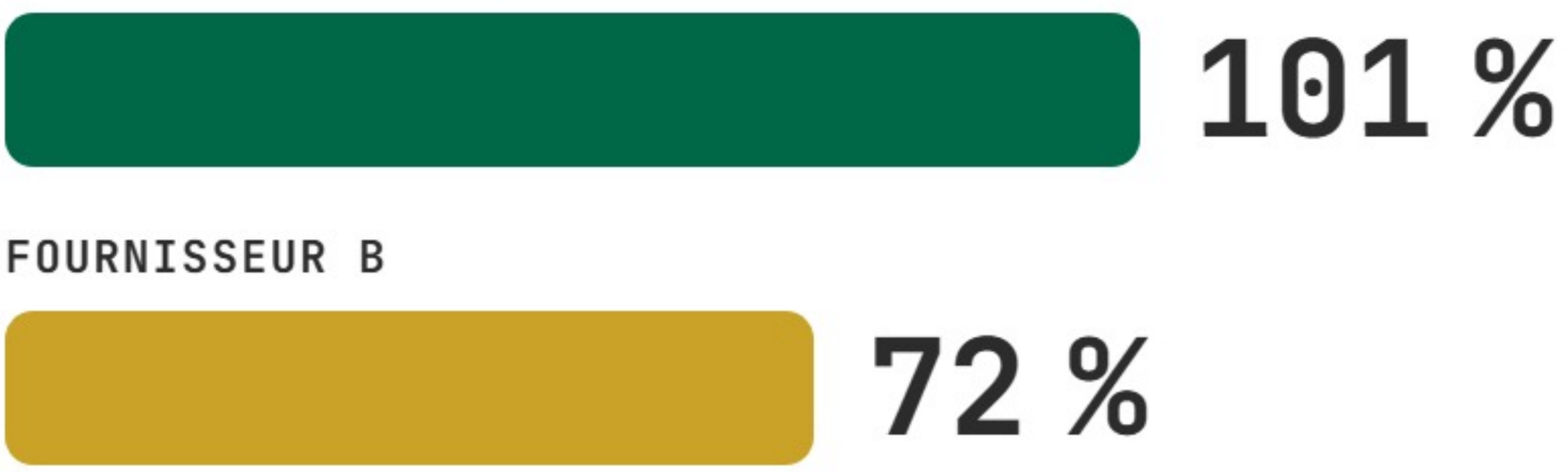
IA · ACHATS · CONTRATS

Réceptionnez votre IA comme une machine.

MÊME MODÈLE, MÊME NOM, DEUX FOURNISSEURS



Part de la performance du modèle de référence.

Le test qui le voit coûte un dollar.
Presque personne ne le fait.

Artificial Analysis, octobre 2026

NABIL EL MAHYAOUI

« Ce modèle, c’est bien celui que je paie ? »

UNE MACHINE-OUTIL	UN MODÈLE D’IA
• Essais sur site	• Une clé d’accès
• Procès-verbal	• Un nom
• Réserves à lever	• Une facture
• Signature	• **Aucun essai**

On achète l’IA comme un service. Elle se comporte comme un équipement : **elle peut changer sans que son nom change.**

Entre le nom et la réponse, six maillons.

1	**Le modèle**	Les paramètres appris. Le seul maillon que le nom désigne.

