

Le jour où une entreprise spatiale est devenue le troisième pôle de l'IA mondiale

20 août 2026 - Denis Favre

Rachat de Cursor à 60 milliards de dollars, Grok 4.6 à la frontière, Colossus loué au concurrent direct : anatomie d'un basculement industriel et de ses conséquences économiques et géopolitiques.

Le 14 août 2026, une entreprise de fusées a acheté un éditeur de texte pour 60 milliards de dollars. Formulée ainsi, la phrase paraît absurde. Elle décrit pourtant l'opération la plus lourde de l'année dans l'économie numérique, et sans doute l'un des basculements industriels les plus structurants de la décennie.

Ce dossier essaie de répondre à quatre questions. Que s'est-il réellement passé ? Qui est le mieux placé pour capter la valeur ? Qu'est-ce que cela change aux rapports de force entre les États-Unis, la Chine, la Russie et l'Europe ? Et pourquoi ces questions n'en forment-elles en réalité qu'une seule ?

Les chiffres à retenir

60 milliards de dollars, intégralement en actions : le prix payé par SpaceX pour Anysphere, éditeur de Cursor. Environ 391 millions d'actions de classe A émises. Le plus gros rachat de start-up de l'histoire.

1,25 milliard de dollars par mois : ce qu'Anthropic verse à SpaceX jusqu'en mai 2029 pour le calcul des grappes Colossus. Environ 15 milliards par an.

2,17 milliards de dollars par mois : le total facturé par SpaceX à ses deux clients calcul, Anthropic et Google.

61 points sur l'indice Artificial Analysis pour Grok 4.6, contre 63 pour Claude Opus 5 et 62 pour Claude Fable 5. Deux points d'écart entre le premier et le troisième mondial.

300 mégawatts : la puissance électrique d'une seule grappe Colossus. L'équivalent de la consommation d'une ville moyenne.

Environ 1 900 milliards de dollars : la capitalisation boursière de SpaceX (Nasdaq : SPCX), cotée depuis le 12 juin 2026, avec un résultat par action encore négatif.

1. Ce qui s'est réellement passé

La chronologie compte, parce qu'elle montre une opération construite, pas une impulsion.

En avril 2026, SpaceX et Anysphere — la société derrière Cursor — signent un partenariat technologique. Cursor obtient l'accès à la capacité de calcul de la division IA de SpaceX ; SpaceX obtient une option d'achat sur la totalité de la société, valorisée 60 milliards de dollars, payable en actions.

Le 12 juin 2026, SpaceX entre au Nasdaq sous le code SPCX, au terme de la plus grosse introduction en bourse jamais réalisée : environ 75 milliards de dollars levés, une valorisation d'introduction de l'ordre de 1 750 milliards, et un premier jour de cotation nettement au-dessus du prix d'offre. L'entreprise dispose désormais d'une monnaie d'échange cotée.

Le 14 août 2026, l'opération devient effective, confirmée par un formulaire 8-K auprès de la SEC. Cursor devient une filiale à 100 % d'une division rebaptisée SpaceXAI. La marque Cursor est appelée à s'effacer progressivement derrière celle de Grok. Les quatre cofondateurs d'Anysphere ressortent avec environ 2,7 milliards de dollars chacun, et la société affichait un chiffre d'affaires annualisé d'environ 2,6 milliards au moment du rachat.

Une précision utile, parce qu'elle circule mal : ce n'est ni Grok, ni X, ni xAI qui a racheté Cursor. C'est SpaceX, société cotée, qui avait absorbé xAI plus tôt dans l'année. L'ensemble opère sous le nom de SpaceXAI. La nuance n'est pas cosmétique : elle signifie que l'opération est financée par des actionnaires publics et adossée à un bilan d'entreprise spatiale.

Deux sorties de modèles encadrent l'opération. Le 7 août, Imagine Image 2.0, modèle d'image centré sur l'édition ciblée — baguette magique par zone, segmentation, jusqu'à cinq images de référence — se classe deuxième mondial sur les deux arènes publiques, texte-vers-image et édition, derrière GPT-Image-2 d'OpenAI. Le 12 août, Grok 4.6 atteint 61 points sur l'indice de référence d'Artificial Analysis, à égalité avec GPT-5.6 Sol, derrière les deux modèles d'Anthropic. Ce n'est pas un entraînement parti de zéro : c'est une itération sur la même base que Grok 4.5, sortie cinq semaines plus tôt, avec un entraînement supplémentaire nourri de données d'ingénierie de meilleure qualité.

