

VECTOR LOGIC

AI & SOFTWARE DEVELOPMENT — ODISHA, INDIA

PREPARED FOR PAPPARTI

Papparti — Report Front Matter

DATE

11 October 2026

CONFIDENTIALITY

Confidential. Prepared by Vector Logic for Papparti. Not for distribution beyond Papparti and its advisers.

Contents

Papparti — Report Front Matter	3
A. Executive summary	3
The numbers	3
The timeline shape	4
The three decisions the client must make	4
The single next step	4
B. How to read this document	5
C. Glossary	6
D. The decisions this document asks for	9

Papparti — Report Front Matter

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, Indian numbering. USD converted at the pinned rate **1 USD = ₹96.7384**. **Status:** Front matter to the client package (00 - QUOTATION.md, 01 - PROPOSAL.md, 02 - ESTIMATE.md, 03 - GPU-DECISION.md, 04 - COMMERCIAL - TERMS.md). Every figure below traces to those documents.

A. EXECUTIVE SUMMARY

Papparti is buying one system with five parts, delivered together:

1. **The marketplace software** — a two-sided on-demand services platform: customer booking, provider onboarding and hours, duration-aware scheduling, payments with a provider split, an immutable ledger, an admin console, a provider app. Delivered as a Progressive Web App; no app store.
2. **A speech recognition model** — custom-trained so the system understands spoken Hindi, English and Hinglish and your service vocabulary.
3. **A text-to-speech model** — producing a voice that belongs to Papparti, recorded in a studio session by Vector Logic.
4. **A reasoning model** — running continuously in the backend, turning a spoken request into a structured, **confirmed** booking.
5. **A talking avatar** — layered on top of the booking engine, not as its foundation.

The numbers

ITEM	AMOUNT (₹)	WHAT IT IS
One-time project fee, excluding GST	78,72,318	Phases 0–3, model work, deployment and handover
GST @ 18%	14,17,017	Charged on top of the one-time fee
One-time fee, including GST	92,89,335	What a fully GST-registered Indian contract totals
Monthly running cost, after go-live	1,26,108	GPU, hosting plumbing, and a support retainer
Optional Phase 4 — real-time lip-synced avatar	13,54,338	Priced separately; not in the one-time total

The timeline shape

Six milestones, each tied to a **delivered, accepted artefact** — never to a date:

MILESTONE	WEEKS (ENGINEERING)	INVOICE INR (EXCL. GST)
Mobilisation advance (10%, recovered at Model work)	—	7,87,232
Phase 0 — feasibility benchmark	3 work-weeks + 1 month GPU	4,47,967
Model work — four workstreams	22	30,90,471
Phase 1 — voice core + booking engine	12	16,25,205
Phase 2 — avatar + polish	8	10,83,470
Phase 3 — second vertical / locality	8	10,83,470
Deployment + handover	4	5,41,735

Before the big spend, Phase 0 measures whether the voice stack actually works in Hindi on one rented 48 GB GPU over one month, against ten recorded pass/fail gates. A measured "no" is a complete answer, and it stops the project at the Phase 0 price point. The monthly charge of ₹1,26,108 begins **only when the system is live and serving real bookings** — not on handover date.

The three decisions the client must make

1. **Proceed as scoped** at ₹78,72,318 excluding GST (₹92,89,335 including GST), or **scope down explicitly** — a ₹20–30 lakh budget covers the model work alone and nothing else.
2. **Pick ONE service vertical and ONE locality for launch.** An unbounded "any service" launch is the measured failure mode of this sector in India.
3. **Decide the Phase 4 option** — the real-time lip-synced avatar at ₹13,54,338 — now or later. It is not a launch dependency.

The single next step

Supply the two blank fields (your registered legal entity name, and who signs and in what capacity) and confirm in writing before 10 November 2026. That is all that stands between this document and a signature-ready contract.

B. HOW TO READ THIS DOCUMENT

This page is for a reader who is not technical. Read it once and the rest of the report will make sense.

The one-time fee versus the monthly charge, and when each starts. The **one-time fee** is what it costs to build and hand over the system — the software, the three trained models, the brand voice recording, the deployment and training. It is invoiced across six milestones, each when its deliverable is accepted, plus a 10% mobilisation advance recovered later. The **monthly charge** is what it costs to *run* the finished system: the GPU it runs on, the database and file stores, the real-time media layer, monitoring, and a support retainer for ongoing engineering. **The monthly charge starts only when the system is live and serving real bookings**, not on the day it is handed over.

What GST is, and that it is charged on top. GST is India's Goods and Services Tax — a government tax added to the price of a supply. Vector Logic's development work is a service, and GST at **18%** is charged **on top of** the one-time fee on each invoice. $\text{₹}78,72,318 \times 18\% = \text{₹}14,17,017$, so the GST-inclusive one-time total is **₹92,89,335**. Where Papparti is GST-registered and uses the service for business, that GST is generally claimable back as input tax credit.

