LIONBRIDGE

AI DATA COLLECTION CASE STUDY

COLLECTING 28,000 DATA POINTS IN A WEEK

1 WEEK
TURNAROUND

8,000
ADDITIONAL DATA SETS DELIVERED

28,000
IMAGES DELIVERED

ABOUT THE CUSTOMER

Lionbridge AI™ supported a customer that helps connect brands and the creative professionals they need for projects. This business runs an innovative platform that connects creative talent and brands that need creative talent for projects. Creative talent can post their portfolios and make them searchable by a few factors. Brands can post jobs or briefs. Overall, this Lionbridge customer facilitates connections and networking between brands and creative talent. They can connect on the platform and build relationships for current and future projects.

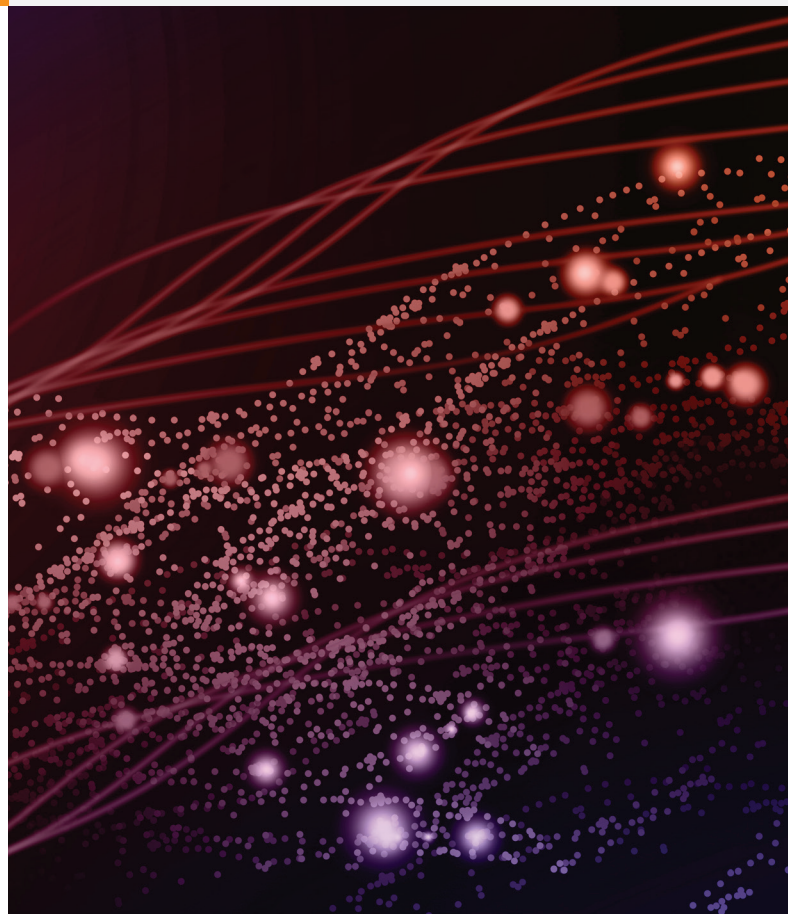
THE CHALLENGE

Lionbridge's customer needed a high volume of data in a tight timeline to train its AI model for object identification. They requested around 20,000 images in a matter of days. The customer needed high-quality, human data, not synthetic images (which can be of much lower caliber and result in notably poorer model performance). In particular, the customer needed images of objects within objects. These would help the AI model learn to identify when an image shows objects within other objects, such as a basket, bag, bowl, etc. They asked for four images in a set:

- Each object separately
- One object in the other
- The empty background with no objects

It's also notable that the customer prioritized data from diverse perspectives. Getting 20,000 images of the same objects would not have helped train their model to perform as well as 20,000 different images. Getting images from participants across many demographics worldwide would help the model understand and perform for audiences of all demographics, too.

People from different countries, age groups, genders, etc., are likely to input many different kinds of images into the customer's website. With a more diverse dataset to train the website's model, it could perform for a broader customer base.



SOME STATISTICS ABOUT OUR CROWD

Speak **350+** languages

Live in **450+** locations worldwide:

- North America: **26%**
- Europe: **29.5%**
- Asia: **27%**
- Africa: **6.5%**
- South America: **8.5%**
- Oceania/Australia: **2.5%**

- Female: **52%**
- Male: **48%**

Ages range from **18 to 60+**

They work in:

- **16%** Business
- **16%** Engineering
- **15%** Humanities
- **13%** Science
- **9%** Technical
- **4%** eCommerce
- **2%** HR
- **2%** Healthcare
- **2%** Fine Arts
- **2%** Legal
- **10%** Other

THE SOLUTION

To service this customer, we used our Lionbridge Aurora AI Studio™ platform. The platform connects us with a crowd of half a million diverse experts from around the world. They display diversity across location, languages spoken, age, gender, and the industry they work in. We engage them to complete many tasks, including data collection, at record speeds.

