

•LIVE

AI for Data Analytics

How To 10x Your Data Analysis
Productivity With (Generative) AI

About me



Tobias Zwingmann

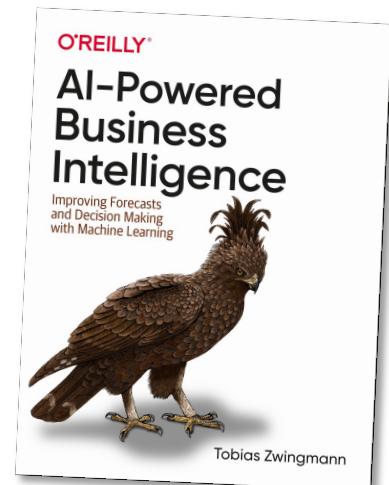
AI Advisor, Author

Managing Partner @ RAPYD.AI

15+ years corporate experience

Author: *AI-Powered Business Intelligence*

Connect: [in/tobias-zwingmann/](#)



Agenda

1. Intro to AI (15m)
2. Use Case: ChatGPT Advanced Data Analysis (10m)
3. Use Case: Problem solving with AI help (10m)
4. Use Case: AI in spreadsheets (10m)
5. Q & A (15m)

Course Overview

Learning goals:

- Understand the potential of AI for data analytics
- Use ChatGPT for automated data analysis
- Leverage AI for data-driven problem solving
- Integrate AI into your workflow with spreadsheets

Course Overview

Prerequisites

- A free OpenAI account for ChatGPT access
(ideally ChatGPT Plus subscription)
- A free Google account to access Google Sheets



Discussion:

**What's your ChatGPT
experience level?**

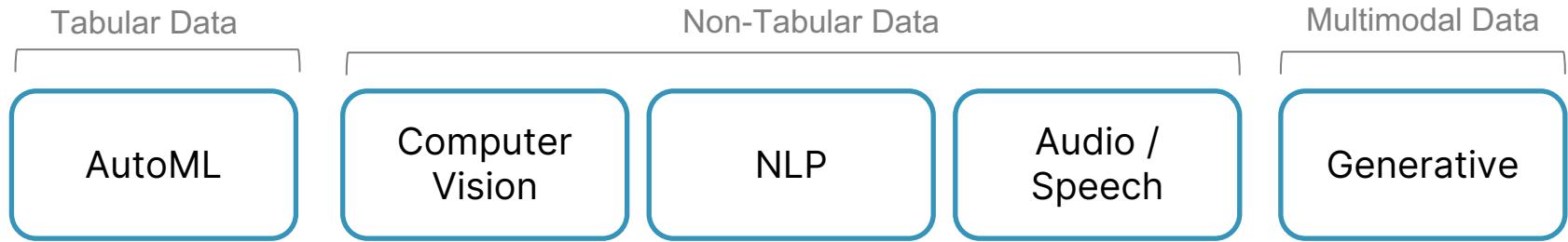


Poll:

- A) I use ChatGPT every day**
- B) I use ChatGPT at least once a week**
- C) I use ChatGPT at least once a month**
- D) I've used it maybe 1-2 times or less in total**

1. AI Fundamentals

The 5 AI Archetypes



Auto ML

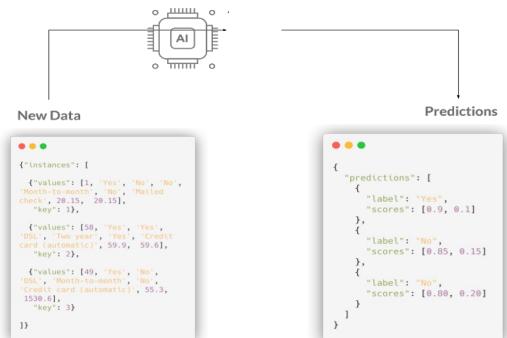
- Automate a supervised machine learning workflow for a given dataset to find the optimal fitting model for a target variable
- “Solve the problem by brute force”

Use cases:

- Predict classes / categories
- Predict numeric data
- Impute missing values
- Time series forecasting
- Recommendation systems
- ...

Variable	Type
tenure	Numeric
PhoneService	Categorical
MultipleLines	Categorical
InternetService	Categorical
Contract	Categorical
PaperlessBilling	Categorical
PaymentMethod	Categorical
MonthlyCharges	Numeric
TotalCharges	Numeric
Churn	Categoric
TARGET	Churn

The table shows a sample of data from a telecommunications dataset. It includes columns for tenure (1 to 30), phone service (No, Yes), multiple lines (DSL, Fiber optic), internet service (DSL, Fiber optic), contract type (Month-to-month, One year, Two years), paperless billing (Yes, No), payment method (Electronic check, Mailed check, Credit card (automatic)), monthly charges (20.85 to 1530.61), total charges (20.85 to 1530.61), and churn status (Yes, No). The TARGET column indicates the actual churn status for each row.

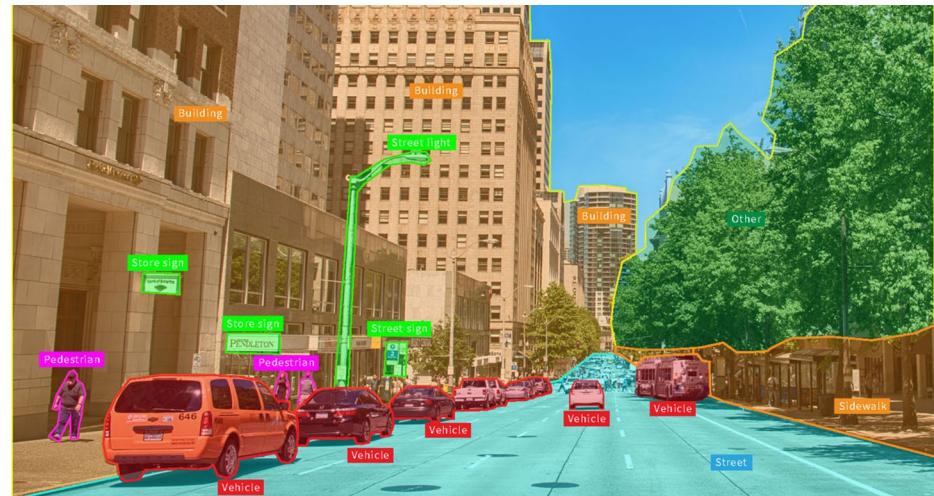


Computer Vision

- Let machines 'see' your images or documents in order to analyze, detect or extract meaningful information
- Tasks: Object detection, face recognition, text extraction, landmark detection, ...
- Driven by development in autonomous driving

Use cases:

- Identify entities for downstream processing (e.g. count objects)
- Classify images
- Remove PII
- ...



