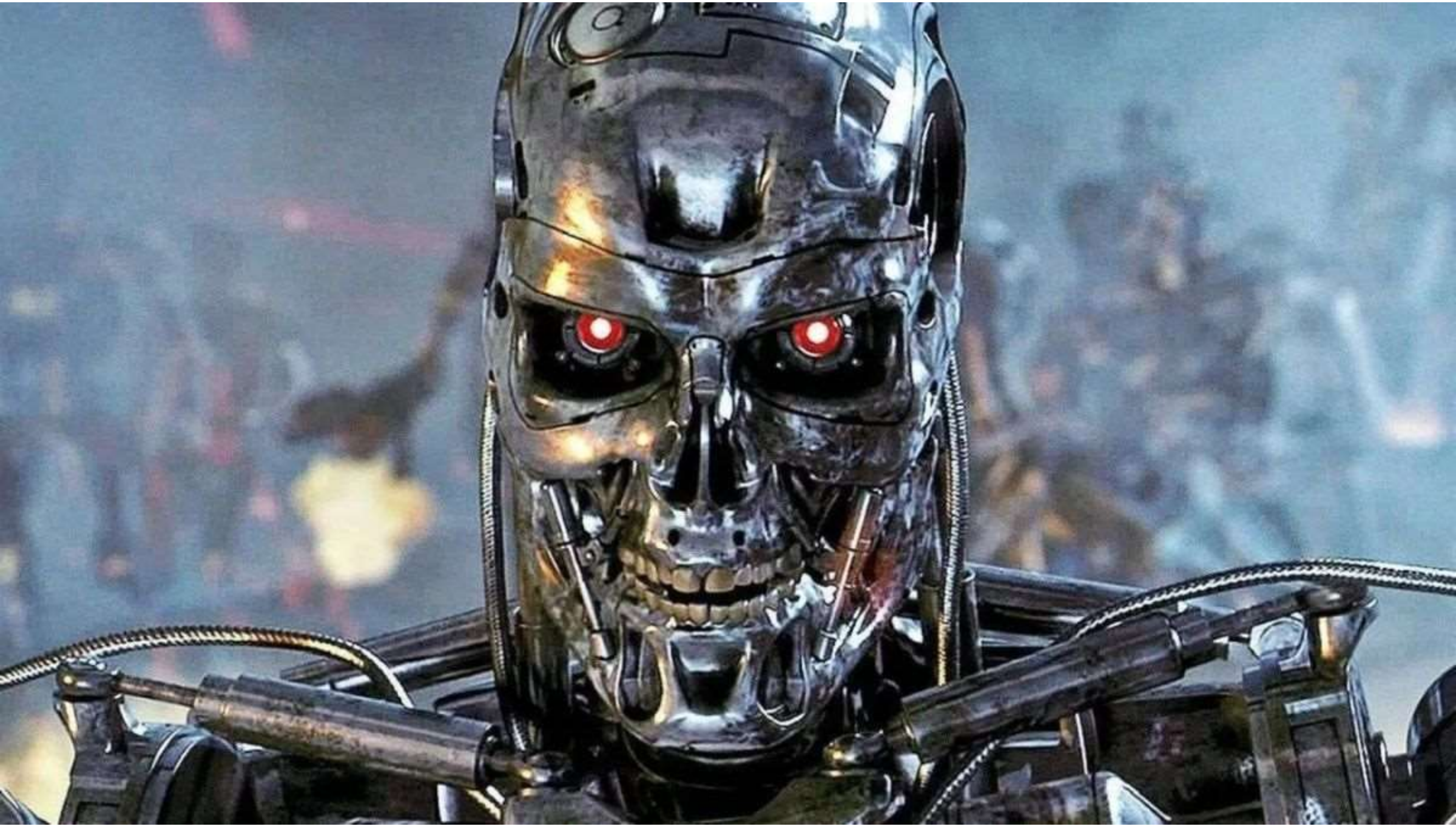
Artificial Intelligence is here

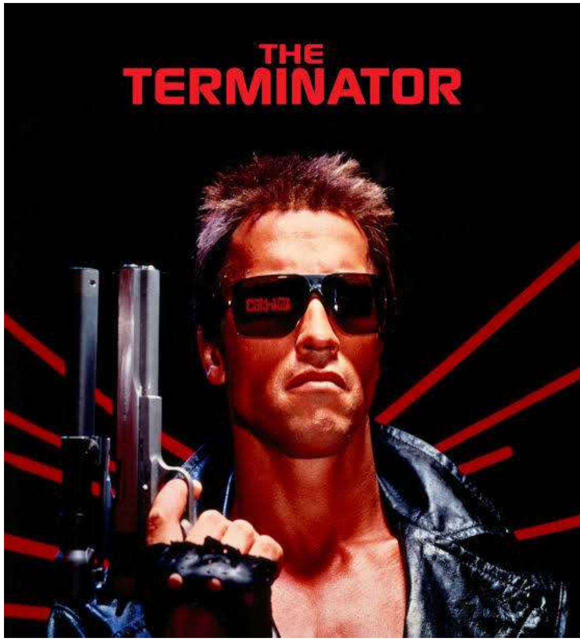
October 1984

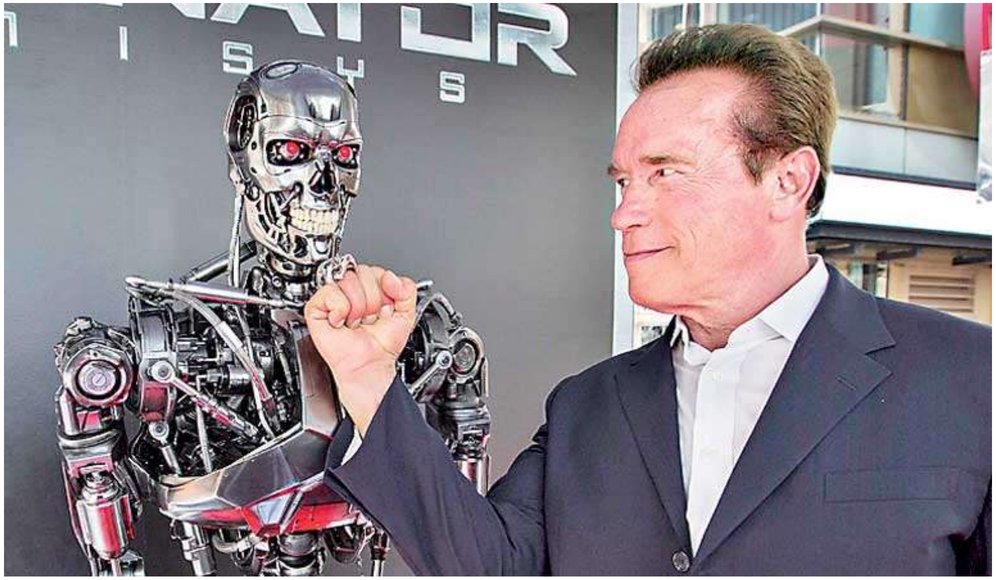
Artificial Intelligence is here



October 1984

Artificial Intelligence is here





What Large Langue Models (LLM) are?

- LLMs can be understood as advanced tools built on the principles of **statistical pattern recognition and prediction**.

- LLMs are designed to **predict the next most probable word ("token**“) in a sequence.

- A "**token" is the fundamental unit of text**. It can be a word, a character, or even a punctuation mark. On average a single word translates to about 0.75 tokens.

How Do Large Language Models (LLMs) Work?

- The term "**sequence**" refers to the context or the "**window**" of text that the model considers when making its predictions. This could be a single sentence, a paragraph, or even a longer body of text like a book chapter.

- Models like ChatGPT-03, the maximum sequence length is 4096 tokens (**July 2022**), which is equivalent to a few pages of text.

Model Version	Maximum Tokens (Context Window)	Notes	Year
GPT-3.5	4,096 – 16,385 tokens	Base model supports 4,096 tokens;	2022
GPT-4	8,192 – 32,768 tokens	Available in 8K and 32K token versions.	2023
GPT-4 Turbo	128,000 tokens	Enhanced version with a significantly larger context window.	May 2024
GPT-4o	128,000 tokens	Multimodal model supporting text, audio, and vision inputs.	Dec 2024
GPT-4.1	1,000,000 tokens	Extensive context window, suitable for processing large datasets.	Feb 2025
GTP-5.0	400,000 tokens	improved reasoning and reduced hallucinations	Aug 2025

How a LLM problem is modelled?

