A Mindmap to grasp AI

L. Michel

Professor and Synchrony Chair for Cybersecurity
School of Computing
University of Connecticut

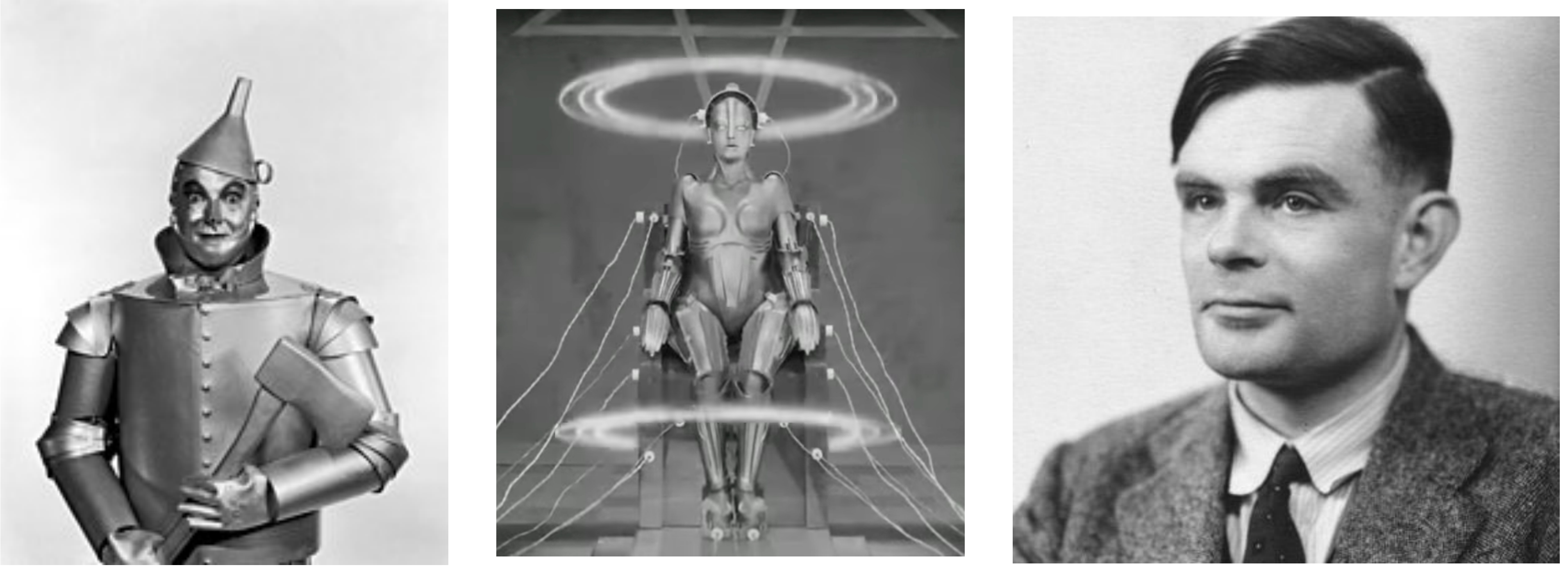
Overview

- AI in the last century
- GOFAI, Modern, and Generative Era
- GOFAI, Modern, and Generative AI… in a Nutshell
- What *is* AI?
- Reality and Hype of AI
- Opportunities, Challenges and Ethical dilemma

AI's Evolution

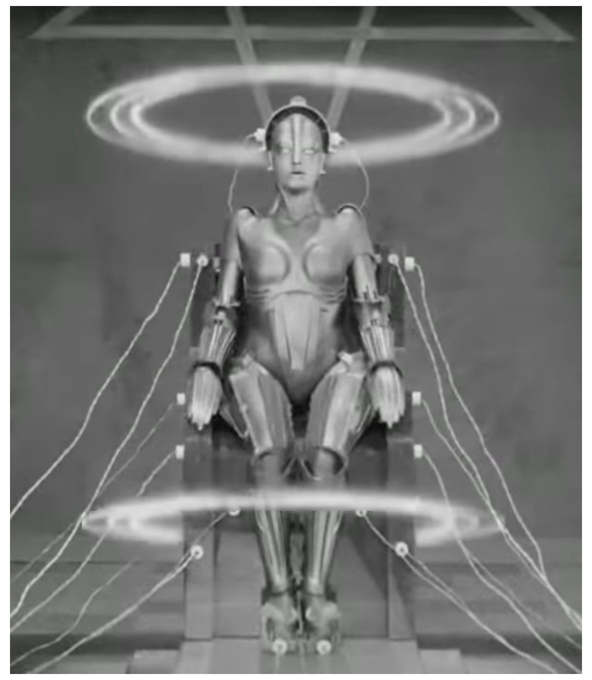
- A cycle of…
  - Boom
  - Bust

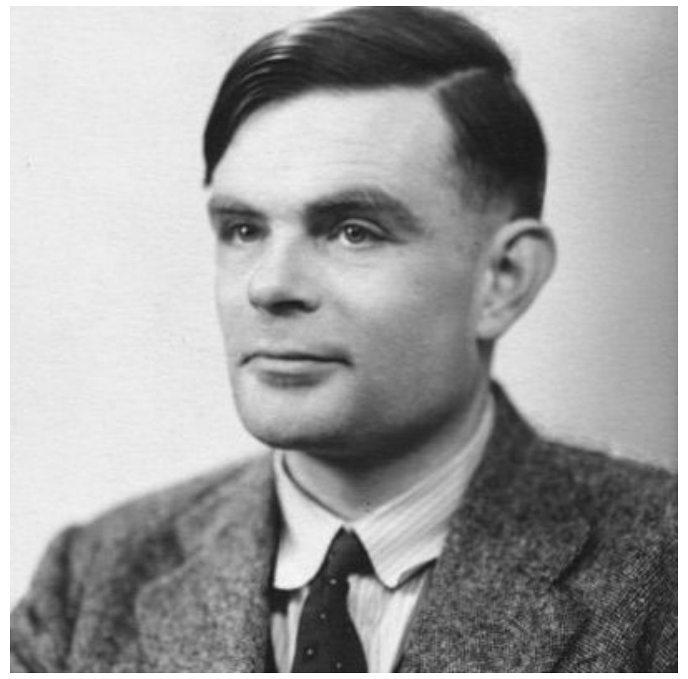
GOFAI



- **Genesis** The vision of early robotics
  - Wizard of Oz (1900)
  - Metropolis (1927)
  - Alan Turing (1940-54) “The Turing test”
    - Thinking machine fooling human
  - 1956: Logic Theorist Program (Newell, Shaw, H. Simon, McCarty and Minsky)
    - Symbolic reasoning, deduction, inference
  - Golden age follows with many successes … The optimistic phase!
    - General problem solving
    - Games that AI can play and beat humans at
    - 1966: Eliza the Psychotherapy ‘bot’ (https://web.njit.edu/~ronkowit/eliza.html)

GOFAI





- **Genesis** The vision of early robotics
  - Wizard of Oz (1900)
  - Metropolis (1927)
  - Alan Turing (1940-54) “The Turing test”
    - Thinking machine fooling human
  - 1956: Logic Theorist Program (Newell, Shaw, H. Simon, McCarty and Minsky)
    - Symbolic reasoning, deduction, inference
  - Golden age follows with many successes … The optimistic phase!
    - General problem solving
    - Games that AI can play and beat humans at
    - 1966: Eliza the Psychotherapy ‘bot’ (https://web.njit.edu/~ronkowit/eliza.html)

Weak AI

Timeline

ARTIFICIAL INTELLIGENCE TIMELINE

1938-1946: Golden Age of Science Fiction

1949: Machester Mark 1, the first storted program computer, is invented

1950: *Can Machines Think?* -Alan Turing

1955: *Logic Theorists*, the first AI program, is invented

1956: Dartmouth Summer Research Project on Artificial Intelligence

1963: DARPA funds AI at MIT

1965: Moore's Law

1968: “By the year 2001 we will have machines with intelligence that matched or exceeded human's” -Arthur Clarke and Steve Kubrik

1970: “From 3-8 years we will have a machine with the general intelligence of a human being” -M. Minsky

1980: Edward Feigenbaum introduces expert systems

1982: Japan's Fifth Generation Computer Project

1986: Navlab, the first autonomous car, is built by Carnegie Melon

1997: Deep Blue defeats Gary Kasparov in chess

1997: First publicly availa speech recongition softwa developed by *Dragon Sys*

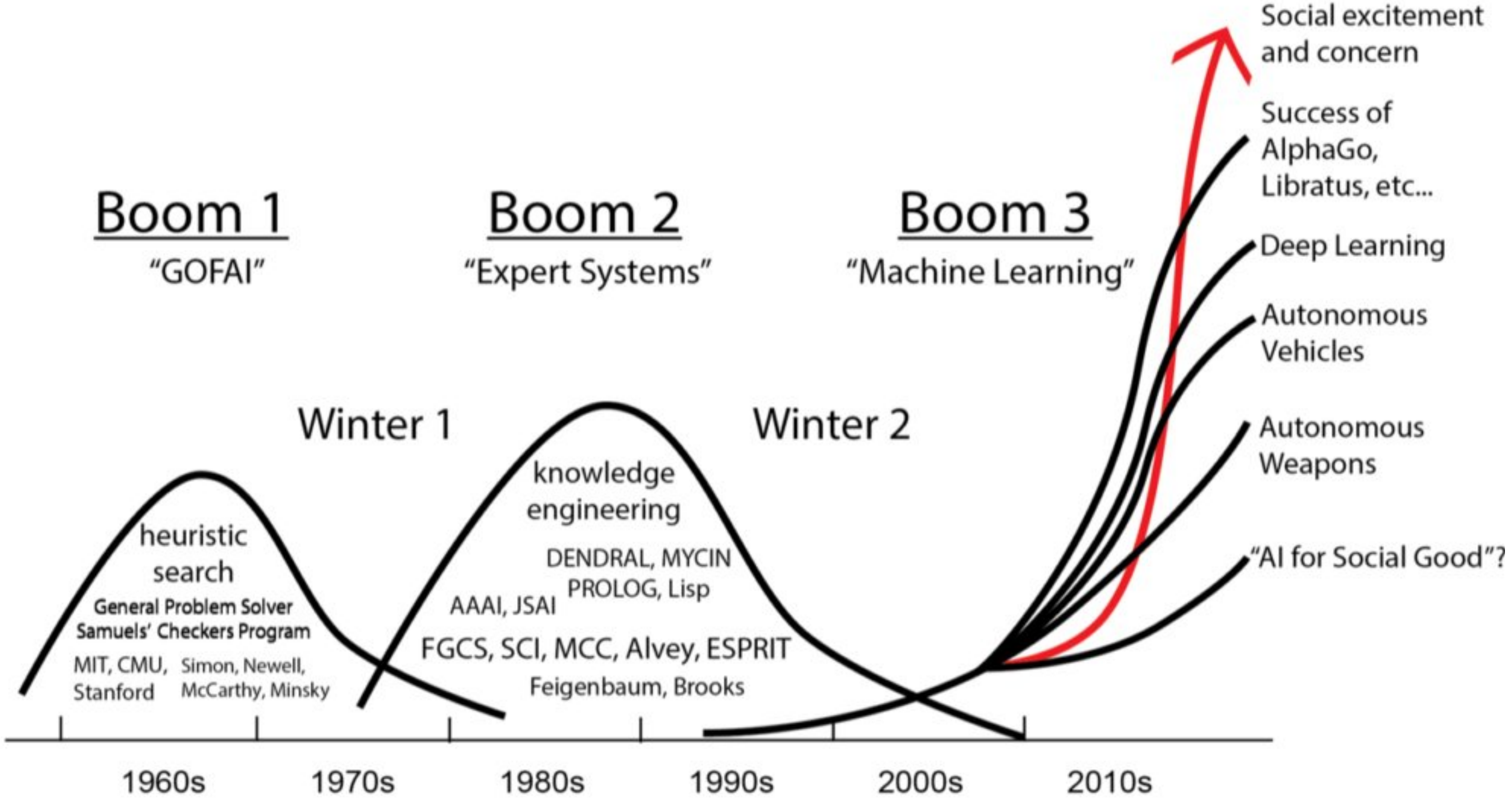
https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Seasons...

- What comes after summer and fall ?

Seasons...

- What comes after summer and fall ?



AI Winter

- First winter 1974-1980
  - Promised too much, delivered too little.
  - Improved understanding of computing and its limits
  - Intractability (S. Cook’s 1971 result on what cannot be done)
  - Computers not powerful enough…

Modern Era

- Expert systems lead the charge
  - Domain specific expert knowledge embedded in query-response system
  - Not general purpose, but effective!
  - Started during the winter phase.
  - Fundamentally: symbolic reasoning and sophisticated deduction
  - The rise of LISP (US) and Prolog (France, started 1972, A. Colmerauer)
  - MYCIN,Caduceus, XCON,...
- The return of Neural Networks (invented 1958, shelved during the winter)
  - From Perceptron to Artificial Neural Networks

Eureka!

- Key realization
  - Intelligence is not just deduction
  - It's a by-product of having
    - Lots of experiential data!
    - Lots of knowledge
- Bottom line
  - It's no longer just about having some experts knowledge + inference
  - It's about pulling from a lot of data
  - Neural networks get a second shot (thanks to faster computers too)

AI Winter

- Second winter : 1987 - 1993
  - Collapse of companies making dedicated hardware (e.g., LISP machine)
  - Rise of general purpose computers: IBM PC & Macs from Apple
  - Fifth generation project (started in Japan) fails to deliver
    - Promised too much... delivered too little [2nd time!]
  - AI oriented companies going bust. Commercially, AI is not (yet) viable

Generative Era (1995 - ...)

- Enablers
  - Computer power is exploding
    - Deep Blue about 10^7 times faster than first chess computer from 1951)
  - Real commercial adoption
  - The Rise of knowledge system (data-centric!)
  - Clean slate of techniques (SVM, Search, Clustering, Classification, xNN,...)
  - Deep learning
  - Generative Models
  - Big victories visible to all...

Big Victories

Big Victories

1997 : Deep blue beats Kasparov

Deep Blue
IBM chess computer

Garry Kasparov
World Chess Champion

Big Victories

1997 : Deep blue beats Kasparov

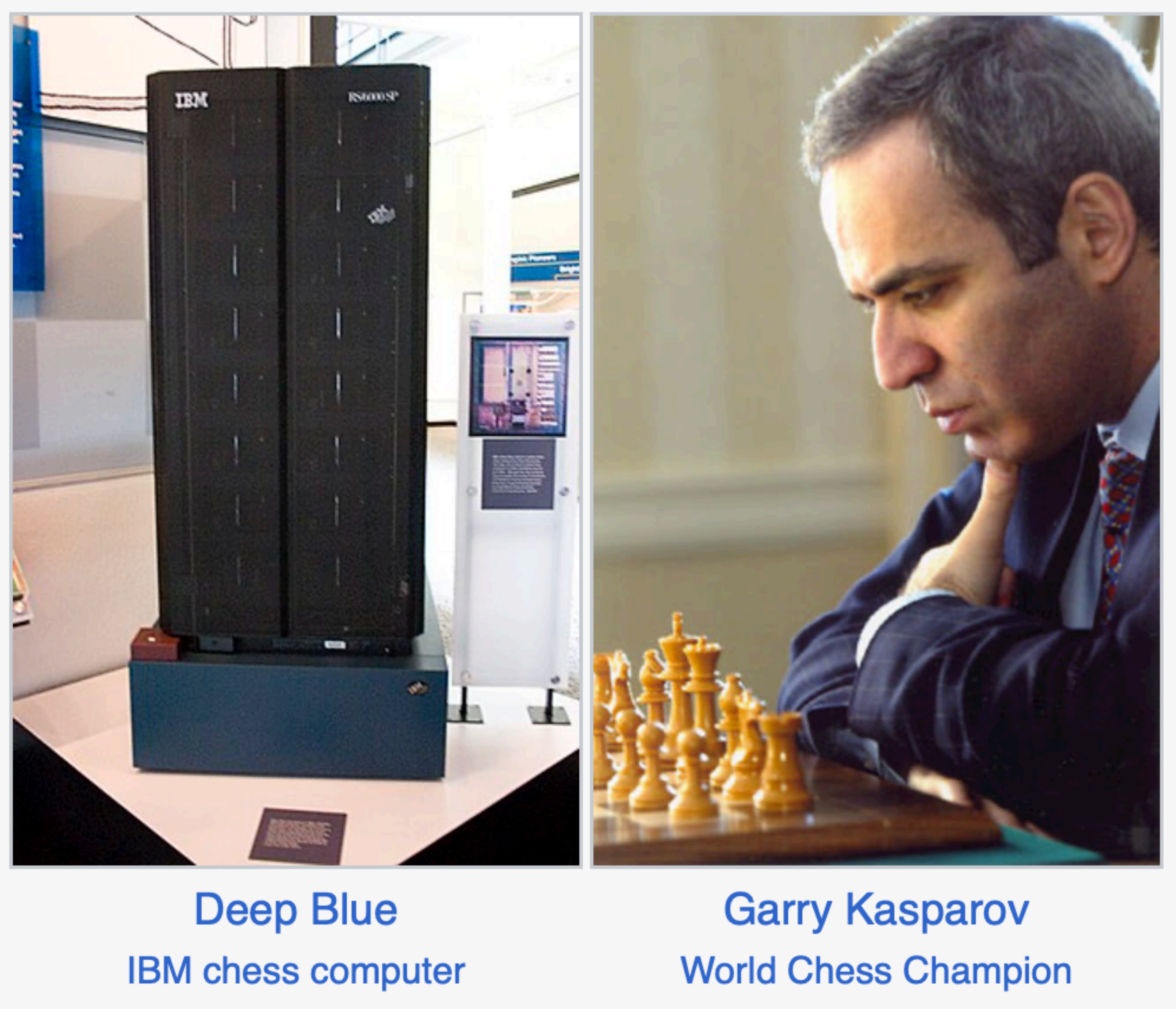
Deep Blue
IBM chess computer

Garry Kasparov
World Chess Champion

2005 : Stanford's autonomous Car

Big Victories

1997 : Deep blue beats Kasparov



Deep Blue
IBM chess computer

Garry Kasparov
World Chess Champion

2011 : Watson wins Jeopardy

2005 : Stanford's autonomous Car

Big Victories

1997 : Deep blue beats Kasparov

Deep Blue

IBM chess computer

Garry Kasparov

World Chess Champion

2005 : Stanford's autonomous Car

2011 : Watson wins Jeopardy

2016 : AlphaGo beats Sedol Lee

Big Victories

1997 : Deep blue beats Kasparov

Deep Blue
IBM chess computer

Garry Kasparov
World Chess Champion

2005 : Stanford's autonomous Car

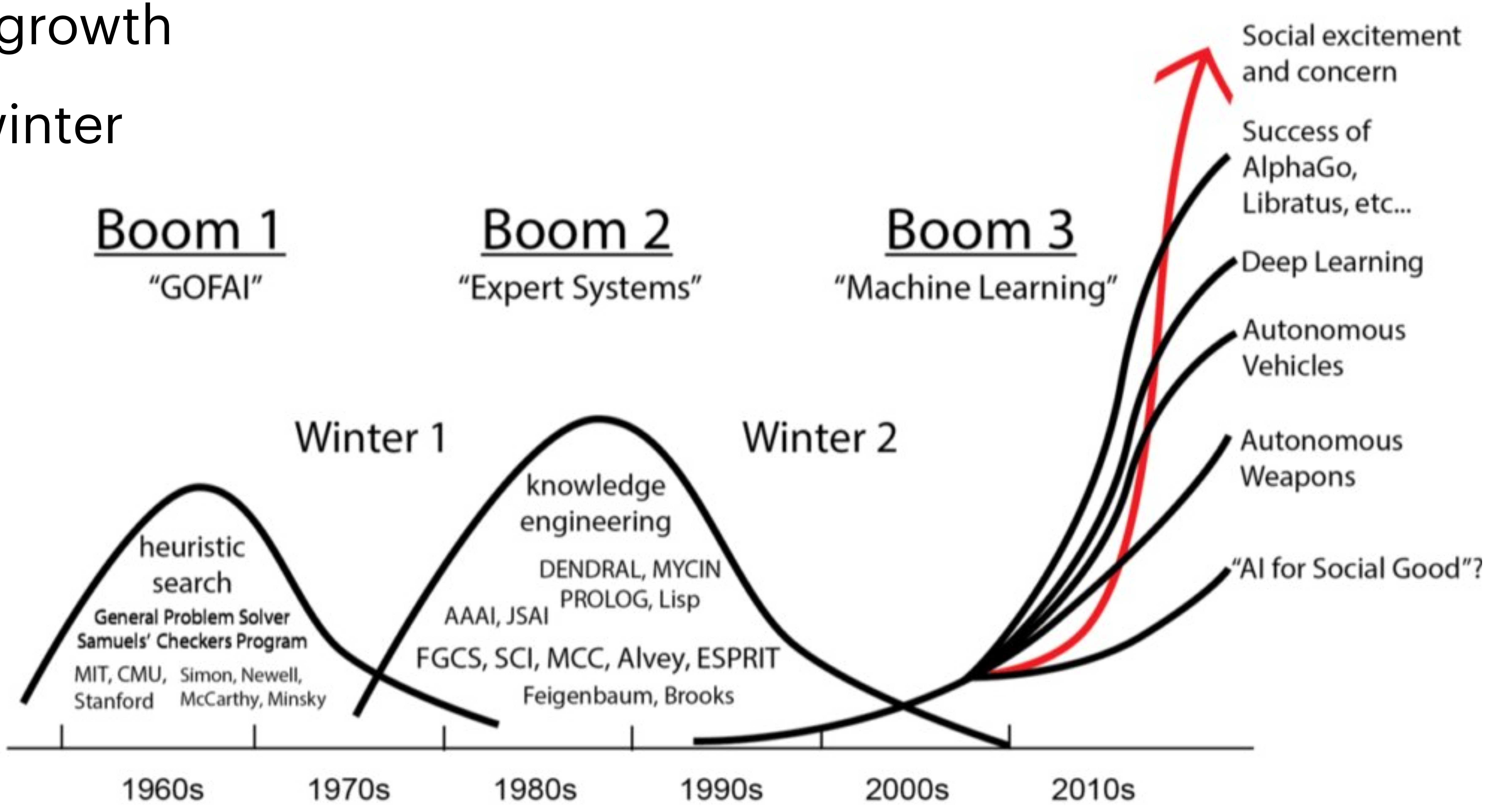
2011 : Watson wins Jeopardy

2016 : AlphaGo beats Sedol Lee

2022 : ChatGPT blows up

Ahead ?

- Remember the curve…
  - Huge growth
  - Or a winter

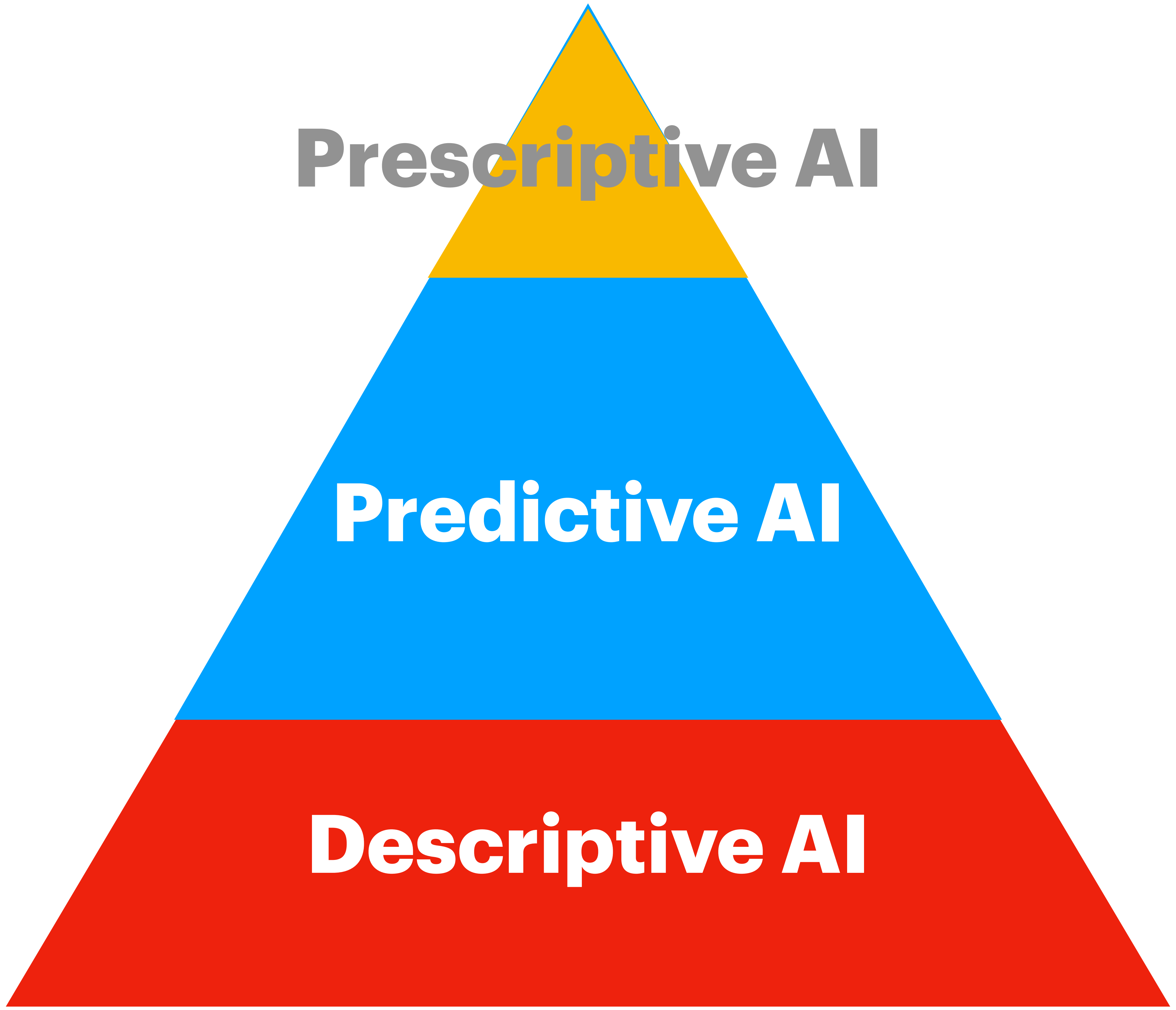


So...

