



LIONBRIDGE 

AI データ収集のケース スタディ

# 1 週間で 28,000 件の データ ポイントを収集

1 週間  
納期

8,000  
追加で提供したデータ  
セットの件数

28,000  
提供した画像の総数

## 課題

お客様は物体認知用の AI モデルをトレーニングさせるため、非常に短期間で大量のデータを必要としていました。具体的にはわずか数日で約 20,000 件の画像が必要でした。またその画像は合成データではなく、人の手で作成された高品質なデータでなければなりません。合成データの中にはときに非常に低品質なデータが含まれ、モデルのパフォーマンスが著しく低下するおそれがあるためです。特に求められていたのは、「物体の中に物体がある」状態の画像です。これはたとえば、かご、バッグ、ボウルなどのような物体の中に別の物体が入っている状況を、AI モデルが識別できるようにするためです。また 1 セットのデータには次の 4 種類の画像が含まれている必要がありました。

- ▶ それぞれの物体が単独で存在する個別の画像
- ▶ ある物体が別の物体の中にある状態の画像
- ▶ 物体が何もない空の背景画像

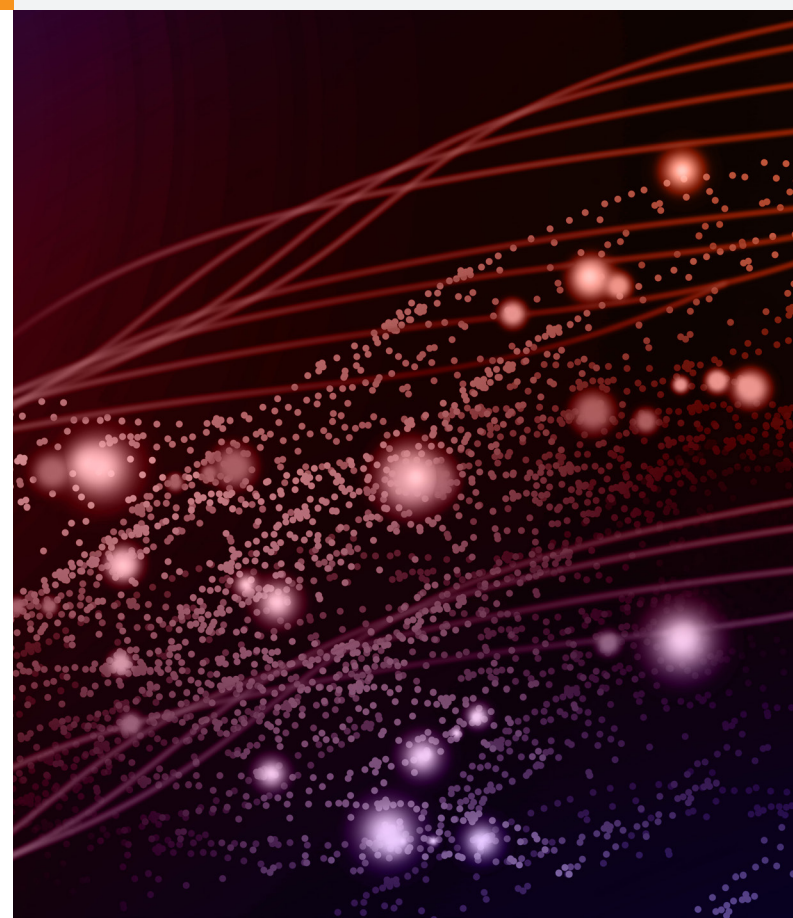
このお客様は多様な視点からのデータを重視していたという点も特徴的です。同じ物体が映っている 20,000 件の画像より、対象の異なる 20,000 種類の画像を収集したほうが、モデルのトレーニングに有用だからです。また世界中の多様な属性の人々から画像を集められれば、モデルがさまざまな属性の利用者を理解し、より適切に動作することにもつながります。

