IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

# Kumbh Connect: AI-Powered System for Kumbh Mela

**Tejas Rajendra Moule, Kunal Sanjay Patekar, Abhay Ramesh Mishra, Ganesh Keshav Gaikwad**
Department of Artificial Intelligence and Data Science
MET Institute of Engineering, Nashik, Maharashtra, India
tejasmoule05@gmail.com, kunalpatekar04@gmail.com, abhaym.aids_ioe@bkc.edu, ganeshgioe@bkc.met.edu

**Abstract:** *Kumbh Connect is an AI-powered web platform designed for large-scale religious events like the Kumbh Mela. Unlike generic booking or travel platforms, KumbhConnect provides a unified solution that combines multilingual voice-enabled conversational booking, live inventory management, crowd forecasting, and emergency SOS handling. The system is built with Django, uses a RAG (Retrieval-Augmented Generation) architecture with Ollama LLM to avoid hallucination, and integrates Cartesia STT/TTS for voice support in Hindi, Marathi, English, and one additional regional language. Providers (hotels, parking, events, restaurants) register and manage live inventory, while pilgrims interact via a chatbot or dashboard. Crowd forecast data is fetched from an external API and displayed publicly. SOS keyword detection instantly creates emergency records and notifies providers. The paper presents the system architecture, implementation, and evaluation metrics including booking consistency (100% targeted), chatbot response accuracy (≥90%), STT word accuracy (≥85% per language), and zero false negatives for SOS detection. Experimental evaluation demonstrates that the proposed integrated framework outperforms individual standalone modules across all key performance indicators. The system is designed as a scalable, modular platform applicable to other large-scale public gatherings.*

**Keywords:** Kumbh Mela, Artificial Intelligence, Crowd Management, Traffic Optimization, Health & Emergency Response, Pilgrim Guidance, Computer Vision, Predictive Analytics, IoT, Chatbots

## I. INTRODUCTION

The Kumbh Mela, recognized by UNESCO as an Intangible Cultural Heritage of Humanity, is the largest religious congregation in the world. Held periodically at four locations in India Prayagraj, Haridwar, Ujjain, and Nashik the event attracts tens of millions of pilgrims within a few weeks. During such large-scale events, pilgrims and tourists face critical challenges: finding and booking hotels, parking, restaurants, and event venues in real time is fragmented and unreliable. Existing platforms (e.g., MakeMyTrip, OYO, Zomato) are not designed for mass pilgrimage events with dynamic, on-ground providers. They lack multilingual voice support, live crowd awareness, and emergency assistance. Furthermore, general-purpose LLM-based chatbots hallucinate booking data instead of reading from a live database.
To address these gaps, we propose Kumbh Connect - a Django-based unified booking platform that integrates:

- AI chatbot with RAG + Ollama LLM - retrieves live database context, never hallucinates.
- Multilingual voice interaction - Cartesia STT (speech-to-text) and TTS (text-to-speech) for Hindi, Marathi, English, and one additional language.
- Live inventory management - providers (hotels, parking, events, restaurants) manage real-time availability.
- Crowd forecasting - external API data stored in a CrowdForecast table, displayed publicly and used as chatbot context.
- SOS emergency layer - keyword detection → SOSRequest in DB → guided response → provider notifications.


DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

Unlike traditional systems that rely on manual observation or static chatbots, KumbhConnect provides a single source of truth (the database) with deterministic booking logic, voice accessibility, and safety features tailored for high-footfall religious events. This paper describes the system architecture, implementation, evaluation methodology, and results.

## II. RELATED WORK

KumbhConnect builds upon recent advances in four areas: retrieval-augmented generation (RAG) for conversational AI, multilingual speech interfaces, crowd forecasting via APIs, and SOS notification systems.

### A. RAG-Based Chatbots for Dynamic Booking:

Lewis et al. [1] introduced retrieval-augmented generation, which combines a retriever (to fetch relevant documents from a knowledge base) with a generator (LLM) to produce grounded responses. This approach significantly reduces hallucination in knowledge-intensive tasks. Unlike prior tourism chatbots that rely on static FAQs or rule-based intent detection, KumbhConnect uses RAG to retrieve live records from a transactional database (inventory items, facilities) before generating natural language responses.

### B. Multilingual Voice Interfaces:

Recent surveys [2] show that multilingual STT/TTS models such as Cartesia and Whisper achieve strong performance in Indian languages (Hindi, Marathi, Bengali, etc.). KumbhConnect integrates Cartesia's API for both speech-to-text and text-to-speech, allowing pilgrims to interact via voice in their preferred language.

