Parameter-Efficient Fine-Tuning for Low-Resource Languages: A Systematic Multi-Task Evaluation on Hausa Language

Abubakar Abdulhamid[1],
Shamsuddeen Hassan Muhammad[2,3], **Idris Abdulmumin**[3,4]
[1]Nasarawa State University, Keffi
[2]Imperial College London, [3]HausaNLP, [4]University of Pretoria
abdulhamid@ab-bkr.com

Abstract

Parameter-efficient fine-tuning methods such as LoRA, Prefix Tuning, and $(IA)^3$ have transformed NLP for high-resource languages by reducing computational costs without sacrificing performance. However, their effectiveness for African languages remains underexplored, despite urgent need for resource-efficient methods in computationally-constrained research environments. We present a systematic comparison of four fine-tuning strategies (Full FT, LoRA, Prefix Tuning, $(IA)^3$) across four base models (AfroXLMR-large, AfroXLMR-base, AfriBERTa-large, XLM-RoBERTa-base) on three HausaNLP tasks: sentiment analysis (AfriSenti), named entity recognition (MasakhaNER 2.0), and machine-generated text detection (Sani et al., 2025). Our 120 experiments yielded three key findings. **First**, AfroXLMR-large combined with LoRA recovers 93.3–96.6% of full fine-tuning performance using only 0.14–0.33% of trainable parameters and achieving up to 2.6× training speedup. **Second**, we identify a novel *amplification effect*: the advantage of Afri-centric models over multilingual baselines is substantially larger under parameter-efficient fine-tuning than under full fine-tuning (∼3× amplification), suggesting that adapters preserve language-specific knowledge that full fine-tuning may overwrite. **Third**, prefix tuning fails dramatically on token-level NER tasks (F1 as low as 0.20), with high seed variance (σ up to 0.07), making it unsuitable for sequence labeling in low-resource African language settings. We release all code, trained adapter weights, and experimental configurations to support resource-efficient African NLP research.

1 Introduction

The rapid advancement of large-scale pretrained language models has transformed Natural Language Processing (NLP), yet the benefits remain unevenly distributed across the world's languages. African languages, despite serving over a billion speakers, remain severely underrepresented in modern NLP systems (Hussen et al., 2025). Recent quantitative analyses document this disparity: Belay et al. (2025) trace the rise of AfricaNLP research over the past two decades using 1.9K papers, while Hussen et al. (2025) demonstrate that fewer than 2% of African languages are supported by current Large Language Models, with persistent challenges including data scarcity, tokenization biases, and prohibitive computational costs.

For Nigerian languages specifically, Inuwa-Dutse (2025) reveal that only 25.1% of NaijaNLP studies contribute new linguistic resources, indicating systemic underproduction of computational resources. Hausa, an Chadic language with over 120 million first-language and 80 million second-language speakers across West Africa validates these challenges(Muhammad et al., 2025). Despite its status as the 19th most-spoken language globally and a regional lingua franca, HausaNLP infrastructure remains nascent, with most existing work focused on full fine-tuning of multilingual models. Full fine-tuning of large multilingual models such as AfroXLMR-large (Alabi et al., 2022) requires updating hundreds of millions of parameters and substantial GPU memory, presenting a significant barrier for African research institutions and practitioners. Parameter-efficient fine-tuning (PEFT) methods address this barrier by training only a small fraction of parameters while keeping the majority of the pretrained model frozen (Houlsby et al., 2019; Hu et al., 2022; Li and Liang, 2021; Liu et al., 2022). While these methods have been extensively benchmarked for English and other European languages, to our knowledge no systematic evaluation exists for Hausa.

This paper presents a comparison of parameter-efficient fine-tuning methods. Our contributions are:

1. **Empirical benchmark:** We evaluated four fine-tuning strategies (Full FT, LoRA, Prefix Tuning, $(IA)^3$) across four base models (AfroXLMR-large/base, AfriBERTa-large, XLM-RoBERTa-base) on three HausaNLP tasks namely: sentiment analysis (AfriSenti, Muhammad et al. 2023), named entity recognition (MasakhaNER 2.0, Adelani et al. 2022), and machine-generated text detection (Sani et al. 2025). Across 120 experiments with three seeds for adapter methods, we provide reliable benchmark results for HausaNLP practitioners.

