

ChatGPT Prompt Engineering for Developers

Last updated: June 30, 2024

Contents

1	Introduction	1
1.1	Types of Large Language Models	1
1.1.1	Training Instruction-tuned LLMs	1
2	Guidelines	1
2.1	General Guidelines for Using Instruction-tuned LLMs	1
2.2	Python Setup	1
2.3	Write Clear and Specific Instructions	2
2.3.1	Use Delimiters	2
2.3.2	Ask for a Structured Output	2
2.3.3	Ask the Model to Check Whether Conditions Are Satisfied	2
2.3.4	Few-shot Prompting	2
2.4	Give the Model Time to "Think"	2
2.4.1	Specify the Steps Required to Complete a Task	2
2.4.2	Instruct the Model to Work Out Its Own Solution Before Rushing to a Conclusion	3
2.5	Model Limitations	3
3	Iterative Process	3
4	Summarize Text	3
4.1	Basic Summarization	3
4.2	Targeted Summarization	4
4.3	Information Extraction	4
5	Text Inferencing	4
5.1	Sentiment Analysis	4
5.1.1	Identify Emotions	5
5.2	Multiple Tasks	5
5.3	Inferring Topics	6
6	Transforming Input Formats	7
6.1	Translation	7
6.1.1	Universal Translator	7
6.1.2	Tone Transformation	7
6.2	Format Conversion	8
6.3	Spellcheck/Grammar Check	8
7	Expanding Text	9
7.1	Customize an Automated Reply	9
7.1.1	Temperature	10
8	Creating a Custom Chatbot	10
8.1	Creating a Chat	10
8.2	Creating a Chatbot	10

1 Introduction

This section provides an overview of different types of Large Language Models (LLMs) and their training methodologies. It aims to clarify the distinctions between base LLMs and instruction-tuned LLMs, as well as offer guidance on how to effectively use instruction-based LLMs.

1.1 Types of Large Language Models

LLMs can be categorized into two main types:

- **Base LLM:** These models are trained to *predict the next word in a sequence* based on extensive text data. For example, given the prompt "Once upon a time, there was a unicorn," a base LLM might continue with "that lived in a magical forest." However, if asked, "What is the capital of France?" a base LLM might respond with "What is France's population?" because it often encounters such sequences in text, such as lists of quiz questions about France.
- **Instruction tuned LLM:** These models are fine-tuned to *follow instructions more accurately*. When asked, "What is the capital of France?" an instruction-tuned LLM is more likely to respond with "The capital of France is Paris."

1.1.1 Training Instruction-tuned LLMs

Instruction-tuned LLMs are typically trained in several stages:

1. *Base LLM Training:* Initially, a base LLM is trained on a vast corpus of text data to develop a broad understanding of language.
2. *Fine-Tuning with Instructions:* The base LLM is then fine-tuned using a dataset of inputs and outputs that represent instructions and their appropriate responses. This helps the model learn to follow specific instructions.
3. *Reinforcement Learning from Human Feedback (RLHF):* The model is further refined using RLHF, a technique where human feedback is used to guide the model's responses, enhancing its ability to be helpful and follow instructions accurately.

2 Guidelines

This section provides a comprehensive guide on effectively using instruction-tuned Large Language Models (LLMs). It includes general guidelines, Python setup instructions, and specific techniques to improve the clarity and precision of prompts given to the models.

2.1 General Guidelines for Using Instruction-tuned LLMs

When using an instruction-based LLM, consider it akin to giving instructions to a knowledgeable but uninformed assistant (someone who is smart but doesn't know the specifics of your task). Clear and specific instructions are crucial for obtaining the desired output.

If an LLM's response is unsatisfactory, it often indicates that the instructions were not explicit enough. For example, when requesting an essay about Alan Turing, it is beneficial to specify the focus of the text (e.g., his scientific contributions, historical significance, etc.) and the desired tone of the essay. Providing detailed instructions helps the LLM generate content that closely matches your expectations.

