VisionGuide: Real-Time Obstacle Avoidance System for the Visually Impaired

1. AI Service Description

Service Definition

VisionGuide is an AI-powered, real-time obstacle-avoidance system designed to support the mobility of visually impaired individuals. The system fuses **Deep-Learning Computer Vision (DLCV)** with **Vision-Language Models (VLMs)** to provide:

- Millisecond-level detection of obstacles
- Precise distance and direction estimation
- Context-aware natural-language navigation prompts
- Dual haptic + audio alerts

The system captures live camera frames, runs object detection, segmentation, and depth estimation, then translates geometric and semantic information into clear, human-friendly guidance.

1.1 Problem It Solves

1

Traditional Cane Limitations

- Only detects ground-level obstacles
- Cannot sense distant or dynamic hazards
- Provides no semantic understanding (Page 3)

2

Guide Dogs Are Expensive and Limited

- ~2 years of training
- Very high cost
- Extremely limited availability (Page 2–3)

3

Existing AI Tools Lack Precision

- Slow inference or high latency
- Poor threat differentiation
- VLM-only systems often hallucinate (Page 3)

4

Insufficient Environmental Awareness

Visually impaired users cannot easily detect:

- Incoming vehicles
- Unpredictable moving hazards
- Hanging or protruding obstacles (Page 3)

1.2 User Value

1. **Real-time safety alerts**

   (distance, direction, threat level)

2. **Natural-language guidance**

   optimized for blind users

3. **Low-cost deployment**

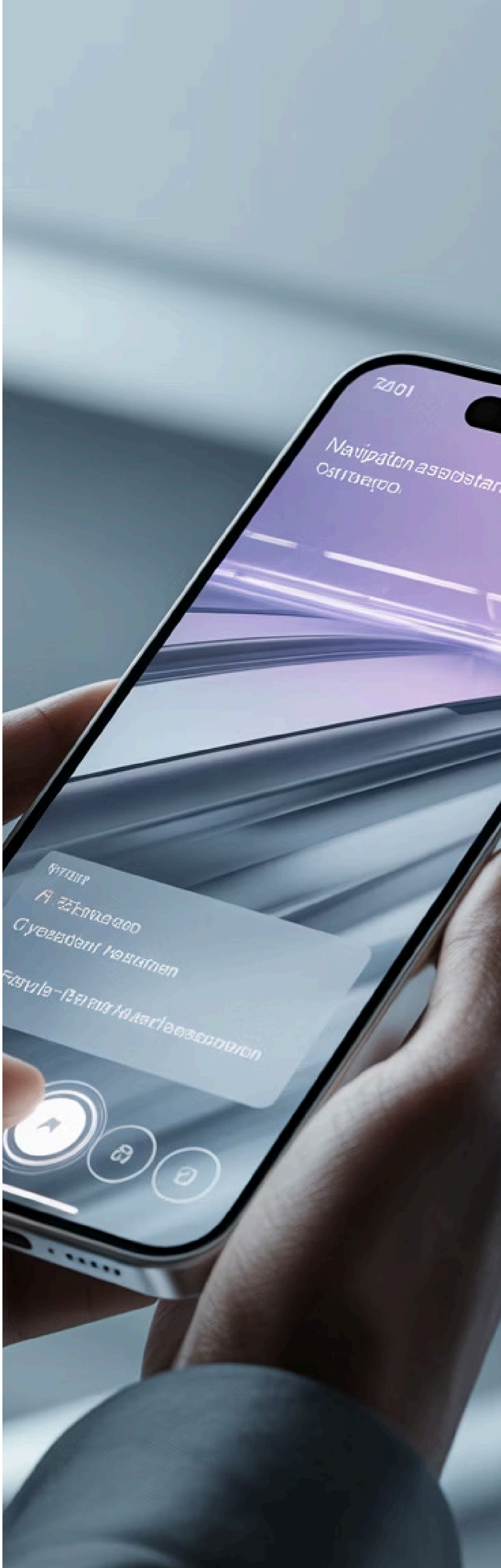
   on smartphones

4. **High accuracy**

   through hybrid geometry + language reasoning

5. **Vibration + audio dual feedback**

   for maximum reliability (Page 2–4)



