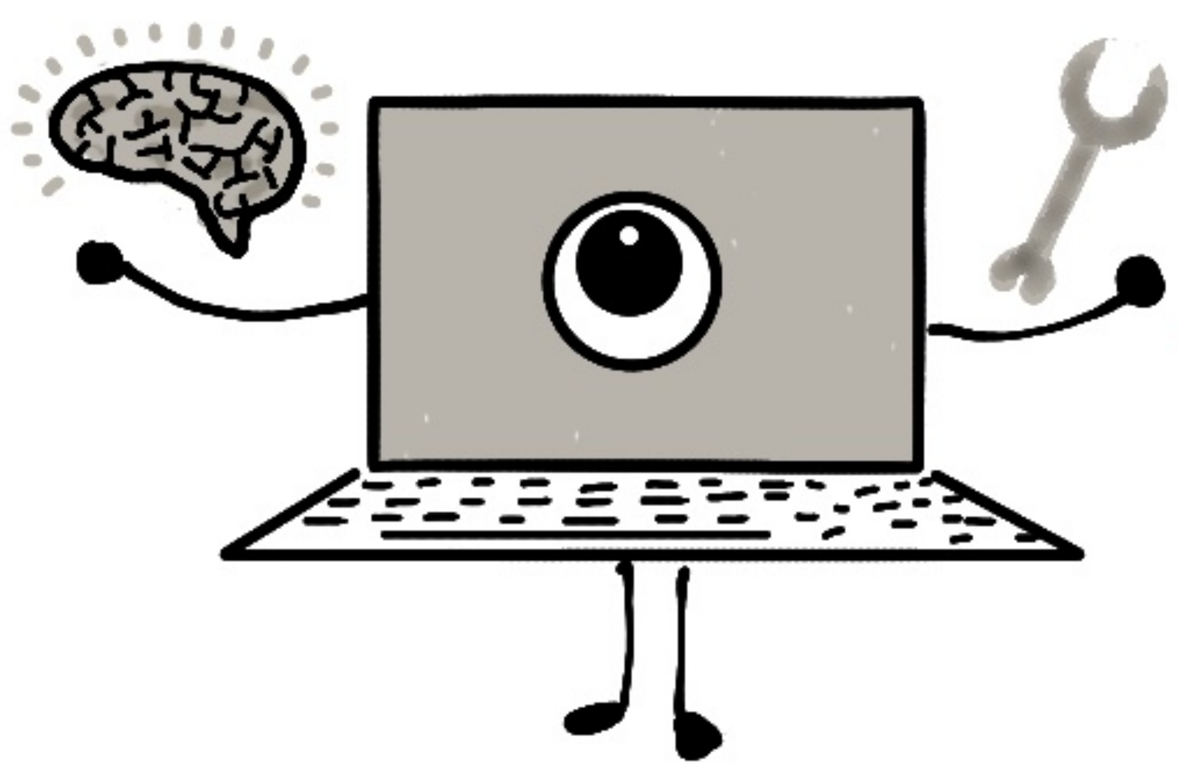
CREATING ARTIFICIAL INTELLIGENCE



By

Gitanjali Venkatraman

INTRODUCTION

WE ARE ALREADY AWARE THAT COMPUTERS TODAY ARE SMART, HELPFUL AND ABLE TO COME UP WITH ANSWERS TO A LOT OF OUR QUESTIONS

THIS BOOK FOCUSES ON THE PUZZLE OF TRYING TO DEFINE INTELLIGENCE. THE INTELLIGENCE OF MACHINES AND THE BRAIN CREATING THEM.

THE HUMAN BRAIN IS EXTREMELY CAPABLE, KIND AND IMAGINATIVE, BUT NOT WITHOUT LIMITATION. MACHINES HAVE NO VALUES OR EMPATHY, BUT CAN BE PROGRAMMED TO DO (ALMOST) ANYTHING.

IN OUR QUEST TO CREATE THINKING MACHINES TO ASSIST, EQUAL OR SURPASS US, WHAT ARE THE INGREDIENTS WE NEED?

LET US EXPLORE HOW THESE MACHINES CAME TO BE AND AND WHAT WE EXPECT OF THEM. HOW DO THEY LEARN TO BE INTELLIGENT? HOW DO WE USE THEM? SHOULD WE FEAR THEM? CAN WE TRUST THEM?

PERHAPS THEY ARE ONLY EVER AS GOOD AS THEIR MAKERS - PERHAPS NOT?

SCOPE

LEARNING MACHINES

- IMPRESSIVE ENCOUNTERS

HUMAN INTELLIGENCE

- DEFINITION/THEORIES

MACHINE INTELLIGENCE

- THE PROBLEM
- DEFINING AI & ITS GOALS

RECENT HISTORY

- 1950-S + DARTMOUTH CONFERENCE
- ENCODED KNOWLEDGE + PERCEPTRONS
- WHY AI WORKS SO WELL NOW

HOW MACHINES LEARN

- VARIOUS WAYS OF LOOKING AT HOW LEARNING HAPPENS

LEARNING ALGORITHMS

- HOW THEY ARE DIFFERENT
- WHAT A MODEL IS
- SOME EXAMPLES

WHY AI IS HARD

- MORAVEC, COMMON SENSE AND A BODY

THE HARDER QUESTIONS

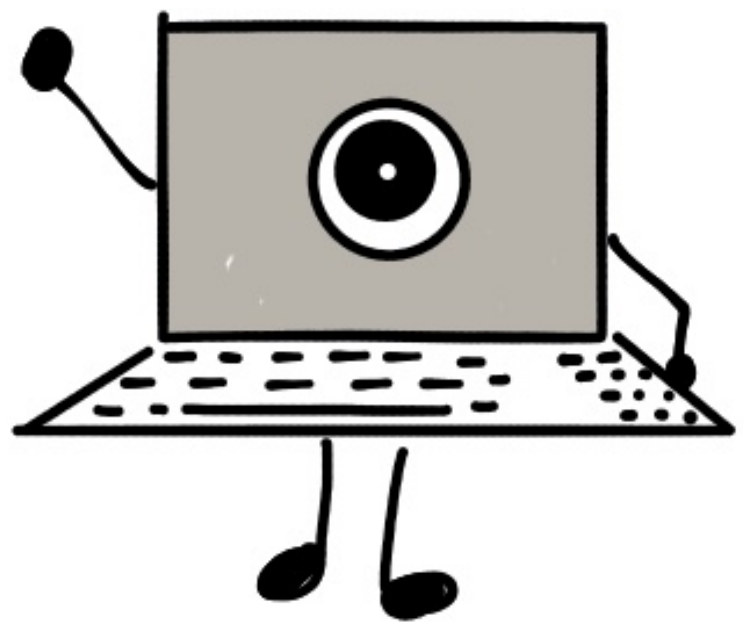
- COST, ETHICS, BLAME, BIAS AND HUMAN-NESS
- BUILDING TRUST

WOMEN IN AI

MORE TO EXPLORE

REFERENCES

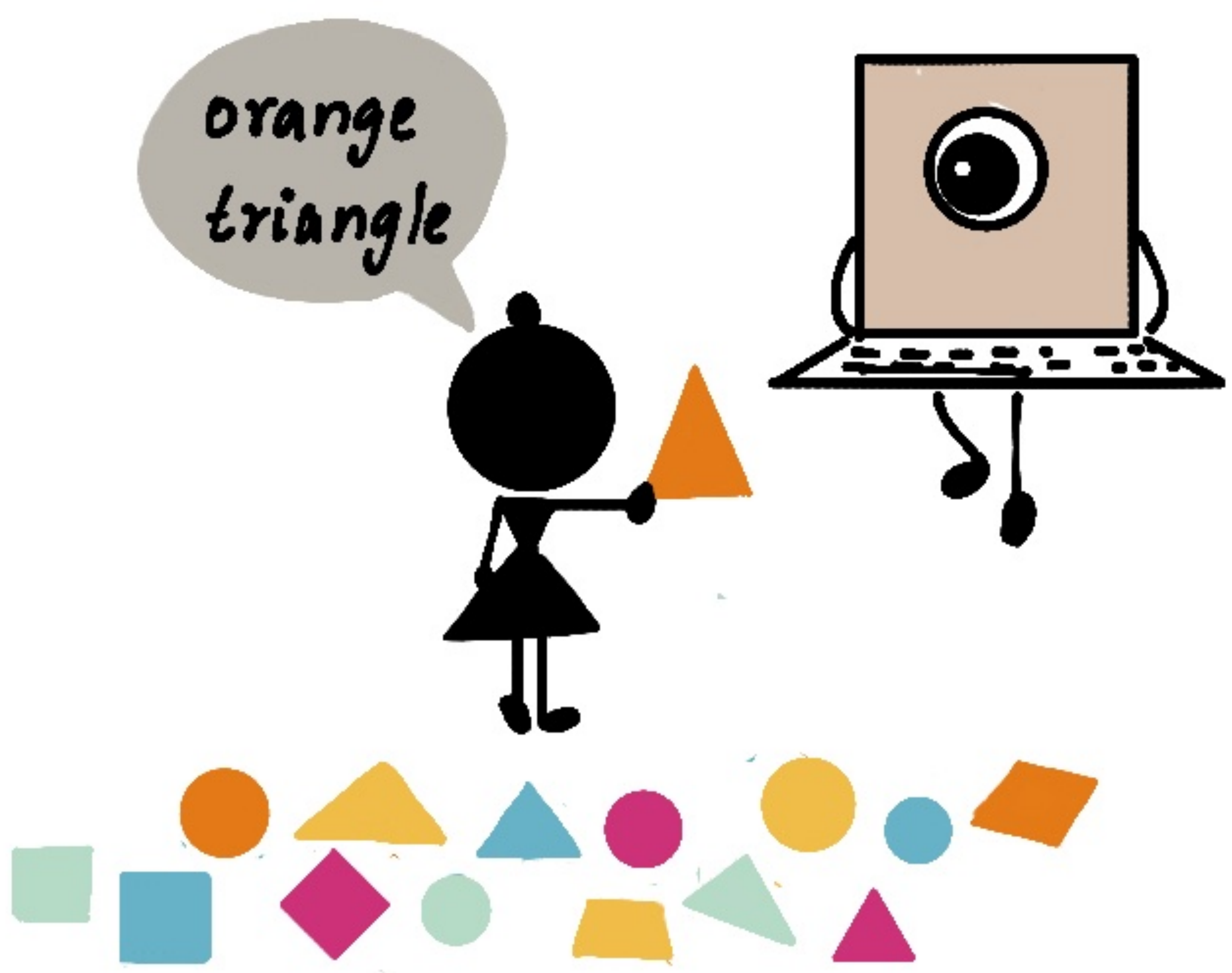
LEARNING MACHINES

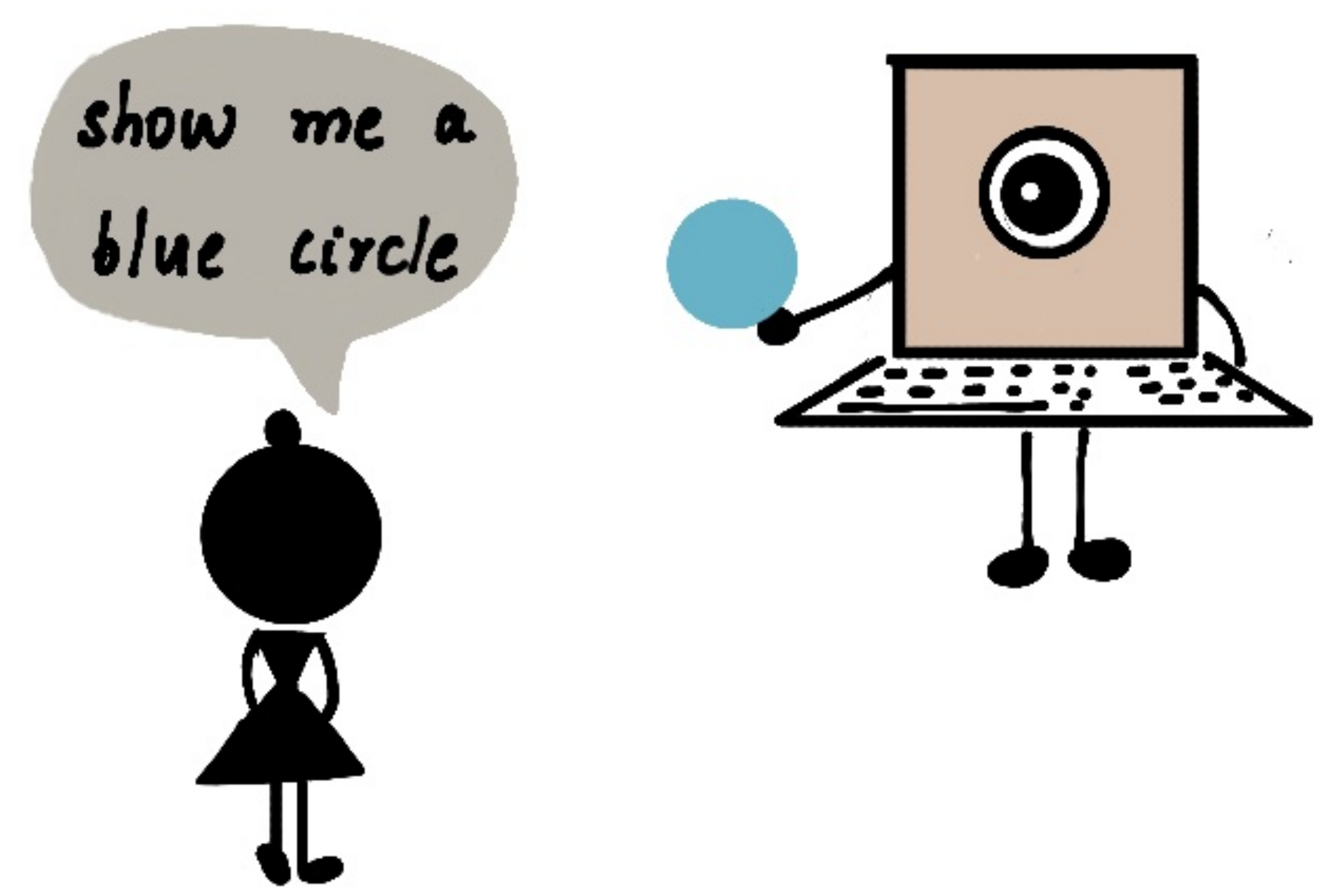


WE HAVE COME A LONG WAY SINCE ALAN TURING'S 1950 PAPER ON COMPUTER INTELLIGENCE.

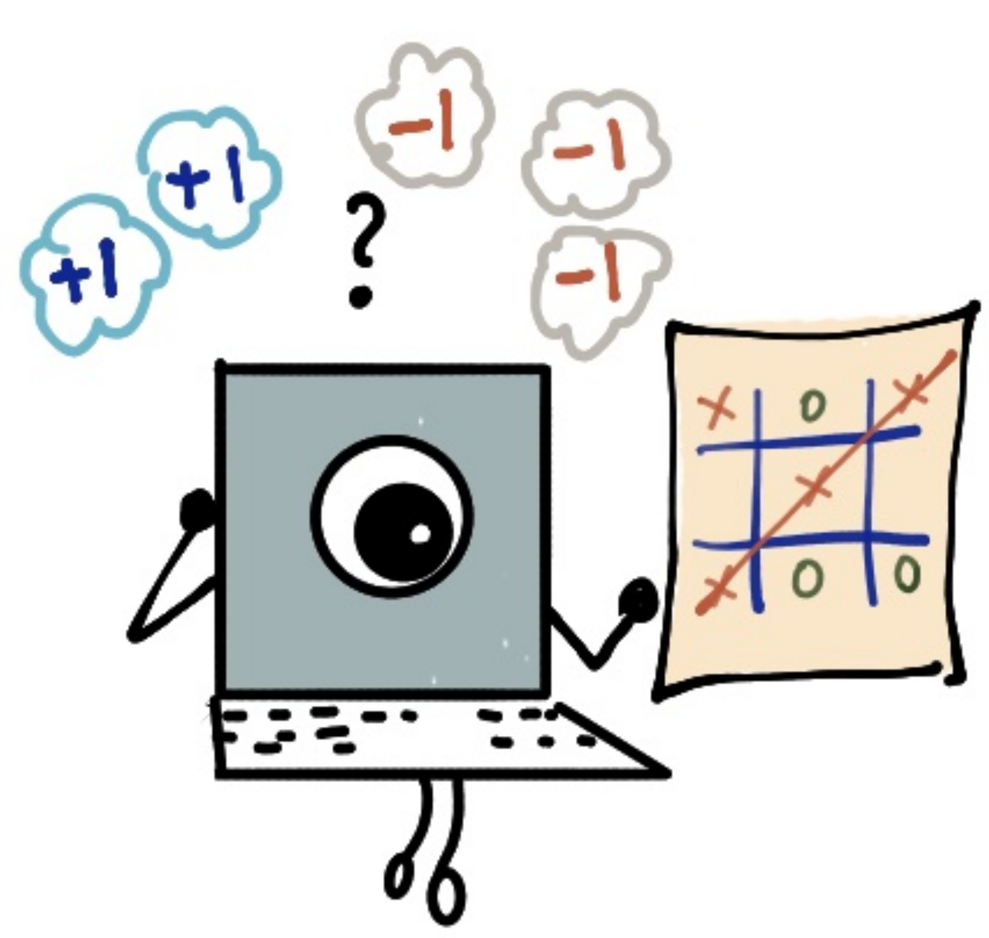
THERE ARE NOW MACHINES THAT CAN LEARN:

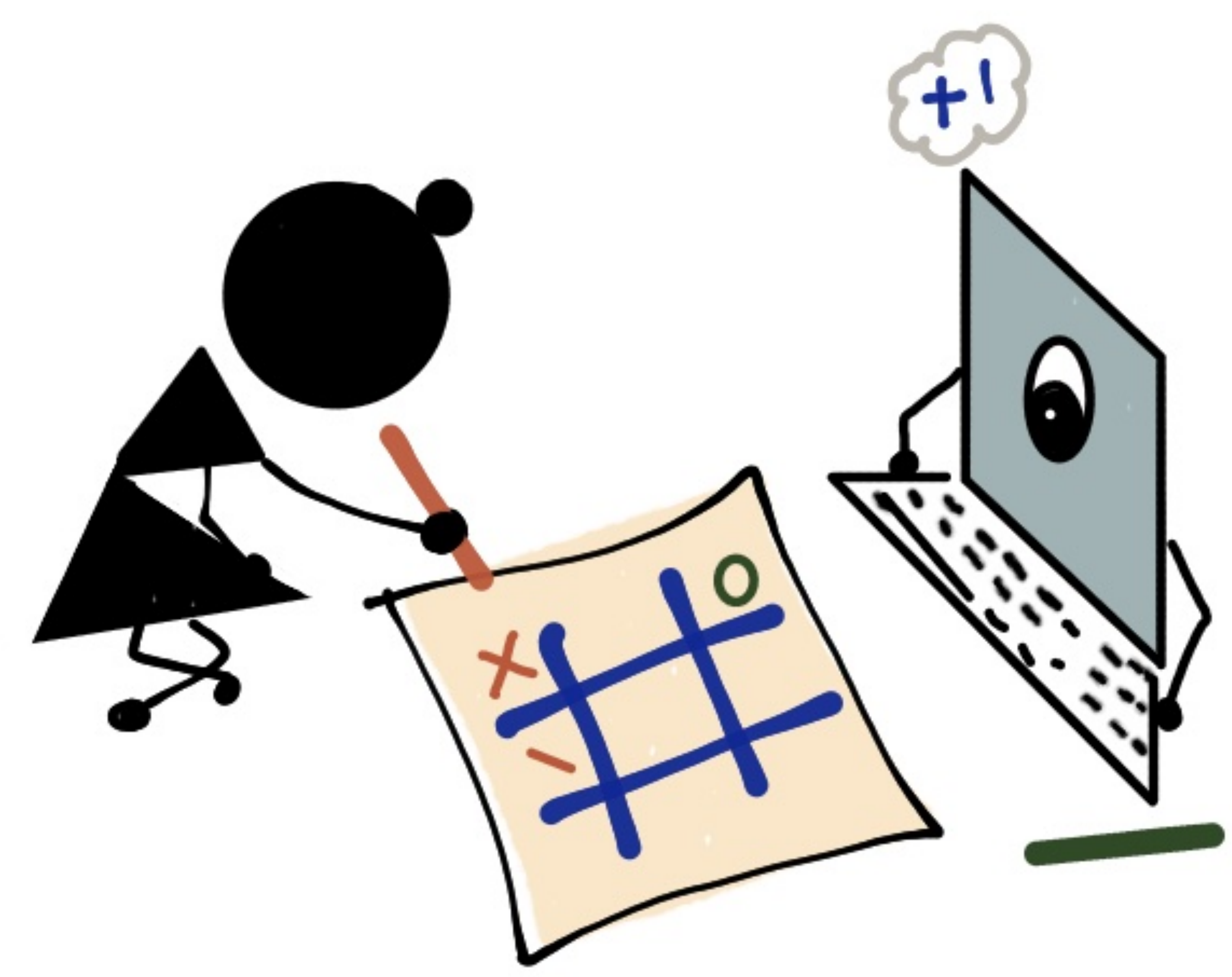
COMPUTERS CAN LEARN BY BEING SHOWN SEVERAL EXAMPLES.





OR THEY LEARN THROUGH REWARDS FOR ACHIEVING GOALS



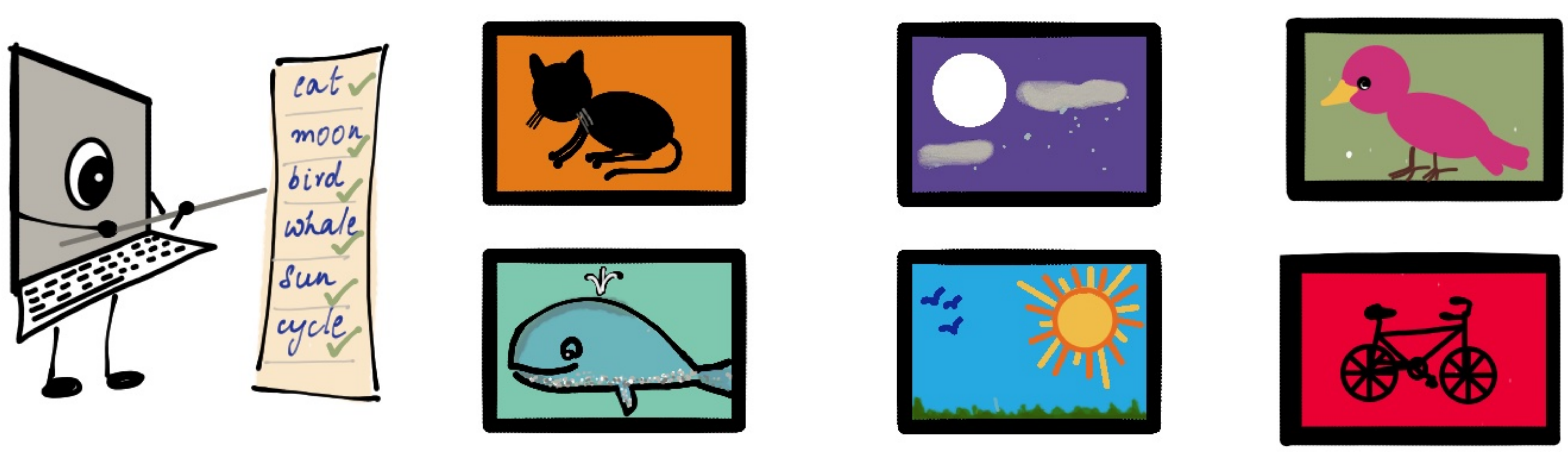


THE LEARNING MACHINE FIGURES OUT THE RULES THAT BUILD UP TO THE GOAL OR FORM THE PATTERN

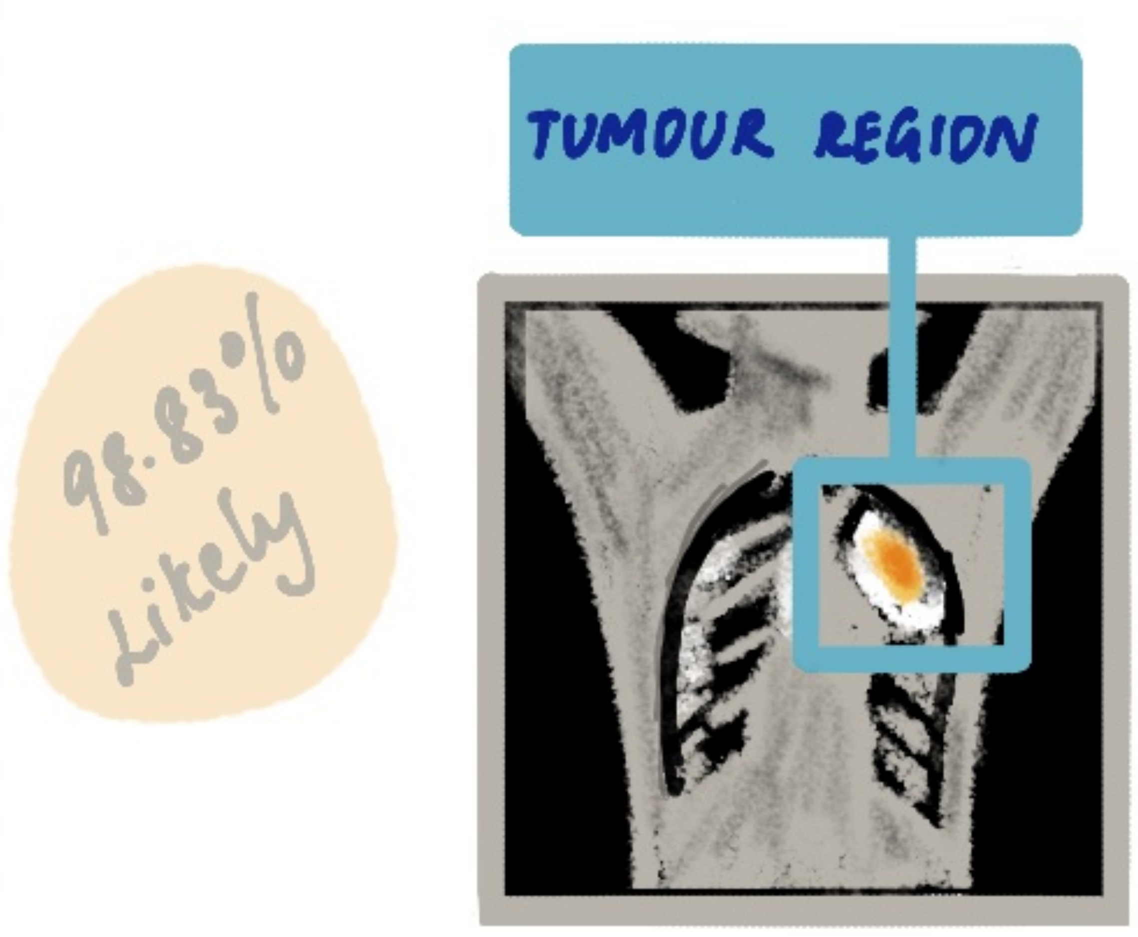
COMPUTERS CAN 'SEE'

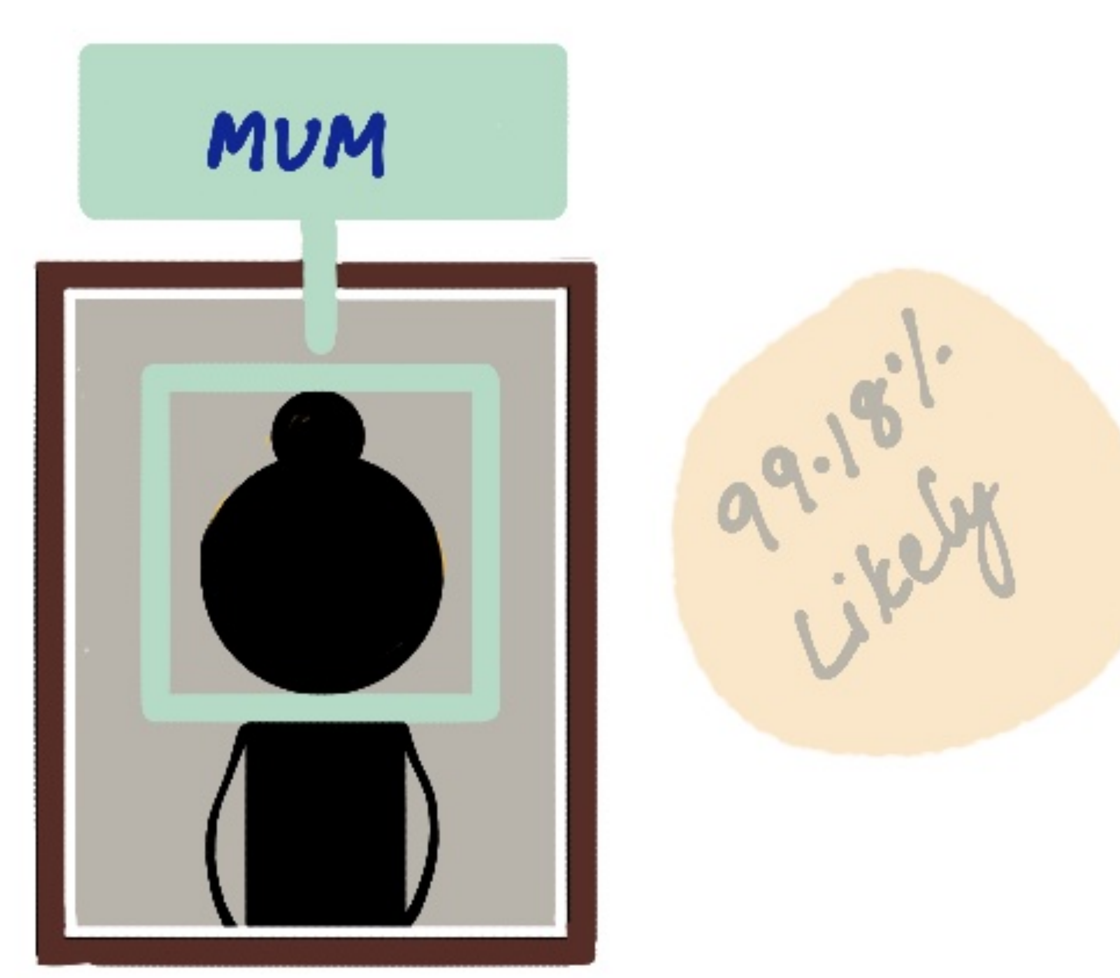
HAVING BEEN TRAINED ON SEVERAL MILLION IMAGES (see Stanford Imagenet)

THEY CAN IDENTIFY OBJECTS WITH A GOOD LEVEL OF ACCURACY



THEY CAN DISCOVER PATTERNS IN DATA, THUS ENABLING THEM TO





REASONABLY PREDICT FROM A SCAN, IF THERE IS DISEASE...

...RECOGNISE FACES AND TRACK MOVEMENTS

THIS ABILITY IS CALLED COMPUTER VISION

COMPUTERS CAN 'HEAR'

WITH ENOUGH TRAINING DATA*, PROGRAMS CAN WRITE UP AUDIO AS TEXT

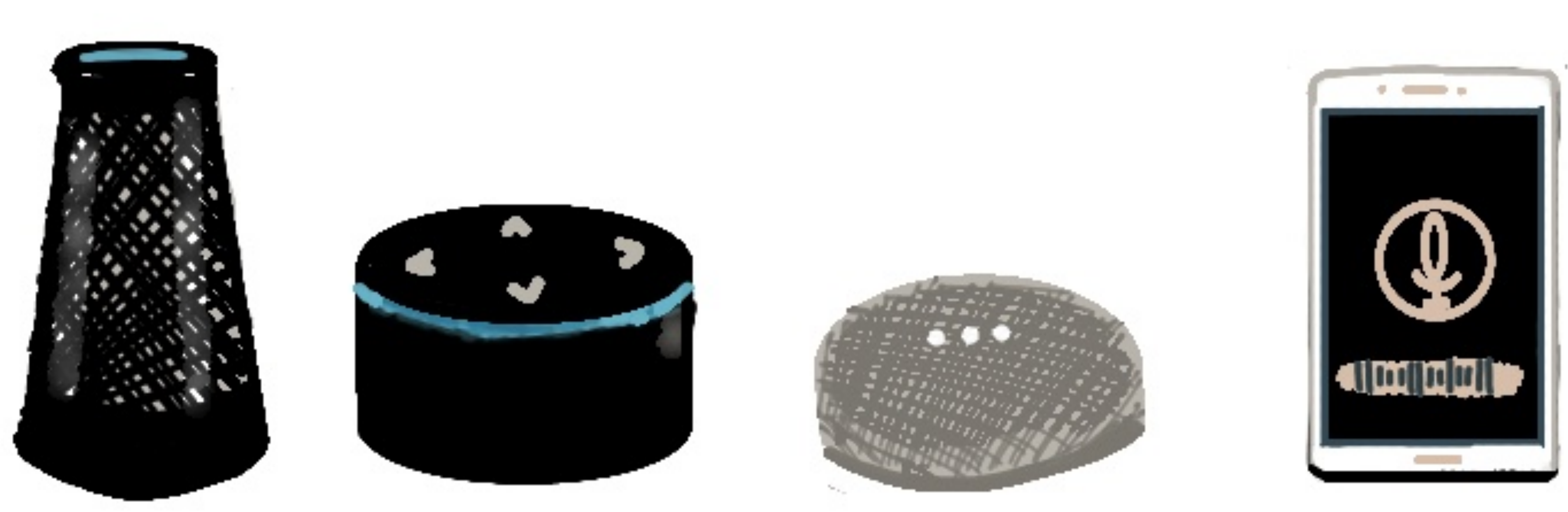
THIS ABILITY IS CALLED SPEECH RECOGNITION OR SPEECH TO TEXT.

* SUCH AS GOOGLE AUDIOSET, OPENSLR.ORG (Ted talks)

gera – nigeria
garhwali – India
gascon – SW france
guajiro – S. america
gallo – france
gagauz – Turkic
gallurese – Italy

IT CAN EVEN BE USED TO SAVE ENDANGERED LANGUAGES

Seneca Indian language / RIT



DEVICES RESPOND TO VOICE COMMANDS AND RECOGNISE INDIVIDUAL VOICES

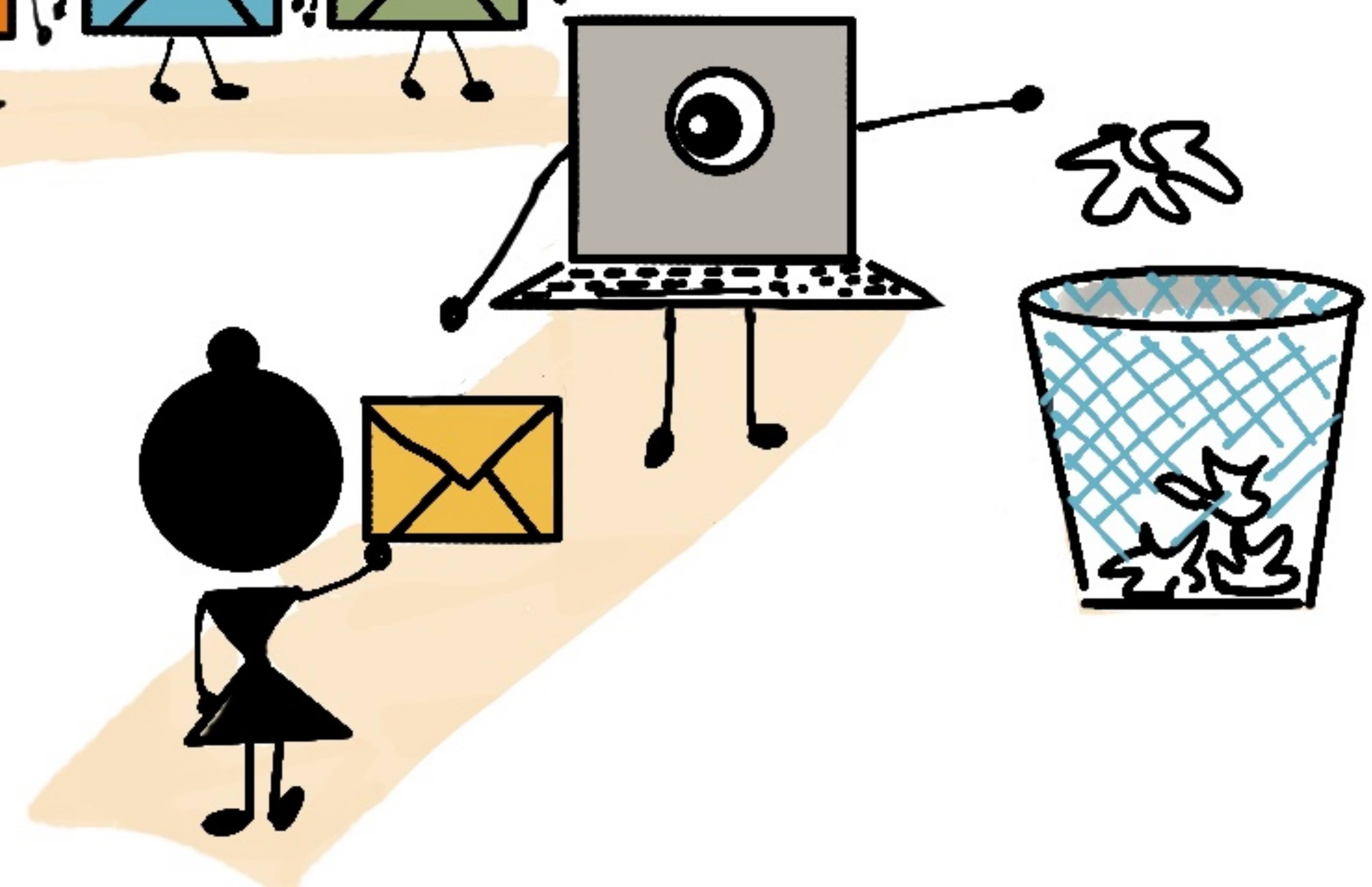
THIS IS CALLED VOICE RECOGNITION OR VOICE IDENTIFICATION

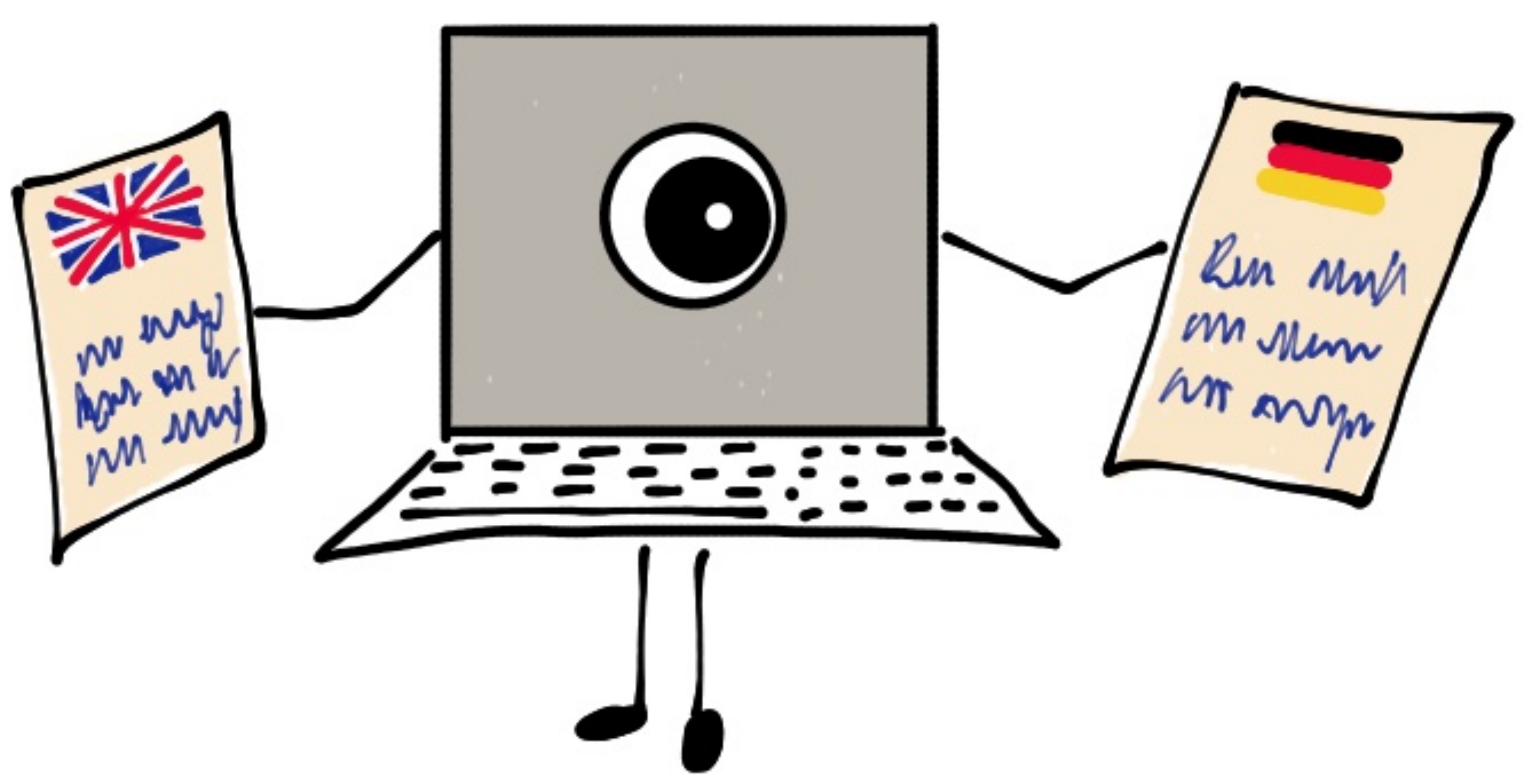
TEXT TO SPEECH (SPEECH SYNTHESIS) ALSO MAKES IT EASIER FOR PEOPLE OF ALL ABILITIES TO INTERACT WITH MACHINES

Here is how to make your own home made slime

COMPUTERS CAN 'UNDERSTAND'

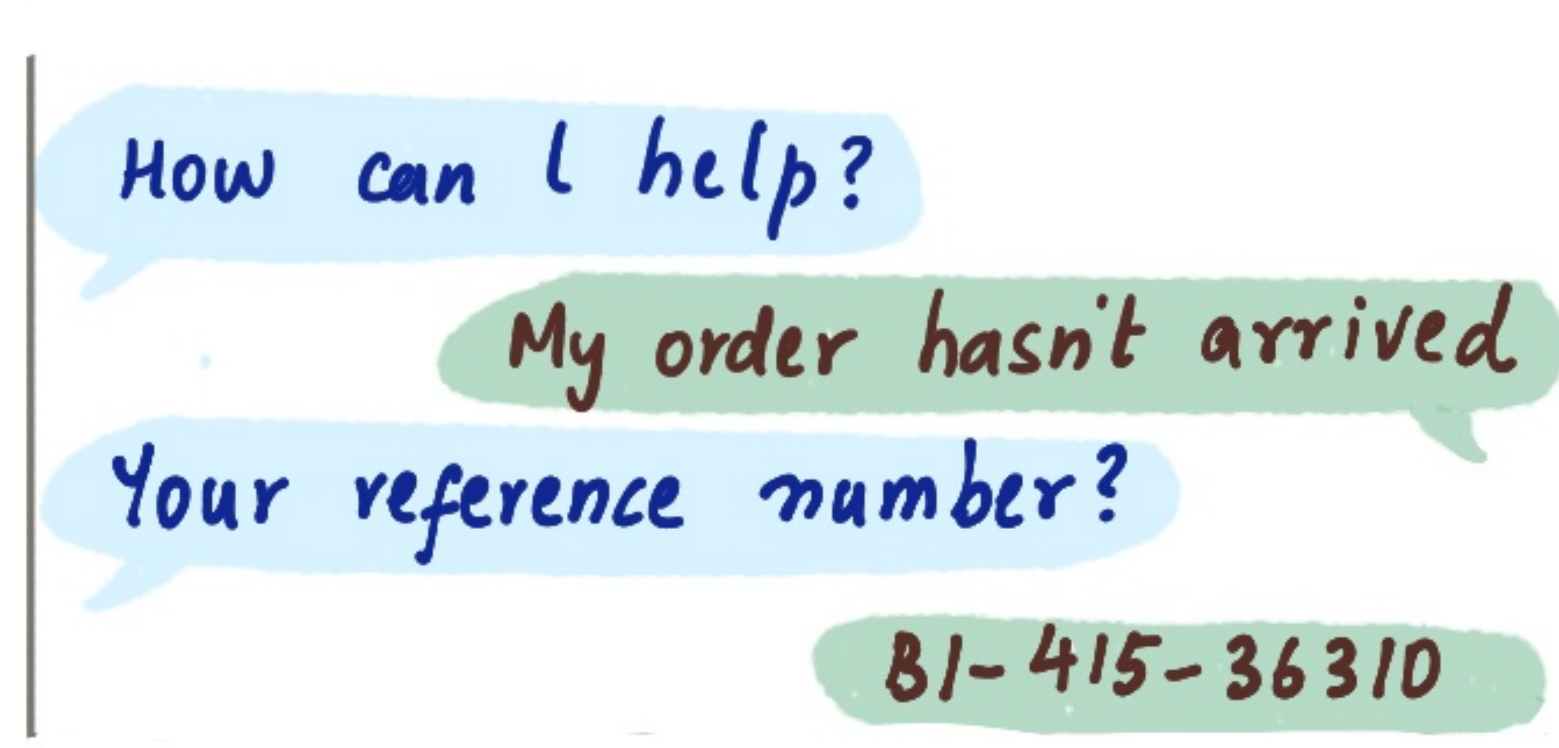
WE ARE FAMILIAR WITH SPAM FILTERS





PROGRAMS THAT TRANSLATE LANGUAGES

CHAT BOTS THAT HELP OUT WITH VARIOUS TASKS



INTERNET SEARCH ENGINES THAT RECOGNISE A FAMOUS PERSON/PLACE

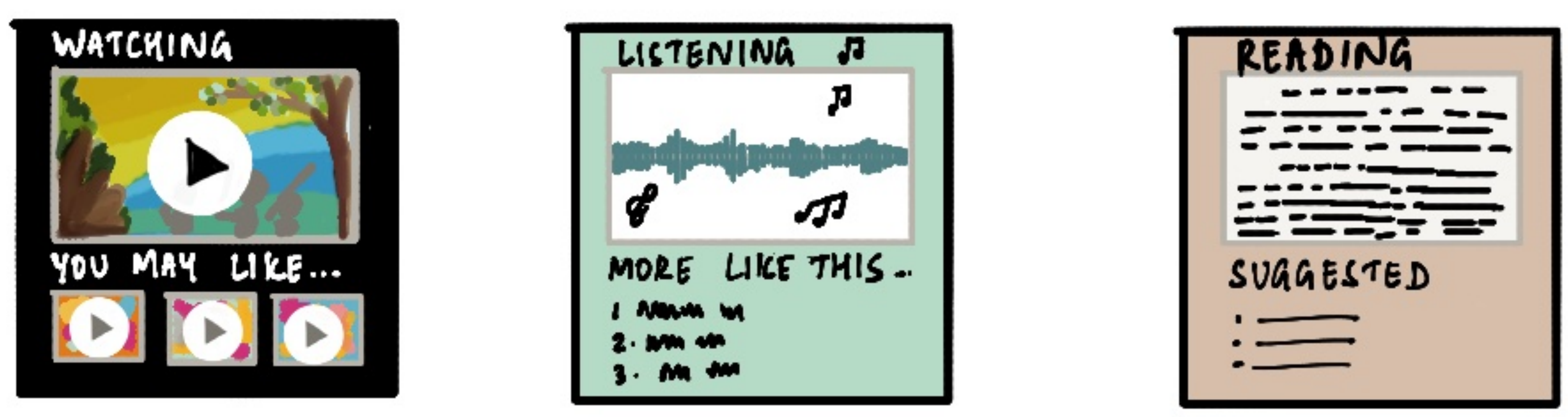
AND HELP WITH SPELLING GRAMMAR CHECKS, TEXT PREDICTION ETC.

