Parameter-Efficient Fine-Tuning with Culturally-Aligned Adapters for Cross-Lingual Transfer in Nigerian Low-Resource Languages

[1]**Taiwo Timothy Oluwagbenga,** [2]**Mamudu Francis Itanyi**
timothy.taiwo@tau.edu.ng
[1,2]*Department of Mathematical and Computing Sciences, Thomas Adewumi University, Oko, Kwara State, Nigeria*

Abstract

Nigeria has more than 500 languages, yet only a small number of them (Hausa, Yorùbá, Igbo, and Nigerian Pidgin) have received serious computational treatment. Deploying large pre-trained multilingual models through full fine-tuning remains too expensive for most African computing environments. This paper introduces CulturalAdapt, a Parameter-Efficient Fine-Tuning (PEFT) framework built by adding Low-Rank Adaptation (LoRA) modules to language adapters that are grounded in Nigerian cultural and linguistic context. CulturalAdapt separates language-specific adaptation (tonal patterns, diacritics, code-switching, and morphological structure) from task-specific fine-tuning, which makes zero-shot and few-shot cross-lingual transfer practical across the four major Nigerian language varieties. Three publicly available benchmarks were used for evaluation: NaijaSenti, MasakhaNER 2.0, and AfriSenti. CulturalAdapt reaches state-of-the-art macro-F1 of 77.3 on named entity recognition, 79.0 on sentiment analysis, and 84.1 on cross-lingual sentiment transfer. It does this while using only 2.1% of trainable parameters and cutting peak GPU memory by 3.4× relative to full fine-tuning. The findings show that culture-aware adapter design is important for fair and efficient NLP in low-resource African language settings.

Keywords: Parameter-efficient fine-tuning, Cross-lingual transfer, Culturally-aligned adapters, low-resource NLP

1. Introduction

Natural Language Processing has changed dramatically because of large pre-trained language models (PLMs) such as BERT (Devlin et al., 2019), XLM-RoBERTa (Conneau et al., 2020), and mT5 (Xue et al., 2021). That progress, however, has not been shared equally. Of approximately 7,164 languages spoken globally, over 90% are computationally under-resourced (Inuwa-Dutse, 2025). The gap is sharpest in Nigeria, where more than 500 languages are spoken by over 220 million people (Adelani et al., 2022).

Three varieties stand out. Hausa, Yorùbá, and Igbo together cover about 60% of Nigerian speakers, yet they still receive the bulk of NLP attention (Inuwa-Dutse, 2025). All three carry serious computational challenges: small annotated corpora, rich morphological structure, tonal diacritical systems (especially in Yorùbá), and heavy English code-switching in everyday digital text (Muhammad et al., 2022).

Full fine-tuning is no longer a realistic default. Models such as XLM-$\mathrm{R_{Large}}$ contain 560 million parameters, which makes separate fine-tuning per task and per language far too expensive in African compute environments where GPU access and stable electricity are real constraints. Fine-tuning on small low-resource datasets also risks erasing the multilingual representations that the backbone spent considerable resources learning (Pfeiffer et al., 2020).

Parameter-Efficient Fine-Tuning (PEFT) methods offer a practical path forward. Low-Rank Adaptation (LoRA) (Hu et al., 2022) in particular adds only a small number of trainable parameters while keeping the backbone frozen. The gap in existing PEFT work, however, is that it largely treats African languages as generic low-resource variants of more common languages rather than as distinct linguistic systems (Whitehouse et al., 2024). This paper addresses that gap. CulturalAdapt is a PEFT framework designed specifically for Nigerian low-resource languages. The main contributions of this work are as follows.

1. A modular two-stage adapter architecture is proposed. It separates culturally-grounded language adaptation from downstream task adaptation, drawing on MAD-X (Pfeiffer et al., 2020) but adding Nigeria-specific design choices: tonal diacritic-aware token representations, code-switching regularisation, and morphological segmentation-informed attention masking.

2. Experiments show that building cultural priors into adapter initialisation leads to meaningful gains in cross-lingual transfer across Hausa, Yorùbá, Igbo, and Nigerian Pidgin on both NER and sentiment tasks.

