AI Debater Project - Technical Development Report

Executive Summary

The AI Debater project is a sophisticated conversational AI system that enables users to engage in medical debates with AI-powered character avatars. The system combines advanced language models, custom voice cloning, lip-sync technology, and interactive web interfaces to create an immersive debating experience. This report details the technical architecture, development processes, and integration strategies used to build this innovative platform.

1. Project Architecture Overview

1.1 System Components

The AI Debater consists of several interconnected components:

- Frontend Interface: Web-based control panel and avatar display system
- Backend API: Node.js server handling agent configuration and API routing
- ElevenLabs ConvAI: Conversational AI engine with custom voice cloning
- Lip-sync Generation: Sync.so platform integration for character introduction and winning speech videos
- Character Management: Multi-personality system with customizable opponents

1.2 Technology Stack

- Frontend: HTML5, CSS3, JavaScript (ES6+) • Backend: Node.js with Express.js
- AI Engine: ElevenLabs Conversational LLM
- Voice Synthesis: ElevenLabs Voice Cloning
- Lip-sync: Sync.so for introduction and winning speech videos
- Deployment: Static file serving with dynamic API endpoints

2. Backend Development

2.1 Node.js Server Architecture

The project implements a Node.js backend solution providing robust API services and static file serving.

Node.js Implementation (server.js)

Key features:

- Express.js framework for API routing
- CORS middleware for cross-origin requests
- Environment-based agent configuration
- Signed URL generation for secure ConvAI access
- Static file serving for frontend assets

2.2 Agent Configuration System

The backend implements a sophisticated agent mapping system that routes different characters to specific AI agents:

- Default Agent: General-purpose medical debater
- Nelson Mandela: Anti-apartheid leader persona with separate Q&A and debate modes
- Michelle (Barbarella): Singaporean model character
- Taylor Swift: Pop icon personality

2.3 API Endpoints

- `/api/signed-url`: Generates secure WebSocket URLs for ConvAI connections
- `/api/getAgentId`: Retrieves agent identifiers for public access
- Static file serving for frontend assets
- Route handling for specialized HTML pages

3. Frontend Development

3.1 Control Panel Interface

The control panel (controls.html) provides a sophisticated user interface for managing debates:

Features:

- Topic Selection: Predefined debate topics with medical focus
- Opponent Selection: Character-based opponent choosing system
- Real-time Status: Connection and speaking status indicators
- Recording Controls: Start/stop debate functionality
- Avatar Management: Integrated avatar window controls

UI/UX Design:

- Glassmorphism Design: Backdrop blur effects with transparent overlays

- Gradient Backgrounds: Multi-layered gradient systems with particle animations
- Responsive Layout: Mobile-first design with adaptive breakpoints
- Interactive Elements: Hover effects and micro-animations

3.2 Avatar Display System

The avatar interface (avatar.html) manages character visualization:

Core Features:

- Video Playback: Lip-sync video display with fallback images
- Speaking Indicators: Visual feedback for AI speech activity
- Responsive Scaling: Automatic sizing based on viewport dimensions
- Fullscreen Support: Immersive viewing experience

Technical Implementation:

- Centered Video Display: Absolute positioning with transform-based centering
- Aspect Ratio Preservation: Object-fit contain for maintaining video proportions
- Z-index Management: Layered display system for video/image switching
- Animation Effects: Subtle scaling and pulsing during speech

4. Voice Cloning and Synthesis

4.1 ElevenLabs Voice Cloning Process

The voice cloning implementation follows a systematic approach:

Data Collection:

- Reference Audio: Collected interviews, speeches, and songs from target personalities
- Audio Quality: High-fidelity recordings with minimal background noise
- Content Diversity: Varied speaking styles and emotional ranges

Voice Training Parameters:

- Stability: Adjusted for consistent character voice output
- Similarity Boost: Fine-tuned to match original speaker characteristics
- Speed Control: Customized speaking rates per character
- Clarity Enhancement: Optimized for clear speech generation **Textual Prosody**

Implementation:

- Rhythm Control: Manipulated speech timing and pacing
- Stress Patterns: Emphasized important debate points
- Intonation Matching: Preserved character-specific speech patterns
- Accent Preservation: Maintained regional and cultural speech characteristics

4.2 Character-Specific Voice Profiles

Each character has unique voice configuration:

