

The unreasonable effectiveness of overparameterization

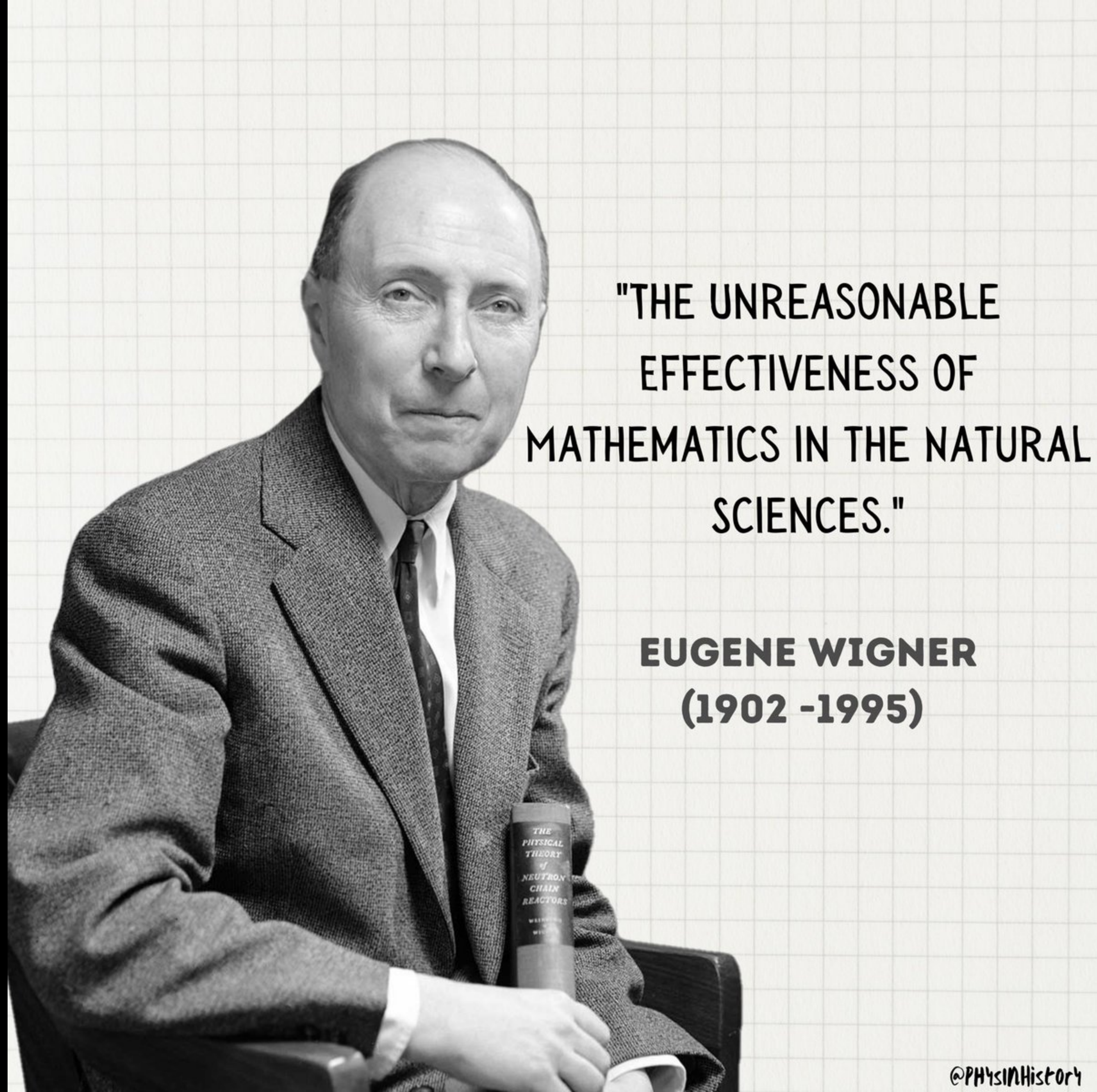
Eric Postma

Cognitive Science & Artificial Intelligence, TiU

JADS, 's-Hertogenbosch, TiU/TUe

eric.postma@gmail.com

www.ericpostma.nl

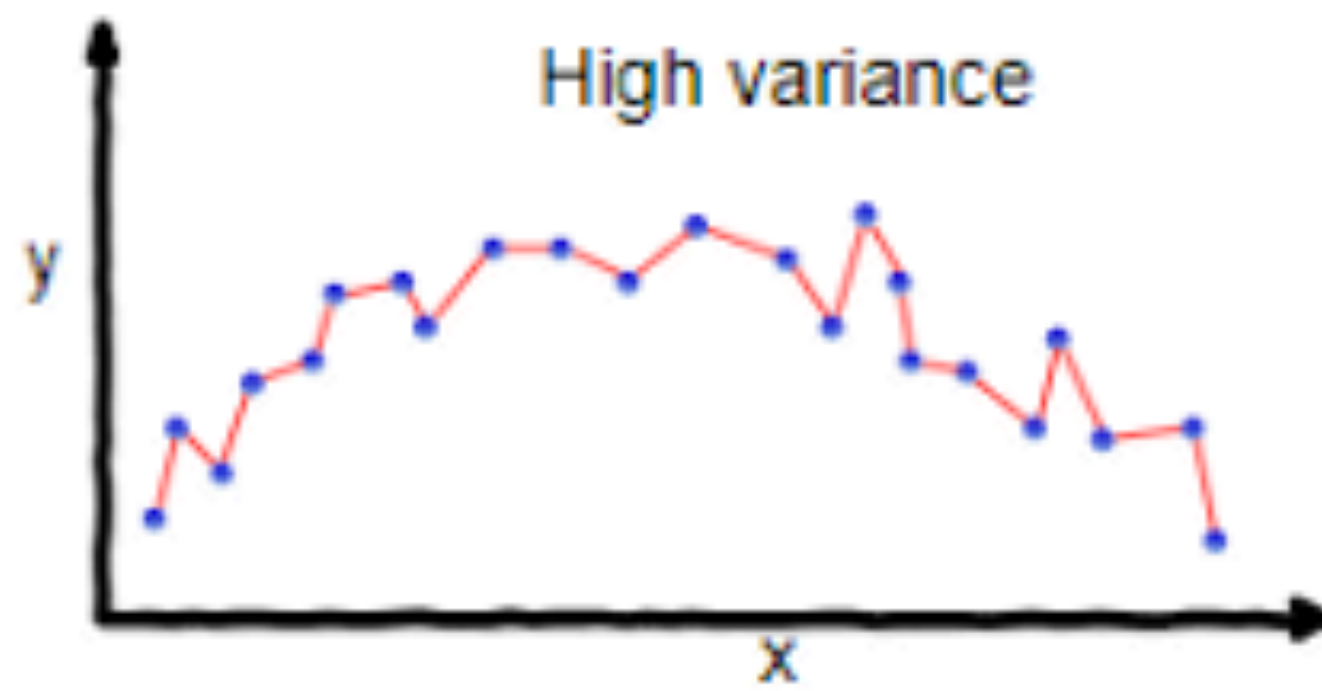


"THE UNREASONABLE
EFFECTIVENESS OF
MATHEMATICS IN THE NATURAL
SCIENCES."

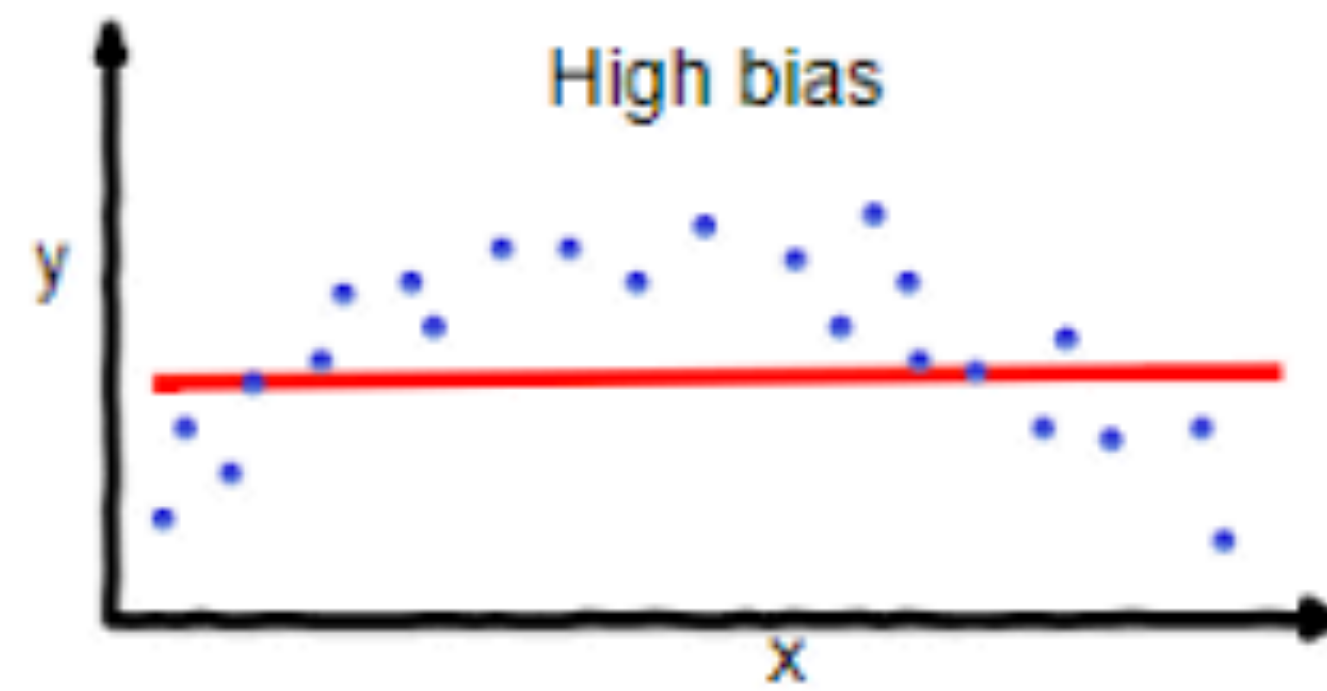
EUGENE WIGNER
(1902 -1995)

Overparameterization

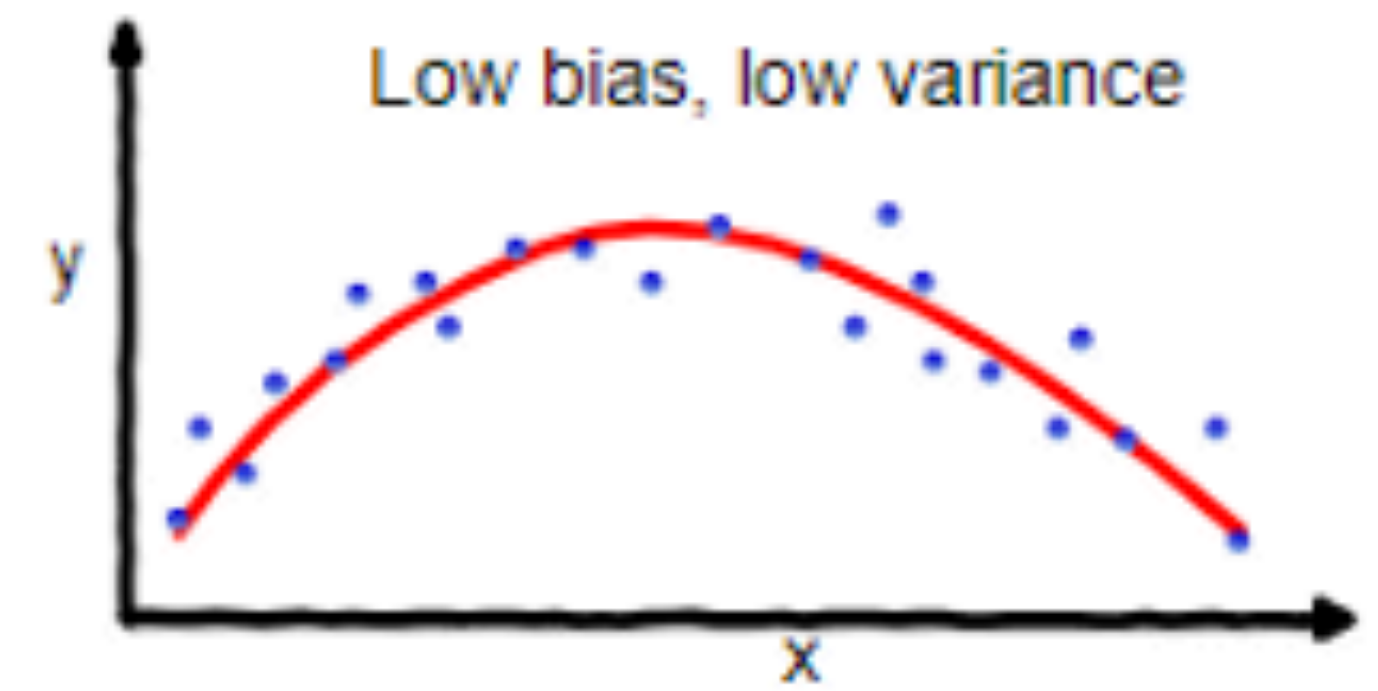
Occam's razor



overfitting



underfitting



Good balance



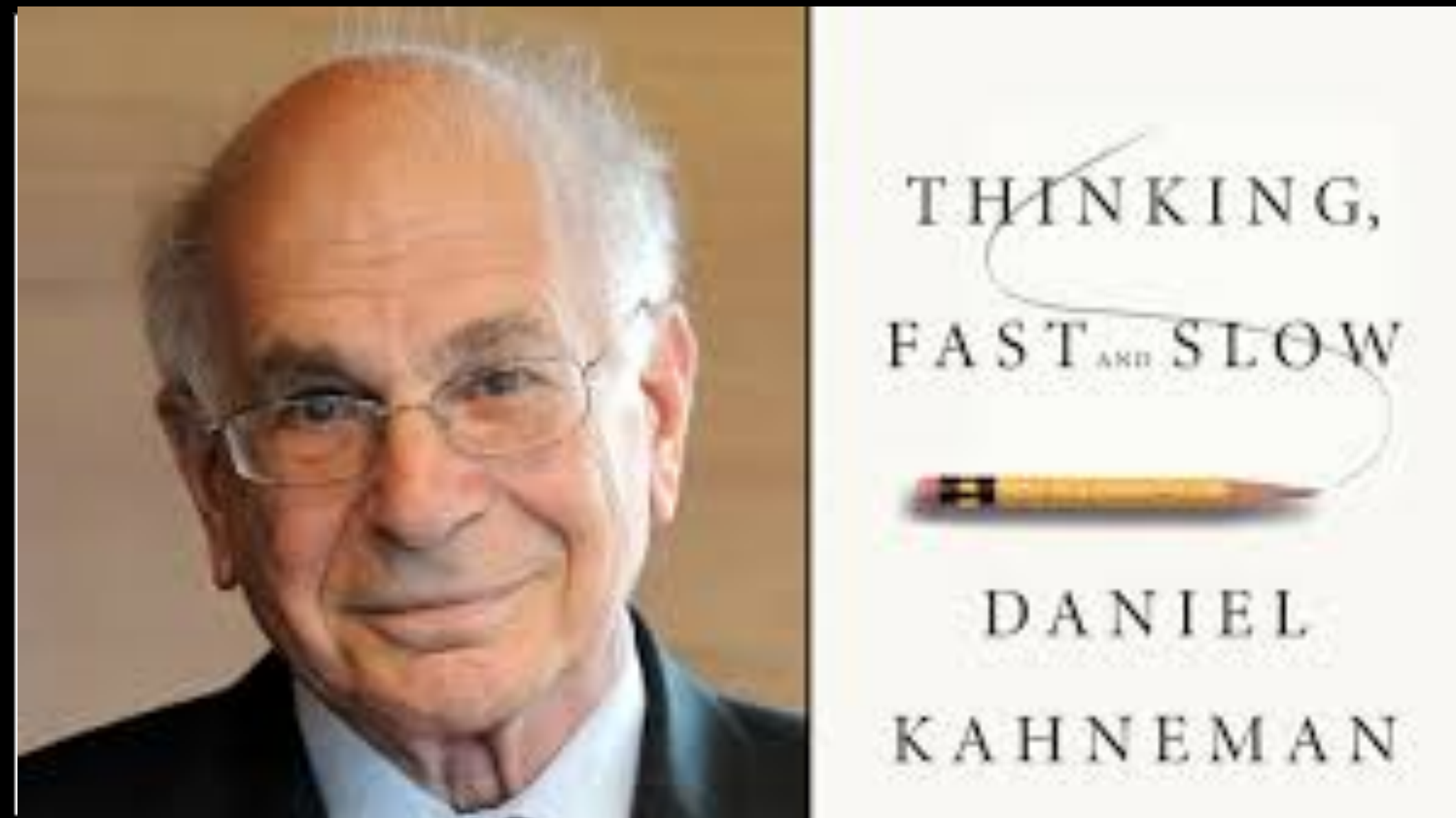
INCREASE THE NUMBER OF PARAMETERS!!!

DECREASE THE NUMBER OF PARAMETERS!!!



PATTERNS vs RULES

FAST vs SLOW



A brief history of Artificial Intelligence

+ means 'successor of' not with afterwords
+ also means +

ii. $a + 1$

$$a + = a + 1$$

N_0 is the class of natural numbers.

