AI REIMAGINE PROJECT · FILE 06

Panel Defence & AI Tool Appendix

FitSense - the argument, the hard questions, and how it was built

Prepared by Abhishek Katkar · June 2026

The argument in one line. FitSense turns Myntra's opaque, history-only size suggestion into a garment-aware, explained, confidence-scored recommendation that finally works at cold-start - and it is honest about what it does not know. Every file in this project serves that one idea. Below are the questions a panel should ask, answered plainly, followed by the AI tools used to build it.

1. Panel Defence

Q. Aren't you just rebuilding Myntra's existing recommendation engine?

No. Their recommendation is purchase-history-only: silent, garment-blind, and absent at cold-start. FitSense deliberately does not try to out-predict the history signal for repeat buyers - that signal is grounded in real fit outcomes and is hard to beat. It wins on a different axis: coverage (first-ever, new-brand, new-category), garment-awareness (it reads the size chart and construction), and trust (every call carries a reason and a confidence level). The two are complementary, not competing.

Q. Cold-start inputs are noisy - 67-73% of shoppers don't measure accurately. Why trust an AI here?

We don't pretend to be certain. The non-negotiable design rule is confidence honesty: when inputs are thin, FitSense says "Low confidence" and invites a refinement instead of bluffing a size. That is why calibration (ECE) is our headline evaluation metric - "High" has to mean high accuracy - and why a confident-wrong answer is the specific failure we engineered against, with the low-confidence fallback you can see in the prototype.

Q. Why hasn't Myntra already built this?

They took the high-coverage, structured-data win first - a history-based recommendation off existing purchase data. The remaining layer (cold-start, garment-awareness, explainability) is harder: it needs construction data extracted from unstructured descriptions and a careful confidence UX, and the per-SKU return on that work looked lower. Naming exactly that gap - high value, structurally unaddressed - is the product judgement this project is built on.

Q. What happens when the AI is wrong or unavailable?

It fails safe, and the prototype shows it. Low confidence becomes an honest nudge plus "Refine my fit"; an unavailable model degrades to the static size chart so the PDP never blocks; a flag control and the post-delivery kept/returned signal feed the learning loop. Above that sit kill-switches: a calibration breach or a fairness-gap breach pulls the recommendation.

Q. Won't an A/B test just lose to the history signal?

A global head-to-head would - and that is the point. The primary metric is scoped to the cold-start / cross-brand cohort, where FitSense's advantage is structural. Size-related return rate on

that cohort, after the return window, is the North Star; calibration and the fairness gap are launch gates. Measuring where the feature actually wins is a deliberate choice, not a hedge.

Q. Isn't using body measurements a privacy risk?

Measurements are strictly opt-in, encrypted in transit and at rest, never shared with sellers, viewable and deletable by the user, minimally retained, and used only to suggest size. The prototype surfaces this in plain language ("How this works" and the data note), because transparency is part of earning trust at the size-decision moment.

Q. Does it work fairly across body types?

Accuracy is evaluated by body-size band, gender, and category, and a maximum accuracy gap is both a launch gate and a kill-switch. We actively collect under-represented fit data so the model does not quietly serve some bodies worse than others.

Q. Why this friction over returns logistics or substitution?

Fit at the point of purchase is the largest AI-suited friction: sizing confusion is roughly half of returns by Myntra's own account, and the existing recommendation leaves clear, nameable gaps. Logistics and inventory frictions are real, but they are not model problems - no recommendation fixes a stock-out.

2. Honest Limitations

Stated plainly, because the feature's credibility comes from not over-claiming.

- **Construction-data coverage is the real constraint.** Garment-awareness is only as good as the per-SKU data we can extract; sparse-data SKUs degrade to confidence-scored chart guidance.
- **The prototype proves the mechanism, not the population.** It runs on one garment with seeded returning- and cold-start profiles; production needs offline calibration on historical orders before any user sees a recommendation.
- **The outcome has a lag.** Return-rate results only mature after the return window, so the experiment must be powered and run long enough to read honestly.

3. AI Tool Appendix

This project was built with AI tools, used deliberately - with every product decision owned by me.

AI tool	How it was used	What I owned
Claude (Anthropic)	Research synthesis and friction analysis; drafting of the PRD, responsible-AI plan, and metrics/GTM; generation of the step-by-step build prompts; and a critical thinking partner that pushed back on weak framing.	Product and flow selection; the friction to solve; the interaction model; the scoped-win positioning; every screenshot; and validation of every claim.
Lovable	AI-assisted, prompt-to-app full-stack build	The build sequence, the Myntra

AI tool	How it was used	What I owned
	of the working prototype (UI and backend).	design-fidelity bar, and acceptance of each step.
Lovable AI Gateway -> Google Gemini	The runtime LLM that powers FitSense's live recommendations in the prototype.	The system prompt and the strict-JSON contract that enforce confidence honesty.

Prompting approach

- Reference-driven: study strong examples before generating, so the output has a standard to meet.
- Step-by-step build prompts, each validated against a running screenshot before moving to the next.
- A verbatim system prompt enforcing strict JSON and confidence honesty - the model's behaviour is pinned, not hoped for.
- Honest framing over flattery: when the PDP turned out to already carry a recommendation, the friction was reframed ("garment-blind, unexplained, absent at cold-start") rather than papered over.

4. Closing

This completes the six-file submission. FitSense is honest about a hard problem: it does not claim to beat a signal it cannot beat, and it says so when it does not know. That restraint - garment-aware, explainable, confidence-scored, and safe when wrong - is what makes it defensible.