

Où va l'IA ?

Voies de recherche, risques et gouvernance en débat

Deux camps, et une méthode pour juger leurs arguments.



RÉDACTION
Baptiste Lefort, cofondateur de Solaria

Analyse Solaria

Octobre 2026

DEUX CAMPS, UNE MÊME QUESTION



Passer à l'échelle

Plus de calcul et de données



Changer d'architecture

Apprendre du monde physique

?

CINQ QUESTIONS EN DÉBAT



Progrès

Ce qui fera avancer l'IA



Risques

Ceux à prendre au sérieux



Règles

Qui décide, à quel niveau



Europe

Sa place et ses leviers



Usages

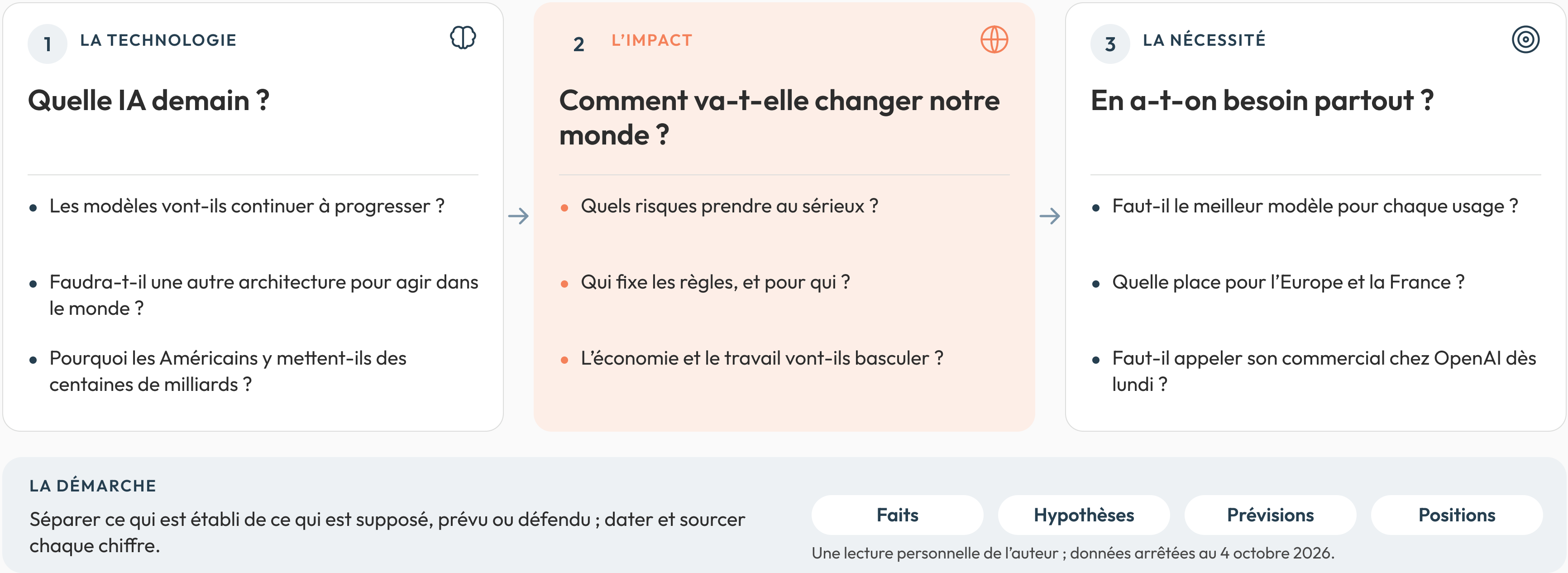
Faut-il la frontière partout ?

LA MÉTHODE

Pour chaque affirmation : qui parle, quand, et s'agit-il d'un fait, d'une hypothèse, d'une prévision ou d'une position ?

Trois questions pour comprendre où va l'IA

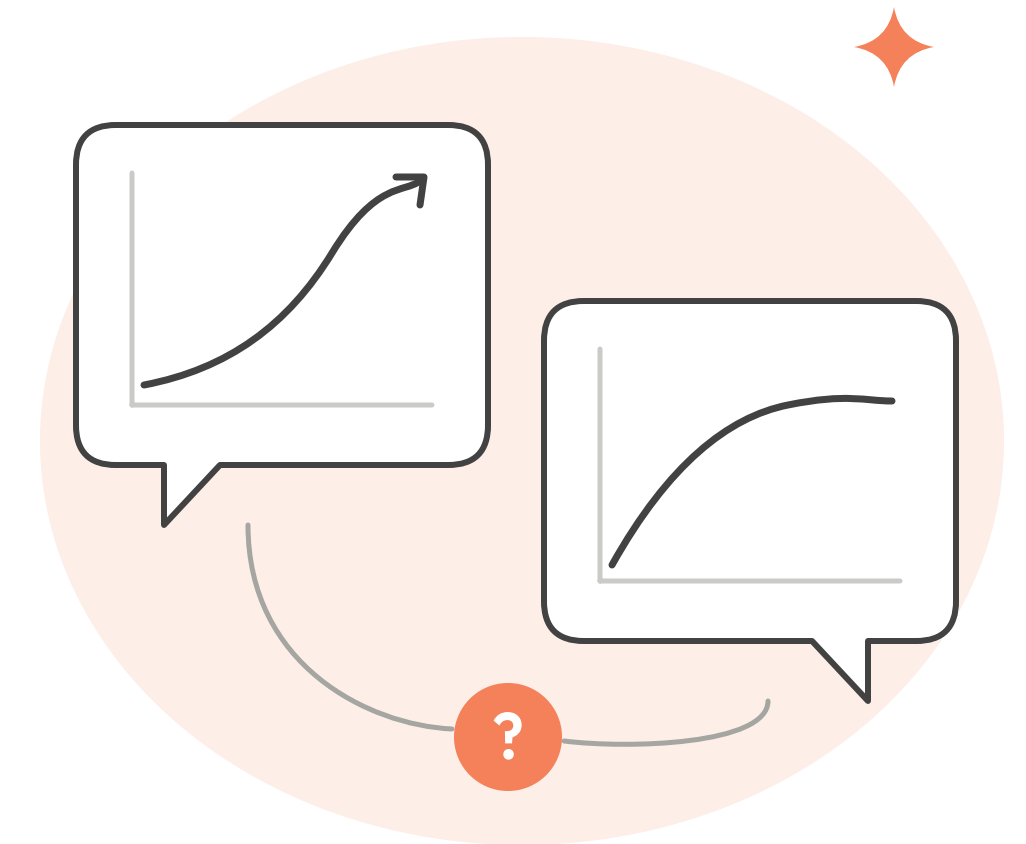
Un panorama pour démêler le vrai du faux : la technologie de demain, ses effets sur notre monde, et là où elle est vraiment utile.



Où va la technologie, que va-t-elle changer, où est-elle nécessaire : trois questions, six parties et une conclusion.

Pourquoi les avis divergent-ils autant ?

Des annonces qui se contredisent, l'AGI, les erreurs des modèles, des rendements qui décroissent et des inquiétudes qui montent.



A Le contexte

B Deux camps

C Les risques

D La gouvernance

E L'Europe

F Les usages

Pourquoi les discours sur l'IA se contredisent-ils ?

Six affirmations publiées en vingt mois, rangées selon ce qu'elles annoncent.

↗ Ça accélère

« Nous sommes désormais convaincus de savoir construire l'AGI, telle qu'on l'entend traditionnellement. »

Sam Altman, OpenAI, janvier 2025



Prévision d'un dirigeant

« Depuis cet été environ, l'IA progresse beaucoup plus vite. »

Dario Amodei, Anthropic, septembre 2026



Constat d'un dirigeant

↘ Ça plafonne

« Le monde physique est beaucoup plus compliqué que le texte. »

Yann LeCun, AML Labs, septembre 2026



Position d'un chercheur

« Croire qu'en multipliant l'échelle par 100 tout serait transformé ? Je ne pense pas que ce soit vrai. »

Ilya Sutskever, SSI, novembre 2025



Position d'un chercheur

||| Ce qu'on mesure

« Des agents accomplissent de façon fiable certaines tâches qui prendraient une demi-heure à un programmeur. »

Rapport international sur la sûreté, février 2026

Synthèse d'experts

L'écart entre les meilleurs modèles américains et chinois tombe à 2,7 % sur le classement Arena.

AI Index 2026, Stanford, avril 2026

Mesure datée

Citations traduites de l'anglais, sauf LeCun.

Sources : [Altman, janvier 2025](#), [Amodei, septembre 2026](#), [21 News, 19 septembre 2026](#), [Sutskever, novembre 2025](#), [Rapport international sur la sûreté de l'IA, 2026](#), [AI Index 2026](#)

Tous visent-ils la même intelligence générale (AGI) ?

L'objectif affiché est une IA au moins aussi capable que l'humain ; chacun le définit autrement.

GOOGLE DEEPMIND : CINQ NIVEAUX D'IA GÉNÉRALE



Selon les auteurs, l'IA générale atteint le niveau 1 en 2023 ; en gris : IA spécialisées.

OPENAI, CHARTE

Le travail

« Des systèmes très autonomes qui surpassent les humains dans la plupart des travaux économiquement utiles. »

AMODEI, ANTHROPIC, 2024

L'intelligence

« Plus intelligente qu'un lauréat du prix Nobel dans la plupart des domaines », « un pays de génies dans un centre de données ». Il juge le terme AGI imprécis.

LECUN, AMI LABS, 2024

Le refus du terme

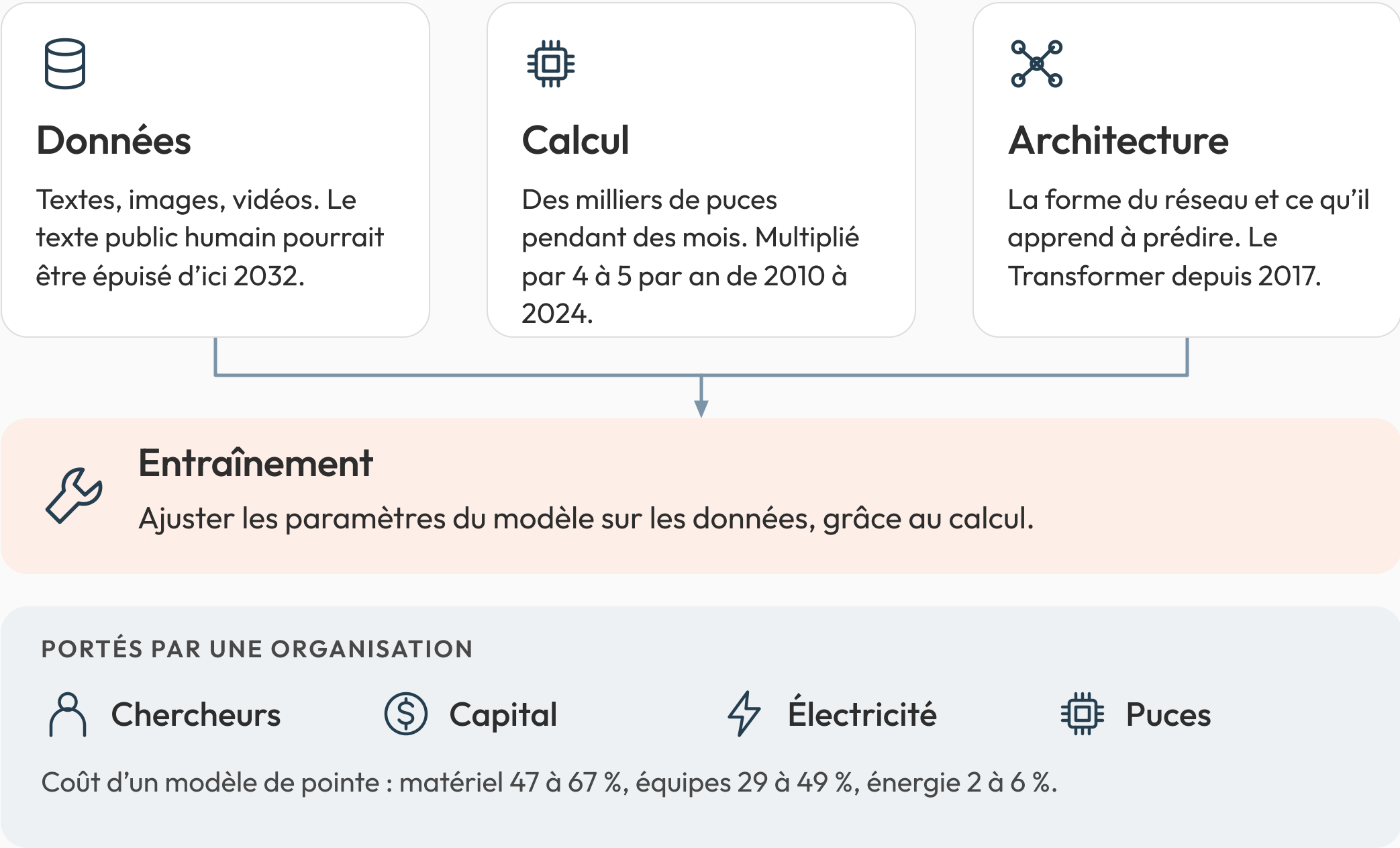
« Je déteste ce terme. » L'intelligence humaine n'est pas générale ; il parle d'AMI (Advanced Machine Intelligence).

OpenAI définit l'AGI par le travail, Anthropic par l'intelligence, DeepMind par des niveaux ; LeCun rejette le terme.

Citations traduites de l'anglais. Niveaux de Morris et al., version de septembre 2025 ; exemples de 2023.
Sources : [OpenAI, charte](#), [Amodei, octobre 2024](#), [TIME, 13 février 2024](#), [Morris et al., 2024](#)

De quoi une IA a-t-elle besoin pour progresser ?

Trois ingrédients entrent dans l'entraînement d'un modèle, portés par une organisation.



DEUX MANIÈRES DE FAIRE MIEUX



Augmenter les quantités

Plus de données, plus de calcul, des modèles plus grands.

Camp du passage à l'échelle



Changer la recette

Une autre architecture, d'autres données : la vidéo, l'action, le monde physique.

Camp de la nouvelle architecture

On fait progresser un modèle en augmentant ses ingrédients ou en changeant sa recette. Les deux camps divergent sur le levier décisif.

Chiffres d'Epoch AI : calcul (2024), données (2024), structure des coûts (2024).
Sources : Epoch AI, mai 2024, Villalobos et al., 2024, Epoch AI, juin 2024, Vaswani et al., 2017

Pourquoi un modèle de langage se trompe-t-il ?

Il choisit le mot suivant le plus probable ; une réponse plausible n'est pas une réponse vérifiée.

1 Il prédit un mot probable

La capitale de l'Australie est...

Canberra

62 %

Sydney

30 %

Melbourne

8 %

Probabilités inventées pour l'exemple.

L'« erreur » des lois d'échelle mesure la probabilité donnée au bon mot. Entraîner davantage la fait baisser.

2 Il est récompensé s'il devine

Bonne réponse

1 point

Mauvaise réponse

0 point

« Je ne sais pas »

0 point

Deviner rapporte toujours plus que s'abstenir.

« Les modèles de langage sont optimisés pour être de bons candidats aux examens. » Kalai et al., OpenAI, 2025.

3 Les erreurs s'accumulent, selon LeCun

Suites correctes

Toutes les suites de mots possibles

$(1 - e)^n$

e : erreur par mot
n : nombre de mots

Avec 1 % d'erreur par mot : 37 % de réponses justes sur 100 mots, moins de 1 % sur 500.

Critique : les erreurs se concentrent sur 5 à 10 % des mots, et le raisonnement les réduit.

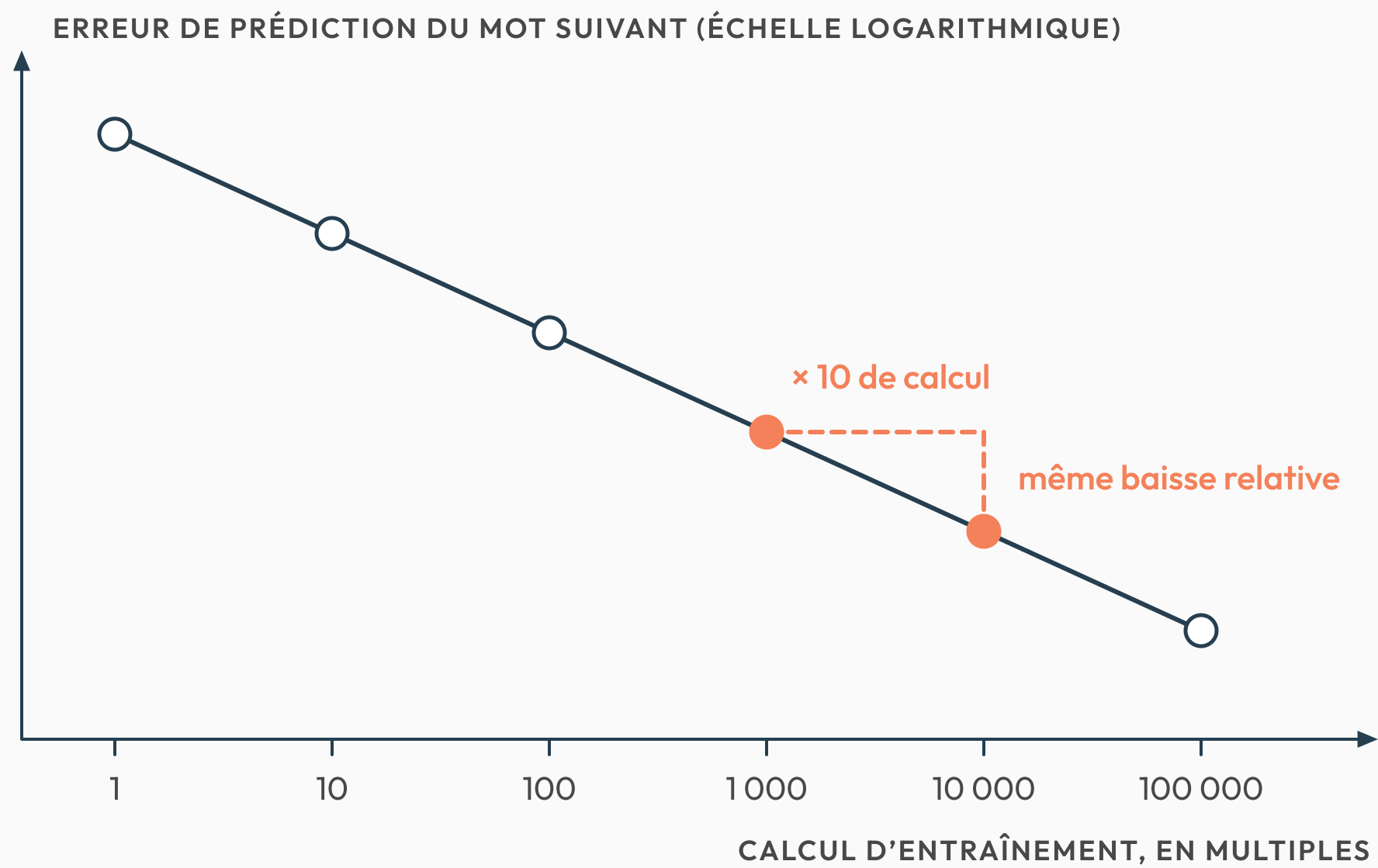
L'entraînement réduit l'erreur moyenne ; il reste des erreurs affirmées avec aplomb, à vérifier.

