David Ifeoluwa Adelani

Personal Data

Place of Birth:	Abeokuta, Nigeria
Address:	Montréal, Canada
email:	david.adelani@mcgill.ca or david.adelani@mila.quebec
website:	dadelani.github.io or On Google Scholar

Education

JAN 2015 - JUN 2023	Preparatory Phase (2015-2018) and Dissertation Phase of Doctoral Studies **Spoken Language Systems Group, Saarland Informatics Campus**, Germany Thesis: "Natural Language Processing for African Languages" Advisor: Prof. Dr. Dietrich Klakow Grade: summa cum laude
JUL 2013 - *DEC* 2014	Master of Science in Computer Science **African University of Science and Technology (AUST)**, Abuja, Nigeria Thesis: "A DEVS-Based ANN Training and Prediction Platform" | Advisor: M.K. Traore Gpa: 3.91/4
OCT 2006 - *OCT* 2010	Bachelor of Science in Computer Science **Federal University of Agriculture**, Abeokuta, Nigeria Thesis: "Securing E-voting with EC-ELGAMAL" | Advisor: A.S. Sodiya Gpa: 4.42/5

Work Experience

Aug 2024 -	Assistant Professor (Tenure-track) *School of Computer Science, Faculty of Science, McGill University* Core Academic Member at Mila & Canada CIFAR AI Chair
Sep 2022 - Present	Senior Research Fellow (DeepMind Academic Fellow) *Department of Computer Science, University College London, UK.* Supported by DeepMind
Jan 2019 - Jun 2023	PhD Researcher (Group: Spoken Language Systems Group) *Department of Language Science & Technology, Saarland University* Supported by the EU H2020 COMPRISE project

Research Grants

Apr 2025 - Mar 2027	LLM for machine translation of endangered and low-resource languages (Co-PI with Fatiha Sadat, Jackie Cheung, and James Crippen, CA$60,000 funded by IVADO)
Feb 2025 - Jan 2026	End-to-end LLM-based speech translation for low-resource languages (Co-PI with Dr. Min Ma (Google), CA$50,000 funded by Google via MILA)
Sep 2024 - Aug 2025	Adapting LLMs for MT evaluation in African languages (Co-PI with Dr. Colin Cherry (Google), CA$50,000 funded by Google via MILA)
Aug 2024 - Jul 2029	Canada CIFAR AI Chair CA $1,020,000 funded by Canadian Institute for Advanced Research (CIFAR)
Apr 2023 - Sep 2024	NiLE: Nigerian Languages for Education (Co-PI with Kola Tubosun, US$45,000 funded by Africa AI4D-EduAI Hub)
Jul 2023 - Mar 2024	Conversational AI and Benchmark datasets for African languages (PI, US $207,814, A Masakhane project funded by Lacuna Fund 2021)
May 2022 - Apr 2023	YorùbáVoice: Developing Speech Corpus for Yoruba language (Co-PI with Kola Tubosun, US$20,000 funded by Translated-Imminent Grant 2021)
May 2021 - DEC 2022	Named Entity Recognition and Parts of Speech for African Languages (Co-PI with Peter Nabende, US$156,500, A Masakhane Project funded by Lacuna Fund 2020)

Research Interests

Multilingual Natural language processing (NLP)
NLP for low-resource languages
Large language model adaptation

Teaching Experience

Teaching	Multilingual Representation Learning – Winter 25 McGill University, Canada
Co-Teaching	Natural language Processing – Fall 24 McGill University, Canada (with Prof. Jackie Cheung)
Co-Teaching	Natural language Processing (2 lectures) – Winter 24 UCL, UK (with Dr. Pontus Stenetorp)
Co-Teaching	Natural language Processing (2 lectures) – Winter 23 UCL, UK (with Dr. Tim Rocktäschel and Dr. Pontus Stenetorp)