Basically, a LLM is aimed to answer the question: **What is p(text)**

Given a sequence of tokens: $\left(x^{(1)}, x^{(2)}, x^{(3)}, , \ldots\ x^{(n)}\right)$ (1)

Then:

$P(x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(N)})$ is the probability of sequence (1)

For a sequence of three tokens the probability is equal to

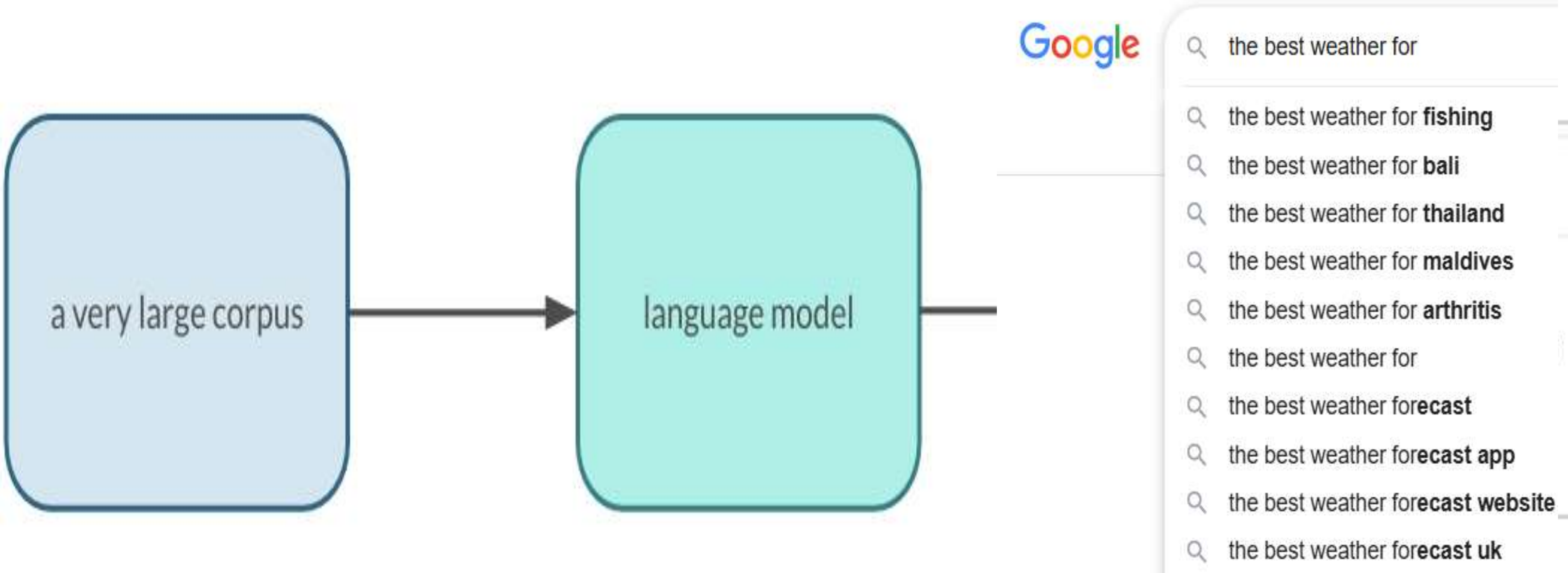
$$P\left(x^{(1)}, x^{(2)}, x^{(3)}\right) = p\left(x^{(1)}\right)p\left(x^{(2)} \mid x^{(1)}\right)p\left(x^{(3)}, \mid x^{(1)}\ x^{(2)}\right)$$

How a LLM problem is modelled?

The general case is defined as follows

$$P\left(x^{(1)}, \dots, x^{(N)}\right) = \prod_{i=1}^{N} p\left(x^{(i)} \mid x^{(i)}, \dots , x^{(i-1)} \right)$$

How a LLM problem is modelled?



“How are you this afternoon? Has your car been broken?” → P(10^{-15})

because the second sentence is **less typical or more unexpected** in a daily basis conversation.

