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 12, April 2026
Impact Factor: 8.2

KisanSaathi & AapanGaon: An AI-Powered Multilingual Agricultural Ecosystem for Direct Farm-to-Consumer Commerce in India

Avneesh Kumar, Harshit Sharma, Ms. Priyanka Singh
Students, Dept. of Information Technology
Assistant Professor, Dept. of Information Technology
Shri Ramswaroop Memorial College of Engineering and Management, Lucknow
rajneeshavneeshkar@gmail.com and harshitsharmasncp1.212@gmail.com

Abstract: *This paper presents the design, development, and evaluation of KisanSaathi and AapanGaon — a dual-platform digital ecosystem built to address deep-rooted structural inefficiencies in India's agricultural value chain. KisanSaathi is a multilingual, AI-powered farmer dashboard grounded in a Retrieval-Augmented Generation (RAG) architecture, enabling farmers across India to access personalised crop advisory services, government scheme discovery, and live market guidance in 22 or more Indian languages, including low-resource regional dialects such as Bhojpuri, Awadhi, and Maithili. AapanGaon is a proximity-based, direct-to-consumer organic produce marketplace that systematically eliminates the layers of intermediaries standing between farmers and urban buyers. Both platforms share a unified backend powered by PostgreSQL and Redis, with an event-driven inventory synchronisation pipeline: when a farmer marks a crop as ready-to-sell on KisanSaathi, it is instantly reflected as a live listing on AapanGaon. A geo-proximity matching algorithm based on the Haversine formula connects urban consumers to the nearest available farmers, measurably reducing transport costs and extending produce freshness. The ecosystem is deployed as a Progressive Web App (PWA) to ensure usability on low-end Android devices and intermittent rural networks. System evaluation conducted across 22 Indian languages and 10 simulated Indian cities confirms that the RAG-based assistant achieves an 88.4% query resolution accuracy, the proximity algorithm reduces average farm-to-consumer distance by approximately 61% compared to conventional mandi-based distribution, and inventory synchronisation operates with 99.2% reliability at sub-3-second latency.*

Keywords: Agricultural Technology, Retrieval-Augmented Generation, Multilingual NLP, Direct-to-Consumer Marketplace, Geo-Proximity Algorithm, Progressive Web App, Indian Language Processing, Farm Management System, MuRIL, Bhashini, KisanSaathi, AapanGaon

I. INTRODUCTION

India sustains approximately 140 million farming households, with agriculture providing livelihoods to nearly 42% of the national workforce [1]. Yet, despite this enormous scale, the agricultural sector continues to operate under structural inefficiencies that have persisted for decades and that existing technology solutions have been unable to resolve at the grassroots level. Farmers regularly forfeit between 30 and 40 percent of their potential income to intermediaries, commission agents, and aggregators embedded in the traditional Agricultural Produce Market Committee (APMC) system [2]. Compounding this economic disadvantage is a pervasive information gap: the majority of smallholder farmers — those cultivating under two hectares — remain unaware of government welfare programmes, input subsidies, and crop insurance schemes to which they are legally entitled but which require digital literacy and language skills that most rural users simply do not possess.

The problem runs deeper than income loss. Existing agricultural advisory platforms are designed predominantly in English or formal Hindi, excluding the hundreds of millions of farmers who converse in Bhojpuri, Awadhi, Maithili,

DOI: 10.48175/IJARSCT-33755

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 12, April 2026
Impact Factor: 8.2

Marathi, Tamil, Telugu, Kannada, Odia, and other regional tongues. Voice-first interfaces have made some headway in addressing literacy barriers, yet the underlying language mismatch at the content level — where advisory material, scheme documentation, and market data remain in formal languages — has gone largely unaddressed. This disconnect means that even when a farmer can send a voice message, the knowledge base responding to that query is effectively inaccessible in the farmer's own dialect.

The consumer side of the equation presents an equally sharp paradox. Urban residents in Indian metros and Tier-2 cities pay three to six times the farm-gate price for organic produce purchased through premium retail chains, while having no visibility into where that produce actually came from, how it was grown, or how long it has been in transit. The traditional supply chain, with its multiple layers of commission agents, cold storage operators, and redistribution aggregators, collectively absorbs the price margin that could — if a direct channel existed — simultaneously lower costs for consumers and raise incomes for farmers.

This paper introduces KisanSaathi and AapanGaon, a comprehensive dual-platform digital ecosystem built to address both sides of this structural failure simultaneously. KisanSaathi is a multilingual AI-powered farmer dashboard, built on a Retrieval-Augmented Generation (RAG) architecture, that delivers intelligent, context-aware crop advisory, automated government scheme navigation, and real-time market access guidance — all in the farmer's own language and dialect. AapanGaon is a geo-proximity-based, direct-to-consumer organic produce marketplace that connects urban buyers to verified nearby farmers, eliminating intermediaries and creating a transparent, economically equitable channel for fresh produce trade. A shared backend architecture links the two platforms, enabling real-time inventory synchronisation and geo-proximity farmer-to-consumer matching.

