AI Driven Personalized Health Assistant: Adaptive Recommendations Based on Individual Health Profiles

SAT5144 - Artificial Intelligence in Healthcare
Team Members: Yaswanth Ganapathi, Ganesh Vannam

Project Goal

- **Key Challenges:**
  - Most apps rely on text input and generic responses
  - Lack of personalization based on patient medical history
  - Require cloud access or high-end hardware
- **Our Goal:**

  To design a lightweight, intelligent assistant that:
  - Listens, understands, and speaks to patients
  - Adapts to individual health data
  - Speaks in the user's preferred language

Literature Review

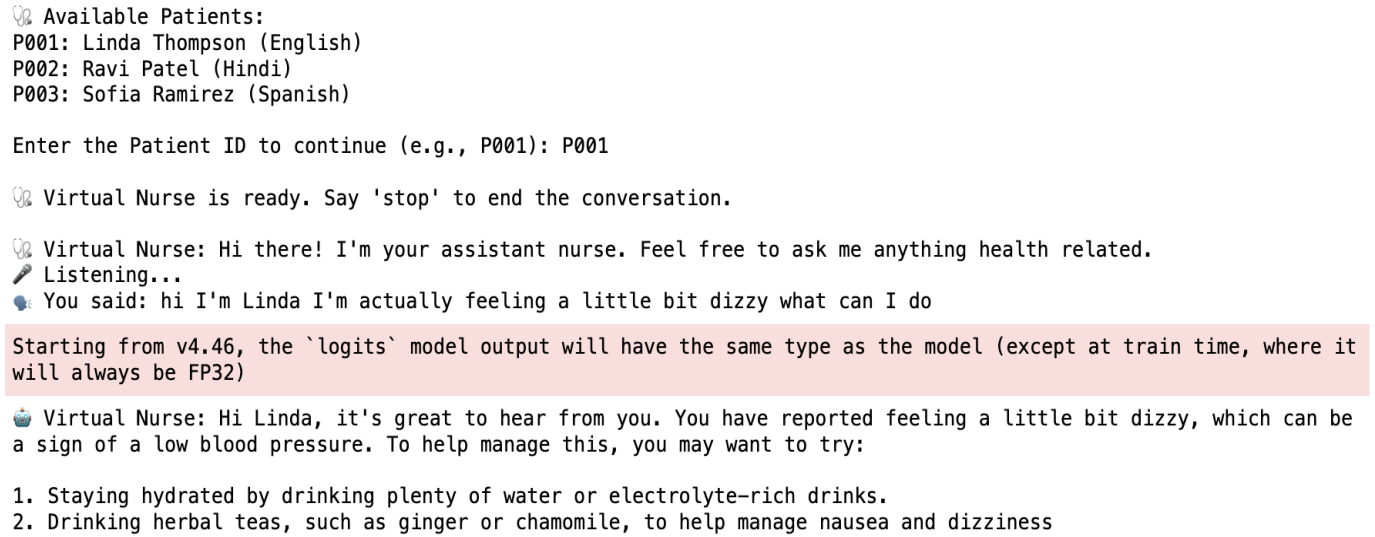
- Foundations That Shaped Our Approach
  - **Hugging Face Transformers:** Provided access to pretrained models like TinyLLaMA, enabling fast integration and testing without custom training.
  - **NLP in Healthcare** – Reinforced the importance of processing health text contextually
  - **LLMs for Health** – Recent work shows small, tuned models can support health applications
- Our Contribution
  - We built a real-time, voice-based assistant that uses structured patient data and a lightweight language model to deliver personalized responses.

Simulated Patient Data

- **Custom Health Profiles**: Demographics like age, gender, language, and reading level.
- **Allergies & Diet**: Food and drug allergies, intolerances, diet type, and restricted foods.
- **Mental Health & Mood**: Daily mood status, stress levels, and depression scores.
- **Clinical Encounters**: Admission/discharge dates, diagnoses, medications, care plans.
- **Lab Results**: BNP, glucose, hemoglobin, and creatinine values with interpretation.
- **PG Health Data**: Daily vitals including weight, blood pressure, heart rate, steps, and sleep.