Awards and Scholarships

Oct. 2024	Dr.-Eduard-Martin-Preis (Dissertation Prize of Fakultät für Mathematik und Informatik) of the Saarland University Society, Universitat des Saarlandes
Nov. 2023	Area Chair Award at the IJCNLP-AACL 2023 for the MasakhaNEWS paper
Jun. 2023	summa cum laude. Ph.D. Graduation Honors, Saarland University, Germany.
May. 2023	Best Paper Award at the AfricaNLP Workshop @ICLR 2023 for the MasakhaNEWS paper
Oct. 2022	Best Paper Award (Grand Challenge KISTI category) at the COLING 2022, South Korea
Apr. 2021	Best Paper Award at the AfricaNLP Workshop @EACL 2021 for the MasakhaNER paper
May. 2019	Best Poster Presentation Award at the Web Conference (formerly WWW) 2019
Jun. 2018	NII International Internship Program, Japan (February - April 2019)
Mar. 2017	MPI-SWS Pre-Doctoral Studies Scholarship (ended December 2018)
Apr. 2015	Saarbrücken Graduate School of Computer Science Full Scholarship (ended Mar 2017)
Jul. 2013	Nelson Mandela Institute (NMI) Full Scholarship for MSc in Computer Science ($22,500)

Invited Talks

Nov 20, 2024	**Invited talk at Network of AI Safety Institues** at San Francisco How good are Large Language Models on African Languages?
Apr 26, 2023	**Invited talk @DeepMind**: NLP for under-resourced African languages
Mar 24, 2023	**Invited talk @McGill**: NLP for under-resourced African languages
Mar 23, 2023	**Invited talk @co:here**: NLP for under-resourced African languages
Feb 16, 2023	**Invited talk @CMU**: NLP for under-resourced African languages
Dec 3, 2022	**Transfer learning4NLP Workshop @NeurIPS**: Cross-lingual Transfer for NER
Nov 28, 2022	**Black in AI Workshop @NeurIPS**: Cross-lingual Transfer for NER
Jul 14, 2022	**Deeplo Workshop @NAACL**: Multilingual Language Model Adaptive Fine-Tuning: A Study on African Languages
Dec 6, 2021	**Tutorial @NeurIPS**: A Journey Through the Opportunity of Low Resourced Natural Language Processing — An African Lens
May 27, 2021	**UCREL CRS**: MasakhaNER: Named entity recognition for African Languages
Apr 19, 2021	**AfricaNLP Workshop @EACL**: MasakhaNER: NER for African Languages
Mar 26, 2021	**CMU-LTI Colloquium**: MasakhaNER: NER for African Languages
Sep 23, 2020	**NLP with Friends** : Development of NLP datasets and models for African Languages
Jul 7, 2020	**Data Skeptic Podcast**: Sentiment Preserving Fake Reviews
Jul 3, 2020	**AI4D Webinar**: Ensuring good text quality in African Language Datasets

Volunteer Service

Director at	Masakhane Research Foundation — grassroots NLP community for Africa
Organizer for	AfricaNLP 2022-2025, Cross-Cultural Considerations in NLP Workshop 2023, Multilingual Representation Learning Workshop 2023-2025
SAC for	ACL 2025
AC for	NAACL 2025, EMNLP 2024, ACL 2024, EACL 2024 & NAACL 2024 ACL 2023 & EMNLP 2023 (Multilingualism and/or Cross-Lingual NLP track)
Reviewer for	AfricaNLP Workshop (2020, 2021), WeaSUL @ICLR 2021, NAACL 2021, ARR 2021 - present, LREC 2022, COLING 2022, EMNLP 2022, EACL 2023

Most significant contributions

Over the past five years, I have made a number of important contributions for natural language processing for under-resourced languages especially African languages, including developing novel datasets and models for over 20 African languages. Many are highly cited (based on Google Scholar and Semantic Scholar), especially considering it is a relatively new area. I present five most important contributions, and details about their impact. For the **References**, I refer to my CV for **(J)**ournal, and **(C)**onference papers.

Participatory research methodology for addressing lack of datasets for under-resourced languages A major challenge in developing language technology for under-resourced languages is the lack of datasets for pre-training natural language processing (NLP) models and evaluating their performance on downstream tasks. Pre-training typically requires large amounts of unlabelled or web data, which are often scarce and of low quality. For evaluation, we require labelled data, which is also limited because annotations are costly in terms of financial resources and time-consuming for native speakers who volunteer for the annotation process. During my PhD, I collaborated with Masakhane, a grassroots organization in Africa working on African languages to address this issue. I proposed the use of participatory research methodology to tackle this challenge. This methodology ensures that various stakeholders (e.g., native speakers, linguists, and NLP experts) are involved in the development of *labelled datasets* for African languages.

