

Team TheSopranos

ABLER

AI THAT SENSES FOR ALL

So that everyone can sense the world.



Telephone
90739 89733
(Team Rep)

Address
Indian Institute of Engineering Science
and Technology, Botanical Garden
Area, Howrah, West Bengal 711103

Feb 2025

TABLE OF CONTENTS

WHAT MAKES ABLER CAPABLE OF
ENABLING EVERYONE TO EXPERIENCE
THE WORLD 100%



01 The Barriers

02 Our Approach

03 Abler System Design

04 Implementing Abler

05 Potential Pitfalls

06 Growth Scope

Team TheSopranos

THE BARRIERS

CHALLENGES FACED BY USERS

01 Visual Impairments:

- Problem: Navigating **complex environments** is challenging without **real-time awareness** of obstacles.
- Insight: Traditional tools lack dynamic object detection and distance estimation needed for safe navigation.

02 Speech & Hearing Impairments:

- Problem: Converting speech to text and text to speech in **noisy or diverse environments** remains unreliable.
- Insight: A robust, bidirectional communication system is essential to bridge this gap.

03 Limited Motor Capabilities:

- Problem: Touchscreen interfaces are often inaccessible for those with **reduced dexterity**.
- Insight: Gesture-based controls must be intuitive, responsive, and accurate for everyday device interaction.



OUR APPROACH

ABOUT OUR VISION AND MISSION

“A world where everyone can sense and interact with their environment effortlessly.”

VISION

Abler, in its development phase, aims to collate high quality varied training datasets to train the various models required for the specified use cases. The hand-picked models, and learning algorithms after training are exposed through an user-friendly and accessibility oriented UI to the users.

