

Springer, Le Monde et pas les autres : les rapports compliqués de la presse avec OpenAI

Description

Le *New York Times* n'a pas trouvé de terrain d'entente avec OpenAI et a préféré porter plainte. Springer en Allemagne et *Le Monde* en France ont signé un accord qui, moyennant rémunération, autorise l'accès à leurs données pour l'entraînement de ChatGPT et valorise les articles de leurs journaux dans les réponses de l'IA. Au risque d'une information à deux vitesses, entre grands et petits éditeurs.

Si l'intelligence artificielle générative a reçu un accueil favorable de la part du public et des investisseurs depuis la mise à leur disposition de ChatGPT le 30 novembre 2022, elle suscite en revanche de nombreuses tensions chez les ayants droit. Après des premières plaintes pour non-respect du droit d'auteur émanant des mondes de l'image et du livre, les éditeurs de presse se mobilisent contre l'exploitation de leurs articles pour entraîner les intelligences artificielles qui, ensuite, proposeront des réponses sur mesure aux internautes, ce qui les dispensera potentiellement d'aller s'informer sur les sites et applications des journaux.

À cet égard, la plainte du *New York Times* déposée contre OpenAI le 27 décembre 2023 est emblématique. Le quotidien new-yorkais de référence, connu pour avoir fortement innové dans le numérique en faisant la preuve de la pertinence des stratégies de *paywall* pour la presse quotidienne en ligne, s'en prend cette fois-ci à une autre entreprise innovante, venue du monde de la recherche en informatique et étrangère aux problématiques qui sont celles des producteurs d'information. Il reproche à OpenAI d'avoir exploité ses articles sans autorisation et sans contrepartie financière pour entraîner son intelligence artificielle. Il conteste que cette exploitation puisse relever du *fair use* en vigueur aux États-Unis. En effet, les réponses apportées par ChatGPT sont en compétition avec les articles du *New York Times* comme source d'information possible pour les internautes.

En outre, elles imitent les articles qui ont servi à l'entraînement, le *New York Times* ayant identifié l'équivalent de « copier-coller ». Le 8 janvier 2024, OpenAI a précisé dans un post de blog qu'il s'agit d'un raté assez rare des IA génératives, qui sera corrigé, sous-entendant que le *fair use* reste, en l'espèce, valable. À l'évidence, l'absence de liens hypertextes vers les articles du *New York Times*, qui a échoué à s'entendre avec OpenAI sur un accord de rémunération, remet en question le modèle d'affaires du quotidien et soulève la question des asymétries, dans la négociation, entre les acteurs de l'IA générative et les éditeurs de presse, surtout quand ceux-ci organisent la concurrence sur l'accès à la rémunération par les

IA. En effet, OpenAI a reconnu qu'il peut être légitime de payer l'accès à des données qualifiées pour entraîner son intelligence artificielle, un accord en ce sens ayant été signé avec Associated Press (AP) en juillet 2023.

Il aurait donc pu trouver un accord avec le *New York Times*, mais il disposait de solutions alternatives proposées par d'autres éditeurs. Pour preuve, le 13 décembre 2023, quinze jours avant que le *New York Times* ne porte plainte, le groupe Springer rendait public un accord avec OpenAI qui le rémunère pour entraîner son intelligence artificielle et, surtout, qui s'engage à mettre à disposition des liens hypertextes vers les articles de presse du groupe quand les réponses apportées sont laborieuses avec les textes de ses journaux, des médias américains tant concernés comme Business Insider et Politico. C'est donc moins le principe d'une rémunération que les conditions souhaitées par le *New York Times* qui ont conduit à l'absence d'accord et au dépôt de plainte.