What **is** AI ?

Pyramidal View : where it applies

- **Descriptive** AI
  - Classic: Describe systems, data, rules, knowledge (KR), in some <u>structured</u> way.
- **Predictive** AI
  - Predict what will happen based on what we know
  - This is where all the excitement is!
- **Prescriptive** AI
  - Decide how to respond to the prediction and make decisions



Technical View : what it applies

- GOFAI
  - Rule-based reasoning: if-then-else

Technical View : what it applies

- GOFAI
  - Rule-based reasoning: if-then-else

- **The Good**
  - Small program with *all* the rules…
  - No need for data
  - Easy to understand and explain
- **The Bad**
  - Typically domain-specific (not general at all)
  - If rules are “missing” we are toast
  - What if it is hard to formally express some rules ?

Technical View : what it applies

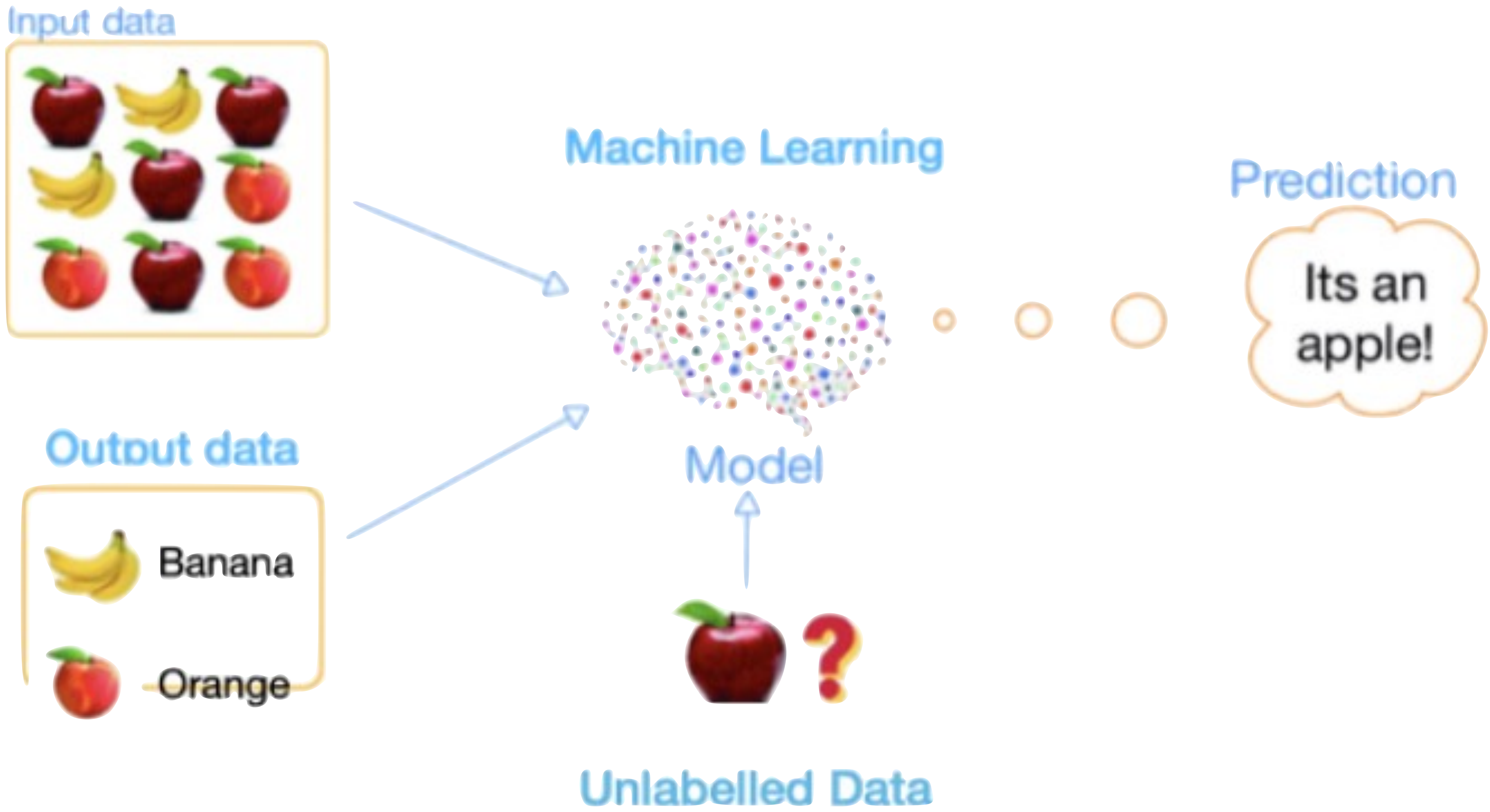
- GOFAI
  - Rule-based reasoning: if-then-else

- **The Good**
  - Small program with *all* the rules…
  - No need for data
  - Easy to understand and explain
- **The Bad**
  - Typically domain-specific (not general at all)
  - If rules are “missing” we are toast
  - What if it is hard to formally express some rules ?

AI Imitates intelligence using rules. Fake it ‘till you make it!

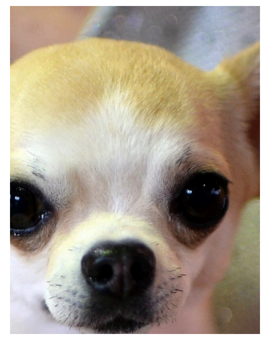
Technical View : what it applies

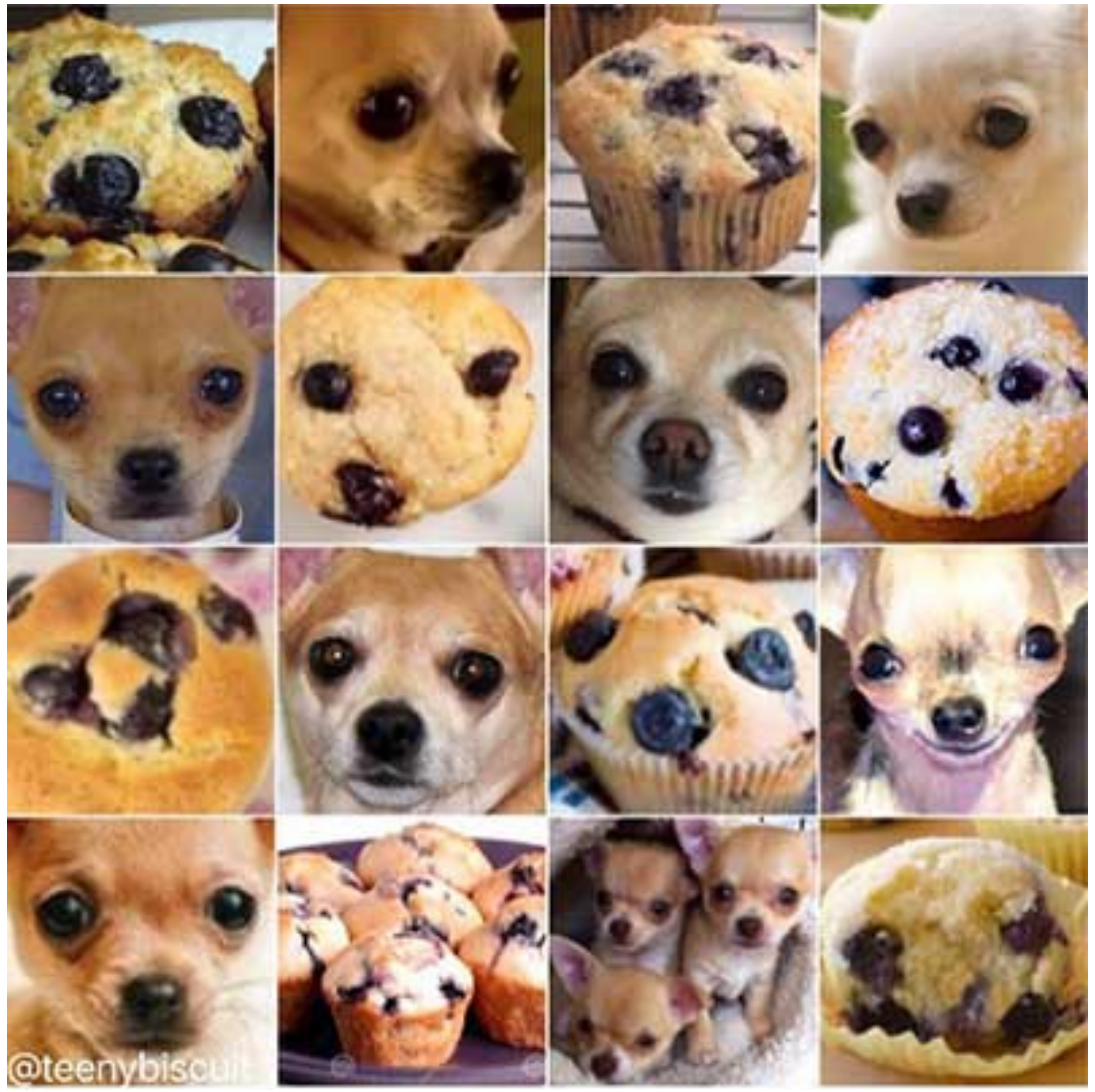
- GOFAI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised** Learning with labels
  - SVM, Neural Networks Machine Learning, Deep Learning

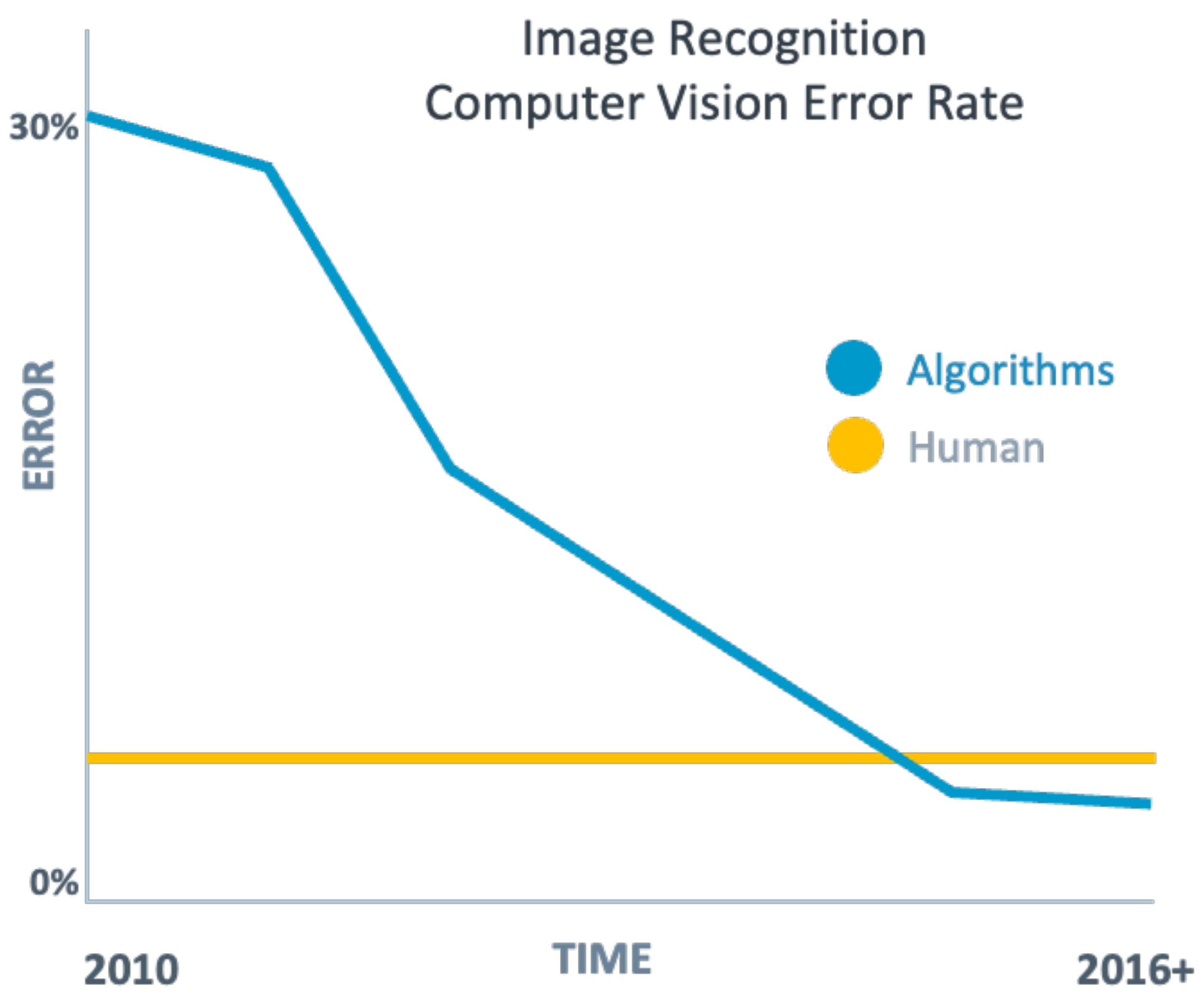


App Example : A Chiuahua or a Muffin ?

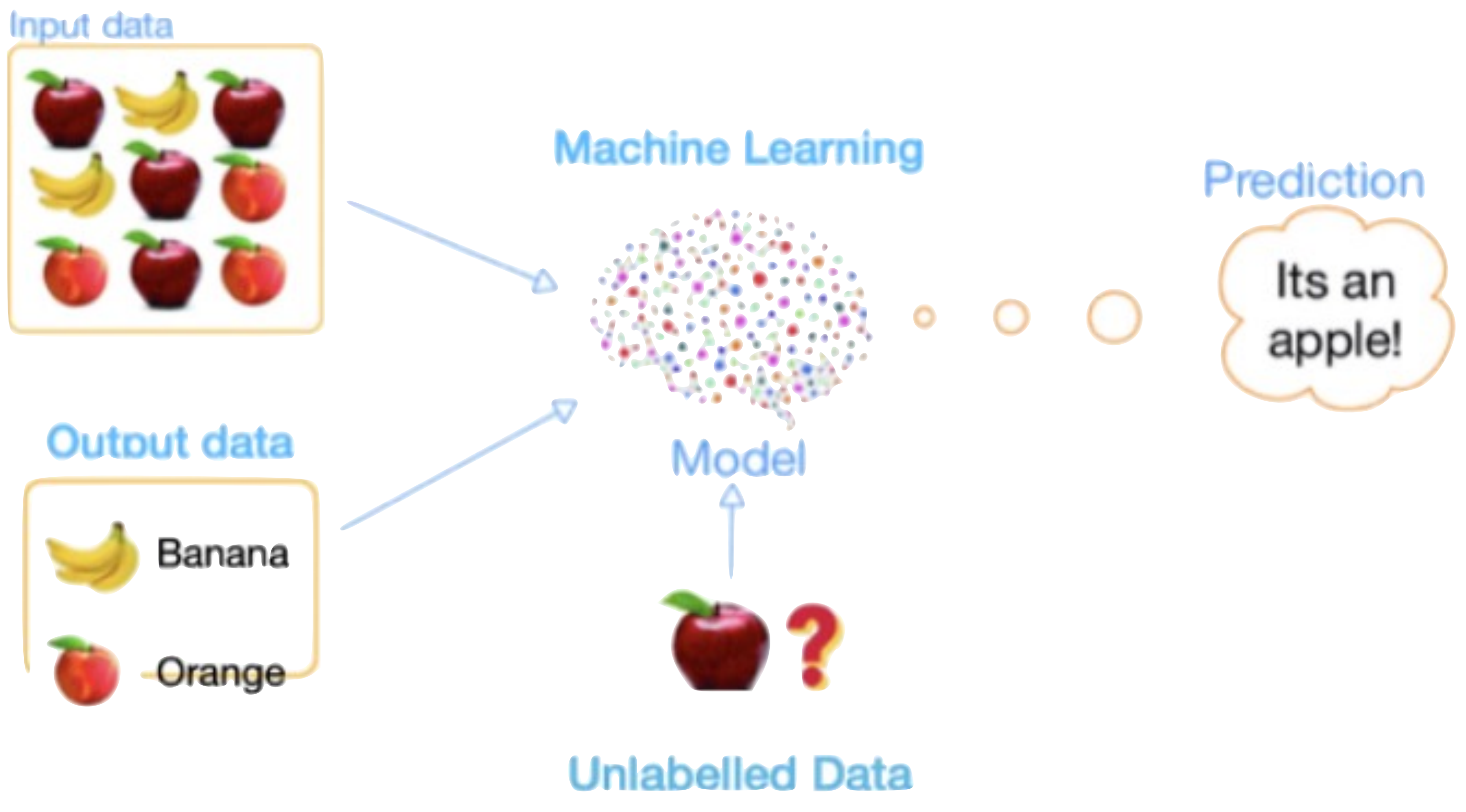
- Training
  - Labeled images of Chiuahua and muffins…
- Prediction
  - What is this ?







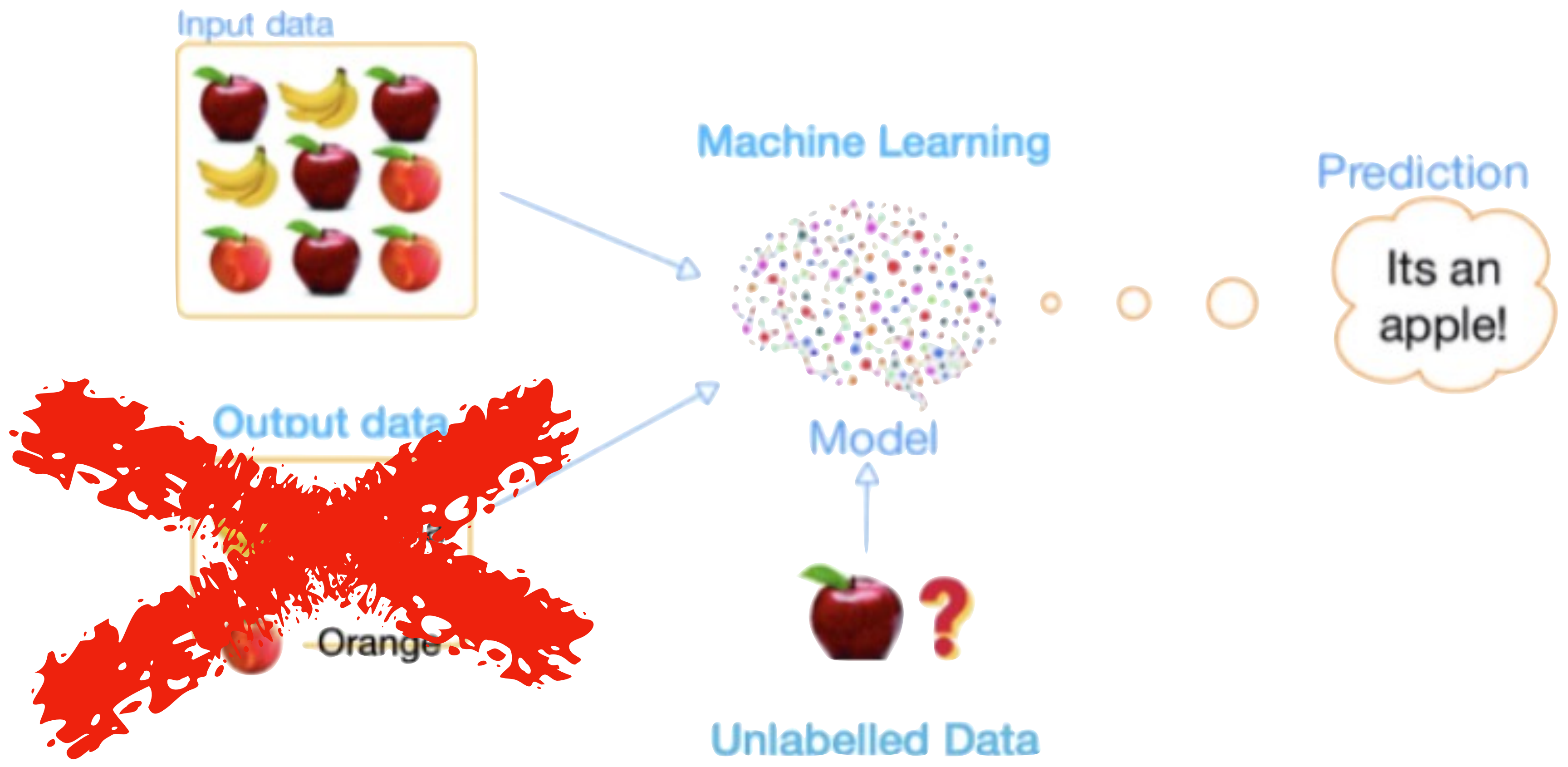
Technical View : what it applies



- GOFAI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised** Learning with labels
  - SVM, Neural Networks Machine Learning, Deep Learning
  - **Unsupervised** Learning without labels
  - Clustering, Anomaly Detection, ...

