

Le lancement de DeepSeek peut-il accélérer la croissance des entreprises de logiciels cloud ?

Publié le 12 février 2025

Baoqi Zhu

Senior Associate, Quantitative Research and Multi Asset Solutions

Elvira Kuramshina

Associate Director, Quantitative Research

- Les techniques de « Mixture of Experts » et d'apprentissage par renforcement pur (Reinforcement Learning, RL) de DeepSeek améliorent l'efficacité, en réduisant les coûts de calcul sans compromettre les capacités.
- L'« effet poisson-chat » qui en résulte contraint les fournisseurs d'intelligence artificielle (IA) à réduire leurs prix, ce qui rend la technologie plus abordable pour les entreprises de logiciels en tant que service (SaaS).
- L'intégration de DeepSeek R1 aux principales plateformes de services cloud atténue les préoccupations en matière de confidentialité et de conformité réglementaire, ce qui permet d'accélérer l'adoption de l'IA au sein des entreprises de logiciels.
- Les entreprises de logiciels cloud devraient bénéficier des avancées de l'IA grâce à la différenciation de leurs produits, à l'optimisation des coûts, ainsi qu'à la création de nouvelles sources de revenus.
- L'évolution vers l'IA agentique et verticale révolutionne le secteur des logiciels cloud, en favorisant le développement de systèmes autonomes et d'applications d'IA spécifiques au secteur, qui offrent un avantage concurrentiel.

DeepSeek fait la une des journaux depuis la veille de l'année du serpent (24/01/2025), tandis que les marchés financiers réagissent de manière marquée et contrastée. En parallèle, les discussions relatives à la confidentialité des données, à la propriété intellectuelle ainsi qu'aux tensions géopolitiques se sont intensifiées. Nous souhaitons néanmoins prendre du recul par rapport à ces débats et nous concentrer sur les avancées technologiques des modèles de DeepSeek, ainsi que sur leur impact potentiel sur le secteur des logiciels.

Quelles innovations DeepSeek a-t-il apportées ?

Reposant sur l'architecture Transformer, le modèle d'intelligence artificielle (IA) de DeepSeek intègre plusieurs améliorations en matière de conception et d'ingénierie afin de renforcer son efficacité. Plutôt que de recourir à des méthodes entièrement nouvelles, DeepSeek combine des innovations existantes et

répond aux défis pratiques, tels que la puissance de calcul limitée, afin d'associer avancées théoriques et exécution technique rigoureuse.

Nous considérons notamment que les modèles DeepSeek, DeepSeek V3 et DeepSeek R1 Zero intègrent deux avancées majeures :

Une architecture « Mixture of Experts » (MOE) associée à une stratégie sans perte auxiliaire :

DeepSeek-V3 divise le modèle en plusieurs modules « experts » pour répartir efficacement la charge de calcul. Pendant l'entraînement, différents experts sont répartis sur différents GPU, améliorant ainsi l'efficacité. Au cours de l'inférence, seulement 37 milliards de paramètres actifs sont sollicités par jeton, contre 671 milliards au total, ce qui permet de réduire de manière significative les coûts de calcul. L'architecture MoE rencontre régulièrement des difficultés liées à un déséquilibre de charge entre les modules experts ; DeepSeek résout cette problématique en utilisant un routage basé sur la pertinence plutôt que des fonctions de perte auxiliaires, garantissant la préservation de tous les jetons au cours de l'entraînement et de l'inférence.

Un apprentissage par renforcement sans rétroaction humaine :

DeepSeek-R1 Zero est un modèle de raisonnement similaire à GPT-o1, qui démontre que l'apprentissage par renforcement pur (Reinforcement Learning, RL) peut développer des capacités de raisonnement sans s'appuyer sur l'apprentissage par renforcement basé sur les commentaires humains (Reinforcement Learning from Human Feedback, RLHF). La plupart des grands modèles de langage (Large Language Models, LLM) reposent sur l'étiquetage et les commentaires humains via le réglage fin supervisé (supervised fine-tuning, SFT) et le RLHF. Or, DeepSeek R1 Zero a choisi de remplacer le SFT et le RLHF par un apprentissage par renforcement pur, sans rétroaction humaine.

