

Farzam Azhar

Senior Full-Stack & AI Engineer: LLMs, Next.js, React, Rails, Python

Lahore, Pakistan (UTC+5), remote · farzam.azhar7777@gmail.com · farzamazhar.com

upwork.com/freelancers/farzamazhar3 · linkedin.com/in/farzam-azhar-25134b132/ · github.com/farzam-dev7777

SUMMARY

Senior full-stack and AI engineer with 9+ years building production software for clients in the US, UK, EU and Australia. Puts LLMs into real products: an AI receptionist with a Claude call grader measured against human reviewers, a multi-provider AI gateway with evals in CI and budget-aware routing, hybrid RAG on pgvector, and real-time LiveKit voice agents. Full-stack across Next.js, React and TypeScript, Ruby on Rails (3 to 8.1), Node.js and Python. 45+ completed Upwork contracts, 5.0 rating on long-term engagements, 8,000+ hours billed. Available 40 hours/week, full time or contract.

EXPERIENCE

Senior Full-Stack & AI Engineer (independent), Upwork and direct clients

2020 – present

Remote

- SurgePoint: lead engineer (879 of 900 commits) on an AI receptionist SaaS. Claude call grader with versioned rubrics and human-agreement tracking, Retell voice agent booking into EMRs, Rails 8 on Kamal.
- pandOS: built the AI platform layer. Multi-provider gateway (Anthropic, OpenAI, Ollama) with circuit breakers and per-call cost logging, versioned prompts, LLM-judge evals gating CI, budget-aware model routing.
- Mentara: hybrid RAG (pgvector + full-text, reciprocal rank fusion, HyDE, multilingual Cohere embeddings) and a Python LiveKit voice agent (Deepgram, Cartesia, barge-in).
- Next.js / React: SEAL cyber-risk portal (Next.js 16, React 19, sole front-end developer), a Next.js 15 + Supabase LLM screening app with a measured model choice and 72 test files, and the PadStats map front end (1,012 commits).
- Document AI and Python: WinAD contract ingestion with vision-based redaction detection via Portkey and an eval harness; a Django + Celery transcription and insights service.
- 45+ completed Upwork contracts, rated 5.0 on long engagements of 2,449, 1,934, 1,748 and 1,168 hours with single clients.

Senior Rails Developer (contract), Curve Tomorrow, Australia

Feb 2021 – Jul 2023

Remote

- Worked across the agency's health portfolio: WeGuide, The Fast 800, Young Stroke Service and Child and Family eHub.
- WeGuide: multi-tenant Rails 6.1 (ros-apartment) with tenant-aware Sidekiq jobs, two-factor auth and Firebase push.
- The Fast 800: Braintree billing on a Rails API, plus Ionic + Capacitor apps for iOS, Android and web.
- Rails 6 → 7 upgrades, with CircleCI and Code Climate gates on every app.

Ruby on Rails Developer, Square63, Lahore

2018 – 2020

Lahore, Pakistan

- Built Rails 5 and 6 applications for international clients at a Lahore software house.
- Brentmark client portal: configurable product content types with drag-and-drop ordering, payment preferences and policy acceptance (351 commits).
- Concierge marketplace API: orders and checkout, Stripe auto top-ups on Sidekiq, OTP login, OAuth with Doorkeeper, rswag API docs.
- Two-factor authentication over SMS and email with backup codes (devise-two-factor, Twilio) for FundingPro.

SELECTED PROJECTS

SurgePoint AI Front Desk thesurgepoint.com

Lead engineer · Voice AI · Reviews automation

An AI receptionist that answers, books and follows up for med-spas and clinics, on a multi-tenant Rails 8 SaaS I rebuilt from Next.js and now own (879 of 900 commits). LLM call grader: Claude scores every call through a forced-tool schema against versioned rubrics, with human-reviewer agreement tracked per rubric version.

Stack: Rails 8, Anthropic Claude, Retell AI, Sidekiq, PostgreSQL, Kamal, Stripe

pandOS AI platform pandos.nl

Built the AI platform layer · Proptech · Netherlands

The AI layer behind an operating system for Dutch real-estate investors: one gateway every model call goes through, with evals, budgets and fallbacks built in. Gateway across Anthropic, OpenAI and local Ollama with idempotency, per-provider circuit breakers, retries and a cost-in-EUR log for every call.