2. **Practical recommendations:** We demonstrate that AfroXLMR-large combined with LoRA achieves 93.3–96.6% of full fine-tuning performance using only 0.14–0.33% of parameters and 1.89–2.64× faster training, providing a clear, actionable recommendation for resource-constrained HausaNLP research.

3. **Novelty of adapter amplification effect:** We identify that the advantage of Afri-centric models over multilingual baselines is substantially *amplified* under parameter-efficient fine-tuning (∼3× amplification on average), suggesting that adapters better preserve language-specific knowledge from pretraining.

4. **Prefix tuning unsuitability:** We document that prefix tuning fails on token-level sequence labeling (NER F1 as low as 0.20) with high variance across seeds, providing critical guidance against using this method for similar low-resource African language tasks.

2 Background and Related Work

2.1 State of HausaNLP

In recent years, we have witnessed substantial growth in HausaNLP (Muhammad et al., 2025; Inuwa-Dutse, 2021). Foundational works include the first Hausa fake news corpus (Imam et al., 2022), subsequent fake news detection via transfer learning (Ibrahim et al., 2024), sentiment analysis benchmarks (Muhammad et al., 2022, 2023), named entity recognition datasets (Adelani et al., 2021, 2022), machine translation resources (Abdulmumin et al., 2022), and most recently, machine-generated text detection (Sani et al., 2025). The HausaNLP Catalog[1] aggregates all these resources.

[1] https://catalog.hausanlp.org

Further, several Afri-centric pretrained models have been developed to address the limited Hausa coverage in multilingual models. **AfroXLMR** (Alabi et al., 2022) extends XLM-RoBERTa via Multilingual Adaptive Fine-Tuning on 17 African languages including Hausa. **AfriBERTa** (Ogueji et al., 2021) pretrains a multilingual model on 11 African languages with limited data. **AfroLM** (Dossou et al., 2022) uses self-active learning across 23 African languages. Sani et al. (2025) report AfroXLMR achieving 99.23% accuracy on Hausa machine-generated text detection, establishing it as a strong baseline.

These efforts unfold against a broader landscape that recent surveys characterize as rapidly growing yet persistently under-resourced. Belay et al. (2025) provided a quantitative analysis of AfricaNLP contributions over 2005–2025, while Hussen et al. (2025) documented that LLMs cover fewer than 2% of African languages. Inuwa-Dutse (2025) reported that only 25.1% of Nigerian NLP studies produce novel resources, and Muhammad et al. (2025) survey Hausa-specific resources and the challenges that remain. Together, these surveys motivate resource-efficient methods that can extend the reach of existing Afri-centric models without the computational cost of full fine-tuning.

2.2 Parameter-Efficient Fine-Tuning

Parameter-efficient fine-tuning (PEFT) methods enable adaptation of large pretrained language models with minimal trainable parameters. These methods have proliferated since Houlsby et al. (2019)'s original bottleneck adapters. Prominent extensions include LoRA (Hu et al., 2022), prefix tuning (Li and Liang, 2021), prompt tuning (Lester et al., 2021), $(IA)^3$ (Liu et al., 2022), and AdapterFusion (Pfeiffer et al., 2021). Recent comparisons (Ding et al., 2023) have benchmarked these methods on English tasks, but systematic evaluation for African languages has remained absent. We evaluate four representative approaches:

Full Fine-Tuning: The traditional approach updates all model parameters. While maximally expressive, full fine-tuning requires storing optimizer states for billions of parameters, presenting both memory and storage challenges, particularly for resource-constrained research environments.

LoRA (Hu et al., 2022): Low-Rank Adaptation introduces small rank-decomposition matrices into the attention layers' query and value projections.

The pretrained weights remain frozen while only the rank-r matrices are trained. The intrinsic dimensionality hypothesis (Aghajanyan et al., 2021) provides theoretical motivation: task adaptation often lies in a low-dimensional subspace.

Prefix Tuning (Li and Liang, 2021): Prefix Tuning prepends learnable continuous vectors ("virtual tokens") to the keys and values at every transformer layer. The model parameters remain frozen while only the prefix vectors are optimized.

(IA)^3 (Liu et al., 2022): Infused Adapter by Inhibiting and Amplifying Inner Activations introduces learned scaling vectors that modulate keys, values, and feedforward activations. Among PEFT methods, (IA)^3 achieves the most extreme parameter efficiency, training fewer than 0.1% of model parameters in our setup.

