

Chatbot!

User types
in a prompt



Gets a
response

A terminal window with a dark background and a title bar with three colored circles (red, yellow, green) on the left. The title bar contains the word 'Terminal'. The terminal shows a prompt '>>' followed by the text 'What is 1+1?'. Below this, the number '2' is displayed as the response.

```
Terminal

>> What is 1+1?

2
```



Terminal

>> What is 1+1?

2

>> And if you add 3
more?

5

Lots of Loaded Terminology

Large Language Model (LLM)

GPT-3

Completion Model

GPT-4

Chat Model

ChatGPT

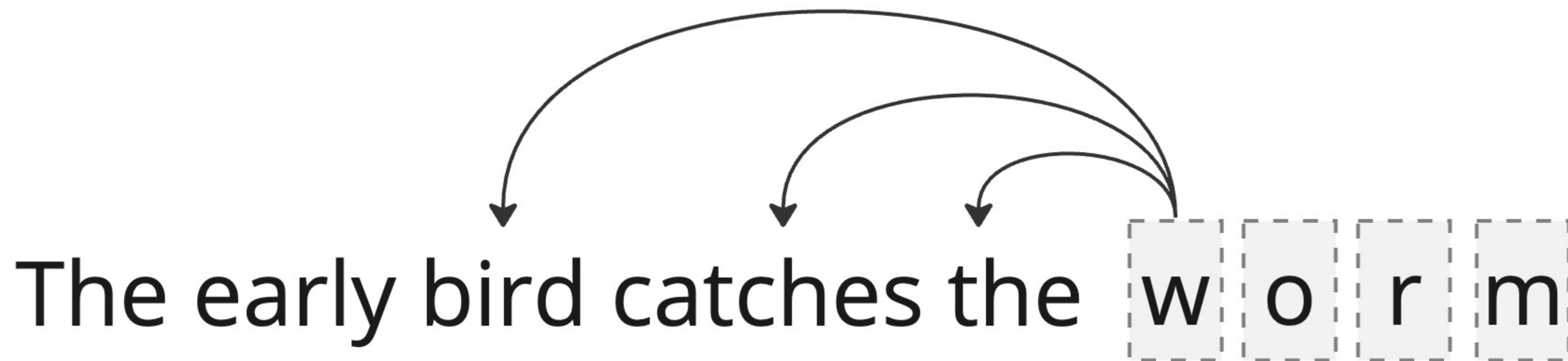
Even if you know the difference,
LangChain introduces a little confusion...

Large Language Model (LLM)

