

AI データ サービス

データ収集と注釈付け



ライオンブリッジでは、高品質のデータ収集と注釈付けを通じて AI 開発をサポートしています。当社のグローバルな人材ネットワークにより、多様性のある、文化的にも適切なデータセットをお届けします。このようなデータセットは、さまざまな言語、地域、現実世界での利用においてモデルを正確に機能させるうえで非常に重要となります。



必要な理由と必要としているチーム・企業 高品質なデータ収集と注釈付けが必要な理由

AI モデルの品質は、モデルのトレーニングに使用されたデータの品質と同程度にしかありません。多くの企業が、正確で信頼できる AI システムを開発するうえで必要となる大規模かつ高品質で多様性のあるデータを入手するのに苦労しています。ライオンブリッジなら、以下のように支援できます。

1

データの不足

大多数の組織では、必要とされる大量の多様なデータセットを収集するための社内リソースが不足しています。エッジケース、情報があまりない地域、または希少言語においては特にこの傾向が顕著です。

2

品質と整合性

ラベル付けの品質が低いデータや不整合のあるデータを使用すると、モデルに偏見が生じ、生成される出力の精度が低くなり、モデルの展開に失敗するという結果につながります。当社の専門家による注釈付けワークフローと厳格な QA を活用いただければ、初日から精度の高いデータを得られます。

3

多言語とマルチモーダルの需要への対応

音声アシスタントからコンピュータービジョンに至るまで、AI システムはさまざまな言語や形式のデータ、多様な環境でも機能する必要があります。当社は文化的に適切な、多言語のマルチモーダル データセットを大規模に提供します。

4

スピードと拡張性

社内でデータチームを立ち上げて維持するには時間もコストもかかります。当社なら迅速に準備を整え、グローバルな拡張にも対応し、迅速なイテレーション サイクルをサポートすることで、お客様のモデルを実稼働させるまでの時間を短縮できます。

5

人間参加型モデルの統合

現時点ではトップクラスのモデルでさえ人間による検証が必要です。当社のサービスは、人によるレビューを経た、パフォーマンスの継続的な向上に効果的なトレーニング データで完結します。

一般的にデータ収集とデータの注釈付けサービスを必要とするのは誰か

当社のサービスは、AI を構築または拡張するさまざまな業界のチームを支援しています。一般的には次のようなチームを対象としています。



マシニングチームやデータサイエンスチーム

モデルのトレーニング、検証、ファインチューニングに必要な、大規模かつ高品質のデータセットを必要とするチーム。



AI 製品チームおよびエンジニアリングチーム

音声アシスタント、コンピュータービジョンツール、検索アルゴリズムなどの機能のリリースにおいて、正確にラベル付けされたデータを必要とするチーム。



NLP チーム、CV チーム、およびマルチモーダル AI チーム

表現抽出、センチメント分析、オブジェクト検出、または書き起こしなどの具体的なタスクで必要となる、専門データ（テキスト、音声、画像、動画）の注釈付けを必要とするチーム。



医療、金融、自動車、およびリテール企業

診断、不正検出、自動化、パーソナライゼーションなどにおいて AI を活用しており、それぞれに分野固有の正確なトレーニングデータを必要とする企業。



生成 AI を開発しているスタートアップや企業

競争優位性のある AI アプリケーションを構築するために、プロンプトデータ、合成入力データ、ファインチューニングされたトレーニングセット、人間によって評価された出力データを必要とする企業。

