

AI Training Dataset Market by Dataset Creation (Data Collection, Data Annotation, Synthetic Data Generation), Dataset Selling (Off-the-Shelf Datasets, Dataset Marketplaces), Data Modality (Text, Image, Video, Audio, Multimodal) - Global Forecast to 2029

<https://marketpublishers.com/r/AD17BE55D414EN.html>

Date: October 2024

Pages: 447

Price: US\$ 4,950.00 (Single User License)

ID: AD17BE55D414EN

Abstracts

The market for AI training datasets is expected to increase from USD 2.82 billion in 2024 to USD 9.58 billion in 2029, experiencing a compound annual growth rate (CAGR) of 27.7% from 2024 to 2029. The demand for AI training datasets is rapidly increasing as various sectors look for more machine learning and AI uses. A key factor driving the growth of the market is the increasing demand for top-notch, varied data collections to properly train AI models, especially in industries such as healthcare, finance, and autonomous vehicles. However, concerns regarding data privacy and compliance with regulations continue to pose a major barrier that could hinder data collection and restrict access to personal data. Businesses encounter difficulties in obtaining and controlling data that comply with performance and regulation requirements, while also harmonizing innovation and ethical factors.

“By offering, dataset creation segment is expected to register the fastest market growth rate during the forecast period.”

The dataset creation segment is expected to have the quickest increase in the market in the forecast period, due to the growing need for top-notch data in different industries. Businesses are realizing the significance of making decisions based on data and are therefore making substantial investments in developing thorough and precise sets of data. This part takes advantage of AI and ML progress, which simplify data collection and processing, enabling businesses to create datasets more quickly and on a larger scale. Additionally, the rapid growth of this sector is fueled by the increasing number of IoT devices, and the growing amount of data produced from digital interactions. Companies are prioritizing the creation of large data sets to conduct predictive analysis,

comprehend customer actions, and devise tailored marketing tactics to improve their results. Rules like GDPR and CCPA have prompted businesses to focus on ethical ways of collecting data, creating a demand for customized datasets that abide by the regulations. Companies require tailored data sets to meet specific business requirements in order to stay competitive in their respective industries and experience market growth.

“By dataset selling, Off-the-Shelf (OTS) datasets segment is expected to have the largest market share during the forecast period.”

The OTS datasets are expected to lead the dataset selling segment in market because of their inexpensive price, easy access, and immediate suitability for various uses. Companies are opting for pre-made datasets more often as they save time on data collection and preparation, enabling a swift adoption of data-driven strategies. The rising demand for data analysis in different sectors such as healthcare, finance, and marketing are pushing this trend further, as companies seek to leverage existing data for improved decision-making and obtaining valuable insights. In addition, the rise of artificial intelligence and machine learning technologies has raised the demand for top-notch data to train models, resulting in a heavier reliance on pre-made datasets. The use of ready-made datasets is expected to rise steadily in the upcoming years as businesses prioritize adaptability and remaining competitive.

“By annotation type, synthetic datasets segment is expected to register the fastest market growth rate during the forecast period.”

Throughout the predicted period, the synthetic datasets segment in the AI training dataset market is expected to experience the most significant increase in growth rate. Synthetic datasets generate abundant data simulating real-world scenarios, solving problems of insufficient data and privacy issues associated with authentic datasets. Customizing synthetic data to suit particular purposes increases its attractiveness, since it can be tailored to fulfill the diverse demands of artificial intelligence models across different industries. Progress in developing models and simulation techniques enhances the accuracy and authenticity of synthetic data, ultimately boosting its efficacy in training machine learning algorithms. The demand for robust and flexible datasets is projected to increase as companies focus on improving their AI capabilities, underscoring the importance of synthetic datasets in future AI projects. This phenomenon is encouraging ethical AI methods by employing artificial data to reduce prejudice and ensure fairer outcomes in AI uses.

“By Region, North America to have the largest market share in 2024, and Asia Pacific is slated to grow at the fastest rate during the forecast period.”

In 2024, North America is expected to dominate the AI training dataset market with the largest market share. The reason for this dominance is the existence of big tech firms, significant investments in AI, and a strong network of data-centric advancements.

Companies in North America are increasingly integrating artificial intelligence to enhance their operations, leading to a demand for high-quality training data. In the meantime, it is expected that the Asia Pacific region will show the highest rate of growth in the predicted period. The rapid expansion is due to additional investments in AI, higher internet usage, and a growing number of AI and machine learning startups. China and India are leading the way in embracing AI technologies, thanks to their abundant data and young population well-versed in technology.

Breakdown of primaries

In-depth interviews were conducted with Chief Executive Officers (CEOs), innovation and technology directors, system integrators, and executives from various key organizations operating in the AI training dataset market.

? By Company: Tier I – 18%, Tier II – 52%, and Tier III – 30%

? By Designation: C-Level Executives – 42%, D-Level Executives – 36%, and others – 22%

? By Region: North America – 42%, Europe – 26%, Asia Pacific – 21%, Middle East & Africa – 4%, and Latin America – 7%

The report includes the study of key players offering AI training dataset solutions. It profiles major vendors in the AI training dataset market. The major players in the AI training dataset market include Google (US), IBM (US), AWS (US), Microsoft (US), NVIDIA (US), Snorkel (US), Gretel (US), Shaip (US), Clickworker (US), Appen (Australia), Nexdata (US), Bitext (US), Aimleap (US), Deep Vision Data (US), Cogito Tech (US), Sama (US), Scale AI (US), Lionbridge Technologies (US), Alegion (US), TELUS International (Canada), iMerit (US), Labelbox (US), V7Labs (UK), Defined.ai (US), SuperAnnotate (US), LXT (Canada), Toloka AI (Netherlands), Innodata (US), Kili technology (France), HumanSignal (US), Superb AI (US), Hugging Face (US), CloudFactory (UK), FileMarket (Hong Kong), TagX (UAE), Roboflow (US), Supervise.ly (Estonia), Encord (UK), TransPerfect (US), Keylabs (Israel), and Data.world (US).

