

## La chaîne de valeur de l'intelligence artificielle : enjeux économiques et place de la France

*Solal CHARDON-BOUCAUD, Arthur DOZIAS, Charlotte GALLEZOT*

- L'intelligence artificielle (IA) désigne l'ensemble des techniques permettant à des machines de simuler l'intelligence humaine. La chaîne de valeur de l'IA est segmentée en trois principaux blocs (cf. Graphique) : (i) les intrants nécessaires au développement de systèmes et services d'IA (capacités de calcul, données, main d'œuvre spécialisée) ; (ii) la modélisation, qui comprend le développement des modèles d'IA généralistes (les modèles de fondation) et leur spécialisation ; et (iii) le déploiement de ces modèles chez les utilisateurs finaux.
- Concernant les intrants aux systèmes d'IA, la France ne dispose pas d'acteurs d'envergure comparable aux leaders mondiaux sur les marchés de la fabrication de puces et de la location de capacités de calcul. Elle bénéficie toutefois d'une main d'œuvre qualifiée et du dynamisme de son écosystème d'innovation.
- Concernant le développement de modèles, quelques acteurs français émergent mais ce segment est dominé par les grands acteurs du numérique, qui préexistaient à l'arrivée de cette technologie. Ces derniers bénéficient d'une intégration verticale, grâce à leurs accès privilégiés aux intrants en amont et aux canaux de diffusion en aval pour déployer leurs solutions d'IA (par exemple, les suites bureautiques). Ils ont par ailleurs noué des partenariats avec des acteurs émergents du secteur qui leur permettent une intégration horizontale des processus de production de l'IA.
- La domination du marché par quelques grands acteurs déjà bien implantés au préalable pose des questions d'efficacité économique, de concurrence et de souveraineté, avec le risque que la diffusion à l'ensemble de l'économie des gains de valeur ajoutée et de productivité liés à l'IA reste limitée.

### Chaîne de valeur de l'intelligence artificielle

Intrants	<ul style="list-style-type: none"><li>• Puissance de calcul : matériel informatique (e.g. GPU), services cloud, supercalculateurs publics</li><li>• Données : création, collecte, nettoyage, traitement</li><li>• Main d'œuvre qualifiée</li></ul>
Modélisation	<ul style="list-style-type: none"><li>• Développement et entraînement des modèles de fondation, e.g. GPT 4-o (OpenAI), Mistral 7B (Mistral AI)</li><li>• Spécialisation (ou réglage fin) des modèles pour des tâches spécifiques</li></ul>
Déploiement	<ul style="list-style-type: none"><li>• Déploiement de modèles d'IA générative pré-entraînés ou spécialisés et commercialisation de services destinés aux utilisateurs finaux : Chat GPT/Dall-E-3 (OpenAI), Mistral Le Chat (Mistral AI)</li></ul>

Source : Autorité de la Concurrence (2024), Avis 24-A-05 relatif au secteur de l'intelligence artificielle générative.

# 1. Intrants aux systèmes d'IA : la France est mal positionnée sur le segment du calcul, mais dispose d'une main d'œuvre qualifiée

Le développement de systèmes d'IA requiert des capacités de calcul importantes, de larges jeux de données de qualité, et de la main d'œuvre spécialisée. Si le marché du calcul (infrastructure ou location) est très concentré autour de quelques entreprises étatsuniennes, la France dispose d'un atout majeur : une main d'œuvre très qualifiée et un écosystème de recherche dynamique. La France pourrait toutefois mieux exploiter certains viviers de données, notamment pour contrebalancer l'influence des données en langue anglaise en matière d'entraînement des modèles.

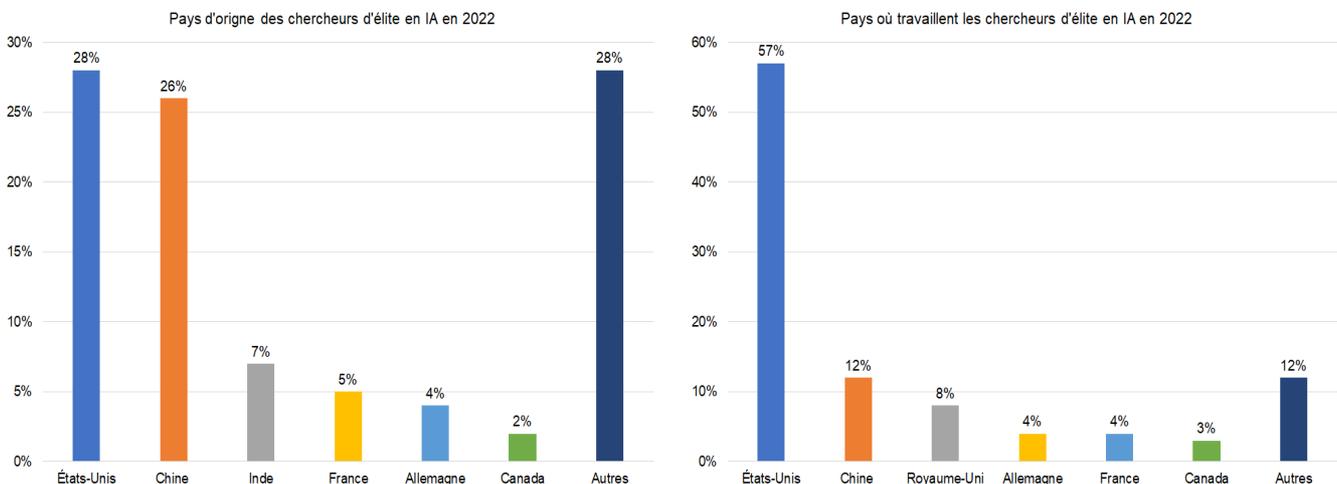
## 1.1 L'écosystème d'innovation et de recherche autour de l'IA est dynamique en France

Le développement de systèmes d'IA repose largement sur la qualité de l'écosystème de recherche et d'innovation. D'une part, le développement de modèles d'IA nécessite une main d'œuvre très qualifiée capable d'innover pour développer des modèles compétitifs et à l'état de l'art, dans un domaine évoluant très vite. À cet

égard, la qualité du système d'enseignement et de recherche qui contribue à former cette main d'œuvre est cruciale. D'autre part, l'émergence d'acteurs innovants est favorisée par l'existence d'un écosystème d'innovation qui rassemble les profils nécessaires et permet aux start-ups de croître et de trouver des financements en fonds propres.

L'écosystème de recherche et de formation est un atout pour la France, qui dispose d'une main d'œuvre bien formée et d'un large vivier de talents en mathématiques, informatique et ingénierie<sup>1</sup>. En 2022, la France était le 4<sup>e</sup> pays d'origine des chercheurs d'élite en IA derrière les États-Unis, la Chine et l'Inde, et le 5<sup>e</sup> pays où ils travaillent, derrière les États-Unis, la Chine, le Royaume-Uni et l'Allemagne<sup>2</sup> (cf. Graphique 1). La recherche française en IA présente en outre un niveau de publications scientifiques satisfaisant à l'échelle de l'Europe : en 2021, en nombre de publications, elle se situait au 4<sup>e</sup> rang européen derrière le Royaume-Uni, l'Allemagne et l'Italie et au 10<sup>e</sup> rang mondial<sup>3</sup>.

**Graphique 1 : Pays d'origine et pays où travaillent les chercheurs d'élite en IA en 2022**



Source : MarcoPolo (2023), *The Global AI Talent Tracker 2.0*.

Note : Les « chercheurs d'élite en IA » sont définis comme les auteurs d'articles sélectionnés pour des présentations orales à NeurIPS, qui représentent la classe d'articles la plus prestigieuse. Le taux d'acceptation des présentations orales était de 1,8 % en 2022.