D'où venaient ces données ? La question contient toute la logique du rachat.

2. Pourquoi 60 milliards pour un éditeur de code

Un éditeur de code, vu de loin, ressemble à un logiciel de confort pour développeurs. Vu de près, c'est une machine à produire un type de données que rien d'autre ne produit.

Quand un développeur travaille dans un environnement assisté par IA, il ne laisse pas seulement du code derrière lui. Il laisse la trajectoire complète du travail : la demande initiale, la proposition du modèle, le rejet, la correction, le test qui échoue, la reformulation, la validation finale. C'est exactement la matière dont on a besoin pour entraîner des systèmes agentiques, c'est-à-dire des modèles qui exécutent des tâches longues au lieu de répondre à des questions. Ce signal ne s'achète pas et ne s'aspire pas sur le web ouvert : il se produit, en continu, par des utilisateurs réels.

Le déséquilibre était donc parfaitement complémentaire. Cursor possédait des milliards de lignes d'interaction entre de vrais professionnels et du code, sans posséder un seul centre de calcul. SpaceXAI possédait Colossus, l'un des plus grands parcs d'accélérateurs du monde, sans avoir d'utilisateurs qui produisent ce signal.

C'est ce qu'on appelle une boucle : les développeurs utilisent l'outil, l'usage génère des données, les données entraînent le modèle suivant, le modèle attire plus d'utilisateurs, les revenus financent le calcul. Anthropic a compris cette mécanique le premier avec Claude Code, OpenAI l'a répliquée avec Codex. SpaceXAI vient de l'acheter au lieu de la construire. En 2026, cinq semaines séparent deux générations de modèles : construire une boucle prendrait deux ans. La racheter prend un trimestre.

Reste l'objection évidente : pourquoi le code justifierait-il un tel prix ? Parce que le code n'est plus une spécialité, c'est le substrat d'exécution universel des agents. Une présentation, un tableau de bord, une analyse financière, une campagne d'emails, un rapport réglementaire : tout cela est aujourd'hui produit par du code que l'utilisateur ne voit jamais. Celui qui contrôle la surface où le code se rédige contrôle la couche agentique — c'est-à-dire l'automatisation de l'essentiel du travail de bureau.

3. L'éléphant blanc de Memphis, et son retournement

Il faut revenir un an en arrière pour mesurer l'ampleur du retournement. Colossus, la grappe de Memphis, avait été construite en un temps record et dimensionnée pour Grok. Problème : peu de gens voulaient de Grok pour travailler. Le parc tournait à environ 11 % d'utilisation, un taux que l'opérateur lui-même avait publiquement qualifié de gênant.

La stratégie a un nom : l'éléphant blanc. Le roi de Siam offrait à ses rivaux un animal sacré, impossible à faire travailler, impossible à refuser, impossible à revendre, et ruineux à entretenir. Colossus, c'était cela : une infrastructure de plusieurs milliards qu'on ne pouvait ni éteindre ni rentabiliser.

Puis, en mai 2026, l'inversion. Anthropic signe pour la totalité de la capacité de Colossus 1 : plus de 220 000 accélérateurs Nvidia, 300 mégawatts, 1,25 milliard de dollars par mois jusqu'en mai 2029. Un mois plus tard, Google signe à son tour pour environ 920 millions mensuels. SpaceX devient, presque par accident, le plus grand loueur de calcul du monde, avec un revenu récurrent d'environ 2,17 milliards de dollars par mois.

Il faut prendre la mesure de la situation ainsi créée. Le laboratoire qui domine actuellement la frontière technologique paye chaque mois plus d'un milliard de dollars à son concurrent direct pour avoir le droit de faire tourner ses propres modèles. Cette dépendance n'est pas une anomalie passagère : c'est la conséquence mécanique du fait que le calcul est devenu plus rare que le talent.

