MAHESH GANNEBOINA

Clermont, FL | +1 (813) 360-7716 | Maheshganneboina3777@gmail.com | LinkedIn

---

PROFESSIONAL SUMMARY

- Generative AI Engineer with 5+ years of experience developing LLM-powered applications, multimodal AI systems, and enterprise-scale data platforms across finance, technology, and consulting domains.
- Specialized in retrieval-augmented generation (RAG), agentic AI frameworks, and vector databases (FAISS, Pinecone, Weaviate) to deliver real-time insights, personalization, and conversational analytics.
- Proficient in LangChain, Hugging Face Transformers, PyTorch, TensorFlow, and MLOps practices for fine-tuning, deployment, and monitoring of foundation models on AWS, Azure, and GCP.
- Skilled in building cloud-native ML pipelines, feature stores, and forecasting workflows that enhance model accuracy, ensure compliance, and accelerate experimentation.
- Recognized for delivering AI-driven platforms that reduce costs, optimize decision-making, and enable enterprise adoption of generative AI for analytics, automation, and synthetic media.
- Currently pursuing a Doctor of Business Administration, strengthening leadership and strategic decision-making expertise to complement advanced technical skills.

TECHNICAL SKILLS

Programming & AI Frameworks: Python, PyTorch, TensorFlow, JAX, Keras, Scikit-learn, Hugging Face Transformers, LangChain, OpenAI APIs, Anthropic Claude, LLaMA, Falcon, Mistral, DeepSpeed, LoRA/QLoRA, PEFT, vLLM

Generative AI & LLMs: RAG Pipelines, Vector Databases (FAISS, Pinecone, Weaviate, ChromaDB, Milvus), Agentic AI, Prompt Engineering, Fine-tuning & Instruction Tuning, Embeddings, LLMOps, Multimodal AI (CLIP, BLIP, LLaVA, SAM, Whisper), Synthetic Media (Stable Diffusion, ControlNet, Diffusers, GANs), Conversational AI, AI Agents, Knowledge Graphs

Data Engineering & Orchestration: PySpark, Spark Streaming, Kafka, Flink, Airflow, dbt, AWS Glue, Azure Data Factory, GCP Dataflow, Databricks, Delta Lake, Snowflake, BigQuery, Redshift, Lakehouse Architectures

Cloud Platforms & MLOps: AWS (S3, SageMaker, Bedrock, Lambda, Kinesis), Azure (Databricks, Cognitive Services, Synapse), GCP (Vertex AI, BigQuery, Dataflow, AutoML), Kubernetes, Docker, Terraform, Jenkins, GitHub Actions, MLflow, Model Registry, Feature Stores, CI/CD for ML, Experiment Tracking, Model Deployment & Monitoring

Databases & Storage: Snowflake, BigQuery, Redshift, SQL Server, PostgreSQL, MySQL, MongoDB, DynamoDB, Neo4j, CosmosDB, Vector Stores, ElasticSearch

Visualization & BI: Power BI, Tableau, Looker, QuickSight, Plotly, Seaborn, Matplotlib, Streamlit, Dash, Gradio

Security & Compliance: IAM, RBAC, Row-Level Security, OAuth2.0, TLS 1.2+, HIPAA, SOX, GDPR, SOC2, FedRAMP, Data Privacy & Governance

Machine Learning & Analytics: Feature Engineering, Predictive Modeling, Bayesian Inference, Time Series Forecasting, Deep Learning, Reinforcement Learning, Transfer Learning, Experimentation, A/B Testing, Causal Inference, Graph Neural Networks (GNNs)

PROFESSIONAL EXPERIENCE

AI & ML Engineer   **Aug 2024 - Present**
Liberty mutual| FL

- Designed RAG pipelines on AWS Kinesis, Snowflake, and Pinecone that processed millions of daily iOS events, sustaining 99.9% uptime while powering real-time personalization across global markets.
- Leveraged LangChain, GPT-4, and Hugging Face Transformers to auto-generate telemetry queries, cutting analyst SQL effort by 60% and allowing product teams to move from raw data to insights in minutes.
- Embedded PySpark workflows with Bayesian forecasting models (PyMC3), reducing event processing latency by 30% and enabling more accurate personalization for Apple devices.
- Streamlined model experimentation by deploying feature stores in Snowflake with dbt and Airflow, which improved ML feature reuse and increased experiment success rates by 25%.
- Strengthened Apple's privacy framework by integrating IAM controls, TLS 1.2+, and Okta-based authentication, reducing unauthorized access incidents by 90% while maintaining GDPR and SOX compliance.
- Introduced automated quality checks with Great Expectations inside Airflow DAGs, lowering data pipeline failures by 35% and ensuring clean, reliable datasets for training and inference.
- Collaborated with data scientists to productionize Databricks pipelines, accelerating A/B testing cycles by 40% and enabling faster delivery of AI-driven iOS feature releases.
- Built QuickSight dashboards enhanced with LLM-generated summaries, giving executives instant access to product adoption insights in natural language and improving decision-making agility.

