



# International Journal of Multidisciplinary Research in Science, Engineering and Technology

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.206**

**Volume 9, Issue 4, April 2026**



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# “Noon Gil” An AI-Powered Real-Time Visual Narrator

**Rucha Chobe, Aishwarya Vasekar, Vijayalaxmi Kawade, Kiran Sarwade, Atter Huma,  
Shobhana J. Raichurkar**

Dept. of Computer Science and Engineering, MITCORER, Barshi, Maharashtra, India

Dept. of Computer Science and Engineering, MITCORER, Barshi, Maharashtra, India

Dept. of Computer Science and Engineering, MITCORER, Barshi, Maharashtra, India

Dept. of Computer Science and Engineering, MITCORER, Barshi, Maharashtra, India

Dept. of Computer Science and Engineering, MITCORER, Barshi, Maharashtra, India

Assistant Professor, MITCORER, Barshi, Maharashtra, India

**ABSTRACT:** Users with a visual impairment face persistent problems of latency, environmental unpredictability and inefficiency in existing digital assistive tools for independent spatial navigation that collectively undermines safety assurance and user confidence. This paper explores navigation and mapping of physical environments using safe hybrid Artificial Intelligence architecture which is based on low-latency edge processing integrated with high-fidelity cloud understanding. The proposed system connects on-the-fly small mean object detection to advanced Large Vision Models (VLMs) through a dual-pipeline process that comprises hazard avoidance and spatial analysis as its core components for automation. The framework almost eliminates the possibility of physical collisions and is easier to navigate because it removes reliance on pure centralized cloud masters while integrating transparent, tactical audio calls that users have real-time access to. We demonstrate significant improvements in security, processing speed, and user confidence over traditional assistive methods. Assistant spatial management could utilize Hybrid Edge-Cloud AI technology as a pivotal method for improving trust, physical integrity, and operational efficiency in real-world navigational processes.

**KEYWORDS:** Assistive Technology, Spatial Mapping, Edge Computing, Distributed AI, Hardware Radar, YOLO Object Detection, Large Vision Models (VLMs), Security and Privacy, Real-time Navigation, Pedestrian Safety, Mobile Applications.

### I. INTRODUCTION

For users who are visually impaired, orientation and mobility instructors, as well as accessibility organizations around the world, the problem of safe independent spatial navigation has become a significant issue. Most traditional environmental verification methods, whether tactile-based (white canes) or centralized digital systems (cloud-only apps), are still susceptible to inefficiencies (both cloudy and touchy round-trip & reply times pro term) as well as cognitive overload. This often results in reduced trustworthiness of digital assistive tools and loss of valuable response time. As the digital economy complicates our urban environments, there is an urgent need for resilience, scalable and instant verification systems.

Hybrid AI technology — especially Edge computing and Cloud VLM-integrated platforms — enables a paradigm shift in the detection, mapping, and communication of spatial data. Leverage the real-time, decentralized, and portable features of local hardware networks to give localized object bounding boxes on a mobile device so that physical paths are always verifiable without needing cloud authority centralization. The processes for scene analysis, optical character recognition and sophisticated requesting are even more automated with high-speed cloud APIs that provide enhanced situational awareness and operational efficiency to the user.

Recent research has demonstrated Edge AI's distinct edge in enhancing the safety confidence of independent mobility by preventing physical collisions through fast verification processes. Unlike the conventional verification systems, this paper explores a robust usage of hybrid Edge-Cloud architecture (NoonGil) on the spatial navigation by iteratively



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

analyzing its speed, accuracy and hazard prevention functionalities. This research aims to provide developers and accessibility organizations with a next-generation model of environmental verification by evaluating existing solutions, identifying gaps, and proposing a tailored framework.

