Who Wrote This? Identifying Machine vs Human-Generated Text in Hausa

Babangida Sani[1,3], **Aakansha Soy**[1], **Sukairaj Hafiz Imam**[2,3], **Ahmad Mustapha**[1,3], **Lukman Jibril Aliyu**[3], **Idris Abdulmumin**[3,4], **Ibrahim Said Ahmad**[2,3,5], **Shamsuddeen Hassan Muhammad**[2,3,6]

[1]Kalinga University, [2]Bayero University, Kano, [3]HausaNLP

[4]DSFSI, University of Pretoria, [5]Northeastern University, [6]Imperial College London

correspondence: bsani480@gmail.com

Abstract

The advancement of large language models (LLMs) has allowed them to be proficient in various tasks, including content generation. However, their unregulated usage can lead to malicious activities such as plagiarism and generating and spreading fake news, especially for low-resource languages. Most existing machine-generated text detectors are trained on high-resource languages like English, French, etc. In this study, we developed the first large-scale detector that can distinguish between human- and machine-generated content in Hausa. We scraped seven Hausa-language media outlets for the human-generated text and the Gemini-2.0 flash model to automatically generate the corresponding Hausa-language articles based on the human-generated article headlines. We fine-tuned four pre-trained Afri-centric models (AfriTeVa, AfriBERTa, AfroXLMR, and AfroXLMR-76L) on the resulting dataset and assessed their performance using accuracy and F1-score metrics. AfroXLMR achieved the highest performance with an accuracy of 99.23% and an F1 score of 99.21%, demonstrating its effectiveness for Hausa text detection. Our dataset is made publicly available[1] to enable further research.

Keywords: *Large Language Model (LLM), Natural Language Processing (NLP), Hausa, Transformer, Gemini, Fine-tune*

1 Introduction

Hausa is among the most spoken Chadic languages, belonging to the Afroasiatic phylum. Over 100 million people are estimated to speak the language, with the majority of speakers living in Northern Nigeria and the Republic of Niger, respectively (Inuwa-Dutse, 2021). However, from computational linguistics, it is regarded as a low-resource language, having insufficient resources to support tasks involving Natural Language Processing (NLP; Adam et al. 2023; Muhammad et al. 2023).

Hausa language is written in either the Latin (or ***Boko***) and Arabic (or ***Ajami***) script (Jaggar, 2006). The *Boko* script, existing since the 1930s, was introduced by the British colonial administration, and is used in education, government, and digital communication. The *Ajami* script, an order writing system of the Hausa language that existed in pre-colonial times, is used mostly in religious, cultural, and informal writing. For the purpose of our work, and as Hausa is widely written nowadays, we scraped and generated data based on the Latin-based script.

Large language models (LLMs) are becoming mainstream and easily accessible, ushering in an explosion of machine-generated content over various channels, such as news, social media, question-answering (QA) forums, educational, and even academic contexts (Wang et al., 2023). The human-like quality of texts generated by LLMs models for different languages including Hausa language is always advancing, allowing them to generate diverse content. LLMs, intentionally or unintentionally, have the potential to be used to create and propagate harmful or misleading content, such as fake news or hate speech (Xie et al., 2024), or even fake or artificial scholarship. To ensure the authenticity, accuracy, and trustworthiness of content, there is a need for machine-generated text detectors. Extensive research has been undertaken to differentiate between machine-generated texts (MGTs) and human-generated texts (HGTs), primarily employing model-based approaches (Wang et al., 2023; Alshammari, 2024; Ji et al., 2024).