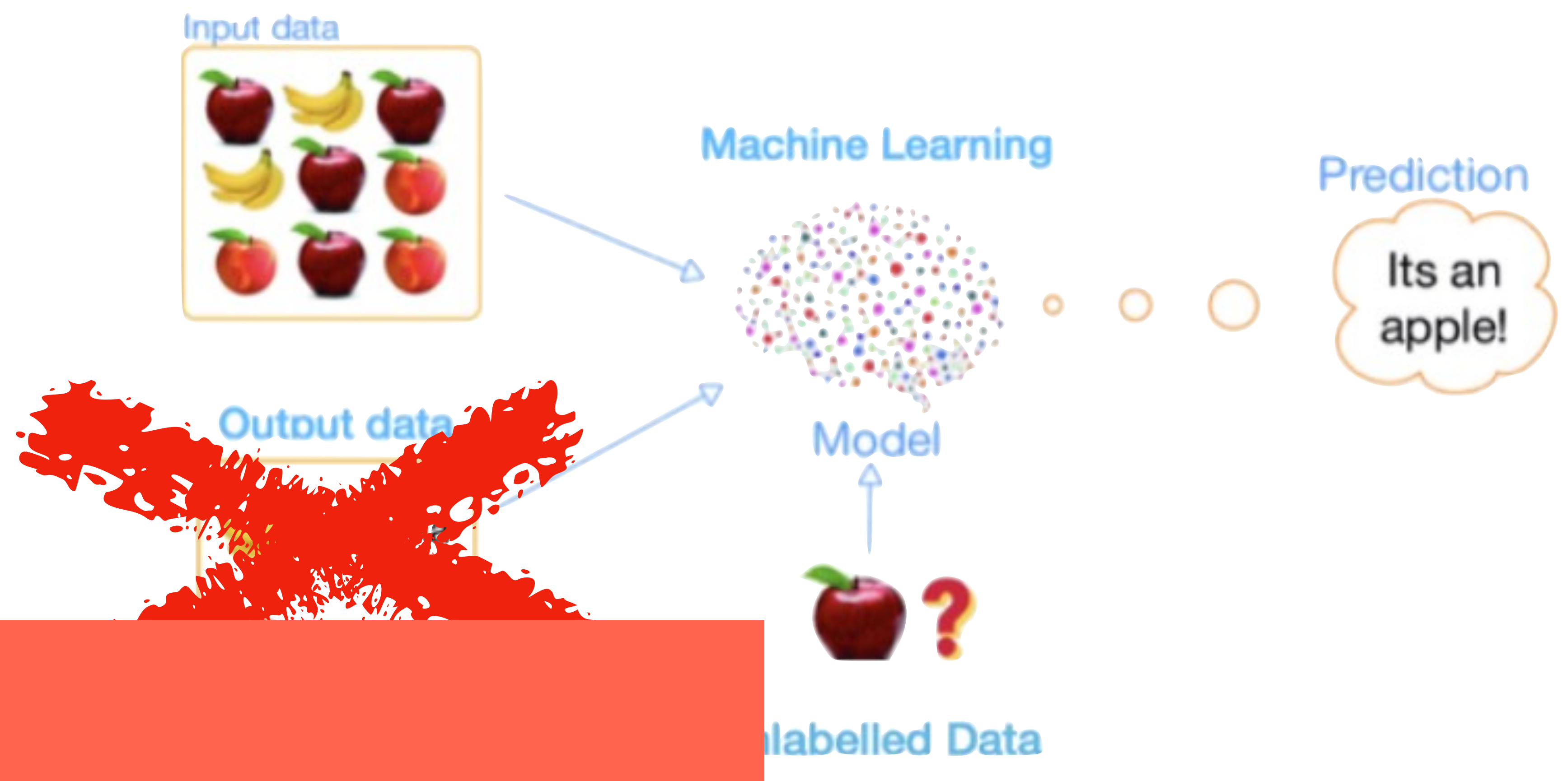
Technical View : what it applies

- GOFAI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised** Learning with labels
  - SVM, Neural Networks Machine Learning, Deep Learning
  - **Unsupervised** Learning without labels
  - Clustering, Anomaly Detection, ...



Technical View : what it applies

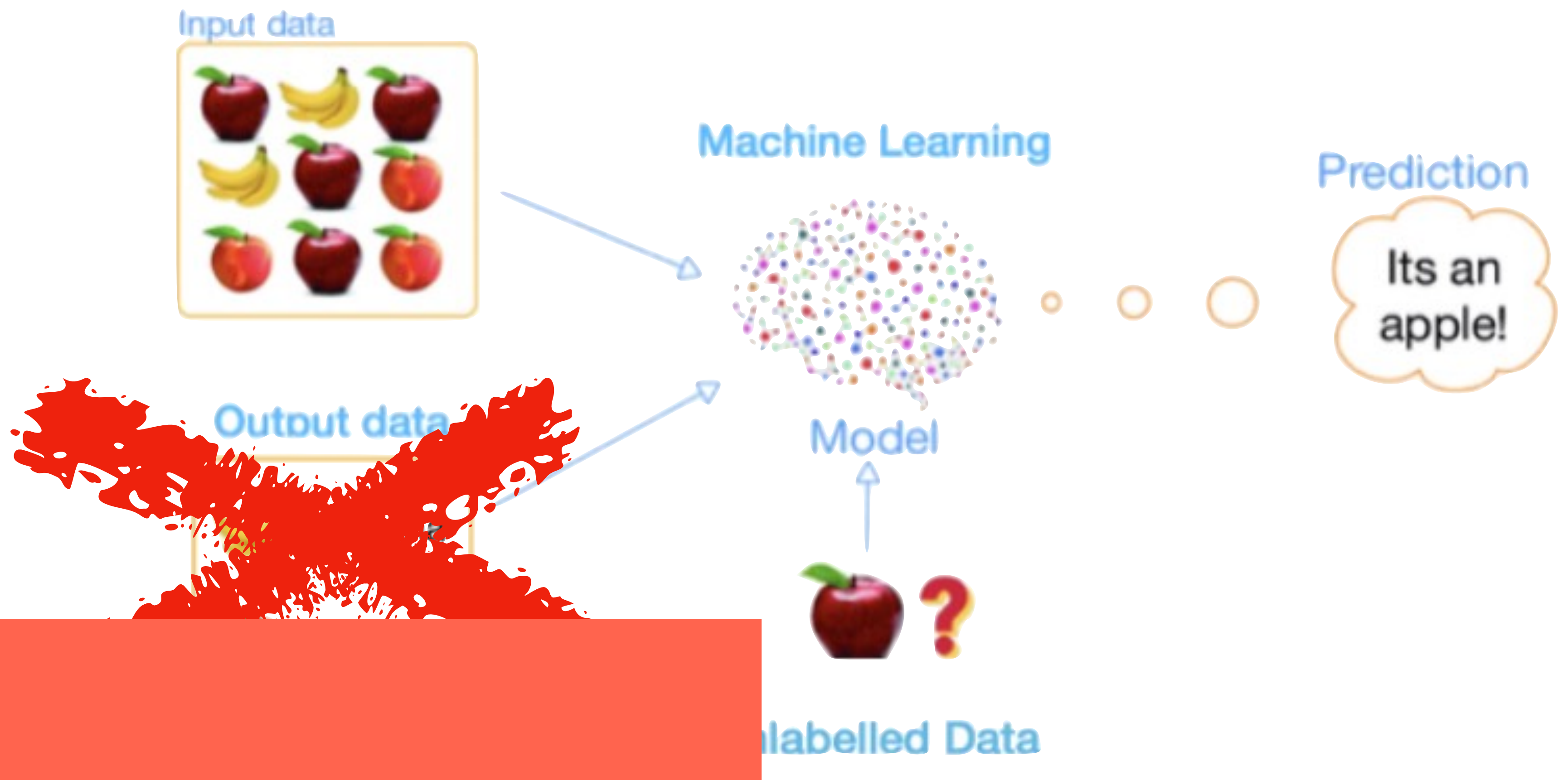
- GOFAI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised**
  - SVM, Neura
    Learning, D
  - **Unsupervis**
  - Clustering,



- **The Good**
  - Programs/techniques are more generic / reusable
  - Applies to more domain
  - Improves at it gets more data
- **The Bad**
  - It needs **LOTS** of data
  - Explainability can become an issue

Technical View : what it applies

- GOFAI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised**
  - SVM, Neura
    Learning, D
  - **Unsupervis**
  - Clustering,

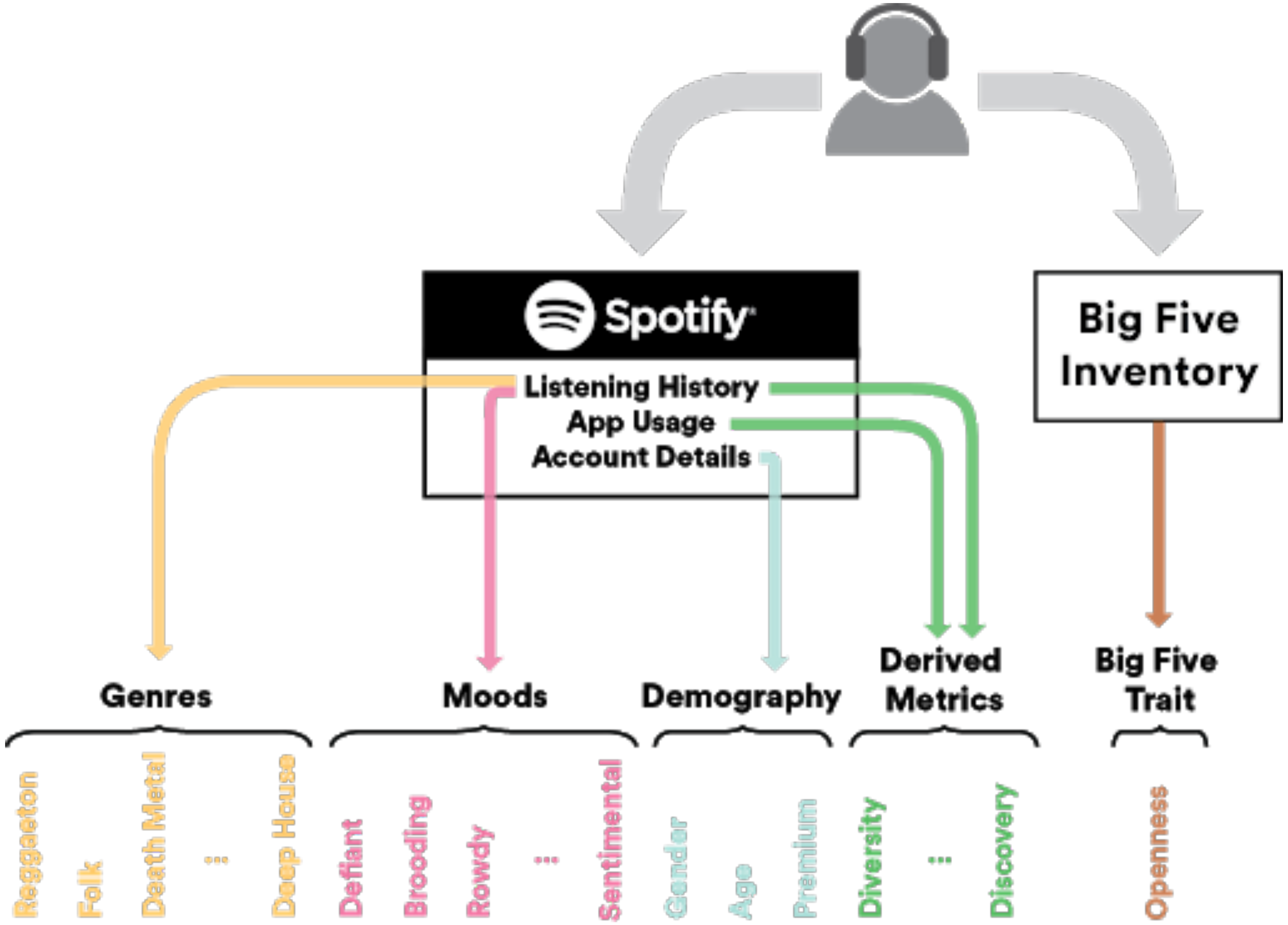


- **The Good**
  - Programs/techniques are more generic / reusable
  - Applies to more domain
  - Improves at it gets more data
- **The Bad**
  - It needs **LOTS** of data
  - Explainability can become an issue

AI recognizes similar “contexts” and extrapolate responses. It is not creative.

App Example : Recommendation Systems

- Training
  - Listening habits of millions of users
- Prediction
  - Clustering of people into “groups”
  - Suggest the right music to keep them listening (mixtape/radio)
  - Lots of similar gigs for marketing purposes

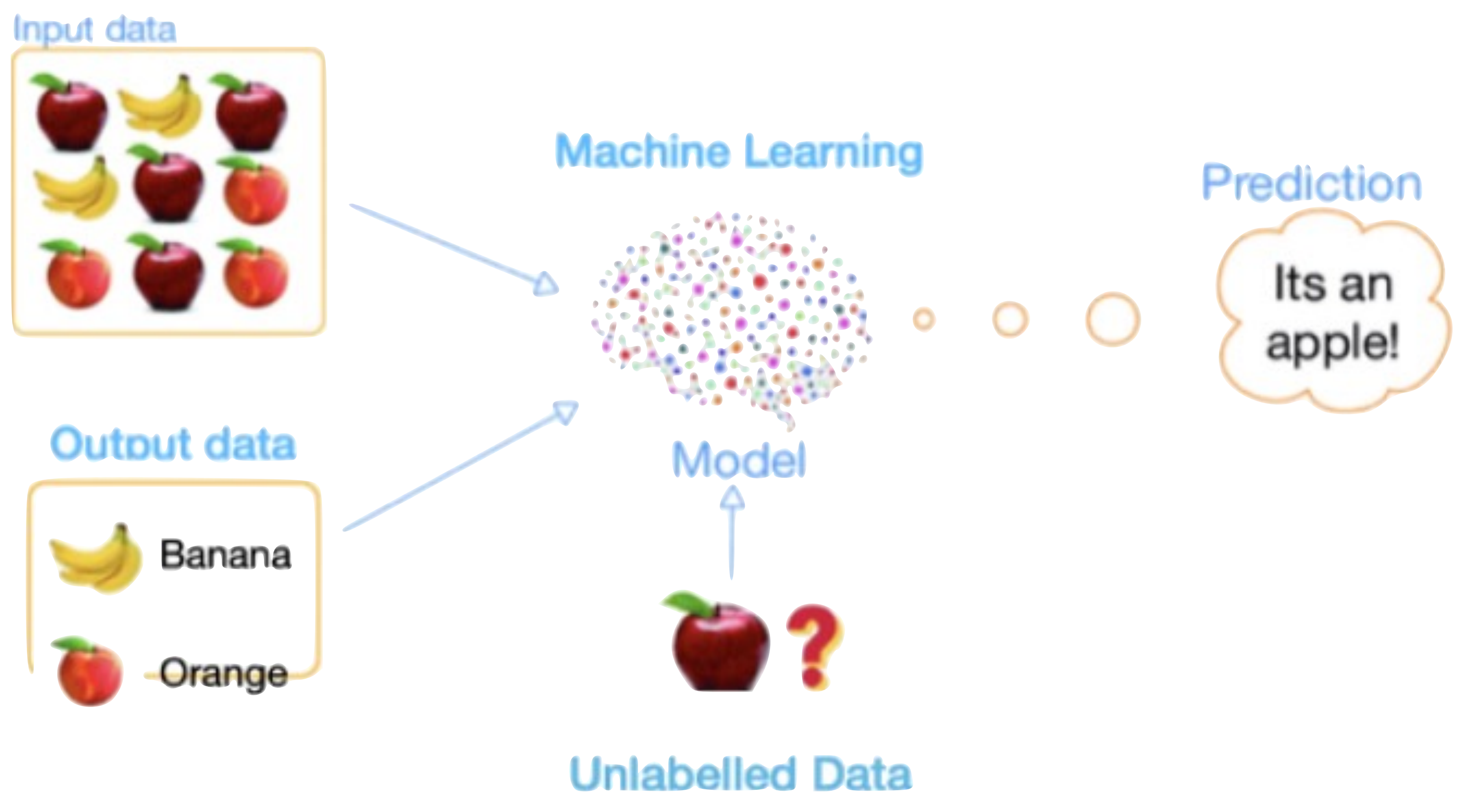


App Example : HR processing

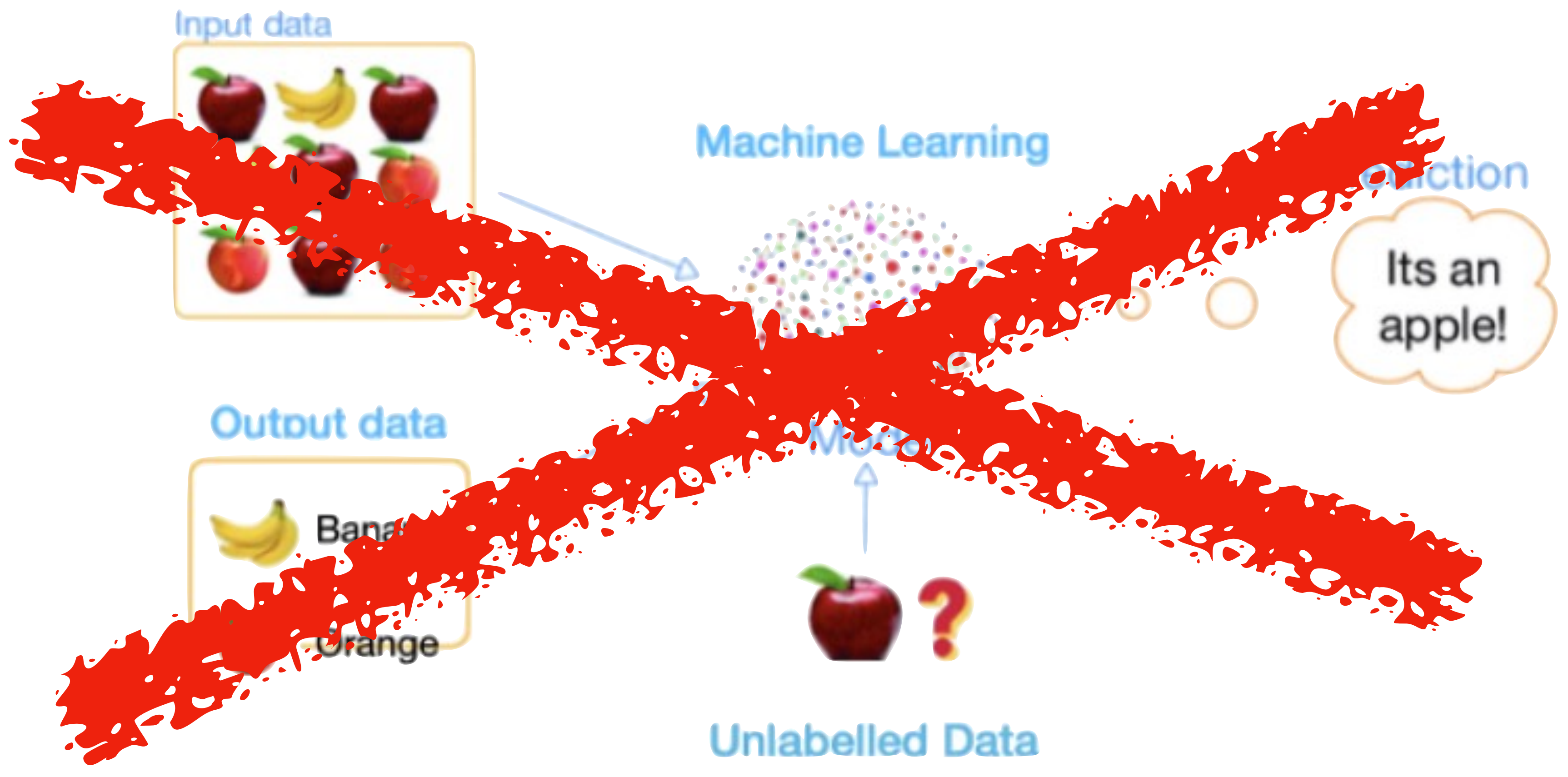
- Training
  - Million(s) of vitae from employees
- Prediction
  - Given a vitae of a candidate...
  - Predict what category he falls in, for which positions,....

Technical View : what it applies

- Classic AI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised** Learning with labels
  - SVM, Neural Networks Machine Learning, Deep Learning
  - **Unsupervised** Learning without labels
  - Clustering, Anomaly Detection, …
- Generative AI
  - Language-based Large Language Models

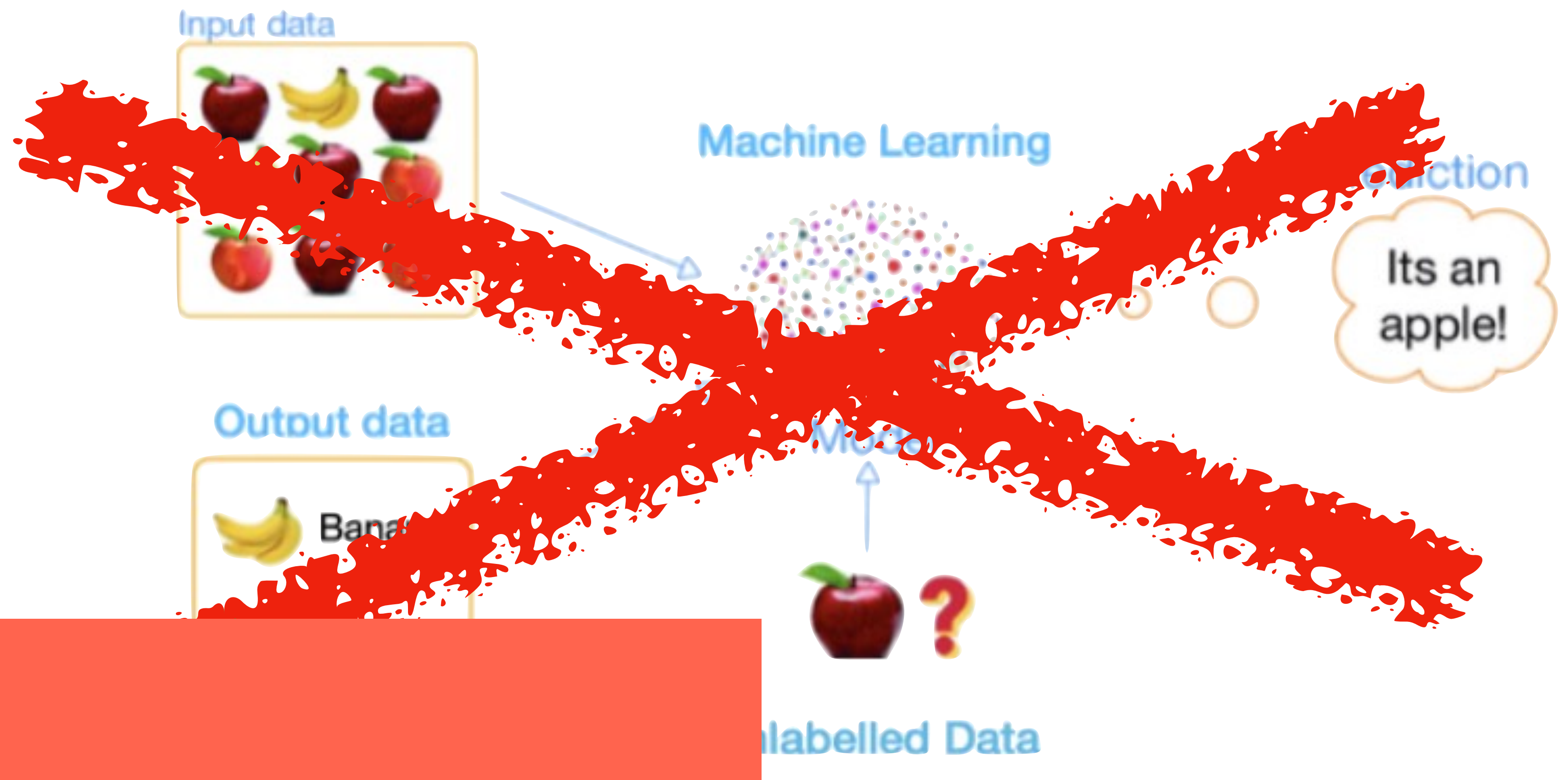


Technical View : what it applies



- Classic AI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised** Learning with labels
  - SVM, Neural Networks Machine Learning, Deep Learning
  - **Unsupervised** Learning without labels
  - Clustering, Anomaly Detection, …
- Generative AI
  - Language-based Large Language Models

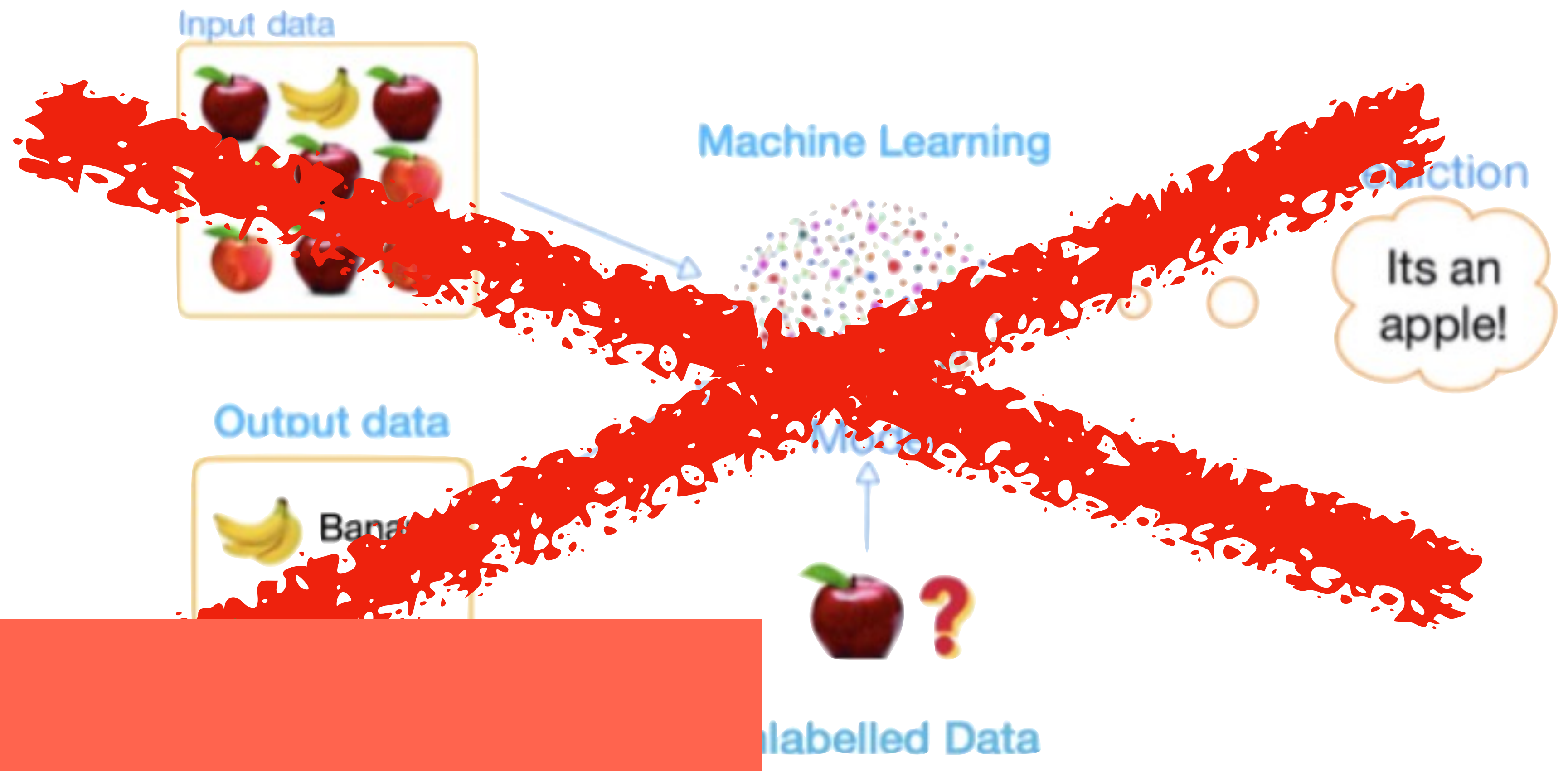
Technical View : what it applies



- Classic AI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised**
  - SVM, Neura
    Learning, D
  - **Unsupervis**
  - Clustering,
- Generative AI
  - Language-based Large Language Models

- **The Good**
  - It’s quite effective. We use language a *lot!*
  - Applies broadly
  - Very versatile
- **The Bad**
  - Explainability is a full-blown issue
  - It’s totally **opaque**
  - It needs huge amounts of data to work!

