How to reason about Machine Learning

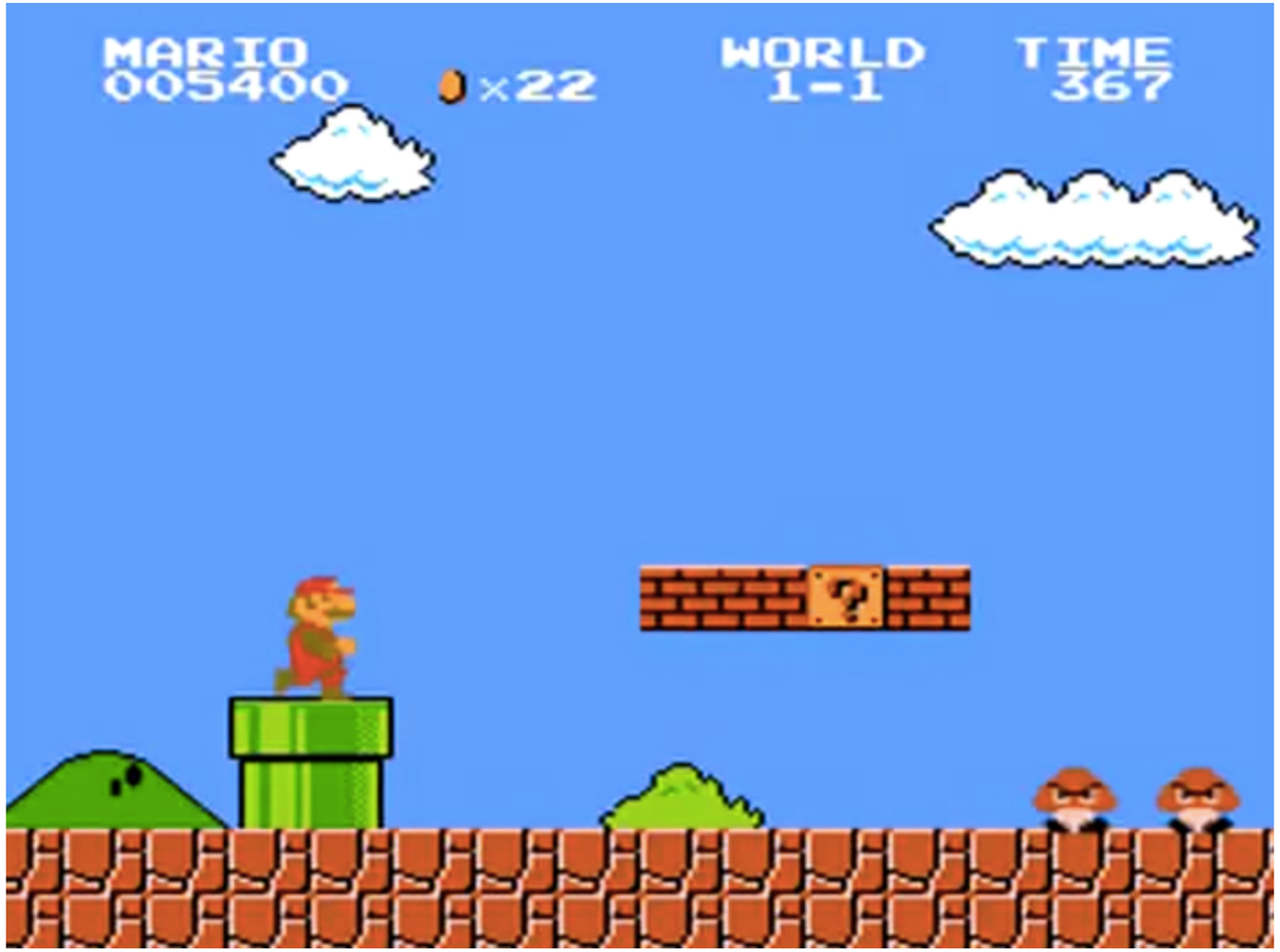
Overview and Introduction

Michael Ngo, April 19th 2025

What’s happening with Machine Learning today?

This section adapts and copies from the first lecture of Intro to ML taught by Professor Wen Sun.
See https://www.cs.cornell.edu/courses/cs4780/2023fa/

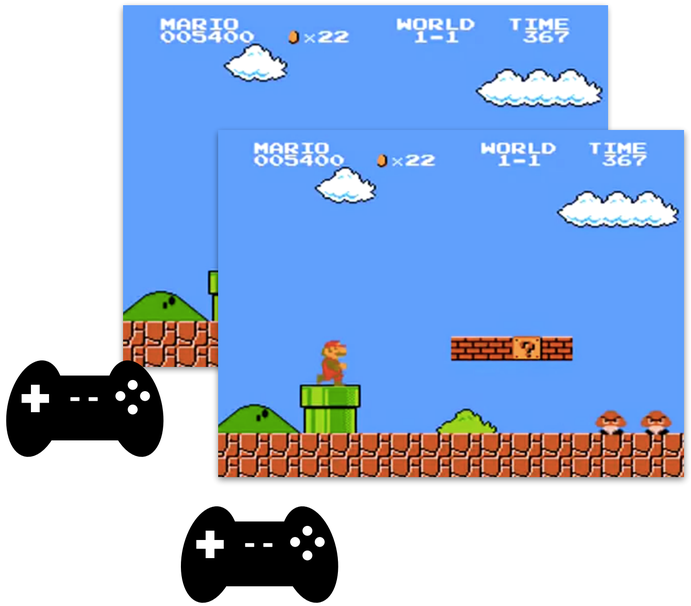
Super Mario Bros



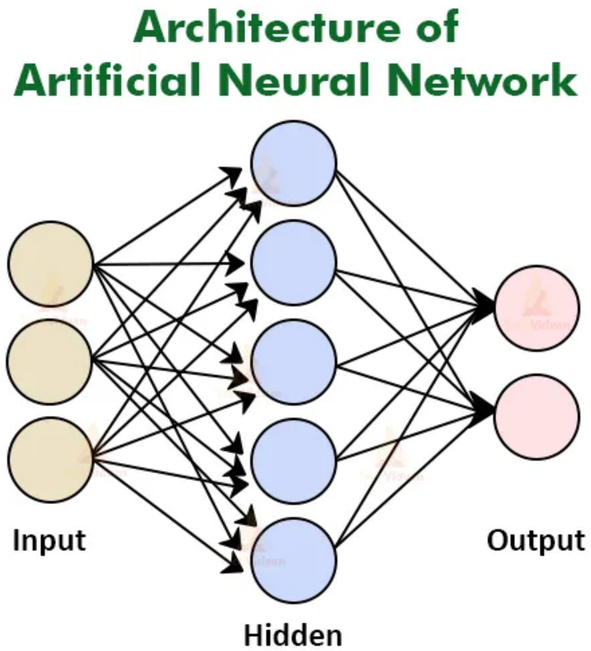
https://www.businessinsider.com/most-expensive-video-game-ever-sold-super-mario-bros-2019-3

Machine Learning makes this easy

Human Demonstrations



ML Algorithm



https://www.businessinsider.com/most-expensive-video-game-ever-sold-super-mario-bros-2019-3
https://blog.knoldus.com/architecture-of-artificial-neural-network/

A computer learned this!

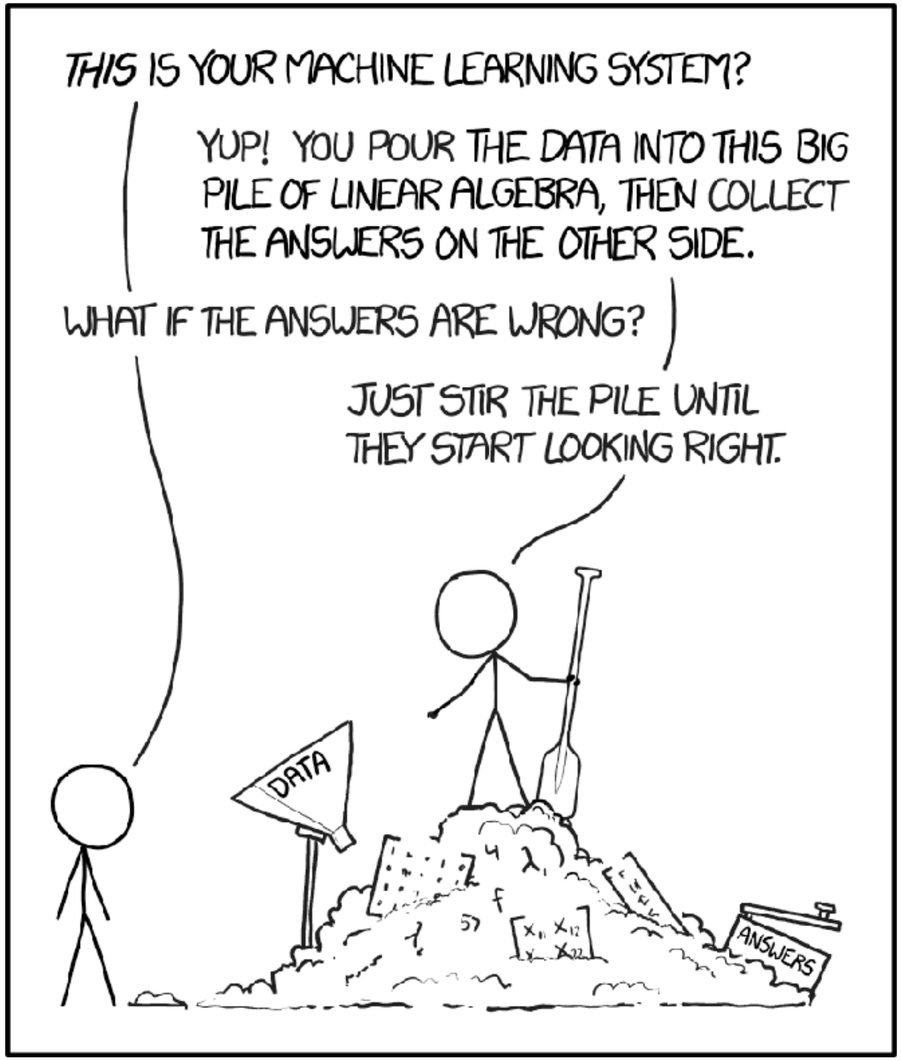
https://pytorch.org/tutorials/intermediate/mario_rl_tutorial.html

