interweb

Gerado a 15/11/2026 00:27

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

Bots de IA a visitar o seu site: robots.txt e como os bloquear

IA no seu site

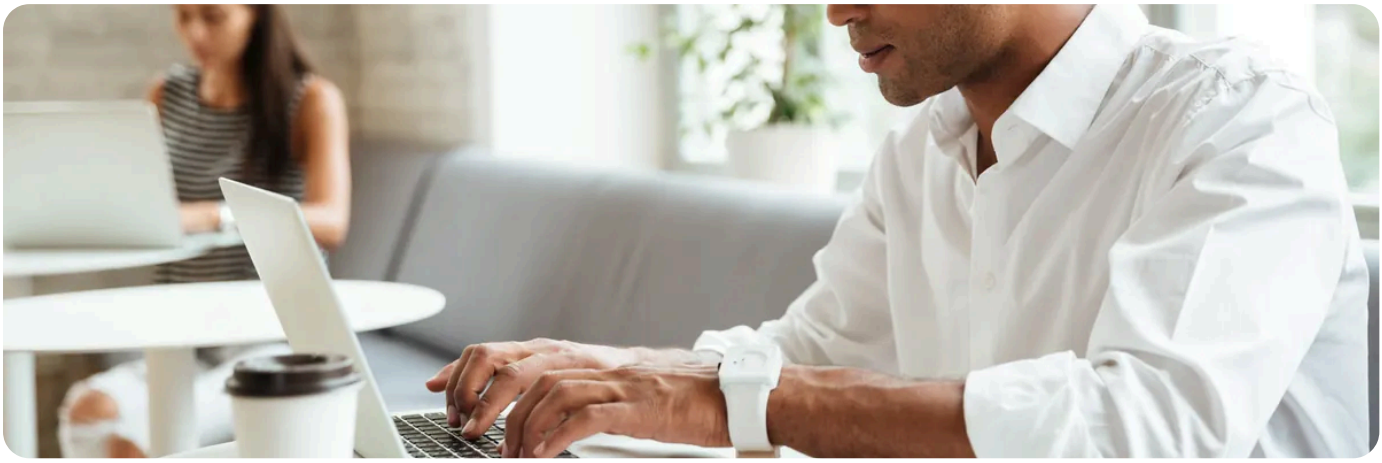
CÓDIGO	VERSÃO	PAÍS · IDIOMA
RLX-AO-KB-950	v2026.10.10	Angola · Português

Interweb Angola
RLX Angola, Lda. · Grupo RLX Holdings
NIF 500 260 5724 · Matrícula 26464-25/250724

https://www.meu.interweb.ao/knowledgebase/950

Bots de IA a visitar o seu site: robots.txt e como os bloquear

IA no seu site · Criado em **10 de outubro de 2026** · 3 min de leitura



Os fornecedores de IA usam robôs (*crawlers*) que percorrem sites para recolher texto, seja para treinar modelos, seja para responder a perguntas em tempo real. Quem não quer isso pode pedir-lhes que não entrem com o **robots.txt**, e pode **recusar o acesso no servidor** pelo nome com que se identificam. O primeiro é um pedido de boas maneiras; o segundo é uma porta fechada, mas só vale contra quem diz a verdade sobre quem é.

Primeiro, veja quem lá anda

1. **Abra os registos de acesso** do site no cPanel e procure o campo do «user agent» (o nome com que o visitante se apresenta).

2. **Procure nomes que reconheça.** Os fornecedores publicam os nomes dos seus robôs na documentação deles: por exemplo, GPTBot (OpenAI), ClaudeBot (Anthropic), CCBot (Common Crawl) e PerplexityBot. Os nomes mudam, por isso confirme sempre na documentação de cada um.

3. **Decida o que quer.** Quem bloqueia deixa de poder aparecer em respostas desses serviços. Quem não bloqueia aceita que o texto seja lido.

Pedir com o robots.txt

O ficheiro `robots.txt` fica na raiz do domínio (`asuaempresa.ao/robots.txt`). Por cada robô, uma linha a dizer quem é e outra a dizer o que não pode ver:

```
User-agent: GPTBot
Disallow: /

User-agent: ClaudeBot
Disallow: /
```

```
User-agent: CCBot
Disallow: /
```

O Google tem um nome à parte, o `Google-Extended`, que, segundo as orientações públicas do Google, controla se o conteúdo é usado para os seus produtos de IA generativa. **Não bloqueie o `Googlebot`**, que é o que indexa o site na pesquisa. Veja **o que fazem o robots.txt e o sitemap**.

Recusar no servidor

O robots.txt é voluntário: cada robô cumpre se quiser. Para recusar de facto, num site em hospedagem partilhada que use o ficheiro `.htaccess`, pode negar o acesso pelo nome:

```
RewriteEngine On
RewriteCond %{HTTP_USER_AGENT} (GPTBot|ClaudeBot|CCBot) [NC]
RewriteRule .* - [F,L]
```

Faça cópia do .htaccess antes de o editar,
e confirme que o site continua a abrir. Um erro de sintaxe põe o site inteiro a dar erro 500. Se o site já tiver outras regras, junte estas sem apagar as existentes. Veja **erros de site explicados**.

Nem o robots.txt nem o bloqueio por nome travam quem mente.
Há robôs que se apresentam como um navegador normal. Para esses, o caminho é bloquear por endereço IP ou usar uma camada de protecção à frente do site, como o Cloudflare. Veja **bloquear um endereço IP** e **o que é o Cloudflare**.

Se o problema é o consumo, e não a privacidade, meça primeiro quanto pedem. Um robô educado raramente pesa no servidor; um agressivo pode pesar, e esse não cumpre o robots.txt.

PRECISA DE AJUDA?

O seu site está lento ou com erros por causa de tráfego estranho? Abra um pedido com a hora e o endereço e vemos o que consta nos registos.

Abrir um pedido de suporte →

VEJA TAMBÉM

robots.txt e sitemap.xml: o que fazem e como os verificar →

Bloquear um endereço IP de visitar o seu site →

Cloudflare: o que é, o que muda e o que fica igual →

interweb

Interweb Angola

RLX Angola, Lda. · Grupo RLX Holdings

NIF	500 260 5724
MATRÍCULA	26464-25/250724
SEDE	Rua Direita do Patriota, Bairro Dangereux Belas, Luanda, Angola
CONTACTOS	info@interweb.ao · +244 942 011 191 · www.interweb.ao
HORÁRIO	2ª. à 6ª. Feira - 08 à 16 Horas Sábado & Domingo - Fechado

RLX-AO-KB-950 · v2026.10.10 · Gerado a 15/11/2026 00:27

Versão online deste documento