

Idris Abdulmumin, PhD

African NLP · Machine Translation · Speech & Language Resources

✉ contact@abumafrim.com

🌐 abumafrim.com

🌐 [linkedin.com/in/abumafrim](https://www.linkedin.com/in/abumafrim)

📞 +27 (0) 67-191-5773

🐙 github.com/abumafrim

🆔 orcid.org/0000-0002-3795-8381

Research Profile

Postdoctoral researcher and educator specialising in Natural Language Processing for African and other low-resource languages. My work spans machine translation, automatic speech recognition, and the analysis of sentiment, emotion, hate, and offensive language, with a strong emphasis on building openly available datasets, benchmarks, and models that serve under-represented communities. I lead and co-organise large multilingual research efforts across SemEval, WMT, IWSLT, and AfricaNLP, and supervise postgraduate research at several African institutions.

Education

- Jan. 2023 **Ph.D. Computer Science, Bayero University, Kano.**
Thesis title: *Domain-Aware Self-Learning and Synthetic Data Transfer for Efficient Back-Translation in Low Resource Neural Machine Translation.*
- Jan. 2017 **M.Sc. IT Security, Nottingham Trent University, Nottingham, UK.**
Dissertation title: *Web Application Security Analysis.*
- Apr. 2013 **B.Sc. Computer Science, Bayero University, Kano.**
Project title: *Plagiarism Checker, File Comparer and Project Library Software.*



Employment

- 2025, 2026 **Visiting Professor,**
African Institute for Mathematical Sciences (AIMS) – South Africa.
- 2023 – 2026 **Postdoctoral Fellow,**
Data Science for Social Impact Research Group, University of Pretoria – South Africa.
- 2023 **Lecturer I,**
Department of Computer Science, Ahmadu Bello University, Zaria, Kaduna–Nigeria.
- 2020 – 2023 **Lecturer II,**
Department of Computer Science, Ahmadu Bello University, Zaria, Kaduna–Nigeria.
- 2017 – 2020 **Assistant Lecturer,**
Department of Computer Science, Ahmadu Bello University, Zaria, Kaduna–Nigeria.



Teaching

- 2025 **African Institute for Mathematical Sciences, South Africa**
Natural Language Processing and LLMs (MSc)
- University of Pretoria**
COS760: Natural Language Processing (Honors)

Teaching (continued)

- 2023 – ...  **Arewa Data Science Academy**
Natural Language Processing
- 2017 – 2023  **Ahmadu Bello University, Zaria**
COSC211: Object-Oriented Programming I
COSC212: Object-Oriented Programming II
COSC406: Advanced Database Systems
COSC403: Software Engineering
COSC802: Software Engineering Methodologies (MSc)

Research Supervision

- 2024 – ...  **University of Pretoria.** Supervising 1 PhD, 6 Master's, and 4 Honours research students in African NLP, machine translation, sentiment and emotion analysis, and large language model evaluation.
- 2017 – 2023  **Ahmadu Bello University, Zaria.** Supervised 16 final-year BSc projects in Computer Science and Software Engineering.

Selected Publications

- 1** **Abdulmumin, I.**, Gwadabe, T., Muhammad, S. H., Adelani, D. I., Khalo, N., Ahmad, I. S., ... Marivate, V. (2026). Afriscience-mt: Towards decolonizing science in africa through text translation. arXiv: 2605.29741 [cs.CL]. Retrieved from  <https://arxiv.org/abs/2605.29741>
- 2** Muhammad, S. H., Ousidhoum, N., **Abdulmumin, I.**, Wahle, J. P., Ruas, T., Beloucif, M., ... Mohammad, S. M. (2025, July). BRIGHTER: BRIdging the gap in human-annotated textual emotion recognition datasets for 28 languages. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Proceedings of the 63rd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 8895–8916).  doi:10.18653/v1/2025.acl-long.436
- 3** **Abdulmumin, I.**, & Marivate, V. (2025). Zasca-sum: A dataset of the south africa supreme courts of appeal judgments and media summaries for legal documents summarization research. *Data in Brief*, 111567.  doi:<https://doi.org/10.1016/j.dib.2025.111567>
- 4** **Abdulmumin, I.**, Mkhwanazi, S., Mbooi, M., Muhammad, S. H., Ahmad, I. S., Putini, N., ... Marivate, V. (2024, November). Correcting FLORES evaluation dataset for four African languages. In B. Haddow, T. Kocmi, P. Koehn, & C. Monz (Eds.), *Proceedings of the ninth conference on machine translation* (pp. 570–578).  doi:10.18653/v1/2024.wmt-1.44
- 5** Muhammad, S., **Abdulmumin, I.**, Ayele, A., Ousidhoum, N., Adelani, D., Yimam, S., ... Arthur, S. (2023, December). AfriSenti: A Twitter sentiment analysis benchmark for African languages. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 13968–13981). Singapore: Association for Computational Linguistics. Retrieved from  <https://aclanthology.org/2023.emnlp-main.862>
- 6** **Abdulmumin, I.**, Beukman, M., Alabi, J., Emezue, C. C., Chimoto, E., Adewumi, T., ... Gwadabe, T. (2022, December). Separating grains from the chaff: Using data filtering to improve multilingual translation for low-resourced african languages. In *Proceedings of the seventh conference on machine translation* (pp. 1001–1014). Abu Dhabi: Association for Computational Linguistics.

- 7 **Abdulmumin, I.**, Dash, S. R., Dawud, M. A., Parida, S., Muhammad, S., Ahmad, I. S., ... Bello, B. S. (2022, June). Hausa visual genome: A dataset for multi-modal english to hausa machine translation. In *Proceedings of the language resources and evaluation conference* (pp. 6471–6479). Marseille, France: European Language Resources Association. Retrieved from <https://aclanthology.org/2022.lrec-1.694>
- 8 **Abdulmumin, I.**, Galadanci, B. S., & Aliyu, G. (2021). Tag-less back-translation. *Machine Translation*, 35(4), 519–549. [doi:10.1007/s10590-021-09284-y](https://doi.org/10.1007/s10590-021-09284-y)

Selected from 50+ peer-reviewed publications. Full list: [ORCID](#) • [Google Scholar](#).

Skills

Core	Consultation: Provide expert advice in African Natural Language Processing (e.g., Machine Translation, Automatic Speech Recognition (ASR) and Text-to-Speech (TTS), Sentiment, Emotion, Hate, and Offensive Language Analysis). Academic research: Conduct interdisciplinary research and publish peer-reviewed papers. Teaching, supervision, and training: Deliver lectures, workshops, mentoring sessions, and research supervision to undergraduate and postgraduate students.
Languages	Hausa: Native. English: Excellent reading, writing, and speaking proficiency. Nigerian Pidgin: Proficient in reading, writing, and speaking.
Coding	Python, Java, php, C++, sql, \LaTeX .
Databases	Mysql.

References

Available on Request