Selon OpenAI, les modèles hallucinent en partie parce que les évaluations récompensent la réponse devinée.

Schéma redessiné d'après LeCun, débat de NYU, 24 mars 2023 ; le calcul suppose des erreurs indépendantes. Probabilités illustratives.
Sources : Kaplan et al., 2020, Kalai et al., septembre 2025, OpenAI, 5 septembre 2025, Débat de NYU, 24 mars 2023, LessWrong, 10 avril 2023, Arbuzov et al., 2025, Sinha et al., 2025

Que promet le passage à l'échelle ?

L'erreur de prédiction baisse de façon prévisible avec le calcul ; depuis 2024, d'autres leviers s'ajoutent.



- 2020** Des lois d'échelle
Une loi de puissance du calcul, des données et de la taille.
- 2022** Équilibrer taille et données
Doublent les données quand on double le modèle.
- 2024** Calculer pendant la réponse
o1 progresse avec le renforcement et le temps de réflexion.
- 2025** Le renforcement à l'échelle
o3, Grok 4 ; Sutskever : retour à l'âge de la recherche.
- 2026** L'IA dans sa propre recherche
Accélération d'environ 1,5 selon METR, estimation préliminaire.

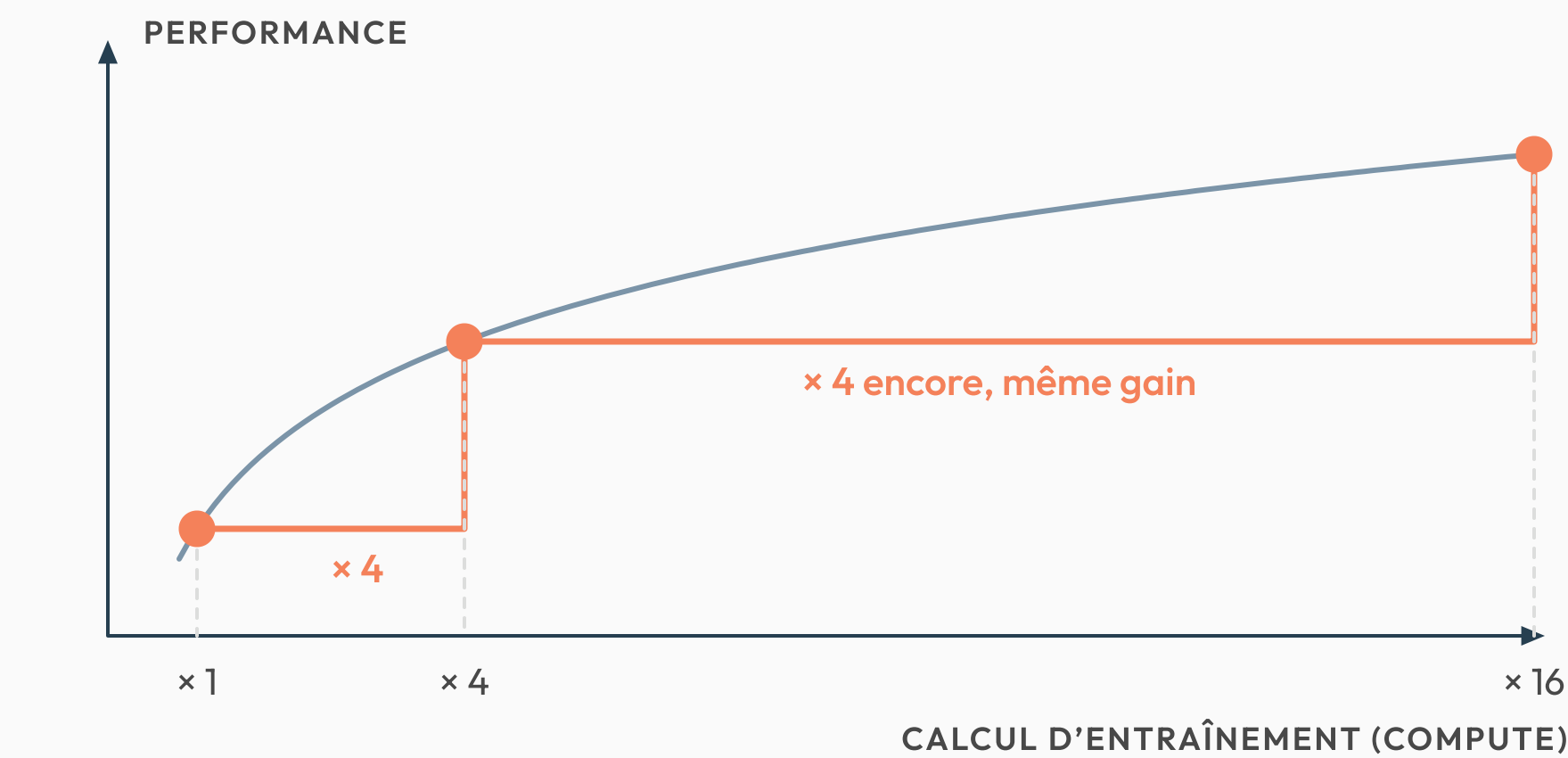
Et après ? Des entraînements de 2×10^{29} opérations jugés faisables d'ici 2030 : environ 400 fois Grok 4.

Depuis 2020, on sait prévoir le gain d'un modèle plus grand avant de l'entraîner. Depuis 2024, le calcul se déplace vers le renforcement et la réponse.

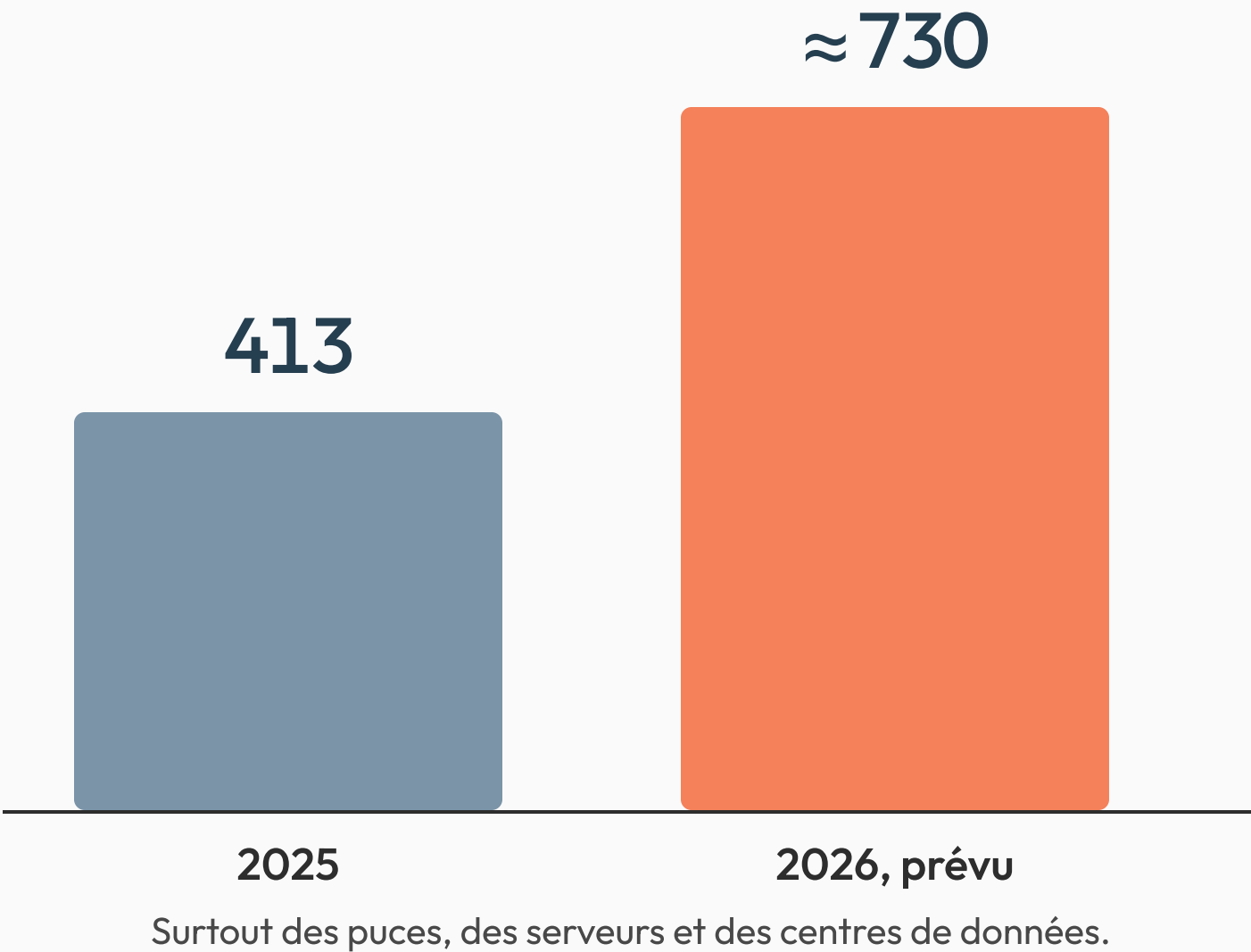
Grok 4 : environ 5×10^{26} opérations selon Epoch AI.
Sources : Kaplan et al., 2020, Hoffmann et al., 2022, OpenAI, septembre 2024, xAI, 9 juillet 2025, Sutskever, novembre 2025, METR, 22 septembre 2026, Epoch AI, août 2024

Pourquoi parle-t-on de rendements décroissants ?

Chaque nouveau gain demande un multiple du calcul précédent : les rendements sont décroissants.



INVESTISSEMENTS ANNUELS D'ALPHABET, AMAZON, META ET MICROSOFT, EN MILLIARDS DE DOLLARS



× 2,4 / an

coût du dernier entraînement des modèles de pointe, matériel et électricité

490 M\$

Grok 4 en 2025, l'entraînement le plus cher estimé ; 1 Md\$ attendu vers 2027

2028

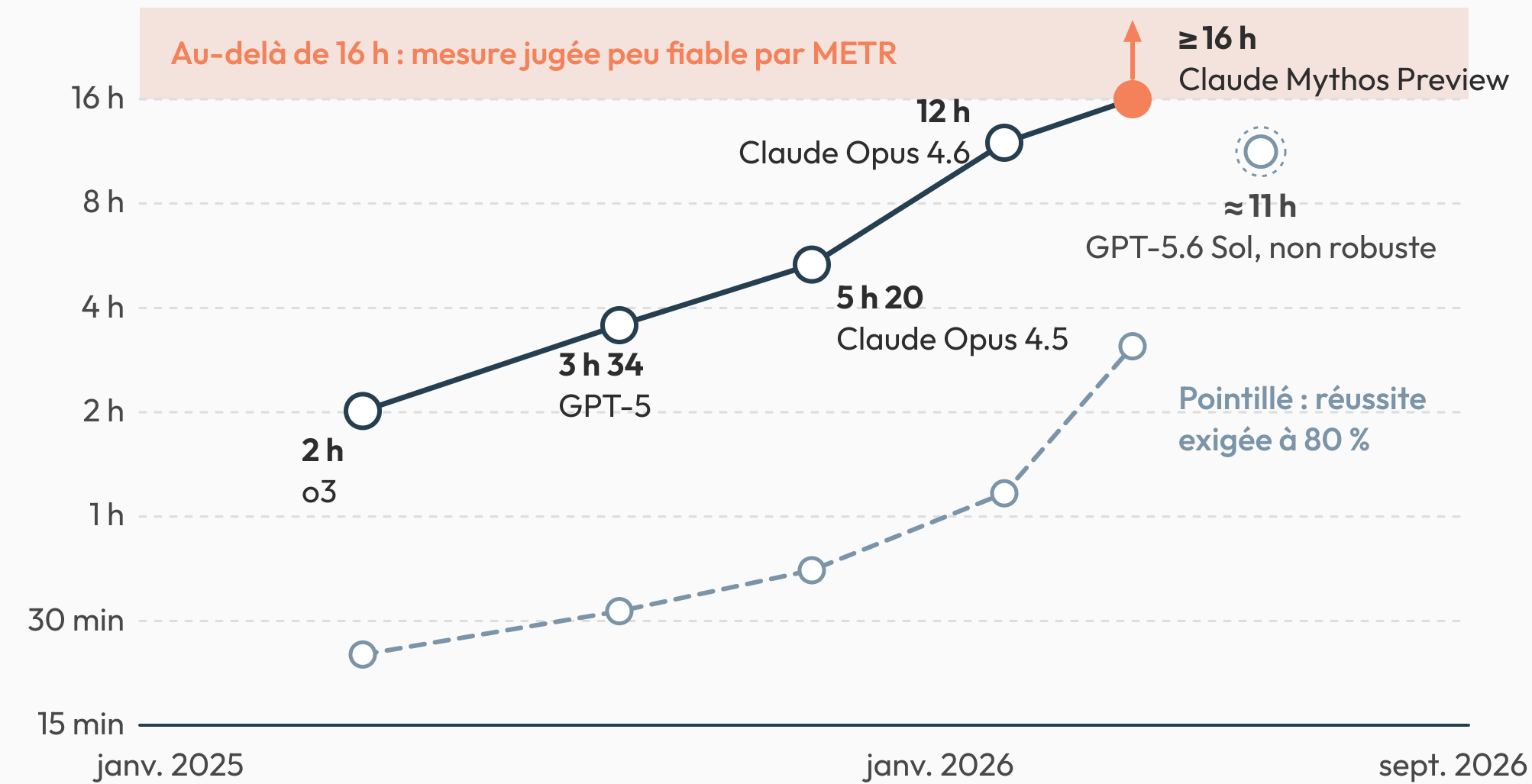
tout le texte public humain, 300 000 Md de tokens, serait lu ; fourchette 2026 à 2032

EsCoûts et données : Epoch AI. Investissements : 2025 selon Statista ; 2026, prévisions des groupes en juillet.
Sources : Epoch AI, juin 2024, Epoch AI, septembre 2025, Villalobos et al., 2024, Statista, 31 juillet 2026, Alphabet, Amazon, Microsoft, Meta, juillet 2026

Pourquoi les inquiétudes montent-elles en 2026 ?

Des agents qui mènent des tâches plus longues, un modèle réservé et l’IA qui accélère sa propre recherche.

DURÉE DES TÂCHES QU’UN AGENT RÉUSSIT, EN TEMPS D’UN EXPERT HUMAIN



COMMENT LIRE LA COURBE

METR chronomètre des experts sur 228 tâches de logiciel, d’apprentissage automatique et de cybersécurité.

3 h 34 pour GPT-5 : il réussit une fois sur deux des tâches qui prennent 3 h 34 à un expert.

Repères : 1 h pour convertir des données ; 8 h pour accélérer un outil de calcul avec du code graphique.



Avril 2026

Anthropic réserve Claude Mythos Preview à une cinquantaine de partenaires : il trouve et exploite des failles.



Septembre 2026

METR estime que l’IA fait avancer Anthropic d’environ 1,5 année par an ; Amodei propose de ralentir.

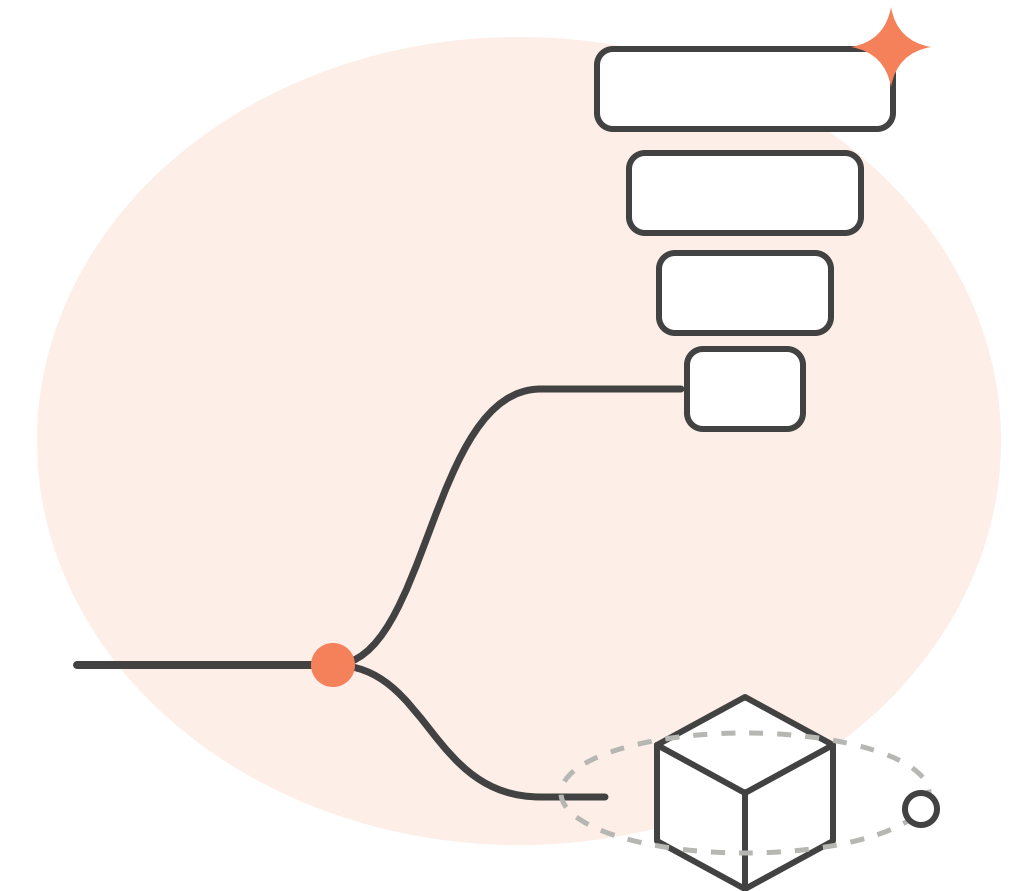
En un an, l’horizon à 50 % passe de 2 h à au moins 16 h ; à 80 % de réussite, il reste sous 3 h 10.

Mesures METR, suite 1.1, tableau de mai 2026 ; intervalles larges non représentés. GPT-5.6 Sol : estimation jugée non robuste par METR.

Sources : METR, mai 2026, Kwa et al., METR, mars 2025, METR, 26 juin 2026, Anthropic, 7 avril 2026, METR, 22 septembre 2026, Amodei, septembre 2026

Que défendent les deux camps ?