Technical View : what it applies

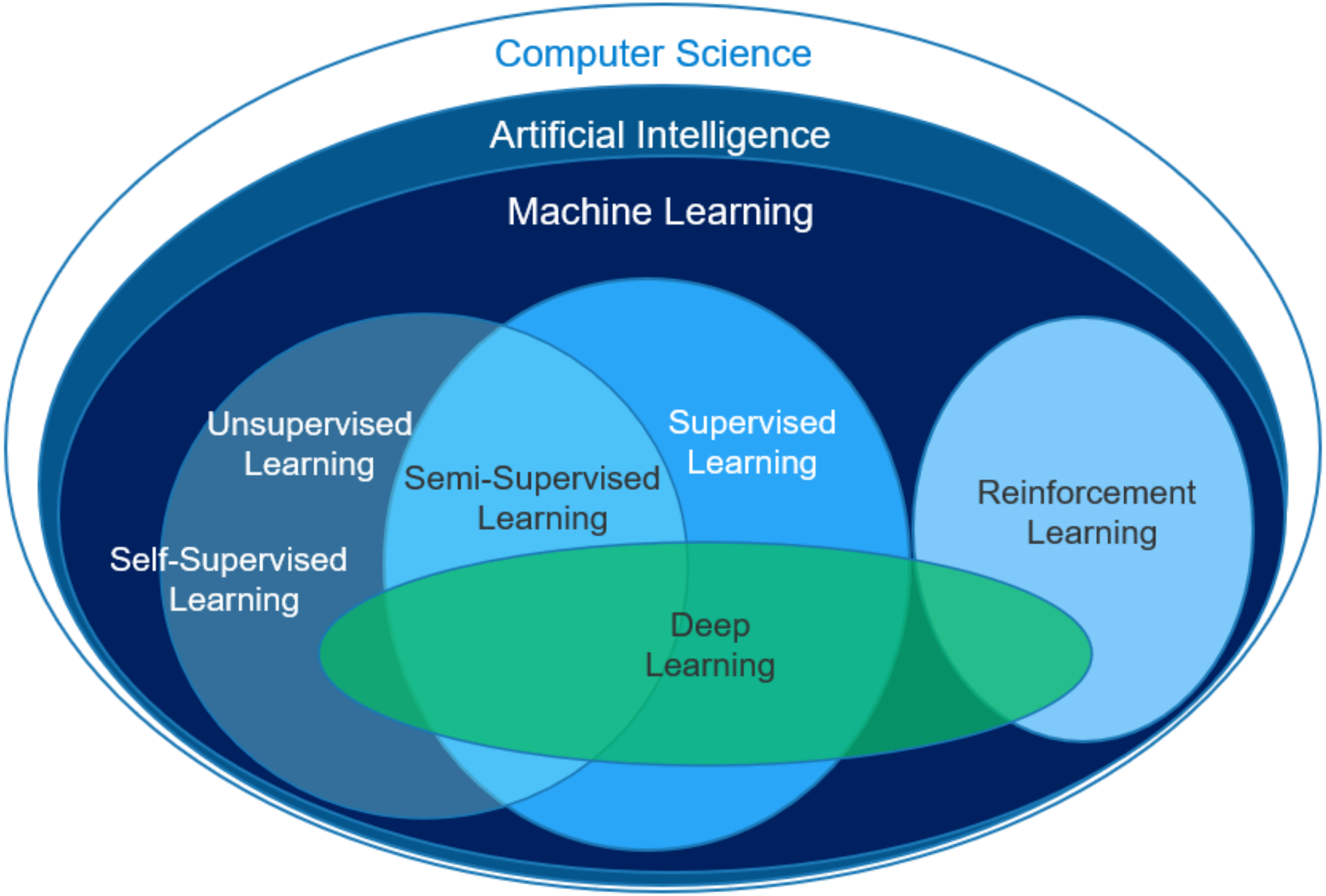


- Classic AI
  - Rule-based reasoning: if-then-else
- Modern AI
  - **Supervised**
  - SVM, Neura
    Learning, D
  - **Unsupervis**
  - Clustering,
- Generative AI
  - La

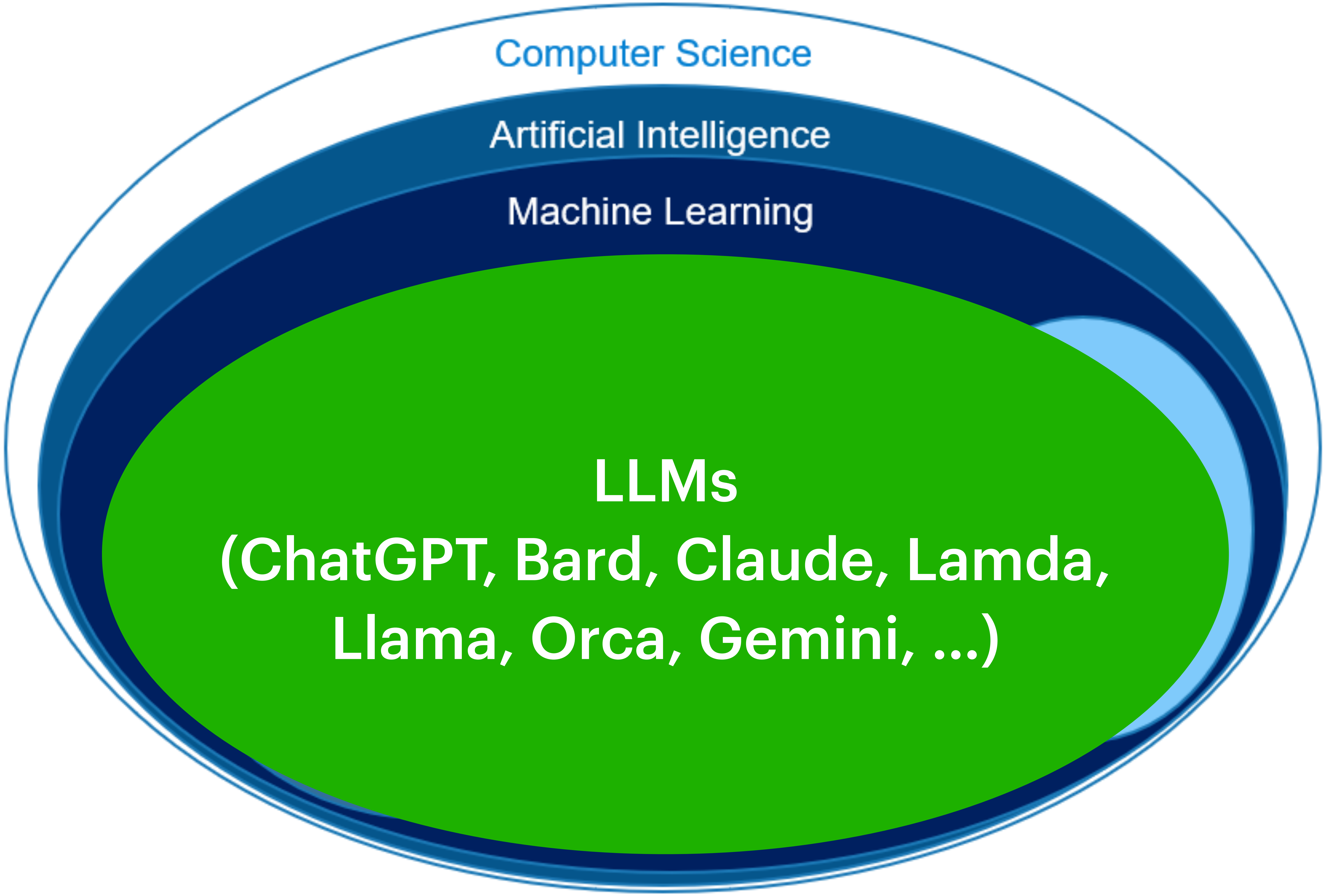
- **The Good**
  - It's quite effective. We use language a *lot!*
  - Applies broadly
  - Very versatile
- **The Bad**
  - Explainability is a full-blown issue
  - It's totally **opaque**
  - It needs huge amounts of data to work!

LLM apply to many application domain. They generate was is **likely.** No sense of "meaning"

Venn Diagram



Venn Diagram



A Side Box

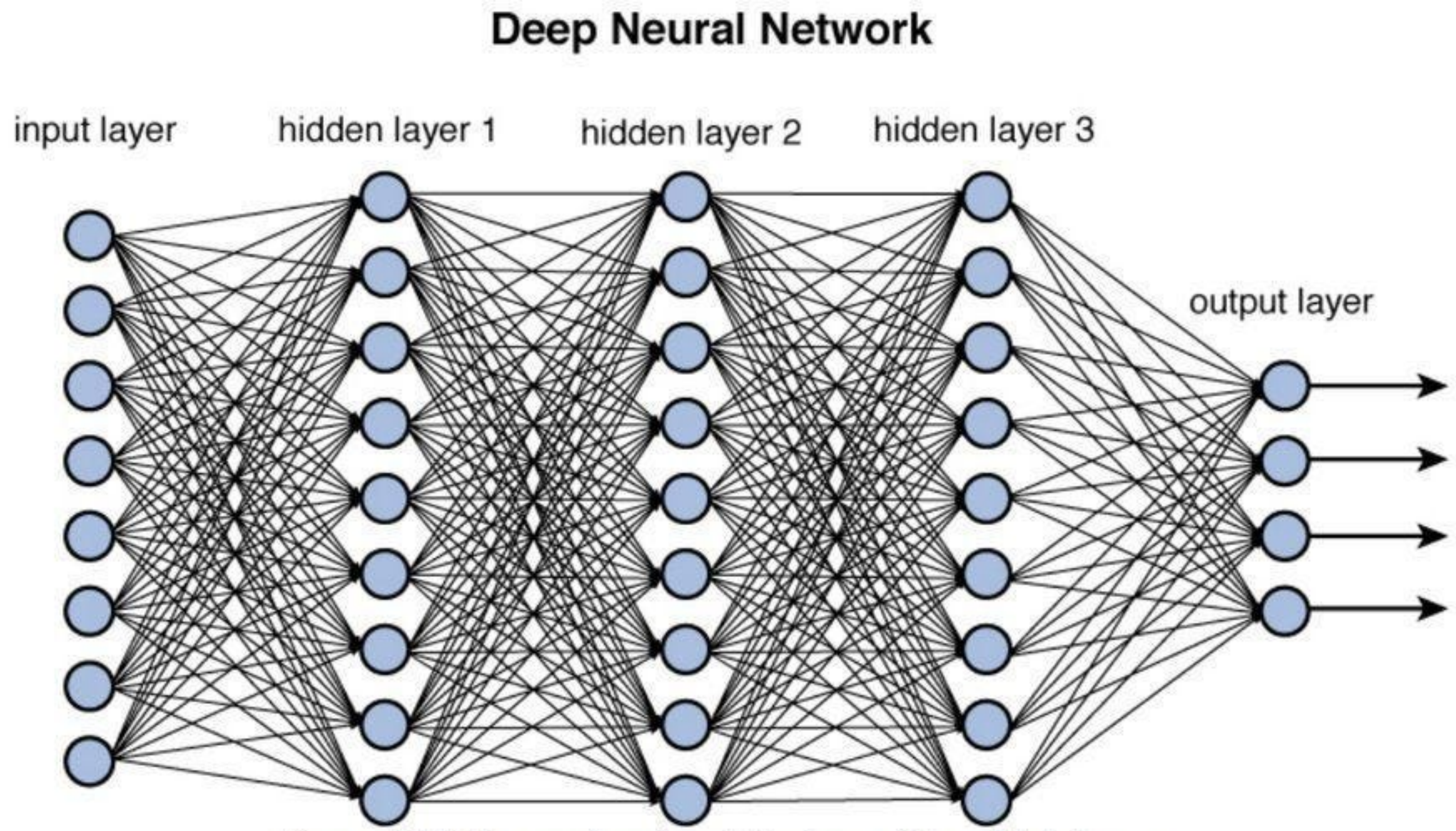
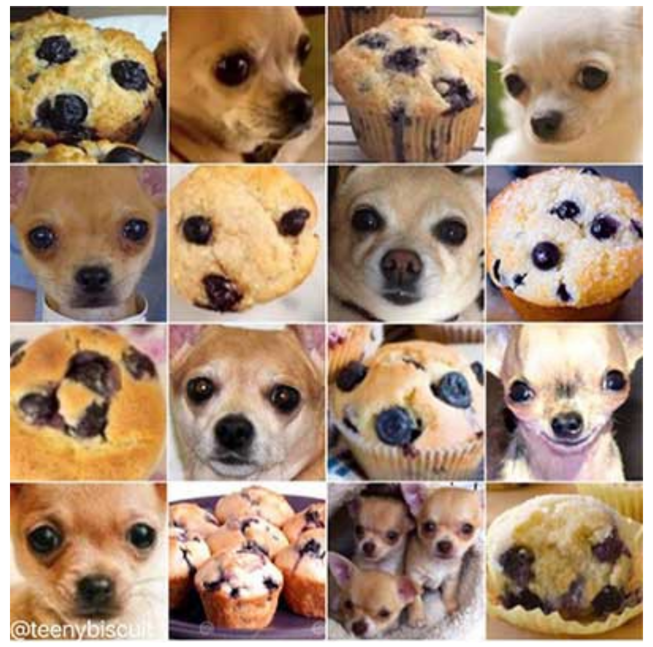


Figure 12.2 Deep network architecture with multiple layers.

- What is Deep learning?
  - A “Network” of “neurons”
  - Taking numerical inputs (encoding)
  - Producing numerical outputs
  - Trained on lots of PAIRS
    - (INPUT,OUTPUT)
    - Training changes the weights to minimize **loss**
  - To drive better predictions!
    - When given an ~INPUT.... it produces ~OUTPUT (or something close!)

The Chiuahua Problem!



- Deep Learning limitations
  - Can only handle the inputs it was defined on!
  - Sometimes susceptible to over-fitting
    - Good “Recall” ability
    - Bad “Prediction” ability
    - i.e., it fails to generalize
  - Trained networks do not apply to other domains
    - e.g., recognizing bananas and plantains
  - Susceptible to “attacks”
    - Feed an image obvious to a human with some near-invisible tweaks....
    - .... and it gets predictions completely wrong.

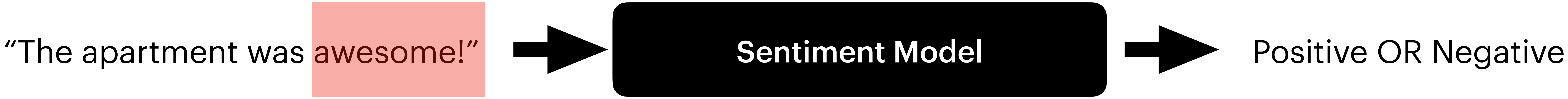
Deep Learning Applied to Natural Language

- Example: sentiment analysis
  - Goal
    - Get a review of a B&B place. Find out if it is a positive review!
    - It's unsupervised learning. Nobody is going to label reviews....
  - Training
    - Input: Lots of reviews of B&B places, locale to visit (as text!)
    - Output: Positive or Negative

"The apartment was awesome!"
Sentiment Model
Positive OR Negative

Deep Learning Applied to Natural Language

- Example: sentiment analysis
  - Goal
    - Get a review of a B&B place. Find out if it is a positive review!
    - It’s unsupervised learning. Nobody is going to label reviews....
  - Training
    - Input: Lots of reviews of B&B places, locale to visit (as text!)
    - Output: Positive or Negative



“The apartment was awesome!” → **Sentiment Model** → Positive OR Negative

Deep Learning Applied to Natural Language

- Example: sentiment analysis
  - Goal
    - Get a review of a B&B place. Find out if it is a positive review!
    - It’s unsupervised learning. Nobody is going to label reviews….
  - Training
    - Input: Lots of reviev t (as text!)
    - Output: Positive or Negative

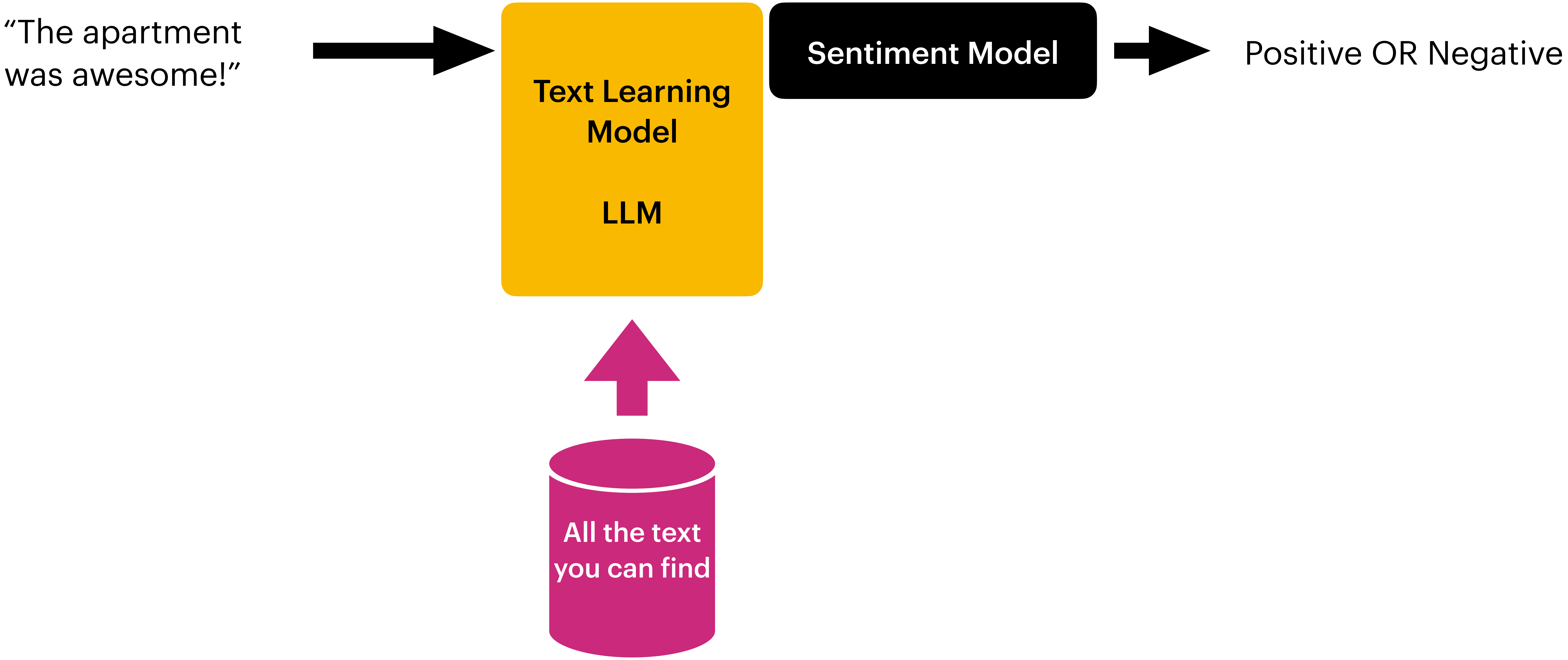
There are big hints in the context!

“The apartment was awesome!”

Sentiment Model

Positive OR Negative

Shifting...



“The apartment was awesome!”

Text Learning Model

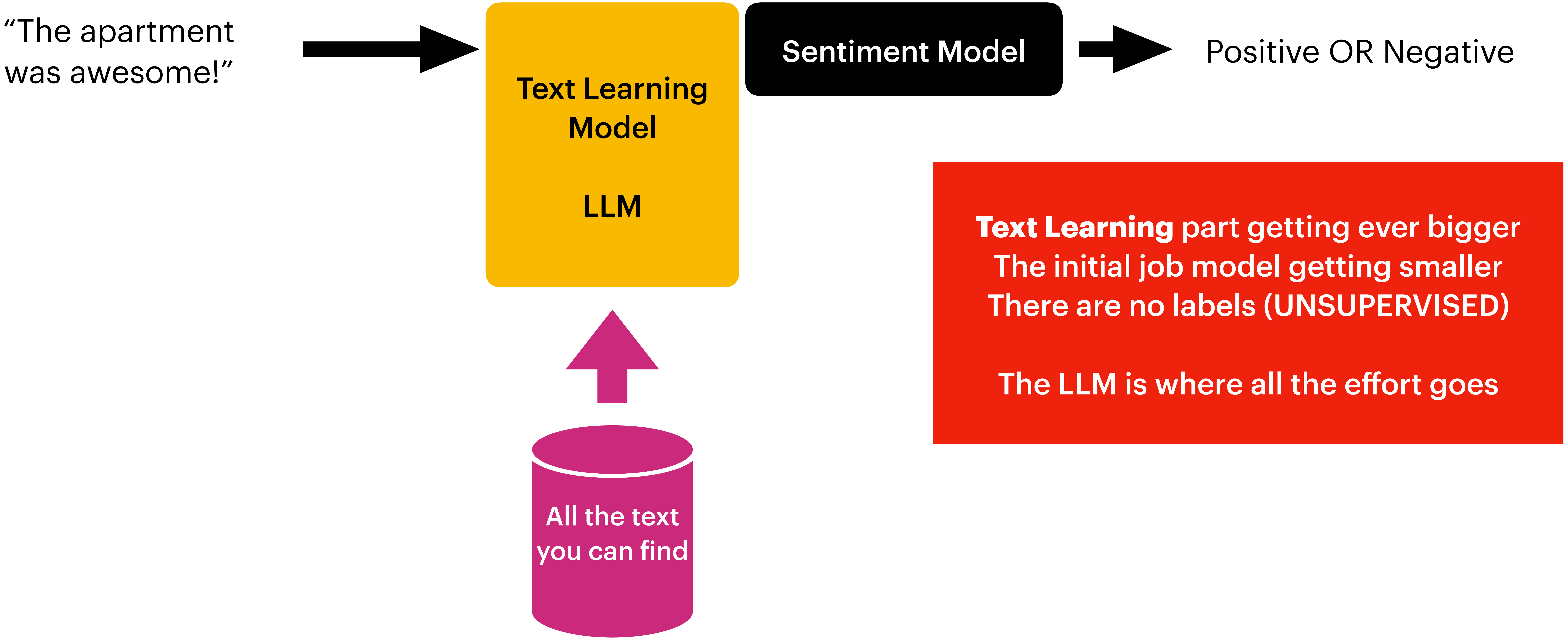
LLM

Sentiment Model

Positive OR Negative

All the text you can find

Shifting…



“The apartment was awesome!”

