

Meet Prajapati

Gujarat, India • +91-93167-61615 • meetprajapatid0207@gmail.com

GitHub • LinkedIn • Portfolio

AI Engineer — Voice agents & agent infrastructure

Experience

Software Developer

Jan 2026 – Present

Vaiu AI • Bengaluru, Karnataka (Remote)

- Built a RAG pipeline on PostgreSQL + pgvector (HNSW, 768-dim) over **6,637 chunks from 40 documents** (Docling/crawl4ai ingestion, BM25 fallback), with OpenAI, Vertex AI, and local ONNX embeddings unified via Alembic; on 200 real queries: **117.7 ms p50 uncached (145.5 ms p95), 1.2 ms cached**.
- Owned Chat-Builder's workflow layer: Pydantic schemas and a **12-rule** validator with auto-repair that turn LLM-generated pathway JSON into executable React Flow graphs, plus the flow canvas, Debug Chat Drawer, and KB search; wrote LangGraph supervisor routing and fixed streaming race conditions that truncated sentences (platform: 41 organizations, 197 pathways).
- Integrated Langfuse tracing through an OpenTelemetry (OTLP) background exporter with queued flush, keeping trace emission off the live conversation path; built a LiveKit TTS caching layer for a Hindi-English voice agent, cutting Google Cloud TTS cost **95%** on bilingual phrases.
- Benchmarked **7 open-source LLMs** (Ollama) on 26 production-prompt checks, selecting qwen3.5-9b-fast (**22/26**); showed n=26 separates only the two tool-incapable Gemma models and wrote an eval v2 spec (Wilson CIs, McNemar tests, pass^k).
- Shipped the deployment layer: JS/React SDK, embed-widget channels (platform: **20 live deployments**), metrics dashboards, and GuardrailsProcessor, a Pipecat frame processor that replaces rule-violating text without dropping the WebRTC stream; GCP deploy scripts and Docker Compose rollback in **under 2 min**.
- Built a 4-stage LLM blog pipeline (Gemini topic discovery, research, generation, image metaphors; Next.js/TypeScript, GCP Cloud Scheduler): **138 published articles**.

Freelance Software Developer

July 2025 – Oct 2025

Bee Pee Coatings Pvt. Ltd. (Berger Paints Subsidiary) • Anand, Gujarat

- Architected an inbound yard-management system (MERN, React 19): a pool of **100 tokens** (up to 100 vehicles in the yard) is issued at entry, stamped at each checkpoint, and recycled at the terminal one, keeping the active MongoDB set at 100 documents with journeys archived to an append-only history.
- Built a no-code checkpoint engine (**6 field types**, 3 trigger modes) with guards rejecting duplicate and out-of-sequence submissions, plus ExcelJS audit reports across **6 preset time windows**.

UI/UX Design Intern, Smallfare • Hyderabad (Remote)

Feb 2025 – Jun 2025

- Designed **35 screens** (events, booking, full user workflow) for the Smallfare app and led **4 design interns**.

Projects

Pipecat Session Continuity

Jul 2026

GitHub • PyPI • Python, FastRTC, WebSockets, Redis, SQLite

- Built idempotent tool-call recovery: duplicate calls after reconnect fell from **80% (16/20) to 0% (0/20)** on an adversarial 20-trial test (gpt-oss-120b via Groq).
- Published a PyPI library (**403 downloads**, mirrors excluded) adding Redis checkpointing and HMAC-signed tokens so dropped Pipecat calls resume; resume adds **2.83 ms** on average (4.57 ms p95) over 1,000 runs.

Sentinel API Checker

Sep 2026

GitHub • GitHub App • TypeScript, Fastify, ts-morph, Groq

- Engineered a ts-morph AST scanner that flags affected consumer files with Groq mitigation summaries: **1,000 TypeScript files in 998 ms, 2,000 in 1.55 s** (55 MB heap, local benchmark).
- Built a GitHub App (Fastify webhook, HMAC-SHA256, Check Runs) and CLI diffing OpenAPI 3.0/3.1 schemas across repos to block breaking merges; **37 merged PRs, 53 unit tests**.

Technical Skills

Languages: Python, TypeScript, JavaScript, SQL

AI / LLM: RAG, pgvector, LangGraph, Pydantic, Langfuse, Ollama, Groq, Gemini, OpenAI / Vertex AI embeddings, ONNX Runtime, Docling

Voice: Pipecat, LiveKit, WebRTC, WebSocket, Google Cloud TTS

Backend & Frontend: FastAPI, Node.js/Express, Fastify, SQLAlchemy, Alembic, OpenAPI, Webhooks, GitHub Apps, React, Next.js

Data & Infra: PostgreSQL, MongoDB, Redis, SQLite, Docker, Git, GCP (VM deployment, Cloud Scheduler)

Education

G H Patel College of Engineering & Technology

2022 – May 2026

B.Tech. in Computer Science and Design • CGPA: **8.6/10** • Provisional degree awarded