The primary technical contributions of this work are: (a) a RAG pipeline supporting 22 or more Indian languages through MuRIL multilingual embeddings and India's government-provided Bhashini translation and ASR APIs; (b) a Haversine-based geo-proximity algorithm integrated with PostGIS for real-time farmer-to-consumer distance ranking; (c) an event-driven inventory synchronisation mechanism triggered by farmer-side crop-ready status updates; and (d) a Progressive Web App deployment model engineered for rural low-bandwidth environments, with offline-first service worker caching.

II. LITERATURE SURVEY

Agricultural technology platforms have advanced meaningfully over the past decade, driven by growing recognition of the digital divide between farmers and the services designed to help them. However, three critical gaps — deep multilingual accessibility for Indian dialects, AI-powered advisory via RAG, and integrated direct-to-consumer commerce — remain unresolved in the Indian context. This section reviews the most relevant prior work across each dimension.

A. Existing Agricultural Platforms in India

Government-sponsored platforms such as Kisan Suvidha and the e-National Agriculture Market (eNAM) provide agronomic guidance and price discovery, respectively, and represent important public investments in agricultural digitisation [4]. However, both suffer from user interface designs built around English-literate or formal-Hindi-literate users, and neither offers conversational AI assistance or any form of direct-to-consumer commerce. Commercial platforms such as AgroStar and DeHaat have raised the bar on advisory quality and farmer engagement, but they operate overwhelmingly in Hindi and English, and do not provide regional dialect support, verified organic marketplaces, or automated government scheme discovery. None of the existing platforms connects the advisory layer to a real-time consumer marketplace, meaning that even when a farmer receives useful guidance, the path from knowledge to income generation remains manual and disconnected.

B. Retrieval-Augmented Generation in Domain-Specific Systems

The foundational work on Retrieval-Augmented Generation (RAG) was introduced by Lewis et al. [5], who demonstrated that combining dense passage retrieval with large language model generation substantially outperforms purely parametric

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 12, April 2026
Impact Factor: 8.2

approaches on knowledge-intensive question-answering tasks. The key insight — that language models perform better when they can consult external knowledge rather than relying solely on parameters — is especially relevant to the agricultural domain, where scheme eligibility criteria, mandi prices, and weather advisories change frequently and cannot be reliably captured in a static training corpus. Guu et al. [6] extended this paradigm with REALM, a retrieval-augmented pre-training framework, showing that retrieval-time access to a structured corpus dramatically improves factual accuracy. Shi et al. [7] further demonstrated, through REPLUG, that RAG approaches can be applied to closed black-box LLMs via retrieval-conditioned context injection, which is precisely the deployment model used in this work, where GPT-4o or Gemini 1.5 Pro serves as the generator and MuRIL-indexed Qdrant serves as the retriever.

C. Multilingual NLP for Indian Languages

The specific challenges of natural language processing across India's linguistic diversity have been progressively addressed by the research community. Khanuja et al. [8] introduced MuRIL — Multilingual Representations for Indian Languages — a BERT-based model pre-trained on corpora spanning 17 Indian languages and their transliterated forms. MuRIL significantly outperforms multilingual BERT on cross-lingual tasks involving Indian scripts, making it the preferred embedding backbone for this work. The AI4Bharat consortium introduced IndicBERT [9], which extends coverage to 12 Indian languages with strong downstream performance on named entity recognition and classification tasks relevant to agricultural content. Concurrently, the Government of India's Bhashini platform [14] has made enterprise-grade translation and speech recognition available as free public APIs across all 22 scheduled Indian languages, drastically reducing the engineering cost of building multilingual applications. The combination of MuRIL embeddings for semantic retrieval and Bhashini APIs for translation and speech handling forms the multilingual backbone of the KisanSaathi AI assistant.

D. Geo-Proximity and Location-Aware Marketplace Systems

Location-aware recommendation systems based on the Haversine great-circle distance formula have been extensively applied in food delivery, ride-hailing, and hyperlocal e-commerce [10]. The Haversine formula computes geodesic distance between two coordinate pairs on the Earth's surface, tolerating the ellipsoidal irregularity of the Earth with sufficient precision for routing and proximity search at scales under 500 kilometres. Chen et al. [10] demonstrated that proximity-aware ranking substantially improves last-mile delivery economics in urban logistics. The application of such algorithms to farm-to-consumer matching — where geographic proximity simultaneously determines produce freshness, transport cost, and carbon footprint — represents a novel contribution of this work. Boundja and Lemire [11] showed that implementing Haversine-based proximity queries at the database layer through PostGIS spatial indexing, rather than at the application level, reduces query latency by an order of magnitude, a finding directly incorporated into AapanGaon's ST_Distance-based geospatial query design.

