

Idris Abdulmumin

✉ abumafirim@gmail.com, iabdulmumin@abu.edu.ng
☎ +27 (0) 67-191-5773 📞 +234 (0) 806-2955-509
🌐 <https://abu.edu.ng/staff/abdulmumin-idris>
🌐 <https://www.linkedin.com/in/abumafirim/>
🌐 <https://github.com/abumafirim>
🌐 <https://orcid.org/0000-0002-3795-8381>



Employment History

- 2017 – ... 📌 **Lecturer,**
Department of Computer Science, Ahmadu Bello University, Zaria, Kaduna–Nigeria.
- 2014 – 2015 📌 **Teacher,**
Intercontinental Schools, Kano–Nigeria.
- 2013 – 2014 📌 **Teacher,**
Al-Ikhlās Private School, Jos–Nigeria.

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.*
- 2008 📌 **WASSCE, College of Arabic and Islamic Studies, Jos Plateau State.**
- 2002 📌 **Primary School Leaving Certificate, St. Anne's Private School Jos Plateau State.**

Research Publications

Journal Articles

- 1 **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
- 2 **Abdulmumin, I., Galadanci, B. S., Isah, A., Kakudi, H. A., & Sinan, I. I.** (2021). A hybrid approach for improved low resource neural machine translation using monolingual data. *Engineering Letters*, 29(4), 1478–1493. Retrieved from 🌐 http://www.engineeringletters.com/issues%5C_v29/issue%5C_4/EL%5C_29%5C_4%5C_21.pdf

Books and Chapters

- 1 **Abdulmumin, I., Galadanci, B. S., Ahmad, I. S., & Abdullahi, R. I.** (2021). Data selection as an alternative to quality estimation in self-learning for low resource neural machine translation. In O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, ... C. M. Torre (Eds.), *Computational science and its applications – iccsa 2021* (pp. 311–326). 🌐 doi:10.1007/978-3-030-87013-3_24
- 2 **Abdulmumin, I., Galadanci, B. S., & Isa, A.** (2021). Enhanced back-translation for low resource neural machine translation using self-training. In S. Misra & B. Muhammad-Bello (Eds.), *Information and communication technology and applications* (pp. 355–371). 🌐 doi:10.1007/978-3-030-69143-1_28

