INITIATIVE ON
Digital Innovation

CGIAR

TECHNICAL REPORT

Speech Recognition, Machine Translation, and Corpus Analysis for Identifying Farmer Demands and Targeting Digital Extension

Eliot Jones-Garcia ✉
International Maize and Wheat Improvement Center (CIMMYT) and University of Nottingham

Abstract The increasing capabilities of Artificial Intelligence-augmented data analytics present significant opportunities for agricultural extension organizations operating in the Global South. In this project, we supported Farm Radio International (FRI) in investigating the possibility of automating the process of translating and analyzing farmers' voice message data. This report reviews several approaches to overcoming technical constraints and then presents a cutting-edge approach that utilizes innovations in unsupervised learning to deliver highly accurate speech recognition and machine translation in a diverse set of languages.

Date: **22 Nov 2022** // Work Package: **5. Platforms & Services** // Partners: **CIMMYT, Farm Radio Int'l**

CIMMYT MR
International Maize and Wheat Improvement Center

FARM RADIO INTERNATIONAL

Table of Contents

Speech Recognition, Machine Translation, and Corpus Analysis for Identifying Farmer Demands and Targeting Digital Extension

Executive Summary

The increasing capabilities of information communication technologies and big data analytics present significant opportunities for agricultural extension organizations operating in the Global South. Existing trajectories, however, are toward large, highly productive farms in the Global North. There is a need for work in AI to expand the application of tools for smallholder farmers in the South while remaining conscious of the implications of the digital divide, to which many have limited access.

Farm Radio International (FRI), a Canadian NGO, has hosted over 700 talk-radio shows in 40 countries across sub-Saharan Africa. Not only do they broadcast discussions between experts on local problems, but they also receive and answer questions from farmers about their practices and demands.

Unfortunately, a significant portion of these responses goes unanswered. It is the intention of this report to investigate the possibility of automating the process of translating and analyzing this data, to propose a tool to seamlessly collect rich information from a broad pool of farmers, and subsequently design better-suited and longer-lasting interventions.

Many of Africa's 2000 languages, however, remain what is called 'low-resource' or those with limited training data upon which to build and train AI models. In this report, we review several approaches to overcoming these constraints, including transfer learning and crowdsourcing data. We then present a cutting-edge approach that utilizes innovations in unsupervised learning to deliver highly accurate speech recognition and machine translation in a diverse set of languages.

We then evidence this through a series of experiments using real-world examples of FRI data in Swahili and Hausa and include a corpus linguistics method for gaining insight from this data at scale.

The Common Voice project employs crowdsourcing for both data collection and data validation, becoming one of the largest corpora in the public domain for SR, both in terms of the number of hours and the number of languages. A team of researchers from some of the largest NLP groups, including Google AI and Hugging Face-built XLS-R. It is a tool based on 436K hours of speech in 128 different languages, using data from across the largest open-source repositories, including Common Voice. These data are 'pre-trained,' allowing the model to recognize patterns in the unlabeled data, then they are fine-tuned for a specific task using a limited set of labeled data. To the best of their knowledge, they state, 'this is the largest effort to date in making speech technology accessible for many more languages using publicly available data, with the largest contribution being toward low and medium source languages.

The report surveys a series of different SR models, including Google as a baseline, with the most accurate results being delivered by the XLS-R. This is followed by an analysis of the keywords and their context within an existing FRI corpus in answer to the question; if you had more power to change things, what would you do to make life better for farming families?

Results indicate that farmers desire significant support and assistance from government ministers to improve their ability to plant quality crops and secure financial loans. This is either predominantly for better fertilizers or seeds. Other themes include supporting young people and ensuring regular markets for produce.

A number of concerns and biases are considered for bringing this project to fruition, and questions are asked on how this can best benefit each party. The project must remain reflexive to a range of information sources while inclusive of different voices. Objectives to be considered going forward include how this can best be rolled out for real-world use, how more languages can be included, and the potential for this to develop into a scientific research delivery.

Introduction

The increasing capabilities of information communication technologies and big data analytics present significant opportunities for agricultural extension organizations operating in the Global South [1]. Their goal is to improve sustainability, boost economic growth, reduce food insecurity, inequality, and poverty, and increase resilience in poorer communities with minimal cost to biodiversity and the planet [2,3]. These have historically been constrained, however, by the practical challenges of accurately gaining insight into farmer demands and providing effective and useful knowledge accordingly [4]. This is of particular concern as a new wave of young, educated farmers enter the industry, who may be motivated by sustainability but lack practical skills. Significant work is being conducted in Artificial intelligence (AI) technologies to support farmers whilst mitigating their challenges [5–7]. This includes machine learning for the prediction of extreme weather events, leading to early warning systems [8]. Precision agriculture is allowing farmers to process immense amounts of data to deliver highly granular treatments to crops, reducing costs and environmental damage [9]. Existing trajectories are, however, toward large, highly productive farms in the Global North [10]. Lowder et al. [11] established that 95% of farmers are smallholders operating on less than 2 hectares of land, 70% of which are based in Sub-Saharan Africa (SSA), Latin America, and South Asia. There is a need for work in AI to expand the application of tools for smallholder farmers in the South whilst remaining conscious of the implications of the digital divide, where many have limited access [12].

According to Heldert [7], of particular concern is the need to build diverse literacies and multiple languages. Audio and visual tools may help in providing information to users with different levels of literacy, limited time, resources, or lack of travel

opportunities to gather information elsewhere. It is suggested that there is only one agricultural extension agent per 3000 clients in SSA [13]. Radio operates as a vital lifeline for Africa's vast rural populations, providing a variety of services, including information sharing, discussion, advocacy, and capacity building. It is also a space for public discourse, garnering significant information for development agencies regarding the perception of their interventions [14]. The advancement of AI for smallholder farmers was a significant theme in the UN's recent report on data governance for food and nutrition [15], and several scholars and development professionals have come to understand analog radio technologies as the vehicle for change [14,16,17].

Farm Radio International (FRI), a Canadian NGO, has hosted over 700 talk-radio shows in 40 countries across sub-Saharan Africa. They work on the principle that 'Farmers have a lot to say', and 'as nations, organizations, and individuals, we all must commit to listening and taking action together' [18]. As such, not only do they broadcast discussions between experts on local problems, but they also receive and answer questions from farmers about their practices and demands. In a recent project, FRI partnered with six radio stations in Burkina Faso, Ghana, Tanzania, and Uganda, to ask small-scale farmers, vendors, processors, marketers, and others how the food system should be changed to meet their needs and the needs of their communities. Nearly 12,000 responses were recorded [18]. Unfortunately, a significant portion of these responses goes unanswered, as without the staff to manually translate each phone call, the information remains inaccessible. This presents a vast and rich dataset, speaking directly to the needs and desires of smallholders. It is the intention of this report to investigate the possibility of automating the process of translating and analyzing this data and to propose a tool

to seamlessly collect rich information from a broad pool of farmers, leading to better-suited and longer-lasting interventions.

Africa alone has over 2000 languages [19], with 75 having at least 1 million speakers [20]. In the development of natural language processing (NLP) for African languages, the AI domain focused on issues of speech and translation, but this has been slow predominantly due to a lack of training data from which the machine can learn. Other problems include limited funding, challenges of discovery and access to existing tools and data, absence of benchmarks, and poor reproducibility of ongoing studies [19]. In tackling these, developing further datasets and working in collaboration with local language experts may speak to the diverse needs of multiple literacies, languages, and cultures [7]. Significant opportunities have arisen in recent years due to the development of unsupervised pretraining techniques [21].

In the following sections, this report will detail ongoing research in speech recognition, machine translation, and computational linguistics before describing and testing an approach to processing and analyzing farmer telephone calls at scale. Swahili and Hausa were selected as test languages, with details included on how to expand this into other under-resourced African languages. It is hoped that this report will serve as a proof of concept for the implementation of the tool by FRI and the potential for a larger research project on the needs of farmers across Africa.

Literature Review

Languages under examination; Swahili and Hausa

The two languages selected for this proof of concept were Swahili and Hausa. These represent the largest languages in the East and West of Africa, respectively, and are

sufficiently large that they can be tested using the best and most reliable data and models.

Swahili has 100 to 150 million speakers [22]. It is a Bantu language that serves as both a first and second language to various groups and incorporates Arabic, Persian, German, Portuguese, English, and French vocabulary [23]. It is spoken in Tanzania, Kenya, Uganda, Rwanda, Burundi, Mozambique, Somalia, and the DRC, each having its own dialect that differs in both vocabulary and structure [22].

Hausa is from the Chadic language family [24] and has more first-language speakers than any other Sub-Saharan African language [25]. It is spoken by 100 million people in Nigeria, the Republic of Niger, Cameroon, Togo, Chad, Benin, Burkina Faso, and Ghana [24]. One-quarter of the vocabulary is drawn from Arabic with English and French influences [24]. It is widely considered to be the Lingua Franca in West Africa [25]. It has its own Latin-based alphabet, known as Boko, and distinguishes between short and long vowels, which can also affect word meaning but cannot be written [24], presenting a problem for NLP interpretation.

