2021-04

Amharic Dialogue Based Expert System on pregnancy

Amir, Ali Mohammed

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES

FACULTY OF COMPUTING

Amharic Dialogue Based Expert System on pregnancy

By

Amir Ali Mohammed

Advisor Name: Tesfa Tegegne (PhD)

Bahir Dar, Ethiopia
April, 2021

Amharic Dialogue Based Expert System on pregnancy

Amir Ali Mohammed

A thesis submitted to the school of Research and Graduate Studies of Bahir Dar Institute of Technology, Bahir Dar University in partial fulfillment of the requirements for the degree of Master of Science in Information Technology in the regular program in the faculty of computing and informatics.

Advisor Name: Tesfa Tegegne (PhD)

Bahir Dar, Ethiopia

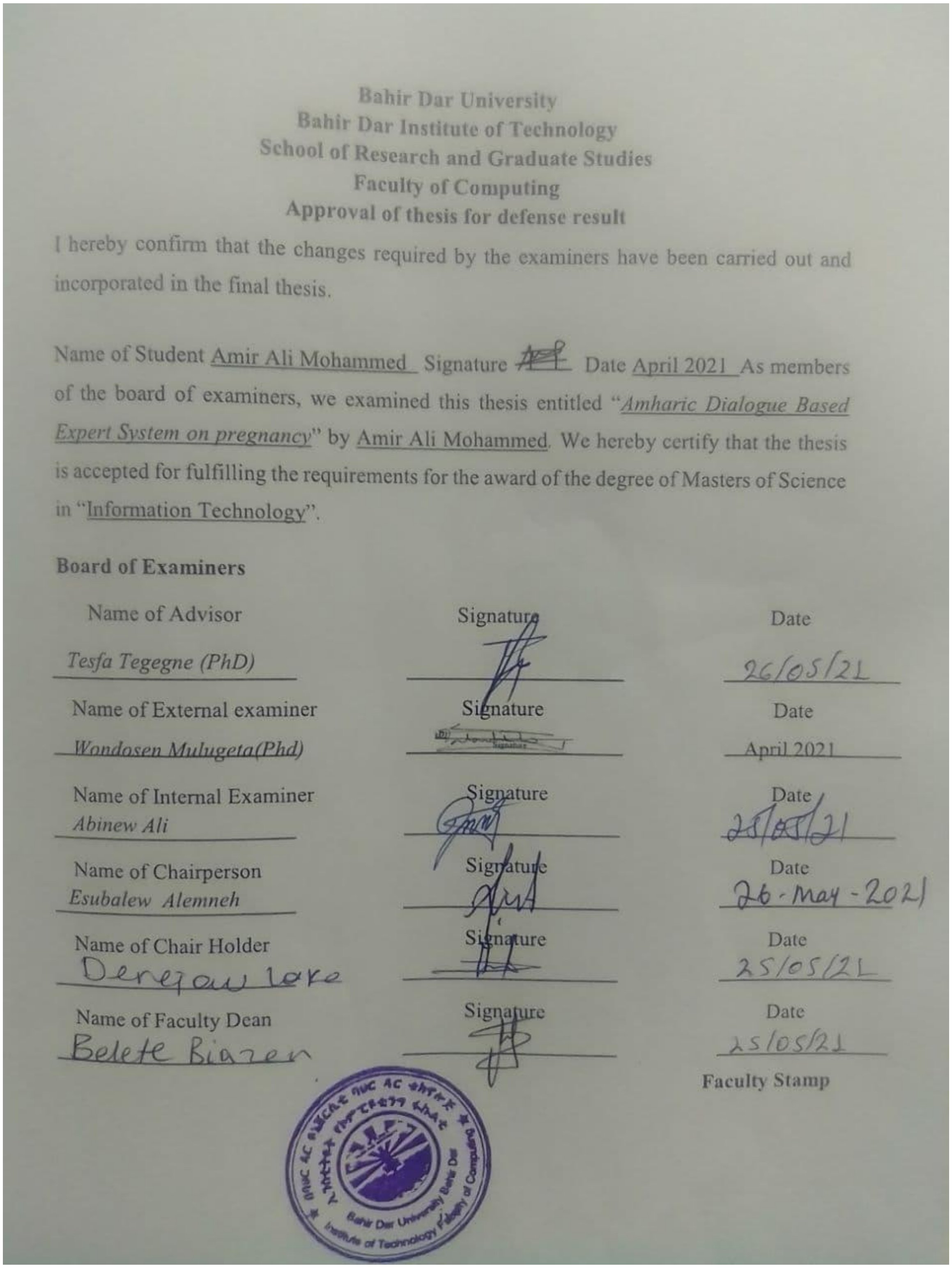
April,2021

Declaration

This is to certify that the thesis entitled “Amharic Dialogue based Expert System On pregnancy”, submitted in partial fulfillment of the requirements for the degree of Master of Science in Information Technology under Computing Faculty, Bahir Dar Institute of Technology, is a record of original work carried out by me and has never been submitted to this or any other institution to get any other degree or certificates. The assistance and help I received during the course of this investigation have been duly acknowledged.

Amir Ali Mohammed April, 2021

Name of the candidate signature Date



Bahir Dar University
Bahir Dar Institute of Technology
School of Research and Graduate Studies
Faculty of Computing
Approval of thesis for defense result

I hereby confirm that the changes required by the examiners have been carried out and incorporated in the final thesis.

Name of Student Amir Ali Mohammed Signature Date April 2021 As members of the board of examiners, we examined this thesis entitled "*Amharic Dialogue Based Expert System on pregnancy*" by Amir Ali Mohammed. We hereby certify that the thesis is accepted for fulfilling the requirements for the award of the degree of Masters of Science in "Information Technology".

Board of Examiners

Name of Advisor	Signature	Date
Tesfa Tegegne (PhD)		26/05/21
Name of External examiner	Signature	Date
Wondosen Mulugeta(Phd)		April 2021
Name of Internal Examiner *Abinew Ali*	Signature	Date 25/05/21
Name of Chairperson *Esubalew Alemneh*	Signature	Date 26-May-2021
Name of Chair Holder Derejaw lake	Signature	Date 25/05/21
Name of Faculty Dean Belete Biazen	Signature	Date 25/05/21

Faculty Stamp

በባሕር ዳር ዩኒቨርሲቲ ባሕር ዳር ቴክኖሎጂ ኢንስቲትዩት ★ የኮምፒዩቲንግ ፋካልቲ
Bahir Dar University Bahir Dar Institute of Technology Faculty of Computing

.

Acknowledgements

First of all, I would like to thank Allah who gave me a strength to finish this study. Then after I want to thank my advisor Dr. Tesfa T. who support me in my work. And my wife, family, friends and others they stand beside me for the success of this work. They appreciate and support me by giving important ideas. I have big respect for all persons who support me either directly or indirectly.

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
ASR	Automatic Speech Recognizer
HMM	Hidden Markove Model
JSON	JavaScript Object notation
NLU	Natural Language Understanding
RNN	Recurrent Neural Network
Seq2Seq	Sequence to Sequence
TTS	Text –to-Speech Synthesis

Table of Content

List of Figures

Abstract

Human beings communicate with different systems in different ways but by considering the growth of technology speech communication becomes familiar and preferable to communicate with systems. In our country patient-physician ratio is not balanced this led to increase the burden of physicians in the health centers. In this domain an expert system needs to be implemented to solve existing problems of our country. Conversational systems which are developed for consulting about published low of Ethiopia and on hotel and restaurant domain text is used for conversation. The proposed study designs end-to-end speech based conversational AI for the domain of healthcare using under resourced language Amharic. To achieve our study, the required dataset is collected through interviewing pregnant women, physicians and nutrition professionals. Besides, we make an intensive literature review and analyze documents (manuals about pregnancy). A total of 560 text and audio data is prepared. We split the data into 80% for training and 20% for testing. The conversational system performs different tasks such as greeting, farewell, giving advice and suggestions, etc. Generally speaking, the proposed system achieved 92% accuracy. Thus, the proposed system shows an encouraging result with limited dataset. As future work developing Amharic speech recognizer that abeles to recognize different people's with respect to age, gender and accent. A total of 560 text and audio data is prepared. We split the data into 80% for training and 20% for testing. The conversational system performs

Keywords: conversational AI, ASR, user and system initiative, task and no-task oriented

Chapter 1: Introduction

1.1. Background

Peoples communicate each other in different ways to exchange information. But speech is the most common way of communication. In the same way humans interact with computers through graphical user interface, command etc. Users give an input for the computer in different ways such as by typing text, speech, navigating through graphical menus, buttons and icons. Based on the user input the system respond for users. The beginning of speech enabled personal assistances such as Apple's Siri, Googles Now and Microsoft Cortana brings unfamiliar medium of interaction which is speech. As a result, ongoing researches are done in speech technology and artificial intelligence to enable computers to understand natural languages like *English.* The main objective is to interact human users with computer in human like manner. The motivation for this is speech communication comes naturally to humans. Spoken human computer interaction led to the emergence of spoken dialogue system (Meena, 2016).

After the emergence of Apple's Siri, googles Now and so on different conversational agents are developed. Chatbot's are the current conversational agents to communicate with users as like as human (Laranjo et al., 2018). Those conversational agents able to have conversation with the users in different domains. The conversational agents play vital role in the health care and medical care by giving advice or consultation for patients when behavioral change happened and they assist clinicians at the time of diagnoses (Nishida, Nakazawa, Ohmoto, & Mohammad, 2014). During conversation with the conversational agent or dialogue system you may use text or speech inputs. Conversational agent which can support speech input simplify the way to communicate with users.

Spoken dialogue systems enables humans to interact with computers to access information and services available on the computer or over the internet using speech as interaction medium between the user and computer (Lee et al., 2010). There is difference between spoken dialogue system and other systems which use speech as input and output those are dictation systems, command-and-control systems, and text-to-speech systems. Dictation systems use speech recognition to transcribe a user’s speech into words and sentences, but do not process the transcribed data further, whereas command-and-control systems interpret user commands, such as “*Make this bold*” or “*Switch on the light*” by translating the word in to action. Similarly, text-to-speech systems, such as screen readers or SMS readers for mobile phones, produce spoken output from a segment of written text but do not engage the user in any interaction (Jokinen & McTear, 2009).

There are some previous works done on conversational AI for Amharic language which is done by (Seyoum, 2015) and (Asimare, 2020). In the first work they try to find the address of hotels and restaurants. In this work the speech recognizer is done only in word level. This makes it difficult to describe or ask something that the users want to. The second work is a chat bot developed to provide clarification about Ethiopian Constitution based on user’s request. In this work there is no Automatic Speech Recognizer (ASR) and Text –to-Speech Synthesis (TTS).

Amharic is one of the Ethiopian Semitic languages, which is subgrouping within the Semitic branch of the Afro-asiatic languages. It is spoken as a first language in Amhara region and as a lingua franca or second language in the remaining regions of Ethiopia. The language serves as the official working language of Ethiopia, and it also used as the official working language in several states of the country. Amharic language use letters which is called *Fidel* those are generated from *Geez,* peoples write using those letters for different purpose.

In Ethiopia there is lack of doctors throughout the country, to simplify the burden of doctors and nurses using Artificial intelligence systems may support healthcare service delivery. One of those artificial intelligent systems are conversational agents or dialogue systems. They play an important role in giving support for clinicians in health institutions. As (Berhan, 2008) described in the 23 years’ period (1984-2006), the highest and lowest physician to population ratios in the public sector were found to be (1: 28,000) in 1989 and (1: 118,000) in 2006, respectively. In 2006, the physician to population ratio in Amhara, Oromia and SNNPR regional states was computed to be 1: 280,000, 1: 220,000, and 1:

230,000, respectively. This study shows the physician to population ratio is not balanced according to the number of populations in the country. To solve the problem, we provide conversational agent or dialogue system to enable pregnant women to get advice during their pregnancy period using speech technology. And by interacting with the system, they can get information, and advice from the system. It can be considered as a virtual doctor.

1.2. Motivation

The ratio of physician and patients throughout the country is not fair and balanced. (Berhan, 2008) put the patient-physician ratio to shows the reality in the ground. Currently the technology is growing time to time including conversational AI and ASR. By considering fast growth of technology we propose to develop Amharic conversational AI for advising mothers during their pregnancy. The reason we chose pregnant women is they have to check themselves by going to hospital when they have unusual behavior or symptoms for the safety of the fetus. At the time of COVID19, everything was closed so pregnant women cannot get the follow up they intended to therefore such kind of system may support the pregnant women to treat themselves. The knowledge of few medical specialists can reach several mothers using Conversational AI. Besides, the system may reduce the burdens of physicians and increases the number of mothers who get medical information. Unlike the text based systems, speech based systems can be used by literate and illiterate people

1.3. Statement of the problem

According to (Berhan, 2008) the patient-physician ratio is not balanced. To support the health sector by simplifying their work using artificial intelligent software's are recommended. Specially we can contribute on information delivery, consultation and others to enable customers to get the information and advice based on their requests. This will reduce the burden of physicians. There are no conversational dialogue systems which is developed for Amharic language in this domain. The studies must focus on this area for the sake of reducing the burden of physicians and to enable patients to get the required information's from the system and enable patients to get advice based on their behavioral change. Those changes happen throughout pregnancy periods. Developing conversational system for under resourced languages is challenging. The main challenge is researchers always begin from scratch because the required components like ASR and TTS are not developed well. As (Solomon et al. 2012) those components are done for the fulfilment of some academic graduation only and those works didn't revise for future use. There are different Amharic speech recognizers and Text to speech was developed, but it was being implemented.

According to Ethiopia there is no conversational AI or conversational agents which give medical support through advice using any language of the country (like Amharic, Afaan Oromo etc.). During period of pregnancy there are different changes which will happen on mothers, it includes both physical or behavioral changes. To handle this and to enable pregnant women treat themselves at their home those conversational agents are important. To get advice about their physical and behavioral change they must go to hospital or clinic to contact physicians. Every time if they want an information about their change they have to wait for doctors. But the feeling they had may be able solved with simple treatment at their home.

Speech based conversation with the conversational AI make the human-machine interaction smooth and the conversation in between mimic as like as human-human conversation. The conversational AI developed by (Asimare, 2020) give response about published laws of Ethiopia. In this work the conversation takes place with single turn and during conversation the input is text based only. (Asimare, 2020) recommended integration of Amharic speech recognizer and text-to-speech are appreciated future works. (Seyoum, 2015) design and develop Amharic dialogue system that works in word level .To fill the gaps in those works we design end-to-end speech conversational AI for having spoken dialogue to give advice for pregnant women. They can ask any thing which is related to pregnancy. The speech recognizer of proposed study can understand words, phrases and sentences rather than words only.

1.4. Research questions

Based on the problems we discussed above the proposed study need to answer the questions below.

- How to design and develop end-to-end speech conversation AI for advising pregnant with Amharic language?
- How to develop Amharic ASR which able to recognize words, phrases and sentences?

1.5. Objectives

1.5.1. General objective

The general objective of this study is to design and develop end-to-end task based task-based Amharic conversational AI for advising pregnant women.

1.5.2. Specific objectives

To achieve the general objective of this study the following specific objectives are set

- To find out the required methods, algorithms and components for end-to-end speech conversational AI
- To prepare conversational AI dataset
- To design Amharic Conversational AI systems
- To develop Amharic Conversational AI systems
- To Identify tasks can be performed by the conversational AI
- To Evaluate the performance of Amharic Conversational AI

1.6. Scope and Limitations

The proposed study targeted to develop prototype for speech based dialogue using Amharic language. This study aimed to support physicians by reducing their burden and pregnant women's by giving the required advice based on their interest. This prototype proposed to able to get inputs either in the form of text or speech and it gives the response in both text and speech form. The proposed Amharic speech recognizer able to recognize words, phrases and sentences not only words. In this work the users ask the system and to make the conversation smooth the system ask users to get some information's based on its necessity. And it gives the required response for the user.

This work is considered for pregnant woman's and the provided corpus is based on the collected data about pregnancy. And it doesn't give any diagnosis and other medical treatments for pregnant and other patients. In this study the developed prototype communicates with users in Amharic language only.

1.7. Methods

Literature review

Different research literatures which are related with our domain and others which are important for our study are revised. Revising such literatures are used to identify the gap of works and to understand what they did in their study and different supportive ideas and knowledge's are collected to solve stated problems of our study.

Data Collection

Data is an important input for the study we proposed. To achieve the objective of our study, we prepared our dataset in both text and audio format. Those provided data are collected through interview and reading different literatures related with pregnancy. Then we constructed conversations according to the information we have collected. The reason we need to prepare such conversations is there is no well documented pregnant-doctor conversation in Amharic language. Then those conversations are prepared as a dataset and used in the study.

Tools

Different tools are used throughout our study those tools are used for different purposes. Python is a programming language used for writing scripts in our study. Smart audio recorder is mobile software used for recording audio with mobile device and a computer application wave pad is used for setting the audios frequency rate as we required. Samsung A2 hardware is used during recording audios.

1.8. Significance of the study

Conducting this study will reduce the work load and the number of physicians required in hospitals and clinics for giving advice for pregnant women's. At the current time everything is changing into digital or computerized way. So the communication between human and computers need to be in speech based than text based (Grover, 2008). This makes users more confident to ask, but in the developed system users can use both text and speech for interaction with the system. Both illiterate and literate pregnant women's able to get advice equally. Pregnant women's able to ask any questions what she have to do during her pregnancy period without fear and she can get the required information and advice for all feelings she had and all behavioral changes that will happen because pregnancy. The study enables each pregnant mother to treat themselves at home and they can get information and awareness for unusual feelings they didn't feel before.

