

****Project Overview**

We need to build an AI system to **consolidate and analyze data from multiple lab instruments** studying exosomes (extracellular vesicles). The system should identify anomalies and patterns across different data types but **NOT interpret** the results—just flag them for researchers.

Input Data Sources

1. Flow Cytometry Data (FCS files)

- .fcs files containing scatter plot data
- Parameters: FSC (Forward Scatter), SSC (Side Scatter), and multiple fluorescence channels (FL1-FL6)
- Each event = one particle
- Need to parse using FlowCytometry libraries

2. Nanoparticle Tracking Analysis (Text files)

- The text file discussed.txt contains ZetaView output
- Size distribution data (particle size in nm, concentration, volume, area)
- Metadata: temperature, pH, conductivity, experimental conditions

3. Electron Microscope Images (TEM data)

- Image files showing exosomes
- Need computer vision to:
 - Detect scale bars
 - Measure particle sizes
 - Filter background noise
 - Identify viable exosomes

4. Western Blot Data (future integration - early 2025)

- Not yet provided, but needs to be architected for
-

Architecture Components

1. Data Ingestion Layer

Sub-components:

- **FCS File Parser:** Use fcsparser or FlowCytometryTools (Python)
- **Text File Parser:** Custom parser for ZetaView format
- **Image Processor:** OpenCV/PIL for TEM images
- **Metadata Extractor:** Parse experimental conditions from all sources

2. Data Preprocessing Layer

Sub-components:

- **Data Normalization:** Standardize units across different instruments
- **Quality Control Module:**
 - Check temperature compliance
 - Validate particle drift

- Filter invalid readings
- **Size Binning Engine:** Group particles by size ranges (40-80nm, 80-100nm, 100-120nm) based on customer-provided thresholds

3. Computer Vision Module (for TEM)

Sub-components:

- **Scale Detection:** Identify and measure scale bars
- **Particle Segmentation:** Separate exosomes from background
- **Size Measurement:** Calculate particle diameters
- **Noise Filtering:** Remove artifacts

4. Multi-Modal Data Fusion Layer

Sub-components:

- **Sample ID Matcher:** Link data from same sample across instruments
- **Feature Extraction:**
 - From FCS: scatter intensities, fluorescence profiles
 - From NTA: size distributions, concentrations
 - From TEM: morphology, size validation
- **Data Alignment:** Temporal and spatial correlation

5. Anomaly Detection Engine

Sub-components:

- **Scatter Plot Analyzer:**
 - Auto-select optimal X/Y axis combinations
 - Detect population shifts between readings
 - Identify outlier clusters
- **Statistical Comparison Module:**
 - Compare repeat measurements
 - Flag significant deviations
 - Cross-validate size data (NTA vs TEM)
- **Pattern Recognition:** Use ML (clustering, PCA) to find unusual patterns

6. Visualization & Reporting Layer

Sub-components:

- **Interactive Plot Generator:** Create scatter plots with highlighted anomalies
- **Comparison Dashboard:** Side-by-side views of multiple readings
- **Alert System:** Flag specific anomalies with timestamps
- **Export Module:** Generate reports in PDF/Excel

7. AI/ML Core

Sub-components:

- **Unsupervised Learning:**
 - K-means/DBSCAN for clustering
 - Autoencoders for anomaly detection
- **Semi-supervised Learning:** Use customer feedback to refine models
- **Feature Importance:** Identify which parameters matter most