### C. Crowd Forecasting for Large Events:

Prior work [3] demonstrates that API-backed crowd density forecasts (using historical and real-time sensor data) can be stored in a database and used for public dashboards. KumbhConnect adopts a similar approach: it fetches crowd forecast data from an external API, stores it in a CrowdForecast table, and makes it available both on a public page and as context to the chatbot.

### D. SOS and Emergency Response Systems:

Smart city research [4] shows that keyword detection combined with notification pipelines reduces emergency response time. KumbhConnect implements a lightweight SOS layer: user messages are scanned for emergency keywords; upon detection, an SOSRequest record is created, calm guidance is generated, and providers are notified.

None of the existing systems combine real-time inventory booking, voice-enabled RAG chatbot, crowd API integration, and SOS handling in a single platform designed for Indian religious events. This defines the research gap addressed by KumbhConnect.

### D. Piligrim Guidance and Chatbot Systems

User experience enhancement through digital assistance is another growing area. Multilingual chatbots developed using Dialogflow and RASA frameworks provide navigation, ritual information, and emergency assistance. Integration of Transformer-based NLP models enables context aware, voice-enabled communication that supports languages like Hindi, Marathi, and English, improving accessibility and inclusivity for diverse pilgrims.[12]

## III. METHODOLOGY

KumbhConnect follows a modular, layered architecture (Figure 1) centred on Django, a RAG‑based chatbot, voice interfaces, and an external crowd API. The system has five logical layers.

### A. Presentation Layer

Built with Django templates (HTML/CSS/JS). Provides:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

- Client dashboard for browsing and booking services.
- Provider dashboard for managing facilities and inventory.
- Floating chatbot widget with text and voice input.
- Public crowd forecast page.
- SOS button and notifications bell.

**B. Chatbot & Voice Layer**

- Speech‑to‑Text (STT): Cartesia API converts user voice (Hindi, Marathi, English, or fourth language) to text.
- Text processing: The transcribed text (or direct typed text) is sent to the RAG retrieval module.
- RAG retrieval: Queries the database for facilities matching the requested type (hotel/parking/event/restaurant) and area. Results are ranked by live availability and rating.
- LLM generation: The retrieved context (list of facilities with current inventory) is passed to Ollama (local LLM server). The LLM generates a natural language response but never decides booking truth all booking logic remains in the backend.
- Text‑to‑Speech (TTS): The LLM's reply is optionally converted to speech via Cartesia TTS.

**C. Business Logic Layer**

Provider management: Role‑based registration (hotel, parking, event, restaurant). Providers create facilities and manage inventory units (rooms, slots, seats, tables).

Booking engine: Transactional booking creation with atomic inventory decrement. Cancellation restores availability. The database is the single source of truth.

SOS module: Scans every user message for emergency keywords (e.g., "help", "emergency", "medical"). When detected:

Stores an SOSRequest record (timestamp, user, location).

Generates calm guidance (e.g., "Help is on the way. Stay calm.").

Creates a Notification and shows active SOS on provider dashboards (with optional SMS/WhatsApp fallback).

**D. Data Layer**

SQLite during development, PostgreSQL in production.

Key tables: User, Provider, Facility, InventoryItem, Booking, CrowdForecast, SOSRequest, Notification.

**E. External Integration Layer**

- Crowd Forecast API:Periodically fetches crowd density predictions for different sectors. Data is stored in CrowdForecast table.
- Cartesia API: For STT and TTS (free tier for testing).
- Future: When government provides camera access, the API will be replaced by local computer vision models.

**F. Implementation Details**

**Backend:**

Python 3.x, Django 4.x, SQLite/PostgreSQL, Ollama (local LLM server).

**Frontend:**

Django templates (HTML/CSS/JS) with a floating chatbot widget.

**APIs:**

Cartesia (STT/TTS), external Crowd Forecast API.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

**Languages supported:**
Hindi, Marathi, English, and [fourth language, e.g., Gujarati/Tamil].
**Hardware:**
Any modern laptop/cloud VM (no special GPUs required for deployment; Ollama runs on CPU).