2.3 Adapter Evaluations in Multilingual Contexts

Cross-lingual adapter applications have been explored for multilingual NLP. The MAD-X framework (Pfeiffer et al., 2020) composes language and task adapters for cross-lingual transfer. Hedderich et al. (2020) examine transfer learning for African languages. However, no prior work has compared adapter architectures across multiple Hausa-specific tasks. Our work fills this gap by providing the first comprehensive PEFT benchmark for Hausa language.

3 Datasets

We evaluate on three diverse HausaNLP tasks selected to span both sentence-level and token-level classification:

Sentiment Analysis (AfriSenti). We use the Hausa subset of AfriSenti (Muhammad et al., 2023), containing 22,152 tweets labeled as positive, negative, or neutral. We use the official splits: 14,172 train, 2,677 development, and 5,303 test instances. We report weighted F1 to account for label imbalance.

Named Entity Recognition (MasakhaNER 2.0). We use the Hausa subset of MasakhaNER 2.0 (Adelani et al., 2022), containing 8,165 sentences with span-level annotations for four entity types: Person (PER), Organization (ORG), Location (LOC), and Date (DATE). We use the official splits: 5,716 train, 816 development, and 1,633 test instances. We report entity-level F1 using `seqeval`.

Authenticity Detection (Sani et al.). We use the dataset from Sani et al. (2025), containing 5,172 Hausa articles labeled as human-written or machine-generated. We use stratified 80/10/10 splits (4,137 train, 517 dev, 518 test) with random seed 42 for reproducibility. We report macro-F1 to weight both classes equally.

4 Experimental Setup

4.1 Base Models

We evaluate four pretrained encoder models:

- **AfroXLMR-large** (560M parameters): The largest Afri-centric model, adapted from XLM-RoBERTa-large via MAFT (Alabi et al., 2022).
- **AfroXLMR-base** (270M parameters): The base-size variant of AfroXLMR.
- **AfriBERTa-large** (270M parameters): Pretrained on 11 African languages from scratch (Ogueji et al., 2021).
- **XLM-RoBERTa-base** (270M parameters): Multilingual baseline without explicit African language adaptation (Conneau et al., 2020).

4.2 Adapter Configurations

We use the HuggingFace PEFT library (Mangrulkar et al., 2022) with the following configurations:

- **LoRA:** rank $r = 8$, $\alpha = 16$, dropout 0.1, applied to query and value projections.
- **Prefix Tuning:** 50 virtual tokens.
- **(IA)^3:** Applied to key, value, and feedforward output projections.

4.3 Training Details

All experiments use the AdamW optimizer with learning rate 1×10^{-4} for adapters and 2×10^{-5} for full fine-tuning, weight decay 0.01, and 10% linear warmup. We train for up to 5 epochs with early stopping (patience 2 on validation F1), batch size 16, and mixed-precision (fp16) on a single NVIDIA RTX 3090 GPU (24 GB). For adapter methods, we report mean and standard deviation across three random seeds (42, 123, 456); for full fine-tuning, we use seed 42 only due to computational cost. Experiments use `transformers==4.46.3`, `datasets==3.6.0`, `peft==0.13.2`, and `huggingface_hub==0.26.5`. We note that

the AfriSenti and MasakhaNER 2.0 datasets rely on legacy loading scripts that have been deprecated in `datasets≥4.0`; reproducing our setup requires `datasets<4.0` with `trust_remote_code=True`.

4.4 Evaluation Metrics

We report task-specific F1 scores (weighted F1 for sentiment, entity-level F1 for NER, macro-F1 for authenticity), trainable parameter percentage, training time (wall-clock seconds), and peak GPU memory consumption.

5 Results

5.1 Main Performance Comparison

RQ1: Which adapter method best preserves full fine-tuning performance?

LoRA preserves full fine-tuning performance best, recovering 93.3–96.6% of full FT F1 with AfroXLMR-large while training only 0.14–0.33% of parameters. It consistently outperforms Prefix Tuning and (IA)³ across all three tasks and trains 2.36× faster than full fine-tuning on average.

Table 1 presents F1 scores across all configurations. Three patterns emerge clearly:

AfroXLMR-large dominates across all tasks. With both full fine-tuning and LoRA, AfroXLMR-large achieves the highest F1 across all three tasks (sentiment: 0.802/0.763, NER: 0.865/0.807, authenticity: 0.988/0.955). This consistency suggests that for HausaNLP, the larger Afri-centric model provides robust gains regardless of fine-tuning strategy.