An algorithm for generating text based upon a prompt

Predicts the next likely bit of text by looking at the preceding text

Large Language Model



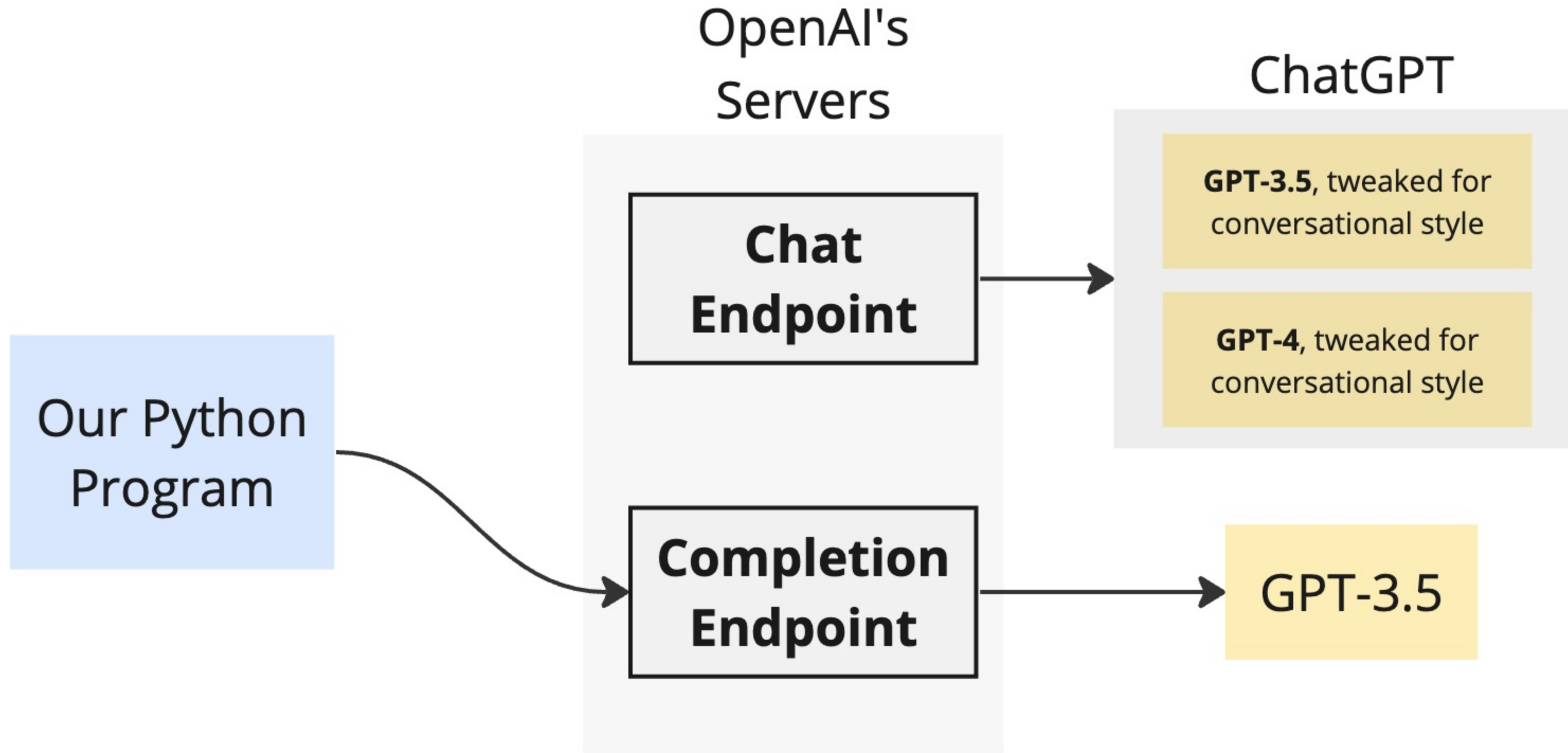
Completion Model

A large language model

Chat Model

A LLM that is customized to generate text in a conversational format

Still, fundamentally, a LLM.



LLM's + LangChain

LLM = Large Language Model.
Algorithm that generates text

There are two styles of interfacing with an LLM - **completion** and **chat**.

LangChain refers to **completion**-based models as '**LLM's**

LangChain refers to **chat**-based models as '**Chat Models**'

LLM's + LangChain

```
1 from langchain.llms import OpenAI
2
3 llm = OpenAI()
```

*Gives us an object that will use
OpenAI's **completion** endpoint*

```
1 from langchain.chat_models import ChatOpenAI
2
3 chat = ChatOpenAI()
```

*Gives us an object that will use
OpenAI's **chat** endpoint*

```
1 from langchain.llms import OpenAI
```

```
2
```

```
3 llm = OpenAI()
```

*Gives us an object that will use
OpenAI's **completion** endpoint*

Completion model



Large Language Model that tries to add text onto the end of a prompt that you give it

Think of it like a really fancy autocomplete

```
1 from langchain.chat_models import ChatOpenAI
2
3 chat = ChatOpenAI()
```

*Gives us an object that will use
OpenAI's **chat** endpoint*

Chat model



Large Language Model designed for
back-and-forth conversational style text
generation

Its still a LLM!

