A Casual Introduction To AI, ML, DL, & LLMs

2024.09.03 Guest Lecture

Ethical Challenges in AI, Ethics Lab @ Georgetown University

Yo Joong "YJ" Choe

Data Science Institute @ University of Chicago

Goals

1. Explain what AI, ML, DL, and LLMs are.

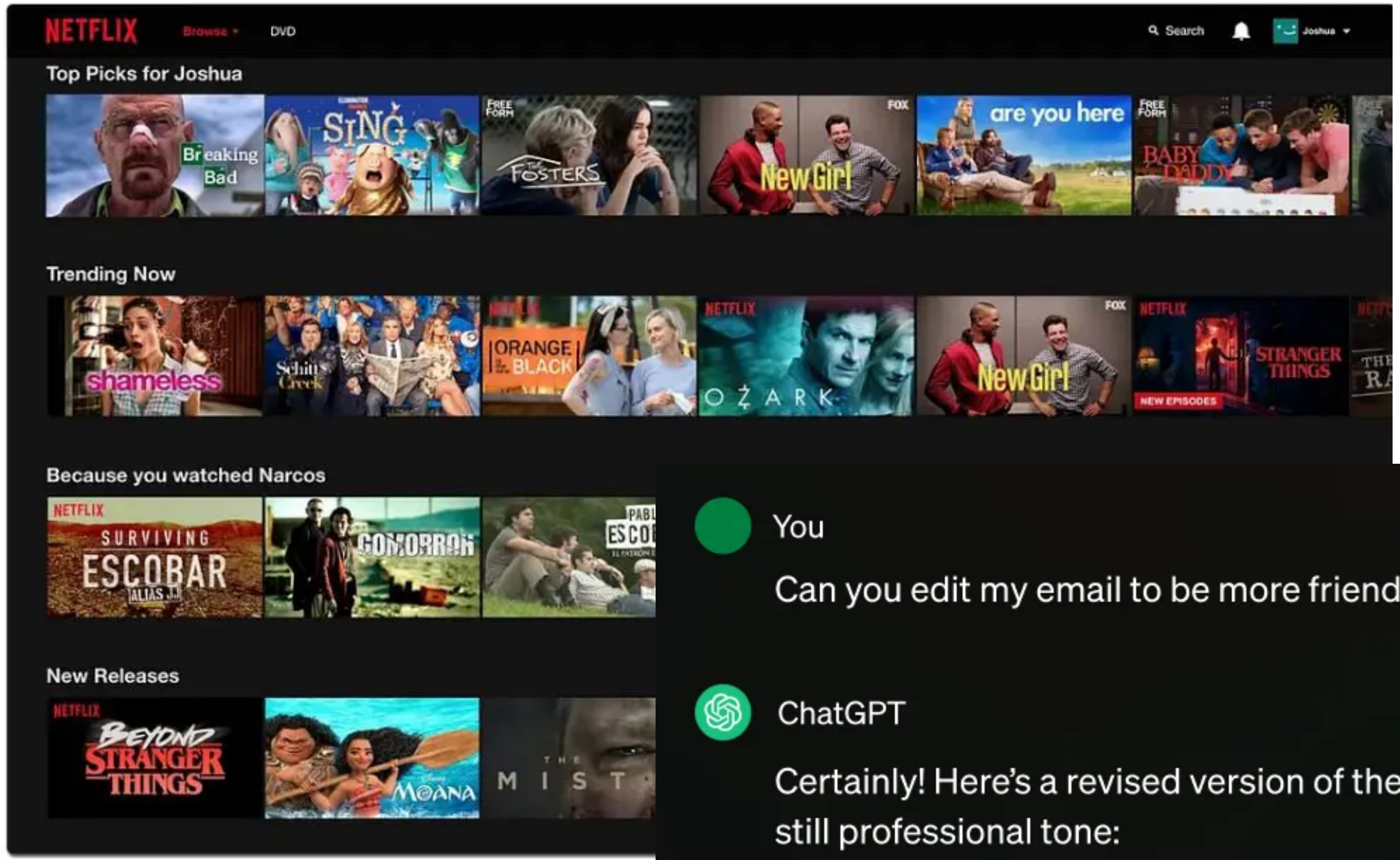
2. Explain how machines actually “learn.”

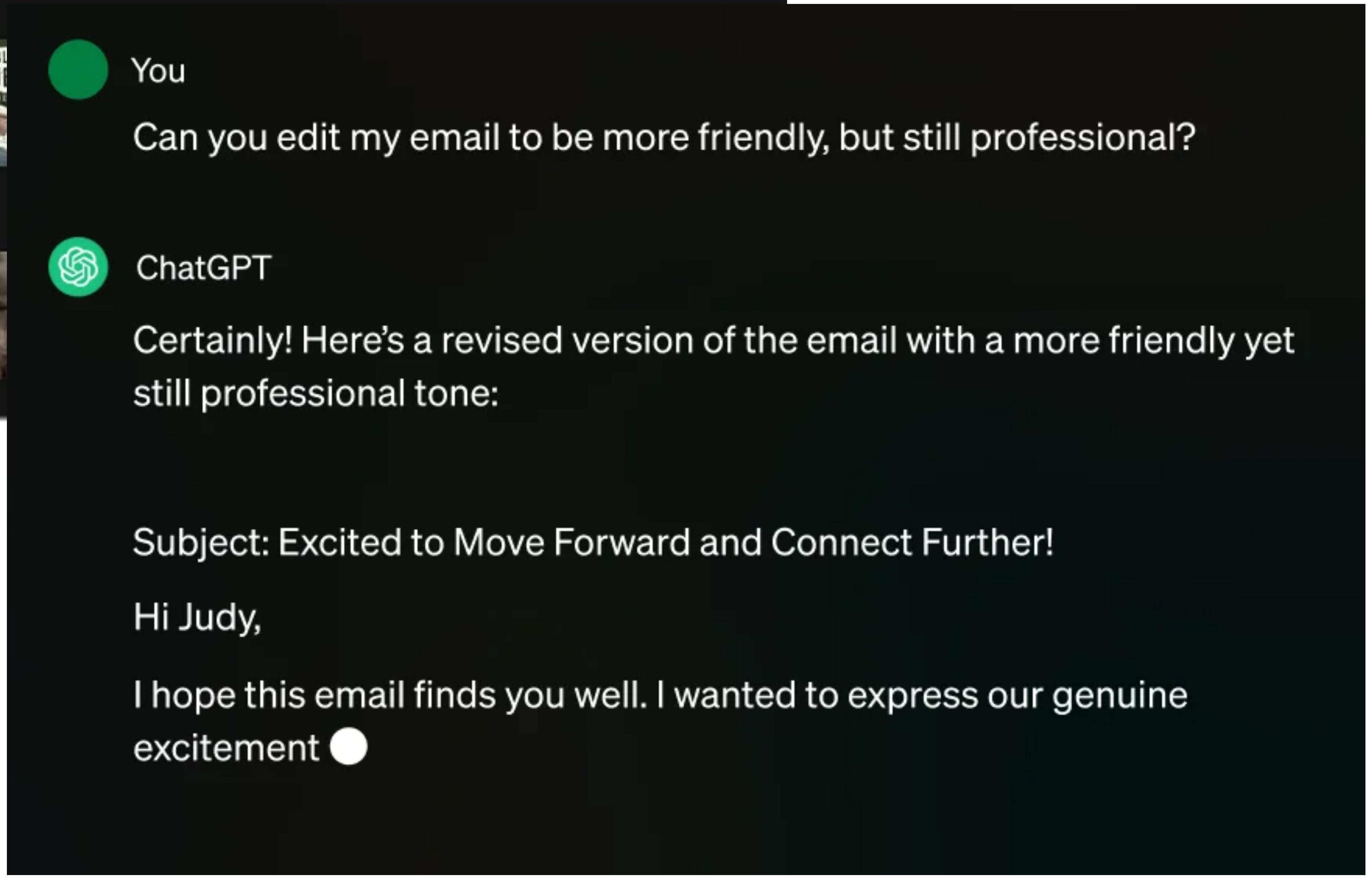
3. Introduce some key applications of AI in the past, present, and future.

   (& Help you think about where the ethical challenges may arise!)

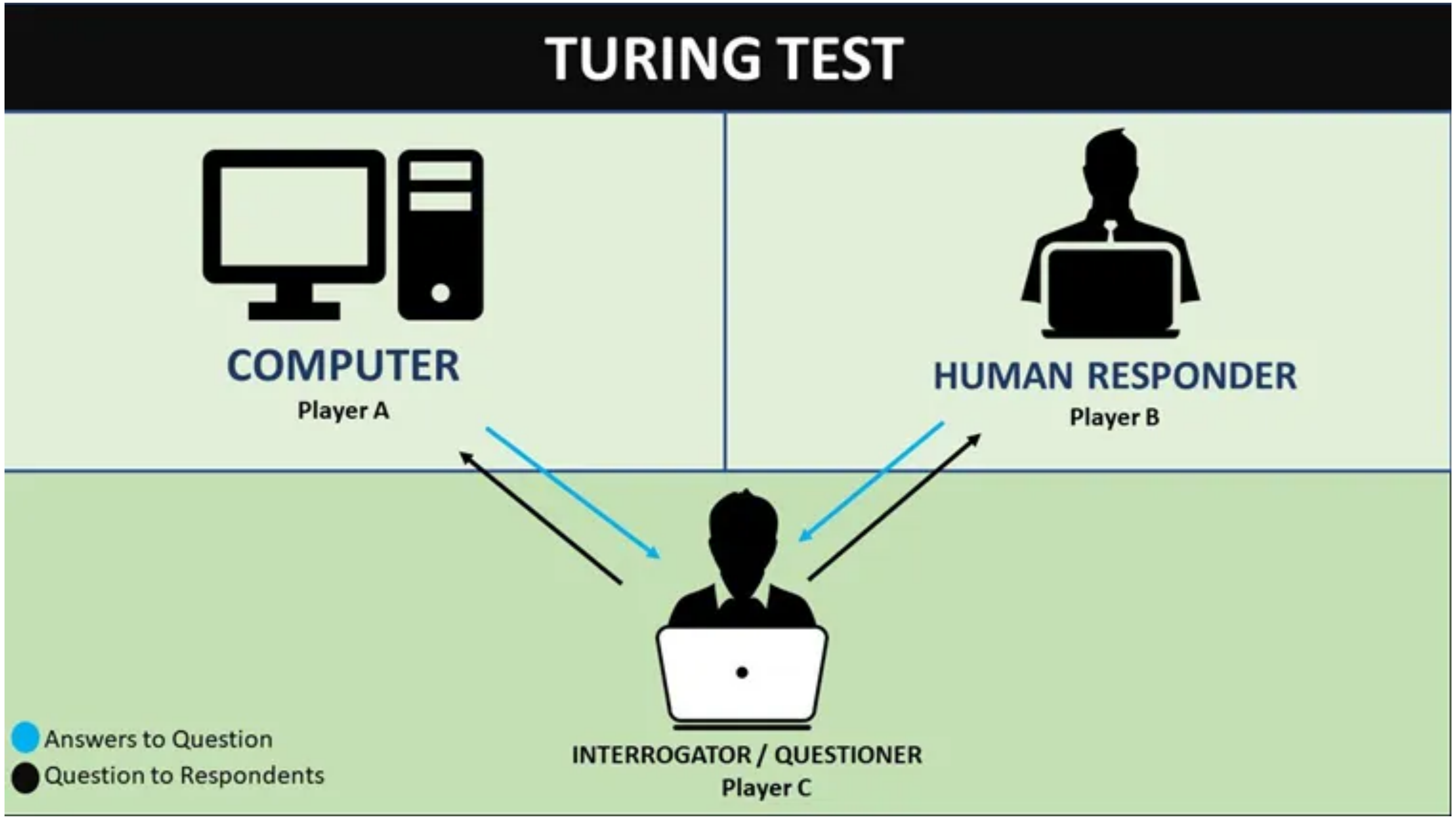
What is AI?

What Is AI?


NETFLIX
Browse
DVD
Search
Top Picks for Joshua
Trending Now
Because you watched Narcos
New Releases


You
Can you edit my email to be more friendly, but still professional?
ChatGPT
Certainly! Here's a revised version of the email with a more friendly yet still professional tone:
Subject: Excited to Move Forward and Connect Further!
Hi Judy,
I hope this email finds you well. I wanted to express our genuine excitement

What Is AI?



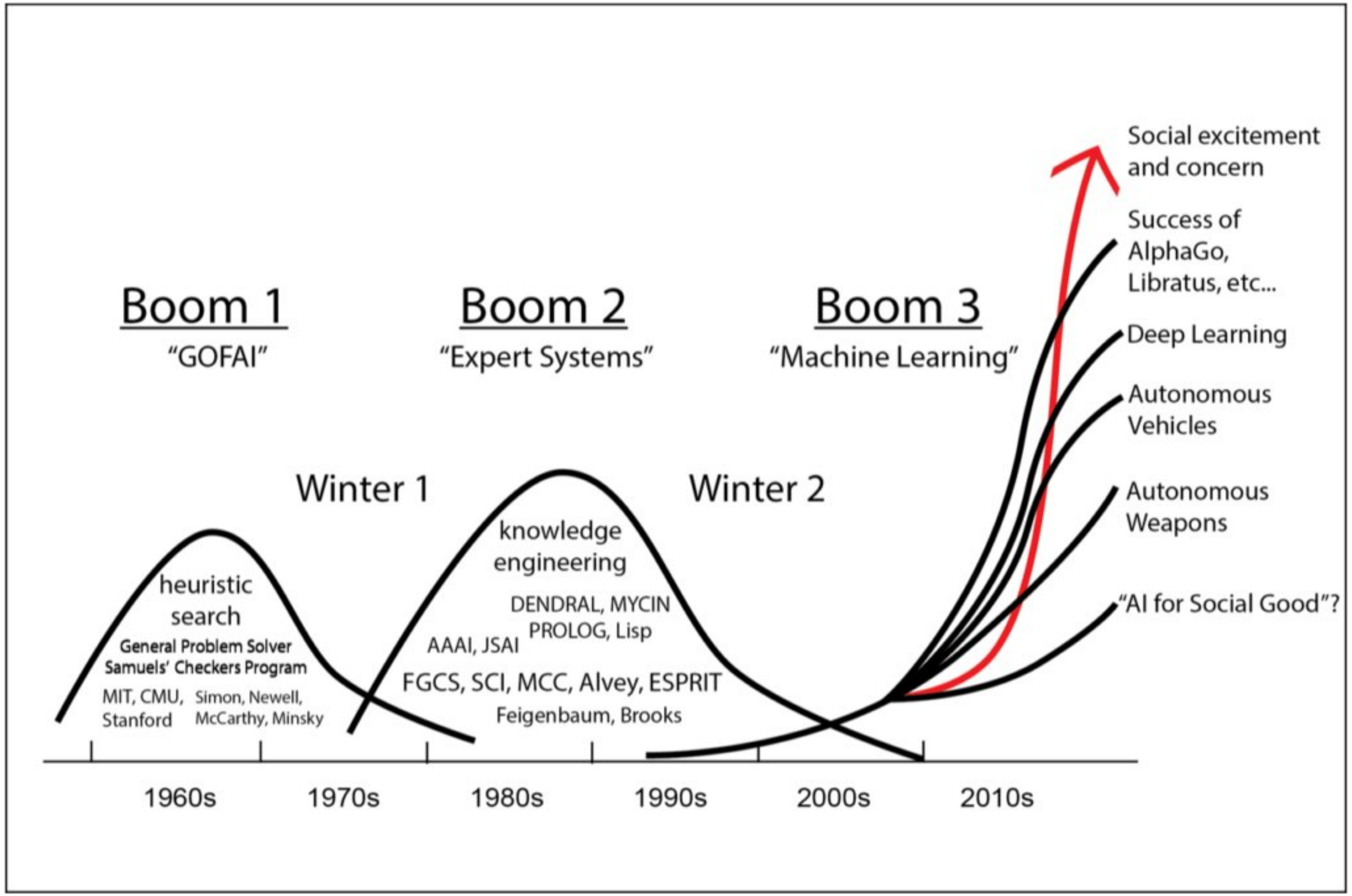
A. Turing (1950)

What Is AI?

“We call programs intelligent if they exhibit **behaviors that would be regarded intelligent if they were exhibited by human beings**.”

- Simon (see also: Minsky, 1968)

