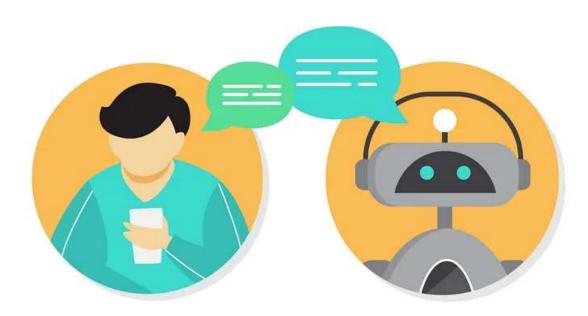


Published in Geek Culture



Using OpenAI GPT-3 AI to build question answer service

Create a question-answer service using GPT-3 and OpenAl

For any existing knowledgebase for your product or service, it is easier than ever to create an AI powered add-on that can search the knowledgebase and answer questions in natural language.

Using GPT-3 (developed by OpenAI), one can easily build an API driven service with great results, without actually learning or implementing any Machine Learning code or putting heavy compute power to support it.

The OpenAI APIs are easy to understand and are backed by a decent documentation. This service recently came out of private beta. You can simply signup at <u>OpenAI website</u> and start using the APIs.

Note: OpenAI is not a free service. You get a free credit of \$18 upon signup. Thereafter, you need to put your credit card for pay-as-you-go billing.

Before we dive further, let's understand some terms at play here:

OpenAl

OpenAI is an artificial intelligence (AI) research laboratory consisting of the for-profit corporation OpenAI LP and its parent company, the non-profit OpenAI Inc. — <u>Wikipedia</u>

OpenAI Inc. was founded in 2015 by Elon Musk, who later resigned in 2018. It's stated objective was to develop friendly AI to help humanity at large. Currently it is supported by Microsoft, while Elon is still a donor.

GPT-3

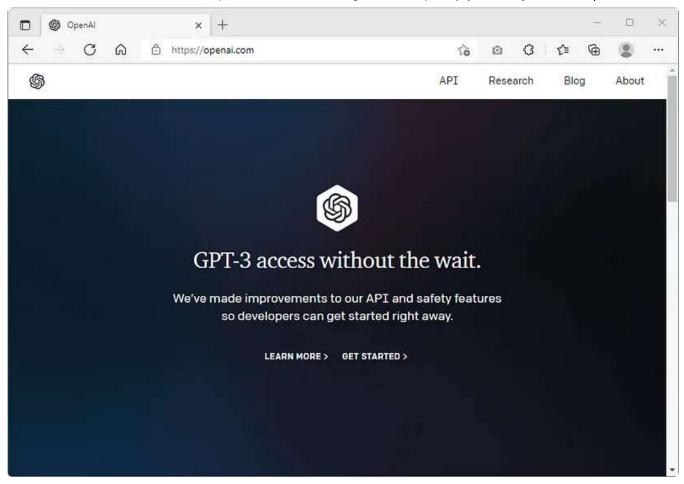
Generative Pre-trained Transformer 3 (GPT-3) is an autoregressive language model that uses deep learning to produce human-like text. — <u>Wikipedia</u>

GPT-3 is the third-generation language prediction model developed by OpenAI. It is said that the quality of text generated by GPT-3 is so high, that it can be very difficult to determine if it was generated by a human or AI. While GPT-3 can be used publicly by signing up with OpenAI, but Microsoft has exclusive rights to the GPT-3's underlying model, which is not made public.