In existing studies, (i) focus has mainly been on high-resource languages like English; (ii) there are no reliable detectors for detecting human vs. AI-generated text in the Hausa language; (iii) ensuring content authenticity is difficult, especially for low resource languages like Hausa (Ji et al., 2024). We

[1]https://github.com/TheBangis/hausa_corpus

Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025), pages 82–88

aim, therefore, to develop an automatic detector to classify human-generated and machine-generated text in Hausa, focused on the news domain, hence filling this gap. The following are our contributions:

- We are the first to develop a Hausa detector that is capable of differentiating HGT and MGT in Hausa. We believe it would help in ensuring content authenticity in digital communication, academia, and mitigating fake news.
- We curated a dataset that consists of human-generated data by scraping seven Hausa media outlets and machine-generated data using Gemini, addressing the lack of high-quality data in the area.
- By focusing on the Hausa language, we contribute to the expanding NLP capabilities for low-resource languages.
- All our resources will be open-source to encourage future academic research in the Hausa language.

2 Related Work

2.1 Detection of MGTs before ChatGPT

Radford et al. (2019) raised concerns regarding using machine-generated text for malicious purposes such as spam, fake news, plagiarism, and disinformation. The GLTR (Giant Language Model Test Room) tool (Gehrmann et al., 2019), released in June 2019, is an open-source system for detecting GPT-2-generated text using baseline statistical methods. Later that year, OpenAI enhanced the Roberta model (Liu et al., 2019) by introducing a dedicated GPT-2 detector (Radford et al., 2019). Another major advancement was the GROVER model (Zellers et al., 2019), which can both generate and detect fake news. With 5,000 self-generated articles and extensive real news content, GROVER achieved a 92% detection accuracy, surpassing models like the Plug and Play Language Model (PPLM) (Dathathri et al., 2019) and BERT (Devlin et al., 2019). Another study by Ippolito et al. (2019) examines the detection of machine-generated texts (MGTs) from GPT-2 with decoding strategies such as top-k, untruncated random sampling, and nucleus sampling in English. They discovered that optimized BERT was best but had poor cross-strategy generalization, whereas automatic classifiers performed better than humans, who misclassified AI text more than 30% of the time. AraGPT-2 (Antoun et al., 2020) introduced the first advanced Arabic language model that aides in distinguishing human-written and machine-generated Arabic text.

2.2 Detection of MGTs after ChatGPT

The launch of ChatGPT in late 2022 and later sequential models like GPT-4, Gemini, Claude, Llama, DeepSeek, etc., have posed new challenges as machine-generated texts (MGTs) mimic human writing styles more effectively than ever before. This raises concerns and the need for detection models to discern between HGTs and MGTs in different fields, such as academia, to mitigate plagiarism. In 2024, a study by Jawaid et al. (2024) presents a systemic approach for discerning between HWTs and MGTs using a combination of deep learning models, textual feature-based models, and machine learning models. Similarly, Xie et al. (2024) used eight traditional machine learning models and integrate statistical analysis, linguistic patterns, sentiment analysis and fact-checking as factors to differentiate between human-generated and machine-generated content across the three different datasets. In another study, Mitrović et al. (2023) examines ChatGPT-generated short text detection using DistilBERT and a perplexity-based classifier on online reviews, creating three datasets: human-written, ChatGPT-generated, and ChatGPT-rephrased. DistilBERT achieved 98% accuracy on original AI-generated text but only 79% on rephrased text, indicating the challenge of detecting AI-rewritten text.

3 Methodology

3.1 Datasets

We used both human-generated text (HGTs) and machine-generated text (MTGTs) for the Hausa language in the news domain. We collected 2,586 HGTs from seven different local and international news outlets and generated equal amounts of texts for the MGTs by leveraging the Gemini-2.0-flash closed-source model. We merged the datasets into a single file and created a source column to label whether a text is a HGT or MGT and then shuffled the data for effective training and evaluation. Table 1 displays selected samples from our HGT and MGT dataset used in the experiments and Table 2 provides information on the composition of our

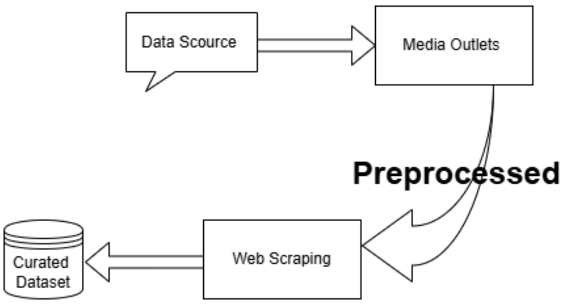

Figure 1: Overview of the pipeline's data collection for human-generated texts.

dataset in terms of the number of sentences, words, and unique words.

