



AVRIL
2025

Startups européennes et IA générative dépasser la domination des *Big Tech*

Hugo LE PICARD



L’Ifri est, en France, le principal centre indépendant de recherche, d’information et de débat sur les grandes questions internationales. Créé en 1979 par Thierry de Montbrial, l’Ifri est une fondation reconnue d’utilité publique par décret du 16 novembre 2022. Elle n’est soumise à aucune tutelle administrative, définit librement ses activités et publie régulièrement ses travaux.

L’Ifri associe, au travers de ses études et de ses débats, dans une démarche interdisciplinaire, décideurs politiques et experts à l’échelle internationale.

Les opinions exprimées dans ce texte n’engagent que la responsabilité de l’auteur.

ISBN : 979-10-373-1030-9

© Tous droits réservés, Ifri, 2025

Couverture : © Image générée à l’aide de FLUX PRO 1.1,
développé par Black Forest Labs

Comment citer cette publication :

Hugo Le Picard, « Startups européennes et IA générative : dépasser la domination des *Big Tech* », *Études de l’Ifri*, Ifri, avril 2025.

Ifri

27 rue de la Procession 75740 Paris Cedex 15 – FRANCE

Tél. : +33 (0)1 40 61 60 00 – Fax : +33 (0)1 40 61 60 60

E-mail : accueil@ifri.org

Site internet : ifri.org

Auteur

Hugo Le Picard est chercheur associé au Centre géopolitique des technologies de l'Ifri et cofondateur de Clone, une start-up spécialisée dans la génération d'images. Il était auparavant chercheur au Centre Énergie et Climat de l'Ifri, qu'il a rejoint en 2018. Son travail se concentrait sur l'accès à l'électricité et le développement d'infrastructures électriques en Afrique subsaharienne, à l'aide de méthodes d'analyse d'images satellite et de *deep learning*. Il a enseigné à l'université Paris-Dauphine, où il donnait des cours d'analyse d'images, d'introduction au *deep learning* et de microéconomie. Il a également contribué à des cours de master à Sciences Po Paris et est titulaire d'un doctorat en économie de l'université PSL.

Résumé

L'Europe est à la croisée des chemins : face à la domination des *Big Tech* américaines sur toute la chaîne de valeur de l'IA générative, des modèles de fondation aux infrastructures *cloud*, en passant par les canaux de distribution et l'*open source*, elle risque un décrochage technologique et économique durable. Néanmoins, l'IA générative pourrait aussi représenter une opportunité de transformation économique majeure, avec un potentiel de valeur estimé à 1,5 fois le PIB français. Mais pour en faire un moteur de renouveau économique, l'Europe doit dépasser sa quête illusoire d'indépendance technologique totale et construire un écosystème capable de tirer parti des ressources des *Big Tech* tout en renforçant ses propres capacités d'innovation.

La véritable opportunité pour l'Europe se situe aujourd'hui plus en aval de la chaîne de valeur de l'IA générative, là où s'opère l'adaptation des modèles aux besoins concrets des différents secteurs, notamment industriels. Forte de son expertise industrielle, réglementaire et technique, l'Europe peut y construire des avantages compétitifs durables. Plutôt que de viser une souveraineté totale en amont, l'enjeu est d'accélérer l'adoption sectorielle de l'IA, en s'appuyant sur ses filières d'excellence et en soutenant le développement d'un tissu de startups capables de transformer la technologie en solutions concrètes.

En amont de la filière, la compétition féroce entre les géants technologiques américains crée paradoxalement des effets d'aubaine pour l'Europe. Les géants technologiques américains se livrent une compétition de plus en plus intense, dont la course aux investissements en est l'un des symptômes les plus visibles. Pris dans un dilemme où aucun ne peut se permettre de ralentir, malgré les risques de surcapacité, ils multiplient les dépenses pour garder leur avance. Cette dynamique concurrentielle force les *Big Tech* à proposer des conditions d'accès plus favorables à leurs outils et infrastructures. Les entreprises européennes peuvent en tirer parti pour accélérer le développement de solutions innovantes, à des coûts réduits, et viser des marchés comme les États-Unis pour les monétiser rapidement.

Abstract

Europe is at a crossroads. Faced with the domination of American *Big Tech* across the entire generative AI value chain, from foundation models to cloud infrastructure, distribution channels, and open source, it risks long-term technological and economic decline. Yet generative AI also represents a major opportunity for economic transformation, with a potential value estimated at 1.5 times France's GDP. To turn it into a driver of renewal, Europe must move beyond the illusion of total technological independence and instead build an ecosystem that leverages *Big Tech* resources while strengthening its own innovation capabilities.

The real opportunity for Europe lies further downstream in the generative AI value chain, where models are adapted to the concrete needs of various sectors, particularly industrial ones. Thanks to its industrial, regulatory, and technical expertise, Europe can carve out lasting competitive advantages here. Rather than chasing an upstream sovereignty, the real challenge is to accelerate AI adoption across sectors by capitalizing on its areas of excellence and nurturing a startup ecosystem that turns cutting-edge technology into practical solutions.

Upstream, the fierce competition between American tech giants paradoxically creates windfall opportunities for Europe. Locked in an investment race, none can afford to slow down despite the risks of overcapacity, pushing them to multiply spending in order to maintain their lead. This dynamic compels *Big Tech* to offer increasingly favorable access to their tools and infrastructure. European companies can seize this moment to develop innovative solutions at lower costs and monetize them quickly by targeting large markets like the United States.

Sommaire

INTRODUCTION	6
LE CONTROLE STRUCTUREL DES <i>BIG TECH</i> SUR L'IA EUROPÉENNE.....	9
Une course technologique dominée par les <i>Big Tech</i>	9
La maîtrise des infrastructures critiques.....	11
Le contrôle des canaux de distribution.....	13
L'influence sur l'écosystème <i>open source</i>	16
Au-delà du mythe de l'indépendance technologique européenne	17
LA CHAÎNE DE VALEUR DE L'IA : IDENTIFIER LES OPPORTUNITÉS POUR L'EUROPE	19
La diffusion tout autant que la production : repenser l'ambition européenne	19
Les outils de développement : un segment stratégique.....	20
La verticalisation : le principal atout européen ?	23
La miniaturisation des modèles : une tendance alignée avec la spécialisation sectorielle	24
UNE NOUVELLE DONNE POUR L'INDUSTRIE TECHNOLOGIQUE.....	26
La fragilité paradoxale des géants du numérique	26
L'intensification de la compétition entre <i>Big Tech</i> et avec la Chine ...	27
Un modèle économique en transformation	28
Le levier du marché américain	30
CONCLUSION	32

Introduction

Face à un risque de décrochage économique historique vis-à-vis des États-Unis, l'intelligence artificielle générative¹ pourrait représenter pour l'Europe soit le levier de son renouveau, soit l'accélérateur de son déclin. Cette technologie représente un potentiel de création de valeur annuel considérable, qui équivaldrait à plus d'une fois et demi le produit intérieur brut (PIB) de la France². Tout retard dans la maîtrise et, surtout, l'adoption de ces technologies pour des usages économiques pourrait rendre irréversible l'écart de développement économique entre les deux continents.

Deux ans après la sortie de ChatGPT, les transformations économiques associées à l'IA générative s'accroissent, soutenues par une dynamique de réduction des coûts sans précédent dans l'histoire des technologies³. En moins de deux ans, l'IA générative s'est intégrée dans le quotidien de dizaines de millions d'utilisateurs, tant dans leurs usages personnels que professionnels⁴. Cette technologie disruptive⁵ a déjà bouleversé certains secteurs établis. Dans l'éducation par exemple, la plateforme éducative en ligne Chegg a perdu 99 % de sa valeur boursière après le lancement de ChatGPT – soit 14,5 milliards de dollars évaporés en quelques mois⁶ –, du fait de la migration massive des étudiants vers l'agent conversationnel d'OpenAI.

À la différence des innovations passées, l'IA générative se distingue par sa polyvalence extraordinaire : elle peut assister et augmenter simultanément la productivité de profils aussi différents que des biologistes, des ingénieurs, des radiologues ou des développeurs, tout en

1. L'IA générative se distingue par sa capacité à créer de nouveaux contenus (textes, images, sons, code) à partir des données sur lesquelles elle a été entraînée, là où l'IA traditionnelle se concentre principalement sur l'analyse, la classification et la prédiction à partir de données existantes. Cette capacité de création lui permet non seulement d'analyser mais aussi de générer du contenu original, d'enrichir et de transformer les *inputs* qu'elle reçoit, ouvrant ainsi un champ d'applications beaucoup plus vaste que les systèmes d'IA classiques.

2. J. Manyika et M. Spence, « The Coming AI Economic Revolution: Can Artificial Intelligence Reverse the Productivity Slowdown? », *Foreign Affairs*, novembre/décembre 2023.

3. Les coûts de l'IA sont divisés par dix en moyenne chaque année ; cf. G. Appenzeller, « Welcome to LLMflation – LLM Inference Cost is Going Down Fast », Andreessen Horowitz (a16z), 12 novembre 2024.

4. D. Gallegos, « The Many Ways WSJ Readers Use AI in Their Everyday Lives », *The Wall Street Journal*, 23 novembre 2024.

5. B. Pajot, « Les risques de l'IA. Enjeux discursifs d'une technologie stratégique », *Études de l'Ifri*, Ifri, juin 2024.

6. M. Kruppa, « How ChatGPT Brought Down an Online Education Giant », *The Wall Street Journal*, 9 novembre 2024.

s'adaptant à des usages quotidiens variés, de la création d'images à l'assistance conversationnelle ou encore à la compagnie virtuelle. Cette polyvalence est renforcée par la capacité des modèles de pointe à gérer simultanément plusieurs modalités (texte, code, audio, image)⁷ et se nourrit d'un volume de connaissances colossal : selon Yann Le Cun⁸, il faudrait plus de 22 000 ans de lecture à un humain pour acquérir l'équivalent du « savoir » de GPT-3.

