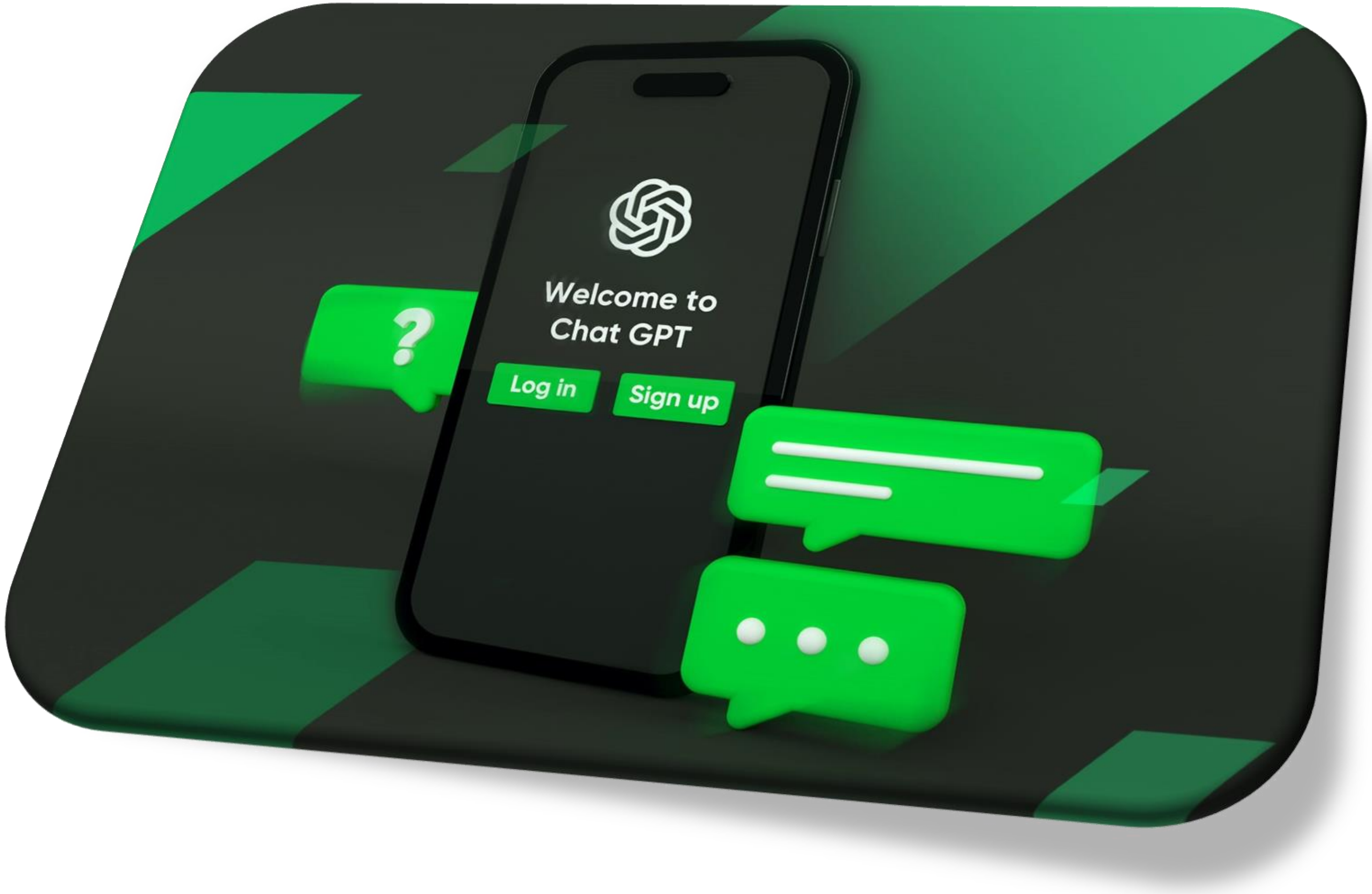
# ChatGPT
# For
# Data Scientists



دایکه
DAYCHE

Soheil Tehranipour

August 2023

# COURSE OUTLINE

| | | |
|---|---|---|
| 1 | **Why Analysts Need to Use AI** | *Explore how AI is changing the way data analysts work, and why leveraging these tools is critical to stay competitive* |
| 2 | **Intro to AI, LLMs & GPTs** | *Introduce the AI landscape and the role of large language models and generative pre-trained transformers like ChatGPT* |
| 3 | **Prompt Engineering Tips** | *Review best practices for creating clear and effective prompts when working with tools like ChatGPT or Bard* |
| 4 | **Common Use Cases** | Excel<br>Google Sheets<br>Power BI<br>MySQL<br>Python |

# سهیل تهرانی پور

- دکتری هوش مصنوعی و رباتیکز
- مدیر عامل شرکت دانش بنیان ساعیان ارتباط

  - هوش مصنوعی در مراکز فرودگاه‌های سراسر کشور، گروه صنعتی انتخاب
  - مدیر پروژه سامانه‌های مخابراتی مترو تهران
  - تولید نرم افزار شناسایی موضوع متون از با استفاده از روش های هوش مصنوعی
  - تولید نرم افزار شناسایی علایم راهنمایی و رانندگی در خودروهای بدون سرنشین
  - تولید Chatbot فارسی در حوزه سامانه‌های پروازی

- مدرس دوره‌های یادگیری ماشین و علم داده در همراه اول، وزارت نفت، بیمه ایران
- هم بنیان گذار **Iran Machine Learning**

OpenAI
ChatGPT

# WHY YOU NEED TO USE AI TOOLS?

## THE WORLD IS CHANGING

Technology like ChatGPT, Google Bard, and other Artificial Intelligence tools allow you to complete tasks that used to take hours in a matter of minutes, making it a **total game changer**

