

Generative AI Ambassadors Immersion Day

How to develop a PoC demo

Heiko Hotz (he/him)

Senior AI/ML Specialist Solutions Architect
AWS



Agenda

Challenges with GenAI

Build a first "Hello world" web app

Build a first chatbot app

Build a first Chain with LangChain

Bringing it all together!

Wrap-Up and Q&A

Challenges with GenAI

- Hallucinations
- Knowledge cutoff
- Legal uncertainties

Overcoming knowledge cutoff



Who is the current prime minister of the UK? Where was he or she born? How far is their birth place from London?



The current Prime Minister of the UK is Boris Johnson. He was born in New York City, United States on June 19, 1964. However, he renounced his US citizenship in 2016.

Boris Johnson's birthplace, New York City, is located approximately 3,458 miles (5,568 kilometers) from London, UK.

LLM supercharged with LangChain

You:

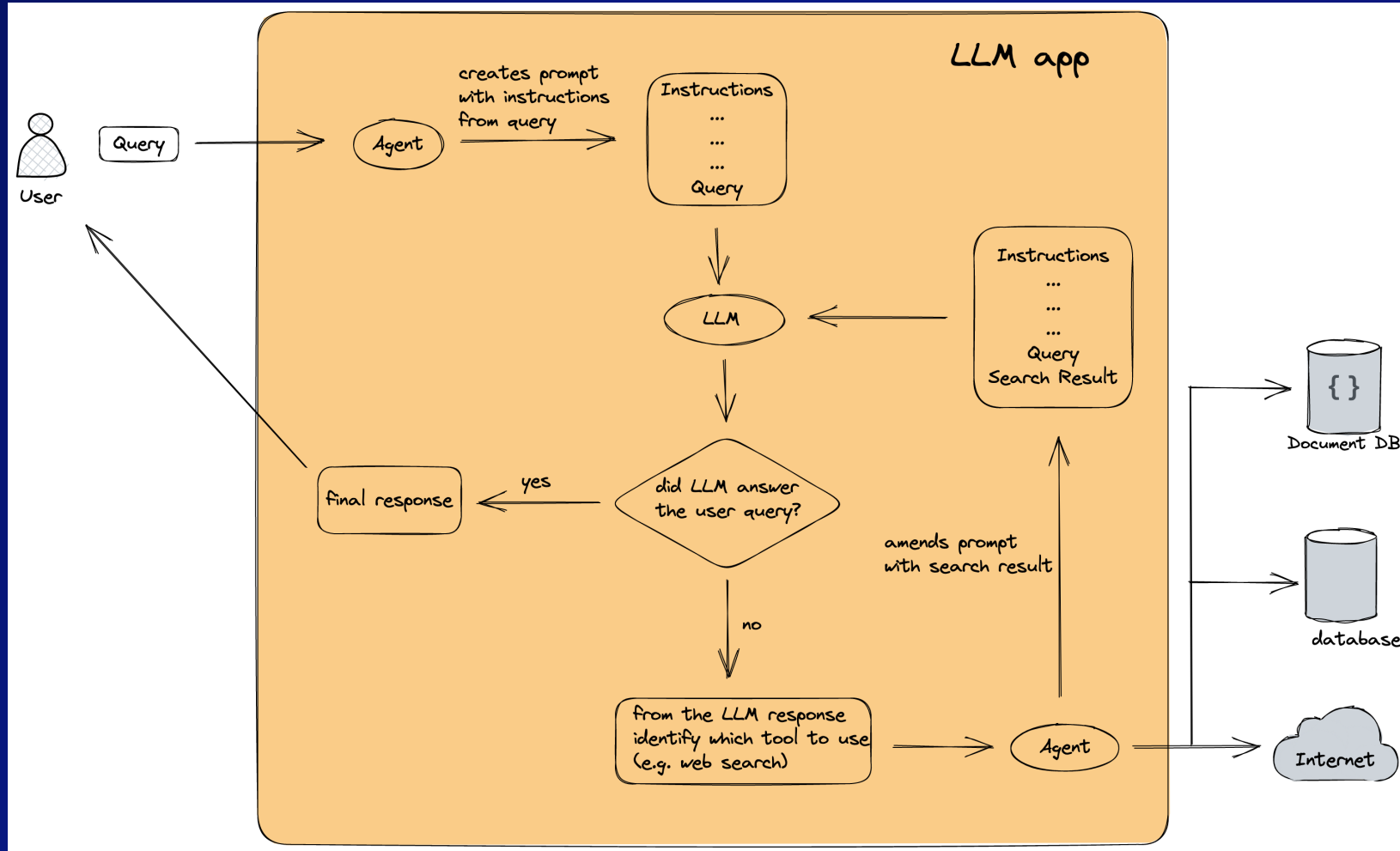


Rishi Sunak was born in Southampton, United Kingdom, which is 69.47 miles from London.

