

# Adhoc du cep

No 5 | 2024

12 mars 2024

## Concurrence dans le domaine de l'IA générative

Étude répondant à l'appel à contribution de la Commission européenne

Anselm Küsters et Matthias Kullas



**L'intelligence artificielle générative (IA générative) va remodeler les économies et la façon dont nous percevons la réalité et interagissons avec elle. Garantir la concurrence dans la chaîne de valeur de l'IA générative n'est donc pas seulement un impératif économique, mais aussi une mesure de précaution pour protéger les valeurs démocratiques et garantir la souveraineté numérique de l'UE. Cet adhoc du cep évalue les possibles restrictions de concurrence dans les trois parties de la chaîne de valeur de l'IA générative : l'infrastructure (données à grande échelle et puissance de calcul), la formation des modèles de base, et les services et applications B2B/B2C en aval.**

- Les infrastructures d'IA générative : en ce qui concerne les données d'entraînement, nous ne voyons actuellement aucun problème de concurrence. Toutefois, les accords de licence exclusifs avec les entreprises médiatiques doivent être surveillés de près. En ce qui concerne la puissance de calcul, la combinaison unique d'une forte demande, d'un petit nombre de fournisseurs de puissance de calcul, de coûts de commutation élevés et d'une intégration verticale croissante présente des risques non négligeables pour la concurrence, pouvant par exemple prendre la forme d'autoréférencement ou de discrimination.
- La formation à l'IA générative : en raison d'importantes économies d'échelle et d'envergure, telles que des coûts fixes élevés pour la formation et des « capacités émergentes » lors de la mise à l'échelle, le marché des modèles de base est un marché où le gagnant empoche le plus, « winner-takes-most ». Il est donc très difficile pour les jeunes entreprises de concurrencer les fournisseurs de modèles de base bien établis.
- Le déploiement de l'IA générative : en ce qui concerne les applications et les services de l'IA générative en aval, les start-ups doivent rivaliser avec les grandes entreprises telles que Microsoft et Google qui ont un accès préférentiel aux modèles avancés et peuvent tirer parti de leur pouvoir de marché et de leurs clients existants. Les modèles de « magasins d'applications » pourraient constituer un obstacle à l'avenir.
- En plus de comprendre la dynamique de chaque niveau de la chaîne de valeur, les responsables de l'application du droit de la concurrence de l'UE doivent adopter une perspective globale pour comprendre le degré d'intégration verticale de certains acteurs du marché. Un exemple clé est celui de Microsoft et de ses partenariats stratégiques avec OpenAI et Mistral AI, qui peuvent être considérés comme une « entreprise » disposant d'un fort pouvoir de marché sur l'ensemble de la chaîne de valeur de l'IA générative.

## Contenu

<b>1</b>	<b>Introduction</b> .....	<b>3</b>
<b>2</b>	<b>Contexte : Droit de la concurrence, marchés numériques et IA</b> .....	<b>4</b>
<b>3</b>	<b>Problèmes de concurrence dans la chaîne de valeur de l'IA générative</b> .....	<b>7</b>
3.1	Infrastructure de l'IA générative .....	8
3.1.1	Données d'entraînement.....	8
3.1.2	Puissance de calcul (« compute ») .....	10
3.2	Formation de l'IA générative (modèles de base) .....	15
3.3	Déploiement d'IA générative : Services en aval.....	18
<b>4</b>	<b>Considérations sur la résilience</b> .....	<b>20</b>
<b>5</b>	<b>Conclusion</b> .....	<b>24</b>

## Chiffres

Fig. 1:	Chaîne de valeur de l'IA.....	3
Fig. 3:	Calcul de formation (FLOP) de certains grands modèles linguistiques et multimodaux, 2019-22	11
Fig. 2:	Utilisation des puces citées dans les articles sur l'intelligence artificielle (représentation logarithmique).....	12

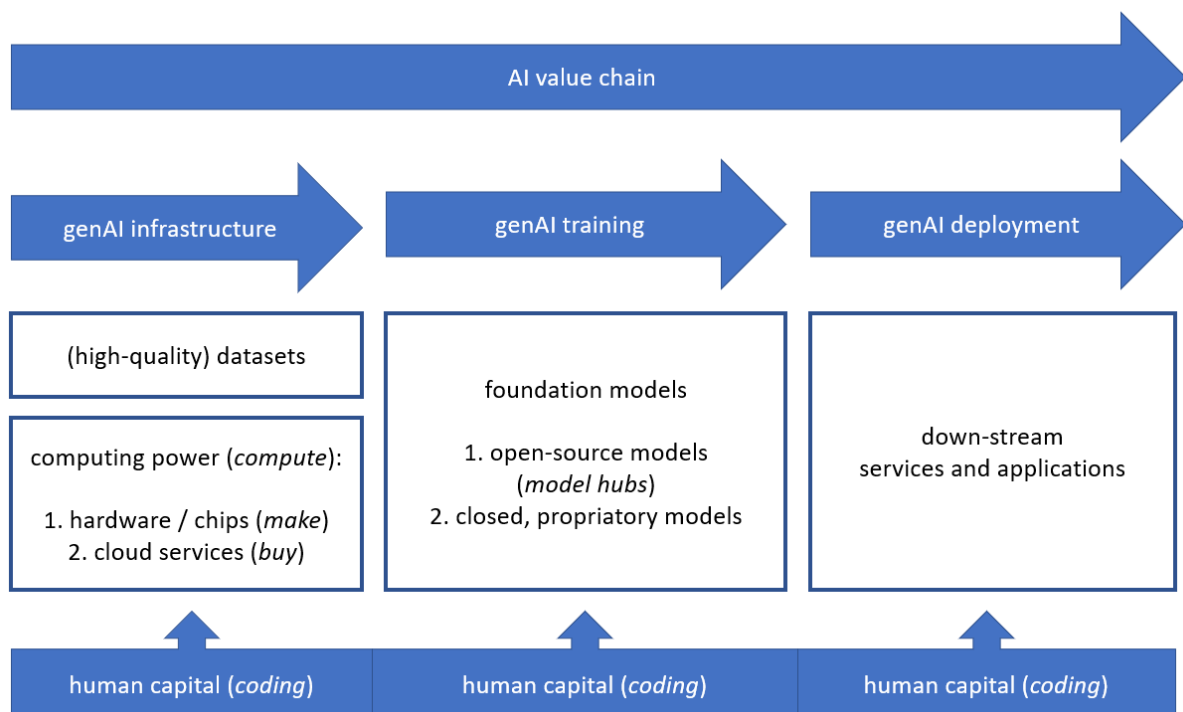
## Tableaux

Tab. 1:	Classement par la foule des modèles de fondation (10 premiers) .....	16
---------	--	----

## 1 Introduction

L'intelligence artificielle générative (IA générative) constitue une technologie transformatrice qui remodèle la concurrence et l'innovation. Dans un avenir proche, elle influencera la manière dont nous accédons à l'information<sup>1</sup> et, de plus en plus, la manière dont nous percevons la réalité et interagissons avec elle, étant donné que les agents de l'IA générative sont intégrés dans un éventail de plus en plus large de produits et de services<sup>2</sup>. Par conséquent, l'évaluation de l'état de la concurrence dans la chaîne de valeur de l'IA générative (figure 1) n'est pas seulement un impératif économique conforme aux dispositions du droit de la concurrence de l'Union européenne (UE), qui couvrent la cartellisation, les positions dominantes et les fusions, mais devrait également être considérée comme un exercice de précaution crucial pour protéger les valeurs démocratiques de l'Union et garantir sa souveraineté numérique à l'avenir.

Fig. 1: Chaîne de valeur de l'IA



Source : Illustration personnelle.

En réponse à l'appel à contributions de la Commission européenne<sup>3</sup>, cet Adhoc du cep vise à fournir une analyse complète de la dynamique concurrentielle au sein de la chaîne de valeur de l'IA générative, également connue sous le nom de « pile technologique » de l'IA. Nous divisons la chaîne de valeur de l'IA générative en trois parties principales : La première partie concerne l'infrastructure de l'IA générative, qui comprend le marché des ensembles de données à grande échelle (de haute qualité) et le marché de la puissance de calcul. Ces deux éléments sont essentiels pour la formation et le

<sup>1</sup> Pour un exemple de la manière dont l'IA générative pourrait modifier le trafic de Google vers les sites web des journaux, voir : [News Publishers See Google's AI Search Tool as a Traffic-Destroying Nightmare - WSJ \(Les éditeurs de presse voient l'outil de recherche IA de Google comme un cauchemar destructeur de trafic\)](#). Voir également l'excellente discussion à l'adresse suivante : « Generative AI's end-run around copyright » : [Generative AI's end-run around copyright won't be resolved by the courts \(aisnakeoil.com\)](#).

<sup>2</sup> C'est la vision, par exemple, de Suleyman (2023), *The Coming Wave*, Londres.

<sup>3</sup> L'appel à contributions est disponible ici : [Concurrence dans les mondes virtuels et l'IA générative \(europa.eu\)](#).

développement de modèles de base, qui constituent la deuxième partie de la chaîne de valeur de l'IA. La troisième partie est le marché (ou les marchés) des services et applications en aval B2B ou B2C qui utilisent un modèle de base comme intrant (figure 1)<sup>4</sup>. Chaque partie de cette chaîne de valeur nécessite des informaticiens pour le codage, ce qui signifie que le capital humain limité pourrait devenir un goulot d'étranglement supplémentaire et une nouvelle barrière à l'entrée.

En particulier, nous examinons toutes les parties de la chaîne de valeur de l'IA générative pour y déceler d'importantes restrictions de concurrence, en nous appuyant sur les trois questions suivantes pour déterminer si ces marchés tendent à être soumis à des problèmes de concurrence :

- 1. L'importance de la taille : Existe-t-il des économies d'échelle, telles que des coûts fixes élevés, ou des effets de réseau susceptibles d'entraîner une tendance à la monopolisation ?**
- 2. Coûts de changement de fournisseur : Existe-t-il des effets de verrouillage qui limitent la concurrence ?**
- 3. Facilités essentielles : Existe-t-il une ressource rare, telle que le capital humain ou les brevets, qui restreint la concurrence parce que seules certaines entreprises y ont accès ?**

Compte tenu de l'évolution rapide des technologies de l'IA générative et de leur impact considérable, il est essentiel pour la Commission d'anticiper et de traiter de manière proactive les problèmes de concurrence potentiels. Cet Adhoc du cep fournit un cadre structuré pour analyser ces problèmes, en proposant des recommandations pour garantir un environnement concurrentiel et innovant pour l'IA générative en Europe. Après avoir brièvement souligné la nécessité d'une application agile du droit de la concurrence dans le domaine de l'IA générative en complément de la loi européenne sur l'IA (section 2), l'analyse se poursuit par une évaluation de la dynamique concurrentielle dans chaque partie de la chaîne de valeur de l'IA générative (section 3). Enfin, nous soulignons certaines préoccupations plus générales ne relevant pas d'une définition étroite du droit de la concurrence, qui pourraient toutefois être pertinentes compte tenu de l'objectif de la Commission de renforcer l'autonomie stratégique de l'Europe<sup>5</sup> dans le domaine du numérique (section 4). Sur cette base, nous dressons notre conclusion (section 5).

## 2 Contexte : Droit de la concurrence, marchés numériques et IA

Au cours de la dernière décennie, les marchés numériques ont de plus en plus attiré l'attention des autorités chargées de l'application du droit de la concurrence et des législateurs, en particulier dans le contexte des technologies émergentes telles que les plateformes numériques bifaces. Cela s'explique en grande partie par les caractéristiques uniques des marchés numériques, qui comprennent souvent une innovation rapide, une grande évolutivité, des effets de verrouillage et d'importants effets de réseau. Si la technologie numérique contribue à débloquent l'innovation, la combinaison de ces attributs pose des problèmes pour les cadres du droit de la concurrence. À la suite de sa communication de novembre 2021 intitulée « Une politique de concurrence adaptée aux nouveaux enjeux »<sup>6</sup>, la Commission a donc adopté plusieurs initiatives législatives au cours des dernières années, notamment de

<sup>4</sup> Dans la définition de la chaîne de valeur de l'IA générique, nous suivons : [Exploring opportunities in the gen AI value chain | McKinsey](#). Aux 6 dimensions mises en évidence par McKinsey, nous ajoutons les données et le capital humain.

<sup>5</sup> Pour une définition et une discussion de ce terme, voir : Armin Steinbach (2023), [EU's Turn to 'Strategic Autonomy' : Leeway for Policy Action and Points of Conflict | European Journal of International Law | Oxford Academic \(oup.com\)](#).

<sup>6</sup> Voir : [Une politique de concurrence adaptée aux nouveaux défis - Commission européenne \(europa.eu\)](#).

nouvelles règles pour la concurrence numérique, en particulier par le biais de la loi sur les marchés numériques (DMA), de nouvelles lignes directrices pour sa panoplie traditionnelle de mesures antitrust (voir, par exemple, la communication révisée sur la définition du marché, qui tient mieux compte des marchés numériques<sup>7</sup>), et le règlement visant à promouvoir l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne<sup>8</sup>. Dans le présent contexte, il convient de souligner la communication sur la promotion d'une approche européenne de l'IA<sup>9</sup>, un examen du plan coordonné sur l'IA,<sup>10</sup> le récent accord politique sur la loi sur l'IA<sup>11</sup>, ainsi que le paquet sur l'innovation en matière d'IA visant à soutenir les start-ups et les PME dans le domaine de l'IA<sup>12</sup>. Suite à l'appel à contributions sur « La concurrence dans les mondes virtuels et l'IA générative », le Commissaire à la concurrence a souligné que puisque l'IA « touchera pratiquement tous les aspects de l'économie, nous devons examiner attentivement l'intégration verticale et les écosystèmes, tenir compte de l'impact de l'IA dans la manière dont nous évaluons les fusions et réfléchir à la manière dont l'IA pourrait conduire à de nouveaux types de collusion algorithmique »<sup>13</sup>.

