International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Voice-Based Virtual Assistant for Emotionally Sensitive Interactions

1st Author
Abhishek Chauhan
B.E-CSE, Chandigarh University
Mohali, India
Sec12pgi@gmail.com

2nd Author
Sarthak Chandi
B.E-CSE, Chandigarh University
Mohali, India
chandi33sarthak@gmail.com

3rd Author
Ashish kumar
B.E-CSE, Chandigarh University
Mohali, India
anshukumaridea3@gmail.com

4th Author
Aryan Vishwa
B.E-CSE, Chandigarh University
Mohali, India
khannaaryan392@gmail.com

5th Author
Pulkit
B.E-CSE, Chandigarh University
Mohali, India
plktsamria@gmail.com

6th Author
Er Priyanka Devi
B.E-CSE, Chandigarh University
Mohali, India
pjangra9105@gmail.com

ABSTRACT

This paper presents Luna, an emotionally intelligent, voice-based virtual assistant built to engage users in meaningful, empathetic conversation. Unlike conventional AI assistants that emphasize efficiency and task execution, Luna is designed to understand human emotions and respond with care, warmth, and contextual understanding.

Luna integrates local large language models (LLMs) using llama-cpp-python, speech recognition via OpenAI's Whisper, and expressive voice output via Coqui's YourTTS.

The assistant interacts through a chat-like GUI designed using Tkinter, offering both voice and text input. The system maintains recent dialogue history and adapts its responses with emotional sensitivity based on user input.

This paper discusses the architecture and technical implementation of Luna, demonstrating how emotion-aware AI can enhance digital companionship in an entirely offline environment.

General Terms

Human-Computer Interaction, Artificial Intelligence, Natural Language Processing, Voice Assistant, Emotion Recognition.

Keywords

Empathetic AI, Human-AI Dialogue, Voice-Based Assistant, Whisper STT, YourTTS, Coqui TTS, Hybrid TTS, Emotion Detection, Local LLMs

1. INTRODUCTION

As AI becomes more integrated into our daily lives, there is growing recognition that virtual assistants should not only perform tasks efficiently but also provide **emotionally responsive communication**. While commercial assistants like Siri, Google Assistant, and Alexa offer reliable task execution, they fall short in addressing the emotional needs of users.

Luna is a voice-based virtual assistant created to fill this gap. Luna is designed to simulate a warm and friendly companion who listens, understands, and responds in a supportive manner. The system features real-time speech recognition, a locally running LLM (LLaMA 3.2B), and emotionally adaptive speech synthesis. Luna supports both functional queries and mental health conversations—delivered through an empathetic lens.

2. RELATED WORK

Research in emotional dialogue systems has evolved with the introduction of datasets like EmpatheticDialogues, DailyDialog, and PersonaChat, which focus on emotionally rich conversations. However, most of these systems are text-only and lack real-world, speech-based implementations.

Text-to-Speech (TTS) systems such as Tacotron2, Glow-TTS, and YourTTS have made strides in emotional speech generation. Coqui's YourTTS stands out for zero-shot speaker cloning and emotional expressiveness. Whisper by OpenAI provides robust multi-language speech-to-text capabilities. Combining these with open-source LLMs like LLaMA and GUI frameworks like Tkinter allows us to develop a fully offline, hybrid assistant that feels conversationally human.

3. SYSTEM ARCHITECTURE

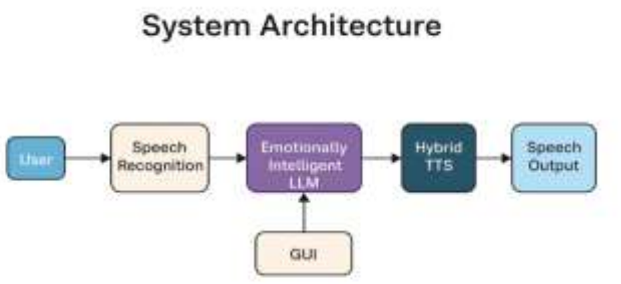


Figure 1: System architecture of Luna, using Whisper STT, a local LLM, and hybrid TTS.

Luna consists of multiple subsystems working in tandem:

3.1 Speech Recognition

Whisper (base model) is used for real-time voice transcription. It records up to 6 seconds of user audio at 16kHz using sounddevice, saves it as a .wav file, and transcribes the content using Whisper's transcribe() function.

3.2 Emotionally Intelligent LLM

Luna uses the LLaMA 3.2B Instruct model loaded through llama-cpp-python. The assistant maintains a rolling history of recent user–assistant exchanges (up to 10 turns). Each prompt to the model is built using a system persona:

You are Luna — a cheerful, emotionally intelligent, and friendly AI assistant...

This ensures consistency in tone and empathy across conversations. Streaming generation (stream=True) is enabled for real-time token output.

