

THOMAS AMOGOLLA TSUMA

Machine Learning Engineer → Research Scientist, Generative & Graph Models

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

SUMMARY

ML engineer with 5+ years shipping production generative and graph-based models (GANs, RNN autoencoders), now building an independent research track spanning generative molecular design, physics-informed learning, and control theory. First-author systematic review of diffusion-based de novo molecular generation in preparation for Briefings in Bioinformatics (Oxford University Press). Comfortable moving between reading and reimplementing theory-heavy papers and shipping models into production pipelines.

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, with emphasis on equivariance, conditioning mechanisms, and evaluation standardization for structure- and ligand-based drug design.

SELECTED INDEPENDENT RESEARCH

Neural Imitation-Learning Control of Kinetic Plasma Instabilities — Independent implementation of an arXiv imitation-learning control paper 2026

- Reproduced a provable imitation-learning result for stabilizing the 1D1V Vlasov–Poisson system from four sparse, noisy sensors, under partial observability.
- Implemented a causal TCN (5 dilated residual blocks, receptive field 125) feeding a 2-layer Transformer encoder (~192K parameters total), trained via behavior cloning against a field-cancellation expert policy.
- Built the full simulation-to-training pipeline: semi-Lagrangian Vlasov–Poisson solver, expert-trajectory generation, and a scale-aware hybrid Fourier-coefficient loss handling target magnitudes spanning several orders of magnitude.

Physics-Informed Neural Network for Binary Distillation Column Simulation — Independent research, chemical-engineering physics 2026

- Embedded governing physics (VLE via Newton–Raphson bubble-point solving, 16-tray MESH component/energy balances) directly as residual penalty terms in the training loss, rather than treating the network as a black-box regressor.
- Designed a sigmoid curriculum shifting the loss from physics-dominant to data-dominant over training, plus residual skip connections to stabilize gradients under large early-epoch physics residuals.

RESEARCH-RELEVANT EXPERIENCE

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

- Designed pix2pix GAN models to generate synthetic Alpha-ATR spectra — a novel-domain generative-modeling application that enabled analysis of 60,000+ soil spectra and secured a USD 3M contract with KALRO.
- Designed an RNN autoencoder for outlier detection in spectral time-series data prior to inference.
- Built a model evaluation framework (RMSE, R^2 , RSC, slope, kurtosis, skewness) for systematic benchmarking across model iterations — the same evaluation-standardization instinct applied at the literature level in the scoping review.
- Owned an automated retraining pipeline (DVC, S3, Azure ML) with a 4-day update turnaround — production infrastructure for continuously retrained generative models.

Co-Founder & ML Lead — Core&Outline

2024 – Present

- Designed graph-based reasoning and LLM-integration architecture for a data analytics platform; built a reinforcement-learning recommendation engine from behavioral data.

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)

TECHNICAL SKILLS

Generative & graph models: GNNs, GANs, VAEs/autoencoders, diffusion & equivariant models (literature + implementation), Transformers, TCNs

Frameworks: PyTorch, TensorFlow, Keras, scikit-learn, NumPy, SciPy

Research methods: PRISMA-ScR systematic review methodology, PDE/ODE simulation, physics-informed loss design, behavior cloning

MLOps & deployment: Azure ML, AWS SageMaker/Lambda/Glue, DVC, Docker, API deployment

Certifications: AWS Certified Machine Learning – Specialty (2024); Microsoft Azure AI Fundamentals