How to reason about Machine Learning

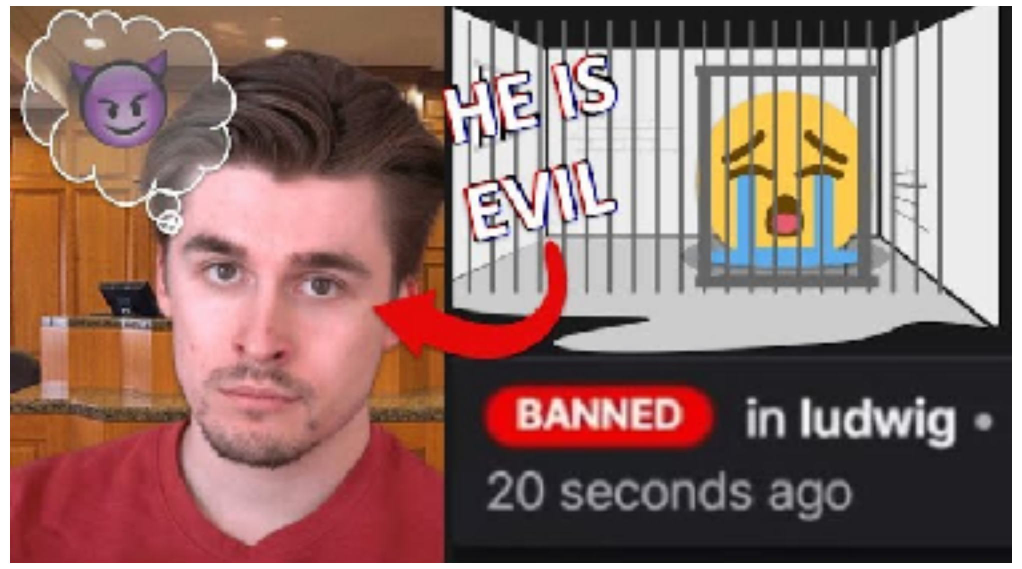
Overview and Introduction

Michael Ngo, April 19th 2025

Hi, I’m Michael Ngo

Pronounced “No”, he/him

- 3rd year undergraduate studying CS
- Research learning theory w/ prof Michael P. Kim
- Teaching assistant for Algorithms (CS 4820)
- Onboarding Chair for Cornell Data Science project team
- Spends way to much time on YouTube



After a really bad haircut

Agenda

1. What is machine learning?
2. Why is machine learning so hot right now?
3. How we can begin to reason about machine learning?
4. What do we have to look out for?

What is machine learning?

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

How to program a computer to complete this level?

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

Super Mario Bros

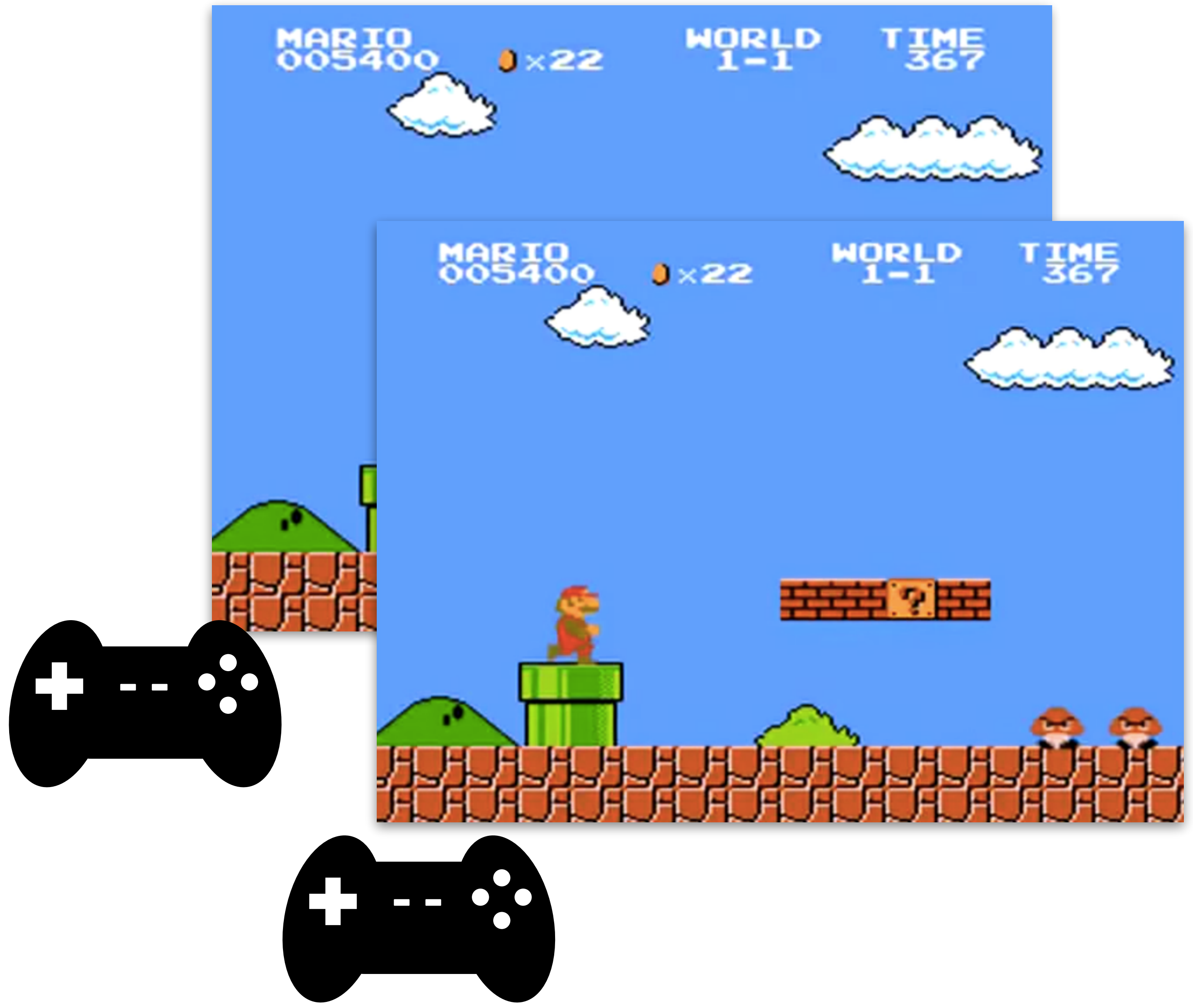
How to program a computer to complete this level?

Naive:
"Move forward,
jump when see enemy,
Enter pipe"

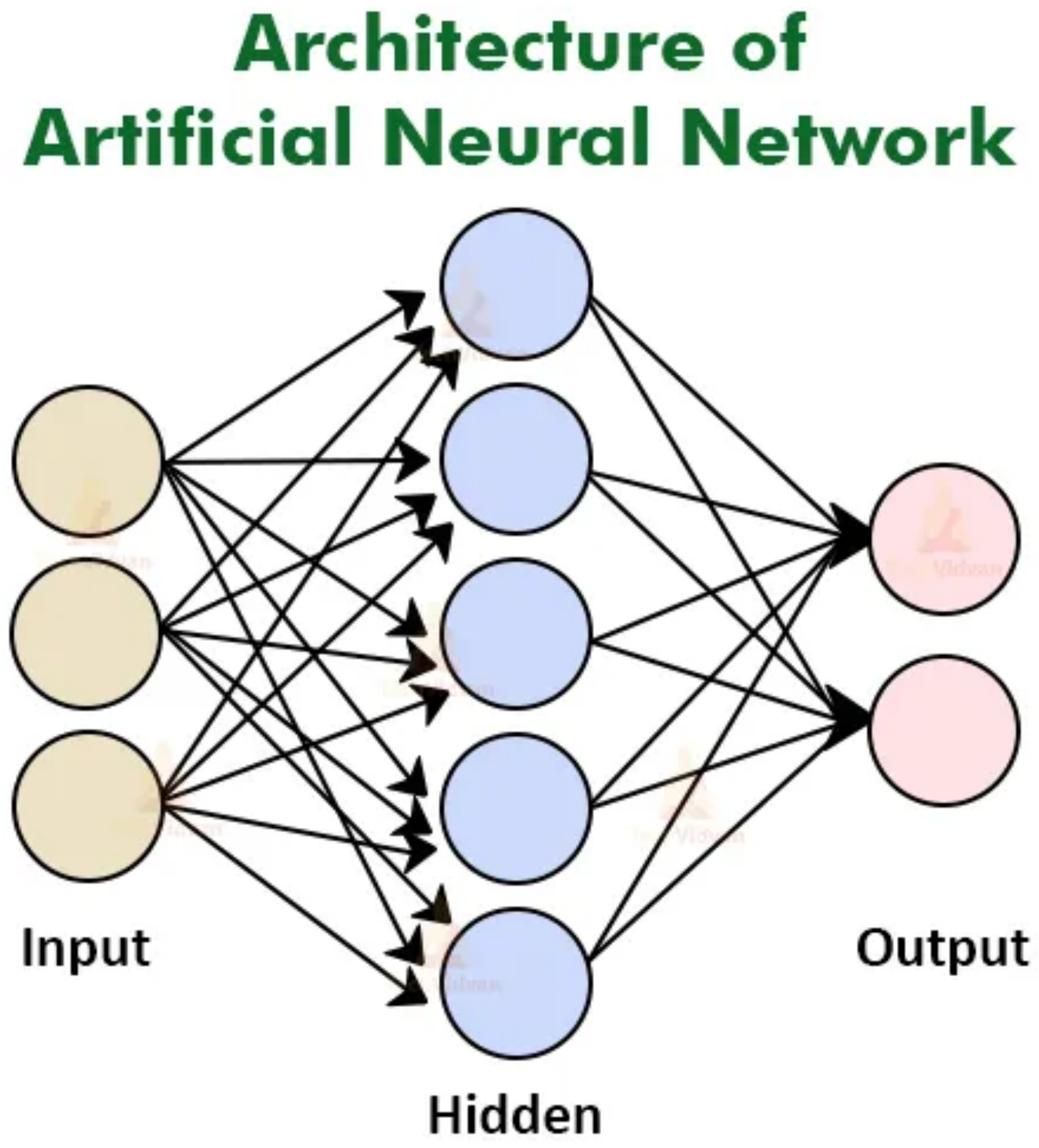
https://www.businessinsider.com/most-expensive-video-game-ever-sold-super-mario-bros-2019-3

ML allows computer to learn like humans

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/

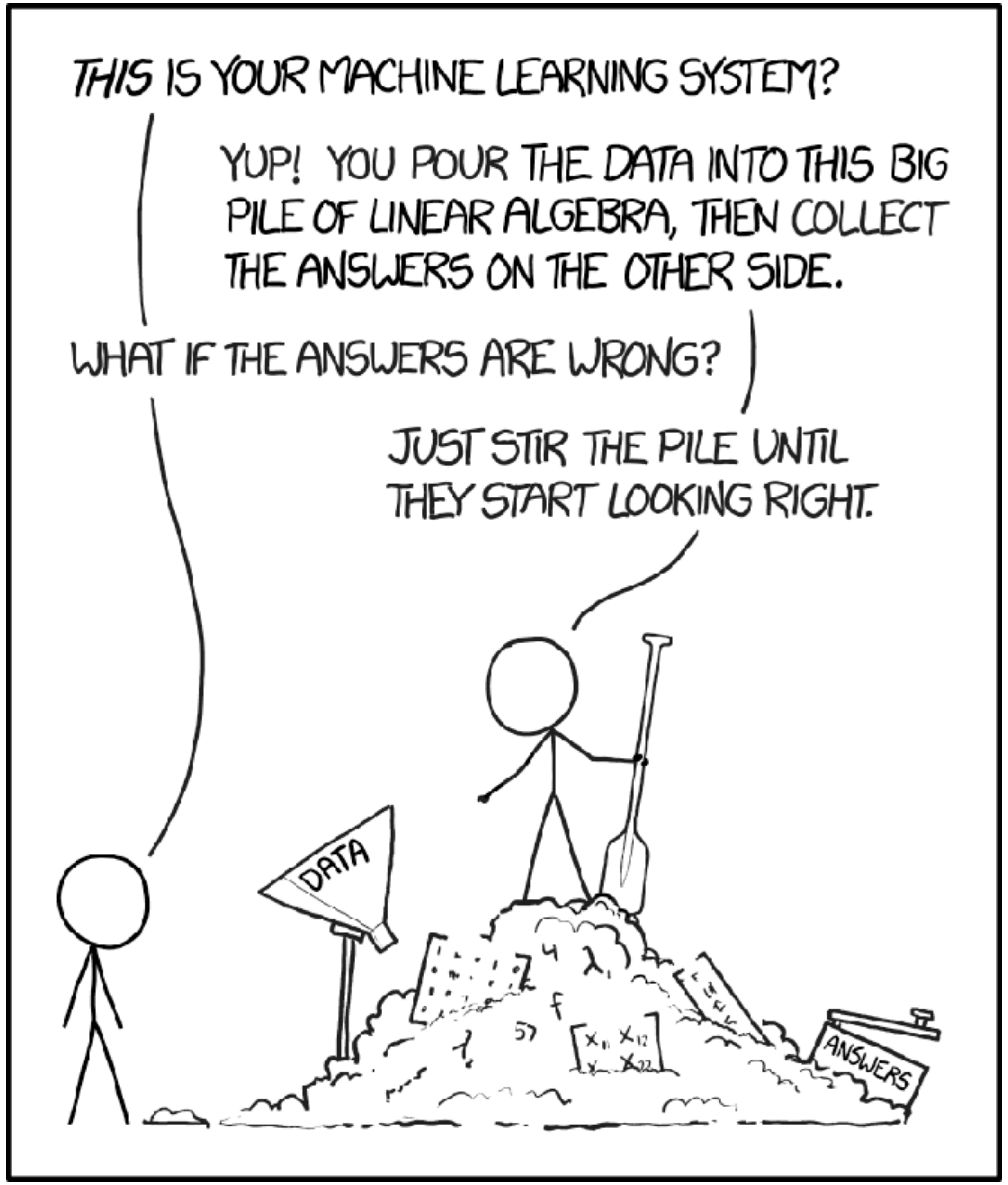
A computer learned this!

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

A computer learned this!

