

Neeraj Kumar

AI Systems Architect & Technology Leader | VP Engineering

+91 7017876527 | neeraj.k2401@gmail.com | [LinkedIn](#)

Professional Summary

VP of AI Engineering with 9+ years architecting enterprise AI systems at scale. Built a cloud ERP platform from the ground up — now serving 500,000+ clients globally — then led its transformation into an AI-native suite driving 70% workflow automation through LLM orchestration, Autonomous Agents, and MCP integration. Leading a 60+ member engineering organization building agentic AI systems for 5,000+ enterprise businesses.

Core Competencies

AI & Agentic Systems: LLM Orchestration, Agentic AI, Model Context Protocol (MCP), RAG, LangChain, LangGraph, LlamaIndex, Claude/Llama APIs, Hugging Face, vLLM, Ollama, Amazon Bedrock, Google Vertex AI, MLflow, PyTorch, TensorFlow, Generative AI, Responsible AI

Vector & Search: Qdrant, Milvus, Weaviate, Faiss, Pinecone, Elasticsearch

Backend Development: Java, Rust, Python, Node.js, Spring Boot, Quarkus, FastAPI, Axum, Actix Web, Django, Flask, Kafka

Frontend Development: Angular, React, Next.js, Flutter

Database Management: MySQL, PostgreSQL, MongoDB, ClickHouse, Redis

DevOps & Cloud: Terraform, Docker, Kubernetes, AWS, GCP, GitHub Actions, GitLab CI/CD, Helm

Monitoring & Observability: OpenTelemetry, Grafana, Prometheus, Loki, Tempo, Langfuse, ELK Stack

Architecture: Microservices, SOA, Event-Driven, Cloud-Native, API Gateway, Distributed Systems, Platform Security

Quality Assurance: Playwright, JMeter, Selenium, K6, Qodana, BrowserStack

Leadership: Engineering Strategy, OKRs, Strategic Roadmapping, Agile Methodology, Team Building (scaled to 60+), Technical Due Diligence, Stakeholder & Investor Communication, Cross-Functional Team Management, Product Development Lifecycle

Professional Experience

VP of AI Engineering | HostBooks | Gurgaon, India

August 2025 — Present

- Leading architecture and engineering strategy for AI-powered ERP Suite serving 5,000+ enterprise businesses globally
- Scaling AI capabilities built during DGM phase — Zeno AI, Enterprise RAG, AP/AR Automation, and MCP — into unified platform-wide experience
- Driving LLM orchestration across the platform using Claude, Llama, and Amazon models via AWS Bedrock and Google Vertex AI
- Managing 60+ member engineering organization; achieved 70% reduction in manual workflows while building AI-first engineering culture

DGM — Engineering & Product | HostBooks | Gurgaon, India

July 2023 — August 2025

- The AI-native transformation phase — designed and shipped every major AI capability: Zeno AI (Aug 2023), AP/AR GenAI pipeline (Jan 2024), Enterprise RAG (Aug 2024), and MCP integration (Jan 2025)
- Led 60+ member engineering team across product development, delivering scalable enterprise solutions
- Transformed HostBooks from cloud-based ERP into AI-native Suite with modern microservices architecture
- Conducted technical due diligence supporting the \$5M Series B funding round
- Collaborated with third-party vendors: Shopify, Shiprocket, HotelLogix, Opera, Amazon, Flipkart

Chief Manager — Engineering & Product | HostBooks | Gurgaon, India

December 2021 — July 2023

- Directed cross-functional teams to deliver multiple high-impact projects including ERP modules and fintech integrations
- Integrated banking solutions (ICICI, HDFC, JP Morgan Chase) and fintech platforms (Plaid, Razorpay, AvaTax); delivered \$5M+ in SaaS modules
- Drove 85%+ automation of back-office workflows across the platform; mentored 30+ tech leads/engineers
- Scaled bank reconciliation ML models for multi-bank, multi-currency data; built Invoice Processing v1 (PaddleOCR + template extraction)

Technical Project Lead / Senior Software Developer | HostBooks | Gurgaon, India

August 2019 — December 2021

- Designed scalable architectures and supervised development teams for enterprise application delivery
- Led development of multiple ERP modules including accounting, payroll, inventory, fleet management, and POS
- Implemented compliance features for GST, TDS, and VAT supporting multi-entity operations
- Engineered core accounting engine handling multi-currency transactions across 10,000+ concurrent tenants
- Deployed the first production ML model inside the ERP — bank reconciliation engine using XGBoost and scikit-learn

Software Developer | Codefire Technologies | Noida, India

August 2017 — July 2019

- Architected HostBooks ERP from ground up as sole technical architect, leveraging SOA principles
- Delivered classical ML solutions for multiple clients: recommendation engine (collaborative filtering), document similarity (TF-IDF), NLP tools for business process automation
- Mentored 10+ Engineers/Interns, improving team efficiency and development tools

