

**DIRECTION DES AFFAIRES FINANCIÈRES ET DES ENTREPRISES
COMITÉ DE LA CONCURRENCE**

Concurrence algorithmique – Contribution de la France

14 juin 2023

Ce document est une contribution écrite soumise par la France au titre de la session 5 de la 140ème réunion du Comité de Concurrence 14-16 juin 2023.

D'autres documents relatifs à cette discussion sont disponibles sur :
<https://www.oecd.org/competition/algorithmic-competition.htm>

Antonio CAPOBIANCO
Antonio.Capobianco@oecd.org, +(33-1) 45 24 98 08

JT03520611

France

Introduction

1. Les algorithmes comptent parmi les leviers technologiques les plus importants dans le processus de numérisation qui a lieu actuellement, en permettant aux entreprises d'être plus innovantes et plus efficaces.
2. Depuis quelques années, un débat a vu le jour sur la question de savoir si et dans quelle mesure les algorithmes pourraient également avoir des effets néfastes sur le fonctionnement concurrentiel des marchés.
3. Dans ce cadre, l'Autorité de la concurrence (ci-après l'« Autorité ») reviendra à titre liminaire sur l'étude conjointe réalisée en 2019 avec le Bundeskartellamt sur les algorithmes et leurs enjeux pour l'application du droit de la concurrence (1). Elle s'intéressera ensuite aux nouveaux modèles de langue conversationnels et aux risques concurrentiels qu'ils sont susceptibles de soulever (2). Elle s'interrogera enfin sur les possibles apports des algorithmes pour les autorités de concurrence (3).

1. L'étude conjointe réalisée par l'Autorité de la concurrence et le Bundeskartellamt en 2019

4. Pour rappel, en 2019, l'Autorité de la concurrence et le Bundeskartellamt ont étudié les potentiels risques concurrentiels pouvant être associés aux algorithmes. Ils ont étudié de manière approfondie la notion d'algorithme ainsi que les différents types et domaines d'application.
5. Dans leur étude commune¹, les deux autorités se sont concentrées sur les algorithmes de prix et les risques de collusion, mais elles ont également envisagé les interdépendances qui peuvent exister entre les algorithmes et le pouvoir de marché des entreprises qui les utilisent, ainsi que les difficultés pratiques rencontrées lors des enquêtes sur les algorithmes.
6. L'étude a notamment analysé trois scénarios différents suivant que les algorithmes sont utilisés comme moyen de soutenir ou de faciliter des pratiques anticoncurrentielles « classiques », que la collusion est basée sur un algorithme entre concurrents impliquant un tiers ou bien que la collusion est induite par l'utilisation parallèle d'algorithmes individuels.
7. Si les ententes algorithmiques figurent parmi les premiers risques identifiés, les algorithmes peuvent également être le support ou le vecteur de potentiels abus de position dominante et les nouvelles générations d'algorithmes en intelligence artificielle suscitent certaines interrogations à ce sujet.

¹ https://www.autoritedelaconcurrence.fr/sites/default/files/Algorithms-and-competition_FR.pdf (version française) ou https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Berichte/Algorithms_and_Competition_Working-Paper.pdf?__blob=publicationFile&v=5 (version anglaise)

2. Les modèles de langue conversationnels

2.1. Présentation

8. Depuis l'ouverture de ChatGPT au public en novembre 2022 les modèles de langue (« *Large Language Model* » ou LLM) conversationnels ont pris une grande place dans l'espace et dans le débat public. Tout internaute peut, depuis lors, avoir une conversation « réaliste » avec ces agents conversationnels, et obtenir des réponses à toutes les questions posées.

