

Abhishek Jaiswal

📍 Bengaluru, India | ✉ ajaiswal9043@gmail.com | 📞 +91-9958045472 | 🌐 in/abhijaiswal18

PROFILE

Senior Software Engineer with 5 years of experience in designing and building backend services, maintaining distributed systems, and developing core networking features. Adept at solving complex problems, leading initiatives, mentoring interns and juniors, and delivering production-grade software in agile environments. Experienced in AI/agent systems and cloud infrastructure.

EDUCATION

B.Tech • CSE

Indian Institute of Technology
Guwahati
2017 – 2021

SKILLS

Languages

◊ Python ◊ TypeScript ◊ C ◊ C++ ◊ Java

Cloud

◊ AWS (EC2, S3, SQS) ◊ GKE ◊ Terraform

Infrastructure

◊ Docker ◊ Kubernetes ◊ KEDA

AI & LLM

◊ AI Agent Dev ◊ RAG ◊ Prompt Engineering
◊ LangChain

Frameworks

◊ FastAPI ◊ Flask ◊ Node.js ◊ Express.js
◊ Spring

DB

◊ PostgreSQL ◊ MySQL ◊ Redis ◊ ChromaDB
◊ MongoDB

Methods

◊ CI/CD ◊ Agile ◊ Unit Testing ◊ Microservices

Miscellaneous

◊ Elasticsearch ◊ Fluent-bit ◊ SD-WAN ◊ TCP/IP
◊ REST ◊ Network Security

LANGUAGES

◊ English (C2) ◊ Spanish (A1)

CERTIFICATIONS

Retrieval Augmented Generation (RAG)

April 2026 – Present

Claude Code: An Agentic Coding Assistant

May 2026

AWS Storage

June 2025

EXPERIENCE

Founding Engineer

Nov 2025 – Jan 2026

Knowl

Bengaluru, India

- Designed Python/FastAPI backend services handling **1M+ API requests/day**, deployed on AWS EC2 and GKE in a hybrid cloud architecture.
- Built an autonomous AI agentic system for language detection using LLM reasoning and multi-strategy fallbacks, achieving **45% accuracy improvement** via prompt engineering.
- Implemented Redis-based rate limiting and throttling for high-volume APIs ensuring stability under P95 loads.
- Developed scalable REST APIs with Pydantic validation in a reactive microservice architecture.

Member of Technical Staff 2

Jul 2021 – Oct 2025

VMware - Arista (Velocloud SD-WAN)

Bengaluru, India

- Built a network interface tracking system enhancing real-time statistics visibility for **200K+ edge devices** via kernel data extraction and user-facing dashboard integration.
- Delivered Advanced Threat Protection (ATP) with IDS/IPS, threat intelligence, and URL filtering, mitigating **10K+ monthly malicious connection attempts**.
- Designed and implemented IPv6 DHCP Prefix Delegation for dynamic prefix allocation across large SD-WAN deployments with zero-NAT global addressing.
- Led introduction of a unit testing framework for legacy C code, scaling coverage from **5% to 90%** and enabling full CI automation.
- Built a packet processing latency visualisation platform, improving performance analysis and debugging workflows.

PROJECTS

Interview Helper

March 2026 – Present

- Built a local Retrieval-Augmented Generation (RAG) application in Python using **Ollama**, **ChromaDB**, and Sentence Transformers to store and query interview experiences with semantic search using vector embeddings, and metadata-based filtering for efficient information retrieval.
- Designed an end-to-end pipeline for ingestion, indexing, retrieval, and LLM-powered question answering, enabling context-aware insights from historical interview data while running entirely on local infrastructure.

Librarian - Intelligent Book Chatbot

May 2026 – Present

- Engineered a multi-book conversational AI platform that ingests PDF documents, builds searchable knowledge bases, and delivers citation-backed responses using Retrieval-Augmented Generation (RAG), local language models.
- Implemented a hybrid search architecture leveraging vector retrieval, **TF-IDF** ranking, **cosine similarity** scoring, and metadata-based filtering for efficient and accurate document discovery.