Research coverage

This research report categorizes the AI training dataset Market by Offering (Dataset Creation and Dataset Selling), by Dataset Creation (Dataset Creation Software, and Dataset Creation Services), by Dataset Selling (Off-The-Shelf (OTS) Datasets, and Dataset Marketplaces), by Annotation Type (Pre-Labeled Datasets, Unlabeled Datasets, and Synthetic Datasets), by Data Modality (Text, Image, Audio & Speech, Video and Multimodal), By Type (Generative AI and Other AI), by End User (BFSI, Software & Technology Providers, Telecommunications, Automotive, Media & Entertainment, Government & Defense, Healthcare & Life Sciences, Manufacturing, Retail & Consumer Goods, And Other End Users) and by Region (North America, Europe, Asia Pacific, Middle East & Africa, and Latin America). The scope of the report covers detailed information regarding the major factors, such as drivers, restraints,

challenges, and opportunities, influencing the growth of the AI training dataset market. A detailed analysis of the key industry players has been done to provide insights into their business overview, solutions, and services; key strategies; contracts, partnerships, agreements, new product & service launches, mergers and acquisitions, and recent developments associated with the AI training dataset market. Competitive analysis of upcoming startups in the AI training dataset market ecosystem is covered in this report.

Key Benefits of Buying the Report

The report would provide the market leaders/new entrants in this market with information on the closest approximations of the revenue numbers for the overall AI training dataset market and its subsegments. It would help stakeholders understand the competitive landscape and gain more insights better to position their business and plan suitable go-to-market strategies. It also helps stakeholders understand the pulse of the market and provides them with information on key market drivers, restraints, challenges, and opportunities.

The report provides insights on the following pointers:

- **Analysis of key drivers** (increasing demand for diverse and continuously updated multimodal datasets for generative AI models, rising demand for multilingual datasets for conversational AI, demand for high-quality labeled data for autonomous vehicles, and Increased used of synthetic data for rare event simulation), restraints (legal risks of web-scraped data due to copyright infringement and limited access to high-quality medical datasets due to HIPAA compliance), opportunities (growing demand for specialized data annotation services in diverse fields, synthetic data generation and privacy-preserving techniques for augmented training data, and creation of customized AI Datasets and specialized formats (3D, AR/VR) for Enterprise Solutions), and challenges (data quality and relevance issues like inconsistency, bias, keeping datasets up to date, and diverse dataset formats and inconsistent annotation practices may hinder integration and reliability).
- **Product Development/Innovation:** Detailed insights on upcoming technologies, research & development activities, and new product & service launches in the AI training dataset market.
- **Market Development:** Comprehensive information about lucrative markets – the report analyses the AI training dataset market across varied regions.
- **Market Diversification:** Exhaustive information about new products & services, untapped geographies, recent developments, and investments in the AI training dataset market.
- **Competitive Assessment:** In-depth assessment of market shares, growth strategies and service offerings of leading players like Google (US), IBM (US), AWS (US), Microsoft (US), NVIDIA (US), Snorkel (US), Gretel (US), Shaip (US), Clickworker (US),

Appen (Australia), Nexdata (US), Bitext (US), Aimleap (US), Deep Vision Data (US), Cogito Tech (US), Sama (US), Scale AI (US), Lionbridge Technologies (US), Alegion (US), TELUS International (Canada), iMerit (US), Labelbox (US), V7Labs (UK), Defined.ai (US), SuperAnnotate (US), LXT (Canada), Toloka AI (Netherlands), Innodata (US), Kili technology (France), HumanSignal (US), Superb AI (US), Hugging Face (US), CloudFactory (UK), FileMarket (Hong Kong), TagX (UAE), Roboflow (US), Supervise.ly (Estonia), Encord (UK), TransPerfect (US), Keylabs (Israel), and Data.world (US) among others in the AI training dataset market. The report also helps stakeholders understand the pulse of the AI training dataset market and provides them with information on key market drivers, restraints, challenges, and opportunities.

Contents

1 INTRODUCTION

- 1.1 STUDY OBJECTIVES
- 1.2 MARKET DEFINITION
 - 1.2.1 INCLUSIONS AND EXCLUSIONS
- 1.3 MARKET SCOPE
 - 1.3.1 MARKET SEGMENTATION
 - 1.3.2 YEARS CONSIDERED
- 1.4 CURRENCY CONSIDERED
- 1.5 STAKEHOLDERS

2 RESEARCH METHODOLOGY

- 2.1 RESEARCH DATA
 - 2.1.1 SECONDARY DATA
 - 2.1.2 PRIMARY DATA
 - 2.1.2.1 Breakup of primary profiles
 - 2.1.2.2 Key industry insights
- 2.2 MARKET BREAKUP AND DATA TRIANGULATION
- 2.3 MARKET SIZE ESTIMATION
 - 2.3.1 TOP-DOWN APPROACH
 - 2.3.2 BOTTOM-UP APPROACH
- 2.4 MARKET FORECAST
- 2.5 RESEARCH ASSUMPTIONS
- 2.6 RESEARCH LIMITATIONS