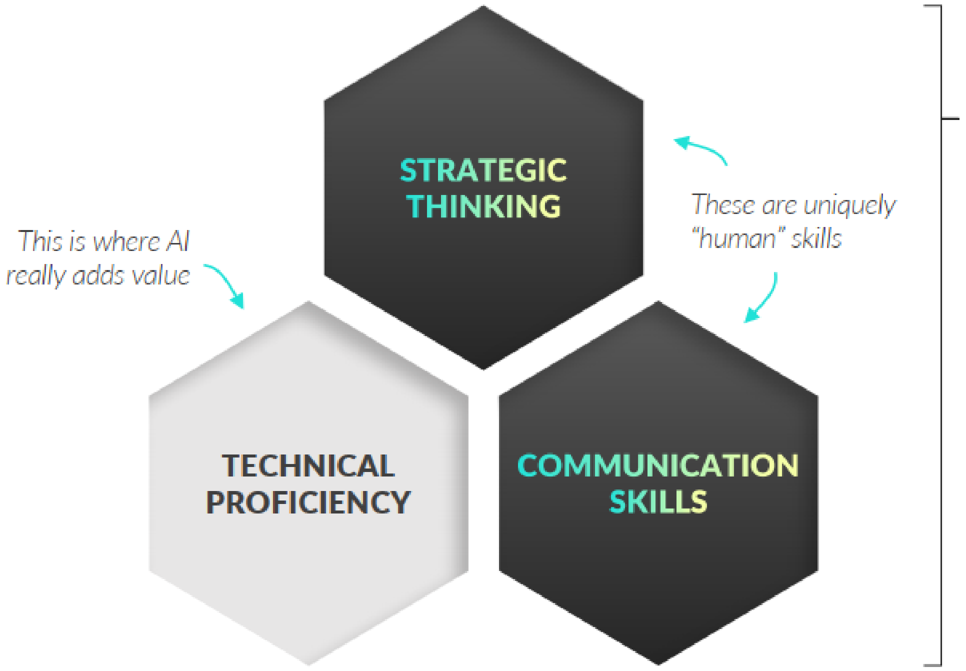
## Reasons you should learn to use AI tools:

- ✓ Improve your performance & efficiency
- ✓ Automate routine, low-value tasks
- ✓ Spend more time on the higher value activities
- ✓ Use AI as an easy way to check your work
- ✓ Leverage it for learning and answering quick questions
- ✓ Stay on the cutting edge and competitive in your field

# Our job at Risk?

## AI WILL NOT REPLACE YOU

Fears of analysts losing jobs to AI are overblown; companies will always need human talent and intuition, but the **skills required** to be an effective analyst will shift substantially


STRATEGIC THINKING
TECHNICAL PROFICIENCY
COMMUNICATION SKILLS
This is where AI really adds value
These are uniquely "human" skills

We call this the "**Analytics Trifecta**", because it represents the three core skills that separate a good analyst from a *great* one

Since AI can fill many technical gaps, analysts with exceptional strategic thinking and communication skills will become more valuable than ever to organizations

**Combine those "human" skills with a strong technical foundation enhanced by AI, and you'll be unstoppable** 💪

# Using AI for being updated

## IF YOU DON'T USE AI SOMEONE ELSE WILL

Imagine being a data analyst today without learning spreadsheets or using the internet. We'll likely be **saying the same thing about ChatGPT and other AI tools** in the near future.

**750M+**
users worldwide

**5B+**
users worldwide

**1B+**
users worldwide

The key to an enduring, successful career in Data Analytics is to build a strong foundational skill set, and **keep it sharp by embracing and adapting to modern technology**

# Where AI will SHINE

While there are *many* potential ways that data professionals can leverage AI, these are a few of the **most common and powerful use cases** we've explored:

Generating code or formulas

Explaining technical concepts

Troubleshooting or debugging

Commenting code

Optimizing queries or formulas

Generating step-by-step tutorials

Providing data visualization tips

Automating manual tasks

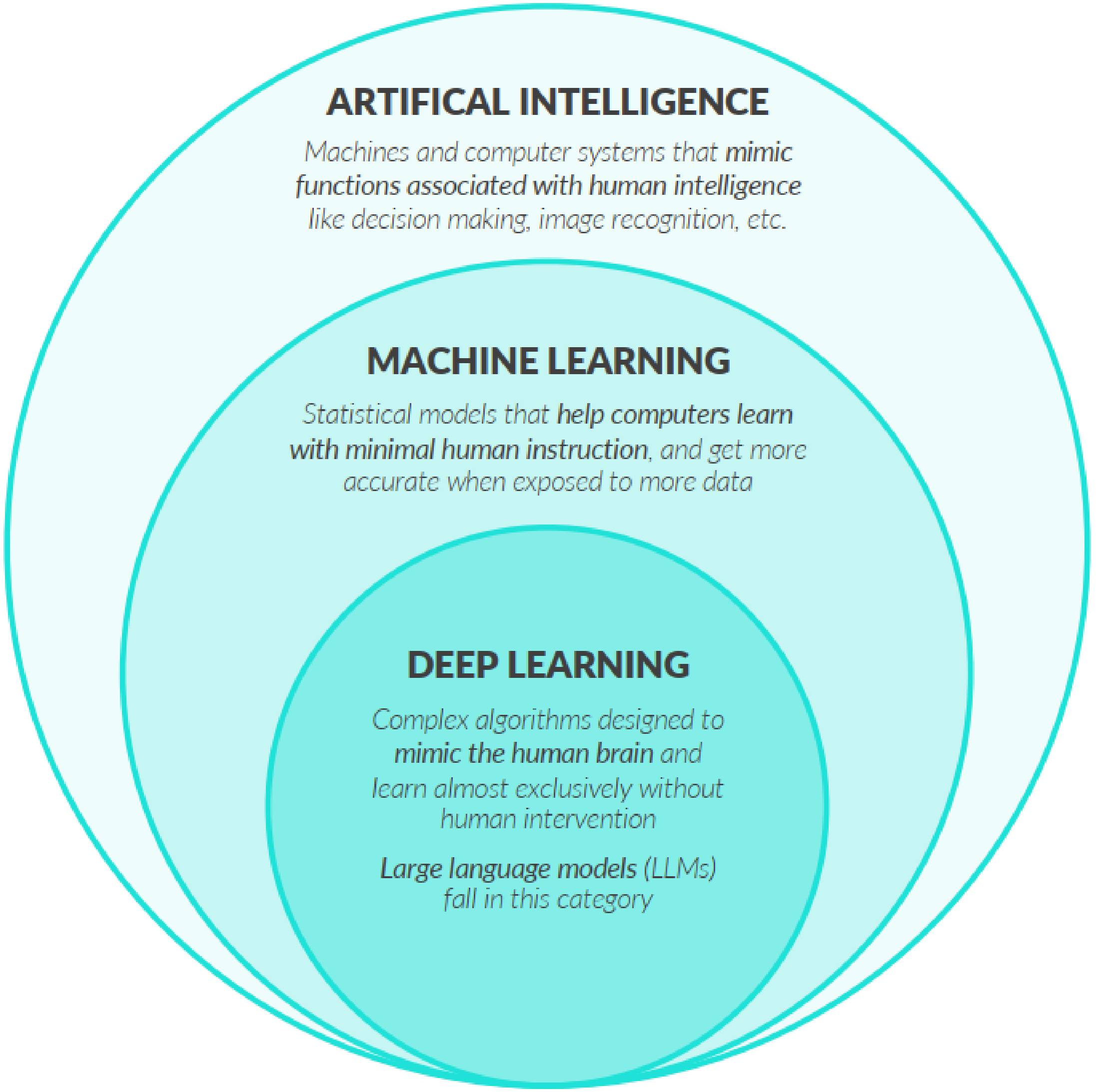
Researching technical topics

Generating data samples

# What is ChatGPT?

INTRO TO AI, LLMS & CHATGPT

# What is ChatGPT?



**ARTIFICAL INTELLIGENCE**

*Machines and computer systems that **mimic functions associated with human intelligence** like decision making, image recognition, etc.*

