Eswar Sai Korrapati
Open Source Contributor | 551-344-5356 | eksai0726@gmail.com | LinkedIn | Github | Portfolio

---

Professional Summary

AI/ML Engineer with over three years' experience designing and deploying scalable machine learning solutions on AWS using SageMaker, MLflow, and Spark. Skilled in building end-to-end pipelines, API integration, and MLOps for reproducibility and automation. Experienced in model explainability with SHAP and LIME, and optimizing models for real-time inference in production.

Professional Experience

Software Engineer – AI/ML
Comcast, San Jose, CA
Aug 2024 – September 2025

- Delivered production-ready ML models in TensorFlow and PyTorch, boosting prediction accuracy by 20% in live streaming systems.
- Built reproducible workflows with Airflow, MLflow, and DVC, enabling versioned pipelines and consistent experiment tracking.
- Deployed models via AWS SageMaker and Lambda, cutting infrastructure costs by 15% and deployment times by 35%.
- Designed SHAP- and LIME-based explainability dashboards to support compliance and data-driven decisions.
- Automated hyperparameter tuning with SageMaker Autopilot and Optuna, improving model F1 score by 18%.
- Built scalable batch inference pipelines with Apache Spark and AWS S3, reducing inference time by 25% across large datasets.
- Integrated LLM-driven chat support using Bedrock API and prompt engineering enhancing customer service automation.
- Fine-tuned domain-specific LLMs using Hugging Face and Bedrock, improving contextual accuracy for enterprise use cases.
- Integrated MCP tools with LLM pipelines to enable cross-tool orchestration and enhance automation capabilities.
- Collaborated in Agile teams with Git, Docker, and Jenkins to streamline CI/CD across ML projects.

Stack: Python, TensorFlow, PyTorch, MLflow, Hugging Face, SageMaker, DVC, Airflow, Spark, AWS (S3, Lambda, CloudWatch, Bedrock, Sagemaker), Docker, Git, Optuna, SHAP, LIME

Software Engineer
Dentsu, Hyderabad, India
May 2020 – Jul 2022

- Developed RESTful APIs and backend services in Django and Flask, improving system throughput and API response times.
- Built automated ML workflows with Airflow, strengthening reliability and scheduling across data ingestion tasks.
- Containerized and deployed web services using Docker and AWS EC2, achieving faster and more resilient rollouts.
- Designed a central Feature Store to streamline feature reuse, speeding up experimentation and improving consistency.
- Applied AutoML for ad campaign forecasting, reducing manual tuning by 40% and increasing accuracy.
- Enhanced reporting with model metrics (AUC, Precision, Recall, F1) and Tableau visualizations for clear stakeholder insights.
- Managed cloud deployments with AWS and Jenkins, improving release reliability through automated workflows.

Stack: Python, Flask, Django, SQL, MongoDB, H2o AutoML, Airflow, Feature Store, Docker, Jenkins, Git, AWS EC2, Tableau

---

Education

Master's in Computer Science
Montclair State University, NJ
Graduated: May 2024 | GPA: 3.65

Projects

- RAG-based PDF Summarizer | Link
- Next Word Prediction with BERT Transformer | Link
- Machine Transliteration | Link
- AI Meeting Preparation Agent | Link
- Job Application Automation | Link

Technical Skills

Programming Languages: Python, C++, SQL, JavaScript
AI/ML & NLP: TensorFlow, PyTorch, Scikit-learn, Keras, Transformers, Hugging Face, LangChain, LangGraph,n8n
MLOps & Deployment: MLflow, SageMaker, DVC, Airflow, Docker, Kubernetes, CI/CD, Feature Store, Jenkins
Explainability & Evaluation: SHAP, LIME, AUC, Precision, Recall, F1 Score, Confusion Matrix
Cloud & DevOps: AWS (S3, EC2, Lambda, SageMaker, CloudWatch, Bedrock), GitHub Actions
Data & Processing: Apache Spark, Pandas, NumPy, MongoDB, MySQL, Tableau
LLMs & APIs: OpenAI API, Prompt Engineering, LangChain, FastAPI, REST APIs, RAG, Fine-tuning(RLHF,LoRA), MCP
Tools & Collaboration: Git, Jupyter, Streamlit, Agile, IntelliJ, VS Code
Core CS: Data Structures and Algorithms