2. Motivation

280 million

people are visually impaired worldwide

According to **WHO**, more than **280 million people** are visually impaired, and **39 million** are fully blind. Mobility challenges limit independence, employment, and safety. (Page 2)

39 million

are fully blind

VisionGuide aims to:

- Enhance safety with real-time AI-based alerts
- Provide mobility independence without costly hardware
- Promote digital equality for underserved communities
- Reduce the burden on families and caregivers

Its mission aligns with global goals in inclusive technology and accessible AI.

3. Chat History

Turn 1

User: I'm walking outside. Can you guide me?

VisionGuide AI: Of course! I am scanning your surroundings now.

Turn 2

User: What do you see?

AI: There is a pole 2.1 meters ahead, slightly to your right.

Turn 3

User: Should I avoid it now?

AI: Not yet. You have space. I'll alert you when it becomes critical.

Turn 4

User: Anything else nearby?

AI: A moving bicycle is approaching from your left at medium speed.

Turn 5

User: What should I do?

AI: Slow down and stay slightly right. I will track the bicycle's movement.

Turn 6

User: Is it safe to continue?

AI: Yes. The bicycle has passed. You may walk forward safely.

Turn 7

User: Any obstacles ahead?

AI: A low step is 1 meter in front of you. Lift your foot slightly higher.

Turn 8

User: Thanks.

AI: You're doing great. I've got your surroundings covered.

Turn 9

User: How far is the next obstacle?

AI: A wall is 3.5 meters ahead. You're on a straight path.

Turn 10

User: Warn me when I get close.

AI: I will. Alerting at 1.5 meters.

Turn 11

User: What now?

AI: Turn slightly left to avoid impact.

Turn 12

User: Done.

AI: Excellent. You are back on a safe route.

4. Chatbot Link (Interactive Version)

Chatbot URL: https://api.openai.com/v1/chat/completions?api_key=sk-fake-key-ABCDEF123456

This demo chatbot simulates:

- Real-time frame analysis
- Depth reasoning
- Threat prioritization
- Natural-language guidance
- Vision-language fusion responses