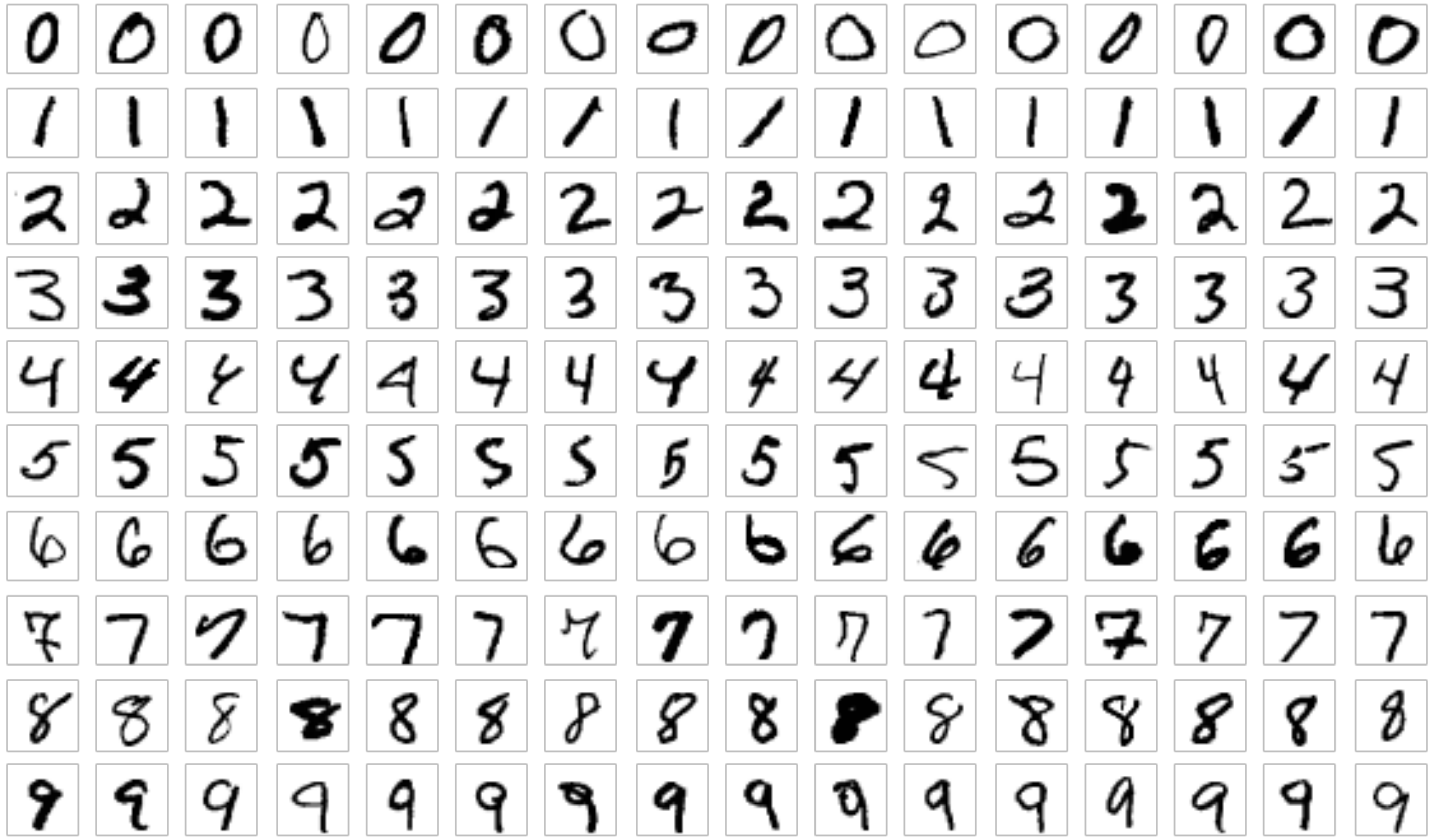
AI: A History of Booms & Busts



https://jaylatta.net/history-of-ai-from-winter-to-winter/

Success Stories of AI

HANDWRITING RECOGNITION



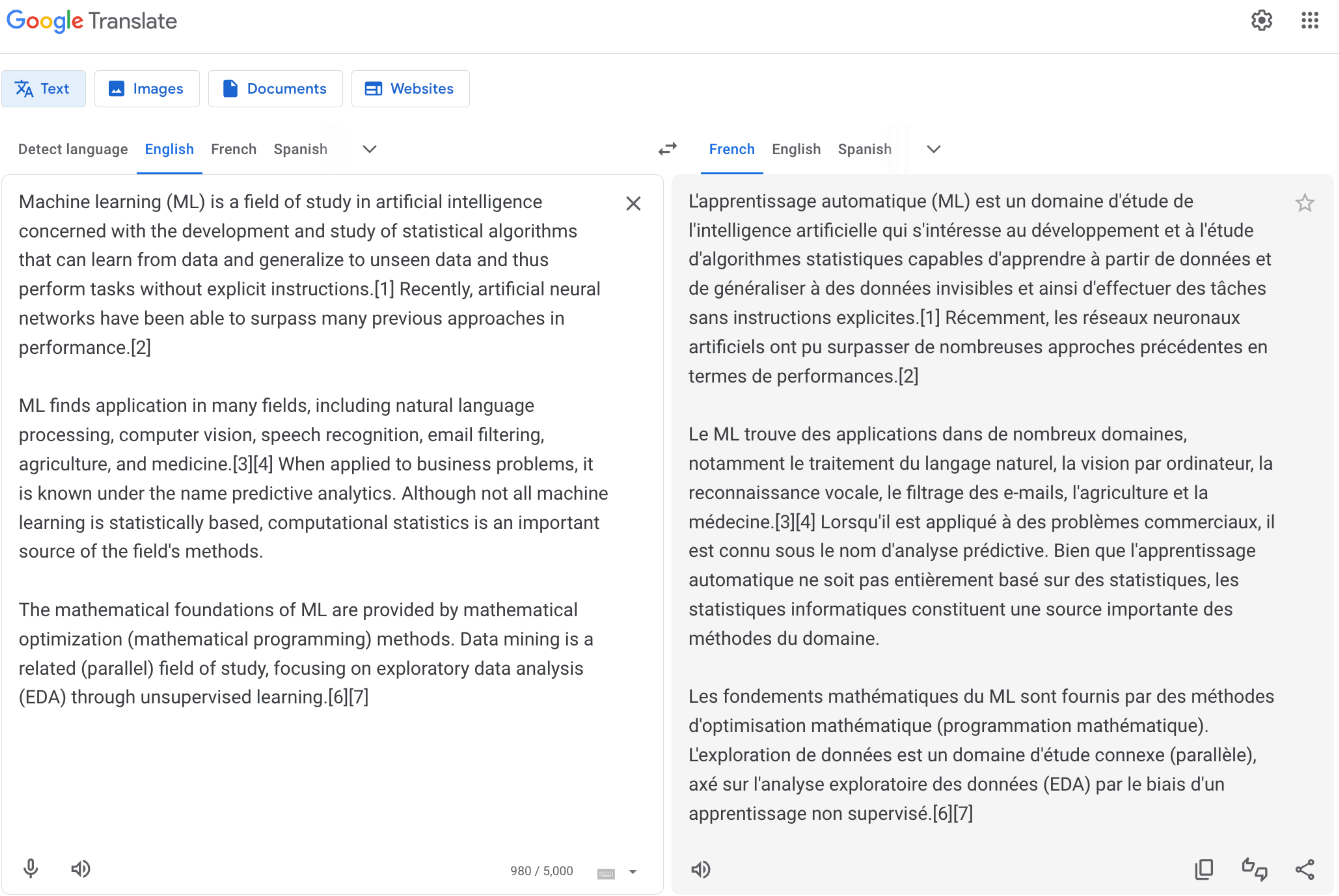
Object Recognition & Detection

Krizhevsky et al. (2012)

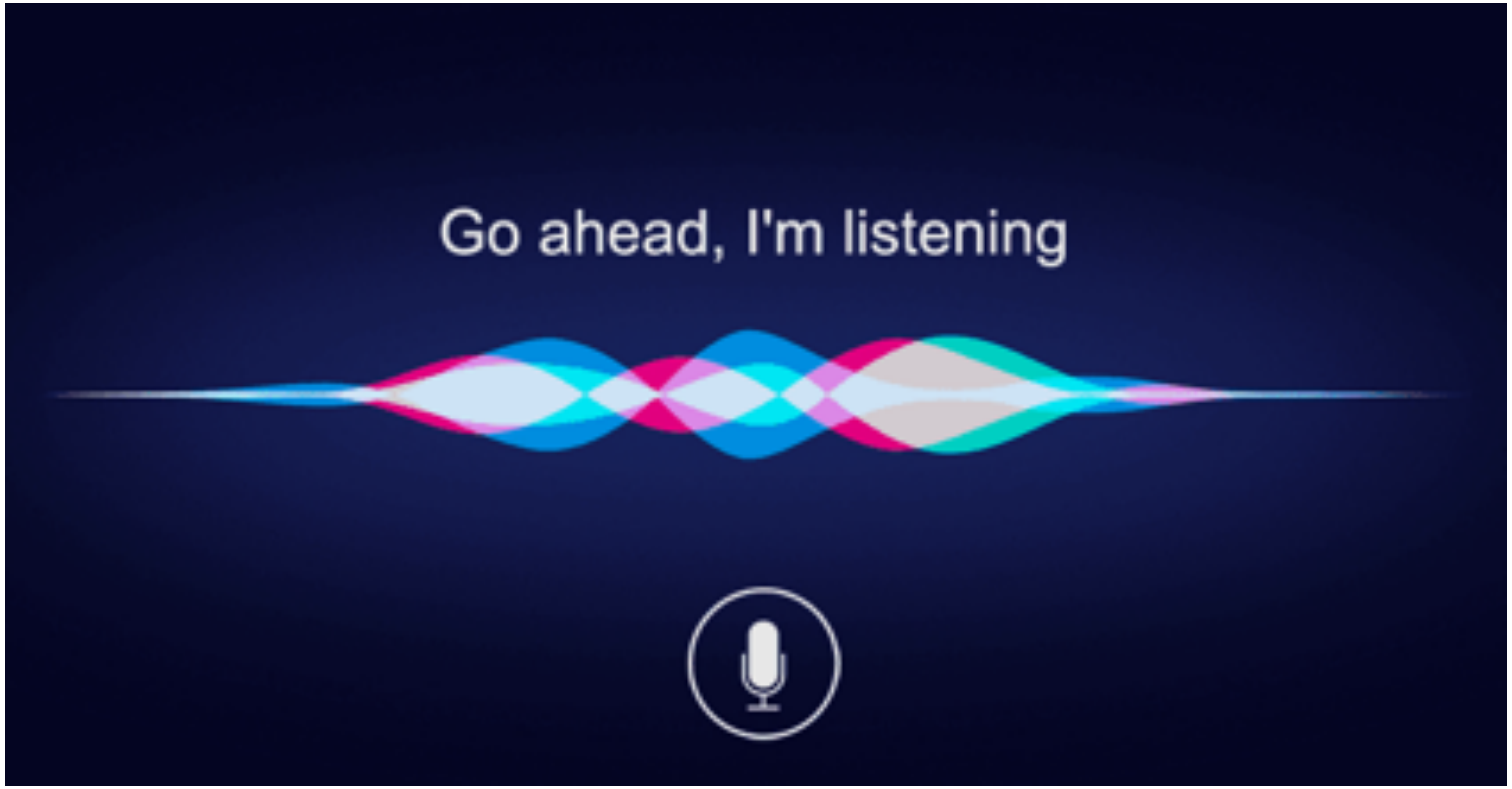
He et al. (2017)

Machine Translation

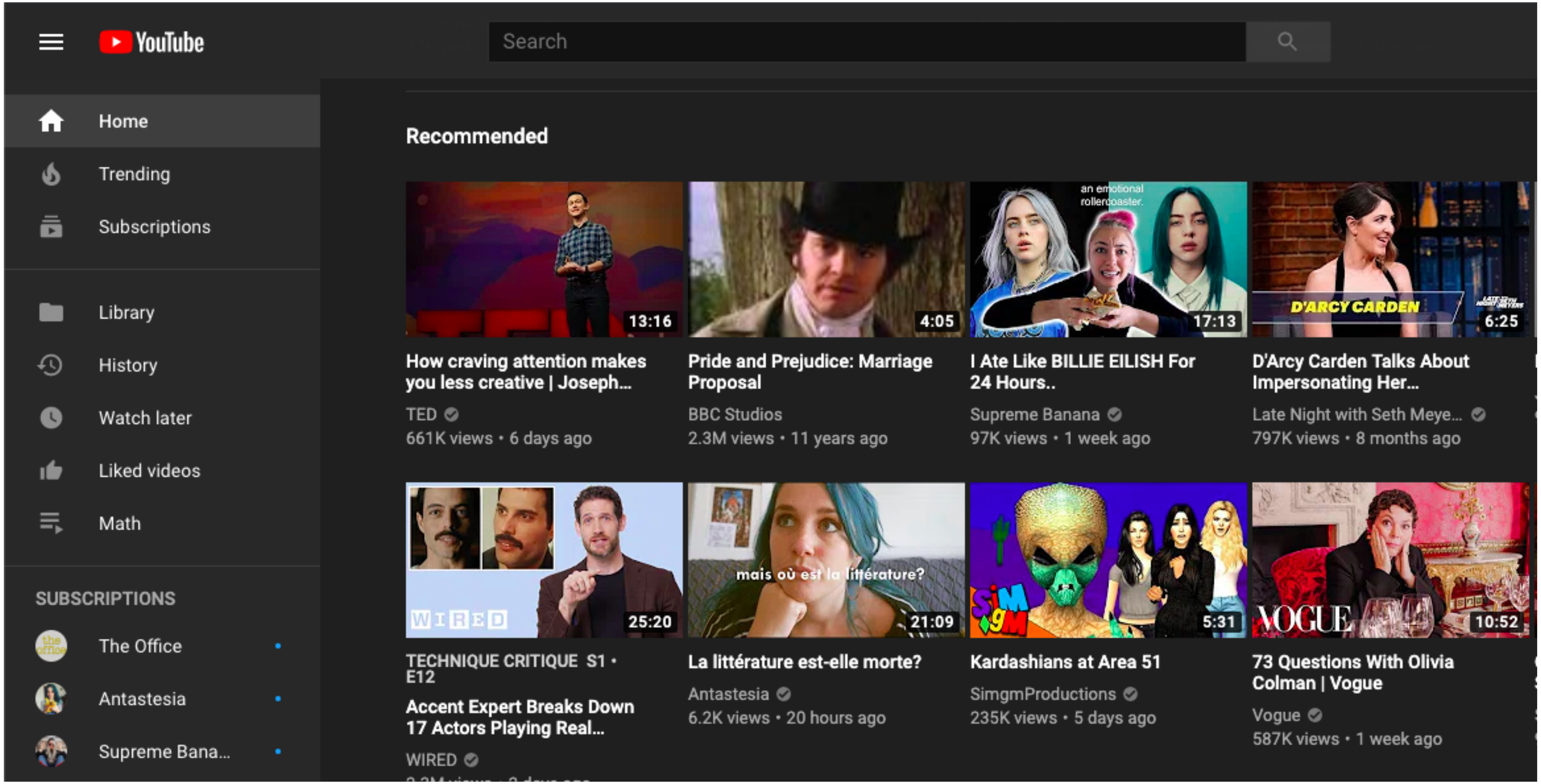
(1954)



Speech Recognition & Synthesis



Recommender Systems



IBM Deep Blue (Chess)

The 1997 "Rematch" between Garry Kasparov & IBM Deep Blue.

AlphaGo & AlphaStar

Go (2016)

Starcraft II (2019)

Libratus (Poker)

Libratus, the poker AI (2017)

Name	Rank	Results (in chips)
Dong Kim	1	-$85,649
Daniel MacAulay	2	-$277,657
Jimmy Chou	3	-$522,857
Jason Les	4	-$880,087
Total:		-$1,766,250

Results of 4 human players against Libratus after 20 days of competition

The Explosion of Generative AI: ChatGPT, Midjourney, Copilot, Sora, …

What is Machine Learning?
(How Machines Actually "Learn")

WHAT IS MACHINE LEARNING?

A. L. Samuel

Some Studies in Machine Learning Using the Game of Checkers

Abstract: Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote
more, it can learn to do this in a remarkably short period of time (8 or 10 hours of
when given only the rules of the game, a sense of direction, and a redundant
parameters which are thought to have something to do with the game, but whose co
weights are unknown and unspecified. The principles of machine learning verified
are, of course, applicable to many other situations.

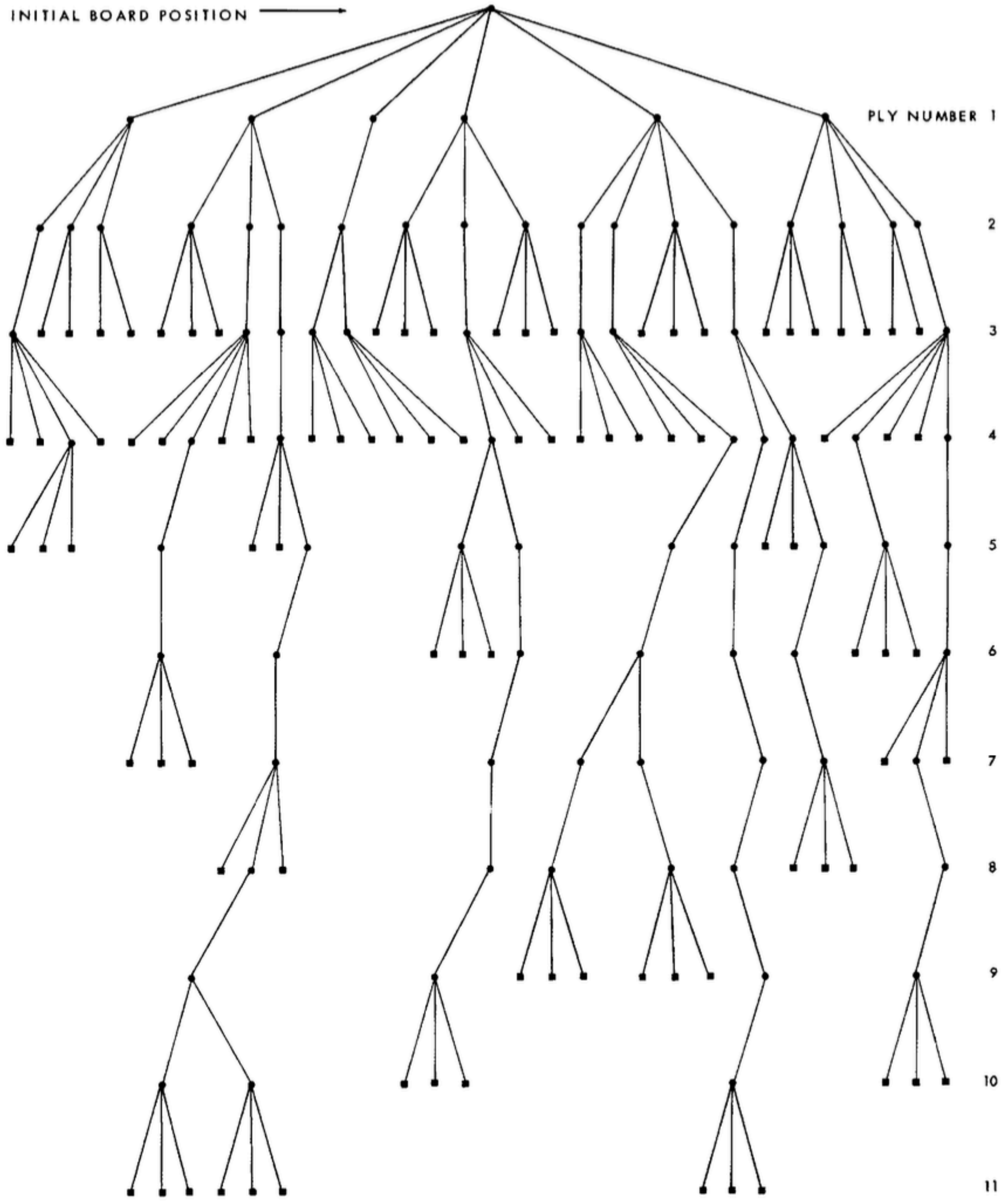


Figure 1 A "tree" of moves which might be investigated during the look-ahead procedure. The actual branchings are much more numerous than those shown, and the "tree" is apt to extend to as many as 20 levels.

What Is Machine Learning?

"Programming computers to learn from experience should eventually **eliminate the need for much of this detailed programming effort.**"

- A. Samuel (1959)

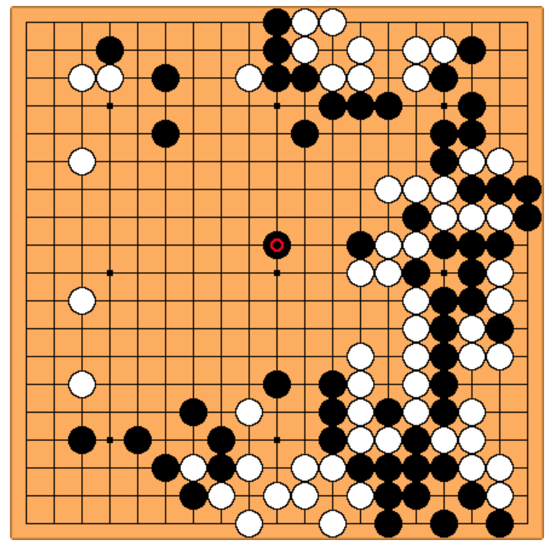
Weekday,
Afternoon,
Sunny, 85F,
10mph Wind,
Nearby Georgetown,
During the Semester,
...

→ **ML** →

Bikeshare
Demand:
80

"Thank you for helping me maintain my weight."

