

# **Data-Centric AI Development Market Forecasts to 2034 – Global Analysis By Component (Tools & Platforms and Services), Data Type, Deployment Mode, Data Lifecycle Stage, Application, End User and By Geography**

<https://marketpublishers.com/r/DE439A410BBBEN.html>

Date: May 2026

Pages: 200

Price: US\$ 4,150.00 (Single User License)

ID: DE439A410BBBEN

## **Abstracts**

According to Statistics MRC, the Global Data-Centric AI Development Market is accounted for \$8.4 billion in 2026 and is expected to reach \$32.1 billion by 2034 growing at a CAGR of 18.2% during the forecast period. Data-centric AI development refers to the systematic methodology of improving artificial intelligence model performance by prioritizing the quality, consistency, labeling accuracy, and representativeness of training datasets over model architecture optimization alone, supported by specialized tooling platforms for data collection, cleaning, annotation, versioning, and quality management throughout the AI development lifecycle. These platforms incorporate active learning frameworks, automated data quality assessment engines, crowdsourced annotation management systems, and data-driven model debugging tools that enable AI engineers to systematically identify and resolve data defects that limit production model accuracy across vision, language, speech, and structured prediction tasks.

Market Dynamics:

Driver:

Production AI accuracy demands

Enterprise deployment of AI systems in high-stakes applications, including medical diagnosis, autonomous vehicle control, financial fraud detection, and industrial quality inspection, is generating rigorous accuracy and reliability requirements that can only be achieved through systematic data quality management rather than model architecture improvements alone. Organizations deploying production AI systems are discovering that 80 percent of model performance problems originate in training data defects rather

than algorithmic limitations, driving systematic investment in data-centric development infrastructure that guarantees consistent annotation quality, eliminates systematic labeling errors, and ensures comprehensive edge case coverage.

Restraint:

Data annotation cost and scale

Producing large volumes of accurately labeled training data for complex AI tasks, including medical image segmentation, autonomous driving scene understanding, and multi-language NLP, requires substantial investment in specialized annotator recruitment, training, quality assurance, and management infrastructure that creates significant cost barriers limiting data-centric AI adoption among smaller organizations. Enterprise AI teams requiring millions of high-precision annotations face annotation cost structures that consume disproportionate shares of AI development budgets, while maintaining annotation quality consistency across large distributed annotator workforces introduces systematic variance that undermines the data quality improvements that data-centric approaches are designed to achieve.

Opportunity:

Synthetic data generation adoption

Advances in generative AI and simulation technology enabling high-fidelity synthetic training data generation for scenarios where real-world data collection is prohibitively expensive, privacy-restricted, or safety-prohibitive represent a transformative opportunity for data-centric AI development platform vendors to expand addressable markets beyond annotation services into integrated data generation and management solutions. Automotive AI developers using synthetic sensor data, healthcare AI companies generating synthetic patient records compliant with privacy regulations, and robotics firms simulating edge case scenarios are driving rapid adoption of synthetic data platforms that integrate directly with data quality management infrastructure.

Threat:

AutoML and foundation models

Rapid advancement of large foundation models pre-trained on internet-scale datasets that achieve strong performance on downstream tasks with minimal fine-tuning data is potentially reducing the volume of custom training data required for many enterprise AI applications, threatening the demand for large-scale data annotation and quality management services that underpin data-centric AI development platform revenue. If foundation model transfer learning capabilities continue improving to the point where enterprise AI applications require only hundreds of high-quality examples rather than millions of annotated samples, the structural demand for extensive data-centric development infrastructure may decline significantly across mainstream AI use cases.

Covid-19 Impact:

The pandemic dramatically accelerated enterprise AI adoption across remote work, e-

commerce, healthcare diagnostics, and supply chain management, which intensified demand for production-quality AI systems requiring rigorous training data infrastructure. Remote work requirements drove the rapid development of distributed annotation workforce management platforms, enabling global data labeling operations. Post-pandemic, enterprise AI maturity has advanced to the stage where production deployment quality and regulatory compliance requirements make data-centric development methodology adoption a strategic necessity rather than an optional best practice.

