

MLOps Final Project - LLM



Luke Schwenke & Aaron Chan
December 4, 2023

Contents

1. HuggingFace Model Overview & Architecture
2. Prompt Engineering:
 - a. Zero-Shot
 - b. Few-Shot
 - c. Chain-of-Thought
 - d. Retrieval Augmented Generation (RAG)
3. Instruction Tuning with MASH-QA dataset



LLM Model: **microsoft/phi-1_5**



Overview:

- Transformer with 1.3 billion parameters train on 30 billion tokens
- phi-1.5 demonstrates a nearly state-of-the-art performance among models with less than 10 billion parameters
- Trained on python code, Q&A forums, textbooks, and more from phi-1.0. Augmented with NLP synthetic texts.
- Synthetic training data helps prevent toxicity and bias

Intended Uses

Given the nature of the training data, phi-1.5 is best suited for prompts using the QA format, the chat format, and the code format. Note that phi-1.5, being a base model, often produces irrelevant text following the main answer. In the following example, we've truncated the answer for illustrative purposes only.

https://huggingface.co/microsoft/phi-1_5

Microsoft's phi-1.5 Architecture



Transformer with 24 layers, 32 heads, and each head has dimension 64. Rotary embeddings with rotary dimension of 32, and context length 2048. Flash-attention [DFE+ 22, Dao23] are used for training speed up, and the tokenizer of codegen-mono [NPH+ 22] is also used.

*How small can a LLM be to achieve certain capabilities
(common sense reasoning)*

[Access the full technical report here](#)

Zero-Shot

Prompt	Response
What types of exercise are best for people with asthma?	Aerobic exercise, such as walking, swimming, or cycling, is best for people with asthma
How is obsessive-compulsive disorder diagnosed?	Obsessive-compulsive disorder is diagnosed through a combination of interviews, psychological tests, and observations.
When are you more likely to get a blood clot?	You are more likely to get a blood clot when you are sitting or lying down for a long time, or when you have a medical condition that affects your blood circulation.
How should you lift objects to prevent back pain?	Bend your knees and keep your back straight.
How can you be smart with antibiotics?	Take them exactly as prescribed by your doctor.

Few-Shot

Prompt	Response
Swimming and walking can be good for people who have asthma. What types of exercise are best for people with asthma?	Swimming and walking are good for people with asthma.
OCD can be diagnosed by obsessions. How is obsessive-compulsive disorder diagnosed?	By obsessions.
Blood clots are more likely in people with high blood pressure. When are you more likely to get a blood clot?	You are more likely to get a blood clot when you have high blood pressure.
Lift objects with your legs to prevent back pain. How should you lift objects to prevent back pain?	Lift objects with your legs to prevent back pain.
Antibiotics should only be taken in specific doses when you need them most. How can you be smart with antibiotics?	Antibiotics should only be taken when you need them most. This means that you should only take antibiotics when you have a bacterial infection, and not when you have a viral infection. Antibiotics are only effective against bacteria, not viruses. Taking antibiotics when you don't need them can lead to antibiotic resistance

Irrelevant Info Example with Few-Shot

Prompt

Lift objects with your legs to prevent back pain. How should you lift objects to prevent back pain?

Response

Lift objects with your legs to prevent back pain.

Once upon a time, in a small town called Greenville, there lived a young girl named Lily. Lily was a curious and adventurous girl who loved exploring the world around her. One day, she stumbled upon an old, dusty book in her grandfather's attic. The book was filled...

Chain-of-Thought

#1

Question: What is a good exercise for people with breathing problems?
Answer: Light exercise like walking can be good for people with these issues.
Question: What types of exercise are best for people with asthma?



Model Response: Aerobic exercise like walking, swimming, or cycling can be good for people with asthma.

#2

Question: How can mental disorders be diagnosed?
Answer: Generally these disorders are diagnosed through various tests and interviews with a psychiatrist.
Question: How is an obsessive-compulsive disorder diagnosed?



Model Response: An obsessive-compulsive disorder is diagnosed through a combination of interviews, psychological tests, and a thorough medical history.

Chain-of-Thought Cont.

#3

Question: When do blood problems usually show up?

Answer: Blood problems are often the result of high blood pressure in people over 30 years old.

Question: When are you more likely to get a blood clot?



Model Response: You are more likely to get a blood clot if you have a blood clotting disorder or if you have had a blood clot in the past.

#4

Question: How should heavy objects be lifted?

Answer: Heavy objects should be lifted with your legs primarily.

Question: How should you lift objects to prevent back pain?



Model Response: You should lift objects with your legs to prevent back pain.

