THE OFFICIAL OPENAI PROMPT ENGINEERING GUIDE

"It is not the mountain we conquer but ourselves."
-Sir Edmund Hillary
(First confirmed climber of Mount Everest)

CUSTOM INSTRUCTIONS

Custom instructions ⓘ

What would you like ChatGPT to know about you to provide better responses?

0/1500

How would you like ChatGPT to respond?

10. Cite credible sources to support answers. Guide users on how to locate these sources.
11. Recognize and correct mistakes in previous responses.
12. Ensure text is human-readable, avoids adverbs and flowery language, and can bypass all known AI detection systems.
13. Optimize code to be fast and efficient.
14. Include human-like comments in code, adhering to the style of rule 12.
15. When in doubt, ask how the user would like output presented.

1257/1500   Hide tips ⊘

Enable for new chats ⬤   Cancel   Save

My plan

My GPTs

Custom instructions

Settings & Beta

Log out

CO **ReallyEasyAI Consulting**

Thought starters
- How formal or casual should ChatGPT be?
- How long or short should responses generally be?
- How do you want to be addressed?
- Should ChatGPT have opinions on topics or remain neutral?

START CLEAN WHEN EXPERIMENTING

When experimenting with different prompting techniques, **always** make sure you clear out custom instructions (make sure to save them somewhere) and start with a fresh chat session. Also, be aware of **system instructions** that you can't see that can influence your tests.

GUIDE OVERVIEW

Search   CTRL K

Prompt engineering

This guide shares strategies and tactics for getting better results from large language models (sometimes referred to as GPT models) like GPT-4. The methods described here can sometimes be deployed in combination for greater effect. We encourage experimentation to find the methods that work best for you.

Some of the examples demonstrated here currently work only with our most capable model, `gpt-4`. In general, if you find that a model fails at a task and a more capable model is available, it's often worth trying again with the more capable model.

You can also explore example prompts which showcase what our models are capable of:

Prompt examples
Explore prompt examples to learn what GPT models can do

Six strategies for getting better results

Write clear instructions

These models can't read your mind. If outputs are too long, ask for brief replies. If

SIX STRATEGIES FOR BETTER RESULTS

- Write clear instructions

- Provide reference text

- Split complex tasks into simpler subtasks

- Give the model time to "think"

- Use external tools

SPECIAL CONSIDERATIONS

- This guide is not exhaustive

- Examples are geared toward OpenAI API developers

- They do not adopt a framework

SYSTEM AND USER ROLES

ChatCompletions ⌄

```
1  from openai import OpenAI
2  client = OpenAI()
3
4  completion = client.chat.completions.create(
5    model="gpt-3.5-turbo",
6    messages=[
7      {"role": "system", "content": "You are a poetic assistant, skilled in explaining complex programming concepts with creative flair."},
8      {"role": "user", "content": "Compose a poem that explains the concept of recursion in programming."}
9    ]
10 )
11
12 print(completion.choices[0].message)
```
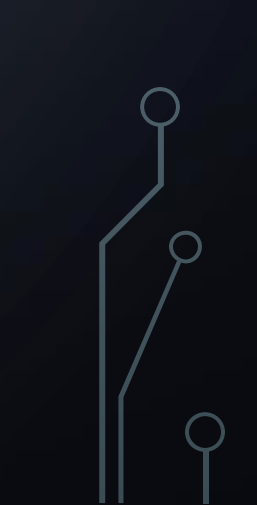
WRITE CLEAR INSTRUCTIONS

THE MODELS CAN'T READ YOUR MIND...YET

Write clear instructions

These models can't read your mind. If outputs are too long, ask for brief replies. If outputs are too simple, ask for expert-level writing. If you dislike the format, demonstrate the format you'd like to see. The less the model has to guess at what you want, the more likely you'll get it.

