Generative AI for Risk and Reliability

Lect 2: How GPT works

Zhiguo Zeng, Professor,

Chaire on Risk and Resilience of Complex Systems,

CentraleSupélec, Université Paris-Saclay

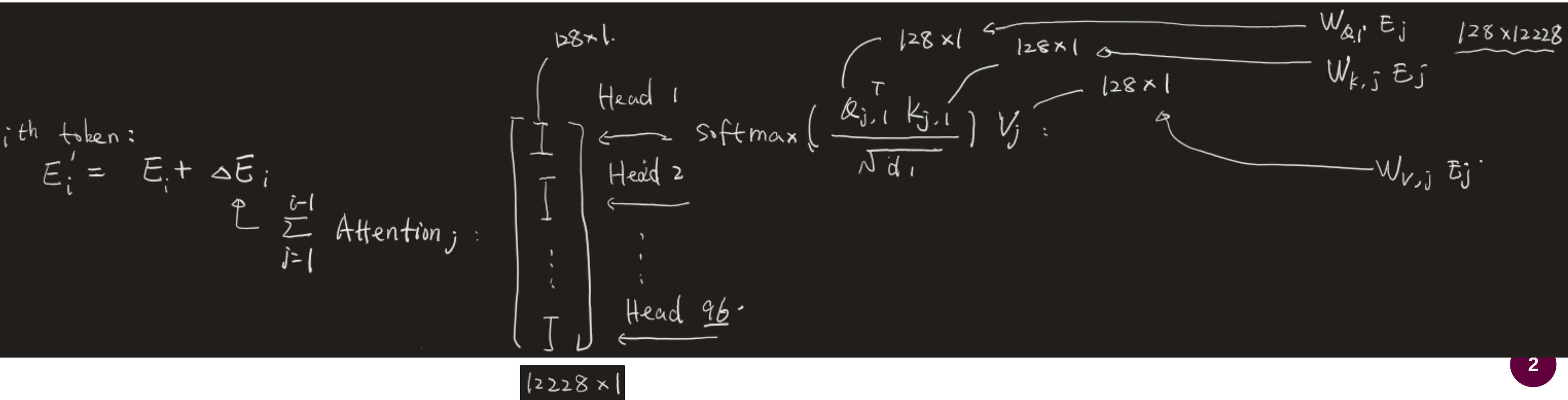
zhiguo.zeng@centralesupelec.fr

03/December/2024

A quick recap – How GPT works

- Let’s walk through how a token flows in a GPT model: https://bbycroft.net/llm

- A mistake in the video:
  - The dimension of the value vector V per attention head is 128, while in the video, I mistaken it to be 12288.
    - In fact, each attention head will produce a vector V of dimension 128×1.
    - The 96 V from the attention heads are concatenated to produce the output value vector, which has a dimension of $128 \times 96 = 12288$.
  - In general, value vector size per attention head =
    - dimension of the embeddings / number of attention heads.



ith token:
$E_i' = E_i + \Delta E_i$
$\sum_{j=1}^{i-1} \text{Attention}_j$:
128×1.
Head 1
Head 2
Head 96.
12228×1
$\text{softmax}\left(\frac{Q_{j,1}^T K_{j,1}}{\sqrt{d_1}}\right) V_j$:
128×1
128×1
128×1
$W_{Q,1} \cdot E_j$
$W_{K,j} E_j$
$W_{V,j} E_j$
128×12228

Coding with English

- Use ChatGPT to:
  - Generate a python code to write a script, and send a chat completion request using GPT-4o-mini, with prompt: “Hello, world”.
  - Get the results, and retrieve the response texts from the results.
  - Print the results on the screen.
- Note that:
  - The openai changed their api. It might be possible that ChatGPT will give you code following the old interface.
  - In this case, you can copy the api reference: https://platform.openai.com/docs/api-reference/chat/create, and input this as a context to your prompt.