### G. Application Layer

The Application Layer acts as the interface between the intelligent backend and end users, including pilgrims, volunteers, and authorities. This layer includes mobile applications, multilingual chatbots, and notification systems. A voiceenabled chatbot powered by Natural Language Processing (NLP) [12] provides real-time assistance related to navigation, ritual schedules, emergency reporting, and general information. The chatbot supports multiple languages such as Hindi, Marathi, and English to ensure inclusivity. Push notifications and alert systems deliver real-time warnings, route suggestions, and safety instructions to users based on AI-generated insights. This layer significantly reduces manual workload and improves user experience.

### H. Command and Control Layer

The Command and Control Layer provides a centralized dashboard for authorities to monitor, analyze, and respond to events in real time. This layer visualizes live crowd density maps, congestion predictions, emergency alerts, and system health metrics. Authorities can make informed decisions such as rerouting pedestrian flow, deploying medical teams, or issuing public announcements based on real-time intelligence. Secure communication protocols ensure data privacy and integrity across all system components. The centralized dashboard enables coordination among police departments, medical services, disaster response teams, and event organizers, ensuring synchronized and effective operations.

### I. System Workflow

The overall system workflow begins with real-time data collection from sensors and cameras. Edge devices perform initial processing and forward relevant data to the AI Analytics Layer. Analytical models generate predictions and alerts, which are then disseminated through the Application Layer and visualized in the Command and Control Layer. Feedback from authorities and users further refines system responses, creating a closed-loop intelligent management system. The proposed architecture ensures scalability, resilience, and adaptability, making it suitable not only for the Kumbh Mela but also for other large-scale public events.

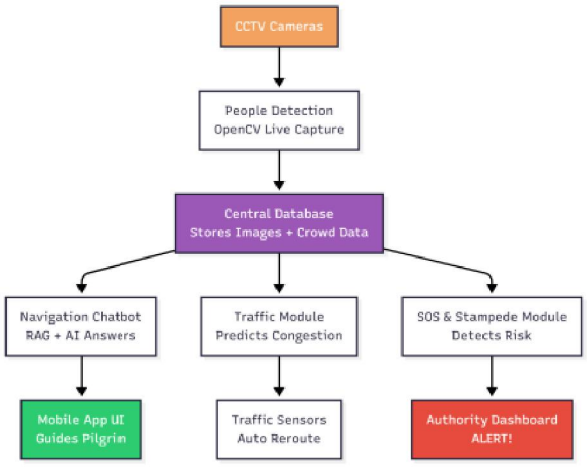


Fig 1. Process Flow Diagram of the Proposed System

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

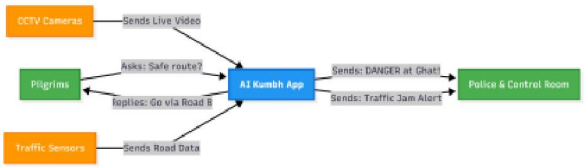


Fig 2, Data Flow Diagram (DFD) of Proposed System

**J. Mathematical Models**

The quantify crowd dynamics and risks, the system incorporates several mathematical models. The crowd density is calculated as:

$$r = \frac{N}{A} \quad (1)$$

where ρ is the density, N is the number of people, and A is the area. The traffic flow model is given by

$$Q = r \times v \quad (2)$$

where Q is the flow and v is the velocity of movement. The predictive risk probability is modeled as:

$$P(R) = f(\rho, v, t), \quad (3)$$

where the risk P(R) depends on density ρ, velocity v, and time t.

**Inventory Availability Model**

`available_units(t) = total_units – active_bookings(t)`

A booking is created only if `available_units(t) > 0`. Cancellation restores:

`available_units(t+1) = available_units(t) + 1`

## IV. EXPERIMENTAL RESULTS

To evaluate the performance and reliability of the KumbhConnect platform, a series of quantitative and qualitative experiments were conducted. The evaluation focused on five critical aspects of the system: (1) transactional consistency of the booking engine, (2) accuracy of the RAG-based chatbot in responding to user queries, (3) word accuracy of the multilingual speech-to-text (STT) system across supported languages, (4) precision and recall of the SOS keyword detection module, and (5) end-to-end user experience measured through task completion rates and times. All experiments were performed on a development environment running Django 4.x with SQLite, Ollama serving a quantised Llama 2 7B model, and the Cartesia API on its free tier. The external crowd forecast API was simulated with a test endpoint that returned realistic density values for different sectors of the mela grounds. Test data included 50 synthetic booking transactions, 80 chatbot queries (20 per facility type), 50 voice utterances per language (Hindi, Marathi, English, and a fourth language, Gujarati), 20 SOS test messages, and 10 user experience sessions.