“How are you this afternoon? It’s good to see you" → P(10^{-5})

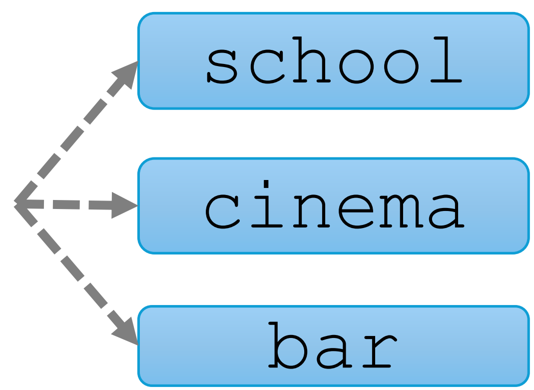
Because **it is more frequent of expected in daily bases conversations**, leading to a higher probability.

Chain rule

https://en.wikipedia.org/wiki/Chain_rule_(probability)

$$P\left(x^{(1)},\ldots,x^{(N)}\right)=\prod_{i=1}^{N}p\left(x^{(i)}\mid \underbrace{x^{(i)},\ldots,x^{(i-1)}}_{\textbf{Context}}\right)$$

The students went to ___



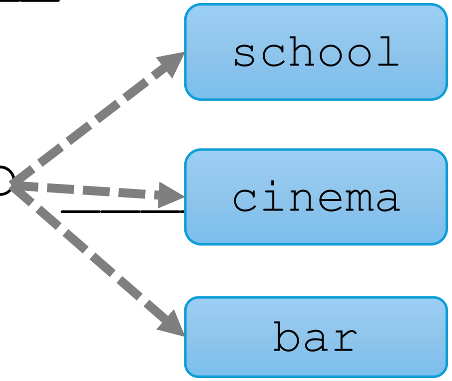
Every time a new token is added to the context, then a new probability distribution is calculated

