RISHABH BATRA

+91 9996254447|| rishabhbatra81@gmail.com| Gurgaon, India
LinkedIn: https://www.linkedin.com/in/rishabh-batra-6190841b3/ | GitHub | Leetcode

Java Backend Developer | Microservices & API Specialist | AI Assisted Development

Results-driven **Backend Developer /FullStack** with **4 years** of experience in building scalable microservices and enterprise-level APIs within the **banking** and **fintech** domain. Recently focused on developing **RAG-based knowledge** retrieval systems and **integrating LLMs** . Working with **AI Assisted tools** Adept at designing high-performance systems using **Java, Spring Boot**, and **WebFlux**, with proven success in reducing latency, improving code quality, and optimizing backend processes. Demonstrates a strong command of secure coding practices, unit testing, and **CI/CD** integrations. Known for cross-functional collaboration, proactive problem-solving, and clear technical

KEY SKILLS

Backend Development | API Design | Microservices Architecture | Code Quality | Vulnerability Remediation | Performance Optimization (Latency Reduction) | Unit Testing | Test Coverage Improvement | Deployment Management (Dev, SIT, UAT, Production) | Agile/Scrum Participation | Cross-Functional Team Collaboration | Technical Documentation & Release Notes | Root Cause Analysis | Technical Writing | Knowledge Sharing | Continuous Improvement | Reusability in Development | Agentic AI |AI Assisted Development Tools | LLM Powered | RAG Pipelines | Prompt Engineering | Agentic AI Engineer

WORK EXPERIENCE

Software Engineer | Wissen Technology, Gurgaon, India | June 2025 – Present

Client: Macquarie Group

- Built a **RAG-based knowledge retrieval and LLM inference pipelines** so new engineers can understand legacy modules and dependencies in minutes, not hours.
- Built messaging with 2 platforms (**Solace PubSub+**) on Java/Spring Boot for a highvolume platform; enabled realtime trade acks and position sync across 3 asset classes (equities, FX, derivatives).
- Automated bulk trade upload validation in Meteor/JavaScript, calling external REST APIs for settlement dates, restricted stock checks, and market calendars fewer manual touchpoints and faster processing.
- Authored 4 doc types—architecture guides, API specs, implementation plans, and runbooks—so dev/support teams could move faster and with fewer escalations.
- Configured **Claude Code** using Amazon Bedrock, enabled sub-accounts, and created a SAM user for secure access. Built a CI/CD pipeline to automate deployment and manage inference profiles for streamlined model integration

Software Engineer | FIS Global, Gurgaon, India | Feb 2022 – Apr 2025

- Reduced **system latency** by **60%** by designing and deploying high-performance reactive microservices using Spring WebFlux, enhancing client experience on the banking platform.
- Improved system reliability by resolving code vulnerabilities identified by SonarQube & Black Duck, ensuring secure and compliant releases.
- **Boosted API performance** and platform efficiency by developing and managing RESTful services, cutting processing time and improving throughput.
- **Enhanced development agility** by building REST APIs aligned with OpenAPI Specification standards, enabling **100%** documentation consistency and faster integration.
- **Streamlined development cycle** by **40%** through reusable backend components and standardized error-handling templates, improving maintainability.
- **Achieved 85%+ unit test coverage** using JUnit5 and Mockito, ensuring API robustness and reducing regression risk.
- Integrated Kafka consumers to process upstream service events and persist data reliably, contributing to real-time insights and system responsiveness.
- Improved **cross-team delivery** by coordinating deployments across Dev, SIT, UAT, and Production environments, ensuring smooth rollout and version control.
- **Accelerated business workflows** by delivering responsive web services for business account opening, enhancing user experience and backend integration.
- Maintained Agile delivery momentum by participating in sprint planning, stand-ups, retrospectives, & driving continuous improvement.
- **Ensured release transparency** by drafting detailed release notes for updates and maintenance deployments, supporting clear stakeholder communication.

Technical Content Writer | GeeksForGeeks (Remote – Freelance) | Jun 2021 – Dec 2021

- Established **thought leadership** by authoring widely read technical articles on problem-solving, data structures, and algorithms.
- Simplified complex concepts into digestible content, improving reader comprehension by **75%** through clear explanations & structured solutions.
- Contributed to increased **platform engagement** and learning effectiveness, receiving positive feedback from a diverse reader base.

