PURNA CHANDER KONDA

(618)-974-7219| kondapurnachander42@gmail.com | LinkedIn

PROFESSIONAL SUMMARY

Results-driven **AI/ML Engineer** with **4+ years of experience** designing and deploying **production-ready ML models** and **scalable AI systems** across **finance, retail, and healthcare**. Expert in **Generative AI**, **RAG**, and **MLOps** using **Python**, **PyTorch**, **TensorFlow**, **Hugging Face**, **LangChain**, and **AWS**. Skilled in **NLP**, **Computer Vision**, and **Predictive Analytics**, with a proven track record of **automating workflows**, **improving efficiency**, and delivering **secure, high-impact business outcomes**. Committed to **Responsible AI** and **regulatory compliance**.

TECHNICAL SKILLS

- **Programming & Scripting:** Python, SQL, JavaScript, R, Java
- **Machine Learning & Evaluation:** Supervised/Unsupervised Learning, Random Forest, XGBoost, LightGBM, CatBoost, SVM, Time-Series (Prophet, LSTM), Graph ML (GNN), Feature Engineering (Pandas, NumPy, Scikit-learn), Metrics (F1, AUC-ROC, BLEU, ROUGE)
- **Large Language Models & Deep Learning:** PyTorch, TensorFlow, Keras, Hugging Face Transformers, GPT, BERT, T5, CLIP, Diffusion Models, GANs, VAEs, Fine-Tuning (LoRA, QLoRA, PEFT), NLP (spaCy, NLTK)
- **Generative AI & Prompt Engineering:** OpenAI API, LangChain, LlamaIndex, Cohere, Anthropic Claude, DreamBooth, Zero/Few-Shot Prompting, Chain-of-Thought, Retrieval-Augmented Generation (RAG), RAGAS, LangSmith
- **MLOps & Model Lifecycle:** MLflow, Kubeflow, KServe, AWS SageMaker, Vertex AI, DVC, Weights & Biases, CI/CD (Jenkins, GitHub Actions), Large-Scale ML Systems, A/B Testing, Optuna, Hyperparameter Tuning
- **Data Engineering & Pipelines:** Kafka, Apache Spark, Apache Airflow, BigQuery, Snowflake, Apache NiFi, ETL Pipelines, Data Warehouses, Distributed Computing
- **Vector Databases & Storage:** FAISS, Pinecone, Weaviate, ChromaDB, Milvus, PostgreSQL, MongoDB, Neo4j
- **Deployment & Cloud:** AWS (Bedrock, S3, Lambda), GCP (Vertex AI), Azure, Docker, Kubernetes, ONNX, TensorRT, Model Quantization (GGUF, AWQ, GPTQ), FastAPI, Flask, gRPC
- **Security, Responsible AI & Compliance:** SHAP, LIME, Fairlearn, Federated Learning, Responsible AI Standards, Model Governance, GDPR, HIPAA, FINRA, FCRA, CCPA, PII Protection
- **Monitoring & Visualization:** Prometheus, Grafana, ELK Stack, Model Drift Detection, Streamlit, Gradio, Tableau, Power BI

PROFESSIONAL EXPERIENCE

Broadridge ***Kansas City, MO***
AI/ML Engineer *Apr 2025 - Present*

- Architected scaled **AI/ML pipelines** using **RAG (LangChain, LlamaIndex)** and vector stores (**Pinecone, Weaviate, Milvus**), boosting document analysis efficiency by **40%** with **RAGAS** benchmarking for accuracy.
- Fine-tuned foundation models **(GPT, LLaMA)** using **LoRA, QLoRA,** PEFT; implemented **Chain-of-Thought** and **RLHF** to reduce hallucinations and improve compliance metrics by **30%.**
- Orchestrated production-grade **MLOps** via **Kubeflow, MLflow, and SageMaker Pipelines,** maintaining **99.9%** uptime and accelerating deployment cycles for financial services by **40%.**
- Engineered high-throughput fraud detection and credit scoring pipelines using **Apache Spark, Kafka,** and **Vertex AI,** boosting anomaly detection by 25% through high-precision predictive modeling.
- Developed **XAI** dashboards with **SHAP, LIME,** and **Fairlearn,** enabling auditors to assess bias and meet strict **FINRA, GDPR**, and **FCRA** regulatory and compliance standards.
- Optimized model inference via **ONNX and TensorRT** with **Quantization (GGUF/AWQ)**, reducing latency by **35%** for high-traffic financial decision systems and real-time workflows.
- Constructed **financial Knowledge Graphs with Neo4j,** mapping complex relationships to uncover hidden fraud rings and enabling analysts to retrieve risk-linked insights **50%** faster.
- Integrated AI microservices into enterprise platforms using **AWS Bedrock, FastAPI, and gRPC**, enabling secure, real-time financial workflows for 15K+ active enterprise users.