Deux fragilités méritent d'être notées, car elles seront décisives dans les dix-huit prochains mois. D'une part, la totalité des revenus calcul de SpaceX provient de deux clients. D'autre part, les contrats prévoient une sortie possible avec 90 jours de préavis après le 31 décembre 2026. Ce n'est pas de l'abonnement récurrent, c'est un bon de commande révocable. La date du 31 décembre 2026 est probablement l'échéance la plus importante du calendrier économique de l'IA.

4. Trois mécanismes économiques déjà à l'œuvre

La déflation de l'intelligence

Le chiffre à surveiller n'est pas le score d'un modèle, c'est le coût d'une tâche accomplie. Sur ce plan, la trajectoire est brutale. Grok 4.6 gagne cinq points d'indice en cinq semaines sans augmenter son tarif affiché — 2 dollars le million de jetons en entrée, 6 en sortie. Mesuré indépendamment, le coût moyen par tâche s'établit autour de 0,84 dollar, contre 1,04 pour GPT-5.6 Sol et nettement plus pour les modèles d'Anthropic.

Plus révélateur encore : l'efficacité par tour. Sur des tâches longues, Grok 4.6 résout en moyenne un problème en une cinquantaine d'échanges et un demi-milliard de jetons d'entrée, là où Claude Opus 5 en consomme environ le double de tours et quatre fois plus de contexte. Sur du travail agentique, cette différence pèse davantage que le prix au jeton.

La lecture économique est simple : le prix d'une unité de raisonnement de niveau professionnel baisse d'un facteur important chaque année, pendant que la qualité monte. Toute activité dont le prix est indexé sur des heures humaines de production documentaire va subir une compression de marge. Ce n'est pas une prédiction, c'est une observation de courbe.

La valeur remonte vers l'amont

Deuxième mécanisme, moins intuitif. Entre le premier et le troisième modèle mondial, il y a deux points d'indice et cinq semaines de calendrier. Autrement dit : la couche des modèles est contestable, donc peu rentable durablement. La rente se déplace vers ce qui est rare et lent à répliquer.

Ce qui est rare, en 2026 : la lithographie avancée, dont une seule entreprise au monde détient la clé ; la fonderie de pointe ; les accélérateurs ; et surtout l'électricité et les raccordements réseau. Trois cents mégawatts pour une seule grappe, c'est un objet d'aménagement du territoire autant qu'un objet informatique. La contrainte limitante de l'IA n'est plus le talent ni même le capital : ce sont les mégawatts, les transformateurs et les permis de construire.

La recomposition du travail de bureau

Troisième mécanisme, le plus discuté et le plus mal mesuré. Les nouveaux tests ne mesurent plus des questions d'examen mais du travail réel : rédaction juridique, analyse, production documentaire. Grok 4.6 y obtient le meilleur score publié sur un test de raisonnement juridique professionnel, à 15,8 %, contre 2,5 % pour GPT-5.6 Sol et 11,3 % pour Claude Fable 5.

Il faut lire ce chiffre honnêtement, dans les deux sens. En relatif, l'écart est spectaculaire. En absolu, 15,8 % reste un score faible : aucun de ces systèmes ne remplace un juriste. Ce qu'ils compriment, c'est la couche intermédiaire de production — la première version, la synthèse, la recherche préparatoire, la mise en forme, la traduction, le code de premier niveau.

Deuxième nuance, décisive : la capacité n'est pas l'adoption. Ce qui freine le déploiement en entreprise n'est presque jamais le benchmark. C'est la responsabilité juridique, la conformité, l'intégration aux systèmes existants et la réorganisation des processus. C'est pourquoi les effets sur l'emploi se manifesteront d'abord par un ralentissement des recrutements sur les postes juniors de production documentaire, plutôt que par des plans sociaux visibles.

5. Les trois acteurs les mieux placés pour rafler la mise

Le critère retenu ici n'est pas « qui a le meilleur modèle ce trimestre » — cette question change de réponse toutes les cinq semaines. Le critère est : qui capte une rente durable, c'est-à-dire un revenu que les autres doivent payer quel que soit le vainqueur de la course aux modèles.