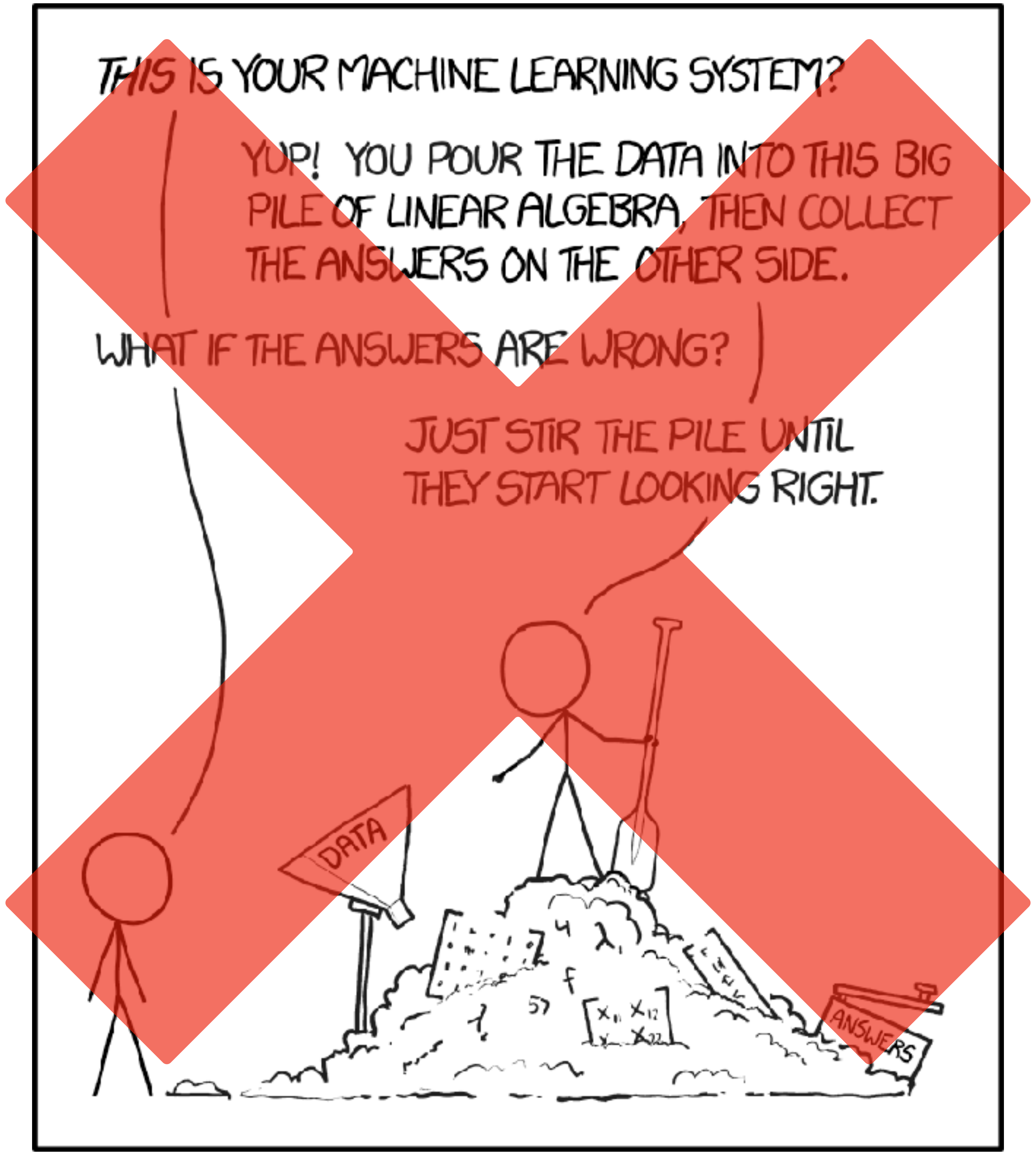
https://pytorch.org/tutorials/intermediate/mario_rl_tutorial.html

A misconception



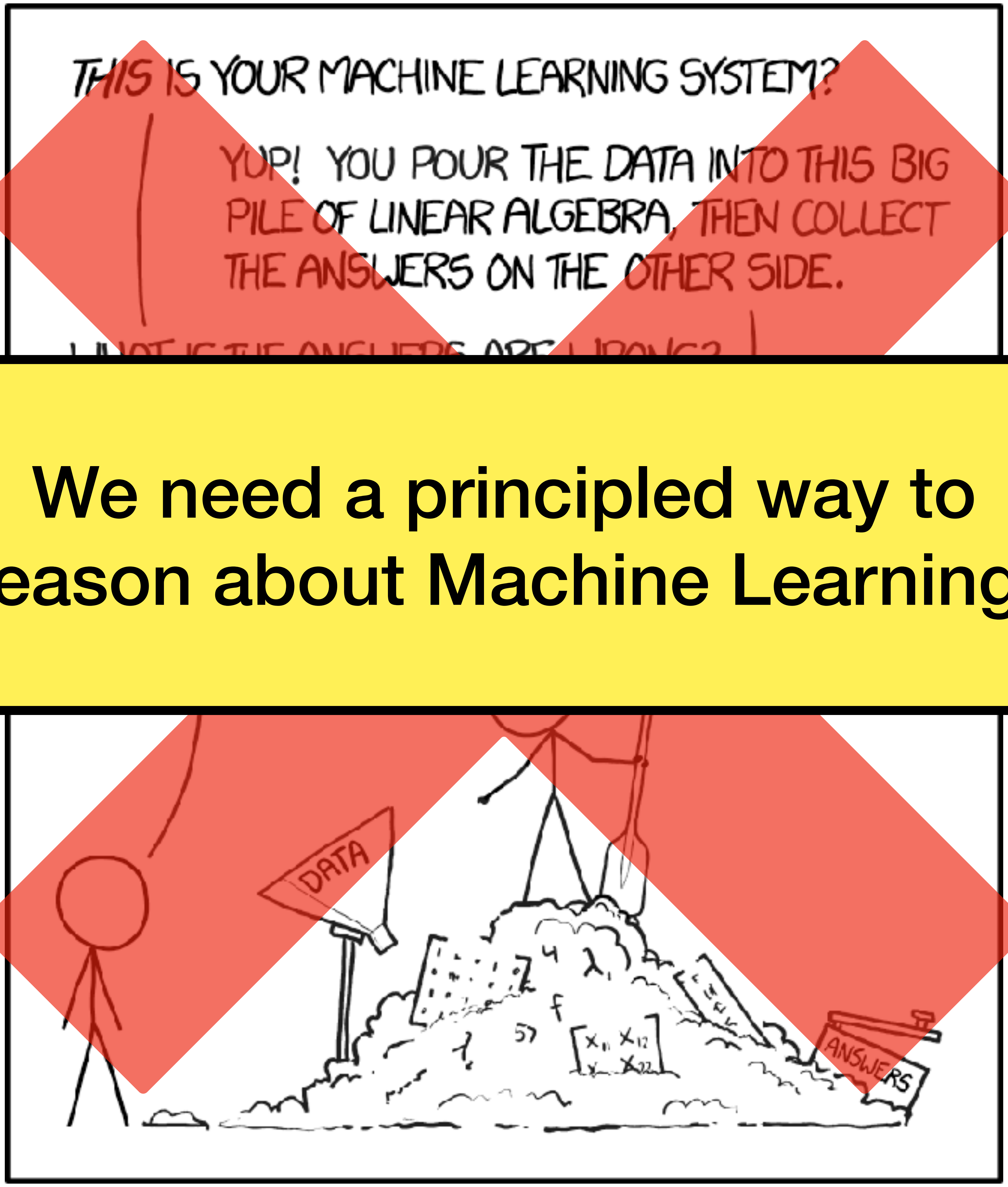
https://xkcd.com/1838/

A misconception



https://xkcd.com/1838/

A misconception



We need a principled way to reason about Machine Learning!

https://xkcd.com/1838/

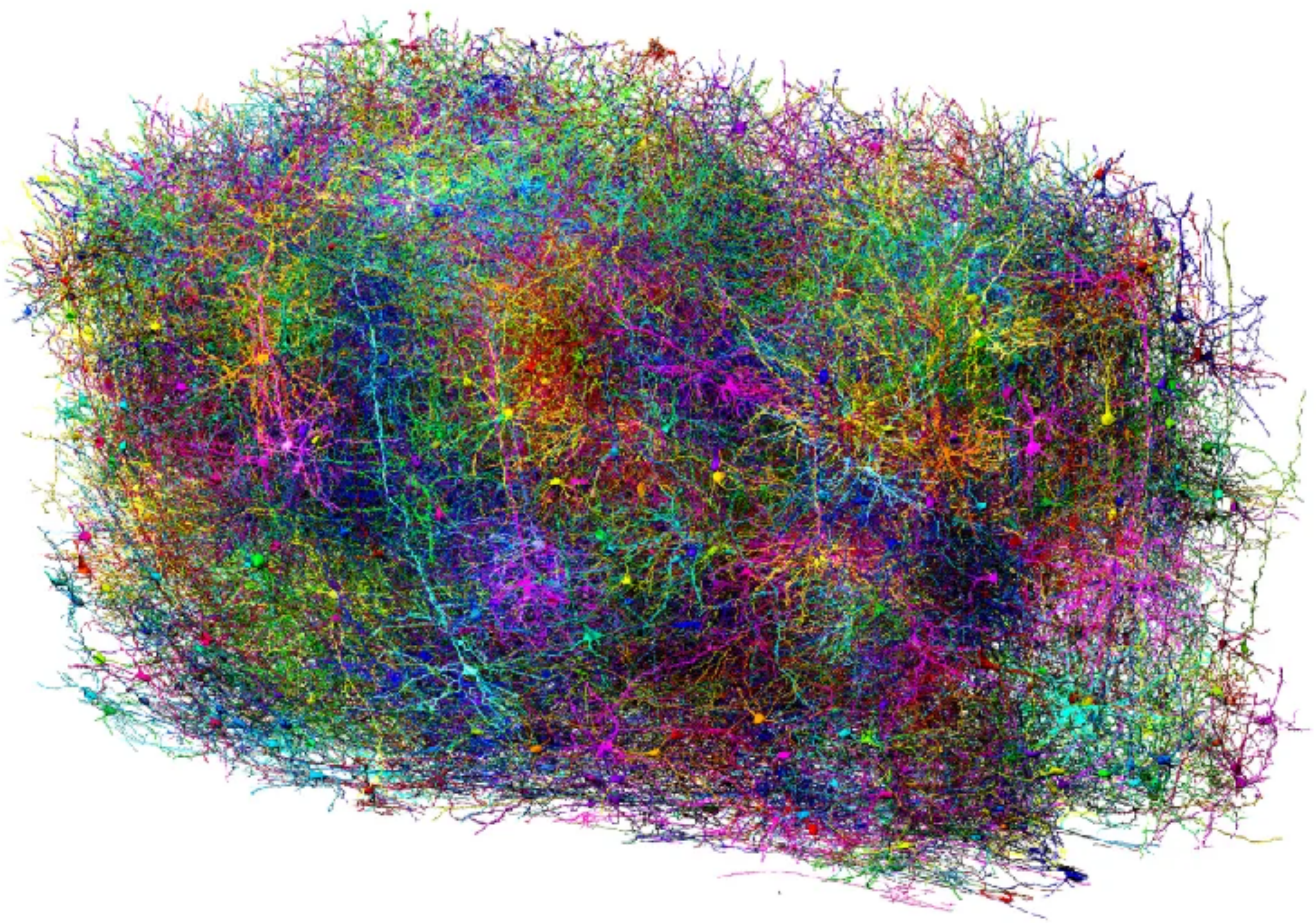
Why should I care about machine learning?

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/

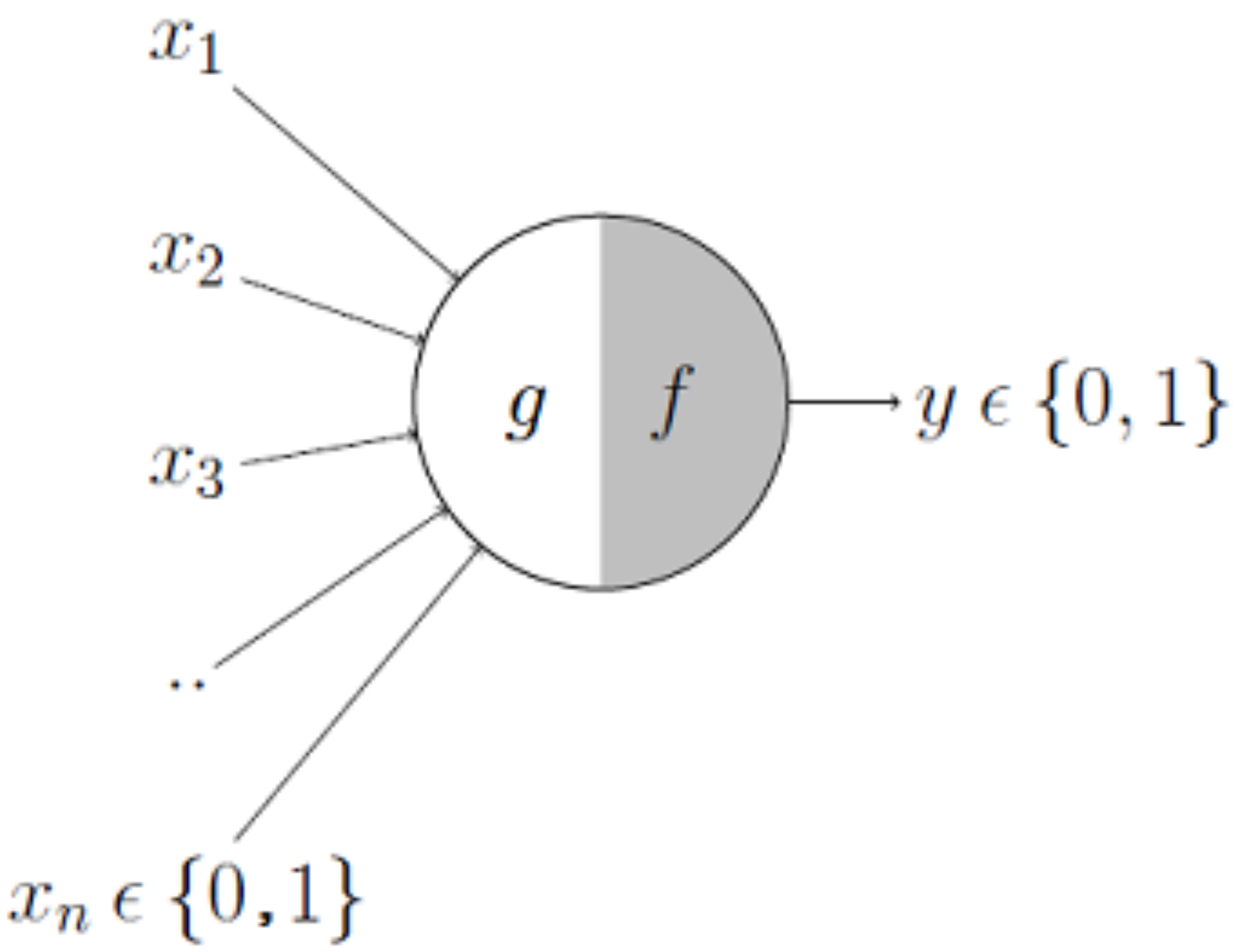
Why should I care about machine learning?