A computer learned this!

https://pytorch.org/tutorials/intermediate/mario_rl_tutorial.html

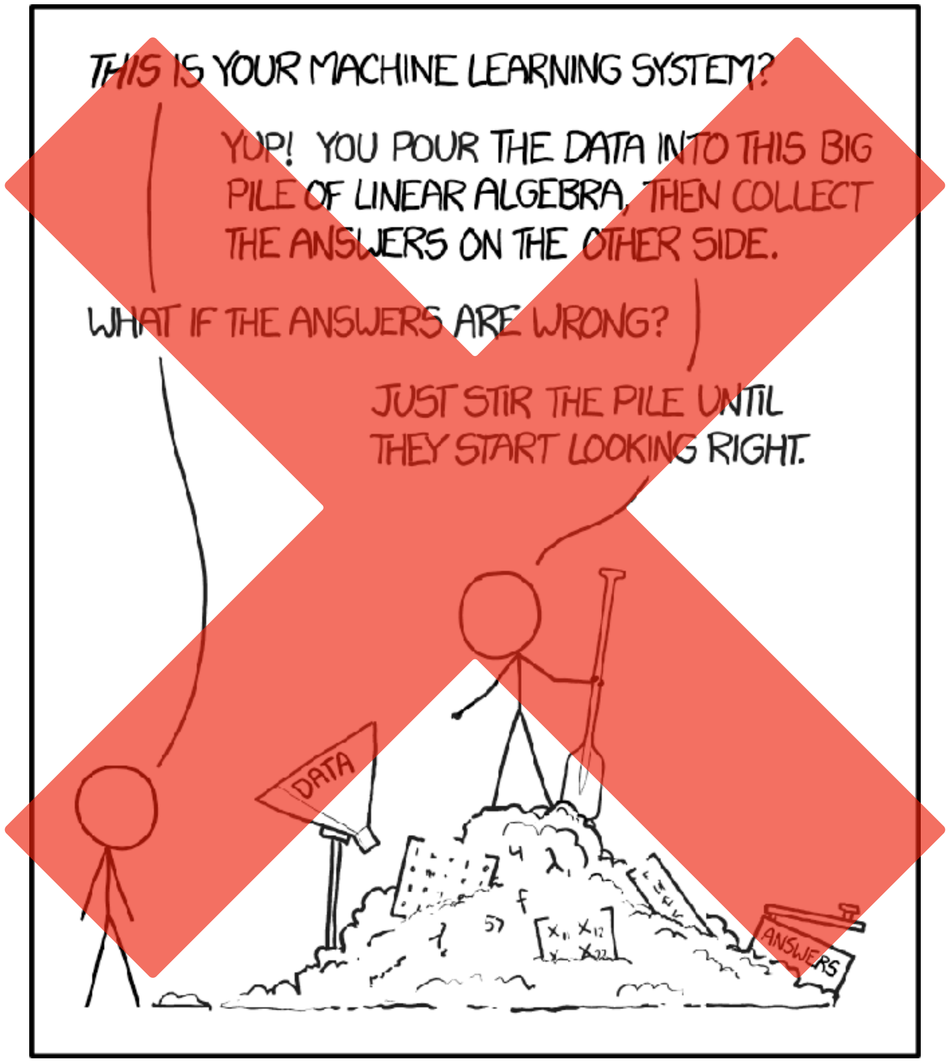
Machine learning means the computer learns from data without explicit programming.

Machine Learning



https://xkcd.com/1838/

Machine Learning



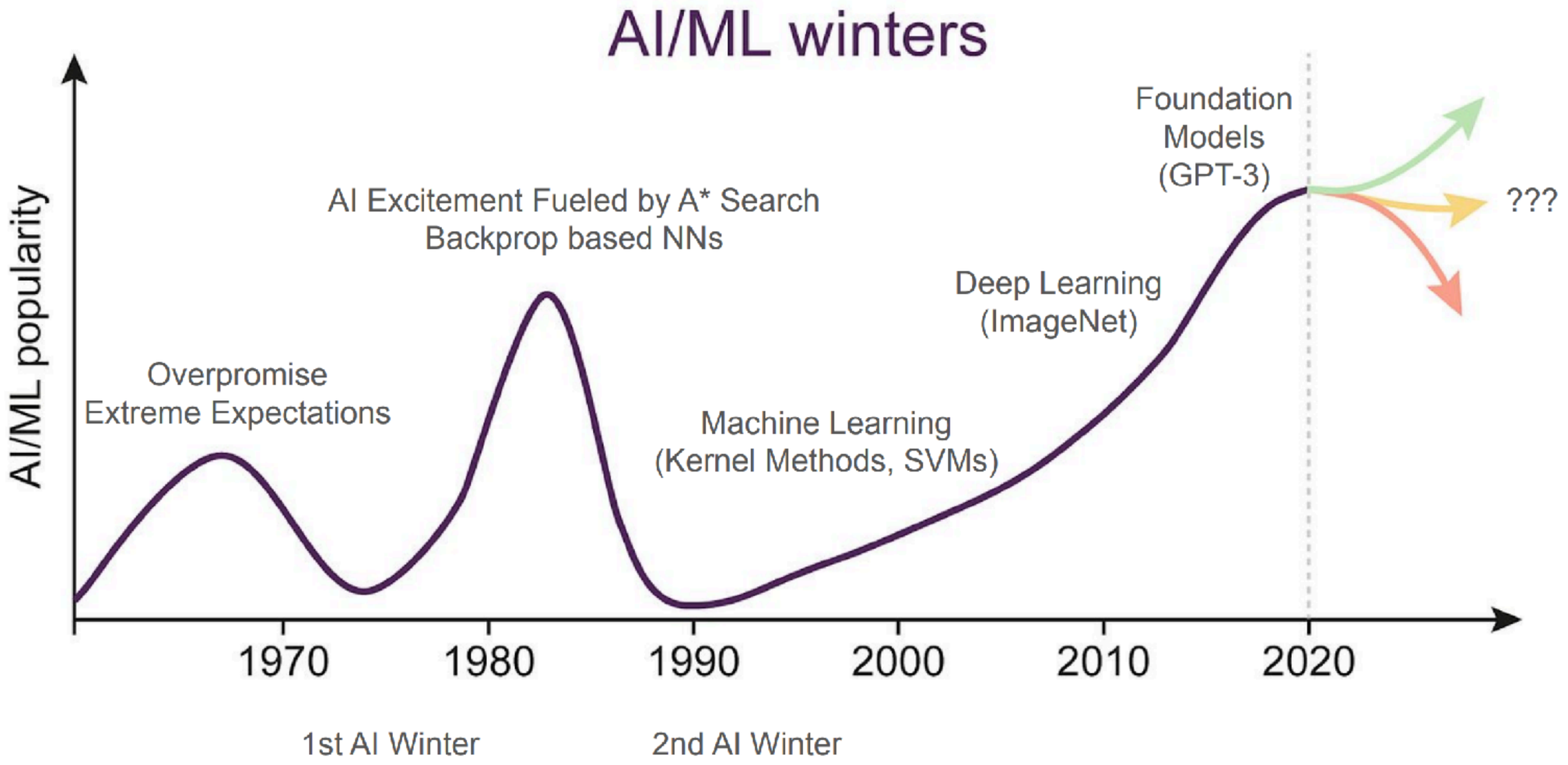
https://xkcd.com/1838/

Machine Learning

Why do we care about ML?
How is it done well?

