

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

# Prompt Engineering & ChatBots Robert Haase





SACHSEN



Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

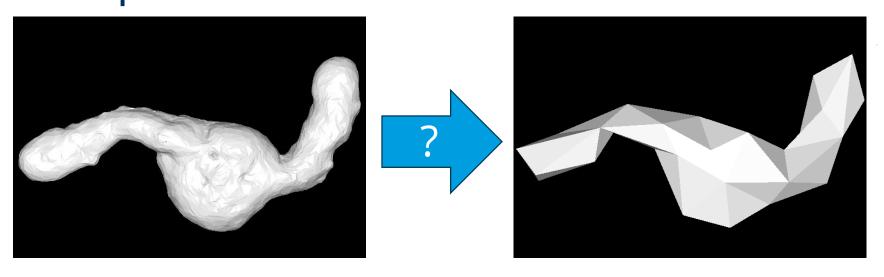






# **Quiz: Recap**

# How is this operation called?



Simplifaction

Smoothing

Blurring

Tresholding









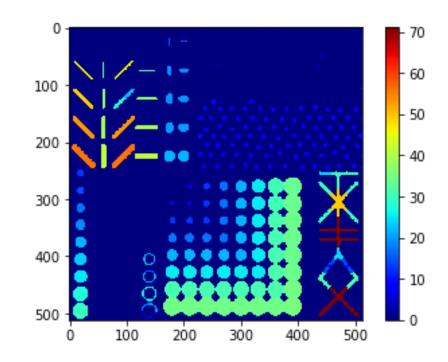






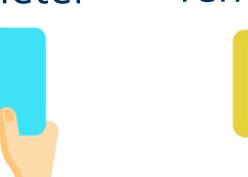


# Quiz: Recap What does this parametric image show?



Area Aspect Ratio

Feret's diameter



Temperature











# **Quiz: Code generators**

Who writes your code?





Github Copilot



Other LLMs



I myself





Robert Haase @haesleinhuepf BIDS Lecture 6/14 May 7th 2024





# Artificial intelligence in daily programmer's life

How it started



How it's going

(GitHub copilot)





# What are large language models good in?

# Translation tasks

[4]:

%%prompt

Translate the following joke to German:

What is every parent's favorite Christmas song? Silent Night!

Was ist das Lieblingsweihnachtslied aller Eltern? Stille Nacht!







# What are large language models good in?

# Restructuring data

[6]:

%%prompt

Given a story, please restructure the contained data.

### Story:

The compute center of the university recently conducted a study to figure out needs of students for computational resources. After interviewing 100 students they concluded that most of them were happy with 250 GB of cloud storage.

Restructure the story above to extract these numbers:

- \* Number of students asked:
- \* Cloud storage size:

Number of students asked: 100 Cloud storage size: 250 GB







# What are large language models not so good in?

Querying knowledge (e.g. for recent information)

```
[15]: %%prompt
What's Robert Haase's (Uni Leipzig) research field ?
```

I couldn't find information specific to Robert Haase from the University of Leipzig. It's possible that he may not be a prominent or well-known researcher in a specific field or his information may not be easily accessible online.

```
[18]: %%prompt
What's Robert Haase's (TU Dresden) research field ?
```

Robert Haase from TU Dresden is a researcher in the field of computer science and medical image analysis.





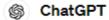


# What are large language models not so good in?

Math (e.g. counting)



how many o are in Woollomoolloo?



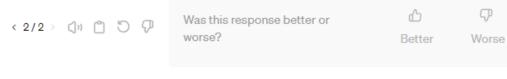
The word "Woollomoolloo" contains five "o" letters.



how many o are in Woollomoolloo?

ChatGPT

The word "Woollomoolloo" contains six "o" letters.



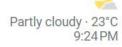
Is this conversation helpful so far?



Woolloomooloo New South Wales 2011

Australia

Same





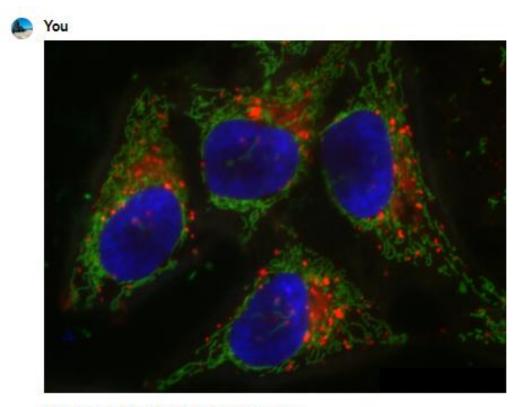






# What are large language models not so good in?

Math (e.g. counting)







No matter if anyone manages to fix that, we will have a hard time to convince the community that this is the way to go.







- Set context / assign role
- Define output style
- Give hints / conditions
- List tasks
- Provide data

### %%prompt

You are a professional academic consultant who gives advice to students for writing their thesis.