Premier : la chaîne du silicium, Nvidia en tête

C'est le pari le moins spectaculaire et le plus solide. Quand Anthropic verse 15 milliards par an à SpaceX, une part majeure de cette somme se convertit en matériel Nvidia, lui-même gravé par une seule fonderie de pointe, sur des machines produites par un seul équipementier au monde, aux

Pays-Bas. Cette couche encaisse quel que soit le laboratoire qui gagne. Elle est protégée par des barrières physiques — capitaux, savoir-faire, délais de plusieurs années — et non par des barrières logicielles répliquables en cinq semaines.

Ses vulnérabilités sont réelles et identifiables : la substitution par du silicium chinois (Ascend) et par les puces conçues en interne par les grands clients ; et le fait qu'une part importante de la demande actuelle est financée par de la dette et par des contrats croisés entre acteurs du même écosystème.

Deuxième : SpaceXAI

Aucune autre entreprise au monde ne réunit aujourd'hui cette combinaison : accès à l'orbite, constellation de satellites et donc distribution et connectivité mondiales, centres de calcul et contrats d'énergie, parc d'accélérateurs, modèle de frontière, et désormais la surface de rédaction de code. Il faut y ajouter une position singulière : SpaceXAI est simultanément le concurrent et le bailleur de deux de ses trois rivaux.

Le risque est du même ordre de grandeur que l'ambition. La dette a été creusée à plus de 60 milliards de dollars par l'intégration de xAI. Le résultat par action reste négatif. Cent pour cent des revenus calcul dépendent de deux clients disposant d'une clause de sortie. Le titre est extrêmement volatil — il a évolué entre 105 et 226 dollars en deux mois. Enfin, l'exposition réglementaire européenne est lourde, et la concentration du pouvoir de décision sur une seule personne est en soi un facteur de risque systémique.

Troisième : Anthropic

Le laboratoire domine l'indice de référence, a construit la boucle du code avant tout le monde, et facture principalement des entreprises — le segment où la valeur se monétise le mieux et où les contrats sont les plus durables. C'est aujourd'hui l'acteur dont la qualité par unité de risque est la plus élevée.

Sa vulnérabilité est structurelle et bien identifiée : il loue son infrastructure à un concurrent direct, pour un montant qui absorbe une part majeure de son chiffre d'affaires. Il est également exposé aux décisions de contrôle à l'exportation de son propre gouvernement — l'épisode de juin 2026, sur lequel nous revenons plus bas, l'a démontré. Et ses modèles font l'objet d'accusations de captation par distillation de la part de laboratoires chinois.

La contre-lecture, qu'il serait malhonnête d'omettre

Deux acteurs pourraient légitimement prendre l'une de ces trois places.

Google, d'abord. C'est le seul acteur intégré verticalement de la puce (ses propres accélérateurs) à la distribution (Android, la recherche, le cloud), sans louer quoi que ce soit à un rival. Si l'on classait sur la seule robustesse structurelle, Google occuperait probablement la première place. S'il n'y figure pas ici, c'est parce que sa capacité à convertir cette intégration en position dominante sur la couche agentique reste, à ce jour, moins démontrée que celle de ses concurrents.

OpenAI, ensuite, qui dispose d'une distribution grand public qu'aucun autre laboratoire n'égale, et d'une boucle de code fonctionnelle. Son handicap relatif tient à sa structure de coûts et à l'ampleur de ses engagements d'infrastructure.

Enfin, un « acteur » qui n'est pas une entreprise : l'écosystème chinois des modèles à poids ouverts. Il ne captera probablement pas la rente financière. Il peut capter quelque chose de plus lourd à long terme : le standard.

6. États-Unis : la puissance déléguée à des entreprises

Les États-Unis n'ont pas construit de programme public d'intelligence artificielle comparable au programme Apollo ou au projet Manhattan. Ils ont délégué l'infrastructure stratégique à une poignée d'entreprises privées. C'est un choix cohérent avec leur modèle, et il produit des effets qui n'ont pas d'équivalent historique.

Le premier effet est un déplacement de la souveraineté. Lorsque la connectivité d'un théâtre d'opérations, la capacité de calcul d'un laboratoire de frontière et l'accès à l'orbite dépendent des décisions d'un même conseil d'administration, la puissance publique s'exerce à travers des entités qu'elle ne contrôle pas complètement. La désactivation d'un lot de terminaux ou la suspension d'un modèle produisent des effets de politique étrangère à la vitesse d'une décision privée.