Tactics:

- 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

INCLUDE DETAILS IN YOUR QUERY

Worse	Better
How do I add numbers in Excel?	How do I add up a row of dollar amounts in Excel? I want to do this automatically for a whole sheet of rows with all the totals ending up on the right in a column called "Total".
Who's president?	Who was the president of Mexico in 2021, and how frequently are elections held?
Write code to calculate the Fibonacci sequence.	Write a TypeScript function to efficiently calculate the Fibonacci sequence. Comment the code liberally to explain what each piece does and why it's written that way.
Summarize the meeting notes.	Summarize the meeting notes in a single paragraph. Then write a markdown list of the speakers and each of their key points. Finally, list the next steps or action items suggested by the speakers, if any.

ASK THE MODEL TO ADOPT A PERSONA, TONE, AND/OR STYLE

SYSTEM	When I ask for help to write something, you will reply with a document that contains at least one joke or playful comment in every paragraph.
USER	Write a thank you note to my steel bolt vendor for getting the delivery in on time and in short notice. This made it possible for us to deliver an important order.

USE DELIMITERS TO CLEARLY INDICATE DISTINCT PARTS OF THE INPUT

USER

Summarize the text delimited by triple quotes with a haiku.

"""insert text here"""

Open in Playground ↗

SYSTEM

You will be provided with a pair of articles (delimited with XML tags) about the same topic. First summarize the arguments of each article. Then indicate which of them makes a better argument and explain why.

USER

<article> insert first article here </article>

<article> insert second article here </article>

Open in Playground ↗

SYSTEM

You will be provided with a thesis abstract and a suggested title for it. The thesis title should give the reader a good idea of the topic of the thesis but should also be eye-catching. If the title does not meet these criteria, suggest 5 alternatives.

USER

Abstract: insert abstract here

Title: insert title here

Open in Playground ↗

SPECIFY THE STEPS REQUIRED TO COMPLETE A TASK

SYSTEM   Use the following step-by-step instructions to respond to user inputs.

Step 1 - The user will provide you with text in triple quotes. Summarize this text in one sentence with a prefix that says "Summary: ".

Step 2 - Translate the summary from Step 1 into Spanish, with a prefix that says "Translation: ".

USER   """insert text here"""

PROVIDE EXAMPLES (SHOTS)

SYSTEM	Answer in a consistent style.
USER	Teach me about patience.
ASSISTANT	The river that carves the deepest valley flows from a modest spring; the grandest symphony originates from a single note; the most intricate tapestry begins with a solitary thread.
USER	Teach me about the ocean.

SPECIFY THE DESIRED LENGTH OF THE OUTPUT

USER

Summarize the text delimited by triple quotes in about 50 words.

"""insert text here"""

Open in Playground ↗

USER

Summarize the text delimited by triple quotes in 2 paragraphs.

"""insert text here"""

Open in Playground ↗

USER

Summarize the text delimited by triple quotes in 3 bullet points.

"""insert text here"""

Open in Playground ↗

PROVIDE REFERENCE TEXT

EVEN MODELS CAN USE SOME HELP

Provide reference text

Language models can confidently invent fake answers, especially when asked about esoteric topics or for citations and URLs. In the same way that a sheet of notes can help a student do better on a test, providing reference text to these models can help in answering with fewer fabrications.

Tactics:

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

INSTRUCT THE MODEL TO ANSWER USING A REFERENCE TEXT

SYSTEM	Use the provided articles delimited by triple quotes to answer questions. If the answer cannot be found in the articles, write "I could not find an answer."
USER	\<insert articles, each delimited by triple quotes\> Question: \<insert question here\>

INSTRUCT THE MODEL TO ANSWER WITH CITATIONS FROM A REFERENCE TEXT

SYSTEM
You will be provided with a document delimited by triple quotes and a question. Your task is to answer the question using only the provided document and to cite the passage(s) of the document used to answer the question. If the document does not contain the information needed to answer this question then simply write: "Insufficient information." If an answer to the question is provided, it must be annotated with a citation. Use the following format for to cite relevant passages ({"citation": ...}).