THIS IS CALLED NATURAL LANGUAGE PROCESSING OR NLP

COMPUTERS CAN 'DO'

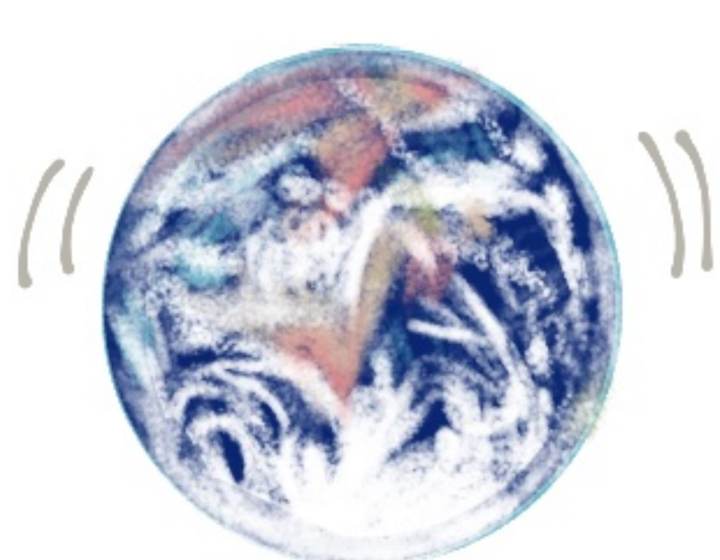
IN ADDITION TO CLEANING FLOORS AND DRIVING CARS/PLANES/DRONES, THEY CAN SUGGEST THE NEXT PRODUCT TO BUY

AND RECOMMEND

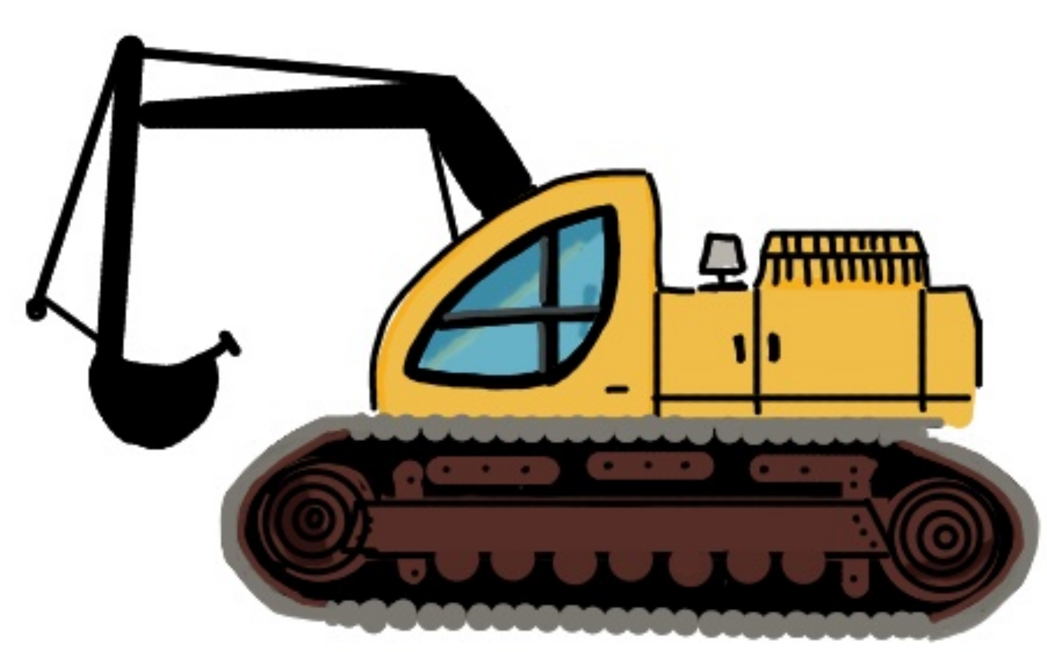


MOVIES MUSIC ARTICLES AND MORE

THEY CAN SENSE EARTHQUAKES



PREDICT FAULTS IN HEAVY MACHINERY



AND DETECT FINANCIAL FRAUDS

Portrait of Edmond Belamy

IT TURNS OUT THAT MACHINES ARE ALSO CAPABLE OF GENERATING ART.

AND MUSIC.

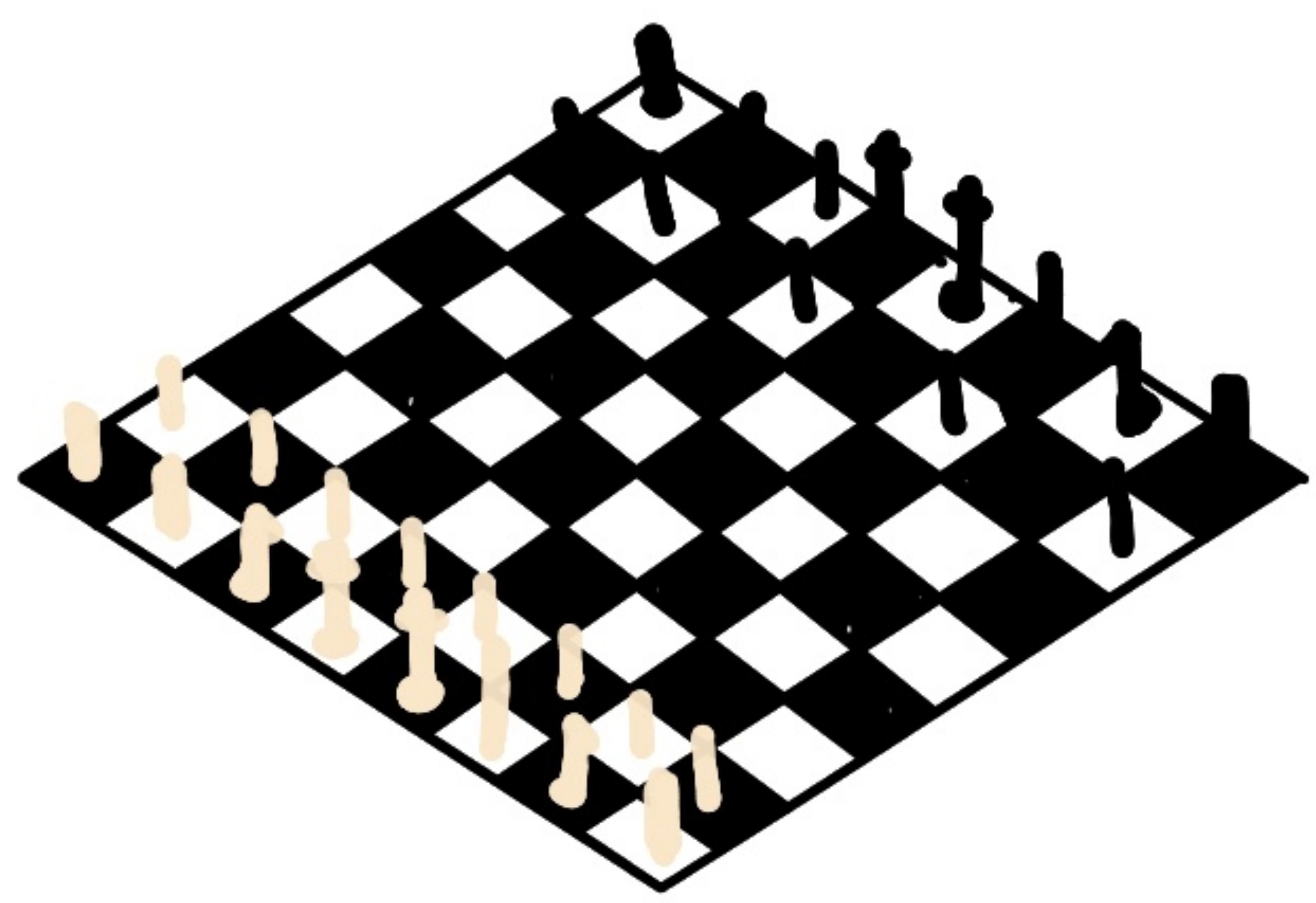
Google "Obvious AI Art"

IMPRESSIVE

ARGUABLY, THESE ATTRIBUTES MAKE A COMPUTER SEEM CLEVERER THAN THE AVERAGE MACHINE.

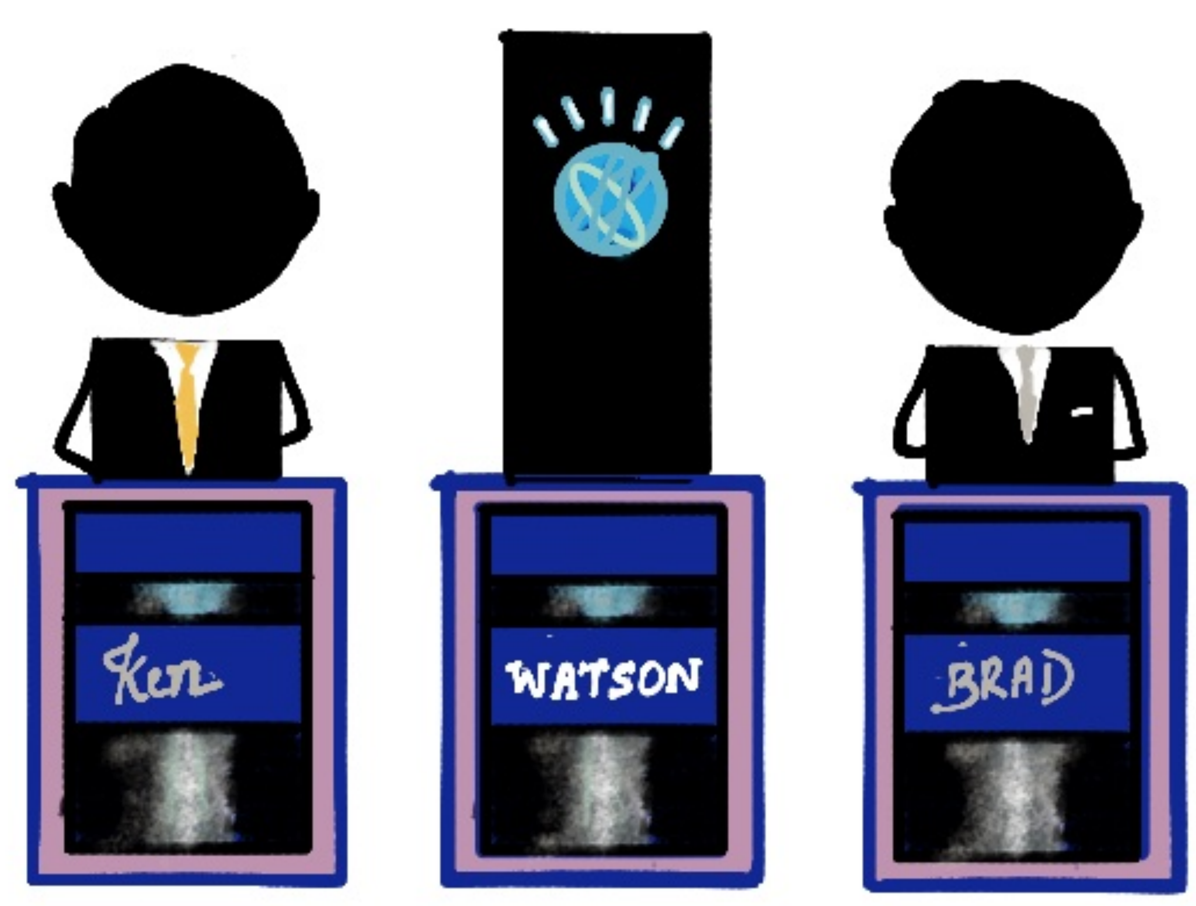
TOASTER

A QUICK LOOK INTO THEIR PAST REVEALS THAT COMPUTERS HAVE EXCELLED IN SKILLS THAT ONLY THE BRIGHTEST HUMAN MINDS POSSESSED



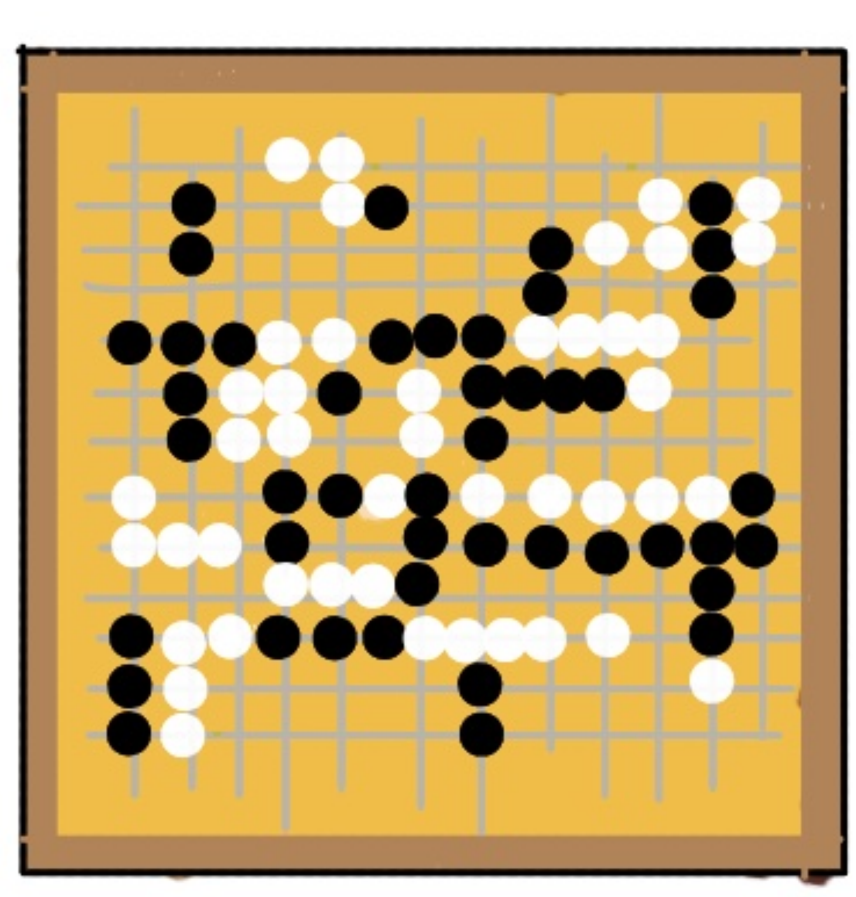
CHESS - 1996

IBM DEEP BLUE BEAT GARY KASPAROV



JEOPARDY QUIZ SHOW 2011

IBM WATSON BEAT KEN JENNINGS & BRAD RUTTER



GO 2016

DEEPMIND ALPHA GO BEAT LEE SEDOL

THE WORD 'INTELLIGENT' IS USED AS AN ADJECTIVE FOR SUCH MACHINES ARE THEY INTELLIGENT? WHAT IS INTELLIGENCE?

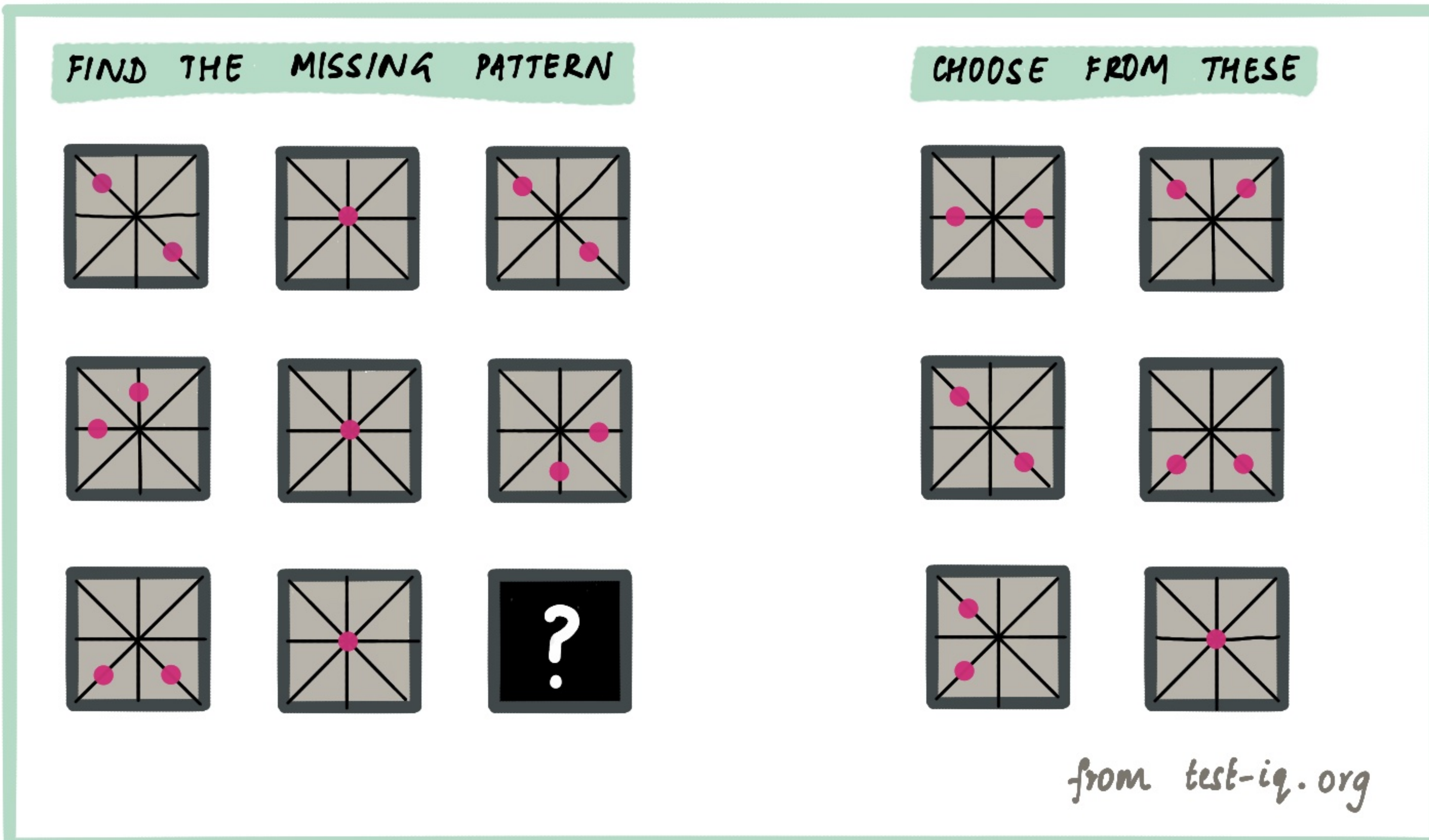
WHAT IS INTELLIGENCE?

MEASURING IT DOWN TO A NUMBER HAD UNFORTUNATE CONSEQUENCES FOR HUMANITY WITH THE DEVELOPMENT OF EUGENICS/EVIL ASSOCIATIONS.

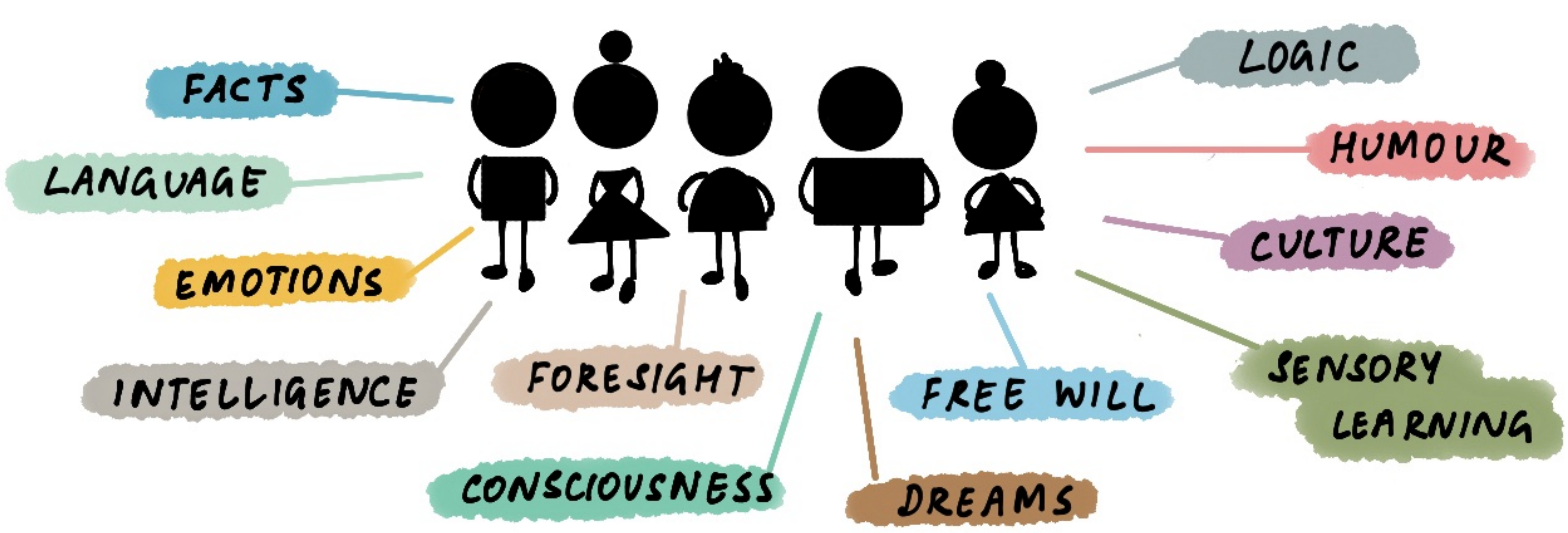
THIS SECTION DOES NOT TRAVEL DOWN THE DARK PATH. WE WILL TREAT IT AS A FASCINATING QUESTION LEADING TO GAINFUL INTROSPECTION.

BUT FIRST, A TEST

THIS IS A COMMON FORMAT FOR AN "INTELLIGENCE TEST"



FOR VARIOUS REASONS: HISTORICAL, ETHICAL & CULTURAL, THESE TESTS HAVE BEEN CONTROVERSIAL. IMPORTANTLY, THEY DO NOT EFFECTIVELY REVEAL ANY TRUE INTELLIGENCE.



SURELY, HUMANS ARE A BIT MORE COMPLEX AND CANNOT BE REDUCED TO ONE MEASUREMENT. WHAT THEN IS A SIGN OF INTELLIGENCE IN HUMANS?

IS IT LANGUAGE?

NOAM CHOMSKY

CHOMSKY, ARGUABLY THE TOP PHILOSOPHER AND LINGUIST THE WORLD HAS PRODUCED, SAYS THAT THE PURPOSE OF LANGUAGE ISN'T COMMUNICATION, BUT THAT

LANGUAGE EVOLVED AS A MODE OF CREATING AND INTERPRETING THOUGHT

THIS ABILITY SEPARATES HUMANS FROM OTHER CREATURES

LANGUAGE PLAYS TWO PARTS

INTERNAL

ACTIVATES PERCEPTUAL SYSTEM

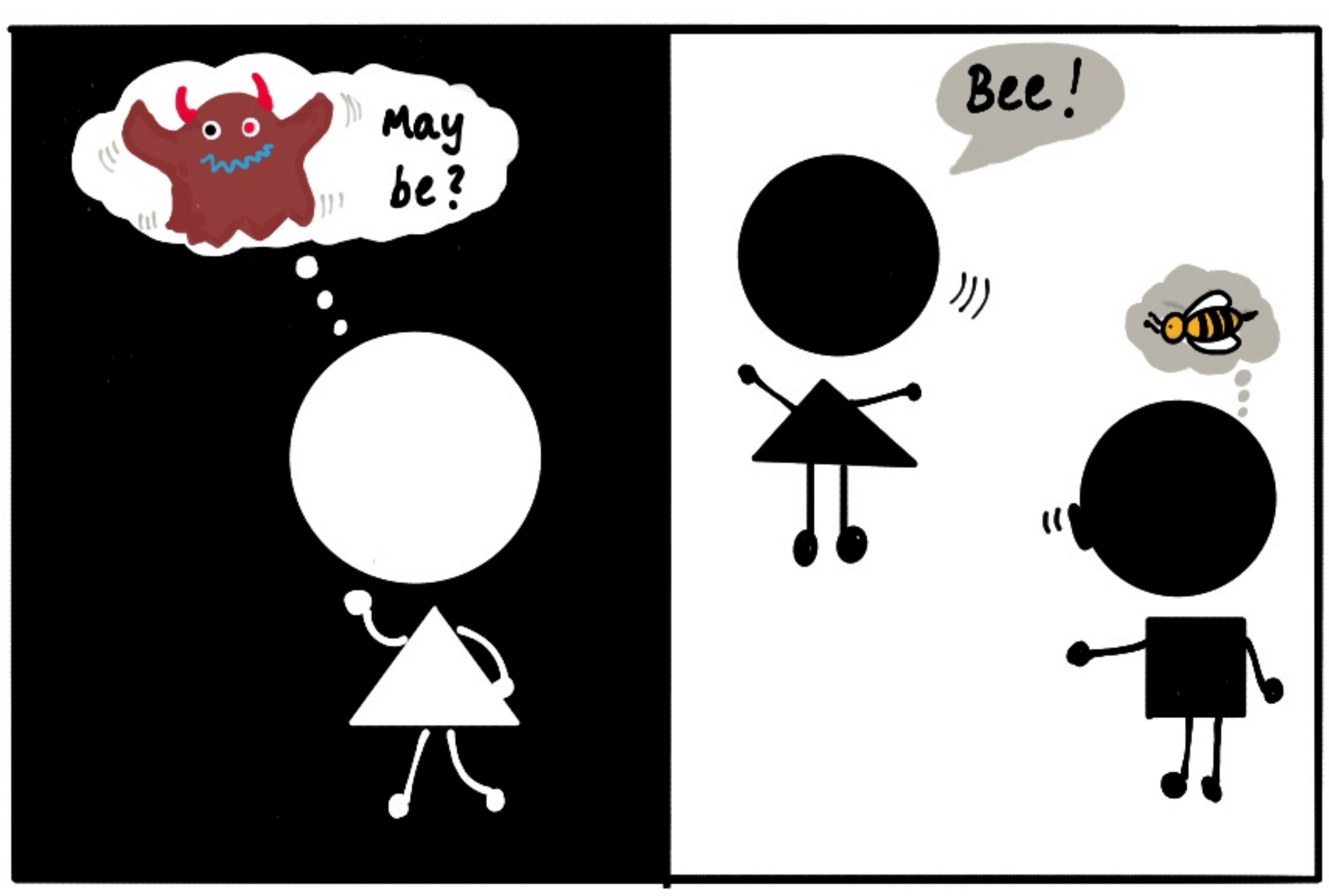
USES IT TO IMAGINE NEW THINGS

EXTERNAL

DESCRIBE THINGS

TELL STORIES

LEARN NEW THINGS



'SO, LANGUAGE IS AT THE CENTRE OF INTELLIGENCE'

– PATRICK WINSTON, MIT

HUMAN INTELLIGENCE

WE CAN AGREE THAT HUMANS CAN BE DESCRIBED AS INTELLIGENT.

THE ABILITY TO ACQUIRE AND APPLY KNOWLEDGE AND SKILLS

DICTIONARY

DEFINITION

A MENTAL QUALITY THAT ALLOWS US TO

SOME PSYCHOLOGISTS

LEARN

SOLVE PROBLEMS

ADAPT TO NEW SITUATIONS

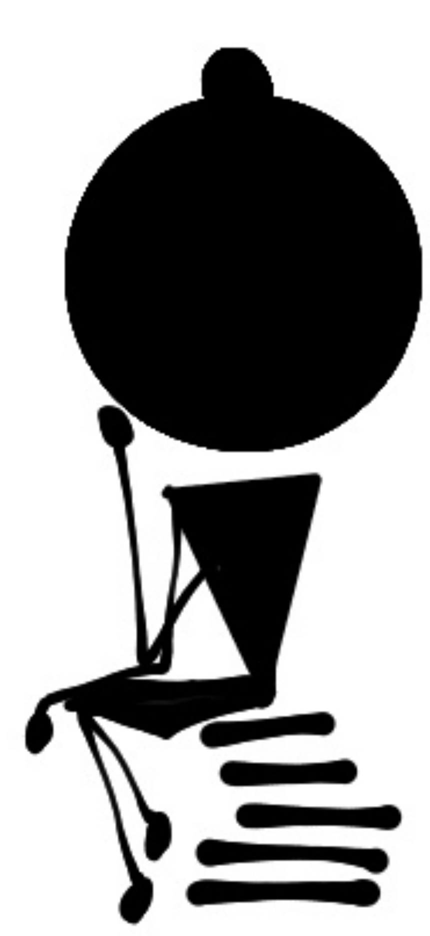
SUGGEST

THERE IS STILL SO MUCH OF THE BRAIN WE DON'T UNDERSTAND

WHAT ABOUT OUR GUT BRAIN?

WHAT CONNECTS INTELLIGENCE AND CONSCIOUSNESS?

HOW DOES MATTER (BRAIN) CREATE NON MATTER (THOUGHT)?



INTELLIGENCE OUTSIDE EARTH: IS IT SIMILAR?

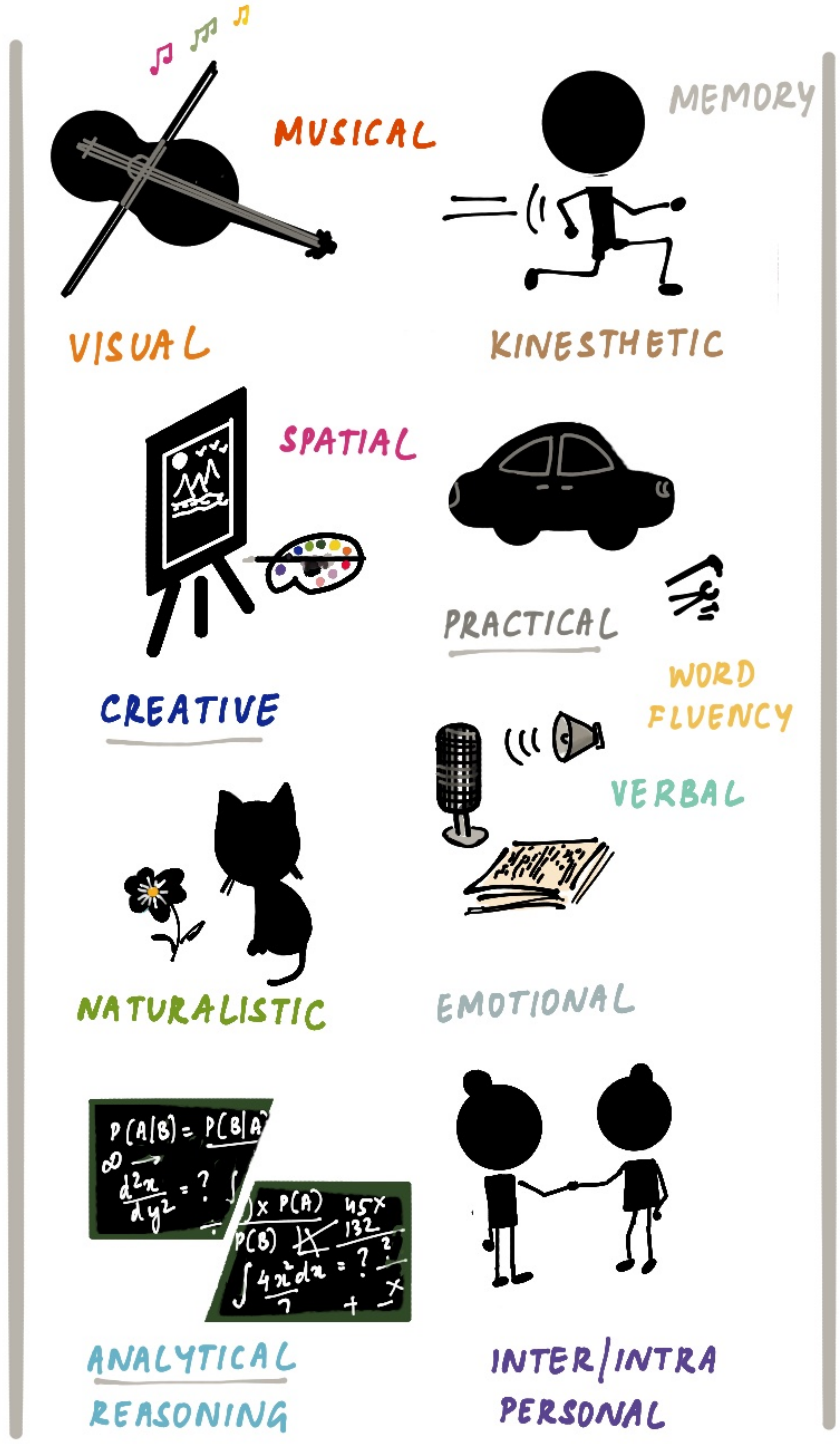
INTELLIGENCE ELUDES PRECISE DEFINITION AND MEASUREMENT

THEORIES OF INTELLIGENCE

THERE ARE MANY THEORIES OF INTELLIGENCE AS WELL.

CHARLES SPEARMAN
1863 – 1945
general intelligence
g factor

HOWARD GARDNER
1943 –
7-9 intelligences



LL THURSTONE
1887 – 1955
7 mental abilities

ROBERT STERNBERG
1949 –
3 intelligences

AND THE DEBATE CONTINUES:

- ARE THEY INTELLIGENCES OR SKILLS?
- ARE THESE GOOD ENOUGH CRITERIA TO 'MEASURE' INTELLIGENCE?
- IS BEING GOOD AT ONE ANY INDICATION OF BEING GOOD AT ANOTHER?
- IS THERE ONE GENERAL INTELLIGENCE OR MULTIPLE INTELLIGENCES?
- IS HUMAN INTELLIGENCE GENERAL OR SPECIALISED FOR HUMAN EXPERIENCE?

MACHINE INTELLIGENCE

HISTORY RECORDS MANY INSTANCES OF MECHANICAL CREATIONS: BOTH REAL AND FICTIONAL. THERE ARE EVEN LITERARY WORKS DESCRIBING THINKING BEINGS. IT IS AGAINST THIS BACKDROP THAT WE MEET RENÉ DESCARTES.

MIND BODY DUALISM

RENÉ DESCARTES

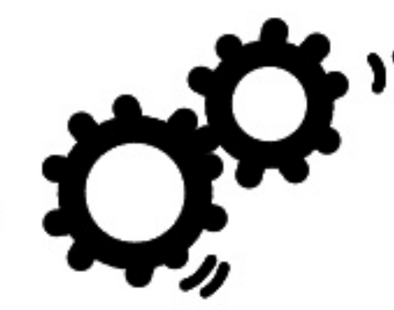
17th century

RENÉ DESCARTES DEFINED **THE MIND-BODY PROBLEM**

ANIMALS ARE WONDERFUL MACHINES.

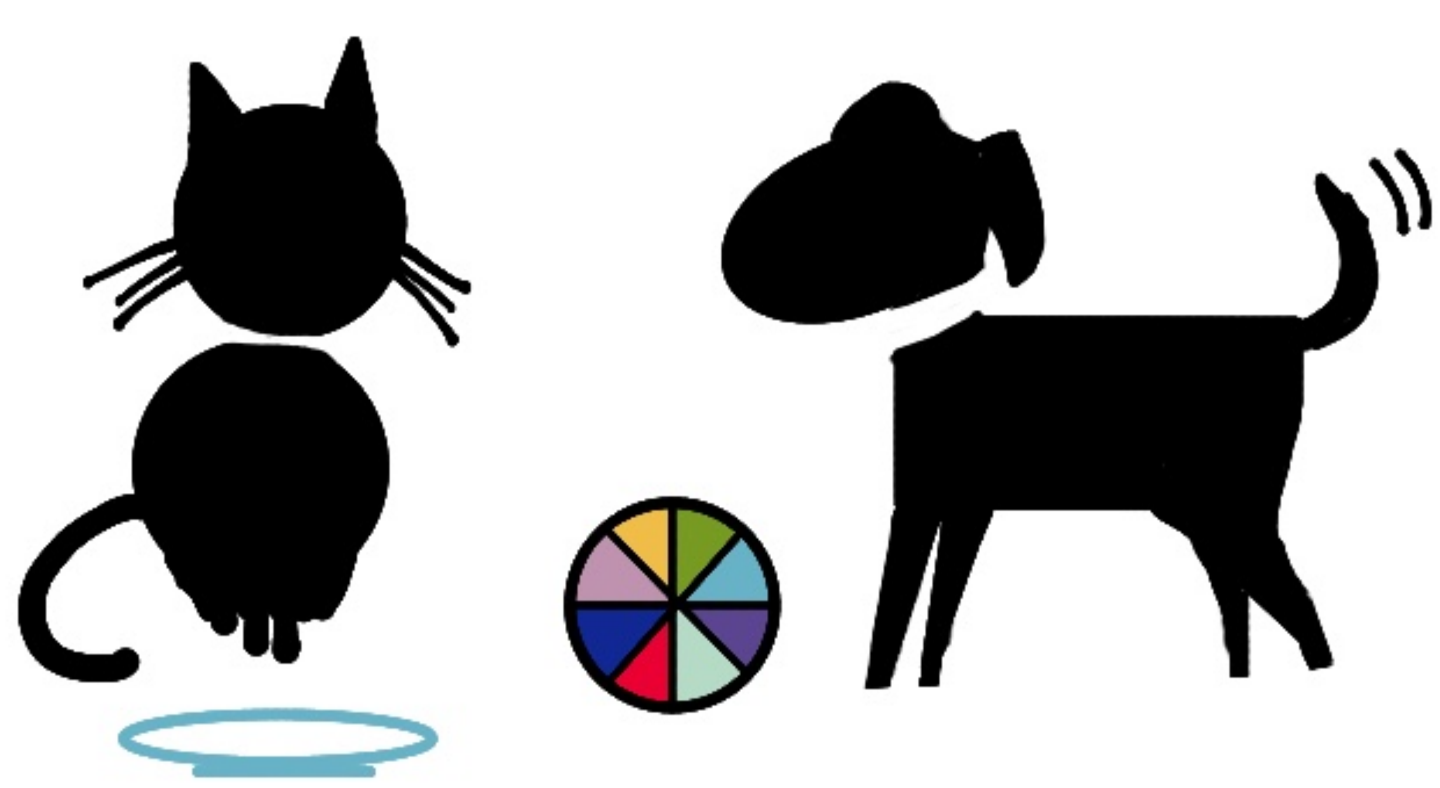
HUMANS TOO. POSSIBLY. EXCEPT WE HAVE MINDS.

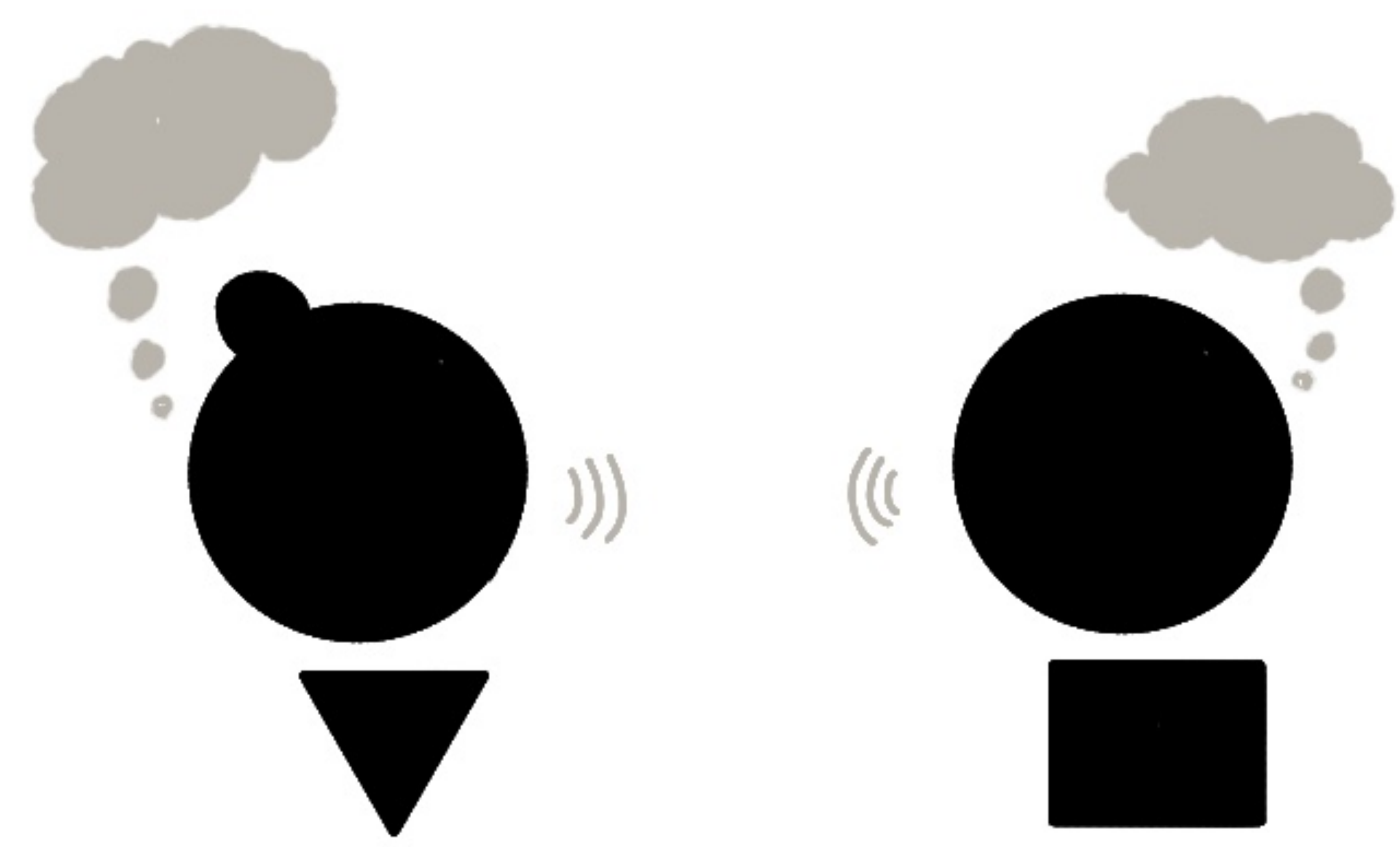
RATIONAL Vs MECHANICAL



THE MECHANICAL COULD BE IMITATED. NOT THE RATIONAL.

ANIMALS MIGHT ALSO HAVE CONSCIOUSNESS, MEMORY AND FEELING.





BUT HUMANS HAVE LANGUAGE AND THE ABILITY TO THINK.