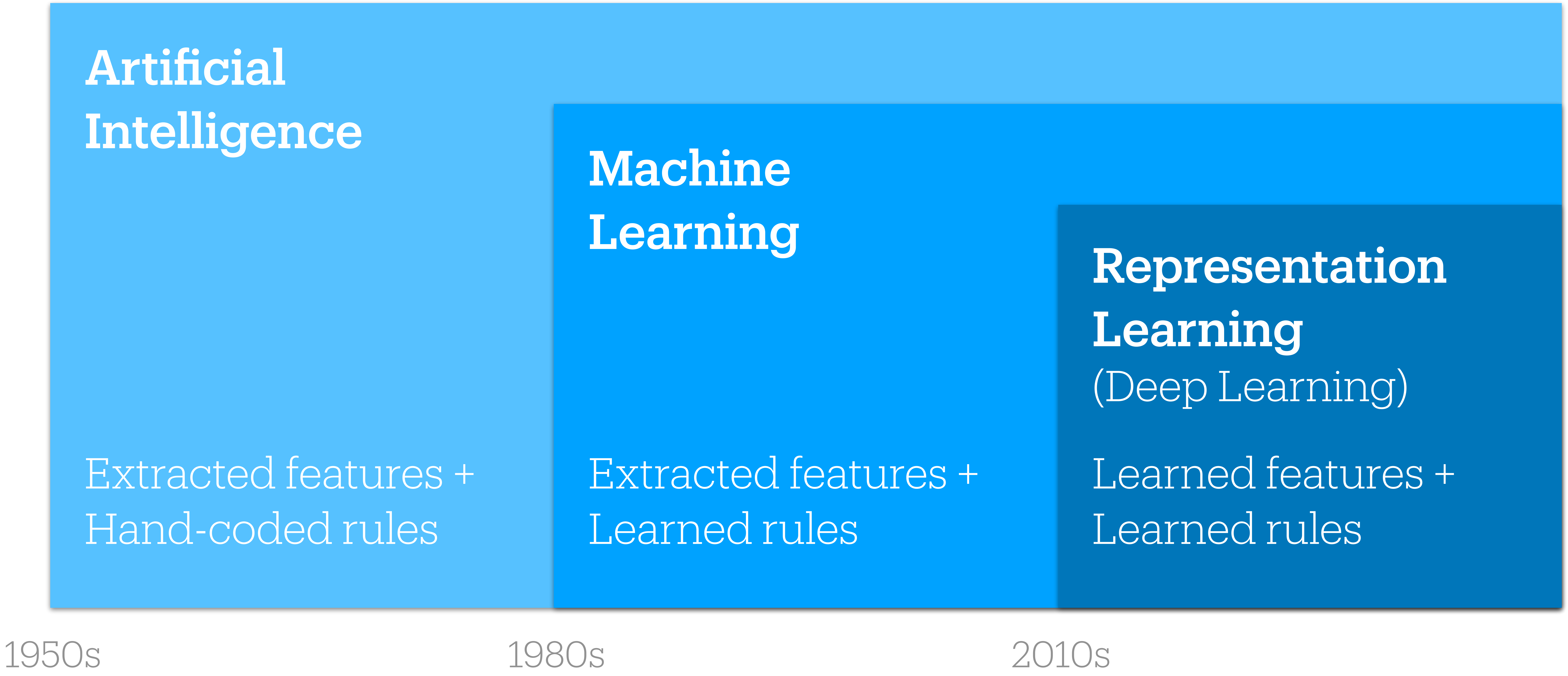
→ **ML** →



→ **ML** →

C16

History of AI = Automation



A (slightly more) Modern Definition

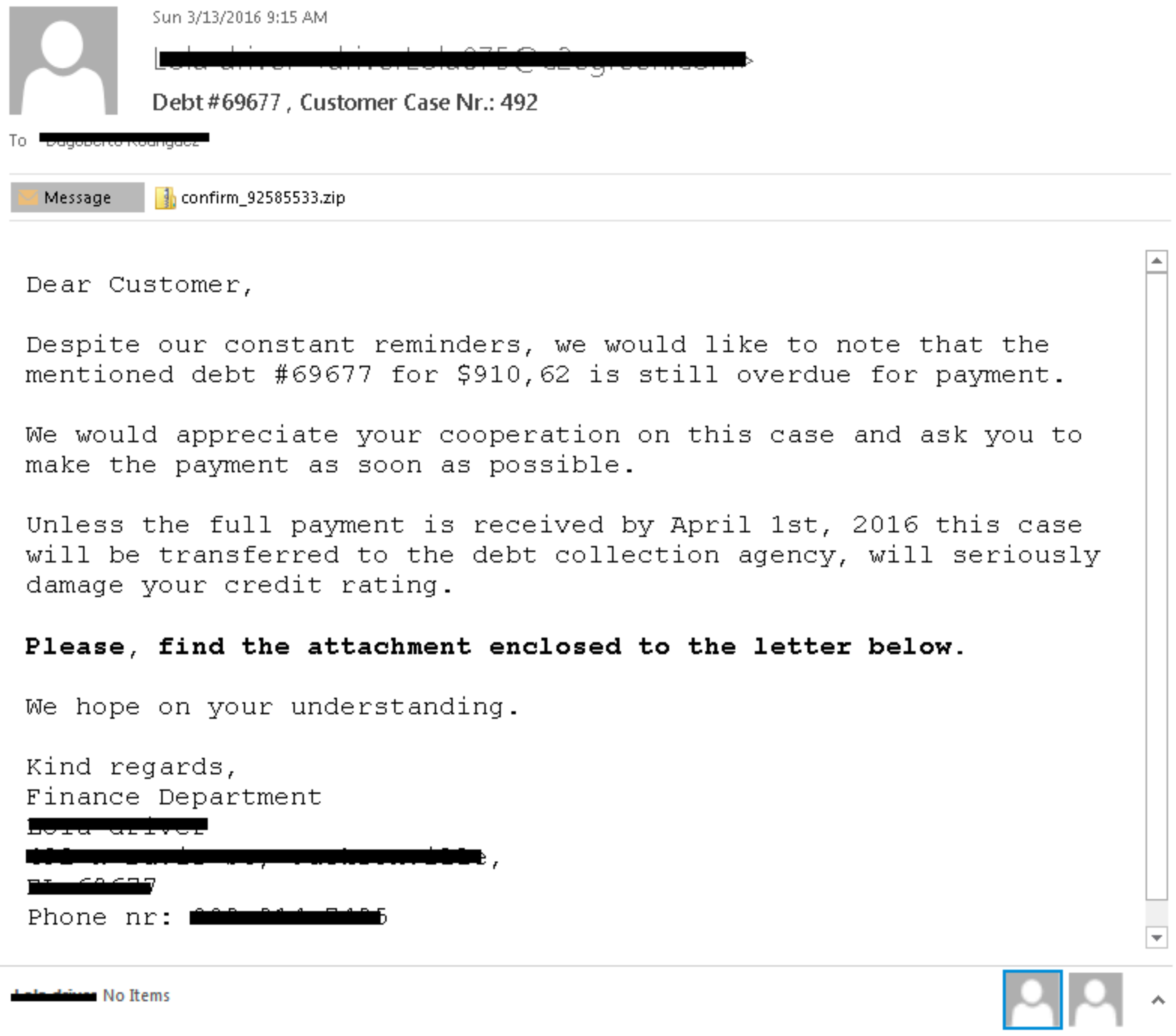
"A computer program is said to learn [...] if
its performance at [given] **tasks**,
as measured by some **performance measure**,
improves with **experience**."

- T. Mitchell (1997)

Key Objectives in ML

- **Generalization**
  - How well can the algorithm process previously unseen examples?
- **Representation**
  - How well can the algorithm capture relevant features of (raw) data?

Example: Spam Classification



Sun 3/13/2016 9:15 AM

[redacted]

Debt #69677 , Customer Case Nr.: 492

To [redacted]

Message | confirm_92585533.zip

Dear Customer,

Despite our constant reminders, we would like to note that the mentioned debt #69677 for $910,62 is still overdue for payment.

We would appreciate your cooperation on this case and ask you to make the payment as soon as possible.

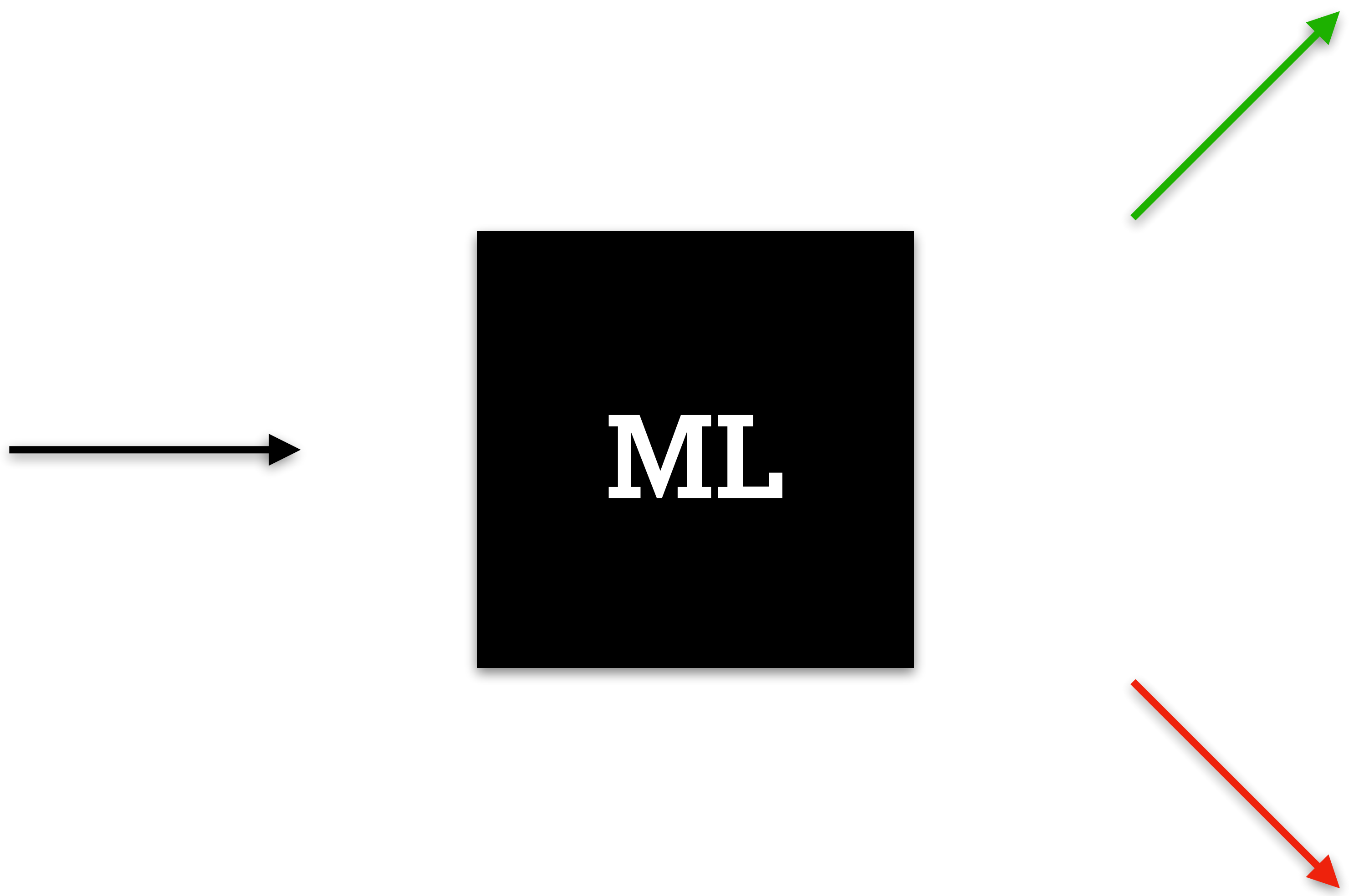
Unless the full payment is received by April 1st, 2016 this case will be transferred to the debt collection agency, will seriously damage your credit rating.

Please, find the attachment enclosed to the letter below.

We hope on your understanding.

Kind regards,
Finance Department
[redacted]
Phone nr: [redacted]

[redacted] No Items



Representing Emails

- **Examples of features in an email:**
  - Number of words
  - Term frequencies
  - The domain of sender's email address
  - Whether the text contains keywords:
    "debt", price ("$3.99"), special characters, all caps, ...
  - **Automatically learned** features for the task

Learning from Experience

ID	# Words	Domain	$	"debt"	"FREE"	Label
0	120	gmail.com	Yes	Yes	Yes	Spam
1	500	gmail.com	No	No	No	Ham
2	400	cmu.edu	Yes	Yes	No	Ham
3	1200	suspicio.us	Yes	No	Yes	Spam

Learning from Experience

1. Define a **loss function** (the "performance measure").
   - 0-1 Loss: 1 if incorrect, 0 otherwise.
2. Define a **model** between features ("input") and the label ("output").
   - *The model often has parameters that can be learned, e.g.:*
   - $\text{Odds}(\text{Spam}/\text{Ham}) = \exp[w_1 \cdot (\#\text{Words}) + w_2 \cdot (\text{KnownDomain}) + w_3 \cdot ('\text{FREE}')]$
3. Make a prediction, and **update** your model according to the loss.

Learning from Experience

- Initial Guess:

$$\text{Odds(Spam/Ham)} = \exp[(0.0)\cdot(\#\text{Words}) + (-1.0)\cdot(\text{KnownDomain}) + (0.0)\cdot(\text{'FREE'})] \approx 0.36$$

- Loss: 1 (incorrect)

- ***Feedback ("gradient"): Add 0.001 to w1, 0.5 to w2, 0.5 to w3***

ID	# Words	Domain	"FREE"	Label
0	120	gmail.com	Yes	Spam

Learning from Experience

- Initial Guess:

$$\text{Odds(Spam/Ham)} = \exp[(0.001)\cdot(\#\text{Words}) + (-0.5)\cdot(\text{KnownDomain}) + (0.5)\cdot(\text{'FREE'})] \approx 1.65$$

- Loss: 1 (incorrect)

- ***Feedback ("gradient"): Add -0.002 to w1, 0.25 to w2, 0.25 to w3***

ID	# Words	Domain	"FREE"	Label
0	120	gmail.com	Yes	Spam
1	500	gmail.com	No	Ham

Learning from Experience

- Initial Guess:

$$\text{Odds(Spam/Ham)} = \exp[(-0.001) \cdot (\#\text{Words}) + (-0.25) \cdot (\text{KnownDomain}) + (0.75) \cdot (\text{'FREE'})] \approx 0.52$$

- Loss: 0 (correct!)
- ***Feedback ("gradient"): no change***

ID	# Words	Domain	"FREE"	Label
0	120	gmail.com	Yes	Spam
1	500	gmail.com	No	Ham
2	400	cmu.edu	No	Ham

Learning from Experience

- Initial Guess:

$$\text{Odds(Spam/Ham)} = \exp[(-0.001)\cdot(\#\text{Words}) + (-0.25)\cdot(\text{KnownDomain}) + (0.75)\cdot(\text{'FREE'})] \approx 0.64$$

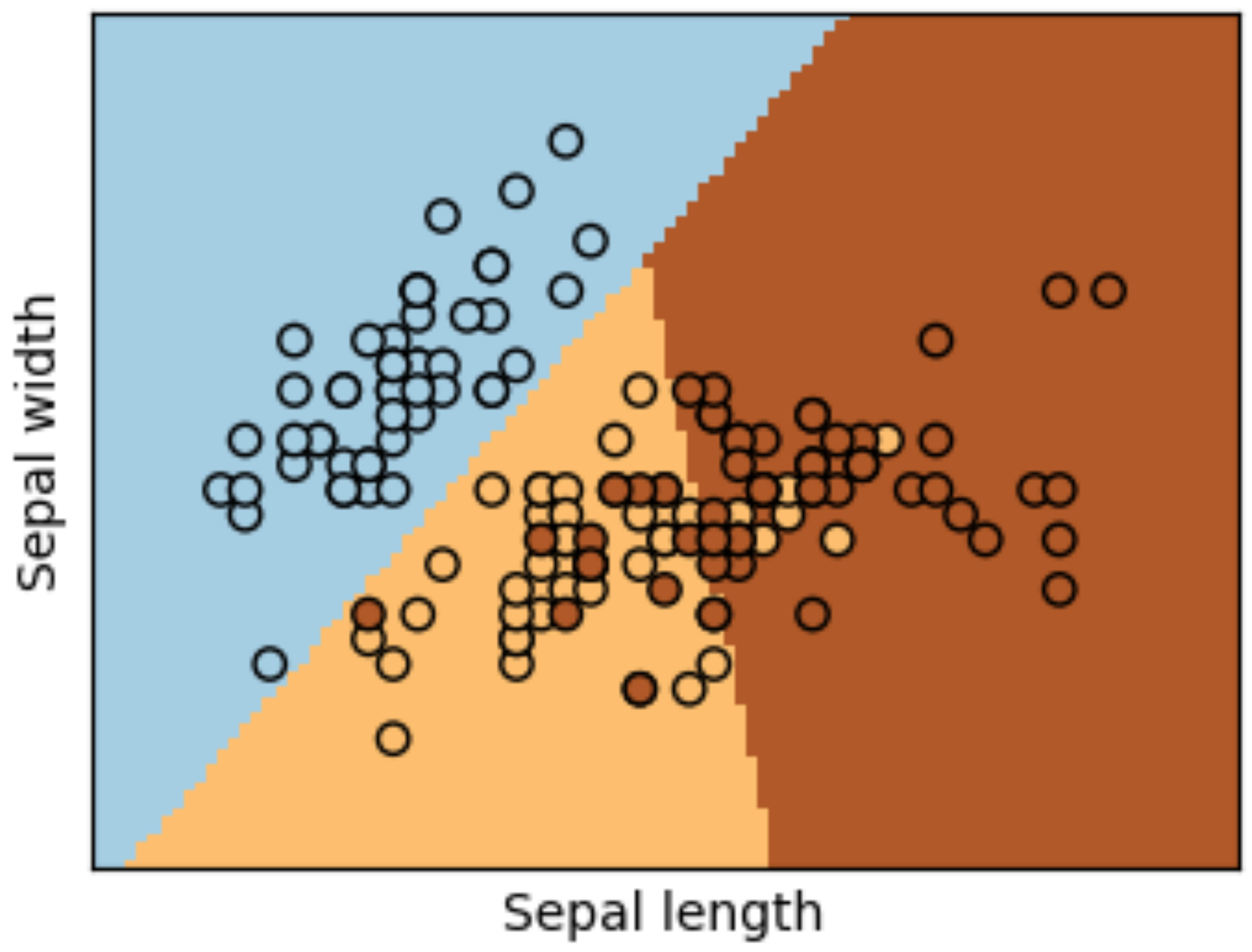
- Loss: 1 (incorrect)

