How to use the async API for LLMs

LangChain provides async support for LLMs by leveraging the asyncio library.

Async support is particularly useful for calling multiple LLMs concurrently, as these calls are network-bound. Currently, OpenAI, PromptLayerOpenAI, ChatOpenAI and Anthropic are supported, but async support for other LLMs is on the roadmap.

You can use the agenerate method to call an OpenAI LLM asynchronously.

```
import time
import asyncio
from langchain.llms import OpenAI
def generate_serially():
   11m = OpenAI(temperature=0.9)
   for _ in range(10):
        resp = llm.generate(["Hello, how are you?"])
        print(resp.generations[0][0].text)
async def async generate(llm):
    resp = await llm.agenerate(["Hello, how are you?"])
    print(resp.generations[0][0].text)
async def generate concurrently():
   11m = OpenAI(temperature=0.9)
   tasks = [async_generate(llm) for _ in range(10)]
    await asyncio.gather(*tasks)
s = time.perf_counter()
# If running this outside of Jupyter, use asyncio.run(generate concurrently())
await generate_concurrently()
elapsed = time.perf counter() - s
print('\033[1m' + f"Concurrent executed in {elapsed:0.2f} seconds." + '\033[0m')
s = time.perf counter()
generate_serially()
```

Skip to main content

```
elapsed = time.perf_counter() - s
print('\033[1m' + f"Serial executed in {elapsed:0.2f} seconds." + '\033[0m')
```

```
I'm doing well, thank you. How about you?
I'm doing well, thank you. How about you?
I'm doing well, how about you?
I'm doing well, thank you. How about you?
I'm doing well, thank you. How about you?
I'm doing well, thank you. How about yourself?
I'm doing well, thank you! How about you?
I'm doing well, thank you. How about you?
I'm doing well, thank you! How about you?
I'm doing well, thank you. How about you?
Concurrent executed in 1.39 seconds.
I'm doing well, thank you. How about you?
I'm doing well, thank you. How about yourself?
I'm doing well, thanks for asking. How about you?
I'm doing well, thanks! How about you?
```

I'm doing well, thank you. How about you?

I'm doing well, thank you. How about yourself?

I'm doing well, thanks for asking. How about you? Serial executed in 5.77 seconds.