TWO CENTURIES LATER (1830s), ADA LOVELACE WROTE OF THE ANALYTICAL ENGINE

ADA LOVELACE

"...HAS NO PRETENSIONS WHATEVER TO ORIGINATE ANYTHING. IT CAN DO WHATEVER WE KNOW HOW TO ORDER IT TO PERFORM"

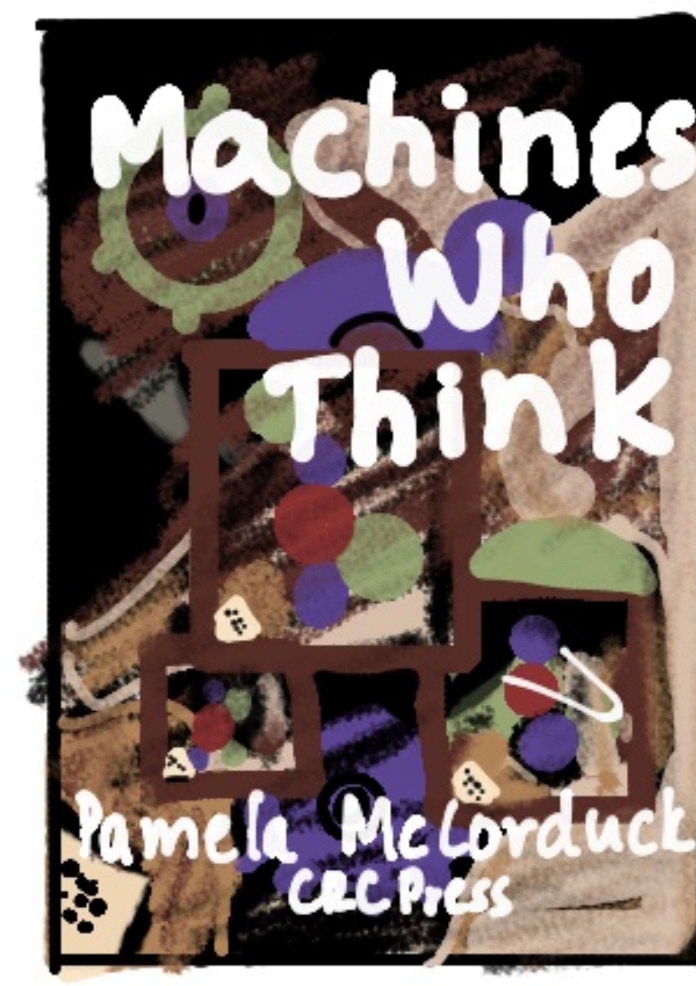
BUT, WHAT IF THERE WAS A MACHINE THAT COULD IMITATE THE MIND?

A SELF COPY

ARTIFICIAL INTELLIGENCE BEGAN WITH AN ANCIENT WISH TO FORGE THE GODS

PAMELA MCCORDUCK

BRASS FOR BRAIN

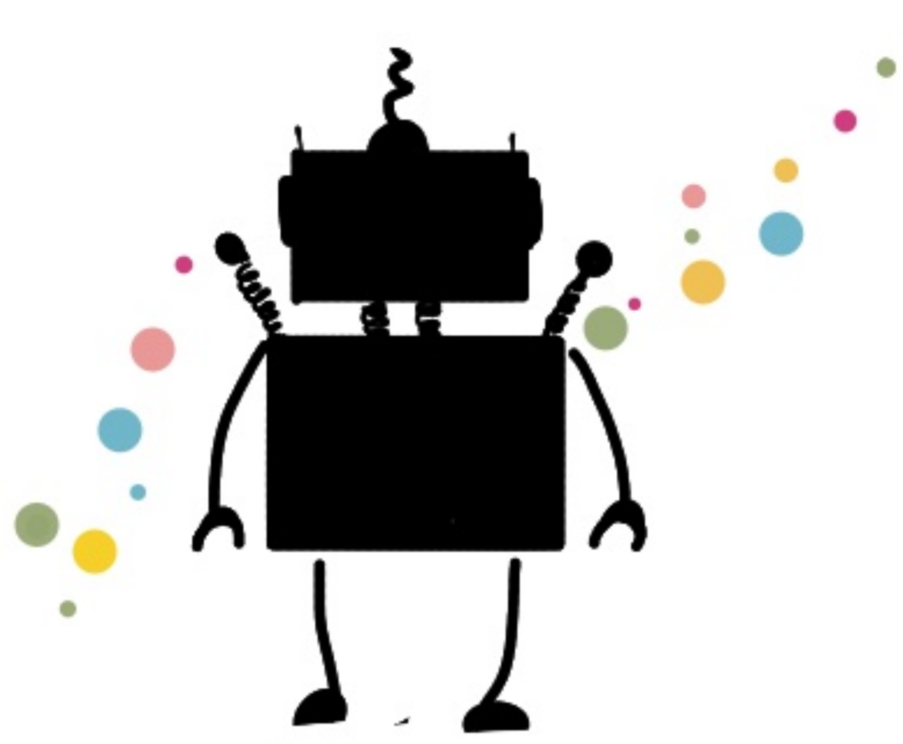


THE DESIRE TO CREATE ARTIFICIAL INTELLIGENCE IS AS OLD AS HUMAN CIVILISATION

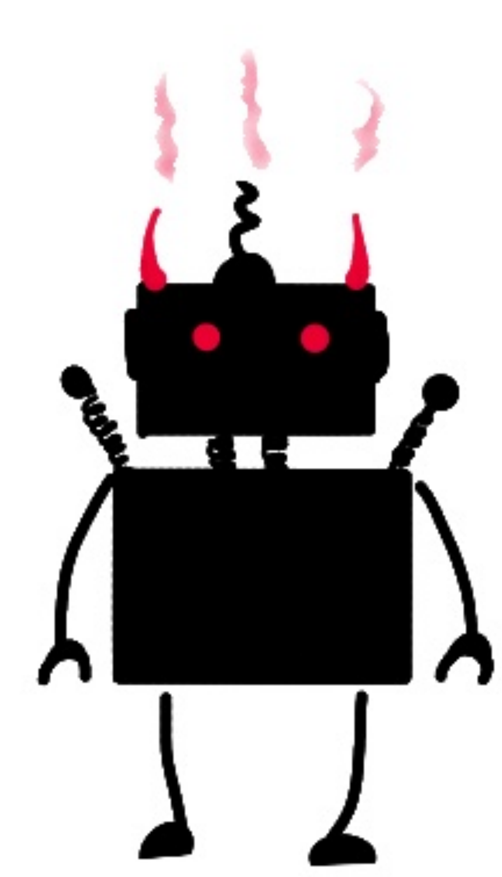
PAMELA MCCORDUCK GIVES EXAMPLES IN THE FIRST EVER BOOK ON THE HISTORY OF AI

CLOCKWORK REPLICAS OF HUMANS EXIST MORE AS STORIES THAN REALITY

MOST SUCH CHARACTERS, WHETHER ROBOTIC ATTENDANTS OR TINMAN, WERE FRIENDLY, OR ACCOMPLISHED SOME TASKS.



THE STORIES CHANGED THEME. EVENTUALLY.



GOOD OR EVIL?

OF THE STORIES OF SUCH HUMAN ATTEMPTS, THE MOST CHILLING ACCOUNT OF ALL, IS PROBABLY THE 1818 STORY OF FRANKENSTEIN'S MONSTER BY MARY SHELLEY

THE PROTAGONIST VICTOR FRANKENSTEIN'S CREATION TURNS OUT TO BE A MONSTER AND STARTS TO THREATEN AND DESTROY HIS LIFE.

FRANKENSTEIN'S MONSTER

BUT IT NEED NOT BE THIS TERRIFYING.

WE CAN DRAW SOME REASSURANCE FROM ISAAC ASIMOV'S IDEA OF THE THREE LAWS OF ROBOTICS THAT PRIORITISE HUMAN WELL-BEING

ISAAC ASIMOV

THREE LAWS OF ROBOTICS

1. A robot must not injure a human or through inaction allow a human to come to harm
2. A robot must obey orders except if it conflicts with 1
3. A robot must protect its own existence unless it conflicts with 1 and 2

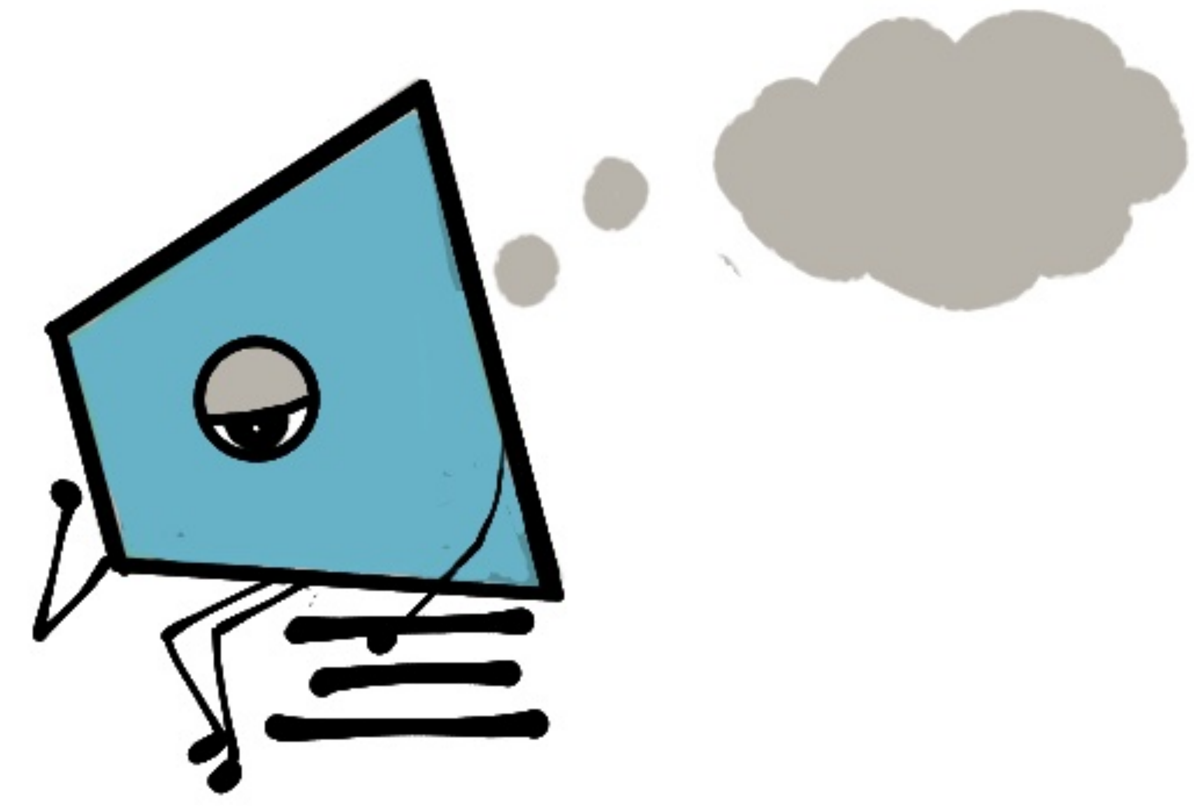
FROM THE NOVEL I, ROBOT 1950

THESE STORIES FORCE US TO CONSIDER QUESTIONS THAT WE OUGHT TO AT LEAST ASK, IN THE PROCESS OF CREATING ARTIFICIAL INTELLIGENCES.

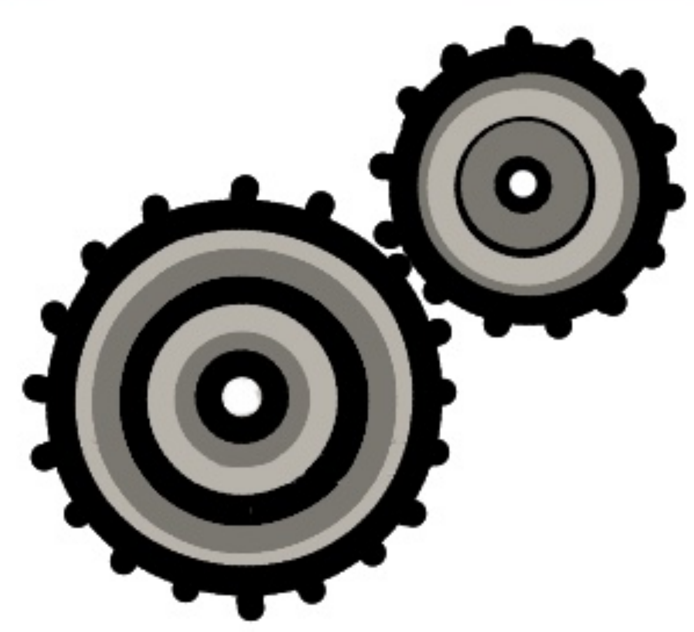
AI IS... WHAT AI DOES

GIVEN WHAT WE UNDERSTAND OF HUMAN INTELLIGENCE, UNSURPRISINGLY, ARTIFICIAL INTELLIGENCE HAS NO AGREED DEFINITIONS

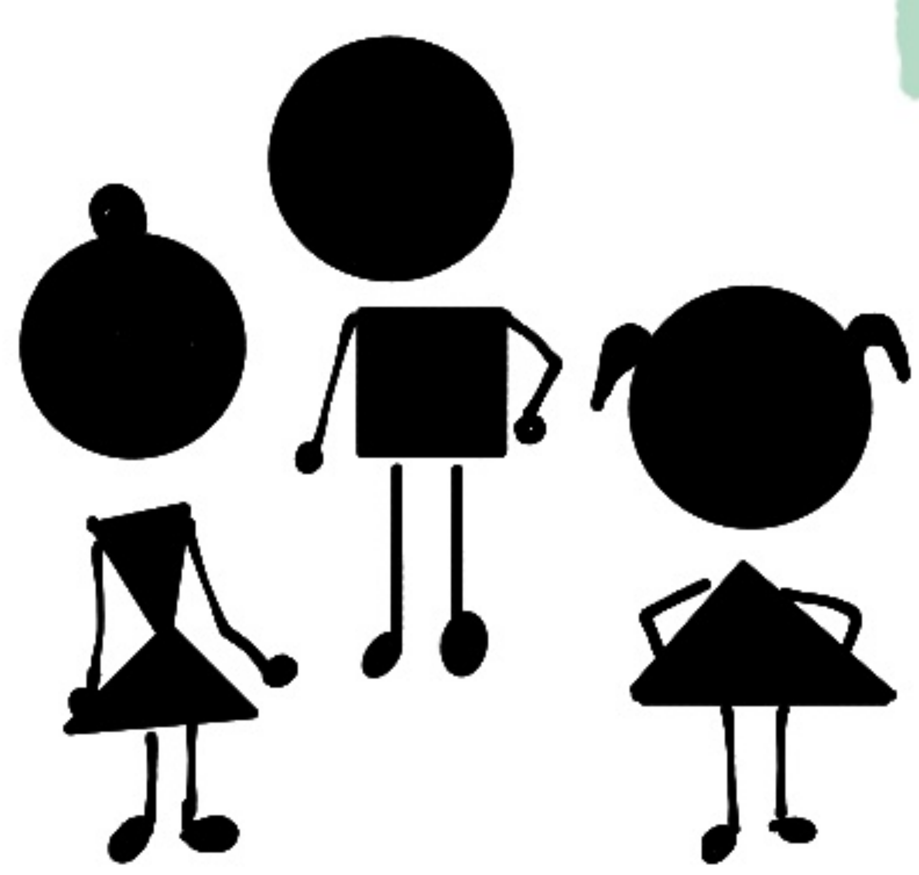
THE THINKING MACHINE IDEA IS VERY CAPTIVATING AND IT REMAINS A DEFINING THEME OF ARTIFICIAL INTELLIGENCE



THE GOAL OF AI IS TO BUILD SYSTEMS



THAT THINK AND ACT



LIKE HUMANS

RATIONALLY

ASIDE FROM THE PUZZLES OF WHAT IT MEANS TO BE HUMAN OR RATIONAL, THERE ARE ASPIRATIONS FOR ACHIEVING SUPERHUMAN INTELLIGENCE

(CAUSING REACTIONS OF FEAR/SKEPTICISM)

Singularity
Intelligence Explosion
Artificial General Intelligence
AGI, AIXI are some terms used for this aspiration

AN ASIDE

ON THE BASIS THAT INTELLIGENCE REQUIRES THE 'AGENT' TO ADAPT SUCCESSFULLY TO A WIDE RANGE OF ENVIRONMENTS

THERE IS EVEN AN EQUATION TO MEASURE A MACHINE'S INTELLIGENCE

A FIRST FORMALISATION

$$\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_{\mu}^{\pi}$$

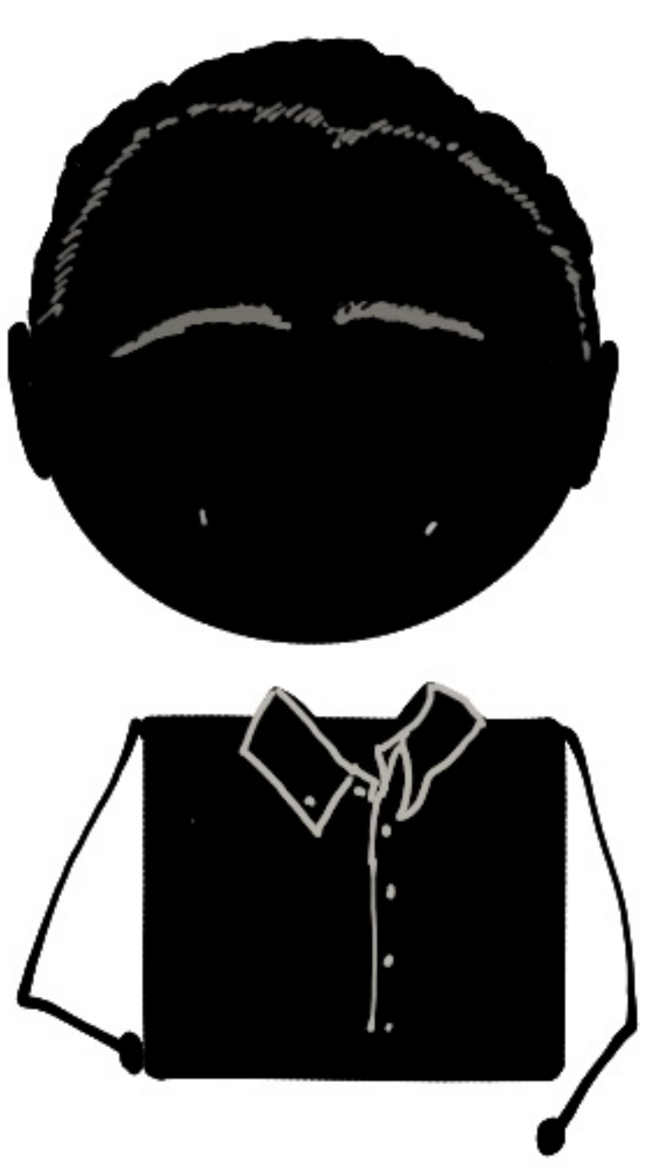
agent

Measure of intelligence

Sum over Environments

complexity function

Value achieved



SHANE LEGG

arxiv.org/abs/0712.3329

Universal Intelligence

A Definition of Machine Intelligence

Shane Legg & Marcus Hutter

MARCUS HUTTER

THE EXPERTS: ON AI

HERE IS WHAT SOME WELL-KNOWN PEOPLE HAVE TO SAY ABOUT AI

AI IS A FIELD THAT IS DEEPLY HUMAN
(with nothing artificial about it)

— FEI FEI LI

AI IS, IN LARGE MEASURE, PHILOSOPHY

— DANIEL DENNETT

AI IS ALMOST A HUMANITIES DISCIPLINE. AN ATTEMPT TO UNDERSTAND HUMAN INTELLIGENCE AND COGNITION

— SEBASTIAN THRUN

AI MEASURES AN AGENT'S ABILITY TO PERFORM WELL IN A WIDE RANGE OF ENVIRONMENTS

— MARCUS HUTTER | SHANE LEGG

THE EFFICIENCY WITH WHICH YOU TURN EXPERIENCE INTO GENERALISABLE PROGRAMS

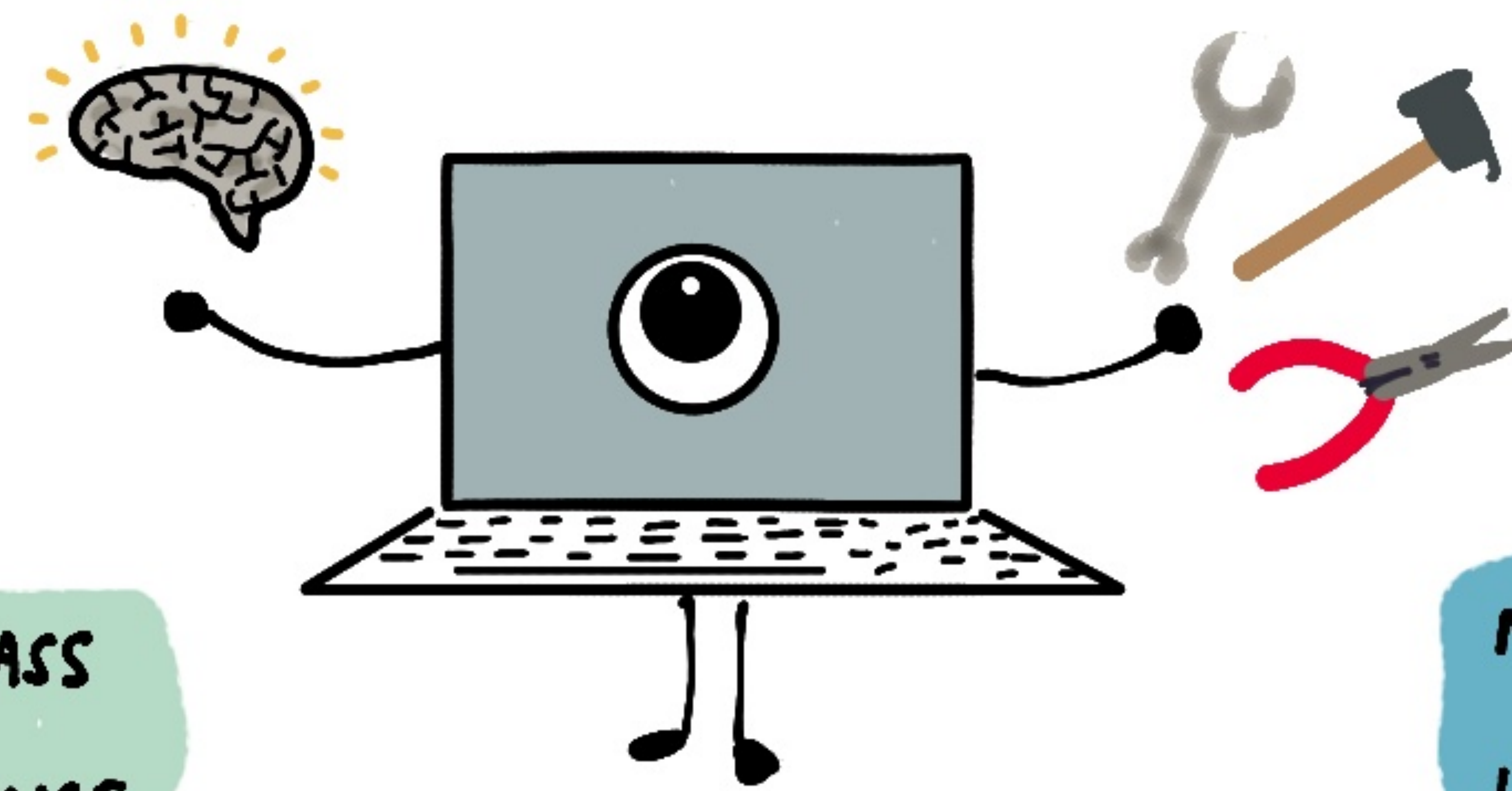
— FRANÇOIS CHOLLET

EACH NEW STEP IN AI MERELY REVEALS WHAT REAL INTELLIGENCE IS NOT

— DOUGLAS HOFSTADTER

GOALS OF AI

VERY BROADLY, AI MIGHT BE SAID TO HAVE TWO (RELATED) GOALS



1

MIMIC AND SURPASS HUMAN INTELLIGENCE

2

MAKE TOOLS USEFUL TO HUMANS

THE DEVELOPMENT OF AI COULD SPELL THE END OF THE HUMAN RACE

-STEPHEN HAWKING

IT IS A FASCINATING ENDEAVOUR TO UNDERSTAND THE NATURE OF THOUGHT, CONSCIOUSNESS AND INTELLIGENCE.

AS IT NEEDS MORE YEARS OF RESEARCH, WE WILL CONSIDER IT OUTSIDE THE SCOPE OF THIS PIECE OF WORK

AS WE SAW EARLIER, THE TOOLS CREATED BY LEARNING MACHINES

CAN

PROCESS/INTERPRET LARGE AMOUNTS OF DATA DIFFICULT FOR HUMANS

AND

AUTOMATE SOME TEDIOUS OR SUBCONSCIOUS TASKS THAT PROGRAMS PREVIOUSLY COULD NOT

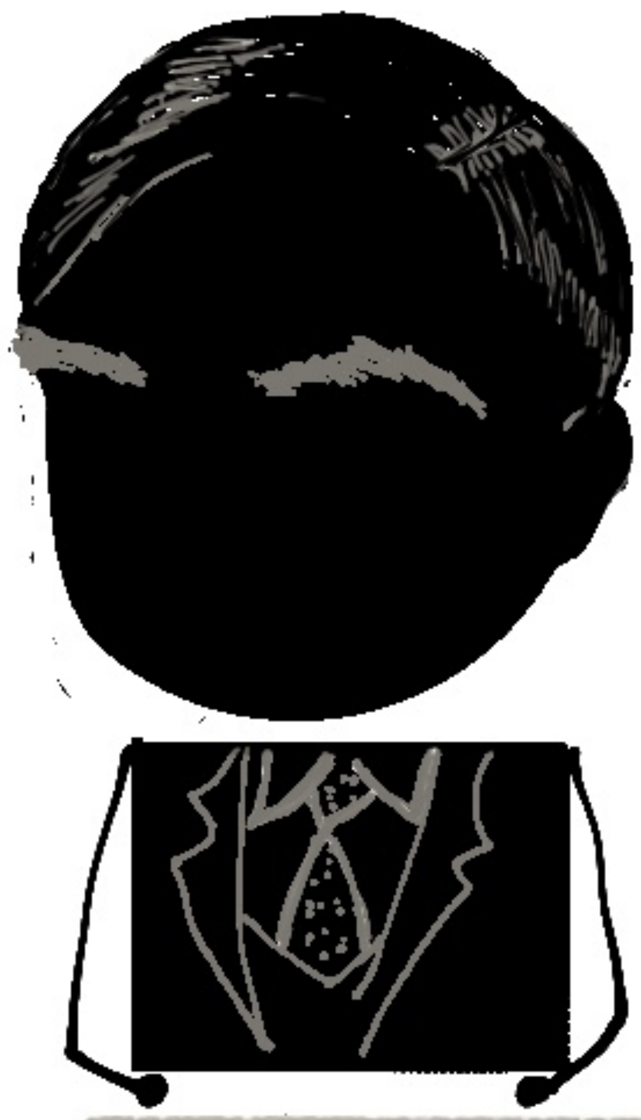
THE SECOND GOAL LEADS TO WHAT CAN BE DESCRIBED AS NARROW INTELLIGENCE

BUT FIRST, A LOOK AT HOW WE GOT HERE.

A RECENT HISTORY

GENERATIONS OF THINKERS FROM ARISTOTLE, GOTTFRIED LEIBNITZ, GEORGE BOOLE, BERTRAND RUSSELL, ALL HAVE ATTEMPTED TO FORMALISE HUMAN THOUGHT AND DECISION MAKING. THE 1950s NOTCHED UP THE EXCITEMENT.

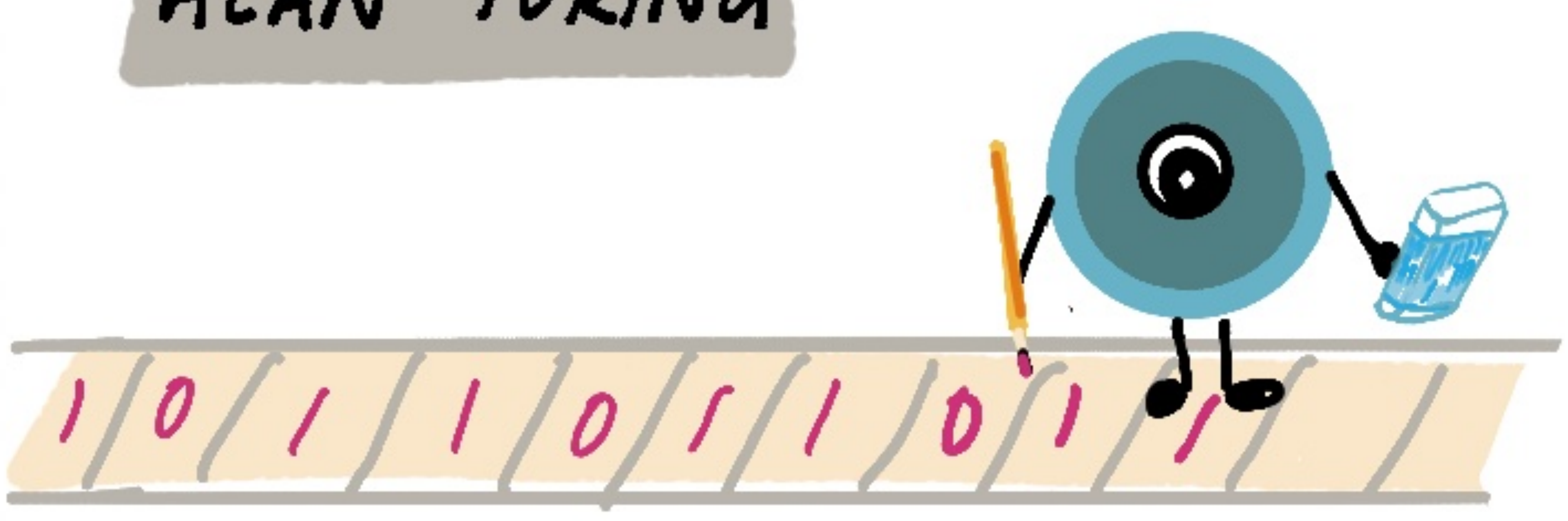
AI FROM THE '50s



ALAN TURING

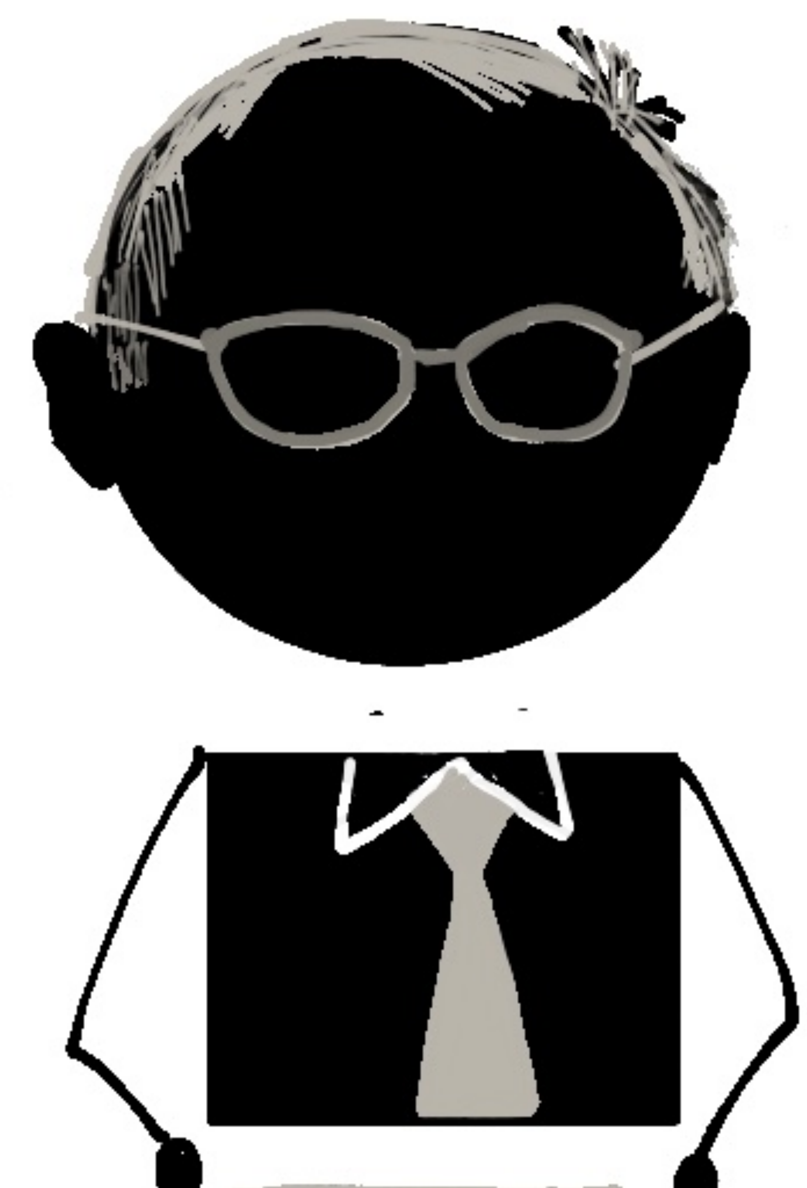
ALAN TURING WONDERED IF A MACHINE COULD EVER CONVINCE HUMANS THAT IT COULD THINK.

AND HE WROTE A PAPER ABOUT IT.

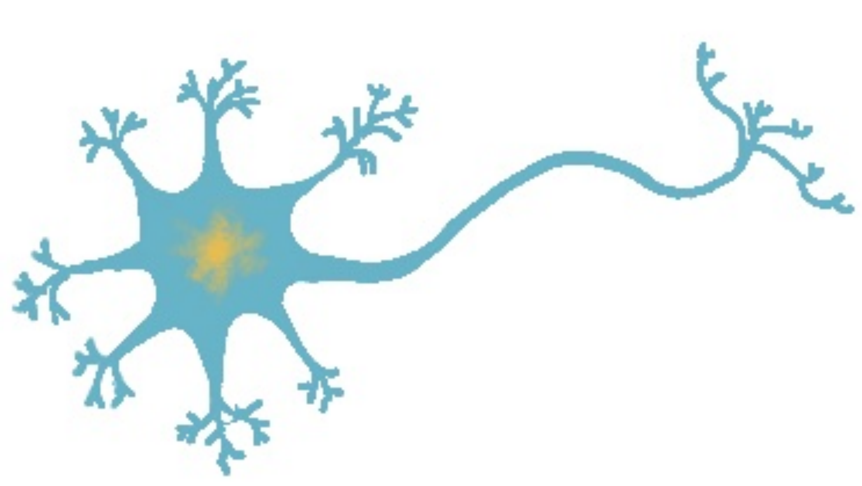


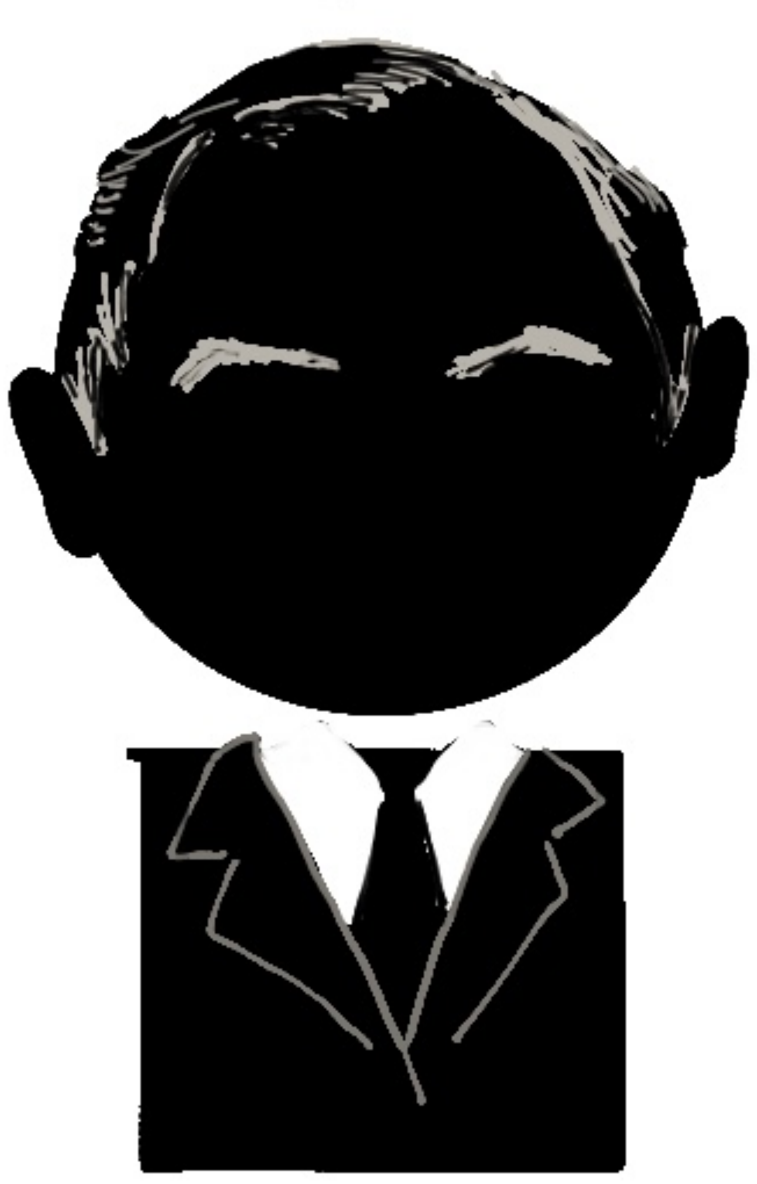
MEANWHILE, MCCULLOCH AND PITTS CAME UP WITH THE IDEA THAT THE BRAIN BEHAVES LIKE A TURING MACHINE

WARREN MCCULLOCH



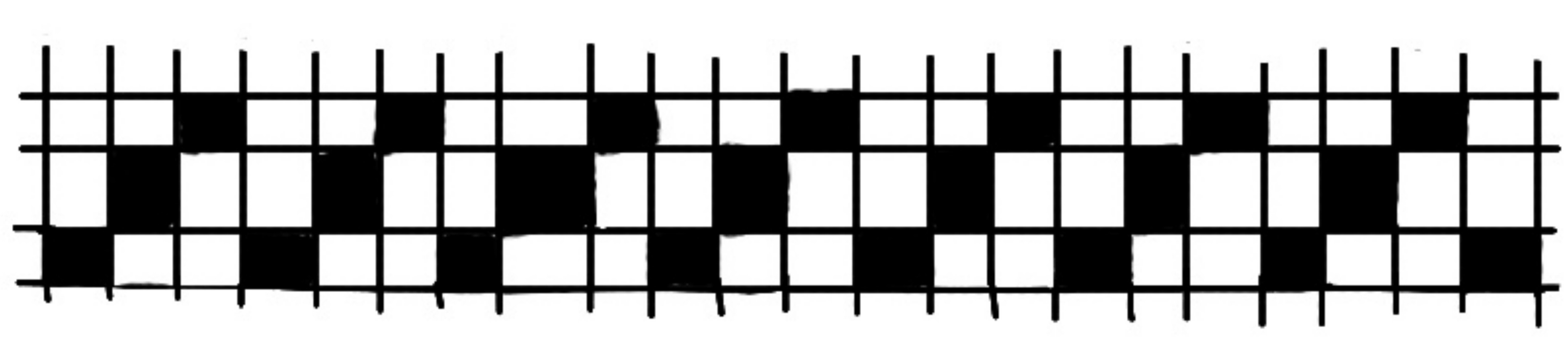
WALTER PITTS





JOHN VON NEUMANN

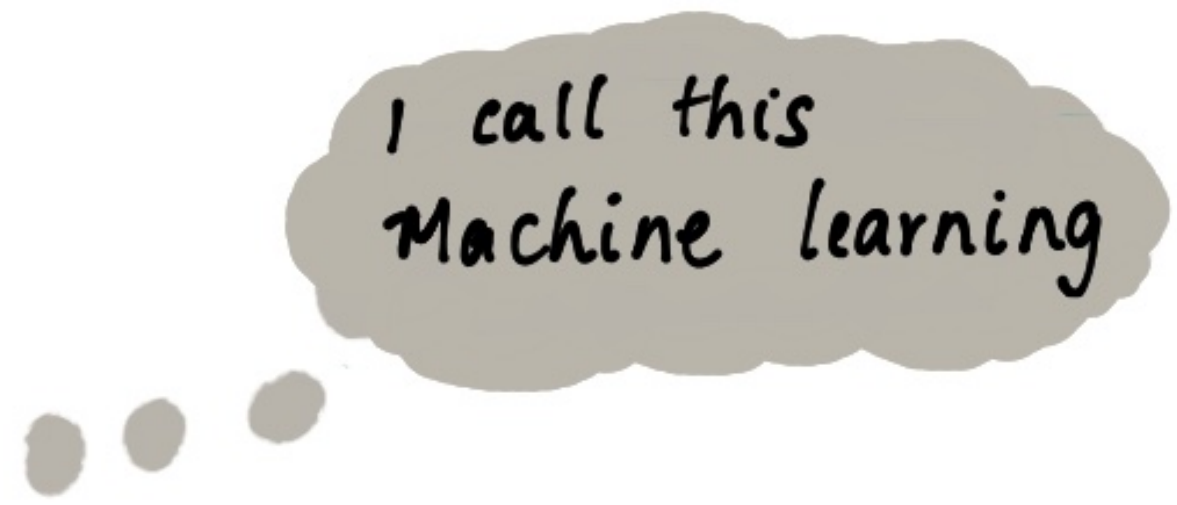
VON-NEUMANN AAREED WITH THE BRAIN-COMPUTER SIMILARITIES- BUT PERHAPS NOT WITH 'THINKING' MACHINES

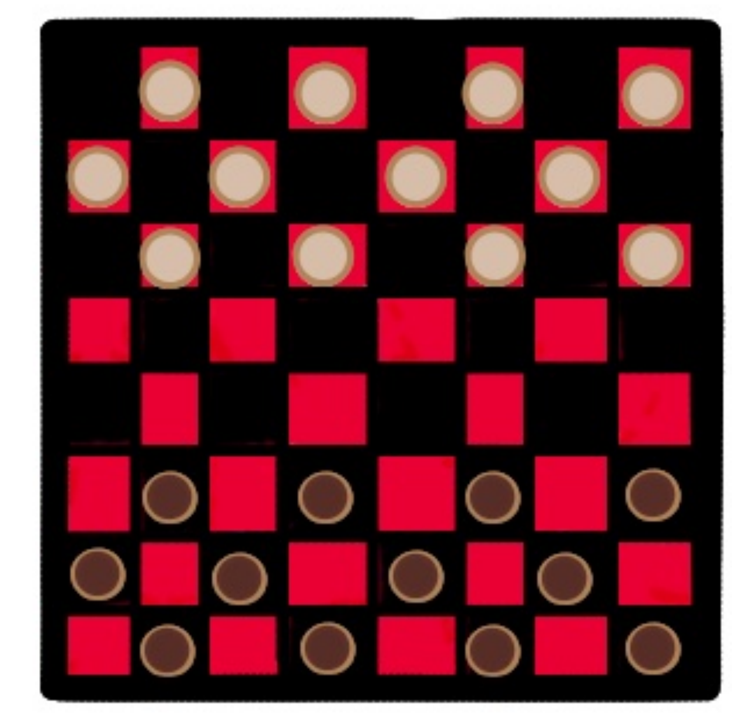


HE HAD HIS OWN IDEA OF CELLULAR AUTOMATA, WHICH WERE SELF-REPLICATING MATHEMATICAL STRUCTURES

AI FROM THE '50s

CHECKERS





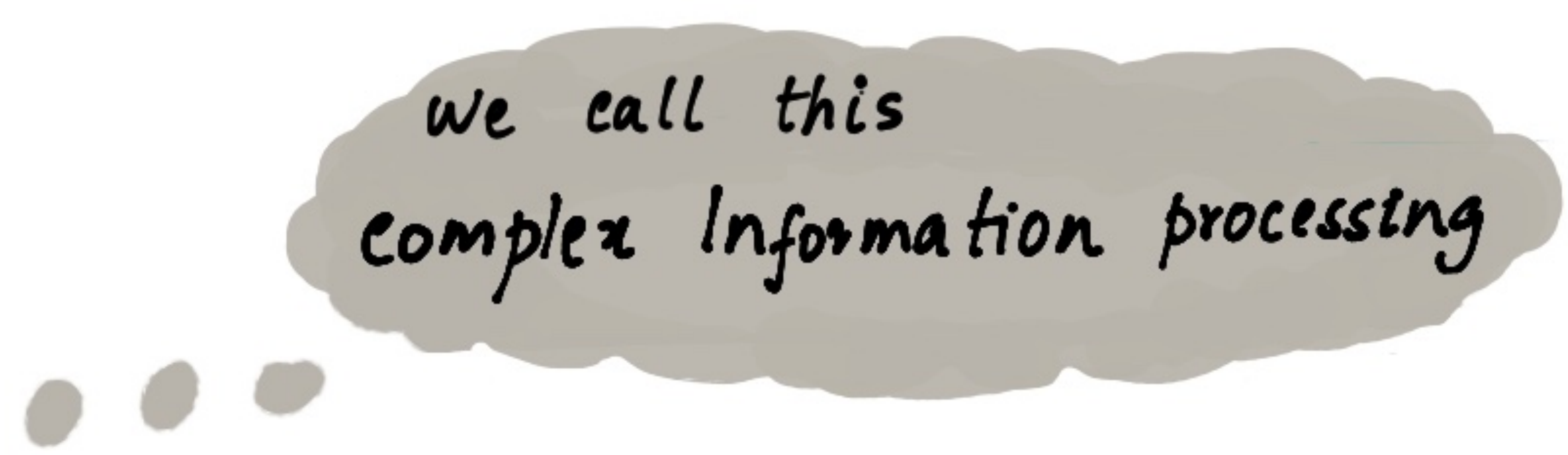
ARTHUR SAMUEL CREATED A CHECKERS PLAYING PROGRAM.