It also minimizes the cost to hire physicians. For pregnant's it saves transportation cost they spent to communicate and consult physicians because pregnant's doesn't have to go to hospital for each and every symptoms of pregnancy because for some symptoms or changes they can treat themselves using the developed system.

1.9. Organization of the thesis

The remaining part of this work is organized as follows.

Chapter two presents

Chapter two presents literature review about conversation al agent and its application to healthcare domain.

Chapter three describes about the methodology employed. In this chapter, how dataset is prepared, tools used and the overall system architecture is presented. The evaluation technique used in the study is also addressed.

Chapter four discuss the experiment we conducted and the results we get from the experiment.

In Chapter five concluding remarks are presented and future direction is presented

Chapter 2: Literature Review

2.1. Introduction

In this chapter there are different concepts which are discussed and literatures related with conversational AI are reviewed. The application and different domains that conversational AI's can be implemented also described. Some of techniques that can be used for conversational AI are discussed.

2.2. Overview of conversational agents

The emergence of new technologies make Conversational agents become more dynamic and the classification of chatbot's has become subjective to the scope of their use. Conversational agents or chatbot's can be classified based on different criteria's for example the knowledge domain, mode of interaction, their usage and the design techniques (response generation method) that are typically employed in building these chatbot's (Hussain, Sianaki, & Ababneh, 2019).

The broad classification of conversational agents is listed below.

- Interaction Mode
- Chatbot Application
- knowledge domain (Domain-Specific or Open-Domain)
- Rule -based or AI

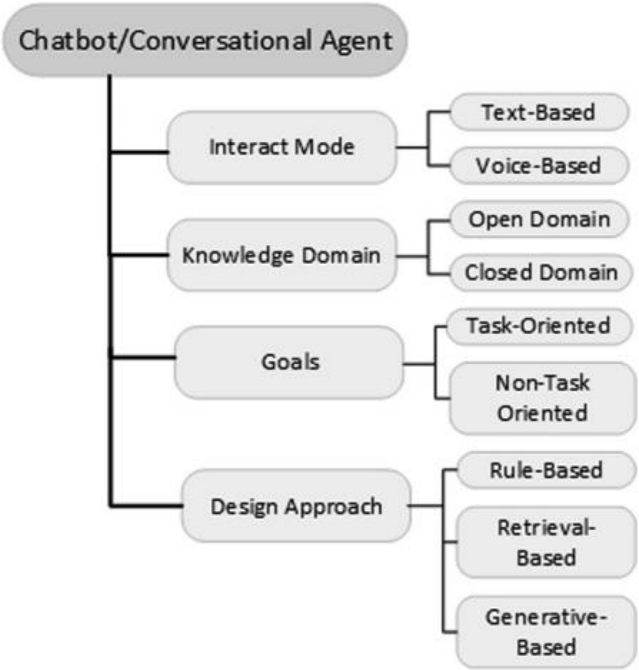


Figure 1 Broad classification of Chatbot's (Hussain et al., 2019)

Chatbot's are classified in to two main classification based on the goals. Which are task oriented and non-task oriented.

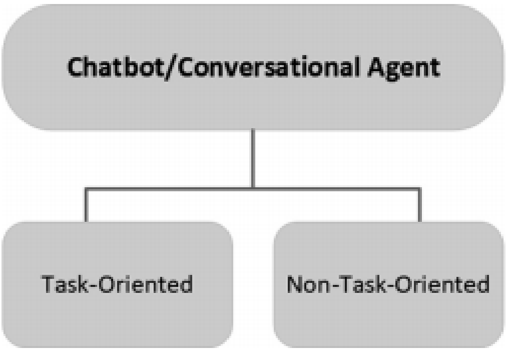


Figure 2 Main classification of chatbot's based on their goal (Hussain et al., 2019)

Task-Oriented conversational agents are a significant conversational agent designed for dealing with specific scenarios. Conversational agents which are intended to perform specific task is called task oriented conversational AI. They give an answer based on the knowledge they have in specific domain they designed for (Budulan, 2018). Those chatbot's are focus only in restricted domain and they didn't have general knowledge. You cannot ask them any trivial questions because they focus on the specific domain and they are targeted to solve user's problem in the specific domain**.** Some tasks are booking a hotel/flight, scheduling an event, or helping users to access some specific information, placing an order for a product, etc. Personal assistants such as Alexa, Siri and Cortana are examples of voice based task-oriented Chatbot/conversational agents who attempt to return an answer for the task they receive. Task-oriented conversational agents answer users question, give information's and support users based on the knowledge they have in that specific domain or task (Hussain et al., 2019)**.**

Non-task-oriented conversational agents are designed for not only specific domain or task. Which enables to have an extended conversation without boundary limit in specific tasks. The conversation with such kind of chatbot's are unstructured those conversational agents are set up to mimic the conversation as human-human without any limitation. Such type of conversational agents has big entertainment value.

Conversational agents are classified as open domain and close-domain based on their knowledge domain (Hussain et al., 2019). Those conversational agents destined to retrieve all sort of questions raise from users with no restriction, some of the questions can be "What year was Salvador Dali born in?", "How many species of frogs exist worldwide?", "What will the weather be like in three days?". Conversational agents which have the ability to respond questions which belongs to in different topics such as general knowledge's which is called open domain conversational agent. Close-domain conversational agents which operates based on the information in the specific domain to answer users question. Such kind of conversational agents are provided for narrow scenarios for example to guide tourists in the museum by giving the required information's. generally those conversational agents perform well in the real environments and they are serving well for users (Budulan, 2018).

According to the way of interaction conversational agents are categorized as text base and voice based. Spoken based conversational AI get input in the form of speech and after processing it gives response for the user. Text base conversational AI it uses text to interact with the users. Users ask the system using text and after processing the input it respond for them (Mhatre, Motani, Shah, & Mali, 2016).

During conversation with conversational agents when the system begins the conversation in between is called system initiative and when the user begins the conversation the system is called user initiative. But if both the system and users able to initiate the conversation is called mixed-initiative (Jurafsky & Martin., 2017).

2.3. Application of conversational agents

The basic application of conversational agents is giving service for users as required by replacing human beings based on the way as peoples want. Users communicate with those conversational AI and they accept users request to perform some tasks. Those conversational AI can be used in the range of business tasks. Those tasks can be ticket boking, requesting hotels and restaurants, asking advice or some other tasks. Users of the conversational agents can be communicating with it using text, speech’s or both of them. This used to imitate the conversation like human-human. The main target which is considered during designing and developing conversational AI’s is simplifying people’s life by making information’s available for peoples in their day to day life’s. Some of their applications are described below.

2.3.1. Application of conversational agent for cultural heritage

To present the cultures and heritages of the country conversational agents play vital roles. Conversational agents enable to searching and present results effectively and efficiently. They can give the required information about the heritages in the form of smother conversation between tourists or anybody who want to know about and the system using natural languages. The communication can be using text or speeches (Machidon, Tavčar, Gams, & Duguleană, 2020).

2.3.2. Application of conversational agent for online market

At this time buying and selling goods are performed online using different websites. Customers need to visit different e-commerce websites like Alibaba, Amazon and eBay. But to perform their task they need to do it in the smooth and interactive way rather than the common way as we know. To achieve the interest of users integrating conversational agents for marketing is much important. Those conversational agents in the market act as salesperson to help companies to advertise products (Pradana, Sing, Kumar, & Applications, 2017).

2.3.3. Application of conversational agent in healthcare

Conversational agents have great contribution in health care. During conversation with conversational AI users feed their symptoms and feelings, by analyzing user input the system generates the appropriate response and solutions from the medical knowledge it learns at the time of training. Those kind of systems enable to conduct diagnosis and giving advice for patients. Additionally, they can suggest you the nearest clinic and gives you an appointment (Gentsch, 2019). By considering such kind of conversational agent's functionality we intend to design and develop a conversational system which used for advising pregnant's that have the ability to communicate with users using speech and text in Amharic language.

2.3.4. Application of conversational agent for education

Conversational agents support students and teachers to achieve their goals. Those agents help to manage learning resources and support students. In distance learning environment, agents can improve the motivation and concentration of the students and, consequently, they can improve the effectiveness of the educational process. They allow to tech students in different forms like question and answering, gamming and others (Landowska, 2010). The telegram bot which have words for learning one of example of applications of conversational agents for education (Ankit Kumar O. I., 2016).

2.3.5. Application of conversational agent for psychologist

Conversational AI systems act as psychologist or give psychological treatment for the person by having conversation. It acts like your own real friend. It helps customers to led healthy life by discussing the conversational AI they can get important information's and they able to discuss on important topics that are important in person's day to day life.

2.3.6. Application of conversational agent as customer service support

Peoples need some help to complete their intended tasks. those conversational agents offer their help for customers to solve their problems simply. The way they help customers can be by giving relevant information, responding customer questions, providing the required assistant for them and they give information about updated service and products. Question and answering help customers by giving information about what they asked they advertise new services (Kuramoto et al., 2018).

2.4. Techniques for designing conversational agent or Chatbot

To develop task-oriented and non-task oriented chatbot's there are different approaches. In this subsection we explore some of those approaches. There are three main categories in which these approaches can be divided (Hussain et al., 2019).

I. Rule-based approach

Before the emergence of artificial intelligence and machine learning conversational systems were developed using rule-based approaches. Rule-based conversational systems generate pre-determined answer for the given questions. In the rule –based approach there is a framework developed by developers to host a logic for word processing, guiding users through developer defined decision making path or decision tree. The early ELIZA and PARRY systems used rule-based Retrieval-based approach. The logic behind giving response for incoming requests is there are number of implemented rules. The input message or request corresponding to certain pattern then script based response will be generated. Rule-based Conversational systems take an action according to the conditions or patterns implemented. Those conversational systems don't give AI capabilities for parsing (syntax analysis) and classifying the incoming request. In the rule-based conversational systems developers code manually for message processing and interaction logic. In a typical conversation to map different dialogue states hundreds of rules required to be defined (Hussain et al., 2019).

II. Retrieval-based approach

Retrieval-based approaches work on the principle of graphs or directed flows. The conversational AI trained with predefined responses but it generates the appropriate response based on users input. Keyword matching, machine learning or deep learning are used techniques in retrieval-based conversational AI for identifying the most appropriate response. Regardless of techniques it used those retrieval-based conversational AI didn't generate new output but it responds predefined response according to the existing information the system gets from user's (Hussain et al., 2019).

III. Generative-based approach

Generative-based conversational AI can generate new response based on the training conversational data. It can dynamically create responses in real-time unlike retrieval-based conversational AI. Techniques used in those conversational AI are a combination of reinforcement, learning unsupervised learning, supervised learning and adversarial learning for multi-step learning.

Conversations in Supervised learning is structured as a sequence-to-sequence problem. User inputs in sequence-to-sequence learning mapped to a computer-generated response. However, this type of learning has a tendency to prioritize high-probability responses.

Supervised learning systems also have trouble including proper nouns into their speech since they appear less in dialogue compared to other words. As a supervised learning chatbot's sound repetitive and cannot promote stable human conversation. To address this issue, developers leverage reinforcement learning to teach chatbot's how to optimize dialogues for some cumulative reward (Hussain et al., 2019).

There are multiple techniques which are implemented by the above approaches. Those are discussed below.

1. Parsing

Parsing is a technique used to analyze the given input text and manipulating it by using a number of natural language processing functions like trees in python NLTK. It is a method of extracting meaningful information's from the given input text by converting that text into a set of more uncomplicated words. Those words can be stored and manipulated easily. Which used to determine the grammatical structure of a given sentence. An earlier Chatbot ELIZA used a simplistic parsing technique to parse the input text for keywords in the sentence. To find the appropriate response or answer for the user keywords matched against to the provided corpus (Weizenbaum, 1966). In the advancement of technology semantic parsing is able to convert users input sentence to machine-understandable representation of its meaning. For example, Dialogue flow is a commercial Chatbot.

2. Pattern matching

Which is commonly used technique in the development of conversational Chatbot in which user's inputs classify as a pattern and provide appropriate response for user input from the provided template. The <pattern> <template> pairs are handcrafted and it is provided based on the domain we focus. This technique is used in both early and modern Chatbot's and the complexity of the algorithm differ from one Chatbot to the other. ALICE use more complex pattern matching rule and when searching the stored categories for a matching pattern it associates some degree of conversational context. The early Chatbot ELIZA uses simplest pattern matching rule than ALICE (Wallace, 2009). This technique has the advantage of flexibility to create conversations. Depending on matching types like natural language enquiries, semantic meaning of enquiries or simple statements the pattern matching technique used commonly in question-answering systems.

3. AIML

To construct Chatbot we need flexible, easy to understand and universal languages. AIML satisfies those requirements and it is widely used technique. AIML stands for Artificial Intelligence Mark-up Language which is designed to create conversational flow in chatbot's. It is derived from XML (Extensible Mark-up Language) and open standard. It is powerful for designing chatbot's conversational flow, easy to use and flexible but it needs NLP (natural language processing) and programming expertise (Marietto et al., 2013).AIML based on the technologies of A.L.I.C.E (Artificial Intelligent Computer Entity) and it represents knowledge put into Chatbot. Data objects used to make AIML are called AIML objects. Those objects are called AIML elements. AINL has the ability to characterize those data objects. Units that AIML objects made up of are called topic and categories. The topic has a name attribute and set of categories related to that topic and it is an optional top-level element. Categories are the fundamental unit of knowledge in an AIML file. Parsed or unparsed data's can be contained under categories. Minimum of two or more elements are included in the category which are pattern and template. Each category contains rules for matching an input and converting it to an output. The pattern element under categories matches against the user input and the 'template' is used to generate the Chatbot's answer (Hussain et al., 2019).

The AIML purpose is to simplify job of conversational modeling in "stimulus-response generation" relation process. It is XML based markup language and depends on tags which are the identifiers that make snippets of codes to send commands into the Chatbot. Modeling conversational patterns is the responsibility of data objects (AIML objects). Each data objects are a language tag that associates with language command. The general structure of data objects as putted by (Marietto et al., 2013)

<command> List of parameters </command>

category, pattern, and template are most important object among the AIML objects category, pattern, and template. Defining the knowledge unit of the conversation is task of category tag and identifying user input is task of the pattern tag. The task of template tag is to respond to specific user input. Category, pattern and templates are the most frequently used tags and those are the bases to design AIML chatbot’s. the structure of those tags are as shown below:

```
<category>
    <pattern> User Input</pattern>
            <template>
                Corresponding Response to input
             </template>
</category>
```

4. Chatscript

Chatscript is an authoring tool to build conversational chatbot’s and accessible as open source, it is a combination of both natural language and dialogue management system. Chatscript technique used when no matches are occurred in the AIML. To build a sensible default answers it focuses on best syntax. It designed for interactive conversation while maintaining user state across the conversations. It is rule based engine and it use dialogue flow script which is a program script used to create rules through process. Scripts stored as a normal text file. For mining user’s conversations logs Machine learning tools can be used which improve the dialogue flow. Chatscript make use of concepts. A concept of all nouns or adverbs can be created (Hussain et al., 2019).the main functionalities of Chatscript are a variable concept, facts, logical and/or.

5. Ontologies

To replace handcrafted domain knowledge with ontological domain knowledge in Chatbot domain ontologies are used. They have been used in specific dialogue system modules which shows use of ontologies is not a new thing, for example it was implemented as the basis of the systemic grammar approach in language generation (Milward & Beveridge, 2003). Using ontologies in Chatbot used to establish relationships between concepts that are used in conversation by exploring concept node of an ontologies and it can also indicate new reasoning (Al-Zubaide & Issa, 2011).

6. Markov Chain Model

Markov chain Model is popular model to create Chatbot. Markov chain is probabilistic model that means it uses probabilities for the rules. It seeks to model probabilities of state transition over time. Markov chain model used to create Chatbot for entertainment purpose to emulate simple conversation as humans rather than complex conversations and it is easy to program. The main key idea of this model is current state, one or more states have fixed probabilities that will go to the next. Chatbot's that implement the Markov chain model able to produce an output that that goes with the state transition. It allows to for constructing responses that are more suitable probabilistically and responses are being different every time but they are more or less coherent (Ramesh, Ravishankaran, Joshi, & Chandrasekaran, 2017).

7. Artificial Neural Networks Models

In the advancement of machine learning specially in artificial intelligence more intelligent bots are developed for different tasks. Those bots are more intelligent than the previous bots. In terms of the response generation Artificial Neural Network bots can use retrieval-based or generative base approaches the current trend of research is shifted in to generative based approaches (Nuez Ezquerra, 2018). There is a difference between rule-based and artificial neural network chatbot's. The main difference is the existence of learning algorithm in artificial neural network. One of the learning algorithms is artificial neural network models which is used in machine learning. Both supervised and unsupervised machine learning algorithms are kind of learning algorithms. Deep learning has the capability to understand and learn from unstructured data. In terms of processing data and creating patterns which is valuable for decision making. We can use artificial neural networks for different type of tasks, some of those tasks are decision making, machine translation, computer vision, social network **fi**ltering, medical diagnosis and speech recognition etc. For natural language processing different artificial neural networks can be used.

