

Abhishek Prakash

+91 8853318817 — Email — Github — LinkedIn

Technical Skills

AI & Machine Learning LLMs, RAG Pipelines, AI Agents, Prompt Engineering, LangChain, LlamaIndex, Pinecone (Vector DB), NLP
Languages Python (Advanced), SQL (PostgreSQL, MySQL), Java, C++, Go, JavaScript
Frameworks & Tools FastAPI, Spring Boot, Docker, Git, Google Cloud Platform (GCP), REST APIs, Microservices
Core CS Data Structures and Algorithms (DSA), DBMS, Operating Systems, OOP, SOLID Principles

Additional Skills

Coursework DSA, Operating Systems, OOP, DBMS, Compiler Design, Cryptography, Agentic AI
Languages English (Professional Working Proficiency), Hindi (Native)

Projects

CogniChat – Production-Grade AI Agent & RAG Platform

July 2025 – Aug 2025

[Link](#)

Tech: Python, FastAPI, LangChain, Pinecone, Gemini, Cohere, Docker

- Engineered a scalable Retrieval-Augmented Generation (RAG) pipeline using **Python** and **FastAPI**, designed to handle complex queries over large-scale unstructured datasets.
- Developed an **Autonomous AI Agent** using **LangChain** to orchestrate multi-step reasoning tasks, including tool-calling for dynamic data retrieval and document summarization.
- Optimized retrieval precision by 30% through a hybrid search approach (Semantic + Keyword) and a **Cohere Re-ranking layer**, integrated with **Pinecone** for high-concurrency vector storage.
- Built robust **data preprocessing scripts** to clean, transform, and validate diverse file formats (PDF, Markdown, TXT) into query-efficient chunks using recursive character splitting.
- Implemented real-time, asynchronous streaming of LLM responses via **Server-Sent Events (SSE)** and utilized **Docker** for containerized, cross-environment deployment.

Personal Knowledge Vault – AI-Powered Second Brain (Java Backend)

Sept 2025 – Present

[GitHub](#)

Tech: Java 17, Spring Boot, Apache Lucene, Ollama API, TailwindCSS

- Architected a high-performance Java backend using **Spring Boot**, following a layered architecture (Controller-Service-Repository) for separation of concerns.
- Developed an end-to-end **RAG pipeline**, utilizing **Apache Lucene** for vector-based semantic search and efficient document retrieval.
- Leveraged **Java Concurrency (ExecutorService)** to enable asynchronous multi-threaded document indexing, reducing ingestion time by 40%.
- Integrated **Apache PDFBox** for robust text extraction and implemented a custom recursive character splitter for chunking.
- Implemented RESTful endpoints with **Spring Web**, adhering to **SOLID principles** and ensuring thread-safe operations during concurrent LLM queries.

Post-Quantum Secure Tunneling (WireGuard & SSH)

Mar 2025 – May 2025

Tech: C, Python, WireGuard, SSH, McEliece Cryptosystem

- Engineered a hybrid cryptographic framework by integrating **McEliece Post-Quantum KEM** into the WireGuard Noise protocol.
- Developed a custom **SSH Application** that tunnels traffic through a quantum-resistant handshake, mitigating "Harvest Now, Decrypt Later" risks.
- Modified low-level C modules to establish a dual-encryption pipeline, combining classical **Curve25519 ECDH** with PQC algorithms.
- Implemented secure session management and port-forwarding logic to ensure robust application-level connectivity.

Experience

Undergraduate Researcher, IIIT Kalyani

Jan 2025 – Present

- Spearheading research on **Post-Quantum Cryptography (PQC)**, specifically optimizing Hermitian code-based McEliece systems.
- Developed and optimized **List Decoding Algorithms** in C++, improving the error-correction capability and security margins of the cryptosystem.
- Automated performance analysis using **Python and Java**, generating data-driven insights on encryption throughput and memory overhead.

Education

IIIT Kalyani

2021 – 2025

B.Tech in Computer Science Engineering — CGPA: 7.80

Nagaji Saraswati Vidhya Mandir, Ballia

2020

12th: 77.46% — 10th: 89.8%

Certificates

- Google Cloud Computing Foundations: Cloud Computing Fundamentals
- Google Cloud Computing Foundations: Infrastructure in Google Cloud