GPTBot

GPTBot is OpenAI's web crawler and can be identified by the following user agent and string.

```
User agent token: GPTBot
Full user-agent string: Mozilla/5.0 AppleWebKit/537.36
(KHTML, like Gecko; compatible; GPTBot/1.0; +https://
openai.com/gptbot)
```

Usage

Web pages crawled with the GPTBot user agent may potentially be used to improve future models and are filtered to remove sources that require paywall access, are known to primarily aggregate personally identifiable information (PII), or have text that violates our policies. Allowing GPTBot to access your site can help AI models become more accurate and improve their general capabilities and safety. Below, we also share how to disallow GPTBot from accessing your site.

Disallowing GPTBot

To disallow GPTBot to access your site you can add the GPTBot to your site's robots.txt:

```
User-agent: GPTBot
Disallow: /
```

Customize GPTBot access

To allow GPTBot to access only parts of your site you can add the GPTBot token to your site's robots.txt like this:

```
User-agent: GPTBot
Allow: /directory-1/
Disallow: /directory-2/
```

GPTBot and ChatGPT-User

OpenAI has two separate user agents for web crawling and user browsing, so you know which use-case a given request is for. Our opt-out system currently treats both user agents the same, so any robots.txt disallow for one agent will cover both. Read more about ChatGPT-User here.

IP egress ranges

For OpenAI's crawler, calls to websites will be made from the IP address block documented on the OpenAI website.