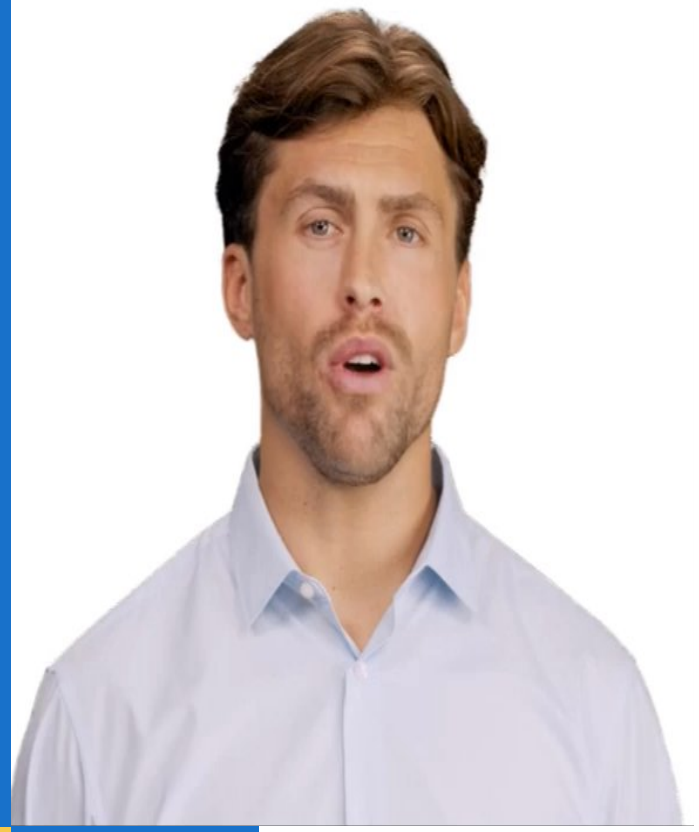


Avatar Lab

Building Smart & Realistic AI Avatars

G331
Nikhilesh Nilagiri

Project Overview

Avatar Lab is a system that converts text into realistic talking head videos in seconds.

Key Innovations :

- Integrates DiffDub for talking head animation and SMALL-E for speech synthesis
- Uses deep learning models to sync speech with facial expressions and emotions

🎯 End Goal : To build an application that generates realistic, interactive AI avatars for digital experiences

Business Problem

Current Challenges :

- AI avatars lack realistic lip-sync and expressions
- Existing solutions are costly and complex.
- Lack of personalization making avatars feel generic

Our Solution :

- ✓ Improves realism with precise lip-syncing and expressions
- ✓ Customizable avatars for personalized, engaging experiences
- ✓ User-friendly platform, no technical skills required

Typical Users

- Virtual Assistants 
- Game Developers 
- Educators 
- Content Creators 

User Roles & Interaction

- Roles Involved :

End Users : Input text, customize avatars/voice, and generate video

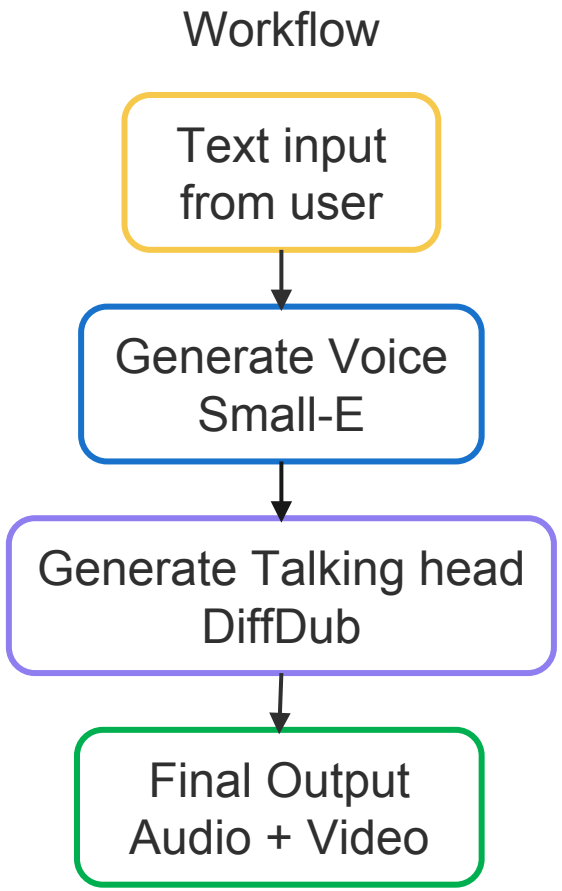
Developers/Admins : Fine-tune AI models and optimize pipelines

- Interaction Flow :

Input : Text script + parameters (e.g. voice tone)

Output : Avatar video with lip-sync, expressions, and voice modulation.

Interface: Web-based platform with a simple text editor and preview panel



Datasets & Preprocessing

Models:

- DiffDub: HDTF
- Small-E: LibriTTS, LibriTTS-R, MLS-English (10k hours), GigaSpeech XL

Preprocessing:

- Verify: Remove misaligned audio-video pairs.
- Normalize: Resample audio to 16kHz, standardize video resolution.
- Validate: Ensure proper audio-video alignment.

Tech Stack

Frontend : Next.js

Backend : Flask (API for model integration)

AI Models : DiffDub + SMALL-E

Database : MongoDB, PostgreSQL

Thank You