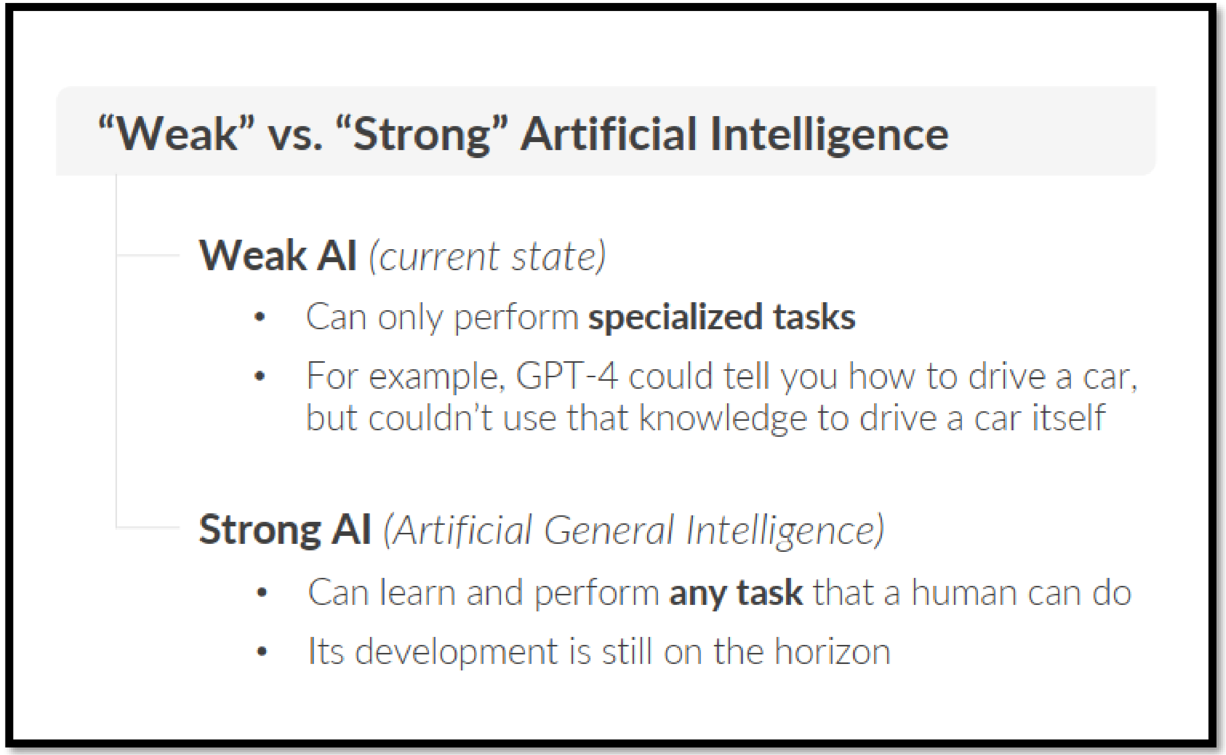
**MACHINE LEARNING**

*Statistical models that **help computers learn with minimal human instruction**, and get more accurate when exposed to more data*

**DEEP LEARNING**

*Complex algorithms designed to **mimic the human brain** and learn almost exclusively without human intervention*

***Large language models** (LLMs) fall in this category*



## "Weak" vs. "Strong" Artificial Intelligence

**Weak AI** *(current state)*

- Can only perform **specialized tasks**
- For example, GPT-4 could tell you how to drive a car, but couldn't use that knowledge to drive a car itself

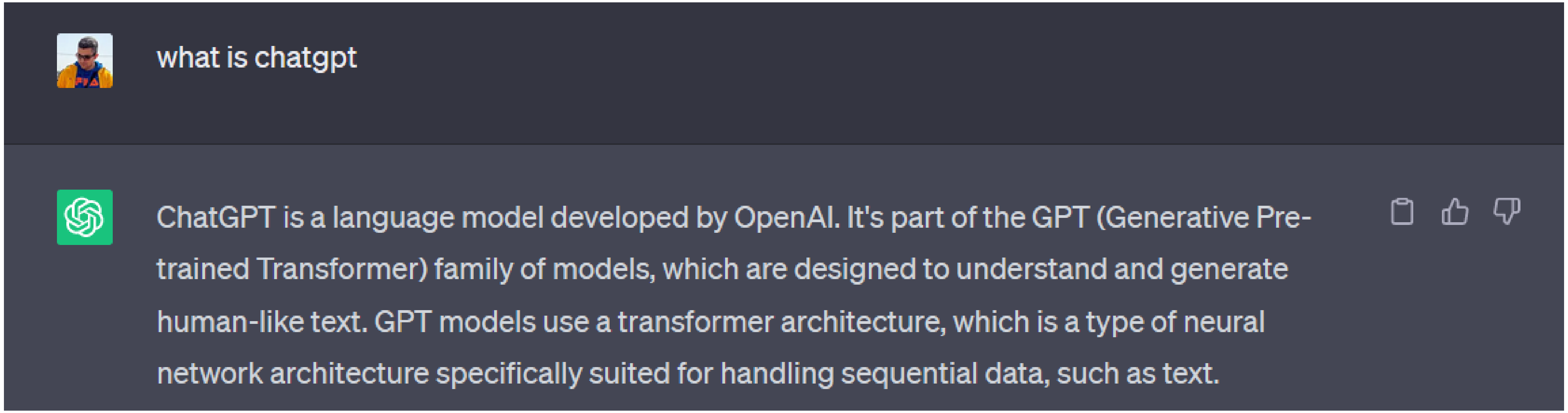
**Strong AI** *(Artificial General Intelligence)*

- Can learn and perform **any task** that a human can do
- Its development is still on the horizon

# GENERATIVE AI & LLMs

**Generative AI** systems are deep learning models capable of generating original text, images and other types of media in response to user prompts

- **Large Language Models** (LLMs) are generative AI models focused on producing text outputs specifically
- Other generative AI models include DALL-E and Midjourney, which are used to generate images



what is chatgpt

ChatGPT is a language model developed by OpenAI. It's part of the GPT (Generative Pre-trained Transformer) family of models, which are designed to understand and generate human-like text. GPT models use a transformer architecture, which is a type of neural network architecture specifically suited for handling sequential data, such as text.