PROJECTS

Persona Studio — AI-Powered Professional Headshot Live Link

Tech Stack: React 19, TypeScript, Vite, Tailwind CSS, Google Gemini AI, Vercel

- **Developed an AI-powered headshot generation web app** using React 19 and TypeScript integrating **Google Gemini 2.5 API** for real-time image generation with 5-15 second processing time
- Designed a **privacy-first architecture** where all image processing occurs client-side, eliminating server storage and ensuring **GDPR-compliant data handling**
- Engineered image compression and optimization utilities reducing upload sizes by 60%+ while maintaining quality for AI processing with **optimized Base64 processing**
- **Integrated Google Gemini 2.5 AI (Flash & Pro models)** for real-time image generation, implementing advanced prompt engineering techniques to preserve facial identity
- **Created 12+ professional style preset**s with **dynamic prompt generation**, plus custom style support using **natural language processing**

AI-Powered Recruitment Automation System Workflow
Youtube Demo Link   GoogleDrive Demo Link

- **Engineered end-to-end recruitment automation platform** using **n8n**, **ElevenLabs** voice AI, and **Google Gemini LLM**, reducing manual screening time by 85% and enabling 24/7 candidate evaluation
- **Built full-stack job portal application prototype using Lovable** with React 18, TypeScript, Tailwind CSS, and **Supabase** backend, featuring dual-portal architecture for recruiters and job seekers with real-time application tracking
- **Developed intelligent resume parsing and ATS scoring system** with weighted evaluation criteria (40% experience, 20% skills, 20% projects, 20% education), achieving 7.5+ threshold accuracy for candidate shortlisting
- **Integrated voice AI screening agents** with ElevenLabs and **Twilio APIs**, automating phone interviews and generating structured transcripts for candidates with real-time conversation analysis.
- **Implemented dynamic workflow orchestration** with automated email notifications, Google Calendar scheduling, and streamlining recruitment pipeline from application to interview scheduling

TECHNICAL SKILLS

Programming Languages: Java, J2EE, C, C++, SQL
Frameworks & Libraries: Spring, Spring Boot, Hibernate, JPA, WebFlux
API Development: RESTful APIs, OpenAPI Specification
Testing Tools: JUnit 5, Mockito
Build & Version Control: Maven, Git, Bitbucket
DevOps & CI/CD: Jenkins, Docker, Kubernetes
Messaging & Streaming: Apache Kafka
Caching & Databases: Redis, MySQL
Other Tools: SonarQube, Black Duck, JIRA , Low Code Tools
Development Practices: Microservices, Reactive Programming, Multithreading, Agile, CI/CD
Concepts: Data Structures, Algorithms, Object-Oriented Programming (OOP), SOLID Principles, Design Patterns
AI & Modern Tech: OpenAI/Claude/Gemini Api/Perplexity/Lovable,Supabase, AI Agents ,n8n workflow, Prompt Engineering

CERTIFICATIONS

- Java Object-Oriented Programming – LinkedIn Learning
- Java Lambdas and Streams – LinkedIn Learning
- Microservices and REST API – Pluralsight
- Learning REST API – LinkedIn Learning

EDUCATION

- **Bachelor of Technology (Computer Science Engineering) |** *Chandigarh Group of Colleges, Landran | 2022*

Hackthon's & Competition

Project: Lavio – Voice-Powered AI Browser Extension
[Demo](#) [Hackathon](#) [Github](#)

- **Engineered end-to-end recruitment automation platform** using **n8n**, **ElevenLabs** voice AI, and **Google Gemini LLM**, reducing manual screening time by 85% and enabling 24/7 candidate evaluation
- **Built full-stack job portal application prototype using Lovable** with React 18, TypeScript, Tailwind CSS, and **Supabase** backend, featuring dual-portal architecture for recruiters and job seekers with real-time application tracking
- **Developed intelligent resume parsing and ATS scoring system** with weighted evaluation criteria (40% experience, 20% skills, 20% projects, 20% education), achieving 7.5+ threshold accuracy for candidate shortlisting
- **Integrated voice AI screening agents** with ElevenLabs and **Twilio APIs**, automating phone interviews and generating structured transcripts for candidates with real-time conversation analysis.
- **Implemented dynamic workflow orchestration** with automated email notifications, Google Calendar scheduling, and streamlining recruitment pipeline from application to interview scheduling

Project: Cronos: AI Reminder App
[Demo](#) [Hackathon](#) [Github](#)

- Built an offline-first **React Native** app using **Zustand** and **MMKV**, ensuring the UI never blocks while waiting for complex AI inference latency or background syncs from the server.
- Implemented a robust bidirectional data sync engine using **Supabase** and **PostgreSQL**. Managed state consistency across iOS and Android by writing timestamp conflict resolution logic and debouncing real-time WebSockets to prevent feedback loops.
- Designed a custom NLP pipeline that streams raw user audio to the **OpenAI Whisper** API, then uses **GPT-4o-mini** to extract structured JSON data. Handled edge cases by building dynamic prompt injections that inject user timezone offsets for accurate scheduling.
- Integrated the **Perplexity API** to automate complex task research, building a centralized orchestration layer that caches identical LLM requests in-memory and enforces rate limiting to drastically cut down on API cost and latency.