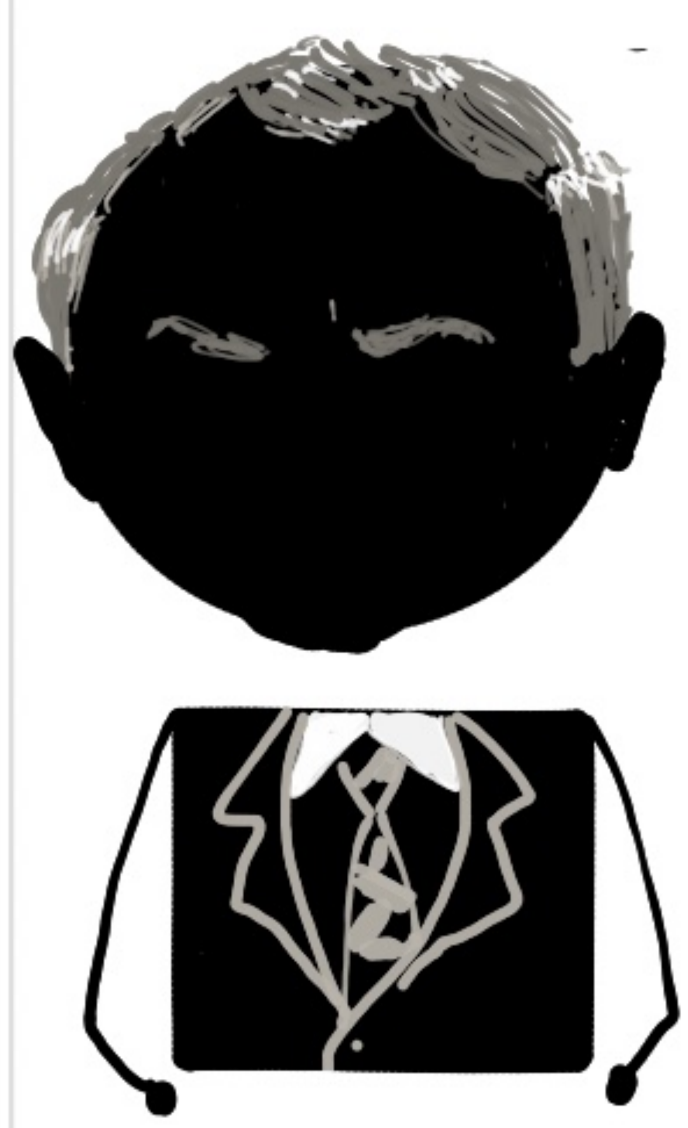
IT CHECKED MANY POSSIBLE MOVES AHEAD, ASSIGNED WEIGHTS TO EACH OUTCOME.

ARTHUR SAMUEL

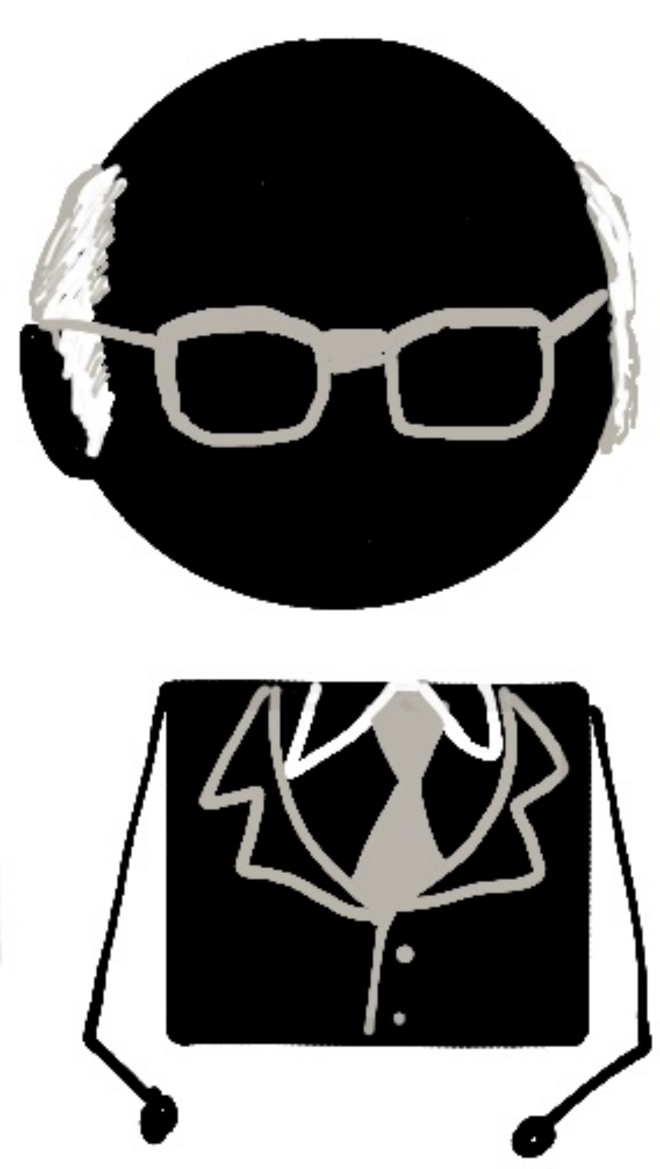
IT IMPROVED ITS GAME FROM PAST EVENTS

LOGIC THEORIST

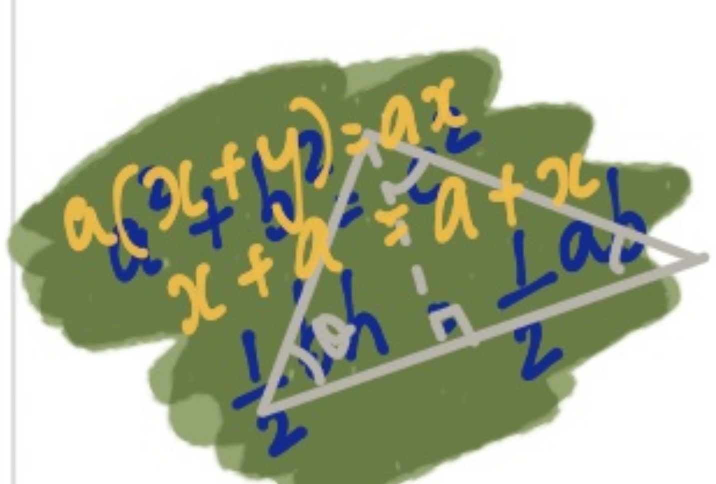




HERBERT SIMON



ALLEN NEWELL



NEWELL AND SIMON, WITH JC SHAW, FAR AHEAD OF THEIR TIME, WROTE A PROGRAM CALLED LOGIC THEORIST.

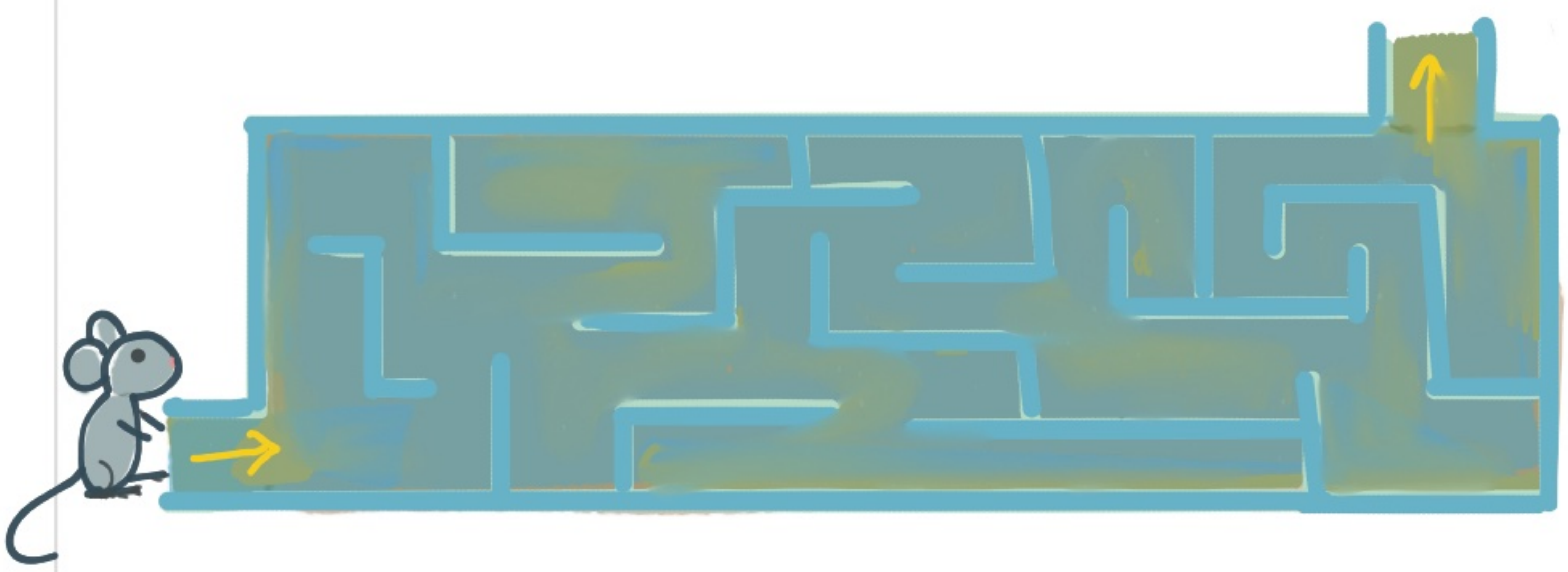
IT PROVED THEOREMS FROM PRINCIPIA MATHEMATICA OF RUSSELL & WHITEHEAD.

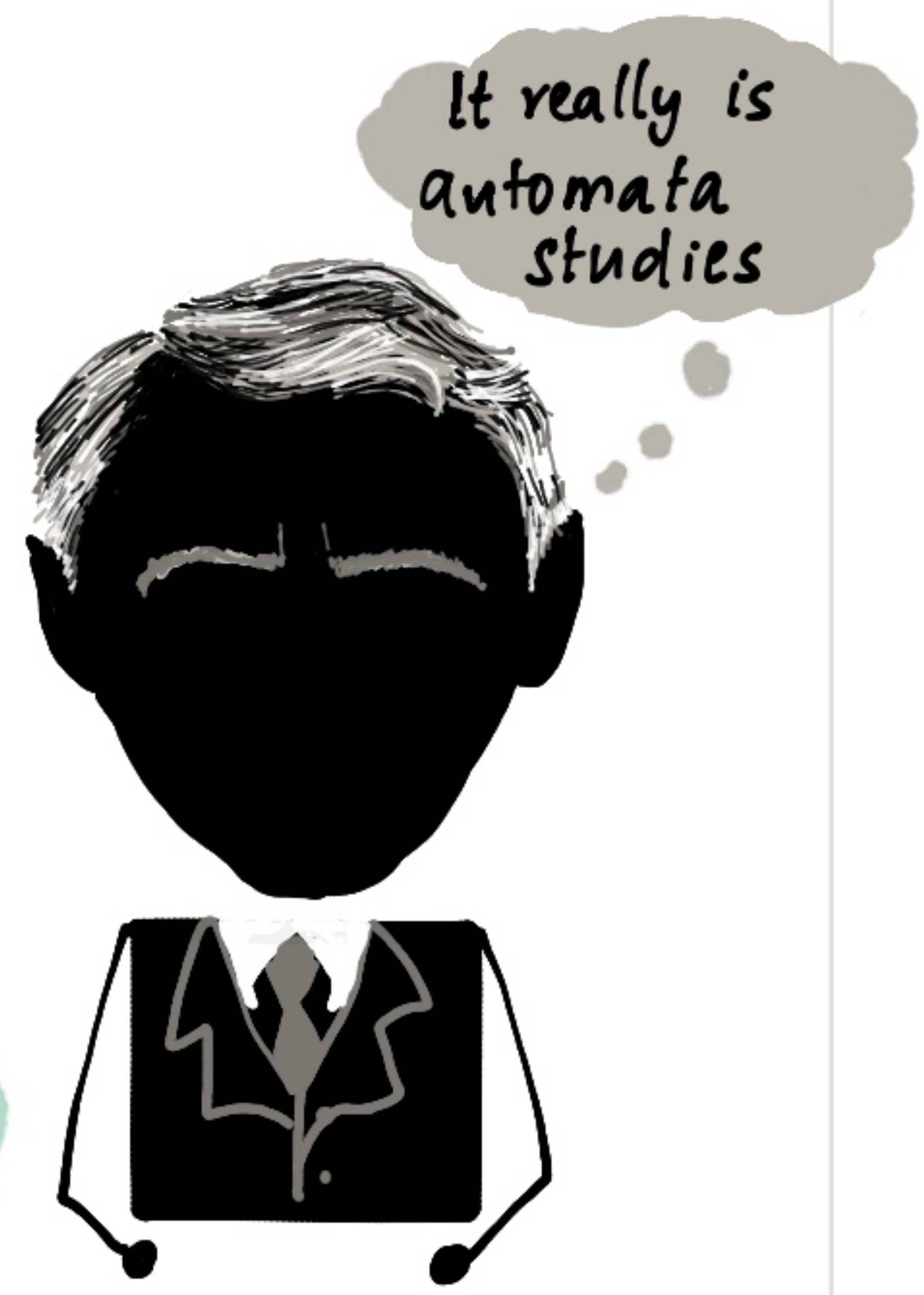
AI FROM THE '50s

NORBERT WIENER USED THE WORD CYBERNETICS TO DESCRIBE THE STUDY OF ANY SOCIAL SYSTEM (HUMAN, ANIMAL OR MACHINE) BASED ON COMMUNICATION AND FEEDBACK

HIS WORK, BOTH TECHNICAL & PHILOSOPHICAL, WAS ALSO VERY INFLUENTIAL.

NORBERT WIENER





CLAUDE SHANNON WROTE A CHESS PLAYING PROGRAM. HE ALSO BUILT A PHYSICAL MAZE AND A MOUSE THAT COULD 'LEARN' ITS WAY OUT OF THE MAZE

CLAUDE SHANNON

MARGARET MASTERMAN CREATED THE LANGUAGE RESEARCH UNIT AND THE WORK SHE DID WAS MACHINE TRANSLATION – YEARS AHEAD OF TIME – AND WITHOUT PROPER CREDIT

MARGARET MASTERMAN

THE DARTMOUTH CONFERENCE

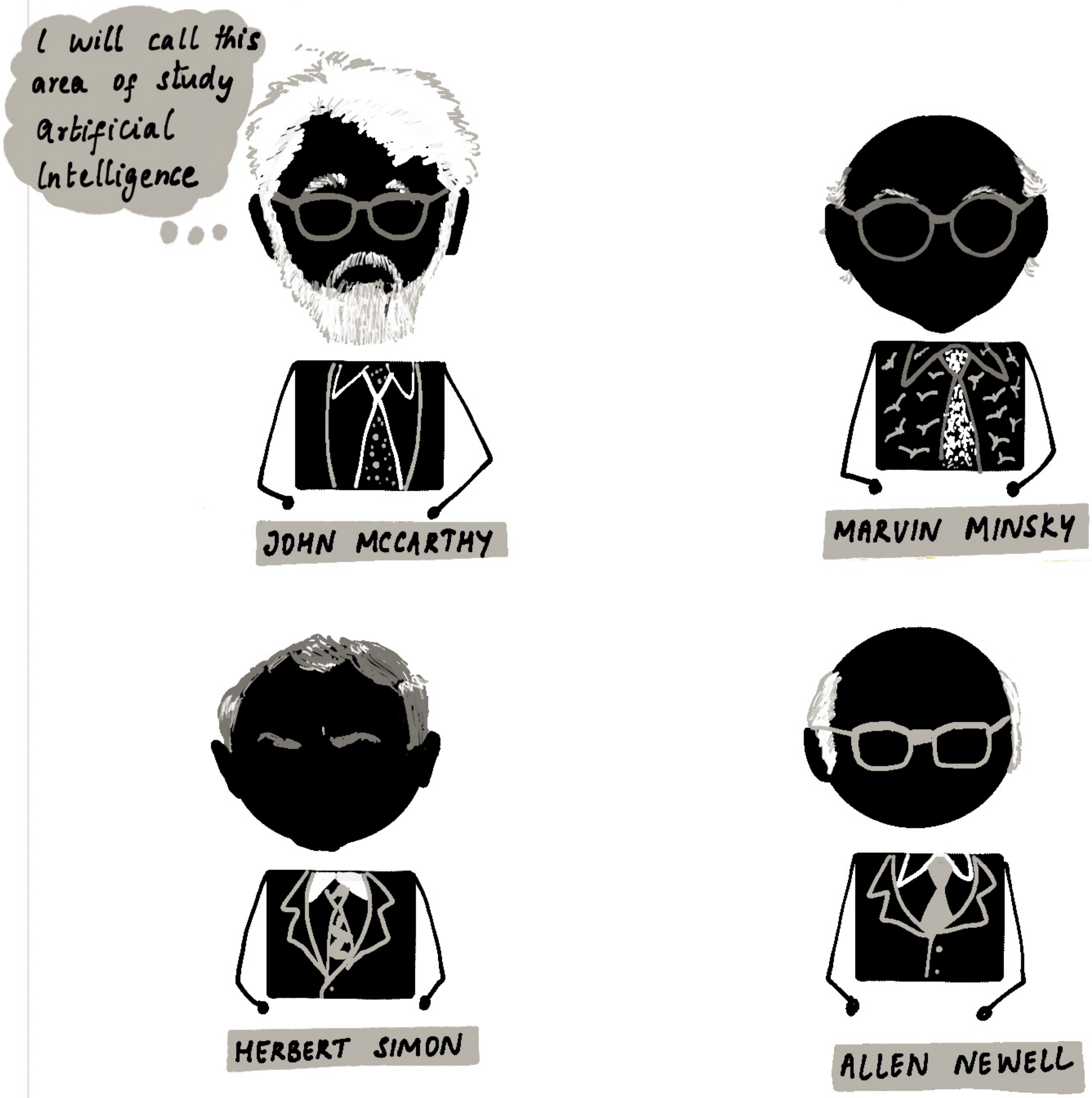
We propose that a two month study be carried out in the Dartmouth college, New Hampshire over the summer of 1956 ....

. . . .

... learning or any other feature of intelligence ..so precisely described that a machine can be made to simulate it.....

– McCARTHY, MINSKY, ROCHESTER, SHANNON

HERE ARE THE FOUR CONSIDERED THE FOUNDERS OF THE FIELD



TO THEM, THE LOGIC-BASED REASONING OF THE BRAIN WAS THE KEY

ENCODED KNOWLEDGE

HERE ARE SOME PROGRAMS DEVELOPED ON A LOGICAL BASIS

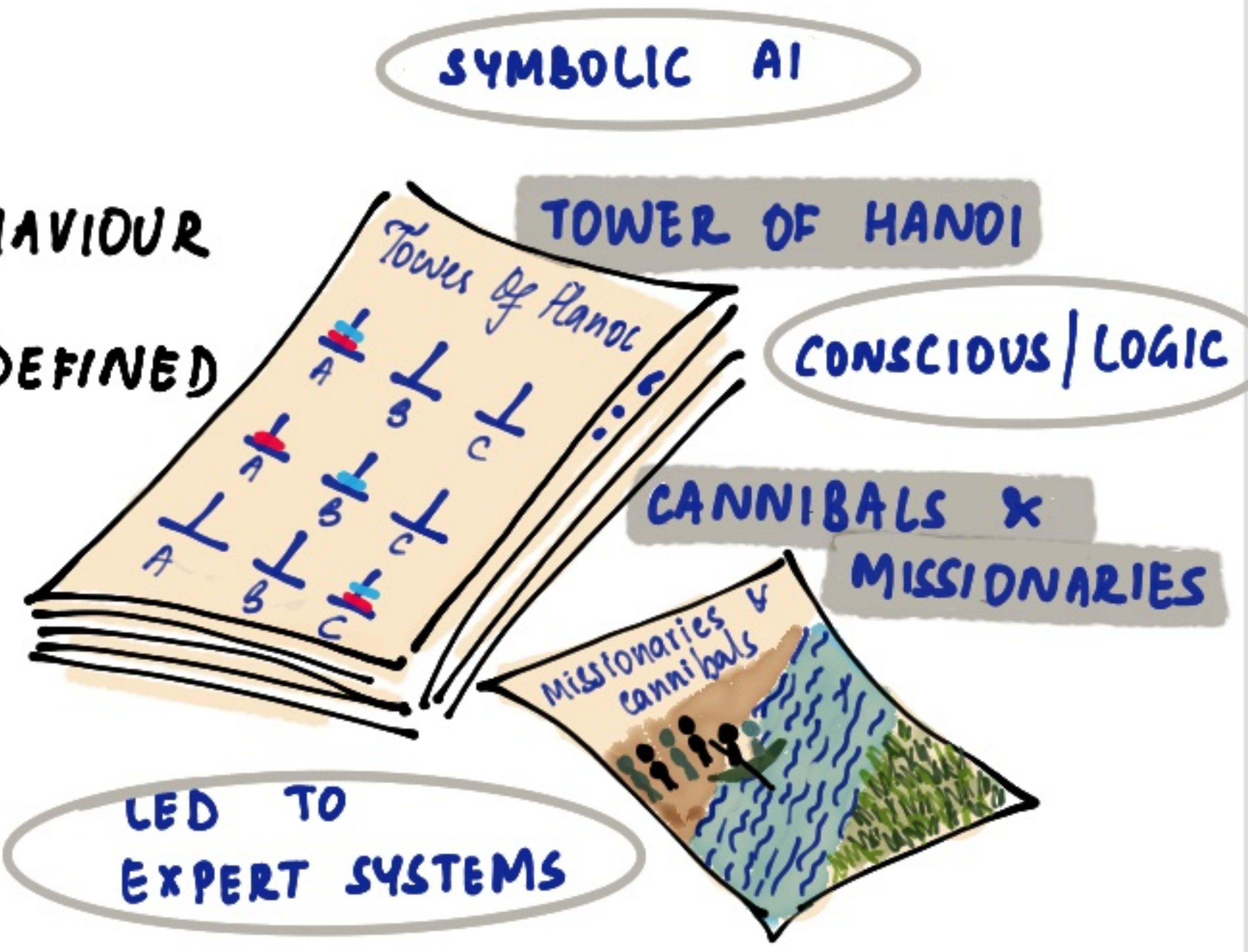
GENERAL PROBLEM SOLVER (1959)

THE FIRST APPROXIMATION OF HUMAN BEHAVIOUR

SOLVED PROBLEMS THAT COULD BE WELL DEFINED

(NOT VERY GENERAL)

CREATED BY NEWELL, SIMON & SHAW



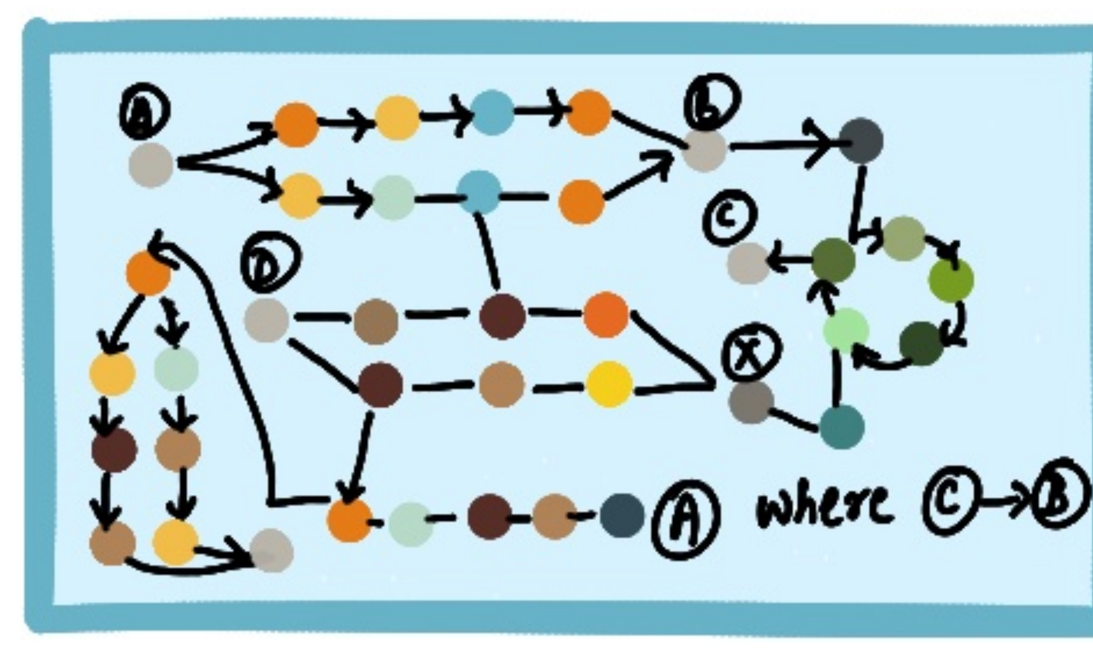
ADVICE TAKER (LATE 1960s)

A CONCEPTUAL MODEL OF A PROGRAM THAT

NOT ONLY SOLVES A LOT OF PROBLEMS,

BUT ALSO TAKES ADVICE WHILE SOLVING THEM.

CONCEPT OF JOHN MCCARTHY



ELIZA (MID 1960s)

THE FIRST CHATBOT

COULD MIMIC A PSYCHOTHERAPIST

NAMED AFTER THE CHARACTER IN PYGMALION

CREATED BY JOSEPH WEIZENBAUM, MIT,

SHAKEY THE ROBOT (LATE 1960s)

THE FIRST MOBILE ROBOT ABLE TO REASON

PROGRAMMED IN LISP

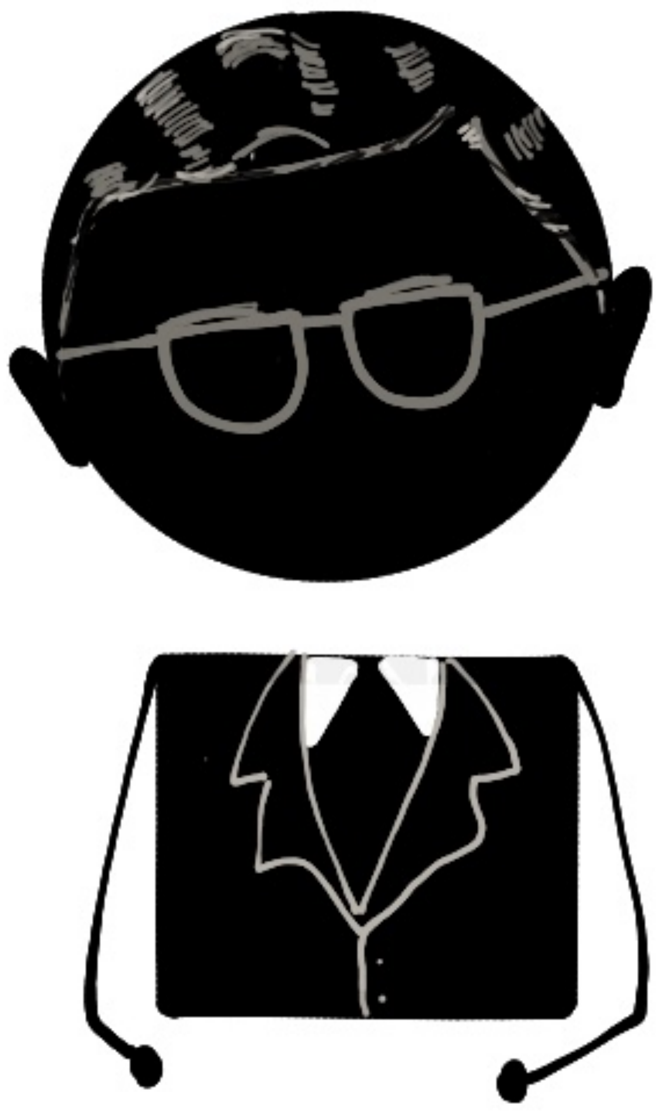
CREATED BY TEAM AT STANFORD | DARPA

(INCLUDING BERT RAPHAEL, NILS NILSSON)

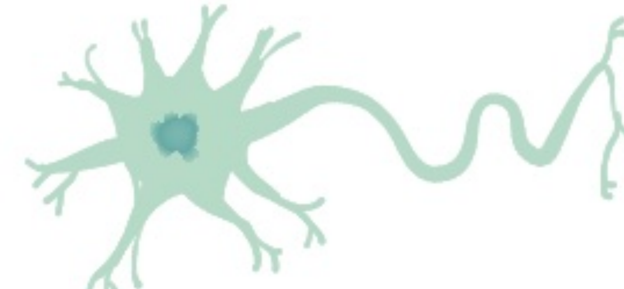
PERCEPTRON

FRANK ROSENBLATT WAS INSPIRED BY MCCULLOCH & PITTS.

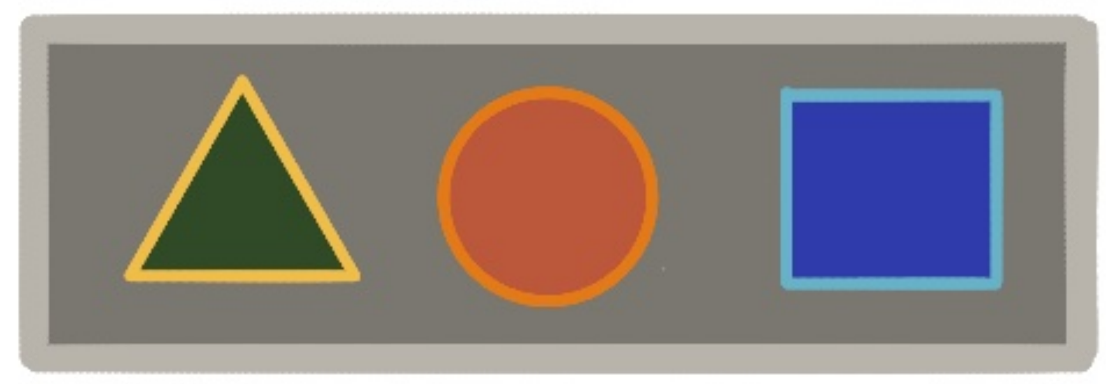
HE DESIGNED A 'NEURON' WITH HARDWARE & ELECTRICALS



FRANK ROSENBLATT



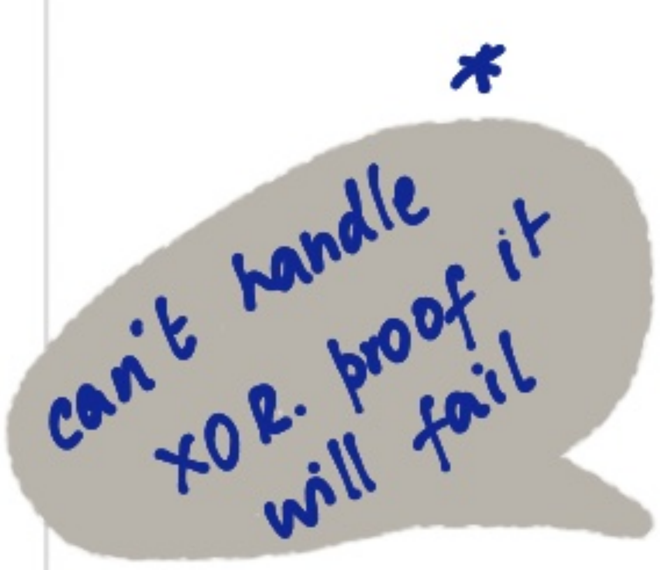
HE CALLED IT A PERCEPTRON (LATE 1950s)

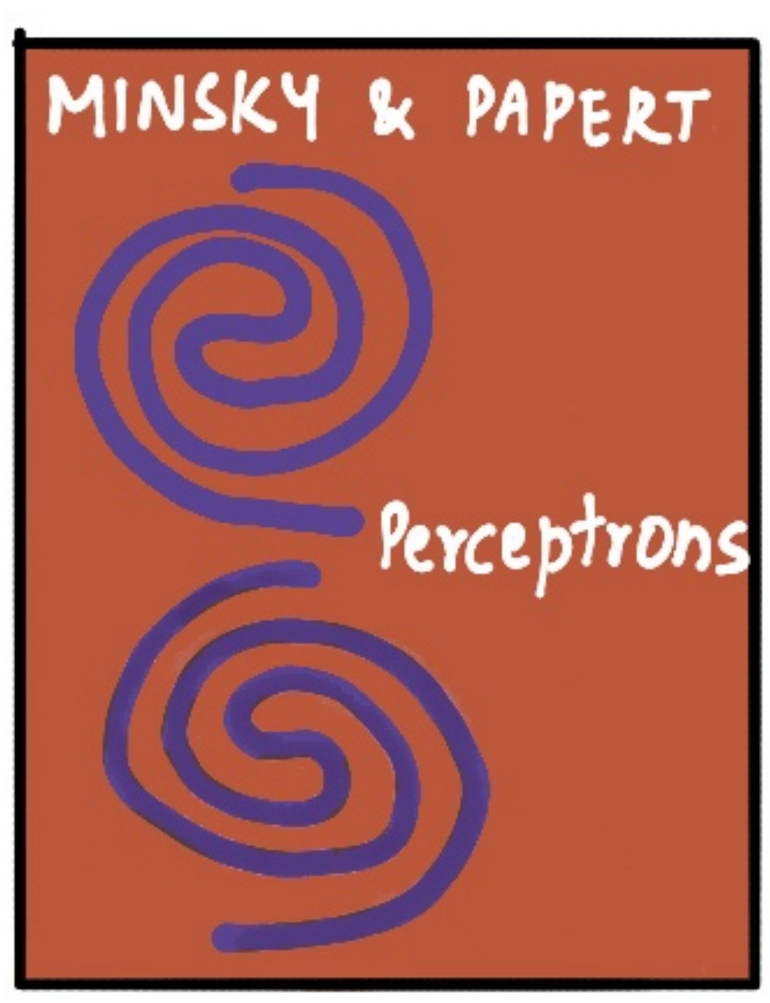


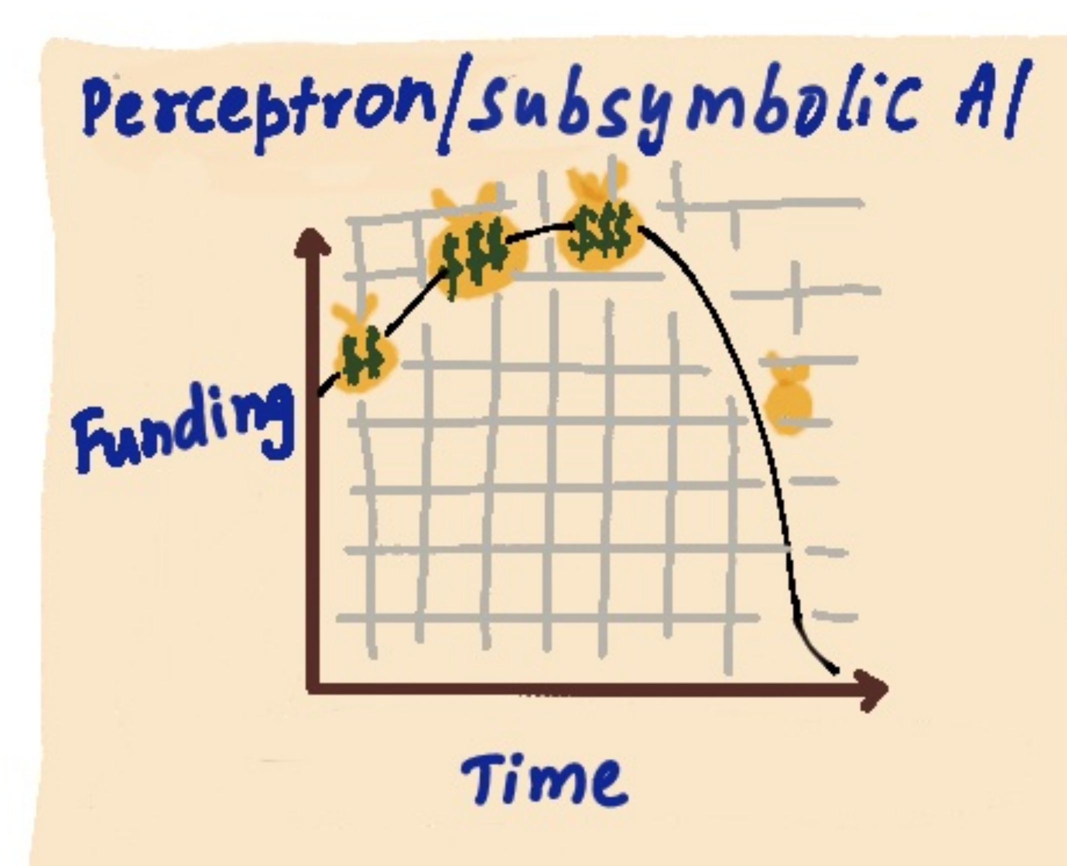
SUBSYMBOLIC AI

HE TRAINED IT TO IDENTIFY SHAPES/LETTERS MUCH LIKE HUMANS SUBCONSCIOUSLY RECOGNISE FACES - (CONTRAST TO THINKING/REASONING TASKS)

AFTER MUCH HYPE SHORTLY FOLLOWED BY THE BELIEF THAT A PERCEPTRON WOULD FAIL AT SIMPLE TASKS, INTEREST WANED.







ROSENBLATT WAS ON THE RIGHT TRACK, THOUGH. HIS APPROACH IS THE ANCESTOR OF ONE OF THE MODERN AI METHODS THAT USES MATHEMATICS - KNOWN AS DEEP NEURAL NETWORKS

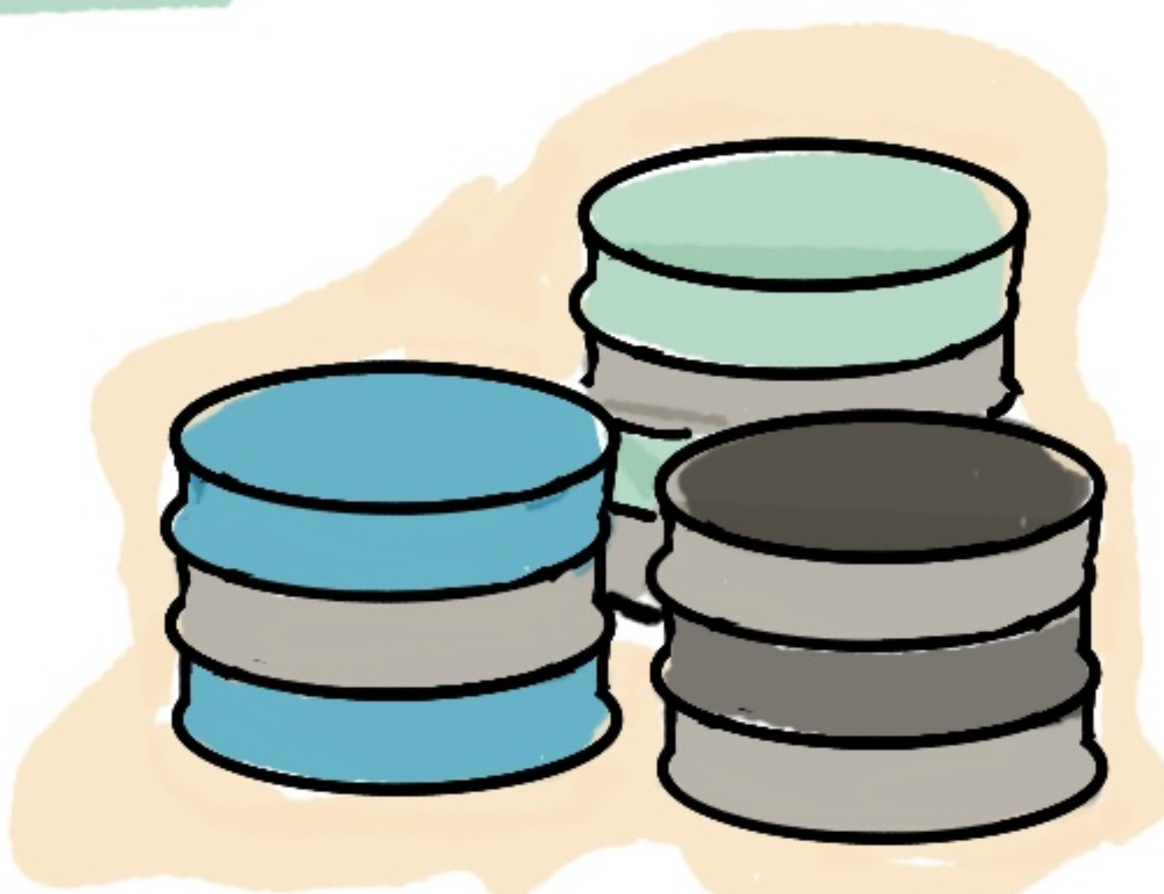
* PAPERT & MINSKY ACKNOWLEDGED (MUCH LATER) THE POTENTIAL OF PERCEPTRONS

WHY AI WORKS SO WELL

AI WENT THROUGH HIGHS OF ACHIEVEMENTS AND LOWS (CALLED AI WINTERS) WITH LACK OF INNOVATION/FUNDING. SYMBOLIC AI DOMINATED - WITH EXPERT SYSTEMS BRINGING REVENUE. ONLY TOWARDS THE 1990s DID AI BOUNCE BACK.

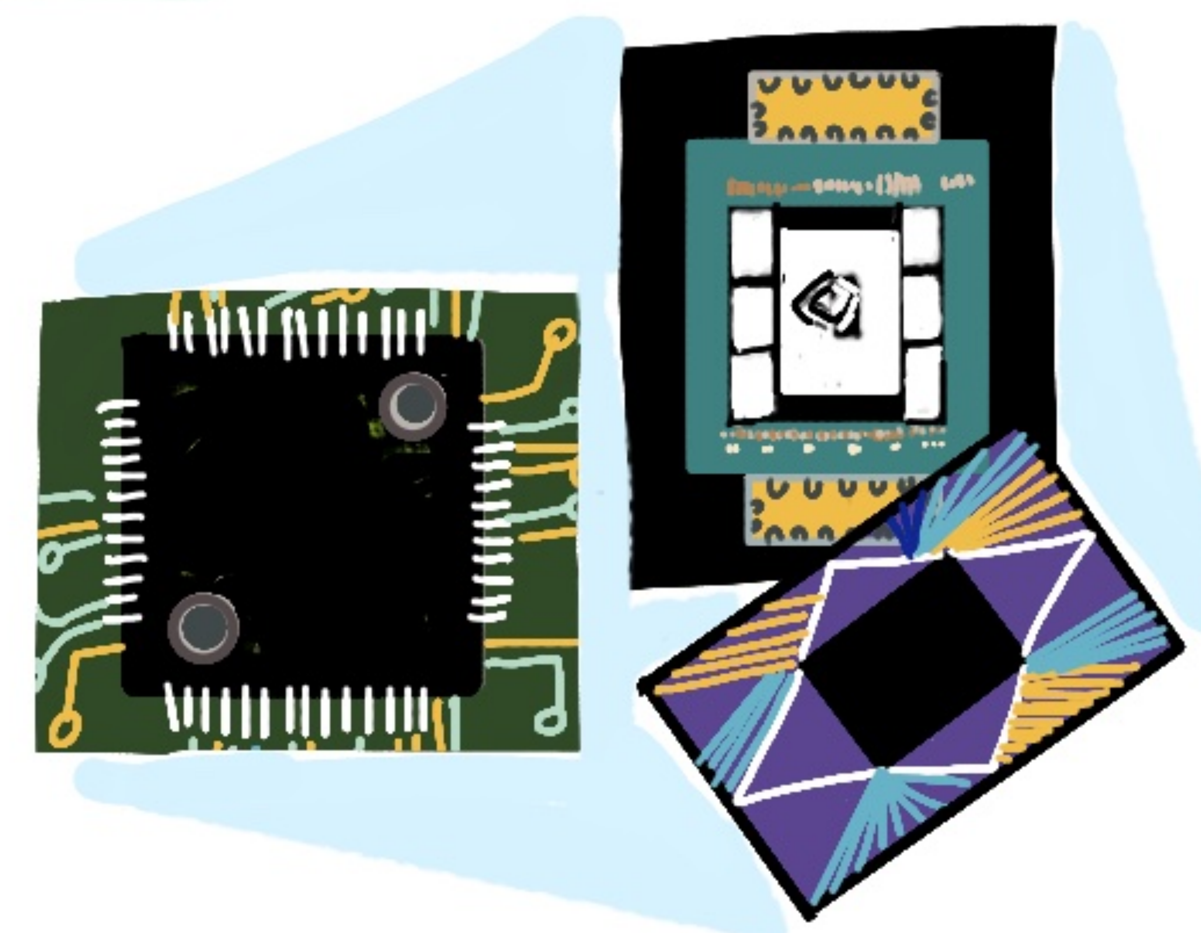
HERE ARE SOME REASONS WHY IT IS NOW A THRIVING FIELD

DATA



MORE DATA AVAILABLE FROM MORE INTERNET USAGE AND TRANSACTIONS

COMPUTE



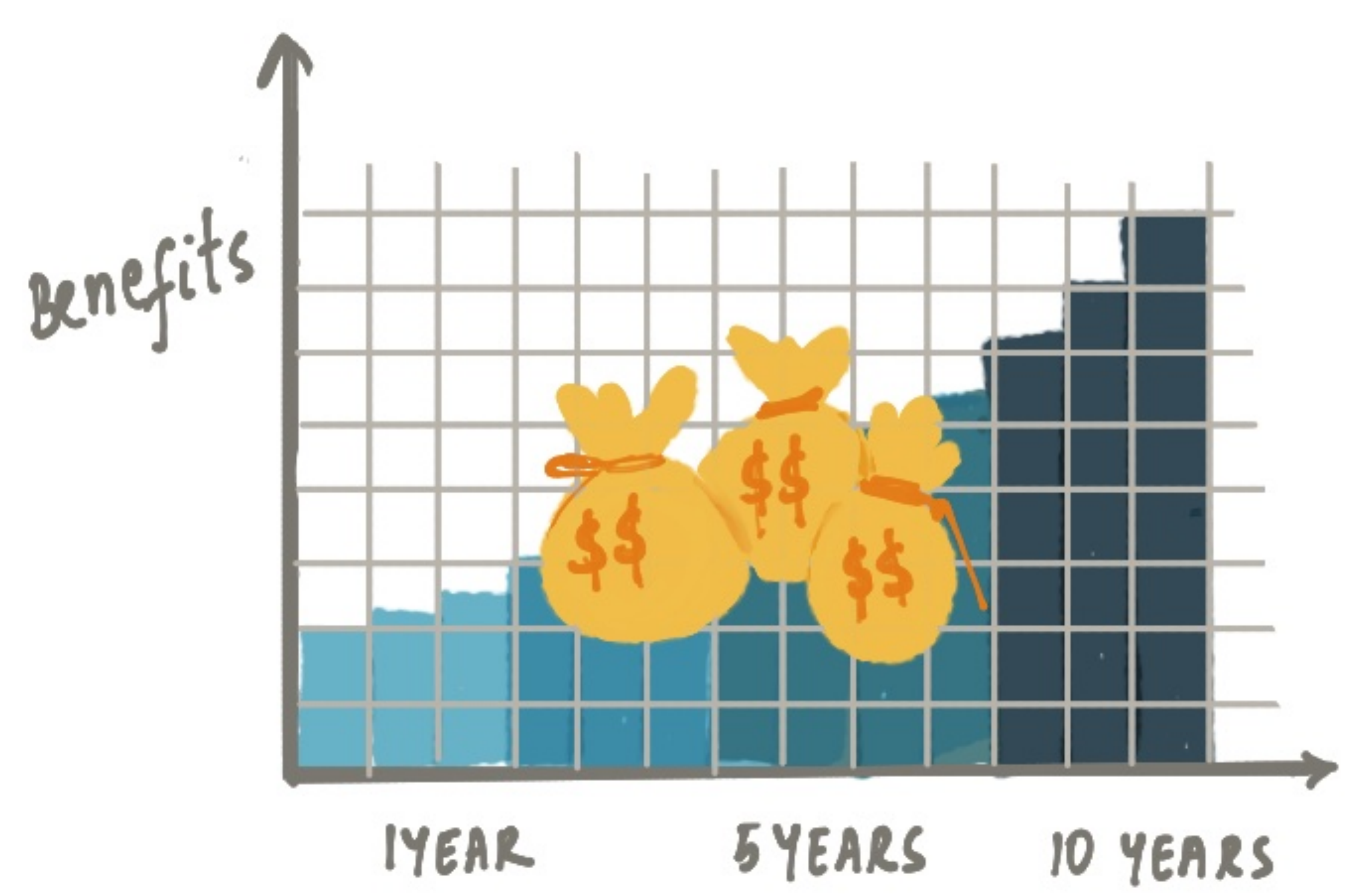
INCREASE IN COMPUTE POWER FROM CPUs, GPUs AND NEUROMORPHIC AND QUANTUM CHIPS

ALGORITHM

- NAIVE BAYES
- BACK PROPAGATION
- K MEANS CLUSTERING
- NEAREST NEIGHBOURS
- REGRESSION
- DECISION TREE
- SUPPORT VECTOR MACHINES
- ... AND MORE!