The services segment is expected to be the largest during the forecast period. The services segment is expected to account for the largest market share during the forecast period, due to the premium value of specialized expertise guiding enterprise organizations through data strategy design, annotation workflow architecture, and production AI deployment that most internal teams lack without external support. Large enterprises undertaking strategic AI transformation programs require comprehensive consulting engagements covering data governance frameworks, annotation vendor selection, quality assurance protocol design, and AI model auditing that generate substantial professional services revenue. Major consulting firms and specialized AI services companies are scaling data-centric AI practices to meet enterprise demand. The structured data segment is expected to have the highest CAGR during the forecast period.

Over the forecast period, the structured data segment is predicted to witness the highest growth rate, driven by the massive expansion of enterprise AI applications in financial services, healthcare records management, supply chain optimization, and customer analytics that rely on structured tabular and transactional data as the primary training input. Financial institutions deploying AI fraud detection, credit risk, and trading systems are investing heavily in structured data quality management infrastructure to meet regulatory model validation requirements. The proliferation of cloud data warehouses is accelerating structured data AI development by centralizing quality management across enterprise data pipelines.

Region with largest share:

During the forecast period, the North America region is expected to hold the largest market share, due to the world's highest concentration of enterprise AI development activity, leading AI research institutions, and data-centric platform startups receiving significant venture capital investment. The United States hosts the largest ecosystem of AI development tooling companies, including Scale AI, Labelbox, and Weights & Biases, that are building a comprehensive data-centric development infrastructure. Enterprise technology companies, including Google, Microsoft, and Amazon, are making substantial investments in data quality and management tooling integrated with their AI development cloud platforms.

### Region with highest CAGR:

Over the forecast period, the Asia Pacific region is expected to exhibit the highest CAGR, driven by the acceleration of enterprise AI adoption in China, India, South Korea, and Japan, combined with government AI development programs that mandate domestic AI capability building, generating substantial institutional demand for data-centric development platforms. China's national AI strategy, which is driving large-scale AI deployment in manufacturing, healthcare, and financial services, is creating enormous training data production requirements. India's growing AI services export industry and domestic digital transformation programs are driving strong investment in data annotation and quality management platforms.

### Key players in the market

Google LLC, Microsoft Corporation, Amazon Web Services Inc., IBM Corporation, Snowflake Inc., Databricks Inc., Scale AI Inc., Appen Limited, Samasource Inc., Alteryx Inc., DataRobot Inc., H2O.ai Inc., Oracle Corporation, SAP SE, Cloudera Inc., Teradata Corporation, and C3.ai Inc..

### Key Developments:

In April 2026, Databricks Inc. expanded its Mosaic AI platform with data-centric model evaluation tools enabling systematic identification and remediation of training data quality issues in large language model fine-tuning pipelines.

In February 2026, Snorkel AI Inc. announced a major enterprise partnership with a leading healthcare provider to deploy programmatic data labeling infrastructure for clinical AI model development across radiology and pathology applications.

In January 2026, Labelbox Inc. introduced integrated synthetic data generation capabilities within its data-centric AI platform, enabling seamless blending of real and synthetic training examples for improved model robustness.

### Solution Types Covered:

Carbon Monitoring Platforms

AI-Based Soil Analytics

Carbon Credit Platforms

MRV (Measurement Reporting Verification) Tools

Predictive Carbon Modeling Systems

Soil Data Intelligence Platforms

### Farm Types Covered:

Row Crop Farms

Permanent Crop Farms

Mixed Farms

Agroforestry Systems

### Technologies Covered:

Machine Learning Models

Remote Sensing & Satellite Analytics

IoT Soil Sensors

Big Data Platforms

Blockchain for Carbon Credits

### Applications Covered:

Carbon Credit Generation

Soil Health Monitoring

Sustainable Farming Planning

Climate Reporting

Regenerative Agriculture

### End Users Covered:

Farmers

Agribusiness Companies

Carbon Credit Developers

Government Organizations

### Regions Covered:

North America

United States

Canada

Mexico

Europe

United Kingdom

Germany

France

Italy

Spain

Netherlands

Belgium

Sweden

Switzerland

Poland

Rest of Europe

### Asia Pacific

China

Japan

India

South Korea

Australia

Indonesia

Thailand

Malaysia

Singapore

Vietnam

Rest of Asia Pacific

### South America

Brazil

Argentina

Colombia

Chile

Peru

Rest of South America

Rest of the World (RoW)