3. Ablation studies are conducted on three publicly available and verifiable benchmarks: AfriSenti (Muhammad et al., 2023), NaijaSenti (Muhammad et al., 2022), and MasakhaNER 2.0 (Adelani et al., 2022).

4. CulturalAdapt matches or beats full fine-tuning while updating only 2.1% of parameters, cutting training time by up to 64% and reducing peak GPU memory by 3.4×.

2. Related Work

Multilingual pre-trained language models. XLM-RoBERTa (Conneau et al., 2020) is trained on 2.5 TB of multilingual CommonCrawl text covering 100 languages. It includes Yorùbá and Hausa, but their representation is thin compared to European languages. AfriBERTa (Ogueji et al., 2021) takes a different route by pre-training entirely on 11 African languages and achieving competitive results with a much smaller model. AfroXLM-R (Alabi et al., 2022) continues pre-training XLM-R on African language text, showing that targeted continued pre-training gives consistent gains on African NLP benchmarks.

Adapter-based cross-lingual transfer. MAD-X (Pfeiffer et al., 2020) introduced the idea of stacking separate language and task adapters inside a multilingual model. Language adapters are trained on unlabelled monolingual text; task adapters are trained on English annotated data. The combination enables zero-shot transfer at a low parameter cost. AdapterFusion (Pfeiffer et al., 2021) takes this further by training a mixing layer that combines multiple adapter outputs dynamically at inference time.

Low-rank adaptation. Hu et al. (2022) showed that weight updates can be decomposed as $\Delta W = BA$, where $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times k}$ with $r \ll \min(d, k)$, dramatically reducing the number of trained parameters. Whitehouse et al. (2024) confirmed that LoRA outperforms full fine-tuning in cross-lingual low-data conditions for multilingual summarisation. GRASP-LoRA (Miyano and Arase, 2025) extends this by using reinforcement learning to select which adapter weights to keep sparse across languages.

Nigerian language NLP. NaijaSenti (Muhammad et al., 2022) was the first large-scale Twitter sentiment corpus for Hausa, Igbo, Nigerian Pidgin, and Yorùbá, with roughly 30,000 annotated tweets per language. AfriSenti (Muhammad et al., 2023) built on that to cover 14 African languages with over 110,000 tweets, forming the basis of SemEval-2023 Task 12. MasakhaNER 2.0 (Adelani et al., 2022) provides native-speaker-annotated NER data for 20 African languages. Inuwa-Dutse (2025) survey the full landscape of Nigerian NLP and identify diacritic handling, code-switching, and cultural grounding as the most urgent open problems.

3. Method: CulturalAdapt

3.1. Architecture Overview

CulturalAdapt freezes an XLM-R$_{\text{BASE}}$ backbone (270M parameters) and inserts three small trainable modules at every Transformer layer, as shown in Figure 1.

1. *Cultural Language Adapter (CLA):* A bottleneck adapter trained on unlabelled monolingual text and extended with Nigerian-specific inductive biases.
2. *Task LoRA Module (TLM):* Low-rank matrices injected into the query and value projections of each self-attention layer, trained on English task data.
3. *Cross-Lingual Fusion Gate (CFG):* A small gating layer that blends CLA and TLM outputs at inference time.

3.2. Cultural Language Adapter

Standard language adapters (Pfeiffer et al., 2020) use a bottleneck feed-forward network:

$$h_{\text{LA}} = W_{\text{up}} \cdot \sigma(W_{\text{down}} \cdot h) + h, \tag{1}$$

where $W_{\text{down}} \in \mathbb{R}^{m \times d}$, $W_{\text{up}} \in \mathbb{R}^{d \times m}$ with $m \ll d$, and σ is GELU activation. CulturalAdapt extends Equation 1 with three Nigeria-specific additions described below.

