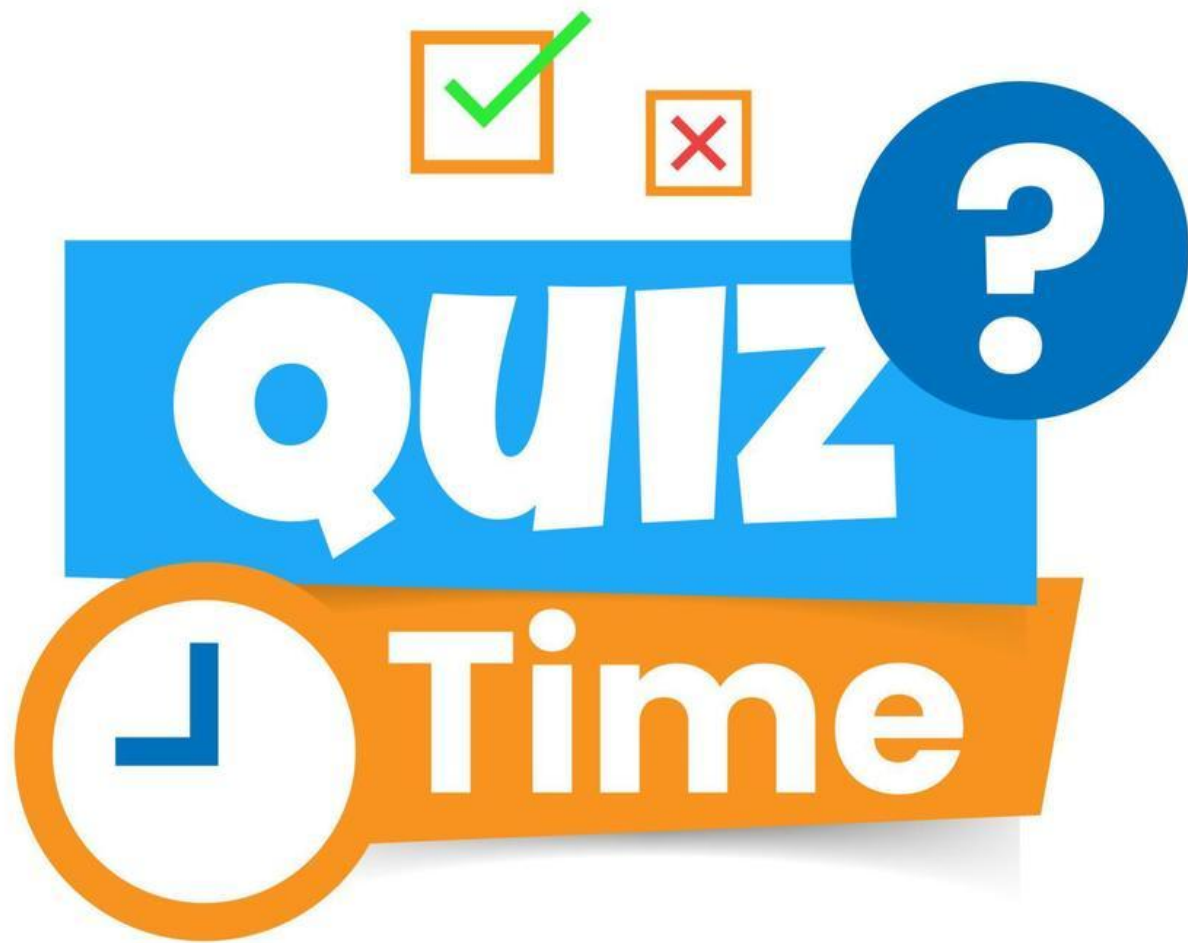


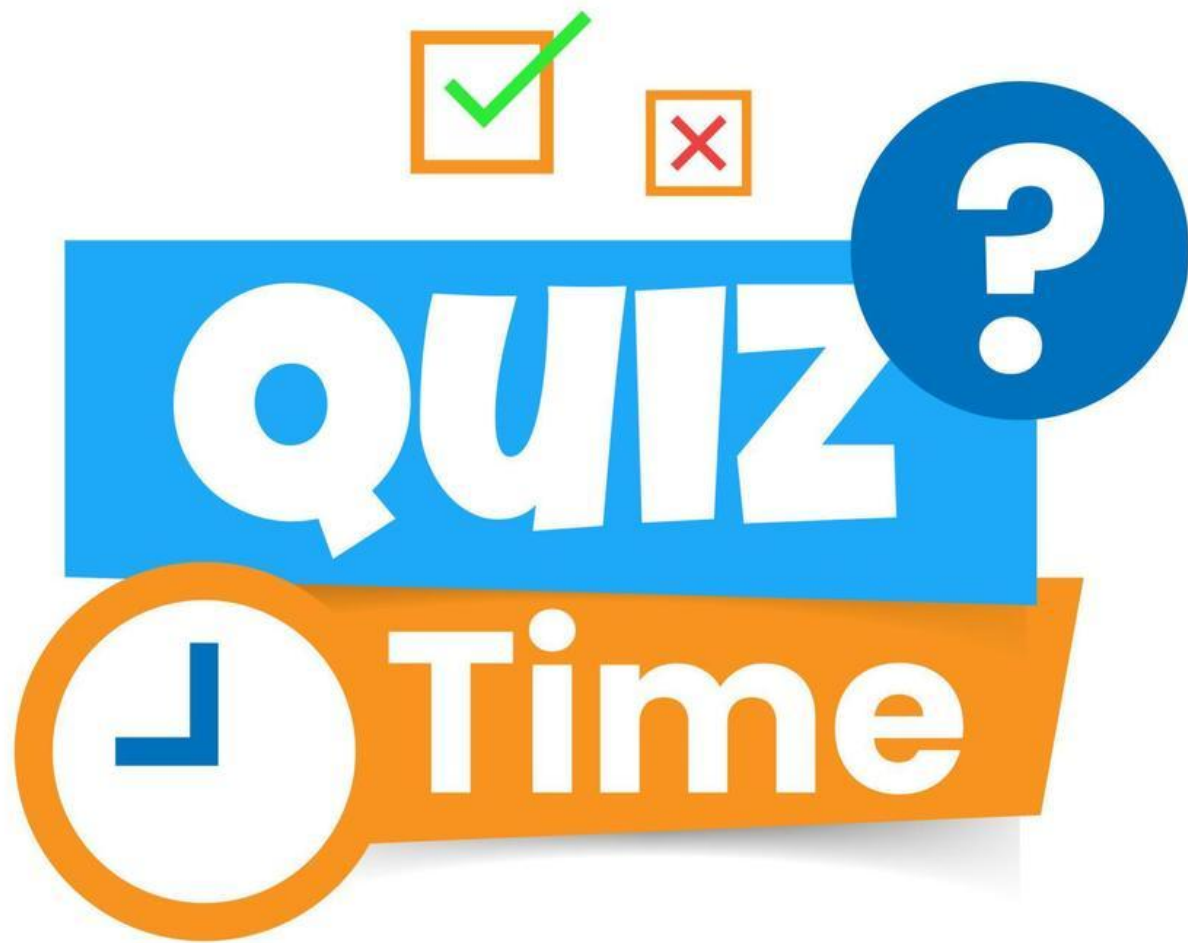


Part II - Prompt Engineering




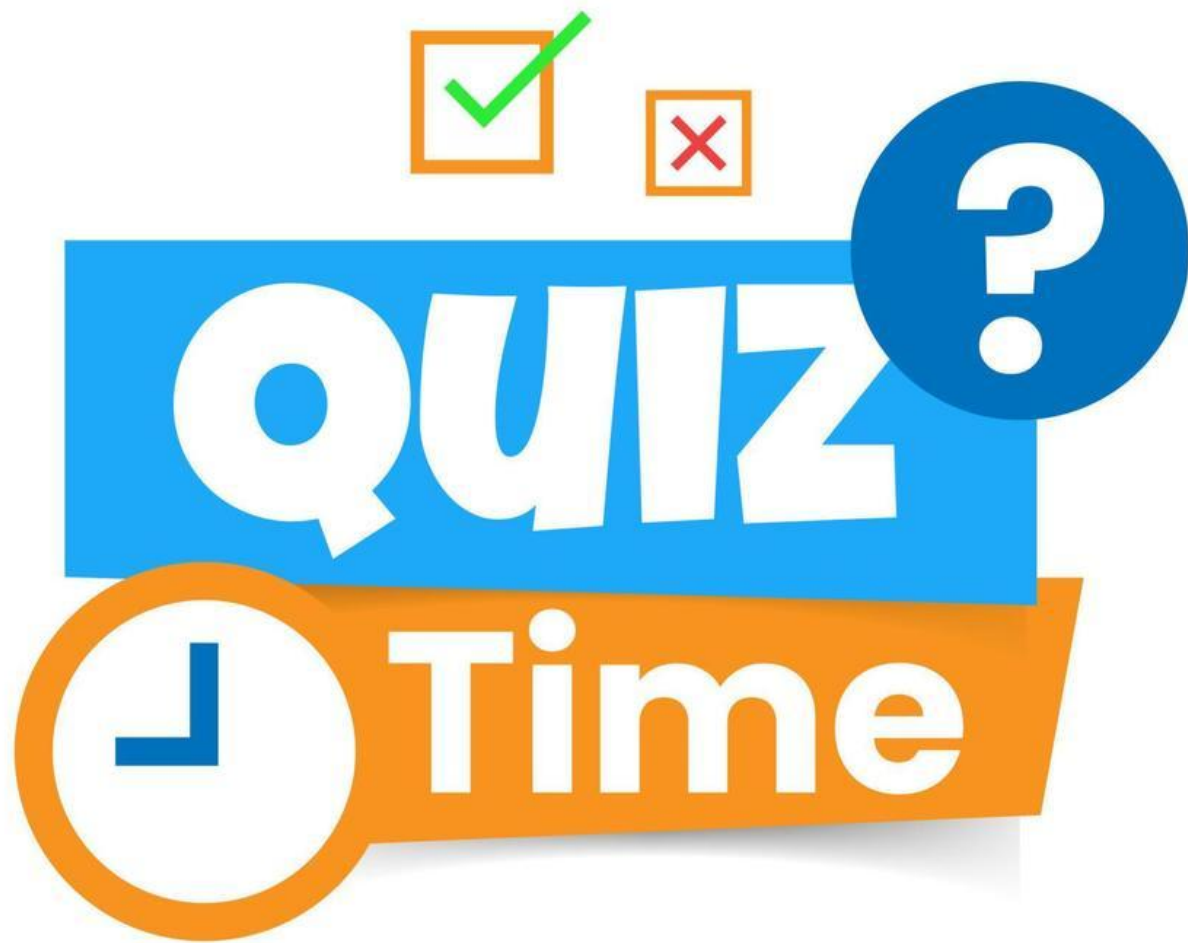
InnovateTech Solutions is developing an AI-driven platform to revolutionize content creation for digital marketing campaigns. The company aims to generate creative and personalized content to enhance user engagement. As the Chief Data Scientist, you are leading the team in selecting the most suitable ML model for this task. Which of the following key ML models is pivotal for InnovateTech Solutions for generating new and personalized content for digital marketing campaigns?

A	Descriptive Model
B	Generative Model



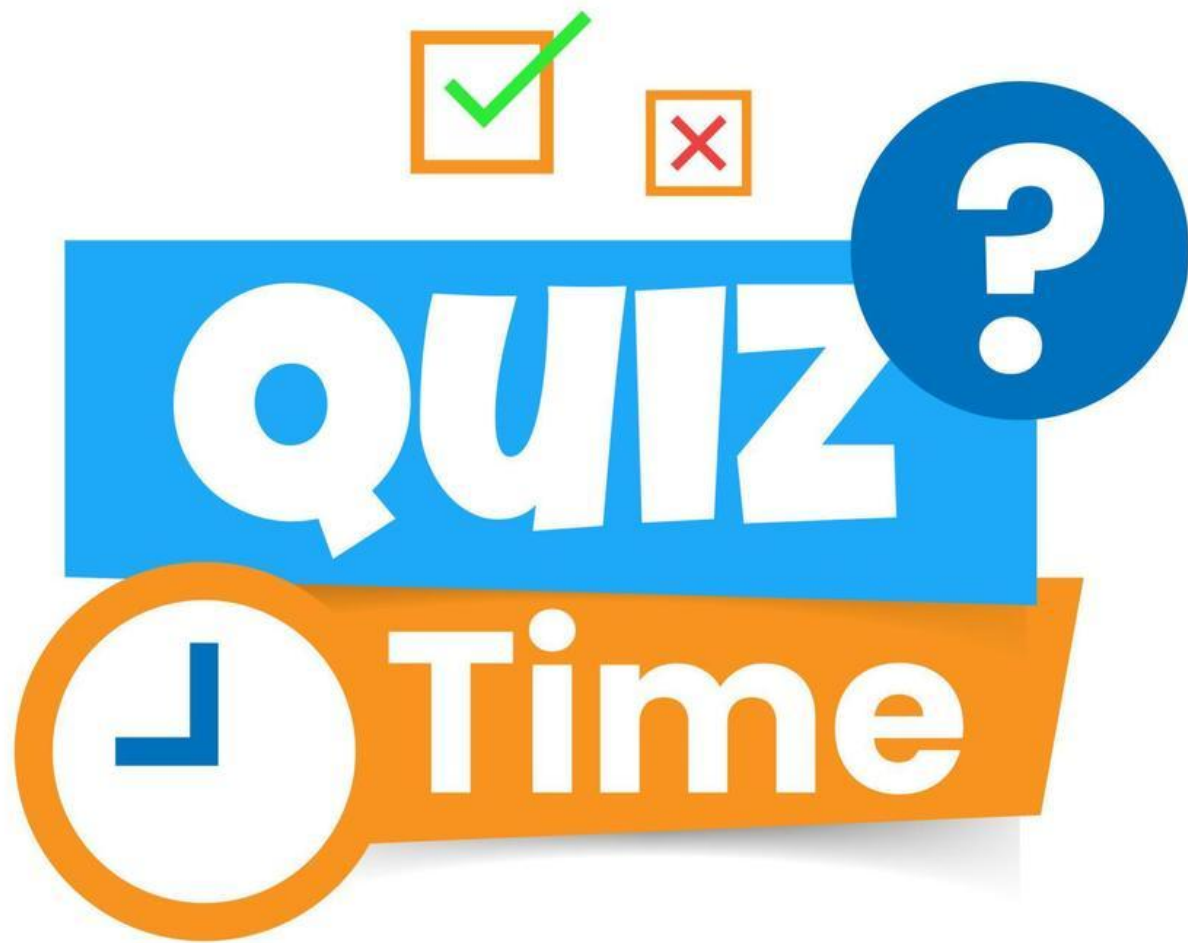
InnovateTech Solutions is developing an AI-driven platform to revolutionize content creation for digital marketing campaigns. The company aims to generate creative and personalized content to enhance user engagement. As the Chief Data Scientist, you are leading the team in selecting the most suitable ML model for this task. Which of the following key ML models is pivotal for InnovateTech Solutions for generating new and personalized content for digital marketing campaigns?

A	Descriptive Model
 B	Generative Model




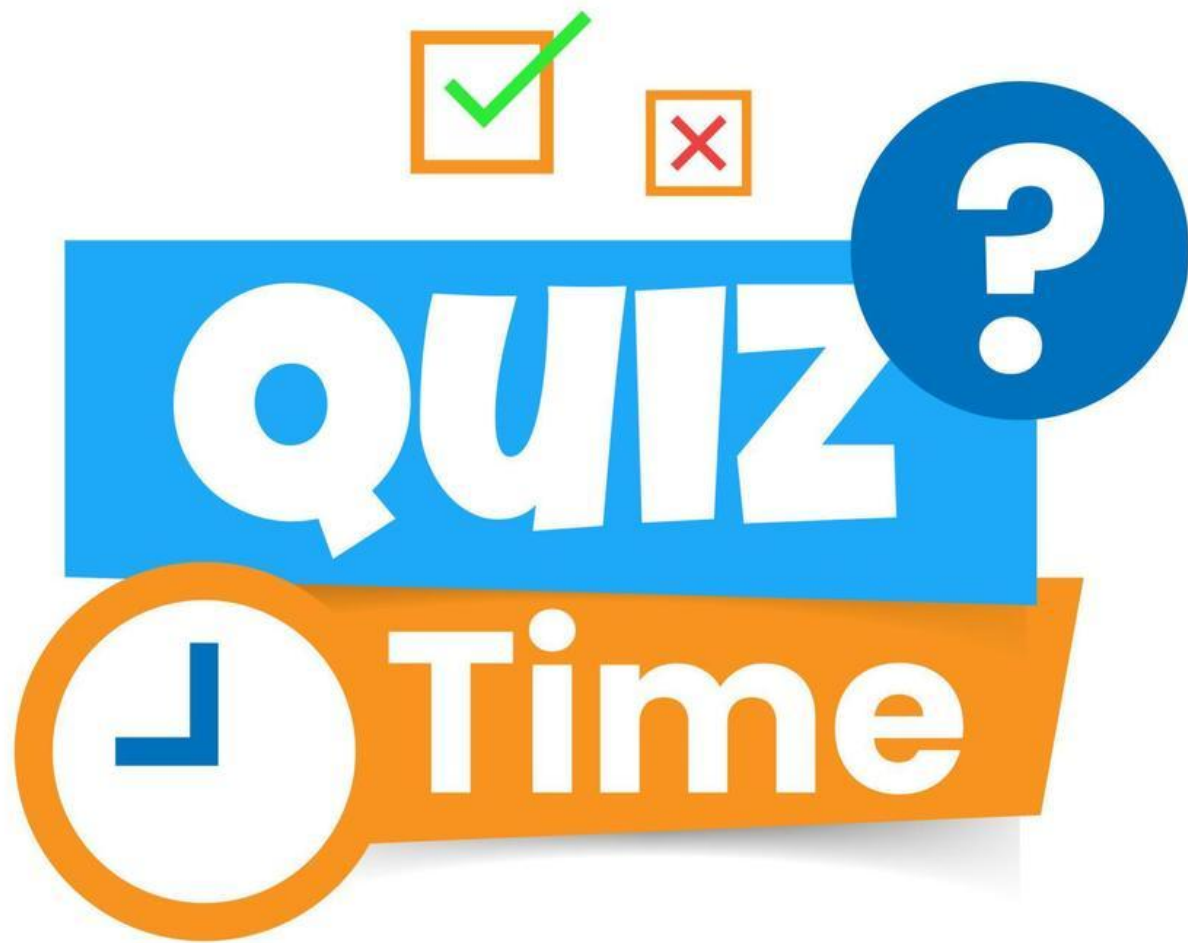
As the Chief Technology Officer (CTO) of TechSolutions Ltd., a leading IT organization specializing in digital transformation, you are tasked with enhancing the user experience and operational efficiency of your company's internal knowledge base. Considering the utilization of Azure Cognitive Services, which of the following use cases would be most impactful for the organization using Azure's AI capabilities?

