Data Sources

- Text Corpora
- Image Datasets
- Audio Databases
- Video Datasets
- Simulated Data
- Sensor Data

File Formats

- XML / YML
- CSV
- JSON
- Custom format

Data Preparation

- Data Validation
- Segment
- Normalize
- Data Cleansing
- Dedupe
- Data Standardization
- Append

Setting up the project

- Understand Business Requirements and Annotation Specifications
- Develop Annotation Guidelines and SOPs
- Annotator Trainings
- Techqniue / Method Iteration Process

Annotation Process

- Pilot Annotation Project
- Initial Annotation Done Through Automation
- Manual Review by Human Annotators
- Consensus Annotation

Quality Checks & Validation

- Annotation Guidelines Review
- Inter-Annotator Agreement (IAA)
- Spot Checks and Audits
- Error Tracking and Correction
- Automated Validation Tools
- Feedback Loops
- Final Review
- Metrics for Continuous Improvement

If OK-> Done

Training Datasets Fed into ML Models and Assess the result

OR

If not found OK-> run the loop again

Training Datasets Dispatched To Client In Specified Formats