



## AI 数据收集案例研究

# 在一周内收集 28,000 个数据点

1 周  
周转期

8,000  
个交付的额外数据集

28,000  
张交付的图像

## 面临的挑战

Lionbridge 的这位客户需要在紧迫的时间内获取大量数据,用以训练其物体识别 AI 模型。他们要求在几天内提供大约 20,000 张图像。客户需要高质量的真人数据,而不是合成图像(合成图像的质量可能会低得多,会导致模型性能显著下降)。客户还特别提出,需要“物体中包含物体”的图像。这类图像将帮助 AI 模型学习识别图像中一个物体位于另一物体内部的情况(例如篮子、袋子、碗等容器)。客户要求每组图像包含四张:

- ▶ 每个物体单独呈现
- ▶ 一个物体位于另一物体的内部
- ▶ 无任何物体的空白背景

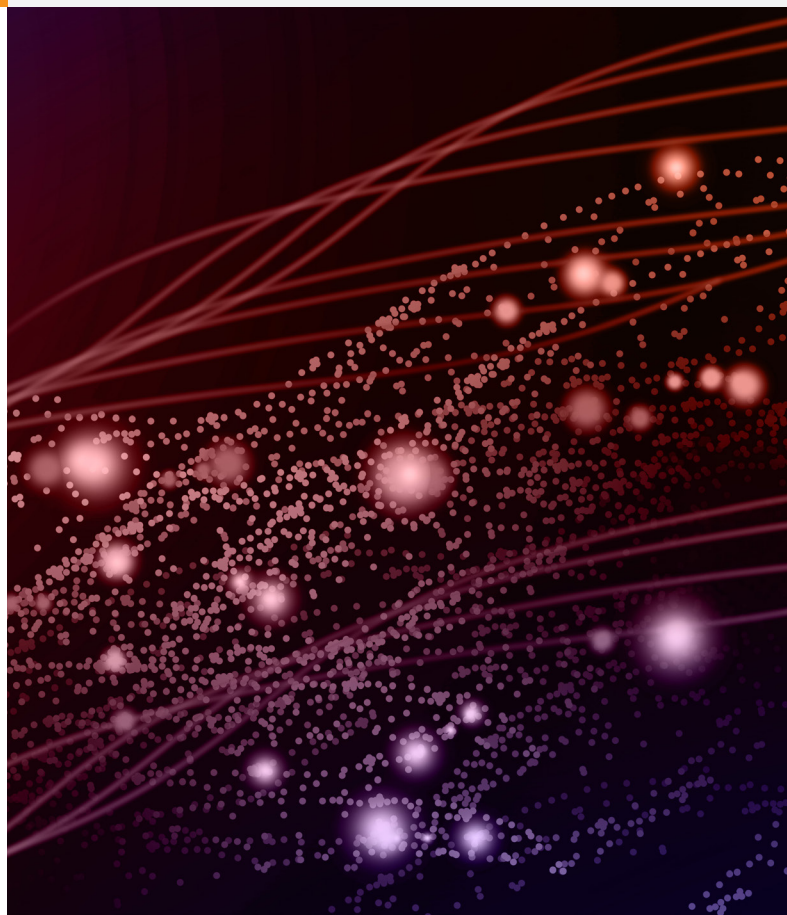
同样值得注意的是,客户特别强调数据应来自不同视角。若 20,000 张图像都是相同物体的,其训练效果远不如使用 20,000 张不同的图像。收集来自世界各地不同人群的图像,有助于模型更好地理解并服务于所有人群。

不同国家/地区、年龄段、性别等的用户可能会向客户网站上传各种类型的图像。使用更多样化的数据集训练网站模型,可以使其更好地服务于更广泛的客户群体。

## 客户简介

Lionbridge AI™ 曾为一位客户提供支持,该客户致力于帮助品牌方与项目所需创意专业人士进行对接。

该企业运营着一个创新平台,专注于连接创意人才与需要创意人才来完成项目的品牌方。创意人才可在平台上发布个人作品集,并设置多项筛选标签,便于品牌方精准查找;品牌方可发布职位或项目需求。总体而言,Lionbridge 的这位客户致力于促进品牌方与创意人才之间的联系与网络构建。双方可在平台上建立联系,为当前及未来的项目搭建合作桥梁。