https://xkcd.com/1838/

Public Perception of ML

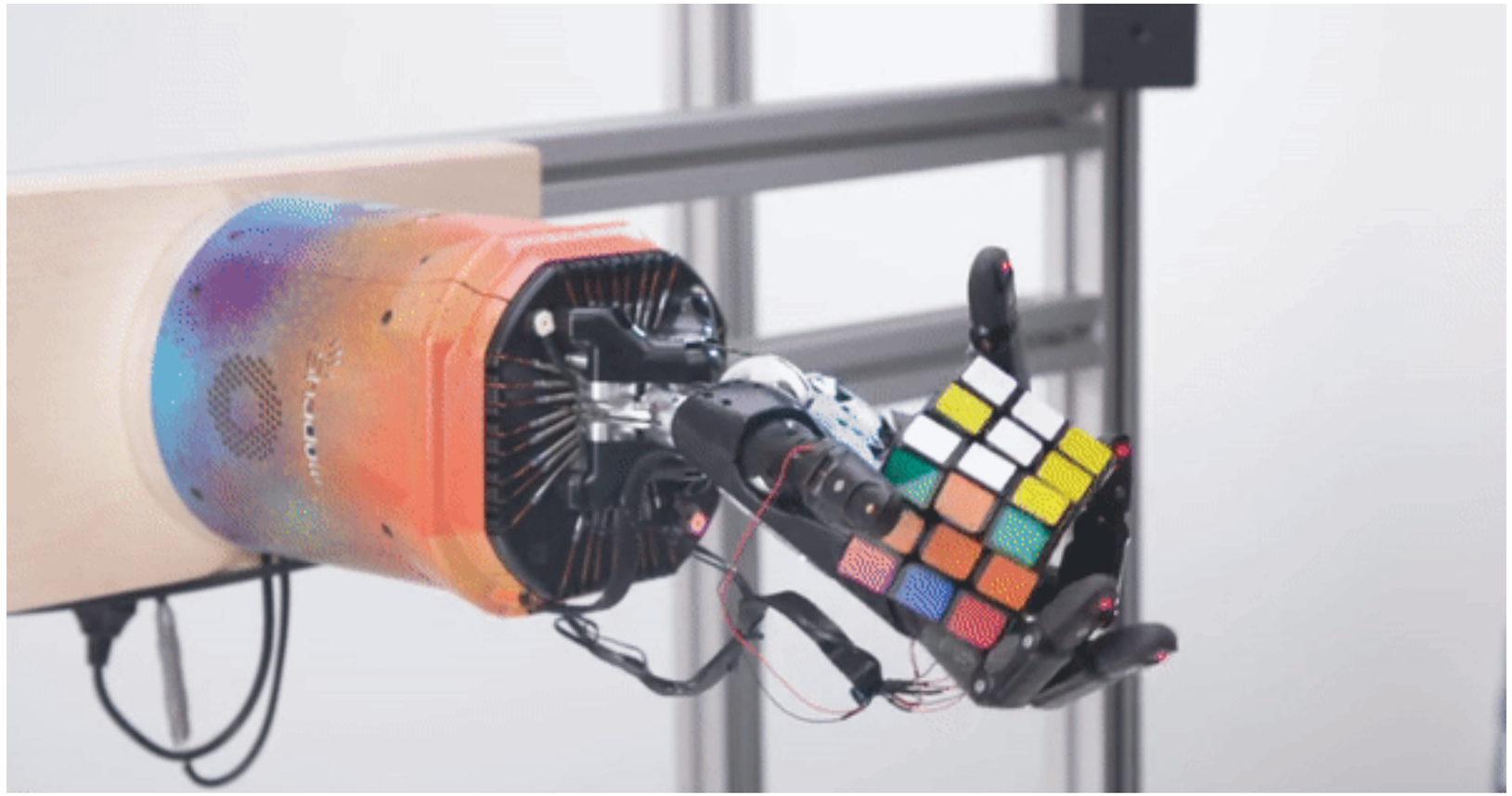


https://www.cs.cornell.edu/courses/cs4782/2025sp/slides/pdf/week1_1.pdf
https://medium.com/ersiliaio/beyond-the-hype-of-ai-ml-in-biomedicine-76c198c07467

AlphaGo Beats Lee Se-dol

https://www.bbc.com/news/technology-35785875

Robot Solves Rubik's Cube from Scratch



https://bgr.com/science/robot-rubiks-cube-openai-puzzle/

DALL-E Image Generation

Prompt

“Tiny potato kings wearing majestic crowns, sitting on thrones, overseeing their vast potato kingdom filled with potato subjects and potato castles.

https://openai.com/index/dall-e-3/

ChatGPT & Other Large Language Models

how to do inline citations in latex with names like: Goldwasser, Micali, et. al. (2020)

To format inline citations in LaTeX with author names like "Goldwasser, Micali, et al. (2020)", you should use the `natbib` package with an appropriate bibliography style.

Steps:

1. **Include the `natbib` package** in your preamble:

latex Copy Edit

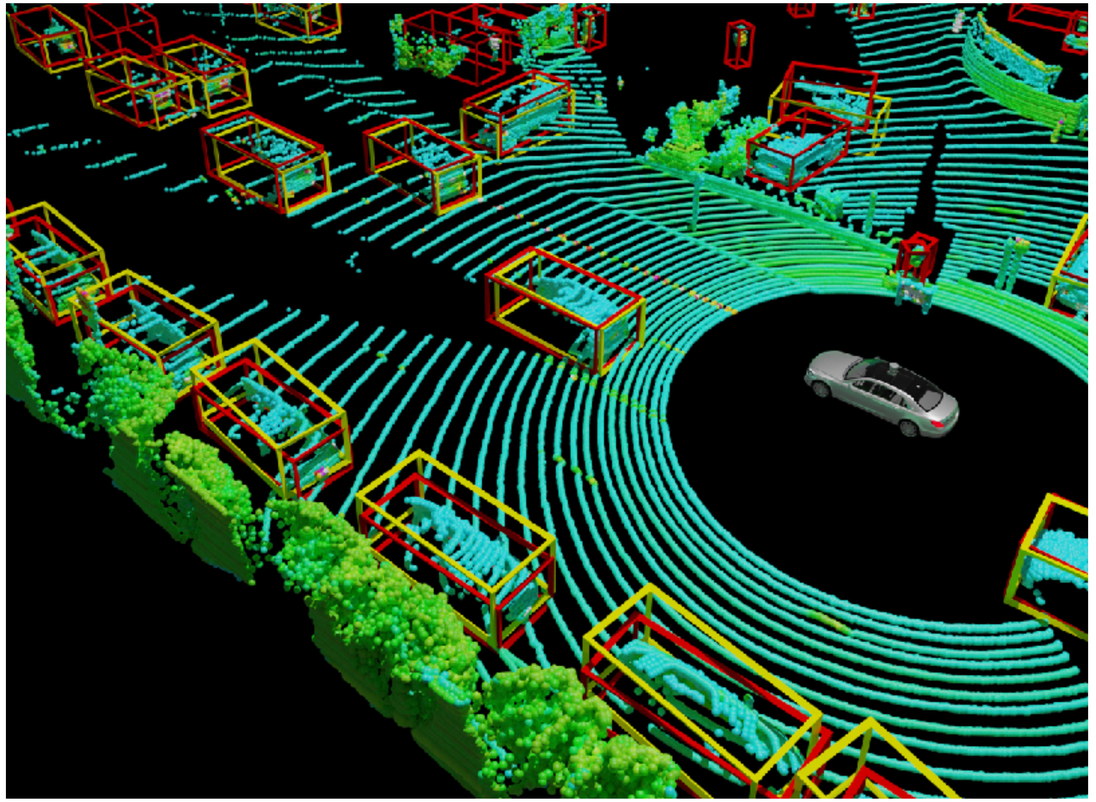
```
\usepackage[natbibapa]{apacite}  % Alternative: \usepackage{natbib
\bibliographystyle{apacite}      % Alternative: \bibliographystyle
```

2. **Use the citation command**:

https://chatgpt.com/share/67f1779a-f4a4-800e-99a7-9378c6f55057

Use in Critical Systems

- Wildfire & earthquake prediction
- Adaptive cancer treatment prediction
- Self-driving cars
- Detecting anomalies in large systems



https://arxiv.org/html/2403.07071v1

ML is powerful!

ML is powerful!

ML is dangerous!

Will AI/ML take over the world?

The ‘father of the internet’ and hundreds of tech experts worry we’ll rely on AI too much

By Clare Duffy, CNN

5 minute read · Published 8:00 AM EDT, Wed April 2, 2025

Do large language models understand the world?

In addition to its practical implications, recent work on “meaning representations” could shed light on some old philosophical questions.

