

برز نتیجه نتیجه نتیجه نتیجه

DEEPMODEL AUDIOS



Outline

Motivation & Introduction

Competitors

Functional & Non functional requirements

AI models

UI /UX

Comparisons

architecture

Demo

Future work & Questions

Disclaimer ! ! :

Any voices used in this presentation are fake and used for educational purposes

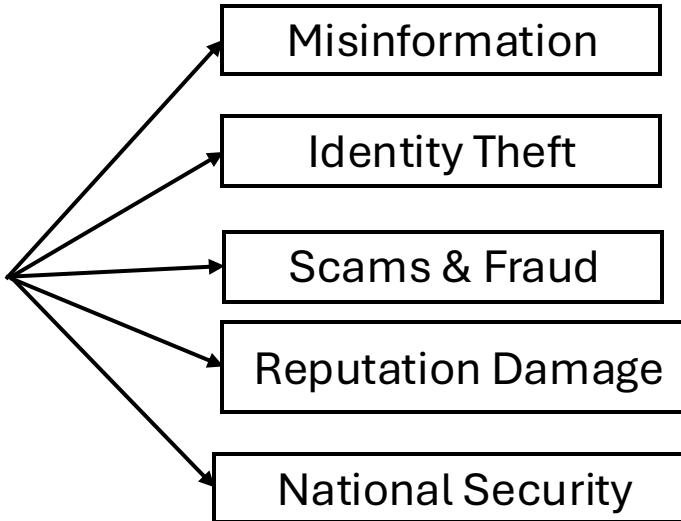
Motivation

Why would someone be interested in our project



What if this audio was trending by social media bombing ?

Fake Audio can provide



These are problems we must address



The Hidden Epidemic

While you've been worried about fake images and videos, audio deepfakes have been quietly infiltrating:

-  **Your phone calls** – Is that really your bank?
-  **Voice messages** – Did your friend actually send that?
-  **News broadcasts** – Are those real celebrity quotes?
-  **Political speeches** – Which politician really said what

**The technology is here.
The defenses are not.**



Introduction

What is deep fake audio?

Deepfake audio is the ability to mimic human-like voice using AI and software, making it hard to distinguish from real voices. It can be used for both legal and illegal purposes.

Problems

- Audio deepfake detection is overwhelmingly English-focused:

Most studies focus on detecting audio deepfake in the English language, not paying much attention to the most spoken languages like Chinese and Spanish, as well as Hindi and Arabic.

- Arabic deepfake audio research is still nascent:

A 2023 IEEE Access study noted that “Arabic speech has not yet been explored with synthetic fake audio,” focusing mainly on imitation attacks and with minimal dialect diversity

In late 2024, a real-time Arabic deepfake detection paper from Egypt-Japan University reported promising results—but overall, the field remains in its infancy regionally

- Specific work on Egyptian-dialect fake audio detection is extremely limited, with only one small GitHub project as an isolated attempt

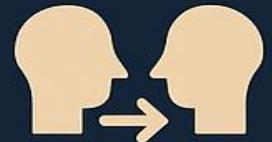
<https://github.com/SaifEKhaled/DetectingFakeAudioOfEgyptianDialect>

Still a python project not an app or website

What we'll offer



Detection



Speech to speech



Text to speech



Interactive Game

competitors

Websites

Detection only

Hi Ahmed Hossam!

Be Protected Against AI Cloned Voices and Deepfakes

We offer an AI tool that can identify if an audio is real or AI generated

- ✓ Integrated background noise remover
- ✓ Integrated background music remover



Select an audio file

[Browse](#)

[Detect Now](#)

OR

[Download our Browser Extension](#)



X

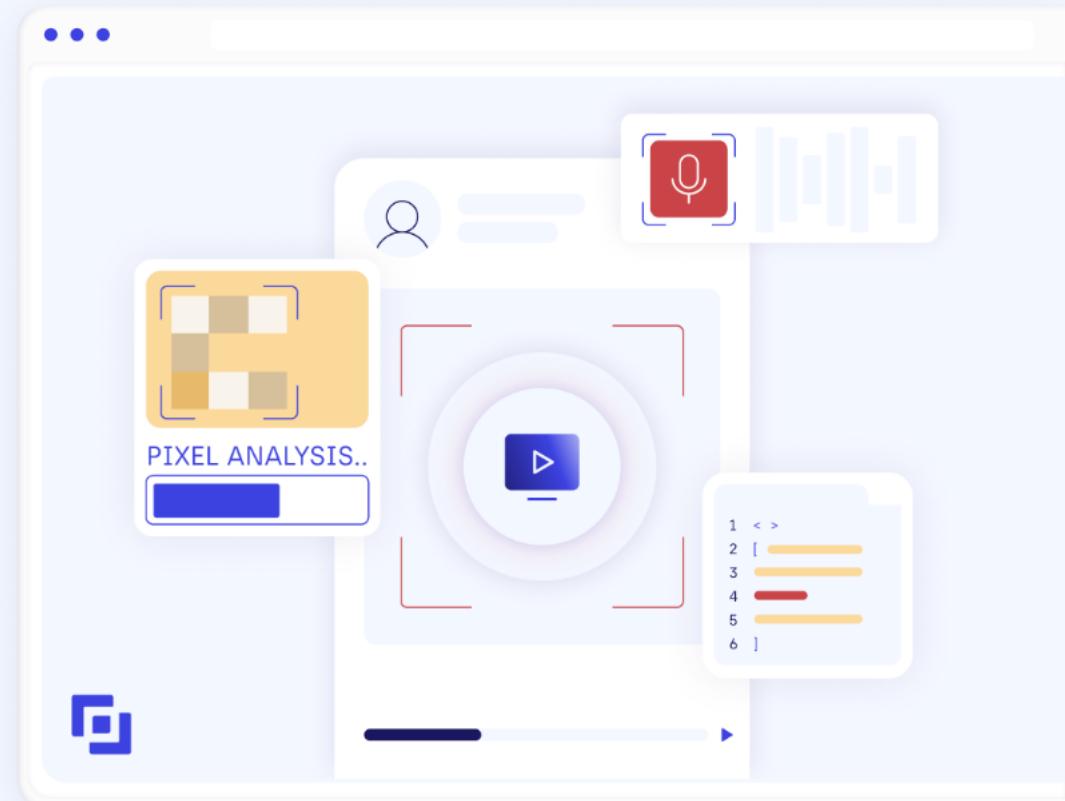
Use it to scan voices on Youtube, WhatsApp, Tiktok, Zoom

Got any questions? I'm happy to help.



Deepfake Detection

Experience a frictionless and user-friendly interface designed for everyone. Simply drag and drop your files and get a deepfake media verification within seconds. Our advanced system uses a multilayer approach, examining pixels, file structures, and voice patterns to deliver the most comprehensive assessment effortlessly.

[Get started](#)[Talk to an expert ➔](#)

Speech to Speech only

Supertone play



Home



Projects



Rewards



Voice

Cloning



API



Feedback

INVITE AND GET AIRPODS MAX!

Weekly draw for referrers & invitees - Check our Instagram & Discord for details

Join now →

Create the content you want freely
with Supertone Play.

Share

English ▾

My Own Cloning Voice



What voice does the character I made with my voice sound like?

Create Cloning Voice >

Supertone Play Voice

HappyUnfriendlyAngryCuriousNormalSuspiciousLovingSadAmusedActive

Text to speech speech and
speech to speech



Localize content across 29 languages with AI dubbing

Translate audio and video while preserving the emotion, timing, tone and unique characteristics of each speaker

[START DUBBING FREE](#)[EXPLORE PLANS](#)

DUBBING STUDIO

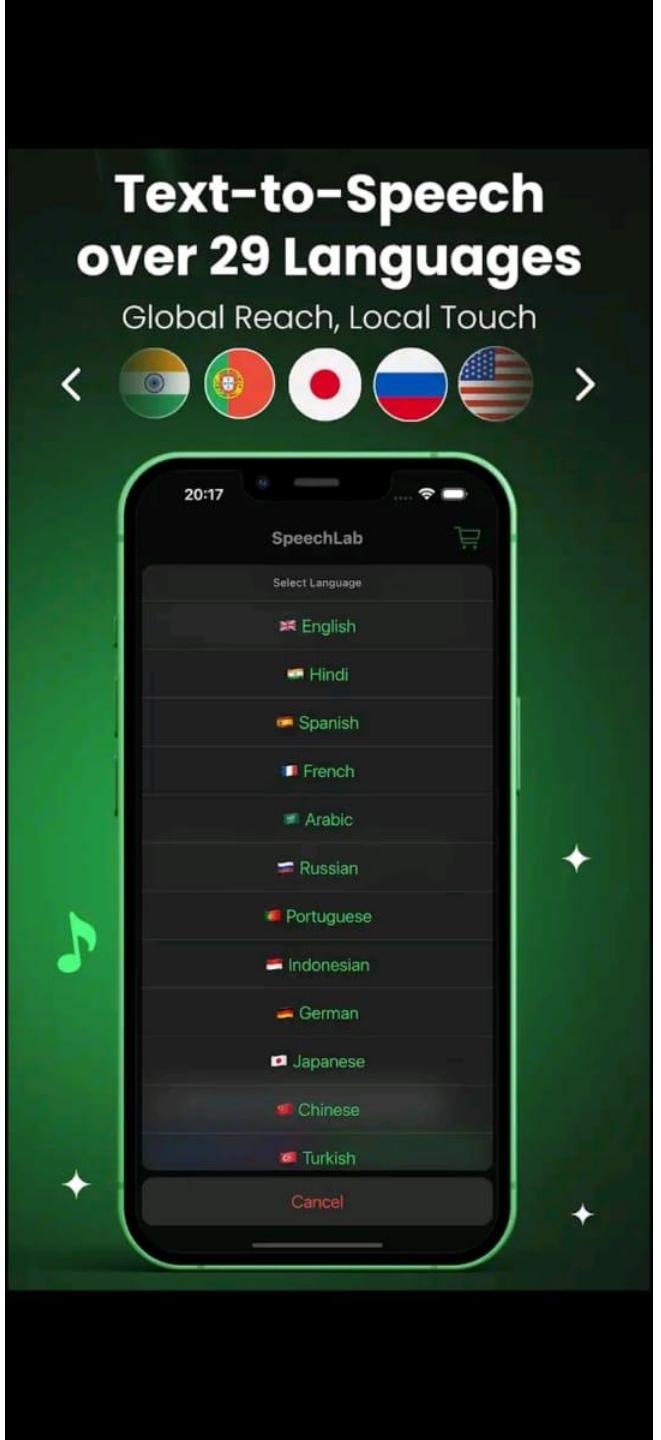
Speaker 1
But that connection is really kind of what I'm trying to describe, so.
Transcribe Audio

Speaker 2
— Yeah, you could send, like, a fax to someone, similar to—
Transcribe Audio

Voice chat

Powered by ElevenLabs Conversational AI

Mobile Application



6:39

Wi-Fi Signal Strength

AI Voice Changer & Dubbing

stringcode ou

In-app purchases

4.1★
3K reviews ⓘ

25 MB

12+

Install

Instant & Crystal Clear Voices

Over 300 Voices Await

Text-to-Speech over 29 Languages

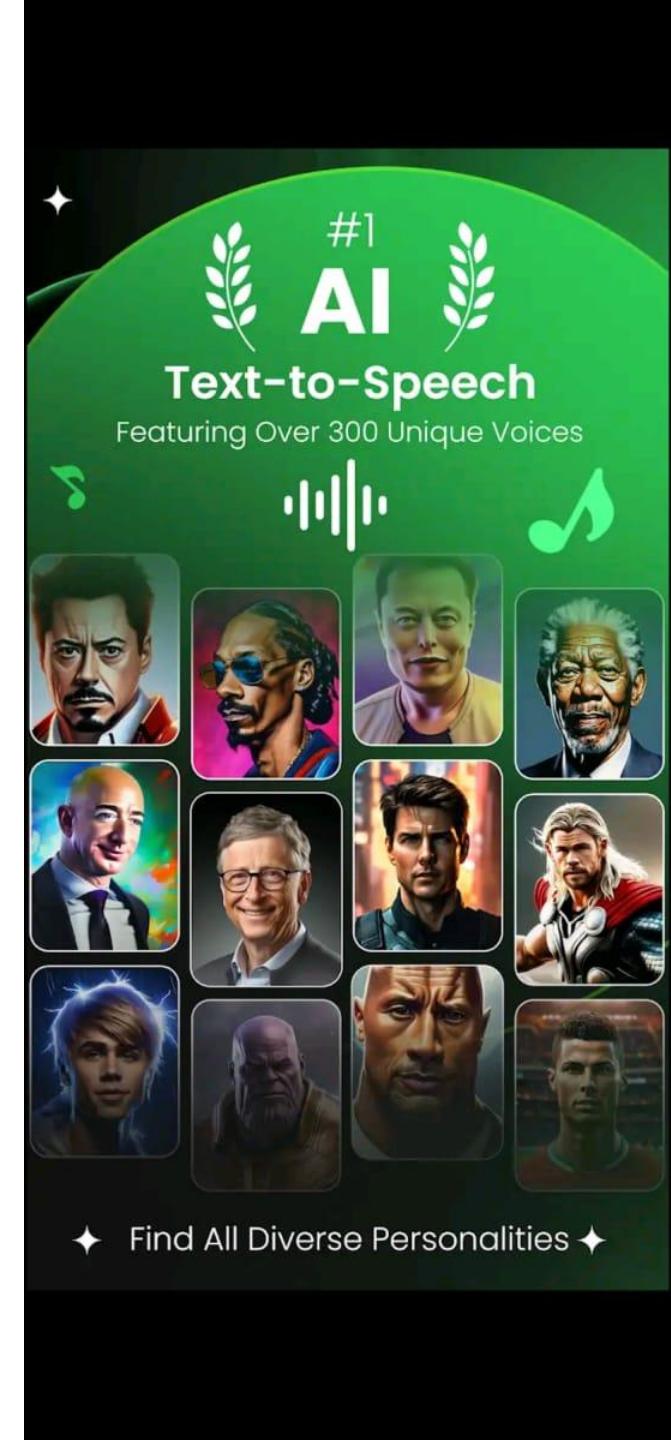
About this app

Celebrity voices and high-quality Text to Speech, powered by ElevenLabs and GPT.