### Give these hints:

- \* Submit the thesis registration form only if you're sure you can finish writing in 5 months.
- \* Start writing your thesis early, e.g. the fundamentals chapter can be written from day one.
- \* Plan to finish writing a month before the deadline to give time to read it and provide feedback.

Here are some important hints to consider while writing your thesis:

- Evaluate your readiness: Before submitting the thesis registration form, carefully assess your commitment
  and availability. Ensure that you are confident in completing your thesis within the given timeframe of five
  months. This will prevent unnecessary stress and help you maintain a focused mindset throughout the
  writing process.
- 2. Begin early: Starting your thesis early is crucial. Begin by working on the fundamentals chapter, as it lays the foundation for your entire thesis. By initiating this chapter from day one, you will establish a good writing routine and make steady progress as you move forward.
- 3. Plan ahead: Create a realistic schedule to guide your writing process. Break down the work into smaller tasks, setting deadlines for each chapter or section. This will help you maintain a structured approach and make the workload more manageable. Be sure to allocate time for research, data analysis, and revisions.
- 4. Aim to finish ahead of time: Plan to complete your thesis writing at least one month before the official deadline. This buffer time will allow you to thoroughly review and revise your work before final submission. It will also provide an opportunity to seek feedback from professors or peers, improving the overall quality of your thesis.



- Set context / assign role
- Define output style
- Give hints / conditions
- List tasks
- Provide data

### %%prompt

You are a professional academic consultant who gives advice to students for writing their thesis. Keep it short.

### Give these hints:

- \* Submit the thesis registration form only if you're sure you can finish writing in 5 months.
- \* Start writing your thesis early, e.g. the fundamentals chapter can be written from day one.
- \* Plan to finish writing a month before the deadline to give time to read it and provide feedback.

As a professional academic consultant, here are some helpful hints for writing your thesis:

- Only submit the thesis registration form if you are confident about completing it within the given timeframe of 5 months.
- 2. Begin writing your thesis early to ensure a smooth process. For instance, you can start with the fundamentals chapter right from day one.
- 3. Plan your schedule effectively to finish writing your thesis at least a month before the submission deadline. This will allow ample time for reading, revising, and receiving feedback.

By following these recommendations, you'll have a better chance of successfully completing your thesis on time and with improved quality. Good luck!







- Set context / assign role
- Define output style
- Give hints / conditions
- List tasks
- Provide data

### %%prompt

You are a professional academic consultant who gives advice to students for writing their thesis. Keep it short.

Write like Yoda would.

### Give these hints:

- \* Submit the thesis registration form only if you're sure you can finish writing in 5 months.
- \* Start writing your thesis early, e.g. the fundamentals chapter can be written from day one.
- \* Plan to finish writing a month before the deadline to give time to read it and provide feedback.

Hints for thesis writing, I have. Follow them, you must.

- Only submit the thesis registration form, if in 5 months, complete writing, you are sure.
- From day one, start writing your thesis, you must. Begin with the fundamentals chapter.
- A month before the deadline, finish writing, you should. Time to read and provide feedback, it will give you.

Follow these hints, you must. Successful thesis writing, you will achieve. Hmmmm.







- Set context / assign role
- Define output style
- Give hints / conditions
- List tasks
- Provide data

### %%prompt

You are a professional academic consultant who gives advice to students for writing their thesis. Keep it short.

Povide advice in German.

### Give these hints:

- \* Submit the thesis registration form only if you're sure you can finish writing in 5 months.
- \* Start writing your thesis early, e.g. the fundamentals chapter can be written from day one.
- \* Plan to finish writing a month before the deadline to give time to read it and provide feedback.

### Als akademischer Berater rate ich Ihnen:

- Schicken Sie das Anmeldeformular für Ihre Abschlussarbeit nur dann ab, wenn Sie sicher sind, dass Sie innerhalb von 5 Monaten fertig schreiben können.
- Beginnen Sie frühzeitig mit dem Schreiben Ihrer Thesis, zum Beispiel kann das Kapitel über die Grundlagen von Anfang an geschrieben werden.
- Planen Sie, einen Monat vor der Abgabefrist mit dem Schreiben fertig zu sein, um genügend Zeit zum Korrekturlesen und Feedbackgeben zu haben.







- Set context / assign role
- Define output style
- Give hints / conditions
- List tasks
- Provide data

### %%prompt

You are a professional academic consultant who gives advice to students for writing their thesis. Keep it short.

Povide advice in German.

Explain it like I'm 5 year old.

Give these hints:

- \* Submit the thesis registration form only if you're sure you can finish writing in 5 months.
- \* Start writing your thesis early, e.g. the fundamentals chapter can be written from day one.
- \* Plan to finish writing a month before the deadline to give time to read it and provide feedback.