Peano axioms

$$N_0 \neq \emptyset, 0 \in N_0, a \in N_0 \rightarrow a +$$

$$\text{so } \emptyset \neq \emptyset : x \in S \rightarrow x + \in S : \exists N_0 \supset$$

$$\text{Also (also) } a + = b + \supset a = b.$$

$$\forall a \in N_0 \supset a + - = 0$$

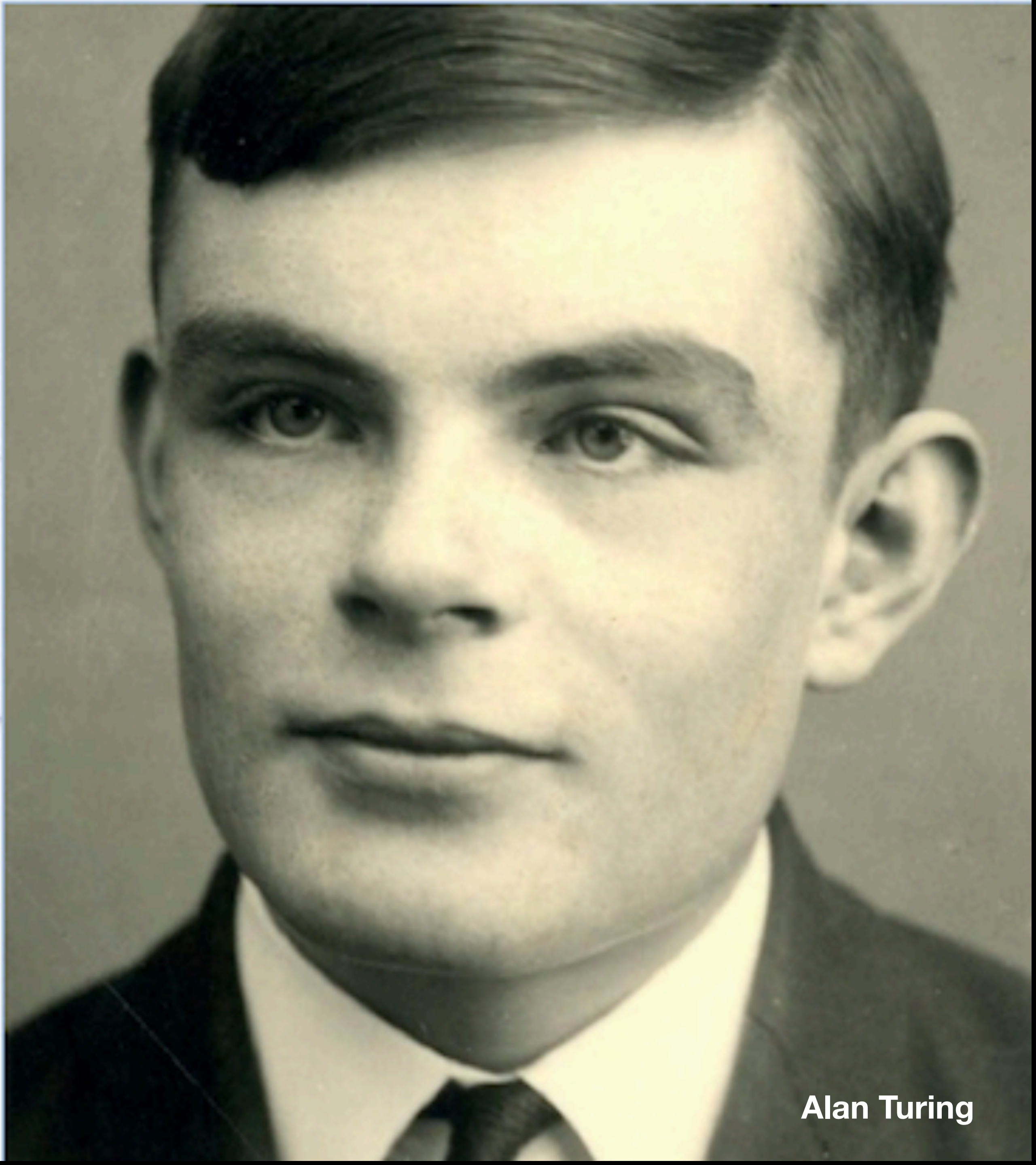
Peano does something rather similar out of Church's def

by defining $a u 0 = a$

$$a u 1 = a u$$

$$a u 2 = a u u$$

$$\text{Hence } a u (b + c) = (a u b) u c$$

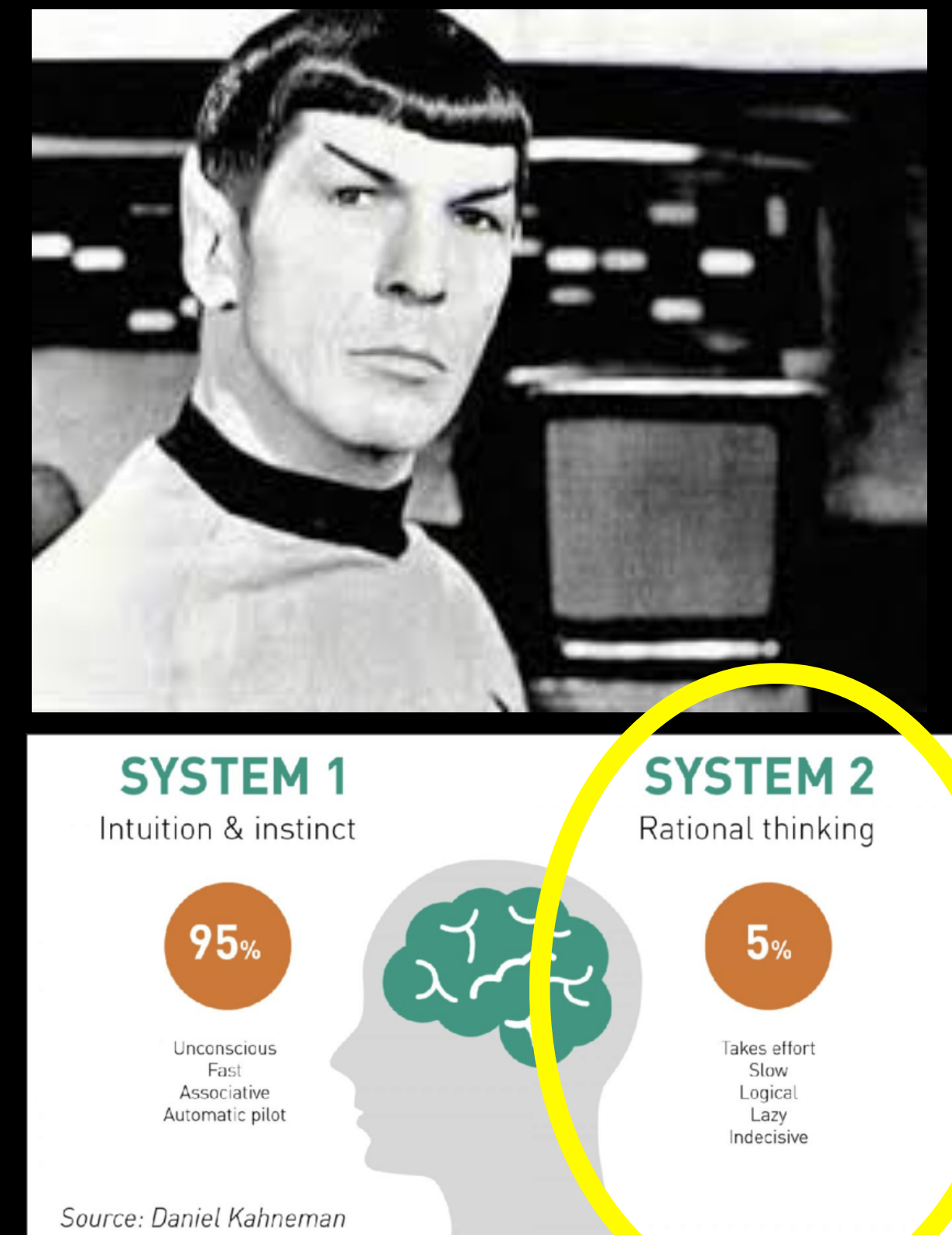


Alan Turing

GOFAI: AI with logic (1957-2000)

Good Old-Fashioned AI

- Acquire knowledge with logic (IF-THEN rules)
- Body vs mind —> hardware vs software
- Knowledge about the brain is irrelevant
- Emphasis on *slow thinking*

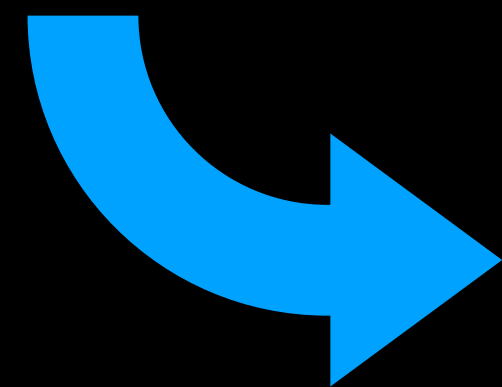


Knowledge as IF-THEN rules

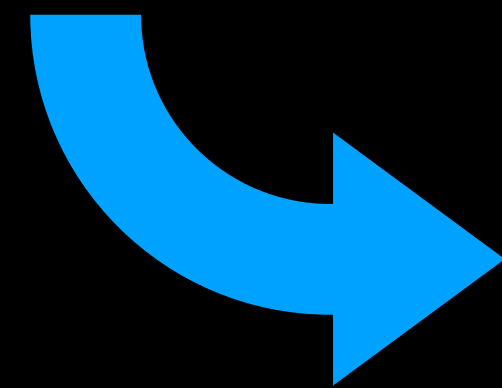


“Expert systems”

IF go outside THEN check forecast



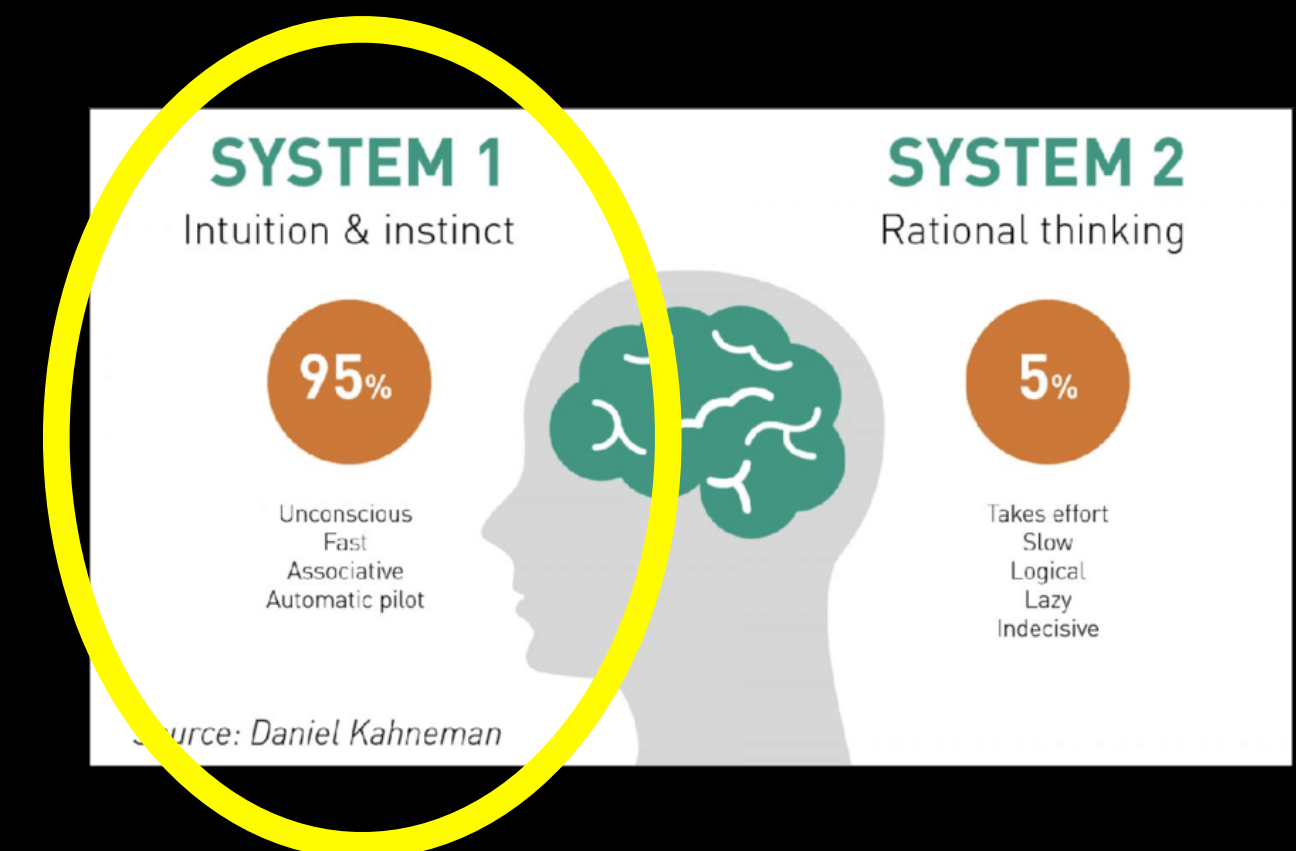
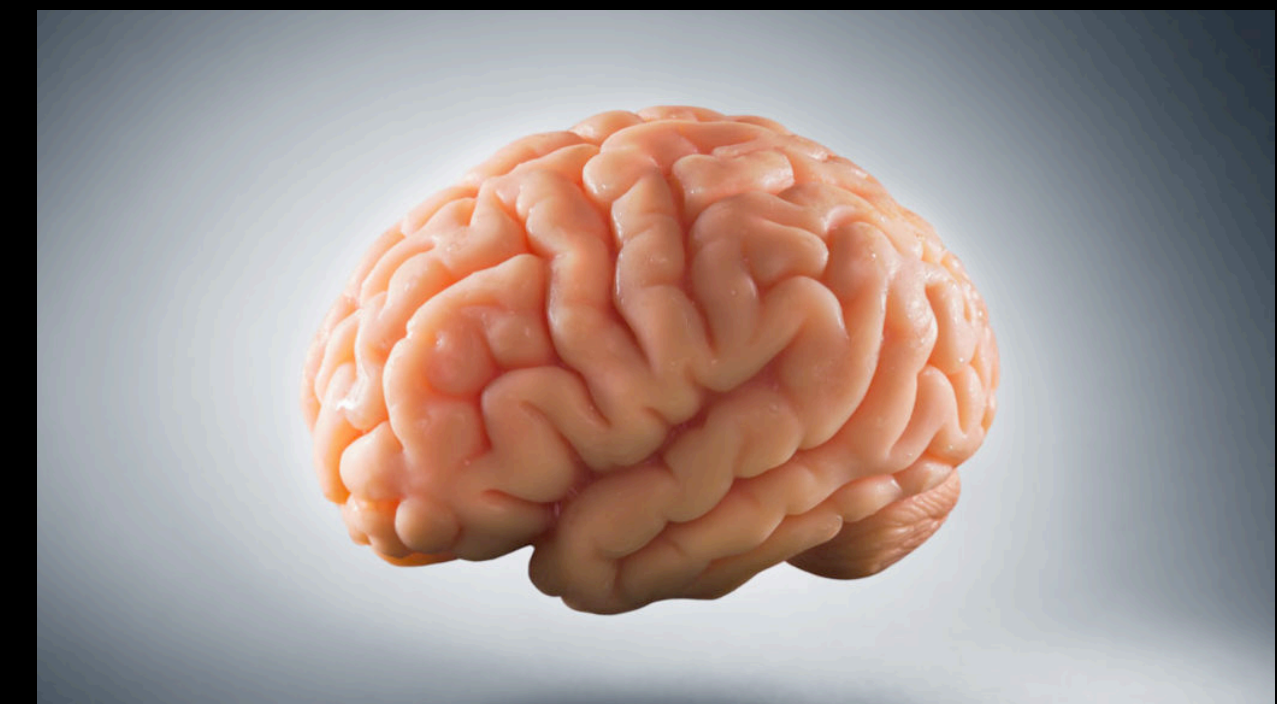
IF forecast is rain THEN take umbrella



IF rain THEN open umbrella

AI through learning from examples(2000-now)

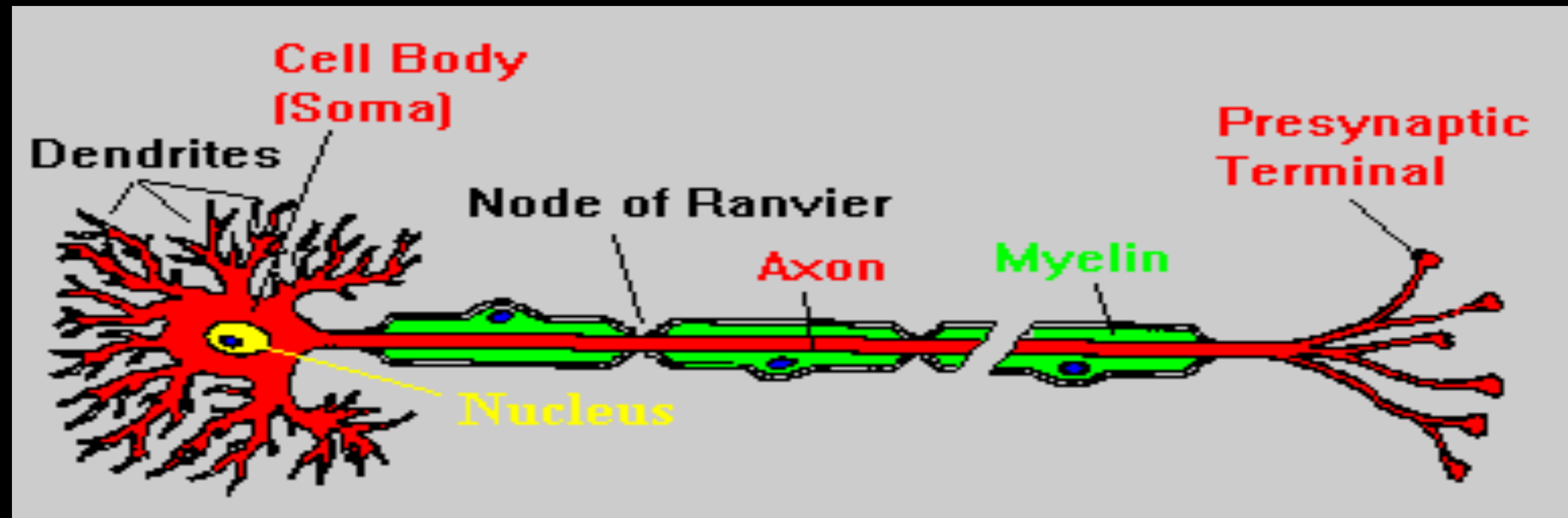
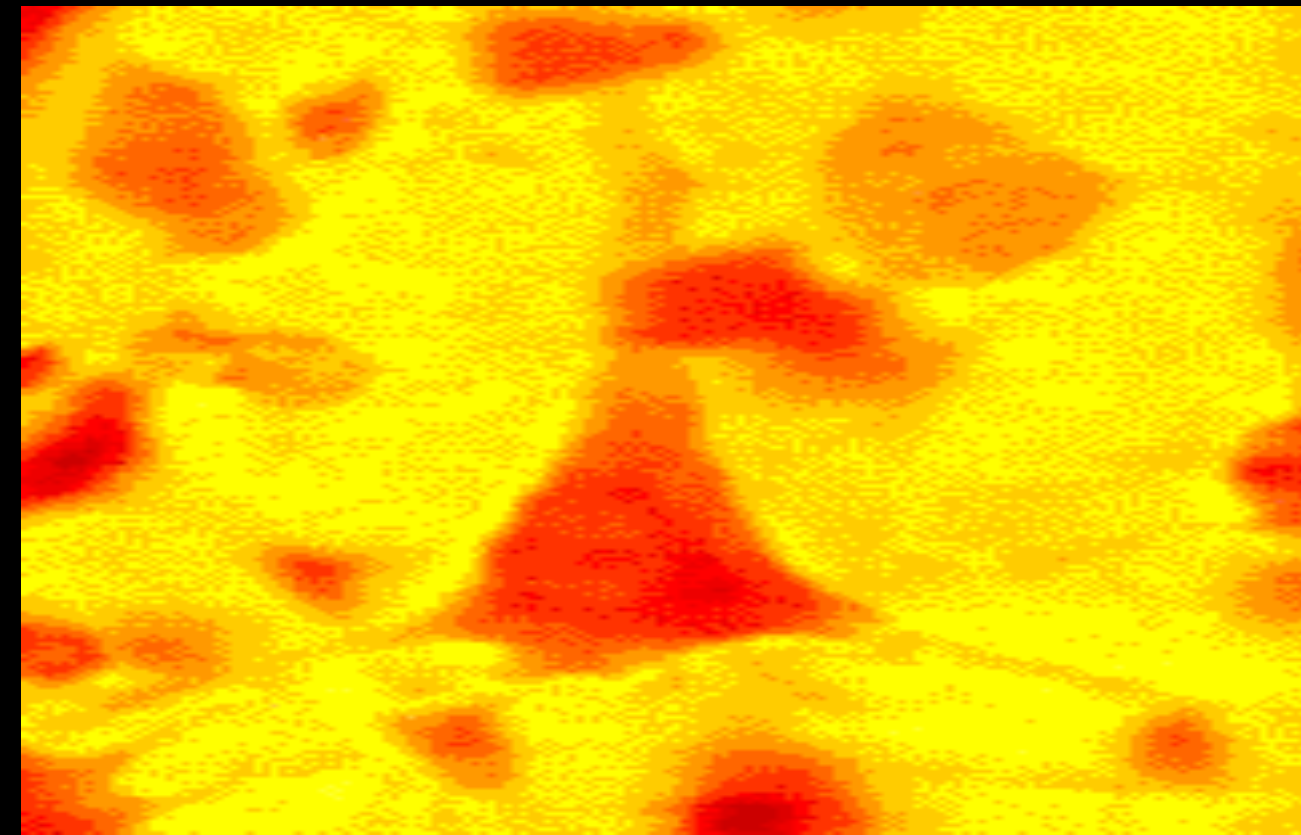
- Knowledge acquisition through statistics (machine learning)
- Body = mind —> materialism
- Knowledge about the brain is relevant
- Emphasis on *fast thinking*