Productivity Entertainment

Games Apps Search Books

A screenshot of the Google Play Store page for the AI Voice Changer & Dubbing app. The page shows a rating of 4.1 stars from 3K reviews, a file size of 25 MB, and an age rating of 12+. A large blue "Install" button is prominent. Below the button, there are four screenshots of the app's interface. A section titled "About this app" provides a brief description of the app's features, mentioning celebrity voices and high-quality Text to Speech powered by ElevenLabs and GPT. The bottom navigation bar includes tabs for Productivity, Entertainment, Games, Apps, Search (which is selected), and Books.



6:41



AI Voice Generator & AI Cover

AppLab Kamil Piekarz

In-app purchases

4.4★

8K reviews

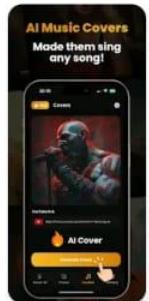


23 MB

3+

Rated for 3

Install



About this app



Voice AI & AI Voice Generator Funny Deepfake
Speech AI & AI Voice clone!

Productivity

Entertainment



Games



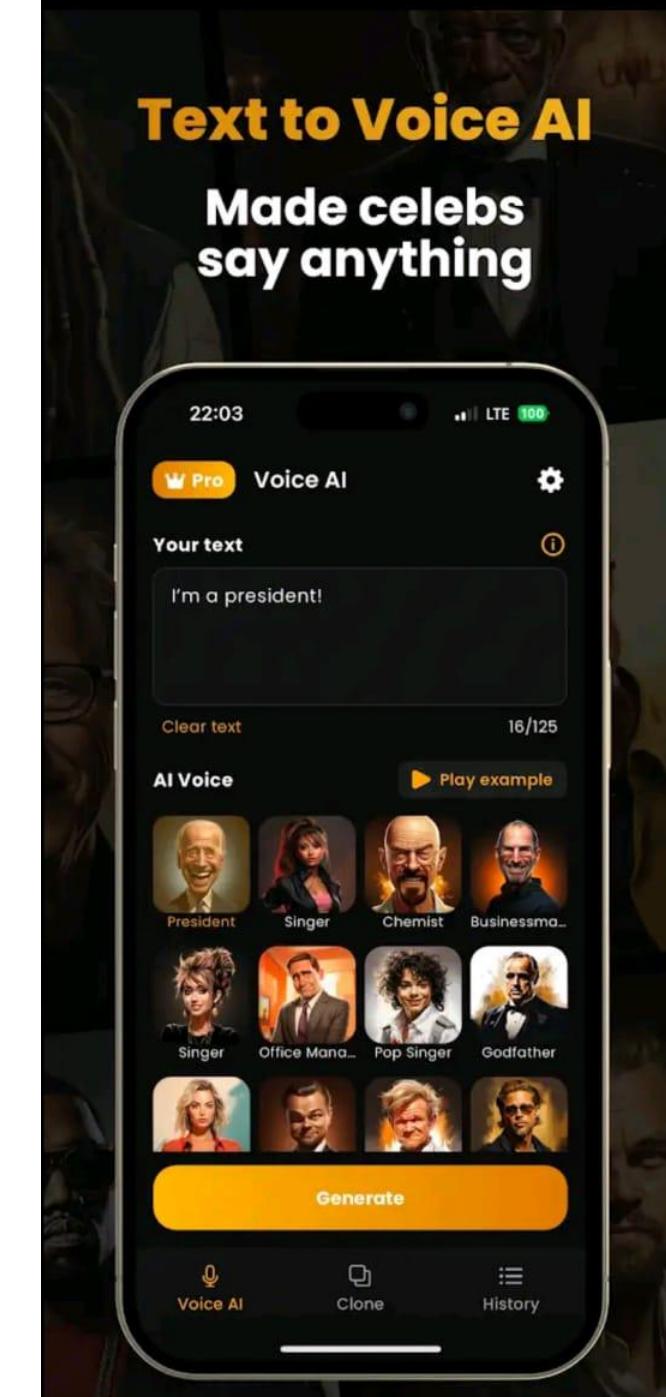
Apps



Search



Books



Comparison

	Detection	Speech to Speech	Text to Speech	Games	Arabic
Ai voice detector (website) <u>Link</u>	✓	X	X	X	X
Sensity(website) <u>link</u>	✓	X	X	X	X
Play Ai(website) <u>link</u>	✓	✓	✓	X	X
NaturalReader(website) <u>link</u>	X	X	✓	X	X
Ai voice changer(APP)	X	X	✓	X	X
Ai voice generator (APP)	X	✓	✓	X	X
DeepFake Audios (website and APP)	✓	✓	✓	✓	✓

Our Website and Mobile Application

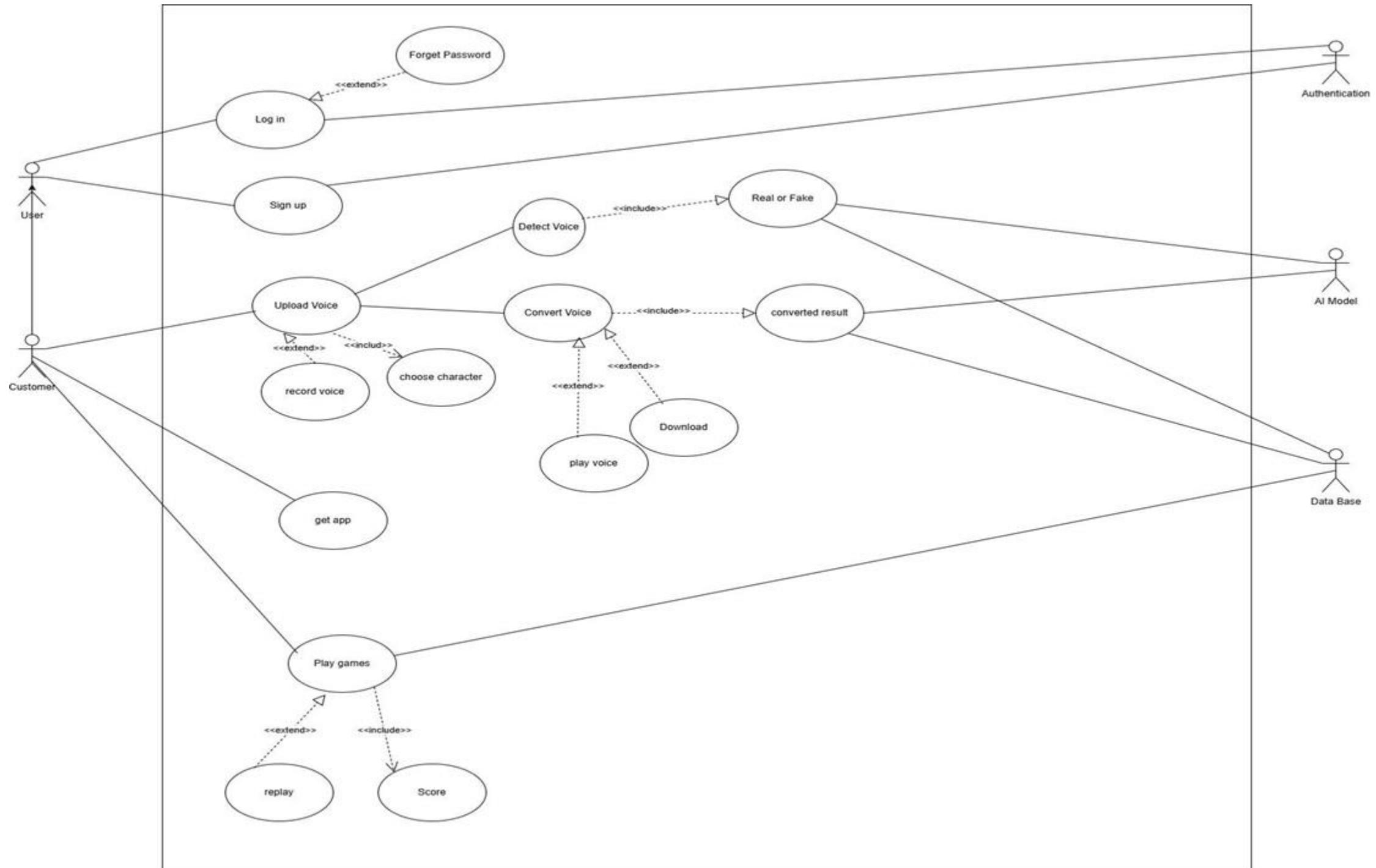
- Detection
- Speech to Speech
- Text to Speech
- Interactive Games

function requirements (Use case diagram (eldok))
& non function requirements (3omda)

Function requirements

- Detection
- Speech to speech
- Text to speech
- Interactive games

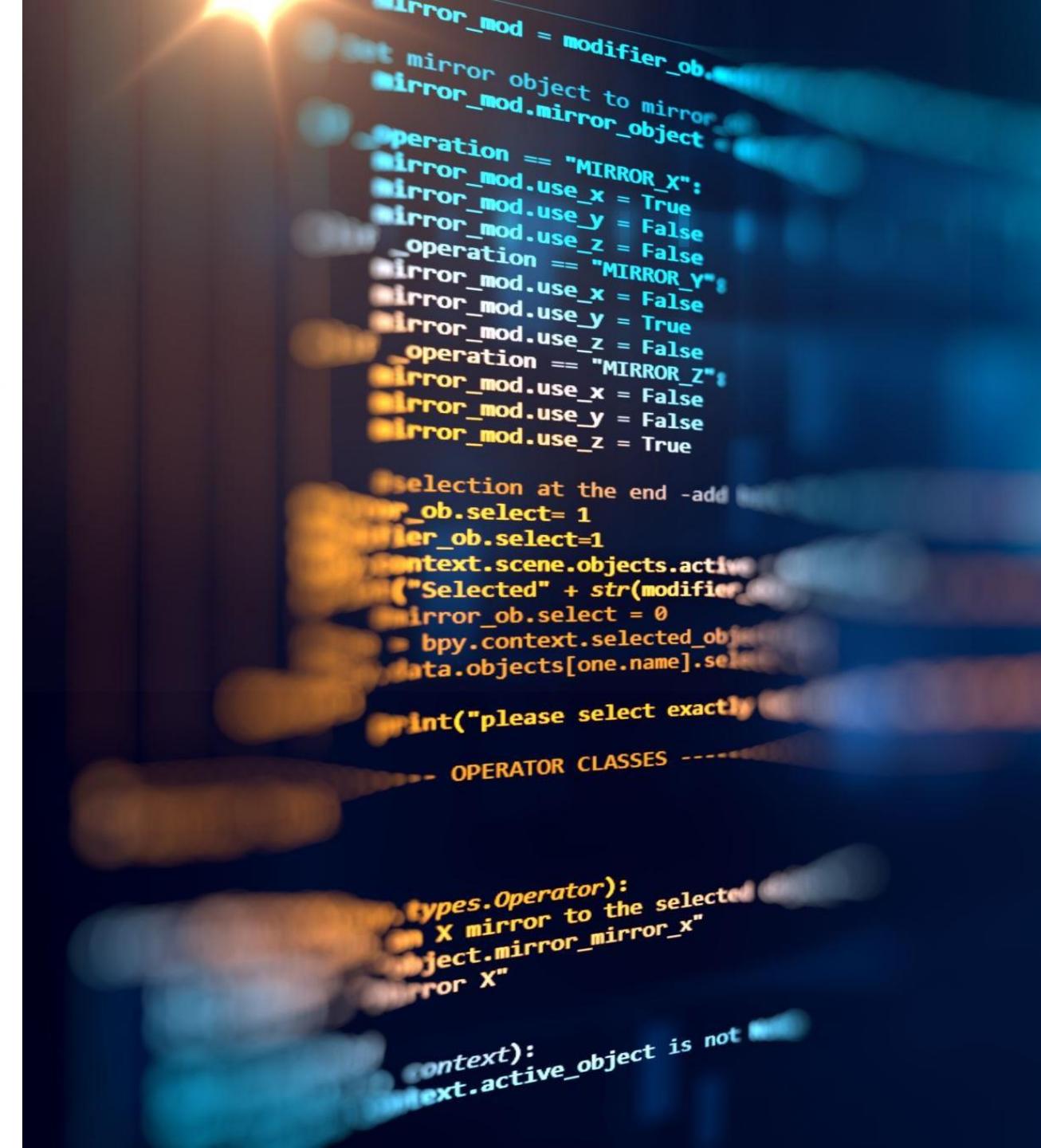
Use Case Diagram



Non function requirements (3omda)

NON-FUNCTIONAL REQUIREMENTS

- Responsive, & Cross-device
- Acceptable App & Database Performance
- Authentication & Authorization
- Security & Protection Against Common Attacks
- Database Scalability
- Accessibility & Localization



AI

I.Detection

Related work

- 1. **Fake Song Generation & Detection (FSD)**— 6 Sep 2023
- 2. **CLAD: Contrastive Learning for Robust Deepfake Detection**— 24 Apr 2024
- 3. **Codecfake Dataset & CSAM Countermeasure**— 15 May 2024

Our AI detection model

Datasets Used :

- ASV Spoof 2019
 - Challenge Tracks:
 - Logical Access (LA): Focuses on synthetic speech and voice conversion attacks.
 - Physical Access (PA): Targets replay attacks using audio recordings.
- Dataset Composition:
 - Speech Data: Genuine and spoofed speech samples.
 - Training/Validation/Testing Splits: Pre-divided for consistent benchmarking.