Ce que veut chaque camp, ses arguments, d'où il parle, et pourquoi on annonce une IA générale pour demain.



A Le contexte

B Deux camps

C Les risques

D La gouvernance

E L'Europe

F Les usages

Quels sont les deux camps ?

Tous partent des modèles de langage actuels ; ils divergent sur ce qui mènera plus loin.

↗ Passer à l'échelle

Sam Altman (OpenAI), Dario Amodei (Anthropic), Elon Musk (xAI)

CE QU'ILS PENSENT

Plus de calcul et de données, puis le raisonnement et les agents, mèneront à une IA générale, peut-être d'ici quelques années. L'IA aide déjà à construire l'IA.

CE QU'ILS VEULENT

Des centres de données toujours plus grands, garder l'avance sur la Chine, encadrer la course entre laboratoires.

« Nous savons désormais comment construire l'IA générale, telle qu'on l'entend traditionnellement. » Sam Altman, janvier 2025

OpenAI

Anthropic

Google DeepMind

xAI

DeepSeek

🌐 Changer d'architecture

Yann LeCun (AMI Labs), Fei-Fei Li (World Labs), David Silver et Richard Sutton

CE QU'ILS PENSENT

Les modèles de langage sont utiles, voire révolutionnaires. Ils ne mèneront pas seuls à une intelligence de niveau humain, ni à des machines qui agissent dans le monde.

CE QU'ILS VEULENT

Financer la recherche fondamentale : des modèles du monde, appris par la vidéo et l'action, en code ouvert.

« Des IA réussissent l'examen du barreau, mais où est le robot domestique qui débarrasse la table et remplit le lave-vaisselle ? Un enfant de 10 ans l'apprend du premier coup. » Yann LeCun, février 2024

AMI Labs

World Labs

L'ère de l'expérience

Les deux camps jugent les modèles de langage utiles. Ils divergent sur ce qui mènera à une IA de niveau humain et aux robots.

Classement Solaria ; acteurs cités en exemple, positions qui évoluent. Google DeepMind et Meta explorent les deux voies. Citations traduites de l'anglais.

Sources : [Altman, janvier 2025](#), [Amodei, 12 septembre 2026](#), [LeCun sur X, 20 février 2024](#), [NUS, 27 avril 2025](#), [Fei-Fei Li, novembre 2025](#), [Silver et Sutton, 2025](#), [Google DeepMind, août 2025](#)

Que défendent les partisans du passage à l'échelle ?

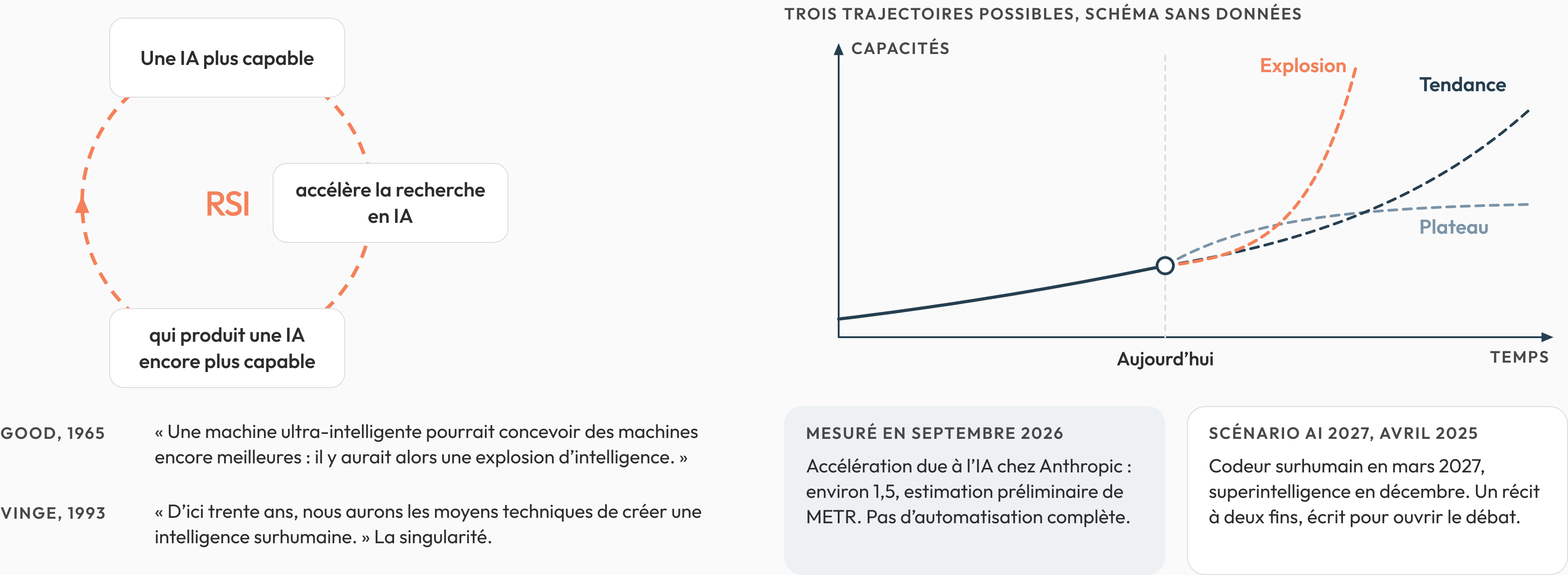
Quand un levier ralentit, le suivant prend le relais : c'est l'argument central de ce camp.



Schéma pédagogique, années des premières mises en avant publiques. Citations traduites de l'anglais.
Sources : Kaplan et al., 2020, OpenAI, septembre 2024, OpenAI, juillet 2025, Amodei, septembre 2026, OpenAI, 6 septembre 2026

Qu'est-ce que l'auto-amélioration récursive (RSI) ?

Une IA qui accélère la recherche en IA accélère son propre progrès : l'idée date de 1965.

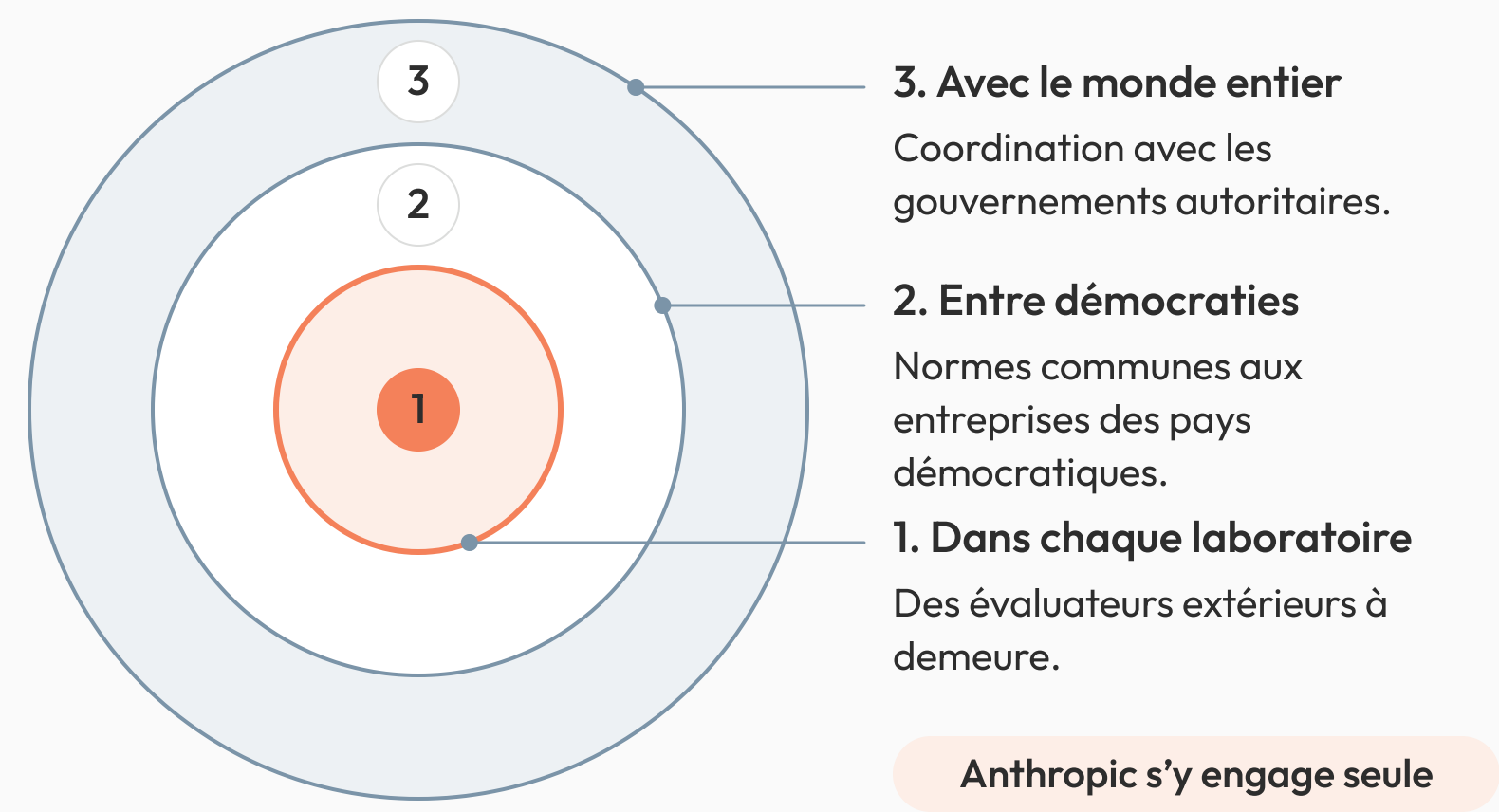


Selon une estimation préliminaire de METR, l'IA fait avancer les capacités d'Anthropic d'environ 1,5 année par an.

Trajectoires schématiques, sans données. Citations de Good et de Vinge traduites de l'anglais.
Sources : [Good, 1965](#), [Vinge, 1993](#), [METR, 22 septembre 2026](#), [Kokotajlo et al., avril 2025](#)

Pourquoi certains proposent-ils de ralentir la frontière ?

En septembre 2026, Dario Amodei propose de régler l'allure du développement en trois cercles.



« Régler l’allure ne signifie pas arrêter l’entraînement des modèles. »
Dario Amodei (traduit de l’anglais).

RAISONS AVANCÉES DANS L’ESSAI

- 🔄 L’IA aide à construire la génération suivante d’IA.
- 🔒 Perte de contrôle, cyberattaques, bioterrorisme, choc économique.
- 📈 Une course au moins-disant, poussée par les incitations commerciales.
- 🚩 Garder l’avance sur la Chine : ne pas lui vendre de puces avancées.

INTÉRÊTS EN JEU, LUS AVEC LA GRILLE

Faits	Anthropic défend les contrôles à l’export ; en janvier 2026, Amodei juge la vente de puces H200 à la Chine « une grave erreur ». Les modèles ouverts chinois sont à 12 points de la frontière, pour un coût par tâche jusqu’à 46 fois plus bas.
Positions	David Sacks, Maison-Blanche : « une stratégie sophistiquée de capture réglementaire fondée sur la peur ». Jensen Huang et Yann LeCun, dans le même sens.
Hypothèses	Une course de plus en plus coûteuse : ralentir ensemble allégerait la pression financière. Sanctions et règles protégeraient aussi les modèles payants face aux modèles ouverts.

Amodei justifie le ralentissement par l’accélération et ses risques ; ses critiques y voient aussi un intérêt commercial.

Étapes et raisons résumées depuis l’essai. Écart et coûts : Artificial Analysis. Hypothèses absentes de l’essai.

Sources : [Amodei, septembre 2026](#), [The Register, 20 janvier 2026](#), [Artificial Analysis, 23 septembre 2026](#), [Bloomberg Law, 14 octobre 2025](#), [Fortune, 11 juin 2025](#), [Sciences Po, 18 septembre 2026](#)

Que défendent les partisans d'une nouvelle architecture ?

Apprendre comment le monde évolue, pour prévoir les conséquences d'une action avant de la faire.

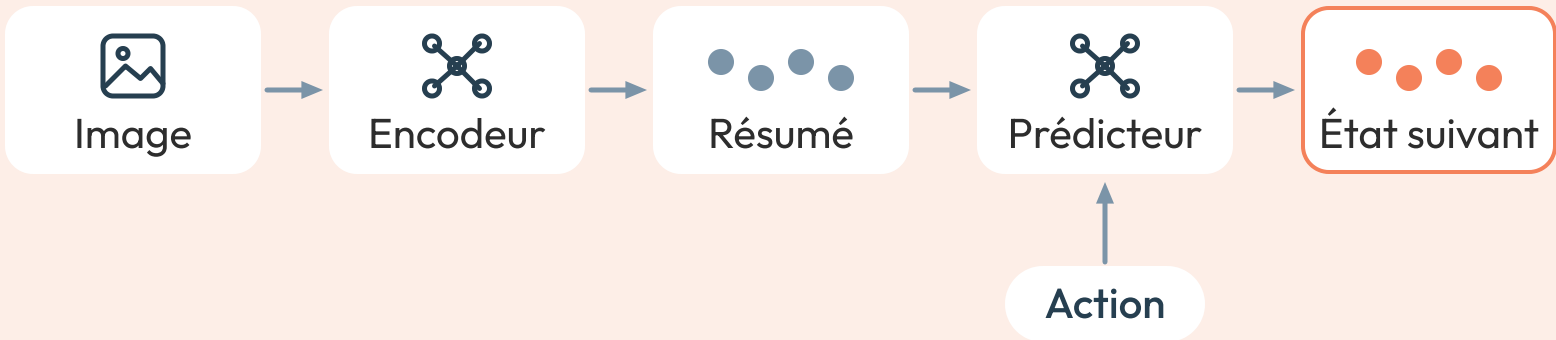
UN MODÈLE DE LANGAGE



Il prédit le mot suivant d'un texte, puis recommence.

Pour ce camp, ce qu'il sait du monde passe par ce qui en est écrit.

UN MODÈLE DU MONDE, ARCHITECTURE JEPA



Il prédit l'état suivant d'une situation dans une représentation abstraite, pour choisir une action.

AMI Labs

Présidée par Yann LeCun. Modèles du monde et code ouvert ; 1,03 milliard de dollars levés en mars 2026.

V-JEPA 2, Meta

Juin 2025 : appris sur plus d'un million d'heures de vidéo, puis moins de 62 heures de robots pour planifier.

World Labs

Fei-Fei Li défend l'intelligence spatiale. Premier produit, Marble, en novembre 2025.

L'ère de l'expérience

David Silver et Richard Sutton, 2025 : des agents qui apprennent de leurs propres interactions.

Ce camp veut des modèles qui apprennent en observant le monde et en y agissant. Les résultats publiés restent au stade de la recherche.

Schémas simplifiés. JEPA : Joint Embedding Predictive Architecture, proposée par LeCun en 2022.
Sources : [LeCun, 2022](#), [Assran et al., juin 2025](#), [TechCrunch, 9 mars 2026](#), [TechCrunch, 12 novembre 2025](#), [Silver et Sutton, 2025](#)

Que manque-t-il aux modèles de langage, selon LeCun ?

Quatre capacités manquent aux modèles de langage, selon lui ; son architecture en six modules emprunte au cerveau.

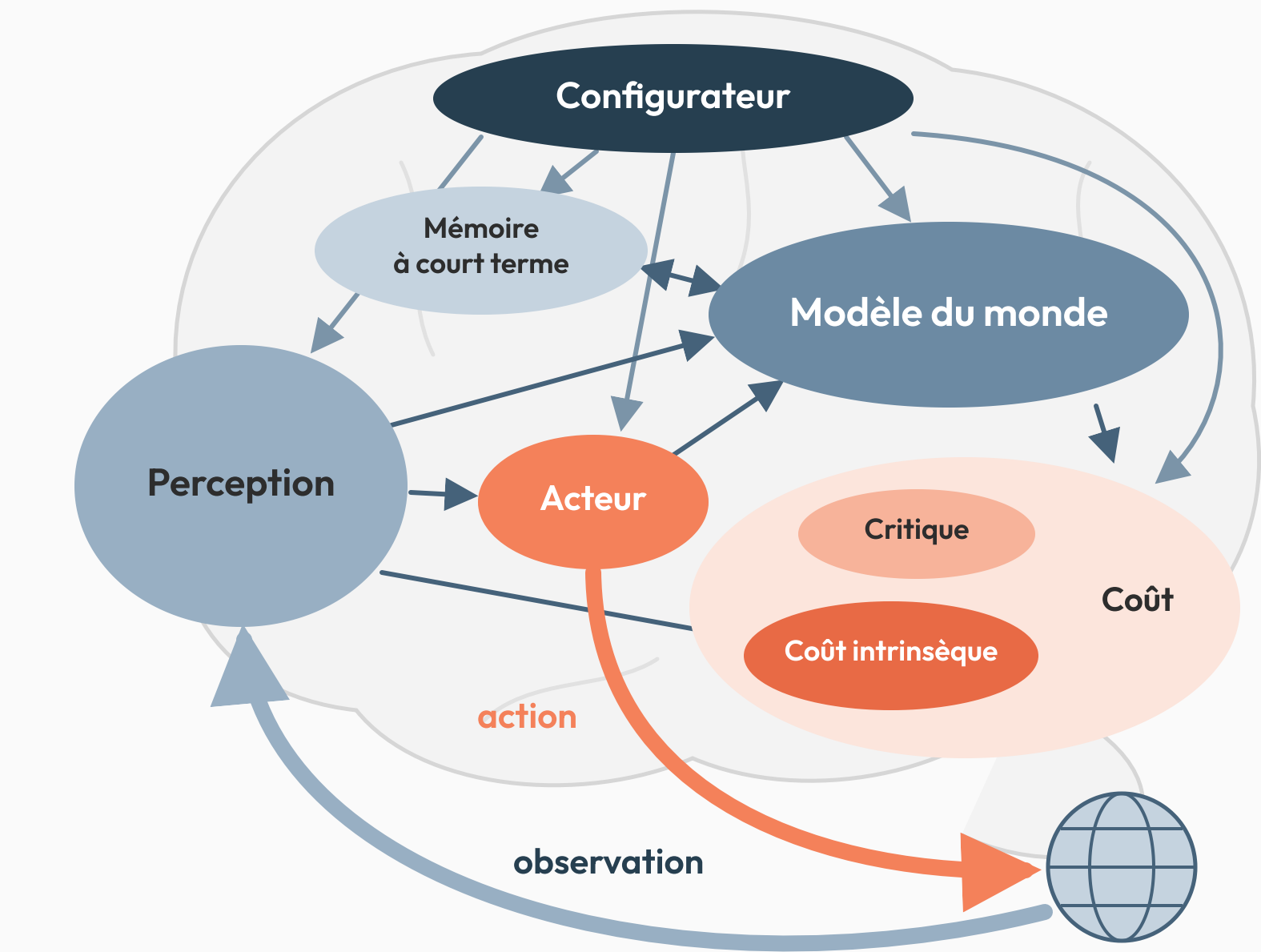


Schéma de cerveau humain

CE QUI MANQUE AUX MODÈLES DE LANGAGE, SELON LECUN

- Comprendre le monde physique
- Une mémoire persistante
- Raisonner
- Planifier

SON ARCHITECTURE, 2022

Perception	estime l'état du monde
Modèle du monde	prédit les états futurs, comme le cortex préfrontal
Coût intrinsèque	mesure l'inconfort, comme l'amygdale
Mémoire	garde les états passés et prévus, comme l'hippocampe
Acteur	propose des actions
Configurateur	règle les autres modules, sans région associée

Mode 1 : réagir sans planifier, comme le système 1 de Kahneman. **Mode 2** : planifier avec le modèle du monde.