3 EXECUTIVE SUMMARY

4 PREMIUM INSIGHTS

- 4.1 ATTRACTIVE OPPORTUNITIES FOR PLAYERS IN AI TRAINING DATASET MARKET
- 4.2 AI TRAINING DATASET MARKET, BY TOP THREE DATA MODALITIES
- 4.3 NORTH AMERICA: AI TRAINING DATASET MARKET, BY ANNOTATION TYPE AND END USER
- 4.4 AI TRAINING DATASET MARKET, BY REGION

5 MARKET OVERVIEW AND INDUSTRY TRENDS

5.1 INTRODUCTION

5.2 MARKET DYNAMICS

5.2.1 DRIVERS

5.2.1.1 Increasing need for diverse and continuously updated multimodal datasets for generative AI models

5.2.1.2 Rising use of multilingual datasets in conversational AI

5.2.1.3 Growing demand for high-quality labeled data for autonomous vehicles

5.2.1.4 Rising adoption of synthetic data for rare event simulation

5.2.2 RESTRAINTS

5.2.2.1 Legal risks of web-scraped data due to copyright infringement

5.2.2.2 Limited access to high-quality medical datasets due to HIPAA compliance

5.2.3 OPPORTUNITIES

5.2.3.1 Growing demand for specialized data annotation services in diverse fields

5.2.3.2 Synthetic data generation and privacy-preserving techniques for augmented training data

5.2.3.3 Creation of customized AI datasets and specialized formats for enterprise solutions

5.2.4 CHALLENGES

5.2.4.1 Data quality and relevance issues

5.2.4.2 Diverse dataset formats and inconsistent annotation practices

5.3 EVOLUTION OF AI TRAINING DATASET

5.4 SUPPLY CHAIN ANALYSIS

5.5 ECOSYSTEM ANALYSIS

5.5.1 DATA COLLECTION SOFTWARE PROVIDERS

5.5.2 DATA LABELING AND ANNOTATION PLATFORM PROVIDERS

5.5.3 SYNTHETIC DATA PROVIDERS

5.5.4 DATA AUGMENTATION TOOL PROVIDERS

5.5.5 OFF-THE-SHELF (OTS) DATASET PROVIDERS

5.5.6 AI TRAINING DATASET SERVICE PROVIDERS

5.6 INVESTMENT AND FUNDING SCENARIO

5.7 IMPACT OF GENERATIVE AI ON AI TRAINING DATASET MARKET

5.7.1 DATA AUGMENTATION FOR IMAGE RECOGNITION

5.7.2 SYNTHETIC TEXT GENERATION FOR NLP

5.7.3 SPEECH AND AUDIO DATA SYNTHESIS

5.7.4 SIMULATED USER INTERACTION DATA

5.7.5 BIAS MITIGATION IN DATASETS

5.7.6 SCENARIO TESTING FOR PREDICTIVE MODELS

5.8 CASE STUDY ANALYSIS

5.8.1 CASE STUDY 1: CLICKWORKER BOOSTS AI TRAINING DATASET FOR AUTOMOTIVE SYSTEMS, IMPROVING SPEECH RECOGNITION ACCURACY

5.8.2 CASE STUDY 2: APPEN ENHANCES MICROSOFT TRANSLATOR WITH COMPREHENSIVE AI TRAINING DATASETS FOR 110 LANGUAGES

5.8.3 CASE STUDY 3: COGITO TECH LLC ENHANCES CARDIAC SURGERY WITH AI-DRIVEN AORTIC VALVE DATASETS

5.8.4 CASE STUDY 4: ENHANCING AI TRAINING DATASETS FOR PAIN REDUCTION THROUGH HINGE HEALTH'S SUCCESS WITH SUPERANNOTATE

5.8.5 CASE STUDY 5: OUTREACH ENHANCES AI TRAINING WITH LABEL STUDIO

5.8.6 CASE STUDY 6: ENCORD ADDRESSES KEY CHALLENGES IN SURGICAL VIDEO ANNOTATION FOR ENHANCED DATA QUALITY AND EFFICIENCY