What "excluding GST" means for the number you saw. When the headline number reads **₹78,72,318 excluding GST**, it means that number is what Vector Logic charges **before** the 18% government tax is added. The amount actually invoiced to a GST-registered client is **₹92,89,335**. The monthly figure of ₹1,26,108 is likewise an excluding-GST figure; GST applies to it as applicable.

Why the monthly figure does not fall when usage is low. The GPU is billed **for the hours it exists, not the hours it is used** — a rented 48 GB card runs at $\$0.59/\text{hour} \times 730 \text{ hours a month} = \text{₹}41,665$, 24×7, busy or idle. The reasoning model also runs continuously by design. If bookings are light, the vendor bill is unchanged. The GPU decision note shows when buying instead would become cheaper (about **21 months** for a used card, **61 months** for a new one) and **recommends renting until Phase 0 measures real utilisation**.

The three labels used throughout these documents.

- **MEASURED** — we ran an instrument and can show you the number. Example: the voice samples' durations and speech energy (dBFS); the avatar's pixel change between frames.

- **ESTIMATED** — a modelled figure with its assumption stated in the same place.
Example: the studio recording cost (₹97,000, the midpoint of a ₹64,000–₹1,30,000 band); the monthly monitoring and storage lines.
- **UNVERIFIED** — we could not confirm it, and it must be confirmed before you rely on it. Example: the SAC tax codes; whether the voice stack produces good Hindi at all — which is exactly what Phase 0 exists to measure.

Why you will see a placeholder for your own company details. The address in these documents reads "Papparti", which is a **brand**, not a registered legal entity. An invoice needs the registered entity's name, and a contract needs an authorised signatory. Neither has been supplied, so rather than guess, the documents carry deliberate blank fields and the placeholder email **contact@papparti.com only** in the demo references. Tell us the real details and we correct them everywhere.

C. GLOSSARY

Every term below appears in the five documents in this package. Definitions are plain English; the "where it matters" note says which document and section the term decides something in.

TERM	PLAIN-ENGLISH MEANING	WHERE IT MATTERS
Adapter	A small trained add-on that sits on top of a frozen base model and changes its behaviour for your data, without retraining the whole model.	Proposal §1; the three models are adapters, not train-from-scratch models.
ASR (Automatic Speech Recognition)	The machine task of turning spoken audio into text. Same job as STT; the two names are interchangeable.	Estimate header; the speech-recognition model.
Change order	Any change to the agreed scope, raised in writing and priced before work begins.	Commercial Terms §5; changes over ₹13,542 are quoted.
Concurrency	How many conversations the system can serve at the same time.	Estimate §5; a hidden driver of whether one GPU card is enough.
Diarisation	Telling apart <i>who</i> is speaking when more than one person talks — e.g. customer versus provider on a call.	Proposal §3.1 / §3.3; relevant once real booking audio is used.

TERM	PLAIN-ENGLISH MEANING	WHERE IT MATTERS
Fine-tune	Training an existing model further on your own data so it performs better on your domain, rather than building a model from nothing.	Proposal §1; the documented meaning of "retrain" in this project.
GST (Goods and Services Tax)	India's national tax on supplies of goods and services; here 18% on the development service.	Quotation §A; adds ₹14,17,017 to the one-time fee.
GSTIN	A business's GST registration number, read from its certificate.	Commercial Terms §4.2; a blank field until supplied.
Go/no-go gate	A pre-agreed pass/fail threshold that decides whether the project continues, changes, or stops.	Proposal §4; the ten gates Phase 0 reports against.
GPU (Graphics Processing Unit)	The specialised processor on which AI models are trained and run.	Quotation §C; the single largest monthly line at ₹41,665.
Hinglish	Hindi and English mixed within the same sentence — the everyday speech pattern of the target user.	Proposal §8.1; the hardest and least-solved part of the voice stack.
HSN / SAC	Standard codes used on an Indian tax invoice to classify what was supplied.	Quotation §A; UNVERIFIED — to be confirmed with a tax adviser.
Input credit (ITC)	GST you paid on your inputs, which a registered business can generally claim back against the GST it charges onward.	Estimate §7c; applies to Vector Logic's GST and to reverse-charge IGST.
Latency	The delay between a user doing something and the system responding.	Proposal §5; Phase 2 gate is the avatar adding under ~800 ms.
LoRA / QLoRA	The standard, efficient methods for fine-tuning a large model on modest hardware — training a small adapter instead of the whole model.	Proposal §1; the industry-standard method this project uses.
Mobilisation advance	A payment drawn at the start of the project against front-loaded costs (GPU, studio booking), recovered later.	Commercial Terms §2.1–2.4; ₹7,87,232, recovered at the Model work milestone.
MOS (Mean Opinion Score)	The average score human listeners give to speech quality — the only real test of whether a voice sounds good.	Proposal §8.1; no MOS exists yet ; the listening test runs in Phase 1.

