MERCEL VUBANGSI

+1-641-233-1997 ⋄ Bear, DE
vmercel@outlook.fr ⋄ linkedin.com/in/mercel-vubangsi-68109b23a

PROFESSIONAL SUMMARY

Senior MLOps & AI Ops Engineer with over 10 years of experience in designing scalable AI systems, data-driven solutions, and MLOps practices across healthcare, finance, government, and robotics sectors. Expertise in developing, deploying, managing, and monitoring ML models in production environments using Python, TensorFlow, PyTorch, and scikit-learn. Proficient in building automated ML pipelines with CI/CD tools like Jenkins, GitHub Actions, and GitLab CI/CD, infrastructure management on cloud platforms including AWS, Azure, and GCP, and utilizing containerization with Docker and Kubernetes. Skilled in applying agentic AI and LLMs to automate IT operations, including log analysis, incident response, and real-time system monitoring. Proven track record in leading cross-functional teams to productionize ML models, integrate them with production systems, and implement monitoring and alerting systems. Eager to apply strong analytical, problem-solving, and communication skills to drive impactful machine learning projects in a federal government setting, with exposure to HPC compute and storage for scientific data tools.

TECHNICAL SKILLS

Programming	Python, Java, Scala, C++, Node.js, MATLAB
Machine Learning	TensorFlow, PyTorch, scikit-learn, XGBoost, Transformer Architectures (BERT, GPT), Distributed Training (Horovod, Ray), Agentic AI (LangGraph, LangChain, AutoGen), Natural Language Processing (NLP), Reinforcement Learning
Data Science	Pandas, NumPy, Scikit-learn, Bayesian Methods, Time-series Analysis, SHAP, A/B Testing
Databases	Relational (PostgreSQL, MySQL), NoSQL (MongoDB), Snowflake, Elasticsearch, Vector DBs (FAISS)
Software Engineering	Microservices, REST/GraphQL APIs, Django, Event-Driven Design, System Design (Load Balancing, Caching), Spring Boot, FastAPI
IT Operations & Observability	Elasticsearch (Log Analysis), Prometheus, Grafana, ChatOps (Slack API, Microsoft Teams API)
Signal Processing	MATLAB, SciPy, FFT, Wavelet Transforms, Real-time Control Systems
Quantum Computing	Qiskit, Quantum Machine Learning, Cirq
Cloud & DevOps	AWS (SageMaker, ECS, Lambda, S3, EFS), Azure (ML services, compute, storage), GCP (Vertex AI, BigQuery), Kubernetes, Docker, Terraform, CI/CD (Jenkins, GitHub Actions, GitLab CI/CD)
Big Data	Apache Spark (Optimization, Delta Lake), Kafka (Event Streaming), Snowflake, Elasticsearch, Vector DBs (FAISS)
MLOps	Model Monitoring (Evidently), MLflow, Kubeflow, TF Serving, SageMaker, DataRobot

EXPERIENCE

Senior MLOps Engineer 07/2022 - 06/2025
AI and Robotics Institute, Nicosia, Cyprus *(Hybrid)*

- Led the development of an MLOps platform using AWS SageMaker, MLflow, and Kubeflow, automating ML pipelines for data ingestion, preprocessing, model training, validation, and deployment with CI/CD practices via Jenkins and GitHub Actions, reducing deployment time from 2 weeks to 2 days for 50+ data scientists in healthcare and robotics teams.
- Productionized ML models by integrating trained TensorFlow and PyTorch models with production systems, ensuring scalability and reliability on AWS and Azure cloud platforms.

- Designed and managed infrastructure for ML workloads using Docker for containerization and Kubernetes for orchestration, optimizing performance and efficiency on AWS compute, storage (S3, EFS), and networking services.
- Implemented monitoring and alerting systems with Evidently and Prometheus to track ML model performance in production, automating alerts for drift and anomalies, improving model reliability by 30%.
- Built automated data pipelines with Apache Spark for data engineering, handling relational (PostgreSQL, MySQL) and NoSQL (MongoDB) databases, processing 1TB/day for financial risk analysis with XGBoost models achieving 94% precision on HPC-optimized clusters.
- Developed an autonomous agentic AI system using LangChain and LangGraph for real-time log analysis, anomaly detection, and incident response. The system processed operational data streams, automated alerts to Slack, and improved mean time to resolution (MTTR) by 25%.
- Integrated Elasticsearch with AI agents to extract, summarize, and visualize operational logs and metrics, creating dashboards for observability that reduced manual monitoring efforts by 40%.
- Created a LightGBM model for a federal government project, deploying a blockchain and AI-enabled digital ID system via FastAPI, integrating with production databases (MongoDB, PostgreSQL), reducing identity fraud by 18% and handling 500K requests/day.
- Designed event-driven pipelines using Kafka and microservices for robotic control, incorporating DevOps principles and CI/CD with GitLab CI/CD, enhancing fault recovery by 35% and scalability.
- Developed a Java Spring Boot backend for a financial application, integrating scikit-learn models and NLP for text analysis, deployed with Docker on Azure, achieving sub-200ms latency.
- Developed custom Plotly dashboards to visualize system health metrics and AI model performance, providing actionable operational insights to cross-functional teams.