Chain

PromptTemplate to use

Language model to use

Output parser to use

Memory to use

Callbacks to use

Completion-Based
PromptTemplate

Completion-Based
Model

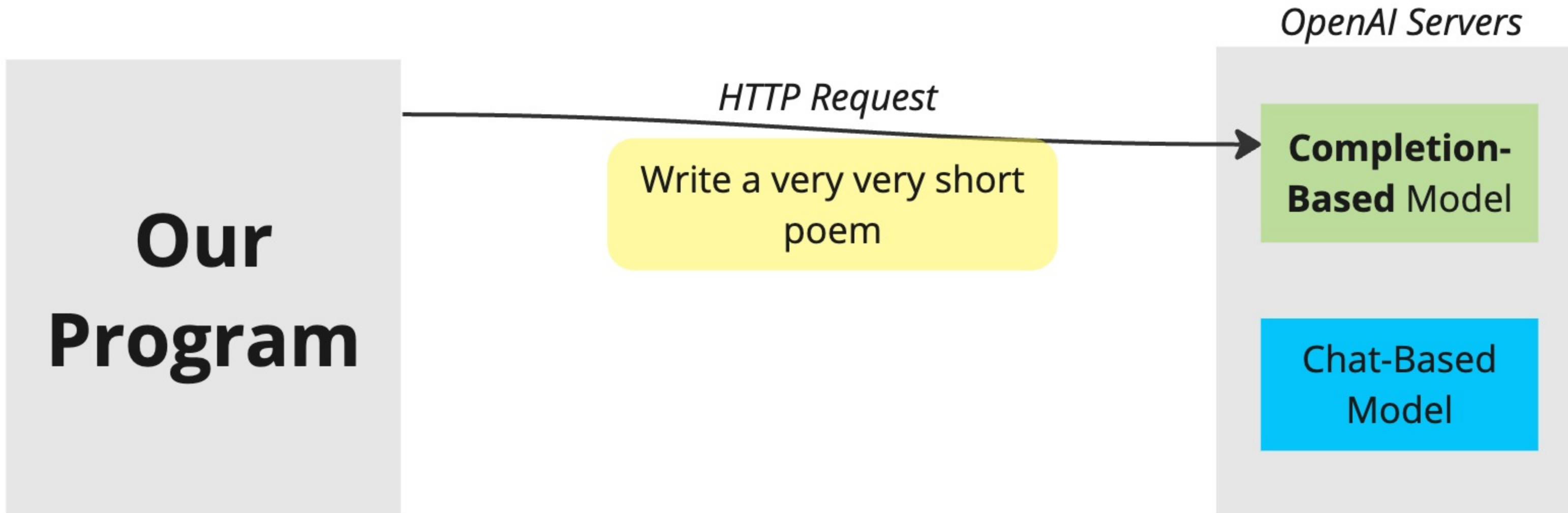
Chat-Based
PromptTemplate

Chat-Based
Model

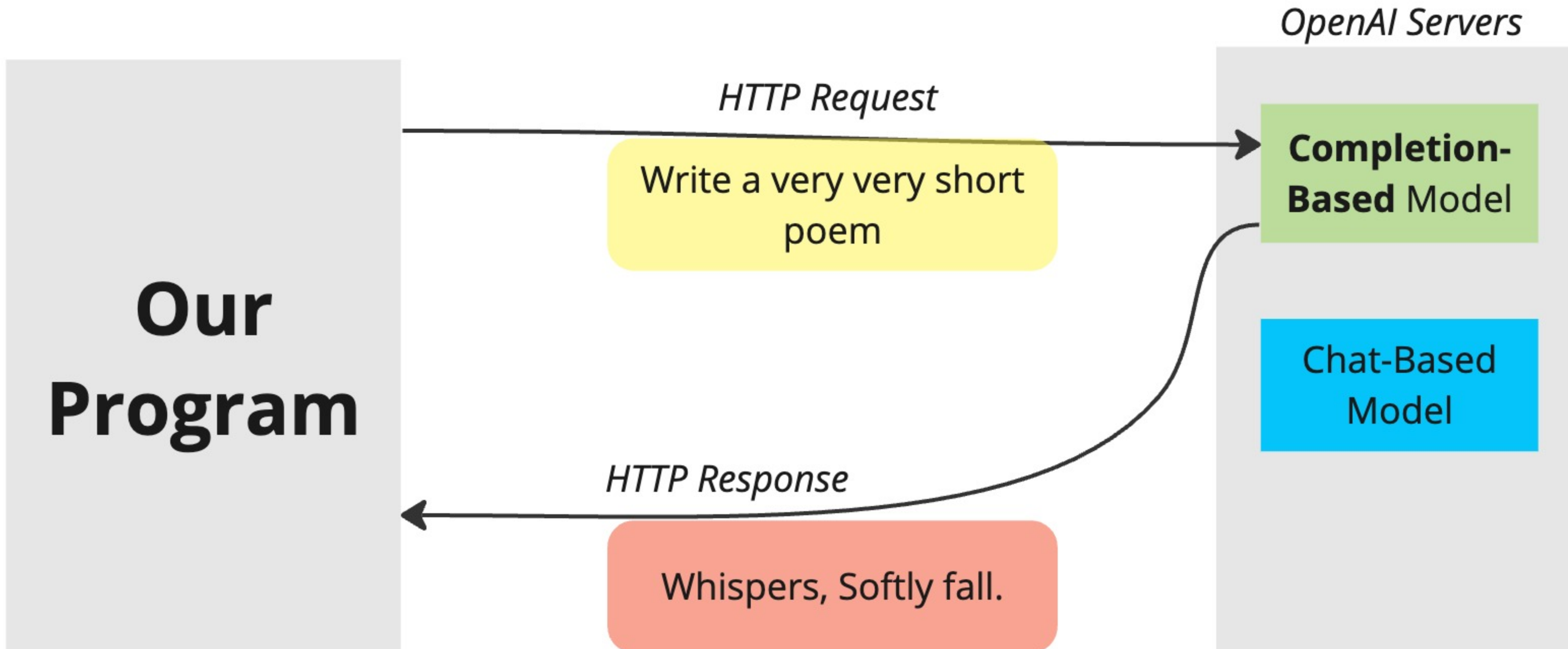
Chat-Based models can be used in a Chain, just like a completion-based model

A Chat-Based Prompt Template is, unfortunately, a little more confusing

Using a Completion-Based Model



Using a Completion-Based Model



Using Chat-Based Completions

The conversation is encoded as a list of messages

To generate text, we send the OpenAI API our list of messages

The API is stateless - the whole list must be sent each time to continue a conversation

SYSTEM


You are a helpful assistant.

USER

What does a comedian say?