LoRA recovers 93–97% of full FT performance. Table 2 quantifies the performance recovery for the best base model (AfroXLMR-large). LoRA achieves 95.1% of full fine-tuning F1 on sentiment, 93.3% on NER, and 96.6% on authenticity—using only 0.14–0.33% of trainable parameters. Figure 1 visualizes this recovery.

Pareto efficiency. Figure 2 shows the Pareto frontier of F1 vs trainable parameters. LoRA configurations consistently dominate the upper-left region (high F1, low parameters), with AfroXLMR-large + LoRA forming the empirical Pareto frontier across all tasks.

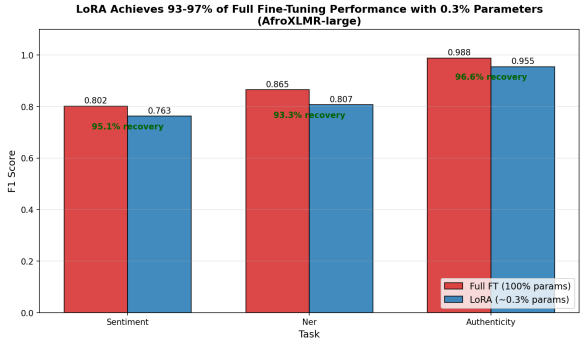

Figure 1: LoRA performance recovery vs full fine-tuning for AfroXLMR-large. LoRA achieves 93–97% of full FT performance with ∼0.3% of parameters across all three tasks.

5.2 The Adapter Amplification Effect

RQ2: How do Afri-centric models compare to multilingual baselines under PEFT?

Afri-centric models outperform the multilingual XLM-R baseline across all tasks, and crucially this advantage amplifies under PEFT: AfroXLMR-large's edge over XLM-R grows from +3.9 F1 (full FT) to +12.5 F1 (LoRA), a 3.2× amplification. This indicates adapters preserve language-specific knowledge from pretraining that full fine-tuning can partially overwrite.

Our most striking finding is that the advantage of Afri-centric pretraining over multilingual baselines is substantially *amplified* under parameter-efficient fine-tuning. Table 3 quantifies this effect. For full fine-tuning, AfroXLMR-large outperforms XLM-R-base by an average of 3.9 F1 points. With LoRA, this advantage grows to 12.5 F1 points and a 3.2× amplification. The effect is consistent across all three tasks: sentiment (2.7×), NER (3.7×), and authenticity (3.5×).

We hypothesize that this amplification arises because adapters preserve the language-specific knowledge encoded in pretrained weights. Full fine-tuning can partially recover from suboptimal pretraining by overwriting weights, narrowing the gap between Afri-centric and multilingual models. In contrast, adapters cannot recover information that was not learned during pretraining: if a model lacks Hausa-specific knowledge (as XLM-R does), parameter-efficient methods cannot synthesize it.

This finding has substantial practical implications: *practitioners working with low-resource lan-*

Task	Base Model	Adapter Method			
		Full FT	LoRA	Prefix Tuning	$(IA)^3$
Sentiment	AfroXLMR-large	**80.2**	**76.3**†	62.1	65.0
	AfroXLMR-base	77.4	69.7	50.0	63.6
	AfriBERTa-large	76.9	71.4	68.9	62.6
	XLM-R-base	74.2	59.9	52.9	55.7
NER	AfroXLMR-large	**86.5**	**80.7**†	50.9	47.1
	AfroXLMR-base	85.6	71.4	24.3	46.9
	AfriBERTa-large	84.2	73.8	55.1	54.9
	XLM-R-base	83.2	68.5	19.7	42.4
Authenticity	AfroXLMR-large	**98.8**	**95.5**†	78.7	94.0
	AfroXLMR-base	96.9	89.5	82.1	90.2
	AfriBERTa-large	98.1	94.1	91.9	93.2
	XLM-R-base	96.3	86.7	80.9	83.1
Mean	Trainable params (%)	100.00	0.32	0.61	0.21
	Training time (s)	232.1	**98.5**	96.9	120.2

Table 1: F1 scores (%) across base models and adapter architectures for three Hausa NLP tasks. For adapter methods, results are averaged over 3 seeds (42, 123, 456). Full FT results are from a single seed. **Bold** indicates the best result per column. † marks LoRA recovery > 90% of full fine-tuning performance.

