Mugume Twinamatsiko Atwine

Health Data Scientist & ML Engineer, based in Kampala, Uganda

twinmugume@gmail.com atwine.github.io ORCID 0009-0007-8183-9412 github.com/atwine LinkedIn

Summary

I work on language technology and data infrastructure for health in Sub-Saharan Africa. I build and evaluate bilingual English–Luganda assistants that answer only from expert-approved or official sources, and open tools that use language models to help research projects harmonise their data.

Experience

2021 – present

Data Science Specialist, eLwazi Open Data Science Platform

Uganda Virus Research Institute, Entebbe · NIH-funded DS-I Africa program

- Leads the eLwazi Data Models, Storage, Registries and Catalogue Working Group, which has had between about 20 and 50 members from DS-I Africa projects.
- Co-developer of the Metadata Harmonisation Tool, an open-source application that uses large language models to map study variables to a target codebook.
- Co-author (ideas and editing) of papers on the institute's bioinformatics computing service (*Bioinformatics*, 2026; *PLOS Computational Biology*, 2025).

2019 – present

Senior Data Scientist and ML Technical Lead

Africa Center of Excellence in Bioinformatics and Data-Intensive Sciences (ACE Uganda), Infectious Diseases Institute, Makerere University, Kampala

- HEAL (2023 – present): team lead for a bilingual assistant that grounds GPT-4-turbo in Uganda's disease-surveillance guidelines, with a fine-tuned English–Luganda translation model. Supported by the Gates Foundation and Code for Africa. First author of the preprint.
- Lacuna Luganda Health Corpus (funded by the Lacuna Fund): first author of the dataset paper, under review, describing 38,644 human-reviewed English–Luganda health sentence pairs released on Hugging Face.
- AfriNorm (2026): lead developer of an open toolkit for restoring diacritics in African-language text, evaluated on six languages. First author of the paper draft.
- Started two studies on Luganda machine translation (2026): how open models cope with speech that mixes Luganda and English (project lead), and predicting which machine translations need a human reviewer.
- Built a model to estimate HIV risk among young people from behavioural data (SIDA Risk AI).
- Developed a time-to-event model for the onset of cryptococcal meningitis in people living with HIV.
- Built a retrieval-augmented question-answering system over long technical manuals.

2020 – present

Capacity Building Lead and Research Assistant Lecturer

ACE Uganda, Makerere University, Kampala

- Developed and taught a Big Data Analytics course for MSc and PhD students.
- Supervised student research projects.

2021 – 2023

Natural Language Specialist (contract)

Academy for Health Innovation, Infectious Diseases Institute, Kampala

- Led the technical build of a clinical decision support system for health facilities.

2021 – 2023

Lead Data Scientist, COVID-19 response

Makerere University, Kampala

- Technical lead for COAST, an English–Luganda COVID-19 chatbot that answers only with expert-approved responses. Built with Rasa and deployed on Kubernetes. Co-first author of the resulting paper in *PLOS Digital Health* (2026).

Education

2011 – 2013

MSc Computer Networking

Limkokwing University of Creative Technology, Malaysia

- Research on workload optimisation and load balancing in virtualised cloud data centres, published as four papers (listed below).

2009 – 2011

BSc Computer Science

Mbarara University of Science and Technology, Uganda

Certificates

Computational Thinking and Big Data, University of Adelaide (edX), 2017

Project Management, University of Adelaide (edX), 2018

Evidence-Based Project Management, Australian National University (edX), 2018

Publications

Journal articles

1. **Atwine MT***, Jjingo D*, Nsubuga M, Serunjogi R, Mbabaali I, Lujumba I, Byansi D, Kintu T, Galiwango R, Kebirungi G. A bi-lingual chatbot implementation for pandemic response using the transformer-based approach. *PLOS Digital Health*. 2026;5(4):e0001256. doi:10.1371/journal.pdig.0001256. *Co-first authors.

