

Un « moment DeepSeek » ?

Benjamin PAJOT

► Points clés

- DeepSeek, érigé en champion de l'IA chinoise, induit moins une révolution qu'une optimisation sensible de l'existant.
- Les doutes persistent quant aux chiffres avancés par la start-up et incitent à relativiser le battage médiatique d'un rattrapage technologique chinois.
- DeepSeek traduit néanmoins la nécessaire remise en cause d'un modèle économique exclusivement centré sur la course à la puissance de calcul.
- En misant sur l'innovation ouverte, l'Europe peut frayer son chemin dans une compétition qui n'a rien d'un jeu à somme nulle.

Introduction

La sortie du modèle d'Intelligence artificielle (IA) générative DeepSeek-R le 20 janvier dernier a créé un vent de panique plus ou moins sincère à travers la Silicon Valley. Affichant des performances supérieures à ChatGPT4 d'OpenAI, la petite start-up chinoise d'alors 143 employés venait confirmer ses capacités, entre-aperçues un mois auparavant avec la publication de DeepSeek v3 fin 2024. Créée et dirigée depuis 2023 par Liang Wenfeng, à la tête du fonds spéculatif High-Flyer dont DeepSeek est une filiale, l'entreprise a ainsi pu bénéficier de solides financements de départ, lui ayant permis de concentrer ses efforts initiaux dans la recherche plutôt que dans la quête de rentabilité. Ses liens avec les autorités chinoises demeurent mal connus et son émergence pourrait avoir été une bonne surprise pour Pékin, qui l'a érigée en champion national dans sa compétition technologique avec Washington.

Mais plutôt qu'un « moment Spoutnik », c'est bien un « moment DeepSeek » auquel font face les leaders américains d'un secteur en pleine recomposition, à l'heure où les modèles dits « de raisonnement » constituent une évolution substantielle des grands modèles de langage (LLM). DeepSeek-R1, inspiré de ses concurrents, serait également symptomatique de la popularisation des techniques de distillation, processus permettant d'extraire les capacités et les « connaissances » d'un autre modèle d'IA en s'entraînant à partir de ses réponses – et donc d'en proposer une déclinaison crédible. Librement accessible, le modèle traduit le positionnement stratégique de la Chine dans l'IA ouverte (*open source*) et questionne la place de l'Europe dans un jeu qui apparaît plus ouvert que figé.

Au-delà du battage médiatique

Un concurrent sérieux mais une panique boursière passagère

Les performances de DeepSeek-R1, soumis à une batterie de tests faisant référence, l'ont propulsé en tête des modèles ouverts (devant Llama3 développé par Meta) et comme concurrent direct des modèles fermés o1 d'OpenAI ou Claude 3.5 Sonnet d'Anthropic, alors considérés comme parmi les plus avancés du marché. Ces résultats, combinés aux chiffres communiqués par la start-up chinoise sur le faible coût de développement du modèle et ses gains d'efficacité majeurs, ont provoqué un choc aux États-Unis. Ils furent interprétés comme la double indication d'un rattrapage technologique chinois et d'une impasse économique à venir pour les acteurs américains du secteur. Il en a résulté une panique boursière d'une rare intensité, qui a vu le concepteur de puces NVIDIA perdre 17 % de sa valorisation (soit 600 milliards de dollars).

Si la question des coûts de développement demeure débattue, les gains d'efficacité semblent réels. Ils résident dans une méthode d'activation sélective : au lieu d'exécuter l'intégralité du modèle pour chaque requête, celui-ci est divisé en sous-modèles « experts » qui ne sont sollicités que lorsque la requête correspond à leur champ de compétence. Ces gains entraînent *a priori* une réduction des coûts de développement et d'exploitation des modèles, ouvrant la voie à leur plus grande accessibilité et donc commercialisation. Cette optimisation, pour habile qu'elle soit, n'en demeure pas moins répliquable, ce qui explique que les acteurs américains aient déployé beaucoup d'énergie à disséquer DeepSeek-R1.