First architecture

Wav2vec

Encoder

Transformer Layers

Fine-tuning

Attention
added=0.1

Hidden
dropout=0.1

Num
Labels=2

Second architecture

- Detailed CNN-Based Model Using MFCC Features
- What is MFCC?
- Mel-Frequency Cepstral Coefficients (MFCC)
- Audio Input: Raw waveforms are preprocessed into MFCC spectrograms (e.g., $1 \times 40 \times T$, where T is time frames).
- CNN Layers
- Convolutional Filters
- Activation Function
- Pooling
- Squeeze-and-Excitation (SE) Blocks

Hyperparameters

- First architecture
 - Epochs=5
 - Learning rate=5e-5
 - Weight decay=0.01
 - Warmup steps=500
- Second architecture
 - Epochs=10
 - Learning rate=1e-3
 - Weight decay=0.01
 - Warmup steps=1000

Ensembling both architectures

- Output-Level Ensembling:
 - Perform late fusion by aggregating the predictions from both models.
 - Techniques include:
 - Averaging: Compute the mean probability scores for each class.
 - Weighted Averaging: Assign weights to each model based on its validation performance.
 - Voting: Use majority or weighted voting for class decisions.

Benefits of Ensembling Both Architectures

- Improved Performance
- Better Generalization
- More Robustness
- Reduced Model Weaknesses
- Higher Confidence in Predictions

Result

- Best result on validation data of ASV spoof 2019 LA part
- API Deployment

Validation Loss: 0.0042 | Validation Accuracy: 99.91%

II.TTS

Why Voice Cloning?

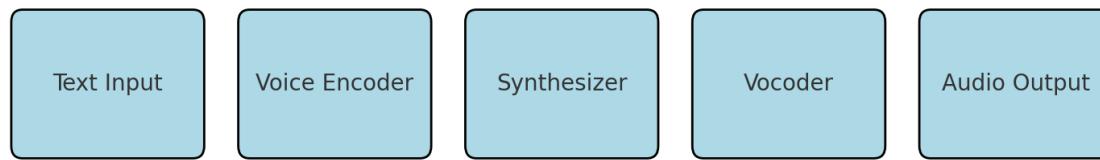
- "Traditional TTS systems lack personalized voice characteristics"
- "Need for customizable character voices in:
 - Educational content
 - Audiobook narration
 - Entertainment media"

Best Model

- Zero-Shot Multilingual Evaluation
- From the evaluation paper (arXiv Jun 7, 2024) comparing XTTS versus other state-of-the-art multilingual TTS models:

Language	Model	CER (↓)	SECS (↑)
Arabic (ar)	YourTTS (Exp.2)	11.17	0.4400
	XTTS (Exp.3)	3.3503	0.5007

XTTS v2 Architecture



How XTTs v2 Works

- **Voice Encoder:**
 - Extracts speaker embeddings from 3-10 sec reference audio
- **Tacotron Synthesizer:**
 - Generates mel-spectrograms from text + embeddings
- **Waveform Decoder:**
 - Converts spectrograms to raw audio (16kHz)

Parameters

- `temperature=0.3,`
- `length_penalty=1.3,`
- `split_sentences=False`

Collecting data

- Four Characters
 - Amr Adib
 - Medhat Shalaby
 - Anwar Elsadat
 - Ibrahim Faiq
- Cleaning audios from noise
- 30 audios each audio 30 seconds

End-to-End Processing

- User selects character (ID 1-4)
- System loads corresponding reference audios
- Text preprocessing (Arabic normalization)
- XTTS generates voice-matched audio
- Output delivered as downloadable WAV

Here Example from the model

- Medhat Shalaby



III.Voice Conversion

AI model Speech-to-speech (SEED-VC)

We adopt **SEED-VC** (Plachtaa, 2024) for zero-shot VC and SVC with real-time support. Key modules and configurations

Why Seed VC?

Zero-Shot Voice Conversion Performance

According to the **Seed-VC** evaluation on LibriTTS test-clean (100 utterances, 8 unseen target speakers), compared to models like OpenVoice and CosyVoice:

Model	SECS↑	WER↓	CER↓	SIG↑	BAK↑	OVRL↑
Ground Truth	1.0000	8.02 %	1.57 %	—	—	—
OpenVoice	0.7547	15.46 %	4.73 %	3.56	4.02	3.27
CosyVoice	0.8440	18.98 %	7.29 %	3.51	4.02	3.21
Seed-VC (full reference – ours)	0.8676	11.99 %	2.92 %	3.42	3.97	3.11

Architecture

Component	V1 (Real-Time)	V1 (Offline/SVC)	V2 (Enhanced VC)
Content Encoder	XLSR-large (multilingual)	Whisper-small	Hubert + BSQ-VAE-small
Backbone	U-ViT diffusion denoiser	U-ViT diffusion denoiser	U-ViT diffusion denoiser with Classifier-Free Guidance (CFG)
Vocoder	HIFT (tiny, 384-dim)	BigVGAN	BigVGAN (CFM) + Autoregressive (AR) module
Sample Rate	22 050 Hz	44 100 Hz (for singing)	22 050 Hz
Model Size	25 M params	98 M params	157 M params
Inference Controls	--diffusion-steps , --length-adjust , --inference-cfg-rate	+ --f0-condition , --semi-tone-shift , --auto-f0-adjust	Same as V1 (Real-Time) with CFG support

Content Encoder Details:

- **XLSR-large**: A cross-lingual wav2vec 2.0 pretrained on 128 languages. Provides robust phonetic features even on low-data speakers.
- **Whisper-small/base**: Transformer encoder trained on 680k hours of multilingual data; excels at noisy and singing inputs.
- **Hubert-BSQVAE**: Discrete latent representations via vector quantization, encouraging cleaner disentanglement of speaker identity.

Diffusion Conversion Backbone:

- Employs a U-shaped Vision Transformer (**U-ViT**) as the denoiser network.
- Denoising steps configurable (4–10 for low-latency; 30–50 for production-quality).
- **Classifier-Free Guidance (CFG)**: Balances content fidelity against timbre divergence via inference CFG rate (0.0–1.0).

Vocoder

HIFT: A lightweight GAN-based vocoder optimized for 22 kHz real-time VC (latent dimension 384, 9 layers).

BigVGAN: High-fidelity GAN for singing voice at 44 kHz, supports conditioning on pitch and timbre for nuanced SVC outputs.

CFM+AR (V2): Augments BigVGAN with a content-fidelity module and an autoregressive refinement network for accent/emotion control.

Dataset Speaker List

For fine-tuning, we collected recordings from the following speakers:

- Amr Adib
- Ahmed Abu Zaid
- Abu Hafeeza
- Mona El-Shazly
- Lamis
- El-Usta Abdo
- Marwan Serry
- Abu Trika
- El-Gayar

Parameters

- --diffusion-steps 40
- --length-adjust 1.0
- --inference-cfg-rate 1.0
- For music We separate music using demucs library

Ngrok Deployment

Here examples from the model

- Normal audio



- With music



UI /UX (eldok)

WEBSITE(FIGMA)

SIGN IN

AUDIO DEEP FAKE



Sign IN

Email Address*

Enter your Email

Password

Enter your password

Remember Me

[Forgot Password?](#)

[Continue](#)

[Don't have an Account? Sign Up](#)

QR Sign In With QR



Use The [Steam Mobile App](#) To Sign
In Via QR Code

SIGN UP

AUDIO DEEP FAKE



Create Your Account

First Name _____

First Name

Second Name _____

Second Name

Email _____

Email

Password _____

password



Country _____

Egypt



"I Am 13 Years Of Age Or Older And Agree To The
Terms Of The Steam Subscriber Agreement And
The Valve Privacy Policy."

Continue

Sign Up with Google

Already Have An Account? [Sign In](#)

HOME PAGE

Experience The Future Of Audio Technology

With Advanced AI, Audio Deep Fake Allows You To Explore Digital Audio's Potential. Identify AI-Generated Voices, Create New Ones, Or Play Interactive Games—Endless Possibilities Await!

Get It Now



Detect Audio

Analyze Whether The Audio Is Real Or AI-Generated In Seconds. Perfect For Ensuring Authenticity In Media Or Just For Fun.



Speech To Speech

AI Technology Allows For Highly Realistic Voice Conversions, Opening Up New Possibilities In Entertainment, Content Creation, And Beyond.



Interactive Games

Challenge Yourself With Our Fun Game Mode To Guess If The Audio Is Human Or AI-Made, And Improve Your Skills While Having Fun.



[Watch How Detect Works](#)



Detection Page

From Audio File

"Upload An Audio File (Max File Size: 50MB) (Wav, Mp3, Flac & Ogg Supported). The AI Will Mimic Your Speaking Style."

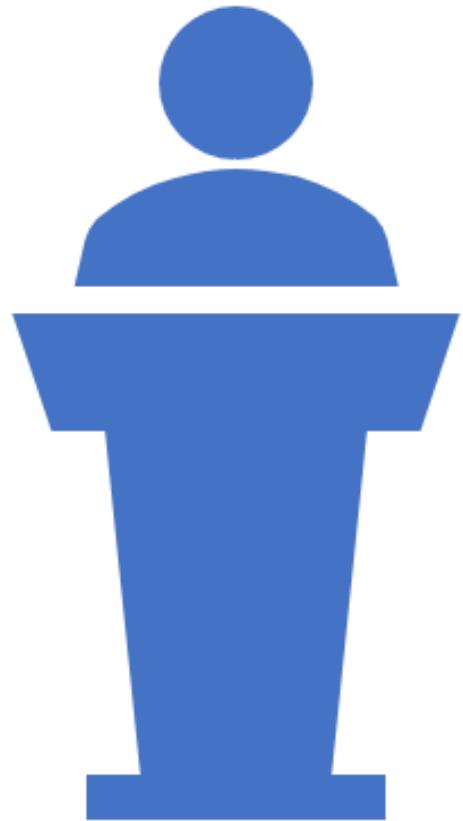
Detect

Help

Results

Clear All

- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Fake!
- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Fake!
- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Real!
- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Real!
- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Real!
- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Fake!
- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Fake!
- "Sherien Mashrabtesh Mn Nebla Clip Mp3" December 04,2023 11:27 AM Fake!



**Speech to Speech and
choose a character**

AUDIO DEEP FAKE



Detect Audio

Speech To Speech

Interactive Games



William(Whispering) ▾



From Live Audio Recording

"Record Yourself Directly From Your Browser.
The AI Voice You Select Will Imitate Your
Speaking Style, Including Cadence, Vocal
Range, Speed, And intonation."



From Audio File

"Upload An Audio File
The AI Will Mimic Your Speaking Style."

(Max File Size: 50MB)
(Wav, Mp3, Flac & Ogg Supported).

Download



DeepFake

AUDIO DEEP FAKE



Detect Audio

Speech To Speech

Interactive Games



William(Whispering) ▾

Create Your Voice Character



Donald Trump



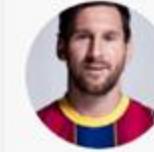
Joe Biden



Mr Beast



Obama



Leo Messi



Elon Musk



Will Smith



Morgan Freeman



Denzel



Elsadat

Help



Speech to Speech (Live Recording)

AUDIO DEEP FAKE



Detect Audio

Speech To Speech

Interactive Games



William(Whispering) ▾

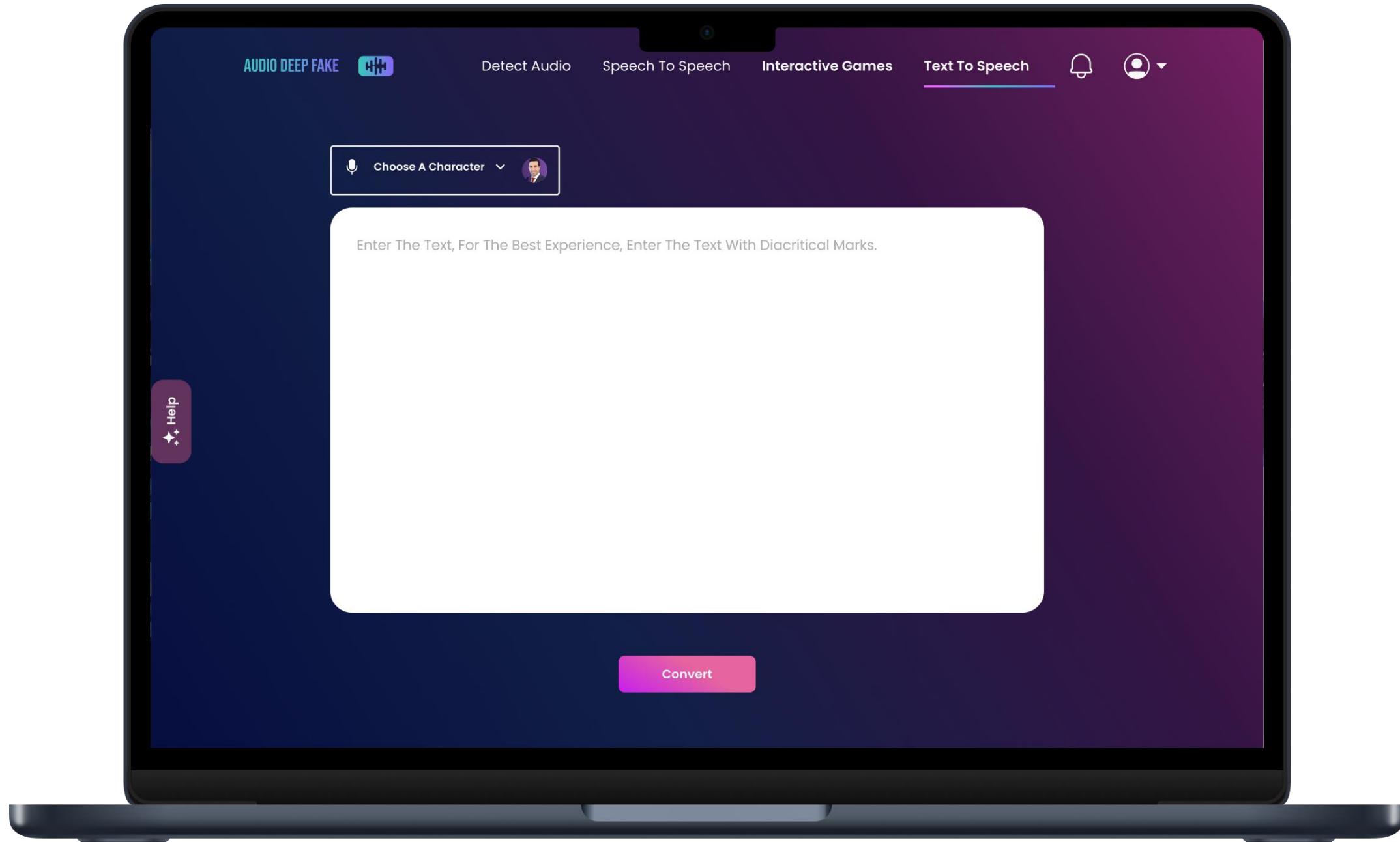
Live Audio Recording

Record

Submit

Help

Text To Speech



AUDIO DEEP FAKE



Detect Audio

Speech To Speech

Interactive Games

Text To Speech



Choose A Character

Create Your Voice Character

X



Amr Adeeb



Medhat Shalaby



Ibrahim Faik



Abo Trika



Elsadat



Amr Adeeb



Medhat Shalaby



Ibrahim Faik



Abo Trika



Elsadat

Help

Interactive Game





Can You Detect The Audio Deep Fake?

