

# Les vraies questions que pose l'affaire du plagiat par IA NewsDayFR.com



**François Avril**

Digital Media Specialist

9 février 2025

*Article écrit le 9 février 2025 avant la condamnation de NewsDayFRCom par le tribunal judiciaire de Paris, le 7 mai 2025 (cf. [Libération](#))*

A l'évidence, le site [NewsDayFR.com](#) est un site de plagiat mal exécuté qui réécrit à l'aide de l'IA générative des informations piochées dans la presse nationale, locale et étrangère sans lien vers les contenus originaux.

Mais pour autant, est-il vraiment nécessaire de [déployer une telle artillerie](#) pour un site à l'audience digne d'un blog (141k visites en décembre 2024 et 92k en janvier 2025 [selon les données SimilarWeb](#)).

Si on rapporte aux 6000 articles pillés, cela représente moins de 20 visites par article. On est loin de la menace systémique !

100k visites mensuelles, c'est bien peu par rapport à l'audience des médias dont le contenu est repris. A titre d'exemple Liberation totalise, chaque mois, [plus de 24M de visites](#).

Alors oui, ce site mérite d'être fermé sans délai, mais cette affaire pose d'autres questions plus systémiques auxquelles il serait intéressant de répondre si les médias ne veulent pas se retrouver relégués au rang de fournisseur de commodités par la disruption provoquée par l'IA générative.

Pour illustrer les possibilités de l'IA, j'ai demandé au principaux modèles de résumer mon article et de donner leur avis. N'hésitez pas à consulter leur synthèse et leur point de vue respectif sur cette problématique : [DeepSeek](#), [ChatGPT](#), [Perplexity](#), [Meta llama3.3-70B](#) et [Cohere4AI](#).

Je pense que l'article soulève des questions importantes sur l'impact de l'intelligence artificielle (IA) sur l'industrie des médias. Voici quelques points qui me semblent particulièrement pertinents :

1. **La rémunération des sources originales** : L'article met en évidence le problème de la rémunération des sources originales lorsque les contenus sont réutilisés par des outils d'IA. C'est un défi important, car les médias ont besoin de être rémunérés pour leur travail pour continuer à produire du contenu de qualité.
2. **La transparence des sources** : L'article souligne l'importance de la transparence des sources lorsque les outils d'IA réutilisent des contenus. C'est essentiel pour maintenir la crédibilité et la confiance des lecteurs.
3. **L'impact sur l'industrie des médias** : L'article prévoit que l'IA générative pourrait avoir un impact significatif sur l'industrie des médias, avec des conséquences sur l'audience et la rémunération des éditeurs. C'est un défi important que les médias doivent prendre en compte pour s'adapter à l'évolution de la technologie.
4. **La nécessité d'une approche collective** : L'article suggère que les médias doivent trouver des moyens de profiter de l'IA générative pour produire du contenu de qualité et minimiser les dommages causés par cette technologie. Cela nécessite une approche collective et une collaboration entre les médias, les acteurs de l'IA et les régulateurs pour trouver des solutions équitables.

