

The logo for LIONBRIDGE, featuring the company name in a bold, black, sans-serif font. A small orange horizontal line is positioned under the letter 'O' in 'BRIDGE'. The logo is centered within a white rounded rectangular box.

LIONBRIDGE

ANALYTIQUE DES DONNÉES D'IA POUR L'ÉVALUATION DES MODÈLES



CATÉGORIES D'ÉVALUATION

L'évaluation des performances des grands modèles de langage (LLM) requiert une approche structurée, tenant compte de la nature multidimensionnelle de la génération de langage. Les catégories d'évaluation suivantes ont été sélectionnées afin de fournir un cadre complet pour évaluer les résultats des modèles en matière de qualité, de fiabilité et de pertinence pour l'utilisateur. Ces catégories permettent de s'assurer que les évaluations ne s'arrêtent pas à l'exactitude apparente, mais qu'elles intègrent bien des aspects tels que la fluidité, l'exhaustivité, la terminologie spécifique au domaine et la pertinence culturelle. Ainsi, les organisations peuvent déterminer en toute confiance si les résultats d'un modèle répondent aux attentes spécifiques de leurs utilisateurs, suivent les objectifs métier et respectent les normes éthiques.



PRÉCISION

Détermine si la réponse est correcte factuellement et dénuée d'erreurs.

Exemple : Cite la bonne loi dans une réponse portant sur du juridique.



EXHAUSTIVITÉ

Détermine si toutes les parties de la question ont bien été traitées.

Exemple : Prend en compte tous les points essentiels dans une demande de synthèse.



PERTINENCE

Détermine si la réponse est dans le thème et répond directement à la question.

Exemple : N'ajoute pas d'informations sans rapport avec le sujet dans la description d'un produit.



COHÉRENCE

Veille à ce que la réponse soit logique en elle-même et dénuée de contradictions.

Exemple : Ne se contredit pas dans une réponse sur plusieurs paragraphes.



FLUIDITÉ

Évalue l'exactitude grammaticale et le flux naturel du langage.

Exemple : Utilise une structure syntaxique et une ponctuation appropriées.



HALLUCINATION (INVERSÉE)

Détermine si le modèle évite de créer des faits. (moins d'hallucinations = meilleur score après inversion)

Exemple : N'invente pas de fonctionnalités du produit qui n'existent pas.



TERMINOLOGIE

Évalue la bonne utilisation des termes spécifiques au domaine et du jargon.

Exemple : Utilise une terminologie médicale précise dans une réponse concernant la santé.



LISIBILITÉ

Détermine dans quelle mesure le texte est facile à lire et à comprendre pour le public cible.

Exemple : Utilise une formulation claire et concise pour un public général.



PERTINENCE CULTURELLE

Détermine si la réponse est appropriée et sensible aux normes culturelles du public cible.

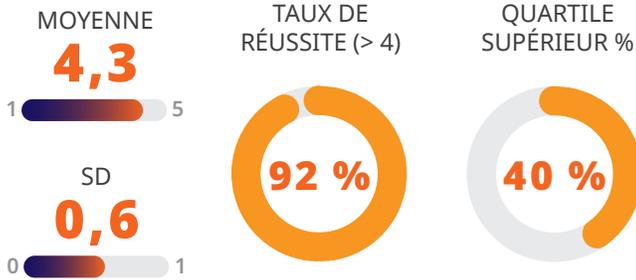
Exemple : Évite les idiomes ou les références qui peuvent être inappropriés dans certains pays.

En appliquant systématiquement ces catégories, notées sur une échelle de Likert à 5 points, les clients bénéficient d'informations exploitables sur les domaines dans lesquels le modèle excelle et sur ceux dans lesquels il peut nécessiter un ajustement, un entraînement supplémentaire ou une supervision par des humains. Cela permet de s'assurer que les déploiements d'IA offrent des expériences cohérentes et de haute qualité dans différentes applications, qu'il s'agisse de support client, de génération de contenu, de recommandations de produits ou d'autres domaines.

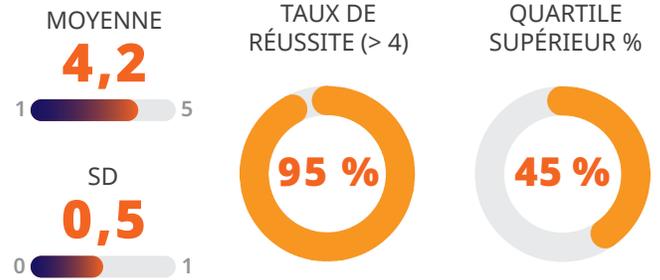
EXEMPLE D'ANALYTIQUE | BÉNÉFICIEZ D'INFORMATIONS PLUS APPROFONDIES

L'analytique des services de données de Lionbridge fournit une visibilité essentielle sur la qualité, la cohérence et l'efficacité des processus d'entraînement de l'IA. Ces informations aident à identifier les schémas dans la diversité des données, les niveaux d'accord des annotateurs, les points forts et les points faibles du modèle, ainsi que les tendances dans les résultats. Cela permet de s'assurer que les systèmes d'IA sont solides, dénués de biais et hautement performants. En s'appuyant sur l'analytique, les organisations peuvent prendre des décisions basées sur les données afin d'optimiser l'entraînement de modèle, d'améliorer les processus d'annotation et d'accélérer le déploiement de l'IA en toute confiance.

Précision



Pertinence



Accord entre les évaluateurs (Kappa de Fleiss)



Corrélation entre les catégories d'évaluation



POUR EN SAVOIR PLUS, RENDEZ-VOUS SUR [LIONBRIDGE.COM/FR](https://lionbridge.com/fr)