Speech recognition

Speech recognition (SR) is the first and foremost process in designing the FRI tool. SR can account for specific words, be focused on identifying specific traits of the speaker, or be used to translate entire conversations into text infinitely [26]. The goal of this report is to take the audio data and transcribe it in the respective language. This presents a challenge as, while more widely spoken western languages have received significant investment in the production of different training datasets, there are limited resources for less spoken languages, including those from the global South. SR has received increased attention in recent years, however, as voice is expected to overtake keyboards in order to overcome issues of

illiteracy and blindness [26]. The best voice recognition software can reportedly achieve as high as 97% accuracy in English, while Google, Microsoft, and IBM Watson are all disclosing accuracy of around 95% and increasing [27].

The process of building an SR model tends to be broken down into 1) preprocessing, 2) feature extraction, and 3) modeling [26], which are explained in detail below.

1. Preprocessing relates to the current form of the data. In the case of audio, this will be the sampling rate, the bit depth, the number of channels it is recorded in, and the overall noise level of the recording. Each of these contributes to the quality of the audio file and, therefore, the accuracy of the resultant model. The sample rate specifies the number of samples to take from an audio source material per second. A high sample rate increases the ability of digital audio to faithfully represent high frequencies. Standard telephone calls are transmitted at 8khz (8000 samples per second), whereas the minimum for effective speech recognition is 16khz, and this can go up to 44.1khz for a good-quality audio sample. Bit depth affects the dynamic range of a given audio sample. A higher bit depth allows the representation of more precise amplitudes. With a mix of loud and soft sounds within the same audio sample, a higher bit depth is necessary to represent those sounds correctly. Channels refer to mono (1) or stereo (2), and modeling requires a single input. Noise refers literally to background noise in the recording [28]. Each of the currently available processors has several options for resampling the files, increasing or reducing bit depth, changing the number of channels, and reducing noise. Depending on the file encoding (WMA, Mp3), however, these processes can introduce more or less a 'loss' of quality from the original file [28,29].

2. Feature extraction refers to the process of obtaining different features such as power, pitch, and vocal tract configuration from the speech signal. These are then transformed into parameters that allow the model to differentiate one utterance from another [26,30]. This begins by digitally representing the audio in its waveform, as shown in Figure 1. Sample rate and bit depth are crucial here to ensure the waveforms represent the proper amplitude across the sound sample [28]. These are then converted into a Mel Spectrogram, an image form based on signal strength or 'loudness,' whilst color becomes the indicator of amplitude [28]. This is based on the Mel scale, a perceptual scale of pitches judged by listeners. For human speech, it is common to take an additional step and convert the Mel Spectrogram into an MFCC (Mel Frequency Cepstral Coefficients). MFCCs produce a compressed representation of the Mel Spectrogram by extracting only the most essential frequency coefficients, which correspond to the frequency ranges at which humans speak [29]. Thus, they are more efficient for training [33].
3. Finally, modeling has grown rapidly since the late 2000s with the introduction of deep learning alongside the advancement of computing power and hardware, allowing researchers to make the neural networks deeper and more powerful, and providing availability and storage of more training data [31]. Examples of models include CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network). Moving forward from traditional machine learning using artificial neural networks, these models allow several layers, each of which can fulfill different functions to handle the data and learn with greater accuracy, more efficiently. They continue to rely upon labeled training data, however, requiring hundreds, if not thousands, of hours of audio files for their transcription to be effective [32]. The accuracy of the

model is then measured using the WER or Word Error Rate (Substitutions + Deletions + Insertions) / Number of Words Spoken), comparing the input with the output of the test set [33].

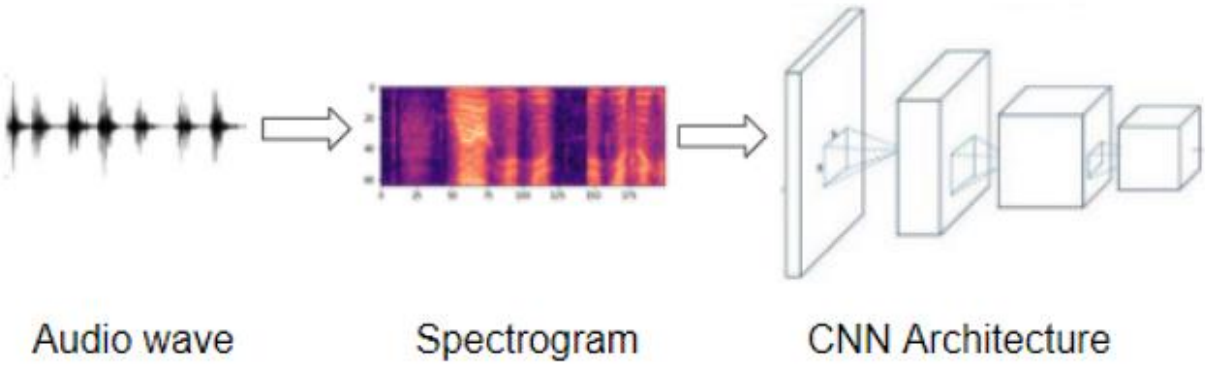

Figure 1 Early stages of the SR model, modified from [29]

A serious problem in the SR technique is code-switching, which occurs when speakers switch to a different language during a conversation, borrowing words, phrases, or sentences [34]. Intonation is very difficult to pick up; the sound of each character could be of a different duration, or there could be gaps and pauses between these characters. Several characters could be merged together or repeated, leading to a lack of clarity in the final transcription. These issues are exacerbated as the models are usually trained and tested on high-resource languages, which means little is known about how they will function in the real world or on smaller data sets [35]

Attempts to build approaches and overcome these problems for SR in less-resourced languages include Woldemariam [33], who attempts to apply transfer learning. This entails training a model on another, similar language to the one that is in view, then using that insight to make deductions about the target language. In the case of Amharic, processors learn from English and Mandarin to aid in setting

the correct parameters for learning that language before removing outside layers of the model and fine-tuning it to the target. This has been seen to produce a WER reduction from 38.72 percent to 24.50 percent.

More often than not, however, attempts have been made to increase the amount of training data available. In the case of Swahili, early efforts to produce SR data and models looked to crowdsourcing data [36]. Efforts to develop open-source datasets include Kencorpus for Swahili, Dholuo, and Luhya, which have a collection of 5,594 items, being 4,442 texts of 5.6 million words and 1,152 speech files worth 177 hours of audio [37]. The Lacuna fund has delivered several openly accessible text and speech resources for NLP research in a range of African languages. By far, the most impactful seems to be the Common Voice project [38] and the AI-sharing community Hugging Face [39].

The Common Voice project employs crowdsourcing for both data collection and data validation, becoming one of the largest corpora in the public domain for SR, both in terms of the number of hours and the number of languages [38]. In March of 2022, the Mozilla Foundation, the group behind Common Voice and the Firefox browser, awarded eight projects, each USD $50,000, for leveraging the Swahili language and voice technology to increase social and economic opportunities for marginalized groups in Kenya, Tanzania, and the DRC. These include:

- Kiazi Bora, or "Quality Potatoes" in Swahili, uses a voice-enabled application that advises vulnerable women living in rural areas and marginalized communities of Tanzania on the nutritional values of Orange Fleshed Sweet Potatoes (OFSP), farming skills for better yields, and detailed market availability for raw or processed OFSP food products, all through a voice data set app.

- LivHealth, a group that aims to correctly identify livestock syndromes and get timely interventions from qualified livestock practitioners. The project will build Kiswahili text-to-speech models for disseminating disease information to marginalized communities. They work closely with their partner, One Health Center in Africa (OHRECA), based at ILRI.
- Imarika, a conversational chatbot offering digital climate advisory services in English and Swahili that will support smallholder farmers to adapt to changing weather patterns.
- Duniacom Group, developers of a text and voice-based platform made available in the language of the underserved to provide wide access, adoption, and usage of digital agricultural advisory and financial services in Tanzania.

In September of 2022, Common Voice announced that their 100th language would be Twi of Ghana, the first language of 18 million Africans, demonstrating their commitment to bringing the benefits of NLP to those outside of the default European-colonial languages, stating even Google and Wikipedia "exclude almost half the African population on the basis of primary language" [20].

Examples of those taking advantage of the site include Babirye [17], who looked to expand SR into some of the less spoken Ugandan indigenous languages, including Runyankore-Rukiga, Acholi, and Lumasaaba, citing monitoring of local radio and dissemination of information for smallholder farmers among their primary motivations. They used available media on a text basis and Common Voice as their starting point for speech. They then described their process of attempting to gather more data by incentivizing communities within their universities to contribute to the existing data sets on Common Voice, creating monetary rewards for top speakers. The communities curated over 200,000 Swahili sentences and 100 hours

of voice contributions from 80 participants in Kenya and 90 in Tanzania. For Luganada, they actually reached a state of 'over contribution,' where there was insufficient text data to satisfy the number of hours of recorded speech.

