Muhammad Bahjat

Lahore, Pakistan | +92 335 4615591 | syedmbahjat848@gmail.com
https://muhammadbahjat.com | https://www.linkedin.com/in/muhammadbahjat | https://github.com/SyedBahjat

PROFESSIONAL SUMMARY

Senior AI Engineer with 2+ years of experience building production multi-agent systems, voice AI platforms, and full-stack applications across PropTech and HRTech for US-based companies. Proven track record of cutting LLM inference costs by ~90% through model selection and prompt re-engineering, and shipping multi-platform products to the App Store, Google Play, and Chrome Web Store. Expert in LangChain, LangGraph, and Mastra agent orchestration, with deep experience deploying scalable services on GCP and AWS using Kubernetes and Docker.

CORE COMPETENCIES

Technical Skills: Multi-Agent Orchestration, Retrieval-Augmented Generation (RAG), Prompt Engineering, LLM Fine-Tuning & Evaluation, Voice AI Systems, Semantic Search & Vector Databases, Full-Stack Web & Mobile Development, REST API Design, Microservices Architecture, Cloud-Native Deployment, CI/CD, Containerization, Web Scraping & Browser Automation

Soft Skills: End-to-End Product Ownership, Cross-Functional Collaboration, Vendor Negotiation, Engineer Mentorship, Independent Decision-Making, Translating Technical Work into Business Outcomes

Tools & Technologies: Python, JavaScript/TypeScript, Next.js, React, React Native (Expo), Node.js, FastAPI, Flask, Django REST, LangChain, LangGraph, Mastra, Agno, CrewAI, OpenAI, Gemini, Claude, Whisper, ElevenLabs, Bland AI, Retell AI, Vapi, Twilio, n8n, PostgreSQL, Supabase, Prisma, MongoDB, pgvector, Pinecone, FAISS, Redis, Celery, Docker, Kubernetes, GCP (GKE, Cloud Run, Vertex AI, Cloud Scheduler), AWS (EC2, S3), Vercel, GitHub Actions, RevenueCat, Playwright, Selenium, Apify

PROFESSIONAL EXPERIENCE

Synapse (AI-powered recruitment platform) | California, USA (Remote)
AI Engineer | Sep 2025 – Present

- **Rebuilt the candidate FitScore evaluation system,** reducing per-candidate evaluation time from ~45 seconds to ~10 seconds while maintaining 80–90% accuracy by switching LLMs by task and re-engineering the prompt stack.
- **Built and shipped a recruiter-facing AI mobile app to both the App Store and Google Play.** Recruiters describe a target role in chat; the agent searches candidates via Apify, evaluates fit with Gemini (with Perplexity and other LLM fallbacks), and returns ranked results in-app.
- **Integrated Gmail and LinkedIn account connections directly into the mobile app,** letting recruiters send personalized outreach emails and connection requests to surfaced candidates without leaving the app.
- **Implemented full RevenueCat monetization across iOS and Android,** including subscription trials, multiple payment tiers, and consumable credit top-ups. Owned the entire submission and review process for both stores.
- **Built the SynapseHire web platform's candidate-search backend with pgvector** semantic search runs against the local Postgres candidate base first, falls back to Apify/LinkedIn scraping only on miss, and writes new candidates back to the database, cutting repeat-search cost to near zero on overlapping queries.
- **Designed n8n workflows automating the full post-screening pipeline:** candidate approval, client submission, and interview-scheduling emails eliminating manual handoffs between recruiters and clients.
- **Owned full-stack delivery on Next.js + Neon Postgres + Prisma,** shipping frontend, REST APIs, and infrastructure integration for the recruiter web platform.

HomeEasy (US real estate technology) | Chicago, IL, USA (Remote)
Senior AI Engineer | May 2024 – Apr 2026

- **Cut monthly LLM spend by ~90%** by re-architecting the agent stack combined model selection with prompt redesign to preserve quality at lower context cost.
- **Built a voice AI system on Bland AI + Retell** that places 1,000+ outbound calls per day to property managers, qualifies commission availability, and books tours, while handling 1,500+ active client conversations concurrently during business hours.
- **Designed and shipped a custom CRM (Next.js + PostgreSQL)** replacing FollowUpBoss; integrated Twilio voice and SMS so 10+ sales agents share a single business number with full call/text history in-app.

