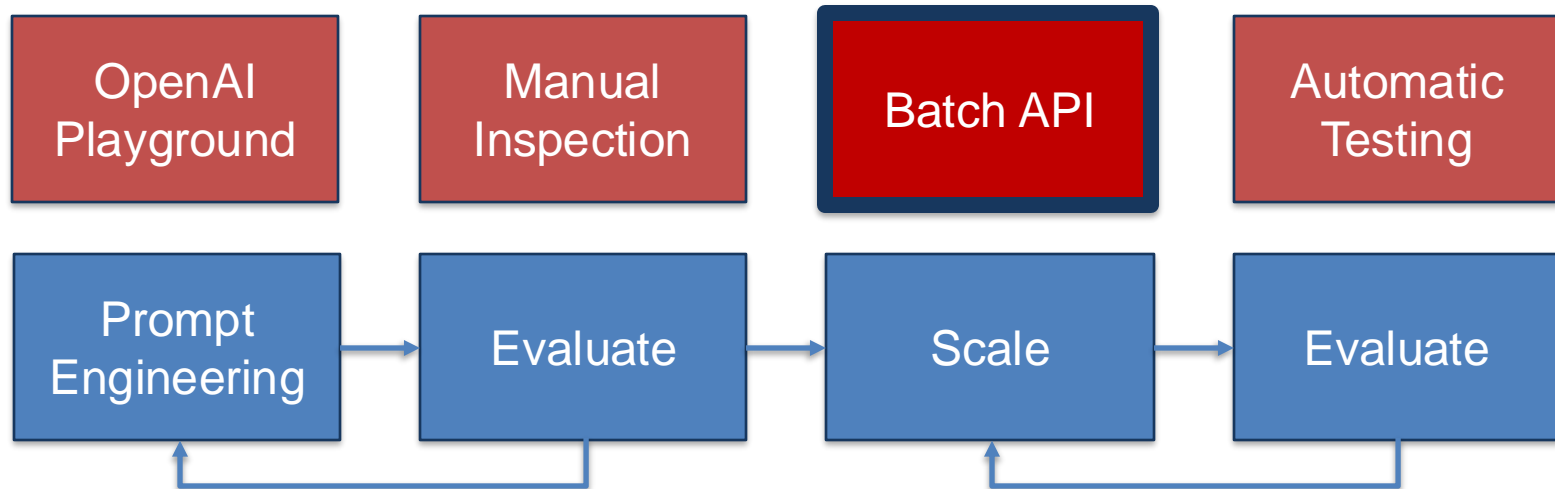


OpenAI Batch API for Scaling

October 17, 2024

Northwestern | Kellogg

Workflow for Scaling Tasks



Batch API

<https://platform.openai.com/docs/guides/batch/overview>

- **Advantages:**
 - Price: 50% of the cost of using the regular API
 - Rate Limits: significantly higher rate limits from a separate pool
 - Generates complete record of all inputs and outputs
- **When to Use:**
 - Whenever you have large jobs that don't require synchronous responses
 - Can live with 24 hour turn around time guarantee

Batch API Workflow



- All these steps except for creating the batch file can be done using a web-based console
- Creating the batch file is the hard part, you have to create one line of JSON per chat completion request
- Each line differs by some parameterized prompt

OpenAI Batch Console

<https://platform.openai.com/batches/>

The screenshot displays the OpenAI Batch Console interface. On the left is a sidebar with navigation options: DASHBOARD, Chat Completions, Assistants, Threads, Batches (selected), Evaluations, Fine-tuning, Storage, Usage, API keys, Cookbook, Forum, and Help. The main area is titled 'Batches' and contains a table of batch records. The first batch, 'batch_670846ed445c81909d26a1ac990082f0', is highlighted and its details are shown on the right. The details panel indicates the batch is 'Completed' and provides metadata such as creation time, endpoint, completion window, and request counts. A timeline at the bottom of the details panel shows the batch's progress from creation to completion. A 'Download output' button is located at the bottom right of the details panel.

