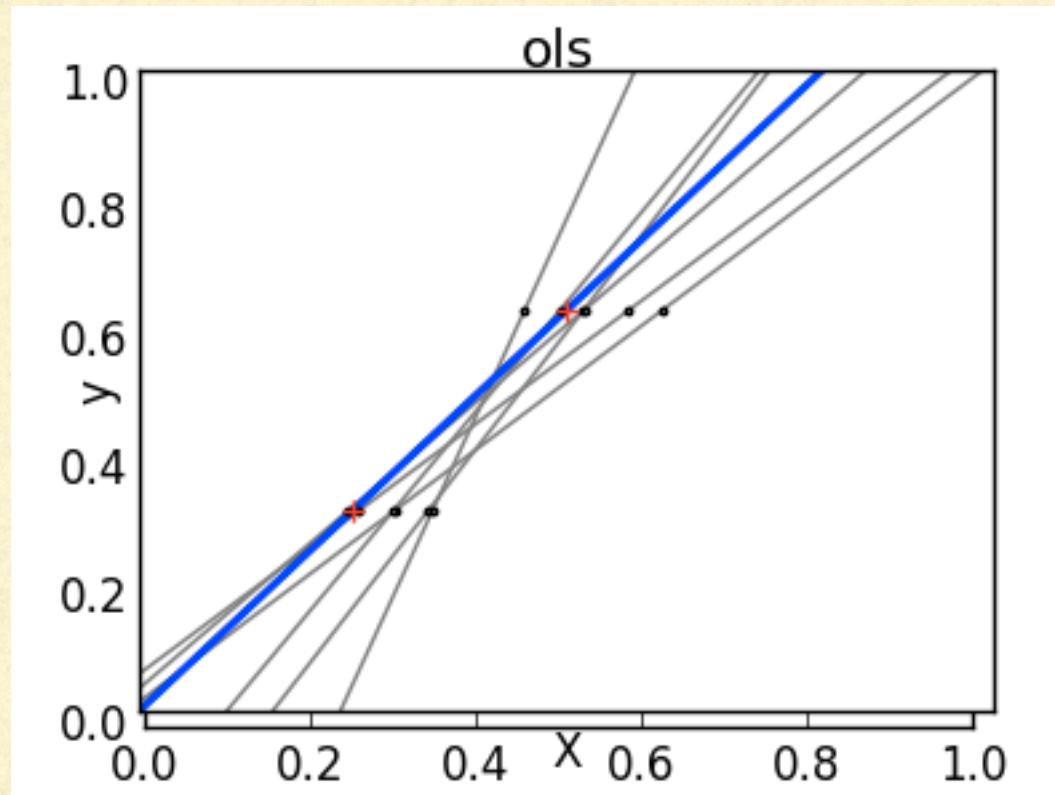
Predicting a continuous-valued attribute associated with an object.



In linear regression, we draw all possible lines going through the points such that it is closest to all.

MACHINE LEARNING - GRADIENT DESCENT

- Instead of trying all lines, go into the direction yielding better results.



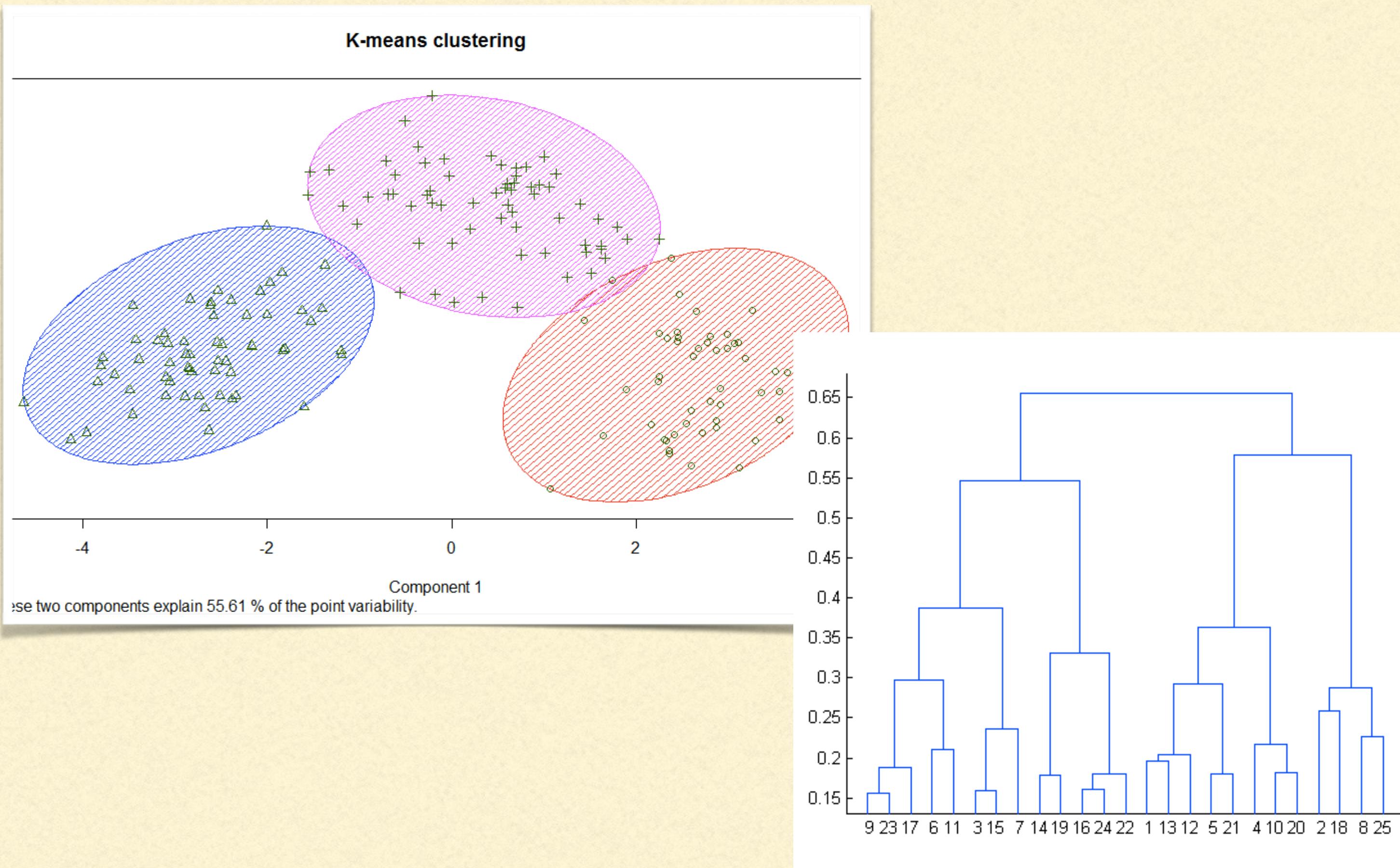
Imagine yourself blindfolded on the mountainous terrain.

