

PREPARED FOR PAPPARTI

Papparti — Project Proposal

What Vector Logic proposes to build for Papparti, and why it is shaped this way.
For the decision maker: scope, approach, phases, and what is deliberately excluded.
Asks for: a go-ahead on the scope, or a list of changes first.

AT A GLANCE

The product	A voice-first on-demand services marketplace
Three trained models	Speech recognition, speech synthesis, reasoning
Runs where	Self-hosted on a dedicated GPU; the models become yours
Delivery	Three software phases, after a measured feasibility phase
One-time fee	78,72,318 excluding GST

DATE 11 October 2026	VALIDITY 30 days; prices re-quoted after 10 November 2026.	CONFIDENTIALITY Confidential. Prepared by Vector Logic for Papparti. Not for distribution beyond Papparti and its advisers.
--------------------------------	--	---

Contents

Papparti — Project Proposal	3
Section 1 — What you are buying	3
What "self-hosted" buys you commercially	3
What is delivered	3
Section 2 — Why voice-first, and why now	4
The evidence that supports the approach	4
The honest finding: an avatar as the WHOLE interface is the weakest commercial case	4
How this proposal mitigates it	4
Section 3 — The three models	5
3.1 Speech recognition ("hearing")	5
3.2 Speech synthesis ("speaking")	5
3.3 Reasoning model ("deciding") — runs 24/7	5
3.4 The voice recording workstream (Vector Logic)	5
Section 4 — What we will prove before we build at scale (Phase 0)	6
The hard gates	6
What a failure means, commercially	6
Cost and duration	6
Section 5 — Scope, phase by phase	7
How the phases fit together	7
Optional add-on, priced separately (not in the total above)	7
Section 6 — Commercials	7
5.0 The demo set, and what each artefact proves	7
6.0 Validity, and the fields we need from you	8
6.1 One-time project fee	8
6.2 Monthly after handover	8
6.3 What is included	9
6.4 What is not included	9
6.5 Payment milestones	9
6.6 On the budget band	9
Section 7 — What you get at handover	10
Section 8 — Risks, stated plainly	10
8.1 Risks with no mitigation	10
8.2 Mitigated risks	11
8.3 Commercial risks you should weigh	12
8.3 The talking avatar — what you are seeing, and what you are not	12
8.4 The voice samples you can listen to now	13
Section 9 — What we need from you	13
Section 10 — Assumptions	14
Explicitly unverified, and excluded from every total	15
What we could not verify	15

Papparti — Project Proposal

Project: Papparti — on-demand any-service marketplace with a self-hosted voice and avatar layer **Prepared for:** Papparti **Prepared by:** Vector Logic — AI & software development, Odisha, India **Date:** 11 October 2026 (IST) **Currency:** All figures in INR. USD prices are converted at a pinned rate of **1 USD = ₹96.7384** (XE mid-market, 11 October 2026, 04:05 UTC) — the rate used in our cost estimate.

SECTION 1 — WHAT YOU ARE BUYING

You are buying four things, delivered together, as one system:

1. **The marketplace software.** A two-sided on-demand services platform: customer booking, provider onboarding and working hours, duration-aware scheduling, payments with a provider split, an immutable ledger, an admin console, and a provider app. Delivered as a Progressive Web App (PWA) — installable from the browser, no app store.
2. **A speech recognition model**, custom-trained on your audio and your service vocabulary.
3. **A text-to-speech model**, producing a voice that belongs to Papparti.
4. **A reasoning model** that runs continuously in the backend, turning a spoken Hindi/Hinglish request into a structured, confirmed booking.

Plus the voice itself: your brand voice is **recorded by Vector Logic** in a studio session (engineering, casting, direction, QC and the recording are a priced line in this proposal — see Section 3).

What "self-hosted" buys you commercially

Speech recognition and speech synthesis run **on infrastructure you control**, not through a third-party cloud speech API. The commercial consequences:

- **You own it.** On handover, the three trained models become Papparti's property, along with the code. You are not renting access to a service that can be repriced or withdrawn.
- **No per-request fee to a third party.** Recognition and synthesis do not meter per call to an external vendor. The marginal cost of one more spoken booking is your own compute, not someone else's rate card.
- **Your users' audio does not leave your stack.** Sending customer voice to an outside speech API would export your users' audio to a third party; self-hosting does not.
- **This is not the cheaper option, and we will not pretend it is.** Our own research is explicit: hosted cloud speech would cost less per booking than running your own GPU tier. Self-hosting is a **strategic and compliance decision, not a cost saving**, and it should be budgeted as one. You are buying control and independence with money you could otherwise spend elsewhere.