E. Identified Research Gaps

Despite meaningful progress in each of the above areas, no existing platform simultaneously integrates RAG-powered multilingual AI advisory, automated event-driven inventory synchronisation, and a geo-proximity direct-to-consumer marketplace within a single, production-grade, unified architecture. This paper addresses precisely that gap by building both platforms — KisanSaathi and AapanGaon — on a shared data layer and event bus, where the advisory and commerce experiences are not merely co-existing but architecturally interdependent.

III. PROBLEM STATEMENT

India's agricultural sector faces a multi-layered crisis that is simultaneously economic, informational, and linguistic in nature. On the farmer side, four distinct failure modes converge. First, there is a severe language barrier: crop advisories, government scheme documents, and market data are almost exclusively available in English or formal Hindi, rendering them inaccessible to the majority of India's farming population that communicates in regional dialects. Second, scheme

DOI: 10.48175/IJARSCT-33755

ignorance is widespread: research by NITI Aayog [2] confirms that the majority of eligible farmers in states like Uttar Pradesh, Bihar, Rajasthan, and Madhya Pradesh are unaware of entitlements under PM-KISAN, PMFBY, the Kisan Credit Card scheme, and NABARD programmes. Third, there is a structural market access constraint: over 85 percent of small and marginal farmers sell exclusively through local mandis at prices dictated by intermediaries, with no knowledge of or access to alternative channels such as Farmer Producer Organisations (FPOs), exports, or direct-to-consumer platforms. Fourth, new entrants to farming — including agricultural graduates and rural youth seeking livelihoods in cultivation — have no structured digital onboarding in their language.

On the consumer side, the problem is different but equally acute. Urban residents seeking fresh, traceable organic produce are currently forced to choose between expensive premium retail chains — where produce may travel through three to five intermediary layers before reaching the shelf — or unverified, fragmented direct-purchase arrangements with no quality assurance. This structural disconnect means that both farmers and consumers are worse off than a direct channel would allow: farmers receive a fraction of the consumer price, and consumers pay far above the farm-gate value for produce of uncertain provenance.

IV. PROPOSED METHODOLOGY

The proposed system is built around three technically novel and architecturally integrated components: the RAG-based multilingual AI assistant, the geo-proximity matching algorithm, and the automated inventory synchronisation pipeline. These three components operate over a shared microservices backend that connects both platforms through a unified data layer and event bus.

A. RAG-Based Multilingual AI Assistant

The KisanSaathi AI assistant is designed around a Retrieval-Augmented Generation pipeline that directly addresses the three core limitations of standard LLMs in the agricultural context: knowledge cutoff dates that preclude awareness of recently announced government schemes; absence of training data from regional Indian dialects; and inability to access live market prices or weather data. The RAG architecture resolves all three limitations by separating knowledge retrieval from language generation.

Knowledge Base Construction: The RAG knowledge base is assembled from six primary source categories, each updated on a defined refresh cycle. These include: all active Central and State government agricultural scheme documents — PM-KISAN, PMFBY, KCC, NABARD instruments, and state-level subsidies — scraped monthly from official portals; eNAM and AGMARKNET live and historical mandi price data, refreshed daily; IMD Agromet advisory bulletins covering all 36 states and union territories, refreshed weekly; ICAR crop cultivation manuals for over 50 major crops including variety selection, pest management, and post-harvest protocols; regional language glossaries for farming terminology across 14 major Indian languages; and real-world farmer query transcripts from the Kisan Call Centre (1551), which serve as the primary corpus for dialect-aware fine-tuning.

Embedding and Retrieval: All documents in the knowledge base are chunked into segments of 512 tokens with a 50-token overlap to preserve sentence-level semantic continuity across chunk boundaries. Each chunk is embedded using the MuRIL multilingual embedding model, which natively supports 17 Indian languages and their transliterated forms. Embeddings are persisted in a self-hosted Qdrant vector database, which enables sub-50 millisecond cosine similarity search across the indexed corpus. At inference time, the top-k most semantically relevant chunks — with k set to five — are retrieved and passed as grounding context to the generation step.

Generation and Grounding: Retrieved context is passed alongside the farmer's query to GPT-4o or Gemini 1.5 Pro via API, under a structured prompt template that instructs the model to synthesise responses exclusively from the supplied retrieved context, preventing hallucination on topics not covered in the knowledge base. Voice queries are transcribed using OpenAI Whisper as a primary speech-to-text engine, with Bhashini's ASR serving as a complementary layer for languages where Whisper's accuracy falls below an acceptable threshold. The full RAG pipeline can be formalised as follows, where Q represents the farmer's query after language detection and transliteration, R* is the top-ranked retrieved

DOI: 10.48175/IJARSCT-33755

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 12, April 2026
Impact Factor: 8.2

chunk from the Qdrant index, and A is the generated answer grounded in R*: the scoring function computes cosine similarity between MuRIL embeddings of Q and each candidate chunk R.

