

LLM 마스터클래스 #3

LLM 서비스 개발: 기초 실습



Part 3-1.

플레이 그라운드 해보기



Playground

<https://platform.openai.com/playground?mode=chat>

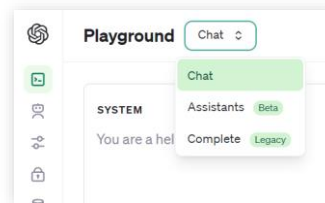
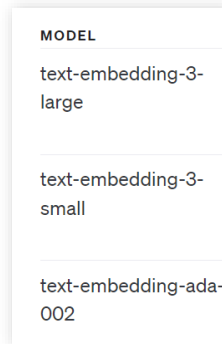
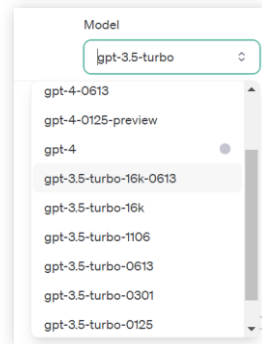
살짝 뚜껑을 뜯어서 안을 볼 수 있습니다!

- 챗/assistant/ completion
- model types
- parameters:
 - temperature
 - max length...



모델들

- GPT4
- GPT3
- Completion, chat, assistant



Models

Overview

The OpenAI API is powered by a diverse set of models with different capabilities and price points. You can also make customizations to our models for your specific use case with [fine-tuning](#).

MODEL	DESCRIPTION
GPT-4 and GPT-4 Turbo	A set of models that improve on GPT-3.5 and can understand as well as generate natural language or code
GPT-3.5 Turbo	A set of models that improve on GPT-3.5 and can understand as well as generate natural language or code
DALL·E	A model that can generate and edit images given a natural language prompt
TTS	A set of models that can convert text into natural sounding spoken audio
Whisper	A model that can convert audio into text
Embeddings	A set of models that can convert text into a numerical form
Moderation	A fine-tuned model that can detect whether text may be sensitive or unsafe
GPT base	A set of models without instruction following that can understand as well as generate natural language or code
Deprecated	A full list of models that have been deprecated along with the suggested replacement



View code

View code

You can use the following code to start integrating your current prompt and settings into your application.

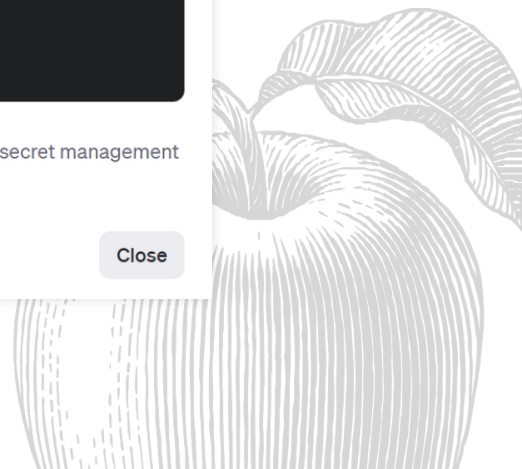
POST /v1/chat/completions

python  Copy

```
1 from openai import OpenAI
2 client = OpenAI()
3
4 response = client.chat.completions.create(
5     model="gpt-3.5-turbo",
6     messages=[
7         {
8             "role": "user",
9             "content": ""
10        }
11    ],
12    temperature=1,
13    max_tokens=256,
14    top_p=1,
15    frequency_penalty=0,
16    presence_penalty=0
17 )
```

Your API Key can be found [here](#). You should use environment variables or a secret management tool to expose your key to your applications.

Close



Part 3-2.

Colab 으로 코드 실습



GPT 리퀘스트 보내기

Colab demo - 링크



Part 3-3.

Whisper API

STT, TTS



음성을 텍스트로, 텍스트를 음성으로 - demo

Whisper API - STT

TTS

- [노트북 링크](#)



Part 3-4.
OpenAI vision,
Image gen



이미지 읽기, 이미지 만들기 - demo

vision

Image generation

- [노트북 링크](#)



Part 3-5.
Langchain
OpenAI 와 다른
모델 쓰기



Langchain

Langchain 설명



랭체인이 뭔가요

- 뚜껑 열고 이것 저것 시도해보는 사람들을 위한 도구 모음
- LLM 개발을 쉽게 하기 위한 소프트웨어 프레임워크
- 오픈소스이고, 아주 활발하게 개발되고 있음
- 랭체인 사용 예: GPT/LLM 위에 앱을 구축하고, 자신의 데이터 (데이터베이스, PDF, 이메일) 등을 쓸 수 있도록 하고, 프롬프트를 쉽게 관리할 수 있도록 하며, 대화 기록을 저장하고 여러가지 도구를 쓰고....
- 언어 모델을 가지고 어떻게든 뭐 해보겠다는 사람들을 위한 도구 모음집



Part 3-6.
Langchain
+ 단순 RAG



Langchain

Langchain 사용해서 RAG 구현하기



Langchain RAG

DuckDuckGo demo

Youtube 내용 기반으로 답변 하기

- Using langchain: DuckDuckGo, Youtube
- 노트북 [링크](#)



PDF 검색해서 답변하기 - demo

GPTs 로 해보기

Bing 으로 해보기

코드 데모



Part 3-7. Function calling

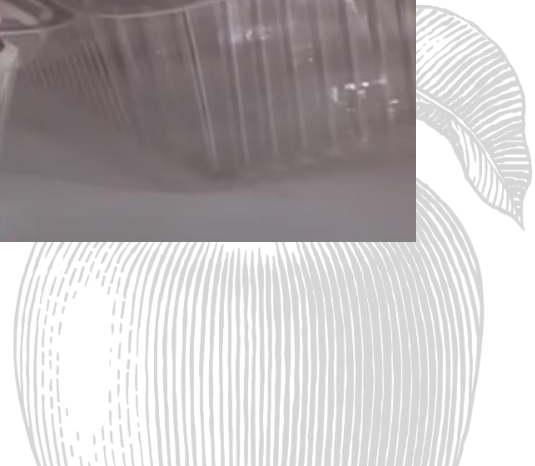


개발자라면 꼭!꼭! 필요함!

Function calling



정확한 규격의 데이터를 필요로 하는 프로세스



Function calling - demo

Function calling - 왜 필요한가

- 노트북 [링크](#)



잠깐 정리

- Model types
- Colab
- 코드 돌려보기
- 리퀘스트 보내기
- 랭체인 알아보기
- 랭체인으로 간단한 RAG 구현하기
- 그 외에 쓸모있는 API - Whisper, Function calling... 써보기