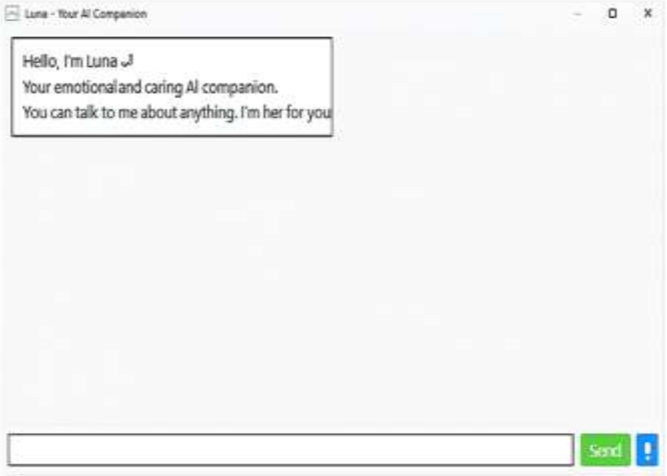
3.3 Hybrid Text-to-Speech (TTS)

Luna features two voice modes:

- **Microsoft Hazel (via pyttsx3):** Offline, fast, neutral voice used for casual or task-oriented replies.
- **Coqui YourTTS:** Emotional, human-like voice based on reference audio (e.g., Nichalia_Schwartz.wav). Luna automatically switches to this voice when user input suggests sadness or vulnerability.

3.4 Graphical User Interface

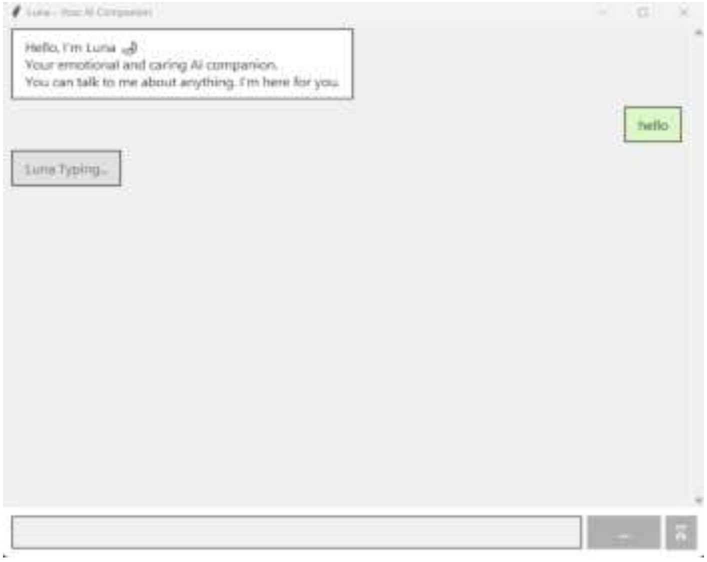
Built in **Tkinter**, the GUI mimics modern chat apps:



Hello, I'm Luna 🎵
Your emotional and caring AI companion.
You can talk to me about anything. I'm her for you

Figure 2: GUI interface of Luna showing the chat window, greeting message, and voice input control.

- Colored message bubbles
- Scrollable conversation area
- Mic button for voice input
- Live updates while Luna is thinking ("Luna Typing...")



Hello, I'm Luna 🎵
Your emotional and caring AI companion.
You can talk to me about anything. I'm here for you.

hello

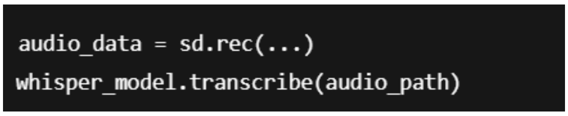
Luna Typing...

Figure 3: GUI interface of Luna Typing.

- Hotkeys for fullscreen, return key for sending text

4. METHODOLOGY

4.1 Speech-to-Text (STT)

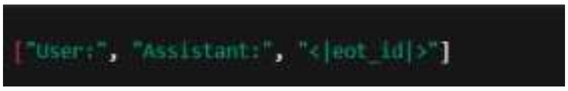


```
audio_data = sd.rec(...)
whisper_model.transcribe(audio_path)
```

Whisper provides robust transcription with support for different accents and background noise. Silence detection is implemented via max amplitude checks.

4.2 Prompt Engineering & LLM Response

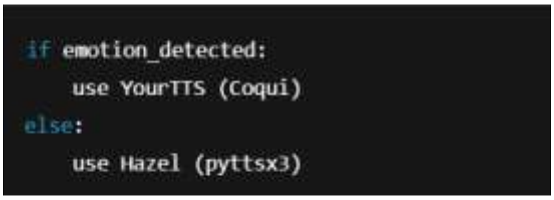
Prompt = [System Prompt] + [Conversation History] +"Assistant:" Luna stops generation at:



```
["User:", "Assistant:", "<|eot_id|>"]
```

Tokens are streamed and formatted into GUI chat bubbles. If keywords like "sad", "alone", or "stressed" appear, the system switches to emotional voice output.

