



Se connecter

Proposer un contenu

Identifiant

Identifiant

Mot de passe

Mot de passe

☐ Connexion automatique

Se connecter

Pas de compte ? S'inscrire...

Lien : Nouvel effet secondaire pervers du pillage massif du web par les entreprises d'IA (robots BnF empêchés de faire leur travail)

Posté par Ysabeau (courriel, site web personnel, Mastodon) le 16 mars 2026 à 14:42.

Étiquettes : bnf, intelligence_artificielle, merdification, archivage, grands_modèles_de_langage



<https://digipres.club/@BertrandCaron/116239>

(13 commentaires).

16 mar. 2026

9

moui

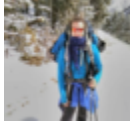
Posté par ChocolatineFlying le 16 mars 2026 à 14:55.

Évalué à -4 (+2/-7).

[^] # Re: moui

Posté par jtreemesay (site web personnel) le 16 mars 2026 à 15:00. Évalué à 6 (+5/-1).

Copier n'est pas voler
Voler un truc, c'est l'retirer
Le copier c'est l'ajouter
Ça sert à ça copier



Copier n'est pas voler
Si je copie le tiens, tu n'le perds pas
Un pour moi et un pour toi
Copier ça sert à ça

Si je vole ton vélo, tu dois prendre le bus
Si je ne fais que le copier, y'a un cycliste en plus !

[source](#)

Répondre

[^] # Re: moui

Posté par Ysabeau (courriel, site web personnel, Mastodon) le 16 mars 2026 à 15:15. Évalué à 4 (+1/-0).

Et en même temps toutes les plateformes qui proposent du texte en accès ouvert, comme Gallica BnF subissent en même temps du [spam très agressif](#).



Ce qui fait qu'elles subissent des problèmes des deux côtés de leurs mission et sont empêchées de la mener à bien pour ça.

--

Je n'ai aucun avis sur systemd

Répondre

[^] # Re: moui

Posté par Faya le 16 mars 2026 à 15:17. Évalué à 3 (+1/-0).

Si je copie le tiens, tu n'le perds pas
Un pour moi et un pour toi



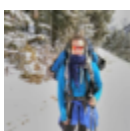
Ouais, enfin là c'est *"un pour toi si tu te déplaces physiquement jusqu'à chez moi et que tu ne repars pas avec"*... Pas vraiment le même esprit que la Wayback Machine.

Répondre

[^] # Re: moui

Posté par jtreemesay (site web personnel) le 16 mars 2026 à 15:23. Évalué à 3 (+2/-1).

j'venais juste pour dire que la contrefaçon n'est pas du vol.



Répondre

[^] # Re: moui

Posté par ChocolatineFlying le 16 mars 2026 à 15:58. Évalué à 0 (+1/-2).

marrant que vous me sautez tous dessus sur ce point que j'aurais du mettre en (humour) mais que le déplacement physique pour regarder une page internet ne vous choque pas plus que cela surtout avec la raison invoqué.



~_(ツ)_/~

heureusement qu'internet archive ne fais pas cela, je vous rappel que gallica estam-pille tous les document scannée avec leur jolie copyright dans leur cgu et interdit la réutilisation copie etc ... , je ne vais pas pleurer sur leur sort pour ma part.

Répondre

[^] # Re: moui

Posté par Ysabeau (courriel, site web personnel, Mastodon) le 16 mars 2026 à 15:16. Évalué à 9 (+7/-1).

Cela s'inscrit dans le cadre de leur mission, comme le dépôt légal et, comme cela a été dit, ce n'est pas du vol.



--

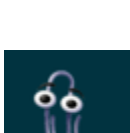
Je n'ai aucun avis sur systemd

Répondre

[^] # Re: moui

Posté par François Chaix (Mastodon) le 16 mars 2026 à 15:37. Évalué à 8 (+7/-0).

La notion de vol n'est pas la bonne, pour un truc public.



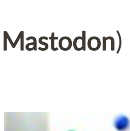
Là, la bonne analogie, ce serait plutôt la horde de gens qui viennent à 500 à ta journée portes ouvertes, avec de la m*** plein les bottes, qui ouvrent tous les robinets de la maison et s'essuient dans tes rideaux, avant de repartir sans t'avoir dit bonjour.

Répondre

[^] # Re: moui

Posté par pulkomandy (site web personnel, Mastodon) le 16 mars 2026 à 17:38. Évalué à 10 (+8/-0).

La BNF collecte toutes sortes de choses, on en avait entendu parler par exemple lors de la fermeture des Skyblogs. Ils ont été archivés mais contiennent beaucoup de données personnelles (photos, etc). Il faut donc trouver l'équilibre entre la préservation de ces pages, et la protection de la vie privée des personnes concernées, le droit à l'oubli, etc.