- Source: <https://appen.com/blog/computer-vision-vs-machine-vision/>

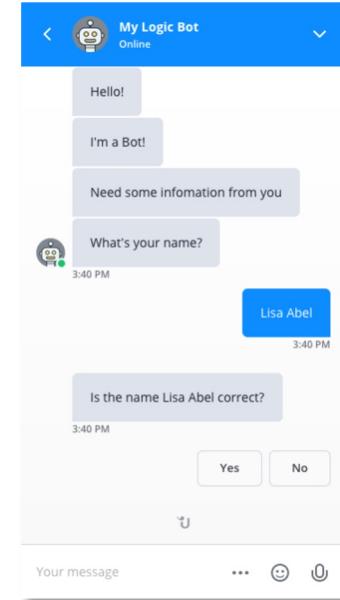
Natural Language Processing

- NLP lets computers deal with human language
- Gives machines the ability to analyze, interpret and generate human language
- Typical tasks: Sentiment analysis, entities, key phrases, summaries, answer retrieval, translation

Use cases:

- Analyze customer feedback
- Categorize content and provide context
- Chat bots
- ...

Next summer **DATE** I want to visit
Barcelona **LOCATION** and **Lion** **LOCATION**.



Audio / Speech

- Gives computers the ability to transcribe voice recordings and vice versa
- Often applied in combination with NLP
- Typical tasks: transcription, audio enhancement, filtering

Use cases:

- Transcribe sales calls
- Listen to articles
- Voice bots
- ...



Generative AI

Our
Focus
Today!

- Generate media files or text data using AI
- Often based on transformer architectures or GANs
- Took off since release of Google's BERT in 2018
(Bidirectional Encoder Representations from
Transformers)

Use cases:

- Text generation / completion
- Speech synthesis
- Realistic video avatar generation
- Video compression
- ...

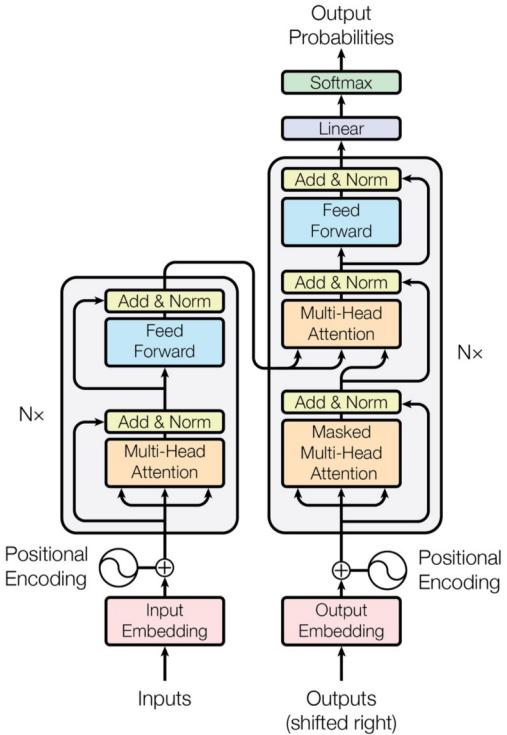
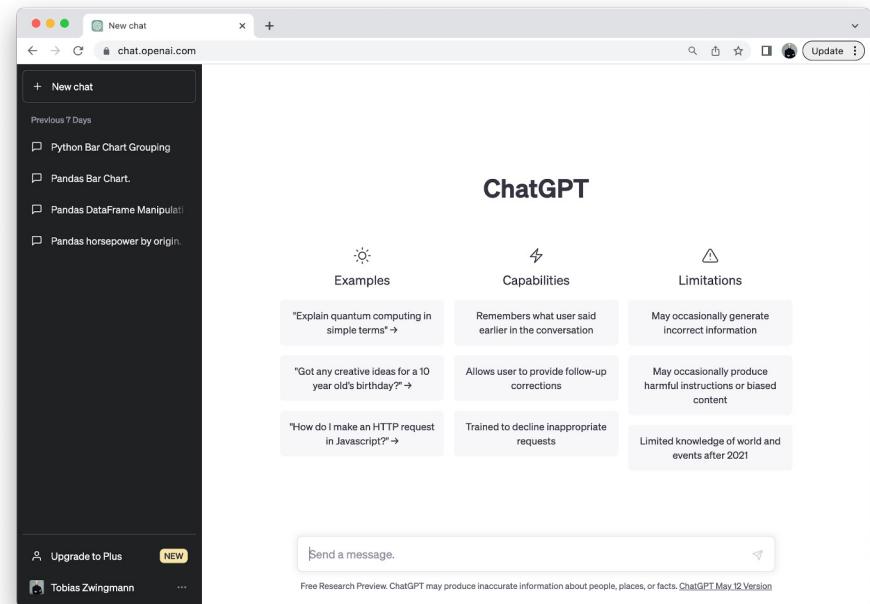


Figure 1: The Transformer - model architecture.

The Most Popular Generative AI App

- ChatGPT is an AI-powered web application by OpenAI, a US-based for-profit company
- It can respond to any instructions
-and sometimes it's right! ;-)
- Knowing how to use it is key to get valuable outcomes!
- ChatGPT is powered by a large language model
 - GPT-3.5-Turbo
 - GPT-4

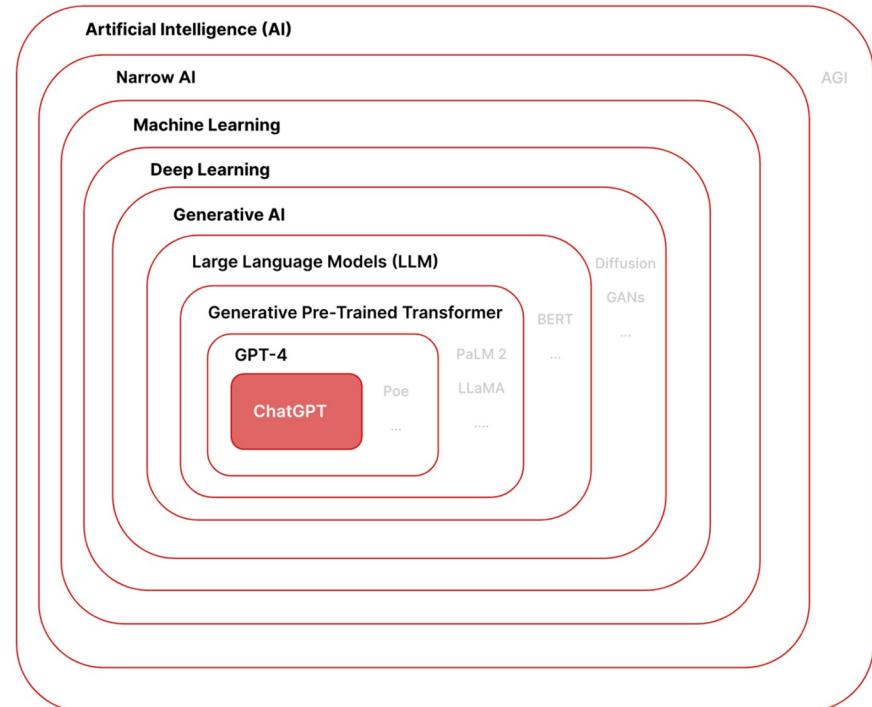


Some Terminology

- **Generative AI:** AI trained to generate digital content (text, image, video, audio).
- **LLMs:** Specific Generative AI application for text
- **GPTs** are a special category of LLMs, trained using the transformer architecture and a huge corpus of text. Examples: GPT-4, PaLM, LLaMA, Claude, ...
- **ChatGPT** is a web app that is built on GPT-3.5/4

