





Jesujoba O. Alabi

✉ Spoken Language Systems Group, Saarland University, 66123 Saarbrücken, Germany

✉ jalabi@lsv.uni-saarland.de

🌐 <https://ajesujoba.github.io/>

Education

- 2022 – 2026  **Ph.D., Computer Science**, Saarland University, Saarbrücken, Germany.
Advisor: *Prof. Dr. Dietrich Klakow*.
Thesis title:
- 2018 – 2021  **M.Sc. Computer Science**, Saarland University, Saarbrücken, Germany.
Thesis title: *Phrase-Enriched Self-Supervised Neural Machine Translation*.
Overall CGPA: 1.7 (Best 1.0, Worst 5.0)
- 2017 – 2021  **M.Sc. Computer Science**, University of Ibadan, Ibadan, Nigeria.
Thesis title: *Transfer Learning for Malaria Detection from Thin-Blood Smear Images*
Overall CGPA: 6.7/7
- 2011 – 2015  **B.Sc. Computer Science**, Bowen University, Iwo, Nigeria.
Overall CGPA: 4.84/5

Employment History

- 01/25 – 02/25  **Visiting PHD Student**, National Institute of Informatics, Tokyo, Japan.
Host: *Prof. Junichi Yamagishi*
Topic: *Multilingual text-to-speech*
- 05/24 – 08/24  **Research Intern**, National Institute of Informatics, Tokyo, Japan.
Advisor: *Prof. Junichi Yamagishi*
Topic: *Self-supervised learning for under-resourced languages*
- 02/22 – 06/22  **Research Engineer**, ALMAnaCH project-team, INRIA, Paris, France.
Advisor: *Dr. Rachel Bawden*
Topic: *Domain adaptation in neural machine translation*
- 08/21 – 01/22  **Research Intern**, Max Planck Institute for Informatics, Saarbrücken, Germany.
Advisor: *Dr. Rishiraj Saha Roy*
Topic: *Unified Question Answering over heterogeneous sources*
- 09/20 – 07/21  **Research Assistant**, Max Planck Institute for Informatics, Saarbrücken, Germany.
- 11/19 – 07/20  **Research Assistant**, German Research Center for Artificial Intelligence, Saarbrücken, Germany.
Advisor: *Dr. Cristina España-Bonet*
Topic: *Self-supervised neural machine translation*
- 01/16 – 10/16  **Graduate Assistant**, Afe Babalola University, Ado-Ekiti, Nigeria.
Advisor: *Dr. Seyi Adeyemo*

Research Publications

Journal Articles

- 1 S. Pramanik, **J. Alabi**, R. S. Roy, and G. Weikum, “Uniqorn: Unified question answering over rdf knowledge graphs and natural language text”, *Journal of Web Semantics*, vol. 83, 2024.
- 2 D. I. Adelani, J. Abbott, G. Neubig, D. D’souza, J. Kreutzer, *et al.*, “MasakhaNER: Named entity recognition for African languages”, *Transactions of the Association for Computational Linguistics*, vol. 9, 2021.