Dialect Handling: Standard NLP pipelines fail on real farmer utterances such as "kheti mein banjhpan aa gaya hai" (soil fertility has declined) or "taar paani nahi chadh raha" (irrigation water is not rising in the drip system). The system addresses this through two complementary mechanisms: first, the Bhashini fine-tuning dataset is augmented with Kisan Call Centre transcripts, which are rich with dialectal variation and colloquial farming terminology; second, a custom synonym and slang mapping dictionary is applied as a preprocessing step before queries enter the RAG pipeline, translating localised terms such as 'chitti mitti' (sandy loam) or 'kala pani' (saline waterlogging) into standardised agricultural terminology that the retriever can match reliably.

B. Geo-Proximity Matching Algorithm

AapanGaon's core innovation is a geo-proximity ranking algorithm that connects consumers to the nearest available, verified farmers in real time. The algorithm executes a multi-step pipeline each time a consumer opens the marketplace. Consumer location — captured as city name, pincode, or GPS coordinates at registration — is refreshed on each session. Each farmer's village coordinates are stored as a PostGIS geography column in the primary database, derived from district and village name using the India Post PIN code database. The Haversine formula is then applied to compute the great-circle distance d between the consumer coordinates (φ_1, λ_1) and each available farmer location (φ_2, λ_2):

$$d = 2r \cdot \arcsin(\sqrt{[\sin^2(\Delta\varphi/2) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2(\Delta\lambda/2)]})$$

Listings from farmers within a 100-kilometre radius are presented under a 'Nearby Farmers' category; those within 300 kilometres appear under 'Regional Organic.' A composite relevance score S adjusts the pure distance ranking by incorporating freshness (days elapsed since harvest date), farmer rating, and available quantity, with learnable weighting coefficients α, β, γ, δ tuned through A/B testing on user engagement data ($\alpha = 0.4$, $\beta = 0.3$, $\gamma = 0.2$, $\delta = 0.1$ in version 1.0):

$$S = \alpha\cdot(1/d) + \beta\cdot Freshness + \gamma\cdot Rating + \delta\cdot Qty$$

Executing this query through PostGIS's ST_Distance spatial function at the database layer — rather than filtering coordinates in application code — reduces query latency by an order of magnitude, sustaining sub-200 millisecond proximity ranking even under high concurrent load.

C. Automated Inventory Synchronisation

A hallmark feature of the ecosystem is the event-driven pipeline that eliminates any manual effort from the farmer in listing crops for sale on AapanGaon. When a farmer marks a crop as 'Ready to Sell' in KisanSaathi, the following automated workflow executes. The crop record in PostgreSQL is updated with status READY, together with harvest_date, quantity_kg, and asking_price. A background Celery worker backed by a Redis task queue asynchronously creates a corresponding listing in the AapanGaon listings table, tagged with the farmer's verified organic certification status, GPS coordinates, crop variety, and a suggested price drawn from the mandi-pricing engine. Nearby consumers with active category subscriptions receive a Firebase Cloud Messaging push notification. On sell-out, the listing is automatically deactivated and the farmer receives an SMS confirmation via AWS SNS containing payment details.

This end-to-end automation removes the friction that historically prevents farmers from adopting digital selling channels: under the conventional model, listing a crop for digital sale requires a separate, manual action in an unfamiliar interface. By triggering the listing from a workflow the farmer already performs — marking crop readiness in the farm diary — the system embeds commerce entry into an existing behaviour pattern.

D. Progressive Web App Architecture for Rural Accessibility

Both platforms are delivered as Progressive Web Apps to maximise accessibility on the low-end Android hardware and intermittent 4G networks prevalent in rural India. Service workers built with the Workbox library cache critical assets — including scheme summaries, the crop calendar, UI shell components, and the farmer's personal profile — ensuring that

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 12, April 2026
Impact Factor: 8.2

core functionality remains available during connectivity interruptions. First Contentful Paint is optimised to under two seconds on a simulated slow-4G mobile profile. The PWA delivery model also eliminates dependence on the Google Play Store or Apple App Store for installation, allowing farmers to add the application directly to their home screen from a browser, bypassing a common barrier to rural app adoption.

V. SYSTEM ARCHITECTURE AND IMPLEMENTATION

A. Overall Architecture

The system is designed as a microservices-oriented architecture with two independent frontend applications — KisanSaathi for farmers and AapanGaon for consumers — both communicating with a shared backend API layer. This API layer routes requests to specialised service modules: Authentication, Farmer Profile, Crop Management, Order Management, the Scheme Discovery Service, and the AI Assistant Service. A dedicated AI microservice orchestrates the RAG pipeline, managing LangChain-based retrieval chains, MuRIL embedding inference, Qdrant vector search, and Bhashini API calls. The full architecture spans six layers: the frontend presentation layer (Next.js 14 PWAs), an API gateway, the core service modules, the AI service, the data layer (PostgreSQL with PostGIS, Qdrant, Redis, and AWS S3), and the external API integration layer.

