

Optimizing GenAl Models



Quick Recap



- GenAl designs various types of creative content, including images, text, and music, emphasizing its role in creative endeavors.
- LLMs like **ChatGPT** are advanced Al models that go beyond understanding human language. It also processes and analyze data of various other formats, such as images, graphs, charts.

Engage and Think



You are working on a document and must quickly summarize a lengthy report. Instead of reading the entire report yourself, copy and paste it into a GenAl writing assistant. With a simple prompt like **Summarize the following text in 3-4 concise paragraphs**, the Al generates a clear and concise summary for you in seconds.

How does the AI system understand and extract the key points from a lengthy text to generate a concise and structured summary from the simple prompt you provided?

Learning Objectives

By the end of this lesson, you will be able to:

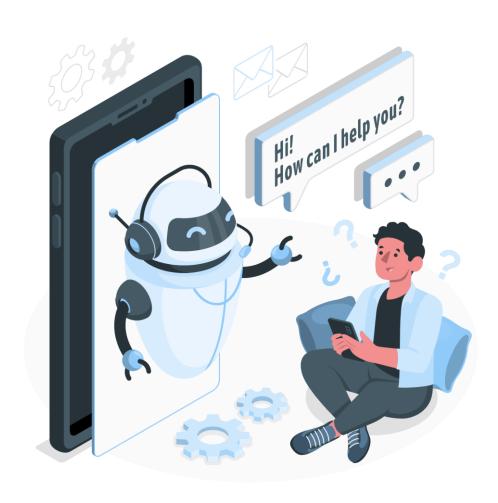
- Generate effective prompts for optimizing interactions with GenAl models
- Apply advanced prompting techniques, such as zero-shot, few-shot, chain-of-thought, self-consistency, and tree-of-thought, to generate the desired output from the GenAI model
- Understand the concepts of pre-training and fine-tuning the GenAl models to enhance domain-specific performance



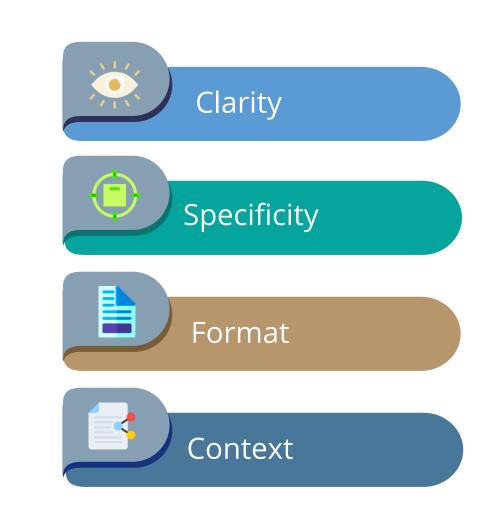
Introduction to Prompt Engineering

Prompt Engineering

Prompt engineering involves instructing conversational AI models, such as ChatGPT, to generate desired outputs.