- ***Feedback ("gradient"): Add 0.0005 to w1, -0.1 to w2, +0.1 to w3***

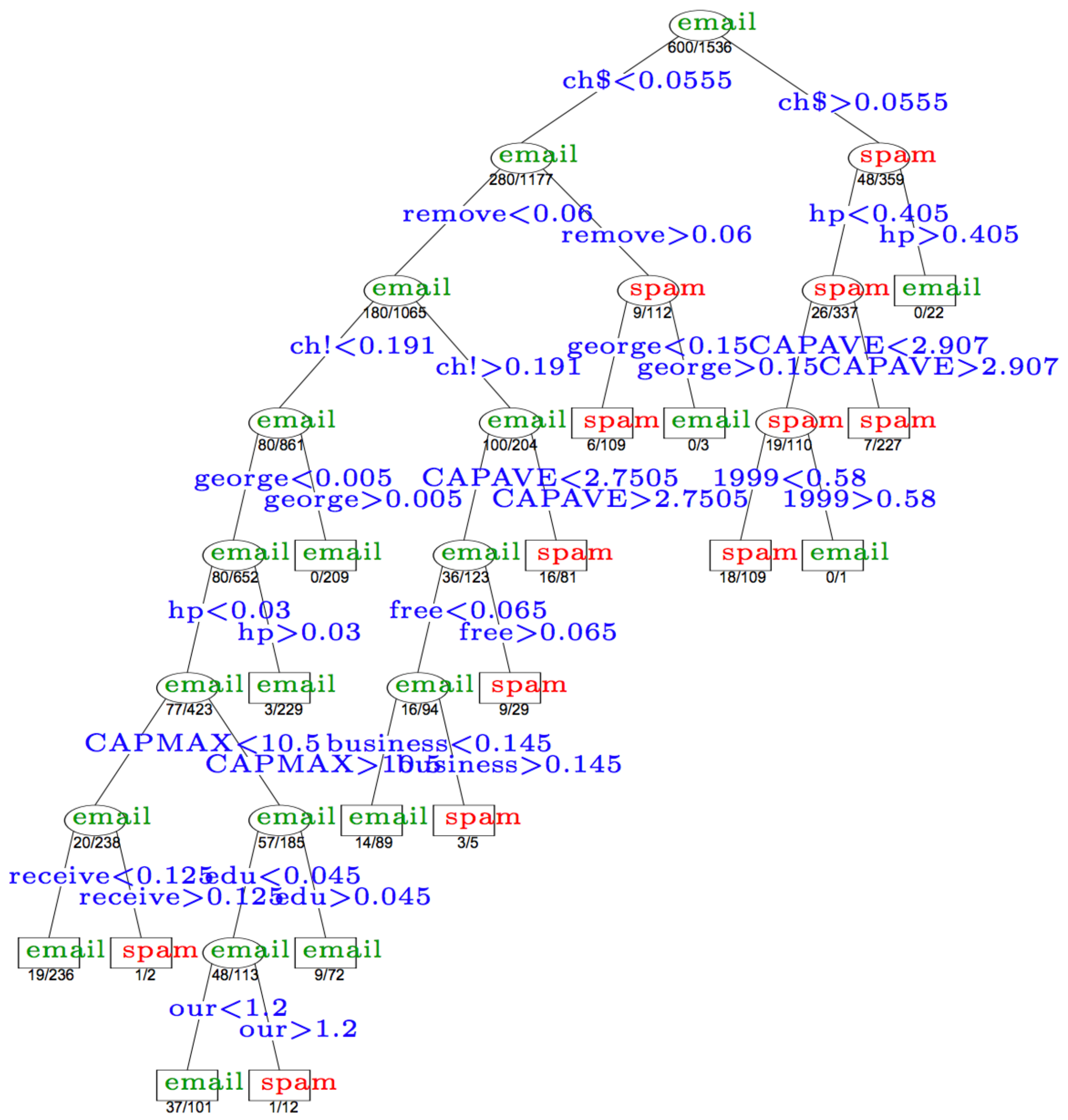
ID	# Words	Domain	"FREE"	Label
0	120	gmail.com	Yes	Spam
1	500	gmail.com	No	Ham
2	400	cmu.edu	No	Ham
3	1200	suspicio.us	Yes	Spam

Supervised ML Methods

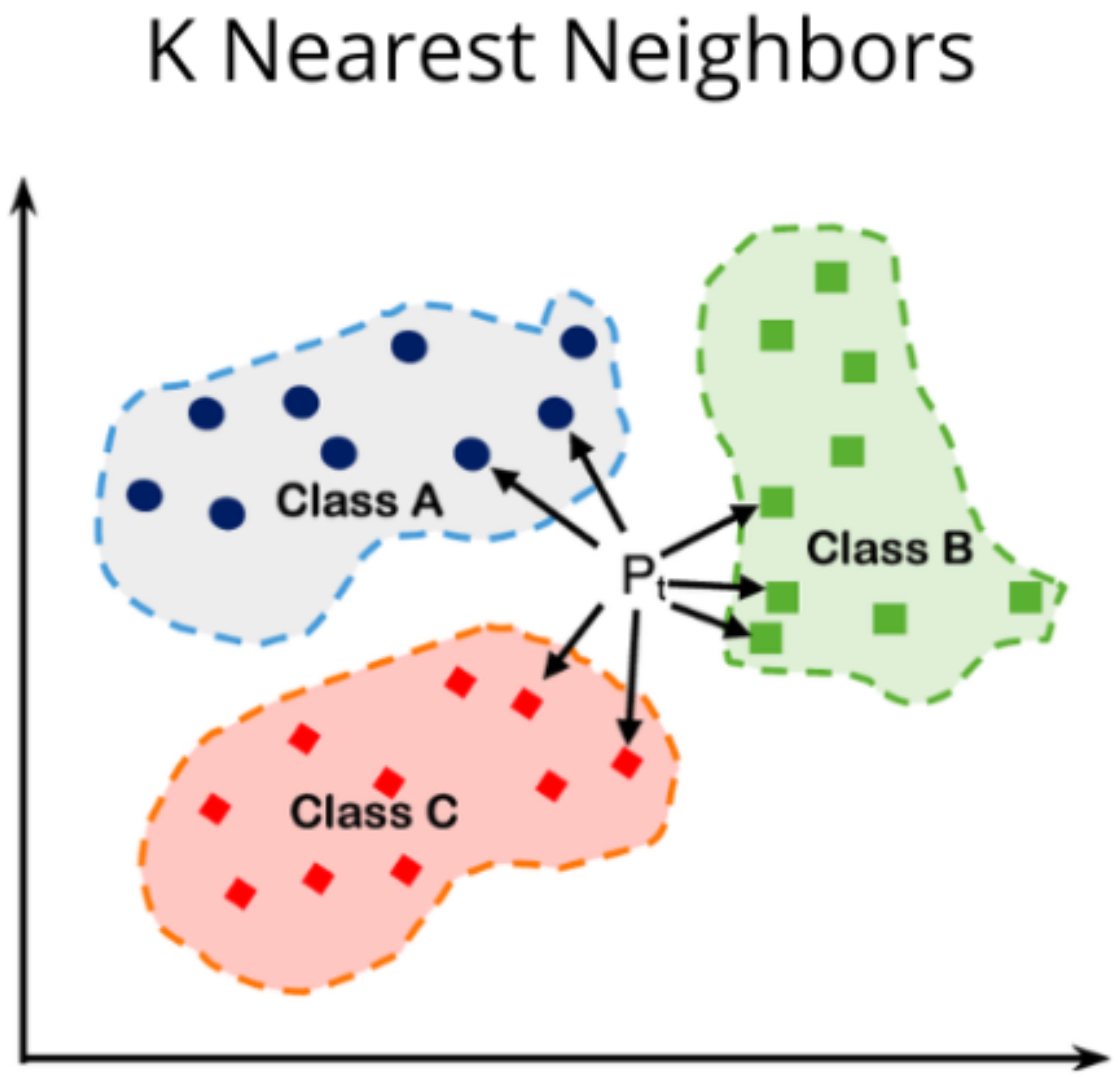
Linear/Logistic Regression



Decision Trees

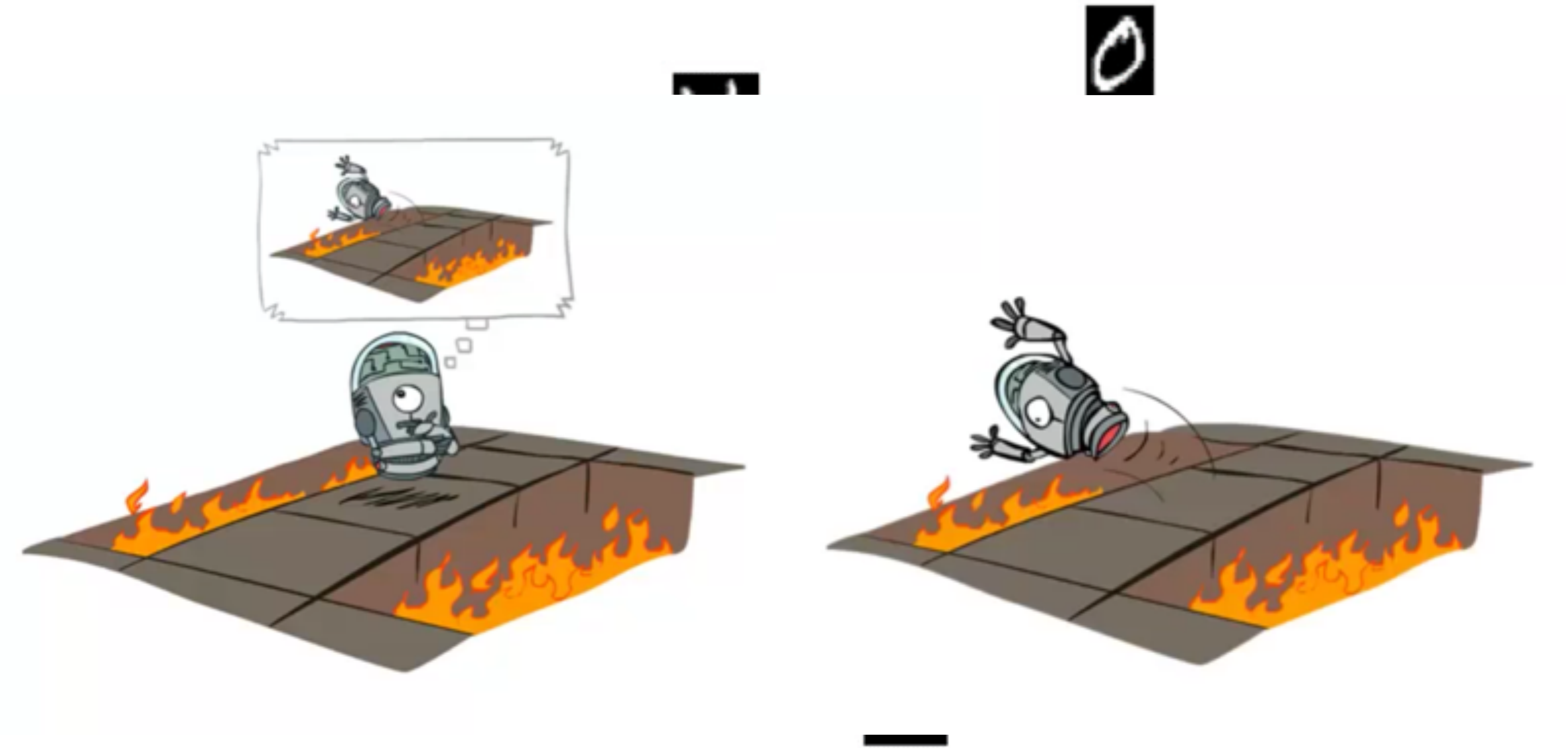


Nearest Neighbors



(...and many, many more)

OTHER TYPES OF ML TASKS



- **Unsupervised Learning:** no labels
- **Semi-supervised Learning**
- **Pre-training & Fine-tuning**
- **Reinforcement Learning:** (input, state) -> (action, next_state)
- **Online Learning, Active Learning, Meta-Learning, …**

Cases When Classical ML Might Be Useful

- **“Tabular data” with meaningful hand-crafted features**
  - Predicting an individual’s credit/loan, political leanings, hiring decisions, etc.
  - Predicting the likelihood of cancer using key health measures
  - Predicting the chance of recidivism using demographic/socioeconomic information
- **Small data scenario**
  - Medical data analysis / genomics / neuroscience / rare events
- **(Arguably) whenever transparency/explanability/interpretability is needed**
  - Would you use AI for policing, hiring, loan decisions, etc.?

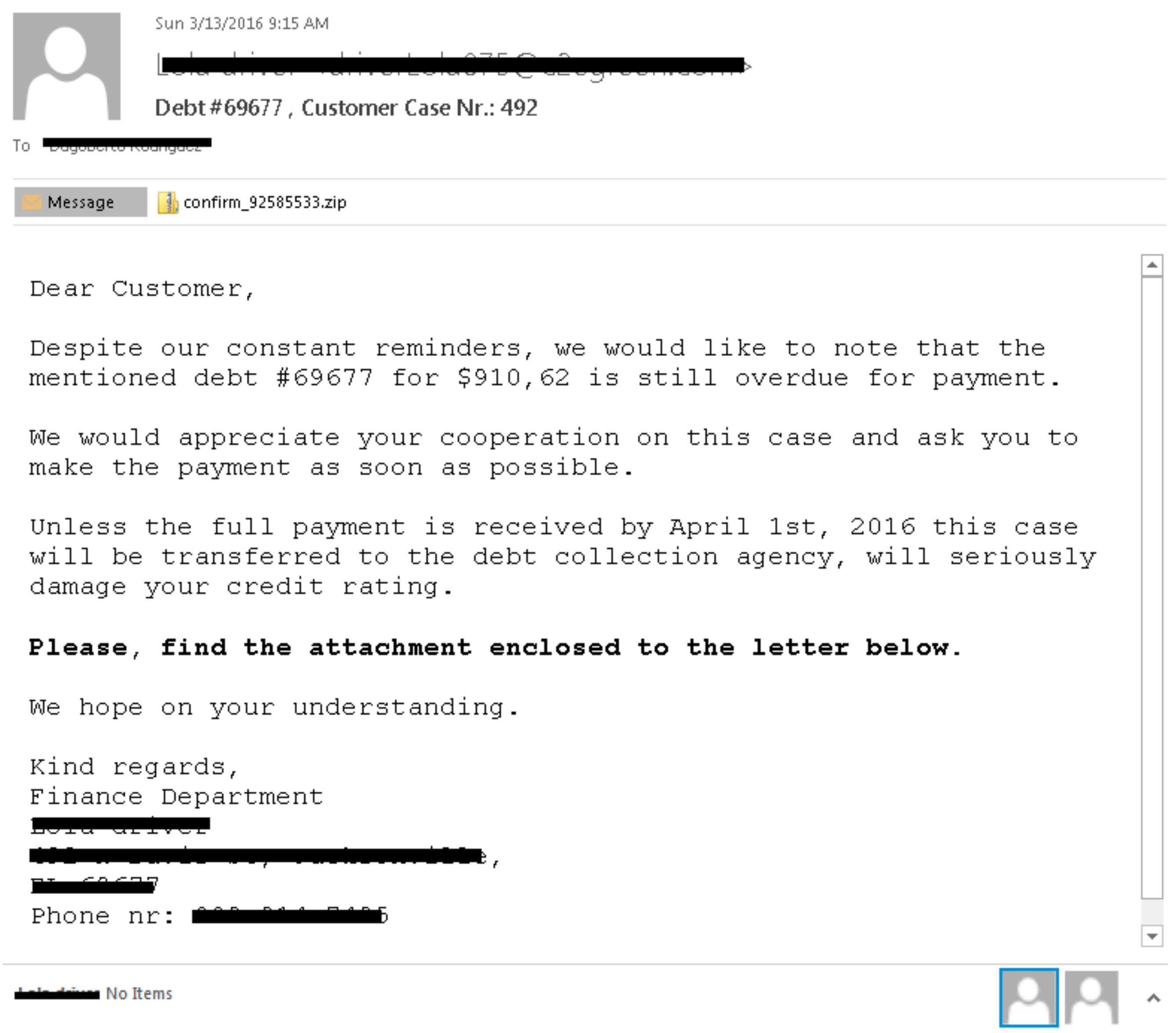
What is Deep Learning?

https://xkcd.com/1425/

These five years are well past us now! (Roughly, 2012-2017)

The Perceptron

(Rosenblatt 1959)



Sun 3/13/2016 9:15 AM

[redacted]

Debt #69677 , Customer Case Nr.: 492

To [redacted]

Message | confirm_92585533.zip

Dear Customer,

Despite our constant reminders, we would like to note that the mentioned debt #69677 for $910,62 is still overdue for payment.

We would appreciate your cooperation on this case and ask you to make the payment as soon as possible.

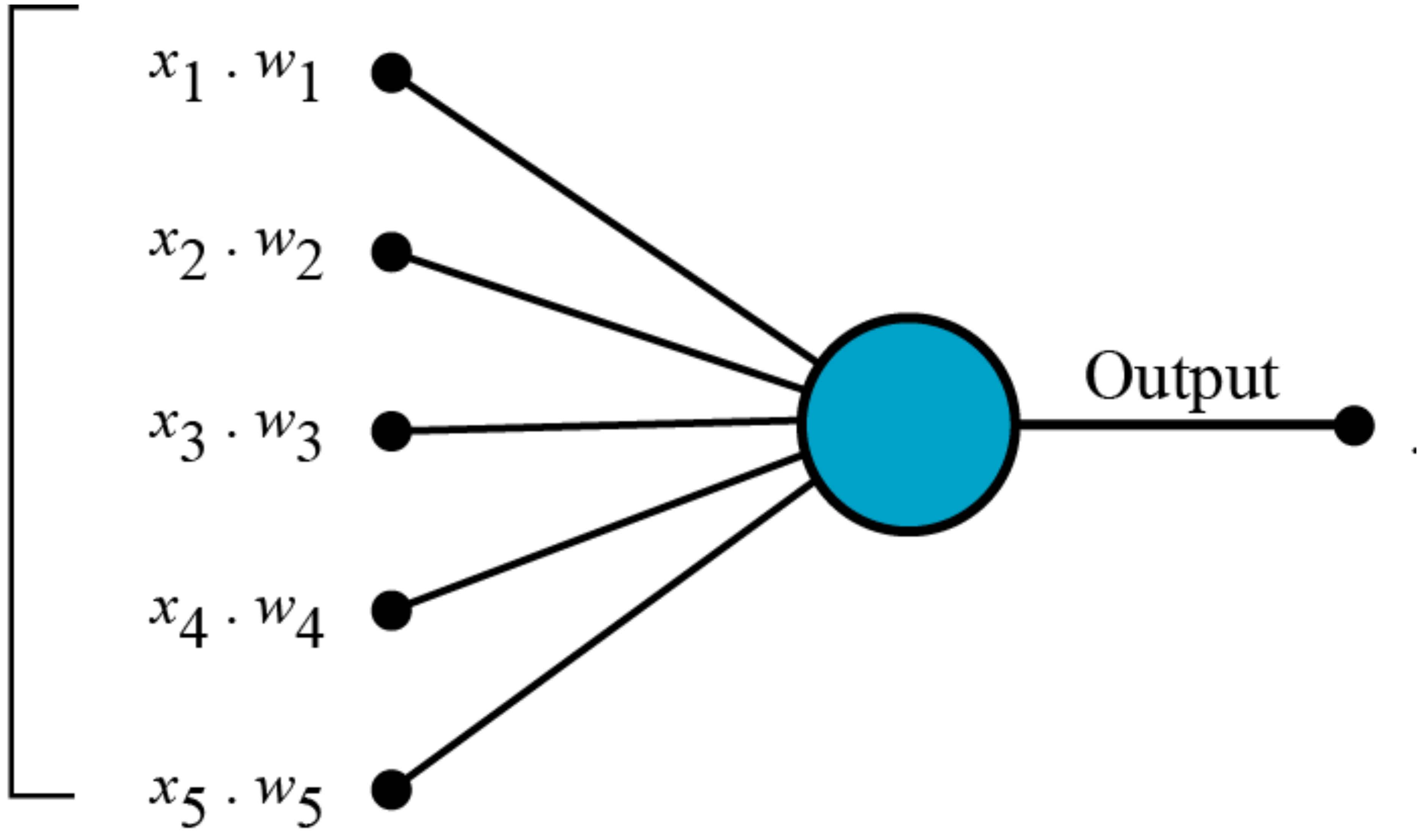
Unless the full payment is received by April 1st, 2016 this case will be transferred to the debt collection agency, will seriously damage your credit rating.

