Project Specifications Report

1. Introduction

This project is developed as a graduation project for TOBB University of Economics and Technology (TOBB ETÜ) in the 2025–2026 Spring Term. Its primary goal is to help new graduates and early-career candidates prepare for real-world interviews by offering a practical, guided mock interview experience and delivering structured feedback that highlights strengths and improvement areas. The platform targets both HR and technical interview practice in a time-bounded format to support repeatable learning and progression.

1.1 Description

This project delivers a web-based AI Mock Interview platform for new graduates and early-career candidates, providing short (10–15 minute) HR or Technical (non-coding) interview simulations through a voice-based AI interviewer. The system offers two interaction modes—Supportive and Neutral—where both modes follow the same evaluation rubric but differ in the interviewer's conversational behavior and guidance intensity. Candidate responses are captured via microphone and transcribed using Speech-to-Text (STT), while the interviewer speaks using Text-to-Speech (TTS). Each session is personalized based on the user's selected target role/position, company context, domain/interest areas, and difficulty level (e.g., junior/intermediate). Before starting, the user can view a sample question preview to align expectations on scope and difficulty; however, the preview does not strictly guarantee that the exact same questions will appear during the interview.

The primary focus is Supportive Mode and the feedback system. In Supportive Mode, the interviewer acts as a coach: when the candidate drifts off-topic, it gently redirects them with empathetic prompts (e.g., "We may be moving slightly away from the question—let's refocus on…"), and when the candidate expresses uncertainty (e.g., "I don't know" / "I'm not sure"), it helps them continue by opening the topic with clarifying sub-questions and small hints rather than penalizing them. In addition, time-bound intervention is applied in both modes: if a response exceeds the allocated time window, the interviewer politely prompts the candidate to summarize and return to the core point. If time permits, the platform may also include optional enhancement features such as question-specific Hint Cards on demand (e.g., STAR structure for HR answers or key points for technical explanations), a human avatar interviewer (visual representation of the interviewer), login and progress tracking/dashboard, and CV-based HR personalization. These optional items are considered stretch goals and are not guaranteed due to project time constraints. After each session, the system produces a scored feedback report combining content evaluation (relevance, clarity, completeness) with measurable communication signals such as fluency indicators (long pauses/silence, filler-word frequency, speaking pace/energy) and camera-based behavioral analysis (with explicit

user consent), including a focus/engagement proxy, head movement patterns, and coarse facial expression cues (e.g., tension vs. positive expression). These signals are aggregated into clear scores and actionable recommendations, enabling candidates to iteratively improve interview performance through evidence-based coaching rather than generic advice.

The system integrates the OpenAI API as one of its core AI components. The system uses the OpenAI API primarily for LLM-based capabilities, such as generating interview questions, managing the interviewer's dialogue, and producing content-level feedback from the transcribed answers. To ensure consistency and prevent scope drift, the platform does not rely on the LLM alone; instead, it applies application-side rules and guardrails (e.g., difficulty boundaries, topic constraints, time-limit interventions, and mode-specific guidance behaviors) that control how the interview progresses and how outputs are presented. In parallel, the behavioral feedback pipeline (e.g., focus/engagement proxy, head movement, coarse facial expression cues, pauses/silence, filler-word frequency, and basic speech rhythm indicators) is planned to be implemented using open-source computer vision and audio signal processing approaches, operating as separate scoring components that complement the LLM's content evaluation.

1.2 Constraints

1.2.1 Economic / Financial Constraints

- **C-01 Budget limitations:** The project must minimize paid API usage during development and testing.
- **C-02 Cost-aware operation:** The system is constrained to short sessions (10–15 minutes) and efficient AI usage to control recurring costs.
- **C-03 Resource allocation:** Team size and limited availability require strict MVP(Minimum Viable Product) prioritization; high-effort features remain optional.

Impact: Encourages limited scope, efficient request design, and careful prioritization of core vs. optional features.

1.2.2 Environmental and Socio-Cultural Constraints

- **C-04 Language adaptation:** The system must support a Turkish interview experience (questions, interviewer voice, and feedback language).

- **C-05 Accessibility and usability:** The web interface must remain usable on standard devices and browsers with a simple setup flow.

- **C-06 Communication style:** Feedback must be constructive and coaching-oriented, especially in Supportive Mode.

Impact: Influences UX, wording, and tone policies, ensuring the experience remains culturally appropriate and supportive.