La panique boursière a connu une rapide accalmie une fois les effets d'annonce dissipés, et les cercles financiers ont retrouvé leurs esprits pour prendre la mesure d'un bouleversement somme toute relatif. Car c'est moins l'agilité intellectuelle et la taille modeste de l'équipe chinoise que sa nationalité qui ont soulevé les inquiétudes. En témoigne la performance analogue de chercheurs américains – parvenus à publier en source ouverte un modèle de raisonnement aux performances comparables à ceux d'OpenAI et DeepSeek, pour moins de 50 dollars – quasi passée sous silence¹. Renvoyés à leur hantise du dépassement technologique, naguère russe puis japonais et aujourd'hui chinois, les États-Unis ont à nouveau semblé en proie au doute. Face à une entreprise adoubée par Xi Jinping pour incarner l'IA « à caractéristiques chinoises » – mélange d'orientations étatiques, d'ingéniosité du secteur privé et de bénéfices tirés de collaborations permises par l'*open source* –, l'Amérique semblait croire à une nouvelle course aux étoiles.

Renvoyés à leur hantise
du dépassement
technologique, les États-
Unis ont à nouveau
semblé en proie au doute

La facilité intellectuelle du « nouveau moment Spoutnik »

Nombreux ont été ceux qui se sont empressés de commenter la sortie du modèle chinois en la qualifiant de « nouveau moment Spoutnik² ». Pour tentante qu'elle puisse paraître, la comparaison avec la période de lancement du premier satellite soviétique en 1957 – qui avait choqué la communauté stratégique américaine et conduit à la création de la National Aeronautics and Space Administration (NASA) – semble être hors de proportion. DeepSeek-R1 ne relève pas d'une révolution technologique ou d'un nouveau paradigme technique mais plutôt une optimisation de l'existant. Il ne s'agit donc pas

1. M. Zeff, « Researchers Created an Open Rival to OpenAI's o1 "Reasoning" Model for under \$50 », *TechCrunch*, 5 février 2025.

2. S. Kauffmann, « Contrairement au "moment Spoutnik" en 1957, il n'y aura probablement pas de moment DeepSeek aux États-Unis », *Le Monde*, 29 janvier 2025.

d'une percée techno-scientifique majeure comme l'était le lancement du premier satellite mondial. Le leadership américain n'en est pas remis en cause, tandis que les moyens colossaux déjà mobilisés par les États-Unis dans l'IA illustrent leur engagement durable dans cette compétition.

La pseudo-surprise stratégique des acteurs américains face à une entreprise chinoise en capacité de concurrencer leur écosystème national rappelle plutôt celle qui avait semblé frapper la Silicon Valley lors du lancement par Huawei de smartphones intégrant des puces de haute qualité (atteignant la barre des 7 nanomètres³). Les séquences d'emballage médiatique, récurrentes aux États-Unis, sont à replacer dans leur historicité. Washington a ainsi régulièrement recours à la désignation, unilatérale ou non, d'un adversaire pour engager un rapport de force productif à même de stimuler la croissance de son secteur technologique.

Un grand bond en avant chinois ?

Consécutive à la sortie du modèle, l'intégration de DeepSeek à des plateformes chinoises majeures (WeChat, Baidu, Huawei Cloud) et ses partenariats avec divers acteurs industriels nationaux (une douzaine de constructeurs motos et automobiles et plusieurs grands ports) traduit une forme de frénésie patriotique au profit de la start-up, érigée en symbole d'une *remontada* face aux États-Unis⁴. Il ne faut toutefois pas minimiser la dynamique en cours, car la montée en puissance d'acteurs chinois semble réelle et profonde, en dépit des manœuvres américaines pour en limiter l'émergence⁵. D'autres modèles performants sont aujourd'hui développés par les principaux acteurs technologiques chinois : Qwen 2.5 (Alibaba), Ernie Bot (Baidu), Doubao 1.5 Pro (ByteDance) ou encore Kimi k1.5 (Moonshot AI) revendiquent ainsi des performances égales ou supérieures aux modèles américains sur certains segments⁶.