O'Reilly Auto Parts ***Springfield, MO***
Machine Learning Engineer *Mar 2024 - Mar 2025*

- Architected demand forecasting engines using **XGBoost** and **Prophet** via **PyTorch**, optimizing regional inventory with **DVC** versioning to reduce stockout events **by 18% a**cross **1,200** hubs.
- Engineered **ETL/ML** pipelines via **Apache Spark, Airflow,** and **Snowflake,** orchestrating **100TB+** of transactional data for production-grade supply chain predictive analytics.

- Orchestrated containerized inference microservices using **Docker** and **Kubernetes,** maintaining **sub-200ms** latency for pricing models via **Terraform**-managed elastic infrastructure and **Helm** deployments.
- Streamlined **MLOps f**rameworks via **AWS SageMaker**, **Optuna**, and **MLflow**, automating hyperparameter optimization and model tracking to accelerate retraining cycles by **35%.**
- Constructed real-time streaming architectures using **Kafka** and **Apache Flink** for proactive anomaly detection, identifying logistics disruptions across **28** distribution centers with **PySpark** integration.
- Fine-tuned **BERT-based NLP models** to categorize 500,000+ unstructured parts, utilizing **Temporal Fusion Transformers** and **Hugging Face** to capture seasonal patterns and reduce overstock by 10%.
- **Automated enterprise CI/CD workflows** via **GitLab CI** and **GitHub Actions**, facilitating seamless model regression testing to maintain **99.9% uptime** for mission-critical retail analytics microservices.
- **Developed dashboards** using **Prometheus**, **Grafana**, and **Tableau**, translating complex model telemetry into actionable supply chain **KPIs** that accelerated operational decision cycles by 30%.

Sahrudaya Healthcare ***Hyderabad, India***
Data Scientist *Aug 2022 - Nov 2023*

- Developed predictive risk models using **XGBoost** and **LightGBM,** increasing patient engagement by **22%** and reducing hospital readmissions by **15%** through advanced clinical feature engineering.
- Fine-tuned **BERT** and **RoBERTa NLP** models via **Hugging Face** to classify **1M+** medical records, improving classification accuracy by **28%** and reducing manual clinician review time by **40%.**
- Architected **HIPAA-compliant Spark** and **Hive ETL workflows** on **Azure** and **GCP** (Vertex AI), processing high-volume clinical data while ensuring secure **PII handling** and **99.9%** data reliability.
- Engineered Computer Vision pipelines using **CNNs** and **OpenCV** for automated medical imaging analysis, improving diagnostic screening speed by **30%** and assisting in early pathology detection.
- Constructed time-series forecasting models using **LSTM** and **Keras** to predict emergency room capacity, improving patient throughput by **20%** via **TensorFlow**-based deep learning.
- Developed model explainability frameworks using **SHAP** and **MLflow**, implementing **automated drift detection** to validate fairness and optimize patient prioritization via **A/B testing**.

Citco ***Hyderabad, India***
Data Scientist *Mar 2021 - Jul 2022*

- Architected **GNN-based** fraud detection models and link analysis pipelines using **Neo4j,** uncovering hidden fraud rings across policyholder data and resulting in a **27%** reduction in false positives.
- Developed high-performance time-series ensembles using **CatBoost, Prophet,** and **LSTM,** improving underwriting accuracy for insurance premium collections through automated hyperparameter tuning.
- Engineered scalable enterprise pipelines via **Databricks** and **Spark MLlib,** cutting reporting turnaround by **40%** and enabling complex daily risk simulations on distributed datasets.
- Optimized distributed query performance by restructuring unstructured data pipelines in **Cassandra** and **MongoDB,** reducing retrieval latency for high-frequency financial markers by **40%.**
- Deployed production ML microservices using **Docker** and **Kubernetes**, integrating **Tableau/Power BI** visual analytics to align model **KPIs** with global risk and compliance strategies.

EDUCATION

Master's in information technology management, Webster University, MO, USA.
Bachelor's in computer science and information technology, CVR College of Engineering, Hyderabad, India.

PROJECT HIGHLIGHTS

Diffusion-Based Image Generation Pipeline | Stable Diffusion, DreamBooth, AWS

- Built a text-to-image synthesis system using **Stable Diffusion** and **DreamBooth**, fine-tuning models on branded product datasets to automate personalized marketing asset creation.
- Orchestrated distributed training on **AWS EC2** using **PyTorch Lightning (DDP)**, reducing training latency by **40%,** and deployed an internal web-tool via **Streamlit** and **Gradio**.

Explainable AI (XAI) Credit Risk Framework | SHAP, Fairlearn, Kubeflow

- Developed a regulatory-aligned credit scoring framework using **Transformers** and **SHAP**, providing auditable model decisions that improved audit report turnaround by **25%.**
- Integrated **Fairlearn** and **LIME** to mitigate algorithmic bias, achieving a **15% reduction** in disparate impact across demographic groups while maintaining model precision.

- Deployed an automated retraining and compliance monitoring pipeline via **Kubeflow** and **Airflow**, ensuring **99% alignment** with model governance and transparency standards.

CERTIFICATIONS

- AWS CERTIFIED DATA ENGINEER - ASSOCIATE
- MICROSOFT CERTIFIED: AZURE AI ENGINEER ASSOCIATE