Le deuxième effet concerne l'instrument principal de la politique américaine : le contrôle des exportations. En juillet 2026, l'administration a publiquement accusé Moonshot AI d'avoir développé son modèle Kimi K3 par distillation à grande échelle d'un modèle d'Anthropic, et soupçonné un contournement des restrictions sur les puces Blackwell via des filiales tierces. Ces accusations, quelle que soit leur issue, révèlent un dilemme insoluble : restreindre la diffusion pour ralentir Pékin, ou maintenir une diffusion suffisamment large pour éviter que les modèles chinois ne deviennent l'option par défaut du reste du monde. On ne peut pas faire les deux.

Le troisième effet est domestique et physique. La contrainte américaine n'est plus le capital ni les ingénieurs, ce sont les mégawatts. Les moratoires locaux sur les centres de données, les délais de raccordement et les oppositions municipales deviennent des variables de politique industrielle nationale.

Le quatrième effet est financier, et c'est probablement le canal de risque le plus sous-estimé. Une entreprise valorisée autour de 1 900 milliards de dollars avec un résultat par action négatif, dont une partie des revenus provient de contrats révocables signés par des concurrents eux-mêmes exposés aux décisions administratives de Washington : cette configuration relie directement une décision politique, une valorisation boursière et l'approvisionnement d'une entreprise européenne.

7. Chine : banaliser plutôt que dépasser

La stratégie chinoise de 2026 ne consiste pas à battre les modèles américains sur leur terrain. Elle consiste à rendre leurs capacités banales.

Le véhicule est le poids ouvert. Qwen chez Alibaba, DeepSeek V4 sorti au printemps sous licence permissive et taillé pour les puces Huawei, Kimi K3 chez Moonshot avec ses 2 800 milliards de paramètres, auxquels s'ajoutent Zhipu et MiniMax. Kimi K3 a pris la première place mondiale sur au moins un classement de génération de code en juillet, et se situe désormais juste derrière le trio de tête sur l'indice composite de référence.

Deux conséquences opérationnelles en découlent, et elles sont plus importantes que les scores.

La première est matérielle : plusieurs modèles chinois de premier plan sont désormais entraînés sur des accélérateurs Ascend, sans matériel américain. Chaque modèle de frontière entraîné sur du silicium chinois réduit d'autant le levier des contrôles à l'exportation. L'instrument principal de Washington perd en efficacité à mesure qu'il est utilisé.

La seconde est commerciale. Des projets de villes intelligentes et des programmes gouvernementaux à Singapour, en Malaisie ou dans le Golfe déploient déjà des modèles Qwen ou des architectures portées par Huawei. L'enjeu n'est pas la marge : c'est le standard, la dépendance d'écosystème, la formation des ingénieurs locaux sur une pile plutôt qu'une autre. C'est exactement la manœuvre que les États-Unis ont menée avec leurs propres systèmes d'exploitation pendant trente ans.

La contre-lecture existe, et elle est solide. Les poids ouverts diffusent mais monétisent mal : la Chine peut gagner la bataille de la diffusion et perdre celle de la rente. Par ailleurs, si les accusations de distillation se confirment, une partie de la trajectoire chinoise dépend de la frontière qu'elle est accusée de copier — ce qui la rend vulnérable à un décrochage si cette frontière s'éloigne réellement.

Ce qu'il faut en retenir pour la question centrale : si des modèles ouverts chinois restent six mois derrière la frontière à un dixième du prix, le pouvoir de fixation des prix des trois acteurs cités plus haut s'érode précisément dans les segments où la rente était censée se former.

8. Russie et Ukraine : l'infrastructure privée comme arme

L'Ukraine est le premier conflit où une constellation satellitaire privée constitue une dépendance militaire critique — pour les deux camps.

Côté ukrainien, des dizaines de milliers de terminaux Starlink assurent les communications de terrain et le pilotage d'une partie des missions de drones, appréciés pour leur stabilité et leur résistance au brouillage. Côté russe, des terminaux Starlink ont été retrouvés montés sur des drones d'attaque longue portée, permettant de contourner les défenses électroniques qui neutralisent les signaux GPS et radio. Kiev a documenté des centaines de cas à partir de janvier 2026, dont une frappe sur un train de passagers.