La montée en puissance d'acteurs chinois semble réelle

Cette tendance pourrait se voir confortée par la création d'un nouveau fonds national d'investissement de 8,2 milliards de dollars dédié au développement de l'IA chinoise⁷. Pékin déploie une politique active en matière de gouvernance internationale afin de renforcer son influence sur les normes et les cadres mondiaux en matière d'IA. Pour autant, la stimulation induite par la concurrence chinoise pourrait bénéficier *in fine* à l'écosystème américain, dont les capacités d'innovation se voient ainsi aiguillonnées.

3. Q. Liu, « How Huawei Surprised the US with a Cutting-edge Chip Made in China », *Financial Times*, 30 novembre 2023.

4. T. Westbrook et S. Shen, « Nicknames Bloom in China's Tech Renaissance », Reuters, 19 février 2025.

5. H. Booth, « How China Is Advancing in AI Despite U.S. Chip Restrictions », *Time*, 28 janvier 2025.

6. Z. Soo, « DeepSeek Has Rattled the AI Industry: Here's a Look at Other Chinese AI Models », *Time*, 28 janvier 2025.

7. B. Jiang, « Tech War: China Creates US\$8.2 Billion AI Investment Fund amid Tightened US Trade Controls », *South China Morning Post*, 20 janvier 2025.

De l'échec réel ou supposé du régime de contrôle des exportations américain

La sortie de DeepSeek-R1 a également suscité de nombreux commentaires quant à la politique américaine de contrôle des exportations à l'encontre de la Chine. Censées prévenir un rattrapage technologique chinois pour des raisons de sécurité nationale, les mesures américaines de restriction d'accès à l'innovation sont centrées sur les semi-conducteurs, en particulier les puces H100 commercialisées par NVIDIA, considérées comme les plus avancées du marché⁸. La capacité de DeepSeek à développer un modèle aussi évolué, supposément sans accès à ces puces, a entraîné dès lors plusieurs hypothèses. Soit la start-up chinoise dit vrai, et est donc parvenue à contourner les limitations techniques imposées par le régime, en optimisant son usage des puces H800 de rang inférieur. DeepSeek-R1 constituerait ainsi un cas d'espèce d'optimisation et d'adaptation face aux mesures de rétorsion américaines. Soit l'entreprise dissimule la possibilité qu'elle avait eue de stocker au préalable ou d'accéder en continu à des puces H100, ce qui signifierait que les réseaux de contrebande permettant leur acheminement en Chine sont plus efficaces qu'estimé. Dans les deux cas, l'efficacité du régime de contrôle des exportations se trouve cependant mise en cause.

Alors que l'une des dernières mesures de l'administration Biden a été d'œuvrer à son renforcement, bon nombre d'acteurs – à commencer par NVIDIA, pour qui le marché chinois est essentiel – tentent à présent d'en réduire la portée et mènent pour cela un lobbying intense auprès des autorités américaines⁹. Il faut toutefois garder à l'esprit que l'équation n'a rien d'un jeu à somme nulle : la conception de DeepSeek-R1 a probablement été stimulée par la volonté de trouver des solutions pratiques face à l'embargo américain, mais ce dernier a tout autant pu limiter la portée « révolutionnaire » du modèle chinois. Alors que la demande en puces NVIDIA continue de bondir en Chine, on pourrait tout aussi bien assister à un renforcement prochain du régime de contrôle des exportations, au nom du principe trumpien de l'*America First*¹⁰.

Des doutes persistent

Des interrogations sur le coût d'entraînement

Si l'enjeu du coût d'entraînement de DeepSeek-R1 a déchaîné les passions, c'est aussi parce qu'il paraissait sous-évalué. La start-up chinoise a opté pour une stratégie

8. M. Velliet, « From Non-proliferation to Strategic Competition: US Export Controls and China », *International Politics*, 6 janvier 2025.