Human-Generated Data The human-generated data was extracted from seven different local and international online news outlet websites written in the Hausa language through web scraping, structured into headlines and content. Initially, we extracted 3,700 news articles; and after automatically filtering out unwanted texts, we preprocessed the dataset to remove rows with empty content, reducing the news articles to exactly 2,586. Figure 1 shows an overview of the pipeline's data collection process for human-generated data.

Machine-Generated Data For the machine-generated text, we used the Gemini-2.0 flash model, through Google AI Studio, to automatically generate corresponding Hausa news articles based on each of the article headlines in the human-generated text. The dataset consists of headlines, content, and a word count for every article so that the text length of the generated article is near or equal to the actual article. To produce the machine-generated articles, we processed the dataset in batches of 10 and checked whether machine-generated text existed and initiated missing values where required. For every batch, the model generated full articles from the headlines, and after every batch, progress was saved so as not to lose data, and a 10-second delay was added between batches to avoid exceeding API limits. This iterative process continued until machine-generated articles were created for all headlines.

Data Processing

The data collected, particularly the human-generated data, was noisy, containing many duplicates, English content, and other markup language symbols. Furthermore, some rows in the generated data produced an error message, while others included headlines that required cleaning. We identified and removed all unwanted content, including URLs, in both datasets before merging them into a single file to ensure effective training of our detection model.

Subdomain Analysis In order to get a clearer view of our dataset, we carried out an analysis of how articles are distributed in different news subdomains. The aim was to see how the content is distributed in terms of categories like politics, health, sports, business, entertainment, religion, and technology. Table 3 shows the distribution of news articles in different subdomains of the dataset.

The dataset is dominated by political content, which makes up more than half, 53.5% of the total articles. Categories such as health, 10.4% and sports, 9.7% have moderate representation, whereas business, 8.8%, entertainment, 7.9%, religion, 5.2%, and technology, 4.5% are less represented in the dataset.

Data Splits and Contamination Avoidance To ensure robust evaluation, we split the dataset into training (80%), validation (10%), and test (10%) sets, with no overlap of headlines across splits. Each headline (and its associated human and machine-generated articles) was grouped and assigned to a single split to avoid contamination. This ensures that the model does not encounter semantically similar content across training and evaluation phases. The dataset consists of 5172 samples (2586 human-generated and 2586 machine-generated), corresponding to 2586 unique headlines. These were divided into 2068 headlines (4136 samples) for training, 258 headlines (516 samples) for validation, and 260 headlines (520 samples) for testing.

4 Experiments setup

For our experiments, we utilized four Afri-centric transformer pre-trained language models: three multilingual and one monolingual. These models were selected due to their prior optimization for African languages, including Hausa. The models are AfriTeVa (Jude Ogundepo et al., 2022), AfroXLMR-76L (Adelani et al., 2023), AfroXLMR (Alabi et al., 2022), and AfriBERTa (Ogueji et al., 2021), respectively. Each pre-trained language model was fine-tuned on our constructed dataset by training it up 3 epochs. The models were optimized with AdamW, with a learning rate

