



Sentinel Vision Navigator

Giving eyes to blind and partially sighted people through live camera understanding, fluent voice guidance, Android handoffs, and opt-in screen control.

AMD Developer Hackathon - Vision and Multimodal AI + AI Agents
Built with AMD GPU training, Akash-hosted Qwen, Expo, Hugging Face, and MI300X research.

Executive Summary

Sentinel Vision Navigator is an assistive AI prototype designed to act as a practical pair of eyes for blind and partially sighted users. It combines live camera analysis, spoken conversation, a 3D-style heads-up display, and an agent layer capable of consent-based handoffs to Android calls, SMS, WhatsApp, email, maps, contacts, location, ride/nearby search workflows, and native Accessibility Service screen actions when the user enables Sentinel Vision Control.

When the user is alone, they can ask: What is in front of me? Guide me out of this room. Is someone approaching? Read that sign. The system responds with short, calm, body-centered guidance. It augments, but does not replace, a cane, guide dog, or orientation and mobility training.

Problem

At least 2.2 billion people globally have near or distance vision impairment, according to WHO reporting, and at least 1 billion cases are preventable or unaddressed. Even with treatment or mobility training, daily navigation still includes difficult moments: unfamiliar rooms, silent obstacles, poorly marked doors, complex sidewalks, people approaching, and text that cannot be read independently.

Existing tools solve fragments. OCR reads text. Maps give long-distance routing. General assistants answer questions. The missing layer is immediate, camera-based understanding of the physical environment.

Solution

Sentinel Vision turns a phone camera into a live multimodal agent. For requests like guide me, get out, avoid obstacles, or which way, it ignores GPS and focuses on what is visible: obstacles, clear path, doors, stairs, walls, signs, people, vehicles, and curbs.

The app retrieves blind-mobility guidance rules, sends the live camera frame to a Qwen multimodal model through Akash, and speaks one safe next step in fluent English.

Core Features

- Live camera-based scene understanding for blind and partially sighted navigation.
- Fluent English voice assistant with short, calm guidance and interruption handling.
- Vision RAG knowledge base for indoor rooms, outdoor sidewalks, stairs, curbs, object finding, and sign reading.
- 3D-style camera HUD with reticle, perspective grid, scan lines, detected object chips, hazard markers, and connector status.
- Android connector setup for camera, speech/mic, contacts, location, email app handoff, WhatsApp handoff, SMS, calls, maps, and native Accessibility settings.
- Multitasking agent actions: call, SMS, WhatsApp handoff, email compose/read handoff, maps directions, nearby search, ride handoff, battery, time, open apps, read visible text, find objects in camera view, describe scenes, and conversation.
- Native Android Accessibility Service: after the user enables Sentinel Vision Control, the APK can read visible screen text, tap visible controls, scroll, use navigation actions, and type into focused fields. It does not secretly access encrypted chats or account APIs.
- Web mode at amdvision.qubitpage.com supports camera, microphone, text fallback, local command handling, and APK download. Browser mode does not control other Android apps.
- Native Android APK path, with Apple/TestFlight path planned after Apple credentials and platform work.

Architecture

Layer	Implementation	Purpose
Web	HTML/CSS/JS, MediaDevices, Web Speech APIs	Immediate browser demo over HTTPS.
Android	Expo SDK 54, React Native, camera, speech, contacts, location, Accessibility Service bridge	Native permissions, OS handoffs, and visible-screen accessibility actions after user opt-in.
AI	Akash ML API, Qwen/Qwen3.6-35B-A3B	Vision reasoning, command routing, fluent responses.
RAG	Embedded navigationKnowledge module	Indoor/outdoor/stairs/text/object safety rules.
Distribution	nginx + HTTPS project hosting	Public app, APK, and document downloads.
Hugging Face	Model page and Space target	Proof-of-concept SentinelBrain checkpoint, deployment target, and build logs.
AMD training	AMD Developer Cloud, MI300X, ROCm/PyTorch	Large-model training experiment, checkpoints, monitoring, and future specialized assistant path.