DEVELOPMENT OF MORE TYPES OF LEARNING ALGORITHMS

INVESTMENT



MORE ATTENTION AND MONEY AVAILABLE AS THE BENEFITS ARE NOW TANGIBLE

HOW TO GET MACHINES TO LEARN

LEARNING TO LEARN

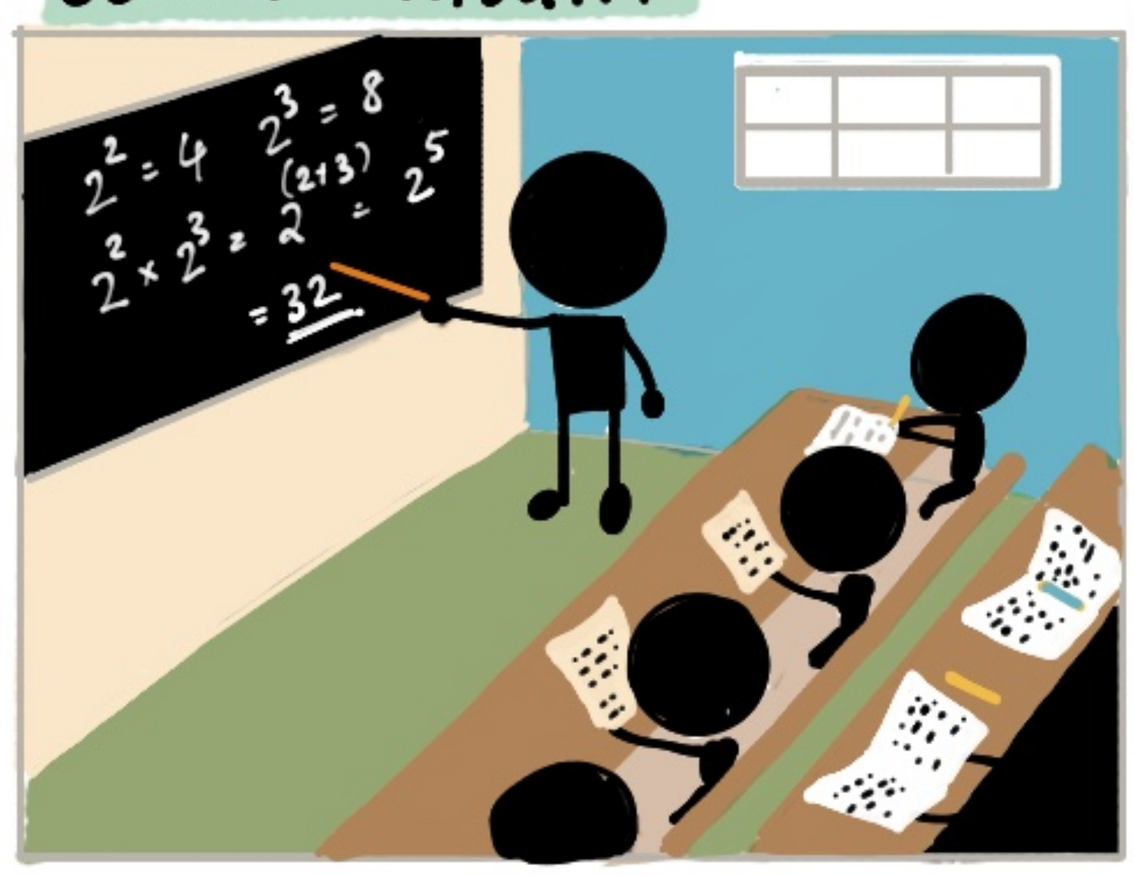
RECALL THAT A GOAL OF AI IS TO BUILD TOOLS FOR HUMANS USING MACHINES THAT CAN TEACH THEMSELVES. EVEN IMPROVE.

THIS ENDEAVOUR, BROADLY, IS **MACHINE LEARNING**

AI SEEKS TO MIMIC HUMAN THOUGHT AND ACTION. HOW IT GETS THERE IS ALSO INSPIRED BY SOME FAMILIAR WAYS

SELF-DISCOVERY

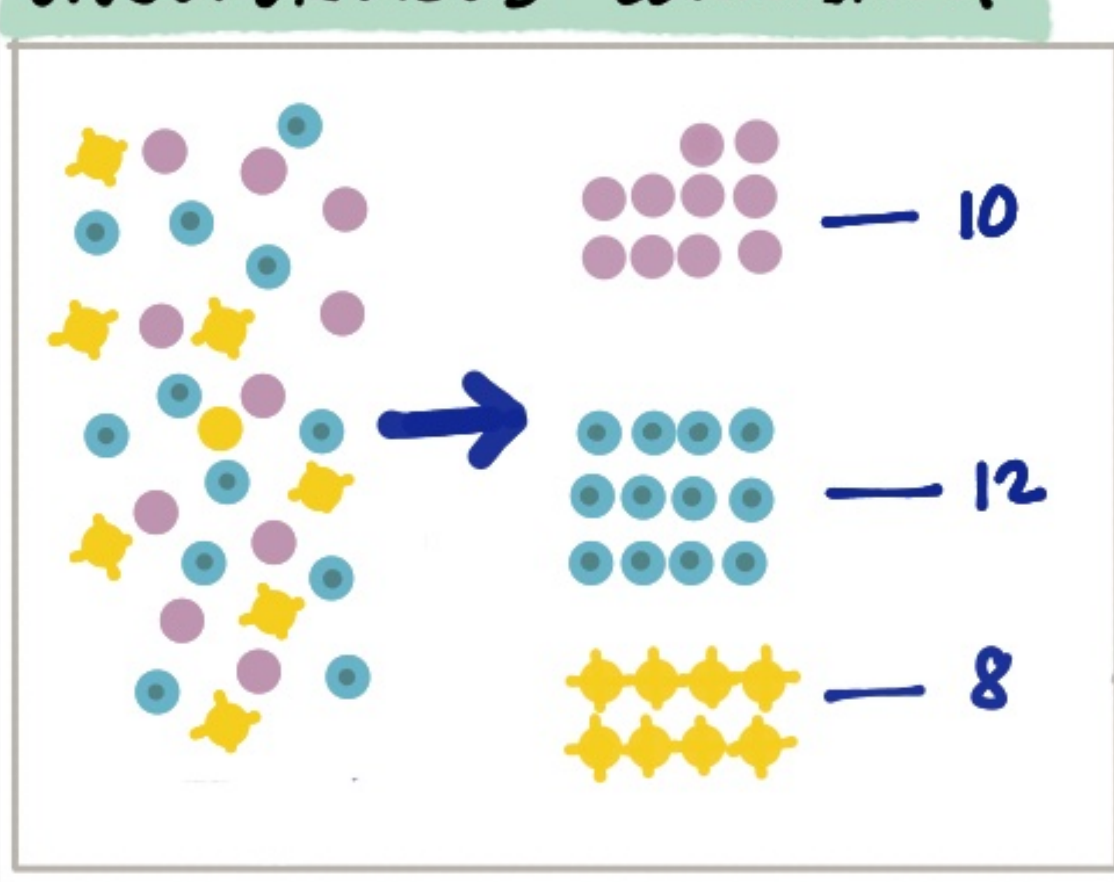
BEING TAUGHT



WITH REWARDS

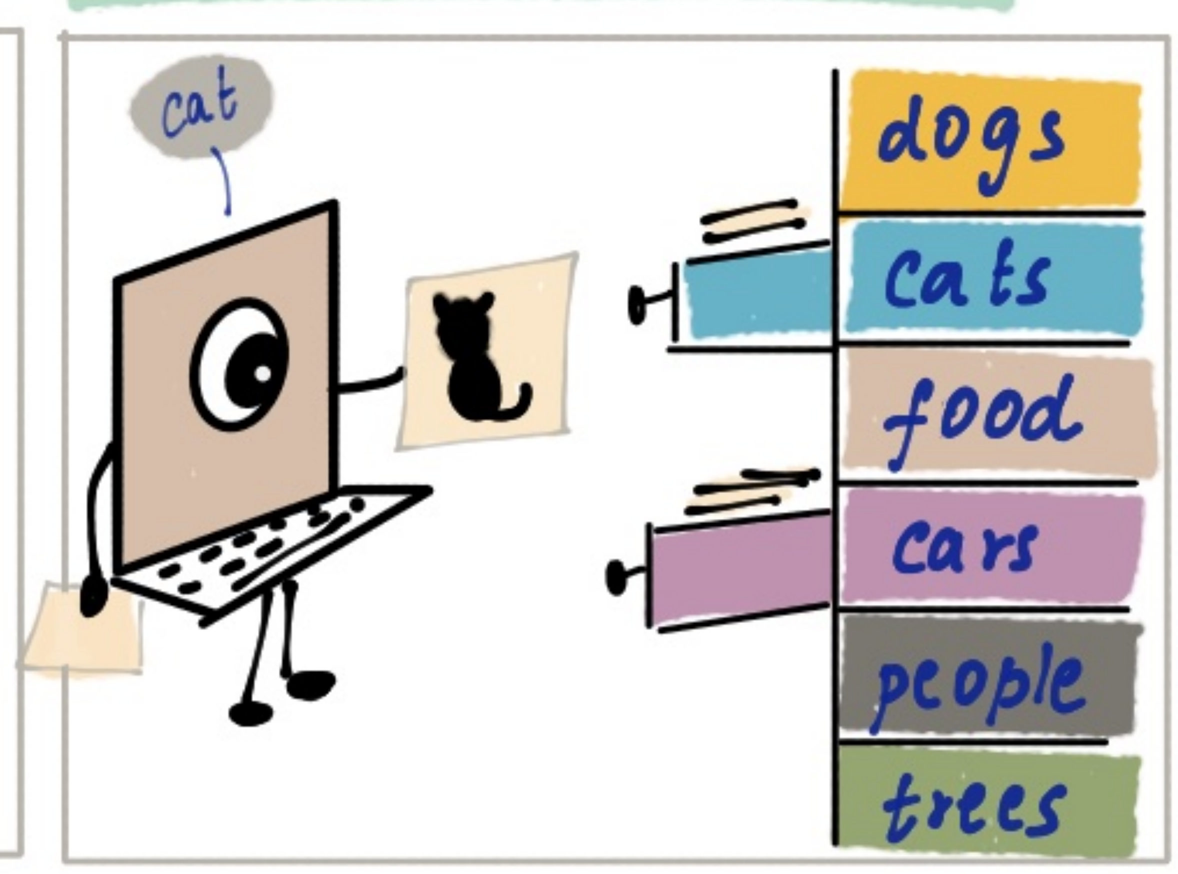
SOME MACHINE LEARNING APPROACHES

* UNSUPERVISED LEARNING



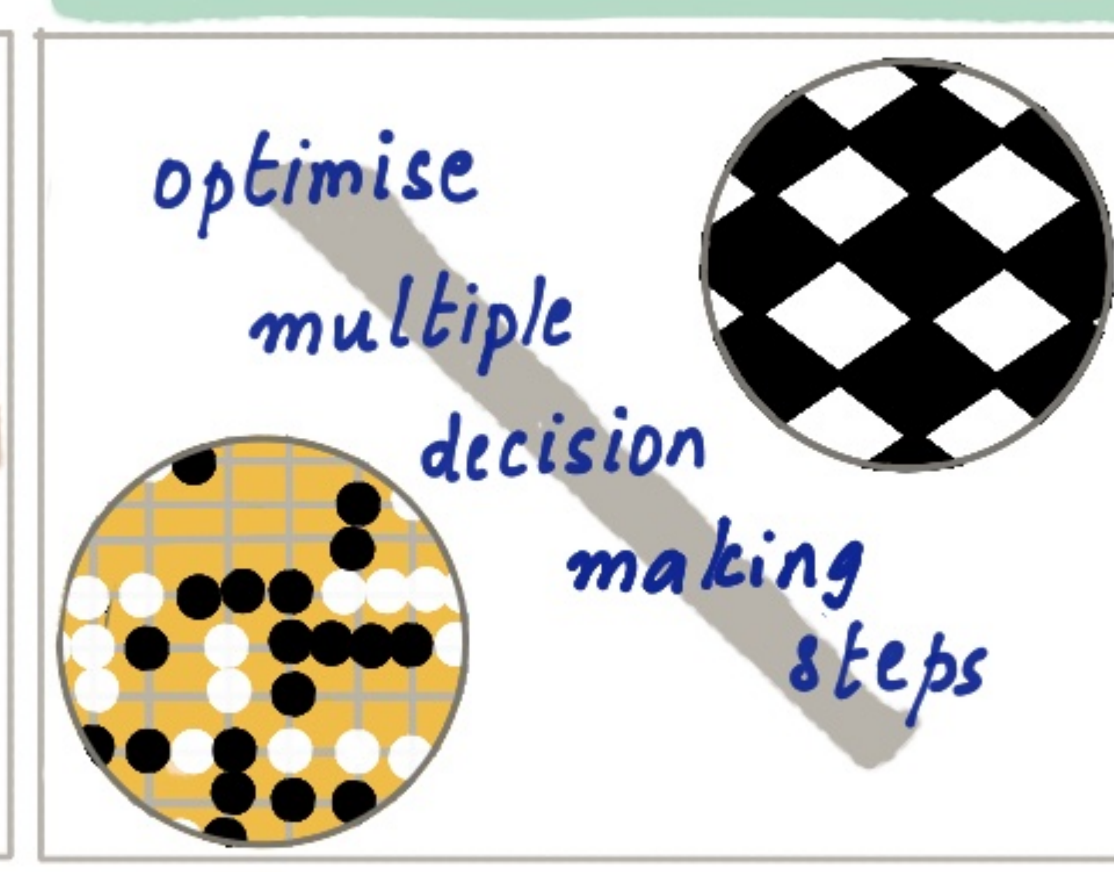
NO KNOWN SOLUTION
+
MAKING INFERENCES

SUPERVISED LEARNING



LEARNING FROM LABELLED DATA

REINFORCEMENT LEARNING



WITHIN CONSTRAINTS
+
REWARD + PUNISHMENT

SEMI SUPERVISED LEARNING

INVOLVES LEARNING FROM A SMALL AMOUNT OF LABELLED DATA AND LOTS OF UNLABELLED DATA.

* Yann Lecunn calls them 'self supervised'

MACHINE LEARNING TRIBES

IN HIS BOOK
'THE MASTER ALGORITHM'
PEDRO DOMINGOS
DESCRIBES

FIVE TRIBES OF MACHINE LEARNING

EACH WITH ITS KEY ALGORITHM — ITS OWN ANSWER TO THE QUESTION 'HOW DO WE LEARN?'

① SYMBOLIST

All humans are mortal
\+ inverse-deduce this bit
⇓
Socrates is a mortal

BASED ON LOGIC & PSYCHOLOGY

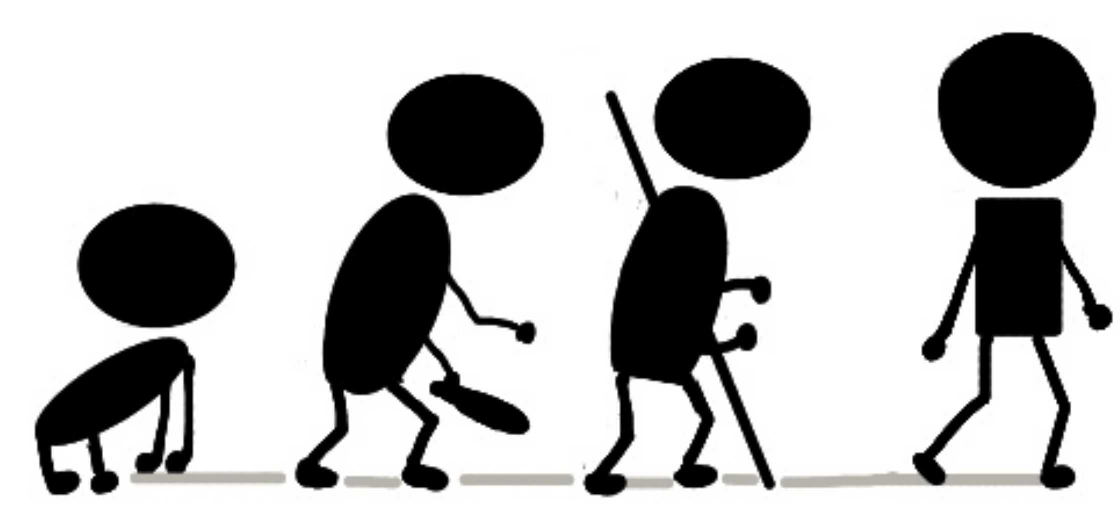
Example - Inverse deduction

② CONNECTIONIST

LOOSELY BASED ON THE SYNAPSES AND THE NETWORK OF NEURONS IN THE BRAIN

Example - Back propagation

③ EVOLUTIONARY



BASED ON EVOLUTIONARY BIOLOGY AND GENETICS

Example - Genetic Algorithms

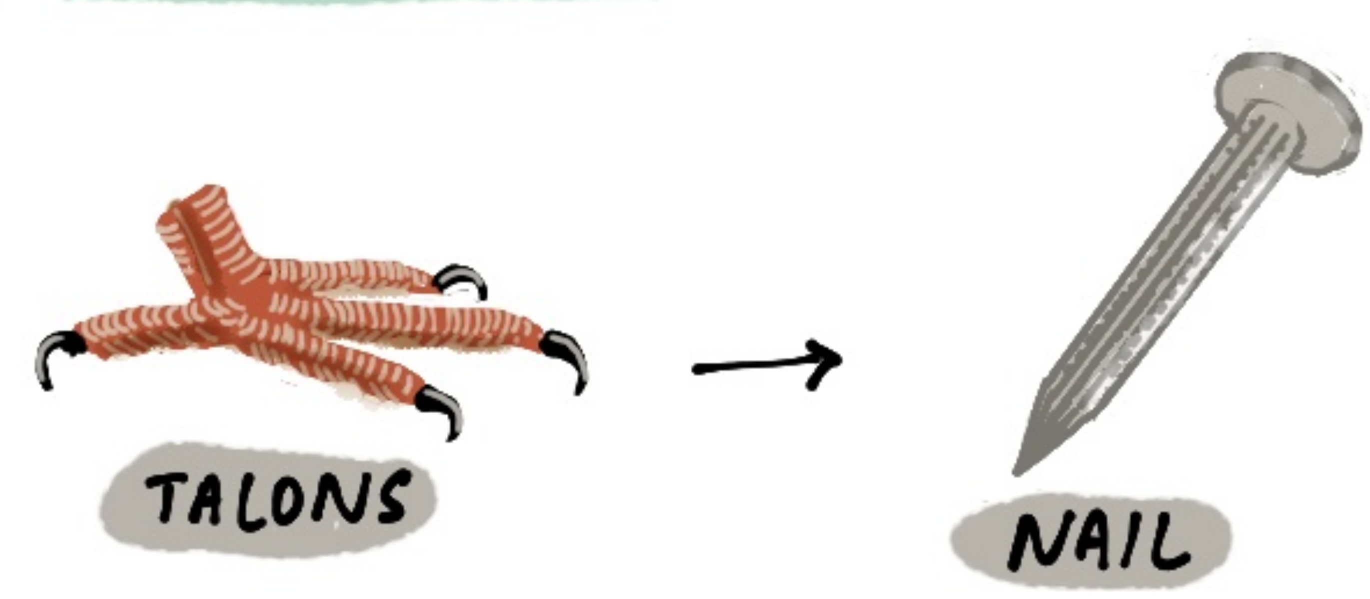
④ BAYESIAN

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

BASED ON UNCERTAINTY AND PROBABILITY AND MAKING INFERENCES USING STATISTICS

Example - Monte carlo methods

⑤ ANALOGISER

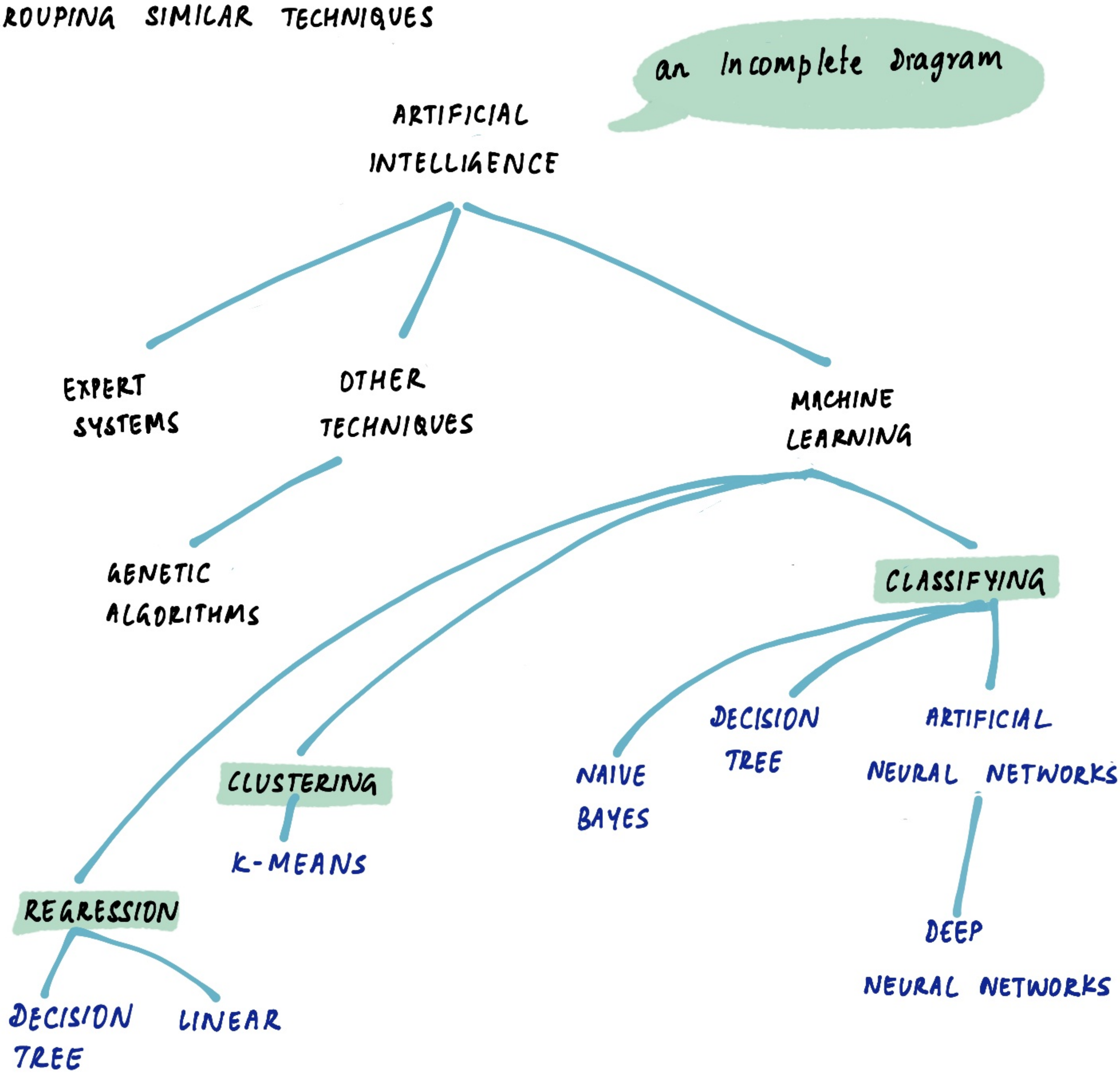


BASED ON FINDING SIMILARITY TO A PREVIOUS EXPERIENCE

Example - Support Vector Machines

AI METHODS

YET ANOTHER WAY TO LOOK AT HOW MACHINES CAN LEARN IS BY GROUPING SIMILAR TECHNIQUES



OBVIOUSLY, THERE ARE MORE TECHNIQUES, EACH WITH MANY ALGORITHMS.

IT HAPPENS THAT ALGORITHMS BECOME SUITED TO SOLVE SPECIFIC TYPES OF PROBLEMS

Source : Becominghuman.ai cheatsheet/ andrew Ng : Coursera

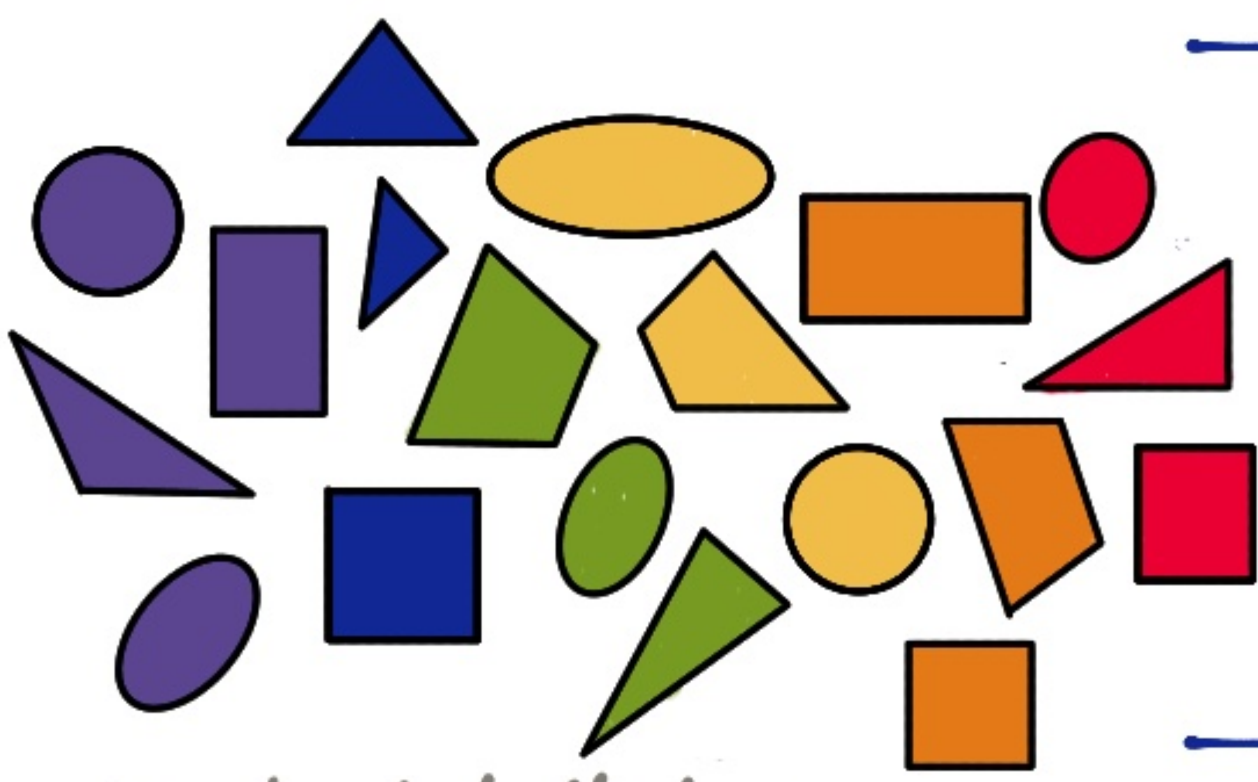
LEARNING ALGORITHMS

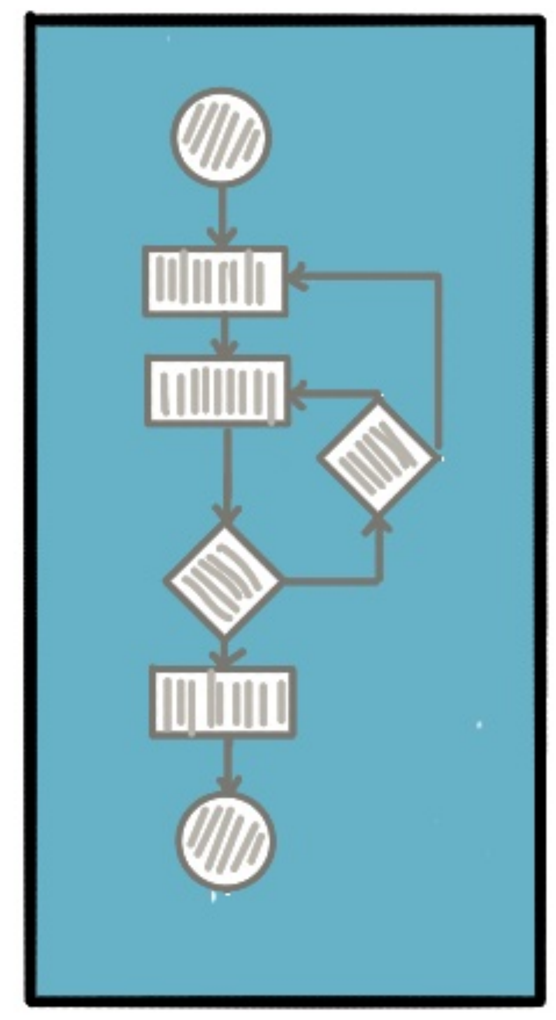
HOW ARE THEY DIFFERENT TO THE OTHER ALGORITHMS WE ENCOUNTER?
WHAT GOES INTO THE MAKING OF THESE?

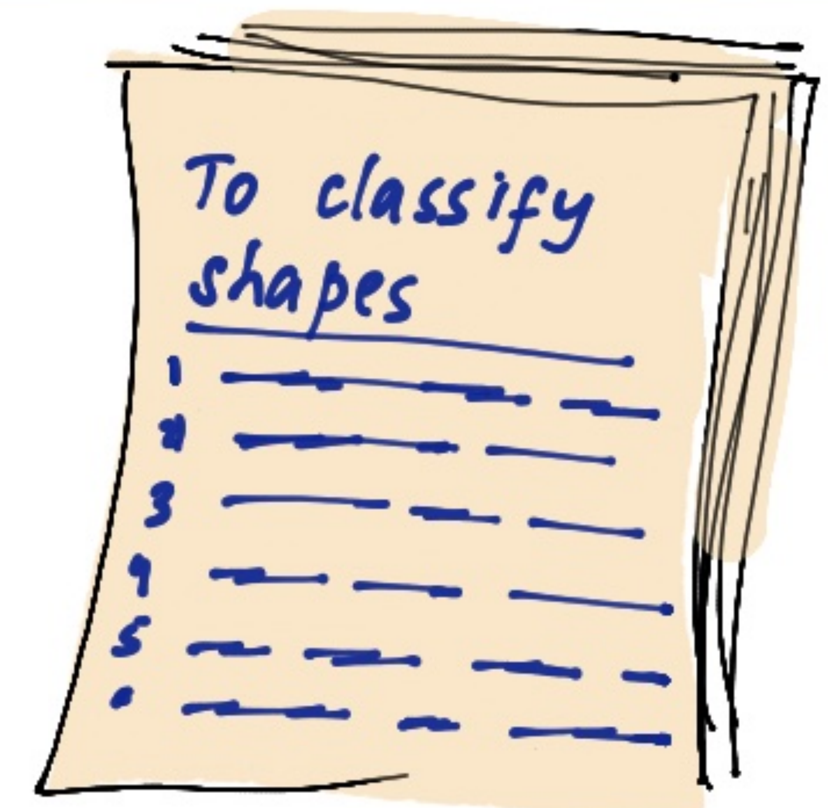
A LEARNING ALGORITHM

A LEARNING ALGORITHM IS DIFFERENT TO THE USUAL ALGORITHMS

TAKES INPUT + ENDGOALS







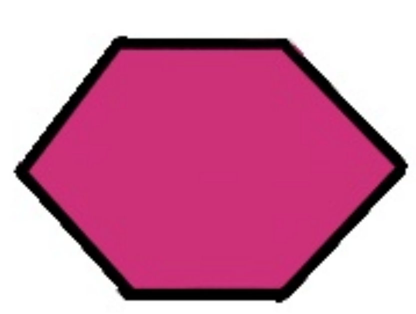
can be labelled or unlabelled input

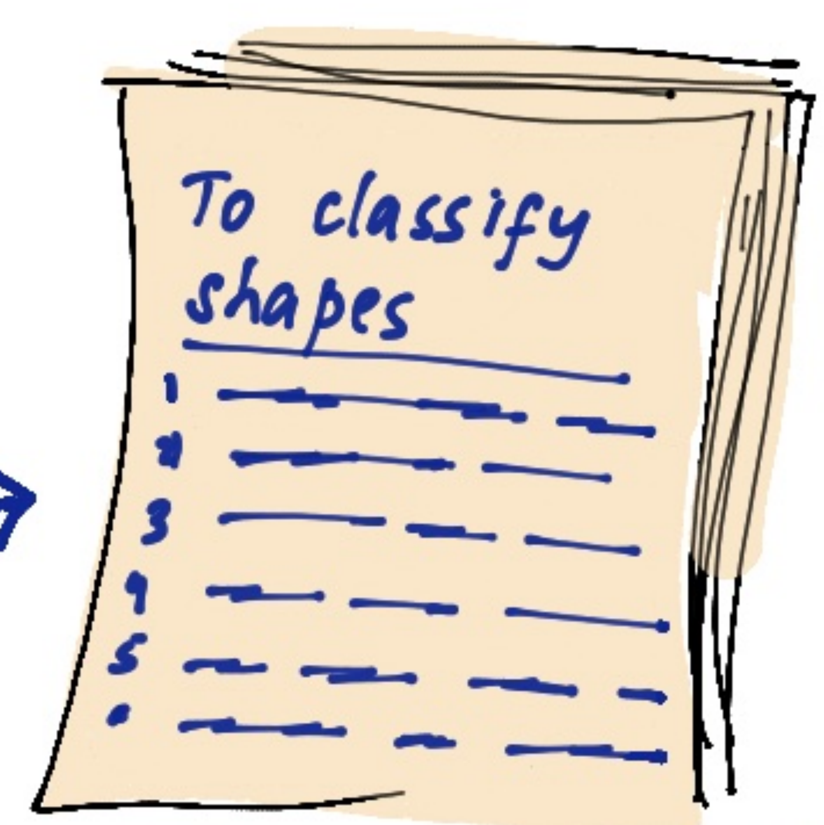
SHAPES + CLASSIFY

OUTPUTS THE 'HEURISTICS'/ SET OF RULES FOR HOW TO ARRIVE AT THE GOAL

USES HEURISTIC TO INTERPRET A NEW INPUT FOR A SIMILAR PROBLEM

UNFAMILIAR INPUT





HEURISTIC/MODEL

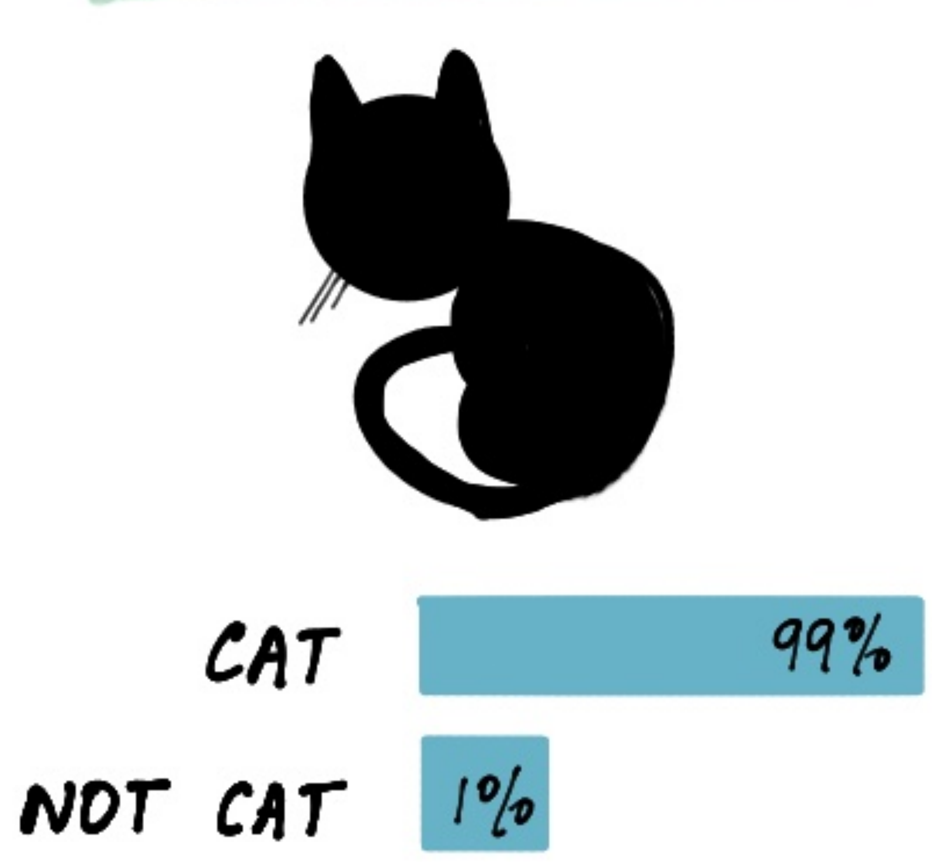
CLASSIFY

A regular 6 sided shape

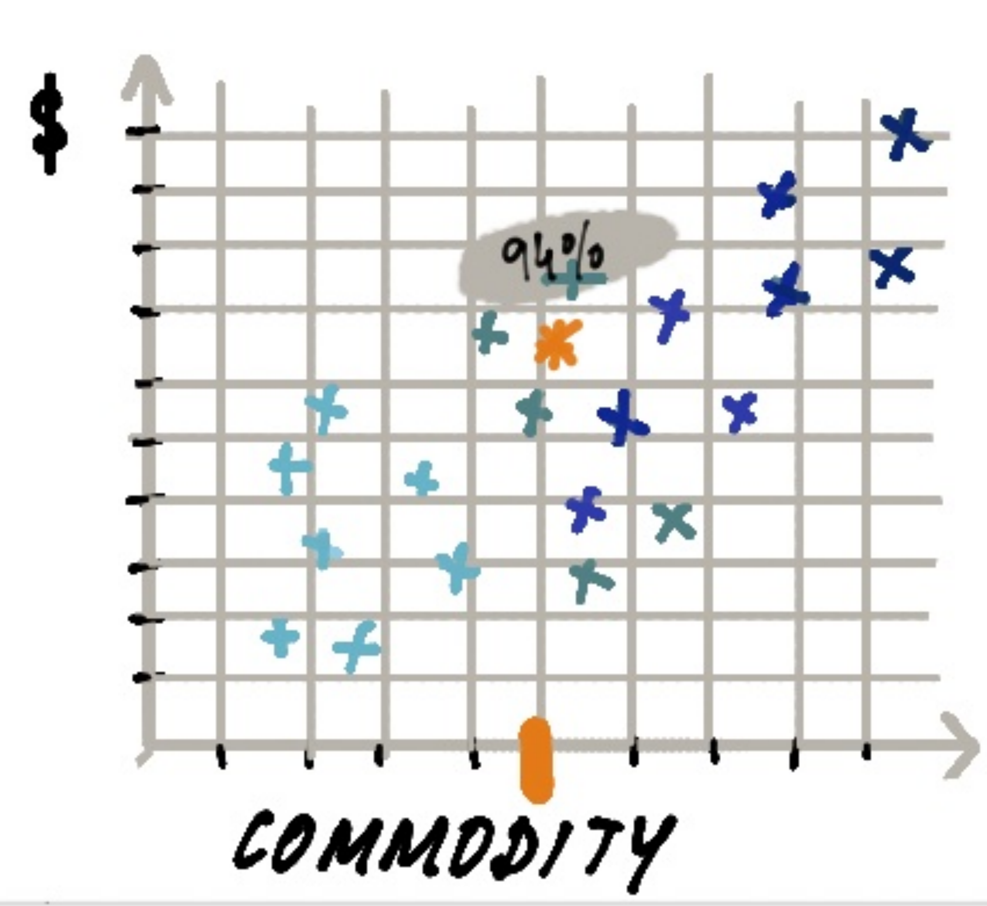
99% accuracy

GIVEN ENOUGH VARIETY AND VOLUME OF INPUTS, THESE

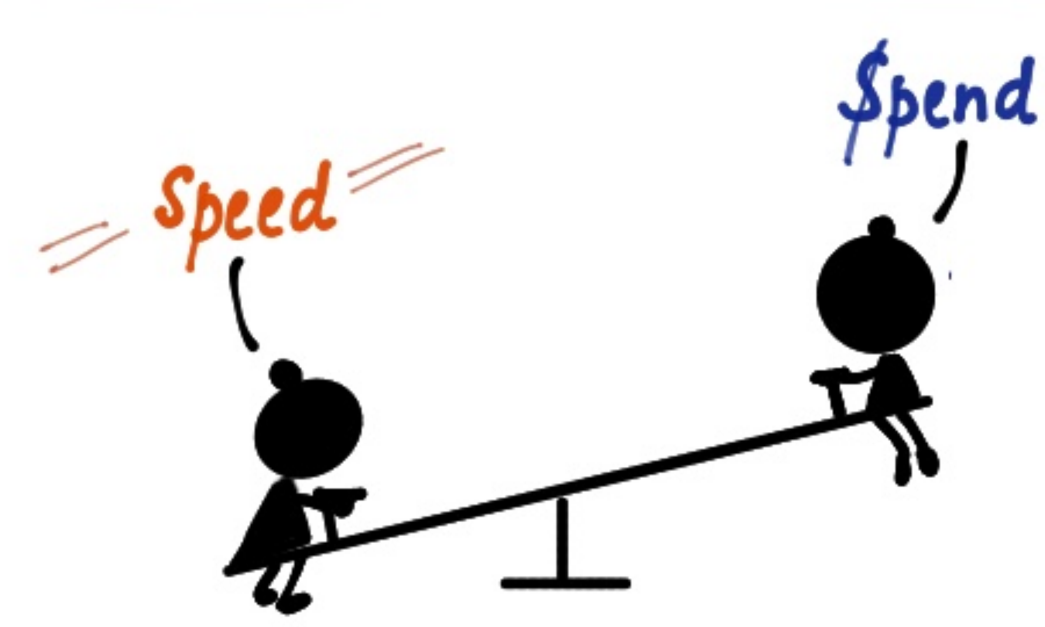
PREDICT CATEGORY



PREDICT VALUE



OPERATE WITHIN CONSTRAINTS



OPTIMISE AND Self-correct

SOME EXAMPLES

- PREDICT A VALUE
- CLASSIFY AN OBJECT

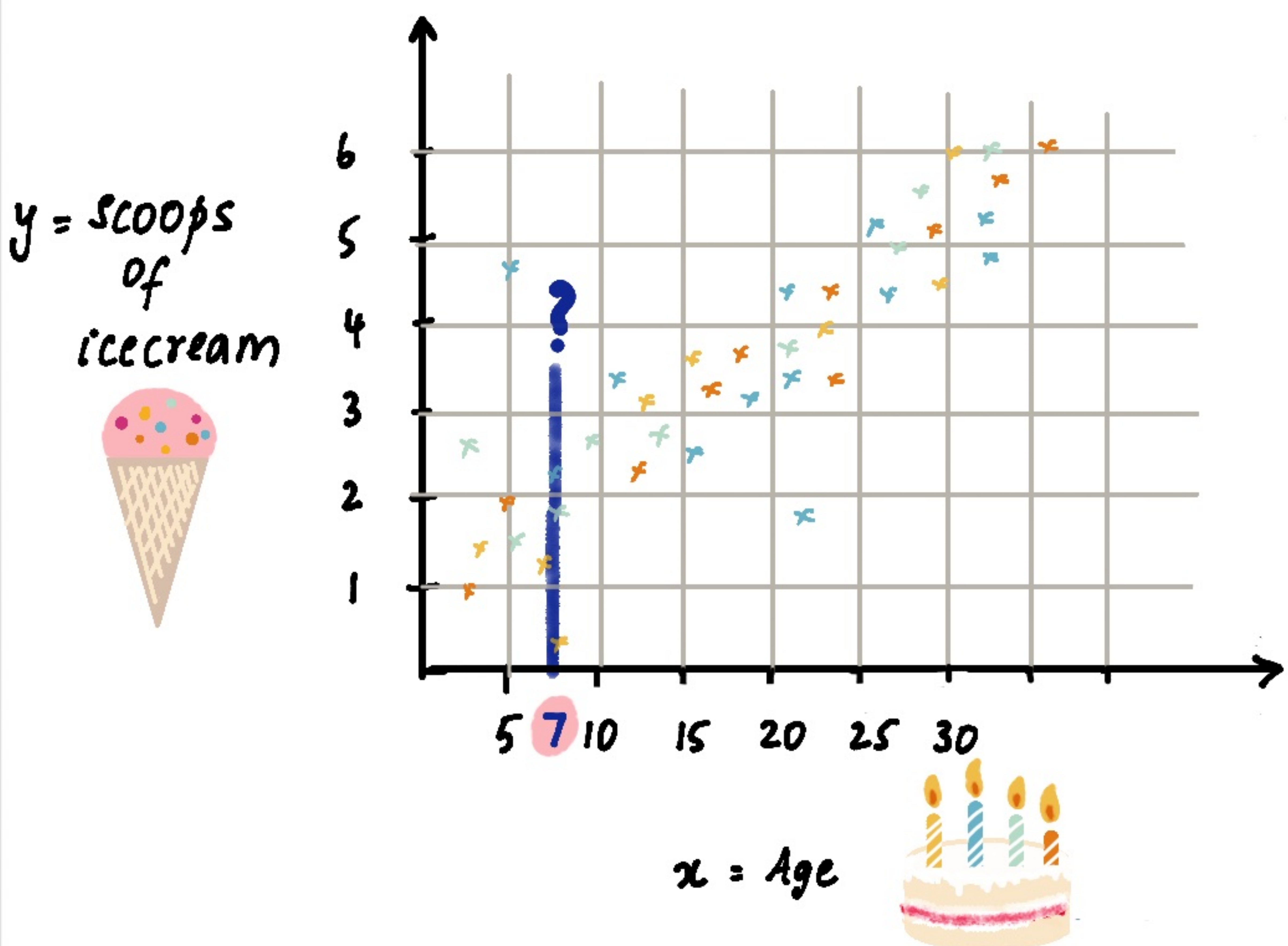
— USING —

- LINEAR REGRESSION
- K NEAREST NEIGHBOURS
- DECISION TREES
- NEURAL NETWORKS

LINEAR REGRESSION

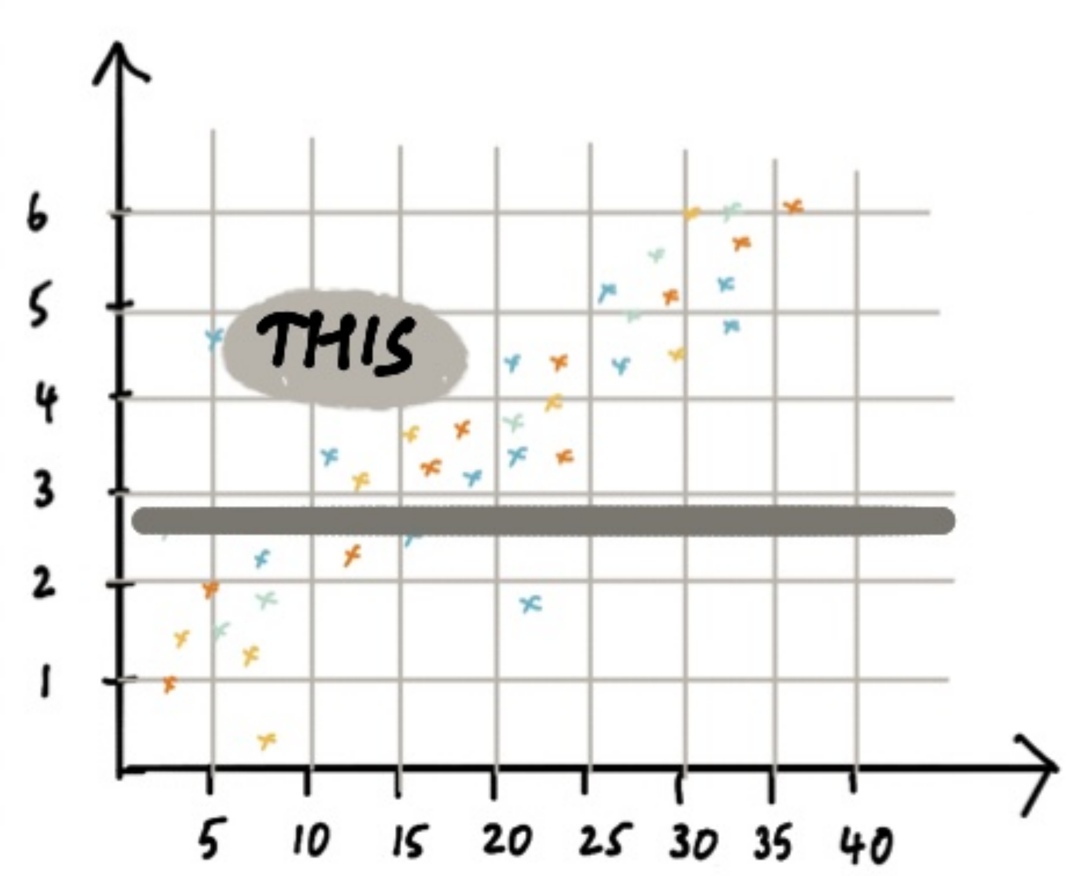
THIS ALGORITHM IS USED TO MAKE A PREDICTION OF LIKELIHOOD. SAY, WE NEED TO PREDICT HOW MUCH ICECREAM A 7 YEAR OLD IS LIKELY TO EAT