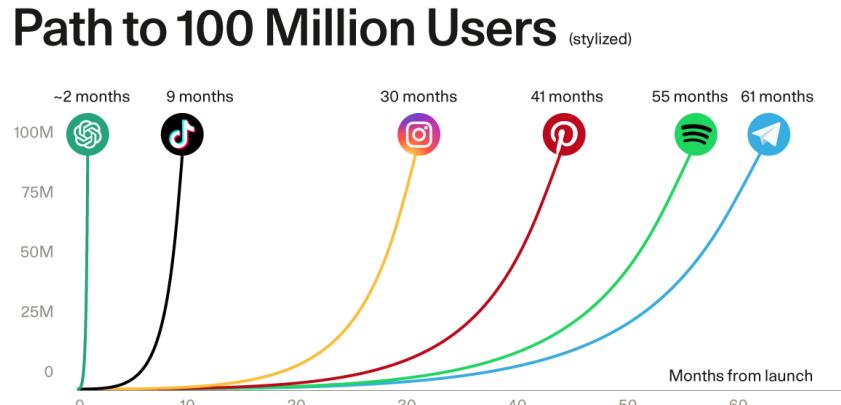
Main features:

- **Contextual Understanding:** The model doesn't just look at individual words, but also the context around them.
- **No Personal Memory:** LLMs don't have personal experiences or memories. They generate responses based on patterns they've learned from the data.



What's so special about ChatGPT?

- ChatGPT acquired 1 Million users in only about 5 days
- Set a record for the fastest app to reach 100 million users, faster than Google+ (1+ year)
- Meanwhile, growth has slowed down, partly due to competition and stronger integration



- Source: <https://www.sequoiacap.com/article/generative-ai-act-two/>

What can it do?

- Answer questions
- Provide recommendations
- Simulate conversations
- Generate stories
- Translate languages
- Offer explanations
- Assist with research
- Summarize articles
- Generate poetry
- Edit text
- Proofread documents
- ...

- Create logos
- Generate product names
- Draft emails
- Design graphics
- Teach languages
- Create quizzes
- Provide technical support
- Generate taglines
- Create social media posts
- Solve math problems
- Generate book titles
- Assist with legal research
- Generate conversation starters
- Write speeches
- Offer nutrition advice
- Provide fitness routines
- Generate song lyrics
- Assist with meditation
- Help with time management
- Create characters
- Generate blog topics
- Assist with financial planning

 Regenerate response

Send a message.



ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT May 12 Version](#)

How was it built?

Data Sources:

- **Common Crawl** corpus contains petabytes of data collected over 8 years of web crawling. The corpus contains raw web page data, metadata extracts and text extracts with light filtering.
- **WebText2** is the text of web pages from all outbound Reddit links from posts with 3+ upvotes.
- **Books1 & Books2** are two internet-based books corpora.
- **Wikipedia** pages in the English language are also part of the training corpus....

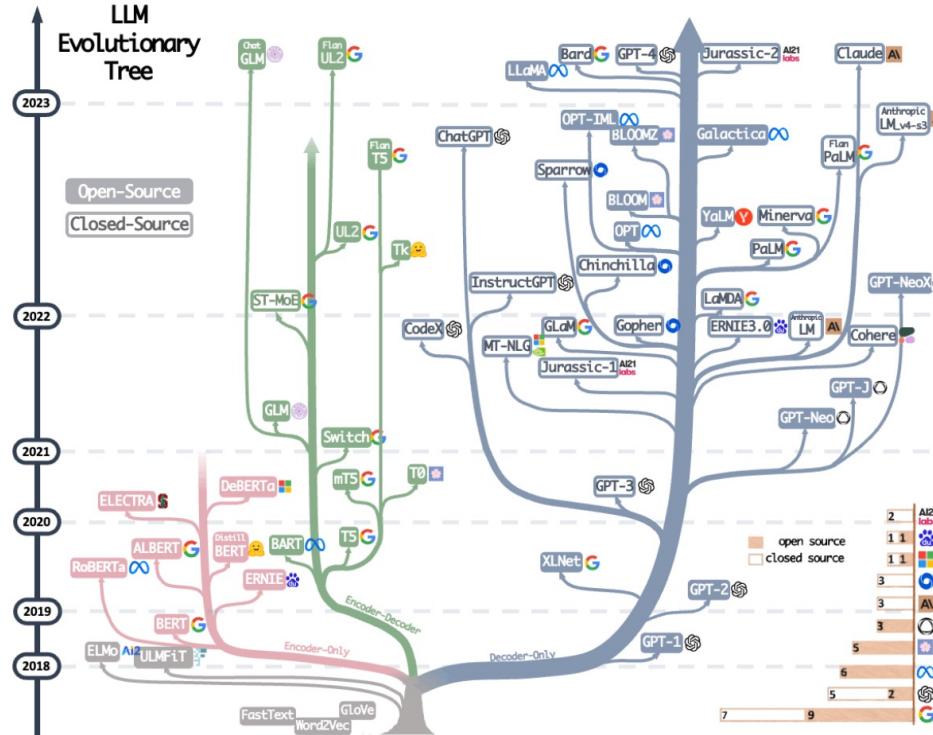
Dataset	Quantity (tokens)	Weight in training mix
<i>Common Crawl (filtered)</i>	410 billion	60%
<i>WebText2</i>	19 billion	22%
<i>Books1</i>	12 billion	8%
<i>Books2</i>	55 billion	8%
<i>Wikipedia</i>	3 billion	3%

Last training: September 2021

Training a model like GPT-4 costs around 100M+ USD.

- Source: <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>

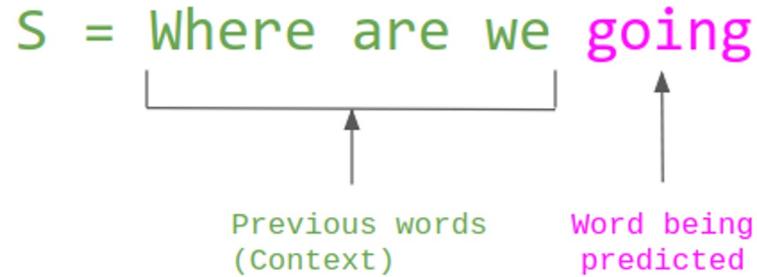
The Evolution of Large Language Models



Source: <https://github.com/Hannibal046/Awesome-LLM>

How does an LLM work?