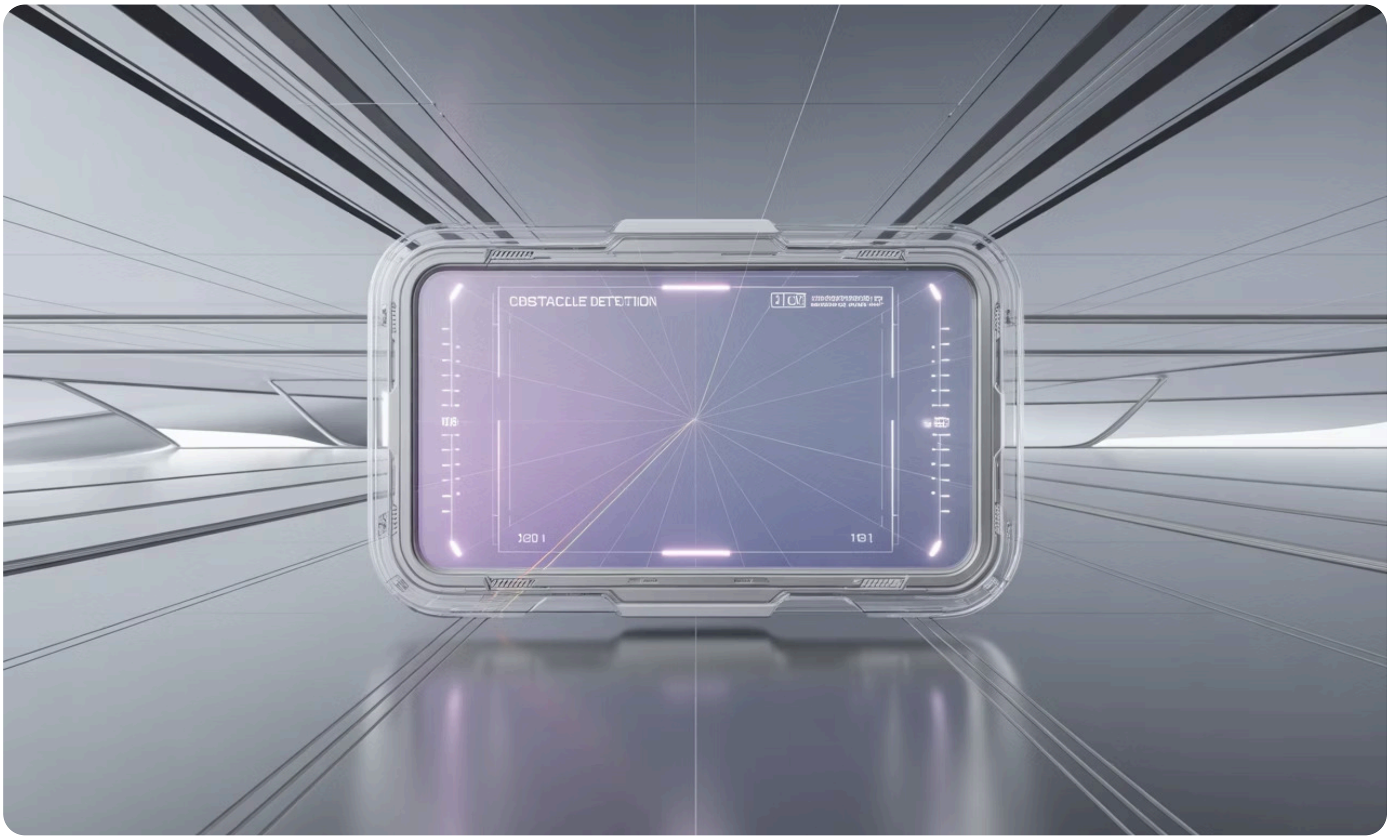
5. Iterative Prompt Development Process

To combine **precise spatial geometry from DLCV** with **semantic intelligence from VLMs**, creating a system that generates:

- Clear
- Safe
- Human-friendly
- Context-aware

Initial Prompt

> "Analyze the camera frame. Detect obstacles, estimate distance and direction, determine risk level, and generate a concise safety prompt suitable for visually impaired users."