Even more significant innovations are coming from within Hugging Face, which has revolutionized SR through Wav2vec [32] and XLS-R [39]. Traditionally, supervised machine learning is defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately according to its label, as opposed to unsupervised methods, which look to discover hidden patterns or data groupings in unlabeled data. Wav2vec takes advantage of unsupervised pretraining, essentially allowing a model to learn from and recognize patterns in unlabeled data. To achieve specific 'downstream' tasks, such as SR of a new dataset, then is simply a matter of fine-tuning the model to the specific language or context under examination using this limited labeled dataset. In this case, Baevski et al. [32] use a linear labeled layer on top of the original model, establishing a method for achieving highly accurate language transcription with as little as ten minutes of labeled data, achieving a word error rate of 4.8/8.2.

A team of researchers from some of the largest NLP groups, including Google AI and Hugging Face has built XLS-R [21,39] on top of Wav2Vec models, as demonstrated in Figure 2. It is a tool based on 436K hours of speech in 128 different languages, using data from across the largest open-source repositories, including Common Voice [38]. These data are 'pre-trained,' allowing the Wav2vec model to recognize patterns in the unlabeled data via several CNNs. Then they are fine-tuned for a specific task using a limited set of labeled data. To the best of their knowledge, they state, 'this is the largest effort to date, in making speech technology accessible for many more languages using publicly available data,' with the largest contribution being toward low and medium source languages. BABEL is

of particular interest as it incorporates noisy telephone conversational data. Examples of applications include Le [40] and Zanon Boito [35]. Le[4 0] used the XLS-R to account for dialectical differences between Coastal and Congolese and Central Swahili and their translation to English and French, achieving a WER of 36.75% and 31.25%, respectively. Denisov [41], in the same competition, achieved 12.5/17.6%.

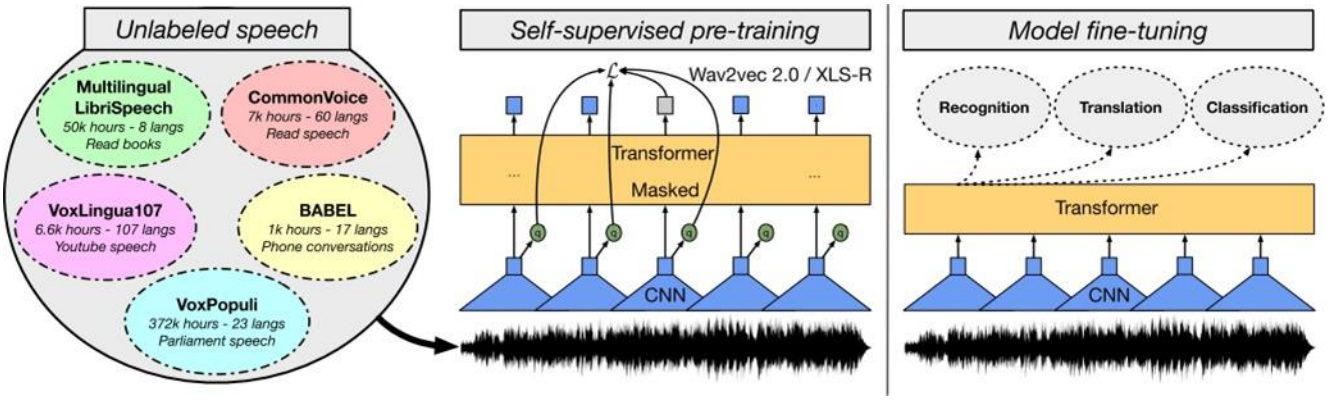

Figure 2 XLS-R model [39]

Applications of SR are wide. Wu [42] show how it can be used in the diagnosis of Parkinson's disease due to the distinct effect on the way sufferers talk, although the study was heavily dependent on the quality of sampling.

In Agriculture, significant attention has been paid to chatbots or question-and-answer systems for farmers that simulate a real conversation. Mostaço [43] developed AgronomoBot, a Telegram-based app for Brazilian farmers, using both text and speech to rapidly and efficiently deliver farmers' data on field conditions, such as air and soil temperature, air relative humidity, soil moisture, rainfall and wind speed. Kung [44] had a similar project for the pig industry in Taiwan and [45] for Malaysian farmers to improve disease diagnosis. In a study with the World Food Program, [46] collect nutritional information in Ethiopia and Kenya using an app

with a voice function to record each user's shopping list. They have been able to achieve a WER of 18% for Swahili audio transcription, dramatically reducing the effort necessary for participants to input their data.

Machine Translation

Machine translation (MT) is the use of computer-aided software to translate speech or text from one language to another. It allows communication at scale, with information being shared between different language groups with ease [22]. This is a significant need in Africa as news concerning the continent is almost exclusively published in English, French, or Arabic and is thereby inaccessible for speakers of only native African languages [47]. Like with SR, however, MT is similarly restricted in appropriate data sets. While the process of data acquisition is slightly easier, there are significant considerations and opportunities to introduce bias.

Languages have different origins, vocabularies, dialects, structures, slang, and sociolect, all factors which affect the accuracy of any machine translation approach [22,48]. Challenges for NLP also include the range of different data types (audio or text, utterances, phrases, related words) and the different requirements for modeling, ranging from tens to hundreds of thousands of sentences [49]. Underrepresentation of certain groups and languages in training corpora, which often disproportionately affects communities that are marginalized, excluded, or less frequently recorded or cultures where the educated are multilingual [48], is a major factor for the lack of engagement.

Swahili, however, has been the subject of machine translation research for over 50 years and is purportedly well supported by mainstream language processing software like Google and Microsoft Translate [22], as is Hausa. Much of the research in MT is toward producing effective models to translate under-resourced

languages using available datasets, that is, those that are not already used by Google or other popular translation services. The disadvantages met by speakers of languages outside of those covered by these services are stated as motivators for such research, especially in terms of exclusion from education resources and the dissemination of science [19].

In this regard, Inuwa-Dutse [25], in their collection of Hausa datasets, identified serious problems with Google Translate, suggesting insufficient data on colloquialisms and local dialects. This can only be remedied by picking up on some of the informal or day-to-day terms used in the language obtained via online social media, not news websites or religious texts. They also provide a framework for obtaining further data.

De Pauw's [50] early efforts to match Google's translation for Swahili found they were able to outperform them in translating from English, taking a deconstructed approach to Swahili vocabulary. The morphology of Swahili is quite complex, a single word representing a sentence of English. By breaking down the Swahili words into 'morphemes,' the machine may better interpret them. However, in the case of more obscure and difficult dialects, it has been shown that transfer learning and the creation of synthetic learning data is an effective strategy for developing translation models [22]. Massively multilingual models such as mBERT [51], XLM-R [21], or mT5 [52] use Wikipedia for pretraining and then fine-tuning downstream NLP tasks [49,51].

Whilst transfer learning has made great leaps, there remain problems when using high-resource European languages to learn low-resource African languages. Improved results have been shown using different African languages to learn from each other. Thus [53] develop MMTAfrica, the first many-to-many translation

system for six African languages. It uses a system of back-translation and reconstruction to train each language of the other.

These models continue to be dependent either on the amount of labeled data or unlabeled content online. Those languages with neither have very little hope of soon becoming the subject of effective computational linguistics [49]. While in their taxonomy of successful NLP languages, [49] Swahili is listed as a 'rising star,' having benefited greatly from unsupervised learning in recent years and having a large online presence, low-resource languages often have little written history, with few language experts or those familiar with NLP. Thus, those attempting to develop MT tools often cannot speak the language they are working with [54].

Nekoto et al. [54] demonstrate the efficiency of a participatory approach for sourcing data in low-resourced African languages. This entails ensuring that the speakers in the MT process originate from the countries where the low-resourced languages are spoken, involving lay persons in crowdsourcing data. However, they suggest citizen science projects involving participants in all stages of research are necessary for achieving quality evaluation, especially in languages unfamiliar to the NLP scientist. In their study, they employed 400 participants from at least 20 countries, enabling them to conduct a human evaluation study of model outputs, which has been one of the key limitations of previous approaches.

Examples of these models applied in the real world include Translators Without Borders, a group that uses machine translation for humanitarian aid, ensuring that local communities with language restrictions have access to the best information. They have been instrumental during the outbreak of the Ebola virus in the DRC in 2019 and across Africa during the COVID-19 pandemic. Their approach allows users to ask questions and receive answers in their own language [22].

Wefarm is a platform to connect farmers to each other for the purpose of knowledge sharing through SMS without the need for internet access. The farmer-to-farmer digital network connects farmers' questions and answers both online and through SMS. Wefarm intermediates the network, using machine learning to understand the request for information and patch it through to the right response. Founded in 2015, it exceeded 2.6 million users by 2020. It operates in collaboration with Amazon Web services (AWS), which offers a variety of services compatible with public and private cloud services and their open-source software base [55].

Corpus linguistics

Corpus linguistics is the study of language based on large collections of "real life" language use stored in corpora (or corpuses)—computerized databases created for linguistic research [56,57]. Corpus linguistics encompasses a number of analysis techniques that can be applied as needed rather than according to a particular protocol [58]. These include Key-word-in-context or concordance, collocations, word distribution, and corpus comparisons. A concordance displays all instances of a given the word in its immediate textual surroundings and helps the researcher to connect words of potential interest to the context [57].

