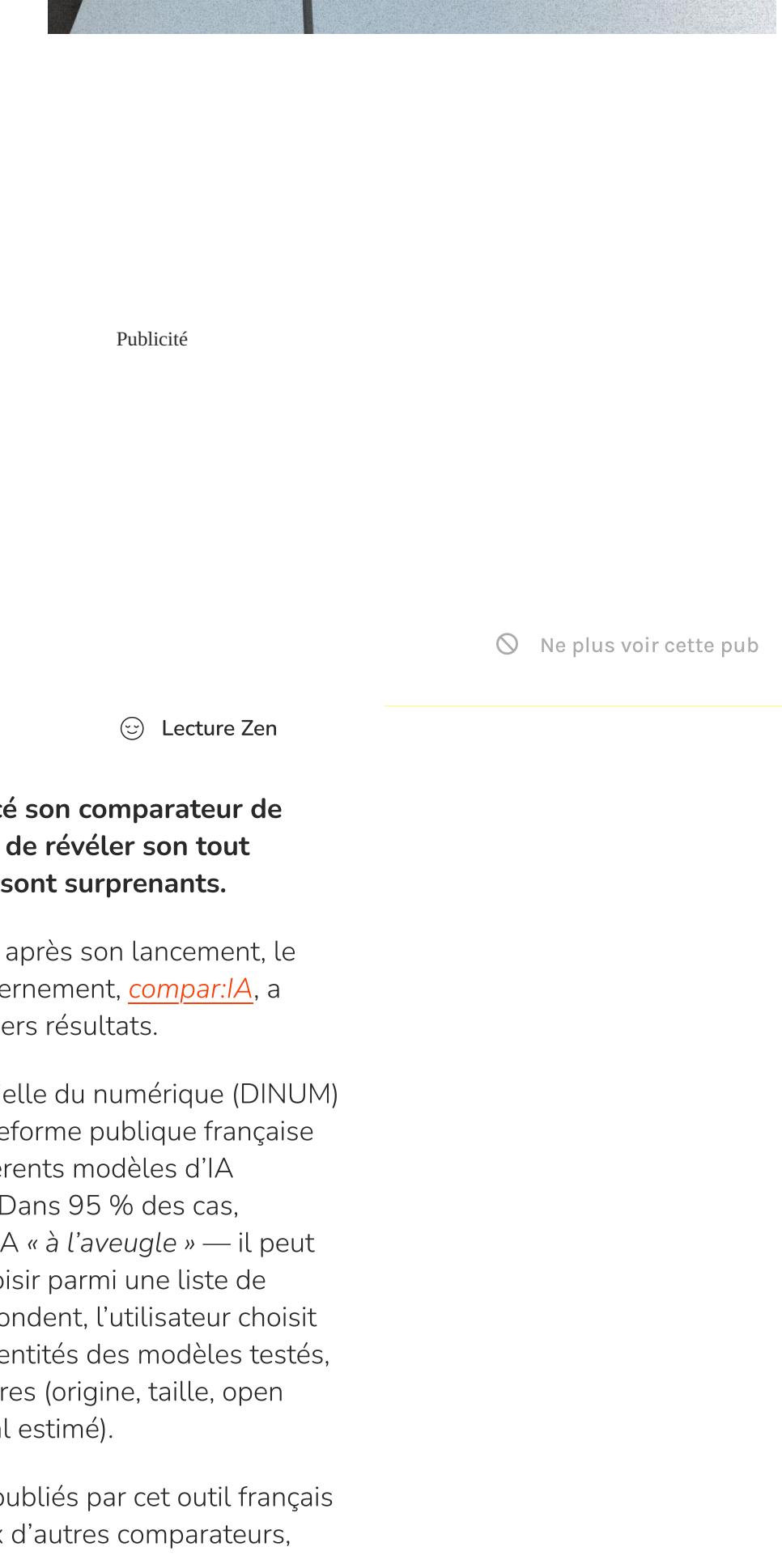


Publié le 04 novembre 2025 à 17h14

Tech Intelligence artificielle

La France dévoile son classement des meilleures IA : le gagnant va vous surprendre

Mistral > OpenAI, Anthropic et Google



🕒 4 min

Lisa Imperatrice

Publicité

Lisa Imperatrice est une journaliste et écrivaine française, spécialisée dans les thématiques de l'art contemporain et de la culture. Elle a travaillé pour divers médias, dont Le Monde et L'Art Magazine.