TERM	PLAIN-ENGLISH MEANING	WHERE IT MATTERS
Parameter class	The rough size of a model, measured in billions of parameters (B). Bigger is generally heavier to run.	Estimate §4; recognition 1–2 B, synthesis 0.3–0.5 B, reasoning ~32 B.
p95	The value below which 95% of measurements fall — a measure of the slow cases, not the average.	Proposal §4; the p95 time-to-first-audio gate is 1,500 ms.
Place of supply	The rule that decides which state's GST (CGST/SGST or IGST) applies to a service supply.	Commercial Terms §4.3; recorded on each invoice.
Retention / holdback	A slice of the fee the client keeps until the delivered system is accepted.	Commercial Terms §2.3; ₹3,93,616 (5%), released at handover acceptance.
RTF (Real-Time Factor)	How long the machine took to process audio relative to the audio's own length — under 1.0 means faster than live.	Proposal §8.1; a measured objective figure, not a quality score.
Self-hosting	Running the models on infrastructure you control rather than through a third-party cloud API.	Proposal §1; a control-and-compliance decision, explicitly not a cost saving .
Sessions per node	How many simultaneous conversations one GPU machine can hold before quality or speed degrades.	Proposal §4; gate is ≥ 10 ; below that the cost model breaks.
SFU / TURN	The infrastructure that carries live audio between phones in real time across real networks.	Quotation §C; the ₹4,837/month real-time media line.
SLA (Service Level)	An agreed expectation of how reliable a service will be.	Commercial Terms §9; liability is bounded, and no accuracy figure is warranted pre-measurement.
STT (Speech-to-Text)	Same as ASR — turning speech into text.	Estimate §4; the speech-recognition workstream.
STT / ASR	The two names for the same recognition task; used interchangeably in these documents.	Estimate header and §4.
TDS (Tax Deducted at Source)	Tax the payer withholds from a payment and remits on the supplier's behalf.	Commercial Terms §4.5; section and rate must be confirmed with a tax adviser.
TTS (Text-to-Speech)	The machine task of turning written text into spoken audio — the brand voice.	Estimate §4; the synthesis workstream.

TERM	PLAIN-ENGLISH MEANING	WHERE IT MATTERS
Vector store	A database that stores text meanings as numbers so the system can find relevant information quickly.	Quotation §C; the ₹5,891/month managed database line includes it.
Voice cloning	Building a synthetic voice that matches a specific recorded speaker. Requires that speaker's signed release.	Proposal §3.4; no release = no clone.
VRAM	The memory on the GPU. Model size is limited by how much VRAM the card has (here 48 GB).	GPU Decision §2b; 48 GB is the stated safe floor for the project.
Warm handoff	Passing a conversation smoothly from the automated system to a human, without dropping it.	Proposal §5 / Commercial Terms §3; the fallback when voice fails.
WER (Word Error Rate)	The percentage of words in a transcript that came out wrong; lower is better.	Proposal §4; the recognition gates are 20% Hinglish and 12% Hindi.
WER gate	The specific WER threshold that decides whether the voice-first design proceeds or becomes text-first.	Proposal §4; failure means stop/redesign, before big spend.
Warranty	The 30-day window after handover in which defects are fixed at no charge.	Quotation §D; starts on signature of the handover acceptance certificate.

D. THE DECISIONS THIS DOCUMENT ASKS FOR

Please answer each of these as a clear yes or no, in writing. They are the only decisions that block a signature.

1. **Do you accept the one-time fee of ₹78,72,318 excluding GST (₹92,89,335 including GST) for the scope as written?** If your budget is capped lower, do you want a written scope-down of the same system, and which part do you protect?
2. **Do you approve Phase 0 first**, at ₹4,47,967, as a measurement you buy before any large development spend — accepting that a measured "no" stops the project at that price point?
3. **Which ONE service vertical and which ONE locality** will launch first?
4. **Do you want the Phase 4 real-time lip-synced avatar** at ₹13,54,338 as a priced option now, or do you defer it entirely?

5. **Do you accept the mobilisation advance of 10% (₹7,87,232)** and the retention of 5% (₹3,93,616) released at handover acceptance?
6. **Will you supply the model-the-voice decision and the signed voice-rights release** in Papparti's favour — perpetual, irrevocable, worldwide, transferable, commercial — before the voice is cloned?
7. **Do you accept that recognition and synthesis run on self-hosted, rented GPU** as a control-and-compliance choice, knowing it is **not** the cheaper option at plausible volumes?
8. **Will you nominate the legal entity, the signatory and designation, the contact email, the GSTIN and the tax adviser** so the contract, the SAC codes and the invoicing can be completed?