

Pratik Verma

Senior Software Engineer — AI Engineering — Backend Systems

☎ +91 8770800807 ✉ pratikverma9691@gmail.com

🌐 linkedin.com/in/pratik-verma-1a0970192 🐙 github.com/elliott09alderson

Professional Summary

Senior Software Engineer with 4+ years of experience architecting and delivering production-grade AI systems, enterprise backend platforms, and cloud-native applications across Insurance, Healthcare, FinTech, Logistics, and EdTech domains.

Specialized in Agentic AI, Retrieval-Augmented Generation (RAG), Multi-Agent Systems, AI Voice Agents, LLM orchestration, and enterprise workflow automation. Experienced in owning end-to-end technical architecture from system design and backend engineering to cloud deployment and production operations using Python, FastAPI, Node.js, React, Docker, AWS, and CI/CD.

Proven track record of building scalable AI applications, optimizing backend performance, automating complex business workflows, and translating enterprise requirements into reliable production systems.

Technical Skills

Programming Languages: Python, JavaScript, TypeScript

Backend Engineering: FastAPI, Node.js, Express.js, REST APIs, Scalable API Design

Frontend Engineering: React.js, Tailwind CSS, Interactive Dashboards

AI Engineering: LLM Applications, Agentic AI, LangChain, LangGraph, Retrieval-Augmented Generation (RAG), Multi-Agent Systems, AI Agents, AI Orchestration, Prompt Engineering, OCR Pipelines, Human-in-the-Loop Workflows

AI Infrastructure & MCP Servers, Web MCP, N8N Automation, LiveKit Cloud, Exotel SIP

Automation: Integration, Enterprise Workflow Automation

Cloud & DevOps: AWS EC2, AWS S3, Google Cloud Platform, Docker, GitHub Actions, CI/CD, Production Deployments

Architecture & Concepts: System Design, Backend Architecture, Microservices, Monolithic Architecture, Distributed Systems, API Design, Monorepo, Turborepo

Security: RBAC, ABAC, CSP Headers, DRM Video Protection, Rate Limiting.

Key Achievements

- Architected and delivered production-grade enterprise AI systems supporting 300+ active users daily across real-world operational workflows.
- Designed scalable RAG and Agentic AI workflows for contextual enterprise knowledge retrieval, intelligent automation, and LLM-powered applications.
- Engineered AI Voice Agent infrastructure integrating real-time LLM orchestration, LiveKit Cloud, and SIP telephony for automated enterprise medical interviews.
- Achieved 99.8%+ OCR accuracy and reduced manual document-processing effort by 86% through AI-powered extraction and Human-in-the-Loop validation.
- Improved backend API response latency by 35% through architecture optimization, database tuning, and scalable service design.
- Owned end-to-end technical delivery across system design, backend engineering, AI integration, CI/CD, cloud deployment, and production operations.

Professional Experience

Nu10 Technology

Senior Software Engineer — July 2025 – Present

- Led technical architecture and development of production-grade Agentic AI platforms, owning key engineering decisions from system design and AI orchestration to backend implementation and production deployment.
- Designed scalable Retrieval-Augmented Generation (RAG) architectures for enterprise knowledge retrieval, enabling contextual LLM applications and reducing dependency on manual information discovery.

- Engineered AI Voice Agent workflows integrating real-time LLM orchestration, SIP telephony, automated reporting, and enterprise workflow automation.
- Designed Human-in-the-Loop AI validation workflows combining automated document intelligence with rule-based verification to improve operational accuracy and reliability.
- Built Docker-based deployment environments and automated CI/CD workflows using GitHub Actions, reducing manual deployment effort and improving release consistency.
- Optimized backend APIs, database interactions, and service architecture, reducing response latency by 35% under production workloads.
- Collaborated with product teams and enterprise stakeholders to translate complex business workflows into scalable AI and backend architectures.

AxonAio Technologies

Full Stack Developer — July 2023 – Feb 2025

- Designed and developed full-stack platforms for FinTech and EdTech products, contributing to backend architecture, API design, frontend engineering, security, and cloud deployment.
- Built scalable REST APIs and modular backend services using Node.js and Express.js to support transaction, administrative, and customer-facing workflows.
- Implemented DRM-based video protection and Content Security Policy (CSP) controls, strengthening application security against unauthorized content distribution, XSS, and related web vulnerabilities.
- Deployed and operated cloud-hosted application services using AWS EC2 and S3, supporting scalable application delivery and production workloads.
- Drove technical implementation across frontend, backend, infrastructure, and production deployments while collaborating with cross-functional stakeholders.

TrustlyPay Pvt Ltd

Junior Software Developer — Aug 2022 – July 2023

- Engineered backend services and React-based operational dashboards for FinTech payment workflows, supporting transaction processing and merchant administration.
- Developed PayIn, PayOut, checkout, merchant onboarding, and transaction-management workflows for an end-to-end digital payment platform.
- Implemented secure authentication and Role-Based Access Control (RBAC), strengthening authorization across merchant and administrative workflows.
- Optimized database queries and backend data-access patterns to improve API performance and reduce processing overhead for transaction-intensive operations.

Selected Projects

BGIL — Enterprise AI Medical Interview Platform

- Architected an end-to-end production AI Voice Agent platform for automated medical tele-interviews, handling 40–45 structured health questions per session.
- Engineered real-time voice infrastructure integrating LiveKit Cloud, Exotel SIP telephony, and AI orchestration to support reliable automated outbound medical interviews.
- Designed automated post-call workflows using N8N to transform interview data into structured CSV reports, PDF Medical Examination Reports (MER), and operational MIS documents.
- Automated report distribution and recording-link delivery to responsible stakeholders, reducing manual post-interview coordination and reporting effort.
- Engineered the production workflow to support 300+ active users daily across a 9-hour operational window with consistent report generation and delivery.
- Owned core technical architecture spanning AI voice orchestration, telephony integration, backend workflows, automation, reporting, and production deployment.

Education

B.Tech in Computer Science

CSVТУ

Graduated: 2021