Début février 2026, SpaceX a désactivé les terminaux utilisés sans autorisation. Des analystes militaires ont estimé que cette mesure avait probablement contribué au succès de contre-attaques ukrainiennes localisées — tout en perturbant, au passage, une partie des systèmes ukrainiens eux-mêmes. Kiev développe depuis un mécanisme n'autorisant que les terminaux enregistrés à fonctionner sur son territoire.

Il faut mesurer ce que cet épisode signifie. Une entreprise privée a modifié le tempo d'une guerre par un changement de configuration. Cette même entreprise possède désormais un modèle de frontière et la couche de rédaction de code. La séparation traditionnelle entre infrastructure civile et capacité militaire n'existe plus.

Pour la Russie, les conséquences sont contrastées. Coupée des puces avancées, elle n'a pas de trajectoire crédible vers la frontière technologique et dépendra de la pile chinoise. Mais — et c'est l'asymétrie centrale de ce dossier — l'IA militairement décisive à court terme n'est pas la plus coûteuse. Un guidage terminal autonome sur les dernières centaines de mètres, capable de fonctionner sans liaison et sans GPS, tourne sur des composants à quelques dizaines de dollars. Il ne requiert ni Colossus, ni 300 mégawatts, ni un modèle à 63 points d'indice.

Autrement dit : la course au calcul et la course à l'autonomie militaire de masse sont deux courses distinctes, faiblement couplées. La première se joue entre trois ou quatre entreprises américaines et chinoises. La seconde se joue dans des ateliers, à coût marginal quasi nul, et elle diffuse beaucoup plus vite. Les contrôles à l'exportation sont efficaces sur la première et presque inopérants sur la seconde.

Pour l'Ukraine, la contrainte a produit un avantage comparatif inattendu : le pays est devenu l'un des principaux laboratoires mondiaux d'autonomie appliquée, non par la qualité de ses modèles mais par sa vitesse d'intégration — le cycle entre un retour de terrain et une modification déployée s'y compte en jours. C'est un savoir-faire que l'Europe n'a pas et qu'elle ferait bien d'acheter plutôt que de

reconstruire.

9. Europe : le pouvoir d'interdire sans le pouvoir de substituer

L'Europe détient dans cette chaîne un actif réellement irremplaçable : la lithographie avancée. C'est un point de passage obligé pour l'ensemble de l'industrie mondiale, y compris chinoise. Il ne faut pas le sous-estimer — c'est le seul véritable levier européen.

Le reste relève de la dépendance, et l'Europe a choisi d'y répondre par la norme. Les instruments sont réels : une amende de 120 millions d'euros infligée à X en décembre 2025, une procédure formelle ouverte le 26 janvier 2026 au titre du règlement sur les services numériques visant spécifiquement l'intégration de Grok à la plateforme, des injonctions de conservation de documents, une interdiction des pratiques de dénudation par IA introduite dans le cadre du règlement sur l'intelligence artificielle, et un plafond de sanction pouvant atteindre 6 % du chiffre d'affaires mondial.

Cette capacité de sanction est authentique. Mais elle produit un paradoxe : l'Europe peut interdire un produit, elle ne peut pas en offrir un substitut. Le résultat concret, mesurable, est que les entreprises européennes travaillent avec six à douze mois de retard sur des outils que leurs concurrents utilisent quotidiennement.

L'épisode le plus révélateur de 2026 n'est pourtant pas européen. Le 12 juin, Anthropic a suspendu l'accès à ses modèles Fable 5 et Mythos 5 pour se conformer à une décision de contrôle à l'exportation du département du Commerce américain. L'accès a été rétabli le 1er juillet, après la levée de ces contrôles. Dix-huit jours pendant lesquels une décision administrative américaine a déterminé avec quoi les entreprises européennes pouvaient calculer. Aucune réglementation européenne n'aurait pu l'empêcher, ni le compenser.

Ce qui se construit malgré tout mérite d'être cité sans complaisance ni dénigrement. Mistral AI vise 200 mégawatts d'ici fin 2027 et un gigawatt d'ici 2030, financés par des engagements pluriannuels convertibles pris par de grands industriels européens — un montage qui revient à faire préfinancer l'infrastructure par ses futurs clients. L'Union européenne, de son côté, a lancé un appel à projets pour un ensemble de gigafactories d'IA, doté de plusieurs dizaines de milliards d'euros, dont les candidatures se ferment en novembre 2026 pour une sélection début 2027.