Collocations denote the co-occurrence of two or more words, uncovering the meaning imbued in words by those words they collocate with. The strength of collocation between two words can be measured and represented statistically by the mutual information score of these two words, or 'faithfulness' [56,57,59]. Word dispersion measures the distribution of words over a number of texts rather than just one. Corpus comparison involves standing one's own text against a very large reference corpus to establish what is 'normal' and what is not, identifying the 'keyness' of certain words. A common comparison is the Brown Corpus [59].

In this regard, while corpus linguistics is largely reliant on machines for analysis, it differs from traditional computer-aided text analysis in that it focuses on lexical patterns rather than on categories and always involves a combination of quantitative and qualitative analysis [58]. It is, therefore, better suited for this analysis, demanding attention and participation from different stakeholders in interpreting and acting upon patterns in the data rather than simply monitoring answers and outputting statistics.

For a full review of linguistic analyses in agriculture, see [60]. To briefly summarize their article, they reviewed applications of text mining in agriculture or to 'extract, organize and classify information from text data… [that] can be presented to the user in a manner in which they can make informed decisions'. They emphasize the lack of development in corpus linguistics for agriculture due to the absence of open data and the continued use of the intellectual property to restrict access to insights gained. Key themes were Information Retrieval, Information Extraction, and Sentiment Analysis. The most common application is disease monitoring, while Information Retrieval/Question–Answering is also prolific, but the systems surveyed fall short of their promise. Of interest to this study, they cite that knowledge discovery has been used in bio-medicine to discover previously unknown treatments, capitalizing on the knowledge of users. Therefore, perhaps similar discoveries should be possible in agriculture using similar techniques. It is not unreasonable to expect to see advances in areas such as herbicide and pesticide development from knowledge extracted from large collections like the FRI data repository.

Specific examples include [61], who analyzed the sentiment of Twitter posts by farmers in Brazil, Russia, India, China, and South Africa (BRICS). They were able to show that positive sentiment was shared among these countries when discussing

agricultural policy and even when discussing the implementation of digital innovations such as blockchain. [62] suggest using linguistic corpora for improving the accessibility of ongoing trends in agricultural sciences to academic staff, for whom English is their second language, thus allowing more efficient dissemination of knowledge.

Combining each of the tools mentioned here, the UN Global Pulse (2016, 2017) has funded and developed, in collaboration with Ugandan country and university partners, a tool for analyzing public radio. They were able to identify local farmers' priorities, how they reflected Sustainable Development Goals (SDGs) such as health, education, or employment, and to link them to time and location. Their aims were to improve reporting of natural disasters, monitoring and tracking of the effectiveness of campaigns and early warning systems in "Ugandan English," Luganda, and Acholi.

The richest of all data sources was shown to be talk-show radio or discussion programs. UN Global Pulse (2016, 2017) noted, however, that the approach is difficult to scale, as their model required a separate configuration for each show they analyzed, in particular, what questions were asked and how the discussion was framed, suggesting the full impact of the project may not be fully accomplished.

They achieved a word error rate of 50% for Luganda and 60% for Acholi. The speech was more easily recognized during news broadcasts, where presenters are trained to articulate clearly, and is less recognized during call-ins, where the quality of the audio is poorer, and the speech is a rapid conversational style. In human analysis and transcription, they revealed that the results were rarely relevant due to a series of biases, largely related to a misunderstanding of context. For example, 23% of identified words were part of a commercial. They found men largely dominated the

conversation, whereas women were rarely heard. Finally, they focused on topic modeling and word filtering, seeking out issues specifically related to the UN's priorities rather than those stated by farmers.

In this study, we suggest that by working directly with the radio stations and using a more qualitative corpus linguistics approach, we might be able to gain a fuller and truer-to-life idea vision of farmers' voices. Not only a form of co-design, working with the organization to design a tool that best suits their needs, but also accessing a broader swath of farmers beyond the framing of experts.

Materials and Methods

Tools

The overall objective of this project was to make an efficient and easy-to-use tool for FRI based on free and open-source software. As such, all analysis was completed using Python programming software, implementing several modules and Application Programming Interfaces (APIs).

Google

Google's speech recognition and translation served as the baseline for this study. They have several different APIs that can be implemented in Python. Those used include SpeechRecognition [64] and Google Cloud [28]. The second is a paid service with a free starter pack, used mainly for their increased parameter tuning and to explore potentially improved tools. For SR, the details of the models are not shared, but they claim to have hundreds of thousands of hours worth of training data and can achieve upwards of 95% accuracy. They provide a measure of accuracy for each piece of data processed. It is assumed that these models also incorporate much of the open data sets used to train the other tools in this study. Both models allow for a specific Tanzanian or Kenyan version of Swahili. For MT, Google Translate was the

only tool used, assuming this is by far the most advanced tool in this regard. However, other tools may be considered in the future, pending human evaluation of translations.

XLR-S

As described above, the XLR-S claims to be the largest attempt to incorporate open data for SR. It is trained on 75 hours of spoken Hausa and 91 hours of Swahili. It has similar accuracy levels as Google and offers other African languages, such as Amharic, Arabic, Ganda, Kabyle, Kinyarwanda, Lingala, Shona, Somali, Twi, Yoruba, and Zulu [39].

SpeechBrain

SpeechBrain is an open-source and all-in-one speech toolkit designed to integrate well with Python, also based on Common Voice for SR. It was designed, however, to comprise a series of different tools for NLP, including speech recognition and detection. It is foremostly built on Pytorch, with emphasis on accessibility, ease of use, and replicability [65].

Corpus linguistics

The phases of analysis drew from [57,59,66]. The analysis began by generating a word frequency list to demonstrate which words occurred the most throughout the corpus. This involved tokenization, reducing sentences to strings of raw data, and lemmatization, the grouping of words based on their inflicted form so that they can be analyzed as one item (e.g., destroyed, destroys = destroy) [57]. In addition, stop words were removed, both common connectives and those unique to this corpus that would affect results. Finally, a quantified list was generated.

Of those most frequently occurring words, a concordance sample was produced, indicating how they appear in context. In a larger corpus, a random sample of 25

combinations is usually provided for each word of interest, however, in this case, a selection of the more meaningful words was used, and all combinations were revealed, providing the context of 5 words to the left and right of the selected 'node words.'

Next, a collocation analysis was conducted to calculate statistics that provide information about the strength of association (SOA) between lexical items. One key term was used ('farm') to measure 'faithfulness', or the probability of seeing the collocate given the presence of the target item. This goes for both the likelihood to occur when the target term is used and the likelihood the target term will appear given the collocate.

Keyness was calculated using the Brown Corpus, testing the percentage difference between the FRI data and that of a 'normal' corpus. Finally, an n-gram analysis was conducted, connecting words together in a series to view which are most influential in a corpus. Noun phrases were identified, or a concordance that will show a repeated phraseology where the verb occurs followed by a noun group [57]. These were produced and visualized using Textblob and Wordcloud, respectively.

Data

Data was collected by the FRI using an Interactive Voice Response (IVR) system called Uliza to present listeners with questions. The radio stations broadcasting the talk shows advertise a telephone number that farmers can call if they have something to comment on. Uliza will then return the call to the farmer such that they do not incur any cost. Callers are then asked a series of multiple-choice questions, which they answer using the dial pad before they answer an open-ended question with their voice.

In past surveys, polls were able to gather stakeholder opinions on nutrition. For example, a higher percentage of women than men said that those in need should eat first, whereas many participants identified moving away from chemical pesticides and fertilizers as a key priority. These findings are contentious and fall outside the general conception or trajectory of conventional agriculture. They reveal a human side to agricultural development, based on local values and desires; over 90% of participants were willing to act to reverse the effects of climate change, whereas 1 in 12 said that the only way to cope would be to move to another place. They even spoke about the little-discussed issue of shrinking villages and changing occupations of farmers in the face of difficult conditions.

The focus of the current work, however, is the voice reply. Questions are related to nutrition, food scarcity, climate change, information access, and how farmers felt they were at risk in each aspect. Answers are aggregated based on country, gender, and age. In general, one-third of calls are from women, while the majority appear to be over 30 years old. The potential audience is reported as 12,339,739 people [18].

Each audio signal was either recorded in 8khz or 22khz and so was resampled to 16khz, the ideal for SR. They were also converted to a single channel, mono-output. This was a hindrance to SR, however, as the low quality of the recordings will not fare as well when converted to a spectrogram. Noise reduction was attempted, but this reduced the quality of the audio until it was unrecognizable. Each of the models described above was applied to each audio file.

The linguistic analysis was performed on a human-transcribed Tanzanian Swahili dataset, as there was much more available data due to ethical and permission constraints.

Results

Speech recognition and translation

Swahili: What is the biggest threat to your family eating enough safe and nutritious food?

