1. Introduction:

Meet BOTX, your new AI buddy designed to have chats that feel just like talking to a friend. It's not your average digital assistant—it's here to offer warmth and comfort through conversation. BOTX is all about making you feel good, with its clever responses that feel just like chatting with someone close to you. It's like having a digital friend who's always there to cheer you up and lend a listening ear. Say hello to a whole new level of friendly AI companionship!

2. Current Features:

- Mimic Human Behavior:

We are using WhatsApp chat history to tell our AI model how people talk and interact. This helps our model understand human conversations better and mimic them. Our goal is to offer personalized and caring support by using advanced AI models. This way, we can help people feel less lonely more effectively.

- Persistent Memory:

We are saving all chat history in a database, and then load it into the AI's memory component as needed. This helps our model remember past conversations without losing the previous information. When you start chatting, the AI can access previous messages quickly, making the conversation smoother.

- Audio Message:

We are using two powerful models. One transcribes your spoken words into text, so the AI can understand them. The other takes the AI's text responses and turns them into natural-sounding speech. This way, when you send a voice message, the AI can hear it, understand it, and respond with a spoken message of its own.

- Voice Cloning:

Our model can clone anyone’s voice. So, when you send a voice message, the chatbot can respond in a voice that sounds just like yours or the person you're chatting with. It adds a whole new level of personalization to the conversation!

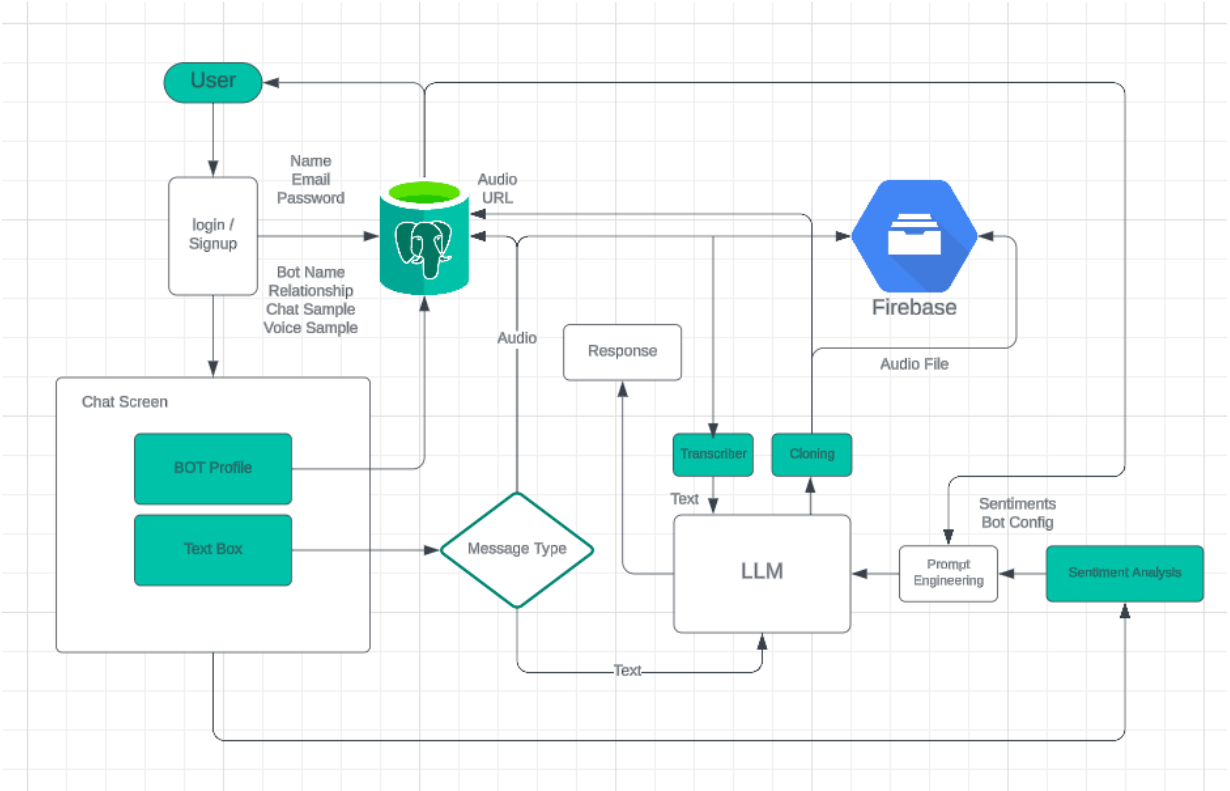
- ## Personal Languages:

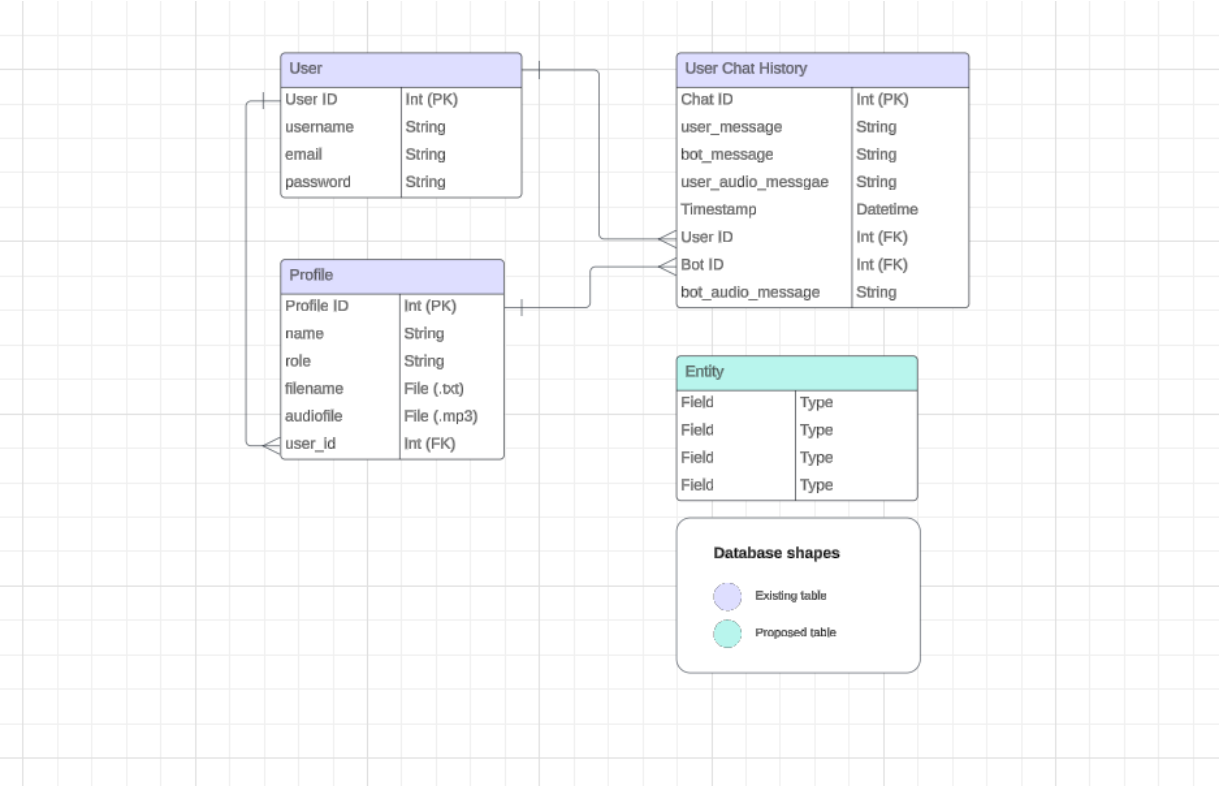
  Our model can talk in English and Romanized Urdu. We are limiting the response by doing some prompt engineering. It can understand and respond appropriately. So, whether you're chatting in English or Romanized Urdu, our system will understand and reply in the same language, ensuring smooth communication.

- ## Sentiment Analysis:

  When you send a message, our system does something cool. It takes a quick snapshot of your face using the camera on your device. Then, it looks at your facial expression to figure out how you're feeling. Based on that, it analyzes your emotions and sends that information to our Large Language Model (LLM). The LLM then adjusts its response to match your current mood. So, whether you're feeling happy, sad, or anything in between, our system can understand and respond accordingly.

3. Architecture Diagram





4. Tech Stack

- Front-End Framework: JavaScript, React, Tailwind and Particle JS
- Backend Technologies: Python, Flask
- Database: PostgreSQL and Firebase File storage
- AI: Langchain and OpenAI
- Models:

- Audio Transcriber: Nova (deepgram)
- Voice Cloning Model: TTS (huggingface)
- Sentiment Analysis Model: Open CV

Future Implementations

Voice Clone Research

We have implemented the Eleven Labs and found out that the results are quite good. We can rate our previous model and this model

Model	Time	Result Similarity
TTS	At least 8 to 10 seconds	30% - 40%
Eleven Labs	Less than 5 seconds	75% - 80%

We can opt for the Pro Subscription plan, priced at $99, which offers several benefits:

1. 500,000 characters per month (though this may still be limited).
2. If we exceed the 500,000-character limit, additional characters will be charged at $0.24 per 1000 characters.
3. Access to create up to 160 custom voices.

This subscription plan provides great value for its features and flexibility, ensuring that our needs are met efficiently.

Custom Avatar Research

We have tried implementing many models, some of which yielded better results. However, to use them, GPUs are required. Otherwise, we face a delay of around 20 minutes in every response because several algorithms work in the background, such as lip synchronization and movement of heads. Therefore, based on our requirements, we have found two models that can provide very good results.

1. D-ID
2. HeyGen

We tried using HeyGen, but its free version didn't meet our requirements. However, we watched several YouTube videos where we found good results.

What are we getting according to our use case by using these models?

- Customize Avatar (same as you)
- Best Lip sync results (that make you feel more realistic)
- Real time response in no time
- More than 40 languages

- More than 100 accents

Both models are good, but I would prefer Heygen over D-ID because of its additional functionalities and better lip sync results, even though it's a bit more expensive.

Good Pricing Plan:

- $150 per month
- 5 custom Avatar
- Voice Cloning (we don’t know about results)
- But we have some confusion in credits, like we don’t know the specific meaning of credits we just know k we will get 30 credits in this.

How much time would it take to implement this feature?

We have thoroughly reviewed the documentation, although we may not have full expertise in this. However, we estimate that it may take us at least 17 days (about 2 and a half weeks), but we can definitely accomplish this.