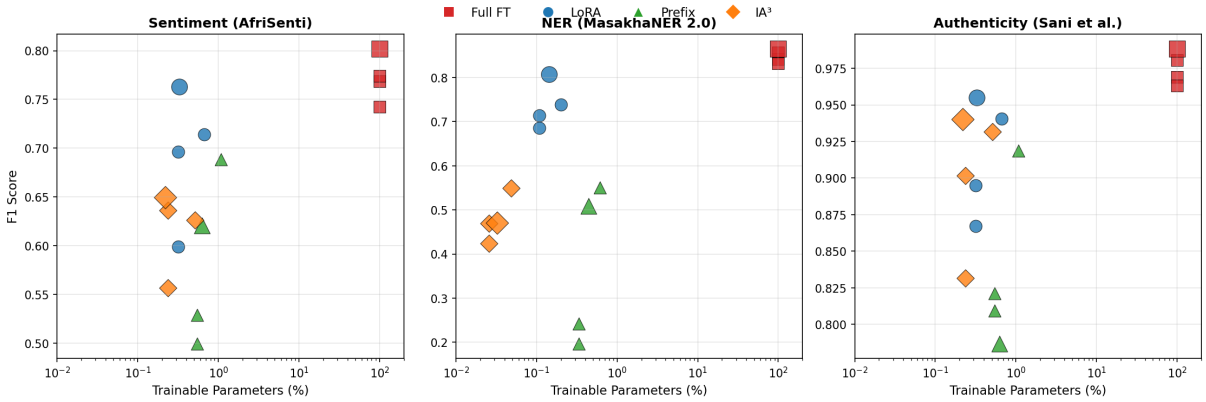

Figure 2: Pareto frontier of F1 score vs trainable parameter percentage. LoRA (blue circles) consistently dominates other parameter-efficient methods, achieving F1 close to full fine-tuning (red squares) with ∼1000× fewer parameters.

Task	Full FT	LoRA	Recovery
Sentiment	0.802	0.763	**95.1%**
NER	0.865	0.807	**93.3%**
Authenticity	0.988	0.955	**96.6%**
Mean	**0.885**	**0.842**	**95.0%**

Table 2: LoRA performance recovery vs full fine-tuning, demonstrating the practical efficacy of parameter-efficient fine-tuning for Hausa NLP. Results shown for the best configuration (AfroXLMR-large).

Task	Full FT ΔF1	LoRA ΔF1	Amplification
Sentiment	+6.0	+16.4	2.7×
NER	+3.3	+12.2	3.7×
Authenticity	+2.5	+8.8	3.5×
Mean	**+3.9**	**+12.5**	**3.2×**

Table 3: F1 score advantage of AfroXLMR-large over XLM-R-base baseline. The Afri-centric advantage is substantially *amplified* when using LoRA, suggesting parameter-efficient methods preserve language-specific knowledge.

guages should prioritize language-specialized base models when using PEFT methods. The cost of using a multilingual baseline is significantly higher under PEFT than under full fine-tuning.

5.3 Prefix Tuning Failures on Sequence Labeling

Figure 3 shows a stark pattern: prefix tuning achieves dramatically lower F1 on NER than other

Method	Mean F1	Params (%)	Time (s)	Speedup
Full FT	0.859	100.000	232.1	1.00×
LoRA	**0.755**	**0.318**	**98.5**	**2.36×**
Prefix	0.582	0.606	96.9	2.40×
$(IA)^3$	0.657	0.211	120.2	1.93×

Table 4: Adapter efficiency comparison averaged across all (model, task) configurations. **LoRA** provides the best balance of performance and efficiency.

Adapter	Mean σ	Max σ	Stability
$(IA)^3$	**0.010**	0.022	Most stable
LoRA	0.019	0.074	Stable
Prefix Tuning	0.030	0.074	Variable

Table 5: Standard deviation across 3 seeds (42, 123, 456) per (task, model) configuration. Lower values indicate more stable, reproducible results. LoRA exhibits low variance across all configurations.

methods. AfroXLMR-base + prefix achieves F1 = 0.243, and XLM-R-base + prefix achieves F1 = 0.197 when compared to LoRA results of 0.714 and 0.685 for the same models.

RQ3: Are there task-specific failure modes?