Let me give a *brief* history…

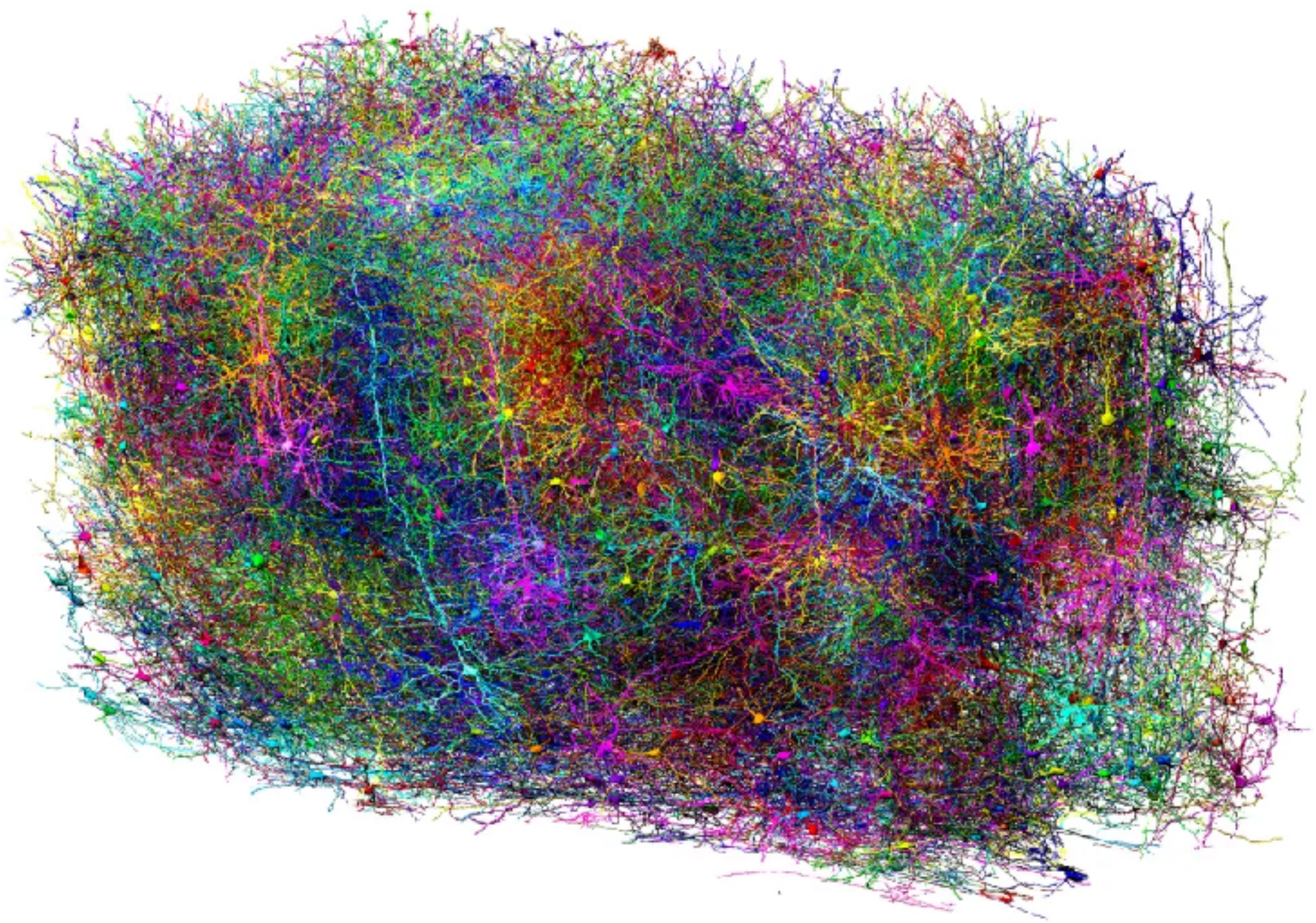
This section adapts and copies from the first lecture of Intro to ML taught by Professor Wen Sun.
See https://www.cs.cornell.edu/courses/cs4780/2023fa/



Artificial Neural Network (1943)



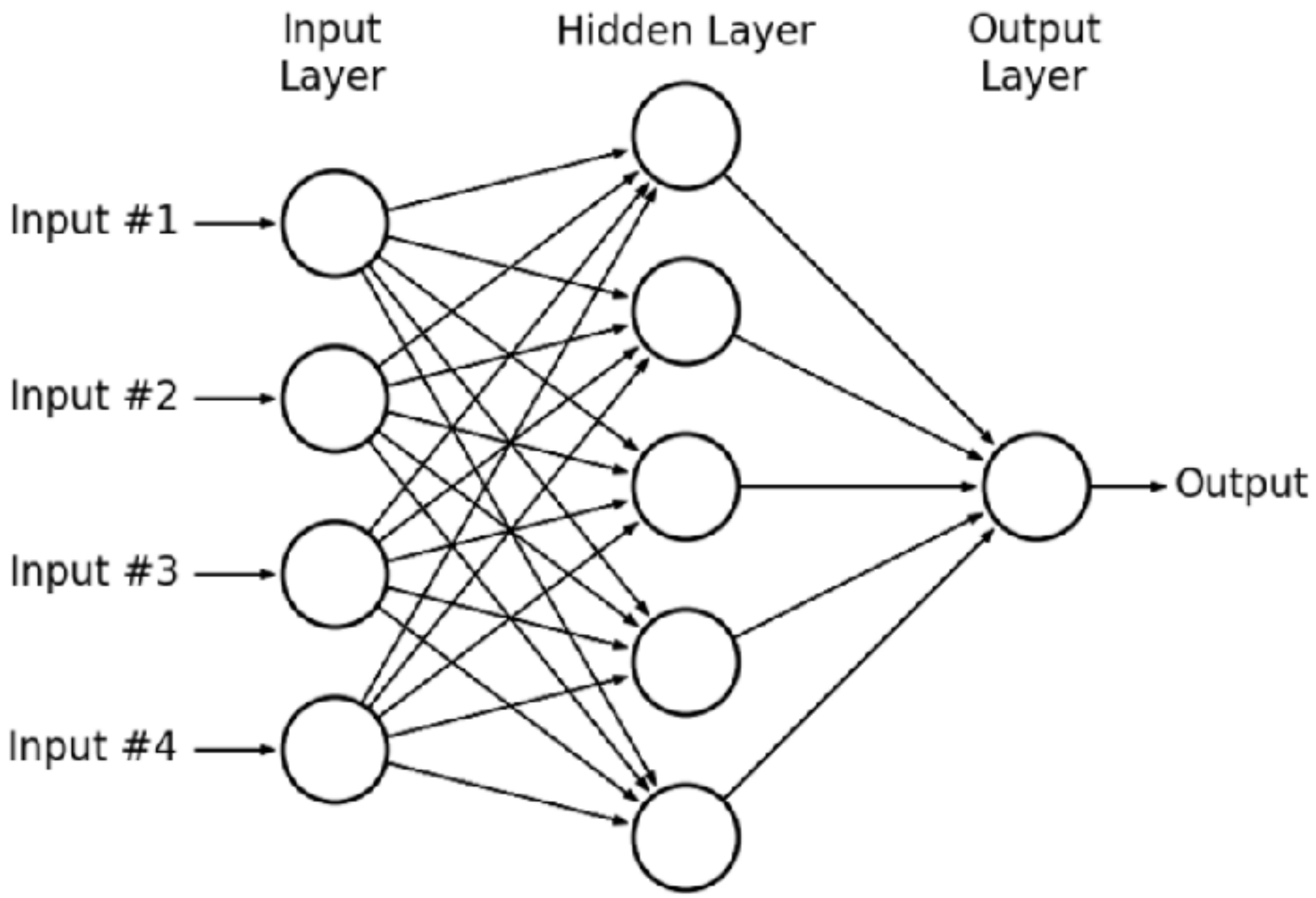
McCulloh-Pitts Neuron



Mouses’s brain

https://www.cs.cornell.edu/courses/cs4782/2025sp/slides/pdf/week1_1.pdf
https://www.cnn.com/2025/04/15/science/3d-brain-map-mouse-mammal-breakthrough/index.html

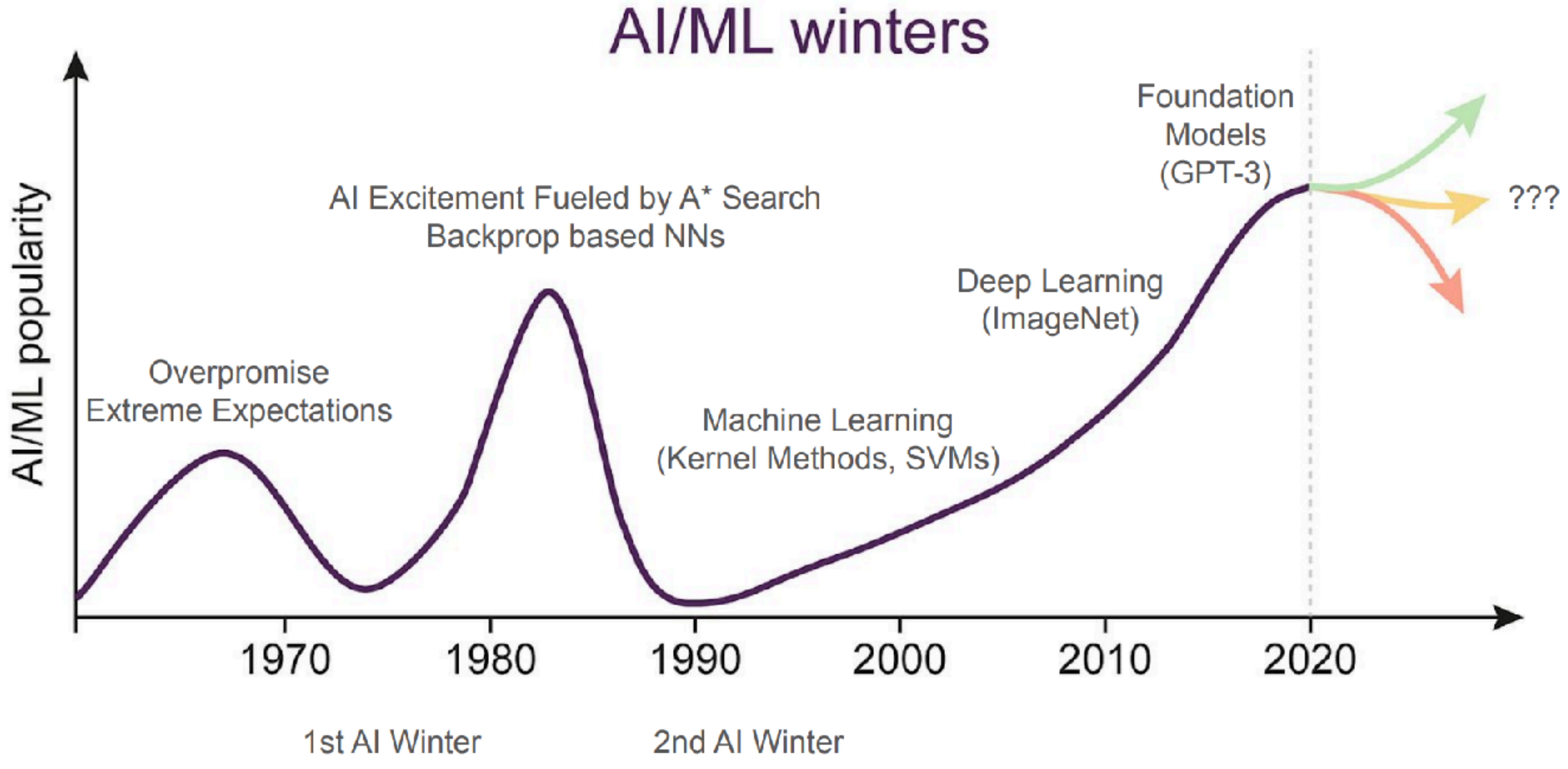
Perceptron (1957), Multi-Layer Perceptron (1965)



Rosenblatt's Multi-Layer Perceptron

https://www.cs.cornell.edu/courses/cs4782/2025sp/slides/pdf/week1_1.pdf
https://www.futurelearn.com/info/courses/machine-learning-for-image-data/0/steps/362737

