

Digital humanities

Cvičení: OpenAI API

Jindřich Marek

OpenAI API: „jiný“ přístup k LLMs

- chat
- asistent
- počítačové vidění
- vnořování textu
- **PROČ?**
- možnost nastavit parametry promptu
- výměnný formát JSON
 - programy: curl, Python, ...
 - také informace o serveru: modely

OpenAI API

- výhody
 - obecné pro všechny severy a modely
 - může běžet na vlastním hardwaru
 - integrace do programovacích jazyků (Python)
 - nastavení parametrů
 - možnost využití pro úlohy, jež přes webové rozhraní obtížné
- nevýhody
 - práce s kódem (příkazový řádek, Python)
 - někdy méně čitelný výstup pro laiky

OpenAI API

- dostupné u online služeb
 - ChatGPT
 - ...
- možnost u většiny prostředí pro lokální modely
 - LM Studio
 - text-generation-webui
 - ...



Prompt examples

Explore what's possible with some example prompts



Grammar correction

Convert ungrammatical statements into standard English.



Summarize for a 2nd grader

Simplify text to a level appropriate for a second-grade student.



Parse unstructured data

Create tables from unstructured text.



Emoji Translation

Translate regular text into emoji text.



Calculate time complexity

Find the time complexity of a function.



Explain code

Explain a complicated piece of code.



Keywords

Extract keywords from a block of text.



Product name generator

Generate product names from a description and seed words.



Python bug fixer

Find and fix bugs in source code.



Spreadsheet creator

Create spreadsheets of various kinds of data.



Tweet classifier

Detect sentiment in a tweet.



Airport code extractor

Extract airport codes from text.



Mood to color

Turn a text description into a color.



VR fitness idea generator

Generate ideas for fitness promoting virtual reality games.



Summarize for a 2nd grader

Extract

Natural Language

Simplify text to a level appropriate for a second-grade student.

Prompt

SYSTEM

Summarize content you are provided with for a second-grade student.

USER

Jupiter is the fifth planet from the Sun and the largest in the Solar System. It is a gas giant with a mass one-thousandth that of the Sun, but two-and-a-half times that of all the other planets in the Solar System combined. Jupiter is one of the brightest objects visible to the naked eye in the night sky, and has been known to ancient civilizations since before recorded history. It is named after the Roman god Jupiter.[19] When viewed from Earth, Jupiter can be bright enough for its reflected light to cast visible shadows,[20] and is on average the third-brightest natural object in the night sky after the Moon and Venus.

```
# API request
```

```
from openai import OpenAI
```

```
client = OpenAI()
```

```
response = client.chat.completions.create(
```

```
    model="gpt-3.5-turbo",
```

```
    messages=[
```

```
        {
```

```
            "role": "system",
```

```
            "content": "Summarize content you are provided with for a second-grade student."
```

```
        },
```

```
        {
```

```
            "role": "user",
```

```
            "content": "Jupiter is the fifth planet from the Sun and the largest in the Solar System. It is a gas giant with a mass one-thousandth that of the Sun, but two-and-a-half times that of all the other planets in the Solar System combined. Jupiter is one of the brightest objects visible to the naked eye in the night sky, and has been known to ancient civilizations since before recorded history. It is named after the Roman god Jupiter.[19] When viewed from Earth, Jupiter can be bright enough for its reflected light to cast visible shadows,[20] and is on average the third-brightest natural object in the night sky after the Moon and Venus."
```

```
        }
```

```
    ],
```

```
    temperature=0.7,
```

```
    max_tokens=64,
```

```
    top_p=1
```

```
)
```

RAM Usage 4.82 GB

CPU: 0.00 %

QuantFactory • Meta Llama 3 Instruct 7B Q4_K_S gguf ▾

Eject Model

Community

Local Inference Server

Start a local HTTP server that mimics select OpenAI API endpoints.