Le RL traditionnel présente un coût de calcul élevé et une évolutivité limitée face à des tâches variées. En introduisant une méthode innovante dénommée « optimisation de la politique relative au groupe » (Group Relative Policy Optimisation, GRPO), DeepSeek R1 Zero répond à ces défis. De plus, les progrès de DeepSeek suggèrent qu'une méthode de RL pur pourrait constituer une voie alternative pour perfectionner en continu les modèles d'IA, étant donné que les données de haute qualité générées par les humains sont limitées et peuvent finir par s'épuiser. Cela renforce la durabilité des améliorations de la capacité du modèle.

Impact sur les entreprises de logiciels

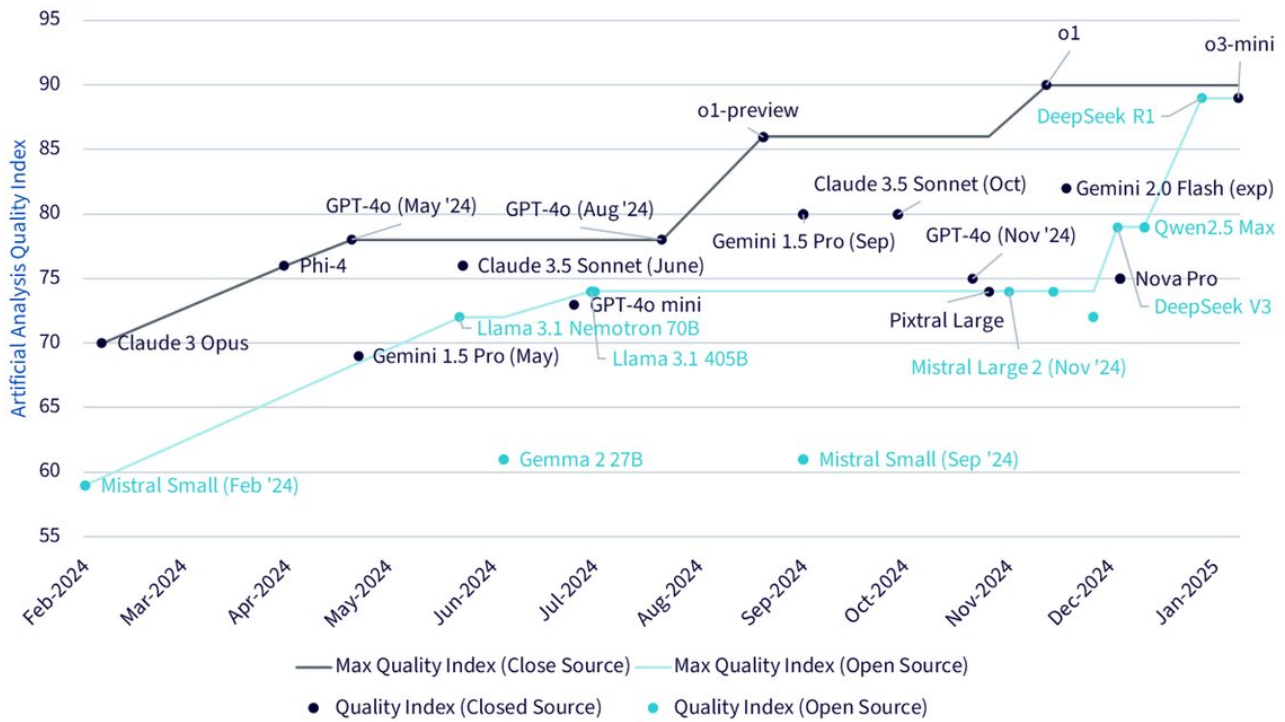
Depuis la sortie de ChatGPT, les investisseurs en logiciels s'interrogent sur une question cruciale : dans quelle mesure l'IA renforce-t-elle la valeur des entreprises SaaS ? Les progrès de l'IA ont suscité des inquiétudes quant à la pérennité du secteur des logiciels, notamment sur la possibilité que cette technologie remplace un jour les logiciels traditionnels. En outre, l'ajout de modèles d'IA dans les offres SaaS engendre de nouveaux coûts d'utilisation, faisant craindre une pression sur les marges. Les

évolutions récentes laissent toutefois entrevoir des perspectives plus favorables, ce qui atténue certaines de ces préoccupations.