"Challenge The AI And Test Your Skills!"

Start The Challenge!

Help

Welcome To The Audio Deepfake Challenge!

In This Exciting Game, You Will Put Your Skills To The Test By Determining Whether Audio Recordings Are Real Or Deep Fakes.

Are You Ready To Challenge The AI?

Click The "Next" Button To Listen, Then Decide If The Audio Is Authentic Or Fake. Each Round Counts Towards Your Score.

AUDIO DEEP FAKE



Detect Audio

Speech To Speech

Interactive Games



Score: 4 / 5

Round 0 Of 5

"Can You Identify If This Audio Is Real Or Fake?"

Next

▶ 0:00 / 1:23



"Correct! The Audio Is Authentic."

Welcome To The Audio Deepfake Challenge!

In This Exciting Game, You Will Put Your Skills To The Test By Determining Whether Audio Recordings Are Real Or Deep Fakes.

Are You Ready To Challenge The AI?

Click The 'Next' Button To Listen, Then Decide If The Audio Is Authentic Or Fake. Each Round Counts Towards Your Score.

Help

AUDIO DEEP FAKE



Detect Audio

Speech To Speech

Interactive Games



Score: 4 / 5

Round 0 Of 5

"Can You Identify If This Audio Is Real Or Fake?"

Next

▶ 0:00 / 1:23 ⏪ ⏴ ⏵



"Incorrect! The Audio Is Authentic."

Welcome To The Audio Deepfake Challenge!

In This Exciting Game, You Will Put Your Skills To The Test By Determining Whether Audio Recordings Are Real Or Deep Fakes.

Are You Ready To Challenge The AI?

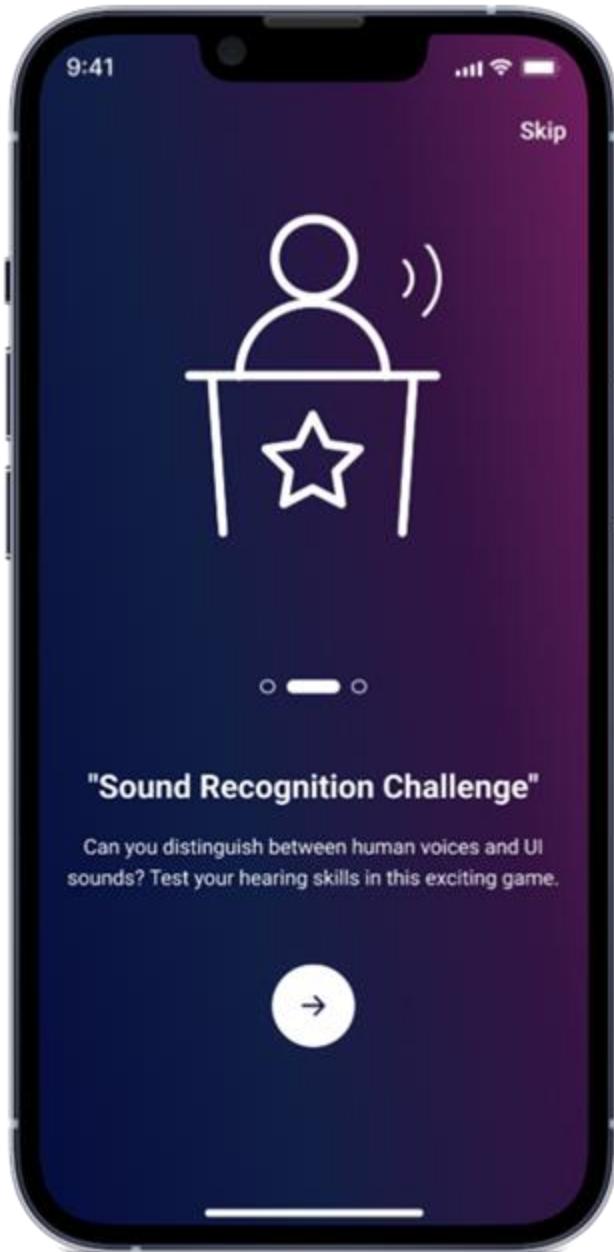
Click The "Next" Button To Listen, Then Decide If The Audio Is Authentic Or Fake. Each Round Counts Towards Your Score.

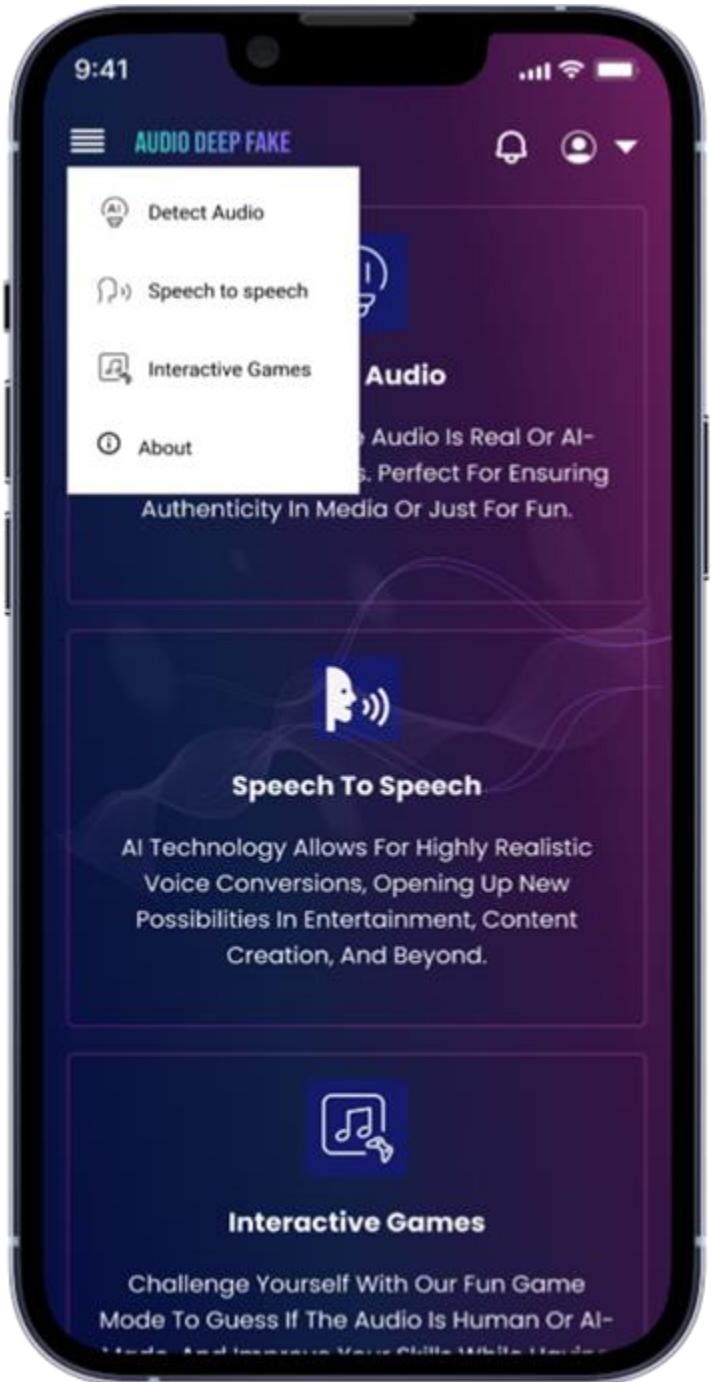
Help

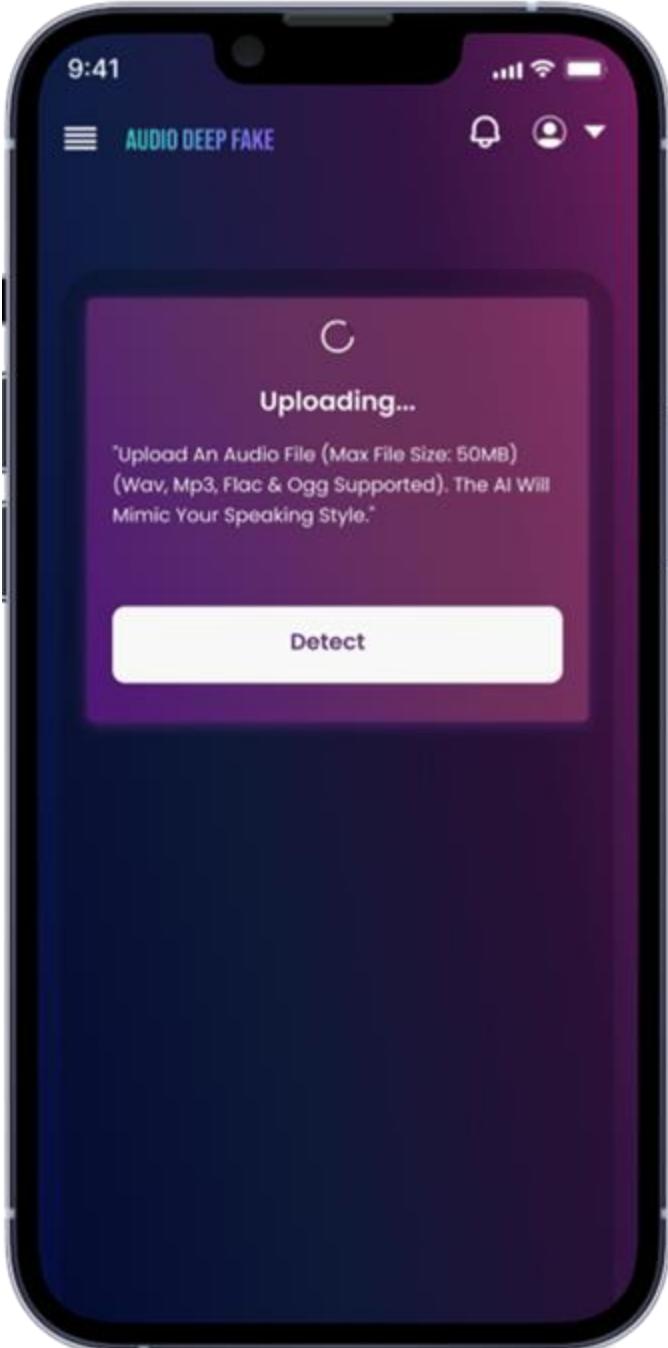
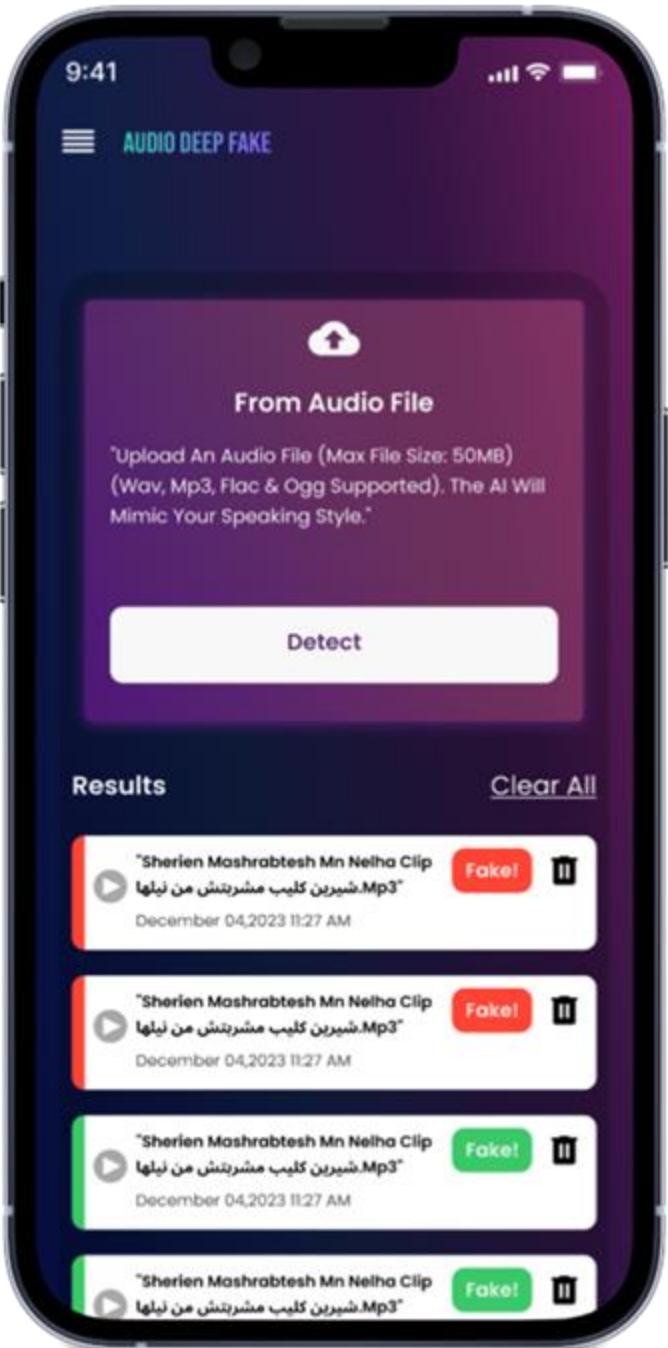