L'IA générative est aujourd'hui perçue comme ce qui pourrait être un moteur majeur de la croissance économique du XXI^e siècle, en résonance avec les théories de la croissance endogène où l'innovation et le capital humain jouent un rôle fondamental. Sa capacité à accélérer simultanément le progrès technique dans de nombreux secteurs transformerait l'équilibre des forces économiques : la puissance d'un pays dépendrait moins de ses ressources ou d'une expertise technologique isolée que de son aptitude à entretenir une innovation constante dans tous les domaines et à les intégrer rapidement dans l'ensemble de sa société, de son économie et de sa sécurité nationale⁹.

Pour l'Europe, cette rupture technologique se produit dans un contexte de risque de déclassement économique durable vis-à-vis des États-Unis, caractérisé par une stagnation de la croissance depuis 2008, un ralentissement de la productivité depuis les années 1990¹⁰ et un vieillissement démographique accru¹¹. Alors que le PIB par habitant en Europe est déjà deux fois inférieur à celui des États-Unis¹², tout retard dans la diffusion de l'IA générative ne peut qu'accentuer ce déclassement.

Face à cet enjeu crucial, l'Europe dispose d'atouts pour tirer parti de cette nouvelle donne technologique. Mais le développement de l'IA générative européenne fait face à un paradoxe : malgré un vivier de talents et l'émergence de startups innovantes comme Mistral AI, il reste profondément dépendant des géants technologiques américains. Cette relation complexe appelle une analyse approfondie, pour définir une stratégie de développement réaliste de l'IA générative en Europe.

Cette note s'articule autour de trois axes complémentaires. Elle examine d'abord la nature et l'étendue de la domination des *Big Tech*¹³ sur

7. J. Manyika et M. Spence, « The Coming AI Economic Revolution: Can Artificial Intelligence Reverse the Productivity Slowdown? », *Foreign Affairs*, novembre/décembre 2023.

8. D. Bavezez, *Bienvenue en économie de guerre*, Paris, Novice, 2024.

9. E. Schmidt, « Innovation Power: Why Technology Will Define the Future of Geopolitics », *Foreign Affairs*, mars/avril 2023.

10. J. Manyika et M. Spence, « The Coming AI Economic Revolution », *Foreign Affairs*, *op. cit.*

11. N. Eberstadt, « The Age of Depopulation: Surviving a World Gone Gray », *Foreign Affairs*, novembre/décembre 2024.

12. « World Development Indicators - GDP per capita (current US\$) », Banque mondiale, données pour 2023, consulté le 22 février 2025.

13. Dans cette étude, les termes « *Big Tech* » et « géants technologiques » sont utilisés indistinctement pour désigner les principales entreprises technologiques américaines qui dominent le secteur : Alphabet (Google), Amazon, Apple, Meta (Facebook), Microsoft et Nvidia.

l'écosystème européen de l'IA, révélant une dépendance structurelle. Elle identifie ensuite les espaces d'opportunité dans la chaîne de valeur émergente de l'IA générative, particulièrement en son aval, où les entreprises européennes peuvent bâtir des avantages compétitifs durables. Enfin, elle analyse comment l'intensification de la concurrence entre les géants américains, et plus récemment chinois, en matière d'IA pourrait être porteuse de nouvelles opportunités pour le Vieux Continent.

Cette progression, d'un constat initial lucide à l'exploration d'axes concrets, vise à dépasser toute vision binaire de domination ou d'indépendance technologique et à esquisser une voie européenne réaliste dans l'IA générative.

Le contrôle structurel des *Big Tech* sur l'IA européenne

La position dominante des *Big Tech* dans l'écosystème de l'IA générative s'étend sur l'ensemble de la chaîne de valeur. Cette première partie analyse quatre pans essentiels où s'exerce cette domination : les modèles de fondation¹⁴ qui constituent le cœur des technologies, l'infrastructure *cloud* nécessaire à leur développement, les canaux de distribution qui permettent leur commercialisation, et les outils *open source* qu'elles soutiennent et influencent stratégiquement pour faciliter le développement et le déploiement de leurs technologies. Cette mainmise structure l'ensemble de l'écosystème technologique¹⁵ européen et explique l'insuffisance d'une simple régulation pour créer les conditions d'une innovation européenne autonome.

Une course technologique dominée par les *Big Tech*

Le développement des modèles de fondation illustre la domination croissante des *Big Tech* dans la course à l'IA générative. La première phase de développement de l'IA générative a en effet accéléré une concentration sans précédent des moyens financiers à leur profit. Depuis 2019, la capitalisation des grands groupes numériques (Microsoft, Apple, Amazon, Alphabet, Meta) a plus que triplé, atteignant 11,8 billions de dollars. Si on ajoute le concepteur américain de semi-conducteurs Nvidia spécialisé dans les processeurs graphiques (GPU), leur valorisation dépasse 1,5 fois celle des 25 plus grandes entreprises américaines traditionnelles (pétrole, pharmacie, banque, etc.)¹⁶. Nvidia a été jusqu'à présent la principale bénéficiaire de la vague de l'IA générative : elle est passée de moins de 100 milliards de dollars de valorisation boursière à plus de 3 billions de dollars en quelques années seulement¹⁷.

Cette puissance financière leur a permis de s'engager dans une course technologique pour développer des modèles génératifs toujours plus

14. Les modèles de fondation sont des modèles d'IA entraînés sur de vastes quantités de données qui servent de base au développement d'applications plus spécialisées. GPT-4 d'OpenAI ou Claude d'Anthropic en sont des exemples emblématiques.

15. L'écosystème technologique désigne l'ensemble des composants interdépendants (infrastructures, outils, canaux de distribution) nécessaires au développement et au déploiement de ces technologies.

16. « Why America's Tech Giants Have Got Bigger and Stronger: Whatever Happened to Creative Destruction? », *The Economist*, 22 août 2024.

17. *Ibid.*

performants, dont l'amélioration repose sur deux leviers principaux : le développement d'algorithmes plus sophistiqués et l'augmentation de la puissance de calcul. Si les avancées algorithmiques sont fondamentales, elles sont aussi moins prévisibles et donc plus difficiles à planifier. Dans leurs investissements massifs en IA, les *Big Tech* ont davantage mis l'accent sur la puissance de calcul, qui offre un levier plus direct et mesurable pour améliorer les performances. Cette stratégie se reflète dans l'explosion des coûts d'entraînement : d'environ 100 millions de dollars fin 2024 pour un modèle de pointe, ils pourraient atteindre 5 à 10 milliards de dollars dès 2025-2026 selon le PDG d'Anthropic¹⁸. Cette escalade des coûts a réduit mécaniquement le nombre d'acteurs capables de développer des modèles de fondation.

L'apparition du modèle chinois DeepSeek suggère néanmoins qu'une autre approche est envisageable. Il aurait atteint des performances presque similaires à ChatGPT tout en s'appuyant sur des processeurs graphiques¹⁹ d'ancienne génération, avec une phase finale d'entraînement dont le coût s'élève, selon les auteurs, à un peu moins de six millions de dollars²⁰. Si ces résultats se confirment, ils pourraient démontrer que l'innovation algorithmique pourrait encore redistribuer les cartes, même face aux ressources considérables des géants technologiques. Cette percée souligne aussi la dualité du développement de l'IA : si la puissance computationnelle demeure un levier maîtrisable et nécessaire pour développer et déployer les modèles, les avancées par exemple dans l'architecture des modèles²¹ ou les techniques d'entraînement peuvent créer des points de rupture technologiques, difficiles à anticiper. Le développement des modèles de fondation pourrait ainsi continuer à être marqué par une dynamique mêlant progression linéaire de la puissance de calcul et bonds technologiques abrupts.

Cette dualité entre innovation algorithmique et puissance computationnelle se reflète dans la stratégie d'investissement des géants technologiques. Au-delà de leurs propres efforts de recherche, ils investissent massivement dans des startups innovantes comme Anthropic ou OpenAI (plus de 24 milliards de dollars cumulés entre Microsoft, Amazon et Google). La Federal Trade Commission (FTC) révèle que ces investissements sont souvent assortis d'accords de *circular spending* : les

18. S. Fiegeman et R. Metz, « AI Startups Struggle to Keep Up With Big Tech's Spending Spree », *Bloomberg*, 6 septembre 2024.

19. Les « processeurs graphiques » ou « cartes graphiques » sont historiquement conçus pour les calculs d'affichage des jeux vidéo. Leur architecture, optimisée pour effectuer de nombreux calculs en parallèle, s'est révélée particulièrement adaptée à l'IA. Les réseaux de neurones artificiels qui constituent le cœur des systèmes d'IA reposent sur des multiplications matricielles qui peuvent être décomposées en calculs parallèles. Là où un processeur classique (CPU) traite les opérations une par une, un GPU peut en traiter des milliers simultanément, accélérant considérablement l'entraînement et l'exécution des modèles d'IA.

20. « DeepSeek-V3 Technical Report », DeepSeek-AI, 18 février 2025.

21. L'architecture d'un modèle désigne le plan ou la structure qui organise les différentes parties du réseau de neurones et détermine comment l'information circule, se transforme et s'apprend au sein du modèle.

startups doivent alors dépenser une large part des fonds reçus en services *cloud* développés par leurs investisseurs²². Cette approche permet aux *Big Tech* d'optimiser leur positionnement stratégique sur les deux segments : se positionner sur l'innovation algorithmique tout en renforçant leur position dominante dans l'infrastructure *cloud*.

La maîtrise des infrastructures critiques

Au-delà du développement des modèles de fondation, les *Big Tech* renforcent leur position dominante grâce à leur mainmise sur les infrastructures essentielles de l'IA. Leurs investissements massifs dans les *data centers* et le *cloud* créent une double dépendance : ils sont non seulement les seuls à pouvoir opérer leurs propres modèles à grande échelle, mais ils contrôlent aussi l'accès aux ressources nécessaires à l'exploitation de tout autre modèle. Ainsi, même si des acteurs comme DeepSeek parviennent à développer des modèles plus efficaces, leur impact reste limité sans accès à ces infrastructures massives. Cette réalité consolide l'hégémonie des *Big Tech* américains qui, grâce à leurs investissements colossaux dans les infrastructures, conservent un avantage stratégique sur l'ensemble de la chaîne de valeur de l'IA.