● ライオンブリッジのサービス ●



データの注釈付け



データの収集



モデルの評価

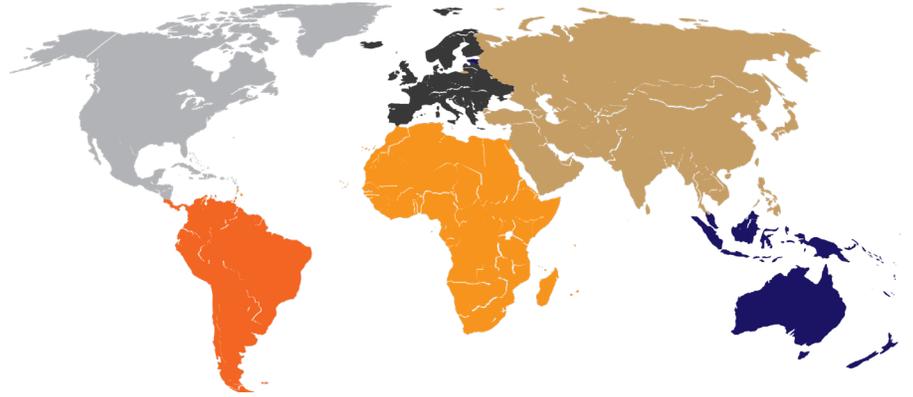
人材ネットワーク

御社の競争優位性を推進する当社の人材ネットワーク

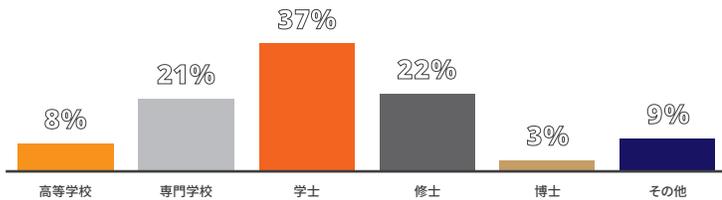
ライオンブリッジが擁するグローバルな人材ネットワークには、各分野の専門家でありながら、日常的なユーザーでもある優秀な人材が多数在籍しているため、大規模な対応が可能であるうえ、お客様はそれぞれの深い専門知識を活用できます。AI モデルをトレーニングする場合でも、複雑な分野固有のデータの注釈付けを行う場合でも、当社はタスクに最適な人材を手配します。専任のプロジェクト マネージャーと厳格な QA ワークフローにより、データが現実世界でのモデルの利用目的に合致し、モデルのパフォーマンスを向上させる、高品質で文化的に適したものであることを保証します。

高度に専門的で多様性があり、
高い関心と専門知識を備えた
グローバル コミュニティ

- 北米: 26%
- 欧州: 29.5%
- アジア: 27%
- アフリカ: 6.5%
- 南米: 8.5%
- オセアニア/オーストラリア: 2.5%



最終学歴



性別



年齢



60 万名
のグローバル
コミュニティ

350+
の対応言語

450+
の地域

業種



自社で構築すべきか購入すべきか。ROI についての現実的なユースケースを通じ、独自に開発するコストと適切な方法で行う場合のコストを比較してみましょう。

社内でデータ運用チームを立ち上げるには、人材、ツール、そして時間の面で数百万ドルがかかるうえ、データの品質や拡張性について保証はありません。あるグローバルなワークフォース分析会社では 30 言語の人材市場データの注釈付けを行う必要があり、これは言語ごとに 1,200 時間を要し、合計では 36,000 時間もの時間がかかる作業でした。

社内対応の場合:

- 言語の専門家にかかる人件費: ~ 1,800 万ドル
- 雇用にかかる時間: 30 名の言語の専門家の採用 + オンボーディングで 3 ~ 4 か月
- プロジェクトの期間: 準備期間と分納で 6 か月超
- トレーニングデータの準備: モデルのトレーニングにデータを使用できるようになるまで計 8 か月超

ライオンブリッジと提携した場合:

- 適格な言語の専門家とバイリンガルの注釈付け担当者を擁する、オンデマンドで活用できるグローバルな人材ネットワーク
- 多層的な QA の実施と当社のプロジェクト マネージャーによる完全な管理体制
- 数か月ではなく、数週間でのプロジェクトの立ち上げ
- 12 週間以内にトレーニングデータを用意し、モデル開発にかかる時間を数か月単位で短縮
- 社内に対応した場合と比べ、40% 迅速な納品と 100 万ドル以上のコスト削減を実現

重要なインサイト

- 30 ~ 50% 迅速化: データのクリーニングとやり直し作業にかかる時間を削減することで、モデル開発にかかる時間を短縮
- 50 ~ 60% 低いコスト: 当社のグローバルな人材ネットワークと効率的なインフラストラクチャにより、社内でデータチームを構築する場合と比べて大幅なコスト削減を実現
- 3 倍迅速化: 時間のかかる採用プロセス、ツールの準備、トレーニングサイクルを回避することで、実稼働までの時間を短縮
- 社員の負担軽減: 御社の ML チームはデータ関連の作業ではなく、モデル開発に専念できる
- AI データの作成を大規模化: 当社は成果重視型の AI データ作成を手ごろな価格で提供するため、お客様はオーバーヘッドの増加なしに成果を得ることができる



詳しくはこちら

[LIONBRIDGE.COM](https://lionbridge.com)