What is delivered

ITEM	DELIVERED
Marketplace software (customer + provider sides, admin, payments, ledger)	Yes
Speech recognition model — trained on your data	Yes
Speech synthesis model — trained on your data	Yes
Reasoning model — trained on your data, runs 24/7	Yes
Brand voice recording (studio session, ~20 h delivered audio)	Yes — Vector Logic workstream
Self-hosted deployment on rented GPU	Yes
Code, model weights, endpoints, runbooks, knowledge transfer	Yes

ITEM	DELIVERED
30-day warranty after handover	Yes

One scope point stated up front, in writing: the three models are **fine-tunes**. The base model's weights are frozen; your data trains an *adapter* on top of it (the industry-standard method, LoRA/QLoRA). This is not a train-from-scratch pretrain. Our own decision record flags this as something that must be closed in writing before signature, because "retrain" can be read to mean more than it delivers. This is what "retrain" means here.

SECTION 2 — WHY VOICE-FIRST, AND WHY NOW

The evidence that supports the approach

- **India's market is conversational, and largely not app-native.** Our research's own framing of the strongest opportunity is an **agent-native, voice-first, vernacular-first marketplace for India's ~98%-offline market, reachable without a smartphone app** — closer to a conversational agent than a photoreal avatar. Hindi/Hinglish speech is the natural interface for that user.
- **Voice ordering has already been proven in India at scale.** Hindi + English voice-first ordering shipped in this market in April 2026. The interface is no longer speculative — which also means it is no longer unique. See the honest finding below.
- **The delivery model is reachable.** A PWA on a WebGL2 browser works on Indian Android phones without WebGPU, without downloading model weights, and without the Play Store. This is what makes a voice-and-avatar product viable on the device band this business must serve.

The honest finding: an avatar as the **WHOLE** interface is the weakest commercial case

Our research measured the market and did not soften its conclusion. You should have it in front of you before you commit money:

- **"An avatar as the entire UI on both sides" is the only genuinely unshipped square in this space — and the research calls it the weakest commercial case in the plan.**
- The evidence against avatar-first commerce is real: **only about 22% of smart-speaker users buy by voice**, and **about 45% will not send a payment through a voice assistant.**
- Voice-first Hindi/English booking is **already shipped in India** (April 2026, 11 languages), so voice alone is not a differentiator.
- Self-hosted Indic speech is feasible but **not unique and invisible to users** — a margin and compliance play, not a customer-facing feature.

How this proposal mitigates it

We build what you asked for, with one structural safeguard:

1. **Confirm-before-price stays in text, always.** The model **proposes** a duration; the customer **confirms** it; **only the confirmed value is priced**. A conventional on-screen confirmation, receipt and ledger stay in place for every booking and every rupee — not voice-only. An inferred duration that becomes a bill is, in our research's words, "the first trust-destroying event," and we treat it as a hard, unbreakable gate.
2. **The avatar is a differentiator, not the foundation.** It is layered on top of a booking engine that works without it. If the avatar underperforms, the business does not.
3. **We test the thesis cheaply before spending the big money.** Section 4's Phase 0 measures whether the voice stack actually works in Hindi before full development spend.

This is not a reason to stop. It is a reason to decide deliberately.

SECTION 3 — THE THREE MODELS

All three are **fine-tuned on Papparti's data** and **self-hosted**. Parameter scales are given; the specific base models are commercially confidential to Vector Logic and are deliberately not named in this document. What we can tell you is the data each is trained on and what it does.

3.1 Speech recognition ("hearing")

Parameter scale	1-2 billion parameters
Trained on	Public Hindi and Hinglish speech corpora (~150-250 hours total: spontane
What it does	Turns spoken Hindi, English and Hinglish into text for the booking flow
Where it runs	Self-hosted on a rented 48 GB GPU. Not a cloud speech API
Licence posture	Trained on commercially-licensed data only

3.2 Speech synthesis ("speaking")

Parameter scale	0.3-0.5 billion parameters
Trained on	Papparti's brand voice — ~20 hours of studio-recorded audio of one ch
What it does	Speaks the booking conversation in Papparti's own voice, in Hindi and Eng
Where it runs	Self-hosted on the same rented 48 GB GPU
Design requirement	Must survive being interrupted mid-sentence (a customer talking over

3.3 Reasoning model ("deciding") — runs 24/7

Parameter scale	~32 billion parameters
Trained on	Synthetic booking dialogues built from a grammar of your booking domain
What it does	Converts a Hindi/Hinglish/English conversation into a validated structured
Where it runs	Self-hosted on the same rented 48 GB GPU
The non-negotiable behaviour	On an ambiguous request ("thoda kaam hai", "jaldi aao") it must a

3.4 The voice recording workstream (Vector Logic)

LINE	INR
Voice production — casting, session direction, QC, retakes, mastering (2 work-weeks)	2,70,868
Studio recording — ~20 h delivered audio (pass-through)	97,000
Voice production total	3,67,868

Vector Logic records the brand voice itself. This converts what was the single biggest unmitigated blocker in the model plan — voice cloning of a not-yet-chosen speaker, blocked on a human decision, a legal act and money — into work we control.

