

ASHWINI SHARMA

812 W Royal Ln, Irving, TX • +1 312-885-7424 • ashwini.gsharma@gmail.com
Architect GenAI Solutions | AI Enthusiast | Analytics Specialist with Global Experience

SUMMARY

Experience: Over 14 years in Information and Communication Technology solutions, currently implementing enterprise knowledge layer for AI Agents utilizing cloud infrastructure, defining art of possible with GenAI, advancing & implementing skills in Generative AI (LLMs/SLMs) & Agentic/RAG

AI Expertise: Unstructured data handling, Retrieval-Augmented Generation (RAG), Agentic AI, Developing human-understandable rules from CNNs, building artificial neural networks with custom loss and gradient calculations, Multiple Chunking Strategies, Embedding Models, Vector Stores, Enrichment (NLP & LLM based), Contextual Retrieval, Re-Ranking, Code Generation, Synthetic Data Generation, LLM/SLM Fine-tuning, Knowledge Graph, Code Indexing (LLM, AST, Merkle tree), GraphDB, LightRAG, Graph RAG, Observability, Performance Metrics, Prompt and Context Engineering, AI Architectures (North Star), AI vendor evaluation, Tool Selection, Proof of Concepts, Deep understanding of ML algorithms from statistical and mathematical perspectives, Tech Strategy & Cloud Infra

Data Analysis & Visualization: Extensive experience with descriptive and predictive analytics, producing actionable insights from enormous volumes of data. Proficient with QlikView, QlikSense, Tableau, Power BI, Grafana, Kibana, Matplotlib, and Altair

Leadership: Led large teams in consulting, implementation, operations, and post-deployment support, ensuring successful project outcomes

Global Projects: Onsite work experience across multiple countries including USA, Sweden, Norway, Romania, Morocco, South Africa, UAE, Iran, China and others across Middle East & North Africa regions

WORK EXPERIENCE

Director – GenAI Architect, Tredence

Oct 2025 - Present

- *Serving in a critical leadership role* defining enterprise North Star architectures for Generative AI and Agentic AI systems, establishing scalable design patterns for multi-agent orchestration, reasoning workflows, RAG-based knowledge systems, and secure enterprise integration.
- *Leading high-impact GenAI initiatives* from executive discovery through production delivery, advising C-level stakeholders on AI strategy while translating cutting-edge GenAI capabilities into organization-wide platforms adopted beyond pilot and PoC stages.
- *Recognized as a subject-matter expert in enterprise GenAI architecture*, driving the evaluation, selection, and large-scale deployment of cloud and open-source GenAI platforms (Azure OpenAI, Google Vertex AI & Gemini Enterprise, Databricks AgentBricks), with emphasis on governance, cost-efficient inference, and LLMops observability.
- *Advancing the state of practice in Agentic AI* by mentoring senior architects, creating reusable GenAI accelerators, and shaping firm-wide standards for responsible AI, autonomous agents, and enterprise-grade GenAI adoption.

Enterprise Architect - GenAI CoE, Tech Strategy, Verizon (End Client)**Nov 2024 - Oct 2025**

- *GenAI Solutions*: Defining North Star architectures for enterprise, designing and deploying enterprise-grade GenAI solutions, evolving from traditional approaches to transformers, RAG, and Agentic AI for enterprise applications.
- *Cloud Ready*: Designing architecture for deployment of AI solutions over cloud, specifically GCP and AWS.
- *Enterprise Knowledge Layer*: Researching, architecting, and developing knowledge layer to ingest code repos to establish inter-repo and system-wide impacts to automate SDLC.
- *RAG Pipeline Development*: Leading enterprise-wide unstructured data handling for RAG-based applications, overseeing ingestion, preprocessing, chunking, enrichment, metadata extraction, topic modeling, embedding, prompt & context engineering and indexing in vector stores.
- *Advanced Retrieval Mechanisms*: R&D on multi-query retrieval strategies, including HyDE-based query generation, intent extraction, keyword + semantic retrieval, re-ranking, and response generation using LLMs.
- *LLM Optimization & Monitoring*: Synthetic data generation for fine-tuning, defining evaluation metrics for ingestion and retrieval processes (e.g., chunk attribution and utilization), and monitoring model performance with LLMOps best practices.
- *Hands-on with Cutting-Edge Tools*: Working in tech strategy, evaluating AI vendors and AI tools, provides hands-on experience over tools like LightRAG, LangFlow, Nvidia Nemo, NeuralSeek, Accenture Aspire, and Uniphore X-Stream for RAG-based AI solutions. Utilized LLMOps tools like Arize for experimentation tracking, LLM-as-a-Judge evaluations, and traces analysis.