- "Fancy text completion" – with some magic!



$$P(S) = P(\text{Where}) \times P(\text{are} \mid \text{Where}) \times P(\text{we} \mid \text{Where are}) \times P(\text{going} \mid \text{Where are we})$$

- Source: <https://thegradient.pub/understanding-evaluation-metrics-for-language-models/>

How does an LLM work?

How does it know which word to predict?

- Attention is all you need
- “Context window”
- **The context is all in the prompt! The LLM has no “shared memory” after it was trained!**

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.pолосухин@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly

How does an LLM work?

How does it know which word to predict?

Until

What is the capital of Germany? Berlin

The

What

A

For

...

How does an LLM work?

How does it know which word to predict?

What is the capital of Germany? The

How does an LLM work?

How does it know which word to predict?

What is the capital of Germany? The capital



No going back!

How does an LLM work?

How does it know which word to predict?

What is the capital of Germany? The capital of

How does an LLM work?

How does it know which word to predict?

What is the capital of Germany? The capital of Germany

How does an LLM work?

How does it know which word to predict?

What is the capital of Germany? The capital of Germany is

How does an LLM work?

How does it know which word to predict?

What is the capital of Germany? The capital of Germany is Berlin

How does an LLM work?

Text completion model:

What is the capital of France?

What is the capital of Germany?

What is the capital of Italy?

How does an LLM work?

Instructional model:

What is the capital of France?

What is the capital of Germany?

The capital of France is Paris.

The capital of Germany is Berlin.

TO

What is the capital of France?

What is the capital of Germany?



The capital of France is Paris.

The capital of Germany is Berlin.

How does an LLM work?

The context controls the output:

Germany was split into two parts after the second world war.

What is the capital of Germany? Until 1990 the capital used to be Bonn.

TO

Germany was split into two parts after the second world war.
What is the capital of Germany?

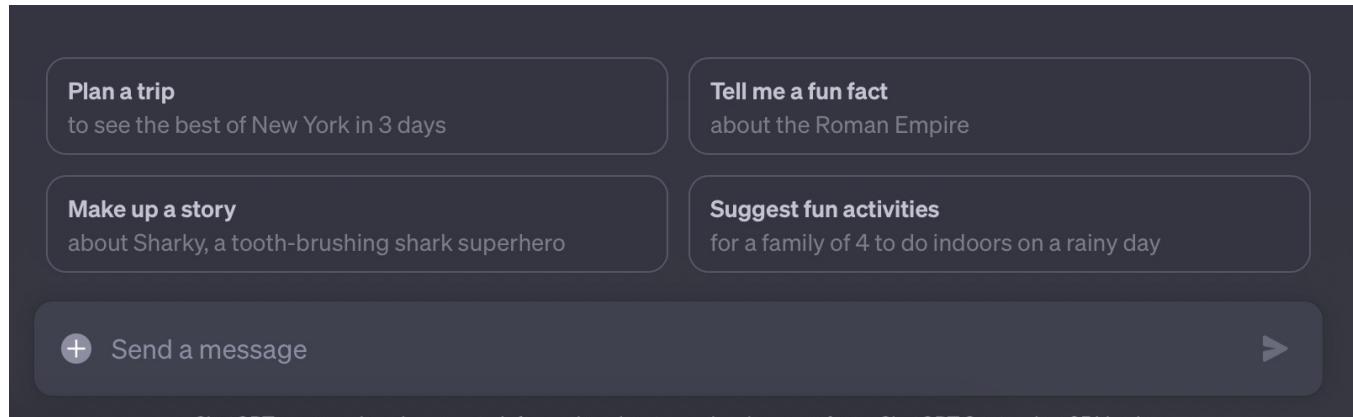


The capital of Germany is Berlin. After the end of World War II, Germany was divided into two parts: East Germany, officially known as the German Democratic Republic (GDR), and West Germany, officially known as the Federal Republic of Germany (FRG). East Germany's capital was East Berlin, while West Germany's capital was Bonn. After the fall of the Berlin Wall in 1989 and the reunification of Germany in 1990, Berlin became the capital of the reunited Germany.



What is a prompt?

- **A question or phrase:** Something you ask or tell the AI to start a conversation or get information.
- **Guides AI response:** The prompt helps the AI understand what you want and shape its reply accordingly.
- **Can be simple or complex:** Prompts can be straightforward questions or more detailed scenarios for the AI to engage with.
- **Like a conversation starter:** Think of it as an opening line that sparks a discussion between you and the AI.



Prompt Engineering: Common mistakes

- Even if ChatGPT is designed to be easy to use, the way you prompt it indeed makes a difference!
 - Most beginner prompts have the following problems:
 - too vague
 - lack context
 - ask for too much
 - not structured properly
 - To get great outputs, there are a few principles that should be executed well!
- **No need to become a “prompt engineer”**

TO

YOUR ROLE: Senior Data Analyst at McKinsey, formerly BCG

YOUR TASK: Help the user perform a Root Cause Analysis (RCA) using the 5 Why methodology.

YOUR APPROACH: Ask the user for a SMART problem statement and a specific aspect of it to analyze. Then work with the user and walk them through the 5 Why methodology.

Important: Help the user step by step! Ask the user to provide data and wait for their feedback before you drill down into the analysis!

CONTEXT: For each step of the 5 Why analysis, ask the user to provide hard data points such as engagement rates, clicks, impressions, budgets, time spent, etc. Compare these data points to industry benchmarks and, if there's a problem, dig deeper to identify the root cause of the problem. Stop after a maximum of 5 why's and conclude with a summary of the entire root cause analysis, leaving the user with actionable feedback to help solve their problem.

Start by asking the user for the SMART problem statement and particular area of interest.

Principle 1: Be as specific as possible

- If your prompt is not specific, it will be mediocre by definition: ChatGPT can't read your mind!

Bad example:

Make this SQL code better

Good example:

I'm using PostgreSQL and would like to optimize the following SQL query. My goal is to reduce the execution time and make the code easier for other developers to understand. Please use comments where appropriate. Before you rewrite the code, please explain the steps you would take to optimize it.

Principle 2: Don't let ChatGPT do the thinking

- Here's the deal: You do the thinking, ChatGPT does the doing!
- Give ChatGPT the mental framework it should use to solve the task

Bad example:

I want to develop a new SaaS application. Give me some ideas.

Good example:

My goal is to build a new SaaS application. To do this, I need to build something that my users will find valuable. My ideal users are social media managers from companies with 5M+ - 50M annual revenue who primarily work with LinkedIn, Facebook, and Instagram. Please use the Jobs-To-Be-Done framework to outline the key features and benefits that would appeal to this target market.