Stack: Rails 8.1, Anthropic, OpenAI, Ollama, pgvector, GitHub Actions, Solid Queue

Mentara (AI Guru) mentara.nl

Built it end to end · AI coaching · RAG + voice

An AI coach that answers from each coach's own material, by chat or by voice. Hybrid retrieval: pgvector semantic search and Postgres full-text fused with reciprocal rank fusion, plus HyDE query expansion and conversational query rewriting.

Stack: Rails 8, Python, LiveKit, pgvector, OpenAI, Cohere, Deepgram, Cartesia

Rental inquiry screening

Sole developer · Proptech · Frankfurt letting agency

Next.js app that reads every rental inquiry a Frankfurt letting agent receives, runs it through screening gates and writes the result back to their CRM. Model chosen by measurement: four models run on the client's 16-case acceptance set, gpt-4.1-mini picked at 14/16.

Stack: Next.js 15, React 19, TypeScript, Supabase, OpenAI, zod, Vercel

PadStats padstats.com

Full-stack engineer · Proptech · Underwriting data

Real-estate intelligence for underwriters: value, rent, crime and hazard exposure on an interactive map. Next.js 14 map: Mapbox GL with viewport-driven H3 hazard layers (flood, fire, storm surge, wind, earthquake).

Stack: Next.js 14, TypeScript, Apollo GraphQL, Mapbox GL, Rails 7, Node.js, Cloud Run

WinAD contract intelligence winad-hq.com

Owned AI data ingestion · Sports · Document AI

Athletic-department intelligence built from coach, vendor and game contracts. Vision model (o4-mini via Portkey) detects redactions and handwritten corrections before extraction.

Stack: Rails 8, Portkey, OpenAI, Unstract, Sidekiq, PostgreSQL

SEAL cyber-risk portal seal.grcx.io

Sole front-end developer · Cybersecurity · Law firms

Cyber-risk assessments for law firms, with separate areas for banks, clients and auditors. Five role-based areas (admin, bank, firm, client, auditor) with NextAuth JWT refresh and TanStack Query.

Stack: Next.js 16, React 19, TypeScript, TanStack Query, NextAuth, Rails 7, OpenAI

Gameplans www.gameplans.com.au

Core developer (1,004 commits) · Proptech · Australia

A property-investment planning platform for Australians: plans, Compass charts, scoreboards and advisor access. OpenAI Assistants with file search and vector stores, with query logging and rate limits.

Stack: Rails 7, OpenAI Assistants, PostgreSQL, Sidekiq, Stripe, Chart.js

MORE WORK

Healthy Brain Hub (Rails 7.1) · **Acton Academy platform** (Rails 8.1 · Claude) · **WellFunded** (Rails 7.1 · Claude) · **Shorts Creator** (Python) · **Claude Code plugin** (Claude Code · Python) · **The Fast 800** (Rails 6.1 · Ionic)

SKILLS

AI engineering: Anthropic Claude (tool use, prompt caching), OpenAI, RAG, pgvector, Hybrid search (RRF, HyDE), Embeddings (OpenAI, Cohere), Evals & LLM-as-judge, Portkey, LangChain, Pinecone, Ollama, Vercel AI SDK, Claude Code

Voice & media AI: LiveKit Agents, Deepgram, Cartesia, Retell AI, Whisper / WhisperX, Vision models

Frontend: Next.js (14–16, App Router), React 19, TypeScript, TanStack Query, Redux Toolkit, Zustand, Apollo GraphQL, Tailwind CSS, shadcn/ui, Mapbox GL, Vue 3, React Native (Expo)

Backend: Ruby on Rails (3 → 8.1), Hotwire, Sidekiq, Node.js, Express, GraphQL, Python, FastAPI, Django, Celery

Data: PostgreSQL, pgvector, PostGIS, H3, Supabase, MongoDB, MySQL, Redis, BigQuery, OpenSearch

Cloud & DevOps: AWS, GCP Cloud Run, Vercel, Cloudflare Workers, Heroku, Docker, Kamal, GitHub Actions, CircleCI, Terraform, Sentry

Testing: RSpec, Capybara, Jest, Vitest, Cypress, Playwright, pytest

Integrations: Stripe, Braintree, Paddle, Twilio, Vonage, SendGrid, Plaid, GoHighLevel, Google APIs

EDUCATION & LANGUAGES

BS Computer Science, FAST-NUCES (National University of Computer and Emerging Sciences), Lahore 2014 – 2018

Certified Ruby on Rails Developer, Upwork

Languages: English (professional), Urdu (native)