Software Engineer | Redian Software | Noida, India

January 2017 — August 2017

- Developed enterprise software solutions and financial applications
- Built backtesting frameworks and wealth management platforms

KEY PROJECTS & ACHIEVEMENTS

Agentic AI ERP Suite (HB Zeno AI)

- Architected Zeno AI — the agentic intelligence layer orchestrating across Enterprise RAG, AP/AR Automation, MCP integration, and Financial Insights Engine
- Serves 5,000+ enterprise businesses, reducing operational overhead by 70% through conversational AI interactions
- Multi-model support: Claude, Llama, and Amazon models via AWS Bedrock and Google Vertex AI with cost-optimized routing

Model Context Protocol (MCP) Integration

- Built core ERP MCP service exposing 17 tools across 6 categories: Master Data, Transactions, Reports, Utilities, Calendar, Company config
- FastMCP framework with SSE transport; dual authentication (internal AI agent headers + external OAuth 2.1 with PKCE)
- Self-healing mechanisms with intelligent retry logic; 17 core tools forming foundation of 50+ tool ecosystem

Enterprise RAG System

- Full-stack RAG platform with hybrid retrieval: Dense + Sparse (SPLADE) + BM25 fusion via Reciprocal Rank Fusion with cross-encoder reranking
- Manages 10M+ document embeddings with low-latency retrieval using Qdrant; 95% query resolution rate
- Multi-provider LLM support (AWS Bedrock, Google Vertex AI, Anthropic, Groq); RBAC & multi-tenancy with JWT auth

Financial Insights Engine

- Real-time financial analytics with AI-driven forecasting (time-series models + LLM-powered narrative generation) and anomaly detection
- Automated financial spreading for loan underwriting and credit analysis — extracting, structuring, and standardizing financial statements for risk assessment
- AI-powered credit term evaluation with automated ratio analysis and borrower risk profiling

Accounts Payable & Receivable Automation

- Reimagined template-dependent PaddleOCR system (v1, 2022) into format-agnostic GenAI pipeline using LLMs + NER (v2, Jan 2024)
- End-to-end AP/AR workflow: invoice capture → PO/GRN matching → approval routing → payment processing
- Invoice cycle times reduced from 30 min to under 20 seconds with 95%+ accuracy; fraud detection layer for anomaly flagging

ML Bank Reconciliation Engine

- First production ML model inside the ERP — auto-matching bank transactions against ledger entries using XGBoost/scikit-learn
- Feature engineering on amounts, dates, descriptions (fuzzy matching), counterparty patterns; confidence-scored routing for human review

Invoice Processing v1 (Template-Based OCR)

- Built automated invoice extraction using PaddleOCR + rule-based template coordinate mapping for AP/AR workflows (~2022)
- Template-per-vendor approach — functional for known formats but not scalable; directly motivated the GenAI reimagination in v2 (Jan 2024)

ML Solutions — Client Projects (CodeFire)

- Product recommendation engine using collaborative filtering for e-commerce and retail clients
- Document similarity detection using TF-IDF vectorization and cosine similarity for automated document matching
- NLP tools for business process automation — tokenization, stemming, named entity extraction pipelines

HostBooks ERP Suite / Cloud360 - Chief Architect (Built from scratch at CodeFire, 2017)

Key Modules Delivered:

- **Accounting ERP (India & Global):** Financial compliance for GST, TDS, US GAAP, VAT with multi-entity consolidation and multi-currency operations
- **Neo Bank & US App Integrations:** Automated cash management, reconciliation, and tax handling through banking and fintech integrations
- **Automated Payroll & HR:** Comprehensive payroll and HRM serving businesses across India, US, and UAE markets
- **Fleet & Transport Management:** Predictive analytics and multi-modal transportation optimization
- **Advanced Inventory & Operations:** Real-time tracking, production planning, and warehouse management
- **Point-of-Sale Solutions:** Multi-mode POS systems for retail and F&B industries
- **HB Super App:** Unified LMS, OMS, DMS, and financial automation tools
- **Asset & Approval Management:** Asset tracking with secure, hierarchical workflows
- **Hb Neo App / Aahar POS:** Offline-first retail POS with multi-outlet sync and financial accounting

Additional Projects:

- Built enterprise reporting & analytics platform integrating Metabase and Superset for real-time business intelligence
- Designed secure remote printing infrastructure enabling document printing across distributed office networks
- Architected lead management system improving sales pipeline conversion and tracking efficiency

EDUCATION

Bachelor of Technology (B.Tech)

- Indian Institute of Technology (BHU), Varanasi
- June 2012 - May 2016

LANGUAGES

Hindi, English

ADDITIONAL INFORMATION

- Supported company growth from bootstrapped to INR 173 Cr valuation and INR 40.6 Cr annual revenue (67% YoY CAGR) across Series A & B funding rounds
- Expertise in enterprise-scale system design, distributed architectures, and high-availability platforms