A Brief History of Machine Learning

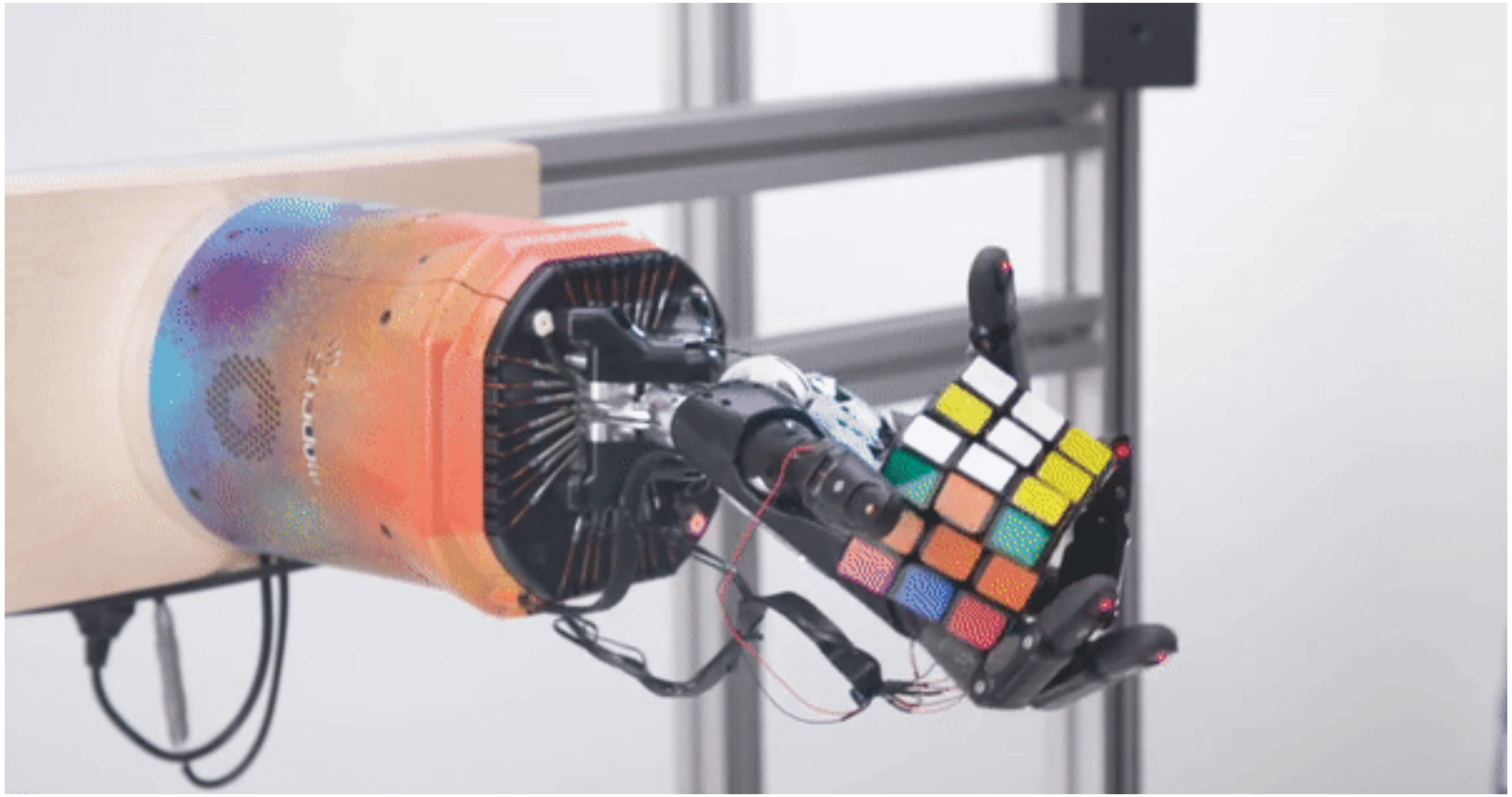


https://www.cs.cornell.edu/courses/cs4782/2025sp/slides/pdf/week1_1.pdf
https://medium.com/ersiliaio/beyond-the-hype-of-ai-ml-in-biomedicine-76c198c07467

AlphaGo Beats Lee Se-dol (2016)

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

Robot Solves Rubik's Cube from Scratch (2019)



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

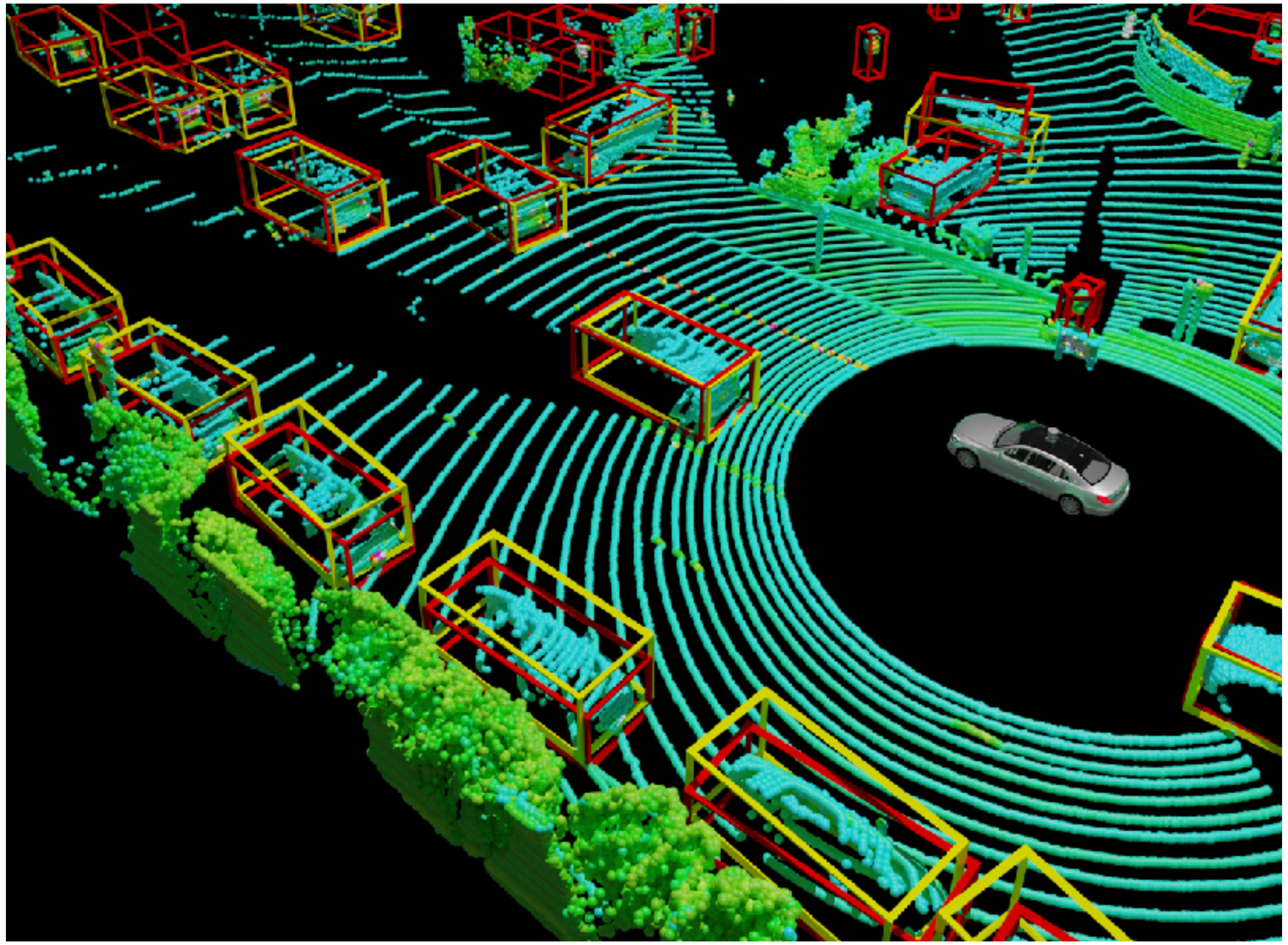
```latex
\usepackage[natbibapa]{apacite}  % Alternative: \usepackage{natbib
\bibliographystyle{apacite}       % Alternative: \bibliographystyl
```

2. **Use the citation command**:

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

Use in Critical Systems

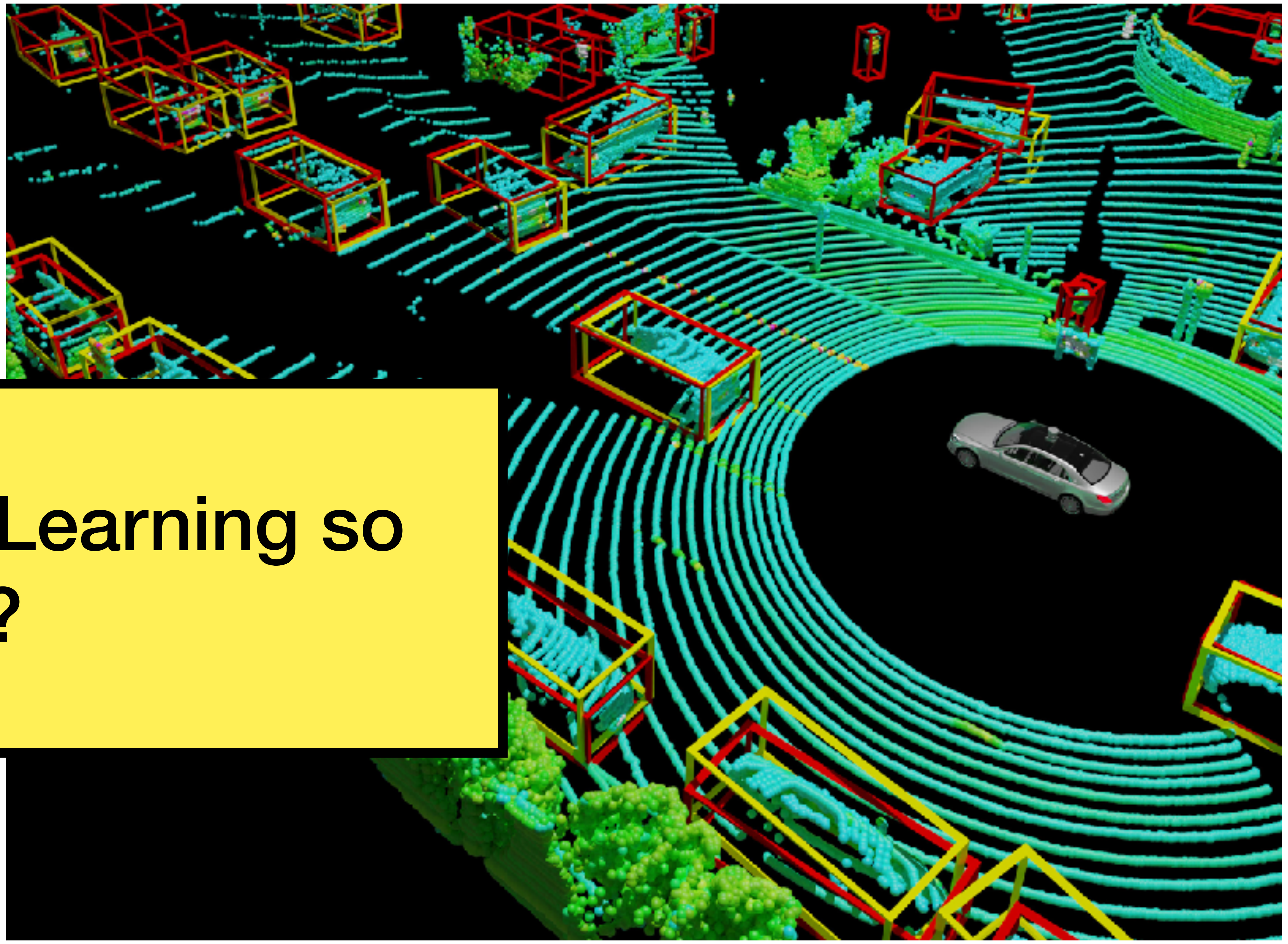
- Wildfire & earthquake prediction
- Self-driving cars



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

Use in Critical Systems

- Wildfire & earthquake prediction
- Self-driving cars



Why is Machine Learning so good?

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

How to reason about machine learning

Back to our scheduled programming…

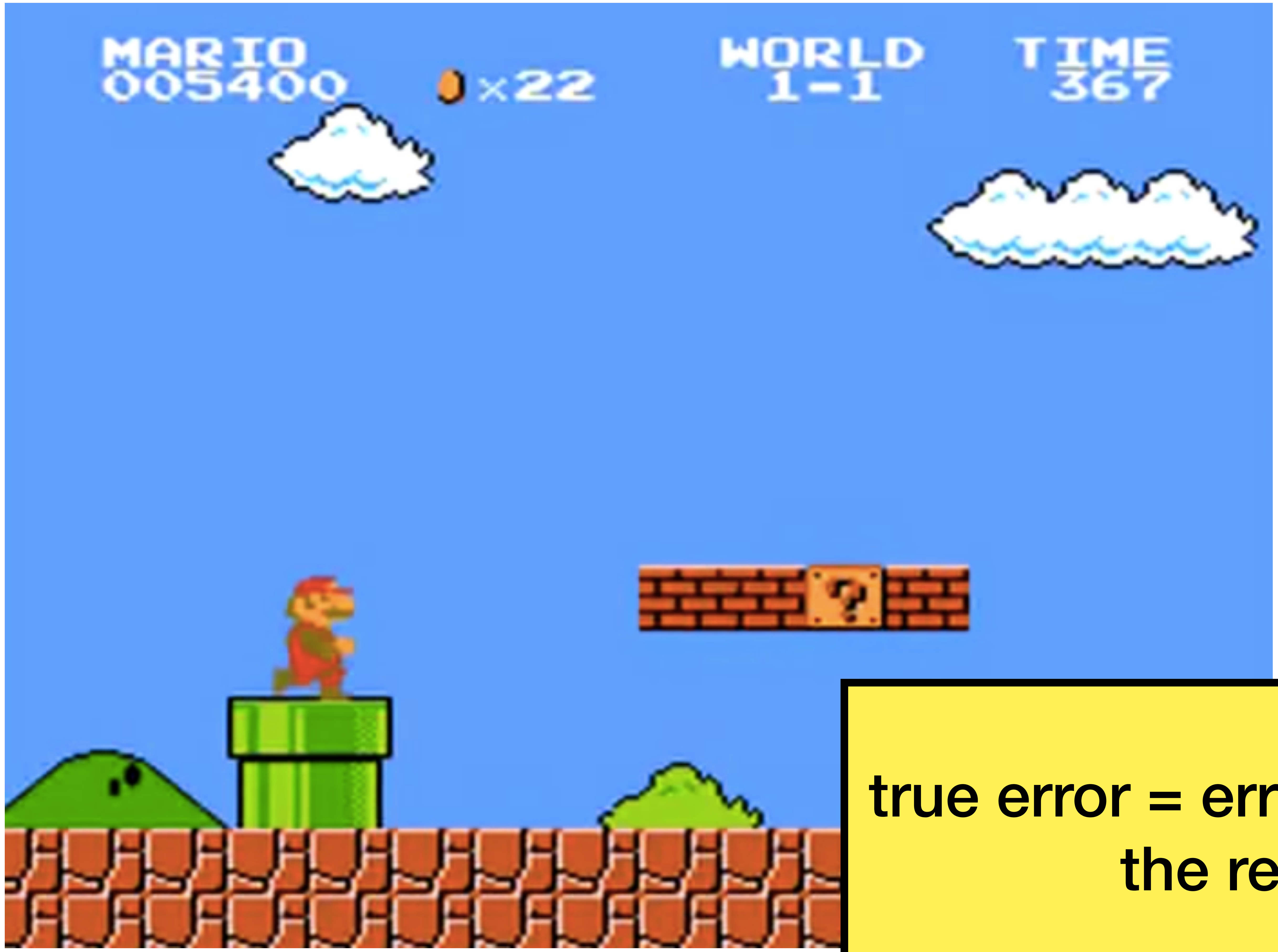
Super Mario Bros

How to measure success?