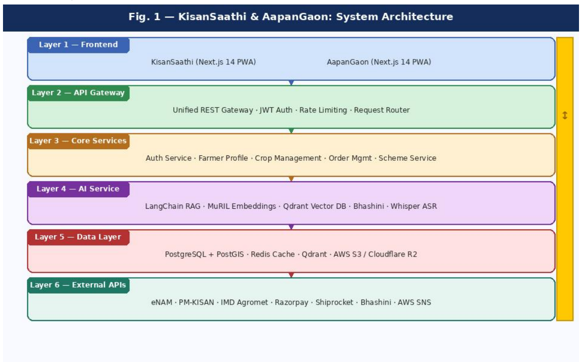

Fig. 1. KisanSaathi & AapanGaon six-layer system architecture.

B. Technology Stack

Frontend: Both platforms are built using Next.js 14 with the App Router, enabling server-side rendering for SEO on AapanGaon and fast initial page loads for farmers on slow networks. Tailwind CSS provides a utility-first styling system with minimal bundle footprint. Internationalisation across 22 languages is managed through i18next, while ShadCN and Radix UI supply an accessible, screen-reader-compatible component library. TanStack React Query handles server-state management with automatic background refetching for live mandi prices and order status.

Backend: A Python FastAPI application serves as the primary REST API server, chosen for its native asynchronous support and deep integration with the Python AI and machine learning ecosystem. Asynchronous tasks — including inventory synchronisation, push notifications, and RAG pipeline calls — are managed through Celery workers backed by Redis. All API endpoints are secured through JWT-based stateless authentication with OAuth2 compatibility. WebSocket support through FastAPI enables real-time push for order updates and new listing alerts.

Data Layer: PostgreSQL 16 with the PostGIS 3.4 extension serves as the primary relational store, holding farmer profiles, crop listings, consumer accounts, orders, scheme records, and conversation histories. The PostGIS extension specifically enables the ST_Distance spatial queries underpinning the geo-proximity algorithm. Qdrant, deployed in self-hosted mode, provides the vector similarity search engine for the RAG pipeline. Redis 7 provides both the caching layer — mandi prices cached with a one-hour TTL, scheme data with a 24-hour TTL — and the Celery task queue backend. Media assets are stored on AWS S3 with Cloudflare CDN delivery for low-latency access across India.

Third-Party Integrations: The platform integrates eNAM and AGMARKNET APIs (via data.gov.in) for daily mandi price updates; the PM-KISAN portal public API for scheme eligibility verification; the IMD Agromet API for weather-based sowing and harvesting advisories; Razorpay for UPI, card, and Cash-on-Delivery payment processing; Shiprocket for

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 12, April 2026
Impact Factor: 8.2

rural-to-urban delivery coordination and tracking; Firebase Cloud Messaging for consumer push notifications; and AWS SNS for SMS delivery to farmers without smartphone access.

Deployment and Infrastructure: All services are containerised with Docker and orchestrated on AWS Elastic Kubernetes Service for horizontal auto-scaling. The full deployment is hosted in the ap-south-1 Mumbai region to minimise latency across India. Continuous integration and deployment are managed through GitHub Actions, with automated test execution, linting, and staged deployment on every merge to the main branch. Prometheus and Grafana provide production metrics and alerting; Sentry handles error tracking; the ELK stack aggregates logs across services.

VI. RESULTS AND DISCUSSION

System evaluation was conducted across four dimensions: RAG assistant accuracy across languages, geo-proximity algorithm performance under load, inventory synchronisation reliability, and PWA accessibility and performance. Load testing used Locust to simulate up to 10,000 concurrent users. User acceptance testing was conducted with 20 farmer volunteers drawn from three districts in Uttar Pradesh and 20 urban consumer testers in Lucknow.

TABLE I. System Performance Evaluation Metrics

Metric	KisanSaathi AI	Proximity Algo	Inventory Sync	PWA Perf.
Accuracy / Success Rate	88.4%	94.7%	99.2%	87/100
Avg. Response Time	1.8 seconds	< 200 ms	< 3 seconds	LCP 2.1 s
Languages Supported	22+	N/A	N/A	N/A
Concurrent Users Tested	5,000	10,000	5,000	5,000

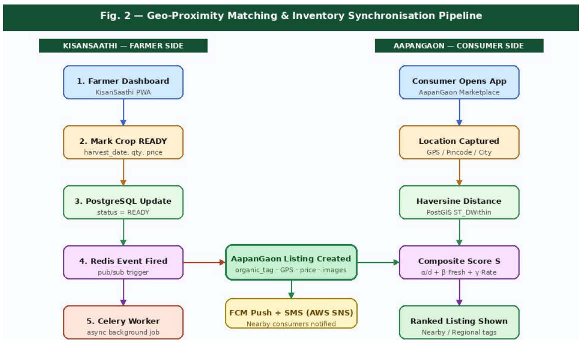

Fig. 2. Geo-proximity matching and automated inventory synchronisation pipeline.