I led a team of native speakers of African languages and NLP experts to build a small set of labeled sentences (i.e., 900 to 3,000) for the named entity recognition (NER) task in 10 African languages [J47], known as MasakhaNER. MasakhaNER has been cited over 100 times since the time it was first published in 2021. While leveraging annotations from volunteers through participatory research often results in a small dataset, we show that by applying the principles of transfer learning, we are able to create a high-quality NER model using the limited labeled data. To develop larger datasets, we typically apply for grants to compensate community members. For example, we developed a second version of MasakhaNER 2.0 with additional languages and greater size using this approach. The participatory approach has been used to create more than 10 new benchmarks for African languages, including topic classification [C14, C24], sentiment classification [C21, C40], machine translation [C38, C46], QA [C20], and text-to-speech [C37], among others. This has significantly improved the representation of African languages in many benchmarks and publications. One of the dataset papers that made use of participatory research for data collection, MasakhaNEWS [C24], received an *Area Chair Award* at the Asia-Pacific Chapter of the Association for Computational Linguistics for building a topic classification dataset for 16 African languages. Furthermore, we developed another topic classification dataset, SIB-200 [C14], based on the Flores-200 machine translation dataset for 200 languages and dialects. This contribution, although published last year has been cited 42 times according to Google scholar and constantly receives thousands of downloads every day on HuggingFace.[1]

Development of SoTA Language Models for African Languages Pre-ChatGPT, massively multilingual language models like multilingual BERT, XLM-RoBERTa, and multilingual T5 models often covered around 100 languages, excluding many low-resource languages. One of my most significant contributions has been the development of state-of-the-art (SoTA) language models for African languages. In 2022, we developed the best multilingual encoder model for African languages by adapting the XLM-R pre-trained language model to 17 native African languages and three widely spoken non-native official languages (English, French, and Arabic). The model we developed, known as AfroXLMR [C33], remains one of the best pre-trained models for fine-tuning on various downstream tasks. Furthermore, AfroXLMR received the *Best Paper Award* at the International Conference on Computational Linguistics (COLING 2022). Additionally, in the shared task on sentiment analysis for African languages, all the winning teams have consistently used our AfroXLMR model. AfroXLMR has been cited over 120 times according to Google Scholar and Semantic Scholar which shows the community is using the model. The model has been open-sourced and receives thousands of downloads on HuggingFace each month.[2].

[1] https://huggingface.co/datasets/Davlan/sib200

[2] https://huggingface.co/Davlan/afro-xlmr-large

Similarly, in 2023, we developed another SoTA model, known as AfriTeVa V2 [C18], a T5 model for 16 African languages and three widely spoken non-native languages (English, French, Arabic, and Portuguese). It achieved impressive performance compared to the multilingual T5.

Machine translation models with few thousand parallel sentences Machine translation often requires hundreds of thousands or even millions of parallel sentences, which are typically absent for low-resource languages. This work investigates how to optimally leverage existing pre-trained models to create low-resource translation systems for new *new languages* and *domains* using only a few thousand high-quality parallel sentences [C38]. The dataset developed was part of the pre-training for the large-scale Meta NLLB project covering 200 languages, a principle that has been followed by several subsequent papers. This paints a promising picture for the development of NLP technology for understudied languages: being able to customize these models for new language of interest with as little as 2K sentences and a few fine-tuning steps, MT developers and users from any language community are less dependent on choices and monetary interest of industry powerhouses from the Global North. This approach is already cited 63 times according to Google scholar and 105 times according to Semantic scholar.

Embedding based machine translation metric for African languages Scaling machine translation models to several languages is important, however, we will not be able to adequately monitor the progress without appropriate evaluation metrics. Machine translation N-gram matching metrics like BLEU show a weaker correlation with human judgments, especially for morphologically rich African languages. I proposed the topic to a PhD student I was co-supervising during my Post-doc to develop an embedding-based metric, *AfriCOMET*, for African languages by leveraging transfer learning with the COMET architecture, high-resource language datasets, and an African-centric pre-trained model. This is the first embedding-based metric for African languages, and more recent papers are now adopting it. We are developing an extension of this model to leverage high-quality African translation to English/French direction to further boost embedding metric performance. This extension project is funded by a grant by Google-Mila collaboration.