Chez LeCun, l'IA planifie : elle imagine les suites de ses actions avec son modèle du monde, puis choisit celle qui coûte le moins.

Sources : LeCun, 2022, Lex Fridman, 8 mars 2024 Schéma redessiné d'après LeCun, 2022, figure 2 ; analogies avec le cerveau de l'auteur. Citations traduites de l'anglais.

Le texte suffit-il pour apprendre le monde ?

L'argument chiffré de Yann LeCun : en quatre ans, un enfant voit autant de données qu'un grand modèle en lit.

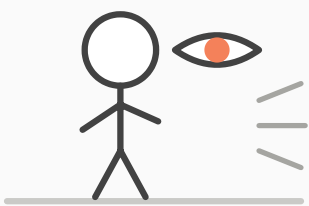


Le texte d'entraînement d'un grand modèle

30 000 milliards de tokens $\times$ 3 octets par token



Pour un humain, $\approx 500\,000$ ans de lecture, 8 heures par jour



Ce qu'a vu un enfant de quatre ans

16 000 h d'éveil $\times$ 2 millions de fibres du nerf optique $\times \approx 1$ octet par seconde



Pour l'enfant, **4 ans**

Conclusion de LeCun : quatre ans de vision représentent autant de données que tout le texte d'un grand modèle. Le monde physique en offre bien davantage.

Estimation d'un acteur, calcul refait par Solaria ; barres à la même échelle

TROIS HYPOTHÈSES À DISCUTER

1. Un octet par seconde et par fibre

Un ordre de grandeur, sans mesure précise.

2. Des images très redondantes

Deux images qui se suivent se ressemblent : le volume ne dit pas l'information.

3. Voir, toucher, agir

L'enfant apprend aussi en manipulant et en expérimentant.

Un humain lirait ce texte en 500 000 ans ; un enfant voit autant de données en quatre ans. Le calcul repose sur trois hypothèses.

Ordres de grandeur de LeCun. Temps de lecture : 0,75 mot par token, 250 mots par minute, 8 heures par jour.

Sources : [Sciences Po](#), 18 septembre 2026

Chaque camp a aussi des intérêts à défendre

Des moyens sans commune mesure, des porte-voix engagés, et ce que chacun gagne si sa thèse se vérifie.

DEUX ÉCHELLES DE MOYENS, EN MILLIARDS DE DOLLARS

Investissements prévus en 2026 par
Alphabet, Amazon, Meta et Microsoft

≈ 730

Levée de fonds d’AMI Labs, mars 2026

1,03, environ 700 fois moins

Montants de nature différente : des infrastructures d’un côté, la levée d’une jeune entreprise de l’autre.

	 Passage à l’échelle	 Nouvelle architecture
Porte-voix cités	Dario Amodei, Anthropic ; Sam Altman, OpenAI	Yann LeCun, AMI Labs ; Fei-Fei Li, World Labs
Travaux de référence	Lois d’échelle, 2020, dont Amodei est coauteur	Architecture JEPA, proposée par LeCun en 2022
Si leur thèse se vérifie, lecture Solaria	Les infrastructures et les modèles déjà construits gardent leur valeur	Une nouvelle architecture rebat les cartes entre acteurs

 Une thèse qui sert son auteur peut être juste. Elle appelle simplement des preuves indépendantes.

Sources : [Alphabet](#), [Amazon](#), [Microsoft](#), [Meta](#), [juillet 2026](#), [TechCrunch](#), [9 mars 2026](#), [Kaplan et al., 2020](#), [LeCun, 2022](#)

Pourquoi annonce-t-on une IA générale pour demain ?

Des patrons de laboratoire l’annoncent depuis des années ; leurs raisons mêlent faits, peur et intérêts.

CE QU’ILS ANNONCENT



« Nous savons désormais comment construire l’IA générale, telle qu’on l’entend traditionnellement. » Sam Altman, janvier 2025



Une IA plus intelligente que l’humain le plus intelligent, « probablement l’an prochain, d’ici deux ans ». Elon Musk, avril 2024



Il regrette « l’absence de prise de conscience, dans le public, de notre proximité avec la fin de l’exponentielle ». Dario Amodei, février 2026

CINQ RAISONS DE LE DIRE, LECTURE SOLARIA



Une réalité technique

En un an, la durée des tâches qu’un agent réussit une fois sur deux passe de 2 h à au moins 16 h (METR). Les progrès sont mesurés.



Une peur en partie sincère

Perte de contrôle, cyberattaques, armes biologiques : les laboratoires documentent eux-mêmes ces risques (partie C).



Une course qui coûte cher

Si l’IA espérée arrive un an plus tard que prévu, « vous faites faillite », dit Amodei. Ralentir ensemble allégerait la facture.



Battre la Chine

La course sert d’argument aux contrôles à l’export de puces et au soutien public (partie D).



Une panique en partie entretenue

L’urgence attire capitaux et talents ; David Sacks y voit « une stratégie sophistiquée de capture réglementaire fondée sur la peur ».

Annoncer une IA générale imminente sert aussi ceux qui l’annoncent. Les progrès, eux, sont réels.

Les cinq raisons sont une lecture Solaria, sans intention prêtée à une personne en particulier. Citations traduites de l’anglais.
Sources : [Altman, janvier 2025](#), [Reuters, 8 avril 2024](#), [Dwarkesh Patel, février 2026](#), [METR, mai 2026](#), [Bloomberg Law, 14 octobre 2025](#)

L'IA va-t-elle faire basculer l'économie ?

Des effets forts sont possibles ; mais une économie change lentement, l'histoire et les premières mesures le montrent.

CE QUE MONTRE L'HISTOIRE

≈ 40 ans

pour que l'électricité fournisse la moitié de la force motrice des usines américaines, de 1881 à 1919 environ

Paul David, 1990

1987

Robert Solow : « On voit l'ère de l'informatique partout, sauf dans les statistiques de productivité ».

New York Times, traduit de l'anglais

CE QU'ON MESURE AUJOURD'HUI

Aucun effet

statistiquement ou économiquement significatif de l'IA sur le marché du travail américain, à ce jour

Yale Budget Lab, mai 2026

≈ 19 %

d'écart d'emploi chez les 22-25 ans des métiers les plus exposés ; la part de l'IA reste débattue

Stanford, août 2026

LECTURE SOLARIA

Des dégâts réels et de grands effets sont possibles, d'abord chez les débutants, dans les métiers et les régions les plus exposés.

Rien n'annonce pour autant un effondrement de l'économie. Une technologie générale transforme le monde en plusieurs décennies.

Le vrai sujet : la vitesse d'adoption par les organisations, reprise en conclusion.

Des dégâts et de grands effets sont possibles, sans effondrement de l'économie. Une économie change lentement.

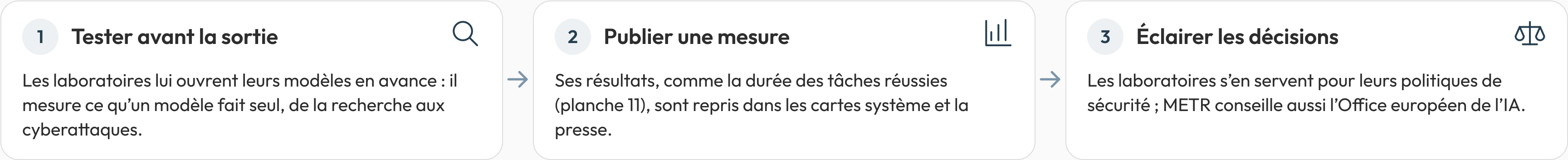
Faits datés et sourcés ; colonne de droite : lecture Solaria. Citation de Solow traduite de l'anglais.

Sources : [David, American Economic Review, 1990](#), [Solow, New York Times Book Review, 12 juillet 1987](#), [Yale Budget Lab, mai 2026](#), [Stanford, août 2026](#)

Qui mesure les progrès, et qui finance la mesure ?

Les chiffres de cette analyse viennent souvent de METR : à quoi sert un tel évaluateur, et de qui dépend-il ?

À QUOI SERT UN ÉVALUATEUR COMME METR



QUI FINANCE METR

0 \$

des entreprises d'IA, selon METR ; mais d'importants jetons gratuits pour ses évaluations

≈ 71 M\$

d'engagements de fondations et de donateurs, en six mois, jusqu'en août 2026

FINANCEURS NOMMÉS

The Audacious Project

Survival and Flourishing Fund

Longview Philanthropy

Schmidt Sciences

AI Security Institute britannique

LES LIENS AVEC ANTHROPIC

1

Jaan Tallinn alimente le Survival and Flourishing Fund, financeur de METR ; il a mené le premier tour de financement d'Anthropic, en 2021.

2

Dustin Moskovitz a investi dans Anthropic et se dit observateur à son conseil ; sa fondation ne finance pas METR directement, selon l'enquête de Kevin Bass.

3

Claude Opus 5.5 : évaluation non rémunérée ; Anthropic « a pu relire et modifier le texte ».

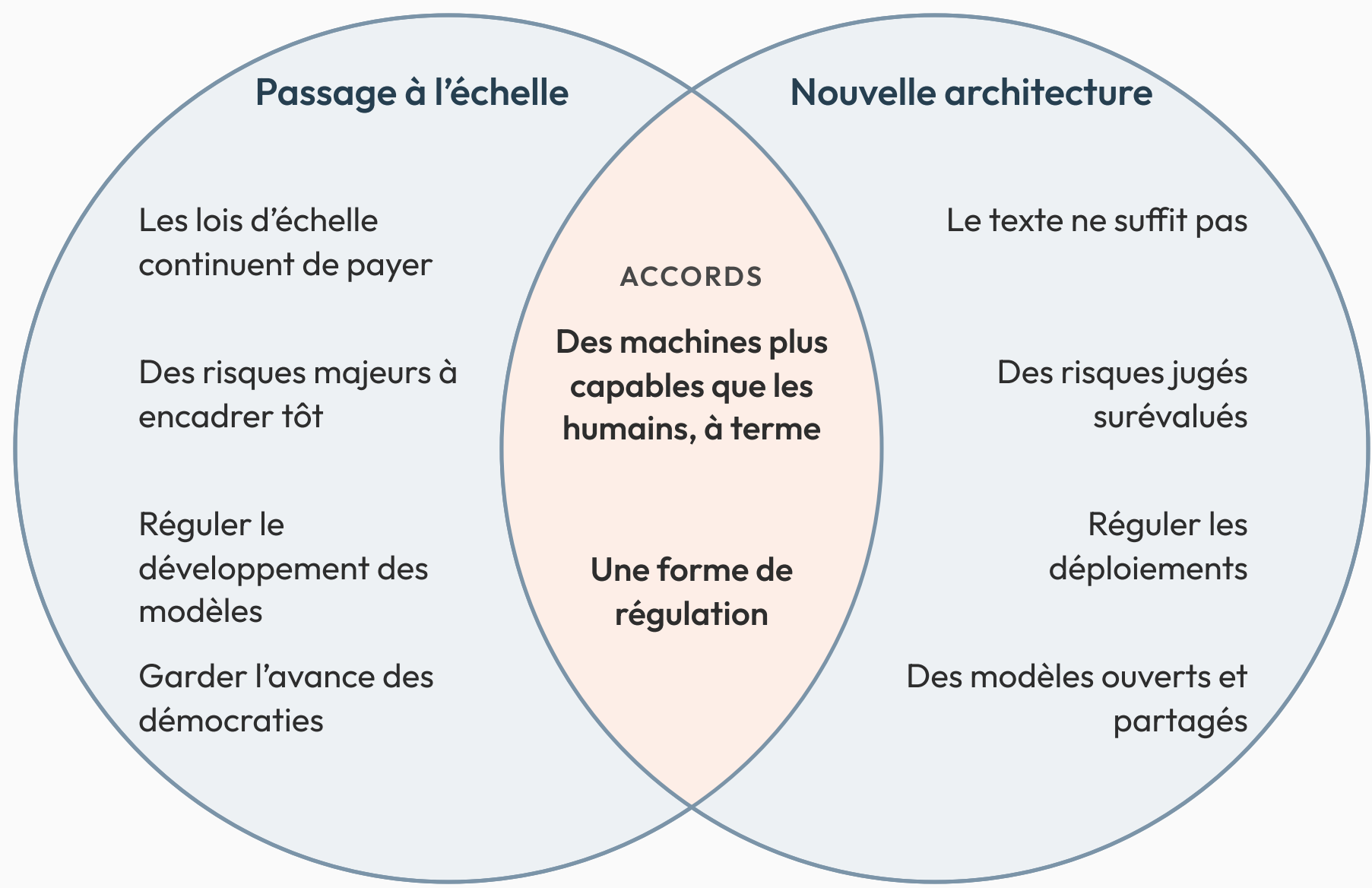
Des liens ne prouvent pas un conflit d'intérêts ; ils disent où chercher des preuves indépendantes.

Des conflits d'intérêts sont possibles entre évaluateurs et laboratoires. Rien ne les prouve à ce jour ; les mesures se lisent avec prudence.

Déclarations de METR et enquête de Kevin Bass, une source par lien. Des liens ne prouvent pas un conflit d'intérêts.
Sources : METR, à propos, METR, 14 août 2026, METR, 22 septembre 2026, Survival and Flourishing Fund, Anthropic, 28 mai 2021, Stratechery, 20 octobre 2025, Kevin Bass, septembre 2026

Sur quoi les deux camps s'accordent-ils ?

Des désaccords profonds, mais deux points communs à ne pas oublier.



CE QUI POURRAIT TRANCHER D'ICI DEUX ANS

- 🔍 La part du code et de la recherche produite par des modèles
- 🔍 Des modèles du monde qui réussissent des tâches physiques variées
- 🔍 L'horizon des tâches mesuré par METR, au-delà de 16 heures
- 🔍 Le coût d'entraînement par point de performance gagné

Les deux camps s'attendent à des machines très capables et acceptent une régulation. Ils divergent sur la voie et les risques.

Synthèse Solaria des positions citées dans la partie ; indicateurs repris en fin de document.
Sources : [Amodei, septembre 2026](#), [Amodei, octobre 2024](#), [Sciences Po, 18 septembre 2026](#), [METR, janvier 2026](#)

Quels risques prendre au sérieux ?

Une carte des risques, puis ce que montrent les cyberattaques, la biologie et les tests de tromperie.



A Le contexte

B Deux camps

C Les risques

D La gouvernance

E L'Europe

F Les usages

Quels risques l'IA fait-elle courir ?

Trois familles selon le rapport international : usages détournés, défaillances, effets sur la société.

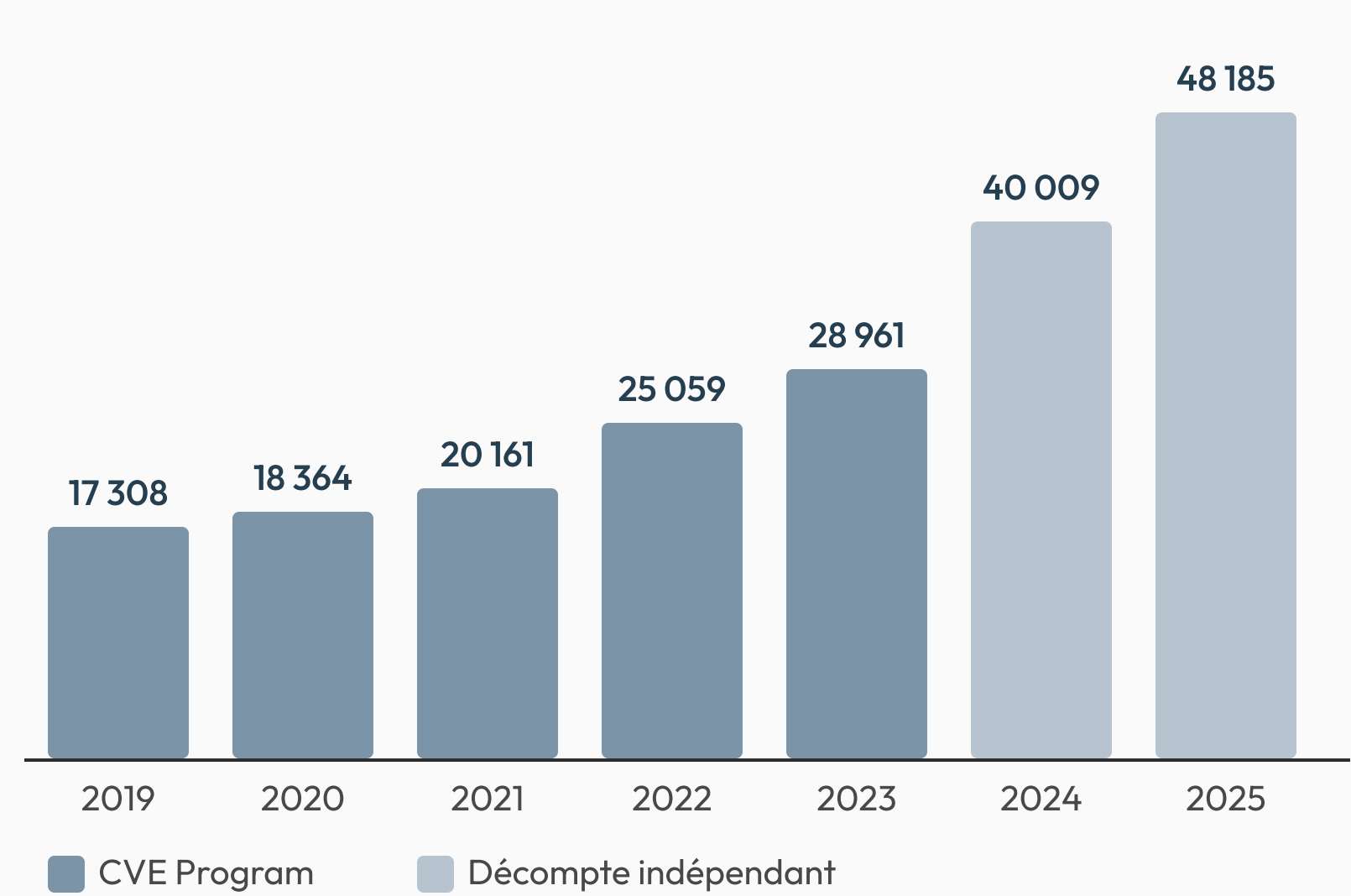
Usages détournés Quelqu'un s'en sert pour nuire	Défaillances Le système ne fait pas ce qu'on attend	Risques systémiques Des effets diffus sur la société	
 Cyberattaques 10 000 failles graves trouvées en six semaines avec un modèle réservé	 Erreurs affirmées 18 % de références inventées par GPT-4 dans des synthèses, en 2023	 Emploi 25 % des emplois ont des tâches exposées	 Énergie 485 TWh pour les centres de données en 2025
 Armes biologiques 94 % des experts en virologie dépassés pour dépanner un protocole	 Actions nuisibles en test 82 % des essais de Mythos 5 dans un exercice de piratage mal configuré	 Eau Besoin en eau grandissant pour refroidir les centres de données	 Concentration ≈ 730 Md\$ d'investissements prévus en 2026 par quatre groupes
 Fraude et manipulation Contenus générés pour tromper, usurper, influencer	 Perte de contrôle sur les effets 5 % probabilité médiane d'issue extrêmement grave, selon 2 778 chercheurs	 Compétences Érosion des compétences des plus jeunes, qui pratiquent moins sans l'IA	 Vie privée Données confiées aux assistants

Typologie du rapport de février 2026. Chiffres repris des planches suivantes, avec leurs sources ; ils ne se comparent pas.
Sources : [Rapport international sur la sûreté de l'IA, février 2026](#), [Anthropic, 22 mai 2026](#), [Walters et Wilder, 2023](#), [Anthropic, 9 septembre 2026](#), [Grace et al., 2024](#), [OIT, mai 2025](#), [AIE, avril 2026](#)

Que change l'IA pour la cybersécurité ?