9. Outre OpenAI (créateur de ChatGPT, et de GPT 3, 3.5 et 4), de nombreux acteurs ont développés leurs propres technologies de LLM, tels que Google (LaMDA, Bard) ou encore Meta (BlenderBot3, LLaMA)

10. Face aux promesses de ces modèles, d'autres acteurs sont plus circonspects et appellent à ce que ces technologies soient régulées², tandis que d'autres prônent leur interdiction, tel que l'agence de protection des données italiennes qui a décidé de bloquer ChatGPT en Italie³.

11. Les modèles de langue tels que ChatGPT sont une combinaison de plusieurs avancées technologiques dans le domaine du traitement du langage naturel (« *Natural Language Processing* » ou NLP). Ils regroupent ainsi la puissance des réseaux de neurones (génératifs et de classification), avec celles des architectures dites « *Transformers* »⁴. Le Pôle d'Expertise de la Régulation Numérique⁵ (PEReN) a publié en avril 2023 une note décrivant précisément le fonctionnement de ces algorithmes⁶.

12. Ces modèles sont dans un premier temps entraînés sur des bases de données d'entraînement, qui leurs permettent d'optimiser leurs « hyper paramètres » (le poids donné à chaque information identifiée pendant la phase d'entraînement). Ils peuvent ensuite faire de l'inférence, à savoir produire des prédictions à partir de données nouvelles, sur lesquelles ils n'ont pas été entraînés.

13. L'entraînement de ces modèles de langue conversationnels demande une quantité élevée de données textuelles. Ces données peuvent provenir de deux types de sources : ouvertes ou privées. Les principales sources de données ouvertes utilisées par ces modèles sont le Common Crawl⁷, le corpus de pages Wikipédia, les fils de discussion Reddit, etc. Ces modèles peuvent également être entraînés sur des données propriétaires (ou une

² <https://www.lesechos.fr/tech-medias/hightech/ia-elon-musk-et-des-experts-appellent-a-une-pause-evoquant-des-risques-majeurs-pour-lhumanite-1920243>

³ https://www.liberation.fr/economie/economie-numerique/litalie-bloque-le-robot-conversationnel-chatgpt-20230331_4CERH7Q7EVC2PH5PWMAIBK3S7E/

⁴ Les réseaux de neurones Transformers sont basés sur un mécanisme d'« attention », leur permettant d'apporter une importance différente aux différentes « tokens » composant une phrase.

⁵ Service à compétence nationale, le PEReN fournit, aux services de l'État et autorités administratives intervenant dans la régulation des plateformes numériques, une expertise et une assistance technique dans les domaines du traitement des données, des data sciences et des procédés algorithmiques. Il s'investit également dans des projets de recherche en science des données à caractère exploratoire ou scientifique.

⁶ https://www.peren.gouv.fr/rapports/2023-04-06_Eclairage%20sur_CHATGPT_FR.pdf

⁷ Le Common Crawl est une initiative d'archivage du Web ouvert. Il met à disposition gratuitement un large ensemble de pages web collectées depuis 2011.

combinaison de données ouvertes et de données propriétaires), et dépendent alors de l'entreprise les développant.

14. Cette phase d'entraînement des modèles de langue conversationnels est très coûteuse, et nécessite une infrastructure informatique très développée. L'algorithme LLaMA⁸ développé par Meta a, par exemple, nécessité l'utilisation de plus de 2 000 processeurs graphiques (GPU) pendant plus de 20 jours. Cela est susceptible de rendre plus difficile l'arrivée de nouveaux entrants sur le marché.