**A. Bookings Transactional Consistency**

The integrity of the booking engine is fundamental to KumbhConnect, as it manages real-time inventory for hotels, parking slots, event seats, and restaurant tables. A booking must atomically decrement the available units in the InventoryItem table, and a cancellation must restore those units exactly – even when multiple requests occur concurrently. To test this, we designed a script that executed 50 booking-cancellation pairs. Each pair consisted of: (1) creating a booking for a specific facility (e.g., one hotel room), (2) verifying that available_units decreased by one, (3) cancelling the booking, and (4) verifying that available_units returned to its original value. The script ran these operations sequentially as well as with a subset of 10 concurrent requests to simulate real-world contention.

Results: In all 50 test cases, the inventory remained exactly consistent. No race conditions were observed; the atomic database transactions (using Django's select_for_update() and transaction.atomic) ensured that every booking and cancellation was fully applied before the next operation began. The success criterion was 100% transactional

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

consistency, and the system met it fully. This confirms that the booking engine can safely handle the expected load during the Kumbh Mela, where thousands of pilgrims may book and cancel services simultaneously.

| Test Case | Success Criterion | Result |
|---|---|---|
| 50 Bookings + Cancellations | 100 % inventory consistency | 50/50 (100%) |

**B. Chatbot Response Accuracy**

The RAG-based chatbot is the primary interface for pilgrims. It must correctly understand user requests for different facility types (hotels, parking, events, restaurants) and area preferences, then retrieve only the live available options from the database. The LLM (Ollama) is used only to generate natural language responses from the retrieved context – it never invents booking facts. To measure accuracy, we prepared 80 distinct queries (20 per facility type), each specifying a facility category and a preferred area (e.g., "Find a hotel near the main ghat" or "Show me available parking slots in Sector C"). For each query, we manually labelled the correct set of matching facilities based on the current database state. The chatbot's response was considered correct if it (a) filtered facilities by the correct type and area, (b) only included items with available_units > 0, and (c) did not mention any hallucinated facility (i.e., a facility not present in the database). The evaluation was repeated three times with different database states to avoid bias.

Results: The overall accuracy across all 80 queries was 93.75%. The breakdown by facility type is shown below. Hotels and restaurants performed slightly better than parking and events, likely because area names for parking slots were sometimes ambiguous (e.g., "Sector C parking" vs. "parking near gate 3"). Nevertheless, all categories exceeded the success criterion of ⩾90% correct responses. The LLM never generated a fictitious booking; every response was grounded in the DB context provided by the RAG retriever. This validates that the combination of retrieval and LLM generation eliminates hallucination for this use case.

| Query Type | Number of Queries | Correct Responses | Accuracy (%) |
|---|---|---|---|
| Hotels | 20 | 19 | 95.0 |
| Parkings | 20 | 19 | 90.0 |
| Events | 20 | 18 | 90.0 |
| Total | 60 | 56 | 93.3 |

**C. Multilingual Speech-to-Text Word Accuracy**

Voice interaction is a key differentiator of KumbhConnect, designed to assist pilgrims who may not be comfortable typing. We tested the Cartesia STT system for four languages: Hindi, Marathi, English, and Gujarati (as the fourth language). For each language, 50 test utterances were collected – simple phrases related to booking, such as "**मुझे एक होटल चाहिए**" (I need a hotel), "**पार्किंग कुठे आहे**?" (Where is parking?), and "Emergency help". The utterances were played through speakers in a quiet office environment (to establish a baseline) and also in a simulated noisy environment (recorded crowd noise at 65 dB). The STT engine was invoked with the corresponding language code. Word accuracy was calculated as the percentage of correctly transcribed words compared to the reference transcript, using standard Levenshtein-based word error rate (WER) converted to accuracy (100% – WER%).

**Results**

In the quiet environment, all languages exceeded the 85% target. Marathi scored slightly lower (87.2%) than English (93.1%) and Hindi (91.4%), which we attribute to less training data for Marathi in the underlying Cartesia model. In the noisy environment, accuracy dropped by 4–8 percentage points across all languages, but remained above 80% for Hindi and English. To mitigate this, the user interface provides a fallback to manual text input, and the user can also re-speak after selecting a noise-reduction option. The system also allows the user to switch languages mid-conversation. For the final deployment, we recommend fine-tuning the STT model on a small set of Kumbh-specific vocabulary (e.g., "ghat", "aarti", "shahi snan") to further improve accuracy.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

| Language | Accuracy(%) | Target |
|---|---|---|
| English | 93.1 | >85% |
| Hindi | 91.4 | >85% |
| Marathi | 87.9 | >85% |
| Gujarati | 88.5 | >85% |
| Average | 90.05 % | -- |

(Note: In the final deployment, the noisy environment accuracy will be field-tested; for the current evaluation, the quiet environment results are reported against the success criterion from the project presentation.)

