

# L'IA va-t-elle signer la fin des journaux et l'avenir de l'info ?

Pas de droits d'auteur, pas d'IA. Les groupes de presse cherchent à assurer leur part de gâteau de l'essor d'intelligences artificielles qui se nourrissent de leurs contenus.

## DÉCODAGE

PHILIPPE LALOUX

**L**a fin des journaux et l'avenir de l'information. Le titre de l'ouvrage, signé en 2009 par Bernard Poulet, ancien rédacteur en chef de *L'Expansion*, avait solidement secoué le secteur de la presse. Montée en puissance des réseaux sociaux dans les pratiques informationnelles, migration des budgets publicitaires vers les Gafam, désaffection des jeunes pour l'écrit, culture du tout-gratuit... L'exercice de prospective n'était pas sot. Mais pas totalement prémonitoire.

Si, depuis, des titres ont certes disparu (dont *L'Expansion*), de nombreux acteurs ont enclenché des processus de transformation numérique dont, souvent, le papier assure encore le financement. Quinze ans plus tard, voilà que l'intelligence artificielle (IA) rebat à nouveau totalement les cartes. Jusqu'à accélérer la fin des sites de presse, cette fois ? Ceux qui n'auront pas intégré les robots dans leur stratégie, peut-être... Avec, en guise d'arbitre, les droits d'auteur pour l'usage de contenus « labellisés ».

Petit exercice de prospective au moment où se nouent les premiers accords entre une poignée d'éditeurs et OpenAI (maison mère de ChatGPT). Ou que d'autres, comme le *New York Times*, déterrent la hache de guerre.

## 1

### L'assistant virtuel, futur de l'information ?

En débarquant sans coup férir le 30 novembre 2022 dans nos vies et tous les pans de la société, ChatGPT a non seulement démocratisé l'IA générative en un temps record, mais aussi donné un sacré coup de vieux à Google. L'issue ne fait quasiment aucun doute : les agents conversationnels vont devenir nos futurs moteurs de recherche. A l'inverse, plutôt que de proposer une « bête » liste de liens, les moteurs de recherche vont tenter de se doper à l'IA. C'est déjà le cas pour Perplexity. Mais aussi pour Bing (Microsoft) ou le Search Generative Experience de Google.

Pour être très concret, il suffira (un jour, donc) de saisir une question (par texte ou commande vocale) pour qu'il contextualise une réponse avec le dernier carat de l'actu. Mieux, l'assistant virtuel pourrait aisément détecter et anticiper nos besoins. Pour générer, par exemple, une newsletter personnalisée, un podcast d'actualité, voire une vidéo usinée à l'IA. Vous êtes fan de foot, passionné de politique, au taquet sur la guerre en Ukraine, féru de questions environnementales ? Votre assistant, qui s'améliorera au fur et à mesure que vous l'utiliserez (comme Spotify ou Netflix connaît nos goûts parfois mieux que nous), vous le servira sur un plateau. Sans passer par la case « libraire » ou le site du média... Prémonitoire ou pas, ce scénario d'assistant virtuel « Monsieur Je-sais-tout » figure clairement dans les plans des plateformes.

## 2

### Les droits d'auteur, le caillou dans la chaussure de l'IA

Pour améliorer la pertinence de leurs

réponses, les agents conversationnels ont deux soucis : ils ne disposent ni de données (les archives des journaux, par exemple) ni de journalistes. Alors, ils « moissonnent ». Des quantités astronomiques de contenus siphonnés sur internet, dont une bonne partie d'œuvres protégées par le droit d'auteur. Sans contrepartie, du moins jusqu'à la signature d'accords de licence avec quelques éditeurs de presse.

Ce *crawling*, comme on dit, profite en réalité d'une faille dans la législation existante, laquelle (là est tout le débat) est antérieure à l'essor de l'IA générative. L'exception prévue dans ce texte s'applique-t-elle encore ? En attendant la réponse, la plupart des sites de presse ont fermé l'accès aux « moissonneuses ». Dans ce cas, le Copyright Act européen impose alors que les outils d'IA demandent une autorisation préalable, éclaircissent les conditions d'utilisation des données. Et garantissent une rémunération aux auteurs. Ce que n'a pas fait OpenAI, dès le départ.

Paniqués à l'idée de recréer un rapport de forces asymétrique avec ces plateformes, comme ce fut le cas avec les Gafam dès les années 2000, cer-

tains éditeurs ont dégainé l'artillerie. Comme le *New York Times*, en procès contre OpenAI. Le législateur, lui aussi, est au taquet. L'Italie, par exemple, avait interdit ChatGPT sur son territoire. Tandis que l'autorité française de la concurrence a sanctionné Google, le 20 mars dernier, pour le non-respect d'engagements pris en 2022 sur les droits voisins. Mais aussi pour avoir entraîné son IA, Bard (rebaptisé Gemini), avec les données de médias.