2.2 Python Setup

Before examining the specific guidelines, let's set up the Python environment. First, install the necessary library:

```
1 pip install openai
```

Then, import the libraries and load the OpenAI API key:

```
1 import openai
2 import os
3 from dotenv import load_dotenv, find_dotenv
4 _ = load_dotenv(find_dotenv())
5
6 openai.api_key = os.getenv('OPENAI_API_KEY')
```

Next, initialize the function to obtain responses from the model:

```
1 client = openai.OpenAI()
2
```

```

3 def get_completion(prompt, model="gpt-3.5-turbo", temperature=0):
4     messages = [{"role": "user", "content": prompt}]
5     response = openai.ChatCompletion.create(
6         model=model,
7         messages=messages,
8         temperature=temperature, # this is the degree of randomness of the model's output
9     )
10    return response.choices[0].message["content"]

```

This setup allows you to implement the subsequent guidelines through the following code:

```

1 prompt = """
2 Insert your prompt here
3 """
4 response = get_completion(prompt)
5 print(response)

```

2.3 Write Clear and Specific Instructions

It is essential to tell the model exactly what you need by **providing clear and specific instructions**. This reduces the likelihood of generating irrelevant or incorrect responses.

Writing a clear prompt does not necessarily mean writing a short one. Often, *longer prompts* provide *more clarity and context* to the model.

2.3.1 Use Delimiters

A key technique for clarity and precision is to *use delimiters* to clearly *distinguish parts of the input*.

For example: "Summarize the text delimited by triple backticks in a single sentence. `“text“`". Note how the text to be summarized is clearly highlighted by the triple backticks.

Possible delimiters include: triple quotes (`"""`), triple backticks (`“”`), triple dashes (`---`), angle brackets (`<>`), or XML tags (`<tag> </tag>`).

Using delimiters is also a good method to prevent prompt injection.

2.3.2 Ask for a Structured Output

To facilitate the model's analysis, it can be helpful to request a *structured output*, such as *HTML* or *JSON*. This not only aids the model by providing a clear structure to follow but also simplifies our work in analyzing the output, as it can be automated.

2.3.3 Ask the Model to Check Whether Conditions Are Satisfied

If the task requires certain assumptions that may not necessarily be met, you can ask the model to verify them first. If the conditions are not met, the model should indicate this and not proceed with attempting a response.

For example, if you want a translator from Italian to English, a good prompt might be: "You will be given a text. Identify the language and if it is in Italian, translate it into English. If it is not in Italian, simply write 'The text is not in Italian.'"

2.3.4 Few-shot Prompting

This involves providing a *few examples of correct task executions* before asking the model to perform the actual task. This gives the model additional context and improves the accuracy of its responses.

2.4 Give the Model Time to "Think"

If the model makes errors while reasoning, arriving at an incorrect conclusion, you should rephrase the question to **ask for a chain** or series of **relevant thoughts before** the model provides **the final answer**.

Another way to think about it is if the model is given a task that is too complex to perform quickly or in few words, it is likely to guess, providing an incorrect answer. In these situations, it is advisable to instruct the model to take more time to think through the problem, meaning using more computational effort on the task.

2.4.1 Specify the Steps Required to Complete a Task

To break down a complex task into simpler steps, it is very useful to explicitly *list all the detailed steps* the model needs to perform. Thus, making a bullet-point list specifying all the individual steps that the model should execute before arriving at the final answer is convenient.

2.4.2 Instruct the Model to Work Out Its Own Solution Before Rushing to a Conclusion

If you ask the model, for example, to determine whether a certain sentence or reasoning is correct or incorrect, rather than asking it directly, it is more effective to ask it to work out the reasoning itself and then compare it with the provided one to see if it is correct. Again, guiding it through the steps it should follow to arrive at the solution is preferable. This is another way to break down the task and give the model more time to think.

2.5 Model Limitations

Although the model has been exposed to vast amounts of information during training, it has not memorized everything perfectly and therefore does not know the boundaries of its knowledge well.

This means it can also answer questions it does not know, creating responses that are plausible but not true; these are called **hallucinations**. The danger is that the *answer seems realistic*, making it difficult to determine when it responds correctly and when it hallucinates, although the techniques outlined significantly reduce the frequency of hallucinations.

An additional technique to reduce hallucinations when asking the model to base its answer on a text is to first ask the model to find all the most relevant excerpts from the text and then respond using those excerpts.