Depuis, OpenAI a reproduit en France le type d'accord signé avec Springer pour ses médias américains et ses médias allemands (*Bild*, *Die Welt*). Évidemment, OpenAI a ciblé un média de référence capable de fournir un nombre conséquent d'articles pour entraîner son IA et capable aussi de sécuriser ses réponses : il s'agit du *Monde*. En effet, l'un des principaux défauts des IA génératives est de proposer parfois des réponses biaisées ou inventées de toutes pièces. Le recours aux articles d'une marque de presse reconnue permet à ChatGPT de développer la confiance dans ses réponses et de corriger, à partir d'un travail humain réalisé par des journalistes, le texte généré automatiquement. Or, ce texte est toujours une synthèse statistique dérivée des données sur lesquelles ChatGPT est entraîné, dont certaines comportent un risque important de carences informationnelles (rappelons que ChatGPT est entraîné à partir des forums de Reddit). *Le Monde* apporte donc à ChatGPT l'autorité qui lui manquait (voir *La rem* n°65-66, p.107), en même temps qu'il s'impose comme le média qui servira de référence en France pour l'une des IA génératives les plus utilisées.

annoncé le 13 mars 2024, cet accord a fait l'objet d'une présentation sur le site du *Monde* par le président de son directoire, Louis Dreyfus, et par son directeur, Jérôme Fenoglio, signe évident de son importance et de la nécessité de l'inscrire dans un contexte qui soit favorable pour *Le Monde*, car il soulève de vraies questions. Présenté comme le premier accord conclu par OpenAI avec un média français, il se veut avantageux pour le titre de presse puisque les réponses fournies par ChatGPT mettent en exergue le logo du *Monde* avec des liens hypertextes renvoyant aux articles du journal, à l'exception de ceux qui reprennent le contenu des dépêches des agences. Le travail des journalistes du *Monde* sera donc bien mis en avant par ChatGPT, que les deux signataires qualifient de « moteur de réponse ».

Le choix du terme n'est pas innocent : qualifier ChatGPT de « moteur de réponse » conduit à faire le rapprochement avec les « moteurs de recherche », que les IA génératives devraient terme remplacer. *Le Monde* aurait ainsi obtenu de sauver sur les « moteurs de réponse » ce qui a été

si utile au développement de la presse en ligne, le référencement des articles sous forme de liens hypertextes dans les pages de résultats de Google Search. Mais, notons-le d'emblée, pour une requête donnée, les moteurs de recherche proposent la liste de toutes les pages du web qui contiennent les mots-clés, donnent donc accès à une diversité très forte de sources d'information. Avec ChatGPT, seuls seront visibles les liens hypertextes qui renvoient au site du *Monde*. Cet écrasement du pluralisme au profit d'un journal présente-t-il «*comme une des références majeures*» pose l'évidence problématique. Mais il ne s'agit que d'un premier accord, soulignent Louis Dreyfus et Jérôme Fenoglio. D'autres journaux finiront en effet par signer eux aussi avec OpenAI, ce qui diversifiera un peu la liste des liens hypertextes sous les réponses de ChatGPT. Ces accords à venir sont en outre rendus possibles par *Le Monde* qui a ainsi réalisé un travail d'approche «*à bâtonnée pour l'ensemble de la profession*».

En signant avec OpenAI, *Le Monde* indique aux éditeurs comme aux autres services d'IA générative que l'exploitation des articles de presse nécessite rémunération. Le lien est donc fait avec les moteurs de recherche qui, en référençant les articles de presse, ont été obligés de financer la production de l'information au titre du droit voisin. D'ailleurs, Louis Dreyfus et Jérôme Fenoglio parlent d'«*une source significative de revenus supplémentaires, pluriannuelle, qui intégre une quote-part au titre des droits voisins*». Cette quote-part n'est pas surprenante dès lors que ChatGPT va proposer également des liens hypertextes renvoyant au site du *Monde*.