Parmi les principaux facteurs de ce changement figure DeepSeek, dont les modèles en source ouverte rivalisent avec les performances des grands modèles de langage les plus avancés, tout en offrant un coût d'utilisation réduit. Cet avantage économique contribue à réduire les obstacles auxquels les entreprises SaaS sont confrontées lorsqu'il s'agit d'ajouter des fonctionnalités d'IA, en leur permettant d'intégrer la technologie dans les entreprises tout en maîtrisant les coûts. En outre, les principales plateformes cloud, notamment Microsoft Azure, AWS et Google Cloud, ont d'ores et déjà ajouté DeepSeek R1 à leurs services. Son déploiement au-delà des frontières de la Chine répond aux préoccupations en matière de confidentialité et de réglementation. Plusieurs entreprises américaines ont adopté DeepSeek. Perplexity, spécialisée dans la fourniture de services de questions-réponses basés sur l'IA, a intégré le modèle DeepSeek R1 dans son moteur de recherche. Cerebras et Groq se sont également lancées dans l'aventure. Dans le même temps, ZoomInfo a fait part de son intérêt pour l'achat d'un accès à DeepSeek R1 via les fournisseurs américains de plateformes cloud².

Les prix compétitifs de DeepSeek pourraient déclencher un « effet poisson-chat », incitant d'autres fournisseurs à publier des modèles plus rentables et moins coûteux. À titre d'exemple, OpenAI a déjà lancé GPT-o3 mini à un prix très attractif, avec un coût de production par million de jetons de seulement 4,4 dollars, soit un tarif inférieur d'environ 63 % à celui de GPT-o1 mini. Cette concurrence profite aux entreprises de logiciels, dans la mesure où elle leur permet d'adopter l'IA tout en préservant leurs marges bénéficiaires. En outre, DeepSeek a partagé ses méthodes d'optimisation, qui contribuent au renforcement des communautés en source ouverte, et qui pourraient combler l'écart entre les modèles à code source ouvert et à code source fermé. Au fil du temps, cette ouverture devrait favoriser la concurrence et contenir les coûts de l'IA à un niveau raisonnable.

Illustration 1 : L'écart de performance entre les modèles à code source fermé et ouvert se réduit



Source : Artificial Analysis. L'indice Artificial Analysis Quality correspond au résultat moyen de nos évaluations couvrant plusieurs paramètres de l'intelligence des modèles. Il inclut actuellement MMLU, GPQA, Math et HumanEval. Les chiffres du modèle OpenAI o1 sont préliminaires et se fondent sur les chiffres fournis par OpenAI. Données du 31 janvier 2025. Les performances historiques ne garantissent pas les performances futures, et tout investissement est susceptible de perdre de la valeur.

Illustration 2 : Les API de DeepSeek sont relativement abordables par rapport aux principaux modèles d'IA

Company	Model	Input Cost Per 1M Tokens (\$)	Output Cost Per 1M Tokens (\$)	Artificial Analysis Quality Index
DeepSeek	DeepSeek R1	0.55	2.19	89.00
DeepSeek	DeepSeek V3	0.27	1.10	79.00
OpenAI	GPT-o1	15.00	60.00	90.00
OpenAI	GPT-o3 mini	1.10	4.40	89.00
OpenAI	GPT-o1 mini	3.00	12.00	84.00
OpenAI	GPT-4o	2.50	10.00	75.00
Anthropic	Claude 3.5 Sonnet	3.00	15.00	80.00
Google	Gemini 2.0	NA	NA	82.00
Google	Gemini 1.5 Pro	1.25	5.00	80.00