5.9 TECHNOLOGY ANALYSIS

5.9.1 KEY TECHNOLOGIES

5.9.1.1 Data labeling and annotation

5.9.1.2 Synthetic data generation

5.9.1.3 Data augmentation

5.9.1.4 Human-in-the-loop (HITL) feedback systems

5.9.1.5 Active learning

5.9.1.6 Data cleansing and preprocessing

5.9.1.7 Bias detection and mitigation

5.9.1.8 Dataset versioning and management

5.9.2 COMPLEMENTARY TECHNOLOGIES

5.9.2.1 Cloud storage and data lakes

5.9.2.2 MLOps and model management

5.9.2.3 Data governance

5.9.2.4 Machine learning frameworks

5.9.3 ADJACENT TECHNOLOGIES

5.9.3.1 Federated learning

5.9.3.2 Edge AI for data processing

5.9.3.3 Differential privacy

5.9.3.4 AutoML

5.9.3.5 Transfer learning

5.10 REGULATORY LANDSCAPE

5.10.1 REGULATORY BODIES, GOVERNMENT AGENCIES, AND OTHER ORGANIZATIONS

5.10.2 REGULATIONS: AI TRAINING DATASET

5.10.2.1 North America

5.10.2.1.1 Blueprint for an AI Bill of Rights (US)

5.10.2.1.2 Directive on Automated Decision-Making (Canada)

5.10.2.2 Europe

5.10.2.2.1 UK AI Regulation White Paper

5.10.2.2.2 Gesetz zur Regulierung Künstlicher Intelligenz (AI Regulation Law - Germany)

5.10.2.2.3 Loi pour une République numérique (Digital Republic Act - France)

5.10.2.2.4 Codice in materia di protezione dei dati personali (Data Protection Code - Italy)

5.10.2.2.5 Ley de Servicios Digitales (Digital Services Act - Spain)

5.10.2.2.6 Dutch Data Protection Authority (Autoriteit Persoonsgegevens)

Guidelines

5.10.2.2.7 The Swedish National Board of Trade AI Guidelines

5.10.2.2.8 Danish Data Protection Agency (Datatilsynet) AI Recommendations

5.10.2.2.9 Artificial Intelligence 4.0 (AI 4.0) Program - Finland

5.10.2.3 Asia Pacific

5.10.2.3.1 Personal Data Protection Bill (PDPB) & National Strategy on AI (NSAI) - India

5.10.2.3.2 The Basic Act on the Advancement of Utilizing Public and Private Sector Data & AI Guidelines - Japan

5.10.2.3.3 New Generation Artificial Intelligence Development Plan & AI Ethics

Guidelines - China

5.10.2.3.4 Framework Act on Intelligent Informatization – South Korea

5.10.2.3.5 AI Ethics Framework (Australia) & AI Strategy (New Zealand)

5.10.2.3.6 Model AI Governance Framework - Singapore

5.10.2.3.7 National AI Framework - Malaysia

5.10.2.3.8 National AI Roadmap - Philippines

5.10.2.4 Middle East & Africa

5.10.2.4.1 Saudi Data & Artificial Intelligence Authority (SDAIA) Regulations

5.10.2.4.2 UAE National AI Strategy 2031

5.10.2.4.3 Qatar National AI Strategy

5.10.2.4.4 National Artificial Intelligence Strategy (2021-2025)- Turkey

5.10.2.4.5 African Union (AU) AI Framework

5.10.2.4.6 Egyptian Artificial Intelligence Strategy

5.10.2.4.7 Kuwait National Development Plan (New Kuwait Vision 2035)

5.10.2.5 Latin America

5.10.2.5.1 Brazilian General Data Protection Law (LGPD)

5.10.2.5.2 Federal Law on the Protection of Personal Data Held by Private Parties - Mexico

5.10.2.5.3 Argentina Personal Data Protection Law (PDPL) & AI Ethics Framework

- 5.10.2.5.4 Chilean Data Protection Law & National AI Policy
- 5.10.2.5.5 Colombian Data Protection Law (Law 1581) & AI Ethics Guidelines
- 5.10.2.5.6 Peruvian Personal Data Protection Law & National AI Strategy

5.11 PATENT ANALYSIS

- 5.11.1 METHODOLOGY
- 5.11.2 PATENTS FILED, BY DOCUMENT TYPE
- 5.11.3 INNOVATION AND PATENT APPLICATIONS

5.12 PRICING ANALYSIS

- 5.12.1 PRICING DATA, BY OFFERING
- 5.12.2 PRICING DATA, BY PRODUCT TYPE

5.13 KEY CONFERENCES AND EVENTS, 2024–2025

5.14 PORTER'S FIVE FORCES ANALYSIS

- 5.14.1 THREAT OF NEW ENTRANTS
- 5.14.2 THREAT OF SUBSTITUTES
- 5.14.3 BARGAINING POWER OF SUPPLIERS
- 5.14.4 BARGAINING POWER OF BUYERS
- 5.14.5 INTENSITY OF COMPETITIVE RIVALRY