Headline	Human-Generated Text	Machine-Generated Text
Man City ta kusan kammala aukar mai tsaron baya Reis	Manchester City ta kusan kammala cinikin matashin an wasan Palmeiras mai tsaron baya, Vitor Reis. City ta auki matakai tana zawarcin an wallon tawagar Brazil, yanzu dai ta kai ga cewa sauran iris ungiyar Etihad ta kammala sayen mai shekara 19. Tun farko Palmeiras ta so rie mai tsaron bayan har zuwa gasar cin kofin duniya, domin a lokacin zai kara tsada...	Manchester City na gab da kammala sayen an wasan baya mai suna Reis daga ungiyar Benfica. Rahotanni sun bayyana cewa, tattaunawa ta yi nisa sosai kuma ana sa ran za a cimma yarjejeniya a cikin 'yan kwanaki masu zuwa. Reis, wanda ya taka rawar gani a gasar Portugal, ya burge koci Pep Guardiola wanda ke neman arfafa layin tsaron ungiyarsa...
Matar Havertz ta samu saonnin cin zarafi a kafar sada zumunta	Matar Kai Havertz ta sanar da samun saonnin cin mutunci da zarafi a kafar sada zumunta, bayan tashi wasan da Manchester United ta fitar da Arsenal a FA Cup ranar Lahadi. Ta kuma ce cikin saonnin da ta samu har da na barazana da aka yi musu da an cikin da ko haifarsa ba a yi ba. Havertz mai shekara 25, ya barar da damar makin da ya kamata ya ci wallayen da zai kai Arsenal zagayen gaba a FA Cup, amma ya yi ta barar da wallaye. Bayan da suka tashi 1-1 har da arin lokacin, sai aka je bugun fenariti, inda mai tsare ragar United, Altay Bayindir ya tare wadda Havertz ya buga...	Matar an wasan wallon afa na asar Jamus, Kai Havertz, wato Sophia Weber ta fuskanci zazzafan cin zarafi a shafukan sada zumunta bayan wasan da ungiyar wallon afa ta Arsenal ta buga da Bayern Munich a gasar zakarun Turai. Masu amfani da shafukan sada zumunta sun yi wa Sophia ruwan zagi da maganganu masu ata rai, inda suka danganta rashin nasarar ungiyar Arsenal da arancin arfin mijinta a filin wasa. Wasu daga cikin saonnin sun yi nuni da cewa Sophia ce sanadiyyar rashin taka rawar gani na Havertz, yayin da wasu suka yi amfani da kalmomi masu zafi da cin mutunci...
Yadda wani farar fata ya kashe mata biyu ya bai wa alade gawarsu a Afirka ta Kudu	Labarin matan nan guda biyu da ake zargin wani farar fata da kashe su a gonarsa sannan ya mia wa alade gawarsu ya cinye na ci gaba da tayar da hankali a asar Afirka ta Kudu. Matan guda biyu da suka haa da Maria Makgato mai shekaru 45 da Lucia Ndlovu mai shekaru 34 dai an yi zargin cewa farar fatar ya harbe su ne a lokacin da suka shiga gonarsa domin neman abinci a kusa da Polokwane da ke arewacin lardin Limpopo na Afirka ta Kudun. An zargi farar fatar da jefa wa alade gawar matan a wani mataki na oarin oye shaidar abin da ya faru....	A wani lamari mai ban tsoro da takaici, wani farar fata ya aikata wani mugun aiki a kasar Afirka ta Kudu, inda ya kashe wasu mata biyu sannan ya bai wa alade gawarsu. Wannan lamari ya girgiza al'ummar kasar, ya kuma haifar da fushin jama'a, musamman a tsakanin bakaken fata. Rahotanni sun bayyana cewa, wanda ake zargin, wanda ba a bayyana sunansa ba tukuna, ya yi amfani da wani makami ne wajen kashe matan biyu a wani gida da ke wani yanki na kasar. Bayan ya aikata wannan aika-aika, sai ya dauki gawarwakin matan ya kai su wani gona da ake kiwon alade, inda ya jefa su a cikin kejin aladun...
Yadda fasinjojin Algeria suka makale a Paris	Yan Algeria 26 da za su koma gida daga Birtaniya sun makale a filin jirgin saman Charles de Gaulle da ke birnin Paris makonni uku da suka gabata. Fasinjojin wadanda suka hada da yara mata biyu da wata tsohuwa mai shekara 75 sun shigo Faransa ne a ranar 26 ga watan Fabareru daga filin jirgin sama na Heathrow...	A ranar Laraba, daruruwan fasinjojin jirgin sama 'yan kasar Aljeriya sun shiga halin kaka-nika-yi a filin jirgin sama na Charles de Gaulle da ke birnin Paris. Fasinjojin, wadanda suka yi niyyar komawa gida bayan ziyara ko tafiye-tafiye daban-daban, sun makale ne sakamakon soke jirage da kamfanonin jiragen sama suka yi ba zato ba tsammani...
Yan wasa 10 da hankali zai karkata kansu a AFCON	Nan da sa'o'i kaan ne, nahiyar Afrika za ta au harama ta ko'ina inda za a ria jin sowar magoya baya sakamakon Gasar cin Kofin nahiyar ta 2023, da za a fara karo na 34 a asar Cote d'Ivoire. Fitattun 'yan wasan nahiyar Afrika da suka yi fice a duniya za su baje-koli, daga ranar 13 ga watan Janairu zuwa 11 ga watan Fabrairu...	Gasar cin kofin nahiyar Afirka ta 2023 (AFCON) na gabatowa, kuma a wannan shekara ma, kamar yadda aka saba, akwai 'yan wasan da za su ja hankalin masoya kwallon kafa a fadin nahiyar da ma duniya baki daya. Daga cikin dubban 'yan wasan da za su fafata a wannan gasa, akwai wasu da ake ganin za su yi fice fiye da sauran saboda irin bajintar da suke nunawa a kungiyoyinsu da kuma kasashensu...
Babu gaskiya a zargin da shugaban sojin Nijar ya yi wa Najeriya - Ribadu	NUHU RIBADU Mai bai wa shugaban Najeriya shawara kan harkokin tsaro Malam Nuhu Ribau, ya nuna takaicinsa kan wasu zarge-zarge da shugaban mulkin sojin Nijar, Janar Abdulrahman Tchiani, ya yi yayin hirarsa da kafar talabijin in asar ranar Laraba. Shi dai Janar Tchiani ya zargi Najeriya da ba asar Faransa hadin kai wajen ba 'yan bindiga mafaka da kuma oarin kafa sansani a arewacin Najeriya, don shirya yadda za su far wa asarsa...	Mai ba shugaban kasa shawara kan harkokin tsaro, Nuhu Ribadu, ya yi watsi da zargin da shugaban sojin Nijar ya yi wa Najeriya, yana mai cewa babu gaskiya a cikin zargin. Ribadu ya bayyana haka ne a wata tattaunawa da manema labarai a Abuja, inda ya yi karin haske kan batun da ya jawo cece-kuce a 'yan kwanakin nan...