The

The students ___

The students went ___

The students went to ___



Introduction to Large Language

Google

I love visiting

I love visiting - Google Search

i love visiting **different places because**

i love visiting **you folks in the city**

i love visiting **new places**

i love visiting **historical places**

i love visiting **museums**

i love visiting **card**

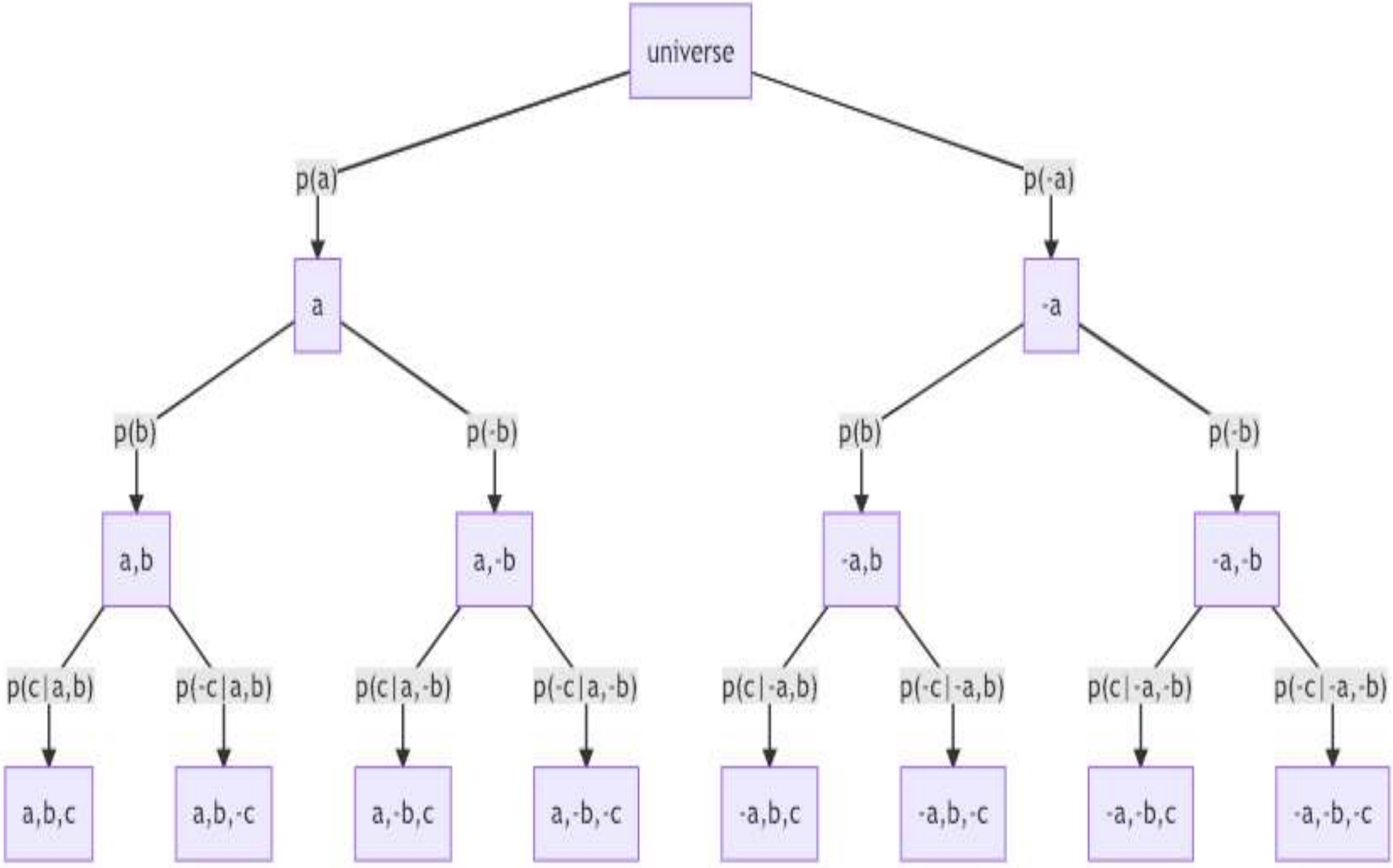
i love visiting **chicago in winter**

i love visiting **the zoo**

i love visiting **you**

$$P\left(x^{(1)}, \ldots, x^{(N)}\right) = \prod_{i=1}^{N} p\left(x^{(i)} \mid x^{(i)}, \ldots , x^{(i-1)} \right)$$

Visual representation of the



How does a Language Model “learn”?

I have a duck whose name is Rocky. I have two cats. They like playing with Rocky.

- `Corpus size:  17`
- `P(Rocky)  = 2/17`
- `P(cats) = 1/17`

I have a dog whose name is Rocky. I have two cats, they like playing with Rocky.

Bigram probability (based on this corpus).

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

$$P(\text{have} \mid \text{I}) = \frac{P(\text{I have})}{P(\text{I})} = \frac{2}{2} = 1$$

$$P(w_2|w_1) = \frac{C(w_1,w_2)}{\sum_w C(w_1,w)} = \frac{C(w_1,w_2)}{C(w_1)}$$

$$P(\text{two} \mid \text{have}) = \frac{P(\text{have two})}{P(\text{have})} = \frac{1}{2} = 0.5$$

$$P(\text{eating} \mid \text{have}) = \frac{P(\text{have eating})}{P(\text{have})} = \frac{0}{2} = 0$$

I have a dog whose name is Rocky. I have two cats, they like playing with Rocky.

Trigram probability (based on this corpus).

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

$$P(a \mid I\ have) = \frac{C(I\ have\ a)}{C(I\ have)} = \frac{1}{2} = 0.5$$

$$P(w_3|w_1\, w_2) = \frac{C(w_1,w_2,w_3)}{\sum_w C(w_1,w_2,w)} = \frac{C(w_1,w_2,w_3)}{C(w_1,w_2)}$$

$$P(several \mid I\ have) = \frac{C(I\ have\ several)}{C(I\ have)} = \frac{0}{2} = 0$$

I have a dog whose name is Rocky. I have two cats, they like playing with Rocky.

N-gram probability (based on this corpus).