15. Le tableau ci-dessous présente les principaux modèles présents sur le marché :

Graphique 1. Principaux modèles de langue développés

	Année	Nb. maximal de paramètres (en mds)	Architecture publique	Modèle entraîné ouvert	Données ouvertes	Conversationnel (RLHF)	Accessible aux utilisateurs (UI ou API)
BigScience BLOOM	2022	175	✓	✓	✓	X	✓
Google GLaM/PaLM	2021 / 2022	1200 / 540	✓	X	X	X	✓ (API payante)
Google LaMDA/Bard	2022	137 / ?	✓ / X	X	X	✓	✓ (UK/US)
Meta OPT	2022	175	✓	✓	✓	X	X
Meta BlenderBot3	2022	175	✓	✓	✓	✓	✓ (US)
Meta LLaMA	2023	65	✓	✓	✓	X	X
OpenAI GPT-3	2020	175	✓	X	✓ (1re version seulement)	X	✓ (option payante)
OpenAI GPT-3.5 (InstructGPT /ChatGPT)	2022	175 / ?	✓ / X	X	X	✓	✓ (option payante)
OpenAI GPT-4	2023	?	X	X	X	✓	✓ (API et UI payantes)

Source : PEReN ([lien](#))

2.2. Potentiels Risques concurrentiels

16. L'Autorité présente ici quelques pistes de réflexion préliminaires sur les potentiels risques concurrentiels que pourraient soulever ces nouveaux types d'algorithmes.

17. Premièrement, le marché recouvre à l'heure actuelle un faible nombre d'entreprises, qui profitent pour la plupart d'une position préférentielle ou dominante sur d'autres marchés numériques. Une attention particulière devra donc être portée aux pratiques pouvant résulter en un appauvrissement de la concurrence.

⁸ <https://ai.facebook.com/blog/large-language-model-llama-meta-ai/>

18. Les barrières à l'entrée pourront être un autre angle d'analyse, puisque les investissements nécessaires à l'entrée sur ce marché sont significatifs (architecture informatique de laquelle dépend notamment la puissance de calcul, accès aux données d'entraînements, etc.) et peuvent avoir tendance à augmenter dans le temps, avec l'augmentation de la taille des jeux de données d'entraînements. Ainsi, le « *cloud computing* » est généralement l'infrastructure privilégiée pour entraîner ces modèles⁹. Or, en France, quelques acteurs concentrent l'essentiel du marché en volume et en croissance¹⁰.

19. Les données sur lesquels ces modèles sont entraînés soulèvent également de nombreuses problématiques. Certaines données d'entraînement proviennent de sources ouvertes mais peuvent poser des questions de données personnelles ou de droits voisins dans le cadre des articles de presse. Cette problématique des droits voisins a d'ailleurs déjà été appréhendée par l'Autorité.¹¹

20. Ensuite, certains modèles peuvent se baser sur des données privées, collectées à un autre titre (par exemple, des données propriétaires d'opérateurs exploitant des plateformes en ligne), pouvant conférer un avantage concurrentiel à ces entreprises vis-à-vis de leurs potentiels concurrents.

21. De plus, ces modèles de langue conversationnels peuvent utiliser l'apprentissage par renforcement pour s'améliorer continuellement sur la base des retours reçus lors des échanges tenus avec les utilisateurs. Ces retours d'utilisateurs peuvent être utilisés pour améliorer les futurs algorithmes. Ainsi, une boucle d'interactions positives se crée : plus un algorithme de langage conversationnel est utilisé, plus il pourra atteindre des niveaux de performance élevés en s'entraînant sur ses propres données. Les premiers entrants sont donc susceptibles de disposer d'un avantage concurrentiel de par les données qu'ils sont en mesure de collecter.

22. Ces algorithmes peuvent également poser des questions quant aux problématiques de *self-preferencing* et de classement (« *ranking* »). En effet, on pourrait imaginer qu'une entreprise développant un modèle de langue conversationnel puisse décider, de manière déterministe, d'utiliser son IA conversationnelle pour mettre en avant ses propres produits au détriment d'autres produits provenant de ses concurrents. Par ailleurs, le fait que les agents conversationnels ne proposent généralement qu'une seule réponse rend cette problématique encore plus prégnante, les places au-delà de la première étant dépourvues d'intérêt, ce qui pourrait accentuer les effets de telles pratiques.