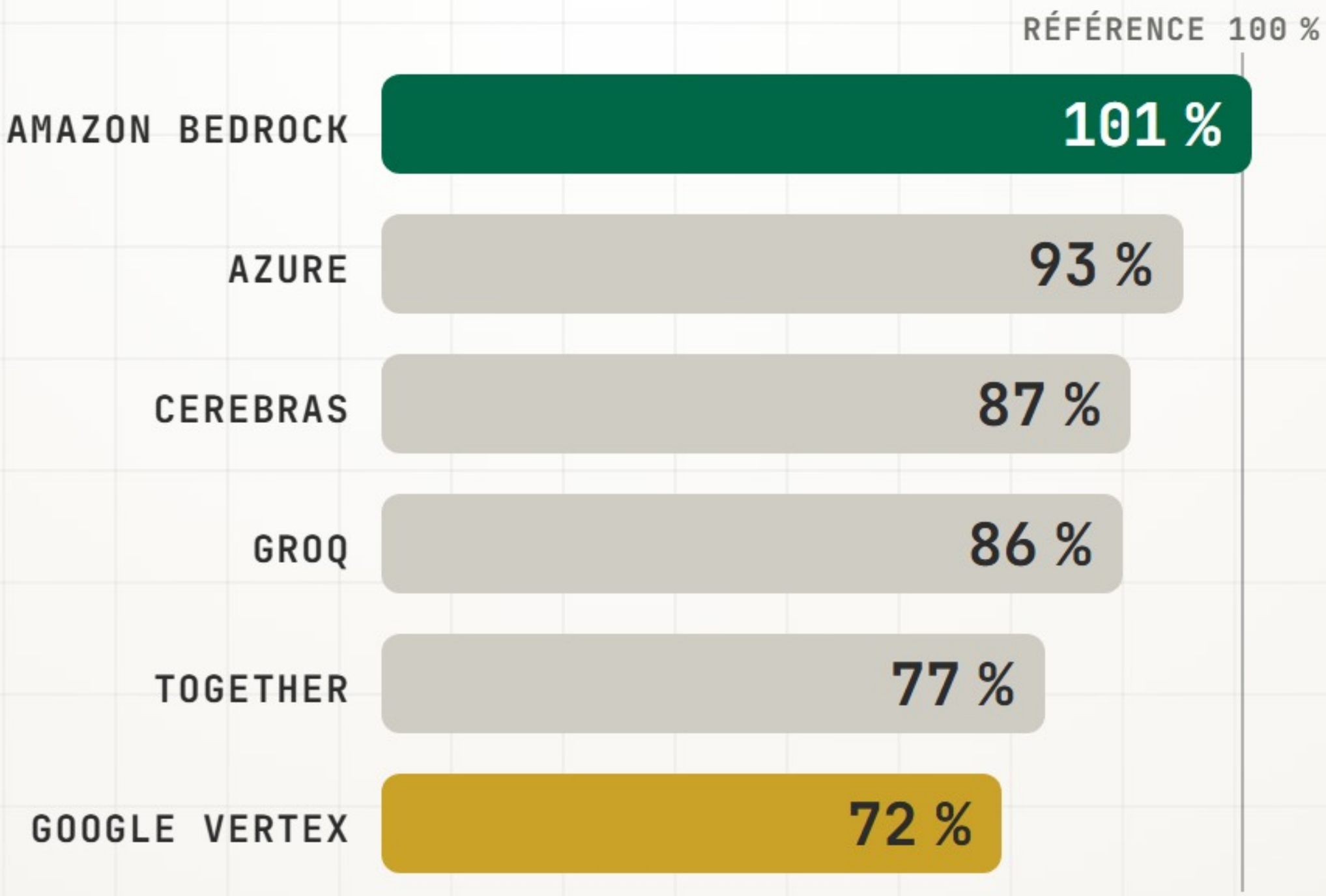
LES CINQ AUTRES PEUVENT BOUGER, LE NOM RESTE

2	**La précision**	Compressé, le modèle coûte moins cher et répond un peu autrement.
3	**Le moteur**	Le logiciel qui le fait tourner sur les puces.
4	**Le tirage**	Les réglages du hasard qui choisit chaque mot.
5	**L’aiguillage**	Il décide quel serveur, parfois quel modèle, vous répond.
6	**L’enveloppe**	Consignes cachées et réglages par défaut : effort, longueur.