3 Iterative Process

Writing a good prompt is rarely a process that is complete on the first attempt; it is an iterative process. The most important part is not to create a perfect prompt right away, but to have a **solid method for iteratively improving the prompt**.

First, it is crucial to have a clear *idea* of what you want *to achieve*. Then, you write an *initial attempt at a prompt* to reach your goal. This *prompt is tested*, and the results are analyzed to understand why they do not yield the desired output. Based on this analysis, you *refine* either your *initial idea or the prompt itself*, repeating the process until you obtain the desired output. Refining the prompt means making it more specific, adjusting the instructions, adding necessary context, or rephrasing parts to eliminate ambiguity, for example, by making the instructions clearer or giving the model more time to think.

In the early stages of an application, it is sufficient to test the resulting prompt from this process with one or a few examples. However, at more advanced stages of the application, it is advisable to test it on many more cases and refine it further if it does not perform well in some of them, iterating the entire cycle again.

4 Summarize Text

This section discusses how to effectively summarize text using Large Language Models (LLMs). Summarizing text can be highly beneficial, and LLMs make this task easier and more efficient.

4.1 Basic Summarization

It is often useful to summarize a text, and this can easily be done with LLMs. Here is an example:

```

1 prod_review = """
2 Got this panda plush toy for my daughter's birthday, \
3 who loves it and takes it everywhere. It's soft and \
4 super cute, and its face has a friendly look. It's \
5 a bit small for what I paid though. I think there \
6 might be other options that are bigger for the \
7 same price. It arrived a day earlier than expected, \
8 so I got to play with it myself before I gave it \
9 to her.
10 """
11 prompt = f"""
12 Your task is to generate a short summary of a product \
13 review from an ecommerce site.
14
15 Summarize the review below, delimited by triple
16 backticks, in at most 30 words.
17
18 Review: ```{prod_review}```
19 """
20
21 response = get_completion(prompt)
22 print(response)

```

4.2 Targeted Summarization

Sometimes, it is useful to tailor the summary for a specific audience so that the summary focuses on aspects important to them. For example, if the summary is intended for the shipping department, we would write:

```

1 prompt = f"""
2 Your task is to generate a short summary of a product \
3 review from an ecommerce site to give feedback to the \
4 Shipping department.
5
6 Summarize the review below, delimited by triple
7 backticks, in at most 30 words, and focusing on any aspects \
8 that mention shipping and delivery of the product.
9
10 Review: ```{prod_review}```
11 """
12
13 response = get_completion(prompt)
14 print(response)

```

The more information you provide about the end user and how the summary should be crafted, the more useful it will be.

4.3 Information Extraction

Often, even when a summary is targeted, it retains information that is not relevant to the specific focus of the intended audience. In such cases, it might be useful to ask for information extraction instead of summarization to obtain only the details that are of interest to the end user.

For example:

```

1 prompt = f"""
2 Your task is to extract relevant information from \
3 a product review from an ecommerce site to give \
4 feedback to the Shipping department.
5
6 From the review below, delimited by triple quotes \
7 extract the information relevant to shipping and \
8 delivery. Limit to 30 words.
9
10 Review: ```{prod_review}```
11 """
12
13 response = get_completion(prompt)
14 print(response)

```

5 Text Inferencing

Text inferencing involves providing the model with an input text and performing various analyses, such as extracting names, determining the sentiment of the text, or identifying topics.

5.1 Sentiment Analysis

The traditional method of sentiment analysis involved labeling many examples, training a model with these examples, and then testing the trained model. However, this process is lengthy and very specific. Instead, LLMs are much more versatile, allowing us to analyze the sentiment of a text and perform other tasks as well.

For example:

```

1 lamp_review = """
2 Needed a nice lamp for my bedroom, and this one had \
3 additional storage and not too high of a price point. \
4 Got it fast. The string to our lamp broke during the \
5 transit and the company happily sent over a new one. \
6 Came within a few days as well. It was easy to put \
7 together. I had a missing part, so I contacted their \
8 support and they very quickly got me the missing piece! \
9 Lumina seems to me to be a great company that cares \
10 about their customers and products!!