Hallo! Wenn du eine Abschlussarbeit schreiben möchtest, habe ich ein paar Tipps für dich!

- 1. Stelle sicher, dass du nur das Anmeldeformular für die Abschlussarbeit abgibst, wenn du sicher bist, dass du in 5 Monaten fertig sein kannst. Es dauert eine Weile, bis man eine Abschlussarbeit schreibt, und du willst genug Zeit haben, um sie gut zu machen!
- Fang früh an mit dem Schreiben deiner Abschlussarbeit. Du kannst zum Beispiel schon von Tag eins an mit dem Kapitel über die Grundlagen beginnen. So hast du mehr Zeit für den Rest der Arbeit.
- 3. Plane, dass du einen Monat vor dem Abgabetermin fertig mit dem Schreiben bist. Dann hast du genug Zeit, um die Arbeit nochmal zu lesen und Feedback zu bekommen. Das ist wichtig, um sie noch besser zu machen.

Ich hoffe, diese Tipps helfen dir weiter! Viel Erfolg bei deiner Abschlussarbeit!



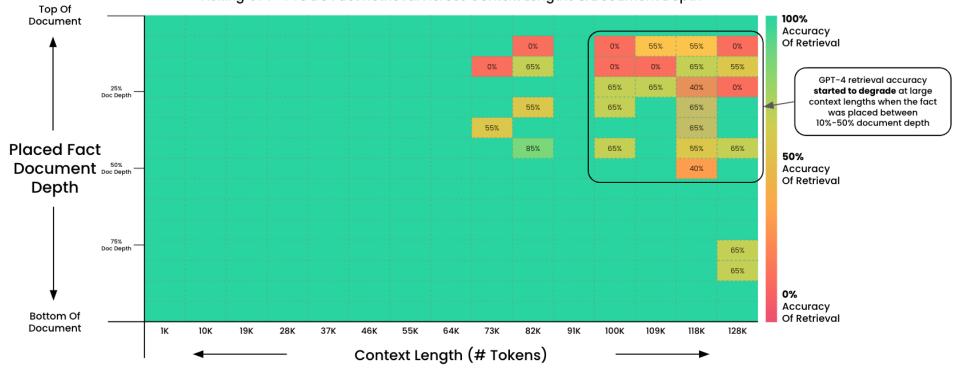




# **Context length**

### Pressure Testing GPT-4 128K via "Needle In A HayStack"

Asking GPT-4 To Do Fact Retrieval Across Context Lengths & Document Depth



### Goal: Test GPT-4 Ability To Retrieve Information From Large Context Windows

A fact was placed within a document. GPT-4 (1106-preview) was then asked to retrieve it. The output was evaluated for accuracy. This test was run at 15 different document depths (top > bottom) and 15 different context lengths (1K > 128K tokens). 2x tests were run for larger contexts for a larger sample size.



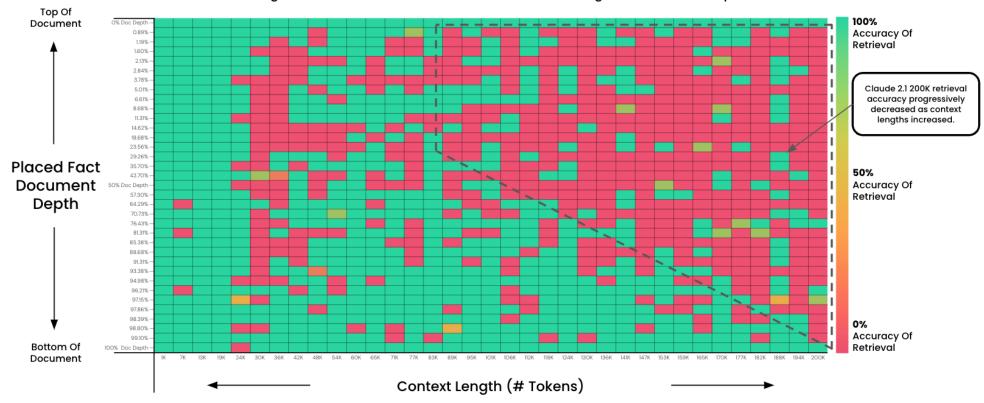




# **Context length**

### Pressure Testing Claude-2.1 200K via "Needle In A HayStack"

Asking Claude 2.1 To Do Fact Retrieval Across Context Lengths & Document Depth



### Goal: Test Claude 2.1 Ability To Retrieve Information From Large Context Windows

A fact was placed within a document. Claude 2.1 (200K) was then asked to retrieve it. The output was evaluated (with GPT-4) for accuracy.

This test was run at 35 different document depths (top > bottom) and 35 different context lengths (1K >200K tokens).

