5 大選擇理由

人工 AI 資料收集

訓練模型需要資料，這些資料可以是合成的、由人類收集的，抑或是兩者的組合。歡迎探索為何人工收集的資料 (至少是您使用的大部分資料) 更有可能讓您取得更強大、更可靠的模型成效表現。

1 品質

品質越高，越能確保更優異的模型效能表現。這是因為高品質的資料通常不但更加正確、即時、乾淨且一致，同時結構良好並含有豐富的情境脈絡資訊。所有這些特質，都有助您更輕鬆地訓練模型更細膩地理解語言及相關主題，進而展現更上層樓的表現。

實例

Lionbridge 為某線上學習解決方案供應商提供支援，審閱超過 300 個以機器聽寫的影片，檢查是否有品質問題。審閱人員修改了 AI 聽寫的字幕，提供了高度正確的影片聽寫。

2 多樣性

擁有極其多樣之全球資料提供人員的供應商，能提供涵蓋範圍更完善的真實情境資料，進而反映多元且沒有偏見的觀點。

實例

為了協助某大型科技公司，Lionbridge 仰賴 Aurora AI Studio™ 及我們的召募團隊，以指定的許多情緒錄製了百萬個句子。這些錄音提供者能說多種語言和方言，並涵蓋各式各樣的人口族群。

3 速度

只要選對供應商，便可能享有飛快的速度。Lionbridge Aurora AI Studio 能讓您與我們超過五十萬資料提供人員組成的全球眾包網連上線。我們擁有強大的招募人員團隊，如有需要亦可在眾包人才網之外尋找資料提供人員。客戶交付的工作往往可在數天或甚至數小時之內受理並完成。

實例

Lionbridge 使用 Aurora AI Studio，協助某智慧型手機製造商收集了來自 8 種語言的大量「實際」對話實例，這些資料由超過 200,000 則對話組成，每則對話的參與人數最多達 5 人。多虧我們的平台，我們得以在 4 個星期內提供所有對話資料。

4 價格

單憑合成資料訓練而得的模型，其效能往往表現不佳，因而必須進行額外的驗證程序，或取得額外的人工收集資料。從第一次開始便選擇使用人工收集的資料，有助降低成本。

實例

Lionbridge 支援某線上影音服務供應商，確保將大量的影片從多種語言翻譯至英文。翻譯人員將含有粗俗、偏見、冒犯或仇恨語言的任何內容標記出來。這些額外的心思及審閱工作，確保了 LLM 能更快地展現不具冒犯性的最佳表現，不需要再另外收集資料或進行驗證。

5 開發流程

開發向來所費不貲，因此更不能浪費開發人員的寶貴時間與人力成本。不妨第一次開始便選擇使用由人工收集的可靠資料。

實例

為了協助某 AI 開發商，我們提供的審閱人員為多個語言提供了 LLM 所需的廣泛學習資料，再針對提示選擇最適合的回應，並依據多個因素評量這些回應。

LIONBRIDGE

想探索 AI 資料收集服務嗎？歡迎立即與 Lionbridge AI™ 聯絡！
如需深入了解，歡迎造訪 lionbridge.com