23. De manière générale, la possible intégration des modèles de langue conversationnels dans les moteurs de recherche pourrait avoir un impact sur leur fonctionnement et leur utilisation. Il est important que les autorités de concurrence anticipent et comprennent ces évolutions pour en mesurer les conséquences pour les activités des plateformes en position dominante (notamment dans la perspective de la mise

⁹ OpenAI et ChatGPT utilisent par exemple l'infrastructure de Microsoft Azure (Source : <https://openai.com/blog/openai-and-microsoft>)

¹⁰ <https://www.autoritedelaconcurrence.fr/fr/communiqués-de-presse/lautorite-de-la-concurrence-ouvre-jusquau-19-septembre-2022-une-consultation> (version française) ou <https://www.autoritedelaconcurrence.fr/en/communiqués-de-presse/autorite-de-la-concurrence-opens-public-consultation-until-19-september-2022> (version anglaise)

¹¹ Voir notamment la Décision 22-D-13 du 21 juin 2022 relative à des pratiques mises en œuvre par Google dans le secteur de la presse ([lien](#))

en œuvre du *Digital Markets Act*¹²) et pour l'économie de la publicité en ligne liée aux recherches, qui a nourri la pratique décisionnelle récente de l'Autorité, aussi bien directement (décisions Meta¹³ et Adloox¹⁴) qu'indirectement (décision Google « droits voisins »¹⁵).

3. Apport bénéfiques des algorithmes pour les autorités de concurrence

24. Enfin, si les réflexions en cours tendent à se focaliser sur les risques concurrentiels liés aux algorithmes, ces derniers peuvent également avoir des apports bénéfiques, notamment pour les autorités de concurrence, afin de les aider dans leurs missions.

25. Pour illustrer cet apport bénéfique, on peut citer le fait que l'Autorité a mis disposition, en mode « *open source* », le premier outil¹⁶ permettant de visualiser les interactions entre ses publications.

26. Développé par le service de l'économie numérique de l'Autorité, en partenariat avec le « *Computational Antitrust* » de CodeX, cet outil, à destination des acteurs du droit de la concurrence (rapporteurs, chercheurs, avocats, etc.), prend la forme d'un graphe de réseau dans lequel les publications de l'Autorité sont représentées et reliées entre elles par les citations qu'elles contiennent. Il permet ainsi d'identifier au premier coup d'œil les interconnexions entre les différentes publications et offre à l'utilisateur une vue d'ensemble de la pratique décisionnelle de l'Autorité.

27. L'Autorité détaille dans un article¹⁷ la méthode et les algorithmes utilisés, afin que cet outil puisse être déployé par d'autres autorités de concurrence.

¹² Règlement (UE) 2022/1925 du parlement européen et du conseil du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (règlement sur les marchés numériques) ([lien](#))

¹³ Décision 22-D-12 du 16 juin 2022 relative à des pratiques mises en œuvre dans le secteur de la publicité sur internet ([lien](#))

¹⁴ Décision 23-MC-01 du 04 mai 2023 relative à une demande de mesures conservatoires de la société Adloox ([lien](#))

¹⁵ Décision 22-D-13 du 21 juin 2022 relative à des pratiques mises en œuvre par Google dans le secteur de la presse ([lien](#))

¹⁶ <https://www.autoritedelaconcurrence.fr/fr/article/lautorite-de-la-concurrence-met-disposition-le-premier-outil-de-visualisation-de-ses> ou <https://www.autoritedelaconcurrence.fr/en/article/autorite-de-la-concurrence-introduces-first-network-visualization-tool-explore-its>

¹⁷ <https://www.autoritedelaconcurrence.fr/sites/default/files/2023-01/Stanford-Computational-Antitrust-en.pdf> (version française) ou <https://www.autoritedelaconcurrence.fr/sites/default/files/2023-01/Stanford-Computational-Antitrust-fr.pdf> (version anglaise)

Graphique 2. Illustration du réseau créé par les publications de l'Autorité