La comparaison des ordres de grandeur est le vrai sujet. Une seule grappe Colossus consomme aujourd'hui 300 mégawatts. L'objectif du champion européen est de 200 mégawatts fin 2027. L'écart n'est pas un écart d'ambition, c'est un écart de calendrier — et dans une industrie où deux générations de modèles séparent cinq semaines, un écart de calendrier de trois ans est un écart de nature.

La bonne question européenne n'est probablement pas « comment avoir notre propre modèle de frontière », qui suppose de gagner une course déjà engagée avec trois ans de retard et cinquante fois moins de capital. C'est : « qu'est-ce qui continue de fonctionner si l'accès est coupé dix-huit jours ? » Cette question-là a une réponse atteignable — modèles à poids ouverts, inférence hébergée en Europe, compétences internes — et elle coûte un ordre de grandeur de moins. Le fait que Mistral héberge désormais des modèles ouverts d'origine chinoise sur sa propre plateforme est de ce point de vue moins une contradiction qu'un aveu de ce que la souveraineté signifie concrètement : la continuité de service, pas la pureté d'origine.

10. Pourquoi toutes ces questions n'en font qu'une

Il est tentant de traiter séparément l'actualité technologique, la guerre en Ukraine, la rivalité sino-américaine et la réglementation européenne. Ce découpage ne tient plus, pour une raison matérielle : ces sujets partagent les mêmes objets physiques.

La chaîne est unique, et chaque maillon est un point d'étranglement contrôlé par quelqu'un : électricité et raccordement réseau, lithographie, fonderie, accélérateurs, centres de calcul, modèles, surface de distribution, données d'usage — et retour au modèle. Une décision prise sur n'importe lequel de ces maillons se propage sur tous les autres.

Trois couplages méritent d'être nommés explicitement.

Le couplage entre finance et géopolitique. La valorisation d'une entreprise cotée américaine repose en partie sur les loyers que lui versent ses concurrents, lesquels sont exposés aux décisions administratives de leur propre gouvernement, lesquelles répondent aux progrès des laboratoires chinois. Une décision à Washington déplace simultanément un cours de bourse à New York et la disponibilité d'un outil dans une PME lyonnaise.

Le couplage entre énergie et souveraineté. Qui ne peut pas raccorder 300 mégawatts ne peut pas héberger un modèle de frontière. La souveraineté numérique se mesure désormais en gigawatts, en transformateurs et en délais d'instruction de permis — c'est-à-dire dans un domaine où les décisions se prennent à l'échelle de la décennie, face à une industrie qui itère à l'échelle du mois.

Le couplage entre le code et le reste. Qui contrôle la surface où le code s'écrit contrôle la couche agentique, donc l'automatisation de tout travail produisant des documents, donc l'essentiel de l'économie tertiaire. C'est la raison pour laquelle un éditeur de texte vaut 60 milliards de dollars, et c'est aussi la raison pour laquelle cette acquisition est un sujet géopolitique et pas seulement un sujet d'entreprise.

Le même satellite fournit l'internet d'un village et le guidage d'un drone. Le même centre de calcul fait tourner un assistant de rédaction et une analyse d'imagerie. Le même contrôle à l'exportation est une mesure commerciale et une mesure de guerre. Il n'y a plus de frontière entre les rubriques.

11. Court terme et moyen terme : ce qui est déjà engagé

Douze à dix-huit mois

Le coût par tâche continue de baisser fortement. Ce qui coûte aujourd'hui le prix d'un modèle de frontière sera accessible au prix d'un modèle intermédiaire d'ici mi-2027.

Consolidation des surfaces agentiques autour de trois ou quatre écosystèmes, avec des coûts de sortie croissants à mesure que les processus internes s'y adossent.

Premiers effets visibles sur l'emploi, par ralentissement des recrutements sur les postes de production documentaire de premier niveau plutôt que par des suppressions de postes annoncées.

