NIRANJAN KUMAR CHINTALA

SENIOR AI/ML CONSULTANT & ARCHITECT

Hyderabad, India | (+91) 9739979440 | niranjankumar.c@outlook.com |
LinkedIn: linkedin/niranjankumar-c **| Blog:** @medium/niranjankumarc

PROFESSIONAL SUMMARY

- As an AI Architect and Strategy & Analytics Senior Consultant with a decade of experience, I specialize in Artificial Intelligence, Machine Learning, Generative AI, and Agentic AI systems. I design and deliver enterprise-grade AI/ML architectures across diverse industries, with a strong focus on Industry 4.0 transformation. My expertise spans architecting and deploying bespoke AI solutions—integrating LLMs, Retrieval-Augmented Generation (RAG), multi-agent systems, and advanced NLP—ensuring alignment with business goals and industry standards.
- I have a proven record of leading high-performance data science teams, driving innovation, and managing the full AI lifecycle from strategy and design to deployment and AIOps/MLOps-based maintenance. Proficient in cloud-native AI platforms, scalable model pipelines, and advanced monitoring frameworks, I excel in solution architecture, project governance, Statements of Work (SOWs), and end-to-end delivery. Passionate about emerging AI technologies, I am committed to crafting future-ready, sustainable AI ecosystems that deliver measurable strategic value.

EDUCATION

Post Graduation in Data Science and Machine Learning | Praxis Business School
Machine Learning, Text Mining, Deep learning, Data Visualization and Big Data – Bangalore India, **2019**

Bachelor of Technology | Amrita Vishwa Vidyapeetham University
Coimbatore India, **2014**

SKILLS

Generative AI:	Prompt Engineering, LLM Fine-tuning, Agentic RAG, Multi-Agent Systems, AI Workflow Orchestration
Core Competencies:	Gen AI, Machine Learning, Deep Learning, AI System Architecture, AIOps/MLOps
Language Models:	OpenAI , LLaMA, BERT, Phi-3.5, Mixtral 7B
Cloud:	Google Cloud (Vertex AI Studio, AutoML, BigQuery, Vertex AI Pipelines & Model Monitoring) Azure (AI Foundry, AI Search, Azure Databricks) & AWS (Bedrock, SageMaker, OpenSearch & Kendra)
NLP:	Text Classification, Named Entity Recognition (NER), Sentiment Analysis, Semantic Search
MLOps:	Git, Docker, MLflow, Azure DevOps, Azure Functions & Databricks Model Serving
Frameworks:	Langchain, Langgraph, llama-index, Crew AI, Auto Gen, OpenAI Swarm

EXPERIENCE

Senior Consultant - Machine Learning & Artificial Intelligence| Deloitte US-India (USI)
April 2021 – Present, Hyderabad – India

Value Engineering: *Framework to Realize Business Value through ERP Enabled Transformations (Chemical & Energy) Client*

- Pioneered an AI-powered KPI forecasting and insight generation platform for a global chemical client, fusing driver-based time series models with Azure OpenAI GPT-4o to deliver KPI driver insights and actionable recommendations.
- Integrated a human-in-the-loop validation framework to ensure precision and strategic relevance. Built in the Azure AI Foundry ecosystem using Azure Functions, Azure Data Factory, and Azure AI Search, achieving 95%+ forecast accuracy, cutting manual reporting by 70%, and driving operational efficiencies worth $8M+ annually.
- Spearheaded the creation of an Agentic RAG-based KPI insights chatbot leveraging LangChain, multi-agent orchestration, and Azure AI Search for contextual retrieval. Connected to enterprise KPI datasets for real-time, domain-specific insights.

Omnichannel Marketing Plan: *Personalized diabetes marketing using data-driven segmentation, channel, & content optimization.*

- Developed AI-driven micro-segments of HCPs based on traits and prescription levels, reallocating interactions to boost engagement and increasing Soliqua & Toujeo prescriptions.