# GENERATIVE PRE-TRAINED TRANSFORMERS

**Generative Pre-Trained Transformers** (GPTs) are a type of large language model trained on massive text datasets, and are designed to generate outputs that mimic human-written text

## GENERATIVE

The model generates **new and original natural language text**, instead of copying and pasting existing data

## PRE-TRAINED

The model **was already trained on a large dataset** before being fine-tuned to perform specific tasks

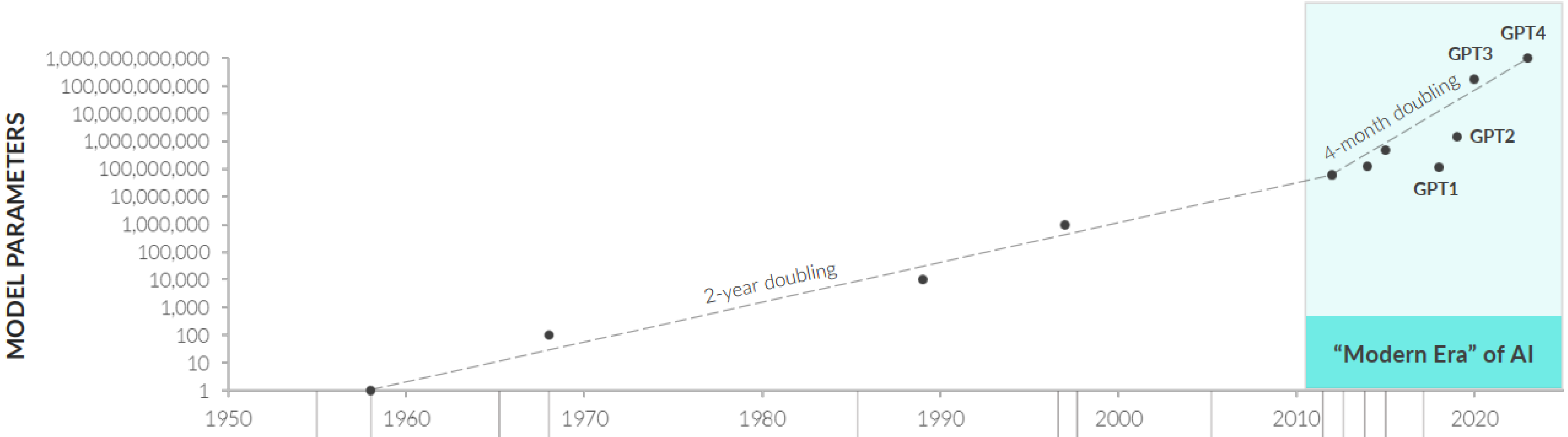
## TRANSFORMERS

A type of deep learning model that can **process sequential inputs** and differentiate the importance of individual parts *(also known as self-attention)*

LLMs like ChatGPT are among the **most sophisticated deep learning models** ever built; GPT-4, which powers premium versions of ChatGPT, has over 1 TRILLION parameters, cost over $100 million dollars, and took 11 months to train

# History of ChatGPT

AI tools like **ChatGPT** became widely popular in late 2022, but owe their success to more than **60 years of research and development** in artificial intelligence systems



*The term "Artificial Intelligence" is used for the first time*

*The first neural network algorithm is implemented*

***ELIZA, the first chatbot, is created at MIT***

*Researchers find major flaws in the Perceptron algorithm*

*Multi-layer Perceptrons & Backpropagation implemented in the same algorithm*

***Deep Blue beats chess champion Gary Kasparov***

***Neural networks outperform other solutions in document recognition***

*The first neural network with deep learning capabilities is developed*

***IBM Watson beats humans at Jeopardy!***

*AlexNet sets a standard for image recognition*

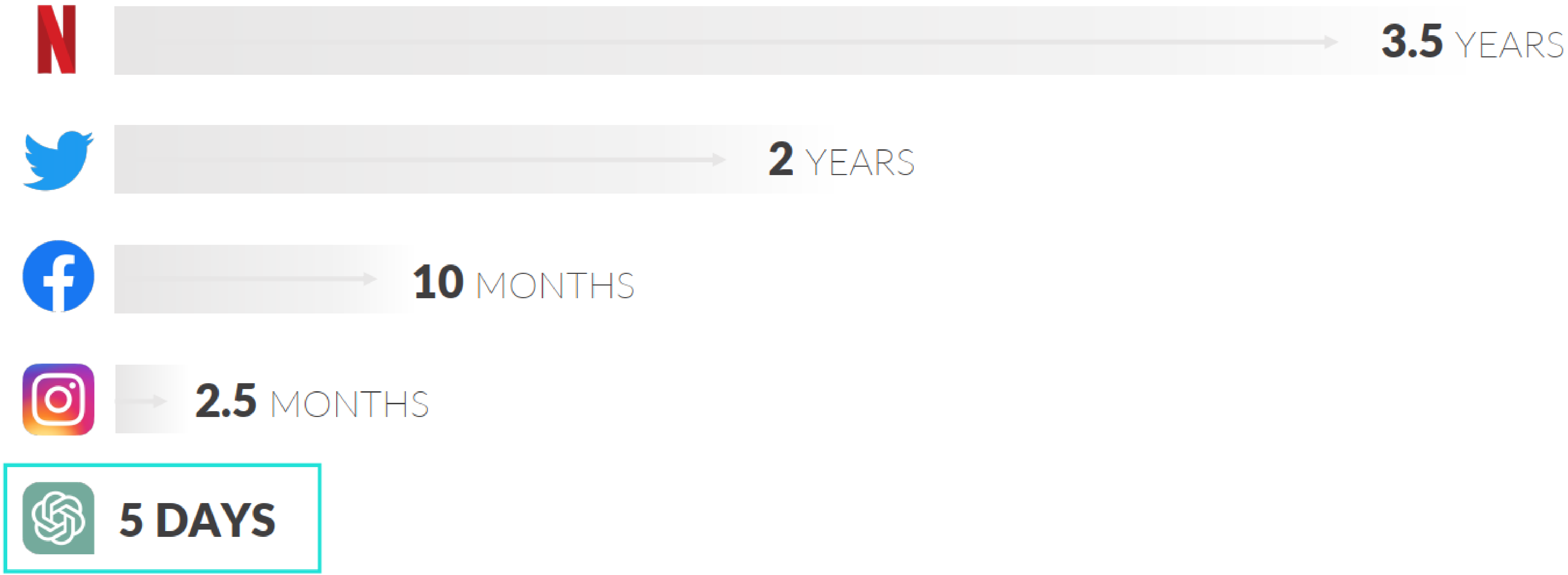
***Siri & Alexa are launched***

***Open AI founded, AlphaGo beats world champions at Go***

*Transformer Layers are introduced in Neural Nets*

# Growth of ChatGPT

The rate of adoption for modern AI tools is unprecedented; ChatGPT became the fastest-growing online product in history, reaching **1 million users in only 5 days**