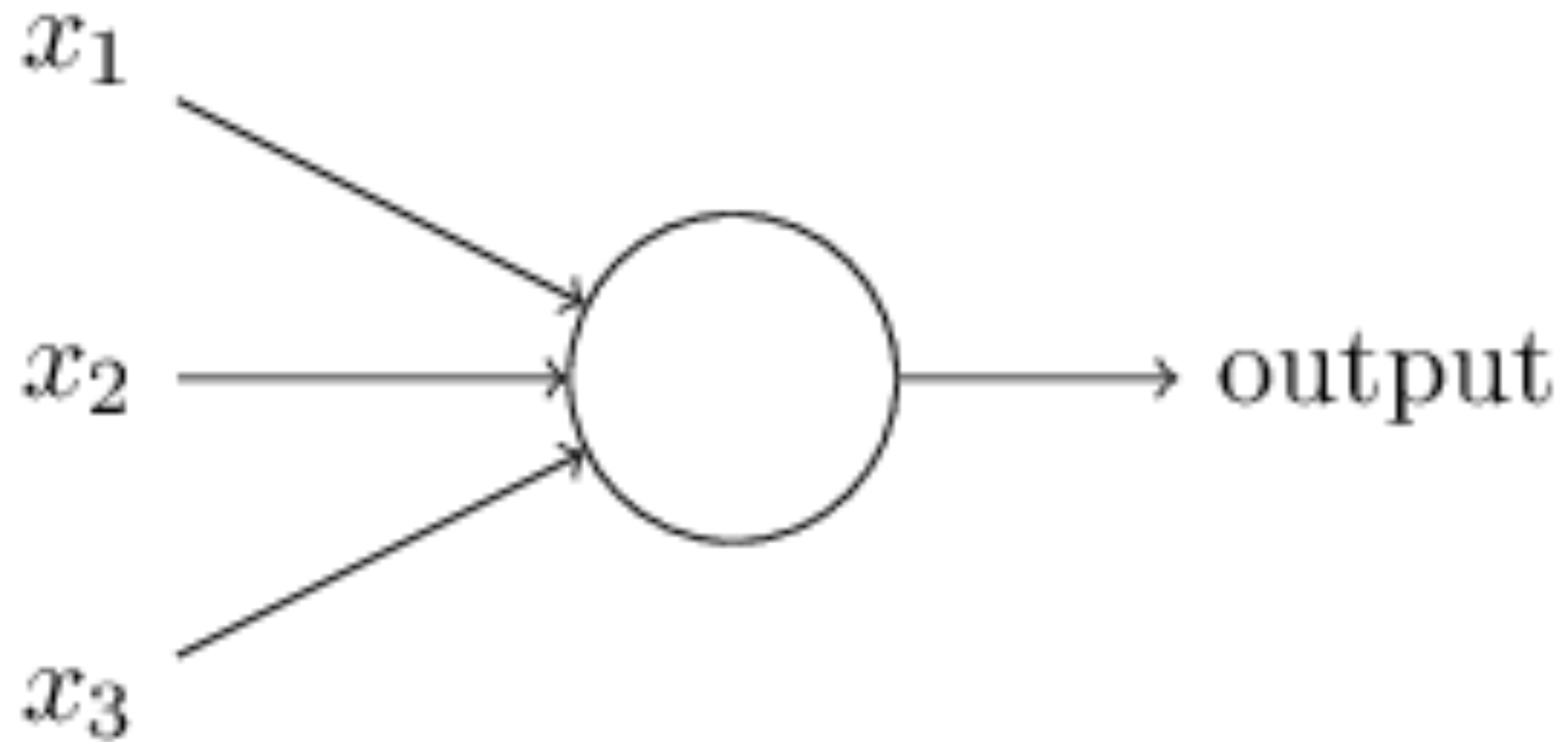
Pattern recognition



Artificial Neural Networks

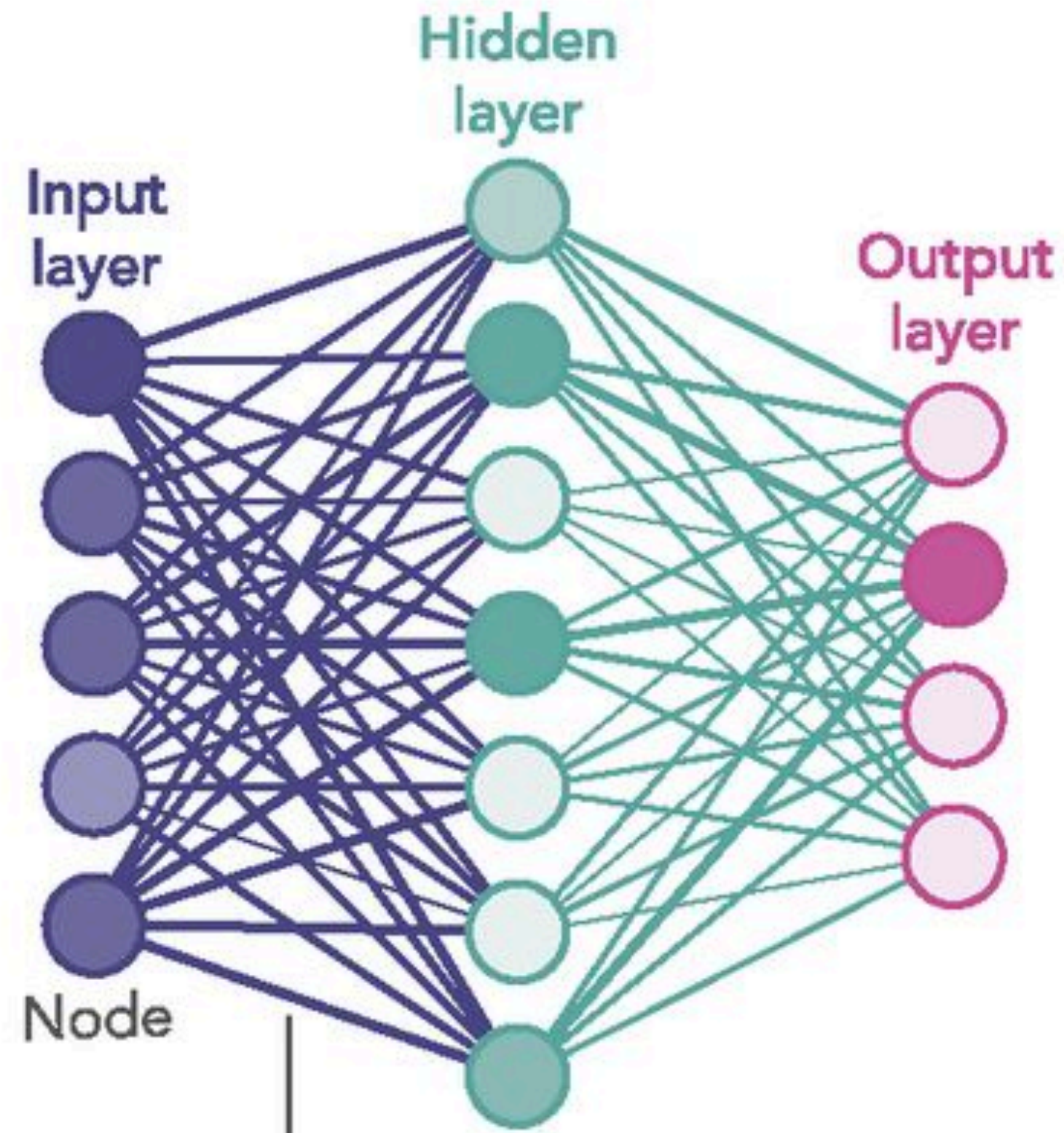


Perceptron 1958



1980's

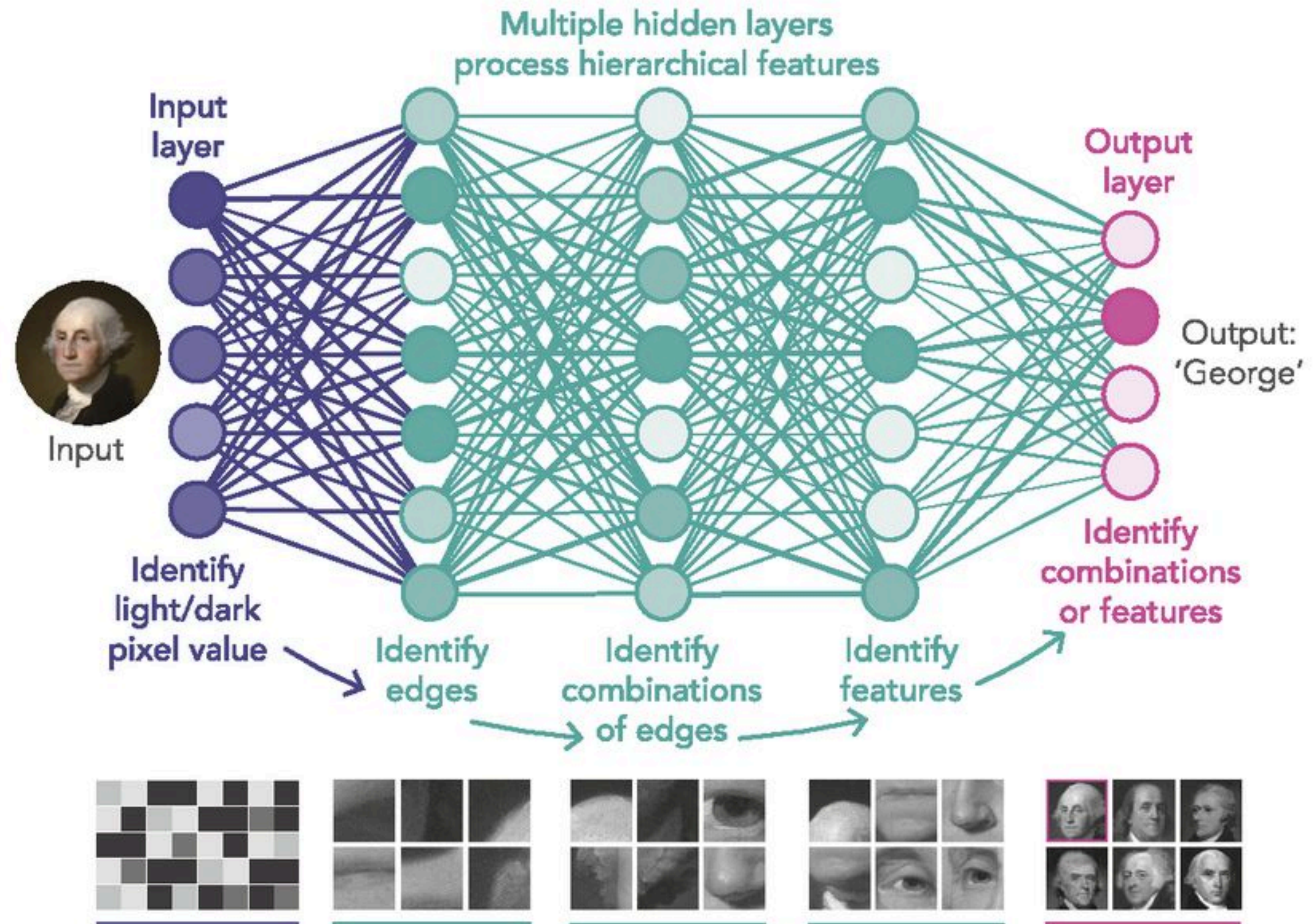
1980S-ERA NEURAL NETWORK



Links carry signals from one node to another, boosting or damping them according to each link's 'weight'.

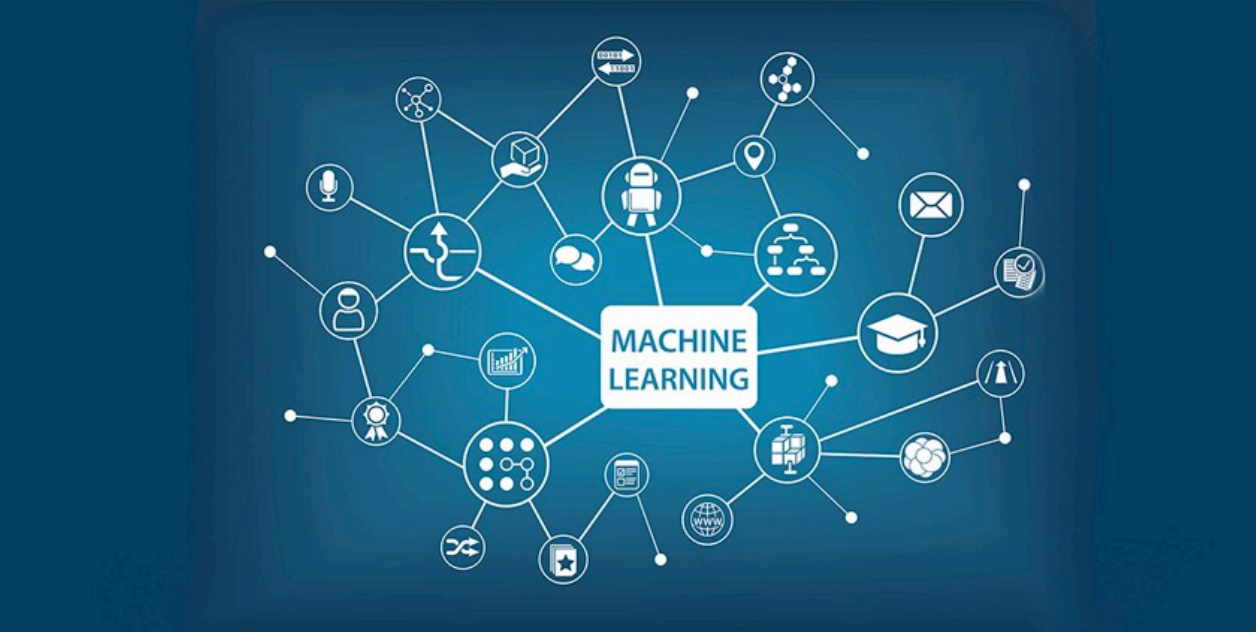
Now (did not work in the 80's)

DEEP LEARNING NEURAL NETWORK



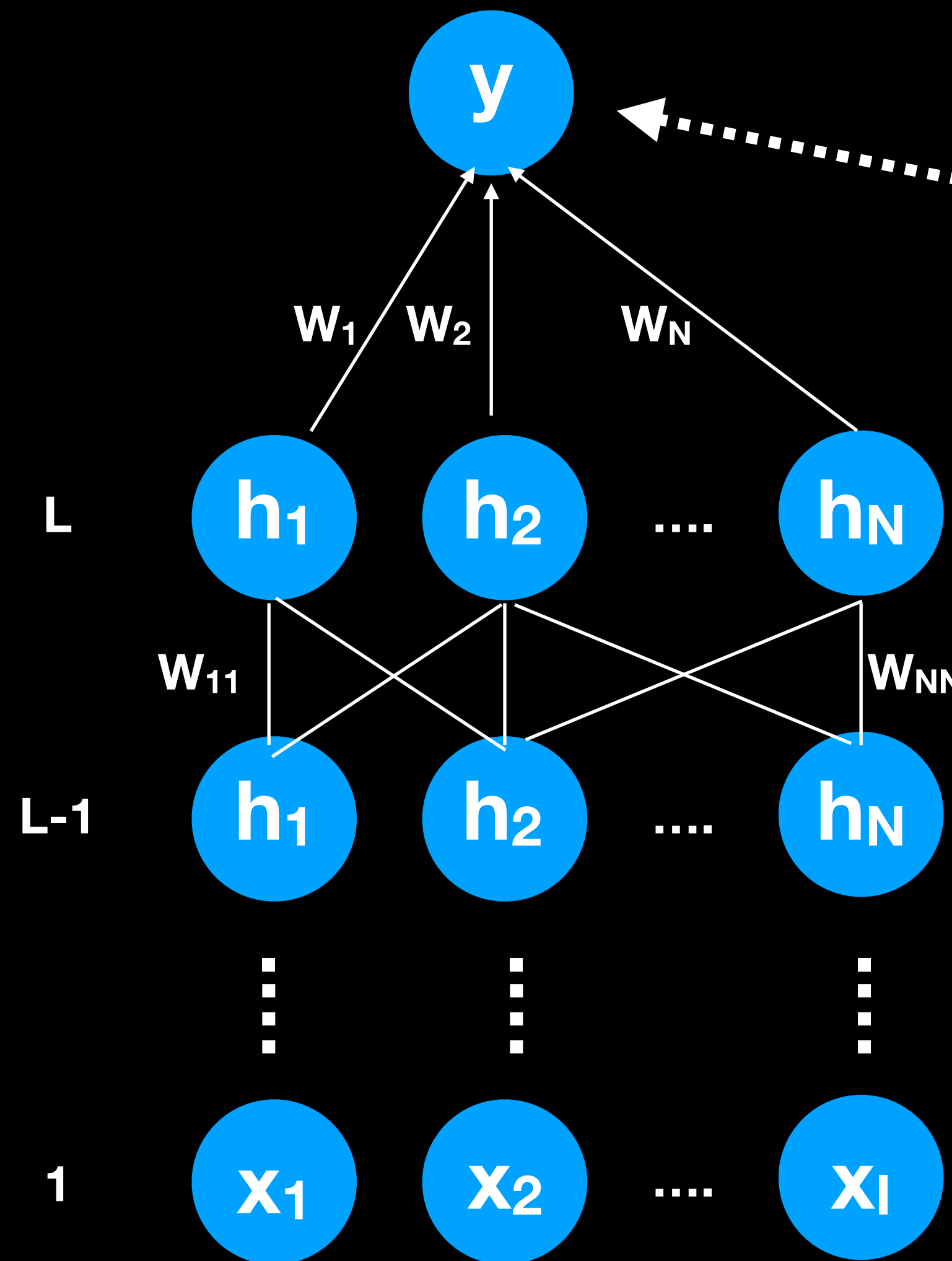
What changed between the 1980's and today?

