

5 RAISONS DE CHOISIR LA COLLECTE DE DONNÉES D'IA PAR DES HUMAINS



Pour entraîner votre modèle, vous avez besoin de données, qu'elles soient synthétiques, collectées par des humains ou les deux. Découvrez pourquoi les données collectées par des humains (du moins, la majorité de celles que vous utilisez) sont plus susceptibles d'offrir des performances du modèle plus fiables et plus efficaces.

1

QUALITÉ

Des données de qualité supérieure améliorent les performances du modèle, car elles sont généralement plus exactes, à jour, claires, cohérentes et structurées, tout en étant riches en informations contextuelles. Toutes ces qualités permettent à un modèle de développer plus facilement une compréhension plus nuancée du langage et des sujets pertinents, et ainsi de livrer de meilleures performances.



EXEMPLE

Lionbridge a aidé un prestataire de solutions de formations en ligne en révisant plus de 300 vidéos transcrites automatiquement, afin de détecter les problèmes de qualité. Les réviseurs ont modifié les sous-titres transcrits par l'IA et ont fourni des transcriptions de vidéo exactes.

2

DIVERSITÉ

Les prestataires possédant une base de contributeurs fortement diversifiée et internationale fournissent des données qui couvrent des scénarios concrets de manière plus complète et reflètent différents points de vue dénués de préjugés.



EXEMPLE

Pour aider une grande société de technologies, Lionbridge a utilisé Aurora AI Studio™ et son équipe de recrutement pour recueillir plus d'un million de phrases enregistrées avec des émotions spécifiques. Les participants parlaient plusieurs langues et dialectes, et couvraient différents secteurs démographiques.

3

VITESSE

Un prestataire efficace peut considérablement accélérer les processus. Lionbridge Aurora AI Studio nous relie à notre base mondiale d'un demi-million de contributeurs. Nous disposons d'une solide équipe de recruteurs qui recherchent des contributeurs en dehors de cette base si besoin. Les tâches sont acceptées et terminées en quelques jours, voire heures.



EXEMPLE

Pour aider un fabricant de smartphones, Lionbridge a utilisé Aurora AI Studio pour fournir un jeu de données volumineux d'exemples de conversations de la vie réelle dans 8 langues, comprenant plus de 200 000 dialogues avec un maximum de 5 participants par conversation. Notre plateforme nous a permis de fournir toutes les données de conversation en 4 semaines.

4

PRIX

Si seules des données synthétiques sont utilisées, le modèle livre souvent des performances médiocres, et des procédures de validation supplémentaires ou l'obtention de nouvelles données collectées par des humains sont nécessaires. Réduisez les coûts en choisissant directement des données collectées par des humains.



EXEMPLE

Lionbridge a aidé un prestataire de services vidéo en ligne à assurer la traduction d'un grand nombre de vidéos à partir de plusieurs langues vers l'anglais. Les traducteurs ont signalé tout le contenu vulgaire, biaisé, offensant ou haineux. Le soin qu'ils ont apporté à la tâche et leurs révisions ont permis de s'assurer que le LLM livrerait plus rapidement des résultats optimaux et non offensants. Il n'a pas été nécessaire de collecter ni de valider des données supplémentaires.

5

PROCESSUS DE DÉVELOPPEMENT

Le développement est un processus onéreux. Ne faites pas perdre de temps aux développeurs et ne gaspillez pas les coûts liés à la main-d'œuvre. Choisissez dès le début des données fiables et collectées par des humains.



EXEMPLE

Pour aider un développeur d'IA, nous avons mis à disposition des réviseurs humains qui ont fourni au LLM les données d'apprentissage étendues requises dans plusieurs langues. Ils ont sélectionné les réponses optimales aux invites et ont noté les réponses selon plusieurs facteurs.