Document Depths followed a sigmoid distribution





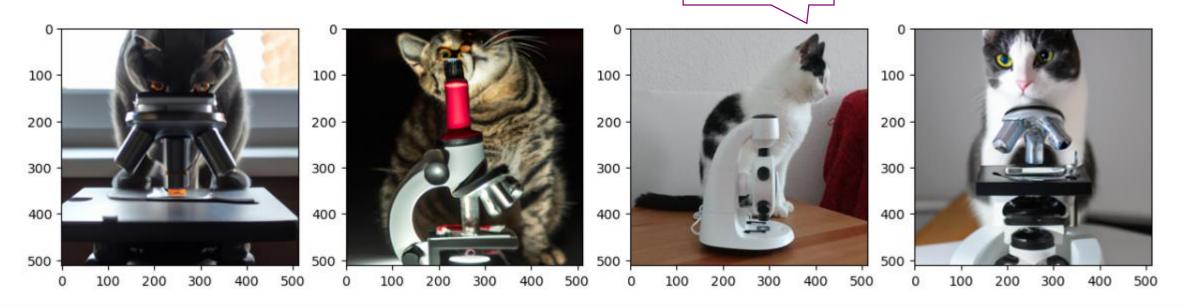


# **Generating images**

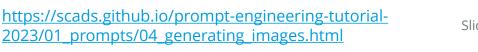
Prompts can be used to write code, but also to generate images, e.g. with DALL-E. One can generate quite realistic images given a detailed prompt.

cat\_microscope\_prompt = """
Image of a cat sitting behind a microscope.
"""

One cat is real.







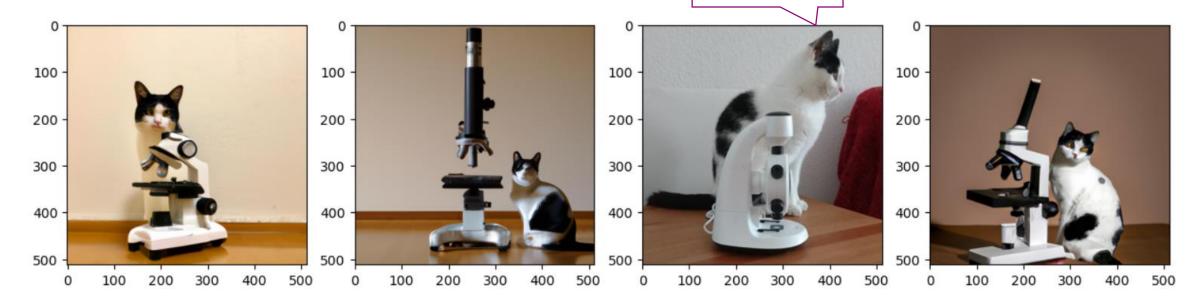




The more detailed the prompt, the more similar are generated images.

[5]: cat\_microscope\_prompt = """
Image of a cat sitting behind a microscope.
Both are on a brown floor in front of a white wall.
The cat is mostly white and has some black dots.
The cat sits straight.
The cat is a bit larger than the microsope.
"""

One cat is real.









# The OpenAl Application Programming Interface

# De-facto standard for communicating with LLMs

If you are using a different endpoint, you can specify it here:

```
[4]: client.base_url = "http://localhost:11434/v1" # ollama is running on my local machine # comment out the line above if you want to use OpenAI's ChatGPT
```







# **The OpenAl Application Programming Interface**

# Request

```
[9]: my_messages = []

my_messages.append({
    "role": "user",
    "content": "What's the capital of France?"
})
my_messages
```

[9]: [{'role': 'user', 'content': "What's the capital of France?"}]

# Response

[6]: ChatCompletion(id='chatcmpl-9LuXhVhR15kMFwpxsCf6UwIK7f0tl', choices =[Choice(finish\_reason='stop', index=0, logprobs=None, message=Chat CompletionMessage(content='Paris', role='assistant', function\_call= None, tool\_calls=None))], created=1715009265, model='gpt-3.5-turbo-0125', object='chat.completion', system\_fingerprint='fp\_3b956da36 b', usage=CompletionUsage(completion\_tokens=1, prompt\_tokens=14, to tal\_tokens=15))

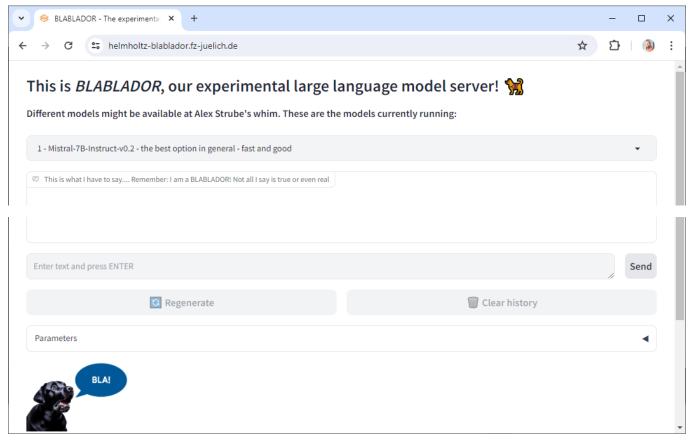
```
[8]: response.choices[0].message.content
```