# ChatGPT rivals

VS
VS
Bard
Powered By Google
Chat GPT
Powered By Open AI
Bing Chat
Powered By Microsoft

# WARNING: COMMON PITFALLS

⚠ LLMs are known to **“hallucinate” facts** with total confidence

- *Remember that YOU are ultimately responsible for verifying the accuracy of model outputs*

⚠ Solutions provided **may be suboptimal or entirely incorrect**

- *AI tools don’t guarantee accuracy, and may provide incorrect or inefficient solutions*

⚠ These tools are broad and often **lack specific domain knowledge**

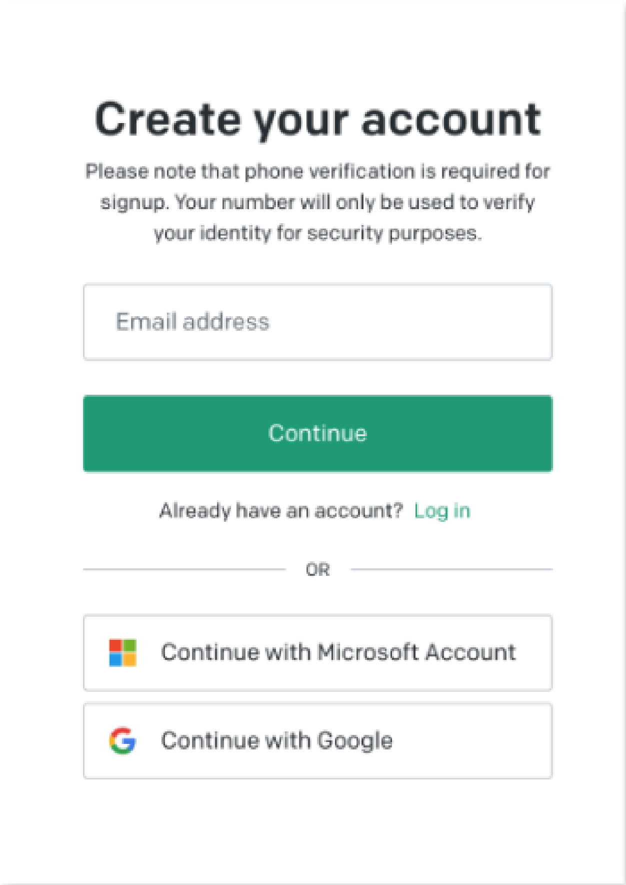
- *LLMs may not understand specific business context or the “why” behind the responses they produce*

⚠ LLMs are NOT capable of **common sense** or **human judgement**

- *Models need specific, objective inputs, and may miss critical context that may seem obvious to humans*

# How to register

You can access ChatGPT for free at **chat.openai.com**



Create your account

Please note that phone verification is required for signup. Your number will only be used to verify your identity for security purposes.

Already have an account? Log in

- Open AI currently allows anyone to sign up and use ChatGPT for free *(note that there is a paid version as well)*
- You can create an account using an email address, or by connecting an existing Microsoft or Google account
- Read more information about ChatGPT on the OpenAI website: **https://openai.com/blog/chatgpt**