AWS, Microsoft Azure et Google Cloud contrôlent ainsi 68 % du marché mondial des services *cloud*²³, tandis que Nvidia détient 90 % du marché des GPU, composants essentiels pour entraîner et exploiter les modèles d'IA²⁴. Cette position dominante se renforce continuellement : depuis 2019, les investissements des *Big Tech* dans les infrastructures d'IA ont explosé. Les cinq *Big Tech* et Nvidia ont ainsi doublé leurs dépenses en capital, atteignant 169 milliards de dollars en 2023 – dépassant largement les 135 milliards investis par les 25 plus grandes entreprises suivantes combinées²⁵. Ces investissements massifs sont d'autant plus critiques que les générations successives de GPU, qui se renouvellent à un rythme soutenu, représentent des coûts fixes considérables qui peuvent devenir obsolètes. Cette course technologique et financière s'est encore accélérée en 2024 : au dernier trimestre, Amazon, Microsoft et Alphabet ont investi 50,6 milliards de dollars, principalement dans leurs infrastructures d'IA²⁶ – l'équivalent du budget annuel de la défense française. Meta prévoit d'investir 10 milliards de dollars dans un *data center* IA en Louisiane²⁷, tandis qu'Elon Musk rejoint la

22. « Partnerships Between Cloud Service Providers and AI Developers – FTC Staff Report on AI Partnerships & Investments 6(b) Study », Federal Trade Commission, Office of Technology Staff, janvier 2025.

23. M. Haranas, « Cloud Market Share For \$84B Q3 2024: AWS, Microsoft, Google Cloud Lead », CRN, 7 novembre 2024.

24. N. Khan, « NVIDIA Crushes Rivals: Secures Unprecedented 90% of GPU Market in Q3 2024 », *Yahoo Finance*, 12 décembre 2024.

25. « Why America's Tech Giants Have Got Bigger and Stronger », *The Economist*, op. cit.

26. M. Kruppa et T. Dotan, « Tech Giants See AI Bets Starting to Pay Off », *The Wall Street Journal*, 1^{er} novembre 2024.

27. « Meta to Invest \$10 Billion for Louisiana Data Center », Reuters, 4 décembre 2024.

course avec xAI, dont le *data center* de Memphis – qualifié de « superordinateur le plus rapide de la planète » par le PDG de Nvidia – devrait doubler sa capacité pour atteindre 200 000 GPU²⁸.

Ces investissements colossaux ne visent pas uniquement l'entraînement des modèles actuels, mais préparent aussi l'infrastructure nécessaire au déploiement futur des services d'IA. Dans un contexte où les GPU restent difficiles d'accès, cette course à l'infrastructure répond directement aux besoins stratégiques des startups.

Pour les startups, la capacité à faire évoluer rapidement leurs services en cas de succès est cruciale. La migration d'un fournisseur d'infrastructure à un autre étant techniquement complexe et coûteuse, elles doivent d'emblée s'associer à des *providers* capables d'accompagner efficacement leur croissance future. Les investissements massifs des géants américains dans leurs infrastructures envoient dès lors un signal fort : ils seront les seuls à même de répondre aux besoins d'expansion rapide des services d'IA. Cette réalité pousse les startups européennes à dépendre des *data centers* américains dès leurs premières phases de développement, renforçant encore la position dominante des *Big Tech* sur l'ensemble de la chaîne de valeur.

Dans ce contexte, les crédits *cloud*, qui permettent d'utiliser gratuitement les infrastructures informatiques offertes par les *Big Tech* aux startups européennes, jouent un rôle paradoxal. Bien que conçus comme des instruments de dépendance à long terme, ces crédits représentent aussi une ressource indispensable pour les startups. Dans un écosystème où les levées de fonds restent difficiles, ils permettent aux startups d'accéder à des infrastructures de pointe sans mobiliser leur trésorerie limitée. Cette situation est emblématique de la complexité du rapport entre l'Europe et les *Big Tech* : si leur domination pose des questions stratégiques légitimes, leurs ressources technologiques et financières constituent, à court terme, un levier essentiel pour le développement de l'innovation européenne.

L'ampleur de cette course aux infrastructures est illustrée par le projet Stargate AI aux États-Unis : OpenAI, SoftBank et Oracle prévoient ainsi d'investir jusqu'à 500 milliards de dollars sur quatre ans dans le développement d'infrastructures d'IA, avec une première phase de 100 milliards de dollars pour un *data center* au Texas.

Face à cette annonce, l'Union européenne dévoile son plan InvestAI, à l'occasion du Sommet pour l'action sur l'intelligence artificielle qui s'est tenu à Paris en février 2025. S'inscrivant dans ce contexte de surenchère mondiale autour de l'IA, la Commission ambitionne de mobiliser 200 milliards d'euros via une alliance entre la Commission européenne et plus de 60 entreprises privées (regroupées sous la bannière de l'*European*

28. B. Jin et M. Bobrowsky, « Elon Musk's xAI in Talks to Raise Funding Valuing It at \$40 Billion », *The Wall Street Journal*, 29 octobre 2024.

AI Champions Initiative)²⁹. La Commission devrait apporter 50 milliards d'euros de financements publics, dont 20 milliards devraient servir à construire de grands *data centers* spécialisés dans l'IA, tandis que le secteur privé devrait contribuer à hauteur de 150 milliards³⁰. Cette initiative européenne fait écho à l'engagement français présenté en marge du même sommet, visant à mobiliser 109 milliards d'euros pour accélérer le développement de l'intelligence artificielle dans les prochaines années³¹.

Il est important de noter que la première vague d'investissements dans les infrastructures critiques de l'IA a été menée par le secteur privé – les *Big Tech* investissant massivement depuis plus de deux ans déjà. Si les récentes initiatives européennes marquent une prise de conscience politique bienvenue, elles doivent avant tout viser à créer un environnement favorable aux investissements privés. Plus qu'une injection de fonds publics, même complétée par des grands groupes, l'enjeu pour les pouvoirs publics européens et français est de catalyser un écosystème capable de s'auto-entretenir. Cela implique de mettre en place les conditions réglementaires, fiscales et industrielles permettant au secteur privé d'investir naturellement et massivement dans les infrastructures dont il a besoin.

Le contrôle des canaux de distribution

La dépendance des startups aux *Big Tech* ne se limite pas aux infrastructures d'IA : elle s'étend également aux canaux de distribution, essentiels pour atteindre leurs utilisateurs finaux. Or, que ce soit pour les applications mobiles grand public ou les solutions professionnelles, les startups européennes doivent passer par les plateformes des géants américains.

En ce qui concerne les applications mobiles, les plateformes de distribution d'applications Google Play et App Store contrôlent 95 % du marché en Europe et dans le monde (hors Chine)³². Cette position dominante leur permet d'imposer des commissions sur toutes les transactions, pour un total de 127,3 milliards de dollars dépensés sur les *stores* en 2024³³. Sur le marché professionnel, l'intégration aux écosystèmes dominants est aussi incontournable : Google Workspace et Microsoft 365 contrôlent à eux seuls 74 % du marché mondial des solutions bureautiques³⁴. Le besoin d'interopérabilité fait que les clients exigent

29. F. Chaaban, « L'UE annonce un plan historique de 200 milliards d'euros pour l'intelligence artificielle », *Toute l'Europe*, 12 février 2025, disponible sur : www.touteleurope.eu.

30. F. Y. Chee, « EU's AI Push to Get 50 Billion Euro Boost, Says von der Leyen », Reuters, 11 février 2025, disponible sur : www.reuters.com.

31. J. Weatherbed, « EU mobilizes \$200 billion in AI race against US and China », *The Verge*, 11 février 2025, disponible sur : www.theverge.com.

32. D. Curry, « App Store Data (2025) », *Business of Apps*, 27 janvier 2025.

33. N. Lefebvre, « L'App Store d'Apple domine le marché avec 91,6 milliards en 2024 (la moitié en abonnement) », *iPhonesoft*, 18 décembre 2024.

34. L. S. Vailshery, « Market Share of Major Office Productivity Software Worldwide as of February 2025 », Statista, février 2025.

souvent une compatibilité avec ces outils qu'ils utilisent au quotidien, faisant des géants technologiques américains des partenaires obligés pour toute startup développant des solutions professionnelles.

La domination des *Big Tech* sur les canaux de distribution prend une dimension particulière avec l'émergence de l'IA générative. Si leur position soulève des inquiétudes légitimes en matière de concurrence, elle offre paradoxalement des protections essentielles face aux nouveaux risques induits par ces technologies. L'explosion des services rendus possibles par ces technologies s'accompagne d'une multiplication des risques : applications frauduleuses, services de qualité médiocre, tentatives d'escroquerie sophistiquées. Là où la régulation traditionnelle peine à suivre le rythme de l'innovation, les *Big Tech* disposent des ressources et des mécanismes nécessaires pour filtrer efficacement les applications. Le processus rigoureux de vérification imposé par Apple à chaque mise à jour de toutes les applications illustre cette capacité à maintenir un niveau élevé de sécurité, une fonction d'autant plus précieuse que l'IA générative facilite la création massive de contenus malveillants³⁵.

Pour leur croissance, les startups dépendent non seulement des canaux de distribution mais aussi des plateformes publicitaires contrôlées par les géants numériques. Dans un marché mondial de la publicité aujourd'hui largement numérisé, estimé à 933 milliards de dollars en 2024, Google, Meta et Amazon captent 61 % des recettes (hors Chine)³⁶. Les startups européennes doivent par ailleurs investir une part significative de leur budget marketing sur ces plateformes (Facebook, Instagram, LinkedIn, YouTube) pour toucher de nouveaux clients³⁷.