Principle 3: Break large tasks into small pieces

- If you ask for too much, you will get nothing useful in return

Bad example:

 Write a strategy for my retail business

 Title: Retail Business Growth Strategy

Executive Summary:

A well-defined strategy is crucial for any retail business to achieve success in today's competitive market. The following strategy focuses on boosting your retail business's sales, profitability, and overall growth by identifying target customers, improving product offerings, enhancing customer experience, leveraging digital channels, and optimizing store operations.

I. Identify Target Customer Segments

1. Define your ideal customer profile based on demographics, preferences, and behaviors.
2. Conduct market research to understand the needs and preferences of these customer segments.
3. Regularly review and update your customer profiles to stay in tune with their evolving preferences.

Good example:

 Write a strategy for my retail business

 Title: Retail Business Growth Strategy

Executive Summary:

A well-defined strategy is crucial for any retail business to achieve success in today's competitive market. The following strategy focuses on boosting your retail business's sales, profitability, and overall growth by identifying target customers, improving product offerings, enhancing customer experience, leveraging digital channels, and optimizing store operations.

I. Identify Target Customer Segments

1. Define your ideal customer profile based on demographics, preferences, and behaviors.
2. Conduct market research to understand the needs and preferences of these customer segments.
3. Regularly review and update your customer profiles to stay in tune with their evolving preferences.

 I want to write a strategy for my retail business. Which strategy frameworks are used by top strategy consultants to come up with a good management strategy for a business?

 Top strategy consultants use a variety of frameworks to develop management strategies for businesses. Here are some widely used frameworks that can help you craft a strategy for your retail business:

1. SWOT Analysis: Assess your business's Strengths, Weaknesses, Opportunities, and Threats to identify areas for improvement and capitalize on opportunities.
2. Porter's Five Forces: Analyze the competitive forces within your industry, including the threat of new entrants, the bargaining power of suppliers, the bargaining power of buyers, the threat of substitute products, and the intensity of competitive rivalry.
3. Value Chain Analysis: Evaluate the series of activities that create and build value in your business, identifying opportunities to improve efficiency, reduce costs, and create a competitive advantage.
4. Ansoff Matrix: Explore growth strategies through market penetration, market development, product development, and diversification, considering the balance between risk and potential rewards.
5. BCG Matrix: Categorize your products or business units into four categories based on

Principle 4: Use structured prompt formats

- Using structured prompt formats can significantly improve the quality of results obtained from ChatGPT and also make it easier to maintain your prompts.

TO

YOUR ROLE: Senior Data Analyst at McKinsey, formerly BCG

YOUR TASK: Help the user perform a Root Cause Analysis (RCA) using the 5 Why methodology.

YOUR APPROACH: Ask the user for a SMART problem statement and a specific aspect of it to analyze. Then work with the user and walk them through the 5 Why methodology.

Important: Help the user step by step! Ask the user to provide data and wait for their feedback before you drill down into the analysis!

CONTEXT: For each step of the 5 Why analysis, ask the user to provide hard data points such as engagement rates, clicks, impressions, budgets, time spent, etc. Compare these data points to industry benchmarks and, if there's a problem, dig deeper to identify the root cause of the problem. Stop after a maximum of 5 why's and conclude with a summary of the entire root cause analysis, leaving the user with actionable feedback to help solve their problem.

Principle 5: Build your prompt library

- Store your best prompts for easy retrieval

The image shows a digital workspace interface. On the left, there is a 'Prompts Library' section titled 'Aa Task' which is 'Locked'. It contains a list of tasks with icons and categories:

Aa Task	Category
Kill my startup	Strategy
Create a growth hacking plan	Marketing
Design a magical customer experience	Design
Generate habit-forming product using the Hooked model	Strategy Design
Create an Agile coach	Delivery
Design a workshop	Problem solving
Develop a customer onboarding program	Marketing Design
Train ChatGPT to write like you	Marketing
Create a Go To Market strategy template	Marketing
Run a pre mortem	Problem solving
Create a product thinking coach	Culture
Pick your wedge	Strategy
Estimate market size	Research
Identify a North Star Metric	Strategy
Design a Business Model Canvas 2.0	Strategy
Prepare for a Jobs To Be Done customer interview	Research
Suggest an Opportunity Solution Tree	Strategy

On the right, there is a 'Prompt' card with the following content:

You are here to help me switch from being primarily a project thinker to a product thinker. Here's the difference:

Project Thinking means valuing output over outcomes. When presented with a new idea or directive, project thinkers default to questions like: When does it need to be finished? Who will do it? How will we do it? Project thinking is centred around understanding expectations, formulating plans, marshalling resources, and coordinating actions.

Product Thinking: Rather than centring around the logistics of the project, product thinking is about understanding goals, motivations, devising solutions, simulating the effects of those solutions, and then choosing the best course of action. Product thinking involves questions like: Why is it important? What are our goals? What else could happen? How will we differentiate? Product thinking means looking ahead to the effects you want to have on the user or audience.

When I come to you with a problem, one Product Thinking framework you can use is WAYRTTD = What are you really trying to do? Meaning: What is it that you're trying to accomplish? Instead of just troubleshooting, let's talk about the goal. Taking this step back can help to recenter the conversation on outcomes can lead to a better, more creative solution.

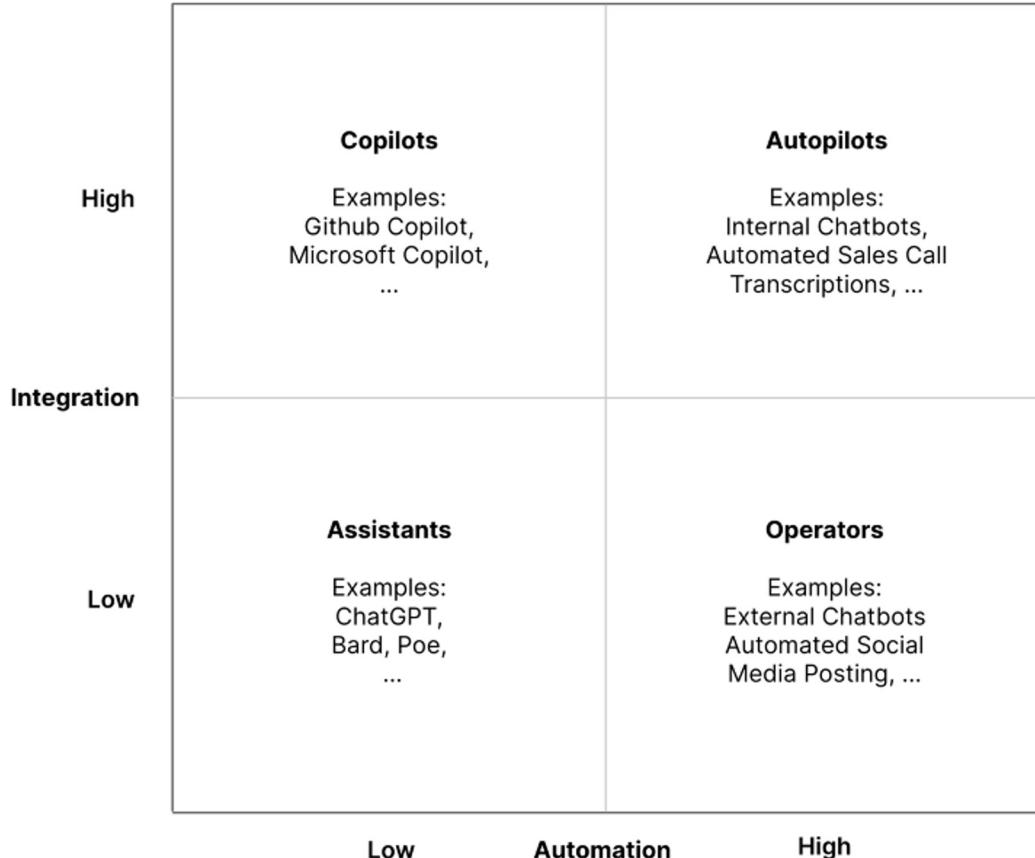
Do you understand these instructions and your role?