Batch ID	Created at
batch_670846ed445c81909d26a1ac990082f0	10/10/2024, 4:28 PM
batch_670842b74bc881908ccda05381d82223	10/10/2024, 4:10 PM
batch_03vXEiTC91UdvIRJnXZ8qmZC	8/24/2024, 3:33 PM
batch_8Np5UsUPLZHf0V9RjJzNfYmJ	8/19/2024, 10:55 AM
batch_FcJK5eejv6jusZI4kdKBFsoH	8/13/2024, 8:34 AM
batch_Km23D9omyqYLcMHY5C61s51r	8/13/2024, 7:51 AM
batch_txclNvyzJFC0K1rLoIaZg7e27	8/13/2024, 7:50 AM

BATCH
batch_670846ed445c81909d26a1ac990082f0

- Status: Completed
- Created at: Oct 10, 2024, 4:28 PM
- Endpoint: /v1/chat/completions
- Completion window: 24h
- Completion time: 1 minute, 42 seconds
- Request counts: 5 completed, 0 failed of 5 total requests

Files

- Input: [batch-requests.jsonl](#)
- Output: [batch_670846ed445c81909d26a1ac990082f0_output.jsonl](#)
- Error:

THURSDAY, OCTOBER 10, 2024

- 16:28:13 Batch created
- 16:28:14 Batch in progress
- 16:29:54 Batch finalizing
- 16:29:55 Batch completed

[Download output](#)

Creating Batch Files

Prompt Template File (JSON)

```
{
  "model": "gpt-4o",
  "messages": [
    {
      "role": "system",
      "content": [
        {
          "type": "text",
          "text": "You are a participant in a marketing study from Monkeys Elbow, Kentucky. Here is your persona: {{Description}}"
        }
      ]
    },
    {
      "role": "user",
      "content": [
        {
          "type": "text",
          "text": "Please select the image you find the most appealing from these three and explain your choice. The image on the left can be called 'kawaii', the image in the middle 'cubs', and the image on the right 'futuristic'."
        },
        {
          "type": "image_url",
          "image_url": [
            {
              "url": "https://kellogg-shared.s3.us-east-2.amazonaws.com/images/select_enhanced.png"
            }
          ]
        }
      ]
    }
  ],
  "temperature": 1,
  "max_tokens": 2840,
  "top_p": 1,
  "frequency_penalty": 0,
  "presence_penalty": 0,
  "response_format": {
    "type": "text"
  }
}
```

```
ID, Age, College, Description
1, 28, No, Hank Rodriguez is a 28 year old black architect who enjoys Star Wars.
2, 45, Yes, David Carter is a 45 year old black nurse who enjoys medieval history.
3, 39, Yes, David Smith is a 39 year old caucasian architect who enjoys medieval history.
4, 38, Yes, Hank Rodriguez is a 38 year old black architect who enjoys cooking.
5, 38, Yes, Sophia Lee is a 38 year old caucasian software engineer who enjoys medieval history.
6, 27, Yes, David Smith is a 27 year old caucasian artist who enjoys painting.
7, 43, Yes, Hank Smith is a 43 year old caucasian teacher who enjoys fitness.
8, 42, Yes, Mia Lee is a 42 year old software engineer who enjoys cooking.
9, 30, Yes, Mia Smith is a 30 year old artist who enjoys painting.
10, 41, No, Aaron Rodriguez is a 41 year old caucasian software engineer who enjoys painting.
11, 30, No, Michael Carter is a 30 year old caucasian truck driver who enjoys outdoors.
12, 38, Yes, Hank Smith is a 38 year old black bartender who enjoys travel.
13, 34, No, Michael Lee is a 34 year old asian bartender who enjoys medieval history.
14, 37, No, Aaron Carter is a 37 year old latino artist who enjoys painting.
15, 28, Yes, Sam Banana is a 28 year old caucasian artist who enjoys Star Wars.
16, 40, No, Michael Rodriguez is a 40 year old caucasian software engineer who enjoys cooking.
17, 36, Yes, David Lee is a 36 year old asian truck driver who enjoys painting.
18, 37, No, David Lee is a 37 year old black bartender who enjoys travel.
19, 38, Yes, Sophia Rodriguez is a 38 year old caucasian nurse who enjoys painting.
20, 32, Yes, Hank Rodriguez is a 32 year old black architect who enjoys cooking.
21, 29, Yes, Sophia Smith is a 29 year old caucasian bartender who enjoys travel.
22, 31, Yes, Aaron Carter is a 31 year old caucasian architect who enjoys cooking.
23, 29, No, Mia Lee is a 29 year old caucasian teacher who enjoys painting.
24, 29, Yes, Aaron Rodriguez is a 29 year old caucasian bartender who enjoys travel.
25, 28, Yes, Sam Smith is a 28 year old caucasian truck driver who enjoys medieval history.
26, 30, Yes, Anna Lee is a 30 year old black bartender who enjoys travel.
```

