

Inclusive Education through AI: A Framework for Swahili - English Audio Access in Tanzanian Libraries

Access to education for blind learners in Tanzania remains severely constrained. Tanzania carries one of the highest burdens of vision loss in East Africa, with an estimated 8.2 million people experiencing some form of vision loss and around 290,000 who are blind. According to the 2022 census, 3.6 million people about 8% of the population reported a disability. Persons with disabilities face disproportionate exclusion: illiteracy among them stands at 48%, nearly double the national average, and only 3.1% are in paid employment. These inequalities extend to libraries and educational resources. Globally, fewer than 5% of published books are accessible to print disabled readers, and fewer than 20% of websites meet accessibility standard. In Tanzania, libraries are poorly equipped, with studies at the University of Dar es Salaam documenting a scarcity of assistive technologies, limited training for staff, and ICT systems that are largely incompatible with screen readers. Assistive technologies remain prohibitively costly, reaching fewer than 15% of those who need them. Government and NGO efforts, such as the distribution of 72 Orbit Reader braille devices, have been limited in scope and geographic coverage. Crucially, audio reading programs in Swahili and English are virtually absent.

This paper proposes a framework for an AI driven, web based audio reading system to address this gap. The framework focus on four sections. First, open source computer vision will be applied for scanning print materials. CNN based optical character recognition (OCR) has surpassed traditional OCR in accuracy to convert text into machine readable form, including Swahili characters. Second, large language models (LLMs) will run text to speech synthesis. Recent work shows that curated crowdsourced datasets enable neural TTS in low resource languages; the Kiswahili dataset built on Mozilla Common Voice contains over 6,300 validated utterances amounting to 15.1 hours of speech. Multilingual LLMs with built-in TTS capabilities will be finetuned to generate natural audio in Swahili and English, ensuring inclusivity. Third, a web architecture will integrate these modules. Libraries will upload scanned pages or books, with the back end handling CNN-based OCR, LLM processing, and TTS output. The expected impact focuses on leveraging open-source datasets and models, the system addresses both affordability and linguistic accessibility. This work directly responds to documented policy gaps in inclusive education and reduces barriers to acquiring assistive technologies in Tanzania.