Single-Page Group Progress Tracking Website Specification

November 03, 2025

Version 1.1 – Simplified from Retail Specification

1 Website Overview

This document specifies a single-page website designed to handle audio inputs of group progress updates through instant recording, uploads, or transcript uploads; transcribe them via Automatic Speech Recognition (ASR); summarize using Large Language Models (LLMs) in point-form; match summaries to a preset specification checklist extracted from a PDF; display coverage with evidence; and generate a follow-up email. The system includes a testing phase to evaluate ASR + LLM combinations. The website is a private-access demo with a simple interface for inputs and result displays.

2 Frontend

The website operates as a single-page application with the following features:

- **Input Area**:
  - Supports instant audio recording (browser-based, max 5 minutes).
  - Allows uploading of audio files (formats: MP3, WAV; max size: 10MB).
  - Permits uploading of transcripts (formats: TXT, DOCX).
  - Validates inputs and displays confirmation upon success.
  - Includes a drag-and-drop interface for uploads.

- **Results Display**:
  - Shows specification coverage in a table with overall score and breakdowns.
  - Displays scoring evidence (e.g., matched summary points with quotes from transcript).
  - Presents LLM-generated point-form summary.
  - Includes auto-generated follow-up email draft.
  - Includes a progress indicator for processing (e.g., ASR, summarization).
  - Provides error messages for invalid inputs or processing failures.

3 Backend

The backend manages storage, audio-to-text conversion, summarization, and specification matching:

- **Storage**:
  - Temporarily saves recorded or uploaded audio files (retained for 24 hours).
  - Saves transcripts (uploaded directly or generated from ASR).

- **ASR Conversion**:
  - Converts audio to text using selected ASR tools (e.g., OpenAI Whisper).

- Handles Cantonese and English by default.

- **Summarization**:

  - Uses LLMs to generate point-form summaries of the transcript.

- **Specification Matching**:

  - Extracts checklist items from an uploaded **PDF specification** (e.g., `Retail_specification_1.pdf`).
  - Matches summary points to extracted items.
  - Computes coverage and collects evidence (e.g., relevant transcript excerpts).
  - Returns match results to the frontend for display.

- **Email Generation**:

  - Auto-generates a follow-up email based on covered and missing items.

4 Testing and Validation

To ensure quality, the system includes a testing phase:

- **Combination Evaluation**:

  - Evaluates ASR + LLM pairs for end-to-end performance.
  - Tests on a dataset of at least 100 audio samples from a single domain.
  - Automatically computes metrics for selected combinations to determine the best (e.g., highest combined score).

5 Backend Testing Page

The backend includes a dedicated testing page for dynamic selection and evaluation of ASR tools, LLMs, and their combinations. This page allows changes for testing and deployment without code modifications.

- **Dynamic Selection Features**:

  - Allows selection of ASR tools from a configurable list (e.g., Whisper, Google Speech-to-Text).
  - Enables selection of LLMs for summarization (e.g., GPT-4, Claude).
  - Supports choosing combinations of ASR + LLM.
  - **Upload the specification PDF** — this is where the preset checklist is input.
  - Triggers automated computation of metrics on test data.
  - Displays comparison results to identify the best combination.

- **Deployment Integration**:

  - Configurations (including uploaded PDF) can be applied to the production backend.
  - Changes are persisted and take effect immediately or upon restart.
  - Includes logging of test results, metrics, and configuration history.

- **Access and Security**:

  - Restricted to authorized users (e.g., admins) via authentication.
  - Provides a user-friendly interface with dropdowns, **PDF upload button**, and preview of extracted items.