A	Implementing Azure Computer Vision to automatically tag and categorize internal scanned images of the document based on their content, facilitating streamlined document management and retrieval
B	Deploying Azure Text Analytics on technical manuals to automatically extract key insights and terminologies, aiding in the creation of a comprehensive knowledge base for IT support
C	Utilizing Azure Personalizer to customize the user interface of the organization's internal applications based on individual employee preferences and usage patterns
D	Implementing Azure QnA Maker to develop an intelligent IT support chatbot, capable of answering employees' queries and providing solutions to common technical issues



As the Chief Technology Officer (CTO) of TechSolutions Ltd., a leading IT organization specializing in digital transformation, you are tasked with enhancing the user experience and operational efficiency of your company's internal knowledge base. Considering the utilization of Azure Cognitive Services, which of the following use cases would be most impactful for the organization using Azure's AI capabilities?

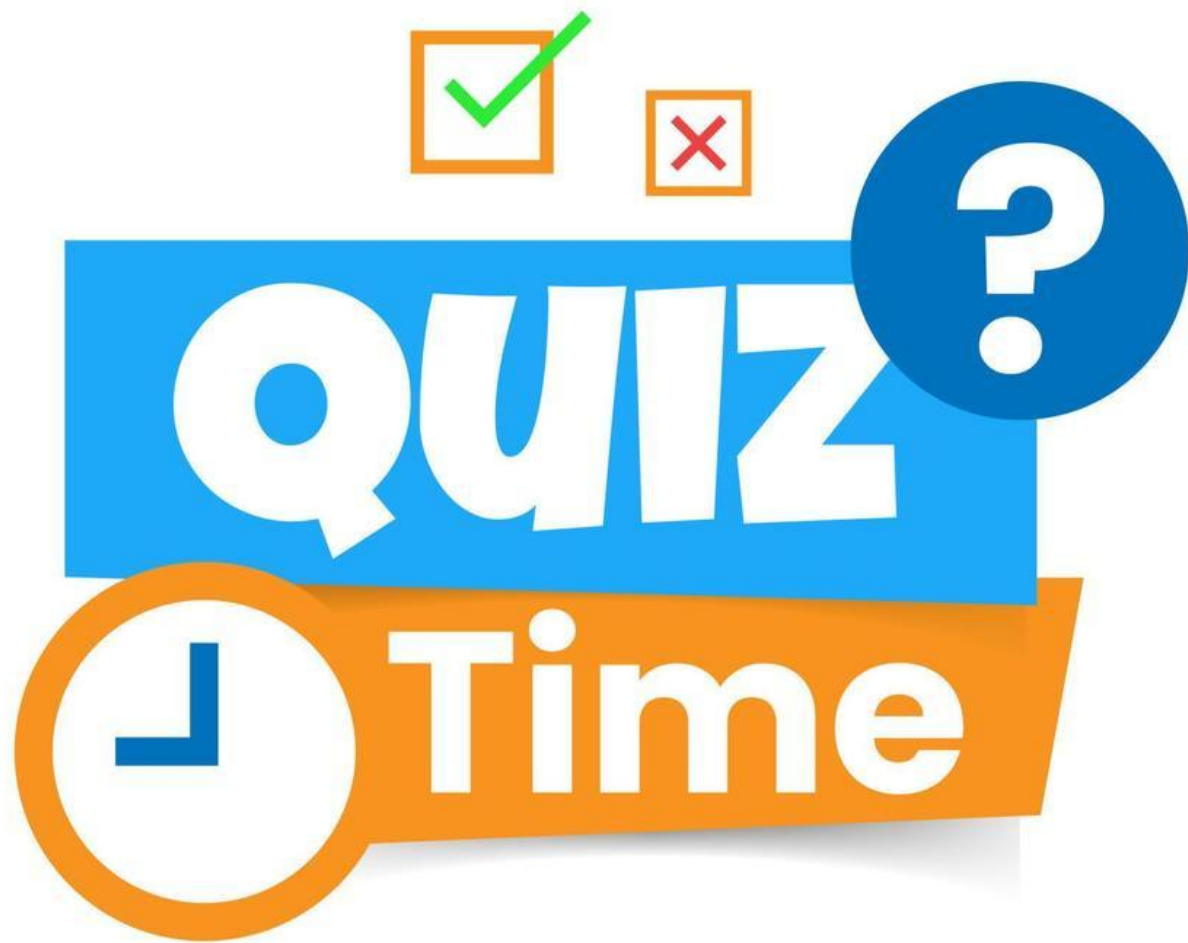
A 	Implementing Azure Computer Vision to automatically tag and categorize internal scanned images of the document based on their content, facilitating streamlined document management and retrieval
B	Deploying Azure Text Analytics on technical manuals to automatically extract key insights and terminologies, aiding in the creation of a comprehensive knowledge base for IT support
C	Utilizing Azure Personalizer to customize the user interface of the organization's internal applications based on individual employee preferences and usage patterns
D	Implementing Azure QnA Maker to develop an intelligent IT support chatbot, capable of answering employees' queries and providing solutions to common technical issues



DataSolutions Tech (a leading IT organization focused on data-driven solutions) aims to boost its software development process by leveraging the best features of Generative AI. As the CIO, you want to explore how Generative AI can contribute to the innovation and efficiency of the organization's coding practices.


Which of the following best describes a key benefit or feature of Generative AI for an IT organization in the context of software development?

A	Utilizing Generative AI to automate routine IT tasks and administrative processes within the organization
B	Integrating Generative AI to analyze and optimize network performance, ensuring seamless connectivity for all employees
C	Applying Generative AI for cybersecurity measures, automatically identifying and mitigating potential threats to the IT infrastructure
D	Incorporating Generative AI to dynamically generate code snippets and assist developers in rapid prototyping and software development



DataSolutions Tech (a leading IT organization focused on data-driven solutions) aims to boost its software development process by leveraging the best features of Generative AI. As the CIO, you want to explore how Generative AI can contribute to the innovation and efficiency of the organization's coding practices.

Which of the following best describes a key benefit or feature of Generative AI for an IT organization in the context of software development?

A	Utilizing Generative AI to automate routine IT tasks and administrative processes within the organization
B	Integrating Generative AI to analyze and optimize network performance, ensuring seamless connectivity for all employees
C	Applying Generative AI for cybersecurity measures, automatically identifying and mitigating potential threats to the IT infrastructure
 D	Incorporating Generative AI to dynamically generate code snippets and assist developers in rapid prototyping and software development

Agenda

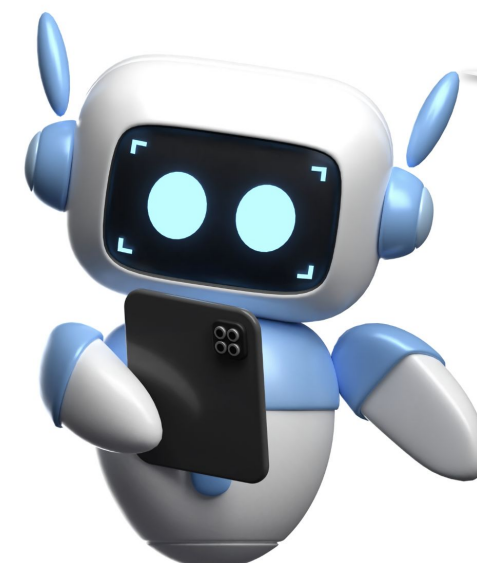
- ● **Recap of LLMs**
- Understanding Prompt Design
- Context, Examples, & Tuning Prompt Parameters
- Prompt Design Best Practices

What is a Language Model?



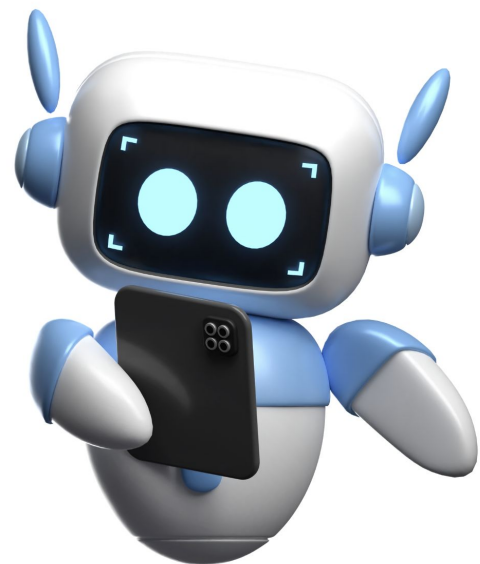
A language model is like a *smart machine that helps you finish sentences*. It's trained to predict and complete sentences, making it a super useful tool when you're typing or texting

Listen to your _____

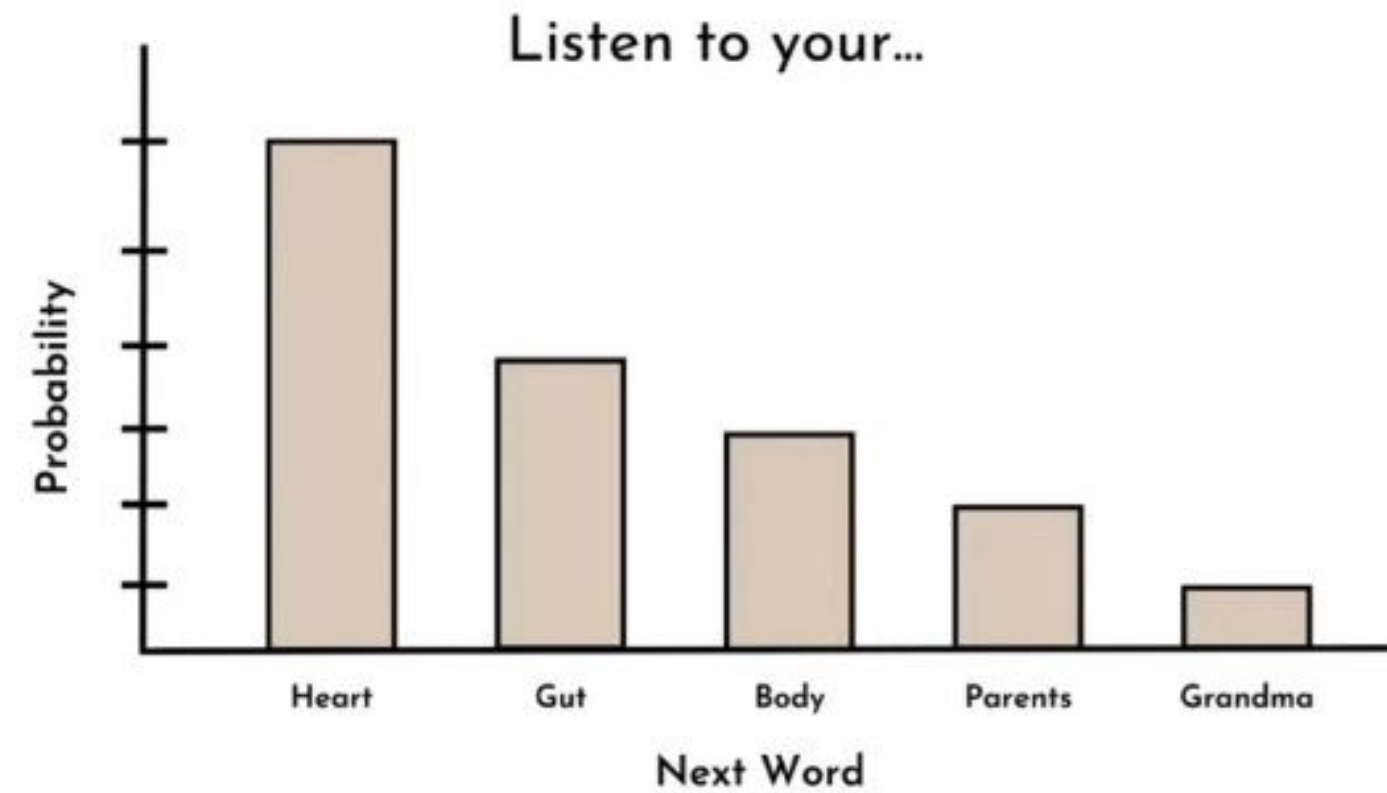


Heart **10.5%**
Gut 9.8%
Body 5.6%
Parents 3.5%
...

What is a Language Model?



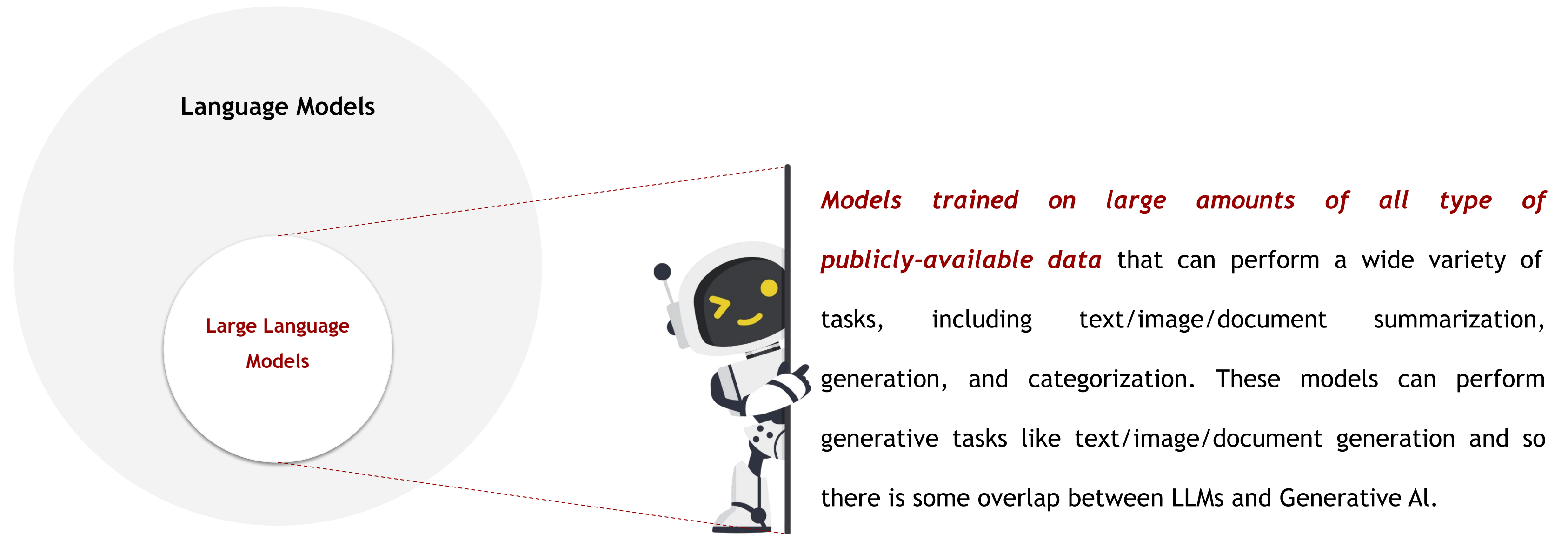
Next-word prediction paradigm



[1] Survey of Large Language Models. [arXiv:2303.18223](https://arxiv.org/abs/2303.18223) [cs.CL]

[3] Radford, A., & Narasimhan, K. (2016). Improving Language Understanding by Generative Pre-Training. (GPT-1 Paper)

What are Large Language Models?



What is “Large” in Large Language Models?

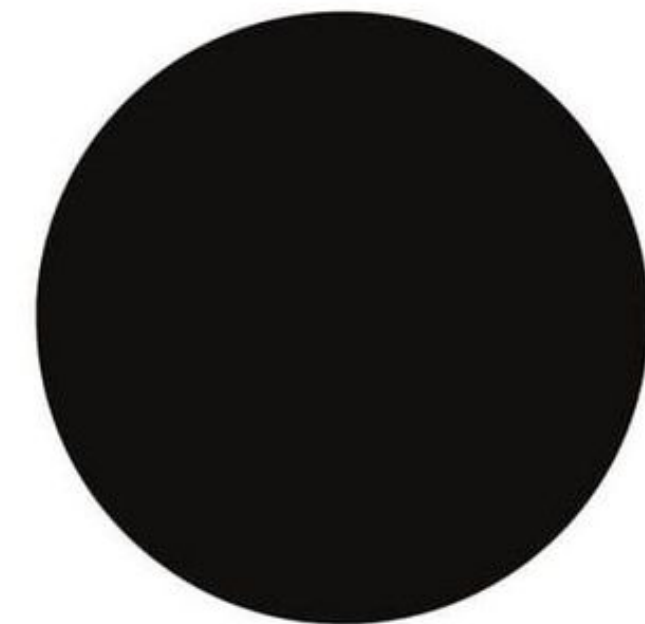
- The term “large” refers to the size of the model in terms of its parameters and the volume of the training data
- Parameters are a part of the model that learns from historical training data
- They help in predicting the next word in a sequence
- One Trillion Parameters have been used in GPT 4

