Shamsuddeen Hassan Muhammad

✉ shamsuddeen2004@gmail.com   🐦 @shmuhammadd
🌐 https://scholar.google.com/citations?user=Ne1nt4gAAAAJ&hl=en
🌐 https://shmuhammadd.github.io
in https://www.linkedin.com/in/shmuhammad/
○ https://github.com/shmuhammadd

Academic and Research Experience

2023 – · · ·   **Google DeepMind Academic Fellow**
                 Imperial College London, UK

2020 – 2023   **NLP Researcher, INESCTEC Porto, Portugal**
                 INESCTEC Porto, Portugal

2012 – · · ·   **Lecturer,**
                 Department of Computer Science, Bayero University, Kano–Nigeria.

2010 – 2012   **Lecturer,**
                 Department of Computer Science, Saadatu Rimi College of Education, Kano State-Nigeria

Education

2023   **Ph.D. in Computer Science, University of Porto, Portugal**.
         Thesis Title: *Domain-Specific and Context-Aware Approaches to Sentiment Analysis*.
         Supervisor: *Pavel Brazdil*

2013   **M.Sc. in Computer Science, University of Manchester, UK**.
         Dissertation Title: *Error-Correcting Codes in Wireless Communication*.
         Supervisor: *Barry Cheetham*

2009   **B.Sc. in Computer Science, Bayero University, Kano**.

Teaching Certificates

2019   **Teaching in Higher Education, University of Porto, Portugal**.
         Certificate for teaching in higher education.

2020   **Certified Instructor, The Carpentries**.
         Certified instructor in teaching foundational computational and data skills to researchers using R and Python programming languages.

Grants and Awards

Grants

2022   Principal-Investigator: Artificial Intelligence for Education Innovation Research Network in Africa. (40,000 USD)

         Principal-Investigator: Lacuna Fund 2021 Language Datasets Award: Development of Hate Speech Dataset for African Languages (147,000 USD)

         Co-Investigator: Lacuna Fund Grant for Developing Nigerian Speech Corpus (120,000 USD)

Grants and Awards (continued)

2021   ▮ Principal-Investigator: Lacuna Fund 2020 Language Datasets Award. Development of Sentiment Analysis Dataset for African Languages (60,000 USD)

▮ Principal-Investigator: Lacuna Fund Grant for SemEval Shared Task. (8,500 USD)

Awards

2023   ▮ Good DeepMind Academic Fellow at Imperial College London, UK

▮ Best Paper Award, AfricaNLP workshop 2023, Kigali Rwanda.

▮ IRCAI, Global Top 100 list 2022, international list of AI solutions for reaching the 17 SGDs.

2022   ▮ Oracle OCI Award. (50,000 USD)

▮ Google exploreCSR Award. (18,000 USD)

▮ Google's TensorFlow Award 2022 (7,000 USD)

▮ Google Africa Travel Award. (3,000 USD)

▮ RStudio Conference Travel Award. (1,500 USD)

▮ EMNLP Conference Travel Award. (1,500 USD)

▮ Faculty of Science, UPorto, PhD Scholarship Award. (52,000 EURO)

▮ Jiangsu University(JSU), China, PhD scholarship, Award (Declined).

2018   ▮ Nigerian TetFund Ph.D. Scholarship Award.

▮ NITDA MSc Scholarship Award (25,000 USD).

▮ Black in AI Travel Award to NeurIPS 2022, Value of Award (1,100 USD)

▮ Kano State Govt-Nigeria , Postgraduate Scholarship to Study MSc in Computer Science.

▮ Second Best Computer Science Graduating Student Award, Bayero University, Kano-Nigeria

Research Publications

Journal Articles

1. Brazdil, P., **Muhammad, Shamsuddeen H.**, Oliveira, F., Cordeiro, J., Silva, F., Silvano, P., & Leal, A. (2022). Semi-automatic approaches for exploiting shifter patterns in domain-specific sentiment analysis. *Mathematics*, *10*(18). 🔗 doi:10.3390/math10183232

2. Kreutzer, J., Caswell, I., Wang, L., Wahab, A., van Esch, D., Ulzii-Orshikh, N., … Sikasote, C. et al. (2022). Quality at a glance: An audit of web-crawled multilingual datasets. *Transactions of the Association for Computational Linguistics*, *10*, 50–72. Retrieved from 🔗 https://doi.org/10.1162/tacl_a_00447

3. Adelani, D. I., Abbott, J., Neubig, G., D'souza, D., Kreutzer, J., Lignos, C., … Osei, S. (2021). MasakhaNER: Named entity recognition for African languages. *Transactions of the Association for Computational Linguistics*, *9*, 1116–1131. 🔗 doi:10.1162/tacl_a_00416

4. Ahmad, I. S., Bakar, A. A., Yaakub, M. R., & **Muhammad, Shamsuddeen Hassan**. (2020). A survey on machine learning techniques in movie revenue prediction. *SN Computer Science*, *1*(4), 1–14.

5. **Muhammad, SH**, Mustapha, A., & Haruna, K. (2016). Massive open online courses: Awareness, adoption, benefits and challenges in sub-saharan africa. *International Journal of ICT and Managemant*, *4*(2), 60–68.