Cette dépendance aux plateformes publicitaires s'avère d'autant plus critique que leur efficacité repose sur des algorithmes de ciblage sophistiqués, permettant aux startups d'optimiser leurs coûts d'acquisition clients. Les récentes évolutions réglementaires européennes, comme l'option de publicités moins personnalisées imposée à Meta³⁸, pourraient paradoxalement renforcer cette dépendance : une hausse des coûts d'acquisition fragiliserait en effet davantage les startups européennes face à

35. Ainsi, si les régulateurs européens cherchent légitimement à stimuler la concurrence sur le marché des applications mobiles, notamment via le *Digital Markets Act*, ils doivent également prendre en compte que l'intérêt des consommateurs ne réside pas uniquement dans l'accès à plus d'applications ou à des prix plus bas. Dans un écosystème numérique rendu plus risqué par l'IA générative, la capacité des plateformes dominantes à agir comme garde-fous contre les dérives représente une valeur significative pour les utilisateurs.

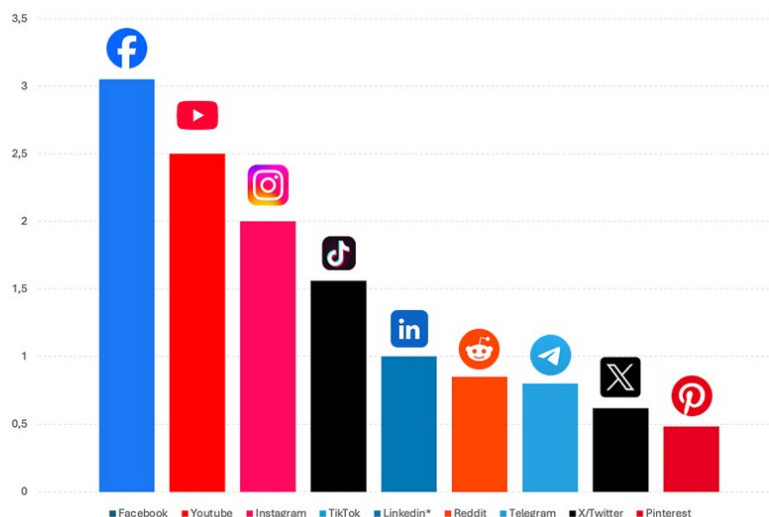
36. M. J. Levin, « Les recettes publicitaires mondiales en hausse de 10 % en 2024, selon Magna », *e-marketing.fr*, 9 décembre 2024.

37. Cette dynamique éclaire d'ailleurs le positionnement de Meta en faveur de l'open source dans l'IA : en rendant ses modèles librement accessibles, l'entreprise encourage la création de nouveaux services par les startups. Ces services, pour se développer, devront à leur tour investir en publicité sur les plateformes de Meta, renforçant ainsi son modèle économique principal basé sur la publicité.

38. S. Schechner et K. Mackrael, « In Europe, Instagram Ads Are About to Get Less Personal », *The Wall Street Journal*, 12 novembre 2024.

des concurrents disposant de ressources marketing plus importantes, voire les pousserait à se concentrer sur des marchés moins régulés, consolidant ainsi la position dominante des acteurs établis sur le marché européen.

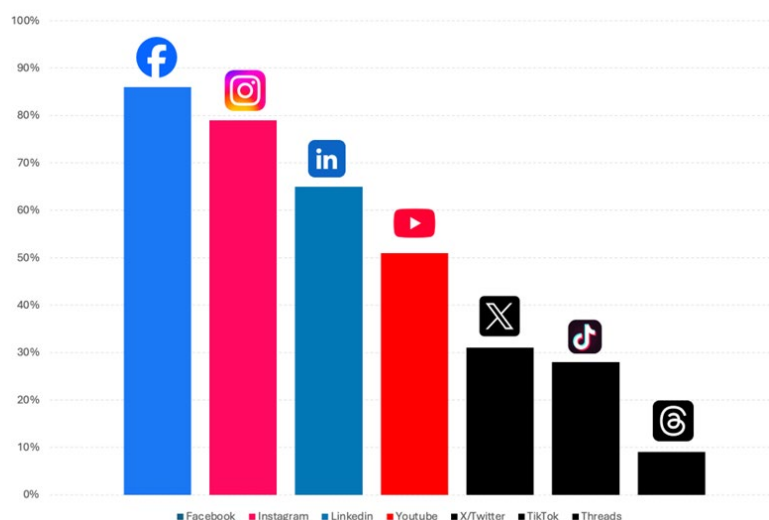
Graphique 1 : Nombre d'utilisateurs actifs mensuels des réseaux sociaux dans le monde en milliards (2024)



* Nombre total de « membres », LinkedIn ne fournissant pas de données sur les utilisateurs actifs mensuels (MAU).

Source : « Social Media User Statistics: Social Media Report 2024 », Blog2Social, 2024.

Graphique 2 : Principales plateformes de médias sociaux utilisées par les marketeurs dans le monde (2024)



Source : Statista Research Department, « Leading Social Media Platforms Used by Marketers worldwide 2024 », Statista, 10 décembre 2024.

L'influence sur l'écosystème *open source*

La dépendance des startups aux *Big Tech* ne se manifeste pas uniquement à travers leurs services commerciaux : elle s'étend également au domaine de l'*open source*, particulièrement critique dans le développement d'outils et de *frameworks* d'IA³⁹. En alimentant ces technologies, les *Big Tech* orientent les standards de l'industrie et facilitent l'adoption de leurs innovations. De TensorFlow, Go et Flutter (Google) à PyTorch (Meta) en passant par Triton Inference Server (Nvidia)⁴⁰, ils développent et assurent la maintenance des outils *open source* devenus des standards de l'industrie. En investissant massivement dans ces technologies, notamment par le biais d'équipes dédiées de développeurs, ils orientent les choix technologiques de l'ensemble de l'écosystème⁴¹. L'*open source* consolide de ce fait l'emprise des géants sur l'ensemble du secteur, une stratégie du reste tout à fait assumée par Mark Zuckerberg : « Les logiciels *open source* deviennent souvent des standards industriels et, quand les entreprises standardisent leur développement sur notre *stack* technologique⁴², il devient plus facile d'intégrer nos nouvelles innovations dans leurs produits.⁴³ »

Cette influence des *Big Tech* sur l'écosystème *open source*, bien que contribuant à leur domination du secteur, présente aussi des avantages significatifs pour l'ensemble de l'industrie. Les startups mais aussi les chercheurs bénéficient d'outils robustes et gratuits, développés et maintenus par des équipes hautement qualifiées, leur permettant de construire plus rapidement leurs services. Les investissements des géants technologiques dans la sécurité de ces outils constituent une protection précieuse pour l'ensemble de l'écosystème face aux risques inhérents au code ouvert (injections de code malveillant, *backdoors*, vers informatiques, bombes à retardement, etc.⁴⁴). À cet égard, l'influence

39. Un *framework* d'IA est un ensemble d'outils et de bibliothèques logicielles qui facilite la conception, l'entraînement et le déploiement de modèles d'intelligence artificielle.

40. Ces outils sont des *frameworks* et bibliothèques de développement essentiels pour l'IA et les applications numériques : TensorFlow et PyTorch permettent de construire et d'entraîner des modèles d'IA, Flutter facilite le développement d'applications mobiles et web, Go est un langage de programmation performant pour les applications *cloud*, et Triton Inference Server est un outil pour optimiser le déploiement des modèles d'IA à grande échelle.

41. A. Pannier, « Sources d'influence. Enjeux économiques et géopolitiques des logiciels open source », *Études de l'Ifri*, Ifri, décembre 2022.

42. Un « *stack* technologique » désigne l'ensemble cohérent des technologies, langages, outils et *frameworks* qu'une entreprise utilise pour développer ses produits ou services numériques.

43. E. Gent, « The Tech Industry Can't Agree on What Open-Source AI Means. That's a Problem. », *MIT Technology Review*, 25 mars 2024.

44. Injections de code malveillant : technique où un attaquant insère du code malveillant dans une application pour exécuter des actions non prévues ; *backdoors* : accès dissimulé laissé dans un système pour y entrer discrètement sans authentification normale ; vers informatiques : programmes autonomes qui se propagent automatiquement d'un système à l'autre, souvent pour causer des dégâts ou voler des données ; bombes à retardement : code malveillant programmé pour s'activer à une date ou un événement précis afin de perturber un système.

majeure des *Big Tech* sur l'*open source* contribue à structurer et sécuriser le secteur, créant un environnement plus stable pour l'innovation des startups.

Au-delà du mythe de l'indépendance technologique européenne

La domination des géants américains sur les infrastructures et les outils de développement impose un constat lucide : l'idée d'une indépendance technologique totale à court terme en Europe n'est ni réalisable ni nécessairement souhaitable. La mainmise des *Big Tech* sur l'écosystème technologique révèle aussi une forme d'interdépendance : si leur position leur permet de capter une part importante de la valeur, leurs investissements et leurs services créent aussi les conditions nécessaires au développement des startups européennes, lesquelles, en retour, permettent de valoriser leurs infrastructures.

La régulation européenne à elle seule ne saurait instaurer une véritable souveraineté numérique : celle-ci exige un écosystème dynamique d'entreprises européennes solides dans ce domaine. En effet, les États-Unis comptent 9 fois plus de licornes technologiques que la Chine et 5,5 fois plus que l'Europe⁴⁵. De plus, en moyenne, les startups européennes emploient six fois moins de salariées que leurs homologues américaines⁴⁶. Ce déficit d'entrepreneuriat n'est pas tant une conséquence du contrôle des *Big Tech* sur l'écosystème que le résultat de faiblesses structurelles européennes. Le marché du capital-risque européen reste fragmenté et sous-dimensionné par rapport aux États-Unis, les systèmes de retraites par répartition et l'absence de grands fonds de pension limitent les capacités d'investissement. La fragmentation du marché européen, combinée aux barrières linguistiques et culturelles, réduit la taille du marché potentiel et limite de fait la capacité des startups à diffuser leurs services à grande échelle.

