THE ART OF DATA ANNOATION

OPTIMIZING MACHINE LEARNING ACCURACY.

About

Carscan is an automobile technology platform that specialises in offering services for car maintenance, diagnostics, and inspection of vehicles using AI.

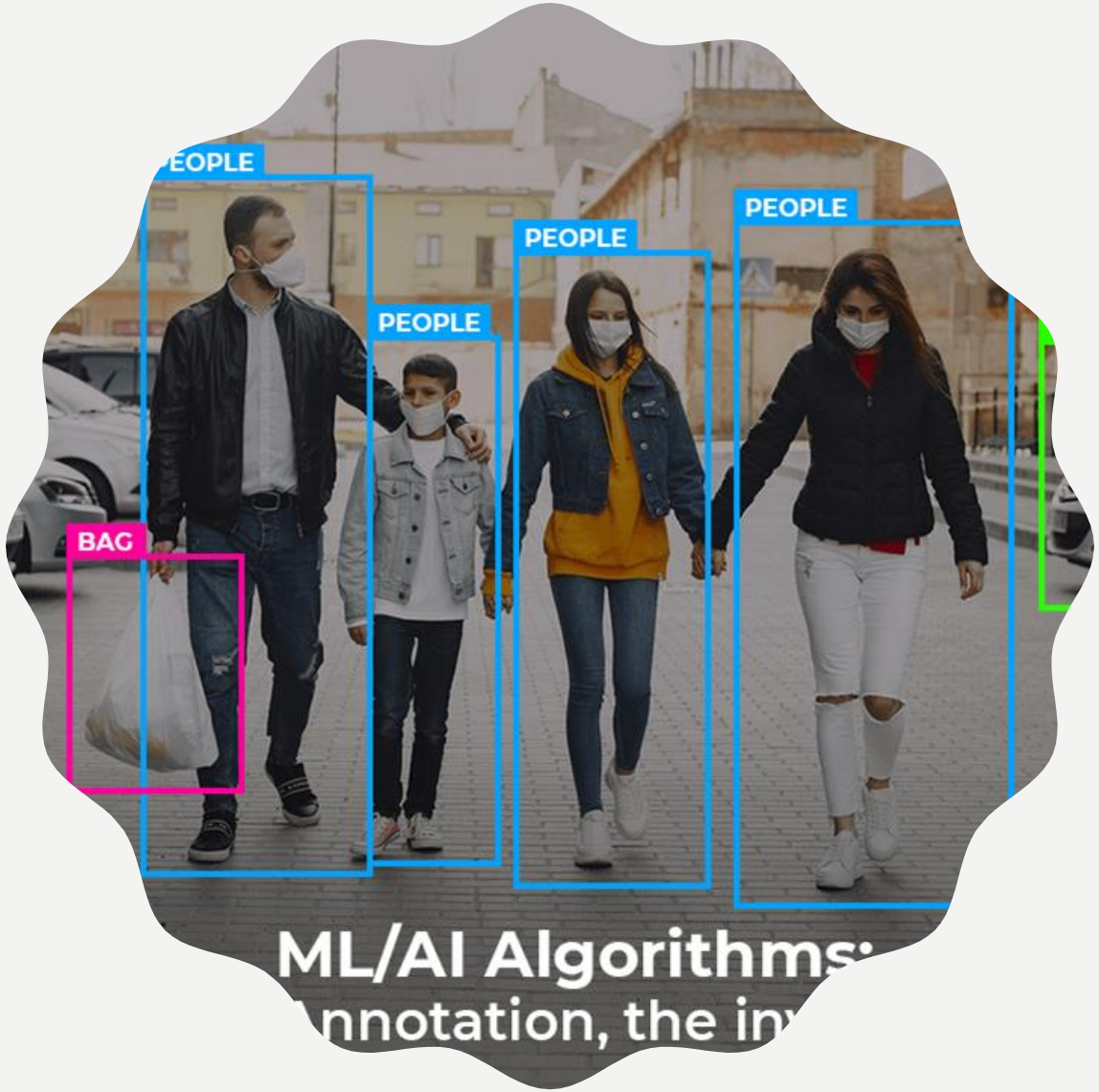
In order to deliver thorough inspections and diagnostics, it analyses photos and videos of automobiles using computer vision algorithms and machine learning models.

