

5 MOTIVOS PARA ELEGIR LA RECOPIACIÓN HUMANA DE DATOS DE IA



El entrenamiento de su modelo requiere datos, ya sean sintéticos, recopilados por humanos o una combinación de ambos. Descubra por qué utilizar datos recopilados por humanos (o, al menos, en su mayoría) se traduce generalmente en un rendimiento del modelo más sólido y fiable.

1

CALIDAD

Los datos de alta calidad garantizan un rendimiento optimizado de los modelos, dado que, generalmente, se trata de datos repletos de información contextual que son más precisos, limpios, coherentes y bien estructurados, además de más actualizados. Todas estas cualidades facilitan que el modelo desarrolle una capacidad de comprensión más depurada del lenguaje y de los temas relevantes, lo que le permite rendir a un nivel superior.



EJEMPLO

Lionbridge ayudó a un proveedor de soluciones de aprendizaje automático con la revisión de más de 300 vídeos transcritos automáticamente para detectar problemas de calidad. Los revisores corrigieron los subtítulos generados por IA y proporcionaron transcripciones de vídeo de gran precisión.

2

DIVERSIDAD

Los proveedores con una base de colaboradores global y muy diversa aportan datos que reflejan con mayor exactitud situaciones del mundo real, con diferentes perspectivas libres de sesgos.



EJEMPLO

Para prestar sus servicios a una gran empresa de tecnología, Lionbridge recurrió a Aurora AI Studio™ y a su equipo de seleccionadores para recopilar más de un millón de frases grabadas que expresaban emociones específicas. Los participantes hablaban múltiples idiomas y dialectos, y abarcaban una amplia variedad de perfiles demográficos.

3

RAPIDEZ

El proceso puede ser extremadamente rápido con los proveedores adecuados. Lionbridge Aurora AI Studio nos conecta con nuestra comunidad global de medio millón de colaboradores. Contamos con un sólido equipo de seleccionadores preparados para buscar participantes fuera de nuestra comunidad, si fuera necesario. Todas las tareas se asignan y se completan en cuestión de días o incluso de horas.



EJEMPLO

Lionbridge utilizó Aurora AI Studio para ayudar a un fabricante de smartphones a obtener una extensa recopilación de datos de ejemplos de conversaciones reales en 8 idiomas y logró recabar más de 200 000 conversaciones con hasta 5 participantes por conversación. Nuestra plataforma nos permitió entregar todos los datos en un plazo de 4 semanas.

4

PRECIO

Los datos sintéticos por sí solos suelen ofrecer un rendimiento deficiente del modelo, lo que implica la necesidad de aplicar procesos adicionales de validación o de recabar más datos recopilados por humanos. Reduzca costes utilizando datos recopilados por humanos desde el principio.



EJEMPLO

Lionbridge colaboró con un prestador de servicios de vídeo en línea ayudándole a traducir un gran volumen de vídeos desde varios idiomas al inglés. Los traductores identificaron el contenido que podría considerarse vulgar, sesgado u ofensivo, o incitar al odio. Esta minuciosa atención al detalle y las revisiones realizadas permitieron que el LLM lograra un rendimiento óptimo y generara contenido no ofensivo en menos tiempo, sin necesidad de recopilar ni validar datos adicionales.

5

PROCESO DE DESARROLLO

El desarrollo es costoso. No desperdicie el tiempo de los desarrolladores ni los costes de mano de obra. Opte por utilizar datos recopilados por humanos fiables desde el primer momento.



EJEMPLO

Para apoyar a un desarrollador de IA, facilitamos revisores humanos capaces de proporcionar al LLM la gran cantidad de datos de aprendizaje necesaria en diversos idiomas. Estos se encargaron de seleccionar las respuestas óptimas a los prompts y evaluaron las respuestas basándose en diversos factores.