AI/ML Consultant - Freelancing (Pursuing Master's)**Nov 2022 - Oct 2024**

AI Consulting: While pursuing Master's in AI, helped one African Telco achieve their goal for moving towards ML base approach for Fraud Detection from traditional static thresholds. With open source stack, created automated pipelines to read, filter, pre-process, enhance and make data ready to be consumed by ML algorithms. Provided interactive GUI for business teams to see real time impact of specific threshold on precision, recall and number of alarms.

GenAI Dev/Deployment: Created North start architecture for an AI start-up who was struggling with scaling due to just prompt based approach. Not only proposed, but developed and deployed the architecture which helped them scale 10x with reduced hallucinations.

Load estimation and hardware sizing: Calculated GPUs estimations for Fine-tuning & Inference. Finetuned 130Million to a few billion parameter models (VGG16, BERT large, Nomic, Llama 3b).

Technical Manager - Risk Assurance and Analytics, Subex Limited**Apr 2019 - Sep 2022**

As a Technical Manager, I led a team of engineers and consultants to deliver real-time analytics on billions of daily records, aligning data solutions with business needs. I ensured 99.9% system uptime (revenue, fraud & risk management), deployed ML models that averted \$5M in fraud losses, and automated workflows that suspended over 200K fraudulent subscribers—significantly improving efficiency and ROI. Extensively worked on python and machine learning to deliver smart cloud (AWS) native solutions.

Manager - Internal Data Systems, Bharti Airtel**May 2018 - Apr 2019**

I led revenue & fraud management initiatives by conducting PoCs and implementing real-time roaming fraud detection solutions. I improved data integrity and governance, leveraged analytics to uncover fraud patterns, and deployed SAS-based controls that prevented \$400K in losses while enhancing access controls over sensitive customer data. Python and Java development, ventured into AI tech.

Consultant – RA (Data Strategy), Notrika Radman Communications**Nov 2017 - May 2018**

As a Consultant for Revenue Assurance (Data Quality & Governance), I defined the scope and strategy for Revenue & Business Assurance by identifying key data sources and requirements. I developed a master data management framework for subscriber data, delivered KPIs to quantify leakage, and improved data quality by identifying anomalies and reducing false positives—resulting in more accurate revenue reporting

Senior Engineer - Revenue and Fraud Ops, Etisalat UAE**Aug 2017 - Nov 2017**

As Senior Engineer – Revenue and Fraud Ops at Etisalat UAE, I conducted in-depth data analysis to uncover new leakage and fraud scenarios, strengthening detection mechanisms. I implemented system controls, expanded data coverage, and established 24x7 monitoring dashboards—enabling predictive maintenance and improving operational responsiveness.

Senior Engineer - Implementation & Data Analytics, Subex Limited**Sep 2014 - Aug 2017**

As Senior Engineer – Analytics, I led the global telecom engagements for development of telecom analytics strategies focused on fraud detection, customer churn, and network optimization. I designed and implemented key metrics to identify high-revenue and high-risk zones, enabling rapid fraud detection through real-time sim-box controls and automated deactivations at signaling level. Statistical analysis to auto threshold suggestion for static rules, was my first exposure to learning algorithms, which gave customers advantages over manual work and enhanced Subex solutions.

