



Chainlit: How to Build a ChatGPT-like User Interface in Python

Chris Alexiuk, Head of LLMs, Machine Learning Maker Space


Greg Loughnane, Founder, Machine Learning Maker Space



Outline

-  Chainlit + arXiv
-  Chainlit
 - Chain of Thought
 - Prompt Playground
 - AskArXiv with LangChain
- Q&A

News! Parameter Efficient Fine-Tuning v0.4.0

 **Sourab Mangrulkar**
@sourab_m

😊 PEFT 0.4.0 is out with exciting features! ✨

- 1 4-bit QLoRA via bitsandbytes
- 2 IA3 method addition
- 3 Support for new tasks: QA and Feature Extraction
- 4 AutoPeftModelForxxx for Simplified UX
- 5 LoRA for custom models & new utils

Release Notes 📖: [github.com/huggingface/pe...](https://github.com/huggingface/peft)

1/7








peft v0.4.0

- Official **QLoRA** support
train adapters on top of 4bit base models
- **IA3** support
new adapter method, more memory efficient than LoRA
- **New task:** Feature Extraction
perform feature extraction using adapters!
- **Auto-mapping** feature
load and use peft models with one line of code
- **LoRA** for any model
apply LoRA and push adapter weights for any pytorch model

5:43 AM · Jul 18, 2023 · 20.6K Views

Let's AskArxiv!

 Spaces |  ml-maker-space / **ArxivChainLitDemo**   like | 0  Running

Chat **Readme**

⚠ Warning ⚠

You will need a GPT-4 API key to use this app due to large context size!

Welcome to AskArxiv powered by Chainlit!

In this app, you'll be able to enter a topic - and then ask ~30 papers from Arxiv about that topic!

Link To Demo

[Hugging Face Space](#)

Chain of Thought Prompting

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27.



Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓

Chain of Thought Reasoning

User
20:45:42


What is the Paris weather forecast for tomorrow? How does it compare to today's?

Used LLMChain ^

LLMChain
20:45:46

Thought: I need to find out the weather forecast for Paris for tomorrow and compare it to today's.
Action:

```
{  
  "action": "Search",  
  "action_input": "Paris weather forecast for tomorrow"  
}
```



Search
20:45:47

Paris, Ville de Paris, France Weather Forecast, with current conditions, wind, air quality, and what to expect for the next 3 days.

LLMChain
20:45:50

I need to compare the forecast for tomorrow to today's.
Action:

<https://docs.chainlit.io/examples/mrkl>

Prompt Playground

- Beyond visualization
- Check specific prompts
- Improve them

User
16:36:55

what is the tallest mountain?

Took 1 step ^


LLMChain
16:37:20

First, we need to define what we mean by "tallest mountain." Do we mean the mountain with the highest peak above sea level, or the mountain with the greatest vertical relief from base to summit?

If we're talking about the highest peak above sea level, then the answer is Mount Everest, which stands at 29,029 feet (8,848 meters) above sea level.

If we're talking about the mountain with the greatest vertical relief from base to summit, then the answer is Mauna Kea in Hawaii, which rises 33,500 feet (10,210 meters) from its base on the ocean floor to its summit. However, much of Mauna Kea is submerged underwater, so it is not always considered a "true" mountain by some definitions.

So, depending on how we define "tallest mountain," the answer could be either Mount Everest or Mauna Kea.

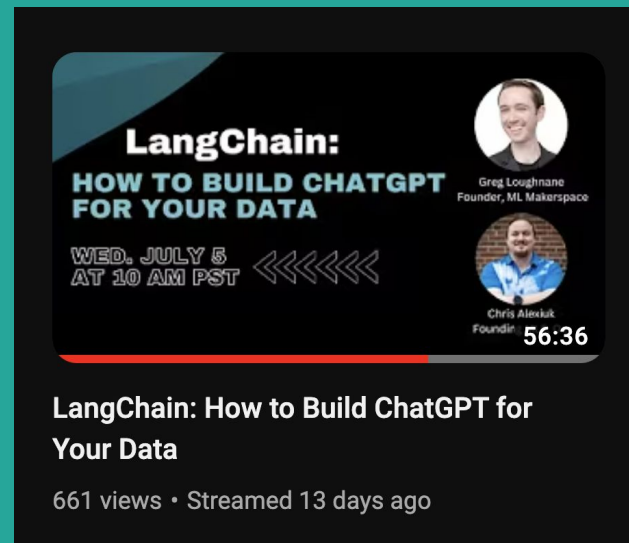


Chatbot
16:37:20

First, we need to define what we mean by "tallest mountain." Do we mean the mountain with the highest peak

LangChain

“The real power comes when you can combine [LLMs] with other sources of computation and knowledge.” ~ Harrison Chase, Creator of LangChain



youtube.com/@MLMakerspace/streams

Building a Chainlit Application with LangChain

1. Create app.py
2. Write application logic
3. Run it!

```
chainlit run app.py
```

```
import os
from langchain import PromptTemplate, OpenAI, LLMChain
import chainlit as cl

os.environ["OPENAI_API_KEY"] = "YOUR_OPEN_AI_API_KEY"

template = """Question: {question}

Answer: Let's think step by step."""

@cl.langchain_factory(use_async=True)
def factory():
    prompt = PromptTemplate(template=template, input_variables=["question"])
    llm_chain = LLMChain(prompt=prompt, llm=OpenAI(temperature=0), verbose=True)

    return llm_chain
```

And don't forget a README.md

chainlit.md

Warning

You will need a GPT-4 API key to use this app due to large context size!