GPT-3



175.000.000.000

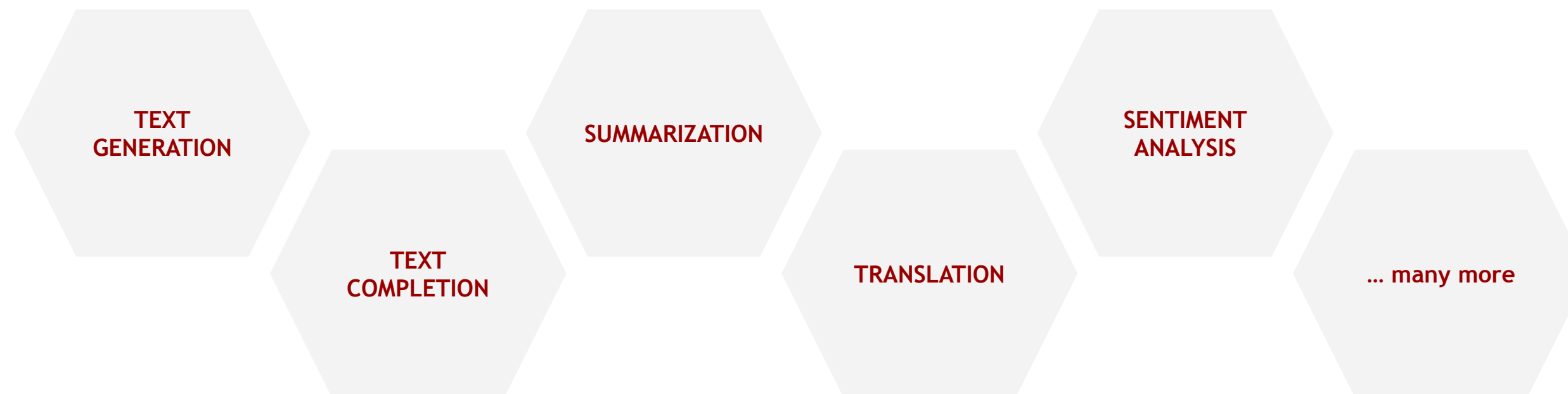
GPT-4



1.000.000.000.000.00

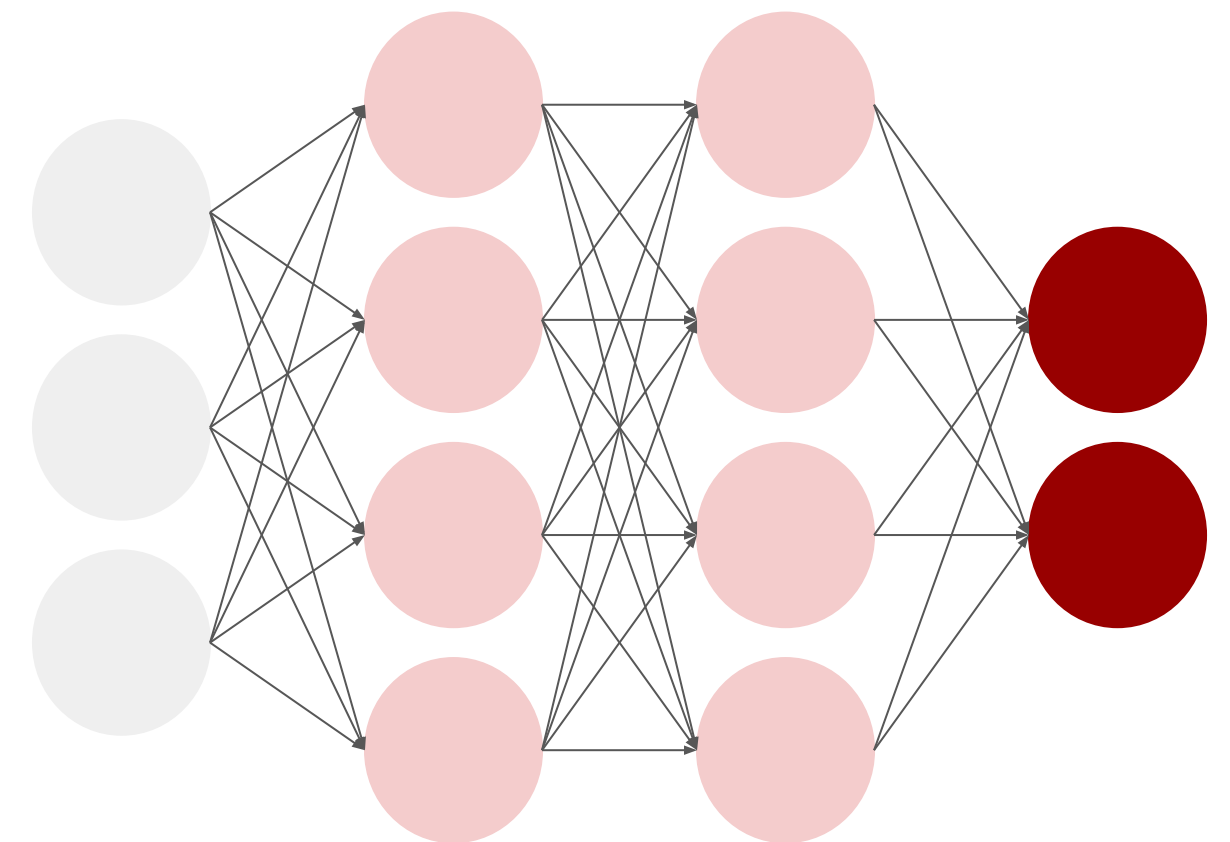
What is “Language” in Large Language Models?

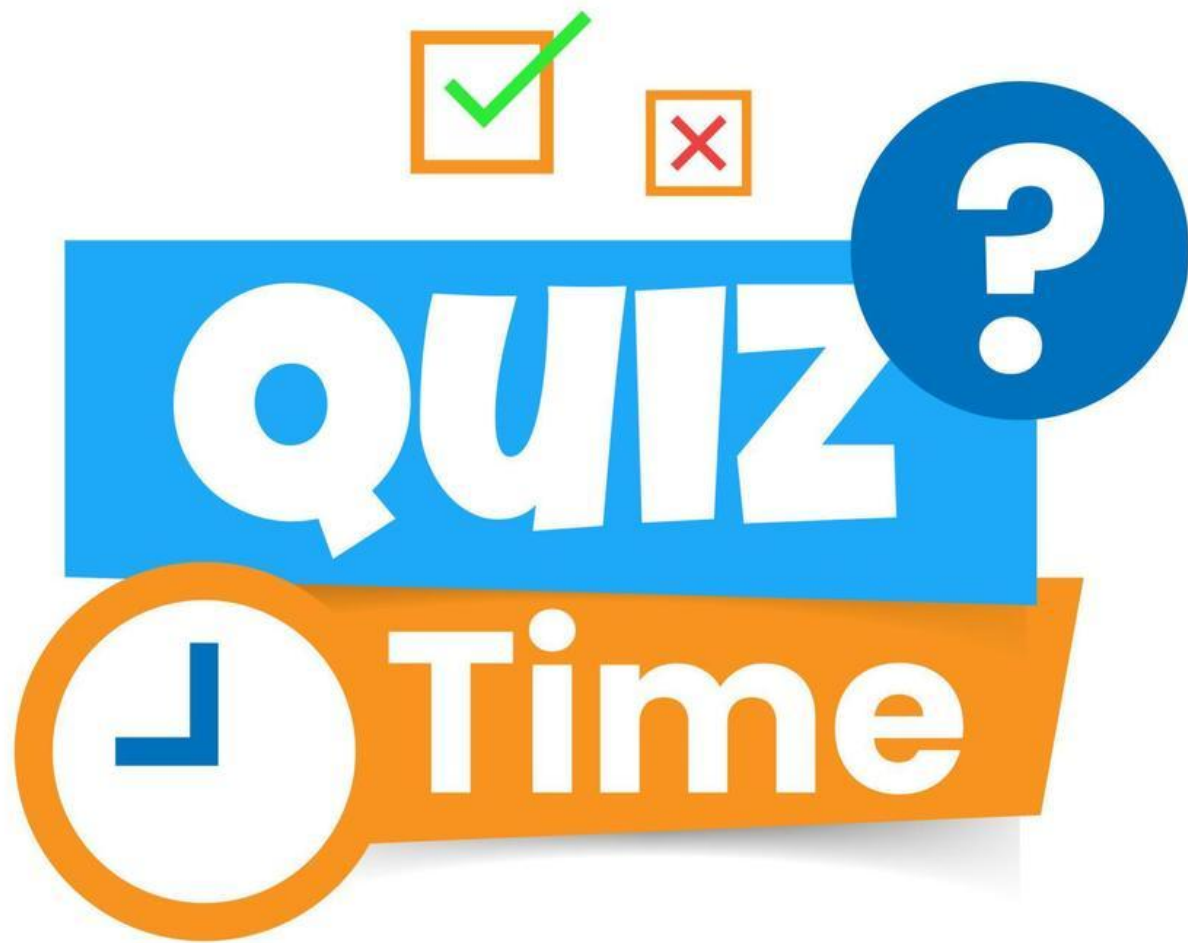
- Human-understandable languages like English
- The models learn statistical patterns of the data
 - Example: Probability of "the" preceding the noun "cat" are very high
- These models can generate new text, mimicking Humans
- Applications:-



What is a “Model” in Large Language Models?

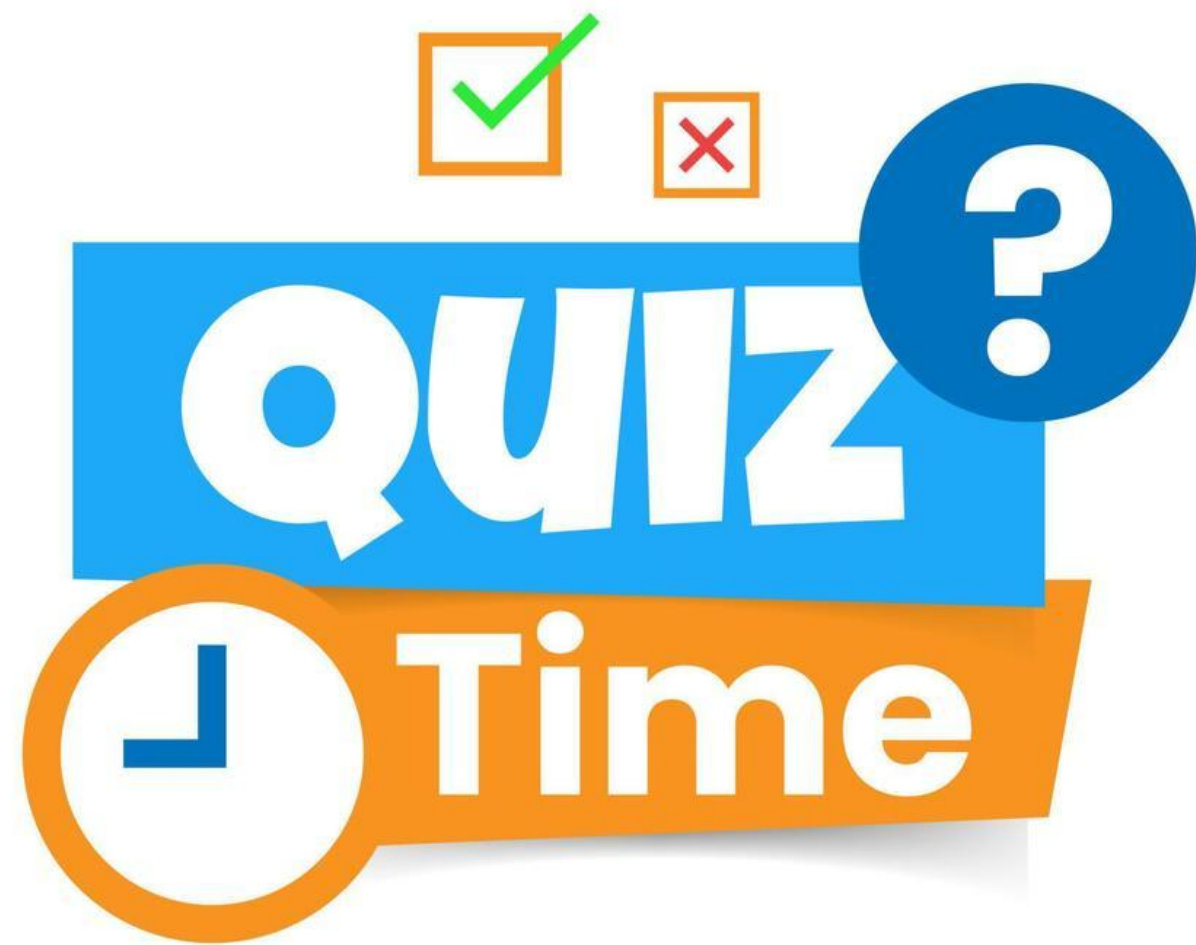
- A **model** refers to a mathematical representation of a real-world process
- It is essentially a program or an algorithm
 - Trained on a dataset to recognize patterns
 - Uses these patterns to make predictions or decisions without being explicitly programmed to perform the task
- ChatGPT uses GPT3 or GPT4 models





In the field of natural language processing, how are Large Language Models (LLMs) commonly applied?

A	Generating human-like text in various languages
B	Analyzing complex mathematical equations
C	Designing computer hardware components
D	Controlling robotic movements in manufacturing

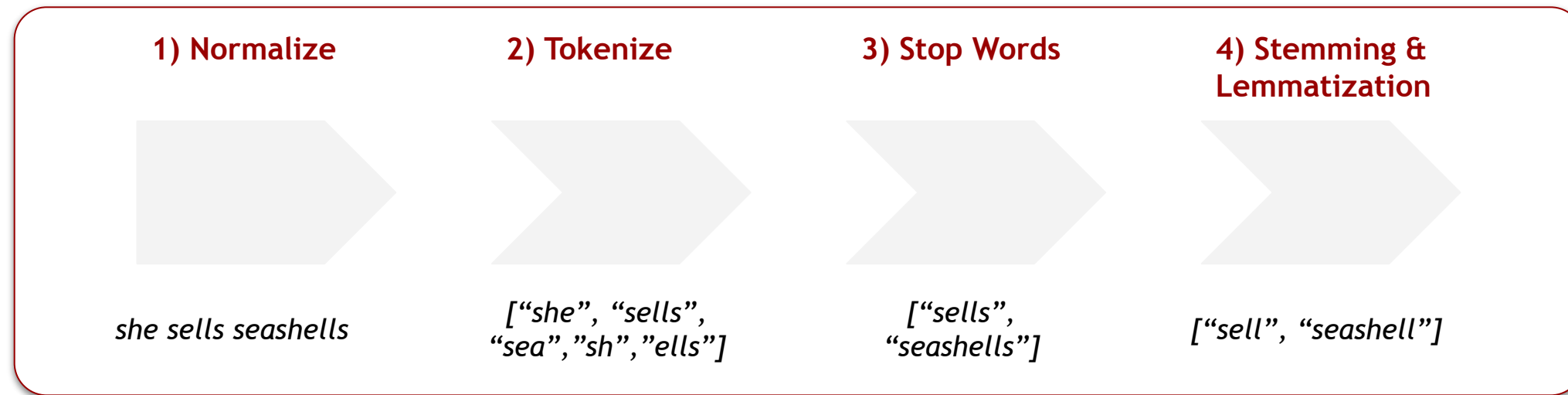


In the field of natural language processing, how are Large Language Models (LLMs) commonly applied?

A ✓	Generating human-like text in various languages
B	Analyzing complex mathematical equations
C	Designing computer hardware components
D	Controlling robotic movements in manufacturing

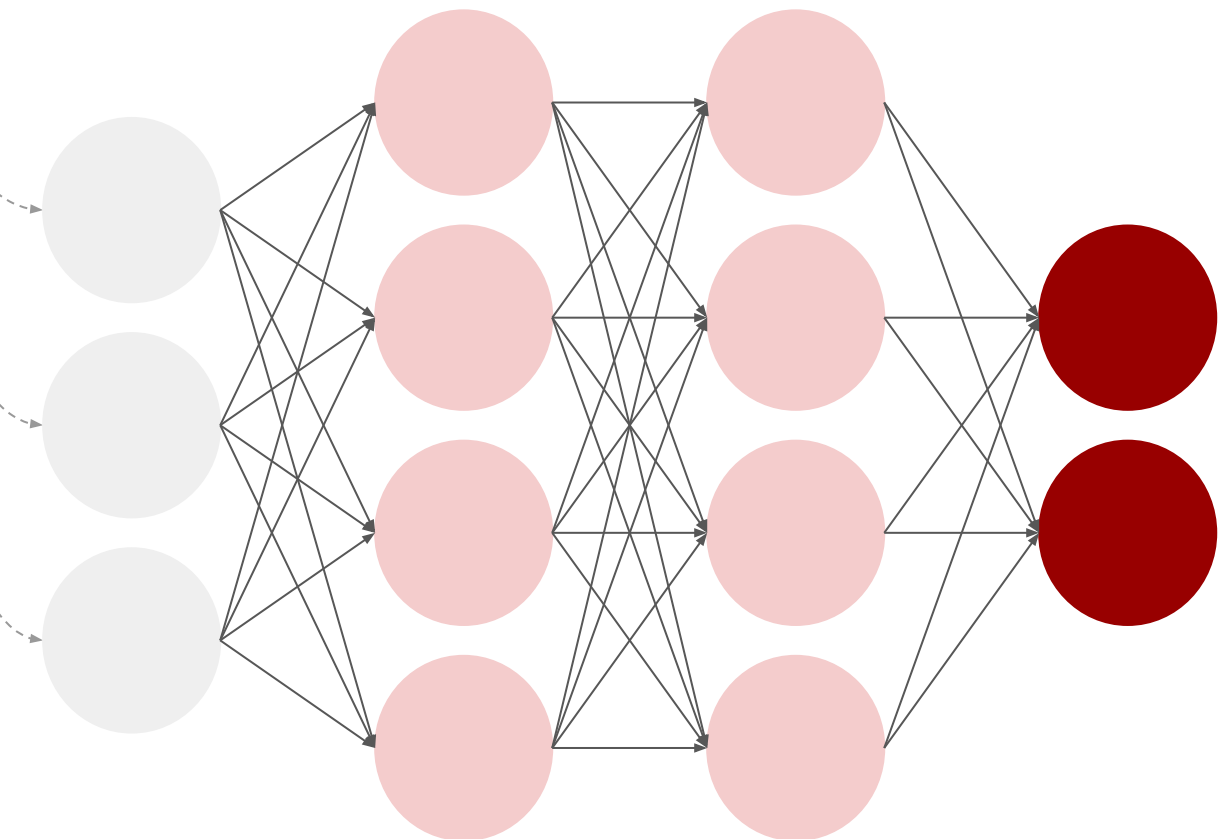
How a Large Language Model Works?

