ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO DI
SCIENZE STATISTICHE "PAOLO FORTUNATI"

9 - OpenAI API and GPT

Giovanni Della Lunga
giovanni.dellalunga@unibo.it

Halloween Conference in Quantitative Finance

Bologna - October 25-26-27, 2023

OpenAI API

Open AI API

- OpenAI is a cutting-edge research organization specializing in artificial intelligence.
- Founded in 2015, OpenAI has made significant contributions to AI research, including language models like GPT-3.5 and GPT-4.
- The OpenAI API is a platform that provides developers with access to OpenAI's powerful language models.
- It allows developers to integrate natural language processing capabilities into their applications, products, and services.
- Through API calls, developers can leverage OpenAI's state-of-the-art models for various language-related tasks.

Open AI API - Key Features

- Natural Language Understanding: The API can analyze and understand text in natural language, making it suitable for a wide range of applications.
- Text Generation: It can generate human-like text based on given prompts, enabling content creation and more.
- Language Translation: The API can translate text from one language to another, promoting global accessibility.
- Summarization: It can automatically summarize long documents, making information more digestible.
- Chatbot Capabilities: Developers can create chatbots and virtual assistants using the API.
- Code Generation: The API can generate code snippets in various programming languages.

Open AI API - How to Access OpenAI API

- Account Registration: Users need to create an OpenAI account to access the API.
- API Key Generation: After registration, users generate API keys, which are used for authentication.
- The OpenAI API uses API keys for authentication.
- Documentation: OpenAI provides extensive documentation with code examples to help users get started.
- SDKs and Libraries: There are SDKs and libraries available for popular programming languages to simplify integration.