Data Scientist / MLOps Specialist 03/2022 - 07/2022
North Prime Group, Nicosia, Cyprus *(Onsite)*

- Developed time-series forecasting models with PyTorch, building ML pipelines for data preprocessing and validation using Apache Spark and Delta Lake, optimizing for 500GB+ financial data and achieving 93% accuracy.
- Implemented CI/CD workflows with GitHub Actions for deploying ML models to production, including infrastructure setup on Azure and monitoring with MLflow.
- Designed REST APIs with FastAPI for real-time insights, integrating relational databases (MySQL) and NoSQL (MongoDB), supporting 100K+ queries/day with low latency.
- Conducted statistical analysis with SHAP and Bayesian methods to interpret and optimize ML models, automating workflows for reproducibility.
- Built quantum-inspired ML models with Qiskit, deploying on cloud platforms with Docker, enhancing risk analysis predictions by 15%.
- Deployed NLP-based healthcare chatbots using BERT and Hugging Face, leveraging AWS and GCP services for scalable document and text processing.
- Integrated Gemini and Vertex AI Agent Builder with Dialogflow to create an intelligent complaint resolution pipeline for e-commerce, utilizing extensions and vector search for enhanced customer support and faster resolution times.

Machine Learning Engineer 12/2019 - 03/2022
Advanced Analytics, Bamenda, Cameroon *(Onsite)*

- Scaled TensorFlow ensemble models on GCP Vertex AI and AWS SageMaker, building end-to-end ML pipelines with CI/CD using Jenkins, processing 10M+ records/month for educational analytics.
- Deployed NLP-based healthcare chatbots using BERT on AWS Lambda, productionizing models with integration to production systems and monitoring for performance.

- Developed data pipelines with Apache Spark, incorporating data storage in PostgreSQL and MongoDB, enhancing feature engineering efficiency by 40%.
- Built Django-based analytics portals with real-time feedback, automating ML tasks and optimizing infrastructure with Kubernetes.
- Designed scalable ML systems for government-related educational projects, ensuring compliance and efficiency.

Data Scientist/ ML Engineer 09/2013 - 12/2019
Quantum Systems Electronics and Signal Processing Lab, University of Dschang, Cameroon *(Hybrid)*

- Developed early ML pipelines in Python for quantum signal processing, and structure/property linkages in materials science using NumPy, Pandas, and scikit-learn for data preprocessing and model training, laying foundation for production deployment.
- Implemented data storage solutions with MySQL and MongoDB for experimental datasets, automating ingestion and preprocessing tasks.
- Optimized MATLAB and Python scripts for signal denoising with wavelet transforms, integrating ML algorithms to improve accuracy by 20%.
- Built Java and Scala-based simulations for quantum systems, incorporating DevOps practices for version control and deployment.
- Developed C++ algorithms for real-time analysis, with experience in containerization using Docker for reproducible environments.
- Migrated pipelines to Delta Lake, reducing costs by 35%, and set up basic monitoring for model performance.

EDUCATION

B.S. Physics — University of Dschang, Dschang, Cameroon 2008 - 2009
- Focused on Search for minimal decoherence pathways for quantum bits for quantum computing

M.S. Computational Physics — University of Dschang, Dschang, Cameroon 2009 - 2013
- Focused on Search for minimal decoherence pathways for quantum bits for quantum computing

Ph.D. Computational Physics — University of Dschang, Dschang, Cameroon 2013 - 2017
- Thesis: "Quantum Simulations for Large-Scale Dynamical Systems using HPC" (Published in Journal of Mathematical Physics)

B.S. Software Engineering — University of Bamenda, Bamenda, Cameroon 2018 - 2020

M.S. Artificial Intelligence Engineering — Near East University, Nicosia, Cyprus 2022 - 2023
- Focused on Deep Learning and Distributed Systems

M.S. Computer Science — Maharishi International University, Fairfield, IA, USA 2024 - 2027
- On-campus coursework completed; remaining courses in distance education (DE) mode

PROJECTS

- **Autonomous IT Operations Agent**: Designed a personal project using LangChain to create agents that query Elasticsearch logs, detect anomalies using LLMs, and automatically post alerts to a Microsoft Teams channel using the Teams API. Visualized operational trends with Matplotlib.
- **Distributed Model Training**: Implemented Horovod + PyTorch on AWS EC2 with Kubernetes for HPC, training 100M-parameter models 3x faster, incorporating CI/CD and monitoring.
- **Vector Search Engine**: Built FAISS-based system with MongoDB integration, improving relevance by 30% through optimized ML pipelines.
- **Agentic Healthcare Chatbot**: Created with LangGraph and NLP, deployed on GCP with Azure ML services for hybrid cloud, processing 20K+ queries/day.

- Built Django REST backend for education platform with Celery, integrating PostgreSQL and MLflow for model management.
- Developed Java/Scala backend for flight booking app, using Deeplearning4j and DataRobot for ML insights, deployed with Docker on AWS.

ACHIEVEMENTS

- Speaker at AI and IoT Conference 2023: "Scaling Transformer Models for Real-Time Applications."
- Best Paper Award, Quantum Electronics Symposium 2022, for "Quantum Signal Processing with Machine Learning."