Plus fondamentalement, l'Europe manque de startups ayant connu une forte croissance, ce qui limite non seulement le vivier de professionnels expérimentés mais aussi le développement plus large de l'écosystème. Les entreprises en forte croissance jouent en effet un rôle structurant à plusieurs niveaux : leurs anciens employés, riches de leur expérience, peuvent créer leurs propres startups ou accompagner la croissance d'autres entreprises en tant que cadres ou consultants. Ces entreprises deviennent également des clients potentiels importants pour les nouvelles startups, contribuent au financement de l'innovation à travers des investissements

45. « Les États-Unis comptent 9 fois plus de licornes technologiques que la Chine et 5,5 fois plus que l'Europe », *Itrnews*, 8 octobre 2024, disponible sur : <https://itrnews.com>.

46. D. Cerdeiro, G. H. Hong et A. Kammer, « How to Awaken Europe's Private Sector and Boost Economic Growth: EU Companies Grow and Innovate Less Than American Counterparts », *IMF Blog*, 11 septembre 2024.

directs, et enrichissent l'écosystème technique local en développant et partageant des outils *open source*, à l'image d'Hugging Face, qui propose des bibliothèques et plateformes facilitant l'utilisation et le partage de modèles d'IA. Cette dynamique vertueuse, bien établie dans l'écosystème américain, peine encore à émerger en Europe, limitant les opportunités de développement pour les nouvelles entreprises.

En définitive, le véritable enjeu n'est pas tant la position dominante des *Big Tech*, qui fournissent des services essentiels à l'innovation et à la croissance des startups, que l'absence d'acteurs européens parmi ces leaders technologiques mondiaux.

La chaîne de valeur de l'IA : identifier les opportunités pour l'Europe

Cette domination américaine invite à repenser le positionnement européen dans l'IA générative. La valeur économique principale de l'IA générative ne se cantonne pas à l'entraînement de modèles de fondation mais réside avant tout dans leur adaptation aux besoins spécifiques de multiples secteurs. Cette perspective ouvre la voie à une forme d'interdépendance plus équilibrée à court terme : si les startups européennes dépendent des *Big Tech* pour un certain nombre de technologies et pour leurs infrastructures, ces derniers ont aussi besoin d'un tissu d'entreprises innovantes pour valoriser pleinement leurs plateformes.

La décomposition de la chaîne de valeur de l'IA générative révèle ainsi une réalité plus nuancée que la simple domination américaine. Si les infrastructures et les modèles de fondation restent l'apanage des géants technologiques, des niveaux sous-jacents émergent : celui des outils de développement et des applications sectorielles verticalisées. C'est dans cet espace que l'Europe peut construire des avantages compétitifs durables. C'est là aussi que se concentre l'essentiel de la valeur économique de l'IA générative.

La diffusion tout autant que la production : repenser l'ambition européenne

La domination des infrastructures et des modèles de fondation ne garantit pas la captation de valeur : un pays peut générer des retombées économiques importantes, non pas en produisant de l'IA mais en excellant dans son adoption et son utilisation. Au-delà de la course technologique visant à développer des modèles d'IA plus performants, l'enjeu économique majeur réside dans la capacité des pays, des secteurs et des entreprises à adopter et intégrer efficacement ces technologies dans leurs activités. L'ampleur du potentiel est considérable : l'IA générative pourrait ajouter jusqu'à 4 000 milliards de dollars par an à l'économie mondiale – soit environ 1,5 fois le PIB de la France –, s'ajoutant aux 11 000 milliards de dollars déjà anticipés pour l'IA non générative et autres automatisations⁴⁷. Cette distinction entre développement et adoption est d'autant plus cruciale

47. J. Manyika et M. Spence, « The Coming AI Economic Revolution », *Foreign Affairs*, *op. cit.*

que les gains de productivité ne proviennent pas de la seule existence d'une technologie, mais de sa diffusion effective au sein du tissu économique. La révolution numérique en est la preuve : les gains de productivité significatifs n'ont émergé qu'une fois les entreprises ayant adapté leurs processus et formé leurs employés.

Dans la production de modèles d'IA, la France possède déjà son champion avec Mistral AI, dont la croissance rapide et la crédibilité technologique démontrent qu'une entreprise européenne peut s'imposer sur la scène mondiale. Plutôt que de disperser les ressources dans une course aux modèles souverains qui profite principalement à Nvidia – l'entreprise américaine prévoyait de générer près de 10 milliards de dollars en 2024 de la vente de GPU aux pays désireux de développer leurs propres modèles d'IA⁴⁸ –, l'Europe pourrait consolider cette position acquise en renforçant Mistral AI comme champion continental, tout en concentrant désormais ses efforts sur l'étape indispensable de la diffusion de l'IA dans son tissu économique et sur l'émergence de nouvelles licornes aux différents niveaux de la chaîne de valeur de l'IA générative.

Les outils de développement : un segment stratégique

La création de valeur dans l'IA ne se limite pas à la course aux grands modèles de langage (LLM). Même si les avancées technologiques des modèles de pointe comme GPT-4 ou Gemini s'arrêtaient aujourd'hui, un immense potentiel résiderait dans leur transformation en applications concrètes⁴⁹. La compétition intense que se livrent les *Big Tech*, tant sur les infrastructures que sur les modèles de fondation, a entraîné une baisse significative des coûts et une amélioration constante des performances. Cette dynamique concurrentielle a aussi permis la structuration de la chaîne de valeur de l'IA générative, faisant émerger de nouveaux segments d'opportunités.

Entre les infrastructures de base et les applications finales, l'échelon des outils de développement joue également un rôle majeur. Ces outils permettent d'adapter, de sécuriser, d'optimiser et de déployer les modèles d'IA dans des usages concrets. Plusieurs acteurs majeurs illustrent la diversité de ces besoins technologiques : Hugging Face, startup française, s'est imposée comme une référence dans le partage et l'adaptation des modèles grâce à ses outils *open source* et sa plateforme communautaire. L'entreprise américaine Langchain a développé des solutions essentielles pour l'orchestration et l'intégration des LLM dans des applications

48. A. Fitch et S. Schechner, « Nvidia's New Sales Booster: The Global Push for National AI Champions », *The Wall Street Journal*, 9 juin 2024.

49. « A Venture Capitalist on Where the AI Opportunities Are for Investors », *The Wall Street Journal*, 27 octobre 2024.

complexes (par exemple, pour créer des assistants capables de répondre à des questions en s'appuyant sur une base de données interne). Dans le domaine de la fiabilité et du suivi des performances, l'américaine Galileo fournit des outils d'évaluation indispensables, tandis que Datadog, entreprise fondée par des Français, est devenue leader dans le *monitoring* des modèles en production (il permet, par exemple, de suivre en temps réel les performances, les erreurs ou les temps de réponse d'un modèle déployé dans une application).

De nouvelles startups françaises émergent également sur des besoins critiques. Pruna AI s'attaque à l'optimisation de la vitesse d'inférence des modèles⁵⁰, un enjeu majeur qui permet de réduire significativement les coûts et la complexité du déploiement. Dans le domaine de la sécurité, Leanear développe des solutions innovantes de chiffrement pour protéger les *datasets* d'IA, devenus aujourd'hui les véritables actifs stratégiques⁵¹, tout en préservant la flexibilité dans le choix des infrastructures *cloud*⁵².

Dans ce segment des outils de développement, où la maîtrise technique prime sur la puissance financière, l'Europe et particulièrement la France disposent d'atouts considérables. La tradition d'excellence en mathématiques et en ingénierie fournit un vivier de talents particulièrement adapté aux défis complexes de l'IA⁵³. Cette expertise technique, combinée à une culture de l'innovation pragmatique, positionne idéalement les entreprises françaises sur ce marché où la différenciation repose moins sur les moyens financiers que sur la capacité à développer des solutions robustes et innovantes.