Yes. Prefix Tuning fails on token-level NER (F1 as low as 0.197 vs. 0.685 for LoRA on the same model), because its prepended virtual tokens disrupt the token-label alignment that sequence labeling requires. The same method performs adequately on sentence-level tasks (sentiment, authenticity), confirming the failure is specific to token-level prediction.

Prefix tuning prepends learned virtual tokens to the input sequence at every transformer layer. For sentence-level tasks (sentiment, authenticity), this provides a global context modification. However, for token-level tasks (NER), the prepended tokens disrupt the alignment between input tokens and their predicted labels.

Variance analysis. Prefix tuning also exhibited substantially higher variance across seeds than other adapters. For NER with AfroXLMR-large, $\sigma = 0.060$ compared to $\sigma = 0.006$ for LoRA on the same configuration. This instability indicates that prefix tuning is sensitive to initialization for sequence labeling, providing further evidence against its use for such tasks in low-resource settings.

5.4 Stability and Reproducibility

RQ4: How stable are adapter results across random seeds?

LoRA is highly stable, with mean standard deviation of 0.019 across three seeds, while Prefix Tuning is markedly less stable (mean $\sigma = 0.030$, reaching 0.060 on NER). $(IA)^3$ shows the lowest variance ($\sigma = 0.010$) but at substantially lower performance, making LoRA the best overall balance of stability and accuracy.

Table 5 summarizes seed variance across adapter methods. $(IA)^3$ shows the lowest variance (mean $\sigma = 0.010$), but at the cost of substantially lower performance. LoRA provides the best balance: stable ($\sigma = 0.019$) while achieving high performance.

6 Discussion

6.1 Recommendations for HausaNLP Practitioners

Based on our findings, we provide the following recommendations:

- **For best performance:** Use AfroXLMR-large with full fine-tuning, accepting the higher computational cost.
- **For best efficiency-performance trade-off:** Use AfroXLMR-large with LoRA (rank=8). This achieves 93–97% of full FT performance with <0.5% of parameters.
- **Avoid:** Prefix tuning for sequence labeling tasks. The combination of poor performance and high variance makes it unsuitable for token-level applications.

6.2 Implications for African Language NLP

Our findings have broader implications for African language NLP research, where computational constraints are common. The 99.7% parameter reduction achieved by LoRA enables adapter checkpoints of <1MB, facilitating distribution and deployment in resource-constrained environments. The 2.36× training speedup translates to substantial reductions in compute time and energy use.

The amplification effect we document shows that investments in language-specific pretraining, such as the AfroXLMR (Alabi et al., 2022) and AfriBERTa (Ogueji et al., 2021) efforts, provide outsized

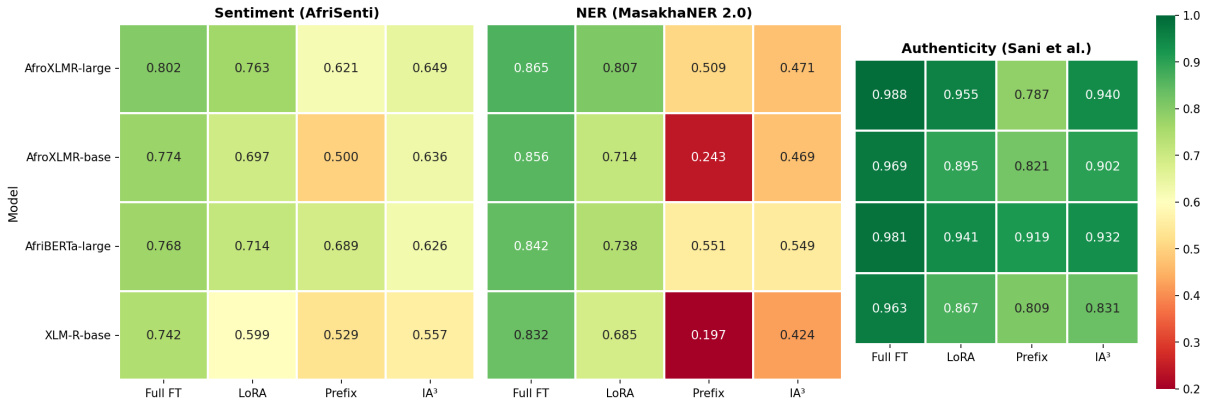

Figure 3: F1 heatmap across all (model, adapter, task) configurations. Prefix tuning shows pronounced failure on NER for AfroXLMR-base (0.243) and XLM-R-base (0.197), in stark contrast to its sentence-level task performance.

benefits when downstream practitioners use PEFT methods. This argues for continued community investment in Afri-centric pretraining.