(1) OCDE (2023), *Education at a glance 2023: OECD indicators*. Table B5.3.

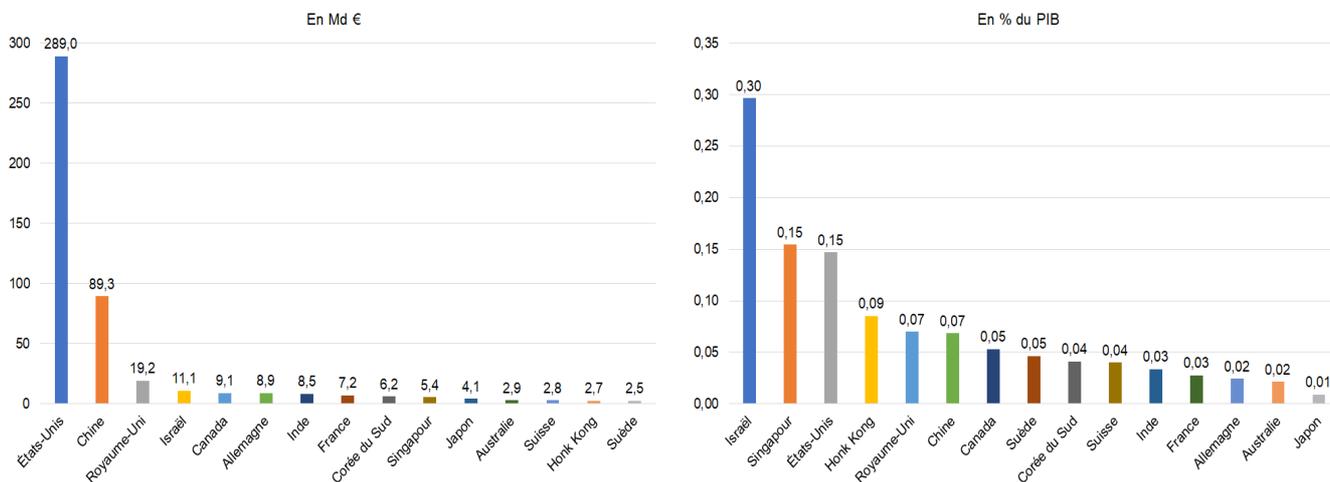
(2) MarcoPolo (2023), *The Global AI Talent Tracker 2.0*.

(3) Source : OECD.AI (2024), d'après des données OpenAlex. D'après la même source, la France occupe les mêmes positions si l'on se concentre uniquement sur les publications à « fort impact ». En nombre de publications tous domaines confondus, la France se classe 9<sup>e</sup> à l'échelle mondiale.

Cet écosystème est structuré autour d'organismes de recherche au rayonnement mondial<sup>4</sup> comme le CNRS ou l'Inria (Institut national de recherche en sciences et technologies du numérique) et de formations d'excellence, notamment en mathématiques<sup>5</sup>. Les centres de recherche français sont toutefois confrontés à des difficultés en matière de rétention des talents car ils pâtissent d'un manque d'attractivité par rapport à certaines institutions étrangères offrant de meilleures rémunérations et perspectives de carrières<sup>6</sup>. Par ailleurs, la « culture de l'innovation » s'est diffusée et approfondie au sein des principaux « pôles d'innovation »<sup>7</sup> français, notamment celui, en développement, de Paris. Ces pôles génèrent des externalités positives liées à la concentration de profils complémentaires, en particulier les chercheurs et les profils plus entrepreneuriaux<sup>8</sup>.

Les start-ups d'IA européennes sont parvenues à obtenir des financements privés assez importants au cours de la dernière décennie (entre 2013 et 2023, 29,5 Md€ pour l'Union européenne, dont 7,2 Md€ pour la France et 8,9 Md€ pour l'Allemagne notamment<sup>9</sup>) [cf. Graphique 2], mais nettement inférieurs aux États-Unis (289,0 Md€), à la Chine (89,3 Md€), et dans une moindre mesure, à certains pays asiatiques. Rapporté au PIB, la France est ainsi au-delà du 10<sup>e</sup> rang mondial en matière de financement des start-ups en IA : elle se place derrière le Royaume-Uni, la Suède, la Suisse, Israël, la Corée du Sud ou l'Inde, mais devance l'Allemagne, le Japon ou l'Australie (cf. Graphique 2).

**Graphique 2 : Investissement privé dans les levées de fonds liées à l'IA par pays entre 2013 et 2023**



Sources : Pour les données relatives aux investissements privés dans les levées de fonds en IA : Stanford University (2024), AI Index Report 2024, données issues du graph 4.3.9. Pour les données relatives au PIB : Banque Mondiale.

Notes : Pour présenter les chiffres en euros courants (et non en dollars courants), on utilise le taux de change moyen euro-dollar sur la période 2013-2023. Comme les données 2023 de PIB ne sont pas disponibles pour certains pays, on utilise comme proxy le PIB sur la période 2013-2022 en y ajoutant une année de PIB 2022 supplémentaire.

Si l'écosystème de capital-risque français et européen contribue de façon importante à ces levées de fonds, et semble performant pour les start-ups en amorçage, la disponibilité de fonds français et européens reste limitée pour les levées supérieures à 50 ou 100 M€<sup>10</sup>,

qui sont en général nécessaires pour accélérer le « passage à l'échelle internationale » (*scale-up*) des entreprises technologiques. En France et en Europe, les entreprises en phase de croissance plus avancée rencontrent ainsi des difficultés importantes à se

(4) HCERES, Rapport d'évaluation du CNRS (2023), Rapport d'évaluation de l'INRIA (2024).

(5) En 2023, 3 universités françaises étaient classées dans le top 10 du classement de Shanghai en mathématiques : Université Paris Saclay (n° 2), Sorbonne Université (n° 4), Paris cité (n° 7).

(6) Commission de l'intelligence artificielle (mars 2024), Rapport « IA : notre ambition pour la France », chapitre 2.2.4.

(7) Regroupement géographique et thématique de diverses ressources et organisations (start-ups, PME, organismes de recherche, etc.) permettant de stimuler la dynamique d'innovation.

(8) de Vasconcelos Gomes L.A., Figueiredo Facin A.L., Salerno M.S., Ikenami R.K. (2018), « Unpacking the innovation ecosystem construct: Evolution, gaps and trends », *Technological Forecasting and Social Change*.

(9) Stanford University (2024), AI Index Report 2024, données issues du graph 4.3.9 pour l'estimation pays et du graph 4.3.10 pour l'estimation UE.

(10) Hafied F., Rachiq C. et Rouleau G. (2021), « Capital-risque et développement des start-ups françaises », *Trésor-Éco*, n° 276.

financer auprès d'investisseurs européens<sup>11</sup>. Entre 2013 et 2023, parmi les levées de fonds des *scale-up* européennes<sup>12</sup>, 82 % avaient ainsi pour fonds d'investissement « lead »<sup>13</sup> un fonds extra-européen. Parmi les *scale-up* européennes qui ont été rachetées, plus de 60 % l'ont été par un acteur extra-européen<sup>14</sup>. Les initiatives Tibi<sup>15</sup>, au niveau français, et *Scale-up* Europe<sup>16</sup>, au niveau européen, cherchent à accroître la taille des fonds *tech* français et européens pour qu'ils puissent investir dans les levées supérieures à 100 M€. Les programmes d'investissement de type InvestEU<sup>17</sup> et, plus largement, le développement de l'Union des marchés de capitaux<sup>18</sup>, devraient contribuer aussi à combler le retard européen sur les levées de fonds *late stage*.