お客様の Web サイトには、さまざまな国、年齢層、ジェンダーなどの人々が、多様な種類の画像を投稿する可能性があります。そのため、多様な属性を含むデータセットで Web サイトのモデルをトレーニングすることで、より幅広い顧客層に対応できるようになります。

## お客様について

今回 Lionbridge AI™ が支援したお客様は、ブランドとブランドが求めるプロのクリエイターを結び付ける事業を展開しています。

このお客様は、プロジェクトの遂行に必要なクリエイターを探しているブランドとクリエイターをつなぐ、革新的なプラットフォームを運営しています。クリエイターがこのプラットフォームに自身のポートフォリオを掲載すると、複数の条件で検索されるようになります。またブランドは求人情報や業務概要を掲載できます。概して、このお客様はブランドとクリエイター間のネットワークを構築し、つながりを促進する役割を担っています。双方がこのプラットフォームを通じてつながり、現在や将来のプロジェクトに向けた関係を構築できます。



## 当社のコミュニティに関する数字



350 以上の言語に対応



450 以上の国や地域で活躍:

- ▶ 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%**



- ▶ 女性: **52%**
- ▶ 男性: **48%**



年齢 **18~60 歳以上**

業界:

- ▶ **16%** ビジネス
- ▶ **16%** エンジニアリング
- ▶ **15%** 人文
- ▶ **13%** 科学
- ▶ **9%** 技術
- ▶ **4%** eコマース
- ▶ **2%** HR
- ▶ **2%** 医療
- ▶ **2%** 芸術
- ▶ **2%** 法律
- ▶ **10%** その他

## 解決策

このお客様からのご要望に対して、当社は **Lionbridge Aurora AI Studio™ プラットフォーム** を使用しました。Lionbridge Aurora AI Studio™ は、当社と世界中で活躍する 50 万名以上の多様な専門家をつなぐプラットフォームです。この専門家コミュニティはさまざまな居住地、使用言語、年齢、ジェンダー、業界の人々から成り、当社はこれらの専門家と連携することで、データ収集など多数のタスクを記録的なスピードで実施しています。

本プロジェクトでは、Aurora AI Studio 上にお客様向けのタスクを設定しました。わずか数時間のうちに 2,000 名を超える人々が集まり、彼らには 8 時間のうちにデータを収集するというタスクが割り当てられました。当社ではこれと並行して、膨大なデータと厳しい納期に対応できるよう、データをチェックするための大規模な社内品質保証チー

## ライオンブリッジのアプローチ

当社は次の 6 段階のプロセスによってお客様をサポートしました。

**ステップ 1:** Aurora AI Studio 上でタスクを作成

**ステップ 2:** 人材を確保

**ステップ 3:** 収集されたデータを受領

**ステップ 4:** データ品質をチェックし、メタデータをラベル付け

**ステップ 5:** 収集したデータをお客様に納品

**ステップ 6:** お客様のご要望に応じて追加でデータを収集・納品

ムを編成しました。チームは収集されたデータが正確で、お客様の要件に沿っていることを確認しました。その後は各画像のメタデータをラベル付けることで、お客様の労力を削減できました。

当社は収集したデータを 3 回に分けて納品しました。お客様ができるだけ早くモデルのトレーニングを開始できるよう、最初のデータセットは期日より前に納品しました。この初回の納品では、3 日間で 12,000 件、2 回目は 5 日後に、残りの 8,000 件の画像を納品しました。

