

AI for Bharat Hackathon

Powered by **aws**



Team Name : **MLOps 4.0**

Team Leader Name : **Aditya Tiwari**

Problem Statement : **AI for Communities, Access & Public Impact**

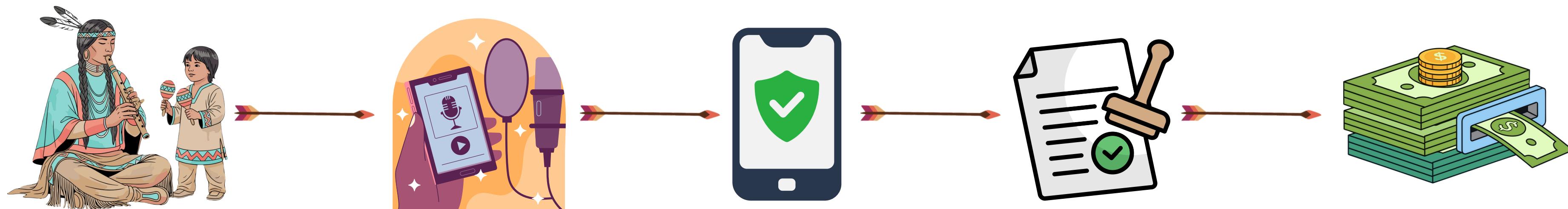
DHAROHAR (धरोहर) - Safeguarding India's Wisdom with Digital Sovereignty

Problem:

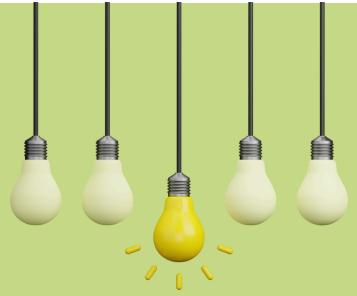
Indigenous communities hold 80% of the world's **biodiversity knowledge**, yet they face economic exclusion due to the lack of 'Digital Sovereignty.' They cannot prove ownership of their **oral remedies (Bio-Piracy)**, **verify the authenticity of their handicrafts (Counterfeits)**, or **monetize their cultural assets (folk music)**, leading to a loss of \$10B+ in potential annual revenue for rural India.

Solution:

Dharohar is India's first "**Heritage-as-an-Asset**" infrastructure. It is a multi-modal AI platform that enables indigenous communities to **Digitize, Validate, and Monetize** their intangible heritage. By combining **GenAI (AWS Bedrock)** for **oral history documentation** and **Computer Vision (AWS Rekognition)**, Dharohar creates a "**Digital Passport**" for rural assets. This transforms vague "oral tradition" into legally defensible "Prior Art Dossiers" and verifiable "artistic ownership" that can be licensed to global industries.

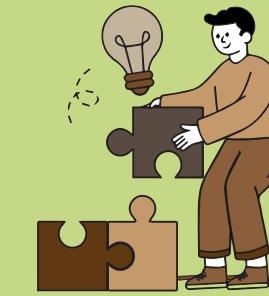


How is it different, how it solves the problem, USP



Proactive vs. Reactive: Unlike existing digital archives (like TKDL), which are static and defensive, Dharohar is Progressive — it actively helps communities license and sell rights.

Multimodal Validation: While others rely on text claims, we use Biometric Weave Analysis (Vision) and Dialect-to-Legal Translation (GenAI).

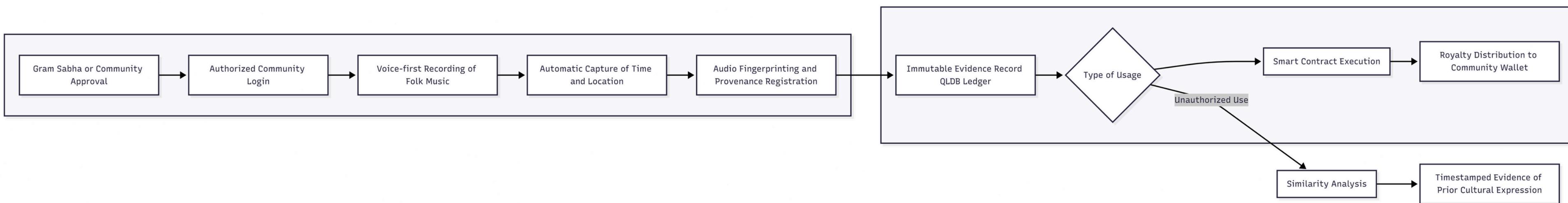


Stops Bio-Piracy: Auto-generates "Form-1" compliant Prior Art dossiers to block unauthorized patents. Restores control and consent to communities