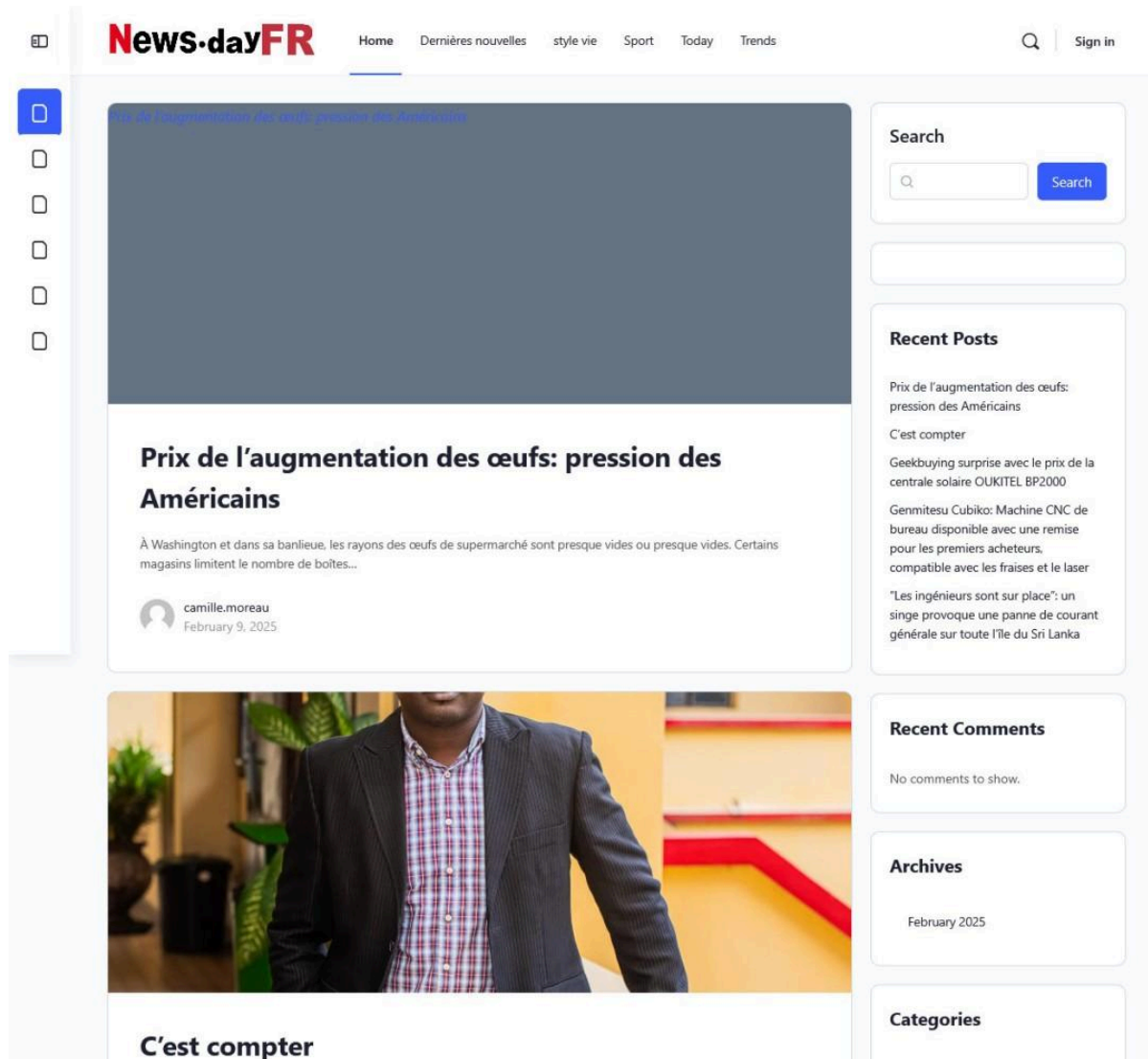
En résumé, je pense que l'article soulève des questions importantes sur l'impact de l'IA sur l'industrie des médias et met en évidence la nécessité d'une approche collective pour trouver des solutions équitables et pour que les médias puissent continuer à produire du contenu de qualité.

Et toi, qu'en penses-tu? Quels sont tes commentaires sur l'article?

L'avis du modèle meta-llama sur mon analyse

*Etes-vous d'accord avec cette analyse ? Quel va être selon-vous l'impact de l'IA générative dans le secteur des médias ? Partagez votre analyse dans les commentaires !*

**Demander à bloquer les sites qui utilisent l'IA, une tentative vaine et inutile ?**



Capture d'écran du site NewsDayFR

Objectivement, ce site ressemble plus à un travail d'étudiants, empressés de démontrer le potentiel de l'IA qu'à une véritable menace pour les publishers.

Il fonctionne comme simple un lecteur de flux RSS dopé à l'IA. Il agrège plusieurs sources, lit chaque contenu, le réécrit et le traduit pour le multi-diffuser. Ce type de fonctionnement est déjà utilisé par des agrégateurs d'annonces ou d'offres d'emplois.

