Venkata Sai Ganesh Chandu Bheesetty

(AI/ML Engineer)

venkatchandu848@gmail.com +91 8207562778 (Whatsapp Only) venkat-chandu-bheesetty venkatchandu848

Wolfsburg, Germany (Open to Relocation)

SUMMARY

Passionate AI/ML Engineer with 2.5+ years of experience at Volkswagen AG, University of Siegen, and Vedanta Limited. Specializes in Machine Learning, Generative AI, and Agentic AI, with strong expertise in scalable AI solutions and research methodology.

Core Stack: Python, C++, SQL, PyTorch, LangChain, OpenCV, Azure/AWS, Docker, MATLAB.

EXPERIENCE

Data Scientist, *Stealth Startup* 10/2025 – 12/2025 | Berlin, Germany

- Automated ETL/ELT pipelines, replacing manual workflows to **reduce processing time by 20%**.
- Scaled web scraping ingestion pipeline to **2x peak loads**, achieving higher uptime and reliability.
- Optimized ML models for forecasting, pattern recognition, and deployment, through rigorous fine-tuning and benchmarking.
- Built LLM-driven workflows through advanced prompting techniques (DSPy) to automate analytical tasks and enhance forecasting predictions, while utilizing distributed programming for faster inference.

AI Engineer, *Volkswagen AG* 12/2024 – 08/2025 | Wolfsburg, Germany

- Developed and Optimized Vision Transformer (ViT) and Vision-Language Model (VLM) pipelines to enhance model explainability for automotive perception tasks.
- Fine-tuned token-pruned Faster R-CNN model with ViT backbone, delivering **3x reduction in FLOPs, while achieving 4% mAP accuracy gain** for object detection task, enabling model deployment in resource-limited environments.
- Conducted unsupervised clusterability analysis of high-dimensional VLM embeddings using NMI, **identifying DINO's 60% superior semantic separability vs baselines**, improving object detection performance and interpretability.
- Conducted comparative analysis between Large Language Models (LLMs) and ViT regarding high-dimensional embedding properties for better feature engineering decisions.
- Utilized Azure ML for experiment tracking, model versioning and reproducibility, streamlining research and analysis workflows.

Machine Learning Engineer, *Volkswagen AG* 05/2024 – 11/2024 | Wolfsburg, Germany

ML model development for ADAS Safety Functionality

- Refined and optimized existing CNN-LSTM pipelines using PyTorch for time-series prediction, anomaly detection and model evaluation in ADAS system to **achieve 7% MCC gain**.
- Performed exploratory time-series analysis on multivariate datasets from 30+ vehicles to identify potential edge cases.
- Analysed temporal features causing misclassifications to enhancing model interpretability and error analysis.
- Applied diverse pre-processing and augmentation techniques to improve model robustness against signal variability.

Research Assistant, *Universität Siegen* 04/2024 – 11/2024 | Germany

- Developed distributed learning pipelines for semantic segmentation across simulated edge devices, optimizing real-time vision models for constrained environments using DeepLabV3 + MobileNetV2 architecture.
- Benchmarked federated learning algorithms (FedAvg, FedProx, FedDyn) under diverse non-IID data conditions, evaluating scalability, robustness, and accuracy trade-offs.
- Integrated a MAN regularizer for activation control, **improving model generalization by 2% in decentralized systems**.

Senior Maintenance Engineer, *Vedanta Limited* 07/2021 – 07/2022 | India

- Led maintenance initiatives for heavy machinery, coordinating with cross-functional teams and leveraging ERP SAP PM/MM for planning and tracking, **improving equipment availability by 20%**.
- Performed Root Cause Analysis (RCA) for long breakdown machinery and documented findings and corrective actions within SAP maintenance notifications to ensure traceability.
- Collaborated with cross-functional teams and external partners to implement process improvements and optimized workflows.

PROJECTS

arXiv Research Assistant RAG System

- Built a production-style arXiv research assistant using FastAPI, PostgreSQL, OpenSearch, Airflow, and Docker, enabling automated ingestion, storage, and retrieval of academic papers.
- Engineered **hybrid BM25 + dense retrieval, boosting precision by 30% and recall@10 by 15%**, outperforming dense-only based retrieval.
- Implemented Agentic RAG workflows with LangGraph (guardrails, multi-attempr retrieval, query rewriting), reducing hallucinations.
- Deployed FastAPI endpoints backed by Ollama and Redis Caching, scaling Q&A throughput with lower latency; integrated Langfuse observability for full pipeline tracing leading to faster debugging time.

TikTok Keyword Trend Detection System

- Developed a scalable multimodal data pipeline using Python, Apache Airflow, and TimescaleDB to automate TikTok trend detection, boosting data processing efficiency and scalability.
- Applied statistical modeling and AI-based keyword extraction on multi-modal data to identify trends and generate actionable insights.
- Built interactive dashboards with Tableau and Streamlit to visualize trend analytics, driving data-based decisions with high reliability.

EDUCATION

Master of Science in Mechatronics, *Universität Siegen* 10/2022 – 09/2025 | Germany

Bachelor of Technology in Mechanical, *IIT Dhanbad* 07/2017 – 05/2021 | India

ACHIEVEMENTS

DEUTSCHLAND STIPENDIUM

Received DEUTSCHLAND STIPENDIUM for excellent grades in academics on Nov 2023.