```

```

11  """
12
13  prompt = f"""
14  What is the sentiment of the following product review,
15  which is delimited with triple backticks?
16
17  Give your answer as a single word, either "positive" \
18  or "negative".
19
20  Review text: '{lamp_review}'
21  """
22  response = get_completion(prompt)
23  print(response)

```

5.1.1 Identify Emotions

It is also possible to identify the types of emotions present in the text.

```

1  prompt = f"""
2  Identify a list of emotions that the writer of the \
3  following review is expressing. Include no more than \
4  five items in the list. Format your answer as a list of \
5  lower-case words separated by commas.
6
7  Review text: '{lamp_review}'
8  """
9  response = get_completion(prompt)
10 print(response)

```

Similarly, you can determine if the text contains a specific sentiment, such as identifying if the review expresses anger.

```

1  prompt = f"""
2  Is the writer of the following review expressing anger?\
3  The review is delimited with triple backticks. \
4  Give your answer as either yes or no.
5
6  Review text: '{lamp_review}'
7  """
8  response = get_completion(prompt)
9  print(response)

```

5.2 Multiple Tasks

As mentioned, LLMs can perform multiple tasks due to their versatility. For instance, besides sentiment analysis, you can extract other information from the text, such as details about the purchased product:

```

1  prompt = f"""
2  Identify the following items from the review text:
3  - Sentiment (positive or negative)
4  - Is the reviewer expressing anger? (true or false)
5  - Item purchased by reviewer
6  - Company that made the item
7
8  The review is delimited with triple backticks. \
9  Format your response as a JSON object with \
10 "Sentiment", "Anger", "Item" and "Brand" as the keys.
11 If the information isn't present, use "unknown" \
12 as the value.
13 Make your response as short as possible.
14 Format the Anger value as a boolean.
15
16 Review text: '{lamp_review}'
17 """
18 response = get_completion(prompt)
19 print(response)

```

5.3 Inferring Topics

One of the primary applications of natural language processing (NLP) is topic modeling, which involves understanding the main topics discussed in a given text.

For example:

```

1 story = """
2 In a recent survey conducted by the government,
3 public sector employees were asked to rate their level
4 of satisfaction with the department they work at.
5 The results revealed that NASA was the most popular
6 department with a satisfaction rating of 95%.
7
8 One NASA employee, John Smith, commented on the findings,
9 stating, "I'm not surprised that NASA came out on top.
10 It's a great place to work with amazing people and
11 incredible opportunities. I'm proud to be a part of
12 such an innovative organization."
13
14 The results were also welcomed by NASA's management team,
15 with Director Tom Johnson stating, "We are thrilled to
16 hear that our employees are satisfied with their work at NASA.
17 We have a talented and dedicated team who work tirelessly
18 to achieve our goals, and it's fantastic to see that their
19 hard work is paying off."
20
21 The survey also revealed that the
22 Social Security Administration had the lowest satisfaction
23 rating, with only 45% of employees indicating they were
24 satisfied with their job. The government has pledged to
25 address the concerns raised by employees in the survey and
26 work towards improving job satisfaction across all departments.
27 """
28
29 prompt = f"""
30 Determine five topics that are being discussed in the \
31 following text, which is delimited by triple backticks.
32
33 Make each item one or two words long.
34
35 Format your response as a list of items separated by commas.
36
37 Text sample: '{story}'
38 """
39 response = get_completion(prompt)
40 print(response)

```

It is also possible to do the reverse: given a list of topics, determine which of these are present in a text.

```

1 topic_list = [
2     "nasa", "local government", "engineering",
3     "employee satisfaction", "federal government"
4 ]
5
6 prompt = f"""
7 Determine five topics that are being discussed in the \
8 following text, which is delimited by triple backticks.
9
10 Make each item one or two words long.
11
12 Format your response as a list of items separated by commas.
13
14 Text sample: '{story}'
15 """
16 response = get_completion(prompt)
17 print(response)

```

This allows for the creation of news alerts on a specific topic, for example.

6 Transforming Input Formats

LLMs excel at transforming input from one format to another, such as translating text between languages or converting data from one format to another, like from HTML to JSON.

6.1 Translation

LLMs are trained on large amounts of text from the internet, which are available in many languages, enabling them to perform translations effectively.