## 关于我们专家群体的一些统计数据



使用的语言**超过 350 种**



分布于全球 **450 多个地区**:

- 北美:**26%**
- 欧洲:**29.5%**
- 亚洲:**27%**
- 非洲:**6.5%**
- 南美:**8.5%**
- 大洋洲/澳大利亚:**2.5%**



- 女性:**52%**
- 男性:**48%**



年龄范围从 **18 岁到 60 岁以上**

他们所处的行业:

- **16%** 商业
- **16%** 工程
- **15%** 人文
- **13%** 科学
- **9%** 技术
- **4%** 电子商务
- **2%** 人力资源
- **2%** 医疗
- **2%** 艺术
- **2%** 法律
- **10%** 其他

## 解决方案

为了服务这位客户,我们使用了 **Lionbridge Aurora AI Studio™** 平台。

该平台连接着全球超过 50 万名背景多元的专家群体,他们在所在地区、使用语言、年龄、性别及所处行业方面均呈现出多样性。我们邀请这些专家以创纪录的速度完成包括数据收集在内的各类任务。

针对此项目,我们在 Aurora AI Studio 平台上为客户设置了任务。短短数小时内,我们便召集了逾 2,000 名参与者。参与者需在 8 小时内完成数据收集任务。与此同时,我们组建了规模更大的内部质量保障团队进行数据审核。鉴于数据量庞大且项目交期紧迫,我们特别扩充了团队规模。该团队严格审核收集到的数据,确保其准确且符合客户的需求。完成质量审核后,我们还为每张图像标注了元数据,为客户节省了大量人工操作时间。

## 方法

**Lionbridge 采用多步流程为该客户提供帮助:**

- 第 1 步:**在 Aurora AI Studio 平台上创建任务
- 第 2 步:**招募参与者
- 第 3 步:**接收收集到的数据
- 第 4 步:**审核数据并标注元数据
- 第 5 步:**将收集到的数据交付给客户
- 第 6 步:**应客户要求收集并交付更多数据

Lionbridge 分三批向客户交付了收集的数据。首批数据提前交付,使客户能尽快启动模型训练,该批次包含 12,000 张图像,在三天内完成交付;第二批 8,000 张图像则在五天内交付。

客户对这些 AI 训练数据极为满意,要求在两天内再提供 8,000 张图像。我们欣然应允,最终在一周内向客户交付了 28,000 张图像,其中最初要求的 20,000 张图像仅用四天就全部交付完毕。





## 多学科团队

项目总监 | 技术 AI 工程师 | QA 主管 | 项目经理 | 10 位内部 QA 专家



## 成果

**Lionbridge AI 为这位客户提供了三方面的帮助。**首先,我们利用 Aurora AI Studio 平台和极其多元化的专家群体,积累了海量的高质量人工采集数据,这些数据反映了来自世界各地的视角和图像。我们可确保客户不会遭遇因完全依赖合成数据而导致模型训练质量低下的后果。Lionbridge 还特别扩充了质量保证团队,在交付前对所有数据进行核验。这些精准优质的数据将助力客户训练模型,使其能够精准支持品牌方及创意人才用户,无论用户来自何处、属于何种人群,亦或作品集中包含何种内容,都能获得支持。

我们提供的第二项助力是:在一周内交付了海量数据(实际数据量甚至超出了客户最初的要求)。在模型训练过程中,快速周转至关重要,任何延误都可能造成巨大损失。Lionbridge 通过 Aurora AI Studio 平台高效完成项目,为客户节省了数千美元的人工成本和业务损失。客户得以确保其平台模型在一周内完成充分训练。

我们令这位客户惊叹的最后一项助力是,超越了他们的初始要求。客户原本只要求 20,000 张图像,我们却提供了 28,000 张。Lionbridge 不仅为客户提供了他们需要的数据,更主动交付了具有潜在价值的图像,我们为此感到自豪。在短短一周内,不仅满足客户所有需求,更超额交付,这无疑是一项重要成就。

更多详情,请访问

**LIONBRIDGE.COM**