Please, find the attachment enclosed to the letter below.

We hope on your understanding.

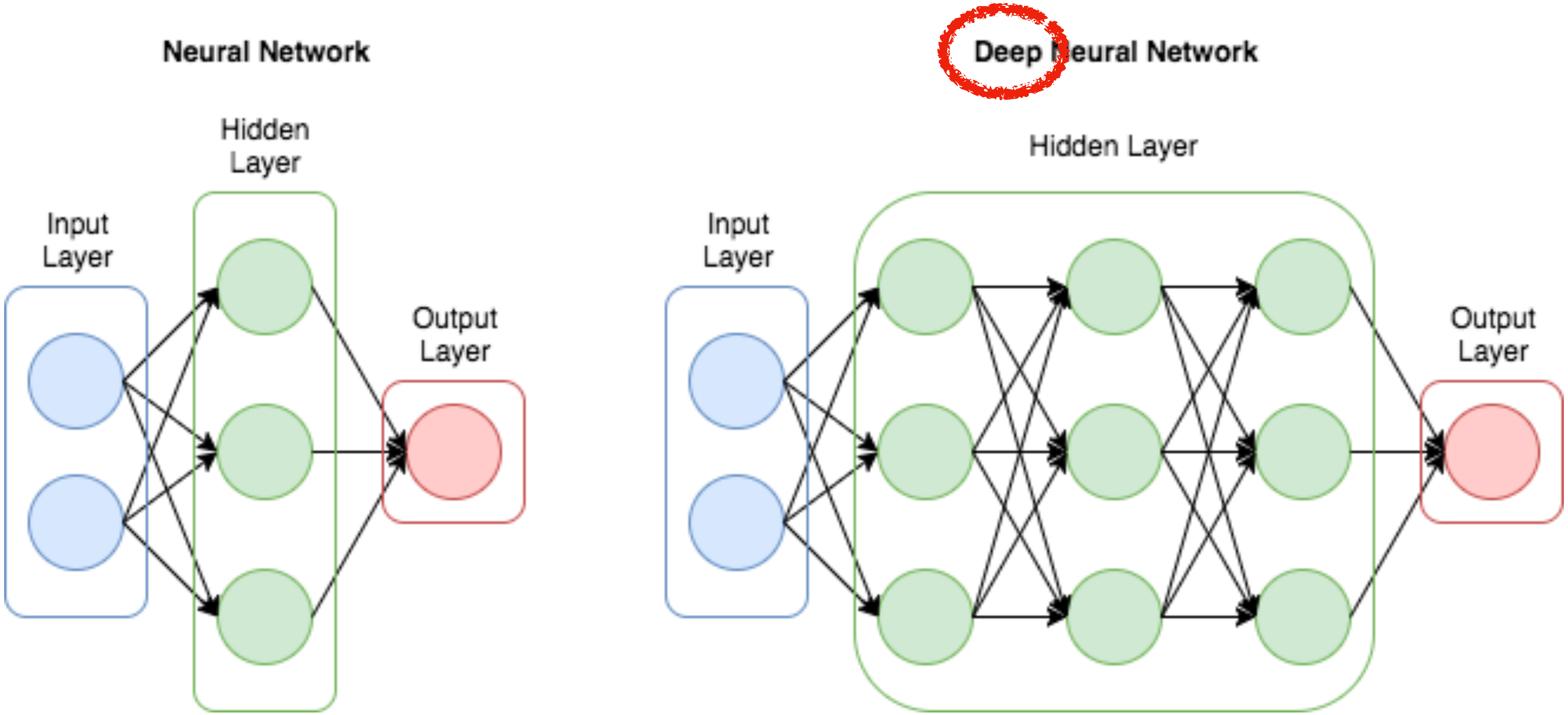
Kind regards,
Finance Department
[redacted],
Phone nr: [redacted]

[redacted] No Items



https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53

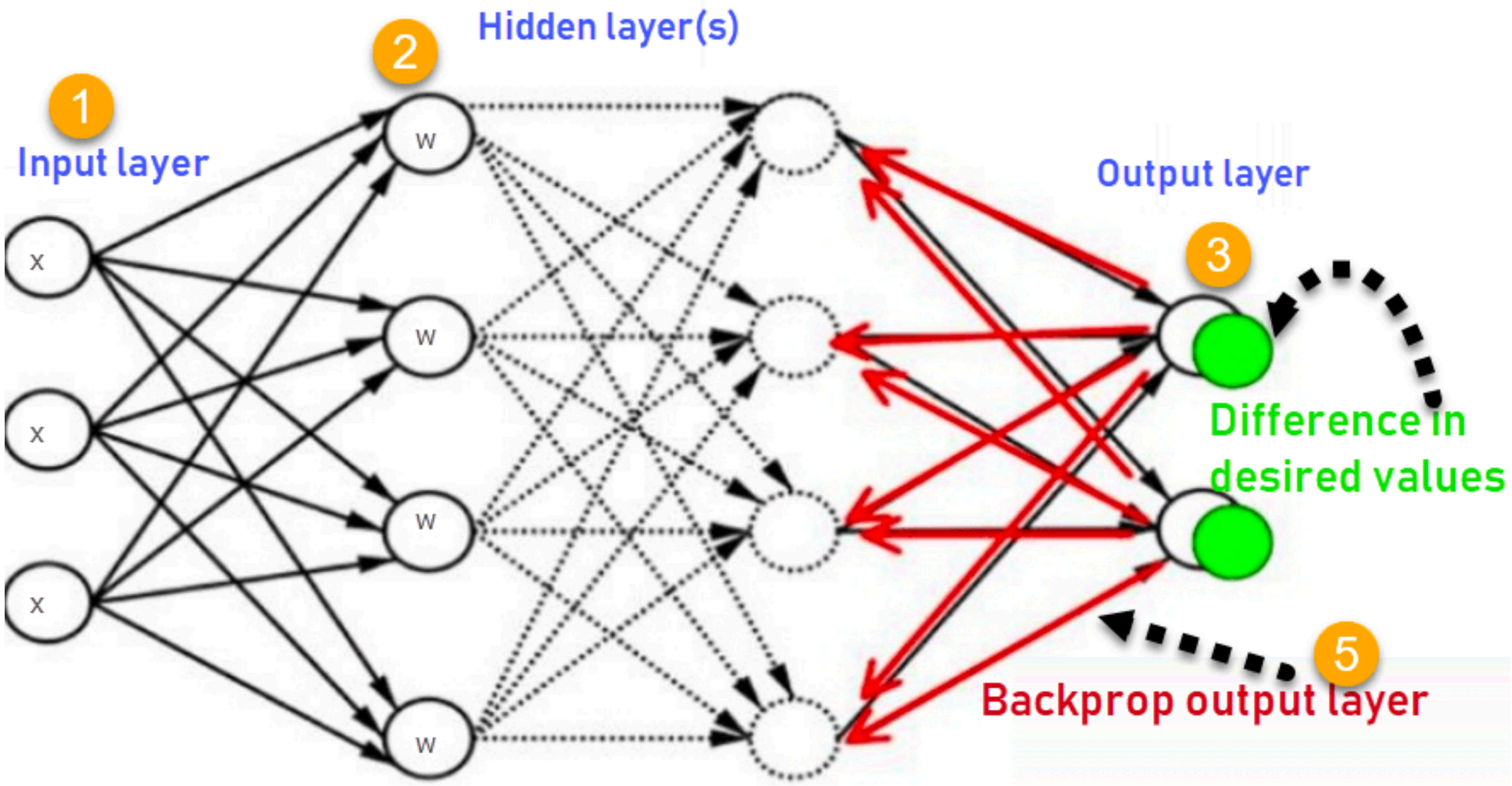
Artificial Neural Networks



http://marubon-ds.blogspot.com/2017/09/simple-tutorial-to-write-deep-neural.html

BACKPROPAGATION

(Rumelhart et al., 1986)

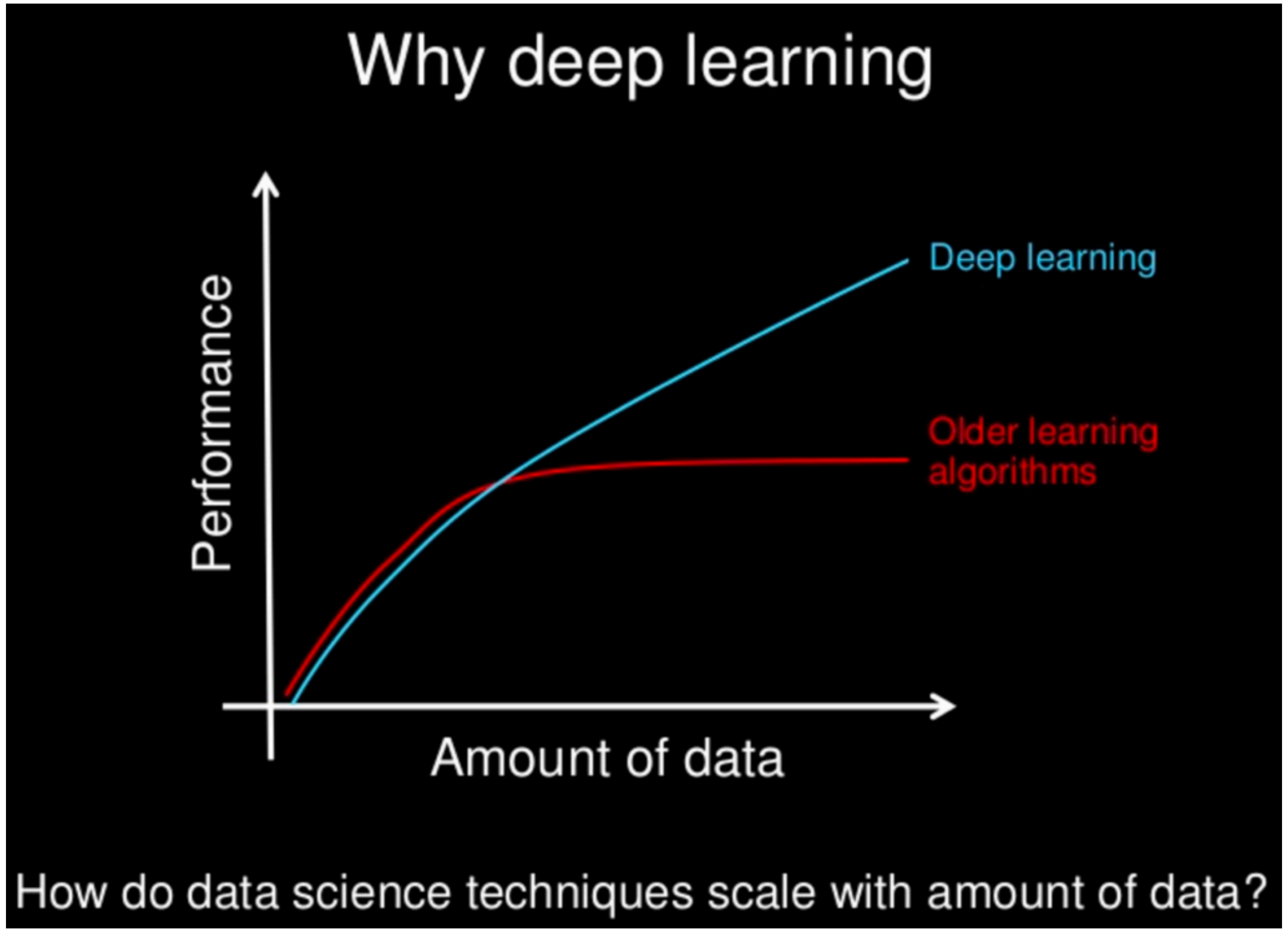


https://www.guru99.com/backpropogation-neural-network.html

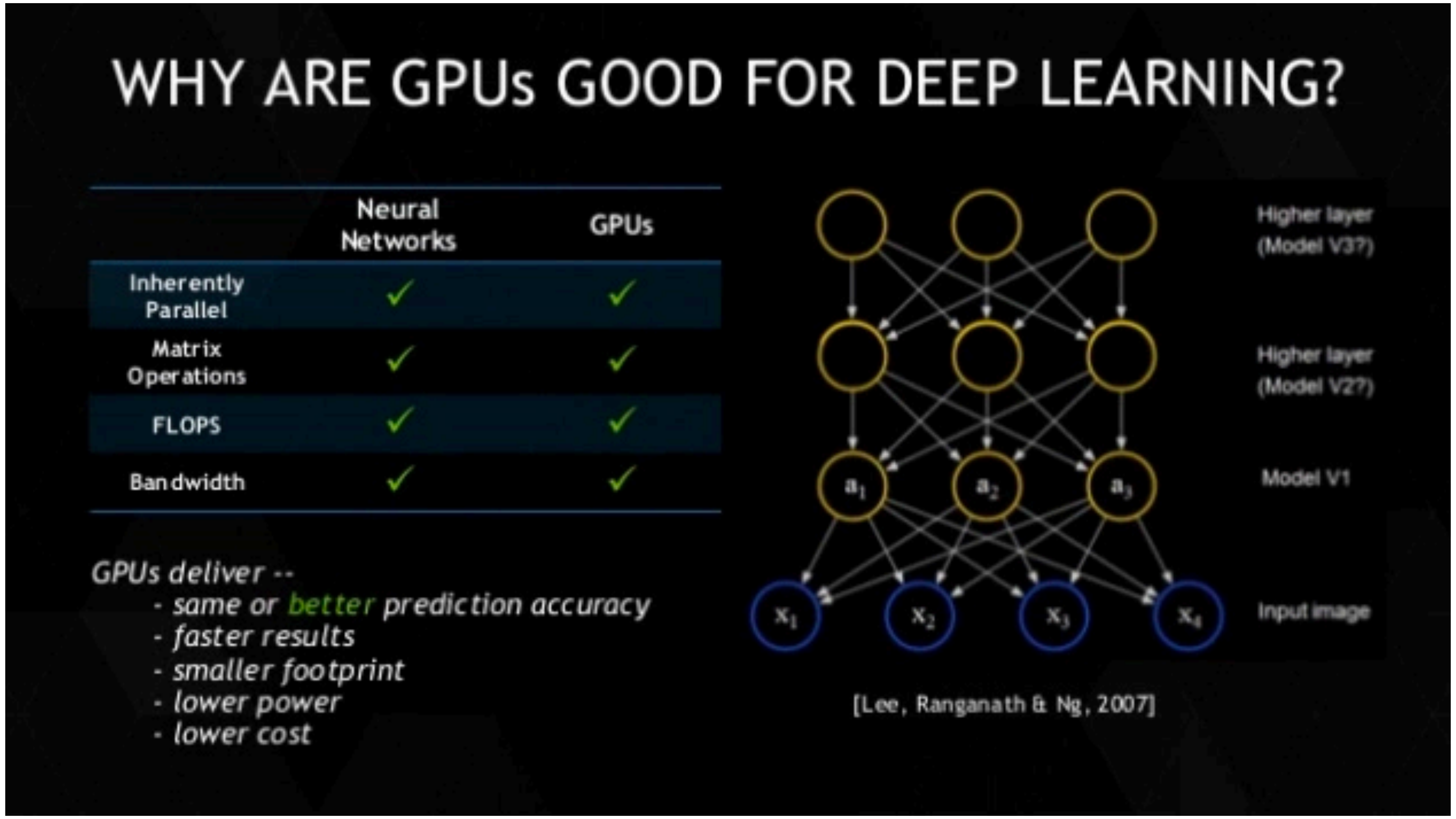
WHY "DEEP" LEARNING?

Deep learning is **data-efficient**.

Deep learning is **modular**, and thus it can take advantage of **GPUs** that can efficiently process billions of simple math operations quickly.

