Generate Prompt

Prompts are the starting point for the LLM generation process. They provide context, set the tone, define objectives, and ultimately shape the quality of the response.

Prompt design is an entire field of its own. See What is Prompt Engineering in AI and How Does it Work?↗.

In this topic, you can learn how to lay out a starting prompt for a basic RAG pipeline. Then read about the CO-STAR technique that can be used to organize prompts for more complex tasks.

Starting prompt

Problem: You have proprietary knowledge you want to accurately impart to readers, using direct examples. You consider this problem, and decide the best solution is to create a chatbot powered by an LLM, with a vector database of your documentation for a backend.

We recommend you start with LangChain's basic prompt template and OpenAI for the embedding and querying.

```
template = """
Answer the question based only on the supplied context. \
If you don't know the answer, say you don't know the answer.
Context: {context}
Question: {question}
Your answer:
"""
```

Further prompt refinement

This is a good prompt to get started, but you can further refine it to solve your problem and improve the docs experience.

How can you improve the first prompt?

- Users want to know they aren't getting bad information or wasting API calls.
- Operators want to know the chatbot is functioning correctly.
- Users who want to know more would like a quick link to supporting documentation.

- Users have deeper questions, and your engineers can help! Maybe they could connect on Slack or Discord.

You decide to add a few things to the prompt template:

```
template = """
You are a tool called IRL Company Chatbot. \
Your purpose is to use the below documentation from the
IRL Company to answer the subsequent documentation questions. \
Also, if possible, give the reference URLs
according to the following instructions. The way to create the URLs is: \
add "https://docs.irl.ai/docs/" before the "slug" value of the document. \
For any URL references that start with "doc:" or "ref:" \
use its value to create a URL by adding "https://docs.irl.ai/docs/" before tha
For reference URLs about release notes add "https://docs.irl.ai/changelog/" \
before the "slug" value of the document. \
Do not use page titles to create urls. \
If the answer cannot be found in the documentation, write "I could not find an
Join our [Slack community](https://www.irl.ai/slackinvite) for further clarifi
Do not make up an answer or give an answer that does not exist in the provided
Context: {context}
Question: {question}
Helpful answer:"""
```

This solves your problem and improves the docs experience.

- Users know they're getting good information, because the sources are clearly identified. Operators can see the bot is functioning correctly.
- Users who want to know more can click the reference links. The chatbot knows how to build links to your docs without manual intervention.
- Users who have deeper questions can connect with your engineers on Slack.

The additional prompting defines context and new objectives for the LLM. However, just like humans, LLMs can get confused if the prompt is poorly organized or contains non-relevant text. We'll further refine the prompt with the CO-STAR framework.

CO-STAR prompt framework

A simple way to further improve quality of generation is to follow the CO-STAR framework, a template that sections out key aspects that most influence response generation. CO-STAR was originally developed by GovTech Singapore's Data Science and Artificial Intelligence Division.

CO-STAR stands for:

- (C) Context: Provide background information on the task.
- (O) Objective: Define what the task is that you want the LLM to perform.
- (S) Style: Specify the writing style you want the LLM to use.
- (T) Tone: Set the attitude of the response.
- (A) Audience: Identify who the response is intended for.
- (R) Response: Provide the response format.

Here's how you would use the CO-STAR framework to organize your previous prompt.

```
prompt_template = """
# CONTEXT #
You are a tool called IRL Company Chatbot. \
You are a technical expert with a specific knowledge base supplied to you via

# OBJECTIVE #
* Answer questions based only on the given context.
* If possible, include reference URLs in the following format: \
add "https://docs.irl.ai/docs" before the "slug" value of the document. \
For any URL references that start with "doc:" or "ref:" \
use its value to create a URL by adding "https://docs.irl.ai/docs/" before tha
For reference URLs about release notes add "https://docs.irl.ai/changelog/" \
before the "slug" value of the document. \
Do not use page titles to create urls. \
* If the answer cannot be found in the documentation, write "I could not find
Join our [Slack Community](https://www.irl.ai/slackinvite) for further clarifi
* Do not make up an answer or give an answer that is not supported by the prov

# STYLE #
Follow the writing style of technical experts.

# TONE #
Professional

# AUDIENCE #
People that want to learn about IRL Company.

# RESPONSE #
The response should be in the following format:
---
answer

url_reference
---
```

```
Context: {context}
Question: {question}
Your answer:
"""
```

You can see how CO-STAR guides the LLM through a structured approach to answering questions. This helps the LLM (and the programmer) solve the problem at hand and reduces the chance of generating non-relevant text.