Data Pipeline Overview

- **Speech Input:** User speaks a question; the system captures audio via microphone.
- **Transcription:** Google Speech API converts the voice input into text.
- **Context Injection:** Relevant structured data is fetched based on Patient ID (profile, labs, etc.) to add context to the query.
- **AI Response:** TinyLLaMA generates a tailored response.
- **Spoken Output:** GTTS (text-to-speech) vocalizes the answer back to the user.



Model Used

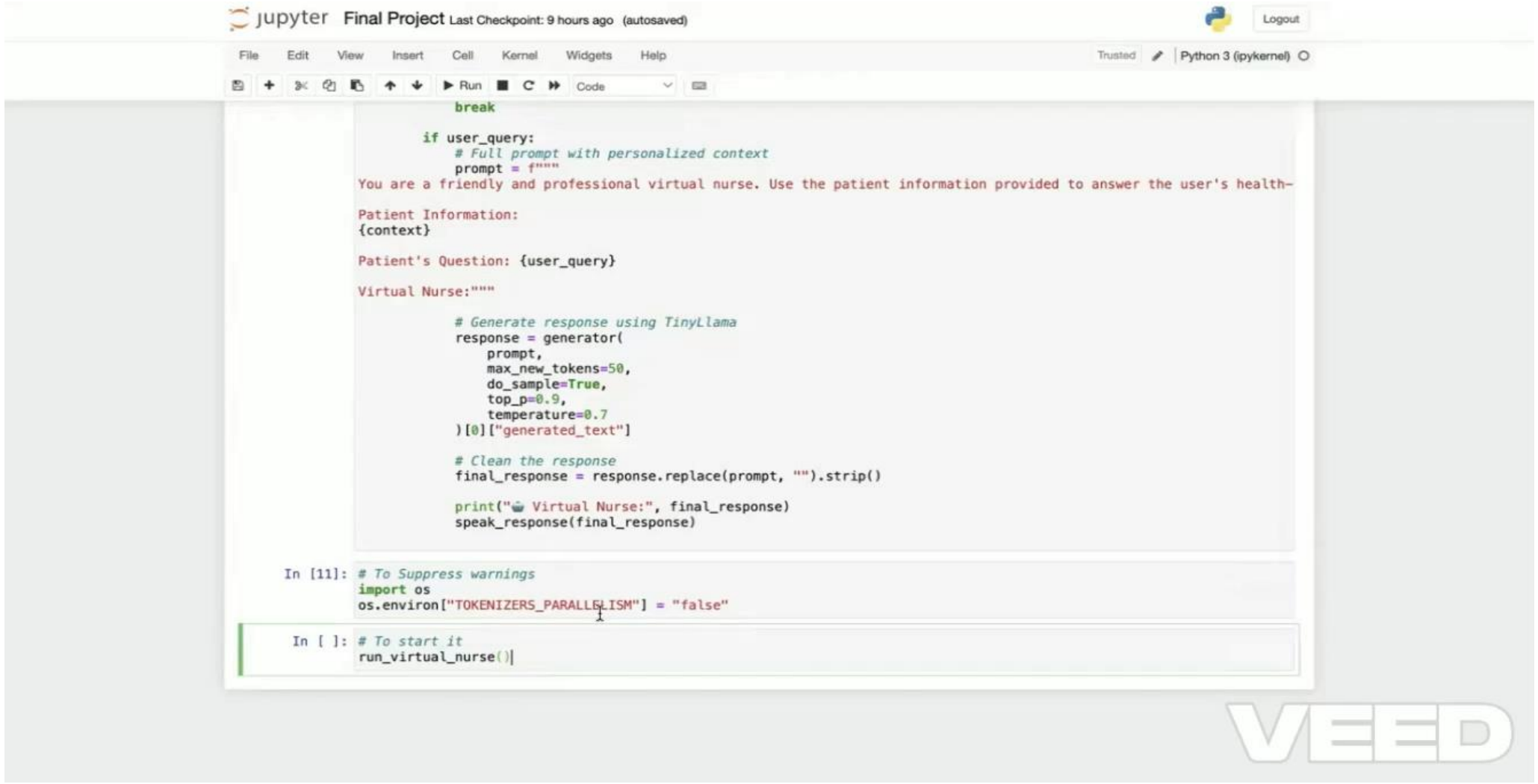
Core Model: TinyLLaMA

- **Model Type**: Decoder-only Transformer
- **Size**: ~1.1 billion parameters
- **Source**: Instruction-tuned, pretrained version from Hugging Face
- **Model ID**: "TinyLlama/TinyLlama-1.1B-Chat-v1.0"
- **Purpose**: Generates conversational responses using structured patient input