Detecting of generated texts from large language models This is one of the earliest papers to highlight the difficulty of detecting texts generated by language models. We demonstrated this using a low-skill attack on the GPT-2 model with 172M parameters [C34]. The generated texts are difficult for humans to distinguish, we provided a decent method to detect LLM-generated texts that many people have built on. This paper has been cited over 170 times since it was first published in 2020.

Publications

(*) Equal contributions

1. **David Ifeoluwa Adelani**, Jessica Ojo, Israel Abebe Azime, Jian Yun Zhuang, Xuanli He Jesujoba O. Alabi, and Millicent Ochiengand Sara Hooker et al. IrokoBench: A New Benchmark for African Languages in the Age of Large Language Models. In *NAACL*, 2025
2. **David Ifeoluwa Adelani**, A. Seza Dogruöz, Iyanuoluwa Shode, and Anuoluwapo Aremu. Does Generative AI speak Nigerian-Pidgin?: Issues about Representativeness and Bias for Multilingualism in LLMs. In *Findings of NAACL*, 2025
3. Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, **David Ifeoluwa Adelani**, Ibrahim Said Ahmad, and et al. AfriHate: A Multilingual Collection of Hate Speech and Abusive Language Datasets for African Languages. In *NAACL*, 2025
4. Genta Indra Winata, Frederikus Hudi, Patrick Amadeus Irawan, David Anugraha, Rifki Afina Putri, ..., **David Ifeoluwa Adelani**, En-Shiun Annie Lee, Shogo Okada, Ayu Purwarianti, Alham Fikri Aji, Taro Watanabe, Derry Tanti Wijaya, Alice Oh, and Chong-Wah Ngo. WorldCuisines: A Massive-Scale Benchmark for Multilingual and Multicultural Visual Question Answering on Global Cuisines . In *NAACL*, 2025

