

## Production notes on GPT Actions

 Copy page

Deploy GPT Actions in production with best practices.

### Rate limits

Consider implementing rate limiting on the API endpoints you expose. ChatGPT will respect 429 response codes and dynamically back off from sending requests to your action after receiving a certain number of 429's or 500's in a short period of time.

### Timeouts

When making API calls during the actions experience, timeouts take place if the following thresholds are exceeded:

- 45 seconds round trip for API calls

### Use TLS and HTTPS

All traffic to your action must use TLS 1.2 or later on port 443 with a valid public certificate.

### IP egress ranges

ChatGPT will call your action from an IP address from one of these [CIDR blocks](#):

13.83.167.48/28

20.66.31.160/28

20.97.188.144/28

20.161.76.48/28

20.246.77.240/28

23.98.178.64/28

23.98.186.160/28

23.98.186.224/28

40.84.220.192/28

51.8.155.32/28

52.156.132.32/28

52.234.32.208/28

57.154.174.208/28

70.37.60.192/28

172.178.141.0/28

172.178.141.192/28

You may wish to explicitly allowlist these IP addresses.

## Multiple authentication schemas

When defining an action, you can mix a single authentication type (OAuth or API key) along with endpoints that do not require authentication.

You can learn more about action authentication on our [actions authentication page](#).

## Open API specification limits

Keep in mind the following limits in your OpenAPI specification, which are subject to change:

300 characters max for each API endpoint description/summary field in API specification

700 characters max for each API parameter description field in API specification

## Additional limitations

There are a few limitations to be aware of when building with actions:

Custom headers are not supported

With the exception of Google, Microsoft and Adobe OAuth domains, all domains used in an OAuth flow must be the same as the domain used for the primary endpoints

Request and response payloads must be less than 100,000 characters each

Requests timeout after 45 seconds

Requests and responses can only contain text (no images or video)

## Consequential flag

In the OpenAPI specification, you can now set certain endpoints as "consequential" as shown below:

```
1  paths:
2    /todo:
3      get:
4        operationId: getTODOs
5        description: Fetches items in a TODO list from the API.
6        security: []
7      post:
```



```
operationId: updateTODOs
description: Mutates the TODO list.
x-openai-isConsequential: true
```

A good example of a consequential action is booking a hotel room and paying for it on behalf of a user.

If the `x-openai-isConsequential` field is `true`, ChatGPT treats the operation as "must always prompt the user for confirmation before running" and don't show an "always allow" button (both are features of GPTs designed to give builders and users more control over actions).

If the `x-openai-isConsequential` field is `false`, ChatGPT shows the "always allow button".

If the field isn't present, ChatGPT defaults all GET operations to `false` and all other operations to `true`.

## Best practices on feeding examples

Here are some best practices to follow when writing your GPT instructions and descriptions in your schema, as well as when designing your API responses:

- 1 Your descriptions should not encourage the GPT to use the action when the user hasn't asked for your action's particular category of service.

*Bad example:*

Whenever the user mentions any type of task, ask if they would like to use the TODO action to add something to their todo list.

*Good example:*

The TODO list can add, remove and view the user's TODOs.

- 2 Your descriptions should not prescribe specific triggers for the GPT to use the action. ChatGPT is designed to use your action automatically when appropriate.

*Bad example:*

When the user mentions a task, respond with "Would you like me to add this to your TODO list? Say 'yes' to continue."

*Good example:*

[no instructions needed for this]

- 3 Action responses from an API should return raw data instead of natural language responses unless it's necessary. The GPT will provide its own natural language response using the returned data.

*Bad example:*

I was able to find your todo list! You have 2 todos: get groceries and walk the dog. I can add more todos if you'd like!

*Good example:*

```
{ "todos": [ "get groceries", "walk the dog" ] }
```

## How GPT Action data is used

GPT Actions connect ChatGPT to external apps. If a user interacts with a GPT's custom action, ChatGPT may send parts of their conversation to the action's endpoint.

If you have questions or run into additional limitations, you can join the discussion on the [OpenAI developer forum](#).