5.15 KEY STAKEHOLDERS AND BUYING CRITERIA

- 5.15.1 KEY STAKEHOLDERS IN BUYING PROCESS
- 5.15.2 BUYING CRITERIA

5.16 TRENDS/DISRUPTIONS IMPACTING CUSTOMER BUSINESS

6 AI TRAINING DATASET MARKET, BY OFFERING

6.1 INTRODUCTION

- 6.1.1 OFFERING: AI TRAINING DATASET MARKET DRIVERS

6.2 DATASET CREATION

- 6.2.1 DATASET CREATION KEY TO DEVELOPING ROBUST AI APPLICATIONS

6.3 DATASET SELLING

- 6.3.1 MONETIZING DATA FOR AI DEVELOPMENT THROUGH ETHICAL DATA SELLING

7 AI TRAINING DATASET MARKET, BY DATASET CREATION

7.1 INTRODUCTION

- 7.1.1 DATASET CREATION: AI TRAINING DATASET MARKET DRIVERS

7.2 DATASET CREATION SOFTWARE

- 7.2.1 DATASET CREATION SOFTWARE FUELING INNOVATIONS ACROSS VARIOUS SECTORS

7.2.2 DATA COLLECTION SOFTWARE

7.2.2.1 Web scraping tools

7.2.2.2 Data sourcing API

7.2.2.3 Crowdsourcing platforms

7.2.2.4 Sensor data collection software

7.2.3 DATA LABELING & ANNOTATION

7.2.3.1 Image annotation

7.2.3.2 Text annotation

7.2.3.3 Video annotation

7.2.3.4 Audio annotation

7.2.3.5 3D data annotation

7.2.4 SYNTHETIC DATA GENERATION SOFTWARE

7.2.5 DATA AUGMENTATION SOFTWARE

7.3 DATASET CREATION SERVICES

7.3.1 CUSTOMIZED DATA CREATION SERVICES FOR OPTIMAL AI MODEL ALIGNMENT

7.3.2 DATA COLLECTION SERVICES

7.3.3 DATA ANNOTATION & LABELING SERVICES

7.3.4 DATA VALIDATION SERVICES

8 AI TRAINING DATASET MARKET, BY DATASET SELLING

8.1 INTRODUCTION

8.1.1 DATASET SELLING: AI TRAINING DATASET MARKET DRIVERS

8.2 OFF-THE-SHELF (OTS) DATASETS

8.2.1 SCALABILITY AND EASE OF DISTRIBUTION MAKE OTS DATASETS APPEALING FOR AI TRAINING

8.3 DATASET MARKETPLACES

8.3.1 DATASET MARKETPLACES ACCELERATE AI INNOVATION BY DEMOCRATIZING ACCESS TO CRITICAL RESOURCES

9 AI TRAINING DATASET MARKET, BY ANNOTATION TYPE

9.1 INTRODUCTION

9.1.1 ANNOTATION TYPE: AI TRAINING DATASET MARKET DRIVERS

9.2 PRE-LABELED DATASETS

9.2.1 HIGH-QUALITY PRE-LABELED DATASETS ACCELERATE AI DEVELOPMENT ACROSS VARIOUS SECTORS

9.3 UNLABELED DATASETS

- 9.3.1 UNLABELED DATASETS ENABLE ROBUST AI MODEL TRAINING
- 9.4 SYNTHETIC DATASETS
 - 9.4.1 ADVANCEMENTS IN GENERATIVE MODELS ENHANCE QUALITY OF SYNTHETIC DATASETS

10 AI TRAINING DATASET MARKET, BY DATA MODALITY

- 10.1 INTRODUCTION
 - 10.1.1 DATA TYPE: AI TRAINING DATASET MARKET DRIVERS
- 10.2 TEXT
 - 10.2.1 BUSINESSES PRIORITIZE CURATING DIVERSE, LABELED TEXT DATASETS TO ENHANCE MODEL ACCURACY
 - 10.2.2 TEXT CLASSIFICATION
 - 10.2.3 CHATBOTS
 - 10.2.4 SENTIMENT ANALYSIS
 - 10.2.5 DOCUMENT PARSING
 - 10.2.6 OTHER TEXT DATA MODALITIES
- 10.3 IMAGE
 - 10.3.1 ADVANCEMENTS IN DEEP LEARNING TECHNIQUES, PARTICULARLY CONVOLUTIONAL NEURAL NETWORKS, ELEVATE ROLE OF IMAGE DATA IN AI DEVELOPMENT
 - 10.3.2 OBJECT DETECTION
 - 10.3.3 FACIAL RECOGNITION
 - 10.3.4 MEDICAL IMAGING
 - 10.3.5 SATELLITE IMAGERY
 - 10.3.6 OTHER IMAGE DATA MODALITIES
- 10.4 AUDIO & SPEECH
 - 10.4.1 RISING POPULARITY OF VOICE-ACTIVATED TECHNOLOGIES FUELS DEMAND FOR DIVERSE, HIGH-QUALITY AUDIO DATASETS
 - 10.4.2 SPEECH RECOGNITION
 - 10.4.3 AUDIO CLASSIFICATION
 - 10.4.4 MUSIC GENERATION
 - 10.4.5 VOICE SYNTHESIS
 - 10.4.6 OTHER AUDIO & SPEECH DATA MODALITIES
- 10.5 VIDEO
 - 10.5.1 SURGE IN DEMAND FOR HIGH-QUALITY LABELED VIDEO DATASETS AS ORGANIZATIONS SEEK TO HARNESS VIDEO CONTENT POTENTIAL
 - 10.5.2 ACTION RECOGNITION
 - 10.5.3 AUTONOMOUS DRIVING

