LIONBRIDGE

AI 데이터 수집 사례연구

일주일 만에 데이터포인트 28,000개 수집

일주일	8,000장	28,000장
처리 시간	추가 데이터세트 전달	전달한 이미지 수

고객사 관련 정보

고객사는 브랜드가 프로젝트를 진행하기 위해 필요한 크리에이티브 전문가를 찾을 수 있도록 돕는 기업으로 **Lionbridge AI™를 활용했습니다.**

이 기업은 브랜드가 프로젝트에 필요한 창의적 인재를 찾을 수 있는 혁신적인 플랫폼을 운영합니다. 창의적 인재는 플랫폼에 자신의 포트폴리오를 게시하고 여러 요소로 검색되도록 할 수 있으며, 브랜드는 채용 공고나 기업 정보를 게시할 수 있습니다. 종합하면 이 라이온브리지 고객사는 브랜드와 창의적 인재의 연결과 네트워킹을 촉진합니다. 플랫폼에서 브랜드와 인재는 현재 및 향후 프로젝트를 위한 관계를 구축할 수 있습니다.

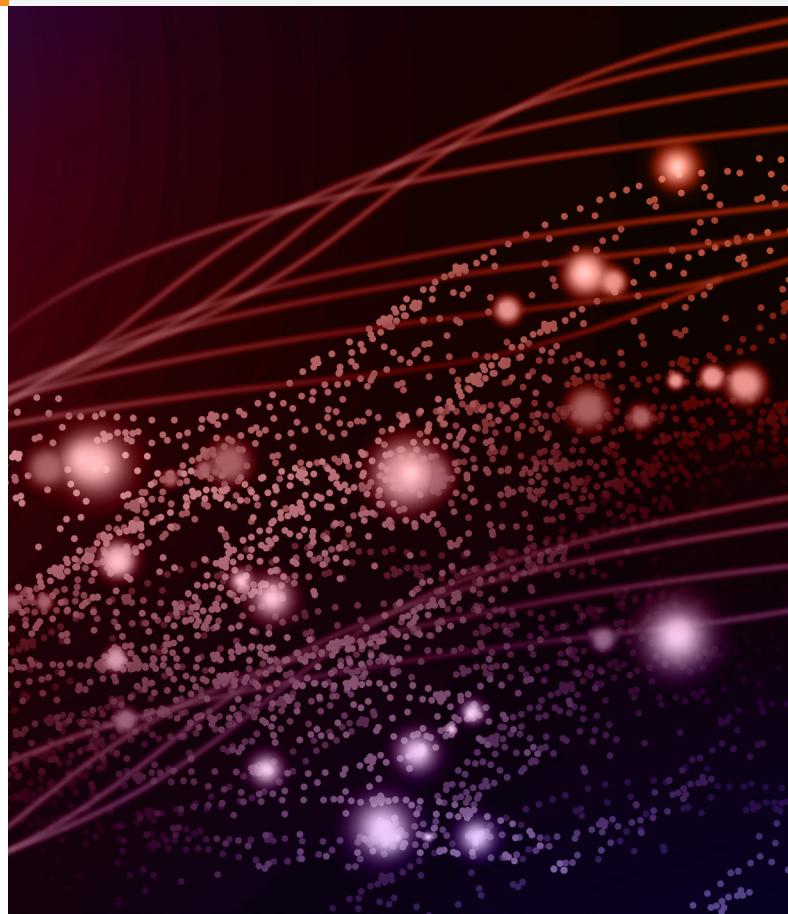
해결 과제

라이온브리지(Lionbridge)의 고객사는 객체 식별을 위한 AI 모델을 학습시키기 위해 짧은 기간 내에 대량의 데이터가 필요했습니다. 고객사는 며칠 만에 약 20,000장의 이미지를 요청했으며, 품질이 낮아 모델 성능이 현저히 저하될 수 있는 합성 이미지가 아닌, 사람이 생성한 고품질의 데이터가 필요했습니다. 특히 고객사는 '객체 내부의 객체' 이미지를 필요로 했습니다. 이러한 이미지는 AI 모델이 바구니, 가방, 그릇 등 물체 안에 다른 물체가 포함된 상황을 학습하는 데 도움이 됩니다. 고객사는 한 세트당 4장의 이미지를 요청했습니다.