Table 1 displays the same Tanzanian Swahili transcription for each model and its translation to English using Google Translate. The final row shows the actual human translation. Google SR is surprisingly superior to Google Cloud, its conscription-based counterpart. XLS-R by Akashpb13 seems to be the most accurate. The Swahili transcription has 65% accurate, according to WER. This can be improved through further fine-tuning and using higher-quality audio in the next iteration of this study; however, if this is to be executed on a large scale, such errors will be lost in aggregated data, as will be shown in the next section.

Table 1 Language model and translation of Swahili

Model	Transcription	Translation
Google SR (confidence: 0.83117056)	hakuna bora na bora ni kupeleka pingamizi kujenga mwili na vipindi vya kulinda mwili wake kutoka kilimanjaro	there is nothing better and the best is to send objections to build the body and periods to protect his body from kilimanjaro
Google Cloud (confidence: 0.7421932)	hakuna bora maboga nikapeleka player mwili kujenga mwili lavington yule mwizi funny mzee yusufu kutoka kilimanjaro	there is no better pumpkin, I sent the player body to build the body, Lavington, the thief funny old man Yusufu from Kilimanjaro
SpeechBrain	uorauaborani haua kuilinda mwili aenga mwili nchainnaitani ya kulinda mwiliapani a fefu kutoka klimanjaro	uorauaborani does not kill to protect the body, he protects the body, the body protects the body, apani a fefu from the beginning

XLS-R (alokmatta)	pa kule cha kula boro chakula bora ni chakulecha kulinda mwizi cha gujenga mwizi na cha trupin na vitani ya kulinda mwizi hafani mzea hivyo sefu kutoka kile majaro	where to eat, the best food is the food to protect the thief, to protect the thief and to protect the thief in the war, he is not old, so safe from the old age.
XLS-R (Akashpb13)	heutaka kula chakula borchakula bora ni chakula cha kulinda mwili cha kujenga mwili na chaprotimi na vita iya kulinda mwilihapani kimzeijosefu kutoka kilimanjaro a	you don't want to eat bad food, the best food is food that protects the body, builds the body, and chaprotimi and war that protect the body.
Human translation	Nataka kula chakula bora, chakula bora ni chakula cha kujenga mwili (cha kujenga mwili) na cha protini na vitamini za kujenga mwili. Hapa ni mzee Joseph kutoka Kilimanjaro	I want to eat good food, good food is body building food (body building) and protein and body building vitamins.

Hausa: What is the biggest threat to your family eating enough safe and nutritious food?

Table 2 displays the Hausa transcription for each model, and it is translated to English using Google Translate. SpeechBrain and Google Cloud were not tested due to their poor performance with Swahili. Hausa has received significantly less development as an SR language. The Google model is only a 'preview' which is reflected in its output. The XLS-R by Mofe seems to be the most meaningful output. Again, human evaluation is necessary for future fine-tuning.

Table 2 Language model and translation of Hausa

Model	Transcription	Translation
Google SR	to barka madara haka sunana raira suna tabarau	To give good milk so saana sing lambs
XLS-R (anuragshas)	oburƙa murorhaƙaɗunarywon otɗa adaɗuna'maƙuron wanan i a ƙashir an pidapiloiluƙaɗuɗaƙonƙuaɗodumbard nuƙito hine ayon vi oɓoroicoviarcewdo'a wanciwocoɗalurorparcon'ƙ rauny unaeamun maƙal li na maɗara loloroidolunacare alurocovnitan'lpa mota muɗavaubutaƙiamagodeƙutalai	It's important to have a good relationship with the country.
XLS-R (Mofe)	cobarka mudor hakasunanrayn motsolemanandaga kaduna nakuro wanan cil a kashar ancibarbinitaɗada uƙuɗuguagangуɗatentambar da nukiditahine nayonjin rabara cobilitar cewa idan mutunada waniciwo otal lurartfagurcolta rauni yana i samunmatala ulina da matsalaraloreilanjinakarɗia lurar cbinancin hatan barfn ramota wtaba tabudatackenana goden gutalefia	Try to think like this, the Muslim man from Kaduna, I have a lot of problems with the development of the city and the city.

Linguistic analysis

If you had more power to change things, what would you do to make life better for farming families?

Table 3 shows the most frequently used words in the corpus. The farm is an expected top return, whereas the minister and government show to whom the callers are looking in order to make their lives better. Support, plant, assistance, and money seem to be the most telling in how that change might take place, and thus were used in the proceeding analysis to put this in context.

Table 4 demonstrates the concordance or 'key-word-in-context' of four of the most frequent words from Table 3. Using 'assist,' respondents appear to be seeking help accessing fertilizers or agrochemicals. Using 'plant,' they want both quality seeds and markets to sell their produce. 'Support' refers to loans and finance, while 'money' seems to tie each of these themes together, what the users seem to see as the bedrock of their activities.

Looking at all the derivatives of the most commonly occurring root word 'farm,' Tables 5 and 6 indicate the level of probability that collocates will occur. Table 5 shows that when the word 'aid' appears in the corpus, there is more than a 150% chance that 'farm' will appear. This is because the farm may appear multiple times in each utterance. 'Materials' appears here for the first time, indicating its significance as a need of farmers.

Table 6 indicates that the word 'to' will appear 67% of the time when the word 'farm 'appears. These words are large to be expected; however, 'support' and 'aid' occur again.

Table 3 Word frequency

Word	Frequency
farm	84
minister	62
support	35
government	15
farmer	14
money	14
plant	14
assist	13
youth	9
good	9
north	8
land	8
maize	8
price	8
agriculture	7
work	7
fertilizer	7
train	6
visit	6
seed	6
buy	5
financial	5
provide	5
concentrate	5
rain	5

Table 4 Concordance

Left Context	Node word	Right Context
agro chemicals to farmers to	assist	them in their farming
that they may need to	assist	them in their farming
on that report i can	assist	them my name is
the small scale farmers and	assist	us with fertilizers because we
with whatever they need to	assist	them do their farm work
i would have help or	assist	the farmers with whatever they
that they may need to	assist	them in their farming
seeds to them	plant	grains form their previous harvest
get the proper seed to	plant	my name is fatau
have seeds and land to	plant	so i need support in
money and quality seeds to	plant	and also allocate corporate buyers
do farming well when we	plant	our crops at the right
radio bar. i normally	plant	maize every season and that
provide funds to farmers to	support	them in their farming
farmers interest free loan to	support	their farming am in
the farmers with loan to	support	them my name is
agric, i would have	support	farmers with loans and provide
villages on regular basis and	support	them with money and other
a vast land for the	support	youth to farming on it
from kintampo, i need	support	for my business my
cropping seasons to educate and	support	them financially to aid them
provided loans to farmers to	support	them in their farming
consider given us loans to	support	our farming and also recommends
to plant so i need	support	in terms of land
farming and provide them with	support	my name is daniel
who are interested in farming	money	to do farming my
don 't have land and	money	you can't do farming
basis and support them with	money	and other farming material
finance should support farmers with	money	to aid them to do
kwadjo poku farmers need	money	to support them my
in their farming without	money	It's difficult to carry
farmers should be supported with	money	to aid them in their
armers will be able make	money	my name is yaw
do well so i need	money	to support my farming
land, seed, and	money	to plant or grow cashew
have supported the farmers with	money	and quality seeds to plant
to farmers and the	money	should come on time
also farming is all about	money	because weeding, planting and
applying fertilizer is all about	money	so the government should provide

Table 5 and 6 Faithfulness of collocation and of target word

Collocate	SOA
aid	1.588235
materials	1.5
loan	1.25
into	1.25
import	1.25
financially	1.25
activities	1.25
assist	1.230769
in	1.191489
buyers	1.125
who	1.076923
do	1.045455
them	1.030769
their	1.016949
provided	1
support	0.96875
to	0.929078
youth	0.888889
supported	0.888889
as	0.875

Collocate	SOA
to	0.678756
the	0.398964
is	0.388601
them	0.34715
my	0.321244
their	0.310881
in	0.290155
name	0.279793
and	0.243523
i	0.233161
have	0.222798
from	0.181347
would	0.160622
support	0.160622
with	0.150259
aid	0.139896
will	0.129534
do	0.119171
on	0.108808
agric	0.103627

Tables 7 and 8 indicate keyness. The results in table 7 indicate that 'farming' occurs with a frequency that is 138,407 % higher in the FRI corpus than the Brown Corpus. Other 'out of the ordinary' terms include 'fertilizers,' 'cocoa,' and 'mosquito.' The results in Table 8 indicate that 'there' has an 89% lower frequency in the FRI corpus than in the Brown Corpus. Again, while these are mostly to be expected, it is interesting that 'years' stands out as an absent theme. In the future, this tool could be more usefully applied, comparing new data with the existing FRI corpus.