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

Super Mario Bros

How to measure success?



true error = error of the model in the real world!

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

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

true error = bias + variance

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

Bias-variance decomposition

Error of model in
the real world

true error = bias + 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

Training Data	Training Algorithm	Model
Ex 1. “We like to drink coffee here!” Ex 2. “Your coffee has caffeine.” …	→ ML Algorithm →	“We’re often sleepy in the morning, so we drink ________________”

Learning to finish sentences

Training Data

Ex 1. “We like to drink coffee here!”

Ex 2. “Your coffee has caffeine.”

…

Training Algorithm

ML Algorithm

Model

“We’re often sleepy in the morning, so we drink a cup of coffee. ”

Learning to finish sentences

Training Data

Ex 1. "We like to drink coffee here!"

Ex 2. "Your coffee has caffeine."

…

Training Algorithm

ML Algorithm

Model

"We're often sleepy in the morning, so we drink a cup of coffee. ____"

This is how ChatGPT works! It learns how to finish the user's prompt.

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.”

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

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.”

true error = bias + 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

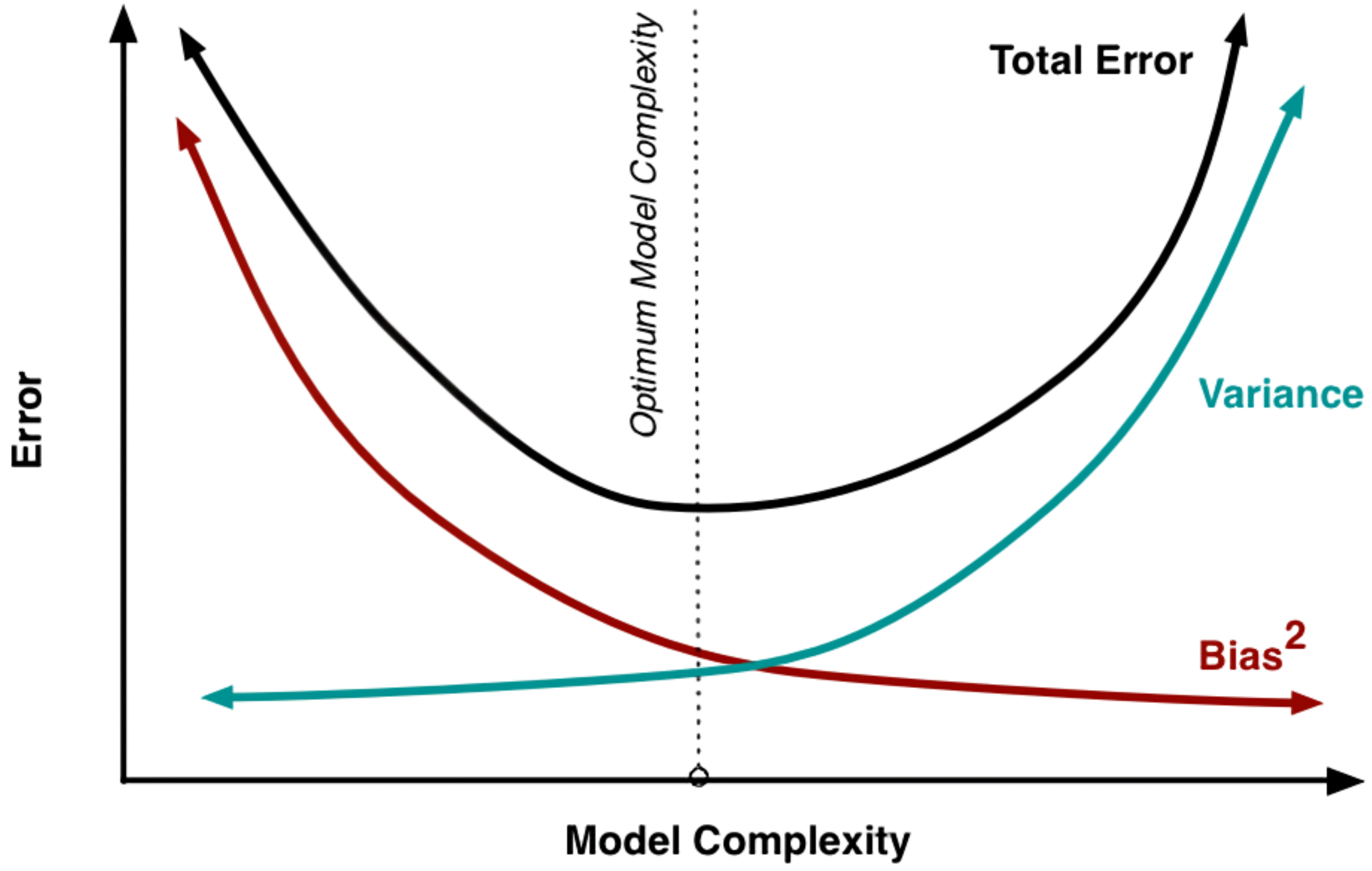
“Four score and seven years ago… our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.”

true error = bias + variance

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

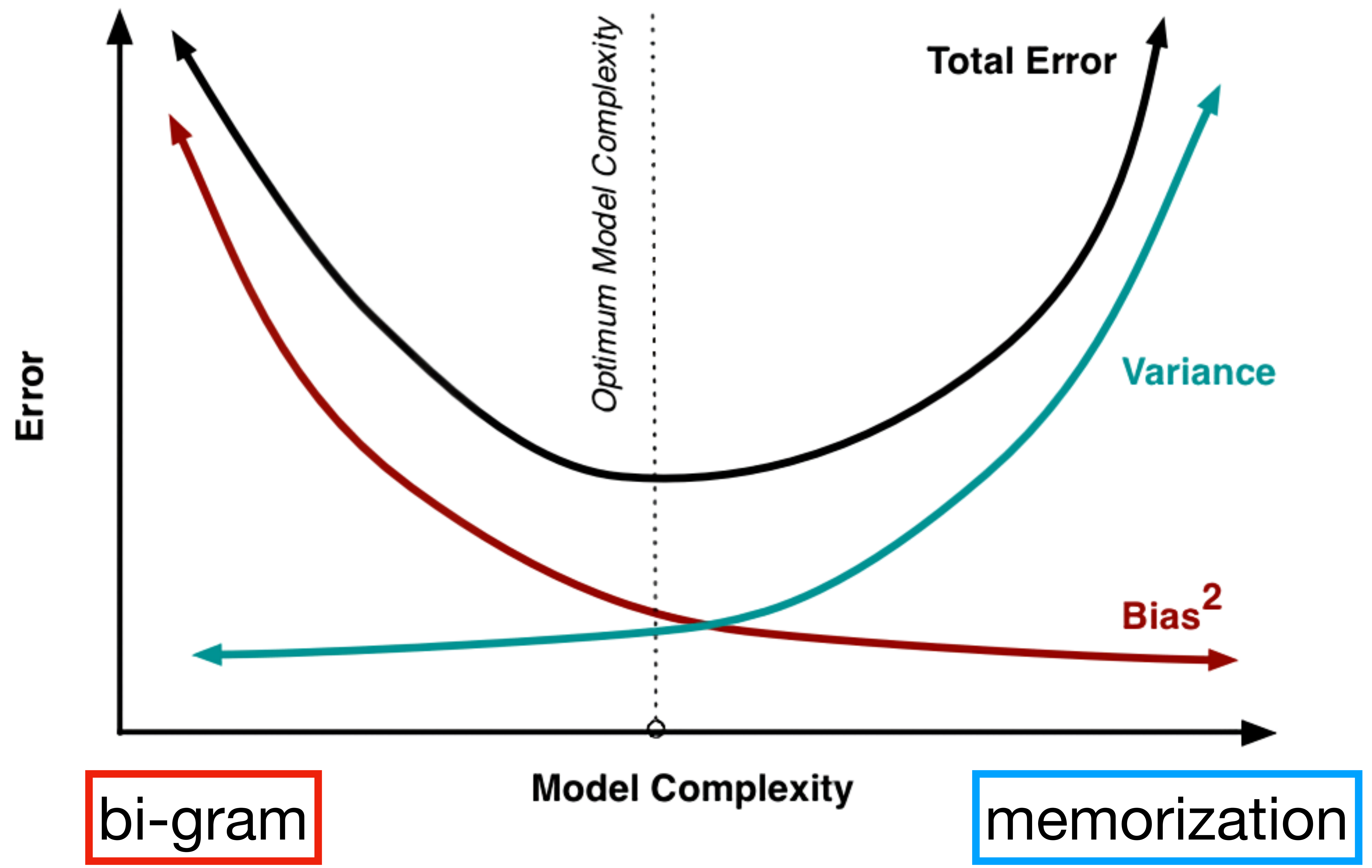
Can only finish sentences in dataset. With limited memory, it will not even be close to finishing every possible sentence.

Bias-variance tradeoff



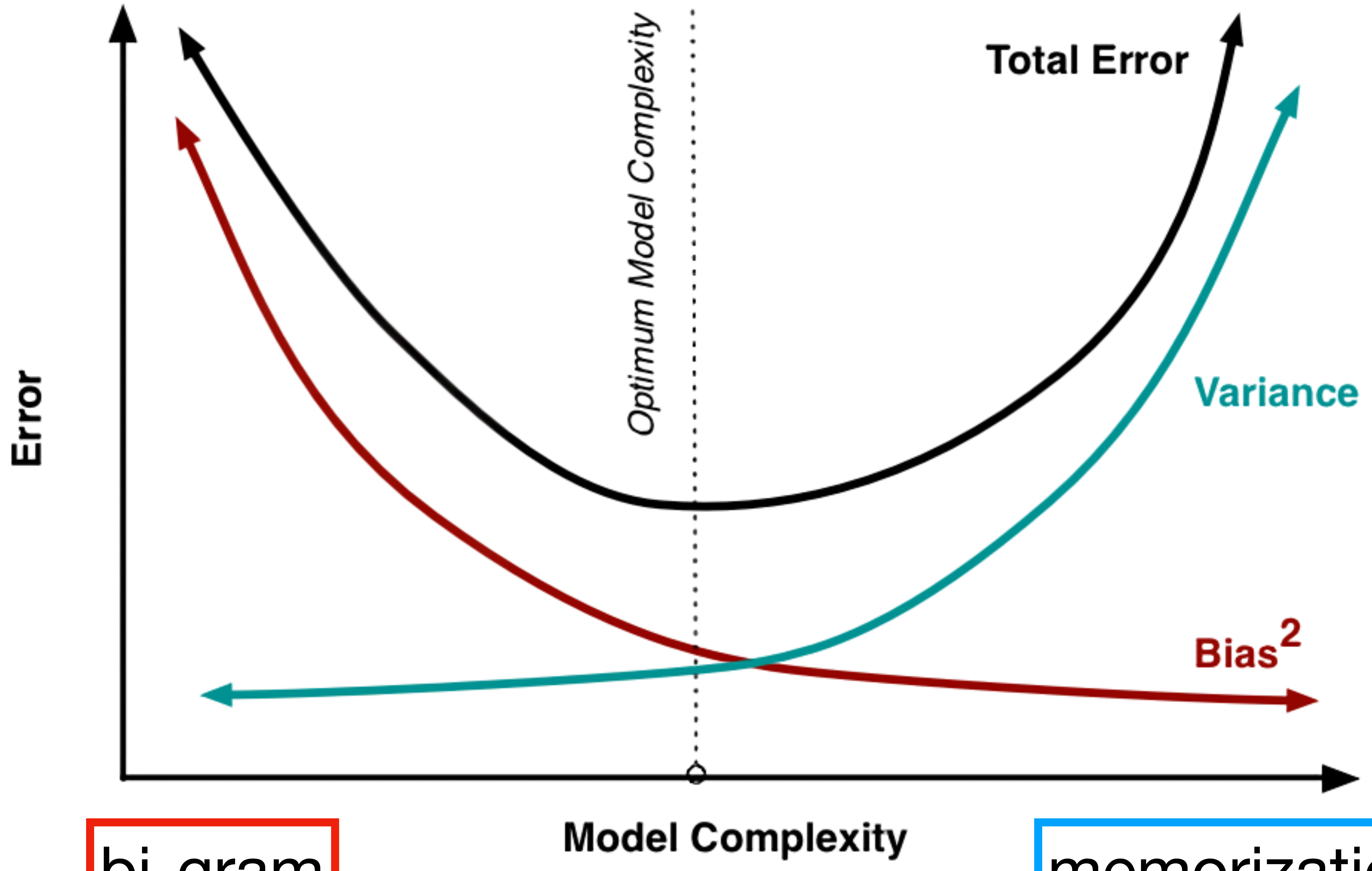
https://scott.fortmann-roe.com/docs/BiasVariance.html