- **Implemented semantic property matching with pgvector** across the inventory database, letting the system reuse already-qualified buildings for new clients with matching neighborhood and budget eliminating redundant scraping and API calls.
- **Owned the full deployment pipeline on GCP** (GKE, Cloud Run, Cloud Storage, VM workloads) with Docker, Celery and Celery Beat for scheduled agents, and Redis for queue/state. Production workloads scaled to peak load with zero downtime.
- **Solved headless-scraping blocks on apartment-listing sites** (Zillow, Apartments.com) by deploying Selenium and Playwright agents inside Docker VNC containers with full GUI simulation.
- **Drove prompt engineering, vendor negotiations with GCP, and supported onboarding for new AI engineers** owned the cost-reduction and reliability roadmap end-to-end.

Groomify | Lahore, Pakistan
Software Engineer | Jan 2024 – May 2024

- **Developed REST APIs for a behavioral incentive platform;** designed and optimized database models, reducing query response times by ~25%.
- **Integrated third-party APIs and authentication mechanisms** for secure user access; collaborated with frontend and UI/UX teams on delivery.

ChatDroid | Lahore, Pakistan (Remote)
Software Engineer | Jun 2023 – Dec 2023

- **Built conversational AI sales and support bots** with LangChain and OpenAI, applying FAISS for retrieval and Whisper for voice input.
- **Improved response accuracy and reduced manual support load** via personalized chatbot dashboards; contributed JavaScript and jQuery frontend work.

KEY PROJECTS

WhatsFlow — WhatsApp Automation & AI Integration Platform | https://whatsappflow.duckdns.org/
Node.js, Express.js, Supabase/PostgreSQL, Socket.IO, MCP, OAuth 2.0, Paddle, Google Cloud

- Built a multi-tenant WhatsApp platform with QR-based account linking, multiple sessions, conversation history, and support for text, images, documents, video, and voice notes.
- Developed an **MCP connector for Claude and ChatGPT**, enabling assistants to search contacts, read conversations, and send messages through OAuth 2.0 with PKCE, session-scoped permissions, and revocable access.
- Exposed REST APIs and configurable webhooks for messaging automation; implemented tenant ownership checks, input validation, and message-send rate limiting.
- Integrated Paddle subscription billing, self-service plan changes, and admin-managed free access, enforcing subscription eligibility across messaging, webhooks, and MCP; deployed on Google Cloud with Caddy and PM2.

SparAI AI Mock Interview Platform (Founder & Engineer) | https://sparai.vercel.app

Mastra, Gemini, Next.js, Vercel

- Built an AI-powered mock interview platform where candidates practice technical and behavioral interviews against an AI interviewer that adapts to any job description. Recruiters can also schedule structured interviews with configurable difficulty.
- Built the agent layer with Mastra (Next.js-native agent framework) so the interviewer runs in the same runtime as the web app no separate Python backend, no cross-service latency.

ApplyFlow Job Application Tracker (Founder & Engineer) | https://applyflow-mauve.vercel.app

Next.js, Chrome Extension API, PostgreSQL

- Built a job application tracker where the Chrome extension captures a job's title, company, and link from the active tab; the saved role appears in the user's ApplyFlow dashboard automatically no manual entry.
- Shipped to the Chrome Web Store, App Store, and Google Play after navigating each platform's full review process.

Crypto Trading Bot (Binance)
Python, TFT, Reinforcement Learning (PPO), Fusion MLP, Binance API

- Built a multi-model trading bot (TFT, RL/PPO, Fusion MLP) trained on Kaggle and unified to live Binance candle data; supports automated entries with take-profit and stop-loss management.

Smart Waiter Robot (Final Year Project)

Raspberry Pi, LiDAR, SLAM, Python

- Designed an autonomous restaurant robot using LiDAR and SLAM for navigation. Customers placed orders via a custom web app; the robot retrieved orders from the kitchen and delivered them to mapped tables.

EDUCATION

COMSATS University Islamabad | Lahore, Pakistan
Bachelor of Science in Computer Engineering | Aug 2020 – May 2024

- Activities: Team Lead, AI Engineering · Executive Member, IEEE Robotics & Automation Society (Apr 2023 – Jan 2024)