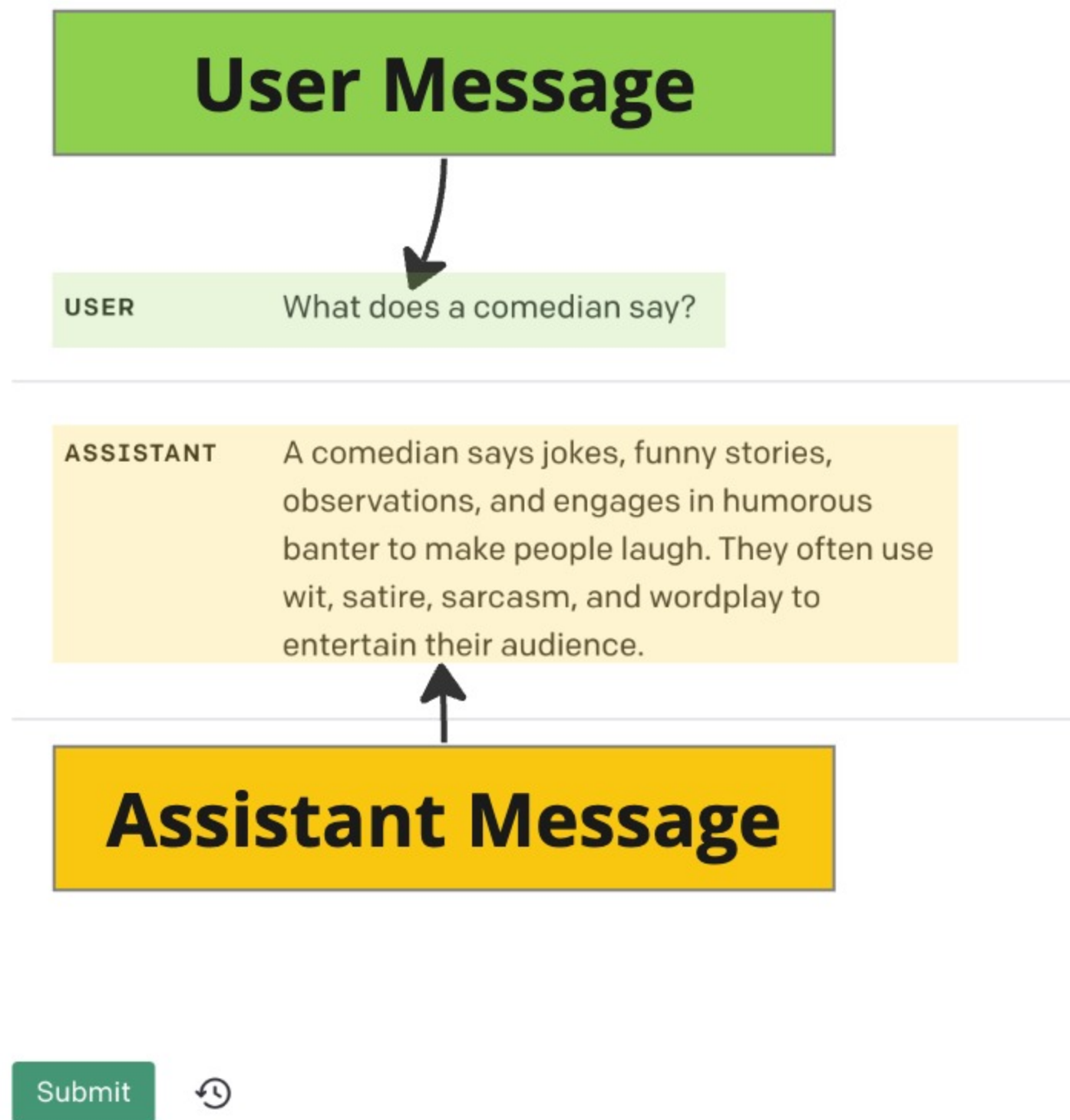
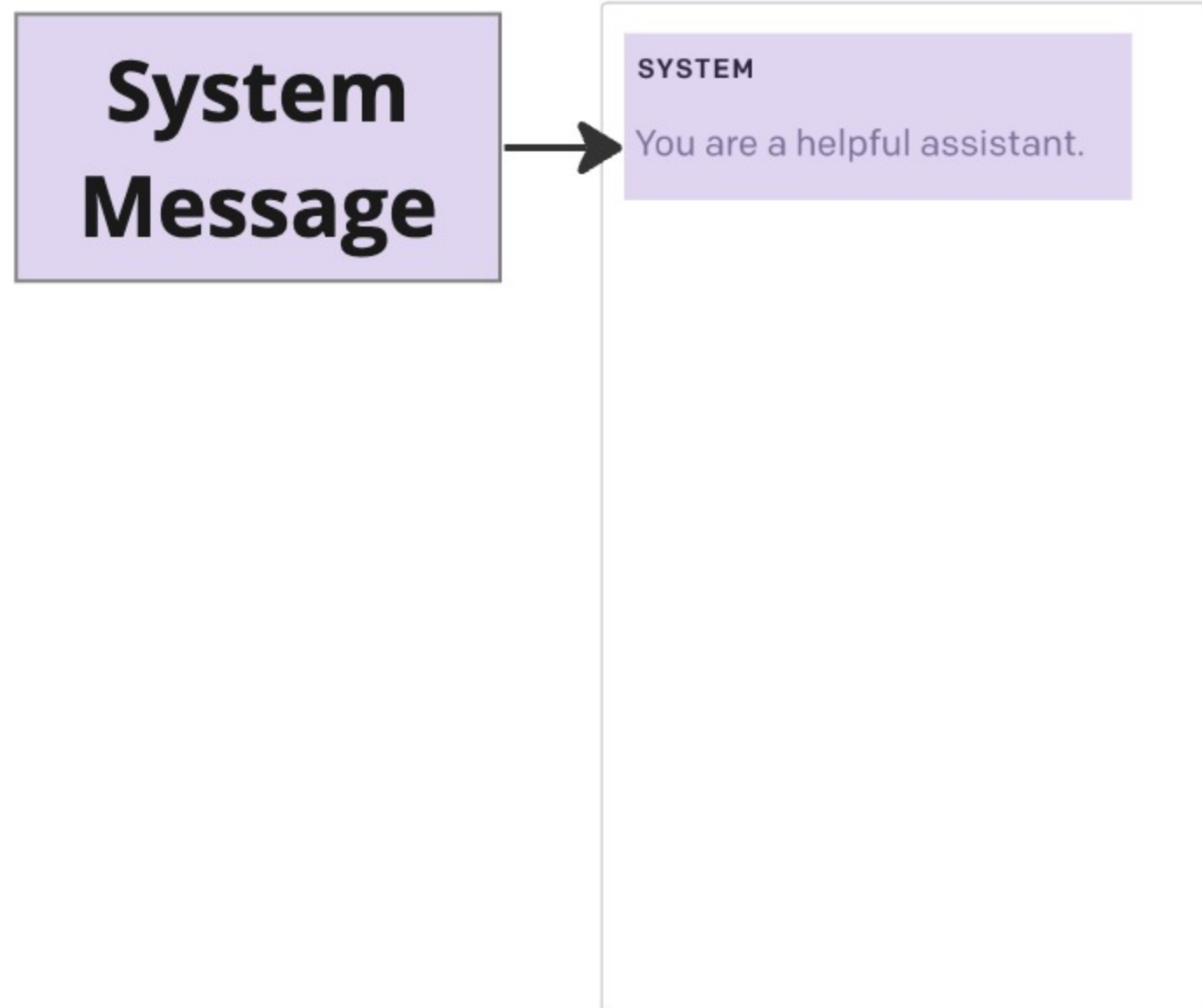
ASSISTANT