7.1. Recurrent Neural Network (RNN)

Recurrent neural network (RNN) initially created in 1980's. It is type of most powerful and robust neural network. This recurrent neural network derived from feed forward networks and shows behavior of how human brain works. RNN is the most popular class of artificial neural network and it is the most appropriate model for natural Language processing tasks. which is a preferable algorithm for sequential data such as text, speech, time series, video, audio, weather and financial data and much more. Recurrent Neural Networks (RNN) works by saving the output of the layer and the saved output is feed to the input for predicting the next output of the layer. RNN has the ability to remember the previous information and that information will be used to perform and process the next tasks (Young, Hazarika, Poria, & Cambria, 2018).

The advantage of RNN is it uses sequential information's but the traditional neural networks are independent and they have not ability to use sequential information's. For example, to predict the coming word in the particular sentence it is better to now the previous one otherwise it is difficult to predict. This approach being suitable for Chatbot to get the context and give the appropriate answer based on the previous word in the sentence (Cahn, Computer, & Science, 2017). RNN have memory to capture the required information that has been calculated in the process. It can use information in long sequences without limitation but it has limitations to looking back except few steps only (BRITZ, 2015).

To see how RNN works clearly we need to have knowledge of sequential data and how feed forward neural network works.

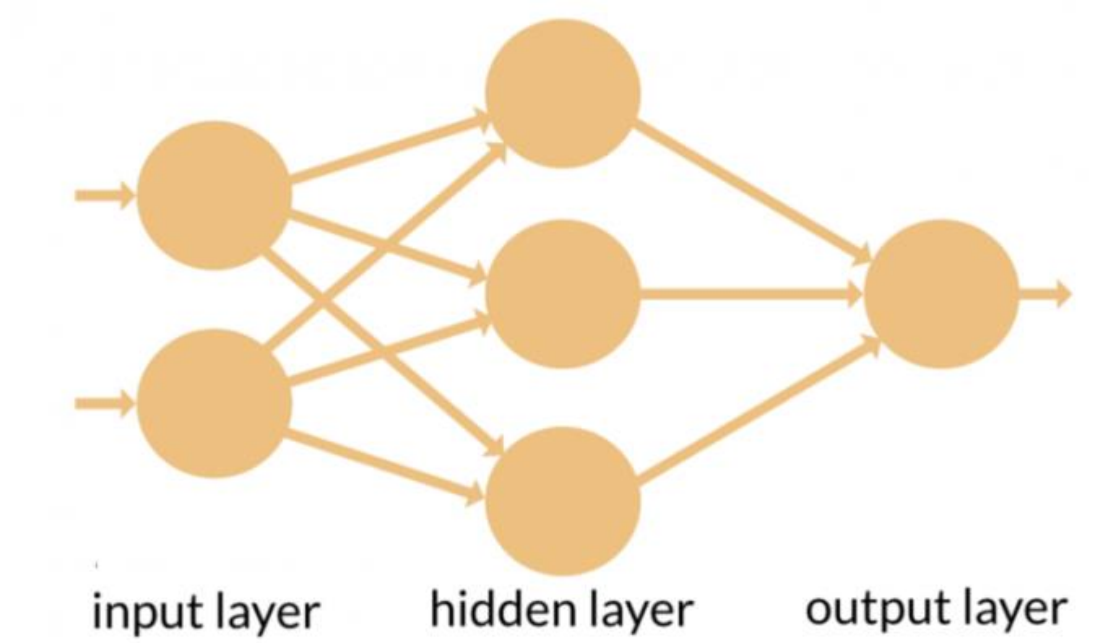


Figure 3 Feed-forward neural network (Donges, June 16 2019)

The name of both RNN and feed-forward networks given from the way they channel information.

Information move in one direction only in the feed-forward recurrent neural network which is from the input layer, through the hidden layer to the output layer. The input information moves straight through the network and never touches a node twice. This feed forward neural network is bad for predicting what will come next and they have no memory for the input they receive. The feed-forward neural network has no notion of order in time and it only consider the current the current input. That means it doesn’t remember what happened in the past except its training.

The current input and the learned information’s from the input it received previously are consider when making decisions in RNN. Information in the RNN cycles through a loop.

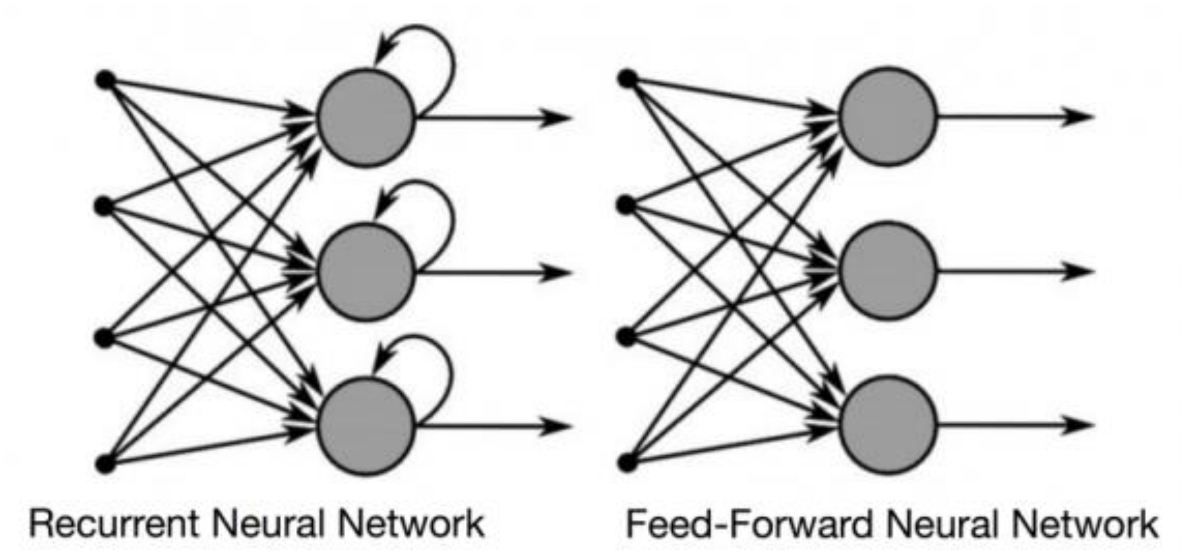


Figure 4 The difference of information flow in both RNN and feed-forward neural network (Donges, June 16 2019)

Recurrent neural network has short-term memory. Those memories used to handle required information about the past for predicting the next. For example, imagine we gave the word "neuron" as an input for the feed-forward neural network and it process the given input character by character. By the time when it reaches the character ‘r’ it has already forgotten about the past two characters “n”, “e” and “u”. this makes difficult for predicting which character would come next. But, because of its internal memory the RNN able to remember those characters. The RNN produce an output, the output would be copied and loops it back into the network.

In the feed-forward neural network for the input a weight matrix would be assigned and then the output produced. RNN apply weights for both current and previous input. It also tweaks the weights for both through gradient descent and backpropagation through time (BPTT). The feed-forward neural networks map one input to one output, but the RNNs can map one to many, many to one (classifying a voice) and many to many (translation) (Donges, June 16 2019).

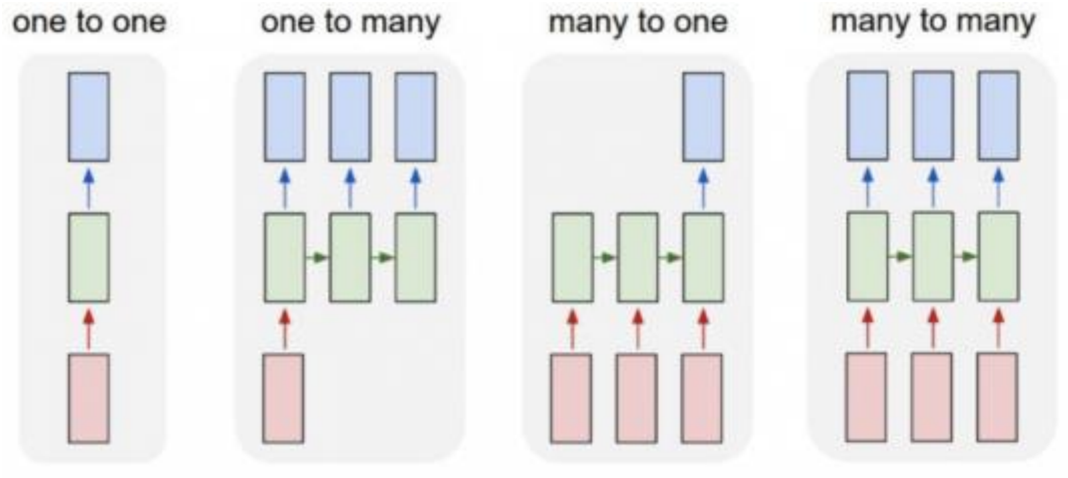


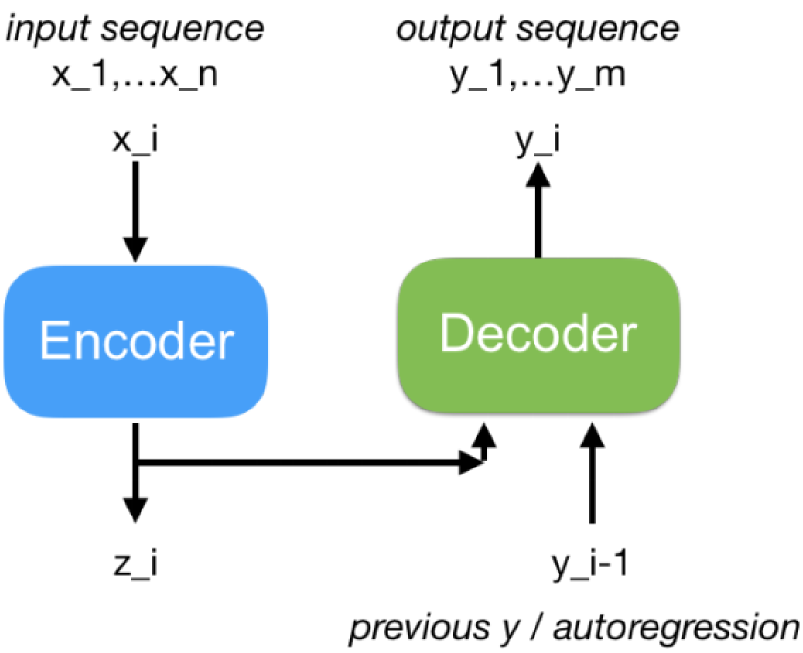
Figure 5 In RNN one input can be mapped to one to one, one to many, many to one and many to many (Donges, June 16 2019)

7.2. Sequence to Sequence (Seq2Seq) Neural Model

Seq2Seq models are popular models used to translate sequence of symbols from one input sequence to an output sequence. Those models are used on different tasks and those models also used to solve complex language problems like video captioning, machine translation, question answering, image captioning etc. Those models are based on RNN architectures and they have both encoder and decoders. Encoders used to process the input and decoders used to process outputs (Cahn et al., 2017).

The figure below shows encoder-decoder neural network architecture for sequence-to-sequence learning. An input sequence Xi is encoded into a sequence vector Zi by the encoder. The decoder produces an output sequence Yi which is auto-regressive (Culurciello, April 16 2019)

.



7.3. Long Short-Term Memory Networks (LSTM)

LSTM is the special kind of RNN and it overcome the long term dependency problem of RNN (Ramesh et al., 2017). Unlike RNN LSTM has the ability to learn long term dependencies. To able to remember information's for long period of time LSTM introduces memory cells and gets. In these memory cells information's can be stored in, write and read from. It uses input gates, forget gates and output gates to make the flow of information controlled (Sak, Senior, & Beaufays, 2014). A well trained LSTM has big capability to perform classification, processing and prediction of time series than RNNN, hidden Markov models, and other sequence learning methods**.** Due to this LSTM is being used in designing Chatbot because it has ability to refer to a piece of distant information in time very frequently (Ramesh et al., 2017).

In the LSTM mechanism of information flow known as cell state. In this LSTMs remember or forget things selectively. The information at a particular cell state has three different dependencies. Those dependencies generalized as:

- The previous cell state (i.e. the information that was present in the memory after the previous time step)
- The previous hidden state (i.e. this is the same as the output of the previous cell)

The input at the current time step (i.e. the new information that is being fed in at that moment) (Strivastava, December 2017).

-

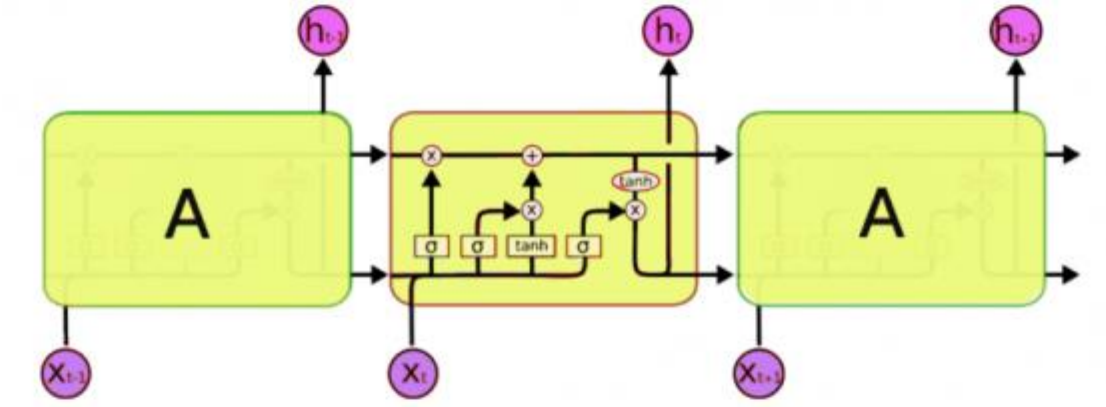


Figure 6 Architecture of LSTM (Strivastava, December 2017)

There are different memory blocks comprised in a typical LSTM network which is called cell. Both cell state and the hidden state are two states that are being transferred to the next cell. In this Remembering things is the responsibility of memory blocks and manipulations to this memory is done through three major mechanisms, called gates (those are Forget Gate, Input Gate and Output Gate).

Forget Gate

Removing information from the cell state is the responsibility of forget gate. The information that is less important or the information that is no longer required for the LSTM to understand things is removed via multiplication of a filter. It is required for LSTM Network to be optimized.

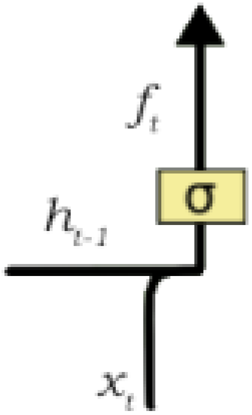


Figure 7 Forget Gate Structure (Strivastava, December 2017)

h_t-1 and x_t are two inputs taken by forget gates. h_t-1 is the output of the previous cell or the hidden state from the previous cell and x_t is the input at that particular time step. The given inputs are multiplied by the weight matrices and a bias is added. Following this, the sigmoid function is applied to this value. The output generated by sigmoid function is vector ranging from 0 to 1 corresponding to each number in the cell state. Deciding which value to be keep and discarded is the responsibility of sigmoid function. For particular value in the cell state if the output is “0” which means the forget gate wants the cell state to forget that piece of information completely. In the same way if the output is “1” it indicates the forget gate wants to remember that entire piece of information. This vector output from the sigmoid function is multiplied to the cell state (Strivastava, December 2017).

Input Gate

For the addition of information to the cell state the input gate is responsible. This addition of information is basically three-step process.

- ➢ Regulating what values need to be added to the cell state by involving a sigmoid function. It is basically very similar to the forget gate and acts as a filter for all the information from h_t-1 and x_t.

- Creating a vector containing all possible values that can be added (as perceived from h_t-1 and x_t) to the cell state. This is done using the tanh function, which outputs values from -1 to +1.
- Multiplying the value of the regulatory filter or the sigmoid gate to the created vector (the tanh function) and then this useful information added to the cell state via addition operation (Strivastava, December 2017)

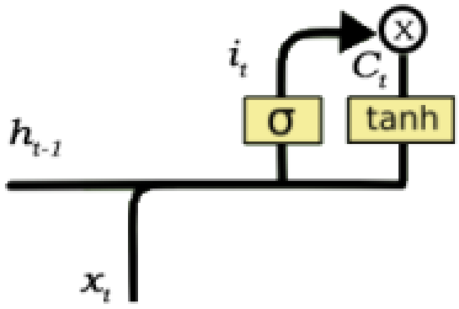


Figure 8 Input Gate Structure (Strivastava, December 2017)

Output Gate

Selecting useful information from the current cell state and showing it out as an output is done via the output gate. Here is its structure:

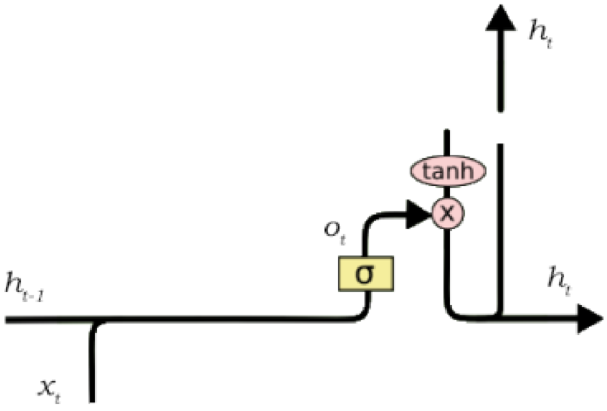


Figure 9 Output Gate Structure (Strivastava, December 2017*)*

The functioning of an output gate can again be broken down to three steps

Step-1. After applying tanh function to the cell state it creates vector, thereby scaling the values to the range -1 to +1.