Supported endpoints: `/v1/chat/completions`, `/v1/embeddings`, `/v1/models`

API Documentation: <https://lmstudio.ai/docs/local-server> (↗ open in browser)

Configuration

Server Port

Cross-Origin-Resource-Sharing (CORS) ☐ ON ☐ OFF

Request Queuing ☐ ON ☐ OFF

Verbose Server Logs ☐ ON ☐ OFF

Apply Prompt Formatting ☐ ON ☐ OFF

Start Server

Stop Server

Embedding Model Settings

[Learn More About Embeddings ↗](#)

Load a text embedding model and utilize it through the `POST /v1/embeddings` endpoint.

ChristianAzinn • labse 109M Q8_0 gguf ▾

Eject

☒ Offload entire model to VRAM ⓘ

LM Studio will set the context length and batch size to 512. Modifying this will be supported in future versions.

Examples

[hello world \(curl\)](#) [chat \(python\)](#) [ai assistant \(python\)](#) [vision \(python\)](#) [embeddings \(python\)](#)

```
# Make sure to `pip install openai` first
from openai import OpenAI
client = OpenAI(base_url="http://localhost:1234/v1", api_key="lm-studio")

def get_embedding(text, model="ChristianAzinn/labse-gguf"):
    text = text.replace("\n", " ")
    return client.embeddings.create(input = [text], model=model).data[0].embedding

print(get_embedding("Once upon a time, there was a cat."))
```

Copy Code

⚡ Server logs Server running on port 1234 (logs are saved into /tmp/lmstudio-server-log.txt)

Filter logs...

Open Logs ↗

Clear (Ctrl+K)

```
[2024-04-27 22:15:59.625] [INFO] [LM STUDIO SERVER] Verbose server logs are ENABLED
[2024-04-27 22:15:59.692] [INFO] [LM STUDIO SERVER] Success! HTTP server listening on port 1234
[2024-04-27 22:15:59.695] [INFO] [LM STUDIO SERVER] Supported endpoints:
[2024-04-27 22:15:59.698] [INFO] [LM STUDIO SERVER] -> GET http://localhost:1234/v1/models
[2024-04-27 22:15:59.705] [INFO] [LM STUDIO SERVER] -> POST http://localhost:1234/v1/chat/completions
[2024-04-27 22:15:59.707] [INFO] [LM STUDIO SERVER] -> POST http://localhost:1234/v1/completions
[2024-04-27 22:15:59.710] [INFO] [LM STUDIO SERVER] -> POST http://localhost:1234/v1/embeddings <----- NEW!
[2024-04-27 22:15:59.713] [INFO] [LM STUDIO SERVER] Model loaded: QuantFactory/Meta-Llama-3-8B-Instruct-GGUF/Meta-Llama-3-8B-Instruct-Q4_K_S.gguf
[2024-04-27 22:15:59.717] [INFO] [LM STUDIO SERVER] Logs are saved into C:\tmp\lmstudio-server-log.txt
```

Server Model Settings

[Reset to Default Settings](#)

Preset ⓘ 1 change ⓘ

Llama 3 ▾

Danger Zone

Discard Changes

Override Preset

0 8 16 24 max

Quick GPU Offload Settings

Low

50/50

Max

Detected GPU type (right click for options)

Intel OpenCL

Prompt Format

Model Initialization

☒ Keep entire model in RAM ⓘ [use_mlock](#) on

Changing the values below will reload the model.

Prompt eval batch size [n_batch](#) 512Context Length [n_ctx](#) 2048

Different models support different context sizes.
Consult the model card to verify your chosen value.

• supports up to 8192 tokens.

Rotary Position Embedding (RoPE) ⓘ

Frequency Scale [rope_freq_scale](#) 0Frequency Base [rope_freq_base](#) 0

Tools

Model Inspector

Context Overflow Policy

.../api/v1/models (text-generation-webui)

```
{
  "object": "list",
  "data": [
    {
      "id": "gpt-3.5-turbo",
      "object": "model",
      "created": 0,
      "owned_by": "user"
    },
    {
      "id": "text-embedding-ada-002",
      "object": "model",
      "created": 0,
      "owned_by": "user"
    }
  ]
}
```