Data File (CSV)

Batch File (JSONL)

```
openai-batch-example / {} batch-requests.jsonl
{"custom_id": "id_0", "method": "POST", "url": "/v1/chat/completions", "body": {"model": "gpt-4o", "messages": [{"content": [{"type": "text", "text": "You are a participant in a marketing study from Monkeys Elbow, Kentucky. He Hank Rodriguez is a 28 year old black architect who enjoys Star Wars."}], {"role": "user", "content": [{"type": "text", "text": "Please select the image you find the most appealing from these three and explain your choice. The image on the left can be called 'kawaii', the image in the middle 'cubs', and the image on the right 'futuristic'."}], {"type": "image_url", "image_url": [{"url": "https://kellogg-shared.s3.us-east-2.amazonaws.com/images/select_enhanced.png"}]}], "temperature": 1, "max_tokens": 2840, "top_p": 1, "frequency_penalty": 0, "presence_penalty": 0, "response_format": {"type": "text"}}}, {"custom_id": "id_1", "method": "POST", "url": "/v1/chat/completions", "body": {"model": "gpt-4o", "messages": [{"content": [{"type": "text", "text": "You are a participant in a marketing study from Monkeys Elbow, Kentucky. He David Carter is a 45 year old black nurse who enjoys medieval history."}], {"role": "user", "content": [{"type": "text", "text": "Please select the image you find the most appealing from these three and explain your choice. The image on the left can be called 'kawaii', the image in the middle 'cubs', and the image on the right 'futuristic'."}], {"type": "image_url", "image_url": [{"url": "https://kellogg-shared.s3.us-east-2.amazonaws.com/images/select_enhanced.png"}]}], "temperature": 1, "max_tokens": 2840, "top_p": 1, "frequency_penalty": 0, "presence_penalty": 0, "response_format": {"type": "text"}}}, {"custom_id": "id_2", "method": "POST", "url": "/v1/chat/completions", "body": {"model": "gpt-4o", "messages": [{"content": [{"type": "text", "text": "You are a participant in a marketing study from Monkeys Elbow, Kentucky. He David Smith is a 39 year old caucasian architect who enjoys medieval history."}], {"role": "user", "content": [{"type": "text", "text": "Please select the image you find the most appealing from these three and explain your choice. The image on the left can be called 'kawaii', the image in the middle 'cubs', and the image on the right 'futuristic'."}], {"type": "image_url", "image_url": [{"url": "https://kellogg-shared.s3.us-east-2.amazonaws.com/images/select_enhanced.png"}]}], "temperature": 1, "max_tokens": 2840, "top_p": 1, "frequency_penalty": 0, "presence_penalty": 0, "response_format": {"type": "text"}}}, {"custom_id": "id_3", "method": "POST", "url": "/v1/chat/completions", "body": {"model": "gpt-4o", "messages": [{"content": [{"type": "text", "text": "You are a participant in a marketing study from Monkeys Elbow, Kentucky. He Hank Rodriguez is a 38 year old black architect who enjoys cooking."}], {"role": "user", "content": [{"type": "text", "text": "Please select the image you find the most appealing from these three and explain your choice. The image on the left can be called 'kawaii', the image in the middle 'cubs', and the image on the right 'futuristic'."}], {"type": "image_url", "image_url": [{"url": "https://kellogg-shared.s3.us-east-2.amazonaws.com/images/select_enhanced.png"}]}], "temperature": 1, "max_tokens": 2840, "top_p": 1, "frequency_penalty": 0, "presence_penalty": 0, "response_format": {"type": "text"}}}]
```

```
batch-responses.jsonl
(llm-env) openai-batch --help
Usage: openai-batch COMMAND

Commands to create, upload, and execute batch jobs using the OpenAI API

Commands
get-batch-results Download batch results to a file if the batch job is completed. If
list-batches List all OpenAI batches for your account
make-batch-file Make a batch file for uploading to OpenAI
start-batch Start a batch job on OpenAI
upload-batch-file Upload a batch file to OpenAI
utils Utility commands for supporting batch jobs
--help, -h Display this message and exit.
--version Display application version.

(llm-env) openai-batch make-batch-file config.json simulated_survey_personas.csv
Processing... 100% 0:00:00
Batch file created: batch-requests.jsonl
```