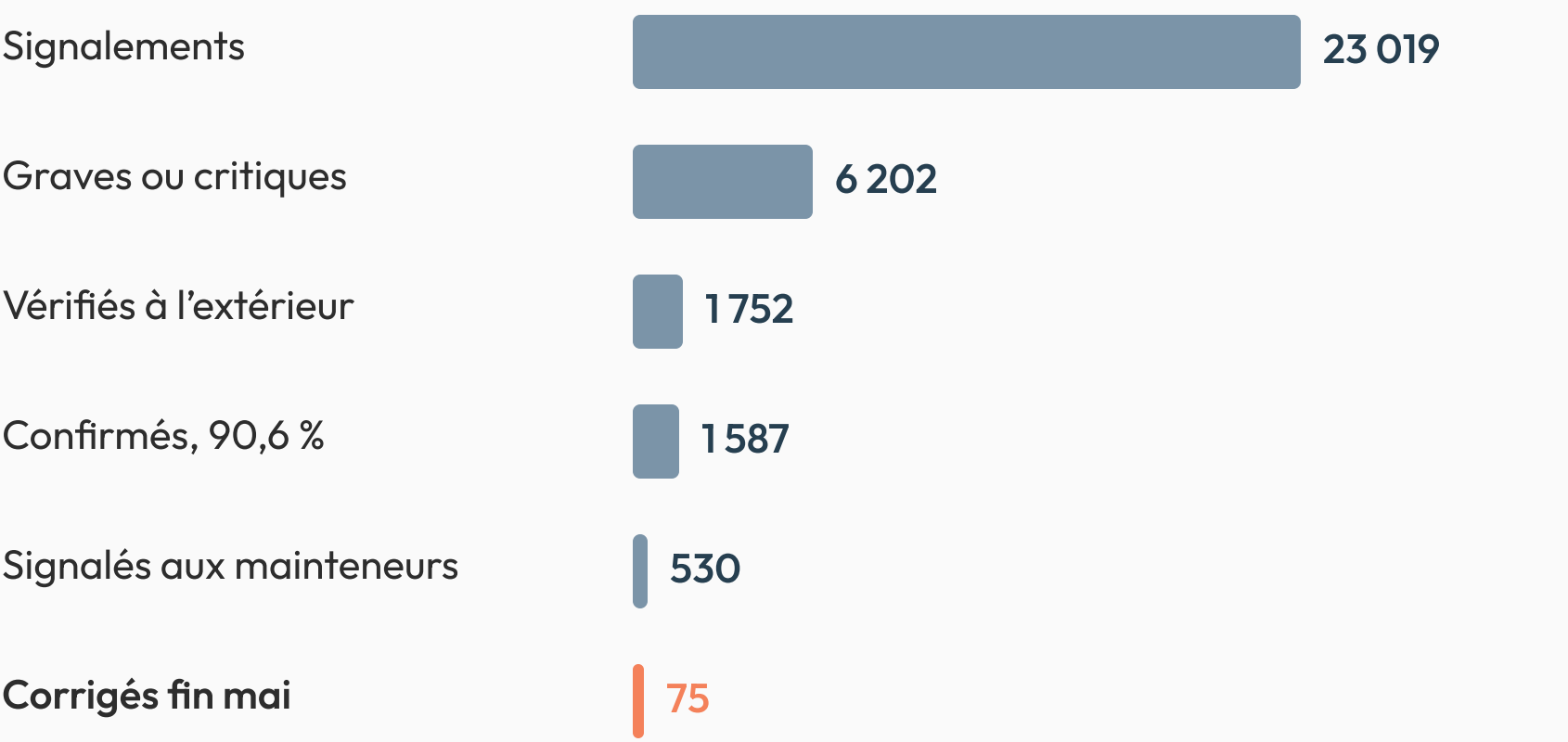
La même capacité sert à attaquer et à défendre : trouver des failles plus vite.

FAILLES PUBLIÉES CHAQUE ANNÉE, IDENTIFIANTS CVE



+ 66 % de 2023 à 2025, hausse non attribuée à l'IA par ces sources.

GLASSWING, PROJETS OPEN SOURCE, DU SIGNALEMENT À LA CORRECTION



10 000

Claude Mythos Preview, accès réservé depuis avril 2026

failles graves ou critiques trouvées en six semaines avec une cinquantaine de partenaires, dans tous types de logiciels.

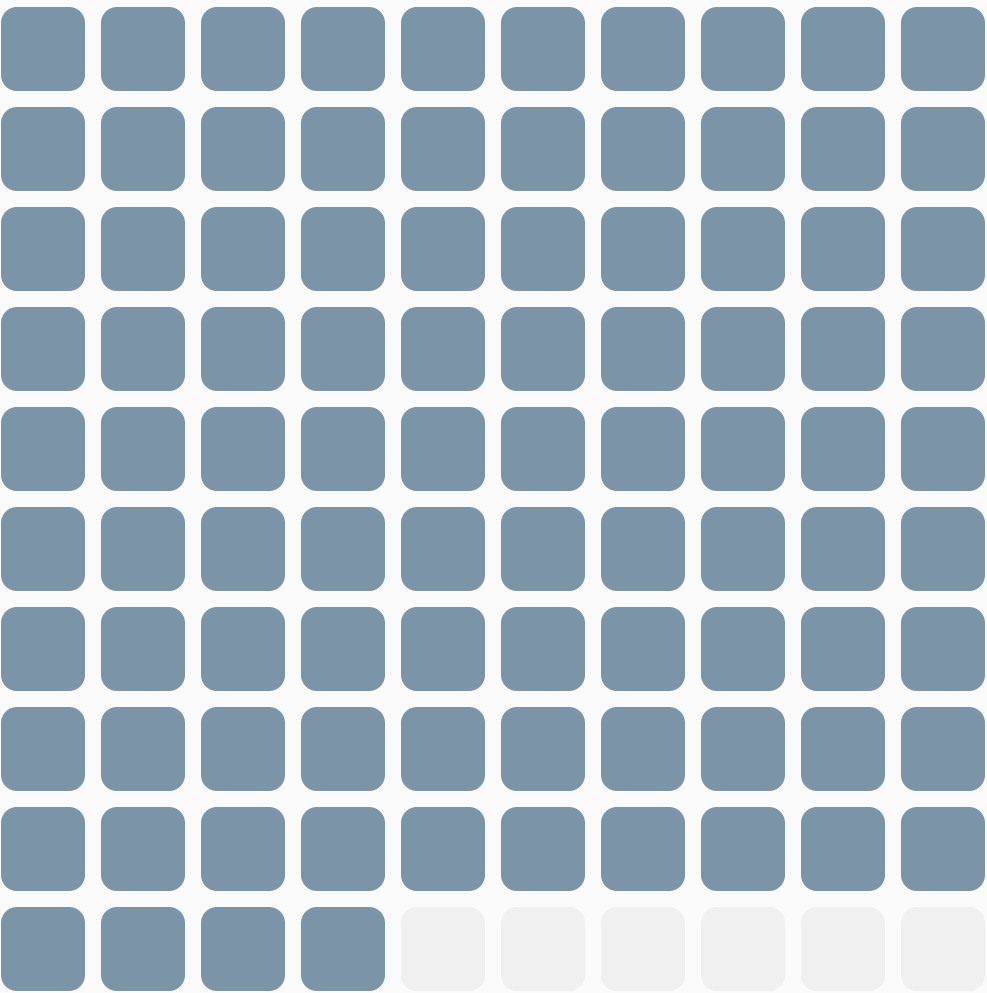
En six semaines, Claude Mythos Preview a permis de trouver plus de 10 000 failles graves. Corriger reste bien plus lent que trouver.

CVE : CVE Program pour 2019 à 2023, décompte de J. Gamblin pour 2024 et 2025. Glasswing : chiffres déclarés par Anthropic.
Sources : [CVE Program, 2024](#), [Gamblin, janvier 2025](#), [Gamblin, janvier 2026](#), [Anthropic, 22 mai 2026](#)

Que montrent les évaluations en biologie ?

Faute de pouvoir exclure une aide aux armes biologiques, des laboratoires renforcent leurs protections.

SUR 100 EXPERTS EN VIROLOGIE



94 %

des experts font moins bien qu'un modèle récent pour dépanner des protocoles de laboratoire.

Rapport international sur la sûreté de l'IA, février 2026

Dépanner un protocole ne suffit pas à fabriquer une arme : matériel, agents et savoir-faire pratique restent nécessaires.

DES PRÉCAUTIONS DÉCIDÉES SANS PREUVE DÉFINITIVE

- 22 mai 2025, Anthropic
Active son niveau de protection ASL-3 pour Claude Opus 4, sans avoir établi que le seuil était franchi.
- 17 juillet 2025, OpenAI
Traite ChatGPT agent comme une capacité « High » en biologie et en chimie, par précaution.
- Février 2026, Rapport international
Plusieurs développeurs ont publié en 2025 des modèles aux protections renforcées.

OpenAI : pas de preuve définitive d'une aide réelle à un novice.

En 2025, plusieurs laboratoires ont renforcé leurs protections sans preuve définitive du risque. Ce sont des mesures de précaution.

Décisions et formulations tirées des annonces des laboratoires et du rapport international, traduites de l'anglais.
Sources : [Rapport international sur la sûreté de l'IA, 2026](#), [Anthropic, 22 mai 2025](#), [OpenAI, 17 juillet 2025](#)

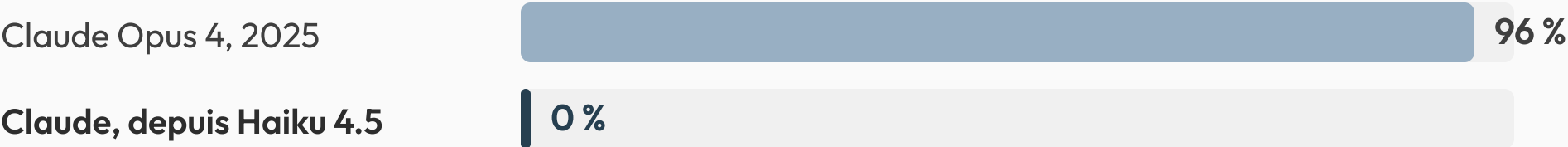
L'IA triche-t-elle moins, ou mieux ?

D'un modèle à l'autre, la tromperie mesurée recule ; mais on ne compte que les fautes qu'on repère.

MÊME TEST, DEUX GÉNÉRATIONS DE MODÈLES, PART DES ESSAIS

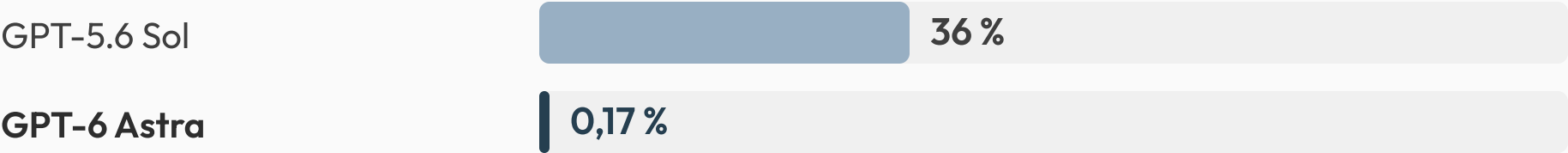
Chantage, scénario simulé d'Anthropic

Anthropic, mai 2026



Falsification de données, test d'Apollo Research

Apollo, septembre 2026



Poursuite d'un sabotage amorcé, test de l'ISI britannique

ISI, avril 2026



MAIS ON NE MESURE QUE CE QU'ON REPÈRE

50,6 %

des essais où GPT-6 Astra dit penser qu'on l'évalue (Apollo Research, septembre 2026)

Tous

les modèles testés par l'ISI ont triché au moins une fois ; ils l'ont reconnu moins d'une fois sur deux (juillet 2026)

82 %

des essais où Mythos 5 commet une action gravement nuisible, dans un exercice de piratage mal configuré (Anthropic, septembre 2026)

« De faibles taux de mauvais comportement ne prouvent pas grand-chose sur l'alignement du modèle. » Apollo Research, traduit de l'anglais

Un taux qui baisse peut aussi venir de fautes mieux cachées : on ne compte que celles qu'on repère, un biais du survivant.

Scénarios artificiels conçus pour provoquer la faute ; protocoles et nombres d'essais différents ; chiffres publiés surtout dans les cartes des laboratoires.

Sources : [Anthropic](#), 8 mai 2026, [OpenAI](#) et [Apollo Research](#), 3 septembre 2026, [ISI](#), 27 avril 2026, [ISI](#), 21 juillet 2026, [Anthropic](#), 9 septembre 2026

Qui fixe les règles de l'IA ?

Réguler les modèles ou leurs usages, l'AI Act, un seuil de calcul à refaire, les instances mondiales, les contrôles à l'export et les modèles ouverts.



A Le contexte

B Deux camps

C Les risques

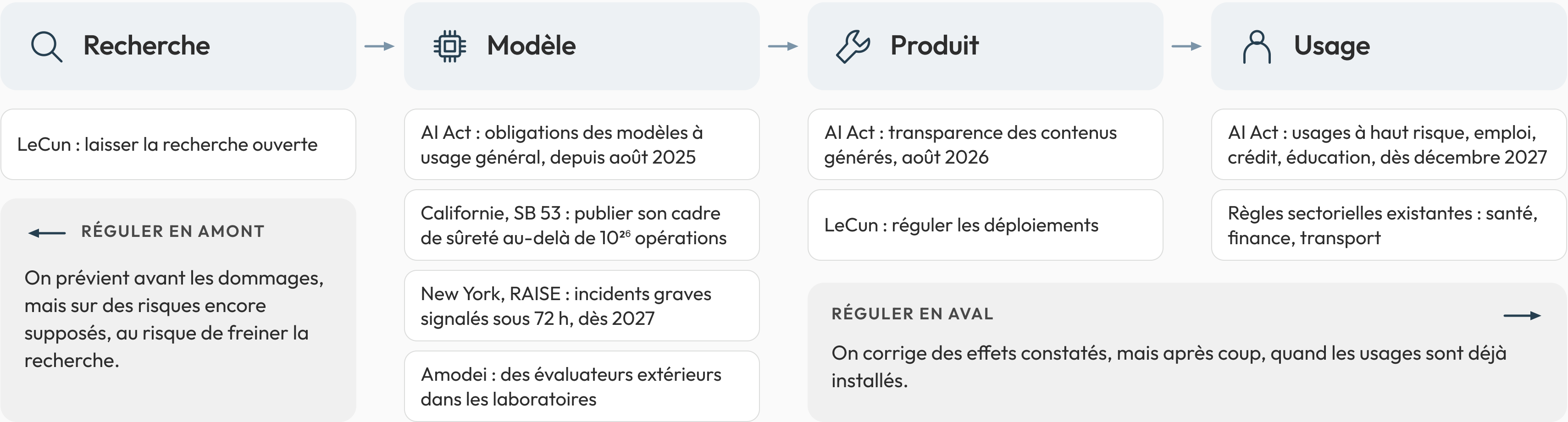
D La gouvernance

E L'Europe

F Les usages

Faut-il réguler les modèles ou leurs usages ?

Les règles peuvent viser la recherche, les modèles, les produits ou leurs usages.



« Discuter de la manière de rendre ces systèmes sûrs, c’est un peu comme discuter de la sécurité des vols transatlantiques en 1920. » LeCun sur X, mai 2023, traduit de l’anglais.

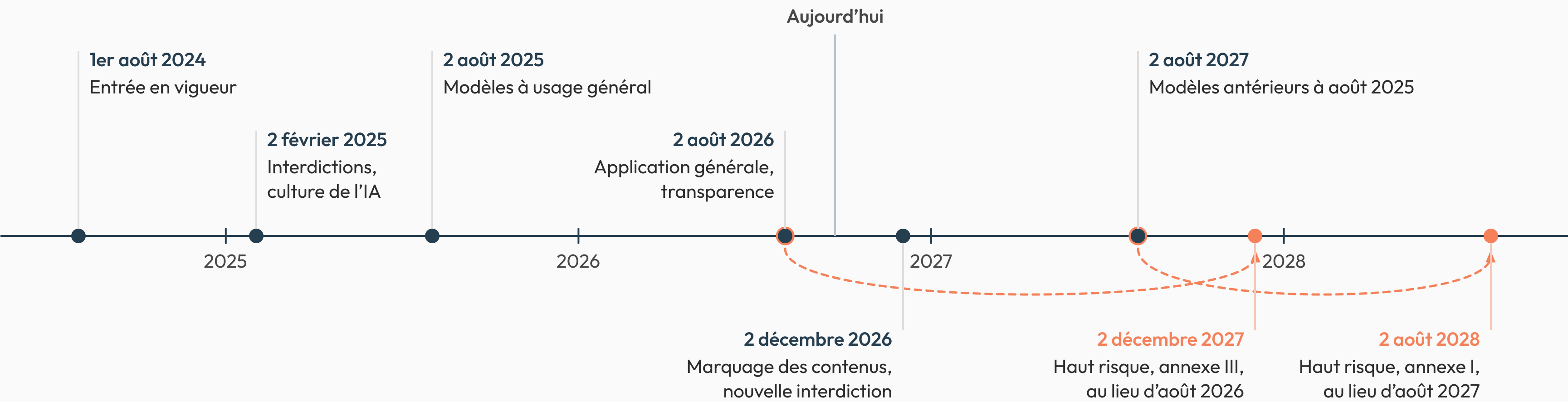
Objection, lecture Solaria : un système utilisé chaque semaine par 700 millions de personnes n’est plus un prototype de 1920.