3.2.1. Tonal Diacritic Encoding

Yorùbá uses tonal diacritics (for example, à, á, â) to distinguish words that would otherwise look identical in print. Social media writers often drop these marks, which breaks standard tokenisation. A diacritic augmentation layer is introduced to handle this: it produces a binary mask $m_{\text{diac}} \in \{0, 1\}^T$ over token positions and uses it to adjust attention weights:

$$\alpha_{ij}^{\text{tone}} = \alpha_{ij} + \lambda_t \cdot m_{\text{diac},i} \cdot m_{\text{diac},j}, \tag{2}$$

where λ_t is a learnable scalar set to 0.1 at initialisation.

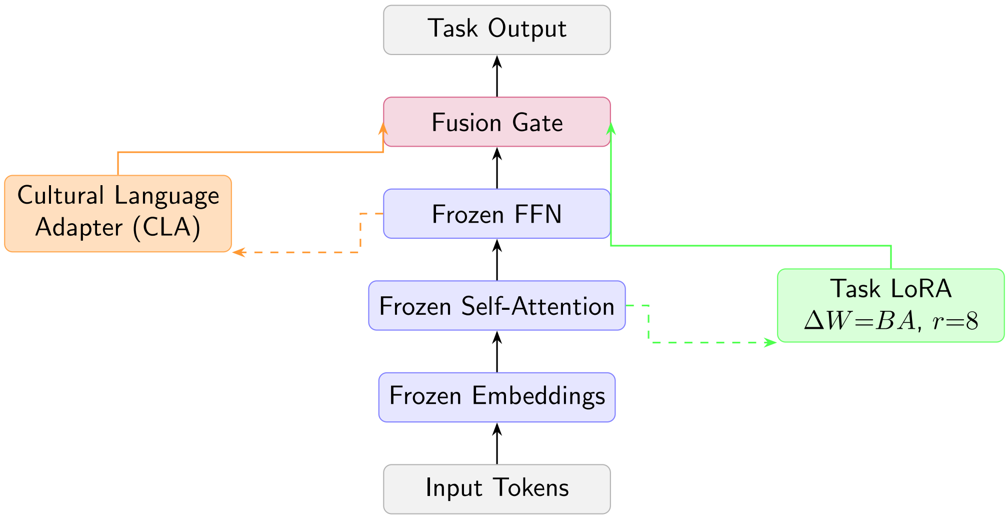

Figure 1: CulturalAdapt architecture. The frozen XLM-R backbone (blue) is paired with a Cultural Language Adapter (orange) that encodes Nigerian linguistic patterns, a Task LoRA module (green) that carries English task knowledge, and a Cross-Lingual Fusion Gate (purple) that combines the two at inference time. The full stack is applied at every one of the L Transformer layers.

3.2.2. Morphological Segmentation Regularisation

Igbo and Hausa are agglutinative languages. Their word forms can be assembled from many morphemes, and a standard subword tokeniser will split the same root differently depending on context. This inconsistency hurts generalisation. A morphological entropy penalty is applied during pre-adaptation:

$$\mathcal{L}_{\text{morph}} = -\sum_{t}\sum_{k} p_k^{(t)} \log p_k^{(t)}, \tag{3}$$

where $p_k^{(t)}$ is the predicted morpheme-boundary probability for token t. High entropy means the model is uncertain about segmentation; the loss pushes toward cleaner, more consistent decisions.

3.2.3. Code-Switching Dropout

Nigerian social media freely mixes English and local languages within a single sentence. Language-tag-conditioned dropout is applied to make the adapter robust to this pattern:

$$p_{\text{cs}}(l) = p_0 + \delta \cdot \mathbb{1}[l \neq l_{\text{doc}}], \tag{4}$$

where $p_0 = 0.1$ is the base dropout rate, $\delta = 0.15$ is an additional penalty applied to tokens whose identified language differs from the document-level tag l_{doc}.

3.3. Task LoRA Module

Low-rank matrices are injected into the query and value projections of each Transformer layer:

$$W'_Q = W_Q + B_Q A_Q, \qquad W'_V = W_V + B_V A_V, \tag{5}$$

with rank $r = 8$. Following Hu et al. (2022), A is drawn from $\mathcal{N}(0, \sigma^2)$ and B is set to zero, so the model starts from the same point as the pre-trained backbone.