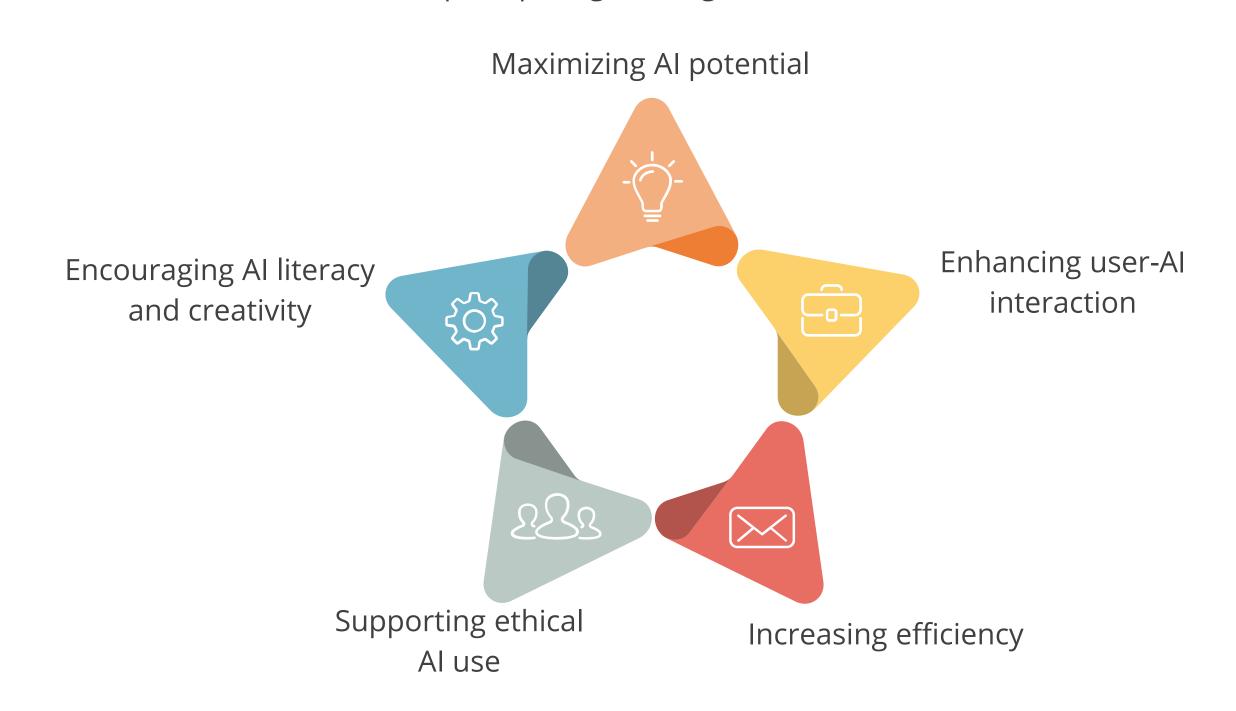


A well-built prompt must have:



Importance of Prompt Engineering

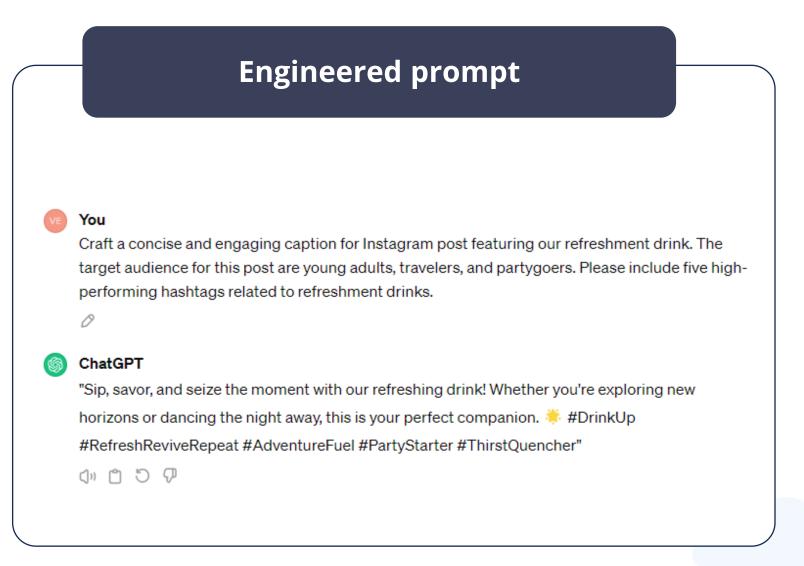
It is increasingly recognized as a crucial skill set in the realm of GenAI. Here are several key reasons why prompt engineering is needed:



Prompts: Examples

Whenever you give a generic prompt, you'll get a random response each time, but engineered prompts will bring back a relevant response.