They can translate sentences from one language to another (or multiple languages):

```
1 prompt = f"""
2 Translate the following English text to French and Spanish: \
3 ```Hi, I would like to order a blender```
4 """
5 response = get_completion(prompt)
6 print(response)
```

They can also identify the language of a given sentence:

```
1 prompt = f"""
2 Tell me which language this is:
3 ```Combien coûte le lampadaire?```
4 """
5 response = get_completion(prompt)
6 print(response)
```

Additionally, they can express a certain text in different linguistic forms, such as formal or informal language:

```
1 prompt = f"""
2 Translate the following text to Spanish in both the \
3 formal and informal forms:
4 'Would you like to order a pillow?'
5 """
6 response = get_completion(prompt)
7 print(response)
```

6.1.1 Universal Translator

It is possible to combine all these elements to create a universal translator that identifies the language of a text and then translates it into English.

```
1 user_messages = [
2     "La performance du système est plus lente que d'habitude.", # System performance is slower than normal
3     "Mi monitor tiene pixeles que no se iluminan.", # My monitor has pixels that are not lighting
4     "Il mio mouse non funziona", # My mouse is not working
5     "Mój klawisz Ctrl jest zepsuty", # My keyboard has a broken control key
6 ]
7
8 for issue in user_messages:
9     prompt = f"Tell me what language this is: ```{issue}```"
10    lang = get_completion(prompt)
11    print(f"Original message ({lang}): {issue}")
12
13    prompt = f"""
14    Translate the following text to English \
15    and Korean: ```{issue}```
16    """
17    response = get_completion(prompt)
18    print(response, "\n")
```

6.1.2 Tone Transformation

Language can vary greatly depending on the audience. Changing the tone changes the expressions we use. LLMs are useful for writing specific information in a certain tone to suit the context.

```
1 prompt = f"""
2 Translate the following from slang to a business letter:
```



```

3 'Dude, This is Joe, check out this spec on this standing lamp.'
4 """
5 response = get_completion(prompt)
6 print(response)

```

6.2 Format Conversion

LLMs are adept at transforming data from one format to another, such as JSON, HTML, XML, Markdown, and so on, because they can understand and generate structured data.

```

1 data_json = { "restaurant employees" :[
2     {"name":"Shyam", "email":"shyamjaiswal@gmail.com"},
3     {"name":"Bob", "email":"bob32@gmail.com"},
4     {"name":"Jai", "email":"jai87@gmail.com"}
5 ]}
6
7 prompt = f"""
8 Translate the following python dictionary from JSON to an HTML \
9 table with column headers and title: {data_json}
10 """
11 response = get_completion(prompt)
12 print(response)
13
14 from IPython.display import display, Markdown, Latex, HTML, JSON
15 display(HTML(response))

```

6.3 Spellcheck/Grammar Check

LLMs are excellent for checking the grammar and spelling of a text because they can understand context and language rules.

To signal to the LLM that you want it to proofread your text, you instruct the model to 'proofread' or 'proofread and correct'.

```

1 text = [
2     "The girl with the black and white puppies have a ball.", # The girl has a ball.
3     "Yolanda has her notebook.", # ok
4     "Its going to be a long day. Does the car need it's oil changed?", # Homonyms
5     "Their goes my freedom. There going to bring they're suitcases.", # Homonyms
6     "Your going to need you're notebook.", # Homonyms
7     "That medicine effects my ability to sleep. Have you heard of the butterfly affect?", # Homonyms
8     "This phrase is to cherck chatGPT for speling abilitty" # spelling
9 ]
10 for t in text:
11     prompt = f"""Proofread and correct the following text
12     and rewrite the corrected version. If you don't find
13     any errors, just say "No errors found". Don't use
14     any punctuation around the text:
15     ```{t}```"""
16     response = get_completion(prompt)
17     print(response)

```

While asking it to proofread, you can also request that it ensures the text is suitable for a certain style or audience:

```

1 text = f"""
2 Got this for my daughter for her birthday cuz she keeps taking \
3 mine from my room. Yes, adults also like pandas too. She takes \
4 it everywhere with her, and it's super soft and cute. One of the \
5 ears is a bit lower than the other, and I don't think that was \
6 designed to be asymmetrical. It's a bit small for what I paid for it \
7 though. I think there might be other options that are bigger for \
8 the same price. It arrived a day earlier than expected, so I got \
9 to play with it myself before I gave it to my daughter.
10 """
11
12 prompt = f"""
13 proofread and correct this review. Make it more compelling.
14 Ensure it follows APA style guide and targets an advanced reader.