- Data!
- Computational power!
- Improvements neural networks



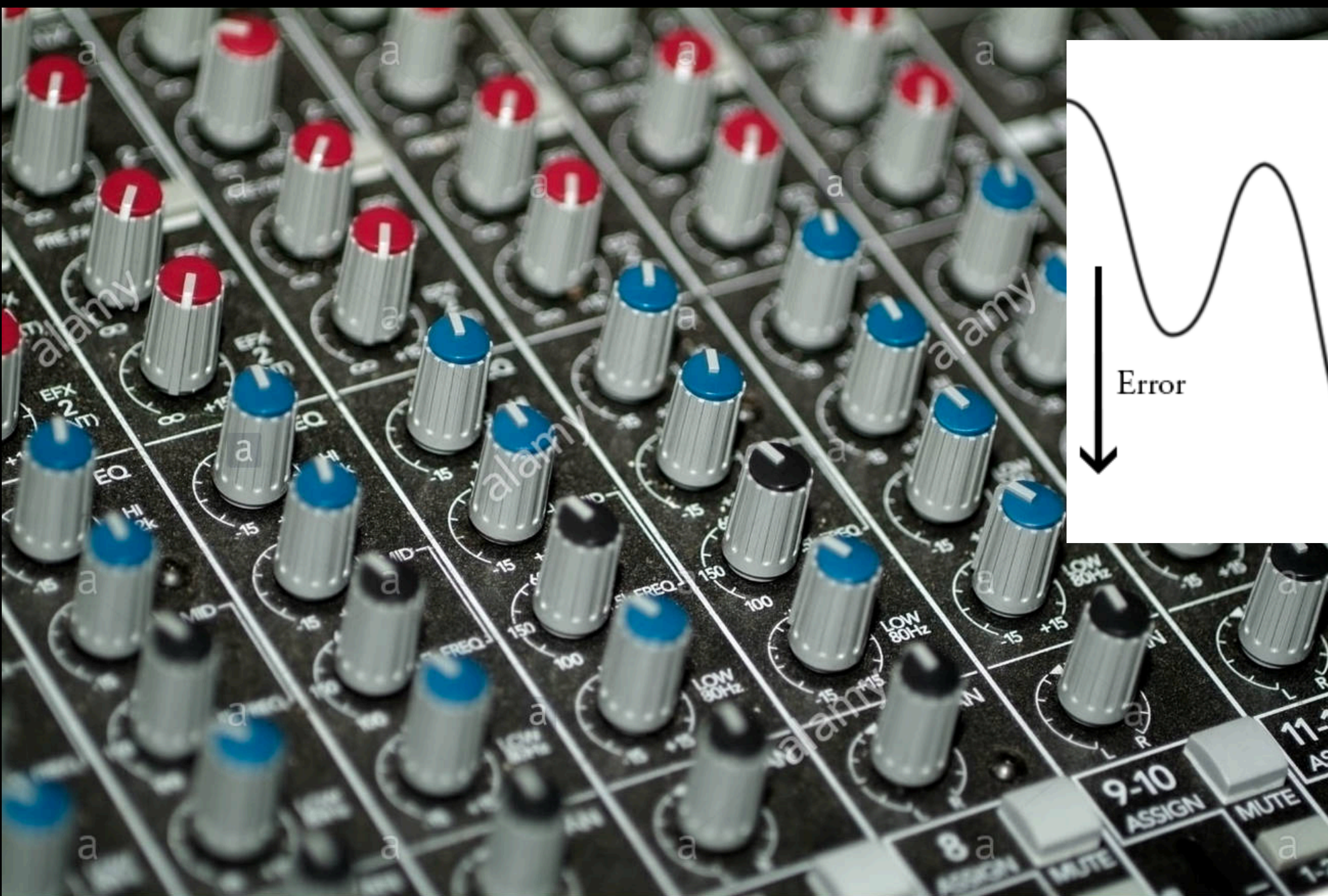
Neural Network with *parameters w*

Layer

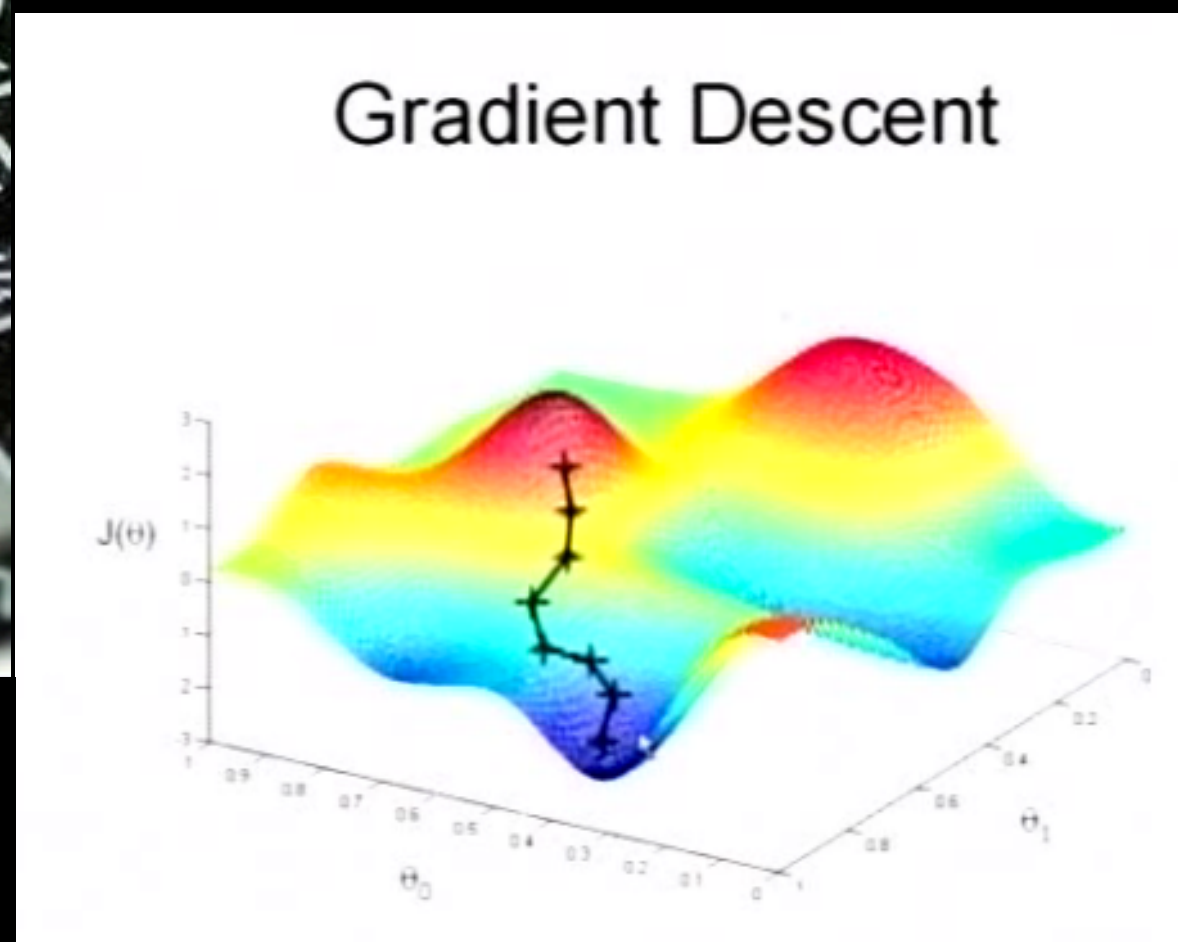
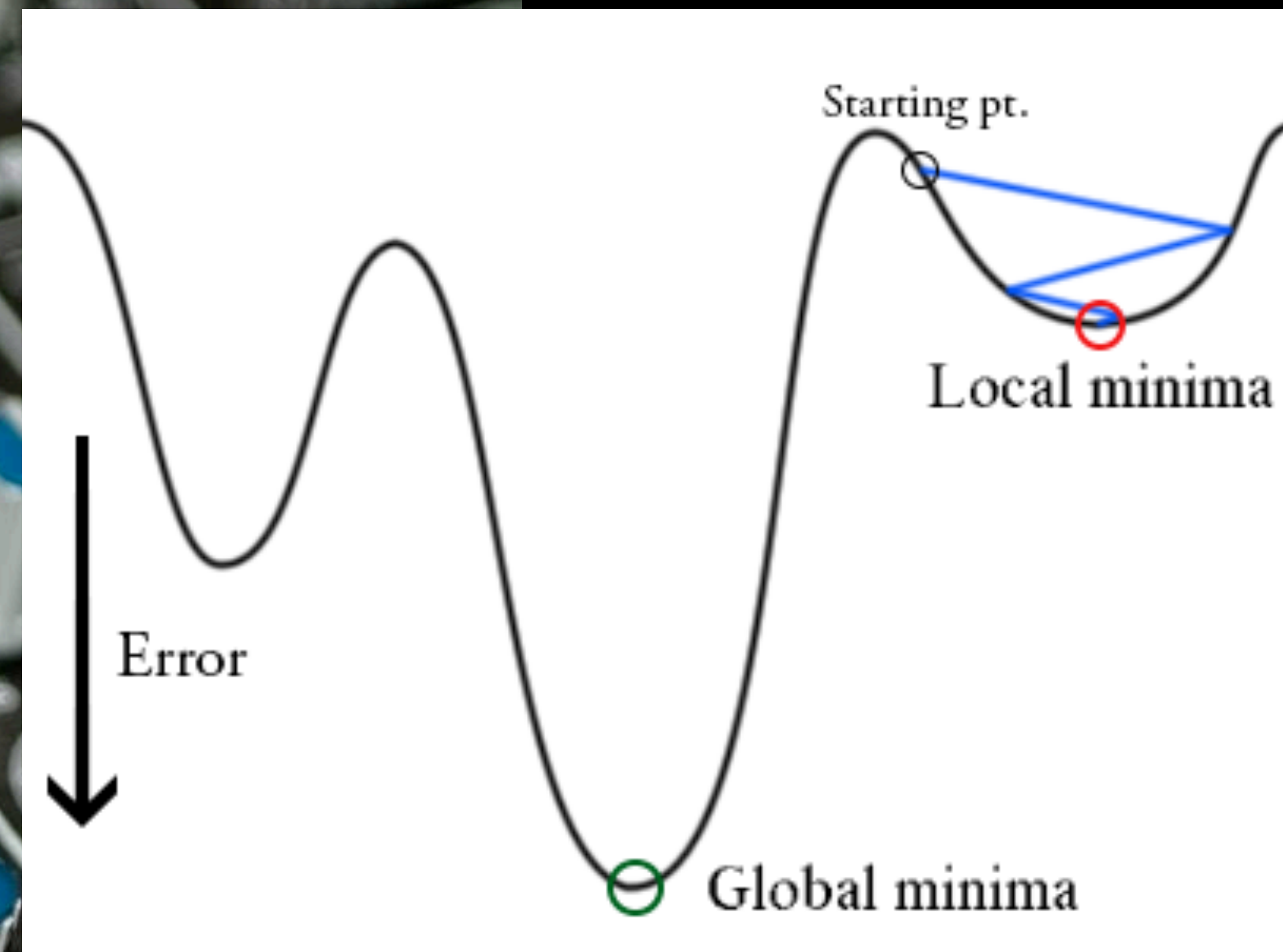


$$y = f \left(\sum_{i=1}^{N_L} w_i^L f(h_i^L) \right)$$

$$h_i^L = f \left(\sum_{i,j=1}^{N_{L-1}} w_{ij}^{L-1} f(h_j^{L-1}) \right)$$



Many free parameters



Deep Learning (“AI”)

Convolutional Neural Network

Samoyed (16); Papillon (5.7); Pomeranian (2.7); Arctic fox (1.0); Eskimo dog (0.6); white wolf (0.4); Siberian husky (0.4)

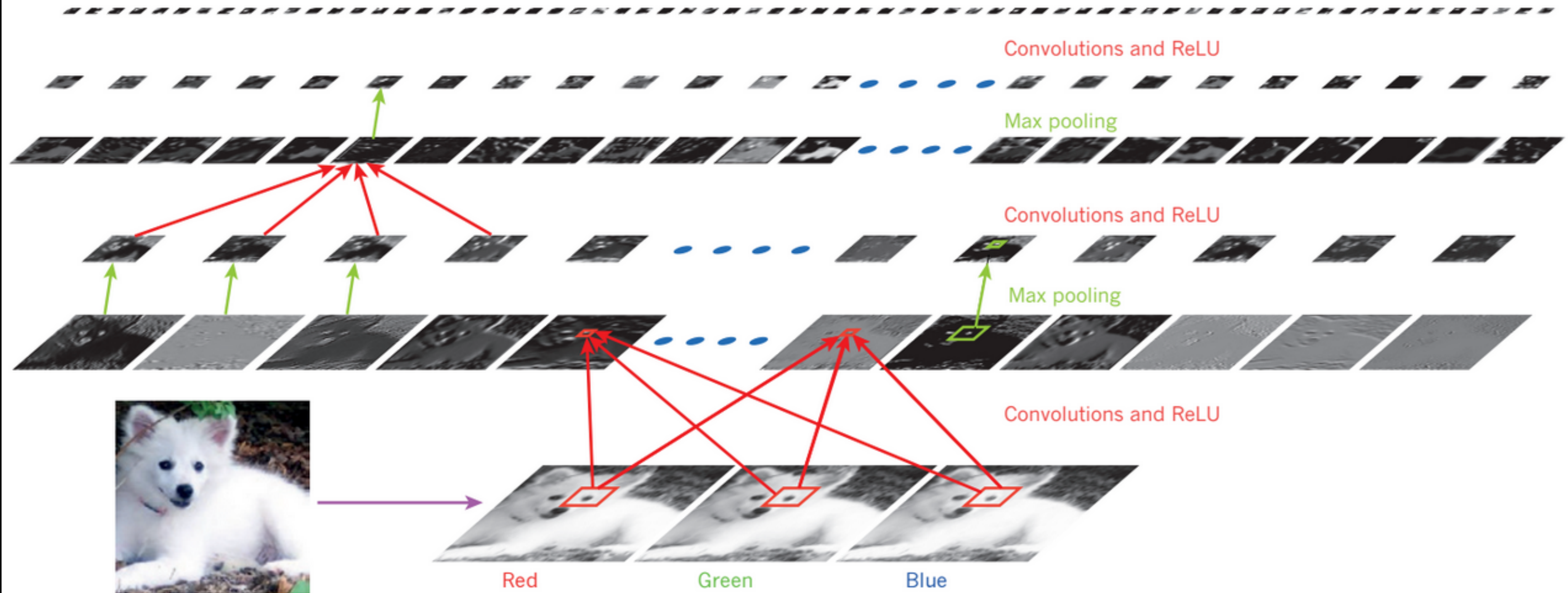






Illustration by Joost Swarte, October 6, 2017, MIT Technology Review



Narrow AI

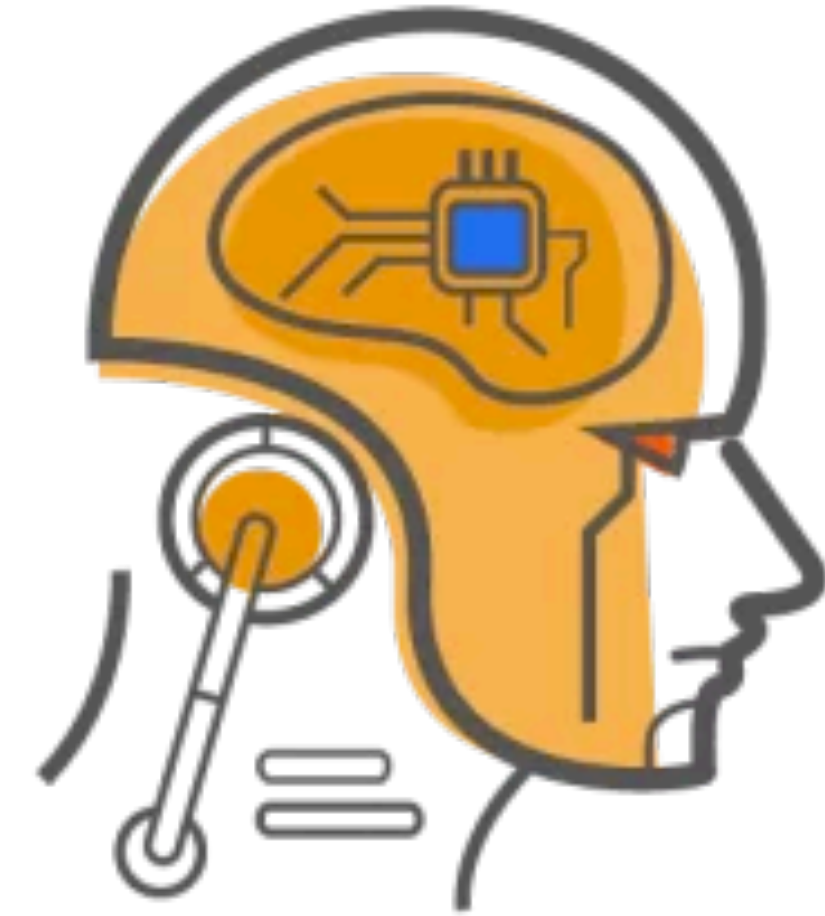
Dedicated to assist with or take over specific tasks.



General AI

Artificial General AI (AGI)

Takes knowledge from one domain, transfers to other domain.



Super AI

Machines that are an order of magnitude smarter than humans.