9. T. Hunnicutt *et al.*, « Trump and Nvidia CEO Discuss DeepSeek, AI Chip Exports during Meeting, Source Says », Reuters, 1^{er} février 2025.

10. F. Potkin et C. Pan, « Exclusive: Nvidia's H20 Chip Orders Jump as Chinese Firms Adopt DeepSeek's AI Models, Sources Say », Reuters, 25 février 2025.

communicationnelle offensive, mentionnant un coût d'entraînement marginal (6 millions de dollars) ne prenant en compte ni le coût des infrastructures sous-jacentes – DeepSeek possédant ses propres centres de données – ni celui de la masse salariale ou de la recherche et développement (R&D) préalable. Les soupçons se sont vu renforcer par la publication d'une firme spécialisée, affirmant que la société chinoise aurait en réalité eut à sa disposition 10 000 puces H100. L'architecture technique de DeepSeek aurait ainsi plutôt nécessité 1,6 milliard de dollars d'investissements, ce qui constitue un ordre de grandeur plus plausible¹¹. Une telle affirmation, si elle était avérée, viendrait de fait réduire à néant une partie des avancées supposées de DeepSeek-R1 et remettrait en cause certaines considérations évoquées (sur le rattrapage chinois, la menace économique ou l'efficacité du régime d'*export control*). Mais elle ne balayerait pas tout : DeepSeek-R1 resterait un modèle *open source* déployable en local, moins gourmand en ressources et énergie, à l'appui d'une architecture composite, avec une approche de « raisonnement » novatrice.

Des inquiétudes sur les pratiques

DeepSeek-R1 a fait l'objet de fortes critiques quant à son ouverture et sa sécurité. Le modèle a bien été publié, mais ses données d'entraînement et son algorithme d'ajustement (*fine-tuning*, étape de spécialisation du modèle pré-entraîné) restent inaccessibles. Entraîné sans supervision humaine, il serait susceptible de renforcer les préjugés au lieu de les corriger. Le modèle ayant été développé à destination initiale d'un public chinois, il est aussi conforme aux attentes de Pékin en matière de propagande et de révisionnisme, comme l'ont démontré ses réponses concernant Tian'anmen, Taïwan ou les Ouïghours.

Sur le plan de la sécurité, l'application DeepSeek capterait des données personnelles usuelles (à l'image de TikTok ou Instagram), qui seraient hébergées en Chine, chez le géant ByteDance. Ceci peut déjà constituer un enjeu de sécurité en soi, dans la mesure où les entreprises chinoises sont soumises à la loi sur le renseignement de 2017 (renforcée en 2023) qui permet à Pékin d'accéder à leurs données. Par ailleurs, cette remontée de données vers les serveurs chinois ne s'effectuerait pas toujours de manière cryptée, et selon des experts, l'application manquerait des protections de base en vigueur¹². L'entreprise américaine de cybersécurité Wiz a détecté et signalé une importante fuite de données exposant les requêtes d'utilisateurs de DeepSeek et leurs informations personnelles¹³. Enfin, sur le plan cyber, DeepSeek-R1 serait à la fois

11. D. Patel *et al.*, « DeepSeek Debates: Chinese Leadership on Cost, True Training Cost, Closed Model Margin Impacts, H100 Pricing Soaring, Subsidized Inference Pricing, Export Controls, MLA », *Semianalysis*, 31 janvier 2025.

12. A. Hoog, « NowSecure Uncovers Multiple Security and Privacy Flaws in DeepSeek iOS Mobile App », *NowSecure*, 6 février 2025.

13. G. Nagali, « Wiz Research Uncovers Exposed DeepSeek Database Leaking Sensitive Information, Including Chat History », *Wiz*, 29 janvier 2025.

perméable à des attaques par *prompt* dans le but de générer du contenu toxique¹⁴, mais montrerait aussi des aptitudes en matière cyberoffensive, qui seraient dues à ses capacités de « raisonnement ». Sa vulnérabilité au *jailbreaking* (le contournement de ses garde-fous par des *prompts* pernecieux) en ferait un vecteur de développement de logiciels malveillants¹⁵. Ces inquiétudes ont dès lors nourri des velléités d'interdire l'application dans divers États (Australie, Corée du Sud, États-Unis, Italie, Taïwan).