## 3

### Pourquoi les robots ont besoin des médias

Pour Yann LeCun, pionnier de l'IA et directeur de recherche chez Meta, les questions de droits d'auteur sont devenues existentielles. « Si l'on déclare que le *crawling* est une violation de copyright, je vous le dis tout de suite, l'industrie de l'IA s'arrête. Ça ne peut pas marcher sans », déclarait-il, en novembre dernier, lors d'un colloque à l'Arcom, le régulateur média français. « Il va falloir trouver un moyen de ménager la chèvre et le chou. »

Mais pour OpenAI, le temps presse. Pas question d'attendre une nouvelle

*Si l'on déclare que [le « crawling »] est une violation de copyright, je vous le dis tout de suite, l'industrie de l'IA s'arrête. Ça ne peut pas marcher sans*

Yann LeCun  
Directeur de recherche chez Meta

”

Très clairement, chez les éditeurs, la réflexion est en cours : comment assurer leur part du gâteau dans la réussite d'IA entraînées grâce à leurs données ? © AFP

législation sur le droit d'auteur. La loi d'internet est sans appel : le premier arrivé rafle la mise. En témoigne le moteur de recherche de Google devenu une montagne infranchissable pour avoir été plus rapide sur la balle et imposer son interface. L'urgence est donc aux accords de partage de revenus, avec des droits et des devoirs de part et d'autre. Mais, pour l'heure, pas avec tout le monde. La stratégie d'OpenAI semble suivre la logique des langues, et non de pays : l'anglais avec l'agence Associated Press, l'allemand avec Axel Springer (*Die Welt...*), l'espagnol avec Prisa (*El País...*). Et le français avec *Le Monde*, qui a donc décidé de jouer cavalier seul. En Belgique, selon nos informations, aucun contact n'a donc été noué. Ces sources de revenus sont inédites. Selon le *Financial Times*, elles pourraient atteindre plusieurs dizaines de millions d'euros dans le cas d'Axel Springer, preuve de l'enjeu stratégique de ces contenus pour l'agent conversationnel.

Concrètement, grâce à ces accords, à partir des requêtes des utilisateurs, ChatGPT pourra donner les réponses en piochant dans des contenus fournis par ces journaux, présentés comme des « labels d'info » fiables. Les références aux articles seront mises en exergue, accompagnées d'un lien hypertexte vers le ou les articles ayant nourri le robot (y compris ceux réservés aux abonnés).

## 4

### Pourquoi les éditeurs ont besoin de robots

Très clairement, chez les éditeurs, la réflexion est en cours : comment assurer leur part du gâteau dans la réussite d'IA entraînées grâce à leurs données ? Les premiers accords conclus plic ploc entre éditeurs et chatbots amorcent déjà un virage dans leur statut : celui de fournisseur officiel de contenus « labellisés » pour de grandes plateformes. Un abonné au *Monde*, pourrait, par exemple, choisir de consulter ses contenus non plus sur le site du *Monde*, mais directement via ChatGPT (qui les lui lirait via son autoradio). « A ce jeu, Netflix pourrait même un jour rafler la mise », nous glisse un observateur averti du secteur des médias. « Qui maîtrise mieux les goûts des gens dans le monde, a accès à autant de foyers, a les moyens de conclure des accords avec des éditeurs et de produire des vidéos et des podcasts ? »

Ce statut de « fournisseur de contenu », même s'il peut paraître juteux pour certains acteurs au vu des montants en jeu, aurait un revers que certains groupes médias auront du mal à avaler : la perte de maîtrise de la diffusion sur leur marché local (bref, le lien direct et historique avec leurs lecteurs). Et son corollaire, un étiolement de leadership sur leur propre paysage informationnel. En ce compris celui de la publicité numérique, un marché dont les Gafam tirent déjà 80 % de la croissance. « Quand la pub arrivera sur ChatGPT, de manière ultrapersonnalisée, on va rire jaune », nous glisse un autre observateur.

La porte de sortie ? Développer leurs propres agents conversationnels en interne, entraînés, en circuit fermé, sur les propres bases de données des éditeurs (ou d'alliances d'éditeurs). Garder le lecteur chez soi, mais lui parler autrement, en somme. « Il y aura sans doute un mélange de plusieurs modèles », renchérit notre interlocuteur : « Des accords de licence chers et vilains avec des plateformes, un service d'information à haute valeur ajoutée sur abonnement en direct et, toujours, un service minimum d'info gratuite, pour l'actu chaude. »

Au passage, le pilier publicitaire sera encore plus effrité. Comme le papier ? « Oui, sans doute », ponctue l'expert. « Mais au moins, le *business model* des médias aura été simplifié, ce qu'il cherche à faire depuis le début du web. »