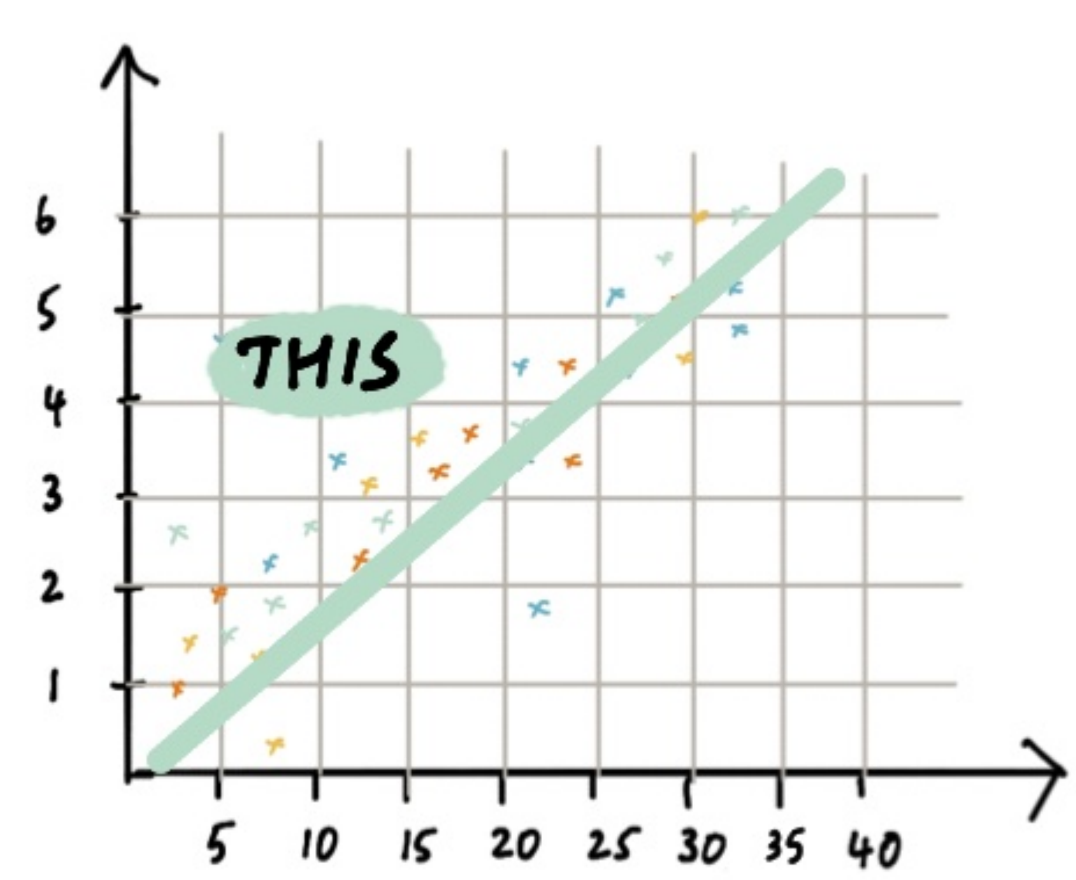
THIS FICTITIOUS TRAINING DATA SHOWS ICECREAM CONSUMED BY AGE GROUPS

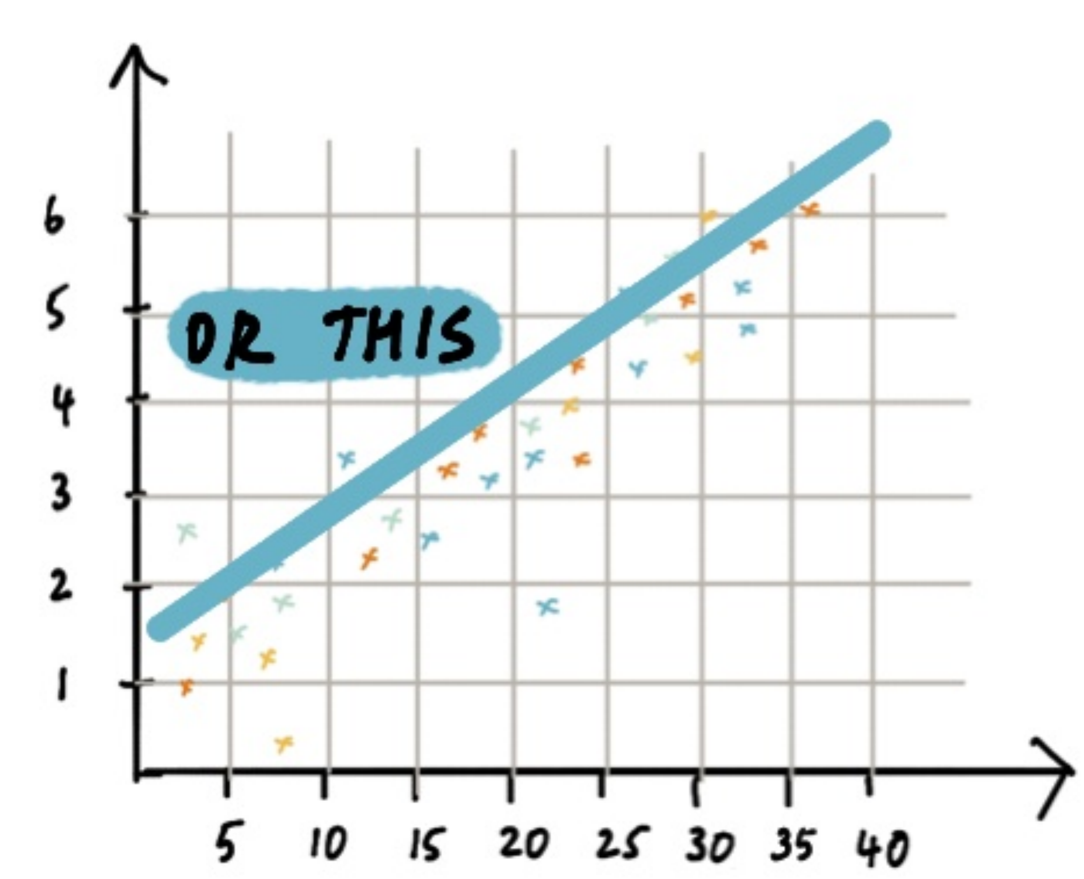


Note:
For simplicity, we consider only one variable 'Age' as influencing the number of scoops of icecreams enjoyed. We assume that 'Hunger' 'sweet-toothedness,' 'Season' are not factors that are relevant for this example

THE MACHINE LEARNING MODEL WILL NEED TO ESTABLISH A 'LINE-LIKE' RELATIONSHIP BETWEEN AGE AND SCOOPS CONSUMED. IT MIGHT LOOK LIKE...







PARAMETERS TO VARY

- WHERE ON THE Y AXIS THE LINE STARTS
- HOW MUCH IT SLOPES

THE WINNING MODEL WOULD BE ONE THAT MINIMISES ERRORS BETWEEN ACTUAL AND PREDICTED SCOOPS OF ICECREAM

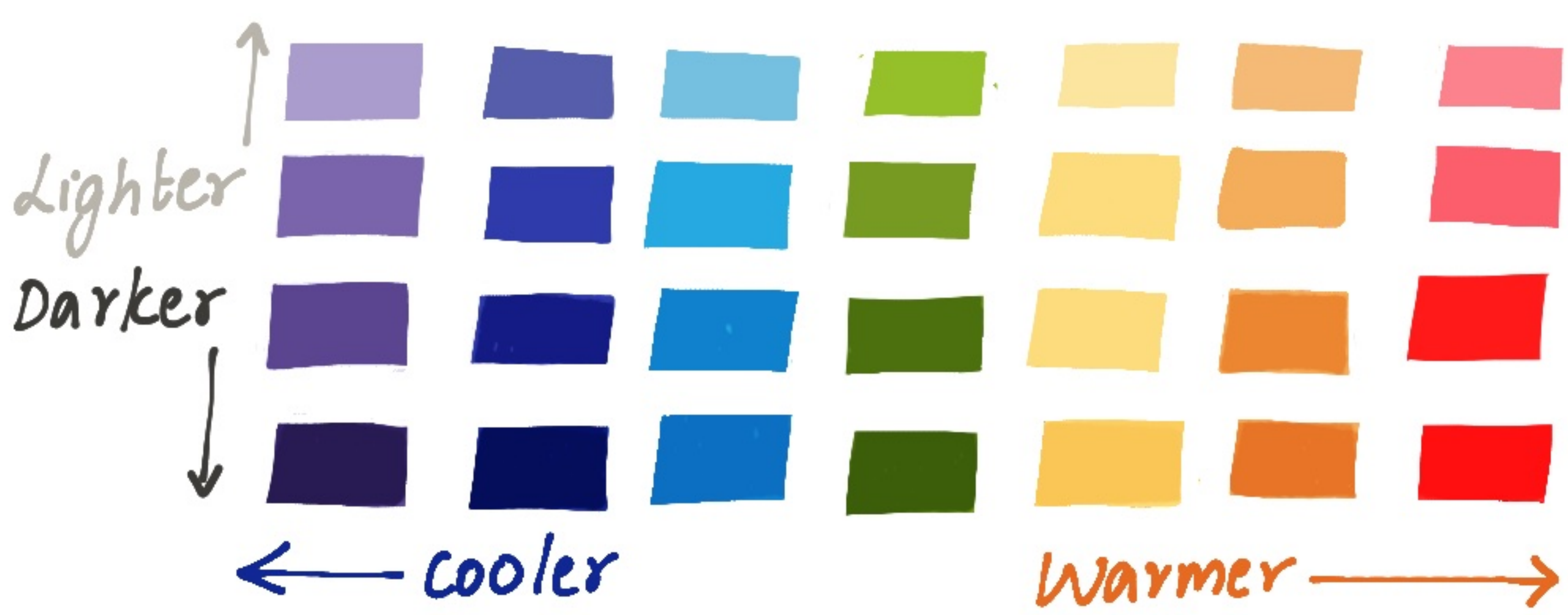
K NEAREST NEIGHBOUR

THIS ALGORITHM GETS USED TO MAKE PREDICTIONS, CLASSIFY & SEARCH.

IT WORKS ON THE ASSUMPTION THAT PROXIMITY = SIMILARITY

FOR EXAMPLE, TO PREDICT WHAT COLOUR IS UNDERNEATH THE GREY PATCH

THE COLOURS APPEAR TO GO COOL → WARM AND LIGHT → DARK



LOOKING AT ITS NEAREST NEIGHBOURING COLOURS, THE ? SHOULD BE ONE OF THE WARMER, DARKER COLOURS AND LIKELY

- NEARNESS IS CALCULATED FROM ONE OF MANY STANDARD METHODS
- K IS HOW MANY NEAREST NEIGHBOURS TO CONSIDER

FOR **NUMERICAL PREDICTIONS**

THE ALGORITHM CONSIDERS THE AVERAGE OF THE NEAREST K VALUES

FOR **CLASSIFYING**

THE RESULT WILL BE THE MOST FREQUENTLY OCCURRING FEATURE/EXAMPLE

DECISION TREES

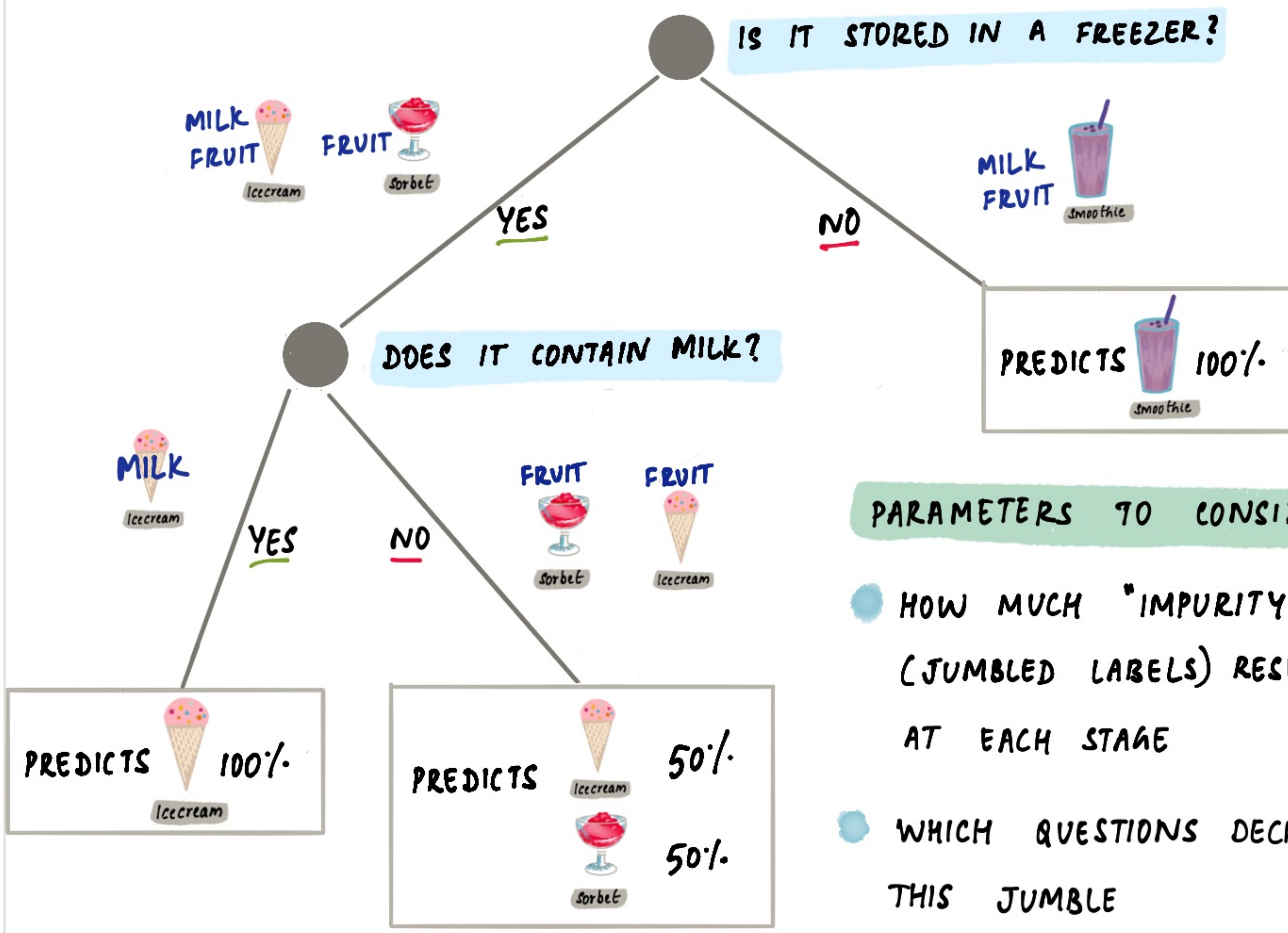
THIS ALGORITHM IS USED FOR BOTH CLASSIFICATION & REGRESSION

IT LEARNS RULES BASED ON FEATURES OF THE DATA TO MAKE DECISIONS

LET'S SAY WE ARE TRYING TO TELL APART 3 COLD DESSERTS

CONTAINS	STORED IN	CLASS
MILK	FREEZER	ICECREAM
FRUIT	FREEZER	SORBET
FRUIT	FREEZER	ICECREAM
MILK	FRIDGE	SMOOTHIE
FRUIT	FRIDGE	SMOOTHIE

THE GOAL IS TO LEARN TO ASK A SERIES OF YES/NO QUESTIONS SO AT EACH STAGE WE GET AS MANY OF THE SAME CLASS AS POSSIBLE



PARAMETERS TO CONSIDER

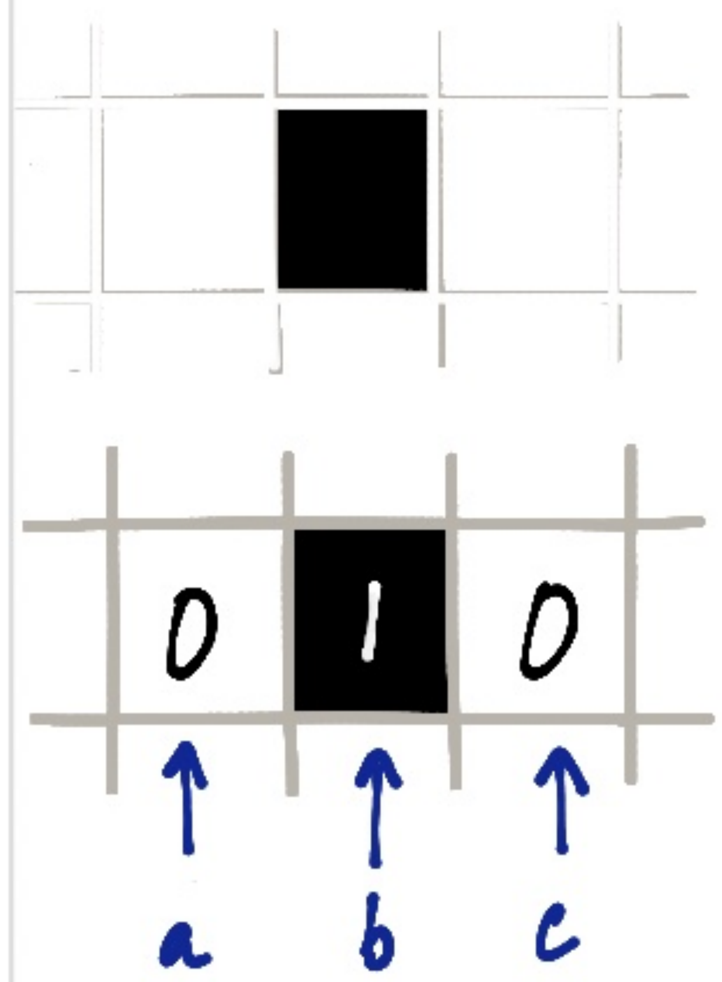
- HOW MUCH "IMPURITY" (JUMBLED LABELS) RESULTS AT EACH STAGE
- WHICH QUESTIONS DECREASE THIS JUMBLE

Decision trees are combined to form RANDOM FORESTS used in classification

NEURAL NETWORKS - 1

NEURAL NETWORKS ARE VERY LOOSELY MODELLED ON HOW THE NEURONS IN OUR OWN BRAIN WORK. THEY CAN BE USED, AMONG OTHER THINGS, TO CLASSIFY OBJECTS. LET US STEP THROUGH A FEW BUILDING BLOCKS TO UNDERSTAND THEM.

A SIMPLE PERCEPTRON CAN DETECT A DOT ON A PLAIN SURFACE.
(Frank Rosenblatt's artificial neuron from the 1950s)



CONVERT THE IMAGE INTO LITTLE SQUARES

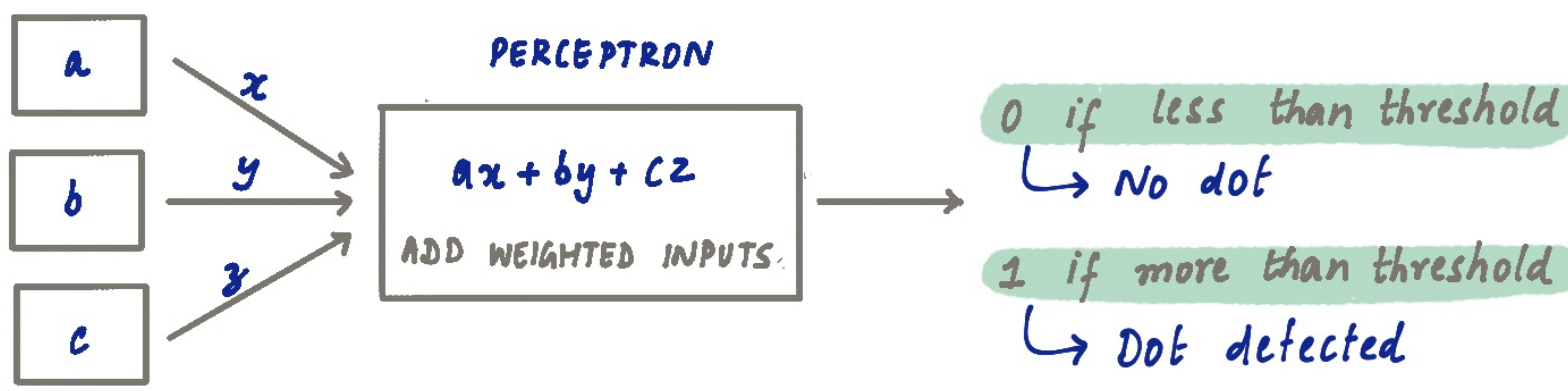
DARKER SQUARES HAVE HIGHER INTENSITY

TURN COLOURS INTO NUMBERS

- WHITE = 0
- BLACK = 1
- GREY = ALL NUMBERS IN BETWEEN

THUS DOTS, EDGES AND SHAPES ARE IDENTIFIED BY A SHARPER CONTRAST TO NEIGHBOURING SQUARES.

THE INPUTS a, b, c ARE GIVEN RANDOM WEIGHTS x, y, z

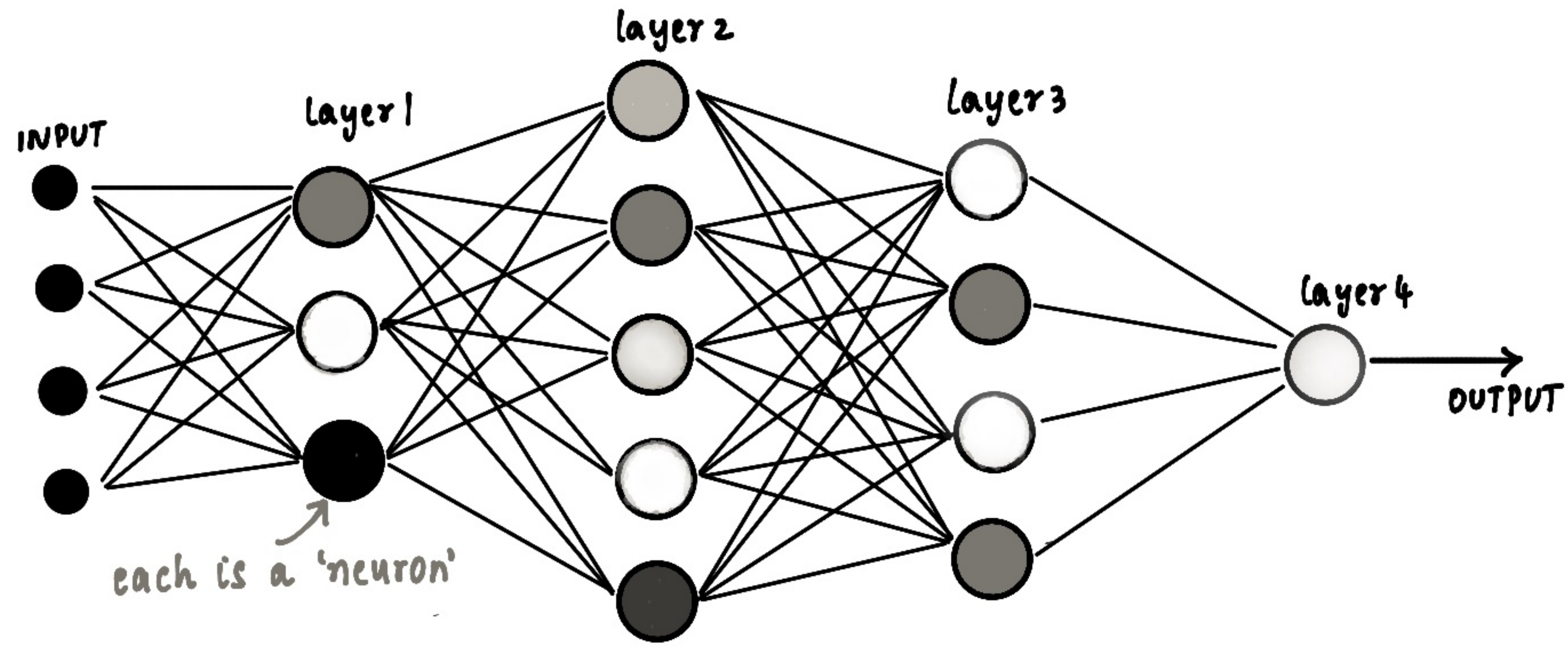


ALTHOUGH EXTRA LAYERS OF PERCEPTRONS COULD DO MORE, HERE ARE SOME REASONS WHY FURTHUR RESEARCH DRIED UP :

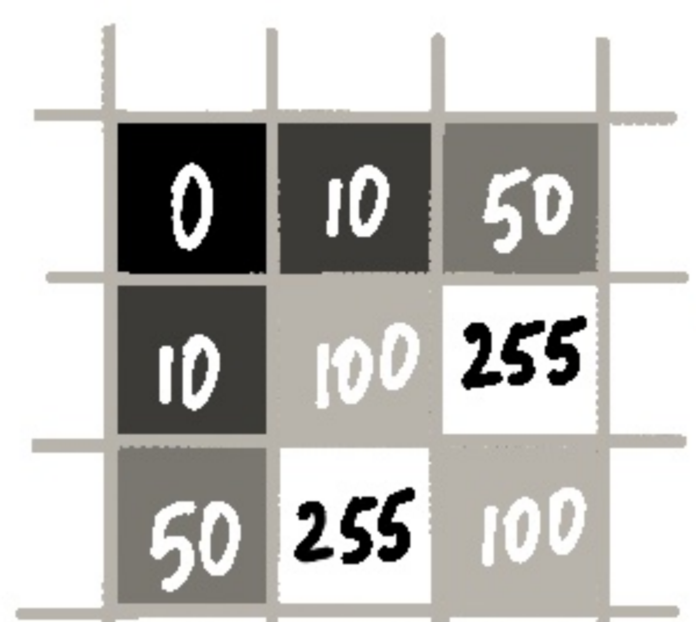
- NO KNOWN MULTILAYER ALGORITHMS
- DIFFICULT TO TRAIN *THIS* MANY WEIGHTS
- CANNOT FORM MEANINGFUL RULES OUT OF THE WEIGHTS

NEURAL NETWORKS – II

A MULTI-LAYER (DEEP) ARTIFICIAL NEURAL NETWORK MIGHT LOOK LIKE THIS



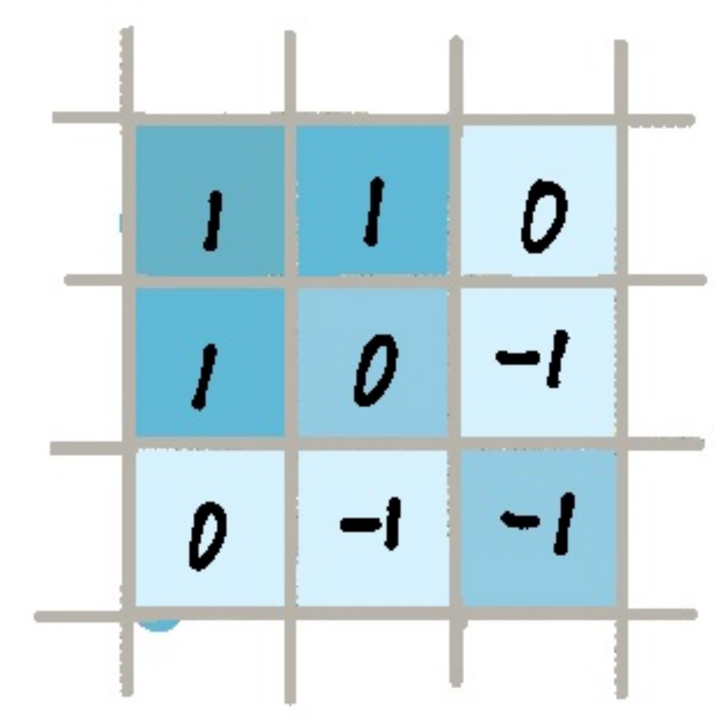
EACH NEURON GETS AN INPUT FROM EVERY MEMBER IN THE PREVIOUS LAYER



Pixel Values

INPUTS COME FROM (e.g.) PIXEL VALUES 0-255

CONNECTIONS TO NEURONS ARE RANDOMLY WEIGHTED



Weights

NEURONS ADD UP WEIGHTED INPUTS + THRESHOLDS

...& CONVERT THIS SUM TO A NUMBER BETWEEN 0-1 BY

APPROXIMATING { HIGH POSITIVES → CLOSER TO 1 → BRIGHTER PIXELS
HIGH NEGATIVES → CLOSER TO 0 → DULLER PIXELS

THIS IS INPUT TO THE NEXT LAYER AND SO ON.

THE OUTPUT OF THE FINAL LAYER IS A CONFIDENCE SCORE OF WHAT IT THINKS THE OBJECT IS. e.g circle or not

REPEAT THIS FOR EVERY LABELLED TRAINING EXAMPLE

NEURAL NETWORKS – III

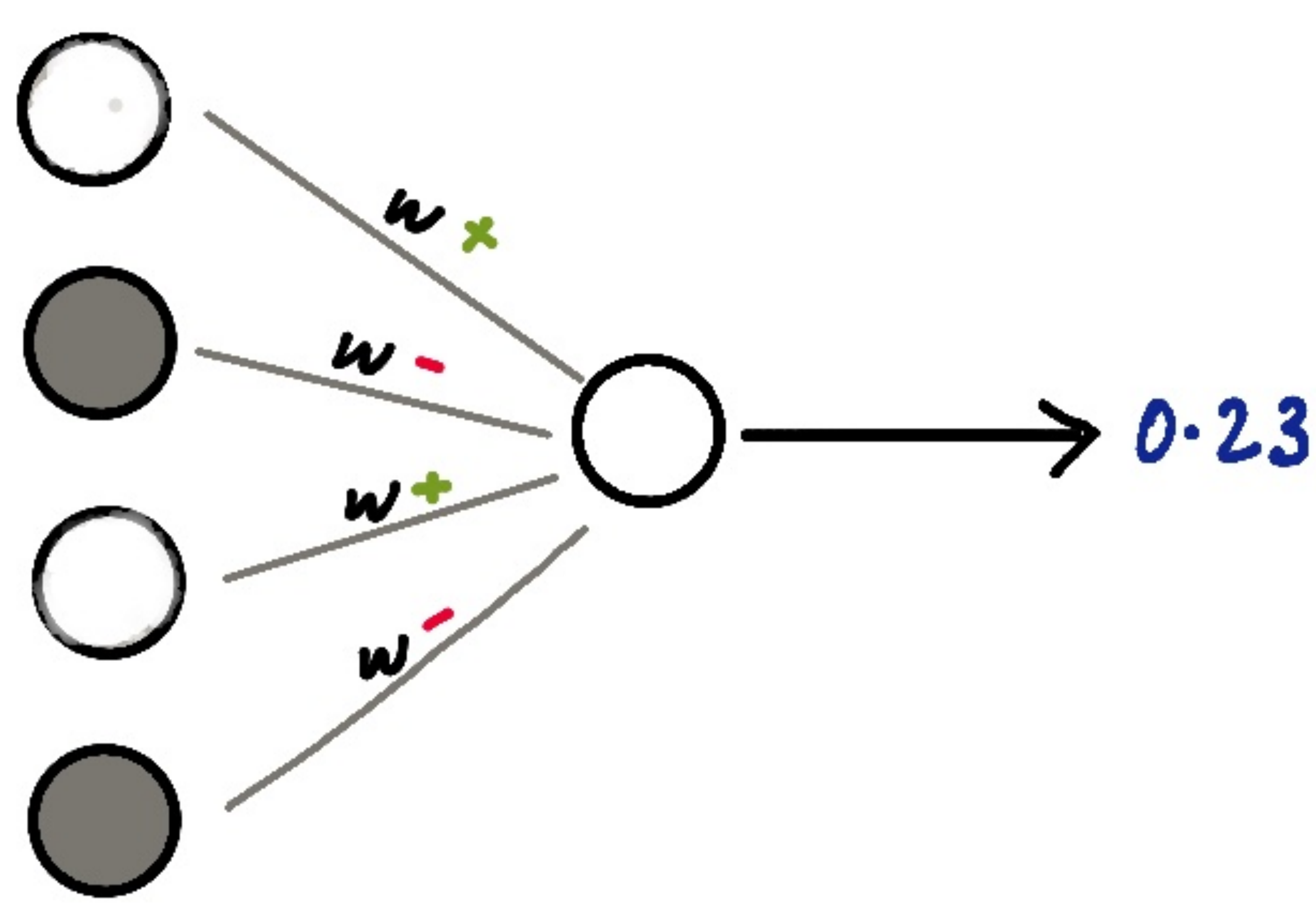
OBVIOUSLY, THE CONFIDENCE SCORE/PREDICTION IS NOT GOING TO BE ACCURATE.

Shape	Predicted	Actual
○	0.23	1
□	0.49	0
○	0.9	1
△	0.02	0
⊔	0.5	0
⬭	0.7	0

THE OBJECTIVE OF THE NEURAL NET IS NOW TO MINIMISE ERRORS FOR ALL TRAINING EXAMPLES
THIS PROCESS IS 'LEARNING'

THIS IS DONE BY TWEAKING WEIGHTS AND THRESHOLDS FOR EVERY NEURON THROUGH AN ALGORITHM CALLED BACKPROPAGATION

TAKE A CIRCLE MARKED 0.23 WHERE IT SHOULD HAVE BEEN 1



TO FIX IT

Previous Layer	BRIGHTER NEURONS	DULLER NEURONS
WEIGHTS	INCREASE	DECREASE
THRESHOLDS	CHANGE	CHANGE

USING THIS SAME IDEA, RECURSIVELY ADJUST THE PARAMETERS OF ALL THE NEURONS ALL THE WAY BACK IN THE NETWORK

REMEMBER THE SAME ADJUSTMENTS SHOULD WORK WELL OVERALL

FOR CIRCLES

FOR NOT-CIRCLES

ALL THIS IS DONE USING LINEAR ALGEBRA & CALCULUS

MEANWHILE, THE MACHINE HAS NO real KNOWLEDGE OF WHAT A CIRCLE IS.

ALL ABOUT MODELS

WHAT DOES IT TAKE TO BUILD A MACHINE LEARNING MODEL? AND HOW DO WE KNOW IF ITS ANY GOOD?

BUILD A MODEL

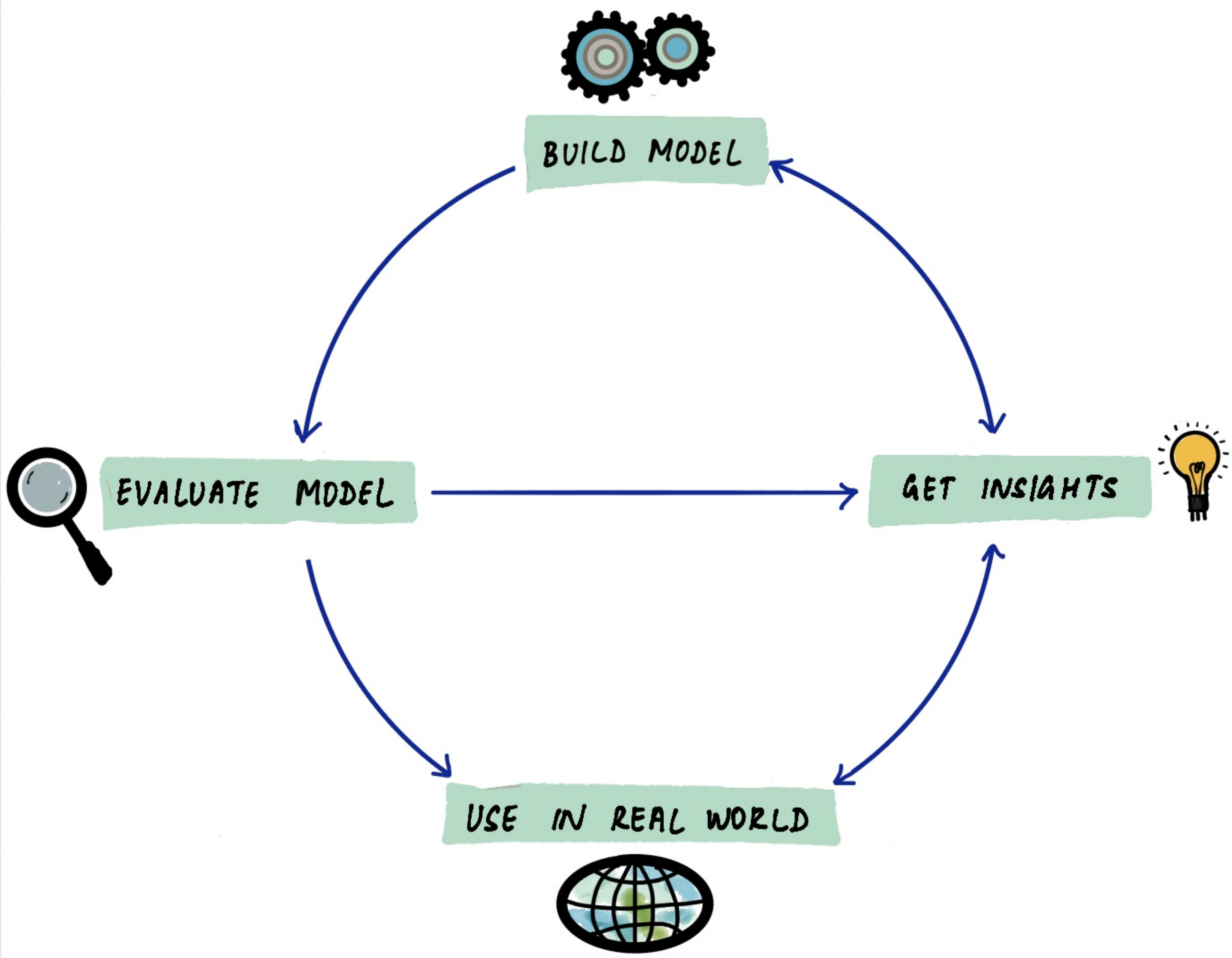
BEGIN WITH A WELL DEFINED QUESTION THAT NEEDS ANSWERING

CLASSIFICATION PROBLEM

- IS THIS A CIRCLE?
- IS THIS A TUMOUR?

REGRESSION PROBLEM

- DO YOUNGER PEOPLE EAT MORE ICECREAM THAN OLDER PEOPLE?
- WHAT AMOUNT TO SPEND ON ADS TO INCREASE MARKET SHARE?

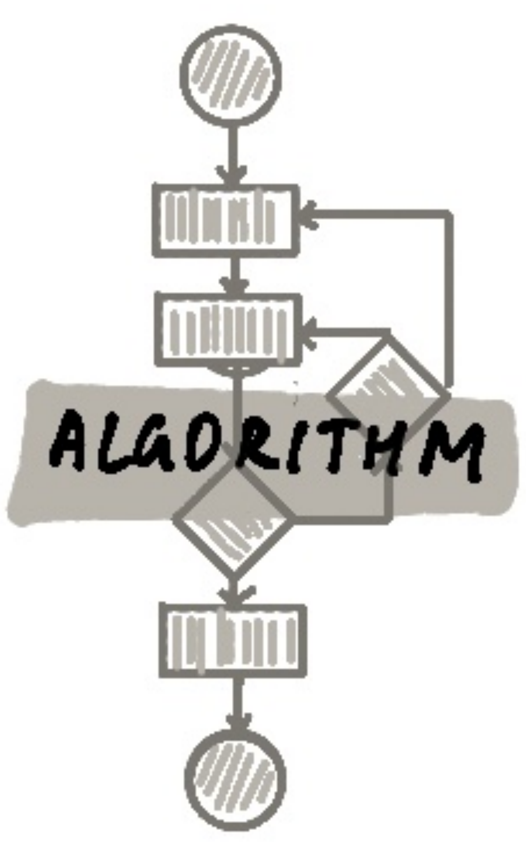


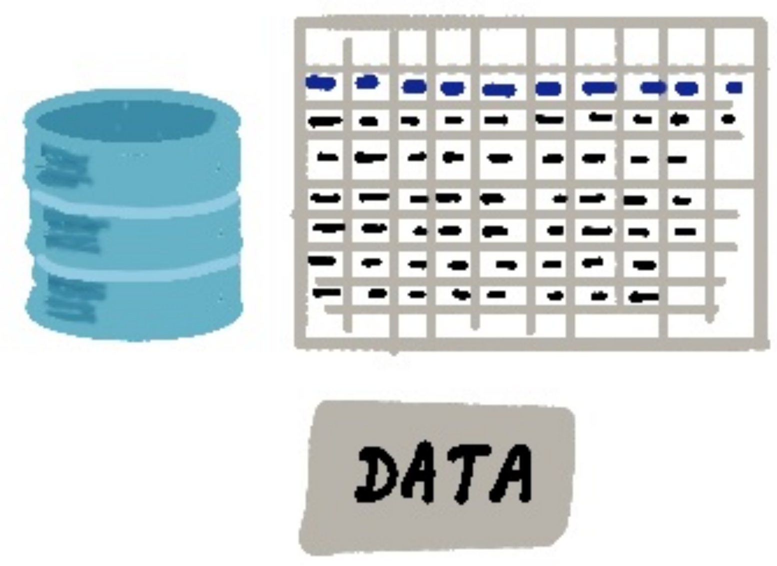
ITERATE/STOP WHEN MEASURABLE GOALS HAVE BEEN ACHIEVED

- WHAT IS THE ACCEPTABLE ERROR LEVEL?
- WHAT CRITEREA EVALUATES THE OUTPUT?
- HOW USEFUL IS IT?