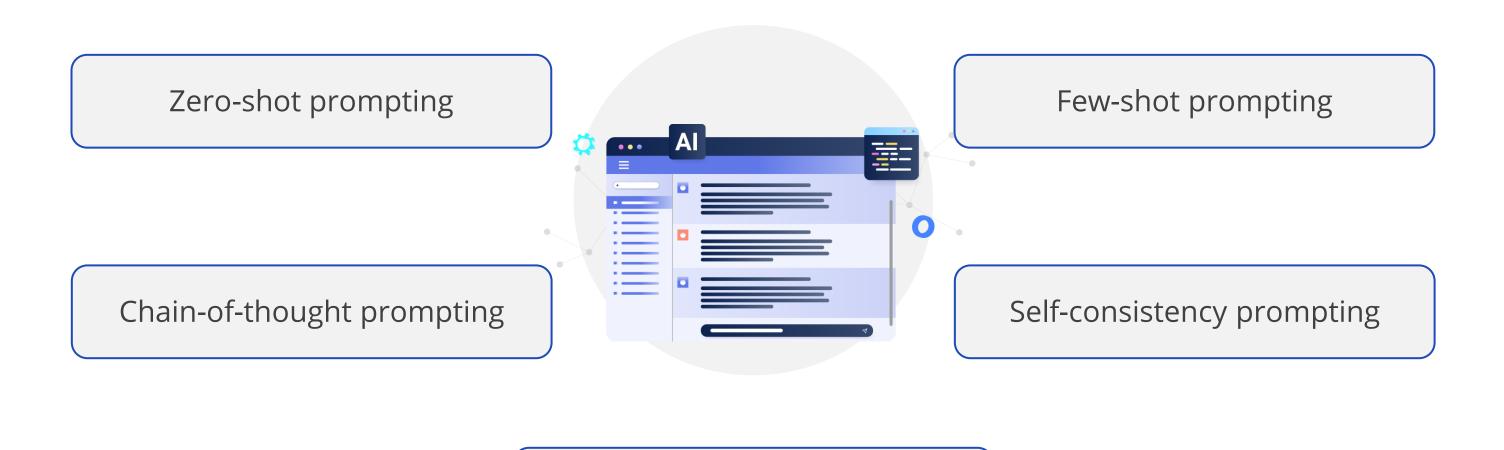
You Write a catchy caption for my Instagram post ChatGPT "Embracing the beauty of simplicity in a chaotic world → #SimpleJoys #FindYourZen"



Advanced Prompting Techniques

Prompting Techniques: From Basics to Advanced

Listed below are some advanced techniques for prompting GenAI models to achieve desired outcomes.



Tree-of-thought prompting

Zero-Shot Prompting

It is a technique that provides a quick response with minimal guidance, helps in setting the context, and gives the desired output without any specific examples.



Its existing knowledge is used for task execution.

Do It on Your Own: Zero-Shot Prompting

Use ChatGPT to try out the prompt below.



Generating social media posts for a brand

Create a social media post for an organic skincare brand, highlighting the natural ingredients and eco-friendly packaging

ChatGPT

🖐 Embrace Nature's Nourishment! 🤟 Our organic skincare line is a radiant ode to the beauty of nature! 🏫 Infused with the purest botanicals and crafted with love, our products harness the power of nature to revitalize your skin! From soothing aloe vera to luxurious coconut oil, each ingredient is carefully selected to pamper your skin and nurture your soul. * And that's not all! We're committed to sustainability every step of the way, from sourcing eco-friendly ingredients to our recyclable packaging. 🔭 Join us on our journey to radiant skin and a healthier planet! 💙 #OrganicSkincare #NaturalBeauty #EcoFriendly #Sustainability #LoveYourSkin

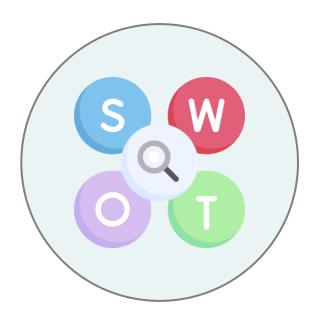






Do It on Your Own: Zero-Shot Prompting

Use ChatGPT to try out the prompt below.



Performing a SWOT analysis

VE

You

Conduct a SWOT analysis for Coca-Cola

(

ChatGPT

Sure, here's a SWOT analysis for Coca-Cola:

Strengths:

- Strong Brand Recognition: Coca-Cola is one of the most recognized brands globally, with a strong reputation built over decades.
- Wide Product Portfolio: Coca-Cola has a diverse range of products including sodas, juices, energy drinks, and water, catering to different consumer preferences.
- Global Presence: The company operates in over 200 countries, giving it a vast international footprint and access to diverse markets.