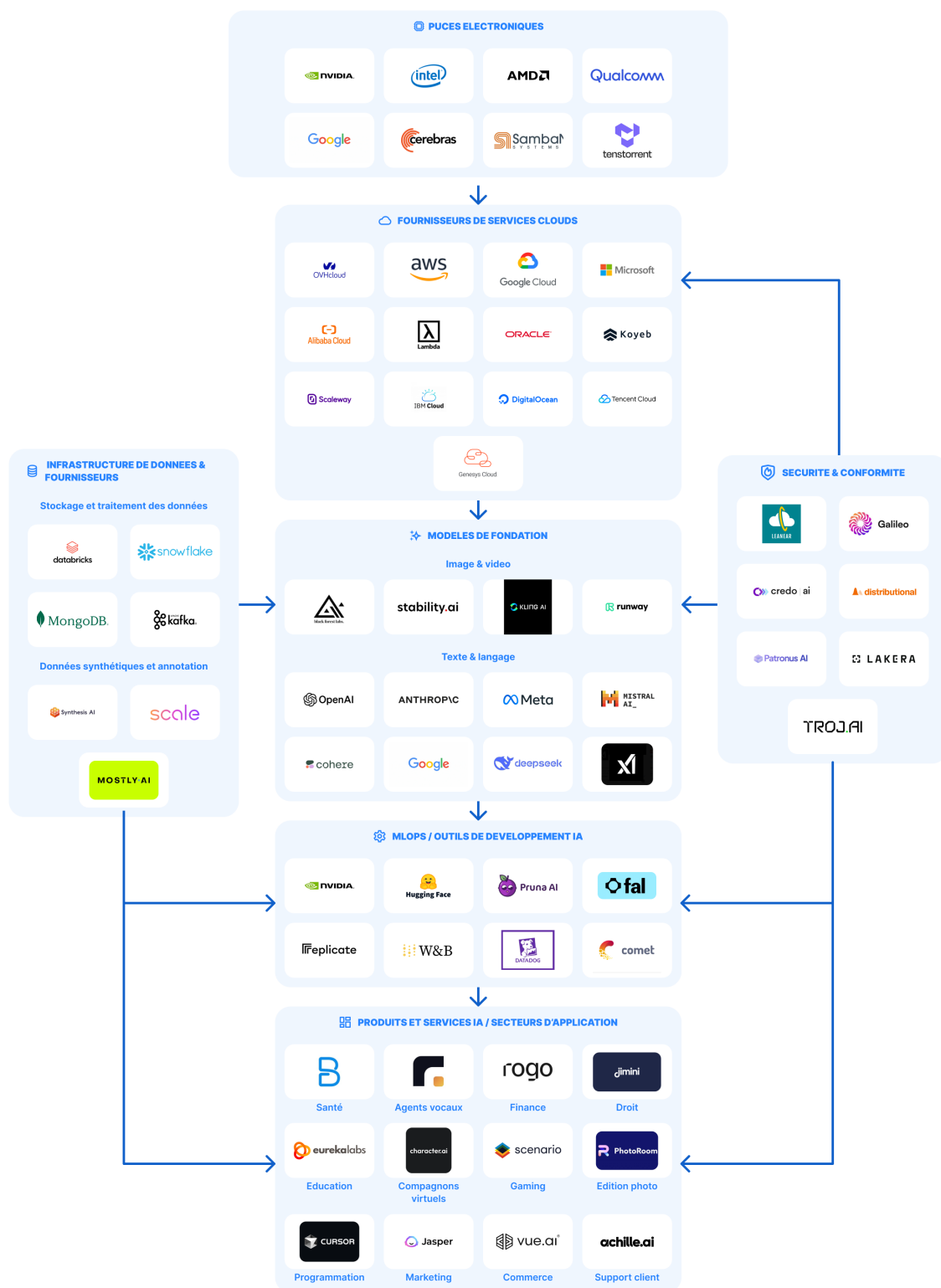
50. L'inférence, c'est-à-dire l'exécution d'un modèle d'IA pour produire des résultats, représente souvent plus de 90 % des coûts opérationnels d'une application d'IA. Optimiser cette vitesse d'inférence est donc crucial : cela permet non seulement de réduire significativement les coûts d'utilisation, mais aussi de rendre possible l'utilisation de modèles d'IA plus sophistiqués dans des applications ayant des contraintes de temps réel ou des ressources limitées.

51. Dans le domaine de l'IA, les jeux de données propriétaires constituent l'actif stratégique le plus précieux des entreprises. Alors que les avancées en matière d'architectures de modèles et d'algorithmes se succèdent rapidement et deviennent la plupart du temps publiquement accessibles, la possession d'une base de données propriétaire offre un avantage concurrentiel pérenne. Les nouvelles avancées algorithmiques peuvent être rapidement intégrées en réentraînant les modèles sur ces données existantes.

52. Les solutions de chiffrement permettent aux industriels européens de sécuriser leurs données, indépendamment de l'infrastructure *cloud* choisie – qu'elle soit américaine (AWS, Google Cloud, Azure), européenne (OVH, Orange) ou même asiatique (Alibaba Cloud). Cette approche leur permet de sélectionner les fournisseurs *cloud* sur des critères purement économiques et techniques tout en conservant un contrôle souverain sur leurs données, même lorsqu'elles sont hébergées hors de l'Union européenne.

53. M. Quiret, « Matthieu Courtecuisse : "La France peut devenir à l'IA ce que l'Inde est pour l'informatique" », *Les Echos*, 16 décembre 2024.

Graphique 3 : La chaîne de valeur de l'IA générative



La verticalisation : le principal atout européen ?

En complément de ces outils de développement, un second niveau d'opportunité émerge pour l'Europe : celui de la verticalisation sectorielle. Quant aux géants comme OpenAI, s'ils captent l'attention médiatique, de nombreuses entreprises se positionnent en développant des applications avec des modèles existants, en « *fine-tuning* » des modèles⁵⁴ ou créant parfois même de plus petits modèles spécialisés. En tirant parti des outils de développement et des modèles existants (*open source* ou API), elles permettent de développer des solutions verticales adaptées aux besoins spécifiques de chaque industrie. Cette approche présente un double avantage : elle réduit significativement les coûts de développement tout en accélérant les cycles d'innovation, une application pouvant être conçue et testée en quelques mois seulement.

L'Europe dispose de deux atouts majeurs pour se positionner sur ces segments. Le premier est son expertise sectorielle : la présence historique de filières industrielles de pointe comme l'automobile, l'aéronautique ou les biotechnologies lui permet de développer des applications d'IA ajustées aux spécificités de chaque secteur. Le second atout est la spécificité du marché européen, caractérisé par des réglementations strictes et une forte proximité avec les clients. Cette connaissance approfondie du terrain et des contraintes locales est particulièrement précieuse dans des domaines réglementés.

Le cas du secteur juridique illustre par exemple la manière dont les spécificités réglementaires européennes créent des barrières naturelles à l'entrée pour des acteurs étrangers. Malgré les moyens considérables d'acteurs américains comme Harvey (200 millions de dollars de levée à ce jour), la division fondamentale entre *common law* et *civil law* crée deux marchés distincts. Cette segmentation naturelle protège les acteurs spécialisés européen et français comme Jimini⁵⁵, qui peuvent se concentrer sur leur marché spécifique sans faire face à la concurrence directe d'acteurs étrangers à court terme.

De nombreuses entreprises françaises démontrent déjà le potentiel de cette approche sectorielle. PhotoRoom s'est imposé comme un leader mondial de l'édition photo, atteignant une valorisation de 500 millions d'euros grâce à ses algorithmes d'IA spécialisés. Dans le domaine du *gaming*, Scénario est devenue une référence dans la création d'*assets* digitaux pour les jeux vidéo, illustrant la capacité des entreprises françaises à s'imposer sur des

54. Le *fine-tuning* consiste à reprendre un modèle d'IA déjà entraîné sur de grandes quantités de données et à l'adapter précisément à une tâche ou un domaine spécifique avec un jeu de données plus restreint. Cette approche permet d'obtenir des performances supérieures sur des usages spécialisés tout en évitant les coûts massifs d'un entraînement complet.

55. Jimini est une startup française qui développe un assistant IA pour la recherche, l'analyse et la rédaction de documents juridiques.

marchés mondiaux. La relation client constitue un autre domaine où l'expertise française se distingue : CallRounded propose des agents vocaux IA pour « réhumaniser » les services clients là où les contraintes de coûts avaient conduit à leur automatisation, tandis qu'Achille.ai optimise le support client grâce à des agents augmentés par l'IA.

Dans le secteur médical, Biolevate illustre l'apport d'une IA sectorielle française en résolvant l'effet « boîte noire » grâce à une IA qui explicite son raisonnement en détail et de façon vérifiable. Cette approche renforce la maîtrise des chercheurs sur leurs travaux, tout en leur faisant bénéficier de la puissance de l'IA. Ces outils trouvent une application particulièrement pertinente dans la recherche sur les maladies orphelines, où les coûts de recherche constituent souvent un frein majeur.

Ces exemples démontrent que la verticalisation permet aux entreprises européennes de créer des solutions d'IA pertinentes et robustes. Cette stratégie, particulièrement adaptée aux forces de l'écosystème européen, ouvre la voie à une innovation technologique ancrée aux besoins réels de chaque industrie.

La miniaturisation des modèles : une tendance alignée avec la spécialisation sectorielle

Cette expertise sectorielle prend une importance croissante alors qu'une nouvelle phase semble émerger dans le développement de l'IA générative. Si en amont de la chaîne de valeur la première étape a été marquée par le gigantisme des modèles de fondation, notamment en termes de besoins en GPU et de nombre de paramètres, il semble se dessiner désormais une tendance à la « miniaturisation » des modèles qui pourrait aller de pair avec la spécialisation sectorielle des applications. Le succès de Deepseek, dont le modèle serait 90 % moins coûteux en inférence que ceux d'OpenAI⁵⁶, pourrait préfigurer de l'émergence de cette nouvelle dynamique.

Cette approche ciblée présente plusieurs avantages significatifs. Premièrement, ces modèles plus petits et spécialisés permettent des temps de réponse plus rapides, réduisant ainsi les coûts tout en améliorant l'expérience client. Deuxièmement, leur taille réduite permet une optimisation fine pour des usages spécifiques, augmentant leur efficacité. Troisièmement, les LLM peuvent servir de base pour créer, par exemple par distillation⁵⁷, des modèles plus petits et spécialisés. Les données de

56. T. Kim, « Wall Street Got DeepSeek All Wrong. Why It's a Boon for Big Tech, Especially Nvidia », *Barron's*, 31 janvier 2025.

57. La distillation de modèles en *deep learning* consiste à entraîner un modèle plus petit (*student*) à imiter un modèle plus grand (*teacher*), en utilisant ses prédictions. Cela permet de réduire la taille du modèle tout en conservant de bonnes performances.

synthèse pourraient jouer ici un rôle crucial en permettant de contextualiser l'information et d'orienter efficacement ces modèles vers leur spécialisation. Cette approche correspond particulièrement bien aux usages sectoriels où l'objectif n'est plus d'atteindre une connaissance universelle, mais plutôt d'exceller dans un domaine précis. Par exemple, un modèle destiné à analyser des rapports d'incidents industriels n'a pas besoin d'intégrer des connaissances en littérature médiévale.

Cette tendance à la miniaturisation ouvre une voie stratégique pour l'Europe. La combinaison de modèles spécialisés avec son expertise sectorielle et ses données industrielles permettrait de développer des solutions plus performantes que les modèles généralistes. L'Europe dispose ainsi d'un avantage naturel sur ce segment en émergence, qui s'aligne parfaitement avec ses atouts industriels et technologiques.

Une nouvelle donne pour l'industrie technologique

Si les géants technologiques américains disposent d'une puissance financière leur permettant d'investir massivement dans l'IA générative, l'intensification de leur rivalité pourrait paradoxalement créer de nouvelles opportunités pour l'Europe. Cette dernière partie analyse comment la compétition acharnée entre les *Big Tech*, motivée par leur peur de manquer le virage de l'IA, pourrait servir les intérêts européens. Au-delà de cette rivalité, le marché américain lui-même, avec son pouvoir d'achat élevé et sa propension à adopter rapidement les nouvelles technologies, pourrait constituer un levier de croissance majeur pour les entreprises européennes qui sauront développer une approche hybride entre les deux continents.