Stops Fakes: Enables ethical licensing aligned with ABS and biodiversity laws. Converts oral knowledge into provable digital evidence



- **Evidence-first, not asset-first**
- Community custodians decide visibility and licensing
- **Zero-Touch Legal Compliance:** Automated "Form-1" generation for the National Biodiversity Authority.
- AI assists documentation, not ownership
- Transparent royalty distribution via smart contracts



List of features offered by the solution

Dharohar-Bio | Protection of Traditional Medicinal Knowledge



Dharohar-Bio enables Gram Sabhas and tribal communities to document their traditional medicinal knowledge through a voice-first interface in local dialects such as Gondi and Bhili. Oral remedies are transcribed, structured, and mapped to scientific taxonomy using AI, creating legally interpretable prior-art and patent-ready documentation that prevents bio-piracy cases like turmeric or neem.

Dharohar-Sonic | Rights & Royalties for Folk Music



Dharohar-Sonic allows communities to register their folk music and oral traditions through audio or video recordings under verified community identity. Each submission is fingerprinted and timestamped to establish copyright-grade provenance, enabling communities to detect unlicensed use, assert ownership claims, and earn royalties through licensed commercial usage.

The Sovereignty Vault | Immutable Rights Registry

The Sovereignty Vault is an immutable ledger that securely records every cultural and biological asset along with its ownership, time, and location metadata. By anchoring records on Amazon QLDB, Dharohar creates verifiable proof of first documentation, authorship, and historical ownership that can be used in legal, academic, or regulatory contexts.

License Marketplace | Ethical Monetization Layer

Dharohar provides a controlled B2B marketplace where researchers, media houses, and institutions can access community knowledge through purpose-bound, time-limited licenses. Smart contracts ensure transparent royalty distribution directly to community wallets, allowing monetization without transferring ownership or exploiting cultural heritage.

Dharohar - A community-owned rights-creation and evidence infrastructure that allows indigenous communities to establish **authorship, provenance, and prior art, and to monetize usage through licensing** – without transferring ownership to the platform.



Key Stakeholders:

Tribal healers, Gram Sabhas, biodiversity boards, ethnobotanists, pharmaceutical researchers, NGOs working on indigenous rights.

Usage:

Communities document oral medicinal knowledge as AI-structured, timestamped evidence to establish prior art, prevent bio-piracy, and enable ethical research access through controlled licensing and benefit sharing.