La solution est de préserver les pages mais de ne pas les rendre accessibles à tout le monde sur internet.

Cette approche est complémentaire de celle d'autres projets comme archive.org, qui eux font le choix de tout exposer publiquement, et s'exposent à des demandes de suppression de données et à diverses attaques en justice pour cette raison. C'est très pratique d'avoir archive.org, et le mode de fonctionnement de la BNF est un peu "vieille école", certes. Mais finalement, les deux se complètent assez bien. Le jour où internet n'existera plus, si la BNF a bien fait son travail, les pages archivées seront toujours consultables. Tout comme l'INA préserve ce qui a été diffusé à la radio et à la télévision, mais ne met pas tout en consultation gratuite, par exemple.

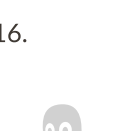
Répondre

Je confirme

Posté par Andréas Livet le 17 mars 2026 à 10:16.

Évalué à 2 (+1/-0).

Actuellement je travaille pour le réseau ISSN (qui travaille avec la BnF) et on vient de sortir notre nouveau portail (portal.issn.org), on s'est mangé un trafic totalement hallucinant provenant de bot IA pas très gentils et utilisant des IP domestiques par dizaines de milliers.



Résultat, notre infra est très vite tombée et on galère à mettre des protections efficaces... On a du genre 500000 r/heure sur un site qui fait des recherches elasticsearch assez lourdes, c'est beaucoup trop pour notre "petite" infra.

Répondre

[^] # Re: Je confirme
Posté par fearan le 17 mars 2026 à 10:55. Évalué à 3 (+0/-0).

les trucs style Anubis ne fonctionnent pas ?



Sinon si tu arrives à les détecter, un sink infini de génération de texte aléatoire pour les empoisonner. (faut bien penser à configurer les robot.txt pour éviter d'empoisonner ceux qui sont bien configuré)

Une autre solution pour les envoyer dans le sink sans identification c'est un lien non visible par les humains, arrivant très tôt dans la page, voir sur une page de garde.

--

Il ne faut pas décorner les boeufs avant d'avoir semé le vent

Répondre

[^] # Re: Je confirme

Posté par pulkomandy (site web personnel, Mastodon) le 17 mars 2026 à 11:35. Évalué à 3 (+0/-0).

Une autre solution pour les envoyer dans le sink sans identification c'est un lien non visible par les humains, arrivant très tôt dans la page, voir sur une page de garde.



Moi j'ai mis un lien visible dont le texte est à peu près "ne cliquez pas ici, ceci est une protection contre les robots, vous allez vous faire bannir du serveur".

Mais les requêtes proviennent d'un tellement grand nombre d'IPs que le bannissement (par cette méthode et d'autres) reste peu efficace. Chaque IP fait moins de requêtes qu'un visiteur normal du site chargeant la page d'accueil et les quelques ressources (images, css, ...) qui y sont liées.

Répondre

[^] # Re: Je confirme

Posté par Andréas Livet le 17 mars 2026 à 15:41. Évalué à 2 (+1/-0).

On a pas encore testé d'installer Anubis, mais ça va peut-être une solution, bien que je n'apprécie pas bien ce genre de mécanismes qui empêche justement d'autres bots tout à fait légitimes...



Répondre

Envoyer un commentaire

Suivre le flux des commentaires

Note : les commentaires appartiennent à celles et ceux qui les ont postés. Nous n'en sommes pas responsables.

Revenir en haut de page

Derniers commentaires

- Site web à la c...
- Re: dépôt github
- Re: dépôt github
- Re: Cluster reader
- Re: Et les moules ?
- Re: IPv4 agnostique ...
- Re: IPv4 agnostique ...
- Re: Et les moules ?
- Le trop méconnu bu...
- Re: Et les moules ?
- Re: Et les moules ?
- Re: Neutralisation d...

Étiquettes (tags) populaires

- intelligence_artificielle
- merdification
- états-unis
- grands_modèles_de...
- administration_fran...
- bigtech
- capitalisme_de_surv...
- politique
- capitalisme
- sortie_version
- google
- souveraineté_numer...

Sites amis

- April
- Agenda du Libre
- Framasoft
- Éditions D-BookeR
- Éditions Eyrolles
- Éditions Diamond
- Éditions ENI
- La Quadrature du Net
- Lea-Linux
- En Vente Libre
- Grafik Plus
- Open Source Initiative

À propos de LinuxFr.org

- Mentions légales
- Faire un don
- L'équipe de LinuxFr...
- Informations sur le s...
- Aide / Foire aux que...
- Suivi des suggestion...
- Wiki du site
- Règles de modération
- Statistiques
- API pour le développ...
- Code source du site
- Plan du site