## 1.2 La France n'est pas aujourd'hui un acteur majeur du calcul

L'entraînement et le déploiement des modèles d'IA requièrent des capacités de calcul informatique importantes et coûteuses, en particulier pour les modèles de fondation. Pour répondre à ces besoins, une entreprise peut (i) acheter directement des capacités de calcul, (ii) louer des capacités de calcul (*cloud computing* ou éventuellement location d'infrastructures dédiées) ou (iii) plus marginalement, utiliser des infrastructures publiques de calcul (e.g. supercalculateur Jean Zay). Les calculs lors de l'entraînement et de l'utilisation de modèles d'IA se font principalement sur des processeurs graphiques (GPU ou *Graphics Processing Unit*), qui permettent des calculs en parallèle, ce qui en accroît la rapidité.

Le marché des GPU est très concentré autour de l'entreprise Nvidia, issue de la Silicon Valley. Celle-ci est en situation de quasi-monopole avec environ 85 % des parts de marché mondial en 2023<sup>19</sup>. Le deuxième fabricant (AMD) détient 10 % des parts de marché. La position dominante de Nvidia est renforcée par son expertise logicielle, en particulier la bibliothèque logicielle *Compute Unified Device Architecture* (CUDA) qui permet d'optimiser les calculs. Plusieurs acteurs travaillent actuellement sur des alternatives aux GPU<sup>20</sup> : des start-ups comme Cerebra Systems mais aussi des grands acteurs du numérique comme Google cherchent à optimiser les puces pour l'IA, notamment pour faire baisser la consommation d'énergie élevée des GPU<sup>21</sup>.

Le marché des infrastructures de calcul et de leur location est aussi très concentré autour de quelques grands acteurs américains, dits *hyperscalers*, face auxquels les acteurs français existants jouent un rôle marginal. En 2021, Amazon (AWS) détenait ainsi 46 % des parts du marché français du *cloud*, Microsoft (Azure) 17 %, et Alphabet (Google Cloud) 8 %<sup>22</sup>. Il existe toutefois des acteurs *cloud* plus spécialisés (*pure players*<sup>23</sup>), pour certains français (e.g. Scaleway, OVHcloud, Outscale), compétitifs par rapport aux grandes entreprises américaines sur certains services *cloud* spécifiques, mais qui pâtissent du caractère incomplet de leur catalogue<sup>24</sup>. En France, Scaleway a annoncé des investissements dans des infrastructures de calcul pour l'IA et un partenariat avec Nvidia pour offrir des services pour l'entraînement de modèles d'IA. Les autres acteurs de *cloud* français se positionnent sur l'inférence, *i.e.* le déploiement et l'utilisation d'un modèle une fois qu'il a déjà été entraîné.

- 
- (11) BEI (2022), « *La BEI soutient l'initiative paneuropéenne Scale-up Europe pour faire émerger des champions technologiques européens* ».
- (12) Définies comme les entreprises dont la valorisation était comprise entre 500 millions et 10 milliards de dollars au moment de la levée.
- (13) Dans le cadre d'un co-investissement en capital investissement, on appelle fonds *lead* celui qui investit le montant le plus élevé et qui a fixé les conditions de l'opération (valorisation, gouvernance, etc).
- (14) BEI (2024), "*The scale-up gap: Financial market constraints holding back innovative firms in the European Union*".
- (15) Lancée en 2019, Tibi est une initiative de place qui a pour objectif d'augmenter la capacité de financement des entreprises technologiques françaises en mobilisant l'épargne des investisseurs institutionnels. Lors de la première phase de l'initiative (période 2020-2022), 6 Md€ de fonds ont été mobilisés et ont permis d'apporter 30 Md€ à l'éco-système tech. Pour la deuxième phase de l'initiative, les investisseurs se sont engagés sur un montant de 7 Md€ d'ici fin 2026.
- (16) Ayant mené au European Tech Champions Initiative (ETCI), fonds de fonds géré par le Fonds européen d'investissement (FEI) doté actuellement de 3,85 Md€ et aidant des fonds à dépasser une taille critique de 1 Md€.
- (17) Successeur du plan Juncker (550 Md€ investis entre 2015 et 2020), le programme InvestEU a pour objectif de mobiliser 372 Md€ d'investissements d'ici à 2027, qui bénéficieront à une large variété d'acteurs, des bailleurs de logements sociaux aux grands équipementiers automobiles, en passant par des start-ups. Cette somme doit être atteinte par effet de levier, à partir de 26 Md€ de garantie budgétaire de l'UE. Il est mis en œuvre par le groupe BEI (dont le FEI), mais l'est aussi par d'autres intermédiaires financiers, comme Bpifrance ou la Caisse des dépôts.
- (18) Cf. Les recommandations de la mission Noyer (2024), « *Développer les marchés de capitaux européens pour financer l'avenir* ».
- (19) Sastry G. *et al.* (2024), "Computing Power and the Governance of Artificial Intelligence"
- (20) Le Monde Informatique (2024), Des start-ups poussent des alternatives au GPU pour l'IA.
- (21) Hardware T. (2023), "Nvidia's H100 GPUs will consume more power than some countries".
- (22) Autorité de la concurrence (juin 2023), *Avis 23-A-08 sur le fonctionnement concurrentiel de l'informatique en nuage ("cloud")*.
- (23) Un *pure player* est une entreprise exerçant dans un secteur d'activité unique non diversifié.
- (24) Autorité de la concurrence (2023), *op. cit.*

Bien que l'absence d'acteurs français de taille comparable aux leaders mondiaux dans la production de puces ou le *cloud computing* n'entrave pas actuellement l'accès des entreprises françaises au calcul (les acteurs français bénéficiant aujourd'hui des mêmes conditions que les entreprises étrangères), la domination de quelques entreprises sur ces secteurs peut soulever un certain nombre de préoccupations.

Au niveau des puces informatiques, l'Europe est vulnérable à des risques de pénurie ou, à long terme, de restriction des exportations<sup>25</sup>, dans un contexte de tensions sur l'approvisionnement dues à la forte augmentation de la demande. Le rapport récent de l'Autorité de la concurrence sur l'IA générative<sup>26</sup> souligne le risque de pratiques anti-concurrentielles de la part de Nvidia, du fait de sa position prépondérante sur le marché (fixation oligopolistique des prix, restrictions de production, conditions contractuelles déloyales, discrimination en faveur de certains gros clients qui accumulent déjà une grande partie des puces disponibles<sup>27</sup>). Des préoccupations émergent aussi concernant le logiciel de programmation de puces CUDA de Nvidia (seul environnement parfaitement compatible avec les GPU devenus incontournables pour le calcul accéléré) et les investissements récents de Nvidia dans des fournisseurs de services *cloud* spécialisés dans l'IA, tels que CoreWeave<sup>28</sup>.

Concernant les services *cloud*, les acteurs dominants du *cloud computing* proposent des offres aux start-ups leur permettant d'accéder à des ressources de calcul à moindre coût (comme les offres de crédits *cloud*, *i.e.* des offres d'essai octroyant un accès gratuit à un client dans un délai défini). Si ces pratiques sont assorties d'un verrouillage financier ou technique (comme des obstacles à la migration), elles posent un risque concurrentiel et pourraient être considérées comme un abus de position dominante<sup>29</sup>.

Dans ce contexte, et face aux difficultés financières et techniques que rencontrent certains acteurs (au sein de la recherche publique ou privée notamment) pour

accéder au calcul<sup>30</sup>, la France et l'Europe investissent dans le développement d'infrastructures de calcul publiques, comme le supercalculateur Jean-Zay en France et les supercalculateurs Euro-HPC dans l'Union européenne (qui incluront la mise en service d'un nouveau supercalculateur « Jules Verne » sur le territoire français en 2026). Le récent rapport Draghi<sup>31</sup> sur la compétitivité de l'économie européenne encourage le développement de ces supercalculateurs à l'échelle européenne.