Mobile Application

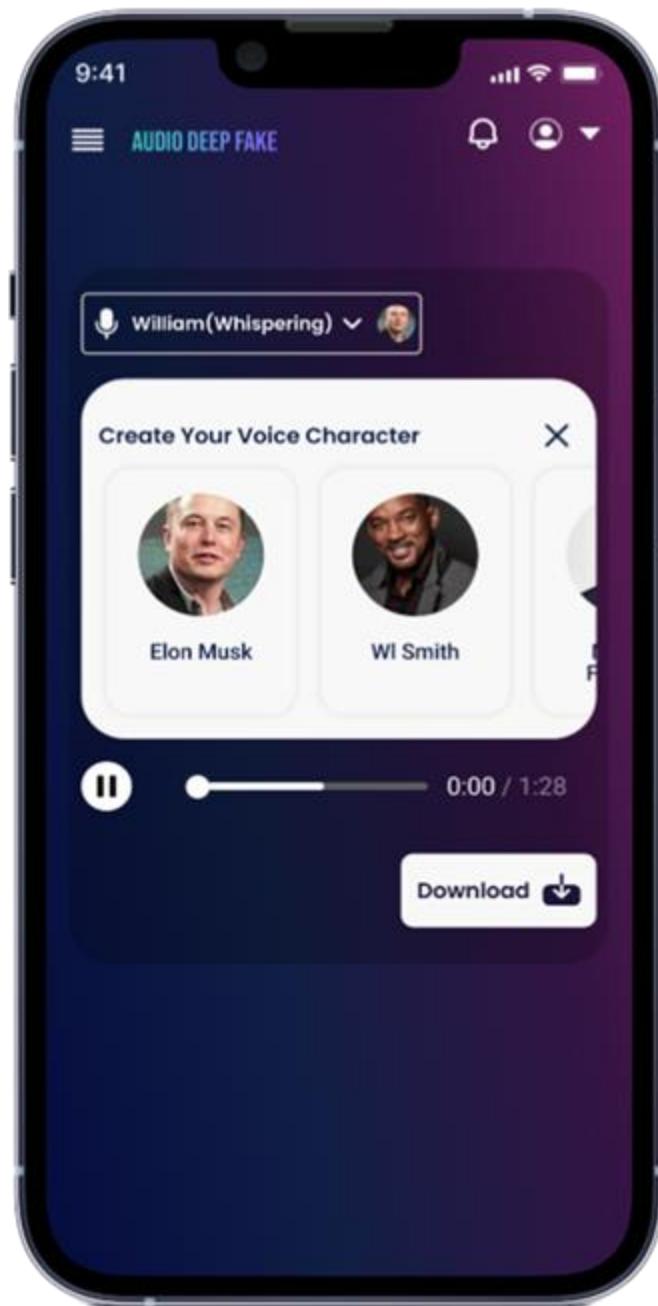
Onboarding





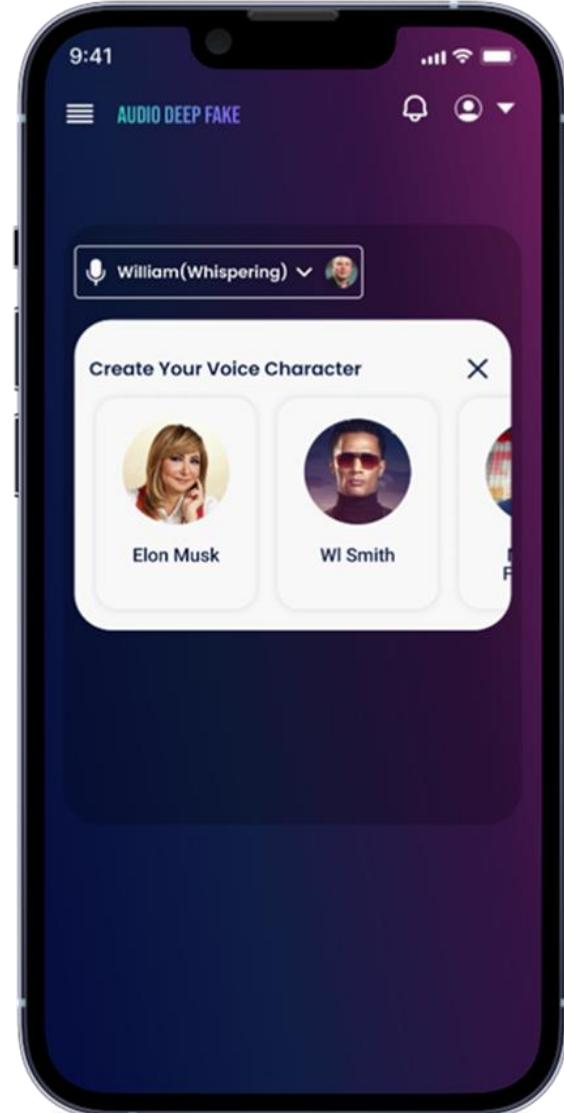


Speech To Speech

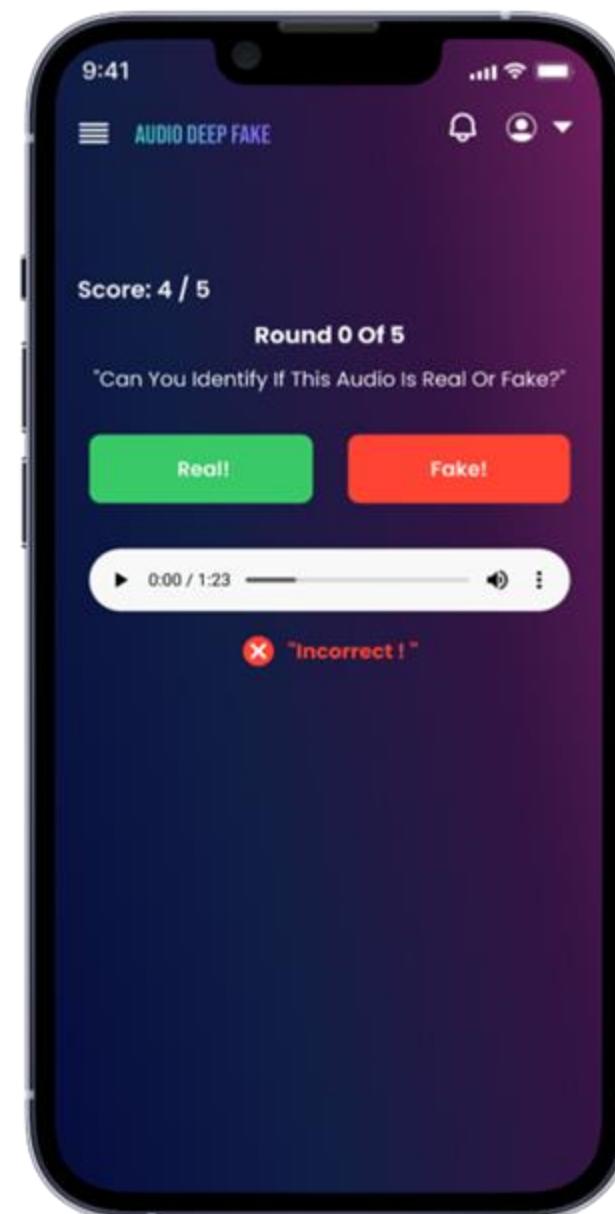
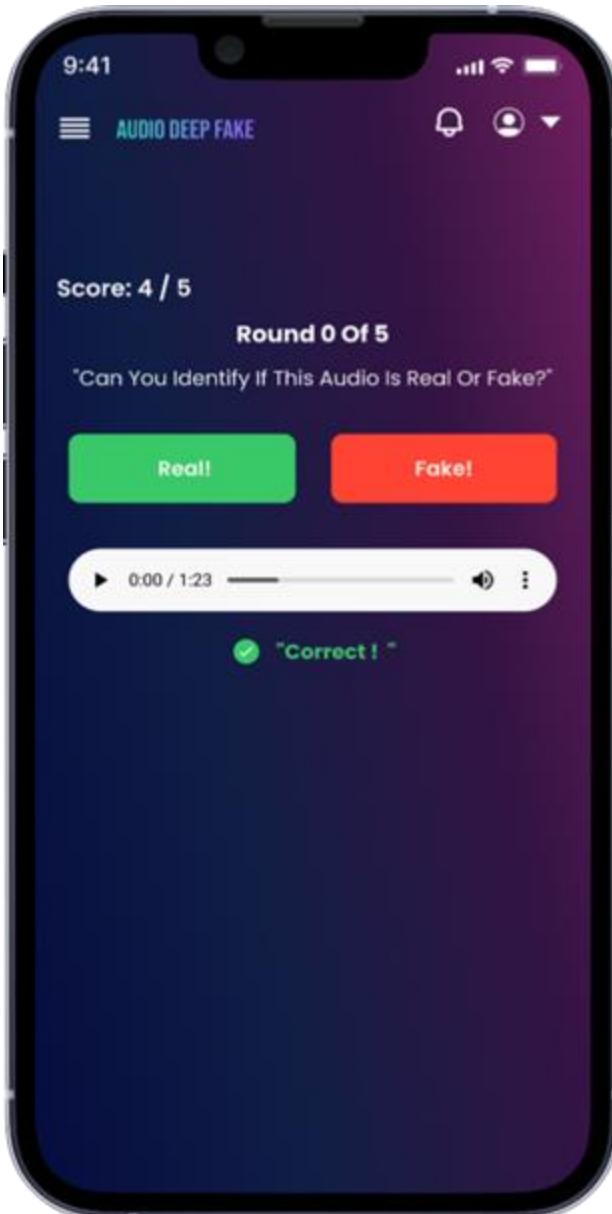
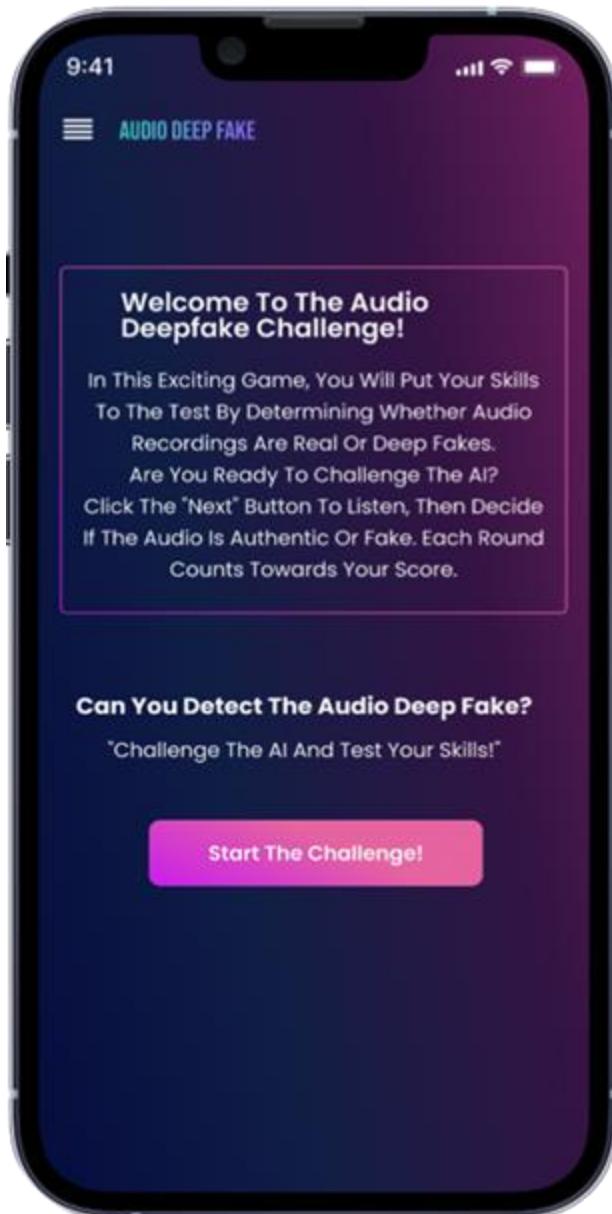