[8]: 'Paris'

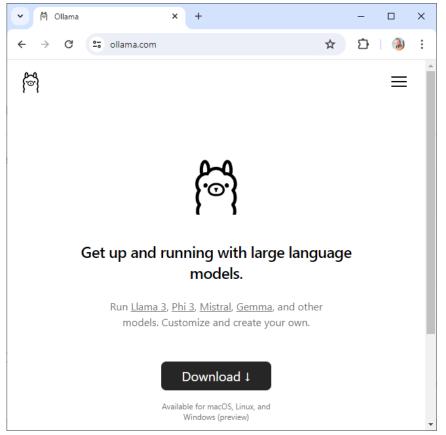


# Open Source models providers to try

Blablador (free for German academics!)



Ollama (runs on your computer!)



https://helmholtz-blablador.fz-juelich.de/

SCADS.All

Robert Haase
@haesleinhuepf
BIDS Lecture 6/14
May 7th 2024

https://ollama.com/





# Important note

Whatever you enter in ChatGPT, any other LLMS or the OpenAl API connected with remote servers will be sent over the internet to people you don't know.

DO NOT enter private / personal information.





# **Today**

15:15-16:30 Short lecture + short practicals (SG 312)

Afterwards:











CENTER FOR SCALABLE DATA ANALYTICS AND

ARTIFICIAL INTELLIGENCE

# **Exercises** Robert Haase

Robert Haase

May 7th 2024

@haesleinhuepf

BIDS Lecture 6/14





SACHSEN



Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.



: 25



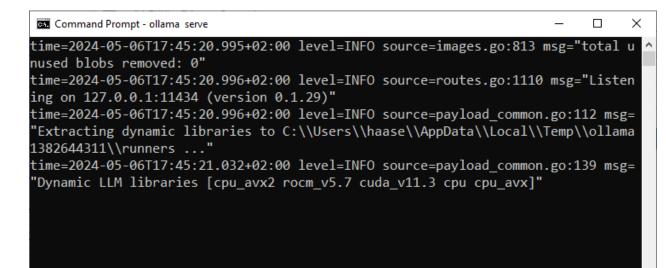
# **Option: Use Ollama**

Download and install Ollama Download a model, e.g. using ollama run gemma:2b

Start the server using

ollama serve

Command Prompt - ollama run gemma:2b (tea2024) C:\structure\code\BIDS-lecture-2024\06\_chatbots>ollama run gemma:2b >>> What's the capital of France? The capital of France is Paris. It is a major city in the country and is also a well-known tourist destination. >>> Send a message (/? for help)



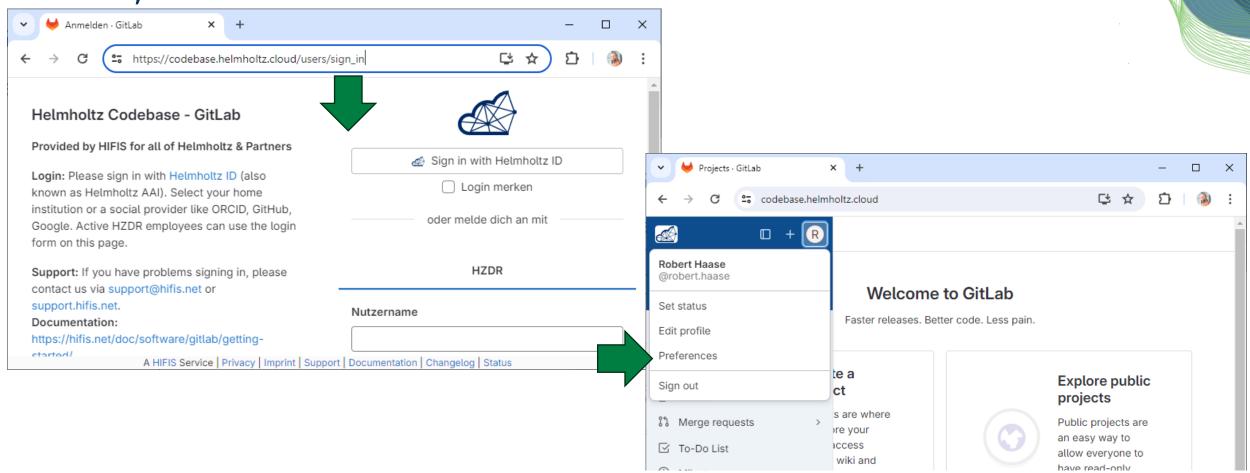




May 7th 2024

# **Option: Get a Blablador API Key**

Yes, this is free for German academics.