Text Learning Model

LLM

Sentiment Model

Positive OR Negative

All the text you can find

Text Learning part getting ever bigger
The initial job model getting smaller
There are no labels (UNSUPERVISED)

The LLM is where all the effort goes

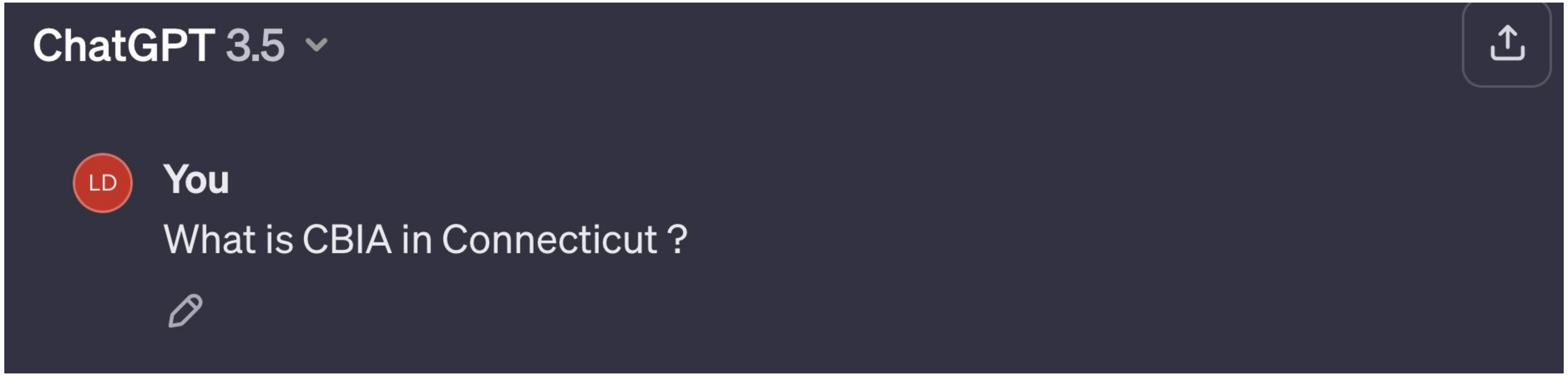
LLMs

Their mission…
Their secrets…
Their Achilles' heel…

Their mission

Demo Time!

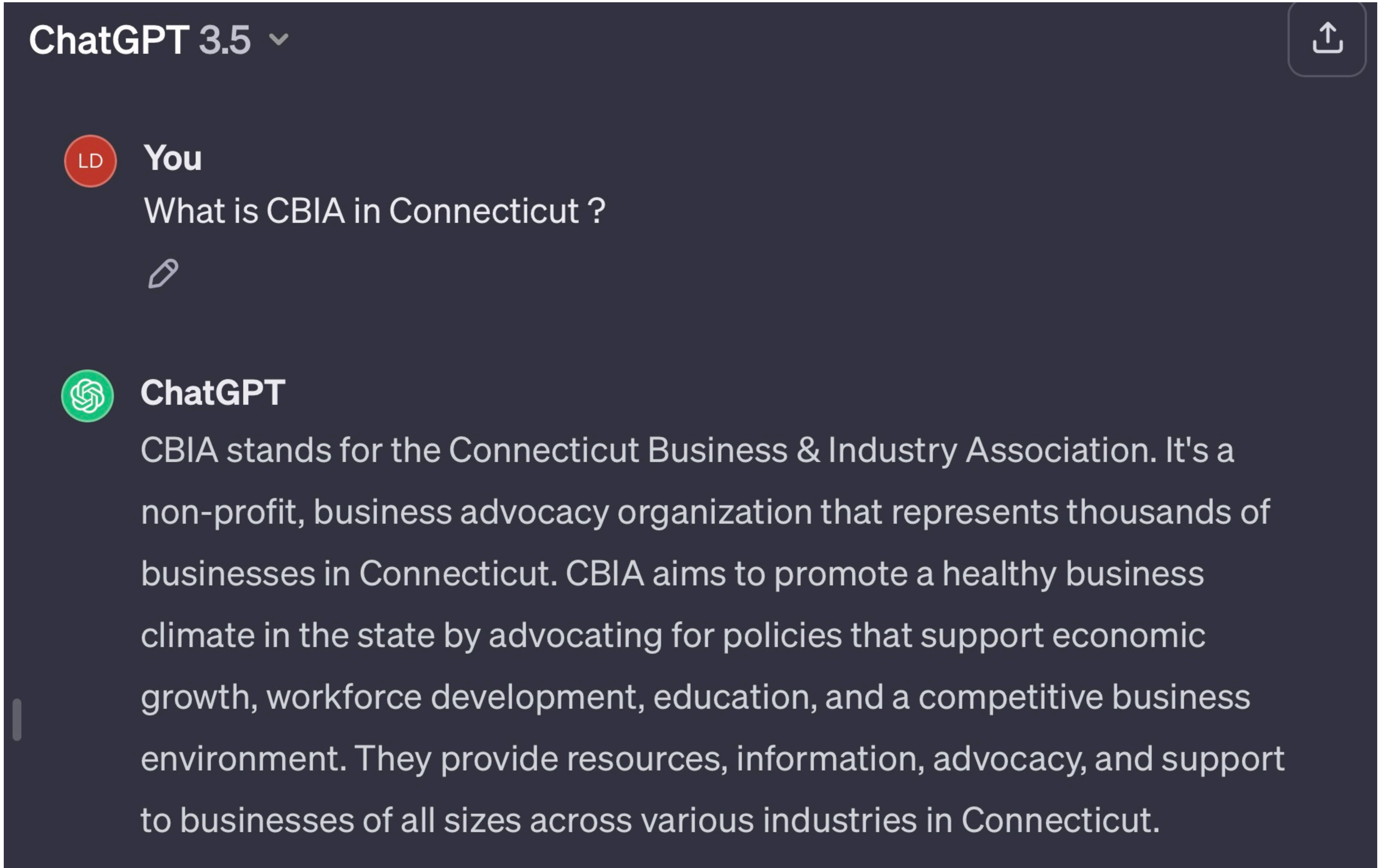
Demo Time!



You

What is CBIA in Connecticut ?

Demo Time!



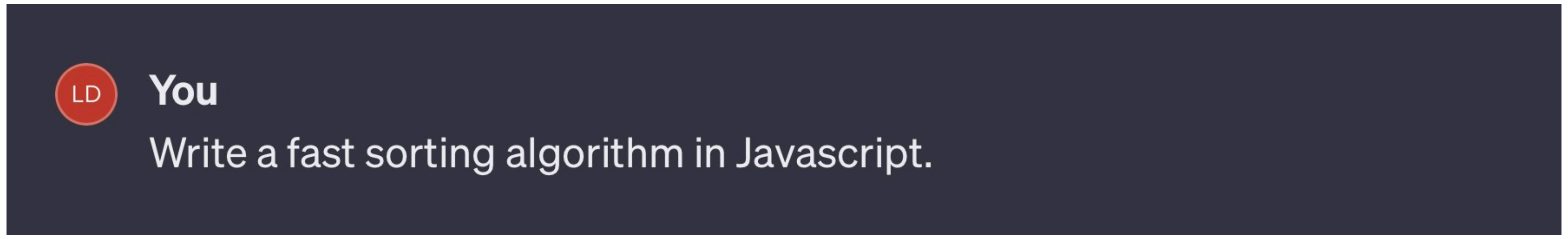
You

What is CBIA in Connecticut ?

ChatGPT

CBIA stands for the Connecticut Business & Industry Association. It's a non-profit, business advocacy organization that represents thousands of businesses in Connecticut. CBIA aims to promote a healthy business climate in the state by advocating for policies that support economic growth, workforce development, education, and a competitive business environment. They provide resources, information, advocacy, and support to businesses of all sizes across various industries in Connecticut.

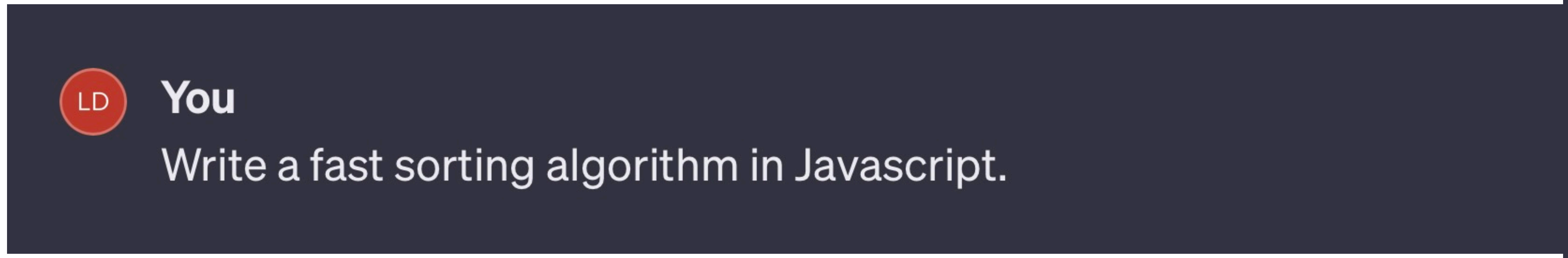
Coding...



LD **You**

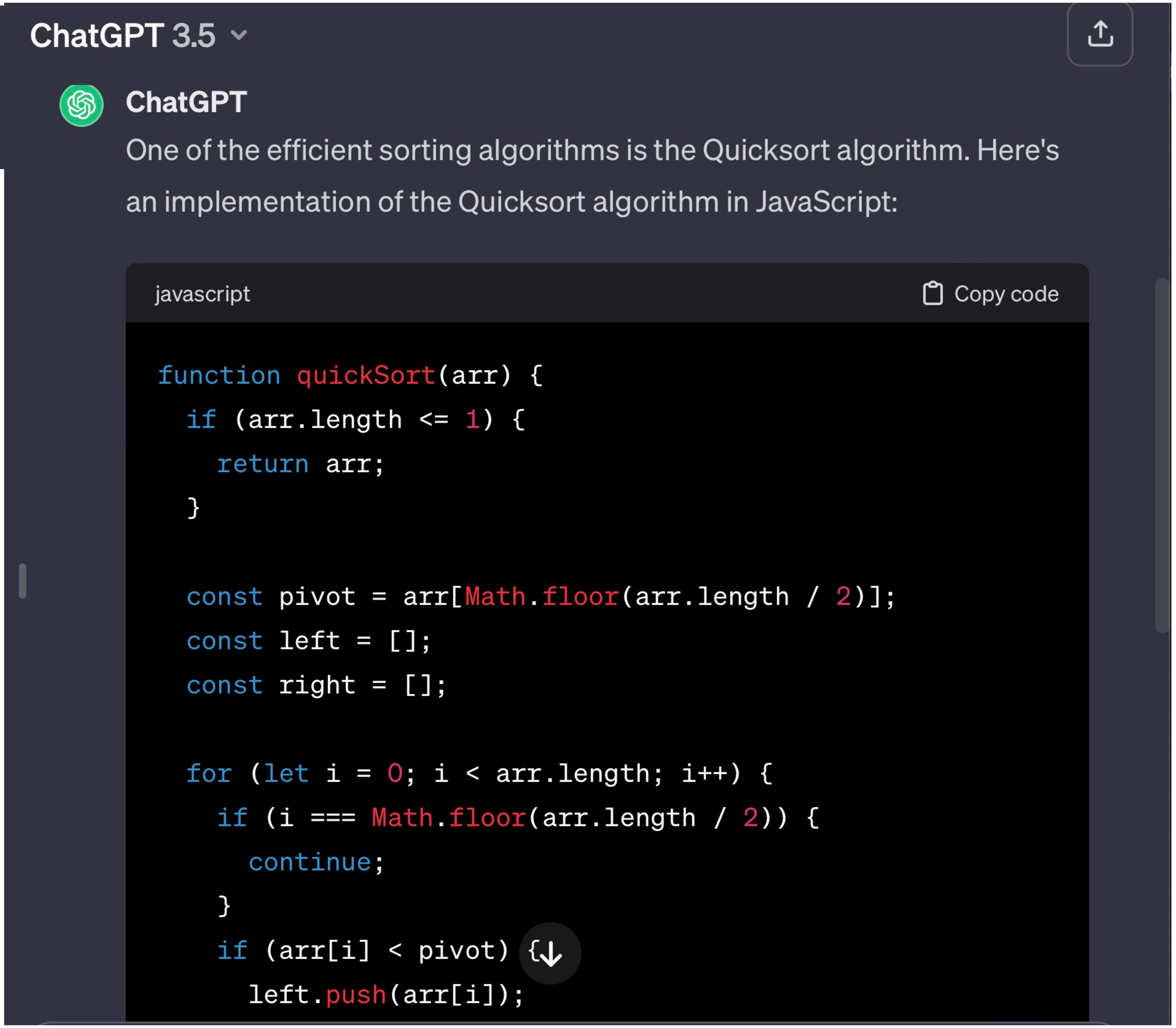
Write a fast sorting algorithm in Javascript.

Coding...



You

Write a fast sorting algorithm in Javascript.

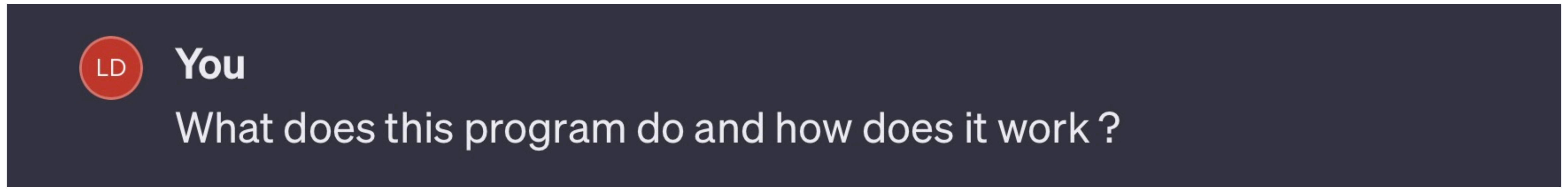


ChatGPT

One of the efficient sorting algorithms is the Quicksort algorithm. Here's an implementation of the Quicksort algorithm in JavaScript:

```javascript
function quickSort(arr) {
  if (arr.length <= 1) {
    return arr;
  }

  const pivot = arr[Math.floor(arr.length / 2)];
  const left = [];
  const right = [];

  for (let i = 0; i < arr.length; i++) {
    if (i === Math.floor(arr.length / 2)) {
      continue;
    }
    if (arr[i] < pivot) {
      left.push(arr[i]);
```

Questioning...


LD **You**

What does this program do and how does it work ?

And the answer is.....

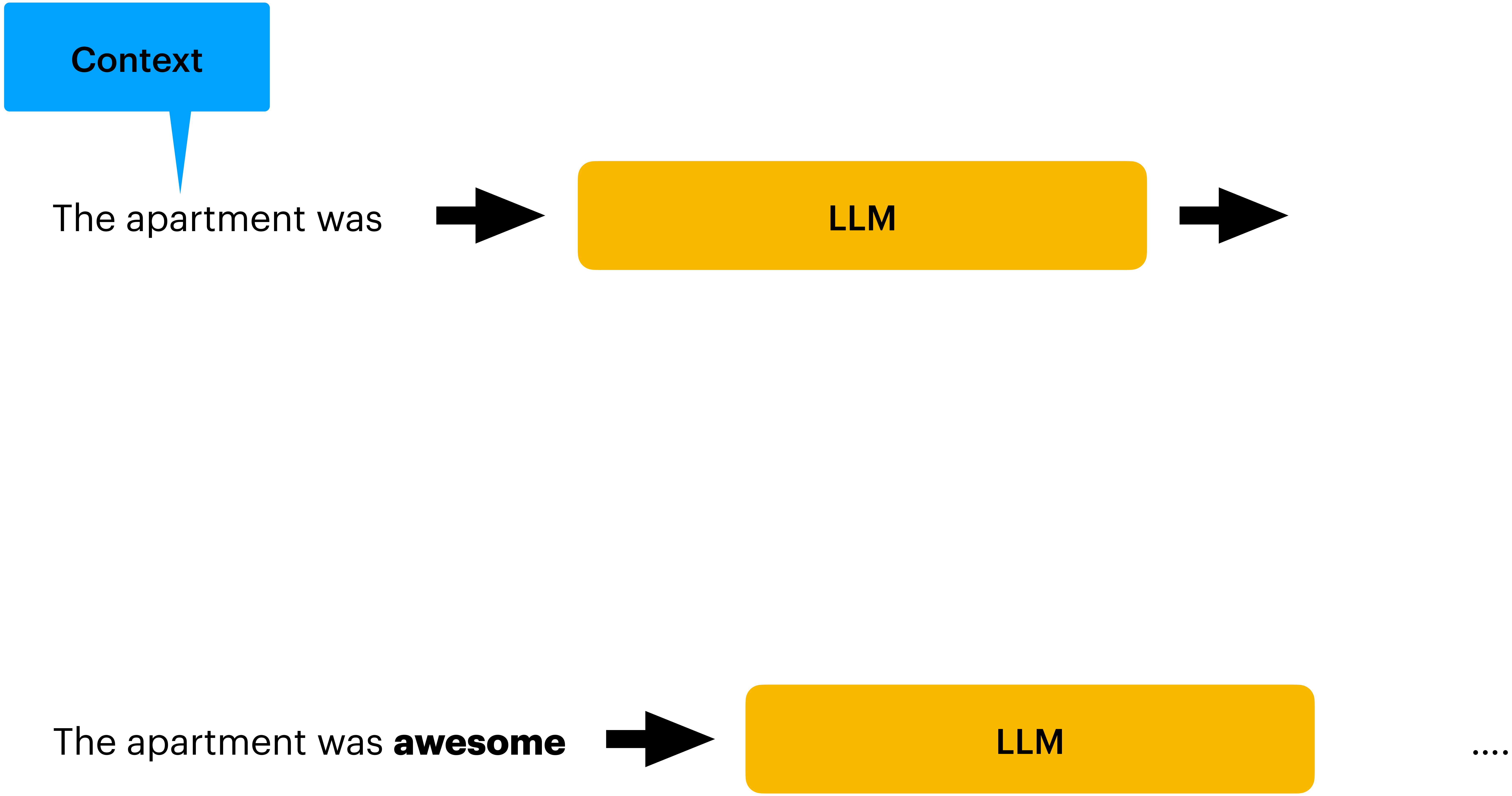
```haskell
{-# LANGUAGE BangPatterns #-}
import Data.Digest.Pure.MD5
import qualified Data.ByteString.Lazy as L
import System.Environment
import Control.Concurrent
import Control.Concurrent.Chan
import Control.Monad (forever, forM_, replicateM_)

nrWorkers = 4

main = do
    files <- getArgs
    str <- newChan
    fileChan <- newChan
    forM_ [1..nrWorkers] (\_ -> forkIO $ worker str fileChan)
    forM_ files (writeChan fileChan)
    printNrResults (length files) str

printNrResults i var = replicateM_ i (readChan var >>= putStrLn)

worker :: Chan String -> Chan String -> IO ()
worker str fileChan = forever (readChan fileChan >>= hashAndPrint str)

hashAndPrint str f = do
        bs <- L.readFile f
        let !h = show $ md5 bs
        writeChan str (f ++ ": " ++ h)
```

Their secrets

An LLM's job...

- It is *still only about predictions!*
    - Given a prompt (a sequence of words), form an **initial context**
    - Predicts the most likely word (first word of answer) following the context
    - It then uses the prompt and the generated word as a **new context**
    - And predicts the next word...
    - Repeat!

In a nutshell

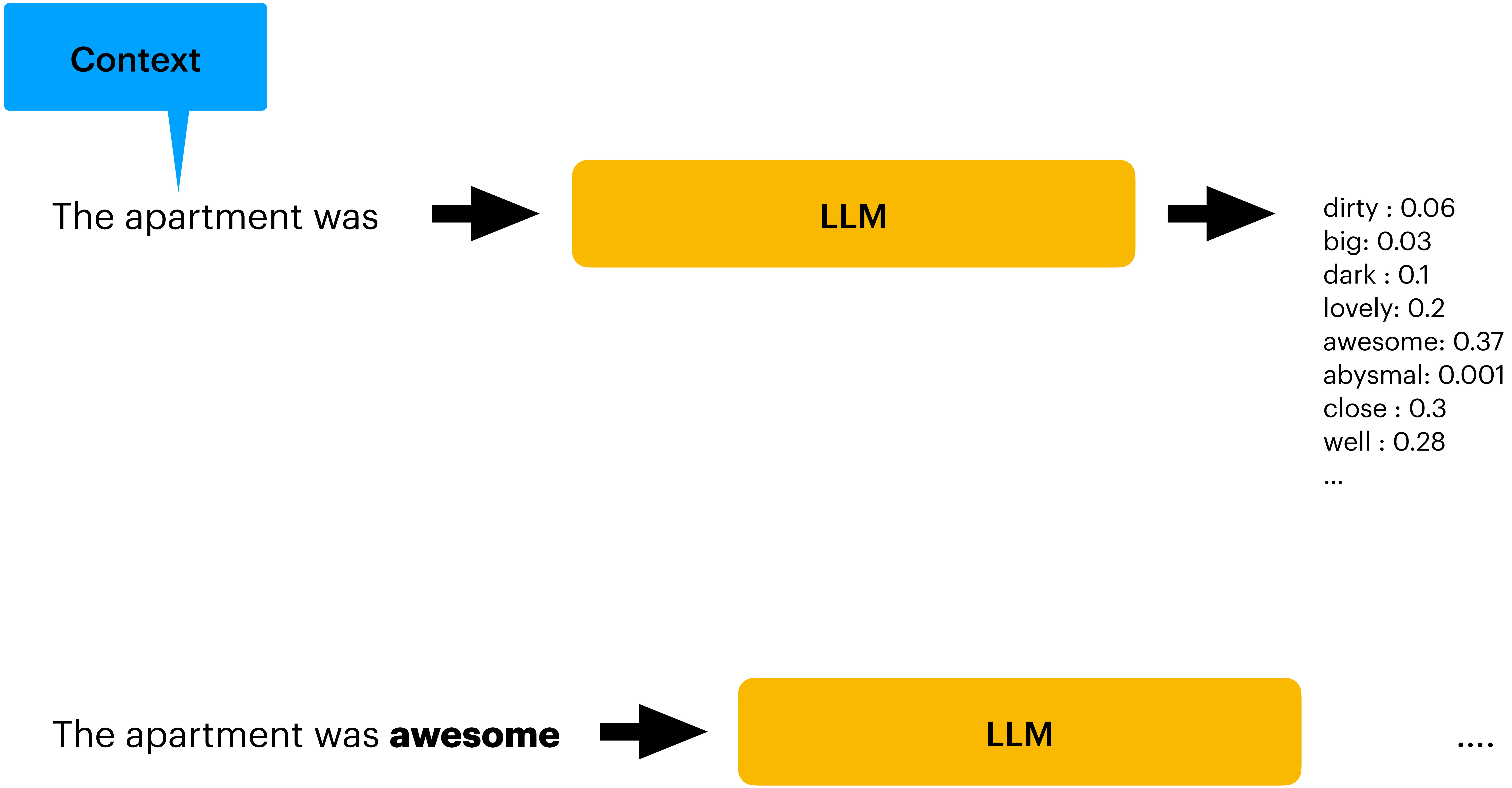


Context

The apartment was → LLM →

The apartment was **awesome** → LLM ....