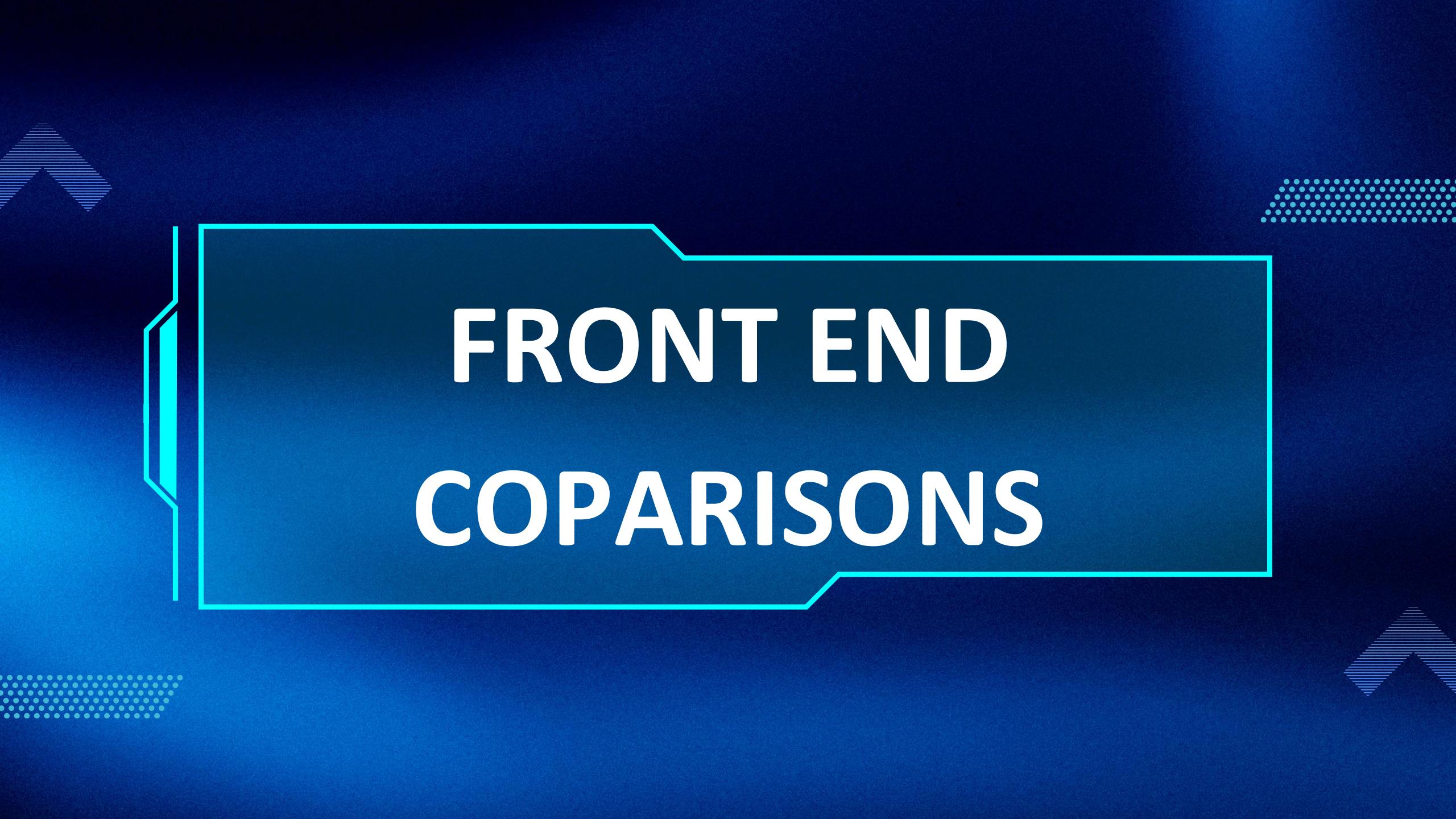
Text To Speech



Interactive Games

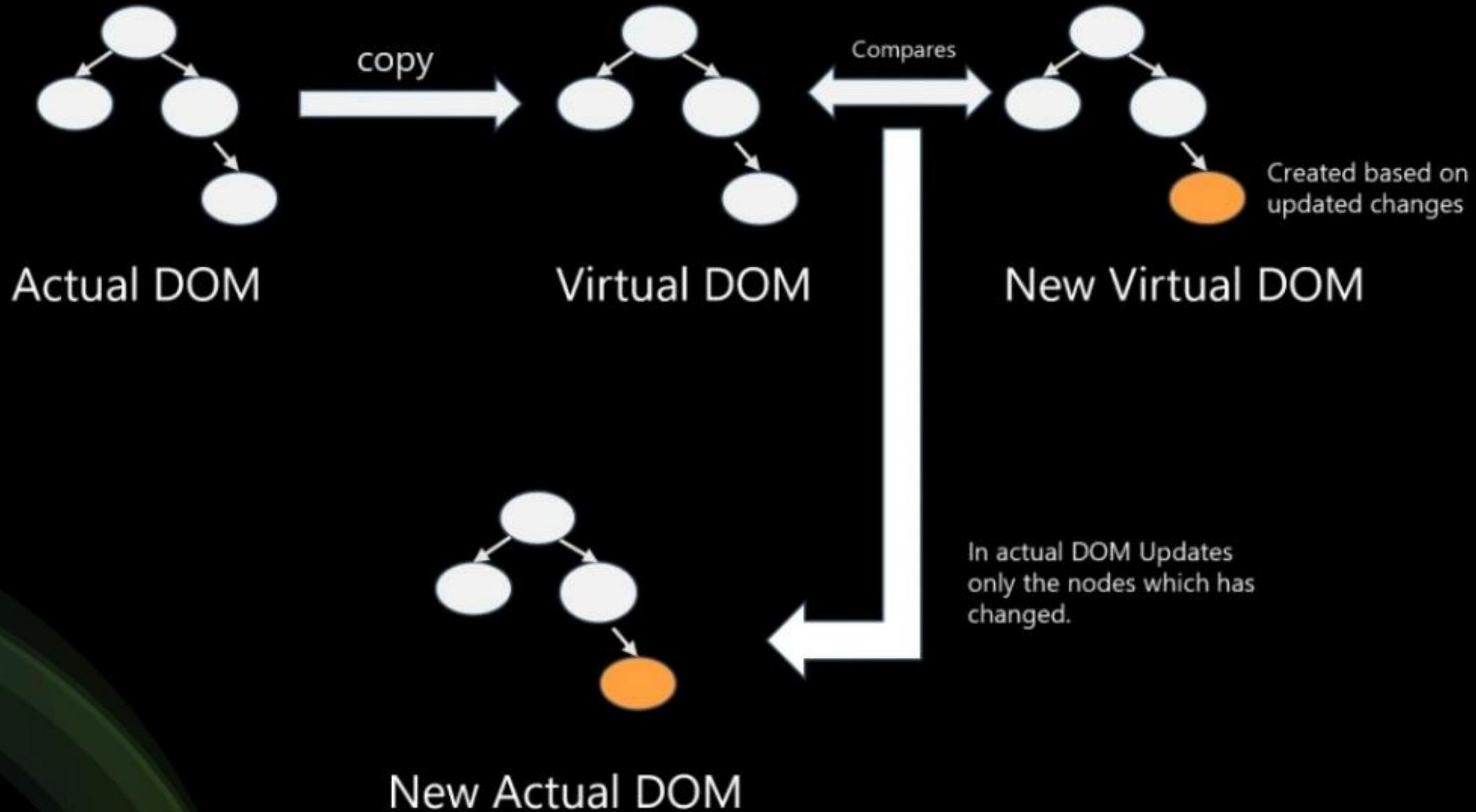


Comparisons (nour - Kiro)



FRONT END COPARISONS

Framework	React	Angular	Vue
Type	Library	Framework	Framework
Rendering	Virtual DOM	Real DOM	Virtual DOM
Routing	✓	✓	✓
Community & Ecosystem	✓	✓	
Performance	✓		✓
Limitations	<ul style="list-style-type: none"> 1. JSX can be a barrier for some developers. 2. Requires additional libraries for full functionality (e.g., state management). 	<ul style="list-style-type: none"> 1. Steeper learning curve, especially for beginners. 2. Can be overkill for small projects due to its complexity. 	<ul style="list-style-type: none"> 1. Smaller community compared to React and Angular. 2. Less enterprise-level support and resources.



MOST FRAMEWORKS USED ACCORDING TO STACK OVERFLOW SURVEY 2024



Backend Development & Security Aspects

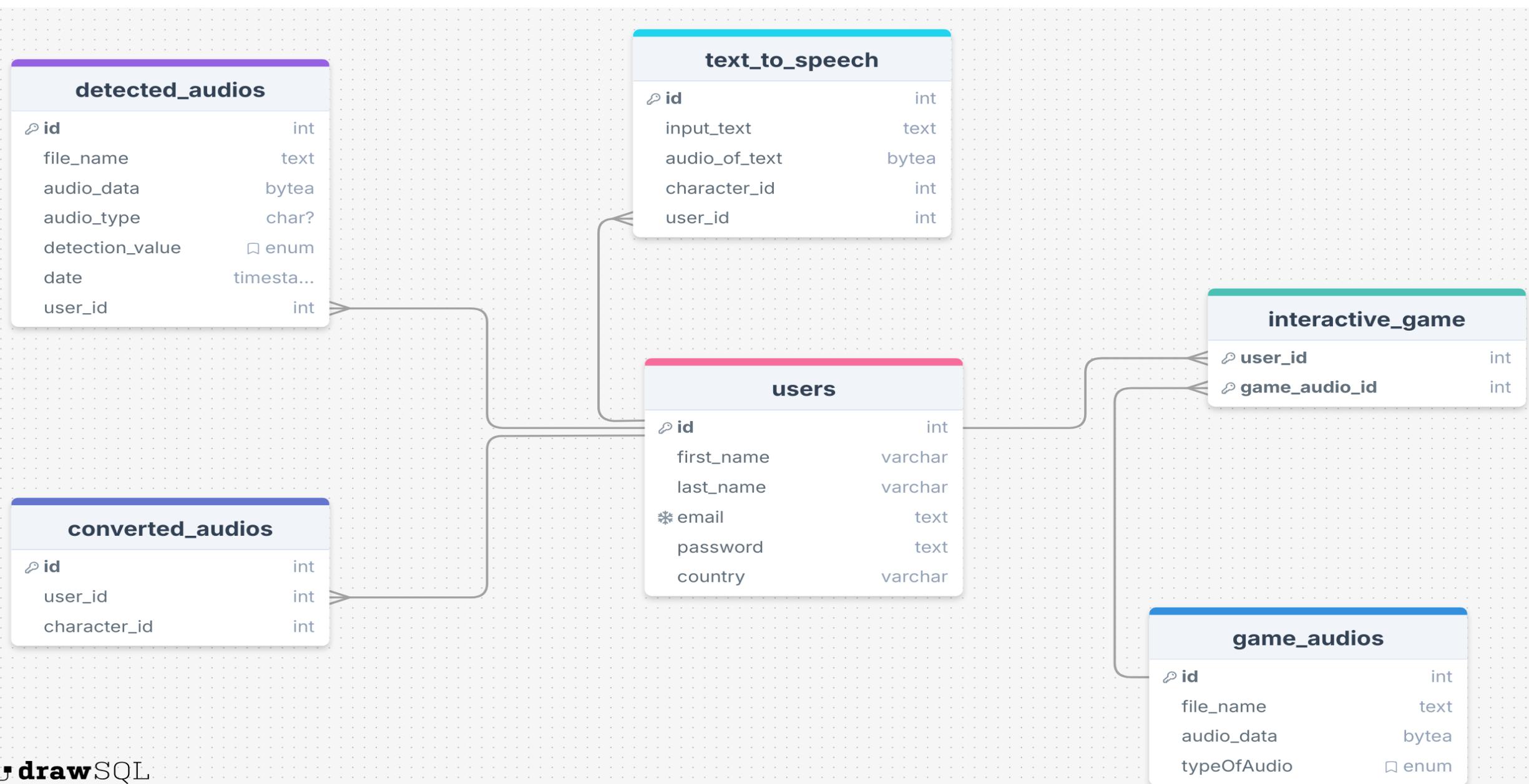
Backend Framework Technologies

Framework	Node.js	Flask
Language	JavaScript	Python
Performance	Good for I/O-bound tasks like APIs and real time apps	Good for CPU-bound tasks like AI models
Community & Support	Huge global community	Tight-knit Python community
Concurrency Model	Single-threaded event loop (non-blocking I/O, handles many requests at once)	Handles one request per worker, can be scaled with multiple workers
Ecosystems	NPM (real-time frameworks like Express)	PyPI (supports AI libraries like TensorFlow, PyTorch, NumPy, etc.)
Database Support	Supports both NoSQL & SQL efficiently	Supports SQL and can handle NoSQL
Limitations	<ul style="list-style-type: none">Callback Complexity "callback hell"CPU-Bound Limits (heavy computations block the event loop and slow all requests)	<ul style="list-style-type: none">Synchronous By Default (can be slower under high concurrent load)GIL Constrains (Python's Global Interpreter Lock limits multi-threaded CPU Performance)

Database Options

Database	PostgreSQL	MySQL	MongoDB
Database Type	Relational + Object-Relational (Hybrid)	Relational (SQL-based)	NoSQL (Document-based)
Schema	Fixed schema, with flexibility via JSON	Fixed Schema	Schema-less, flexible
Query Language	SQL with advanced extensions	SQL (Structured Query Language)	MongoDB Query Language(MQL)
Scalability	Horizontal and Vertical Scaling	Vertical Scaling	Horizontal Scaling(Sharding)
Performance	Balanced for both transactional and analytical workloads	Optimized for structured data, transactional apps	Optimized for unstructured data, real-time apps
Use Cases	Analytics, hybrid systems, modern apps	E-commerce, Traditional websites	Rabidly Changing data models, real-time apps

Database ERD Model



Security Aspects



Password Hashing



Environment-Backed Secrets



CORS Restrictions



Session Management



Passport Google OAuth2

Mobile Technologies

Options:

- Native

- Kotlin

- Swift

- Cross Platform

- Flutter

- React Native

Why excluding the native options

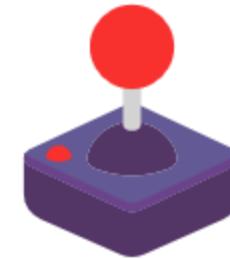
Building one app shouldn't feel like building two.