5. David Romero, Chenyang Lyu, Haryo Akbarianto Wibowo, Teresa Lynn, Injy Hamed, Aditya Nanda Kishore, Aishik Mandal, Alina Dragonetti, Artem Abzaliev, Atnafu Lambebo Tonja, Bontu Fufa Balcha, Chenxi Whitehouse, Christian Salamea, Dan John Velasco, **David Ifeoluwa Adelani**, David Le Meur, Emilio Villa-Cueva, and et al. CVQA: Culturally-diverse Multilingual Visual Question Answering Benchmark. In *NeurIPS D & B*, 2024
6. Orevaoghene Ahia, Anuoluwapo Aremu, Diana Abagyan, Hila Gonen, **David Ifeoluwa Adelani**, Daud Olamide Abolade, Noah A. Smith, and Yulia Tsvetkov. Voices unheard: Nlp resources and models for yorùbá regional dialects. In *EMNLP*, 2024
7. Kenza Benkirane, Laura Gongas, Shahar Pelles, Naomi Fuchs, Joshua Darmon, Pontus Stenetorp, **David Ifeoluwa Adelani**, and Eduardo Sánchez. Machine translation hallucination detection for low and high resource languages using large language models. In *EMNLP Findings*, 2024
8. Genta Indra Winata, Ruochen Zhang, and **David Ifeoluwa Adelani**. Miners: Multilingual language models as semantic retrievers. In *EMNLP Findings*, 2024
9. Jiayi Wang, **David Ifeoluwa Adelani**, Sweta Agrawal, Ricardo Rei, Eleftheria Briakou, Marine Carpuat, Marek Masiak, Xuanli He, Sofia Bourhim, Andiswa Bukula, Muhidin A. Mohamed, ..., and Pontus Stenetorp. AfriMTE and AfriCOMET: Empowering COMET to Embrace Under-resourced African Languages. In *NAACL*, 2024
10. Comfort Ilevbare, Jesujoba Alabi, **Adelani, David Ifeoluwa**, Firdous Bakare, Oluwatoyin Abiola, and Oluwaseyi Adeyemo. EkoHate: Abusive language and hate speech detection for code-switched political discussions on Nigerian Twitter. In *Workshop on Online Abuse and Harms (WOAH) @NAACL*, 2024
11. **Adelani, David Ifeoluwa**, A. Seza Dogruoz, André Coneglian, and Atul Kr. Ojha. Comparing LLM prompting with cross-lingual transfer performance on indigenous and low-resource Brazilian languages. In *AmericasNLP Workshop @NAACL*, 2024
12. Tolulope Ogunremi, Kola Tubosun, Anuoluwapo Aremu, Iroro Orife, and **Adelani, David Ifeoluwa**. ÌròyìnSpeech: A multi-purpose Yorùbá speech corpus. In *LREC-COLING*, 2024
13. Garry Kuwanto, Eno-Abasi E. Urua, Priscilla Amondi Amuok, Shamsuddeen Hassan Muhammad, Anuoluwapo ... Aremu, **Adelani, David Ifeoluwa**, Derry Tanti Wijaya, and Anietie Andy. Mitigating translationese in low-resource languages: The storyboard approach. In *LREC-COLING 2024*, 2024
14. **David Ifeoluwa Adelani**, Hannah Liu, Xiaoyu Shen, Nikita Vassilyev, Jesujoba Oluwadara Alabi, Yanke Mao, Haonan Gao, and Annie En-Shiun Lee. Sib-200: A simple, inclusive, and big evaluation dataset for topic classification in 200+ languages and dialects. In *EACL*, 2024
15. Jessica Ojo, Kelechi Ogueji, Pontus Stenetorp, and **David Ifeoluwa Adelani**. How good are Large Language Models on African Languages? 2023
16. Anuoluwapo Aremu, Jesujoba Oluwadara Alabi, and **David Ifeoluwa Adelani**. YORC: Yoruba Reading Comprehension dataset. 2023
17. Yihong Chen, Kelly Marchisio, Roberta Raileanu, **David Ifeoluwa Adelani**, Pontus Stenetor, Sebastian Riedel, and Mikel Artetx. Improving Language Plasticity via Pretraining with Active Forgetting. In *NeurIPS*, 2023
18. Akintunde Oladipo, Mofetoluwa Adeyemi, Orevaoghene Ahia, Abraham Owodunni, Odunayo Ogundepo, **David Adelani**, and Jimmy Lin. Better Quality Pre-training Data and T5 Models for African Languages. In *EMNLP*, 2023
19. Sebastian Ruder, J. Clark, Alexander Gutkin, Mihir Kale, Min Ma, Massimo Nicosia, Shruti Rijhwani, Parker Riley, Jean Michel A. Sarr, Xinyi Wang, John Wieting, Nitish Gupta, Anna Katanova, Christo Kirov, Dana L. Dickinson, Brian Roark, Bidisha Samanta, Connie Tao, **David Ifeoluwa Adelani**, Vera Axelrod, and et al. XTREME-UP: A User-Centric Scarce-Data Benchmark for Under-Represented Languages. In *EMNLP*, 2023
20. Odunayo Ogundepo, Tajuddeen Rabiu Gwadabe, Clara Rivera, J. Clark, Sebastian Ruder, **David Ifeoluwa Adelani**, Bonaventure F. P. Dossou, Abdoulahat Diop, Claytone Sika-

sote, Gilles Hacheme, Happy Buzaaba, and et al. AfriQA: Cross-lingual Open-Retrieval Question Answering for African Languages. In *EMNLP*, 2023

21. Shamsuddeen H. Muhammad, Idris Abdulmumin, Abinew A. Ayele, Nedjma Djouhra Ousidhoum, **David Ifeoluwa Adelani**, Seid M. Yimam, Ibrahim Said Ahmad, Meriem Beloucif, Saif M. Mohammad, Sebastian Ruder, and et al. AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages. In *EMNLP*, 2023 Also accepted at AfricaNLP 2023 @ICLR)

22. Francesco Tinner, **Adelani, David Ifeoluwa**, Chris Emezue, Mammad Hajili, Omer Goldman, ..., Rahul Aralikatte, Najoung Kim, and Duygu Ataman. Findings of the 1st Shared Task on Multi-lingual Multi-task Information Retrieval at MRL 2023. In Duygu Ataman, editor, *MRL Workshop @EMNLP*, 2023

23. Kathleen Siminyu, Jade Abbott, K lá Túb sún, Aremu Anuoluwapo, Blessing K. Sibanda, Kofi Yeboah, **David Adelani**, and et al. Consultative engagement of stakeholders toward a roadmap for African language technologies. In *Patterns*, 2023

