Sakayo Toadoum Sari

62300 Lens, Hauts-de-France | sakayo@cril.fr | +33 07 59 09 55 83 | Toadoum.github.io

linkedin.com/in/toadoum | github.com/Toadoum | orcid.org/0000-0003-2617-081X

PROFESSIONAL SUMMARY

Multilingual NLP researcher and engineer with a strong background in low-resource language modeling, machine translation, and applied machine learning. Experienced in speech technologies (ASR/TTS), interpretability, human evaluation, and responsible AI. Recently worked with UNICEF Chad as a consultant in Risk Communication and Community Engagement (RCCE), leading national field research to inform inclusive, data-driven communication strategies. Currently starting a PhD at CRIL Lab (University of Artois, France), focusing on discourse-level semantics, intertextuality, and causal reasoning in text. Passionate about building transparent, community-centered language technologies for the Global South, with a publication record across African NLP and multilingual research venues.

EDUCATION

PhD in Artificial Intelligence, CRIL Lab, École Doctorale STS – Picardie, Artois University – Lens, France
October 2025 – September 2028
Thesis: *Intertextuality and Causalities: Towards a Semantic Analysis of Discourses*
Funding: *Artois University and CRIL Lab.*
Supervised by: *Pr. Lakhdar Saïs, Pr. Said Jabbour, and Dr. Fabien Delorme.*

African Master's of Machine Intelligence (AMMI), African Institute for Mathematical Sciences (AIMS) – Mbour, Senegal
December 2021 – June 2023
Thesis: *Ngambaye-French Neural Machine Translation*
Funding: *Google and Meta*
Supervised by: *Dr. Angela Fan*

Co-operative Master in Industrial Mathematics with interest in Data Sciences, African Institute for Mathematical Sciences (AIMS) – Limbe, Cameroon
October 2018 – February 2020
Thesis: *Hybrid Architecture Combining Long Short-Term Memory and Autoregressive Models: Application to Time Series Data for Birth Prediction*
Funding: *MasterCard Foundation*
Supervised by: *Pr. MARCELLIN Atemkeng*

Master in Physics, Major Electronics, Electrical engineering, Automation, University of N'Gaoundéré – N'Gaoundéré, Cameroon
December 2017 – November 2019
Thesis: *Optimization of Energy Transfer in a Photovoltaic System*
Funding: *Self-funded*
Supervised by: *Pr. EFFA Jean Yves*

RESEARCH INTERESTS

Discourse and Semantic Analysis, Intertextuality and Causality in NLP, Low-Resource and Multilingual NLP, African Languages, Speech Recognition and Synthesis (ASR/TTS), Machine Translation, Data-Centric and Inclusive AI, Human-Centered Evaluation, Responsible and Ethical AI, Language Technologies for Development and Communication

EXPERIENCE

Teaching Assistant (64h – TD Equivalent), Université d'Artois – Lens, France
October 2025 – September 2028

- Delivered 64 hours of undergraduate teaching (tutorials/TD) in the 2025–2026 academic year as part of a doctoral contract mission.

- Assisted in course delivery, exercises, and student supervision in computer science or related modules.

Consultant – Risk Communication and Community Engagement (RCCE), UNICEF – N'Djamena, Chad
21 July 2025–22 December 2025

- Led a national mapping of RCCE actors, grassroots organizations (CBOs), and local influencers across multiple regions of Chad.
- Designed and implemented qualitative and quantitative tools including surveys, interview guides, and focus group protocols.
- Conducted in-depth analysis of community dynamics, trust structures, and participatory mechanisms.
- Developed actionable recommendations to enhance UNICEF's RCCE programming and strengthen community ownership.
- Coordinated technical validation workshops with stakeholders from government, NGOs, and community representatives.
- Produced a comprehensive final report and interactive actor map for use in future coordination and programming.

Machine Learning Engineer, CDIAL – Lagos, Nigeria (Remote)   April 2025– July 2025
- Perform large scale pre-processing of NLP and speech data.
- Train state-of-the-art ASR and TTS models for African languages.
- Coordinate large scale human evaluation of model performance among users.
- Collaborate with cross-functional teams as needed to meet project goals.
- Evaluate and identify differences in data distribution that could potentially alter model performance in real-world applications.
- Stay current with recent developments and techniques in the generative Artificial Intelligence industry.