Table 1: Sample entries from our Human and Machine-Generated Hausa dataset

Statistic	Count
Total Sentences	6,737
Total Words	3,376,976
Unique Words	49,883

Table 2: Statistics of the dataset used in Hausa Machine-generated text detection.

of 1e-5, a batch size of 8, and a maximum sequence length of 512, with evaluation performed after each epoch. Table Table 4 shows the combinations of hyperparameters used to train the four models. The experiments were performed using PyTorch and Hugging Face Transformers. Upon completion, each fine-tuned model and tokenizer was saved and pushed to the Hugging Face Hub.

Subdomain	Proportion (%)
Politics	53.50
Health	10.38
Sports	9.72
Business	8.79
Entertainment	7.94
Religion	5.19
Technology	4.49

Table 3: Distribution of the news articles across different subdomains.

5 Results and Discussion

5.1 Results

Table 5 shows the performance of the fine-tuned models. The models' good performances indicate

Hyperparameter	Value
optimizer	AdamW
epochs	3
batch size	8
learning rate	1e-5

Table 4: Hyperparameters used for training the pre-trained language models.

their capabilities to distinguish between machine- and human-generated news texts written in Hausa language. Consistent in many downstream tasks, AfroXLMR performed the best, with an accuracy of 0.9923 and an F1 score of 0.9921. This is followed by AfriTeVa and AfriBERTa with an accuracy of 0.9884 and 0.9807 and an F1 score of 0.9881 and 0.9805, respectively, while AfroXLMR-76L had the lowest performance with an accuracy of 0.9672 and an F1 score of 0.9674. However, the overall performance indicates that all the developed models are very capable of detecting texts that are automatically generated from human-written news articles.