For this project, we set up the task for the customer in Aurora AI Studio. Within just a few hours, we had engaged over 2,000 participants. Participants were given eight hours to complete the data collection task. Concurrently, we built a larger internal quality assurance team to review the data. We created a larger team due to the volume and the project's tight deadlines.
This team reviewed the data we collected for accuracy and adherence to the customer's needs. After this quality review, we named the metadata for each image, saving the customer hours of labor.

Lionbridge delivered the collected data to the customer in three batches. We delivered the first set early, so the customer could begin training their model as quickly as possible. This batch consisted of 12,000 images over three days. The second batch, delivered in five days, was the remaining 8,000 images.

The customer was so pleased with these batches of collected AI training data that they asked for an additional 8,000 images in two more days.
We were happy to comply and delivered 28,000 images to the customer — all within just one week. The initially requested batch of 20,000 images was all delivered in just four days.

METHODOLOGY

Lionbridge used a multi-step process to support the customer:

Step 1: Create the task in Aurora AI Studio

Step 2: Engage participants

Step 3: Accept collected data

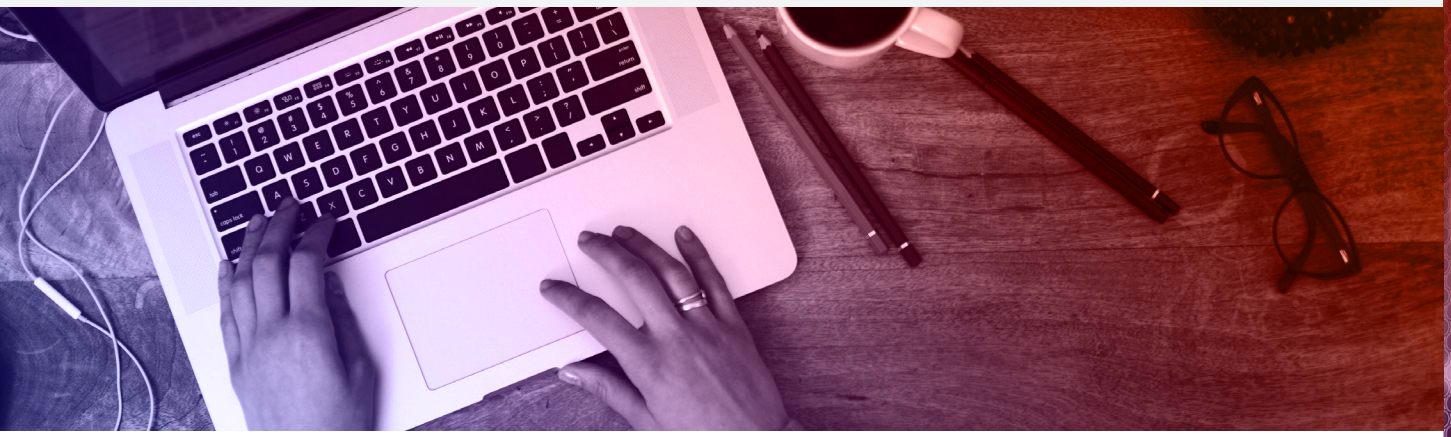
Step 4: Review data and name metadata

Step 5: Deliver collected data to customer

Step 6: Collect and deliver more data, as requested by the customer

MULTIDISCIPLINARY TEAM

Program Director | Technical AI Engineer | QA Lead | Project Manager | 10 Internal QA Specialists



THE RESULTS

Lionbridge AI helped this customer in three key ways. Firstly, using our Aurora AI Studio platform and our extremely diverse crowd, we amassed a colossal volume of high-quality human-collected data that reflected perspectives and images from around the globe. We ensured the customer would avoid the consequences of relying solely on synthetic data, which can lead to a poorly trained model. Lionbridge even called in an expanded quality assurance team to verify the data before delivery. This impeccable data will help train their model to adeptly support their brand and creative talent users, no matter where they're from, their demographics, or what their portfolios include.

The second way we assisted was by delivering this very high volume of data (in fact, even higher than the customer originally requested) in a week. Quick turnaround times are vital when training models. Every day lost can be extremely costly. Lionbridge saved the customer thousands in labor and lost business costs by using our Aurora AI Studio platform to complete the project so quickly. They could ensure their platform's model was adequately trained in just a week.

The last way we wowed this customer was by going beyond their initial request. While the customer originally requested 20,000 images, we procured 28,000. Lionbridge was proud to provide not only the data the customer needed, but also the data they didn't even know they wanted. It was a considerable achievement to give a customer everything they wanted and more — all in just one week.