Conference Proceedings

- 1 Adelani, D. I., Masiak, M., Azime, I. A., Alabi, J., Tonja, A. L., Mwase, C., ... Stenetorp, P. (2023). Masakhanews: News topic classification for african languages. In *Proceedings of the 13th international joint conference on natural language processing and the 3rd conference of the asia-pacific chapter of the association for computational linguistics* (pp. 144–159). Nusa Dua, Bali: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2023.ijcnlp-long.10>
- 2 Aliyu, S. M., **Abdulmumin, I.**, Muhammad, S. H., Ahmad, I. S., Salahudeen, S. A., Yusuf, A., & Lawan, F. I. (2023). HausaNLP at SemEval-2023 task 10: Transfer learning, synthetic data and side-information for multi-level sexism classification. In *Proceedings of the the 17th international workshop on semantic evaluation (semeval-2023)* (pp. 1983–1987). Toronto, Canada: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2023.semeval-1.273>
- 3 Muhammad, S. H., **Abdulmumin, I.**, Yimam, S. M., Adelani, D. I., Ahmad, I. S., Ousidhoum, N., ... Ruder, S. (2023). SemEval-2023 task 12: Sentiment analysis for African languages (AfriSenti-SemEval). In *Proceedings of the the 17th international workshop on semantic evaluation (semeval-2023)* (pp. 2319–2337). Toronto, Canada: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2023.semeval-1.315>
- 4 Parida, S., **Abdulmumin, I.**, Muhammad, S. H., Bose, A., Kohli, G. S., Ahmad, I. S., ... Kakudi, H. (2023). HaVQA: A dataset for visual question answering and multimodal research in Hausa language. In *Findings of the association for computational linguistics: Acl 2023* (pp. 10162–10183). Toronto, Canada: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2023.findings-acl.646>
- 5 Adelani, D. I., Masiak, M., Azime, I. A., Alabi, J. O., Tonja, A. L., Mwase, C., ... Stenetorp, P. (2023). AfricaNEWS: News topic classification for african languages. In *4th Workshop on African Natural Language Processing*. Retrieved from <https://openreview.net/forum?id=sj45chH1F1>
- 6 Muhammad, S. H., Yimam, S. M., **Abdulmumin, I.**, Ahmad, I. S., Adelani, D. I., Ruder, S., ... Imam, A. A. (2023). Afrisenti-semeval: Sentiment analysis for low-resource african languages using twitter dataset. In *Task proposals for the 17th international workshop on semantic evaluation*. Retrieved from https://openreview.net/forum?id=EOUUC0wOq_r
- 7 **Abdulmumin, I.**, Beukman, M., Alabi, J., Emezue, C. C., Chimoto, E., Adewumi, T., ... Gwadabe, T. (2022). 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.
- 8 Adelani, D. I., Neubig, G., Ruder, S., Rijhwani, S., Beukman, M., Palen-Michel, C., ... Klakow, D. (2022). MasakhaNER 2.0: Africa-centric Transfer Learning for Named Entity Recognition. In *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 4488–4508). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2022.emnlp-main.298>
- 9 Aliyu, S., Wajiga, G., Murtala, M., Muhammad, S., **Abdulmumin, I.**, & Ahmad, I. (2022). Herdphobia: A dataset for hate speech detection against fulani herdsmen in nigeria. In *Proceedings of the sixth workshop on widening natural language processing*, Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. Retrieved from <https://arxiv.org/abs/2211.15262>
- 10 Muhammad, S., Brazdil, P., & **Abdulmumin, I.** (2022). Domain-specific lexicon-based sentiment analysis using contextual shifter patterns. In *Proceedings of the sixth workshop on widening natural language processing*, Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2022.winlp-1.0>
- 11 Adelani, D., Alabi, J., Fan, A., Kreutzer, J., Shen, X., Reid, M., ... Manthalu, S. (2022). A few thousand translations go a long way! leveraging pre-trained models for African news translation. In *Proceedings of the 2022 conference of the north american chapter of the association for computational linguistics: Human*

language technologies (pp. 3053–3070). Seattle, United States: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2022.naacl-main.223>

- 12 **Abdulmumin, I.**, Dash, S. R., Dawud, M. A., Parida, S., Muhammad, S., Ahmad, I. S., ... Bello, B. S. (2022). 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>
- 13 Muhammad, S. H., Adelani, D. I., Ruder, S., Ahmad, I. S., **Abdulmumin, I.**, Bello, B. S., ... Brazdil, P. (2022). Naijasenti: A nigerian twitter sentiment corpus for multilingual sentiment analysis. In *Proceedings of the language resources and evaluation conference* (pp. 590–602). Marseille, France: European Language Resources Association. Retrieved from <https://aclanthology.org/2022.lrec-1.63>
- 14 **Abdulmumin, I.**, Galadanci, B. S., Hassan Muhammad, S., & Aliyu, G. (2022). Quantity vs. quality of monolingual source data in automatic text translation: Can it be too little if it is too good? In *2022 ieee nigeria 4th international conference on disruptive technologies for sustainable development (nigercon)* (pp. 1–5). [doi:10.1109/NIGERCON54645.2022.9803137](https://doi.org/10.1109/NIGERCON54645.2022.9803137)
- 15 Emezue, C. C., Nigatu, H. H., Thinwa, C., Louis, L., **Abdulmumin, I.**, Oyerinde, S. G., ... Samuel, O. (2022). The african stopwords project: Curating stopwords for african languages. In *3rd workshop on african natural language processing*. Retrieved from <https://openreview.net/forum?id=B0GEqcGV8-5>
- 16 Ibrahim, R. A., & **Abdulmumin, I.** (2022). NECAT-CLWE: A simple but efficient parallel data generation approach for unsupervised and semi-supervised neural machine translation. In *3rd workshop on african natural language processing*. Retrieved from <https://openreview.net/forum?id=ruW4c9fNLZ5>
- 17 **Abdulmumin, I.**, & Galadanci, B. S. (2019). Hauwe: Hausa words embedding for natural language processing. In *2019 2nd international conference of the ieee nigeria computer chapter, nigeriacomputconf 2019*, Zaria, Kaduna-Nigeria. Retrieved from <http://ieeexplore.ieee.org/document/8949674/>