PRE-PROCESSING (Example Text: She Sells Seashells)



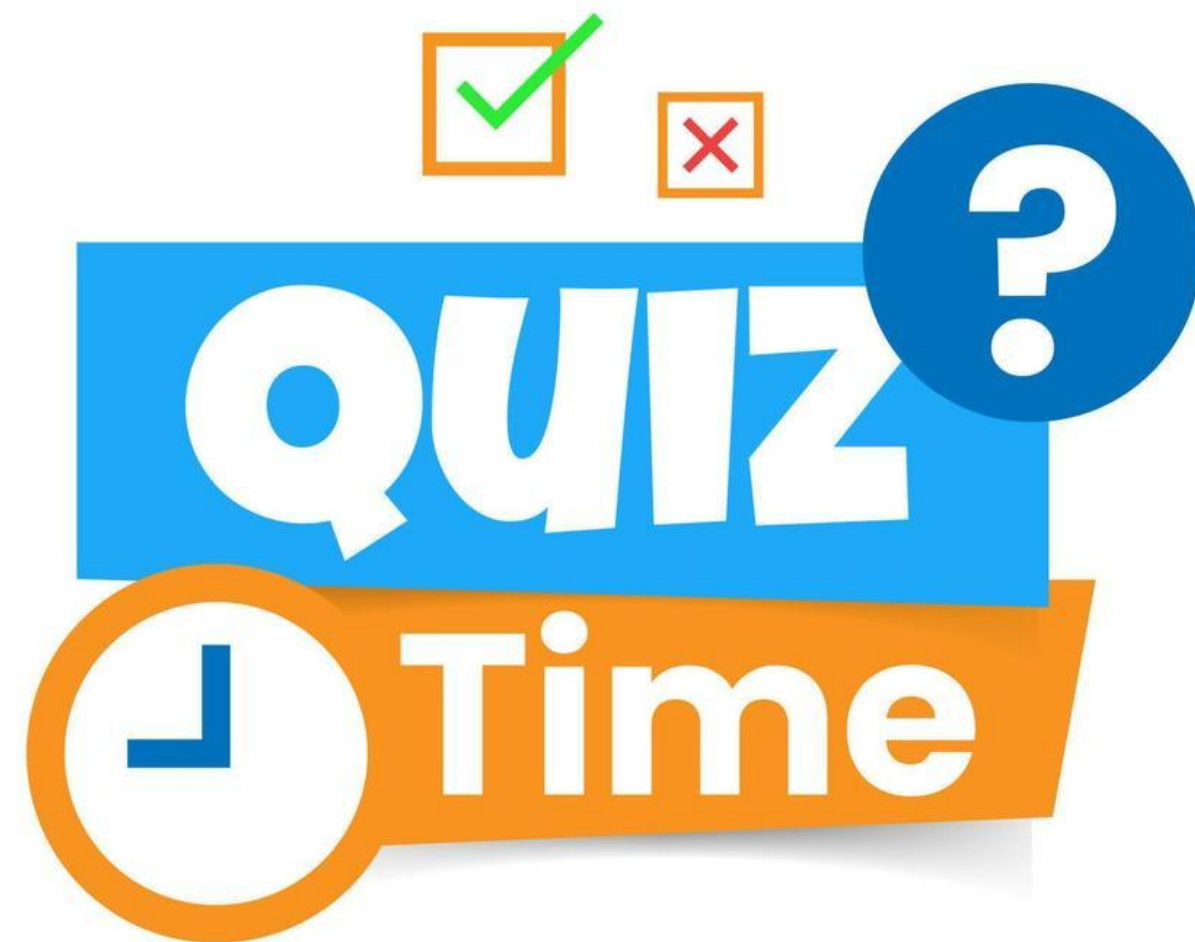
- **Text Normalization** is the process of converting text to a standard format, such as to lowercase, removing special characters, and converting numbers to their written form
- **Tokenization** is the process of breaking down text into individual units, such as words or phrases. This is an important step in preparing text data for NLP tasks
- **Stop Words** are common words that are usually removed during text processing, as they do not carry much meaning and can introduce noise or affect the results of NLP tasks. Examples of stop words include "the," "a," "an," "in," and "is"
- **Stemming and Lemmatization** are techniques used to reduce words to their base form. This helps reduce the dimensionality of the data and improve the performance of models

- LLMs are trained using a process called **unsupervised learning**
- This involves **feeding the model massive amounts of text data**, such as books, articles, and websites, and having the model learn the patterns and relationships between words and phrases in the text
- The model is then **fine-tuned on a specific task**, such as language translation or text summarization



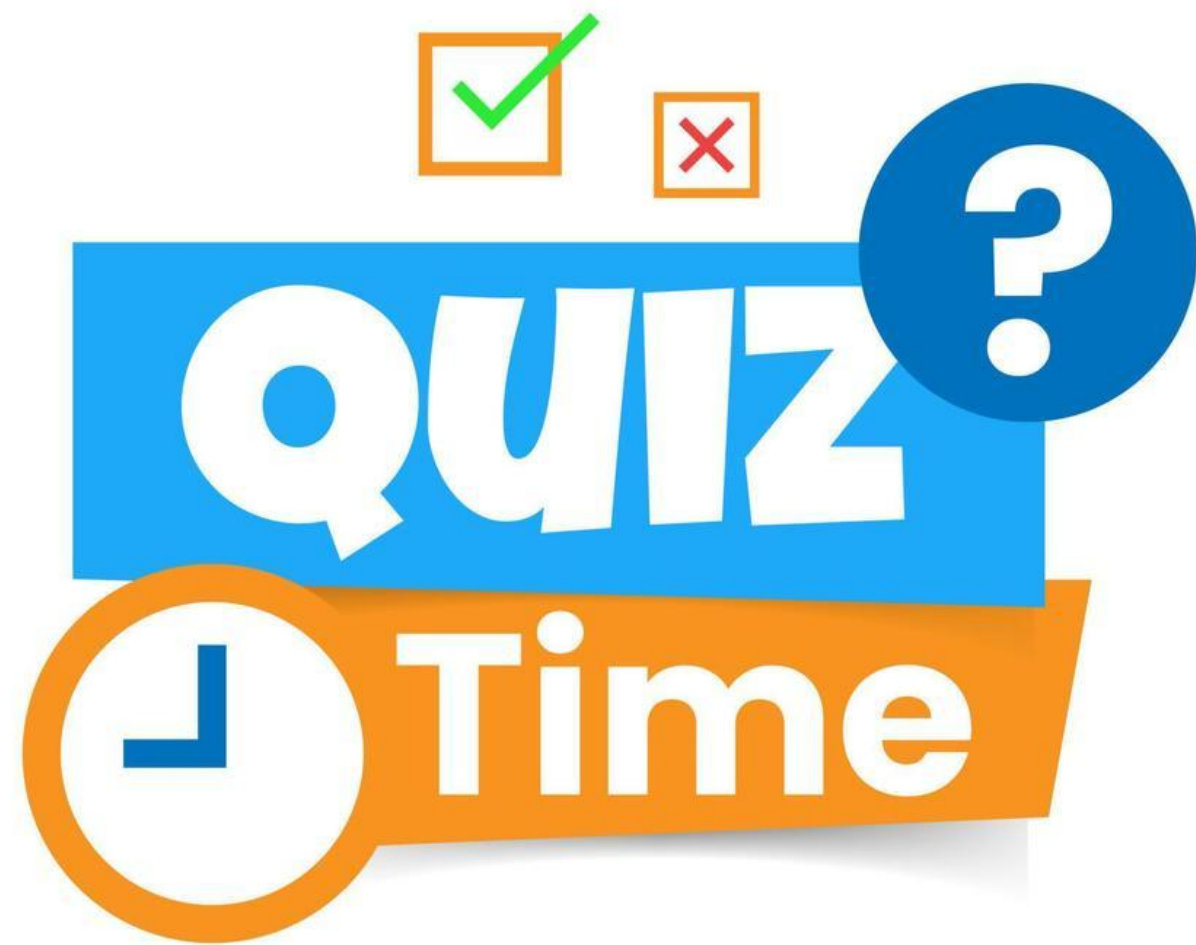
Challenges with Large Language Model

- One of the main challenges with LLMs is the **potential for offensive language**, as the models learn from the patterns found in the training data
- Unethical considerations, such as **gender and racial biases**
- **Amount of computational resources** needed to train and run LLMs, which can be expensive and energy-intensive
- **Making up things** which are not even facts
- While large language models have shown impressive performance on a variety of NLP tasks, they may not perform as well on specific tasks, such as those that require a **deeper understanding of the underlying context**

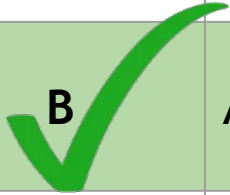


When Generative AI systems just make something up that isn't true, it is called what?

A	A lie
B	A hallucination
C	A bug
D	It is not possible for that to happen



When Generative AI systems just make something up that isn't true, it is called what?

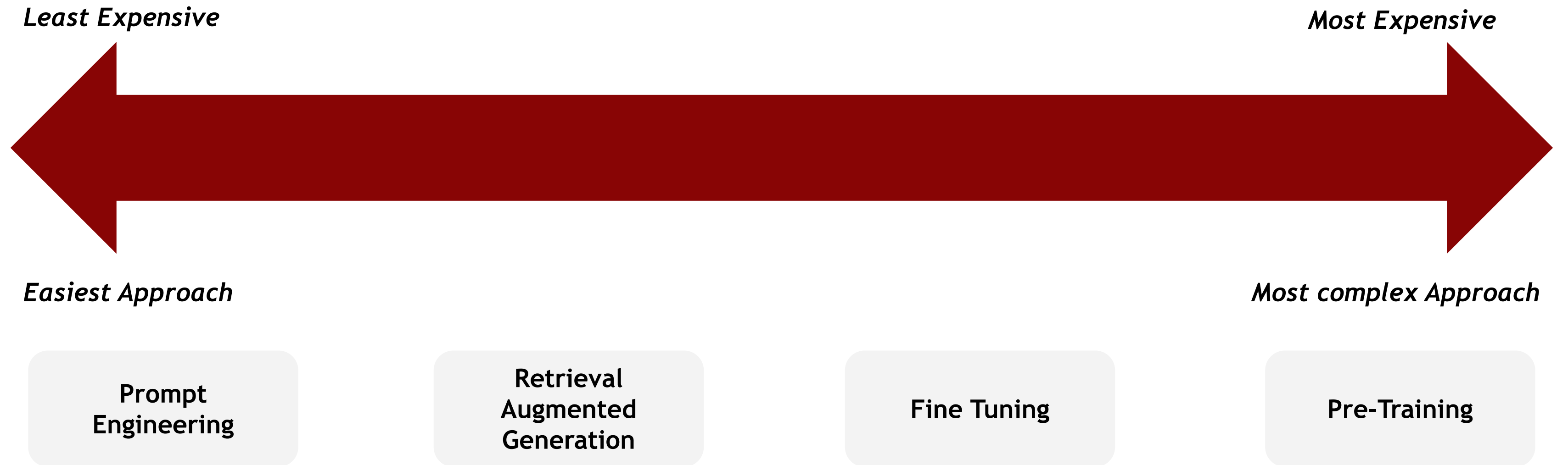
A	A lie
 B	A hallucination
C	A bug
D	It is not possible for that to happen

Agenda

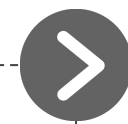
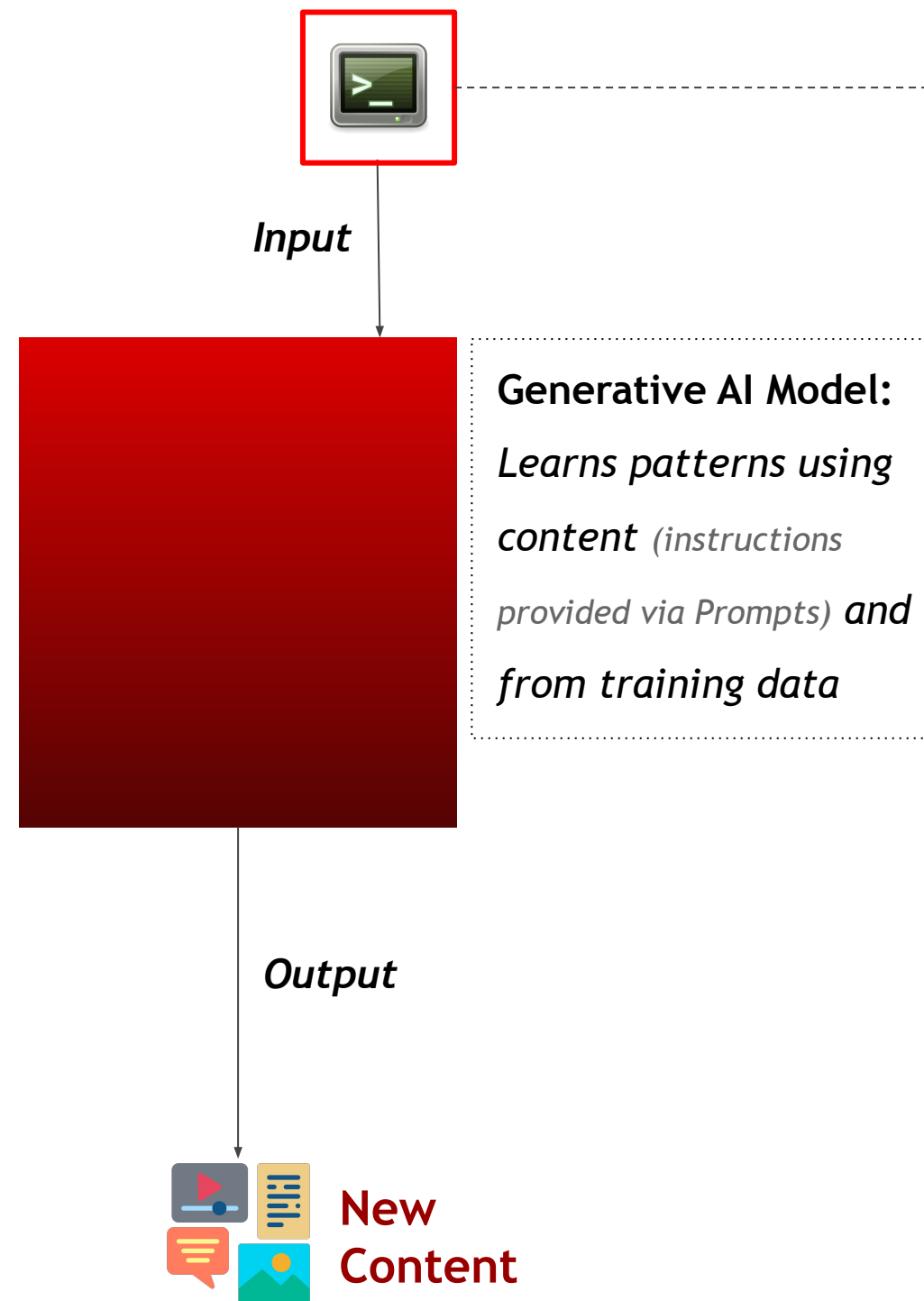
- Recap of LLMs
- ● **Understanding Prompt Design**
 - Context, Examples, & Tuning Prompt Parameters
 - Prompt Design Best Practices

Ways to customize Large Language Models

Accuracy Requirements, Skillset, Compute Resources, & Data Availability influence the choice of approach



What is a Prompt?



A prompt refers to a specific instruction, query, or input given to an AI model to generate a response or perform a task



It serves as cue, guidance, or a starting point for the AI system to produce an output based on its learned knowledge and language comprehension

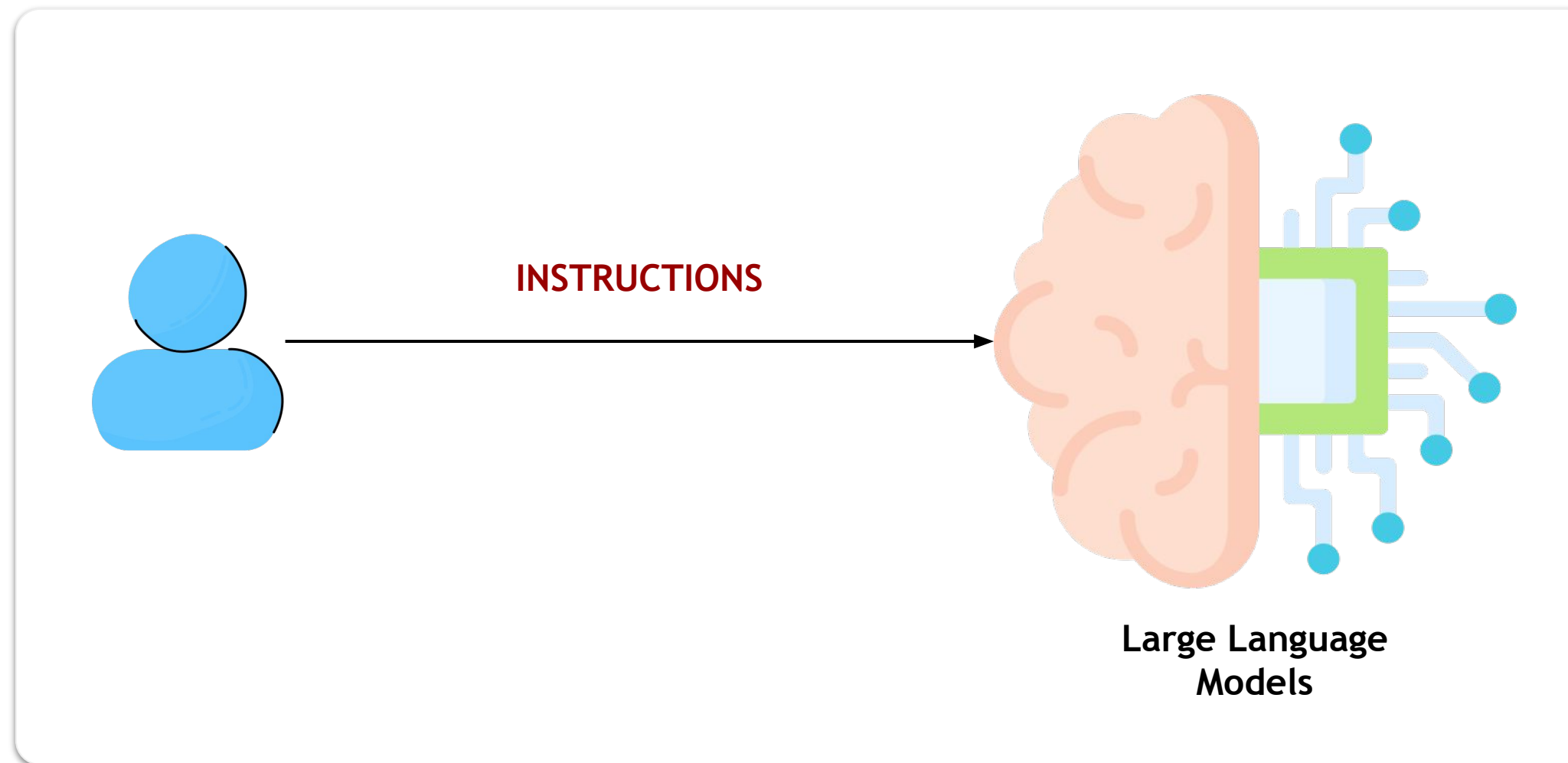


The form and tone of a prompt varies depending on the AI application and the desired outcome



Well-defined, context-driven, and unambiguous prompts bring AI output closer to the intended results

What is Prompt Engineering?



Prompt Engineering is an art of asking the right question to get the best response from an LLM

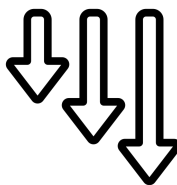
Why is Prompt Engineering Important?