Docs et post-mortems Anthropic, 2025-2026

Même modèle, 29 points d’écart.

gpt-oss-120b, le modèle ouvert d’OpenAI : mêmes poids chez les 14 vendeurs. Part de la performance de référence que chacun conserve :

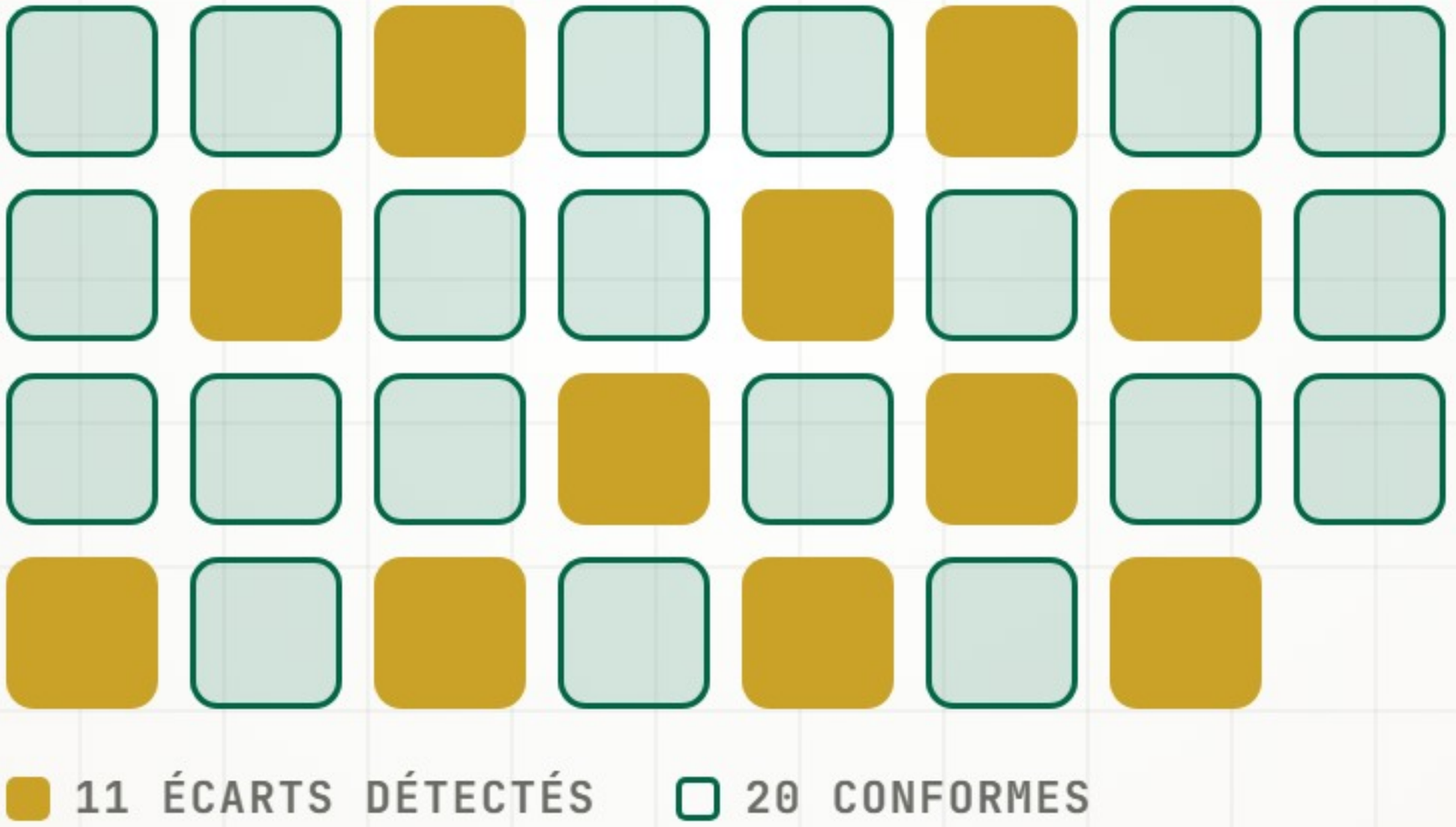


Six des quatorze points d’accès mesurés. Référence : la copie hébergée par Artificial Analysis.

Artificial Analysis, octobre 2026

NABIL EL MAHYAOUI

11 sur 31 servaient autre chose que le modèle publié.



Le Llama de Meta, vendu par neuf fournisseurs. Deux seulement avaient déclaré une modification.

Coût de chaque audit : moins d’un dollar.

