Sandeep Kumar Amgothu

Machine Learning Engineer

sandeepkumaramgothu3@gmail.com | +1 (203) 997-4125

linkedin.com/in/sandeepkumaramgothu | github.com/Sandeepkumaramgothu | Google Scholar | Portfolio

Professional Summary

Machine Learning Engineer with 5+ years building production-grade AI systems specializing in LLM security, computer vision, and scalable ML infrastructure. Expert in adversarial AI red-teaming, responsible AI governance, and MLOps automation, with proven ability to deploy models processing 15M+ records monthly while reducing latency by 40%. Published researcher (ICUAS 2025) bridging academic innovation and enterprise-grade ML deployment. Seeking to leverage deep expertise in AI safety, model hardening, and cloud-native ML systems to drive transformative AI solutions in 2025 and beyond.

Core Technical Skills

Machine Learning & AI: PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, LangChain, OpenAI API, Anthropic Claude

LLM & Generative AI: GPT-4/4o, LLaMA-2/3, BERT, T5, RAG Pipelines, Prompt Engineering, Fine-tuning (LoRA/QLoRA)

AI Safety & Security: Adversarial Red-Teaming, Jailbreak Detection, Toxicity Evaluation, Bias Mitigation, Llama Guard

MLOps & Infrastructure: MLflow, Weights & Biases, DVC, Docker, Kubernetes, CI/CD, Model Monitoring, A/B Testing

Data Engineering: Python, SQL, Apache Airflow, PySpark, Pandas, AWS (S3, Glue, Lambda, SageMaker, Bedrock), Snowflake

Vector Databases & Search: FAISS, Pinecone, Milvus, Chroma, Elasticsearch

Cloud & DevOps: AWS, Azure, GCP, Terraform, GitHub Actions, Model Deployment (REST APIs, Serverless)

Analytics & Visualization: Power BI, Tableau, Matplotlib, Seaborn, Plotly, Streamlit

Professional Experience

Texas A&M University–Corpus Christi (TAMUCC) **Corpus Christi, USA**

Graduate Research Assistant – AI/ML Research Jan 2024 – May 2025

- Led development of production-ready adversarial AI testing framework processing 3,000+ LLM interactions with automated taxonomy-driven safety evaluation and database-backed persistence for enterprise-scale model governance
- Architected end-to-end MLOps pipeline for audio-based UAV detection: engineered STFT spectrogram preprocessing workflows reducing computation time by 66% (15h → 5h), trained ensemble deep learning models (CNN/RNN/CRNN/VGG19) achieving 97%+ F1-score on 50K+ samples
- Built production RAG systems integrating LangChain, Hugging Face Transformers, and FAISS vector stores for context-aware document retrieval, enabling sub-200ms query latency at scale for enterprise knowledge management applications
- Designed and deployed automated LLM red-teaming infrastructure (PYRIT Crescendo, jailbreak templates) reducing attack success rates by 20–50% through safety fine-tuning and weighted ensemble scoring (Llama Guard, Detoxify, DeBERTa)
- Published peer-reviewed research at ICUAS 2025 demonstrating 90%+ classification accuracy for real-time UAV threat detection using deep learning on acoustic signatures

Tata Consultancy Services (TCS) **Hyderabad, India**

Data Engineer Nov 2021 – Jul 2023

- Architected cloud-native ETL pipelines on AWS (S3, Redshift, Glue, Lambda) processing 15M+ records/month with 40% latency reduction through optimized data partitioning, parallel processing, and incremental load strategies
- Led AWS cloud migration for Fortune 500 client (Xerox), delivering 99.9% data availability SLA and enabling real-time analytics dashboards serving 500+ enterprise users across legal/financial operations
- Engineered Python-based intelligent web scrapers (BeautifulSoup, Selenium) with NLP-powered content extraction, automating ingestion of 10K+ legal documents monthly and reducing manual processing by 35%
- Collaborated with ML teams to build feature engineering pipelines and data quality frameworks, improving downstream model accuracy by 15% through schema validation, anomaly detection, and automated data profiling
- Designed executive-facing Power BI/Tableau dashboards with real-time KPI tracking and predictive analytics, enabling data-driven decisions that reduced document turnaround time by 35%