5.2 Discussion

The results of our experiments reveal the efficacy of pre-trained language models in detecting between human-generated text (HGTs) and machine-generated text (MGTs) in the Hausa language. AfroXLMR, was the best-performing model, achieved an accuracy of 99.23% on the test set and an F1 score of 99.21%, indicating its efficacy in identifying text origins with minimal misclassification. This suggests that multilingual pre-trained language models optimized for African languages can be fine-tuned effectively for low-resource language tasks such as MGT detection. Relative to the other three models, AfriTeVa, AfriBERTa, and AfroXLMR-76L showed different levels of performance. AfriTeVa achieved an accuracy of 98.84% on the test set, followed by AfriBERTa with 98.07% accuracy on the test set and lastly the AfroXLMR-76L achieved the lowest accuracy of 96.72% on the test set. This difference can be due to model architecture, pretraining data, and optimization methods.

6 Conclusion and Future Work

In this paper, we introduced the first large-scale effort to develop a detector capable of distinguishing between human-generated text (HGT) and machine-generated text (MGT) in the Hausa language. The study consists of two main parts. Firstly, we created a dataset consisting of both human-generated and machine-generated. Next, we developed and evaluated the detectors by fine-tuning four Afri-centric pre-trained language models on the dataset. The models are AfriTeVa, AfriBERTa, AfroXLMR, and AfroXLMR-76L. We trained the models multiple times to optimize hyperparameters and enhance performance. The experimental results revealed the efficacy of the proposed models, with AfroXLMR outperforming the other models, achieving an accuracy of 99.23% and an F1 score of 99.21%.

Model	Accuracy	F1 Score
AfriTeVa	0.9884	0.9881
AfriBERTa	0.9807	0.9805
AfroXLMR	**0.9923**	**0.9921**
AfroXLMR-76L	0.9672	0.9674

Table 5: Results and performance of the fine-tuned models. The best-performing model is highlighted in bold.

This study not only advances the detection of human-generated text and machine-generated text in a low-resource language such as Hausa but also shows that multilingual models optimized for African languages can be effectively adapted for detecting machine-generated text in low-resource languages. Support for low-resource languages is continuously improving across various large language models (LLMs). As a result, effective detection is important to prevent the spread of misinformation and disinformation, which are often facilitated by these models. We anticipate that this study will offer a comprehensive assessment of detection capabilities and enhance the ongoing academic discourse on identifying content generated by language models especially in underserved languages.

For future research, we aim to extend the dataset to cover diverse domains beyond news articles such as social media posts, academic writing, and books, as well as increasing the dataset size for better model generalization. Secondly, we aim to create real-time detection frameworks for use on digital platforms to help mitigate the propagation of AI-driven misinformation. Thirdly, exploring the use of the GPTs and other large language models in identifying machine-generated Hausa text. Using these models, with their high-level knowledge of language and context, detection accuracy could be enhanced. Lastly, to expand detection capabilities to other low-resource African languages, future research might explore cross-language transfer learning.

7 Limitations

Our study also has some limitations. First, we focused only on one domain when creating our dataset, which is news articles. The training was limited to three epochs, and a small batch size of 8 across all the models, which may impact the models' performance. Another limitation is that machine-generated texts were created using only the Gemini-2.0-flash model. While this model is high-performing, relying solely on a single source may limit the stylistic diversity of generated texts.

Acknowledgments

We thank HausaNLP Community for generously providing access to the Google Colab GPU Premium Version, greatly enhancing our model's training efficiency and supporting this research.

References

Fatima Muhammad Adam, Abubakar Yakubu Zandam, and Isa Inuwa-Dutse. 2023. Detection of offensive and threatening online content in a low resource language. *arXiv preprint arXiv:2311.10541*.

