

Temporal Interview Profiling System (TIPS) – Synopsis

1. Project Title

Temporal Interview Profiling System (TIPS)

2. Problem Statement

Modern interviews increasingly occur online, where HR teams and recruiters must evaluate not only the content of a candidate's responses but also their *behavioral signals*—confidence, stress, clarity, eye contact, and overall communication quality. Human evaluation is subjective, inconsistent, and time-consuming. Existing automated assessment systems rely on superficial metrics and do not analyze multimodal cues (audio, video, speech, behavioral patterns) over time.

There is a need for an intelligent, temporal, multimodal system that can automatically profile a candidate's interview performance based on synchronized audio, video, and speech patterns.

3. Objectives

The main objectives of TIPS are:

- To build an automated system that evaluates interview performance using temporal analysis.
 - To predict three key behavioral scores: **confidence**, **stress**, and **clarity**.
 - To extract multimodal features from audio, video, and speech transcripts.
 - To apply Transformer-based temporal modeling for sequential behavior understanding.
 - To generate an automated, rule-based natural-language summary of the candidate's performance.
-

4. Scope of the Project

The system will:

- Process recorded interviews (audio + video) locally.
- Extract per-second behavioral features such as pauses, gaze direction, facial expressions, speaking rate, and tone.

- Produce numerical performance scores.
- Produce an automatically generated textual summary.

The system will **not**:

- Perform identity recognition.
 - Judge domain knowledge or correctness of answers.
 - Access or upload data to cloud servers (fully local system).
-

5. Methodology

5.1 Data Collection

Interviews of 5–15 minutes are recorded with synchronized audio and video. Metadata is maintained for each session.

5.2 Feature Extraction

A multimodal feature extraction pipeline is used:

- **Audio features:** MFCCs, energy, pitch, speaking activity, pause ratios.
- **Text features:** words per second, filler-word ratio, lexical diversity.
- **Visual features:** facial landmarks, head pose, gaze direction, eye openness, expression probabilities.

Features are computed on 1-second windows with 0.5-second hop.

5.3 Temporal Modeling

A Transformer-based temporal encoder processes the full sequence of features for each interview session. The model outputs three continuous values:

- **Confidence (0–100)**
- **Stress (0–100)**
- **Clarity (0–100)**

5.4 Summarization Module

A rule-based natural-language generator converts predictions and detected behavioral patterns into a structured interview performance summary.

6. Expected Outcomes

- Accurate prediction of behavioral scores for each candidate.
 - Clear, human-readable performance summary.
 - Detailed timeline of detected events (pauses, stress spikes, gaze deviations).
 - Improved consistency and objectivity in interview evaluation.
-

7. Applications

- HR screening and online recruitment systems.
 - Mock interview and training platforms.
 - Soft-skills assessment tools for students and professionals.
 - Behavioral research and communication analytics.
-

8. Hardware & Software Requirements

Hardware:

- GPU-enabled system
- Webcam and microphone

Software:

- Python 3.x
 - PyTorch / TensorFlow
 - MediaPipe
 - FFmpeg
 - Whisper/Vosk for STT
 - NumPy, Pandas, Matplotlib
-

9. Conclusion

The Temporal Interview Profiling System (TIPS) provides a comprehensive, multimodal, and temporally aware solution for analyzing interview behavior. By combining audio, visual, and linguistic signals through advanced temporal modeling, TIPS delivers objective behavioral scoring and an automated performance summary. This system improves evaluation efficiency, reduces bias, and supports both recruitment and interview training ecosystems.