And you have to find the best lowest point.

If your last step went higher, you will go in opposite direction.

Other, you will keep going just faster

ML - CLUSTERING?



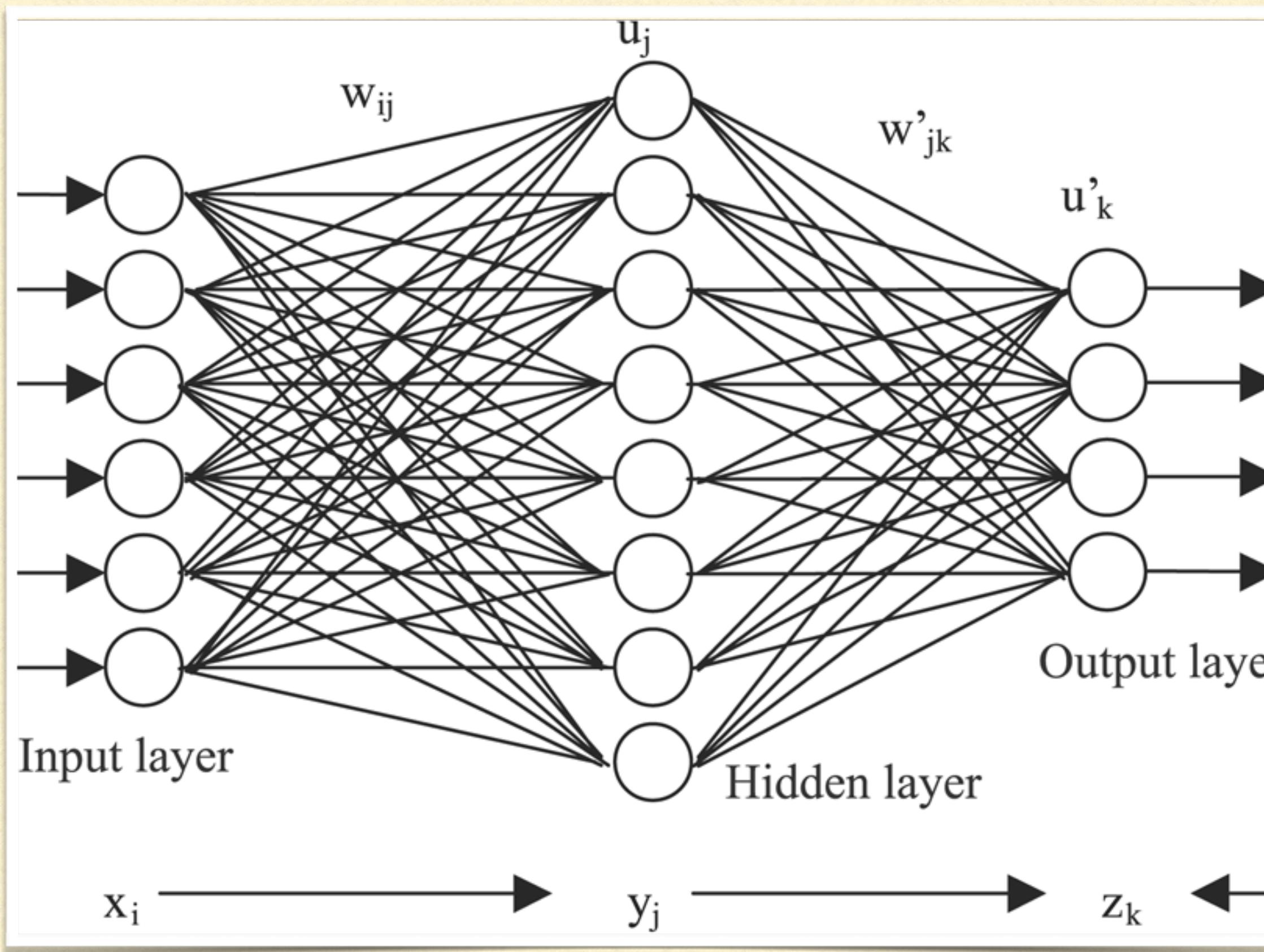
- To form a cluster
- based on some definition of nearness



ML - ARTIFICIAL NEURAL NETWORK (ANN)

