# David Ifeoluwa Adelani

## Personal Data

Place of Birth:	Abeokuta, Nigeria
ADDRESS:	London, United Kingdom
EMAIL:	d.adelani@ucl.ac.uk
WEBSITE:	dadelani.github.io or On Google Scholar

### **EDUCATION**

JAN 2015 - JUN 2023	
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 Affliated with MILA - Quebec AI Institute, Canada
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
Jan 2019 - Apr 2019	Research Intern (Group: Echizen's Content Security Lab) National Institute of Informatics (NII)
Mar 2017 - Dec 2018	Research Assistant (Group: Networked Systems Research Group) Max Planck Institute for Software System (MPI-SWS)
Aug 2016- Mar 2017	Research Assistant (Group: Intelligent User Interfaces) German Research Centre for Artificial Intelligence (DFKI)
Jul 2012 - Apr 2015	Computer Scientist II (Lead of Database Group) Federal Ministry of Communication Technology, Nigeria
Apr 2011 - Jul 2012	Professional Service Consultant (Junior) NCR (NIGERIA) PLC.

## **RESEARCH INTERESTS**

Natural language processing (NLP) for under-resourced languages Multilingual representation learning Transfer learning for NLP Privacy-preserving machine learning

## **TEACHING EXPERIENCE**

<b>CO-TEACHING</b>	Natural language Processing (2 lectures) – Winter 23
	UCL, UK (with Dr. Tim Rocktäschel and Dr. Pontus Stenetorp)
TEACHING ASSISTANT:	Project Seminar for Data Science and AI - Summer Semester 22
	Saarland University (with Prof. Dr. Jens Dittrich)
TEACHING ASSISTANT:	Neural Networks: Implementation and Application - Winter Semester 17/18
	Saarland University (with Prof. Dr. Dietrich Klakow)
TEACHING ASSISTANT:	Statistical Natural Language Processing - Summer 2017
	Saarland University (with Prof. Dr. Dietrich Klakow)

#### **Research Grants**

Apr 2023 - Sep 2024	NiLE: Nigerian Languages for Education
	(Co-PI with Kola Tubosun, \$45,000 funded by Africa AI4D-EduAI Hub )
Jul 2023 - Mar 2024	Conversational AI and Benchmark datasets for African languages
	(PI, \$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, \$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, \$156, 500, A Masakhane Project funded by Lacuna Fund 2020)
Jun 2020 - Dec 2020	MENYO-20k: A Multi-domain English-Yorùbá Corpus for MT
	(PI, \$2,000 funded by AI4D-Knowledge4All Fellowship)

## AWARDS, SCHOLARSHIPS AND CERTIFICATES

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
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)
DEC. 2012	Oracle Database 11g Certified Professional (OCP)
Aug. 2012	Oracle Database 11g Certified Associate (OCA)

## **VOLUNTEER SERVICE**

Area chair for	ACL 2024, EACL 2024 & NAACL 2024 (Efficiency in NLP)
	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
Organizer for	AfricaNLP 2022, 2023, Cross-Cultural Considerations in NLP Workshop 2023,
-	Multilingual Representation Learning Workshop 2023
Director at	Masakhane Research Foundation – grassroots NLP community for Africa

#### **INVITED TALKS**

Apr 26, 2023 Mar 24, 2023 Mar 23, 2023 Feb 16, 2023	Invited talk @DeepMind: NLP for under-resourced African languages Invited talk @McGill: NLP for under-resourced African languages Invited talk @co:here: NLP for under-resourced African languages 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

#### PUBLICATIONS

- (\*) Equal contributions
  - 1. 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
  - 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. 2023
  - 3. Jessica Ojo, Kelechi Ogueji, Pontus Stenetorp, and David Ifeoluwa Adelani. How good are Large Language Models on African Languages? 2023
  - 4. Anuoluwapo Aremu, Jesujoba Oluwadara Alabi, and David Ifeoluwa Adelani. YORC: Yoruba Reading Comprehension dataset. 2023
  - 5. 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
  - 6. 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
  - 7. 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

- 8. Odunayo Ogundepo, Tajuddeen Rabiu Gwadabe, Clara Rivera, J. Clark, Sebastian Ruder, David Ifeoluwa Adelani, Bonaventure F. P. Dossou, Abdoulahat Diop, Claytone Sikasote, Gilles Hacheme, Happy Buzaaba, and et al. AfriQA: Cross-lingual Open-Retrieval Question Answering for African Languages. In *EMNLP*, 2023
- 9. 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)
- 10. Francesco Tinner, David Ifeoluwa Adelani, 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
- 11. 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
- 12. 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)
- 13. 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
- 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
- 15. Cheikh M. Bamba Dione<sup>\*</sup>, **David Ifeoluwa Adelani**<sup>\*</sup>, 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
- 16. 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
- 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
- 18. 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
- 19. 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
- 20. 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

- 21. 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
- 22. 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
- 23. 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
- 24. 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
- 25. 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
- 26. 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
- 27. Miaoran Zhang, Marius Mosbach, **David Ifeoluwa Adelani**, Michael A. Hedderich, and Dietrich Klakow. MCSE: Multimodal Contrastive Learning of Sentence Embeddings. In *NAACL-HLT*, 2022
- 28. 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
- 29. 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
- 30. En-Shiun Annie Lee, Sarubi Thillainathan, Shravan Nayak, Surangika Ranathunga, David Ifeoluwa Adelani, and Ruisi Su andArya D. McCarthy. Pre-Trained Multilingual Sequenceto-Sequence Models: A Hope for Low-Resource Language Translation? In *Findings of ACL*, 2022
- 31. 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
- 32. Iyanuoluwa Shode, **David Ifeoluwa Adelani**, and Anna Feldman. YOSM: A new Yorùbá Sentiment Corpus for Movie reviews. In *AfricaNLP 2022 Workshop at ICLR*, 2022
- 33. **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
- 34. **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)

- 35. 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)
- 36. 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
- 37. 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
- 38. 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)
- 39. David Ifeoluwa Adelani, Ali Davody, Thomas Kleinbauer, and Dietrich Klakow. Privacy guarantees for de-identifying text transformations. In *Interspeech*, 2020
- 40. 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
- 41. **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)
- 42. 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
- 43. 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
- 44. 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
- 45. 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
- 46. 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)
- 47. 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

- 48. **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
- 49. 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