La fragilité paradoxale des géants du numérique

Derrière leur apparence de toute-puissance, les *Big Tech* présentent des profils contrastés, avec des forces et des vulnérabilités distinctes. Chacun d'entre eux doit composer avec des fragilités internes, qu'il s'agisse de difficultés à réussir en dehors de son cœur de métier (Google dans le *hardware*, Amazon dans l'alimentaire)⁵⁸ ou de paris coûteux, comme la réalité virtuelle pour Meta⁵⁹ ou la voiture électrique pour Apple⁶⁰.

Les géants technologiques sont marqués par l'histoire de leurs prédécesseurs déçus, et hantés par la crainte de passer à côté de la prochaine révolution technologique majeure. Les exemples sont nombreux : Fairchild Semiconductor, qui dominait l'industrie des semi-conducteurs dans les années 1950 ; IBM, l'entreprise la plus rentable d'Amérique en 1983 ; et plus récemment, Nokia, qui s'est effondré pour n'avoir pas anticipé la révolution des smartphones⁶¹.

58. A. Beard, « Can Big Tech Be Disrupted? A Conversation with Columbia Business School Professor Jonathan Knee », *Harvard Business Review*, janvier-février 2022.

59. S. E. Needleman, « Meta Reports Record Revenue, Slower Digital Advertising Growth: The Social-Media Giant Is Spending Billions on Artificial Intelligence and Products for Virtual and Augmented Reality », *The Wall Street Journal*, 31 octobre 2024.

60. M. Acton et P. Campbell, « Apple Parks Its Electric Car Project in Sign of EV Industry's Struggles », *Financial Times*, 28 février 2024.

61. « Big Tech's Supersized Ambitions: From Metaverses to Quantum Computing », *The Economist*, 22 janvier 2022.

Contrairement aux secteurs traditionnels comme l'énergie ou l'industrie lourde, qui nécessitent d'importants investissements en actifs physiques, la valeur des entreprises du numérique repose principalement sur des actifs immatériels : logiciels, brevets, bases d'utilisateurs et publicité en ligne avec un coût marginal proche de zéro. Cette légèreté structurelle, si elle permet une rentabilité remarquable, les rend aussi paradoxalement plus vulnérables aux ruptures technologiques : bien que les barrières à l'entrée soient significatives, elles sont moins insurmontables que dans les industries nécessitant de lourdes infrastructures.

Cette dynamique se traduit par l'ascension fulgurante de nouveaux acteurs, un scénario impensable dans des secteurs plus traditionnels. Facebook a ainsi été créé et propulsé en quelques années par un étudiant de moins de 20 ans devenu l'un des hommes les plus puissants de la planète. De la même façon, l'irruption d'OpenAI et la popularité fulgurante de ses modèles de langage rappellent combien il est possible, pour un nouvel entrant, d'ébranler la domination de géants bien établis depuis plusieurs décennies.

L'intensification de la compétition entre *Big Tech* et avec la Chine

Les géants technologiques s'engagent dans une course effrénée aux investissements en IA, perçue comme la révolution technologique majeure des prochaines décennies. Les dépenses en infrastructures ont bondi de 66 % au dernier trimestre 2024 par rapport à l'année précédente⁶², créant un « dilemme du prisonnier »⁶³ où aucune entreprise n'ose réduire ses investissements malgré les risques de surcapacité⁶⁴. Au-delà de simples acquisitions, les *Big Tech* américains deviennent les architectes de l'écosystème IA dès ses premiers stades, Microsoft, Amazon, Google et Nvidia dominant le financement du secteur et éclipsant progressivement les fonds de capital-risque traditionnels⁶⁵.

Cette dynamique de compétition est d'autant plus frappante que la Chine émerge comme un concurrent majeur dans l'IA générative, malgré les restrictions américaines. Alibaba a démontré sa capacité d'innovation

62. M. Kruppa et T. Dotan, « Tech Giants See AI Bets Starting to Pay Off: Microsoft, Google and Amazon Report Strong Growth in Cloud Revenue, but Warn of Increased Spending », *The Wall Street Journal*, 1^{er} novembre 2024.

63. Le dilemme du prisonnier, concept clé de la théorie des jeux, décrit une situation où des acteurs, bien que conscients qu'une coopération serait collectivement bénéfique, choisissent rationnellement une stratégie non-coopérative par crainte d'être désavantagés. Dans le cas des *Big Tech*, bien qu'une modération collective des investissements en IA pourrait être plus profitable, chacune est contrainte d'investir massivement par peur de perdre son avantage compétitif si elle était la seule à se restreindre.

64. S. Fiegman et R. Metz, « AI Startups Struggle to Keep Up With Big Tech's Spending Spree », *Bloomberg*, 6 septembre 2024.

65. J. Thornhill, « The AI Revolution Is Generating Some Investor "Hallucinations" Too: Big Tech Firms Are Muscling In on Funding the Future, Pushing Silicon Valley VCs Out of the Way », *Financial Times*, 25 janvier 2024.

avec son modèle Qwen 2.5, confirmant la montée en puissance des acteurs chinois dans ce domaine. L'annonce de DeepSeek, quant à elle, a eu un impact spectaculaire sur les marchés financiers américains, effaçant temporairement près de 1 000 milliards de dollars de valorisation des géants technologiques⁶⁶, dont 600 milliards pour Nvidia seul⁶⁷.

Cette percée chinoise s'accompagne d'une stratégie *open source* délibérée, qui contraste avec le modèle fermé d'OpenAI ou d'Anthropic et vise à démocratiser l'accès aux modèles performants et à empêcher toute concentration excessive du marché autour des acteurs américains. En diffusant largement leurs innovations, notamment leurs avancées en matière d'optimisation et d'efficacité computationnelle, les entreprises chinoises contribuent à maintenir le marché ouvert et à limiter la capacité des géants américains à imposer leurs standards propriétaires comme seule option viable.

Cette intensification de la compétition dans le domaine de l'IA s'inscrit dans une tendance plus large de concurrence accrue entre les grands groupes du numérique. Depuis 2015, la part de leur chiffre d'affaires provenant de marchés où ils se concurrencent directement a doublé, passant de 20 % à 40 %. Cette rivalité croissante traduit leur crainte de voir leur position historique s'éroder, les poussant à investir massivement dans de nouveaux territoires⁶⁸. La course à l'IA représente ainsi l'apogée de cette dynamique de compétition entre les géants du numérique, mais aussi entre les deux premières puissances économiques mondiales, chacun cherchant à y établir rapidement une position dominante.

Un modèle économique en transformation

Si les *Big Tech* se sont historiquement distingués par leur modèle économique léger, l'IA générative pourrait marquer un tournant vers un secteur beaucoup plus intensif en capital. Cette transformation soulève des questions fondamentales sur la rentabilité future de ces entreprises, alors même que les coûts des infrastructures de base – GPU, énergie, réseaux – demeurent structurellement élevés.

L'escalade des dépenses des *Big Tech* prend une telle ampleur qu'elle commence à susciter des inquiétudes à Wall Street et soulève des questions sur le retour sur investissement de ces montants colossaux. Si Microsoft, Meta, Amazon et Alphabet tentent de rassurer les investisseurs

66. T. Kim, « Wall Street Got DeepSeek All Wrong », *Barron's*, *op. cit.*

67. B. Alpert, « After DeepSeek Claim, Export Controls Should Be Enforced, Not Abandoned », *Barron's*, 27 janvier 2025.

68. « Big Tech's Supersized Ambitions: From Metaverses to Quantum Computing », *The Economist*, 22 janvier 2022.

en soulignant les bénéfices de l'IA sur leurs services existants et leurs coûts opérationnels, ces avantages restent difficiles à quantifier face à l'explosion très concrète des dépenses en infrastructures⁶⁹. Bien que ces investissements massifs témoignent de la volonté des géants technologiques de privilégier la croissance à long terme plutôt que les profits immédiats, rien ne garantit que les sommes considérables investies dans les capacités d'IA générative trouveront des applications assez rentables pour couvrir ces coûts exorbitants⁷⁰.

La montée en puissance de l'IA générative pourrait ainsi marquer un tournant fondamental dans la structure économique du secteur technologique et ouvrir une nouvelle ère plus complexe pour les géants de la *tech*. Le cas d'OpenAI est révélateur des défis à venir. Son PDG Sam Altman décrit l'entreprise comme « la startup la plus intensive en capital de l'histoire de la Silicon Valley⁷¹ ». Malgré des revenus annuels estimés entre 3,5 et 4,5 milliards de dollars, l'entreprise fait face à des coûts opérationnels estimés à 8,5 milliards de dollars pour 2024, dont 4 milliards uniquement pour la puissance de calcul et 3 milliards pour l'entraînement des modèles⁷². Cette structure de coûts pourrait préfigurer l'avenir du secteur. Certains géants commencent même à investir dans l'énergie nucléaire pour faire face aux exigences énergétiques croissantes de ces technologies⁷³.

Dans ce contexte de transformation profonde du modèle économique des entreprises technologiques liée à l'IA générative et à ses coûts d'infrastructure élevés, les récentes orientations stratégiques américaines suggèrent qu'au-delà de la course aux modèles, le contrôle des infrastructures pourrait jouer un rôle décisif à moyen terme. Si des avancées comme le modèle chinois DeepSeek démontrent que la compétition reste ouverte sur le plan technologique, la capacité à déployer massivement ces technologies dépendra largement de la maîtrise de la chaîne d'infrastructures – des GPU aux centres de données, des réseaux de communication à l'approvisionnement énergétique compétitif. Les investissements massifs de Washington dans des projets comme Stargate et ses nouvelles politiques énergétiques témoignent de cette vision où la domination technologique reposera tout autant sur la maîtrise de ressources critiques que sur l'excellence des algorithmes.

