Bihan Banerjee

banerjeebihan456@gmail.com | +91 62969 53887 | bihanbanerjee.com | linkedin | github

SUMMARY

Full-Stack AI Engineer who builds agentic systems from 0 to production across Python and TypeScript. Strong across the full stack — orchestration, memory, real-time backends, and frontend — with a focus on complete, deployable systems over prototypes.

PROOF OF WORK

Buildable — AI Prompt-to-App Builder

Python | FastAPI | LangGraph | Next.js | E2B | PostgreSQL | Docker | Terraform

- Architected a 2-agent LangGraph pipeline that transforms plain-English prompts into deployment-ready React apps inside isolated E2B sandboxes, cutting time-to-working-prototype from hours to **under 2 minutes**.
- Engineered an autonomous build-validation loop that detects and patches compile failures across up to **3 repair attempts**, eliminating manual debug cycles on generated apps.
- Cut E2B sandbox cold-start latency **83%** (30 s → 5 s) by designing a custom template with pre-installed dependency snapshots and a locked base image.

Holdmind — Memory-Augmented Conversational AI

Python | FastAPI | Next.js | PostgreSQL | Qdrant | SQLite | D3.js | OpenRouter

- Built a full-stack conversational AI that constructs a per-user belief graph from conversation history, personalizing every response with user-specific semantic context from turn one.
- Engineered a real-time SSE inference pipeline that retrieves top-k relevant beliefs from Qdrant before each generation step and persists newly extracted claims afterward, maintaining accurate belief state across multi-session conversations.
- Shipped a production-grade security surface: AES-encrypted API key vault, rate-limited endpoints, refresh-token auth, and an interactive D3.js belief graph for live memory introspection.

re-collect — Belief-Centric Memory Layer for AI Agents

Python | LangChain | LangGraph | FAISS | Qdrant | Pinecone | SQLite | pytest

- Designed and shipped an open-source belief-centric memory library to PyPI supporting typed beliefs, confidence scores, and semantic graph memory across **4 relationship types** (supports, contradicts, derives, similar) — giving AI agents auditable, structured long-term memory.
- Engineered a pluggable LLM memory updater with **3 vector backends** (FAISS, Qdrant, Pinecone), keeping SQLite as the authoritative source-of-truth and decoupling fast retrieval from durable storage.
- Implemented cycle-safe explanation chain traversal enabling complete belief provenance — any stored claim fully traceable to its source evidence in O(depth) time.

VeloxTrading — Real-Time Crypto Trading Platform

TypeScript | Next.js | WebSocket | Redis | PostgreSQL | TimescaleDB | Docker | Turborepo

- Architected a **7-service microservices** platform with live Binance market feeds, supporting leveraged positions, real-time liquidation, and crash recovery in a single cohesive system.
- Built a high-throughput in-memory liquidation engine processing **1,000+ price ticks/second per asset** with event-sourced state and snapshot-based recovery, achieving **sub-15 s** full state restoration after failures.
- Eliminated floating-point rounding errors across the entire trading stack by implementing BigInt arithmetic at 10^8 scale, ensuring zero precision drift on all position and PnL calculations.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL, C++
AI / ML: LangGraph, LangChain, LangSmith, Langfuse, Multi-agent Orchestration, RAG, Belief-Graph Memory, Prompt Engineering, PyTorch, TensorFlow, scikit-learn, NumPy
Backend: FastAPI, Node.js, Express, Bun, SQLAlchemy, Prisma, WebSockets, Redis Streams, REST, Microservices
Frontend: React, Next.js, Tailwind CSS, TanStack Query, D3.js
Infrastructure: PostgreSQL, TimescaleDB, SQLite, Qdrant, Redis, FAISS, Pinecone, Docker, Terraform, AWS, CI/CD

PUBLICATIONS

Banerjee, B., et al. Monkeypox detection from skin lesion images using CNN models with Beta function-based normalization. *PLOS ONE*, 18(4), 2023. [DOI]

Banerjee, B., et al. MSENet: Mean and standard deviation based ensemble network for cervical cancer detection. *Engineering Applications of AI*, 123, 2023. [DOI]

EDUCATION

Bachelor of Engineering, Computer Science (Specialization: Artificial Intelligence) 2019 – 2023
University Institute of Technology, Burdwan University CGPA: **9.01 / 10**