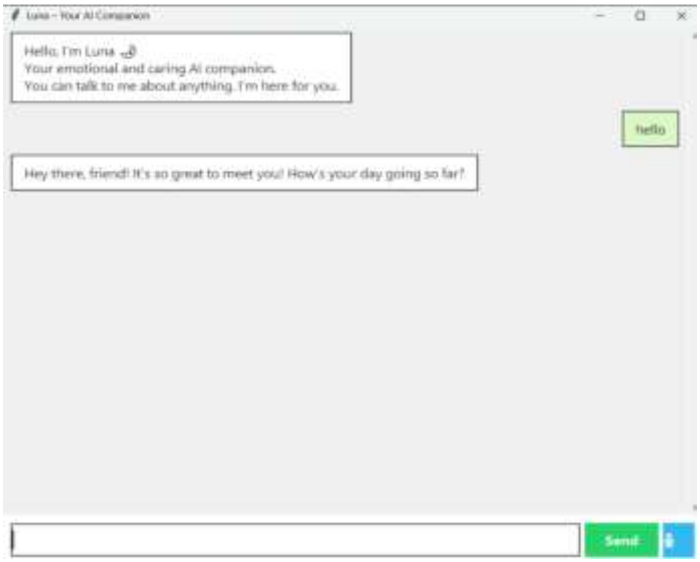
4.2 Hybrid TTS Logic



```
if emotion_detected:
    use YourTTS (Coqui)
else:
    use Hazel (pyttsx3)
```

Generated text is sanitized with regex to remove markdown (**bold**), notes, and non-ASCII characters for clean speech delivery.

4.3 GUI Interaction



Hello, I'm Luna
Your emotional and caring AI companion.
You can talk to me about anything. I'm here for you.

hello

Hey there, friend! It's so great to meet you! How's your day going so far?

Figure 4: GUI interface of Luna responds to user.

Each message is processed through:

- add_message() — Displays in GUI
- speak_text() — Triggers audio
- update_ui_state() — Refreshes button states

Threads are used to separate UI, voice input, and TTS to ensure responsiveness.

5. RESULTS AND FEATURES

Feature	**Description**
Voice Recognition	Accurate speech-to-text using Whisper (base model)
Emotional Voice	Expressive, human-like voice using Coqui YourTTS and reference audio
Natural Persona	Luna speaks in a caring, emotionally intelligent tone
GUI Interface	Tkinter-based interface with chat bubbles, mic button, and status indicators
Multimodal Input	Users can interact via voice or keyboard
Offline Mode	Entire system runs locally (LLM, TTS, STT) without internet
Text Formatting	Cleans up LLM output before speaking (removes markdown, notes, etc.)
Threaded Design	Separate threads for UI, audio input/output, and LLM calls for responsiveness
Context Memory	Remembers recent user–assistant turns for coherent responses

User Feedback
Testers appreciated Luna's voice warmth and understanding, especially in conversations related to sadness, loneliness, or emotional stress. The assistant was described as "comforting," "relatable," and "non-judgmental."

6. PROPOSED ENHANCEMENTS

- **Memory Persistence and Personalization:**
  Luna will be enhanced with short-term memory to remember user-specific details such as names, preferences, or previous moods during the session. This would allow for more personalized and context-aware conversations.

- **Voice Emotion Detection:**
  While Luna currently infers emotion based on text, incorporating voice-based emotion recognition using deep learning (e.g., CNNs trained on audio spectrograms) can allow more accurate mood detection even if the user speaks in a neutral tone.

- **Mood History Visualization**
  The assistant will track emotional trends over time and display them in a user-friendly visual format (graphs or mood calendars), offering insights into the user's emotional wellbeing.

- **Guided Wellness Modules:**
  Luna can be extended to include calming features such as guided meditations, breathing exercises, or motivational affirmations, triggered contextually when signs of stress or sadness are detected.

- **Multilingual Emotional Support:**
  Using Whisper's multilingual STT and LLaMA's language capabilities, Luna can offer emotionally supportive conversations in multiple languages, especially regional ones, increasing inclusivity.

- **Custom Voice Personalization:**
  Users will have the option to customize Luna's voice, selecting from different emotional tones, voice styles, or even cloning their own preferred voice via YourTTS.

- **Contextual Memory with Vector Databases:**
  Implementing local vector storage (e.g., FAISS, ChromaDB) will enable Luna to recall relevant past conversations and maintain context across sessions, enhancing empathy and continuity in dialogue.