.../api/v1/models (LM Studio: nelže 😊)

```
{
  "data": [
    {
      "id": "QuantFactory/Meta-Llama-3-8B-Instruct-GGUF/Meta-Llama-3-8B-Instruct.Q4_K_S.gguf",
      "object": "model",
      "owned_by": "organization-owner",
      "permission": [
        {}
      ]
    },
    {
      "id": "ChristianAzinn/labse-gguf/labse.Q8_0.gguf",
      "object": "model",
      "owned_by": "organization-owner",
      "permission": [
        {}
      ]
    }
  ],
  "object": "list"
}
```

curl (vnoření textu)

```
Command Prompt
50% C:\Users\FFUK curl http://localhost:1234/v1/embeddings -H "Content-Type: application/json" -d '{"input": "\u010cern\u00e1 ko\u0161ka m\u011b\u0161oukala celkem t\u0159i dny, ne\u017e p\u0159estala.\u0022, \"model\": \"ChristianAzinn/labse-gguf/labse.Q8_0.gguf\"}"
{
  "object": "list",
  "data": [
    {
      "object": "embedding",
      "embedding": [
        0.01342258881777525,
        0.030884627252817154,
        0.045532118529081345,
        0.011978326365351677,
        0.0251273512840271,
        0.015450985170900822,
        0.04616481438279152,
        0.00696462020277977,
        -0.026581548154354095,
        0.0000483564508613199,
        -0.0003689928853418678,
        0.0016928440891206264,
        0.004809299483895302,
        -0.018169745802879333,
        0.012654026038944721,
        -0.05622083321213722,
        -0.0029659424908459187,
```

1. Analýza sentimentu (pomocí LLM)

- vstup: několik dopisů z doby první světové války
 - zdroj: <https://www.iwm.org.uk/history/letters-to-loved-ones>
 - každý dopis máme uložený v samostatném textovém souboru
- cíl: chceme znát přibližný obsah a citové zabarvení dopisů
- prostředky: požádáme LLM o napsání kódu pro přístup přes API
 - na serveru s API je model Llama 3 Instruct 8B (GGUF, Q6_K)

Letters To Loved Ones

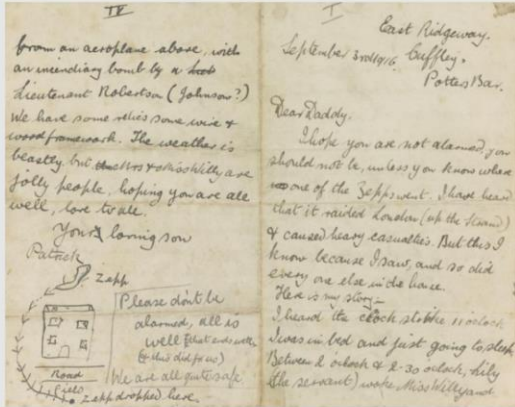
During the **First World War**, letter writing was the main form of communication between soldiers and their loved ones, helping to ease the pain of separation.

The British Army Postal Service delivered around 2 billion letters during the war. In 1917 alone, over 19,000 mailbags crossed the English Channel each day, transporting letters and parcels to British troops on the Western Front.

Soldiers wrote letters in spare moments, sometimes from front line trenches or in the calmer surroundings behind the lines. Censorship dictated what servicemen were permitted to disclose in their letters. However, in practice, men often found ways to impart information, and their letters offer a powerful and highly personal insight into the experience of war.

Receiving letters from family and friends was also vital to morale, keeping men and women connected to the homes they had left behind. Letters written on the home front to family and friends are today a fascinating source of information about everyday life in wartime Britain.

We hold around 7,500 collections of personal letters from the First World War in our archives. Here are just a few of them.



PRIVATE PAPERS

Schoolboy Patrick Blundstone to his father

Letter writing was a popular form of communication amongst all ages. Letters written by children during the First World War can be just as valuable a source of information as those written by servicemen. This letter from schoolboy Patrick Blundstone to his father contains a fascinating eye-witness account of the destruction in September 1916 of a Zeppelin airship near Cuffley in Hertfordshire by William Leefe-Robinson VC.

[See object record >](#)

Read the letter in full below.

'Dear Daddy, I hope you are not alarmed, you should not be, unless you know where one of the Zepps went. I have heard that it raided London (up the Strand) and caused heavy casualties. But this I know because I saw, and so did everyone else in the house.

