

Version en ligne : aipm-standard.com/aipm/cas/revendications-clients · édition du 30/09/2026. Le statut de preuve peut évoluer : la page en ligne fait foi.

Cas approfondi

Douze entreprises, des chiffres précis, une seule voix

+17 % de fidélité au brief, « plus de 10 % » d'affirmations incorrectes en moins, +20 % de bugs détectés, +20 % sur les processus longs. Douze entreprises nommées, des pourcentages à l'unité près, tous publiés par le fournisseur du modèle. Comment un référentiel traite ce type de matière.

Toute la banque de cas (aipm-standard.com/aipm/cas) · La règle de preuve (aipm-standard.com/aipm/methode-de-preuve)

Statut de preuve : Annoncé/projeté. Ce statut n'est pas une opinion sur l'entreprise ou sur la technologie : il décrit ce que les sources consultées permettent d'affirmer, et rien de plus. La règle qui l'attribue est publiée sur la page Méthode de preuve (aipm-standard.com/aipm/methode-de-preuve). Il porte sur le constat central de la fiche ; les chiffres cités à l'intérieur conservent le statut de leur propre source, et ceux qui proviennent du seul éditeur restent « annoncés », y compris dans une fiche confirmée.

Le contexte

Qui, quand, quel système

La page professionnelle de GPT-6 Astra cite nommément douze entreprises et leur attribue des résultats chiffrés : Cognition, Databricks, Hebbia, Box, Figma, Thomson Reuters Labs, Datacurve, Basis, CodeRabbit, XTX Markets, Navan, Avalara.

Les chiffres sont précis. Hebbia : des présentations « 17 % plus fidèles au brief » et des citations correctes « 19 % plus souvent ». Box : « plus de 10 % » de risque en moins d'affirmer avec assurance des conclusions incorrectes. CodeRabbit : « 20 % de bugs détectés en plus », et un taux doublé sur les revues complexes. Basis : « amélioration de 20 % » sur les processus longs. Datacurve : 74 % sur DeepSWE v1.1.

La décision, et qui l'a prise

La Directive de fait

Chacune de ces entreprises a mesuré quelque chose de réel sur son propre produit. Rien n'indique une fabrication, et il n'y a aucune raison de le supposer.

Mais la publication est unique et intéressée : c'est le fournisseur du modèle qui choisit quels clients citer, quels chiffres retenir, et sur quelle période. Aucune des mesures n'est accompagnée de son protocole, de son échantillon, de sa période ; la base de comparaison n'est donnée que par Hebbia, qui se situe « par rapport au meilleur modèle suivant ». Aucune n'est reproductible en l'état.

Ce que le système a réellement fait

Un détail de forme mérite l'attention, parce qu'il revient dans toutes les communications de ce genre : tous les chiffres vont dans le même sens. Sur douze entreprises et une quinzaine de mesures, aucune régression, aucun résultat neutre, aucun cas où le modèle précédent faisait mieux.

Un relevé honnête portant sur des systèmes complexes produit presque toujours des résultats mêlés — c'est d'ailleurs ce qu'a produit l'évaluateur indépendant sur le même modèle la veille, avec une régression nette sur le test le plus proche du travail réel. L'unanimité n'est pas une preuve de fausseté ; c'est un indice de sélection.

L'état de la preuve

Ce que les sources établissent, et ce qu'elles n'établissent pas

Ce que la source établit

- Les entreprises citées existent et sont nommées, ce qui est vérifiable
- Les chiffres sont précis, ce qui les rend contestables — donc utiles
- L'éditeur ne dissimule pas être la source de la publication
- Aucune des mesures ne va dans le sens défavorable

Ce que la source n'établit pas

- Le protocole de chaque mesure : échantillon, période, base de comparaison
- Si les entreprises citées ont validé la formulation qui leur est attribuée
- Combien de clients ont mesuré sans être cités
- Si ces gains persistent au-delà de la période d'évaluation

Sources consultées : OpenAI — GPT-6 Astra pour le travail

Ce que ce cas dit du référentiel

Ce cas est l'illustration la plus nette du **principe 7** dans son sens le plus littéral : l'autorité empruntée. Citer « Box a constaté plus de 10 % d'affirmations incorrectes en moins » dans une note interne, c'est emprunter à la fois la crédibilité de Box, celle d'OpenAI, et celle d'une mesure dont on ignore tout.

Il fixe une règle simple pour la **banque de cas** : un chiffre client relayé par le fournisseur entre au statut « annoncé », quel que soit le prestige du client. Il ne passe à « confirmé » que si le client le publie lui-même avec son protocole, ou si un tiers le reproduit. Cette règle

est mécanique et ne dépend d'aucun jugement sur l'entreprise citée.

C'est cette règle qui explique pourquoi la banque de cas d'AIPM compte autant de lignes « à consolider » : la matière disponible publiquement est, pour l'essentiel, de cette nature.

La consigne

Ce qu'un lecteur fait différemment après avoir lu cette fiche

Renversez la charge. Quand un fournisseur vous cite un client, demandez à parler au client. S'il est cité publiquement, il assume ; l'entretien coûte une heure et remplace une croyance par un constat.

Cherchez la mesure manquante. Devant une liste de résultats tous favorables, la question utile n'est pas « est-ce vrai ? » mais « qu'est-ce qui n'est pas dans la liste ? ». Ici, la réponse était publique, publiée le 9 septembre : une régression mesurée par un tiers sur le test le plus proche du travail facturable.

Refaites la mesure chez vous, en petit. Vingt dossiers traités deux fois, une fois avec, une fois sans, comparés par quelqu'un qui ne sait pas lequel est lequel. C'est deux jours de travail et cela vaut n'importe quel pourcentage annoncé, parce que c'est le vôtre.

Ce qui invaliderait cette lecture. La publication, par l'une des entreprises citées, de sa propre mesure avec son protocole. Ce chiffre-là passerait alors à « confirmé » et sortirait de cette fiche — qui porte sur la chaîne de publication, pas sur la véracité des résultats.

Cette fiche suit le gabarit des cas approfondis d'AIPM : huit questions, toujours les mêmes. Voir les autres fiches (aipm-standard.com/aipm/cas-approfondis).

© 2026 DYONYSOS — Tous droits réservés · Bexo® est une marque déposée de DYONYSOS · aipm-standard.com. Reproduction interdite sans autorisation, hors courtes citations avec mention de la source (aipm-standard.com).