

DATA ANNOTATION

AS A STRATEGIC ASSET IN THE **BANKING INDUSTRY**



Author

Paresh Ashara
Vice President, Data Analytics

QUINTE FINANCIAL TECHNOLOGIES

Contents

Introduction	3
Forecasting Comprehensive Market Trends	4
Why you need Data Annotation	5 - 6
Types of Data Annotation	7
Techniques in Data Annotation	8

Introduction

Financial institutions (FIs) have been the early adopters of technology. The use of technology has enabled them to streamline data collection processes by reducing manual effort and improving efficiencies. The FIs are sitting on a treasure trove of data collected from their customers/members. By leveraging new age technologies in artificial intelligence (AI) – deep learning and machine learning (ML), and now generative AI(GenAI), FIs can use their in-house data to their advantage to cater to various aspects of operations that could improve both their employees' and customers'/members' experience.

Data annotation refers to the systematic process of adding meaningful tags or labels to various types of data—such as images (photos, scanned documents), text (standard documents, transcripts, speech-to-text), audio (customer/member service related), and video (property mortgage underwriting) —making it easier for ML algorithms to interpret and analyze the information. For instance, in computer vision applications like self-driving cars, data annotation helps distinguish between objects (e.g., pedestrians and vehicles) by providing labeled examples from which the algorithms can learn.

ML and deep learning algorithms need data points to be called out meaningfully to understand the context. FIs have humongous data gathered over their lifespan. With the buzz on AI and large language models (LLMs) and now with the focus shifting to small language models (SLMs) or company function-specific language models (FLMs), FIs need to annotate their data to create their SLMs or FLMs. The annotated data will be a reliable source of company-specific information that could result in better efficiency and productivity of their resources.

Forecasting a Comprehensive Look at Market Trends and Regional Insights

The data annotation market is experiencing significant growth, driven by the increasing demand for ML and AI applications across various industries. Here are the key insights regarding its size and projected growth:

Current Market Size



The global Data Annotation Tools Market was valued at approximately USD 1.27 billion in 2023 and is expected to reach around USD 7.17 billion by 2031, growing at a CAGR of 24.55% during the forecast period from 2024 to 2031



Another report estimated that the market size was at USD 3.9 billion in 2023, with projections to grow to USD 15.2 billion by 2032, reflecting a CAGR of 6.5%



A more aggressive forecast predicts the data annotation and labeling market will achieve a valuation of USD 2.07 billion by the end of 2024, escalating to USD 29.58 billion by 2033, indicating a remarkable CAGR of 34.4% regional insights



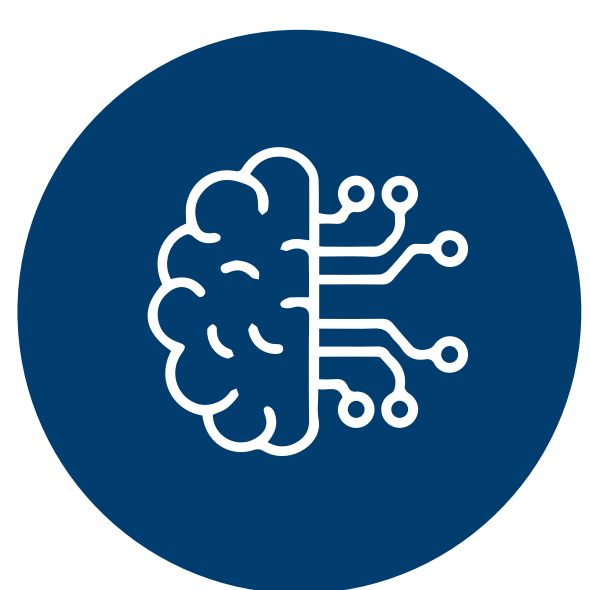
North America held a significant share of the data annotation market, accounting for around 36.08% in 2023, which was attributed to its strong technological infrastructure and investment in AI technologies

Why You Need Data Annotation



Managing Unstructured Data

A significant portion of data in FIs is unstructured (e.g., scanned images, documents, emails, videos, etc.). Data annotation transforms this unstructured data into a format that AI models can effectively utilize.



