ChatGPT Workshop for Data Scientists

8 academic hours

Workshop Description:

This comprehensive workshop provides an in-depth introduction to OpenAI APIs and their application in Machine Learning. Participants will gain the knowledge and skills necessary to harness the power of OpenAI models, particularly focusing on the Chat API. Through practical hands-on exercises and real-world examples, participants will learn how to leverage OpenAI API to build their own intelligent applications and enhance their software with natural language interfaces.

The workshop begins by introducing the participants to the different available models offered by OpenAI, including their capabilities and use cases. Participants will gain an understanding of the strengths and limitations of each model and how they can be applied to solve various language processing tasks.

Next, participants will dive into the specifics of using the OpenAI Chat API to develop their own applications. They will learn how to leverage the API to answer questions based on a set of documents, summarize texts, label text documents for machine learning models, and perform sentiment analysis on different documents. Additionally, participants will explore the exciting possibility of creating conversational agents, such as chatbots, to provide natural language interfaces for internal and external customers.

The workshop then delves into the process of fine-tuning ChatGPT, OpenAI's flagship language model, to meet specific department or organization needs. Participants will discover how to prepare their datasets for fine-tuning and explore advanced usage scenarios to maximize the model's performance. They will also learn how to utilize Weights & Biases for experiment tracking, ensuring efficient experimentation and model development.

Another key aspect of the workshop is leveraging OpenAI word embeddings for various machine learning and data preprocessing tasks. Participants will understand how to convert texts into word embeddings and visualize them in two-dimensional space. They will learn how to use embeddings as text feature encoders for supervised and unsupervised machine learning algorithms, enabling tasks such as zero-shot classification, recommendation systems, and text clustering. Participants will also explore how to perform text searches using embeddings, enhancing the search capabilities of their applications.

Furthermore, the workshop covers OpenAI's Whisper API, which enables speech-to-text transcription and translation. Participants will learn how to transcribe audio in different languages and translate the transcriptions into English, opening up possibilities for multilingual audio processing.

Finally, the workshop addresses content moderation using OpenAI API. Participants will discover how to utilize the API to check whether content is classified as inappropriate, empowering them to take appropriate actions based on identified classes such as hate speech, threatening language, self-harm, sexual content, violence, and graphic material.

Who Should Attend:

This workshop is relevant for data scientists, ML engineers and from all disciplines, who are seeking to better understand ChatGPT and its capabilities for assisting with everyday ML tasks.

Required Skills:

Experience and background in Python and Machine Learning

Course Contents:

- Introduction to OpenAI API's and the different available models
- How to use the OpenAI Chat API for building your own applications:
  - ✓ Answer questions about a set of documents
  - ✓ Summarize texts
  - ✓ Label text documents for ML models
  - ✓ Sentiment analysis – get the sentiment for each document
  - ✓ Create conversational agents (Chat bots) for internal / external customers.
  - ✓ Give your software a natural language interface
  - ✓ Translate languages
- Fine-tune ChatGPT to your department / organization needs. ChatGPT will learn your data and terminology and could provide more accurate and relevant solutions:
  - ✓ How to prepare your dataset for fine-tuning
  - ✓ Advanced usage
  - ✓ Weights & Biases for experiment tracking
- Leverage OpenAI word embeddings for various Machine learning and data preprocessing tasks
  - ✓ Convert texts to word embedding
  - ✓ Data visualization in 2D
  - ✓ Use embedding as a text feature encoder for Supervised and Unsupervised Machine Learning algorithms
  - ✓ Zero-shot classification
  - ✓ Recommendations using embeddings
  - ✓ Clustering texts
  - ✓ Text search using embeddings
- Speech to text – Use the whisper API for transcribing or translating audio
  - ✓ Transcribe audio into whatever language the audio is in
  - ✓ Translate and transcribe the audio into English
- Moderation – Use OpenAI API to check whether a content is classified as inappropriate so you could take an action (Included classes: hate\threatening\self-harm\sexual\violence\ graphic)