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

$$P(w_i \mid w_1, w_2, \dots, w_{i-1}) = \frac{C(w_1, w_2, \dots, w_{i-1}, w_i)}{C(w_1, w_2, \dots, w_{i-1})}$$

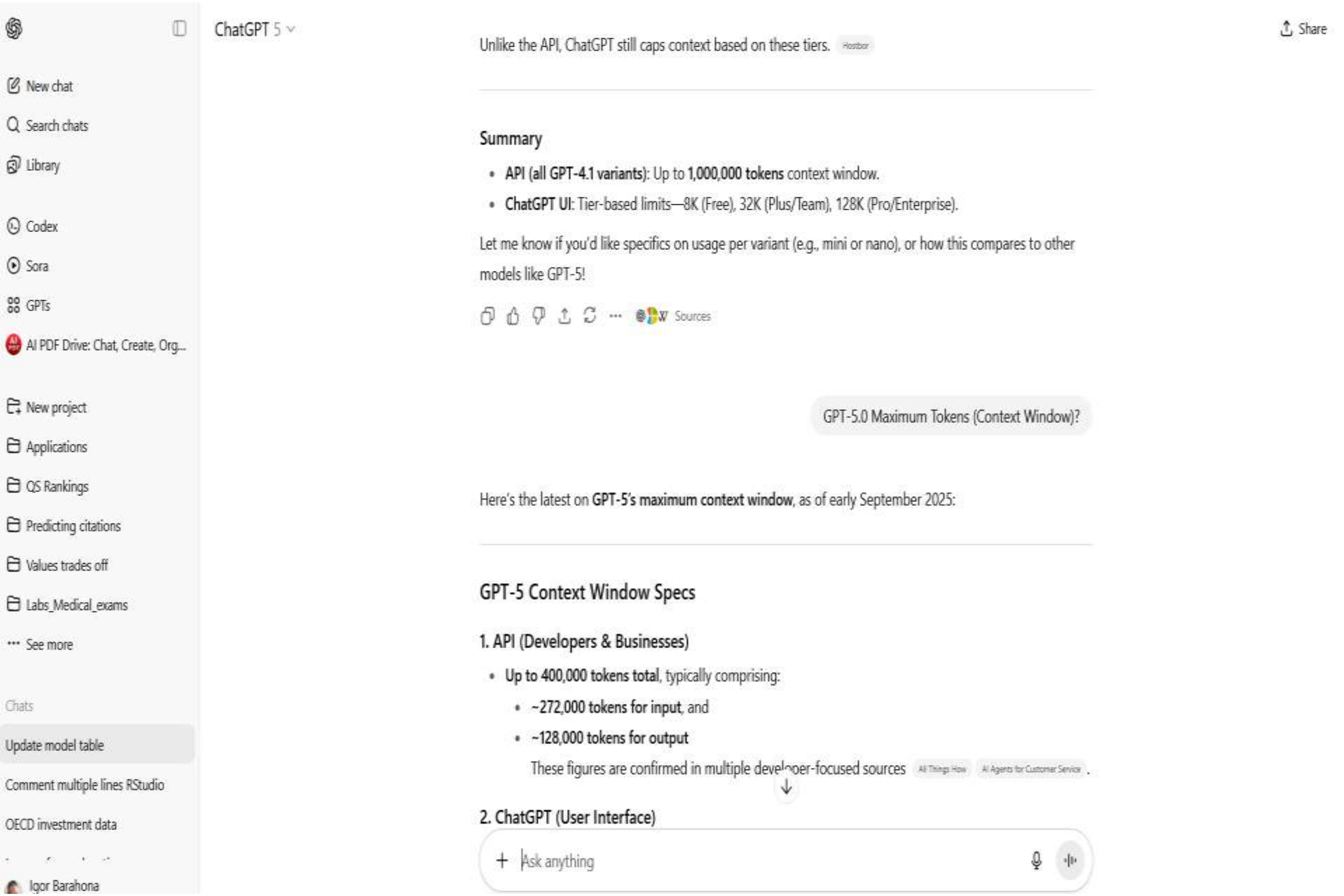
Prompt Engineering

- Prompt engineering means tailoring your questions and input so you can get the most out of an LLM

- The output of a LLM is determined by both what the **system has been trained** on and **what information you give it**

- Prompts can take many forms, from instructing the LLM to take on a role (e.g. a helpful teacher, a pirate) or guiding the way it should process its output (e.g. "chain of thought" or a particular method)