Do It on Your Own: Zero-Shot Prompting

Use ChatGPT to try out the prompt below.



Building a to-do list



Develop a to-do list for organizing a small business conference

ChatGPT

Certainly! Here's a to-do list for organizing a small business conference:

1. Define Goals and Objectives:

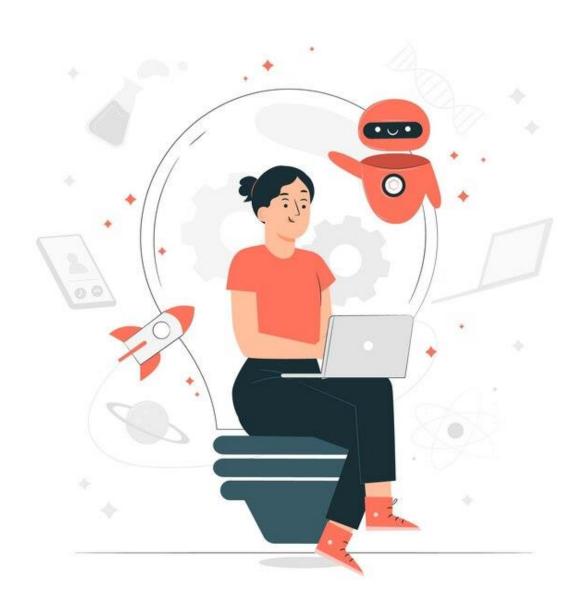
- . Determine the purpose of the conference.
- . Set clear goals and objectives you want to achieve.

2. Budget Planning:

- · Allocate funds for venue, speakers, catering, marketing, etc.
- . Keep track of expenses to stay within budget.

Few-Shot Prompting

It is a style of guiding GenAI models with specific examples to achieve desired responses.



Provides the model with a limited number of specific examples and a definite context

Focuses on providing training sessions to the GenAl bot, enhancing the relevance and precision of the output

Offers a clearer idea of the specific task or style while leveraging the model's existing knowledge

Do It on Your Own: Few-Shot Prompting

Use ChatGPT to try out the prompt below.



You

A Flummoxibbly could be a fantastical creature that inhabits the enchanted forests of a far-off land. An example of a sentence that uses the word Flummoxibbly is:

As I wandered through the mystical forest, I caught a glimpse of a Flummoxibbly darting between the trees, leaving behind a trail of sparkling dust.

The Quizzbizzler is a magical device challenges its users with perplexing riddles. Give an example of a sentence that uses the word Quizzbizzler.





ChatGPT

In the depths of the ancient library, I stumbled upon the Quizzbizzler hidden amidst dusty tomes, its intricate design hinting at the challenging riddles it held within.



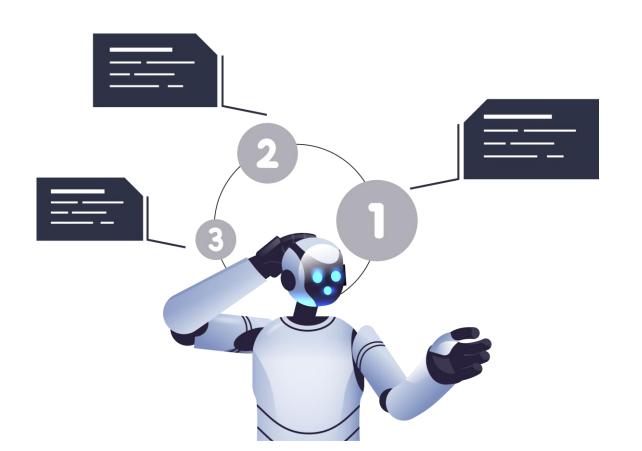






Chain-of-Thought Prompting (CoT)