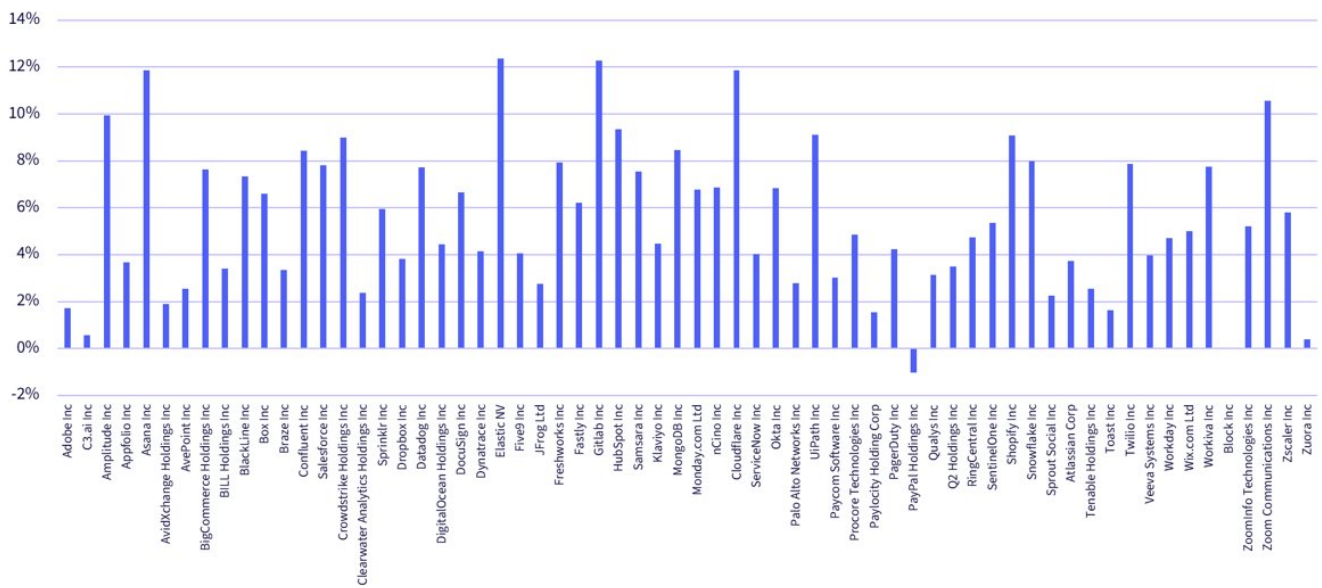
Source : Bloomberg, DeepSeek, OpenAI, Anthropic, Google, Artificial Analysis. Les prix des API sont fournis par les propriétaires de modèles et peuvent être différents de ceux proposés par des fournisseurs tiers.

L'indice Artificial Analysis Quality correspond aux résultats moyens de nos évaluations couvrant plusieurs paramètres de l'intelligence des modèles. Il inclut actuellement MMLU, GPQA, Math et HumanEval. Les chiffres du modèle OpenAI o1 sont préliminaires et se fondent sur les chiffres fournis par OpenAI. Données du 31 janvier 2025. **Les performances historiques ne garantissent pas les performances futures, et tout investissement est susceptible de perdre de la valeur.**

Les logiciels cloud peuvent-ils surfer sur la vague DeepSeek ?

L'indice BVP Nasdaq Emerging Cloud sélectionne plusieurs opportunités pure play dans le secteur du cloud computing, en se concentrant sur celles qui surperforment le marché global en termes de croissance du chiffre d'affaires. Nous avons utilisé cet indice pour mesurer le sentiment des investisseurs et avons évalué les réactions du marché à la sortie de DeepSeek les 27 et 28 janvier. Hormis PayPal, toutes les entreprises ont réagi positivement, avec un rendement médian de 4,9 % sur la période (voir illustration 3). Dans la mesure où aucun rapport sur les bénéfices n'a été publié pendant cet intervalle de temps, nous attribuons principalement cette réaction à l'annonce de DeepSeek. Par ailleurs, la réunion du Comité fédéral des marchés ouverts (Federal Open Market Committee, FOMC) qui s'est déroulée du 28 au 29 janvier a modéré l'optimisme en fin de semaine, faisant chuté le rendement médian sur cinq jours à 2,7 %.

Illustration 3 : Rendements des entreprises de cloud computing suite à la sortie de DeepSeek



Au cours des deux dernières années, les entreprises de logiciels cloud n'ont pas été considérées comme des bénéficiaires directs de l'IA, contrairement aux fournisseurs d'infrastructures d'IA et aux développeurs de modèles. Elles étaient perçues comme agissant en aval de la révolution de l'IA, ce qui les a contraintes à intégrer l'IA dans leurs produits avant d'en extraire de la valeur. En outre, des inquiétudes ont été soulevées quant à l'impact des coûts liés à l'IA sur les marges, en particulier dans un contexte où le pouvoir de fixation des prix est faible en raison des vents contraires macroéconomiques et des budgets limités consacrés aux logiciels.

