

Kavin B M

+91 9940147548 | kavinbm16@gmail.com | [LinkedIn](#) | [GitHub](#) | Chennai, India

SUMMARY

Applied AI Engineer focused on building reliable AI systems that bridge models and real-world applications. Strong foundation in machine learning fundamentals and backend engineering, with hands-on experience across retrieval systems, AI orchestration, and production deployments in robotics environments.

EDUCATION

College of Engineering, Guindy (Anna University) <i>Bachelor of Engineering in Computer Science Engineering — CGPA: 8.0/10</i>	Chennai, India 2021 – 2025
Chennai Public School <i>CBSE Higher Secondary Education — Percentage: 93%</i>	Chennai, India 2019 – 2021

EXPERIENCE

AI Engineer <i>Machani Robotics</i>	Jan 2026 – Present Bangalore, India
- Developing production AI infrastructure for robotics applications and intelligent systems deployment. - Building hybrid gRPC-MCP server architectures for scalable service and tool integration. - Contributing to real-time speech-to-speech conversational pipelines. - Developing internal AI dashboards and observability tooling for model monitoring and evaluation. - Working across backend systems, AI orchestration layers, and deployment infrastructure.	
AI Developer Intern <i>Brixbit Technologies</i>	Jan 2025 – Jun 2025 Chennai, India
- Built and automated end-to-end AI workflows using n8n and Replit for rapid AI prototype deployment. - Developed Retrieval-Augmented Generation (RAG) applications using LangChain and LlamaIndex. - Built agentic workflows capable of contextual retrieval and dynamic response generation. - Contributed to rapid prototyping and deployment of AI-driven solutions.	
Software Engineer Intern <i>PreludeSys India Pvt Ltd</i>	Jun 2024 – Jul 2024 Chennai, India
- Collaborated with engineering teams in the design, development, testing and debugging of software applications. - Participated in software development lifecycle activities across implementation and validation phases.	

PROJECTS

Spotify gRPC-MCP Server | *Go, gRPC, MCP, Protobuf*

- Built a hybrid gRPC and Model Context Protocol server exposing Spotify functionality as AI-native tools.
- Designed protobuf contracts and communication interfaces for low-latency tool invocation.
- Enabled LLM agents to search, control playback and interact with Spotify through standardized tool interfaces.

Pulse Observability Framework Contribution | *Go, OpenTelemetry*

- Contributed to a production-grade observability framework providing unified telemetry for modern applications.
- Implemented improvements to structured logging and telemetry pipelines across tracing and metrics modules.
- Worked with OpenTelemetry data models and distributed observability concepts.

Google API Linter VSCode Extension Contribution

- Contributed fixes and improvements to the Visual Studio Code extension ecosystem surrounding Google's API Linter tooling.

TECHNICAL SKILLS

Languages: Go, Python, SQL

AI Systems: Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Agentic Workflows, Model Evaluation, Prompt Engineering, Vector Databases

Backend Engineering: gRPC, Protocol Buffers (Protobuf), GraphQL, REST APIs, WebSockets, Model Context Protocol (MCP), A2A, Langchain, Opentelemetry (Grafana)

Infrastructure & Observability: Redis, NATS, Distributed Tracing, Metrics Collection, Structured Logging

Developer Tools: Docker, Git, Bazel, Linux