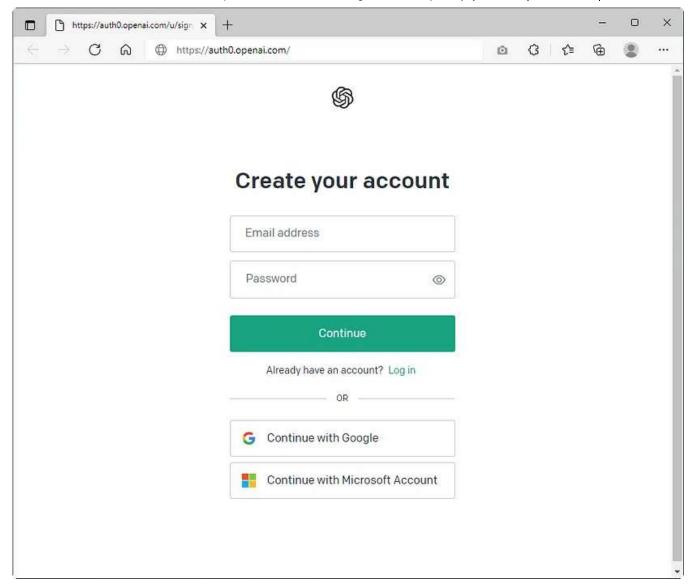
Now, let's see how we can use GPT-3 to create an answering service on top of our existing knowledgebase. This guide explains the step-by-step process.

Signup at OpenAI and get an API Key

This is pretty much obvious. We essentially need to signup at <u>OpenAI's website</u> to create an account. It recently came out of private beta, so you can signup and get access instantly.

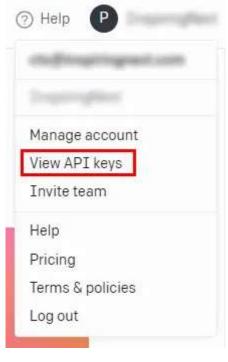


OpenAl's Home Page



OpenAl's Signup Page

You can signup with your email and phone number. After verifying your e-mail address and phone, you will instantly get access to the platform.



OpenAPI Menu

Check the menu on the top right and click on 'View API Keys' to view your API Key.

An API Key is created for you by default. You can, however, create multiple API keys to differentiate usage in different applications.

API keys

Your secret API keys are shown below. Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically rotate any API key that we've found has leaked publicly.



Default organization

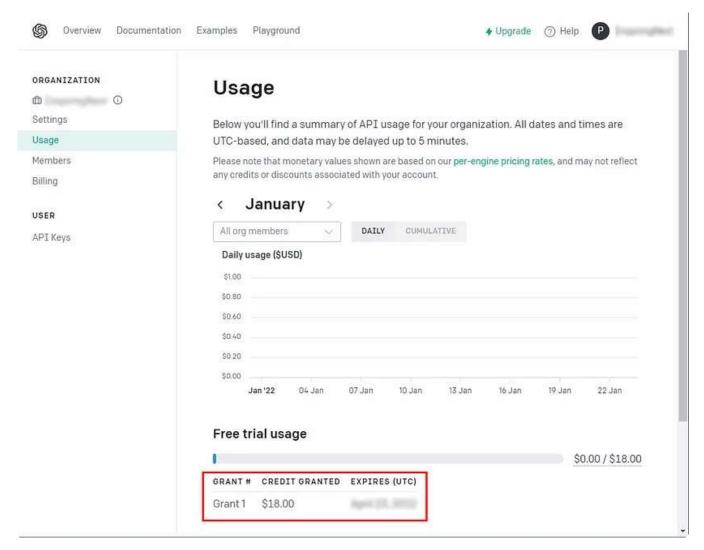
If you belong to multiple organizations, this setting controls which organization is used by default when making requests with the API keys above.



Note: You can also specify which organization to use for each API request. See Authentication to learn more.

View OpenAI API Keys

As you signup for a new account, you are given some trial credit (\$18) to begin experimentation. In my experience, this \$18 is more than enough to play around. Check your usage on 'Usage' page as shown below:



Check Usage and Trial Credit

Ok, that's enough about OpenAI. Let's move on to some other exciting stuff we'll be doing here.

Preparing Feed from the Knowledgebase

Your existing knowledgebase may be a bunch of documents or a WordPress Website with numerous articles. In whatever form this knowledgebase exist, you need to find a way to extract the content in plain-text. Generate separate text chunks for each article.