1.2.3 Manufacturability, Health, and Safety Constraints

- **C-07 Psychological safety:** The system must avoid humiliating or harmful feedback and present coaching-oriented guidance.

- **C-08 Safe interpretation of signals:** Behavioral feedback must be framed as observational coaching (non-clinical).

- **C-09 Session ergonomics:** Short sessions reduce cognitive load and enable repeated practice cycles.

Impact: Constrains how behavioral signals are interpreted and how feedback is phrased.

1.2.4 Legal, Ethical, Temporal, Technological, and Sustainability Constraints

- **C-10 Consent-first operation (KVKK):** Microphone and optional camera access require explicit consent and transparency.

- **C-11 Minimal retention by default:** Raw audio/video should not be stored persistently unless optional account/history features are implemented with clear retention rules.

- **C-12 Temporal constraints:** The project must fit within the academic timeline; optional features may be excluded.

- **C-13 External dependency constraints:** Availability and latency of external AI services may vary; the system must degrade gracefully.

- **C-14 Sustainability:** Resource-efficient processing and reduced redundant AI calls are preferred.

Impact: Drives privacy-by-design, robust error handling, and scope control.

1.2.5 Design Constraints

- **C-15 Web application constraint:** The system shall be delivered as a web application accessible via standard modern browsers.
- **C-16 Backend integration constraint:** The system shall perform external AI calls through a backend service; API keys shall not be present in frontend code.
- **C-17 Privacy constraint:** The system shall apply consent-first access to microphone/camera and follow minimal data retention by default.
- **C-18 Scope constraint:** The system shall exclude coding interviews and senior-level depth as a deliberate design limitation.
- **C-19 Language constraint:** The system shall support Turkish as the primary interview language for the MVP(Minimum Viable Product).

Impact: Keeps the MVP secure, web-accessible, and realistically scoped.

1.3 Professional and Ethical Issues

Consent & transparency: Users must explicitly grant camera/microphone permissions and be informed about what is processed (audio/video), how it is used (feedback), and whether it is stored.

KVKK alignment: Minimize personally identifiable data. Prefer processing without storing raw audio/video when possible. If storage is added (optional, tied to login), retention must be time-limited and clearly stated.

Bias and fairness: Feedback should be framed as coaching, avoid sensitive inferences (e.g., personality/mental health claims), and focus on observable behaviors (pauses, relevance, eye-contact proxy).

Safety & user well-being: Supportive Mode should avoid harsh or humiliating language. The system should present feedback as improvement-oriented suggestions rather than definitive judgments.

2. Requirements

2.1 Functional Requirements (FR)

FR-01: The system shall allow the user to select an interview type (HR or Technical) before starting a session.

FR-02: The system shall allow the user to provide interview context, including target role/position, company or industry context.

FR-03: The system shall allow the user to select a domain/interest area and a difficulty level suitable for junior and intermediate candidates.

FR-04: The system shall provide two interaction modes: Supportive and Neutral.

FR-05: The system shall conduct a voice-based interview session with a target duration of 10–15 minutes.

FR-06: The system shall capture candidate speech via microphone and convert it to text (STT) to support interview flow and analysis.

FR-07: The system shall generate interview questions and follow-up prompts consistent with the selected context and difficulty level.

FR-08: The system shall enforce time boundaries in all modes by prompting the candidate to summarize when answer duration exceeds the allocated window.

FR-09: The system shall, in Supportive Mode, gently steer the candidate back to the question when answers drift off-topic.

FR-10: The system shall, in Supportive Mode, provide supportive guidance when the candidate expresses uncertainty (e.g., “I don’t know / I’m not sure”) using clarifying sub-questions and small hints.

FR-11: The system shall generate a post-session feedback report including scores and actionable recommendations based on both content quality and communication/behavioral signals.

FR-12: The system shall provide interviewer speech output using text-to-speech (TTS).

FR-13: The system shall use camera input (with explicit user consent) to extract behavioral signals for feedback (e.g., focus/engagement proxy, head movement, and coarse facial expression cues).

2.2 Nonfunctional Requirements (NFR)

NFR-01 Usability: The system shall provide a simple, guided flow enabling users to start an interview in a small number of steps.

NFR-02 Responsiveness: The system shall provide an interactive experience with conversational responses produced within a reasonable time for live usage.