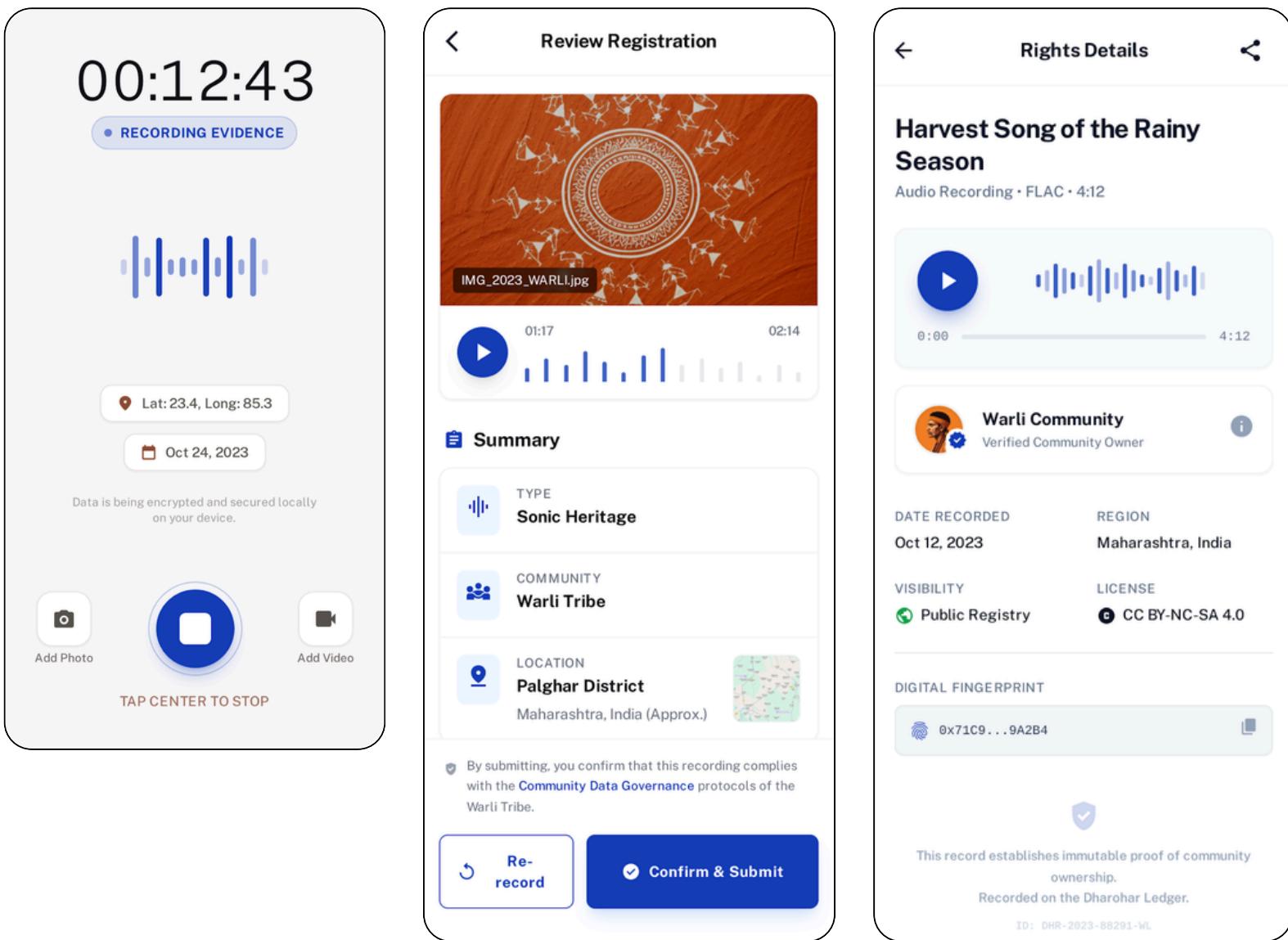
Key Stakeholders:

Folk artists, cultural communities, Gram Sabhas, music labels, film and media producers, copyright organizations, NGOs and cultural authorities.

Usage:

Communities register folk music and oral traditions under verified identity, generate provenance and audio fingerprints, detect unlicensed commercial use, and earn royalties through consent-based licensing.