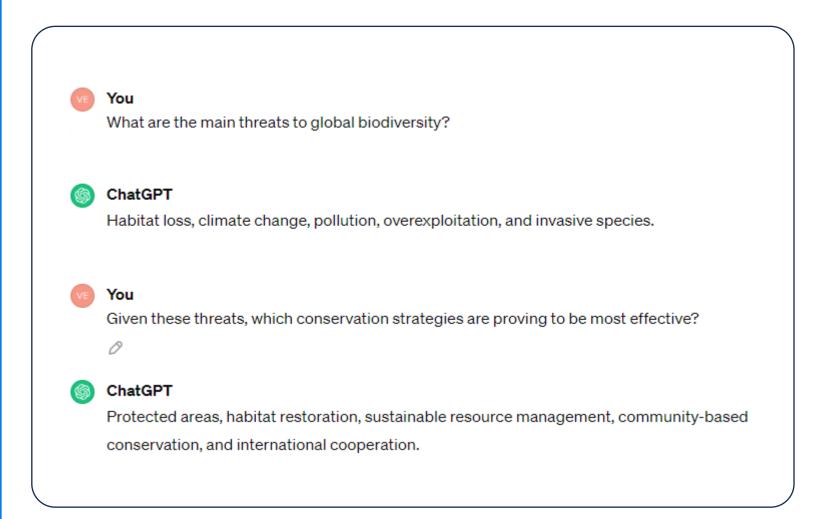
It is a method of providing sequential prompts to the GenAl model to generate sequential outputs, rather than getting the entire output at once.

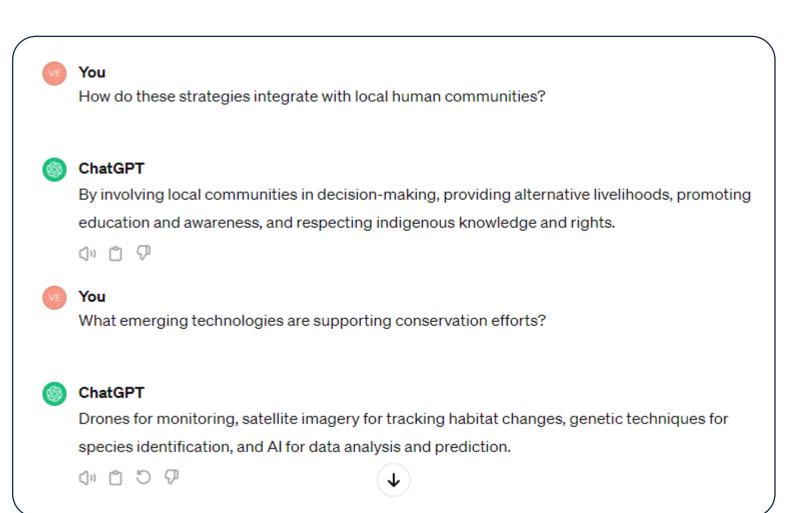


It encourages more thoughtful and detailed responses by avoiding the overload of multiple questions at once.

Do It on Your Own: Chain-of-Thought Prompting

Use ChatGPT to try out the prompt below.





Self-Consistency Prompting

It ensures that a GenAI bot delivers reliable and similar responses when asked the same questions repeatedly.



Example

Repeatedly prompting "What's the capital of France?" should consistently yield the answer *Paris*.

This process makes the AI more reliable, reduces confusion, enhances the user experience, and paves the way for efficient learning.

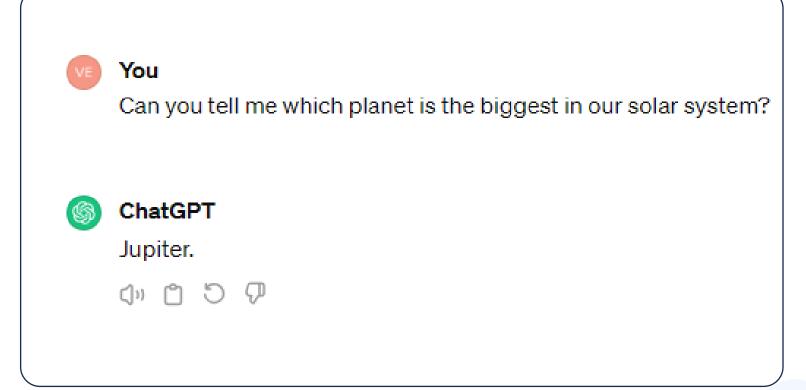
Do It on Your Own: Self-Consistency Prompting

Use ChatGPT to try out the prompt below.