Dans le contexte de cette révision législative générale et de l'adaptation à la révolution numérique, l'IA générative ne devrait pas être considérée comme un changement de paradigme dans l'application du droit de la concurrence en tant que tel, mais comme un nouvel exemple d'une nouvelle technologie qui pose des problèmes inédits, similaires aux précédents changements de technologie numérique. Nous saluons donc le fait que la Commission s'engage très tôt dans ce dernier développement technologique, afin de prévenir ex ante l'émergence de problèmes de concurrence, tels que les écosystèmes dans d'autres marchés numériques. La difficulté pour les décideurs politiques de suivre le rythme de l'IA générative est apparue clairement lors des négociations finales de la loi européenne sur l'IA, qui a nécessité l'inclusion de nouvelles règles pour les modèles de fondation. Étant donné que le développement rapide de la technologie va souvent plus vite que les réponses réglementaires, il est nécessaire d'appliquer le droit de la concurrence de manière plus agile et plus informée en complément de la formulation de nouvelles règles.

Cependant, lorsqu'on examine comment utiliser la boîte à outils existante du droit de la concurrence dans le contexte de l'IA générative, plusieurs défis apparaissent. Pour commencer, il est nécessaire d'aller au-delà des méthodes traditionnelles de définition du marché<sup>14</sup>. Pour les marchés de l'IA générative, on pourrait s'appuyer sur la définition révisée du marché mentionnée ci-dessus, selon laquelle les « dépendances liées aux données », par exemple « les coûts de la portabilité des données » et « l'accès aux données », peuvent constituer des obstacles importants à la substitution et aux coûts de changement. Ces barrières sont encore plus importantes si les nouveaux entrants et les concurrents

---

<sup>7</sup> Commission (2024), Communication de la Commission sur la définition du marché en cause aux fins du droit de la concurrence de l'Union, Bruxelles, 8.2.2024, C(2023) 6789 final.

<sup>8</sup> Pour une discussion, voir : [La souveraineté numérique de l'Europe : Comment apprivoiser le pouvoir numérique \(com-mongrounedeurope.eu\)](#).

<sup>9</sup> [Communication intitulée « Favoriser une approche européenne de l'intelligence artificielle - Façonner l'avenir numérique de l'Europe » \(europa.eu\)](#).

<sup>10</sup> [Plan coordonné sur l'intelligence artificielle - Façonner l'avenir numérique de l'Europe \(europa.eu\)](#).

<sup>11</sup> [Proposition de règlement établissant des règles harmonisées en matière d'intelligence artificielle | Façonner l'avenir numérique de l'Europe \(europa.eu\)](#).

<sup>12</sup> [Communication sur la stimulation des startups et de l'innovation dans le domaine de l'intelligence artificielle digne de confiance - Façonner l'avenir numérique de l'Europe \(europa.eu\)](#).

<sup>13</sup> Discours du 19 février 2024, Bruxelles, à : [Renouveler l'Europe au Parlement européen \(europa.eu\)](#).

<sup>14</sup> L'article de référence est le suivant : Graef, Inge, Market Definition and Market Power in Data : The Case of Online Platforms (8 septembre 2015). World Competition : Law and Economics Review, Vol. 38, No. 4 (2015), p. 473-506, disponible sur SSRN.

existants doivent réaliser « des investissements en capital spécifiques ou des investissements spécifiques dans les processus de production, l'apprentissage et le capital humain »<sup>15</sup>. Dans le contexte de l'IA générative, il est crucial de prendre en compte des facteurs tels que l'accès aux données de formation et la puissance de calcul (que nous analysons dans la section suivante)<sup>16</sup>. En se concentrant sur les obstacles qui réduisent la concurrence potentielle, au lieu d'utiliser les outils antitrust traditionnels axés sur les prix, les autorités de contrôle peuvent atténuer les problèmes liés à la définition du marché en donnant une image plus réaliste des contraintes concurrentielles sur les marchés numériques<sup>17</sup>. Par exemple, lorsque des paramètres autres que le prix sont particulièrement pertinents pour l'évaluation de la substitution (comme cela pourrait être le cas dans la chaîne de valeur de l'IA générative), la Commission peut se concentrer sur la « portabilité des données et les caractéristiques des licences »<sup>18</sup>, ce qui est pertinent compte tenu de la stratégie actuelle d'OpenAI consistant à octroyer des licences pour des données de formation à l'actualité de haute qualité auprès de grandes sociétés de médias (voir notre analyse à la section 3.1 ci-dessous).

De même, il existe un risque que les entreprises dominantes sur les marchés de l'IA générative se livrent à des pratiques anticoncurrentielles telles que l'autoréférencement, la vente liée ou la tarification prédatrice. Par exemple, « une plateforme de cloud dominante pourrait regrouper ou promouvoir son propre modèle de base au détriment des offres concurrentes, tandis qu'un fournisseur de modèle de base dominant pourrait couper ou dégrader l'accès aux interfaces de programmation d'applications (API) pour les développeurs d'applications rivaux »<sup>19</sup>. La Federal Trade Commission (FTC) a récemment examiné plusieurs exemples de ventes groupées et liées à l'intersection des marchés du cloud et de l'IA. Par exemple, de grandes entreprises technologiques comme Google et Amazon vendent apparemment l'accès à leurs outils d'IA les plus puissants exclusivement aux développeurs qui utilisent les services de cloud de l'entreprise<sup>20</sup>. Sur le marché de l'IA générative, cela pourrait se traduire par un accès restrictif à l'infrastructure du cloud ou à certains « modèles frontières » qui donnent aux utilisateurs en aval des capacités de pointe pour développer ou mettre à jour leurs produits et services basés sur l'IA. Pour éviter ce scénario, certains chercheurs ont recommandé d'encourager, voire d'imposer, des modèles de base ouverts, c'est-à-dire des modèles dont les poids sont largement disponibles, car « les développeurs de modèles fermés exercent un plus grand pouvoir pour définir et restreindre les cas d'utilisation qu'ils jugent inacceptables, alors que les consommateurs en aval des modèles de base peuvent mieux prendre ces décisions par eux-mêmes avec des modèles ouverts »<sup>21</sup>.

Enfin, il faudra peut-être repenser certains éléments du contrôle des fusions. Alors que l'acquisition de jeunes entreprises innovantes par des acteurs établis est une caractéristique commune des marchés

<sup>15</sup> Commission (2024), Communication de la Commission sur la définition du marché en cause aux fins du droit de la concurrence de l'Union, Bruxelles, 8.2.2024, C(2023) 6789 final, pp. 25f.

<sup>16</sup> Comme l'a noté l'autorité britannique de la concurrence et des marchés (CMA), plusieurs développeurs de modèles de base (FM), tels que Microsoft, Amazon et Google, « possèdent des infrastructures clés pour produire et distribuer des FM, telles que des centres de données, des serveurs, des infrastructures de réseau et des référentiels de données ». CMA (2023), AI Foundation Models : Initial Report, 18 septembre 2023, p. 16.

<sup>17</sup> Tone Knapstad (2023) Digital dominance : assessing market definition and market power for online platforms under Article 102 TFEU, European Competition Journal, DOI : 10.1080/17441056.2023.2280334.

<sup>18</sup> Commission (2024), Communication de la Commission sur la définition du marché en cause aux fins du droit de la concurrence de l'Union, Bruxelles, 8.2.2024, C(2023) 6789 final, p. 39.

<sup>19</sup> Max von Thun (2024), [L'UE ne doit pas attendre la loi sur l'IA pour agir - Euractiv](#).

<sup>20</sup> Voir la discussion sur la réunion de la FTC à l'adresse suivante : [Generative AI Concerns Fuel Federal Trade Commission Scrutiny of Cloud Services - Center for Journalism & Liberty \(journalismliberty.org\)](#).

<sup>21</sup> Rishi Bommasani, Sayash Kapoor, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Daniel Zhang, Marietje Schaake, Daniel E. Ho, Arvind Narayanan, Percy Liang (2023), Considerations for Governing Open Foundation Models, [Governing-Open-Foundation-Models.pdf \(stanford.edu\)](#), p. 5.



numériques depuis une dizaine d'années, la stratégie actuelle dans le paysage de l'IA générative semble être la formation de partenariats stratégiques entre des entreprises Big Tech établies et des jeunes entreprises innovantes. À ce jour, de tels partenariats ont été conclus entre Microsoft/OpenAI et Mistral ; Hugging Face/Amazon ; Cohere/Google et Nvidia ; Stability AI/Amazon ; et Inflection AI/Microsoft et Nvidia<sup>22</sup>. Comme l'a noté l'autorité britannique de la concurrence, « les relations verticales peuvent également se produire lorsque les entreprises ont recours à des partenariats à long terme et à des investissements stratégiques comme alternative aux acquisitions pures et simples et à l'intégration verticale »<sup>23</sup>. Étant donné que ces partenariats offrent souvent un accès prioritaire à la puissance de calcul ainsi que des tarifs plus avantageux, ils peuvent néanmoins étouffer l'innovation et réduire la concurrence, ce qui nécessite un examen minutieux. Comme l'a récemment souligné le commissaire européen à la concurrence, les fortes barrières à l'entrée et le manque de talents en matière d'IA font qu'il est peu probable que la prochaine vague de perturbation de l'IA soit « menée par une poignée d'étudiants qui ont abandonné leurs études et qui parviennent à surpasser Open AI, le partenaire de Microsoft, ou DeepMind, celui de Google »<sup>24</sup>. En ce sens, on pourrait qualifier les partenariats tels que celui entre Microsoft/OpenAI de nouveau type d'« acquisition meurtrière » au sens du droit européen de la concurrence<sup>25</sup>. Toutefois, si ces partenariats ne relèvent pas du règlement sur les concentrations de l'UE, comme l'indique l'examen préliminaire de la Commission concernant le partenariat Microsoft/OpenAI, « la Commission devrait explorer d'autres pistes, y compris les examiner en tant qu'accords anticoncurrentiels en vertu de l'article 101 du traité sur le fonctionnement de l'Union européenne (TFUE) ou simplement mettre à jour le règlement sur les concentrations pour s'assurer qu'il englobe ce type d'investissements et de partenariats »<sup>26</sup>.

Dans l'ensemble, la Commission peut s'appuyer sur les trois piliers traditionnels du droit européen de la concurrence - accords, positions dominantes et fusions - pour contrer les comportements anticoncurrentiels émergents dans le domaine de l'IA générative. En outre, il convient de souligner que la DMA pourrait également être appliquée à bon nombre des problèmes examinés ci-dessous, si la Commission décidait, par exemple, de désigner les services de cloud dominants comme des gardiens tout en incluant les modèles de fondation dans le champ d'application de la législation<sup>27</sup>.

### 3 Problèmes de concurrence dans la chaîne de valeur de l'IA générative

Dans le paysage en évolution dynamique de l'IA générative, une chaîne de valeur distincte a commencé à émerger, composée de trois parties qui contribuent à leur développement et à leur déploiement. Cette chaîne de valeur, tout en présentant des similitudes avec les structures traditionnelles de l'IA, introduit des éléments uniques, en particulier dans le domaine des modèles de fondation. Alors que la DG Concurrence de la Commission cherche à comprendre l'IA générative, nous suggérons d'analyser trois parties intégrantes différentes qui sous-tendent sa chaîne de valeur : Premièrement, l'infrastructure de l'IA générative composée d'ensembles de données (de haute qualité) et de puissance de calcul (nous subdivisons la puissance de calcul en marchés des puces et des plates-formes de cloud). Deuxièmement, le marché des modèles de base et, troisièmement, les services et applications en aval qui

<sup>22</sup> [Le pouvoir des monopoles est l'éléphant dans le débat sur l'IA | TechPolicy.Press.](#)

<sup>23</sup> CMA (2023), AI Foundation Models : Initial Report, 18 septembre 2023, p. 18, 35.

<sup>24</sup> Discours du 19 février 2024, Bruxelles, à : [Renouveler l'Europe au Parlement européen \(europa.eu\).](#)

<sup>25</sup> Radsch (2023), [The real story of the OpenAI debacle is the tyranny of big tech | Courtney Radsch | The Guardian.](#)

<sup>26</sup> Max von Thun (2024), [L'UE ne doit pas attendre la loi sur l'IA pour agir - Euractiv.](#)

<sup>27</sup> Max von Thun (2024), [L'UE ne doit pas attendre la loi sur l'IA pour agir - Euractiv.](#)

utilisent les modèles de base comme intrants. Dans ce qui suit, nous souhaitons mettre en lumière le rôle de chaque composante et la dynamique concurrentielle au sein de l'écosystème de l'IA générative.