10.5.4 VIDEO SURVEILLANCE

10.5.5 VIDEO CONTENT MODERATION

10.5.6 OTHER VIDEO DATA MODALITIES

10.6 MULTIMODAL

10.6.1 RISING DEMAND FOR MULTIMODAL DATASETS BOOSTS INNOVATION
AND ADVANCES IN AI APPLICATIONS

10.6.2 SPEECH-TO-TEXT

10.6.3 CONTENT RECOMMENDATION

10.6.4 VISUAL QUESTION ANSWERING (VQA)

10.6.5 MULTIMODAL ANALYTICS

10.6.6 OTHER MULTIMODALITIES

11 AI TRAINING DATASET MARKET, BY TYPE

11.1 INTRODUCTION

11.1.1 TYPE: AI TRAINING DATASET MARKET DRIVERS

11.2 GENERATIVE AI

11.2.1 GENERATIVE AI REVOLUTIONIZES CREATIVITY ACROSS INDUSTRIES
THROUGH DIVERSE TRAINING DATASETS

11.2.2 LLM EVALUATION

11.2.3 RAG OPTIMIZATION

11.2.4 LLM FINE TUNING

11.2.5 CONVERSATIONAL AGENTS

11.2.6 CONTENT CREATION

11.2.7 CODE GENERATION

11.2.8 OTHER GENERATIVE AI

11.3 OTHER AI

11.3.1 RISING ROLE OF NLP AND COMPUTER VISION IN ENTERPRISE AI
APPLICATIONS TO BOOST OTHER AI DATASET DEMAND

11.3.2 NATURAL LANGUAGE PROCESSING (NLP)

11.3.2.1 Text classification

11.3.2.2 Named entity recognition (NER)

11.3.2.3 Sentiment analysis

11.3.2.4 Document parsing and extraction

11.3.3 COMPUTER VISION

11.3.3.1 Image classification

11.3.3.2 Object detection

11.3.3.3 Video analysis

11.3.3.4 Optical character recognition (OCR)

11.3.4 PREDICTIVE ANALYTICS

11.3.4.1 Time series forecasting

11.3.4.2 Anomaly detection

11.3.4.3 Customer behavior prediction

11.3.4.4 Risk scoring and management

11.3.5 RECOMMENDATION SYSTEMS

11.3.5.1 Product and content recommendations

11.3.5.2 Personalized marketing and ads

11.3.5.3 Collaborative filtering

11.3.6 SPEECH AND AUDIO PROCESSING

11.3.6.1 Speech recognition

11.3.6.2 Audio classification

11.3.6.3 Voice command recognition

11.3.6.4 Speech-to-text transcription

11.3.7 OTHER TYPES

12 AI TRAINING DATASET MARKET, BY END USER

12.1 INTRODUCTION

12.1.1 END USER: AI TRAINING DATASET MARKET DRIVERS

12.2 BFSI

12.2.1 FINANCIAL INSTITUTIONS LEVERAGE AI TRAINING DATASETS TO ENHANCE FRAUD DETECTION AND RISK MANAGEMENT

12.2.2 BANKING

12.2.3 FINANCIAL SERVICES

12.2.4 INSURANCE

12.3 TELECOMMUNICATIONS

12.3.1 TELECOM COMPANIES BOOST PERFORMANCE AND CUSTOMER SERVICES WITH AI-POWERED INTELLIGENT SYSTEMS

12.4 GOVERNMENT & DEFENSE

12.4.1 AI TRAINING DATASETS PROPEL ADVANCES IN NATIONAL SECURITY AND DEFENSE OPERATIONS

12.5 HEALTHCARE & LIFE SCIENCES

12.5.1 AI TRAINING DATASETS SPEARHEAD TRANSFORMATIVE BREAKTHROUGHS IN PRECISION MEDICINE AND DIAGNOSTICS

12.6 MANUFACTURING

12.6.1 AI TRAINING DATASETS DRIVE EFFICIENCY IN MANUFACTURING WITH AUTOMATION AND PREDICTIVE MAINTENANCE

12.7 RETAIL & CONSUMER GOODS

12.7.1 RETAILERS ENHANCE PERSONALIZED CUSTOMER EXPERIENCES WITH AI-DRIVEN RECOMMENDATIONS AND OPTIMIZED SUPPLY CHAINS

12.8 SOFTWARE & TECHNOLOGY PROVIDERS

12.8.1 INNOVATION ACCELERATES AS SOFTWARE AND TECHNOLOGY PROVIDERS HARNESS AI TRAINING DATASETS FOR CUTTING-EDGE SOLUTIONS

12.8.2 CLOUD HYPERSCALERS

12.8.3 FOUNDATION MODEL/LLM PROVIDERS

12.8.4 AI TECHNOLOGY PROVIDERS

12.8.5 IT & IT-ENABLED SERVICE PROVIDERS

12.9 AUTOMOTIVE

12.9.1 RAPID ADVANCEMENTS IN AUTONOMOUS VEHICLE DEVELOPMENT FUELED BY AI TRAINING DATASETS CAPTURING REAL-WORLD DRIVING BEHAVIORS AND CONDITIONS

12.10 MEDIA & ENTERTAINMENT

12.10.1 AI TRAINING DATASETS FUEL INNOVATION IN CONTENT CREATION ACROSS MEDIA, GAMING, AND ENTERTAINMENT INDUSTRIES

12.11 OTHER END USERS

13 AI TRAINING DATASET MARKET, BY REGION

13.1 INTRODUCTION

13.2 NORTH AMERICA

13.2.1 NORTH AMERICA: AI TRAINING DATASET MARKET DRIVERS

13.2.2 NORTH AMERICA: MACROECONOMIC OUTLOOK

13.2.3 US

13.2.3.1 Reliance of companies across various sectors on large, diverse datasets to improve accuracy and performance of AI algorithms to drive market

13.2.4 CANADA

13.2.4.1 Government focus on gathering insights from stakeholders to maximize AI investment benefits to drive market

13.3 EUROPE

13.3.1 EUROPE: AI TRAINING DATASET MARKET DRIVERS

13.3.2 EUROPE: MACROECONOMIC OUTLOOK

13.3.3 UK

13.3.3.1 Rising demand for quality data and innovative solutions from various sectors to drive market

13.3.4 GERMANY

13.3.4.1 Industry demand, government support, and data privacy regulations to drive

market

13.3.5 FRANCE

13.3.5.1 Increasing adoption of AI solutions by tech companies and startups to maintain competitive edge

13.3.6 ITALY

13.3.6.1 Advances in data collection and management enable companies to access diverse datasets tailored to various AI applications

13.3.7 SPAIN

13.3.7.1 Strategic government initiatives and industry innovation to drive market

13.3.8 NETHERLANDS

13.3.8.1 Focus on ethical AI and expanding digital infrastructure to accelerate demand for high-quality, diverse training datasets