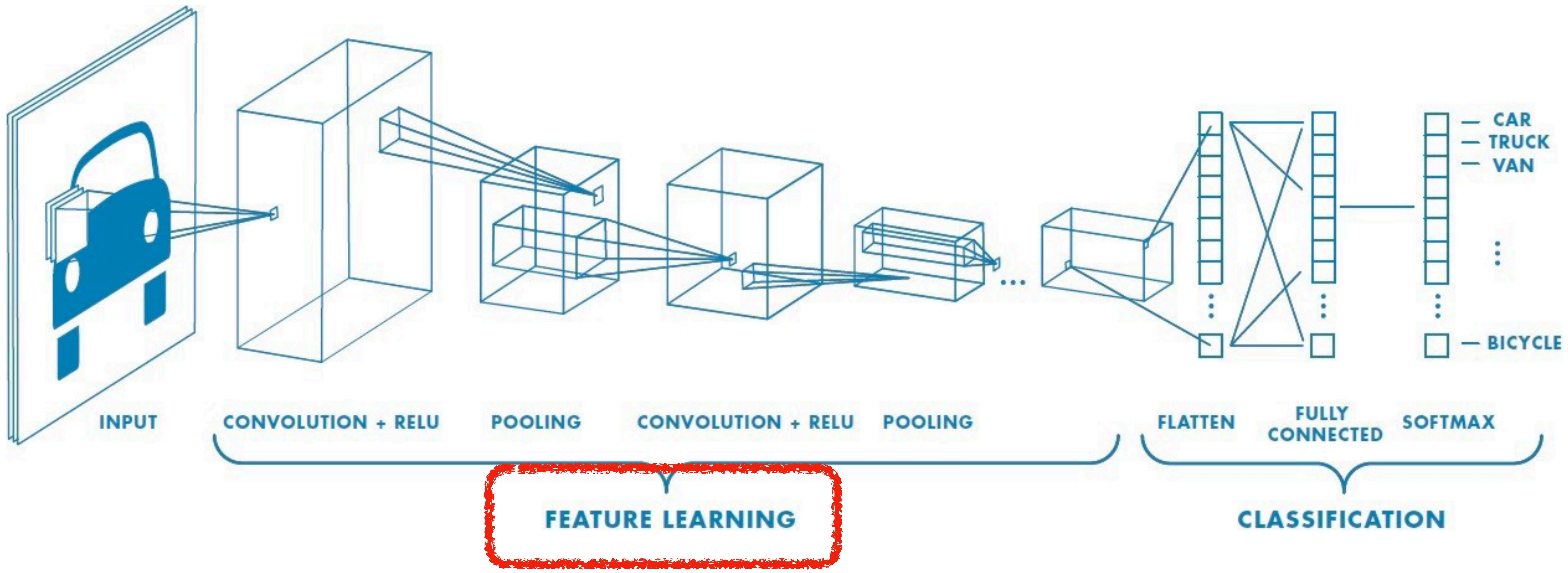


Andrew Ng's Slides



NVIDIA's Slides

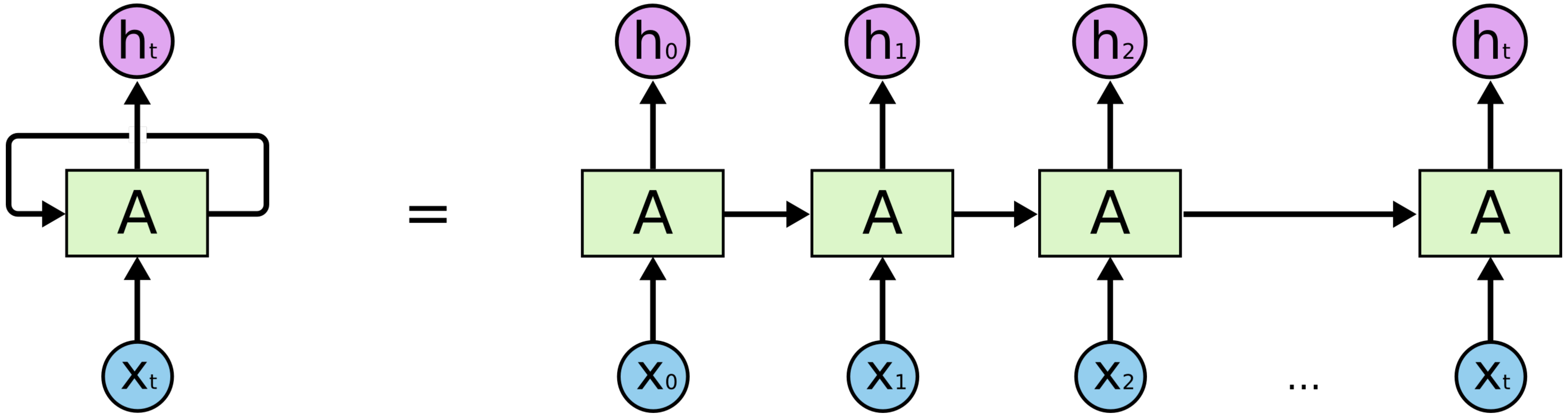
Convolutional Neural Networks (CNNs)



Deep learning models learn representations!

https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

RECURRENT NEURAL NETWORKS (RNNS)



https://colah.github.io/posts/2015-08-Understanding-LSTMs/

For $\bigoplus_{n=1,\dots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

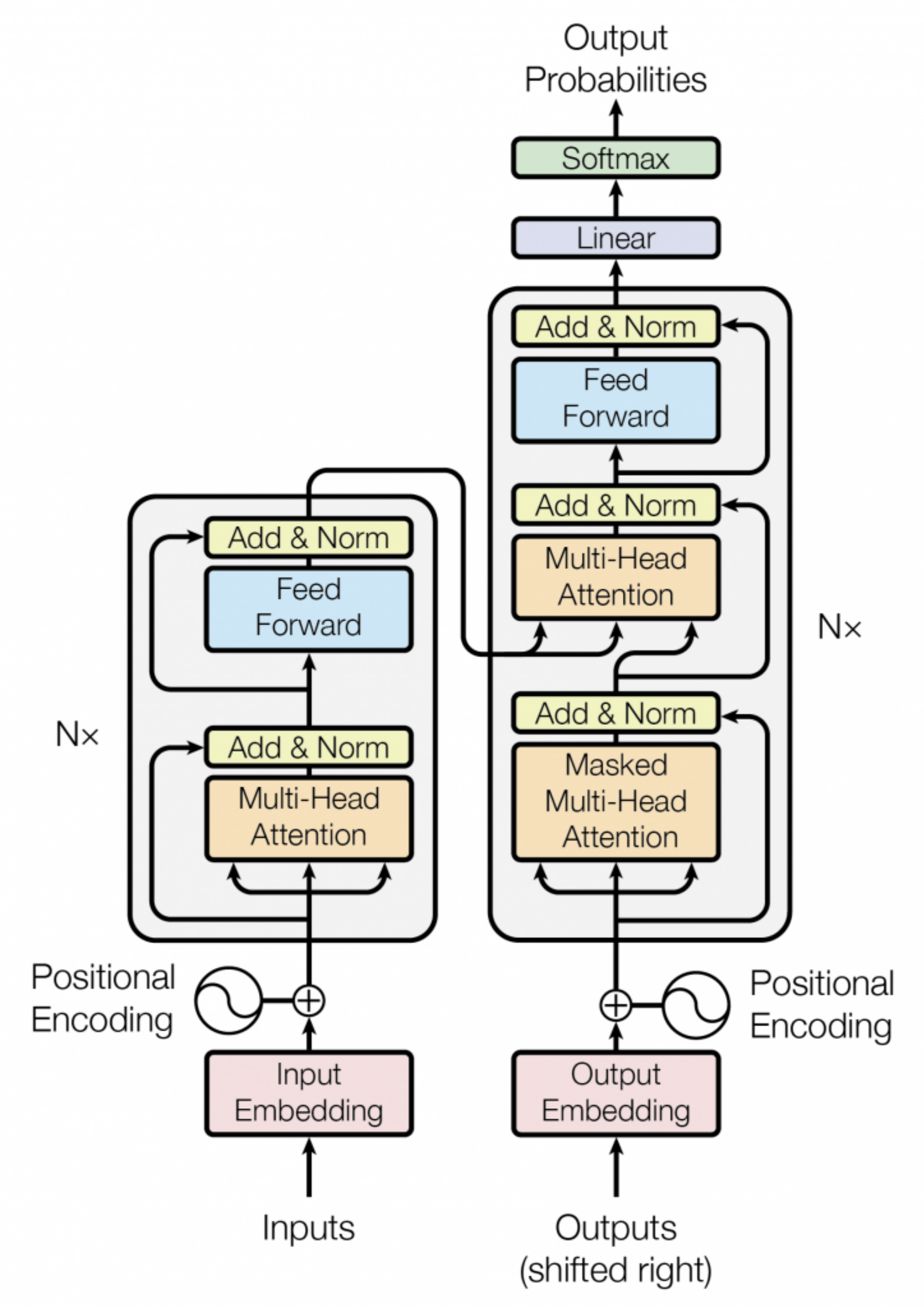
is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

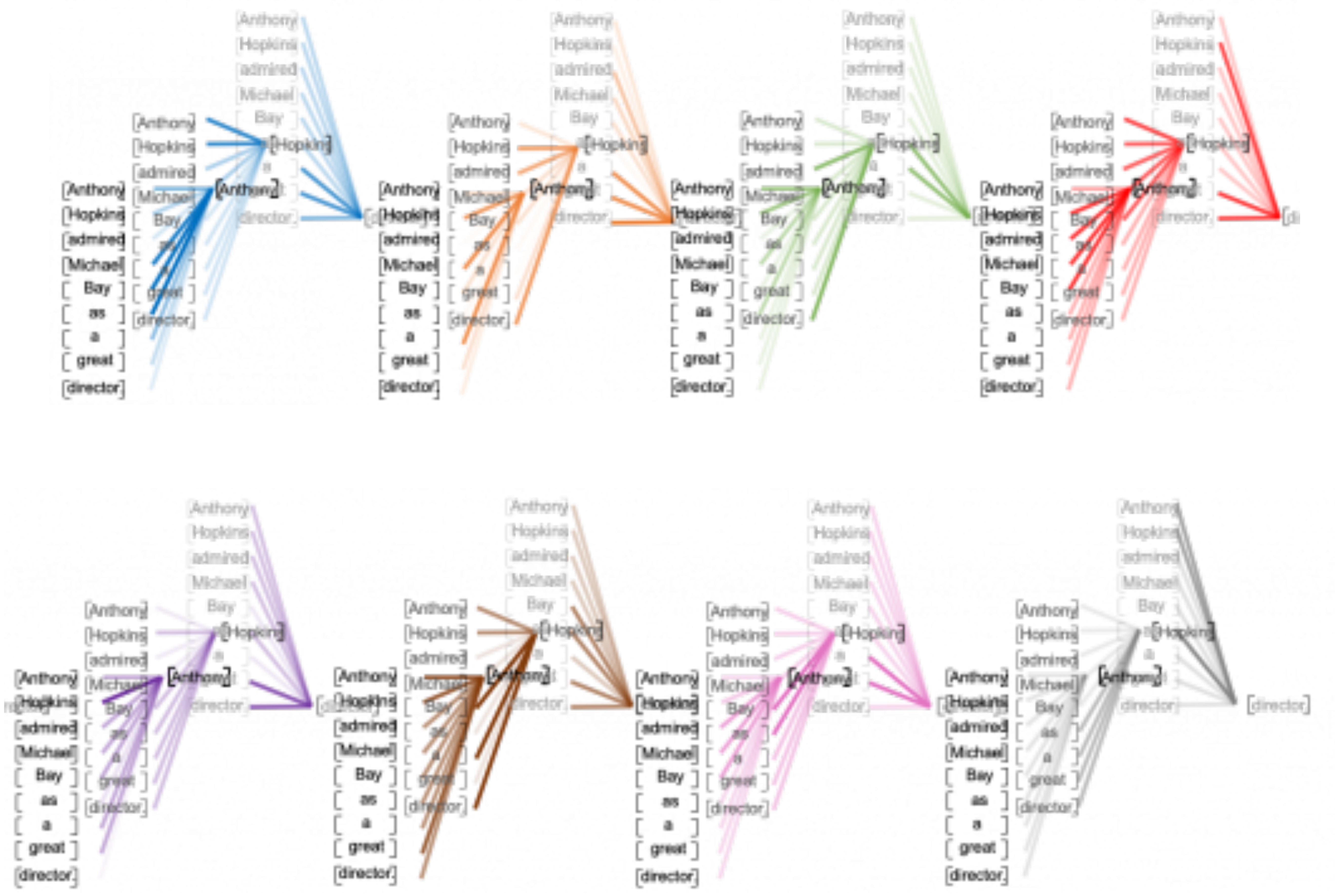
Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

TRANSFORMERS (SELF-ATTENTION NNS)



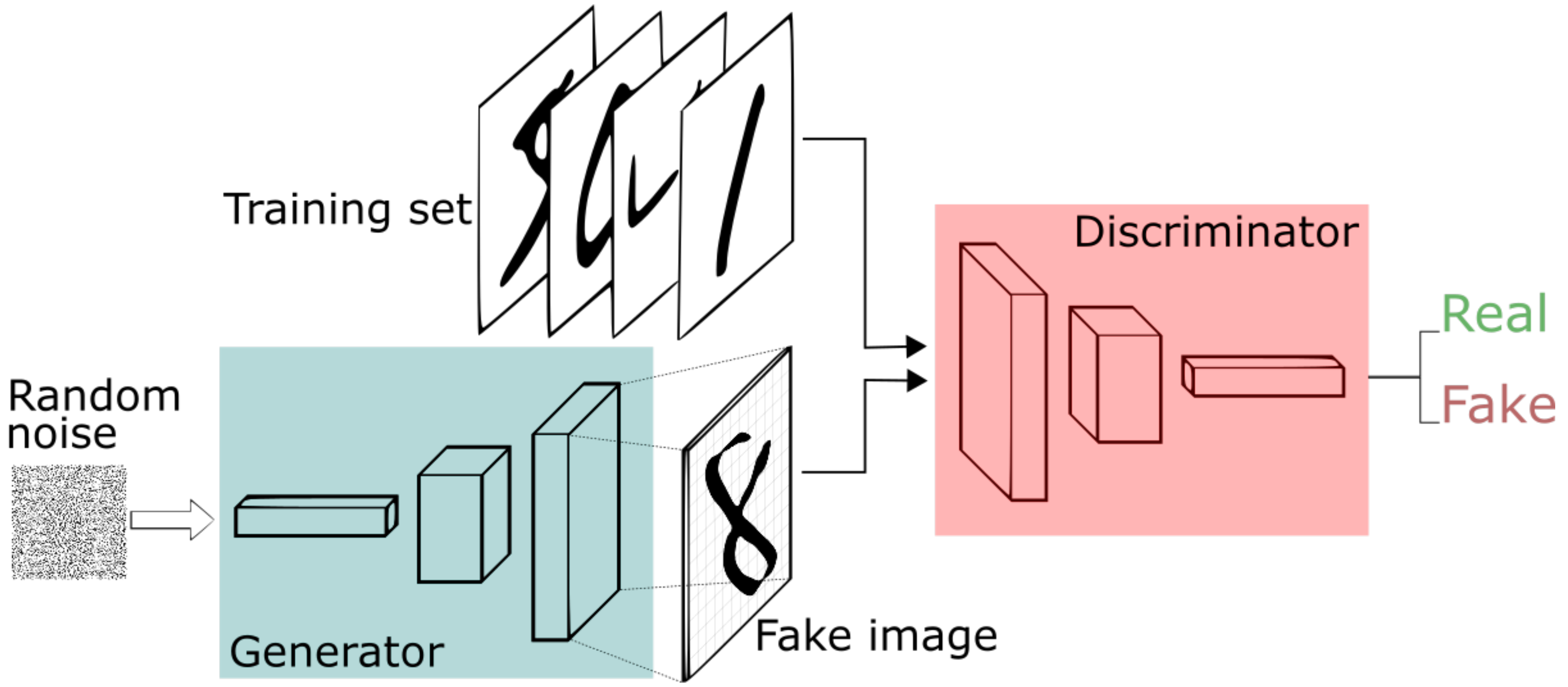


https://data-science-blog.com/blog/2021/04/07/multi-head-attention-mechanism/

Vaswani et al. (2017). “Attention is all you need.”

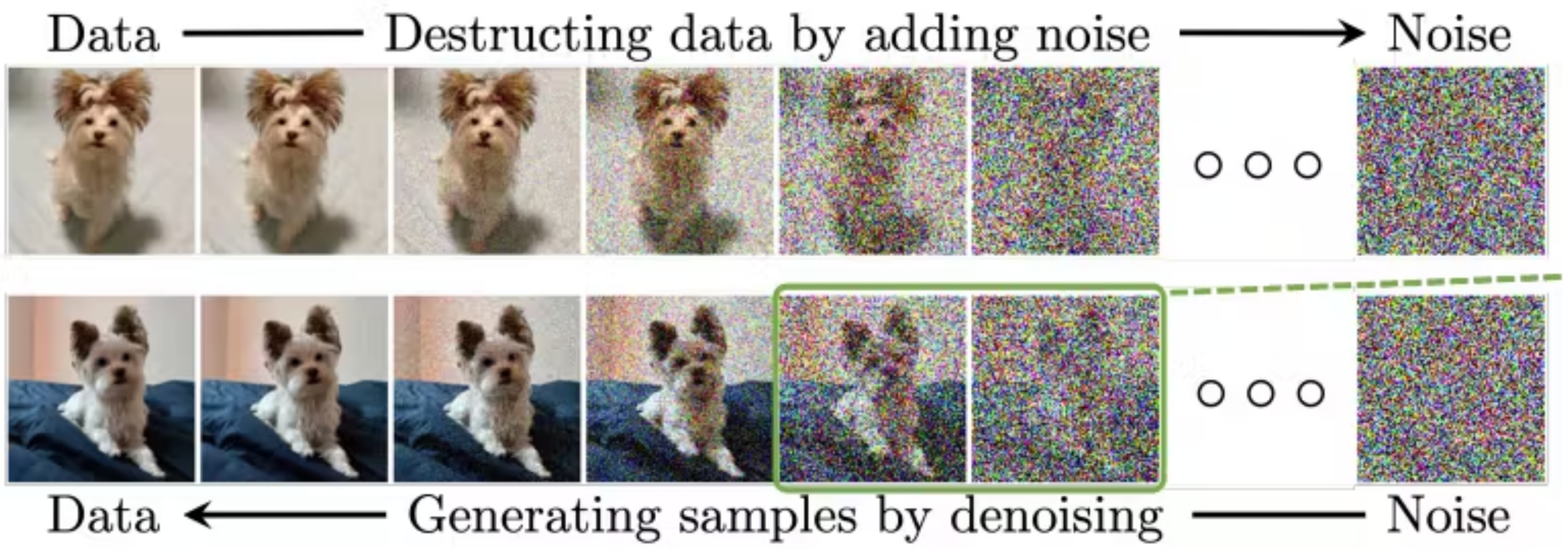
DEEP GENERATIVE MODELS

Generative Adversarial Networks (GANs)



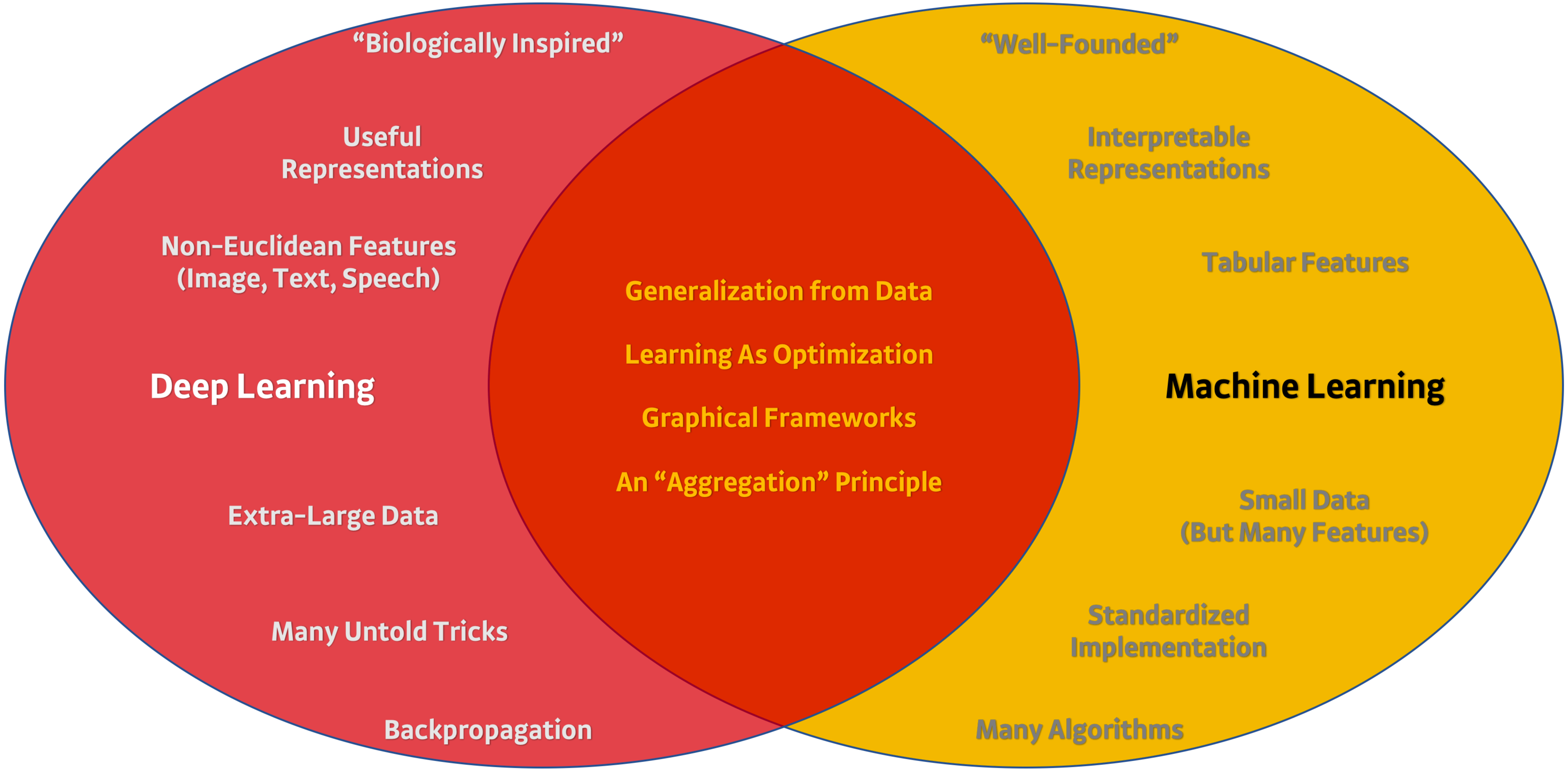
https://medium.freecodecamp.org/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394