Table 7 and 8 Keyness of most and least occurring words

Word	% difference
farming	138407.2443958447
mohammed	110705.79551667576
fertilizers	73770.53034445051
farmers	59709.94644366022
harvested	55302.89775833788
emmanuel	55302.89775833788
cocoa	55302.89775833788
smallscale	55302.89775833788
fertilizer	48377.53553854564
allocate	36835.265172225256
crops	33757.32640787316
minister	28055.570991942208
portia	27601.44887916894
buyers	27601.44887916894
cropping	27601.44887916894
kwame	27601.44887916894
dominic	27601.44887916894
enoch	27601.44887916894
mosquito	27601.44887916894
fatima	27601.44887916894

Word	% difference
there	-89.84550994165362
had	-89.20652683453383
out	-86.7899623847549
an	-85.21801020321828
one	-83.19596670963362
been	-77.5878245314167
which	-76.66263784400257
most	-76.11944062140607
no	-74.83973762109996
its	-74.3505102970658
after	-74.0865772879617
by	-73.89610923561162
many	-73.10538943770007
years	-72.32622489593513
a	-71.47611236056056
back	-71.32355188491826
about	-69.47498746097087
as	-69.47078233457067
down	-69.04866047020228
people	-67.29462942246879

Finally, Figures 3 and 4 are visualizations of Table 1, displaying the highest-frequency words and a new analysis of significant noun phrases in the form of bigrams and trigrams. Whilst Figure 3 reiterates the themes of minsters, aid, and support for farmers, Figure 4 puts some of these into context, for example, 'support farmers' and 'small scale farmers.' What stands out is 'regular basis' and 'corporate buyer' from the rest of the findings, suggesting the callers' desire for reliability and officiality in their support.

Figure 3 Word cloud of most frequent words

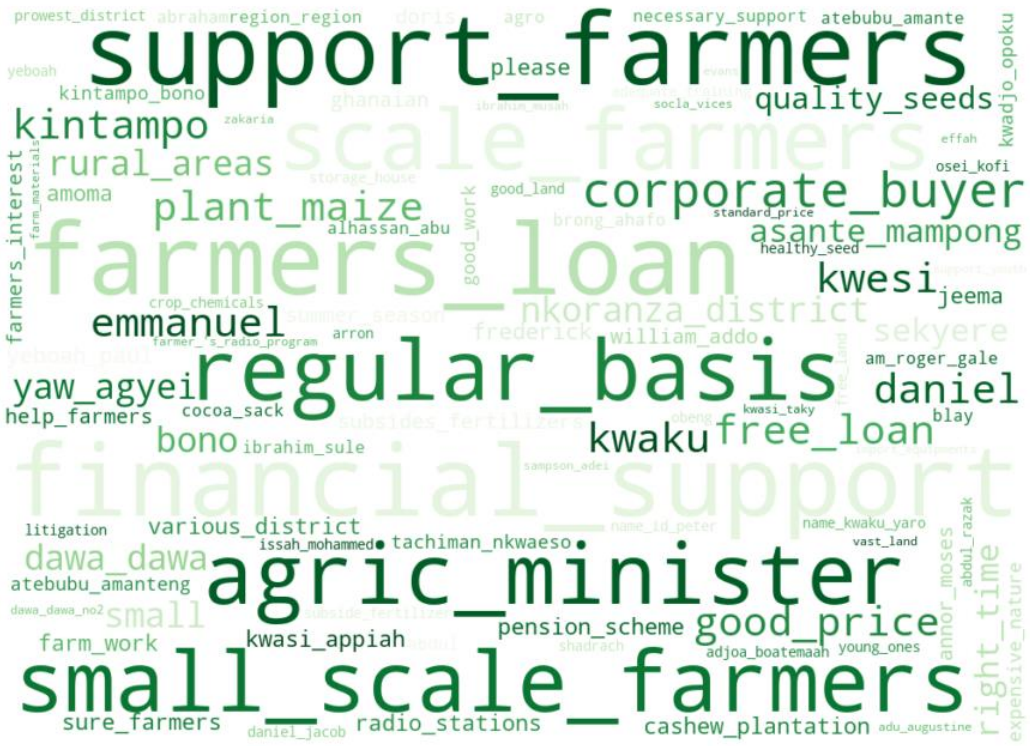

Figure 4 Word cloud of significant noun phrases

Discussion

Findings thus far demonstrate a proof-of-concept tool for transcribing, translating, and analyzing incoming calls from farmers in two African languages; Swahili and Hausa. Moving forward, there are several considerations on how to put this in practice and what each party seeks to gain from the collaboration. Foremost thoughts on the practicalities of further data analysis include limited computing power and limited quality of recordings. To process a large number of calls, there will be a need to access a larger processor with increased RAM and memory capacity. The current capacity is slow, inefficient, and struggles with longer audio files. Fortunately, the quality of recordings is already being increased to 22khz, which should dramatically increase the accuracy of SR in the future.

The work plan for the coming two years is intended to satisfy the following criteria:

1. How can the output be more useful to FRI? This report seeks to suggest ways in which each of the tools mentioned may be improved for the needs of FRI e.g., SR accuracy may be fine-tuned to existing FRI data, increasing accuracy, and statistics such as keyness may be more useful in comparing different FRI corpora. Both of these possibilities, however, are dependent on further data collection and having increased amounts of labeled data, each gaining greater representability and validity as the repository increases in size. Whilst there have been indications that the tools would be helpful in monitoring successes to funders, what would be more helpful to the intentions of FRI - for example - are there intentions to use this to respond to each farmer? Would this be helpful if re-packaged in a user-friendly interface? – there have been previous discussions with AWS to do this.

2. How can this proof-of-concept be made more impactful for farmers on the ground? In the discussion regarding the proof-of-concept, there has been considerable emphasis on how we can improve this tool to include more voices to account for extremely low-resource languages and dialects of the many farmers that might listen to FRI's programs. A number of options have been identified; the first and most economical might be to search the web or other known repositories for existing training data not yet applied to the models mentioned; the second might be to pursue AI approaches identified above, such as transfer learning, although this may require bringing on extra expertise; the third option, and likely the most reliable and to have the greatest contributions, is in the same vein as Babiyre [17], to incentivize and encourage the contribution of native language speakers to repositories on Common Voice.

   The existing infrastructure would provide an ideal space for building this data while also allowing its use by others, increasing CGIAR's role in the world of open data. It appears that the development of these kinds of technologies will almost certainly be led by academic-industry collaboration, and perhaps the only hope for African NLP is the backing of Agricultural research and development organizations with a genuine interest in building rural economies.
3. How can we develop this into a research delivery for CGIAR and beyond? The potential for this work to deliver a scientific contribution is considerable. Few studies have been able to listen to the voice of farmers on the scale proposed here. Not only that, but until very recently, this would have been an extremely slow and expensive ordeal, contacting each farmer individually and completing human transcription, translation, and analysis. Combining

each of these tools presented is a baseline for discussing how we can frame the most impactful output, whether that is by comparing different regions and demographics and how they respond to the question being asked or by applying further tools like sentiment analysis. Finally, there is a need for further discussion on how these insights might be best applied and how each organization involved might strategize according to the data produced.

Historically the relationship between farmers and extension organizations has not always been engaging and trustful, to say the least, with the introduction of new technologies and methods sparking the greatest amount of suspicion [4]. Issues of trust, accountability, and transparency must be paramount in forwarding this project. To aid in thinking about these questions, [48] have collected a series of biases apparent in the development of language processing models that should be mitigated in future work:

I. Discrimination, hate speech, and exclusion are of particular interest here considering the low-resource languages under examination, but also how certain opinions may be overrepresented that are harmful. Reports or allegations made on the radio are not necessarily true, for example, where the speaker is motivated by sensationalist or political objectives.
II. Information hazards, including accidentally revealing private data, be it trade secrets or those of users mentioning sensitive information whilst being recorded, may then be disclosed to a third party.
III. Misinformation harms; as opposed to disinformation, this risk is caused by poor quality or misleading data, which may cause incorrect insights without direct malintent by the user.
IV. Malicious uses; this is an intentional use of language modeling to spread disinformation, fraud, or malware.

V. Human-Computer Interaction Harms; this refers to 'conversational agents' or, as commonly proposed for farmers, chatbots and the difficulties incurred in user interactions, potential biases they generate, and the power they have in nudging, deception and manipulation.
VI. Environmental and Socioeconomic harms; this includes both the use of energy in building and operating language models and the environmental cost incurred and of exacerbating social inequalities by delivering uneven benefits or undermining and removing career opportunities.

Points I, III, and VI must play a significant role in how the current project moves forward. The work must remain reflexive to opinions stated by callers, in the understanding that certain ways of thinking may be overrepresented due to the biases already present in those who have access to the radio or can call in. Ensuring that all incoming data is of the best quality is essential, perhaps exploring how to get callers to speak slowly and clearly, so as not to fall into the old principle of 'garbage in, garbage out.' The intention of the work is to listen to farmers and deliver benefits equally. Agricultural extension has a long history of elite capture and damaging residual effects on local communities. Thus, the work should strive to be inclusive of and complementary to a wide range of actors.

Finally, reflecting on efforts by the UN Global Pulse [63], FRI has been successful in designing a system that incurs no cost for the user, allowing even the poorest members of a community to contribute to a discussion. There remain considerations, however, as to how talk show hosts frame the discussion and how this might affect chosen topics and useful answers. Some topics are less likely to be discussed on the radio because of social stigma or fear of retribution, adding to selection bias. It is unclear how FRI chooses topics, so they may be missing rarely discussed topics or those that they are not already aware of.

