



Stakeholder Defence & AI Tool Log

The decision, defended to engineering, design, legal, finance and the NGO's scholars, plus who did what

Jain Sanyam Kamlesh, Product Manager & builder Built with Claude (Anthropic)

1. Written defence

1.1 The friction I picked, and why

I picked **F1, writing questions by hand**. It costs the host about an hour per show and sits at the start of every one. I chose it over F2 (right answers marked wrong), F3 (one difficulty for a mixed-age hall) and F4 (no explanation after the reveal) because it is their common root. Whoever writes the question also decides its difficulty, its explanation and its accepted spellings. So fixing the writing step fixes four frictions. F5 (joined but never answered) is a small interface fix that needs no AI.

1.2 The AI feature on one page

In the Studio's Content tab (File 04, Fig. 4.1) the host chooses a source book, a page range of up to 20 pages, a mix of easy, medium and hard, the language and, for typed rounds, the NGO's own "two clues, one answer" style (Fig. 4.2). The browser sends only those pages to Google Gemini's free tier, which drafts questions in the round's own format. Each draft has a PDF page, a short quote, a difficulty and an explanation, and typed answers come with English-letter spellings (Fig. 4.3–4.4). A second call answers every question **without seeing the answer**, and the code compares the two using the same matcher the phones use. The host then edits, re-levels or unticks, and only ticked drafts are added, as ordinary editable items (Fig. 4.5). If Google is busy the tool walks down a list of free models. If all of them fail it says so plainly and points to manual entry (Fig. 4.6). Workflow: File 02, section 5.

The user moment it changes: the night before a show the host goes from *writing* to *reviewing*, and on stage every reveal can now teach.

1.3 One responsible-AI decision I will defend

The blind check, plus mandatory human approval. My first version asked the second AI "is this answer supported by the book?". It agreed with nonsense. A two-clue riddle whose clues pointed to two different words was marked "found in the book". I replaced it with a blind test: the checker answers each question, and each clue separately, without seeing our answer, and the code decides whether they match. **What I traded:** a second call (about 30–60 extra seconds and twice the free quota per batch), and some good drafts that arrive unticked and cost the host a few seconds each. **What I preserved:** nothing reaches a hall unless an independent answer matched and a person approved it. For religious teaching in front of children and elders, a slower tool that is right beats a faster one that is confidently wrong.

1.4 North Star and kill switch

North Star: AI-drafted questions that hosts approved and that were played in a live show, per month. It only counts drafts that earned trust and reached the hall. **Kill switch:** I pause the feature immediately if any AI-written question is confirmed wrong after being played. I also pause it if a scholar's audit finds a checker false-pass rate above 2%, if hosts approve fewer than 40% of drafts over three batches in a row, or if any cost appears anywhere. Pausing takes about two minutes with the normal deploy, and questions already added remain usable. A remote on/off switch is next in the backlog.

1.5 The hardest question I expect

"A scholar stands up in the hall and says an AI question was wrong, and your checker had agreed with it. Why should an NGO trust a machine with dharma at all?"

Honestly: the checker is an AI too, so its mistakes can line up with the drafter's. It reduces errors; it does not remove them. That is why the person who approves the question is the real gate, and why every draft carries a page number. In Phase 1 a scholar checks every AI item *before* it is played. If one slips through anyway, we correct it in the room and pause the feature until we know why. But the comparison is not "AI versus perfection". On 19 Sep a **hand-written** question asked for "भरत क्षेत्र के एक तीर्थंकर का नाम तथा जिन्होंने आदिनाथ मुनिराज को प्रथम आहार दिया". The expected answer was श्रेयांस. But the first clue points to Shreyansnath, the tirthankar, and the second to King Shreyans, so the 6 players who wrote श्रेयांसनाथ or श्रेयांस कुमार were not wrong. Only 1 of 36 scored. Human questions have errors too. The difference is that an AI draft arrives with its source attached, so the error is easier to catch before the show, not after.