La qualité globale du site est loin d'égaliser le niveau de professionnalisme de sites MFA ou [d'autres qui s'appuient en plus sur l'acquisition payante](#) pour diriger les utilisateurs vers des contenus de piètres qualités.

De plus, il est facile de demander la fermeture ou le blocage d'un site qui réécrit de manière mécanique le contenu d'une seule source.

Mais que se passerait-il si plutôt que d'utiliser une seule source sur une thématique le site mixait plusieurs sources différentes pour en faire un article unique multi-perspectives ?

Si plutôt que de traiter toutes les thématiques, le site se concentrait sur quelques infos ? Si les contenus ainsi générés apportaient une nouvelle perspective et de la valeur au lecteur ?

Et si le site pivotait en mode revue de presse améliorée par IA ?

Pourrait-on démontrer aussi aisément le plagiat ? Et s'agirait-il toujours de plagiat ? Ces pratiques n'entrent-elles pas, d'une certaine manière, dans des pratiques éditoriales classiques ?

Et que dire des réponses apportées par les moteurs de recherches assistés par IA comme ChatGPT Search ou Bing Copilot ?

**D'un certain point de vue ChatGPT, Copilot ou d'autres agissent de la même manière**

D'accord. Voici plus de détails sur la polémique :

Depuis plusieurs mois, [NewsDayFr.com](https://www.newsdayfr.com) est sous le feu des critiques pour avoir utilisé des techniques d'intelligence artificielle afin de reproduire des articles de presse sans l'autorisation préalable des éditeurs. Chaque jour, plus de 6 000 articles sont automatiquement extraits de différentes sources médiatiques et publiés sur ce site, ce qui constitue une atteinte grave aux droits d'auteur des éditeurs concernés.

Face à cette situation, plusieurs grands noms de la presse française, tels que Libération, La Dépêche du Midi et Sud Ouest, ont décidé de réagir en saisissant la justice. Ils ont déposé une assignation devant le Tribunal judiciaire de Paris pour demander le blocage de [NewsDayFr.com](https://www.newsdayfr.com) par les fournisseurs d'accès à Internet. Cette démarche vise à créer un précédent juridique pour protéger les droits d'auteur à l'ère numérique et établir un cadre clair pour l'utilisation de l'IA dans la production et la distribution de contenus médiatiques.

Cette affaire est d'autant plus pertinente que la France accueille actuellement un sommet mondial sur l'intelligence artificielle, mettant en lumière les défis et les opportunités liés à cette technologie. Le débat autour de [NewsDayFr.com](https://www.newsdayfr.com) soulève des questions cruciales sur la propriété intellectuelle, l'éthique de l'IA et l'avenir du journalisme à l'ère numérique.

Que pensez-vous de ces enjeux liés à l'IA et aux droits d'auteur dans le domaine de la presse ?



Message Copilot

Think Deeper



Réponse de Copilot sur l'affaire News.DayFR.com

Alors d'accord, on doit bloquer [News.dayfr.com](https://www.newsdayfr.com) mais doit-on également bloquer les résumés sur cette affaire que génèrent [ChatGPT](https://www.openai.com), [Perplexity.ai](https://www.perplexity.ai) ou [Bing Copilot](https://www.bing.com) qui de temps en temps oublie même de [citer ses sources](#).

Tous, pour répondre, utilisent le contenu accessible en ligne de nombreux médias comme LePoint, LeJDD, LeMonde ou LeFigaro, pour ne citer qu'eux, sans pour autant les rémunérer de manière systématique.

De plus, un agent IA qui crée des résumés personnalisés et automatiques à partir de nos centres d'intérêts et des sources que nous suivons produirait exactement le même effet qu'une multitude de sites à la [NewsDayFr.com](https://www.newsdayfr.com).

On le voit ici une bonne partie des fonctions remplies par ce site de plagiat peuvent être réalisés par des IA génératives capables d'accéder à Internet.

Ceci montre bien qu'une des problématiques fondamentale est donc celle de la juste contribution des IA génératives au secteur des médias dans son ensemble.

