

BHUMIL MODI

Forward Deployed AI Engineer · Full-Stack

Ankleshwar, Gujarat, India | +91 81410 52176 | bhumilmodi2002@gmail.com
linkedin.com/in/bhumil-modi-430148190 | github.com/BhumilModi | bhumilmodi.dev

SUMMARY

Full-stack engineer who builds AI systems embedded with the customer — working POC two weeks after kickoff, live beta inside five months. Three years of engineering experience, the most recent at an applied AI research company across 20+ client products in ten domains: on weekly calls with client stakeholders, then writing the agent pipelines, Python services and React frontends myself. Comfortable owning ambiguity, output quality, and the client relationship at once.

EXPERIENCE

TheAgentic — Forward Deployed AI Engineer

May 2025 – Present

Applied AI research company building and launching vertical AI companies · theagentic.ai

- Shipped 20+ client products, 10+ of them from an empty repository to live beta — working POC within 2 weeks of kickoff, functional application by month 2–2.5, beta with client-selected users by month 4–5.
- Embedded on weekly stakeholder calls alongside the product manager, turning ambiguous requirements into architecture and working software; led teams of up to four in 2–5 person client pods, project lead on six products.
- Delivered across ten domains — clinical communications, architecture and interior design, insurance renewals, conversational agents, generative media, CAD and manufacturing, sales outreach, grant discovery, equity research and biopharma.
- Built a scan-to-drawing pipeline turning room photographs into dimensioned plan, elevation and axonometric sheets — monocular metric depth, multi-frame registration, and linework marking measured boundaries solid and assumed ones dashed — then retired that deterministic geometry stack for image-model rendering with a content-addressed cache once it beat the hand-built one.
- Designed a fact → plan → write → verify LLM orchestrator with focus gating and cross-step deduplication for a US clinical platform — engineered to need no manual review, with human-in-the-loop only where the workflow calls for it by design.
- Hardened that platform for production: process-wide provider concurrency gate with rate-limit backoff, fail-closed audio assembly so partial renders never ship, per-job caps and stage-scoped render slots, a full error taxonomy with retryability, and a middleware auth-bypass patch.
- Shipped v1.0 of an embeddable agent SDK to public release — framework-free telemetry core, React-free WebSocket transport, consumer-resolvable type declarations, bundle budgets, a multi-channel release pipeline, and supply-chain hardening with secret scanning.
- Built the frontend foundation for an insurance renewals product: design system on brand tokens, signed-in shell, book-of-business and account surfaces, an analysis-run state machine, and a voice interview driven by a checkable reducer.

TheAgentic — Senior Software Engineer

Nov 2024 – May 2025

- Built the entire UI for TheAgentic Console — an LLM developer console with model management, prompt playground, observability dashboards and request/session tracing.
- Delivered React frontends and API integrations across seven concurrent client engagements, plus the architecture diagrams and demo documentation used in client presentations; contributed to the frontend of CortexON (450+ GitHub stars) and TheAgenticBench.

ScaleGenAI — Senior Software Engineer

Feb 2024 – Oct 2024

- Sole owner of the frontend architecture for the flagship product — React, TypeScript, Redux Toolkit, React Router and ShadCN UI — including SuperTokens authentication.

Tally Group — Software Engineer

Oct 2023 – Jan 2024

- Built a billing-simulation web application for Australian utility clients (React + TypeScript frontend, .NET backend), delivering reusable UI components and API integrations across agile sprints.

SleevesUp — Software Engineer Intern

Mar 2023 – Aug 2023

- Full-stack intern on two products, using Spring Boot (Kotlin) services with a React + TypeScript frontend.

TECHNICAL SKILLS

AI / LLM: Agentic workflow design, LLM pipeline orchestration, provider abstraction, prompt engineering, output verification and retry-with-feedback loops, image-model rendering, ElevenLabs TTS

Vision & Geometry: Monocular metric depth, point-cloud registration, segmentation masks, scan-to-CAD drawing generation

Languages: Python, TypeScript, JavaScript, SQL

Backend: FastAPI, Node.js, Spring Boot (Kotlin), .NET

Frontend: React (Vite, Next.js), Redux Toolkit, ShadCN UI, Tailwind CSS, WCAG AA

Platform: PostgreSQL, Docker, AWS S3, Auth0, SuperTokens, npm SDK release engineering, supply-chain hardening

EDUCATION

Vellore Institute of Technology

2019 – 2023

B.Tech, Computer Science and Engineering — specialisation in Artificial Intelligence and Machine Learning | GPA 9.08 / 10