Model Integration

- Integrated TinyLLaMA using Hugging Face’s transformers in Python
- Wrapped it into a function that connects voice input, patient data, and response generation
- Prompt includes both user query and structured health context
- Final output is passed to gTTS for spoken response

Project Demonstration



Michigan Tech

```
🩺 Available Patients:
P001: Linda Thompson (English)
P002: Ravi Patel (Hindi)
P003: Sofia Ramirez (Spanish)

Enter the Patient ID to continue (e.g., P001): P001

🩺 Virtual Nurse is ready. Say 'stop' to end the conversation.

🩺 Virtual Nurse: Hi there! I'm your assistant nurse. Feel free to ask me anything health related.
🎤 Listening...
🗣️ You said: hi I'm Linda I'm actually feeling a little bit dizzy what can I do
Starting from v4.46, the `logits` model output will have the same type as the model (except at train time, where it will always be FP32)
🤖 Virtual Nurse: Hi Linda, it's great to hear from you. You have reported feeling a little bit dizzy, which can be a sign of a low blood pressure. To help manage this, you may want to try:

1. Staying hydrated by drinking plenty of water or electrolyte-rich drinks.
2. Drinking herbal teas, such as ginger or chamomile, to help manage nausea and dizziness
🎤 Listening...
❌ Sorry, I couldn't understand that.
🎤 Listening...
🗣️ You said: teach me anything to eat for dinner tonight according to my health records
🤖 Virtual Nurse: (smiling) Of course! I'm happy to help. Based on your health records, I recommend the following h
ealthy dinner options for tonight:

- Vegetable stir-fry with brown rice
- Grilled chicken breast with roasted sweet potato
- Quinoa salad with mixed greens, cherry tomatoes, and cucumber

I hope these suggestions help you feel more comfortable and satisfied with your meal choices. Remember,
🎤 Listening...
```

Performance Metrics

- **Conversational Quality:** TinyLLaMA provided fluent, relevant, and context-aware replies (observed ~8/10)
- **Context Awareness:** Effectively used patient-specific data like lab results, allergies, and mood in responses
- **Voice Recognition Accuracy:** Google Speech API used for transcription (~90–92% accuracy based on known benchmarks)
- **System Efficiency:** Ran smoothly on standard systems (~8-16 GB RAM usage)

Conclusion

- Built an interactive, voice-enabled clinical assistant using AI
- Integrated voice input, patient datasets, and TinyLLaMA for response generation
- Leveraged structured data (labs, diet, mood, etc.) to personalize conversations
- Used Hugging Face pipeline for model deployment and gTTS for voice output
- Delivered fast, context-aware health replies in under 5 seconds

References

- Hugging Face. (2024). *TinyLLaMA-1.1B-Chat-v1.0*. Retrieved from https://huggingface.co/TinyLlama/TinyLlama-1.1B-Chat-v1.0
- Wang, Y., Wang, L., Rastegar-Mojarad, M., Liu, S., Shen, F., Liu, H. (2018). *Clinical information extraction applications: A literature review*. Journal of Biomedical Informatics, 77, 34–49. https://doi.org/10.1016/j.jbi.2017.11.011
- Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., et al. (2023). *Large Language Models Encode Clinical Knowledge*. arXiv preprint arXiv:2305.09617. https://arxiv.org/abs/2305.09617
- Zhai, H., et al. (2021).*AI-enhanced healthcare decision-making: Integrating structured and unstructured data.* https://doi.org/10.1177/1460458211011234

Thank You