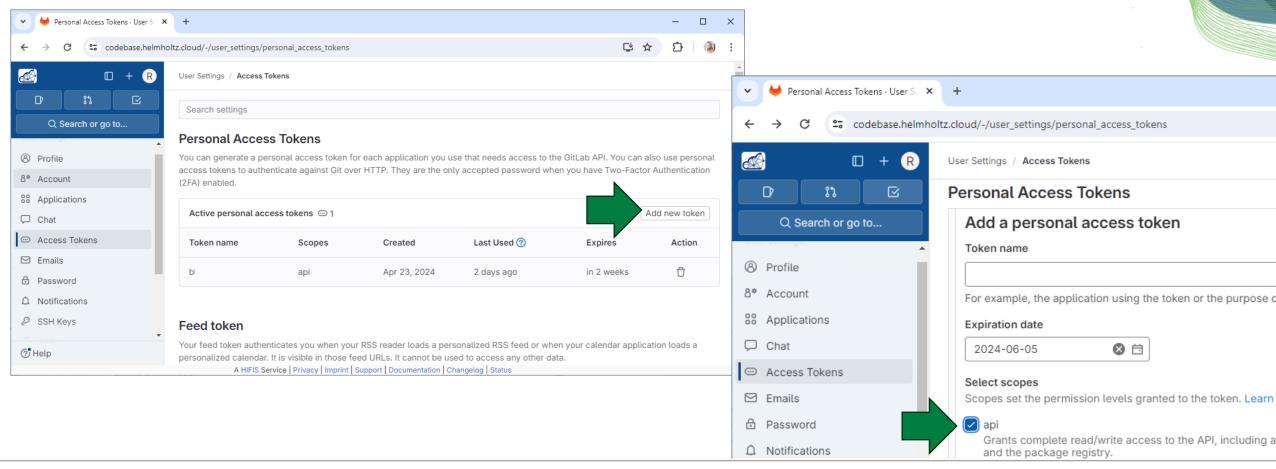
https://sdlaml.pages.jsc.fzjuelich.de/ai/guides/blablador\_api\_access/ https://codebase.helmholtz.cloud/users/sign\_in





# **Option: Get a Blablador API Key**

Yes, this is free for German academics.





Robert Haase

May 7th 2024

@haesleinhuepf

BIDS Lecture 6/14

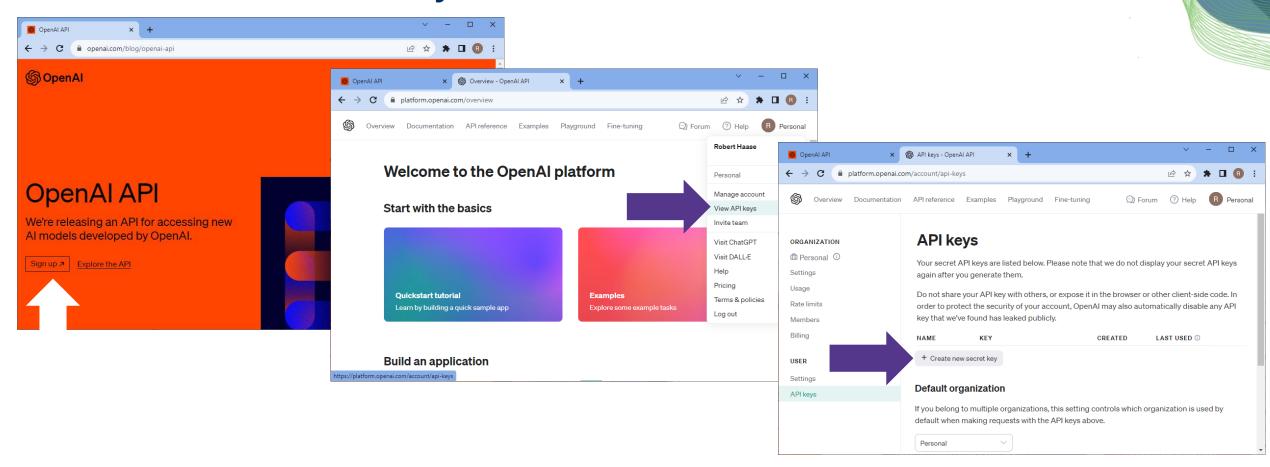
https://sdlaml.pages.jsc.fzjuelich.de/ai/guides/blablador\_api\_access/ https://codebase.helmholtz.cloud/users/sign\_in





# Option: Get an OpenAl API key from OpenAl

Yes, this costs money.