- Analyzed and optimized channel effectiveness, reallocating resources to top-performing channels, resulting in a 15% increase in HCP engagement rates.
- Implemented a data-driven engagement model that optimized channel interaction frequency per HCP, increasing engagement rates by 18% and reducing resource spend by 12% per planning cycle.

Health Care Knowledge Tool: *GenAI-powered tool to enable efficient search and attribute discovery*.

- Boosted data catalog search efficiency by 60-70% by leading the design, development, and deployment of an AI service using Azure OpenAI Studio for data entity search.
- Reduced manual metadata generation workload for Tables and Columns by 30-40% by implementing a generative AI tool with non-sensitive sample data and prompt engineering

Consent Form Analyzer: *Developed an OpenAI chatbot for informed consent analysis to extract patient rights and study details*

- Enhanced GPT 3.5 Turbo's performance on medical queries through fine-tuning with domain-specific data.
- Implemented an OpenAI GPT 3.5-powered chatbot that achieved 93% relevant answers for patient consent form inquiries.

AI Driven Regulatory Compliance: *Incorporated Generative AI to automate and streamline compliance processes.*

- Leveraged Generative AI to automate regulatory tasks such as standard creation, SOP updates, and gap analysis, achieving up to 75% increase in process efficiency and reducing manual review time by up to 50%.
- Achieved a 50% increase in efficiency by using Generative AI to identify stakeholders in regulatory documents, unlocking significant financial benefits.

Plant of the Future: *Computer vision driven collision avoidance system for Real-time process and safety compliance monitoring*

- Led the effort in scaling, deploying and sustaining computer vision solution across multiple cameras to identify collision risks and non-compliance with safety procedures. Integrated real-time safety alerts & monitoring through a centralized Dashboard.
- Spearheaded the complex transition from nine legacy models to a single YOLO model, improving violation detection rates by 30-50%, reducing re-training time by 50-60%, and contributing to a 15-20% reduction in safety incidents.

Senior Consultant – Data Science| Allstate India Private Ltd
October 2019 – April 2021, Bangalore - India

- Developed the fraud prevention model to identify and report potential fraudulent medical claims from first-party motor vehicle accidents using XGBoost, resulting in 85% automation of customer claim assessments.
- Established agile development practices within the team, including sprint planning, daily stand-ups, and code reviews, to ensure efficient project delivery and continuous improvement.
- Implemented NLP techniques to analyse unstructured text data from medical records and insurance claims. This helped identify potential fraudulent patterns in medical claim descriptions, leading to a 9% reduction in fraudulent claims payouts.
- Created an automated clustering framework to automate the process of building clusters from the client's financial data. Framework uses silhouette score to automatically recommend the ideal number of clusters.
- Developed and implemented an automated reporting dashboard using Tableau, which increased the efficiency of a new marketing campaign by 30% through real-time tracking and analysis of key performance indicators

Risk Analytics Intern| HSBC India
April 2019 – October 2019, Bangalore - India

- Developed credit risk models that achieved a 90% accuracy in predicting the expected credit losses for retail banking portfolios.
- Constructed a credit limit optimization model using a decision tree, resulting in a 15% increase in approved credit limits for customers.

Senior Software Engineer| Infosys Private Limited
June 2014 – June 2018, Mysore - India

- Built an end-to-end instance segmentation mask-RCNN model that achieved an 87% accuracy in identifying cracks on concrete surfaces. Deployed the trained model as a web service using Flask on an AWS spot instance.
- Developed a clustering model to segment bank customers based on purchasing habits, monthly average purchases, and credit quality. Evaluated cluster performance with Silhouette and inertia scores and visualized insights for a 4-cluster solution using Tableau and Python.
- Constructed a predictive maintenance model to predict time to failure, resulting in a 15% reduction in flight delays and a 10% increase in operational efficiency.

CERTIFICATIONS

- Google Cloud Certified Machine Learning Engineer
- Google Cloud Certified Cloud Digital Leader
- Deep Learning with Python & Pytorch
- Certified Big Data Analyst – Hadoop & Spark