# Problem statement.

The advancement of technology has paved a way to Artificial Intelligence (AI) being preferred as technology that solves a number of human problems(). The subset of Artificial Intelligence which is Machine Learning (ML) is the leading of this advancement. This advancement has led to the development of tools that assist humans with different tasks such as transcribing and translation, with ease, accuracy and fast execution().

Focusing on transcribing, ML has enabled people in need of this service to finally be able to acquire the resources and services at a much cheaper price and with faster execution and hence a lot of time is saved. Transcription tools are available in multiple languages, even Swahili, but with the Swahili language there is a low accuracy of the AI model. AI This is may because of the fact that the data for the manipulation of the model did not originate from the Swahili people perse. The existing models fail even to capture simple Swahili words like “ujenzi”, “maisha”, “uzoefu” and lots of more other words that can frequently be heard from the swahili people especially in the swahili speaking regions in Tanzania like Dar Es Salaam.

With the Swahili transcription tool, we will have the first transcription tool built by Tanzanians and with the implementation of speech recognition optimization a normal Swahili accent will be used as a benchmark for the rest of accents to fall through, thus increasing the accuracy of this tool far greater than the ones built by Swahili scholars abroad who are not well acquainted with the Swahili language than the indigenous people of Tanzania where Swahili is largely used compared to other parts of the world.

# Objectives.

## Main objectives;

- To develop a swahili transcription tool using machine learning.

### Specific Objectives;

1. To create a swahili words dataset for transcription.
2. To create a Swahili transcription model using natural language processing algorithm
3. To Implement the model using web-based system