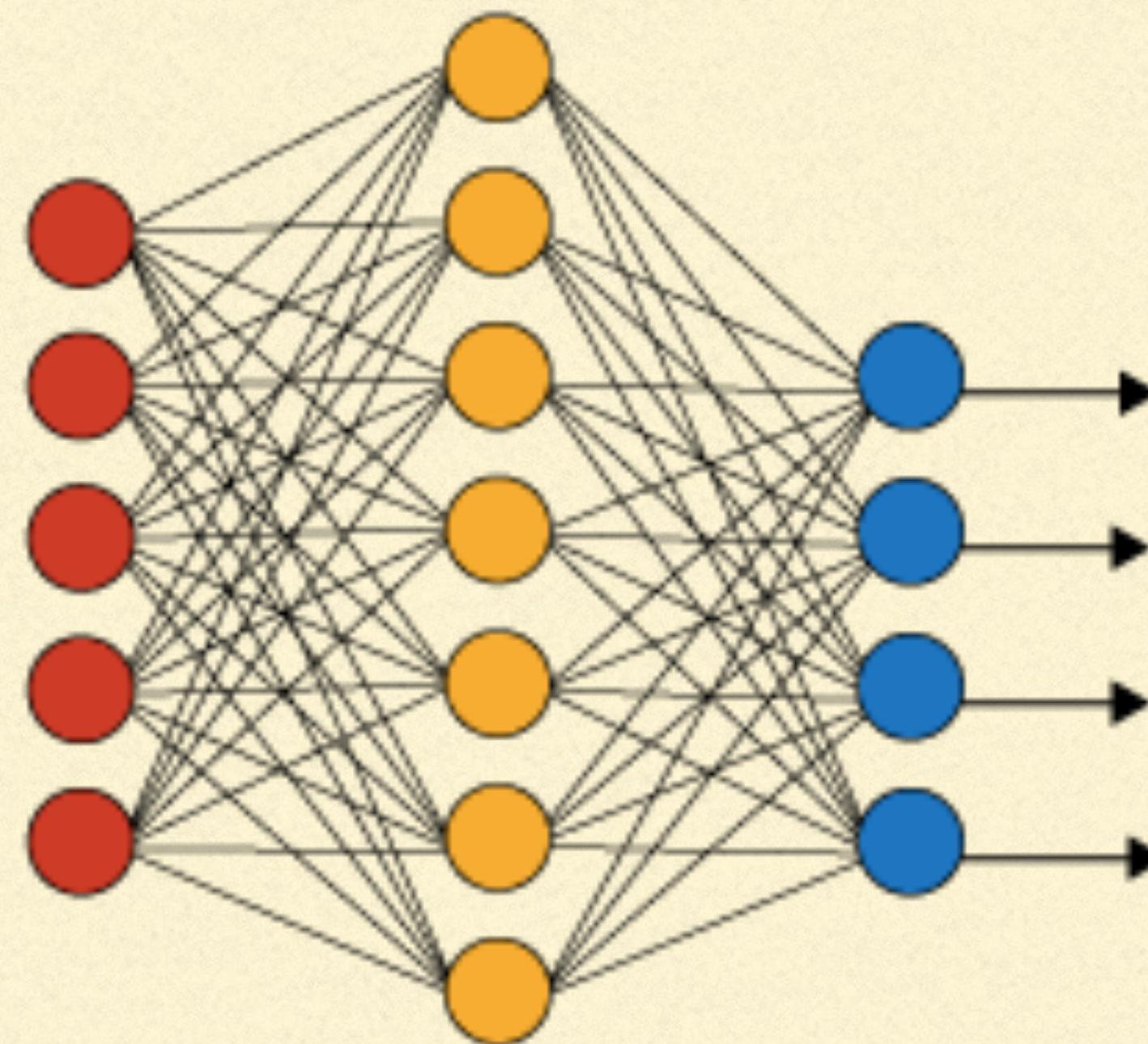
Computing systems inspired by the biological neural networks that constitute animal brains.

ML - ARTIFICIAL NEURAL NETWORK (ANN)



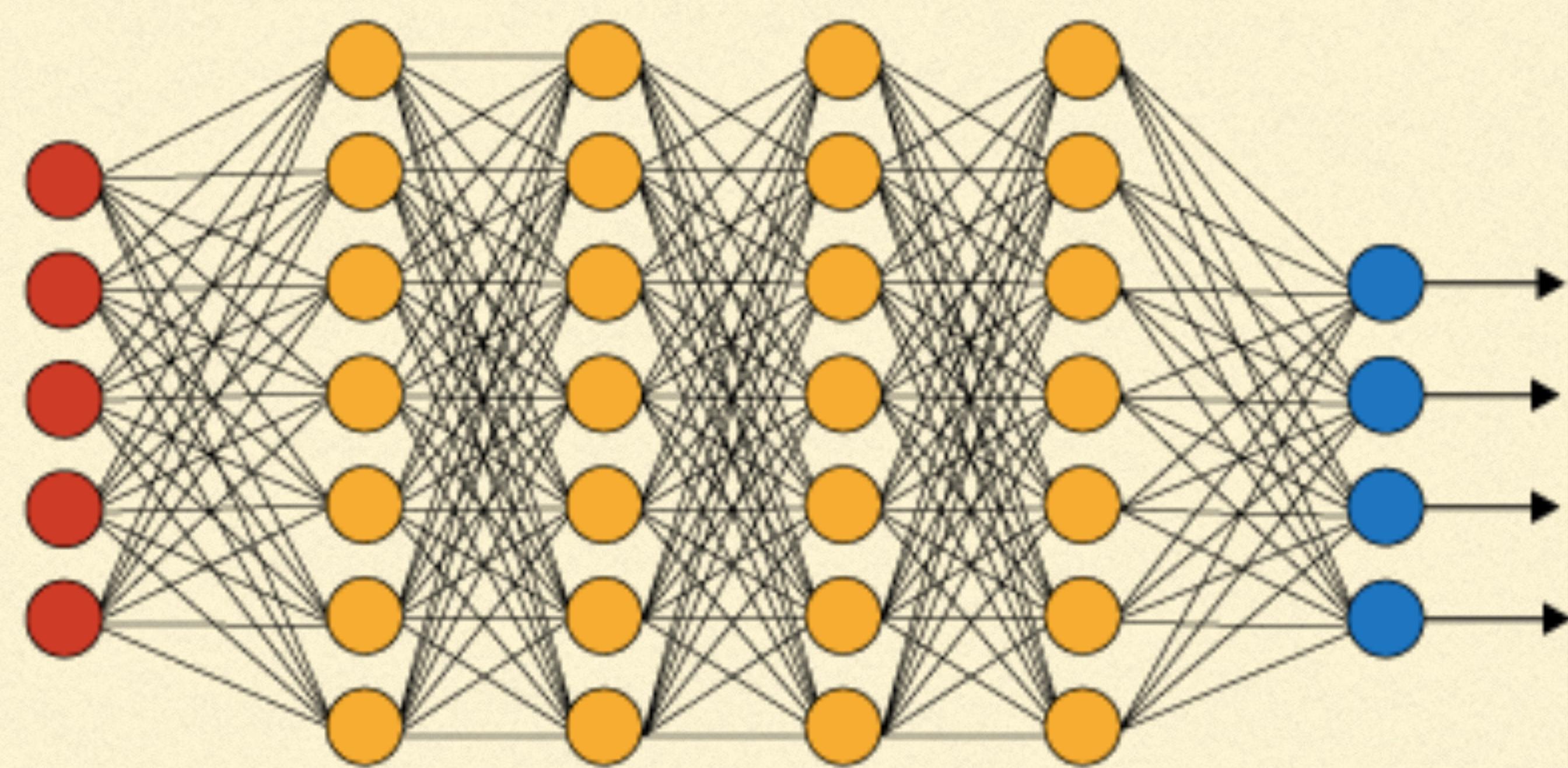
- Learn (progressively improve performance)
- To do tasks by considering examples
- Generally without task-specific programming
- Example: Based on image - cat or no cat?

