

Towards Inclusive NLP: Benchmarking and Mitigating Bias in Named Entity Recognition for African Languages and Entities.

Named Entity Recognition (NER) is a foundational task in Natural Language Processing (NLP), underpinning applications such as information retrieval, knowledge extraction, and conversational systems. However, existing NER systems are disproportionately trained and evaluated on high resource languages and Western centric datasets, resulting in limited performance and cultural misrepresentation when applied to African languages and entities. This gap exacerbates inequities in global AI development, where African users encounter frequent model errors, erasures of local names and places, and systematic underrepresentation in NLP research.

In this work, we present the first large scale benchmark dataset for African NER across multiple languages and domains, capturing person names, locations, organizations, and culturally specific entities. The dataset spans both low resource and moderately resourced African languages, with attention to code-switching and dialectal variation. Beyond raw annotation, we propose a new evaluation framework, the African Entity Recognition Bias Index (AERBI), to systematically measure cultural and linguistic bias in NER systems. This framework quantifies disparities in recognition performance across language groups, entity types, and sociocultural contexts, enabling reproducible fairness audits.

Methodologically, we explore data efficient transfer learning and semi-supervised annotation strategies tailored to African contexts. In particular, we investigate leveraging phonetic similarity of African names and cross lingual embeddings to improve generalization in low-resource settings. We benchmark widely used NER models (BiLSTM-CRF, mBERT, XLM-R, AfroXLM-R) against our dataset and framework, revealing both severe performance gaps and patterns of bias. We further propose bias mitigation approaches, such as adaptive reweighting of training samples and culturally informed data augmentation, and demonstrate significant improvements in recognition accuracy and fairness metrics.

Our contributions are threefold:

1. Dataset Contribution – A large scale, culturally grounded benchmark for African NER.
2. Bias Analysis Framework – A reusable index (AERBI) to quantify and compare NER fairness across African languages and entities.
3. Methodological Innovation – Transfer learning and bias mitigation techniques that improve NER in low resource, culturally diverse settings.

We conclude that current NER systems are not only underperforming but systematically biased against African entities, and that targeted dataset construction, fairness frameworks, and lightweight adaptation methods are crucial steps towards inclusive NLP. By introducing this benchmark and methodology, we aim to catalyze broader research on African languages, bridge representation gaps in global NLP, and lay the foundation for equitable language technologies that serve diverse communities.