Un pavé dans la mare économique

Une remise en cause de la stratégie du gigantisme

Le véritable choc qu'a constitué l'arrivée de DeepSeek-R1 tient au fait qu'il bouscule le modèle économique dominant du secteur. En réduisant les coûts d'inférence et surtout la puissance de calcul nécessaire pour entraîner et faire tourner son modèle, DeepSeek a remis en cause la stratégie du gigantisme poursuivie par les leaders américains du secteur. Celle-ci revenait à développer des modèles d'IA toujours plus volumineux, gourmands en données et puissance de calcul, et donc en puces et énergie. Ce gigantisme, qui permettait aussi d'ériger des barrières à l'entrée vis-à-vis de nouveaux entrants, était jusqu'ici financé essentiellement par des levées de fonds et des augmentations de capitalisation. Faute de concurrence, les acteurs dominants ne s'étaient pas préoccupés d'améliorer leur efficacité énergétique et économique. Ils voient à présent leur potentiel de rentabilité menacé, et leurs actionnaires craindre pour leurs retours sur investissement. Le fait que DeepSeek ait été publié en source ouverte crée aussi de potentielles opportunités pour le développement de services à moindre coût, ouvrant la voie à des acteurs aux moyens plus modestes, tout en forçant les leaders à se repositionner¹⁶. Alors que la distillation devient la norme et que la capacité de développeurs à dupliquer des modèles propriétaires pour les rendre ensuite disponibles en *open source* semble réelle, c'est tout un modèle économique qui se voit bousculé¹⁷. Ceci pourrait entraîner une réponse juridique de la part des acteurs propriétaires (OpenAI, Anthropic et Google au premier chef), avec le soutien des autorités américaines¹⁸.

L'arrivée de DeepSeek-R1 bouscule le modèle économique dominant du secteur

14. D. Goodin, « DeepSeek iOS App Sends Data Unencrypted to ByteDance-controlled Servers », *Ars Technica*, 6 février 2025.

15. N. Miles, « DeepSeek Deep Dive Part 1: Creating Malware, Including Keyloggers and Ransomware », *Tenable*, 13 mars 2025.

16. R. Waters, « Big Tech Is Moving on from the DeepSeek Shock », *Financial Times*, 14 février 2025.

17. B. Edwards, « Hugging Face Clones OpenAI's Deep Research in 24 Hours », *Ars Technica*, 5 février 2025.

18. M. Sweney et D. Milmo, « OpenAI "Reviewing" Allegations That Its AI Models Were Used to Make DeepSeek », *The Guardian*, 29 janvier 2025.

Les Big Tech à même de tirer leur épingle du jeu

Si DeepSeek a pu apparaître comme une mauvaise nouvelle pour les géants de la Silicon Valley, qui fournissent la majorité des infrastructures matérielles nécessaires à l'entraînement des IA, le tableau pourrait se révéler moins sombre qu'il n'en a l'air. Conformément au paradoxe de Jevons, ils pourraient directement bénéficier d'une adoption facilitée des modèles d'IA, qui viendrait *in fine* compenser la baisse de la demande matérielle consécutive aux gains d'efficacité générés par DeepSeek-R1¹⁹. Les besoins en calcul de haute intensité pourraient même croître proportionnellement à la commercialisation de l'IA générative. NVIDIA, principale cible de la panique boursière, a ainsi retrouvé des couleurs à mesure que cette hypothèse semblait s'imposer dans les milieux financiers. De fait, l'entreprise devrait continuer de dominer le secteur car elle peut s'adapter aux évolutions en cours (moindre coût d'inférence, puces aux besoins de plus en plus spécifiques des modèles de raisonnement) et répondre à une demande accrue. De même, Google,

Les géants de la Silicon Valley pourraient céder à la tentation d'exploiter la panique générée par DeepSeek

Meta et OpenAI disposent déjà de l'infrastructure et de l'expertise nécessaires pour intégrer les techniques d'optimisation employées par DeepSeek, et donc s'adapter.

