GNANI · GREAT INDIAN AI INTERNSHIP CHALLENGE · SOLUTION PITCH

Soundstage: *your film crew, in a voice note.*

soundstage-site.vercel.app

The problem

Indian D2C brands need a new reel every week. The agency route takes about three weeks, a crew of a dozen and roughly ₹80,000 a reel (illustrative). Briefs arrive as WhatsApp voice notes in Hinglish, and most AI video tools want a polished English prompt.

The solution

Speak a brief in any Indian language; get an on-brand 9:16 reel, scored before you see it. Soundstage is a studio for D2C teams and small agencies: one voice note becomes a script, approved shots, a rendered and voiced reel, and a reviewer's score.

How it works

1. Speak the brief	A voice note in Hindi, Hinglish or English, rambling and all. Gnani Prisma v3.0 transcribes it (code-mixed: English stays Latin, Hindi stays Devanagari).
2. It writes the script	A director agent turns the transcript and the brand catalogue into a 9:16 script and shot plan. It may only claim what the catalogue says.
3. You approve every shot	A still for every scene before anything renders. One typed note ("cafe mein karo") changes the plan in 12 seconds.
4. It renders and voices it	Shots, word-timed captions, cards and pack shots, voiced by Gnani Timbre v2.5 (Kaveri en-IN, Poorvi hi-en, Bhavna hi-IN, Asmita ta-IN).
5. A strict reviewer scores it	A separate AI reviewer watches the render: 10 weighted criteria, 9 hard gates (no invented claim, no blank frames…). Below 85 it writes timestamped fixes and the reel renders again.

Gnani AI models used

- **Prisma v3.0** (speech to text, `api.vachana.ai`, `transcribe` format): every voice brief, Hindi, Hinglish and English, including noisy notes.
- **Timbre v2.5** (text to speech): every voice in every reel, in the language the brand picks; a Hindi note can become a Hinglish or English reel.

Also in the stack: an LLM director agent, a Gemini video reviewer, and a HyperFrames/ffmpeg render engine. The launch film's narration is a separate voice; all reel voices are Gnani.

Results (real runs, synthetic voice notes)

`3 min 36 s`	from voice note to a reel ready to post (one real run, filmed unedited)
`91.8 / 100`	Brewhouse reel from a 14 s Hinglish note, passed in one round
`67.5 → 93.8`	a rambling 35 s Hindi note: sent back once, passed on round 2
`83.8 → 96`	PulseFit reel: the reviewer's three timestamped fixes, applied

Who uses it

D2C brand marketing teams and the small agencies that make their weekly reels. Brewhouse (cold coffee) and PulseFit (gym) are demo brands built from a real-style catalogue.

Links and team

Live site: soundstage-site.vercel.app · Demo video: in the submission form

Built by Mohit Goyal (product, engine, studio, reviewer, launch film).