AVATAR LABS
BUILDING SMART AND REALISTIC AVATARS

G331

V VISHAL RAJ

23BD1A05BU

SCOPE OF THE PROJECT

■ **WHAT DOES THIS PROJECT AIM TO ACHIEVE?**
AI-driven portrait animation has significantly advanced, enhancing visual storytelling.

■ **AI-DRIVEN SPEECH-TO-VIDEO**
Convert audio/text into realistic talking head animations.
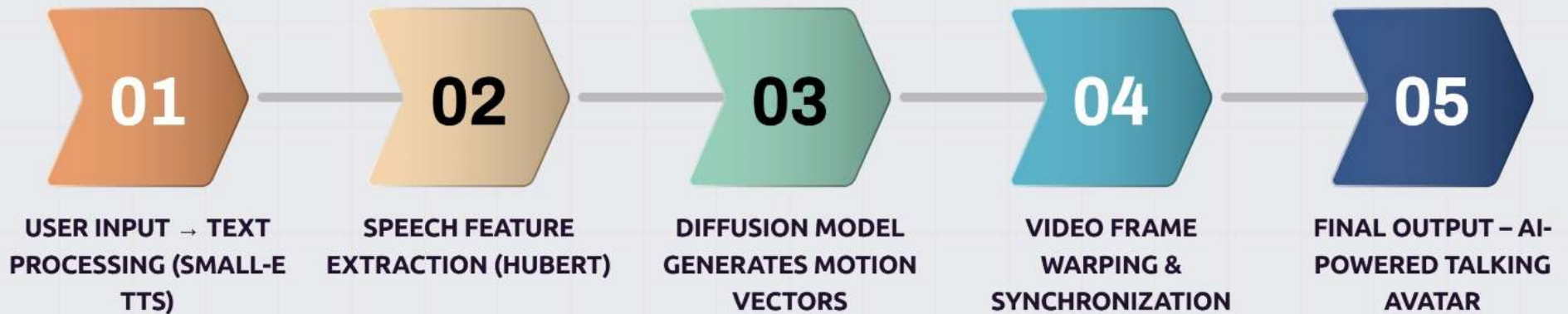
■ **SEAMLESS INTEGRATION**
Combine Small-e TTS (audio generation) with DiffDub (motion synthesis).

■ **SCALABLE & DEPLOYABLE**
Build a production-ready system for real-world use.

UNIFIED PIPELINE: SMALL-E TTS + DIFFDUB

Flow of Data Between Speech Generation & Motion Synthesis

01

USER INPUT → TEXT PROCESSING (SMALL-E TTS)

02

SPEECH FEATURE EXTRACTION (HUBERT)

03

DIFFUSION MODEL GENERATES MOTION VECTORS

04

VIDEO FRAME WARPING & SYNCHRONIZATION

05

FINAL OUTPUT – AI-POWERED TALKING AVATAR

INNOVATIVE BUSINESS MODEL & TARGET USERS

Exploring Paths to Profitability Through Innovation

SUBSCRIPTION PLANS

Offering diverse tiers: Free with watermarked avatars, Pro for HD video, and Enterprise for custom integrations.

FREE TIER

Allows users to create avatars with watermarks, promoting initial engagement and trial.

PRO TIER

Includes full HD video capabilities and real-time API access for advanced users.

ENTERPRISE TIER

Custom integration solutions designed specifically for studios, catering to large-scale needs.

COMPETITIVE EDGE

Combines DiffDub and Small-e TTS technology for superior lip-sync and expressiveness.

EFFICIENCY

Our solution is faster than traditional animation tools, saving time and increasing productivity.

CONTENT CREATORS

Generate lip-synced videos from text/audio.

MARKETING & ADVERTISEMENT

AI-generated avatars for branding.

EDUCATION & E-LEARNING

Virtual tutors with natural speech & expressions.

FILMMAKERS & ANIMATORS

AI-powered dubbing & animation.

TECH STACK

- **DIFFDUB - VIDEO GENERATION PIPELINE**

- **TEXT-TO-SPEECH GENERATION - SMALL-E TTS MODEL**

- **NEXT.JS FOR FRONTEND**

- **BACKEND - FLASK API**

- **DATABASE - POSTGRESQL, MONGODB**