

LAKSHIT WASAN

New Delhi ◇ lakshitwasan31@gmail.com ◇ LinkedIn: [lakshitwasan](#) ◇ GitHub: [lakshitwasan](#)

PROFILE

Software Engineer specializing in backend architecture and AI-enabled systems. Experienced in building scalable microservices, LLM-powered APIs, RAG pipelines, and multi-agent systems deployed on cloud infrastructure. Proven track record of owning production systems end-to-end—from system design and API contracts to deployment, observability, and performance optimization.

EDUCATION

Bachelor of Technology in Computer Science and Engineering Sept 2021 – May 2025
BML Munjal University, Haryana
Cumulative GPA: 8.4 / 10.0

SKILLS

Core Backend Skills	Languages: Python, JavaScript, C++ Backend: FastAPI, Node.js, Express.js Databases: MSSQL, PostgreSQL, MongoDB, Snowflake Systems: REST APIs, Microservices, Caching, Async Processing, Containerization Cloud & DevOps: Docker, AWS, GCP, CI/CD, GitHub Actions
AI Systems & LLMOps	Frameworks: LangChain, LangGraph, Google ADK, Snowflake Cortex, Autogen RAG Pipelines: Embedding generation, chunking, vector indexing, hybrid retrieval Vector Search: FAISS, PGVector, Pinecone, ChromaDB, metadata filtering LLM Integration: Prompt engineering, multi-agent systems, A2A protocols, safety layers
Frontend (Supporting)	React.js (dashboards & internal tools)

EXPERIENCE

Software Engineer (Backend & AI Systems) Aug 2025 – Present
Nebula9.ai, Gurgaon, India

- Designed and deployed backend services powering LLM-driven enterprise workflows, reducing manual operations time by **30–40%** across 3+ client platforms.
- Built modular AI service layers integrating RAG components (embeddings, vector indexing, hybrid search), improving retrieval accuracy and lowering inference latency by **20%**.
- Owned multi-tenant system delivery end-to-end—writing HLD/LLD docs, creating API contracts, and ensuring **99.9% uptime** through staged rollouts and observability tooling.
- Developed conversational analytics agents over Snowflake semantic views using Cortex Agents and LLM routing, enabling natural language querying of multi-domain enterprise datasets.

Technologies Used: FastAPI, PostgreSQL, LangChain, LangGraph, Google ADK, Snowflake Cortex, Vector DBs, Docker, AWS

Software Engineer Intern (AI-Enabled Backend Systems) Oct 2024 – July 2025
Nebula9.ai, Gurgaon, India

- Developed backend components for RAG-powered analytics, abstracting LLM calls behind secure REST APIs and enabling real-time insights for financial and operational dashboards.
- Migrated legacy MERN architecture to **FastAPI + PostgreSQL**, improving performance and maintainability by **~35%** and enabling async model-serving flows.
- Implemented ingestion, embedding generation, and document-chunking pipelines for vector search, improving retrieval throughput and relevance.

Technologies Used: FastAPI, PostgreSQL, LangChain, LangGraph, Docker, AWS, MERN Stack

Web Developer Intern
Coding Blocks, Delhi, India

June 2022 – Aug 2022

- Delivered **6+ production-ready** applications with secure authentication, optimized data models, and improved MongoDB query efficiency by **up to 25%**.

Technologies Used: Node.js, Express.js, MongoDB, PassportJS, Joi, Mongoose

PROJECTS

LLM-Powered Conversational Analytics Platform *Snowflake, Streamlit, Cortex Agents, LLM Routing, Prompt Engineering*

Built a conversational analytics platform enabling natural language interaction over multi-domain marketing & retail datasets on Snowflake. Developed strategy-specific LLM agents over Snowflake semantic views (UMM, MTA, Sonic, Retail, EDM) with intent routing and dynamic tool selection. Converted simulation & optimization workflows into LLM-driven conversational execution, deployed via Streamlit on Snowflake.

AgentZero — Autonomous AI Coding Agent *Multi-Agent Systems, AST Parsing, Docker, Vector Search, LLM Routing*

Built a local-first autonomous coding agent using a multi-agent state-machine architecture covering planning, execution, and validation phases. Implemented AST-aware semantic retrieval (Tree-sitter + vector DB) for **high-precision code context understanding** at the symbol level. Designed a self-healing execution loop with secure Docker sandboxing and adaptive model routing for reliability and cost efficiency.

Enterprise Event Logistics Platform *FastAPI, PostgreSQL, Microservices, Docker, AWS*

Architected microservices for large-scale event logistics with independently scalable booking and user services and **zero production downtime**. Integrated optional natural-language interfaces for querying schedules using LLM APIs. Deployed FastAPI services on AWS with observability tooling, reducing debugging time by **30%**.

AI-Driven Financial Data Platform *FastAPI, PostgreSQL, RAG Pipelines, Vector DBs, React*

Developed a production-grade RAG system for real-time financial insights, implementing ingestion, vector indexing, hybrid search, and LLM orchestration APIs. Enhanced embedding retrieval and batching, reducing model-query latency by **25%**. Delivered production architecture with FastAPI backend and React frontend.

CERTIFICATION & PUBLICATIONS

- Introduction to Artificial Intelligence (AI) — IBM
- Natural Language Processing with Classification and Vector Spaces — DeepLearning.AI
- **Explainable ML for Peptide Therapeutics** — Co-authored a peer-reviewed study presenting an interpretable machine learning framework for identifying antimicrobial peptide candidates. The model leverages physicochemical properties such as hydrophobicity and dipeptide motifs, and powers the **Pred-AHCP** web server used for therapeutic peptide design and validation. [Published Paper](#) — [Pred-AHCP Web Server](#).