NLP Engineer, Batazia – Rotterdam, Netherlands (Remote)   October 2023 – Present
- Developing cutting-edge NLP models, algorithms, and tools tailored for African languages;
- Identifying vulnerabilities and developing a comprehensive information strategy;
- Collaborating with engineering and product teams to ensure the successful integration of NLP solutions into products;
- Ensuring the ethical and responsible use of data in compliance with industry standards and data privacy regulations;
- Implementing and maintaining rigorous quality assurance processes for NLP models and systems, including testing, evaluation, and continuous improvement;
- Ensuring that NLP solutions meet high accuracy, reliability, and scalability standards.

Independent Consultant, Digital Services & GMS Consulting SARL – Douala,   March – June 2024
Cameroon
- Teaching an Introduction to Machine Learning to professionals pursuing a Data Analyst Certificate in Douala, Cameroon, covering key ML concepts, algorithms, and practical applications.

Junior Consultant in Machine Learning, STEM-Away – Santa Clara, USA   December 2022 – March
(Remote)   2023
- Extracting skills from job descriptions using NLP techniques to enhance recruitment and job-matching systems.

Summer Research Intern in NLP, École Polytechnique Fédérale de Lausanne   June – August 2023
(EPFL) – Lausanne, Switzerland
- Modeling and predicting online polarization using NLP techniques.

Lead Trainer in Data, Tec4Tchad (Simplon/Université de Ndjamena) – Ndjamena,   November 2022 – October
Chad   2023
- Managing and teaching courses including Python programming, Introduction to Machine Learning, and Git & GitHub.

Project Lead Machine Learning Intern, STEM-Away – Santa Clara, USA (Remote)   June – September 2022
- Developing a question recommendation and classification system for a forum discussion platform using NLP techniques, leveraging Sentence-BERT and Pinecone for vector search capabilities.

Web Developer and Machine Learning Intern, Limbe City Council – Limbe,   July 2019 – January 2020
Cameroon

- Design and implement a database management system with a web interface for managing civil status documents, including birth, death, and marriage certificates.

PEER-REVIEW ACTIVITIES

- **Review Date:** 2024-01-28
  Type: Review
  Role: Reviewer
  Review Identifier: OTHER-ID: f866d66b-ff53-4087-96f2-dabeb594eb02
  Convening Organization: IGI Global (Hershey, US)
- **Review Date:** 2023-08-13
  Type: Review
  Role: Reviewer
  Review Identifier: OTHER-ID: 30802a02-82e9-41a0-8a8a-9cdd1b653550
  Convening Organization: IGI Global (Hershey, US)

SCHOLARSHIPS AND AWARDS

- African Master's of Machine Intelligence (AMMI) full scholarship, Google and Facebook Scholar
- African Institute for Mathematical Sciences (AIMS) full scholarship, MasterCard Foundation Scholar

PUBLICATIONS

- Pretraining Strategies using Monolingual and Parallel Data for Low-Resource Machine Translation, African NLP 2025, https://openreview.net/pdf?id=FSa9WBovfD
- Ngambay-French Neural Machine Translation (sba-Fr), NLP4TIA, RANLP Processing 2023, arxiv.org/abs/2308.13497
- Error Correction Based Deep Neural Networks for Modeling and Predicting South African Wildlife-Vehicle Collision Data. Journal: MDPI, mdpi.com
- MASAKHANEWS: News Topic Classification For African Languages. AfricaNLP workshop at ICLR2023, arxiv.org/abs/2304.09972
- AfriMTE and AfriCOMET: Empowering COMET to Embrace Under-resourced African Languages, preprint, arxiv.org/abs/2311.09828

ACADEMIC CONFERENCE PRESENTATIONS

- **Ngambay-French Neural Machine Translation (sba-Fr)**
  Presented at RANLP / Natural Language Processing tools and resources for translation and interpreting applications (NLP4TIA) Conference 2023
  Link to Presentation

REFERENCES

- **Prof. Lakhdar SAÏS**
  Full Professor, Centre de Recherche en Informatique de Lens (CRIL)
  Université d'Artois | France
  sais@cril.fr
  Mobile: (+33) 6 27 72 73 13
- **Prof. Said Jabbour**
  Full Professor, Centre de Recherche en Informatique de Lens (CRIL)
  Université d'Artois | France
  jabbour@cril.fr
  Mobile: (+33) 6 59 74 26 39
- **Prof. Marcellin ATEMKENG**
  Centre for Radio Astronomy Techniques & Technologies (RATT), Department of Mathematics, SA
  Rhodes University | South Africa
  m.atemkeng@ru.ac.za