Conference Publications

- 1 D. I. Adelani, J. Ojo, I. A. Azime, J. Y. Zhuang, **J. O. Alabi**, *et al.*, “IrokoBench: A new benchmark for African languages in the age of large language models”, in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*, Apr. 2025.
- 2 **J. O. Alabi**, X. Liu, D. Klakow, and J. Yamagishi, “Afrihubert: A self-supervised speech representation model for african languages”, in *Interspeech*, 2025.
- 3 L. Kiefer, **J. Alabi**, T. Vakili, H. Dalianis, and D. Klakow, “Instruction-tuning llama for synthetic medical note generation in swedish and english”, in *Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing*, Sep. 2025.
- 4 H. Yu, **J. O. Alabi**, A. Bukula, J. Y. Zhuang, E.-S. A. Lee, *et al.*, “INJONGO: A multicultural intent detection and slot-filling dataset for 16 African languages”, in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, Jul. 2025.
- 5 D. I. Adelani, H. Liu, X. Shen, N. Vassilyev, **J. O. Alabi**, *et al.*, “SIB-200: A simple, inclusive, and big evaluation dataset for topic classification in 200+ languages and dialects”, in *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics*, Mar. 2024.
- 6 **J. Alabi**, M. Mosbach, M. Eyal, D. Klakow, and M. Geva, “The hidden space of transformer language adapters”, in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Aug. 2024.
- 7 M. Zhang, V. Gautam, M. Wang, **J. Alabi**, X. Shen, *et al.*, “The impact of demonstrations on multilingual in-context learning: A multidimensional analysis”, in *Findings of the Association for Computational Linguistics: ACL 2024*, Bangkok, Thailand, Aug. 2024.
- 8 D. I. Adelani, M. Masiak, I. A. Azime, **J. Alabi**, A. L. Tonja, *et al.*, “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*, Nov. 2023.
- 9 C. M. B. Dione, D. I. Adelani, P. Nabende, **J. Alabi**, T. Sindane, *et al.*, “MasakhaPOS: Part-of-speech tagging for typologically diverse African languages”, in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, Jul. 2023.
- 10 A. Krishnan, **J. O. Alabi**, and D. Klakow, “On the n-gram approximation of pre-trained language models”, in *Interspeech*, 2023.
- 11 D. I. Adelani, **J. O. Alabi**, A. Fan, J. Kreutzer, X. Shen, *et al.*, “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*, Jul. 2022.
- 12 **J. O. Alabi**^{*}, D. I. Adelani^{*}, M. Mosbach, and D. Klakow, “Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning”, in *Proceedings of the 29th International Conference on Computational Linguistics*, Oct. 2022.
- 13 E. Chang, **J. O. Alabi**, D. I. Adelani, and V. Demberg, “Few-shot pidgin text adaptation via contrastive fine-tuning”, in *Proceedings of the 29th International Conference on Computational Linguistics*, Oct. 2022.
- 14 D. I. Adelani^{*}, D. Ruiter^{*}, **J. O. Alabi**^{*}, D. Adebajo, A. Ayeni, *et al.*, “The effect of domain and diacritics in Yoruba–English neural machine translation”, in *Proceedings of Machine Translation Summit XVIII: Research Track*, Aug. 2021.

- 15 **J. O. Alabi***, K. Amponsah-Kaakyire*, D. I. Adelani, and C. España-Bonet, “Massive vs. curated embeddings for low-resourced languages: The case of Yorùbá and Twi”, Proceedings of the Twelfth Language Resources and Evaluation Conference, May 2020.
- 16 M. A. Hedderich, D. I. Adelani, D. Zhu, **J. Alabi**, U. Markus, *et al.*, “Transfer learning and distant supervision for multilingual transformer models: A study on African languages”, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Nov. 2020.

Workshop Publications

- 1 **J. O. Alabi** and R. Bawden, *Exploring inline lexicon injection for cross-domain transfer in neural machine translation*, Proceedings of the First International Workshop on Knowledge-Enhanced Machine Translation, Jun. 2024.
- 2 C. Ilevbare*, **J. Alabi***, D. I. Adelani, F. Bakare, O. Abiola, *et al.*, *EkoHate: Abusive language and hate speech detection for code-switched political discussions on Nigerian Twitter*, Proceedings of the 8th Workshop on Online Abuse and Harms (WOAH 2024), Jun. 2024.
- 3 A. M. Olawole, **J. O. Alabi***, A. B. Sakpere, and D. I. Adelani, *YAD: Leveraging t5 for improved automatic diacritization of yorùbá text*, 5th Workshop on African Natural Language Processing, 2024.
- 4 M. Zhang, M. Wang, **J. Alabi**, and D. Klakow, *AAdAM at SemEval-2024 task 1: Augmentation and adaptation for multilingual semantic textual relatedness*, Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), Jun. 2024.
- 5 I. Akinade*, **J. O. Alabi***, D. I. Adelani, C. Odoje, and D. Klakow, *Varepsilon kú mask: Integrating Yorùbá cultural greetings into machine translation*, Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP), May 2023.
- 6 I. Abdulmumin*, M. Beukman*, **J. Alabi***, C. C. Emezue, E. Chimoto, *et al.*, *Separating grains from the chaff: Using data filtering to improve multilingual translation for low-resourced African languages*, Proceedings of the Seventh Conference on Machine Translation (WMT), Dec. 2022.
- 7 **J. Alabi**, L. Nishimwe, B. Muller, C. Rey, B. Sagot, *et al.*, *Inria-ALMANaCH at WMT 2022: Does transcription help cross-script machine translation?*, Proceedings of the Seventh Conference on Machine Translation (WMT), Dec. 2022.
- 8 S. Tchistiakova, **J. Alabi**, K. Dutta Chowdhury, S. Dutta, and D. Ruiter, *EdinSaar@WMT21: North-Germanic low-resource multilingual NMT*, Proceedings of the Sixth Conference on Machine Translation, Nov. 2021.
- 9 S. Dutta*, **J. Alabi***, S. Bandyopadhyay, D. Ruiter, and J. van Genabith, *UdS-DFKI@WMT20: Unsupervised MT and very low resource supervised MT for German-Upper Sorbian*, Proceedings of the Fifth Conference on Machine Translation, Online, Nov. 2020.  URL: <https://aclanthology.org/2020.wmt-1.129/>.




