Kavach — a second listener for scam calls

Catches scam calls in Telugu and Hindi, warns the listener, alerts the family and seals the evidence. Built entirely on Gnani AI.

THE PROBLEM

Indians lost **₹22,845 crore** to cyber fraud in 2024 (Ministry of Home Affairs, Lok Sabha, July 2025) — about **₹62 crore a day**. The typical attack is a phone call: someone posing as a bank, a power company or the police asks for an OTP, a payment or a remote-access app. The person most at risk is alone on the line, often elderly, and the call is in Telugu or Hindi. English-only, text-only fraud filters never hear it, and the family finds out after the money is gone.

THE SOLUTION

Kavach sits on the call as a second listener. When the caller uses a scam tactic it **warns the listener in Telugu or Hindi**, **alerts the family** with a text and voice note, **drafts the 1930 cybercrime complaint**, and **seals the evidence on Bitcoin** so no one can alter it later. Afterwards the family can ask Kavach what happened, in Telugu.

HOW IT WORKS: EIGHT AGENTS

1. **Listener** (Prisma): who said what, Telugu/Hindi/English/code-mixed
2. **Prosecutor** (Evon): scam tactics, each quoting the caller
3. **Defender** (Evon): signs the call is genuine
4. **Judge** (code): a claim counts only if the caller's own words prove it
5. **Guardian** (Evon + Timbre): spoken warning; family alert
6. **Reporter** (Evon): the 1930 complaint
7. **Notary**: SHA-256 hash chain, OpenTimestamps on Bitcoin
8. **Family assistant** (Evon over MCP): answers from sealed evidence

WHY IT CAN BE TRUSTED

- **The LLM argues, code decides.** An LLM judge flagged every genuine call; with the prosecutor/defender split and a code Judge, no genuine test call has been flagged since.
- **No caller words reach the listener.** Warnings and family alerts are built from verified tactics only.
- **Private by design.** Only a 32-byte hash leaves the device to be timestamped.
- **Evidence that holds.** A one-byte edit fails verification against the Bitcoin block.

100% GNANI · SOVEREIGN AI

Prisma v2.5	speech-to-text; separates caller from listener; Telugu, Hindi, English, code-mixed
Evon v3.3	open-weight LLM, self-hosted (Q4_K_M on 2× T4): prosecutor, defender, warning text, complaint, assistant
Timbre	text-to-speech: the warning to the listener and the family's voice note, in Telugu and Hindi

RESULTS SO FAR (HONEST, SYNTHETIC CALLS)

22 / 22
genuine Telugu–English calls left alone (44-call set)

3 / 3
Telugu calls right end to end with real Evon

9–12 s
to a Telugu verdict, laptop → self-hosted Evon

120
automated tests pass; seals confirmed on Bitcoin

Scams caught on the 44-call set: 16/22 — Evon still misses some tactics; red-line rules cover OTP/PIN and remote-access requests.

LIMITS AND NEXT STEPS

Today calls are analysed after recording; next is streaming Prisma for true mid-call warnings. Family alerts are built, voiced and queued, but no SMS/WhatsApp gateway is connected yet. All test calls are synthetic (scripted, voiced by Timbre). Next: a gateway, more Indian languages, and a small pilot with real families.

Code: github.com/lucky07-07/kavach-scam-call-guardian · **Videos:** data/demo/ in the repo · Gnani AI Great Indian AI Internship Challenge 2026 · #GnaniAI