A comedian says jokes, funny stories, observations, and engages in humorous banter to make people laugh. They often use wit, satire, sarcasm, and wordplay to entertain their audience.

 Add message

Submit





User Message



A message produced by the **user**

Assistant Message



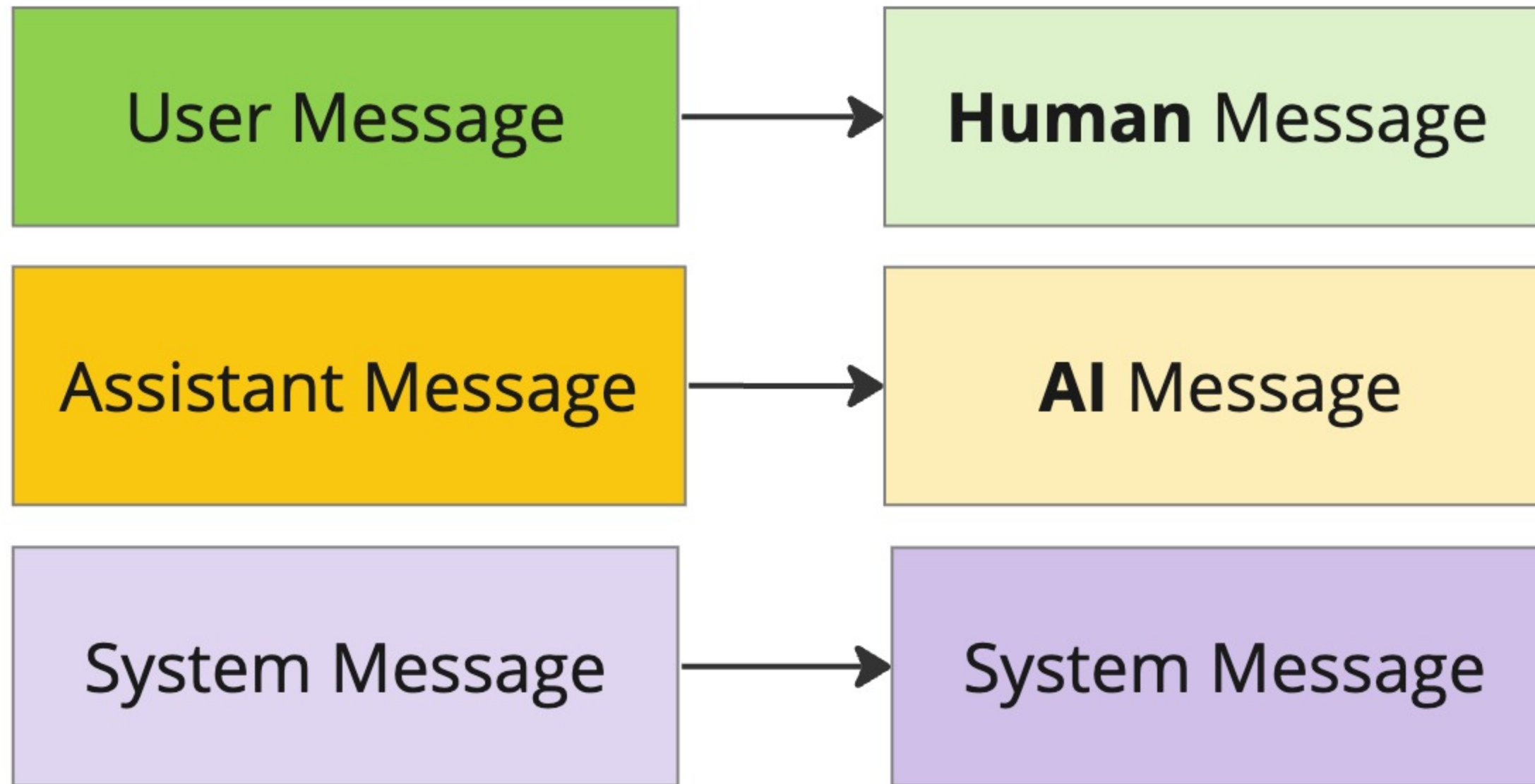
A message produced by the **chat model**

System Message



A message that gives some initial directions to the chat model
Usually produced by us (the developers)

LangChain uses some different terminology



SYSTEM

You are a hilarious chat bot.
Always respond with a terrible pun.