Improve the Accuracy



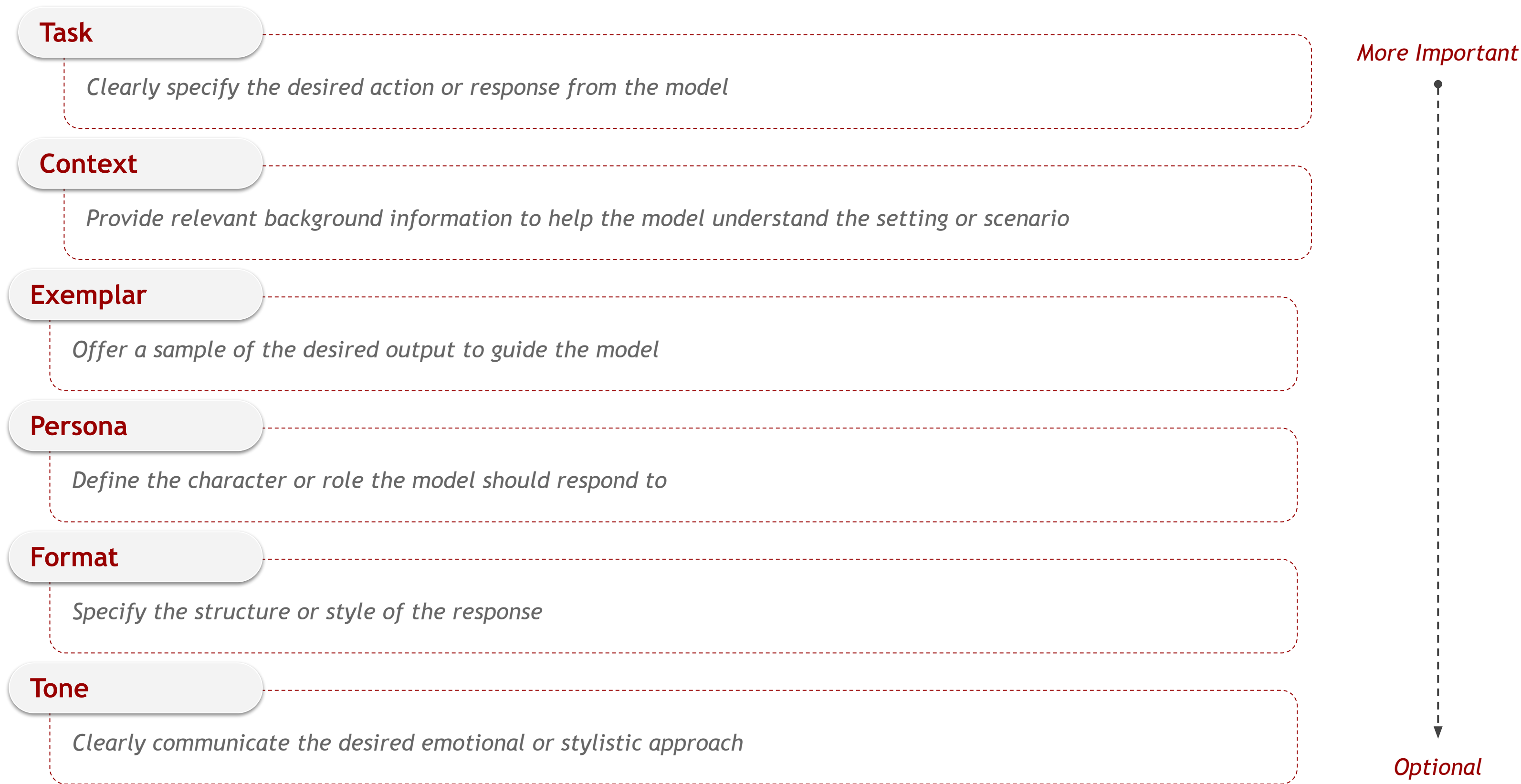
Improve the usefulness of AI generated content



Reduce the risk of generating harmful and biased content

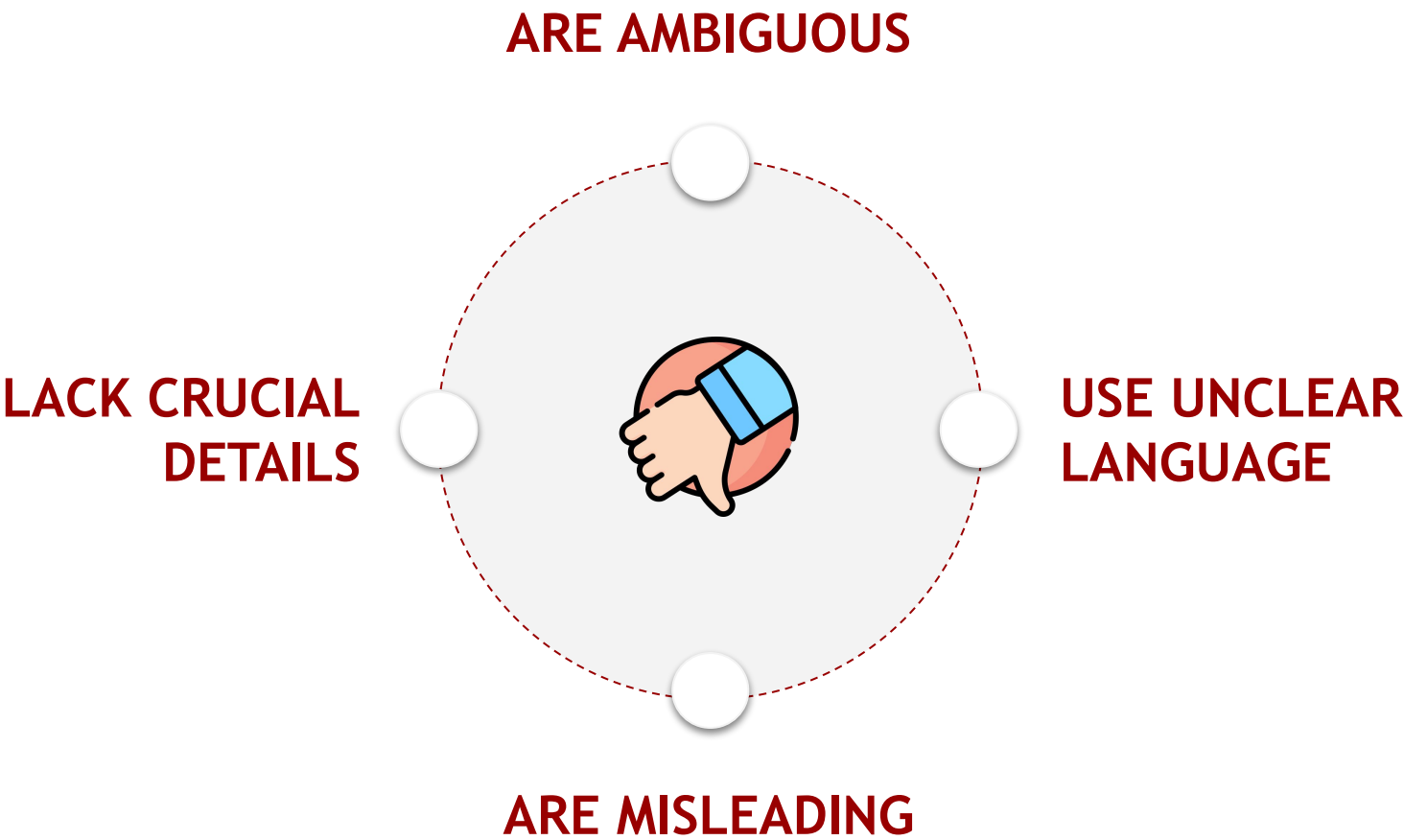
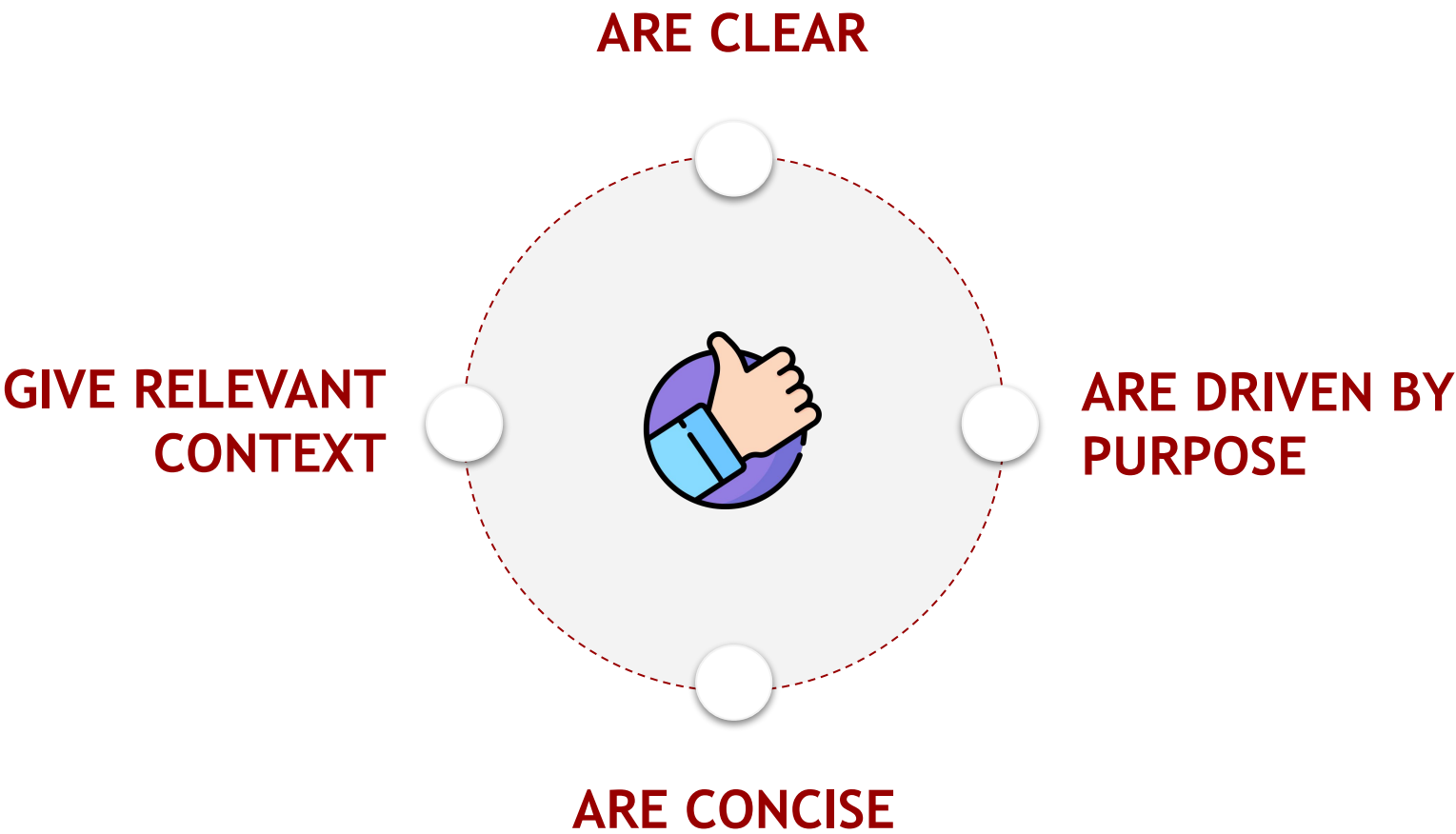
Key elements of a Prompt

Let’s look at aspects that make up a good prompt:

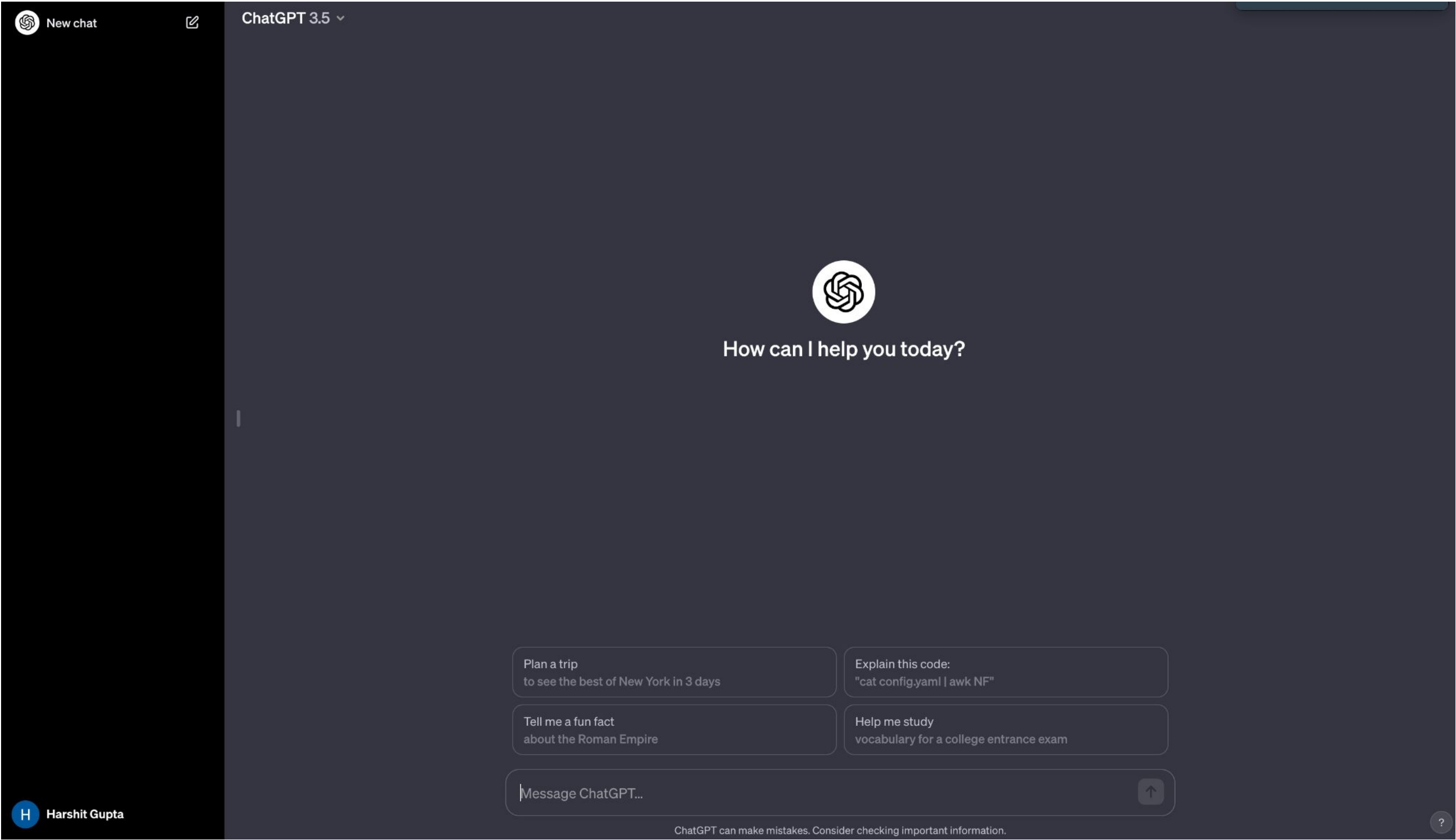


Good Prompts vs. Bad Prompts

The quality of the prompt significantly influences the output generated by the model. Here are the traits of Good & Bad Prompts:

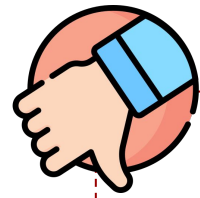


ChatGPT UI



Tips for designing Prompts

Write Clear Instructions



Tell me something about Generative AI



Act as an AI Engineer working for a multinational Consumer Packaged Goods (CPG) organization. One of the CXOs have asked you to explain about “What is Generative AI?” through a concise article, just to keep up with the latest trends in the market. Now provide me with an explanation of Generative AI using following instructions:

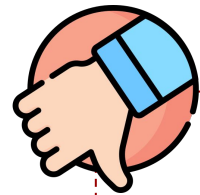
Change your prompt as per following to make it clearer:

- *Include details in your query to get more relevant answers*
- *Ask the model to adopt a persona*
- *Use delimiters to clearly indicate distinct parts of the input*
- *Specify the steps required to complete a task*
- *Provide examples*
- *Specify the desired length of the output*

1. *It should be easy to understand and must be tweaked as per CPG Industry*
2. *Overall length of article should not be more than 500 words*
3. *Add relevant examples and some cases with the source links*

Tips for designing Prompts

Ask to provide reference text when required



Share some case studies related to the adoption of Generative AI in consumer packaged goods field.



Share some case studies related to the adoption of Generative AI in consumer packaged goods domain. You can take references from websites like gartner.com, mckinsey.com, forbes.com

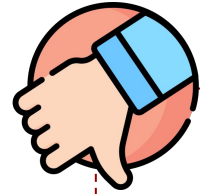
Also add proper source links for the case studies shared in the response.

Change your prompt like this to make it clearer:

- *Instruct the model to answer using a reference text/website*
- *Instruct the model to answer with citations from a reference text*

Tips for designing Prompts

Split a complex task into simpler subtasks



In order to propel our Consumer Packaged Goods (CPG) company into the next era of operational excellence, we are seeking a comprehensive strategy to evolve from a Data-Driven Organization to an AI-Driven Organization. Please provide a detailed plan

Use the following tactics to improve the prompt:

- *Use intent classification to identify the most relevant instructions for a user query*
- *For dialog-driven applications that require very long conversations, summarize or filter previous dialog(s)*
- *Summarize long documents piecemeal and construct a complete summary, recursively*



To propel our CPG organization towards an AI-driven future, we seek a detailed strategy for transitioning from a Data-Driven to an AI-Driven Organization. Key focus areas include:

- *Data Integration and Enrichment:*
 - *Propose methods for enhanced data integration, ensuring seamless interoperability and enrichment from external sources*
- *Advanced Analytics Capabilities:*
 - *Specify tools to extract actionable insights, surpassing traditional analytics, and gaining predictive capabilities*
- *Machine Learning Applications:*
 - *Identify use cases for machine learning in decision-making, supply chain management, and demand forecasting, with a scalable implementation plan*
- *Automation and Process Enhancement:*
 - *Propose AI-driven automation solutions to streamline tasks, enhance efficiency, and address integration challenges*
- *Cultural Shift and Skill Development:*
 - *Devise a plan for fostering AI adoption culture, promoting awareness, enthusiasm, and workforce skill development*
- *Performance Monitoring and Iterative Improvement:*
 - *Establish KPIs for measuring AI initiative success, implement a monitoring framework, and detail a feedback loop for continuous improvement*

Provide a comprehensive, phased roadmap aligning with organizational goals, emphasizing smooth transition and sustained evolution. The plan should prioritize data integration, advanced analytics, machine learning, automation, cultural adaptation, and iterative improvement for a successful shift to an AI-Driven Organization

Tips for designing Prompts

Test changes systematically. Playground is a nice way to test for Developers

Evaluate model outputs with reference to gold-standard answers



You will be provided with text delimited by triple quotes that is supposed to be the answer to a question. Check if the following pieces of information are directly contained in the answer:

- Neil Armstrong was the first person to walk on the moon.*
- The date Neil Armstrong first walked on the moon was July 21, 1969.*

For each of these points perform the following steps:

- 1 - Restate the point.*
- 2 - Provide a citation from the answer which is closest to this point.*
- 3 - Consider if someone reading the citation who doesn't know the topic could directly infer the point. Explain why or why not before making up your mind.*
- 4 - Write "yes" if the answer to 3 was yes, otherwise write "no".*

Finally, provide a count of how many "yes" answers there are. Provide this count as {"count": <insert count here>}.

""Neil Armstrong is famous for being the first human to set foot on the Moon. This historic event took place on July 21, 1969, during the Apollo 11 mission.""

Techniques in Prompt Engineering

Here, we see more intricate strategies that require a deeper understanding of the model's behavior

ZERO-SHOT PROMPTING

This technique involves providing the model with a task it hasn't seen during its training. It tests the model's ability to generalize and produce relevant outputs without relying on prior examples

Classify the text into neutral, negative or positive.

Text: I think the vacation was okay.

