CHAITANYA KHARCHE

LinkedIn | Github | kharche.c@northeastern.edu | 857-313-0545

Full-stack software engineer with 3 years' experience in machine-learning, Gen AI & RAG pipelines + scalable web & cloud deployments. Available immediately to join full-time & drive high-performance solutions.

PROGRAMMING SKILLS

Languages: JavaScript, Python, Java, SQL, Kotlin
Frameworks & Tools: React, Node.js, Spring Boot, Django, Flutter, RESTful APIs, MySQL, PostgreSQL MongoDB
Cloud & Tools: AWS (EC2, Lambda, S3), Docker, Git, Azure Devops (Azure Monitor, Pipelines)
AI/ML: LangChain, Agentic AI, Vector Databases (Pinecone, FAISS), Azure OpenAI, RAG Systems, LLMs

CERTIFICATIONS

Amazon Web Services - AWS Certified Solutions Architect **| Sept. 2025 -** Certificate available on 09/23/2025
Amazon Web Services - AWS Certified Cloud Practitioner **| April 2025**

EXPERIENCE

FounderWay.ai (Dium Inc.) | Software Engineer Co-op   Remote, US | 09/2024 - 12/2024
◦ Built & Fixed frontend assets using React, Next.js, and Formik, fixing components, designs and bugs.
◦ Engineered Azure Open AI Vector Flow RAG, enhancing platform interactivity by 50% for 1000+ users.
◦ Made a LLama based co-founder matching algorithm, with NLP inputs to boost accuracy by 40%.
◦ Developed backend in Java Spring Boot & managed PostgreSQL tables, tuning indexing for faster queries.

Bredcrums AI | Software AI Developer Intern   Remote, US | 06/2023 - 08/2023
◦ Developed Gen-AI platform in Bubble.io, enhancing data analysis & research for over 100+ researchers.
◦ Used LangChain's Chains and Prompt Templates to streamline workflows using GPT-4 for processing.
◦ Enhanced research processes by embedding into Pinecone DB, for better prompt engineering output.
◦ Executed custom SQL queries refining data retrieval & storage, reducing compute costs by 30%.
◦ Transitioned to microservices via AWS Lambda and API Gateway, for dealing with complex calls to AWS API.

SDK Infotech | Software Engineer - II   Pune, IN | 08/2021 - 08/2022
◦ Optimized REST APIs using Express.js & Node.js, enhancing system integration and data throughput by 14%
◦ Conducted testing using Mocha and Chai, achieving 98% coverage and decreasing bugs by 10%.
◦ Managed Docker containers on AWS, designing scalable systems that improved resource utilization by 20%.
◦ Wrote webhooks in Node & Express.js to automate API responses, improving system responsiveness by 12%.

SDK Infotech | Software Engineer - I   Pune, IN | 09/2020 - 08/2021
◦ Built a Python application that uses a Django backend, streamlining payroll processing capabilities.
◦ Integrated MongoDB to sustain rapid CRUD operations under peak loads to handle millions of rows of data.
◦ Refactored loop-based MongoDB queries into a single indexed calls, reducing load time from 48.2s to 19s.
◦ Supported cloud operations by revamping AWS EC2 instances for data handling and application servers.

CUBYNE | Lead Developer   Mumbai, IN | 07/2020 - 05/2022
◦ Co-founded a freelance development team of 4 engineers, delivering web & mobile solutions to 10+ clients.
◦ Built GreenSat (5K+ downloads), a precision farming app with ML-based disease detection, & 5 production platforms using Flutter, Next.js, and AWS. Managed full product lifecycle via REST APIs, Firebase & CI/CD.

PROJECTS

CurriculAI: Built a **curriculum optimization system** for personalized 32-course degree plans. The system uses an **Agentic RAG pipeline** with **BAAI/bge embeddings** and **NetworkX** dependency graphs to **semantically** analyze 200+ courses. Its core **Llama-3.1-70B** model was trained on an **Nvidia H200 GPU**.

ProfitBook: Achieved a 40% backtested win rate with an AI trading system using specialized **sentiment, financial, and instruction-following LLMs** for real-time analysis. Deployed as **FastAPI microservices**, the system is optimized with **4-bit quantization** and makes decisions in under **60 seconds**.

LegalEaze: Engineered a **NLP-based LEGAL text simplification system** that achieves a **7th-grade reading level** (6.9 FKGL score). Generated a high-quality 500k+ sample dataset from 59k txt files using a **Llama-3.3-70B** model, then **fine-tuned a specialized** facebook/bart-large-cnn **model** with **LoRA** on Nvidia **A100/H200 GPUs**.

EDUCATION

Northeastern University, M.S Computer Science   Boston, MA
Relevant Courses: Mobile App Dev, DBMS, Algorithms, OOD, Data Structures   **Expected** - 05/2025

Savitribai Phule Pune University, B.E Civil Engineering