If building a more diverse voice training dataset, it is important that we keep track of the metrics around the voice contributors, such as age group and gender, whilst also remaining respectful of their privacy [17].

It is hoped that these thoughts will form the basis for future farmer-centered design, putting the voice of farmers first and foremost in the design of further Interaction, i.e., what are the lives they aspire to live [67] and how do they constitute success beyond yield [68]? - These are critical questions if we want to build lasting sustainability and complementary rather than coercive AI tools [7].

References

1. Sambasivn N, Holbrook J. for the Next Billion Users. Interactiona. 2019.
2. Velten S, Leventon J, Jager N, Newig J. What Is Sustainable Agriculture? A Systematic Review. 2015;7: 7833–7865. doi:10.3390/su7067833
3. Garzón Delvaux PA, Riesgo PA, Gomez L. Sustainable agricultural practices and their adoption in sub-Saharan Africa - A selected review. Seville, Spain; 2020. doi:10.2760/360761
4. Cook BR, Satizábal P, Curnow J. Humanising agricultural extension: A review. World Dev. 2021;140: 105337. doi:10.1016/j.worlddev.2020.105337
5. Chandra R, Collis S. Digital agriculture for smallscale producers. Commun ACM. 2021;64: 75–84. doi:10.1145/3454008
6. Orn D, Duan L, Liang Y, Siy H, Subramaniam M. Agro-AI Education: Artificial Intelligence for Future Farmers. SIGITE 2020 - Proc 21st Annu Conf Inf Technol Educ. 2020; 54–57. doi:10.1145/3368308.3415457
7. Heldert C. What Does AI mean for Smallholder farmers? A Proposal for Farmer-centred ai research. Angew Chemie Int Ed 6(11), 951–952. 2021.
8. Ait Issad H, Aoudjit R, Rodrigues JJPC. A comprehensive review of Data Mining techniques in smart agriculture. Eng Agric Environ Food. 2019;12: 511–525. doi:10.1016/j.eaef.2019.11.003
9. Finger R, Swinton SM, El Benni N, Walter A. Precision Farming at the Nexus of Agricultural Production and the Environment. 2019. doi:10.1146/annurev-resource-100518
10. Birner R, Daum | Thomas, Pray | Carl. Who drives the digital revolution in agriculture? A review of supply-side trends, players and challenges. 2021. doi:10.1002/aepp.13145
11. Lowder SK, Skoet J, Raney T. The number, size, and distribution of farms, smallholder farms, and family farms Worldwide. World Dev. 2016;87: 16–29. doi:10.1016/j.worlddev.2015.10.041
12. Mehrabi Z, McDowell MJ, Ricciardi V, Levers C, Martinez JD, Mehrabi N, et al. The global divide in data-driven farming. Nat Sustain. 2020. doi:10.1038/s41893-020-00631-0
13. Feder G, Willett A, Zijp W. Agricultural Extension: Generic Challenges and the Ingredients for Solutions. Knowl Gener Tech Chang. 2001; 313–353. doi:10.1007/978-1-4615-1499-2_15
14. UN Global Pulse. Public radio content analysis tool. Tool Ser. 2016; 1–2.

15. HLPE-UN. Data collection and analysis tools for food security and nutrition. 2022. Available: https://www.fao.org/fsnforum/cfs-hlpe/discussions/data-collection-analysis

16. Silvestri S, Richard M, Edward B, Dharmesh G, Dannie R. Going digital in agriculture: how radio and SMS can scale-up smallholder participation in legume- based sustainable agricultural intensification practices and technologies in Tanzania. 2020. doi:10.1080/14735903.2020.1750796

17. Babirye C, Nakatumba-Nabende J, … Building Text and Speech Datasets for Low Resourced Languages: A Case of Languages in East Africa. … Lang Process. 2022; 1–11. Available: https://openreview.net/forum?id=SO-U99z4U-q%0Ahttps://openreview.net/pdf?id=SO-U99z4U-q

18. FRI. LISTENING TO RURAL PEOPLE 2021. 2021. https://www.ifad.org/documents/38714170/43721925/onair_dialogues_full_e.pdf/fe73de8c-99be-36ce-2e04-f69cdc0228f7?t=1631881790801

19. Martinus L, Abbott JZ. A Focus on Neural Machine Translation for African Languages. 2019. Available: http://arxiv.org/abs/1906.05685

20. Onukwue A. Ghana's Twi added to Mozilla's Common Voice language project — Quartz Africa. In: Quartz Africa [Internet]. 2022 [cited 23 Sep 2022]. Available: https://qz.com/ghana-s-most-popular-language-will-be-available-to-more-1849572359

21. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, et al. Unsupervised Cross-lingual Representation Learning at Scale. 2020; 8440–8451. doi:10.18653/V1/2020.ACL-MAIN.747

22. Oktem A, Deluca E, Bashizi R, Paquin E, Tang G. Congolese Swahili Machine Translation for Humanitarian Response. 2021. Available: http://kevindonnelly.org.uk

23. Oirere AM, Deshmukh RR, Shrishrimal PP, Waghmare VB. Swahili Text and Speech Corpus: a Review. Asian J Comput Sci Inf Technol J homepage. 2012;2: 286–290. Available: http://www.innovativejournal.in/index.php/ajcsit

24. Schlippe T, Djomgang EGK, Vu NT, Ochs S, Schultz T. Hausa Large Vocabulary Continuous Speech Recognition. 3rd Work Spok Lang Technol Under-Resourced Lang SLTU 2012. 2012; 11–14.

25. Inuwa-Dutse I. The first large scale collection of diverse Hausa language datasets. 2021; 1–10. Available: http://arxiv.org/abs/2102.06991

26. Swetha P, Srilatha J. Applications of Speech Recognition in the Agriculture Sector: A Review. ECS Trans. 2022;107: 19377–19383. doi:10.1149/10701.19377ecst

27. LeadDesk. Speech-to-text: A full guide for contact centers wanting to harness data. In: 2022 [Internet]. [cited 26 Sep 2022]. Available: https://leaddesk.com/blog/speech-to-text-guide-for-contact-centers/

28. Google Cloud. Introduction to audio encoding | Cloud Speech-to-Text Documentation | Google Cloud. 2022 [cited 26 Sep 2022]. Available: https://cloud.google.com/speech-to-text/docs/encoding

29. Doshi K. Audio Deep Learning Made Simple: Automatic Speech Recognition (ASR), How it Works . In: Towards Data Science [Internet]. 2021 [cited 9 Sep 2022]. Available: https://towardsdatascience.com/audio-deep-learning-made-simple-automatic-speech-recognition-asr-how-it-works-716cfce4c706

30. Lee C, Hyun D, Choi E, Go J, Lee C. Optimizing feature extraction for speech recognition. IEEE Trans Speech Audio Process. 2003;11: 80–87. doi:10.1109/TSA.2002.805644

31. Kumar A, Verma S, Mangla H. A Survey of Deep Learning Techniques in Speech Recognition. Proc - IEEE 2018 Int Conf Adv Comput Commun Control Networking, ICACCCN 2018. 2018; 179–185. doi:10.1109/ICACCCN.2018.8748399

32. Baevski A, Zhou H, Mohamed A, Auli M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. 2020. Available: https://github.com/pytorch/fairseq

33. Woldemariam Y. Transfer Learning for Less-Resourced Semitic Languages Speech Recognition: the Case of Amharic. Proc 1st Jt Work Spok Lang Technol Under-resourced Lang Collab Comput Under-Resourced Lang. 2020; 61–69. Available: https://aclanthology.org/2020.sltu-1.9

34. Kleynhans N, Hartman W, Van Niekerk D, Van Heerden C, Schwartz R, Tsakalidis S, et al. Code-switched English Pronunciation Modeling for Swahili Spoken Term Detection. Procedia Comput Sci. 2016;81: 128–135. doi:10.1016/J.PROCS.2016.04.040

35. Zanon Boito M, Ortega J, Riguidel H, Laurent A, Barrault L, Bougares F, et al. ON-TRAC Consortium Systems for the IWSLT 2022 Dialect and Low-resource Speech Translation Tasks. 2022; 308–318. doi:10.18653/v1/2022.iwslt-1.28

36. Gelas H, Besacier L, Pellegrino F. DEVELOPMENTS OF SWAHILI RESOURCES FOR AN AUTOMATIC SPEECH RECOGNITION SYSTEM. 2011 [cited 11 Sep 2022]. Available: http://news.bbc.co.uk/2/hi/africa/4527876.stm

37. Wanjawa B, Wanzare L, Indede F, McOryango O, Ombui E, Muchemi L. Kencorpus: A Kenyan Language Corpus of Swahili, Dholuo and Luhya for Natural Language Processing Tasks. ArXiv. 2022.

38. Ardila R, Branson M, Davis K, Henretty M, Kohler M, Meyer J, et al. Common voice: A massively-multilingual speech corpus. Lr 2020 - 12th Int Conf Lang Resour Eval Conf Proc. 2020; 4218–4222.