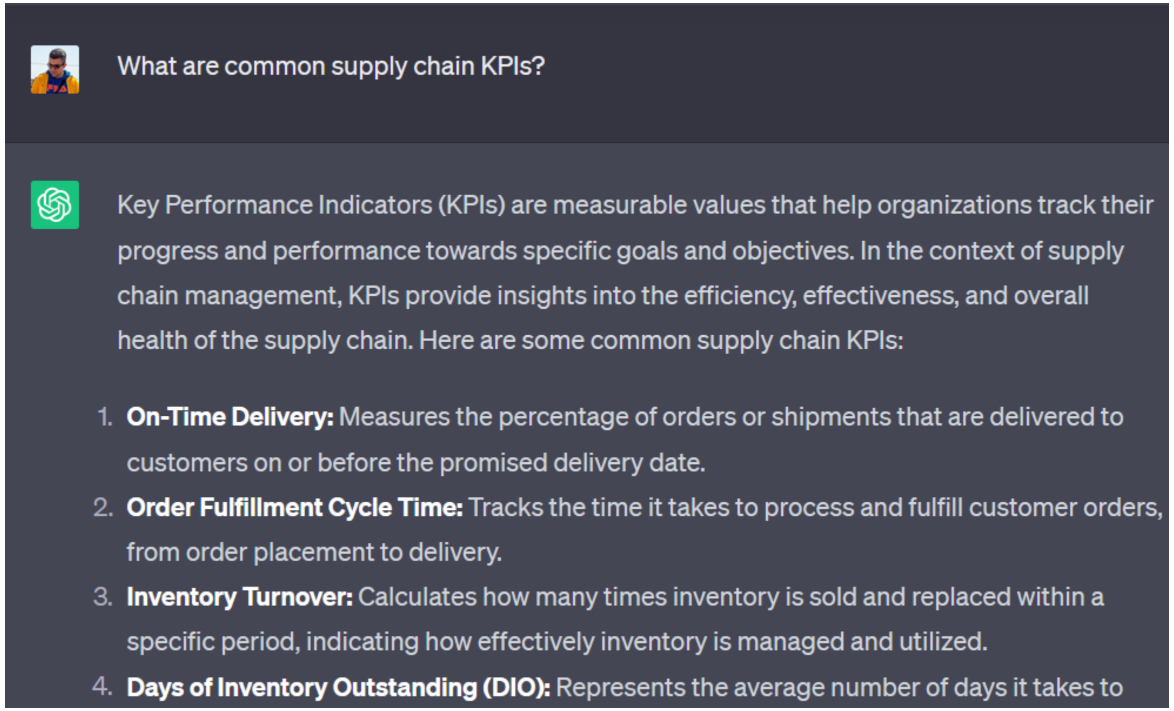
# PROMPT ENGINEERING

# Prompt Engineering – Provide Context

ORIGINAL PROMPT

VS

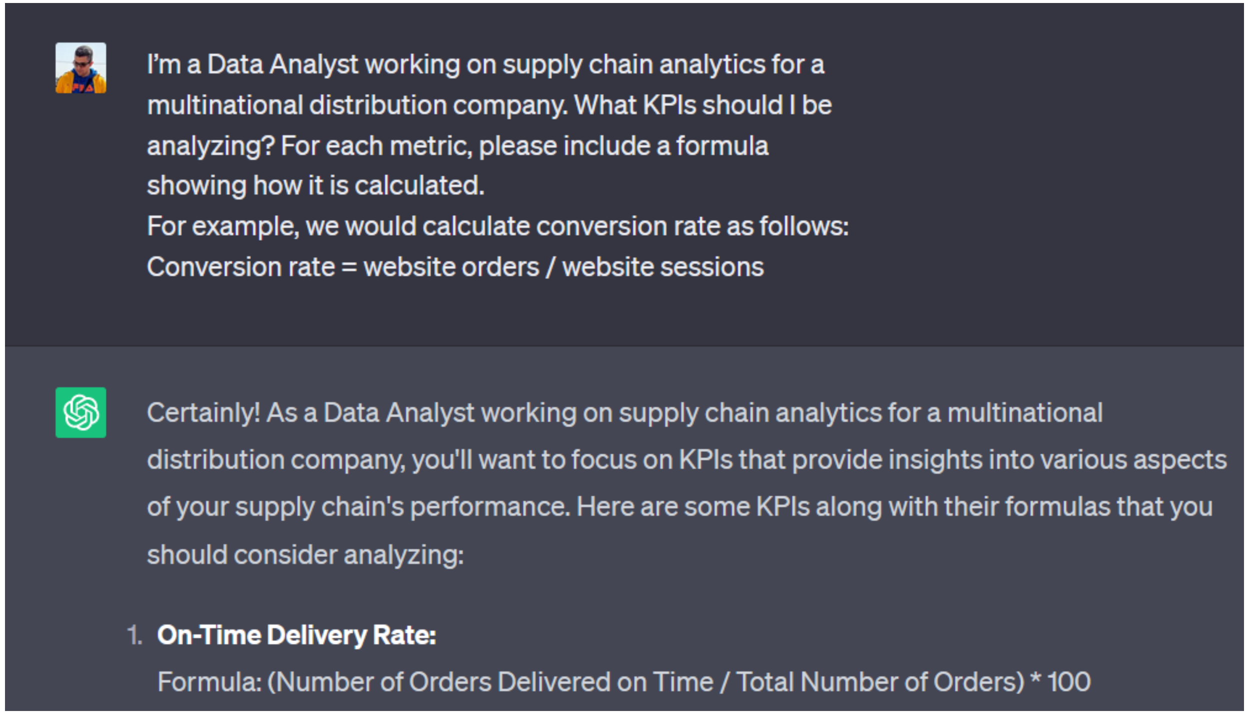
BETTER PROMPT



What are common supply chain KPIs?

Key Performance Indicators (KPIs) are measurable values that help organizations track their progress and performance towards specific goals and objectives. In the context of supply chain management, KPIs provide insights into the efficiency, effectiveness, and overall health of the supply chain. Here are some common supply chain KPIs:

1. **On-Time Delivery:** Measures the percentage of orders or shipments that are delivered to customers on or before the promised delivery date.
2. **Order Fulfillment Cycle Time:** Tracks the time it takes to process and fulfill customer orders, from order placement to delivery.
3. **Inventory Turnover:** Calculates how many times inventory is sold and replaced within a specific period, indicating how effectively inventory is managed and utilized.
4. **Days of Inventory Outstanding (DIO):** Represents the average number of days it takes to



I'm a Data Analyst working on supply chain analytics for a multinational distribution company. What KPIs should I be analyzing? For each metric, please include a formula showing how it is calculated.
For example, we would calculate conversion rate as follows:
Conversion rate = website orders / website sessions

Certainly! As a Data Analyst working on supply chain analytics for a multinational distribution company, you'll want to focus on KPIs that provide insights into various aspects of your supply chain's performance. Here are some KPIs along with their formulas that you should consider analyzing:

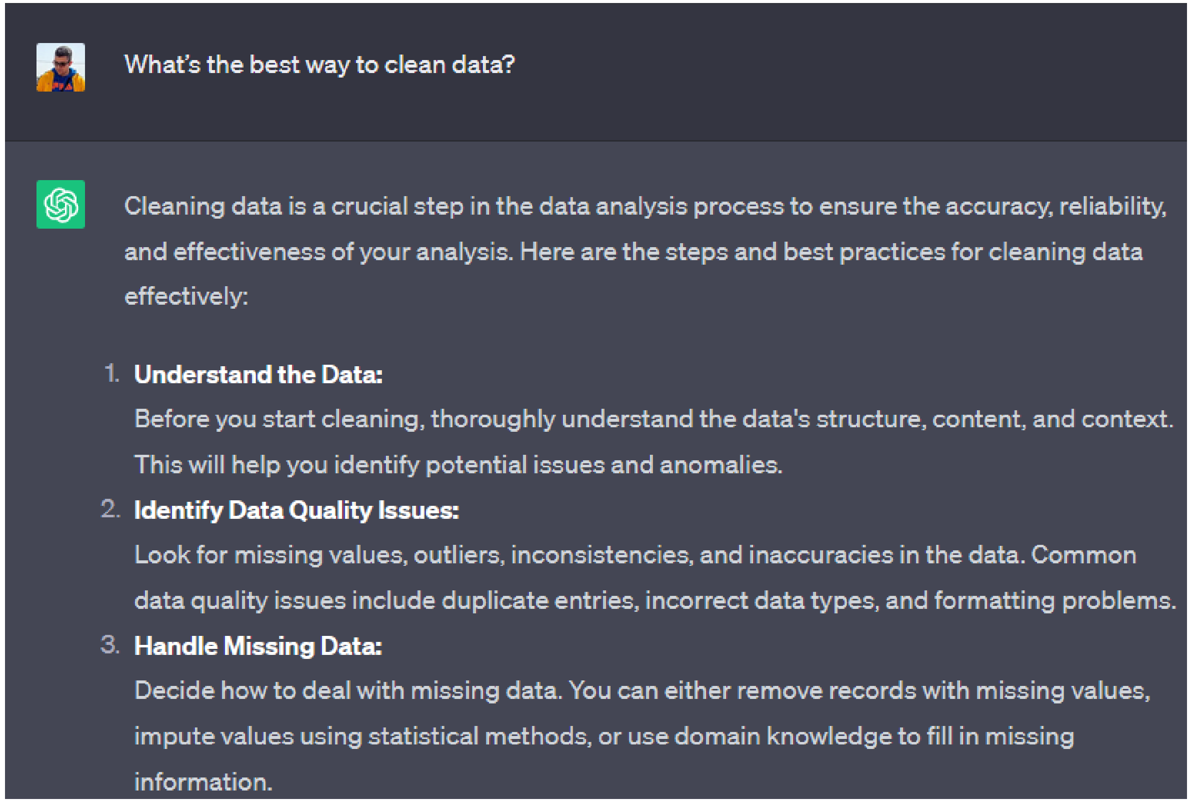
1. **On-Time Delivery Rate:**
   Formula: (Number of Orders Delivered on Time / Total Number of Orders) * 100