Si le développement de ces infrastructures n'a pas vocation à remplacer les acteurs privés, de telles capacités peuvent être utiles pour garantir l'accès au calcul à la recherche publique ou privée. Ainsi, le calculateur Jean Zay est accessible aux acteurs privés qui effectuent des travaux de recherche ouverte impliquant la publication des résultats obtenus. En outre, le développement de ressources de calcul souveraines semble nécessaire pour certains usages sensibles de l'IA (*e.g.* défense, santé) où il existe des risques élevés d'atteintes à la sécurité nationale, au secret industriel ou à la vie privée<sup>32</sup>.

### 1.3 Les viviers de données français et européens pourraient être mieux exploités

L'accès à des jeux de données de bonne qualité et en quantité suffisante est une condition *sine qua non* pour le développement des modèles d'IA, aussi bien pour les modèles de fondation que pour leurs applications. Les modèles de fondation sont avant tout entraînés à partir de grandes bases de données généralistes, contenant du texte, des images ou de la vidéo. La spécialisation d'un modèle pré-entraîné pour une application spécifique, que ce soit *via* la technique du *fine-tuning* ou celle du « *retrieval augmented generation* » (RAG)<sup>33</sup>, nécessite quant à elle des données sectorielles plus spécifiques. Dans les deux cas, le nettoyage et le traitement des données constituent un facteur différenciant, les acteurs devant les filtrer pour ne garder que le contenu de qualité.

(25) De telles restrictions sont actuellement mises en place aux États-Unis pour la vente de puces à destination de la Chine.

(26) Autorité de la concurrence (2024), Avis 24-A-05 relatif au secteur de l'intelligence artificielle générative.

(27) Sherwood (2024), "Just four companies are hoarding tens of billions of dollars worth of Nvidia GPU chips".

(28) Autorité de la concurrence (2024), *op. cit.*

(29) Autorité de la concurrence (2024), *op. cit.*

(30) Comme souligné par le récent rapport Draghi (*cf.* note de bas de page 32).

(31) Draghi M. (2024), "The future of European competitiveness - A competitiveness strategy for Europe".

(32) La Lettre (2024), *Le ministère des armées finalise le lancement de son cloud privé* & Rapport IGAS (2023), *Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé*.

(33) Dans le cadre du *fine tuning*, une partie du modèle (dont les paramètres) est réentraînée sur un nouveau jeu de données représentatif du contexte visé. À l'inverse, le RAG consiste à coupler le modèle avec un moteur de recherche dans des bases de connaissances externes, ce qui permet d'obtenir des résultats plus pertinents sans réentraîner les paramètres du modèle.

Si certaines grandes bases de données sont accessibles publiquement et gratuitement (par exemple, l'organisation Common Crawl met à disposition des archives Web depuis 2008), les acteurs disposant de données propres sont avantagés dans le développement de modèles de fondation ou d'applications. C'est le cas des grandes entreprises, telles que Google, Apple, Facebook, Amazon ou Microsoft (les GAFAM), qui disposent à la fois d'un accès privilégié à de larges volumes de données *via* les contenus qu'ils hébergent (YouTube offre par exemple à Google une source majeure de données d'entraînement pour les modèles d'IA) ou *via* les données associées à l'utilisation de leurs services. L'avantage qu'ils en retirent est d'autant plus important que la quantité globale de données gratuites, accessibles et de qualité semble avoir atteint un pic, dans un contexte où les modèles exigent toujours plus de données pour leur apprentissage<sup>34</sup>.

Dans ce cadre, les GAFAM, mais aussi la plupart des réseaux sociaux ou des groupes de presse, tendent à verrouiller l'accès aux données qu'ils publient : YouTube interdit par exemple l'utilisation de ses vidéos par des applications tierces. De même, certains acteurs<sup>35</sup> utilisent diverses stratégies pour limiter l'accès des tiers aux données de leurs utilisateurs en faisant un usage parfois abusif de règles juridiques comme la protection des données personnelles, ou encore de préoccupations de sécurité. Les grandes entreprises du numérique peuvent enfin utiliser leur puissance financière pour conclure des accords avec des propriétaires de données tierces leur garantissant l'exclusivité des données, ce qui est susceptible de constituer des pratiques anti-concurrentielles.

Face à cet avantage, plusieurs initiatives publiques visent à faciliter l'accès large des entreprises aux données, qu'elles soient générales ou sectorielles, et à les rendre exploitables (harmonisation des nomenclatures entre différentes sources, *etc.*). Concernant les données issues des entités publiques ou parapubliques, la France est à la pointe en matière d'*open data* au niveau européen et mondial<sup>36</sup>. La plateforme [data.gouv.fr](https://data.gouv.fr) met ainsi à disposition plus de 40 000 jeux de données, concernant par exemple les transactions immobilières (base « Demande de valeurs foncières »), les cadastres, les remboursements par

l'assurance-maladie, la biodiversité, la pollution, *etc.* Concernant les données sectorielles, utilisées davantage pour les applications, la France soutient l'initiative « Espaces européens communs de données » (*Common European Data Spaces*, ou CEDS) qui vise à mettre à disposition des bases de données intégrées dans certains secteurs (*e.g.* santé, mobilité, énergie), tout en garantissant la protection de ces données<sup>37</sup>. Des initiatives similaires ont par ailleurs été portées dans le cadre de l'appel à projets « Communs numériques pour l'IA générative » de France 2030. Dans le domaine de la santé, plus spécifiquement, la France pourrait, en complément de sa participation au CEDS, s'inspirer des mesures mises en place par certains pays comme les États-Unis ou le Royaume-Uni, qui ont développé un cadre qui facilite la réutilisation des données tout en assurant leur confidentialité.

Au-delà de l'accès aux données, garantir la qualité et la diversité des données utilisées pour l'entraînement de l'IA est un enjeu important. L'origine culturelle et la langue des données utilisées déterminent partiellement leur représentation, ce qui peut entraîner une moindre performance des modèles dans la compréhension et le traitement de l'information dans certaines langues. Les données accessibles pour l'entraînement de modèles sont aujourd'hui très majoritairement en anglais, même si le français est relativement bien représenté parmi les autres langues. Sur *Hugging Face*, plateforme de mise à disposition d'outils, modèles et bases de données d'IA, 57 % de données d'entraînement sont en anglais, contre 3 % pour le français (*cf.* Graphique 3). Ceci implique que les modèles généralistes d'IA entraînés sur les données accessibles publiquement sont optimisés pour traiter l'information en anglais et qu'ils peuvent se montrer moins performants pour les langues moins représentées, limitant l'utilité de l'IA pour leurs locuteurs.

Il y a donc un enjeu à disposer de données d'entraînement en langue française pour augmenter le nombre de références francophones au sein des modèles d'IA, et disposer de modèles performants pour des usagers français. C'est dans cet objectif que l'initiative « Villers-Cotterêts », lancée en décembre 2023 par l'INA, la BNF et le CNRS, réunira une vaste quantité de textes en français exploitables par des

(34) Villalobos P. *et al.* (2024), "Will we run out of data? Limits of LLM scaling based on human-generated data".

(35) Autorité de la concurrence (2023), [Publicité sur applications mobiles iOS : le rapporteur général indique avoir notifié un grief au groupe Apple](#).