69. R. Waters et T. Bradshaw, « Wall Street Frets Over Big Tech's \$200bn AI Spending Splurge », *Financial Times*, 1^{er} novembre 2024.

70. R. Armstrong et A. Reiter, « Big Tech's Big Investments », *Financial Times*, 31 juillet 2024.

71. « LLM Price War, Black Forest's Open Image Generator, The High Cost of AI Leadership, Machine Translation Goes Agentic », *The Batch AI News and Insights*, n° 262, DeepLearning.AI, 14 août 2024.

72. *Ibid.*

73. A. Salzman, « Meta Announces \$10 Billion Data Center in Louisiana », *Barron's*, 4 décembre 2024.

Le levier du marché américain

Au-delà d'une approche pragmatique centrée sur les forces sectorielles européennes, l'écart de développement économique entre les États-Unis et l'Europe pourrait paradoxalement constituer un atout. Dans un contexte de productivité en baisse et de vieillissement démographique, le marché américain, caractérisé par un pouvoir d'achat élevé et une forte propension à adopter rapidement les technologies, représente une source de revenus substantielle pour les solutions d'IA européennes. Un modèle d'affaires prometteur émerge ainsi : développer et valider les services en Europe, où les coûts sont plus faibles et l'excellence technique reconnue, pour ensuite les commercialiser aux États-Unis où ils peuvent être rapidement monétisés⁷⁴.

Cette dynamique est d'ailleurs déjà renforcée par la fragmentation du marché européen en comparaison des États-Unis. Plutôt que de se développer en Europe, où chaque pays nécessite une adaptation spécifique et coûteuse des solutions d'IA aux langues et contextes locaux, les startups européennes privilégient souvent une commercialisation directe sur le marché américain, plus unifié et plus vaste.

« Pour être un champion européen, il faut d'abord être un champion aux États-Unis.⁷⁵ » Cette réalité du marché de l'IA dessine un modèle hybride, tirant parti des atouts complémentaires des deux continents. L'Europe, et particulièrement Paris, dispose d'une concentration remarquable de talents en IA mais peine à transformer cette excellence académique en succès commerciaux. Face à ce constat, certaines entreprises optent pour une approche bicéphale : établir leur siège social aux États-Unis pour bénéficier de l'écosystème de financement mature, tout en maintenant une base technique en Europe où les coûts de développement restent compétitifs. Cette stratégie permet de contourner les défis du marché du travail américain, où l'influence des géants technologiques a créé une forte inflation salariale. Les entreprises peuvent ainsi combiner le meilleur des deux mondes : l'accès aux capitaux et au marché américains, tout en maintenant des coûts opérationnels maîtrisés grâce à des équipes techniques basées en Europe.

En définitive, l'Europe peut transformer ses apparentes faiblesses – fragmentation du marché, moindre capacité d'investissement dans les infrastructures – en avantages stratégiques. En combinant son excellence technique et ses coûts opérationnels compétitifs avec un accès au marché et aux capitaux américains, elle peut construire une forme innovante de souveraineté numérique. Celle-ci ne reposerait pas sur une illusoire

74. M. Quiret, « Matthieu Courtecuisse : “La France peut devenir à l'IA ce que l'Inde est pour l'informatique” », *Les Echos*, 16 décembre 2024.

75. F. Douetteau, « Building European champions in AI Panel, AI Action Summit at Station F », Youtube, 11 février 2025, disponible sur : www.youtube.com.

indépendance technologique totale, mais sur sa capacité à développer des solutions d'IA spécialisées et à les déployer efficacement des deux côtés de l'Atlantique. Cette stratégie permettrait à l'Europe de renforcer sa position dans la chaîne de valeur mondiale de l'IA, tout en préservant son autonomie.

Conclusion

L'histoire économique est ponctuée de ruptures technologiques qui ont transformé radicalement la production et la distribution des biens et services. Si l'informatisation a révolutionné le calcul et internet la distribution, l'IA générative marque aujourd'hui une nouvelle étape en réduisant drastiquement les coûts associés au langage, au raisonnement et à la création⁷⁶. Cette innovation pourrait ainsi ouvrir la voie à un nouveau cycle majeur de transformation économique⁷⁷.

Néanmoins, l'ampleur de cette rupture technologique ne se mesure pas uniquement à la puissance des modèles d'IA développés, mais surtout à leur capacité à être intégrés dans des usages économiques concrets. Cette réalité invite à repenser la création de valeur dans l'IA : les gains de productivité ne seront réels que si ces technologies sont effectivement déployées dans le tissu industriel et commercial⁷⁸.

L'écosystème européen évolue aujourd'hui dans un contexte contraint : toutes les startups du secteur de l'IA générative dépendent, de près ou de loin, des services fournis par les *Big Tech* américains, qui contrôlent une part significative de la chaîne de valeur, des infrastructures aux modèles de fondation. Face à cette réalité, l'Europe doit adopter une approche pragmatique : s'appuyer à court terme sur ces plateformes tout en capitalisant sur ses atouts propres. Si l'Europe peut compter sur Mistral AI, qui, malgré des financements nettement inférieurs à ceux de ses concurrents directs, s'est imposé comme un champion dans la production de modèles, elle bénéficie aussi surtout d'une expertise sectorielle reconnue dans de nombreuses industries. Cette excellence technique, particulièrement visible en aval de la chaîne de valeur de l'IA générative, lui permet de se positionner précisément là où se situe l'essentiel de la valeur économique.

À court terme, l'enjeu pour l'Europe n'est donc pas nécessairement de s'affranchir de la présence dominante des *Big Tech* – qui fournissent aujourd'hui des services essentiels à l'innovation – que de maximiser les retombées économiques de l'IA générative. Deux opportunités se dessinent clairement dans ce contexte. D'une part, la stratégie des géants américains, concentrés sur les infrastructures massives et les modèles génériques, ouvre des niches stratégiques où la connaissance approfondie d'un secteur prime.

76. « A Venture Capitalist on Where the AI Opportunities Are for Investors », *The Wall Street Journal*, 27 octobre 2024.

77. *Ibid.*

78. J. Ding, « The Innovation Fallacy – In the U.S.-Chinese Tech Race, Diffusion Matters More Than Invention », *Foreign Affairs*, 19 août 2024.

Les entreprises européennes spécialisées dans la santé, l'industrie ou la finance peuvent ainsi tirer parti de leur expertise pour proposer des solutions IA ciblées, moins couvertes par les *Big Tech*. D'autre part, la rivalité croissante entre ces acteurs crée une dynamique favorable aux entreprises européennes. Leur surenchère d'investissements dans les infrastructures et les modèles génériques les pousse à proposer des conditions d'accès toujours plus avantageuses pour conquérir des parts de marché. Cette concurrence permet aux entreprises européennes d'accéder à des technologies de pointe à moindres coûts, renforçant ainsi leur capacité à développer des solutions innovantes dans leurs domaines d'expertise.

À moyen terme, le défi européen dépasse la simple question de la dépendance aux *Big Tech*. Le véritable enjeu réside dans l'absence d'acteurs européens parmi les leaders mondiaux de la technologie. Alors que les États-Unis et la Chine ont su faire émerger leurs propres géants technologiques, l'Europe peine encore à transformer son excellence technique en succès commerciaux d'envergure mondiale. Ce retard s'explique par des différences structurelles profondes, comme le souligne le FMI : depuis 2005, la productivité des entreprises technologiques américaines a augmenté de 40 %, tandis qu'elle stagne en Europe⁷⁹. Plusieurs facteurs expliquent cette divergence : des coûts de conformité réglementaire élevés et surtout un sous-investissement chronique du secteur privé dans les startups⁸⁰.

Le problème réside donc avant tout dans la faible capacité de l'écosystème technologique européen à s'autofinancer. Celui-ci dépend encore trop souvent de l'impulsion publique, même lorsque des fonds privés importants sont mobilisés dans le cadre de partenariats public-privé. Les récentes annonces de 50 milliards d'euros combinés à 150 milliards de capitaux privés en sont une illustration. Ces investissements, bien qu'importants, restent un ordre de grandeur très largement en dessous de ce que l'écosystème technologique américain a investi depuis le début de la vague de l'IA générative. Rien que les quatre principales *Big Tech* – Microsoft, Alphabet, Amazon et Meta – prévoient de dépenser jusqu'à 320 milliards de dollars combinés en technologies d'IA et en construction de centres de données en 2025⁸¹. Leurs dépenses totales en capital s'élevaient déjà à 230 milliards de dollars en 2024⁸² et 151 milliards en 2023⁸³.

79. D. Cerdeiro, G. H. Hong et A. Kammer, « How to Awaken Europe's Private Sector and Boost Economic Growth – EU companies grow and innovate less than American counterparts », *IMF Blog*, 11 septembre 2024.

80. B. Martens, « Draghi Disappoints on Digital – The Draghi Report's Emphasis on Hardware and Telecoms is Out of Touch with Modern Digital Developments », Bruegel, 11 septembre 2024.

81. S. Subin, « Tech Megacaps Plan to Spend more than \$300 Billion in 2025 as AI Race Intensifies », *CNBC*, 8 février 2025, disponible sur : www.cnbc.com.

82. *Ibid.*

83. S. Morris et R. Uddin, « Big Tech Lines Up Over \$300bn in AI Spending for 2025 », *Financial Times*, 7 février 2025, disponible sur : www.ft.com.

L'avenir technologique de l'Europe ne dépend pas tant de sa capacité à s'affranchir ou à rivaliser frontalement avec les *Big Tech* que de sa faculté à tracer sa propre voie. Pour cela, elle doit capitaliser sur ses forces distinctes tout en corrigeant ses faiblesses structurelles. C'est à cette condition que l'Europe pourra non seulement bénéficier des retombées économiques de l'IA générative, mais aussi devenir un acteur majeur du paysage technologique mondial.



27 rue de la Procession 75740 Paris cedex 15 – France

Ifri.org