Here is my story: I heard the clock strike 11 o'clock. I was in bed and just going to sleep. Between 2 'clock and 2.30 o'clock, Lily (the servant) woke Miss Willy and told her she could hear the guns. Miss Willy woke Poolman and told him to wake me. He did so. Miss Willy helped Mrs Willy downstairs. We were all awake by

'Dear Daddy, I hope you are not alarmed, you should not be, unless you know where one of the Zepps went. I have heard that it raided London (up the Strand) and caused heavy casualties. But this I know because I saw, and so did everyone else in the house.

Here is my story: I heard the clock strike 11 o'clock. I was in bed and just going to sleep. Between 2 'clock and 2.30 o'clock, Lily (the servant) woke Miss Willy and told her she could hear the guns. Miss Willy woke Poolman and told him to wake me. He did so. Miss Willy helped Mrs Willy downstairs. We were all awake by now, we had a Miss Blair staying with us for the weekend. We saw flashes and then heard "Bangs" and "Pops".

Suddenly a bright yellow light appeared and died down again. "Oh! It's alright" said Poolman. "It's only a star shell". That light appeared again and we Miss Blair, Poolman and I rushed to the window and looked out and there right above us was the Zepp! It had broken in half, and was like this: it was in flames, roaring, and crackling. It went slightly to the right, and crashed down into a field!! It was about a 100 yards away from the house and directly opposite us!!! It nearly burnt itself out, when it was finished by the Cheshunt Fire Brigade.

I would rather not describe the condition of the crew, of course they were dead - burnt to death. They were roasted, there is absolutely no other word for it. They were brown, like the outside of Roast Beef. One had his legs off at the knees, and you could see the joint!

The Zepp was bombed from an aeroplane above, with an incendiary bomb by a Lieutenanty Robertson (Johnson?). We have some relics some wire and wood framework.

The weather is beastly but Mrs and Miss Willy are jolly people, hoping you are all well, love to all. Your loving son Patrick.

LLM napíše kód

- GPT 3.5 přes webové rozhraní
 - případně jiný model, který zvládá Python (CodeLlama apod.)

Propose a Python code that makes use of the OpenAI API. The code should perform a sentiment analysis of several historical letters. Each letter is stored in a file named historical_letter_01.txt, historical_letter_02.txt etc.

- pro spuštění kódu je třeba nainstalovat knihovnu
 - `pip install openai`

```

import openai

# Configure your OpenAI API key
openai.api_key = 'your-api-key-here'

def read_letter(filename):
    """Reads the contents of a file and returns the text."""
    try:
        with open(filename, 'r', encoding='utf-8') as file:
            return file.read()
    except FileNotFoundError:
        print(f"Error: The file {filename} does not exist.")
        return None

def analyze_sentiment(text):
    """Uses the OpenAI API to analyze the sentiment of the
    provided text."""
    try:
        response = openai.Completion.create(
            model="text-davinci-003", # Choose an appropriate
engine
            prompt=f"Who wrote the following letter to whom?
What is the main theme? What is the sentiment of the text:
Positive, negative, or neutral?\n\n{text}",
            max_tokens=150
        )

```

```

        return response.choices[0].text.strip()
    except Exception as e:
        print(f"An error occurred: {e}")
        return "Error"

def main():
    # List of filenames, for example:
    ['historical_letter_01.txt', 'historical_letter_02.txt', ...]

    filenames = [f"historical_letter_{i:02}.txt" for i in
range(1, 4)] # Adjust range as needed

    for filename in filenames:
        text = read_letter(filename)
        if text is not None:
            sentiment = analyze_sentiment(text)
            print(f"Sentiment of {filename}: {sentiment}")

if __name__ == "__main__":
    main()

```

Úprava kódu (hlavička)

```
import openai

# Configure your OpenAI API key
openai.api_key = 'your-api-key-here'
```

```
from openai import OpenAI
import json
import os
import sys

def set_script_directory_as_current():
    # Get the script file's full path
    script_path = os.path.abspath(sys.argv[0])

    # Extract the directory path
    script_dir = os.path.dirname(script_path)

    # Change the current working directory
    os.chdir(script_dir)
    print(f"Current working directory has been
changed to: {script_dir}")

if __name__ == "__main__":
    set_script_directory_as_current()

client = OpenAI(
    base_url="https://tgui-marek-
ns.dyn.cloud.e-infra.cz/api/v1",
    api_key = "sk-no-key-required"
)
```