Sentiment:

Neutral

Techniques in Prompt Engineering

Here, we see more intricate strategies that require a deeper understanding of the model's behavior

FEW-SHOT PROMPTING

Here, the model is given a few examples (shots) to guide its response. By providing exemplars, the model can better understand and generate the desired output. For example, showing a model several examples of translated sentences before asking it to translate a new one

Craft a compelling tagline for social media introducing our upcoming product, a sustainable packaging initiative for our soft-drink line. Highlight the eco-friendly features and emphasize the positive impact on the environment.

Example1: Every sip is a step towards a greener future

Example2: Good for you, great for the Earth

Example3: Because flavor shouldn't cost the Earth. Share the taste, share the planet

Sip sustainably, savor responsibly

Techniques in Prompt Engineering

Here, we see more intricate strategies that require a deeper understanding of the model's behavior

CHAIN OF THOUGHTS (COT)

It involves guiding the model through a series of reasoning steps. By breaking down a complex task into intermediate steps or a "chains of reasoning," the model can achieve better language understanding and more accurate outcomes

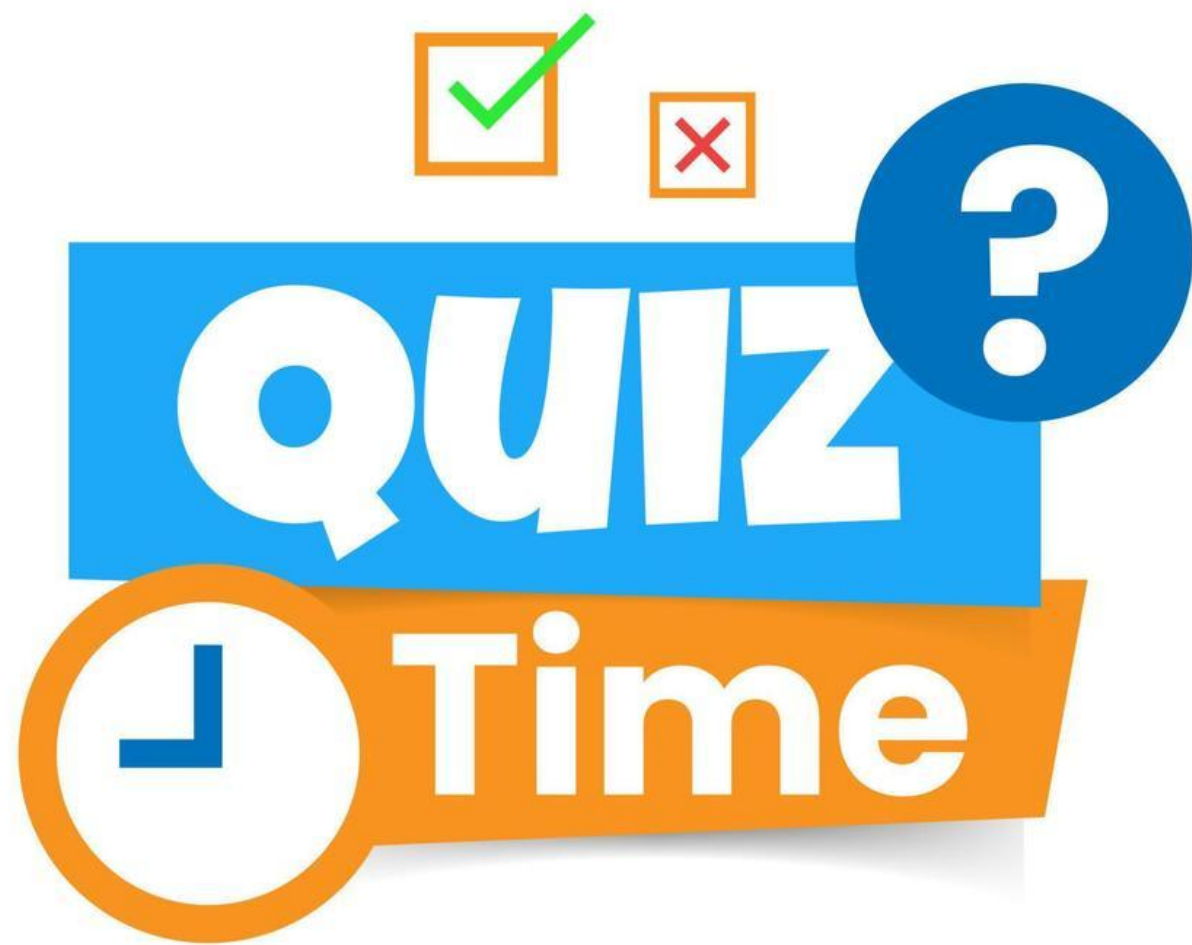
Process the problem step-by-step and answer the question asked:

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

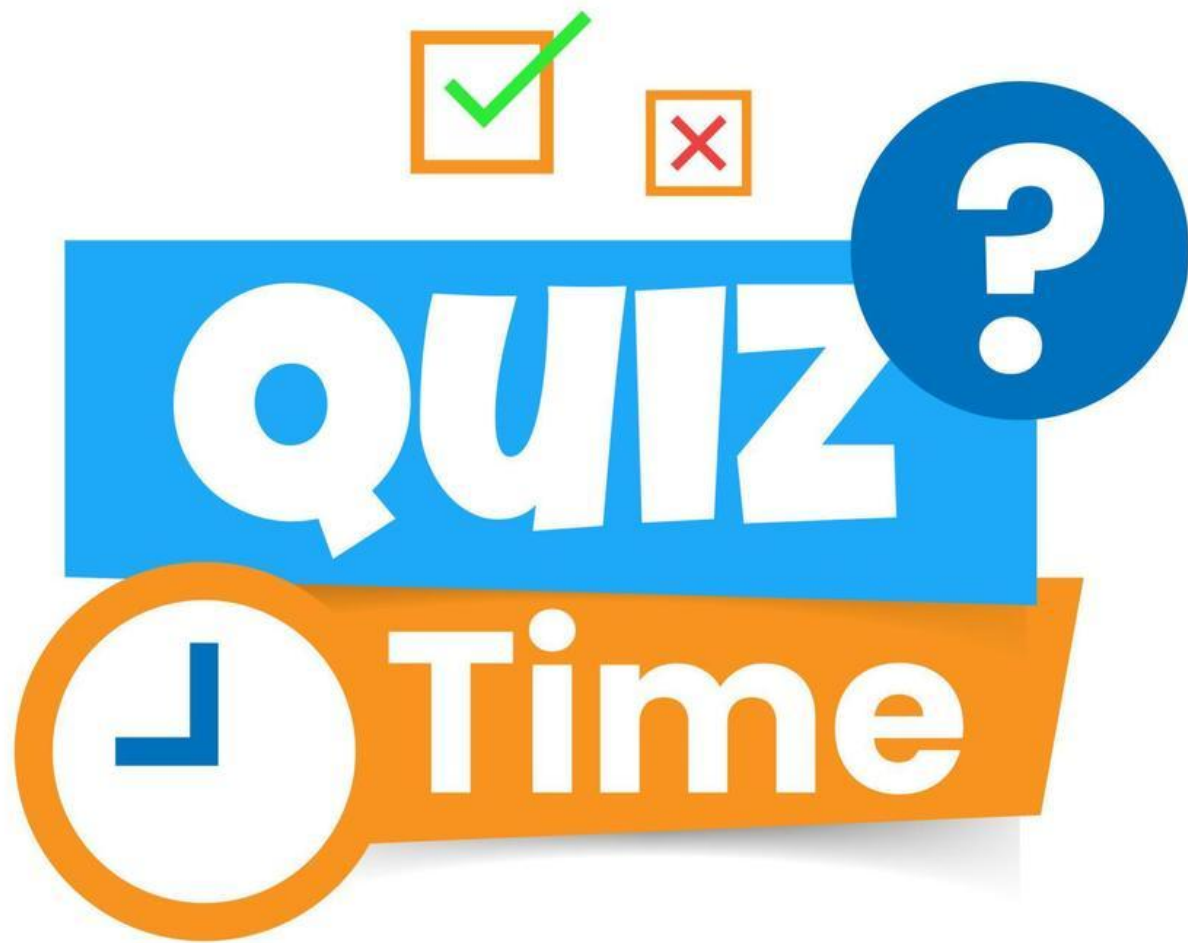
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9

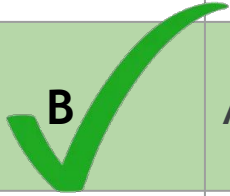


If you want the model to write in your style what should you do?

A	Use clear language
B	Add examples
C	Add context
D	Add purpose



If you want the model to write in your style what should you do?

A	Use clear language
 B	Add examples
C	Add context
D	Add drive purpose

Agenda

- Recap of LLMs
- Understanding Prompt Design
- ● **Context, Examples, & Tuning Prompt Parameters**
- Prompt Design Best Practices

OpenAI Developer Playground

The side menu bar includes options for *Playground, Assistant, Fine Tuning, API Keys, Files, Usage Stats, and Settings*

Get started

Enter an instruction or select a preset, and watch the API respond with a completion that attempts to match the context or pattern you provided.

You can control which model completes your request by changing the model.

KEEP IN MIND

Use good judgment when sharing outputs, and attribute them to your name or company. [Learn more.](#)

Requests submitted to our API and Playground will not be used to train or improve future models. [Learn more.](#)

Our default models' training data cuts off in 2021, so they may not have knowledge of current events.

Playground

Chat

Your presets

Save

View code

Share

...

SYSTEM

You are a helpful assistant.

Add the context and scenarios for the system to provide the responses in certain tone

USER

Enter a user message here.

+ Add message

Provide user question or inputs for which you are seeking responses

Model

gpt-3.5-turbo

Temperature 1

Maximum length 256

Stop sequences

Enter sequence and press Tab

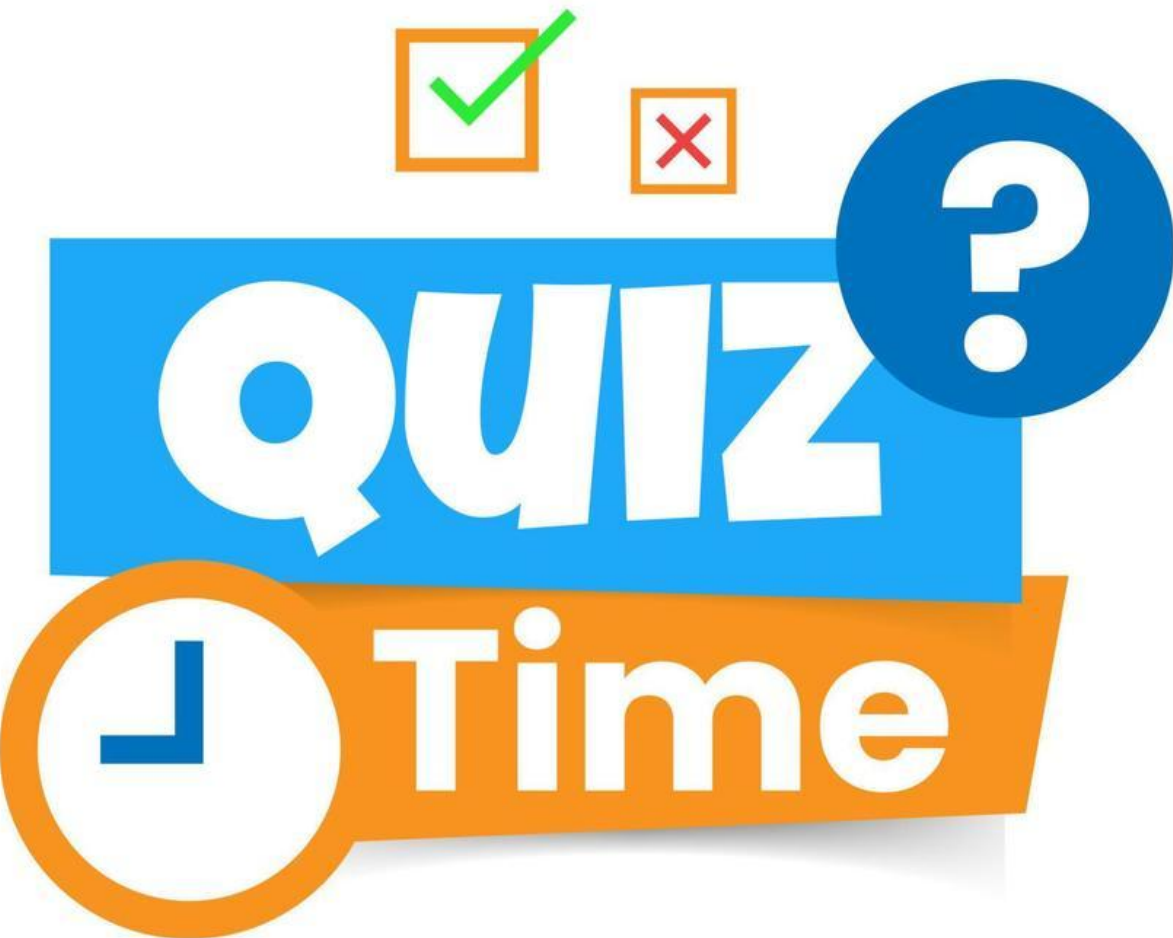
Top P 1

Frequency penalty 0

Presence penalty 0

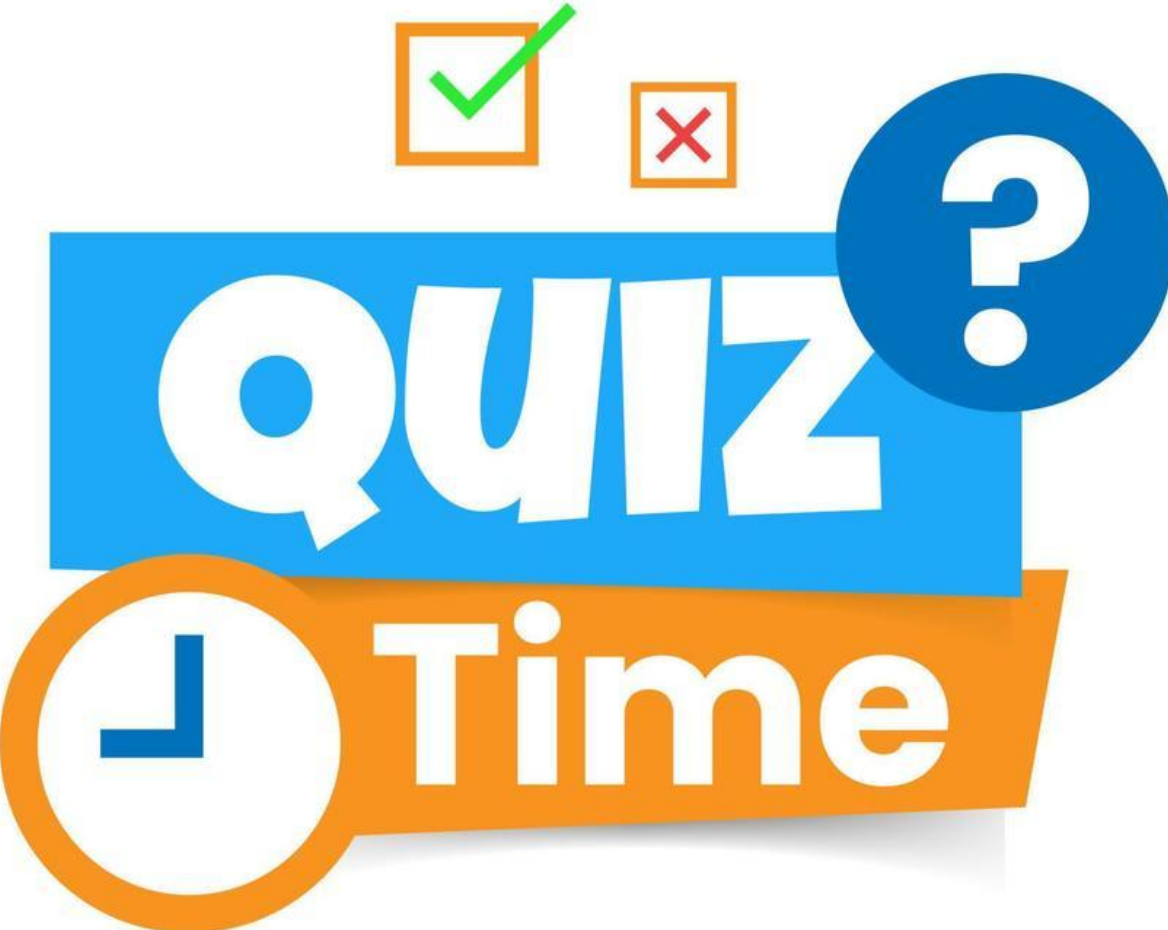
API and Playground requests will not be used to train our models. [Learn more](#)

Tune the responses using these parameters

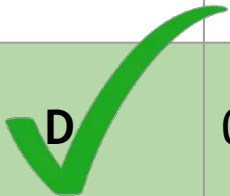


Which setting for temperature would make the model more creative when answering prompts?

A	-1
B	0
C	0.1
D	0.9



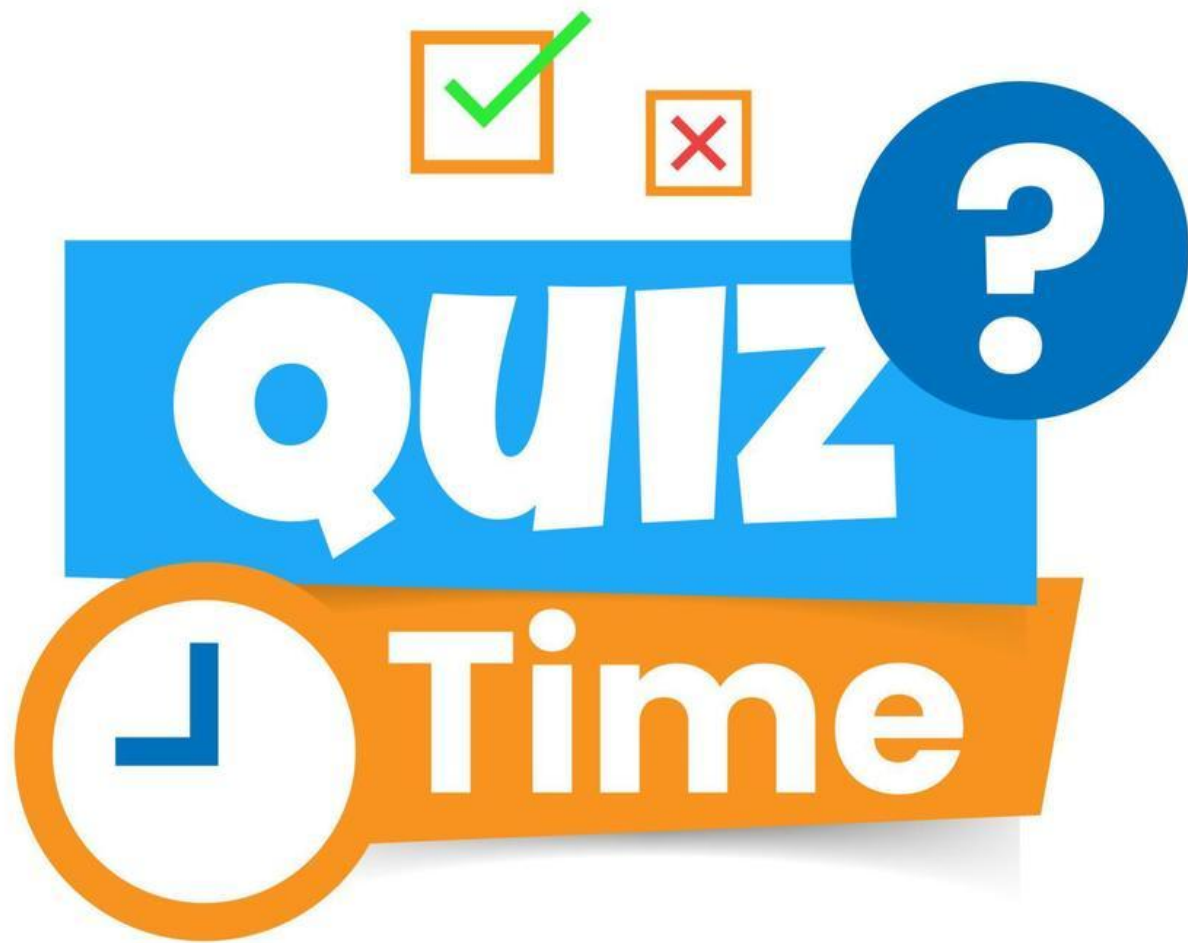
Which setting for temperature would make the model more creative when answering prompts?