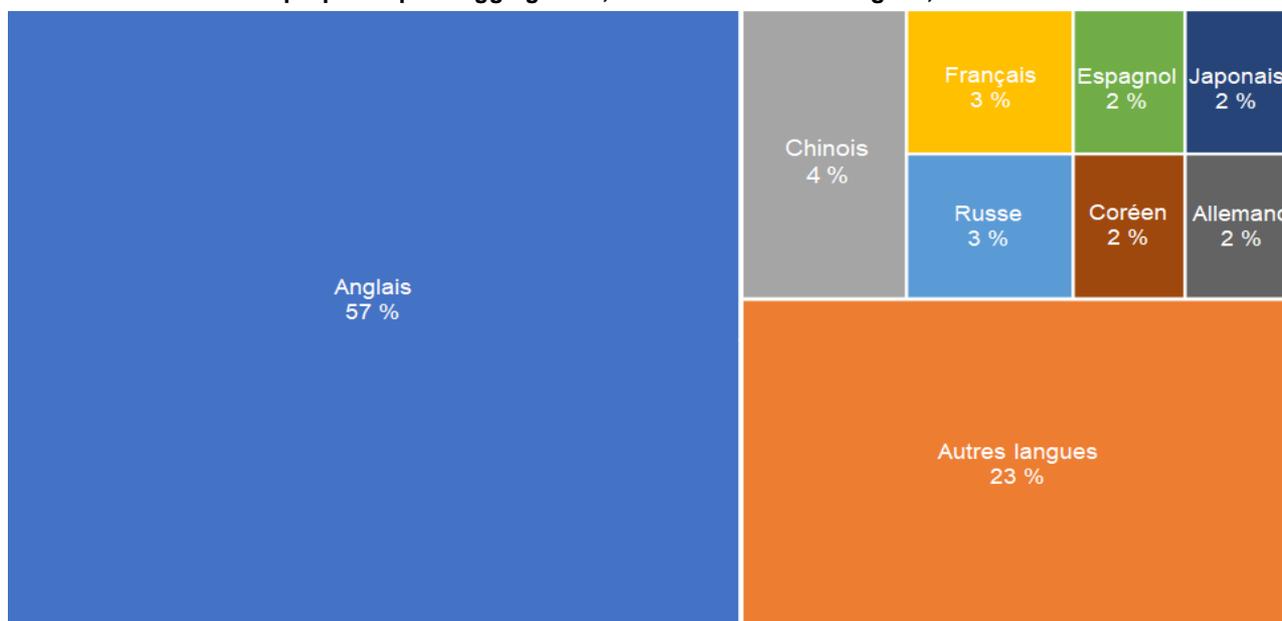
(36) Commission européenne (2023), *Open Data maturity report 2023*. OCDE (2023), *OECD Digital Government Index 2023*.

(37) Le succès de l'initiative CEDS reposera sur la participation et la maturité des entreprises en matière de gestion des données qu'elles détiennent, afin de définir les normes liées aux partages de données intra-sectorielles. À ce niveau, l'Allemagne semble avoir pris de l'avance dans l'implication de ses entreprises.

modèles d'IA. Le lancement prochain d'un consortium européen ALT-EDIC (*Alliance for Language Technologies European Digital Infrastructure*

*Consortium*)<sup>38</sup>, qui permettra d'accroître la disponibilité de données dans diverses langues européennes, s'inscrit aussi dans cette perspective.

**Graphique 3 : Répartition des langues, en pourcentage, dans les ensembles de données d'entraînement d'IA proposés par Hugging Face, sur un total de 225 langues, en 2024**



Source : OCDE (2024), *Perspectives de l'économie numérique de l'OCDE 2024 (Volume 1)*, graphique 2.2.

Note : Ce graphique représente la distribution linguistique dans les ensembles de données d'entraînement d'IA (*training datasets*) proposés par Hugging Face, sur un total de 225 langues, en 2024. Les ensembles de données multilingues et de traduction sur Hugging Face contiennent plus d'une langue et sont donc comptés deux fois. Toutes les langues référencées dans « Autres langues » représentent moins de 2 % du total.

### Encadré 1 : Quel équilibre entre accès aux données d'IA et protection de la propriété intellectuelle ?

L'utilisation de données culturelles pour l'entraînement des modèles de fondation pose des questions importantes de respect du droit d'auteur. D'un point de vue économique, l'utilité et la valeur d'un modèle d'IA reposent sur l'accès et la mobilisation de données nombreuses et de qualité. Dans ce cadre, négocier individuellement avec chaque propriétaire de données ou ayant droit peut constituer un coût important pour les producteurs d'IA, quand bien même le maintien d'une incitation à créer du contenu francophone de qualité est bénéfique à la fois pour les ayants droit et pour les producteurs d'IA.

En Europe, l'utilisation de données protégées pour l'entraînement d'un système d'IA est encadrée par la directive européenne de 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique<sup>a</sup>. Bien que cette directive établisse une exception pour la fouille de textes et de données (dite "*Text and Data Mining*"), utilisés par les fournisseurs de systèmes d'IA, l'ayant droit de données peut exercer un droit d'opposition à leur utilisation (« *opt-out* »). Cette option doit permettre aux ayants droit de recouvrer une capacité de négociation. Ces derniers rencontrent toutefois des difficultés pratiques du fait de l'opacité sur les données d'entraînement utilisées par les producteurs d'IA.

L'*AI Act* européen voté en février 2024 prévoit une obligation de transparence s'agissant des sources d'entraînement des systèmes d'IA. Comme les données constituent un élément de différenciation important pour les fournisseurs de systèmes d'IA, cette transparence est susceptible de nuire à leur développement. La mise en œuvre de l'obligation de transparence de l'*AI Act* devra par conséquent trouver un équilibre entre les coûts pour les fournisseurs de modèles d'IA et la protection des ayants droit.

a. Directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

(38) Un EDIC (European Digital Infrastructure Consortium) est un instrument mis à la disposition des États membres dans le cadre du programme de la décennie numérique à l'horizon 2030 afin d'accélérer et de simplifier la mise en œuvre de projets multinationaux.

## 2. Développement de modèles d'IA : malgré de nouveaux acteurs innovants, la domination préexistante des géants du numérique soulève des enjeux concurrentiels

Le développement des modèles de fondation d'IA est directement ou indirectement dominé par les grandes entreprises du numérique de la Silicon Valley, même si la France dispose d'une start-up (Mistral AI) performante dans le domaine. Les GAFAM, de par leur taille et leur présence sur divers segments de l'économie numérique, sont avantagés par leur accès aux ressources nécessaires (données, puissance de calcul, main d'œuvre spécialisée, financement). Cet accès constitue une barrière à l'entrée et leur permet de se développer dans le secteur de l'IA, notamment en acquérant ou en nouant des partenariats avec des entreprises innovantes. Pour la spécialisation des modèles en aval, les barrières à l'entrée sont moindres, même si les GAFAM bénéficient de leur intégration avec leurs produits existants. La domination d'ensemble des grands acteurs du numérique soulève de nombreux enjeux concurrentiels et économiques : il existe un risque que la diffusion des gains économiques liés à l'IA (valeur ajoutée, profits, productivité) soit atténuée ou biaisée par des pratiques anti-concurrentielles.