Middle East

Saudi Arabia

United Arab Emirates

Qatar

Israel

Rest of Middle East

Africa

South Africa

Egypt

Morocco

Rest of Africa

What our report offers:

Market share assessments for the regional and country-level segments

Strategic recommendations for the new entrants

Covers Market data for the years 2023, 2024, 2025, 2026, 2027, 2028, 2030, 2032 and 2034

Market Trends (Drivers, Constraints, Opportunities, Threats, Challenges, Investment Opportunities, and recommendations)

Strategic recommendations in key business segments based on the market estimations

Competitive landscaping mapping the key common trends

Company profiling with detailed strategies, financials, and recent developments

Supply chain trends mapping the latest technological advancements

#### Free Customization Offerings:

All the customers of this report will be entitled to receive one of the following free customization options:

##### Company Profiling

Comprehensive profiling of additional market players (up to 3)

SWOT Analysis of key players (up to 3)

##### Regional Segmentation

Market estimations, Forecasts and CAGR of any prominent country as per the client's interest (Note: Depends on feasibility check)

##### Competitive Benchmarking

Benchmarking of key players based on product portfolio, geographical presence, and strategic alliances

## Contents

### **1 EXECUTIVE SUMMARY**

- 1.1 Market Snapshot and Key Highlights
- 1.2 Growth Drivers, Challenges, and Opportunities
- 1.3 Competitive Landscape Overview
- 1.4 Strategic Insights and Recommendations

### **2 RESEARCH FRAMEWORK**

- 2.1 Study Objectives and Scope
- 2.2 Stakeholder Analysis
- 2.3 Research Assumptions and Limitations
- 2.4 Research Methodology
  - 2.4.1 Data Collection (Primary and Secondary)
  - 2.4.2 Data Modeling and Estimation Techniques
  - 2.4.3 Data Validation and Triangulation
  - 2.4.4 Analytical and Forecasting Approach

### **3 MARKET DYNAMICS AND TREND ANALYSIS**

- 3.1 Market Definition and Structure
- 3.2 Key Market Drivers
- 3.3 Market Restraints and Challenges
- 3.4 Growth Opportunities and Investment Hotspots
- 3.5 Industry Threats and Risk Assessment
- 3.6 Technology and Innovation Landscape
- 3.7 Emerging and High-Growth Markets
- 3.8 Regulatory and Policy Environment
- 3.9 Impact of COVID-19 and Recovery Outlook

### **4 COMPETITIVE AND STRATEGIC ASSESSMENT**

- 4.1 Porter's Five Forces Analysis
  - 4.1.1 Supplier Bargaining Power
  - 4.1.2 Buyer Bargaining Power
  - 4.1.3 Threat of Substitutes
  - 4.1.4 Threat of New Entrants

- 4.1.5 Competitive Rivalry
- 4.2 Market Share Analysis of Key Players
- 4.3 Product Benchmarking and Performance Comparison

## **5 GLOBAL DATA-CENTRIC AI DEVELOPMENT MARKET, BY COMPONENT**

- 5.1 Tools & Platforms
  - 5.1.1 Data Labeling Tools
  - 5.1.2 Data Versioning Platforms
  - 5.1.3 Data Quality Management Tools
- 5.2 Services
  - 5.2.1 Data Annotation Services
  - 5.2.2 AI Consulting Services
  - 5.2.3 Data Engineering Services

## **6 GLOBAL DATA-CENTRIC AI DEVELOPMENT MARKET, BY DATA TYPE**

- 6.1 Structured Data
- 6.2 Unstructured Data
  - 6.2.1 Text Data
  - 6.2.2 Image Data
  - 6.2.3 Video Data
- 6.3 Semi-Structured Data

## **7 GLOBAL DATA-CENTRIC AI DEVELOPMENT MARKET, BY DEPLOYMENT MODE**

- 7.1 On-Premises
- 7.2 Cloud-Based
- 7.3 Hybrid Deployment

## **8 GLOBAL DATA-CENTRIC AI DEVELOPMENT MARKET, BY DATA LIFECYCLE STAGE**

- 8.1 Data Collection
- 8.2 Data Cleaning & Preparation
- 8.3 Data Labeling & Annotation
- 8.4 Model Training & Optimization