13.3.9 REST OF EUROPE

13.4 ASIA PACIFIC

13.4.1 ASIA PACIFIC: AI TRAINING DATASET MARKET DRIVERS

13.4.2 ASIA PACIFIC: MACROECONOMIC OUTLOOK

13.4.3 CHINA

13.4.3.1 Increasing demand for high-quality data for training models from various sectors to drive market

13.4.4 JAPAN

13.4.4.1 Supportive government policies and strategic corporate initiatives to drive market

13.4.5 INDIA

13.4.5.1 Increasing demand for AI solutions across various sectors to drive market

13.4.6 SOUTH KOREA

13.4.6.1 Increasing AI adoption and necessity for high-quality datasets to drive market

13.4.7 AUSTRALIA

13.4.7.1 Demand for quality data and ethical standards to drive market

13.4.8 SINGAPORE

13.4.8.1 Initiatives like Infocomm Media Development Authority (IMDA) promote data literacy and use of AI

13.4.9 REST OF ASIA PACIFIC

13.5 MIDDLE EAST & AFRICA

13.5.1 MIDDLE EAST & AFRICA: AI TRAINING DATASET MARKET DRIVERS

13.5.2 MIDDLE EAST & AFRICA: MACROECONOMIC OUTLOOK

13.5.3 MIDDLE EAST

13.5.3.1 UAE

13.5.3.1.1 Initiatives by healthcare sector to build vast medical datasets for

predictive analytics and disease detection to drive market

13.5.3.2 Saudi Arabia

13.5.3.2.1 Launch of Saudi Open Data Platform and partnership with global tech firms to accelerate AI training dataset development

13.5.3.3 Qatar

13.5.3.3.1 Strategic investments in startups specializing in streaming data to drive market

13.5.3.4 Turkey

13.5.3.4.1 Government initiatives and increasing demand for high-quality datasets from various sectors to drive market

13.5.3.5 Rest of Middle East

13.5.4 AFRICA

13.5.4.1 Increasing potential for AI application in various sectors to drive market

13.6 LATIN AMERICA

13.6.1 LATIN AMERICA: AI TRAINING DATASET MARKET DRIVERS

13.6.2 LATIN AMERICA: MACROECONOMIC OUTLOOK

13.6.3 BRAZIL

13.6.3.1 Growth in IT and healthcare sectors to drive market

13.6.4 MEXICO

13.6.4.1 Government initiatives and private sector investments to drive market

13.6.5 ARGENTINA

13.6.5.1 Government transparency initiatives and startup support to drive market

13.6.6 REST OF LATIN AMERICA

14 COMPETITIVE LANDSCAPE

14.1 OVERVIEW

14.2 KEY PLAYER STRATEGIES/RIGHT TO WIN, 2021–2024

14.3 REVENUE ANALYSIS, 2019–2023

14.4 MARKET SHARE ANALYSIS, 2023

14.4.1 MARKET RANKING ANALYSIS

14.5 PRODUCT COMPARATIVE ANALYSIS

14.5.1 AWS SAGEMAKER (AWS)

14.5.2 AI DATA PLATFORM (APPEN)

14.5.3 SAMA PLATFORM (SAMA)

14.5.4 DATA ENGINE, SCALE GEN AI PLATFORM (SCALE AI)

14.5.5 IMERIT PLATFORMS (IMERIT)

14.6 COMPANY VALUATION AND FINANCIAL METRICS, 2024

14.7 COMPANY EVALUATION MATRIX: KEY PLAYERS, 2023

- 14.7.1 STARS
- 14.7.2 EMERGING LEADERS
- 14.7.3 PERVASIVE PLAYERS
- 14.7.4 PARTICIPANTS
- 14.7.5 COMPANY FOOTPRINT: KEY PLAYERS, 2023
 - 14.7.5.1 Company footprint
 - 14.7.5.2 Region footprint
 - 14.7.5.3 Offering footprint
 - 14.7.5.4 Data modality footprint
 - 14.7.5.5 End user footprint
- 14.8 COMPANY EVALUATION MATRIX: STARTUPS/SMES, 2023
 - 14.8.1 PROGRESSIVE COMPANIES
 - 14.8.2 RESPONSIVE COMPANIES
 - 14.8.3 DYNAMIC COMPANIES
 - 14.8.4 STARTING BLOCKS
 - 14.8.5 COMPETITIVE BENCHMARKING: STARTUPS/SMES, 2023
 - 14.8.5.1 Detailed list of key startups/SMEs
 - 14.8.5.2 Competitive benchmarking of key startups/SMEs
- 14.9 COMPETITIVE SCENARIO
 - 14.9.1 PRODUCT LAUNCHES AND ENHANCEMENTS
 - 14.9.2 DEALS

15 COMPANY PROFILES

- 15.1 INTRODUCTION
- 15.2 KEY PLAYERS
 - 15.2.1 GOOGLE
 - 15.2.1.1 Business overview
 - 15.2.1.2 Products/Solutions/Services offered
 - 15.2.1.3 Recent developments
 - 15.2.1.3.1 Product launches and enhancements
 - 15.2.1.3.2 Deals
 - 15.2.1.4 MnM view
 - 15.2.1.4.1 Key strengths
 - 15.2.1.4.2 Strategic choices
 - 15.2.1.4.3 Weaknesses and competitive threats
 - 15.2.2 MICROSOFT
 - 15.2.2.1 Business overview
 - 15.2.2.2 Products/Solutions/Services offered

