



Santosh Kumar KMN

AI Architect / Fullstack Developer / ML Engineer / Data Scientist

✉ grey.ethics@gmail.com

☎ (+91) 974 193 4583

Profile

A multi-disciplinary technology professional with strong expertise across nearly every major domain of Computer Science including Full Stack Web Development, AI & ML Engineering, Generative AI, Data Science, Cloud Architecture, Automation, and Cybersecurity.

- With over 12+ years of hands-on experience in the tech industry, I bring a wealth of knowledge and expertise to every project.

Work Experience

AI Architect

Jan 2026 – Present

Zealogics

AI Architect & Fullstack Developer

- Worked in an Agile team where I led daily standups, gathered team updates, and owned modules through development and deployment.
- Interacted directly with clients to understand their requirements and implement new features that addressed their business needs.

Key Projects/Works:

KPMG Workbench (KPMG's Enterprise Application)

- Architected and enhanced KPMG Workbench, an enterprise AI platform providing secure access to Anthropic Claude, OpenAI GPT, and Google Gemini models through Azure APIM and Azure Front Door.
- Developed an AI-powered Fullstack Application which enabled users to securely test and consume LLM services, manage API credentials and monitor usage.
- Automated the process of identifying new releases and deprecated AI models, automatically notifying the company to onboard new models and to retire unsupported ones.
- Collaborated directly with clients and stakeholders to gather requirements, define AI solution architecture, and deliver production-ready features.

ZeaLearn (Fullstack AI-Training Application)

- Designed and developed an end-to-end Agentic AI-powered Document-to-Course platform using React, FastAPI, PostgreSQL, OpenAI GPT models, and cloud-native infrastructure.
- Built autonomous AI workflows capable of extracting knowledge from uploaded documents, generating structured learning paths, quizzes, assessments, summaries, and educational content.
- Implemented Retrieval-Augmented Generation (RAG), prompt engineering, structured outputs, and context-aware content generation to improve response quality and consistency.
- Led feature development, production support, bug resolution, GitHub task management, and deployment activities throughout the project lifecycle.

AI Architect

Jul 2025 – Dec 2025

Healthark Insights

AI Architect & Team Lead

- Designed end-to-end application architectures (code + cloud) and advised the company on best engineering practices.
- Led hiring and technical interviews, helping build a high-performing software development team aligned with company goals.

Key Projects/Works:

Curie (Fullstack Multi-Agent AI Application)

- Architected and developed Curie, a full-stack Agentic AI platform with a ChatGPT-style conversational interface, enabling users to interact with enterprise knowledge systems through natural language queries and AI-powered workflows.

- Built a Retrieval-Augmented Generation (RAG) framework integrating enterprise documents and knowledge repositories, delivering context-aware, grounded responses through semantic search and intelligent information retrieval.
- Implemented multi-tool-calling AI workflows, including an automated Resume-to-Job Description Matching Engine that analyzes resumes against job requirements, generates compatibility scores, and identifies skill gaps and candidate suitability.
- Developed a Document Intelligence and Template Automation Engine capable of consolidating multiple documents, extracting relevant information, and generating standardized output documents based on configurable business templates, significantly reducing manual document processing effort.

Intas (Fullstack Automated Web Application)

- Architected and developed INTAS, a full-stack healthcare and pharmaceutical management platform supporting multiple user roles including System Administrators, Regional Managers, Branch Managers, Doctors, and Patients through a unified web application.
- Designed and implemented automated workflow and notification systems for appointment scheduling, medication reminders, patient engagement, doctor communications, and business operations, significantly reducing manual coordination efforts.
- Built role-based dashboards, task management modules, performance monitoring features, and automated alert mechanisms that enabled managers to track activities, receive performance updates, and improve operational efficiency across branches and regions.
- Developed scalable backend services, business process automation, secure user management, and communication workflows integrating email and messaging capabilities to streamline interactions between patients, doctors, and organizational stakeholders.

Senior ML Engineer

Mar 2022 – Jun 2025

KSK Ventures

Artificial Intelligence & Machine Learning Engineer

- Managed all Python and AI/ML projects across the company, leading a 4-member team to ensure successful delivery of AI solutions.
- Independently completed multiple projects for various clients, delivering AI Software & Architect Solutions in the NLP & Gen AI domain.