Bias-variance tradeoff



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

Bias-variance tradeoff



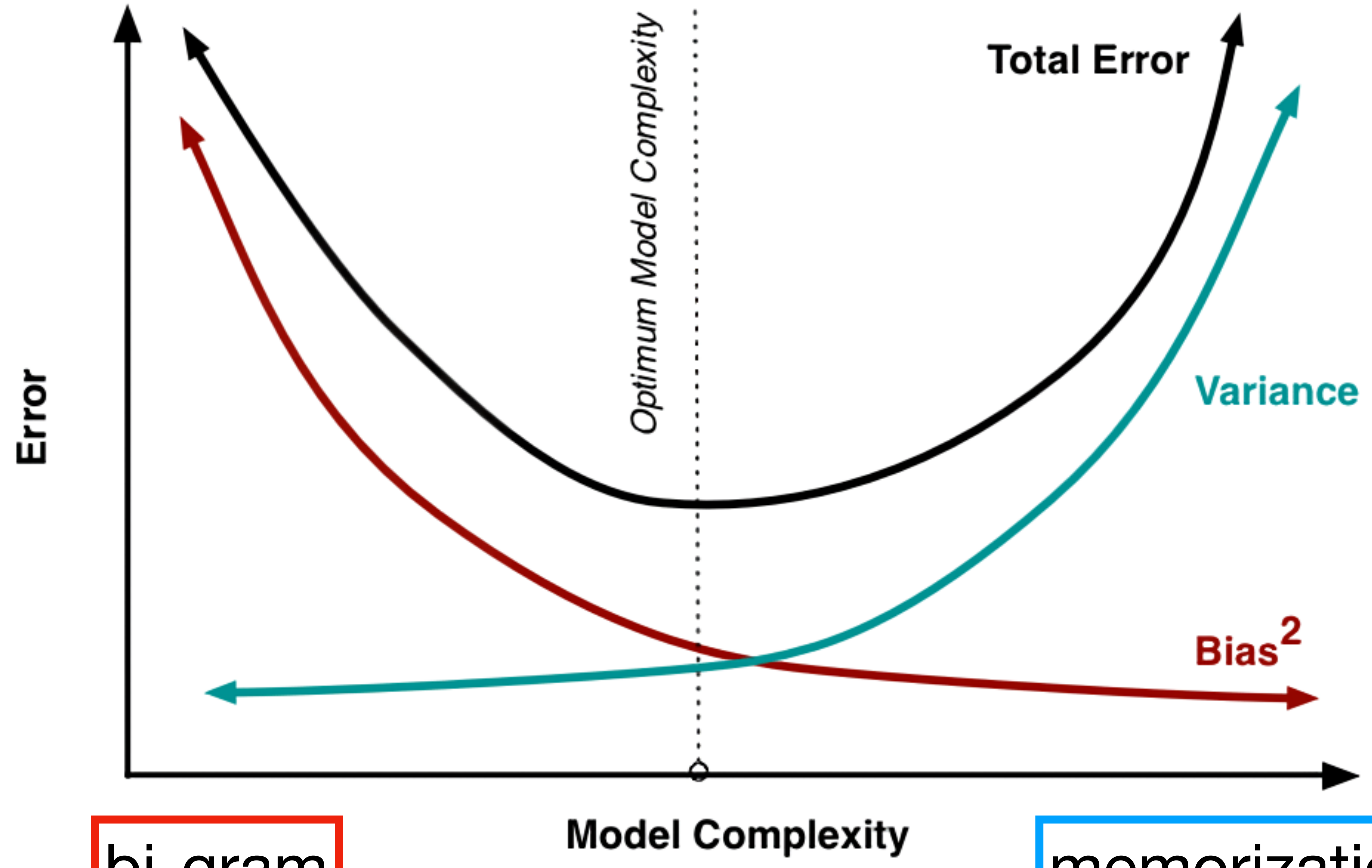
bi-gram

memorization

Minimum true error is too large!

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

Bias-variance tradeoff



bi-gram

memorization

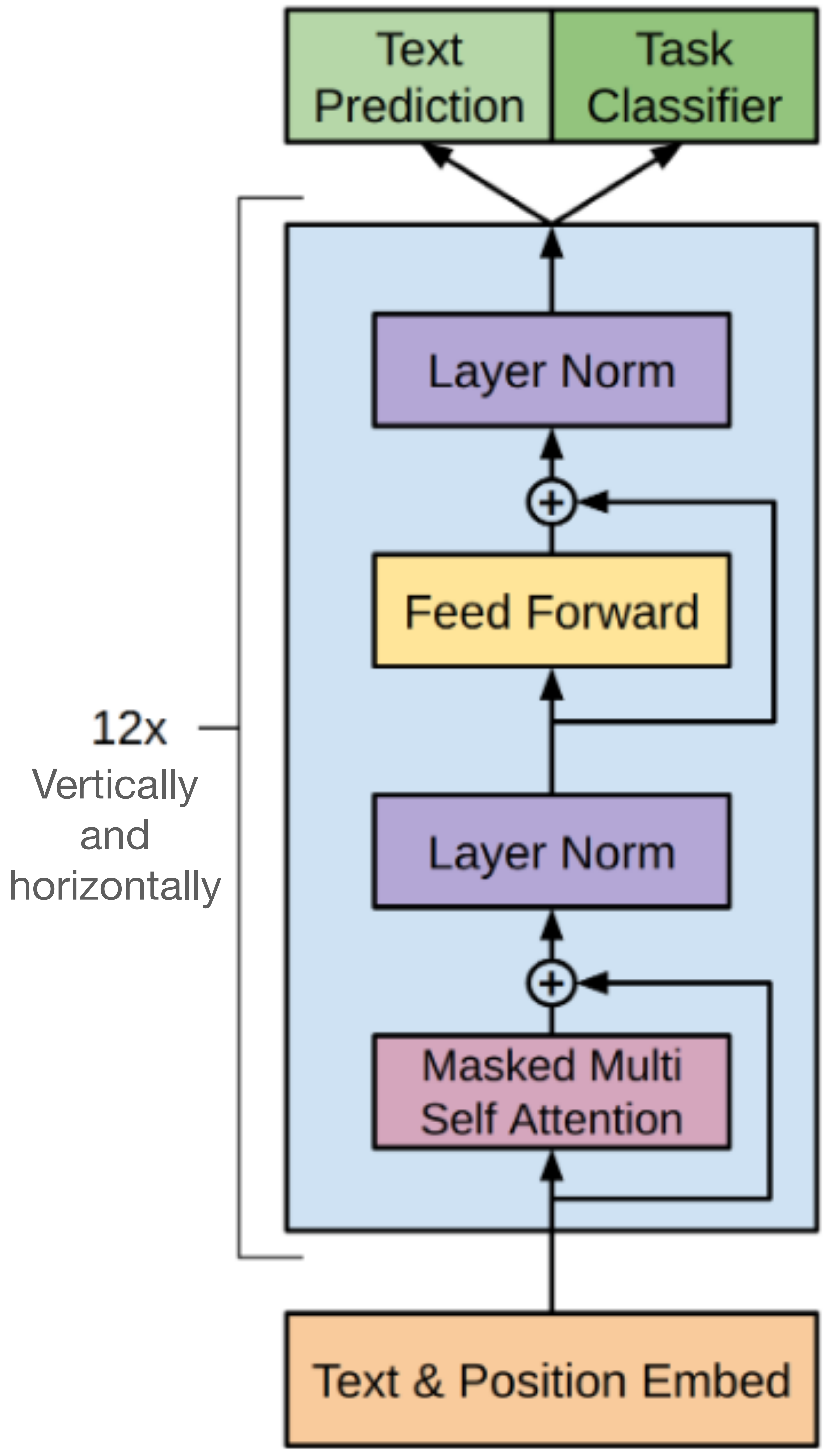
ChatGPT

Minimum true error is too large!

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

How ChatGPT cheats

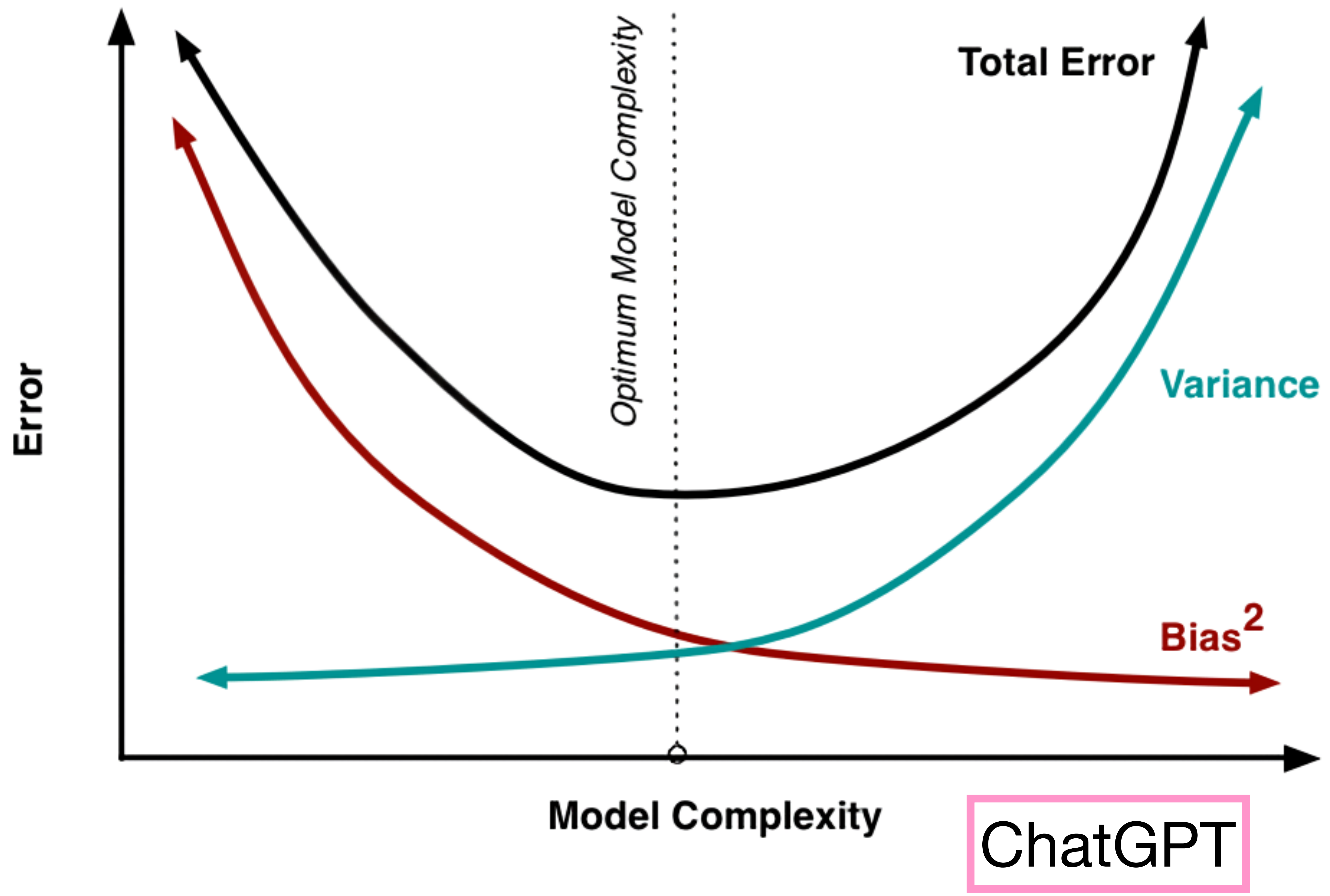
- GPT-4 has ~1 trillion parameters
- Push variance to the right via
  - Train on the a massive dataset (the internet)
  - Use a transformer-based architectures which allows for really good parallelization with GPUs.
  - Spend > $100 million



GPT-3 architecture

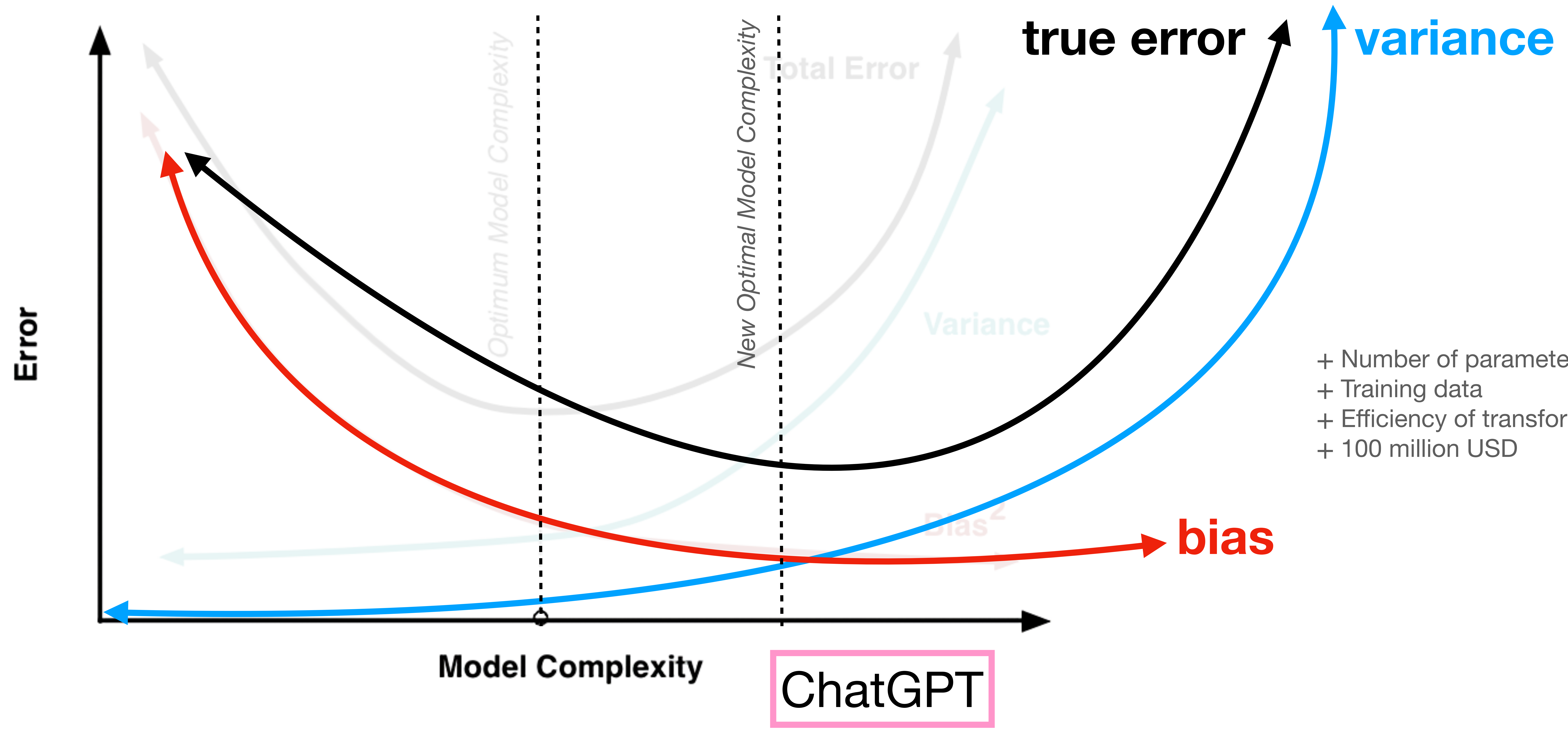
https://en.wikipedia.org/wiki/ChatGPT

Bias-variance tradeoff for GPT



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

Bias-variance tradeoff for GPT



+ Number of parameters
+ Training data
+ Efficiency of transformers
+ 100 million USD

