

FEM GODA SKÄL ATT SAMLA IN AI- DATA MANUELLT



Det krävs data för att träna en språkmodell, oavsett om de är syntetiska eller manuellt insamlade eller både och. Läs om varför manuellt insamlade data (åtminstone merparten av dem du använder) ökar chanserna att modellen producerar effektiva och mer tillförlitliga resultat.

1

KVALITET

Data av högre kvalitet leder till bättre resultat från modellen, eftersom de vanligtvis är mer korrekta, uppdaterade, rena, enhetliga och välstrukturerade och erbjuder ett tydligt sammanhang. Det är egenskaper som gör det enklare för modellen att få mer nyanserad förståelse för språket och relevanta ämnen, så att den kan prestera bättre.



EXEMPEL

Lionbridge hjälpte en leverantör av eLearning-lösningar genom att granska över 300 maskintranskriberade videor för att upptäcka kvalitetsproblem. Språkgranskarna förbättrade de AI-transkriberade undertexterna och såg till att videotranskriberingen blev helt korrekt.

2

MÅNGFALD

Leverantörer med en global och mångfasetterad community kan tillhandahålla data som ger en mer heltäckande bild av verkliga förhållanden och speglar vitt skilda och fördomsfria perspektiv.



EXEMPEL

Lionbridge hjälpte ett stort teknikföretag genom att använda Aurora AI Studio™ och vårt rekryteringsteam för att samla in över en miljon inspelade meningar med specifika känslouttryck. Deltagarna talade flera språk och dialekter och kom från många olika demografiska områden.

3

LEVERANSTID

Med rätt leverantörer kan det gå otroligt snabbt. Lionbridge Aurora AI Studio skapar kontakter med vår globala crowd med en halv miljon bidragsgivare. Vid behov kan vårt duktiga rekryteringsteam även hitta bidragsgivare utanför vår crowd. Uppdragen accepteras och slutförs på bara några dagar eller till och med några timmar.



EXEMPEL

Lionbridge hjälpte en smartphone-tillverkare genom att använda Aurora AI Studio för att samla in en stor datauppsättning med samtalsexempel från verkliga livet på åtta språk. Det blev sammanlagt över 200 000 dialoger med upp till fem deltagare per samtal. Tack vare vår plattform kunde vi leverera alla samtalsdata på mindre än fyra veckor.

4

PRIS

Om enbart syntetiska data används fungerar modellen ofta dåligt, vilket gör att det krävs ytterligare valideringsprocedurer eller manuell insamling av data. Minska kostnaderna genom att välja manuell datainsamling redan från början.



EXEMPEL

Lionbridge hjälpte en onlineleverantör av videotjänster genom att översätta ett stort antal videor från flera språk till engelska. Översättarna flaggade innehåll som var vulgärt, fördomsfullt, stötande eller hatiskt. Den noggranna granskningen säkerställde att språkmodellen producerade optimerat och fördomsfritt innehåll snabbare än väntat, utan att företaget behövde samla in eller validera ytterligare data.

5

UTVECKLINGSPROCESS

Utveckling kostar pengar. Slösa inte med utvecklarnas tid och arbetskostnad. Välj tillförlitliga, manuellt insamlade data redan från början.



EXEMPEL

För att hjälpa ett AI-utvecklingsföretag erbjöd vi mänskliga granskare som kunde förse den stora språkmodellen med omfattande träningsdata på flera språk. De valde ut de bästa resultaten av promptar och poängsatte resultat utifrån flera faktorer.