What is the largest planet in our solar system?

ChatGPT Jupiter.



Tree-of-Thought Prompting (ToT)

It is a method used in GenAI models that helps solve complex problems.



It forms a *tree* with various thought paths or steps, resembling branches on a tree.

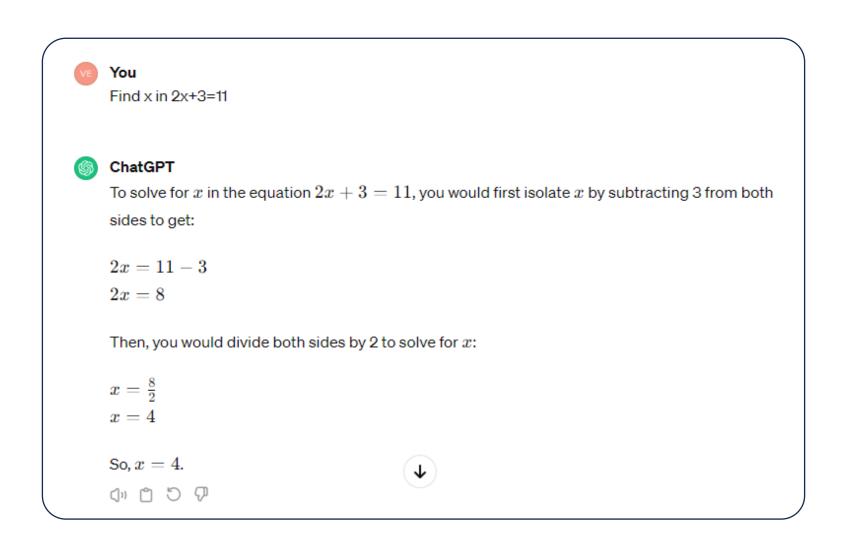
Each branch represents a potential step or idea toward solving a problem.

The AI evaluates these branches to identify the best solution.

It is particularly useful for tasks that require strategic thinking or involve multiple steps.

ToT Prompting: Example

Use ChatGPT to try out the prompt below.



Note: The Al uses ToT to break down the problem into smaller steps, evaluate each step, and arrive at the correct solution.

Best Practices for Creating Prompts



Dos

- Begin the requests with clear and specific framing.
- **Example:** List 5 Paris tourist attractions with 2-3 sentence descriptions.
- Offer context by providing background details.
- **Example:** I'm writing an article on sustainable farming practices, so I need information on the benefits of crop rotation and how it improves soil health.



Dos

- Ask follow-up questions if the response lacks completeness.
- **Example:** Your explanation of photosynthesis was helpful, but could you also briefly explain how plants use the energy produced during that process?
- Provide guidelines early on regarding tone, length, and format.
- **Example:** For this blog post, I need around 500 words written in a conversational, lighthearted tone with examples or analogies to explain complex concepts. Please use headings to break it into 3-4 sections.



Don'ts

- Don't provide unnecessary details.
- **Example:** Just give me the ingredients list for a basic spaghetti recipe, nothing more. I don't need details on preparation steps, cooking times, or serving suggestions.
- Don't give it too many tasks to perform at the same time.
- **Example:** First, draft an outline for the introduction section of the report. Then, research and write the background information. After that, analyze the data and summarize the key findings. Let's work on one component at a time.



Don'ts

- Don't provide any contradictory statements while prompting.
- **Example:** Describe a utopian society where everyone lives in harmony, but also include details about conflicts and divisions between different groups.

Note: Don't rely on the information given by the tool, as the GenAI model might generate an inaccurate response.