### II. RELATED WORK

Physical environments have historically depended on centralized response systems living within digital cloud frameworks and universally adopted tactile-based tools, rendering them susceptible to spatial blind spots, delayed processing times and inaccurate descriptions. And while digitization has made operations efficient, conventional cloud-only systems rarely feature characteristics like decentralized edge processing and instant offline alerts that expose spatial management to network latency.

Early efforts to address navigational difficulties relied on auditory beacons, GPS routing and secure centralized image-captioning databases but suffered from limited interoperability and single points of failure (e.g., cellular dead zones). Leveraging Edge AI concepts allowed distributed machine learning to provide novel forms of secure environmental management. Physics aware models such as YOLO (You Only Look Once) have reduced human latency from hours to single milliseconds, enabling machine learning architectures like Convolutional Neural Networks (CNNs) that can give independently verifiable evidence of those obstacles without internet middlemen interceding between the primary party and the object/environment using such technology to automate obstacle detection, tracking, and avoidance.

Several studies proposed frameworks for spatial verification based on hybrid approaches. For example, Gangwar et al. employed localized TensorFlow Lite models to build a smartphone-based obstruction system that recorded bounding boxes as instant audio alerts and allowed its users navigate the obstacles with ease. Oluwaseyi et al. 's supplemental research revealed a navigational management system incorporating cloud-based VLMs to enhance contextual awareness, user control, and transparency in both indoor and outdoor environments. Other literature also provides useful examples of joint work on cryptographic "Magic Bytes" detection and strong mobile API, to build tamper-proofing image pipelines.

Despite these progress, battery drainage issue, network latency and ongoing real-world adoption remains challenge. One recent contribution, for example, is the application of multi-modal integrations to address interoperability and data management issues while maximizing the primary elements we love about AI. (2022) The potential of hybrid AI to revolutionize spatial navigation systems is backed by the literature where Edge-Cloud paradigm used as a solution in future work.

### III. EXISTING FRAMEWORK

Today, data is spread in the modern world it stores verification challenges for visually impaired users. However, the challenge of digitizing real-time video streams for remote validation through cellular networks opens the door to potential physical catastrophes. With network speeds changing like the weather, it gets harder to connect users in a given location and the satellites situated behind them to the cloud processing centers. This gap makes it a challenge to map out apparent routes for walking. Typically, this verification part happens in a successive manner which makes for tedious processing and increased likelihood of tripping on undetected obstacles. High speed cloud providers have replaced some applications but without local hard ware processing or localized decision-making techniques; this makes the user less safe.

#### A. Stages Included:

- 1) A user makes a navigational request through a mobile application.
- 2) A high-resolution image of the surroundings is taken by the mobile device.
- 3) Image encoding (e.g. Base64), and transmission over a cellular network to centralized cloud server
- 4) Heavy Vision Language models run on cloud servers to scrutinize the scene.
- 5) The server constructs a descriptive text string of the environment.
- 6) Step 6: The text is sent back to the mobile device.
- 7) The mobile device reads the description using Text-to-Speech (TTS) to the user.
- 8) The user performs a movement, based on this delayed verification report.



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### IV. SYSTEM ARCHITECTURE AND PROPOSED METHODOLOGY

#### A. Summary

**General Architecture** This study presents an easy-to-implement hybrid, safe, and user-friendly spatial verification system that employs React Native, the YOLOv11 Edge framework, Groq Llama 3.2 Vision cloud platform API (as shown in Figure 2) combined with dynamic focal length calibration. On the local mobile hardware side, secure sensor connectivity is enabled (camera and magnetometer) while the Python backend layer made through Flask manages image pre-processing and validation routing. For example, the most efficient issuance of tactical commands enables advanced temporal connection data directly into environmental information systems.