L'électricité devient un sujet politique local : l'implantation d'un centre de calcul devient un enjeu d'élection municipale, en Amérique du Nord d'abord, en Europe ensuite.

Échéance à surveiller : le 31 décembre 2026, date après laquelle les grands contrats de calcul deviennent résiliables avec 90 jours de préavis. Une sortie de client provoquerait une réévaluation brutale.

Escalade réglementaire européenne, avec une sanction possible calculée sur le chiffre d'affaires mondial — et la question, immédiatement politique, des représailles.

Trois à cinq ans

Deux scénarios structurent la période, et ils s'excluent.

Dans le premier, la frontière continue de s'écarter des modèles ouverts. La rente se consolide alors chez trois ou quatre acteurs, la couche du calcul et du silicium capte l'essentiel de la valeur, et la dépendance des États à ces entreprises devient un sujet constitutionnel plus qu'économique.

Dans le second, les modèles ouverts comblent l'écart. La rente se déplace alors vers la distribution et les applications, les marges se compriment partout, et le gagnant n'est plus celui qui possède le modèle mais celui qui possède la relation client et les données propriétaires. Ce scénario est nettement plus favorable à l'Europe, aux PME et aux professionnels indépendants.

Dans les deux cas, trois évolutions semblent robustes. La fragmentation en deux ou trois blocs technologiques aux standards distincts, avec une vaste zone non alignée — Golfe, Asie du Sud-Est, Afrique, Amérique latine — qui constitue le véritable enjeu de la compétition. Le déplacement du débat sur l'emploi, de « les postes vont-ils disparaître » vers « qui capte le gain de productivité » : les entreprises, les salariés, ou la couche de calcul qui facture les deux. Et la diffusion vers le bas de l'autonomie militaire, à bas coût, indépendamment de la course à la frontière.

12. Ce qui pourrait faire dérailler ce scénario

Un dossier qui n'énonce pas ses conditions de réfutation n'est pas une analyse, c'est une prophétie. Cinq facteurs peuvent invalider ce qui précède.

Financier. Une part significative de la dépense d'infrastructure est financée par de la dette et par des contrats croisés entre acteurs du même écosystème. Une correction sur ces valorisations retarderait l'ensemble de la trajectoire de deux à trois ans.

Physique. Électricité, transformateurs, eau de refroidissement, permis : ces contraintes se déplacent lentement et peuvent plafonner la croissance indépendamment de la demande.

Politique. Un incident marquant — scandale de contenus synthétiques, accident industriel attribué à un système agentique — produit de la réglementation plus vite que n'importe quel score de benchmark.

Technique. Rien ne garantit que les gains mesurés sur les tests se traduisent en adoption. Un score de 15,8 % sur une tâche juridique reste un score faible en valeur absolue.

Concentration. Une entité unique qui touche à l'accès à l'orbite, à la connectivité mondiale, au calcul, aux modèles, aux paiements et au transport constitue une cible antitrust évidente, et un risque de personne-clé difficile à couvrir.

Conclusion

La formule qui circule depuis des années — ne jamais parier contre Elon Musk — est un slogan, pas une analyse. La question intéressante n'est pas de savoir s'il gagne. Elle est de constater que la structure de l'industrie a changé de forme en un trimestre.

Il y a un an, la frontière technologique se jouait entre deux boucles. Il y en a trois aujourd'hui, ce qui accélère mécaniquement le rythme et écrase les prix. La rente s'est déplacée vers l'amont, vers ce qui ne se réplique pas en cinq semaines : le silicium, l'énergie, l'accès physique. Et les États exercent désormais leur puissance à travers des entreprises qu'ils ne contrôlent qu'imparfaitement, ce qui

donne à des décisions de conseil d'administration des effets de politique étrangère.

Pour une entreprise, un indépendant ou un professionnel, la conclusion opérationnelle est peu spectaculaire mais solide. Le coût de l'intelligence artificielle baisse et continuera de baisser. L'avantage concurrentiel ne viendra donc pas de l'accès au meilleur modèle — cet accès sera universel et bon marché — mais de la qualité de ses propres processus, de ses propres données et de sa capacité à réorganiser son travail autour de ces outils. C'est précisément la partie que personne ne peut acheter à 60 milliards de dollars.

Denis Favre — La Réussite à Domicile - DENIS8.COM