### 3.1 Infrastructure de l'IA générative

#### 3.1.1 Données d'entraînement

En particulier, dans les premières phases de développement des modèles (phase de pré-entraînement), les entreprises d'IA générative utilisent de grands ensembles de données, souvent des centaines ou des milliers de gigaoctets, pour construire les connaissances de base de leurs modèles. Ces ensembles de données sont généralement basés sur l'exploration du web<sup>28</sup> ou créés synthétiquement. Les entreprises d'IA générative disposent de différentes possibilités pour obtenir ces ensembles de données. Elles peuvent soit acheter ces ensembles de données, ce qui n'est généralement pas un problème étant donné que de nombreuses entreprises proposent des données (de haute qualité)<sup>29</sup>, soit explorer le web elles-mêmes afin d'obtenir suffisamment de données. Certains ensembles de données peuvent même être utilisés gratuitement. Dernièrement, l'utilisation de données propriétaires exclusives est également devenue courante.

Les ensembles de données populaires souvent utilisés à ce stade de pré-entraînement comprennent l'ensemble de données C4, compilé par AllenAI à partir de la vaste collection Common Crawl, et The Pile, une collection de 22 ensembles de données de haute qualité compilés par EleutherAI à partir de diverses sources, y compris PubMed et GitHub<sup>30</sup>. En outre, le corpus du Projet Gutenberg, qui contient plus de 50 000 livres du domaine public, et les ensembles de données LAION, qui contiennent des millions de paires image-texte, sont des ressources essentielles pour l'entraînement des modèles de langage et des algorithmes de génération d'images. La disponibilité immédiate de ces sources de données montre que la disponibilité des données n'est pas actuellement une barrière suffisante à l'entrée. Bien qu'il y ait des économies d'échelle, il y a une forte concurrence entre les entreprises qui offrent des données. Comme il n'y a pas de coûts de changement pour les entreprises qui achètent ces données pour entraîner leurs modèles de base, il n'y a pas de blocage apparent.

Toutefois, des entreprises comme OpenAI se tournent de plus en plus vers des données propriétaires (provenant par exemple de grandes sociétés de médias), qui peuvent être plus difficiles (voire impossibles, dans le cas d'une licence exclusive) d'accès pour les start-ups et les rivaux émergents. Aujourd'hui déjà, des actions en justice intentées par des artistes et des journaux tels que le NYT montrent que des données propriétaires de grande qualité ont peut-être été utilisées illégalement pour entraîner les modèles, mais elles étaient souvent disponibles en ligne et pouvaient donc être récupérées par crawling, même par les nouveaux venus. Si, à l'avenir, ces données sont mieux protégées et font l'objet d'une licence exclusive accordée à un seul fournisseur de modèles linguistiques, comme OpenAI, cela pourrait limiter la capacité des nouveaux venus à former leurs propres modèles. En fait, plusieurs articles de presse circulent actuellement et suggèrent que cette stratégie est en train de devenir la stratégie dominante d'OpenAI, peut-être pour évincer les concurrents. OpenAI a conclu un accord pluriannuel avec Axel Springer qui n'engage apparemment aucune des parties à l'exclusivité, mais qui permet

<sup>28</sup> L'exploration du web implique l'utilisation de programmes automatisés pour identifier et collecter des données à partir de pages web nouvellement publiées ou révisées à des fins de formation au modèle.

<sup>29</sup> Pour cette argumentation, voir : [Rohstoff Daten : Eine falsche Analogie bremst die europäische Digitalpolitik aus - Tagespiegel Background](#).

<sup>30</sup> Pour l'argumentation de ce paragraphe, voir : CMA (2023), AI Foundation Models : Initial Report, 18 septembre 2023, p. 11.



à OpenAI d'utiliser le contenu de l'éditeur, y compris les articles de Business Insider et Politico, pour entraîner ses modèles d'IA et améliorer ChatGPT. En retour, Axel Springer recevra des paiements non divulgués de la part d'OpenAI et prévoit de soutenir ses projets axés sur l'IA en utilisant la technologie d'OpenAI<sup>31</sup>. Une licence similaire a déjà été négociée pour une partie des archives de l'Associated Press<sup>32</sup>. Des pourparlers sont également en cours avec des sociétés de médias telles que CNN, Fox Corp et Time<sup>33</sup>. Selon les médias, OpenAI offre entre 1 et 5 millions de dollars par an pour la licence de ces ensembles de données d'entraînement<sup>34</sup>. De même, Google a conclu un accord de 60 millions de dollars avec Reddit pour utiliser leurs données afin d'entraîner l'IA<sup>35</sup>. Pour OpenAI ou Google, ces coûts sont marginaux, mais ils pourraient dissuader les nouveaux entrants dans le cadre d'une stratégie visant à « augmenter les coûts pour les rivaux »<sup>36</sup> et entraîner des problèmes d'installations essentielles. Ce problème semble plus important à la lumière des prévisions selon lesquelles les développeurs d'IA seront à court de données linguistiques de haute qualité inutilisées d'ici 2026 (définies comme le stock total de données non étiquetées disponibles sur l'internet), ce qui signifie que toutes les données communément disponibles sur le web auront été utilisées pour la formation à ce moment-là et que toutes les données supplémentaires, qui pourraient être nécessaires pour l'étape suivante des modèles d'IA générative, doivent provenir de sources privées et exclusives telles que les données internes des entreprises<sup>37</sup>. Cela suggère qu'à moins que de nouvelles sources de données (synthétiques) n'apparaissent, la croissance des grands modèles d'IA générative reposant sur de vastes ensembles de données en ligne pourrait ralentir et que l'IA ouverte continue de s'appuyer de plus en plus sur des données propriétaires concédées sous licence par quelques entreprises dominantes.

Cependant, il existe des raisons théoriques, fondées sur l'état actuel de la recherche, de ne pas s'inquiéter outre mesure des problèmes d'accès aux données. En particulier, les recherches récentes indiquent clairement que l'efficacité des données et l'architecture des modèles d'IA s'amélioreront considérablement, ce qui signifie que les LLM avancés nécessiteront de moins en moins de données pour atteindre des performances supérieures<sup>38</sup>. Cela contraste avec l'idée reçue selon laquelle l'augmentation du nombre de paramètres du modèle et de la taille des données d'apprentissage accroît considérablement la qualité et la capacité des LLM. Par exemple, un ensemble de données synthétiques d'histoires courtes en langage simple peut être utilisé pour former et évaluer des modèles linguistiques qui sont beaucoup plus petits que les modèles de pointe (c'est-à-dire qu'ils ont moins de 10 millions de paramètres au total, par rapport à un total d'environ 1,76 trillion de paramètres pour le GPT-4 d'OpenAI<sup>39</sup>), ou qui ont des architectures de réseau beaucoup plus simples (avec un seul bloc de transformateur<sup>40</sup>), tout en produisant un texte long et cohérent et en démontrant des capacités de

<sup>31</sup> [OpenAI conclut un accord avec Axel Springer sur l'octroi de licences pour l'entraînement de modèles | TechCrunch.](#)

<sup>32</sup> [ChatGPT-maker OpenAI signe un accord avec AP pour la licence des articles de presse | AP News.](#)

<sup>33</sup> [OpenAI est en pourparlers avec de grands médias pour l'octroi de licences : Report \(businessinsider.com\).](#)

<sup>34</sup> [Les contrats d'OpenAI avec les éditeurs de presse s'élèveraient à 5 millions de dollars par an - The Verge.](#)

<sup>35</sup> [Exclusif : Reddit conclut un accord de licence avec Google pour le contenu de l'IA | Reuters.](#)

<sup>36</sup> Salop, S. C. et Scheffman, D. T. (1983). Raising Rivals' Costs. *The American Economic Review*, 73(2), 267-271.

<sup>37</sup> Les données linguistiques et les données d'image de faible qualité dureront beaucoup plus longtemps, entre 2030 et 2050 et 2030 et 2060, respectivement. Villalobos et al. (2022), Will we run out of data ? An analysis of the limits of scaling datasets in Machine Learning, [2211.04325.pdf \(arxiv.org\)](#).

<sup>38</sup> Voir, par exemple : Williams (2024), [100x less compute with GPT-level LLM performance](#), Tech Radar.

<sup>39</sup> [Le mélange est tout ce dont vous avez besoin : Une alternative moins chère et meilleure au LLM à un trillion de paramètres \(arxiv.org\).](#)

<sup>40</sup> Il s'agit de la quintessence de l'architecture responsable de l'engouement actuel pour l'IA et des progrès réalisés, à savoir les modèles de langage « transformateurs » pré-entraînés. L'article fondateur est le suivant : [\[1706.03762\] Attention Is All You Need \(arxiv.org\)](#).

raisonnement<sup>41</sup>. De plus, des recherches récentes suggèrent que lorsque ces petits modèles sont combinés (ce qu'on appelle le « blending »), ils peuvent même être plus performants que les grands modèles. Par exemple, les chercheurs notent que la combinaison de trois modèles de taille modérée peut égaler, voire dépasser, les performances d'un modèle beaucoup plus grand comme ChatGPT. Plus généralement, si l'on veut se faire une idée de l'orientation future du domaine, il est utile d'examiner les actes du récent « BabyLM Challenge », qui visait à déterminer comment entraîner un « BabyLM » sur seulement 100 millions de mots (inspiré par le fait que les LLM actuels sont désormais entraînés sur >1000 fois plus de données linguistiques qu'un enfant)<sup>42</sup>. Les projets gagnants de ce concours ont obtenu des performances supérieures à celles des modèles formés sur des trillions de mots<sup>43</sup>. D'autres soumissions remarquables ont obtenu des résultats significatifs soit en s'entraînant sur des séquences d'entrée plus courtes, soit en utilisant une approche d'entraînement élève-enseignant, où un modèle plus petit (élève) apprend à partir d'un modèle plus grand pré-entraîné (enseignant). Bien que tous les efforts n'aient pas été couronnés de succès, les progrès récents de la recherche indiquent clairement que les données de formation ne constitueront pas un obstacle important à l'entrée sur le marché des entreprises d'IA générative.

**Dans l'ensemble, nous ne voyons pas actuellement de problèmes de concurrence sur le marché des données de haute qualité, car les recherches récentes indiquent que les LLM avancés nécessiteront de moins en moins de données pour atteindre une performance supérieure. Néanmoins, les accords de licence exclusive entre les fournisseurs d'informations en ligne ou d'autres propriétaires d'ensembles de données de haute qualité devraient être surveillés de près, car ils pourraient rendre plus difficile l'obtention de données de haute qualité par de nouveaux concurrents.**

### 3.1.2 Puissance de calcul (« compute »)

Une fois qu'un candidat à l'entrée sur le marché a collecté ou acheté suffisamment de données d'entraînement, il doit trouver suffisamment de puissance de calcul pour traiter les données et créer un modèle de base. Les modèles d'IA modernes nécessitent d'importantes ressources de calcul pour le pré-entraînement, le réglage fin et l'inférence, ce qui requiert souvent du matériel spécialisé tel que les GPU (Graphical Processing Units) ou les TPU (Tensor Processing Units) équipés de puces accélératrices. Alors que les GPU ont été initialement conçus et utilisés pour les graphiques en 3D, ils sont aujourd'hui devenus des piliers de l'apprentissage automatique moderne, en raison de leur rapidité de calcul parallèle. En revanche, les TPU sont des circuits intégrés moins flexibles et spécifiques à une application, développés par Google, qui peuvent être plus rapides ou plus lents que les GPU, en fonction du cas d'utilisation<sup>44</sup>. Ces composants sont donc essentiels pour gérer les immenses demandes de traitement de données des modèles d'IA générative. Au fil du temps, le calcul d'entraînement pour les grands modèles linguistiques et multimodaux a eu tendance à augmenter (figure 3) - une tendance qui se maintient<sup>45</sup>.

<sup>41</sup> [\[2305.07759\] TinyStories : How Small Can Language Models Be and Still Speak Coherent English ? \(arxiv.org\)](https://arxiv.org/abs/2305.07759).

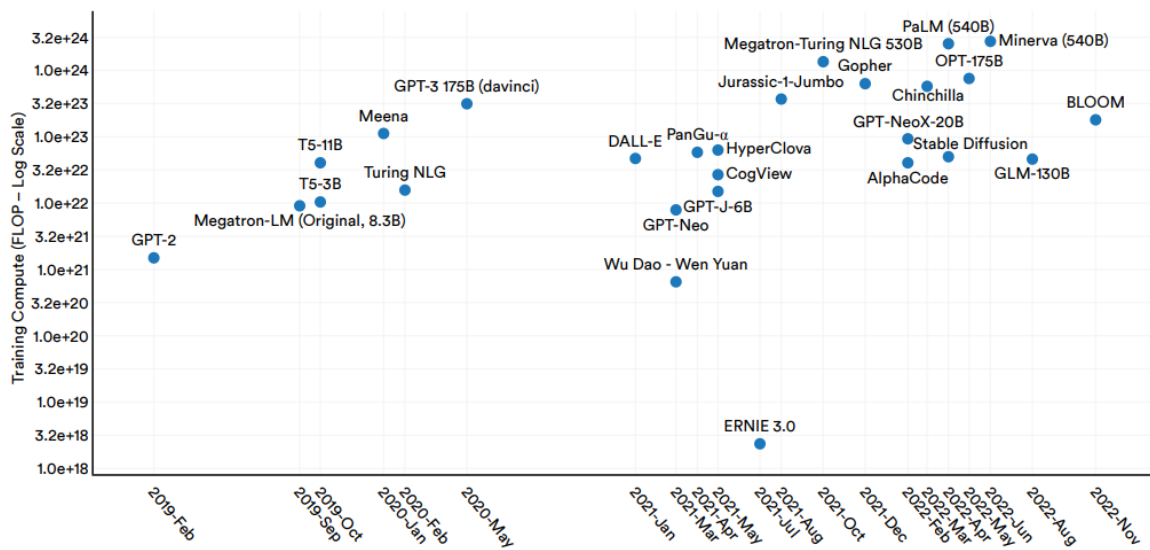
<sup>42</sup> Le compte-rendu du défi BabyLM est disponible ici : [Compte rendu du défi BabyLM à la 27e conférence sur l'apprentissage automatique du langage naturel - Anthologie ACL](https://aclanthology.org/2023.conll-babylm.1.pdf).