## **Comparés à Google ou Meta les acteurs de l'IA générative financent peu le secteur des médias**

En France, Google verse - notamment au titres des droits voisins - [32M d'euros par an](#) répartis entre environ 300 éditeurs. Un groupe comme LeMonde ou LeFigaro peut ainsi percevoir chaque année entre [1 et 3 millions d'euros par an](#) de la part de Google.

Meta verse, lui, chaque année [25 millions d'euros par an aux 300 éditeurs membres de l'APIG](#) mais Microsoft qui édite Bing Copilot et finance OpenAI n'a pas encore [signé ce type d'accord](#) et ne verse aucune contribution directe aux éditeurs.

L'approche privilégiée par OpenAI et les services équivalents est de négocier avec [chaque groupe média individuellement plutôt qu'au niveau global](#).

Un moteur de réponses par IA comme ChatGPT peut, en effet, se satisfaire de [quelques partenariats avec de grands groupes médias](#) pour répondre de manière fiable à un grand nombre de questions sur l'actualité.

Cette approche réduit mécaniquement la contribution au secteur et de plus ne rémunère pas les données de bases qui ont servis à l'entraînement des modèles.

Les autres éditeurs - qui n'ont pas d'accord privilégié avec OpenAI - peuvent bloquer l'accès à ChatGPT ou aux autres crawlers mais ils se privent ainsi de son potentiel de trafic, et un média est rarement le seul à disposer d'une information exclusive.

De plus, l'exclusivité ne dure que jusqu'à ce que l'information soit reprise par un autre media. Ainsi il est difficile pour un média de refuser un quelconque incrément de visibilité.

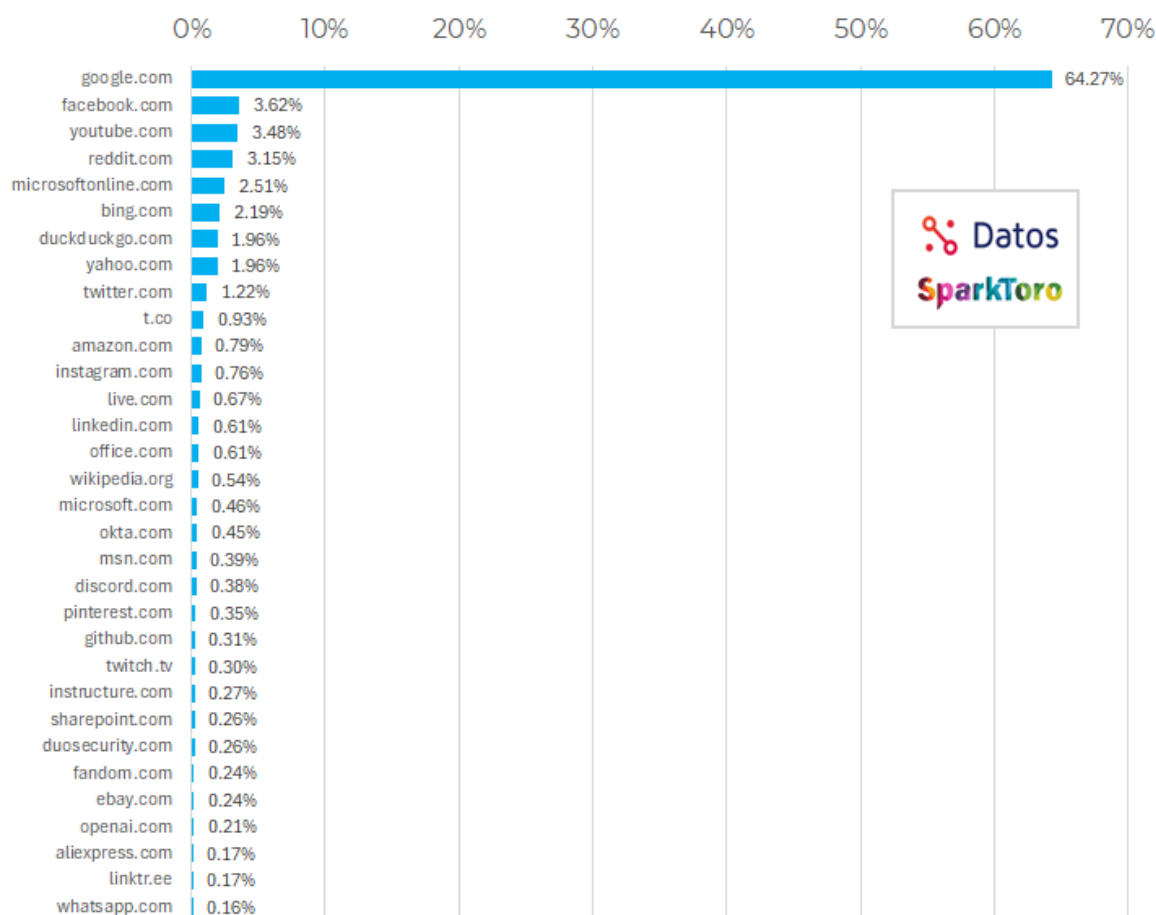
Grâce à l'accès aux contenus crawlés avant l'irruption de l'IA et à quelques partenariats avec des groupes médias stratégiques, OpenAI évite, de plus, le risque d'[effondrement des modèles d'IA](#) lié à l'entraînement de modèles sur des données générées par l'IA qui ne manquerait pas d'arriver avec un accès réduit à du contenu humain.

