

WORK EXPERIENCE

- Senior AI Engineer:** *Analytical Intelligence International, LLC* *Jan 2025 – Present*
- Orchestrated autonomous, multi-agent enterprise architectures leveraging frontier models (Gemini, GPT, Claude). Engineered deep agents and proprietary A2A communication protocols using AutoGen and the Google Agent Development Kit (GADK), and the Model Context Protocol (MCP). Standardized complex tool calling across models to dynamically route complex reasoning tasks, driving a 40% reduction in manual processing time and delivering transformative operational cost savings at enterprise scale.
 - Pioneered self-evolving cognitive architectures that continuously optimize policy reasoning through automated Langfuse/Opik observability loops. Integrated these feedback mechanisms with hybrid-search RAG (Pinecone, Weaviate), enabling agents to dynamically self-correct memory states, boosting retrieval precision by 35% and virtually eliminating hallucinations in mission-critical queries.
 - Architected zero-downtime MLOps and forward-deployment pipelines (Kubernetes, Docker, Azure DevOps, Gitlab, Terraform, Crossplane) to securely push autonomous AI assets into strictly isolated, on-premise client environments. Established rigorous, automated model validation gates that slashed deployment cycles by 60% while guaranteeing 99.9% system reliability at enterprise scale.
- ML Data Engineer:** *Shodh AI Pvt Ltd* *July 2024 - Jan 2025*
- Architected a petabyte-scale multi modal data lake using Apache Spark and Delta Lake, optimizing ingestion for diverse data types (text, image, audio). Leveraged Apache Parquet for storage optimization and Z-Ordering for query performance, resulting in a 45% reduction in processing latency while ensuring strict ACID compliance and data quality via automated Great Expectations validation suites.
 - Engineered robust MLOps orchestration using Kubeflow and Airflow, automating the full life cycle from feature engineering to model serving. Integrated DVC (Data Version Control) and Weights & Biases for experiment tracking, and utilized A/B testing and Canary deployments to reduce deployment time by 70%. These safeguards increased production reliability by 40% through proactive drift detection and automated retraining triggers.
 - Developed high-performance distributed training frameworks utilizing Ray and PyTorch FSDP (Fully Sharded Data Parallel) to optimize LLM fine-tuning on multi-node GPU clusters. Implemented memory-efficient techniques such as LoRA (Low-Rank Adaptation) and DeepSpeed ZeRO-3 optimizations, achieving a 3x speedup in training throughput and a 25% reduction in compute costs without sacrificing model accuracy.
- Data Engineer:** *Pravaig Dynamics Pvt Ltd* *Feb 2023 - July 2024*
- Oversaw ML projects in the EV R&D division by constructing reliable pipelines for Battery Management System (BMS) and Predictive Maintenance using Python, XGBoost, and Scikit-Learn. Developed automated feature engineering for State of Charge (SOC) and State of Health (SOH) estimation, improving performance analytics accuracy by 40% and reducing the R&D testing cycle from weeks to days.
 - Designed and developed advanced Sensor Fusion models using PyTorch and TensorFlow to synthesize high-frequency data from CAN-BUS, Radar, GPS, and LiDAR/Cameras. Implemented Kalman Filters and Convolutional Neural Networks (CNNs) to process 50GB+ of daily telemetry, achieving 99.5% accuracy in object detection and vehicle localization for autonomous driving simulations.
 - Created high-throughput data streaming pipelines using Apache Kafka and MQTT, seamlessly integrated with the ROS2 (Robot Operating System) framework for edge-to-cloud synchronization. Optimized data serialization via Protobuf, achieving sub-100ms end-to-end latency for over 500 concurrent sensor streams, enabling real-time safety interventions and remote vehicle diagnostics.
- Data Engineer:** *Vumonic Data Labs Pvt Ltd* *Oct 2022 - Jan 2023*
- Designed and deployed high-volume data pipelines on Azure Databricks and Snowflake, processing over 500GB of daily telemetry with 99.9% uptime. Implemented Dead Letter Queues (DLQ) for automated error recovery and integrated Datadog/Prometheus for real-time alerting, ensuring data consistency and eliminating data loss during peak burst loads.
 - Streamlined complex data orchestration by building dynamic Apache Airflow DAGs and optimized PySpark jobs utilizing broadcast joins and predicate pushdown. This automation improved processing efficiency by 35% and eliminated 20+ hours of weekly manual intervention, enabling seamless, low-latency integration between Amazon S3 and downstream analytics platforms.
- Big Data Engineer:** *Tata Consultancy Services* *Apr 2022 - Oct 2022*
- Optimized large-scale Hadoop (HDFS/YARN) and Spark clusters by fine-tuning executor memory, core allocation, and parallelism settings. Implemented Spark Adaptive Query Execution (AQE) and handled data skew using salt-key techniques, resulting in a 30% increase in processing efficiency and a significant reduction in job failures for mission-critical batch workloads.
 - Developed a comprehensive data quality framework using AWS Deequ and Great Expectations, integrating automated schema validation and statistical anomaly detection into the ingestion layer. Developed custom circuit breakers and data lineage tracking, which reduced ingestion errors by 45% and ensured high-fidelity datasets for mission-critical ML model training and feature stores.

