**VENKATA ADITYA CHINTALA**

9806216227 | ch.aditya95@gmail.com

**OBJECTIVE**

Experienced Senior AI Engineer with over 5 years of experience in AI and Machine Learning, seeking to leverage my expertise in generative AI, LLMs, and predictive modeling in an AI Engineer role. Skilled in designing and implementing AI models and pipelines, with a strong foundation in MLOps, cloud platforms, and software development best practices. Looking to contribute to a dynamic environment by driving innovative AI solutions and enhancing business outcomes.

**EDUCATION**

**Texas A&M University,** College Station, TX*May 2019*

*Master of Science, Computer Science(Data Science Track)* *GPA: 4.0/4.0*

**WORK AUTHORIZATION**

**• Legally eligible** to work in the United States(**Permanent Resident**). Does **not** require sponsorship.

**CAREER SUMMARY**

* Over 5 years of experience as a Senior Data Engineer, specializing in AI and machine learning with a focus on generative AI, large language models (LLMs), and predictive modeling.
* Proficient in designing, developing, and managing AI models and pipelines, with hands-on experience in implementing retrieval-augmented generation pipelines and LLM-supported use cases.
* Proficient with emerging AI tools and technologies, including LangChain, LlamaIndex, Haystack, and Azure AI Studio, and adept at integrating these tools into AI pipelines.
* Extensive expertise in MLOps, automating machine learning pipelines in cloud environments (Azure, AWS, GCP), and utilizing orchestration tools like Airflow, Kubeflow, and AWS Step Functions
* Extensive experience in data warehousing systems, data systems integration, and cloud data solutions including data lakes or managed services such as Snowflake and/or Databricks.
* Knowledge of CI/CD tools (e.g., Jenkins), version control (Git), orchestration/DAGs tools (AWS Step Functions, Airflow, Luigi, Kubeflow, etc.)
* Proficiency in Python and SQL across multiple platforms, and exposure to the Python ML ecosystem (numpy, pandas, scikit-learn, tensorflow, etc.)
* Strong understanding of software development best practices including CI/CD, unit and integration testing, and code reviews, ensuring robust and scalable AI solutions
* Experienced in working with popular data engineering tools and technologies such as Apache Airflow, Domino Data Science Platform, Apache Spark, and Apache Hive to build, manage and optimize high-performing data pipelines and workflows in a distributed computing environment.

**KEY SKILLS**

• **AI & Machine Learning:** Expertise in generative AI, LLMs (Claude, GPT, Llama 3.2), TensorFlow, PyTorch, and RAG pipelines, achieving significant performance improvements in production systems.

• **MLOps & Cloud Platforms:** Designed and deployed automated ML workflows using tools like MLflow, Vertex AI, Azure Machine Learning, and AWS SageMaker, with measurable improvements in deployment speed and model reliability.

• **Data Engineering**: Built scalable, end-to-end data pipelines using Apache Spark, Kafka, SQL, and NoSQL databases, increasing data processing efficiency by up to 85% in large-scale projects.

• **Cloud Expertise:** Hands-on experience with AWS (Bedrock, Lambda, SageMaker), Azure (AI Studio, Synapse), and GCP (Vertex AI, BigQuery), enabling cross-cloud integrations and scalability**.**   
• **Synthetic Data Expertise**: Applied techniques such as VAEs, GANs, and SMOTE to expand training datasets up to 10x, improving model performance and robustness by as much as 17%.  
• **Machine Learning**: Developed and deployed machine learning models in production environments (TensorFlow, PyTorch,), achieving a 12% boost in accuracy through synthetic data augmentation.  
• **Orchestration & Automation**: Implemented CI/CD pipelines (Jenkins) and workflow orchestration tools (Airflow, Kubeflow, AWS Step Functions), reducing deployment time by 40%.  
• **Containerization**: Used Docker and Kubernetes for scalable, efficient model deployment, reducing downtime by 30%.  
• **Programming**: Proficient in Python, Java, and Scala, with extensive use of Python ML libraries to streamline complex workflows.

**TECHNICAL SKILLS**

**• Tools:** Python, Airflow, SnowFlake, Azure, AWS, Matillion, **• Databases:** SnowFlake, Redis, MongoDB, PostGreSQL.

**• Packages:** Kubeflow, Pandas, PySpark, Spacy, Kafka. PyTorch **• AWS Tools:** Lambda, S3, EC2, Glacier, RedShift, Athena.