6.3 Connection to Downstream Applications

The strong performance on text authenticity detection (LoRA F1 = 0.955) indicates that adapter methods are practical for content moderation and misinformation applications in Hausa, addressing a critical real-world need given the rapid spread of LLM-generated misinformation. Our findings inform ongoing work on extending these methods to fine-grained propaganda technique detection, building on existing binary fake news detection efforts (Imam et al., 2022) toward more nuanced analyses.

7 Conclusion

We presented the systematic comparison of parameter-efficient fine-tuning methods for HausaNLP, evaluating four adapter strategies across four base models on three Hausa downstream tasks. Our 120 experiments yield three principal findings: (1) AfroXLMR-large with LoRA achieves 93–97% of full fine-tuning performance with $<0.5\%$ of parameters; (2) Afri-centric models' advantage over multilingual baselines is amplified $\sim 3\times$ under PEFT, suggesting adapters preserve language-specific knowledge; and (3) prefix tuning fails dramatically on token-level NER tasks. Our findings inform practitioners working with low-resource languages and motivate continued investment in language-specialized pretraining.

Limitations

Task scope. We evaluated three HausaNLP tasks. While these span sentence-level and token-level classification, generalization to generation tasks (machine translation, summarization, question answering) requires further investigation. Our findings on token-level prefix tuning failure may extend to other sequence labeling tasks (e.g., POS tagging) but this remains untested. While our focus on Hausa is motivated by its scale and importance (200M+ speakers), generalization of findings to other African languages particularly those from non-Afro-Asiatic families remains an empirical question. Future work may extend this analysis to Yorùbá, Igbo, Swahili, Amharic, and others.

We use standard hyperparameter configurations from the literature (LoRA rank=8, prefix length=50). Broader hyperparameter sweeps could reveal additional patterns, particularly for prefix tuning where instability suggests sensitivity to initialization. Our efficiency analysis is based on single-GPU (RTX 3090) training. Multi-GPU and distributed training scenarios are not directly evaluated. Memory-bound versus compute-bound analyses may shift under different hardware.

Our parameter efficiency analysis focuses on fine-tuning costs. The substantial costs of pretraining the base models are not included, as we use publicly available checkpoints. The amplification effect we document depends on the existence of these models. We use three random seeds for adapter methods and one for full fine-tuning. While this captures key variance patterns, formal statistical significance testing was not conducted due to computational cost. Reported recovery percentages should be interpreted as point estimates.

Ethics Statement

This work uses publicly available datasets (AfriSenti, MasakhaNER 2.0, Sani et al. 2025 corpus) collected and annotated by the original authors following ethical guidelines. We do not collect any new data or interact with human subjects. The computational efficiency benefits demonstrated here support more equitable access to NLP research capabilities, particularly benefiting researchers in resource-constrained environments common to African research institutions. We release all code and trained models openly to maximize community benefit.

Acknowledgments

We thank the AfricaNLP and HausaNLP communities for foundational resources and ongoing collaboration. We thank Sani et al. (2025) for making the Hausa authenticity corpus publicly available.

References

Idris Abdulmumin, Satya Ranjan Dash, Musa Abdullahi Dawud, Shantipriya Parida, Shamsuddeen Hassan Muhammad, Ibrahim Said Ahmad, Subhadarshi Panda, Ondřej Bojar, Bashir Shehu Galadanci, and Bello Shehu Bello. 2022. Hausa visual genome: A dataset for multi-modal English to Hausa machine translation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6471–6479.

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, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, and 42 others. 2021. MasakhaNER: Named entity recognition for African languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131.

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, Andiswa Bukula, Rooweither Mabuya, Bonaventure F. P. Dossou, Blessing Sibanda, Happy Buzaaba, Jonathan Mukiibi, Godson Kalipe, Derguene Mbaye, and 28 others. 2022. MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4488–4508.

Armen Aghajanyan, Luke Zettlemoyer, and Sonal Gupta. 2021. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, pages 7319–7328.

Jesujoba O Alabi, David I 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.

