

Capturer gptbot et lui faire indexer ce que l'on veut

Oros 2024-06

Moi

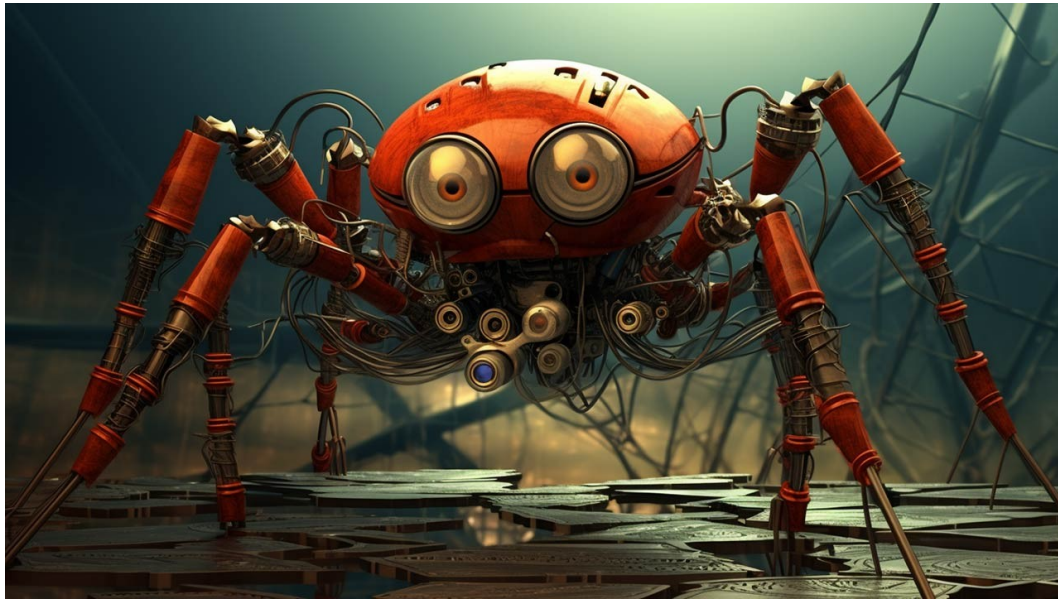
- Oros
- Hacker DevSecOps
- Breizh Entropy, Hackerspace de Rennes

Capturer gptbot et lui faire indexer ce que l'on veut.



Qu'est-ce que gptbot ?

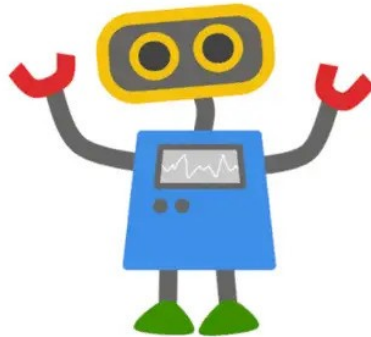
- Le crawler d'openAI qui se nourrit d'internet pour alimenter chatGPT.



crawler ?

- Un robot qui parcours le web pour l'indexer.
- Exemple :

Googlebot

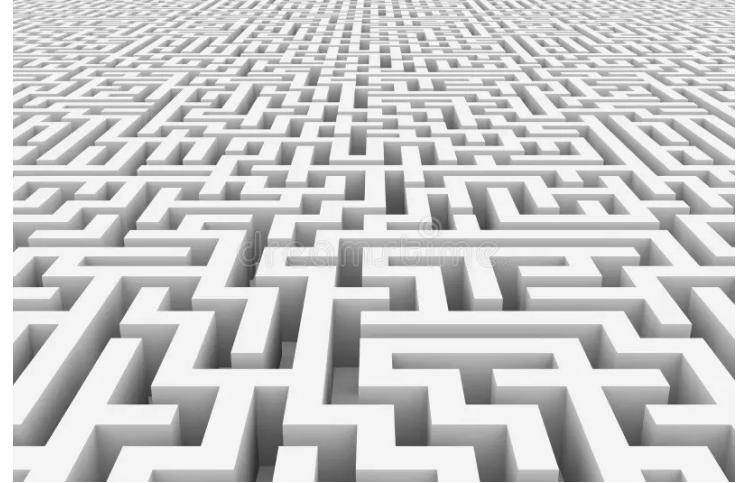


Création du piège

infini.php

```
```php
<?php
for ($i=0; $i < 100; $i++) {
 printf('%d
', rand(), rand());
 // affiche 1325428233

}
```
```



Le piège

Une boucle avec du texte random qui point sur la page elle-même.

```
```html
1224277605