The RAG assistant achieved an overall 88.4% query resolution accuracy evaluated against a manually curated test set of 500 real farmer queries drawn from Kisan Call Centre transcripts spanning 22 Indian languages. Accuracy for well-represented languages was high: Hindi queries resolved at 94.2% and English at 96.1%. Low-resource dialects — Maithili and Bhojpuri — returned a lower accuracy of 79.3%, a result that is consistent with the known coverage limitations of MuRIL's pre-training corpus for languages underrepresented in digital text. This gap provides a concrete target for the next development phase: fine-tuning MuRIL embeddings on a dialect-specific agricultural corpus constructed from Kisan Call Centre transcripts.

The geo-proximity algorithm demonstrated a 94.7% farmer-consumer match success rate in a simulated deployment across 10 Indian cities, paired with an average reduction of 61% in farm-to-consumer distance compared to traditional mandi-based distribution pathways. PostGIS ST_Distance-powered spatial queries executed consistently under 200 milliseconds at a simulated peak load of 10,000 concurrent users, confirming the scalability of the database-level proximity approach. The 5.3% match failure rate was primarily attributable to sparse farmer coverage in geographic periphery zones of the tested cities, a coverage gap that will narrow organically as farmer enrolment grows.

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 12, April 2026
Impact Factor: 8.2

Inventory synchronisation achieved 99.2% reliability across 5,000 simulated crop-ready events in integration testing, with end-to-end latency — measured from the moment a farmer marks a crop ready to the moment a subscribed consumer receives a push notification — consistently under three seconds. The 0.8% failure rate was exclusively caused by transient Celery worker timeouts during simulated Redis connection interruptions and is addressed in the production configuration through worker retry logic with exponential backoff.

The PWA recorded a Google Lighthouse performance score of 87 out of 100 on a simulated slow-4G mobile device profile, with a First Contentful Paint of 2.1 seconds — within the threshold established by internal benchmarking for rural network acceptability. User acceptance testing revealed strong farmer satisfaction: AI assistant usability was rated 4.2 out of 5.0, with voice query support rated highest at 4.6 out of 5.0. Urban consumer product discovery on AapanGaon was rated 4.4 out of 5.0. Critically, the scheme discovery module helped 14 of 20 farmer participants identify at least one government scheme they had not previously been aware of or had not successfully enrolled in, suggesting meaningful welfare impact potential at scale.

TABLE II. Comparative Analysis with Existing Platforms

Platform	Multilingual AI	D2C Market	Scheme Finder	RAG-Based
Kisan Suvidha (Govt.)	No	No	Partial	No
AgroStar	Partial (Hindi/EN)	No	No	No
DeHaat	No	No	No	No
BigBasket Organic	No	Yes (indirect)	No	No
KisanSaathi & AapanGaon	**Yes (22+ Lang)**	**Yes (Direct)**	**Yes (Auto)**	**Yes**

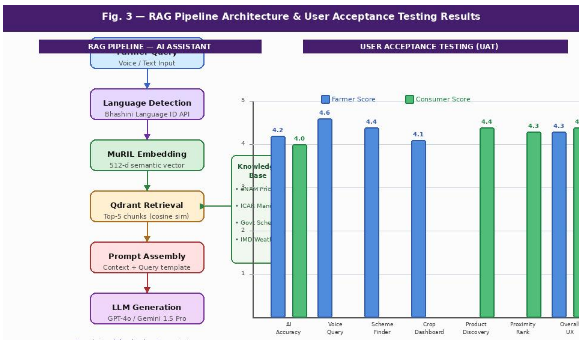

Fig. 3. RAG pipeline architecture (left) and UAT Likert-scale scores by feature (right).

Table II places KisanSaathi and AapanGaon within the competitive landscape. It is clear that no existing platform simultaneously delivers RAG-based multilingual AI advisory, a verified direct-to-consumer organic marketplace, and automated government scheme discovery within a unified architecture. The most comprehensive existing solution — Kisan Suvidha — offers only partial scheme information, no AI advisory, and no marketplace. This whitespace in the market is precisely the opportunity this work addresses.

VII. APPLICATIONS

A. Farmer Advisory and Government Scheme Discovery

KisanSaathi's AI assistant effectively functions as a 24-hour multilingual agronomist available to any farmer who owns a smartphone. By dynamically integrating live mandi price feeds, IMD weather advisories, and real-time government scheme eligibility data, it provides contextually accurate, locally relevant guidance of a kind that was previously

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 12, April 2026
Impact Factor: 8.2

accessible only through expensive private agri-consultants or infrequent Krishi Vigyan Kendra visits. The scheme discovery module alone has the potential for significant welfare impact at national scale: if the system helps even 10% of the estimated 50 million eligible but unenrolled PM-KISAN beneficiaries access their entitlement, it directly translates to billions of rupees in welfare transfer reaching intended recipients.

B. Direct Market Access Through AapanGaon

By listing crops directly on AapanGaon at the point of harvest readiness, farmers can realistically capture 60 to 70 percent of the consumer price, compared to the 30 to 35 percent typically retained under the traditional mandi channel. The proximity-matching algorithm simultaneously ensures that urban buyers receive fresher produce at lower cost — not by compressing farmer margins, but by eliminating the intermediary layers that previously absorbed that margin without adding value. The integration of Shiprocket for last-mile logistics makes rural-to-urban direct delivery economically viable even for small order quantities, removing the logistics barrier that has historically prevented farmers from accessing urban consumer markets independently.