Demo: Refining Prompts



Duration: 15 minutes

In this demo, we will continue iterating on the prompt, demonstrating techniques like:

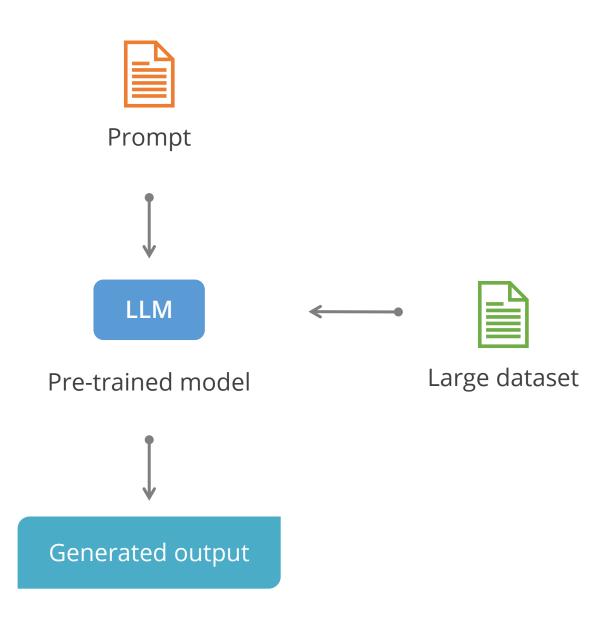
- Asking clarifying questions about elements that are still vague or unclear
- Providing examples to illustrate the desired type of response
- Adding or removing specificity as needed
- Setting guidelines for length, tone, structure, and more

DEMONSTRATION

Introduction to Fine-Tuning

Pre-Training

Pre-training is a process of training a GenAl model on a large corpus dataset.



Fine-Tuning

Fine-tuning involves taking pre-trained models and training them more on smaller, specific datasets to enhance performance in a particular task or domain.



Why is network traffic encrypted?



Because it protects data during transmission.



Why is network traffic encrypted?



Network traffic is encrypted using complex algorithms to prevent unauthorized interception. This safeguards sensitive data from potential attackers, ensuring confidentiality and integrity.

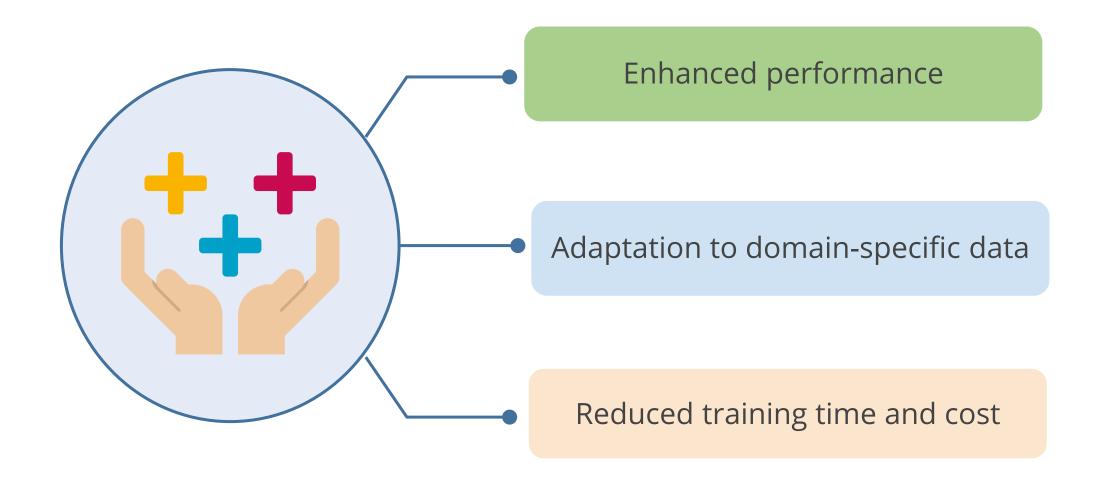
Common encryption protocols include SSL/TLS, IPsec, and WPA2 for web, VPN, and wireless traffic, respectively.