Let's take an example to simplify the remaining of this guide. Let's assume that we have a platform that offers online courses. Our documentation consists of articles that explains how to signup for a course, what is the refund policy, what are the payment options available. Here are some sample articles from our knowledgebase:

Sample Article 1

Title: How to enroll for a course?

Content: To enroll for a course, simply create an account with us by clicking on 'Signup' link from our home page. After signing up, you can search from our catalog of over 500 courses. To enroll in a course, click on 'Enroll' button on the course page.

Sample Article 2

Title: What is the refund policy?

Content: We offer 14-day refund policy. If you are not satisfied with any of our course, just send us an e-mail and we will refund 100% of your course fee. No questions asked.

Sample Article 3

Title: What payment methods are supported?

Content: We support all types of debit cards, credit cards and bank transfers. You can also pay in installments for courses having fee more than \$49.

In order to prepare the question-answer service, you can merge title and content into a single string, but its not necessary. For the 3 articles listed above, you should end up with 3 unique strings. Combining all three of these, let's call this input as our **Feed**.

How does it work?

This is what happens in a nutshell:

We send the *Feed* and a *Question* to the *Answers API*. The API returns an *Answer* to the *Question*, based on the *Feed*.

The answer returned may not be a part of your original **Feed**. It can be completely different in terms of words and sentences. It, however, will be a relevant answer to your question, based on the **Feed**.

Understanding OpenAl Answers API

OpenAI offers a bunch of APIs to do different things. Since we are looking at a question-answer service here, we will use the *Answers API*.

The Answers API Endpoint is: https://api.openai.com/v1/answers
The request method is POST
The API Key must go into Authorization header (prefixed with 'Bearer')

Request Payload Format:

Don't get overwhelmed, we'll break it down in just a bit.

```
documents: [list of strings from knoweledgebase],
  question: "This is your question?",
  search_model: "ada",
  model: "curie",
  examples_context: "There are nine planets and one star in the
solar system.",
  examples: [["How many planets are there in the solar system?",
"There are 9 planets in the solar system"]],
  max_tokens: 20
}
```

Here's a breakdown of each item in the request payload:

documents

7

This is simply the list of strings that represent your knowledgebase articles. In case of our sample knowledgebase defined above, it will look like this:

```
"How to enroll for a course? To enroll for a course, simply create an account with us by clicking on 'Signup' link from our home page. After signing up, you can search from our catalog of over 500 courses. To enroll in a course, click on 'Enroll' button on the course page.",

"What is the refund policy? Content: We offer 14-day refund policy. If you are not satisfied with any of our course, just send us an e-mail and we will refund 100% of your course fee. No questions asked.",

"What payment methods are supported? We support all types of debit cards, credit cards and bank transfers. You can also pay in installments for courses having fee more than $49."
```

question

This is the question you want to ask against your knowledgebase. As an example, we'll ask "How many courses are available in your catalog?" in following section.

we are the state of the state of





Sign In





and also differ in pricing. The common GPT-3 models are: "ada", "babbage", "curie", and "da-vinci". To remember the names, you can note the names begin with a, b, c and d. "Da-Vinci" is the most capable and expensive model, while "ada" is the fastest and cheapest one.

The Answers API work in two steps. In the first step, an appropriate document is selected from the list of documents (*Feed*) supplied by you. In second step, an answer is composed from the most relevant document. *search_model* value specifies which model to use for searching the list of documents. Usually "ada" works just fine for this.

model

This value specifies which model to use for generating the answer. You can experiment with different models for your use case. Depending on the type of content you have, you can decide on any one of the four models.

example_context and example

These values have nothing to do with your documentation. These are simply instructing the API on what type of answer we are expecting from it. You can provide any logical context and a sample answer. OpenAI will use these values to determine the *tone* and answer *format*.

max_tokens

This value can be used to specify the maximum length of the answer. It's not an exact representation of number of words or letters. You can try with different values to find the most suitable one for your use case.