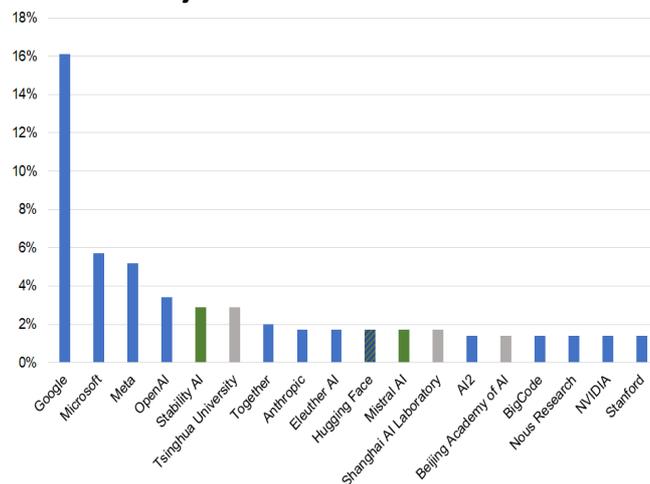
### 2.1 L'émergence de l'IA s'inscrit dans un paysage particulier, marqué par la domination des géants du numérique, qui s'illustre notamment sur les modèles de fondation

Le paysage de l'IA est marqué par le développement rapide de start-ups proposant des modèles très performants, à l'instar d'OpenAI, l'entreprise ayant développé ChatGPT. En France, une start-up (Mistral AI) développe des modèles de fondation parmi les plus performants au monde<sup>39</sup>, après avoir levé plus d'un milliard d'euros lors de trois tours de tables successifs en 2023 et 2024.

Toutefois, le développement des modèles de fondation reste largement dominé par les grandes entreprises américaines du numérique (notamment Google, Microsoft et Meta). Presque un tiers des modèles récents ont été développés par ces trois acteurs (cf. Graphique 4). En considérant la performance des modèles, d'après le *LLM Leaderboard*<sup>40</sup>, Google est à ce jour en avance sur ses concurrents *via* son modèle Gemini, qui devance le modèle GPT-4o d'OpenAI. En outre, Microsoft a noué un partenariat stratégique avec

OpenAI (cf. *infra*), qui conforte le rôle clé joué par les GAFAM sur le segment des modèles de fondation.

**Graphique 4 : Part de différentes organisations dans le nombre total de modèles de fondation entre le 1<sup>er</sup> janvier 2022 et le 4 octobre 2024**



Source : Stanford ecosystem graph for foundation models, à la date du 4 octobre 2024.

Notes : Les organisations américaines sont représentées en bleu, les organisations asiatiques en gris et les organisations européennes en vert. Hugging Face est représentée en vert hachuré de bleu, car il s'agit d'une entreprise franco-américaine, créée originellement à Paris par trois français, mais dont le siège social est aujourd'hui situé à New York, avec une immatriculation dans le Delaware. Au total, 348 modèles sont recensés. Seules les organisations ayant développé 5 modèles ou plus au cours de la période sont présentées sur le graphique. Lorsqu'un modèle est développé conjointement par plusieurs organisations, il est compté une fois pour chaque organisation.

La place prépondérante des GAFAM dans le développement de modèles de fondation s'explique tout d'abord par les ressources dont ils disposent, issues de nombreuses années de développement dans la sphère numérique : ressources financières importantes pour investir, mais aussi capacité à attirer les meilleurs chercheurs (salaires, conditions de travail).

L'intégration verticale et conglomérale des géants du numérique tout au long de la chaîne de valeur des modèles de fondation (cf. Graphique 5) leur confère également un avantage concurrentiel naturel :

- En amont de la conception d'un modèle, ces entreprises bénéficient de leur accès direct à la puissance de calcul nécessaire, dont ils disposent

(39) Chatbot Arena LLM Leaderboard, au 20 novembre 2024.

(40) Chatbot Arena LLM Leaderboard, *op. cit.*

naturellement pour leurs autres activités. Les géants du numérique ont en outre investi dans la conception et la fabrication de certaines puces d'IA (cf. partie 1) et le contrôle de bibliothèques logicielles indispensables pour les développeurs, comme Pytorch (Meta) ou TensorFlow (Google), qui leur confère un pouvoir normatif sur le développement des modèles de fondation<sup>41</sup>. Par ailleurs, ils disposent également d'un accès à de très larges volumes de données, facteur différenciant pour les modèles de fondation.

- En aval, les GAFAM disposent de canaux de diffusion préexistants pour distribuer leurs modèles. D'une part, la vente de l'accès à des modèles est complémentaire à la vente d'accès aux infrastructures de calcul, dont le *cloud* (avec de fortes synergies puisque l'utilisation de modèles requiert des infrastructures de calcul conséquentes). Les GAFAM distribuent ainsi des modèles vendus comme des services d'IA (*Model-as-a-service*, *MaaS*) dans des places de marché adossées à leur services *cloud* (e.g. Model Garden adossée à Google Cloud, ou Amazon Bedrock adossée à AWS). D'autre part, les modèles de fondation peuvent directement être intégrés à des produits et

services existants (e.g. suites logicielles, moteurs de recherche, réseaux sociaux) disposant d'une large base d'utilisateurs et générant des effets d'échelle et de réseau.

*In fine*, l'IA a la particularité d'être une révolution technologique où des acteurs préexistants jouent d'emblée un rôle prépondérant<sup>42</sup>, alors que la dynamique d'innovation est généralement portée par des nouveaux entrants qui viennent contester la domination d'acteurs historiques ou créer de nouveaux marchés<sup>43</sup>. À court terme, l'investissement d'acteurs établis dans l'IA est positif pour le développement de cette technologie et permet de proposer rapidement des fonctionnalités innovantes. À moyen terme, cette situation peut limiter le développement et la diffusion de l'IA, par rapport à de nouveaux entrants potentiellement plus innovants. L'émergence des grands modèles de langage (LLM) illustre ce paradoxe : les modèles d'IA des GAFAM (Gemini, Llama, etc.) sont très performants et se développent rapidement. Toutefois, la diffusion de ces modèles pour le grand public et les entreprises a été fortement accélérée par les nouveaux entrants qui contestent le marché (e.g. lancement de ChatGPT par OpenAI en novembre 2022)<sup>44</sup>.

**Graphique 5 : Position des grands acteurs du numérique sur la chaîne de valeur de l'IA**



Source : Autorité de la concurrence britannique, *AI foundation models: Update report*.

Sur une dynamique de plus long terme, l'existence de données en accès libre et le développement de modèles de fondation « frugaux », qui utilisent moins

de ressources (e.g. puissance de calcul et données), pourraient également permettre de réduire certaines barrières à l'entrée identifiées actuellement, et ainsi

(41) Un modèle de fondation est plus performant s'il bénéficie de très grandes quantités de données et de puissance de calcul (effets d'échelle) ; sa valeur augmente s'il est utilisé par d'autres utilisateurs (effets de réseau), car leurs requêtes permettront d'améliorer le fonctionnement du modèle.

(42) Le Monde (2024), « L'IA est la première technologie à être d'emblée dominée par des grands acteurs ». Entretien avec Benoît Coeuré, président de l'Autorité de la concurrence.

(43) Aghion *et al.* (2005), "Competition and Innovation: an Inverted-U Relationship".

(44) Cf. Synthèse du Séminaire Nasse, « Intelligence artificielle : enjeux concurrentiels, enjeux économiques et applications pratiques dans le secteur des médias », organisé par la DG Trésor et l'Autorité de la concurrence, 19 septembre 2024.

favoriser l'émergence de nouveaux acteurs. Le développement de modèles frugaux apparaît d'ailleurs d'autant plus nécessaire que la consommation en énergie des services d'IA augmente exponentiellement ces dernières années : certains fournisseurs d'IA contractent d'ailleurs des accords avec des fournisseurs d'énergie décarbonée, notamment nucléaire<sup>45</sup>.