## **9 GLOBAL DATA-CENTRIC AI DEVELOPMENT MARKET, BY APPLICATION**

- 9.1 Natural Language Processing
- 9.2 Computer Vision
- 9.3 Speech Recognition
- 9.4 Recommendation Systems
- 9.5 Fraud Detection

## **10 GLOBAL DATA-CENTRIC AI DEVELOPMENT MARKET, BY END USER**

- 10.1 Enterprises
- 10.2 AI Startups
- 10.3 Research Institutions

## **11 GLOBAL DATA-CENTRIC AI DEVELOPMENT MARKET, BY GEOGRAPHY**

- 11.1 North America
  - 11.1.1 United States
  - 11.1.2 Canada
  - 11.1.3 Mexico
- 11.2 Europe
  - 11.2.1 United Kingdom
  - 11.2.2 Germany
  - 11.2.3 France
  - 11.2.4 Italy
  - 11.2.5 Spain
  - 11.2.6 Netherlands
  - 11.2.7 Belgium
  - 11.2.8 Sweden
  - 11.2.9 Switzerland
  - 11.2.10 Poland
  - 11.2.11 Rest of Europe
- 11.3 Asia Pacific
  - 11.3.1 China
  - 11.3.2 Japan
  - 11.3.3 India
  - 11.3.4 South Korea
  - 11.3.5 Australia
  - 11.3.6 Indonesia

- 11.3.7 Thailand
- 11.3.8 Malaysia
- 11.3.9 Singapore
- 11.3.10 Vietnam
- 11.3.11 Rest of Asia Pacific
- 11.4 South America
  - 11.4.1 Brazil
  - 11.4.2 Argentina
  - 11.4.3 Colombia
  - 11.4.4 Chile
  - 11.4.5 Peru
  - 11.4.6 Rest of South America
- 11.5 Rest of the World (RoW)
  - 11.5.1 Middle East
    - 11.5.1.1 Saudi Arabia
    - 11.5.1.2 United Arab Emirates
    - 11.5.1.3 Qatar
    - 11.5.1.4 Israel
    - 11.5.1.5 Rest of Middle East
  - 11.5.2 Africa
    - 11.5.2.1 South Africa
    - 11.5.2.2 Egypt
    - 11.5.2.3 Morocco
    - 11.5.2.4 Rest of Africa

## **12 STRATEGIC MARKET INTELLIGENCE**

- 12.1 Industry Value Network and Supply Chain Assessment
- 12.2 White-Space and Opportunity Mapping
- 12.3 Product Evolution and Market Life Cycle Analysis
- 12.4 Channel, Distributor, and Go-to-Market Assessment

## **13 INDUSTRY DEVELOPMENTS AND STRATEGIC INITIATIVES**

- 13.1 Mergers and Acquisitions
- 13.2 Partnerships, Alliances, and Joint Ventures
- 13.3 New Product Launches and Certifications
- 13.4 Capacity Expansion and Investments
- 13.5 Other Strategic Initiatives

## 14 COMPANY PROFILES

- 14.1 Google LLC
- 14.2 Microsoft Corporation
- 14.3 Amazon Web Services Inc.
- 14.4 IBM Corporation
- 14.5 Snowflake Inc.
- 14.6 Databricks Inc.
- 14.7 Scale AI Inc.
- 14.8 Appen Limited
- 14.9 Samasource Inc.
- 14.10 Alteryx Inc.
- 14.11 DataRobot Inc.
- 14.12 H2O.ai Inc.
- 14.13 Oracle Corporation
- 14.14 SAP SE
- 14.15 Cloudera Inc.
- 14.16 Teradata Corporation
- 14.17 C3.ai Inc.

## List Of Tables

### LIST OF TABLES

Table 1 Global Data-Centric AI Development Market Outlook, By Region (2023-2034) (\$MN)

Table 2 Global Data-Centric AI Development Market Outlook, By Component (2023-2034) (\$MN)

Table 3 Global Data-Centric AI Development Market Outlook, By Tools & Platforms (2023-2034) (\$MN)

Table 4 Global Data-Centric AI Development Market Outlook, By Data Labeling Tools (2023-2034) (\$MN)

Table 5 Global Data-Centric AI Development Market Outlook, By Data Versioning Platforms (2023-2034) (\$MN)