Note: To keep it simple, I have intentionally ignored some other keys that can be passed in the payload. You can have a finer control over the answer API by adding more keys in your payload. Some more keys are explained at end of this article. Refer to OpenAI Documentation for full API reference.

Let's hit the Answers API

Just to recap, at this stage, we have all these items ready:

- 1. Signed up at OpenAI and got an API Kev
- 2. Prepared strings from our Knowledgebase. One string for each article.
- 3. Understood the Answers API and various payload keys.

Using our sample articles, here's a the payload with documents and a question "How many course are available in your catalog?".

```
{
    "documents": ["How to enroll for a course? To enroll for a course, simply create an account
    with us by clicking on 'Signup' link from our home page. After signing up, you can search
    from our catalog of over 500 courses. To enroll in a course, click on 'Enroll' button on the
    course page.", "What is the refund policy? Content: We offer 14-day refund policy. If you are
    not satisfied with any of our course, just send us an e-mail and we will refund 100% of your
    course fee. No questions asked.", "What payment methods are supported? We support all types
    of debit cards, credit cards and bank transfers. You can also pay in installments for courses
    having fee more than $49."],
    "question": "How many courses are available in your catalog?",
    "examples_context": "There are nine planets and one star in the solar system.",
    "examples": [["How many planets are there in the solar system?", "There are 9 planets in the
    solar system"]],
    "search_model": "ada",
    "model": "curie",
    "max_tokens": 20
```

OpenAl Answers API Sample Payload

Here's the answer received:

```
answers": [
   "We have over 500 courses in our catalog.
"completion": "cmpl-4TYJhhAHZhGgnecUYYnbS2mKjH26z",
"model": "curie:2020-05-03",
"object": "answer",
"search_model": "ada",
"selected_documents": [
       "document": 0,
       "text": "How to enroll for a course? To enroll for a course, simply create an account
           with us by clicking on 'Signup' link from our home page. After signing up, you
           can search from our catalog of over 500 courses. To enroll in a course, click on
           'Enroll' button on the course page, "
   },
       "document": 2,
       "text": "What payment methods are supported? We support all types of debit cards,
           credit cards and bank transfers. You can also pay in installments for courses
           having fee more than $49. "
   },
       "document": 1,
       "text": "What is the refund policy? Content: We offer 14-day refund policy. If you
           are not satisfied with any of our course, just send us an e-mail and we will
           refund 100% of your course fee. No questions asked. "
```

Response from OpenAl Answers API

As you can see, the answer correctly states that "We have over 500 courses in our catalog." Now this particular text is not part of any of our documents in the Feed. The answer is composed by GPT-3 by interpreting the meaning of your input.

Fine Tuning

The above is a very simple example. You can further fine-tune the response by specifying additional values in your request. Let's quickly cover a couple more here.

temperature

Temperature specifies how much creative freedom the AI should use. The value can be anything between 0 to 1. 0 being no creative and 1 being most creative. In most of my use cases, a value between 0.5 to 0.7 works best.

n

A value for key "n" specifies how many answers should be generated. The default

value is 1.

Conclusion

OpenAI offers incredible power to developers without actually getting into the complexities of training models and learning machine learning algorithms to utilize the power of AI. OpenAI does all the heavy lifting for us. We can simply use the APIs to build truly AI powered applications.

To read all my stories, along-with thousands of other writers on medium, consider becoming a medium member. If you join using my link, I will get some part from your joining fee. Good Luck!

Al Machine Learning Programming API Application Development

Sign up for Geek Culture Hits

By Geek Culture

Subscribe to receive top 10 most read stories of Geek Culture — delivered straight into your inbox, once a week. <u>Take a look.</u>

Your email (1)

We couldn't process your request. Try again, or contact our support team.



By signing up, you will create a Medium account if you don't already have one. Review our <u>Privacy Policy</u> for more information about our privacy practices.

About Help Terms Privacy

Get the Medium app



