

THOMAS AMOGOLLA TSUMA

Curriculum Vitae — PhD Applicant, Computational Drug Design / AI for Science

Malaysia • +60 19 762 3483 • tommytsuma7@gmail.com • github.com/TomiTsuma • linkedin.com/in/tomitsuma • [portfolio](#)

RESEARCH INTERESTS

Generative and graph neural architectures for de novo molecular design; diffusion and equivariant models for structure-based drug design; physics-informed neural networks; imitation learning for physical control systems.

EDUCATION

MSc, Data Science — INTI International University, Malaysia (Expected 2026)

Thesis research: Graph neural networks for drug discovery; massive activations in attention-based networks.

BSc, Computer Science (Computational Science), Second Class Upper Honours — Multimedia University of Kenya (2017–2021)

Undergraduate projects: Lung cancer detection in CT scans (CNN); hate speech detection using NLP.

PUBLICATIONS & MANUSCRIPTS

- Tsuma, T., Amin, Z. "A Systematic Scoping Review of Diffusion-Based De Novo Molecular Generation: Architectures, Evaluation, and Open Challenges for Computational Drug Discovery." Manuscript in preparation for submission to Briefings in Bioinformatics (Oxford University Press). PRISMA-ScR methodology; concept-centric synthesis of 90+ primary sources spanning GNN, transformer, GAN, autoencoder, flow-matching, and diffusion architectures for molecular generation, with emphasis on equivariance, conditioning mechanisms, and evaluation standardization.
- "Coca-Cola Bottle Image Recognition." Analytics Vidhya Data Science Blogathon — ranked Top 30 of 400+ global submissions. Colour segmentation, transfer learning, and image classification in TensorFlow/OpenCV.
- "Lung Cancer Diagnosis: Making Machines That Think Like Human Beings." Analytics Vidhya. CNNs and transfer learning for detecting cancerous nodules in lung CT scans.
- "AI Agents for Supercharged Supply Chain and Inventory Management." Core&Outline Blog. Deep Q-Learning with shaped rewards for inventory optimization.

INDEPENDENT RESEARCH PROJECTS

Neural Imitation-Learning Control of Kinetic Plasma Instabilities — Independent research 2026

- Implemented a causal Temporal Convolutional Network + Transformer controller (~192K parameters) reproducing an arXiv imitation-learning result for stabilizing the 1D1V Vlasov–Poisson system from four sparse, noisy density sensors.
- Built the full data pipeline: semi-Lagrangian Vlasov–Poisson simulator, expert-policy trajectory generation, and a scale-aware hybrid Fourier-coefficient loss.
- Engineering context: relevant to tokamak fusion-plasma control and to partial-observability control problems more broadly.

Physics-Informed Neural Network for Binary Distillation Column Simulation — Independent research 2026

- Designed a PINN (17→4, residual skip connections) with governing-equation residuals (VLE via Newton–Raphson bubble-point solving, 16-tray MESH mass balance, tray-level energy balance) embedded directly in the loss function.
- Implemented a sigmoid curriculum transitioning training from physics-dominant to data-dominant loss weighting, targeting 95% component purity under sensor noise and feed/pressure perturbations.

RESEARCH-RELEVANT PROFESSIONAL EXPERIENCE

Machine Learning Engineer — Crop Nutrition Laboratory Services (Cropnuts) Sept 2022 – Present | Limuru, Kenya

- Designed pix2pix GAN models generating synthetic Alpha-ATR spectra, enabling analysis of 60,000+ soil spectra and securing a USD 3M contract with KALRO — direct generative-modeling research applied to a novel scientific data domain.
- Designed an RNN autoencoder for outlier detection in spectral data prior to inference.
- Built and published a model evaluation framework (RMSE, R^2 , RSC, slope, kurtosis, skewness) for benchmarking model iterations.

TECHNICAL SKILLS

ML/DL frameworks: PyTorch, TensorFlow, Keras, scikit-learn

Architectures: GNNs, GANs, VAEs/autoencoders, Transformers, diffusion/equivariant models (literature + implementation), TCNs, PINNs

Scientific computing: NumPy, SciPy, ODE/PDE simulation (semi-Lagrangian solvers), Newton-Raphson methods

Languages: Python, SQL, R, C, C++, JavaScript

Research methods: PRISMA-ScR systematic review methodology, thematic/concept-centric literature synthesis

CERTIFICATIONS & HONORS

- AWS Certified Machine Learning – Specialty (MLS-C01), 2024
- Microsoft Certified: Azure AI Fundamentals (AI-900)
- Top 30 of 400+ global submissions, Analytics Vidhya Data Science Blogathon