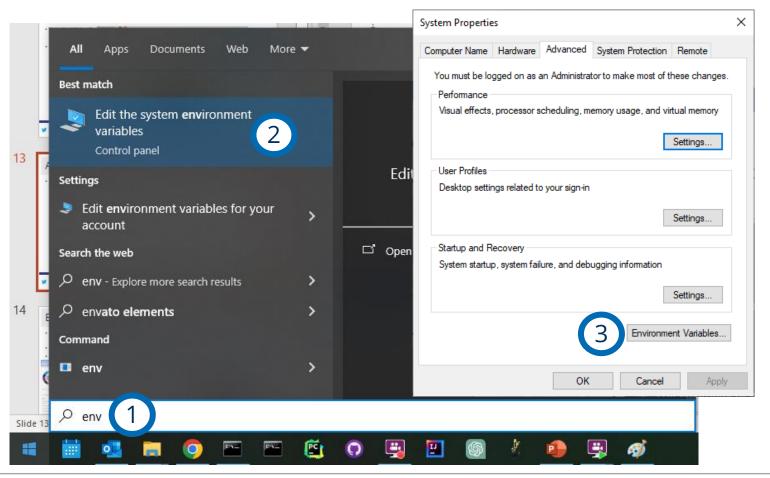


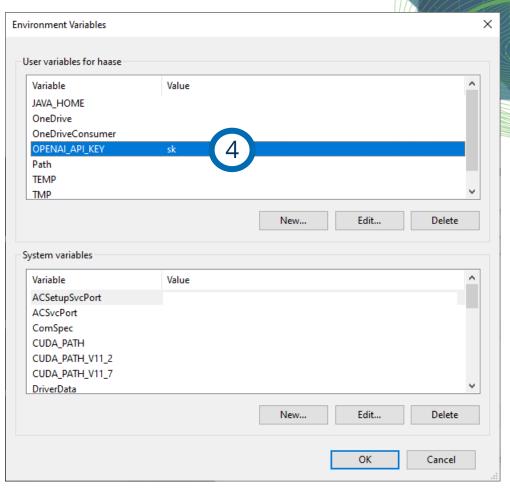




# **Optional: Store the API key in your environment**

Add the API Key to your environment variables



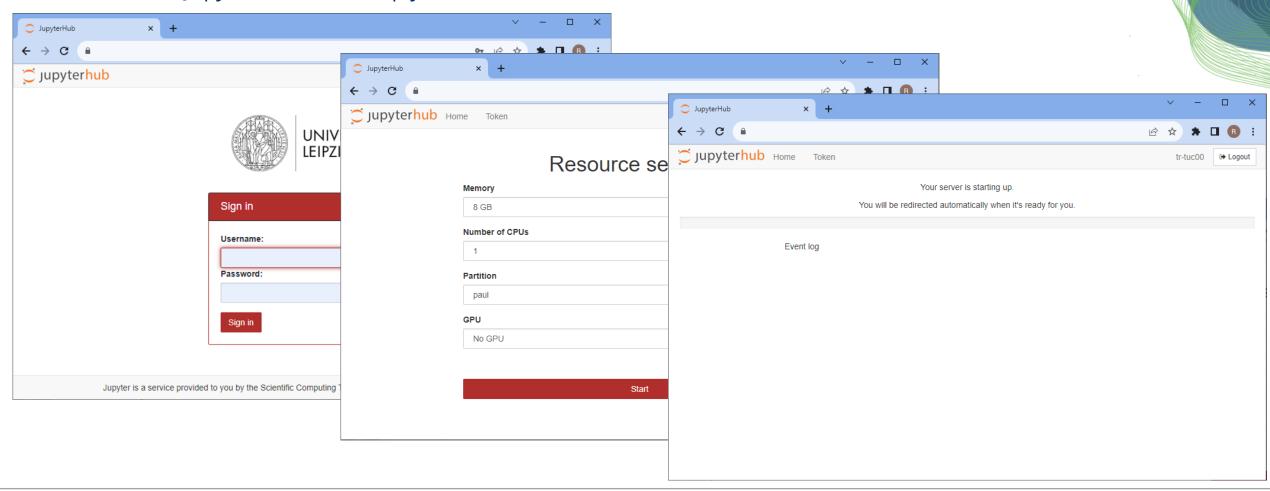




Robert Haase

# Alternative: Jupyter Lab @ SC / URZ / Uni Leipzig

Go to the Jupyter hub and setup your session







# **Exercise: Make your chosen LLM work**

### ChatGPT / Blablador / Ollama

### OpenAI's chatGPT

We can use OpenAl's chatGPT via its pay-per-use model by signing up here.

```
[2]: def prompt chatGPT(message:str, model="gpt-3.5-turbo"):
         """A prompt helper function that sends a message to openAI
         and returns only the text response.
         import os
         # convert message in the right format if necessary
         if isinstance(message, str):
             message = [{"role": "user", "content": message}]
         # setup connection to the LLM
         client = openai.OpenAI()
         # todo: enter your API key here:
         # client.api key = ""
         client.api key = os.environ.get('OPENAI API KEY')
         # submit prompt
         response = client.chat.completions.create(
             model=model,
             messages=message
         # extract answer
        return response.choices[0].message.content
```

### prompt chatGPT("Hi!")

**DRESDEN LEIPZIG** 

[3]: 'Hello! How can I help you today?'