Step-2. By using the values of h_t-1 and x_t it makes a filter. Then the value need to be output from vector created above can be regulated. This filter again employs a sigmoid function.

Step-3. Multiplying the value of this regulatory filter to the vector created in step 1, and sending it out as an output and also to the hidden state of the next cell (Strivastava, December 2017).

2.5. Related Works

In the real world conversational agents are used to solve problems and simplify peoples' life. As we discuss above conversational AI have the capability to accomplish its intended task based on the knowledge it gets from training. Task-oriented conversational agents focusing on a single domain and which is different from non-task oriented conversational AI according to the data it takes and processing management.

In the healthcare domain task-oriented conversational system was developed for automatic diagnosis (Wei et al., 2018). In which the system able to make automatic diagnosis through conversation with the user. To train the system they build the required dataset from an online medical forum by extracting symptoms from both conversational data between patient and doctor and patients self-reports. The dataset is collected from a popular website which is called Chinese online health care community from pediatric department. It used for users to inquire with doctors online. In this web site users provide self-report about his/her condition. Then the doctor initialize conversation with the patient to collect more related and important information's with user's problem. Based on the conversational data and self-reported data the diagnosis would be performed. Four type of diseases are chosen for annotation those re children functional dyspepsia, children's bronchitis, infantile diarrhea and upper respiratory infection.

The study conducted for identifying the performance and ability of smartphone based conversational agents for answering questions about mental health, interpersonal violence and physical health (Miner et al., 2016). In this study different smartphone based conversational agents are tested such as Sir (Apple), Cortana (Microsoft), Google Now and S Voice (Samsung). In this pilot study they use 68 phone samples from 7 manufacturers and it is conducted from December 2015 to January 2016. Investigators asked for the smartphones 9 questions, 3 each in mental health, physical health and interpersonal violence in their natural language. The conversational agents asked questions repeatedly until no new answer. The responses generated from those conversational agents characterized based on -1- recognize crisis, -2- responding respectful natural language, and -3- appropriate helpline. When the conversational agents asked simple question about mental health, physical health and interpersonal violence S Voice, Siri, Cortana and Google Now responded inconsistently and incompletely. This shows to get full and effective response improving the performance of those conversational agents required. By considering such problems, we focused on providing an end-to-end speech conversational framework for specific domain. Which help to provide appropriate and required data for conversational systems. It enables them to respond users request appropriately and effectively.

There are three main components are included in this work those are Natural language understanding (NLU) for detecting user's intent and slots with the value from utterance, dialogue manager (DM) which tracks the dialogue state and take system actions and natural language generation (NLG) generates natural language given the system actions. The conversation between the doctor and the patient takes place with text, because it doesn't include speech recognizer and text-to-speech synthesizer (TTS). The proposed conversational framework provided for Chinese language only it doesn't support other languages including Amharic.

The work done before by (Seyoum, 2015) is an end to end speech Amharic spoken dialogue system. The developed prototype focused on hotel and restaurants for giving information about location of restaurants and type of foods the hotels serve. In this study an end-to-end

Amharic spoken dialogue system prototype was developed. But the speech recognizer able to recognize Amharic words only. Using words only during conversation is not relevant because peoples can’t express well about what they want. To solve this problem, the proposed system, provide a good Amharic speech recognizer which can able to recognize words, phrases and sentences.

There is study conducted by (Asimare, 2020), which is done on conversational system for consulting Ethiopian published laws. The system gives the required response for users according to their question. Users use this system to get consultation and to know what Ethiopian constitution stated about the issues they want to know. In this study mode of communication is text, the conversation between the system and the user takes place using text. The user asks the system in the form of text then the system respond in the form of text too. So to make the conversation between users and the system as like human-human the proposed system design and implement an end-to-end speech conversational AI prototype. The proposed proto type support dual mode of interaction which is text and speech and it support multi turns during conversation no single turn like as (Asimare, 2020). with the proposed prototype users able to talk to the system using speech input and the system responds to the user in the form of speech.

Chapter 3: Methodology

3.1. Introduction

The chapter discuss and describe in detail about methods, tools we used in this study and the architecture of the proposed study. We put clearly how the datasets we used for training and testing Amharic ASR and the conversational system are prepared and how the conversational system works from getting an input to generating an output from and to users.

3.2. Data collection

In this study we collect the required data from different sources. The data collection is performed for single domain which is about pregnant women's. Making an interview was the mechanism used for getting the required data. Six pregnant's are selected randomly and we make an interview to get an information about the feeling they had, symptoms seen, their behavioral changes, their way of feeding and others. Then we organized responses of pregnant's. To validate responses of pregnant's we cross check with different literatures and mobile applications. Those literatures and mobile applications described about pregnancy from the beginning up to the end of pregnancy period. In these what pregnant women's have to do and don't, what they feel, different things that will happen during pregnancy period and others are discussed in detail. Mobile applications used are provided in Amharic Language. Generally, literatures and mobile applications give more information about pregnancy to enable peoples to get awareness about it.

Additionally, we communicate some peoples who are physician and professionals in nutrition to ask and understand about the general behaviors of mothers during pregnancy and way of feeding during pregnancy period. Consulting and asking professionals help to understand the real feeling and behaviors of pregnant women's. Which is additional input to understand the whole thing that will happen on pregnant women's.

3.3. Dataset (Speech Corpus and Text Corpus)

There are datasets provided for the proposed study. In our study two datasets are provided based on the domain we focus and the task need to be performed. The first one is for Amharic speech recognizer and the second one is for developed conversational system. For end-to-end speech based conversational AI ASR is the main component of the system, because the communication with the system takes place through speech. For that Amharic ASR is developed according to health care domain for pregnant's only. To do this we have to prepare our dataset or corpus based on our domain. For Amharic ASR we used both text data and speech data in "*.wav*" format. The text data is prepared according to the information we get from the interview and reviewed literatures during data collection. The total number of text data is 560. Then the audio data is provided by recording the text data. Recording is not performed under controlled environment. We recorded the audio at home in any time of the day by selecting single women randomly at the age of 19 and reading the provided text. The frequency rate of the recorded audio is 16KHz.

For the conversational AI the dataset is provided in the form of text. According to the appropriateness and relation of sentences in the intents the dataset is labeled in the form of pattern and response. There are intents provided from the collected data. Each intent has their own tag, pattern and response. For the provided intents tags used as an ID to identify each intent uniquely from the other intents. Patterns are sentences which is expected questions or conversations that the user will raise or enter it as input to the system and they can be one or more in a single intent but they are related. The other one is response, which is the provided response for the given pattern of the intent. The dataset is saved as JavaScript Object notation (JSON) format. All of the pattern-response combinations are relevant according to the literatures about pregnancy. From this data provided a machine learning model would be created which used for classifying the input from the user.

3.4. Tools

In our study there are tools we used to construct the proposed system. The programming language that we use to write the code is python programming. Python programming language is the most powerful language and it is easy to learn and implement. It is preferable for data science and there are different packages and libraries that we can import and integrate easily based on our interest into python. At the current time for machine learning related works python programming is most preferable. Machine learning Libraries that we use in this work are Tensorflow, TFLearn and nltk. TFLearn is built on top of Tensorflow and which is high-level API it enables building and training of neural network fast and easy. The other one is Tensorflow which is an open source end to end platform for machine learning. It is comprehensive, flexible ecosystem of tools, libraries and community resources. The nltk toolkit used for tokenizing the input text or the transcribed speech in to list of words.

For building Amharic ASR we use Poketsphinx, Sphinxbase and Sphinxtrain by implementing some instructions for creating acoustic model, training and testing. Poketsphinx is most advantageous for recognizing continuous speeches.

3.5. Architecture of proposed prototype

The architecture of proposed system shows the whole process of end-to-end speech conversational AI and it includes different components like Amharic ASR and TTS. The conversation between the system and the user performed both in speech and text form. Amharic speech recognizer is responsible for transcribing the input speech into text. Then the conversational AI process the transcribed text and generate the appropriate response for user input. The text to speech synthesizer convert the generated text to speech form. Finally, the user gets the speech output for his/her input. But the conversational AI able to get input and generate output in the form of text too. This indicates it has dual mode of interaction which are text and speech.

The figure below shows the acoustic model for Amharic speech recognizer generated from the training audio data.

.

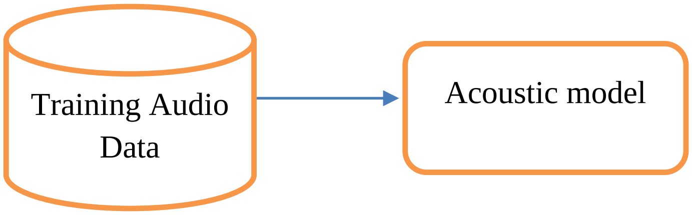


Figure 10 Acoustic model created from audio data

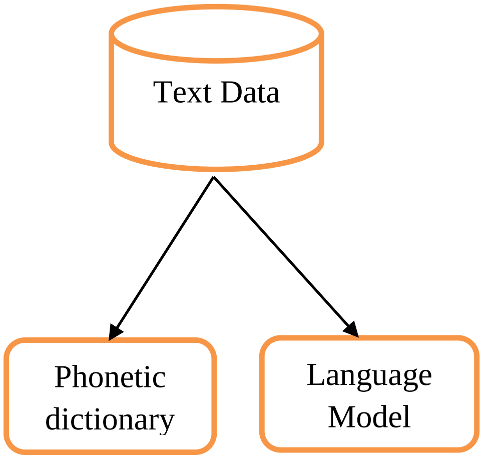


Figure 11 Phonetic dictionary and Language model created from text

Figure 11 shows both the language model and phonetic dictionary generated from the provided training text data. The generated language model identifies which word follow after previously identified word and phonetic dictionary which is word to phone mapped dictionary from text data.

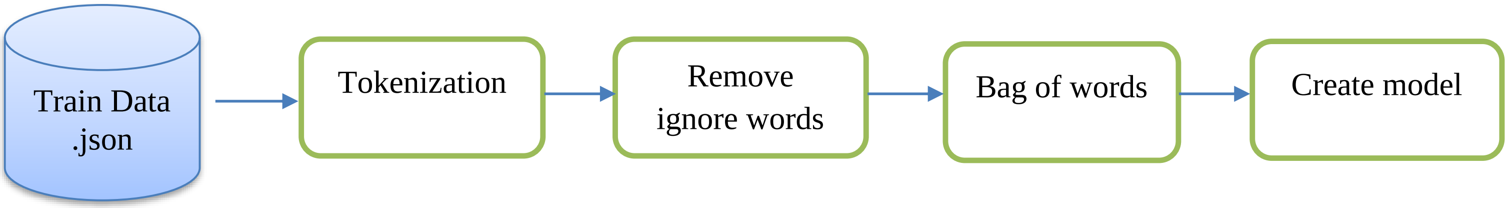
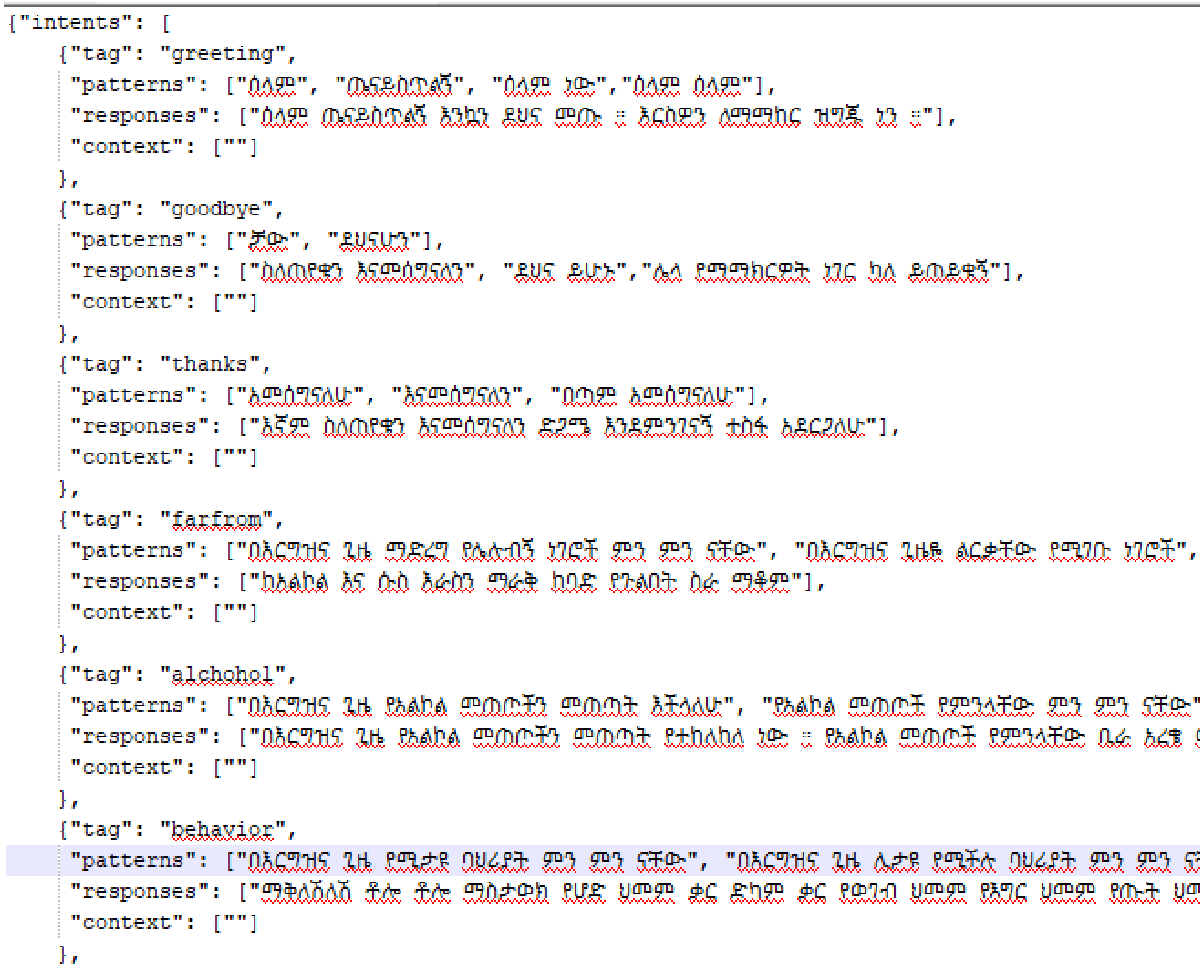


Figure 12 Generating model from intents for classifying user input

The created model generated from the provided intents to train the conversational AI through the process of tokenization, ignore word removing and generating bag of words.

The created model contains the whole information of which intent belongs to in which class and which response would be appropriate for each pattern. This model would be used for classifying user input during conversation. The process of creating the model shown in the figure below.



```
{"intents": [
    {"tag": "greeting",
     "patterns": ["ሰላም", "ጤናይስጥልኝ", "ሰላም ነው","ሰላም ሰላም"],
     "responses": ["ሰላም ጤናይስጥልኝ እንኳን ደህና መጡ ። እርስዎን ለማማከር ዝግጁ ነን ።"],
     "context": [""]
    },
    {"tag": "goodbye",
     "patterns": ["ቻው", "ደህናሁን"],
     "responses": ["ስለጠየቁን እናመሰግናለን", "ደህና ይሁኑ","ሌላ የማማክርዎት ነገር ካለ ይጠይቁኝ"],
     "context": [""]
    },
    {"tag": "thanks",
     "patterns": ["አመሰግናለሁ", "እናመሰግናለን", "በጣም አመሰግናለሁ"],
     "responses": ["እኛም ስለጠየቁን እናመሰግናለን ድጋሜ እንደምንገናኝ ተስፋ አደርጋለሁ"],
     "context": [""]
    },
    {"tag": "farfrom",
     "patterns": ["በእርግዝና ጊዜ ማድረግ የሌለብኝ ነገሮች ምን ምን ናቸው", "በእርግዝና ጊዜዬ ልርቃቸው የሚገቡ ነገሮች",
     "responses": ["ከአልኮል እና ሱስ እራስን ማራቅ ከባድ የጉልበት ስራ ማቆም"],
     "context": [""]
    },
    {"tag": "alchohol",
     "patterns": ["በእርግዝና ጊዜ የአልኮል መጠጦችን መጠጣት እችላለሁ", "የአልኮል መጠጦች የምንላቸው ምን ምን ናቸው"
     "responses": ["በእርግዝና ጊዜ የአልኮል መጠጦችን መጠጣት የተከለከለ ነው ። የአልኮል መጠጦች የምንላቸው ቢራ አረቄ
     "context": [""]
    },
    {"tag": "behavior",
     "patterns": ["በእርግዝና ጊዜ የሚታዩ ባህሪያት ምን ምን ናቸው", "በእርግዝና ጊዜ ሊታዩ የሚችሉ ባህሪያት ምን ምን
     "responses": ["ማቅለሽለሽ ቶሎ ቶሎ ማስታወክ የሆድ ህመም ቁር ድካም ቁር የወገብ ህመም የእግር ህመም የጡት
     "context": [""]
    },
```

Figure 13 Number of intents

Figure 13 shows number of intents for creating the model during training the conversational AI. Then the proposed system responds for users based on the knowledge it get from training data.