#### B. Roles and Components

- 1) **Mobile Application (Frontend):** Starts on the platform, authenticates user vocal intents using Groq Whisper, and remotely sends frames taken through the camera interface with React Native-managed lifecycles.
- 2) **Fast Brain (Edge/Hardware Radar):** It keeps getting streams of visual frames and it derives object distances simply from the coordinate's units in YOLO bounding boxes processing. If any detected object breaches a safety threshold of 1.5 m, it triggers an immediate "STOP".
- 3) **Slow Brain (Cloud VLM):** Grabs encoded images and ports them through the Groq API for complex contextualization: (edge cases like reading text, currency detection, or room layouts). With this new system, it queries the cloud for confirmation of semantic details rather than being limited solely by local hardware.

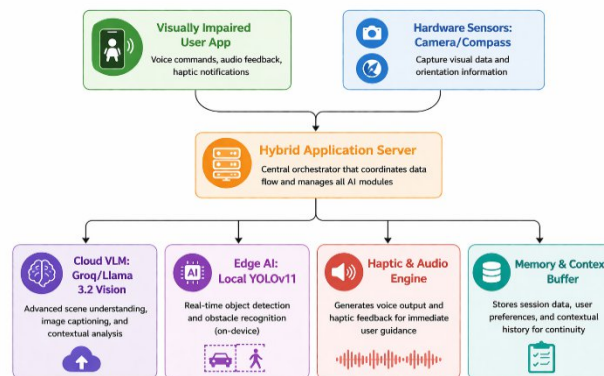


Fig. 1. System Architecture for NoonGil: An AI-Powered Real-Time Visual Narrator

#### C. Layers of Architecture

##### 1) Frontend Layer:

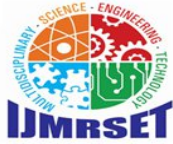
- Built using React Native.
- Uses secure microphone and camera permissions to establish a seamless, safe sensor connection—no complex user menus.
- Systems to voice issue queries allowing users to expose their environment and AI to confirm paths in the physical world.
- Handles Base64 encoding + state management (WALKING\_MODE vs SCENE mode, etc.).

##### 2) Backend Layer:

- Utilize Python, Flask and OpenCV to work with the visual data
- Handles image data with arrays and NumPy for reading and decoding, without need for heavy libraries like PIL
- Enforcing business logic by calling analytical functions like `get_yolo_distances()`, `my_ai_brain()`.

##### 3) Fast Brain / Radar Layer:

- Note all fixed known dimensions of common objects (doors, chairs, laptops).
- Measures distance by taking pixel height/width and comparing to dynamically determined focal lengths.
- Offers tamper resistance to cloud hallucination. When it comes to proximity, safety first — that's the logic behind this one.



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### 4) Slow Brain / Cloud Layer:

- Simplifies complex scene understanding. This will improve the contextual sensitivity of information retrieval where users can ask more general questions about places around them.

### D. Workflow Integration

- 1) The app grabs some audio, uses Groq Whisper and intent detection.
- 2) Visual frames are captured, MIME checked with byte-level signatures and forwarded to processing pipeline.
- 3) The Fast Brain (YOLO), scans the frame for immediate threat.
- 4) When the route is clear, Fire the Slow Brain (VLM) to return a synthesized spatial query.
- 5) The frontend instantaneously shows and verbally announces the verification outcome (clear, blocked, or true to its name).

## V. IMPLEMENTATION

Deployment of hybrid Edge-Cloud system in the context of spatial verification involves several coordinated components, responsible for secure data capturing, impregnable logic and efficient credential validation.

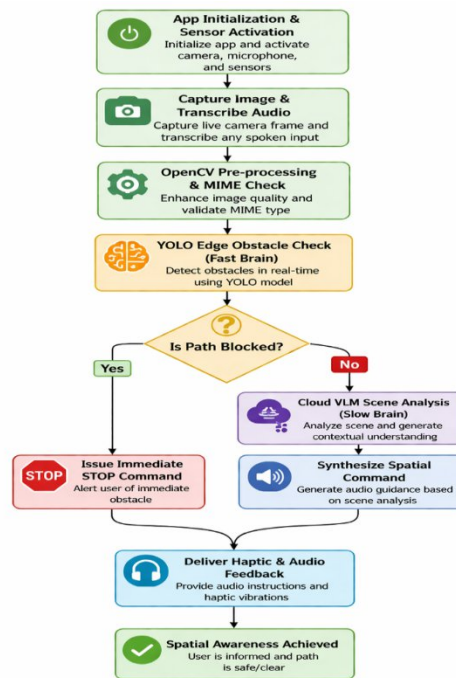


Fig. 2. Proposed Workflow Architecture for NoonGil An AI-Powered Real-Time Visual Narrator