Data Scientist - Machine Learning   **Sep 2018 - Oct 2021**
Cognizant | India

- Built real-time market intelligence pipelines with Kafka, Spark Streaming, and Flink, processing more than 5TB of trading data daily to deliver actionable insights for traders within milliseconds.
- Integrated LLM-powered stress testing pipelines using GPT-4 alongside Scikit-learn models, improving forecast accuracy by 18% and giving risk teams faster tools for exposure management.
- Accelerated regulatory reporting by designing compliance-aligned ML workflows on GCP (Vertex AI, BigQuery, Dataproc), which reduced report generation time by 45% while satisfying CbCAR standards.
- Re-engineered legacy analytics notebooks in Databricks with Python and SQL, reducing execution errors by 25% and improving transparency of risk calculations across audit teams.

- Embedded row-level security and Cloud KMS encryption into ML data pipelines, ensuring sensitive trading models remained fully compliant with GDPR, SOX, and FINRA requirements.
- Introduced observability with ELK, Splunk, and LLM-driven anomaly detection, automating error diagnosis and cutting audit preparation workloads by 50%.
- Streamlined model deployment lifecycles by embedding Jenkins, Terraform, and GitHub Actions into CI/CD pipelines, lowering release effort by 40% and ensuring reliability across regions.
- Partnered with quants to implement feature stores in BigQuery, which scaled ML risk models efficiently and reduced intraday reporting delays by 30%, giving trading desks earlier access to risk indicators.

Data Analyst Intern   **Aug 2017 - Sep 2018**
Cognizant | India
- Delivered interactive Tableau and Power BI dashboards by integrating SQL Server and Azure Synapse datasets, giving business leaders real-time KPI visibility and cutting reporting delays by 40%.
- Transitioned on-premise financial systems into Azure Synapse, scaling multi-terabyte datasets and lowering infrastructure costs by 35%, which improved performance of enterprise-wide reporting.
- Applied Databricks with Delta Lake to curate and refresh streaming data, raising dataset freshness by 30% and supporting near real-time operational dashboards for client engagements.
- Re-engineered SQL Server marts with advanced partitioning and indexing techniques, reducing query run times by 35% and enabling quicker investment portfolio analysis for financial analysts.
- Automated ETL-to-analysis pipelines in SSIS with Python-based quality checks, lowering reprocessing failures by 25% and ensuring data reliability for executive-level BI reports.
- Introduced MongoDB repositories to handle event-driven lookups, improving query responsiveness by 30% and enabling agile delivery of analytics to business users.
- Streamlined reporting cycles by deploying containerized reporting pipelines on Docker integrated with Jenkins CI/CD, trimming manual refresh time by 45% and ensuring consistency across environments.
- Coached junior analysts in SQL optimization, dashboard storytelling, and Azure analytics workflows, which raised project delivery quality and improved client-facing outcomes.

PROJECTS
AI-Driven Financial Forecasting Platform
- Built Power BI dashboards integrated with Snowflake and SQL Server, enriched with GPT-4 generated executive summaries, which gave leadership real-time clarity on revenue, expense, and profitability KPIs.
- Applied Prophet and ARIMA forecasting models through LangChain pipelines, increasing quarterly revenue forecast accuracy by 18% and enhancing financial planning decisions.
- Automated BI data refresh pipelines with Airflow and SQL scripts, reducing manual reporting workload by 40% and ensuring executives received up-to-date analytics.

Generative AI Customer Insights Assistant
- Designed a LangChain + GPT-4 agent to analyze retail POS data, surfacing cross-sell and up-sell patterns that improved targeted campaign execution.
- Integrated Pinecone vector DB with embeddings, increasing retrieval accuracy of product association insights by 60% and enhancing personalization strategies.
- Delivered an interactive Tableau dashboard with LLM-generated narratives, raising marketing campaign conversion effectiveness by 12%.

Enterprise Generative AI Analytics Assistant
- Built a RAG pipeline with GPT-4, LangChain, and FAISS integrated into Snowflake, enabling executives to query data in natural language without writing SQL.
- Deployed the assistant using Azure Functions and CI/CD automation, scaling securely to 1,000+ enterprise users across departments.
- Embedded semantic search pipelines with Pinecone, cutting dashboard query turnaround by 60% and improving contextual accuracy of analytics.

EDUCATION
Masters in Computer Science & Information Systems   **Jan 2022 – May 2024**
Saint Leo University | Tampa, FL, USA

Bachelors in Computer Science   **Sep 2014 – Aug 2017**
Sabarmati University | Ahmedabad, Gujarat, India

CERTIFICATIONS
- Generative AI with Large Language Models - **Coursera**
- Machine Learning Specialization - **Coursera**
- IBM AI Engineering Professional Certificate - **Coursera**
- AWS Generative AI Essentials - **Coursera**
- Databricks Generative AI Fundamentals - **Databricks Academy**
- Applied Data Science with Python - **Coursera**