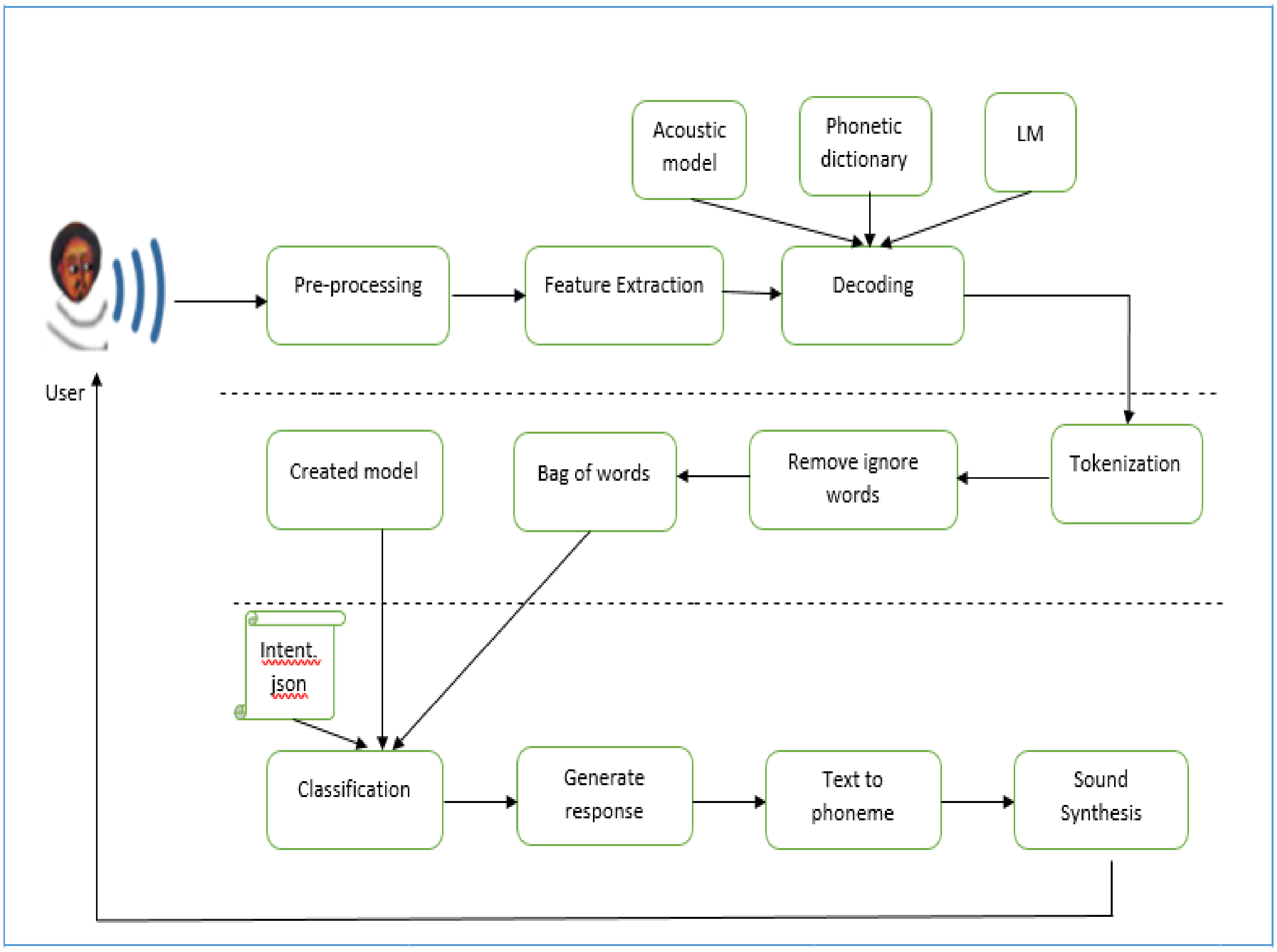


Figure 14 The overall architecture of proposed prototype

3.5.1. Amharic Automatic Speech Recognizer (Amharic ASR).

For our speech based conversational AI we need to have a good speech recognizer that can able to recognize and transcribe user's speech input into text. Recognizing user's speech properly help the conversational AI system to generate appropriate response according to their input. If the ASR didn't recognize user's speech input well the conversational system will generate unappropriated response. The automatic speech recognizer (ASR) of this system is developed for Amharic language only. The recognizer is user dependent, it is not trained with multiple speech data of multiple users with the variety of age, gender and accent. To solve the problem of the speech recognizer which is stated in chapter1 under the statement of problem and to improve its recognition ability the Amharic ASR is developed.

The previous Amharic speech recognizer developed for recognizing words only (Seyoum, 2015) because the speech recognizer trained with words but the proposed Amharic ASR targeted to able to recognize words, phrases and sentences.

Speech is a complex phenomenon and it is a dynamic process with no clearly distinguished part. It is the continuous stream of audio. The naïve perception about speech is it is built from words and each words consists of phones, but the reality is different. Descriptions about speech are probabilistic which means there is no exactly known boundary between words or between units (Otander, 2015).

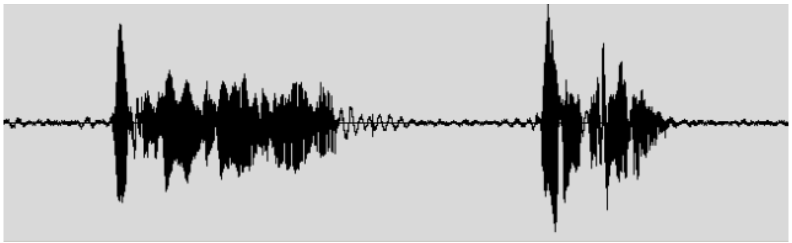


Figure 15 Here is for example the speech recording in an audio editor (Otander, 2015)

During recognition process to understand what is being said we take wave form, split it at utterances by silence and try to understand. To do this all combination of words are taken and try to match them with the audio. In matching process there are three important concepts which are features, models and matching process itself. By dividing speech into frames there are numbers which is calculated from the speech which is called features. Model is mathematical object that able to gathers common attributes of the spoken word. Hidden Markov Model (HMM) is speech model and it describe sequential process e.g. speech. Process in this model considered as a sequence of states and they are changed each other with certain probability.

The speech recognition system predict which phone would be uttered based on an input audio signal. Phones derived using different algorithms for processing speech well. Both hidden Markove model (HMM) and deep neural network are main approaches used in acoustic model for mapping the most likely phones. In this study we used Hidden Markove model.

HMM is the simplest variant of a dynamic Bayesian network and a stochastic signal model (Champ & Shepherd, 2007). The name HMM sprung from its ongoing process, not visible during calculation. In speech recognition for each of phones the acoustic model is the representation of HMM. Let U be an unknown utterance generated by any of the models a_1, a_2………… a_n in the acoustic model, then calculating p(U|a_i) can be used to select the phone most likely to have been uttered. Because of the ability to train HMMs, they are often adopted to map the relationship between audio signals and phones.

Hidden Markove Model has transition probability parameter for representing the probability to move from one state to the other and output parameter, which is set of output probability densities.

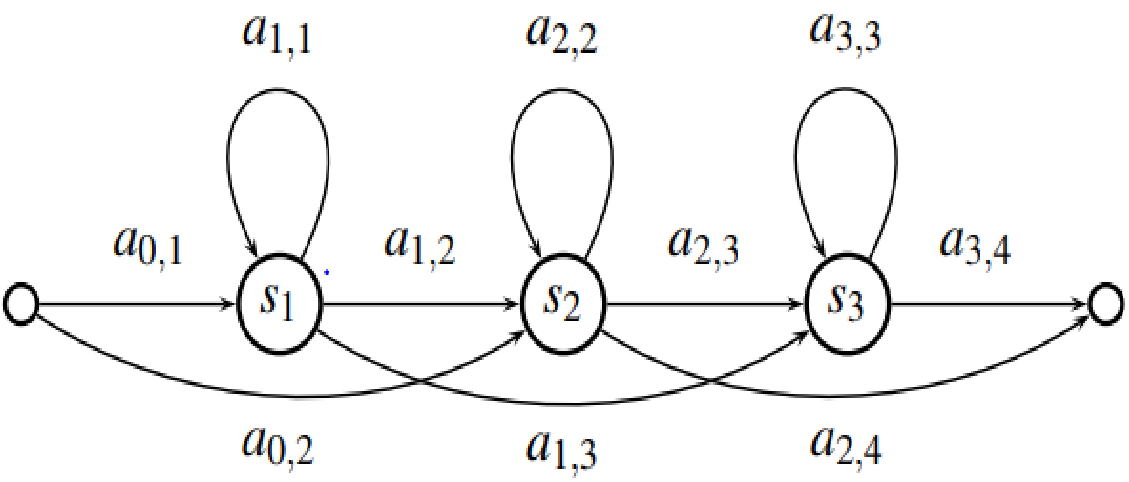


Figure 16 A Hidden Markove Model

The probability to move from one state to the other (or back to the same state) represented by set of transition probability parameters $a_{1,\ 1............a3,4}$. For each states S_1, $S_{2\ ....,S}3$ the probability density of acoustic observation X calculated by the function $P(X|S_i)$. in this HMM calculating phones with varying length is possible because it iterates the same state several times. Based on the audio signal the HMM used to find out most popular phones (Blunsom, 2004)

There are three models used to do the match in speech recognition. These are acoustic model, phonetic dictionary and language model.

Acoustic model - $p(phones|speech)$

With acoustic model data from an audio signal translated to the most probable phones uttered. It contains statistical mappings for each phone (or phonemes) and describes the sound of words. There are context-dependent models those are built from senones with context and context-independent models that contain properties that is most probable feature vectors for each phone. Basically acoustic models contain the acoustic properties of each senone (Otander, 2015).

Phonetic dictionary - $p(words|phones)$

It is a dictionary which maps the words and its phone relation. Because of pronunciation difference the dictionary may contain several variants of some words. If this model is very large it can affect the decoding time of speech recognition.

language model - $p(sentences|words)$

It contains statistical information about which words should follow the other in a sentence and it used the matching process by removing words that are not probable (Otander, 2015). This model is important for determining the probability of sentences and to select the one that represent the audio signal most likely. Because words in different context has different meaning. Some words are very similar out of their context like the word “two” and “too” but both words are quite different. It is difficult to differentiate both words in speech. when calculating the likelihood of words, the static number of previous words often considered. Models using this technique is called N-gram models, in this the context is taken into account by looking at the (N-1) previous words.

To recognizing the users input speech there are three steps that have to be passed, those are preprocessing, feature extraction and decoding.

Preprocessing: in which the number of channels need to be defined and frequency rate of input speech is need to be 16 KHz (16000Hz) in our context. In the preprocessing silences removed from the beginning or at the end of speech and noises would be removed to make it ready for next step.

Feature Extraction: From audio file huge amount of data's obtained for this reason filtration is required before decoding is begin**.** Feature extraction **is** dividing the input speech into overlapping frames of about 25ms. The extracted frames from the speech data referred to as feature vector. To represent the audio parameters are extracted for future use during decoding.

Decoding: in this step to represent the feature vector most likely sentences would be calculated. For example, the feature vector Y $=y_1$, y_2, y_3 ……. y_n and to be the uttered sentence the sequence of words $S=s_1$, s_2, s_3 ……... $s_{n,}$ it uses Bayes Rule to calculate the most probable sentence to represent the feature vector which is $w^* = argmax_w\{p(Y|s)\, p(s)\}$. For decoding the audio, the acoustic model, phonetic dictionary and language model are required. Acoustic model contains the hidden Markove model states for each phone which used to calculate phones.

3.5.1. Text to speech

Text-to-speech (TTS) is converting the given text into its equivalent sound. Through speech synthesis process an artificial human sound produced. In the proposed study responses for user input generated in the form of both text and speech. The speech output is generated by synthesizing the text response. The text-to-speech synthesis is performed by integrating espeakNG into the proposed conversational AI. The espeakNG synthesizer support over 70 languages including Amharic and it uses formant synthesis method. This method allows many languages to be provided with small size. The TTS generate sound for the response generated through converting the text into phoneme. Text-to-phoneme translation is the first step for converting text to speech. In which the generated text as a response for users input translated into pronunciation phonemes then the pronunciation phonemes are synthesized into sound. The speech sound is clear but, it is not as natural as synthesizers trained with human speech recordings. Because of this it generates robotic sound and it has lack of smoothness.

The formant synthesis is a sort of source-filter-method that is based on human speech organ mathematic models. The approach pipe is modeled from a number of resonances with resemblance to the formants in natural speech. Formants are frequency bands with high energy in voices. The first electronic voices Vocoder, and later on OVE and PAT, use formant synthesis to produce sounds. The sound they speak were totally synthetic and electronic.

3.6. The proposed system the overall work flow

The developed speech conversational system in this study can get input from users in both speech and text form, but the developed ASR recognize only Amharic language. The Amharic ASR after getting the speech input form the user it transcribes into text through preprocessing, feature extraction and decoding then the system generates the appropriate response by analyzing the input. The response which is generated from the system is based on the knowledge learned from the training dataset of our domain. When we train there is a model created based on the data we feed into the system and the system used this model to give response for users input.

For creating the required model there is provided conversational dataset for our domain. The provided dataset is saved in the JSON format. The JSON file contains a bunch of messages that users most probably may will use in the conversation with the provided system. Intents in the JSON file have their own tag which is a unique identifier of each intent. Using the provided dataset, we train neural network to take sentence of words and to classify it. The training process begin by loading the dataset to the system. After the data is loaded it tokenized into words using nltk tokenizer. Those tokenized words have to be saved into bag of words with no repetition.

Both neural network and machine learning algorithms needs numerical inputs. For this reason, we use bag of words to represent sentences in the dataset with numbers. Then each sentences represented with a list the length of the amount of words in our models vocabulary and each position in the list represent a word from our vocabulary. In which “1” represent the word is exited in the position of the list otherwise the position in the list would be “*0”*. This bag of word used to get words in our model. If the sequence of words in the sentence is lost, we can know the existence of words in our models vocabulary. All words in the bag of words should be exist in the intents of the dataset.

To make sense our neural network as we formatting our input we need to format our output. So output lists created which are the length of the amount of labels or tags we have in our dataset. Each position in the created list represent labels or tags, like bag of words “*1”* in any position of the list show the tag which is represented.

There are irrelevant characters (“” □ □ ‘ ’ □ □ ? / ! .) which have to be ignored during the training and when the user enter an input to the system during conversation those characters are note being considered. But those characters have their own meaning when peoples are using them for writing letters or some other texts. The created model is trained using TFLearn by transforming conversational intents that we provide into Tensorflow model then the model saved as Pickle format. TFLearn is high-level API which make building and training of neural networks fast and easy. TFLearn is built on top of Tensorflow and it is modular and transparent deep learning library. It was designed to provide high-level API to Tensorflow for facilitating and speed up experimentations.

The system generates a response for user inputs based on the trained model. The model doesn’t take the whole sentence that users enter, but it takes bag of words. So users input need to be transform into bag of words. The trained model cannot spit out sentences, but for all our classes it generates a list of probabilities. To respond users, request the system convert user input into bag of words then get prediction from the model and find the most probable class in which it belongs too. Then the response would be picked from that class.

Finally, the generate response from the system synthesized in the form of sound for the user. Users get response for their input in the form of sound. Actually can users get the response in both text and speech form. The generated response is synthesized by espeakNG.

3.7. Evaluation

Evaluation is the process of examining the proposed framework. For evaluating the proposed prototype, we considered word error rate, accuracy, dialogue success rate and precision and they are used to evaluate different components as their appropriateness. The Amharic speech recognizer evaluated according to the word error rate and accuracy it generates from the test audio and text data. Word error rate is standard to identify how much the recognizer able to recognize the input speech or sound. This is computed by using the text which is hypothesized by the proposed ASR from the test audio data and by taking the original transcribed text of the given test audio data. The other one is the accuracy of the proposed ASR could be analyzed. Then the overall end-to-end speech conversational AI dialogue success rate and its precision's are evaluated by considering the task it performs or the correct response it responds for user's according to their input.

Chapter 4: Experiment and Result

4.1. Introduction

In this chapter we described about the proposed Amharic ASR and the end-to-end speech conversational AI. Poketsphinx, Sphinxbase and Sphinxtrain are used as a tool for Amharic speech recognizer. After we trained our Amharic speech recognizer using both text and audio data we have to test how much the trained Amharic ASR able to recognize the given input. To do this we prepare test data in both text and audio. The other experiment is identifying how much the end-to-end speech conversational AI able to perform tasks based on user’s request.

4.2. Data usage

There are datasets prepare for training and testing. As we described in the chapter three those data’s are prepared from different sources like literatures and other reading materials. During preparing the dataset we classified into training and testing dataset which are used for training and testing our proposed prototype. For our proposed study we had the audio data in addition to text data. The audio data recorded by selecting single women and recording is performed by reading the provided training and testing text data which is used for testing and training Amharic ASR. From the total of 560 both text and audio data’s 448 0r 80% used for training and the remaining 112 or 20% used for testing Amharic speech recognizer.

The conversational AI trained with the data provided which contain list of intents. Those intents have their own tag, pattern and response. From the training dataset the system generates a model which will be used for classifying users input and generating the proper answer. Tags are a unique identifier of each intent from the other, patterns include questions or issues raised by users during conversation and the corresponding solutions for user’s issues are putted as a response.

