

SageMakerEndpoint

This notebooks goes over how to use an LLM hosted on a SageMaker endpoint.

```
!pip3 install langchain boto3
```

```
from langchain.docstore.document import Document
```

```
example_doc_1 = """
Peter and Elizabeth took a taxi to attend the night party in the city. While in
the party, Elizabeth collapsed and was rushed to the hospital.
Since she was diagnosed with a brain injury, the doctor told Peter to stay besides
her until she gets well.
Therefore, Peter stayed with her at the hospital for 3 days without leaving.
"""

docs = [
    Document(
        page_content=example_doc_1,
    )
]
```

```
from typing import Dict

from langchain import PromptTemplate, SagemakerEndpoint
from langchain.llms.sagemaker_endpoint import ContentHandlerBase
from langchain.chains.question_answering import load_qa_chain
import json

query = """How long was Elizabeth hospitalized?
"""

prompt_template = """Use the following pieces of context to answer the question at
the end.

{context}

Question: {question}
Answer: """
PROMPT = PromptTemplate(
    template=prompt_template, input_variables=["context", "question"]
)
```

[Skip to main content](#)

```
class ContentHandler(ContentHandlerBase):
    content_type = "application/json"
    accepts = "application/json"

    def transform_input(self, prompt: str, model_kwargs: Dict) -> bytes:
        input_str = json.dumps({prompt: prompt, **model_kwargs})
        return input_str.encode('utf-8')

    def transform_output(self, output: bytes) -> str:
        response_json = json.loads(output.read().decode("utf-8"))
        return response_json[0]["generated_text"]

content_handler = ContentHandler()

chain = load_qa_chain(
    llm=SagemakerEndpoint(
        endpoint_name="endpoint-name",
        credentials_profile_name="credentials-profile-name",
        region_name="us-west-2",
        model_kwargs={"temperature":1e-10},
        content_handler=content_handler
    ),
    prompt=PROMPT
)

chain({"input_documents": docs, "question": query}, return_only_outputs=True)
```