USER

Why do I pay taxes?

ASSISTANT

Because it's a taxing way to support the country!

USER

Why are oranges sweet?

ASSISTANT

Because they don't want to be too citrus-tive!

⊕

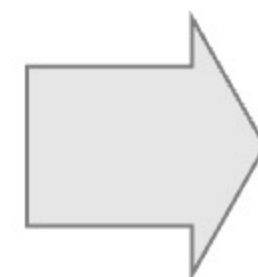
Add message

Submit

↺

🚩

Give us feedback



List of Messages

System Message

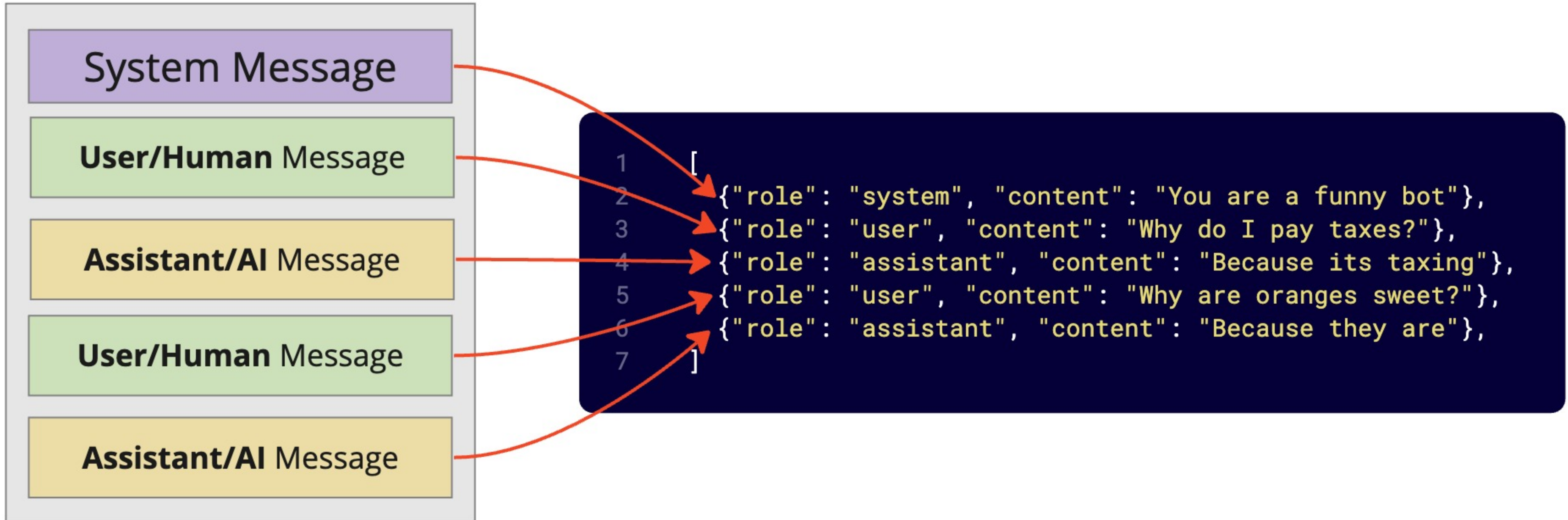
Human Message

AI Message

Human Message

AI Message

List of Messages



Using a **Chat**-Based Model

**Our
Program**

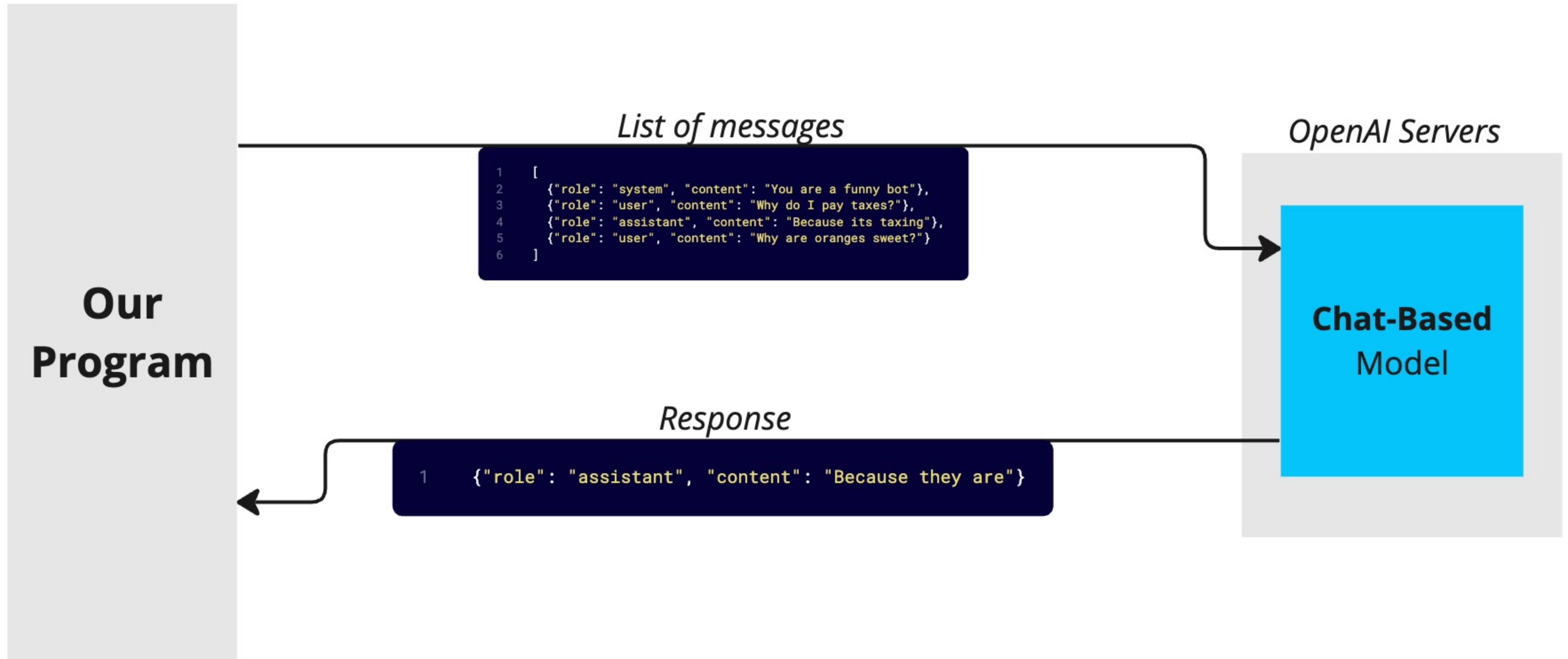
List of messages

```
1  [  
