

BALASUBRAMANIAN M

Agentic AI Engineer

Bangalore, India | balasnest@gmail.com | balacc.dev | [Linkedin](https://www.linkedin.com/in/balasnest)

PROFESSIONAL SUMMARY

Agentic AI Engineer with 10+ years of software engineering experience building production LLM applications, agentic workflows, RAG systems, and AI evaluation platforms. Experienced in designing reliable agent systems, multi-agent workflows, tool-calling pipelines, and LLM evaluation and observability systems using modern AI frameworks and infrastructure.

TECHNICAL SKILLS

- **AI / Agentic AI:** Python, LLMs, Agentic AI, Agent Orchestration, LangGraph, LangChain, CrewAI, Multi-Agent Systems, Tool Calling, MCP, Structured Outputs, Guardrails, Pydantic, RAG, RAGAS, DSPy, LangSmith, Temporal, LiteLLM, OpenAI, LLM Evaluation, LLM Observability
- **Retrieval / Knowledge:** Embeddings, Semantic Retrieval, Chunking Strategies, Vector Databases, pgvector, RAG Pipelines, Retrieval Optimization
- **Backend / Cloud:** BFF, RestAPI, FastAPI, Docker, AWS, Redis, Vercel
- **Mobile & Web:** React Native (TypeScript, JavaScript, Expo), React, iOS (Swift, Objective-C)
- **Engineering:** Technical Leadership, GitHub Actions, CI/CD, Agile, Scrum

PROFESSIONAL EXPERIENCE

Engineering Lead | Sentient Labs

11/2025 – 08/2026 | Remote | AI research company building reasoning systems, AI agents, and infrastructure for agent success management.

- **Built** LLM-powered agent systems using tool calling, structured outputs, state management, and guardrails, enabling iterative task execution across context reasoning, tool selection, and result processing.
- **Designed** production RAG pipelines using embeddings, semantic retrieval, and chunking strategies to ground agent responses in enterprise knowledge.
- **Designed** multi-step and multi-agent workflows using LangGraph, Temporal, and FSM-based orchestration for stateful and long-running agent execution.
- **Built** agent evaluation pipelines using DSPy and LLM-as-a-Judge to assess response quality, decision correctness, RAG performance, and regression across evaluation scenarios.
- **Developed** multi-provider LLM infrastructure using LiteLLM for OpenAI and Anthropic, supporting model routing, retries, cost and latency tracking, structured outputs, and distributed tracing.

Software Development Lead | Zeta Private Limited

05/2022 – 11/2025 | Bangalore | PayZapp mobile payment and financial services application for HDFC Bank

- Led architecture, solution design, and end-to-end delivery for mobile product initiatives across PayZapp.
- Led a team of 5 engineers to build and deliver PixelCard on PayZapp for HDFC Bank.
- Automated CI/CD workflows using GitHub Actions and Bitrise, improving development and release processes.
- Improved React Native application performance using Hermes and Flipper by addressing JavaScript-thread bottlenecks and FPS issues.
- Resolved 200+ production defects, reducing customer-reported issues by 30%.
- Built and established a new engineering team while adopting a new technology stack for cross-platform product development.

Technical Lead | mfine, Novocura Tech Health Services

08/2019 – 05/2022 | Bangalore | AI-powered healthcare platform connecting patients with doctors

- Led mobile application architecture, performance optimization, and production deployment for an AI-powered healthcare platform connecting patients with doctors.
- Drove technical execution across mobile product development, production releases, and engineering delivery.

Software Engineer II | Compassites Software Pvt Ltd

02/2018 – 08/2019 | Bangalore

- Developed React Native and iOS applications, contributing to mobile engineering initiatives and production product delivery.

iOS Developer | Globussoft Technology

11/2015 – 01/2018 | Bangalore

- Developed and supported iOS applications across product development and maintenance activities.

Application Support Engineer | WebDesignMagics

08/2013 – 10/2015 | Bangalore

- Provided application support and engineering assistance for production software systems.

PROJECTS

- **Refund Assistant Agent** | <https://support.balacc.dev/> | https://github.com/Balasnest/agent_reliability_lab
Repetitive refund requests drain support teams and frustrate customers. Built a conversational AI agent that reasons over refund policies, checks eligibility, handles multi-turn conversations, and escalates complex or out-of-policy requests to human support.
- **Sentient Chat | AI-Native LLM Application**
Built a production AI-native application centered on LLM conversations, reasoning workflows, and agent capabilities using React Native and Expo

EDUCATION

B.E. in Information Science | New Horizon College of Engineering, VTU | 2009 – 2013 | Bangalore

AWARDS

- **Spot Award** | Q2 2023, Q1 2024 | Recognized for exceptional leadership and contributions to the team.
- **Shining Star Award** | Q4 2024, Q2 2026 | Recognized for high ownership of individual and team goals and contribution to collaboration.

CERTIFICATIONS & LEARNING

- **Proficient AI Engineer:** Builder / Coder / Leadership / Core / Agentic / MLOps Tracks
- **AI Engineer Core Track:** LLM Engineering, RAG, Agents
- **AI Engineer Agentic Track:** The Complete Agent & MCP
- **AI Engineer Production Track:** Deploy LLMs & Agents