By Matthew Trager, Stefano Soatto

February 15, 2024

Guest

AGI is coming faster than we think — we must get ready now

Gary Grossman, Edelman

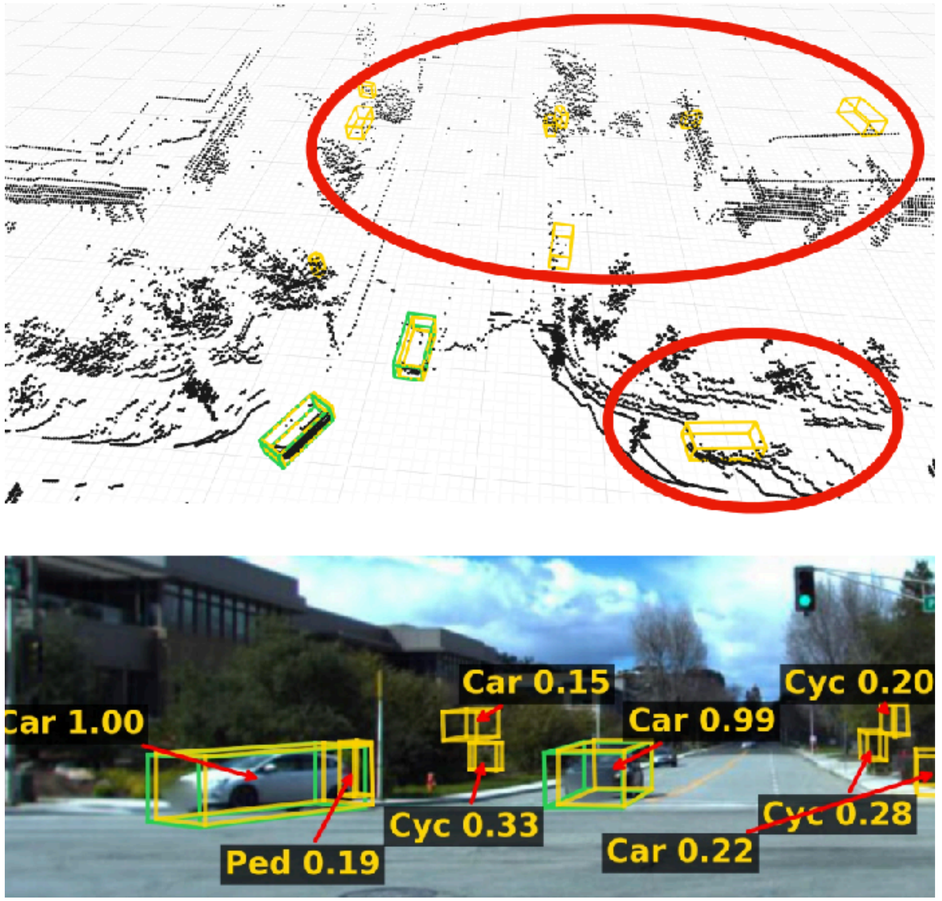
@garyg02

November 10, 2024 12:15 PM

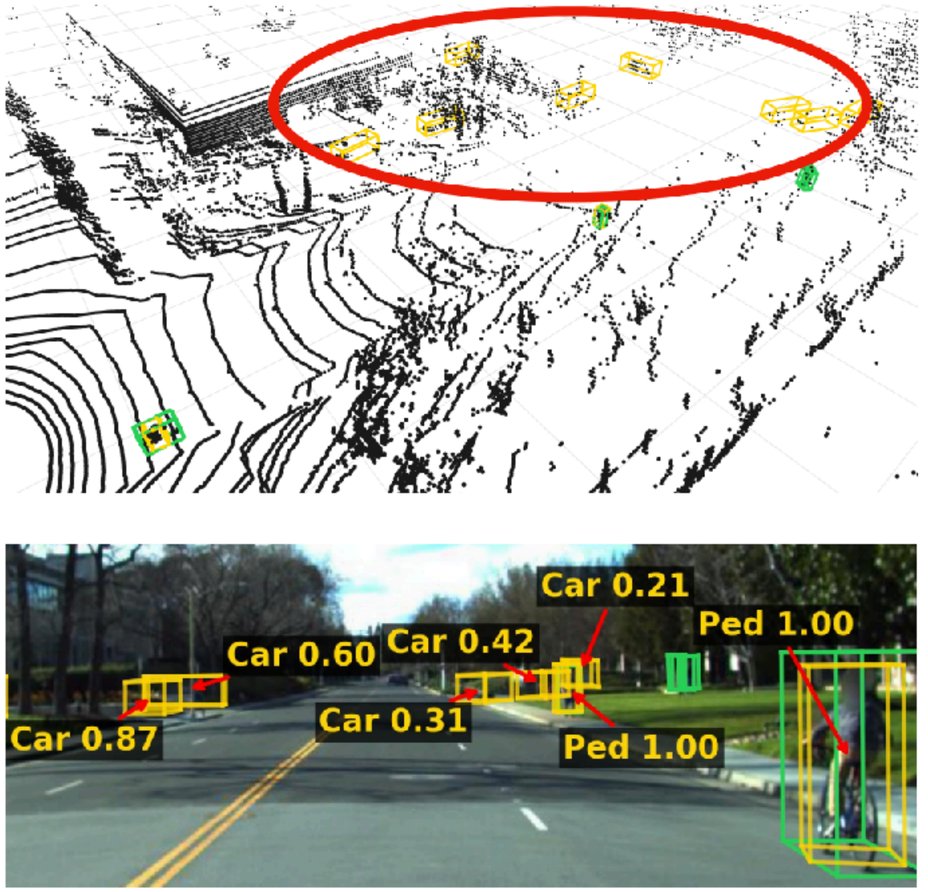
https://www.cnn.com/2025/04/02/tech/ai-future-of-humanity-2035-report/index.html
https://www.ibm.com/think/news/agi-right-goal
https://venturebeat.com/ai/agi-is-coming-faster-than-we-think-we-must-get-ready-now/

ML can make mistakes

Example Scene 1



Example Scene 2



https://arxiv.org/abs/1812.04244

ML confuses correlation for causation

ML algorithm thinks: ***brake light on*** => ***brake***

ML confuses correlation for causation

ML algorithm thinks: ***brake light on*** => ***brake***

Correlation ≠ Causation

This is called **overfitting.**

Even GPT-4o struggles at reasoning

Solid choice — you placed your **X** in position **8**. Here's the updated board:

markdown Copy code

```
 0 | X | 0
-----------
 0 | X | 6
-----------
 7 | X | X
```

I'll go with **6**:

markdown Copy code

```
 0 | X | 0
-----------
 0 | X | 0
-----------
 7 | X | X
```

Your turn! Only positions **7** are left.

https://chatgpt.com/c/67f159e6-0468-800e-9fba-e928d6d51524

Will AI/ML take over the world?

https://www.cnn.com/2025/04/02/tech/ai-future-of-humanity-2035-report/index.html
https://www.ibm.com/think/news/agi-right-goal
https://venturebeat.com/ai/agi-is-coming-faster-than-we-think-we-must-get-ready-now/

Will AI/ML take over the world?

Good news:
AI is nowhere close to general intelligence.

Bad news:
AI doesn't have to be smarter than us to be harmful.

https://www.cnn.com/2025/04/02/tech/ai-future-of-humanity-2035-report/index.html
https://www.ibm.com/think/news/agi-right-goal
https://venturebeat.com/ai/agi-is-coming-faster-than-we-think-we-must-get-ready-now/

Bias and fairness issues in ML

https://www.cnn.com/2024/10/04/politics/video/ai-elections-jake-tapper-lead-digvid

ML-based AI can generate fake information

Can you tell the difference? Jake Tapper uses his own deepfake to show how powerful AI is

The Lead

How AI deepfakes polluted elections in 2024

DECEMBER 21, 2024 · 5:00 AM ET

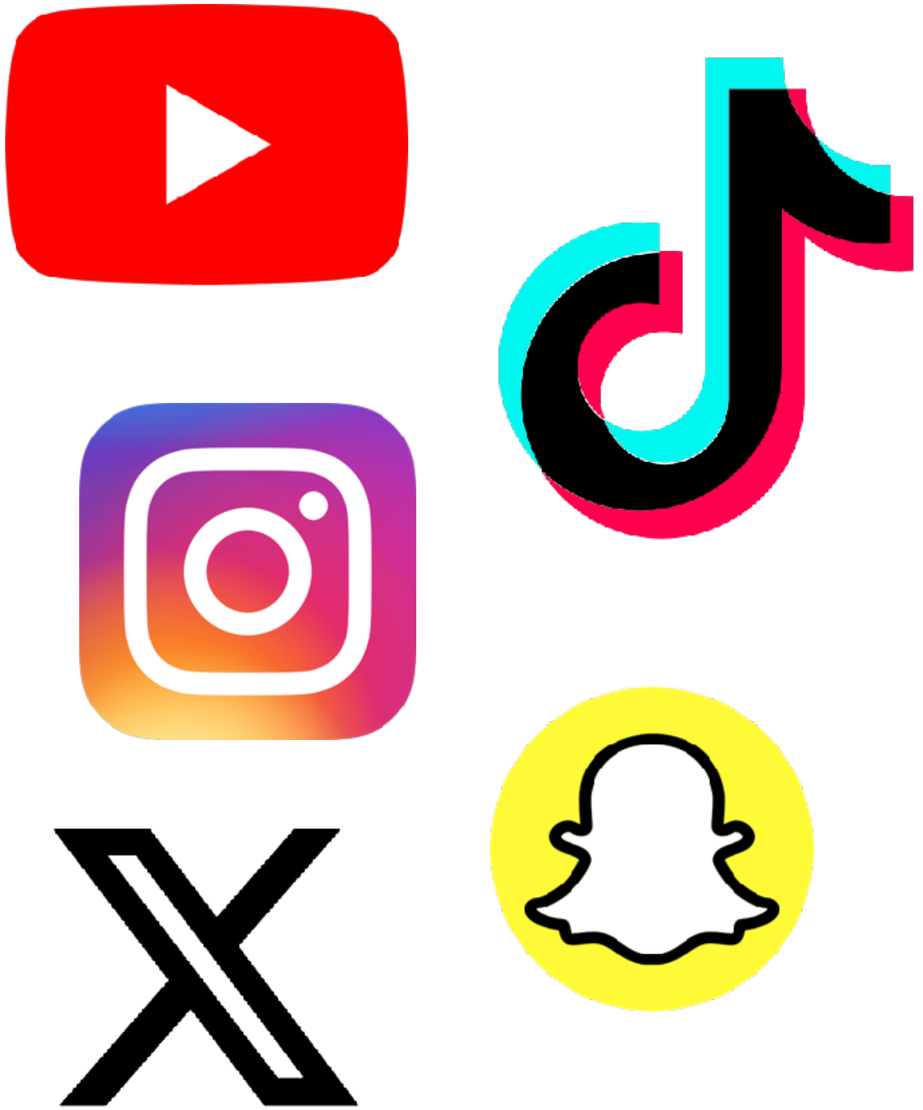
HEARD ON ALL THINGS CONSIDERED

Shannon Bond

https://www.cnn.com/2024/10/04/politics/video/ai-elections-jake-tapper-lead-digvid
https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

ML-powered advertising

- Goal: Learns from your habits to induce a change to stay on app longer and click more ads.
- Doesn’t promote content that we want. Promotes content that we will engage with.
- Effect: wasting time, plays up fear and anxiety and mis/disinformation.



https://www.cnn.com/2024/10/04/politics/video/ai-elections-jake-tapper-lead-digvid
https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

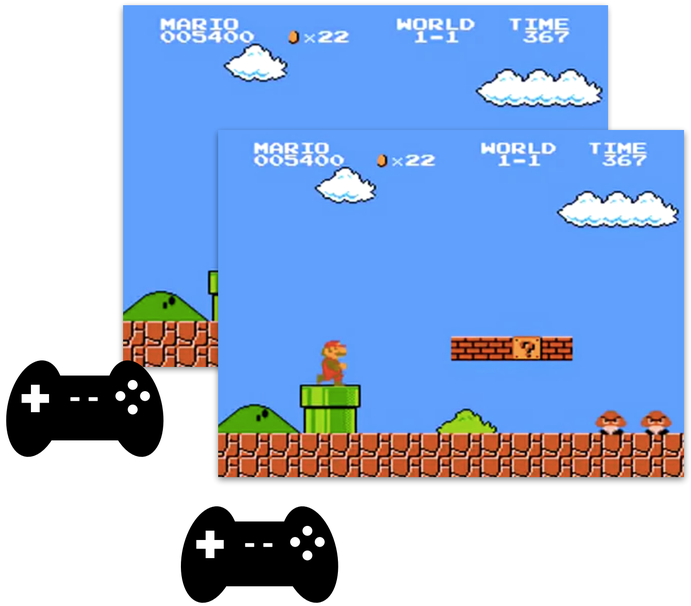
Machine learning is learning from data without explicit programming.

It’s powerful, and it can be harmful. We need to be mindful where and how to us this technology!