2    {"role": "system", "content": "You are a funny bot"},  
3    {"role": "user", "content": "Why do I pay taxes?"},  
4    {"role": "assistant", "content": "Because its taxing"},  
5    {"role": "user", "content": "Why are oranges sweet?"}  
6  ]
```

OpenAI Servers

**Chat-Based
Model**

Using a **Chat**-Based Model



Reference to an
earlier message



What is $1+1$?



The sum of 1 and 1 is 2.



And 3 more?



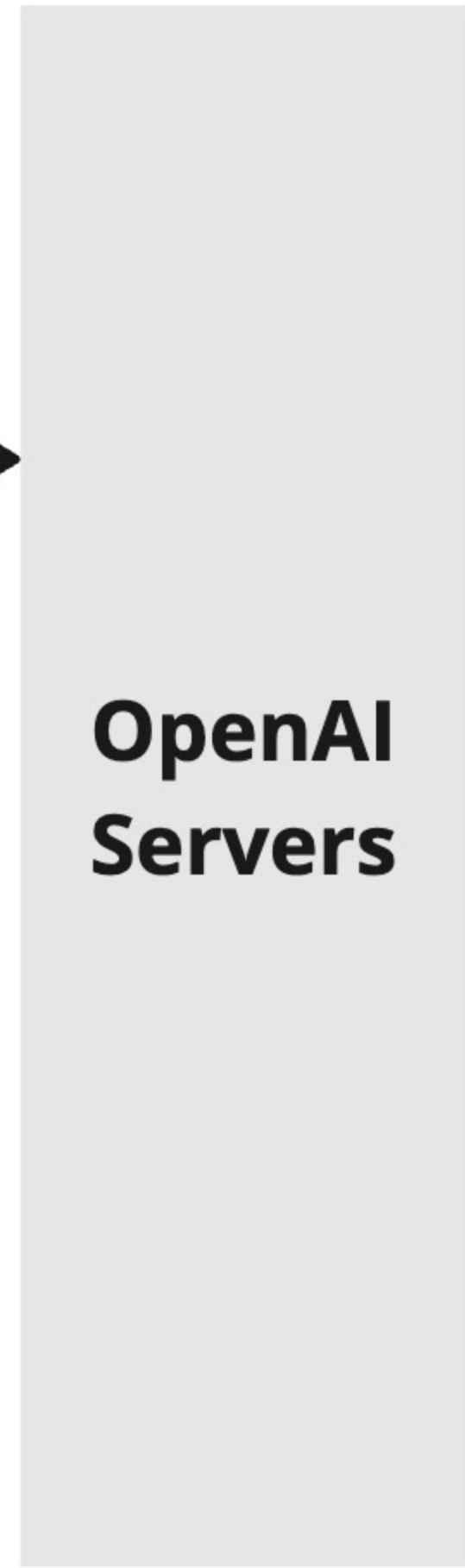
If you add 3 more to the previous result of 2, the sum would be 5. So, $1 + 1 + 3 = 5$.