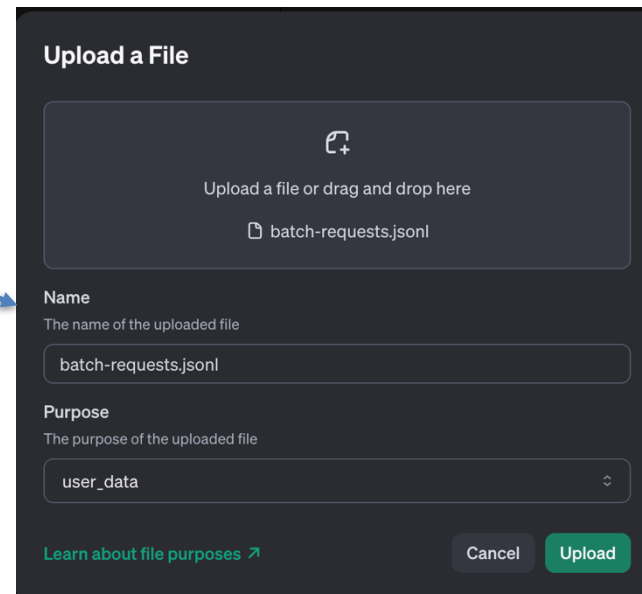
openai-batch command line interface

Uploading Batch Files

```
$ openai-batch upload-batch-file batch-requests.jsonl
Uploaded batch file: batch-requests.jsonl
FileObject(id='file-poped6fkTfRDVQutlBjmXPyi', bytes=88120, created_at=1728602591, filename='batch-requests.jsonl', object='file', purpose='batch', status='processed', status_details=None)
$
```

You can either use the openai-batch command line interface

Or the OpenAI web-based console



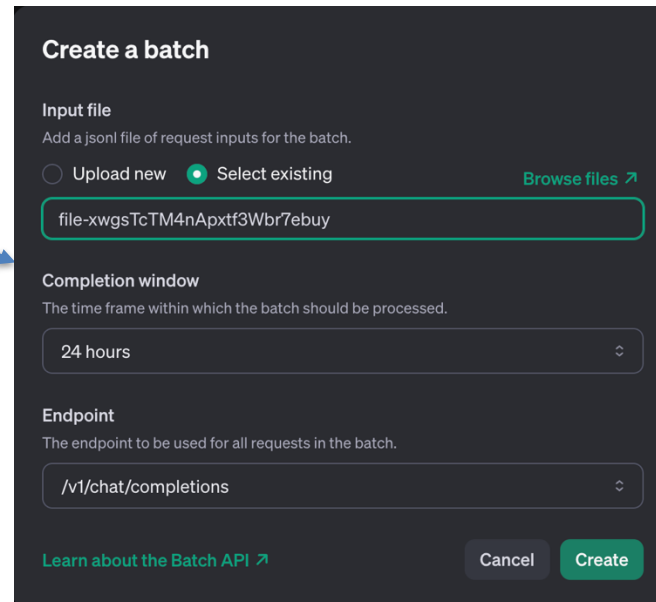
The screenshot shows the 'Upload a File' interface in the OpenAI console. It features a dark theme with a central upload area containing a plus icon and the text 'Upload a file or drag and drop here'. Below this, the filename 'batch-requests.jsonl' is displayed. The 'Name' field is pre-filled with 'batch-requests.jsonl' and has a description 'The name of the uploaded file'. The 'Purpose' field is a dropdown menu set to 'user_data' with a description 'The purpose of the uploaded file'. At the bottom, there is a link 'Learn about file purposes' and two buttons: 'Cancel' and 'Upload'.