In a nutshell



Context

The apartment was

LLM

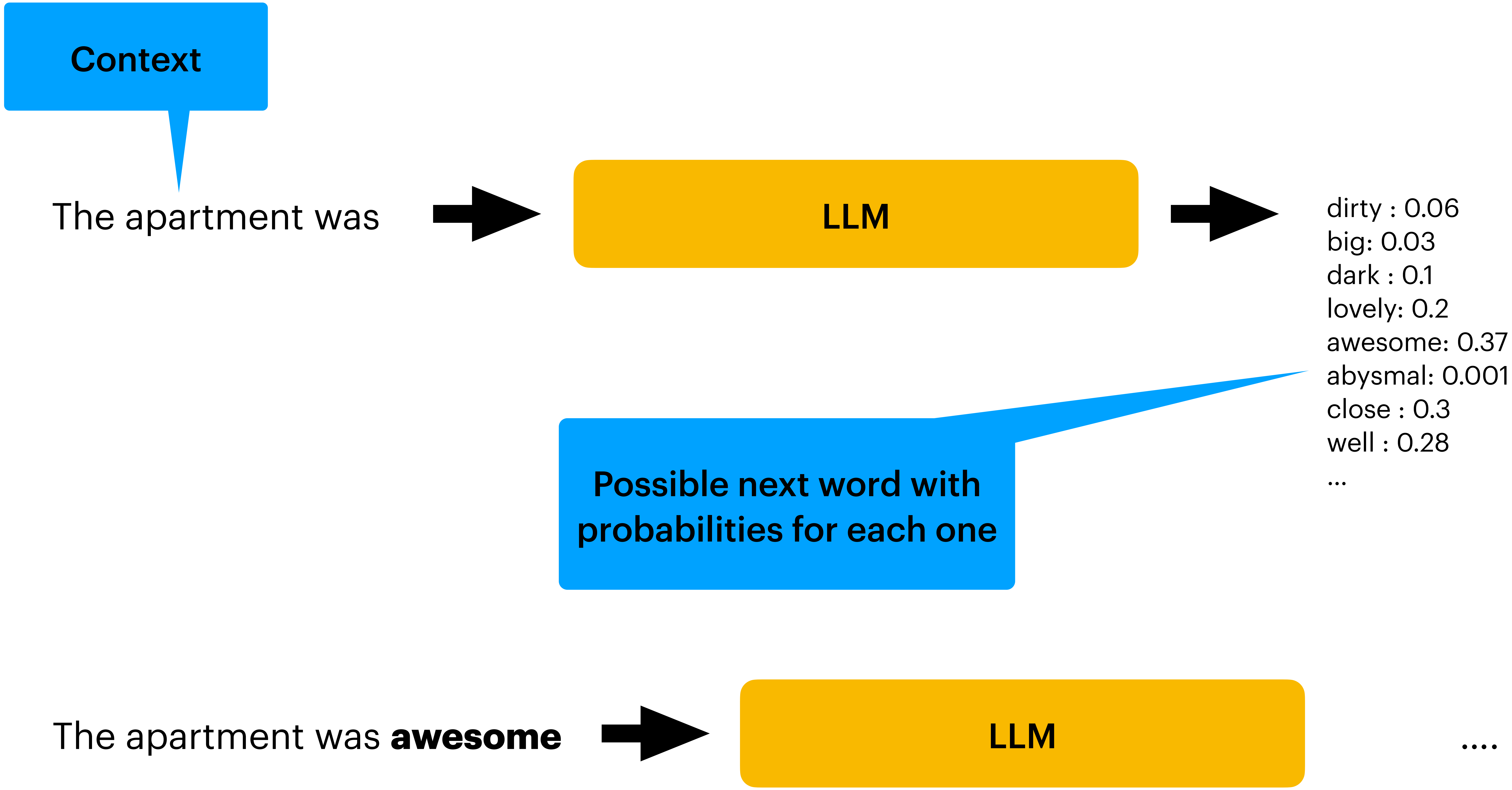
dirty : 0.06
big: 0.03
dark : 0.1
lovely: 0.2
awesome: 0.37
abysmal: 0.001
close : 0.3
well : 0.28
...

The apartment was **awesome**

LLM

....

In a nutshell



Context

The apartment was

LLM

dirty : 0.06
big: 0.03
dark : 0.1
lovely: 0.2
awesome: 0.37
abysmal: 0.001
close : 0.3
well : 0.28
...

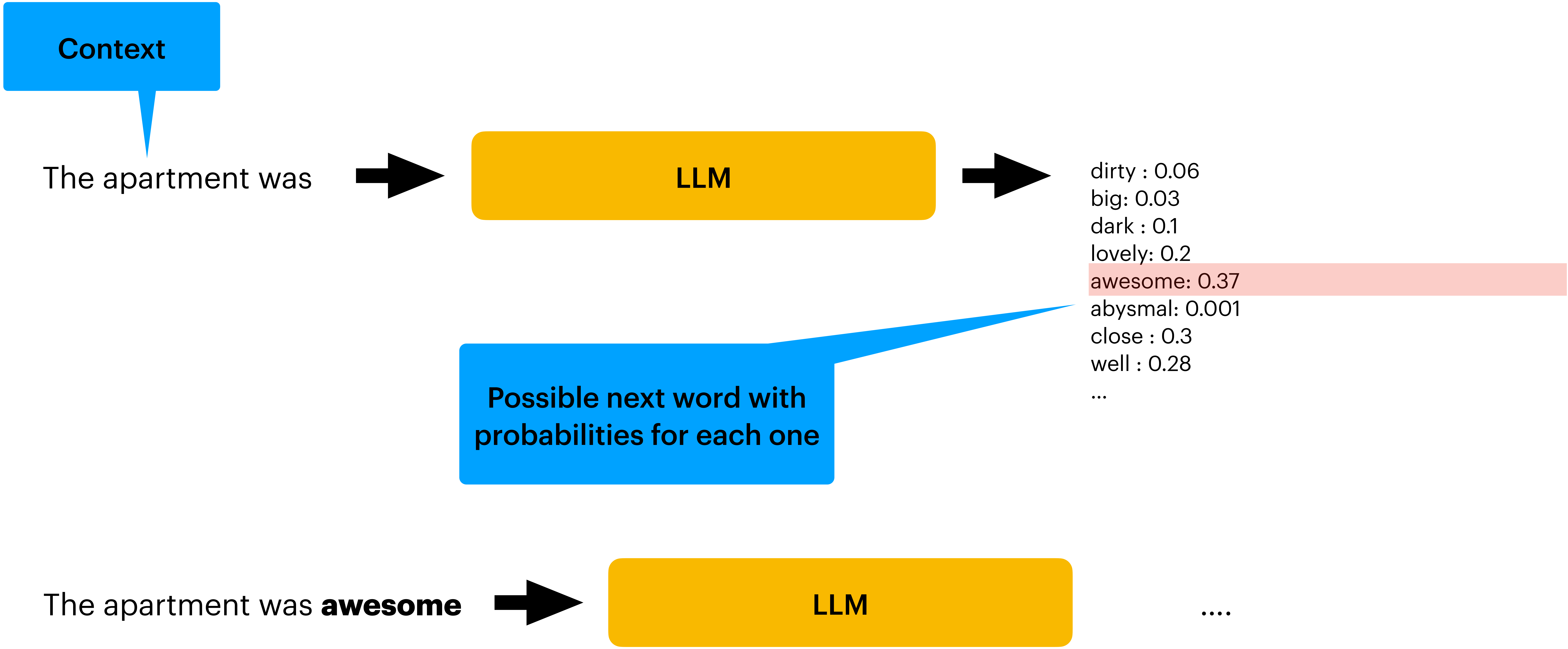
Possible next word with probabilities for each one

The apartment was **awesome**

LLM

....

In a nutshell



Context

The apartment was → LLM →

dirty : 0.06
big: 0.03
dark : 0.1
lovely: 0.2
awesome: 0.37
abysmal: 0.001
close : 0.3
well : 0.28
...

Possible next word with probabilities for each one

The apartment was **awesome** → LLM ....

In a nutshell

Context

The apartment was → LLM →

dirty : 0.06
big: 0.03
dark : 0.1
lovely: 0.2
awesome: 0.37
abysmal: 0.001
close : 0.3
well : 0.28
…

Possible next word with probabilities for each one

The apartment was **aw**

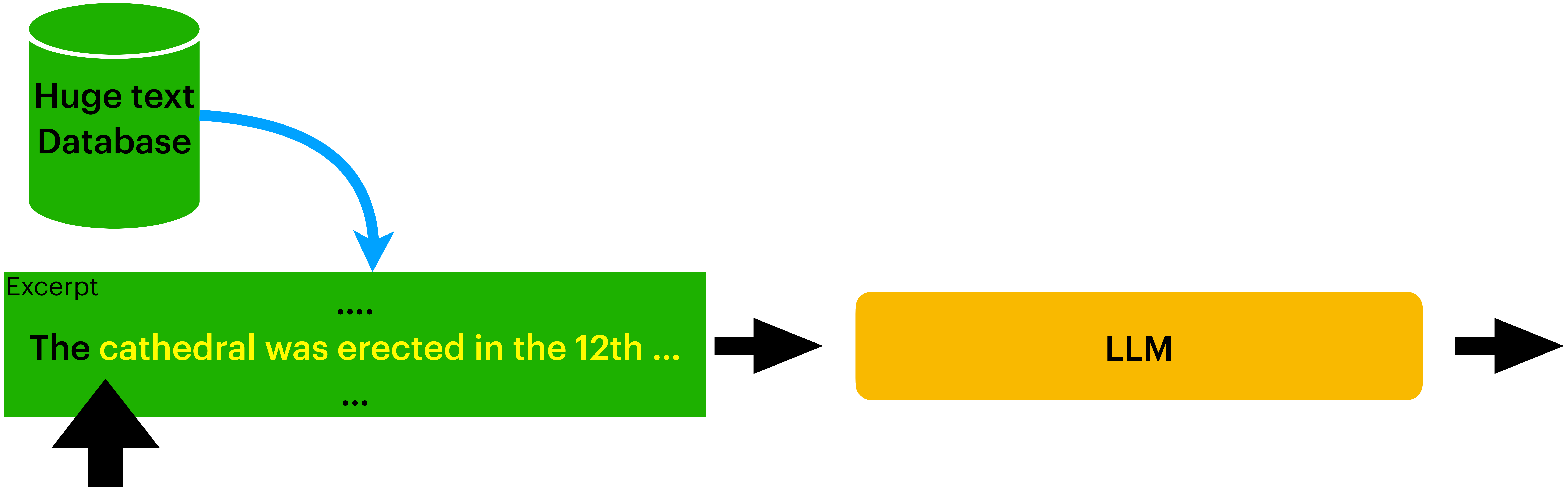
So it is very simple…
And very “mechanical”

There is no “sentience” here!

.

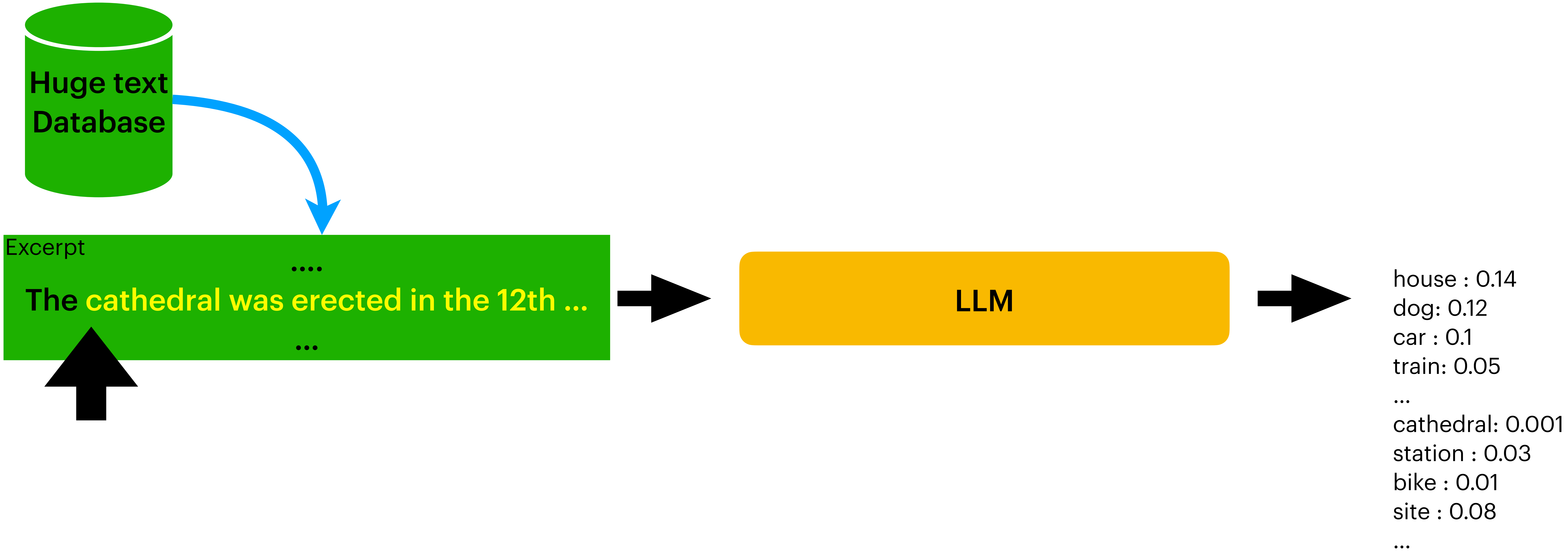
LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ....
- Learn to predict the next word in any prefix of any sentence!



LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ....
- Learn to predict the next word in any prefix of any sentence!



Huge text Database

Excerpt

....

The cathedral was erected in the 12th ...

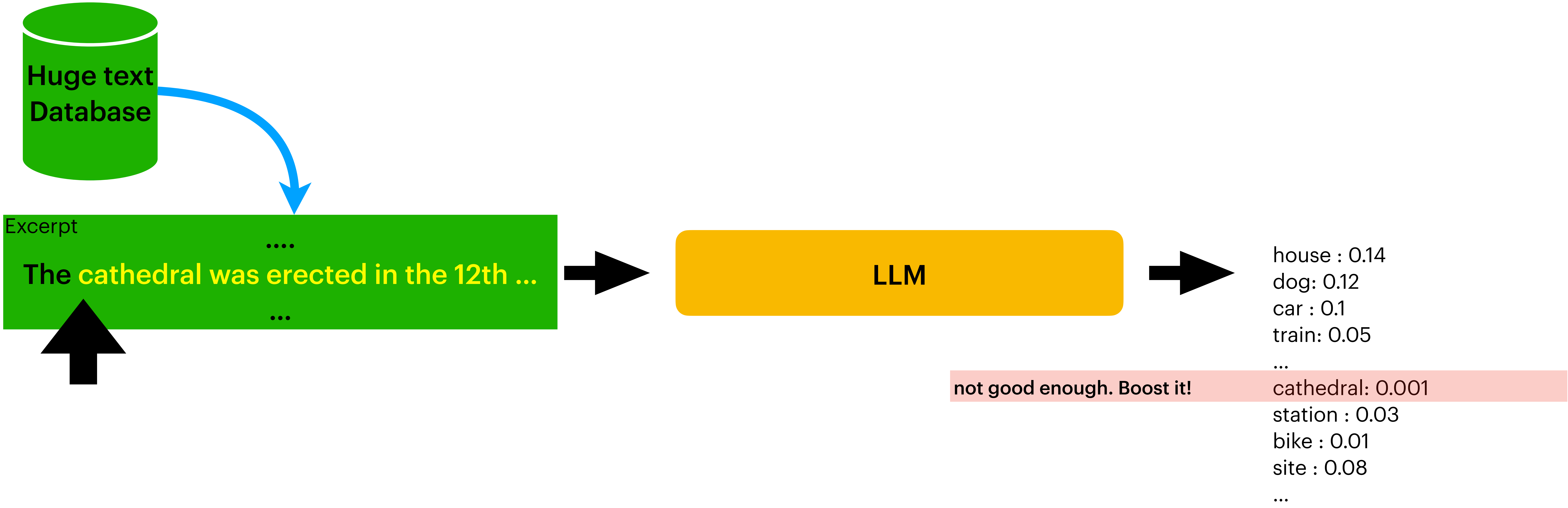
...

LLM

house : 0.14
dog: 0.12
car : 0.1
train: 0.05
...
cathedral: 0.001
station : 0.03
bike : 0.01
site : 0.08
...

LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ....
- Learn to predict the next word in any prefix of any sentence!



Huge text Database

Excerpt

....

The cathedral was erected in the 12th ...

...

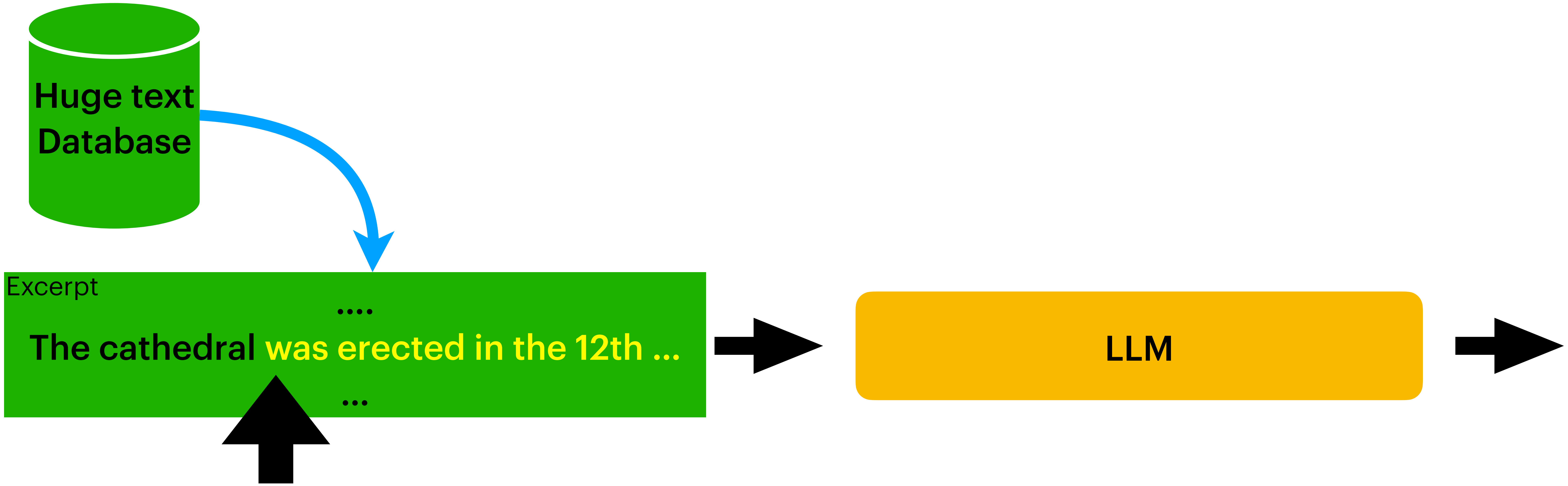
LLM

house : 0.14
dog: 0.12
car : 0.1
train: 0.05
...
cathedral: 0.001
station : 0.03
bike : 0.01
site : 0.08
...

not good enough. Boost it!

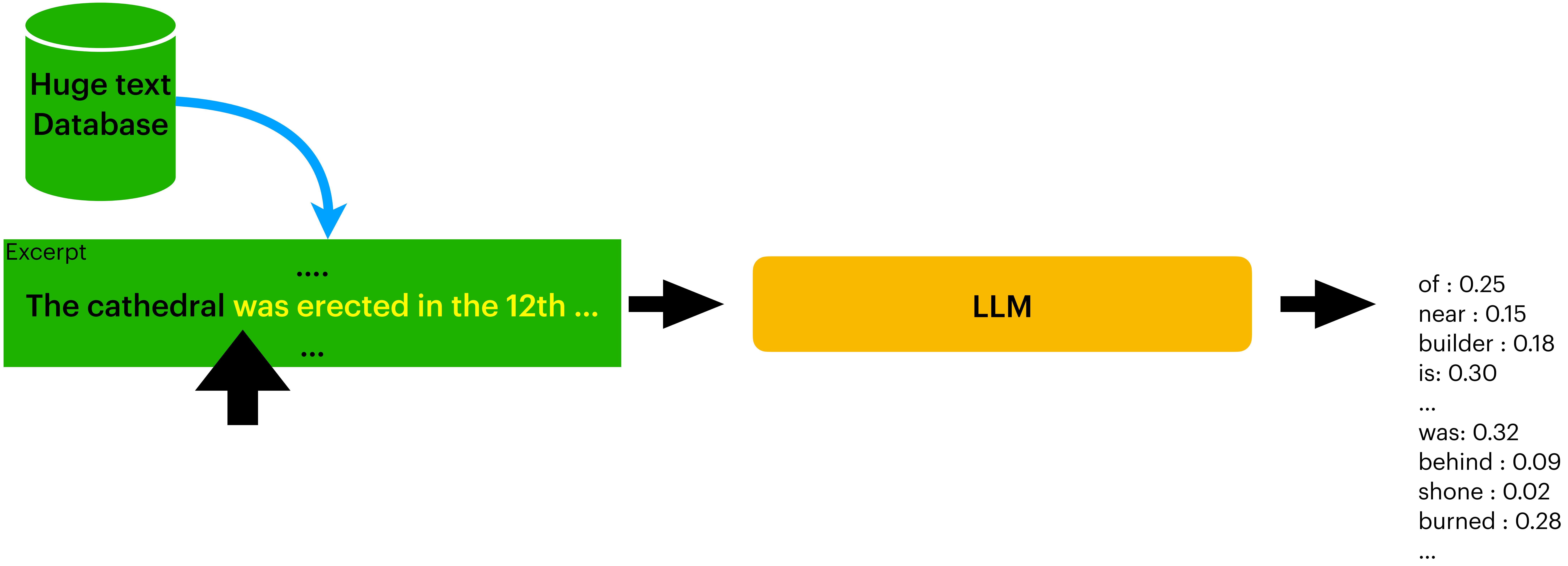
LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ....
- Learn to predict the next word in any prefix of any sentence!



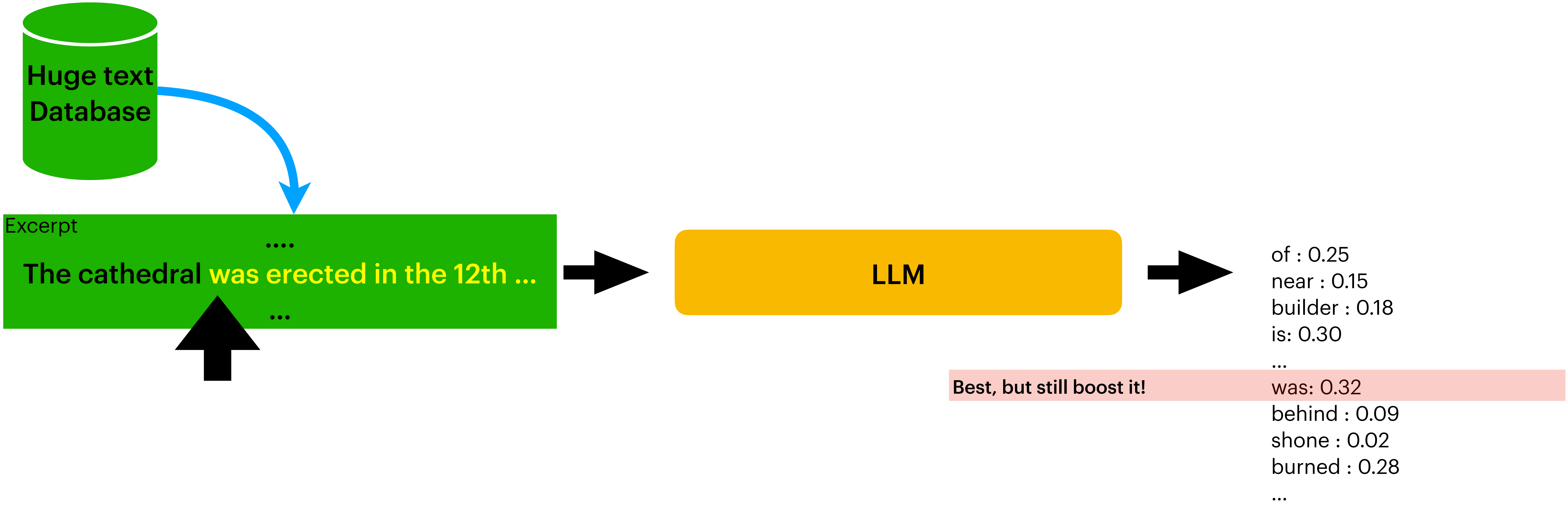
LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ....
- Learn to predict the next word in any prefix of any sentence!



LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ….
- Learn to predict the next word in any prefix of any sentence!



Huge text Database

Excerpt

….