**Our
Program**

List of messages

Role	Content
user	What is 1+1?



**OpenAI
Servers**

**Our
Program**

Role	Content
user	What is 1+1?

**OpenAI
Servers**

Role	Content
assistant	The result of 1+1 is 2

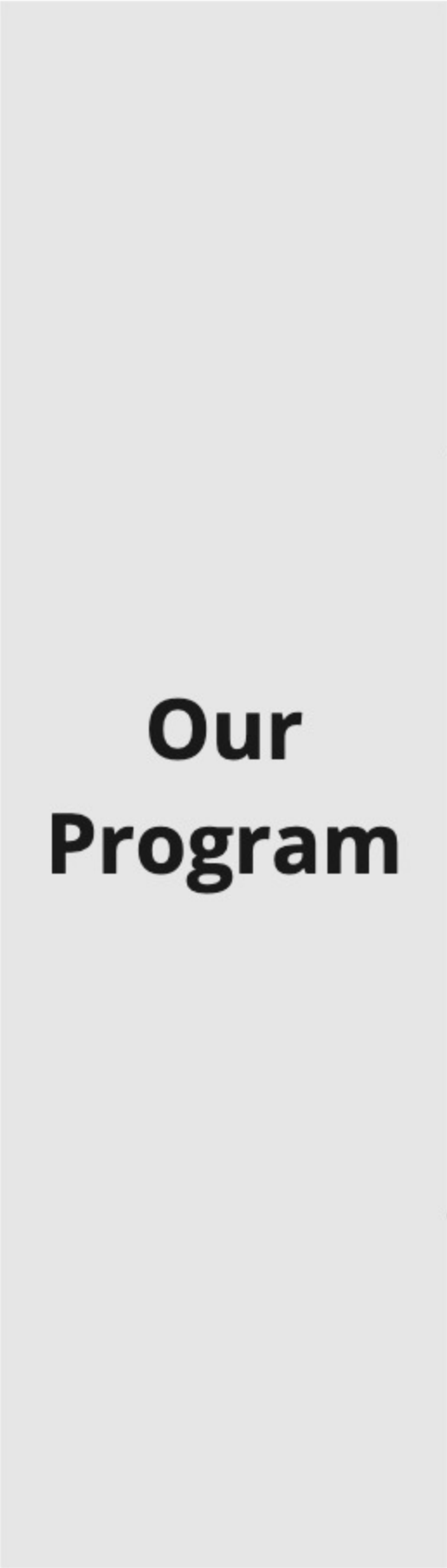
Our Program

Role	Content
user	And 3 more?

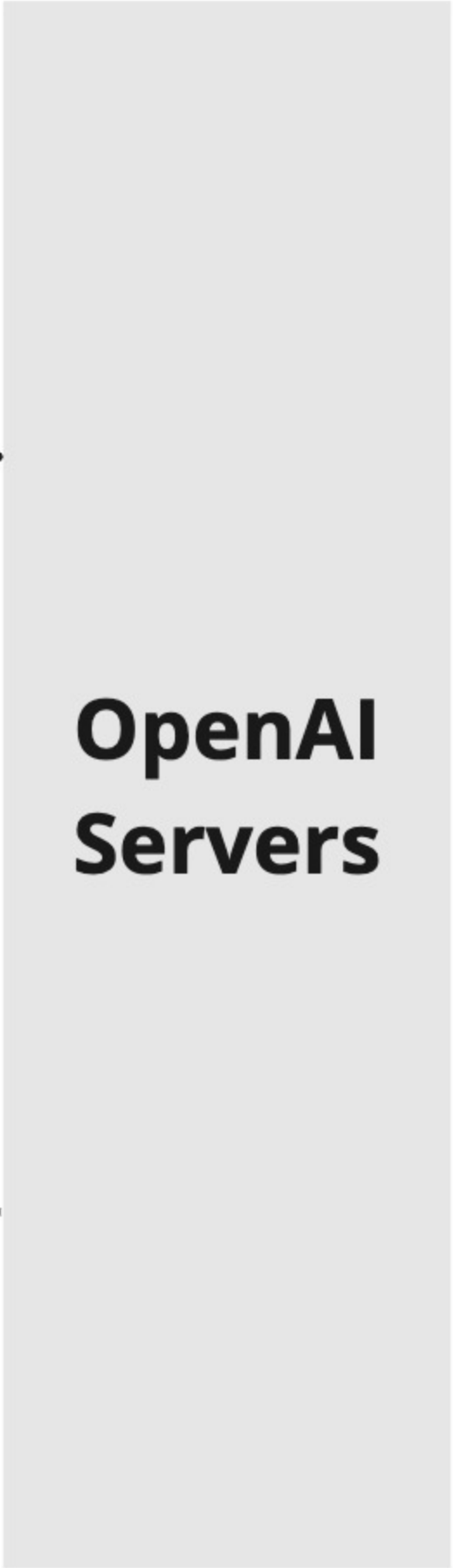
ChatGPT has no
idea what we're
referring to

OpenAI Servers

Role	Content
assistant	??



Role	Content
user	What is 1+1?
assistant	The result of 1+1 is 2
user	And 3 more?



Role	Content
assistant	2 + 3 is 5

