

CASE STUDIES PORTFOLIO

Building High-Quality Human Data for the Next Generation of AI

MYK DataWorks is an AI Data Operations company specializing in multilingual transcription, data annotation, audio annotation, RLHF, dataset preparation, and human-in-the-loop workflows for machine learning and AI applications.

10M+ Data Points Processed	20+ Languages Supported	100+ Contributors Network	95–98% Quality-Focused Delivery
--	---	---	---

www.mykdataworks.com

contact@mykdataworks.com

LinkedIn: MYK DataWorks

ABOUT MYK DATAWORKS

Specialized AI Data Operations

MYK DataWorks is a specialized AI Data Operations company dedicated to helping organizations build high-quality datasets for artificial intelligence and machine learning systems.

Our expertise spans multilingual transcription, image annotation, audio annotation, dataset preparation, RLHF, and human-in-the-loop workflows, enabling AI companies to accelerate model development through reliable and scalable data operations.

Unlike traditional outsourcing models, MYK DataWorks follows a managed delivery approach — coordinating skilled contributors, structured quality assurance processes, and standardized workflows to ensure consistent project execution.

OUR MISSION

To empower AI innovation by delivering accurate, scalable, and high-quality human data that enables organizations to build better intelligent systems.

CORE SERVICES

- Multilingual Transcription
 - Audio Annotation
 - RLHF
 - Translation & Localization
- Image Annotation
 - AI Training Data Creation
 - Dataset Preparation
 - Human-in-the-Loop Workflows

WHY MYK DATAWORKS

- 10M+ Data Points Processed
- Support for 20+ Languages
- 100+ Contributors Network
- Managed AI Data Operations
- Structured Quality Assurance
- Scalable Workforce Model

Multilingual Speech Transcription for AI Training Dataset

- Multilingual Transcription
- Speech AI
- AI Training Data

PLATFORM

DataForce

PROJECT DURATION

2 Months

PROJECT OVERVIEW

MYK DataWorks supported a multilingual speech transcription project through the DataForce platform, contributing high-quality training data for Automatic Speech Recognition (ASR) systems.

The project involved converting speech recordings into accurately transcribed, timestamped text while maintaining language-specific consistency and formatting standards for AI model training.

By leveraging a dedicated multilingual workforce and structured review process, MYK DataWorks delivered scalable transcription operations across multiple Indian languages.

BUSINESS CHALLENGE

The client required reliable multilingual speech datasets that could be directly integrated into AI training pipelines. Key requirements included:

- High transcription accuracy
- Timestamp consistency
- Standardized formatting
- Language-specific quality validation
- Scalable execution within strict timelines

MYK DATAWORKS SOLUTION

MYK DataWorks established a dedicated multilingual transcription workflow supported by experienced transcriptionists and linguistic reviewers. Execution included:

- Speech-to-text transcription
- Timestamp alignment
- Linguistic validation
- Dataset formatting
- Multi-level quality assurance

SERVICES DELIVERED

500+ Hours of Speech Transcription

Timestamped Transcript Creation

Multilingual Dataset Preparation

Dataset Formatting

Quality Assurance & Validation

TEAM STRUCTURE

25+

Transcriptionists

Responsible for multilingual speech transcription and primary dataset creation.

5+

Linguistic Reviewers

Performed language-specific validation and consistency checks.

3

QA Supervisors

Oversaw workflow management, quality control, and final dataset verification.

LANGUAGES COVERED

Hindi

Marathi

Tamil

Telugu

WORKFLOW

01

Audio Collection

02

Speech Transcription

03

Timestamp Alignment

04

Linguistic Review

05

Quality Validation

06

Dataset Formatting

07

Final Delivery

OUTCOME

The project successfully produced AI-ready multilingual speech datasets that supported Automatic Speech Recognition model development across multiple Indian languages. Through structured workflows, multilingual expertise, and dedicated quality assurance, MYK DataWorks delivered consistent and scalable transcription outputs aligned with client requirements and machine learning standards.

Image Annotation for Computer Vision Models

- Image Annotation
- Computer Vision
- AI Training Data

PLATFORM
TELUS AI

PROJECT DURATION
2 Months

PROJECT OVERVIEW

MYK DataWorks delivered image annotation services through the TELUS AI platform, supporting high-quality computer vision datasets for machine learning applications.

The project focused on accurately identifying and labeling objects using bounding box annotation techniques while maintaining consistency across diverse image datasets.

The resulting annotations were prepared for direct integration into AI model training pipelines.

BUSINESS CHALLENGE

The client required a scalable annotation workflow capable of producing accurate object labels while maintaining strict quality standards across a large image dataset. Key requirements included:

- Precise object localization
- Consistent annotation guidelines
- Multi-level quality validation
- Efficient workforce coordination
- Production-ready dataset delivery

MYK DATAWORKS SOLUTION

MYK DataWorks established a structured annotation workflow supported by trained image annotators and dedicated quality reviewers. Execution included:

- Bounding Box Annotation
- Object Identification
- Annotation Validation
- Quality Assurance
- Final Dataset Preparation

SERVICES DELIVERED

Bounding Box Annotation

Object Detection Labeling

Annotation Validation

Multi-Level Quality Review

AI Dataset Preparation

TEAM STRUCTURE

20+

Image Annotators

Responsible for accurately identifying and labeling objects across image datasets while following project-specific annotation guidelines.

4

Quality Reviewers

Performed secondary validation to ensure annotation consistency, quality compliance, and production-ready outputs.