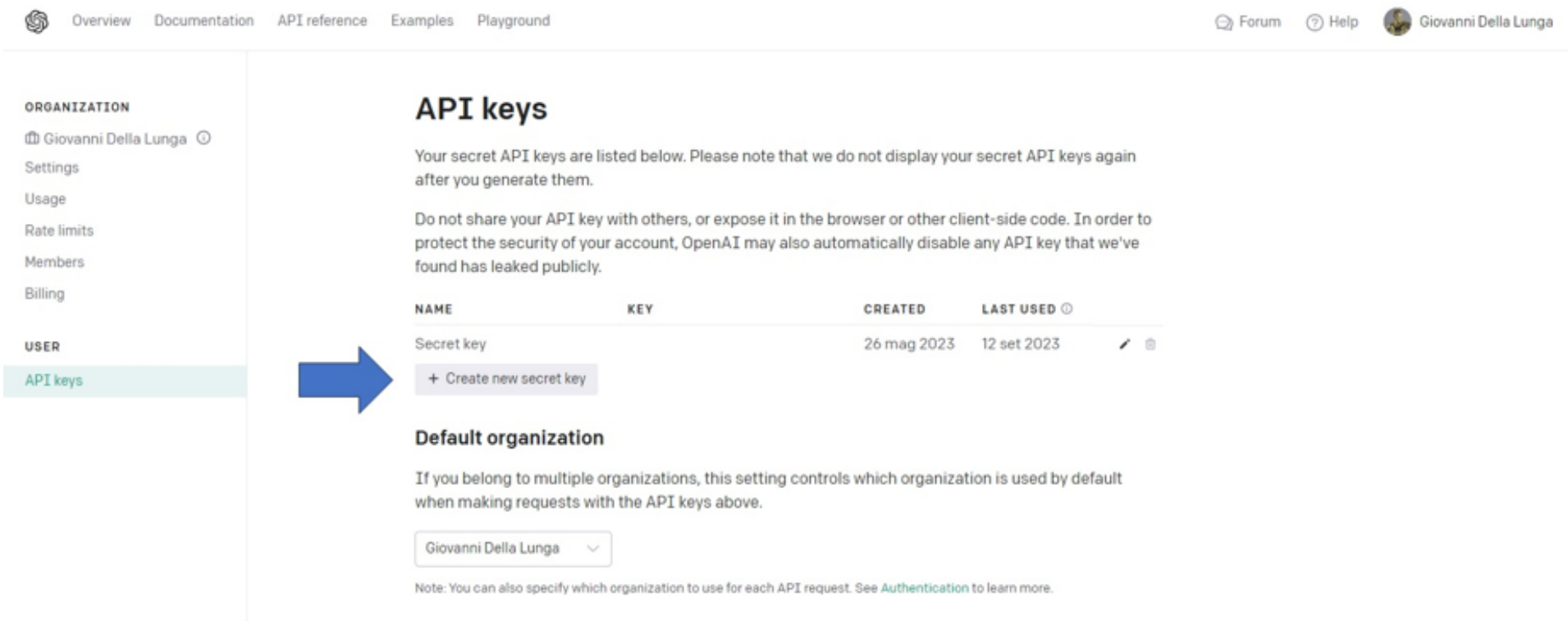
Open AI API - How to Access OpenAI API

- You can interact with the API through HTTP requests from any language, via OpenAI official Python bindings;
- To install the official Python bindings, run the following command:

```
pip install openai
```

Open AI API - Create API Key

To create API keys, sign up to OpenAI, and click on your profile and on "View API keys". There you can create a new key.



Open AI API - Create API Key

- A good practice to keep credentials outside the code so the API keys are not committed to your repository. Save the keys in a .env file or in an environment variable.
- Now, you can either load the environment variables with source .env or you can use a package like python-dotenv to load the environment variables automatically in your notebook or Python script.

GPT Models

- To use a GPT model via the OpenAI API, you'll send a request containing the inputs and your API key, and receive a response containing the model's output.
- Latest models, gpt-4 and gpt-3.5-turbo, are accessed through the chat completions API endpoint.

	MODEL FAMILIES	API ENDPOINT
Newer models (2023–)	gpt-4, gpt-3.5-turbo	https://api.openai.com/v1/chat/completions
Updated base models (2023)	babbage-002, davinci-002	https://api.openai.com/v1/completions
Legacy models (2020–2022)	text-davinci-003, text-davinci-002, davinci, curie, babbage, ada	https://api.openai.com/v1/completions

Completion and Chat Completion

- In the context of GPT (Generative Pre-trained Transformer) programming, **completion** typically refers to the process of generating or completing a sequence of text based on an initial input or prompt.
- GPT models, including GPT-3 and GPT-4, are powerful language models that are trained on a vast amount of text data, allowing them to generate coherent and contextually relevant text based on the input they receive.

Chat Completion

- Chat models take a list of messages as input and return a model-generated message as output.
- Although the chat format is designed to make multi-turn conversations easy, it's just as useful for single-turn tasks without any conversation.
- An example Chat completions API call looks like the following:

python Copy

```python
response = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are a helpful assistant."},
        {"role": "user", "content": "Who won the world series in 2020?"},
        {"role": "assistant", "content": "The Los Angeles Dodgers won the World Series in 20
        {"role": "user", "content": "Where was it played?"}
    ]
)
```

Chat Completion Response Format

An example Chat completions API response looks as follows:


```
{
  "choices": [
    {
      "finish_reason": "stop",
      "index": 0,
      "message": {
        "content": "The 2020 World Series was played in Texas at Globe Life Fie
        "role": "assistant"
      }
    }
  ],
  "created": 1677664795,
  "id": "chatcmpl-7QyqpwdfhqwajicIEznoc6Q47XAyW",
  "model": "gpt-3.5-turbo-0613",
  "object": "chat.completion",
  "usage": {
    "completion_tokens": 17,
    "prompt_tokens": 57,
    "total_tokens": 74
  }
}
```

Completion API

- The completions API endpoint received its final update in July 2023 and has a different interface than the new chat completions endpoint.
- Instead of the input being a list of messages, the input is a freeform text string called a prompt.
- An example API call looks as follows:

python Copy

```python
response = openai.Completion.create(
  model="gpt-3.5-turbo-instruct",
  prompt="Write a tagline for an ice cream shop."
)
```

ChatCompletion and Completion

- The difference between these APIs derives mainly from the underlying GPT models that are available in each.
- The chat completions API is the interface to the most capable model (gpt-4), and to the most cost effective model (gpt-3.5-turbo).
- See the full API reference documentation to learn more.