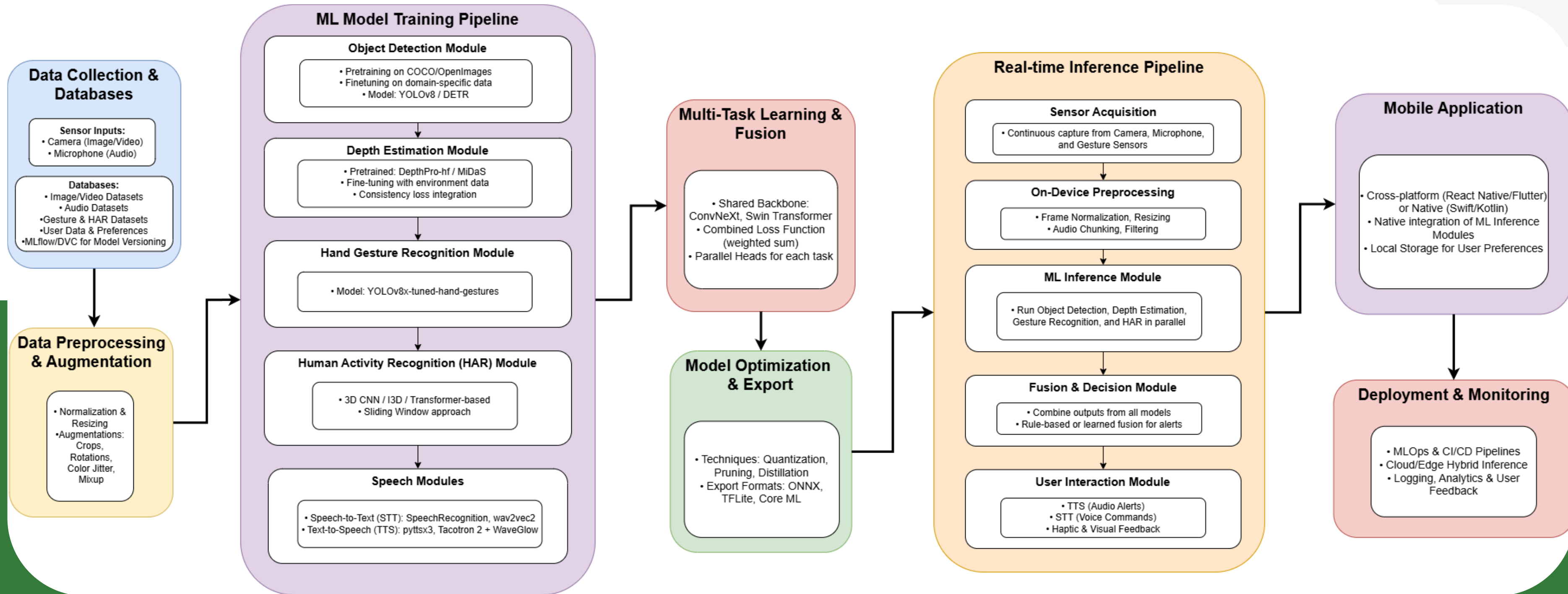
MISSION

Integrate proven open-source models with newly trained, fine-tuned algorithms to create a multimodal AI assistant that:

- Detects and identifies objects in real time using advanced object detection and depth estimation.
- Recognizes hand gestures with high accuracy to enable contactless control.
- Converts speech and text seamlessly for effective communication.

Leverage state-of-the-art training methodologies and continuous testing to ensure robust performance in diverse real-world conditions.

ABLER SYSTEM DESIGN



IMPLEMENTING

ABLER ISN'T JUST A CASTLE IN THE AIR.

ABLER is a fully integrated, end-to-end AI solution that transforms accessibility by merging robust, multi-task machine learning with seamless mobile deployment. Our implementation is built around three core components—each enhanced with modern training techniques and deployed through a modular, scalable architecture.

01 Gesture-Detection

- Uses a **modified YOLOv8x-tuned-hand-gestures** model fine-tuned on **8,000+ gestures** via **curriculum learning** and **knowledge distillation**.

02 Object & Depth Detection

1. Object Detection:

- Dual-Model Strategy: Combines a **high-speed YOLOv8 variant** for rapid obstacle detection with **Facebook's DETR** (ResNet-50 backbone, trained on the COCO 2017 dataset with **118k images**) for precise localization.

2. Depth Estimation:

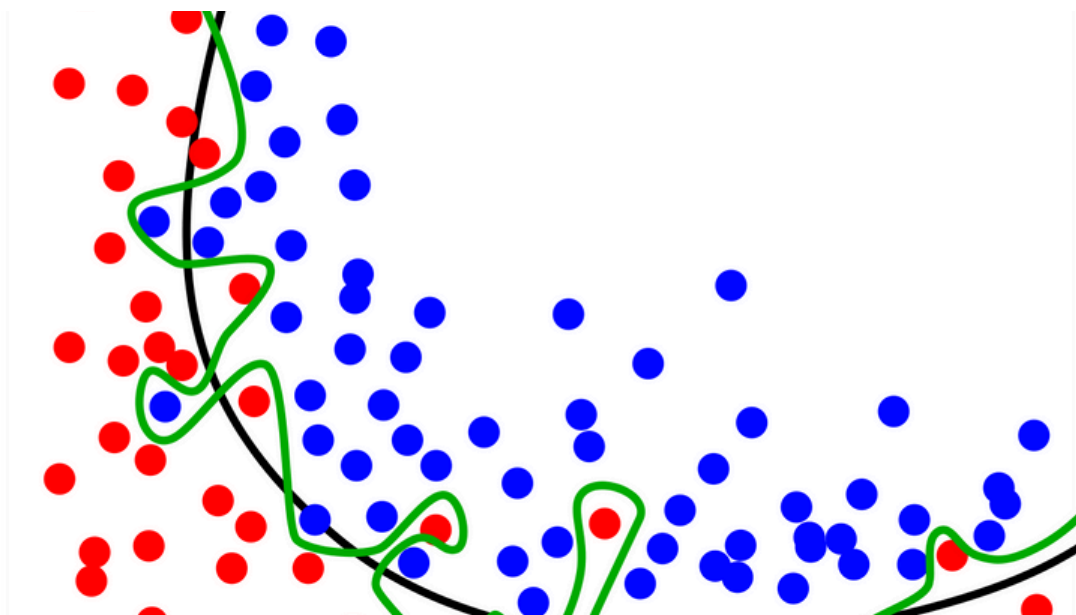
- Enhanced Spatial Awareness: Employs the **Apple DepthPro-hf** model to create real-time depth heatmaps, merging detection outputs with depth data to compute distances and spatial relationships.