One obligation this creates, and it is not optional: the voice artist must sign a **perpetual, irrevocable, worldwide, transferable commercial release** in Papparti's favour. The model is handed to you; a revocable release would break your product the day the artist withdrew. Legal review of that release is a Phase 0 item.

SECTION 4 — WHAT WE WILL PROVE BEFORE WE BUILD AT SCALE (PHASE 0)

The single largest unverified assumption under this entire plan is that the chosen models actually produce good Hindi. No report has tested it. Every quality figure in our research is a vendor claim or a model-card number, not a measurement on our own hardware.

Phase 0 exists to replace those claims with our own measurements, on one rented 48 GB GPU, over one month, before the big development spend.

The hard gates

If a gate fails, the project changes or stops — it does not proceed on hope.

GATE	THRESHOLD	WHAT A FAILURE MEANS
Hinglish code-switch recognition accuracy (before any fine-tune)	worse than 20% word error rate	Stop/redesign. The interface becomes text-first from day one; the voice-first UX is cut
Hindi recognition accuracy (best base, after fine-tune)	worse than 12% word error rate	Change the base model — the chosen one cannot carry Hindi
Fine-tune improvement over the untrained base	under 15% relative gain	The fine-tune has not earned its place; ship the base and re-scope
Voice naturalness (human listeners, ≥30 raters)	below 3.5 on a 1–5 scale	The voice is not trustworthy enough to be the interface; fall back and de-prioritise the brand voice
Voice number/duration intelligibility	below 100%	Hard gate. A mis-spoken price is a money error. Do not ship synthesis on the money path
Booking tool-call accuracy (Hindi/Hinglish)	below 85% exact match	Change the model
Ambiguous-duration inference rate	above 0%	Hard stop on the reasoning model as specified. Inferred-duration pricing is the catastrophic failure
General-capability regression after fine-tuning	above 3%	The fine-tune is over-fitting; retrain with more anchor data
Concurrent sessions per GPU node	below 10	The per-node cost model breaks; re-size the fleet
p95 time to first audio	above 1,500 ms	The avatar is silent too long; "feels broken"

Phase 0 also settles the concurrency number under real load (a 3× error here triples the infrastructure bill) and resolves two contradictory Hindi accuracy figures carried in our research by measuring them ourselves.

What a failure means, commercially

We tell you, and you do not spend the big money. Phase 0 is priced so that a negative result is cheap. If the voice stack cannot be made to work in Hindi at the quality the product needs, you find out at Phase 0 cost — before Phase 1–3 development and before the full model workstream — and we say so in writing.

Cost and duration

LINE	DETAIL	INR
GPU rental — 1× 48 GB dedicated	730 hr @ \$0.59/hr	41,665
Engineering — discovery, benchmark harness, Hindi STT/TTS/LLM evaluation, written feasibility report	3 work-weeks	4,06,301
Phase 0 total		4,47,967

Duration: one month of GPU time plus 3 work-weeks of engineering.

What you receive: a written feasibility report with our own measured numbers and a go/no-go on each model.

SECTION 5 — SCOPE, PHASE BY PHASE

PHASE	WHAT IT DELIVERS	WEEKS	INR (EXCL. GST)
Phase 0	Discovery + Hindi/AI feasibility benchmark — one rented 48 GB GPU, one month	3 work-weeks + 1 month GPU	4,47,967
Model work	Four workstreams: speech recognition, speech synthesis, reasoning LLM, voice production	22 engineering weeks	30,90,471
Phase 1	Voice core + booking engine, audio-only — self-hosted speech in and out, duration-aware booking, payments, provider split, immutable ledger, provider onboarding and hours, PWA shell	12	16,25,205
Phase 2	Avatar + polish — the avatar interaction layer on both sides (rendered on the phone), UX polish, hardening	8	10,83,470
Phase 3	Second vertical / locality on the same platform — catalogue, pricing rules, provider cohort, localisation	8	10,83,470
Deployment + handover	Self-hosted bring-up, CI/CD, monitoring, runbooks, admin training, 30-day warranty	4	5,41,735
ONE-TIME TOTAL (excl. GST)			₹78,72,318

How the phases fit together

- **Phase 1 gate:** an end-to-end real booking by voice, paid, the provider paid on completion, the ledger reconciles, **and two overlapping bookings are correctly rejected by the database, not by application logic.**
- **Phase 2 gate:** the avatar adds under ~800 ms to p95 latency; captions land inside 500–800 ms.
- **Phase 3 gate:** only starts after vertical 1 is per-order profitable **including GPU cost.**

Optional add-on, priced separately (not in the total above)

OPTION	WEEKS	INR
Phase 4 — real-time lip-synced avatar	10 (starting estimate)	13,54,338

Phase 4 is the highest-risk, least-proven scope: true real-time neural lip-sync is server-video-cost territory (~9–11 MB per user-minute against ~9 KB/min for the on-phone avatar). It is priced as an option with a defined starting estimate and **excluded from the one-time total**. It is not a launch dependency.