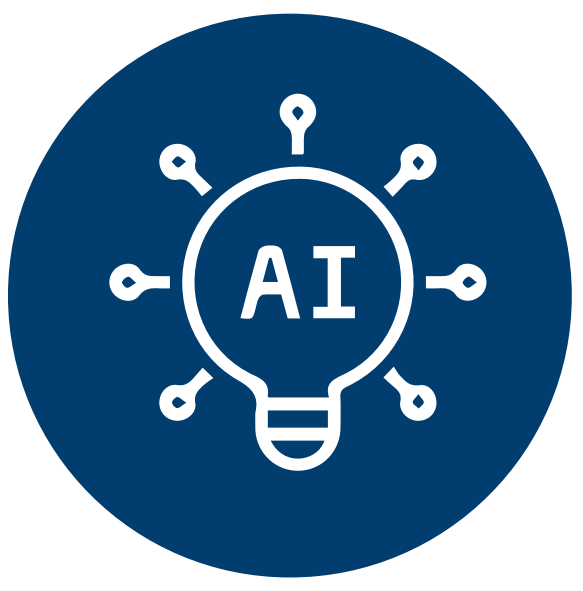
Training Machine Learning Models

Annotated datasets are essential for supervised learning, where algorithms learn from labeled examples. These enable models to recognize patterns and make accurate predictions. For example, building an AI fraud detection model or an AI Credit scoring model. By tagging and classifying various regulatory texts and compliance requirements, AI models can accurately recognize and interpret complex regulations, helping institutions stay aligned with legal standards and avoid potential violations.



Enhancing Model Performance

High-quality annotated data improves the accuracy and efficiency of AI models, allowing them to learn faster and deliver more reliable results across various applications. For example, if the AI fraud detection model has been trained on company-specific annotated data, the model's accuracy can be far superior to a traditional model. For that matter, in SLM or FLM, there are no hallucinations when answering company-specific questions.



Generative AI and RAG-based Applications

Annotated documents and images used in GenAI use cases leveraging Retrieval-Augmented Generation (RAG) approaches can immensely improve the effectiveness of GenAI applications. By labeling data with relevant intents and entities, models can retrieve the most pertinent information from knowledge bases and generate more accurate responses. This ensures that the AI can differentiate between similar queries and respond appropriately based on context.



Enhancing User Trust and Eliminating Hallucinations

AI hallucinations occur when a generative model perceives patterns or produces output that does not exist. For instance, a language model might create plausible sounding but false information, such as invented product features or non-existent services. These outputs can mislead users and undermine trust in AI systems. A well-annotated data can provide more reliable and personalized interactions. Users' confidence in the system increases When they receive accurate responses based on their specific queries. This is vital for applications like customer/member support chatbots, where user confidence is essential for engagement.



Types of

Data Annotation

Data annotation encompasses various techniques tailored to specific data types.



Text Annotation

Includes tasks such as customer/member sentiment analysis, where text is labeled based on the emotions expressed or named entity recognition, which identifies specific entities within the text.

Annotating news articles, social media posts, and market reports enables real-time sentiment analysis, helping FIs uncover market trends and predict future volatility to inform investment strategies.



Image Annotation

Involves tagging images with labels or bounding boxes to identify objects within them. It aids in automating the extraction of critical information from various financial documents, such as scanned invoices and contracts, streamlining operations and reducing manual workload.



Audio Annotation

Involves transcribing speech or labeling audio clips based on the contact center channel. The customer/member interaction can be evaluated for their needs and satisfaction. This annotation type is vital for developing voice recognition systems and real-time agentic support leveraging GenAI.



Video Annotation

Like image annotation but applied to video content, where frames are labeled to track objects over time. The application could be in mortgage and commercial real estate financing for automated market value estimation and determining insurance needs.

Predictive AI models can be built on such annotated video data to derive business benefits.

Techniques in Data Annotation

Manual Annotation

Human annotators label data directly. While this method ensures high quality, it can be time-consuming and labor-intensive.

Automatic Annotation

Utilizes pre-trained models to annotate large datasets quickly. Although faster, the quality depends on the underlying model's accuracy.

Active Learning

The model identifies which data points would be most beneficial for learning and requests annotations for those specific points, optimizing the annotation process.

Data annotation is integral to transforming raw financial data into actionable insights, enhancing operational efficiency, ensuring compliance, and improving customer/member experiences in the banking industry.

About Author

Paresh Ashara is a Vice President at Quinte Financial Technologies, overseeing Data Analytics, with over 25 years of IT services and product engineering experience in BFSI industry. He can be reached at paresh.ashara@quinteft.com

Sources

<https://www.kingsresearch.com/data-annotation-tools-market-259>

<https://www.custommarketinsights.com/press-releases/data-annotation-tool-market-size/>

<https://finance.yahoo.com/news/data-annotation-labelling-market-expected-144500834.html?guccounter=1>

<https://aws.amazon.com/marketplace/pp/prodview-z7v6bdzb2z4vo>

<https://www.marketsandmarkets.com/Market-Reports/data-annotation-and-labelling-market-20349022.html>

<https://labelyourdata.com/articles/trends-in-data-annotation-market-forecast>

<https://www.grandviewresearch.com/industry-analysis/data-annotation-tools-market>

<https://www.gminsights.com/industry-analysis/data-annotation-tools-market>