03 Speech Processing (STT & TTS)

- **Speech-to-Text (STT)**: Utilizes advanced **ASR models** like **wav2vec2** (via the **SpeechRecognition API**) with **CMUSphinx** as a fallback for accurate transcription in challenging conditions.
- **Text-to-Speech (TTS)**: Converts text into natural, clear speech using engines such as **pyttsx3**.

POTENTIAL PITFALLS

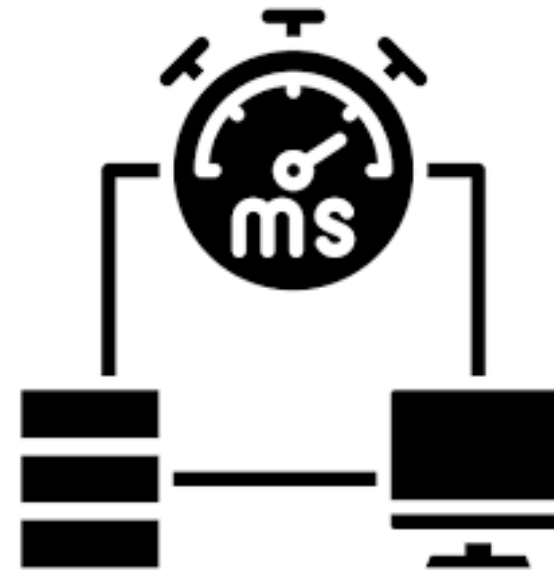
AND HOW ABLER MITIGATES THEM



Overfitting & Data Imbalance:

Issue: Limited domain-specific data may cause models to overfit.

Mitigation: Use extensive data augmentation, regularization, and cross-validation.



Real-time Latency:

Issue: High computational load might delay predictions on mobile devices.

Mitigation: Optimize models via quantization, pruning, and, where necessary, hybrid edge-cloud inference



Environmental Variability:

Issue: Changing lighting, background clutter, and ambient noise can degrade performance.

Mitigation: Incorporate adaptive filtering, online learning, and sensor fusion (e.g., additional depth sensors).

GROWTH SCOPE

AND ABLER WILL GROW!

01 Data Expansion:

Broaden training datasets to include diverse environments, languages, and user demographics for better generalization.

02 Enhanced Sensor Fusion:

Integrate additional sensors (e.g., LiDAR, infrared) to improve depth estimation and refine gesture precision.

03 Adaptive & Online Learning:

Implement real-time model adaptation based on user interactions to continuously enhance accuracy.

04 Multilingual & Cultural Adaptation:

Extend support for multiple languages and regional dialects in both STT and TTS systems.

05 UI/UX Innovations:

Refine the user interface to further reduce cognitive load, improve accessibility, and offer customizable experiences.

06 Advanced Analytics & Feedback Loops:

Incorporate analytics to monitor usage, detect potential issues, and rapidly iterate on improvements based on real-world feedback.

ABOUT US

GET TO KNOW THE TEAM BEHIND ABLER

Our team comprises of
Aritro Shome, Angshuman Roy and Soham
Halдар.

Team TheSopranos

THANK YOU

● SEE YOU IN THE NEXT PHASE!

Telephone

9073989733 (Team
Rep)

Address

Indian Institute of Engineering
Science and Technology, Shibpur,
Howrah

February 2025