SECTION 6 — COMMERCIALS

5.0 The demo set, and what each artefact proves

ARTEFACT	WHAT IT DEMONSTRATES	WHAT IT DOES NOT CLAIM
architecture.html	Where every piece runs — the phone, our self-hosted GPU service, and your existing booking backend — and why the avatar renders client-side (no per-frame server cost)	Any measured latency or throughput. Those come from Phase 0.
avatar/high-tier-*.png, low-tier-*.png	One character in two rendering tiers, so a low-end and a high-end phone both get a talking avatar	Photoreal. This is a vector rig, chosen because a photoreal real-time face is not deliverable on a mid-range phone.
avatar/avatar-speaking.mp4	The face audibly speaking Hindi — the real rig, the real synthesised voice, muxed	Lips shaped by the actual phonemes. The mouth is driven by a speech envelope, so it is in time with the speech , not shaped by it .

ARTEFACT	WHAT IT DEMONSTRATES	WHAT IT DOES NOT CLAIM
voice/conversation/*	A three-turn dialogue with a time and a price in it — the two places a mistake becomes a money error	A live model output, or a human recording. It is a scripted demo; both roles are synthesised.
voice/*.mp3	Two Hindi voices measured on the same sentences	The final brand voice. That is produced in the studio session priced in this estimate, which has not run yet.

We have deliberately labelled every one of these with what it is not. A demo that implies more than it shows costs more to correct in Phase 1 than it saves in Phase 0.

6.0 Validity, and the fields we need from you

This pricing is valid for 30 days from 11 October 2026. The reason is commercial, not legal: the monthly figures rest on live cloud GPU and hosting prices read on 11 October 2026, and those move. After 30 days we re-price the affected lines from then-current vendor prices. The engineering rate (USD 35/hr, ₹3,385.84/hr) is the part that does not move — it is our rate, not a vendor's.

Two fields must be filled before this is signature-ready, and we have deliberately left them blank rather than guessing:

FIELD	WHY IT MATTERS
Your legal entity name as it should appear on a contract and an invoice	The address in this document reads "Papparti", which is a brand. An invoice
Who signs, and in what capacity	So acceptance is unambiguous and the right person receives each milesto

We have also used contact@papparti.com only as a placeholder in the demo references — tell us the address that reaches you and we will correct it everywhere before sending anything further.

The full commercial annexure — payment milestones, acceptance criteria per milestone, what happens if a Phase 0 gate fails, invoicing and tax, scope changes, and termination — is a separate document: 04 - COMMERCIAL - TERMS.md.

6.1 One-time project fee

	INR
One-time total, excluding GST	₹78,72,318
GST @ 18% (SAC 998314, software development service): ₹78,72,318 × 0.18	₹14,17,017
One-time total, including 18% GST	₹92,89,335

The ₹14,17,017 GST line is arithmetic on the ₹78,72,318 excluding-GST total: ₹78,72,318 × 0.18 = ₹14,17,017 (rounded to the rupee), and ₹78,72,318 + ₹14,17,017 = ₹92,89,335 — the including-GST figure quoted in our cost estimate.

6.2 Monthly after handover

ITEM	INR / MONTH
GPU serving — 1× 48 GB dedicated, 24/7	41,665
Managed Postgres + vector search	5,891
Redis	1,935
Real-time media server (SFU/TURN)	4,837
Object storage	1,161
Monitoring + CI	2,902
Support retainer (0.5 engineering work-week / month)	67,717
MONTHLY TOTAL	₹1,26,108

The support retainer is deliberately larger than the entire GPU bill. A live 24/7 voice system needs ongoing engineering, not just hardware. The GPU line assumes **one** 48 GB card; a production serving tier with real concurrency may need a second (~₹83,000/month).

6.3 What is included

- All six phases in Section 5 (Phase 0 through deployment and handover).
- All four model workstreams, including the brand voice recording session.
- Self-hosted deployment on rented GPU.
- Runbooks, CI/CD, monitoring wiring and admin training.
- A **30-day post-handover warranty**.
- Full handover of code, trained model weights and endpoints.