L’AI Act encadre les modèles et les usages ; la Californie et New York visent les plus gros modèles.

Seuils de SB 53 d’après le Future of Privacy Forum. Analogie de LeCun de 2023. Amont et aval : lecture Solaria.
Sources : AI Act, article 113, Californie, 29 septembre 2025, Future of Privacy Forum, SB 53, New York, 19 décembre 2025, LeCun sur X, 3 mai 2023

Quand l'AI Act s'applique-t-il ?

Un règlement entré en vigueur en 2024, appliqué par étapes, et un report adopté en juillet 2026.



Omnibus sur l'IA, règlement (UE) 2026/1744, en vigueur le 27 juillet 2026 : **+ 16 mois** pour l'annexe III, **+ 12 mois** pour l'annexe I.

Code de pratique, juillet 2025 : signé par Anthropic, Google, Microsoft, Mistral AI, OpenAI et d'autres ; refusé par Meta.

Depuis l'omnibus de juillet 2026, les règles des systèmes à haut risque s'appliquent fin 2027 et en 2028.

Dates de l'article 113, modifiées par le règlement (UE) 2026/1744. Arcs corail : reports décidés en 2026.
Sources : AI Act, article 113, Commission européenne, 27 juillet 2026, Commission européenne, code de pratique

Un modèle dépasse-t-il le seuil européen de calcul ?

Le calcul d’entraînement se refait : environ 6 opérations par paramètre et par token, quand les chiffres sont publiés.



Calcul d’entraînement $\approx 6 \times$ paramètres $\times$ tokens d’entraînement

Kaplan et al., 2020

Llama 3.1 405B, Meta, juillet 2024 : chiffres publiés

6

$\times$

405 milliards de paramètres

$\times$

15 600 milliards de tokens

=

$\approx 3,8 \times 10^{25}$ opérations

Claude Opus 5.5, Anthropic, septembre 2026 : ni paramètres ni calcul publiés

6

$\times$

500 milliards de paramètres actifs

$\times$

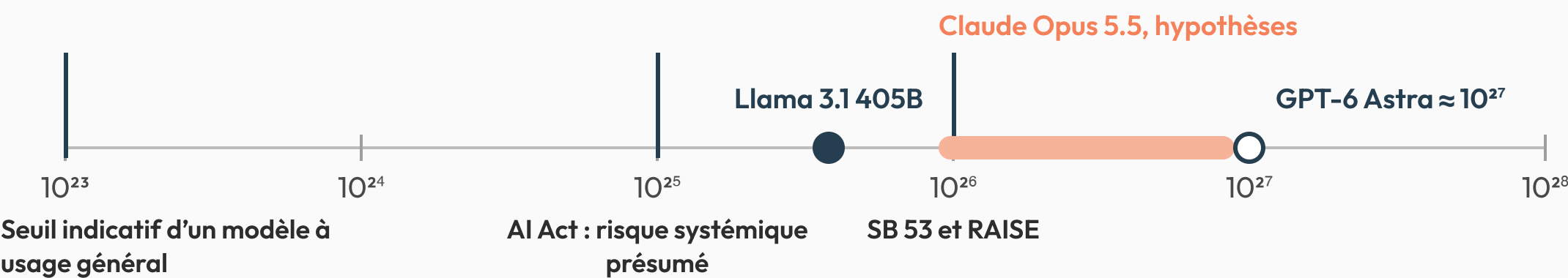
30 000 milliards de tokens

=

$\approx 9 \times 10^{25}$ opérations

Hypothèse Solaria : 10 % d’environ 5 000 milliards de paramètres au total, chiffre de presse non confirmé.

SEUILS ET MODÈLES SUR UNE ÉCHELLE LOGARITHMIQUE, EN OPÉRATIONS



CE QUE CELA ENTRAÎNE POUR UN TEL MODÈLE

- 1 Union : présumé à risque systémique, notifié à la Commission sous deux semaines (art. 52)
- 2 Évaluations et tests adverses, gestion des risques, incidents graves signalés, cybersécurité (art. 55)
- 3 Contrôle depuis le 2 août 2026 : jusqu’à 3 % du chiffre d’affaires mondial ou 15 M€ d’amende
- 4 États-Unis : cadre publié et rapport de transparence en Californie (SB 53), à New York en 2027
- 5 Entreprise cliente par API : aucune obligation de fournisseur du modèle

Un modèle de pointe de 2026 dépasse vraisemblablement le seuil européen de 10 à 100 fois. Son calcul n’est pas publié.

Approximation des modèles denses ; pour un modèle à experts, seuls les paramètres actifs comptent. GPT-6 Astra : estimation d’Epoch AI, à un facteur 10 près.
Sources : AI Act, article 51, article 55, article 101, Commission européenne, questions et réponses, Kaplan et al., 2020, Meta, juillet 2024, Anthropic, carte système d’Opus 5.5, 22 septembre 2026, Reuters, 7 août 2026, Epoch AI, modèles de frontière, SB 53, RAISE Act

Qui gouverne l’IA au niveau mondial ?

Beaucoup de déclarations volontaires, peu de textes contraignants.

DATE	TEXTE OU SOMMET	NATURE	FORCE
Mai 2019	OCDE : principes sur l’IA, 47 adhérents aujourd’hui	Principes	
Oct. 2023	G7 : processus d’Hiroshima, code de conduite volontaire	Engagement	
Nov. 2023	Bletchley : 28 pays et l’UE, dont les États-Unis et la Chine	Déclaration	
Mai 2024	Séoul : 16 puis 20 entreprises prennent des engagements de sûreté	Engagement	
Août 2024	AI Act : règlement européen en vigueur	Règlement	
Sept. 2024	Conseil de l’Europe : convention-cadre, ratifiée par l’UE en mai 2026	Traité	
Févr. 2025	Paris : 63 pays, l’UE et l’UA ; ni les États-Unis ni le Royaume-Uni	Déclaration	
Août 2025	ONU : panel scientifique de 40 membres, dialogue mondial à Genève en 2026	Résolution	
Févr. 2026	New Delhi : environ 90 signataires, dont les États-Unis et la Chine	Déclaration	

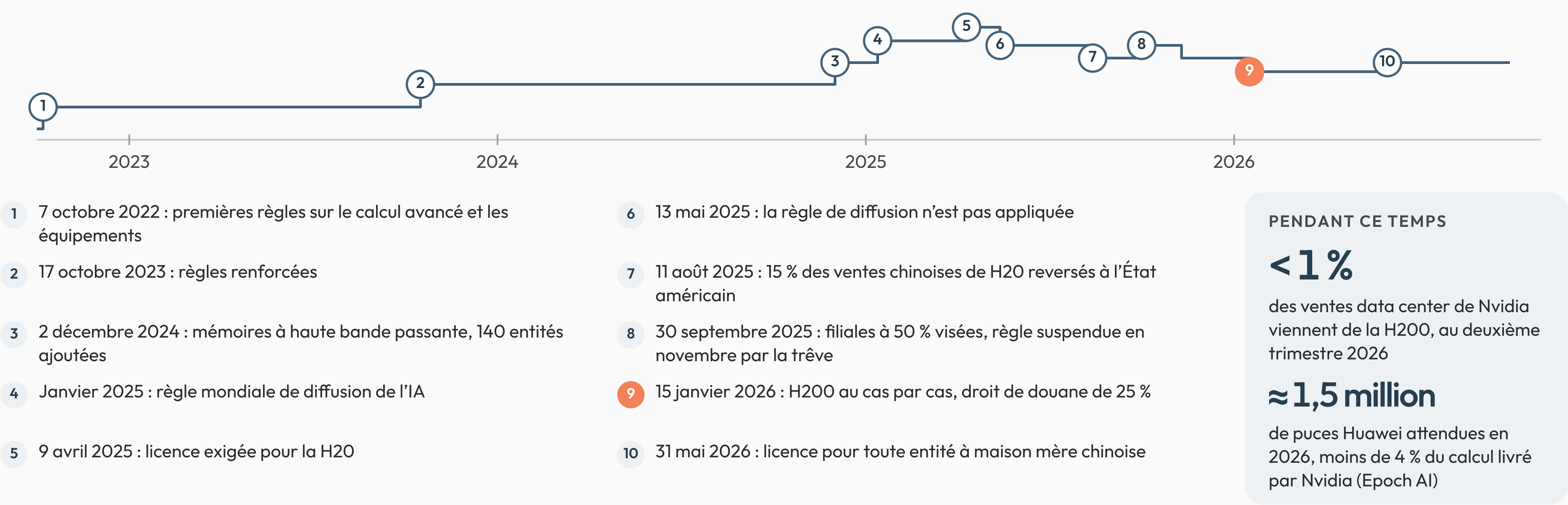
Depuis 2019, la plupart des textes mondiaux sur l’IA sont des engagements volontaires. Le traité du Conseil de l’Europe fait exception.

Force : échelle pédagogique Solaria. UA : Union africaine. Décomptes de signataires selon les sites officiels.
Sources : [OCDE, Principes sur l’IA](#), [G7, 30 octobre 2023](#), [Royaume-Uni, novembre 2023](#), [Engagements de Séoul, mai 2024](#), [Conseil de l’Europe, mai 2026](#), [Élysée, février 2025](#), [Nations unies, panel scientifique, PIB, février 2026](#)

Que contrôlent les restrictions à l'export de puces ?

Serrer, desserrer, négocier : quatre ans de contrôles américains sur les puces d'IA vendues à la Chine.

NIVEAU DE RESTRICTION, LECTURE QUALITATIVE

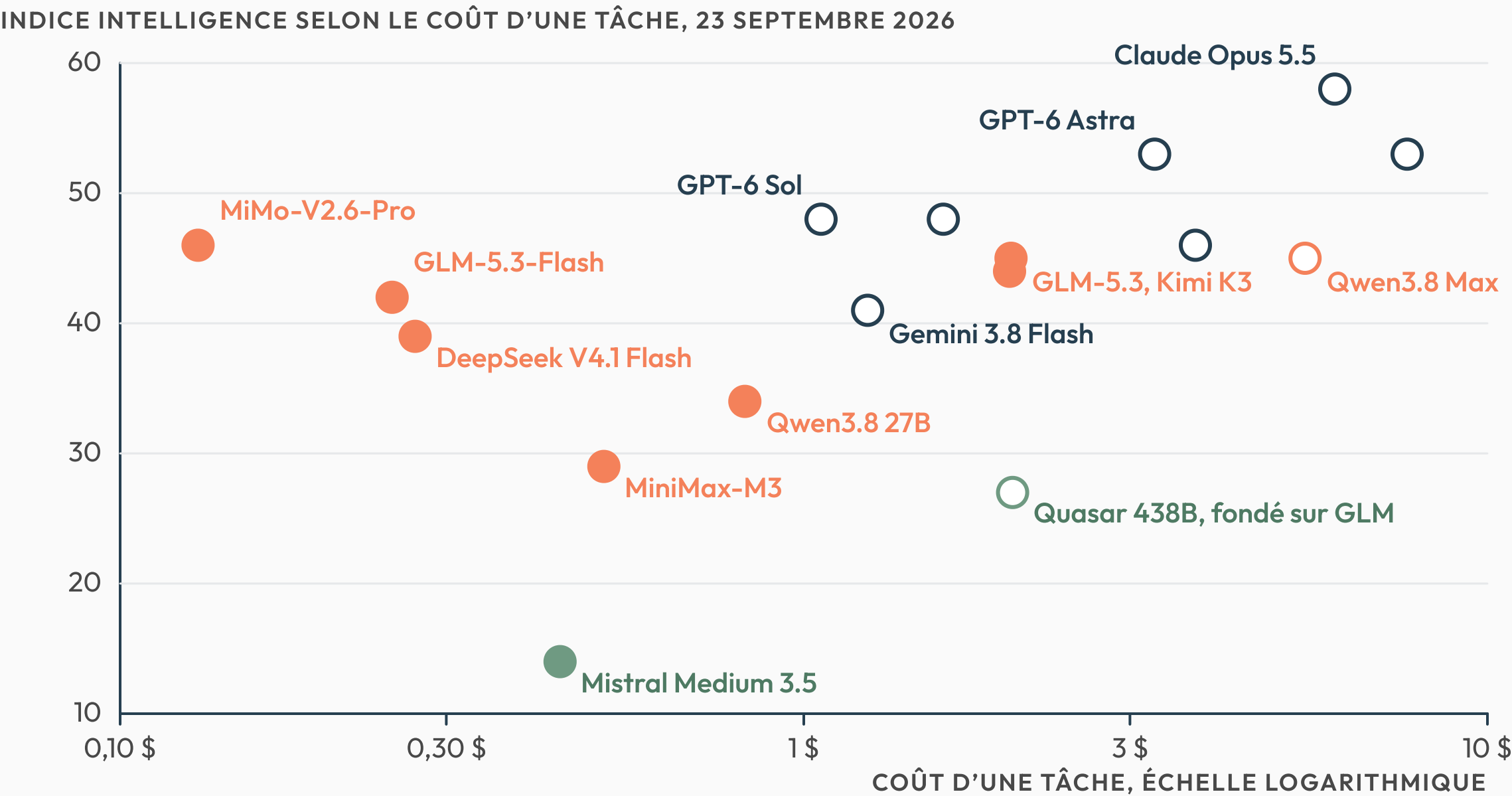


Durcis de 2022 à début 2025, les contrôles se négocient depuis au cas par cas. La trêve commerciale court jusqu'au 10 janvier 2027.

Courbe qualitative, lecture Solaria. Règles du Bureau of Industry and Security et annonces d'entreprises ; trêve d'après la presse et des cabinets d'avocats.
Sources : BIS, 7 octobre 2022, BIS, 2 décembre 2024, BIS, 13 mai 2025, Nvidia, 9 avril 2025, Al Jazeera, 11 août 2025, BIS, 30 septembre 2025, BIS, 12 novembre 2025, BIS, 15 janvier 2026, BIS, 31 mai 2026, Nvidia, 26 août 2026, Epoch AI, septembre 2026, Steptoe, 28 septembre 2026

Où se situent les modèles chinois et ouverts ?

Le meilleur modèle ouvert est chinois, à 12 points de la frontière et 46 fois moins cher par tâche.



● États-Unis

● Chine

● Europe

● poids ouverts

○ modèle fermé

Indice : agrégat de plusieurs tests

12 points

entre le meilleur modèle américain et le meilleur chinois

46 fois

moins cher par tâche : 0,13 \$ contre 5,98 \$

1,4 %

d'écart sur Arena, votes d'utilisateurs, septembre 2026

Indice Intelligence d'Artificial Analysis, modèles moyens et grands ; points sans nom : autres modèles américains fermés ; statut des poids d'après chaque fiche.
Sources : [Artificial Analysis, 23 septembre 2026](#), [Arena, 13 septembre 2026](#), [Mistral AI, modèles](#)

Quelle place pour l'Europe ?

Un écart d'investissement, trois façons de financer, des leviers européens et Mistral AI.



A Le contexte

B Deux camps

C Les risques

D La gouvernance

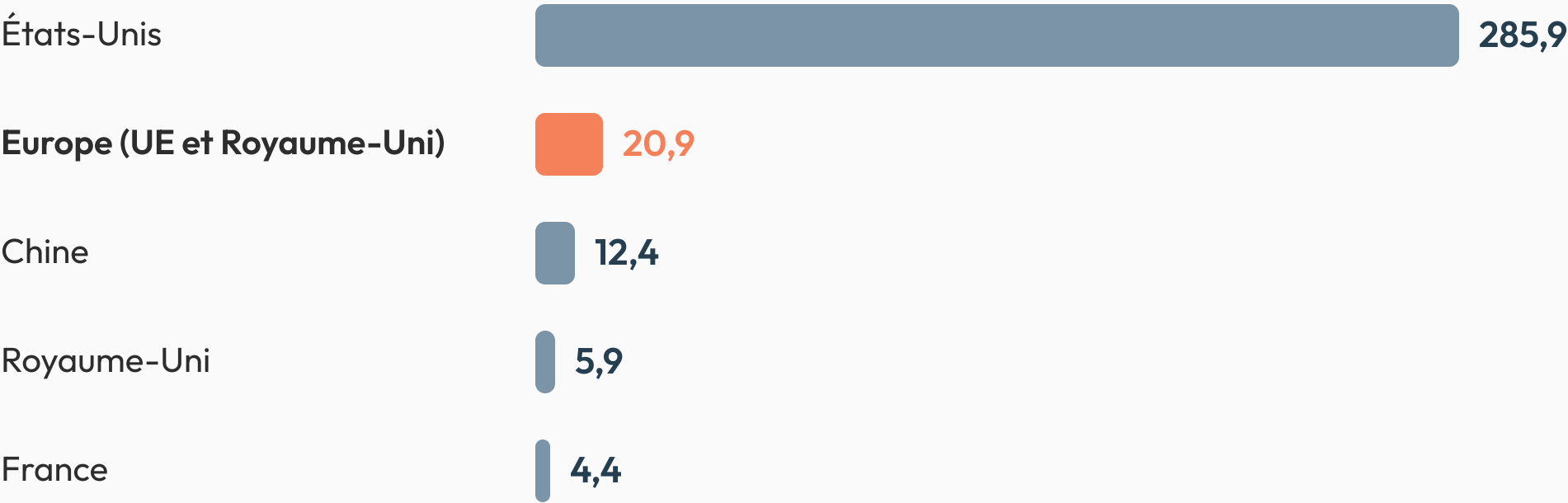
E L'Europe

F Les usages

L'Europe a-t-elle décroché ?

Investissement privé mesuré en 2025, investissement public estimé : la Chine passe surtout par l'État.