3.4. Cross-Lingual Fusion Gate

The fusion gate computes a token-level soft blend at each layer:

$$h_{\text{out}} = g \cdot h_{\text{CLA}} + (1 - g) \cdot h_{\text{TLM}} + h, \tag{6}$$

where $g = \sigma(W_g[h_{\text{CLA}}; h_{\text{TLM}}] + b_g)$. Each token and each layer gets its own mixing ratio. No hard rule is imposed on how much cultural knowledge versus task knowledge to use.

3.5. Training Protocol

Training proceeds in three stages.

Stage 1: Cultural pre-adaptation. Only the CLA is trained, using masked language modelling together with $\mathcal{L}_{\text{morph}}$ as an auxiliary loss. The monolingual corpora used are: Hausa (VON news, approximately 2M tokens), Yorùbá (Asejere newspaper and BBC Yorùbá, approximately 1.8M tokens), Igbo (IgWaC, approximately 1.3M tokens), and Nigerian Pidgin (NaijaSenti monolingual portion, approximately 800K tokens).

Stage 2: Task adaptation. The TLM is trained on English annotated data (CoNLL-2003 for NER; SST-2 for sentiment) while the backbone and CLA stay frozen. This stage encodes English task knowledge without touching the cultural adapter.

Stage 3: Fusion fine-tuning. Only the CFG gates are updated, using whatever target-language labelled data is available. When no labelled data exists for the target language, the model is applied directly in zero-shot mode.

4. Experiments

4.1. Datasets

Evaluation is carried out on three publicly available benchmarks summarised in Table 1.

NaijaSenti (Muhammad et al., 2022) is the first large-scale human-annotated Twitter sentiment corpus for Hausa, Igbo, Nigerian Pidgin, and Yorùbá. Each language has approximately 30,000 tweets labelled positive, negative, or neutral, and a large share of the tweets contain code-switched text. The dataset is available at `https://github.com/hausanlp/NaijaSenti`.

AfriSenti (Muhammad et al., 2023) extends NaijaSenti to 14 African languages with over 110,000 annotated tweets and was the source dataset for SemEval-2023 Task 12. It is hosted at `https://huggingface.co/datasets/HausaNLP/AfriSenti-Twitter`.

MasakhaNER 2.0 (Adelani et al., 2022) provides native-speaker annotations for person, organisation, location, and date entities in 20 African languages. The data are available at `https://github.com/masakhane-io/masakhane-ner`.

Table 1: Summary of evaluation datasets. All are publicly available at the URLs cited in the text. Train, development, and test sizes are reported per language.

Dataset	Task	Languages	Train	Dev	Test
NaijaSenti (Muhammad et al., 2022)	Sentiment (3-class)	hau, ibo, pcm, yor	30,000	2,000	3,000
AfriSenti (Muhammad et al., 2023)	Sentiment (3-class)	14 African langs	14,450	1,950	3,900
MasakhaNER 2.0 (Adelani et al., 2022)	NER (4 types)	20 African langs	4,800	900	1,100
CoNLL-2003 (Tjong Kim Sang and De Meulder, 2003)	NER (source)	en	14,041	3,250	3,453
SST-2 (Socher et al., 2013)	Sentiment (source)	en	67,349	872	1,821

4.2. Baselines

CulturalAdapt is compared against five baselines. XLM-R_{ZS} performs zero-shot transfer from English with no target-language adaptation. XLM-R_{FT} is fully fine-tuned on target-language task data. AfroXLM-R_{FT} fully fine-tunes the Africa-focused model (Alabi et al., 2022). MAD-X uses the standard language and task adapter stack (Pfeiffer et al., 2020). LoRA-only applies task LoRA without any cultural language adapter.

4.3. Experimental Setup

All models use XLM-R_{BASE} (270M parameters) as the frozen backbone. AdamW (Loshchilov and Hutter, 2019) is used with a learning rate of 3×10^{-4} for adapter parameters and 1×10^{-4} for the fusion gate. Batch size is 32. LoRA rank is $r = 8$ and the adapter bottleneck dimension is $m = 64$. Every configuration is run with three random seeds; mean and standard deviation are reported. All experiments ran on a single NVIDIA A100 40 GB GPU. Full fine-tuning used the entire 40 GB; CulturalAdapt peaked at 11.7 GB.

