

# Ibrahim Said Ahmad, PhD

POSTDOCTORAL RESEARCH FELLOW · NORTHEASTERN UNIVERSITY

100 Fore Street, Portland, ME 04101, United States

☎ (857)-397-3275 | ✉ ibrahimsaidahmad@outlook.com | 🏠 isahmadbbr.github.io | 🌐 isahmadbbr

## Summary

I am a seasoned researcher and educator in Natural Language Processing (NLP) with a decade of experience in teaching, research, and interdisciplinary collaboration. My work focuses on low-resource language processing, where I have led large-scale, cross-continental projects spanning Africa, Asia, and Europe. As co-founder of Arewa Data Science, I am dedicated to expanding access to data science and machine learning, fostering advancements in AI research. My research contributions include developing state-of-the-art models for emotion analysis, hate speech detection, and sentiment analysis across 18 languages, as well as pioneering predictive modeling techniques for box-office revenue using social media data. With a strong publication record in high-impact journals, extensive grant funding experience, and active engagement in international conferences, I am committed to advancing equitable AI through teaching and research, mentoring future scholars, and bridging the gap between computational linguistics and real-world applications.

## Education

### Universiti Kebangsaan Malaysia

PHD COMPUTER SCIENCE

*Dissertation:* Purchase Intention Mining Approach for Movie Revenue Prediction.

Malaysia

2016 — 2020

### The University of Nottingham

MSC INFORMATION TECHNOLOGY

*Dissertation:* Engaging Children with ADHD using Mobile Games.

Malaysia

2013 — 2015

### Bayero University Kano

BSc COMPUTER SCIENCE

Nigeria

2006 — 2012

## Work Experience

### Northeastern University

POSITION: POSTDOCTORAL RESEARCH FELLOW, INSTITUTE FOR EXPERIENTIAL AI

100 Fore Street, Portland, ME 04101, USA.

United States

July 05 2023 — Present

### Bayero University Kano

LATEST POSITION: ASSISTANT PROFESSOR

Faculty of Computing, BUK Newsite, Gwarzo Road, Kano 700241, Nigeria.

Nigeria

Dec. 31 2014 — June 30 2023

## Research Interests

- Natural Language Processing and Computational Social Science.
- Cross-cultural collaborative research in NLP.
- Ethical AI and bias mitigation in multilingual NLP.
- Predictive Modeling for Business Applications.
- Democratization of data science and machine learning.

## Research Experience

---

### **Bridging the Gap in Text-Based Emotion Detection**

**Postdoctoral Research Fellow, Northeastern University**

**Sep. 2023 — Present**

- Developed large and high-quality emotion dataset for 18 African languages, which can be used for various NLP tasks, such as emotion detection, emotion generation, emotion analysis, and emotion synthesis.
- Conducting monolingual and multi-lingual experiments on multi-label emotion detection, emotion intensity classification and cross-lingual emotion detection.
- Organized SemEval shared task titled "SemEval 2025 Task 11: Bridging the Gap in Text-Based Emotion Detection."

### **Cultural Representations of Emotions in LLMs through Mixed Emotion Survey**

**Postdoctoral Research Fellow, Northeastern University**

**Sep. 2023 — Present**

- Designed and conducting experiments to study the emotion representations of Western and East Asian cultures, by studying the emotional responses of LLMs to various mixed-emotion situations.
- Using SOTA LLMs and human evaluators, studied the similarities and differences in which Western cultures (American, French, Spanish and German) and East Asian (Japanese, Vietnamese, Korean and Chinese) cultures experience mixed emotions, focusing on the co-occurrence of positive and negative emotions experienced by cultures in plausible situations of their daily lives.
- Found that the five different LLMs: Mistral, Gemma, Llama, GPT3.5, GPT4 demonstrate limited alignment with mixed emotions phenomena described in a previous study.

### **Exploring Cross-Cultural Alignment in LLMs Responses**

**Postdoctoral Research Fellow, Northeastern University**

**Sep. 2023 — June 2024**

- Designed and conducted experiments to investigate the cultural and emotional alignment of ChatGPT responses with those of native Hausa speakers.
- Crafted 37 culturally relevant prompts, validated by a psycholinguistic expert in Hausa language and culture.
- Collected native Hausa speakers and ChatGPT's responses to the prompts.
- Also collected and analyzed people's ratings of the responses generated by ChatGPT in terms of their alignment to Hausa culture.
- Used sentiment analysis to analyze the emotional content of responses from both ChatGPT and human participants.
- Applied BERTScore and METEOR to assess the textual similarity between ChatGPT and human responses.