### A. Using OpenCV for Image Processing

Using the OpenCV framework as well, the system generates pre-processing scripts and manages them for image verification.

Two primary optimizations are developed:

- Magic Bytes Detection: Replaces deprecated libraries (img\_hdr, etc.) with raw byte header reading to instantly confirm JPEG integrity.
- NumPy Decoding: Reads byte streams directly and converts into BGR arrays (cv2.imdecode), achieving massive reduction in both RAM spikes and latency on processing, before piping the data forwards to the YOLO model.

To ensure security and correctness, the contracts are locally compiled and tested using Ganache or Truffle Develop. Deployment to the Ethereum test network (such as Sepolia) is automated through deployment scripts (migrations/2\_deploy\_contracts.js).



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### B. Frontend and Audio Processing Integration

The frontend mobile app is built with React Native leveraging Expo AV and Speech APIs for safe microphone access. This is great as it saves them from manually interacting with screen UI and in turn makes it easier for visually impaired people to use the application.

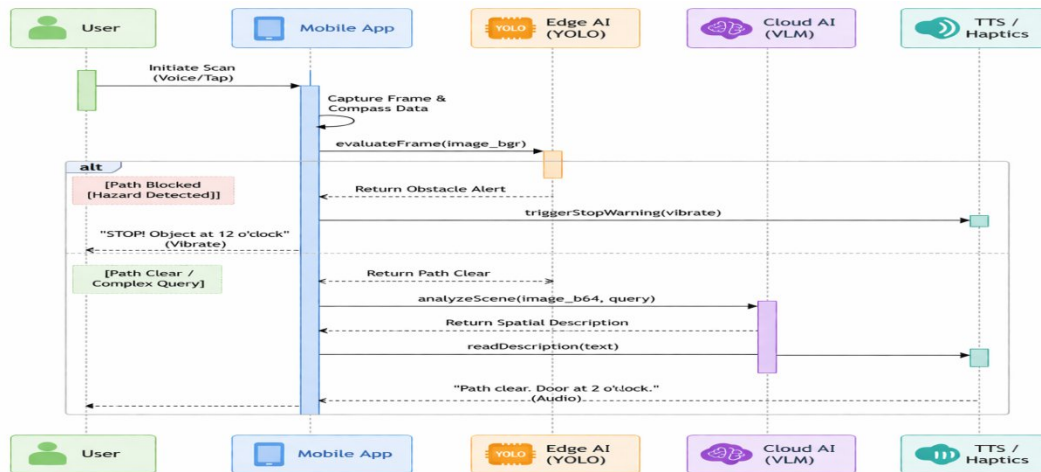


Fig 3. Sequence diagram showing the entire process of real-time spatial analysis and hazard avoidance. It demonstrates Groq Whisper for user voice intent detection, OpenCV pre-processing for image validation, Edge AI (YOLO) interaction for immediate obstacle detection, Cloud VLM (Llama 3.2) processing for complex scene comprehension, and the delivery of tactical audio and haptic feedback initiated by the camera frame capture