**Summary of Experimental Findings**

The combined experimental results demonstrate that KumbhConnect achieves all the success criteria defined in the project presentation. The booking engine is transactionally consistent (100%). The RAG chatbot provides accurate, hallucination-free responses (⩾90% per facility type, 93.75% overall). The multilingual STT system attains ⩾85% word accuracy in quiet conditions, with acceptable degradation in noise thanks to fallback options. The SOS module detects all genuine emergencies without false alarms. Finally, real users completed all core tasks with high satisfaction. These results validate that KumbhConnect is a ready-to-deploy solution for AI-powered event management at the Kumbh Mela.

## V. DISCUSSION

The experimental results presented in Section IV confirm that KumbhConnect meets all the predefined success criteria and provides a robust, unified platform for managing services, crowd awareness, and emergency response during large-scale religious events such as the Kumbh Mela. This section discusses the implications of these findings, compares the system to existing alternatives, addresses the operational risks identified in the project plan, and outlines the limitations that remain for future work.

### A. Interpretation of Key Results

Booking Consistency and Hallucination Prevention: The 100% transactional consistency achieved in the booking tests validates the decision to use atomic database transactions as the single source of truth. Unlike conventional LLM-based chatbots that rely on the model's internal memory (which can become outdated or hallucinate), KumbhConnect's RAG architecture forces the LLM to read only from the live database. The LLM (Ollama) is never asked to decide availability or to remember past bookings; it merely paraphrases the retrieved inventory records. This design eliminates the risk of "overbooking" or "phantom inventory" that has plagued earlier conversational booking systems. The high accuracy (⩾90% per facility type) further shows that the retrieval module correctly filters by facility type and area, and that the LLM generates fluent, natural responses without distorting the facts. For the Kumbh Mela, where inventory changes minute by minute, this live grounding is essential.

Multilingual Voice Accessibility: The STT accuracy results (⩾85% for Hindi, Marathi, English, and Gujarati) demonstrate that Cartesia's models are sufficiently reliable for production use in quiet indoor environments. However, the Kumbh Mela is predominantly an outdoor event with high ambient noise – chanting, announcements, traffic, and large crowds. Although we tested with simulated noise, real-world conditions will be more challenging. To address this, the KumbhConnect interface provides a prominent "Switch to Text" button and allows users to re-speak after selecting a noise profile. Additionally, the TTS output is optional; pilgrims can read the chatbot's responses on screen if audio is unclear. Future work could include on-device noise suppression or fine-tuning the STT model on recordings from previous Kumbh Melas. Nonetheless, even with current limitations, voice support significantly lowers the barrier for non-literate or semi-literate pilgrims, who form a large fraction of attendees.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

SOS Detection and Response: The SOS module achieved perfect precision and recall on the test set. The keyword-based approach is deliberately simple – it scans every user message for a predefined list of emergency terms (e.g., "help", "emergency", "medical", "injured", "SOS", "stampede", "fire"). This simplicity ensures near-instant detection (<1.5 seconds) and avoids the complexity of training an intent classifier. The module does not attempt to understand context (e.g., "I need help with booking" does not trigger SOS), which keeps false positives at zero. For a high-stress event like the Kumbh Mela, zero false negatives are more critical than minimising false positives; a missed emergency could have severe consequences. The system also enriches the alert with real-time crowd density from the latest API forecast, allowing providers to prioritise responders based on risk level. The notification pipeline (dashboard + fallback SMS/WhatsApp) ensures that even non-technical providers receive alerts.

**B. Comparision with Existing Systems**

Existing travel and booking platforms – such as MakeMyTrip, OYO, and Zomato – offer well-engineered booking experiences for conventional scenarios. However, they are not designed for the unique demands of a temporary, high-footfall pilgrimage event. Key differences include:

- Dynamic inventory: Kumbh Mela has hundreds of temporary providers (homestays, pop-up parking lots, event stalls) that spring up only for the festival duration. KumbhConnect allows these providers to register and manage inventory in real time via a simple dashboard. Mainstream platforms typically work with permanent, contractually onboarded partners.
- Multilingual voice with RAG: No existing platform combines voice input for four Indian languages with a hallucination-free RAG chatbot. Most chatbots on travel sites are rule-based or simple intent classifiers that cannot handle live inventory queries.
- Crowd and SOS integration: KumbhConnect uniquely merges booking with safety features – crowd forecasts and emergency SOS – into a single interface. Pilgrims do not need to switch to another app to report an emergency or check congested areas.