- 각 객체의 개별 이미지
- 한 객체가 다른 객체 안에 포함된 이미지
- 객체가 없는 빈 배경 이미지

또한 고객사는 다양한 관점에서의 데이터에 우선순위를 부여했습니다. 동일한 객체에 대한 이미지 2만 장보다는 서로 다른 이미지 2만 장이 모델 훈련에 훨씬 유용합니다. 전 세계 다양한 인구 집단의 참여자들로부터 이미지를 확보하면 모델이 모든 인구학적 특성의 고객층을 이해하고 그에 맞춰 작동하는 데 도움이 됩니다.

다양한 국가, 연령대, 성별의 사람들은 고객사의 웹사이트에 다양한 종류의 이미지를 입력할 가능성이 높습니다. 보다 다양한 데이터세트를 활용해 웹사이트 모델을 학습시키면, 더 광범위한 고객층을 대상으로 모델이 작동할 수 있습니다.



라이온브리지 크라우드에 대한 몇 가지 통계

350개 이상의 언어 사용

전 세계 **450개 이상**의 지역에 거주
- 북미: **26%**
- 유럽: **29.5%**
- 아시아: **27%**
- 아프리카: **6.5%**
- 남미: **8.5%**
- 오세아니아/호주: **2.5%**

- 여성: **52%**
- 남성: **48%**

연령대: **18세에서 60세 이상**

전문 분야:
- **16%** 비즈니스
- **16%** 엔지니어링
- **15%** 인문학
- **13%** 과학
- **9%** 기술
- **4%** 전자상거래
- **2%** 인사
- **2%** 의료
- **2%** 순수 예술
- **2%** 법률
- **10%** 기타

솔루션

라이온브리지는 이 고객사를 지원하기 위해 Lionbridge Aurora AI Studio™ 플랫폼을 활용했습니다. 이 플랫폼에서는 다양한 지역, 사용 언어, 연령, 성별, 종사 산업 분야에 걸쳐 전 세계 50만 명에 달하는 다양한 전문가 집단에 연결될 수 있습니다. 당사는 이들을 활용해 데이터 수집을 포함한 여러 작업을 기록적인 속도로 완료할 수 있습니다.

당사는 이 프로젝트를 위해 Aurora AI Studio에서 고객의 작업을 설정했습니다. 불과 몇 시간 만에 2,000명 이상의 참가자를 확보했습니다. 참가자들은 데이터 수집 작업을 완료하는 데 8시간이 주어졌습니다. 동시에 당사는 데이터를 검토하기 위해 더 큰 규모의 내부 품질 보증 팀을 구성했습니다. 데이터 양과 프로젝트의 촉박한 일정을 고려할 때 대규모 팀이 필요했습니다. 이 팀은 수집된 데이터의 정확성과 고객 요건 충족 여부를 검토했습니다. 품질 검토 후에는 고객사가 수작업 시간을 줄일 수 있도록 각 이미지의 메타데이터를 명명했습니다.

라이온브리지는 수집된 데이터를 고객사에 세 차례에 걸쳐 전달했습니다. 고객사가 모델 학습을 최대한 빨리 시작할 수 있도록 첫 번째 세트를 먼저 전달했습니다. 이 데이터세트는 3일 동안 수집한 12,000장의 이미지로 구성되었습니다. 5일 만에 전달한 두 번째 세트에는 나머지 8,000장의 이미지가 포함되었습니다.

고객사는 이렇게 수집된 AI 학습 데이터에 매우 만족하여 이틀 안에 8,000장의 이미지를 추가로 수집할 것을 요청했습니다. 당사는 기꺼이 이러한 요청에 응하였으며, 결과적으로 총 28,000장의 이미지를 단 일주일 만에 고객사에 전달했습니다. 처음에 요청한 이미지 20,000장은 단 4일 만에 모두 전달되었습니다.

방법론

라이온브리지는 다단계 절차를 통해 고객사를 지원했습니다.