PROJECTS

Dog Emotion Recognition Using Deep Learning (AIP Published)

Developed a specialized emotion classification system using three distinct CNN architectures and a proposed MBCC-CNN model, achieving superior accuracy in emotional recognition tasks that resulted in the successful publication of findings in the American Institute of Physics (AIP) journal. [Link](#)

Product Photography with Generative AI

Designed an automated background generation tool using FastSAM and Stable Diffusion, establishing an efficient pipeline for precise image segmentation that enabled scalable, high-quality product photography generation for e-commerce applications. [Link](#)

Math Solver with Retrieval-Augmented Generation (RAG)

Created a modular mathematical problem solver by implementing Retrieval-Augmented Generation (RAG) with the Pythia LLM, significantly improving solution accuracy and reliability through advanced context-aware retrieval mechanisms deployed for complex problem-solving. [Link](#)

KEY SKILLS

LLMs & Generative AI: OpenAI GPT/ChatGPT, Claude, Gemini, LLAMA, Mistral, Deepseek, RAG, Fine-tuning, PEFT, LORA, Prompt Engineering Chain-of-Thought, Function Calling, Embeddings, MCP, A2A, Contextual Engineering.

AI Frameworks & Tools: LangChain, LlamaIndex, Hugging Face, Transformers, Ollama, AutoGen, CrewAI, Semantic Kernel, Vector DBs, Stable Diffusion.

ML/DL & Classical Modeling: Python, PyTorch, TensorFlow, Lightning, Scikit-learn, XGBoost, OpenCV, NumPy, Pandas, TensorRT, Quantization, C++, Java.

AI Agents & Orchestration: Multi-Agent Systems, Tool Use, ReAct, AutoGen, CrewAI, LangChain, Vector DBs.

Cloud & MLOps: AWS (SageMaker, EC2, S3, Lambda, EKS), Azure (DevOps, Resources, AI Foundry, AKS, etc), Docker, Kubernetes, MLflow, Jenkins, Circle CI, Airflow, CI/CD, A/B Testing.

Big Data & Data Engineering: Hadoop, Spark, Kafka, Hive, ROS2, Distributed Computing, PostgreSQL, MongoDB, Redis, Elasticsearch, FastAPI, ETL Pipelines.

Data & APIs: Postgres, Clickhouse, Aurora RDS, FastAPI, REST/GraphQL, SQL, MongoDB, Redis, Elasticsearch, Data Pipelines, ETL, Real-time Processing, API Security, AWS WAF

EDUCATION

MSc, Big Data Analytics

St. Joseph's University, Bengaluru

2020 - 2022

BCA, Computer Applications

Seshadripuram College, Bengaluru

2016 - 2019