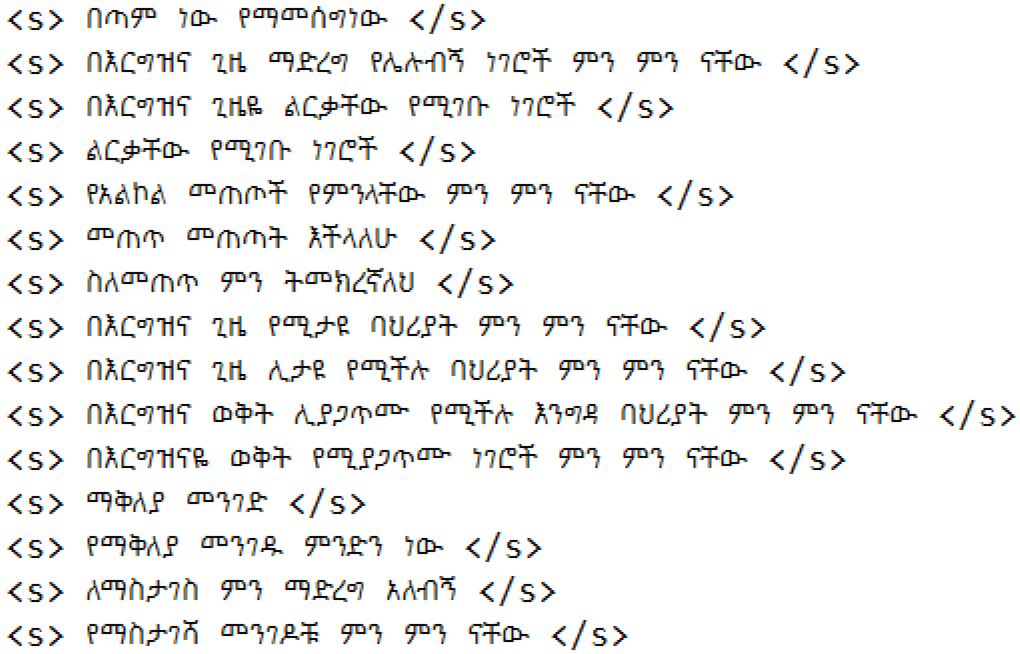
4.3. Experiment and Results of Amharic ASR

The Amharic speech recognizer of the proposed study is trained with the trained data to generate the acoustic model for future use during decoding the test data.

For training the Amharic speech recognizer we used both text and audio data from the provided dataset. Each text data in the training dataset labeled with its recorded audio. The audio file have ‘’.wave’’ format. For training the speech recognizer the phonetic dictionary and language models also used. Phonetic dictionary mapped from word to phones. This phonetic dictionary is generated from the text data that we prepare to use for training. All the words and phones in the dictionary must be exist in the training text. The language model also generated from the training text data. Statistical language models contain the probability of words and combination of words. There are different tools like CMU language modeling toolkit, ARPA and online quick web service to build statistical language model. Those toolkits used for generating the language model based on the data we feed. In this study we used quick online web service because our data is not much larger. The CMU language modeling toolkit recommended to use for large datasets.

There are three different formats that language models can be loaded and stored. These are binary BIN format, text ARPA format and binary DMP format. The *.lm.bin* is the file extension of binary files The binary format load faster and take less space. The ARPA format it takes more space and it is possible to edit. The “*. lm”* is the file extension of ARPA file. The last one is DMP which is not recommended and obsolete.

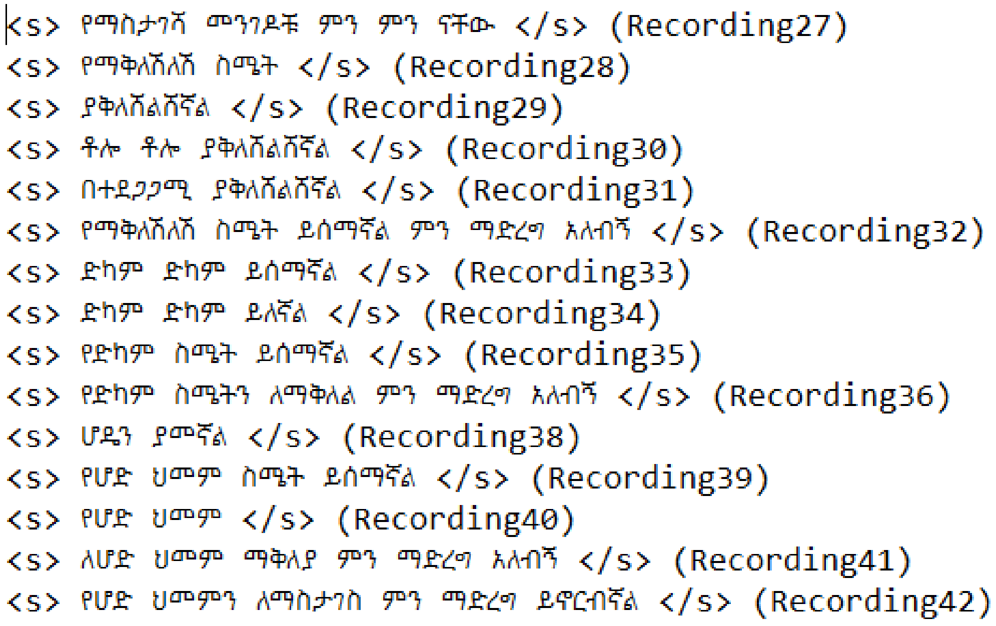
To generate statistical language model we used the training text data. During preparing the training text data Arabic numbers should be converted to words, abbreviations are expanded and also non-word items are cleaned. Texts used for generating language model is labeled as displayed in the figure below. It is in the form of normalized text file and each sentence begins with ‘’<s>’’ tag and end with ‘’</s>’’ tag.



```
<s> በጣም ነው የማመሰግነው </s>
<s> በእርግዝና ጊዜ ማድረግ የሌሉብኝ ነገሮች ምን ምን ናቸው </s>
<s> በእርግዝና ጊዜዬ ልርቃቸው የሚገቡ ነገሮች </s>
<s> ልርቃቸው የሚገቡ ነገሮች </s>
<s> የአልኮል መጠጦች የምንላቸው ምን ምን ናቸው </s>
<s> መጠጥ መጠጣት እችላለሁ </s>
<s> ስለመጠጥ ምን ትመክረኛለህ </s>
<s> በእርግዝና ጊዜ የሚታዩ ባህሪያት ምን ምን ናቸው </s>
<s> በእርግዝና ጊዜ ሊታዩ የሚችሉ ባህሪያት ምን ምን ናቸው </s>
<s> በእርግዝና ወቅት ሊያጋጥሙ የሚችሉ እንግዳ ባህሪያት ምን ምን ናቸው </s>
<s> በእርግዝናዬ ወቅት የሚያጋጥሙ ነገሮች ምን ምን ናቸው </s>
<s> ማቅለያ መንገድ </s>
<s> የማቅለያ መንገዱ ምንድን ነው </s>
<s> ለማስታገስ ምን ማድረግ አለብኝ </s>
<s> የማስታገሻ መንገዶቹ ምን ምን ናቸው </s>
```

Figure 17 Prepared text for generating Statistical language model

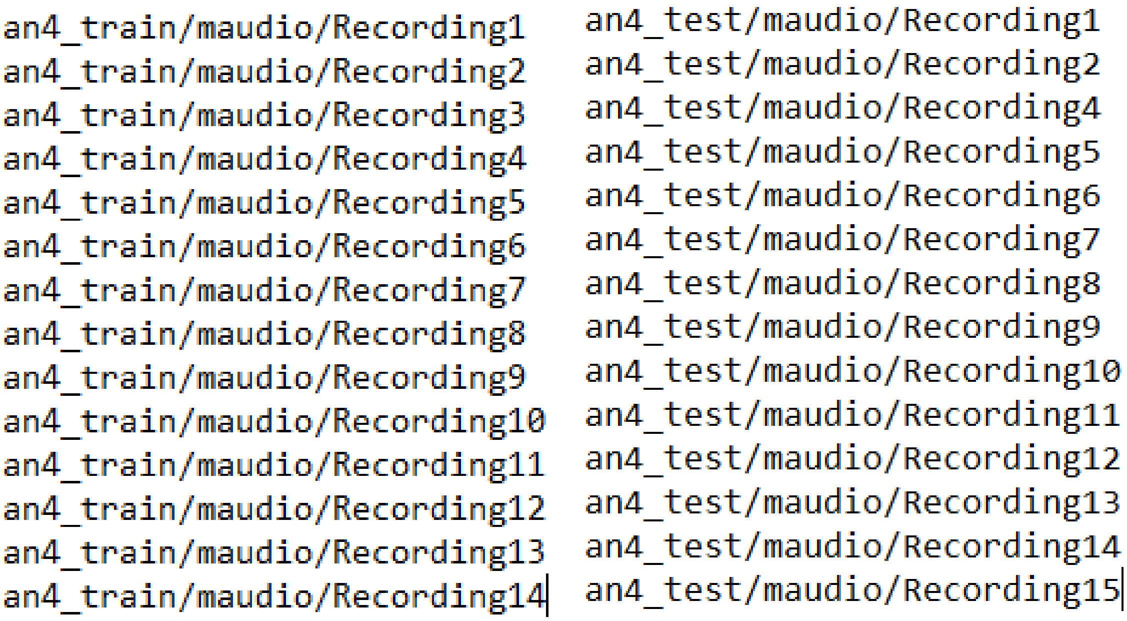
Before we train Amharic ASR we have to label each sentences in the training and test data with its corresponding wave file as shown in the figure below.



```
<s> የማስታገሻ መንገዶቹ ምን ምን ናቸው </s> (Recording27)
<s> የማቅለሽለሽ ስሜት </s> (Recording28)
<s> ያቅለሸልሸኛል </s> (Recording29)
<s> ቶሎ ቶሎ ያቅለሸልሸኛል </s> (Recording30)
<s> በተደጋጋሚ ያቅለሸልሸኛል </s> (Recording31)
<s> የማቅለሽለሽ ስሜት ይሰማኛል ምን ማድረግ አለብኝ </s> (Recording32)
<s> ድካም ድካም ይሰማኛል </s> (Recording33)
<s> ድካም ድካም ይለኛል </s> (Recording34)
<s> የድካም ስሜት ይሰማኛል </s> (Recording35)
<s> የድካም ስሜትን ለማቅለል ምን ማድረግ አለብኝ </s> (Recording36)
<s> ሆዴን ያመኛል </s> (Recording38)
<s> የሆድ ህመም ስሜት ይሰማኛል </s> (Recording39)
<s> የሆድ ህመም </s> (Recording40)
<s> ለሆድ ህመም ማቅለያ ምን ማድረግ አለብኝ </s> (Recording41)
<s> የሆድ ህመምን ለማስታገስ ምን ማድረግ ይኖርብኛል </s> (Recording42)
```

Figure 18 Transcription text labeled with its corresponding recorded audio

For the system to able to find the recorded audio for the transcribed text we have to specify the location where each audio of transcription text is located. The figure below shows the assigned path of recorded audio for both training and testing.

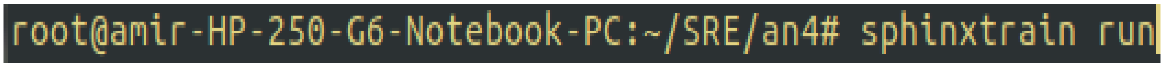


```
an4_train/maudio/Recording1     an4_test/maudio/Recording1
an4_train/maudio/Recording2     an4_test/maudio/Recording2
an4_train/maudio/Recording3     an4_test/maudio/Recording4
an4_train/maudio/Recording4     an4_test/maudio/Recording5
an4_train/maudio/Recording5     an4_test/maudio/Recording6
an4_train/maudio/Recording6     an4_test/maudio/Recording7
an4_train/maudio/Recording7     an4_test/maudio/Recording8
an4_train/maudio/Recording8     an4_test/maudio/Recording9
an4_train/maudio/Recording9     an4_test/maudio/Recording10
an4_train/maudio/Recording10    an4_test/maudio/Recording11
an4_train/maudio/Recording11    an4_test/maudio/Recording12
an4_train/maudio/Recording12    an4_test/maudio/Recording13
an4_train/maudio/Recording13    an4_test/maudio/Recording14
an4_train/maudio/Recording14    an4_test/maudio/Recording15
```

Figure 19 Locating the corresponding audio record for each text transcription

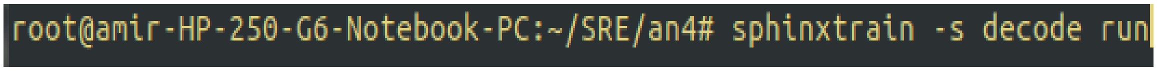
Through the process of training Amharic ASR, the acoustic model is generated then it will be used with phonetic dictionary and language model to decode user's input speech. Here is the command we used for training Amharic ASR and for decoding the test data.

For training Amharic ASR



```
root@amir-HP-250-G6-Notebook-PC:~/SRE/an4# sphinxtrain run
```

For testing Amharic ASR



```
root@amir-HP-250-G6-Notebook-PC:~/SRE/an4# sphinxtrain -s decode run
```

Amharic speech recognizer tested using test data to know how much the recognizer able to recognize. During testing of Amharic ASR both the word error rate and accuracy are considered to know how much the recognizer able to recognize. Word error rate is a common way for evaluating speech recognizer accuracy. It is a way for measuring errors occurred during the audio signal translation. Word error rate calculated by summarizing the total number of errors in the hypothesis and dividing it by the total number of words in the correct sentence. An error is either in the deletion, insertion or in an incorrect substitution. Word error rate is calculated as (Otander, 2015):

$$\mathbf{WER = (I + D + S) / N}$$

Let's consider that we have two texts which are an original text and recognized one with length of N words. In the formula to calculate word error rate *I* indicates the number of inserted words, *D* indicates the number of deleted words and *S* for substituted words. Word error rate expressed in percent.

Accuracy show the accuracy level that the recognizer able to recognize the given speech input and it is generated after testing. It doesn't take insertion into account. As described by (Otander, 2015) the equation for calculating the accuracy is:

$$\mathbf{Accuracy = (N - D - S) / N}$$

Based on the above consideration from 20% (112 both text and audio) test data we have got an accuracy of 90% and the remaining is an error. Which indicates 90 percent from the test data is recognized correctly by Amharic ASR and the remaining is not recognized correctly. The implication of this result is the developed Amharic speech recognizer can give appropriate transcribed input for the conversational system during conversation. If we have accurate speech recognizer the system able to generate the required response for users input.

4.4. Experiment and result of the overall end-to-end conversational AI

The proposed prototype is an end-to-end speech conversational AI, this proposed prototype communicate with users using speech and it gives output for users in the form of sound, but also it can able to get an input and respond from and to users in the form of text. There are number of tasks listed which can be performed by users. Those tasks are greeting, goodbye, thanks, get information about what she has to do, from what she has to far, discussing about what kind of physical exercise she can do, discussing pregnancy symptoms and the self-treatments she can do, what kind of physical fitness exercises she doesn’t have to do, asking information about what kind of vaccines and medical treatments she have to get during her pregnancy and others. According to those tasks the conversational AI would be evaluated by considering both dialogue success rate and precision of the provided prototype. To evaluate the dialogue success rate and precision we used 27 set of dialogues and 102 natural language queries. Dialogue success rate for each set calculated as:

Dialogue success rate for each set = number of answers or responses generated by the system / Number of turns issued by the user

Number of turns are the total number of issues raised by the user and the response given to the system during conversation and number of answers are answers generated by the system to respond for user’s questions or input. By calculating total of dialogue success rate of each set we would find the dialogue success rate.

The total dialogue success rate obtained from 27 dialogue set is 24.84. Total turns in 27 dialogue sets is 102. Each dialogue consisted of 1 to 6 queries for performing a single task. the dialogue success rate for the conversational AI calculated as:

*Dialogue success rate = (∑Dialogue success rate for each set / Number of sets of dialogues) * 100*

*= (24.84 / 27) * 100*

= 92%

The result of the dialogue success rate shows the conversational AI respond 92% from the given dialogue sets successfully. The remaining 8% is responded incorrectly. The reason can be the speech recognizer didn’t recognize user’s speech, or the trained data limitation.

The precision is defined as:

*Precision = (Number of correct answers given by the system / number of answers given by the system) * 100.*

*= (94 / 98) * 100   0.959183≈0.96*

*= 0.96 * 100*

=96%

The system generates 98 answers from total of 102 user inputs. From those generated answers 94 was answered correctly. Then the precision is calculated.

[□□□]

□□□□□□□

[□□□□]

□□□ □□□□□□□ □□□□□ □□□ □□ □ □□□□□ □□□□□ □□□ □□ □

[□□□]

□□□□□□ □□ □□□ □□□□ □□□□□ □□ □□ □□□

[□□□□]

□□□□□□ □□ □□ □□□□□ □□□ □□□ □□ □□□ □□ □□□□ □□□ □□□□ □□□
□□□ □□□ □□□□
□□ □□□□ □□□□□ □□□ □□□□ □□□□□ □

[□□□]

□□□□□□ □□ □□□□ □□□□

[□□□□]

□□□□□ □□□□□□ □□ □□□□ □□□ □□?