How to reason about machine learning

The goal of machine learning

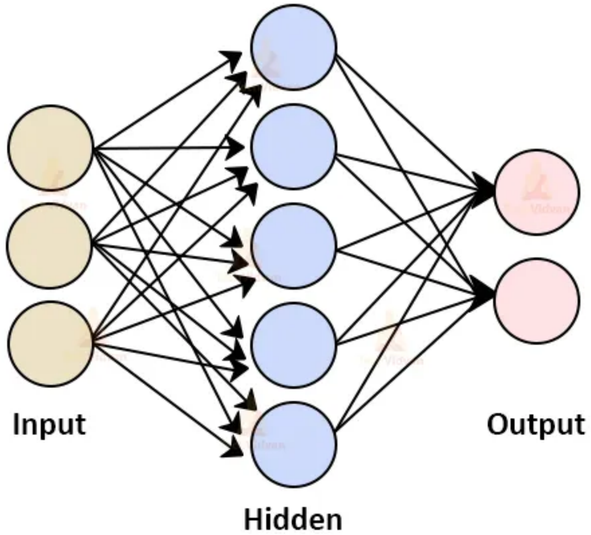
Training Data



Training Algorithm

ML Algorithm

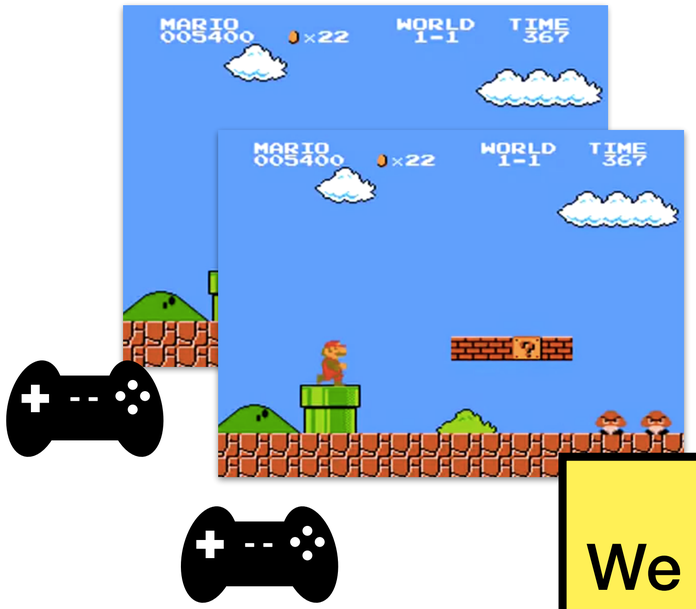
Model



https://www.businessinsider.com/most-expensive-video-game-ever-sold-super-mario-bros-2019-3
https://blog.knoldus.com/architecture-of-artificial-neural-network/

The goal of machine learning

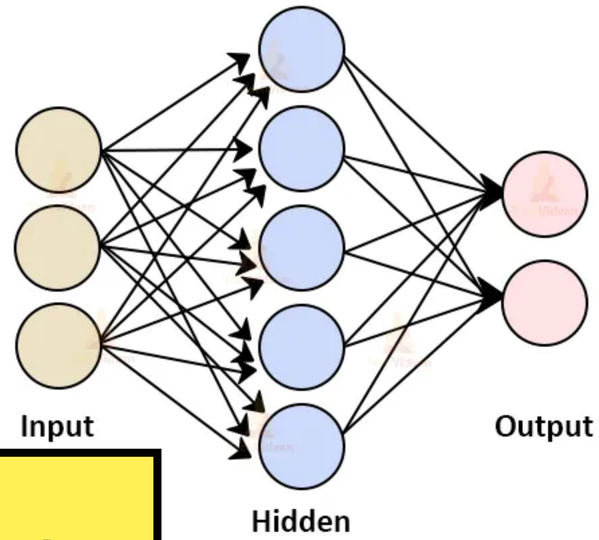
Training Data



Training Algorithm

ML Algorithm

Model



We want to minimize the **true error**, so what is one way to reason about it?

https://www.businessinsider.com/most-expensive-video-game-ever-sold-super-mario-bros-2019-3
https://blog.knoldus.com/architecture-of-artificial-neural-network/

Bias-variance decomposition

true error = bias + variance

Bias-variance decomposition

Error of model in the real world

$$\text{true error} = \text{bias} + \text{variance}$$

Bias-variance decomposition

Error of model in the real world

$$\text{true error} = \text{bias} + \text{variance}$$

Absolute best our ML algorithm can do with unlimited data and computation

Bias-variance decomposition

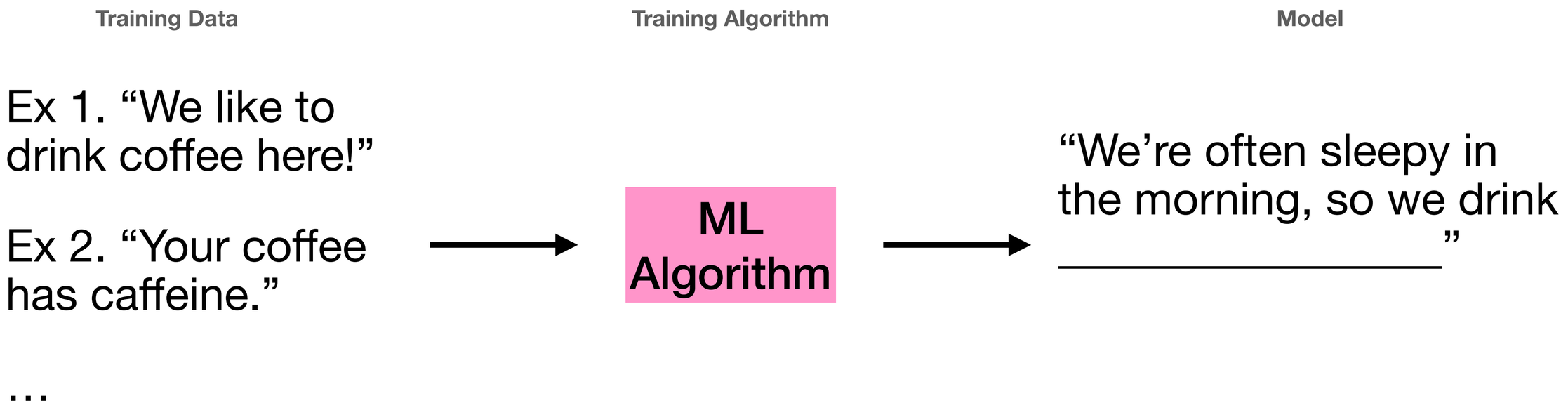
Error of model in the real world

$$\text{true error} = \text{bias} + \text{variance}$$

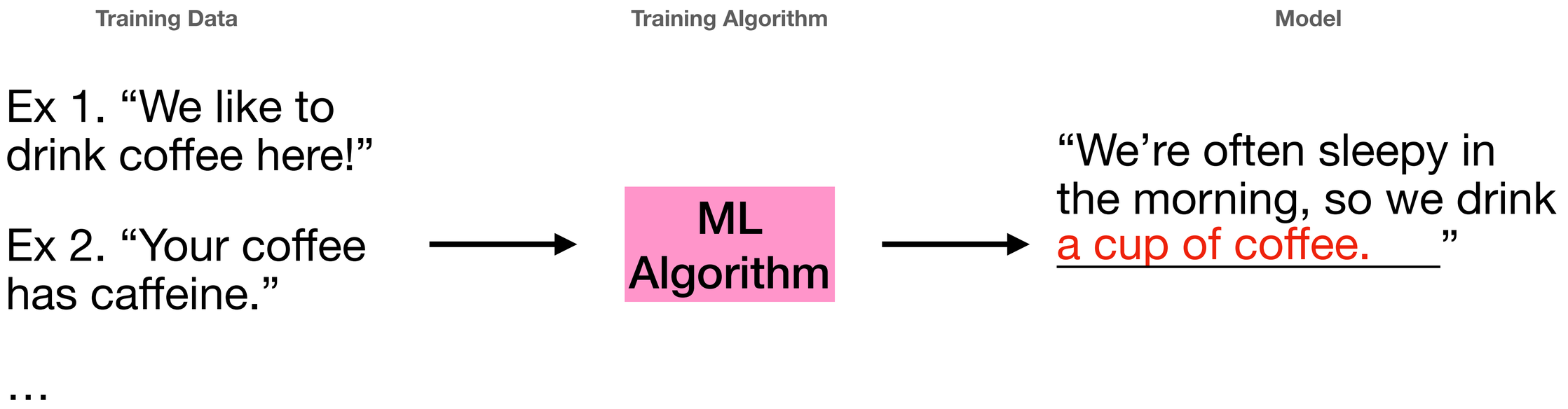
Absolute best our ML algorithm can do with unlimited data and computation

Variance from optimal model. Approximation error from not having enough data or compute power.