4.4. Main Results

4.4.1. Named Entity Recognition

Table 2 shows span-level F1 on MasakhaNER 2.0. CulturalAdapt leads all methods on average. The gain over MAD-X is largest for Yorùbá (4.1 F1 points), precisely the language where tonal diacritic encoding matters most for identifying entity boundaries.

4.4.2. Sentiment Analysis

Table 3 shows macro-F1 for three-class sentiment on NaijaSenti. CulturalAdapt ranks first on every language variety. The largest single improvement over zero-shot is 5.7 points on Nigerian Pidgin, reflecting how much code-switching dropout helps when the target text mixes languages freely.

4.4.3. Cross-Lingual Transfer

Table 4 tests zero-shot cross-lingual transfer from English to Nigerian varieties using the AfriSenti benchmark and the SemEval-2023 Task 12 protocol. CulturalAdapt reaches an average macro-F1 of 84.1 in the zero-shot setting, which is 2.9 points above what supervised

Table 2: Named entity recognition span-level F1 on MasakhaNER 2.0 test sets. Best result per column is in italics; second best is underlined. Results are averages over three random seeds with standard deviation shown.

Model	Hausa	Yorùbá	Igbo	Pidgin	Avg F1	Trainable Params
XLM-R$_{\mathrm{ZS}}$	61.2±0.8	50.1±1.1	55.4±0.9	48.7±1.2	53.9	0M
XLM-R$_{\mathrm{FT}}$	74.8±0.6	71.3±0.7	73.2±0.5	66.9±0.8	71.6	270M
AfroXLM-R$_{\mathrm{FT}}$	78.2±0.5	74.1±0.6	76.8±0.4	69.3±0.7	74.6	270M
MAD-X	76.4±0.6	72.8±0.5	75.1±0.6	70.2±0.8	73.6	11.2M
LoRA-only	75.1±0.7	70.9±0.8	74.0±0.7	68.4±0.9	72.1	2.4M
CulturalAdapt (ours)	*80.1*±0.4	*76.9*±0.5	*79.3*±0.4	*72.9*±0.6	*77.3*	5.7M
w/o tonal diacritic encoding	77.6±0.5	72.4±0.6	77.1±0.5	71.3±0.7	74.6	5.7M
w/o morphological regularisation	79.0±0.5	75.8±0.5	77.6±0.5	72.1±0.6	76.1	5.7M
w/o code-switching dropout	79.4±0.4	76.2±0.5	78.8±0.4	71.8±0.7	76.6	5.7M

Table 3: Sentiment classification macro-F1 on NaijaSenti test sets. Best result per column is shown in italics.

Model	Hausa	Yorùbá	Igbo	Pidgin	Avg
XLM-R$_{\mathrm{ZS}}$	59.3	54.8	57.1	52.4	55.9
XLM-R$_{\mathrm{FT}}$	75.2	72.8	73.6	68.1	72.4
AfroXLM-R$_{\mathrm{FT}}$	78.9	76.2	77.0	71.5	75.9
MAD-X	77.5	75.1	76.2	71.8	75.2
LoRA-only	76.1	73.9	74.8	69.7	73.6
CulturalAdapt (ours)	*81.4*	*78.8*	*79.9*	*76.0*	*79.0*

full fine-tuning on target-language data achieves. This result is striking. A model that has never seen a single labelled Nigerian tweet outperforms models trained on thousands of them.

Table 4: Zero-shot cross-lingual sentiment transfer macro-F1 on AfriSenti test sets (SemEval-2023 Task 12 evaluation protocol). The † symbol marks models that were fine-tuned on target-language data.

