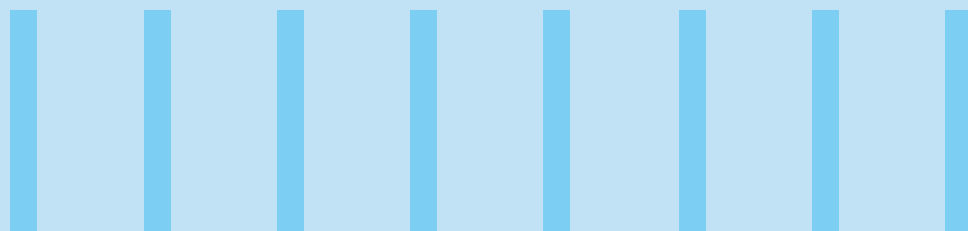# DATA ANNOTATION SERVICES: WHAT DRIVES THE CUSTOMERS' INTEREST?

Abstract

Organisations are unable to maintain speed at which available raw data is growing exponentially. AI-powered projects can maximise their potential and offer better performance and higher accuracy rates with the help of data annotation. This POV discusses how data annotation is useful for the functioning of AI and ML models across organisations.

Infosys®
Navigate your next

What is data annotation?

Data annotation, also known as digital annotation, is the process of categorising and labelling data in various formats such as text, image, audio, or video for machine learning (ML) and artificial intelligence (AI) applications. With high-quality, human-annotated data, organisations can build and improve AI algorithms with greater quality, speed, and accuracy. This would result in improved customer experience and higher engagement with the brand.[*]

The global adoption of ML and AI technologies in organisations has fuelled innovation in data annotation tools and services. According to a study, a majority of customer interactions will filter through technologies such as ML, virtual assistants or chatbots, and mobile messaging. ML and AI algorithms thrive on data; the more they feed on data, the more accurate their outcomes are. While developing an AI-based project, organisations spend about 80% of their time preparing the data. This is a crucial task because the success of algorithms relies heavily on the right kind of data, and any error in data can translate into inaccurate results.

An AI model gains more and more intelligence as we provide it better training data. The key to an efficient and effective AI model is the identification of patterns in available data. However, for an AI model to be as accurate in identifying these patterns as humans, it requires human involvement and data annotation solutions to understand a given context properly. Only then can we trust that the end products are as accurate and representative as possible.[1][2]

Types of data annotation

Data annotation helps AI or ML models function effectively. The more dynamic and comprehensive we want the model to be, the more data it needs. This implies the need for data in various formats too. Various data types aid in connecting different environments and understanding them as much as possible.[1]

1. **Text annotation:** Text annotation is a fundamental annotation type. Over 70% of companies use text annotation in various forms, such as sentiment, intent, query, semantic, and named entity annotation.[1]

a. **Sentiment annotation:** Sentiment analysis can gauge attitudes, emotions, and opinions in a text. It provides insights into whether a thought or a sentiment is positive, negative, or neutral. Sentiment analysis is an extremely powerful tool in marketing, customer service, and social media because it can help assess the general public's opinion or response to a certain product or a piece of content. Sentiment annotation involves human annotators as they can best recognise the nuances of opinions, conventions, and emotions.

b. **Intent annotation:** It is crucial for machines to understand the user's intent in communicating and performing intelligent tasks. Intent annotation analyses the need behind texts and categorises them as requests, recommendations, commands, bookings, confirmations, or approvals. Slangs, jargons, and colloquialisms are terms best understood by humans through continued usage. Intent annotation focuses on these natural language factors to assess the intent of the user.

c. **Named entity annotation:** To accurately identify and label named entities, AI uses named entity recognition (NER). In the case of unstructured and raw data, NER can prove helpful in categorising text and targeting social media advertisements. Annotators can understand the context of a text and help better the quality of the NER models for NLP algorithms. Named entity annotation is manual. The algorithm requires large quantities of manually annotated training data. Relevant information such as names and entities of individuals, companies, locations, times, and organisations is extracted from text documents and analysed.

d. **Semantic annotation:** Semantic annotation is the process of tagging text documents with relevant metadata to create lists and other text that is manageable. It is best suited for search engine result optimisation, training chatbots, and applications where the context of a title or a query impacts the relevance of the labelled entity. Semantic annotation works by interpreting unstructured content and enriching the content with information that links the content to the extracted concepts.

2. **Audio annotation:** Audio annotation is the transcription and time stamping of audio data. It analyses the specific pronunciation, inflexion, and intonation, as well as identifies language, dialect, and speaker demographics. This has applications in various fields, such as safety hotlines, which can detect any dangers based on the sounds in the background or from the caller's tone.

3. **Image annotation:** Image annotation helps an ML model process and interpret images by categorising image data and labelling them. Image annotation has several applications, including computer vision, robotic vision, facial recognition, and more. Annotators use identifiers, captions, or keywords to help the models identify the image parameters to train AI-based algorithms.

4. **Video annotation:** Video annotation is vital for applications such as self-driving cars that need to precisely identify and outline moving objects such as vehicles, traffic signs, and pedestrians. Bounding boxes and human-annotated data can help the AI models learn the movements and patterns of moving objects. Video annotation can also help implement motion blur and object tracking in the models.

What are the benefits of data annotation?