Starting Batch Jobs

```
$ openai-batch start-batch file-poped6fkTfRDVQutLBjmXPyi
Batch(
  id='batch_6708634c0ba081909ca21a5f182ce840',
  completion_window='24h',
  created_at=1728602956,
  endpoint='/v1/chat/completions',
  input_file_id='file-poped6fkTfRDVQutLBjmXPyi',
  object='batch',
  status='validating',
  cancelled_at=None,
  cancelling_at=None,
  completed_at=None,
  error_file_id=None,
  errors=None,
  expired_at=None,
  expires_at=1728689356,
  failed_at=None,
  finalizing_at=None,
  in_progress_at=None,
  metadata={'description': 'batch job'},
  output_file_id=None,
  request_counts=BatchRequestCounts(completed=0, failed=0, total=0)
)
```

You can either use the openai-batch command line interface

Or the OpenAI web-based console



Create a batch

Input file
Add a jsonl file of request inputs for the batch.

☐ Upload new ☒ Select existing [Browse files ↗](#)

file-xwgsTcTM4nApxtf3Wbr7ebuy

Completion window
The time frame within which the batch should be processed.

24 hours

Endpoint
The endpoint to be used for all requests in the batch.

/v1/chat/completions

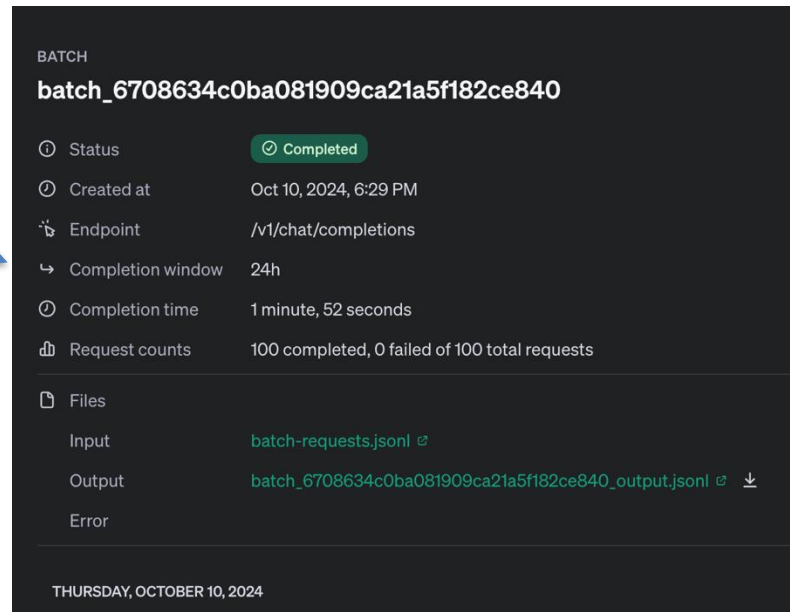
[Learn about the Batch API ↗](#)

Accessing Batch Results

```
$ openai-batch get-batch-results batch_6708634c0ba081909ca21a5f182ce840
Batch(
  id='batch_6708634c0ba081909ca21a5f182ce840',
  completion_window='24h',
  created_at=1728602956,
  endpoint='/v1/chat/completions',
  input_file_id='file-poped6fkTfRDVQutlBjmXPyi',
  object='batch',
  status='completed',
  cancelled_at=None,
  cancelling_at=None,
  completed_at=1728603068,
  error_file_id=None,
  errors=None,
  expired_at=None,
  expires_at=1728689356,
  failed_at=None,
  finalizing_at=1728603062,
  in_progress_at=1728602957,
  metadata={'description': 'batch job'},
  output_file_id='file-efTbwFbPPvqMFDvMjLXqIp7d',
  request_counts=BatchRequestCounts(completed=100, failed=0, total=100)
)
writing json output to batch-responses.jsonl
```

You can either use the
openai-batch command
line interface

Or the OpenAI web-based console



BATCH

batch_6708634c0ba081909ca21a5f182ce840

Status	Completed
Created at	Oct 10, 2024, 6:29 PM
Endpoint	/v1/chat/completions
Completion window	24h
Completion time	1 minute, 52 seconds
Request counts	100 completed, 0 failed of 100 total requests
Files	
Input	batch-requests.jsonl
Output	batch_6708634c0ba081909ca21a5f182ce840_output.jsonl ↓
Error	

