

# A Comprehensive Guide to Al Data Collection

## Have you ever wondered? How do devices as simple as a watch accurately predict heart

attacks in humans? How is it possible that cars and automobiles have always required a driver to suddenly go driverless on roads? How do chatbots make us believe that we are talking to another human on the other side?

If you observe the answer to every question, it boils down to just one

element - DATA. Data lies at the center of all AI-specific operations and processes. It is data that helps machines understand concepts, process inputs, and deliver accurate results.

Types of Data











# your ML Models?

Acquiring AI Training Data for















# effective Budget 03 02

Factors to consider for an

Scope

**Project** 

**Volume** 

Data

**Pricing** 

**Strategy** 

Sourcing Method

## Al experts or machine learning veterans, they might share any permutation of high-quality data is anything that is -

Well, like the very data we feed into our systems, quality has a lot of

factors and parameters associated with it as well. If you reach out to

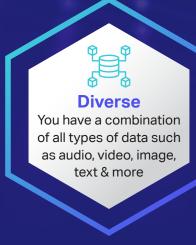
Data Quality

Uniform Consistent Comprehensive Uniformity in datasets Every single byte Data that covers all that are sourced from of data is similar in possible scenarios your

nature



multiple sources





system is intended to work on

## Vendors understand your market segment, use cases, demographics, to fetch you the right datasets. **Data Diversity** Vendors source datasets such as images, videos, text, audio files, or all of these from diverse sources across the globe.

Vendors structure and tag the data with attributes

and insights that machines understand.

Training Data service provider

Relevant Datasets

**Data Curation** 

### SME's from specific domains take care of annotating crucial pieces of information for greater accuracy.

objective results and inferences.

**Data Annotation** 

**Data Bias** 

Adhere to HIPAA, GDPR, and other industry compliances to eliminate legal complexities.

**Data De-identification** 

Eliminate bias from datasets, ensuring you have



Choosing the right Al Data