Good practice:

- Be precise:
  - Specify your tasks,
  - How your data are structured
  - Variable type.
- Split the tasks into clear steps.
- Provide context.
- If error occurs, don’t worry, iterate back-and-forth.

Coding with English

Generate a python code to write a script and send a chat completion request using GPT-4o-mini, with prompt: “Hello, world”. Get the results and retrieve the response texts from the results and print the results.

Generate a python script: First, send a prompt “hello, world” to the chat completion api of OpenAI. Use “GPT-4o-mini” as the model and get the response. Then, extract the generated texts from the response. Finally, print the extracted texts on the screen.

Coding with English

Generate a python script: First, send a prompt “hello, world” to the chat completion api of OpenAI. Use “GPT-4o-mini” as the model and get the response. For this step, please use the api chat.completions.creat(). Below is an example of usage (delimited by '“):

“’ …

‘”

Then, extract the generated texts from the response. The response is an object defined by

“’…

‘”

Finally, print the extracted texts on the screen. Please generate the code only based on the context here.

I have an error message "An error occurred: 'ChatCompletionMessage' object is not subscriptable"

Exercise: Investigate the impact of temperature

- Develop a python script to:
  - Take an arbitrary question from the train.csv, and send it to gpt-4o-mini to generate an answer. Then, ask the model to write no more than 300 words to explain the choice.
  - Change the parameter “temperature” 0, 0.5, 2. See how the response changes.
  - Put your answer here:
  - respones_with_diff_temperature.xlsx

Good practice:

- Be precise:
  - Specify your tasks,
  - How your data are structured
  - Variable type.
- Split the tasks into clear steps.
- Provide context.
- If error occurs, don’t worry, iterate back-and-forth.

CentraleSupélec

Course project:

Develop an AI agent to pass the Certified Reliability Engineer exam

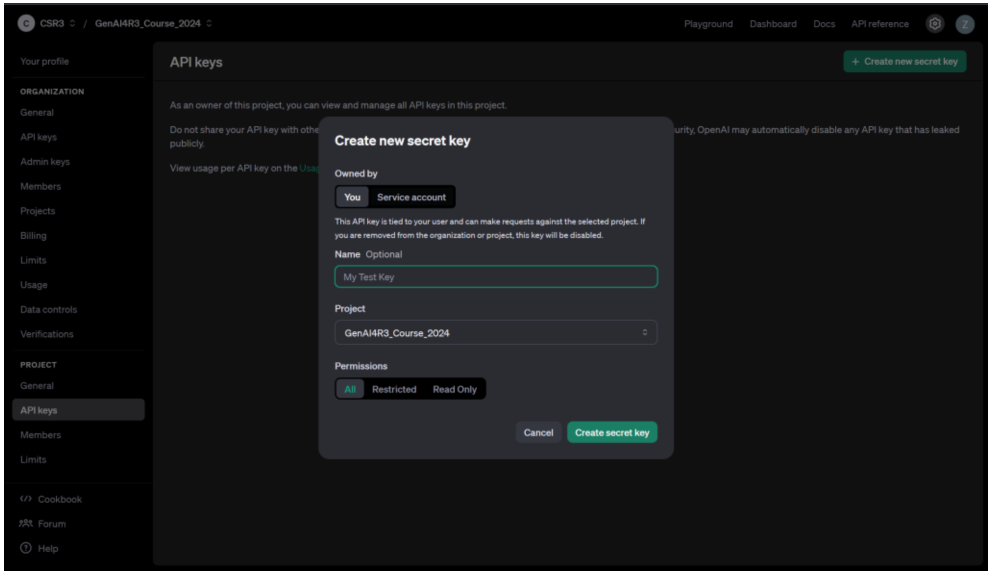
12/6/2024

Objectives

- Objectives: Develop an AI agent that is able to fulfil the basic requirement of a reliability engineer.
- Evaluation dataset:
  - 49 questions collected from the certified reliability engineer exam from ASQ.
  - All multiple choice questions.
  - 25 are given correct answers for training, 24 for testing.
  - The goal is to have the AI agent predict correctly the answers for the test dataset.
- You will be able to use OpenAI's GPT-4o-mini as your base model.
- Deadline: 23:59, 17/01/2025
- Details: https://www.kaggle.com/competitions/generative-ai-for-reliability-engineering