THURSDAY, OCTOBER 10, 2024

Batch Results

```
1 [{"id": "batch_req_670863b63f4c8190a8ff330fe47232d4", "custom_id": "id_0", "response": {"status_code": 200, "request_id":  
"5592039fb02417b4c6efe26894b25c5b", "body": {"id": "chatcmpl-AGwsrtHDAXWY0QCFmbRCzi9LIusq", "object": "chat.completion", "created":  
1728602961, "model": "gpt-4o-2024-05-13", "choices": [{"index": 0, "message": {"role": "assistant", "content": "As Hank Rodriguez, a  
28-year-old black architect who enjoys Star Wars, I find the \"futuristic\" image on the right the most appealing. This image  
resonates with my interest in science fiction and futuristic themes, which align closely with the Star Wars aesthetic. The modern and  
sleek design, combined with the dramatic and advanced cityscape in the background, captures a sense of innovation and forward-thinking  
that appeals to me both professionally as an architect and personally as a sci-fi enthusiast. The overall vibe of the image aligns  
well with my tastes and interests."}, "refusal": null}, "logprobs": null, "finish_reason": "stop"}], "usage": {"prompt_tokens": 1541,  
"completion_tokens": 111, "total_tokens": 1652, "prompt_tokens_details": {"cached_tokens": 0}, "completion_tokens_details":  
{"text_tokens": 0, "reasoning_tokens": 0, "audio_tokens": 0}}, "system_fingerprint": "fp_568f3bfff7"}, {"error": null}  
2 [{"id": "batch_req_670863b64bb48190ad3ce268b65e2f19", "custom_id": "id_1", "response": {"status_code": 200, "request_id":  
"a43771a9aa0b8950cad0d55b05a782d8", "body": {"id": "chatcmpl-AGwsv0QFULUbh6SFpnozHtYmptWRV", "object": "chat.completion", "created":  
1728602965, "model": "gpt-4o-2024-05-13", "choices": [{"index": 0, "message": {"role": "assistant", "content": "I find the  
\"futuristic\" image on the right the most appealing. This image stands out due to its sleek and modern aesthetic, which suggests a  
forward-thinking and innovative approach. The contrast between the industrial, science fiction-inspired background and the familiar,  
iconic red Coca-Cola can creates a striking visual effect that draws my attention. Additionally, the unique setting and design  
elements associated with the future fascinate me as they evoke a sense of exploration and curiosity about what lies ahead."},  
"refusal": null}, "logprobs": null, "finish_reason": "stop"}], "usage": {"prompt_tokens": 1541, "completion_tokens": 95,  
"total_tokens": 1636, "prompt_tokens_details": {"cached_tokens": 0}, "completion_tokens_details": {"text_tokens": 0,  
"reasoning_tokens": 0, "audio_tokens": 0}}, "system_fingerprint": "fp_568f3bfff7"}, {"error": null}  
3 [{"id": "batch_req_670863b660988190aab69843b88a23a5", "custom_id": "id_2", "response": {"status_code": 200, "request_id":  
"330a93bccb0698c2d39d6f22af54ae45", "body": {"id": "chatcmpl-AGwsymu5YpDwzWuKQgr7P7jF4KMpg", "object": "chat.completion", "created":  
1728602968, "model": "gpt-4o-2024-05-13", "choices": [{"index": 0, "message": {"role": "assistant", "content": "As a 39-year-old  
architect with an interest in medieval history, I find the image on the right (\"futuristic\") the most appealing. It presents a  
stark, intriguing contrast between the modern Coca-Cola can and an elaborate, otherworldly landscape. The architectural elements in  
the background evoke thoughts of grand, monumental structures, perhaps correlating with my fascination for the grand designs of  
medieval architecture. This image stimulates imagination and contemplation, resonating with both my professional interest in  
architecture and my personal interest in rich, historical narratives.\n\nAdditionally, the other images, while vibrant and fun, don't  
quite align with my tastes as much as the futuristic one does. The \"kawaii\" image is playful and cute but feels more suited to a  
younger audience or those with an interest in cute, animated designs. The middle image (\"cubs\") evokes a sense of a classic and  
lively setting, but it doesn't capture the same intrigue and architectural allure that draws me to the futuristic image."}, "refusal":  
null}, "logprobs": null, "finish_reason": "stop"}], "usage": {"prompt_tokens": 1542, "completion_tokens": 195, "total_tokens": 1737,  
"prompt_tokens_details": {"cached_tokens": 0}, "completion_tokens_details": {"text_tokens": 0, "reasoning_tokens": 0, "audio_tokens":  
0}}, "system_fingerprint": "fp_568f3bfff7"}, {"error": null}]
```