Business Analyst- Mediation, IBM**Feb 2014 - Jun 2014**

As a Mediation analyst, I developed and implemented production rules for Mediation system to efficiently process billions of call detail records, ensuring near real-time delivery to multiple destination systems. This approach enhanced data quality and governance while providing a fail-safe mechanism critical to maintaining reliable data operations.

Executive, Tishitu Research & Consultancy**Jun 2013 - Jan 2014**

Embedded and robotics project design and development for Department of Science & Tech.

Engineer - Deployment & Data Migration, Huawei Technologies**Oct 2011 - May 2013**

As a delivery engineer, I led data migration initiatives for product version upgrades, managing metadata and data catalogues to ensure seamless transitions. I deployed and configured BSS stack products including rating, charging, billing, and mediation systems. Also got exposure to implement smart memory databases (SMDB) which enabled real-time call rating & charging for telecom customers, which was SOTA of the time.

PUBLICATION**Human Interpretable Rule Generation from Convolutional Neural Networks Using RICE (Rotation Invariant Contour Extraction)****Aug 2024**

Publisher: University of North Texas

Type: Thesis

Link: <https://digital.library.unt.edu/ark:/67531/metadc2356140/>

Keywords: AI, eXplainable Artificial Intelligence, Convolutional Neural Network, Feature Maps

Generating Human-Interpretable Rules from Convolutional Neural Networks**Mar 2025**

Publisher: MDPI ([Advances in Explainable Artificial Intelligence, 2nd Edition](#))

Type: Article (Research Paper)

Link: <https://doi.org/10.3390/info16030230>

Keywords: AI, eXplainable Artificial Intelligence, Convolutional Neural Network, Feature Maps

EDUCATION

Master of Science in Artificial Intelligence - Thesis Track

Jan 2023 - May 2024

University of North Texas - USA

- Thesis on Explainable Deep Neural Networks, proposed "*RICE: Rotation Invariant Contour Extraction*", producing human interpretable rules explaining global decision-making of CNNs.
- Implemented fully connected neural networks for binary and multi-class classification in text datasets from scratch with custom penalties and loss functions.
- Conducted literature reviews on 50+ top AI/ML research papers, contributing insights to the thesis.

Bachelor of Technology in Computer Science & Engineering

Aug 2007 - Jul 2011

Technical University Rajasthan - India

- Relevant coursework in Computer Science

ADDITIONAL INFORMATION

- **Technical Skills:** Data Modelling, Data analytics, SQL, Oracle architecture, Hive, Shell, Python, Debugging, TensorFlow, Computer vision, Exploratory Data Analysis (EDA), Visualization, Supervised and Unsupervised learning, Scikit learn, Hyperparameter tuning, NLP, Statistical Analysis and Hypothesis Testing, Feature Engineering, Feature Selection, eXplainable AI, Ensemble Learning, Data Quality Assurance, Database design, Governance strategies, Risk and Gap analysis, Root cause analysis, Cloud Infra Architectures (AWS/GCP/GKE), JIRA, Agile, PL/SQL, Graph DBs, PostgreSQL, Data Engineering
- **AI Tools:** DataStax LangFlow (RAG+Agentic), Vector stores: DataStax Astra Vector Store, Elastic Vector DB, Nono DB, NetworkX, NeuralSeek (RAG & Agentic), LLMOps: Arize, Galileo, RAG: Uniphore X-Stream, Accenture Aspire, Notebook based Retrieval Pipelines, Nvidia NeMo services including curator, customizer and retriever(embeddings and re-rankers), Docling, Unstructured IO, NV-Ingest, LightRAG, GraphRAG, Knowledge Graph, Knowledge Base, Models: Gemini Pro & Flash, OpenAI, Anthropic, Llama, Cohere