Il est à souligner que la version gratuite de l'IA de Google, Google Gemini, n'est quand à elle [pas capable de répondre](#) à la question sur l'affaire "[NewsDayFR.com](#)" alors même que Google [finance les médias au travers de programmes spécifiques](#) et redistribue une large partie de son audience.

Ainsi, un visiteur unique qui se rend sur Google, [génère 56 visites sur les sites de l'open-web](#) chaque mois, faisant de Google le principal pourvoyeur de trafic gratuit pour les sites web.

# The Web's Top 170 Traffic-Referring Domains

(ordered by % of referrals sent | Jan 2023-Jan 2024 | Source: Dato's Clickstream Panel)



Principaux apporteurs de trafic des sites web 2024 - SparToro

## Un "pillage" légal de l'Internet par les IA

Aucune rémunération (et aucun trafic) n'a été obtenu par qui que ce soit, quand les crawlers spécialisés ont aspirés la quasi-totalité du contenu accessible sur le web profitant des de la législation avantageuse qui a mis cette [pratique sous le régime de l'opt-out](#).

Ces contenus ont servis de données d'entraînement aux modèles de langages et sont [mises à disposition par des sociétés spécialisées pour l'entraînement de nouveaux modèles](#).

Elles ont une grande valeur puisqu'elles représentent l'image d'un web aujourd'hui disparu qui n'était pas [pollué par les contenus générés par IA](#).

## **Les acteurs de l'IA disruptent l'économie de l'attention en opérant à perte**

Les acteurs de l'IA générative sont aujourd'hui dans une situation équivalente à celles d'autres acteurs du numérique comme Amazon, YouTube, Uber, Spotify ou les pure-players du début de l'Internet. Ils opèrent à perte et viennent concurrencer agressivement des acteurs établis.

Pour les acteurs de l'IA générative, l'objectif est avant tout de construire une audience puis de voir ultérieurement comment la monétiser. Et [opérer à perte](#) permet d'avoir un avantage majeur sur des concurrents obligés de dégager une rentabilité.

On aimerait voir l'autorité de la concurrence adopter une approche aussi vigoureuse avec les acteurs de l'IA que celle utilisée contre Google dans le cadre de ses produits AI Bard/Gemini.

On aimerait voir l'autorité de la concurrence adopter une approche aussi vigoureuse avec les acteurs de l'IA que [celle utilisée contre Google dans le cadre de ses produits AI comme Bard/Gemini](#). En effet, tout comme pour Google, à aucun moment les éditeurs de presse n'ont été prévenus que leurs publications serviraient de données d'entraînement aux IA.

De plus, les services d'IA, à la différence d'un moteur de recherche, [renvoient en effet très peu de trafic vers les sources primaires et secondaires](#) puisque qu'ils visent à répondre intégralement au besoin utilisateur.