Gao, Liang, Guestrin · Stanford, ICLR 2025

Deux fois, les utilisateurs l'ont vu avant les tests.

SEPTEMBRE 2025

Trois bugs d'infrastructure. Jusqu'à 16 % des requêtes touchées à la pire heure.

AVRIL 2026

Trois changements de réglages et de consignes. Des semaines de dégradation, sans toucher au modèle.

> *« Nos évaluations n'ont tout simplement pas capté la dégradation que signalaient les utilisateurs. »*
>
> ANTHROPIC, SEPTEMBRE 2025 · MA TRADUCTION

Si le fabricant l'apprend par ses clients, l'acheteur ne le verra pas à l'usage.

Anthropic, post-mortems 09/2025 et 04/2026

NABIL EL MAHYAOUI

Le nom est garanti. Le comportement, non.

GARANTI

- Un identifiant daté sert toujours le même modèle (Anthropic).
- Un préavis avant le retrait d’un modèle.

PAS GARANTI

- Ce que l’alias court désignera demain.
- La précision et le moteur chez qui vous le sert.
- Le comportement : l’infrastructure peut le modifier, même à identifiant et poids inchangés.

Le nom vous protège d’un changement de modèle. Rien ne vous protège d’un changement de tout le reste.

Docs Anthropic et OpenAI, lues le 19/10/2026

NABIL EL MAHYAOUI

Le test de réception coûte un dollar.

1 **Le jour de la réception**

25 questions de votre métier, 10 réponses à chacune. C’est l’empreinte : on la fige.

2 **Chaque mois, et après chaque incident**

Les mêmes questions, de nouvelles réponses.

3 **Un test compare les deux piles**

Méthode publiée par Stanford. Pas de note, pas d’expert, un comptage. Verdict : pareil, ou alerte.

CE QUE J’AI VU EN LE FAISANT

Sur le fournisseur que j’utilise, le test distingue deux modèles au premier coup d’œil. Mais des réponses tirées dans la même minute se ressemblent trop : **construisez l’empreinte sur plusieurs jours**, sinon il crie au loup.

Trois décisions, avant la prochaine facture.

Épingler

Un identifiant daté dans tout ce qui compte. Jamais l'alias court : c'est le choix de la semaine du fournisseur.

Sonder

L'empreinte à la réception, le test chaque mois. S'il échoue, c'est une réserve : une date, et une conversation avec le fournisseur.

Contractualiser

Quatre questions écrites : quelle précision, quel moteur, l'aiguillage peut-il substituer un autre modèle, où est le journal des changements ?

La grille à garder.

Épingler

01

QUI
La DSI

QUAND
À la mise en production

COÛT
Zéro

AU CONTRAT
L’identifiant daté de chaque modèle servi.

Sonder

02

QUI
L’équipe data

QUAND
À la réception, puis chaque mois

COÛT
Moins d’un dollar par audit

AU CONTRAT
Un droit d’audit et un seuil d’alerte convenus.

Contractualiser

03

QUI
Achats et juridique

QUAND
À la signature, au renouvellement

COÛT
Une clause

AU CONTRAT
Préavis sur précision, moteur, consignes et aiguillage ; un recours si l’audit échoue.

D’où viennent ces chiffres.

72 À 101 %

Artificial Analysis, Endpoint Accuracy Index, gpt-oss-120b, octobre 2026.

11 SUR 31, MOINS D’UN DOLLAR

Gao, Liang, Guestrin, « Model Equality Testing », Stanford, ICLR 2025.

LES DEUX ÉPISODES

Anthropic, post-mortems du 17/09/2025 et du 23/04/2026.

CE QUI EST GARANTI

Documentation des versions de modèles, Anthropic et OpenAI.

LIMITES

Les écarts mesurés concernent des modèles ouverts revendus par des tiers ; chez le fabricant, ce sont les réglages qui ont bougé, pas le modèle. Mon test est un essai, pas une étude.

POUR FINIR

Un modèle d’IA sans test de réception n’est pas un équipement.

C’est un abonnement à une promesse.

Qui, chez vous, signe la réception d’un modèle d’IA ?

À envoyer à la personne qui achète vos modèles d’IA.