NFR-03 Reliability: The system shall complete the interview and provide feedback even if some non-critical components degrade (e.g., partial loss of specific behavioral signals), while maintaining the core voice-based flow.

NFR-04 Privacy and security: The system shall require explicit consent for microphone/camera usage and shall not expose API credentials in the client application.

NFR-05 Maintainability: The system shall be modular to allow optional features to be added or removed without impacting the core interview flow.

2.3 Optional Requirements (OR) — Not Guaranteed

OR-01 Login / User Accounts (Optional): The system may provide authentication to enable cross-device access and persistent session history.

OR-02 Dashboard / Progress Tracking (Optional): The system may provide a dashboard summarizing past sessions and score trends if accounts are enabled.

OR-03 CV-Based HR Personalization (Optional): The system may allow CV upload or CV text input to tailor HR questions and follow-ups to candidate background.

OR-04 Hint Cards UI (Optional): The system may provide an optional panel showing short, question-specific hints (e.g., STAR structure, key points) when the user requests it in Supportive Mode.

OR-05 Human Avatar Interviewer (Optional): The system may render an optional human avatar for the interviewer; the default remains voice-only.

2.4 External Interface Requirements

2.4.1 User Interface Requirements (UIR)

UIR-01: The system shall provide a web-based setup interface for selecting interview type, context, difficulty, and mode.

UIR-02: The system shall provide an interview interface that presents the current question and supports voice-based interaction.

UIR-03: The system shall provide a post-session interface that displays feedback scores and recommendations.

UIR-04 (Optional): The system may provide login and dashboard interfaces if optional account features are implemented.

2.4.2 Hardware Interface Requirements (HIR)

HIR-01: The system shall interface with a microphone to capture user speech.

HIR-02: The system shall interface with a camera (with explicit user consent) to extract behavioral signals used for feedback.

2.4.3 Software Interface Requirements (SIR)

SIR-01: The system shall interface with external AI services for language reasoning, speech-to-text, and text-to-speech generation.

SIR-02: The system shall route external AI calls through a backend service to protect API keys and credentials.

SIR-03 (Optional): The system may interface with a persistent database if login and session history are implemented.

2.5 Performance Requirements

2.5.1 Responsiveness and Timing (PR)

PR-01: The system shall support completing a full interview session within 10–15 minutes.

PR-02: The system shall provide interviewer responses with acceptable perceived latency suitable for live conversation.

PR-03: The system shall produce a post-session feedback report within a reasonable time after the session ends.

2.5.2 Robustness Under Partial Availability (PR)

PR-04: The system shall continue operating and provide feedback even if some behavioral signal extraction is degraded, while preserving the core voice-based interview flow.

PR-05: The system shall limit redundant external AI calls during a session to control cost and maintain responsiveness.

2.6 Pseudo requirements

PSR-01 Guided interview flow: The system is expected to maintain a coherent interview progression, ensuring that questions, follow-up prompts, and timing interventions support a realistic interview experience.

PSR-02 Mode-consistent behavior: The interviewer's interaction style is expected to remain consistent with the selected mode (Supportive or Neutral) throughout the session, while applying the same evaluation criteria in both modes.

PSR-03 Balanced guidance: In Supportive Mode, guidance behaviors such as redirection or clarification are expected to assist the user without interrupting the interview flow or revealing direct answers.

PSR-04 Controlled AI behavior: AI-driven components are expected to operate within predefined scope boundaries, including difficulty level, topic relevance, and session timing, through rule-based guardrails.

PSR-05 Transparent feedback presentation: Feedback is expected to be presented in a clear and interpretable manner, allowing users to understand how scores and observations relate to their interview performance.

PSR-06 Non-intrusive behavioral analysis: Behavioral signal extraction is expected to operate unobtrusively during the interview and be used solely for post-session feedback, subject to user consent.

PSR-07 Graceful degradation: If certain analysis components become unavailable or produce incomplete data, the system is expected to continue the interview and provide feedback based on available information.

3. References

1. ISO. *ISO 9001:2015 — Quality Management Systems* (Quality and process principles referenced in constraint analysis).
2. ISO/IEC/IEEE. *29148: Systems and software engineering — Requirements engineering* (Requirement statement conventions and specification structure).
3. Republic of Türkiye. *Law No. 6698 on the Protection of Personal Data (KVKK)* (Consent, data processing, and retention principles).