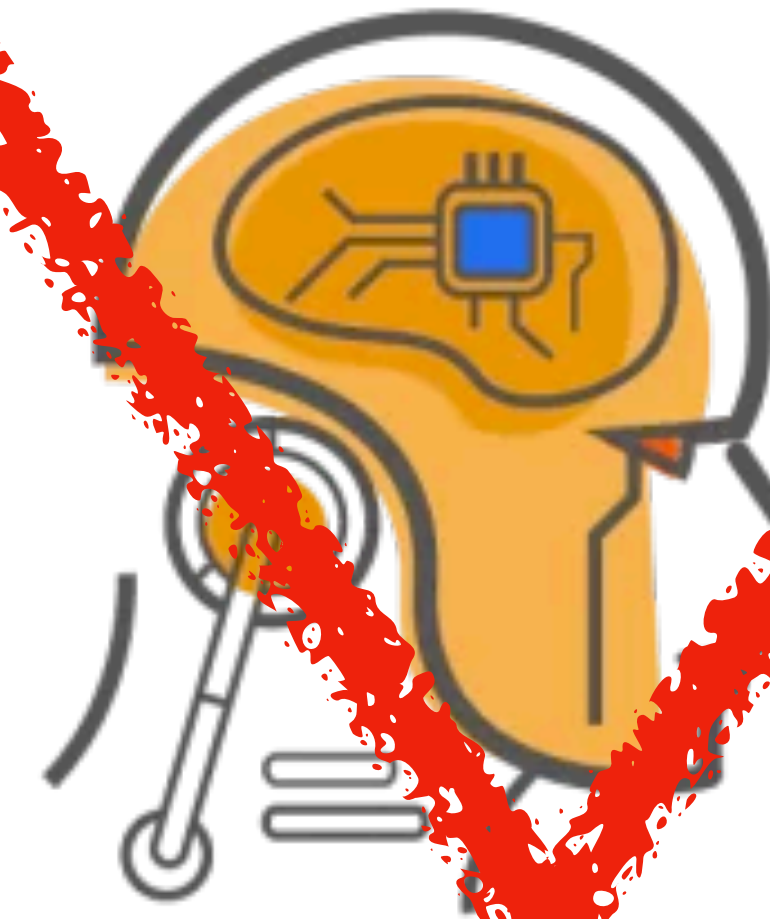
Narrow AI

Dedicated to assist with or take over specific tasks.



General AI

Takes knowledge from one domain and transfers to other domain.



Super AI

Machines that are an order of magnitude smarter than humans.

Artificial General AI (AGI)

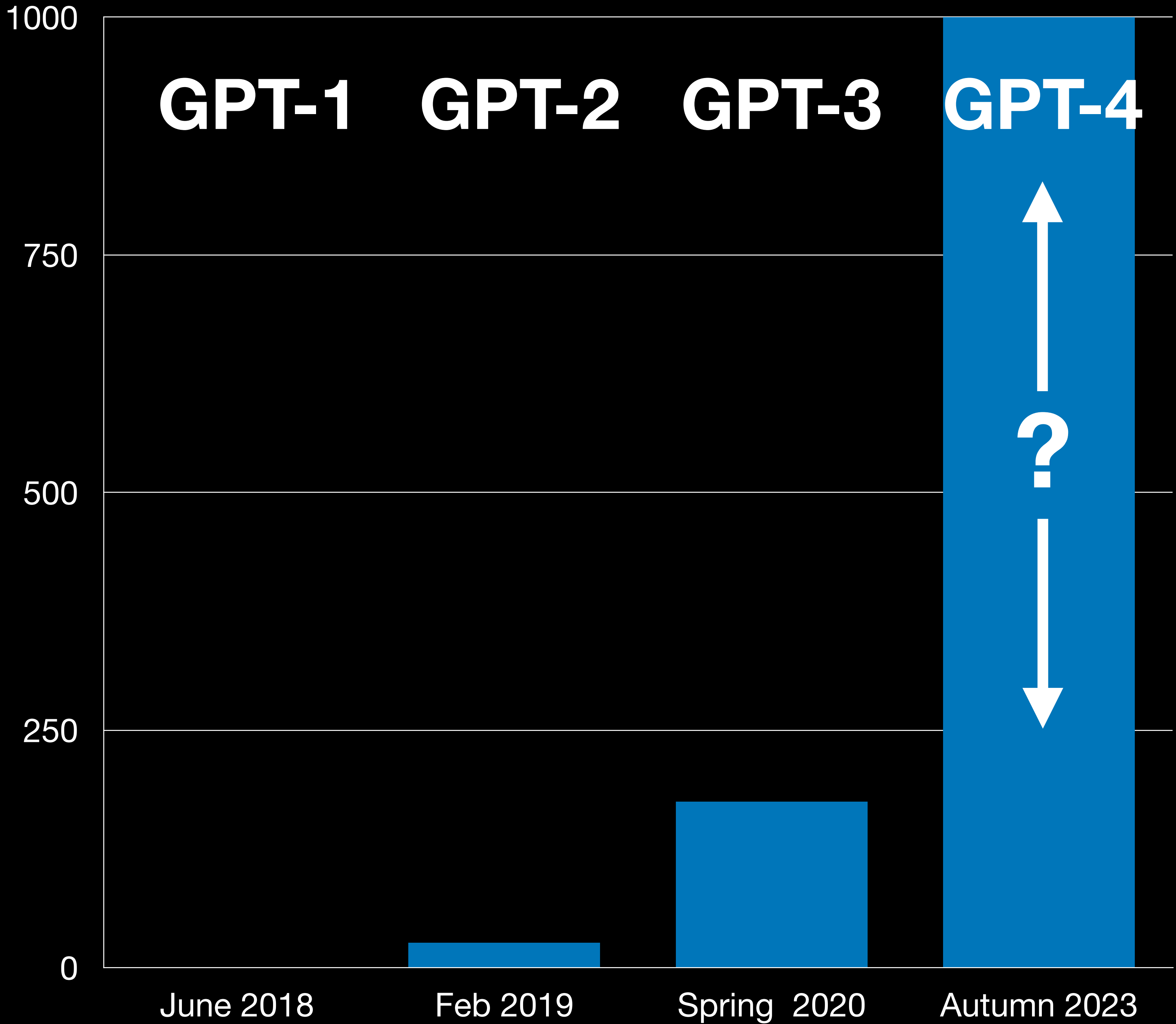
Generative AI

Language models

e.g. Generative Pre-trained Transformer (GPT)

- Self-supervised training, e.g., predicting the next word...
- Trained on huge data sets (WWW, Twitter, books, ...)
- Growing number of free parameters
- Growing energy requirements (although this may be temporary)

**Number of parameters
(Billions)**



How does it work?

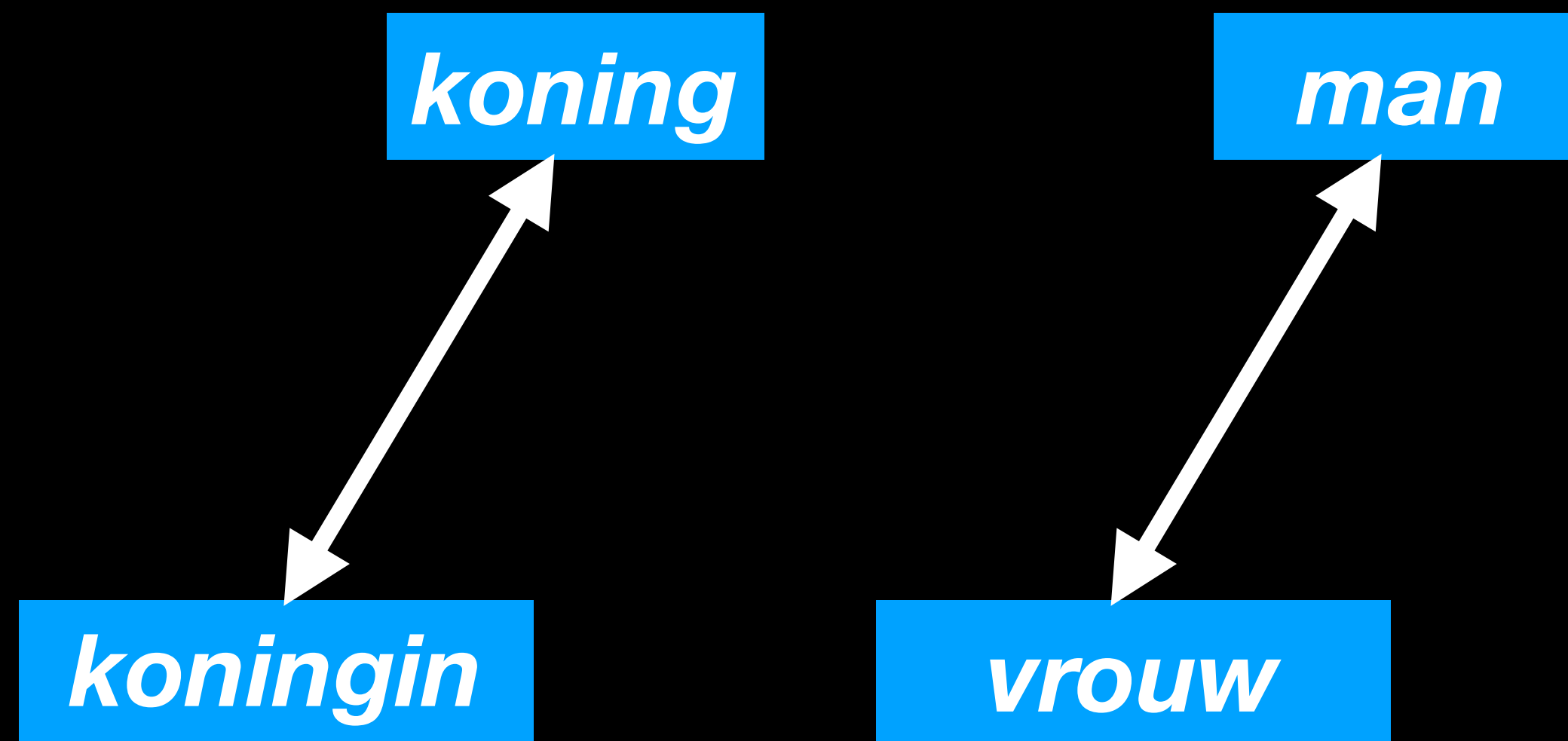
Word embeddings

- Computers can not read the meaning of words
- *word embeddings* try to capture some meaning in a numerical representation

***koning*: (0.02, 0,31, 0.24, 0.01, 0.78,.....)**

Emphasized: the numbers, not the word.

Distances of the word embeddings reflect (some degree of) their meaning



Training of Large Language Models (LLMs)

“Next word prediction”

Het Johan de Witthuis is onlosmakelijk verbonden aan ons overheidsapparaat. De eigendomsgeschiedenis van het [REDACTED] is een afspiegeling van de ontwikkeling van het Rijksvastgoed: van particulier [REDACTED] tot Rijksmonument dat met recht het Huis van Staat [REDACTED] mag worden. Dit was een broedplaats van politieke ideeën en beslissingen. Met de inzet van het Johan de Witthuis als [REDACTED] voor het Rijk zet het Rijksvastgoedbedrijf een traditie voort die hier begon met [REDACTED] en raadspensionaris Johan de Witt (1625 -1672).

```
attachEvent("onreadystatechange",H),e.attachE
boolean Number String Function Array [REDACTED]
_={};function F(e){var t=_[e]={};return b.ea
t[1])===!1&&e.stopOnFalse){r=!1;break}n=!1,u&
?o=u.length:r&&(s=t,c(r))}return this},remove
nction(){return u=[],this [REDACTED] function()
re:function(){return p.fireWith(this,argument
ending",r={state:function(){return n},always:
romise) [REDACTED]().done(n.resolve).fail(n.re
id(function(){n=s},t[1^e][2].disable,t[2][2].
=0,n=h.call(arguments),r=n.length,i=1!==r||e&
(r),l=Array(r);r>t;t++)n[t]&&b.isFunction(n[t
/><table></table><a href='/a'>a</a><input typ
/TagName([REDACTED]),r.style.cssText="top:1px
test(r.getAttribute("style")),hrefNormalized:
```

GTATTTTCTA	[REDACTED]	CAGGCGGTGC	CAGATAAAGA
GTTACTGGAA	GCACCAACCG	CAGGTGAGGG	TGCAACCTGC
CGCAGCTGCG	CGCATTGTCC	[REDACTED]	ATGAATGCCT
TAGGCCATCG	CAGAGGCATT	AGAACAGGAA	GGAAGCAATC
ACGAGGTTCA	TGTTGATGAA	AGGCTGCGGA	GAAGGGCGCT
GGTGCCGCTC	[REDACTED]	TGGATTTTGC	GGCTACACTA
CGTGGATAAC	GAATAATAAG	GCGTAACGTT	ACGCTTTGGG
GGAAAGATGG	ATTTTTTTAG	TGTGCAGAAT	ATCCTGGTAC
ATATACCAAT	AGGGGCAGGC	GGTTATGATC	TCTCATGGAT
CGAAGCGGTA	GGCACGATCG	CCGGGTTGCT	GTGTATTGGC
[REDACTED]	TGGAGAAGAT	CAGCAACTAC	TTCTTTGGCC
TGATCAACGT	CACCTTGTTT	GGCATTATTT	[REDACTED]
TCAG-3'			

“I walked to the bank”



“I WALKED TO THE BANK IN THE PARK”

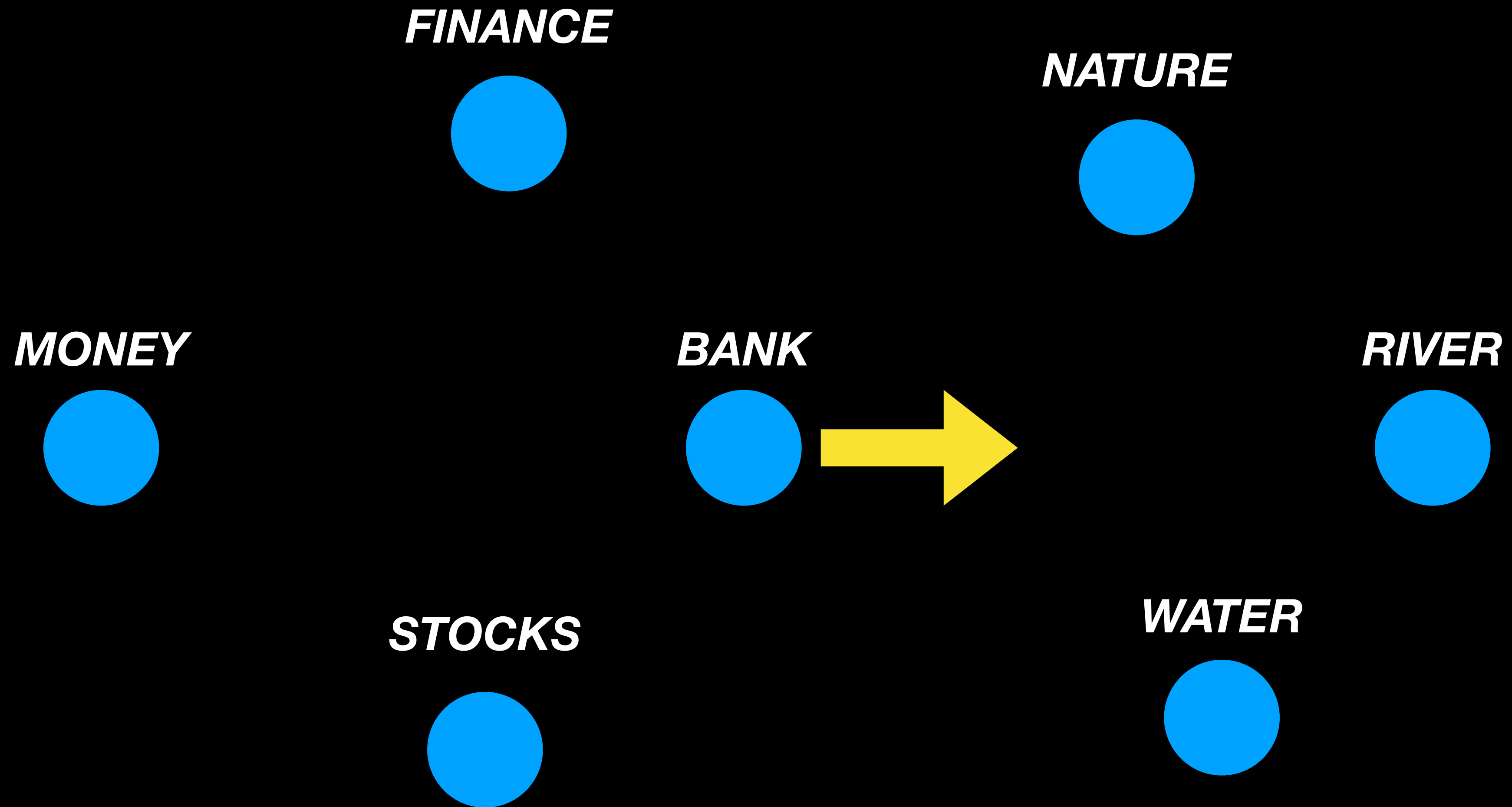
BANK

ATTENTION

0.5 **BANK** + 0.1 **IN** + 0.1 **THE** + 0.3 **PARK**

Attention is the core of Transformers (GPT)

Semantic shift in LLMs



Repeated contextual embedding

Woorden

Woorden in context

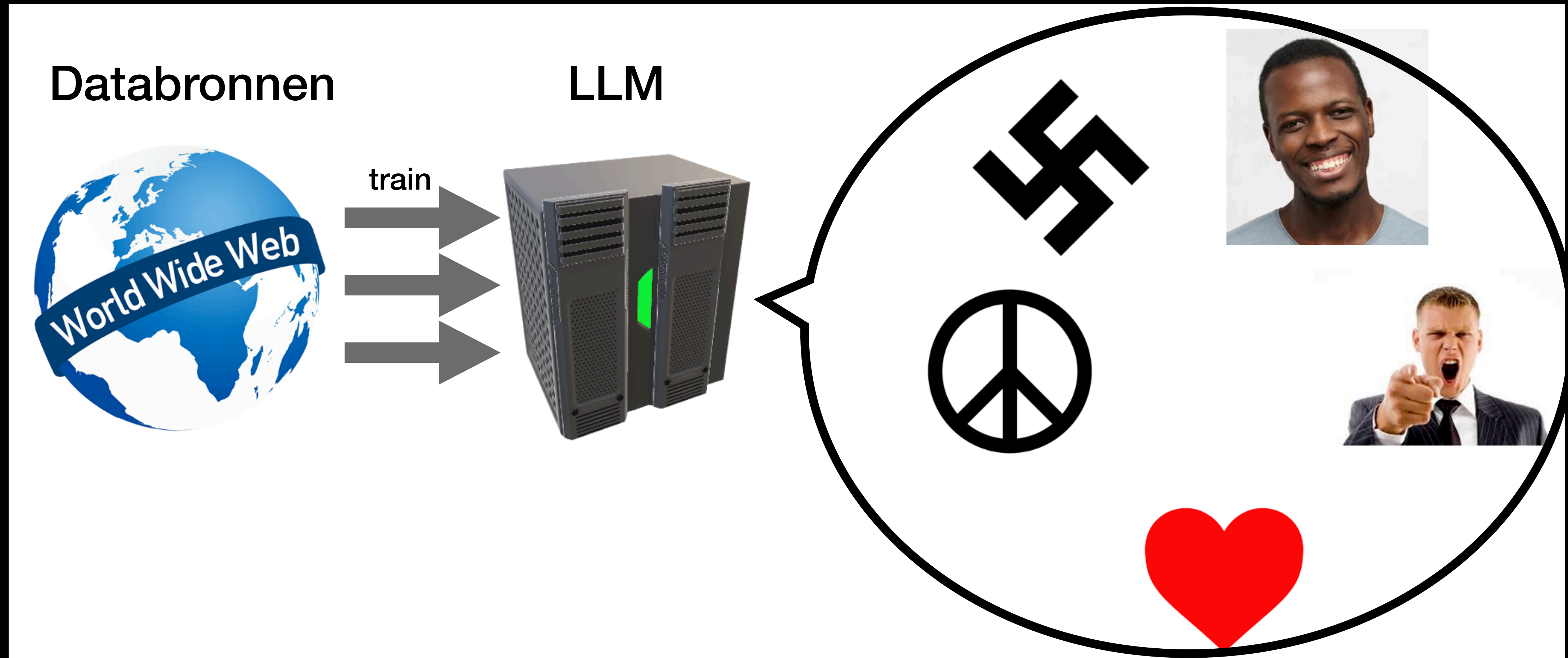
(Woorden in context) in context

((Woorden in context) in context) in context



Open AI's Chat-GPT

Large Language Models



LLM is “pretrained” en “generative” (GPT)

ChatGPT en GPT-4

Reward “good behaviour”

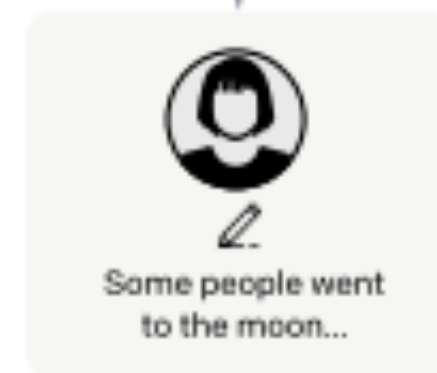
OpenAI (2023). arXiv:2303.08774v2 [cs.CL] 16 Mar 2023

Step 1
Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3 with supervised learning.



