Verbal Communication Skills Trainer

1. Introduction

Verbal communication is a critical skill in both personal and professional settings. Our project, the **Verbal Communication Skills Trainer**, leverages an **open-source large language model (LLM)** to help users improve their speaking skills through **interactive coaching, voice analysis, and structured feedback**. This report outlines the key features, technical decisions, optimizations, and future enhancements of the project.

2. Project Overview

The **Verbal Communication Skills Trainer** is a standalone application that enables users to enhance their communication skills through **chat-based interactions and voice input analysis**. The app provides real-time coaching, structured exercises, and personalized feedback on clarity, tone, fluency, and structure.

Key Features:

1. **Text-Based Coaching:** Users engage in a conversation with an AI coach that provides feedback on verbal clarity, coherence, and engagement.
2. **Voice Input Analysis:** Users can record and submit speeches, which are transcribed and evaluated for pacing, pronunciation, and structure.
3. **Skill-Training Modules:** Structured exercises focus on impromptu speaking, storytelling, and conflict resolution scenarios.
4. **Presentation Assessment:** Users receive detailed feedback on their speeches, including structural and linguistic improvements.
5. **Progress Tracking:** The application stores user progress in a local database for personalized learning.

3. Model Selection & Justification

For this project, we selected **Mistral-7B**, an open-source LLM, due to its balance between performance and efficiency. The model provides high-quality responses while maintaining a manageable memory footprint.

Reasons for Choosing Mistral-7B:

- **Lightweight compared to larger models** (e.g., LLaMA-13B), reducing computational costs.
- **Optimized for instruction-following**, ensuring relevant and insightful feedback.
- **Supports quantization**, making deployment more efficient.

Additionally, we integrated **OpenAI's Whisper** for **speech-to-text transcription**, offering:

- **High accuracy across different accents and noise levels.**
- **Multilingual support.**

- **Efficient audio chunk processing for real-time feedback.**

4. Application Features

4.1 Interactive Communication Modes

- **Chat-Based Interaction:** Users type responses, and the AI provides constructive feedback.
- **Voice-Based Interaction:** Users record their speech, which is transcribed and analyzed by the AI.

4.2 Skill Training Modules

1. **Impromptu Speaking:** Users receive a random topic and are assessed on fluency and logical structuring.
2. **Storytelling:** The AI evaluates storytelling elements like engagement and coherence.
3. **Conflict Resolution:** Simulated real-world scenarios test users' diplomacy and problem-solving abilities.

4.3 Presentation Assessment

- Users submit either **text scripts or voice recordings**.
- The AI evaluates **content structure, pacing, filler words, and persuasiveness**.
- A structured **feedback report** is generated, providing scores and actionable improvement suggestions.

5. Optimization Strategies

To ensure efficiency and smooth performance, the following optimizations were applied:

1. **Model Quantization:**
   o Used **8-bit quantization (via Bitsandbytes)** to reduce memory consumption while maintaining performance.
2. **Response Caching:**
   o Implemented caching using **functools.lru_cache** to speed up responses for frequently used prompts.
3. **Efficient Speech Processing:**
   o Whisper **processes audio in chunks**, reducing processing latency and improving real-time performance.
4. **Database Optimization:**
   o **SQLite was used for tracking user progress**, ensuring fast and lightweight storage.

6. Challenges & Solutions

Challenge 1: Latency Issues

- **Problem:** LLM inference is computationally expensive, causing delays.

- **Solution: Preloading the model** and **reusing it across sessions** reduced initialization delays.

Challenge 2: Speech Recognition Accuracy

- **Problem:** Background noise affected transcription quality.
- **Solution:** Applied **audio preprocessing** techniques, such as noise reduction and normalization.

Challenge 3: Memory Constraints

- **Problem:** Large models can cause **Out-of-Memory (OOM) errors**.
- **Solution:** Implemented **batch inference, quantization, and memory-efficient model loading**.

7. Future Improvements

While the project is fully functional, the following enhancements could further improve user experience:

1. **Fine-Tuning the LLM** on a dataset specific to public speaking and communication for more accurate coaching.
2. **Integrating a Text-to-Speech (TTS) system** to provide spoken feedback.
3. **Developing an API-based version** for integration into third-party platforms.
4. **Expanding Multi-User Support**, allowing different users to track their unique progress.

8. Conclusion

The **Verbal Communication Skills Trainer** is an AI-powered tool designed to help users improve their speaking abilities through **interactive coaching and structured feedback**. By leveraging **open-source LLMs and speech-processing models**, the application provides an effective, scalable, and accessible solution for communication training.

Through **text and voice-based interactions**, users receive **real-time insights** into their speaking patterns and personalized feedback to enhance their verbal skills. The **project demonstrates the potential of AI-driven education and personal development applications**, paving the way for further advancements in AI-assisted training programs.

Thank you for reading!