Diffusion Models



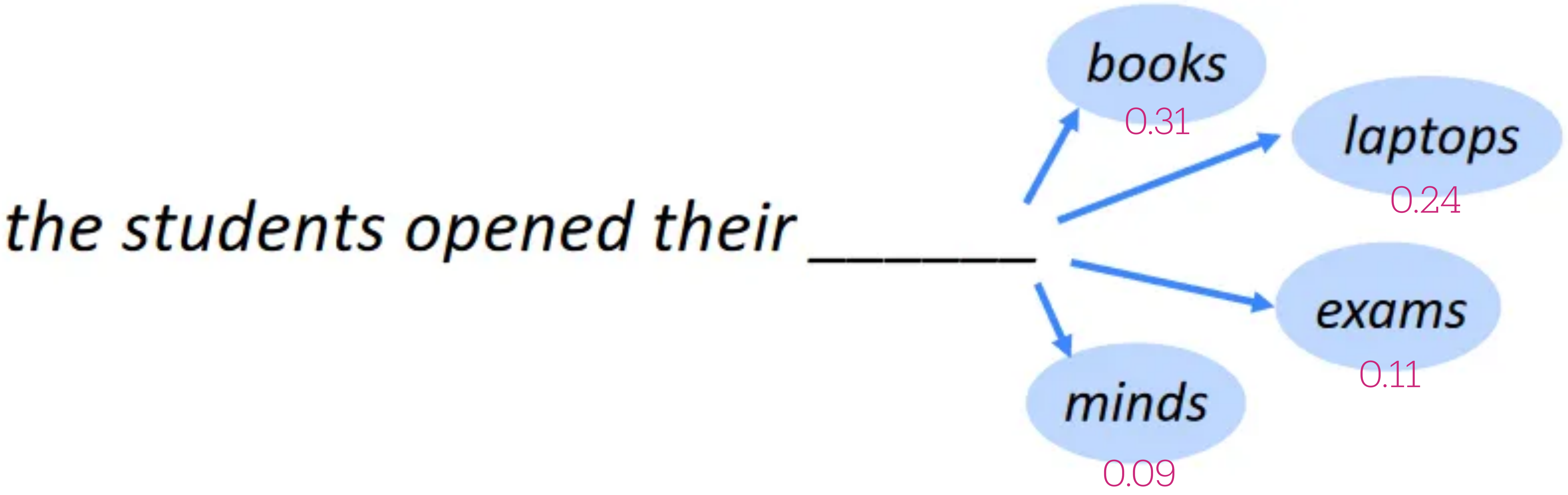
Yang et al. (2022). "Diffusion Models: A Comprehensive Survey of Methods and Applications"

Deep Learning vs. Classical Machine Learning



Large Language Models (LLMs)

Language Modeling*



Task: Estimate the probability of the next word given each context

*More formally, (forward) statistical language modeling

LM ISN'T *JUST* A SUPERVISED TASK

- On the surface, language modeling appears to be a **supervised** task. (Input: context, Output: next word)

- But we don't need to manually provide labels (a **self-supervised** task).

  - A sentence consisting of N words turns into (essentially) N data points.

The	**students**				
The	students	**opened**			
The	students	opened	**their**		
The	students	opened	their	**books**	
The	students	opened	their	books	**.**

This one sentence contains 5 examples for the "next word prediction" task.

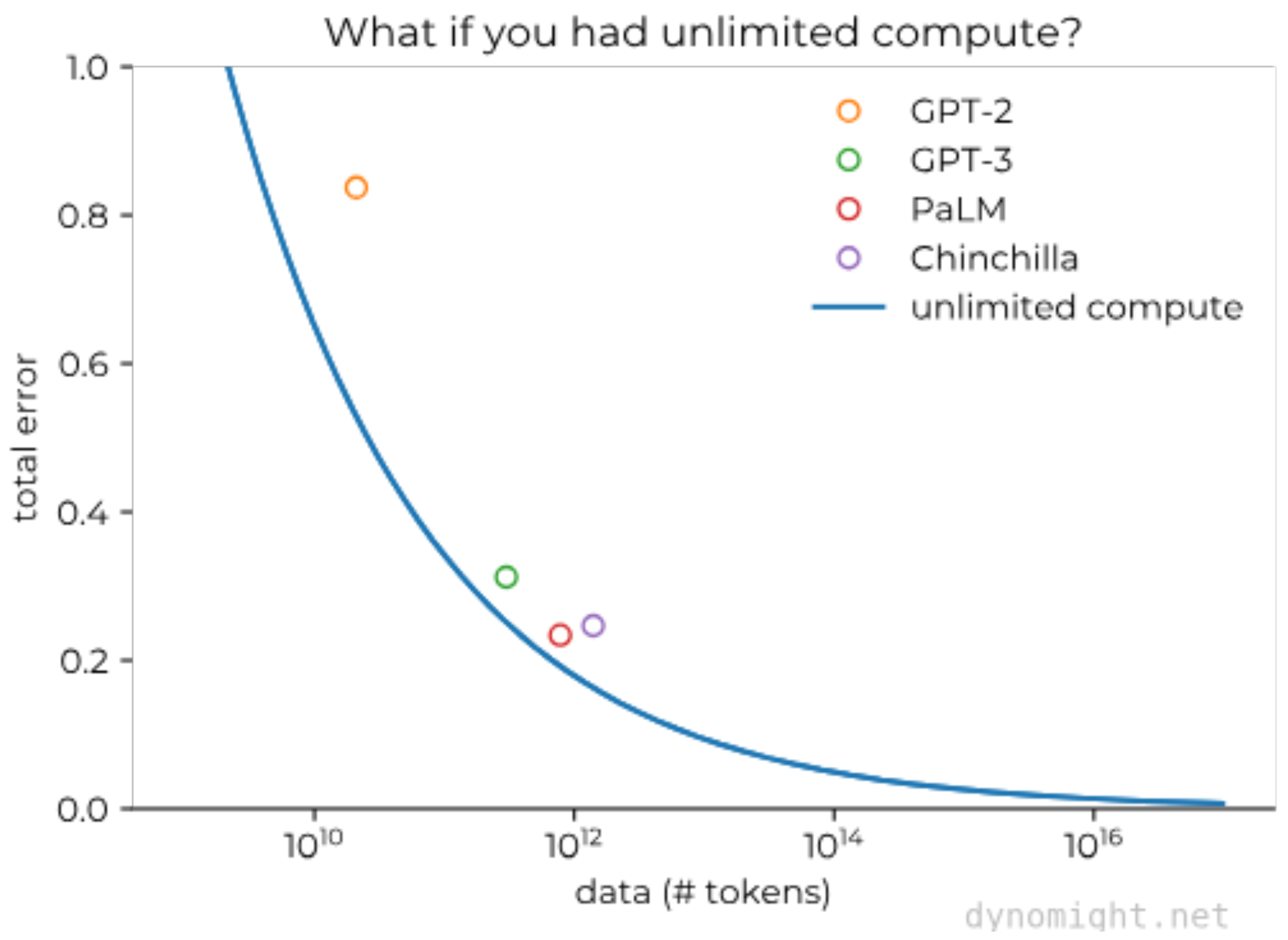
This also depicts the ***generative*** *process (after training)!*

"LARGE" LANGUAGE MODELING: A RECIPE

1. Get a LOT of text-form **data** (trillions of words/symbols).

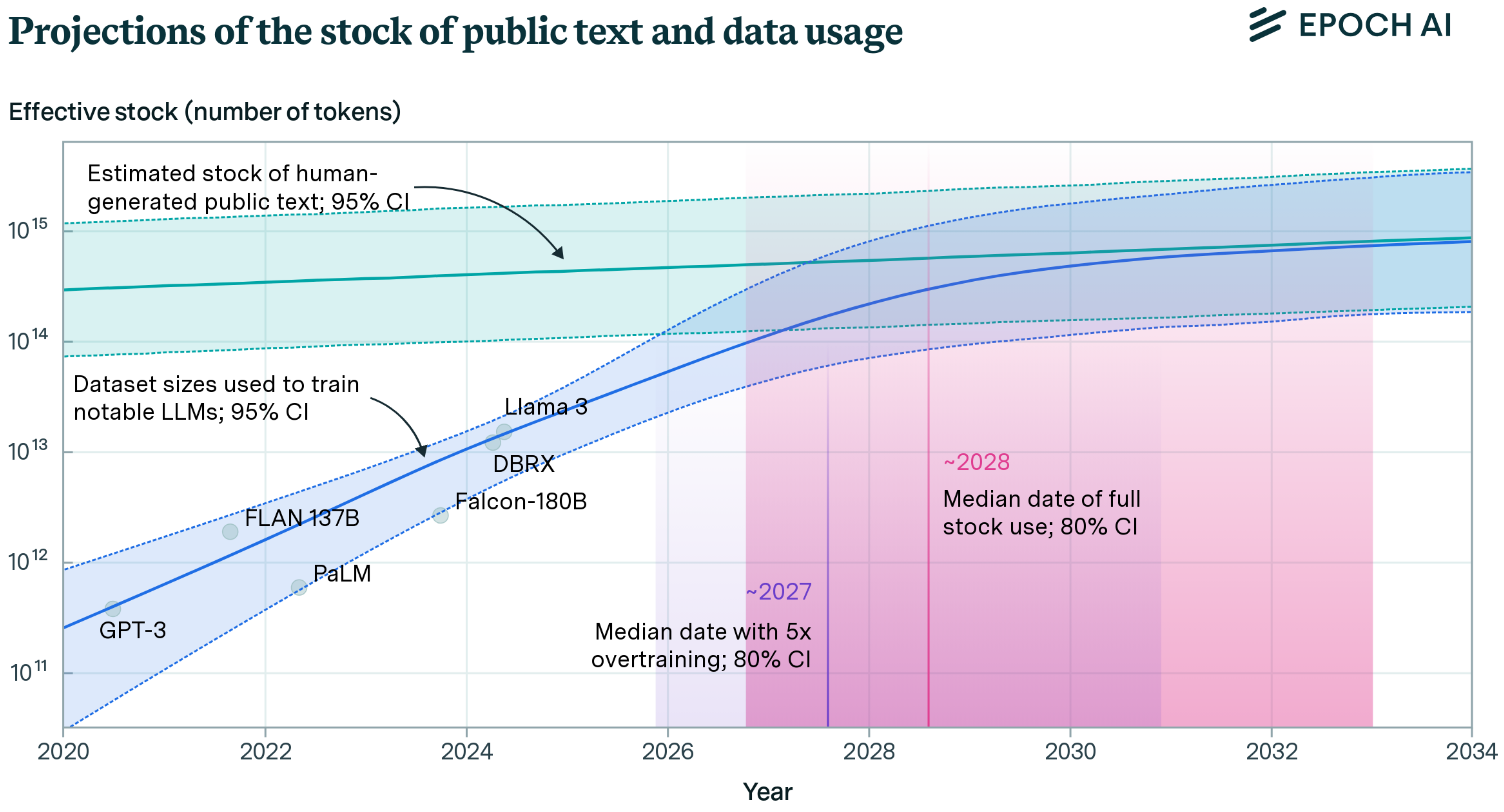
2. Set up a large *Transformer* model (billions of **parameters**).

3. Train it on the language modeling task with many, many **compute-hours** on GPU.



https://dynomight.net/scaling/

Where we might be at in terms of data

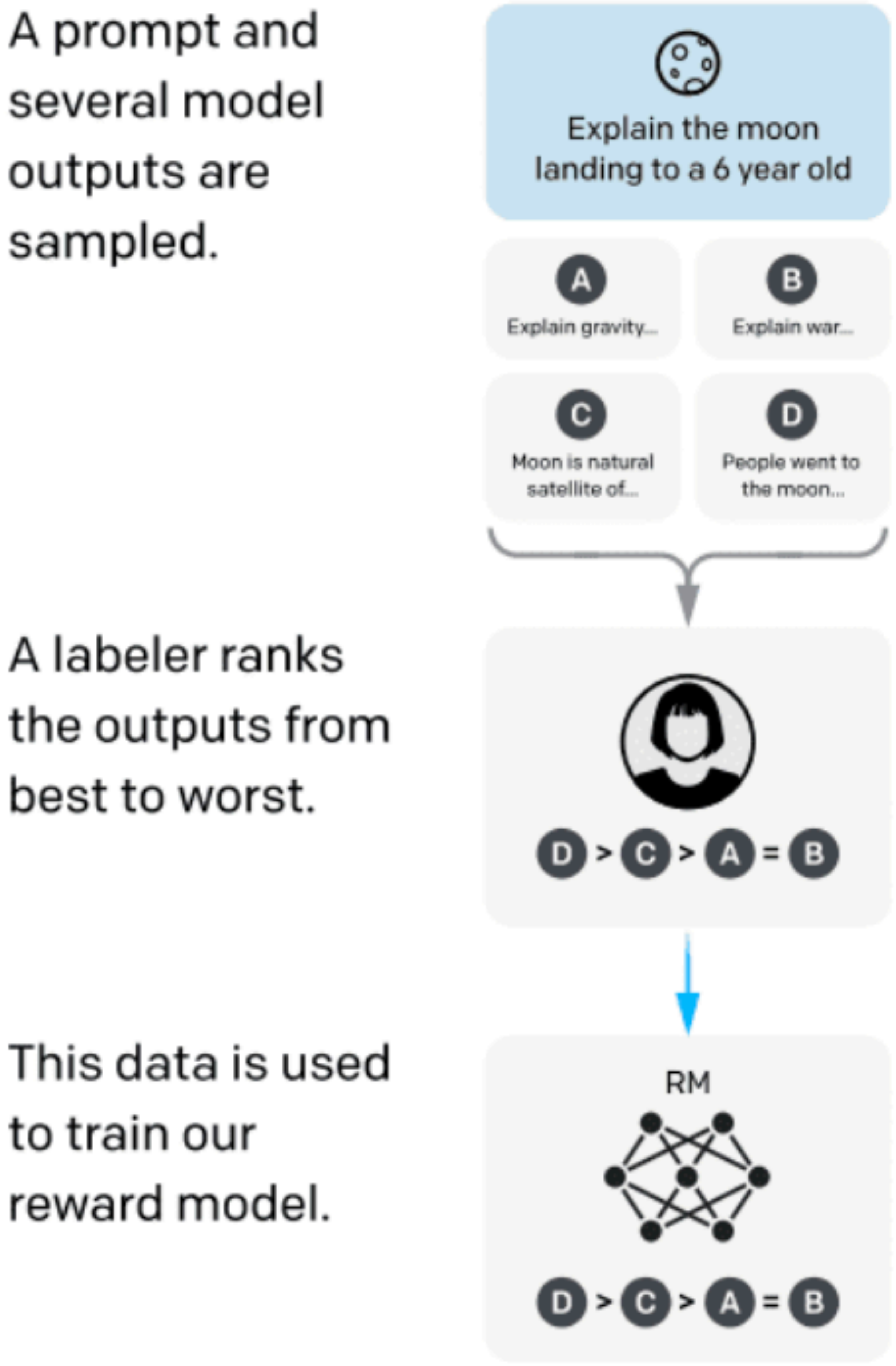


Are LLMs "Intelligent"?

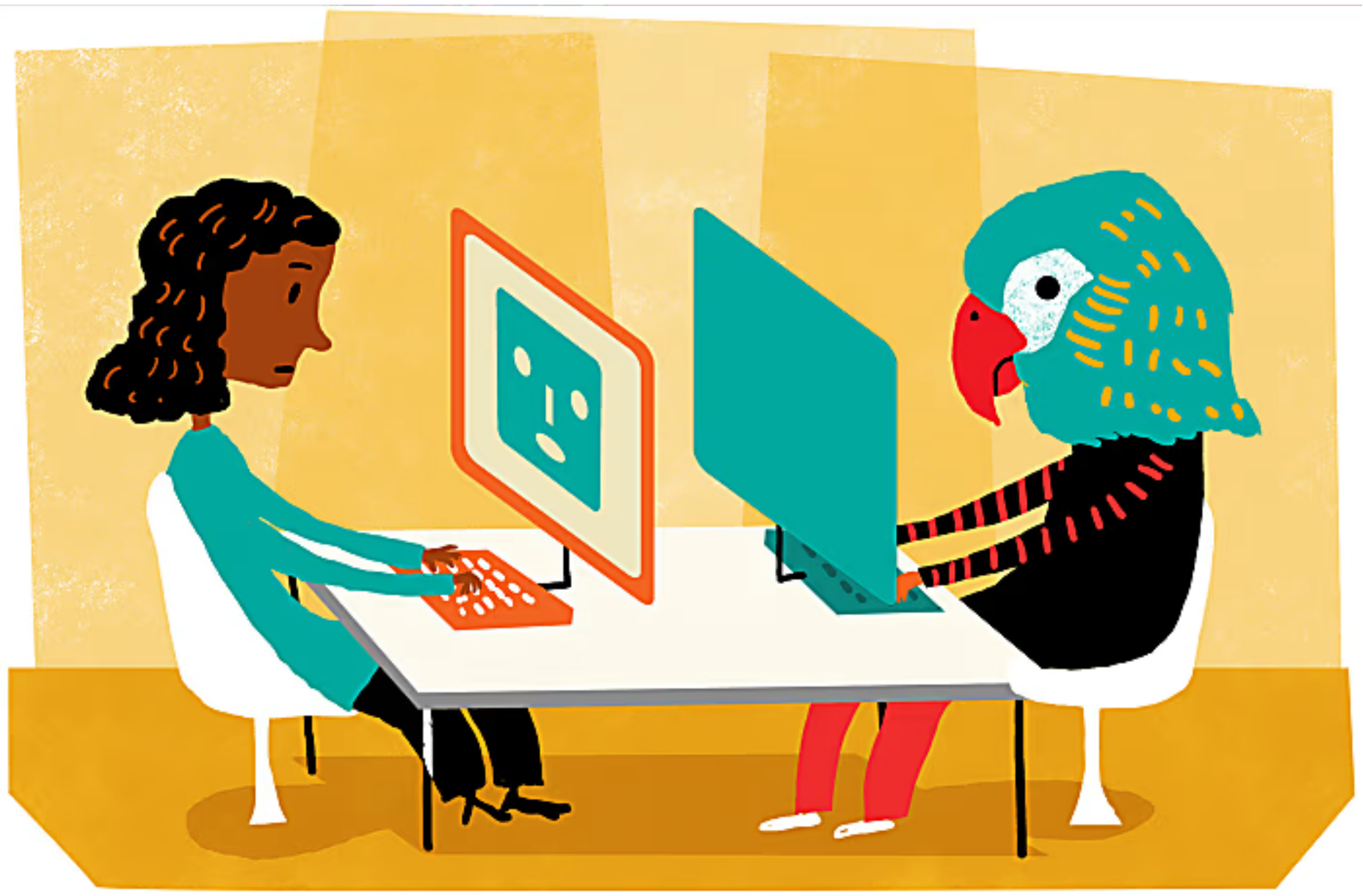
- Recall, a language model is a **generative** model that can predict entire sequences of words, one-word-at-a-time, via next word prediction.
- *Maybe*, a good chunk of human intelligence **can** be posed as (a sequence of) next word prediction problems:
  - *"The first president of the United States was [___]."*
  - *"If I compose two hydrogen atoms with one oxygen atom, I get [___]."*
- ***So, is language modeling merely a (self-)supervised learning task, or is it a task of (crudely) modeling human intelligence?***

ALIGNMENT TO HUMAN PREFERENCES (RLHF)

- It turns out that, when you train a model on loads of random text from the Internet, it will learn to say all kinds of things, **good *and* bad**.
  - *It can be helpful, or it can be sarcastic.*
  - *It can be polite/nice, or it can be hurtful/harmful.*
  - *It can make things up when it doesn't "know" the answer.*
- An *ad hoc* strategy: fine-tune LLMs using **human preferences**.
- **To think about:** Which human values are we aligning to?