Simple Neural Network



● Input Layer

Deep Learning Neural Network



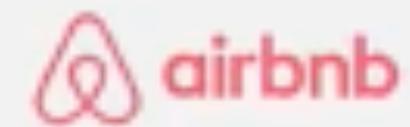
● Hidden Layer

● Output Layer

ML - DEEP LEARNING?

Multiple layers of neurons

Prices



A screenshot of the Airbnb mobile application interface. At the top, there's a navigation bar with icons for messaging, reviews, and account settings. Below it, a header shows the date as May 2015. The main content area displays a graph and a table showing price history over time. A green progress bar indicates a price drop of 15% compared to the previous week. The table includes columns for 'Price' and 'Weeks Ago'. At the bottom, there are buttons for 'Done' and 'See cheaper'.

BuzzFeed

Food



31 Delicious Things To Cook In May

What Should You Make For Dinner Tonight?
Don't worry, we're here to help.
A collection by Buzzfeed Food

10 Classic Movie Movies That Are Totally Improvised

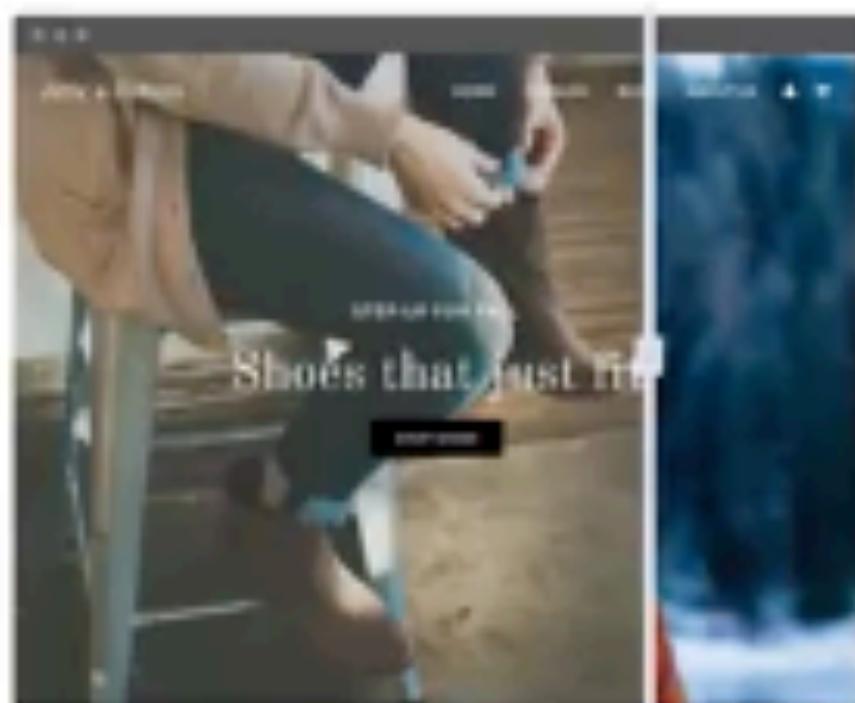
Alums are renowned for their improvisational skills, but it's their improvisational skills that they make them brilliant. May, 2015 — you got that?

Sprout

Fry Up These Absolutely Delicious Eggplant Parmesan Bites Today
Get the hyper recipe!
A collection by Buzzfeed Food