ANNOTATION TYPE

Bounding Box Annotation

KEY HIGHLIGHTS

- High-Quality Object Labeling
- Structured Annotation Workflow
- Dedicated Quality Review Process
- AI-Ready Computer Vision Datasets
- Scalable Workforce Execution

WORKFLOW

01

Dataset Intake

02

Image Annotation

03

Quality Review

04

Validation

05

Dataset Formatting

06

Final Delivery

OUTCOME

The project successfully produced structured image datasets for computer vision applications, enabling reliable object detection model training through accurate annotations and rigorous quality validation. By combining experienced annotators with standardized workflows, MYK DataWorks delivered production-ready datasets aligned with enterprise AI requirements.

Multilingual Audio Annotation for Speech AI Systems

Audio Annotation

Speech AI

AI Data Operations

PLATFORM
DataCompute

PROJECT DURATION
5 Months

PROJECT OVERVIEW

MYK DataWorks supported a multilingual audio annotation project through the DataCompute platform, contributing high-quality metadata annotations for speech AI and machine learning applications.

The project involved reviewing short-duration audio clips and classifying each sample across multiple predefined attributes according to client guidelines.

These structured annotations were used to create AI-ready datasets for speech technology and language understanding systems.

BUSINESS CHALLENGE

The client required a scalable workforce capable of accurately reviewing and classifying thousands of short audio recordings across multiple languages with strict consistency. Key requirements included:

- Metadata classification
- Speaker attribute identification
- Accent recognition
- Language verification
- Consistent annotation quality
- Scalable workforce coordination

MYK DATAWORKS SOLUTION

MYK DataWorks assembled a multilingual contributor network and implemented structured annotation workflows to ensure reliable and standardized outputs across all supported languages.

Each audio clip was manually reviewed and labeled according to project-specific instructions before undergoing internal quality validation and final submission.

SERVICES DELIVERED

Audio Metadata Annotation

Speaker Gender Identification

Accent Classification

Language Verification

Multi-Attribute Audio Labeling

Quality Validation

LANGUAGES COVERED

English (US)

English (India)

Hindi

Thai

French

Dutch

Arabic (Egypt)

TEAM STRUCTURE

40+

Audio Annotators

Responsible for reviewing short-duration audio samples and assigning metadata labels according to predefined annotation guidelines.

ANNOTATION FRAMEWORK

Each audio recording was evaluated across approximately 14 predefined annotation segments, including:

Speaker Gender

Speaker Accent

Language Identification

Audio Characteristics

Additional Project-Specific Metadata

WORKFLOW

01

Audio Sample Review

02

Metadata Classification

03

Gender Identification

04

Accent Classification

05

Language Verification

06

Quality Validation

07

Final Dataset Submission

OUTCOME

By leveraging a distributed multilingual workforce of more than 40 annotators, MYK DataWorks successfully delivered structured audio metadata annotations across multiple languages. The completed datasets supported speech AI development by providing consistent, scalable, and high-quality labeled audio data suitable for machine learning and voice technology applications.

AI Training Data Creation – Question & Answer Dataset

- AI Training Data
- NLP
- Large Language Models

PLATFORM

DataForce

PROJECT DURATION

Project-Based

PROJECT OVERVIEW

MYK DataWorks contributed to an AI training data creation project through the DataForce platform, supporting the development of structured Question & Answer (Q&A) datasets for NLP and Large Language Model applications.

The objective was to transform publicly available information into high-quality, contextually accurate question-answer pairs suitable for AI model training and evaluation.

BUSINESS CHALLENGE

The client required a scalable and standardized workflow capable of generating reliable question-answer datasets while maintaining factual accuracy, consistency, and formatting quality. Key requirements included:

- Information extraction from multiple sources
- Context-aware question generation
- Accurate answer creation
- Standardized dataset formatting
- Consistent quality validation

MYK DATAWORKS SOLUTION

MYK DataWorks deployed a dedicated content operations team supported by structured review processes to generate AI-ready datasets. The project execution focused on:

- Information Extraction
- Question Generation
- Answer Generation
- Dataset Structuring
- Quality Review
- Final Dataset Validation

SERVICES DELIVERED

Information Extraction

Structured Question Generation

Context-Aware Answer Creation

Dataset Formatting

Quality Validation

AI Training Dataset Preparation

TEAM STRUCTURE

10

Content Annotators

Responsible for researching source material and generating structured question-answer pairs following project guidelines.

2

Dataset Reviewers

Performed quality validation, formatting checks, and consistency reviews before final submission.

WORKFLOW

01

Source Material Collection

02

Information Extraction

03

Question Generation

04

Answer Generation

05

Quality Review

06

Dataset Formatting

07

Final Delivery

OUTCOME

The completed dataset provided structured question-answer pairs that supported Natural Language Processing and Large Language Model development. By combining human expertise with managed quality

LET'S BUILD BETTER AI TOGETHER

Powering the Next Generation of AI with High-Quality Human Data

At MYK DataWorks, we help AI companies, research organizations, and technology teams build reliable datasets through managed data operations, multilingual expertise, and structured quality assurance.

Whether you're developing speech recognition systems, computer vision models, or large language models, our scalable contributor network and human-in-the-loop workflows are designed to support every stage of your AI journey.

10M+

Data Points Processed

20+

Languages Supported

100+

Contributors Network

Managed

AI Data Operations

CORE CAPABILITIES

Multilingual Transcription

Image Annotation

Audio Annotation

AI Training Data Creation

RLHF & Human Feedback

Dataset Preparation

Human-in-the-Loop Operations

CONTACT US

WEBSITE

www.mykdataworks.com

EMAIL

contact@mykdataworks.com

LINKEDIN

linkedin.com/in/mykdataworks