



intel ai Hackathon

From Ideas to Impact, Accelerate with OPEA.

Powered by **H2S**
HACK2SKILL

Team Name: unnamed_v2 🧑💻

Team Leader Name: Ratul Pal

Theme: Smart Workplace Assistant ✨




Team Members

1. Ratul Pal
2. Sumit Kumar
3. Prateek Ralhan



Brief About the Solution

Objective:

To empower visually impaired individuals to query, comprehend and analyze visual content (images, videos, surroundings) using an intuitive, voice-driven interface backed by multimodal AI -> **Pyxis**  .

How it works:

- User speaks a query related to their visual environment or static image/video feed.
- System performs ASR → Vision analysis → LLM-based Q&A (reasoning models) → Speech response
- Uses both real-time camera feeds and static content (images/videos)

Key Capabilities:

- ✓ Real-time vision understanding
- ✓ Multimodal reasoning (text + vision + speech)
- ✓ Voice-based interaction & feedback
- ✓ Deployable on Intel Edge Devices/cloud infrastructure!

Motivation/Stats/References

💡 Blindness affects over 253 million people globally (WHO) – <https://pmc.ncbi.nlm.nih.gov/articles/PMC5820628/>

📊 90% of visually impaired people live in low-income settings – <https://www.disabled-world.com/disability/types/vision/#:-:text=Over%2090%25%20of%20those%20with%20vision%20loss%20live,do%20not%20have%20access%20to%20eye%20care%20services>

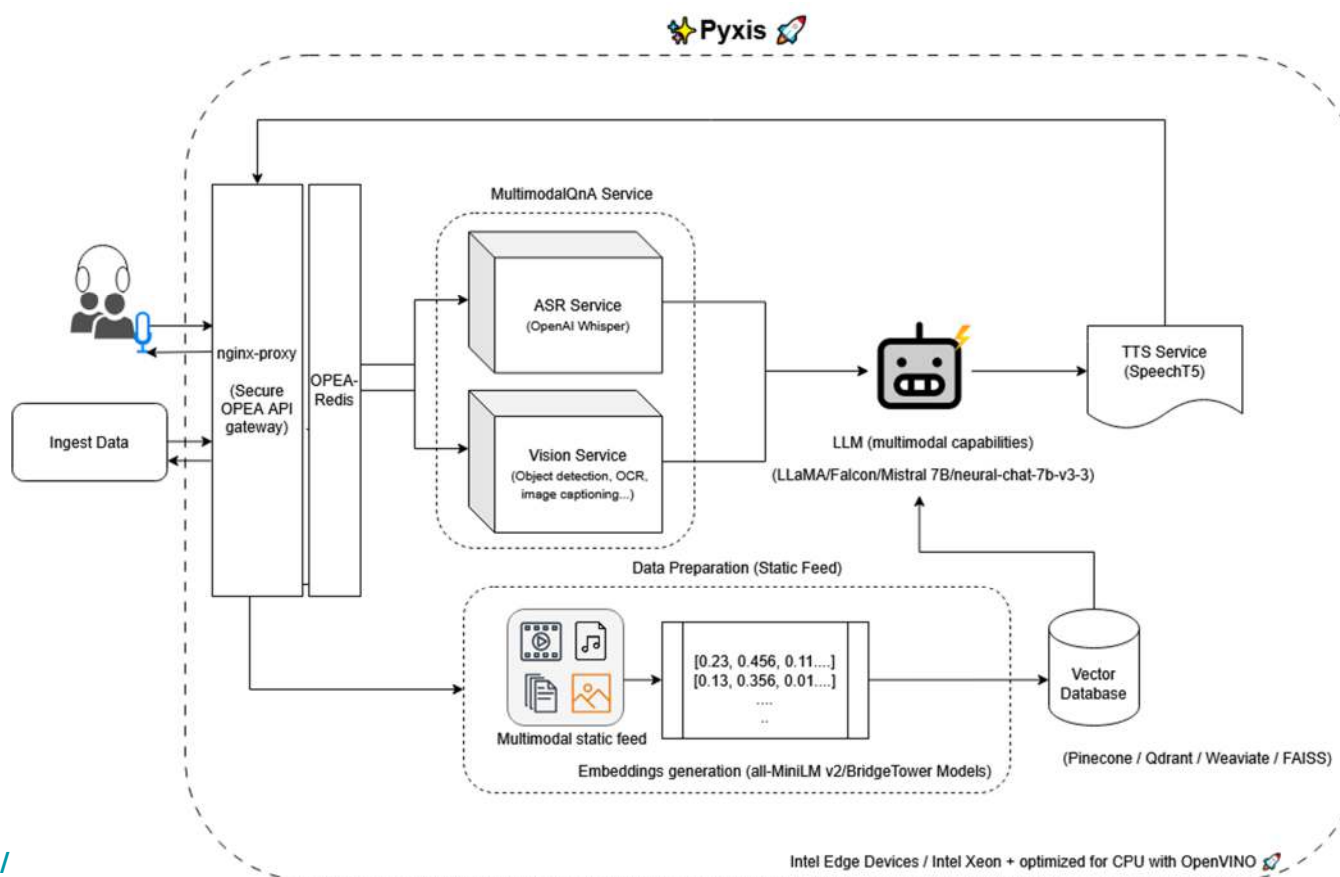
🕒 Navigating and understanding visual context is a daily challenge – <https://www.jessejacquesphoto.com/post/the-essential-skill-of-visual-literacy-a-guide-to-understanding-visual-content>