Headlines

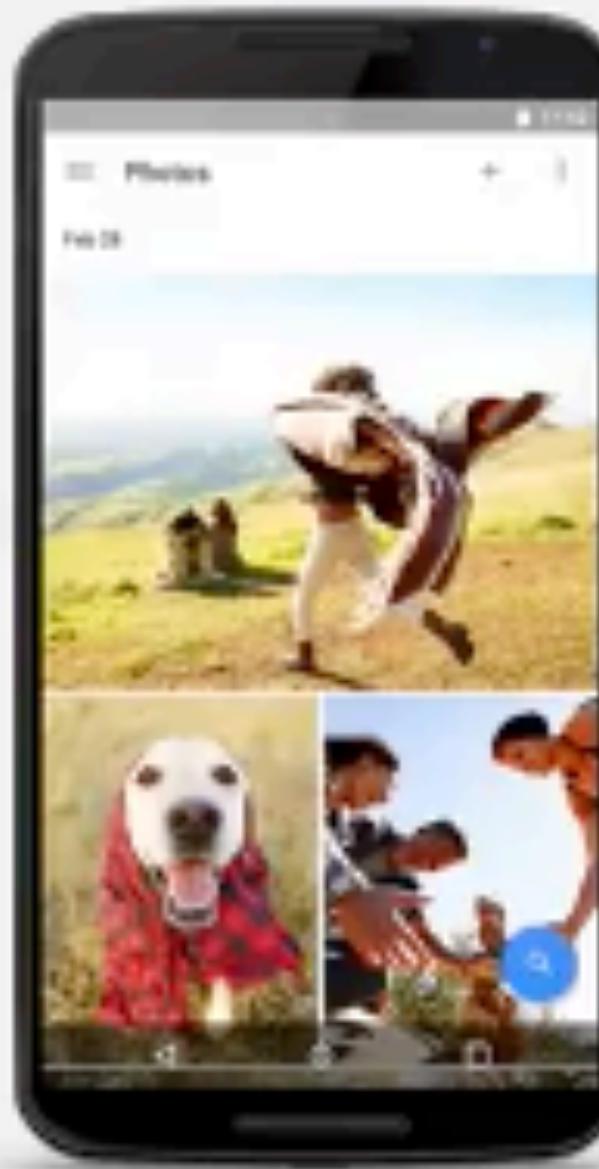
optimizely



WHO IS USING?

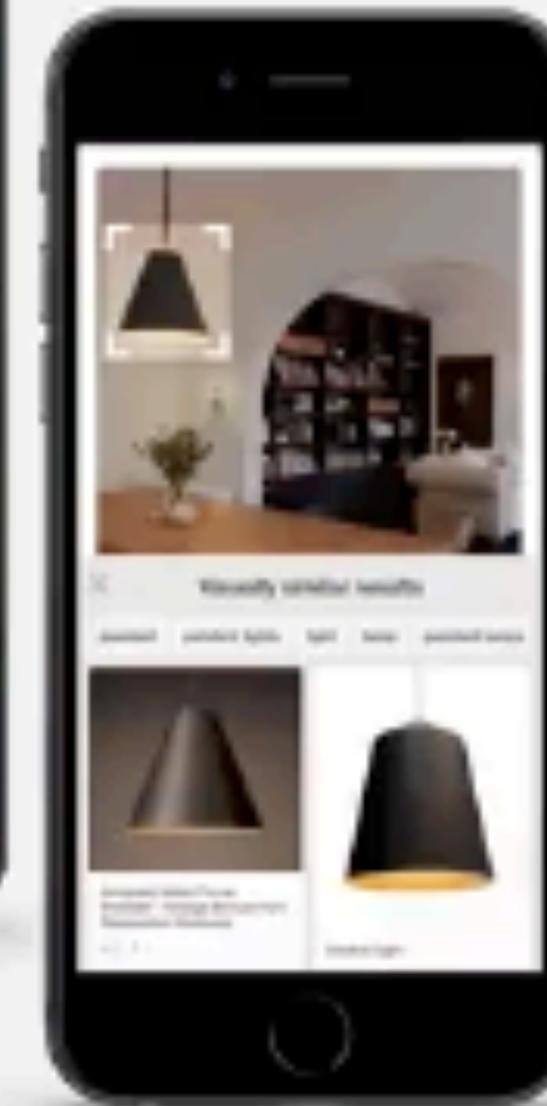
Almost all.