Par ailleurs, des acteurs comme Microsoft, Amazon ou Apple, moins avancés jusqu'ici dans le champ de l'IA générative, profiteront eux aussi à la fois des moindres coûts de développement, et des possibilités offertes par la disponibilité des modèles en source ouverte. Apple – qui conçoit ses propres puces à mémoire unifiée facilitant l'exécution de modèles plus légers sur ses produits – pourrait même revenir sur le devant de la scène par le biais de l'intégration de services d'IA pour le grand public²⁰. S'ils devraient donc n'être que mollement bousculés, les géants de la Silicon Valley pourraient toutefois céder à la tentation d'exploiter la panique générée par DeepSeek aux États-Unis. Que ce soit en prônant la dérégulation maximale du secteur ou en appelant à sa consolidation sous leur aile et à leur bénéfice exclusif, leur activisme potentiel fait l'objet de craintes légitimes²¹.

Vers un éclatement de la bulle de l'IA ?

Cette remise en cause de la stratégie du gigantisme questionne indirectement la rationalité des dépenses effectuées jusqu'ici par les plus gros acteurs de l'IA. Dans la

19. Du nom d'un économiste britannique du XIX^e siècle, ce paradoxe lie optimisation technique et consommation de ressources additionnelles.

20. E. Sayegh, « DeepSeek Hype Vs. Nvidia And Apple: Will Calmer Minds Prevail? », *Forbes*, 28 janvier 2025.

21. T. Wheeler, « DeepSeek Is Not a Good Reason for Big Tech to Become More Powerful », *Brookings*, 11 février 2025 ;

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

mesure où ces entreprises sont très déficitaires, les doutes sur la viabilité économique du secteur en sortent renforcés²². À mesure que la miniaturisation des semi-conducteurs devient plus difficile, que les LLM – de l’aveu même de leurs concepteurs – atteignent leurs limites et que leur entraînement sur des données synthétiques ou parcellaires crée toujours plus de difficultés (hallucinations, auto-intoxication, inexplicabilité totale), les inquiétudes sur l’éclatement potentiel de la bulle de l’IA ressurgissent. Et ce d’autant plus que la consommation électrique croissante des IA – que les gains d’efficacité seront en peine de compenser – induit non seulement une augmentation des coûts mais pourrait également limiter l’exploitation jusqu’à 40 % des centres de données, faute d’énergie disponible²³. À cet égard, le plan « Stargate » annoncé par la Maison-Blanche risque d’aggraver la situation et de se trouver rapidement contraint par la réalité des ressources disponibles, tant la tension entre les différents besoins énergétiques semble déjà forte dans les territoires concernés²⁴. Pour autant, l’éclatement de la bulle n’est pas garanti, car les investissements massifs annoncés aux États-Unis, en Europe ou même en France, combinés à ceux du capital-risque, devraient contribuer à maintenir le secteur à flot après une année 2024 record²⁵.

Des leçons à tirer pour l’Europe

En finir avec la vision du « gagnant raflant la mise »

Cette séquence est venue rappeler avec force que la « course à l’IA » est un phénomène mondial n’ayant rien d’un tableau figé, et encore moins d’une table de casino où le gagnant serait seul à remporter la mise. Cette vision du « *winner takes all* », prépondérante aux États-Unis et en Chine, est en réalité assise à dessein, afin de décourager toute concurrence une fois un écart technique creusé. Or DeepSeek prouve à nouveau que les statuts d’innovateurs dominants n’ont rien de pérennes : la possibilité d’innover n’est pas réductible à la seule puissance capitaliste ou à la hiérarchie boursière, et les pistes d’évolution de l’IA demeurent nombreuses. Son développement ne se résumant plus à la seule accumulation de puissance de calcul, mais prenant à présent en compte les enjeux d’efficacité et d’accessibilité, les acteurs européens ont davantage d’opportunités de concurrencer leurs homologues américains et chinois. Nombre d’entre eux sont en effet investis dans la conception de modèles de petite taille et spécialisés : les économies d’échelle permises par l’optimisation pourraient donc leur être favorables.