Accenture **Hyderabad, India**

Associate Data Engineer – AI/Analytics Trainee Jun 2021 – Oct 2021

- Automated recurring analytics workflows using Python, SQL, and Power BI, eliminating 25% of manual reporting efforts while reducing error rates by 30% through robust data validation and testing frameworks
- Developed production-grade Python scripts for data preprocessing, feature engineering, and quality assurance across 500K+ records, supporting ML model training pipelines and improving model accuracy by 15%
- Delivered 5+ client data analytics projects in cross-functional teams, strengthening expertise in SQL optimization, ETL best practices, and agile data engineering methodologies

Key Machine Learning Projects

Automated Taxonomy-Driven LLM Red-Teaming Framework | GitHub Jul 2025 – Dec 2025

- Architected enterprise-grade adversarial testing framework extending AdversaFlow with automation-first design: orchestrated 3,000+ multi-regime attacks (PYRIT Crescendo, jailbreak templates) across 6 model configurations with database-backed persistence
- Engineered weighted ensemble scorer combining Llama Guard, DeBERTa, Detoxify, and RoBERTa for taxonomy-aware safety evaluation, achieving 20–50% reduction in attack success rates (ASR) through iterative safety fine-tuning of LLaMA-3.2-1B
- Implemented production MLOps pipeline with 4-bit quantization (bitsandbytes), PyTorch GPU acceleration, and CSV-based artifact logging enabling reproducible ablation studies and longitudinal model comparison
- Identified critical residual vulnerabilities in misinformation and human-chatbot interaction harms (59–68% ASR post-defense), demonstrating taxonomy-aware evaluation superiority over blanket safety metrics

UAV Audio Detection using Deep Learning | GitHub Jun 2024 – Jan 2025

- Developed production-ready spectrogram-based classifiers (CNN, RNN, CRNN, VGG19) utilizing STFT transformations achieving 97.2% F1-score and 90%+ accuracy for real-time unauthorized UAV detection across 8 drone types
- Engineered scalable data augmentation pipeline (pitch shift, time-stretch, noise injection, SpecAugment) processing 50K+ audio samples with parallelized transformations, improving model generalization by 12%
- Optimized model architecture and hyperparameters using MLflow experiment tracking and automated hyperparameter tuning, reducing inference latency to <100ms for edge deployment on resource-constrained devices

Supply Chain Analytics Dashboards | Power BI + Python May 2024 – Aug 2024

- Delivered enterprise business intelligence solution integrating automated Power BI dashboards with SQL

Server backend and Python preprocessing pipelines (Pandas, NumPy) for real-time supply chain KPI monitoring

- Reduced manual reporting by 30% while providing actionable insights on inventory optimization, delivery performance, and quality metrics to 200+ operational stakeholders

Education

Texas A&M University–Corpus Christi — M.S., Computer Science Aug 2023 – May 2025
GPA: 3.93/4.0 | Coursework: Machine Learning, Deep Learning, Computer Vision, NLP, AI Safety
Leadership Scholarship Recipient

Vasavi College of Engineering — B.E., Information Technology Aug 2017 – Aug 2021
GPA: 3.64/4.0

Certifications & Publications

- **Microsoft Certified: Power BI Data Analyst Associate** | ID: 8C3D63-F1C5AB Valid: Jun 2025 – Jul 2026
- **Microsoft Certified: Azure Administrator Associate** | ID: 959B1A-3WBE29 Valid: Feb 2023 – Feb 2026
- **Research Publications:**
  - UAV Audio Detection and Identification Using STFT Spectrograms with Deep Learning, ICUAS 2025
  - Automated Taxonomy-Driven Red Teaming for Large Language Models (In Preparation, 2025)
  - UAV Detection Systems, TAMUCC Student Innovation Symposium (2025)