# Digital Twin Project Requirements Document

## 1. Project Overview

The **Digital Twin** project aims to create an AI-powered virtual representation of an individual (the user) to enhance productivity, automate meeting participation, and replicate the user’s communication style. The project is divided into three phases:

- **Phase 1**: An agent monitors the user’s mailbox, joins all calendar-invited meetings (via Google Meet) records sessions, transcribes and summarizes them, and stores outputs in cloud storage.

- **Phase 2**: AI models analyze recordings to learn the user’s meeting participation patterns, emotional expressions, and discussion styles, generating deep intelligence about meetings.

- **Phase 3**: A voice-cloned digital twin represents the user, engaging in conversations via a dialog box or text-to-voice converter, replicating the user’s voice, emotions, and meeting behavior.

## 2. Objectives

- Automate meeting attendance and documentation to save time.

- Develop AI-driven insights into the user’s meeting behaviors and preferences.

- Create a realistic digital twin capable of representing the user in conversations and meetings.

- Ensure secure, scalable, and user-friendly implementation across all phases.

## 3. Scope

**Included:**

- Mailbox monitoring and calendar integration.

- Meeting recording, transcription, summarization, and cloud storage.

- AI training for behavioral analysis.

- Voice cloning and conversational digital twin with text-to-voice capabilities.

- Support for Google Meet.

## 4. Requirements

**4.1 Phase 1: Mailbox Monitoring and Meeting Automation**

**Functional Requirements:**

- **FR1.1**: The agent must monitor the user’s mailbox (e.g., Gmail, Outlook) in real-time for new calendar invites.

- **FR1.2**: The agent must extract meeting details (e.g., date, time, platform, URL) from invites across Google Calendar.

- **FR1.3**: The agent must automatically join every meeting via Google Meet using the provided URL, authenticating with the user’s credentials.

- **FR1.4**: The agent must record audio and video of each meeting session.

- **FR1.5**: The agent must transcribe recordings using speech-to-text technology, ensuring >95% accuracy for clear audio.

- **FR1.6**: The agent must generate a concise summary (100–200 words) of each meeting, highlighting key points, decisions, and action items.

- **FR1.7**: The agent must store recordings, transcriptions, and summaries in a secure cloud storage solution.

- **FR1.8**: The system must provide a user interface (e.g., web dashboard) to view, search, and download stored meeting data.

- **FR1.9**: The agent must handle multiple concurrent meetings, prioritizing them based on user-defined rules (e.g., meeting organizer, topic).

**4.2 Phase 2: AI Training and Behavioral Analysis**

**Functional Requirements:**

- **FR2.1**: The system must ingest all Phase 1 recordings, transcriptions, and summaries for AI training.

- **FR2.2**: AI models must analyze the user’s speech patterns, including tone, pace, and vocabulary, to quantify participation style.

- **FR2.3**: AI models must detect emotional expressions (e.g., confidence, hesitation, enthusiasm) using audio and text analysis.

- **FR2.4**: The system must identify the user’s meeting roles (e.g., leader, contributor, listener) based on participation frequency and content.

- **FR2.5**: The system must generate insights per meeting, including:

- User’s contribution level (e.g., percentage of speaking time).

- Key discussion points led or influenced by the user.

- Emotional trends (e.g., positive/negative sentiment).

- **FR2.6**: The system must provide a report dashboard summarizing behavioral patterns across meetings (e.g., average speaking time, common topics).

- **FR2.7**: AI models must update incrementally as new meeting data is added, improving accuracy over time.

**4.3 Phase 3: Voice-Cloned Digital Twin**

**Functional Requirements:**

- **FR3.1**: The system must create a voice clone of the user using at least 30 minutes of high-quality audio from Phase 1 recordings.

- **FR3.2**: The voice clone must replicate the user’s tone, pitch, and emotional nuances with >95% perceptual similarity.

- **FR3.3**: The digital twin must engage in conversations via a dialog box (text-based) or text-to-voice converter, responding in real-time.

- **FR3.4**: The digital twin must use Phase 2 insights to emulate the user’s meeting behavior, including vocabulary, participation style, and emotional responses.

- **FR3.5**: The digital twin must join meetings on behalf of the user, responding to questions or prompts based on learned patterns.

- **FR3.6**: The system must allow the user to review and edit the digital twin’s responses before or during meetings.

- **FR3.7**: The digital twin must integrate with Google Meet for meeting participation.

- **FR3.8**: The system must provide a log of the digital twin’s interactions, including conversation transcripts and meeting outcomes.

## 5. Assumptions

- The user has a single mailbox and calendar for all meeting invites.

- Meeting platforms provide API access for joining and recording.

- The user consents to recording meetings and complies with local laws (e.g., notifying participants).

- Sufficient high-quality audio is available for voice cloning.

- Cloud storage and AI processing infrastructure are available (e.g., AWS, Azure).

## 6. Risks and Mitigation

- **Risk 1**: Inaccurate transcriptions due to poor audio quality.

- **Mitigation**: Use noise-canceling algorithms and fallback to manual review for low-confidence transcriptions.

- **Risk 2**: Privacy breaches from stored meeting data.

- **Mitigation**: Implement end-to-end encryption and role-based access controls.

- **Risk 3**: Meeting platforms blocking automated agents.

- **Mitigation**: Use official APIs and maintain compliance with platform terms of service.

- **Risk 4**: Voice clone sounding unnatural or failing to capture emotional nuance.

- **Mitigation**: Train on diverse audio samples and incorporate user feedback for refinement.

- **Risk 5**: Ethical concerns about digital twin impersonation.

- **Mitigation**: Clearly disclose the digital twin’s identity in meetings and obtain participant consent.

## 7. Deliverables

- **Phase 1**: Mailbox monitoring agent, meeting joiner, recording system, transcription and summarization engine, cloud storage integration, user dashboard.

- **Phase 2**: AI training pipeline, behavioral analysis models, insight reports, dashboard enhancements.

- **Phase 3**: Voice clone model, digital twin conversation interface (dialog box and text-to-voice), meeting integration module, interaction logs.

## 8. Success Criteria

- **Phase 1**: The agent joins 100% of the meetings invited, with >95% transcription accuracy and summaries stored securely within 10 minutes post-meeting.

- **Phase 2**: AI models achieve >90% accuracy in identifying user roles and emotions across 50 meetings, with reports generated within 5 seconds.

- **Phase 3**: The digital twin replicates the user’s voice with >95% similarity, responds appropriately in 90% of meeting scenarios, and logs all interactions accurately.

## 9. Dependencies

- **External APIs**: Google Calendar, RecallAI (Only paid API that we are allowed and that only for allowing the bot to join meeting, recording and transcription)

- **Cloud Services**: Something open source for uploading the recordings (but keep in mind that it should be open source as well)

- **Third-Party Tools**: Voice cloning via Open Voice Model.

- **User Data**: Access to mailbox, calendar, and sufficient audio samples.