When to use LLMs

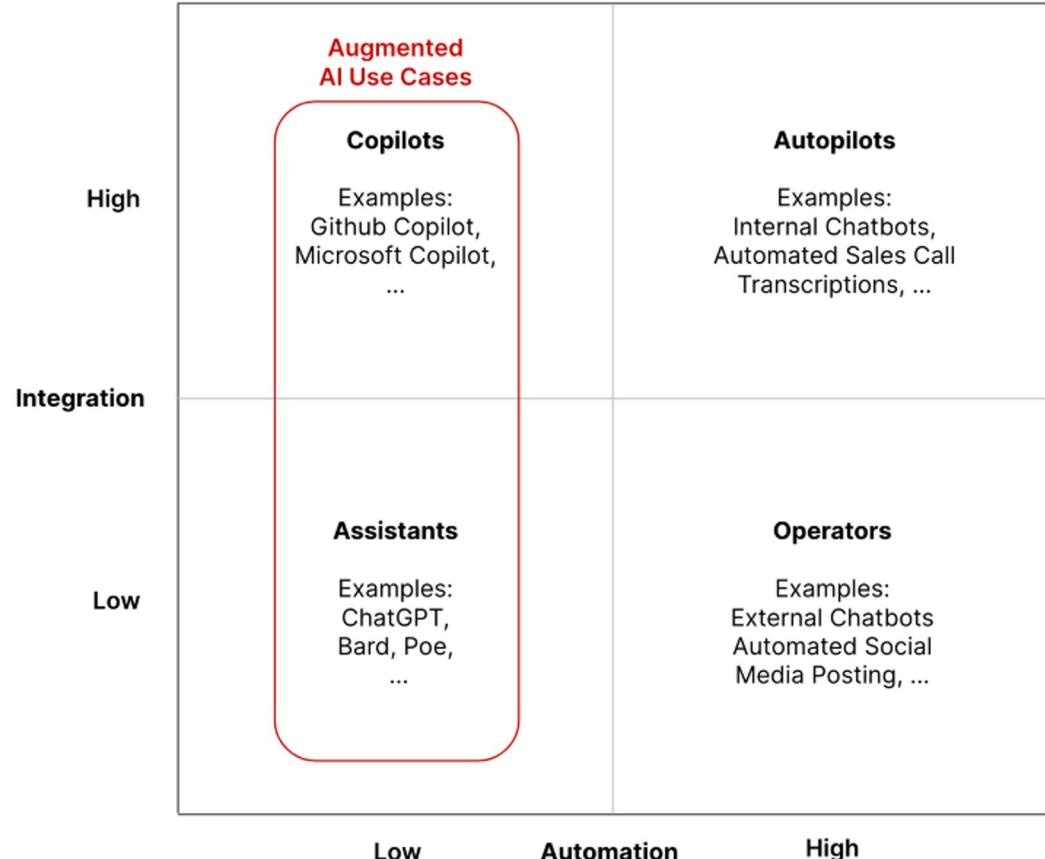
- ✓ Creative use cases
- ✓ No single true answer
- ✓ Discussions

- ✗ Factual use cases
- ✗ There is a single right answer
- ✗ Math (only with plugins)

Use Case Matrix



Use Case Matrix





Mark Tenenholz 
@marktenenholz

Before we start...

- Don't be too dogmatic about LLMs
- Approach them from a practical angle
- See what works for you

→ **Be Group 1**

The core of so many arguments about LLMs:

Group 1 is builders who have pragmatic expectations of LLMs, and use them accordingly. They're quite happy with them.

Group 2 is shitfluencers who are jumping on the hype wave and ruining it for everyone.

Group 3 hates group 2 and loves showing the deficiencies of LLMs to get back at them. Some are realists, others just have very high expectations of AI.

Group 1 doesn't understand group 3, because they just ignore group 2 and make use of it. They feel like group 3 is being unfair to LLMs and cherry-picking.

Now, group 3 starts arguing with group 1. "LLMs are not reliable! They can't always perform complex reasoning!" they shout.

"We agree!" group 1 responds. "But they're still useful!"

And that argument continues, between two groups that agree more than they realize.

Meanwhile, group 2 frolics about, ignoring this entire debate, continuing to post "10 ways to make \$10k/mo with ChatGPT"

And the cycle continues.

Use Case 1: Automated Data Analysis with ChatGPT (Plus)

Automated Data Analysis

Upload data to ChatGPT Advanced Data Analysis

What is Advanced Data Analysis?