39. Babu A, Wang C, Tjandra A, Lakhotia K, Xu Q, Goyal N, et al. XLS-R: SELF-SUPERVISED CROSS-LINGUAL SPEECH REPRESENTATION LEARNING AT SCALE. 2021 [cited 8 Sep 2022]. Available: https://huggingface.co/models?other=xls_r

40. Le H, Barbier F, Nguyen H, Tomashenko N, Mdhaffar S, Gahbiche SG, et al. ON-TRAC' systems for the IWSLT 2021 low-resource speech translation and multilingual speech translation shared tasks. 2021; 169–174. doi:10.18653/v1/2021.iwslt-1.20

41. Denisov P, Mager M, Vu NT. IMS' Systems for the IWSLT 2021 Low-Resource Speech Translation Task. 2021; 175–181. doi:10.18653/v1/2021.iwslt-1.21

42. Wu K, Zhang D, Lu G, Guo Z. Influence of sampling rate on voice analysis for assessment of Parkinson's disease. J Acoust Soc Am. 2018;144: 1416–1423. doi:10.1121/1.5053681

43. Mostaço GM, Campos LB, Cugnasca CE. AgronomoBot: a smart answering Chatbot applied to agricultural sensor networks Development of 5G communication networks View project Projeto de Rastreabilidade para a Indústria Vinícola Brasileira View project. 2018. Available: https://www.researchgate.net/publication/327212062

44. Kung HY, Yu RW, Chen CH, Tsai CW, Lin CY. Intelligent pig-raising knowledge question-answering system based on neural network schemes. Agron J. 2021;113: 906–922. doi:10.1002/agj2.20622

45. Windiatmoko Y, Rahmadi R, Hidayatullah AF, Ong RJ, [1,2] RAAR, Sudin[1,2] S, et al. A Review of Chatbot development for Dynamic Web-based Knowledge Management System (KMS) in Small Scale Agriculture. J Phys Conf Ser. 2021;1755: 012051. doi:10.1088/1742-6596/1755/1/012051

46. Getaneh Biruk. Amharic and Swahili Language Speech Recognition — Speech-to-Text . In: Medium [Internet]. 2022 [cited 11 Sep 2022]. Available: https://medium.com/@birukgetaneh/amharic-and-swahili-language-speech-recognition-speech-to-text-257e00a317b6

47. Adelani D, Alabi J, Fan A, Kreutzer J, Shen X, Reid M, et al. A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation. 2022; 3053–3070. doi:10.18653/v1/2022.naacl-main.223

48. Weidinger L, Uesato J, Rauh M, Griffin C, Huang P Sen, Mellor J, et al. Taxonomy of Risks posed by Language Models. ACM Int Conf Proceeding Ser. 2022; 214–229. doi:10.1145/3531146.3533088

49. Joshi P, Santy S, Budhiraja A, Bali K, Choudhury M. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. 2020; 6282–6293. doi:10.18653/v1/2020.acl-main.560

50. De Pauw G, Wagacha PW, De Schryver G-M. Towards English-Swahili Machine Translation. Res Work Isr Sci Found Mach Transl Morphol Lang. 2011. Available: http://www.mt-archive.info/MTMRL-2011-DePauw.pdf%0Ahttps://biblio.ugent.be/publication/1851705

51. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pretraining of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conf North Am

Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf. 2019;1: 4171–4186.

52. Xue L, Constant N, Roberts A, Kale M, Al-Rfou R, Siddhant A, et al. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. 2021 [cited 18 Sep 2022]. Available: https://pypi.org/project/langdetect/

53. Emezue CC, Dossou BFP. MMTAfrica: Multilingual Machine Translation for African Languages. Proc Sixth Conf Mach Transl. 2021; 403–416.

54. Nekoto W, Marivate V, Matsila T, Fasubaa T, Kolawole T, Fagbohungbe T, et al. Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages ∀ ∗ ,. 2020.

55. Sekinah T. How Wefarm is using NLP and SMS powered by the cloud to help farmers share knowledge. In: diginomica [Internet]. 2020 [cited 11 Sep 2022]. Available: https://diginomica.com/how-wefarm-using-nlp-and-sms-powered-cloud-help-farmers-share-knowledge

56. Kubler S, Zinsmeister H. Corpus Linguistics and Linguistically Annotated Corpora. Bloomsbury. London: Bloomsbury Academic; 2015.

57. Brookes G, McEnery T. Corpus linguistics. Routledge Handb English Lang Digit Humanit. 2020; 378–404. doi:10.4324/9781003031758-20

58. Pollach I. Taming textual data: The contribution of corpus linguistics to computer-aided text analysis. Organ Res Methods. 2012;15: 263–287. doi:10.1177/1094428111417451

59. Kyle K. This is a list of Python mini-tutorials | Introduction to Corpus Analysis With Python 3. In: GitHub [Internet]. 2020 [cited 24 Sep 2022]. Available: https://kristopherkyle.github.io/corpus-analysis-python/py_index.html

60. Drury B, Roche M. A survey of the applications of text mining for agriculture. Comput Electron Agric. 2019;163: 104864. doi:10.1016/j.compag.2019.104864

61. Hooda A. Sentiment Analysis of Recent Tweets for Agriculture from BRICS Countries. 2020. doi:10.13140/RG.2.2.17830.68166

62. Melnikov M, Mallyamova E, Morozova M. Linguistic corpus as a means of adaptation of modern scientific agricultural approaches. BIO Web Conf. 2020;17: 00182. doi:10.1051/BIOCONF/20201700182

63. UN Global Pulse. Using Machine Learning to Analyse Opportunities for Sustainable Development. Glob Pulse. 2017. https://unsdg.un.org/latest/blog/using-machine-learning-accelerate-sustainable-development-solutions-uganda

64. Zhang A. Speech Recognition (version 3.8). In: Github [Internet]. 2017. Available: https://github.com/Uberi/speech_recognition#readme

65. Ravanelli M, Parcollet T, Plantinga P, Rouhe A, Cornell S, Lugosch L, et al. SpeechBrain: A General-Purpose Speech Toolkit. 2021; 1–34. Available:

http://arxiv.org/abs/2106.04624

66. Benjamin B, Bilbro R, Ijeda T. Applied Text analysis with Python. Sebastopol: O'Reilly; 2018.

67. Flachs A. Cultivating knowledge. Angew Chemie Int Ed 6(11), 951–952. 2021; 5–48.

68. Leach M, Scoones I, Stirling A. Dynamic sustainabilities; Technology, environment, social justice. 2010.

Appendix

Proposed Work 2023-2024

Having completed two full iterations of our human-centered design process – consulting FRI to discover their needs and we, as CGIAR, can best apply our skills to meet them – we have agreed the best path forward would be a) to engage in a capacity building exercise, handing over the skills to do the above analysis such that FRI staff can begin to use it b) to collaborate in a field study focusing on gender, measuring the combined impact of radio and the translation tool.

Capacity building

This tool has extended the number of languages available from the most commercially available software, and there is a significant skills gap in who can operate and run it. The next step, therefore, is for staff to transfer those skills to technicians operating on the ground, doing the day-to-day analysis of FRI and understanding how they can make use of it. Within the next months, the principal investigator will fly to one of two destinations where the technicians operate, Addis Abba or Accra. There they will demonstrate to FRI staff how to run the code whilst learning about other potential needs and building trust.

Field study

The second step is to gather some additional data, with ethical and participant approval, to draw some scientific insight from the farmers. Not only will this give voice to farmers, likely being among the largest studies of African farmers' qualitative data, but FRI has also suggested gender as a focus, hoping to understand how they can better design their shows to reach more and achieve greater impact on female farmers.

Gender inequality continues to threaten food security in several African countries, restricting women's access to land, their participation in decision-making, and their ability to benefit from profits. FRI has provided examples of its approach to and impact on these issues. They aspire to respond to the communication and information needs of women, engaging in capacity building with local radio stations to ensure the production of programs that are aimed at both men and women and facilitating radio listening and participation by women.

The Her Farm Radio project, which finished in 2017, partnered with 13 radio stations, providing them with 49 days of training, and provided wind-up radios to 134 community listening groups

The ongoing Scaling Her Voice on Air project has reportedly been broadcast on 73 radio stations and reached 15 million people. They have made and broadcasted almost 700 radio shows that discuss women's rights, decision-making, and sharing the workload within the household. They also broadcast radio dramas that carry a positive message. Impact surveys suggested that 90% of participants found reduced violence and an increase in decision-making and access to land.

FRI's newest project, which is about to launch, is named On-Air for Gender-Inclusive Nature-based Solutions. The focus is on developing climate adaptation strategies according to the needs of women and youth. FRI intends to develop a series of 200 radio documentaries, aiming to shift inequitable social norms at household, community, and national levels, as well as the systems and structure that (re)produce them. It is here that the audio analytic tool comes into play; not only can it be used to more easily and efficiently gather insights on what topics are most useful to listeners and how that should be framed, but also learning from that data, how it compares to previous, male-dominated data sets, leading to a truly user-centered design of the radio show.