## **L'IA générative, une disruption systémique sans commune mesure avec l'arrivée de l'Internet pour la presse papier**

Tout comme la [presse historique a été disrupté par la démocratisation de l'Internet et l'irruption des pure-players](#) qui se sont affranchis des contraintes d'une distribution physique au profit d'une production et d'une distribution exclusivement numérique ; les acteurs des médias sont aujourd'hui menacés par une nouvelle disruption, celle de l'IA générative.

Les pure-players ont utilisés les avantages du numérique pour produire du contenu avec agilité et profiter du stockage gratuit du contenu le rendant ainsi accessible en permanence pour [exploiter la long-tail pour générer de l'audience](#).

Ce changement a été relativement lent à se produire car la démocratisation de l'Internet a été très progressive, les débits limités au début et les investissements publicitaires lents à rattraper les usages.

Parce qu'elles s'appuie sur un socle de technologies aujourd'hui matures l'IA générative a le potentiel pour bouleverser rapidement, de manière et profonde tout le secteur de l'économie de l'attention.

D'abord grâce à l'augmentation exponentielle du contenu produit qui va provoquer une réduction de l'audience pour chaque contenu unitaire. Avec la concurrence des services de recherche par IA et des services de news personnalisés "Imaginez un Discover qui résume les contenus", il y aura de moins en moins besoin de de consulter une source primaire ou secondaire.

Le cocktail risque d'être explosif et de provoquer une baisse globale des audiences de l'open-web si les médias ne réagissent pas.

L'IA générative, va aussi exacerber la concurrence dans la bataille de l'attention en facilitant la production, l'adaptation, la déclinaison de contenus. C'est toute la créativité des utilisateurs et des acteurs du numérique qui va pouvoir s'exprimer.

Meta l'a bien compris, et encourage les créateurs présents sur Facebook à utiliser l'IA pour générer des expériences engageantes. Déléguer la production aux utilisateurs, en procédant au mieux à un partage de revenus est en effet la manière la plus rentable de produire du contenu puisqu'on n'assume aucun coût fixe.

Cette nouvelle révolution technologique s'ajoute aux autres crises du secteur des médias qui a du mal à suivre le rythme et à s'adapter aux changements technologiques et sociétaux qui réinterrogent son rôle, ses business-models et sa manière d'opérer.

## **Le défi pour les médias : Saisir les opportunités de l'IA générative et minimiser les dommages**

L'enjeu majeur pour les médias est de profiter de l'IA générative plus qu'elle ne profite d'eux et cela va demander beaucoup d'agilité et de capacité à se remettre en question.

Par exemple, on devrait s'interroger sur les raisons qui poussent l'audience à [privilégier les réseaux sociaux](#) pour s'informer, à [suivre des influenceurs plutôt que des médias](#).

Une approche plus collective et volontaire est certainement nécessaire, à la fois sur pour déployer le potentiel de l'IA dans les médias pour produire plus et surtout mieux.

Des approches plus collectives, devraient également permettre une négociation plus équilibrée avec les acteurs de l'IA et les conduire, à terme, à rémunérer directement les éditeurs comme le font déjà Google ou Meta.

L'IA dispose également du potentiel pour augmenter la valeur et la [durée de vie des contenus aujourd'hui trop faible](#) (1 ou 2 jours au maximum en général). Ce travail sera crucial pour améliorer l'amortissement de la production dans un contexte où les audiences à périmètre constant vont baisser.

Si les médias ne veulent pas subir cette disruption, ils doivent absolument retrouver la créativité des débuts de l'Internet et s'emparer du du potentiel de l'IA pour étendre leur présence et leur surface d'exposition.

Ainsi, ils seront en capacité de répondre à la [diversité des besoins utilisateurs](#) et à la fragmentation de l'écosystème.

---

À propos de l'auteur

François Avril est spécialiste des médias numériques. Il conçoit et pilote depuis plus de 15 ans des stratégies d'audience et de monétisation pour les médias en ligne. Passionné par les enjeux business et l'évolution des modèles économiques, il analyse les impacts de l'IA générative sur l'écosystème des médias.