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

Scaling laws

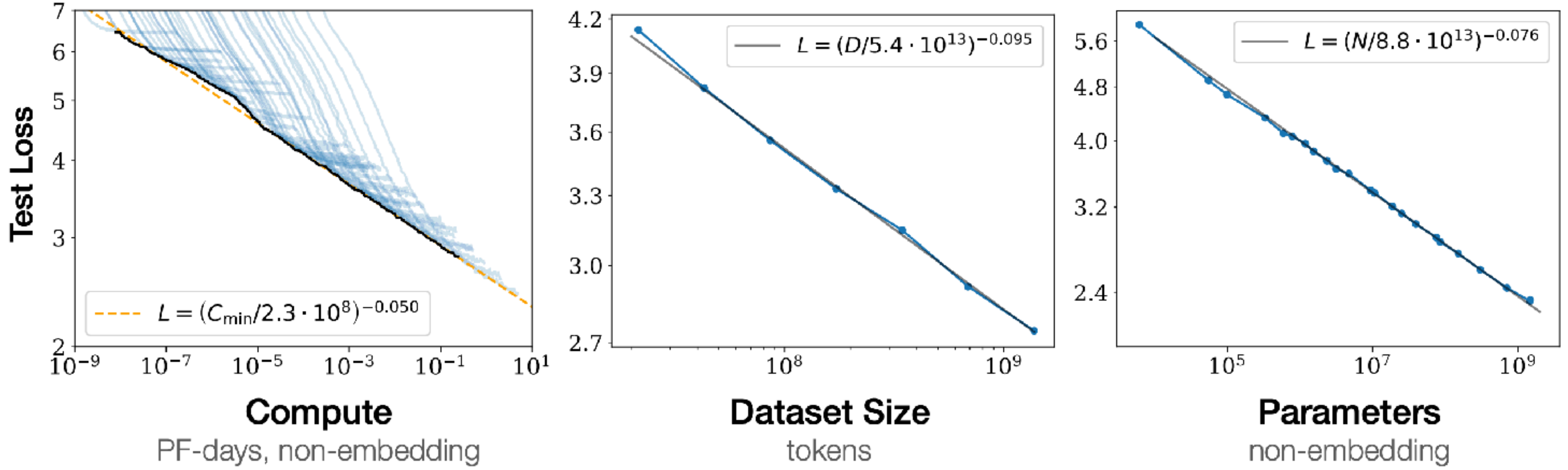


Figure 1 Language modeling performance improves smoothly as we increase the model size, datasetset size, and amount of compute[2] used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when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

https://arxiv.org/abs/2001.08361

ML is powerful!

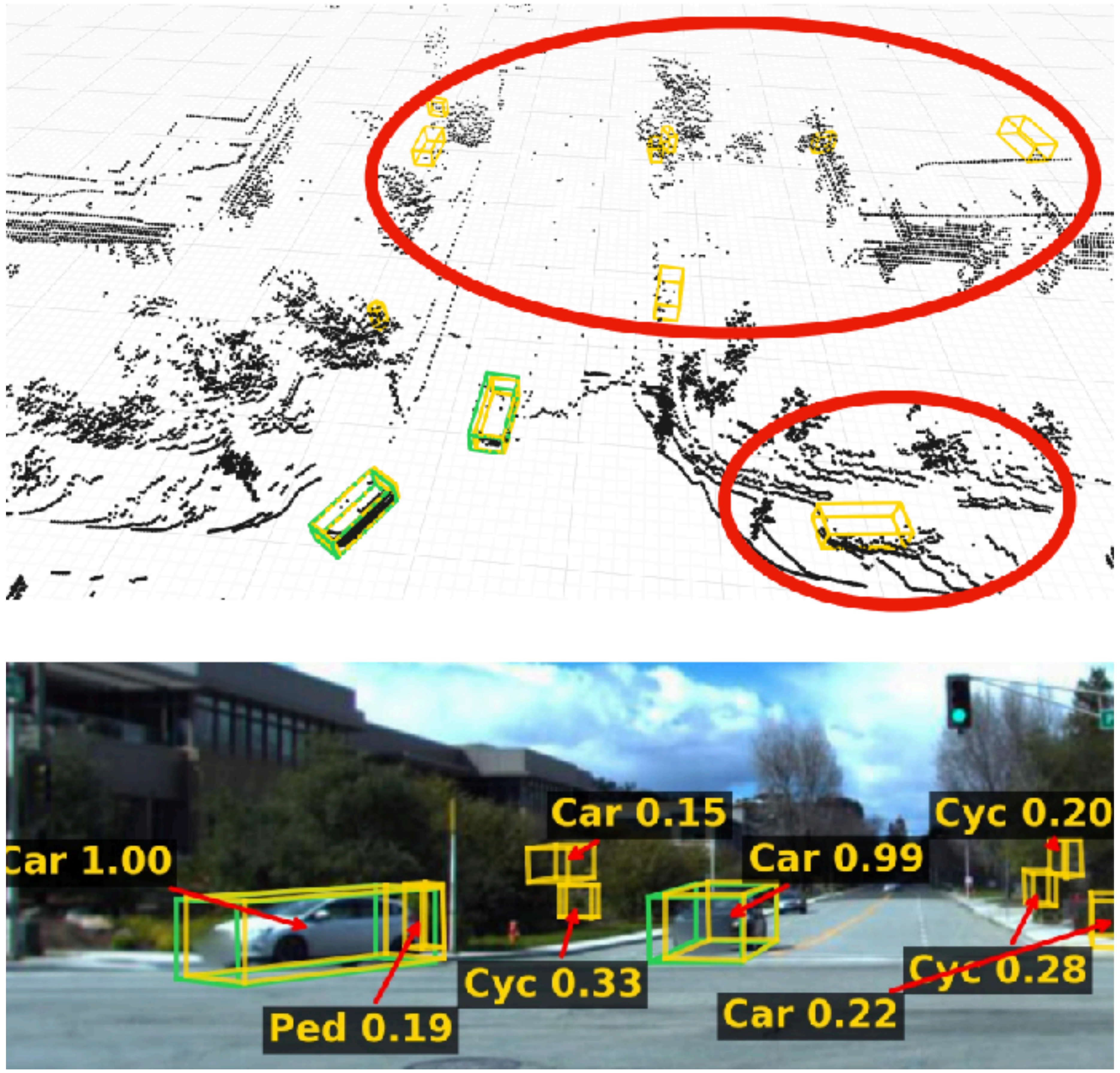
ML is powerful!

↓

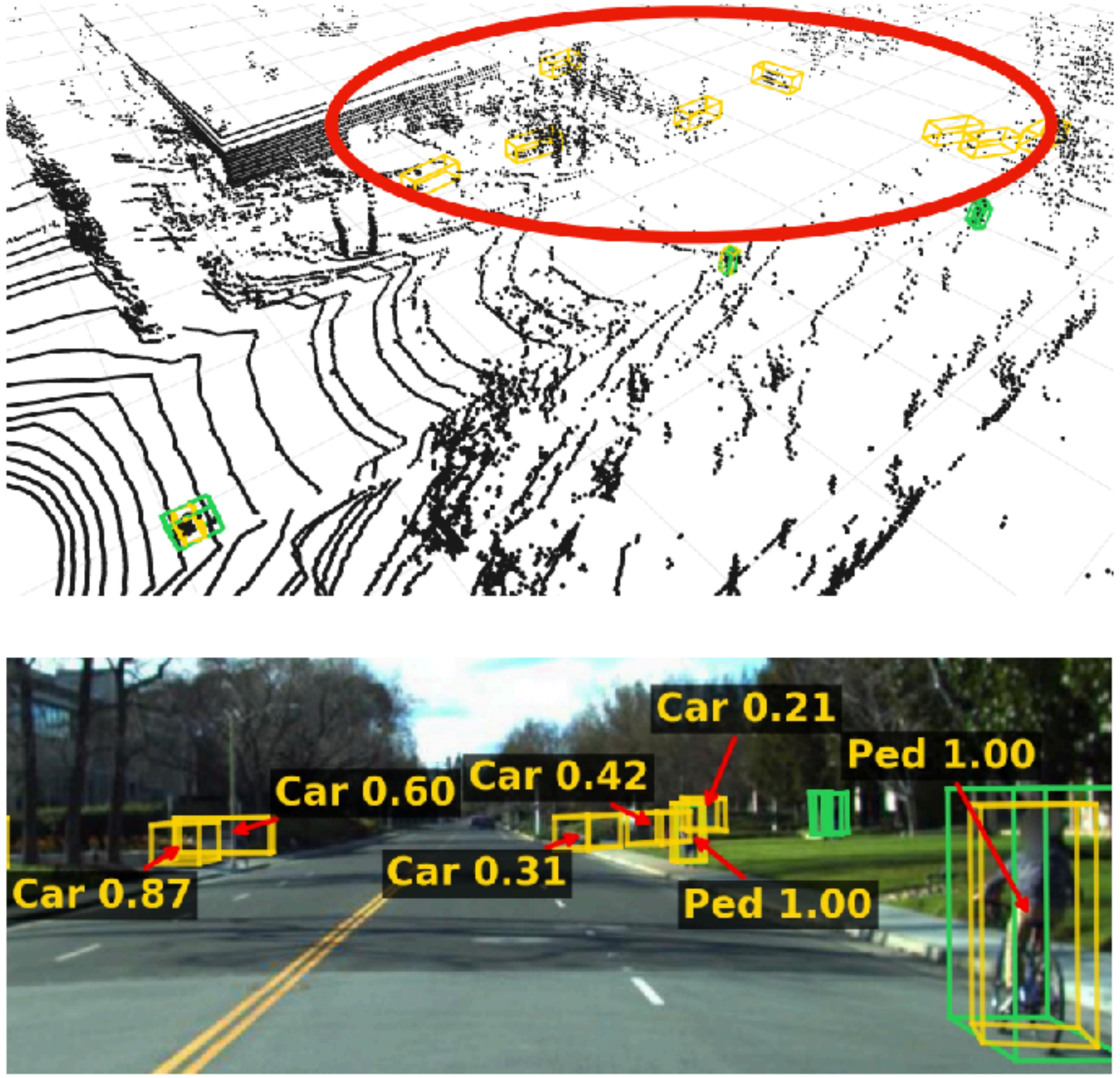
ML is dangerous!

ML can make mistakes

Example Scene 1



Example Scene 2



https://arxiv.org/abs/1812.04244

ML aren’t smart

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

ML aren't smart

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

Correlation $\neq$ Causation

This is called **overfitting.**

ML aren’t smart

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

Correlation ≠ Causation

This is called **overfitting.**

Try talking to ChatGPT! Does it feel like a person?

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

AI is nowhere close to human intelligence. But it 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/