Google



Similar Images

Pinterest

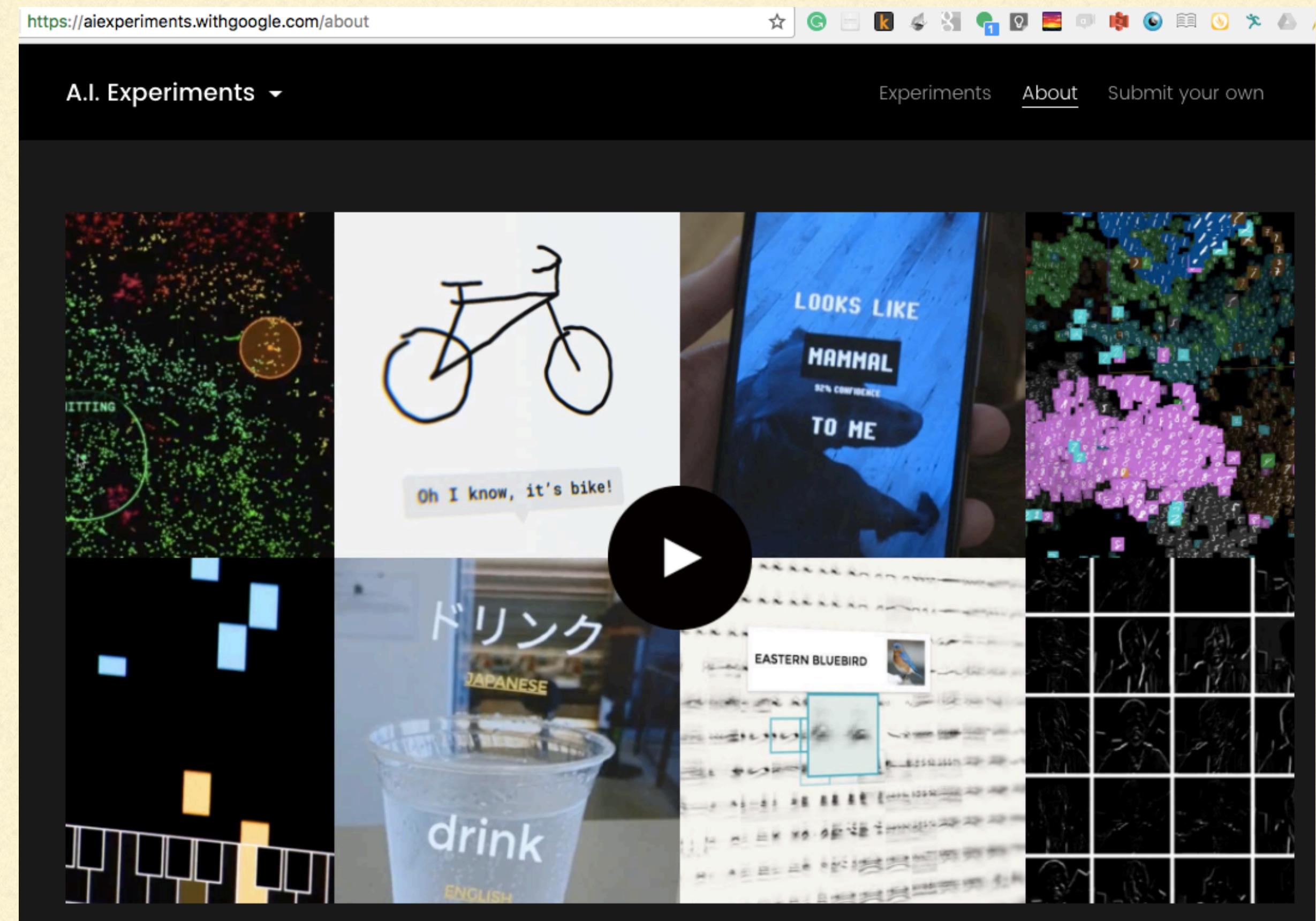
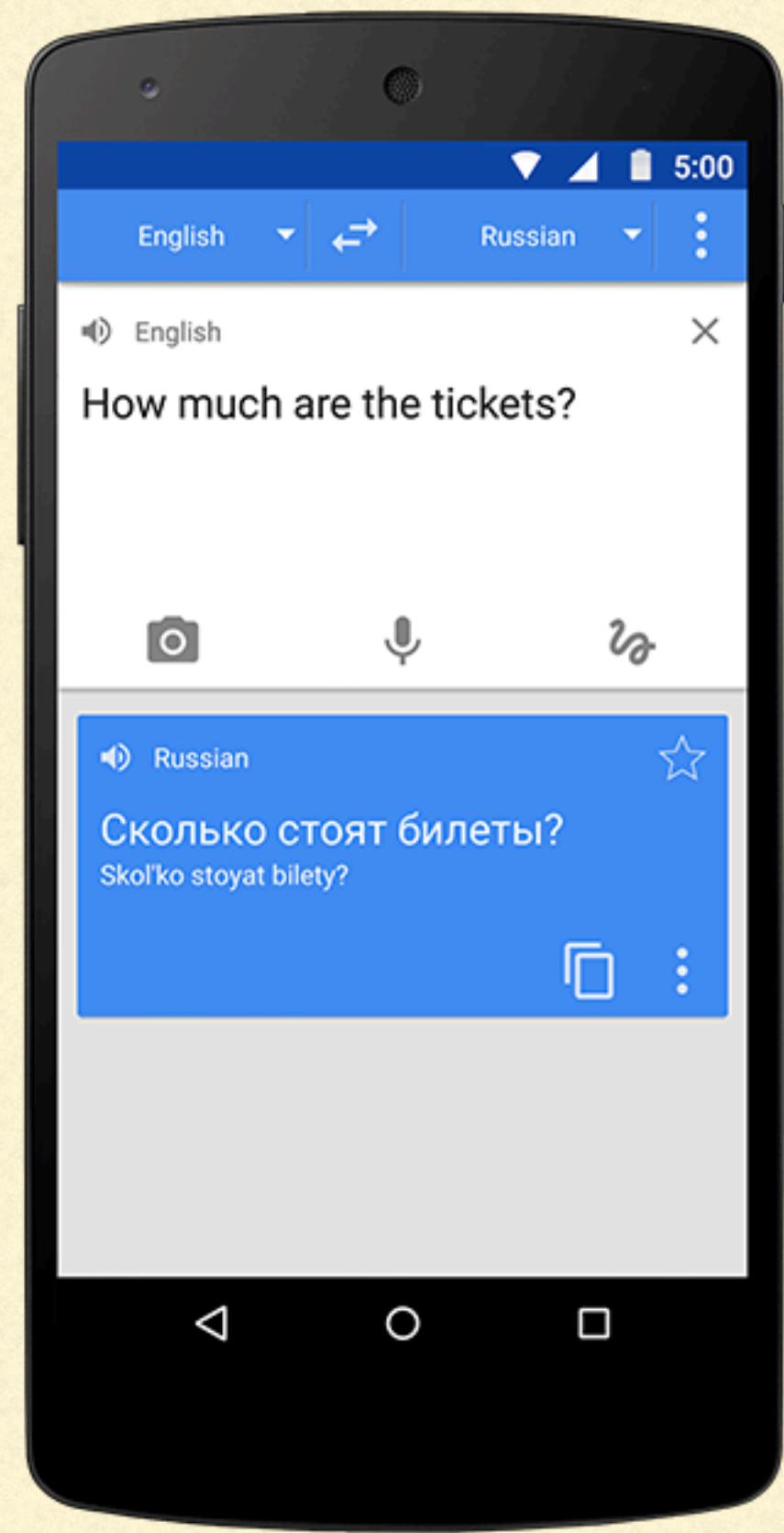


facebook.