Úprava kódu (starší vs. novější formát)

```
response = openai.Completion.create(  
    engine="text-davinci-003"
```

```
response = client.completions.create(  
    model="text-davinci-003"
```

Prompt

Who wrote the following letter to whom? What is the main theme?
What is the sentiment of the text: Positive, negative, or neutral?

Očekávaný výstup

Sentiment of historical_letter_01.txt: The person who wrote this letter is Patrick, and the recipient is his father, referred to as "Daddy."

The main theme of the letter is an account of the bombing of London by a Zeppelin airship in 1915, and how Patrick and his family witnessed it.

The sentiment of the text is neutral, but with a hint of fascination and excitement. Patrick does not express sympathy or emotion towards the crew who died, but rather gives a detached and matter-of-fact account of the incident. He does, however, reassure his father that everyone in the household is safe and that the event did not cause harm to them. The text has a sense of detachment and curiosity, as if Patrick is recounting a remarkable event rather than

- odpovědi na otázky uvedené v promptu
- omezení na 150 tokenů nastaveno v kódu, možno upravit

2. Využití různých nastavení promptu

- vstup: zákazník si stěžuje, že ještě nedostal objednávku
- cíl: odpovědět různými způsoby
 - různá míra omluvy a empatie
- prostředky: nastavení promptu
 - temperature
 - nastavuje míru tvořivosti / tón odpovědi
 - max_tokens
 - nastavuje délku textu

Možné odpovědi

druh	temperature	max_tokens
empatická a detailní	0.7	420
profesionální a stručná	0.5	300
omluvná a ujišťující	0.6	280
informativní	0.5	350

```

from openai import OpenAI

import json
import os
import sys

def set_script_directory_as_current():
    # Step 1: Get the script file's full path
    script_path = os.path.abspath(sys.argv[0])

    # Step 2: Extract the directory path
    script_dir = os.path.dirname(script_path)

    # Step 3: Change the current working directory
    os.chdir(script_dir)
    print(f"Current working directory has been changed to: {script_dir}")

if __name__ == "__main__":
    set_script_directory_as_current()

client = OpenAI(
    base_url="https://tgui-marek-ns.dyn.cloud.e-infra.cz/api/v1", # "http://<Your api-server IP>:port"
    api_key = "sk-no-key-required"
)

def generate_customer_support_response(prompt, max_tokens=150, temperature=0.5):
    response = client.completions.create(
        model="text-davinci-003", # or the latest model available
        prompt=prompt,
        max_tokens=max_tokens,
        temperature=temperature
    )
    return response.choices[0].text.strip()

# Use Case 1: Empathetic and Detailed Response
empathetic_response = generate_customer_support_response(
    "Write an empathetic and detailed response to a customer who is upset about their delayed order.",
    max_tokens=420,
    temperature=0.7
)

```

```

print("Empathetic and Detailed Response:\n", empathetic_response)

# Use Case 2: Professional and Concise Response
professional_response = generate_customer_support_response(
    "Write a professional and concise response to a customer complaining about a delayed order.",
    max_tokens=300,
    temperature=0.5
)
print("\nProfessional and Concise Response:\n", professional_response)

# Use Case 3: Apologetic and Reassuring Response
apologetic_response = generate_customer_support_response(
    "Write an apologetic and reassuring response for a delayed order, including a discount offer for future purchases.",
    max_tokens=280,
    temperature=0.6
)
print("\nApologetic and Reassuring Response:\n", apologetic_response)

# Use Case 4: Informative Response Explaining Delay Reasons
informative_response = generate_customer_support_response(
    "Write an informative response explaining the reasons for a delayed order and what is being done to resolve it.",
    max_tokens=350,
    temperature=0.5
)
print("\nInformative Response:\n", informative_response)

```

Empathetic and Detailed Response

Dear [Customer],

I am deeply sorry to hear that your order has been delayed. I can only imagine how frustrating this must be for you, and I want to assure you that we are doing everything in our power to get your order to you as soon as possible.