2. Anticipated tough questions

#	Question	Asked by	My answer
1	The key sits in a browser and calls Google directly. Isn't that insecure? And the free tier will fail on show day.	Engineering	The key is the host's own, stored only in that browser, and tied to a Google project with no billing . The worst case is someone using up a free quota; no NGO or player data can be reached with it. For show day: prepare the day before, and rely on automatic fallback across up to 8 free models. In testing, four models were busy at once and the batch still finished. The manual paths never depend on AI.
2	Hosts will just tick all ten. The review screen becomes a rubber stamp.	Design	Drafts the checker disagreed with start unticked and show the checker's answer. Batches from a lighter model carry a warning. We log approval and edit rates. If approval is close to 100% with no edits while audits find errors, the review is too easy, and I would require each card to be opened before it can be added.
3	The books say "For Private & Personal Use Only", and free-tier inputs may be used by Google. Is this allowed?	Legal	Use is non-commercial religious teaching by an NGO. Only the chosen pages are sent, quotes are capped at 10 words, and we store no pages. The screen tells the host what is sent. Before Phase 3, when partners outside the NGO join, I would ask the publisher (Jain e-Library) for written permission and offer a list of cleared books.
4	It's free today. What happens when Google starts charging, or cuts the quota?	Finance	The NGO's rule is ₹0, so nothing is billed automatically: the key's project has no billing at all. If the free tier ends, the feature fails safely and hosts go back to manual entry. The NGO committee then decides between fewer pages per batch or a small donor-funded key. It never runs on the billed Firebase project, and the blind check is never removed to save a call.
5	Whose interpretation does the AI follow? Our tradition, or whatever it read on the internet?	NGO scholar	The book's. The prompt says "Use ONLY what is written in these pages... Follow the book's own view. No comparison with or criticism of other traditions." The host chooses the source. A scholar checks AI items in Phase 1 and audits them monthly after that.
6	How do you <i>know</i> it saves time and keeps engagement?	NGO committee	I don't yet. The 1-hour baseline is the host's own estimate, and the feature has not been used in a live show. File 05 sets out a within-show test over 3 shows, with timed prep, and response rate as the guardrail. Until those results exist the feature stays labelled Beta.

3. AI tool usage log

Who did what. Me: the NGO context and the problem, running the three shows, collecting the feedback, choosing the friction and features, setting the constraints (free only, human approval), supplying the book, creating the key, testing, and approving every release and every number in these files. **Claude (Anthropic), through Claude Code:** writing and testing the code to my specification, analysing the show data, capturing screenshots, and first drafts of these six documents.

Google Gemini: the AI inside the product.

Tool	Purpose	Prompt used (my words, verbatim where available)	How I reviewed or changed the output
Claude Code (Anthropic), Sept 2026	Building the platform: Studio, host panel, projector, phone play, hybrid tambola, security rules	Iterative build requests over several sessions. The standing constraint, in my words: <i>"we don't come to pay any money we have to everything in free"</i> and <i>"if that will charge then leave it find another way"</i> .	Ran every version in rehearsal mode and then in three live shows (19, 20, 25 Sep). Kept only Hosting and Realtime Database (₹0). Asked for fixes after the shows.
Claude Code , 14 Oct	Designing the AI feature	<i>"tell me how we can add ai into this that give real value to game. (like If I give pdf of source book and it will create questions from that file by analyzing book set level low medium or hard)"</i>	Chose to build it into the live Studio rather than as a separate prototype ("go with a"). Supplied the source book. Confirmed the 1-hour baseline. Set "we have to go for free everything".
Claude Code , 14 Oct	Building and testing 🔥 Draft with AI	Same thread: build in the live Studio, test with <i>बृहद् द्रव्य संग्रह</i> pages 41–60, then <i>"publish"</i> .	Created the free key myself in a no-billing project. Reviewed test drafts, which led to the blind check, option shuffle, difficulty rule and model fallback. Approved the deploy only after the rehearsal tests.
Claude Code , 14–15 Oct	Analysing show data (read-only, aggregated)	<i>"I think 19th game has been deleted by me by mistake, is there any option to recover or if it presents user data so see it."</i>	It found the game had only been renamed, and exported the player list to a private local file. I confirmed which sessions were real shows and the head-counts. No personal data appears in these files.
Claude Code , 15 Oct	Screenshots and first drafts of these 6 files	<i>"go in this file you can mention that it is build with help of Claude. if needed."</i>	Checked every claim against the data and the product. The quotes are my players' words, unchanged. Figures 4.2–4.5 replay a real AI run so they print sharply; Figure 4.6 is labelled as a simulated outage.
Google Gemini (flash family, free tier)	Inside the product: drafting questions, and the blind check	The product's own prompts (File 04, section 3): "Use ONLY what is written in these pages...", the difficulty rubric, the two-clue examples from our own game.	Every draft goes through the blind check and then human approval. Spot-check of the 15 Oct batch (by Claude, to be confirmed by the NGO's scholar): 8 of 10 usable as written; #5 had a garbled quote and #9 awkward wording.
Playwright + Microsoft Edge	Capturing screenshots in rehearsal mode (no AI)	n/a (automation script)	Test players are marked "(test)". No live show data was touched.

4. What I would do next

Prove it

Run the 3-show test in File 05 with timed prep. Run shadow scoring on the 19 Sep answers. Name a scholar for Phase 1 checks.

Make review faster

Show a thumbnail of the cited page next to each draft. Count drafts generated, so the approval rate is exact. Add a remote on/off switch.

Close the other frictions

F5: a "first question coming" nudge on phones. A "report this question" button for players. Large-text mode for elders (F3).

Six files: 01 Product Audit & Friction Analysis · 02 AI Feature Spec & PRD · 03 Responsible AI & Risk Plan · 04 Build Evidence & Wireframes · 05 Metrics, Experimentation & GTM · 06 Stakeholder Defence & AI Tool Log. Product live at sarvodayahinsa.web.app (audience page: code entry only; hosts sign in).