Chain-of-Thought Cont.

#5

Question: When should you take medication?

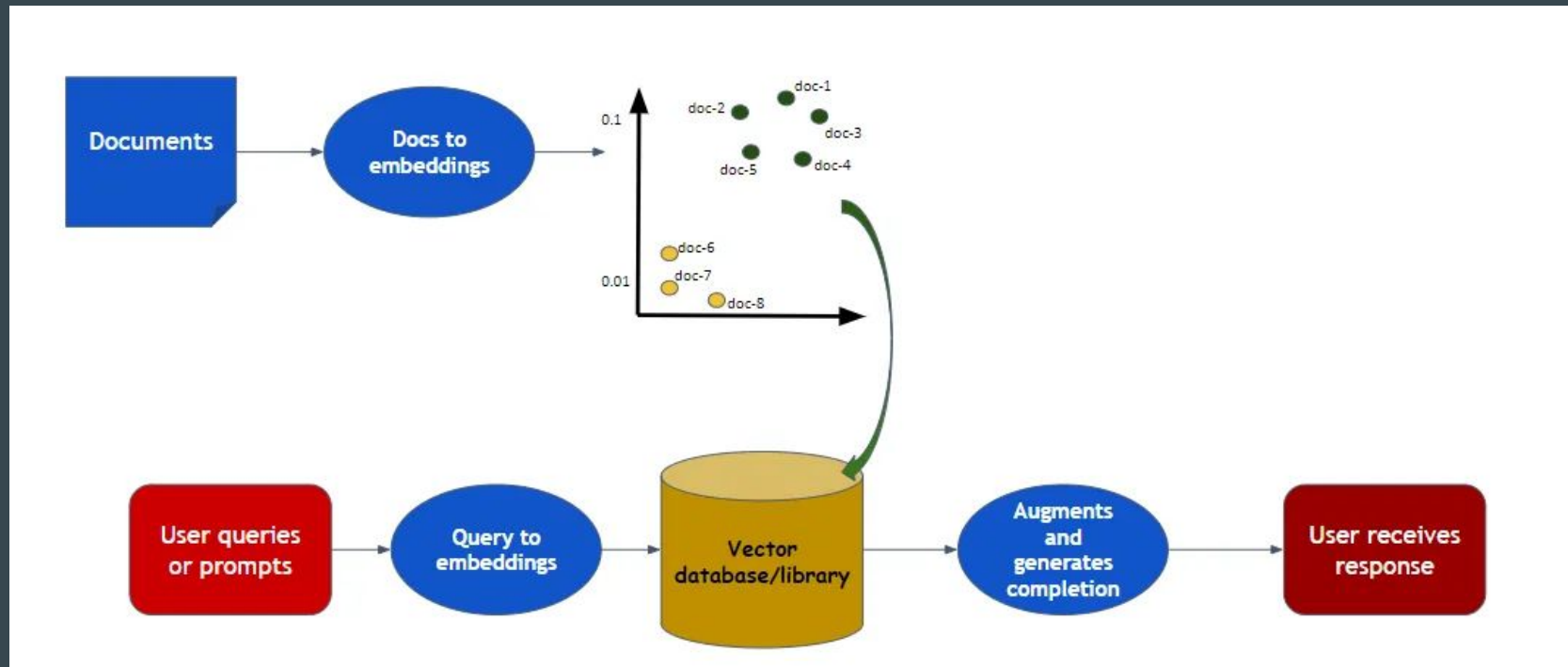
Answer: Only take medication when necessary and when prescribed by your doctor.

Question: How can you be smart with antibiotics?



Model Response: Take antibiotics exactly as prescribed and finish the entire course, even if you start feeling better before the medication is finished.

Retrieval Augmented Generation (RAG)



Retrieval Augmented Generation (RAG)

Process:

1. Gather documents that we want be able to reference

<https://huggingface.co/datasets/databricks/databricks-dolly-15k>

Next, perform transformations. Example: splitting into smaller chunks

2. Transform documents into text embedding vectors
3. Store embeddings into a vector database for retrieval
4. Select an LLM that you want to be able to prompt ([`microsoft/phi-1_5`](#))
5. A retrieval step combines the prompting of an LLM with a retriever to obtain relevant documents related to a query

Retrieval Augmented Generation (RAG)