Preprints

- 1 Muhammad, S. H., **Abdulmumin, I.**, Ayele, A. A., Ousidhoum, N., Adelani, D. I., Yimam, S. M., ... Arthur, S. (2023). AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages. [doi:10.48550/ARXIV.2302.08956](https://doi.org/10.48550/ARXIV.2302.08956)
- 2 Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., ... Wolf, T. (2022). BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. [doi:10.48550/ARXIV.2211.05100](https://doi.org/10.48550/ARXIV.2211.05100)

Skills








Languages	📌 Hausa: Native. English: Strong reading, writing and speaking competency. Nigerian Pidgin: Reading, writing and speaking competency.
Coding	📌 Java, PHP, Python, C++, SQL, \LaTeX .
Databases	📌 MySQL.
Misc.	📌 Academic research, teaching, training, consultation, \LaTeX typesetting and publishing.

Miscellaneous Experience





Awards and Achievements

- 2022 📌 Co-Investigator: Artificial Intelligence for Education Innovation Research Network in Africa. (40,000 USD)
- 📌 Co-Investigator: Google exploreCSR Award. (18,000 USD)

Miscellaneous Experience (continued)

- 2021  Co-Investigator: Lacuna Fund 2021 Language Datasets Award. (147,000 USD)
-  Co-Investigator: Oracle OCI Award. (50,000 USD)
-  Beneficiary: Google Africa Travel Grants. (3,000 USD)
- 2018  Co-Investigator: Lacuna Fund 2020 Language Datasets Award. (60,000 USD)
- 2018  Scholar: National Information Technology Development Education Fund Ph.D. Scholarship Award.
- 2015  Scholar: National Information Technology Development Education Fund M.Sc. Scholarship Award.
- 2011  Beneficiary: Emerging Markets Telecommunications Services Limited (Etisalat) Indigene Award for Excellence.


Certification

- 2018  **Mathematics for Machine Learning: Multivariate Calculus.** Awarded by Coursera.
 <https://www.coursera.org/account/accomplishments/certificate/JBZ4RBRZ2VSC>
-  **Mathematics for Machine Learning: Linear Algebra.** Awarded by Coursera.
 <https://www.coursera.org/account/accomplishments/certificate/DWBUYZZ2C9QC>