1661310683

1621603347

[...]
```

```
https://exemple.com/infini.php?r=1592157817
https://exemple.com/infini.php?r=1713975801
https://exemple.com/infini.php?r=1664556516
```

# Camouflage

---

Conf nginx

```

```
server {  
    [...]  
    location /infinity/ {  
        rewrite ^(/infinity/.*)$ /infinity.php?p=$1 last;  
    }  
    [...]  
}
```

```

https://exemple.com/infinity/<blablabla>

== https://exemple.com/infinity.php?p=<blablabla>



# Camouflage

---

```
infinity.php
```

```
```
```

```
<?php
```

```
// nos liens random
```

```
for ($i=0; $i < random_int(10,2000); $i++) {
```

```
    $bytes = bin2hex(random_bytes(20));
```

```
    echo '<a href="/infinity/'.$bytes.'">'.$bytes.'</a><br>\n';
```

```
}
```

```
// du texte random pour l'entropy
```

```
for ($i=0; $i < random_int(10,20000); $i++) {
```

```
    // du text bidon
```

```
    echo myGenerateRandomString();
```

```
}
```

```
// Texte de notre truc à indexer
```

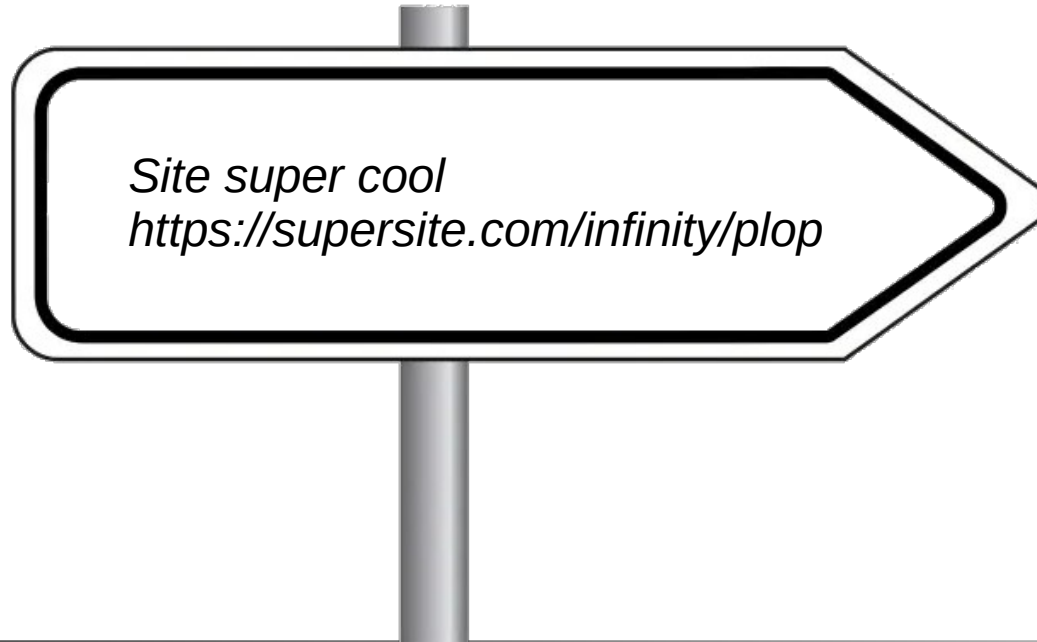
```
echo myPayload();
```

```
```
```

# L'appât

---

Poster le lien de notre site infini sur un site qui a du trafic



# Check des logs

---

...

```
20.171.206.141 - - [21/Aug/2024:13:42:59 +0000] "GET
/infinity/e9980a8440ba7f1c0940c623eb5016e43a409490 HTTP/1.1" 200 379 "-"
"Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko; compatible; GPTBot/1.2;
+https://openai.com/gptbot)"
```

...

\o/

# Alimenter le bot


---

Pensez à mettre à jour myPayload()

# Bots capturés

---

Index 6 à 7 pages / min !!!  
Non-stop



Googlebot/2.1

GPTBot/1.2

AhrefsBot/7.0

MJ12bot/v1.4.8

YandexBot/3.0

facebookexternalhit/1.1

bingbot/2.0

DuckDuckBot/1.1

Googlebot-Image/1.0

coccocbot-web/1.0

Google-Read-Aloud

GoogleOther

meta-externalagent/1.1

YandexImages/3.0

YandexRenderResourcesBot/1.0

wpbot/1.1

SemrushBot/7~bl

GPTBot/1.0

Python-urllib/3.8

DuckDuckBot-Https/1.1

NetcraftSurveyAgent/1.0

# Merci

---

Amusez-vous bien ;-)



Oros <https://ecirtam.net>