L'ère de modèles d'IA plus abordables qu'annonce le lancement de DeepSeek pourrait marquer un tournant et permettre aux entreprises de logiciels cloud de se différencier, d'optimiser leurs coûts et d'explorer de nouvelles sources de revenus. Les tendances récentes révèlent un changement stratégique dans l'univers des logiciels cloud, les entreprises se concentrant de plus en plus sur l'IA agentic (systèmes autonomes qui effectuent des tâches avec une intervention humaine minimale) et l'IA verticale (modèles d'IA hautement spécialisés dans un secteur ou un domaine spécifique, tel que la santé, le droit ou la finance). L'émergence d'entreprises cloud natives de l'IA souligne encore davantage cette évolution, qui permet aux fournisseurs de cloud de profiter plus directement des avancées de cette technologie.

Le lancement de DeepSeek, qui a déclenché une nouvelle vague d'innovations dans le domaine de l'IA, devrait en soutenir l'adoption. Face à cette accélération de l'adoption de l'IA, les entreprises se préparent à investir davantage dans les logiciels pour répondre à la hausse de la demande d'utilisation de l'intelligence artificielle dans tous les domaines. À plus court terme, une politique monétaire américaine plus stricte est susceptible de constituer un vent contraire pour les fournisseurs de logiciels cloud. Les résultats publiés en février et au début du mois de mars pourraient néanmoins donner un aperçu des projets des entreprises visant à tirer parti de cette transformation du secteur portée par l'IA.

1 Dans la version suivante, DeepSeek-R1, le SFT a été appliqué à l'entraînement du modèle, mais son objectif consistait à améliorer la lisibilité et la cohérence des réponses. Les capacités de raisonnement et de prise de décision du modèle ont été principalement affinées grâce à l'apprentissage par renforcement (RL).

2 The Information: <https://www.theinformation.com/articles/deepseek-attracts-surge-of-business-users>.

Important Risks Related to this Article

Informations importantes

Communications commerciales publiées dans l'EEE Ce document est publié et approuvé par WisdomTree Ireland Limited, une société autorisée et réglementée par la Central Bank of Ireland.

Communications commerciales émises dans des juridictions en dehors de l'EEE Ce document est publié et approuvé par WisdomTree UK Limited, une société autorisée et réglementée par la Financial Conduct Authority du Royaume-Uni.

WisdomTree Ireland Limited et WisdomTree UK Limited sont toutes les deux désignées comme « WisdomTree » (le cas échéant). Notre Politique sur les conflits d'intérêts et notre Inventaire sont disponibles sur demande.

Réservé aux clients professionnels uniquement. Les informations figurant dans ce document sont fournies à titre informatif et ne constituent pas une ore de vente, ou une sollicitation d'achat de titres ou d'actions. Ce document ne doit pas être utilisé comme fondement d'une décision d'investissement. La valeur des investissements peut fluctuer et vous êtes susceptible de perte tout ou partie du montant investi. La performance passée ne constitue pas nécessairement une indication des performances futures. Toute décision d'investissement doit être fondée sur les informations figurant dans le prospectus approprié et sur des conseils indépendants en matière d'investissement, fiscaux et juridiques.

L'application des réglementations et lois fiscales peut souvent conduire à des interprétations diérentes. Tous les points de vue ou opinions exprimés dans cette communication représentent les points de vue de WisdomTree et ne doivent pas être interprétés comme des conseils réglementaires, fiscaux ou juridiques. WisdomTree ne donne aucune garantie ou représentation quant à l'exactitude des vues ou opinions exprimées dans cette communication. Toute décision d'investissement doit être fondée sur les informations contenues dans le prospectus approprié et après avoir sollicité des conseils indépendants en matière d'investissement, fiscaux et juridiques. Ce document n'est pas et ne doit en aucun cas être interprété comme une publicité ou une ore publique d'actions ou de titres aux États-Unis ou dans toute province ou tout territoire des États-Unis. L'introduction, la transmission et la distribution (directes ou indirectes) de l'original ou d'une copie de ce document sont interdites aux États-Unis.

Bien que WisdomTree s'efforce d'assurer l'exactitude du contenu de ce document, WisdomTree ne peut en garantir l'exactitude. Les fournisseurs de données tiers sollicités pour obtenir les informations contenues dans le présent document ne donnent aucune garantie ou représentation de quelque sorte en rapport avec ces données. Lorsque WisdomTree exprime ses propres opinions concernant le produit ou l'activité du marché, ces opinions sont susceptibles de changer. WisdomTree, ses alliés et leurs dirigeants, directeurs, partenaires ou employés respectifs déclinent toute responsabilité pour toute perte directe ou indirecte découlant de l'utilisation de ce document ou de son contenu.