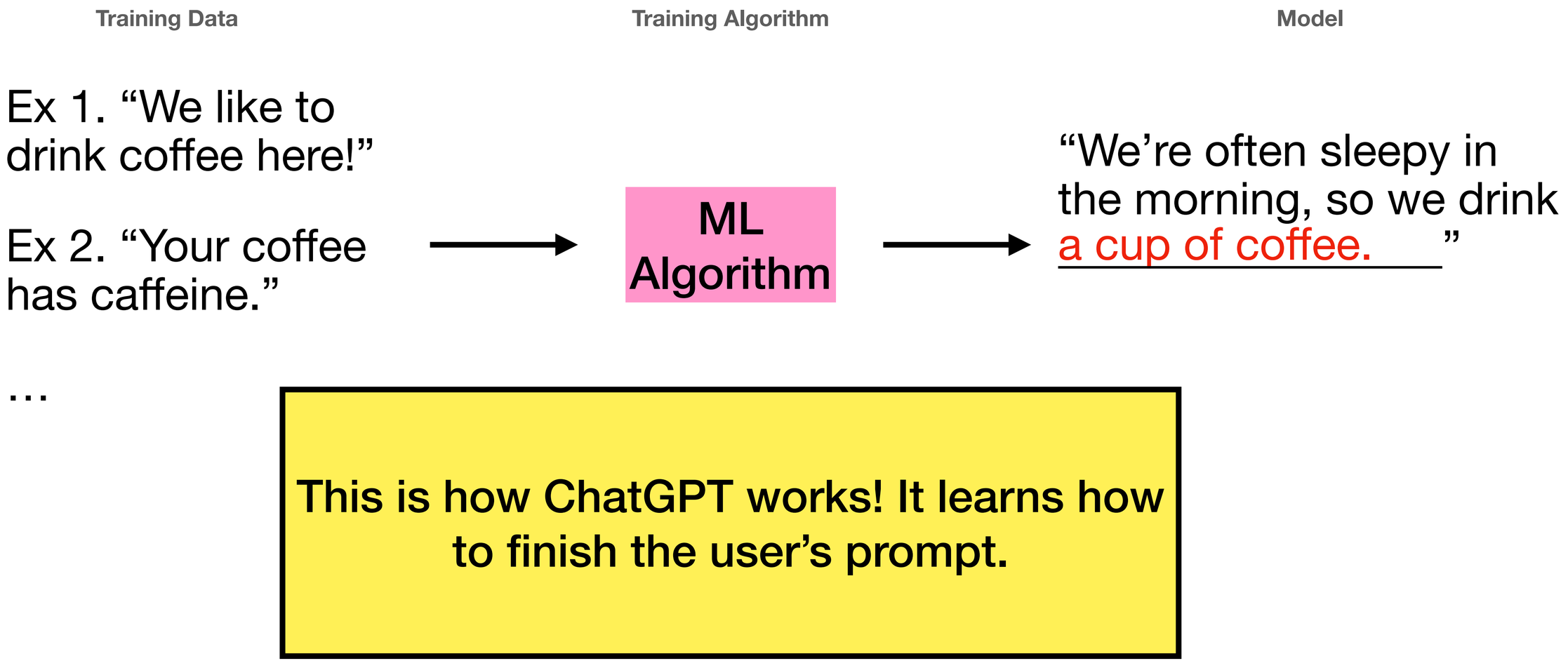
Learning to finish sentences



Learning to finish sentences



Learning to finish sentences



Machine learning on bi-grams

Ex 1. “We like to drink coffee here!” → (“We”, “like”), (“like”, “to”), (“to”, “drink”), (“drink”, “coffee”), (“coffee”, “here”), (“here”, “!”)

Ex 2. “Your coffee has caffeine.” → (“Your”, “coffee”), (“coffee”, “has”), (“has”, “caffeine”), (“caffeine”, “.”)

Model
converts each sentence into a bi-gram and use each bi-gram to predict the next word.

“We’re tired in the morning, so we ….”

Machine learning on bi-grams

Ex 1. “We like to drink coffee here!” → (“We”, “like”), (“like”, “to”), (“to”, “drink”), (“drink”, “coffee”), (“coffee”, “here”), (“here”, “!”)

Ex 2. “Your coffee has caffeine.” → (“Your”, “coffee”), (“coffee”, “has”), (“has”, “caffeine”), (“caffeine”, “.”)

Model
converts each sentence into a bi-gram and use each bi-gram to predict the next word.

“We’re tired in the morning, so we …. like to drink coffee has caffeine.”

Machine learning on bi-grams

Ex 1. “We like to drink coffee here!” → (“We”, “like”), (“like”, “to”), (“to”, “drink”), (“drink”, “coffee”), (“coffee”, “here”), (“here”, “!”)

Ex 2. “Your coffee has caffeine.” → (“Your”, “coffee”), (“coffee”, “has”), (“has”, “caffeine”), (“caffeine”, “.”)

Model
converts each sentence into a bi-gram and use each bi-gram to predict the next word.

“We’re tired in the morning, so we …. like to drink coffee has caffeine.”

Bi-grams have high bias

“We’re tired in the morning so... we drink coffee we work hard during the morning so we drink coffee we are always tired in the evening we work we're happy.”

$$\text{true error} = \text{bias} + \text{variance}$$

Bi-grams have high bias

“We’re tired in the morning so... we drink coffee we work hard during the morning so we drink coffee we are always tired in the evening we work we're happy.”

true error = bias + variance

Finishing a sentence one word at a time will almost never make sense.

Bi-grams have high bias

“We’re tired in the morning so... we drink coffee we work hard during the morning so we drink coffee we are always tired in the evening we work we're happy.”

true error = bias + variance

Finishing a sentence one word at a time will almost never make sense.

Over enough sentences, the actually most frequent of bigrams will always be used. So there’s not much variance from the optimal.

Machine learning by memorization

Ex 1. “We like to drink coffee here!

Ex 2. “Your coffee has caffeine.”

Model
Checks if the prompt matches any of the sentences in the training set, and finishes accordingly.

“We like to drink…. ”

“We’re tired in the morning, so we …. ”

Machine learning by memorization

Ex 1. “We like to drink coffee here!

Ex 2. “Your coffee has caffeine.”

Model
Checks if the prompt matches any of the sentences in the training set, and finishes accordingly.

“We like to drink…. coffee here! ”

“We’re tired in the morning, so we …. ”

Machine learning by memorization

Ex 1. “We like to drink coffee here!

Ex 2. “Your coffee has caffeine.”

Model
Checks if the prompt matches any of the sentences in the training set, and finishes accordingly.

“We like to drink…. coffee here! ”

“We’re tired in the morning, so we …. [NO OUTPUT] ”

Memorization has high variance

“Four score and seven years ago… our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.”

$$\text{true error} = \text{bias} + \text{variance}$$

Memorization has high variance

“Four score and seven years ago… our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.”

true error = bias + variance

With access to every possible conceivable sentence and the memory to store it, it will always finish correctly!

Memorization has high variance

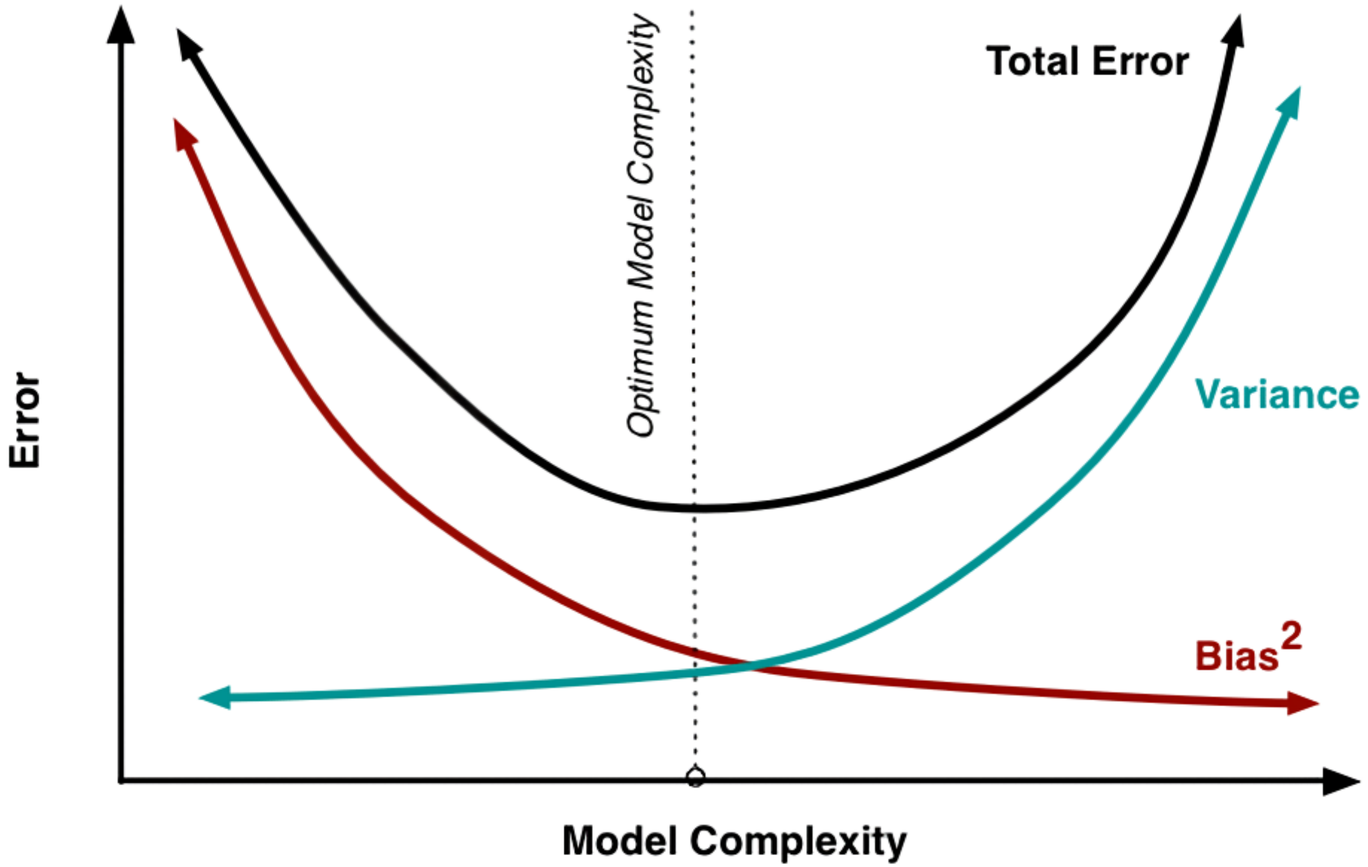
“Four score and seven years ago… our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.”

true error = bias + variance

With access to every possible conceivable sentence and the memory to store it, it will always finish correctly!