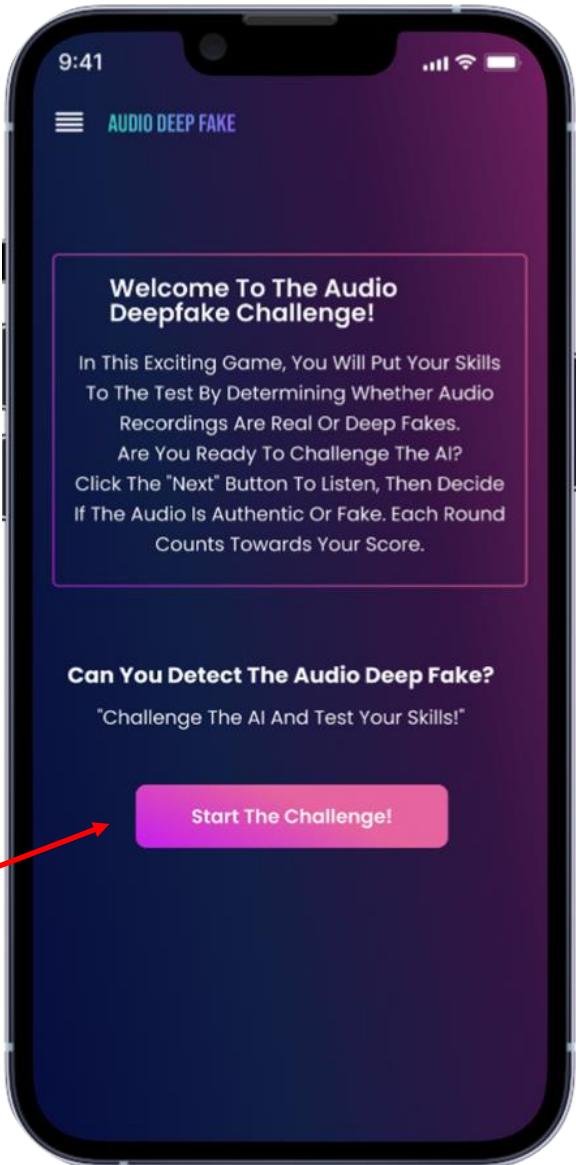
-  **Two Codebases** → Android (Kotlin) + iOS (Swift) = Double effort
-  **Slower Development** → Features take twice the time
-  **Higher Cost** → More dev time = More money
-  **Hard to Stay Aligned** → UI and features may drift apart
-  **Less Flexibility** → Difficult to iterate fast or share logic
-  We needed one fast, flexible, unified solution.
→ So we moved forward with **cross-platform frameworks**.

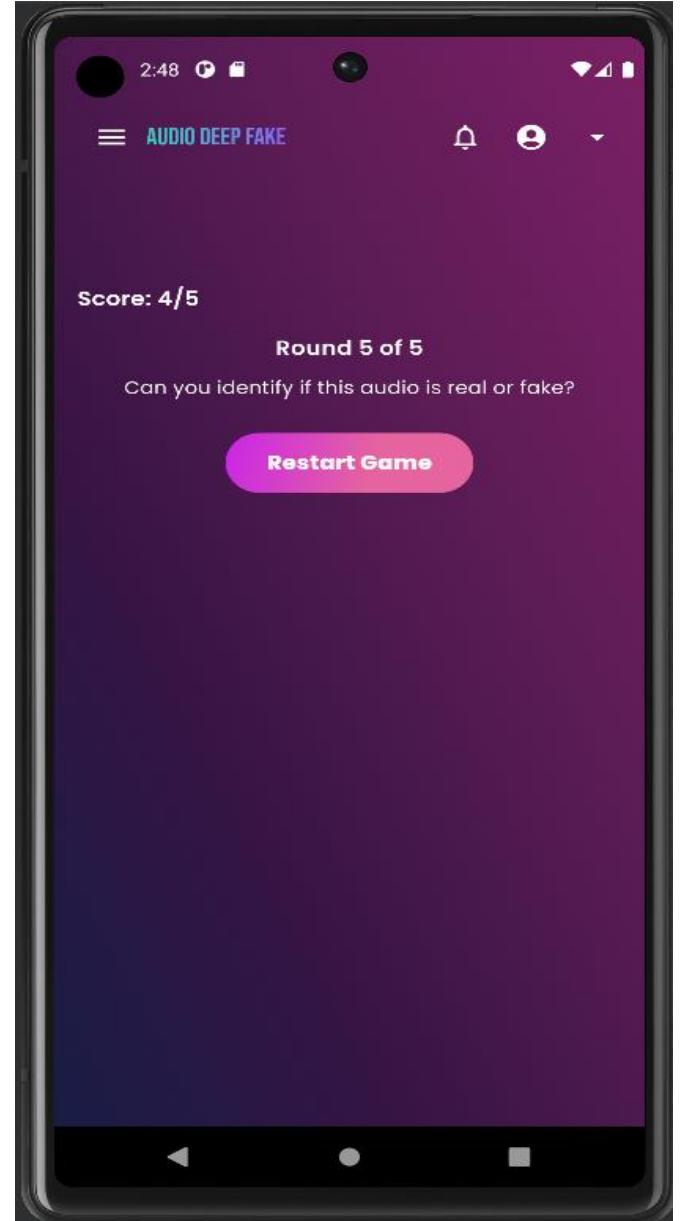
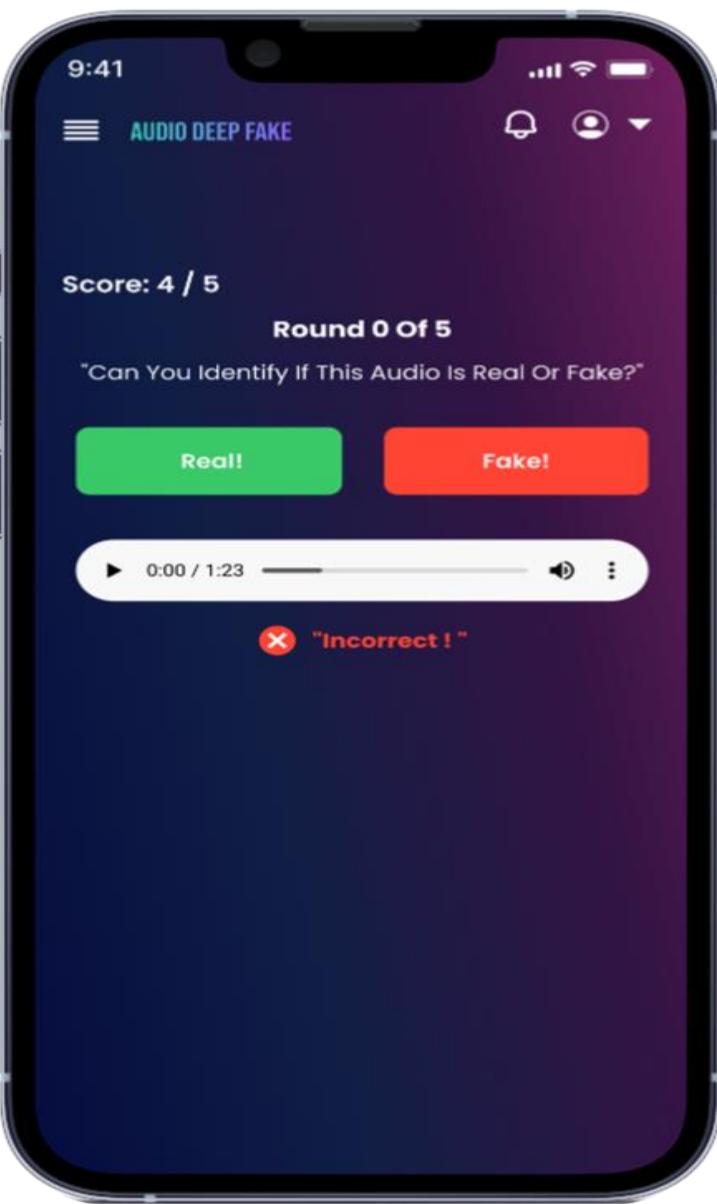
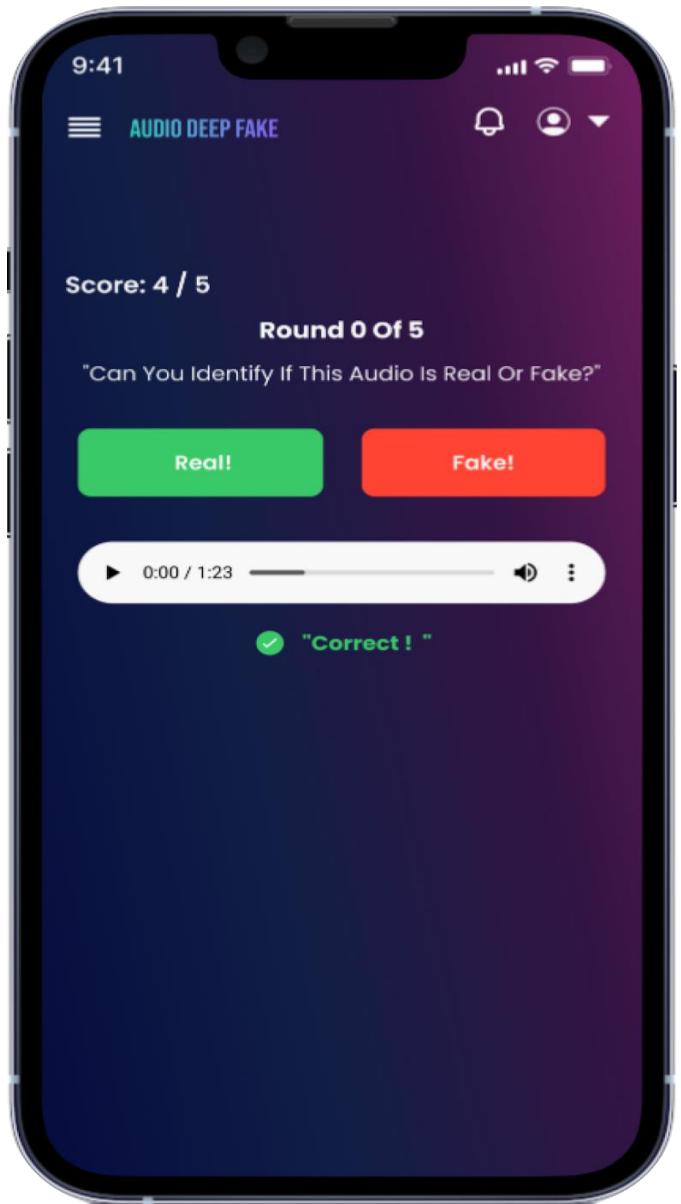
Flutter vs. React Native

	Flutter	React Native
 Made by	Google...	Facebook...
 Codebase	One code for Android & iOS	One code for both platforms
 Design	Consistent UI everywhere	Can feel different on each OS
 Speed & Performance	Fast & smooth—great for audio work	Slightly slower under heavy audio load
 Voice Tools	Excellent plugin support	Decent but requires more setup
 Google Integration	 Beginner-friendly with Firebase, ML, TTS	 Works but needs extra configuration
 Community Support	Growing fast with strong Google backing	Solid but less focused on audio tools

Game



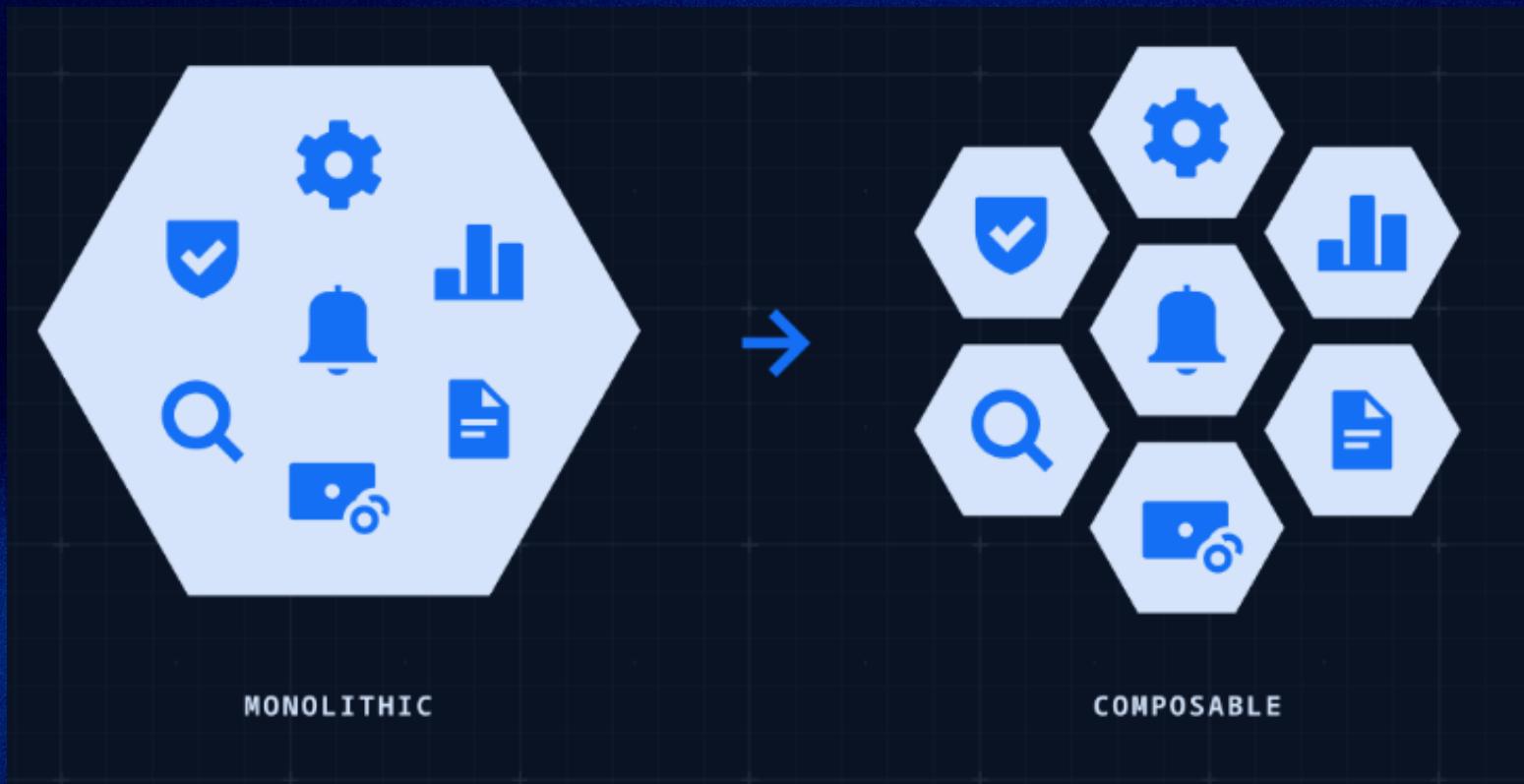




architecture (3omda)

GOALS

Scalability
Maintainability
Performance
Reusability
Testability
independent of specific technologies
making the code more organized and
understandable.



