



## Glossary of Key Terms

- **3D Annotation:** The process of labeling specific points or areas on a 3D model, essential for quality control and assembly of complex geometries in manufacturing.
- **Active Learning:** A machine learning approach that prioritizes labeling the most informative or ambiguous data samples to improve model performance efficiently.
- **AI (Artificial Intelligence):** The simulation of human intelligence processes by machines, especially computer systems.
- **Annotation Guidelines:** Established rules and instructions for consistent data labeling across a project.
- **Annotation Pipeline:** The sequence of steps involved in labeling data, including preprocessing, feature engineering, and quality assurance.
- **Anomaly Detection:** The process of identifying data points, events, or observations that deviate significantly from the expected pattern in a dataset.
- **Bounding Box:** A rectangular box drawn around an object in an image or video frame to identify its location and size.
- **Convolutional Neural Networks (CNNs):** A type of deep learning model particularly effective for image-based tasks, such as visual inspection and defect detection.
- **Contextual Analysis:** The process of examining data within its broader context to extract more meaningful insights, especially in manufacturing environments.
- **Cross-validation:** A technique used to assess model performance and prevent overfitting by testing the model on multiple subsets of the data.
- **Data Annotation:** The process of labeling data to make it usable for machine learning models.
- **Data Augmentation:** The process of creating synthetic data or modifying existing data to increase dataset diversity and improve model robustness.
- **Data Labeling:** Another term for data annotation, involving adding meaningful tags or information to raw data.



- **Data Mining:** The process of discovering patterns in large datasets using methods at the intersection of machine learning, statistics, and database systems.
- **Data Preprocessing:** The initial stage of data preparation, involving cleaning, normalization, and formatting of raw data for annotation or model training.
- **Data Scientist:** A professional who uses data to understand and solve complex problems, often developing machine learning models.
- **Domain Expert:** A professional with specialized knowledge in a specific field, crucial for providing context and insights in value-added data annotation.
- **Enrichment:** The process of adding value to data by incorporating additional context or information.
- **Feature Engineering:** The process of creating relevant features from raw data to improve machine learning model performance.
- **Ground Truth Generation:** The process of creating accurate, verified data labels that serve as a reference for training and evaluating machine learning models.
- **Hierarchical Labeling:** An annotation technique that organizes information in a structured, tree-like manner, enabling detailed classification and easy navigation of data.
- **Image Annotation:** The process of labeling images with relevant information, such as object classifications or locations.
- **Isolation Forest:** A machine learning algorithm used for anomaly detection, particularly effective in identifying outliers in datasets.
- **K-Means Clustering:** An unsupervised learning algorithm used for grouping similar data points, useful in identifying natural patterns in manufacturing processes.
- **Kalman Filter:** A mathematical algorithm used for estimating unknown variables in time-series data, often applied in sensor data processing.
- **Key Points:** Specific points marked on an image to identify important features or landmarks.
- **Long Short-Term Memory (LSTM) Networks:** A type of recurrent neural network architecture designed to capture long-term dependencies in sequential data.



- **Machine Learning (ML):** A subset of AI that focuses on the development of algorithms and statistical models that enable computer systems to improve their performance on a specific task through experience.
- **Machine Learning Pipeline:** The end-to-end process of building a machine learning model, from data collection to deployment and monitoring.
- **Metadata:** Additional information about data that provides context or description.
- **ML Engineer:** A professional who focuses on taking theoretical data science models and helping scale them out to production-level models that can handle terabytes of real-time data.
- **Model Drift:** The degradation of model performance over time due to changes in the underlying data distribution or real-world conditions.
- **Multi-label Annotation:** An annotation technique that involves assigning multiple labels to a single data point, useful for complex products with multiple attributes.
- **Overfitting:** A situation where a machine learning model performs well on training data but poorly on unseen data, often due to excessive complexity.
- **Performance Metrics:** Measures used to evaluate the effectiveness of machine learning models, such as accuracy, precision, recall, and F1-score.
- **Point Cloud:** A set of data points in three-dimensional space, often used to represent the external surface of an object.
- **Polygon Segmentation:** An annotation technique that involves outlining objects or areas with a series of connected points, allowing for precise shape definition.
- **Predictive Maintenance:** The practice of using data analysis techniques to identify potential equipment failures before they occur, allowing for proactive maintenance scheduling.
- **Process Optimization:** The use of data-driven techniques to improve manufacturing processes, increase efficiency, and reduce waste.
- **Quality Assurance (QA):** The process of verifying the accuracy and consistency of annotated data.
- **Quality Control:** The process of ensuring that products meet specified quality standards, often involving data annotation for defect detection and classification.
- **Random Forest:** An ensemble learning method that constructs multiple decision trees and merges them to get a more accurate and stable prediction.



- **Reinforcement Learning:** A type of machine learning where an agent learns to make decisions by interacting with an environment and receiving feedback.
- **Root Cause Analysis:** A problem-solving method used to identify the underlying causes of faults or problems in manufacturing processes.
- **Scalability:** The capability of a system or process to handle a growing amount of work or its potential to be enlarged to accommodate that growth.
- **Semantic Labeling:** The process of assigning meaningful labels or categories to data points, often used in image and video annotation.
- **Semantic Segmentation:** A deep learning algorithm that associates a label or category with every pixel in an image.
- **Supervised Learning:** A machine learning approach where models are trained on labeled data to predict outcomes for new inputs.
- **Support Vector Machines (SVM):** A supervised learning algorithm used for classification and regression tasks, particularly effective in high-dimensional spaces.
- **Tagging:** The process of attaching keywords or terms to a piece of data to help describe or categorize it.
- **Temporal Annotation:** The process of marking changes or events over time in sequential data, crucial for analyzing processes that change over time.
- **Text Annotation:** The process of labeling or categorizing text data for use in natural language processing tasks.
- **Time-Series Data:** Sequential data points collected over time, often used in manufacturing for sensor readings, production metrics, and quality control measurements.
- **Unsupervised Learning:** A machine learning approach that identifies patterns in unlabeled data, useful for discovering hidden structures in datasets.
- **Value-Added Annotation:** An advanced form of data annotation that goes beyond basic labeling by incorporating domain expertise and contextual information to improve AI model performance and insights.
- **Video Annotation:** The process of labeling objects, events, or areas in video footage, often used for analyzing manufacturing processes and improving efficiency.

