

Himanshu Raj

+917009557681 himanshuraj.tn@gmail.com [linkedin.com/in/hr7657316](https://www.linkedin.com/in/hr7657316) github.com/hr7657316

Education

Lovely Professional University

Bachelor of Technology in Computer Science

Completed May 2026

Phagwara, Punjab

- **Relevant Coursework:** Data Structures and Algorithms (C++), Prob & Stat in CS (Python), Intro to CS II (C++), Linear Algebra w/Computational Applications (Python)

Experience

Shopflo(Pine Labs)

Associate Software Engineer

Apr 2026 – Oct 2026

Bengaluru, Karnataka

- Resolved 150+ production issues across **Java/Spring Boot** microservices (Checkout, Discount, Catalog & Shipping).
- Integrated **Shopify GraphQL** and optimized REST APIs, improving sync for **1,000+ product variants**.
- Built weight- and price-based shipping conflict detection, improving reliability for **1,200+ merchants**.
- Developed **MCP tooling** for custom ETD configuration and automated workflows with **Claude MCPs & Agent Skills**.
- Partnered with **Solutions** to resolve end-to-end merchant issues from diagnosis to production fixes.

TetraNoodle Technologies

AI Engineer (Forward Deployed Engineer)

Dec 2025 – Apr 2026

Vancouver, Canada

- Built **Triton**, a private AI chat platform powered by **MCPs, private connectors & Composio**, replacing the company's OpenAI ChatGPT platform for organization-wide use.
- Built AI document automation, reducing manual processing by **70%** across **500+ documents/week**.
- Improved extraction accuracy by **35%** through structured parsing and validation workflows.
- Worked directly with clients to translate requirements into production AI solutions, owning delivery from development to deployment.
- Deployed AI services across **Azure and AWS**, with CI/CD pipelines for reliable production releases.

Projects

Kintsua | MCP, GitHub, APIs

- Self-maintaining integrations for AI agents: Kintsua finds every external API and MCP tool an agent repo depends on, watches those upstreams for breaking changes, opens a GitHub issue within an hour of one landing, and opens a tested fix PR for common cases.

GBDM Query System | Python, FastAPI, React.js, SQLite, Groq API, Llama 3.3, Vite

- Built a graph-based system visualizing SAP O2C flows as a force-directed graph with 1,262 nodes and 4,732 edges.
- Implemented a two-pass NL-to-SQL engine using Llama 3.3 70B for querying 18 tables and 21K+ rows via natural language.
- Designed guardrails for domain filtering, prompt injection detection, and SQL validation ensuring read-only execution.

INFO BOX – AI Document Intelligence Platform | Python, Flask, PyTorch, Pinecone, Groq, Gemini, PyMuPDF

- Automated railway document workflows using OCR and RAG pipelines, cutting manual handling time by 60%.
- Orchestrated dual vector retrieval with Pinecone to power semantic search and department-level knowledge access.
- Leveraged LLM inference for summarization, translation, and contextual Q&A with confidence scoring and source attribution.

Technical Skills

Languages: C, C++, Java, JavaScript, TypeScript, Python, SQL, RegEx

Backend: Spring Boot, Django, Flask, FastAPI, Node.js, GraphQL, REST APIs

AI/ML: Generative AI, LangChain, TensorFlow, NLP, Computer Vision, Scikit-learn, Pandas, NumPy

AI Engineering: Claude Code, MCPs, AI Agents, Automations, Composio

Cloud: AWS Lambda, API Gateway, DynamoDB, EC2, S3, Docker

Databases: PostgreSQL, MySQL

Tools: Git, GitHub, IntelliJ IDEA, VS Code, Postman, cURL, Jira, Slack, Google Colab