AVATAR LAB
CONSTRUCTING REALISTIC & AI-POWERED AVATARS

G331
AKULA.YASHWANTH

23BD1A05A2

INTRODUCTION & PROBLEM STATEMENT

What is Avatar Lab?

> AI-powered tool that converts text into realistic talking head videos.

> Uses deep learning for facial animation and speech synthesis.

Cutting-Edge Developments

Leverages Diff Dub for high-fidelity facial motion synthesis and SMALL-E for neural speech generation, ensuring precise lip-sync and expressive vocal output.

End Goal :

To build an application that generates realistic, interactive AI avatars for digital experiences

MARKET CHALLENGE

> Video production is expensive and time-intensive.

> Scalable, high-quality video creation remains a challenge.

> Limited AI solutions for realistic, customizable, and cost-effective avatars.

TYPICAL USERS

> Content Creators

> Educators

> Virtual Assistants & Game Developers

OUR SOLUTION

> Improves realism with precise lip-syncing and expressions

> Customizable avatars for personalized, engaging experiences

> Enhanced emotion-based facial expressions.
> User-friendly platform, no technical skills required.

FUTURE SCOPE

> Multi-language support for global accessibility.

> Real-time avatar streaming for live interactions.

USER ROLES & INTERACTION

User Roles

General Users: Input text, customize avatar appearance & voice, generate videos.

Developers: Fine-tune AI models and optimize system performance.

User Interaction

Input: Text script + optional emotion/voice parameters.

Processing: AI models generate lip-sync + speech synthesis.

Output: Downloadable AI-generated talking head videos.

Interface: Web-based platform with real-time preview.

TECH STACK & AI PIPELINE

Frontend: Next.js (User Interface)
Backend: Flask (API & Model Integration)
AI Models:

- **DiffDub:** Advanced facial animation and lip-sync.

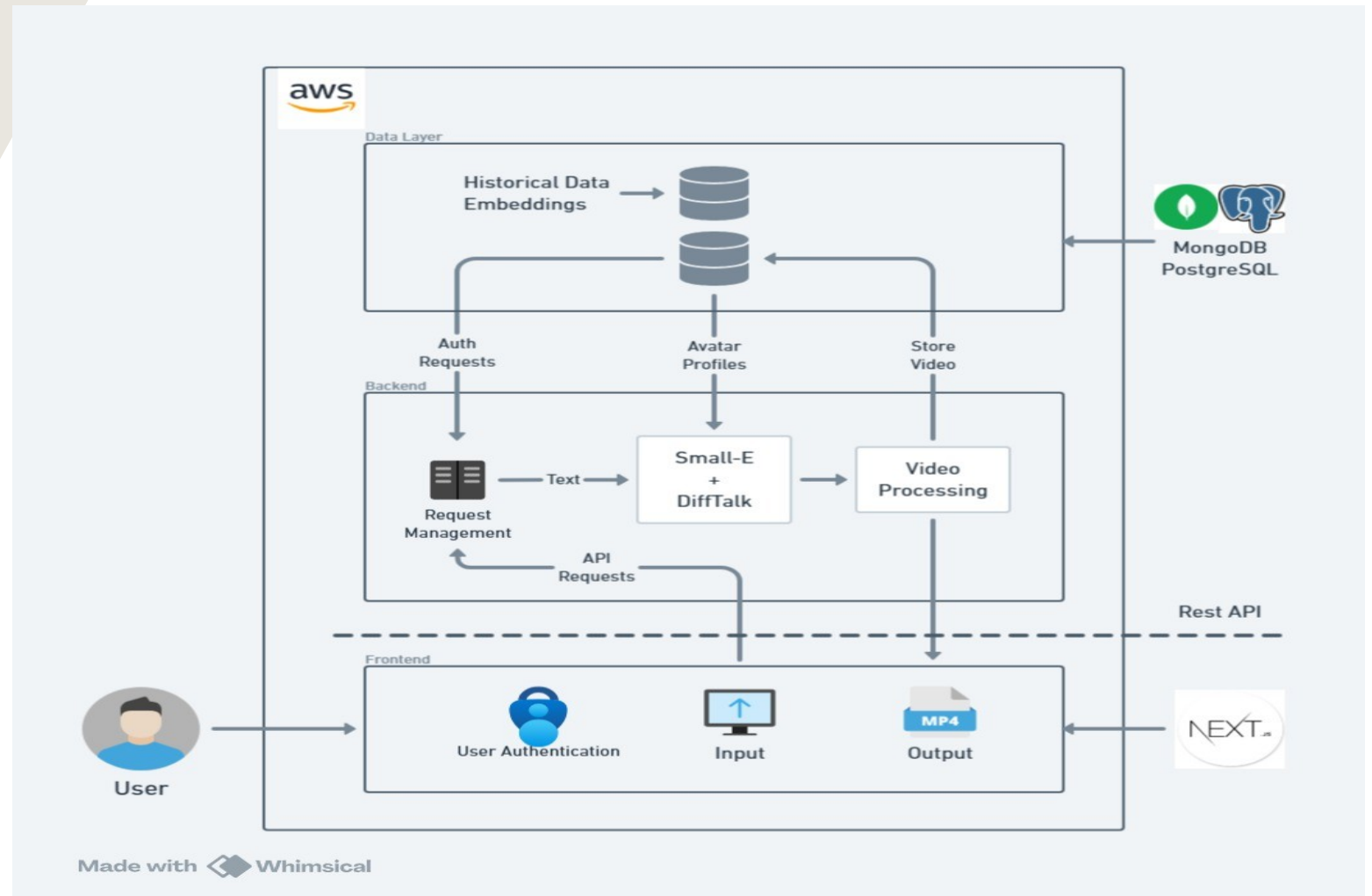- **SMALL-E:** High-quality text-to-speech synthesis.
Database: MongoDB (User Data) + PostgreSQL (Structured Metadata)
Cloud Deployment: AWS/GCP for scalable inference.

Preprocessing Steps:

- Remove misaligned audio-video pairs.

- Standardize audio (16kHz) and video resolution.

- Validate synchronization between speech and lip movements.

ARCHITECTURE DIAGRAM

THANK YOU