Tadesse Destaw Belay, Kedir Yassin Hussen, Sukairaj Hafiz Imam, Ibrahim Said Ahmad, Isa Inuwa-Dutse, Abrham Belete Haile, Grigori Sidorov, Iqra Ameer, Idris Abdulmumin, Tajuddeen Gwadabe, Vukosi Marivate, Seid Muhie Yimam, and Shamsuddeen Hassan Muhammad. 2025. The rise of AfricaNLP: Contributions, contributors, and community impact (2005-2025). *arXiv preprint arXiv:2509.25477*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451.

Ning Ding, Yujia Qin, Guang Yang, Fuchao Wei, Yang Zonghan, Su Yusheng, Hu Shengding, Chen Yulin, Chan Chi-Min, Chen Weize, Yi Jing, Zhao Weilin, Wang Xiaozhi, Liu Zhiyuan, Zheng Hai-Tao, Chen Jianfei, Liu Yang, Tang Jie, Li Juanzi, and Sun Maosong. 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 5(3):220–235.

Bonaventure FP Dossou, Atnafu Lambebo Tonja, Oreen Yousuf, Salomey Osei, Abigail Oppong, Iyanuoluwa Shode, Oluwabusayo Olufunke Awoyomi, and Chris Chinenye Emezue. 2022. AfroLM: A self-active learning-based multilingual pretrained language model for 23 African languages. In *Proceedings of The Third Workshop on Simple and Efficient Natural Language Processing (SustaiNLP)*, pages 52–64.

Michael A Hedderich, David Adelani, Dawei Zhu, Jesujoba Alabi, Udia Markus, and Dietrich Klakow. 2020. Transfer learning and distant supervision for multilingual transformer models: A study on African languages. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2580–2591.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In *International Conference on Machine Learning*, pages 2790–2799.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Kedir Yassin Hussen, Walelign Tewabe Sewunetie, Abinew Ali Ayele, Sukairaj Hafiz Imam, Eyob Nigussie Alemu, Shamsuddeen Hassan Muhammad, and Seid Muhie Yimam. 2025. The state of large language models for African languages: Progress and challenges. *arXiv preprint arXiv:2506.02280*.

Abba Abdullahi Ibrahim, Hadiza Umar Ali, Ismail Zahraddeen Yakubu, and Ibrahim A. Lawal. 2024. Fake news detection in Hausa language using transfer learning method. *International Journal of Innovative Science and Research Technology*, 9(10).

Sukairaj Hafiz Imam, Abubakar Ahmad Musa, and Anil Choudhary. 2022. The first corpus for detecting fake news in Hausa language. In *Emerging Technologies for Computing, Communication and Smart Cities*, volume 875 of *Lecture Notes in Electrical Engineering*, pages 647–657. Springer Singapore.

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

Isa Inuwa-Dutse. 2025. NaijaNLP: A survey of Nigerian low-resource languages. *arXiv preprint arXiv:2502.19784*.

Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3045–3059.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, pages 4582–4597.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. 2022. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. In *Advances in Neural Information Processing Systems*.

Sourab Mangrulkar, Sylvain Gugger, Lysandre Debut, and Younes Belkada. 2022. PEFT: State-of-the-art parameter-efficient fine-tuning methods. https://github.com/huggingface/peft.

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, and 10 others. 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.

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, Alípio Jorge, and Pavel Brazdil. 2022. NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 590–602.

Shamsuddeen Hassan Muhammad, Ibrahim Said Ahmad, Idris Abdulmumin, Falalu Ibrahim Lawan, Babangida Sani, Sukairaj Hafiz Imam, Yusuf Aliyu, Sani Abdullahi Sani, Ali Usman Umar, Tajuddeen Gwadabe, Kenneth Church, and Vukosi Marivate. 2025. HausaNLP: Current status, challenges and future directions for hausa natural language processing. In *Proceedings of the Fifth Workshop on New Frontiers in Summarization (NewSumm), EMNLP 2025*. ArXiv:2505.14311.

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.

Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. 2021. AdapterFusion: Non-destructive task composition for transfer learning. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pages 487–503.

Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020. 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.

Babangida Sani, Aakansha Soy, Sukairaj Hafiz Imam, Ahmad Mustapha, Lukman Jibril Aliyu, Idris Abdulmumin, Ibrahim Said Ahmad, and Shamsuddeen Hassan Muhammad. 2025. Who wrote this? identifying machine vs human-generated text in Hausa. *arXiv preprint arXiv:2503.13101*.