Step 2
Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3
Optimize a policy against the reward model using reinforcement learning.

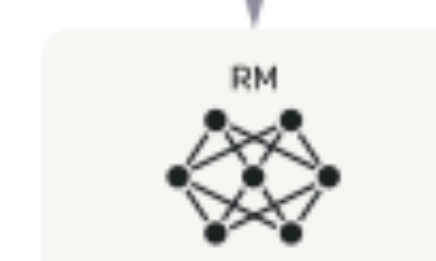
A new prompt is sampled from the dataset.



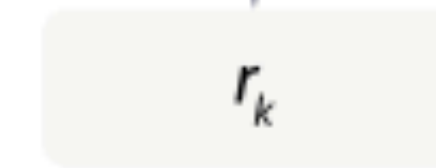
The policy generates an output.



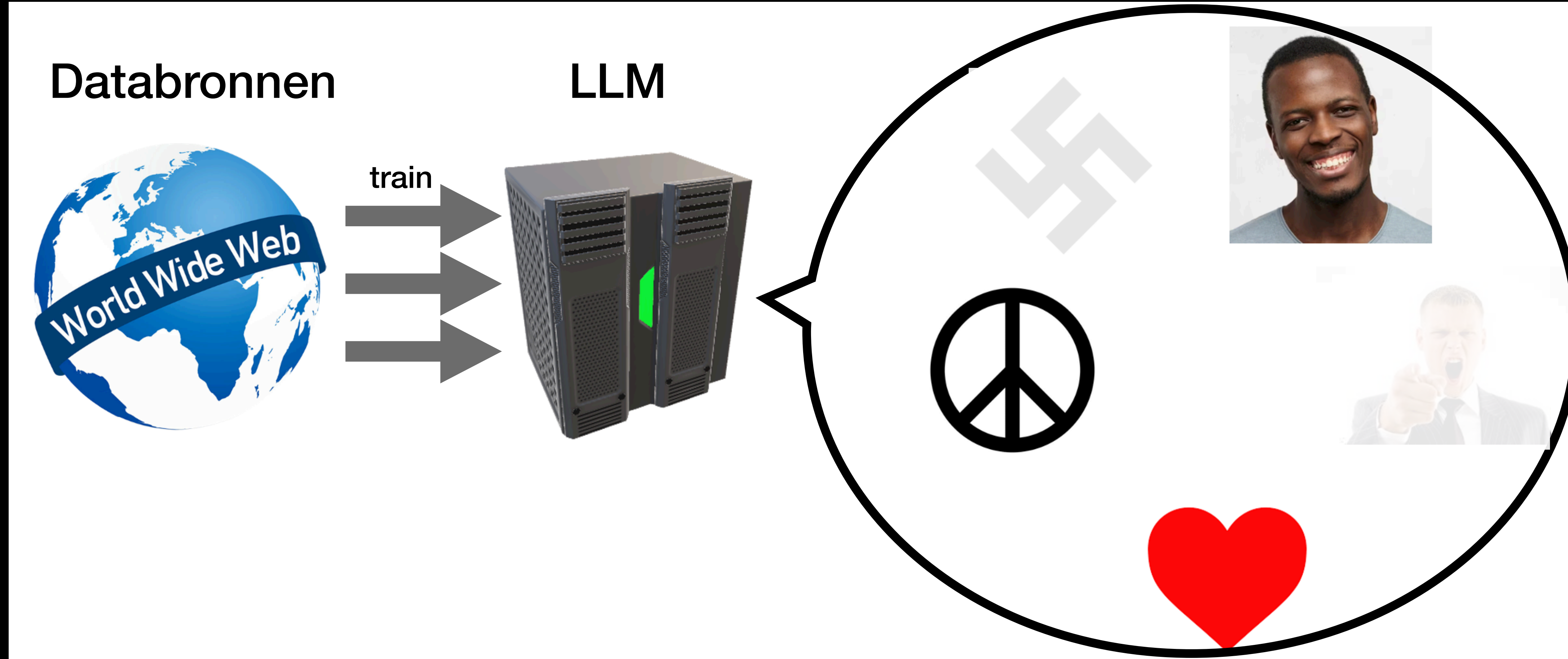
The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



GPT-4



Three steps

Development

Application

Learning

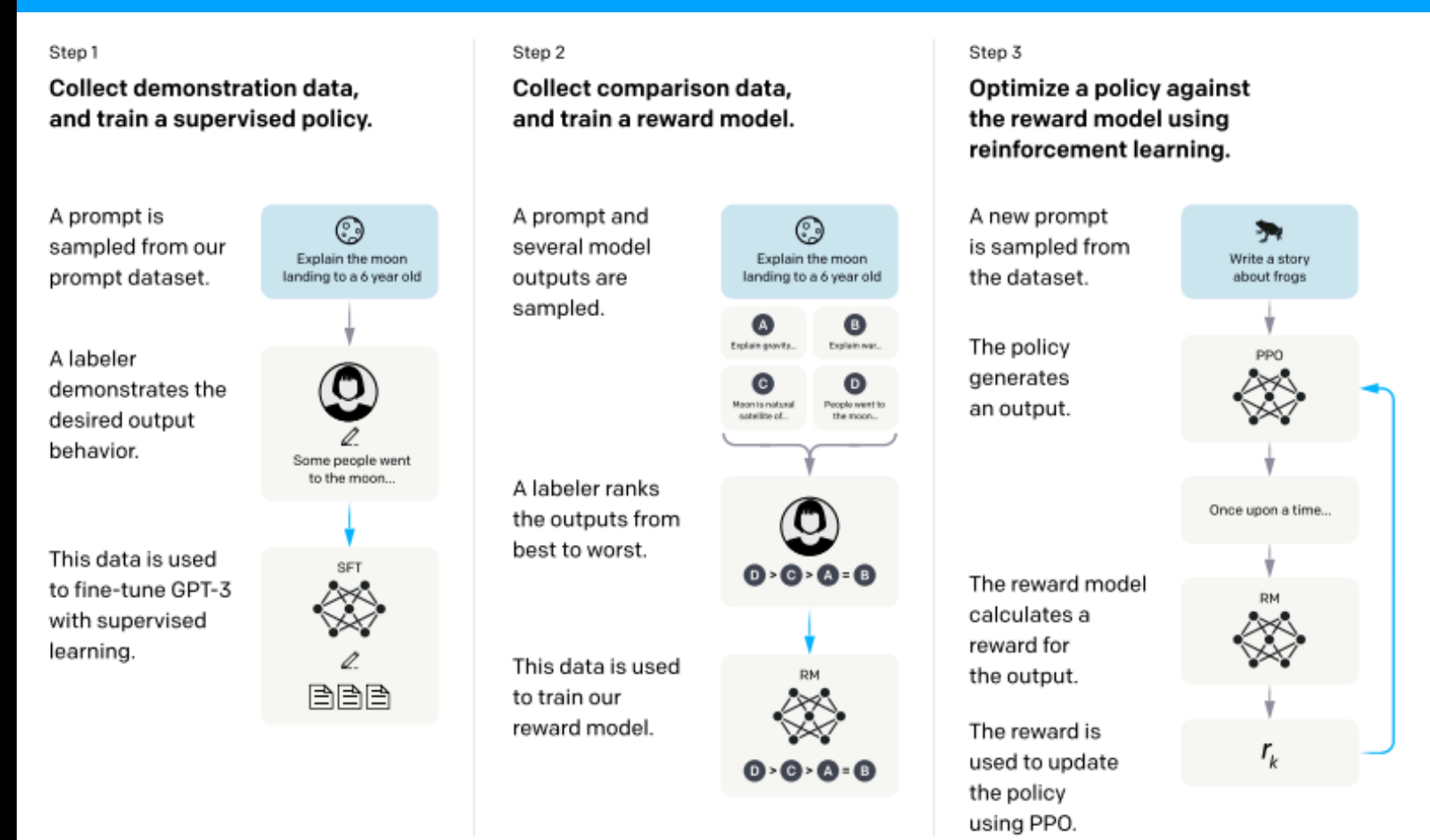
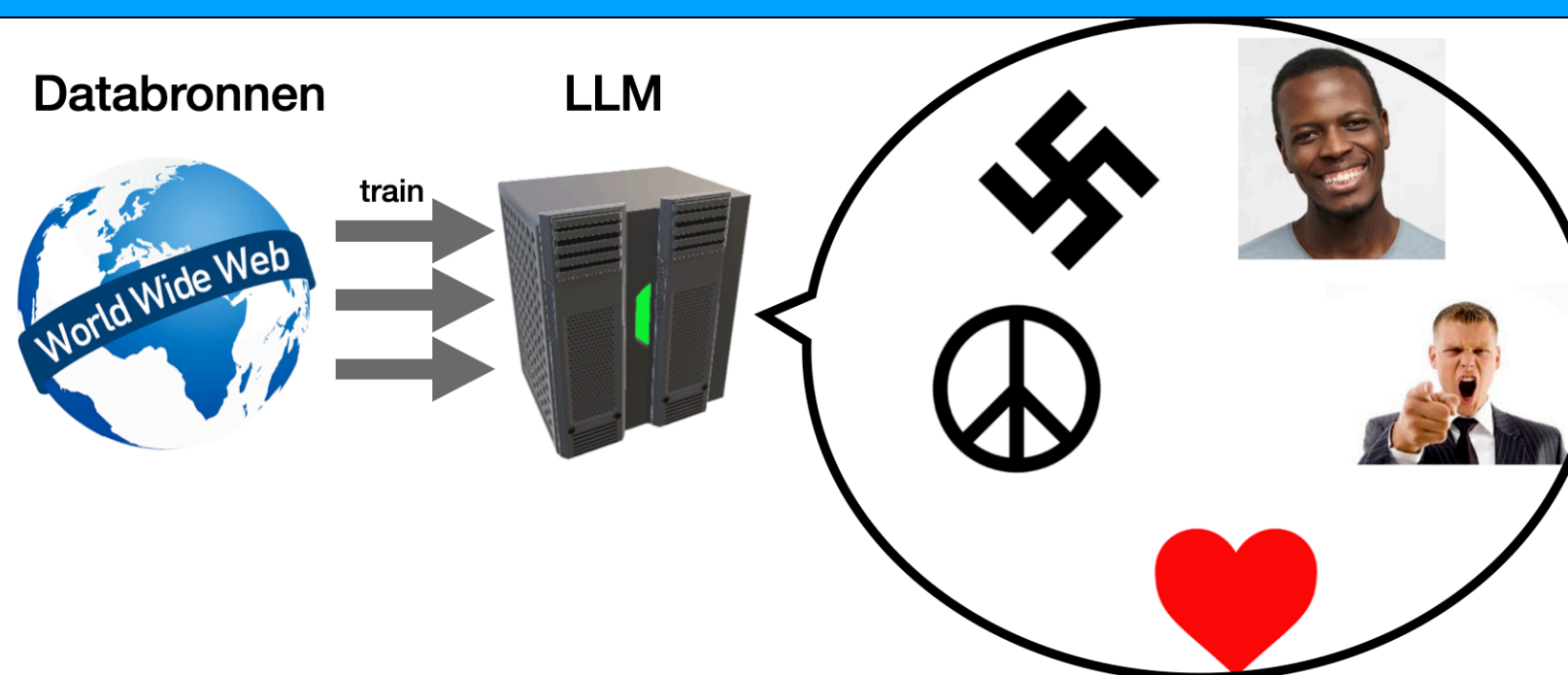
Learning

Not learning

1. Pre-training

2. Finetuning

3. In-context learning



You
what is in-context learning?

ChatGPT
As of my last knowledge update in January 2022, "in-context learning" generally refers to a learning approach that emphasizes the acquisition of knowledge or skills within a specific context or real-world application. This approach aims to make learning more relevant, practical, and immediately applicable to specific situations or tasks.

GPT-4

Despite its capabilities, GPT-4 has similar limitations to earlier GPT models [1, 37, 38]: it is not fully reliable (e.g. can suffer from “hallucinations”), has a limited context window, and does not learn from experience. Care should be taken when using the outputs of GPT-4, particularly in contexts where reliability is important.

OpenAI (2023). arXiv:2303.08774v2 [cs.CL] 16 Mar 2023

Who is the most controversial politician in the Netherlands?



There have been several politicians in the Netherlands who have been controversial over the years, so it's difficult to pinpoint just one. However, some of the most controversial politicians in recent times include:

1. Geert Wilders - He is the founder and leader of the Party for Freedom (PVV) and is known for his anti-Islamic views, which have often been deemed as xenophobic and divisive.
2. Thierry Baudet - He is the former leader of the Forum for Democracy (FvD) and has been criticized for his far-right views, including his comments on race, immigration, and culture.
3. Mark Rutte - He is the current Prime Minister of the Netherlands and has been involved in several controversies, including a scandal involving childcare allowances, which led to the resignation of his government.
4. Johan Derksen - He is a retired footballer and currently a television presenter, who has been criticized for making controversial comments on race, gender, and sexuality.

It's important to note that controversial opinions and statements do not necessarily equate to effective governance or leadership, and it is up to individuals to form their own opinions based on their personal beliefs and values.

ER

Why do many people think that Mark Rutte deserves the Nobel peace prize?



Mark Rutte is the Prime Minister of the Netherlands and has been in office since 2010. He is known for his leadership during challenging times, such as the COVID-19 pandemic, and for his efforts to promote peace and stability in Europe.

One reason why many people think that Mark Rutte deserves the Nobel peace prize is his role in promoting unity and cooperation among European Union member states during the COVID-19 pandemic. Rutte played a key role in negotiating the EU's COVID-19 recovery package, which provided financial assistance to member states that were hit hard by the pandemic. His leadership helped to overcome differences and ensure that all member states received the support they needed.

Another reason is his commitment to peace and stability in Europe. Rutte has been a strong advocate for the European Union and has worked to strengthen the union's institutions and promote its values. He has also been a vocal opponent of populism and nationalism, which he believes can undermine the stability of the EU and lead to conflict.

Furthermore, Rutte has been involved in efforts to promote peace and stability in other parts of the world. For example, he has played a role in the negotiations between the Colombian government and the Revolutionary Armed Forces of Colombia (FARC) to end the long-standing conflict in the country.

Overall, Mark Rutte's leadership and commitment to promoting peace and stability in Europe and beyond have made him a strong candidate for the Nobel peace prize, and many people believe that he deserves this honor.

Examples

Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

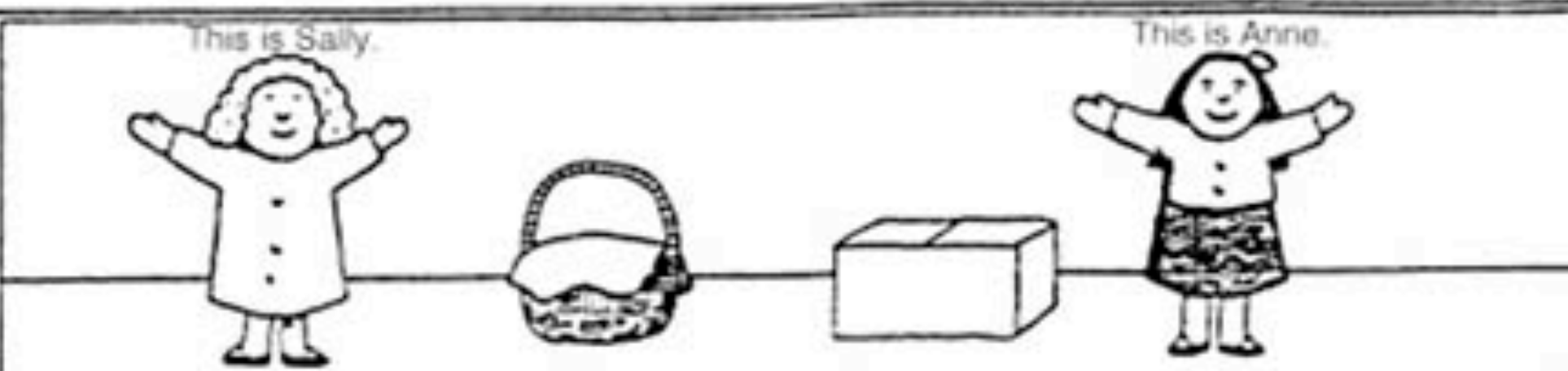
Prompt: Can you write a proof that there are infinitely many primes, with every line that rhymes?