Getting start – Prepare your API key

- Prepare your OpenAI api
  - Go to https://platform.openai.com/settings/proj_HrzJhPdpQk6pBgzvQaxwDAZ0/api-keys.
  - Choose "Create an api key"
  - You will not be able to see the api key again. Please save it.
  - NEVER SHAER IT WITH OTHERS!
  - Export it to an environment variable: https://help.openai.com/en/articles/5112595-best-practices-for-api-key-safety



API keys

As an owner of this project, you can view and manage all API keys in this project.

Do not share your API key with othe… …urity, OpenAI may automatically disable any API key that has leaked publicly.

View usage per API key on the Usag…

Create new secret key

Owned by: You | Service account

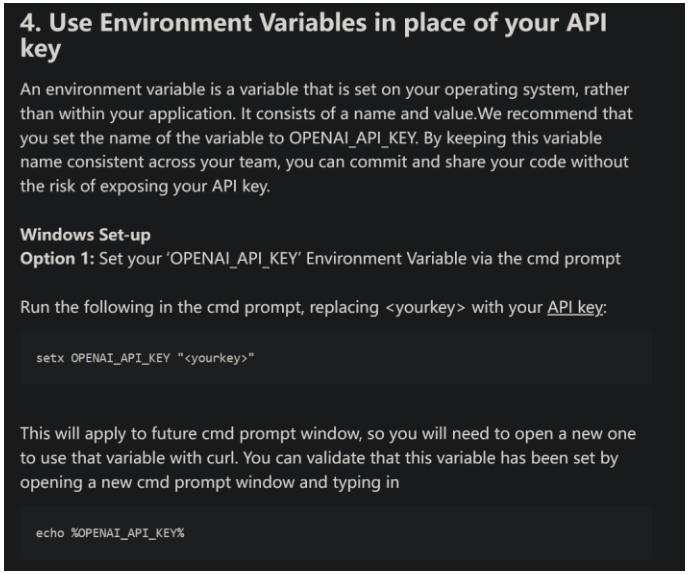
This API key is tied to your user and can make requests against the selected project. If you are removed from the organization or project, this key will be disabled.

Name Optional: My Test Key

Project: GenAI4R3_Course_2024

Permissions: All | Restricted | Read Only

Cancel | Create secret key



4. Use Environment Variables in place of your API key

An environment variable is a variable that is set on your operating system, rather than within your application. It consists of a name and value.We recommend that you set the name of the variable to OPENAI_API_KEY. By keeping this variable name consistent across your team, you can commit and share your code without the risk of exposing your API key.

Windows Set-up
Option 1: Set your 'OPENAI_API_KEY' Environment Variable via the cmd prompt

Run the following in the cmd prompt, replacing <yourkey> with your API key:

```
setx OPENAI_API_KEY "<yourkey>"
```

This will apply to future cmd prompt window, so you will need to open a new one to use that variable with curl. You can validate that this variable has been set by opening a new cmd prompt window and typing in

```
echo %OPENAI_API_KEY%
```