**• AI Tools:** LangChain, LlamaIndex, Vertex AI, Haystack **• Data Processing:** Apache Spark, PySpark, Matillion

**PROFESSIONAL EXPERIENCE**

**North Highland, Senior AI Engineer(***April 2024 – Present)*

* Deployed an end-to-end chatbot system with a React-based front-end and Python-powered back-end, integrating Retrieval-Augmented Generation (RAG) pipelines on AWS for real-time interaction, reducing query resolution time by **30%** and enhancing user satisfaction by **20%.**
* Leveraged AWS Bedrock with Llama 3.2 models for scalable ML model deployment, achieving a **25%** improvement in API efficiency and **70%** faster response times for inferencing, ensuring **99.9%** uptime for high-demand enterprise applications.
* Engineered a Knowledge Graph-based Explainable AI framework leveraging AWS Neptune, ensuring **100%** traceability and transparency across pipelines, and reducing compliance validation efforts.
* Automated ingestion pipelines with SageMaker and GPU-accelerated parallel processing for 20 data sources, reducing processing time by **60%** and manual efforts by **50%**.
* Designed and deployed MLOps pipelines using Azure Machine Learning and MLflow, reducing manual effort by **40%** and improving deployment speed by **35%**, while ensuring continuous delivery of production-ready models.
* Integrated RAG-based AI pipelines with advanced LLMs such as Claude and GPT, achieving a **50%** improvement in data retrieval precision and reducing average user interaction time by **20%**.
* Deployed ML models as managed services on both AWS and Azure, enabling high-availability systems and reducing deployment cycles by **35%**, with seamless integration into existing CI/CD pipelines
* Overcame bottlenecks in content extraction workflows by deploying ML models on AWS Bedrock as APIs for real-time responses. Parallelized parsing processes to handle large-scale documents, reducing processing time by **60%** and increasing throughput by **5x**
* Utilized advanced techniques such as bounding box algorithms to dynamically extract and classify relevant content types (e.g., graphs, images, and tables) from PDF documents, achieving a **40%** increase in accuracy and reducing manual processing time by **50%**.

**Trubridge Inc, Senior ML Data Engineer(***April 2023 – April 2024)*

* Engineered end-to-end machine learning solutions on Azure Databricks, from data ingestion to model deployment, resulting in a **50%** increase in model accuracy and efficiency.
* Deployed AI-powered chatbots(**RAGs**) using OpenAI models on Azure, enhancing customer interaction capabilities and improving response accuracy by **40%**.
* Utilized Azure Managed Serving Endpoints for deploying and managing multiple machine learning models, ensuring high availability and scalability while reducing deployment time by **30%**.
* Implemented LangChain and LlamaIndex for Retrieval-Augmented Generation, achieving a **45%** improvement in data retrieval speed and accuracy for LLM-supported use cases.
* Designed and managed MLOps pipelines using Azure Machine Learning and MLflow, automating model training, testing, and deployment processes, leading to a **40%** reduction in manual effort and a 35% improvement in deployment speed.
* Leveraged Spark DataFrames within Python for large-scale data processing, achieving an **85%** boost in data ingestion and processing speeds..

**Apex Systems, Contracting, Senior Machine Learning Engineer(***April 2022 – April 2023)*

* Engineered end-to-end AI and machine learning solutions using Azure Databricks and Airflow, enhancing data processing reliability by **55%** and minimizing downtime by 40**%**.
* Developed and deployed predictive analytics models with Azure Machine Learning, resulting in a **30%** improvement in decision-making accuracy.
* Enhanced NLP capabilities using Azure Cognitive Services, leading to a **40%** increase in text analysis accuracy and efficiency.
* Optimized warehouse size and reduced costs using Matillion by dynamically altering the size for batch loads and reducing processing time by **35%**, resulting in a **20%** reduction in overall data storage costs.
* Leveraged PySpark DataFrames in DataBricks for ETL processes and captured streaming data, writing it efficiently to SnowFlake, resulting in a **50%** reduction in ETL processing time.
* Managed tasks cost-effectively using Serverless Tasks in SnowFlake, resulting in a **20%** reduction in processing time.
* Improved ingestion and processing speed of large streaming datasets using Spark DataFrames in Python, resulting in an **85%** increase in performance.
* Utilized Azure Data Factory for orchestrating complex data workflows, increasing data pipeline efficiency by 45% and reducing ETL processing time by **30%**.