Providers and Models

- AMD Developer Hackathon stack: AMD Developer Cloud, AMD Instinct MI300X, ROCm path, and AMD/Akash/Hugging Face ecosystem.
- Akash ML API: hosted model API used by web and Android applications.
- Qwen/Qwen3.6-35B-A3B: primary multimodal assistant model for visual understanding, command interpretation, and English responses.
- SentinelBrain-14B-MoE-v0.1: proof-of-concept Hugging Face training artifact. It is part of the competition story and future roadmap, not the production vision model currently serving the app.
- Public distribution site: HTTPS web demo plus APK/PDF downloads.
- Expo/EAS: Android APK build and Apple iOS build path.

SentinelBrain Training Experiment

SentinelBrain-14B-MoE-v0.1 is included as a proof-of-concept AMD GPU training experiment for the competition. The public model page documents the specialized assistant roadmap and preserves training artifacts connected to Sentinel Vision, while the shipped app currently uses Akash-hosted Qwen multimodal inference for live camera reasoning.

The training work used AMD Developer Cloud with AMD Instinct MI300X-class hardware, ROCm/PyTorch, BF16 checkpoints, long-context datasets, watchdog monitoring, and recovery-oriented checkpointing. The architecture explored a 14B+ parameter model with 24 layers, d_model 4096, 32 heads, 4 experts, sequence lengths up to 6K, sharded binary datasets, batch size 1 with gradient accumulation 16, and a planned 8,544-step SFT pass over about 45,578 packed sequences.

The experiment showed the AMD platform is stable and practical for multi-day large-model work, but the model is not claimed as production-ready. One planned 6K SFT pass needs roughly 45-60 hours on a single MI300X-class GPU, and a stronger production model would need more continued training, benchmark gates, safety testing, and separate multimodal/vision/audio components.

Benefits

- Environmental confidence: ask what is ahead without waiting for another person.
- Obstacle awareness: chairs, stairs, curbs, walls, cables, vehicles, people, and blocked paths.

- Social awareness: ask if someone is approaching or standing nearby.
- Text access: read signs, labels, screens, door numbers, and instructions aloud.
- Phone independence: calls, SMS, WhatsApp handoff, email handoff, maps, nearby search, ride handoff, contacts, app launching, and visible-screen Accessibility Service actions become voice-accessible after consent and platform enablement.
- Lower hardware barrier: runs on commodity smartphones and browsers, not only expensive specialized glasses.

Safety and Limitations

Sentinel Vision is an assistive prototype, not a medical device and not a replacement for a cane, guide dog, human assistance, or formal orientation and mobility training. AI vision can be wrong, delayed, or uncertain. The app is designed to be conservative: when unclear, it should tell the user to stop, hold steady, and pan slowly rather than invent a route.

Current connectors use Android app handoffs for email and WhatsApp rather than direct account APIs. Direct Gmail, Outlook, calendar, WhatsApp Business, emergency contacts, and caregiver APIs require OAuth, platform review, and production backend connectors. Web mode can demonstrate camera, voice, text, and local routing, but it cannot control other Android apps; full screen control requires the Android APK and the user-enabled Sentinel Vision Control Accessibility Service.

SentinelBrain-14B-MoE-v0.1 is also not claimed as a finished production model. It is a training experiment and model-progress artifact; production readiness requires more training, benchmark evaluation, safety testing, and modular vision/audio integration.

Roadmap

- Publish and keep current the Hugging Face Space and SentinelBrain model page.
- Continue SentinelBrain training: roughly 45-60 hours for the planned 8,544-step 6K SFT pass, then deeper benchmark gates.
- Add Gmail/Outlook OAuth, calendar, emergency contacts, caregiver dashboard, and WhatsApp Business where approved.
- Add depth estimation, optical flow, and on-device fallback models for safety prompts.
- Run blind and low-vision field testing for latency, accuracy, wording, and false-positive safety warnings.
- Build iOS through EAS and distribute via TestFlight/App Store after Apple Developer setup.

Hackathon Qualification

- Vision and Multimodal AI: primary track because the app processes camera, voice, text, and environment context.
- AI Agents and Agentic Workflows: strong fit because it coordinates perception, conversation, and OS actions.
- Fine-Tuning on AMD GPUs: implemented as a proof-of-concept training experiment through SentinelBrain-14B-MoE-v0.1, AMD Developer Cloud, ROCm/PyTorch, MI300X logs, checkpoints, and training reports.
- Hugging Face Special Prize: target Space deployment under the event organization.
- Build in Public: whitepaper, pitch deck, model-card logs, web demo, APK download, and technical updates.