USER
"""<insert document here>"""

Question: <insert question here>

GPT / ASSISTANT STRATEGY

Knowledge

If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

Upload files

SPLIT COMPLEX TASKS
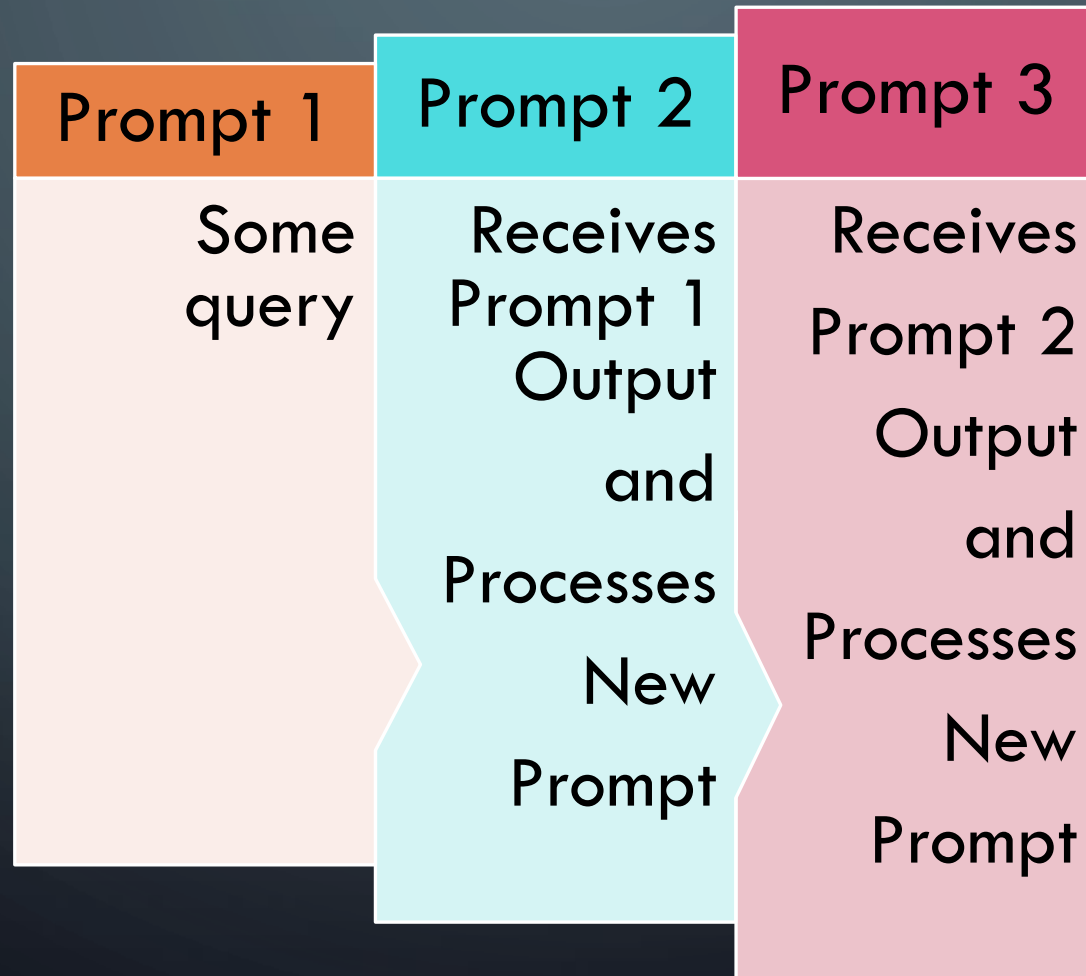INTO SIMPLER SUBTASKS

DECOMPOSE DIFFICULT TASKS

Split complex tasks into simpler subtasks

Just as it is good practice in software engineering to decompose a complex system into a set of modular components, the same is true of tasks submitted to a language model. Complex tasks tend to have higher error rates than simpler tasks. Furthermore, complex tasks can often be re-defined as a workflow of simpler tasks in which the outputs of earlier tasks are used to construct the inputs to later tasks.