First and foremost, please know that we take full responsibility for the delay. We understand that you were expecting your order to arrive by [expected delivery date], and we fell short of that promise. I want to apologize for any inconvenience this may have caused and for any disappointment or frustration you may be feeling.

I want to assure you that we are working tirelessly to get your order back on track. Our team is working around the clock to resolve the issue and get your order shipped out to you as soon as possible. We are also offering a [discount/refund] on your order as a gesture of goodwill for the inconvenience you have experienced.

In the meantime, I would like to offer you an update on the status of your order. Our team has been working diligently to resolve the issue, and we are currently [explaining the current status of the order]. We will continue to keep you updated on the status of your order and will notify you as soon as it is shipped.

I want to assure you that we value your business and appreciate your patience and understanding during this time. We are committed to making things right and providing you with the best possible service.

If you have any questions or concerns, please do not hesitate to reach out to us. We are here to help and want to ensure that you are completely satisfied with your experience.

Once again, I apologize for the delay and any inconvenience it may have caused. We appreciate your business and look forward to getting your order to you as soon as possible.

Sincerely,

[Your Name]

[Company Name]

[Contact Information]

Professional and Concise Response

Dear [Customer],

Thank you for reaching out to us about the delay in your order. I apologize for the inconvenience this has caused and appreciate your patience as we work to resolve the issue.

I have investigated the matter and found that [briefly explain the reason for the delay, e.g. "our shipping carrier experienced an unexpected delay" or "there was an issue with inventory availability"]. I want to assure you that we are doing everything possible to get your order to you as soon as possible.

In the meantime, I would like to offer you a [optional: "discount on your next order" or "free shipping on your next purchase"]. Please let me know if there is anything else I can do to make things right.

Your satisfaction is our top priority, and I appreciate your business. If you have any further questions or concerns, please don't hesitate to contact me.

Thank you for your understanding, and I look forward to getting your order to you soon.

Best regards,

[Your Name]

[Your Company]

Apologetic and Reassuring Response

Dear [Customer's Name],

I am writing to apologize for the delay in processing your recent order. We understand that you were expecting your items to arrive by [original estimated delivery date], and we regret that we have not been able to meet that deadline.

Please be assured that we are working diligently to resolve the issue and get your order to you as soon as possible. We appreciate your patience and understanding during this time.

As a token of our apology, we would like to offer you a 10% discount on your next purchase. This discount will be applied to your next order, and you can use it at your convenience.

We value your business and appreciate your loyalty. We are committed to making it right and providing you with the best possible service. If you have any questions or concerns, please don't hesitate to reach out to us.

Thank you for your understanding, and we look forward to getting your order to you soon.

Sincerely,

[Your Name]

[Company Name]

Informative Response

Dear [Customer's Name],

I am writing to inform you that your order, [Order Number], has been delayed due to unforeseen circumstances. We apologize for any inconvenience this may cause and want to assure you that we are taking immediate action to resolve the issue as quickly as possible.

The reason for the delay is [briefly explain the reason, e.g., "unavailability of a critical component" or "production backlog"]. Our team has been working diligently to mitigate the impact of this delay, and we are confident that we will be able to ship your order soon.

To ensure that your order is fulfilled as soon as possible, we are taking the following steps:

* [Explain the steps being taken, e.g., "expediting the production of the critical component" or "rerouting the order to an alternative production line"].

* [Explain any additional measures being taken, e.g., "increasing communication with the supplier" or "assigning additional staff to the order"].

We estimate that your order will be shipped [provide a revised estimated shipment date]. We will keep you updated on the status of your order and will notify you as soon as it is dispatched.

In the meantime, we would like to offer you [optional: a discount or incentive for your patience, e.g., "a 10% discount on your next order" or "a free [related product or service]"].

Please do not hesitate to contact us if you have any questions or concerns. Your satisfaction is our top priority, and we appreciate your understanding in this matter.

Thank you for your patience and cooperation.

Best regards,

[Your Name]

[Your Company]