DAVID ZAHEMEN YEBOAH

AI Engineer & Software Developer

Accra, Ghana | davidzahemenyeboah@gmail.com | Portfolio | GitHub | +233598020802

PROFESSIONAL SUMMARY

AI/ML Engineer specializing in temporal sequence modeling and predictive system architecture. Expert in designing robust ML pipelines for noisy, low-resource environments (LMICs) with a focus on Transformer-based forecasting and Bayesian uncertainty estimation. Proficient in end-to-end deployment including quantized inference (ONNX), MLOps on AWS/GCP, and high-performance frontend integration. Proven track record in translating complex multi-modal data into actionable predictive insights for high-stakes operational environments.

CORE COMPETENCIES

Modeling & Architecture: Time-Series Forecasting, Transformers, Bayesian Calibration (Monte Carlo), Supervised Contrastive Learning (SupCon), RAG, Agentic Decision Logic.
MLOps & Engineering: PyTorch, TensorFlow, Hugging Face (PEFT/LoRA), ONNX Runtime, SageMaker, Vertex AI, ETL Pipeline Design (SageMaker/GitHub Actions), Anomaly Detection.
Development & Deployment: FastAPI, Next.js, React, Firebase, Supabase, Vercel, Render, Docker, CI/CD, Git, High-Performance Visualization (ECharts/Waveform).

EDUCATION

BSc. Computer Science — Expected June 2026
Kwame Nkrumah University of Science and Technology (KNUST) — Kumasi, Ghana

- Cumulative Weighted Average: 78.0% (approx. 3.9/4.0 GPA equivalent).
- Research & Projects Committee Member, KNUST AI & Data Science Club.
- Coursework: Statistics, Multivariate Calculus, Linear Algebra, DSA, SQL, OOP.

EMPLOYMENT HISTORY

ML Engineer — August 2025 – Present
Asaphic Ltd. — Remote (Nigeria)

- Architected deep learning models for real-time audio source separation and signal enhancement.
- Engineered automated MLOps pipelines on AWS (SageMaker, Lambda, S3) for scalable, low-latency inference.
- Implemented adaptive tuning methods to reduce cloud inference costs by 35% while maintaining audio fidelity.

Data Science Intern — November 2024 – December 2024
Boston Consulting Group (BCGX) — Remote (US)

- Developed multi-GPU training pipelines for 3B-parameter models, reducing training time by 65% via DDP.
- Orchestrated financial data extraction pipelines, processing 120M+ tokens from complex PDF/earnings reports.
- Applied memory-saving techniques (4/8-bit quantization, LoRA) to reduce VRAM utilization by 70% during fine-tuning.

TECHNICAL PROJECT PORTFOLIO

GhSL Sign2Text | Temporal Sequence Modeling & Uncertainty Estimation

- Developed a temporal Transformer architecture for multi-modal sequence translation, achieving 74.7% accuracy.
- Engineered a data-quality gating layer using **motion-energy thresholds** ($\tau = 7.5e^{-4}$) and **torso-centric normalization** to handle inconsistent camera intrinsics in low-resource environments.
- Implemented **Bayesian confidence calibration** and **DTW (Dynamic Time Warping)** re-ranking to manage prediction uncertainty and ambiguity in real-time noisy input streams.
- Optimized inference via ONNX and FastAPI, achieving sub-200ms latency for edge-compatible deployment.

Robin AI | Agentic Decision Infrastructure & Autonomous Workflows

- Architected a multi-agent system with **"Ask → Apply" autonomous execution flows**, utilizing RAG to contextualize system states and reducing workflow redundancy by 75%.
- Engineered a high-reliability **Virtual File System (VFS)** with atomic checkpoint/rollback protocols and RLS-backed persistence to ensure 99.9% state consistency during automated system mutations.
- Developed infrastructure for real-time artifact streaming and **observer-pattern state management**, enabling low-latency feedback loops between agentic triggers and production environments.

FinSightAI | Fine-tuned Predictive Financial Model

- Developed a financial LLM tuned on 75K+ conversations using context-aware chunking and text-augmented pipelines.
- Quantized models for real-time deployment on consumer-grade hardware, reducing inference latency by 60%.

BioQuery | NASA Space Apps Winner (Galactic Impact Award, NASA Space Apps 2025)

- Created an AI research engine connecting NASA bioscience data with mission outcomes via Knowledge Graphs.
- Integrated Cohere embeddings and RAG framework to automate discovery in high-dimensional scientific datasets.

LEADERSHIP & RECOGNITION

- **Committee Lead:** Spearheading AI research and student-led initiatives for the KNUST AI & Data Science Club.
- **Community Activist:** Active member of Google Developer Student Club (GDSC) and KNUST Robotics Club.