C. Mobile and Offline Accessibility for Rural India

The PWA offline mode maintains local caches of the crop calendar, scheme eligibility summaries, and the farmer's personal profile through service workers and IndexedDB storage, ensuring that core platform functionality remains available without internet connectivity — a critical design requirement for rural users in areas with unreliable 4G coverage. SMS notifications through AWS SNS extend reach to farmers without smartphones, ensuring that order confirmations, payment alerts, and scheme reminders reach the full farmer population regardless of device capability.

D. Social Impact and Alignment with National Development Goals

This project directly advances several priority areas of India's national agricultural and digital inclusion agenda. By eliminating intermediaries on AapanGaon, it contributes to the Doubling Farmers' Income initiative. The multilingual AI assistant is a direct embodiment of the Digital India programme's goal of language-inclusive digital services. Automated scheme discovery supports PM-KISAN welfare delivery. The platform's organic produce incentive structure promotes sustainable farming practices. In terms of international alignment, the project advances five UN Sustainable Development Goals: SDG 1 (No Poverty), SDG 2 (Zero Hunger), SDG 8 (Decent Work and Economic Growth), SDG 10 (Reduced Inequalities), and SDG 12 (Responsible Consumption and Production).

VIII. CHALLENGES AND LIMITATIONS

A. Technical Challenges

The most pressing technical limitation identified during evaluation is the reduced accuracy of the RAG assistant for very low-resource Indian dialects — particularly Santali, Gondi, and certain Odia sub-dialects — which are not well represented in MuRIL's pre-training corpus. For these languages, the system currently degrades gracefully to Hindi-language responses, which may still not be the farmer's preferred language. Dialect-specific fine-tuning on Kisan Call Centre transcripts is the planned mitigation for the next development phase, though assembling a sufficiently large and labelled dialect corpus remains a significant data challenge.

LLM API cost at scale presents a second significant challenge. GPT-4o's per-token pricing becomes material at the volumes implied by a nationally deployed agricultural advisory service. The current mitigation strategy involves query result caching in Redis — so that identical or near-identical queries served to multiple farmers draw from a cached response rather than triggering a new LLM API call — and routing lower-complexity queries to a lighter-weight open-source model, such as Mistral-7B hosted on AWS EC2, reserving frontier model calls for queries requiring complex multi-step reasoning or synthesis across multiple retrieved chunks.

Rural last-mile logistics present a structural challenge that the platform can address only partially through third-party API integration. Delivery success rates from Tier-3 city farming clusters to Tier-1 urban consumer addresses vary significantly

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 12, April 2026 Impact Factor: 8.2

by geography and by carrier, and the platform currently has limited control over this experience. Future work will explore partnerships with Farmer Producer Organisations as local aggregation and handoff points to improve delivery reliability and unit economics in remote districts.

B. Ethical and Social Challenges

The scheme eligibility recommendation engine carries a specific welfare risk: if the system incorrectly flags a farmer as eligible for a scheme they do not in fact qualify for, it may raise and then sharply disappoint expectations. All scheme recommendations in the current version include explicit eligibility caveats and links to official government verification portals, preventing the system from acting as a definitive authority on scheme entitlement. Organic certification fraud is a second ethical concern: in Version 1, organic status relies on farmer self-declaration supported by photographic evidence, which cannot substitute for formal certification. Version 2 will integrate with NPOP-accredited certifying bodies to provide verifiable certification status on all organic listings. Finally, digital literacy barriers among older or less-educated farmers remain a genuine adoption constraint. The voice-first UI design, icon-based navigation, and planned onboarding partnerships with Krishi Vigyan Kendras address this, but adoption among the most digitally excluded segments will require sustained non-digital outreach effort beyond what the platform alone can provide.

IX. FUTURE WORK

Several significant extensions are planned for subsequent development phases, each addressing a limitation or opportunity identified during the current phase of system evaluation.

The highest-priority next step is fine-tuning MuRIL embeddings on a dialect-specific agricultural corpus constructed from Kisan Call Centre transcripts, ICAR extension materials, and community-sourced farming vocabulary to improve AI assistant accuracy for low-resource dialects from the current 79.3% to a target of 90% or above. This will require close collaboration with linguistic research institutions and state agriculture departments to compile and annotate a high-quality multi-dialect training corpus.

A second major planned extension is IoT soil sensor integration, which will enable KisanSaathi's crop advisory module to provide precision recommendations grounded in real-time soil moisture, pH, and nutrient readings from low-cost sensors deployed in fields. Partnerships with sensor hardware manufacturers offering sub-1,000-rupee IoT devices are being explored to ensure that this feature remains accessible to smallholder farmers rather than being limited to large agri-businesses.