6.4 What is not included

- **Payment-gateway fees.** These are transaction-linked, not fixed. The gateway charges a flat 2% **platform fee per successful domestic transaction, plus 18% GST on that fee** — no setup, AMC or refund fees. On ₹1,00,000 of settled volume that is ₹2,000 + 18% GST = **₹2,360**. This is not in the monthly total above.
- **GST on imported cloud/GPU services.** Cloud and GPU services bought from foreign providers are an **import of services**. The Indian recipient must self-assess **18% IGST under the Reverse Charge Mechanism**, pay it in cash, and generally claim it back as input tax credit. India-hosted GPU providers invoice domestically and avoid this mechanic, at a higher per-hour price.
- **Counsel and CA fees** — the aggregator classification opinion, the payment/GST posture, and the voice-release legal review are obtained by Papparti. Not modelled here.
- **Phase 4** (real-time lip-synced avatar) — a separate ₹13,54,338 option.
- **Native iOS/Android apps** — the product is PWA-first.
- **Languages beyond Hindi and English.**
- **Multi-vertical launch** — Phase 3 adds a second vertical to the same platform.
- **The second 48 GB serving GPU**, if concurrency demands it.

6.5 Payment milestones

Priced per phase, invoiced as each phase completes:

MILESTONE	INR (EXCL. GST)
Phase 0 — feasibility benchmark delivered	4,47,967
Model work — four workstreams	30,90,471
Phase 1 — voice core + booking engine	16,25,205
Phase 2 — avatar + polish	10,83,470
Phase 3 — second vertical / locality	10,83,470
Deployment + handover	5,41,735
Total	78,72,318

6.6 On the budget band

Our cost estimate's verdict is that the expectation of a **₹20–30 lakh** one-time fee is **not defensible** — it is roughly one third of the real cost of this scope. The arithmetic:

- ₹78,72,318 is 3.15× the **₹25 lakh midpoint** of that band.

- The band's **ceiling (₹30 lakh)** would cover only the **model work (₹30,90,471, four workstreams)** and **nothing else at all** — the trained models plus the voice recording they require exceed the entire ₹30 lakh ceiling on their own, before a single line of marketplace software, booking engine, payments, ledger, admin, provider app, deployment or handover is written.
- To fit ₹30 lakh, scope would have to be cut to roughly **one** of: the model work, or Phases 0–1 of the software. It cannot cover both.

Where the money goes, as a share of the one-time total:

COMPONENT	INR	SHARE
Software development (Phases 1–3)	37,92,145	51%
Model work (4 workstreams, incl. voice production)	30,90,471	39%
Feasibility benchmark (Phase 0)	4,47,967	6%
Deployment + handover	5,41,735	7%
Total	78,72,318	100%

The band was priced like a marketplace app. It is a marketplace app **plus three custom-trained AI models plus a self-hosted voice stack**.

Quote: ₹78,72,318 excluding GST (₹92,89,335 including 18% GST) for the defined scope, ₹13,54,338 for the Phase 4 option, and ₹1,26,108 per month to run it.

SECTION 7 — WHAT YOU GET AT HANDOVER

DELIVERED ITEM	DETAIL
Source code	The full marketplace software — customer and provider sides, admin, pay
Model weights	All three trained models, including the trained adapters, become
Endpoints	Self-hosted serving endpoints for speech recognition, speech synthesis an
Infrastructure-as-code	The self-hosted stack defined as code, on rented GPU
CI/CD	Build and deployment pipelines
Monitoring	Wired monitoring for the live service
Runbooks	Operational documentation for running the stack
Admin training	Training for your team on the admin console and operations
Knowledge transfer	Structured handover of how the system works and how to run it
30-day warranty	Post-handover defect warranty
Voice-rights release	A perpetual, irrevocable, worldwide, transferable commercial release sign

Ownership: on handover the models and the code are yours. You are not renting access to them. This is the "100% first-party development" promise made legible: you buy the asset.

SECTION 8 — RISKS, STATED PLAINLY

8.1 Risks with no mitigation

These are stated honestly. None of them has a mitigation that removes the risk.

1. Hinglish accuracy may not be solvable to product grade with this method.

Our research's evidence is that even dedicated fine-tuning leaves code-switched word error around 10–14%, and the problem is *not solved*. There is **no mitigation that makes Hinglish speech recognition as reliable as monolingual recognition**. The only real mitigations are product design (confirm in text) and scope (accept the residual error). If voice-perfect Hinglish is a requirement, that is not something this project can deliver, and we are saying so now rather than later.