- Available to Plus users
- Python sandbox to write and run code
- Subsequent calls can build on top of each other
- Supports file uploading and downloading

The Good

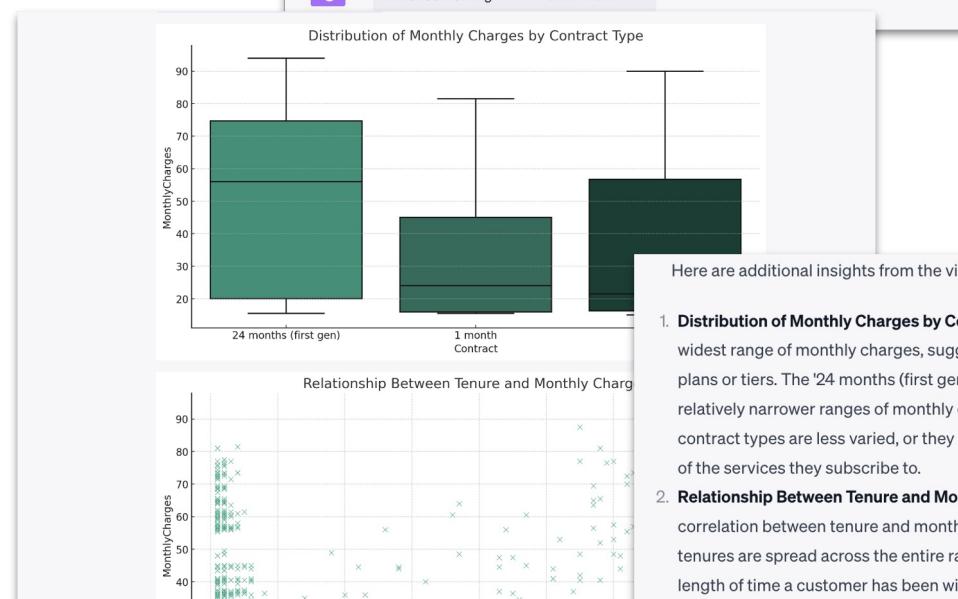
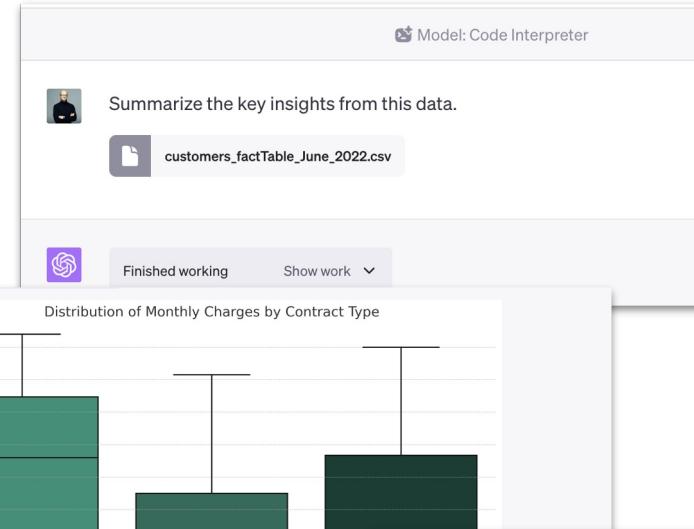
- A data analyst in your pocket

The Bad

- Potentially wrong (need to double check!)

The Ugly

- Probably not allowed



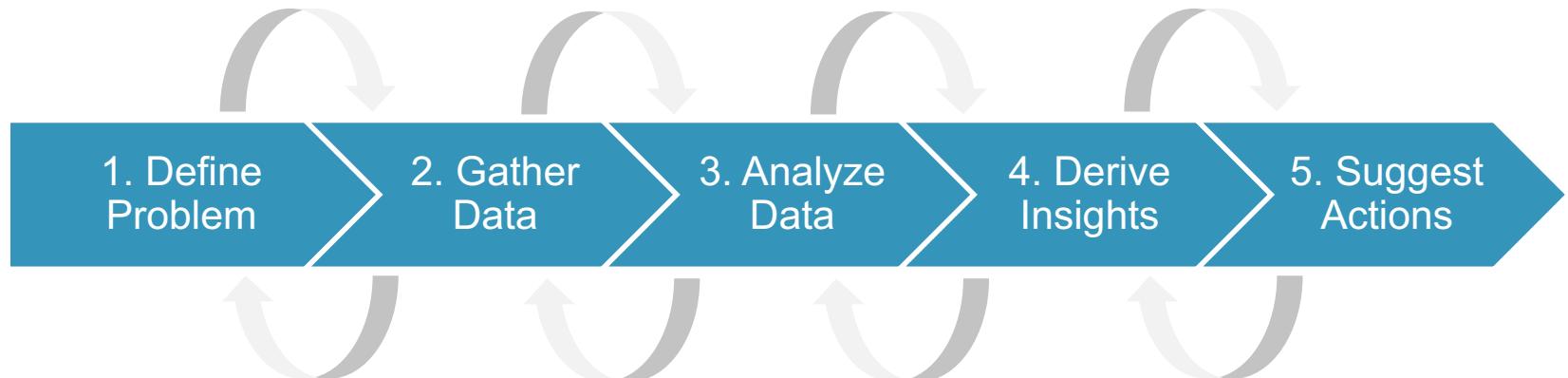


Demo

Use Cases 2: AI-Powered Problem Solving

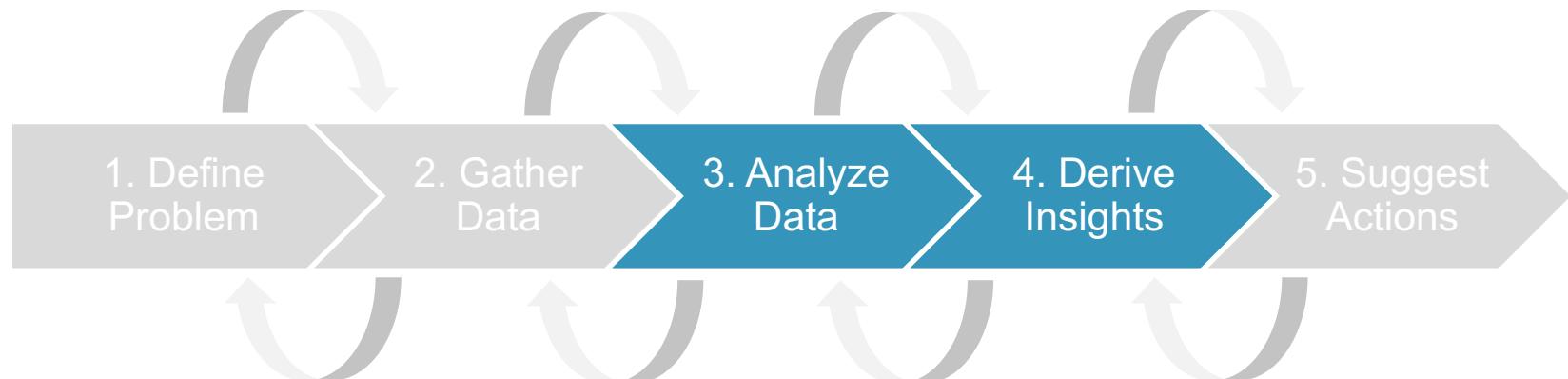
Recap: The Data Analytics Process (simplified)