# Prompt Engineering – Provide Roles

ORIGINAL PROMPT

VS

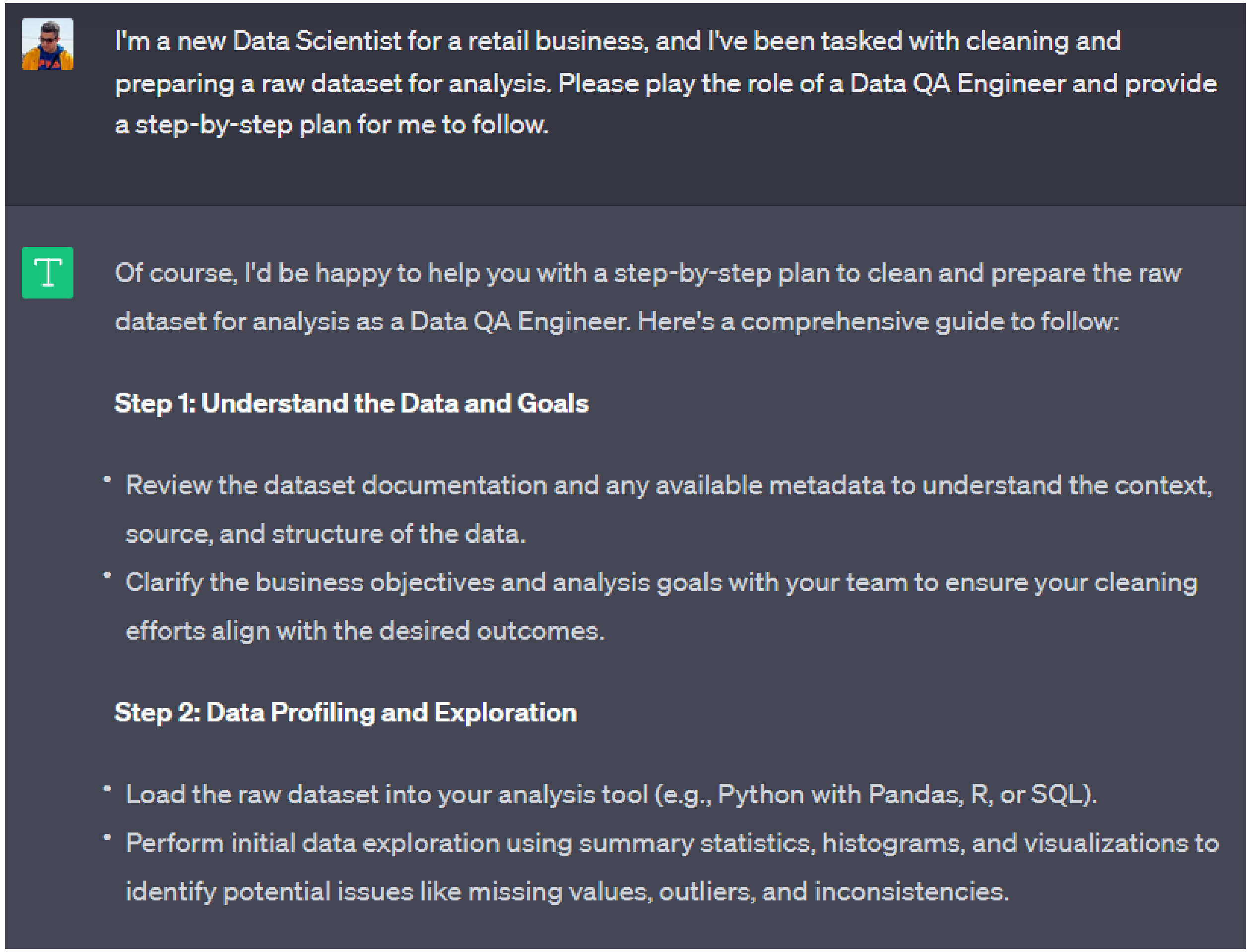
BETTER PROMPT



What's the best way to clean data?

Cleaning data is a crucial step in the data analysis process to ensure the accuracy, reliability, and effectiveness of your analysis. Here are the steps and best practices for cleaning data effectively:

1. **Understand the Data:**
   Before you start cleaning, thoroughly understand the data's structure, content, and context. This will help you identify potential issues and anomalies.
2. **Identify Data Quality Issues:**
   Look for missing values, outliers, inconsistencies, and inaccuracies in the data. Common data quality issues include duplicate entries, incorrect data types, and formatting problems.
3. **Handle Missing Data:**
   Decide how to deal with missing data. You can either remove records with missing values, impute values using statistical methods, or use domain knowledge to fill in missing information.



I'm a new Data Scientist for a retail business, and I've been tasked with cleaning and preparing a raw dataset for analysis. Please play the role of a Data QA Engineer and provide a step-by-step plan for me to follow.

Of course, I'd be happy to help you with a step-by-step plan to clean and prepare the raw dataset for analysis as a Data QA Engineer. Here's a comprehensive guide to follow:

**Step 1: Understand the Data and Goals**

- Review the dataset documentation and any available metadata to understand the context, source, and structure of the data.
- Clarify the business objectives and analysis goals with your team to ensure your cleaning efforts align with the desired outcomes.