2. If you supply no real audio and no real transcripts, the real-world accuracy of the models cannot be stated.

The synthetic-to-real gap can only be measured against real held-out data. If none exists, we ship models whose real-world accuracy we **cannot state**. This is a data-availability limitation, not a technical one — no amount of engineering removes it.

3. Catastrophic forgetting can be bounded but never eliminated.

A fine-tune **will** shift the base model's behaviour; general capability can regress. We can keep the regression under 3% and measure it before and after, but we cannot guarantee the model keeps every general capability it started with. The honest position is "bounded and measured," not "none."

4. No listening test has been run, and no MOS (mean opinion score) exists for our voice output.

We have generated audio and measured it objectively — duration, peak level, RMS, the proportion of non-silent frames, and the real-time factor on this hardware — but **no human listener has rated it, and we quote no quality score**. Any model-card quality figure you may have seen elsewhere is not a substitute for a listening test on your own script and your own brand voice. That test is part of Phase 1, and we will run it before the voice is finalised.

5. The two conflicting Hindi accuracy figures in our research cannot be reconciled from the documentation.

They can only be replaced by our own measurement. Until Phase 0 runs, any Hindi number quoted from a model card is unreliable, and that uncertainty is irreducible until we spend the measurement.

8.2 Mitigated risks

RISK	SEVERITY	MITIGATION
Hinglish speech recognition is genuinely hard	High	Fine-tune on real and synthetic code-switched data; Phase 0 measures the pre/post gap; if it fails, ship text-first
Base Hindi quality is unverified	High	This is what Phase 0 exists for. No fleet spend before the measurement
If Hinglish cannot be fixed, the voice-first UX is unsound	High	Confirm-before-price in text, always. The voice cost becomes a UX cost, not a money-safety cost
Catastrophic forgetting	Medium	Mix general Hindi anchor data into training; before/after capability probe; regression gated under 3%
Voice-cloning consent and legal exposure	High	A signed voice release is mandatory; we clone only the chosen speaker; the synthesis model watermarks its output
Client data too small	Medium	Synthetic-first bootstrapping; the synthetic-to-real gap is measured, not assumed
Share-alike data entering shipped weights	Medium	Prefer permissively-licensed corpora; counsel confirms before any share-alike data trains the shipped adapter; per-source training ledger kept
A base model's fine-tune licence differs from its base licence	Medium	Confirm the licence before shipping; otherwise fine-tune from a cleanly-licensed base ourselves
Synthetic data teaches synthetic artefacts	Medium	Real code-mixed phrasing; scenario-disjoint splits; a real held-out slice synthetic tuning never touches
Number/duration mispronunciation in synthesis	High	All numbers spelled as words in the frozen eval; 100% intelligibility gate ; numeric-safety grammar check before synthesis
Synthesis prosody fails under interruption	Medium	Clause-level chunking plus an interruption-naturalness test

RISK	SEVERITY	MITIGATION
Over-fitting the evaluation	Medium	Test sets frozen before training and stored read-only; independent adversarial verification each phase
Statistical evaluation too small	Medium	Minimum listener panel of 30+; word-error sets of 5+ hours audio; confidence intervals reported, not point estimates

8.3 Commercial risks you should weigh

- **Self-hosting is not cheaper.** Our research states it plainly: at plausible volumes, hosted cloud speech would cost less than running your own GPU tier. Break-even needs millions of bookings per month once an engineer is counted. Self-hosting is a strategic/compliance decision, not a cost saving.
- **The commission take-rate is under threat.** Per-booking commission is table stakes and is under attack by zero-commission competitors in the Indian market.
- **The payment gateway fee sets a floor on your take rate.** At a 2% platform fee plus 18% GST on that fee, if Papparti's per-booking commission is below ~2.4% of booking value, the gateway fee alone exceeds the platform's take on that transaction. Model the commission against this before fixing the take rate.
- **"Any service" is the trap.** Our research measured this: a broad any-service launch is the failure mode of this sector in India. A bounded first vertical is the sane MVP, and Phase 3 extends the platform to a second vertical deliberately.

8.3 The talking avatar — what you are seeing, and what you are not

Two rendering tiers of one character ship with this proposal, as four images:

FILE	TIER	STATE	WHY IT EXISTS
avatar/tierA-idle.png	High-configuration	Resting	What the avatar looks like on a capable phone
avatar/tierA-talking.png	High-configuration	Mid-utterance	The same face while speaking
avatar/tierB-idle.png	Low-configuration	Resting	The same character on a budget phone
avatar/tierB-talking.png	Low-configuration	Mid-utterance	Speaking, on a budget phone

What these are. The avatar is a **layered vector rig that animates on the client device** — not a rendered video and not a photoreal face. That is a deliberate choice, and it matters commercially: a rig ships as a few hundred kilobytes, animates at full frame rate on a mid-range phone with no GPU, needs no server round-trip per frame, and costs you nothing per conversation. A photoreal animated face cannot do any of those things, and **we are not going to show you a photoreal demo and then deliver a rig.**

What these are not. They are **not** a photoreal likeness, and they are **not** a finished art direction. They demonstrate two things only: (1) the same character can be rendered at two cost tiers so a budget-tier phone does not get a degraded experience, just a simpler one; and (2) the rig is built so the face actually redraws while speaking rather than being a still image that toggles.