Who is the current prime minister of the UK? Where was he or she born? How far is their birth place from London?



How does RAG work?



Agenda

Challenges with GenAI

Build a first "Hello world" web app

Build a first chatbot app

Build a first Chain with LangChain

Bringing it all together!

Wrap-Up and Q&A

Hello world web app

Hello Streamlit App

Please enter your name:

Submit

Hello Heiko!

<https://docs.streamlit.io/library/get-started/create-an-app>

Hello world web app

```
import streamlit as st

# Setting the title of the app
st.title("Hello Streamlit App")

# Creating a text input box
text_input = st.text_input("Please enter your name:")

# Creating a submit button
if st.button("Submit"):
    # Display the hello message
    st.write(f"Hello {text_input}!")
```


Agenda

Challenges with GenAI

Build a first "Hello world" web app

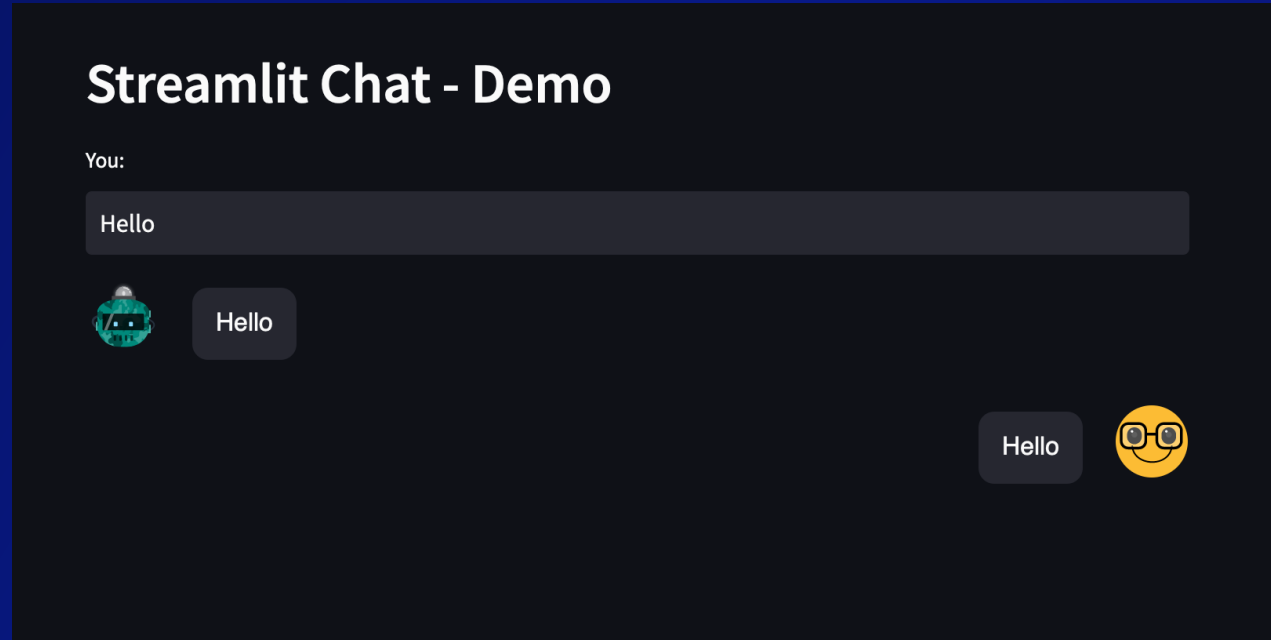
Build a first chatbot app

Build a first Chain with LangChain

Bringing it all together!