Tactics:

- Use intent classification to identify the most relevant instructions for a user query
- For dialogue applications that require very long conversations, summarize or filter previous dialogue
- Summarize long documents piecewise and construct a full summary recursively

MEANT TO TAKE THE OUTPUT FROM ONE PROMPT AND USE IT AS INPUT FOR ANOTHER

Prompt 1	Prompt 2	Prompt 3
Some query	Receives Prompt 1 Output and Processes New Prompt	Receives Prompt 2 Output and Processes New Prompt

USE INTENT CLASSIFICATION TO IDENTIFY THE MOST RELEVANT INSTRUCTIONS FOR A USER QUERY

Tactic: Use intent classification to identify the most relevant instructions for a user query

For tasks in which lots of independent sets of instructions are needed to handle different cases, it can be beneficial to first classify the type of query and to use that classification to determine which instructions are needed. This can be achieved by defining fixed categories and hardcoding instructions that are relevant for handling tasks in a given category. This process can also be applied recursively to decompose a task into a sequence of stages. The advantage of this approach is that each query will contain only those instructions that are required to perform the next stage of a task which can result in lower error rates compared to using a single query to perform the whole task. This can also result in lower costs since larger prompts cost more to run (see pricing information).

SYSTEM
You will be provided with customer service queries. Classify each query into a primary category and a secondary category. Provide your output in json format with the keys: primary and secondary.

Primary categories: Billing, Technical Support, Account Management, or General Inquiry.

Billing secondary categories:
- Unsubscribe or upgrade
- Add a payment method
- Explanation for charge
- Dispute a charge

Technical Support secondary categories:
- Troubleshooting
- Device compatibility
- Software updates

Account Management secondary categories:
- Password reset
- Update personal information
- Close account
- Account security

General Inquiry secondary categories:
- Product information
- Pricing
- Feedback
- Speak to a human

USER
I need to get my internet working again.

SYSTEM

You will be provided with customer service inquiries that require troubleshooting in a technical support context. Help the user by:

- Ask them to check that all cables to/from the router are connected. Note that it is common for cables to come loose over time.
- If all cables are connected and the issue persists, ask them which router model they are using
- Now you will advise them how to restart their device:
-- If the model number is MTD-327J, advise them to push the red button and hold it for 5 seconds, then wait 5 minutes before testing the connection.
-- If the model number is MTD-327S, advise them to unplug and replug it, then wait 5 minutes before testing the connection.
- If the customer's issue persists after restarting the device and waiting 5 minutes, connect them to IT support by outputting {"IT support requested"}.
- If the user starts asking questions that are unrelated to this topic then confirm if they would like to end the current chat about troubleshooting and classify their request according to the following scheme:

<insert primary/secondary classification scheme from above here>

**USER**

I need to get my internet working again.

FOR DIALOGUE APPLICATIONS THAT REQUIRE VERY LONG CONVERSATIONS, SUMMARIZE OR FILTER PREVIOUS DIALOGUE

Tactic: For dialogue applications that require very long conversations, summarize or filter previous dialogue

Since models have a fixed context length, dialogue between a user and an assistant in which the entire conversation is included in the context window cannot continue indefinitely.

There are various workarounds to this problem, one of which is to summarize previous turns in the conversation. Once the size of the input reaches a predetermined threshold length, this could trigger a query that summarizes part of the conversation and the summary of the prior conversation could be included as part of the system message. Alternatively, prior conversation could be summarized asynchronously in the background throughout the entire conversation.