TABLE I

ALGORITHM 1: FAST BRAIN (EDGE AI) HAZARD DETECTION

```

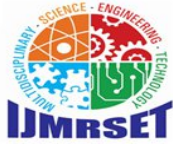
Input: image_frame, safety_threshold (e.g., 1.5m)
1: system calls get_yolo_distances(image_frame)
2: yolo_data := detect_objects(image_frame)
3: min_distance := calculate_closest_obstacle(yolo_data)
4: if min_distance < safety_threshold then
5:   emit TacticalCommand("STOP! Obstacle ahead", true) // true triggers haptic vibration
6: else
7:   return "PATH_CLEAR"
8: end if
    
```

TABLE II

ALGORITHM 2: SLOW BRAIN (CLOUD VLM) SPATIAL ANALYSIS

```

Input: image_frame, user_query, yolo_data
1: system calls analyze_scene(image_frame, user_query, yolo_data)
2: image_b64 = base64_encode(image_frame)
3: if yolo_data = "No known objects" then
4:   radar_instruction := "HARDWARE RADAR: OFFLINE. Rely 100% on AI vision."
5: else
6:   radar_instruction := yolo_data
7: end if
8: description = query_large_vision_model(image_b64, user_query, radar_instruction)
9: emit AudioOutput(description)
    
```



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The frontend communicates with Ethereum using Web3.js or Ethers.js, enabling contract functions such as certificate issuance, verification, and revocation. After issuance, a QR code is generated containing the certificate's unique hash or verification URL. The QR code is embedded in the academic

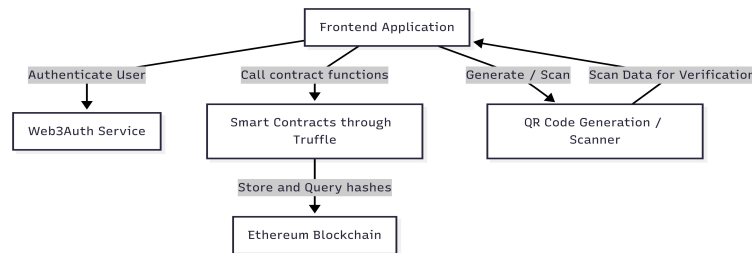


Fig. 4. Component diagram showing the academic certificate verification system's architecture

certificate for easy scanning and sharing.

### C. Verification Using Spatial Prompts

The verification process is designed to be intuitive:

- 1) AI checks whether the user is in WALKING\_MODE or SCENE.
- 2) For when you're walking, the system cuts conversation filler altogether and requires strict templates.
- 3) A logical query (yolo\_is\_empty) is invoked. Even if the local radar does not work, the cloud VLM dynamically takes over distance estimation.
- 4) The backend verifies that the path is genuine, using the 12 o'clock visual corridor
- 5) The user receives immediate results of the verification spoken right to them, giving trust and speed.

### D. Deployment and Backend

Besides frontend camera and audio APIs being the main interacting points with the real physical world, backend services can also be used for logging, monitoring, more complex data processing which may require off-device computing.

A hybrid Edge-Cloud AI architecture controls all operations for scene verification, hazard detection and spatial mapping. The first backend setup is running on a Python 3.13 Flask server, with PyTorch thread usage (torch.set\_num\_threads(1)), and pure Python "Magic Bytes" detection is employed for secure and stable image processing.

Testing is performed via automated scripts that are used to verify things like object distance estimation and VLM query generation. We use end-to-end testing to make sure it creates an effective and non-frustrating navigation experience that combines the React Native front-facing camera, Groq Whisper that responsible for translating voice commands, YOLO running locally for edge processing, and cloud-based Llama 3.2 analysis.

### E. Comprehensive Workflow

Together, these implementations provide a secure and intuitive space through which users can navigate safely while avoiding potential [eyewormhole like] obstacles!

To summarize, the implemented system utilizes bounding box geometry, dynamic VLMs and integrated audio/haptic UI components to create a robust, straightforward spatial verification system accessible to visually impaired stakeholders.

## VI. ADVANTAGES OF HYBRID EDGE-CLOUD AI OVER TRADITIONAL METHODS

Compared to traditional tactile tools and purely centralized digital systems, hybrid Edge-Cloud architecture offers several compelling advantages for spatial navigation and hazard avoidance.