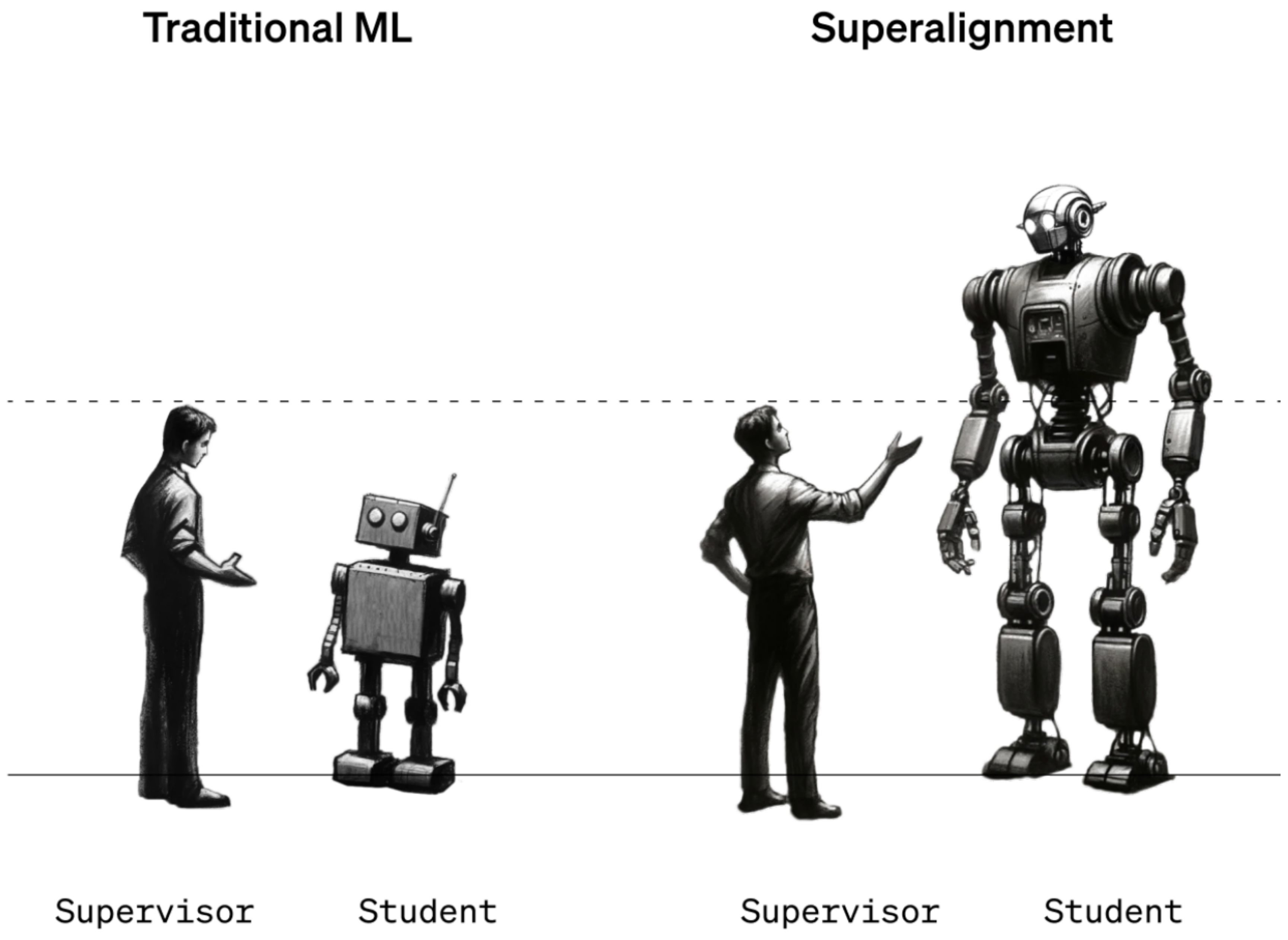


Some (very) different views of LLMs



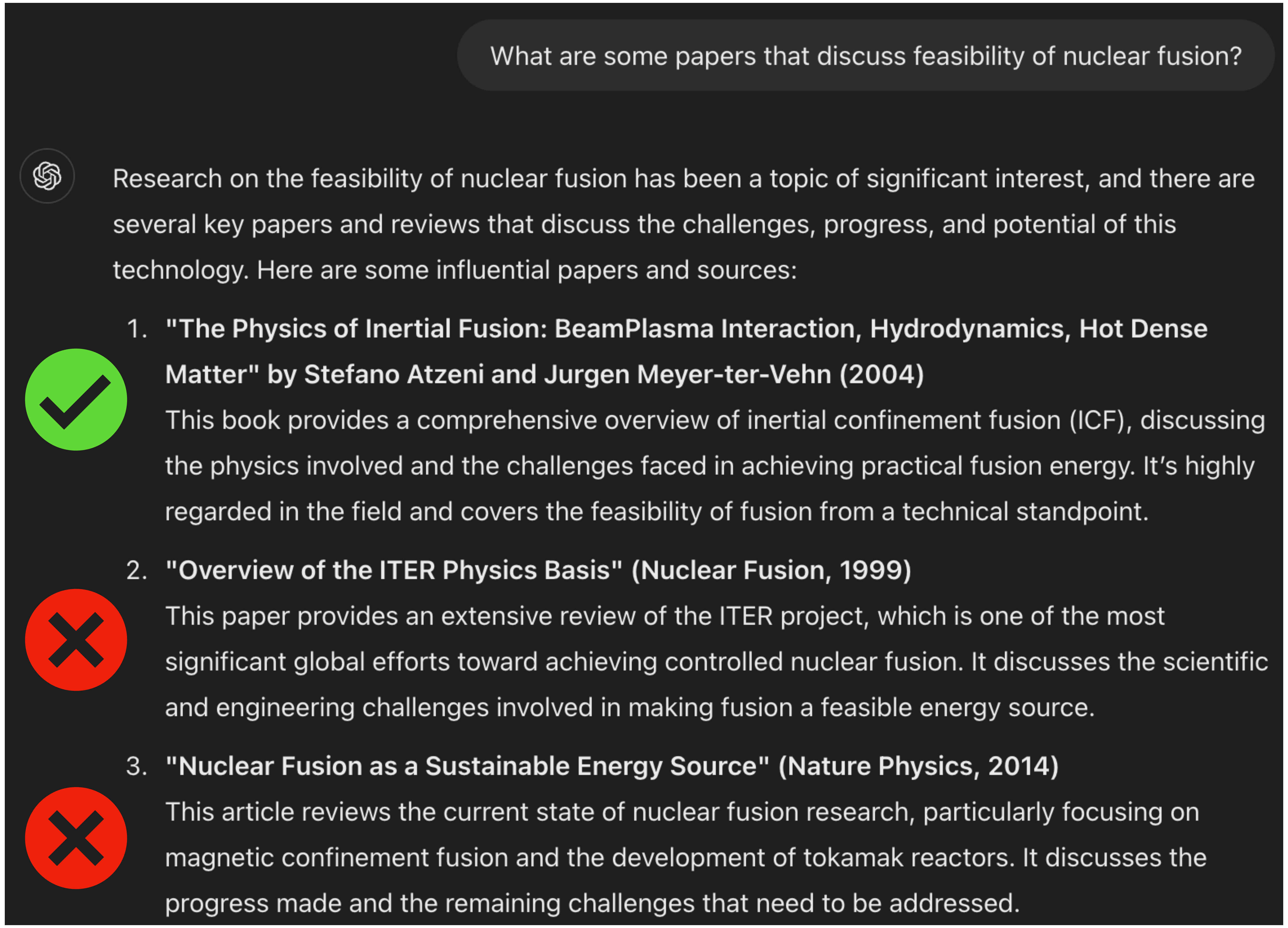
"Stochastic Parrots"
(Bender et al., 2021;
image from Wall Street Journal)

"Einstein in Your Basement"
(Henrik Kniberg, 2024;
recreated image from DALL-E)



"Superintelligence"
(OpenAI, 2023)

HALLUCINATION

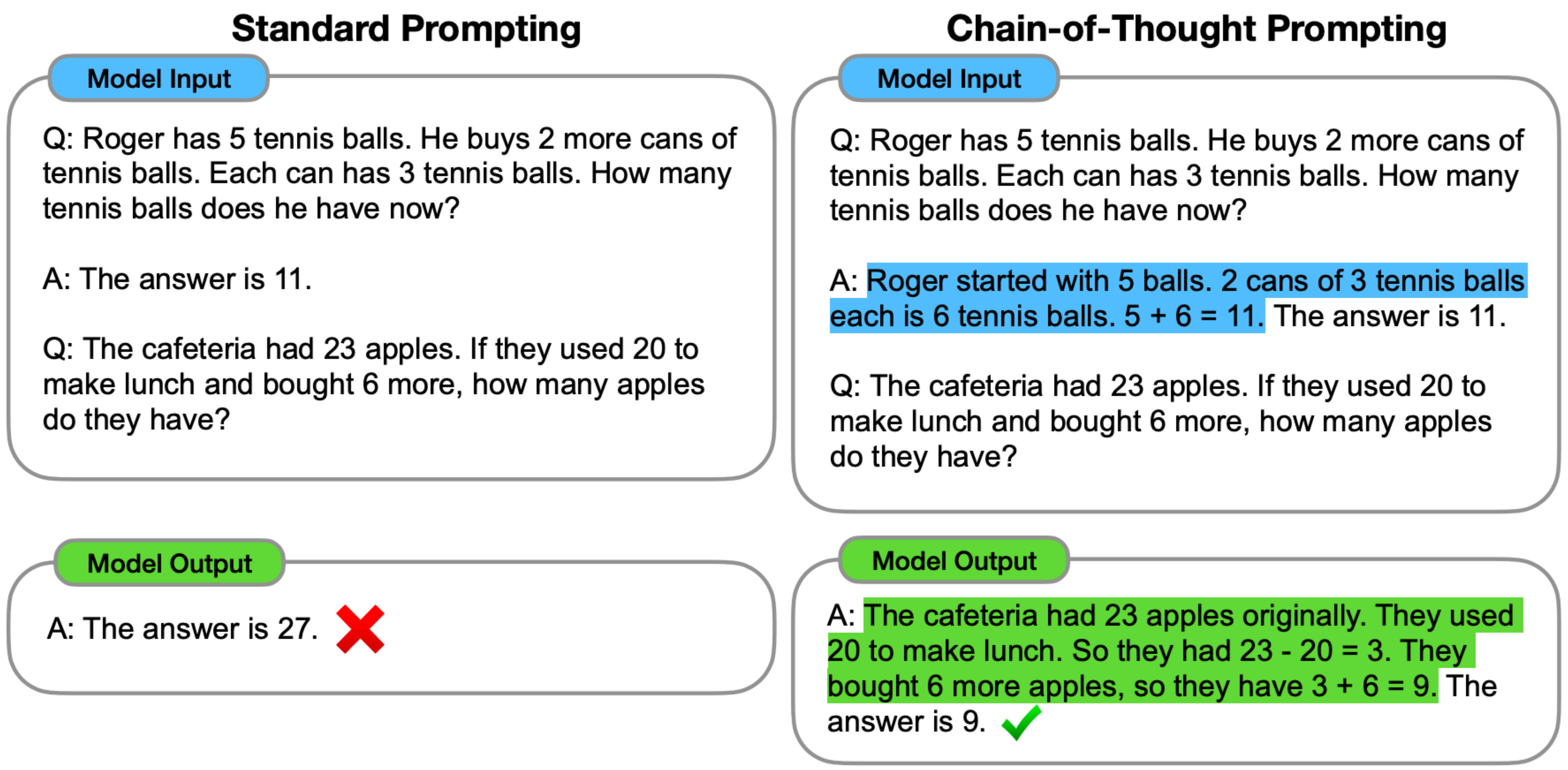


ChatGPT-4o, 09/2024

- Generally speaking, neural networks are very good **pattern matchers**.
  - *An undertrained LLM may know how to put together what* looks like *a citation.*
  - *But then, it may not know what name to put as the next token, so it just predicts the token with the highest probability.*
- **Retrieval-Augmented Generation (RAG)** is the go-to remedy, although it's clearly not perfect.

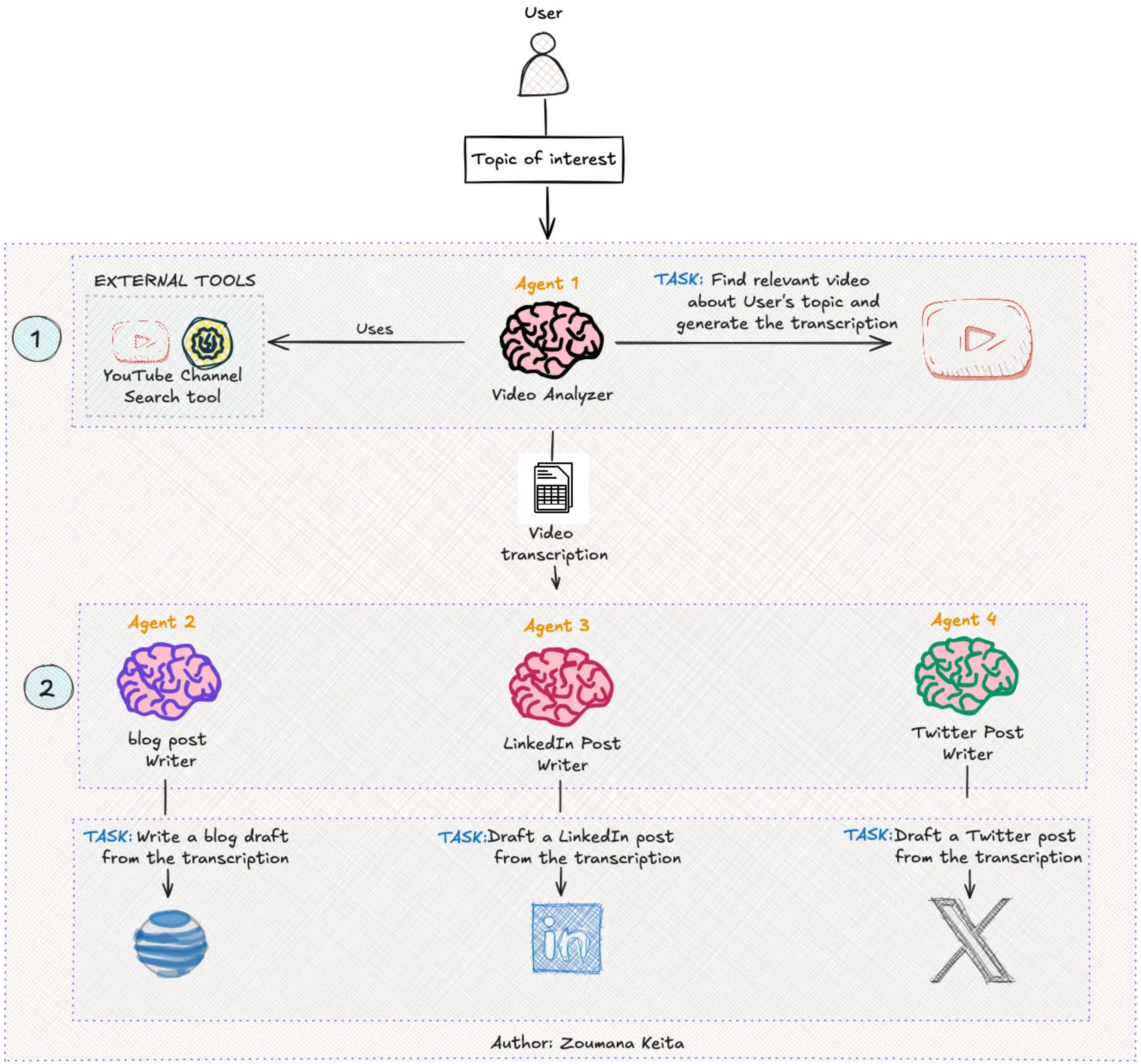
From Pattern Matchers to Reasoners?

- LLMs generally require lots of examples to learn.
- Yet, humans can often learn from only few examples; this is arguably because we can **reason**.
- **Chain-of-Thought (CoT) prompting** is an *ad hoc* technique that leads the model toward a correct reasoning path.



Wei et al. (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models."

"AGENTIC" AI



AUTONOMOUS AI AGENTS INTERACTION FROM
ANALYZING USER'S REQUEST TO CREATING SOCIAL MEDIA CONTENT

https://towardsdatascience.com/ai-agents-from-concepts-to-practical-implementation-in-python-fb26789b1560

What's Next?

- Many, many AI applications: autonomous driving, protein/drug modeling, medical AI, personalized assistants/companions, …
- Multi-modal learning: videos & sensory data; "World Models"
- Explainability & interpretability
- Long-term decision making
- "AGI" & Superalignment
- Unlimited potential and/or risk?

Takeaways

- **Machine Learning** finds generalizable patterns from data with trial & error.
- **Deep Learning** finds useful representations from complex data sources.
- **Large Language Models** (& other "GenAI" models) leverage massive DL models & computation, allowing them to learn from massive amounts of text/image/video/etc.

To think about:

- **(General) Intelligence**
- **Understanding** (what does it mean to “know” something?)
- **Alignment** (whose values are we aligning to?)
- **Agency/Responsibility/Morality/Honesty/…**
- …

Thank You

Any questions?