

Rafiq: Arabic-Speaking AI Assistant for Blind Navigation

Final Project Report

Mohammad Yusif Albeladi Rashed Alrushod Hasan Shahid Kamal

Abdulrahman Almejna Hamad Alkalbani

Under the supervision of

Muhammad Mubashar

August 2025

Abstract

The Rafiq project successfully developed a real-time, Arabic-speaking AI assistant to enhance the mobility and environmental awareness of visually impaired individuals. The system architecture features a dual-mode operation: a low-latency passive mode for continuous obstacle detection and an on-demand active mode that leverages a Vision-Language Model (VLM) for detailed scene description. Key technical achievements include the implementation of chunked Text-to-Speech (TTS) for responsive audio feedback, pure-math object localization, and the integration of a MiDaS depth model to solve critical navigation challenges like wall detection. The project culminated in a fully functional and optimized backend service connected to an intuitive Flutter application. The final product delivers a robust, high-performance assistive tool, demonstrating a significant advancement in accessible technology.

Index Terms—Assistive Technology, Computer Vision, Blind Navigation, Vision-Language Model (VLM), Real-Time Systems, Depth Estimation, Flutter.

CONTENTS

1 Group Member Contributions **3**

2 Project Objectives and Goals **3**

3 Final Milestones & Achievements **3**

4 Timeline and Schedule **4**

5 Challenges and Obstacles **4**

5.1 Real-Time TTS with Chunked Responses 4

5.2 Calculating Object Positions 4

5.3 Muting Passive TTS 4

6 Conclusion **4**

7 Resources and Tools Used **5**

1 Group Member Contributions

- **Mohammad Yusif Albeladi:** Served as the project lead.
  - Led the overall program development, distributed final tasks for optimization and project finalization, and supervised the work to ensure all project goals were met.
- **Rashed Alrushod:** Led project branding and presentation.
  - Designed the official logo for the Rafiq application.
  - Created the project poster for the final presentation.
- **Hasan Shahid Kamal:** Optimized core navigation logic.
  - Refined the `detector.py` module and implemented new navigation features.
  - Solved critical wall detection issues by integrating a MiDaS depth model, significantly improving directional guidance.
- **Abdulrahman Almejna:** Developed the user-facing application.
  - Created the complete Flutter application for the user interface, ensuring a seamless user experience.
- **Hamad Alkalbani:** Focused on backend performance.
  - Optimized the Vision-Language Model (VLM) pipeline to reduce response time and improve the efficiency of the Active Mode.

2 Project Objectives and Goals

The primary goal of the Rafiq project was to empower visually impaired users with real-time, context-aware navigation assistance. Key objectives included:

- Empowering visually impaired users with real-time Arabic obstacle alerts and directional guidance.
- Achieving low-latency passive feedback and rapid on-demand VLM responses.
- Seamlessly muting passive alerts when answering user queries.
- Supporting both frame-by-frame guidance and free-form, query-driven narration.

3 Final Milestones & Achievements

With the project now complete, the team successfully delivered on all core objectives and several key enhancements:

- **Project Completion:** The core functionalities of both Passive (obstacle alert) and Active (VLM query) modes were fully implemented and integrated.

- **Performance Optimization:** Significant improvements were made to the VLM pipeline and detection algorithms, reducing latency and enhancing responsiveness.

- **User Interface Development:** A complete Flutter application was developed, providing an intuitive interface for users.

- **Advanced Navigation:** Wall detection and directional guidance were greatly improved by integrating a MiDaS depth model.

4 Timeline and Schedule

The project followed its planned schedule, culminating in a final week of optimization and delivery.

Week	Focus
Week 1	Repo & Colab scaffold; YOLO+BLIP prototype
Week 2	Quick Mode end-to-end on stub; passive pipeline verified
Week 3	Best Mode prototype; API integration start
Week 4	Streaming TTS, object-side math, audio control
Week 5	Final Optimization, Flutter App Dev, Poster Design & Project Completion

5 Challenges and Obstacles

5.1 Real-Time TTS with Chunked Responses

Synchronizing partial ChatGPT streams with immediate TTS playback required redesigning the `tts.py` worker to handle HTTP stream chunks and spawn `mpv` in a cancel-capable manner.

5.2 Calculating Object Positions

Accurately classifying left/center/right from raw bounding boxes called for midpoint, overlap ratio, and zone heuristics (`_side_by_zones`), all without external geometry libraries.

5.3 Muting Passive TTS

Ensuring that passive alerts are instantly silenced when Active Mode kicks in forced a rethink of our cancellation logic and the introduction of a global mute timer.

6 Conclusion

The Rafiq project has successfully achieved its goal of creating a real-time, Arabic-speaking AI assistant for blind navigation. The final system integrates passive obstacle alerts and on-demand VLM-powered scene description through a user-friendly Flutter application. Key innovations, including chunked TTS streaming, pure-math object classification, and depth-

model-enhanced navigation, have resulted in a robust and responsive solution. The project concludes with a fully functional prototype ready for user testing, laying a strong foundation for future enhancements in assistive technology.

7 Resources and Tools Used

- **Languages & Frameworks:** Python, Flutter, FastAPI, WebSockets, `asyncio`, OpenCV, PyTorch
- **Models:** YOLOv8-Nano, MiDaS (small), GroundingDINO, BLIP-2, Whisper-Tiny
- **APIs:** ElevenLabs (TTS), OpenAI ChatGPT (GPT-4o, GPT-4V)
- **Utilities:** `mpv` for streaming audio, `uvicorn` server, `python-dotenv`, `opencv-python`, `ultralytics`