<sup>43</sup> Utilisation de l'architecture LTG-BERT. Voir le résumé dans : [2023.conll-babylm.1.pdf \(aclanthology.org\)](https://aclanthology.org/2023.conll-babylm.1.pdf).

<sup>44</sup> Voir : [CPU vs GPU vs TPU : Comprendre la différence entre les deux \(serverguy.com\)](https://serverguy.com).

<sup>45</sup> Voir : [HAI AI-Index-Report 2023.pdf \(stanford.edu\)](https://stanford.edu/HAI-AI-Index-Report-2023.pdf), p. 61.

Fig. 2: Calcul de formation (FLOP) de certains grands modèles linguistiques et multimodaux, 2019-



22

Source : [Epoch, 2022 | Graphique : Rapport sur l'indice IA 2023.](#)

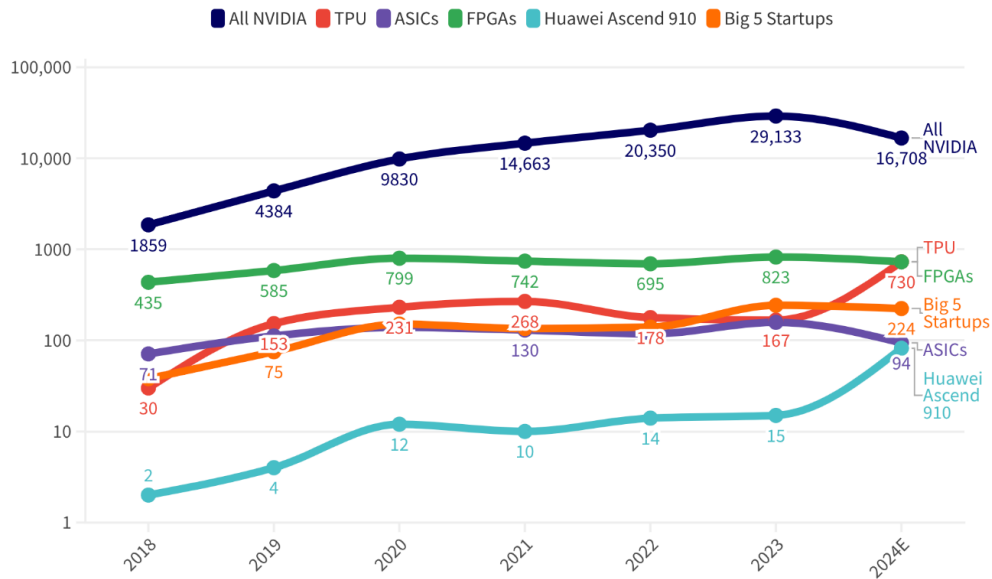
Du point de vue de la chaîne de valeur de l'IA générative, le marché des processeurs avancés adaptés est actuellement dominé par quelques acteurs clés tels que NVIDIA (en particulier le GPU H100 de NVIDIA) et Google, la fabrication étant largement concentrée entre les mains d'entités telles que Taiwan Semiconductor Manufacturing Company Limited (TSMC). Les barrières à l'entrée élevées en termes d'investissement en R&D et d'expertise technique rendent ce segment relativement impénétrable pour les nouveaux entrants. Les parties prenantes notent que la puissance de calcul requise pour l'inférence de l'IA « peut être particulièrement intensive à l'échelle »<sup>46</sup>, ce qui signifie que les nouveaux entrants pourraient avoir du mal à étendre leurs services car ils sont confrontés à des limites en termes de capacités de calcul. À l'avenir, une certaine concurrence pourrait venir des nouvelles puces informatiques récemment annoncées par Intel<sup>47</sup>, y compris Gaudi3, une puce pour logiciel d'IA générative, ainsi que des nouveaux accélérateurs et processeurs d'AMD destinés à l'exécution de grands modèles de langage<sup>48</sup>.

<sup>46</sup> CMA (2023), AI Foundation Models : Initial Report, 18 septembre 2023, p. 14.

<sup>47</sup> [Intel dévoile la puce d'IA Gaudi3 pour concurrencer Nvidia et AMD \(cnbc.com\).](#)

<sup>48</sup> [AMD lance de nouvelles puces pour accélérer l'apprentissage de l'IA - The Verge.](#)

**Fig. 3: Utilisation des puces citées dans les articles sur l'intelligence artificielle (représentation logarithmique)**



Source: [State of AI Report Compute Index and Zeta Alpha](#)

Source : [State of AI Report Compute Index et Zeta Alpha](#). Notes : Cette figure présente le nombre d'articles sur l'IA en source ouverte qui citent l'utilisation de puces d'IA spécifiques selon l'analyse de Zeta Alpha. Les chiffres pour 2024 sont extrapolés à partir de Q4/23.

À l'heure où nous écrivons ces lignes, NVIDIA - dont le cours de l'action a considérablement augmenté au cours de la ruée actuelle sur l'IA - est la société la plus dominante. Lors d'un récent sommet technologique organisé par la FTC, Corey Quinn, économiste en chef de l'informatique dématérialisée au sein du Duckbill Group, a souligné que « tous les chemins mènent à une seule entreprise, à savoir NVIDIA » et qu'il n'y a aucune transparence dans la répartition des actifs clés qu'elle détient<sup>49</sup>. Pour préserver la concurrence sur les marchés des puces informatiques utilisées dans les centres de données, la FTC envisage actuellement de bloquer les acquisitions prévues par NVIDIA<sup>50</sup>. Cette position dominante des puces NVIDIA peut être évaluée à partir du State of AI Report « Compute Index », qui suit l'utilisation de diverses puces d'IA dans les documents de recherche sur l'IA (figure 2). Cet indice repose sur l'hypothèse que l'utilisation des puces dans les articles de recherche sur l'IA (représentant les premiers utilisateurs et l'état de l'art scientifique) est un indicateur avancé de l'utilisation par l'industrie. D'après la figure 2, plus de 16 000 articles de recherche sur l'IA en 2023 utilisaient un type quelconque de matériel NVIDIA. En revanche, la TPU de Google compte 730 articles et la somme des cinq grandes start-ups de l'IA s'élève à 224 articles.

À cet égard, l'annonce récente de Mark Zuckerberg, PDG de Meta, concernant le développement du modèle de fondation Llama-3, souligne un changement important dans la dynamique des ressources de calcul au sein de l'industrie. L'acquisition prévue par Meta de 350 000 GPU H100 pour soutenir cet objectif est une décision stratégique. Étant donné qu'il ne devrait y avoir que 2 à 2,5 millions de H100 dans le monde d'ici à la fin de 2024, Meta contrôlera 14 % du marché mondial des GPU H100<sup>51</sup>. Cette concentration substantielle de ressources ne fait pas qu'amplifier les capacités de calcul de Meta, elle

<sup>49</sup> [Plus de cloud, est-ce l'avenir que nous voulons ? Une dépêche du sommet de la FTC sur la technologie de l'IA | TechPolicy.Press.](#)

<sup>50</sup> Voir : [Nvidia/Arm, dans l'affaire de la Federal Trade Commission \(ftc.gov\).](#)

<sup>51</sup> D'après les calculs de : [Engels' pause ; Silicon sensitivity ; GPUs galore ; Afrobeat ++ #457 \(exponentialview.co\).](#)

constitue également une barrière économique redoutable pour les entreprises émergentes dans le domaine de l'IA générative. Comme l'a noté l'AMC, les quelques entreprises qui dominent le matériel informatique « créent des dépendances pour les startups et les autres entreprises qui développent et déploient des outils d'IA », notamment parce que « les architectures GPU modernes rencontrent des goulets d'étranglement avec les transformateurs » et qu'« il y a une pénurie de GPU pour serveurs à destination des IA »<sup>52</sup>. Il est intéressant de noter que Sam Altman, le directeur d'OpenAI, tente actuellement de réunir 7 000 milliards de dollars pour un nouveau projet de puce d'IA<sup>53</sup>.

Récemment, la Commission a reconnu plus explicitement ce lien entre la puissance de calcul et le développement de l'IA comme une priorité stratégique. L'UE a lancé une enquête antitrust sur le marché des puces dominé par NVIDIA et prévoit des évaluations complètes des risques pour les technologies des semi-conducteurs et de l'IA dans le cadre de sa stratégie de sécurité économique<sup>54</sup>. Selon un paquet d'initiatives en matière d'IA présenté fin janvier 2024, la Commission vise à donner aux PME et aux start-ups européennes un accès privilégié au réseau national de superordinateurs<sup>55</sup>. À cette fin, la Commission modifiera le règlement européen sur le calcul à haute performance. Elle prévoit de faire des usines d'IA un nouveau pilier des activités de l'entreprise commune Supercomputer de l'UE, dont l'utilisation est gratuite. Toutefois, selon les médias, la demande est supérieure à la capacité, de sorte que les start-ups et les PME n'ont pas eu suffisamment de chances<sup>56</sup>. En outre, l'association allemande pour l'IA souligne que les projets précédents menés avec des superordinateurs européens utilisaient beaucoup moins de GPU que ce qui est nécessaire pour la dernière génération d'IA (c'est-à-dire l'IA générative), et demande donc une infrastructure de supercalculateurs plus puissante et plus flexible<sup>57</sup>. La tendance vers des modèles plus petits décrite ci-dessus, qui est déjà évidente dans la recherche mais qui n'a pas encore été suffisamment reconnue dans le discours politique et médiatique, suggère que ces mesures seront moins importantes pour un environnement de marché concurrentiel que ce qui est généralement prévu, au moins à moyen et à long terme. À cet égard, une autre ligne de recherche notable a été récemment présentée par la Fondation Linux, dont la nouvelle architecture de modèle d'IA (appelée « valeur de clé pondérée par la réception ») pourrait être en mesure de réduire jusqu'à 100 fois les besoins en GPU de calcul pour la formation des LLM<sup>58</sup>. Jusqu'à présent, cependant, cette approche se heurte à certains défis techniques qui ne sont pas encore complètement résolus.

Compte tenu des coûts élevés et de la rareté du matériel informatique de pointe, les jeunes entreprises achètent généralement de la puissance informatique à de plus grandes sociétés en utilisant des plateformes de cloud qui offrent aux entreprises un accès évolutif et rentable à la puissance informatique nécessaire. Des pays comme les États-Unis et le Japon, ainsi que des entreprises privées comme Amazon, Google et Alibaba, ont depuis longtemps pris de l'avance sur leurs concurrents grâce à des investissements considérables dans l'infrastructure de cloud nécessaire<sup>59</sup>. Le marché des services de cloud est actuellement dominé par quelques grands fournisseurs, dotés d'une vaste infrastructure et d'un accès privilégié au matériel, à savoir Amazon Web Services (AWS), Microsoft Azure et Google Cloud

<sup>52</sup> CMA (2023), AI Foundation Models : Initial Report, 18 septembre 2023, p. 33.

<sup>53</sup> [Le PDG d'OpenAI, Sam Altman, chercherait à obtenir des milliards de dollars pour son projet de puce d'IA \(cnbc.com\)](#).

<sup>54</sup> [L'UE entame une enquête préliminaire sur les abus du marché des puces d'IA que Nvidia domine - Bloomberg](#).

<sup>55</sup> Commission, COMMUNICATION sur la stimulation des startups et de l'innovation dans le domaine de l'intelligence artificielle de confiance, Bruxelles, 24.1.2024, COM(2024) 28 final.

<sup>56</sup> Selon des rapports en : Europe.Table # 617 / 25. Januar 2024.

<sup>57</sup> [LEAM-MBS KIBV webversion mitAnhang\\_V2\\_2023.pdf](#).

<sup>58</sup> Williams (2024), [100x less compute with GPT-level LLM performance](#), Tech Radar.

<sup>59</sup> Voir les preuves recueillies dans : Vili Lehdonvirta (2022), Cloud Empires : How Digital Platforms Are Overtaking the State and How We Can Regain Control.

Platform (GCP) (communément appelés « hyper-scalers »). Selon un récent rapport britannique de l'Ofcom, AWS et Microsoft détenaient, sur le marché britannique, une part de marché combinée de 70 % à 80 % en 2021, la part de Google se situant entre 5 % et 10 %<sup>60</sup>. Au début de l'année 2024, AWS conservera la plus grande part de marché au niveau mondial, avec 32 %, suivie de Microsoft Azure (23 %) et de Google Cloud (10 %), tandis qu'Alibaba Cloud et Tencent Cloud sont des acteurs importants sur le marché de l'Asie-Pacifique<sup>61</sup>. Les chiffres plus élevés pour le Royaume-Uni, comparés aux parts plus équitables au niveau mondial, suggèrent qu'en Europe, les hyper-scalers sont particulièrement dominants. Cette structure de marché concentrée pourrait entraîner des dépendances problématiques à l'avenir, étant donné que des investissements importants en temps sont nécessaires pour remanier une application ou d'autres services d'IA lorsqu'une entreprise passe d'un fournisseur de cloud à un autre. D'autres effets de verrouillage sont créés en raison des systèmes d'actualisation et des coûts de changement élevés<sup>62</sup>. Dans ce contexte, le partenariat de 13 milliards de dollars entre Microsoft et OpenAI a des implications anticoncurrentielles en raison de la dépendance d'OpenAI à l'égard de l'infrastructure informatique de Microsoft<sup>63</sup>.