Considering the volume of data that ML models consume, data annotation needs to deliver a high-functioning and feasible process. Here is what data annotation offers:[3]

- **Improved accuracy:** Data annotation can significantly improve the accuracy of an AI algorithm's output by training the algorithm with big data sets. As the algorithm gains more and more intelligent, it can also learn to identify various parameters that will further help it extract only the relevant information from the huge amounts of data.

- **Higher quality and faster outputs:** With ML- and AI-based models, all the automated processes can yield faster and highly accurate outputs, thanks to training data and annotation. The quality of the results is higher with every new data set, and the algorithm's overall success rate significantly increases.

- **Enhanced customer experience:** ML- and AI-powered models, such as chatbots and virtual assistants, can help in delivering a more user-centric experience by analysing the user's queries and needs and responding with the most relevant information. This process, when automated, can offer almost immediate solutions to customers, resulting in a satisfactory experience.

Where can data annotation be used?

With the advantages that data annotation clearly entails, one must also know how best we could implement it. We can expect data annotation implementation in sensitive and globally important industries such as:[4]

- **Healthcare:** From diagnostic support to drug development and patient engagement, AI in healthcare can improve with the help of data annotation. Image annotation can increase the efficiency and quality of data and lower the costs of diagnostic tests. Image annotation can be useful in MRI, X-ray, and CT scans to reduce the chances of human error and improve their speed and accuracy. Similarly, thermal image annotation can help diagnose diseases such as breast cancer in earlier stages with higher accuracy. NLP and data annotation can also contribute to drug development by interpreting huge data sets such as research papers, clinical trials, and records. We can use the data to analyse relationships between drugs and various factors such as genes, symptoms, and diseases.

- **Finance:** Financial institutions such as banks, credit bureaus, and payment processors can use data annotation to enhance customer experiences by analysing market fluctuations and transactions and delivering custom advice and reports to the customers.

- **Government and law:** Data annotation, including NER, semantic annotation, sentiment analysis, and audio annotation, can help in providing a secure and efficient solution to sensitive data for government and military applications. Image and video annotation can be used for reliable localisation of target objects. Annotated data can also enrich legal documents and interpret volumes of law research and document discoveries.

- **Autonomous vehicles:** Bounding boxes and video annotation can aid in more efficient self-driving cars and other autonomous vehicles such as drones. The automobile sector could integrate data annotation across computer vision and NLP.

- **Commerce:** Sentiment annotation and intent annotation can help enhance user experience by identifying customer needs and curating products and services relevant to the customers.

- **Agriculture:** Annotated data can study crop health, pest and weed infestation, and soil quality. It can also predict the yield based on weather conditions. Annotating computer vision models can help in crop monitoring and automating regular tasks such as harvesting, picking, seeding, and spraying.

- **Social media:** NLP integration with data annotation can improve social media marketing and increase user conversions. Using sentiment analysis, brands can understand their audiences' response to a new product or a service.

What does data annotation offer?

Consumers in the ML sector are looking for exponential improvement in the existing processes. While AI and ML implementations are still evolving, customers' expectations from data annotation engines will only rise. In such a scenario, we could consider the following significant factors in data annotation:[5]

- **High-quality training data:** The quality of the AI-based models is directly related to the quality and the quantity of the training data used. If the training data isn't reliable and well-processed, the model cannot develop to its full potential. Labelled and qualified training data curated for specific applications is a great starting point for any ML or AI algorithm implementation.

- **Flexibility and scalability:** While the purpose of all annotated data is implementation without human intervention, it must adapt to a change in the data set or intent. Labelled data is sometimes ambiguous and fluctuating. The ML model needs to operate in changing environments and should maintain its performance while discovering new possibilities. So, flexibility is essential for credibility. Similarly, the annotated data should also be scalable through varying workflows.

- **Speed:** ML and AI algorithms that best serve their purpose should be able to replace human effort. A data set that makes the algorithm more efficient and faster piques customer interest. Using pre-labelled data, ML- and AI-powered solutions can improve performance, speed, and capacity without compromising on accuracy.

Conclusion

A large chunk of online communications and networking is moving towards AI and ML. The current volume of customer data is beyond organisations' consumption capabilities. At this rate, we can expect AI to contribute significantly to the tune of $13 trillion by 2030. Data requirements for such growth will be astronomical and efficient annotation services will be in high demanded. Scalability, industry-specific functionality, and cost-efficiency will drive customer interest.[5]

*For organisations on the digital transformation journey, agility is key in responding to a rapidly changing technology and business landscape. Now more than ever, it is crucial to deliver and exceed organisational expectations with a robust digital mindset backed by innovation. Enabling businesses to sense, learn, respond, and evolve like a living organism, will be imperative for business excellence going forward. A comprehensive yet modular suite of services is doing exactly that. Equipping organisations with intuitive decision-making automatically at scale, actionable insights based on real-time solutions, anytime/anywhere experience, and in-depth data visibility across functions leading to hyper-productivity, Live Enterprise is building connected organisations that are innovating collaboratively for the future.

References

1. *https://appen.com/blog/data-annotation/*

2. *https://www.techslang.com/definition/what-is-data-annotation/*

3. *https://customerthink.com/data-annotation-types-tools-benefits-and-applications-in-machine-learning/*

4. *https://imerit.net/data-annotation/*

5. *https://medium.com/nerd-for-tech/data-annotation-services-a-potential-industry-behind-ai-be5611e0a375*

For more information, contact infosysbpm@infosys.com

Infosys
Navigate your next

Infosysbpm.com

Stay Connected