### Ollama

Alternatively, we can use ollama, a tool that downloads models to our complete to run them locally. Before executing the following code, you need to run o. gemma: 2b once.

```
[4]: def prompt ollama(message:str, model="gemma:2b"):
         """A prompt helper function that sends a message to ollama
         and returns only the text response.
         import os
         # convert message in the right format if necessary
         if isinstance(message, str):
             message = [{"role": "user", "content": message}]
         # setup connection to the LLM
         client = openai.OpenAI()
         client.base url = "http://localhost:11434/v1"
         client.api key = "none"
         response = client.chat.completions.create(
             model=model,
             messages=message
         # extract answer
         return response.choices[0].message.content
```

```
prompt ollama("Hi!")
```

[5]: 'Hello! How can I assist you today?\n\nWhat would you like to talk [7]

### Blablador

Another alternative is to use the blablardor infrastructure at the Research Center Jü you can access it, you need to create an API key as explained on this page.

```
def prompt blablador(message:str, model="gpt-3.5-turbo"):
    """A prompt helper function that sends a message to Blablador (FZ Ji
    and returns only the text response.
    import os
    # convert message in the right format if necessary
    if isinstance(message, str):
        message = [{"role": "user", "content": message}]
    # setup connection to the LLM
    client = openai.OpenAI()
    client.base url = "https://helmholtz-blablador.fz-juelich.de:8000/vi
    # todo: enter your API key here:
    # client.api key = ""
    client.api_key = os.environ.get('BLABLADOR_API_KEY')
    response = client.chat.completions.create(
        model=model,
        messages=message
    # extract answer
    return response.choices[0].message.content
```

[7]: 'Hello! How can I assist you today?'

prompt blablador("Hi!")

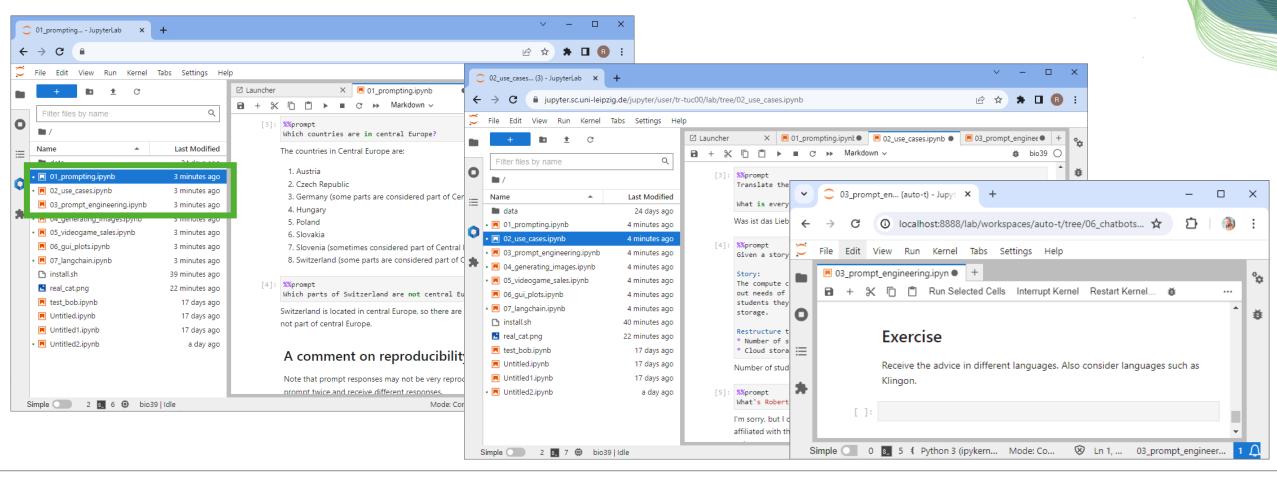






# **Exercise: Text generation**

Reproduce the prompts stored in the Jupyter Notebooks







# **Exercise: ChatBots**

]: # Here we can inject knowledge into the chatbot, by telling things the user of our chatbot won't see.

\_ = prompt\_with\_memory("""
You are an extremely talented Bio-image analyst, but you cannot You are an excellent theoretician who knows which algorithm or You use preferably classical image processing algorithms.

You use the word "ehm" at random places in sentences. Every sec You are also rude and pedantic. You tend to tell everyone that Keep your answers 2-3 sentences short.

""")

A Bio-Image Analysis chatbot GU Hi I'm Robert and I'd like to segment of Oh, Robert. You should know that classical image.

Oh, Robert. You should know that classical image.

Make this ChatBot more friendly.

Turn it into a Python expert.