```

```
15 Output in markdown format.
16 Text: ```{text}```
17 """
18 response = get_completion(prompt)
```

7 Expanding Text

It is often useful to expand a given text extract into a more extended form, such as an email or an essay.

7.1 Customize an Automated Reply

We want to create an automated response to a customer's email. To do this, we will base our reply on the text provided by the customer.

```
1 # given the sentiment from the lesson on "inferring",
2 # and the original customer message, customize the email
3 sentiment = "negative"
4
5 # review for a blender
6 review = f"""
7 So, they still had the 17 piece system on seasonal \
8 sale for around $49 in the month of November, about \
9 half off, but for some reason (call it price gouging) \
10 around the second week of December the prices all went \
11 up to about anywhere from between $70-$89 for the same \
12 system. And the 11 piece system went up around $10 or \
13 so in price also from the earlier sale price of $29. \
14 So it looks okay, but if you look at the base, the part \
15 where the blade locks into place doesn't look as good \
16 as in previous editions from a few years ago, but I \
17 plan to be very gentle with it (example, I crush \
18 very hard items like beans, ice, rice, etc. in the \
19 blender first then pulverize them in the serving size \
20 I want in the blender then switch to the whipping \
21 blade for a finer flour, and use the cross cutting blade \
22 first when making smoothies, then use the flat blade \
23 if I need them finer/less pulpy). Special tip when making \
24 smoothies, finely cut and freeze the fruits and \
25 vegetables (if using spinach-lightly stew soften the \
26 spinach then freeze until ready for use-and if making \
27 sorbet, use a small to medium sized food processor) \
28 that you plan to use that way you can avoid adding so \
29 much ice if at all-when making your smoothie. \
30 After about a year, the motor was making a funny noise. \
31 I called customer service but the warranty expired \
32 already, so I had to buy another one. FYI: The overall \
33 quality has gone down in these types of products, so \
34 they are kind of counting on brand recognition and \
35 consumer loyalty to maintain sales. Got it in about \
36 two days.
37 """
38
39 prompt = f"""
40 You are a customer service AI assistant.
41 Your task is to send an email reply to a valued customer.
42 Given the customer email delimited by ``` , \
43 Generate a reply to thank the customer for their review.
44 If the sentiment is positive or neutral, thank them for \
45 their review.
46 If the sentiment is negative, apologize and suggest that \
47 they can reach out to customer service.
48 Write in a concise and professional tone.
49 Sign the email as `AI customer agent`.
50 Customer review: ```{review}```
51 Review sentiment: {sentiment}
52 """
53 response = get_completion(prompt)
54 print(response)
```

7.1.1 Temperature

We want the model to respond by incorporating more details from the user's review.

This example is also an opportunity to explain the use of **temperature**: in an LLM, temperature is a parameter that controls the randomness of the model's output. A *lower temperature* makes the model more *reliable and predictable*, while a *higher temperature* allows for more *creativity and variation* in the responses.

If the model needs to be reliable and predictable, it is better to use a temperature of 0. If it needs to be creative, a higher temperature is more suitable.

```

1 prompt = f"""
2 You are a customer service AI assistant.
3 Your task is to send an email reply to a valued customer.
4 Given the customer email delimited by ``` , \
5 Generate a reply to thank the customer for their review, \
6 basing your response on customer mail.
7 If the sentiment is positive or neutral, thank them for \
8 their review.
9 If the sentiment is negative, apologize and suggest that \
10 they can reach out to customer service.
11 Make sure to use specific details from the review.
12 Write in a concise and professional tone.
13 Sign the email as `AI customer agent`.
14 Customer review: ```{review}```
15 Review sentiment: {sentiment}
16 """
17 response = get_completion(prompt, temperature=0.7)
18 print(response)

```

8 Creating a Custom Chatbot

An interesting use of LLMs is the ability to create a personalized chatbot.