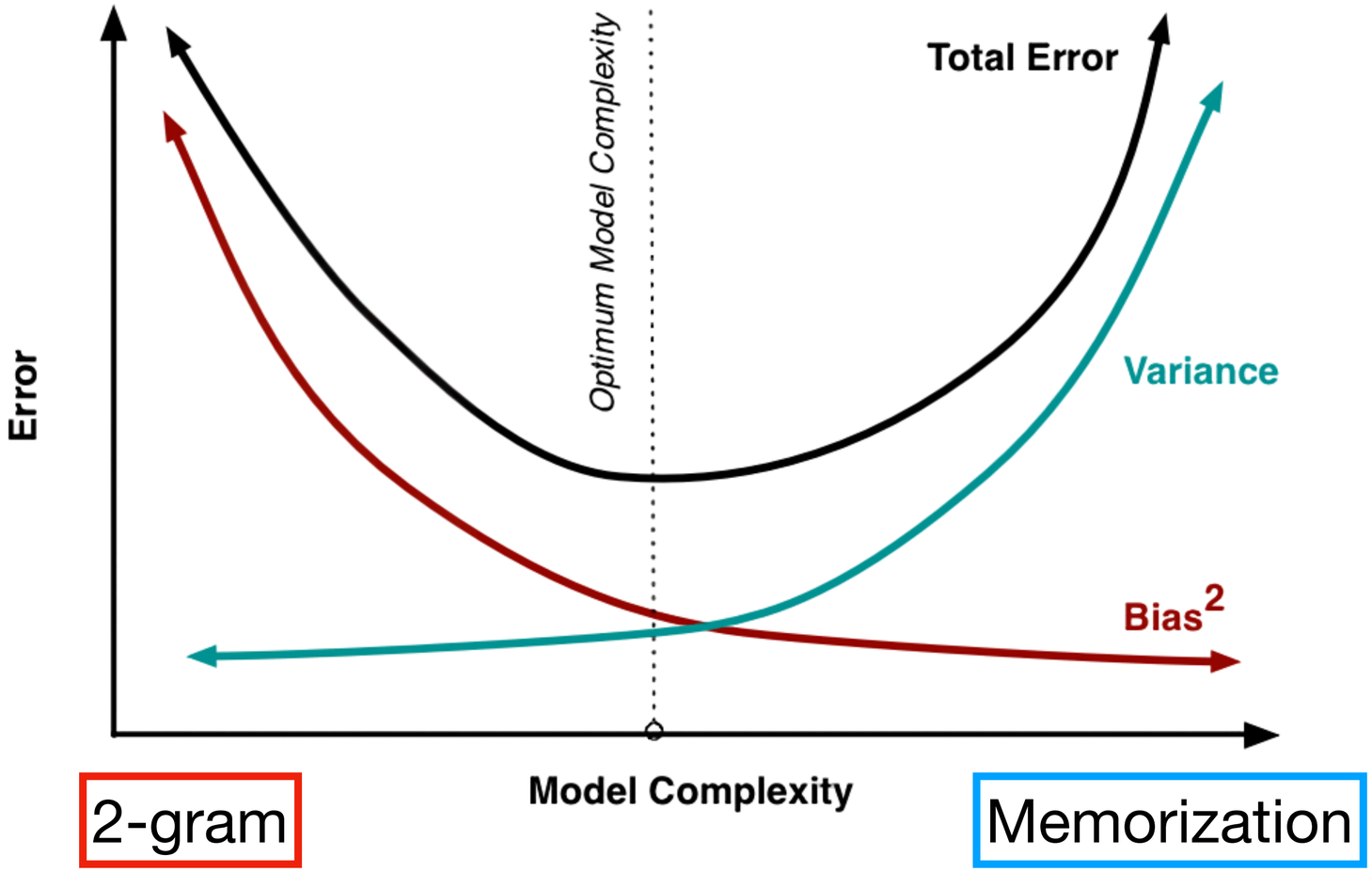
Can only finish sentences in dataset. With limited memory, it will not even be close to finishing every possible sentence.

Bias-variance tradeoff



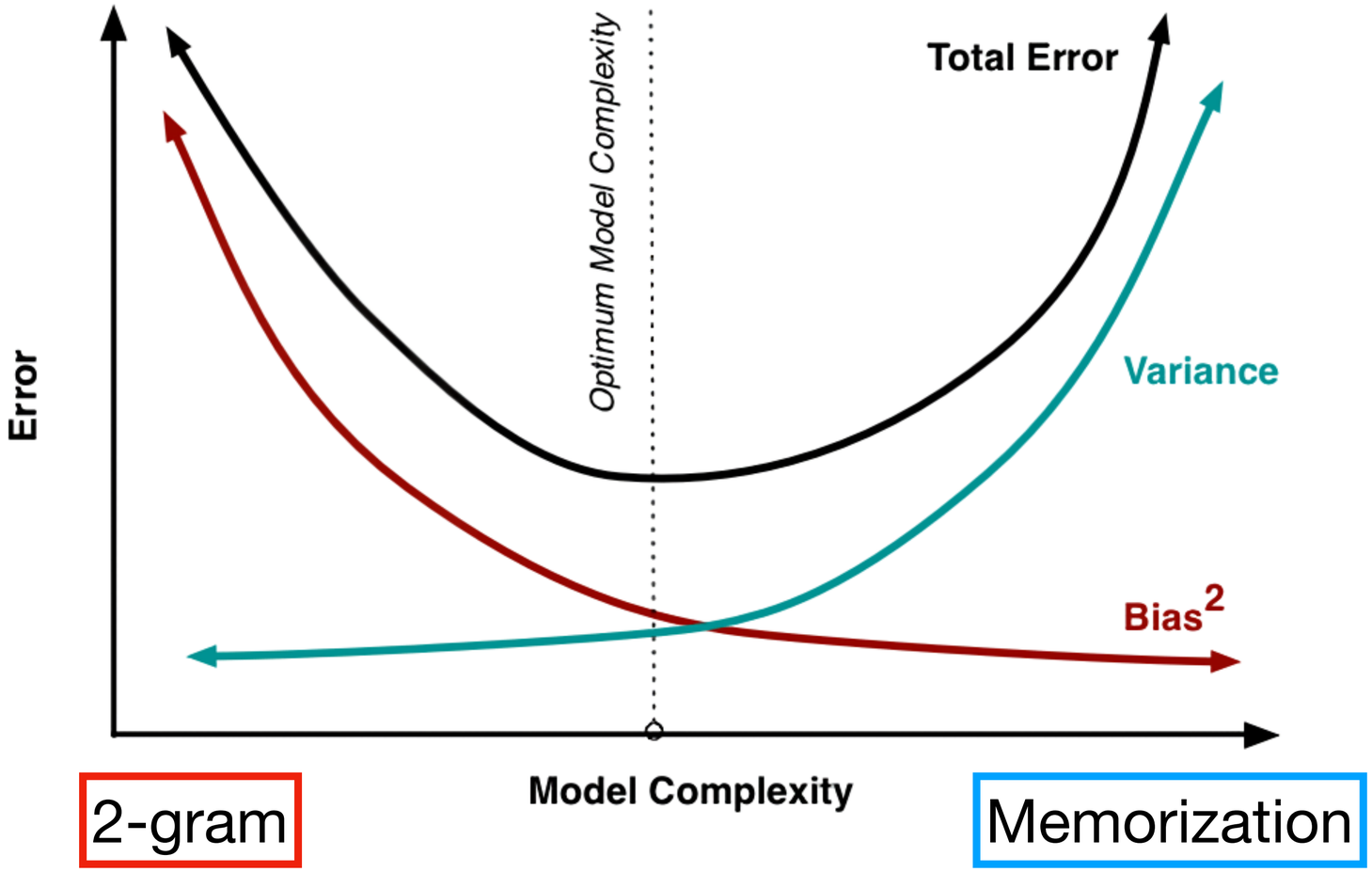
https://scott.fortmann-roe.com/docs/BiasVariance.html