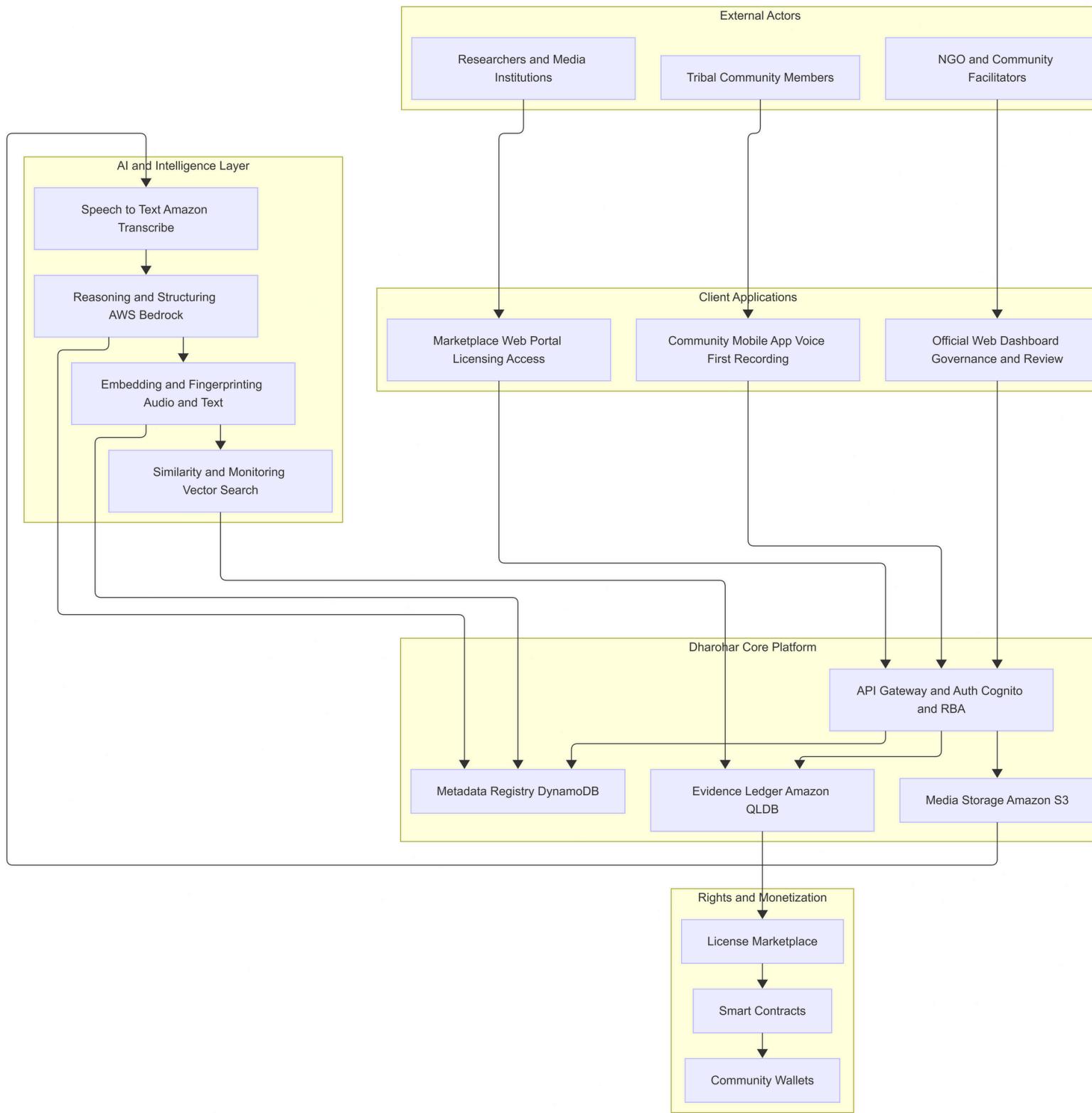
User Interface



Dharohar (Gram Sabha Mobile App) -
Allow communities to publish/ record their
bio heritage or music

Dharohar (NGO/ Authority dashboard) - Allow them for patent approval/ registrations/ monitor the published heritage usage monitoring

Architecture diagram of the proposed solution:



External Actors & Entry Points

At the foundation are three primary actors: tribal community members, NGO or community facilitators, and researchers or media institutions. Each actor interacts with the system only through a dedicated application layer, ensuring clear separation of access, responsibility, and control.

Client Applications Layer

1. The Community Mobile App enables authorized community members to record folk music or medicinal knowledge through a voice-first interface.
2. The Official Web Dashboard is used by NGOs and experts for governance, verification, and oversight.
3. The Marketplace Web Portal allows researchers and institutions to request licensed access to registered cultural or biological assets.
4. All client applications authenticate through a centralized API and role-based access system.

Core Platform Layer

1. The Dharohar Core Platform acts as the backbone of the system.
2. Media files are securely stored in Amazon S3, while structured metadata is maintained in DynamoDB for fast access and querying.
3. The Evidence Ledger built on Amazon QLDB stores immutable records of submissions, reviews, and access events, establishing legal-grade provenance and auditability.

AI & Intelligence Layer

1. The AI layer transforms raw community knowledge into protected digital assets.
2. Amazon Transcribe converts dialect speech into text.
3. AWS Bedrock performs reasoning and structuring, converting oral narratives into semantically meaningful and legally interpretable representations.
4. Audio and text embeddings are generated for fingerprinting, enabling similarity detection.
5. Vector search continuously supports monitoring for unlicensed or unauthorized usage of registered cultural assets.
6. This layer operates transparently, assisting documentation without replacing community or expert authority.

Rights & Monetization Layer

Once evidence is registered, assets flow into the rights and monetization layer. The License Marketplace exposes only approved, purpose-bound access to institutions. Smart contracts manage licensing terms and royalty logic. Community wallets receive automated benefit sharing without transferring ownership of knowledge.