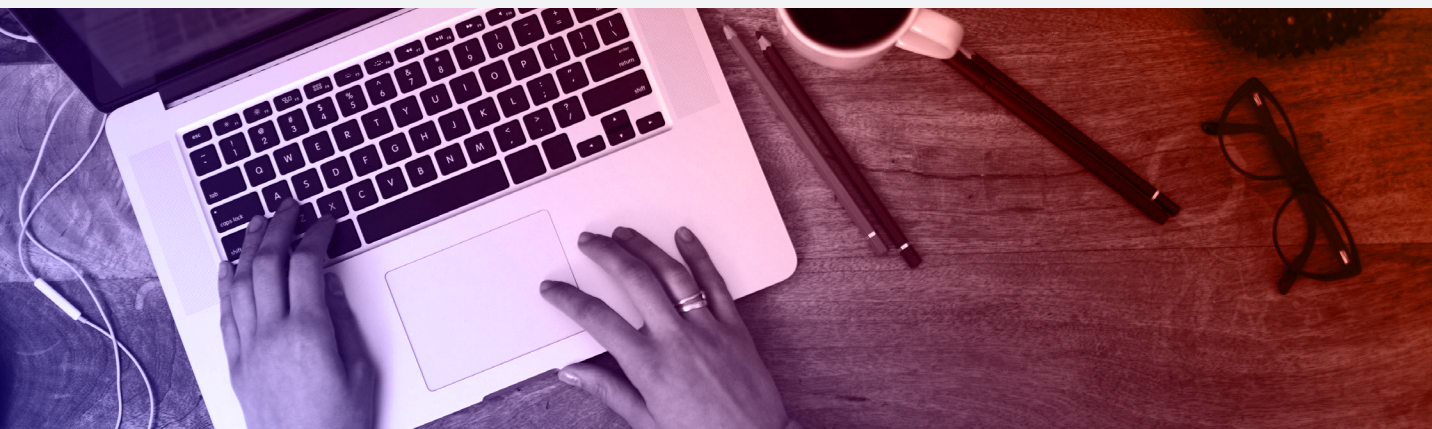
お客様はこの 2 回の納品で、当社の AI トレーニング用データの品質とスピードに大変満足され、さらに 8,000 件の画像を 2 日間の納期で依頼されました。当社はもちろんこれに対応し、合計 28,000 件の画像をわずか 1 週間以内に提供しました。当初依頼されていた 20,000 件の画像は、4 日間のうちにすべて納品しています。





## ● 複数の部門から編成されたチーム

プログラム ディレクター | テクニカル AI エンジニア | QA 責任者 | プロジェクト マネージャー | 社内の QA スペシャリスト (10 名)



## ● 結果

このお客様の支援にあたり、Lionbridge AI は3つの主要な役割を担いました。1つ目は、Lionbridge Aurora AI Studio プラットフォームと多様性に富む当社のコミュニティを活用し、膨大な量のデータを集めた点です。人の手で収集されたこのデータは世界各地の多様な視点を反映しており、高い品質が保たれています。合成データのみを使ったモデルは学習精度が低下する可能性があります。当社のこのデータにより、お客様はそうした弊害を回避できました。さらに当社は大規模な品質保証チームを編成し、納品前のデータを検証しました。こうして完全性が高められたデータは、ブランドやクリエイターの所在地、属性、ポートフォリオの内容に関わらず、それぞれを的確にサポートできるようモデルをトレーニングする際にも役立てることができました。

2つ目は、非常に膨大なデータ（正確には、当初お客様から依頼されていた量を上回るデータ）をわずか1週間で提供した点です。モデルのトレーニングではわずか1日の差が大きなコスト損失につながる可能性があるため、迅速な対応が極めて重要です。当社は Aurora AI Studio プラットフォームを用いて迅速にプロジェクトを完了することで、お客様の人件費や機会損失によるコストを数千ドル規模で削減しました。お客様はわずか1週間で、自社プラットフォームのモデルを十分にトレーニングすることができました。

3つ目は、当初予定されていた量以上のデータを問題なく提供した点です。当初の予定は20,000件の画像でしたが、最終的には28,000件を納品でき、お客様には大変満足していただきました。お客様が必要としていたデータだけではなく、お客様自身も必要だと気付いていなかったデータまで提供できたことは当社の誇りです。期待に応えるだけでなく、それ以上のデータをわずか1週間で提供できたことは非常に大きな成果といえます。



詳しくはこちら  
**LIONBRIDGE.COM/JA**