- **Task Scheduling and Reminders:**
  Luna will include offline task management capabilities such as setting reminders, alarms, or creating to-do lists—all while maintaining an emotionally supportive interface.

7. CONCLUSION AND FUTURE WORK

Luna represents a significant step forward in creating emotionally intelligent, voice-based virtual assistants that prioritize human connection over mere task execution. By integrating local LLMs, Whisper STT, and a hybrid TTS system, Luna offers a fully offline, multimodal experience that feels personal, empathetic, and human-like.

Unlike conventional assistants, Luna is designed to engage with users not just functionally, but emotionally—responding with warmth, care, and contextual understanding. Real-time voice streaming, adaptive tone switching, and a friendly GUI interface ensure a seamless user experience.

Looking ahead, Luna will continue to evolve with the addition of advanced features such as real-time voice emotion detection, memory persistence, multilingual emotional dialogue, and wellness-oriented modules. These enhancements will deepen Luna's ability to act not only as a helpful assistant, but as a comforting digital companion that truly understands and supports the user.

Future Directions:

□ **Integrated Real-Time Emotion Detection from Voice and Text:**
Combining voice tone analysis with semantic sentiment detection will help Luna better recognize subtle emotional states and respond more empathetically.

□ **Long-Term Memory with Context Awareness:**
Implementing contextual memory using vector databases (e.g., ChromaDB or FAISS) will allow Luna to remember past conversations and adapt future interactions accordingly, enabling more personalized experiences.

□ **Wellness Tracking Dashboard:**
A visual interface showing emotional trends over time will allow users to track mood fluctuations and receive tailored self-care recommendations.

□ **Mental Health Crisis Protocols:**
Luna will be enhanced with crisis detection capabilities, including identifying red-flag phrases. In extreme cases, it can provide emergency support resources or connect with a pre-defined support network.

□ **Multimodal Interaction:**
Future versions will support multimodal communication (e.g., facial emotion recognition, gesture interpretation), offering even deeper human-like engagement.

□ **Offline Diary & Note Logging:**
Users will be able to speak freely, and Luna will log and summarize key points from their thoughts into a secure offline diary system.

□ **Customizable Assistant Personality:**
Giving users the ability to adjust Luna’s voice, tone, personality, and response style will enhance emotional resonance and user attachment.

□ **Multi-Platform Support:**
Developing Luna as a cross-platform assistant (PC, mobile, browser extension) will increase accessibility and usability across devices

8. ACKNOWLEDGMENTS

We thank our mentors and faculty at **Chandigarh University** for their continuous support and encouragement throughout this emotionally rewarding project.

9. REFERENCES

[1] Rashkin, H., Smith, E. M., Li, M., & Boureau, Y. L. (2019). Towards Empathetic Open-domain Conversation Models. arXiv preprint arXiv:1811.00207. Retrieved from https://arxiv.org/abs/1811.00207

[2] Li, Y., Su, H., Shen, X., Li, W., Cao, Z., & Niu, S. (2017). DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. Proceedings of IJCAI. Retrieved from https://arxiv.org/abs/1710.03957

[3] Kim, J., Lee, S., & Kim, J. (2022). YourTTS: Towards Zero-Shot Multi-Speaker Text-to-Speech Synthesis. arXiv preprint arXiv:2205.03072. Retrieved from https://arxiv.org/abs/2205.03072

[4] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-training. OpenAI preprint. Retrieved from https://openai.com/research/language-unsupervised

[5] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30. Retrieved from https://arxiv.org/abs/1706.03762

[6] OpenAI. (2022). Whisper: Robust Speech Recognition via Large-Scale Weak Supervision. Retrieved from https://github.com/openai/whisper

[7] Abetlen. (2023). llama-cpp-python: Python Bindings for llama.cpp. Retrieved from https://github.com/abetlen/llama-cpp-python

[8] Coqui AI. (n.d.). Coqui TTS: A Deep Learning Toolkit for Text-to-Speech Synthesis. Retrieved from https://github.com/coqui-ai/TTS

[9] Python Software Foundation. (n.d.). Tkinter Documentation. Retrieved from https://docs.python.org/3/library/tkinter.html

[10] SoundDevice Developers. (n.d.). Sounddevice: Audio Playback and Recording. Retrieved from https://python-sounddevice.readthedocs.io

[11] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems, 32.

[12] Romić, M. (2020). Artificial Intelligence in Emotion Recognition: A Critical Review. AI & Society, 35(3), 627–638. https://doi.org/10.1007/s00146-019-00911-6

[13] Picard, R. W. (1997). Affective Computing. MIT Press. ISBN: 9780262661157.

[14] Facebook AI Research. (2023). LLaMA: Open and Efficient Foundation Language Models. Retrieved from https://ai.facebook.com/blog/large-language-model-llama-meta-ai/