Final Report: Development of an Emergency Senior Assistant with NLP and Speech Technologies

1. Introduction

The project addresses the challenge of ensuring effective communication for elderly individuals during emergencies. The goal was to develop a digital assistant using Natural Language Processing (NLP) and speech recognition technologies to help seniors issue voice commands for emergency assistance. This system combines speech-to-text processing, intent recognition, and a chatbot interface for actionable responses, thereby enhancing safety and independence for elderly users.

---

2. Key Components

2.1. Speech Recognition System

- **Objective:** To convert spoken commands into actionable text for emergencies.
- **Implementation Steps:**
    - **Data Preparation:**
        - Collected and preprocessed audio samples of emergency commands.
        - Noise reduction, segmentation, and normalization were applied for clarity.
        - Speech transcription using Google Web Speech API.
    - **Real-Time Recognition:** The system was tested for real-time input through a microphone and pre-recorded audio.
    - **Challenges and Improvements:**
        - Background noise and varying speech patterns affected accuracy, addressed partially through noise reduction and data diversification.

2.2. Intent Recognition and Chatbot

- **Objective:** To map recognized text to predefined emergency actions.
- **Implementation Details:**
    - Used a BERT-based sequence classification model for intent detection.
    - A Gradio-based chatbot was developed for a user-friendly interface.
    - Mapped intents to actions such as notifying caregivers or emergency services.
    - Recommendations include fine-tuning the BERT model on a dedicated emergency dataset and improving error handling for unexpected inputs.
- **Future Enhancements:**
    - Voice input/output for accessibility.
    - Integration of APIs like Twilio for real-time alerts.

- Multi-turn conversation support for better interaction.

2.3. Text-to-Speech (TTS) System

- **Objective:** To generate audible responses for elderly users.
- **Implementation Steps:**
  - Developed a Seq2Seq model in PyTorch to generate Mel spectrograms from input text.
  - Converted spectrograms into waveforms using the Griffin-Lim vocoder.
  - Challenges included limited dataset diversity and the basic nature of the vocoder, suggesting the use of advanced models like Tacotron 2 and WaveGlow for future improvements.

---

3. Results

- A functional prototype was developed, capable of recognizing emergency speech commands and responding through a chatbot.
- Real-time and offline modes for speech recognition were implemented.
- The chatbot successfully maps recognized intents to actions and offers a scalable design for future features.
- The TTS system provides basic speech synthesis but requires improvements for naturalness and clarity.

---

4. Recommendations and Future Work

1. **Fine-Tuning Models:** Fine-tune the intent recognition model using labeled emergency datasets to improve accuracy.
2. **Enhanced Features:** Incorporate multilingual support, real-time API integrations, and improved TTS models.
3. **Testing and Deployment:** Test the system with diverse real-world scenarios and deploy it on secure, scalable cloud platforms.
4. **Accessibility Enhancements:** Implement voice-based interaction and text-to-speech with advanced naturalness for elderly users.

---

5. Conclusion

The project demonstrates the feasibility of integrating speech recognition, NLP, and TTS technologies into a cohesive system for emergency support. With iterative improvements and real-world testing, this system holds the potential to significantly enhance the safety and quality of life for elderly individuals.