Étant donné que la plupart des PME ont besoin de la technologie avancée des services de cloud computing pour développer leurs propres services d'IA générative, l'engouement actuel pour l'IA a eu pour conséquence que les grands hyperscalers ont pu récemment augmenter leurs revenus grâce à leurs plateformes cloud - une tendance qui va probablement se poursuivre. Au cours du deuxième trimestre 2023, la division Intelligent Cloud de Microsoft a généré près de 24 milliards de dollars de revenus, tandis qu'AWS a produit des revenus supérieurs à 22 milliards de dollars au cours de la même période<sup>64</sup>. Au dernier trimestre 2023, les revenus de Microsoft provenant de sa plateforme cloud Azure et des services associés ont même augmenté de 30 %<sup>65</sup>. Google Cloud arrive en troisième position, avec des revenus dépassant les 8 milliards de dollars au deuxième trimestre 2023. Bien sûr, ces récentes augmentations ne peuvent s'expliquer uniquement par le fait que les PME développent leurs services d'IA, mais il est peu probable qu'il s'agisse d'une coïncidence totale. Dans l'ensemble, le marché mondial du cloud hyperscale devrait atteindre 1 261,15 milliards de dollars d'ici 2028<sup>66</sup>.

**En résumé, nous constatons des problèmes de concurrence sur le marché de la puissance de calcul en raison d'une pénurie de puces adéquates. Cette pénurie est due au fait que le marché des puces de pointe est dominé par une entreprise (NVIDIA). En raison de contraintes technologiques, cette entreprise ne peut pas augmenter facilement la production de ces puces. La pénurie de puces adéquates et le manque d'entrants sur le marché de l'informatique du cloud<sup>67</sup> font que les grandes entreprises technologiques s'approprient la majeure partie de cette ressource rare. Bien qu'elles offrent aux jeunes entreprises l'accès à cette ressource par le biais de services d'informatique de cloud, cela ne résout pas le problème. Dans une certaine mesure, cela aggrave même le problème pour**

<sup>60</sup> Ofcom (2023), Cloud services market study : final report, [Cloud services market study final report \(ofcom.org.uk\)](https://www.ofcom.gov.uk/consult/condocs/cloud/cloud_services_market_study_final_report/), p. 3.

<sup>61</sup> Données extraites de : [2024 Cloud Market Share Analysis : Decoding Industry Leaders and Trends \(hava.io\)](https://hava.io/2024-cloud-market-share-analysis/).

<sup>62</sup> [Plus de cloud, est-ce l'avenir que nous voulons ? Une dépêche du sommet de la FTC sur la technologie de l'IA | TechPolicy.Press.](https://techpolicy.press/plus-de-cloud-est-ce-lavenir-que-nous-voulons-une-dpêche-du-sommet-de-la-ftc-sur-la-technologie-de-lia/)

<sup>63</sup> [Des groupes de la société civile demandent au Royaume-Uni d'enquêter sur le partenariat monopolistique de Microsoft avec OpenAI - Open Markets Institute.](https://openmarketsinstitute.com/des-groupes-de-la-société-civile-demandent-au-royaume-uni-d-enquêter-sur-le-partenariat-monopolistique-de-microsoft-avec-openai/)

<sup>64</sup> Données extraites de : [Cloud hyperscaler quarterly revenue by vendor 2023 | Statista.](https://www.statista.com/chart/1234567/cloud-hyperscaler-quarterly-revenue-by-vendor-2023/)

<sup>65</sup> [Le chiffre d'affaires de Microsoft Azure a augmenté de 30 %, grâce à l'IA, et le géant technologique a dépassé ses prévisions - GeekWire.](https://www.geekwire.com/2024/microsoft-azure-revenue-increased-30-percent/)

<sup>66</sup> [Marché mondial du Cloud Hyperscale : Analysis By End-User, By Region Size and Trends with Impact of COVID-19 and Forecast up to 2028 \(researchandmarkets.com\).](https://www.researchandmarkets.com/researchandmarkets.com/marché-mondial-du-cloud-hyperscale-analysis-by-end-user-by-region-size-and-trends-with-impact-of-covid-19-and-forecast-up-to-2028/)

<sup>67</sup> Il y a peu de capital-risque pour les entreprises intéressées par la mise en place d'une infrastructure de cloud coûteuse. [Plus de cloud, est-ce l'avenir que nous voulons ? A Dispatch from the FTC AI Tech Summit | TechPolicy.Press.](https://techpolicy.press/plus-de-cloud-est-ce-lavenir-que-nous-voulons-a-dispatch-from-the-ftc-ai-tech-summit/)



trois raisons : Premièrement, il est à craindre que les start-ups de premier plan telles qu'OpenAI, et les opérateurs historiques tels que les grandes entreprises technologiques, soient plus susceptibles d'obtenir un accès prioritaire à ces services de cloud pour la formation à l'IA, étant donné qu'ils concluent des accords pour détenir de plus grandes grappes de calcul<sup>68</sup>. Deuxièmement, les jeunes entreprises sont confrontées à des coûts de changement si elles souhaitent changer de fournisseur de services de cloud. Troisièmement, certains fournisseurs de services de cloud développent leurs propres puces et/ou leurs propres modèles d'IA générative, ce qui augmente le risque d'intégration verticale et l'émergence d'écosystèmes avec des effets de verrouillage. Par exemple, AWS commence maintenant à développer ses propres modèles de base internes (par exemple, Amazon Titan LLM)<sup>69</sup>, ce qui signifie qu'ils deviendront des concurrents pour les développeurs de modèles externes. Cela rappelle le comportement d'Amazon sur le marché, où il fournit l'infrastructure tout en étant en concurrence avec une partie des utilisateurs. Ce problème se pose déjà pour Google et Microsoft, en raison de leur position unique dans la chaîne de valeur de l'IA générative. Cette combinaison unique d'une forte demande, d'un petit nombre de fournisseurs de puissance de calcul, de coûts de commutation élevés et d'une intégration verticale croissante présente, selon notre évaluation, des risques non négligeables pour la concurrence, par exemple en favorisant des formes d'auto-référencement ou de discrimination.

### 3.2 Formation de l'IA générative (modèles de base)

Au cœur de l'écosystème de l'IA générative se trouvent les modèles de base. Il s'agit de modèles d'apprentissage profond étendus, pré-entraînés sur de vastes ensembles de données et capables d'être adaptés à un large éventail de tâches. Le développement de ces modèles nécessite une expertise dans de nombreux domaines, notamment la préparation des données, l'architecture du modèle et l'ajustement continu. En outre, les coûts fixes élevés de la formation des modèles de base et leurs faibles coûts marginaux de déploiement entraînent d'importantes économies d'échelle, les coûts unitaires diminuant au fur et à mesure que le déploiement augmente. Enfin, il existe sur ce marché des économies d'échelle et des avantages pour le premier arrivé, associés à des obstacles tels que la rareté du capital humain, des données spécifiques de haute qualité, de la puissance de calcul et de la propriété intellectuelle, qui augmentent encore les avantages de la concentration et pourraient même conduire à des « monopoles naturels »<sup>70</sup>.

Selon le rapport annuel de 2023 de l'AI Index, il y avait en 2022 32 modèles de fondation importants produits par l'industrie contre seulement trois produits par les universités<sup>71</sup>. Si, à première vue, 35 fournisseurs ne semblent pas constituer un problème concurrentiel important, des problèmes de concurrence pourraient se poser lorsque les différences de qualité sont si importantes que les entreprises en aval ne peuvent utiliser qu'un ou deux de ces modèles si elles ne veulent pas être désavantagées par rapport à d'autres (marché de type winner-takes-most). Selon le rapport State of AI Report 2023, le modèle de base « à la plus large capacité générale » est le GPT-4 d'OpenAI, qui bat tous les autres LLM à la fois sur les benchmarks classiques et sur les examens conçus pour évaluer les humains (bien

<sup>68</sup> CMA (2023), AI Foundation Models : Initial Report, 18 septembre 2023, p. 36.

<sup>69</sup> [AWS Bedrock se démarque de Microsoft et Google dans la course à l'IA générative | ITPro.](#)

<sup>70</sup> Vipra et Korinek (2023), [Market concentration implications of foundation models : The Invisible Hand of ChatGPT | Brookings.](#)

<sup>71</sup> Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Parli, Yoav Shoham, Russell Wald, Jack Clark et Raymond Perrault (2023), « The AI Index 2023 Annual Report », AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

que les modèles supérieurs aient récemment rattrapé leur retard, voir le tableau 1 ci-dessous)<sup>72</sup>. En outre, il convient de noter que les modèles de base améliorent considérablement leurs performances lorsqu'ils sont mis à l'échelle (c'est-à-dire en augmentant le nombre de données d'apprentissage et/ou le nombre de paramètres dans le réseau neuronal)<sup>73</sup>, récompensant ainsi les fournisseurs des plus grands modèles. En ce sens, la « taille compte » sur ce marché. Ce phénomène est souvent appelé « capacités émergentes », définies comme de nouvelles capacités qui ne peuvent être prédites en extrapolant simplement les performances de modèles plus petits<sup>74</sup>. En langage clair, cela signifie que, par exemple, un modèle plus petit aurait des difficultés en mathématiques, alors qu'un modèle plus grand pourrait soudainement apprendre cette caractéristique et ainsi connaître un « bond » dans ses capacités. En d'autres termes, cette caractéristique des modèles de base implique qu'une mise à l'échelle supplémentaire pourrait encore élargir l'éventail des capacités des modèles de langage. C'est certainement ce qu'espèrent de nombreux acteurs du marché actuel, en particulier OpenAI.

Néanmoins, pour contrecarrer cette tendance à la concentration, il existe actuellement plusieurs modèles de base qui peuvent être utilisés pour d'importantes tâches (industrielles) en aval, certains propriétaires et « boîte noire », d'autres open source et transparents<sup>75</sup>. Huggingface, par l'intermédiaire de son « Hugging Face Hub », propose plus de 60 000 modèles, tous à code source ouvert et accessibles au public, sur une plateforme en ligne<sup>76</sup>. Pour donner un exemple, considérons qu'une récente recherche de Stanford sur un « indice de transparence des modèles de fondation » a évalué OpenAI (GPT-4), Anthropic (Claude 2), Google (PaLM 2), Meta (Llama 2), Inflection (Inflection-1), Amazon (Titan Text), Cohere (Command), AI21 Labs (Jurassic-2), Hugging Face (BLOOMZ ; en tant qu'hôte de BigScience), et Stability AI (Stable Diffusion 2), sans trouver un fournisseur qui domine toutes les catégories<sup>77</sup>. Au contraire, chacun de ces modèles présente des avantages et des inconvénients. En outre, la tendance vers des méthodes de formation plus efficaces et des modèles plus petits et spécifiques à certaines tâches pourrait progressivement démocratiser ce niveau de la chaîne de valeur de l'IA générative et ainsi éroder la domination actuelle du marché par OpenAI et son modèle GPT-4<sup>78</sup>.

**Tab. 1: Classement par la foule des modèles de fondation (10 premiers)**

Rang	Modèle de fondation	Score Elo de l'arène	Organisation	Licence	Critères d'évaluation des connaissances
1	GPT-4-0125-preview	1253	OpenAI	Propriétaire	2023/4
2	GPT-4-1106-preview	1252	OpenAI	Propriétaire	2023/4
3	Bard (Gemini Pro)	1224	Google	Propriétaire	En ligne
4	GPT-4-0314	1190	OpenAI	Propriétaire	2021/9
5	GPT-4-0613	1162	OpenAI	Propriétaire	2021/9
6	Mistral Moyen	1150	Mistral	Propriétaire	Inconnu

<sup>72</sup> [Bienvenue dans le rapport sur l'état de l'IA en 2023.](#)

<sup>73</sup> Sur les caractéristiques émergentes de l'IA générative, voir : [The Unpredictable Abilities Emerging From Large AI Models | Quanta Magazine](#). Sur les risques émergeant lors d'une polycrise, voir : Küsters (2023), [cepAdhoc AI as Systemic Risk in a Polycrisis.pdf](#).

<sup>74</sup> [\[2206.07682\] Emergent Abilities of Large Language Models \(arxiv.org\)](#).

<sup>75</sup> La publication des modèles de fondation est un gradient, voir : Rishi Bommasani, Sayash Kapoor, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Daniel Zhang, Marietje Schaake, Daniel E. Ho, Arvind Narayanan, Percy Liang (2023), Considerations for Governing Open Foundation Models, [Governing-Open-Foundation-Models.pdf \(stanford.edu\)](#), p. 3.

<sup>76</sup> [Présentation du Hub privé : Une nouvelle façon de construire avec l'apprentissage automatique \(huggingface.co\)](#).

<sup>77</sup> Rishi Bommasani, Kevin Klyman, Shayne Longpre, Sayash Kapoor, Nestor Maslej, Betty Xiong, Daniel Zhang, Percy Liang (2023), The Foundation Model Transparency Index, [\[2310.12941\] The Foundation Model Transparency Index \(arxiv.org\)](#).

<sup>78</sup> Pour un bon exemple, voir : [Phi-2 : la puissance surprenante des petits modèles de langage - Microsoft Research](#).