Core competencies of CARSCAN

- Proprietary AI platform with a database of 2M+ images to scan damages and identify anomalies with precision.
- State of art technology to scan the key internal vehicle details of odometer, engine , gearbox condition with the power of AI.
- Identification of defects from any smart phone through self assessment pattern analysis
- Records vehicle identities and identify fraud through VIN license matching and geo tag mapping.
- Provides real-time vehicle health and obtains real-time competitive quotes directly through your app from local repairs, workshops and service centres.

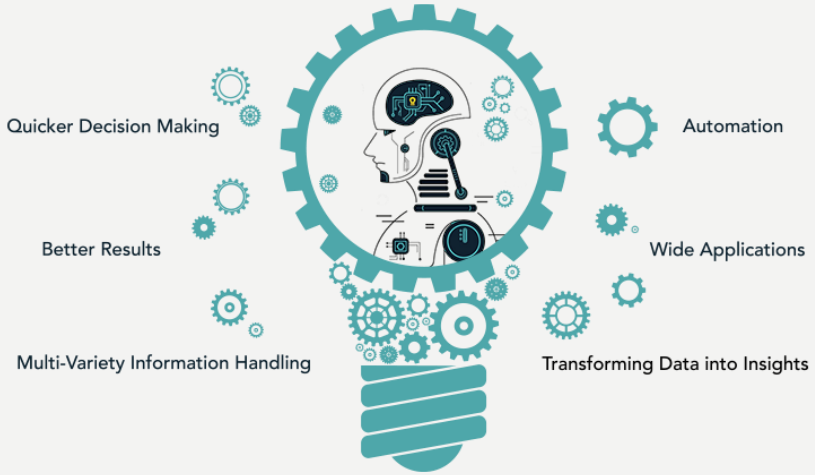
WHAT IS DATA ANNOTATION?

- Data annotation is the categorization and labelling of data for AI applications.
- With high-quality, human-powered data annotation, companies can build and improve AI implementations.
- The result is an enhanced customer experience solution such as product recommendations, relevant search engine results, computer vision, speech recognition, chatbots, and more.



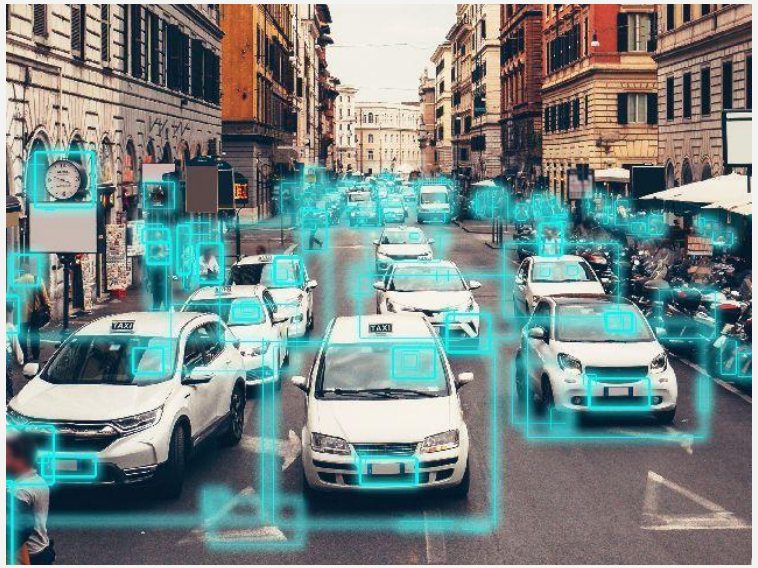
WHY IS IT IMPORTANT?

- Businesses are adopting AI technology to **automate decision-making** and benefit from new business opportunities
- Machine learning models have a wide variety of critical applications (e.g., healthcare) where **erroneous AI/ML models can be dangerous**
- Finding high-quality annotated data is one of the primary challenges of **building accurate machine-learning models**



HOW EASY OR DIFFICULT IS IT?

- Data annotation is especially important when considering the amount of unstructured data that exists in the form of text, images, video, and audio. By most estimates, unstructured data accounts for **80% of all data generated**.