Teamwork and Collaboration with Others

- 2023  **Co-organizer, Tutor & Mentor:** Google #exploreCSR–Arewa DataScience 2023 Fellowship.
-  **Co-organizer:** SemEval 2024 Task 1, title: “Semantic Textual Relatedness for African and Asian Low Resource Languages”.
-  **Co-organizer:** Large Scale Multilingual African Track at the WMT 2023.
-  **Experimental Lead:** Machine Translation, Lacuna Funded Decolonise Science Project.
- 2022  **Co-organizer:** "Bird of the Feather Workshop—Challenges for Hate Speech Detection in Low-resource African languages and Beyond at EMNLP 2022".
-  **Co-Lead:** Hausa Visual Question Answering Project (completed, submitted to [ACL 2023](#)).
-  **Co-organizer & Technical Lead:** SemEval 2023 Task 12, title: “AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages using Twitter Dataset”.
-  **Co-investigator & Project Manager:** Creation of language resource, title: “AfriHate: Hate and Offensive Speech Dataset for African Languages”. (Lacuna 2021 Grant)
-  **Collaborator & Technical Lead:** [Ndyuka Translate](#) - Automatic Translation System for Ndyuka, Pamaka & Aluku to/from English and French.
-  **Project Lead:** [MasaKhane submission to the WMT22](#) – Large-scale Multilingual African track (completed, [published](#)).
-  **Researcher & Collaborator:** [MasakhaNER 2.0](#): Curating named-entity recognition datasets for African languages.
-  **Researcher & Collaborator:** [LAFAND-MT](#): Curating machine translation datasets for African languages.
-  **Researcher & Collaborator:** [African stopwords](#) project.
- 2021  **Project Lead:** Hausa Visual Genome: A Dataset for Multi-Modal English to Hausa Machine Translation (completed, [published](#)).
-  **Project Manager:** Corpus, Sentiment, and Hate Speech Lexicon for Major Nigerian Languages, Lacuna Language 2020 Funded Project (completed, [published](#)).



Miscellaneous Experience (continued)

- 2019  **Project Lead:** HauWE: Hausa Words Embedding for Natural Language Processing (completed, [published](#)).






Conferences/Summits Attended

- 2023  **Author:** The 61st Annual Meeting of the Association for Computational Linguistics, Toronto, 9-14 July 2023
-  **Author:** The 17th International Workshop on Semantic Evaluation (SemEval-2023), Toronto, 9-14 July 2023
- 2022  **Speaker:** Deep Learning IndabaX Uganda 2022, 28 October 2022
-  **Panelist:** [Meetup](#) on Low Resource Machine Translation, 21 October 2022
-  **Participant, Speaker & Workshop Presenter:** Deep Learning Indaba 2022, Tunis, 21-26 August 2022
-  **Author:** 13th Conference on Language Resources and Evaluation (LREC 2022), Marseille, 20-25 June 2022
-  **Author:** 4th International Conference on Disruptive Technologies for Sustainable Development, Abuja, 17-19 May 2022
-  **Author & Program Committee Member:** 2022 AficaNLP Workshop, Virtual, 25 April 2022
- 2021  **Author:** International Conference on Computing and Advances in Information Technology (ICCAIT 2021), Zaria, Nigeria, November 15 – 17.
-  **Participant:** ODI Summit 2021 – Data <> People, hosted by The Open Data Institute, Virtual, November 2.
-  **Author:** International Conference on Computational Science and its Applications (ICCSA 2021), Cagliari, Italy, September 13 – 16.
- 2020  **Author:** International Conference on Information and Communication Technology and its Applications (ICTA 2020), Minna, Nigeria, November 24 – 26.
- 2019  **Author:** 2nd International Conference of the IEEE Nigeria Computer Chapter (IEEEEnig-ComputConf'19), Zaria, Nigeria, October 14 – 17.

Professional Affiliation

- 2020 – . . .  **Member**, International Association of Engineers
- 2022 – . . .  **Member**, [European COST Action LITHME \(Language in the Human Machine Era, WGs 1 & 6\)](#)

Peer-Reviewed For

-  SemEval 2024 Task Proposals
-  International Conference on Computing and Advances in Information Technology (ICCAIT)
-  IEEE/ACM Transactions on Audio, Speech, and Language Processing
-  SN Computer Science
-  PLOS ONE

References

Professor Bashir Shehu Galadanci

Professor of Software Engineering
Bayero University, Kano
Kano-Nigeria.

✉ bsgaladanci.se@buk.edu.ng

Dr Aminu Bello Usman

Associate Head for Computer and Data Science
York St John University
Yorkshire-United Kingdom.

✉ a.usman@yorks.j.ac.uk

Professor Bettina Migge

Professor of Linguistics
School of Languages, Cultures and Linguistics
University College Dublin, Ireland.

✉ bettinamigge@ucd.ie