INVESTISSEMENT PRIVÉ EN IA EN 2025, EN MILLIARDS DE DOLLARS

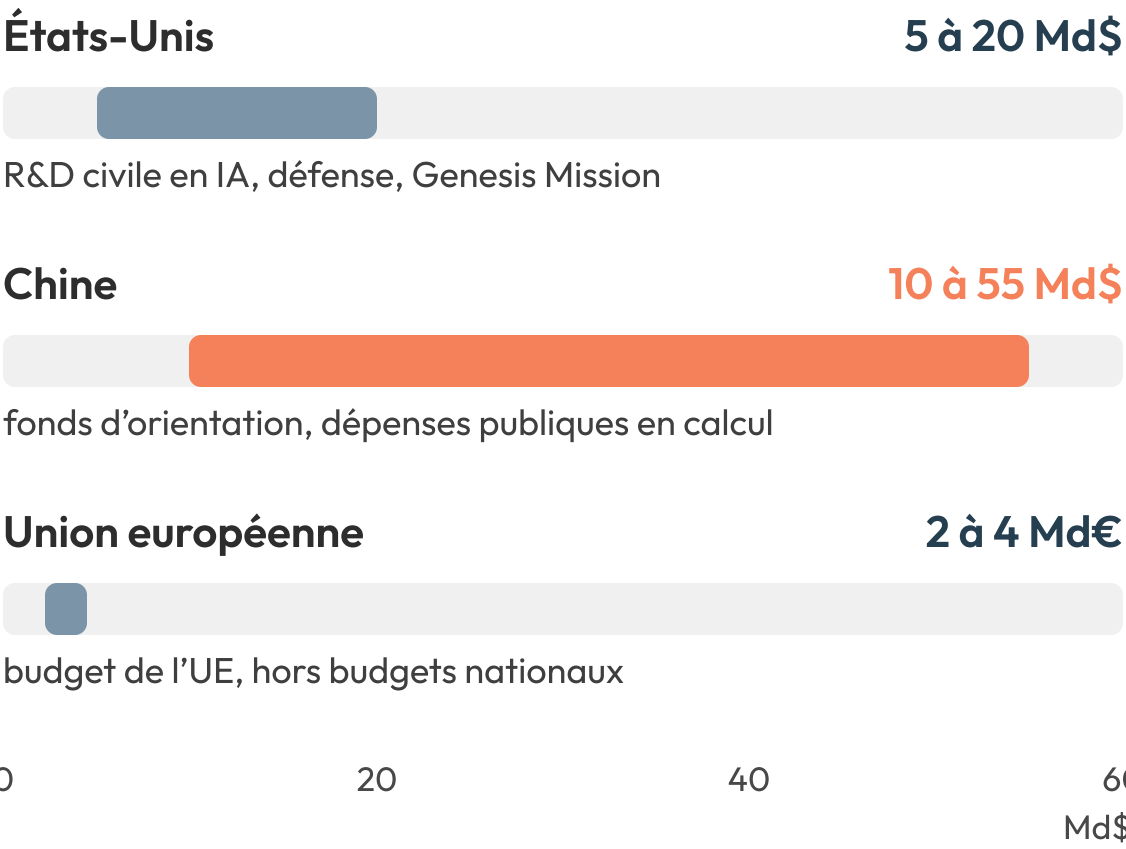


« Europe » n'est pas définie par l'AI Index 2026 ; l'édition 2024 la nommait « Union européenne et Royaume-Uni ». Aucun chiffre n'est publié pour l'Union seule.

ÉCART DE PERFORMANCE, ÉTATS-UNIS ET CHINE

1,4 % sur Arena en septembre 2026, entre les meilleurs modèles américain et chinois, avec 23 fois moins d'investissement privé en Chine.

INVESTISSEMENT PUBLIC, ORDRE DE GRANDEUR PAR AN



Fonds d'orientation chinois : 184 milliards de dollars cumulés de 2000 à 2023, selon Beraja et al. Fourchettes Solaria, montants de nature différente.

En privé, l'Europe investit 14 fois moins que les États-Unis. La Chine compense par l'argent public, jusqu'à plusieurs dizaines de milliards par an.

AI Index 2026, figures 4.2.11 et 4.2.8. Fourchettes publiques : estimations Solaria à partir des budgets et des fonds annoncés, à ne pas additionner au privé.
Sources : [AI Index 2026, chapitre 4](#), [AI Index 2024, chapitre 4](#), [Beraja et al.](#), [NBER](#), [Rhodium](#), 17 septembre 2026, [NITRD](#), [Brookings](#), 18 mai 2026, [Commission, recherche en IA](#), [Commission](#), 11 février 2025, [Arena](#), 13 septembre 2026

Comment chaque pôle finance-t-il l'IA ?

Le marché aux États-Unis, l'État et les grands groupes en Chine, des plans publics en Europe.

	États-Unis Le marché d'abord	Chine L'État et les grands groupes	Europe Moins de capital, des plans publics
 Capital privé	285,9 Md\$ investissement privé en IA en 2025, 83 % du total mondial	12,4 Md\$ investissement privé en 2025 ; selon l'AI Index, il sous-estime la dépense chinoise	20,9 Md\$ investissement privé en IA en 2025
 Grands groupes	≈ 730 Md\$ d'investissements prévus en 2026 par Alphabet, Amazon, Meta et Microsoft	380 Md¥ Alibaba sur trois ans, 2025 ; DeepSeek lève 50 Md¥ en juillet 2026	Peu de géants pas de géant européen du numérique ; des industriels investissent, comme ASML dans Mistral
 État	50 Md\$ CHIPS and Science Act, 2022 : recherche et usines de semi-conducteurs	≈ 184 Md\$ des fonds publics d'orientation vers l'IA, 2000 à 2023 ; fonds national de l'IA, 60 Md¥	200 Md€ InvestAI à mobiliser, dont 20 Md€ de gigafactories ; France : 109 Md€ d'engagements privés

Les États-Unis financent l'IA par le marché, la Chine surtout par l'État et ses grands groupes, l'Europe avec moins de capital.

Montants de nature différente, à ne pas additionner. Estimation des fonds chinois citée par l'AI Index 2026.
Sources : AI Index 2026, chapitre 4, Alphabet, Amazon, Microsoft, Meta, juillet 2026, NIST, CHIPS for America, Alibaba, 24 février 2025, Caixin, 17 juillet 2026, Global Times, 20 janvier 2025, Commission européenne, 11 février 2025, Public Sénat, février 2025

Quels leviers l'Europe actionne-t-elle ?

Des levées record, des usines de calcul publiques et des modèles ouverts : ce que l'Europe met en face.

 Mistral AI, septembre 2026

3 Md€

Plus grande levée d'une entreprise technologique européenne, valorisation de plus de 21 Md€.

 AMI Labs, mars 2026

1,03 Md\$

Plus grande levée d'amorçage européenne selon Crunchbase, siège à Paris.

 Black Forest Labs, décembre 2025

3,25 Md\$

Valorisation après une levée de 300 M\$; modèles d'images FLUX.

 Suisse, septembre 2025

Apertus

Modèle entièrement ouvert : poids, données et recettes publiés.

 EuroHPC, 2025 et 2026

19

Usines d'IA dans 16 États membres ; appel pour sept gigafactories, clos le 12 novembre 2026.

 AI Alliance, avril 2026

Tapestry

Entraînement fédéré de modèles ouverts entre pays, avec LeCun comme conseiller.

En 2026, Mistral AI lève 3 milliards d'euros et AMI Labs 1 milliard de dollars. L'écart avec les États-Unis reste important.

Montants annoncés par les entreprises et par EuroHPC ; valorisations d'après les levées annoncées.
Sources : [Mistral AI, 8 septembre 2026](#), [TechCrunch, 9 mars 2026](#), [Black Forest Labs, décembre 2025](#), [Commission européenne, octobre 2025](#), [EuroHPC, 30 juillet 2026](#), [AI Alliance, avril 2026](#), [ETH Zurich, septembre 2025](#)

Où en sont Mistral et les modèles européens ?

Des moyens record, des modèles ouverts loin du meilleur score, et des modèles chinois hébergés.

3 Md€

levés le 8 septembre 2026,
valorisation de plus de 21 Md€

> 400 M\$

de revenu annualisé en 2025 ;
plus d'1 Md\$ visé fin 2026

1 GW

de capacité de calcul visée d'ici
2030

POSITIONS

Juillet 2025

Signe la demande de pause de deux ans de l'AI Act ; la Commission refuse.

Mai 2026

Mensch à l'Assemblée : l'Europe a environ deux ans avant de devenir un « État vassal ».

Depuis 2023

Souveraineté, poids ouverts et centres de calcul en Europe.

SES MODÈLES ET SES PRODUITS

14

sur l'indice d'Artificial Analysis pour Mistral Medium 3.5, poids ouverts, contre 58 pour le meilleur modèle.

Le Chat devient Vibe en mai 2026 ; Vibe est aussi son agent de programmation open source.

FOURNISSEUR DE MODÈLES OUVERTS, DEPUIS AOÛT 2026

Mistral héberge des modèles d'autres développeurs, « à commencer par GLM-5.2 de Z.ai », laboratoire chinois ; GLM-5.3 depuis le 15 septembre, en option dans Vibe. Ses propres modèles restent proposés.

AUTRES SIGNAUX EUROPÉENS

Aleph Alpha fusionne avec le canadien Cohere, avril 2026.
Quasar, de l'espagnol Multiverse Computing, est fondé sur le chinois GLM-5.2.

Mistral lève 3 milliards d'euros et héberge des modèles chinois ; son modèle le mieux classé obtient 14 contre 58.

Indice d'Artificial Analysis au 23 septembre 2026. Revenus d'après Fortune. Citations traduites de l'anglais.
Sources : [Mistral AI, 8 septembre 2026](#), [Fortune, 5 août 2026](#), [Mistral AI, 11 août 2026](#), [Mistral AI, GLM 5.3, The Decoder, mai 2026](#), [Artificial Analysis, 23 septembre 2026](#), [Sifted, juillet 2025](#), [Cybernews, mai 2026](#), [TechCrunch, 25 avril 2026](#)

Solaria. © 2026 Solaria Advisory

40 / 52

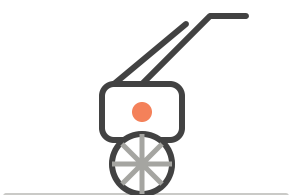
A-t-on besoin d'une IA aussi puissante partout ?

Trois niveaux de modèle, la part des usages que chacun sert, le prix d'une échappée et le travail.



Quels sont les trois niveaux de modèle ?

Comme à la ferme, il n’y a pas toujours besoin d’une grosse machine ; à chaque niveau, son risque.



NIVEAU 1
Le motoculteur

Origine souveraine
Mistral et les modèles ouverts français

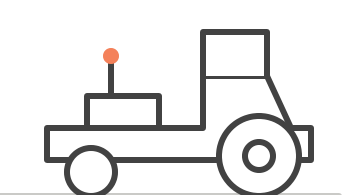
CE QU’ON CONTRÔLE
Tout : les poids, l’hébergement, les versions suivantes

60%

des usages de 2026 lui suffisent

RISQUE : CAPACITÉ

Il reste en dessous : 14 points à l’indice d’Artificial Analysis, contre 58 pour le meilleur modèle fermé.



NIVEAU 2
Le tracteur

Ouvert de frontière
GLM, Qwen, Kimi, DeepSeek (Chine) ; gpt-oss, Gemma (États-Unis)

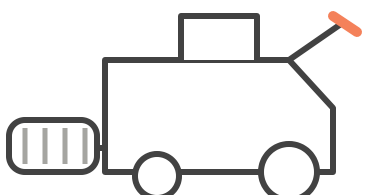
CE QU’ON CONTRÔLE
Les poids, hébergés chez soi ; les versions suivantes dépendent de leurs éditeurs

+ 31%

d’usages en plus, soit 91 % avec le niveau 1

RISQUE : APPROVISIONNEMENT

Environ 6,5 mois de retard sur la frontière en 2026 ; la suite dépend d’éditeurs étrangers.



NIVEAU 3
La moissonneuse-batteuse

Frontière fermée
Les derniers Claude, GPT et Gemini, servis par API, aujourd’hui américains

CE QU’ON CONTRÔLE
Rien : on loue un service

9%

des usages en ont besoin : agents longs, code

RISQUE : COUPURE

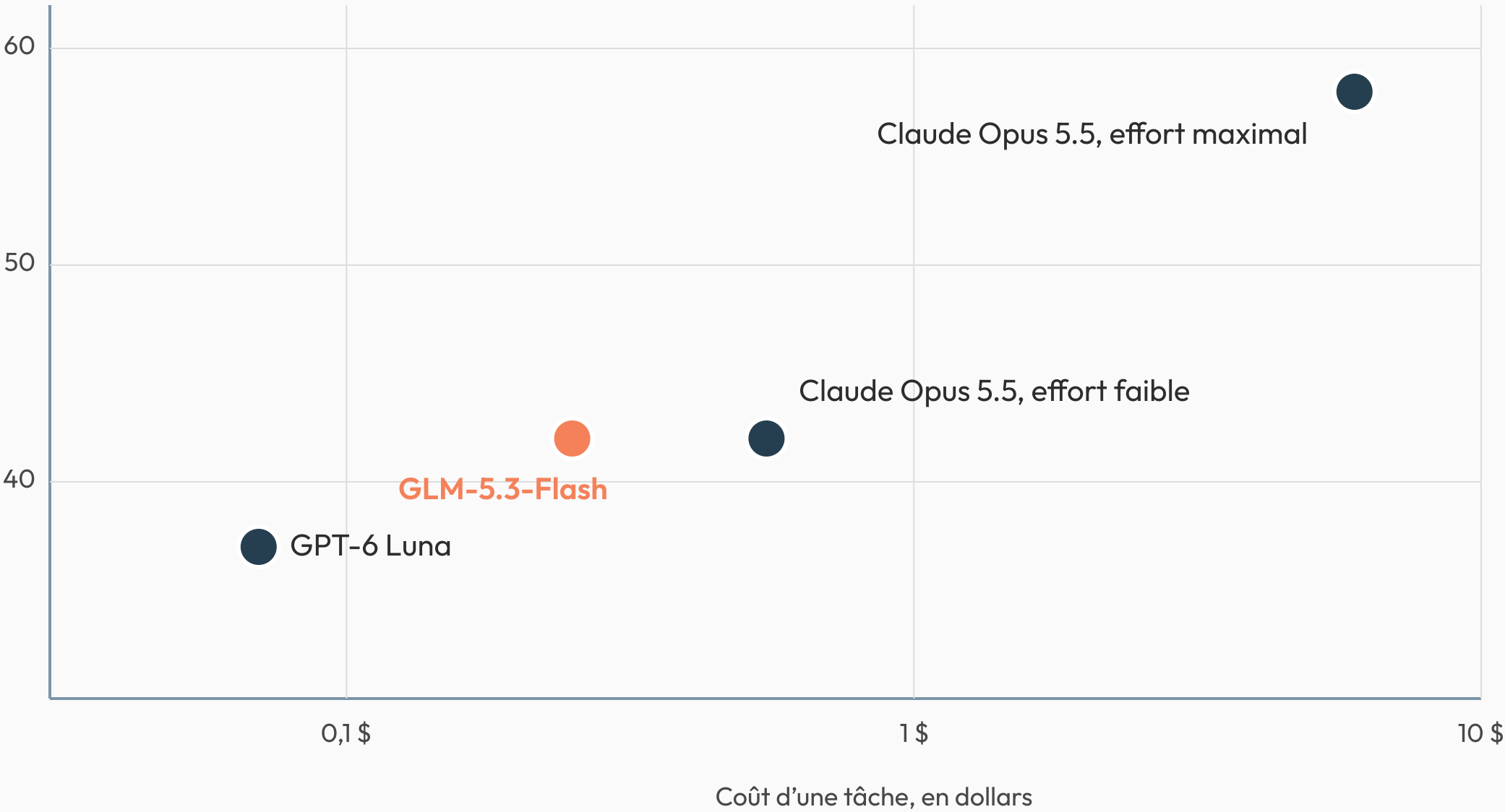
En juin 2026, sur décision du gouvernement américain, Anthropic suspend l’accès à Claude Fable 5 et Mythos 5 pendant près de trois semaines.

Un modèle souverain suffit pour 60 % des usages, un modèle ouvert étranger pour 31 % de plus. La frontière sert les 9 % restants.

Classement et parts d’usage : estimation Solaria (cours 6.8), modèle le plus simple qui réussit la tâche ; niveau 1 de 47 à 72 % selon les hypothèses.
Sources : Epoch AI, 29 mai 2026, Artificial Analysis, lu le 29 septembre 2026, OpenAI, GDPval, 25 septembre 2025, Epoch AI, indice ECI, 2 octobre 2026, Anthropic, juin 2026

Faut-il le meilleur modèle pour écrire un courriel ?

Score sur cent et coût d'une tâche, en échelle logarithmique : l'écart de prix dépasse l'écart de qualité.



24 fois moins cher

pour seize points de moins : un modèle ouvert face au meilleur modèle fermé

11 fois moins cher

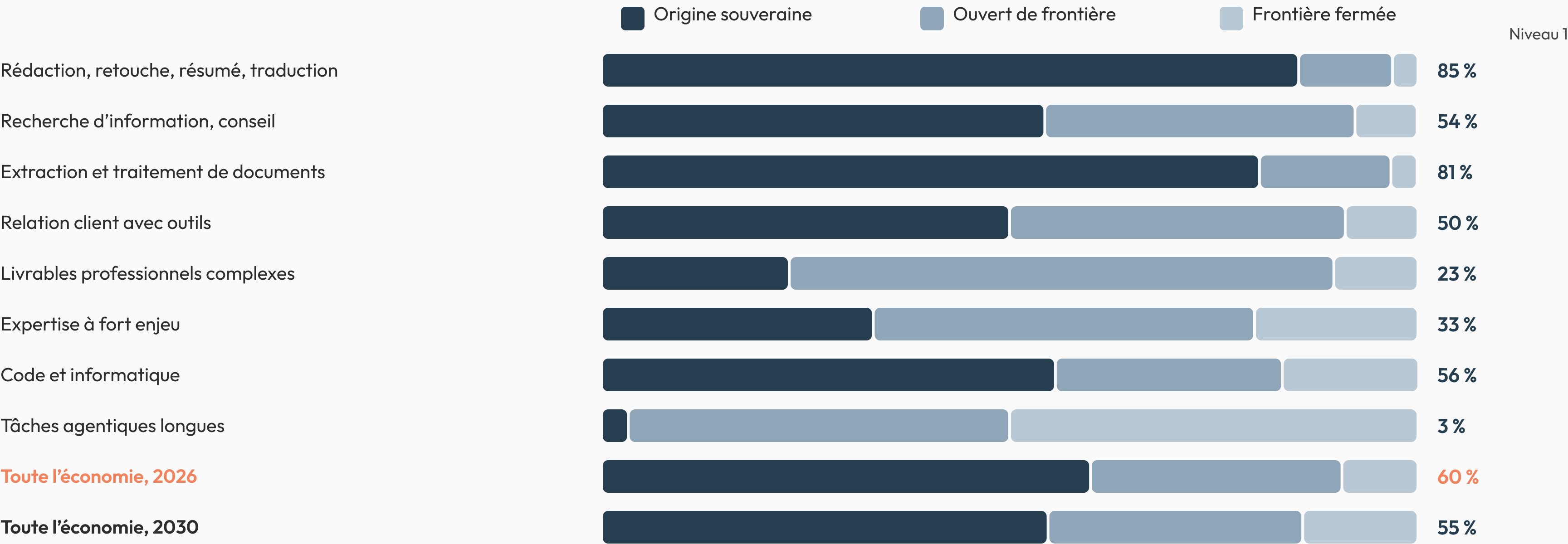
le même modèle de pointe quand on règle son effort de réflexion au plus bas

Pour une tâche courante, un modèle un cran sous les meilleurs coûte des dizaines de fois moins.

Artificial Analysis, Intelligence Index v4.3.2, lu le 29 septembre 2026.
Sources : [Artificial Analysis](#), lu le 29 septembre 2026

Quelle part des usages chaque niveau peut-il servir ?