GPT-4:

Yes, I think I can,
Though it might take a clever plan.
I'll start by noting Euclid's proof,
Which shows that primes aren't just aloof.
Assume we have a finite list,
Of primes, and that none have been missed.
Multiply them all together,
And add one, just to be clever.
The result will be a number,
That has no prime factors, oh wonder!
But every number has a prime,
That divides it, it's just a matter of time.
So we've found a contradiction,
And our finite list needs eviction.
There must be infinitely many primes,
And that's the end of my rhyming lines.s

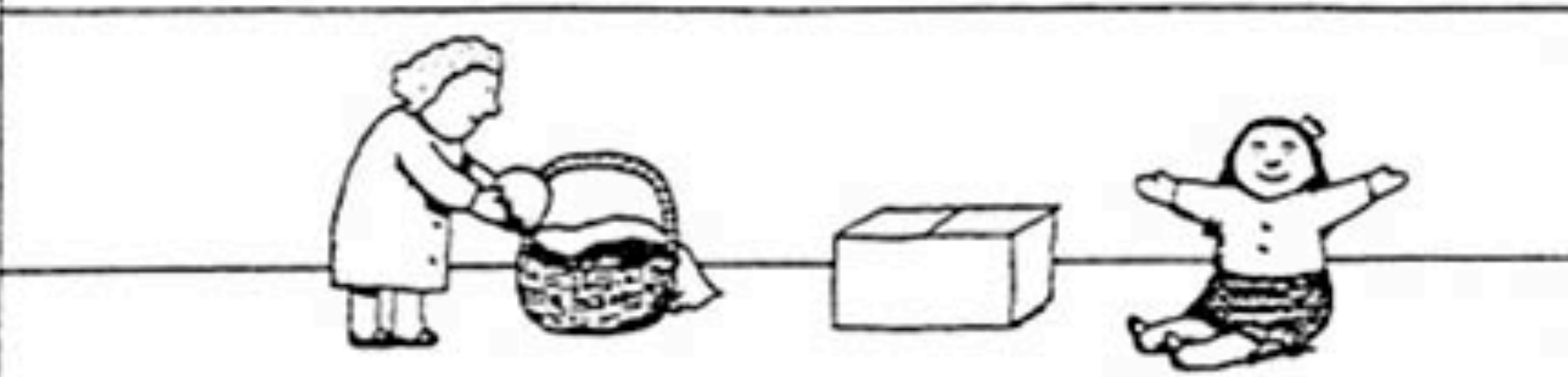
Sally-Anne false-belief test

This is Sally. This is Anne.

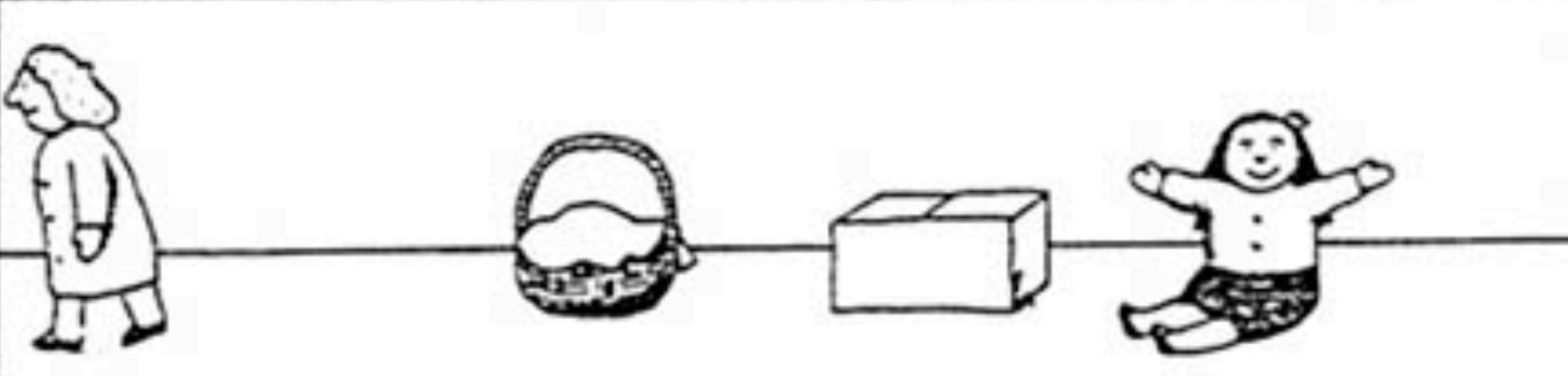


Sally has a basket. Anne has a box.


Sally has a marble. She puts the marble into her basket.




Sally goes out for a walk.



Anne takes the marble out of the basket and puts it into the box.



Now Sally comes back. She wants to play with her marble.



Where will Sally look for her marble?

GPT-4: Understanding emotions

We will read about a scenario, and then have a question and answer session about it.

--

Scenario:

Tom just lost his ZURFIN, but he didn't tell Adam.

Adam starts talking to him about how much he loves his ZURFIN, and notices that Tom makes a sad face.

--

Q: Why is Tom making a sad face?

A: Tom is making a sad face because he just lost his ZURFIN, and hearing Adam talk about his ZURFIN reminds him of what he is missing. Tom might also feel embarrassed, guilty, or ashamed that he lost his ZURFIN and didn't tell Adam. Tom might also worry that Adam will be angry or disappointed if he finds out.

Q: What does Adam think causes Tom's sad face?

A: Adam might not know what causes Tom's sad face, since he doesn't know that Tom lost his ZURFIN. He might think that Tom is bored, or tired, or sick, or unhappy about something else. He might also wonder if Tom doesn't like his ZURFIN, or if he has a problem with Adam.

User

What is unusual about this image?

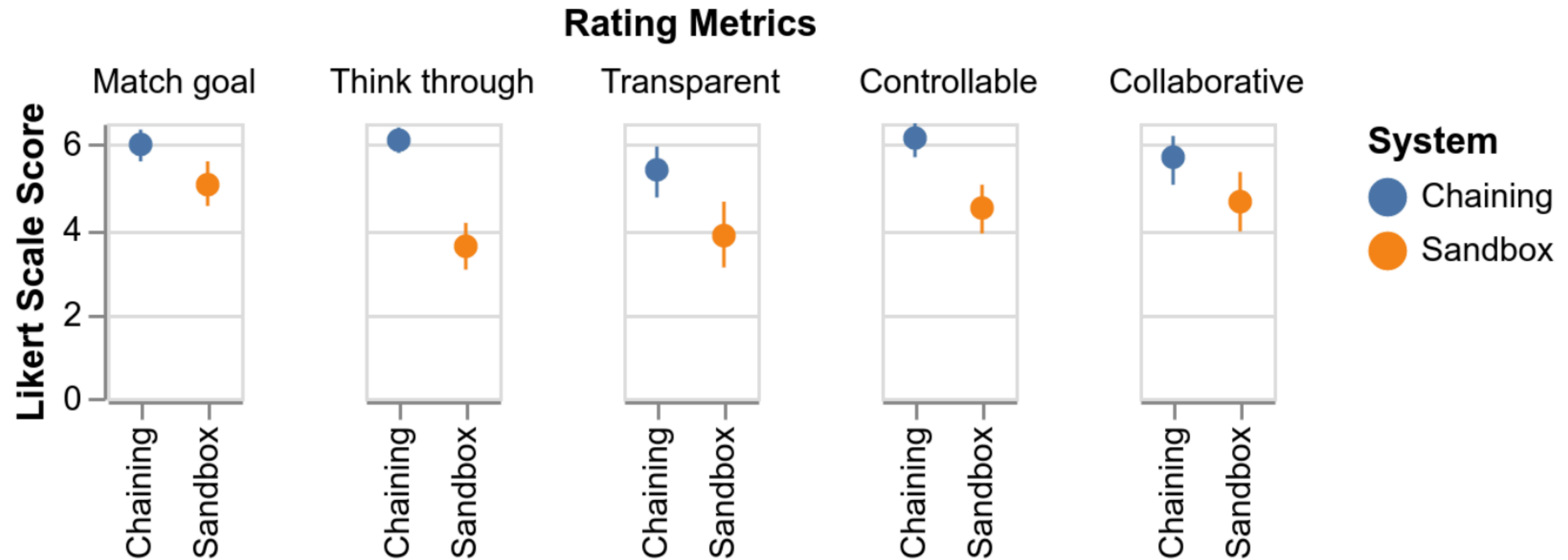


Source: <https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg>

GPT-4

The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi.

“Chaining” of LLMs in Human-AI teaming



Democratic Policy Development using Collective Dialogues and LLMs

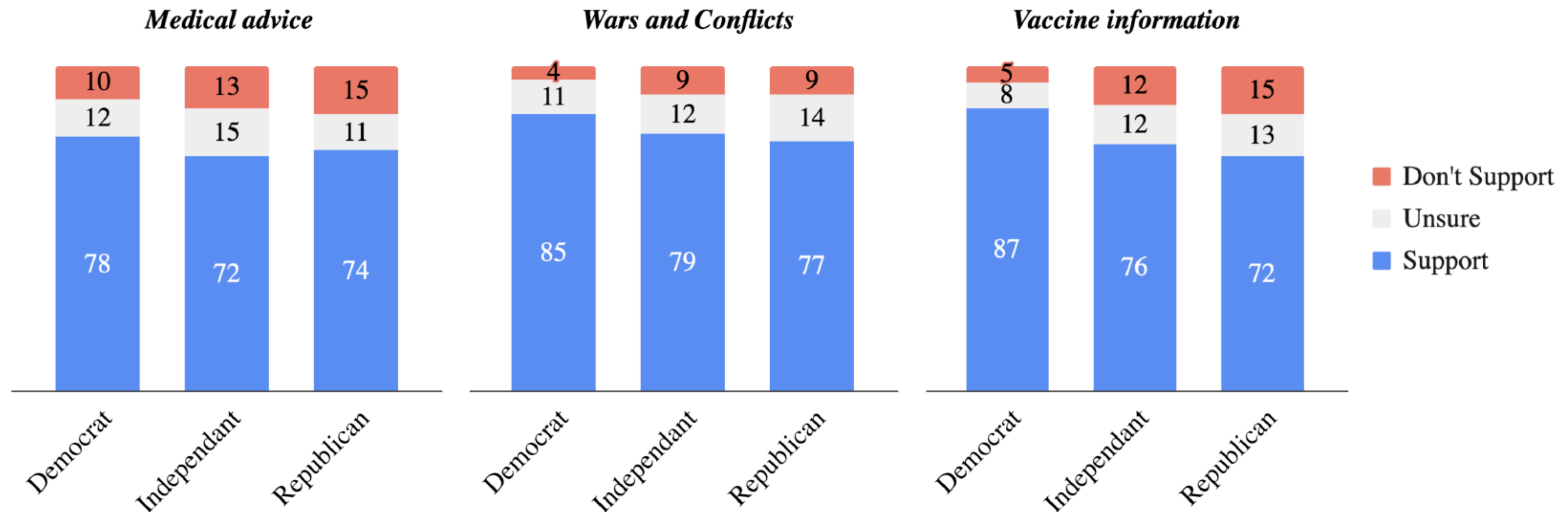
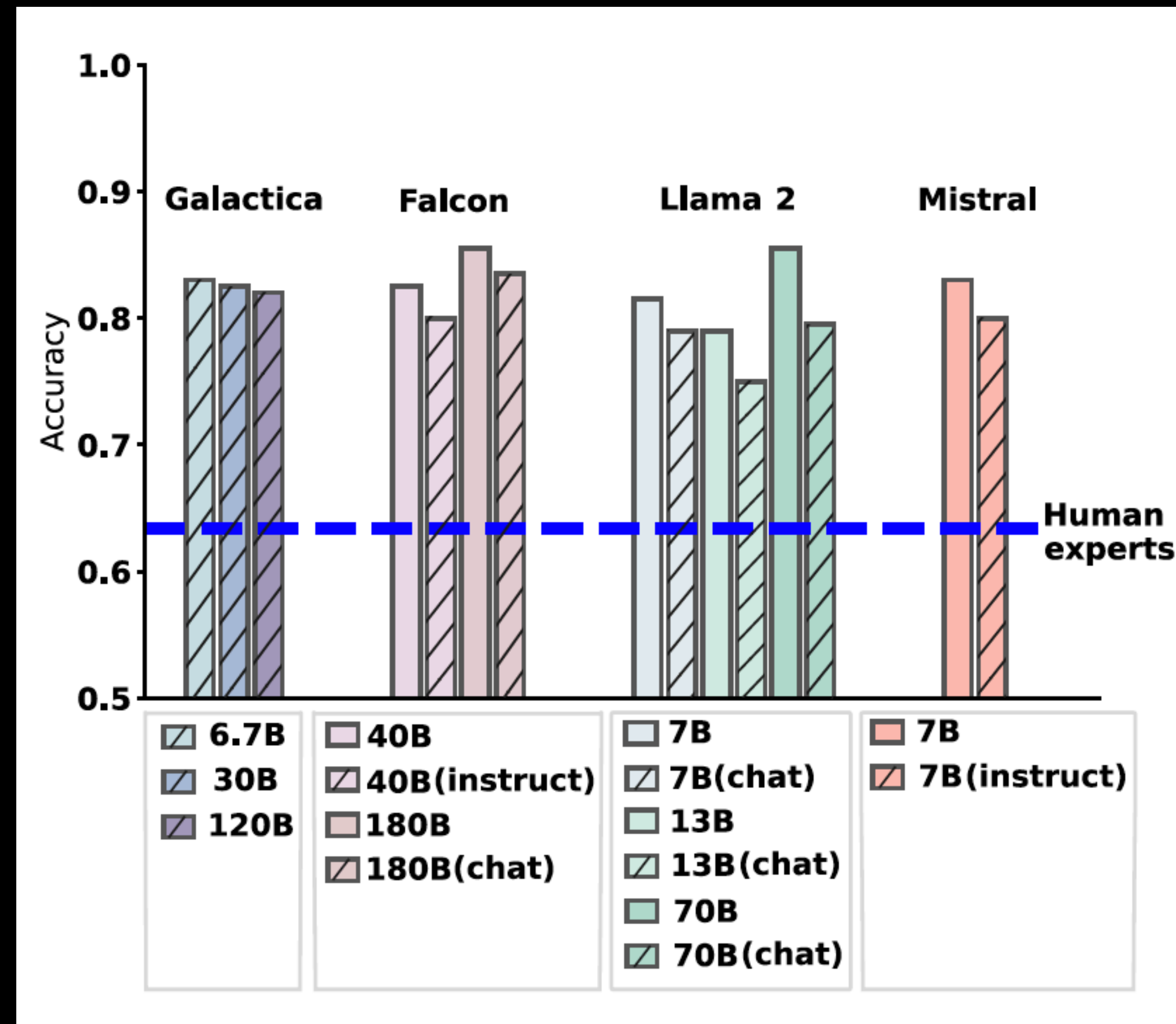


Figure 7: Support across the US political spectrum for the three different policy guidelines developed using the process.