- Nelson Mandela: Deep, authoritative tone with measured pacing
- Michelle (Barbarella): Singaporean accent with modern inflection
- Taylor Swift: American accent with expressive emotional range

5. Lip-Sync Technology Integration

5.1 Sync.so Platform Implementation

The lip-sync generation process is utilized for specific video content:

Use Cases:

- Taylor Swift Introduction Video: Guard of Honour presentation
- Winning Speech Videos: Victory presentations for all three characters (Nelson Mandela, Michelle/Barbarella, Taylor Swift)

Workflow:

1. Audio Processing: ElevenLabs generates character-specific audio for introduction and winning speeches
2. Video Generation: Sync.so creates mouth movements matching audio
3. Facial Animation: Realistic expressions synchronized with speech
4. Quality Optimization: Fine-tuning for natural appearance

Technical Specifications:

- Video Format: MP4 with H.264 encoding
- Resolution: Optimized for web delivery
- Frame Rate: 30 FPS for smooth animation
- Audio Sync: Precise timing alignment

5.2 Avatar Animation System

The frontend avatar system manages video display for special content:

Features:

- Introduction Video Playback: Taylor Swift Guard of Honour presentation
- Winning Speech Display: Character-specific victory videos
- Fallback Handling: Graceful degradation to static images
- Performance Optimization: Efficient video loading and playback
- Cross-browser Compatibility: Consistent experience across platforms

6. LLM Integration and Prompt Engineering

6.1 ElevenLabs ConvAI Integration

The conversational AI system utilizes:

Core Components:

- Conversational LLM: Context-aware response generation
- Retrieval-Augmented Generation (RAG): External knowledge base integration
- Real-time Processing: Low-latency response generation
- Context Management: Maintaining conversation history and character consistency

RAG Implementation:

- Knowledge Base Linking: URLs connected to external medical databases
- Information Retrieval: Accurate and relevant fact-checking
- Source Grounding: Arguments backed by credible information
- Dynamic Updates: Real-time knowledge integration

6.2 Prompt Engineering Strategy

Sophisticated prompt design ensures character authenticity:

Character Consistency:

- Personality Traits: Detailed character backgrounds and speaking styles
- Historical Context: Accurate representation of real personalities
- Debate Strategy: Character-appropriate argumentation approaches
- Response Length: Optimized for natural conversation flow **Medical Domain**

Expertise:

- Topic Specialization: Focused on healthcare and medical ethics
- Argument Structure: Logical reasoning and evidence-based responses

- Counterargument Handling: Sophisticated debate tactics
- Factual Accuracy: Medically sound information and analysis

7. User Interface and Experience Design

7.1 Visual Design System

The interface employs modern web design principles:

Design Elements:

- Color Palette: Medical-themed blues with vibrant accent colors
- Typography: Inter font family for readability and professionalism
- Iconography: Consistent SVG icons for all interface elements
- Animations: Subtle transitions and hover effects

Responsive Design:

- Mobile-First: Optimized for smartphone and tablet usage
- Flexible Layouts: CSS Grid and Flexbox for adaptive layouts
- Touch-Friendly: Appropriately sized buttons and interactive elements
- Performance: Optimized CSS and minimal JavaScript overhead

7.2 Interaction Design

User experience focuses on intuitive operation:

Navigation Flow:

- Character Selection: Simple button-based opponent choosing
- Debate Initiation: One-click start functionality
- Status Feedback: Real-time connection and speaking indicators
- Avatar Controls: Integrated window management **Accessibility Features:**

- Keyboard Navigation: Full keyboard accessibility
- Screen Reader Support: Semantic HTML structure
- High Contrast: Readable text and UI elements
- Responsive Text: Scalable typography for visual accessibility

8. Special Video Content

8.1 Introduction Videos

Taylor Swift Guard of Honour: A specially created introduction video using sync.so technology to present Taylor Swift's character entry into the debate arena.

8.2 Winning Speech Videos

Victory Presentations: Individual winning speech videos for each character:

- Nelson Mandela: Inspirational victory speech in his characteristic style
- Michelle (Barbarella): Confident and engaging winning presentation
- Taylor Swift: Celebratory and charismatic victory speech

These videos utilize sync.so's lip-sync technology to create realistic facial animations synchronized with ElevenLabs-generated character voices, providing an immersive conclusion to successful debates.