

Sakayo Toadoum Sari

Multilingual NLP Researcher & PhD Candidate — CRIL,
Artois University

62300 Lens, France
+235 66 04 90 94
sakayo@cril.fr
Toadoum.github.io
toadoum
Toadoum
0000-0003-2617-081X

Professional Summary

Multilingual NLP researcher and engineer specializing in low-resource language technologies across machine translation, ASR/TTS, and human-centered evaluation. Experienced in responsible AI, dataset design, and field deployments within humanitarian contexts. Former UNICEF Chad consultant leading national RCCE actor mapping and data collection. Beginning a PhD at CRIL (Université d'Artois) on discourse-level semantics, intertextuality, and causal reasoning. Passionate about transparent, inclusive language technologies for the Global South.

Skills

Programming	Python, PyTorch, TensorFlow, Keras, scikit-learn, Hugging Face, Gradio, Docker, Git/GitHub
NLP/ML	MT (NMT), ASR, TTS, tokenization, data curation, evaluation & human studies, interpretability, causal/discourse analysis
Data	Large-scale preprocessing, annotation guidelines, benchmarks, quality assurance, error analysis
Languages	Fluent: English, French
Other	Teaching/training, technical writing, community engagement (Masakhane, Africa NLP)

Experience

Oct 2023 – Present	NLP Engineer, Batazia, Rotterdam, Netherlands (Remote) <ul style="list-style-type: none">Develop models, algorithms, and tools tailored to African languages; integrate into products with strong QA.Lead on ethics, data governance, and privacy-compliant pipelines; define reliability and scalability criteria.
Jul 2025 – Dec 2025	Consultant — RCCE, UNICEF, N'Djamena, Chad <ul style="list-style-type: none">Led national mapping of RCCE actors (CBOs, influencers) across multiple regions; facilitated validation workshops.Designed mixed-methods instruments (surveys, interviews, FGDs) and produced final report & interactive actor map.
Apr 2025 – Jul 2025	Machine Learning Engineer, CDIAL, Lagos, Nigeria (Remote) <ul style="list-style-type: none">Trained SOTA ASR & TTS models for African languages; coordinated large-scale human evaluations.Diagnosed data distribution shifts impacting real-world performance; proposed mitigations.
Jun 2023 – Aug 2023	Summer Research Intern (NLP), EPFL, Lausanne, Switzerland <ul style="list-style-type: none">Modeled and predicted online polarization using NLP techniques.
Nov 2022 – Oct 2023	Lead Trainer in Data, Tec4Tchad (Simplon/Univ. N'Djamena), N'Djamena, Chad <ul style="list-style-type: none">Managed and taught Python, Intro ML, and Git/GitHub to 120+ learners; developed hands-on curricula.
Dec 2022 – Mar 2023	Junior Consultant (ML), STEM-Away, Santa Clara, USA (Remote) <ul style="list-style-type: none">Built NLP skill-extraction pipelines from job descriptions to improve recruitment matching.
Jun 2022 – Sep 2022	Project Lead Intern (ML), STEM-Away, Santa Clara, USA (Remote) <ul style="list-style-type: none">Question recommendation/classification for forums using Sentence-BERT and vector search (Pinecone).
Jul 2019 – Jan 2020	Web Dev & ML Intern, Limbe City Council, Limbe, Cameroon <ul style="list-style-type: none">Built DBMS + web interface for civil status documents (birth, death, marriage) management.

Education

- Oct 2025 – Sep 2028 **PhD in Artificial Intelligence**, *CRIL Lab, Université d'Artois*, Lens, France
 Thesis: *Intertextuality and Causalities: Towards a Semantic Analysis of Discourses*.
 Supervisors: Pr. Lakhdar Saïs, Pr. Said Jabbour, Dr. Fabien Delorme.
- Dec 2021 – Jun 2023 **African Master's of Machine Intelligence (AMMI)**, *AIMS*, Mbour, Senegal
 Thesis: *Ngambay-French Neural Machine Translation*.
 Funding: Google & Meta. Supervisor: Dr. Angela Fan.
- Oct 2018 – Feb 2020 **Co-operative Master in Industrial Mathematics (Data Science)**, *AIMS*, Limbe, Cameroon
 Thesis: *Hybrid LSTM + AR Models for Birth Prediction*.
 Funding: MasterCard Foundation. Supervisor: Pr. Marcellin Atemkeng.
- Dec 2017 – Nov 2019 **Master in Physics (Electronics, Automation)**, *University of N'Gaoundéré*, Cameroon
 Thesis: *Optimization of Energy Transfer in a Photovoltaic System*.

Publications (selected)

- 2025 **Pretraining Strategies using Monolingual and Parallel Data for Low-Resource MT**. AfricaNLP. PDF
- 2023 **Ngambay-French Neural Machine Translation (sba-Fr)**. NLP4TIA @ RANLP. arXiv
- 2022 **Error-Correction DNNs for South African Wildlife-Vehicle Collision Data**. MDPI. Link
- 2023 **MASAKHANEWS: News Topic Classification for African Languages**. AfricaNLP @ ICLR. arXiv
- 2023 **AfriMTE & AfriCOMET: Empowering COMET for African Languages**. Preprint. arXiv

Scholarships & Awards

- AMMI Full scholarship — Google & Meta
- AIMS Full scholarship — MasterCard Foundation

Peer-Review Activities

- IGI Global Reviewer (2023, 2024): IDs f866d66b-ff53-4087-96f2-dabeb594eb02; 30802a02-82e9-41a0-8a8a-9cdd1b653550

References

Dr. Marcellin Atemkeng
 Centre for Radio Astronomy Techniques & Technologies (RATT), Rhodes University, South Africa.
 m.atemkeng@ru.ac.za
 Tel: (+237) 67 753 90 04

Dr. Rémi Philippe Lebreton
 Research Scientist, EPFL, Switzerland.
 remi.lebret@epfl.ch
 Tel: +41 21 693 12 57

Dr. Ismaila Seck
 CTO, Lengo AI, Dakar, Senegal.
 iseck@aimsammi.org