An alternative solution is to dynamically select previous parts of the conversation that are most relevant to the current query. See the tactic "Use embeddings-based search to implement efficient knowledge retrieval".
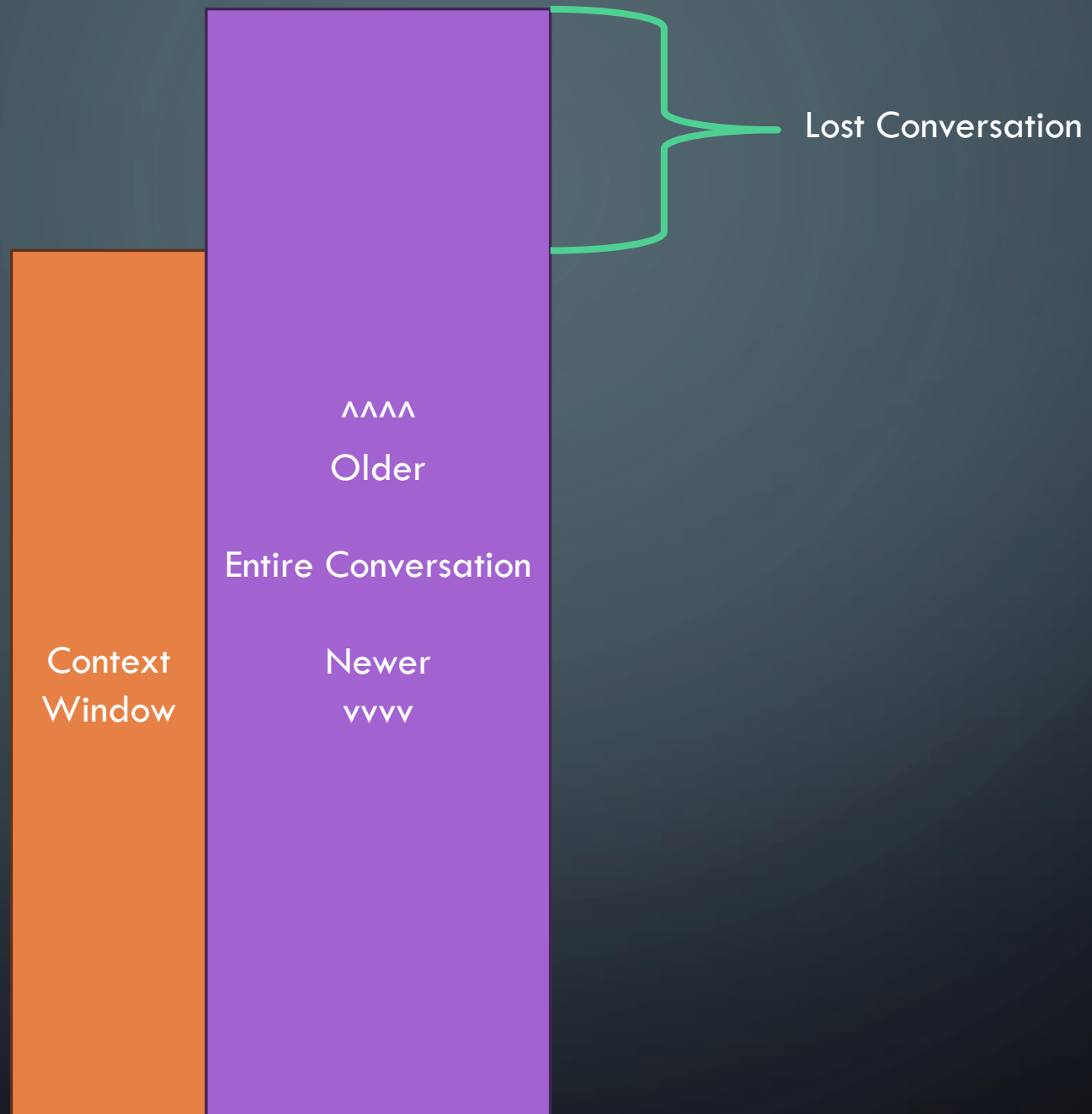
SUMMARIZE LONG DOCUMENTS PIECEWISE AND CONSTRUCT A FULL SUMMARY RECURSIVELY

Tactic: Summarize long documents piecewise and construct a full summary recursively

Since models have a fixed context length, they cannot be used to summarize a text longer than the context length minus the length of the generated summary in a single query.

To summarize a very long document such as a book we can use a sequence of queries to summarize each section of the document. Section summaries can be concatenated and summarized producing summaries of summaries. This process can proceed recursively until an entire document is summarized. If it's necessary to use information about earlier sections in order to make sense of later sections, then a further trick that can be useful is to include a running summary of the text that precedes any given point in the book while summarizing content at that point. The effectiveness of this procedure for summarizing books has been studied in previous research by OpenAI using variants of GPT-3.

GIVE THE MODEL TIME TO "THINK"

RODIN WOULD BE PROUD