**Data Collection & Storage**

**Goal:** Store event photos in a structured format for further processing.  
 **How?**

* Upload images to **Google Cloud Storage**.
* Use **TFX ExampleGen** to pull images into the pipeline.
* **Storage:** Data will be stored in a repository for version control and tracking

A screenshot of a computer program

Description automatically generated

**Data Preprocessing & Feature Extraction**

**Goal:** Process images and extract face embeddings.  
**How?**

* Use **TFX Transform** for image normalization and resizing.
* Apply a pre-trained **FaceNet** or **DeepFace** model for feature extraction.

A screenshot of a computer program

Description automatically generated

**Data Validation & Quality Check**

**Goal:** Ensure images are clean and usable.  
**How?**

* Use **TFDV** to check for missing or corrupted images.
* Detect anomalies using **Deequ** or **Great Expectations**.

A screenshot of a computer

Description automatically generated

**Data Versioning**

**Goal:** Track dataset changes over time.  
**How?**

* Use **DVC** to manage versions of image datasets.
* Store metadata with **Git LFS**.

**Commands:**

A close-up of a computer screen

Description automatically generated

**ML Pipeline Setup (Automation)**

**Goal:** Automate data flow using TFX & ZenML.  
**How?**

* Use **ZenML** to set up a modular pipeline.

A screenshot of a computer program

Description automatically generated

**Feature Store for Face Search**

**Goal:** Store face embeddings for fast lookup.  
 **How?**

* Use **Feast** as a feature store.
* Store embeddings in **Faiss** for efficient searching.

A screenshot of a computer code

Description automatically generated

**Deploy as a Web App**

**Goal:** Allow users to upload photos & search for people in events.  
 **How?**

* Use **FastAPI** for backend API.
* Build a simple **React.js frontend**.

A screenshot of a computer program

Description automatically generated