TYPES OF DATA ANNOTATIONS

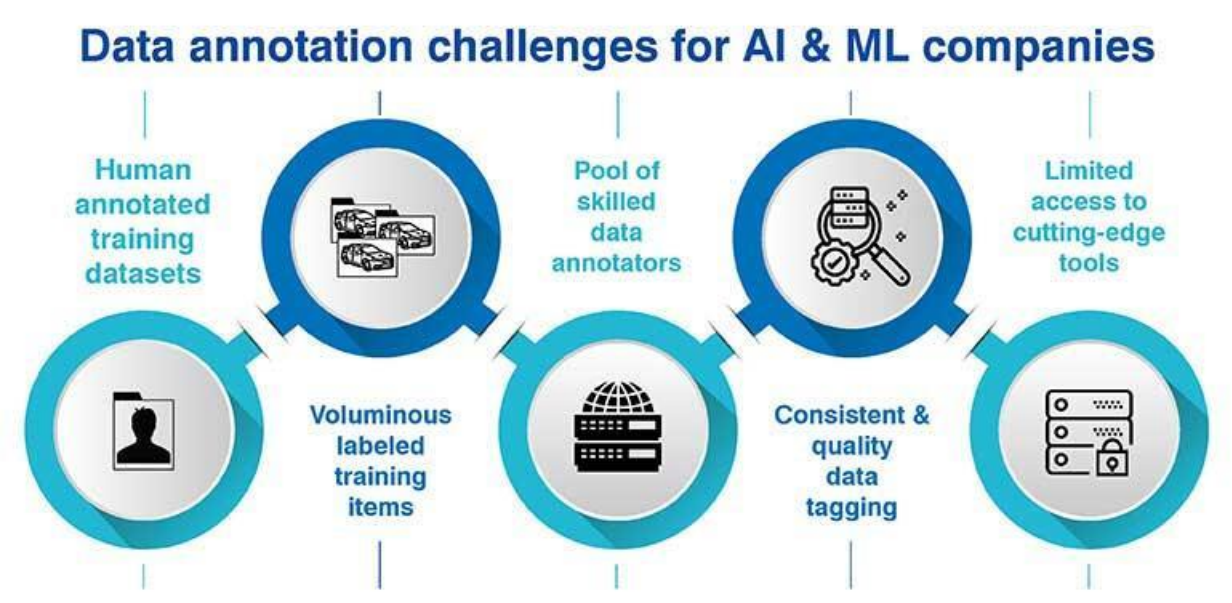
- **Medical data annotation:** Medical data annotation is used to annotate data such as medical images (MRI scans), and clinical notes, etc. This type of data annotation helps develop computer vision-enabled systems for disease diagnosis and automated medical data analysis.

- **Retail data annotation:** Retail data annotation is used to annotate retail data such as product images, customer data, and sentiment data.
- **Automotive data annotation:** This industry-specific annotation is used to annotate data from autonomous vehicles, such as data from cameras and lidar sensors. This annotation type helps develop models that can detect objects in the environment and other data points for autonomous vehicle systems.

- **Industrial data annotation:** Industrial data annotation is used to annotate data from industrial applications, such as manufacturing images, maintenance data, safety data, quality control, etc

USE OF DATA ANNOTATION

- For example, in a supervised learning problem of email spam classification, the algorithm would be trained using a labelled dataset where each email is labelled as "spam" or "not spam." The algorithm would learn the characteristics of spam emails and use that knowledge to classify new, unseen emails as either spam or not spam.