Il n'est pas sûr qu'elle soit liée directement, dans le contrat entre *Le Monde* et OpenAI, à la reconnaissance du droit voisin des éditeurs de presse pour l'entraînement des IA, sans reprise des liens hypertextes dans les réponses des IA. Si tel est le cas, les IA génératives devraient payer pour toutes les données utilisées pour les entraîner, ce qui remettrait en question la «*fouille*» commerciale, l'une des exceptions au droit d'auteur reconnues par l'Union européenne dans la directive du 17 avril 2019 relative au droit d'auteur et aux droits voisins dans le marché unique numérique. ChatGPT peut donc continuer à s'entraîner sur les articles des autres sans les rémunérer, aussi longtemps qu'ils ne lui signifient pas qu'ils souhaitent ne pas être utilisés ou être utilisés sous condition de rémunération. Il faudra ensuite que ChatGPT trouve une raison de s'entendre avec eux pour reprendre leurs articles et les rémunérer en conséquence.

À l'évidence, l'accord entre *Le Monde* et OpenAI interdit de voir émerger une revendication commune des éditeurs de presse vis-à-vis des services d'IA générative. L'intervention de Louis Dreyfus et de Jérôme Fenoglio se comprend mieux, car il est peu probable que les autres titres aient apprécié la manœuvre solitaire du quotidien du soir. Il fallait arrondir les angles d'autant plus que, en septembre 2023, l'Alliance de la presse d'information générale (APIG) avait recommandé à ses membres d'interdire aux IA génératives l'accès à leurs articles pour s'entraîner, afin de se mettre en situation de négocier collectivement des conditions favorables pour les éditeurs de presse. L'accord *Le Monde* «*OpenAI exclut cette solution et rappelle que les IA génératives n'auront pas besoin*

de toute la presse pour produire des réponses pertinentes, mais ces réponses automatisées avec l'aide de certains titres viendront, en revanche, concurrencer le travail de tous les journalistes.

En effet, les IA génératives ne misent pas sur l'exhaustivité, comme les moteurs de recherche. Elles n'ont besoin que de quelques articles, précis et fiables, pour corriger leurs réponses en lien avec l'actualité. *Le Monde* seul ne suffira pas, mais un panel limité de médias de référence devrait faire l'affaire. Ainsi se profile, comme lors des premiers accords sur la réglementation au titre du droit voisin, une logique de « coopération » entre presse de référence et plateformes, associées pour l'occasion, et d'exclusion des autres médias, moins connus, moins puissants pour s'imposer dans la liste de ceux qui méritent d'être référencés. Facebook (News) et Google (Showcase) avaient signé avec les « grands » éditeurs pour proposer des services d'information adaptés en échange d'une réglementation au titre du droit voisin. Il a fallu ensuite l'intervention de l'Autorité de la concurrence pour que Google mette fin à cette pratique et négocie avec l'ensemble des éditeurs et de leurs représentants une réglementation pour le seul droit voisin (voir *La rem* n°60, p.5). Une intervention du régulateur sera peut-être encore nécessaire pour mieux organiser les relations entre éditeurs de presse et IA génératives.

Sources :

- Alcaraz Marina, « Springer signe avec OpenAI un accord inédit », *Les Échos*, 14 décembre 2023.
- Alcaraz Marina, « Le *New York Times* poursuit ChatGPT et Microsoft », *Les Échos*, 28 décembre 2023.
- Loignon Stéphane, « OpenAI accuse le *New York Times* d'avoir « manipulé » ChatGPT », *Les Échos*, 10 janvier 2024.
- Cohen Claudia, « Face à OpenAI, la presse française s'organise », *Le Figaro*, 10 janvier 2024.
- Alcaraz Marina, Freyssenet Elsa, « La presse cherche une alternative au pillage par l'IA », *Les Échos*, 11 janvier 2024.
- Dreyfus Louis, Fenoglio Jérôme, « Intelligence artificielle : un accord de partenariat entre *Le Monde* et OpenAI », *lemonde.fr*, 13 mars 2024.
- Alcaraz Marina, « Ce qui implique l'accord entre OpenAI et *Le Monde* », *lechos.fr*, 13 mars 2024.

Categorie

1. Economie

date création

23 juillet 2024

Auteur

alexandrejoux