7	Claude-1	1149	Anthropic	Propriétaire	Inconnu
8	Claude-2.0	1132	Anthropic	Propriétaire	Inconnu
9	Gemini Pro (Dev API)	1120	Google	Propriétaire	2023/4
10	Claude-2.1	1119	Anthropic	Propriétaire	Inconnu

Source : Evaluation adéquate basée sur <https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard>. Notes : Le score Elo de l'Arena est un système de notation utilisé sur la plateforme Chatbot Arena pour évaluer les niveaux de compétence de différents LLM en fonction de leurs résultats dans des affrontements anonymes et aléatoires, à l'instar du système de notation Elo aux échecs.

C'est surtout au cours du dernier semestre que la rivalité entre les modèles de fondation propriétaires s'est intensifiée, entraînant une convergence notable de leurs capacités. Sur le classement de Chatbot Arena, qui utilise un système d'évaluation ELO (un système d'évaluation largement utilisé aux échecs et dans d'autres jeux de compétition<sup>79</sup>) reflétant des centaines de milliers d'évaluations humaines de LLM, les différentes itérations de GPT-4 Turbo sont toujours en tête, mais au cours des derniers mois, d'autres modèles ont rattrapé leur retard. En particulier, Bard Pro se trouve (à l'heure où nous écrivons ces lignes) en deuxième position. Dans ce contexte, le site web « Chatbot Arena » sert de plateforme d'analyse comparative pour évaluer les performances de divers LLM dans des scénarios réels, offrant ainsi un moyen utile d'évaluer la structure actuelle du marché. En testant rigoureusement les LLM dans une gamme variée de tâches et de défis, l'Arena permet aux utilisateurs de recevoir des mesures objectives et quantifiables qui reflètent les capacités, les forces et les faiblesses de chaque modèle. Bien qu'elle ne donne qu'un aperçu dynamique de l'évolution du marché, la Commission peut exploiter les informations de cet outil d'étalonnage pour prendre des décisions éclairées et anticiper les tendances futures - par exemple en vérifiant si les modèles en place améliorent leur avantage au fil du temps ou si d'autres modèles les rattrapent - dans le secteur des modèles de fondations qui évolue rapidement. Comme le montre également le tableau, Mistral a dévoilé un modèle qui, selon les rumeurs, rivaliserait avec le GPT-4 en termes de performances. Gemini Ultra, qui vient d'être rendu public par Google, ne figure pas encore dans le tableau. Tout comme les modèles de pointe d'OpenAI, il est multimodal (c'est-à-dire qu'il peut mélanger du texte, des images et de l'audio dans la même session) et, contrairement à GPT-4 pour l'instant, il peut traiter des entrées allant jusqu'à un million de jetons<sup>80</sup>. Dans l'ensemble, la position dominante de GPT-4 semble stable pour le moment, mais elle pourrait bientôt être menacée, en particulier par Gemini Ultra et Mixtral. En d'autres termes, on peut s'attendre à ce qu'il n'y ait pas un seul modèle dominant « pour les dominer tous ».

Après avoir prétraité les données d'entraînement et trouvé une puissance de calcul suffisante, que ce soit directement par l'acquisition de matériel ou la location d'espace cloud, un futur entrant sur le marché a également besoin d'expertise humaine pour développer des modèles de base sur ces fondations. L'IA générative a considérablement influencé le marché de l'emploi en 2023, comme en témoigne l'augmentation de plus de 1000 % des offres d'emploi liées à l'IA générative sur Upwork au deuxième trimestre 2023<sup>81</sup>. Parmi les plans d'embauche actuels, un intérêt particulier est porté au traitement du langage naturel, à TensorFlow, au traitement d'images et à d'autres expertises liées à l'IA<sup>82</sup>. Cependant, le nombre de développeurs d'IA au niveau mondial est limité, les estimations de

<sup>79</sup> Le système de notation met à jour les notes en fonction des performances de chaque modèle dans ces batailles, avec des améliorations pour les résultats supérieurs aux attentes et des réductions pour les résultats inférieurs aux attentes, tout en garantissant une évaluation évolutive et progressive des modèles traitant des problèmes ouverts.

<sup>80</sup> [Présentation de Gemini 1.5, le modèle d'IA de nouvelle génération de Google \(blog.google\)](#).

<sup>81</sup> [10 compétences en IA générative les plus demandées | CIO](#).

<sup>82</sup> [10 compétences en IA générative les plus demandées | CIO](#).

2019 allant de 22 000 spécialistes de l'IA hautement qualifiés à 300 000 chercheurs et praticiens de l'IA au sein d'équipes techniques plus larges<sup>83</sup>.

Dans cette compétition pour le capital humain, les entreprises européennes sont confrontées à une bataille ascendante importante, car il est très difficile d'attirer des talents en matière d'IA et cela dépend des secteurs de recherche et des activités commerciales initiales d'une région. Les dynamiques « superstar » et « winner-takes-most » qui prévalent dans les économies numériques impliquent que seul un nombre limité de sites peut faciliter de manière prédominante la croissance à moyen terme liée à l'IA<sup>84</sup>. Une analyse historique a révélé que 56 % des nouvelles technologies ayant le plus d'impact économique provenaient de deux sites américains seulement, la Silicon Valley et le corridor nord-est, d'autant plus que les sites où les nouvelles technologies ont été mises au point sont restés cruciaux pour les emplois hautement qualifiés liés à ces technologies pendant des décennies<sup>85</sup>. Cette tendance semble également s'appliquer à la chaîne de valeur de l'IA générative : Un rapport Brookings de 2023 a révélé que près de la moitié des offres d'emploi pour des postes dans le domaine de l'IA générative au cours des 11 mois précédents étaient concentrées dans seulement six grandes zones métropolitaines américaines : San Francisco, San José (Californie), New York, Los Angeles, Boston et Seattle<sup>86</sup>.

Dans l'ensemble, **nous constatons des problèmes de concurrence sur le marché de la formation à l'IA générative pour trois raisons : Premièrement, il existe d'importantes économies d'échelle et de gamme, telles que des coûts fixes élevés pour la formation, qui rendent l'entrée sur le marché coûteuse. Deuxièmement, les LLM améliorent considérablement leurs performances lorsqu'ils passent à l'échelle supérieure, notamment en raison de leurs « capacités émergentes », ce qui récompense les fournisseurs les plus importants. Par conséquent, le marché des modèles de fondations est un marché où le plus grand nombre gagne, car les utilisateurs veulent généralement utiliser le meilleur modèle de fondations pour leur objectif. Il est donc très difficile pour les jeunes entreprises de concurrencer les fournisseurs de modèles de base bien établis et cela renforce notre point de vue selon lequel le partenariat stratégique entre OpenAI et Microsoft devrait être considéré comme une forme d'intégration verticale susceptible de nuire à la concurrence. Troisièmement, il y a une pénurie d'expertise humaine pour développer des modèles de base. Cela limite encore la concurrence sur ce marché et le rend particulièrement difficile pour les entreprises européennes. Cette évaluation est certainement valable à court et à moyen terme, mais notre examen de la littérature actuelle et les preuves du rattrapage récent par des modèles de fondation concurrents (Gemini, Mistral) suggèrent que ces restrictions de concurrence pourraient s'atténuer à plus long terme.**

### 3.3 Déploiement d'IA générative : Services en aval

Les applications et services IA générative sont des solutions logicielles spécifiques développées à partir de modèles de base, adaptées à des tâches distinctes telles que le service à la clientèle, la génération de contenu ou le conseil spécialisé. La clé du succès dans ce domaine réside dans la capacité à affiner les modèles de base avec des données de niche ou propriétaires, ce qui permet de créer des applications hautement spécialisées et compétitives<sup>87</sup>. Ce secteur verra probablement l'entrée de fournisseurs de services d'IA existants et de nouveaux acteurs de niche, notamment parce qu'il existe de

<sup>83</sup> [Chapitre 6 : La guerre des talents \(stateofai2019.com\)](#).

<sup>84</sup> [La géographie de l'IA | Brookings](#).

<sup>85</sup> [The Diffusion of New Technologies | NBER](#).

<sup>86</sup> [Construire des villes d'IA : How to spread the benefits of an emerging technology across more of America | Brookings](#).

<sup>87</sup> Voir : [Exploring opportunities in the gen AI value chain | McKinsey](#).

nombreux modèles en libre accès (voir ci-dessus). Toutefois, les applications et services les plus performants, et donc les plus compétitifs, nécessiteront des modèles de pointe, actuellement GPT-4. Dans ce contexte, il sera de plus en plus difficile, sur le marché du déploiement de l'IA générative, de concurrencer les grandes entreprises technologiques telles que Microsoft et Google, qui peuvent inclure ces modèles de pointe dans tous leurs services et produits grâce aux partenariats décrits ci-dessus et qui ont souvent déjà un pouvoir de marché sur ces marchés en aval (par exemple, pour les logiciels de bureautique et les moteurs de recherche).

Dans ce contexte, d'autres problèmes de concurrence pourraient apparaître à l'avenir en raison d'un nouveau format de « magasin d'applications » récemment inauguré par OpenAI, qui invite les développeurs et les entreprises à créer des « GPT » personnalisés qui ajoutent des fonctionnalités au chatbot (les GPT peuvent être configurés en définissant des invites et des paramètres spécifiques dans ChatGPT afin d'obtenir un résultat particulier)<sup>88</sup>. Cette récente transformation de ChatGPT en « plateforme d'applications » pourrait introduire une nouvelle dynamique dans la distribution du pouvoir de marché dans le secteur en aval des services et produits basés sur l'IA générative. En fournissant des fondations centralisées pour le déploiement et l'accès aux applications alimentées par GPT, ChatGPT facilite un écosystème rationalisé où les développeurs peuvent atteindre un large public avec un minimum de friction. Toutefois, cette centralisation soulève des inquiétudes quant au potentiel de transfert du pouvoir de marché en aval, en particulier si la plateforme commence à reproduire la fonctionnalité d'applications tierces réussies (comme Amazon l'a fait pour certains de ses vendeurs les plus populaires<sup>89</sup>). En outre, les effets de réseau inhérents à la plateforme, caractérisés par l'augmentation de la valeur du service avec chaque utilisateur et application supplémentaire, pourraient conduire à des effets de verrouillage importants, d'autant plus qu'OpenAI pourrait utiliser le nombre accru d'« invites et de réponses » pour améliorer encore son modèle dans les itérations futures (GPT-5, GPT-6, etc.). Si un tel format de « magasin d'applications » devenait courant, cela aurait un impact sur la fidélisation des utilisateurs, sur l'attrait des LLM individuels et, par conséquent, sur les barrières à l'entrée pour les concurrents potentiels.

**Dans l'ensemble, s'il existe de nombreux modèles de base qui facilitent l'accès au marché des applications et services d'IA générative en aval, la dynamique du « plus fort prend le plus faible » signifie que seule une poignée de fournisseurs connaîtra le succès à long terme. Dans ce contexte, les positions dominantes actuelles des grandes entreprises technologiques sur bon nombre de ces marchés en aval (et leurs partenariats avec les principaux développeurs de modèles de base) signifient qu'elles vont bientôt s'emparer de ces marchés et offrir elles-mêmes ces services. En d'autres termes, cela crée deux désavantages pour les nouveaux arrivants potentiels sur le marché : Premièrement, les jeunes entreprises doivent rivaliser avec les grandes sociétés telles que Microsoft et Google qui ont un accès préférentiel aux modèles avancés. Deuxièmement, les entreprises en place font de la nouvelle technologie une caractéristique de leurs services plus anciens et tirent ainsi parti de leur pouvoir de marché et de leurs clients existants sur ces marchés en aval. À l'avenir, un troisième obstacle pourrait résulter de la diffusion de modèles de « magasins d'applications » assez semblables à la position de Google sur le marché des logiciels pour smartphones, où Google propose un système d'exploitation et un magasin d'applications, ainsi que des applications importantes.**

<sup>88</sup> Voir le rapport dans : [OpenAI's New App Store Could Turn ChatGPT Into an Everything App | WIRED](#). Pour un point de vue plus sceptique, voir : [Le GPT Store n'est pas l'« app store » de ChatGPT - mais il est tout de même important pour les spécialistes du marketing \(econsultancy.com\)](#).

<sup>89</sup> Voir : [Amazon a copié des produits et truqué des résultats de recherche, selon des documents \(reuters.com\)](#).

**Nous pensons qu'il est probable qu'une telle situation se produise également dans l'écosystème de l'IA générative si les autorités de la concurrence n'interviennent pas rapidement.**

## 4 Considérations sur la résilience

Si l'objectif premier du droit européen de la concurrence est de garantir la liberté économique et de protéger les intérêts des consommateurs, la nature évolutive de l'IA générative, avec ses graves implications géopolitiques, politiques et sécuritaires, exige une perspective plus large. Comme l'a récemment fait remarquer Paul Tucker, « si le bien-être futur du consommateur (grossièrement, des prix plus bas pour des biens de qualité équivalente) est le seul critère, alors le régime ne prendra pas en compte les risques et les coûts potentiels pour la stabilité du système politique d'un pouvoir économique privé concentré se transformant en un pouvoir politique concentré »<sup>90</sup>. Ce raisonnement n'est pas étranger au droit européen de la concurrence qui, depuis sa création, a été influencé par la pensée ordolibérale sur les relations entre le pouvoir économique, la stabilité politique et la liberté individuelle<sup>91</sup>. Cependant, la ligne à suivre est très fine : D'une part, l'application du droit de la concurrence doit toujours être fondée sur des règles et ne pas être trop discrétionnaire. D'autre part, d'éminents chercheurs ont mis en garde contre le fait que « le rôle de l'antitrust dans la promotion de la concurrence pourrait bien être sapé si l'antitrust est appelé à traiter des problèmes qui ne sont pas directement liés à la concurrence ou si l'on attend de lui qu'il le fasse »<sup>92</sup>.