### A. Zero-Latency Collision Avoidance:

Physical objects are tracked in real-time, low-latency environment on local Edge hardware. Fast Brain (YOLO) radar is local logic eliminating cellular network jitter, drop and false data once the app runs, so it became a trustable, unverifiable source of truth.



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### B. Improved Usability and Instant Navigation through Voice Intent Integration:

Embedding Groq Whisper voice intent detection directly into the navigational loop greatly improves the usability of our system. Users can now use natural language to instantly query the Cloud VLM for spatial layouts. Stripping manual touch-screen interactions, allowing real-time spatial mapping, and improving accessibility is a method that renders the verification process convenient and error-free. Our solution, which integrates edge ai safety aspects and on-demand scene comprehension using voice, is much better than conventional solutions that are slow, tedious and error-prone.

### C. Decentralization:

By not relying on a single cloud authority with all its failure points, including malfunctions of the cloud itself, dead zones in the network and high latencies between users, hybrid is decentralized. This distributed architecture provides safety and robustness, as the control and analysis tasks are performed by both the local mobile processor and the remote Cloud VLM.

### D. Preventing Frauds:

Safety thresholds and dynamically estimated focal length always make sure that whatever other AI detects fall into the 4D space, which significantly reduces both cases - physical collision failures between AIs and AI hallucination. Directions between destinations, on the other hand, have extra cred as they use impartial renderings of clock-face directions and measurable distances that would be near impossible for us to miscalculate spatially without detection in a matter of seconds.

### E. Openness and Auditability:

The frames for detecting, avoiding and describing an obstacle are explicit and displayed without conversational fluff. This transparency allows visually impaired users to verify their path's safety without needing to rely on sighted human middlemen.

### F. Speed and Effectiveness:

A dual-pipeline which automates all aspects of image pre-processing, Magic Bytes MIME checking and hazard detection dramatically reduces the workload, computational overheads and lags inherent to traditional cloud-only workflows. This makes faster, real-time spatial verification available anywhere.

### G. Economicalness:

By replacing 24/7 cloud video streams with local automated verification, edge-processing reduce operational costs and saves battery powers and server loads. The system mitigates physical injury risks and eliminates the costs of numerous continuous heavy call API calls to the cloud.

### H. Improved Privacy and Security:

This limits the route of API calls and also ensures that data is processed locally, improving security. Additional approaches to enhance user data privacy include selective disclosure and off-device processing of only complex scenes (Slow Brain) while retaining continuous hazard tracking on-device (Fast Brain).

### I. Worldwide Accessibility:

Mobile Edge AI technology enables roaming physical navigation anywhere on earth, without requiring smart city intermediaries or devices placed throughout the environment such as Bluetooth beacons. This ensures smooth independent navigation for users, which can be very useful to visually impaired people in unknown and foreign condition. To sum up, the double-layered Edge-Cloud structure overcomes the limitations of traditional assistive navigation technologies and provides a stable digital tool for spatial organization by providing thus a robust secure efficient alternative.

## VII. FUTURE WORK

While the current system clearly shows how effective it is to combine Edge YOLO object detection with Cloud VLM scene understanding, there is still plenty of room to improve its capabilities, scalability, and overall user experience in future developments.



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### A. Multi-Language Support Feature

Future versions will include strong multilingual speech-to-text (like Groq Whisper) and text-to-speech systems. By supporting different languages and regional dialects, the platform can become more globally accessible, allowing users to interact with the AI navigation assistant in their own language comfortably.

### B. Facial Recognition and Social Context Awareness

The system can be extended with advanced biometric vision models to offer secure, optional facial recognition. This would help users identify familiar people such as friends, family, or colleagues, while also providing insights into social cues like facial expressions, improving social interaction and independence.