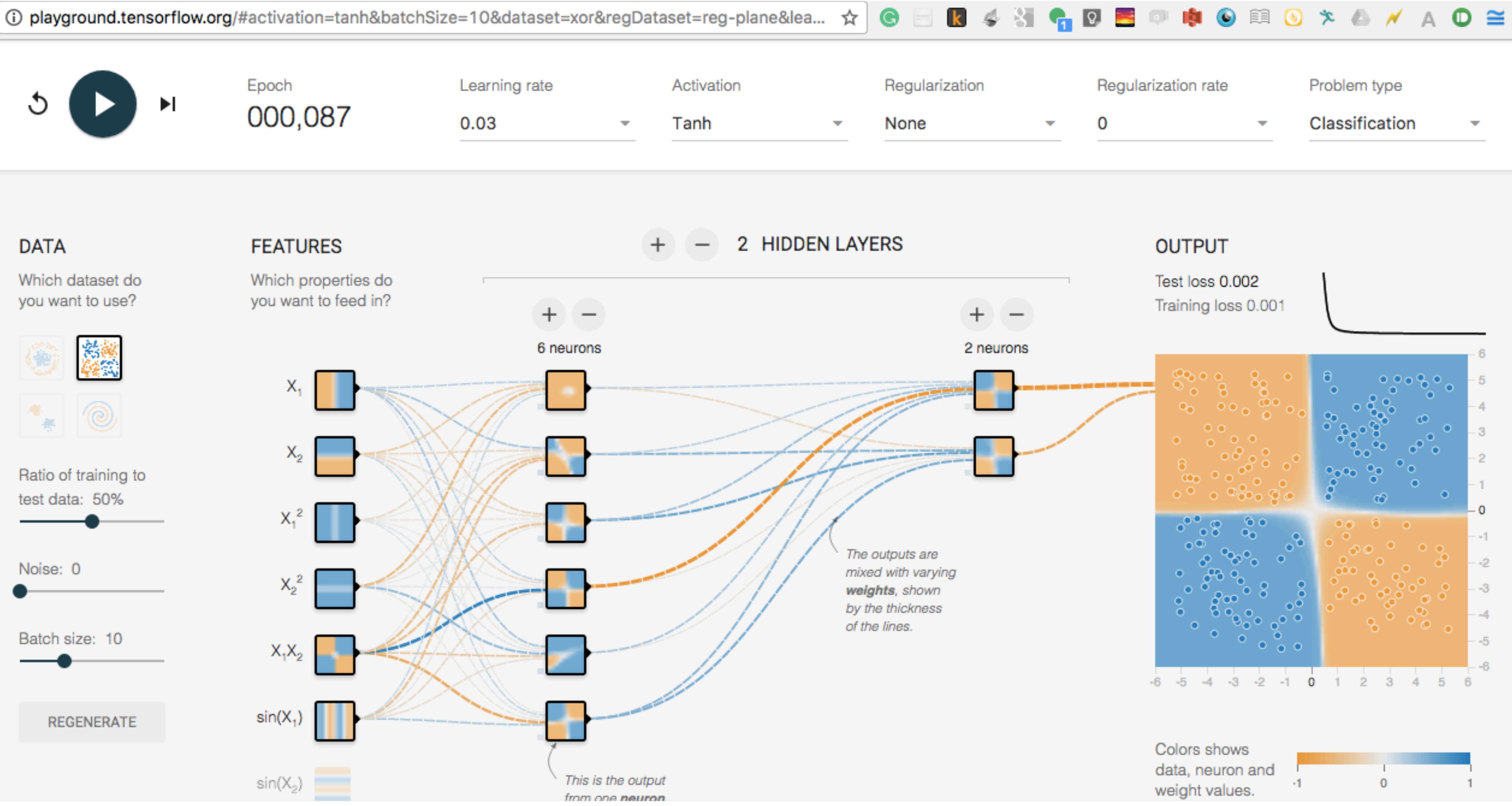
Auto Reply

Face Recognition



GOOGLE TRANSLATE & AUTO DRAW

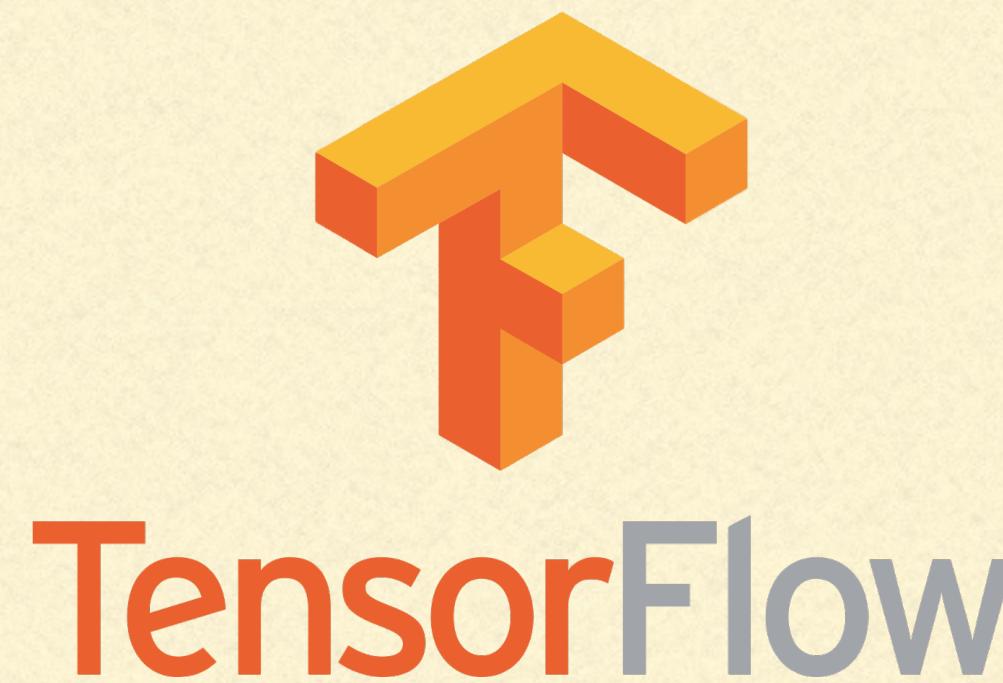
More usecases: <https://aiexperiments.withgoogle.com/>



