

The top half of the image features a city skyline at night, with various skyscrapers illuminated. Overlaid on this are numerous vertical lines of dots in shades of blue, green, and pink, each topped with a small arrow pointing upwards. These lines represent data streams or digital signals. The overall color palette is dominated by dark blues, purples, and oranges.

LIONBRIDGE

AI DATA SERVICES

DATA COLLECTION AND ANNOTATION



Lionbridge powers AI development with high-quality data collection and annotation. Our global crowd delivers diverse, culturally relevant datasets — helping models perform accurately across languages, regions, and real-world scenarios.



WHY AND WHO

Why Do You Need High-Quality Data Collection and Annotation?

AI models are only as good as the data on which they're trained. Many companies struggle to access the required scale, quality, and diversity of data to develop accurate, reliable AI systems. Lionbridge assists with:

1

Data Scarcity

Most organizations lack the in-house resources to collect the needed massive, diverse datasets — especially for edge cases, underrepresented regions, or rare languages.

2

Quality and Consistency

Poorly labeled or inconsistent data leads to model bias, low accuracy, and failed deployments. Our expert annotation workflows and rigorous QA ensure precision from day one.

3

Multilingual and Multimodal Demands

From voice assistants to computer vision, AI systems must work across languages, formats, and environments. We deliver culturally relevant, multilingual, and multimodal datasets at scale.

4

Speed and Scalability

Building internal data teams is slow and costly. We accelerate your models to production by ramping quickly, scaling globally, and supporting rapid iteration cycles.

5

Human-in-the-Loop (HITL) Integration

Even top-tier models need human validation. Our services close the loop with human-reviewed training data that continuously improves performance.

Who Typically Needs Data Collection and Annotation Services?

Our services support teams across industries that build or scale AI. Common users include:



Machine Learning and Data Science Teams

Need large, high-quality datasets to train, validate, and fine-tune models.



AI Product and Engineering Teams

Rely on accurately labeled data to launch features, such as voice assistants, computer vision tools, and search algorithms.



NLP, CV, and Multimodal AI Teams

Require specialized data (text, audio, image, video) annotated for specific tasks, such as entity recognition, sentiment analysis, object detection, or transcription.



Healthcare, Finance, Automotive, and Retail Companies

Use AI for diagnostics, fraud detection, automation, personalization, and more — each requiring domain-specific, accurate training data.



Startups and Enterprises Developing GenAI

Depend on prompt data, synthetic inputs, fine-tuned training sets, and human-evaluated outputs to build competitive AI applications.

OUR SERVICES



Data Annotation



Data Collection



Model Evaluation

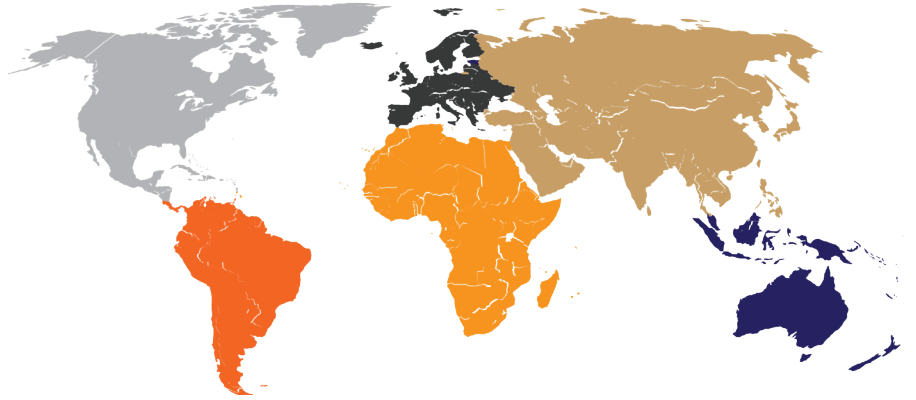
CROWD

Our Crowd is Your Competitive Edge

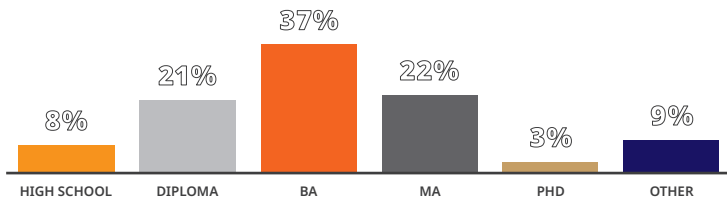
Lionbridge's global crowd combines everyday users with subject matter experts — giving you both scale and depth. Whether you're training an AI model or annotating complex domain-specific data, we source the best contributors for the task. Backed by dedicated project managers and robust QA workflows, we ensure high-quality, culturally relevant data that aligns with real-world use and drives better model performance.

It takes a global community with highly specialized and diverse, expertise and interests.

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



Highest Education Level



Gender



Age



600K
Global Community



350+
Languages



450+
Locales

Industry



Build or Buy? The Cost of Doing It Alone vs. Doing It Right. Real World ROI Use Case.

Standing up an internal data ops team can cost millions in headcount, tools, and time — with no guarantee of quality or scale. A global workforce analytics company needed to annotate labor market data across 30 languages, requiring 1,200 hours of work per language — a total of 36,000 hours of annotation.

Estimated In-House Approach:

- **Linguist Costs:** ~ \$1.8 million
- **Hiring Time:** Recruiting 30 linguists + onboarding, taking 3-4 months
- **Project Duration:** ~6+ months due to ramp-up and staggered delivery
- **Training Data Ready:** ~8+ months total before data could be used to train models

The Lionbridge Approach:

- On-demand global crowd of qualified linguists and bilingual annotators
- Fully managed by Lionbridge PMs with layered QA
- Project launch within weeks, not months
- Training data ready in under 12 weeks, accelerating model development by several months
- **Delivery 40% faster, with savings of over \$1M compared to in-house execution**

KEY INSIGHTS

- **30-50% faster** model deployment by reducing time spent on data cleaning and rework
- **50-60% lower cost** vs. building an in-house data team, thanks to our global contributor network and efficient infrastructure
- **3x faster** time-to-production, avoiding lengthy hiring, tooling, and training cycles
- **Reduced internal burden**, freeing up your ML teams to focus on model development, not data wrangling
- **We make AI data production scalable**, affordable, and outcome-focused. You'll get results, not overhead.



LEARN MORE AT
LIONBRIDGE.COM