Welcome to AskArxiv powered by Chainlit!

In this app, you'll be able to enter a topic - and then ask ~30 papers from Arxiv about that topic!

Link To Demo

[Hugging Face Space](#)

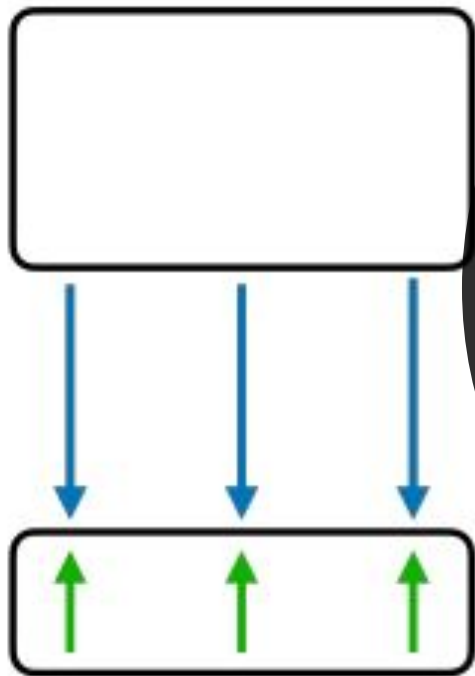
Our Chainlit-arXiv Application Flow



1. Specify topic for arXiv to search for papers on
 - a. “Please enter a topic to begin!”
2. Get top 3 papers (PDFs)
 - a. “Downloading and chunking articles...”
3. Convert papers to document embeddings and store in Vector Database
4. Ask specific questions related to content
 - a. “We found a few papers about {X} you can now ask questions!”
5. Return answers to questions **with** sources



Full Finetuning (No Adapters)

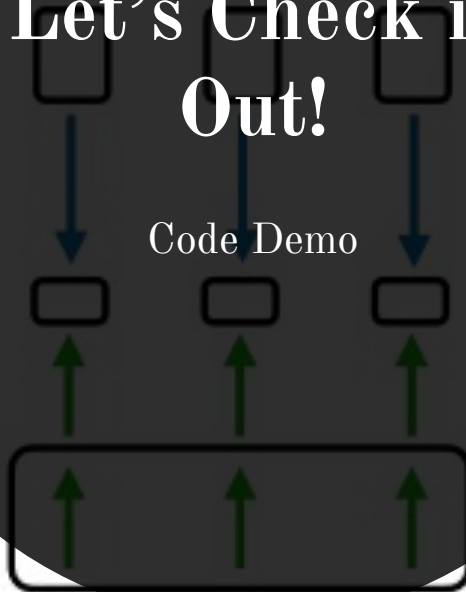


16-bit Transformer

LoRA

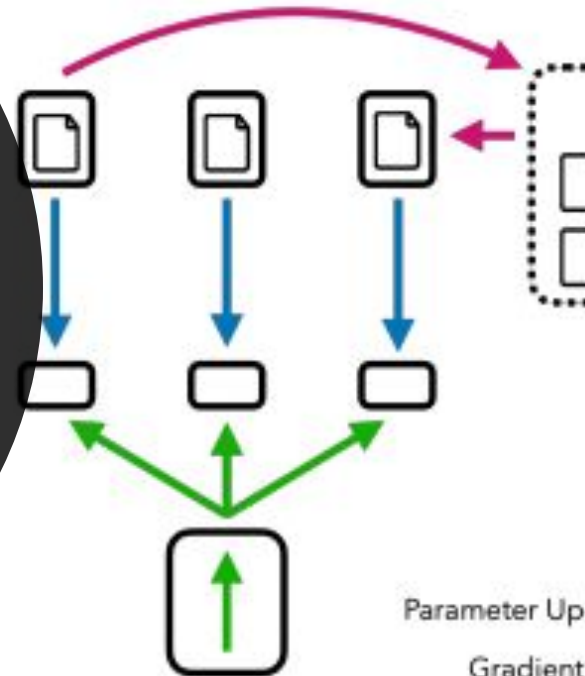
Let's Check it Out!

Code Demo



16-bit Transformer


QLoRA



4-bit Transformer

Parameter Up
Gradient
Paging

Conclusions, Chainlit

- A new best-practice tool for LLM UI MVPs
- CoT Reasoning & Prompt Playground help you dial in flows and performance
-  Built with LangChain in mind!
- App.py + chainlit.md is all you need to get going!




Streamlit

 **Chainlit**

Thank You!

- GitHub Repository
 - [Introduction to Chainlit](#) (AskArXiv)
- AskArXiv Demo App!
 - [Demo!](#)
- Reach out directly with any questions!
 - greg@mlmaker.space
 - chris@mlmaker.space

<https://maven.com/mlmakerspace/llmops>