Software Requirements Specification (SRS): Teach by Doing (TbD) V3.0 - "Insight"

Status: Proposed **Author:** Greg Burns **Date:** November 23, 2025 **Previous Version:** V2.1 (Cloud Native MVP)

1. Introduction

1.1 Purpose

The purpose of this document is to define the functional and non-functional requirements for **Teach by Doing (TbD) V3.0**. This release, codenamed **"Insight"**, transitions the system from a basic data ingestion engine to an intelligent observation platform capable of semantic understanding.

1.2 Scope

V3.0 is a comprehensive upgrade to the "Brain" (Processing Pipeline) of the system. It retains the scalable V2 infrastructure (GCP Cloud Run/PubSub) but completely replaces the heuristic-based computer vision logic with a robust, multi-modal AI pipeline.

Key Capabilities:

- **Precision Vision:** Sub-pixel active region detection using SSIM.
- **Action Classification:** Distinguishing between clicks, scrolls, and typing using Optical Flow.
- **Semantic Understanding:** Generating human-readable step descriptions using Vertex AI (Gemini 2.5 Pro).

1.3 Strategic Goals

- **Eliminate "UNKNOWN_TEXT" Errors:** Achieve >85% successful OCR extraction rate.
- **Contextual Awareness:** Move beyond "User clicked X" to "User clicked X to perform Y."

2. Functional Requirements (FRs)

2.1 Enhanced Visual Processing (The "Eyes")

- **FR-01 (SSIM Differencing):** The system shall identify the "Active Region" of interaction by computing the **Structural Similarity Index (SSIM)** between the start and end frames of a segment to generate a precise bounding box of visual change.
- **FR-02 (Spatial OCR):** The system shall perform Optical Character Recognition (OCR) on the full frame but MUST filter results to include only text elements that geometrically intersect with the Active Region bounding box.
- **FR-03 (Optical Flow Analysis):** The system shall calculate dense optical flow vectors (using Farneback or similar algorithms) between frames to detect motion patterns.
- **FR-04 (Action Classification):** The system shall classify the user's action into one of the following categories based on visual evidence:
    - `click` (Localized change, low motion energy)
    - `scroll` (Global vertical motion vectors)
    - `type` (Rapid, localized high-frequency change)
    - `drag` (Sustained motion vectors across a specific path)

2.2 Semantic Enrichment (The "Brain")

- **FR-05 (Multimodal Reasoning):** The system shall integrate with **Google Cloud Vertex AI** to access the **Gemini 2.5 Pro** model.
- **FR-06 (Context Generation):** For each detected action node, the system shall construct a prompt containing:
    1. The Key Frame image.
    2. The extracted OCR text.
    3. The classified Action Type.
- **FR-07 (Description Output):** The system shall capture the LLM's response as the `semantic_description` field, providing a natural language summary of the step (e.g., "User clicked the 'Save' button to commit changes").

2.3 Data Output & Schema

- **FR-08 (V3 Schema Compliance):** The system shall output a JSON file conforming to the **PAD Schema v0.2**, which includes new fields for `action_class`, `semantic_description`, and `active_region_confidence`.

3. Non-Functional Requirements (NFRs)

3.1 Performance & Latency

- **NFR-01 (Processing Time):** The introduction of GenAI and advanced CV shall not increase total processing time beyond **4x the video duration** (e.g., a 1-minute video processes in <4 minutes).

- **NFR-02 (Parallelization):** Semantic enrichment calls to Vertex AI MUST be executed asynchronously/in parallel to minimize pipeline blocking.

3.2 Reliability & Accuracy

- **NFR-03 (OCR Accuracy):** The system shall achieve a minimum text extraction accuracy of 85% on standard SaaS interfaces (black text on white background).
- **NFR-04 (Hallucination Control):** The GenAI prompt temperature shall be set to `0.0` or `0.1` to minimize creative hallucinations and ensure factual descriptions.

3.3 Cost Management

- **NFR-05 (Model Efficiency):** The system shall default to **Gemini 2.5 Pro** for maximum capability but must be configurable to fallback to **Gemini 1.5 Flash** via environment variables if cost/quota limits are reached.

4. Interface Requirements

4.1 Output Interface

- **IR-01:** The final `Pathway.json` shall be uploaded to the existing GCS Output Bucket defined in V2.
- **IR-02:** The JSON structure shall be backward compatible with V1/V2 consumers (i.e., existing fields remain, new fields are additive).