

Rohit Choudhary

Principal Software Engineer

+91 9971940271 | mrohitchoudhary@gmail.com | [LinkedIn](#) | [Github](#)

PROFESSIONAL SUMMARY

Principal Engineer with 12+ years of experience in software development, specializing in MLOps, GenAI platforms, and scalable cloud-native architectures. Expertise in designing distributed systems, leading cross-functional teams, and delivering impactful AI-integrated solutions.

TECHNICAL SKILLS

Programming Languages: Python, Golang, Node.js, JavaScript, TypeScript | **Frameworks:** React, Redux RTK, FastAPI, Remix, Apache Airflow, Kubebuilder, GraphQL | **Frontend:** React.js, TypeScript | **Backend:** Node.js, Python, Golang, Kubernetes, Docker, Jenkins, CI/CD | **Databases:** PostgreSQL, MySQL, DynamoDB, Redis, Kafka | **GenAI & AI Tools:** Hugging Face, LangChain, OpenAI API, PyTorch, TensorFlow, Pinecone, RAG, Prompt Engineering, MLflow, Model Serving, LLM Fine-Tuning, LLM Inference Services, Synthetic Data Generation | **Platforms:** AWS, Azure, Google Cloud | **Methodologies:** Agile, Scrum, Kanban, MECE, 80/20 Rule, SCQA | **Others:** Project Management, Microservices, Serverless, Event-Driven Design, Hexagonal Architecture, Orchestration

PROFESSIONAL EXPERIENCE (12+ Years)

Principal Software Engineer, ML Studio & Platforms | 2022 - Present

SambaNova Systems | Bangalore

- Designed and launched a scalable MLOps and GenAI platform, optimizing model training, deployment, and management with microservices and inference solutions.
- Developed a Bring Your Own Model (BYOM) feature, enabling customers to deploy custom AI models, reducing integration time by 90%.
- Engineered Kubernetes CRDs with KubeBuilder, reducing operational overhead for scalable, long-running training jobs by 40%.
- Built high-performance inference APIs with stream support, supporting >1000Tokens/Sec
- Led the design and development of a next-gen UI using TypeScript, React, and Redux RTK, boosting customer satisfaction (CSAT) by 70%.
- Contributed to the AI Starter Kit open-source project on GitHub, accelerating enterprise-grade AI application development.
- Developed a Kubernetes-based BYOC solution for AI model conversion and deployment, reducing conversion time from weeks to minutes.
- Improved Helm charts, logging, and monitoring for both air-gapped and non-air-gapped environments, enhancing platform reliability.
- Mentored developers, fostering a culture of best practices and technical excellence.
- Collaborated with Customer Engineering, Product, and ML teams to deliver impactful features and solve real-world challenges.
- Managed sprints, daily stand-ups, and release cycles to ensure timely and high-quality feature delivery.

Digital Specialist 2, Build By McKinsey | 2018 - 2022

McKinsey & Company | Bangalore

- Consulted for Fortune 500 clients (eBay, KPC Indonesia, TechcomBank, Telkomsel), designing and implementing cloud-native architectures and integrating digital tools.
- Introduced Apollo-based GraphQL patterns with Node.js and Python, enabling scalable, maintainable systems for diverse business needs.
- Streamlined ETL pipelines with Apache Airflow DAGs and implemented a Control Tower for centralized monitoring and logging.
- Architected a high-performance microsite for eBay's FlashSale events, handling over 300 orders per minute.
- Led the development of a Mortgage Loan System for TechcomBank, reducing loan disbursement time from 3 months to 1 week.

- Managed a team of 12+ developers across Vietnam, Indonesia, and India, delivering complex projects on time and within scope.

Lead Software Engineer | 2017 - 2018

Intelligent | Gurgaon

- Led end-to-end product engineering for Expertise.com, a curated platform connecting users with vetted experts across industries.
- Owned full-cycle development across backend (Python, Ruby, Node.js) and frontend (React.js), ensuring scalable, maintainable codebases.
- Optimized DevOps workflows using Chef and custom automation scripts, reducing deployment times and improving infrastructure reliability by 20%.
- Developed background automation jobs, improving operational efficiency and reducing manual intervention by 30%.

Senior Associate L1 | 2016 - 2017

Publicis Groupe | Gurgaon

- Served as JavaScript SME for the UK's largest retail chain; led and mentored a 6-member team on React.js, TDD, Agile, and microservices, increasing code coverage by 30% and reducing deployment time by 80%.
- Delivered a high-performance retail buying and sourcing tool, improving product time-to-market by 25%.

Software Development Engineer 2 | 2014 - 2016

Myntra Jabong | Gurgaon

- Developed and maintained Cart, Checkout, and Authentication modules for Jabong's Mobile Web platform, improving mobile shopping experience.
- Built a custom authentication system with Passport.js and Express-Redis, boosting active customer engagement by 40%.
- Contributed to Artemis, a SOA-based rebuild of Jabong.com, delivering a faster and more secure customer experience.

PROJECTS

- **AI Platform Development:** Designed and led the implementation of scalable MLOps platforms for training, deploying, and managing private and open-source LLMs.
- **Cloud AI Solutions:** Built high-performance, cost-effective AI inference platforms for startups and enterprises.
- **Applied AI Applications:** Developed and launched production-ready AI/ML products for document analysis, financial insights, and real-world problem solving.
- **Developer Enablement:** Created robust CLI-based tools and SDKs to streamline and automate ML workflows for engineering teams.
- **Others:** OIDC Auth Module, RTCM, Control Tower

EDUCATION

Master's in Computer Application | 2012

Punjab Technical University

Bachelor's in Computer Application | 2009

IGNOU

CERTIFICATIONS

- AWS Solutions Architect , Azure AI Engineer Associate

LANGUAGES

- Hindi, English, Bengali