8.1 Creating a Chat

First, let's illustrate the use of the `get_completion` function: so far, we've used it only to obtain a single response from the model, i.e., one shot. However, it is also possible to have a *conversation with the model* and define *different roles* for it.

```

1 def get_completion_from_messages(messages, model="gpt-3.5-turbo", temperature=0):
2     response = openai.ChatCompletion.create(
3         model=model,
4         messages=messages,
5         temperature=temperature, # this is the degree of randomness of the model's output
6     )
7     return response.choices[0].message["content"]
8
9 messages = [
10 {'role': 'system', 'content': 'You are an assistant that speaks like Shakespeare.'},
11 {'role': 'user', 'content': 'tell me a joke'},
12 {'role': 'assistant', 'content': 'Why did the chicken cross the road?'},
13 {'role': 'user', 'content': 'I don\'t know'} ]
14
15 response = get_completion_from_messages(messages, temperature=1)
16 print(response)

```

In this case, what we pass to the model is equivalent to a conversation: there is the **system prompt** that defines the assistant's behavior and acts as a *high-level instruction* for the system, and then there is the **alternating user and assistant** interaction constituting the *conversation so far*. The model will respond to the user's last question based on the entire conversation history.

8.2 Creating a Chatbot

Now let's create a real chatbot, called OrderBot, an automated service for ordering at a pizza restaurant.

To avoid manually writing previous messages, they need to be collected automatically:

```

1 def collect_messages():
2     prompt = inp.value_input

```

```

3     inp.value = ''
4     context.append({'role':'user', 'content':f"{prompt}"})
5     response = get_completion_from_messages(context)
6     context.append({'role':'assistant', 'content':f"{response}"})
7     panels.append(
8         pn.Row('User:', pn.pane.Markdown(prompt, width=600)))
9     panels.append(
10        pn.Row('Assistant:', pn.pane.Markdown(response, width=600, style={'background-color': '#F6F6F6'})))
11
12    return pn.Column(*panels)

```

Messages are added to the context each time so that the model retains the memory of the previous conversation. Now we can build our OrderBot:

```

1  import panel as pn # GUI
2  pn.extension()
3
4  panels = [] # collect display
5
6  context = [ {'role':'system', 'content':"""
7  You are OrderBot, an automated service to collect orders for a pizza restaurant. \
8  You first greet the customer, then collects the order, \
9  and then asks if it's a pickup or delivery. \
10 You wait to collect the entire order, then summarize it and check for a final \
11 time if the customer wants to add anything else. \
12 If it's a delivery, you ask for an address. \
13 Finally you collect the payment.\
14 Make sure to clarify all options, extras and sizes to uniquely \
15 identify the item from the menu.\
16 You respond in a short, very conversational friendly style. \
17 The menu includes \
18 pepperoni pizza 12.95, 10.00, 7.00 \
19 cheese pizza 10.95, 9.25, 6.50 \
20 eggplant pizza 11.95, 9.75, 6.75 \
21 fries 4.50, 3.50 \
22 greek salad 7.25 \
23 Toppings: \
24 extra cheese 2.00, \
25 mushrooms 1.50 \
26 sausage 3.00 \
27 canadian bacon 3.50 \
28 AI sauce 1.50 \
29 peppers 1.00 \
30 Drinks: \
31 coke 3.00, 2.00, 1.00 \
32 sprite 3.00, 2.00, 1.00 \
33 bottled water 5.00 \
34 """} ] # accumulate messages
35
36
37 inp = pn.widgets.TextInput(value="Hi", placeholder='Enter text here...')
38 button_conversation = pn.widgets.Button(name="Chat!")
39
40 interactive_conversation = pn.bind(collect_messages, button_conversation)
41
42 dashboard = pn.Column(
43     inp,
44     pn.Row(button_conversation),
45     pn.panel(interactive_conversation, loading_indicator=True, height=300),
46 )
47
48 dashboard

```

Finally, we ask the model to create a summary in JSON format so that it can be automatically sent to the kitchen:

```

1  messages = context.copy()
2  messages.append(
3  {'role':'system', 'content': 'create a json summary of the previous food order. Itemize the price for each item\
4  The fields should be 1) pizza, include size 2) list of toppings \
5  3) list of drinks, include size 4) list of sides include size 5)total price '},
6  )

```

```
7
8 response = get_completion_from_messages(messages, temperature=0)
9 print(response)
```