

# Overview

## Purpose of the Voice Playbook

This playbook provides **practical, ethical, and scalable guidance** for collecting **voice datasets for African languages**, particularly **low-resource languages with limited digital presence**.

The playbook enables:

- Community-led voice dataset creation
- Standardized speech dataset methodologies
- Ethical and consent-based recording
- Scalable workflows for NLP training
- Reusable processes for future language initiatives

The playbook supports **Automatic Speech Recognition (ASR)**, **Text-to-Speech (TTS)**, and **multimodal language models**.

African languages remain severely underrepresented in speech datasets, limiting their participation in modern AI systems. Many African languages lack large-scale audio datasets necessary for training speech technologies.

“ **This playbook lowers barriers by enabling grassroots communities, universities, NGOs, and language activists to collect high-quality speech data.** ”

---

Revision #2

Created 2026-03-05 07:04:03 UTC by Angela

Updated 2026-03-05 07:51:35 UTC by Angela