2. Lukyamuzi E, Kimbowa Wamala T, Ssekagiri A, Galiwango R, Kebirungi G, **Atwine M**, Nsubuga M, Maslamoney S, Panji S, Mulder NJ, Jjingo D, Kayondo J. A scalable HPC framework for bioinformatics in resource-limited settings: design principles, implementation, and sustainability from the UVRI experience. *Bioinformatics*. 2026;42(4):btag149. doi:10.1093/bioinformatics/btag149

3. Jjingo D, Walakira A, Hashim S, et al. (**Atwine MT**, 36th of 67 authors). Pathways, outputs and impact of NIH-supported bioinformatics and genomics graduate trainees in Africa. *Briefings in Bioinformatics*. 2026;27(1):bbag046. doi:10.1093/bib/bbag046

4. Galiwango R, Whalen CJ, Kebirungi G, **Atwine MT**, et al. (26 authors). Ten simple rules for building and maintaining sustainable high-performance computing infrastructure for research in resource-limited settings. *PLOS Computational Biology*. 2025;21(9):e1013481. doi:10.1371/journal.pcbi.1013481

Preprints, papers under review and in preparation

5. **Atwine MT**, Kintu TM, Lujumba I, Mbabali I, Muwonge K, Muwonge L, Nakiire L, Ntono V, Nsubuga M, Galiwango RM, Lunkuse S, Kebirungi G, Jjingo D. Development and evaluation of HEAL: a bilingual English–Luganda RAG chatbot for disease surveillance in Uganda. *Research Square* [preprint]. 2026. doi:10.21203/rs.3.rs-10357473/v1. Under review.

6. **Atwine MT**, Kintu TM, Kebirungi G, Muwonge L, Muwonge K, Lujumba I, Mutegeki H, Kyagambiddwa T, Jjingo D. A human-reviewed English–Luganda parallel corpus for natural language application in biomedical communication. Under review. Dataset: doi:10.57967/hf/10682

7. **Atwine MT**, Muwonge L, Kintu T, Muwonge K, Galiwango R, Kebirungi G, Jjingo D. AfriNorm: an open, extensible system for diacritic restoration in African languages. In preparation.

8. Adams R, Adeleke F, Junck L, et al. (**Atwine MT**, 52nd of 75 authors). Mapping the potentials and limitations of using generative AI technologies to address socio-economic challenges in LMICs [version 1]. *VeriXiv* [preprint]. 2025. doi:10.12688/verixiv.948.1

Earlier publications (MSc research, published as Twinamatsiko AM)

9. **Twinamatsiko AM**, Abdulhussein ANA, Joshi JH, Lashkari AH, Sadeghi M. Dynamic workload performance optimization model for multiple-tenancy in cloud systems. *International Journal of Computer Science Issues*. 2013;10(6):271.

10. Parhizkar B, Abdulhussein ANA, Joshi JH, **Twinamatsiko AM**. A common factors analysis on cloud computing models. *International Journal of Computer Science Issues*. 2013;10(2 Pt 1):523.

11. Joshi JH, Abdulhussein ANA, **Twinamatsiko AM**, Lashkari AH, Sadeghi M. An enhanced quality model for IaaS provider with non-quantifiable key performance indicators. *International Journal of Computer and Communication Engineering*. 2014;3(2).

12. Abdulhussein ANA, Joshi JH, **Twinamatsiko AM**, Lashkari AH, Sadeghi M. An efficient load balancing algorithm for virtualized cloud data centers. In: *Recent Advances in Electrical and Computer Engineering* (conference proceedings).

Software and data

Lacuna Luganda Health Corpus. 38,644 English–Luganda health sentence pairs with a review history for every pair and fixed training, development and test splits. Hugging Face, 2026. CC BY 4.0. doi:10.57967/hf/10682

Metadata Harmonisation Tool, version 0.4.3. Open-source application that uses large language models to map study variables to a target codebook. Zenodo, 2026. doi:10.5281/zenodo.18978529; code at github.com/atwine/Metadata-Harmonisation-Tool

COAST chatbot dataset. Expert-approved COVID-19 questions and answers in English and Luganda. Zenodo, 2025. doi:10.5281/zenodo.14608441

Technical skills

Language models and NLP: retrieval-augmented generation, the GPT-4 API, fine-tuning translation models (mBART) and byte-level models (ByT5), machine translation evaluation, Rasa

Machine learning: classification, time-to-event (survival) models, model evaluation with domain experts

Engineering: Python, PyTorch, Hugging Face Transformers, Streamlit, Flask, React, Docker, Kubernetes (K3s)

Data: data harmonisation, FAIR metadata, research data pipelines

References

Available on request.