ADARSH GORREMUCHU

Tysons, VA • (703) 627-0235 • adarshg15102000@gmail.com • linkedin.com/in/adarsh-gorremuchu

PROFESSIONAL SUMMARY

AI/ML Engineer with an M.S. in Data Analytics Engineering and 4+ years of experience designing, fine-tuning, and deploying LLMs, generative AI, and AI Agents on AWS (Bedrock, SageMaker, Lambda, S3, ECS, EC2). Expert in prompt engineering, Python, PyTorch, TensorFlow, scikit-learn, and scalable MLOps pipelines. Proven track record translating business requirements into production ML models and APIs — with strong command of data privacy, containerization (Docker), and cloud security best practices.

EXPERIENCE

AI/ML Engineer – LLM & AWS Integration Jun 2025 – Present

Ampcus Inc Chantilly, VA

- Designed, developed, and fine-tuned LLMs and generative AI models (OpenAI, Anthropic, Cohere, Hugging Face) for domain-specific use cases; optimized prompt engineering strategies to improve model performance, relevance, and output quality across production workflows.
- Integrated and deployed ML models via AWS Bedrock, SageMaker, Lambda, S3, ECS, and EC2; built scalable data pipelines and RESTful APIs in Python to support end-to-end ML workflows — reducing manual ops effort by 70–85%.
- Developed and deployed AI Agents for business problems using LangChain and Semantic Kernel; architected multi-agent orchestration with classification, planning, and retrieval sub-agents — accelerating throughput by 45% via parallel execution.
- Applied MLOps practices for deployment, monitoring, and iteration: configured CI/CD pipelines, trace logging, and CloudWatch observability; enforced data privacy and security via IAM roles, KMS encryption, SCP policies, and VPC networking.
- Collaborated with business stakeholders to translate requirements into ML models and prompt-based solutions; communicated technical concepts to non-technical audiences through Power BI dashboards and executive presentations.
- Leveraged Docker containerization and Terraform IaC for consistent, reproducible production deployments; reduced rework from incorrect model actions by 30% in compliance-sensitive workflows through evaluation guardrails and access controls.

Cloud Data Engineer & ML Developer May 2024 – Nov 2024

Community Informatics Lab (George Mason University) Fairfax, VA

- Built LLM-assisted analytics pipelines using RAG and prompt engineering in Python; trained and evaluated ML models (PyTorch, TensorFlow, scikit-learn) with statistical and quantitative modeling methodologies — improving model quality by 35%.
- Designed scalable data pipelines and APIs on AWS (Lambda, S3, Step Functions) and Azure; automated MLOps with Airflow and Databricks — accelerating runtimes by 55% while sustaining 99.9% reliability across 20+ concurrent projects.
- Developed ML solutions using Python, PySpark, Pandas, and SQL on Snowflake and BigQuery; applied AI/statistical modeling for forecasting and anomaly detection; monitored model performance and iterated based on feedback and metrics.

Data Engineer & AI Developer Dec 2020 – Dec 2022

TriSX Global India Pvt Ltd Hyderabad, India

- Applied predictive ML with Python, TensorFlow, Pandas, scikit-learn, and R; improved forecast accuracy by 20% through feature engineering and cross-validation; built Power BI dashboards to communicate insights to non-technical stakeholders.
- Built ETL pipelines and scalable data APIs on Databricks, Spark, Redshift, and Snowflake; maintained Git-based CI/CD workflows and stayed current with emerging AI/cloud technologies — reducing reporting latency by 55%.

KEY PROJECTS

Serverless LLM & AWS Bedrock Integration Platform *AWS Bedrock · SageMaker · Lambda · S3 · ECS · EC2 · Docker · Terraform · LangChain*

- Deployed production LLM inference on AWS Bedrock and SageMaker with fine-tuned foundation models; built scalable APIs and data pipelines using Lambda, S3, and ECS — containerized via Docker, provisioned with Terraform IaC.

AI Agent Orchestration – RAG & Generative AI Pipeline *Python · OpenSearch · Pinecone · Prompt Engineering · MLOps · OpenAI · Anthropic*

- Engineered multi-agent AI workflows with RAG retrieval and optimized prompt engineering for 40% relevance gain; monitored model performance with evaluation metrics and feedback-driven iteration — achieving sub-second latency at enterprise scale.

TECHNICAL SKILLS

LLMs & GenAI: LLMs, Generative AI, OpenAI, Anthropic, Cohere, Hugging Face, Prompt Engineering, Fine-Tuning, Foundation Models, AI Agents

ML Libraries: PyTorch, TensorFlow, scikit-learn, Pandas, NumPy, LangChain, Semantic Kernel, RAG, NLP, R

AWS Services: Bedrock, SageMaker, Lambda, S3, ECS, EC2, Step Functions, CloudWatch, IAM, KMS, VPC

Languages: Python, SQL, JavaScript, PySpark, Bash

MLOps & DevOps: Docker, Terraform, Jenkins, Git, CI/CD, Airflow, Databricks, Spark, Monitoring, Observability

Data & Cloud: Snowflake, BigQuery, Redshift, NoSQL, Azure AI, Azure OpenAI, GCP, REST APIs, Power BI

Security: Data Privacy, IAM, SCP, KMS Encryption, VPC Networking, Cloud Security Best Practices

EDUCATION

M.S., Data Analytics Engineering Jan 2023 – Dec 2024

George Mason University, Fairfax, VA | Focus: Machine Learning, Statistics, Data Engineering, Cloud Computing

B.Tech, Electronics & Communications Engineering Aug 2018 – Jun 2022

CVR College of Engineering, Hyderabad, India | Focus: Software Systems, Data Structures, Mathematics