CATEGORIES OF PROBLEMS

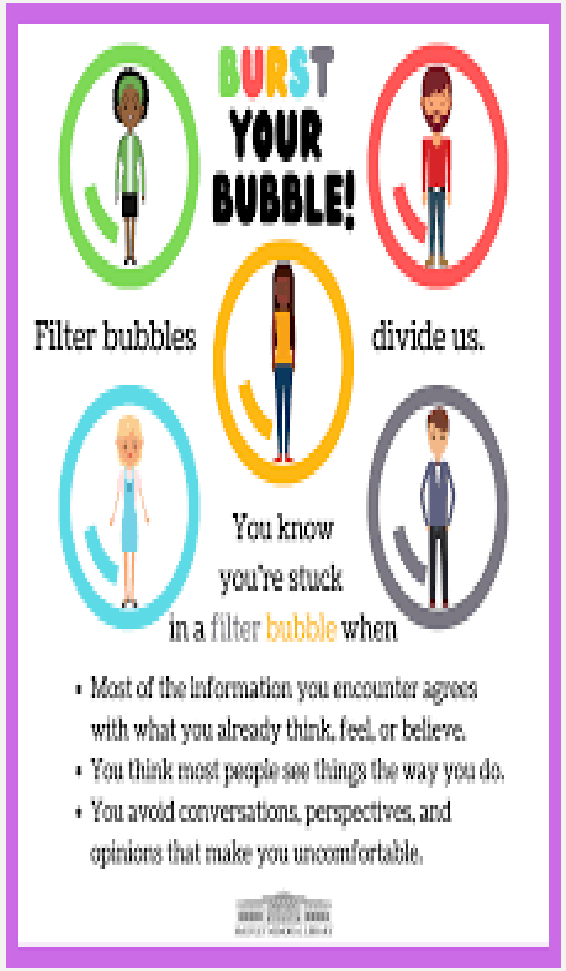
- The problem solved by supervised machine learning model can be viewed in two broad categories:

- **Classification:** Assigning test data into specific categories. For instance, predicting whether a patient has a disease and assigning their health data to “disease” or “no disease” categories is a classification problem.

- **Regression:** Establishing a relationship between dependent and independent variables. Estimating the relationship between the budget for advertising and the sales of a product is an example of a regression problem.

THREAT POSED BY AI

- **The bubble effect :**
- The emergence of social media has introduced algorithms that tailor content to match our interests. While this creates a personalized experience, it also creates a "bubble" of similarity or confirmation bias, reinforcing our existing beliefs and filtering out opposing viewpoints that challenge our perspectives
- As humans, we're highly susceptible to this bubble effect. And since AI relies on human input and intervention, there's always a risk that personal bias "bubbles" filter through to data, despite genuine attempts at objectivity.
- By understanding bias, the sources from which it originates, and the processes by which they infiltrate data, we can actively design systems to avoid, minimize, and – as much as is humanly and mechanically possible – eliminate them.



SOCIAL SURVEILLANCE WITH AI TECHNOLOGY

- In addition to its more existential threat, Ford is focused on the way AI will adversely affect privacy and security. A prime example is China's use of facial recognition technology in offices, schools and other venues. Besides tracking a person's movements, the Chinese government may be able to gather enough data to monitor a person's activities, relationships and political views.
- Another example is U.S. police departments embracing predictive policing algorithms to anticipate where crimes will occur. The problem is that these algorithms are influenced by arrest rates, which disproportionately impact Black communities. Police departments then double down on these communities, leading to over-policing and questions over whether self-proclaimed democracies can resist turning AI into an authoritarian weapon.
- "Authoritarian regimes use or are going to use it," Ford said. "The question is, How much does it invade Western countries, democracies, and what constraints do we put on it?"