Bias-variance tradeoff



https://scott.fortmann-roe.com/docs/BiasVariance.html

Bias-variance tradeoff



2-gram

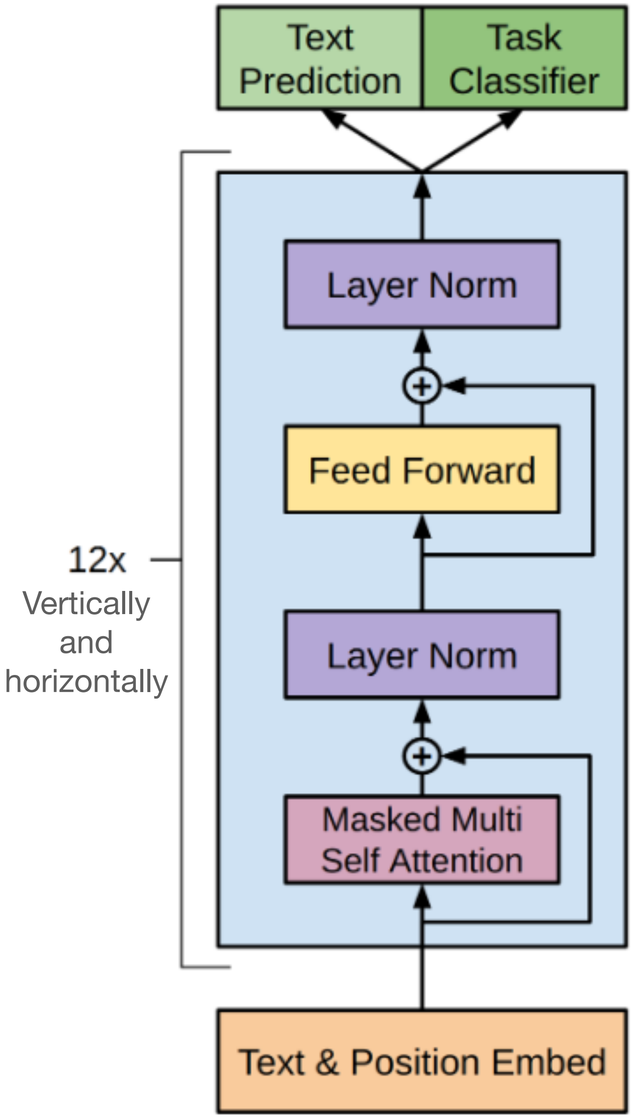
Memorization

ChatGPT

https://scott.fortmann-roe.com/docs/BiasVariance.html

How ChatGPT cheats

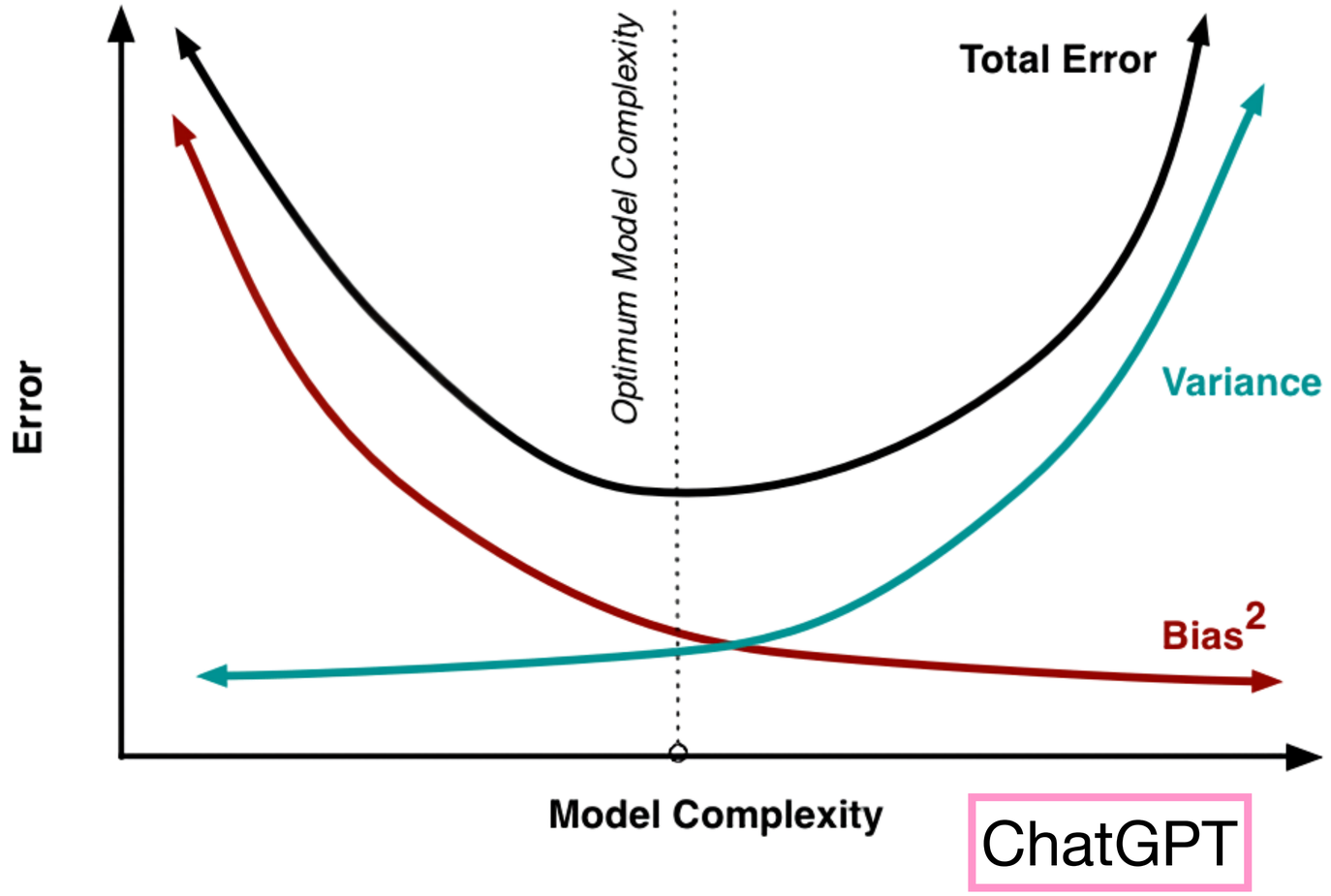
- GPT-4 has ~1 trillion parameters
- Push variance to the right via
  - Train on the a massive dataset (the internet)
  - Use a transformer-based architectures which allows for really good parallelization with GPUs.
  - Spend > $100 million



GPT-3 architecture

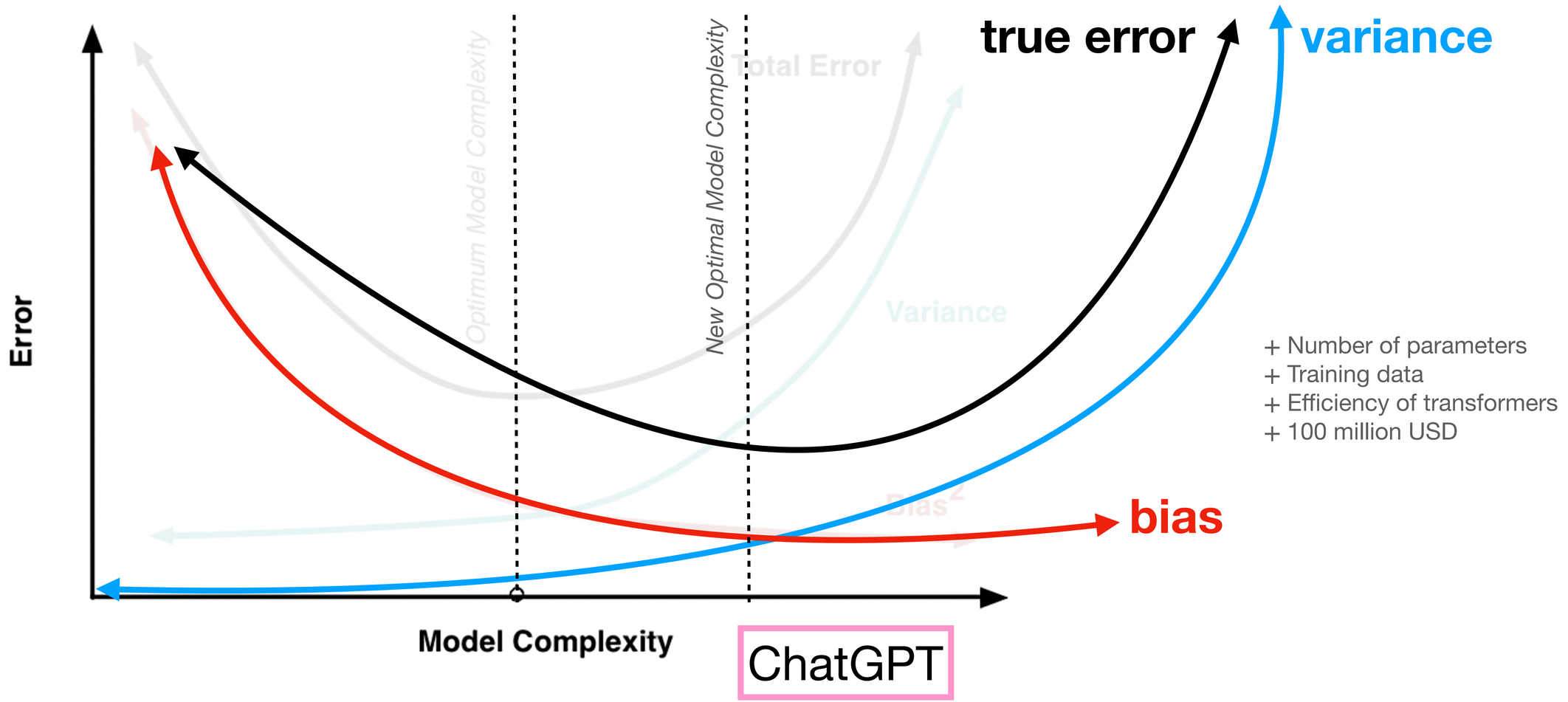
https://en.wikipedia.org/wiki/ChatGPT

Bias-variance tradeoff for GPT



https://scott.fortmann-roe.com/docs/BiasVariance.html

Bias-variance tradeoff for GPT



https://scott.fortmann-roe.com/docs/BiasVariance.html

Scaling laws

[J. Kaplan et al. (2020)]

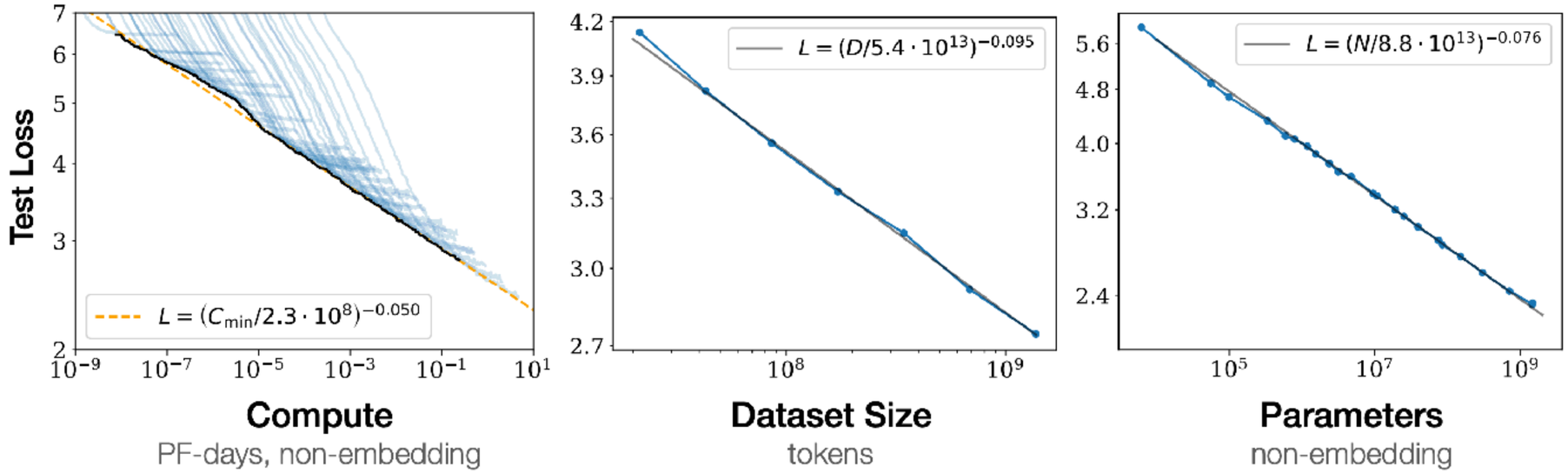


Figure 1 Language modeling performance improves smoothly as we increase the model size, datasetset size, and amount of compute[2] used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.

https://arxiv.org/abs/2001.08361

Summary

- ML is learning from data without explicit programming.
- ML-based AI has exploded in the last few years, especially generative AI for natural language tasks and image or video generation.
- Bias-variance decomposition gives a principled way to evaluate machine learning algorithms. Keep using it!