David Ifeoluwa Adelani, Hannah Liu, Xiaoyu Shen, Nikita Vassilyev, Jesujoba O Alabi, Yanke Mao, Haonan Gao, and Annie En-Shiun Lee. 2023. Sib-200: A simple, inclusive, and big evaluation dataset for topic classification in 200+ languages and dialects. *arXiv preprint arXiv:2309.07445*.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. 2022. Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4336–4349, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Hamed Alshammari. 2024. *AI-Generated Text Detector for Arabic Language*. Ph.D. thesis, University of Bridgeport.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. Aragpt2: Pre-trained transformer for arabic language generation. *arXiv preprint arXiv:2012.15520*.

Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2019. Plug and play language models: A simple approach to controlled text generation. *arXiv preprint arXiv:1912.02164*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Sebastian Gehrmann, Hendrik Strobelt, and Alexander M Rush. 2019. Gltr: Statistical detection and visualization of generated text. *arXiv preprint arXiv:1906.04043*.

Isa Inuwa-Dutse. 2021. The first large scale collection of diverse hausa language datasets. *arXiv preprint arXiv:2102.06991*.

Daphne Ippolito, Daniel Duckworth, Chris Callison-Burch, and Douglas Eck. 2019. Automatic detection of generated text is easiest when humans are fooled. *arXiv preprint arXiv:1911.00650*.

P J Jaggar. 2006. Hausa. *Elsevier Ltd*, pages 222–225.

Urwah Jawaid, Rudra Roy, Pritam Pal, Srijani Debnath, Dipankar Das, and Sivaji Bandyopadhyay. 2024. Human vs machine: An automated machine-generated text detection approach. In *Proceedings of the 21st International Conference on Natural Language Processing (ICON)*, pages 215–223.

Jiazhou Ji, Ruizhe Li, Shujun Li, Jie Guo, Weidong Qiu, Zheng Huang, Chiyu Chen, Xiaoyu Jiang, and Xinru Lu. 2024. Detecting machine-generated texts: Not just" ai vs humans" and explainability is complicated. *arXiv preprint arXiv:2406.18259*.

Odunayo Jude Ogundepo, Akintunde Oladipo, Mofetoluwa Adeyemi, Kelechi Ogueji, and Jimmy Lin. 2022. AfriTeVA: Extending ?small data? pretraining approaches to sequence-to-sequence models. In *Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing*, pages 126–135, Hybrid. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Sandra Mitrović, Davide Andreoletti, and Omran Ayoub. 2023. Chatgpt or human? detect and explain. explaining decisions of machine learning model for detecting short chatgpt-generated text. *arXiv preprint arXiv:2301.13852*.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id

Ahmad, Meriem Beloucif, Saif M. Mohammad, Sebastian Ruder, Oumaima Hourrane, Pavel Brazdil, Alipio Jorge, Felermino Dário Mário António Ali, Davis David, Salomey Osei, Bello Shehu Bello, Falalu Ibrahim, Tajuddeen Gwadabe, Samuel Rutunda, Tadesse Belay, Wendimu Baye Messelle, Hailu Beshada Balcha, Sisay Adugna Chala, Hagos Tesfahun Gebremichael, Bernard Opoku, and Stephen Arthur. 2023. AfriSenti: A Twitter sentiment analysis benchmark for African languages. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13968–13981, Singapore. Association for Computational Linguistics.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. Small data? no problem! exploring the viability of pretrained multilingual language models for low-resourced languages. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 116–126, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Chenxi Whitehouse, Osama Mohammed Afzal, Tarek Mahmoud, Toru Sasaki, et al. 2023. M4: Multi-generator, multi-domain, and multi-lingual black-box machine-generated text detection. *arXiv preprint arXiv:2305.14902*.

Yaqi Xie, Anjali Rawal, Yujing Cen, Dixuan Zhao, Sunil K Narang, and Shanu Sushmita. 2024. Mugc: Machine generated versus user generated content detection. *arXiv preprint arXiv:2403.19725*.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. *Advances in neural information processing systems*, 32.