CHOOSING MODELS

MODEL BUILDING CONSISTS OF USING THE RIGHT COMBINATION OF ALGORITHM AND TRAINING DATA, TWEAKED TO MAKE DESIRED 'PREDICTIONS'. THIS PROCESS IS DESCRIBED AS BEING 'PART ART, PART SCIENCE'





THESE CHOICES ARE CLOSELY LINKED AS SEEN BELOW:

ALGORITHM / LEARNING STYLE	SUPERVISED	UNSUPERVISED	REINFORCEMENT
NEURAL NETWORKS	✓	✓	✓
K NEAREST NEIGHBOUR	—	✓	—
—	—	✓	—

ALGORITHM / NATURE OF DATA	NUMERICAL DATA	UNSTRUCTURED DATA	SOME DATA	LOTS OF DATA
LINEAR REGRESSION	✓	—	—	✓
NEURAL NETWORKS	✓	✓	—	✓
—	—	—	✓	—

IT ALSO HELPS TO KNOW PATTERNS OR RELATIONSHIPS IN DATA. FOR THIS,

TAKE A SAMPLE

Run a hypothesis test

MAKE AN ASSUMPTION

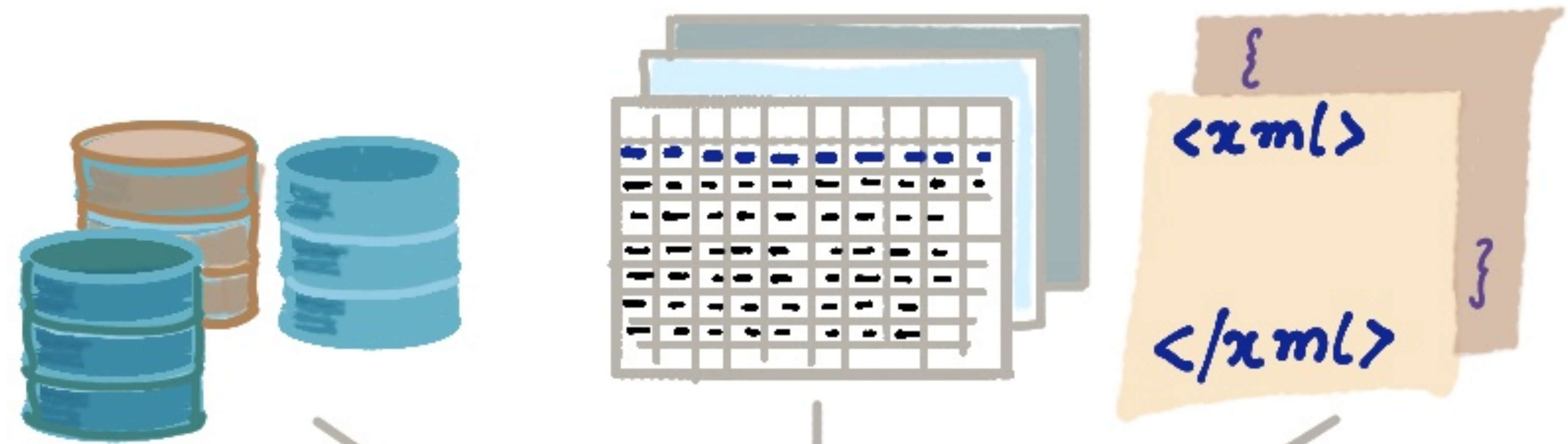
TEST

MEASURE ERROR

ACCEPT OR REJECT ASSUMPTION

PREPARING DATA

DATA MAY COME FROM DIFFERENT SOURCES

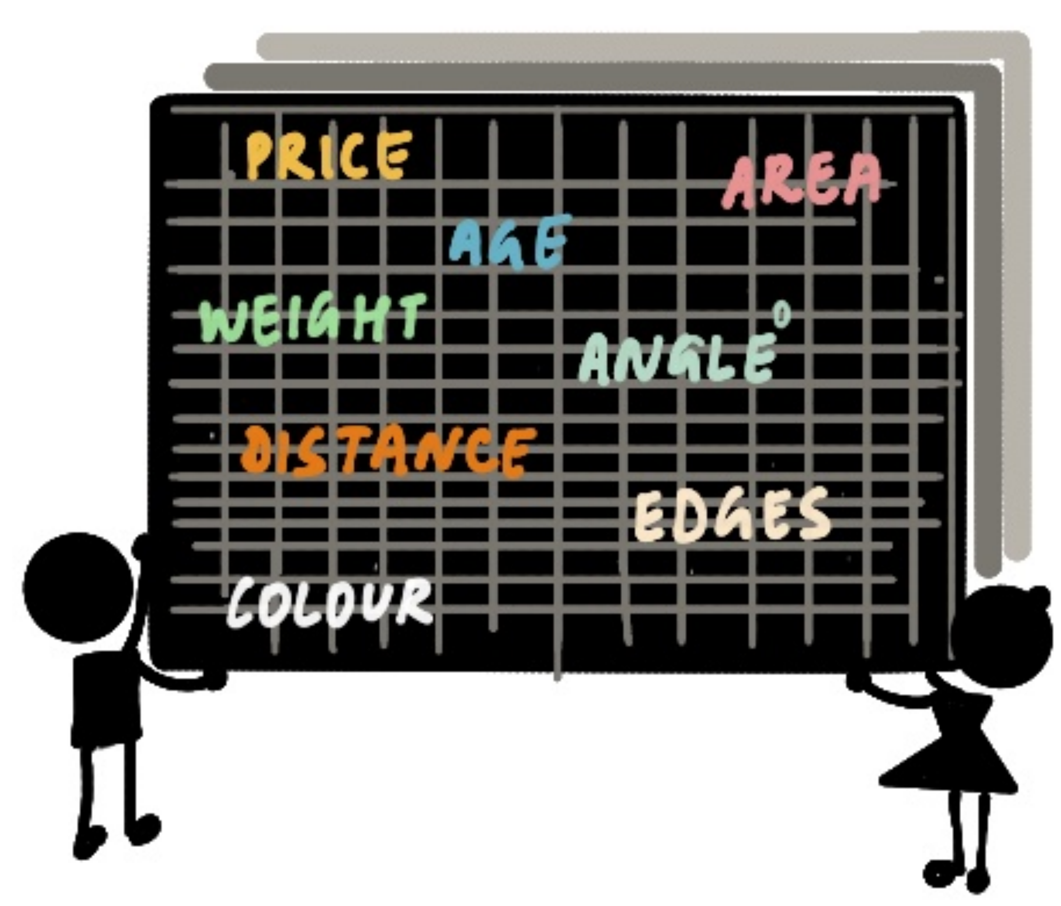


TO BE USEFUL IN TRAINING MODELS, AS FAR AS POSSIBLE, DATA NEEDS TO BE ...

... A CONSISTENT FORMAT

- ERROR FREE
- WITHOUT MISSING BITS

THE DATA MUST BE PRESENTED USEFULLY DESPITE THE FACT...



... THAT IT COULD BE HUGE – BIG DATA

BIASED

UNBIASED

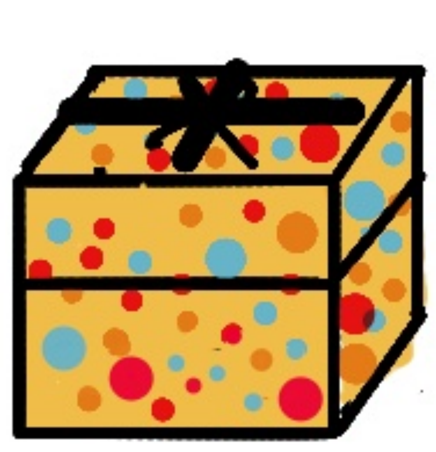
THERE IS ALSO THE BALANCE BETWEEN

REMOVING OUTLIERS

HAVING ENOUGH REPRESENTATIVE DATA

ALL THIS IS APTLY NAMED DATA WRANGLING

ALSO IMPORTANT TO KEEP SEPARATE



TRAINING DATA

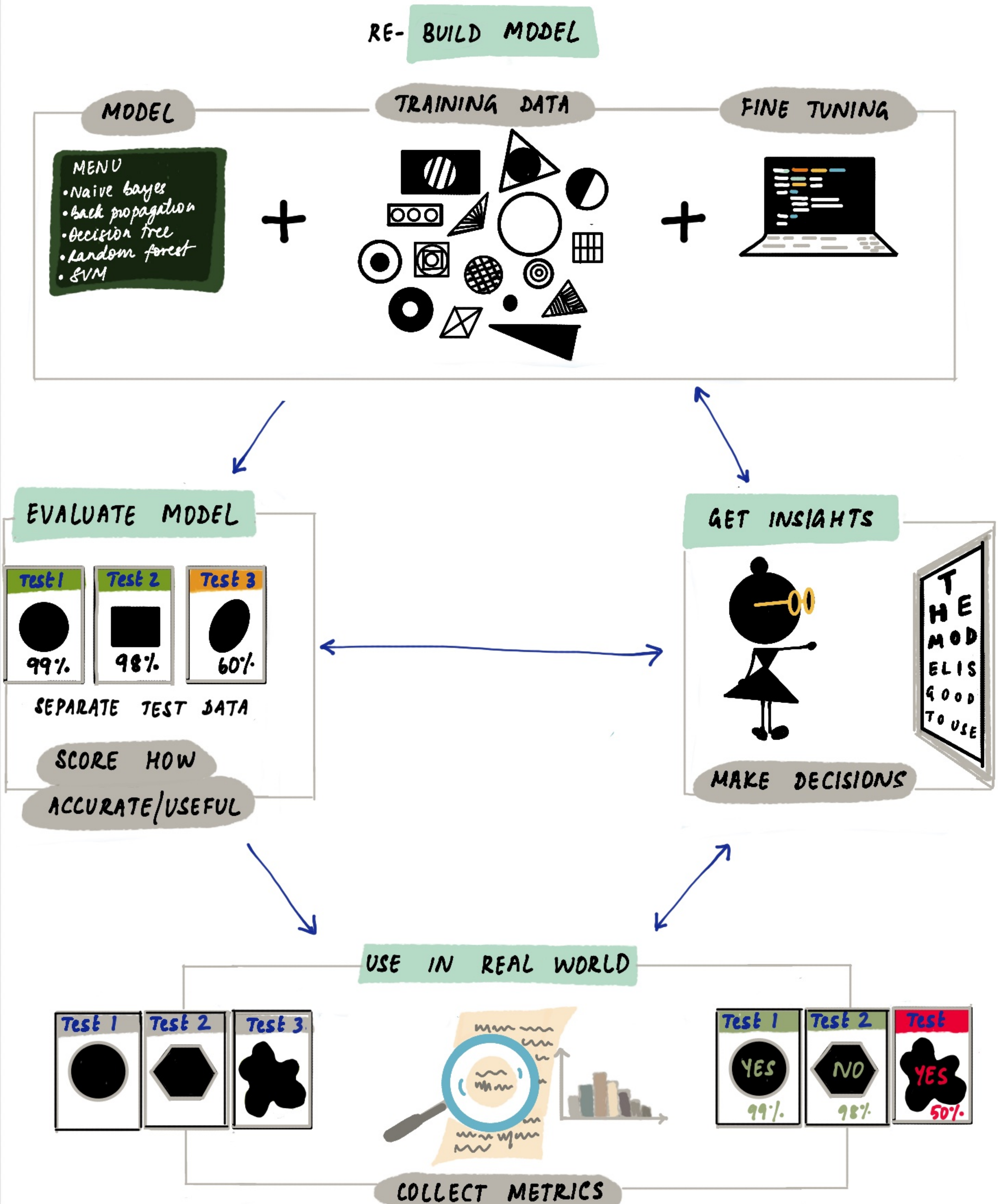
VALIDATION DATA

TEST DATA

for comparing multiple models

IS THIS A CIRCLE?

TO BUILD A MACHINE LEARNING TOOL TO SAY, IDENTIFY A CIRCLE, HERE IS WHAT NEEDS TO HAPPEN – AT A VERY HIGH LEVEL

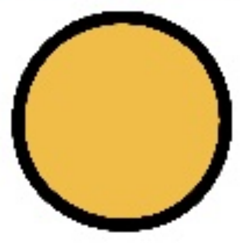


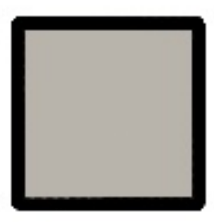
AutoML: A new area of research that aims to automate this process within a limited computational budget once data has been collected arxiv 1810.13306

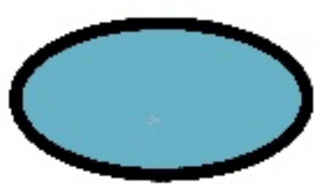
GETTING INSIGHTS

THERE ARE MANY STATISTICAL METRICS TO EVALUATE A MODEL. THEY MEASURE HOW RIGHT THE CORRECT PREDICTIONS ARE AND HOW WRONG THE ERRORS ARE.

LOOK FOR FALSE NEGATIVES AND FALSE POSITIVES

 → CIRCLE → TRUE POSITIVE (TP)

 → NOT CIRCLE → TRUE NEGATIVE (TN)

 → CIRCLE → FALSE POSITIVE (FP)

 → NOT CIRCLE → FALSE NEGATIVE (FN)

PUT THESE NUMBERS INTO THE MATRIX AS BELOW FOR AN OVERVIEW

CONFUSION MATRIX		ACTUAL	
		CIRCLE	NOT
PREDICTION	CIRCLE	# TP	# FP
	NOT	# FN	# TN

% OF CORRECT IDENTIFICATIONS

ACCURACY = # OF CORRECT PREDICTIONS / TOTAL # OF PREDICTIONS

all errors are equally critical

% OF CORRECT POSITIVE IDENTIFICATIONS

PRECISION = # CORRECT POSITIVES / # PREDICTED POSITIVES

false positives are critical

% OF POSITIVES CORRECTLY IDENTIFIED

RECALL = # PREDICTED POSITIVES / # CORRECT POSITIVES

false negatives are critical

EASY TO SEE THAT INCREASING RECALL WILL REDUCE PRECISION, AND THAT NOT ALL METRICS ARE USEFUL IN ALL CIRCUMSTANCES

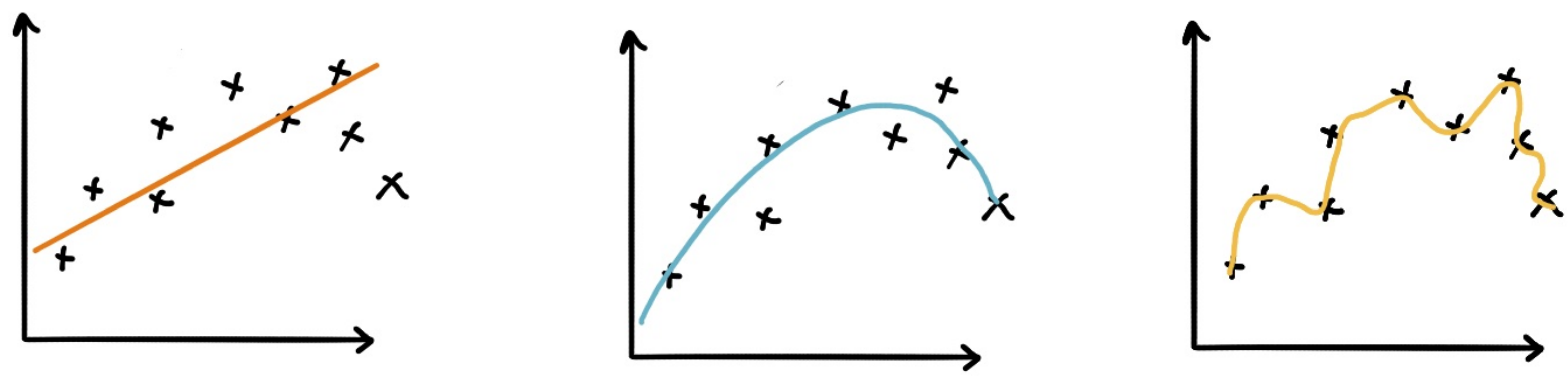
FOR REGRESSION PROBLEMS **MEAN SQUARED ERROR** IS ONE OF MANY HELPFUL METRICS. IT HIGHLIGHTS LARGER ERRORS

$$\text{MEAN SQUARED ERROR} = \frac{1}{N} \times \text{SUM OF ALL } (\text{ACTUAL} - \text{PREDICTED VALUE})^2$$

GETTING INSIGHTS

LOOK OUT FOR UNDERFITTING OR OVERFITTING THE MODEL

HERE IS A VERY COMMON VISUAL EXAMPLE FOR THIS



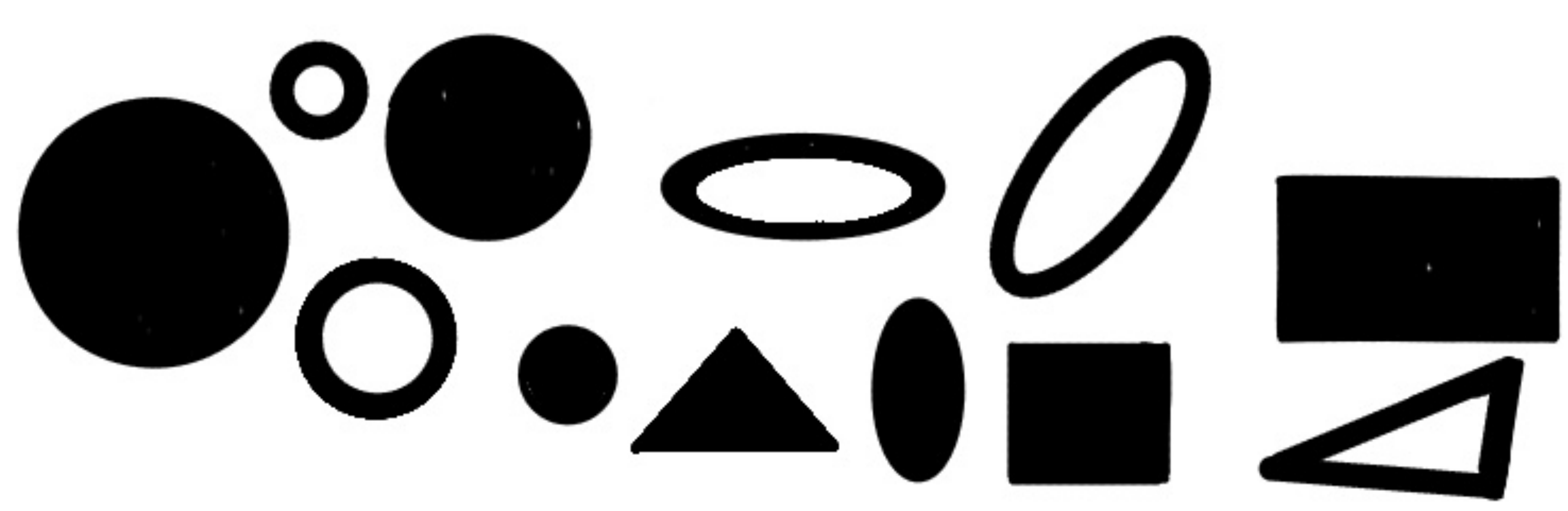
- UNDERFITTED
- HIGH BIAS
- OVERSIMPLIFIES DATA
- TOO MANY ASSUMPTIONS

- JUST RIGHT

- OVERFITTED
- HIGH VARIANCE
- FOLLOWS NOISE
- CANNOT GENERALISE

SO IN OUR EXAMPLE OF 'IS THIS A CIRCLE',

GIVEN: **TEST DATA** FOR A MODEL TO TRAIN ON.



WHEN THE MODEL RUNS ON (NEW) TEST DATA:

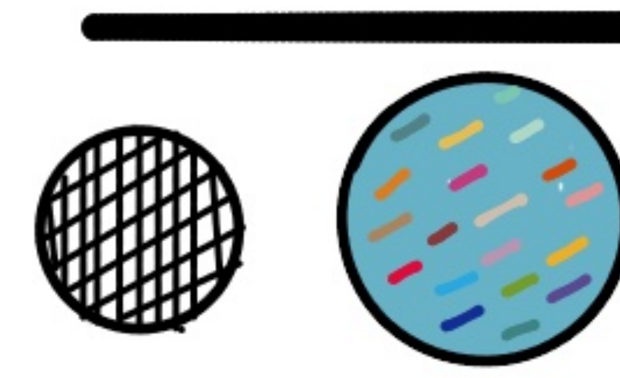
UNDERFITTING

MIGHT MEAN ANYTHING LOOPY OR CLOSED

= CIRCLE

OVERFITTING

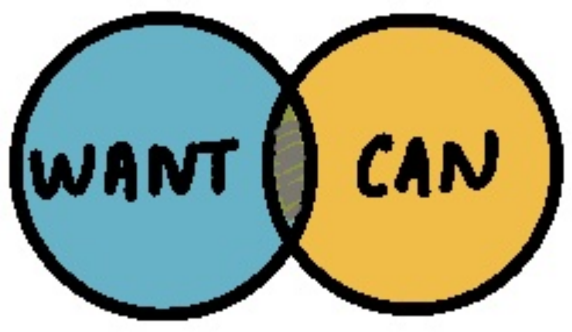
MIGHT MEAN ANY CIRCLE NOT IN THE TRAINING SET



= NOT CIRCLE

HOW GOOD IS THE MODEL?

FEASIBLE



IS THE PROBLEM WELL-DEFINED?

UNBIASED

ARE THE DATA AND ALGORITHMS REPRESENTATIVE & FAIR?

ACCURATE

ARE THE PREDICTIONS CONSISTENT WITH WHAT IS EXPECTED?

UNDERSTANDABLE

ARE THE ACTIONS OF THE ALGORITHM EASY TO EXPLAIN?

SECURE

IS THE DATA, ALGORITHM & CODE TAMPER-PROOF?

PRIVATE

IS SENSITIVE DATA KEPT SAFE FROM PRYING EYES?

A REASONABLY GOOD MODEL ANSWERS YES TO MOST OF THESE QUESTIONS

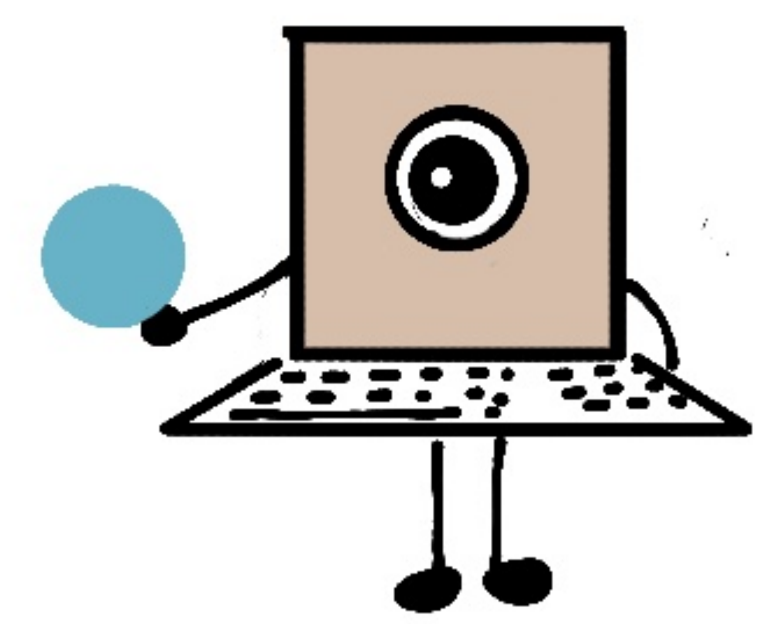
THE HARD PROBLEM THAT IS AI

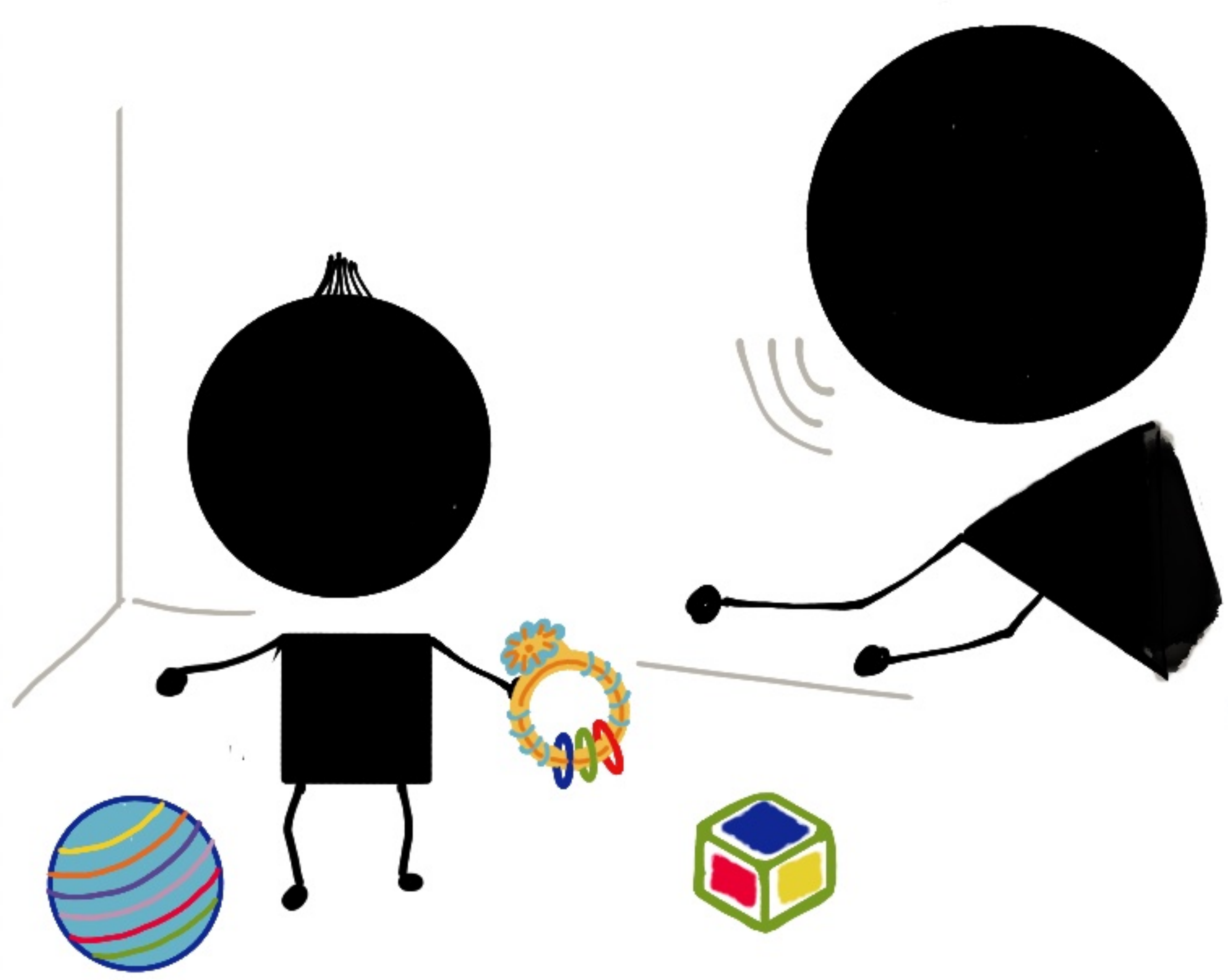
IT APPEARS THAT GETTING AN ALGORITHM TO PERFORM WELL ON A BIG DATASET WITH VERY LITTLE LABELLED DATA IS A TRICK WE HAVEN'T YET MASTERED. BUT THAT IS NOT ALL.

EASY THINGS ARE HARD

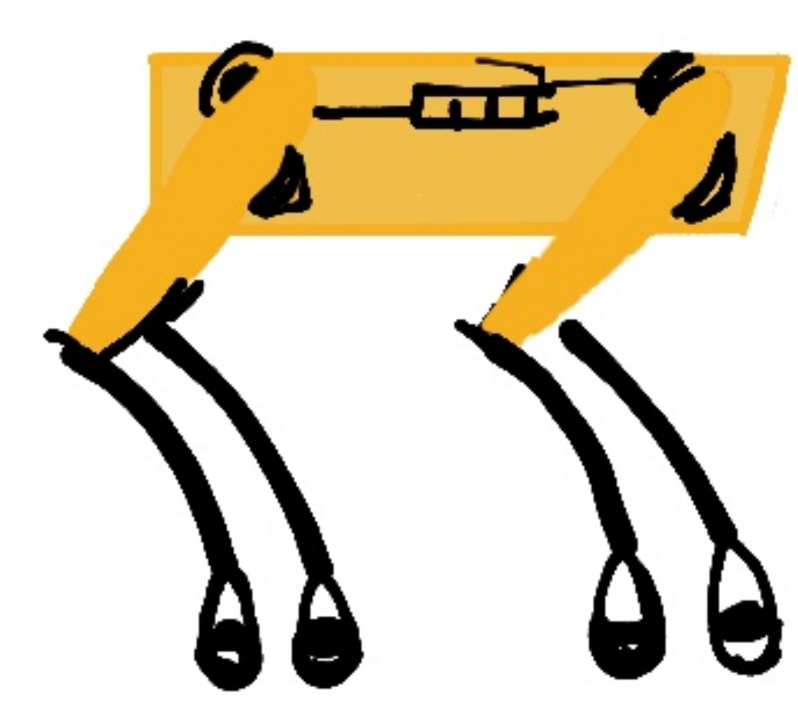
'EASY THINGS ARE HARD'
— MORAVEC'S PARADOX

EVEN FROM AN EARLY AGE, HUMANS (DUE TO MILLENIA OF EVOLUTION) ARE REMARKABLY GOOD AT SENSORY AND MOTOR SKILLS.





IN CONTRAST, A PROGRAM WOULD NEED MILLIONS OF EXAMPLES, COUNTLESS HOURS OF TRAINING MUCH INVESTMENT AND RESEARCH TO PULL OFF THE SAME FEAT



ROBOT DOG

WE CAN MOVE AROUND, PICK UP AND USE OBJECTS, RECOGNISE FACES AND VOICES (EMOTIONS, NON VERBAL CUES) WITH RELATIVE EASE.

'AN IMPERFECT RULE OF AI IS THAT ANYTHING A HUMAN CAN DO IN LESS THAN A SECOND OF MENTAL THOUGHT, AI WILL ALSO BE ABLE TO DO'
- ANDREW NG

MORAVEC'S PARADOX CONTINUES TO BE TRUE SINCE THE EIGHTIES.

COMMON SENSE

WHAT IS IT?

"THE TROPHY DOESN'T FIT IN THE SUITCASE BECAUSE IT WAS TOO SMALL"

"THE TROPHY DOESN'T FIT IN THE SUITCASE BECAUSE IT WAS TOO BIG"

PEOPLE CAN EASILY WORK OUT WHAT THE IT REFERS TO IN BOTH CASES UNLIKE A PROGRAM, WHICH HAS NO FAMILIARITY WITH PACKING THINGS

AN ALTERNATE TURING TEST BASED ON THESE WINOGRAD PAIRS OF SENTENCES HAS BEEN SUGGESTED, TO ASSESS MACHINE INTELLIGENCE

"COMMONSENSE, THE DARK MATTER OF AI"

— OREN ETZIONI

THERE HAVE BEEN ATTEMPTS TO CODIFY COMMONSENSE.

2010s

MACHINE COMMON SENSE

DARPA MCS

LATE 1990s

OPEN MIND COMMON SENSE

MIT OMCS

1980s

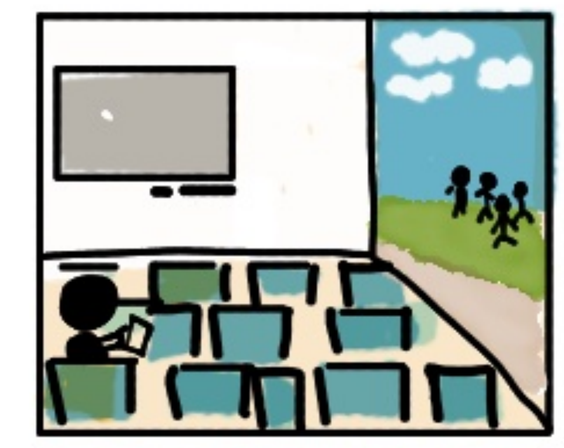
WORLD ENCYCLOPEDIA OF COMMON SENSE

CYCORP CYC

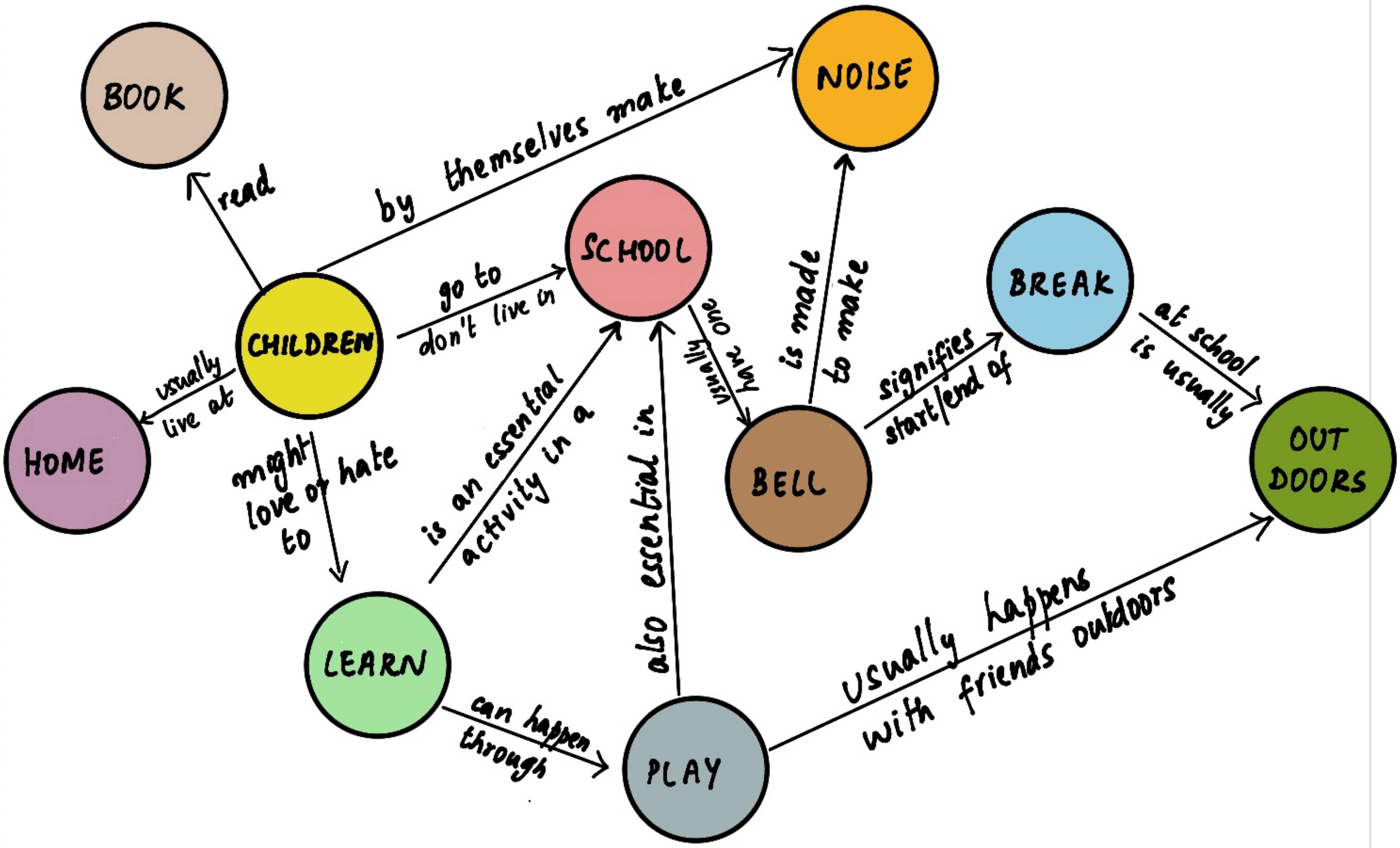
THERE HAS BEEN MUCH DEBATE ON WHETHER THEY ARE EVER FINISHED, USEFUL, OR HAVE EVEN MADE WORTHY CONTRIBUTIONS TO AI AT ALL

EMBODIMENT

THE BELL RANG. ALL THE KIDS RAN OUT.
ONLY ONE STAYED IN WITH A BOOK.



IMPLICIT IN THESE SENTENCES IS A HUGE SET OF PRIORS



THERE IS A VIEW THAT AI CANNOT ACHIEVE INTELLIGENCE UNLESS IT HAS A MORE PHYSICAL-SENSE-EXPERIENCE LIKE HUMANS.

ANY AMOUNT OF ENCODED KNOWLEDGE WILL NOT SUFFICE TO CONVEY CAUSALITY: THAT ONE EVENT OR STATE MAY BE RESPONSIBLE FOR ANOTHER.

NOT ENOUGH TO HAVE THE
RESOURCEFULNESS OF THE THIRSTY CROW
NOT ENOUGH TO MAKE A MACHINE
WONDER WHY THE KID STAYED BACK.

DOES THAT MEAN THE ARTIFICIAL MIND NEEDS AN ARTIFICIAL BODY TOO?

THE HARDER QUESTIONS

HERE WE LOOK AT SOME OF THE MORE OPEN-ENDED, UNRESOLVED QUESTIONS THAT EVERY THINKING PERSON MIGHT COME UP WITH IN THE PROCESS OF CREATING OR INTERACTING WITH AI

COST OF AI

BUILDING A GOOD MACHINE TOOL IS NOT CHEAP. NOT ALWAYS.

IT REQUIRES TREMENDOUS AMOUNTS OF COMPUTING POWER, TONS OF DATA..

... AND STORING THEM IN DATA 'CENTRES'

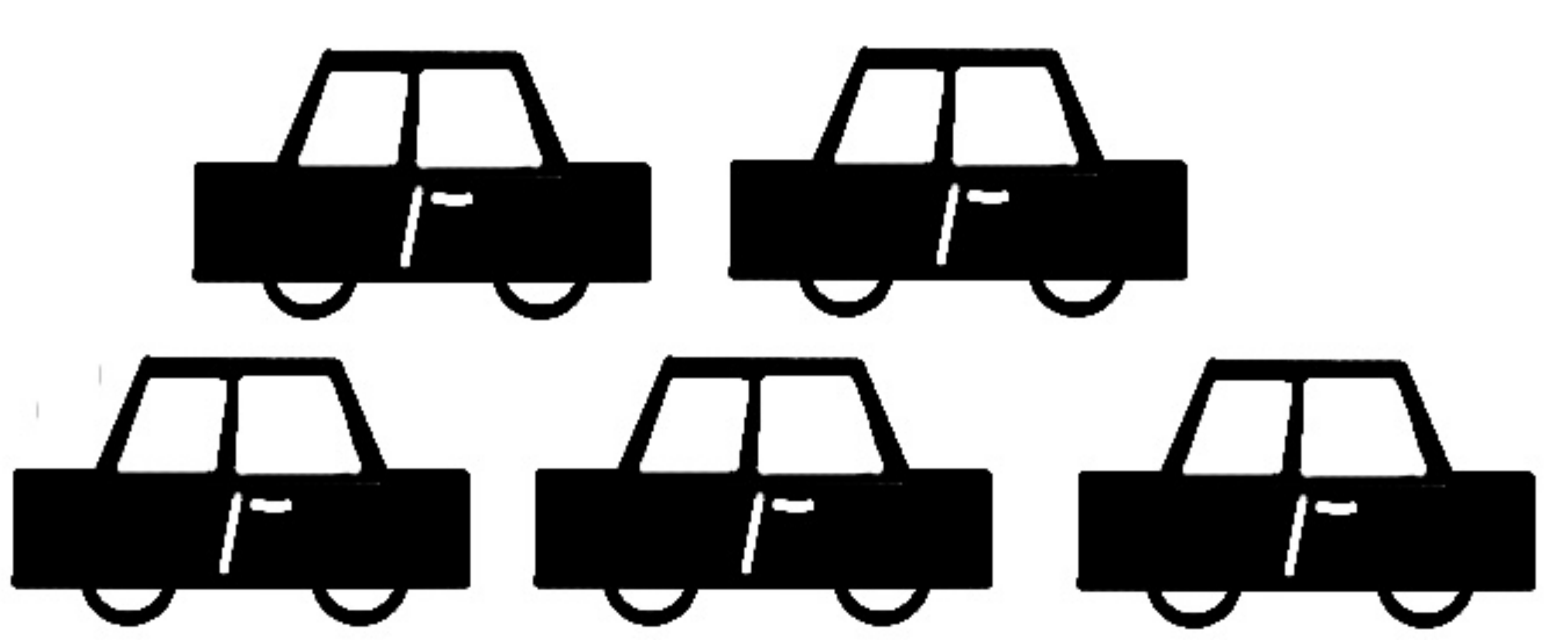
THE COST OF CARRYING OUT THESE COMPUTATIONS IS HIGH ENOUGH TO SEEM LIKE A REASON NOT TO DO THEM.

THIS MIGHT SLOW DOWN INNOVATIONS OR IMPROVEMENTS IN EXISTING ONES

SUCH TECH ALSO COMES WITH AN ENVIRONMENTAL COST

CARBON EMISSIONS

ASSOCIATED TO TRAINING AND DEPLOYING A MODEL HAVE BEEN LIKENED TO...



FIVE TIMES THE LIFE TIME EMISSIONS OF AN AVERAGE CAR

arxiv.org/abs/1906.02243

GREEN TECH ?

THERE ARE NEW AREAS OF RESEARCH AND A PUSH TOWARDS

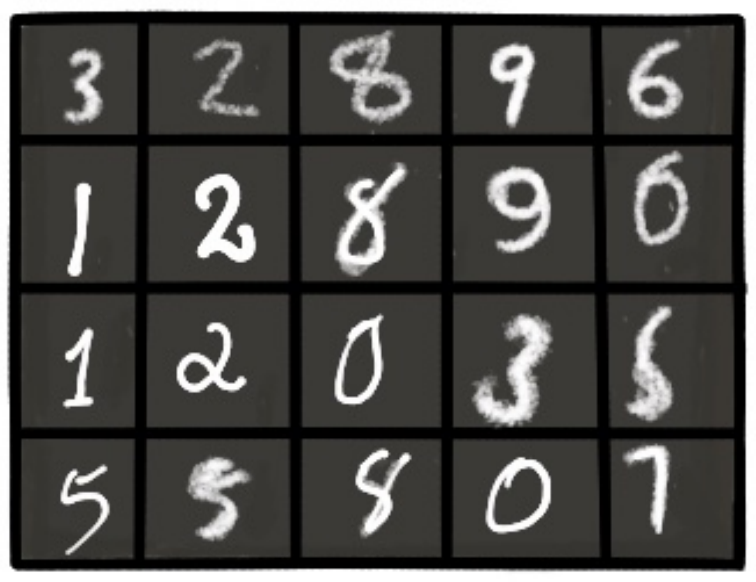
- ✓ EFFICIENT HARDWARE
- ✓ EFFICIENT ALGORITHMS
- ✓ USING PRETRAINED MODELS
- ✓ TRACKING EMISSIONS
- ✓ DISCLOSING EMISSIONS
- ✓ USING SUSTAINABLE CLOUD PROVIDERS

WHAT WILL THE COST OF PROGRESS IN AI BE?