Enabling independent visual understanding through accessible and low-cost multimodal AI can significantly enhance quality of life, aid in education, mobility, and information access for the visually impaired.

Recent breakthroughs in multimodal LLMs (Mistral, LLaMA, BridgeTower, OpenAI Whisper) make contextual visual reasoning possible!

With **Pyxis** 🚀, we bring the power of multimodal AI to the visually impaired – so they can **ask**, **understand** and **respond** to the world around them, independently!

System Architecture







Designed using:
<https://app.diagrams.net/>

Tech Stack








- 🛠️ OPEA Platform: (API Gateway (*nginx-proxy*), Redis Broker (*opea-redis*))
- 🧠 LLMs: LLaMA, Falcon, Mistral, Neural Chat 7B
- 🎙️ ASR: OpenAI Whisper
- 👁️ Vision: Object Detection, OCR, Image Captioning
- 🔊 TTS: SpeechT5
- 📦 Embeddings: all-MiniLM v2, BridgeTower
- 🧠 Vector DB: Pinecone / Qdrant / Weaviate / FAISS
- ⚙️ Intel Edge Devices: (Xeon CPUs) with OpenVINO toolkit for accelerated CPU inference

Assumptions

-  **Input Modality:** User provides spoken queries; system assumes device has microphone access.
-  **Edge Constraints:** Solution designed for low-latency edge inference.
-  **Vision Feed Availability:** System assumes access to either a live camera or uploaded/stored visual content.
-  **Use Case Scope:** Prioritizing general object understanding, text extraction, and scene explanation.

OPEA platform and microservices usage

How OPEA powers Pyxis :

-  **API Gateway (*nginx-proxy*), Redis Broker (*opea-redis*):** Ensures secure, scalable microservice orchestration, acts as load balancer and prevents overuse/abuse of API calls
-  **ASR Service (OpenAI Whisper):** Converts speech to text
-  **Multimodal LLM (LLaMA, Falcon, Mistral, Neural Chat 7B):** Processes combined text + vision context for Q&A
-  **Vision Service:** Object Detection, OCR (text in images), Image Captioning
-  **TTS Service (SpeechT5):** Responds to the user in natural speech
-  **Embeddings (all-MiniLM v2, BridgeTower) + Vector Search (Pinecone / Qdrant / Weaviate / FAISS):** Enables semantic context matching
-  **Edge Optimization:** OpenVINO models run efficiently on Intel Xeon processors

Novelty and Expected Results

Novelty:

- End-to-end multimodal flow (speech → vision → LLM → speech)
- Optimized for edge deployment (real-time on low-cost hardware)
- Built entirely using OPEA-compliant microservices
- Modular architecture enabling plug-and-play enhancements

Expected Results:

- ~90% ASR + vision Q&A accuracy on benchmark visual datasets
- End-to-end latency under 2s on Intel Xeon edge servers
- Positive User feedback (pilot testing with visually impaired users)



intel ai Hackathon

From Ideas to Impact, Accelerate with OPEA.

Thank you