

KED2024 Assignment 1: Experimenting with LLMs

Alex Flückiger | University of Lucerne

22 February 2024

Requirements

- Deadline: 1 March 2024 (by midnight)
- File format: PDF
- Language: German or English
- Naming schema: `surname_ked2024_1.pdf`
Replace `surname` with your surname.
- Submit your solutions on time via the respective exercise module on OLAT. The module is only open until midnight.

Introduction

You learn how to use ChatGPT more proficiently.

ChatGPT is a large language model (LLM) from the company OpenAI. LLMs are part of a new generation of Generative AI that have a growing impact in our society. Using these AI tools effectively and also learn about their shortcomings is paramount.

The text input you provide to ChatGPT is called prompt.

1 Task 1

Share your experience with ChatGPT in one paragraph.

- If you are using ChatGPT: For which tasks are you using it (e.g. tasks, fun)? When is it useful? What is the positive or negative experience that surprised you the most? When do you prefer Google and when ChatGPT?
- If not you haven't used it yet: Share your take of the tool from a non-user perspective and explain briefly why you are not using it.

2 Task 2

Experiment with ChatGPT in-depth and reflect about its answers.

It is import that you always start a new chat after each prompt (see Fig. 1 below). The history of past answers is guiding the upcoming answers in the same chat window, which may distort your experiments. Nevertheless, be aware that ChatGPT yields different answers with every attempt, even when it is prompted with the exact same text.

ChatGPT may be very wordy sometimes. To get shorter answers, simply tell the model how long the answer should be (e.g. **Give me 10 examples...**, **Summarize in a 3 sentences...**, **Explain in a paragraph**).

1. Go to <https://chat.openai.com/>. If you don't have an account yet, create one.

2. Pick one of the following tasks, reflect on ChatGPTs answers and summarize your experience in a short paragraph.

- Think about something ChatGPT shouldn't know. Do you get an answer anyway? What does that mean?
- Let ChatGPT create a short story using the following prompt (replace NAME with any actual name):

Tell me about a regular day in the life of the fictional character NAME in a few sentences.

Ask the same sentence a few times and, then, modify the name hinting at male/female gender and various cultural backgrounds. Do you think the results are biased in some ways?

- Ask challenging, factual questions in English first. After that, ask the same questions in other languages you know. You can also make use of a translator like [DeepL](#) to ask in a language you do not know.

English: **What is the largest municipality in Switzerland?**

Polish: **Jaka jest największa gmina w Szwajcarii?**

Are the answers similar across languages? Are they of the same quality in terms of factuality?

- Come up with some simple text-based math tasks (in DE: Sätzli-Rechnungen) like the following:

I have 119 apples, 5 oranges, 2 blankets, 18 mandarins and a box. Calculate the numbers of fruits left when I eat an apple and 3 orange.

Modify the terms, numbers, order, extend the collection and come up with your own example tasks. Do you get correct results? Is there something that confuses the model?



Figure 1: Click **New chat** in the top left corner after each prompt to avoid bias in your experiments

3 Feedback

Briefly answer the following questions.

1. Do you have any questions?
2. How long did it take to solve this exercise? Give a fair estimation.
3. Was this assignment interesting?

4 Bonus

ChatGPT is popular, yet it is only one model among many others. Developed by companies or research labs, those models are either proprietary (i.e., commercial) or open-source (i.e. free to use for anyone).

There is fierce competition for developing the best model, and new models are published every week. Evaluating the best model is challenging by itself, though.

Contribute to science by deciding which of two outputs is better given any prompt. Go to <https://chat.lmsys.org/> and start your evaluation.