A	-1
B	0
C	0.1
 D	0.9

Freeform vs Structured Prompts in Azure OpenAI Studio

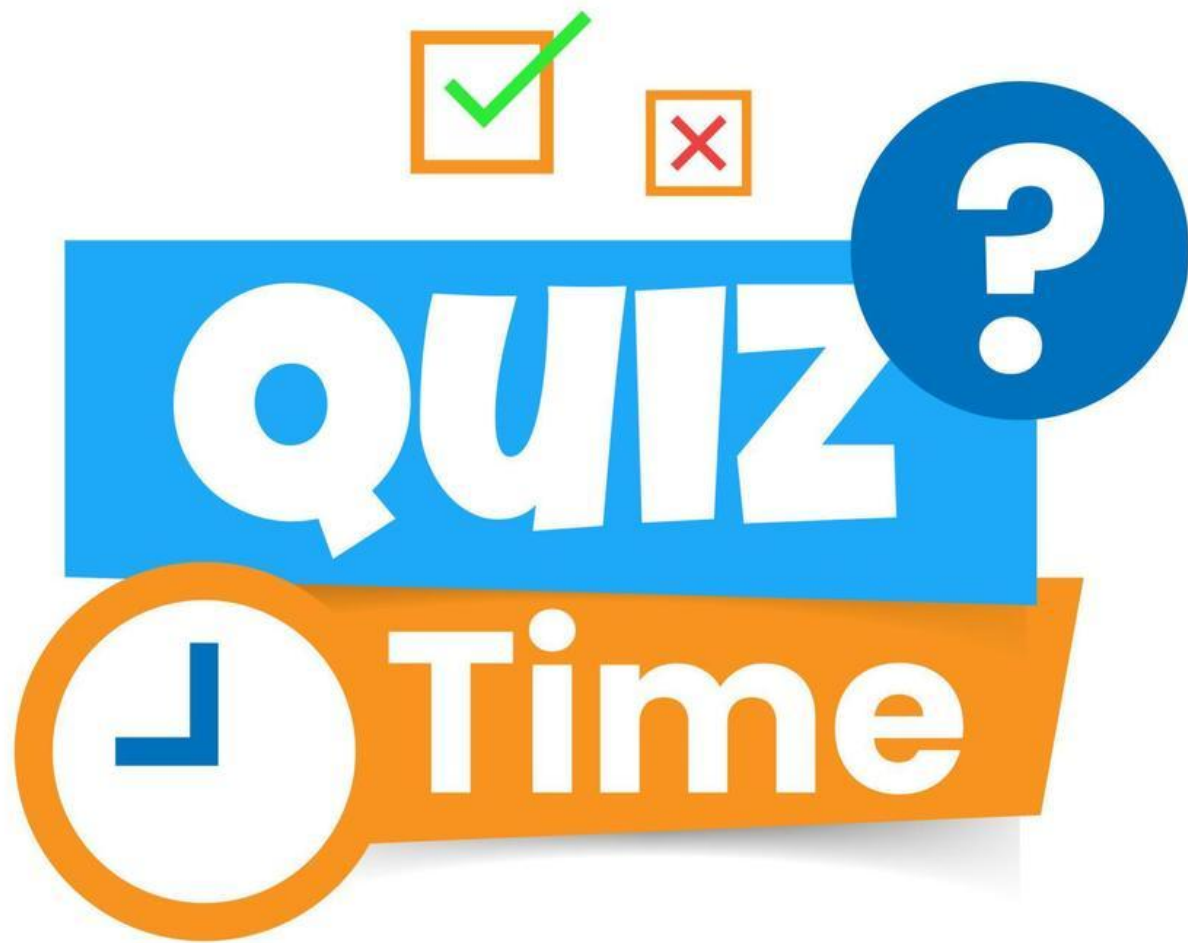
- So far in class, all the examples shown have been freeform prompts
 - Consist only of the prompt
- Structured prompts in Azure OpenAI Studio allow you to add system messages(context) and examples
 - Context tells the model more about your use case characteristics
 - Examples show the model how to structure the response





What is difference between Freeform and Structured prompts?

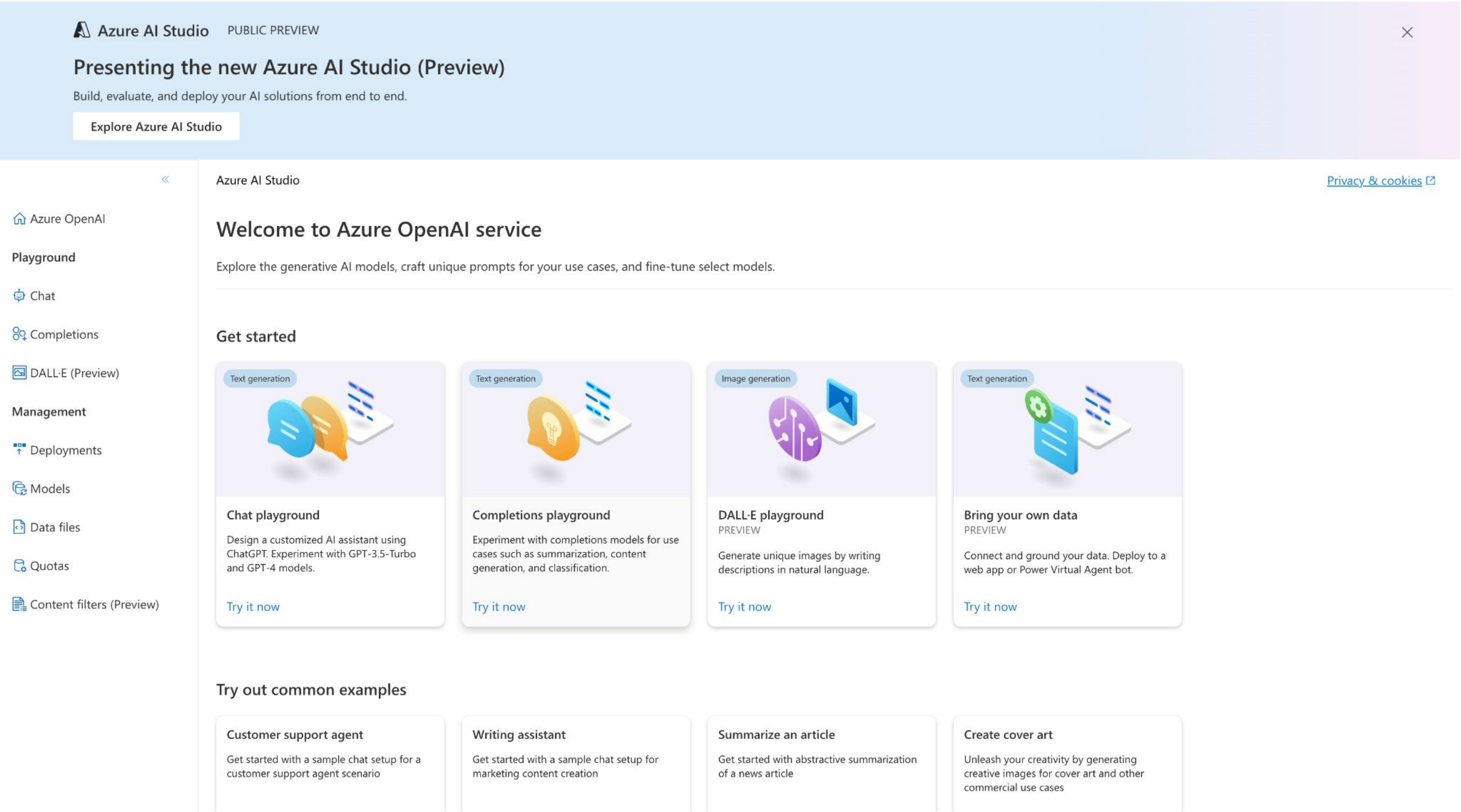
A	Freeform prompts restrict input options, while Structured prompts lack context
B	Freeform prompts and Structured prompts have no notable differences
C	With structured prompts, you can add context and examples
D	Freeform prompts are more contextually rich, while Structured prompts limit input variations



What is difference between Freeform and Structured prompts?

A	Freeform prompts restrict input options, while Structured prompts lack context
B	Freeform prompts and Structured prompts have no notable differences
C	With structured prompts, you can add context and examples
D	Freeform prompts are more contextually rich, while Structured prompts limit input variations

Azure OpenAI Studio UI for Prompts



LogIn using your Azure OpenAI Credentials at: <https://oai.azure.com/>

Azure OpenAI Studio UI for Prompts

Azure AI Studio

PUBLIC PREVIEW

<<

Azure OpenAI

Playground

Chat

Completions

DALL-E (Preview)

Management

Deployments

Models

Data files

Quotas

Content filters (Preview)

Azure AI Studio


Privacy & cookies

Welcome to Azure OpenAI service

Explore the generative AI models, craft unique prompts for your use cases, and fine-tune select models.

Get started

Text generation




Chat playground

Design a customized AI assistant using ChatGPT. Experiment with GPT-3.5-Turbo and GPT-4 models.

Try it now

Text generation




Completions playground

Experiment with completions models for use cases such as summarization, content generation, and classification.

Try it now

Image generation




DALL-E playground
PREVIEW

Generate unique images by writing descriptions in natural language.

Try it now

Text generation



Bring your own data
PREVIEW

Connect and ground your data. Deploy to a web app or Power Virtual Agent bot.

Try it now

Try out common examples

Customer support agent

Get started with a sample chat setup for a customer support agent scenario

Writing assistant

Get started with a sample chat setup for marketing content creation

Summarize an article

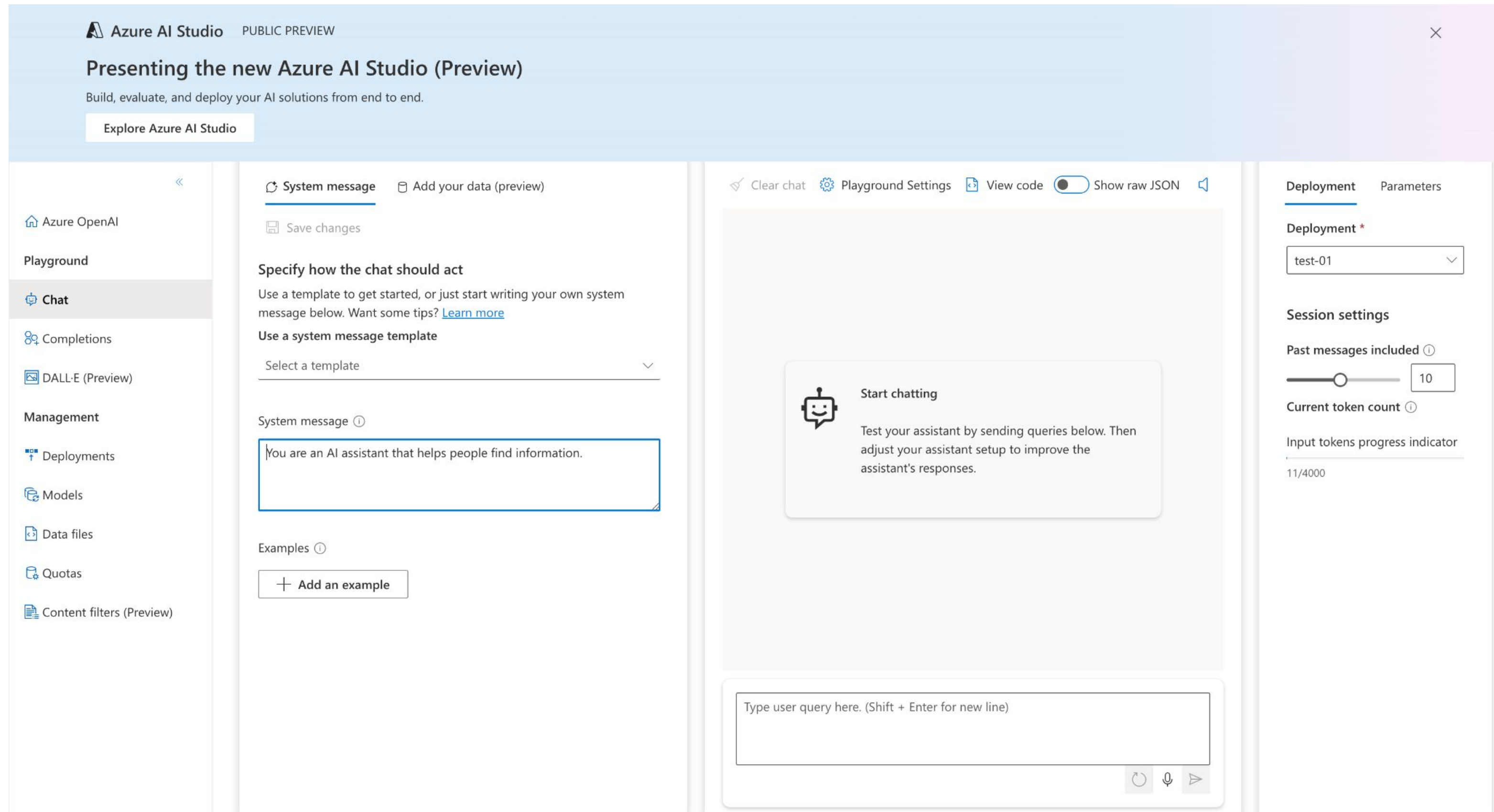
Get started with abstractive summarization of a news article

Create cover art

Unleash your creativity by generating creative images for cover art and other commercial use cases

Select the **Chat** option out of many options available on the left hand side section of the page

Azure OpenAI Studio UI for Prompts



Azure OpenAI Studio UI for Prompts

Azure AI Studio > Chat playground

Chat playground

[Privacy & cookies](#)

Deploy to

Import setup Export setup Show panels

Examples demonstrate
desired output

Assistant setup

Save changes

Specify how the chat should act

Use a template to get started, or just start writing your own system message below. Want some tips? [Learn more](#)

Use a system message template

Select a template

System message

You are an AI assistant that helps people find information.

Examples

+ Add an example

System message
adds information
about the use case

Prompt and response

Chat session

Clear chat



Playground
Settings



View
code



Show raw
JSON



Start chatting

Test your assistant by sending queries below. Then adjust your assistant setup to improve the assistant's responses.

Type user query here. (Shift + Enter for new line)

Configuration

Deployment

Parameters

Max response

800

Temperature

0.7

Top P

0.95

Stop sequence

Stop sequences

Frequency penalty

0

Presence penalty

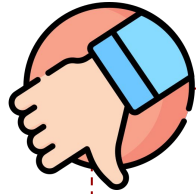
0

Quality of the output can be improved by providing system message to the prompt

- System message can be a few words or a paragraph
- You can specify information about you or your organization
- Specify a genre or style for the results

Tips for designing Prompts

Give model time to “Think”



Determine if the My solution is correct or not.

Problem Statement: I'm building a solar power installation and I need help working out the financials.

- Land costs \$100 / square foot*
- I can buy solar panels for \$250 / square foot*
- I negotiated a contract for maintenance that will cost me a flat \$100k per year, and an additional \$10 / square foot*

What is the total cost for the first year of operation as a function of the number of square feet.

My Solution: Let x be the size of the installation in square feet.

- 1. Land cost: $100x$*
 - 2. Solar panel cost: $250x$*
 - 3. Maintenance cost: $100,000 + 10x$*
- Total cost: $100x + 250x + 100,000 + 10x = 360x + 100,000$*



First work out your own solution to the problem. Then compare your solution to My solution and evaluate if my solution is correct or not. Don't decide if my solution is correct until you have done the problem yourself.

Problem Statement: I'm building a solar power installation and I need help working out the financials.

- Land costs \$100 / square foot*
- I can buy solar panels for \$250 / square foot*
- I negotiated a contract for maintenance that will cost me a flat \$100k per year, and an additional \$10 / square foot*

What is the total cost for the first year of operations as a function of the number of square feet.

My Solution: Let x be the size of the installation in square feet.

- 1. Land cost: $100x$*
 - 2. Solar panel cost: $250x$*
 - 3. Maintenance cost: $100,000 + 10x$*
- Total cost: $100x + 250x + 100,000 + 10x = 360x + 100,000$*

Suggest the following to the model to reach a better and optimal solution:

- *Instruct the model to work out its own solution before rushing to a conclusion*
- *Use inner monologue or a sequence of queries to hide the model's reasoning process*
- *Ask the model if it missed anything on previous passes*

Here are the results of a prompt with no context

Chat session



Clear chat



Playground Settings



View code



Show raw JSON



Write for me an ad about our new orange vest



Introducing our brand new orange vest! If you're looking for a high-visibility garment that can keep you safe while you work, look no further. Our vest is made from durable, lightweight materials that are perfect for any job site. The bright orange color ensures that you'll be visible to others, even in low light conditions. The vest is also designed with multiple pockets to keep your tools and gear organized and easily accessible. Whether you're a construction worker, road worker, or just need a vest for outdoor activities, our orange vest is the perfect choice. Order yours today and stay safe while you work!