### **AfriHate: A Multilingual Hate and Offensive Corpus for Africa**

**Postdoctoral Research Fellow, Northeastern University**

**Sep. 2023 — Present**

- Developed AFRIHATE the first labelled dataset for hate in 18 African languages from 6 countries.
- Benchmark baseline models (using SOTA pre-trained LLMs) for each language.
- Investigating various cultural, linguistic and technical challenges of working across so many diverse and low-resource languages on an inherently subjective concept of hate.

### **A Benchmark of WikiArt with Captions in 28 Languages**

**Postdoctoral Research Fellow, Northeastern University**

**July 2023 — Present**

- Collected 200K emotion labels and affective textual explanations in 25 languages on 2000 images (with ~140 annotations per image).
- Proposed a benchmark to evaluate the Zero-Shot, Few-Shot, and One-vs-All Zero-Shot performance of Multimodal models.
- Adapted and benchmark four contemporary Vision and Language models to work on our multilingual setup.
- Studied pairwise language transfer revealing insights on cultural differences in emotional perception and expression.

### **Analyzing COVID-19 Vaccination Sentiments in Nigerian Cyberspace**

**Lecturer, Bayero University Kano**

**Aug. 2021 — Dec. 2022**

- Twitter (now X) tweets collected using the discontinued Twitter Academic API with keywords related to covid-19 vaccination.
- Annotated tweets manually by three independent annotators based on covid-19 sentiments contained in the text.
- Analyzed the data and derived novel insights from the data using multiple exploratory data analysis techniques like topic modelling..
- Conducted experiments using classical machine learning models (e.g., SVM, Logistic Regression) and fine-tuned large language models (e.g., BERT, Afro-XLMR, XLM-RoBERTa).

### **A Nigerian Twitter Sentiment Corpus for Multilingual Sentiment Analysis**

**Lecturer, Bayero University Kano**

**December 2020 — July 2022**

- Developed the first publicly available large-scale manually annotated Twitter sentiment dataset for the four main Nigerian languages.
- Introduced a manually annotated sentiment lexicon in three languages (Hausa, Igbo, and Yorùbá).
- Presented benchmark experiments on Twitter sentiment dataset using state-of-the-art pre-trained language models and transfer learning.

### **Purchase Intention Mining Approach with Novel Features From Social Media for Movie Revenue Prediction**

**Doctoral Researcher, Universiti Kebangsaan Malaysia**

**March 2016 — April 2020**

- Developed a box-office revenue prediction corpus by collecting and analyzing movie trailer reviews, meta-data, and financial information from YouTube and Box Office Mojo.
- Developed a purchase intention lexicon (PIL) for classifying movie reviews based on purchase intention.
- Proposed the MRPI approach to extract purchase intention from movie reviews.
- Conducted extensive experiments with the corpus and evaluated the performance of prediction algorithms (MLR, SVM, multilayer perceptron NN, and RF) for box-office revenue. prediction.

## **Skills** \_\_\_\_\_

### **Technical Proficiencies:**

- Natural Language Processing (NLP): Low-resource language processing, sentiment analysis, hate speech detection, emotion analysis.
- Machine Learning: Deep learning, transfer learning, predictive modeling.
- Large Language Models: Development, fine-tuning, and adaptation for low-resource languages.

- Programming: Python, PyTorch, TensorFlow, Pandas, Numpy, Matplotlib
- Data Science: Statistical analysis, data visualization, feature engineering
- Tools: Git, Microsoft Office, Google Cloud, Jupyter Notebooks, Hugging Face Transformers, Notion

### Languages:

- Fluent: English, Hausa
- Working proficiency: English, Hausa, Malay, Arabic
- Linguistic analysis expertise in English and 20 low-resource languages.

### Research and Academic:

- Experimental design and implementation.
- Academic writing and publication.
- Grant writing and management.
- Cross-cultural collaborative research.

### Ethical AI and Responsible Development:

- Bias detection and mitigation in multilingual NLP.
- Ethical framework development for AI applications.
- Fairness and accountability in machine learning models.

### Project Management and Leadership:

- Team leadership and mentoring.
- Cross-cultural communication and collaboration.