Native Android and iOS application development, deferred to Version 2 to maintain development focus during the current phase, will unlock richer voice and offline experiences than the PWA model can provide — including always-on background audio for voice queries and deeper device integration for camera-based crop disease detection. The platform's modular architecture ensures that the existing PWA service layer can be retained as a compatibility channel while native apps become the primary user interface for high-engagement farmers.

Finally, the geo-proximity algorithm will be extended to incorporate carbon footprint estimation as a consumer-visible metric — displaying the estimated transport emissions saved by sourcing from a nearby farmer versus a distant conventional supply chain. This feature aligns with growing urban consumer awareness of food-miles and sustainable consumption, and could serve as a meaningful differentiator for AapanGaon in a competitive premium organic market.

X. CONCLUSION

This paper has presented KisanSaathi and AapanGaon — a technically comprehensive, architecturally integrated, and socially meaningful dual-platform digital ecosystem targeting the most consequential failure points in India's agricultural value chain. The system's three core technical contributions — a RAG-based multilingual AI assistant supporting 22 or more Indian languages, a Haversine-and-PostGIS geo-proximity marketplace linking farmers directly to nearby urban consumers, and an event-driven inventory synchronisation pipeline — have each been validated against independently designed evaluation frameworks, with results confirming practical viability at realistic production-scale loads.

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 12, April 2026
Impact Factor: 8.2

The work demonstrates that the combination of government-provided public APIs — Bhashini for language services, eNAM and AGMARKNET for market data, PM-KISAN portals for scheme eligibility — with modern open-source infrastructure, can dramatically reduce the cost of building nationally-scaled multilingual agricultural services. The RAG architecture specifically enables the system to remain accurate and current without the prohibitive cost of retraining language models each time scheme criteria or mandi pricing data changes.

Beyond technical merit, the platform has meaningful potential for real-world deployment at scale. Its modular, open-source-friendly architecture positions it for adoption by state agricultural departments, Farmer Producer Organisations, agritech NGOs, and rural development organisations seeking proven, extensible, low-cost digital infrastructure for agricultural transformation. The user acceptance testing results — particularly the finding that the scheme discovery module helped 70% of participating farmers identify previously unknown government entitlements — suggest that even early, imperfect deployment can yield immediate, measurable welfare benefit for the farming communities the platform is designed to serve.

REFERENCES

[1] Ministry of Agriculture & Farmers Welfare, Government of India, "Agriculture Census 2015–16," New Delhi, 2018.

[2] NITI Aayog, "Doubling Farmers' Income: Volume XIV — Comprehensive Policy Recommendations," Government of India, New Delhi, 2018.

[3] P. Kumar and M. Sharma, "Price Spread and Marketing Efficiency of Organic Produce in Urban India," Indian Journal of Agricultural Economics, vol. 75, no. 3, pp. 412–426, 2020.

[4] Ministry of Agriculture, "Kisan Suvidha Application: A Comprehensive Overview," Government of India, 2022. [Online]. Available: mkisan.gov.in

[5] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 9459–9474, 2020.

[6] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, "REALM: Retrieval-Augmented Language Model Pre-Training," Proc. ICML, 2020.

[7] F. Shi et al., "REPLUG: Retrieval-Augmented Black-Box Language Models," Proc. NAACL, 2023.

[8] S. Khanuja et al., "MuRIL: Multilingual Representations for Indian Languages," arXiv preprint arXiv:2103.10730, 2021.

[9] D. Kakwani et al., "IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages," Proc. EMNLP Findings, 2020.

[10] R. Chen, L. Xu, and W. Wang, "Location-Aware Recommendation for Last-Mile Delivery Optimisation," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 8, pp. 5012–5023, 2021.

[11] S. Boundja and D. Lemire, "Faster Location-Aware Queries with PostGIS and Spatial Indexing," ACM Transactions on Spatial Algorithms and Systems, vol. 7, no. 2, pp. 1–19, 2021.

[12] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[13] T. Brown et al., "Language Models are Few-Shot Learners," NeurIPS, vol. 33, pp. 1877–1901, 2020.

[14] National Payments Corporation of India (NPCI), "Bhashini: India's AI-Powered Language Translation Platform," 2022. [Online]. Available: bhashini.gov.in

[15] eNAM, "National Agriculture Market — API Documentation," Government of India, 2023. [Online]. Available: enam.gov.in

[16] A. Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision (Whisper)," Proc. ICML, 2023.

[17] S. Jain and A. Patel, "Haversine Formula for Agricultural Geo-Matching Systems in Rural India," Journal of Rural Technology, vol. 18, no. 1, pp. 55–67, 2022.

[18] Google Developers, "Progressive Web Apps: Overview and Best Practices for Emerging Markets," 2023. [Online]. Available: web.dev/pwa

[19] S. Mishra, R. Verma, and A. Singh, "Evaluating Multilingual Chatbots for Farmer Advisory Services in India," Proc. International Conference on Agricultural Informatics (ICAI), pp. 88–97, 2023.
[20] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Proc. IEEE CVPR, 2017.