DEMO

<http://playground.tensorflow.org/>

DEEP LEARNING FRAME WORKS

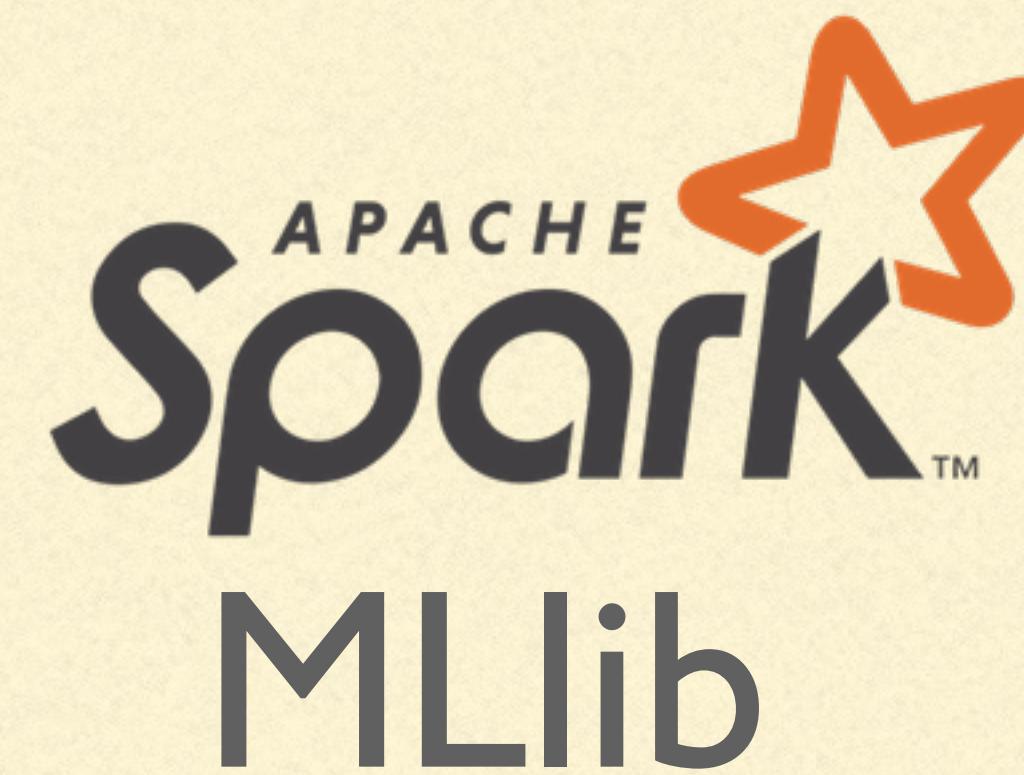


theano  Keras

Caffe



MACHINE LEARNING FRAME WORKS



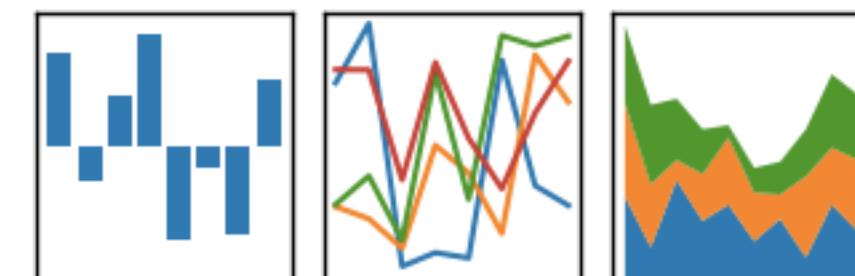
Distributed



Simple

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

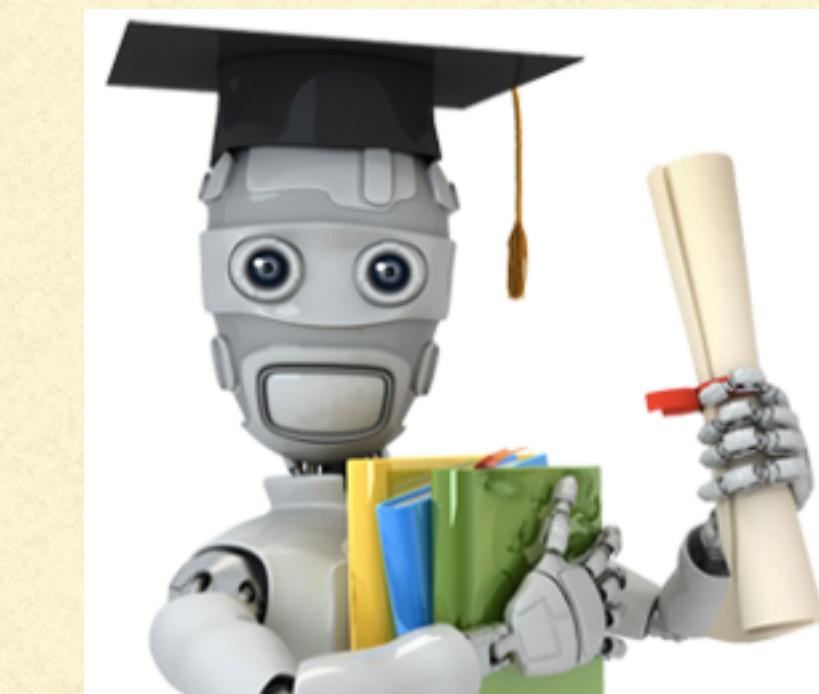


LEARN MORE ?

Machine Learning Courses



Machine Learning with python



coursera

Thank You!