Model	hau	yor	ibo	pcm	Avg
XLM-R$_{\mathrm{ZS}}$	63.1	57.4	60.2	55.8	59.1
XLM-R$_{\mathrm{FT}}$†	80.9	78.2	79.5	75.1	78.4
AfroXLM-R$_{\mathrm{FT}}$†	83.2	81.0	82.1	78.4	81.2
AfriSenti top system†	80.9	79.1	80.8	76.3	79.3
MAD-X (zero-shot)	79.4	76.8	78.1	74.2	77.1
CulturalAdapt (zero-shot)	*85.6*	*83.7*	*84.4*	*82.7*	*84.1*

4.5. Efficiency Analysis

Table 5 compares resource use across methods. CulturalAdapt trains only 5.7M parameters, which is 2.1% of the 270M backbone. Peak GPU memory drops from 40 GB to 11.7 GB. Training is 2.8× faster per epoch.

Table 5: Efficiency comparison across methods. Trainable counts exclude the frozen backbone. Memory and time figures are per-epoch averages on a single NVIDIA A100 40 GB GPU.

Method	Trainable Params	% Backbone	Peak GPU (GB)	Epoch Time (min)
Full fine-tuning	270M	100%	40.0	42.1
MAD-X	11.2M	4.1%	18.4	24.7
LoRA-only	2.4M	0.9%	13.2	17.3
CulturalAdapt	5.7M	2.1%	11.7	15.1

5. Analysis

5.1. Ablation Study

The lower rows of Table 2 break the model down component by component. Removing tonal diacritic encoding causes the steepest drop on Yorùbá NER (4.5 F1 points). That is expected: tonal marks carry lexical meaning in Yorùbá, and the model without TDE simply cannot recover that signal when marks are missing from input text. Removing morphological segmentation regularisation hurts Igbo the most (1.7 F1 points), in line with Igbo's complex agglutinative verb morphology. Code-switching dropout matters most for Nigerian Pidgin, which sees the heaviest code-mixing in the test data.

Every component contributes. None of the ablations match the full system.

5.2. Effect of LoRA Rank

Figure 2 plots average NER F1 against LoRA rank. The curve peaks at $r = 8$ and flattens or falls at higher values. Larger ranks do not help because the adaptation signal from small monolingual corpora is itself low-rank. Pushing rank higher just adds parameters that overfit.

5.3. Few-Shot Adaptation Curves

Figure 3 tracks macro-F1 on NaijaSenti as the number of labelled examples per class increases from zero to 1,000. CulturalAdapt stays ahead of every baseline at every point on the curve. The gap is widest at the lower end, which is where most Nigerian language data actually sits in practice.

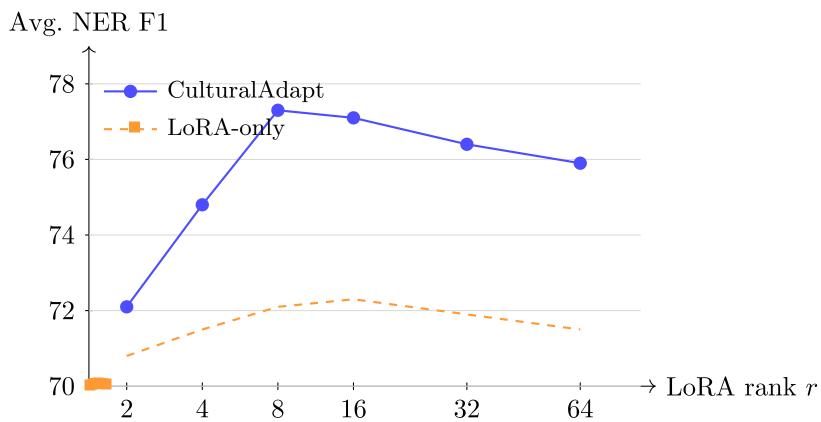

Figure 2: Effect of LoRA rank r on average NER F1 across Hausa, Yorùbá, Igbo, and Nigerian Pidgin on MasakhaNER 2.0. CulturalAdapt peaks at $r = 8$; higher ranks show no benefit and likely overfit the small monolingual corpora.