Table 6 Global Data-Centric AI Development Market Outlook, By Data Quality Management Tools (2023-2034) (\$MN)

Table 7 Global Data-Centric AI Development Market Outlook, By Services (2023-2034) (\$MN)

Table 8 Global Data-Centric AI Development Market Outlook, By Data Annotation Services (2023-2034) (\$MN)

Table 9 Global Data-Centric AI Development Market Outlook, By AI Consulting Services (2023-2034) (\$MN)

Table 10 Global Data-Centric AI Development Market Outlook, By Data Engineering Services (2023-2034) (\$MN)

Table 11 Global Data-Centric AI Development Market Outlook, By Data Type (2023-2034) (\$MN)

Table 12 Global Data-Centric AI Development Market Outlook, By Structured Data (2023-2034) (\$MN)

Table 13 Global Data-Centric AI Development Market Outlook, By Unstructured Data (2023-2034) (\$MN)

Table 14 Global Data-Centric AI Development Market Outlook, By Text Data (2023-2034) (\$MN)

Table 15 Global Data-Centric AI Development Market Outlook, By Image Data (2023-2034) (\$MN)

Table 16 Global Data-Centric AI Development Market Outlook, By Video Data (2023-2034) (\$MN)

Table 17 Global Data-Centric AI Development Market Outlook, By Semi-Structured Data (2023-2034) (\$MN)

Table 18 Global Data-Centric AI Development Market Outlook, By Deployment Mode

(2023-2034) (\$MN)

Table 19 Global Data-Centric AI Development Market Outlook, By On-Premises

(2023-2034) (\$MN)

Table 20 Global Data-Centric AI Development Market Outlook, By Cloud-Based

(2023-2034) (\$MN)

Table 21 Global Data-Centric AI Development Market Outlook, By Hybrid Deployment

(2023-2034) (\$MN)

Table 22 Global Data-Centric AI Development Market Outlook, By Data Lifecycle Stage

(2023-2034) (\$MN)

Table 23 Global Data-Centric AI Development Market Outlook, By Data Collection

(2023-2034) (\$MN)

Table 24 Global Data-Centric AI Development Market Outlook, By Data Cleaning & Preparation (2023-2034) (\$MN)

Table 25 Global Data-Centric AI Development Market Outlook, By Data Labeling & Annotation (2023-2034) (\$MN)

Table 26 Global Data-Centric AI Development Market Outlook, By Model Training & Optimization (2023-2034) (\$MN)

Table 27 Global Data-Centric AI Development Market Outlook, By Application (2023-2034) (\$MN)

Table 28 Global Data-Centric AI Development Market Outlook, By Natural Language Processing (2023-2034) (\$MN)

Table 29 Global Data-Centric AI Development Market Outlook, By Computer Vision (2023-2034) (\$MN)

Table 30 Global Data-Centric AI Development Market Outlook, By Speech Recognition (2023-2034) (\$MN)

Table 31 Global Data-Centric AI Development Market Outlook, By Recommendation Systems (2023-2034) (\$MN)

Table 32 Global Data-Centric AI Development Market Outlook, By Fraud Detection (2023-2034) (\$MN)

Table 33 Global Data-Centric AI Development Market Outlook, By End User (2023-2034) (\$MN)

Table 34 Global Data-Centric AI Development Market Outlook, By Enterprises (2023-2034) (\$MN)

Table 35 Global Data-Centric AI Development Market Outlook, By AI Startups (2023-2034) (\$MN)

Table 36 Global Data-Centric AI Development Market Outlook, By Research Institutions (2023-2034) (\$MN)

Note: Tables for North America, Europe, APAC, South America, and Rest of the World (RoW) Regions are also represented in the same manner as above.

## I would like to order

Product name: Data-Centric AI Development Market Forecasts to 2034 – Global Analysis By Component (Tools & Platforms and Services), Data Type, Deployment Mode, Data Lifecycle Stage, Application, End User and By Geography

Product link: <https://marketpublishers.com/r/DE439A410BBBEN.html>

Price: US\$ 4,150.00 (Single User License / Electronic Delivery)

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

[info@marketpublishers.com](mailto:info@marketpublishers.com)

## Payment

To pay by Credit Card (Visa, MasterCard, American Express, PayPal), please, click button on product page <https://marketpublishers.com/r/DE439A410BBBEN.html>