ChatGPT


ChatGPT 5
Share
New chat
Search chats
Library
Codex
Sora
GPTs
AI PDF Drive: Chat, Create, Org...
New project
Applications
QS Rankings
Predicting citations
Values trades off
Labs_Medical_exams
See more
Chats
Update model table
Comment multiple lines RStudio
OECD investment data
Igor Barahona
Unlike the API, ChatGPT still caps context based on these tiers.
Hostbor
Summary
API (all GPT-4.1 variants): Up to 1,000,000 tokens context window.
ChatGPT UI: Tier-based limits—8K (Free), 32K (Plus/Team), 128K (Pro/Enterprise).
Let me know if you'd like specifics on usage per variant (e.g., mini or nano), or how this compares to other models like GPT-5!
Sources
GPT-5.0 Maximum Tokens (Context Window)?
Here's the latest on GPT-5's maximum context window, as of early September 2025:
GPT-5 Context Window Specs
1. API (Developers & Businesses)
Up to 400,000 tokens total, typically comprising:
~272,000 tokens for input, and
~128,000 tokens for output
These figures are confirmed in multiple developer-focused sources
AI Things How
AI Agents for Customer Service
2. ChatGPT (User Interface)
Ask anything

How does AI "learn"?

Sustainability (vs) Cost

What do you guys might select?

How operations managers make value trade-offs decisions that involve sustainability and cost?

How elicitation methods impact common biases, such as sensitivity?

Motivations

Are managers capable to make consistent value trade-offs that reflect rational prioritizations?

Are importance weights and trade-offs considered consistent when attribute ranges are explicitly specified?

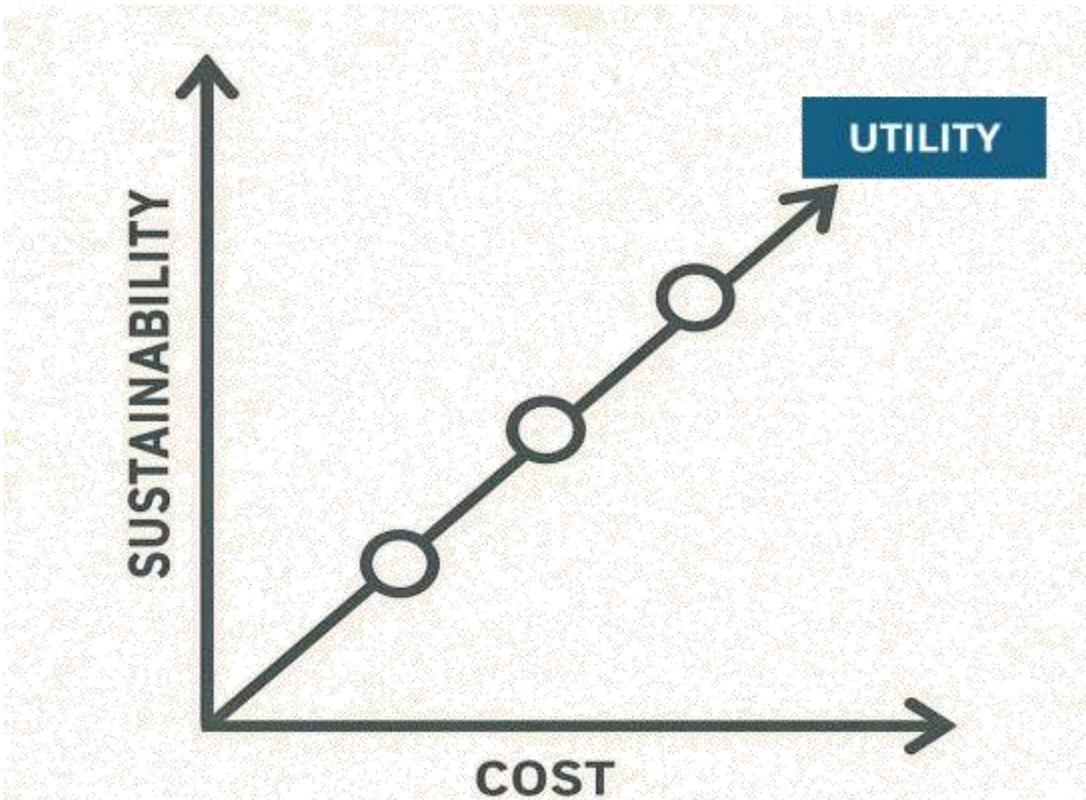
Introduction

Are managers' prioritizations between sustainability and cost conditioned on whether:

a) Personal preferences or

b) Ranges of the underlying attributes.