**Step 2: Data Profiling and Exploration**

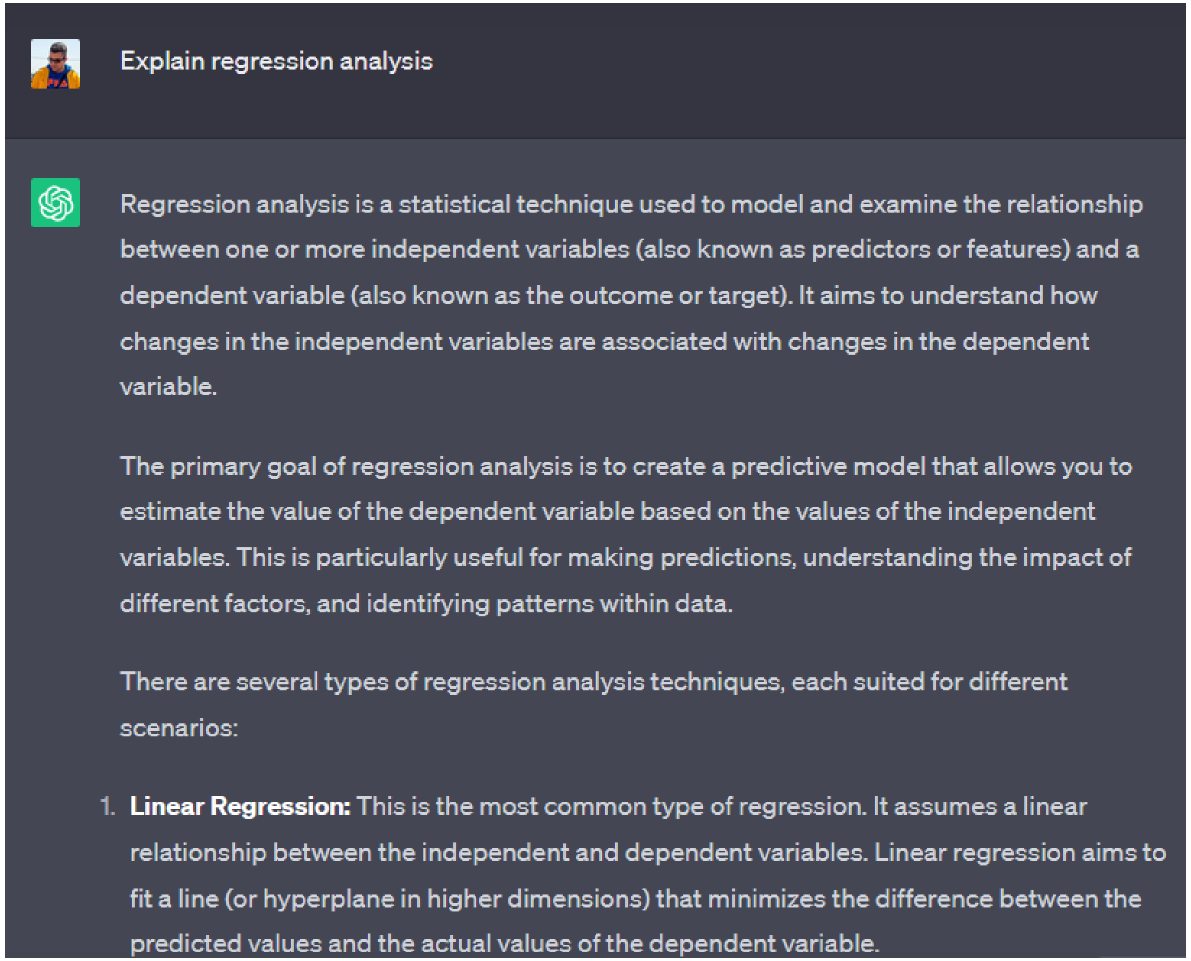
- Load the raw dataset into your analysis tool (e.g., Python with Pandas, R, or SQL).
- Perform initial data exploration using summary statistics, histograms, and visualizations to identify potential issues like missing values, outliers, and inconsistencies.

# Prompt Engineering – Provide the Tone

ORIGINAL PROMPT

VS

BETTER PROMPT



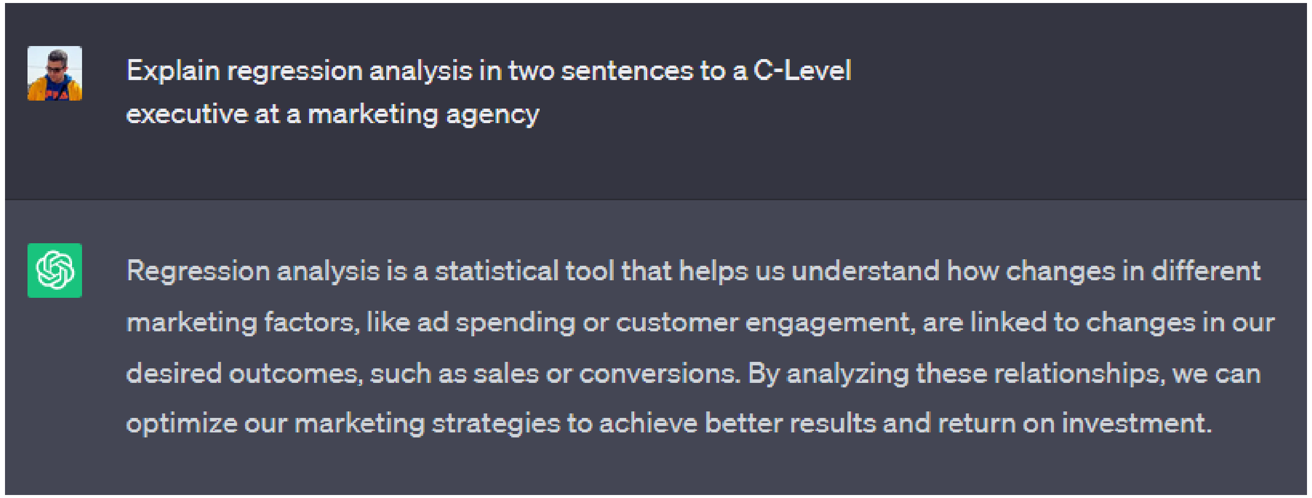
Explain regression analysis

Regression analysis is a statistical technique used to model and examine the relationship between one or more independent variables (also known as predictors or features) and a dependent variable (also known as the outcome or target). It aims to understand how changes in the independent variables are associated with changes in the dependent variable.

The primary goal of regression analysis is to create a predictive model that allows you to estimate the value of the dependent variable based on the values of the independent variables. This is particularly useful for making predictions, understanding the impact of different factors, and identifying patterns within data.

There are several types of regression analysis techniques, each suited for different scenarios:

1. **Linear Regression:** This is the most common type of regression. It assumes a linear relationship between the independent and dependent variables. Linear regression aims to fit a line (or hyperplane in higher dimensions) that minimizes the difference between the predicted values and the actual values of the dependent variable.



Explain regression analysis in two sentences to a C-Level executive at a marketing agency

Regression analysis is a statistical tool that helps us understand how changes in different marketing factors, like ad spending or customer engagement, are linked to changes in our desired outcomes, such as sales or conversions. By analyzing these relationships, we can optimize our marketing strategies to achieve better results and return on investment.

**PRO TIP**:
Use prompts like *"explain it to me as if you were speaking to a child"* to adjust the tone and complexity of the response

# Let’s get our hands dirty…

1. Explain the python code
2. Add some Comment to the code
3. Optimize the code
4. Troubleshooting code errors

ممنون
Dankie
Gracias
Спасибо
شكراً
Merci
Takk
Köszönjük
Terima kasih
Grazie
Dziękujemy
Dėkojame
Ďakujeme
Vielen Dank
Paldies
Kiitos
Täname teid
谢谢
Thank You
Tak
感謝您
Obrigado
Teşekkür Ederiz
감사합니다
Σας Ευχαριστούμ
ขอบคุณ
Bedankt
Děkujeme vám
ありがとうございます
Tack