## Grants & Honors

---

### GRANTS

2023	PI, Google Research exploreCSR award	\$18,000
2023	PI, Google Research TensorFlow award	\$10,000
2022	CO-PI, Artificial Intelligence for Education Innovation Research Network in Africa	\$40,000
2022	CO-PI, Oracle OCI Award	\$50,000
2022	CO-PI, Lacuna Fund for Labeled datasets for Language for Sub-Saharan Africa	\$107,895.84
2022	PI, Google Research exploreCSR award	\$18,000
2022	PI, TensorFlow award	\$10,000
2021	CO-PI, Lacuna Fund Grant for SemEval Shared Task	\$8,500
2020	CO-PI, Lacuna Fund for Labeled datasets for Language for Sub-Saharan Africa	\$60,923.00
2020	CO-PI, Google Covid-19 Research Grant	\$1,500.00
2020	CO-PI, TETFUND NRF Research Grant	\$50,000.00

### HONORS

2024	<b>Best paper mention</b> , 2024 12th International Conference on Affective Computing and Intelligent Interaction (ACII)
2024	<b>Best paper honorable mention</b> , SemEval 2024 for organizers of an individual shared task
2023	<b>Best Paper Award</b> , AfricaNLP workshop 2023, Kigali Rwanda
2023	<b>International list of AI solutions for reaching the 17 SGD</b> , IRCAI, Global Top 100 list 2022

## Supervision

2020-2022	<b>Amina Yusuf Umar</b> , Fake News Detection Based on A Hybrid CNN-GRU Deep Learning Model	MS Computer Science
2020-2022	<b>Boryanka Todorova Mashi</b> , An Improved LGBM Algorithm Based on Bayesian Optimization for Vulnerability Exploitation Prediction	MS Computer Science
2021-2024	<b>Abdulkadir Bello Bichi</b> , Lexical-Semantic Based Machine Learning Model for Fake News Detection	MS Computer Science
2021-Date	<b>Aisha Mustapha</b> , Negative Purchase Intent Identification from Social Media Texts	MS Computer Science

## Courses taught at Bayero University Kano

2022 & 2023	<b>Artificial Intelligence</b>	Graduate
2021 & 2022	<b>Database Systems</b>	Undergraduate
2015	<b>Introduction to probability</b>	Undergraduate
2015	<b>Multimedia techniques</b>	Undergraduate

## Selected Publications

**40+** peer-reviewed publications, among them **10+** first author papers. **Google Scholar**.

**Ibrahim Said Ahmad**, Ibrahim Said Ahmad, Shiran Dudy, Resmi Ramachandranpillai, and Kenneth Church. Are generative language models multicultural? A study on Hausa culture and emotions using ChatGPT. In Vinodkumar Prabhakaran, Sunipa Dev, Luciana Benotti, Daniel Hershcovich, Laura Cabello, Yong Cao, Ife Adebara, and Li Zhou, editors, Proceedings of the 2nd Workshop on Cross-Cultural Considerations in NLP, pages 98–106, Bangkok, Thailand. Association for Computational Linguistics. URL <https://aclanthology.org/2024.c3nlp-1.8>. (2024).

Youssef Mohamed, Runjia Li, **Ibrahim Said Ahmad**, Kilichbek Haydarov, Philip Torr, Kenneth Church, and Mohamed Elhoseiny. No Culture Left Behind: ArtELingo-28, a Benchmark of WikiArt with Captions in 28 Languages. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 20939–20962, Miami, Florida, USA. Association for Computational Linguistics (2024).

Shiran Dudy, **Ibrahim Said Ahmad**, Ryoko Kitajima, and Agata Lapedriza. "Analyzing Cultural Representations of Emotions in LLMs through Mixed Emotion Survey." arXiv preprint arXiv:2408.02143 (2024).

**Ibrahim Said Ahmad**, Lukman Jibril Aliyu, Abubakar Auwal Khalid, Saminu Muhammad Aliyu, Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Bala Mairiga Abduljalil, Bello Shehu Bello, and Amina Imam Abubakar. "Analyzing COVID-19 Vaccination Sentiments in Nigerian Cyberspace: Insights from a Manually Annotated Twitter Dataset." arXiv preprint arXiv:2401.13133 (2024).

Abdulrashid, Ismail, **Ibrahim Said Ahmad**, Aminu Musa, and Mohammed Khalafalla. "Impact of social media posts' characteristics on movie performance prior to release: an explainable machine learning approach." Electronic Commerce Research (2024).