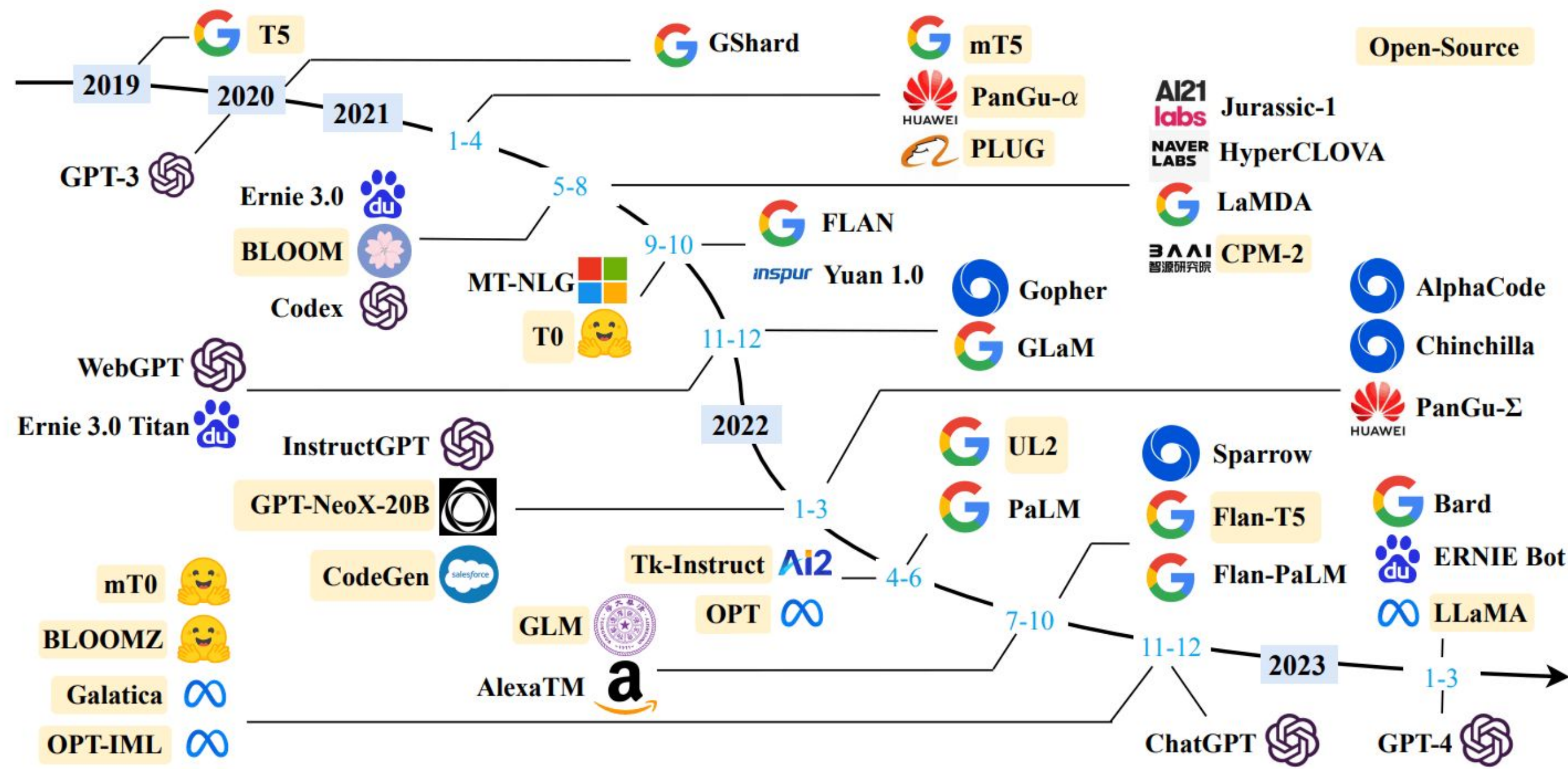


Type user query here. (Shift + Enter for new line)



Various Large Language Models

The most widely used foundation models and their Timeline



54% Of companies are investing heavily in LLMs after recently realizing the capabilities of LLMs

A timeline of existing large language models (having a size larger than 10B) in recent years. We mark the open-source LLMs in yellow color

Let's see how adding context changes the output

System message ⓘ

You work writing ads in the marketing department for a hunting and fishing supply company. The name of the company is Deep Creek Outfitters. The website is www.deepinthecreek.com.

The phone number is: 9876543210

Chat session

 Clear chat  Playground Settings  View code ☒ Show raw JSON 



Attention all hunters! Looking for a high-quality, durable vest that will keep you safe and visible during your next hunting trip? Look no further than Deep Creek Outfitters! Our brand new orange hunting vest is the perfect addition to your hunting gear collection.

Made with premium materials, this vest is designed to withstand even the toughest outdoor conditions. It's lightweight and breathable, so you won't feel weighed down or uncomfortable while wearing it. And best of all, its bright orange color ensures that you'll be visible to other hunters, even in low-light situations.

Whether you're hunting deer, elk, or any other game, our orange vest is a must-have for any serious hunter. Visit our website at www.deepinthecreek.com to order yours today, or call us at 9876543210 to speak with one of our knowledgeable customer service representatives. Trust us, you won't regret it!

Type user query here. (Shift + Enter for new line)



Let's try a different context

System message ⓘ

You write ads for the marketing department of a company called StyleUp that sells clothing targeted towards teens and adults in their 20's.

you sell everything online. The website is www.StyleUp.cool

Chat session

 Clear chat  Playground Settings  View code  Show raw JSON 



Introducing the newest addition to our collection - The Orange Vest! Perfect for those chilly autumn mornings or as an added layer during the winter season. This vest is not only functional, but it's also stylish and trendy. Made with high-quality materials, it's lightweight and comfortable, so you won't feel weighed down while wearing it.

Whether you're running errands, walking your dog, or going on a hike, this vest is perfect for any occasion. The bright orange color will make you stand out in the crowd and add a pop of color to any outfit.

Don't wait any longer, visit our website www.StyleUp.cool and get your hands on this must-have piece. It's time to upgrade your wardrobe with StyleUp's Orange Vest!

Let's try requesting a style of output

System message ⓘ

You write ads for the marketing department of a company called StyleUp that sells clothing targeted towards teens and adults in their 20's.

you sell everything online. The website is www.StyleUp.cool

your ads are always written in the style of Dr. Seuss.



Oh, have you heard? Have you seen?
StyleUp's new orange vest is quite the dream!
It's perfect for teens and young adults,
For those who want to stand out from the crowd.

This vest is cool, it's quite the sight,
It'll make you feel like a shining bright light.
Wear it with jeans, wear it with skirts,
It's versatile enough for any fashion expert.

The color is bold, it's vibrant and fun,
It'll make you feel like you're soaking up the sun.
And the best part, it's available online,
Just visit our website, it's oh so fine.

So what are you waiting for, my friend?
Get yourself a vest that'll never bend.
Order now, before it's too late,
Don't miss out on this amazing fate.

Examples demonstrate what the output should look like

- Use a history of what you have written in the past to get the model to output content in your style:
 - Zero-shot prompting means there are no examples
 - One-shot prompting means there is one example
 - Few-shot prompting means there are multiple examples

Let's add some context and some examples and see how it affects output

System message ⓘ

You work for a chef specializing in southern barbecue with a large audience on Youtube and social media. You post on social media about the latest videos. You like to post with a southern style

Examples ⓘ

User:



Write a post about our latest smoked brisket video

Assistant:

Get ready for a flavor-packed ride down south with my latest youtube video: Smoked Brisket Low and Slow! Slow-smoked to juicy perfection, seasoned with secret spices. Mouthwatering slides included! Join me on the grill ya'll #SouthernBBQ #SmokedBrisket #BBQPassion

+ Add an example

Write a post about our new video about smoking Prime Rib for the holidays.



The holidays are just around the corner, and what better way to impress your family and friends than with a perfectly smoked prime rib? 🍖🔥
Check out my latest youtube video: Smoking Prime Rib for the Holidays. From seasoning to smoking, I'll show you step-by-step how to get that tender, juicy, and flavorful prime rib that will have everyone asking for seconds. 😋👨🍳
Don't forget to share your own mouth-watering photos with us! #SouthernBBQ #SmokedPrimeRib #HolidayFeast

Use the other model parameters to further customize the results

- Max response
- Temperature
- Top P
- Stop sequence
- Frequency penalty
- Presence penalty

Configuration ×

Deployment

Parameters

Max response ⓘ

800

Temperature ⓘ

0.7

Top P ⓘ

0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ

0

Presence penalty ⓘ

0

[Learn more](#) ↗

Max response determines the maximum amount of tokens for each prompt

- The range is from 1 to 4000
 - The default value is 800
- A token is a basic unit of text that the LLM understands
 - When using the OpenAI API, a token is approximately four characters
- A token is a string of characters that represents a concept or object in GenAI
 - For example, the token "dog" represents the concept of a dog

Configuration



Deployment

Parameters

Max response ⓘ



800

Temperature ⓘ



0.7

Top P ⓘ



0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ



0

Presence penalty ⓘ



0

[Learn more](#) ↗

Temperature controls the degree of randomness in token selection

- Range from 0 to 1
- Lower temperatures are good for prompts that require correctness
 - More likely to select the most expected next token
- Higher temperatures can lead to more diverse or unexpected results
 - More creative
- A temperature of 0 is deterministic
 - The highest probability token is always selected
 - Will always return the same result for a given prompt
- Default is a temperature of .7
 - Start there and experiment with your results

Configuration ×

Deployment Parameters

Max response ⓘ

800

Temperature ⓘ

0.7

Top P ⓘ

0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ

0

Presence penalty ⓘ

0

[Learn more](#) ↗

Top-P also changes how the model selects tokens for output

- Range is from 0 to 1
- Tokens are selected from most probable to least until the sum of their probabilities equals the top-P value
 - For example, if tokens A, B, and C have a probability of .3, .2, and .1 and the top-P value is .5, then the model will select either A or B as the next token
- The default top-P value is .95
- The higher the value the more likely the model would be to select a token that is not the most probable one

Configuration ×

Deployment Parameters

Max response ⓘ



800

Temperature ⓘ



0.7

Top P ⓘ



0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ



0

Presence penalty ⓘ



0

[Learn more](#) ↗

Stop sequence describes the point at which the model ends the response

- Make the model end its response at a desired point.
- The model response will end before the specified sequence, so it won't contain the stop sequence text.
- You can include as many as four stop sequences.

Configuration ×

Deployment Parameters

Max response ⓘ



800

Temperature ⓘ



0.7

Top P ⓘ



0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ



0

Presence penalty ⓘ



0

[Learn more](#)

Frequency penalty puts a penalty on a token based on its frequency

- Reduce the chance of repeating a token proportionally based on how often it has appeared in the text so far.
- This decreases the likelihood of repeating the exact same text in a response.
- The default value is 0

Configuration ×

Deployment Parameters

Max response ⓘ

800

Temperature ⓘ

0.7

Top P ⓘ

0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ

0

Presence penalty ⓘ

0

[Learn more](#) ↗

Presence penalty puts a penalty on a token based on its presence

- Reduce the chance of repeating any token that has appeared in the text at all so far.
- This increases the likelihood of introducing new topics in a response.
- The default value is 0.
- This the opposite of frequency penalty

Configuration



Deployment

Parameters

Max response ⓘ



800

Temperature ⓘ



0.7

Top P ⓘ



0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ



0

Presence penalty ⓘ



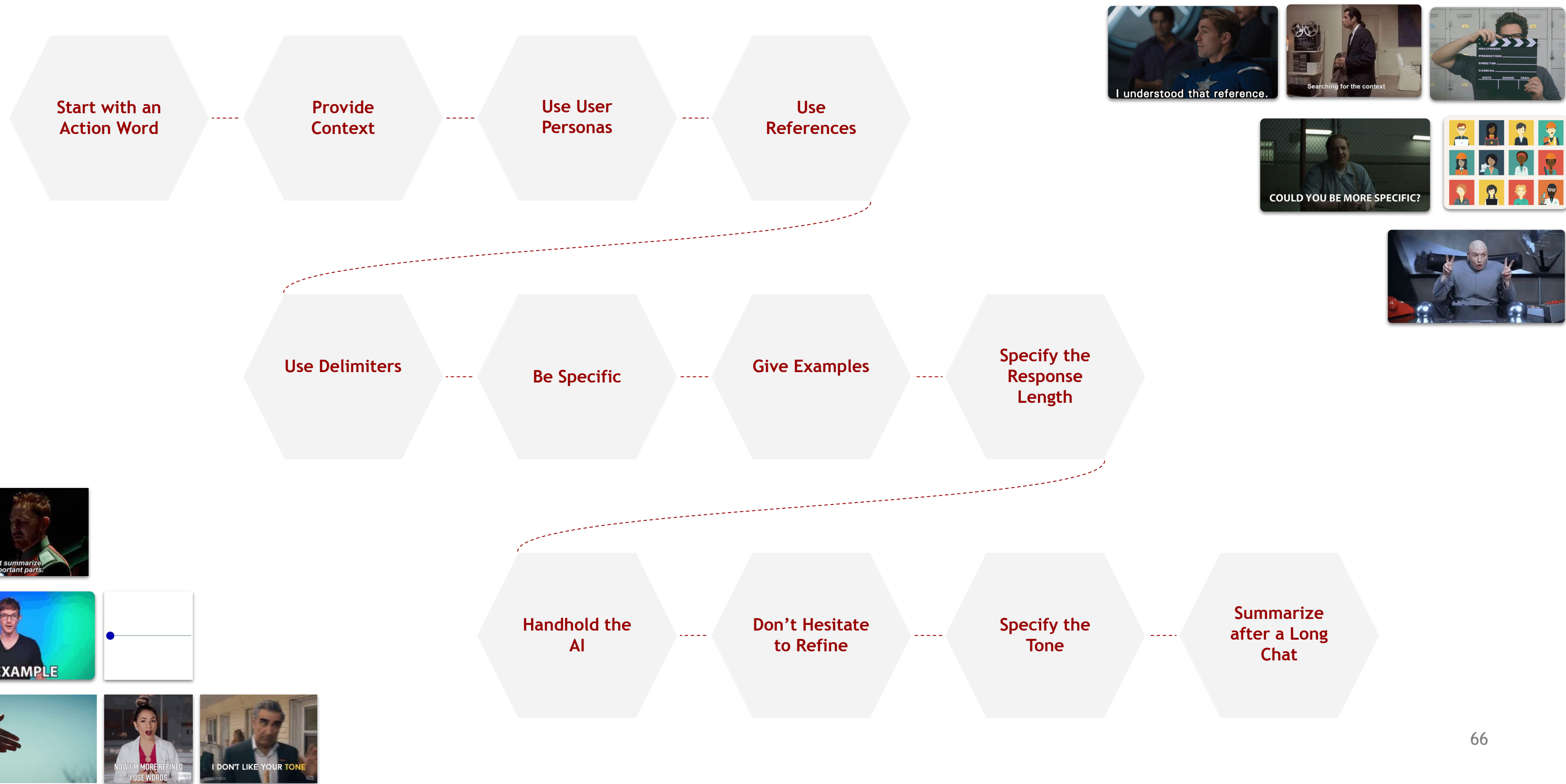
0

[Learn more](#) ↗

Agenda

- Recap of LLMs
- Understanding Prompt Design
- Context, Examples, & Tuning Prompt Parameters
- ● **Prompt Design Best Practices**

Summarizing Prompt Engineering Best Practices

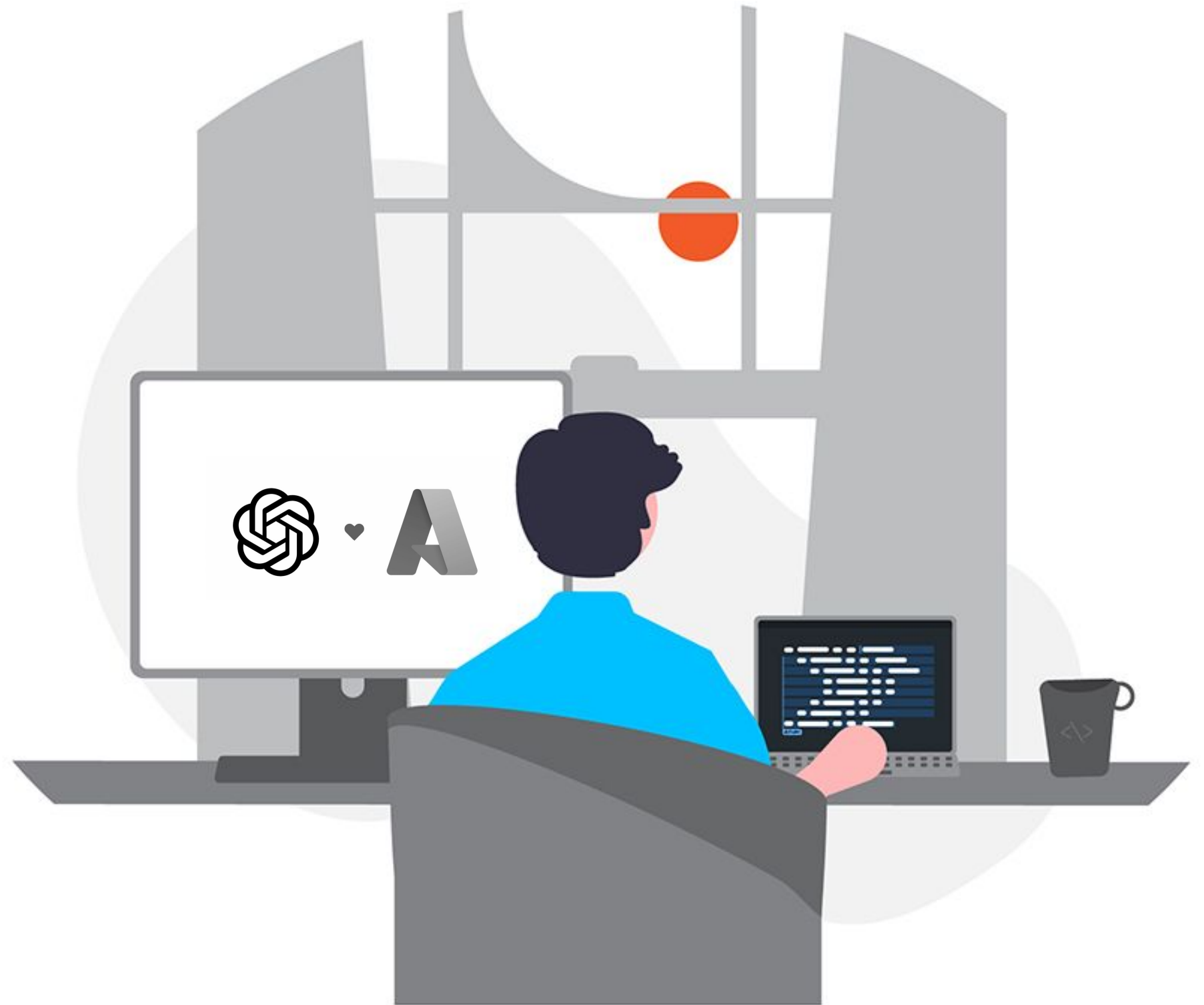


Demo of Text Generation in Azure OpenAI Studio



30 Minutes

Prompt design for Text Generation in Azure OpenAI
Studio



Agenda

- Recap of LLMs
- Understanding Prompt Design
- Context, Examples, & Tuning Prompt Parameters
- Prompt Design Best Practices



Thank You!!

Scan to Stay Connected!

