1 Temporal Interview Profiling System (TIPS)

1.1 Automated Interview Analysis System

- Processes pre-recorded video interviews

- Extracts behavioral signals from audio and video

- Evaluates candidate responses using LLMs

- Generates time-evolving performance scores

- Compares answers against job descriptions

2 Abstract

- TIPS is an automated interview analysis system for video interviews

- Multi-stage backend pipeline processes recordings through 6 stages

- Extracts audio signals: transcription, pitch, energy, speech rate

- Analyzes video signals: face detection, head pose, gaze stability

- Uses Qwen2.5-3B LLM for semantic relevance scoring

3   Problem Statement

- **Subjectivity**: Human interviewers have inherent biases and varying standards

- **Limited Analysis Depth**: Manual evaluation misses vocal and temporal patterns

- **Scalability Issues**: Time for evaluations grows linearly with candidate volume

- **Lack of Standardization**: Comparing candidates across interviewers is difficult

- **No Temporal Insights**: Traditional scoring gives only aggregate scores

4 Motivation

- Modern recruitment faces bottlenecks in interview evaluation

- Organizations invest significant resources with inconsistent results

- Automated systems can capture signals humans often miss

- Modular design allows independent optimization of each stage

- Functions as addon to existing interview platforms

5 System Architecture

5.1 Pipeline Flow

1. **Input**: Interview Recording + Job Description

2. **Stage 0**: Timebase Establishment

3. **Stage 1**: Parallel Signal Extraction (Audio/Video)

4. **Stage 2**: Temporal Segmentation (Q&A Pairing)

5. **Stage 3**: Behavioral Metrics Computation

6. **Stage 4-5**: LLM Semantic Scoring + Verdict

5.2 Components

- Interview UI Layer (WebRTC Recording)

- Backend Pipeline (6-Stage Processing)

- Dashboard (Visualization - Planned)

6 Methodology

6.1 Pipeline Overview

- **Stage 0**: Establishes canonical timebase from video metadata

- **Stage 1**: Parallel extraction of audio features, video features, transcription

- **Stage 2**: Identifies speaking segments and pairs Q&A

- **Stage 3**: Computes behavioral metrics (confidence, fluency, eye contact)

- **Stage 4-5**: LLM scoring and final verdict aggregation

6.2 Key Processing Stages

- **Faster-Whisper**: Speech-to-text with word-level timestamps

- **WebRTC VAD**: Voice activity detection for speech segments

- **MediaPipe**: Face detection, landmark tracking, head pose estimation

- **Librosa**: Audio features (pitch, energy, speech rate)

- **Qwen2.5-3B**: 4-bit quantized LLM for semantic evaluation

7 Results

7.1 Output Artifacts

- **timeline.json**: Master timebase synchronization

- **qa_pairs.json**: Question-answer mappings with timing

- **candidate_behavior_metrics.json**: Confidence, fluency metrics

- **relevance_scores.json**: LLM-evaluated relevance per Q&A

- **candidate_score_timeline.json**: Time-evolving performance scores

8   Evaluation and Validation

- Verdict categories: STRONG_HIRE, HIRE, BORDERLINE, NO_HIRE

- 5 competency dimensions: Technical depth, System design, Production experience, Communication clarity, Problem solving

- Matched keywords extraction from job description

- Incremental verdict progression throughout interview

- Confidence level provided with final recommendation

9 Future Enhancements

- Dashboard visualization for result presentation

- Real-time streaming analysis capability

- Multi-language support for transcription

- Enhanced video analytics (micro-expressions, body language)

- Integration with existing ATS platforms

10   Conclusion

- TIPS provides objective, data-driven candidate assessment

- Multi-stage pipeline enables comprehensive interview analysis

- LLM integration allows semantic evaluation against job requirements

- Temporal tracking shows candidate performance evolution

- Modular design facilitates future enhancements and optimizations