[□□□]

□□□□□□□ □□□

[□□□□]

□□□□ □□□□□ □□□ □□ □□□□□ □□□□ □ □□□ □□□ □□□□ □ □□ □□□□ □□□□ □□ □□□ □□ □□□□□□ □□□□□□□ □□□□ □□□□□□ □□□□□ □

[□□□]

□□□□□□□

[□□□□]

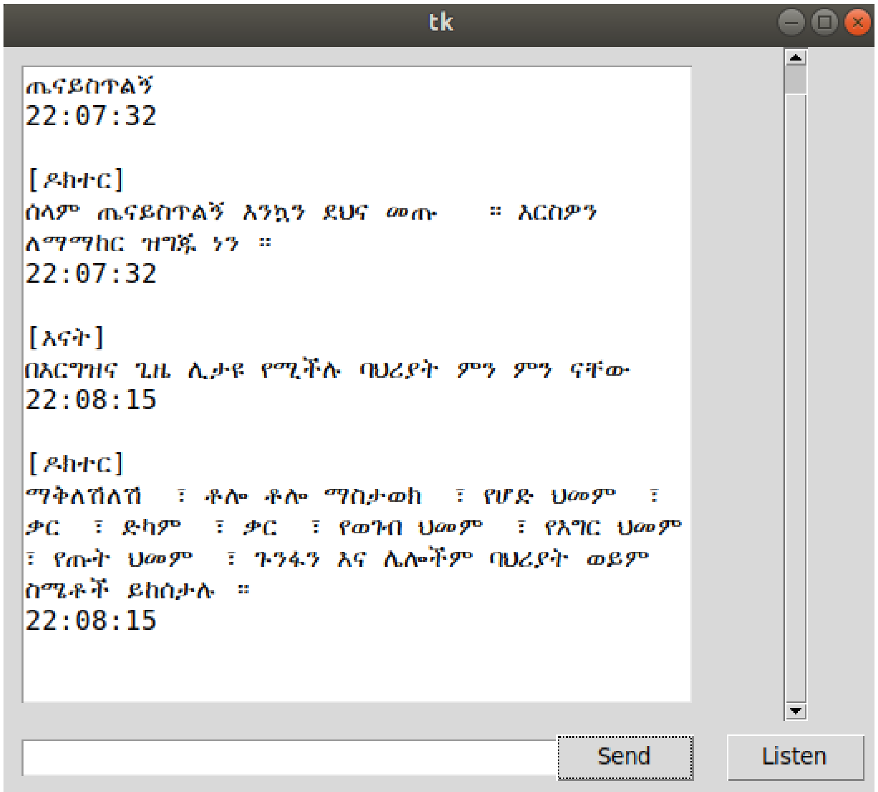
□□□ □□□□□□ □□□□□□□□ □□□ □□□□□□□□ □□□ □□□

[□□□]

□□

[□□□□]

□□□□□□ □□□□□□□□



```
ጤናይስጥልኝ
22:07:32

[ዶክተር]
ሰላም ጤናይስጥልኝ እንኳን ደህና መጡ    ። እርስዎን
ለማማከር ዝግጁ ነን ።
22:07:32

[እናት]
በእርግዝና ጊዜ ሊታዩ የሚችሉ ባህሪያት ምን ምን ናቸው
22:08:15

[ዶክተር]
ማቅለሽለሽ   ፣ ቶሎ ቶሎ ማስታወክ   ፣ የሆድ ህመም   ፣
ቃር   ፣ ድካም   ፣ ቃር   ፣ የወገብ ህመም   ፣ የእግር ህመም
፣ የጡት ህመም   ፣ ጉንፋን እና ሌሎችም ባህሪያት ወይም
ስሜቶች ይከሰታሉ ።
22:08:15
```

Figure 20 sample conversation with the proposed system

The prototype provided in our study have the ability to perform the tasks which is described above. For each task the prototype gives the appropriate and required answer according to our domain we focused on and considered. During an experiment we gave an input in both text and speech form to perform tasks. So the prototype performed tasks as required for both inputs. The above conversation is sample task about discussing pregnancy symptoms and their treatment mechanism. The developed Amharic ASR able to recognize 85% of the test speech data and the conversational AI hast 92% of accuracy.

Chapter 5: Conclusion and Recommendations

5.1. Introduction

This chapter discus about the conclusion we find out from the proposed study and recommendations that we suggest. The stated conclusions generalize the whole work of the proposed study and it puts the main points that the proposed study contains and focused on. Under the recommendation of this chapter the main possible and advisable works that can be done in this domain for the future are listed and described. Which help researchers who are interested to work on this domain by giving some research idea and hint.

5.2. Conclusion

This study is conducted to fill the gap of previous works related with our study. After identifying and analyzing gaps we motivated to fill the holes of those works, specially related works which are done for Amharic language. For solving the identified problems and gaps we use different mechanisms, tool and others. Then we identified domain we need to focus, this help us to get and collect the required data based on our interest. The domain we worked on is healthcare domain about pregnancy. The reason that we select pregnant’s is for supporting them and physicians. Because the physician-patient ratio in Ethiopia is not fair, and pregnant’s have to get consultation every time when they need. So by developing such kind of systems we can support physicians by reducing some work load and help pregnant’s through giving advice.

Generally, the main focus of the proposed study is designing an effective and efficient end-to-end speech conversational AI. Speech based communication is becoming popular because the interaction of human and machine mimic human-human. By considering such things end-to-end speech conversational AI prototype is designed. For speech based system ASR is the main thing we have to consider. To make speeches recognized by the proposed Amharic ASR relevant we collect required data’s by reading literatures and based on that data both the Amharic ASR and conversational AI are trained. Then the speech recognizer has 90% of accuracy during testing and accuracy of the conversational AI is 92%. According to the overall end-to-end speech conversational AI the proposed prototype performs different tasks like greeting, goodbye, thanks, discussing about pregnancy symptoms and their treatment and etc. Based on tasks it performs we have got92% dialogue success rate and 96% of precision at the end of the study we designed the prototype of the required end-to-end speech conversational AI.

5.3. Contribution of the study

The proposed study contributes for the domain we focused on. By filing the gaps of different literatures which are described in chapter two the proposed study able to provide Amharic ASR for conversational AI in the domain of healthcare specially for pregnant’s. This proposed Amharic ASR unlike the works done before able to understand words, phrases and sentences. The other contribution is the provided a prototype is fully end-to-end –speech conversational AI but it can support text also that means the proposed prototype is dual mode of communication. The other thing that we have consider is there was no such kind of study in our country about the conversational AI in the domain of healthcare

5.4. Recommendation

The end-to-end speech conversational AI play a greater role in our day to day life and they can be implemented in different domains including healthcare. In the context of our country conducting study in this domain is important and necessary. But there are different works that have to be done in the future for improving the study and to enabling the study to solve more problems in the healthcare domain. Some of works that have to be focused for the future are listed below.

- Improving the Amharic ASR: The Amharic ASR in this study is user-dependent. So we need to have user independent Amharic speech recognizer through developing and training the speech recognizer with multiple person’s audio data with respect to age, gender and accent. Which helps the system to have the ability to communicate with different users who have difference in their age, gender and the place they live.

- Adding visual outputs like images and videos for users question which help users to understand what the system is saying about the issue they raise.
- Including diagnosis process and Considering other health related issues

Reference

Al-Zubaide, H., & Issa, A. A. (2011). *Ontbot: Ontology based chatbot.* Paper presented at the International Symposium on Innovations in Information and Communications Technology.

Asimare, H. (2020). *Designing and Implementing Adaptive Bot Model to Consult Ethiopian Published Laws Using Ensemble Architecture with Rules Integrated.* (Masters of computerscience), Bahir Dar University,

Berhan, Y. J. E. m. j. (2008). Medical doctors profile in Ethiopia: production, attrition and retention. In memory of 100-years Ethiopian modern medicine & the new Ethiopian millennium. *46*, 1-77.

Blunsom, P. J. L. n., August. (2004). Hidden markov models. *15*(18-19), 48.

BRITZ, D. (2015). Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs. Retrieved from http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

Budulan, S. (2018). Chatbot Categories and Their Limitations. Retrieved from https://dzone.com/articles/chatbots-categories-and-their-limitations-1

Cahn, J. J. U. o. P. S. o. E., Computer, A. S. D. o., & Science, I. (2017). CHATBOT: Architecture, design, & development.

Champ, C. W., & Shepherd, D. K. (2007). Encyclopedia of statistics in quality and reliability.

Grover, A. S. P., Madelaine; Barnard, Etienne; Kuun, Christiaan. (2008). HIV Health Information Access using Spoken Dialogue Systems: Touchtone vs Speech.

Hussain, S., Sianaki, O. A., & Ababneh, N. (2019). *A survey on conversational agents/chatbots classification and design techniques.* Paper presented at the Workshops of the In6ternational Conference on Advanced Information Networking and Applications.

Jokinen, K., & McTear, M. J. S. L. o. H. L. T. (2009). Spoken dialogue systems. *2*(1), 1-151.

Kuramoto, I., Baba, J., Ogawa, K., Yoshikawa, Y., Kawabata, T., & Ishiguro, H. (2018). *Conversational Agents to Suppress Customer Anger in Text-based Customer-support Conversations.* Paper presented at the Proceedings of the 6th International Conference on Human-Agent Interaction.

Landowska, A. J. r. B. L., Postępy e-edukacji, Wydawnictwo PJWSTK, Warszawa. (2010). Application of Intelligent Conversational Agents in E-learning Environments. 357-366.

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., . . . Lau, A. Y. J. J. o. t. A. M. I. A. (2018). Conversational agents in healthcare: a systematic review. *25*(9), 1248-1258.

Lee, C., Jung, S., Kim, K., Lee, D., Lee, G. G. J. J. o. C. S., & Engineering. (2010). Recent approaches to dialog management for spoken dialog systems. *4*(1), 1-22.

Machidon, O.-M., Tavčar, A., Gams, M., & Duguleană, M. J. J. o. C. H. (2020). CulturalERICA: A conversational agent improving the exploration of European cultural heritage. *41*, 152-165.

Marietto, M. d. G. B., de Aguiar, R. V., Barbosa, G. d. O., Botelho, W. T., Pimentel, E., França, R. d. S., & da Silva, V. L. J. a. p. a. (2013). Artificial intelligence markup language: a brief tutorial.

Meena, R. (2016). *Data-driven methods for spoken dialogue systems: Applications in language understanding, turn-taking, error detection, and knowledge acquisition.* KTH Royal Institute of Technology, Mhatre, N., Motani, K., Shah, M., & Mali, S. J. I. J. o. C. A. (2016). Donna interactive chat-bot acting as a personal assistant. *140*(10).

Milward, D., & Beveridge, M. (2003). *Ontology-based dialogue systems.* Paper presented at the Proc. 3rd Workshop on Knowledge and reasoning in practical dialogue systems (IJCAI03).

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. J. J. i. m. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *176*(5), 619-625.

Nishida, T., Nakazawa, A., Ohmoto, Y., & Mohammad, Y. (2014). *Conversational informatics*: Springer.

Nuez Ezquerra, A. (2018). *Implementing ChatBots using Neural Machine Translation techniques.* Universitat Politècnica de Catalunya,

Otander, J. (2015). Basic concepts of speech recognition. Retrieved from https://cmusphinx.github.io/wiki/tutorialconcepts/

Pradana, A., Sing, G. O., Kumar, Y. J. I. J. o. C. I. S., & Applications, I. M. (2017). SamBot-Intelligent Conversational Bot for Interactive Marketing with Consumer-centric Approach. *6*(2014), 265-275.

Strivastava, P. (December 2017). Essentials of Deep Learning : Introduction to Long Short Term Memory. Retrieved from https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/

Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017). *A survey of design techniques for conversational agents.* Paper presented at the International Conference on Information, Communication and Computing Technology.

Sak, H., Senior, A. W., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling.

Seyoum, F. (2015). *Developing Amharic Spoken Dialogue System: A Hybrid Approach.* (Degree of Master of Science), Adiss Ababa,

Wallace, R. S. (2009). The Anatomy of ALICE In: Parsing the Turing test. Part III. In: Springer, Netherlands.

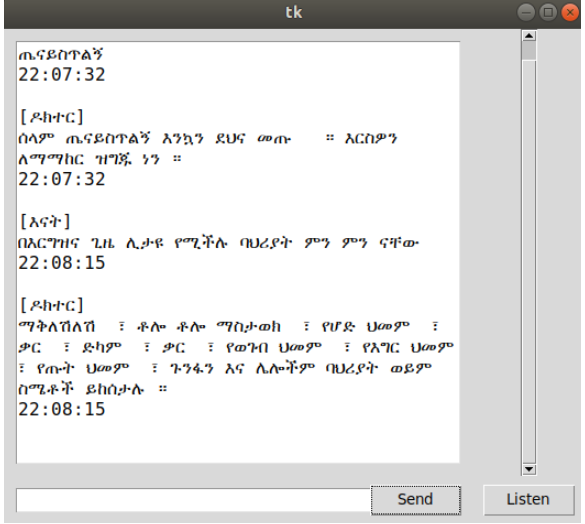
Weizenbaum, J. J. C. o. t. A. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *9*(1), 36-45.

Wei, Z., Liu, Q., Peng, B., Tou, H., Chen, T., Huang, X.-J., . . . Dai, X. (2018). *Task-oriented dialogue system for automatic diagnosis.* Paper presented at the Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).

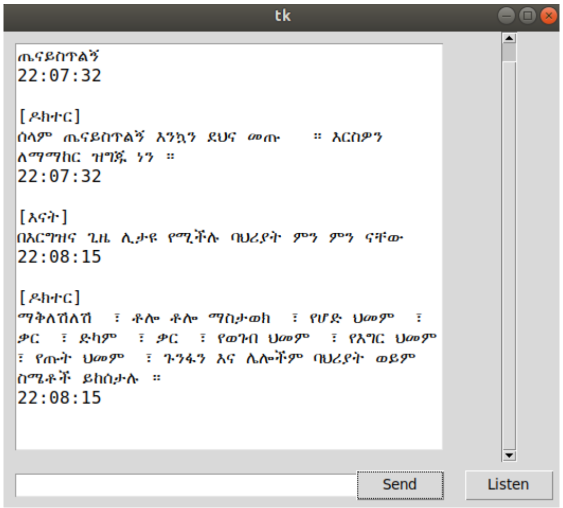
Young, T., Hazarika, D., Poria, S., & Cambria, E. J. i. C. i. m. (2018). Recent trends in deep learning based natural language processing. *13*(3), 55-75.

Appendix

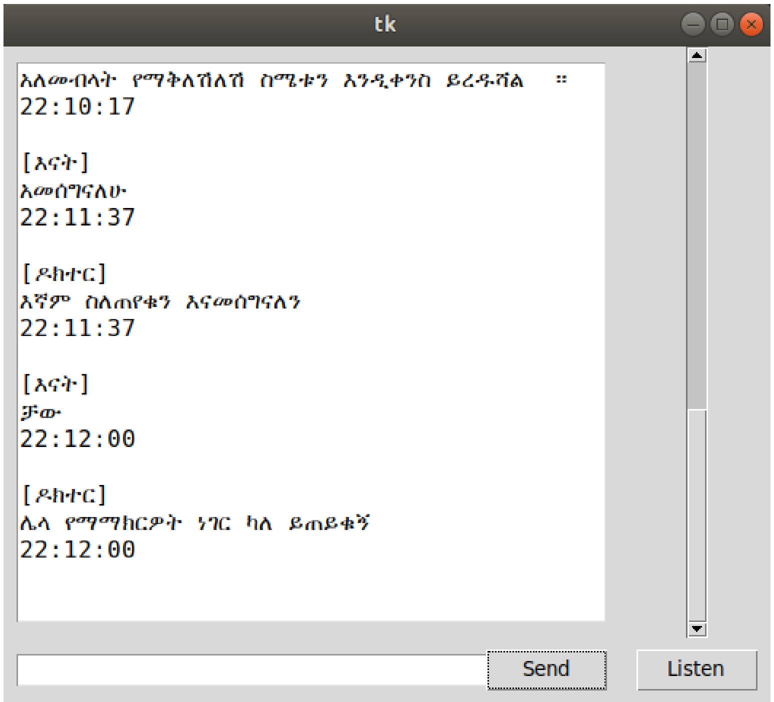
Appendix A Sample Conversation with the proposed system



```
ጤናይስጥልኝ
22:07:32

[ዶክተር]
ሰላም ጤናይስጥልኝ እንኳን ደህና መጡ    ። እርስዎን
ለማማከር ዝግጁ ነን ።
22:07:32

[እናት]
በእርግዝና ጊዜ ሊታዩ የሚችሉ ባህሪያት ምን ምን ናቸው
22:08:15

[ዶክተር]
ማቅለሽለሽ  ፤ ቆሎ ቆሎ ማስታወክ  ፤ የሆድ ህመም  ፤
ቃር  ፤ ድካም  ፤ ቃር  ፤ የወገብ ህመም  ፤ የእግር ህመም
፤ የጡት ህመም  ፤ ጉንፋን እና ሌሎችም ባህሪያት ወይም
ስሜቶች ይከሰታሉ ።
22:08:15
```