The cathedral was erected in the 12th …

…

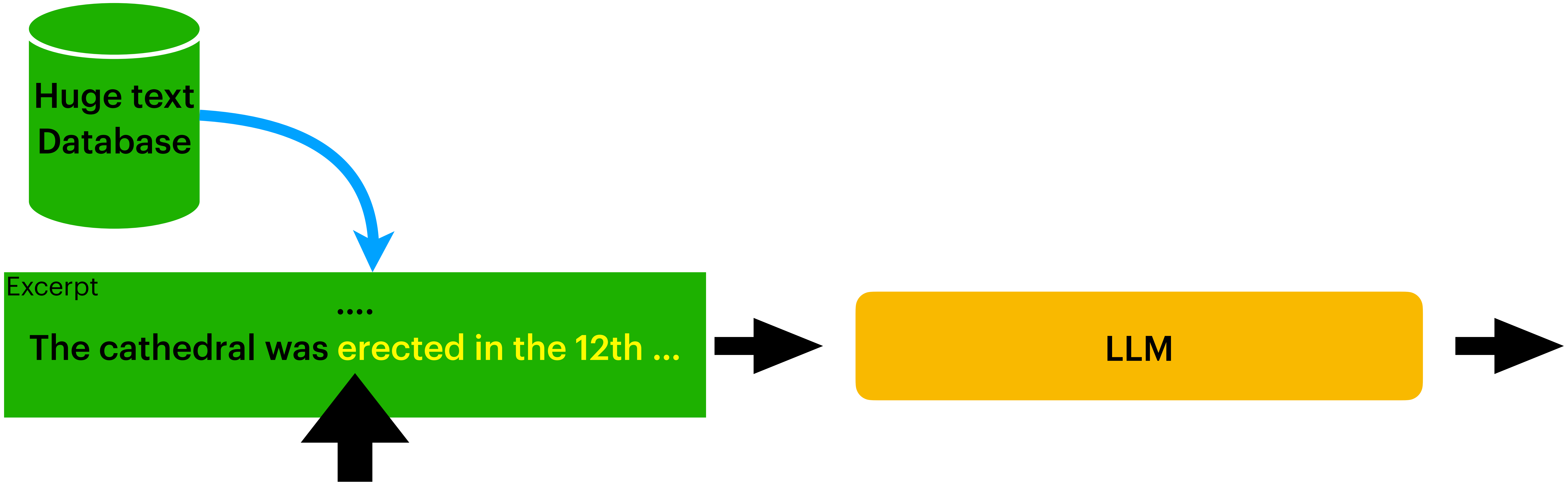
LLM

of : 0.25
near : 0.15
builder : 0.18
is: 0.30
…
was: 0.32
behind : 0.09
shone : 0.02
burned : 0.28
…

Best, but still boost it!

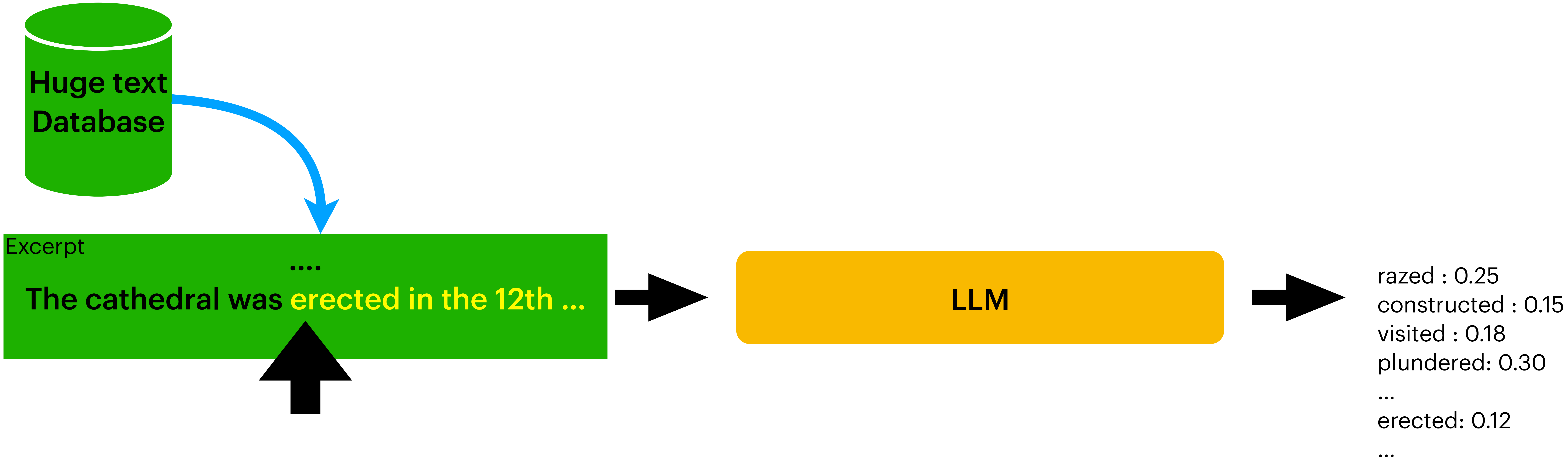
LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ….
- Learn to predict the next word in any prefix of any sentence!



LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ....
- Learn to predict the next word in any prefix of any sentence!



Huge text Database

Excerpt

....

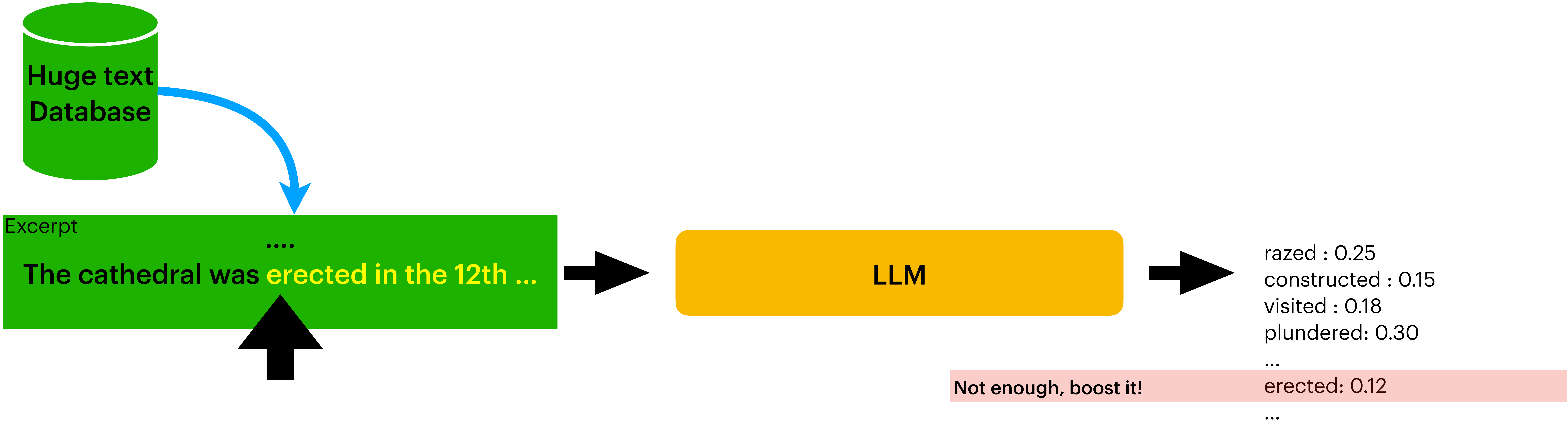
The cathedral was erected in the 12th ...

LLM

razed : 0.25
constructed : 0.15
visited : 0.18
plundered: 0.30
...
erected: 0.12
...

LLMs

- Use vast amounts of texts from many sources
  - Web / Digital libraries / News / Patents / Laws / Books ....
- Learn to predict the next word in any prefix of any sentence!



Huge text Database

Excerpt

....

The cathedral was erected in the 12th ...

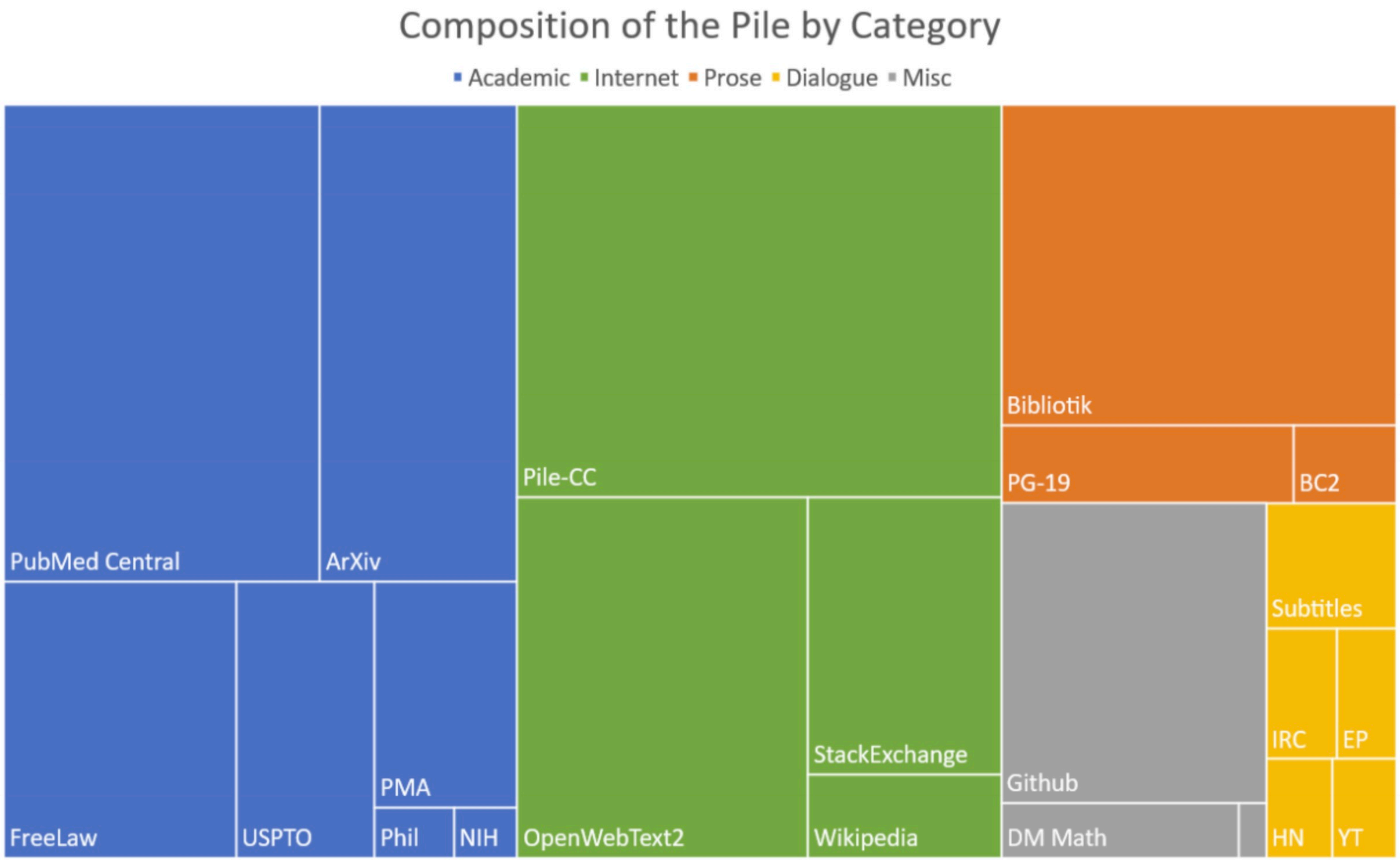
LLM

razed : 0.25
constructed : 0.15
visited : 0.18
plundered: 0.30
...
erected: 0.12
...

Not enough, boost it!

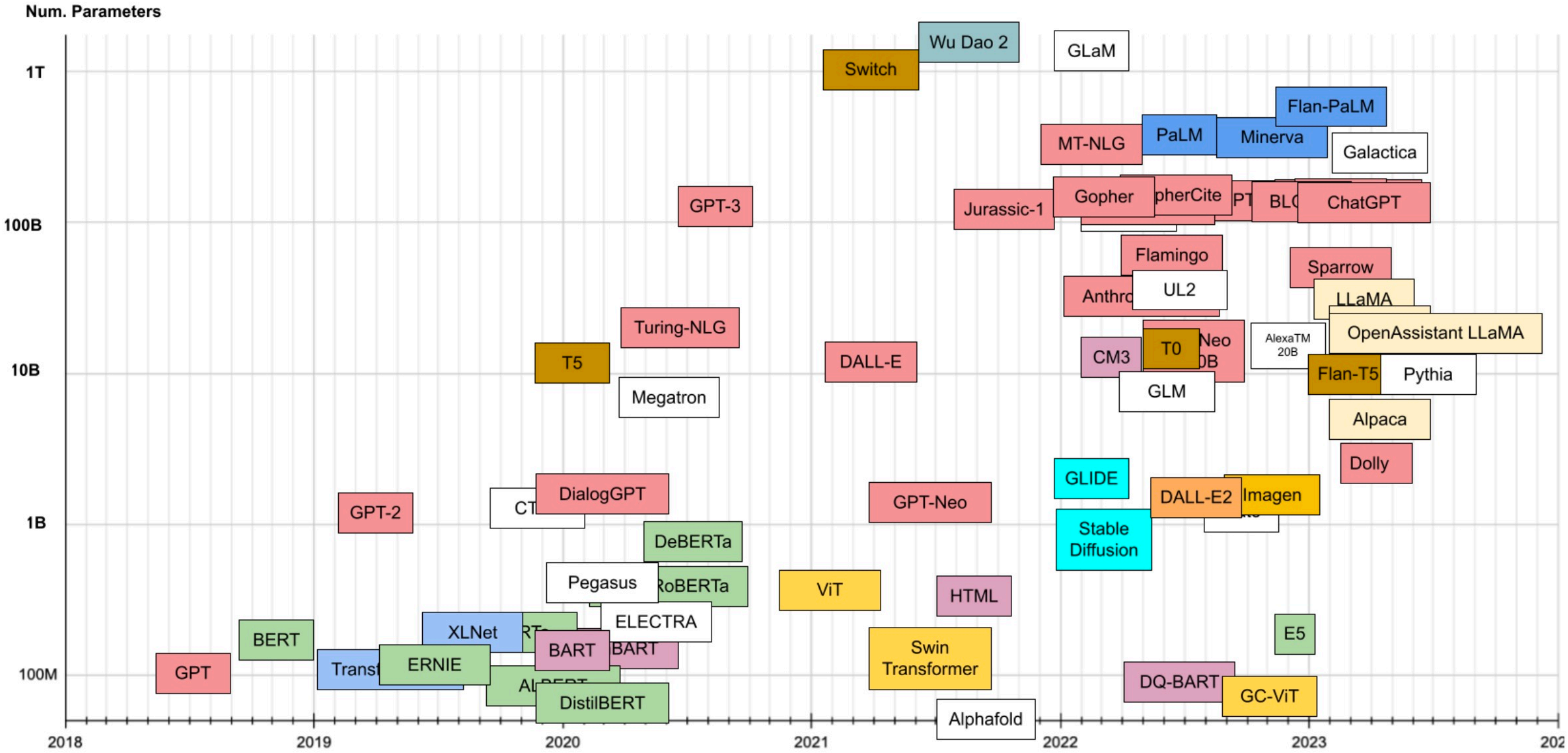
What LLMs train on…

- We do not really know.
- But there are large public sets like…
- The “Pile” (https://paperswithcode.com/dataset/the-pile)
- 800 GB of data.



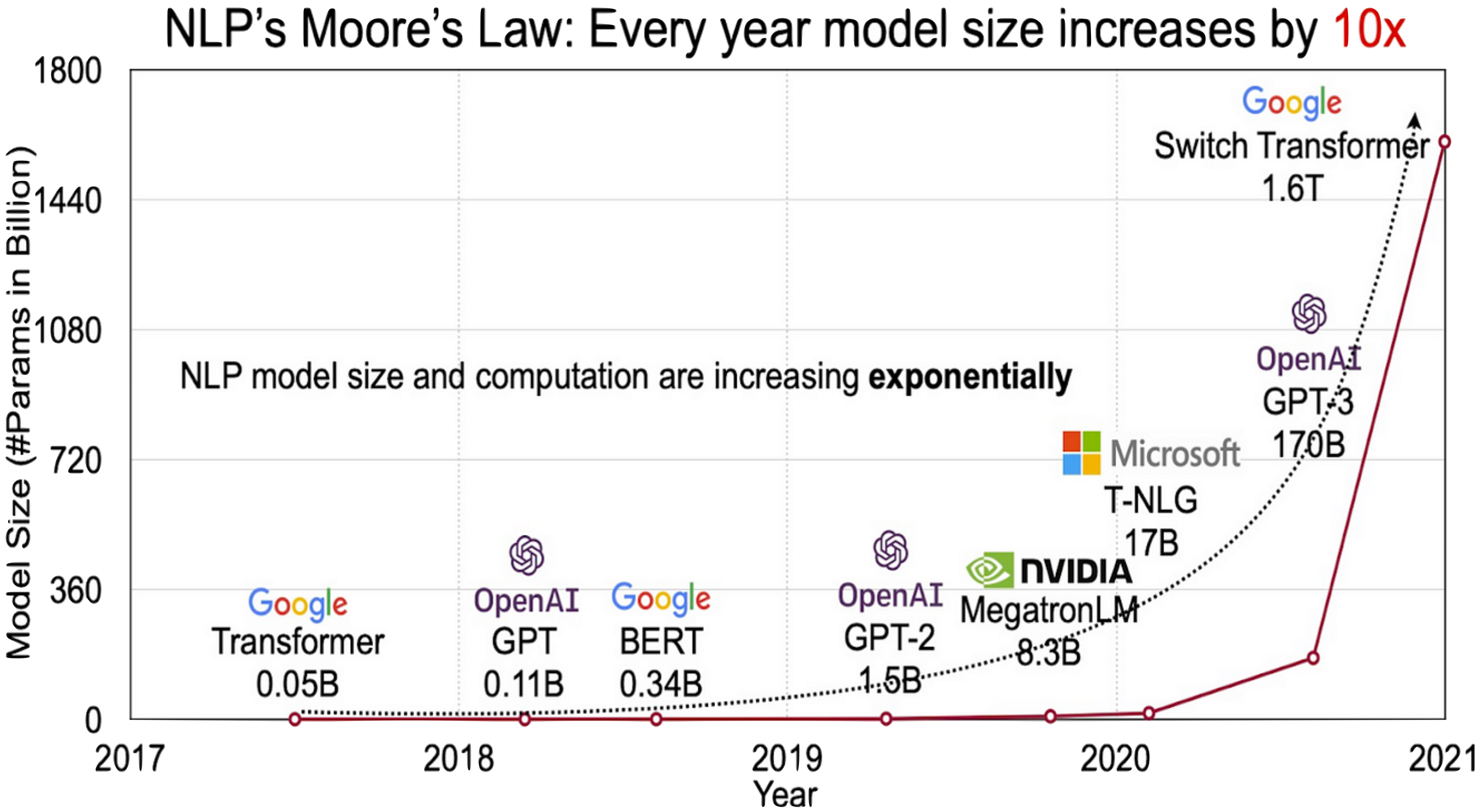
LLMs ranked by size

https://arxiv.org/pdf/2302.07730.pdf



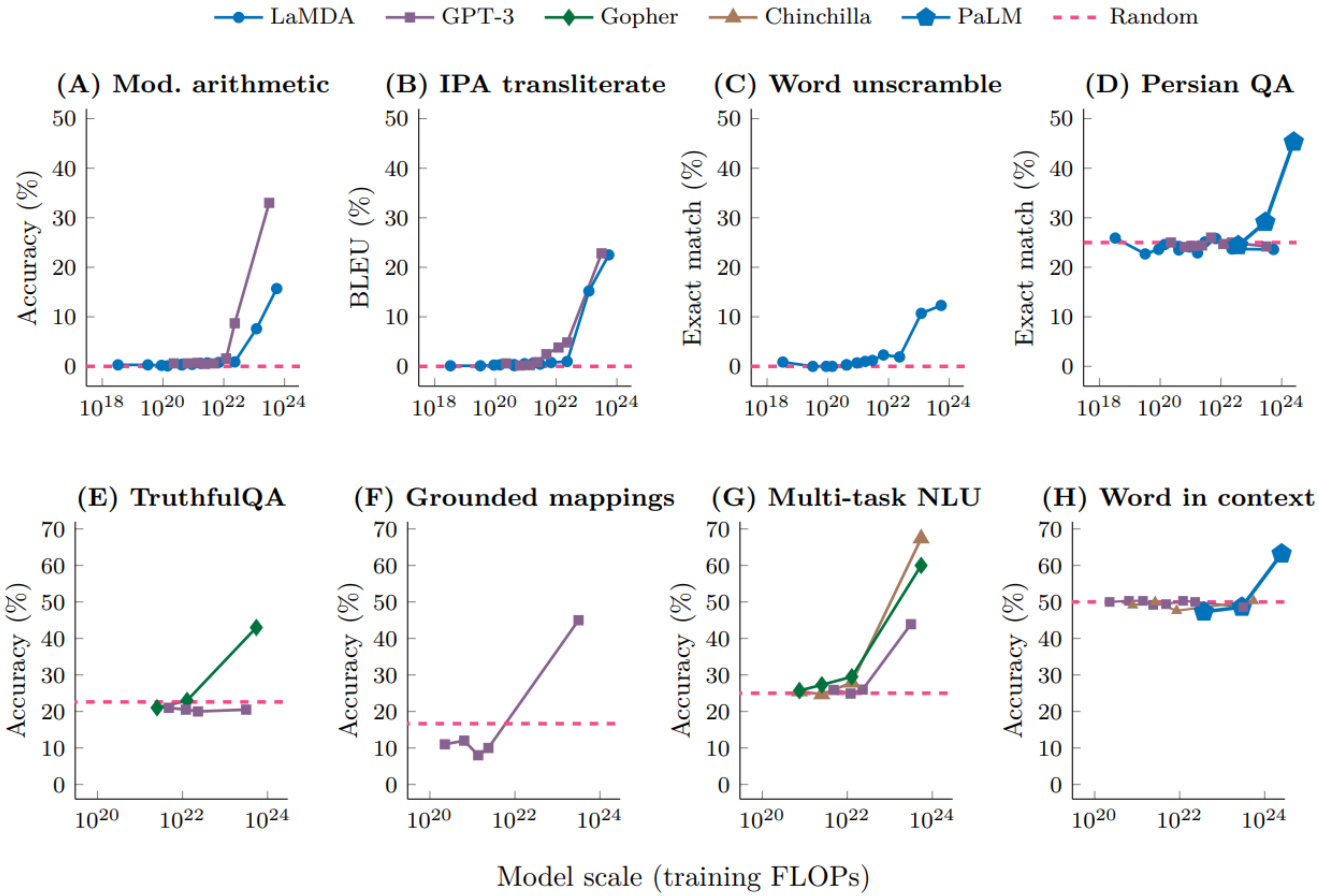
Size of LLMs

- Incredible growth in size!
  - Hundreds of billions at the moment
  - Why?



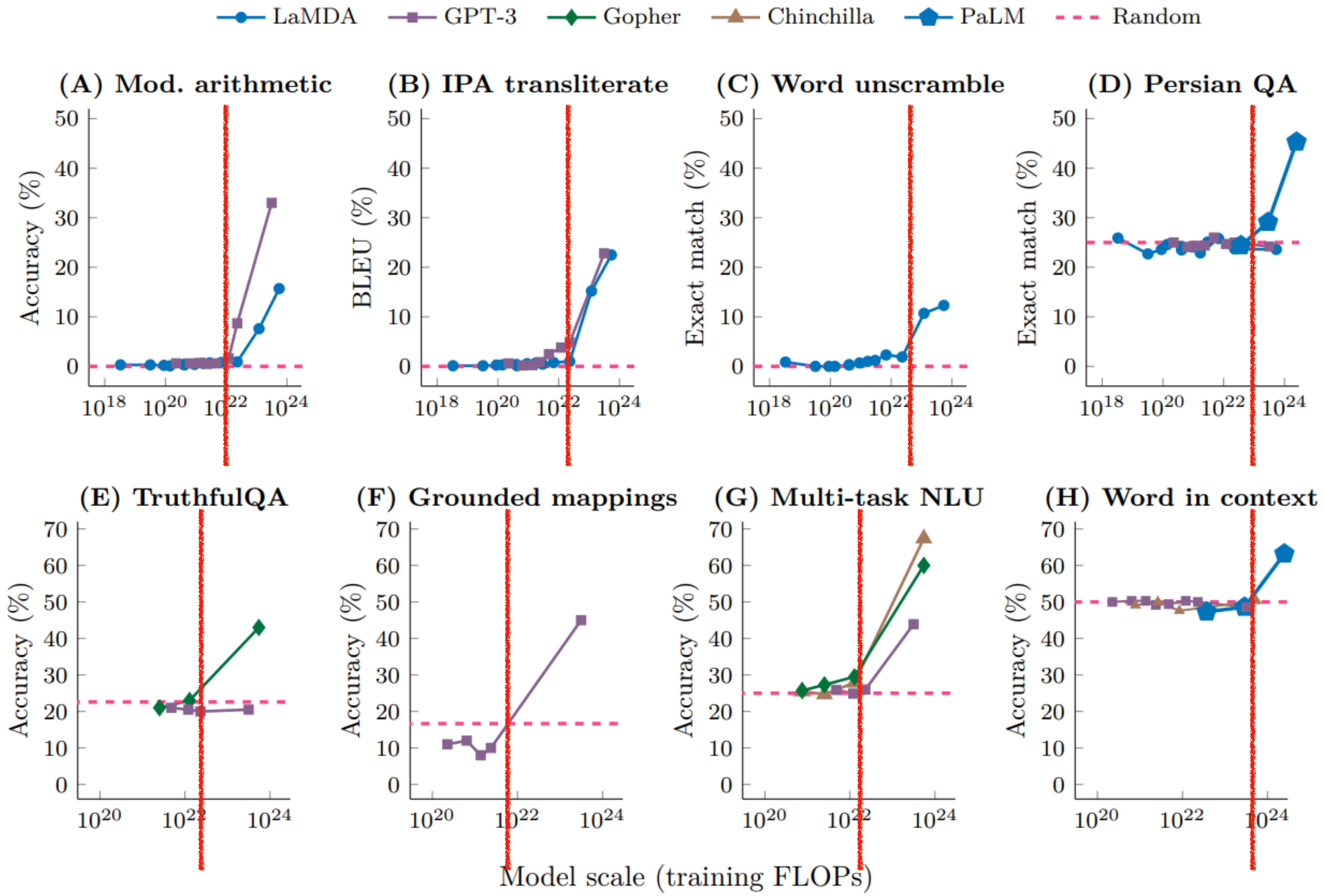
Bigger is Better

- Threshold effect….