Wrap-Up and Q&A

First chat app



<https://github.com/AI-Yash/st-chat/blob/main/examples/chatbot.py>

First chat app

```
1  import streamlit as st
2  from streamlit_chat import message
3
4  st.set_page_config(
5      page_title="Streamlit Chat - Demo",
6      page_icon=":robot:"
7  )
8
9  st.header("Streamlit Chat - Demo")
10
11  if 'generated' not in st.session_state:
12      st.session_state['generated'] = []
13
14  if 'past' not in st.session_state:
15      st.session_state['past'] = []
16
17
18  def get_text():
19      input_text = st.text_input("You: ", key="input")
20      return input_text
21
22
23  user_input = get_text()
24
25  if user_input:
26      output = "Hello"
27      st.session_state.past.append(user_input)
28      st.session_state.generated.append(output)
29
30  if st.session_state['generated']:
31      for i in range(len(st.session_state['generated'])-1, -1, -1):
32          message(st.session_state["generated"][i], key=str(i))
33          message(st.session_state['past'][i], is_user=True, key=str(i) + '_user')
34
```

Agenda

Challenges with GenAI

Build a first "Hello world" web app

Build a first chatbot app

Build a first Chain with LangChain

Bringing it all together!

Wrap-Up and Q&A

First Chain

```
A good name for a company that makes colorful socks could be "Rainbow Socks Co." or "Colorful Steps".
```

https://python.langchain.com/en/latest/modules/chains/generic/llm_chain.html

<https://python.langchain.com/en/latest/modules/models/llms/integrations/sagemaker.html>

First Chain

```
1  from langchain import PromptTemplate, LLMChain
2  from langchain.llms.sagemaker_endpoint import LLMContentHandler, SagemakerEndpoint
3  import json
4  from typing import Dict
5
6
7  prompt_template = "What is a good name for a company that makes {product}?"
8  endpoint_name = "vicuna-13b-1-1-2023-06-03-15-48-50-242"
9
10
11  class ContentHandler(LLMContentHandler):
12      content_type = "application/json"
13      accepts = "application/json"
14      len_prompt = 0
15
16      def transform_input(self, prompt: str, model_kwargs: Dict) -> bytes:
17          self.len_prompt = len(prompt)
18          input_str = json.dumps({"inputs": prompt, "parameters": {"max_new_tokens": 50, "do_sample": False, "repetition_penalty": 1.1}})
19          return input_str.encode('utf-8')
20
21      def transform_output(self, output: bytes) -> str:
22          response_json = output.read()
23          res = json.loads(response_json)
24          # ans = res[0]['generated_text']
25          ans = res[0]['generated_text'][self.len_prompt:]
26          return ans
27
28
29  content_handler = ContentHandler()
30  llm = SagemakerEndpoint(
31      endpoint_name=endpoint_name,
32      region_name="us-east-1",
33      content_handler=content_handler,
34      credentials_profile_name="default"
35  )
36
37  llm_chain = LLMChain(
38      llm=llm,
39      prompt=PromptTemplate.from_template(prompt_template)
40  )
41
42  print(llm_chain("colorful socks")["text"])
```