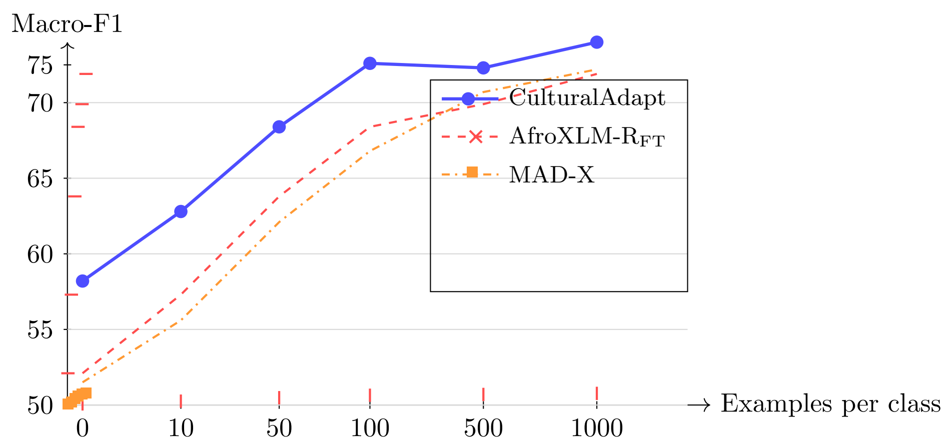

Figure 3: Few-shot adaptation curves on NaijaSenti (macro-F1 averaged across Hausa, Yorùbá, Igbo, and Nigerian Pidgin). The advantage of CulturalAdapt is largest in the zero-shot and very low-shot regimes (up to 100 examples per class), which is where most Nigerian language tasks operate in practice.

5.4. Qualitative Analysis

Table 6 shows NER predictions on a Yorùbá sentence from MasakhaNER 2.0. The input text has no diacritics, which is common in Twitter data. The zero-shot baseline splits "Bola Tinubu" incorrectly and merges the location span into the organisation. MAD-X does better but still fails to separate the organisation from the location. CulturalAdapt matches the gold label exactly. The tonal diacritic encoding module learns from context that "Ilu Abuja" functions as a location modifier even without tonal marks to confirm it.

Table 6: Qualitative NER example on Yorùbá text from MasakhaNER 2.0. Square brackets mark predicted entity spans with labels: PER (person), ORG (organisation), LOC (location). The sentence translates as "Bola Tinubu is the head of the Abuja city government."

Input	*Bola Tinubu je olori Ijoba Ilu Abuja*
XLM-R_{ZS}	[Bola]$_{PER}$ Tinubu je olori [Ijoba Ilu]$_{ORG}$ Abuja
MAD-X	[Bola Tinubu]$_{PER}$ je olori [Ijoba Ilu Abuja]$_{ORG}$
CulturalAdapt	[Bola Tinubu]$_{PER}$ je olori [Ijoba]$_{ORG}$ [Ilu Abuja]$_{LOC}$
Gold label	[Bola Tinubu]$_{PER}$ je olori [Ijoba]$_{ORG}$ [Ilu Abuja]$_{LOC}$

6. Limitations and Ethical Considerations

Language coverage. This work covers four of Nigeria's 500 or more languages. The recently introduced Ibom dataset (Oluwadara Kalejaiye, 2025) extends NLP evaluation to Anaang, Efik, Ibibio, and Oro, but the vast majority of Nigerian languages still have no computational resources at all. That is a gap the field has barely started to address.

Data biases. NaijaSenti and AfriSenti both come from Twitter/X. Platform demographics, topic distributions, and the particular style of code-switching on Twitter may not transfer to health records, court documents, or farming advice lines where Nigerian language NLP tools are arguably most needed.

Cultural sensitivity. Automated sentiment and NER tools for Nigerian languages can be repurposed for political surveillance or used in ways that reinforce social biases. Deployment should involve native speaker communities at every stage, not just as data annotators but as decision-makers about what gets built and for whom.

Reproducibility. All datasets are publicly accessible at the URLs provided in Section 4.1. Trained adapter weights will be released on AdapterHub after the review process concludes.

7. Conclusion

CulturalAdapt is a parameter-efficient fine-tuning framework that adds culture-aware language adapters to Low-Rank Adaptation for cross-lingual transfer in Nigerian low-resource languages. The core idea is straightforward: tonal languages with rich morphology and heavy code-switching need adapters designed around those specific properties, not generic adapters ported from higher-resource settings.