En ce qui concerne la concurrence dans l'IA générative, cette conception plus large du droit de la concurrence implique que les responsables de l'application du droit de la concurrence devraient également prendre en compte des préoccupations générales qui ne relèvent pas d'une définition étroite du droit de la concurrence, mais qui peuvent également être pertinentes compte tenu de l'objectif de la Commission de promouvoir l'autonomie stratégique dans la sphère numérique (« souveraineté numérique »)<sup>93</sup>. En fait, la commissaire européenne chargée de la concurrence a récemment fait remarquer que « les marchés numériques ont une grande portée, affectant parfois l'économie d'une manière inattendue », ce qui implique que « la politique de la concurrence doit travailler de concert avec la réglementation numérique dans cette économie dynamique qui évolue rapidement »<sup>94</sup>. Nous soulignons ici plusieurs préoccupations qui se recoupent à l'intersection de la concurrence et de la gouvernance de l'IA, à savoir les risques de sécurité existentielle posés par les systèmes d'IA, les risques pour le discours politique et la démocratie, la montée des « deepfakes » et de la désinformation, le potentiel des services malveillants et les applications militaires de l'IA. Pour chacun de ces risques, la DG COMP devrait examiner si une concurrence accrue pourrait les réduire, voire les exacerber. Nous passons en revue ces risques ci-dessous.

En substance, notre recommandation suit des stratégies similaires qui ont été proposées pour une modeste reconceptualisation de « l'antitrust vert », un mouvement qui préconise d'assouplir l'application de la loi pour permettre des collaborations favorables à la durabilité avec des effets

<sup>90</sup> Voir : [Un entretien avec Sir Paul Tucker - The Platform Law Blog](#).

<sup>91</sup> Voir : Anselm Küsters (2023), *The Making and Unmaking of Ordoliberal Language. A Digital Conceptual History of European Competition Law*, Studien zur europäischen Rechtsgeschichte 340, Francfort-sur-le-Main : Klostermann 2023.

<sup>92</sup> Shapiro, Carl, (2018), *Antitrust in a time of populism*, International Journal of Industrial Organization, 61, issue C, p. 714-748, ici : p. 716.

<sup>93</sup> Armin Steinbach (2023), [EU's Turn to 'Strategic Autonomy' : Leeway for Policy Action and Points of Conflict | European Journal of International Law | Oxford Academic \(oup.com\)](#).

<sup>94</sup> Discours du 19 février 2024, Bruxelles, à : [Renouveler l'Europe au Parlement européen \(europa.eu\)](#).



anticoncurrentiels<sup>95</sup>. Si la durabilité ne doit pas être l'objectif premier du droit de la concurrence (qui relève de la liberté économique et du droit de la consommation), il convient néanmoins d'examiner quels types de mesures d'application pourraient également avoir des retombées positives sur la durabilité ou, du moins, des avantages indirects pour l'environnement<sup>96</sup>. En effet, en discutant des compromis potentiels entre l'application du droit de la concurrence et le Green Deal de l'UE, la Commission elle-même a demandé si « nous pouvons faire plus, pour appliquer nos règles d'une manière qui soutienne mieux le Green Deal »<sup>97</sup>. De même, la Commission devrait se demander dans quels cas l'application du droit de la concurrence, centrée sur la technologie, pourrait contribuer à réduire les risques non économiques plus larges de l'IA générative, tels que les risques existentiels, et dans quels cas il conviendrait d'être plus prudent avec une application rapide et d'examiner plutôt les éventuels effets secondaires négatifs en dehors du secteur économique qui seraient alimentés par une pression concurrentielle accrue. En d'autres termes, la Commission devrait renforcer son application dans les cas qui permettent de sanctionner des comportements qui sont à la fois anticoncurrentiels et qui entraînent des dommages sociétaux plus importants. Il est possible d'obtenir beaucoup en sélectionnant correctement les cas et en fixant des priorités<sup>98</sup>, sans modifier la substance de la loi.

**Risques existentiels** : À mesure que les modèles de fondation améliorent leurs performances lors de la mise à l'échelle, ils augmentent la probabilité de phénomènes imprévisibles, y compris de risques graves<sup>99</sup>. Le développement et le déploiement de systèmes d'IA générative dans toute l'Europe pourraient ainsi poser, ou accroître, des risques existentiels potentiels, y compris les conséquences involontaires de la prise de décision autonome. La concurrence peut jouer un double rôle à cet égard. D'une part, une concurrence accrue pourrait stimuler les progrès rapides et la diffusion des technologies de l'IA, dépassant éventuellement les cadres réglementaires et éthiques tels que la loi européenne sur l'IA. D'autre part, un environnement concurrentiel pourrait favoriser l'innovation en matière de sécurité et d'éthique de l'IA, les entreprises cherchant à se différencier par des pratiques responsables. Les autorités chargées de faire respecter la concurrence devraient encourager la transparence (comme l'exige la loi sur l'IA pour les modèles à haut risque) et les normes éthiques en tant que facteurs concurrentiels. Alors que la réglementation fournira toujours une solution de second choix, une concurrence accrue et une responsabilité plus claire peuvent optimiser le compromis entre la limitation des risques et l'exploitation des opportunités en fonction des préférences, comme l'a fait valoir le cep dans une étude récente sur la théorie des jeux<sup>100</sup>.

**Risques pour le discours politique et la démocratie** : La technologie de l'IA générative pourrait conduire à un contrôle centralisé des technologies linguistiques et à une influence analogue sur le discours politique et la formation de l'opinion publique dans les démocraties. Lors d'une récente table ronde au Parlement européen, des experts de l'industrie ont souligné le rôle crucial du droit de la concurrence dans la protection de la démocratie dans le secteur de l'IA, se sont inquiétés du potentiel de l'IA

<sup>95</sup> Voir, par exemple : [Sustainability, antitrust and the EU Green Deal | Netherlands | Global law firm | Norton Rose Fulbright](#).

<sup>96</sup> Marco Colino, Sandra, L'empreinte environnementale de l'antitrust : Redefining the Boundaries of Green Antitrust (19 février 2024). North Carolina Law Review, Vol. 103, No. 1, 2024, The Chinese University of Hong Kong Faculty of Law Research Paper No. 2024-01, Disponible sur SSRN.

<sup>97</sup> [conférence 2021 - Commission européenne \(europa.eu\)](#).

<sup>98</sup> Voir : Brook, Or et Cseres, Kati, Policy Report : Priority Setting in EU and National Competition Law Enforcement (28 septembre 2021). Disponible à l'adresse SSRN : [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3930189](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3930189).

<sup>99</sup> Sur les caractéristiques émergentes de l'IA générative, voir : [The Unpredictable Abilities Emerging From Large AI Models | Quanta Magazine](#). Sur les risques émergeant lors d'une polycrise, voir : Küsters (2023), [cepAdhoc AI as Systemic Risk in a Polycrisis.pdf](#).

<sup>100</sup> Küsters et Vöpel (2024), Weniger KI-Risiken durch mehr Wettbewerb, cepInput.

à exacerber les problèmes existants dans le discours numérique et ont appelé à une réglementation et à une application garantissant l'équité concurrentielle afin de protéger les valeurs sociétales et démocratiques<sup>101</sup>. En effet, l'un des principaux risques liés aux LLM qui sous-tendent les services d'IA générative est leur capacité à saper la démocratie en concentrant les technologies linguistiques entre les mains de quelques entreprises privées qui pourraient dicter l'avenir du discours politique et de la délibération publique<sup>102</sup>. Comme décrit ci-dessus, les modèles linguistiques sont formés à partir de diverses sources de données telles que les actualités et les livres, mais qui décide des données à inclure - et de celles à omettre ? Si les applications d'IA générative telles que ChatGPT deviennent la prochaine « plateforme internet », remplaçant les services de recherche traditionnels tels que Google, elles auront un impact considérable sur la formation de l'opinion publique et l'accès à l'information. Les conclusions des informaticiens montrent que les modèles linguistiques préformés actuels présentent des biais politiques qui renforcent la polarisation présente dans les données d'apprentissage sous-jacentes, propageant des biais sociaux dans les détecteurs de discours haineux et de désinformation<sup>103</sup>. Par conséquent, un nombre limité de modèles linguistiques avec leurs biais inhérents, motivés par des choix idéologiques à l'origine de leur développement, représente un risque pour le processus démocratique en privatisant ce qui constitue le discours public et en déformant potentiellement les informations auxquelles les citoyens peuvent avoir accès. Un bon exemple est la motivation derrière le chatbot Grok d'Elon Musk, formé à partir de données Twitter et avec un parti pris explicite « anti-woke ». Un autre exemple concerne le récent accord d'OpenAI avec Springer, mentionné plus haut, qui prévoit que les utilisateurs de ChatGPT recevront des résumés d'articles « sélectionnés » dans les publications d'Axel Springer, parfois même s'ils se trouvent derrière un paywall<sup>104</sup>. En conséquence, les autorités européennes de la concurrence devront peut-être examiner à l'avenir les implications de la sélection des données, les biais inhérents et la possibilité que les technologies de l'IA générative deviennent des plateformes dominantes pour l'accès à l'information, ce qui pourrait nécessiter de nouveaux cadres pour garantir la diversité dans les espaces de communication numérique.

**Deepfakes et désinformation alimentée par l'IA générative** : En lien avec la question plus large du discours politique, les progrès rapides de l'IA dans la génération de textes et d'images pourraient conduire à une escalade de la désinformation politique, ce qui est pertinent à la lumière des nombreuses élections qui auront lieu en 2024. Selon le dernier rapport de Freedom House, au moins 47 gouvernements ont déployé des commentateurs pour manipuler les discussions en ligne en leur faveur au cours de la période couverte, tandis que des outils d'IA générative ont été utilisés dans au moins 16 pays pour semer le doute, salir les opposants ou influencer le débat public<sup>105</sup>. L'utilisation de l'IA par les candidats à la présidence en Argentine lors de la campagne de 2023, la « première élection de l'IA »<sup>106</sup>, souligne encore l'influence croissante de cette technologie dans les arènes politiques. Compte tenu de la capacité de l'IA générative à créer des deepfakes réalistes et à diffuser des informations erronées, ce qui pourrait miner la confiance sociale et l'intégrité de l'information, une dynamique où tout le monde gagne sur un marché oligopolistique pourrait exacerber le problème en incitant les développeurs de modèles dans une « course aux rats »<sup>107</sup> pour permettre la production d'un contenu plus

<sup>101</sup> Voir le résumé dans : [Garantir la concurrence dans l'IA préservera également la démocratie, selon les experts - Euractiv](#).

<sup>102</sup> Pour cet argument, voir l'essai : [Celui qui contrôle les modèles linguistiques contrôle la politique - HANNES BAJOHR](#).

<sup>103</sup> [\[2305.08283\] Des données de pré-entraînement aux modèles de langage et aux tâches en aval : Tracking the Trails of Political Biases Leading to Unfair NLP Models \(arxiv.org\)](#).

<sup>104</sup> [OpenAI conclut un accord avec Axel Springer sur l'octroi de licences pour l'entraînement de modèles | TechCrunch](#).

<sup>105</sup> [Le pouvoir répressif de l'intelligence artificielle | Freedom House](#).

<sup>106</sup> [L'Argentine est-elle la première élection de l'I.A. ? - The New York Times \(nytimes.com\)](#).

<sup>107</sup> Voir à ce sujet notre étude : Küsters et Vöpel (2024), Weniger KI-Risiken durch mehr Wettbewerb, cepInput.

attrayant, bien qu'il soit potentiellement trompeur. Cependant, un paysage concurrentiel pourrait également encourager le développement de contre-technologies pour la détection des deepfakes et la vérification des informations comme paramètre supplémentaire permettant aux fournisseurs de modèles de se distinguer par la qualité de leurs produits<sup>108</sup>. Comme indiqué ci-dessus, les autorités chargées de l'application du droit de la concurrence devraient donc surveiller la structure du marché de l'IA générative afin d'empêcher un contrôle monopolistique de ces technologies, en garantissant l'accessibilité et la diversité des outils de lutte contre la désinformation. Cependant, les recherches existantes montrent que le pouvoir de persuasion des publicités politiques et de la désinformation en ligne est souvent surestimé et suggèrent que la politique devrait se concentrer sur la prévention des abus dans les élections plus petites et plus locales (où l'impact est plus important) et sur l'atténuation des préjugés<sup>109</sup>.

**Services malveillants** : Une récente analyse approfondie de plus de 200 services malveillants concrets utilisant l'IA a mis en évidence le rôle des modèles de fondation, en particulier ceux développés par OpenAI, dans l'alimentation de divers outils et services illicites tels que BadGPT, XXXGPT et Evil-GPT<sup>110</sup>. Sur la base de données provenant de 13 353 inscriptions sur neuf places de marché et forums clandestins, l'étude a révélé qu'une grande majorité (93,4 %) des services malveillants offraient des capacités de génération de logiciels malveillants, certains échappant même à la détection des virus. L'étude a révélé que les modèles GPT-3.5 et GPT-4 d'OpenAI étaient les backends les plus couramment exploités dans ces services malveillants, ce qui implique que ces modèles semblent être actuellement les plus efficaces et qu'ils ne peuvent pas, pour l'instant, être assimilés à des modèles open source gratuits développés par des start-ups. Les résultats suggèrent que les autorités européennes de la concurrence pourraient être amenées à examiner la domination et l'utilisation des modèles fondamentaux d'OpenAI dans la création de services d'IA malveillants, afin de s'assurer qu'il n'y a pas d'abus de pouvoir de marché préjudiciable aux consommateurs. Le fait que les activités illégales aient été menées avec des modèles fermés, tels que ceux d'OpenAI, vraisemblablement via un accès API, et non avec des modèles à accès ouvert facilement disponibles, tels que Llama2, suggère que l'argument commun selon lequel l'accès ouvert et l'interopérabilité des modèles contribueront à réduire les abus n'est pas tout à fait correct.

