Verifiable AI Decisions

(Hackathon Round 2 Submission (Healthcare Focus)

Round 2 Objective

We go beyond the concept and prototype the project's primary technological innovation in Round 2:

- Show how decisions made by AI can be verified later.
- Display blockchain immutability and cryptographic integrity.
- Make sure there is no exposure of patient data.
- Establish a functional workflow for verification

This round is not just about user interface, but also about architecture clarity, auditability, and proof-of-concept implementation.

Problem Recap

AI systems are utilized extensively in healthcare to support diagnosis, anticipate risks, plan treatments, and make insurance decisions.

Once an AI system makes a decision, it is impossible to determine with certainty what decision was taken at that precise moment.

Because AI models are often modified and non-deterministic, re-executing them is unreliable for verification.

- Conventional logs and databases: Adaptable and erasable
- rely on trust inside the company
- Don't offer cryptographic integrity.

Due to privacy and legal constraints, healthcare data is highly sensitive and cannot be stored on public blockchains.

Current blockchain methods are unsuccessful due to one of the following reasons:

- Save sensitive data directly, or
- Do not endorse independent verification

- These constraints generate substantial hazards in:
- Health care audits
- Insurance and legal disputes
- Regulatory adherence
- Responsibility for decisions made with the help of AI

Round 2 Solution Summary

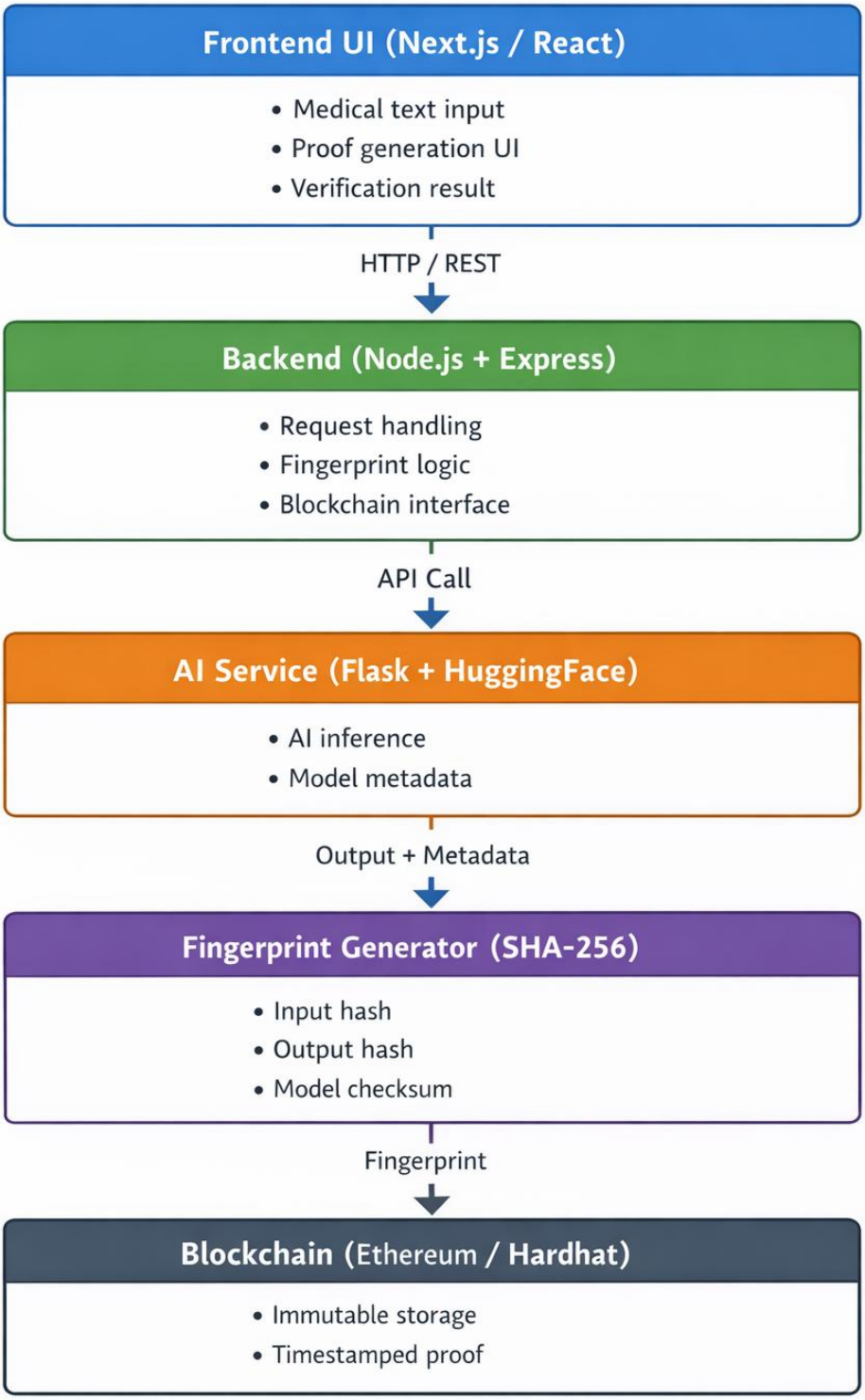
Cryptographic Decision Fingerprinting
Instead of storing data or models, we store a cryptographic fingerprint of:

- AI input (hashed)
- AI output
- Model identity
- Model checksum
- Inference parameters

Stored on blockchain as an immutable receipt
What This Achieves

- Verifiable AI decisions
- No medical data leakage
- No model re-execution required
- Independent third-party audits possible

System Architecture (Round 2)



Step 1: User Input

- Doctor or hospital uploads / enters a medical report
- Input is sent from Frontend (Next.js) to backend via REST API

Step 2: AI Inference

- Backend forwards medical text to AI Service (Flask + HuggingFace)
- AI model generates:
  - Prediction / risk score
  - Model metadata (model ID, checksum, parameters)

Step 3: Fingerprint Generation

- Backend generates a cryptographic fingerprint (SHA-256) using:
  - Hash of input data
  - AI output
  - Model metadata
  - Inference parameters
- This fingerprint uniquely represents the AI decision

Step 4: Blockchain Registration

- Generated fingerprint is stored on Blockchain (Ethereum / Hardhat) with:
  - Decision ID
  - Timestamp
  - Model ID
- No patient data or model weights are stored

Step 5: Verification Request (Later)

- Auditor / hospital submits decision data for verification
- Backend regenerates fingerprint from provided data

Step 6: Integrity Verification

- Regenerated fingerprint is compared with blockchain record
- Result:
  - Match → Decision is authentic
  - Mismatch → Decision has been tampered

Round 2 Features Implemented

Cryptographic Fingerprinting

- Implemented SHA-256 based fingerprint generation for AI decisions
- Fingerprint includes:
  - Hashed input data
  - AI output
  - Model ID and checksum
  - Inference parameters
- Ensures deterministic and tamper-proof representation of each decision

AI Inference Integration

- Integrated AI inference service using Flask and Hugging Face models
- Automatically captures:
  - Prediction result
  - Model metadata
  - Configuration parameters
- Supports extensibility for multiple AI models

Blockchain Storage & Verification

- Deployed Ethereum smart contract using Hardhat
- Stores only cryptographic fingerprints, not sensitive data
- Records:
  - Decision ID
  - Fingerprint hash
  - Timestamp
- Enables independent verification of decisions

Decision Verification System

- Regenerates fingerprint from provided decision data
- Compares regenerated fingerprint with on-chain record
- Clearly identifies:
  - Authentic decision
  - Tampered decision

Privacy-Preserving Design

- No medical data stored on blockchain
- No AI model weights exposed
- Fully compliant with healthcare privacy principles

Modular & Scalable Architecture

- Clear separation of frontend, backend, AI service, and blockchain layers
- Mock mode available when blockchain is unavailable
- Easy to extend with batching, dashboards, and multi-hospital support

Demonstration-Ready Capabilities

- Live proof generation
- Tamper detection by modifying AI output
- Real-time verification feedback

Real world use

- Medical audits
- Insurance disputes
- Legal malpractice cases
- AI regulatory compliance
- Clinical research integrity