INVENT OR DECEIVE

GENERATIVE ADVERSARIAL NETWORKS OR GANs ARE NEURAL NETWORKS THAT CAN GENERATE NEW DATA THAT RESEMBLE TRAINING DATA

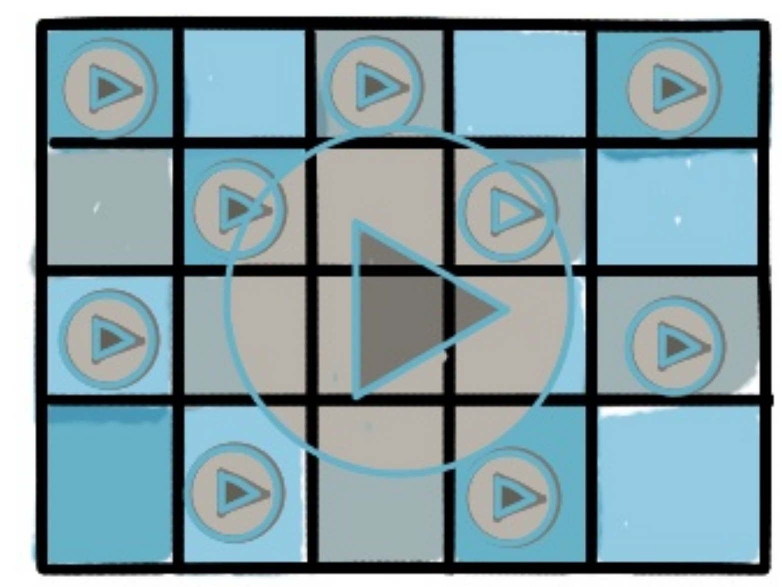
THIS COULD BE



TEXT

IMAGES

AUDIO

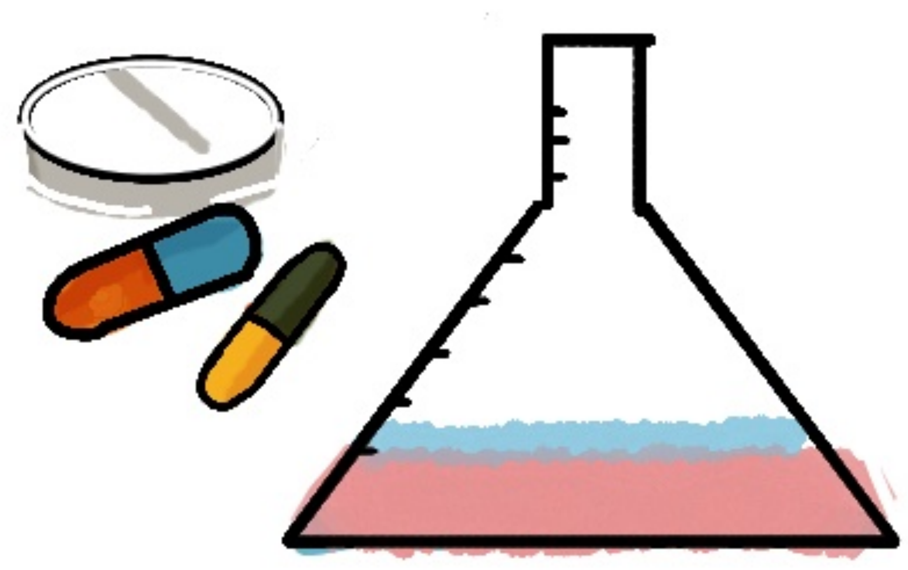


VIDEO

THIS ENABLES AI TO CREATE ART OR COMPOSE MUSIC

GANS CAN BE USED IN SCIENTIFIC RESEARCH

TO ANALYSE EFFECTIVENESS OF A DRUG,



FOR DENTAL RECONSTRUCTION (groundai.com) ORAL-3D

GENERATE TEST DATA FOR QUICKPATH TYPING ON DEVICES (deepai.org)

IN FRAUD PREVENTION AND IN CYBERSECURITY (arxiv.org 1907.03355)

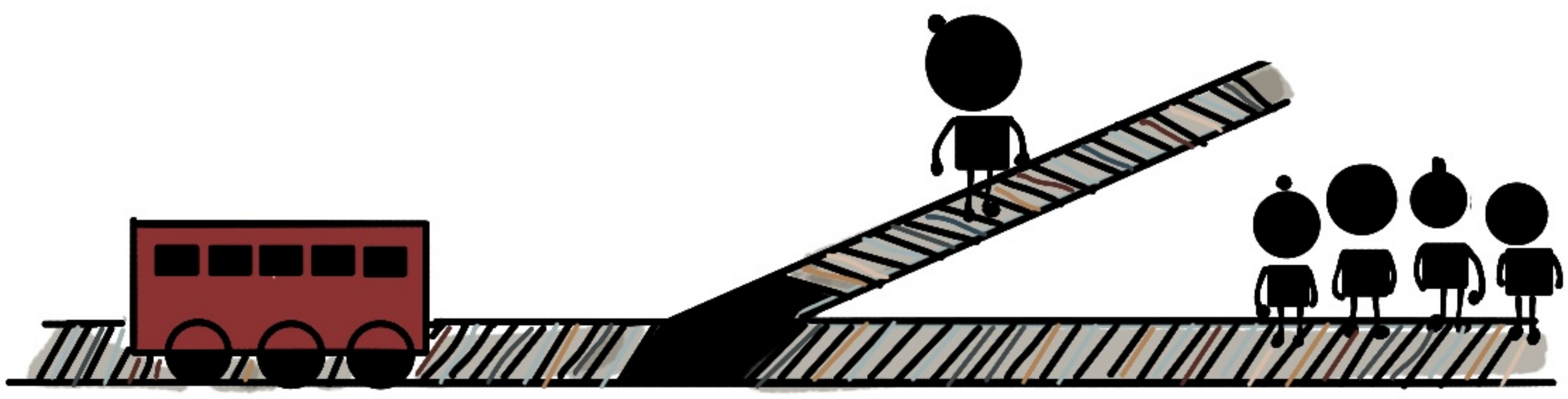
THIS IS ALSO THE TECHNOLOGY BEHIND DEEPFAKES - THOSE VERY CONVINCING VIDEOS ON THE INTERNET USED TO MISLEAD AND MANIPULATE

WHAT CONTROLS OR TOOLS DO WE HAVE TO HELP US BE A BIT MORE DISCERNING ?

WHEN THE AI DOES INVENT SOMETHING USEFUL, WHO GETS THE CREDIT?

TRANSPORT OR HURT

A RUNAWAY TRAIN CAR IS HURTLING DOWN THE TRACKS AND CAN'T STOP



DO NOTHING ⟶ AND 4 PEOPLE ON THE TRACK GET KILLED

DIVERT IT ⟶ AND IT KILLS ONE PERSON ON THE TRACK

WHAT IS TO BE DONE?

THIS DILEMMA IS CALLED THE TROLLEY PROBLEM IN PHILOSOPHY

THIS IS A PROBLEM THAT APPLIES TO A LOT OF THE AI-POWERED DECISION MAKING SYSTEMS.

IN THE CASE OF SELF DRIVING CARS
HOW WOULD A SIMILAR DECISION BE MADE?

IN CASE OF AN EMERGENCY WHO WOULD THE AI SAVE?

THE PASSENGER
OR
THE PEDESTRIAN

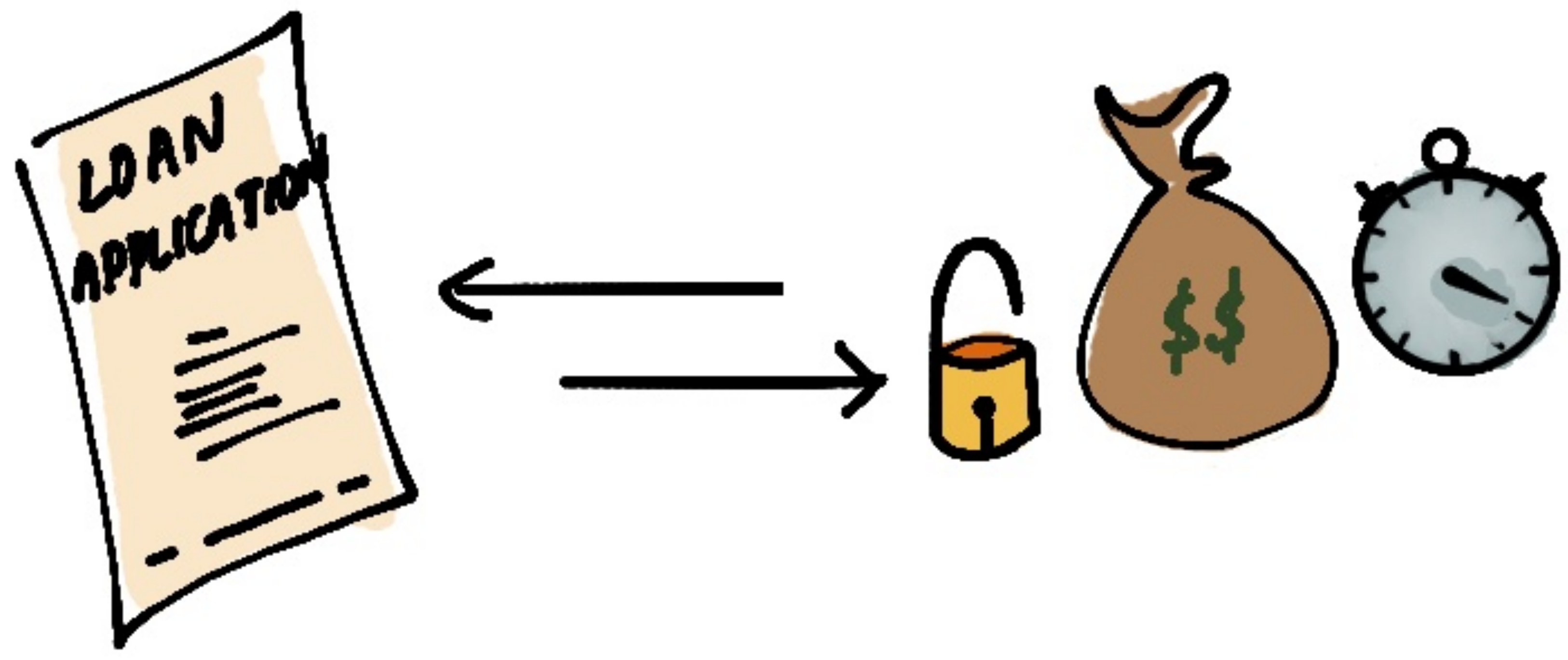
IN CASE OF AN ACCIDENT WHO GETS THE BLAME?

THE OCCUPANT
OR
THE SOFTWARE

DECISIONS & ETHICS

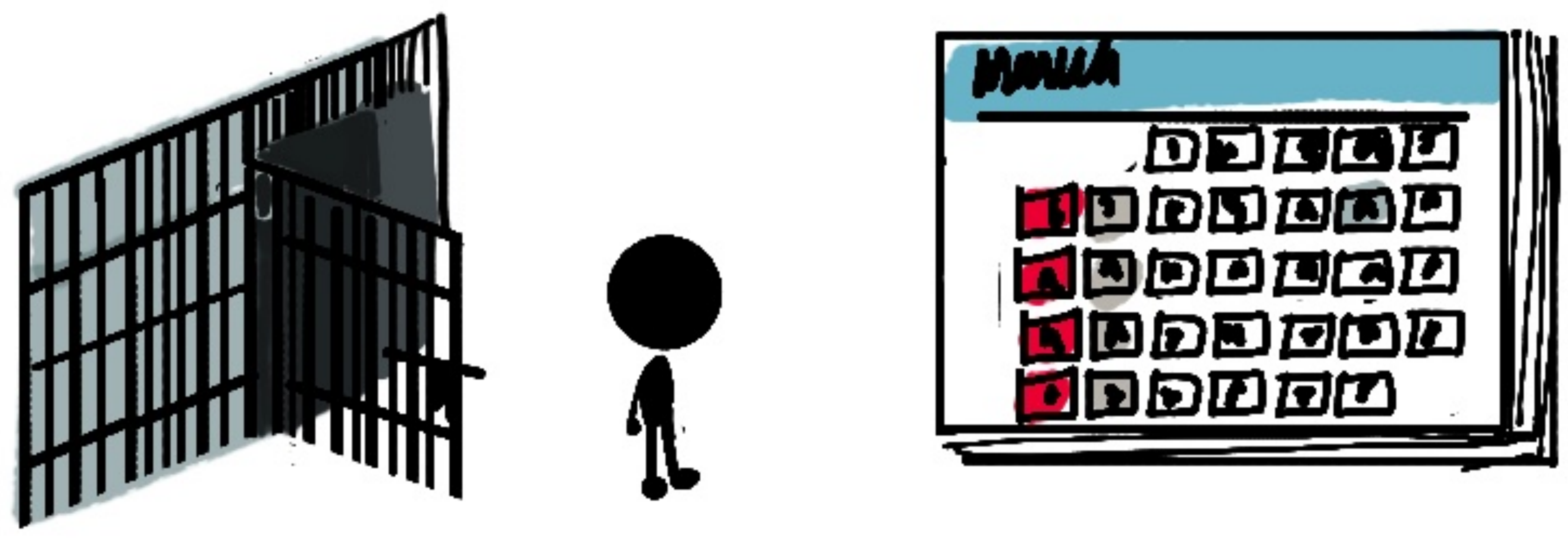
AI TOOLS ARE USED TO MAKE LIFE ALTERING DECISIONS. EVERY SUCH ALGORITHM NEEDS TO BE EXAMINED IN ORDER THAT IT CAN BE TRUSTED

APPROVE LOANS



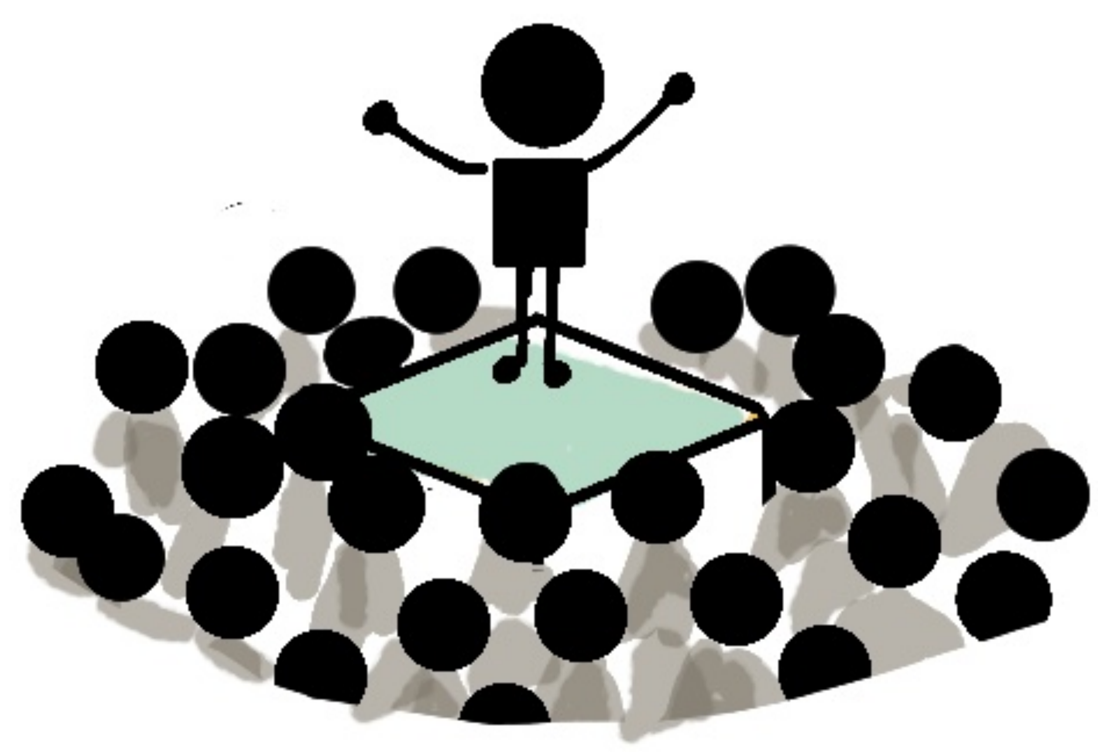
WHO THEN, DOES IT REJECT AND WHY?

GRANT PAROLE



WHO IS DENIED PAROLE AND IS THAT FAIR?

HIRE OR PROMOTE



WHAT FACTOR(S) LED TO THE NON-HIRING DECISION?

GIVE A DIAGNOSIS

WHOSE MEDICAL DATA IS LEAKED TO AN INSURANCE COMPANY?

WHOSE INTEREST IS THE ALGORITHM PROTECTING?

WHAT HARM MIGHT THE DECISION CAUSE AND TO WHO?

WHAT IS FAIR?

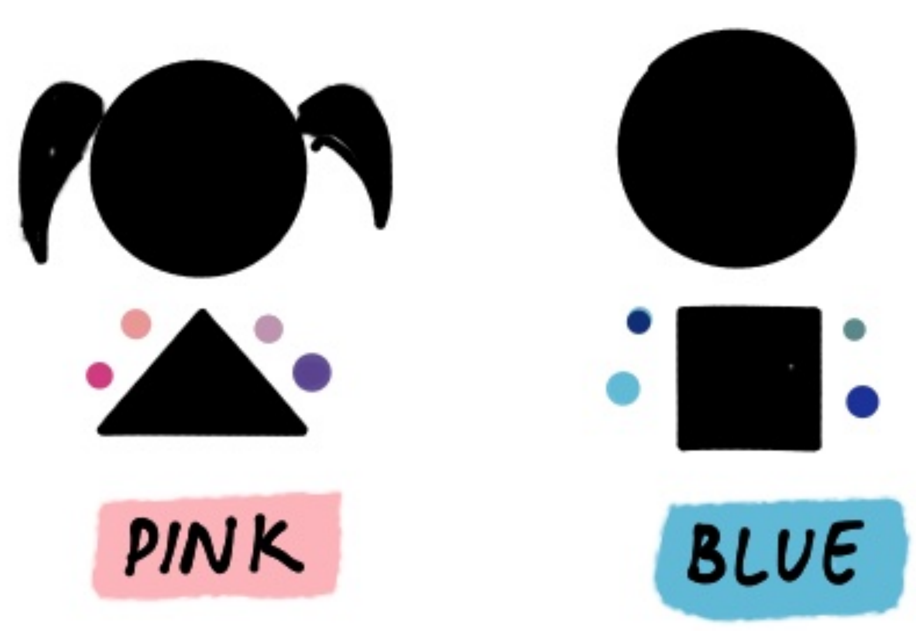
IF THE DATA IS REPRESENTATIVE, IS THE ALGORITHM FAIR OR BIASED?

THESE ARE QUESTIONS NOT JUST FOR PROGRAMMERS TO ENCODE, BUT ALSO FOR THE LAWMAKERS, THE EXPERTS AND THE PUBLIC TO GET INVOLVED.

THE HUMAN FACTOR

REMEMBER WE SAID THAT A GOAL OF AI IS TO THINK LIKE HUMANS AND ACT RATIONALLY. A TALL ORDER, GIVEN HUMAN THOUGHT AND ACTION ARE NOT PREDICTABLE OR STRUCTURED

SOCIETAL BIAS



WE CARRY BIASES

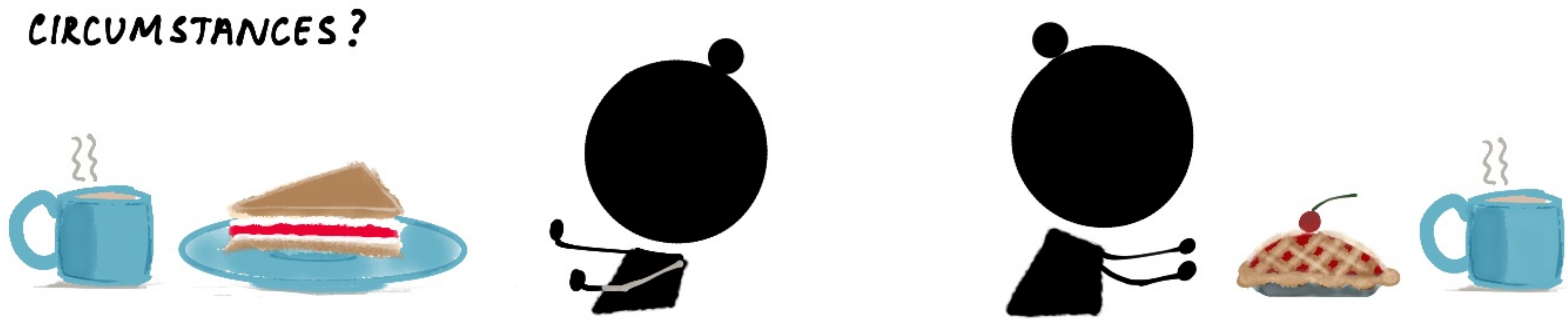
ARE WE ALSO ENCODING THEM?

AUTOMATION BIAS

DO WE ACCEPT 'COMPUTERISED' DECISIONS WITHOUT QUESTION AND COMMON SENSE?

IRRATIONAL BEHAVIOUR

ARE WE SURE OF OUR OWN CONSISTENT BEHAVIOUR IN SIMILAR CIRCUMSTANCES?



FAIRNESS IS UNDEFINED

FAIRNESS TO ONE GROUP
MIGHT COME
AT THE COST
OF FAIRNESS TO ANOTHER

ACCURACY IS NEVER FULLY ACHIEVABLE

ACCURATE PREDICTIONS $\xrightarrow{\text{need}}$ MORE DATA

MORE DATA $\xrightarrow{\text{need}}$ MORE MONEY

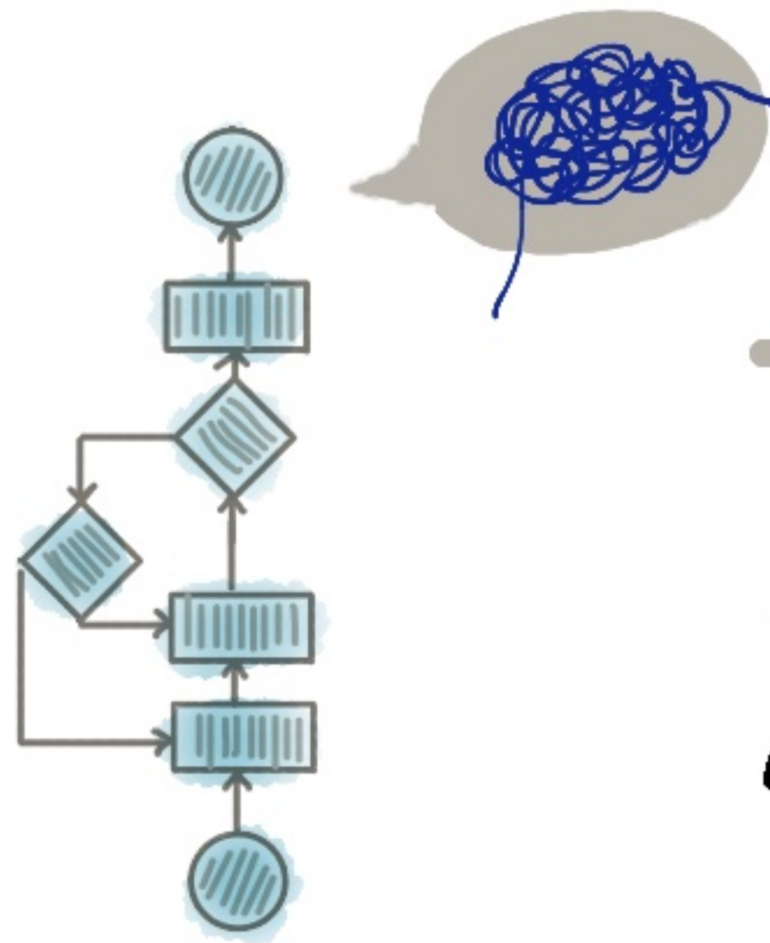
That is a tradeoff we make

HOW TO TRUST AI

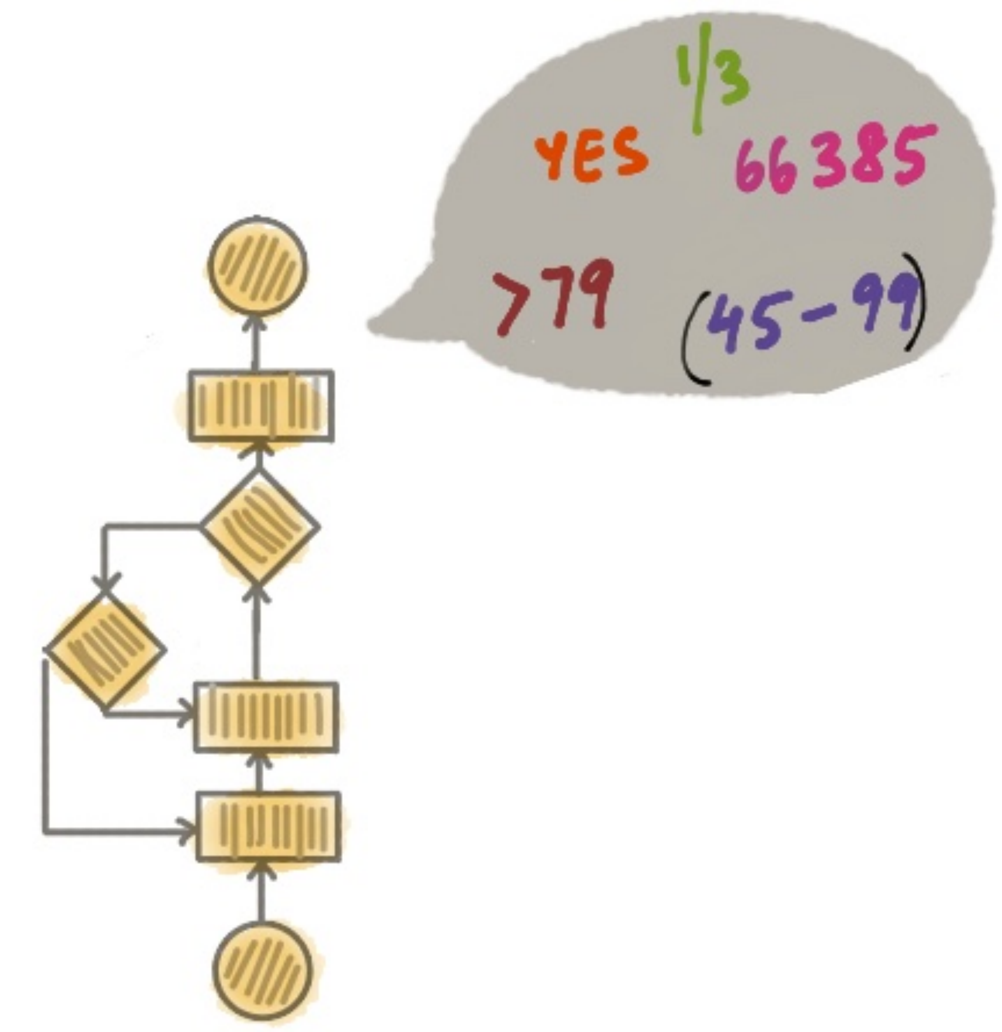
SUPERINTELLIGENT HUMAN-LIKE-OR-SURPASSING AI HAS NOT TAKEN OVER THE WORLD (YET!). BUT WE ARE ALREADY APPLYING 'INTELLIGENT' TOOLS TO MAKE IMPORTANT DECISIONS ON OUR BEHALF.

HOW DO WE HOLD THESE ALGORITHMS TO ACCOUNT?

THESE MODELS ARE NOT EASY TO INTERPRET



BUT THEY CAN BE PROBED TO GET INSIGHTS



FOR EXAMPLE - IN THE ICECREAM CASE

WEIGHTS GIVEN TO EACH FEATURE

Hair Colour	--
Time in queue	+
Education	-
Age	++

INFLUENCE OF EACH FEATURE TO OUTCOME

Appetite — high
Age — higher
No. of arms — medium
Hair Colour — low-non

>= 3 scoops

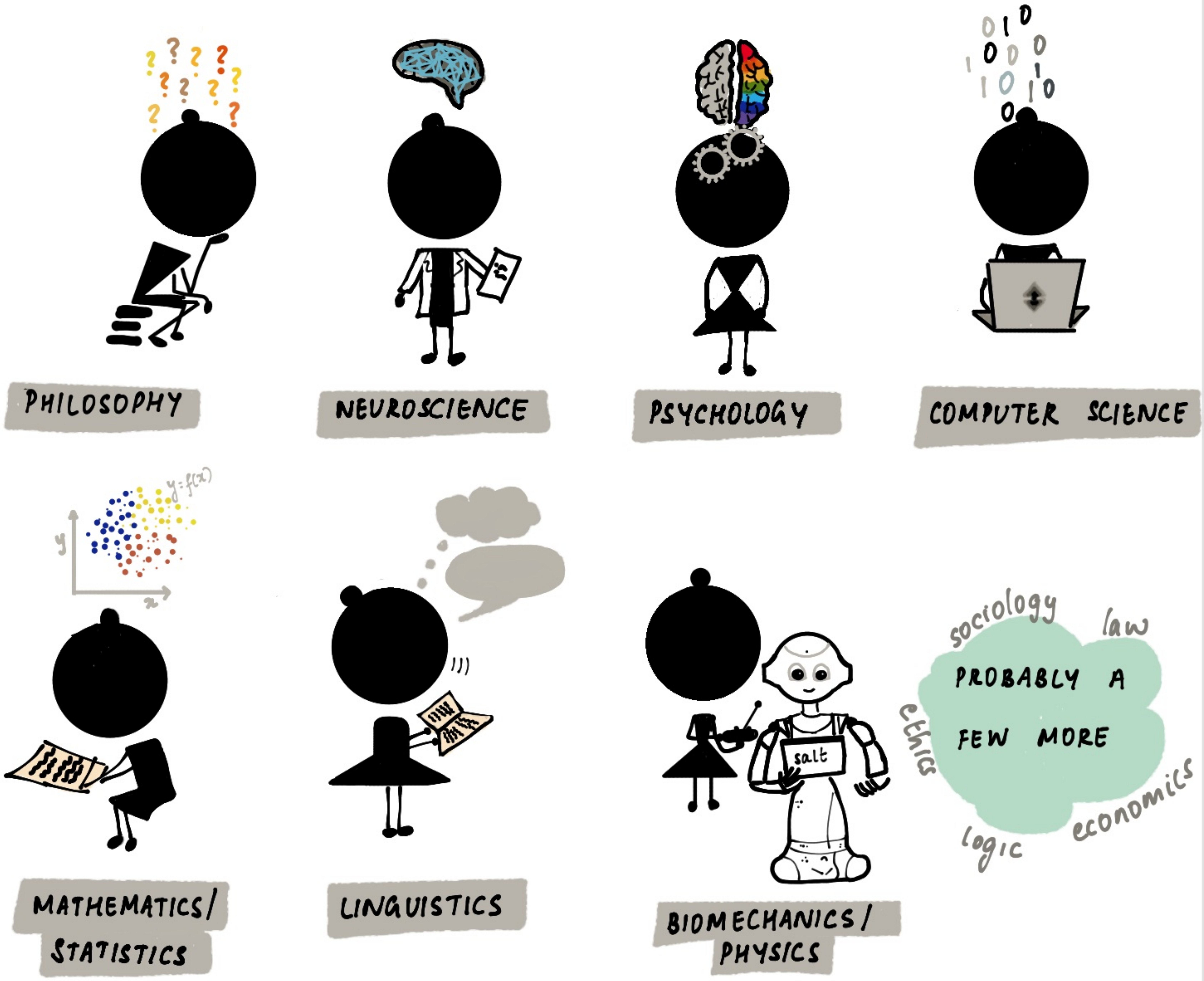
THESE RESPONSES HELP TO DETERMINE THE BASIS FOR A DECISION - TO SEE WHETHER A CERTAIN GENDER, RACE OR QUALIFICATION HAD MORE OR LESS ADVANTAGE IN GETTING A MORE DESIRABLE OUTCOME.

+ GOVERNANCE AUDITS +

THIS - EXPLAINABLE AI - IS A NEW AREA OF RESEARCH. IN COMBINATION WITH GOOD GOVERNANCE, REGULAR AUDITS, AND A GOOD MODEL, BECOMES RESPONSIBLE AI - MORE DESERVING OF OUR TRUST.

A CULMINATION

AI WORKS BY WORKING TOGETHER WITH OTHER DISCIPLINES



HOWEVER, THE OUTPUT FROM AI IMPACTS NEARLY EVERY ASPECT OF OUR INDIVIDUAL AND COLLECTIVE LIVES. THE INFLUENCE OF NARROW AI NEEDS MORE ATTENTION (THAN FUTURE AGI) IN TERMS OF BREACHES IN SECURITY, PRIVACY & TRUST AND OF MANIPULATING OUR THOUGHTS WITH ALGORITHM-CHOSEN CONTENT.

AI CAN IMPACT POSITIVELY WHEN CREATED IN COLLABORATION WITH POLICY MAKERS, EXPERTS AND THOSE FOR WHO IT IS INTENDED.

AI IS MORE PROFOUND THAN ELECTRICITY OR FIRE

— SUNDAR PICHAI

WOMEN IN AI

WE NEED TO KNOW, SHARE AND HONOUR THEIR ACHIEVEMENTS

AND WHO KNOWS — IT MAY INSPIRE ANOTHER!

WOMEN IN AI

FEI-FEI LI

REVOLUTIONISED COMPUTER VISION WITH IMAGENET

ACM FELLOW, SPEAKER AND MENTOR

ISABELLE GUYON

COINVENTER OF SUPPORT VECTOR MACHINES & SIAMESE NEURAL NETWORKS

RUZENA BAJCSY

PERCEPTION METHODS IN ROBOTICS AND IN MEDICAL IMAGE ANALYSIS

ACM ALLEN NEWELL AWARD

LATANYA SWEENEY

KNOWN FOR DATA PRIVACY LAB AND FOR THE IDEA OF K-ANONYMITY (TO RE-IDENTIFY ANONYMISED DATA)

MARGARET BODEN

THINKER OF THE MIND'S COMPUTATION MODELS & PHILOSOPHY OF ARTIFICIAL LIFE

HAVA SIEGELMANN

RAN LIFE LONG LEARNING MACHINE PROGRAM @ DARPA

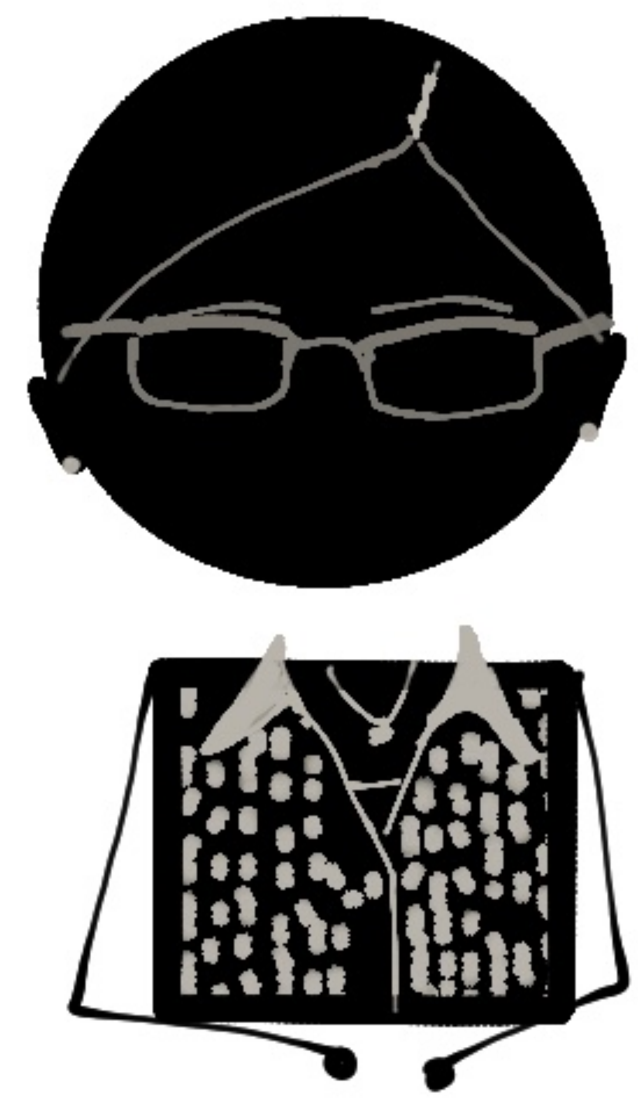
INVENTED SUPER-TURING COMPUTATION

WOMEN IN AI

ANIMA ANANDKUMAR

PIONEER OF TENSOR ALGORITHMS, PROMOTER OF DEMOCRATISING AI MENTOR, SPEAKER.

VIVIENNE SZE



KNOWN FOR RESEARCH IN ENERGY EFFICIENT MACHINE LEARNING

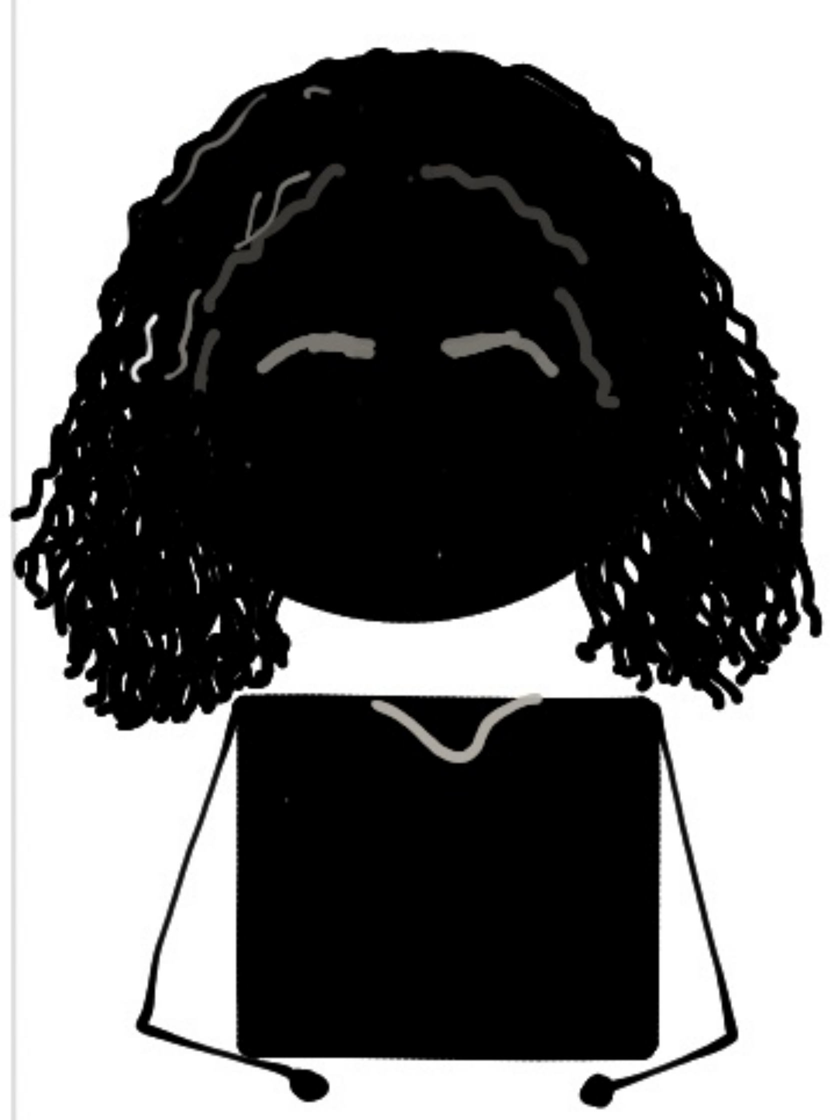
DAPHNE KOHLER

CO-CREATOR OF MOOC AND USES MACHINE LEARNING FOR DRUG DISCOVERY

TIMNIT GEBRU

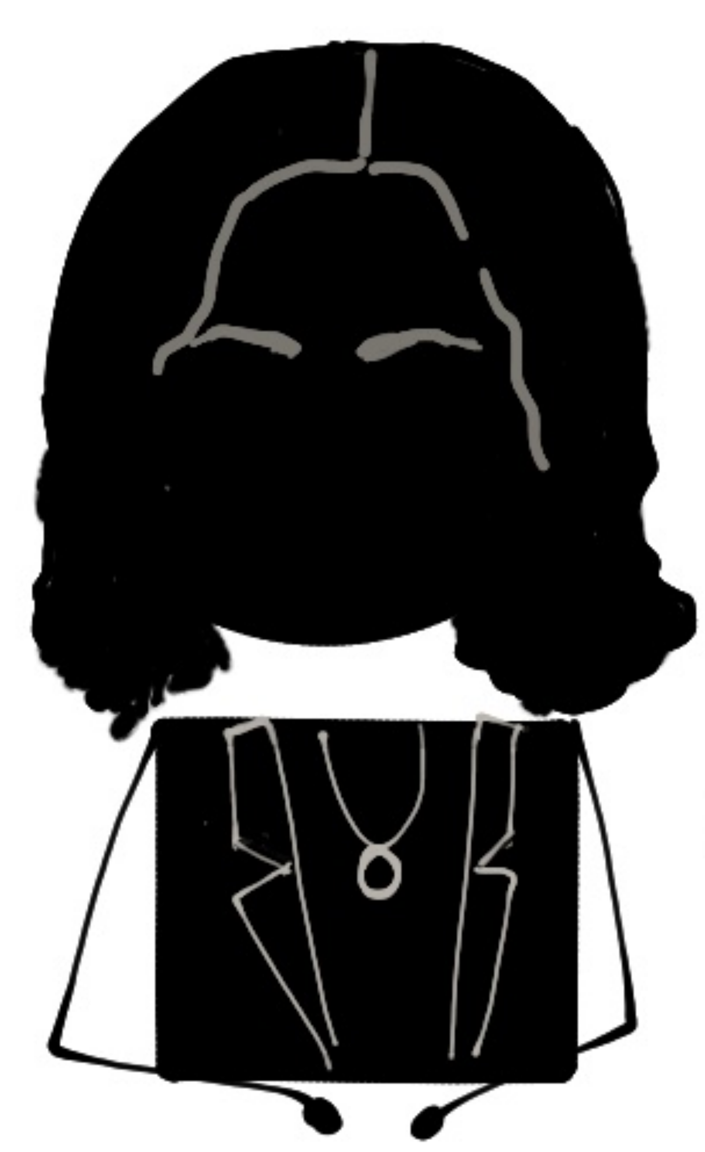
KNOWN FOR RESEARCH IN ETHICAL ARTIFICIAL INTELLIGENCE, SPEAKER@TED

TABITHA GOLDSTAUB



CHAIR OF UK GOV AI COUNCIL TO CHAMPION RAPID RESPONSIBLE AI

HILARY MASON



ML/DATA SCIENTIST ADVISOR TO BUSINESSES MENTOR TO STUDENTS

MORE TO EXPLORE

WITH A TOPIC AS VAST AS AI, IT IS HARD TO BE ABLE TO COVER EVERY ASPECT OF IT WITHOUT VEERING AWAY FROM THE MAIN THREAD

HERE ARE A FEW INTERESTING ASIDES I WANT TO HIGHLIGHT SHOULD YOU WISH TO EXPLORE

- OTHER WAYS TO LEARN – TRANSFER LEARNING, GENETIC ALGORITHMS
- ADVERSARIAL ATTACKS ON MODELS
- USE SYMBOLIC AI WITH TODAY'S TECH
- ROBOTICS & NEW LAWS FOR ROBOTS
- QUANTUM MACHINE LEARNING
- CONSCIOUSNESS IN CONTEXT OF HUMAN-LIKE AI
- CONTRIBUTIONS OF:

DAVID RUMELHART JAMES MCCLELLAND GEOFFREY HINTON JUDEA PEARL RAY KURZWEIL HANS MORAVEC LESLIE KAELBLING YANN LE CUNN YOSHUA BENGIO RAY SOLOMONOFF OLIVER SELFRIDGE TESSA LAU HOD LIPSON TRENCHARD MORE VLADIMIR VAPNIK JOY BUOLAMWINI MICHAEL I JORDAN CHRISTOPHER STRACHEY DIETRICH PRINZ DEMIS HASSABIS IAN GOODFELLOW JOSEPH WEIZENBAUM GERALD SUSSMAN ANDREW NG MOOJAN ASGHARI ADOLFO GUZMAN JOHN KOZA, CAROL E REILEY DANIEL BOBROW AYANNA HOWARD TOM MITCHELL ROSS QUINLAN STEVE MUGGLETON RANA KALIOUBY RODNEY BROOKS DAVID MERR NICK BOSTROM ARTHUR C CLARKE ANDREA FROME JOHN SEARLE STANLEY KUBRICK JEFF HAWKINS STEPHEN WOLFRAM DAWN SONG DAVID CHALMERS MELANIE MITCHELL OLGA RUSSAKOVSKY REGINA BARZILAY CHRISTOF KOCH CATHY O'NEIL ILYA SUTSKEVER PETER NORVIG SEBASTIAN THRUN ANIL K JAIN ANDREJ KARPATHY RACHEL THOMAS JOHN HOLLAND DAVID HECKERMAN PETER HART DOUGLAS HOFSTADTER

MY REFERENCES

Online courses

MIT opencourseware - Patrick Winston's lecture series
Stanford - Andrew Ng - on coursera
AI for everyone - Deeplearning.ai Andrew Ng
IBM on Artificial intelligence - coursera

Books

Machines who think - Pamela McCorduck
Hello World - Hannah Fry
Artificial Intelligence: A Guide for thinking humans - Melanie Mitchell
Master Algorithm - Pedro Domingos
Life 3.0 - Max Tegmark
Weapons of Math Destruction - Cathy O'Neil

Intelligence

R Sternberg Human intelligence - britannica.com
Introduction to Psychology - opentextbc.ca
plato.stanford.edu
 - Innateness and language
 - Chomsky on language
 - Artificial intelligence
Chomsky and Origins of language - news.mit.edu
Chomsky - Language and Thought - youtube interview
Shane Legg - Talk on youtube
Daniel Dennet - Wired Interview
Will AI achieve consciousness? - Wired
Common sense to computers - quantamagazine.org
A tribute to Alan Turing - thoughtworks.com

MY REFERENCES

Concepts/Algorithms/Trends

- machinelearningmastery.com
- becominghuman.ai
- towardsdatascience.com
- kdnuggets.com
- stateof.ai

Tools A brief history of ASR (automatic speech recognition) - medium.com

Cost Prepare for AI to produce less wizardry - Wired

Privacy/Security. Perfectly privacy preserving AI - towardsdatascience.com

Learning Models Continuous Delivery for Machine Learning - Thoughtworks.com

Sustainability

Training a model and emission five times a car's - MIT technology review
Environmental impact of AI - Forbes

Explainability

christophm.github.io/interpretable-ml-book/

Talks/Podcasts/Videos

Bias in Algorithms - Joy Buolamwini
Limitations in AI - Timnit Gebru
Algorithmic Fairness, Privacy and Ethics - Michael Kearns, Lex Fridman
Measure of Intelligence - Francois Chollet Lex Fridman
On AI and machine learning - AnitaB.org Fei Fei Li
Human centred AI - Fei Fei Li