

FÜNF GRÜNDE FÜR DAS ZUSAMMENSTELLEN VON KI-DATEN DURCH MENSCHEN



Zum Trainieren eines KI-Modells werden Daten benötigt. Diese können synthetisch sein, von Menschen zusammengestellt werden oder aus einer Kombination beider Verfahren resultieren. Erfahren Sie, warum von Menschen zusammengestellte Daten (sofern diese den Großteil der verwendeten Daten ausmachen) die Modellleistung verbessern und für zuverlässige Ergebnisse sorgen.

1

QUALITÄT

Daten von höherer Qualität sorgen für bessere Modellleistung, weil sie in der Regel genauer und aktueller, sauberer, konsistenter, gut strukturiert und reich an Kontextinformationen sind. All diese Merkmale erleichtern es einem Modell, als Voraussetzung für optimale Performance ein differenzierteres Verständnis von Sprache und relevanten Themen zu entwickeln.



BEISPIEL

Lionbridge hat für einen Anbieter von E-Learning-Lösungen mehr als 300 maschinell transkribierte Videos auf Qualitätsprobleme kontrolliert. Die Prüfer korrigierten die mittels KI transkribierten Untertitel und lieferten exakte Videotranskriptionen.

2

DIVERSITÄT

Anbieter mit einer großen Community von Mitwirkenden hoher Diversität auf der ganzen Welt liefern Daten, die reale Szenarien umfassender abdecken und vielfältige, vorurteilsfreie Standpunkte widerspiegeln.



BEISPIEL

Im Auftrag eines großen Tech-Unternehmens konnte Lionbridge mithilfe von Aurora AI Studio™ und unseres Rekrutierungsteams über eine Million aufgezeichneter Sätze mit den angegebenen Emotionen erfassen. Die Teilnehmer sprachen etliche Sprachen und Dialekte und deckten verschiedenste demografische Gruppen ab.

3

GESCHWINDIGKEIT

Mit den richtigen Anbietern kann es richtig schnell gehen. Lionbridge Aurora AI Studio stellt die Verbindung zu unserem globalen Netzwerk mit einer halben Million Mitwirkenden her. Wir verfügen über ein starkes Team für die Anwerbung, das bei Bedarf auch außerhalb unserer Community nach Mitwirkenden sucht. So können Tätigkeiten oft in Tagen oder sogar Stunden abgewickelt werden.



BEISPIEL

Bei der Zusammenarbeit mit einem Smartphone-Hersteller nutzte Lionbridge Aurora AI Studio, um eine große Datenmenge mit „realistischen“ Konversationsbeispielen in acht Sprachen zusammenzustellen, die mehr als 200 000 Dialoge mit bis zu fünf Teilnehmern pro Konversation umfasste. Dank unserer Plattform konnten wir alle Konversationsdaten innerhalb von vier Wochen liefern.

4

KOSTEN

Synthetische Daten allein führen oft zu einer schlechten Modellleistung, sodass zusätzliche Validierungsverfahren erforderlich werden oder von Menschen zusammengestellte Daten beschafft werden müssen. Senken Sie die Kosten, indem Sie von Anfang an von Menschen zusammengestellte Daten verwenden.



BEISPIEL

Lionbridge hat für einen Anbieter von Onlinevideoservices eine große Zahl an Videos aus mehreren Sprachen ins Englische übersetzt. Die Übersetzer kennzeichneten jeglichen Content, der als vulgär, voreingenommen, beleidigend oder hasserfüllt beurteilt wurde. Dank dieser aufmerksamen Beurteilung funktionierte das LLM schneller optimal und ohne Anstößigkeiten. Eine zusätzliche Datenzusammenstellung oder -validierung war nicht erforderlich.

5

ENTWICKLUNGSPROZESS

Entwicklungsarbeiten sind kostspielig. Verschenden Sie nicht die Zeit Ihrer Entwickler und die damit verbundenen Arbeitskosten. Entscheiden Sie sich von Anfang an für zuverlässige, von Menschen zusammengestellte Daten.



BEISPIEL

Wir haben für einen KI-Entwickler menschliche Korrektoren bereitgestellt, die umfangreiche Trainingsdaten in mehreren Sprachen für das LLM des Kunden zusammenstellten. Sie wählten die besten Antworten auf relevante Prompts aus und bewerteten die Antworten anhand verschiedener Faktoren.