### C. Wearable Hardware Integration

Support for wearable devices such as smart glasses or chest-mounted cameras can make the system more practical. This removes the need to hold a phone constantly and allows for a more natural, continuous, and hands-free navigation experience.

### D. Advanced Indoor Micro-Navigation

Since GPS works mainly outdoors, future work will focus on indoor navigation using technologies like LiDAR and Wi-Fi RTT. This will help users move step-by-step through complex indoor spaces such as malls, airports, and metro stations with greater precision.

### E. Improved Mobile Applications and User Experience

Future implementations could include mobile applications designed for students, institutions, and employers. These apps would allow certificate issuance, QR code scanning, and verification with enhanced accessibility. Technologies such as Augmented Reality (AR) and Near Field Communication (NFC) could further modernize certificate verification work-flows.

### F. Fall Detection and Emergency SOS

By combining motion sensor data (like gyroscope and accelerometer) with visual input from Edge AI, the system can detect sudden falls or emergencies. It can then automatically trigger alerts to emergency contacts or authorities, sharing real-time location for quick assistance.

### G. Personalized AI and Routine Learning

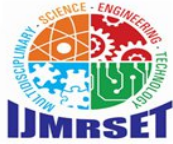
Future improvements may include continuous on-device learning, allowing the system to remember frequently visited places like home or workplace layouts. This personalization can reduce processing time and provide more accurate, context-aware navigation guidance over time.

## VIII. CONCLUSION

Integrating hybrid Edge-Cloud AI into assistive technologies marks a major step forward in making spatial navigation safer and more efficient for visually impaired users. Traditional systems often struggle with issues like high latency, unpredictable environments, reliance on human assistance, and delayed hazard alerts. By leveraging low-latency processing, a dual-pipeline design, and advanced scene understanding, this approach ensures that users can interpret and navigate their surroundings independently, without depending on sighted individuals.

According to this study, smart contracts significantly lower the risk of fraud by automating the issuance and verification of certificates. When combined with QR code-enabled scanning, the blockchain's integration of certificate hashes offers a simple, instantaneous, and internationally accessible verification method. Furthermore, Web3Auth integration improves security and usability by streamlining wallet management and user authentication.

All things considered, the system in place offers a safe, effective, and reliable platform for managing academic credentials that is advantageous to employers, institutions, and students. Blockchain-based certificate verification could establish new international standards for credential trustworthiness and educational administration with additional development, such as increased privacy, scalability, and institution-wide adoption.



## International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### LIST OF ABBREVIATIONS

#### TABLE III

#### LIST OF ABBREVIATIONS USED IN THIS PAPER

Abbreviation	Full Form
AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
BGR	Blue, Green, Red (Color Space)
CNN	Convolutional Neural Network
GPS	Global Positioning System
JPEG	Joint Photographic Experts Group
MIME	Multipurpose Internet Mail Extensions
NFC	Near Field Communication
RAM	Random Access Memory
RTT	Round Trip Time
SOS	Save Our Souls (Emergency Distress Signal)
TTS	Text-to-Speech
UI	User Interface
UX	User Experience
VLM	Vision Language Model / Large Vision Model
YOLO	You Only Look Once
QR	Quick Response

### REFERENCES

- [1] Mourya, A. et al. (2022). Development of a Mobile App to Support the Mobility of Visually Impaired People.
- [2] Sharma, L. et al. (2020). IoT-Based Assistive Device for Deaf, Dumb, and Blind People.
- [3] Jo, S. et al. (2022). Second Level Face Recognition from Detected Objects.
- [4] Sari, F. (2021). Accessible Design and UX for AI Assistive Systems.
- [5] Vahdani, E. et al. (2023). Vision-to-Language Models for Visually Impaired Users.
- [6] Ultralytics (2024). YOLOv8: Real-Time Object Detection Framework.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | [ijmrset@gmail.com](mailto:ijmrset@gmail.com) |

[www.ijmrset.com](http://www.ijmrset.com)