

5 MOTIVI PER SCEGLIERE LA RACCOLTA DATI AI TRAMITE LAVORO UMANO



L'addestramento di un modello richiede dati, che siano sintetici, raccolti tramite lavoro umano o una combinazione di entrambi i tipi. Scoprite perché i dati raccolti da persone (almeno come componente principale) hanno maggiori probabilità di assicurare prestazioni del modello più efficaci e affidabili.

1

QUALITÀ

Dati di qualità superiore assicurano prestazioni migliori del modello perché sono in genere più accurati e aggiornati, più puliti, coerenti, ben strutturati e ricchi di contesto. Tutte queste caratteristiche rendono più facile per il modello sviluppare una comprensione del linguaggio e degli argomenti rilevanti tenendo conto anche delle sfumature, così da offrire prestazioni migliori.



ESEMPIO

Lionbridge ha supportato un fornitore di soluzioni di eLearning rivedendo oltre 300 video trascritti automaticamente per individuare i problemi di qualità. I revisori hanno corretto i sottotitoli generati dall'AI e hanno fornito trascrizioni video estremamente accurate.

2

DIVERSITÀ

I fornitori che collaborano con una community globale e altamente diversificata offrono dati che coprono in modo più completo gli scenari reali e riflettono punti di vista diversi e privi di bias.



ESEMPIO

Per supportare una grande azienda tecnologica, Lionbridge si è affidata ad Aurora AI Studio™ e al proprio team di selezione delle risorse per raccogliere oltre un milione di frasi registrate con emozioni specifiche. I partecipanti sono stati scelti in modo da coprire un ampio spettro di lingue, dialetti e caratteristiche demografiche.

3

VELOCITÀ

Con i fornitori giusti, il processo può essere estremamente rapido. Lionbridge Aurora AI Studio si connette alla nostra community globale di mezzo milione di collaboratori. Disponiamo inoltre di un team di reclutatori esperti in grado di individuare i collaboratori anche al di fuori della community, se necessario. Le attività vengono avviate e completate in pochi giorni o persino in poche ore.



ESEMPIO

Per un produttore di smartphone, Lionbridge ha usato Aurora AI Studio per acquisire un ampio set di dati di esempi di conversazioni "reali" in 8 lingue, composto da oltre 200.000 dialoghi con fino a 5 partecipanti per conversazione. La piattaforma ci ha permesso di consegnare tutti i dati delle conversazioni nel giro di 4 settimane.

4

PREZZO

L'uso esclusivo di dati sintetici spesso porta a prestazioni del modello non soddisfacenti, con la necessità di svolgere ulteriori procedure di convalida o ottenere dati raccolti tramite lavoro umano. Scegliendo fin da subito dati raccolti da persone, è possibile ridurre i costi.



ESEMPIO

Lionbridge ha supportato un fornitore di servizi video online nella traduzione in inglese di un volume elevato di video in diverse lingue di origine. I traduttori hanno segnalato eventuali contenuti volgari, discriminatori, offensivi o incitanti all'odio. Grazie alla grande attenzione e a revisioni aggiuntive, è stato possibile ottenere prestazioni ottimali dell'LLM in tempi più brevi, con la sicurezza di non generare contenuti offensivi e senza necessità di ulteriori attività di raccolta o convalida dei dati.

5

PROCESSO DI SVILUPPO

Lo sviluppo è costoso. Non sprecate il tempo degli sviluppatori e i costi di manodopera. Scegliete fin dall'inizio dati affidabili, raccolti tramite il lavoro umano.



ESEMPIO

Per supportare uno sviluppatore AI, abbiamo fornito revisori umani in grado di offrire all'LLM un grande volume di dati di apprendimento in più lingue, selezionando le risposte migliori ai prompt e classificandole in base a diversi fattori.

