

PARVEZ AHMED

AI Engineer · Agentic Systems · Backend & LLM Applications

Bengaluru, Karnataka 560093, India · +91 78988 15419 · 126ahmedparvez@gmail.com
parvez-ahmed.com · LinkedIn · GitHub

SUMMARY

AI Engineer with production experience building agentic LLM systems, Retrieval-Augmented Generation (RAG) pipelines, and machine learning models in enterprise banking. Combines hands-on model training and multi-model orchestration with deep backend engineering across Java, Spring Boot, and Python. Currently a Senior Analyst at Northern Trust, architecting LLM-driven knowledge retrieval over legacy systems and training ML models to automate financial reconciliation at scale.

PROFESSIONAL EXPERIENCE

Senior Analyst — AI Engineer

Apr 2026 – Present

Northern Trust · Bengaluru, Karnataka

- **Built an agentic RAG system over a legacy AS/400 codebase** using Azure OpenAI (GPT) models and a Neo4j knowledge graph — implementing LLM-driven Cypher query generation, graph retrieval, and inferencing through a custom chat interface that lets developers query decades-old business logic in natural language.
- **Trained a machine learning model to automate financial reconciliation** on historical manual-match data, learning to resolve one-to-many and many-to-many break patterns that the TLM auto-matching engine cannot — reducing manual reconciliation workload across operations teams.
- **Upgraded 16 Java microservices from Java 8 to Java 21 and Spring Boot 2.x to 3.x** with zero production downtime, and remediated 800+ CVEs by hardening dependencies and container configurations across enterprise banking infrastructure.

Analyst, Software Engineer

Aug 2024 – Mar 2026

Northern Trust · Bengaluru, Karnataka

- **Led zero-downtime migration of 21 Spring Boot microservices (v2.1.x → v2.7.x)**, resolving dependency conflicts across enterprise banking infrastructure.
- **Built a Python report-conversion tool** transforming large TLM HTML reports into optimised XLSX files — cutting load time from several minutes to under 5 seconds; adopted company-wide by operations teams.
- **Engineered an ISO-to-CAMT.053 conversion pipeline** parsing large financial XML files into optimised chunks, reducing processing time from ~9 minutes to under 50 seconds (82% reduction).
- **Automated CI/CD pipelines via GitHub Actions**, cutting deployment time by 40% and improving release reliability across PCF-hosted services.
- **Handled RHEL7 → RHEL8 upgrade of Unix production servers** for a system with 30+ upstream dependencies.

Front-End Engineering Intern

Jun 2023 – Sep 2023

Northern Trust · Bengaluru, Karnataka

- **Improved frontend Lighthouse scores by 25%** through targeted audits, code-splitting, and accessibility fixes; integrated REST APIs powering dynamic financial dashboards.

Full-Stack Developer Intern

Oct 2022 – Mar 2023

Skoda Volkswagen Training Academy · Pune, Maharashtra

- **Built a full-stack Decision Support System (Django, React, PostgreSQL, Docker)** for sales and HR analytics across 100+ dealership branches, replacing manual CSV aggregation with on-demand real-time reporting.

PROJECTS

DocVal – Document Intelligence for bank statements

- **Built a deterministic parser for native PDFs**, and a vision model for scans with a per-page fallback that rescues the dense statements a full-document pass garbles. Every row then has to survive arithmetic — opening balance + credits – debits must equal closing, per page and per statement.

ReconMatch – Reconciliations tool for Ledger and statement side transactions

- **Matching runs in tiers** — exact, date-windowed, then split matching for the one-to-many cases that break naive joins. Nothing is a black box: every match carries its reasoning, and every unmatched line is classified as a typed break with a suggested resolution

Planner / Implementer Agentic Coding Workflow — Multi-Model Orchestration

- **Designed a two-tier agentic coding system orchestrating multiple LLMs** — a high-capability planner model generating structured task specs that drive a separate implementer model for execution — coordinated through a shared spec file with persistent story files, session logs, and model-tier assignment for cost-efficient autonomous development.

agentguides.dev — Personal Blog on the AI Agent Ecosystem

- **Author and maintain a personal blog on the AI agent ecosystem** published weekly — covering up-to-date LLM leaderboards, hands-on reviews of new models and frameworks, head-to-head comparisons of AI coding tools and IDEs, and emerging industry trends. Built with Astro/Starlight and deployed on Cloudflare Pages.

TECHNICAL SKILLS

AI / ML: RAG pipelines, LLM agents & agentic systems, multi-model orchestration, ML model training, prompt engineering, vector search, Azure OpenAI, Ollama

Languages: Python, Java, TypeScript, JavaScript, SQL, Shell Scripting

Frameworks & Infra: FastAPI, Spring Boot, Django, React, Node.js, Docker, GitHub Actions, Cloudflare Pages

Databases: PostgreSQL, MySQL, Snowflake, Neo4j, Pinecone

EDUCATION

B.Tech, Computer Science Engineering

Sep 2020 – May 2024

Vishwanath Karad MIT World Peace University · Pune, Maharashtra

- CGPA: 9.55 / 10 · Top-6 finalist, Northern Trust Intern Hackathon

Class XII (CBSE)

2020

Ryan International School · Raipur, Chhattisgarh

- 95.6%