Proceedings

1. **Muhammad, Shamsuddeen**, 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 17th international workshop on semantic evaluation (semeval-2023)*, Toronto, Canada. Retrieved from 🔗 https://arxiv.org/abs/2304.06845

2. 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

3. **Shamsuddeen Hassan Muhammad**, 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=EOUUCOwOq_r

4. 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. Retrieved from 🔗 https://arxiv.org/abs/2210.10692

5. Aliyu, S., Wajiga, G., Murtala, M., **Shamsuddeen Muhammad**, 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://aclanthology.org/2022.winlp-1.0

6. **Muhammad, Shamsuddeen**, 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

7. 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

8. Abdulmumin, I., Dash, S. R., Dawud, M. A., Parida, S., **Muhammad, Shamsuddeen**, 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

9. **Shamsuddeen Hassan Muhammad**, 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

10. Abdulmumin, I., Galadanci, B. S., **Hassan Muhammad, Shamsuddeen**, & 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

11. 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

12 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

13 Meyer, J., Adelani, D., Casanova, E., Öktem, A., Whitenack, D., Weber, J., … **Shamsuddeen Hassan Muhammad**. (2022). Bibletts: A large, high-fidelity, multilingual, and uniquely african speech corpus. In H. Ko & J. H. L. Hansen (Eds.), *Interspeech 2022, 23rd annual conference of the international speech communication association, incheon, korea, 18-22 september 2022* (pp. 2383–2387). 🔗 doi:10.21437/Interspeech.2022-10850

14 Brazdil, P., Silvano, P., Silva, F., **Muhammad, Shamsuddeen**, Oliveira, F., Cordeiro, J., & Leal, A. (2021). Extending general sentiment lexicon to specific domains in (semi-)automatic manner. In S. Carvalho & R. R. Souza (Eds.), *Salld-1: Proceedings of the workshop on sentiment analysis & linguistic linked data*, CEUR, Vol. 3064.

15 **Shamsuddeen Hassan Muhammad**, Brazdil, P., & Jorge, A. (2020). Incremental approach for automatic generation of domain-specific sentiment lexicon. In *European conference on information retrieval* (pp. 619–623). Springer.

Preprints

1 **Muhammad, Shamsuddeen Hassan**, 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

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

Under Review

1 Aliyu, S., Abdullahi, S., **Shamsuddeen Hassan Muhammad**, Ahmad, I. S., Abdulmumin, I., Yusuf, A., & Lawan, F. (n.d.). *HausaNLP at SemEval-2023 Task 10: Transfer Learning, Synthetic Data and Side-information for Multi-level Sexism Classification.*

2 **Muhammad, Shamsudeen Hassan**, Brazdil, P., & Jorge, A. (n.d.). *Combining symbolic and deep learning approaches for sentiment analysis.*

3 Parida, S., Abdulmumin, I., **Shamsuddeen Hassan Muhammad**, Bose, A., Kohli, G. S., Ahmad, I. S., … Kakudi, H. (n.d.). *HaVQA: A dataset for visual question answering and multimodal research in hausa language.*

4 **Muhammad, Shamsuddeen Hassan**, Brazdil, P., & Jorge, A. M. (2023). *Leveraging symbolic and deep learning techniques for explainable sentiment analysis.*

Academic Service

Organizing Conference

2022	🔖 Co-Organizer for the, AfricanNLP2023.
	🔖 Co-organizer for the "Bird of the Feather: Challenges for Hate Speech Detection in Low-resource African languages and Beyond at EMNLP 2022".
	🔖 Co-organizer, AfriSenti-SemEval Shared Task 2023 The First SemEval Shared Task for African Languages.
2021	🔖 Co-Organizer for the, AfricanNLP2021.
2020	🔖 Co-Organizer, SautiYetu AfricanNLP Unconference October 2020.

Service to the University

- ◼ Senate Member, Bayero University, Kano
- ◼ University Examination Committe Member, Bayero University, Kano
- ◼ BSc Computer Science with Economics course Coordinator Bayero University, Kano

Courses Taught

2021 –··· ◼ **Lead TA: Deep Learning at Neuromatch Academy**.
As a Lead Teaching Assistant for a group of approximately 100 students in a Deep Learning course.

2012 – 2017 ◼ **Undergraduate courses**.
Linear Algebra,
Python Programming,
Data Mining,
Data Structure and Algorithm,
Discrete Mathematics.

2012 – 2017 ◼ **Postgraduate courses**.
Research Methods,
Cloud Computing.

Supervision

2012 – 2018··· ◼ **Undergraduate Supervision**.
Abdulrasheed Mustapha (Software Engineer at Intuit Inc)
Saheed A. Salahudeen (Lecturer Kaduna State University)

Community Engagement

Organizing Conference

2022 ◼ Founder, Arewa Data Science: https://arewadatascience.github.io
A community for empowering marginalized students to reach their full potential in data science.

◼ Founder, HausaNLP: https://hausanlp.github.io
HausaNLP is an open-source community dedicated to advancing natural language processing in the Hausa language, which is the second most widely spoken language in Africa.

◼ Active Member of MasakhaneNLP and Black in AI Community.