Vargas, Francielle, Samuel Guimarães, Shamsuddeen Hassan Muhammad, Diego Alves, **Ibrahim Said Ahmad**, Idris Abdulmumin, Diallo Mohamed, Thiago Pardo, and Fabrício Benevenuto. "HausaHate: An Expert Annotated Corpus for Hausa Hate Speech Detection." In Proceedings of the 8th Workshop on Online Abuse and Harms (WOAH 2024), pp. 52-58. 2024.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Sebastian Ruder, **Ibrahim Said Ahmad**, et al. "NaijaSenti: A Nigerian Twitter Sentiment Corpus for Multilingual Sentiment Analysis," In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 590–602, Marseille, France. European Language Resources Association, 2022.

**Ibrahim Said Ahmad**, Azuraliza Abu Bakar, and Mohd Ridzwan Yaakub. "Movie revenue prediction based on purchase intention mining using YouTube trailer reviews." Information Processing & Management 57, no. 5 (2020): 102278.

**Ibrahim Said Ahmad**, Azuraliza Abu Bakar, Mohd Ridzwan Yaakub, and Shamsuddeen Hassan Muhammad. "A survey on machine learning techniques in movie revenue prediction." SN Computer Science 1, no. 4 (2020): 235.

**Ibrahim Said Ahmad**, Azuraliza Abu Bakar, Mohd Ridzwan Yaakub, and Mohammad Darwich. "Sequel movie revenue prediction model based on sentiment analysis." *Data Technologies and Applications* 54, no. 5 (2020): 665-683.

**Ibrahim Said Ahmad**, Azuraliza Abu Bakar, Mohd Ridzwan Yaakub, and Mohammad Darwich. "Beyond sentiment classification: A novel approach for utilizing social media data for business intelligence." *International Journal of Advanced Computer Science and Applications* 11, no. 3 (2020).

Ansari Abdollah, **Ibrahim Said Ahmad**, Azuraliza Abu Bakar, and Mohd Ridzwan Yaakub. "A hybrid metaheuristic method in training artificial neural network for bankruptcy prediction." *IEEE access* 8 (2020): 176640-176650.

**Presentations** \_\_\_\_\_

**CONTRIBUTED PRESENTATIONS**

**Ahmad, I. S.** January 2021. How to prepare for a proposal defense. A postgraduate seminar series presented at the Faculty of Computer Science and Information Technology, Bayero University Kano.

**Ahmad, I. S.** January 2021. Significant Tips for a new researcher. A webinar presented virtually for new postgraduate students.

**Ahmad, I. S.**, Bakar, A. A., & Yaakub, M. R. 2018. Predicting Movie Success from Trailer Reviews. In *Proceeding of 1st International Postgraduate Conference*. Nilai, Malaysia.

**Ahmad, I. S.** January 2017. Sentiment Analysis: Algorithms and Applications. A paper presented at a seminar organized by Faculty of Computer Science and Information Technology, Bayero University Kano.

**Ahmad, I. S.** January 2017. Sentiment Analysis: Algorithms and Applications. A paper presented at a seminar organized by Faculty of Computer Science and Information Technology, Bayero University Kano.

**Ahmad, I. S.** November 2015. Engaging Children with ADHD using Mobile Based Games. A paper presented at a seminar organized by Faculty of Computer Science and Information Technology, Bayero University Kano.

**Outreach & Professional Development** \_\_\_\_\_

**COMMUNITY ENGAGEMENT**

- 2024 **Organizer**, SemEval 2025 Task 11: Bridging the Gap in Text-Based Emotion Detection
- 2024 **Organizer**, International Conference on Spoken Language Translation (IWSLT) low-resource track shared task
- 2023 **Organizer**, SemEval-2023 Task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval)

**PROFESSIONAL MEMBERSHIPS**

- 1. **Member:** Internet Society (ISOC) 2020 — Date
- 2. **Member:** World Institute for Nuclear Security (WINS) 2020 — Date

**PEER REVIEW**

I served as reviewer for prominent journals including Information Processing & Management (Elsevier), Neural Computing and Applications (Springer Nature), SN Computer Science (Springer Nature) and IEEE Access. I also contributed to major conferences such as ACL and LREC. I am selected for the Program Committee of the 31st International Conference on Computational Linguistics (2025).