Key Architectural Principles

1. Communities retain ownership; platforms provide infrastructure
2. NGOs govern and verify; they do not extract value
3. AI assists understanding, protection, and monitoring
4. Monetization is consent-based and transparent

Technologies to be used in the solution:

Mobile App (Community Participants)

Framework: **React Native (Android-first)**

State management: **Zustand or Redux Toolkit**

Offline support: **AsyncStorage + local queue**

Backend & APIs

Core Backend:

Runtime: **Node.js (TypeScript)**

Framework: **AWS Lambda + API Gateway**

Workflow orchestration: **AWS Step Functions**

Auth & access control: **AWS Cognito + custom RBA layer**

File uploads: **Pre-signed S3 URLs**

Data Storage:

Media (audio/video): **Amazon S3**

Metadata & registries: **Amazon DynamoDB**

Trust, Evidence & Rights Layer

Immutable Evidence Ledger

Audit & provenance: **Amazon QLDB**

Submission timestamps

Review actions

Access logs

Blockchain (Minimal, Purpose-Driven)

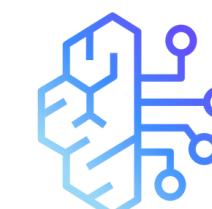
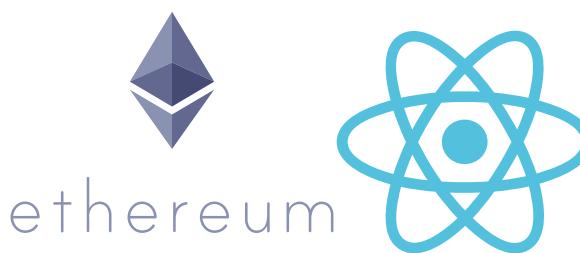
Service: **Amazon Managed Blockchain**

Usage:

Anchor evidence hashes (proof of existence)

Execute **license smart contracts**

Network: **Polygon / Ethereum (Testnet for now)**



AI & Intelligence Layer

Speech & Language (Bio + Sonic)

Speech-to-text: **Amazon Transcribe (custom vocab for dialects)**

Reasoning & structuring: **AWS Bedrock (Claude / Titan)**

Dialect → **structured remedy**

Folk lyrics → **semantic representation**

Embeddings: **Bedrock Embeddings / custom model**

Audio Intelligence (Sonic)

Audio fingerprinting: **Custom Python service (MFCC / spectral embeddings)**

Similarity detection: **Vector similarity search (OpenSearch / Faiss)**

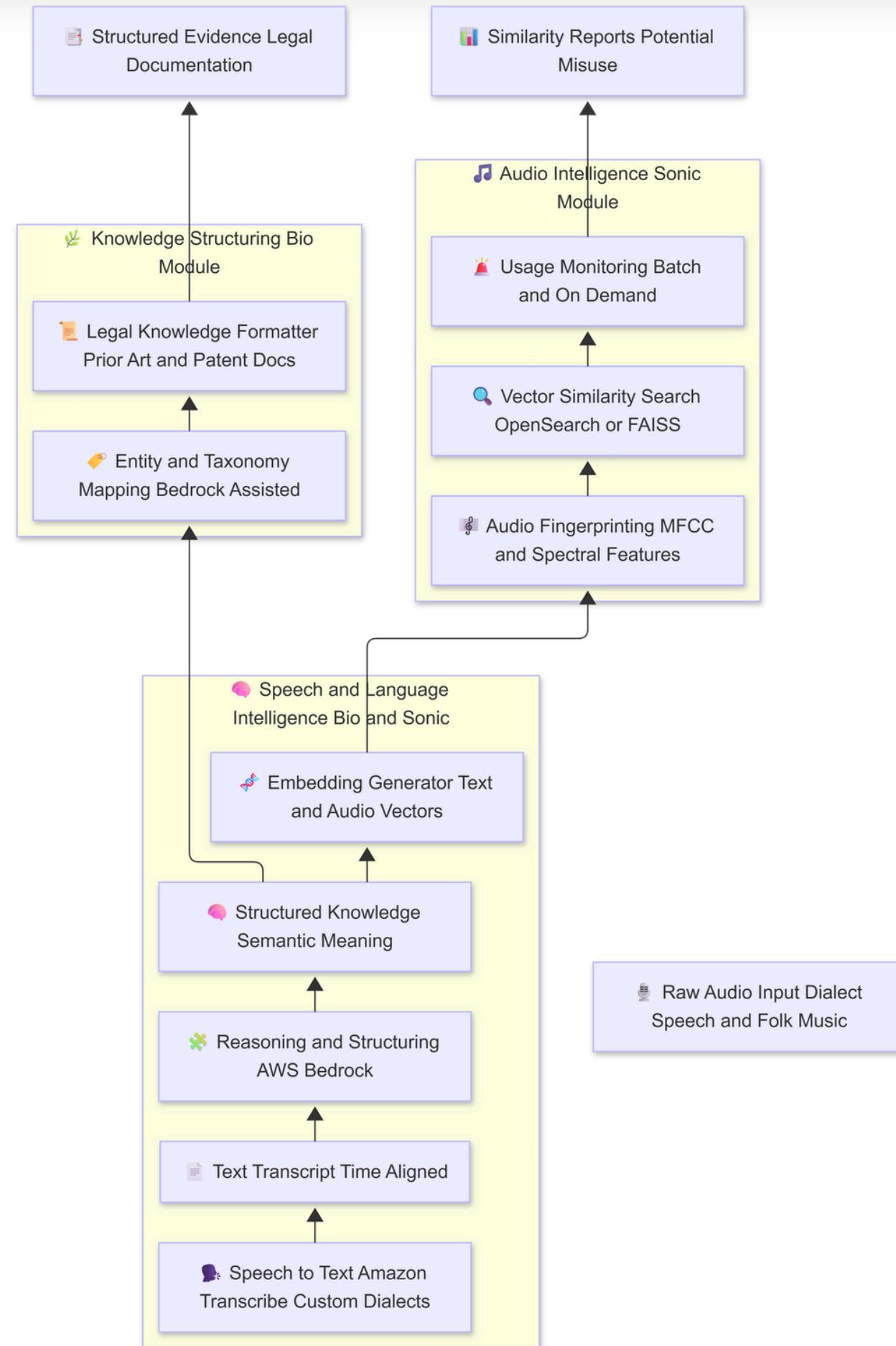
Monitoring pipeline: **Batch + on-demand similarity checks**