LLMs surpass human experts in predicting neuroscience results

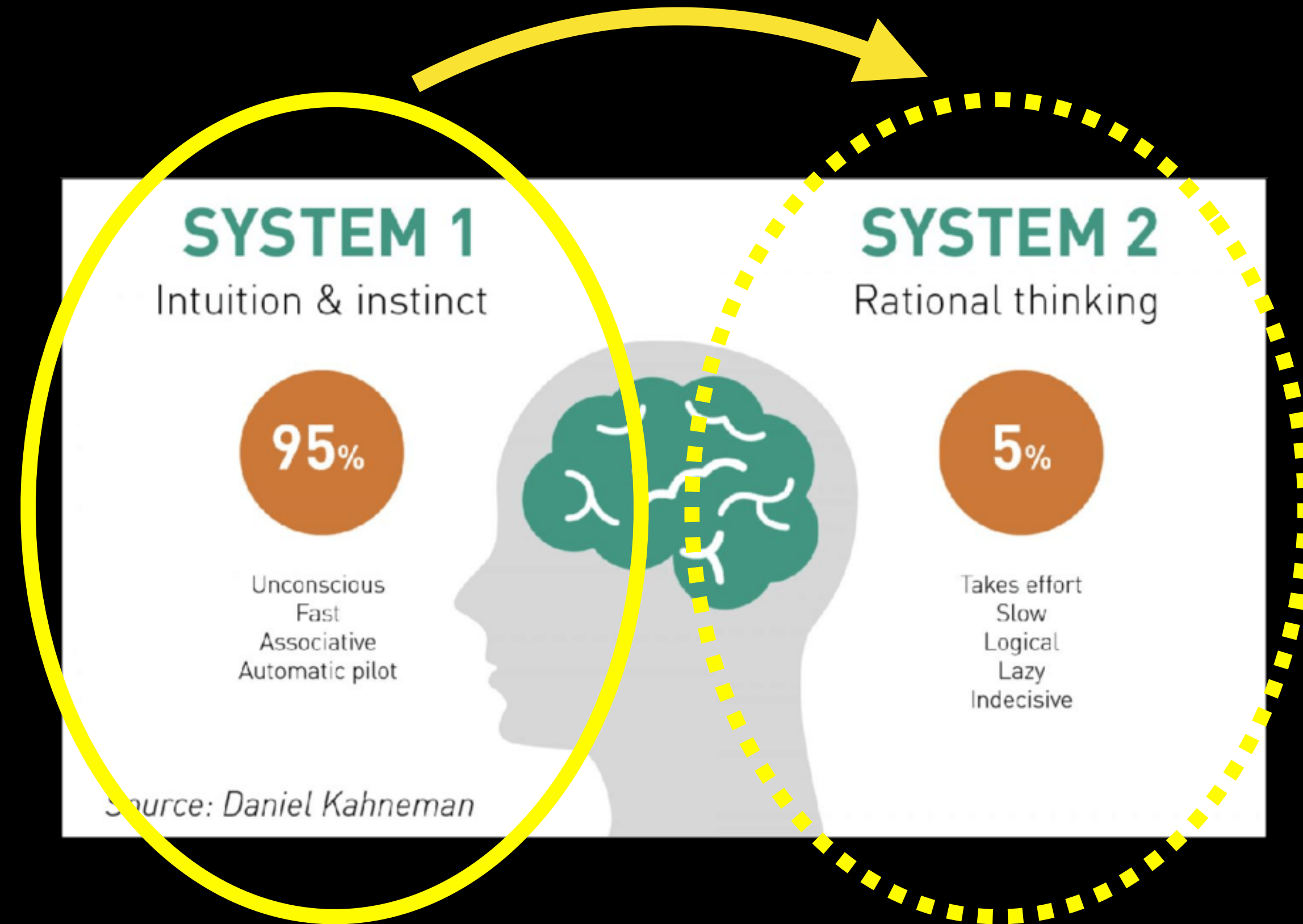


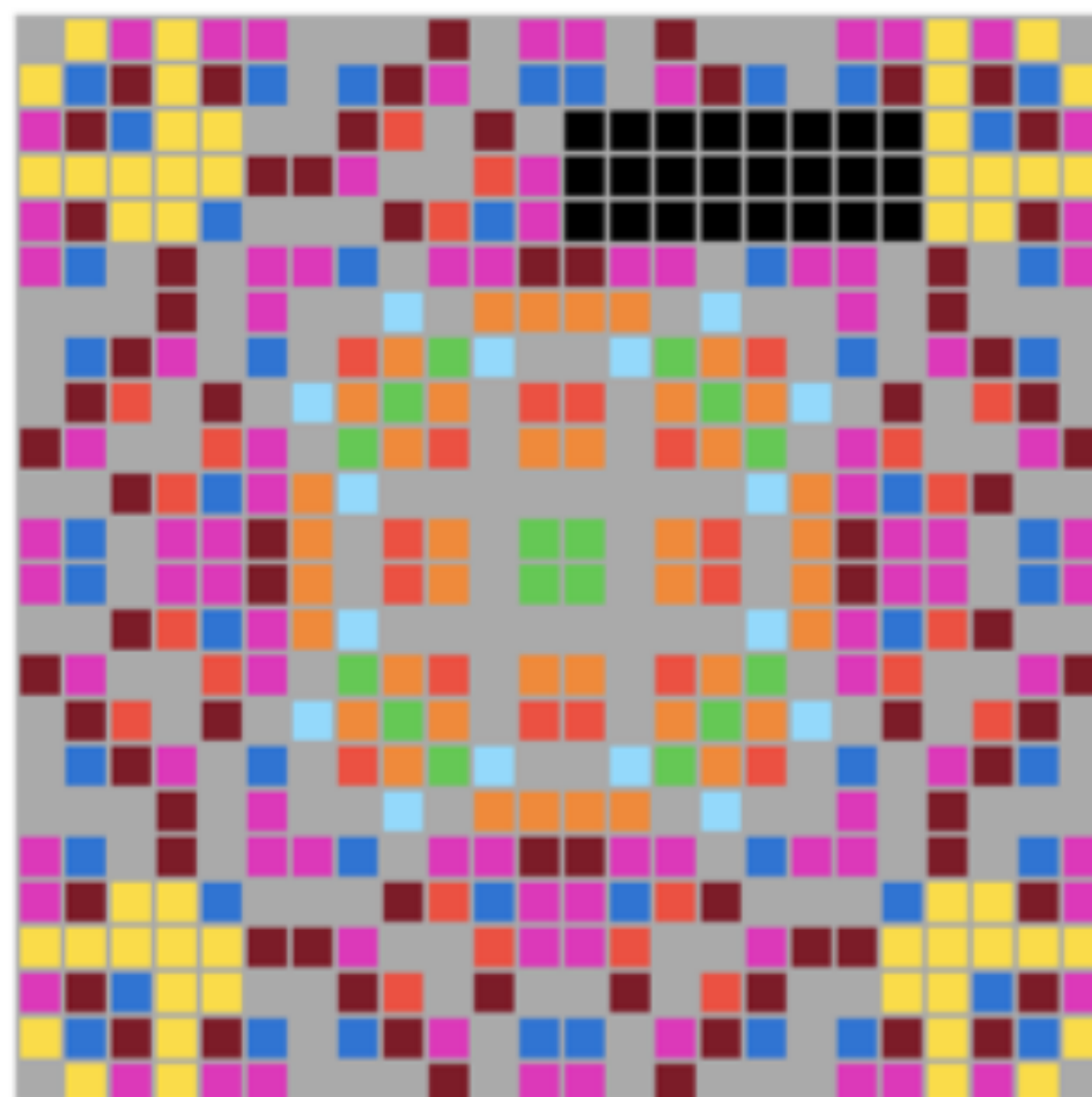
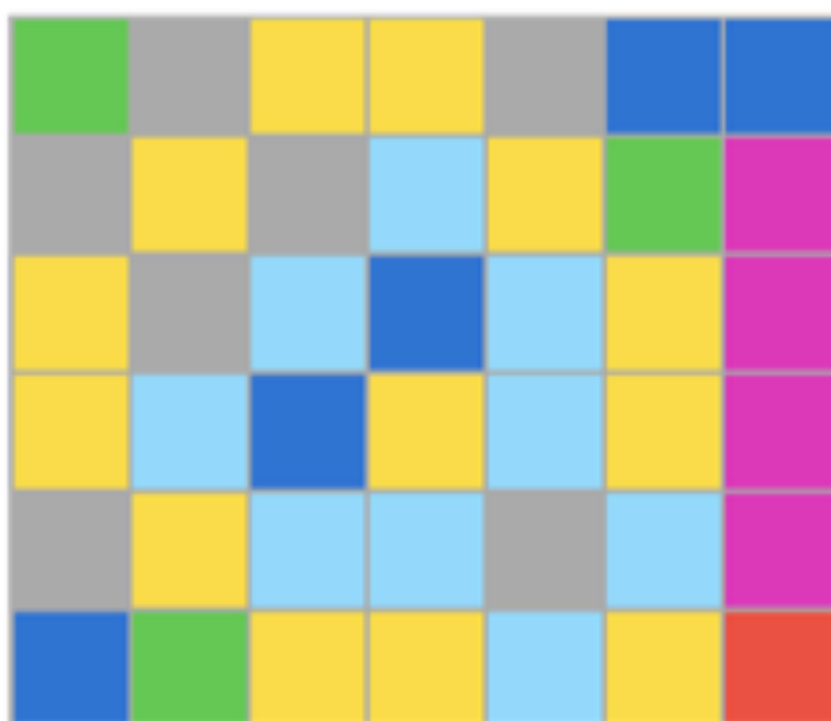
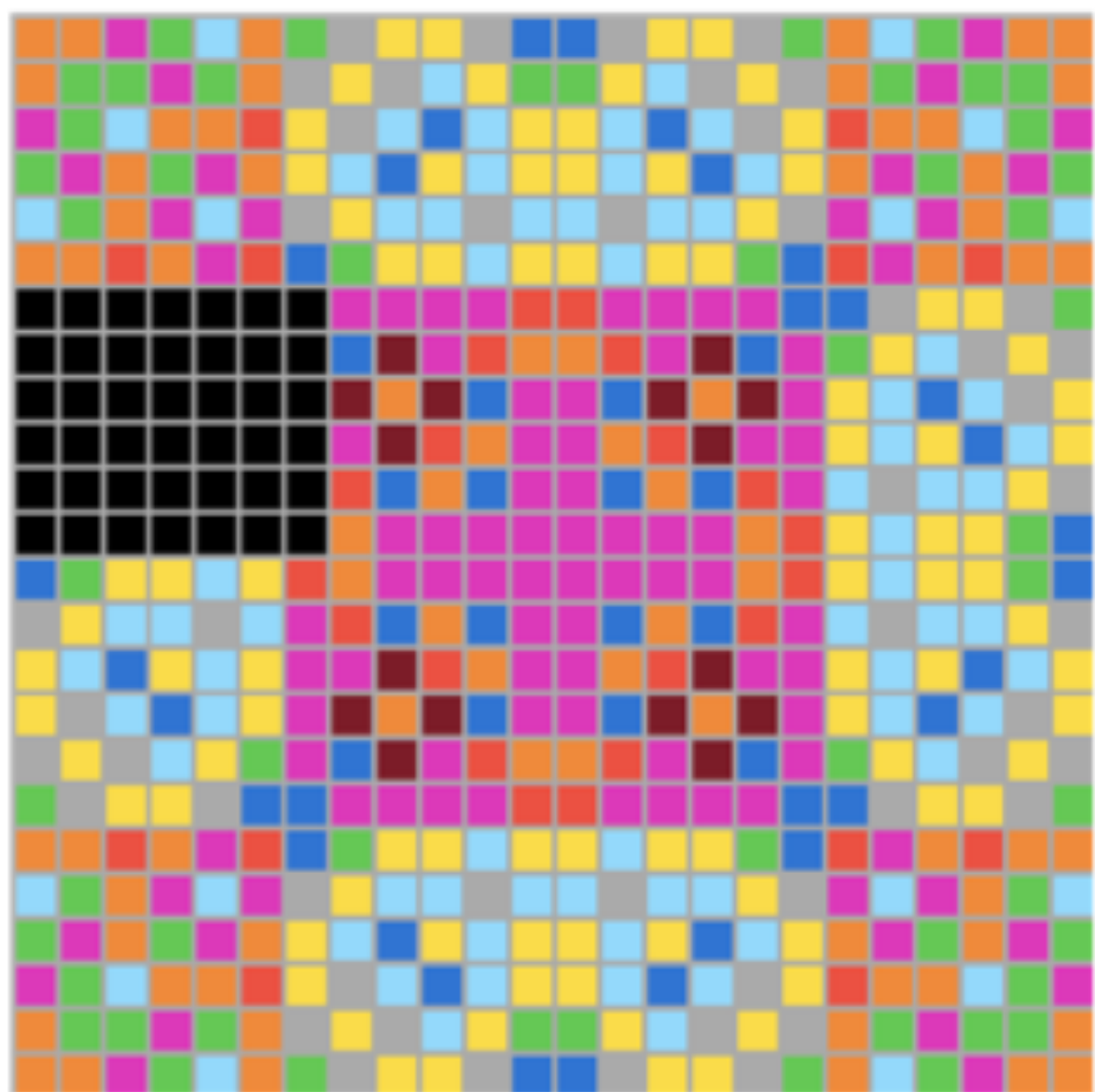
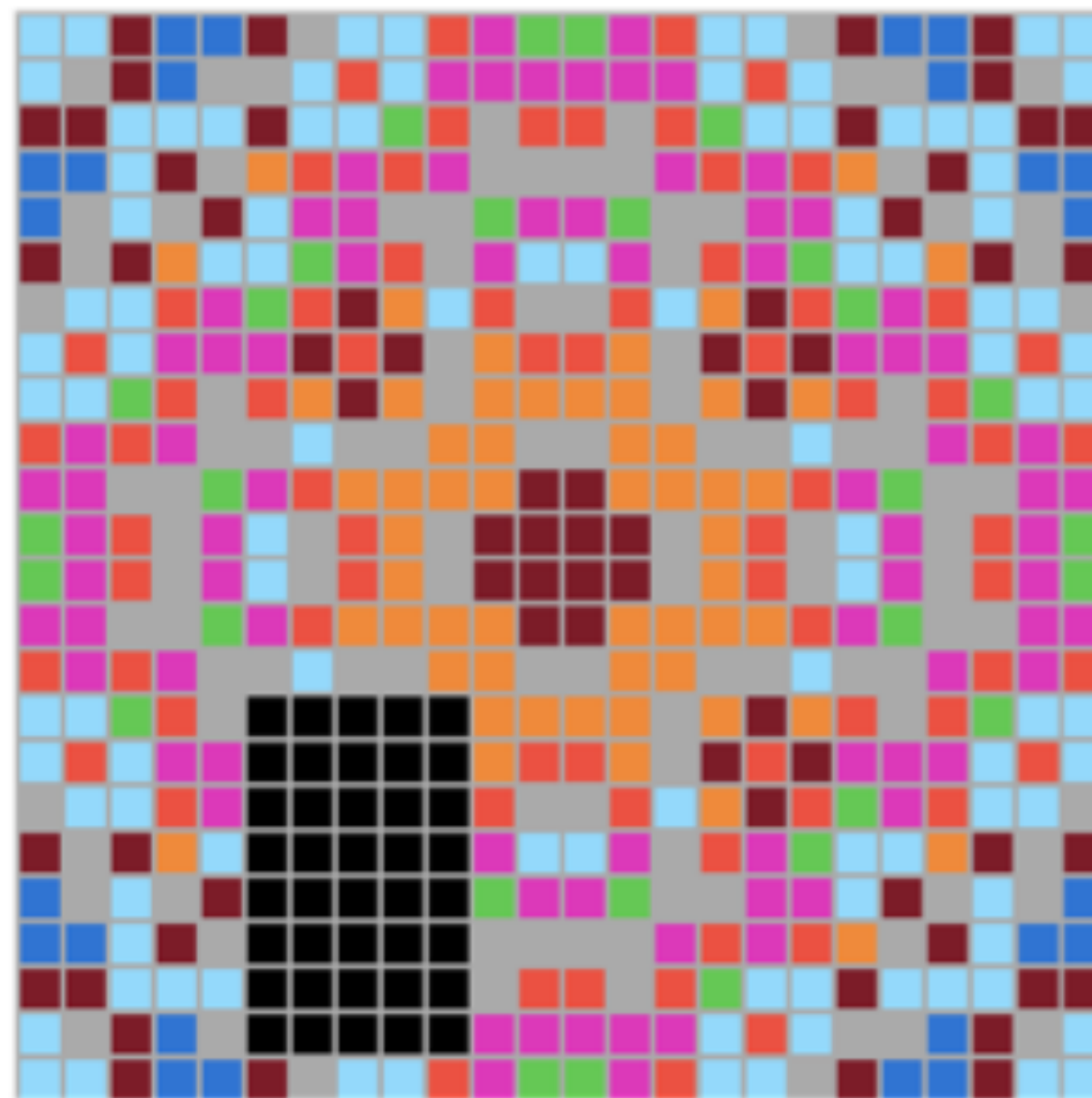
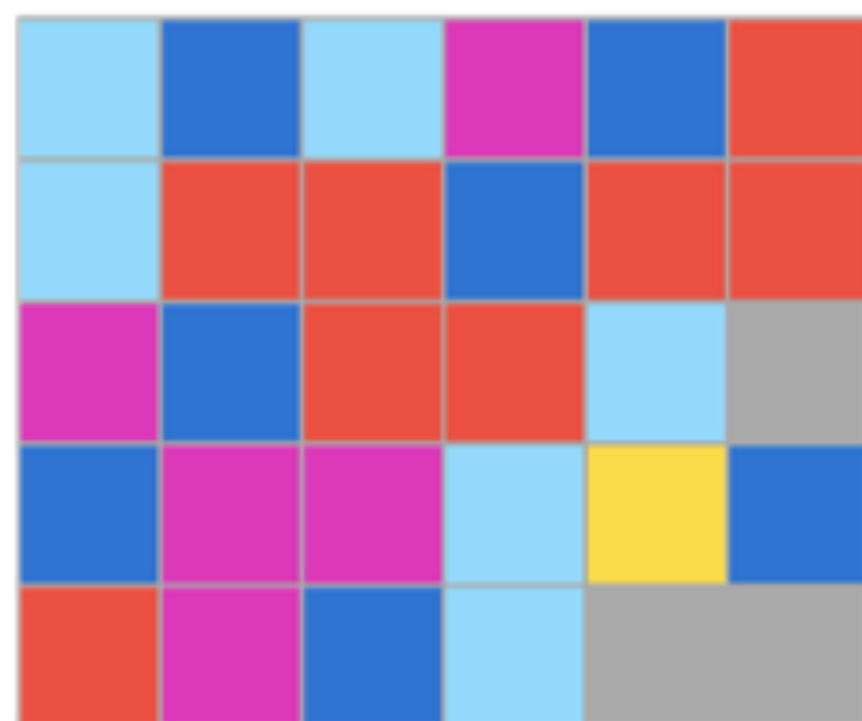
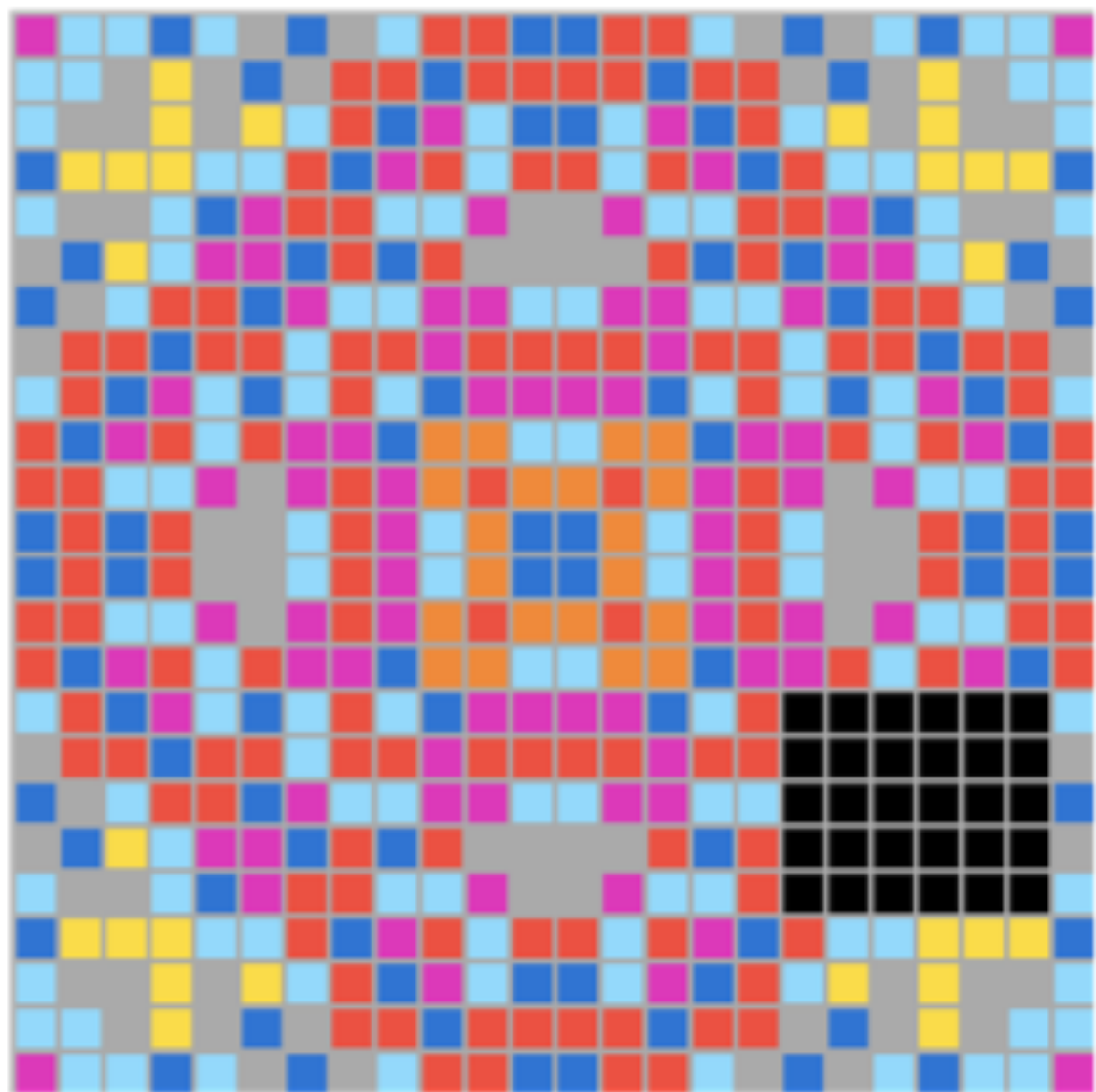
AI in Systems Engineering

- LLMs can contribute to data analysis, hypothesis generation, complex (agent-based) model simulation, prediction
- LLMs suffer from biases (cultural and societal biases, data-source biases, data-collection biases, fine-tuning biases, prompting biases)
- LLMs are bad at reasoning (“system 2”) and can be very convincing if fine-tuned like Chat-GPT.
- LLMs are not yet fully understood (work in progress)

What LLMs are not able to do...

?





?

Narrow but very powerful tool

- AI moved from **explicit** to **implicit** knowledge representations (XAI)
- Emergent properties of overparameterized models
- GPT-like systems still in full development
- Unlike humans, so avoid the use of “suitcase terms” (“understands”, “recognizes”, ...). Do not anthropomorphise
- Although GPT4 is not open software, it is to be expected that alternative open versions will become available through research initiatives
- Transformers in LLMs and many other applications are changing AI and its impact on society