24. **David Ifeoluwa Adelani**, Marek Masiak, Israel Abebe Azime, Jesujoba Oluwadara Alabi, Atnafu Lambebo Tonja, Christine Mwase, Odunayo Ogundepo, Bonaventure F. P. Dossou, Akintunde Oladipo, ..., and Pontus Stenetorp. MasakhaNEWS: News Topic Classification for African languages. In *IJCNLP-AACL*, 2023 Also accepted at AfricaNLP 2023 @ICLR)

25. Iyanuoluwa Shode, **David Ifeoluwa Adelani**, Jing Peng, and Anna Feldman. NollySenti: Leveraging Transfer Learning and Machine Translation for Nigerian Movie Sentiment Classification. In *ACL*, 2023

26. Zheng X. Yong, Hailey Schoelkopf, Niklas Muennighoff, Alham F. Aji, **David Ifeoluwa Adelani**, Khalid Almubarak, M Saiful Bari, Lintang Sutawika, Jungo Kasai, Ahmed Baruwa, Genta I. Winata, Stella Rose Biderman, Dragomir R. Radev, and Vassilina Nikoulina. BLOOM+1: Adding Language Support to BLOOM for Zero-Shot Prompting. In *ACL*, 2023

27. Cheikh M. Bamba Dione*, **David Ifeoluwa Adelani***, Peter Nabende, Jesujoba Alabi, Thapelo Sindane, Happy Buzaaba, Shamsuddeen Hassan Muhammad, Chris Chinenye Emezue, Perez Ogayo, and Dietrich Klakow. MasakhaPOS: Part-of-Speech Tagging for Typologically Diverse African languages . In *ACL*, 2023

28. Shamsuddeen H. Muhammad, Idris Abdulmumin, Seid M. Yimam, **David Ifeoluwa Adelani**, Ibrahim Said Ahmad, Nedjma Djouhra Ousidhoum, Abinew Ali Ayele, Saif M. Mohammad, Meriem Beloucif, and Sebastian Ruder. SemEval-2023 Task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval). In *SEM Eval Workshop @ACL*, 2023

29. Idris Akinade, Jesujoba Oluwadara Alabi, **David Ifeoluwa Adelani**, Clement Odoje, and Dietrich Klakow. ε KÚ: Integrating Yorùbá cultural greetings into machine translation. In *C3NLP Workshop @EACL (Also accepted AfricaNLP Workshop @ICLR*, 2023

30. Rendani Mbuvha, **David Ifeoluwa Adelani**, Tendani Mutavhatsindi, Tshimangadzo Rakhuhu, Aluwani Mauda, Tshifhiwa Joshua Maumela, Andisani Masindi, Seani Rananga, Vukosi Marivate, and T. Marwala. MphayaNER: Named Entity Recognition for Tshivenda. In *AfricaNLP Workshop @ICLR*, 2023

31. **David Ifeoluwa Adelani**, Md Mahfuz Ibn Alam, Antonios Anastasopoulos, Akshita Bhagia, Marta R. Costa-jussà, Jesse Dodge, Fahim Faisal, Christian Federmann, Natalia Fedorova, Francisco Guzmán, and et al. Findings of the WMT'22 Shared Task on Large-Scale Machine Translation Evaluation for African Languages . In *Proceedings of the Seventh Conference on Machine Translation at EMNLP*, 2022

32. **David Ifeoluwa Adelani**, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, Jesujoba O. Alabi, Shamsuddeen H. Muhammad, Peter Nabende, Cheikh M. Bamba Dione, ..., and Dietrich Klakow. MasakhaNER 2.0: Africa-centric Transfer Learning for Named Entity Recognition . In *EMNLP*, 2022

33. Jesujoba Oluwadara Alabi*, **David Ifeoluwa Adelani***, Marius Mosbach, and Dietrich Klakow. Adapting Pre-trained Language Models to African Languages via Multilingual Adaptive Fine-Tuning. In *COLING (Also accepted at AfricaNLP Workshop at ICLR 2022)*, 2022

34. Ernie Chang, Jesujoba Alabi, **David Ifeoluwa Adelani**, and Vera Demberg. Few-Shot Pidgin Text Adaptation via Contrastive Fine-Tuning. In *COLING (Also accepted at AfricaNLP Workshop at ICLR 2022)*, 2022

35. Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ili'c, Daniel Hesslow, Roman Castagn'e, Alexandra Sasha Luccioni, Franccois Yvon, Matthias Gallé, Jonathan Tow, Alexander M. Rush, ..., David Ifeoluwa Adelani, and Thomas Wolf. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. *ArXiv*, abs/2211.05100, 2022

36. Hugo Laurençon, Lucile Saulnier, Thomas Wang, Christopher Akiki, Albert Villanova del Moral, Teven Le Scao, ..., **David Ifeoluwa Adelani**, ..., Margaret Mitchell, Sasha Luccioni, and Yacine Jernite. The BigScience ROOTS Corpus: A 1.6TB Composite Multilingual Dataset . In *NeurIPS (Datasets and Benchmarks Track)*, 2022

37. Josh Meyer, **David Ifeoluwa Adelani**, Edresson Casanova, Alp Öktem, Daniel Whitenack Julian Weber, Salomon Kabongo, Elizabeth Salesky, Iroro Orife, Colin Leong, Perez Ogayo, Chris Emezue, Jonathan Mukiibi, Salomey Osei, and et al. BibleTTS: a large, high-fidelity, multilingual, and uniquely African speech corpus. In *Interspeech*, 2022

38. **David Ifeoluwa Adelani**, Jesujoba Oluwadara Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, , and et al. A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation. In *NAACL-HLT*, 2022

39. Miaoran Zhang, Marius Mosbach, **David Ifeoluwa Adelani**, Michael A. Hedderich, and Dietrich Klakow. MCSE: Multimodal Contrastive Learning of Sentence Embeddings. In *NAACL-HLT*, 2022

40. Shamsuddeen H. Muhammad, **David Ifeoluwa Adelani**, Sebastian Ruder, Ibrahim S. Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris C. Emezue, Saheed S. Abdullahi, Anuoluwapo Aremu, Alipio Jeorge, and Pavel Brazdil. NaijaSenti: A Nigerian Twitter Sentiment Corpus for Multilingual Sentiment Analysis. In *LREC*, 2022

41. Ali Davody, **David Ifeoluwa Adelani**, Thomas Kleinbauer, and Dietrich Klakow. TOKEN is a MASK: Few-shot Named Entity Recognition with Pre-trained Language Models. In *In Proceedings of the 25th International Conference on Text, Speech and Dialogue (TSD)*, 2022

42. En-Shiun Annie Lee, Sarubi Thillainathan, Shravan Nayak, Surangika Ranathunga, **David Ifeoluwa Adelani**, and Ruisi Su andArya D. McCarthy. Pre-Trained Multilingual Sequence-to-Sequence Models: A Hope for Low-Resource Language Translation? In *Findings of ACL*, 2022

43. Dawei Zhu, Michael A. Hedderich, Fangzhou Zhai, **David Ifeoluwa Adelani**, and Dietrich Klakow. Task-Adaptive Pre-Training for Boosting Learning With Noisy Labels: A Study on Text Classification for African Languages. In *Workshop on Insights from Negative Results in NLP 2022 at ACL (Also accepted at AfricaNLP Workshop at ICLR 2022)*, 2022

44. Iyanuoluwa Shode, **David Ifeoluwa Adelani**, and Anna Feldman. YOSM: A new Yorùbá Sentiment Corpus for Movie reviews. In *AfricaNLP 2022 Workshop at ICLR*, 2022

45. **David Ifeoluwa Adelani**, Miaoran Zhang, Xiaoyu Shen, Ali Davody, Thomas Kleinbauer, and Dietrich Klakow. Preventing author profiling through zero-shot multilingual back-translation. In *EMNLP*, 2021

46. **David Ifeoluwa Adelani** and Dana Ruiter and Jesujoba O. Alabi and Damilola Adebonojo and Adesina Ayeni and Mofe Adeyemi and Ayodele Awokoya and Cristina España-Bonet . The Effect of Domain and Diacritics in Yorùbá-English Neural Machine Translation. In *MT Summit*, 2021 (Also accepted at AfricaNLP Workshop @EACL)

47. **David Ifeoluwa Adelani**, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, and et al. MasakhaNER: Named Entity Recognition for African Languages . In *TACL*, 2021 (Also accepted at AfricaNLP Workshop @ EACL 2021)