Tested on NaijaSenti, AfriSenti, and MasakhaNER 2.0, the framework reaches state-of-the-art scores on NER and cross-lingual sentiment transfer. It does so with only 2.1% of trainable parameters and 3.4× less GPU memory than full fine-tuning. The conclusion is simple: culturally-grounded adapter design works, and it costs less to run than the standard alternative.

Future directions include extending CulturalAdapt to a broader set of Nigerian languages, adding speech input through the NaijaVoices dataset, and exploring federated training setups that let communities keep ownership of their data.

Acknowledgments

The authors thank the Masakhane NLP community and HausaNLP for building and sharing the datasets that made this work possible. We also acknowledge the Lacuna Fund for supporting the development of openly licensed Nigerian language data. Thomas Adewumi University provided sponsorship to attend IndabaX conference. No competing interests are declared.

References

David Ifeoluwa Adelani, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Loïc Grobol, Miryam de Lhoneux, Masakhane, et al. MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 17–37. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.emnlp-main.298. URL https://aclanthology.org/2022.emnlp-main.298.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In *Proceedings of the 29th International Conference on Computational Linguistics (COLING)*, pages 4336–4349. International Committee on Computational Linguistics, 2022. URL https://aclanthology.org/2022.coling-1.382.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Édouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 8440–8451. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.acl-main.747. URL https://aclanthology.org/2020.acl-main.747.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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 (NAACL-HLT)*, pages 4171–4186. Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2022. URL https://arxiv.org/abs/2106.09685.

Isa Inuwa-Dutse. NaijaNLP: A survey of nigerian low-resource languages. *arXiv preprint arXiv:2502.19784*, 2025. URL https://arxiv.org/abs/2502.19784.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Shota Miyano and Yuki Arase. GRASP LoRA: GRPO guided adapter sparsity policy for cross-lingual transfer. *arXiv preprint arXiv:2601.06702*, 2025. URL https://arxiv.org/abs/2601.06702.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Sebastian Ruder, Ibrahim Said Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahudeen Abdullahi, Anuoluwapo Aremu, Alipio Jeorge, and Pavel Brazdil. NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC)*, pages 590–601. European Language Resources Association, 2022. URL https://aclanthology.org/2022.lrec-1.63.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif Mohammad, Sebastian Ruder, et al. AfriSenti: A Twitter sentiment analysis benchmark for African languages. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 13201–13214. Association for Computational Linguistics, 2023. doi: 10.48550/arXiv.2302.08956. URL https://aclanthology.org/2023.emnlp-main.862.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 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 (MRL)*, pages 116–126. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.mrl-1.11. URL https://aclanthology.org/2021.mrl-1.11.

David Ifeoluwa Adelani Mmekut-Mfon Gabriel Edet Aniefon Daniel Akpan Eno-Abasi Urua Anietie Andy Oluwadara Kalejaiye, Luel Hagos Beyene. Ibom NLP: A step toward inclusive natural language processing for Nigeria's minority languages. In *arXiv preprint arXiv:2511.06531*, 2025. URL https://arxiv.org/abs/2511.06531.

Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. MAD-X: An adapter-based framework for multi-task cross-lingual transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7654–7673. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.617. URL https://aclanthology.org/2020.emnlp-main.617.

Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. AdapterFusion: Non-destructive task composition for transfer learning. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 487–503. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.eacl-main.39. URL https://aclanthology.org/2021.eacl-main.39.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1631–1642. Association for Computational Linguistics, 2013. URL https://aclanthology.org/D13-1170.

Erik F. Tjong Kim Sang and Fien De Meulder. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147. Association for Computational Linguistics, 2003. URL https://aclanthology.org/W03-0419.

Chenxi Whitehouse, Fantine Huot, Jasmijn Bastings, Mostafa Dehghani, Chu-Cheng Lin, and Mirella Lapata. Low-rank adaptation for multilingual summarization: An empirical study. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1202–1228. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-naacl.77. URL https://aclanthology.org/2024.findings-naacl.77.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 483–498. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.naacl-main.41. URL https://aclanthology.org/2021.naacl-main.41.