Biased data implies biased machine learning

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

ML can be used for bad

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

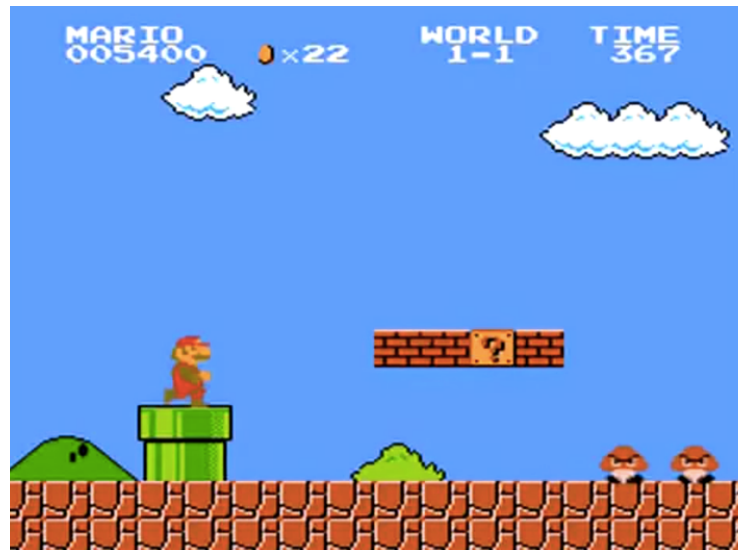
HEARD ON ALL THINGS CONSIDERED

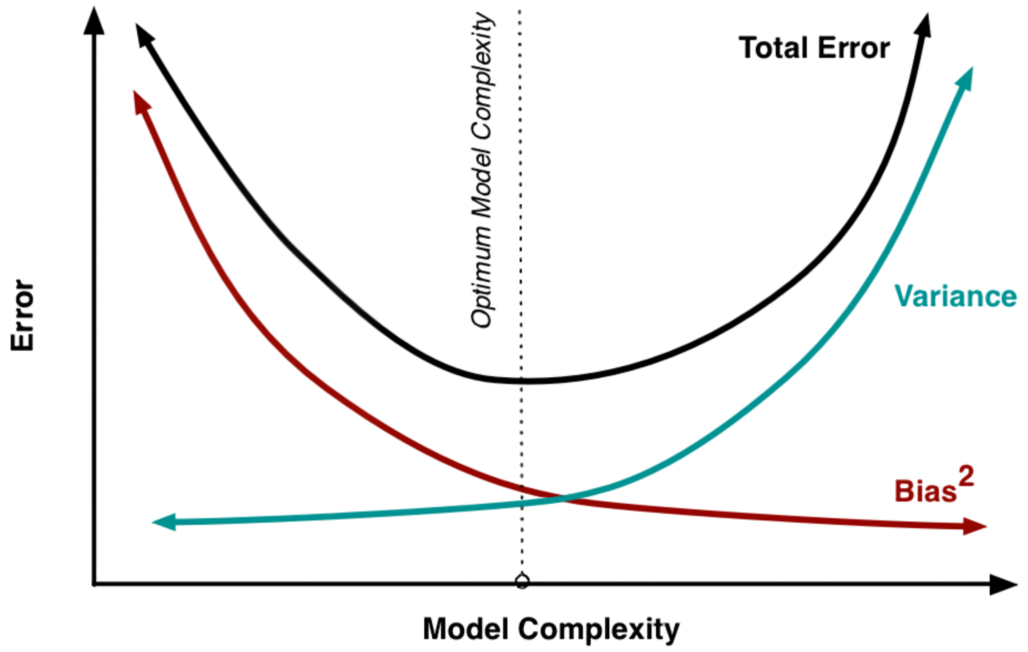
Shannon Bond

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

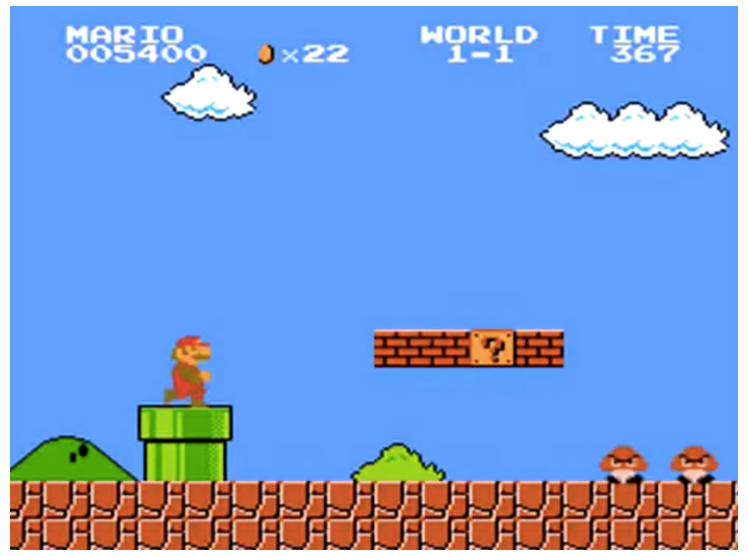
ML-powered advertising

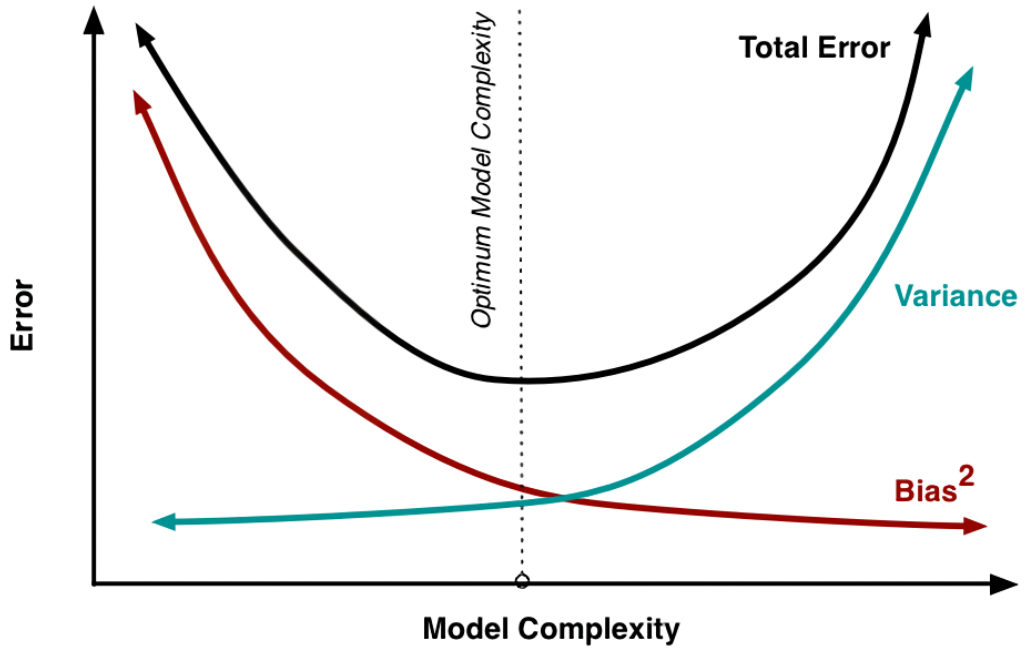
https://www.cnn.com/2024/10/04/politics/video/ai-elections-jake-tapper-lead-digvid
https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections


MARIO
005400
x22
WORLD
1-1
TIME
367


Error
Optimum Model Complexity
Total Error
Variance
Bias^2
Model Complexity

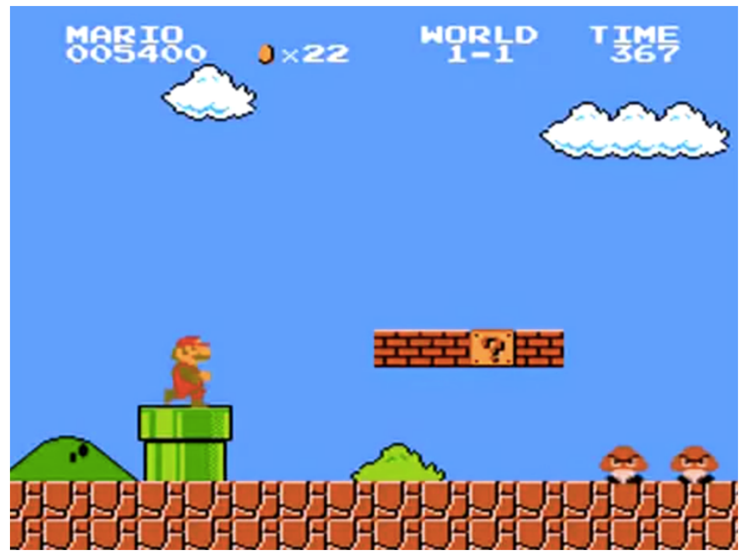
Summary

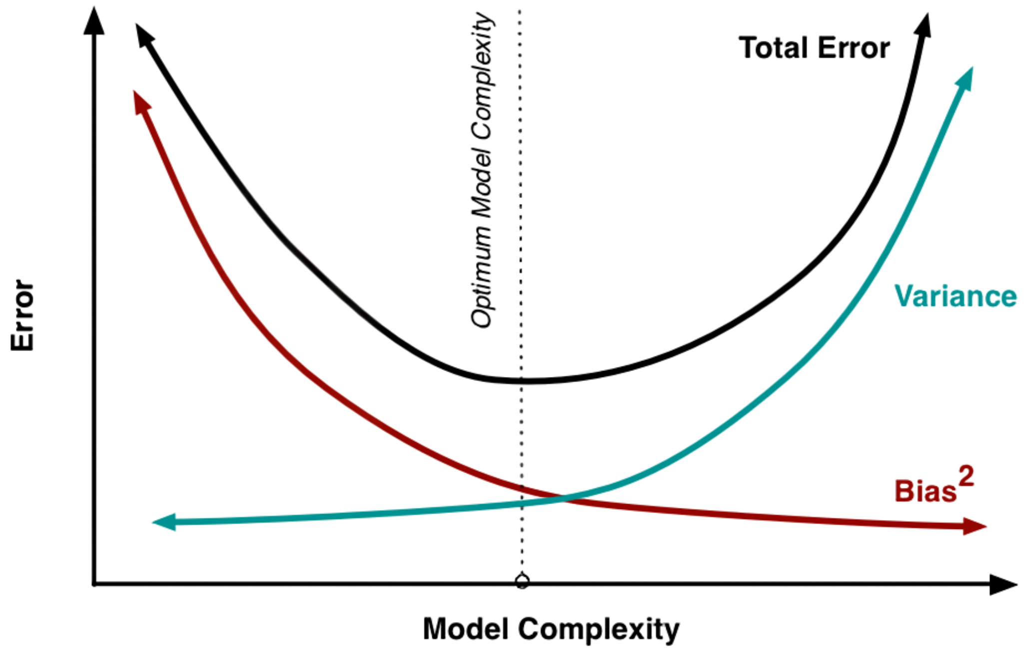




- ML is allows computers to learn from data.

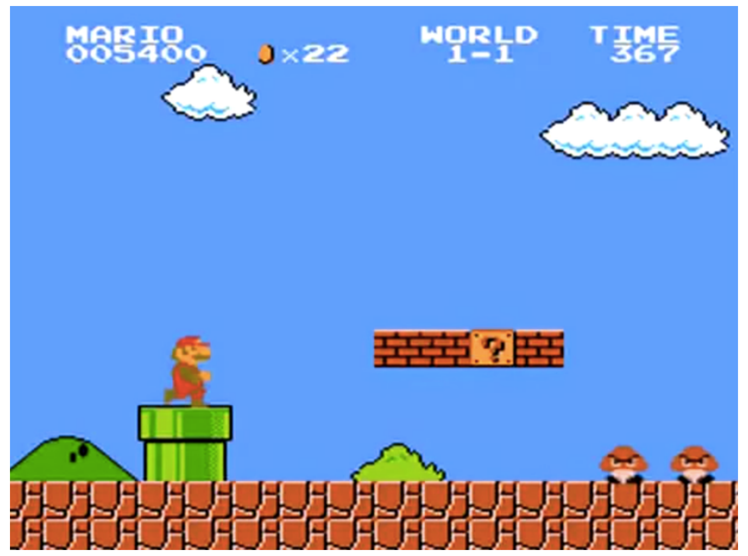
Summary

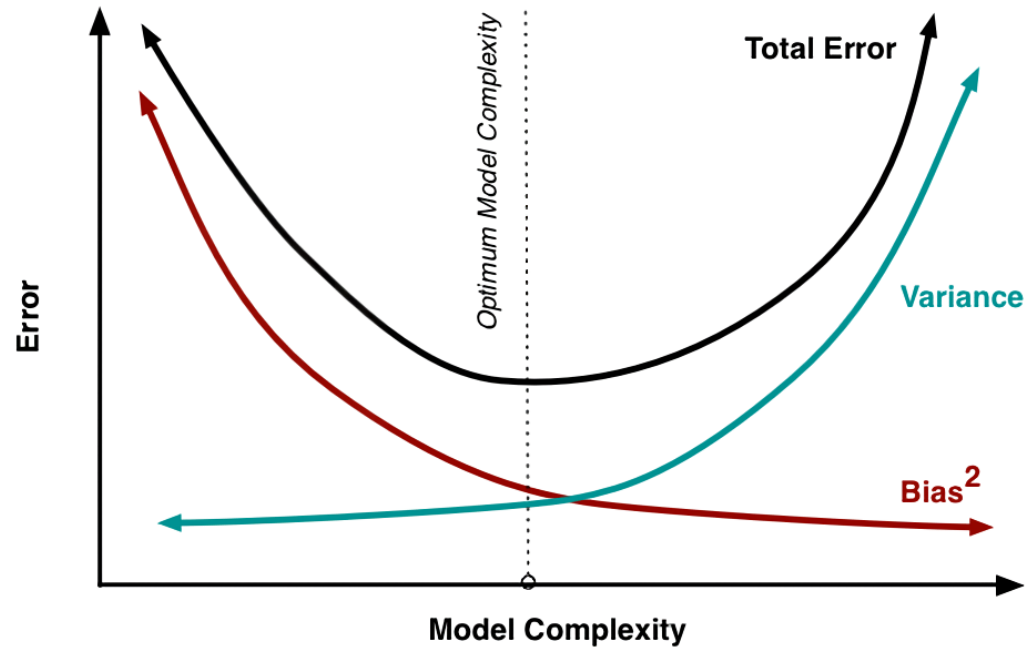




Summary

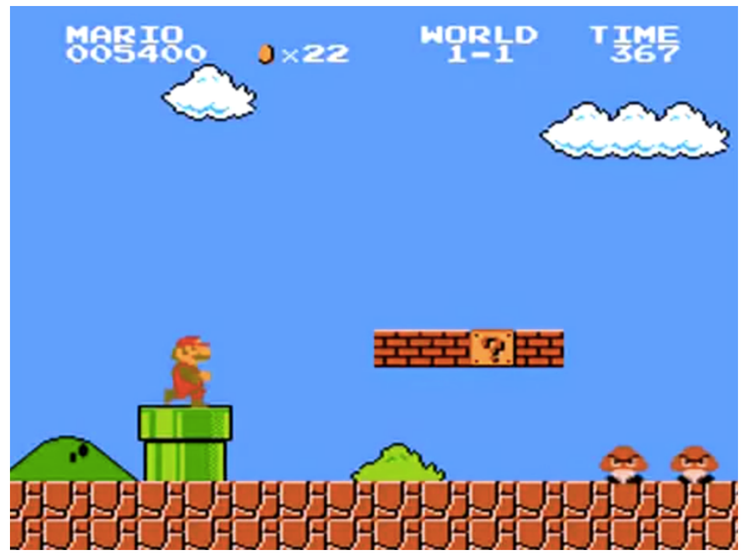
- ML is allows computers to learn from data.
- ML-based AI has exploded in the last few years, especially generative AI for natural language tasks and image or video generation.

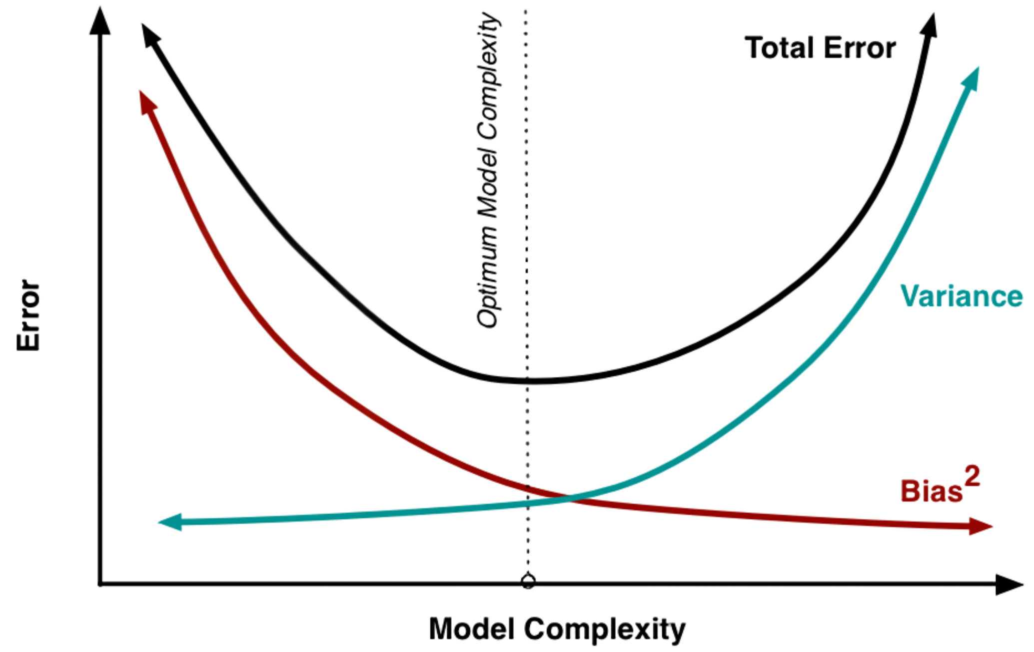




Summary

- ML is allows computers to learn from data.
- 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!





Summary

- ML is allows computers to learn from data.
- 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!
- We need to be careful with how we use it!