48. Kathleen Siminyu, Godson Kalipe, Davor Orlic, Jade Abbott, Vukosi Marivate, Sackey Freshia, Prateek Sibal, Bhanu Neupane, **David I. Adelani**, Amelia Taylor, Jamiil Toure

ALI, Kevin Degila, Momboladji Balogoun, Thierno Ibrahima DIOP, and et al. AI4D – African Language Program. In *AfricaNLP Workshop at EACL*, 2021

49. Michael A. Hedderich, **David Ifeoluwa Adelani**, Dawei Zhu, Jesujoba O. Alabi, Udia Markus, and Dietrich Klakow. Transfer Learning and Distant Supervision for Multilingual Transformer Models: A Study on African Languages. In *EMNLP*, 2020
50. **David Ifeoluwa Adelani**, Ryota Kobayashi, Ingmar Weber, and Przemyslaw A. Grabowicz. Estimating community feedback effect on topic choice in social media with predictive modeling. In *EPJ Data Science*, 2020 (Also accepted at IC2S2 2021 as Extended Abstract)
51. **David Ifeoluwa Adelani**, Ali Davody, Thomas Kleinbauer, and Dietrich Klakow. Privacy guarantees for de-identifying text transformations. In *Interspeech*, 2020
52. Aleena Thomas, **David Ifeoluwa Adelani**, Ali Davody, Aditya Mogadala, and Dietrich Klakow. Investigating the impact of pre-trained word embeddings on memorization in neural networks. In *In Proceedings of the 23rd International Conference on Text, Speech and Dialogue (TSD)*, 2020
53. **David Ifeoluwa Adelani***, Michael A. Hedderich*, Dawei Zhu*, Esther van den Berg, and Dietrich Klakow. Distant supervision and noisy label learning for low resource named entity recognition: A study on Hausa and Yorùbá, Practical ML for Developing Countries Workshop at ICLR 2020 (Also accepted at AfricaNLP 2020 @ICLR)
54. Iroro Orife, **David I. Adelani**, Timi Fasubaa, Victor Williamson, Wuraola Fisayo Oyewusi, Olamilekan Wahab, and Kola Tubosun. Improving Yorùbá diacritic restoration, AfircaNLP Workshop at ICLR 2020
55. Ernie Chang, **David Ifeoluwa Adelani**, Xiaoyu Shen, and Vera Demberg. Unsupervised Pidgin Text Generation By Pivoting English Data and Self-Training, AfircaNLP Workshop at ICLR 2020
56. Jesujoba O. Alabi, Kwabena Amponsah-Kaakyire, **David I. Adelani**, and Cristina España-Bonet. Massive vs. curated word embeddings for low-resourced languages. the case of Yorùbá and Twi. In *12th International Conference on Language Resources and Evaluation (LREC)*, 2020
57. **David Ifeoluwa Adelani**, Haotian Mai, Fuming Fang, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. Generating sentiment-preserving fake online reviews using neural language models and their human- and machine-based detection. In *34th International Conference on Advanced Information Networking and Applications (AINA)*, 2020
58. Zijian Wang, Scott Hale, **David Ifeoluwa Adelani**, Przemyslaw Grabowicz, Timo Hartman, Fabian Flöck, and David Jurgens. Demographic inference and representative population estimates from multilingual social media data. In *The Web conference*. ACM, 2019 (Also accepted at IC2S2 2019 as Extended Abstract)
59. **David Ifeoluwa Adelani**, Fabian Flöck, Przemyslaw Grabowicz, Scott Hale, Timo Hartmann, and David Jurgens. Agendas on social media: Inferring policy attention from non-representative data. International Conference on Computational Social Science (IC2S2) 2018. Extended Abstract
60. **David Ifeoluwa Adelani** and Mamadou Kaba Traore. Enhancing the reusability and interoperability of artificial neural networks with DEVS modeling and simulation. *International Journal of Modeling, Simulation, and Scientific Computing*, 7(3), 2015
61. A.S. Sodiya, S.A. Onashoga, and **D.I. Adelani**. A secure e-voting architecture. In *Information Technology: New Generations (ITNG), 2011 Eighth International Conference on*, pages 342–347. IEEE, 2011

Favourite Quote

Anything you find your hands doing, do it with all your might