MAYUKH BANERJEE

Bengaluru, India | +91 9620868983 | mayukhbanerjee011@gmail.com
linkedin.com/in/mayukh-banerjee | github.com/MayukhBanerjee | mayukh-banerjee.vercel.app

Summary

Cut LLM API costs **90%+** on a live B2B pipeline at 100K-lead scale. Published research reducing AI fabrication from **15.3% to 2.0%** (Cohen's $d = 1.31$). Owns the entire AI layer across five live SaaS products at Easehawk Technologies: architecture, prompt design, and production deployment, end-to-end. Builds agentic systems, RAG pipelines, and LLM orchestration from scratch without relying on pre-built wrappers. Seeking full-time roles in AI/ML Engineering, Applied AI, or SDE at 15–20 LPA, remote or Bengaluru on-site.

Technical Skills

Programming: Python, JavaScript, TypeScript, Java, C++
AI / LLM Engineering: LangChain, LangGraph, Model Context Protocol (MCP), Agentic AI, RAG, Schema-Conditioned Generation, Context Engineering, Prompt Engineering, Prompt Chaining, Prompt Caching, Deterministic Processing, Manual LLM Evaluation, Cron-Based Batch Processing
Model APIs / Platforms: OpenAI API, Anthropic API, AWS Bedrock, xAI Grok, Groq API, Gemini API, Hugging Face
ML Frameworks: PyTorch, Keras, scikit-learn, XGBoost
Backend / Application: FastAPI, Node.js, Express.js, Pydantic, REST APIs, Next.js
Vector Search / Caching: Qdrant, FAISS, Chroma, Pinecone, Redis
Databases / Data Processing: DynamoDB, Oracle SQL, NumPy, Pandas
Cloud / DevOps: AWS Lambda, S3, Cognito, Transcribe, Comprehend, Amplify, CloudWatch, API Gateway, Polly, Docker, CI/CD, GitHub Actions
Tools / Platforms: EnergyPlus, Streamlit, HeyGen, ElevenLabs, Git, GitHub, Google Apps Script

Experience

AI Engineer May 2026 – Present
Easehawk Technologies *Bengaluru, India*

Sole owner of AI architecture, prompt design, and production deployment across five live products: Moxsend (B2B outreach), MoxReply (inbox automation), MoxCRM (lead management), MoxCampus (EdTech), and EaseHawk Avatar Chatbot.

- Built a **LangGraph** email-generation workflow with prompt chaining, prompt caching, and retries; cron-triggered batch jobs ran until all campaign emails were generated. Cohort-based orchestration cut LLM API costs by **90%+** for campaigns of up to **100K leads**.
- Built **LangChain–Qdrant RAG** over user documents and web pages to ground personalised emails; fact-summary preprocessing recorded **0% URL hallucinations** in evaluation and supported Arabic localisation for GCC markets.
- Classified incoming replies by intent, sentiment, lead temperature, and sales signals using LLM analysis and deterministic rules; generated prompt-free, context-aware reply drafts from full email threads. Deterministic lead classification across **6 sectors and 20+ roles** recorded **100% accuracy** in evaluation.
- Developed analytics dashboards and configurable copilots for pipeline logs, performance insights, and workflow tuning; integrated **Redis caching**. Researched AI-driven CRM capabilities for pipeline intelligence, lead forecasting, and deal coaching.
- Built **FastAPI/LangChain** microservices for an AI-powered school platform, generating class insights, recommendations, intervention plans, and parent reports. Integrated a **HeyGen** avatar receptionist with OpenAI voice integration for synchronised navigation.

Technical & Research Team Lead 2024 – 2025
E-Cell VIT *Vellore, India*

- Automated updates across **100+ real-time data sheets** using Python and Google Apps Script for an eight-hour flagship event: **5,000+ applicants**, 200+ participants, Rs. 1L prize pool, zero downtime.
- When the technical team could not deliver on D-Day, independently built and deployed the full event

application in under 24 hours. Ran live for **200+ participants** with zero failures.

Projects

Evoke – AI Personality Preservation for Digital Legacy github.com/MayukhBanerjee/Evoke

Schema-Conditioned RAG, AWS Serverless, LangChain, AWS Bedrock, ElevenLabs

VIT, 2026 | Research Paper

- Reduced measured confident fabrication from **15.3% to 2.0%** (Cohen's $d = 1.31$, $p < .001$) using a schema-conditioned RAG pipeline and an epistemic-humility gate that prevents the model from inventing opinions the person never held. Authenticity score up 1.47 points on a 1–5 Likert scale over an unconditioned baseline.
- Architected a **10-service AWS serverless pipeline**: Lambda, S3, DynamoDB TTL, Cognito, Transcribe, Comprehend, Amplify, CloudWatch, API Gateway, and Polly. Consent and data expiry enforced as architectural invariants; runs at **$0 marginal cost** within AWS Free Tier.

EcoLoop – Agentic HVAC Control for Buildings github.com/MayukhBanerjee/EcoLoop

MCP, LLM Orchestration, EnergyPlus, Groq, Streamlit, Python

VIT, 2026 | Honeywell Campus Hackathon

- Achieved **33.8% lower electricity use** (114.1 vs. 172.5 kWh) and a **10.1-point comfort gain** against a physics-identical uncontrolled EnergyPlus baseline. Every number pulled from simulation artifacts.
- Built an MCP tool-calling architecture with **five Pydantic-validated tools**; queued writes until end-of-timestep and applied a deterministic clamp layer enforcing **seven safety invariants** before any actuator value was changed.

Scheduler AI – Conversational Booking for Local Businesses github.com/MayukhBanerjee/Scheduler-AI

LangGraph, Agentic AI, FastAPI, Next.js, MongoDB

VIT, 2026

- Designed a six-node LangGraph StateGraph that confines the LLM to two boundary roles while search, ranking, and slot verification run as deterministic code; supports an AI receptionist for businesses and a guided booking concierge for customers.

LeBy – Legal Rights in Plain English github.com/MayukhBanerjee/LeBy

RAG, FAISS, Gemini API, LangChain, FastAPI, React

VIT, 2026

- Built a FAISS-backed RAG application answering Indian constitutional and case-law questions in under five seconds, classifying documents across four legal domains with near-zero hallucinations. Full-stack FastAPI and React.

Education

Vellore Institute of Technology Vellore, India

B.Tech, Information Technology | GPA: **8.56 / 10** Aug 2023 – Jun 2027

Maintained while working full-time as AI Engineer at Easehawk Technologies.

Certifications & Courses

Machine Learning A-Z: AI, Python & R | Complete Python Developer

IBM DevOps & Agile Designs | AWS Cloud Architect