Yashwanth Reddy Vasireddy

AI Engineer

Glassboro, NJ | +1 (856) 347-9336 | yashwanth1294@careerattainment.com | Linkedin

Summary

AI/ML Engineer and Data Scientist with 3+ years of experience building production ML systems for complex enterprise environments. Specialized in developing NLP pipelines and fine-tuning Transformer models to handle sensitive, large-scale unstructured data. Deeply skilled in deploying RAG architectures and optimizing inference latency using Python, PySpark, and AWS. Proven track record of translating messy real-world data into scalable solutions that automate workflows and drive operational efficiency.

Skills

Languages:	Python, R, SQL
Libraries & Models:	TensorFlow, PyTorch, Keras, Scikit-learn, Pandas, NumPy, Seaborn, Matplotlib, Beautiful Soup, Hugging Face, Lang Chain, Lang Graph, NLTK, OpenCV, Spacy, ScispaCY, Transformers (BERT, GPT, RoBERT), Crew AI
Databases:	MySQL, PostgreSQL, MongoDB, ChromaDB
Cloud Services:	AWS, IBM Cloud Services (Auto AI, Watson Assistant), Azure, GCP (Big Query)
Artificial Intelligence:	Supervised & Unsupervised Learning, Reinforcement Learning, Deep learning Architectures (Neural networks, Transfer Learning), NLP (Semantic Analysis, Text segmentation), ML Ops (ML Flow), LLMs
Data Engineering:	PySpark, Airflow, Databricks, ETL/ELT Pipelines, Data Modeling, Hadoop, Teradata, Snowflake
Tools:	DataBricks, Tableau, Power BI, Spark, Docker, Google Colab, GitHub
Others:	SDLC, Agile, Jira, CI/CD Pipelines

Experience

AI Engineer | EY, USA **| Jan 2025 – Present**

- Designed and deployed Python based AI model to automatically detect and classify sensitive data (PII, PHI, financial records) across emails, documents, and unstructured text, helping reduce unintended data exposure by 45%.
- Built NLP driven classification pipelines using transformer models to understand content context, significantly reducing false positive DLP alerts by 30% and improving trust in automated security controls.
- Implemented supervised and hybrid machine learning models to analyze large scale data movement patterns and identify abnormal data flows linked to potential leakage incidents.
- Engineered contextual features using custom tokenizers and embeddings, handling noisy unstructured text from emails to improve classification accuracy on sensitive documents.
- Finetuned BERT based transformer models for sensitive data identification, improving classification precision and recall while reducing inference latency by 25% for near realtime security scanning.
- Integrated the AI powered DLP classifier with security dashboards and alerting workflows, enabling security teams to monitor risk trends, investigate alerts faster, and respond efficiently to potential data exfiltration events.

Data Scientist | HCL, India **| Feb 2021 – Jul 2023**

- Built and deployed predictive ML models (Random Forest, Logistic Regression, KNN, K-Means, PCA) to identify fraud risk and customer behavior patterns, reducing fraudulent claims by ~25%.
- Automated SQL-based data pipelines to refresh daily customer profiles, policy eligibility, and quote history datasets, enabling near real-time model scoring and analytics.
- Designed and analyzed A/B experiments comparing rule-based vs. AI-driven recommendations, resulting in a 12–15% increase in offer acceptance rates.
- Built executive-facing Tableau dashboards to track model performance, fraud trends, and recommendation impact, improving leadership visibility and data-driven decisions.
- Developed recommendation models using Python and TensorFlow to personalize insurance product suggestions, improving relevance and decision accuracy.
- Applied NLP techniques to analyze customer service transcripts and sentiment trends, contributing to a 10% uplift in customer satisfaction scores.
- Conducted exploratory data analysis to uncover seasonality, claim surges, lapse risks, and fraud indicators, directly influencing personalization and risk models.
- Performed model validation, feature selection, and hyperparameter tuning to ensure stable performance, regulatory compliance, and reliable production outcomes.

Education

Master of Science in Data Science | *Rowan University, Glassboro, NJ, USA* **May 2025**

Bachelor of Technology in Artificial Intelligence | *Vidya Jyothi Institute of Technology, Hyderabad, India* **May 2023**

Projects

Intelligent RAG & Agentic Q&A System | Stack: LangChain, HuggingFace, ChromaDB, Python, Vector Databases

- Built an LLM RAG system using ChromaDB and flan-t5, achieving high precision (>95%) in tool selection for complex queries.
- Optimized vector embeddings and text splitting and implemented strict context-windowing to minimize hallucinations and ensure fact-based retrieval.

Retrieval-Augmented Generation (RAG) Pipeline | Stack: Python, ETL Pipelines, Data Modeling, Embedding Models, Open-Source LLMs

- Built an end-to-end RAG pipeline using multiple embedding models and open-source LLMs for contextual query answering.
- Evaluated performance on a custom QA dataset and analyzed hallucination cases.
- Compared RAG vs. non-RAG outputs to measure factual accuracy and retrieval effectiveness.