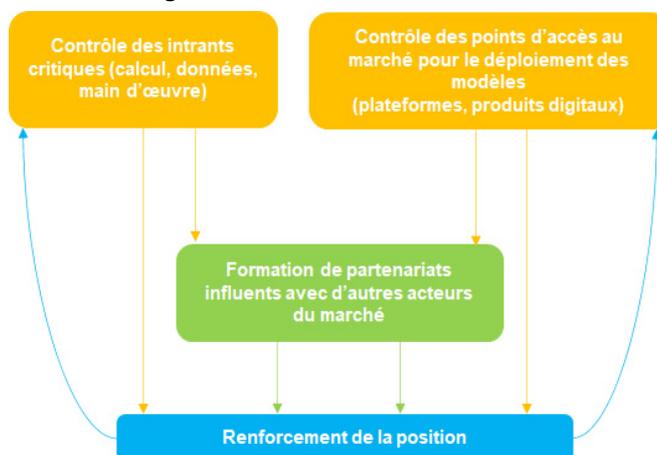
## 2.2 La domination des géants du numérique peut engendrer des risques pour la structuration du marché et le développement de l'IA

Les régulateurs, et notamment les autorités de concurrence (aussi bien en Europe qu'aux États-Unis<sup>46</sup>), surveillent le développement de l'IA dans ce contexte particulier de domination des GAFAM, qui pourrait consolider davantage la structure oligopolistique des marchés numériques. L'Autorité de la concurrence française recommande notamment de prêter une attention particulière aux stratégies qui peuvent être mises en place par les grands acteurs du numérique pour consolider leur pouvoir de marché ou tirer parti de celui-ci pour se développer dans le secteur de l'IA (cf. Graphique 6). Cette vigilance doit s'appliquer aux éventuelles pratiques anti-concurrentielles liées aux intrants ou aux canaux de distribution<sup>47</sup>, mais aussi à celles qui pourraient passer par l'acquisition d'entreprises innovantes spécialisées dans l'IA<sup>48</sup>, ou par la conclusion de partenariats influents avec des acteurs émergents du développement de modèles.

Dans certains partenariats, les grandes entreprises du numérique font en effet bénéficier leurs partenaires de leurs ressources (financières<sup>49</sup>, mais aussi en termes de données ou de calcul) en échange d'un accès privilégié aux modèles développés par les start-ups. Cela permet ensuite à ces grandes entreprises de distribuer ces modèles *via* leurs plateformes ou de les

intégrer dans leurs produits. C'est notamment la logique qui prévaut dans le cadre du partenariat entre OpenAI et Microsoft – Microsoft ayant investi au total plus de 13 milliards de dollars dans l'entreprise développant les modèles GPT. Comme signalé par les autorités de concurrence<sup>50</sup>, un d'accord de ce type présente des risques concurrentiels significatifs : il pourrait en effet affaiblir l'intensité concurrentielle entre les deux entités, entraîner des effets verticaux ou un verrouillage de certains acteurs. Il convient toutefois de noter que ce type d'accord peut aussi permettre à des start-ups de se développer, et plus en amont créer des incitations à innover dans la perspective d'une revente ou d'un partenariat futur.

**Graphique 6 : Dynamiques concurrentielles sur le segment des modèles de fondation**



Source : Autorité de la concurrence britannique, *AI foundation models, update report*.

Les grandes entreprises du numérique peuvent également, plus simplement, embaucher le personnel de start-ups, sans acheter l'entreprise elle-même<sup>51</sup>. Une telle approche est possible sur un segment où le capital humain représente le principal actif. Elle peut aussi constituer un moyen pour ces entreprises de

(45) Le Monde (2024), Les géants du numérique se convertissent au nucléaire pour étancher les besoins énergétiques toujours plus importants de l'IA.

(46) Aux États-Unis, la *Federal Trade Commission* (FTC) a notamment ouvert une enquête concurrentielle sur Microsoft et Open AI, tandis que le *Département de la Justice* (DOJ) a fait de même sur Nvidia. Au Royaume-Uni, la *Competition and Markets Authority* (CMA) a rendu un rapport en septembre 2023 sur les modèles de fondation, et a ouvert une enquête sur le partenariat Microsoft-OpenAI. En France, l'Autorité de la concurrence a publié en juin 2024 un avis sur le secteur de l'intelligence artificielle (mentionné à plusieurs reprises dans cette étude). En octobre 2024, à Rome, le sommet réunissant les représentants des autorités de la concurrence et les décideurs politiques du G7 a également souligné l'importance des enjeux concurrentiels liés à l'IA.

(47) Cf. Partie 1 pour la puissance de calcul et les données. Cf. *infra* pour l'accès à une main d'œuvre qualifiée.

(48) CB Insights (2021), *The Race For AI: Which Tech Giants Are Snapping Up Artificial Intelligence Startups*.

(49) Certaines start-ups ont d'ailleurs récemment pu demander à leurs investisseurs une exclusivité en matière de financement, sans que l'on sache à ce jour s'ils l'ont obtenue. Voir Financial Times (2024), "OpenAI asks investors not to back rival start-ups such as Elon Musk's xAI".

(50) Autorité de la concurrence (2024), *op.cit.* *Competition Markets Authority* (2021), *op. cit.* Les deux autorités soulignent dans leurs rapports les risques potentiels induits par le partenariat entre Microsoft et OpenAI.

(51) Microsoft a embauché des employés de Inflection, Google des employés de Character.AI, et Microsoft a proposé d'embaucher Sam Altman et d'autres employés d'OpenAI dans le cadre d'un différend entre Altman et le comité de gouvernance d'OpenAI. Ces pratiques potentiellement anti-concurrentielles ne semblent toutefois pas soulever de préoccupations particulièrement fortes parmi les acteurs du secteur à ce stade (source : Autorité de la concurrence, *op. cit.*).

maintenir leur position dominante, mais les règles de concurrence sont alors plus difficiles à appliquer<sup>52</sup>.

La position dominante des grandes entreprises du numérique pèse aussi sur les modèles d'affaires et l'organisation de la chaîne de valeur. À moyen terme, le marché des modèles d'IA pourrait se « plateformiser », c'est-à-dire se concentrer autour de quelques plateformes regroupant des écosystèmes d'outils facilitant le développement et l'utilisation des modèles de fondation<sup>53</sup>. De fortes incertitudes subsistent toutefois encore sur les modèles d'affaires qui pourraient être déployés. Des plateformes telles qu'Amazon Bedrock, Hugging Face ou Google Cloud mutualisent déjà l'accès à de multiples ressources et outils (e.g. modèles, bases de données, bibliothèques) pour les développeurs de modèles de fondation et de leurs applications. Elles bénéficient d'effets de réseaux de par leur rôle d'intermédiaire entre les développeurs et les entreprises utilisatrices de ces applications. De façon analogue aux systèmes d'exploitation et magasins d'applications sur smartphones, seules quelques plateformes, celles qui fourniront l'écosystème le plus complet, pourraient survivre à moyen terme. Un certain nombre de pratiques potentiellement anticoncurrentielles liées à ces plateformes (par exemple liées au fait que les entreprises les détenant puissent y favoriser leurs propres modèles) sont par ailleurs déjà identifiées par les autorités de concurrence. Pour les prévenir, l'Autorité de la concurrence recommande<sup>54</sup> que les modèles utilisés comme des services d'IA (*Model-as-a-service*, MaaS) soient inclus dans les services de plateformes régulés par le règlement sur les marchés numériques (*Digital markets act*, DMA)<sup>55</sup>.