What is measured and what is not. The tier difference is a measured fact, not an impression: a straight-line read of the skin's colour across the cheek returns **30 distinct tonal steps spanning 131 of 255** on the high tier against **1 distinct step spanning 0** on the low tier — a genuine 30× difference, checked by a script that fails the build if it ever collapses. The mouth region redraws by **35.8%** between the idle and talking frames. Horizontal overflow is **0 px** at 1920 / 1440 / 1024 / 390 px widths. **No human being has yet judged the art direction** — that is a decision for your team, and these four images are the input to it.

8.4 The voice samples you can listen to now

Two voice candidates were produced and shipped with this proposal. **They are real model output, not a studio mock-up and not a stock clip** — generated on our own CPU hardware and measured, not described. They are the *tone* reference for the brand voice, not a final deliverable: the shipping voice will be recorded in a studio session with a chosen speaker, which is a priced workstream (Section 6).

FILE	SPEAKER	LENGTH	WHAT TO LISTEN FOR
voice/papparti-voice-pratham.mp3	Male	~5.5 s	A short Hindi greeting — naturalness of the opening
voice/papparti-voice-priyamvada.mp3	Female	~6.8 s	The same greeting — pick a preferred register
voice/papparti-number-pratham.mp3	Male	~6 s	A sentence containing a price and a duration
voice/papparti-number-priyamvada.mp3	Female	~6 s	The same, female

Listen to the two number samples first. They are the harder problem and the one that carries money risk: if a system mishears or mis-speaks a price or a duration, the customer is quoted the wrong amount and the booking is wrong. Our own measurement found the number samples carry **lower speech energy than the prose samples** (–16.2 dBFS against –15.4 dBFS for the male voice), which is exactly why numerals get their own hard gate in Phase 0 rather than being assumed to work.

voice/README.md records how each file was produced and every number measured on it.

Our recommendation, and the reasoning, so you can disagree with it. Of the two voices, we recommend the **male** voice for the brand-voice recording session. The basis is measured, not a matter of taste: on the sentence that carries a price and a duration — the case where a mis-spoken numeral becomes a money error — the male voice sits further from clipping (–16.2 dBFS against –12.8 dBFS). Both voices run at the same speed, so performance does not decide it. This is a recommendation from a **fallback** engine, not a listening test: **your ear, on your own script, decides** — which is precisely why the studio session is a separate, priced line item rather than something we assume.

SECTION 9 — WHAT WE NEED FROM YOU

#	OBLIGATION	WHY IT BLOCKS	WHEN
1	Voice session scheduling and approvals — make the chosen speaker available for the studio recording, approve casting, approve the master	The brand voice does not exist until this happens; the synthesis model cannot train without it	Before model work completes
2	The service taxonomy — the actual service categories for launch	Drives the speech recognition vocabulary and the reasoning model's structured output. Generic lists do not carry the real booking phrases	Phase 0 onwards
3	The locality names — the launch locality's real place-names and their Hindi pronunciations	Drives the recognition vocabulary and the reasoning model's locality field	Phase 0 onwards
4	Real booking data once running — export transcripts and the confirmed booking records	Closes the synthetic-to-real gap; without it, real-world accuracy cannot be stated	After bookings run
5	Consent and notice handling under DPDP — data-retention and consent notice for stored audio and transcripts	Voice is sensitive personal data. Required before you store real customer audio	Before real data collection
6	A signed voice-rights release	No release = no clone. The release must be in Papparti's favour and irrevocable	Before the voice is cloned

#	OBLIGATION	WHY IT BLOCKS	WHEN
7	Business facts and a real contact email	The current holding page carries no business facts because none have been supplied. Nothing past a holding page is possible without them	Immediately
8	Counsel engagement — the aggregator classification, the payment/GST posture, and the voice-release review	Gates the financial schema and the payment half of Phase 1. The data model must be built migratable from day one	Phase 0
9	Pick ONE service vertical and ONE locality for launch	"Any service" from day one is the measured failure mode of this sector	Phase 0
10	Answer the existing-audio question — do call recordings, voice notes, or any real Hindi/Hinglish audio exist?	If you have zero audio, that is workable but must be known; it changes Phase 0's test design	Phase 0

Anything you can supply changes the result for the better. Real audio and real transcripts are the only way the shipped models' real-world accuracy can be stated at all.

