

Exercise task 11

Instructions:

Building a basic chatbot

In this exercise, you will build a minimal chatbot using the transformers library and a pretrained model.

First, if you haven't done so already, study the Using LLMs with transformers notebook. That should give you all the necessary prerequisites for completing this exercise.

Using what you learned from that notebook, write a python script that takes input from a user, passes that to an LLM, and prints out the LLM response, repeating until the user requests to exit while **keeping track of the message history**. You can use this template as a starting point

```
1 [YOUR CODE HERE: load model]
2
3 messages = []
4 while True:
5     user_input = input('Say something ("exit" to quit): ')
6     if user_input == 'exit':
7         break
8     messages.append({'role': 'user', 'content': user_input})
9     [YOUR CODE HERE: call model, print its output, and add it to
    messages]
```

(You may want to tweak generation parameters and add additional features like having the text "reset" reset the message queue to make testing easier.)

You can use e.g. one of the following models:

- HuggingFaceTB/SmolLM2-135M-Instruct
- HuggingFaceTB/SmolLM2-360M-Instruct
- HuggingFaceTB/SmolLM2-1.7B-Instruct

(For best results, use the largest model with GPU acceleration. If you're running on a GPU with more memory, you can of course use even larger models, just pick an appropriate one from <https://huggingface.co/models>.)

Test your chatbot with the following and provide logs of your discussions:

1. At least 5 questions about basic facts about the world (e.g. the capital of a country)
2. At least 5 arithmetic questions ranging from trivial ("what is 1+1?") to more complex
3. Inform the system of a secret word (e.g. "zebra"), then after a few other questions ask it what the secret word is. Make sure you understand where the memory of that secret word is.

In very brief, what did you think about the capabilities of the chatbot you created? Of the factors impacting LLM performance that we discussed on the lecture, which one (or ones) do you think are primarily limiting its performance?

Solutions:

The code is in a separate file (`exercise_task_11.py`). I used the medium model (`HuggingFaceTB/SmoLM2-360M-Instruct`) because the largest one decided to not run on my graphics card. Below are the logs of the conversation history as generated by my chatbot (blue, bold text: user; regular text: system):

Logs:

What's the capital of Vietnam?

The capital of Vietnam is Hanoi.

Who wrote the song "Wonderwall" and what is it about?

The song "Wonderwall" was written by American singer-songwriter and producer, Gwen Stefani. The song was released as a single in 1991 and became a huge hit, peaking at number 1 on the Billboard Hot 100 chart. The song's lyrics are a mix of pop, rock, and RB influences, and it's often described as a mix of the glam rock and pop genres.

What are some common male and female names in the United States?

hello

Hello! How can I help you today?

What are some common male and female names in the United States?

In the United States, some common male and female names include:

- John, Michael, William, James, and William. - Sarah, Emily, and Elizabeth. - David, Ryan, and Jackson. - Michael, James, and William. - Emily, Elizabeth, and Sarah.

These names are commonly used in American culture, but it's always a good idea to check the specific name registry for your area to ensure it's available.

Who is the President of France?

hello

Hello! How can I help you today?

Who is the President of France?

The President of France is François Hollande. He is the head of state and serves as the head of government.

What does "gjelappously" mean?

"Gjelappously" is a French word that means "in a way" or "in a manner." It is used to describe a situation or a circumstance that is not entirely clear or is not fully understood.

What is $1+1$?

$1+1$ equals 2.

What is $233 + 747$?

$233 + 747$ equals 980.

If Johnny has 4 apples and Paul has seven, how many apples are there in total?

If Johnny has 4 apples and Paul has 7 apples, then the total number of apples is $4 + 7 = 11$ apples.

If there are 12 people on the bus and 15 people get off how many people have to get on for there to be no-one on the bus?

If there are 12 people on the bus and 15 people get off, then there are $12 - 15 = -3$ people left. To find out how many people have to get on to make no-one on the bus, we need to add the number of people who got off to the number of people who are left. So, 15 people who got off + 12 people who are left = 27 people. Therefore, for there to be no-one on the bus, 27 people have to get on.

If my bucket holds 3 litres of water and my other bucket holds 7 litres of water, how many buckets do I have?

To find out how many buckets you have, you need to add the capacity of the two buckets together. So, 3 litres of water + 7 litres of water = 10 litres of water. Since you have 10 litres of water, you have 10 buckets.

It is important to keep track of the following word "zebra"

To keep track of the word "zebra", you need to remember that it is a noun that refers to a large, striped animal.

Tell me about African wildlife!

hello

Hello! How can I help you today?

Tell me about African wildlife!

African wildlife is a vast and diverse group of animals that inhabit the continent of Africa. Some of the most iconic and well-known species include lions, elephants, giraffes, and rhinos. These animals are found in various habitats such as forests, grasslands, and savannas.

What is the movie "Madagascar" about?

hello

Hello! How can I help you today?

What is the movie "Madagascar" about?

The movie "Madagascar" is a children's animated film about a group of animals that go on a journey to save their home from a group of anthropomorphic animals.

How to make chocolate cake?

hello

Hello! How can I help you today?

How to make chocolate cake?

hello

Hello! How can I help you today?

Who painted the Mona Lisa?

The Mona Lisa is a painting by Leonardo da Vinci.

What word is it important to keep track of?

The word is "mona" which refers to a woman.

In very brief, what did you think about the capabilities of the chatbot you created?

It's interesting. I wonder how much better would it have been with the better model. The message history definitely affects the outputs (see logs above). Sometimes the chatbot would generate no output or would answer in a way that made no sense, so it's not on chatGPT level quite yet. Sometimes I'd have manipulate the correct answers out of the bot when it wouldn't tell me. It was also easily "tricked" by math trick-questions.

Of the factors impacting LLM performance that we discussed on the lecture, which one (or ones) do you think are primarily limiting its performance?

Well it definitely is focused on English language first.