La concentration des modèles de fondation autour des grandes entreprises du numérique pose par ailleurs le risque d'une diffusion limitée des gains de productivité liés à l'IA<sup>56</sup>. En effet, si à moyen terme un fournisseur est incité à proposer son modèle en accès libre car les innovations autour d'un modèle ouvert lui bénéficient et participent à la construction d'un écosystème autour de

son architecture<sup>57</sup>, à plus long terme, une fois ses positions consolidées, il est incité à fermer l'écosystème et à le rentabiliser. Les gains de productivité tirés de l'utilisation du système pourraient alors être capturés sous forme de rentes par les fournisseurs de modèles, avec une diffusion plus limitée au sein de l'économie et des coûts plus élevés pour les utilisateurs. Dès lors, la mise en œuvre d'une politique de la concurrence proactive est indispensable pour s'assurer à la fois de l'innovation dans le secteur de l'IA et de sa diffusion à l'ensemble de l'économie.

La question de la domination des grandes entreprises du numérique sur la chaîne de valeur doit également être appréhendée à l'aune des problématiques plus larges sur la compétitivité de l'économie européenne. Le récent rapport Draghi sur le sujet<sup>58</sup> montre en effet que l'écart de productivité croissant entre les États-Unis et l'Europe s'explique avant tout par le numérique : l'Europe est surtout positionnée sur des technologies déjà matures, tandis que les États-Unis le sont sur des nouvelles technologies innovantes, en particulier les technologies numériques, qui permettent une croissance économique plus importante<sup>59</sup>.

Enfin, le fait que les leaders mondiaux actuels soient des acteurs extra-européens soulève des questions quant à la localisation et à la taxation des profits et des salaires liés à l'IA. Dans un contexte de sous-imposition structurelle des multinationales, la présence de leaders français ou européens contribuerait à renforcer notre souveraineté fiscale, assurant ainsi une imposition plus équitable et adaptée des revenus générés. Les réformes en cours de négociation à l'OCDE concernant la fiscalité des multinationales devraient contribuer à apporter des solutions à cette problématique, sachant qu'une harmonisation de la fiscalité doit être appréhendée à l'échelle la plus large possible (*a minima* européenne), afin de garantir son efficacité. À ce titre, des réformes visant à harmoniser l'environnement fiscal des entreprises sont également en cours de négociation au niveau européen.

(52) D'autant plus que la mise en place de clauses de non débauchage ("*non poaching*" ou "*non compete*"), qui permettrait d'empêcher l'embauche massive de personnels, peut aussi sous certains aspects être considérée comme anti-concurrentielle.

(53) Competition & Markets Authority (2023), *AI Foundation Models: Initial Report*.

(54) Autorité de la concurrence (2024), *op. cit.* Recommandation n°1.

(55) L'auto-préférence (*self-preferencing*) a été classée à l'article 6 du DMA, qui regroupe les pratiques des contrôleurs d'accès pouvant être interdites, au cas par cas, après examen.

(56) Besson L. *et al.* (2024), « *Les enjeux économiques de l'intelligence artificielle* », *Trésor-Éco*, n° 341.

(57) Meta (Facebook) a ouvert en accès libre son modèle Llama 2 à la mi-juillet 2023.

(58) Draghi M. (2024), "The future of European competitiveness": "The key driver of the rising productivity gap between the EU and the US has been digital technology ("tech")", partie A, chapitre 2.

(59) Fuest C., Gros D., Mengel P.L., Presidente G., Tirole J. (2024), "EU Innovation Policy: How to Escape the Middle Technology Trap", *EconPol Policy Reports*.

## 2.3 Le segment des modèles spécialisés présente moins d'obstacles « naturels » au bon fonctionnement concurrentiel du marché

Les éléments structurels pouvant expliquer le caractère déjà fortement oligopolistique de la chaîne de valeur des modèles de fondation s'appliquent moins au segment des modèles spécialisés. En effet, sous réserve que les principaux modèles de fondation restent accessibles (*i.e.* en l'absence de stratégies de verrouillage de la part des principaux acteurs détenant les modèles de fondation), les applications à partir de ces modèles devraient rester peu coûteuses à développer. En effet, la spécialisation d'un modèle pré-entraîné pour une application spécifique nécessite moins de données et de puissance de calcul que l'entraînement d'un modèle de fondation.

Le principal obstacle à l'entrée de nouveaux acteurs sur ce segment pourrait être les supports de vente, où les GAFAM bénéficient de leur intégration avec leurs produits existants comme les suites bureautiques ou les réseaux sociaux). Cet accès privilégié à des canaux de distribution essentiels crée *de facto* des barrières à l'entrée pour le déploiement de modèles spécialisés. Le fait que les grandes entreprises du numérique puissent favoriser leurs propres modèles au sein de produits qui représentent une grande partie du marché potentiel fait

par ailleurs peser le risque d'abus de position dominante.

La France dispose de nombreuses start-ups<sup>60</sup> et grandes entreprises<sup>61</sup> qui investissent dans les nouveaux cas d'usage spécialisés de l'IA. Un des facteurs principaux de différenciation sur ce marché sera l'utilisation de données sectorielles de qualité, adaptées à des besoins spécifiques. Dans ce cadre, la France peut capitaliser sur ses entreprises industrielles, leurs données et leurs cas d'usage pour stimuler le développement d'applications, d'autant plus que le faible coût relatif<sup>62</sup> et la qualité de la main d'œuvre qualifiée française en matière d'IA (*cf.* 1.1) constituent des atouts vis-à-vis de ses partenaires. Ces grandes entreprises peuvent également jouer un rôle pertinent dans le cadre du partage de données intra-filière avec externalités positives.

Il convient toutefois de souligner qu'indépendamment de la nationalité des entreprises développant ces modèles spécialisés, la question de leur adoption au sein du tissu économique, notamment parmi les PME, reste un enjeu majeur. L'usage d'outils d'IA performants par les entreprises représente en effet un gisement de croissance potentiellement plus important que les gains directs liés au développement d'entreprises françaises de l'IA<sup>63</sup>.

(60) France Digitale (2024), Mapping 2023 des start-ups françaises de l'IA.

(61) L'Usine Nouvelle (2024), *L'IA générative dans l'industrie française en 15 cas d'usage*.

(62) Selon Glassdoor, le salaire annuel moyen d'un ingénieur Machine Learning débutant est de 55 k€ bruts en France contre 150 k\$ aux États-Unis. Ce niveau de rémunérations peut toutefois également constituer un obstacle pour attirer les meilleurs talents.

(63) Besson L. *et al.* (2024), *op. cit.*

### Éditeur :

Ministère de l'Économie,  
des Finances  
et de l'Industrie  
Direction générale du Trésor  
139, rue de Bercy  
75575 Paris CEDEX 12

### Directeur de la Publication :

Dorothee Rouzet  
tresor-eco@dgtresor.gouv.fr

### Mise en page :

Maryse Dos Santos  
ISSN 1777-8050  
eISSN 2417-9620

## Derniers numéros parus

### Novembre 2024

N° 353 La finance climat publique française vers les pays en développement : état des lieux et enjeux

Nicolas Krakovitch, Pierre Marc, Victoria Seignez

N° 352 Quel risque de « japonisation » de l'économie chinoise ?

Louis Bertrand, Thomas Carré, Patrick Kanda Tunda, Eloïse Villani

### Octobre 2024

N° 351 Les minerais dans la transition énergétique

Alban Aubert, Sary Zoghely, Xavier Le Guennec

<https://www.tresor.economie.gouv.fr/Articles/tags/Tresor-Eco>



Direction générale du Trésor



@DGTresor

Pour s'abonner à *Trésor-Éco* : [bit.ly/Trésor-Eco](https://bit.ly/Trésor-Eco)

*Ce document a été élaboré sous la responsabilité de la direction générale du Trésor et ne reflète pas nécessairement la position du ministère de l'Économie, des Finances et de l'Industrie.*