Pre-trained model

Fine-tuned Model

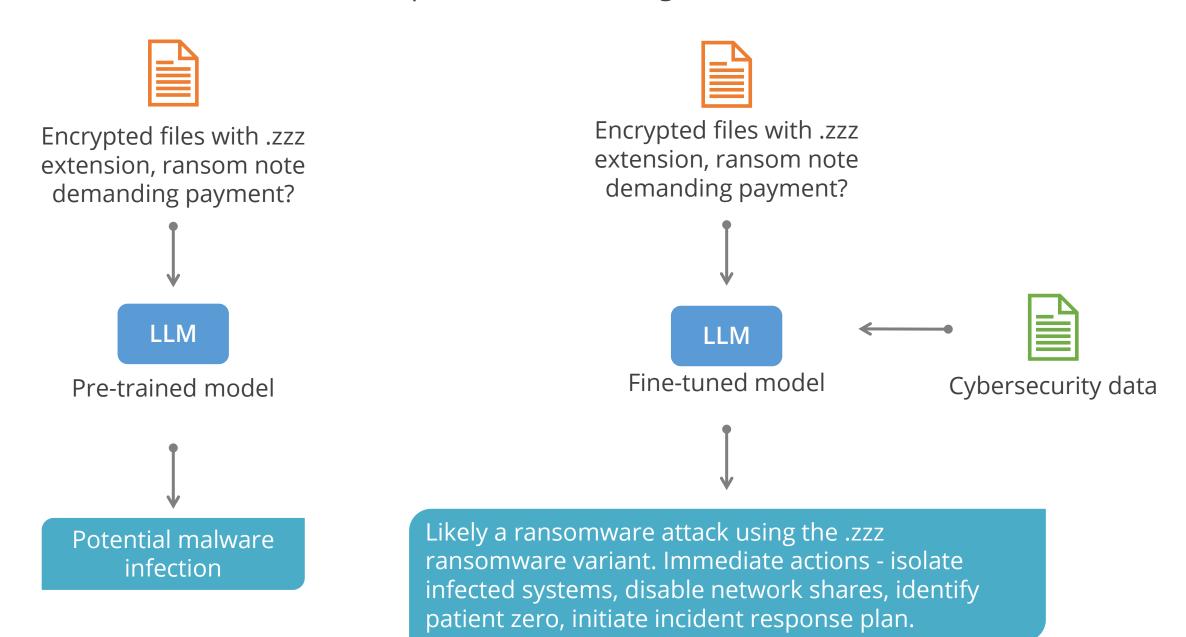
Benefits of Fine-Tuning

Fine-tuning ChatGPT offers several benefits, including:



Improved Performance and Domain Adaptation

Fine-tuning GPTs on task-specific data helps them understand the subtle details and complexities of the target domain.



Considerations for Fine-Tuning Language Models

Fine-tuning isn't always the best choice. Consider this scenario: You've fine-tuned a base model using your Slack messages dataset.

Write an essay analyzing the impacts of climate change.





Sure, I can work on that essay tomorrow and have it ready for you.

No, I need you to write it now. I'm on a tight deadline.





Understood. Let me begin drafting that essay analyzing climate change impacts right away.

Challenges of Fine-Tuning

Overfitting

Undesirable fine-tuning behavior where an LLM excels on training data but struggles with new tasks or domain-specific data.

Catastrophic forgetting

Fine-tuning may cause the LLM to forget all pre-training knowledge, leading to catastrophic forgetting.

Expensive

Fine-tuning LLMs for specific tasks demands substantial data, which is difficult to collect and expensive. Complex computing infrastructure adds to the cost.

Bias

Pre-trained LLMs may carry biases, which can worsen during fine-tuning. Bias mitigation measures should be in place before fine-tuning.

Note: The process of fine-tuning is time-consuming and expensive to implement

Quick Check



Which of the following is a challenge that may occur while fine-tuning GenAl models?

- A. Enhanced essay writing skills
- B. Preservation of general capabilities
- C. Risk of forgetting formal composition skills
- D. No impact on adaptability to new use cases

GUIDED PRACTICE

Guided Practice



Overview Duration: 30 minutes

In this activity, you will craft an effective prompt by applying advanced prompting techniques. This activity will help you gain hands-on experience with writing prompts.

Key Takeaways

- Prompt engineering involves creating clear and effective instructions to guide large language models in generating desired outputs.
- Advanced techniques in prompt engineering such as zero-shot and few-shot methods assist in creating effective prompts.
- Fine-tuning customizes large language models for specific tasks or fields using specialized data.



Additional Resources

• How to Become a Prompt Engineer: Your Path to Al Mastery



Q&A