Le gouvernement français, qui a lancé son comparateur de modèles d'IA en octobre 2024, vient de révéler son tout premier classement. Et les résultats sont surprenants.

Le classement peut surprendre : un an après son lancement, le comparateur de modèles d'IA du gouvernement, [comparAI](#), a dévoilé le 3 novembre 2025 ses premiers résultats. Imaginée par la Direction interministérielle du numérique (DINUM) et le ministère de la Culture, cette plateforme publique française permet depuis un an de comparer différents modèles d'IA conversationnelle, de façon anonyme. Dans 95 % des cas, l'utilisateur pose une question à deux IA « à l'aveugle » — il peut aussi, plus rarement (5 % des cas), choisir parmi une liste de questions prédefinies. Les deux IA répondent, l'utilisateur choisit celle qu'il préfère, puis découvre les identités des modèles testés, avec quelques données complémentaires (origine, taille, open source ou non, impact environnemental estimé).

À notre grande surprise, les résultats publiés par cet outil français n'ont pas grand-chose à voir avec ceux d'autres comparateurs, comme LM Arena.

Total modèles	Total votes	Mise à jour le 04/11/2025	Télécharger les données	Rechercher un modèle
1	1000	Mise à jour le 04/11/2025	Télécharger les données	Rechercher un modèle
2	1000	Confiance (0)	Total votes	Confiance moyenne
3	1000	-0/+0	912	N/A
4	1000	-0/+0	1016	N/A
5	1000	-0/+1	1016	N/A
6	1000	-0/+2	1016	N/A
7	1000	-0/+4	1016	N/A
8	1000	-0/+7	1016	N/A
9	1000	-0/+4	1016	N/A
10	1000	-0/+4	1016	N/A
11	1000	-0/+4	1016	N/A
12	1000	-0/+2	1016	N/A
13	1000	-0/+7	1016	N/A
14	1000	-0/+7	1016	N/A
15	1000	-0/+7	1016	N/A
16	1000	-0/+5	1016	N/A
17	1000	-0/+9	1016	N/A
18	1000	-0/+7	1016	N/A

L'outil est disponible depuis octobre 2024 : voici son classement au bout d'un an. // Source : comparAI

Un petit modèle Mistral en tête du classement, GPT-5 et Gemini 2.5 Pro largués... Le classement de la France étonne

Le podium a de quoi faire lever un sourcil. [En tête, un modèle français — Mistral-medium-3.1](#) — qui n'est même pas le plus performant de la marque. Cet « intermédiaire » est avant tout pensé pour offrir un bon compromis coût/performance dans le cloud. Il devance au classement Gemini-2.5-Flash et Gemini-2.0-Flash, deux modèles allégés conçus pour la rapidité plutôt que la profondeur, qui sont moins performants que Gemini-2.5-Pro, en tête des classements internationaux. La présence de Gemini 2.0 Flash dans le top 3, malgré son ancieté, a de quoi surprendre : comment trois modèles « allégés », comme Mistral Medium 3.1, Gemini 2.5 Flash et Gemini 2.0 Flash peuvent battre les fleurons du secteur ?

Du côté d'OpenAI, le constat est tout aussi surprenant. Le premier modèle GPT n'apparaît qu'à la septième place — et il ne s'agit même pas de GPT-4.5 ou GPT-5, mais de gpt-oss-120b, une version open source publiée en août 2025. Les autres modèles d'OpenAI suivent : gpt-4.1-mini (22e) et gpt-5-mini (26e). À titre de comparaison, dans la catégorie « Text Arena » de LM Arena — dédiée aux performances textuelles des modèles —, les trois premières places sont occupées par Gemini-2.5-Pro, Claude Opus 4.1 et Claude Sonnet 4.5. Des résultats bien plus en phases avec ce que disent les experts de l'industrie.

Le classement de la partie Text Arena de LM Arena, le comparateur international, avec des résultats très différents. // Source : LM Arena

Prudent, le Ministère de la Culture rappelle qu'il dispose de très peu d'informations sur le profil des votants — un choix volontaire lié à la protection des données personnelles. Impossible, donc, de savoir si les participants étaient des experts, des curieux, ou de simples utilisateurs venus tester [les IA](#).