22. C. Weinberg, « OpenAI Projections Imply Losses Tripling to \$14 Billion in 2026 », *The Information*, 9 octobre 2024 ; D. Moyo, « Three Reasons Why AI’s Momentum Could Stall in 2025 », Project Syndicate, 23 janvier 2025.

23. « Gartner Predicts Power Shortages Will Restrict 40% of AI Data Centers By 2027 », Gartner, 12 novembre 2024.

24. C. Albon, « Power Generation Challenges Could Overshadow Stargate AI Initiative », *Defense News*, 24 janvier 2025.

25. I. Lunden, « AI Investments Surged 62% to \$110B in 2024 While Startup Funding Overall Declined 12% », TechCrunch, 11 février 2025.

Il serait toutefois illusoire de considérer que les investissements dans les infrastructures matérielles ne constituent plus une priorité. En dépit des gains d'efficacité, la dépendance du secteur à l'égard des semi-conducteurs avancés demeure inchangée. Si la distillation autorise la déclinaison à peu de frais de modèles plus onéreux, elle ne permet pas jusqu'ici de repousser les limites de l'innovation en créant des modèles bien supérieurs à ceux déjà présents sur le marché. DeepSeek a rappelé en creux tout l'avantage de disposer de ses propres centres de données. Les annonces lors du Sommet de Paris de plans français et européen de développement des infrastructures tentent d'apporter une réponse, même si celle-ci reste questionnable sur les plans financier, énergétique et écologique. Le risque est de tomber dans les mêmes écueils du gigantisme qu'aux États-Unis, alors que l'heure est plutôt à la mutualisation des capacités de calcul et de recherche. Mais ceci suppose une forme de planification industrielle qui fasse fi des nationalismes technologiques au sein de l'Union européenne (UE), une perspective qui semble encore lointaine à ce jour.

Miser davantage sur l'innovation ouverte

Dans la lignée d'une politique nationale visant à s'appuyer sur l'*open source* pour renforcer l'autarcie technologique de la Chine, il semble que les acteurs chinois œuvrent aujourd'hui au développement de l'IA en source ouverte à des fins stratégiques²⁶. L'enjeu est d'attirer les talents tout autant que de faciliter l'adoption de modèles chinois, autorisant la captation de données et d'investissements, mais aussi l'exercice d'un *soft power* en ingénierie, notamment à destination des pays émergents, dans la continuité de la Belt and Road Initiative²⁷. Cette politique vise à prendre le contrepied de l'écosystème américain propriétaire et à contourner les barrières à l'innovation établies par celui-ci. Invitant à considérer l'IA comme une ressource nationale voire un bien public plutôt qu'un actif d'entreprise, une telle stratégie ferait sens en Europe.

DeepSeek se situe aujourd'hui là où des acteurs européens auraient dû être

DeepSeek se situe aujourd'hui là où des acteurs européens auraient dû être : berceau historique et contributrice majeure de l'*open source*, l'Europe est elle aussi en capacité de déployer des innovations de cette ampleur. Là où l'*open source* « à caractéristiques chinoises » demeure étroitement délimitée – le gouvernement encourageant la collaboration dans le cadre d'un système contrôlé où les entreprises soutenues par l'État et les géants de la technologie mènent la danse – l'Europe peut et doit faire valoir une autre approche, fondée sur ses valeurs et principes.

26. A. Panier, « Sources d'influence. Enjeux économiques et géopolitiques des logiciels open source », *Études de l'Ifri*, Ifri, 12 décembre 2022.

27. P. Bloom, « DeepSeek: How China's Embrace of Open-source AI Caused a Geopolitical Earthquake », *The Conversation*, 12 février 2025.