**Applications militaires de l'IA générative** : enfin, l'utilisation de l'IA générative dans des applications militaires soulève des questions d'éthique et de sécurité. Cette question est devenue aiguë, car OpenAI a récemment levé l'interdiction d'utiliser ses outils d'IA à des fins militaires, à la suite d'une recherche sur l'IA menée conjointement avec le ministère américain de la défense<sup>111</sup>. Une étude empirique récente portant sur les risques d'escalade des actions entreprises par les grands modèles de langage et les agents d'IA basés sur ces modèles dans des wargames simulés a révélé que « les modèles ont tendance à développer une dynamique de course aux armements, conduisant à un conflit plus important et, dans de rares cas, même au déploiement d'armes nucléaires »<sup>112</sup>. La loi européenne sur l'IA exempte les applications militaires de son champ d'application. Si une concurrence accrue dans le

<sup>108</sup> Voir à ce sujet notre étude : Küsters et Vöpel (2024), Weniger KI-Risiken durch mehr Wettbewerb, cepInput.

<sup>109</sup> Voir l'étude bibliographique dans : Scott Babwah Brennen & Matt Perault (2023), The new political ad machine : Policy frameworks for political ads in an age of AI, [GAI-and-political-ads.pdf \(unc.edu\)](#).

<sup>110</sup> Zilong Lin, Jian Cui, Xiaojing Liao, et XiaoFeng Wang (2024), Malla : Demystifying Real-world Large Language Model Integrated Malicious Services, [2401.03315.pdf \(arxiv.org\)](#).

<sup>111</sup> [OpenAI lève discrètement l'interdiction d'utiliser ses outils d'IA à des fins militaires \(cnbc.com\)](#).

<sup>112</sup> Juan-Pablo Rivera, Gabriel Mukobi, Anka Reuel, Max Lamparth, Chandler Smith, Jacquelyn Schneider (2024), Escalation Risks from Language Models in Military and Diplomatic Decision-Making, [2401.03408.pdf \(arxiv.org\)](#).

secteur de l'IA générative pouvait conduire à une prolifération de systèmes d'IA militaires avancés et potentiellement déstabiliser la dynamique de la sécurité mondiale, elle pourrait également stimuler l'innovation dans les technologies d'IA défensives en améliorant la sécurité. Bien qu'il ne soit certainement pas du ressort des autorités chargées de l'application du droit de la concurrence de trouver cet équilibre délicat, le contexte géopolitique actuel suscitera probablement des appels à prendre en compte les implications stratégiques de l'IA générative dans la sécurité nationale et internationale lors de la prise de décisions sur des cas concrets.

Dans l'ensemble, notre aperçu de la « résilience » des facteurs non économiques qui pourraient jouer un rôle dans les futures affaires d'IA générative, repose sur l'argument suivant : Au lieu de se concentrer uniquement sur la possibilité qu'une concurrence accrue par l'application de la législation antitrust qui bloque complètement le développement des services européens d'IA générative ou résolve tous les problèmes émergents de l'IA générative (qu'ils soient techniques ou sociétaux), la Commission européenne devrait réfléchir aux façons dont l'application de la loi pourrait avoir des effets positifs (indirects) sur ces questions plus vastes en jeu. Peut-être qu'une pression concurrentielle accrue sur le marché oligopolistique de l'IA générative pourrait conduire à des modèles plus sûrs, en réduisant les incitations à la « course aux rats »<sup>113</sup> ? Que cette hypothèse résiste ou non à un examen théorique et empirique plus approfondi, le fait d'aborder activement ces compromis plus vastes en publiant des orientations et des analyses pourrait contribuer à fixer les attentes et à façonner le discours politique sur l'IA.

## 5 Conclusion

La chaîne de valeur de l'IA générative est un écosystème à multiples facettes caractérisé par trois parties. La première partie concerne l'infrastructure de l'IA générative, qui comprend le marché des ensembles de données à grande échelle et le marché de la puissance de calcul. Ces deux éléments sont essentiels pour la formation et le développement de modèles de base, qui constituent la deuxième partie de la chaîne de valeur de l'IA. La troisième partie est le marché (ou les marchés) des services et applications en aval B2B ou B2C qui utilisent un modèle de base comme intrant. Pour chacune de ces parties, cet Adhoc du cep a étudié les problèmes de concurrence potentiels. En particulier, nous avons cherché à savoir si la taille est importante (par exemple, des coûts fixes élevés), s'il existe des coûts de changement significatifs et si des ressources rares limitent la concurrence. Quelles sont les principales conclusions de notre analyse ?

### 1. Infrastructure de l'IA générative

En ce qui concerne les données d'entraînement, nous ne voyons actuellement pas de problèmes de concurrence car des recherches récentes indiquent que les LLM avancés nécessiteront de moins en moins de données pour atteindre des performances supérieures. Néanmoins, les accords de licence exclusive entre les fournisseurs d'informations en ligne ou d'autres propriétaires d'ensembles de données de haute qualité devraient être surveillés de près car ils pourraient rendre plus difficile l'obtention de données de haute qualité pour les nouveaux concurrents.

En ce qui concerne la puissance de calcul, nous constatons des problèmes de concurrence dus à une pénurie de puces adéquates. Cette pénurie est due au fait que le marché des puces de pointe

---

<sup>113</sup> Voir à ce sujet notre étude : Küsters et Vöpel (2024), Weniger KI-Risiken durch mehr Wettbewerb, cepInput.

est dominé par une entreprise (NVIDIA). En raison de contraintes technologiques, cette entreprise ne peut pas augmenter facilement la production de ces puces. La pénurie de puces adéquates et le manque d'entrants sur le marché de l'informatique du cloud font que les grandes entreprises technologiques s'approprient la majeure partie de cette ressource rare. Bien qu'elles offrent aux jeunes entreprises l'accès à cette ressource par le biais de services de cloud, cela ne résout pas le problème. Dans une certaine mesure, cela aggrave même le problème pour trois raisons : Premièrement, il est à craindre que les start-ups de premier plan telles qu'OpenAI, et les opérateurs historiques tels que les grandes entreprises technologiques, soient plus susceptibles d'obtenir un accès prioritaire à ces services de cloud pour la formation à l'IA, étant donné qu'ils concluent des accords pour détenir de plus grandes grappes de calcul. Deuxièmement, les jeunes entreprises sont confrontées à des coûts de changement si elles souhaitent changer de fournisseur de services de cloud. Troisièmement, certains fournisseurs de services de cloud développent leurs propres puces et/ou leurs propres modèles d'IA générative, ce qui augmente le risque d'intégration verticale et l'émergence d'écosystèmes avec des effets de verrouillage. Cette combinaison unique d'une forte demande, d'un petit nombre de fournisseurs de puissance de calcul, de coûts de changement élevés et d'une intégration verticale croissante présente, selon notre évaluation, des risques non négligeables pour la concurrence, pouvant par exemple prendre la forme d'autoréférencement ou de discrimination.

## **2. Formation de l'IA générative (modèles de base)**

En ce qui concerne les modèles de fondation, nous constatons des problèmes de concurrence pour trois raisons : Premièrement, il existe d'importantes économies d'échelle et de gamme, telles que des coûts fixes élevés pour la formation, qui rendent l'entrée sur le marché coûteuse. Deuxièmement, les LLM améliorent considérablement leurs performances lorsqu'ils passent à l'échelle supérieure, notamment en raison de leurs « capacités émergentes », récompensant ainsi les plus grands fournisseurs. Par conséquent, le marché des modèles de fondations est un marché de type « winner-takes-most », car les utilisateurs veulent généralement utiliser le meilleur modèle de fondations pour leur objectif. Il est donc très difficile pour les jeunes entreprises de concurrencer les fournisseurs de modèles de base bien établis et cela renforce notre point de vue selon lequel le partenariat stratégique entre OpenAI et Microsoft devrait être considéré comme une forme d'intégration verticale susceptible de nuire à la concurrence. Troisièmement, il y a une pénurie d'expertise humaine pour développer des modèles de base. Cela limite encore la concurrence sur ce marché et le rend particulièrement difficile pour les entreprises européennes. Cette évaluation est certainement valable à court et à moyen terme, mais notre examen de la littérature actuelle et les preuves du rattrapage récent par des modèles de fondation concurrents (Gemini, Mistral) suggèrent que ces restrictions de concurrence pourraient s'atténuer à plus long terme.

## **3. Déploiement de l'IA générative (applications et services)**

En ce qui concerne les applications et les services d'IA générative en aval, la dynamique du « winner-takes-most » signifie que seule une poignée de fournisseurs connaîtra le succès à long terme. Dans ce contexte, les positions dominantes actuelles des grandes entreprises technologiques sur bon nombre de ces marchés en aval (et leurs partenariats avec les principaux développeurs de modèles de base) signifient qu'elles vont bientôt s'emparer de ces marchés et offrir elles-mêmes ces services. En d'autres termes, cela crée deux désavantages pour les nouveaux arrivants potentiels sur le marché : Premièrement, les jeunes entreprises doivent rivaliser avec les grandes

sociétés telles que Microsoft et Google qui ont un accès préférentiel aux modèles avancés. Deuxièmement, les entreprises en place font de la nouvelle technologie une caractéristique de leurs services plus anciens et tirent ainsi parti de leur pouvoir de marché et de leurs clients existants sur ces marchés en aval. À l'avenir, un troisième obstacle pourrait résulter de la diffusion de modèles de « magasins d'applications » assez semblables à la position de Google sur le marché des logiciels pour smartphones, où Google propose un système d'exploitation et un magasin d'applications ainsi que d'importantes applications. Nous pensons qu'il est probable qu'une telle situation se produise également dans l'écosystème de l'IA générative si les autorités de la concurrence n'interviennent pas rapidement.

En fin de compte, la compréhension de la dynamique de chaque niveau de la chaîne de valeur est cruciale pour les responsables de l'application du droit de la concurrence de l'UE lorsqu'ils élaborent leur stratégie pour s'attaquer au marché de l'IA générative. En outre, ils doivent également **adopter une perspective holistique pour comprendre le degré d'intégration verticale de certains acteurs du marché**. Un exemple clé est celui de Microsoft et de ses partenariats stratégiques avec OpenAI et Mistral AI, qui peuvent être considérés comme un seul acteur du marché ou, au vu du droit de la concurrence, comme une « entreprise ». Selon notre évaluation, cette entreprise occupe une position forte dans la chaîne de valeur en raison de sa possession de données de haute qualité, de serveurs de cloud, de capital humain, de son contrôle sur plusieurs modèles de fondations de premier plan et de son pouvoir de marché existant dans les services bureautiques en aval. Sur ce dernier point les applications d'IA générative deviendront très pertinentes et lucratives sur le plan commercial. Cela s'applique également, mais dans une moindre mesure, à Google, Meta et Amazon.

À mesure que cette technologie continue d'évoluer, il sera impératif que la DG Concurrence, par son analyse et la formulation de lignes directrices, joue un rôle central en façonnant activement le paysage concurrentiel de cette technologie transformatrice. Si l'objectif premier du droit européen de la concurrence est de préserver la liberté économique et le bien-être des consommateurs, les caractéristiques uniques de la chaîne de valeur de l'IA générative exigent une approche plus large. En considérant les risques existentiels, l'impact sur l'intégrité de l'information et peut-être même les implications pour la sécurité, la DG COMP peut aligner ses mesures d'application sur les objectifs plus larges de la Commission en matière de souveraineté numérique et de bien-être de la société. Cette approche nécessite une compréhension nuancée de l'interaction entre la concurrence et ces externalités, garantissant que l'application de la concurrence contribue positivement à la gestion des risques et à l'exploitation des opportunités présentées par l'IA générative.





**Auteur :**

Anselm Küsters, LL.M., chef de la division Numérisation et nouvelles technologies

[kuesters@cep.eu](mailto:kuesters@cep.eu)

Matthias Kullas, chef de la division Marché intérieur et concurrence

[kullas@cep.eu](mailto:kullas@cep.eu)

Traduit depuis l'allemand par Thomas Plancq, chargé de communication

**Centrum für Europäische Politik** FREIBURG | BERLIN

Kaiser-Joseph-Straße 266 | D-79098 Freiburg

Schiffbauerdamm 40 Raum 4315 | D-10117 Berlin

Tél. + 49 761 38693-0

Le **Centrum für Europäische Politik** FREIBURG | BERLIN, le **Centre de Politique Européenne** PARIS, et le **Centro Politiche Europee** ROMA forment le **réseau des Centres de Politique Européenne** FREIBURG | BERLIN | PARIS | ROMA.

Exempt d'intérêts particuliers et neutre sur le plan politique, le réseau des centres de politique européenne fournit une analyse et une évaluation de la politique de l'Union européenne, dans le but de soutenir l'intégration européenne et de défendre les principes d'un système économique de libre-échange.