Du côté d'OpenAI, le constat est tout aussi surprenant. Le premier modèle GPT n'apparaît qu'à la septième place — et il ne s'agit même pas de GPT-4.5 ou GPT-5, mais de gpt-oss-120b, une version open source publiée en août 2025. Les autres modèles d'OpenAI suivent : gpt-4.1-mini (22e) et gpt-5-mini (26e). À titre de comparaison, dans la catégorie « Text Arena » de LM Arena — dédiée aux performances textuelles des modèles —, les trois premières places sont occupées par Gemini-2.5-Pro, Claude Opus 4.1 et Claude Sonnet 4.5. Des résultats bien plus en phases avec ce que disent les experts de l'industrie.

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Tous nos articles sont aussi sur notre profil Google : [suivez-nous pour ne rien manquer !](#)

Tout comprendre à La guerre de l'IA

La France dévoile son classement des meilleures IA : le gagnant va vous surprendre

OpenAI change de nom pour faire du OpenAI sans toucher à OpenAI

« Détection de ressemblance », YouTube lance un outil pour traquer les chaînes IA qui usurpent l'identité des créateurs, mais on reste loin du compte

Pour contrer ChatGPT, Google songe à armer l'un de ses rivaux en lui prêter sa puissance de calcul

OpenAI dévoile ChatGPT Atlas, un navigateur pour tuer Google Chrome

Crédit photo de la une : Numerama Signaler une erreur dans le texte

Partager l'article X f ln +

France Google Gemini gouvernement IA lr

tech La Fnac et Darty jouent déjà le jeu du Black Friday : voici les meilleures...

04.11.2025 15:39

tech consoles de jeu

Nintendo vend plus de Switch 2 que prévu : la console a presque dépassé la...

04.11.2025 12:27

tech informatique

Windows 11 : vérifiez si votre PC aura encore accès aux mises à jour après l...

04.11.2025 12:06

Quelles sont les meilleures alternatives gratuites à ChatGPT ?

Merci à nos Fondateurs +

Bastian Fabre Dhay

RomainBilly dwalar

MaxZedoyer Dany

Ne plus voir cette pub

En gras, les modèles présents dans les deux classements

Intervalle de confiance : faut-il se fier au classement français ?

Pour interpréter ces résultats, la DINUM et le ministère de la Culture insistent sur un indicateur clé : l'intervalle de confiance. Il mesure la solidité d'une position dans le classement. Il ne s'agit pas d'un simple détail statistique, mais d'un outil essentiel pour mesurer la fiabilité des positions. En pratique, il indique le degré d'incertitude associé au score d'un modèle : plus cet intervalle est étroit, plus la position du modèle est considérée comme solide et stable. À l'inverse, un intervalle large signifie que le modèle reçoit des votes très variés — certains utilisateurs l'adorent, d'autres beaucoup moins. Concrètement, si l'on regarde [deepseek-chat](#): 3,1 (470 votes), il apparaît actuellement à la huitième place sur comparAI, mais avec un intervalle de confiance relativement large, de -10 à +7. Autrement dit, sa position pourrait facilement fluctuer selon les prochains votes : il pourrait aussi bien grimper de plusieurs rangs que redescendre.

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le logo de DeepSeek. // Source : Numerama

Ce type d'écart traduit une forte hétérogénéité des réponses et souligne que le classement reste, à ce stade, vivant et évolutif, plutôt qu'une photographie figée des performances des IA. Le modèle de MistralAI, première position, dispose quant à lui d'un intervalle de confiance de -0/+0 : l'incertitude autour du score du modèle est donc extrêmement faible, voire nulle.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée sur d'éventuels biais linguistiques, culturels ou nationaux dans les réponses des modèles. Pour l'heure, comparAI se présente avant tout comme une base d'observation, ouverte à la recherche et à l'analyse.

Toute l'actu tech en un clin d'œil

Ajoutez Numerama à votre écran d'accueil et restez connectés au futur !

Installer Numerama

Le classement sera mis à jour chaque semaine, nous indique le ministère de la Culture. Trop tôt, donc, pour parler d'un biais « patriotique » en faveur de Mistral AI : aucune étude approfondie n'a encore été menée



[À propos](#) • [Mentions légales](#) • [Données personnelles](#) • [Politique Cookies](#) • [Contact](#) • [Gérer UTIQ](#)

Humanoid © 2025 - Numerama, tous droits réservés