Bigger is Better

- Threshold effect….



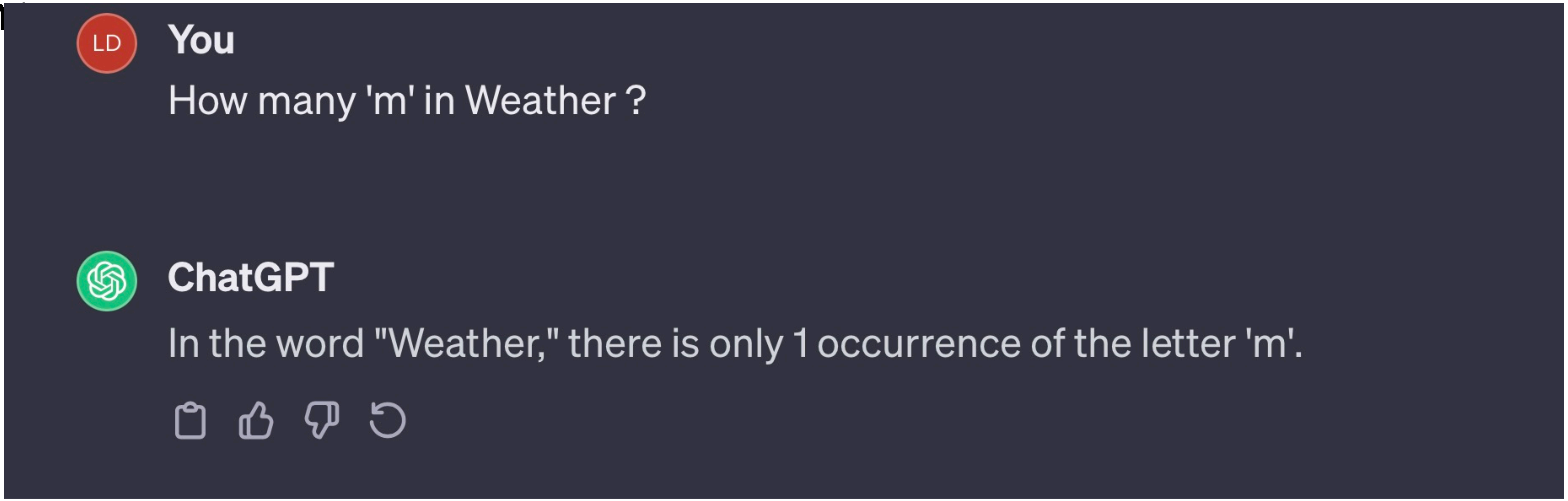
Their Achilles’ heel

LLM Hallucinations

- Phenomenon that happens regularly
  - Many causes
    - Not enough training data
    - Silly prompt
    - Overfitting
    - Biases
    - Lack of grounding (experiential)

LLM Hallucinations

- Phenomenon that happens regularly
  - Many causes
    - Not enough training data
    - Silly prompt
    - Overfittin
    - Biases
    - Lack of g



You

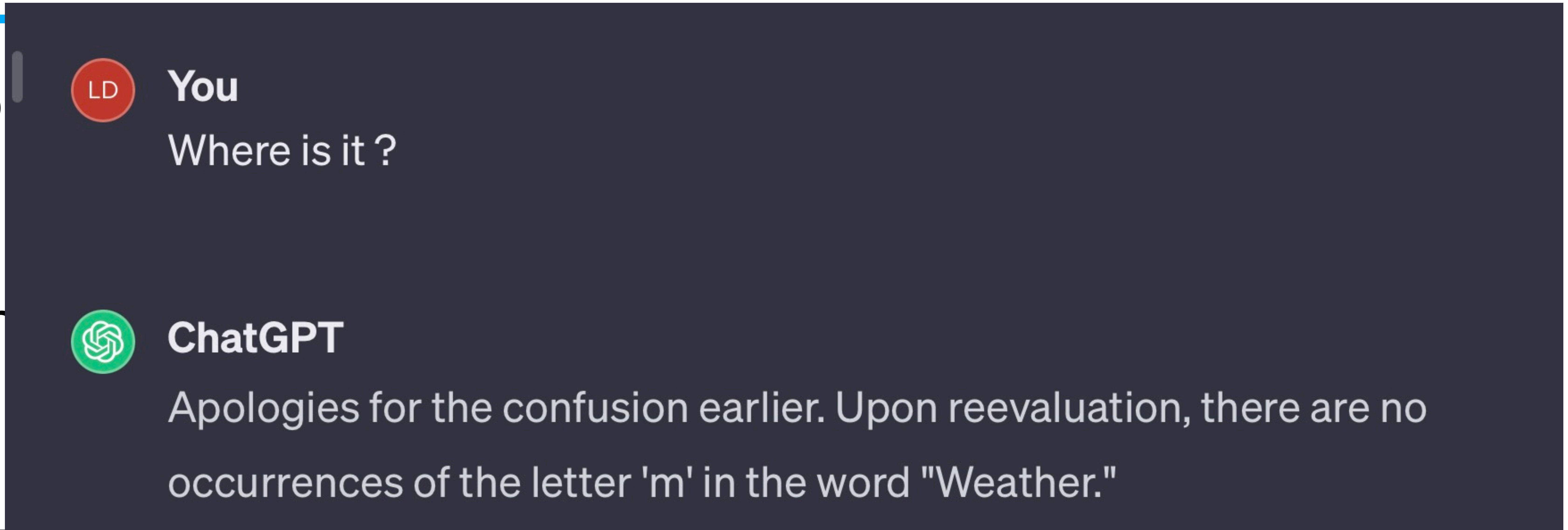
How many 'm' in Weather ?

ChatGPT

In the word "Weather," there is only 1 occurrence of the letter 'm'.

LLM Hallucinations

- Phenomenon that hap
  - Many causes
    - Not enough trainin
    - Silly prompt
    - Overfittin
    - Biases
    - Lack of g

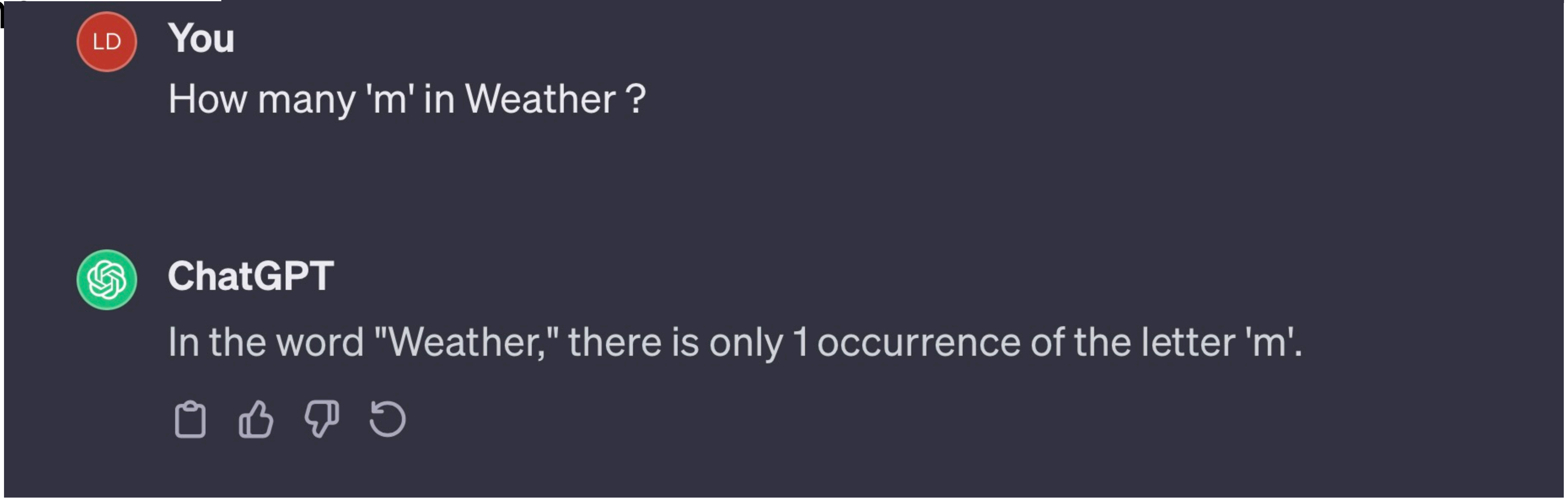


You

Where is it ?

ChatGPT

Apologies for the confusion earlier. Upon reevaluation, there are no occurrences of the letter 'm' in the word "Weather."



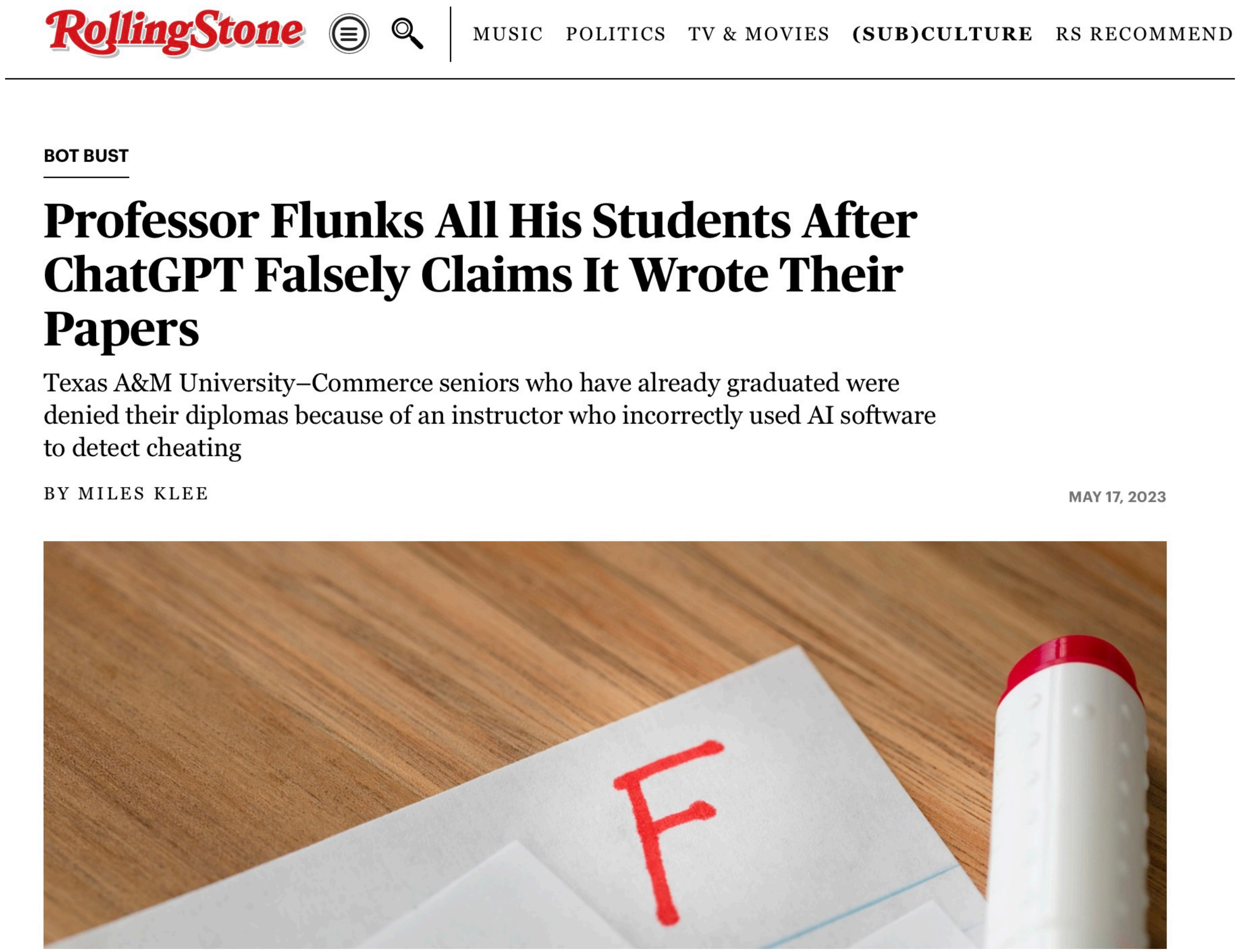
You

How many 'm' in Weather ?

ChatGPT

In the word "Weather," there is only 1 occurrence of the letter 'm'.

Hallucinations in the World



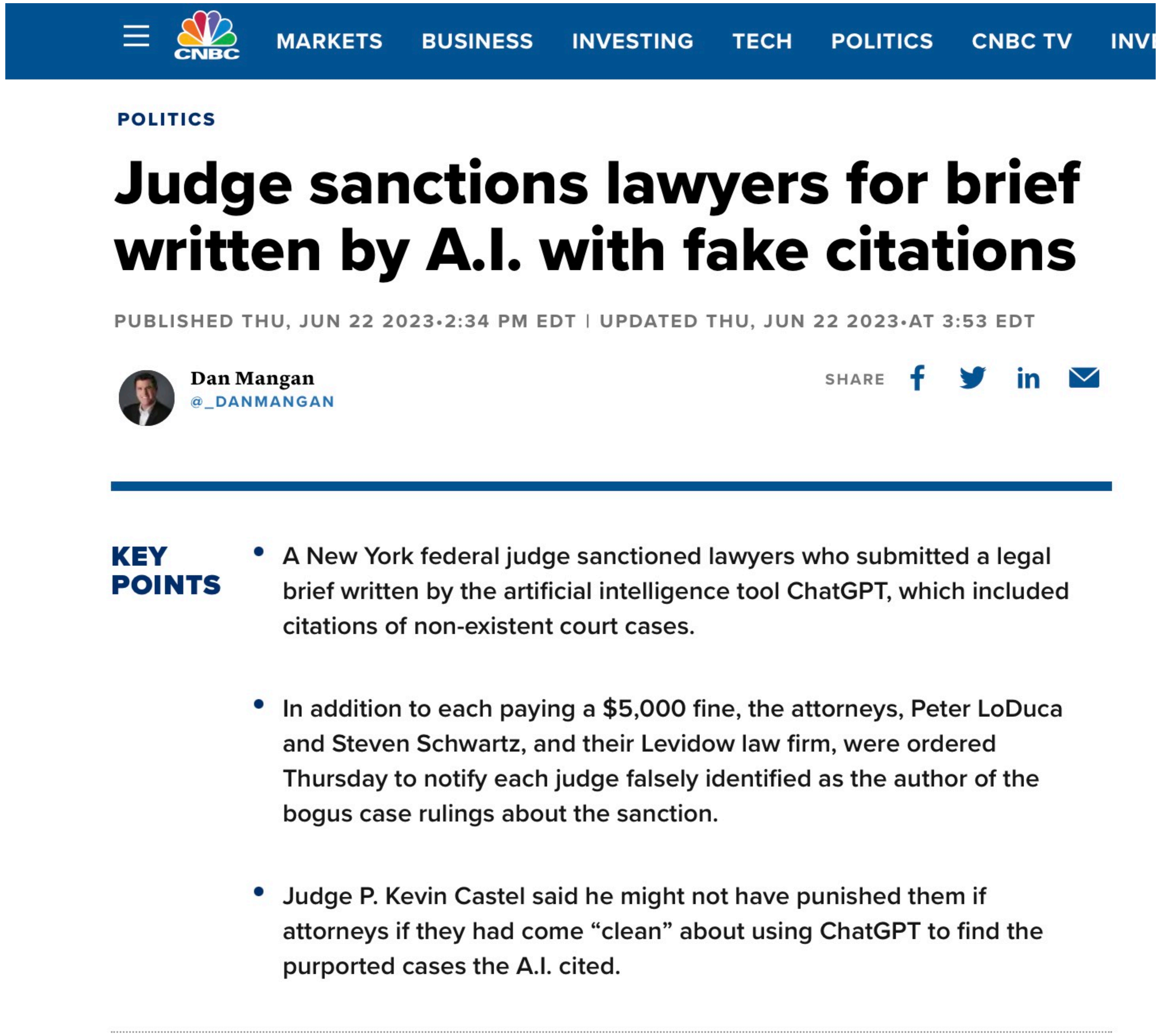
BOT BUST

Professor Flunks All His Students After ChatGPT Falsely Claims It Wrote Their Papers

Texas A&M University–Commerce seniors who have already graduated were denied their diplomas because of an instructor who incorrectly used AI software to detect cheating

BY MILES KLEE

MAY 17, 2023



POLITICS

Judge sanctions lawyers for brief written by A.I. with fake citations

PUBLISHED THU, JUN 22 2023•2:34 PM EDT | UPDATED THU, JUN 22 2023•AT 3:53 EDT

Dan Mangan
@_DANMANGAN

KEY POINTS

- A New York federal judge sanctioned lawyers who submitted a legal brief written by the artificial intelligence tool ChatGPT, which included citations of non-existent court cases.
- In addition to each paying a $5,000 fine, the attorneys, Peter LoDuca and Steven Schwartz, and their Levidow law firm, were ordered Thursday to notify each judge falsely identified as the author of the bogus case rulings about the sanction.
- Judge P. Kevin Castel said he might not have punished them if attorneys if they had come “clean” about using ChatGPT to find the purported cases the A.I. cited.

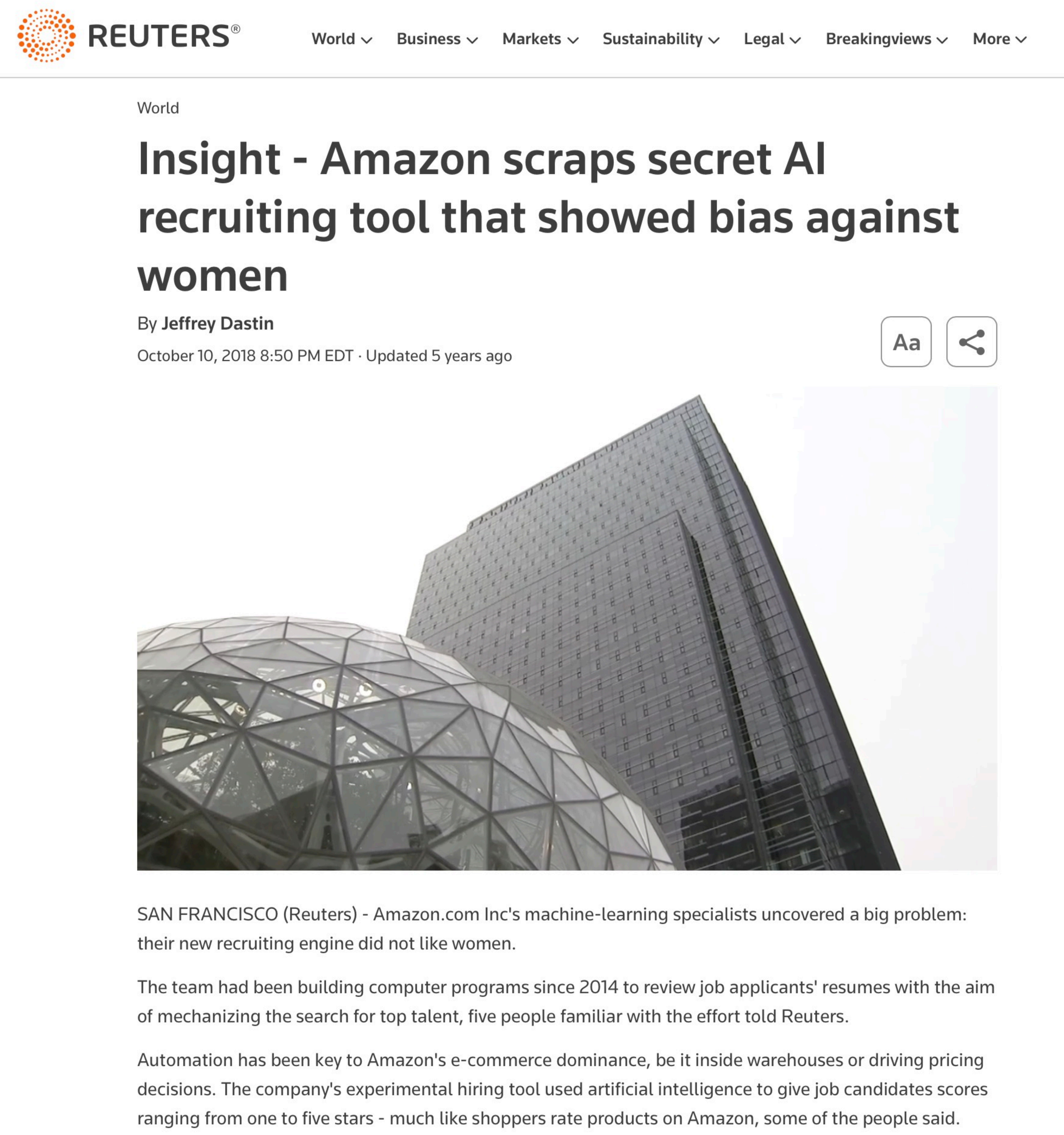
Training LLM

- Soon…
  - LLM content will be published online
  - And picked up by other LLMs for training
  - Machine can generate content faster than humans….

What happens when LLMs content dwarves human content and is used for LLM training?

Using LLMs...

- Only produce the most likely content
  - Even if it is inappropriate
  - It will reproduce the biases in its training data!
  - This is true of ***any*** deep-learning technique



World

Insight - Amazon scraps secret AI recruiting tool that showed bias against women

By **Jeffrey Dastin**

October 10, 2018 8:50 PM EDT · Updated 5 years ago

SAN FRANCISCO (Reuters) - Amazon.com Inc's machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said.

LLM Explanation

- Inability to explain the text generated
  - It’s probabilistic
  - It was not based on the underlying meaning
  - It reproduces variants of texts it has seen
  - No understanding of the underlying domain!

So What ?

The AI Era : Light Side

- AI will continue to evolve
  - Better handling corner case
  - Trained on more data
- LLMs are very good at…
  - Repetitive, error-prone tasks with *many examples* available in text form
    - Porting code
    - Documenting code
    - Summarizing text

The AI Era : Dark Side

- LLMs & challenges…
  - Not creative
  - No underlying semantics
  - Lack of explanations
  - Hallucinations (How do you vet?, How do you prevent?)
  - Bias in training data
  - Increasingly closed source (you can’t see how it’s trained)
  - Expensive to train!
  - Ethical issues will arise (already do!)

Homework

- Stay curious
- Keep learning about AI’s directions
- It’s all about your data now!
- Beware
  - We “bestow” intelligence/sentience based on what we see.
  - Once you know how it works, it dispels beliefs
  - Informed consumers are better consumers!
  - Keep learning about it.

Credits

- Sameer Singh
  - Associate Professor, UCI, Irvine
  - Good talk on LLM!
- Serdar Kadıoğlu
  - Group VP, AI Center of Excellence, Fidelity
  - Adj. Assoc. Prof., Dept. of Computer Science, Brown
- UConn AI Faculty
  - J. Bi, Derek A., Qian Y, Shiri D., …

Questions?