- 15.2.2.3 Recent developments
 - 15.2.2.3.1 Product launches and enhancements
- 15.2.2.4 MnM view
 - 15.2.2.4.1 Key strengths
 - 15.2.2.4.2 Strategic choices
 - 15.2.2.4.3 Weaknesses and competitive threats
- 15.2.3 AWS
 - 15.2.3.1 Business overview
 - 15.2.3.2 Products/Solutions/Services offered
 - 15.2.3.3 Recent developments
 - 15.2.3.3.1 Product launches and enhancements
 - 15.2.3.3.2 Deals
 - 15.2.3.4 MnM view
 - 15.2.3.4.1 Key strengths
 - 15.2.3.4.2 Strategic choices
 - 15.2.3.4.3 Weaknesses and competitive threats
- 15.2.4 APPEN
 - 15.2.4.1 Business overview
 - 15.2.4.2 Products/Solutions/Services offered
 - 15.2.4.3 Recent developments
 - 15.2.4.3.1 Product launches and enhancements
 - 15.2.4.3.2 Deals
 - 15.2.4.4 MnM view
 - 15.2.4.4.1 Key strengths
 - 15.2.4.4.2 Strategic choices
 - 15.2.4.4.3 Weaknesses and competitive threats
- 15.2.5 NVIDIA
 - 15.2.5.1 Business overview
 - 15.2.5.2 Products/Solutions/Services offered
 - 15.2.5.3 Recent developments
 - 15.2.5.3.1 Product launches and enhancements
 - 15.2.5.4 MnM view
 - 15.2.5.4.1 Key strengths
 - 15.2.5.4.2 Strategic choices
 - 15.2.5.4.3 Weaknesses and competitive threats
- 15.2.6 IBM
 - 15.2.6.1 Business overview
 - 15.2.6.2 Products/Solutions/Services offered
- 15.2.7 TELUS INTERNATIONAL

- 15.2.7.1 Business overview
- 15.2.7.2 Products/Solutions/Services offered
- 15.2.8 INNODATA
 - 15.2.8.1 Business overview
 - 15.2.8.2 Products/Solutions/Services offered
 - 15.2.8.3 Recent developments
 - 15.2.8.3.1 Product launches and enhancements
- 15.2.9 COGITO TECH
 - 15.2.9.1 Business overview
 - 15.2.9.2 Products/Solutions/Services offered
- 15.2.10 SAMA
 - 15.2.10.1 Business overview
 - 15.2.10.2 Products/Solutions/Services offered
 - 15.2.10.3 Recent developments
 - 15.2.10.3.1 Product launches and enhancements
- 15.2.11 CLICKWORKER
- 15.2.12 TRANSPERFECT
- 15.2.13 CLOUDFACTORY
- 15.2.14 IMERIT
- 15.2.15 LIONBRIDGE TECHNOLOGIES
- 15.2.16 SCALE AI
- 15.3 STARTUPS/SMES
 - 15.3.1 SNORKEL AI
 - 15.3.2 GRETEL
 - 15.3.3 SHAIPI
 - 15.3.4 NEXDATA
 - 15.3.5 BITEXT
 - 15.3.6 AIMLEAP
 - 15.3.7 ALEGION
 - 15.3.8 DEEP VISION DATA
 - 15.3.9 LABELBOX
 - 15.3.10 V7LABS
 - 15.3.11 DEFINED.AI
 - 15.3.12 SUPERANNOTATE
 - 15.3.13 TOLOKA AI
 - 15.3.14 KILI TECHNOLOGY
 - 15.3.15 HUMANSIGNAL
 - 15.3.16 SUPERB AI
 - 15.3.17 HUGGING FACE

- 15.3.18 FILEMARKET
- 15.3.19 TAGX
- 15.3.20 ROBOFLOW
- 15.3.21 SUPERVISELY
- 15.3.22 ENCORD
- 15.3.23 KEYLABS
- 15.3.24 LXT
- 15.3.25 DATA.WORLD

16 ADJACENT AND RELATED MARKETS

16.1 INTRODUCTION

16.2 DATA ANNOTATION AND LABELING MARKET

16.2.1 MARKET DEFINITION

16.2.2 MARKET OVERVIEW

- 16.2.2.1 Data annotation and labeling market, by component
- 16.2.2.2 Data annotation and labeling market, by data type
- 16.2.2.3 Data annotation and labeling market, by deployment type
- 16.2.2.4 Data annotation and labeling market, by organization size
- 16.2.2.5 Data annotation and labeling market, by annotation type
- 16.2.2.6 Data annotation and labeling market, by application
- 16.2.2.7 Data annotation and labeling market, by vertical
- 16.2.2.8 Data annotation and labeling market, by region

16.3 SYNTHETIC DATA GENERATION MARKET

16.3.1 MARKET DEFINITION

16.3.2 MARKET OVERVIEW

- 16.3.2.1 Synthetic data generation market, by offering
- 16.3.2.2 Synthetic data generation market, by data type
- 16.3.2.3 Synthetic data generation market, by application
- 16.3.2.4 Synthetic data generation market, by vertical
- 16.3.2.5 Synthetic data generation market, by region

17 APPENDIX

17.1 DISCUSSION GUIDE

17.2 KNOWLEDGESTORE: MARKETSANDMARKETS' SUBSCRIPTION PORTAL

17.3 CUSTOMIZATION OPTIONS

17.4 RELATED REPORTS

17.5 AUTHOR DETAILS

I would like to order

Product name: AI Training Dataset Market by Dataset Creation (Data Collection, Data Annotation, Synthetic Data Generation), Dataset Selling (Off-the-Shelf Datasets, Dataset Marketplaces), Data Modality (Text, Image, Video, Audio, Multimodal) - Global Forecast to 2029

Product link: <https://marketpublishers.com/r/AD17BE55D414EN.html>

Price: US\$ 4,950.00 (Single User License / Electronic Delivery)

If you want to order Corporate License or Hard Copy, please, contact our Customer Service:

info@marketpublishers.com

Payment

To pay by Credit Card (Visa, MasterCard, American Express, PayPal), please, click button on product page <https://marketpublishers.com/r/AD17BE55D414EN.html>