**Client Command, Senior Big Data Engineer(***October 2020 – April 2022)*

* Implemented Apache Airflow to manage and monitor ETL workflows, creating custom DAGs that increased data processing efficiency by **55%** and reduced error rates by **35%**
* Designed and implemented real-time data pipelines to process semi-structured data by integrating **50 millions raw records** from multiple sources using PySpark and stored processed data in SnowFlake.
* Developed machine learning models for predictive analytics, using Azure Machine Learning, improving prediction accuracy and operational efficiency by **30%**.
* Enhanced data quality checks with automated processes in Databricks notebooks and Matillion, ensuring high data quality and timely anomaly alerts..
* Implemented API calls to store data into Redis Database for caching for faster accessibility for Front End Applications which reduced the loading time of modules by **40%**.
* Utilized Spark DataFrames in Python for distributed data processing on large streaming datasets to improve ingestion and processing speed of the data by **85%**.

**Decooda International, Data Engineer(***May 2018 – September 2020)*

***Deep Learning Inferencing Service for Historical Data*** ***May 2020-September 2020***

* Solved the problem of Co-Reference Resolution at scale while processing millions of historical documents by utilizing serverless architecture using OpenFaas, Docker and Kubernetes containers resulting in **50%** improved time efficiency.
* Re-Trained a pre-trained NeuralCoref Model to improve accuracy and performance while resolving rare words in financial documents.
* Solved the problem of rapidly inferencing deep learning models while streaming data resulting in about **70% savings** on AWS Costs.
* Improved the processing time from **80 documents/second** to **200 documents/second** by using Asynchronous data pipelines.
* Implemented RESTful API to collect data arriving in S3 to offer near real-time scoring of accurate data.
* Created a Data Pipeline to deploy any Machine Learning and Deep Learning Models in Python using Matillion and SnowFlake.
* Improved the accuracy of predicting rarewords by **5%** while training on Financial Data for the Co-Reference Resolution model.

***Deep Learning Inferencing at scale for Streaming Data*** ***Jan 2020 – April 2020***

* Designed a Data pipeline which utilized serverless architecture for Deep Learning Inferencing for streaming data which improved the processing speed from from **7 Documents/Second** to **80 Documents/Second**.
* Experimented with Serverless technologies Amazon RDS Proxy, AWS Lambda, Amazon Aurora, DynamoDB for this use-case.
* Created an AWS Lambda function which listens to an S3:PUT event which asynchronously processes each file arriving in S3 parallelly.
* Improved average processing speed from **7 documents/second** to **80 documents/second** by leveraging distributed serverless processing.
* Created an distributed pipeline leveraging Kafka, Snowflake, S3 and Kubernetes resulting in **23%** improvement of processing speed.

***SnowFlake DataWarehousing******Feb 2019 – Dec 2019***

* Spearheaded the project which calculated metrics for a major US-based Multinational hedge-fund to predict movement of stock prices.
* Explored and experimented with latest Technologies for the requirement using RedShift, SnowFlake, BigQuery and PostGreSQL.
* Designed and automated a data flow process to migrate billions of rows of real-time data from PostGres to SnowFlake.
* Improved the overall performance speed of existing data pipelines by **50%** using code modification.
* Migrated Data from PostGreSQL, Prolog and Redis to SnowFlake using the ETL tool Matillion.
* Used ETL(Matillion) tool to develop automated jobs to calculate metrics every week **50 times** faster than existing system.
* Improved the processing speed of metrics by a factor of **100** by leveraging data locality in SnowFlake and Matillion.
* Onboarded and Mentored newcomers on the ETL processing project. Wrote and maintained instructional documents for setting up a Development environment.
* Supervised and lead a team of **10+** members as technical lead for a fortune 500 financial client in hedge-fund industry.
* Led a team that increased confluence usage by **60%** by reorganizing/rewriting all content in a more effective fashion.
* Solved the problem of manually giving table names by using orchestration and transformation variables, thereby resulting in hundreds of hours of office time saved.
* Improved processing of real-time financial data by a factor of **50%** which also powers PowerBI Dashboards for Financial Analysts.

**Decooda International, Data Science Intern** ***May 2018 – Dec 2018***

* Performed data extraction using Python to extract data from different sources on like DBPedia, NEWSAPI etc.
* Used ETL (SSIS) to develop jobs for extracting, cleaning, transforming and loading data into data warehouse and implementation of web service task.
* Created machine learning models to Predict documents based on topics and neural networks to establish semantic-linking.
* Designed, tested, and delivered a Topic-Modelling machine learning system.