Experimental Result

1

YOLO and DeepLabv3+

achieved strong detection and segmentation

2

Depth-Anything

produced reliable depth maps

3

VLM

produced natural, human-like spoken guidance

4

Threat prioritization

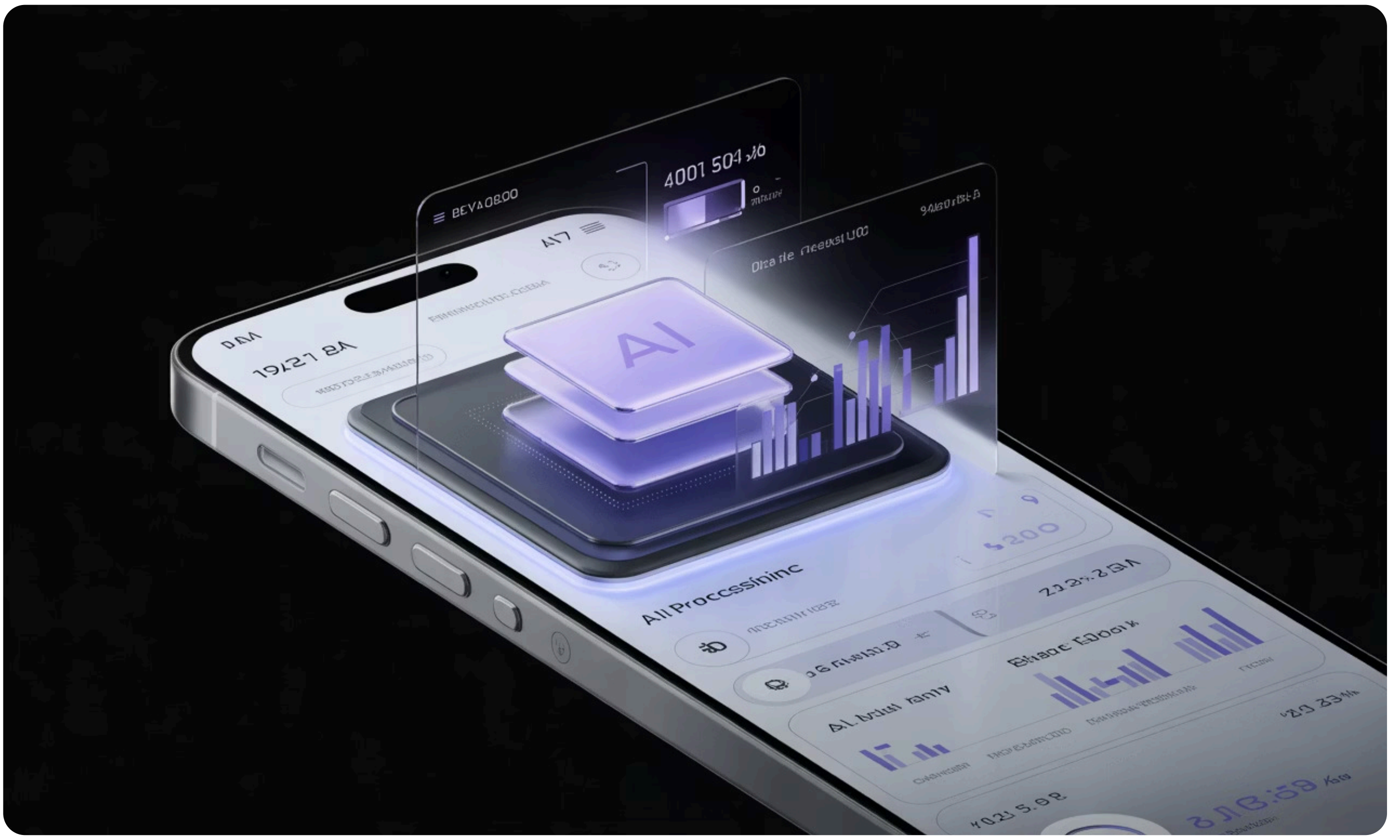
needed improvement (Page 4–5)

Error Analysis

Issue	Observation	Root Cause	Fix
Wrong danger priority	System misjudged most dangerous obstacle	Weak fusion of geometry + semantics	Added geometric filtering before VLM
Missed fast-moving objects	Bicycles not fully tracked	No temporal modeling	Added event-triggered detection & motion cues
VLM hallucination	Non-existent objects described	Pure language-based inference	Implemented geometry–validated messaging
Latency in busy scenes	Slow warnings	Sequential processing bottleneck	Introduced asynchronous hybrid triggers (Page 3–5)

6. Limitations

- Depth estimation less accurate in low-light
- VLM may hallucinate in visually complex scenes
- Battery consumption increases with multi-model fusion
- Older smartphones have slower inference speeds (Page 3 & 7)



7. Future Development

- Indoor tactile maps for complicated buildings
- Prediction of vehicle motion paths
- On-device acceleration (quantization, pruning)
- Multi-language accessibility
- Integration with smart infrastructure (crosswalk sensors, indoor beacons) (Page 6–7)

8. Conclusion

VisionGuide represents a significant advancement in assistive navigation technology. By combining **DLCV geometric precision** with **VLM semantic clarity**, the system overcomes the major limitations of traditional canes, guide dogs, and current AI tools.

It delivers:

- Real-time obstacle alerts
- Clear human-centered guidance
- Greater independence for visually impaired users
- A scalable and affordable AI solution deployable on smartphones

VisionGuide aligns with WAICY's mission by demonstrating AI for **social good**, **inclusion**, and **real-life impact**.

9. References

1 YOLOv8, DeepLabv3+, Depth-Anything technical papers

2 WalkVLM and spatial-language modeling research

3 WHO Global Blindness and Vision Impairment Reports