Dans cette perspective, le Sommet de Paris sur l'IA a permis d'identifier des pistes intéressantes, qui nécessitent d'être explorées plus avant. Le renforcement de l'initiative Open Euro LLM et la création de la fondation CurrentAI, dédiées au financement et au développement d'IA *open source* et frugales, constituent des instruments prometteurs pour encourager l'innovation ouverte. Elles invitent à repenser l'IA comme un bien collectif, à caractère public, en lien étroit avec les communs numériques.

Explorer d'autres voies

Le plus grand apport de DeepSeek est de rouvrir le champ des possibles. Plusieurs pistes sont aujourd'hui à envisager pour espérer peser sur de futurs développements de l'IA. Sur le plan matériel, des puces photoniques (utilisant la lumière plutôt que l'électricité pour la transmission des données) aux processeurs neuromorphiques (dont la structure s'inspire du fonctionnement neuronal), diverses innovations pourraient réduire la consommation d'énergie des IA. Des investissements dans la R&D de matériaux alternatifs pour les semi-conducteurs et des progrès dans les pratiques de recyclage de métaux et terres rares pourraient réduire leur impact environnemental. Enfin, la conception de centres de données à échelle locale et leur meilleure répartition à travers les territoires – pour réutiliser la chaleur qu'ils produisent et pondérer les ressources hydriques nécessaires à leur refroidissement – constituent une autre piste.

Le plus grand apport de DeepSeek est de rouvrir le champ des possibles

Sur le plan logiciel, des optimisations de techniques de calcul comme la « multiplication à complexité linéaire » (L-Mul), promettent d'importants gains d'efficacité énergétique. Alors que les modèles « de raisonnement » donnent un second souffle à une économie des LLM en perte de vitesse, certains acteurs, tels Meta, travaillent à concevoir de « grands modèles conceptuels » (*Large Concept Model* ou LCM). Censés disposer de capacités d'abstraction, ils ouvrent la voie à une compréhension contextuelle indépendante de la langue utilisée pour les entraîner ou interagir avec eux.

Conclusion

Ni anecdotique, ni moment Spoutnik, le « moment DeepSeek » constitue plutôt une invitation pour l'Europe à prendre sa part au jeu qui se déroule sous ses yeux, pour y affirmer davantage sa place. Car derrière la mise en scène de leur rivalité, la Chine et les États-Unis ont en partage une stratégie de développement de l'IA moins motivée par l'innovation et la préservation de l'intérêt général que par le contrôle économique et politique. En dépit des optimisations qu'il a entreprises, DeepSeek-R1 illustre ainsi des choix d'ingénierie privilégiant encore l'accroissement de capacités (échelle, vitesse, efficacité) à la bonne gestion des impacts sociétaux de l'IA.

Si la sortie du modèle chinois questionne la viabilité économique de tout un secteur, il reste beaucoup à faire pour mettre l'IA au service de l'intérêt général. À charge pour l'Europe de prendre acte des mutations en cours pour tenter d'orienter le développement de l'IA et d'exploiter le potentiel de ces technologies. Elle peut pour cela compter sur la richesse de son propre écosystème et sur l'innovation ouverte, gage de l'émergence de l'IA à « caractéristiques européennes ».

Benjamin Pajot est chercheur associé au Centre géopolitique des technologies de l'Ifri et chercheur indépendant en géopolitique numérique. Il travaillait précédemment comme chargé de mission sur les enjeux numériques et cyber au Centre d'analyse, de prévision et de stratégie (CAPS) du ministère de l'Europe et des Affaires étrangères.

Comment citer cette publication :

Benjamin Pajot, « Un “moment DeepSeek” ? », *Briefings de l'Ifri*, Ifri, 26 mars 2025.

ISBN : 979-10-373-1024-8

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

© Tous droits réservés, Ifri, 2025

Couverture : © Algi Febri Sugita/Shutterstock.com



27 rue de la Procession
75740 Paris cedex 15 – France

lfri.org