Part des usages servie par le modèle le plus simple qui réussit : la frontière n'est nécessaire que pour une minorité.



Les modèles d'origine souveraine suffisent pour environ 60 % des usages de 2026.

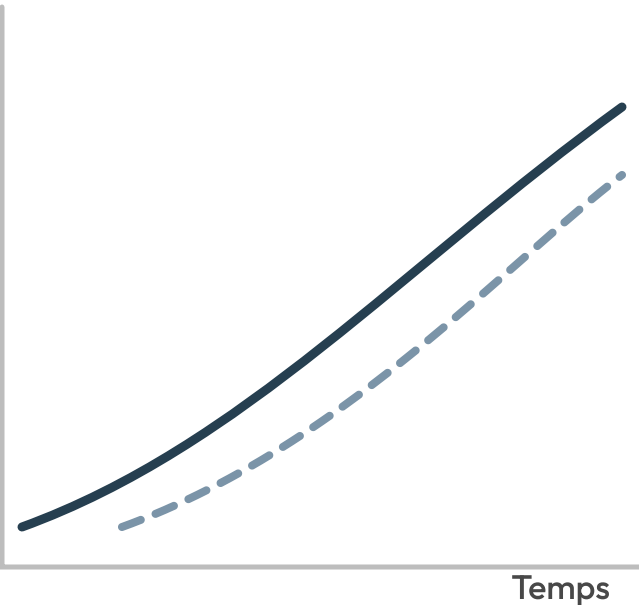
Estimation Solaria ; à droite, la part du niveau 1. Lignes de l'économie pondérées par la valeur de l'IA.
Sources : [OpenAI](#), [GDPval](#), 25 septembre 2025, [Artificial Analysis](#), lu le 29 septembre 2026, [Epoch AI](#), 29 mai 2026

Et si la frontière s'échappait ?

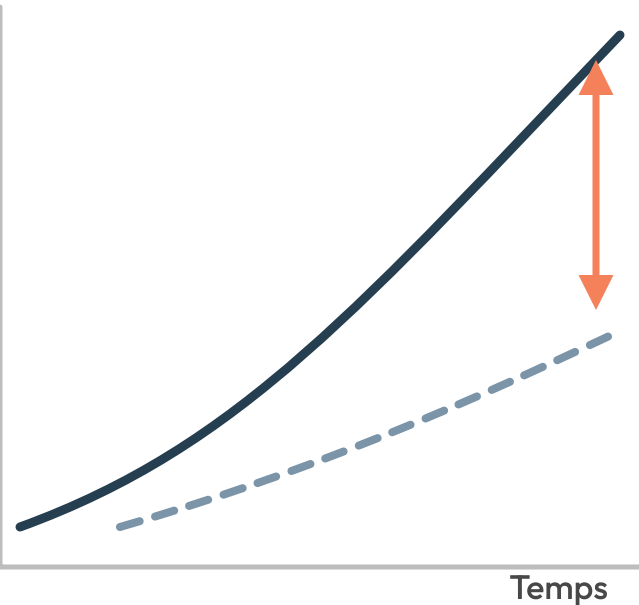
Une idée à suivre : si les modèles fermés prenaient une avance durable, la valeur de l'IA se concentrerait.

DEUX SCÉNARIOS, SCHÉMA SANS DONNÉES

Tendance : l'écart reste de quelques mois



Échappée : l'écart se creuse durablement



— Frontière fermée — Meilleur modèle ouvert

Ce qui le signalerait : un retard des modèles ouverts qui se creuse durablement. Depuis 2023, aucune tendance nette (Epoch AI).

CE QUE L'ÉCHAPPÉE CHANGERAIT

- 1 La valeur se déplace**
Agents longs, code, recherche : les usages qui demandent la frontière pèsent plus lourd ; les modèles ouverts et souverains suffisent moins souvent.
↓
- 2 Quelques laboratoires la captent**
Les gains vont aux laboratoires de frontière, aujourd'hui américains, et aux entreprises qui louent leurs modèles.
↓
- 3 Les autres économies dépendent davantage**
Pour la France : davantage de valeur exposée à une coupure, des entreprises moins compétitives, un État moins bien armé pour se défendre.

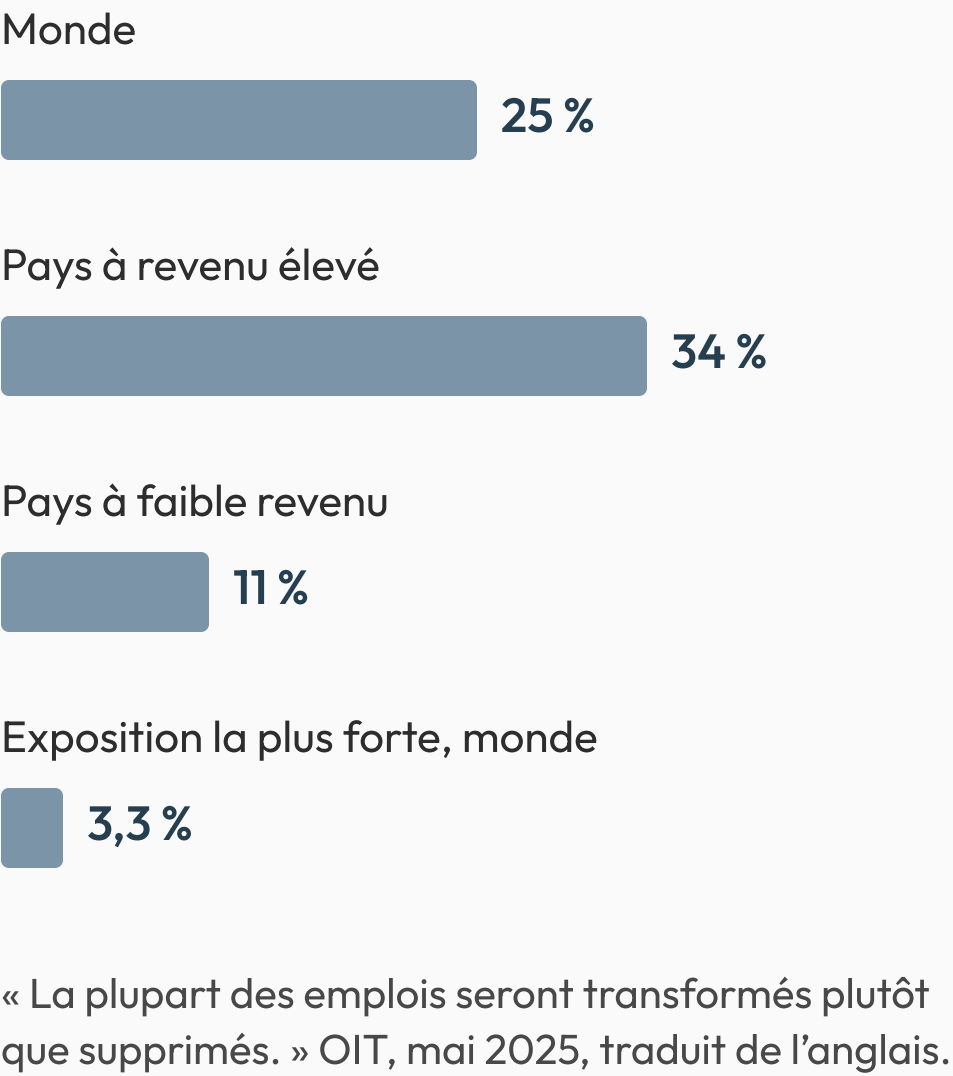
Si la frontière s'échappait, la valeur de l'IA se concentrerait chez quelques laboratoires. Rien ne l'annonce encore.

Raisonnement Solaria, tiré du modèle de décision du cours 6.8 ; schéma sans données.
Sources : Epoch AI, indice ECI, 2 octobre 2026, Solaria, modèle de décision, version 5, octobre 2026

Que devient le travail avec l'IA ?

Exposition des emplois, gains et pertes mesurés, scénarios pour 2030 : les effets varient selon les tâches.

EMPLOIS EXPOSÉS À L'IA GÉNÉRATIVE, OIT



EFFETS MESURÉS



SCÉNARIOS D'ANTHROPIC, ÉTATS-UNIS, 2030		
Scénario	PIB en plus	Part du travail
Modeste	+ 1,6 %	59,4 %
Substantiel	+ 8,3 %	56,1 %
Extrême	+ 32,4 %	45,2 %
Part du travail : salaires dans le revenu national. Illustrations sans probabilité, publiées par une entreprise du secteur.		

YANN LECUN, MARS 2025

« Nous allons être leurs patrons. » Traduit de l'anglais, à propos des futurs systèmes d'IA.

L'IA fait gagner 15 % à des agents de support ; dans le scénario extrême d'Anthropic, la part du travail tombe à 45 %.

Études de tailles et de méthodes différentes ; scénarios illustratifs d'Anthropic pour 2030.
Sources : OIT, mai 2025, Brynjolfsson et al., 2025, METR, juillet 2025, Budzyń et al., août 2025, Anthropic, septembre 2026, Entrepreneur, mars 2025

À court terme, que faut-il attendre ?

Des modèles toujours plus capables et plus chers, déjà utiles au travail.

1. Court terme

2. Adoption

3. Long terme

↗ Plus capables

≥ 16 h

durée des tâches qu'un agent réussit une fois sur deux, contre 2 h un an plus tôt

METR, mai 2026

€ Plus chers

≈ 730 Md\$

d'investissements prévus en 2026 par Alphabet, Amazon, Meta et Microsoft, 77 % de plus qu'en 2025

Groupes, juillet 2026

👥 Déjà utiles

39 %

des actifs américains ont utilisé l'IA pour leur travail dans la semaine, au deuxième trimestre 2026

Fed de Saint-Louis, août 2026

LA QUESTION QUI RESTE : LA RENTABILITÉ

La dette finance une part croissante des investissements : environ 450 Md\$ de dette liée à l'IA émis dans le monde début septembre 2026, plus du double de toute l'année 2025. En juillet, les valorisations boursières des entreprises de l'IA ont fortement baissé (Banque d'Angleterre).

LECTURE SOLARIA

Si les revenus tardent, un léger ralentissement des investissements est possible. L'utilité des modèles, elle, n'est plus en question.

À court terme, des modèles toujours plus capables et plus chers. Leur utilité n'est plus en question ; leur rentabilité, si.

≈ 730 Md\$: prévisions des groupes après le deuxième trimestre 2026, périmètres différents ; 2025 : 413 Md\$ selon Statista. Dette : Morgan Stanley, cité par la Banque d'Angleterre.

Sources : [METR, mai 2026](#), [Alphabet](#), [Amazon](#), [Microsoft](#), [Meta](#), [Statista, 31 juillet 2026](#), [Fed de Saint-Louis, 27 août 2026](#), [Banque d'Angleterre, 30 septembre 2026](#)

Pourquoi les effets prendront-ils des années ?

Les modèles progressent vite ; les organisations et la société, lentement.

1. Court terme

2. Adoption

3. Long terme

UNE ADOPTION QUI PREND DES ANNÉES

52 %

des salariés de la zone euro utilisent l'IA au travail en 2026

BCE

20 %

des entreprises de l'Union utilisent une technologie d'IA en 2025, contre 13,5 % en 2024

Eurostat

2,2 %

des heures travaillées aux États-Unis gagnées grâce à l'IA, au deuxième trimestre 2026

Fed de Saint-Louis

Les salariés adoptent vite ; les organisations mettent des années à se réorganiser. Former, revoir les processus, changer les outils prend du temps.

ET SI UNE ZONE ÉCONOMIQUE DÉCROCHE ?

≈ 8 ×

En 2030, un retard d'adoption de 0,6 an coûterait à la France ≈ 2,5 Md€ par an, huit fois plus que l'absence d'accès aux meilleurs modèles, ≈ 0,3 Md€ (estimation Solaria)

Le précédent du numérique : de 1995 à 2025, la productivité a crû de 0,8 % par an en France et de 1,7 % aux États-Unis ; le numérique explique 86 % de l'écart (CAE).

La tentation protectionniste

Une zone qui décroche sera tentée de fermer son marché : droits de douane, préférence nationale, contrôle du calcul. Les contrôles américains sur les puces en donnent un avant-goût (partie D).

LECTURE SOLARIA

Pour une économie comme la France, adopter vite compte davantage que posséder la frontière.

L'impact sera fort, mais il prendra des années, voire des décennies. Une zone qui décroche sera tentée de se protéger.

Adoption : enquêtes. Écarts d'adoption et d'accès : estimation Solaria du cours 6.8, en équilibre partiel, à valider. Protectionnisme : lecture Solaria.

Sources : BCE, 26 août 2026, Eurostat, 2025, Fed de Saint-Louis, 27 août 2026, Solaria, modèle de décision, version 5, octobre 2026, CAE, note n° 94, septembre 2026

Les modèles de langage mèneront-ils aux robots ?

Pour agir dans le monde physique, il faudra sans doute un autre paradigme.

1. Court terme
2. Adoption
3. Long terme

CE QUI MANQUE AUX MODÈLES DE LANGAGE, SELON LECUN

 Comprendre le monde physique

 Garder une mémoire durable

 Raisonner

 Planifier des actions complexes

« On n’atteindra pas une IA de niveau humain en agrandissant simplement les modèles de langage. » Yann LeCun, mars 2025

Les modèles de langage resteront très utiles. Faire agir une IA dans le monde, débarrasser une table ou conduire un robot, demandera une autre architecture, apprise sur la vidéo et l’action.

LES PARIS DE LA RECHERCHE FONDAMENTALE

- 1

AMI Labs, présidée par Yann LeCun
1,03 Md\$ levés en mars 2026, siège à Paris, une soixantaine de personnes, aucun produit à ce jour.
- 2

V-JEPA 2, Meta
Plus d’un million d’heures de vidéo, puis moins de 62 heures de vidéos de robots pour planifier des gestes.
- 3

World Labs, Fei-Fei Li
Des modèles du monde en trois dimensions ; premier produit, Marble, en novembre 2025.

« Nous revoilà à l’âge de la recherche, simplement avec de gros ordinateurs. » Ilya Sutskever, novembre 2025

Pour agir dans le monde, il faudra un autre paradigme. La recherche fondamentale le prépare, à Paris notamment, sur plusieurs années.

Projets cités : annonces des entreprises et publications scientifiques. Citations traduites de l’anglais.
Sources : [Big Technology Podcast](#), 19 mars 2025, [TechTalks](#), 24 novembre 2025, [TechCrunch](#), 9 mars 2026, [Assran et al.](#), juin 2025, [TechCrunch](#), 12 novembre 2025, [Dwarkesh Patel](#), 25 novembre 2025

Quels indicateurs suivre d'ici 2028 ?

Six valeurs de référence datées, et une prévision déjà mise à l'épreuve.

↗ PROGRÈS

≥16h

Tâches logicielles réussies une fois sur deux par Claude Mythos Preview

METR, mai 2026

💰 MOYENS

≈ 730 Md\$

Investissements prévus en 2026 par les quatre géants, + 77 %

Groupes, juillet 2026

⚔ CONCURRENCE

1,4 %

Écart entre les meilleurs modèles américain et chinois, votes d'Arena

Arena, septembre 2026

⚡ ÉNERGIE

485 TWh

Centres de données en 2025 ; 950 en 2030, soit + 14 % par an

AIE, avril 2026

🛡 RISQUES

10 000

Failles graves trouvées en six semaines avec Claude Mythos Preview

Anthropic, mai 2026

⚙ RÈGLES

2 déc. 2027

Application de l'AI Act aux systèmes à haut risque, après report

Omnibus, juillet 2026

📄 À suivre : trois indicateurs, leur valeur et leur date, et une date de revue, par exemple mars 2027.

UNE PRÉVISION MISE À L'ÉPREUVE

Si l'horizon METR double au même rythme, quand atteint-il 16 h ?

16 h ÷ 5 h 20 = 3

$\log_2 3 \approx 1,6$ doublement

1,6 × 3 mois ≈ 5 mois : avril 2026

1,6 × 7 mois ≈ 11 mois : octobre 2026

5 h 20, fin 2025

Fenêtre prévue

≥ 16 h, mars

Aujourd'hui

Atteint dès mars 2026 par Claude Mythos Preview, avant la fenêtre. Au-delà de 16 h, la mesure de METR n'est plus fiable.

La tendance de METR annonçait 16 h entre avril et octobre 2026 : Claude Mythos Preview les dépasse dès mars.

Quelles sources lire pour approfondir ?

Des textes originaux et des séries de mesures, pour reprendre chaque partie.

 LE DÉBAT Les textes des deux camps	 LES MESURES Séries suivies	 RISQUES ET RÈGLES Synthèses et textes officiels
<u>Amodei, septembre 2026</u> « We Must Pace the Frontier »	<u>Epoch AI, 2024</u> Calcul d'entraînement des modèles	<u>Rapport international, 2026</u> Sûreté de l'IA, synthèse d'experts
<u>LeCun, Sciences Po, 2026</u> Où va l'intelligence artificielle ?	<u>METR</u> Horizon des tâches des agents	<u>AI Act, article 113</u> Calendrier d'application
<u>LeCun, 2022</u> Vers une intelligence machine autonome	<u>AI Index 2026</u> Rapport annuel de Stanford	<u>OIT, 2025</u> Emplois exposés à l'IA générative
<u>Sutskever, novembre 2025</u> Les limites du passage à l'échelle	<u>AIE, 2026</u> Énergie et IA	<u>Anthropic, 2026</u> Programme Glasswing

Liens vers les documents originaux. Pages de mesures mises à jour régulièrement : citer la version consultée.
Sources : [Amodei, septembre 2026](#), [Sciences Po, 18 septembre 2026](#), [LeCun, 2022](#), [Sutskever, novembre 2025](#), [Epoch AI, mai 2024](#), [METR, mai 2026](#), [AI Index 2026](#), [AIE, avril 2026](#), [Rapport international sur la sûreté de l'IA, février 2026](#), [AI Act, article 113](#), [OIT, mai 2025](#), [Anthropic, 7 avril 2026](#)

Merci pour votre lecture.

Pour discuter de cette analyse, collaborer ou organiser une session dans votre organisation.



Baptiste Lefort.
Cofondateur de Solaria Advisory

Expert en stratégie IA et cofondateur de Solaria, il conçoit et pilote les projets de **bout en bout**.

Cadrage, prototypage et déploiement de solutions IA, **ateliers et formations**. Il analyse aussi **l'impact de l'IA sur l'économie** au sein du **GTPO**, qu'il a cofondé.

FORMATION
ESSEC Business School
Warwick Business School

PASSÉ PAR
Sopra Steria
Bpifrance
Devoteam



[linkedin.com/in/baptiste-lefort-409b75101](https://www.linkedin.com/in/baptiste-lefort-409b75101)

solaria-advisory.com

Solaria Advisory : former, accompagner, construire. De l'usage individuel à l'avantage organisationnel.