Key Projects/Works:

- Developed a voice-to-voice ChatBot using Azure, GPT-4o, Twilio, Python for a restaurant which can take customer orders.
- Built an opensource 'Agentic App' to completely automate LinkedIn Profile Targeting, Personalized Messaging, and Replying.
- Developed an end-to-end 'Dialogue Segmentation App' using Open-Source LLM models from LM Studio and with Whisper.
- Trained and developed a 'Computer Vision Model' using TensorFlow, OpenCV, and Python to detect Driver Drowsiness.

Senior Software Developer

Jul 2019 – Jan 2022

E-Mech Technology Solutions

Senior Software Developer

- Played a key role in achieving multiple project milestones and contributing to the company's overall growth and success.
- Trained and mentored junior back-end developers, guiding them in best practices for designing and developing APIs.

Key Projects/Works:

- Built RESTful APIs using Node.js and Express.js to serve front-end applications for multiple businesses and clients.
- Collaborated with the Data Science team to develop and deploy a Product Recommendation Algorithms, boosting user engagement.
- Developed a Real-Time Live Chat feature using WebSockets, enabling one-on-one communication via a floating chat window.

Software Engineer

Jan 2017 – Apr 2019

Reno Plas Pvt. Ltd.

Software Development Engineer

- Developed and maintained software applications, ensuring clarity and ease of use for future developers.
- Worked closely with team members, senior developers and project managers to contribute to planning and decision-making.

Associate Software Developer

Sep 2014 – Jan 2017

Space Age Infotech

Associate Software Developer

- Actively participated in software testing and debugging processes, ensuring the delivery of high-quality products.
- Participated in code reviews with team members and gained feedback from senior developers.

Other Notable Projects

Artificial Intelligence & Machine Learning

- Developed a Face Detection Model using Python, TensorFlow, OpenCV and Albumentations

- Developed an Open Source NLP-Solutions Model by using Falcon by Huggingface and Langchain
- Developed a Deep-Fake Model Using InsightFace and OpenCV
- Developed a Language Translation Model using the Transformers Library and Gradio UI

Data Scientist

- Conducted Data-Driven Customer Behavior Analysis to enhance customer acquisition
- Implemented Employee Turnover Prediction and Retention Strategy
- Tableau Project: Created a Sales Comparison Dashboard
- Power BI Project: Created a Amazon Prime Videos Dashboard

Cloud Architect

- Deployed an Online Movie Watching Application on the Cloud (AWS Solutions Architect : Professional Level)
- Deployed Azure Protected Geo-Redundant Solution with Path-Based Routing (Azure AZ : 305 Level)
- Set-up and Monitoring of a WordPress Instance (AWS Solutions Architect : Associate Level)
- Implemented Azure IaaS (Azure AZ : 104 Level)

Technical Skills

- | | | | |
|-----------------|--------------|-------------------|-------------------|
| • Python | • Javascript | • Java | • HTML & CSS |
| • Azure | • AWS | • GCP | • IBM Cloud |
| • FastAPI | • Flask | • Node Js | • Express Js |
| • PostgreSQL | • MongoDB | • Pinecone | • Milvus |
| • Generative AI | • AI Agents | • NLP | • Computer Vision |
| • TensorFlow | • PyTorch | • Keras | • OpenCV |
| • MERN Stack | • REST APIs | • CRUD Operations | • Git |
| • Docker | • Kubernetes | • Nginx | • Apache Server |
| • Tableau | • PowerBI | • Excel | • Much More... |

Education

- | | |
|--|---------------------|
| Post Graduation in Cyber Securities - (Currently Pursuing)
MIT University | Jun 2025 – Jun 2026 |
| Post Graduation in Artificial Intelligence & Machine Learning
Caltech University | Apr 2023 – Apr 2024 |
| B.Sc Computer Science
IIMTE Autonomous | Apr 2006 – Apr 2009 |

Certifications

- Full Stack Development : MERN Stack**  — 12 Months Professional Course
- Cloud Architect : Masters Program**  — 11 Months Masters Program
- Data Scientist : Masters Program**  — 11 Months Masters Program