Preprints

- 1 **J. O. Alabi**, I. A. Azime, M. Zhang, C. España-Bonet, R. Bawden, *et al.*, “Afridoc-mt: Document-level mt corpus for african languages”, EMNLP (to appear), 2025.
- 2 **J. O. Alabi**, M. A. Hedderich, D. I. Adelani, and D. Klakow, “Charting the landscape of african nlp: Mapping progress and shaping the road ahead”, EMNLP (to appear), 2025.
- 3 L. H. Beyene, V. Verma, M. Ma, **J. O. Alabi**, F. D. Schmidt, *et al.*, “Msteb: Massively multilingual evaluation of llms on speech and text tasks”, ASRU (to appear), 2025.
- 4 A. Aremu, **J. O. Alabi**, D. Abolade, N. F. Aguobi, S. H. Muhammad, *et al.*, “Naijarc: A multi-choice reading comprehension dataset for nigerian languages”, 2024.








- 5 E. Bayes*, I. A. Azime*, **J. O. Alabi***, J. Kgomo, T. Eloundou, *et al.*, “Uhura: A benchmark for evaluating scientific question answering and truthfulness in low-resource african languages”, 2024.

Miscellaneous Experience

Research Grants


- 03/25 – date  AfriFact & AfriGuard: Dataset for Fact-Checking and Robust Refusal (Co-PI with Israel Abebe Azime, **\$113,332.50** funded by Lacuna Fund)
- 05/23 – 08/24  NiLE: Using language technologies to spread educational content in Nigerian languages (Co-PI with Kola Tubosun, **CAD \$45, 000** funded by Africa AI4D-EduAI Hub)
- 10/23 – 08/24  MASAKHANE DAMT: Domain adaptation datasets for African languages (Co-PI with Clement Odoje, **\$66, 875** Masakhane project funded by Lacuna Fund 2022)

Awards and Scholarships





- 2025  **Outstanding Paper Award**, NAACL 2025, New Mexico
- 2024  **Best System Description Paper Award**, SemEval 2024, NAACL 2024, Mexico
- 2023  **Area Chair Award(Resources and Evaluation)**, AACL 2023, Indonesia
- 2022  **Best Paper Award (Grand Challenge KISTI category)**, COLING 2022, South Korea
- 2021  **1st prize**, AI4D Takwimu Lab - Machine Translation Challenge
- 2020  **STIBET III DAAD Graduation Grant**, Saarland Scholarship for International Students
- 2019  **Black in AI travel grant**, 2019 Neurips at Vancouver

Teaching Experience and Academic Service



Teaching Experience

- 2023  Co-teaching, Seminar on Machine Learning for Natural Language Processing and beyond (with Prof. Dr. Dietrich Klakow)


Guest and Keynote Lectures

- 13/09/2025  **Keynote Talk**, Workshop on Advancing NLP for Low-Resource Languages at RANLP (Bulgaria)
Title: TBD
- 26/08/2025  **Keynote Talk**, Erasmus LCT Summer School (Germany)
Title: Adaptation at Scale: Making Multilingual Models Work for Underrepresented Languages
- 31/07/2025  **Keynote Talk**, AfricaNLP Workshop at ACL (Austria)
Title: Mapping Progress in African NLP
- 17/10/2024  **Guest Lecture**, Collaborative Research, Houghton University, USA (Online)
Title: NLP for African Languages

Thesis Supervision

- 2024  **Lotta Elisa Kiefer**, Instruction-Tuning LLaMA for Synthetic Medical Note Generation: Bridging Data Privacy and Utility in Downstream Tasks. (Masters)
-  **Sambit Bhaumik**, Leveraging Multilingual FAQ Resources for Enhanced Cross-Lingual Information Retrieval. (Masters)

Teaching Experience and Academic Service (continued)

2023  **Jeremy Mika Büdinger**, Investigating Adapter-based Fine-Tuning for Luxembourgish.
(Bachelors)

References

Available on Request