A naïve perspective assumes a linear relationship between **Sustainability** and **Cost**



Our experiments show that the relationship is far from linear

Methodology

- CRT assesses a person's tendency to override an initial intuitive (but wrong) answer and engage in deeper, reflective reasoning.
- The classic CRT consists of 3 questions (Frederick, 2005), but has been expanded in later versions to address issues like low score variability.
- Performance can be influenced by numerical skills, leading to concerns about confusing effects in interpretation.

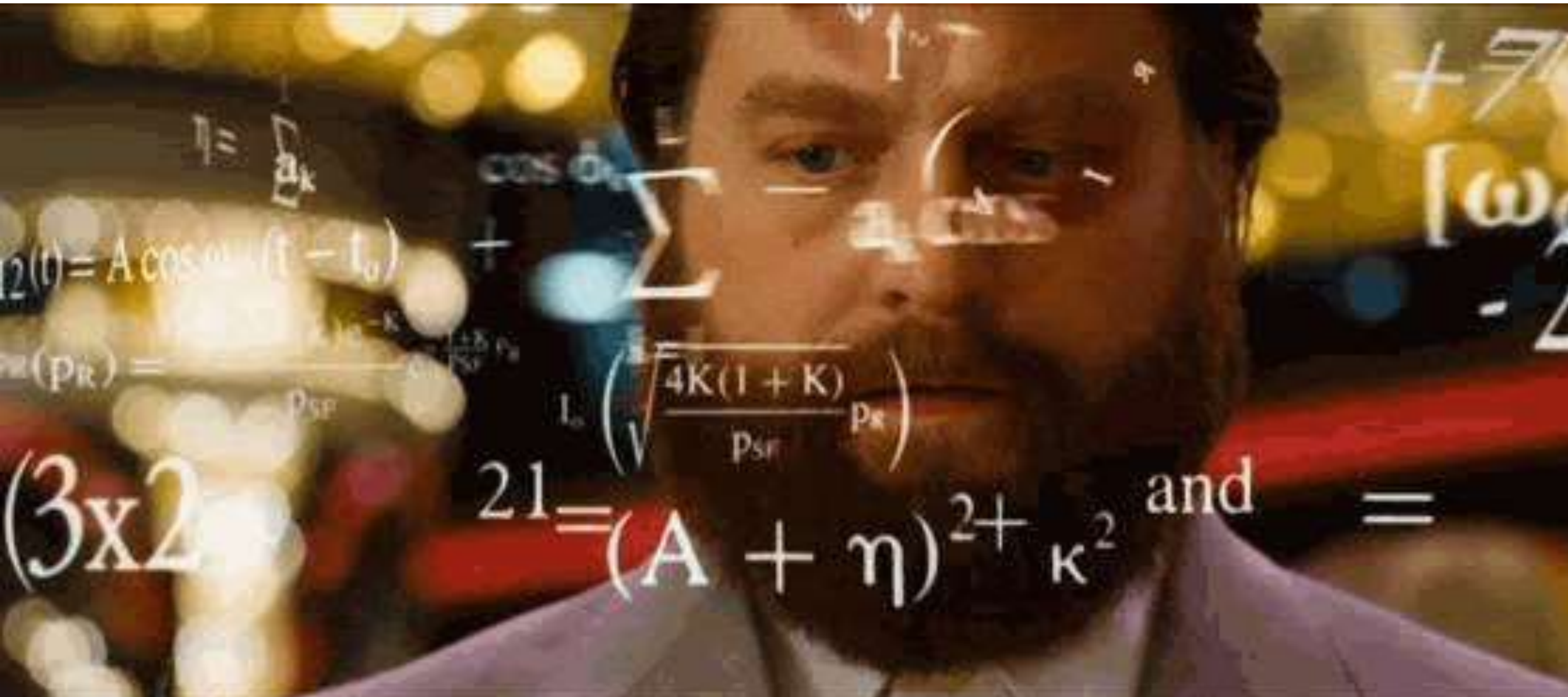
The organization has set up two strategic goals

1) Minimize emissions of the car fleet

2) Minimize acquisition costs of the fleet.

The range of costs per car is between \$30k and \$130k.

The range of emissions is between 0.0 g-CO2/km (for a fully electric vehicle) and 260.0 g CO2/km (for a large gasoline vehicle).


4K(1 + K)
(3x2
21
and
(A + η)2+ κ2