

Jesujoba O. Alabi

✉ Spoken Language Systems Group, Saarland University, 66123 Saarbrücken, Germany
✉ jalabi@lsv.uni-saarland.de
🌐 <https://ajesujoba.github.io/>

Education and Degrees

- 2022 – Present █ **Ph.D. in Computer Science**, Saarland University, Saarbrücken, Germany.
Advisor: *Prof. Dr. Dietrich Klakow*
- 2018 – 2021 █ **M.Sc. in 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. in 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. in Computer Science**, Bowen University, Iwo, Nigeria.
Overall CGPA: 4.84/5

Professional Experience

- 01/25 – 02/25 █ **Visiting Ph.D. 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. Oluwaseyi A. Adeyemo*

Research Publications

* indicates joint first authorship

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 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”, in *IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2025*, 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”, in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 2025.
- 3 **J. O. Alabi**, I. A. Azime, M. Zhang, C. España-Bonet, R. Bawden, *et al.*, “AFRIDOC-MT: Document-level MT corpus for African languages”, in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 2025.
- 4 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*, 2025.
- 5 **J. O. Alabi**, X. Liu, D. Klakow, and J. Yamagishi, “Afrihubert: A self-supervised speech representation model for african languages”, in *Interspeech*, 2025.
- 6 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*, 2025.
- 7 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*, 2025.
- 8 **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*, 2024.
- 9 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*, 2024.
- 10 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*, 2024.
- 11 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*, 2023.
- 12 A. Krishnan, **J. O. Alabi**, and D. Klakow, “On the n-gram approximation of pre-trained language models”, in *Interspeech*, 2023.
- 13 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*, 2023.
- 14 **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*, 2022.
- 15 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*, 2022.

- 16 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*, 2022.
- 17 D. I. Adelani*, D. Ruiter*, **J. O. Alabi***, D. Adebonojo, 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*, 2021.
- 18 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)*, 2020.
- 19 **J. O. Alabi***, K. Ampsonah-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, 2020.

Workshop Publications

- 1 T. Nakai, R. chikkala, L. Oberkircher, N. Jennings, N. Skachkova, *et al.*, *TReplina: Layer-wise CKA+REPINA alignment improves low-resource machine translation in aya-23 8b*, 1st Workshop on MMLoSo: Multimodal Models for Low-Resource Contexts and Social Impact @ ACL, 2025.
- 2 **J. O. Alabi**, C. Gong, E. Cooper, Y. Jiang, D. Klakow, *et al.*, *Submission from ZMM-TTS for Blizzard Challenge 2025*, The Blizzard Challenge 2025, 2025.
- 3 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), 2024.
- 4 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), 2024.
- 5 **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, 2024.
- 6 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.
- 7 I. Akinade*, **J. O. Alabi***, D. I. Adelani, C. Odoje, and D. Klakow, *Varepsilonilon kú mask: Integrating Yorùbá cultural greetings into machine translation*, Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP), 2023.
- 8 I. A. Azime, S. Al-azzawi, A. L. Tonja, I. Shode, **J. Alabi**, *et al.*, *Masakhane-afrisenti at SemEval-2023 task 12: Sentiment analysis using Afro-centric language models and adapters for low-resource African languages*, Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), 2023.
- 9 **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), 2022.
- 10 I. Abdulkumain*, 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), 2022.
- 11 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, 2021.

- 12 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, 2020.

Preprints

- 1 K. Uemura, D. Guzmán, Q. P. Nguyen, **J. O. Alabi**, E.-s. A. Lee, *et al.*, “Merlin: Multi-stage curriculum alignment for multilingual encoder-llm integration in cross-lingual reasoning”, to appear EACL 2026.
- 2 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.
- 3 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 – Present | ■ AfriFact & AfriGuard: Dataset for Fact-Checking and Robust Refusal.
(Co-PI with Israel Abebe Azime, USD \$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.
(PI, USD \$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 |

Summer School and Workshop

- | | |
|-------|---|
| 03/25 | ■ HumanCLAIM Workshop , University of Göttingen, Germany |
| 10/24 | ■ GCND doctoral school/hackathon , Ghent University, Belgium |
| 09/24 | ■ Athens NLP Summer School , NCSR Demokritos Campus, Greece |
| 07/20 | ■ Lisbon Machine Learning School , Online |

Peer Review

- | | |
|------|---|
| 2026 | ■ Area chair - LREC 2026 |
| 2025 | ■ Emergency area chair - ACL ARR (May, October) |
| | ■ Reviewer - COLING, HRI '25, ACL ARR (May, July), AfricaNLP Workshop |
| 2024 | ■ Reviewer - LREC-COLING, AfricaNLP Workshop |

Miscellaneous Experience (continued)

- 2023 ┣ Reviewer - AfriCHI, AfricaNLP Workshop, EMNLP
- 2022 ┣ Reviewer - WMT

Teaching Experience and Academic Service

Teaching Experience

- 2023 ┣ Co-taught 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 (Varna, Bulgaria)
Title: Bridging the Gap: Low-Resource NLP for African Languages
- 03/09/2025 ┣ **Guest Lecture**, Machine Translation, University of Ibadan, Nigeria (Online)
Title: Neural Machine Translation
- 26/08/2025 ┣ **Invited Talk**, Erasmus LCT Summer School, DFKI (Saarbrücken, Germany)
Title: Adaptation at Scale: Making Multilingual Models Work for Underrepresented Languages
- 31/07/2025 ┣ **Keynote Talk**, AfricaNLP Workshop at ACL (Vienna, Austria)
Title: Mapping Progress in African NLP
- 17/10/2024 ┣ **Guest Lecture**, Collaborative Research, Houghton University, USA (Online)
Title: NLP for African Languages
- 04/09/2020 ┣ **Invited Talk**, Data Science for Social Impact Research Group, University of Pretoria, South Africa (Online)
Title: Massive vs. Curated Embeddings for Low-Resourced Languages: the Case of Yorùbá and Twi
- 03/07/2020 ┣ **Invited Talk**, AI4D Africa Webinar Series (Online)
Title: Yoruba and beyond: NLP from an African perspective

Thesis Supervision

- 2025 ┣ **Katharina Trinley**, Bias in LLMs. (Masters, Ongoing)
- ┣ **Lena Sophie Oberkircher**, Text-to-speech for Saarländisch. (Masters, Ongoing)
- 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)
- 2023 ┣ **Jeremy Mika Büdinger**, Investigating Adapter-based Fine-Tuning for Luxembourgish. (Bachelors)

References

Available on Request