Knowledge Structuring (Bio)

NER & taxonomy mapping: **Bedrock + rule-based post-processing**

Patent / prior-art formatting: **Custom template engine (PDF generation)**

The AI Intelligence Layer acts as the core enabler of **Dharohar** by converting **informal, oral, and unstructured community knowledge** into structured, legally interpretable digital assets. Through **speech understanding, semantic reasoning, fingerprinting, and continuous monitoring**, AI establishes provenance, prevents misappropriation, and enables ethical access while keeping ownership and control firmly with the community.



Estimated implementation cost:

Component	Technology / Service	Pilot Usage Assumption	Estimated Monthly Cost (INR)
Backend APIs	AWS Lambda, API Gateway	Serverless APIs with low to medium traffic	3,000 – 5,000
Media Storage	Amazon S3	Audio-heavy uploads with limited video	2,000 – 3,000
Speech to Text	Amazon Transcribe	Dialect transcription for Bio and Sonic modules	5,000 – 7,000
AI Reasoning and Structuring	AWS Bedrock	Semantic reasoning, structuring, embeddings	4,000 – 6,000
Audio Fingerprinting	Custom Python Service	MFCC extraction and fingerprint generation	2,000 – 3,000
Vector Similarity Search	OpenSearch / FAISS	Limited similarity queries and monitoring	1,500 – 2,500
Immutable Evidence Ledger	Amazon QLDB	Evidence records and audit trails	1,000 – 2,000
Blockchain Anchoring	Managed Blockchain / Testnet	Evidence anchoring and license records	1,000 – 2,000
Authentication and Security	AWS Cognito, AWS KMS	Role-based access control and encryption	1,000 – 1,500
Monitoring and Logs	CloudWatch, X-Ray	Basic observability and alerting	500 – 1,000
Total Estimated Monthly Cost	All Services Combined	Pilot with 5–10 communities	22,000 – 30,000

How Dharohar Will Be Sold (Go-to-Market)

1. Adopted through **partnerships with NGOs, cultural organizations, and government-linked bodies** already working with tribal and rural communities
2. Not sold directly to communities; positioned as a **rights-protection and compliance infrastructure**
3. Primary customers include **biodiversity boards, cultural ministries, research institutions, and media organizations**
4. Revenue generated through **institutional licensing, research access fees, and commercial usage licenses**
5. Community participation remains free, with benefits returned through **transparent royalty and benefit-sharing mechanisms**

Innovation partner **H2S**
HACK2SKILL

Media partner **YOURSTORY**

AI for Bharat Hackathon

Powered by **aws**

Thank You