1단계: Aurora AI Studio에서 작업 생성

2단계: 참여자 확보

3단계: 수집된 데이터 수락

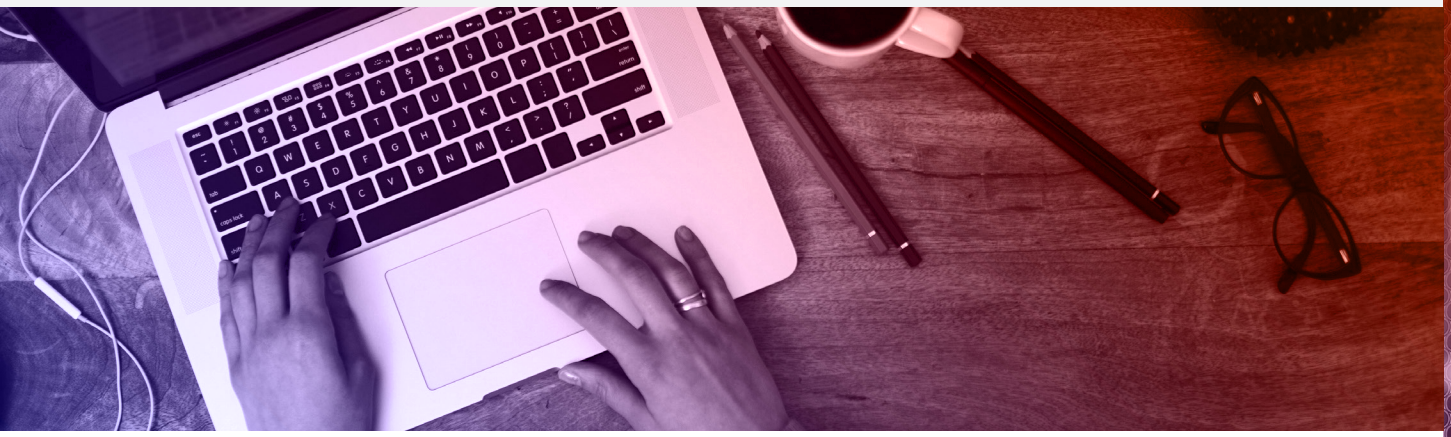
4단계: 데이터 검토 및 메타데이터 명명

5단계: 고객에게 수집된 데이터 전달

6단계: 고객의 요청에 따라 추가 데이터 수집 및 전달

통합전문가 팀

프로그램 디렉터 | AI 기술 엔지니어 | QA 담당 | 프로젝트 담당자 | 내부 QA 전문가 10명



결과

라이온브리지 AI는 세 가지 핵심 방식으로 고객사를 지원했습니다. 첫째, 당사는 Aurora AI Studio 플랫폼과 다양성이 매우 높은 당사의 크라우드 소싱 인력을 활용하여 전 세계 다양한 관점과 이미지가 반영된 방대한 양의 고품질 데이터를 수집했습니다. 이러한 데이터는 사람이 직접 수집함으로써, 고객이 합성 데이터만 사용할 경우 모델 학습 품질이 저하될 수 있는 문제를 방지했습니다. 또한 라이온브리지는 데이터 전달 전 대규모 품질보증 팀까지 투입해 데이터를 검증했습니다. 이처럼 무결점 데이터를 확보함으로써 고객사는 사용자의 출신 지역, 인구학적 특성, 포트폴리오 구성과 무관하게 브랜드와 창의적 인재를 원활히 지원할 수 있도록 모델을 학습시킬 수 있습니다.

두 번째, 당사는 이처럼 방대한 양의 데이터(실제로 고객이 처음에 요청한 양보다 더 많았습니다)를 일주일 만에 전달했습니다. 모델 학습 시 신속한 처리 시간은 매우 중요합니다. 하루라도 지연되면 막대한 비용이 발생할 수 있습니다. 라이온브리지는 Aurora AI Studio 플랫폼을 활용해 프로젝트를 신속히 완료함으로써 고객사가 인건비와 영업 손실 비용으로 수천 달러를 절감할 수 있도록 했습니다. 이를 통해 고객사는 단 일주일 만에 플랫폼 모델을 적절히 학습시킬 수 있었습니다.

마지막으로, 라이온브리지는 고객의 처음 요청을 뛰어넘는 성과를 거두어 고객을 감동시켰습니다. 고객은 원래 20,000장의 이미지를 요청했으나 당사는 28,000장을 조달했습니다. 라이온브리지는 자랑스럽게도 고객이 필요로 하는 데이터뿐만 아니라, 고객조차 필요하다는 것을 몰랐던 데이터까지 제공했습니다. 단 일주일 만에 고객이 원했던 결과 이상을 제공한 것은 상당한 성과였습니다.

자세한 내용:
LIONBRIDGE.COM