SECTION 10 — ASSUMPTIONS

Every assumption below is copied from our cost estimate, with its basis and whether it is citable.

#	ASSUMPTION	VALUE	BASIS	CITABLE?
1	FX rate	1 USD = ₹96.7384, XE mid-market, 11 Oct 2026, 04:05 UTC	Live rate, pinned. All INR figures use it. Re-run if the rate moves materially	Yes — citable
2	Engineering rate	\$35/hour (₹1,35,434 per 40-hour work-week)	India AI-agency premium tier; Clutch lists India at \$25–49/hr; Eucalipse defines an "India (Premium)" tier at \$35/hr	Published bands — citable; the rate itself is our internal assumption
3	Work-week	40 hours, 5 days	Standard	Yes
4	Phases priced at midpoint	Phase 1 at 12 of 10–14 weeks; Phases 2–3 at 8 of 6–10 weeks each	Midpoint of the estimated range	Yes — it is a choice, stated
5	GPU hours per month	730 (24×7 dedicated)	24 × 7 × 365/12	Yes
6	Training/serving GPU	48 GB dedicated @ \$0.59/hr (₹41,665/month at 730 hr)	Live vendor price, read 11 Oct 2026. Provider selection is an implementation detail	Yes — citable
7	Model parameter counts	Speech recognition 1–2 B, synthesis 0.3–0.5 B, reasoning ~32 B	Design decision	Yes. Base models confidential
8	Support retainer	0.5 engineering work-week per month	Agency rate, \$6	Yes
9	Monitoring/CI and storage volume	~₹2,902/month; ~500 GB storage → ~₹1,161/month	Estimated, not quoted. The storage <i>rate</i> is live; the <i>volume</i> is assumed	No — assumption, not a quote
10	Payment-gateway fees	Excluded from the fixed monthly total; modelled as variable	Gateway 2% + 18% GST on the fee	Fee structure citable; it is variable, not fixed
11	Training GPU-hour counts	Speech recognition 40, synthesis 24, reasoning 180 GPU-hours	Not a vendor quote. Anchored to published fine-tuning benchmarks; the rate is live, the hours are estimated. The cost is robust anyway — GPU compute is ~0.5% of each workstream	No — estimate, not a quote
12	Studio recording cost	~₹97,000 for ~20 h at ~₹4,850/delivered hour	No studio quote obtained. Modelled band of ₹64,000–₹1,30,000 from India studio rates; midpoint used	No — estimate, not a quote

#	ASSUMPTION	VALUE	BASIS	CITABLE?
13	A second 48 GB serving GPU	Not priced into the totals	Flagged as a possibility if concurrency demands it	No — unpriced possibility
14	Phase 4 real-time lip-synced avatar	₹13,54,338	Starting estimate, not a quote. Excluded from the one-time total	No — estimate

Explicitly unverified, and excluded from every total

The following are assumptions or published-benchmark estimates, **not live quotes**. They are not added to any verified sum; where they appear in a table they are labelled, and the totals in Section 6 stand independently of them:

- Training GPU-hour counts for the three models (40 / 24 / 180 hours) — not a vendor quote.
- Studio recording cost (~₹97,000 for 20 h) — no studio quote obtained.
- Monitoring + CI monthly cost (~₹2,902) — an assumption, not a quote.
- Object-storage volume (~500 GB → ~₹1,161/month) — the rate is live; the volume is assumed.
- Any second 48 GB serving GPU — flagged as a possibility, not priced.
- Phase 4 real-time lip-synced avatar — open-ended; ₹13.54 lakh is a starting estimate.

What we could not verify

- **A published Indian dev-agency rate card.** No vendor publishes a fixed public rate for AI engineering; the \$35/hour figure rests on published India rate bands.
- **A quoted monthly or annual price from any GPU vendor.** GPU vendors publish per-hour prices only; monthly figures here are per-hour × 730.
- **Vendor-specific latency for India-served inference.** Not measured; this is why the pricier India-hosted GPU option is kept as a contingency, not the default.
- **Concurrency capacity of a single 48 GB card under real load.** Unmeasured; the monthly GPU line assumes single-stream serving. **Phase 0 measures this** — it is one of the hard gates in Section 4.
- **Any Hindi speech quality number from a model card.** Every quality figure our research holds is a vendor claim or a model-card number, not a measurement. **This is precisely what Phase 0 exists to fix**, and it is why we are proposing to measure before we build.

Prepared by Vector Logic, 11 October 2026. All figures traceable to our cited cost estimate. Base model identities are commercially confidential to Vector Logic and are not disclosed in this document.