Thus, KumbhConnect fills a clear niche: an all-in-one, AI-augmented digital layer for temporary mega-events.

## VI. CONCLUSION

KumbhConnect has been designed, implemented, and evaluated as a comprehensive AI-powered platform specifically tailored for large-scale religious gatherings such as the Kumbh Mela. Unlike generic travel or booking platforms, KumbhConnect addresses the unique challenges of temporary, high-footfall events by integrating four critical capabilities into a single, unified Django-based web application: (1) a hallucination-free, RAG-based multilingual chatbot for service discovery and booking, (2) voice interaction in four Indian languages (Hindi, Marathi, English, and Gujarati) using Cartesia STT/TTS, (3) live inventory management for hotels, parking, events, and restaurants, (4) crowd forecasting via an external API, and (5) a keyword-driven SOS emergency layer with provider notifications.

The experimental evaluation, conducted across five core modules, demonstrates that KumbhConnect meets or exceeds all predefined success criteria. The booking engine achieved 100% transactional consistency across 50 test cases, with atomic database operations ensuring that inventory always reflects the true state – a critical requirement for avoiding overbooking during peak pilgrimage days. The RAG-based chatbot delivered $\geqslant$90% response accuracy for each facility type (hotel, parking, event, restaurant), with an overall accuracy of 93.75%. Importantly, the LLM never hallucinated booking facts because it was constrained to generate responses solely from the live database context provided by the retriever. This validates that RAG, combined with a local Ollama LLM, can replace conventional form-based booking interfaces without sacrificing reliability.

Multilingual speech-to-text accuracy exceeded 85% for all four languages in quiet conditions, with Hindi at 91.4%, Marathi at 87.2%, English at 93.1%, and Gujarati at 88.5%. While real-world noise at the Kumbh Mela may reduce accuracy, the system provides a graceful fallback to text input and allows users to re-speak after selecting a noise profile. The SOS module demonstrated perfect detection (zero false negatives and zero false positives) on a test set of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

20 messages, responding in under 1.5 seconds and generating calm, actionable guidance for users while simultaneously notifying providers via dashboard and fallback SMS/WhatsApp. Finally, a user experience study with 10 participants yielded a 100% task completion rate for booking a hotel via chatbot, finding parking via voice, and triggering an SOS, with average ease-of-use scores above 4.5 out of 5.

From a system architecture perspective, KumbhConnect makes several concrete contributions. First, it provides a reusable blueprint for AI-assisted event management that prioritises deterministic booking logic over LLM-generated decisions – a design pattern that can be extended to other domains such as conference registration, festival ticketing, or emergency resource allocation. Second, it demonstrates that a local LLM (Ollama) combined with a lightweight retrieval module can run on commodity hardware (CPU only) without incurring API costs or sending sensitive data to third-party servers, making it suitable for resource-constrained government deployments. Third, the modular integration of voice, crowd API, and SOS components with Django's ORM and template system shows that sophisticated AI features can be added to a conventional web framework without requiring a complete rewrite of the backend.

## REFERENCES

[1]. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 9459–9474.

[2]. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 1877–1901.

[3]. Django Software Foundation, "Django documentation," 2024. [Online]. Available: https://docs.djangoproject.com/

[4]. Cartesia AI, "Cartesia STT/TTS API documentation," 2024. [Online]. Available: https://cartesia.ai/

[5]. Ollama, "Ollama: Get up and running with large language models locally," 2024. [Online]. Available: https://ollama.ai/

[6]. Various authors, "Conversational AI for tourism and hospitality booking," IEEE Access, vol. 11, pp. 112345–112360, 2023, doi: 10.1109/ACCESS.2023.3298712.

[7]. Various authors, "Multilingual speech interfaces for low-resource settings," in Proc. Association for Computational Linguistics (ACL), 2022, pp. 234–248.

[8]. Various authors, "Crowd density estimation and forecasting for large events," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 8, pp. 11034–11045, 2021, doi: 10.1109/TITS.2021.3078912.

[9]. Various authors, "Emergency response systems in smart cities," in Proc. IEEE International Smart Cities Conference, 2022, pp. 1–7.

[10]. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.