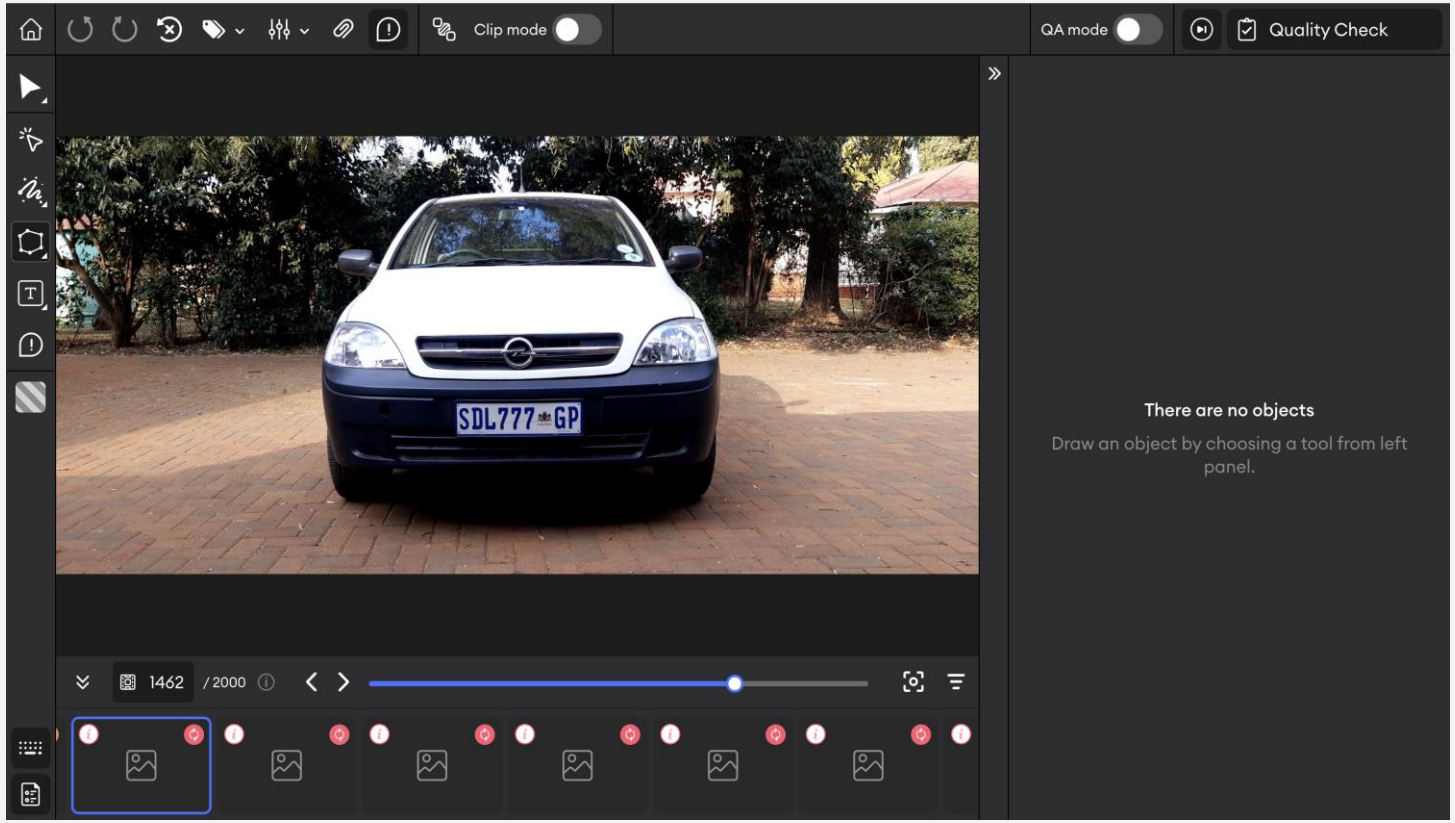
SUPER ANNOTATE TOOL

SuperAnnotate.com is a powerful cloud-based platform that revolutionizes the process of image annotation and dataset management for computer vision tasks. We will look at how this platform has revolutionized the industry.

USAGE AND INTERFACE

Image Annotation Tools:

- SuperAnnotate.com provides a wide range of annotation tools which include bounding boxes, polygons, points, lines, and more.
- The platform’s annotation tools are intuitive and easy to use, allowing annotators to efficiently label datasets without extensive training or expertise.



Using Machine Learning Techniques to reduce data annotation time

The review discusses different approaches that have been proposed to tackle this challenge

ACTIVE LEARNING

- Algorithms aim to intelligently select the most informative samples from a large dataset for annotation
- The review explores various active learning strategies, such as uncertainty sampling, query-by-committee, and information density-based approaches

Semi-Supervised Learning

- This approach leverages both labeled and unlabeled data to train models.
- The review discusses techniques such as self-training, co-training, and multi-view learning.

Transfer Learning

- Transfer learning involves utilizing pre-trained models on large-scale datasets and fine-tuning them on a target task with limited annotated data

- The literature review examines transfer learning Techniques, including domain adaptation and model adaptation

Weakly Supervised Learning

- This approach aims to train models with only partial or noisy annotations
- The review explores techniques such as multiple instance learning, co-training with noisy labels

THANK YOU