Agenda

Challenges with GenAI

Build a first "Hello world" web app

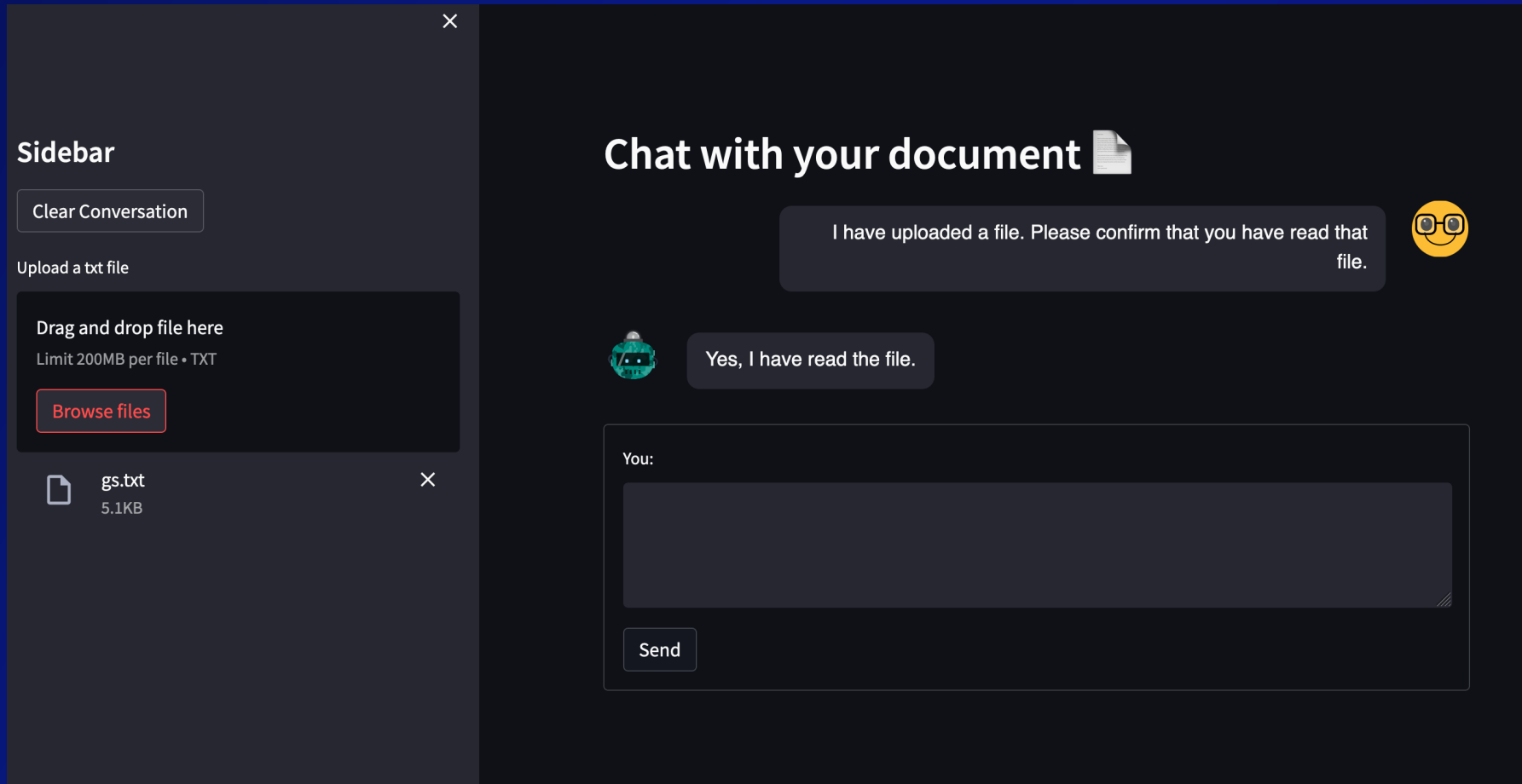
Build a first chatbot app

Build a first Chain with LangChain

Bringing it all together!

Wrap-Up and Q&A

Document chatbot



https://github.com/marshmellow77/document_chatbot/

Agenda

Challenges with GenAI

Build a first "Hello world" web app

Build a first chatbot app

Build a first Chain with LangChain

Bringing it all together!

Wrap-Up and Q&A