- Don't start with the data, start with the problem



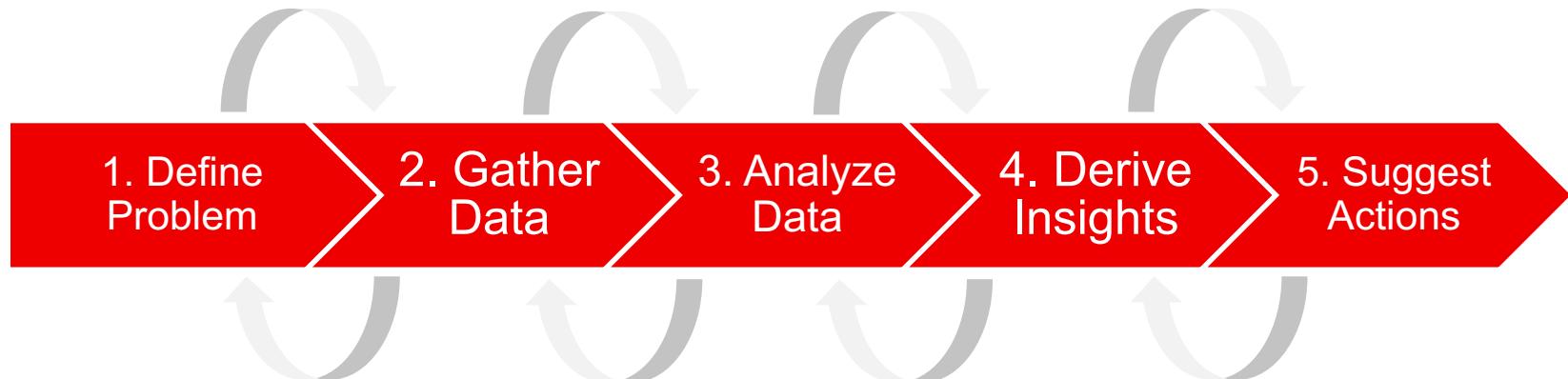
ChatGPT For Data Analysis

→ What people think where ChatGPT can help them:



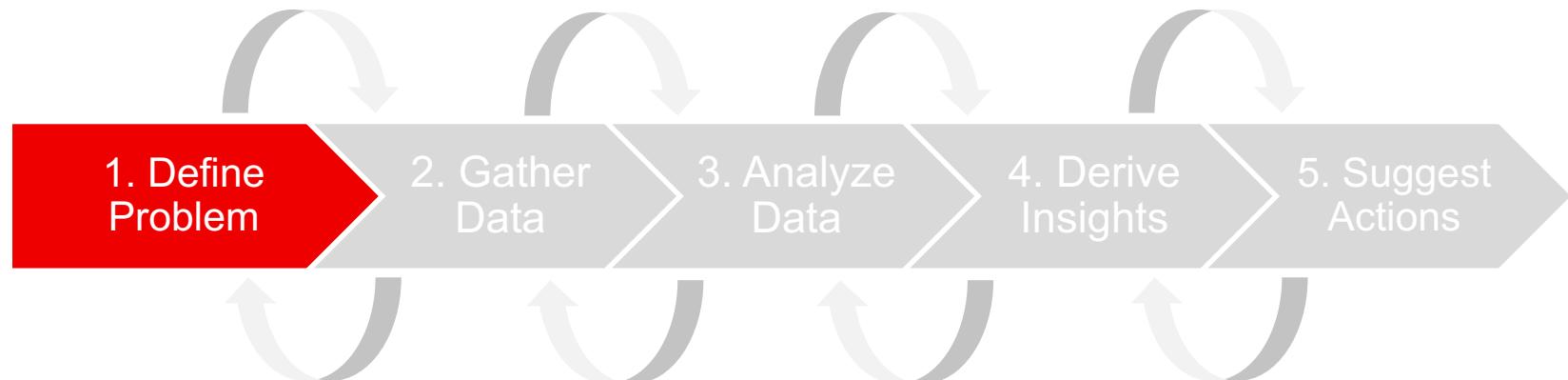
ChatGPT For Data Analysis

→ Where ChatGPT can actually help them:



ChatGPT For Data Analysis

→ Where we will start:

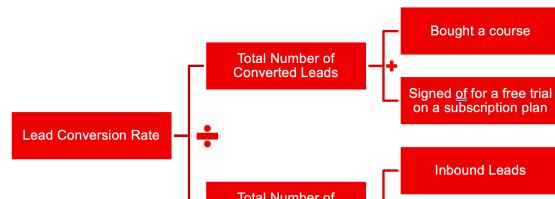


3 Components for Problem Solving

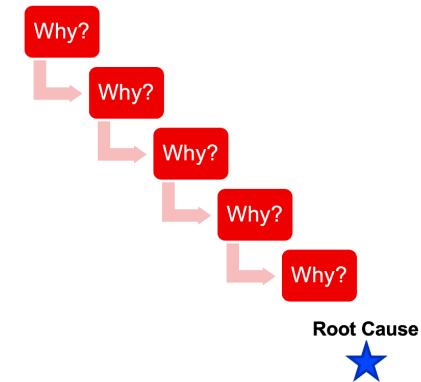
→ Problem Statement,
ideally SMART

What opportunities exist for Edu X to increase its lead conversion rate to 45% over the next 3 months through an improved marketing strategy in alignment with the business objective of selling more than 1,000 courses per quarter.

→ Issue Tree,
ideally MECE



→ Analysis design,
ideally scientific



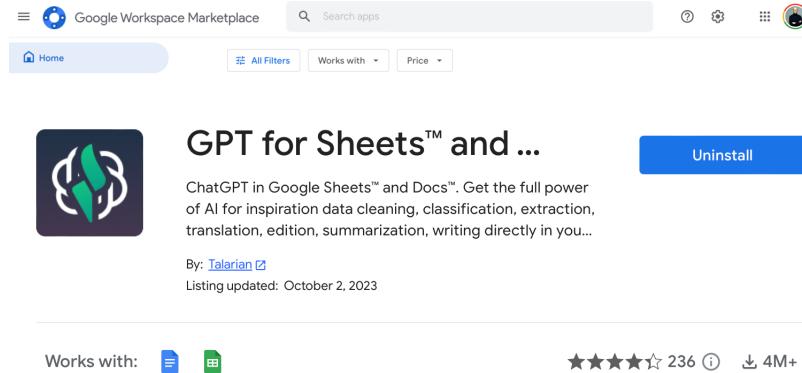


Demo

Use Case 3: AI in Spreadsheets

How to use ChatGPT in Google Sheets (and Excel)

- Google Sheets offers a free plugin, **GPT for Sheets and Docs**, to integrate GPT functionality via an OpenAI API key.
- Use the plugin for **non-critical data**; it's not ideal for sensitive information.
- For secure handling of sensitive data, **deploy GPT-4 on Azure** and use custom functions in Google Sheets or Python/R in Excel.
- Or: Wait for **Microsoft Copilot**.
- Testing with the plugin is a **good first step** before moving on to more advanced, secure options.





Demo

Q&A

Thank you!

**Subscribe to my free
newsletter →**



Let's connect!

linkedin.com/in/tobias-zwingmann

tobias@rapyd.ai

ai4bi.rocks