WEB & MOBILE

WEB ARCHITECTURE

Component-Based

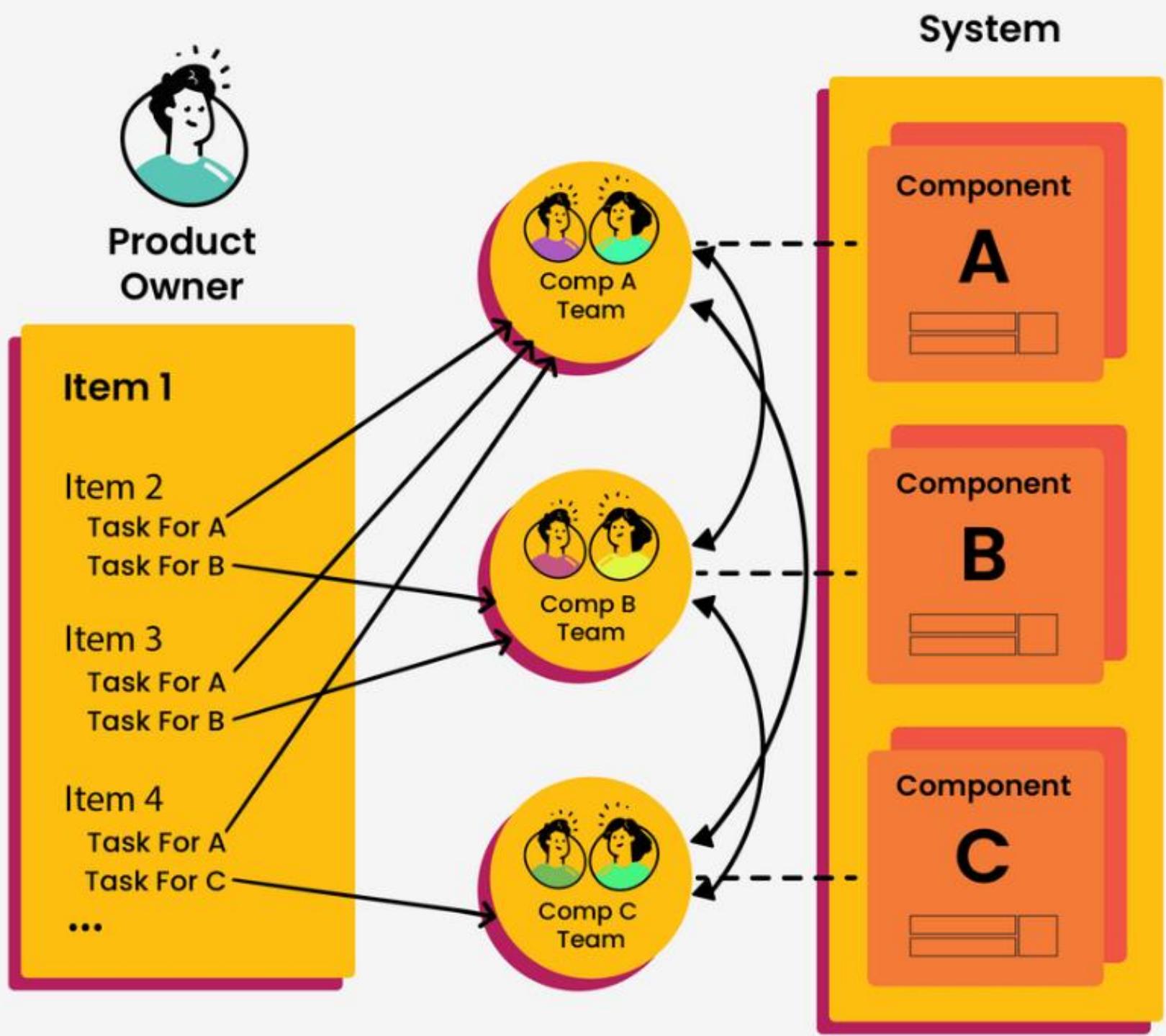
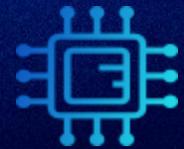
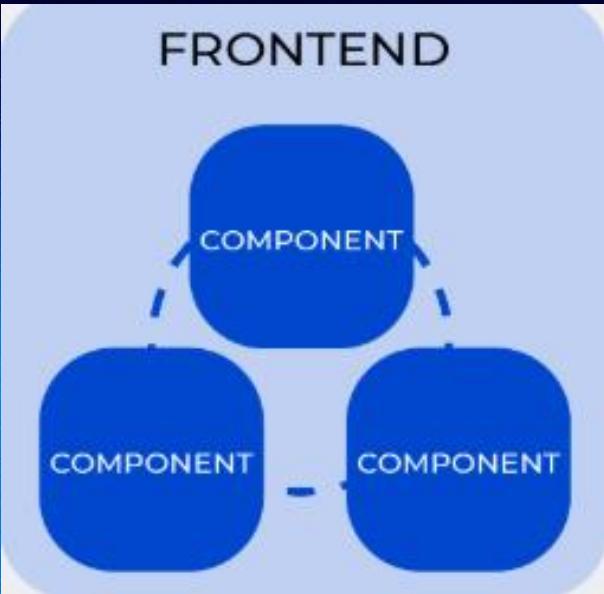
Component-Based Architecture structures software as a collection of reusable, self-contained components. Each component encapsulates specific functionality and communicates through well-defined interfaces.

MOBILE ARCHITECTURE

Clean Architecture

Clean Architecture emphasizes separation of concerns by organizing code into layers with clear boundaries. It promotes testability, scalability, and independence from frameworks, UI, and databases.

Component-Based



System

LAYERS

Widgets ←

Screens ←

State management ←

Presentation

Domain



Use cases

Entities

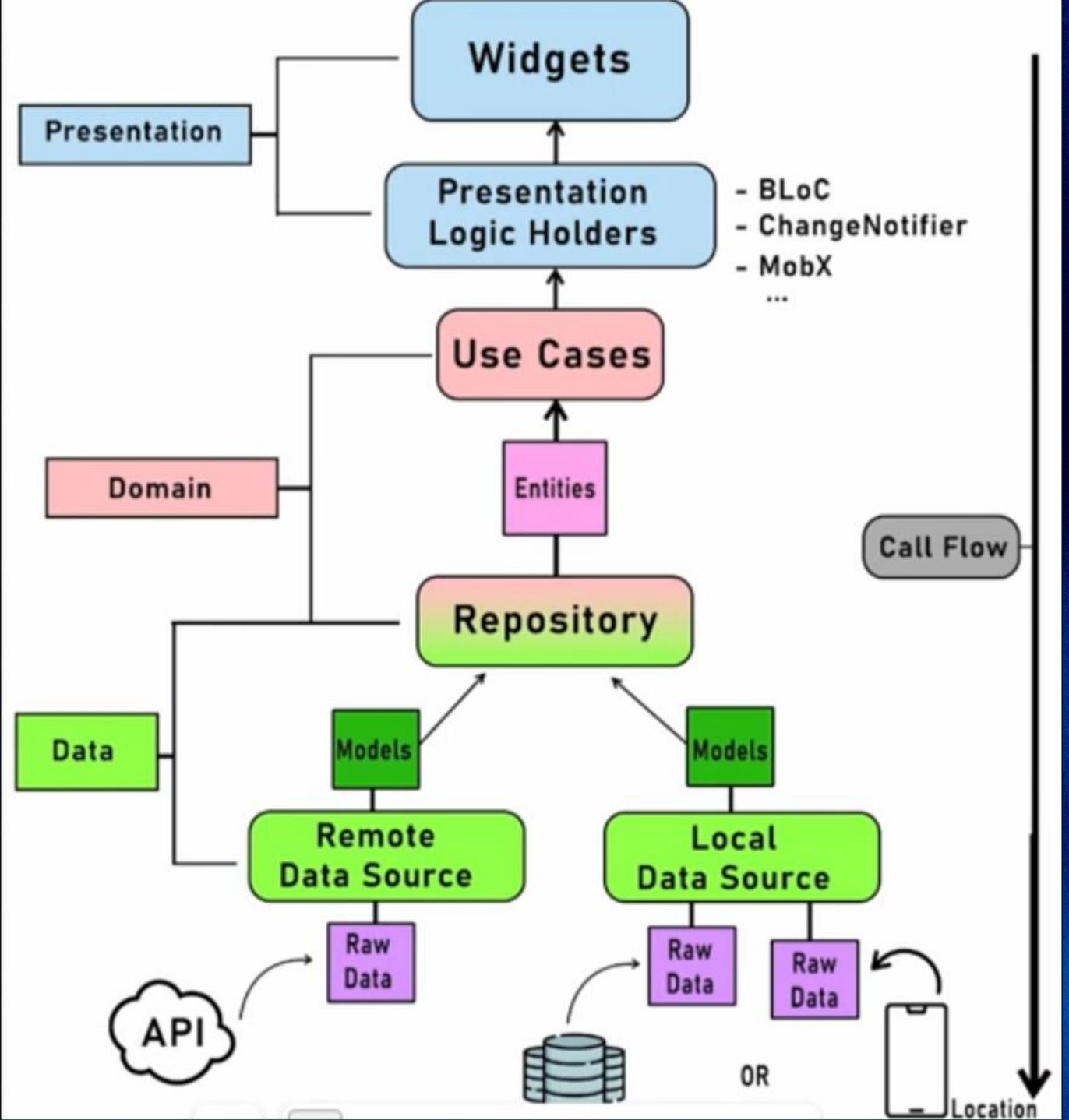
Repositories

Repositories ←

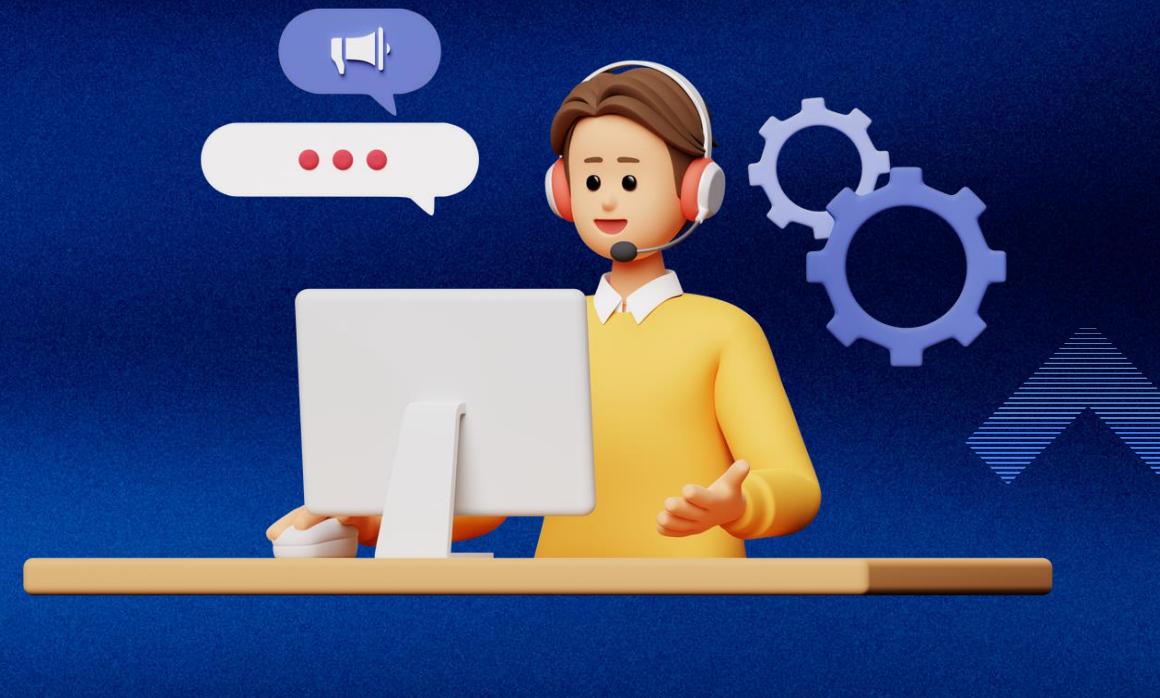
Models ←

Data sources ←

Data



HOW IT WORKS



Demo (kiro- nour)

Future work (ali taha)

any question (ali taha)