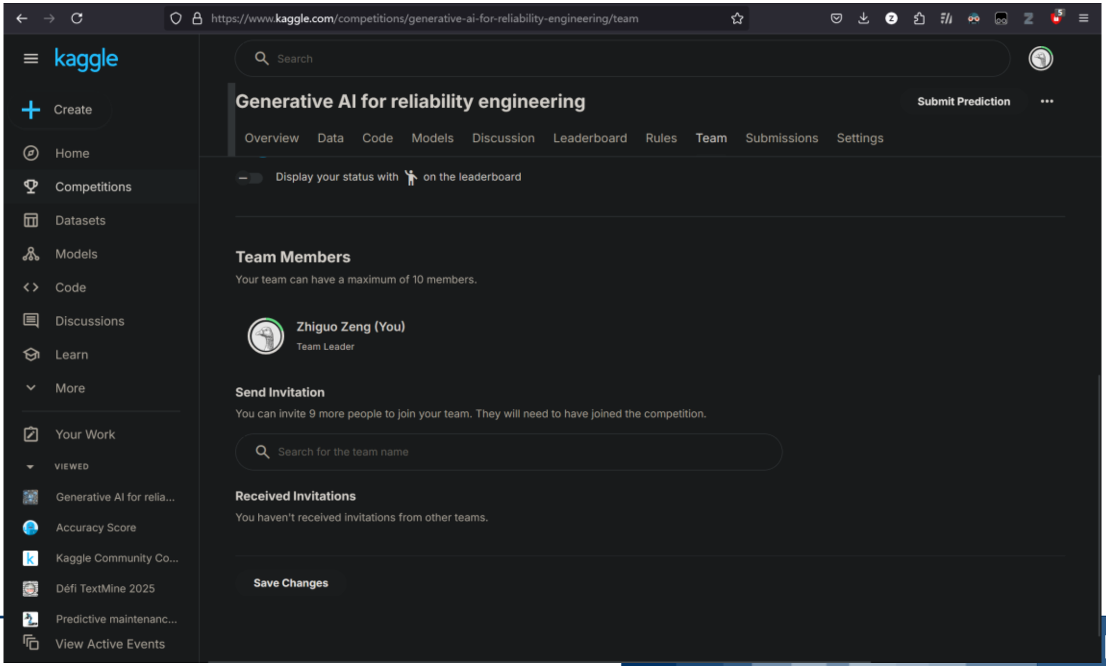
Exercise: Now, let’s move to develop a basic program for the data challenge

- Download the dataset from Kaggle.
- Instruct ChatGPT to write a program:
  - Read the test data from a csv file.
  - Then, create a loop to get each “question” from the corresponding column.
  - For each equation, create an api request:
    - Define your own system prompt
    - Use the question as user prompt
    - Get the answer
  - Extract the letter corresponding to the answer.
  - Output all the predicted answers following the structure of “submission.csv”.

- To design your system prompt:
  - https://platform.openai.com/playground/chat?models=gpt-4o-mini
  - A small exercise:
    - Design a system prompt, to ask the model only return the letter corresponding to the answer.

Make a first submission!

- Create your team in kaggle: https://www.kaggle.com/t/ee8c716b5f374ae1b0e1ae0e09c84c54
- Submit your answer in Kaggle, and see your results.
- If you have sometime:
  - Try on the training dataset.
  - Log all the wrong answers generated by the LLM.
  - Try to discover some patterns and propose directions to improve.



https://www.kaggle.com/competitions/generative-ai-for-reliability-engineering/team

Generative AI for reliability engineering

Display your status with on the leaderboard

Team Members

Your team can have a maximum of 10 members.

Zhiguo Zeng (You)
Team Leader

Send Invitation

You can invite 9 more people to join your team. They will need to have joined the competition.

Search for the team name

Received Invitations

You haven't received invitations from other teams.

Exercise: Let’s log the wrong answers.

- Write a python script:
  - Get all the questions from “train.csv”. Get the correct answers as well.
  - Ask the model to generate an answer. Ask the model to generate an explanation as well
  - Compare the generated answer to the correct answers.
  - If the answer is wrong, save the wrong prediction, correct answer, and the model explanation in a dataframe.
  - When all the questions are answered, output the dataframe to a csv file named “failure_log.csv”.

Thank you! Questions?