

# AI 資料服務

資料收集與註解



Lionbridge AI 發展。我們遍佈全球的眾包人才提供多元且符合在地文化的資料集，進一步協助模型依據不同的語言、地區及真實情境交出正確的結果。



## 使用理由與對象

### 為何您會需要高品質的資料收集與註解？

AI 模型的好壞，取決於它們訓練時所使用的資料。然而許多公司所能取得的資料，在規模、品質和多樣性等方面都難以讓他們開發出正確、可靠的 AI 系統。Lionbridge 可以在以下方面提供您所需的協助：

1

#### 資料短缺

大多數的組織因為缺乏內部資源，所以難以收集龐大、多樣的資料集，尤其是極端案例、代表性不足的地區或是罕見語言等情況。

2

#### 品質與一致性

標記不當或不一致的資料，容易導致模型產生偏見、正確性低或部署失敗等問題。我們專業的註解工作流程及嚴格的 QA 能從一開始便確保資料精確無誤。

3

#### 多語言及多模態需求

從語音助理到電腦視覺，AI 系統必須要能跨不同語言、格式和環境運作。我們能交付與在地文化相關、多語言且多模態的大量資料集。

4

#### 速度與擴充能力

於組織內部建立資料團隊既耗時又成本高昂。我們能迅速展開作業、向全球拓展範疇，並支援快速的迭代週期，進而加快您模型投入生產的腳步。

5

#### 人機迴圈 (HITL) 整合

即使是最頂級的模型，也需要由人類驗證。我們的服務能提供經人類審閱、能持續改善效能的訓練資料來解決這個問題。

## 一般而言，誰會需要使用資料收集與註解服務？

我們的服務能為各個產業的團隊提供建置或拓展 AI 技術的支援。常見的使用者包括：



### 機器學習與資料科學團隊

需要大量高品質的資料集來訓練、驗證和微調模型。



### AI 產品與工程團隊

要仰賴正確標記的資料來啟動功能，例如語音助理、電腦視覺工具及搜尋演算法等。



### NLP、CV 與多模態 AI 團隊

需要針對例如實體辨識、情感分析、物件偵測或聽寫等特定工作來註解專門的資料 (文字、音訊、影像、視訊)。



### 醫療保健、金融、汽車與零售業等公司

將 AI 應用於診斷、詐騙偵測、自動化、個人化以及其他工作，每一種應用都需要特定領域且正確的訓練資料。



### 開發 GenAI 的新創公司和企業

需要有提示資料、合成輸入、微調過的訓練資料集以及經人類評估過的輸出，來建置具競爭力的 AI 應用程式。

## 我們的服務



資料註解



資料收集



模型評估

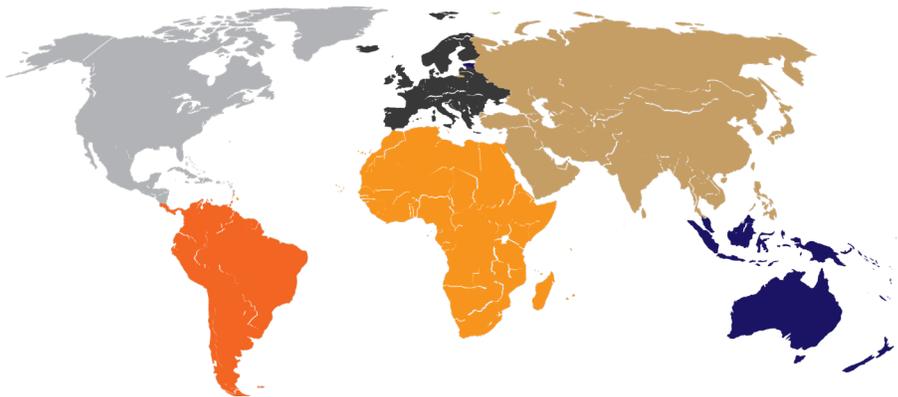
## 眾包服務

### 我們的眾包服務，就是您的競爭優勢

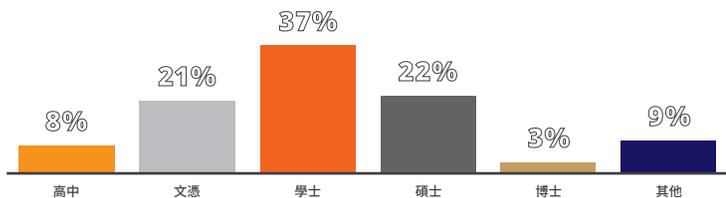
Lionbridge 的全球眾包人才網囊括了日常使用者與主題專家，能提供您規模與深度兼具的服務。無論是要訓練 AI 模型，還是要註解複雜的領域專屬資料，我們都能找到最符合您工作需求的人員。我們有專責的專案經理及健全的 QA 工作流程，確保您的資料品質優異、與在地文化相關，同時亦與實際使用情況相符，進而推動更優秀的模型表現。

這需要一個具備高度專業、多元專長與興趣的全球社群。

- 北美:26%
- 歐洲:29.5%
- 亞洲:27%
- 非洲:6.5%
- 南美:8.5%
- 大洋洲/澳洲:2.5%



### 最高教育程度



### 性別



### 年齡



60 萬人  
全球社群

350+  
語言

450+  
地區

### 產業



### 該建置一個模型還是買現成的？自行建置跟採用最合適的做法，哪個成本比較高？我們提供一個實際的 ROI 使用案例供您參考。

在內部維持一個資料營運團隊，往往會因人力、工具和時間而耗費數百萬元，但即便成本如此高昂，也不能保證其品質和規模能如您所願。某間全球人力分析公司需要註解 30 種語言的勞動市場資料，每種語言需要高達 1,200 小時的工作時間，也就是說註解作業總共需要 36,000 個小時。

#### 採用內部做法的估計值：

- 語言相關成本：約 180 萬美元
- 招聘相關時間：招募 30 名語言專家 + 入職培訓，需要花 3-4 個月
- 專案時間：因需經歷人力配置調整與分階段交付流程，約需 6 個月以上
- 訓練資料準備完成：資料可以用於訓練模型約需 8 個月以上

#### 選用 Lionbridge 的服務：

- 依需求隨選由合格語言專家及雙語註解人員組成的全球眾包人才網
- 由 Lionbridge PM 充分控管，並實施多層級品質管控機制
- 幾週（而非數個月）內便可展開專案
- 不到 12 週即可交付可供使用的訓練資料，可將模型開發時間縮短數個月
- 交付時程縮短 40%，與採用內部執行做法相比可節省超過 100 萬美元的成本

### 重要深入見解

- 30-50% 更快的模型開發速度，因為花在資料清理和重做的時間變少了
- 節省 50-60% 的成本（與建立內部資料團隊相比），因為我們有優秀的全球人才網路以及高效率的基礎設施
- 投入生產作業的時間快上 3 倍，免去曠日廢時的招聘、購買工具及訓練週期
- 減少內部負荷，讓您的 ML 團隊得以專注於模型開發而非資料整理
- 我們能讓 AI 資料製作變得更彈性靈活、經濟實惠並以結果為導向。您將可享有優異的結果，而非高額的間接成本。



如需深入了解，歡迎造訪  
**LIONBRIDGE.COM**