Prompt	Response	Associated Document
What types of exercise are best for people with asthma?	Exercise is important for people with asthma, but it is important to choose the right type of exercise. Aerobic exercise, such as running, swimming, or cycling, is generally recommended for people with asthma. This type of exercise helps to improve lung function and strengthen the respiratory muscles. It is also important to warm up before exercising and to cool down afterwards to prevent any potential asthma symptoms from worsening.	An acute asthma exacerbation is commonly referred to as an asthma attack. The classic symptoms are shortness of breath, wheezing, and chest tightness. The wheezing is most often when breathing out. While these are the primary symptoms of asthma, some people present primarily with coughing,
How is obsessive-compulsive disorder diagnosed?	OCD is diagnosed based on the presence of: a) Sensory integration and processing difficulties b) Anxiety problems c) Attention deficit hyperactivity disorder (ADHD) d) Behavioural disorders e) Autism spectrum disorders	Symptoms may vary according to the disorder's type and subtype present. Sensory integration and processing difficulties can be a feature of a number of disorders, including anxiety problems, attention deficit hyperactivity disorder (ADHD), food intolerances, behavioural disorders, and particularly, autism spectrum.....
When are you more likely to get a blood clot?	When you are more likely to get a blood clot is when you have a history of high blood pressure, smoking, diabetes mellitus, lack of exercise, obesity, high blood cholesterol, poor diet, excessive alcohol consumption, and poor sleep.	The underlying mechanisms for developing cardiovascular diseases (CVD) vary depending on the disease. It is estimated that dietary risk factors are associated with 53% of CVD deaths. Coronary artery disease, stroke, and peripheral artery disease involve atherosclerosis. This may be....
How should you lift objects to prevent back pain?	You should lift objects with your legs, not your back. Explanation: Lifting objects with your back can put unnecessary strain on your spine and lead to back pain. Instead, it is recommended to lift with your legs, keeping your back straight and using your leg muscles to lift the object. This helps distribute the weight evenly and reduces the risk of injury.	Spinal manipulation, which chiropractors call "spinal adjustment" or "chiropractic adjustment", is the most common treatment used in chiropractic care. Spinal manipulation.....
How can you be smart with antibiotics?	Answer: You can be smart with antibiotics by using them only when they are needed, and by following the instructions of your doctor or pharmacist. You should not share your antibiotics with others, or use them for viral infections, such as colds or flu, because they are ineffective against viruses. You should also finish the full course of antibiotics, even if you start feeling better, because this will help prevent antibiotic resistance, which is when bacteria become resistant to the effects of antibiotics and can cause more	Fleming grew the mould in a pure culture and found that the culture broth contained an antibacterial substance. He investigated its anti-bacterial effect on many organisms, and noticed that it affected bacteria such as staphylococci and many other Gram-positive pathogens that cause scarlet fever, pneumonia, meningitis and diphtheria, but not typhoid fever or paratyphoid fever, which are caused by Gram-negative bacteria, for which he was seeking a cure at the time. It also affected Neisseria gonorrhoeae, which causes

Instruction Tuning

- Completed using the MASH-QA (Multiple Answer Spans Healthcare Question Answering) dataset
- The dataset was parsed into a format acceptable by the LLM model we chose
- Utilized a QLoRA (Quantized Low-Rank Apadaptation) approach to fine-tune the model
- Once we train the QLoRA adapter, we merge it with our pre-trained model to perform inference
- Trained on a 2k subset of the total dataset due to resource constraints
- Instruction Tuning did not alter the model architecture

Instruction Tuning

Prompt	Response
What types of exercise are best for people with asthma?	Answer: Aerobic exercise: Aerobic exercise is a type of exercise that gets your heart rate up and makes you breathe harder. It's good for people with asthma because it helps your lungs get stronger. Aerobic exercise can also help you los....
How is obsessive-compulsive disorder diagnosed?	Answer: Your doctor will ask you about your symptoms and your family history. You may also be asked to fill out a questionnaire. Your doctor may also order a blood test to check for a chemical imbalance in your brain. Your doctor may also order a brain scan, such as an MRI or CT scan,
When are you more likely to get a blood clot?	Answer: You're more likely to get a blood clot if you have: -- A family history of blood clots. You're more likely to get a blood clot if you're: -- A smoker. -- A woman who's had a blood clot in the past. -- A man who's had a blood clot i...
How should you lift objects to prevent back pain?	Answer: If you're lifting something heavy, bend your knees and keep your back straight. You should also keep your back close to your body. Don't lift with your back. Instead, bend your knees and lift with....
How can you be smart with antibiotics?	Answer: If you have a bacterial infection, your doctor may prescribe antibiotics. You can help your body fight the infection by: Taking your medicine exactly as directed. Not sharing your medicine with others. Not stopping your medicine before your doctor tells you to. Not taking your medicine for the fu...