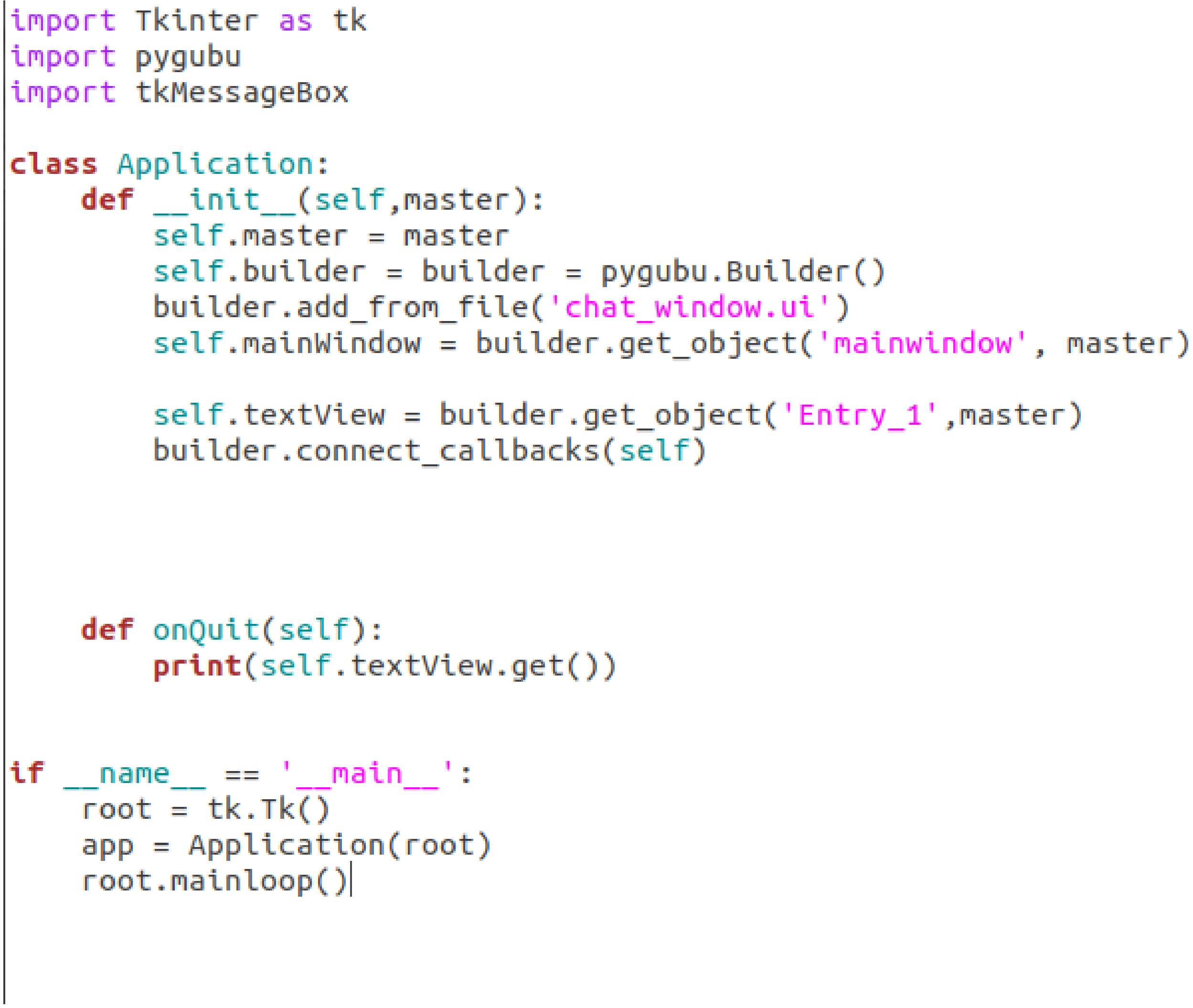


```
ጤናይስጥልኝ
22:07:32

[ዶክተር]
ሰላም ጤናይስጥልኝ እንኳን ደህና መጡ    ። እርስዎን
ለማማከር ዝግጁ ነን ።
22:07:32

[እናት]
በእርግዝና ጊዜ ሊታዩ የሚችሉ ባህሪያት ምን ምን ናቸው
22:08:15

[ዶክተር]
ማቅለሽለሽ  ፤ ቆሎ ቆሎ ማስታወክ  ፤ የሆድ ህመም  ፤
ቃር  ፤ ድካም  ፤ ቃር  ፤ የወገብ ህመም  ፤ የእግር ህመም
፤ የጡት ህመም  ፤ ጉንፋን እና ሌሎችም ባህሪያት ወይም
ስሜቶች ይከሰታሉ ።
22:08:15
```

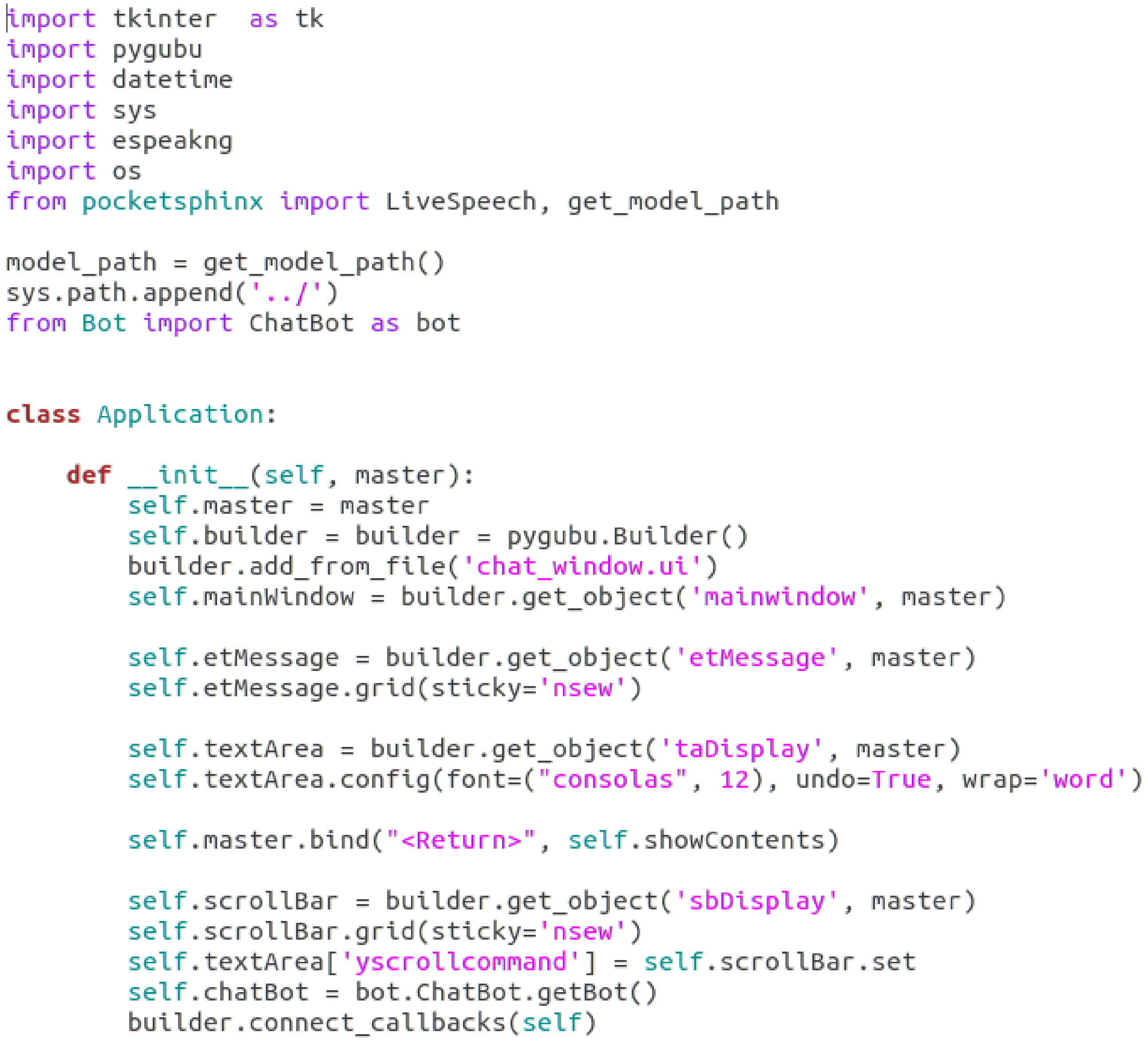


```
አለመብላት የማቅለሽለሽ ስሜቱን እንዲቀንስ ይረዱሻል    ።
22:10:17

[እናት]
አመሰግናለሁ
22:11:37

[ዶክተር]
እኛም ስለጠየቁን እናመሰግናለን
22:11:37

[እናት]
ቻው
22:12:00

[ዶክተር]
ሌላ የማማከርዎት ነገር ካለ ይጠይቁኝ
22:12:00
```

Appendix B Code for designing an interface

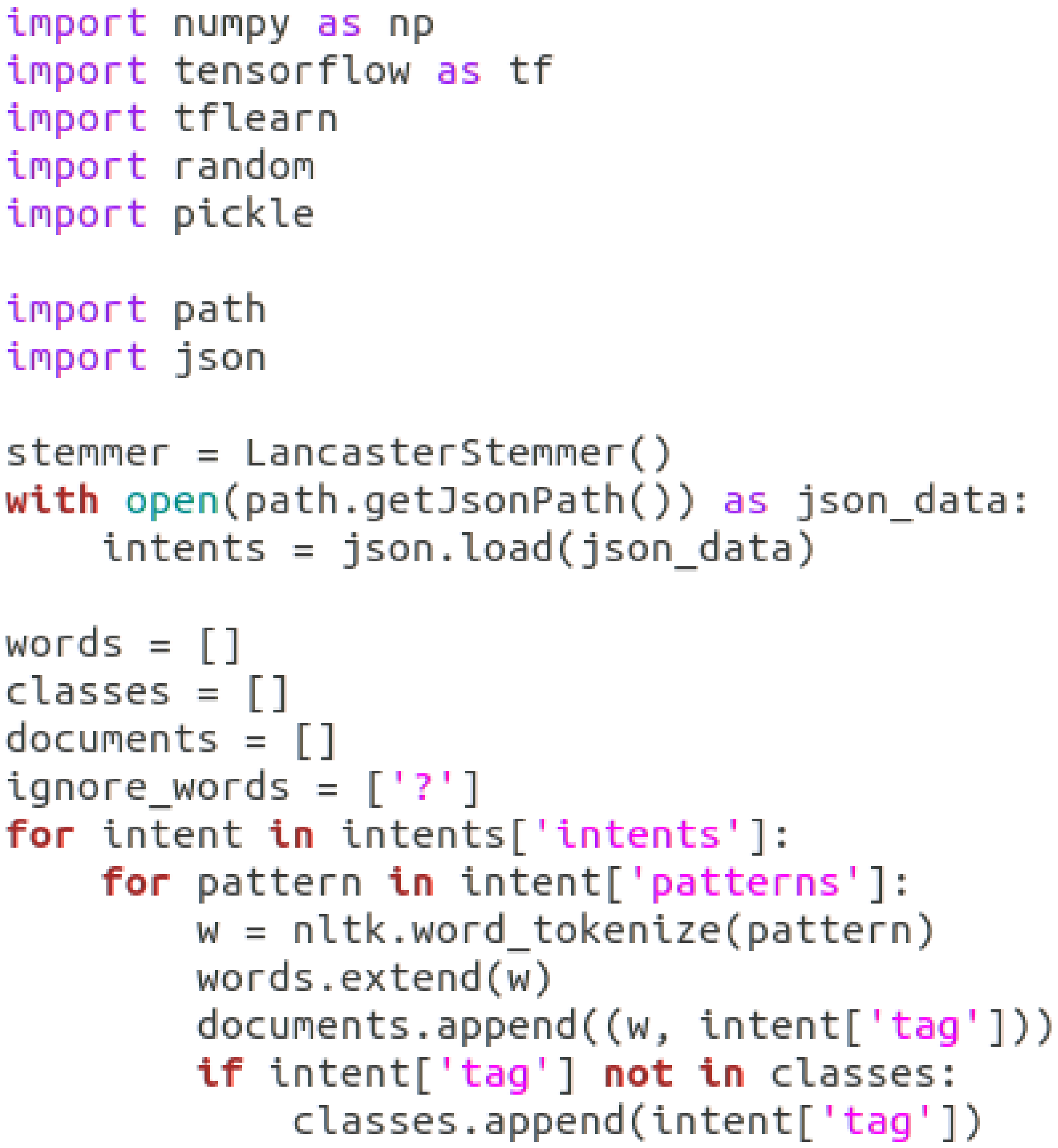


```
import Tkinter as tk
import pygubu
import tkMessageBox

class Application:
    def __init__(self,master):
        self.master = master
        self.builder = builder = pygubu.Builder()
        builder.add_from_file('chat_window.ui')
        self.mainWindow = builder.get_object('mainwindow', master)

        self.textView = builder.get_object('Entry_1',master)
        builder.connect_callbacks(self)




    def onQuit(self):
        print(self.textView.get())


if __name__ == '__main__':
    root = tk.Tk()
    app = Application(root)
    root.mainloop()
```



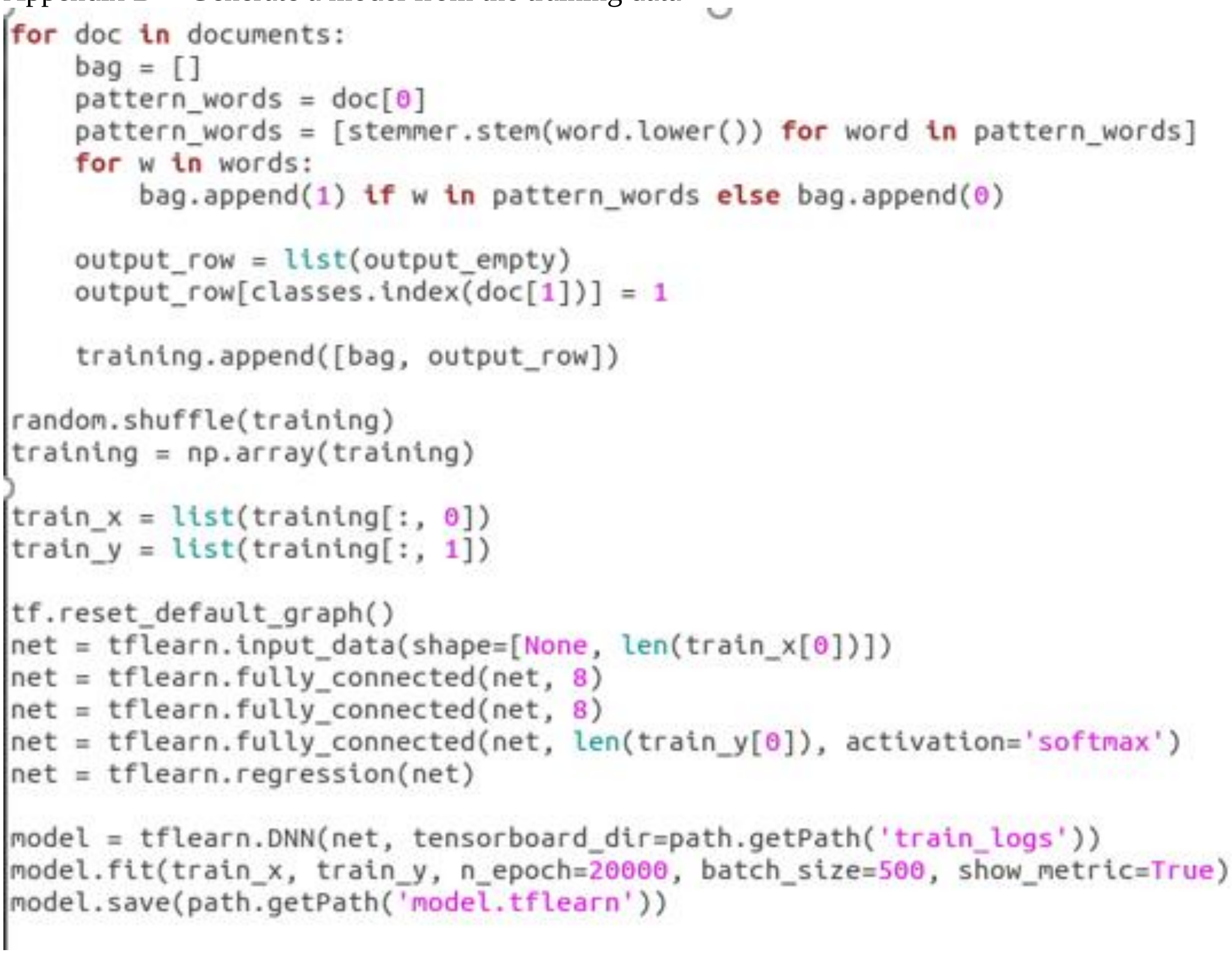
```
import tkinter  as tk
import pygubu
import datetime
import sys
import espeakng
import os
from pocketsphinx import LiveSpeech, get_model_path

model_path = get_model_path()
sys.path.append('../')
from Bot import ChatBot as bot


class Application:

    def __init__(self, master):
        self.master = master
        self.builder = builder = pygubu.Builder()
        builder.add_from_file('chat_window.ui')
        self.mainWindow = builder.get_object('mainwindow', master)

        self.etMessage = builder.get_object('etMessage', master)
        self.etMessage.grid(sticky='nsew')

        self.textArea = builder.get_object('taDisplay', master)
        self.textArea.config(font=("consolas", 12), undo=True, wrap='word')

        self.master.bind("<Return>", self.showContents)

        self.scrollBar = builder.get_object('sbDisplay', master)
        self.scrollBar.grid(sticky='nsew')
        self.textArea['yscrollcommand'] = self.scrollBar.set
        self.chatBot = bot.ChatBot.getBot()
        builder.connect_callbacks(self)
```

Appendix C Loading the dataset for training



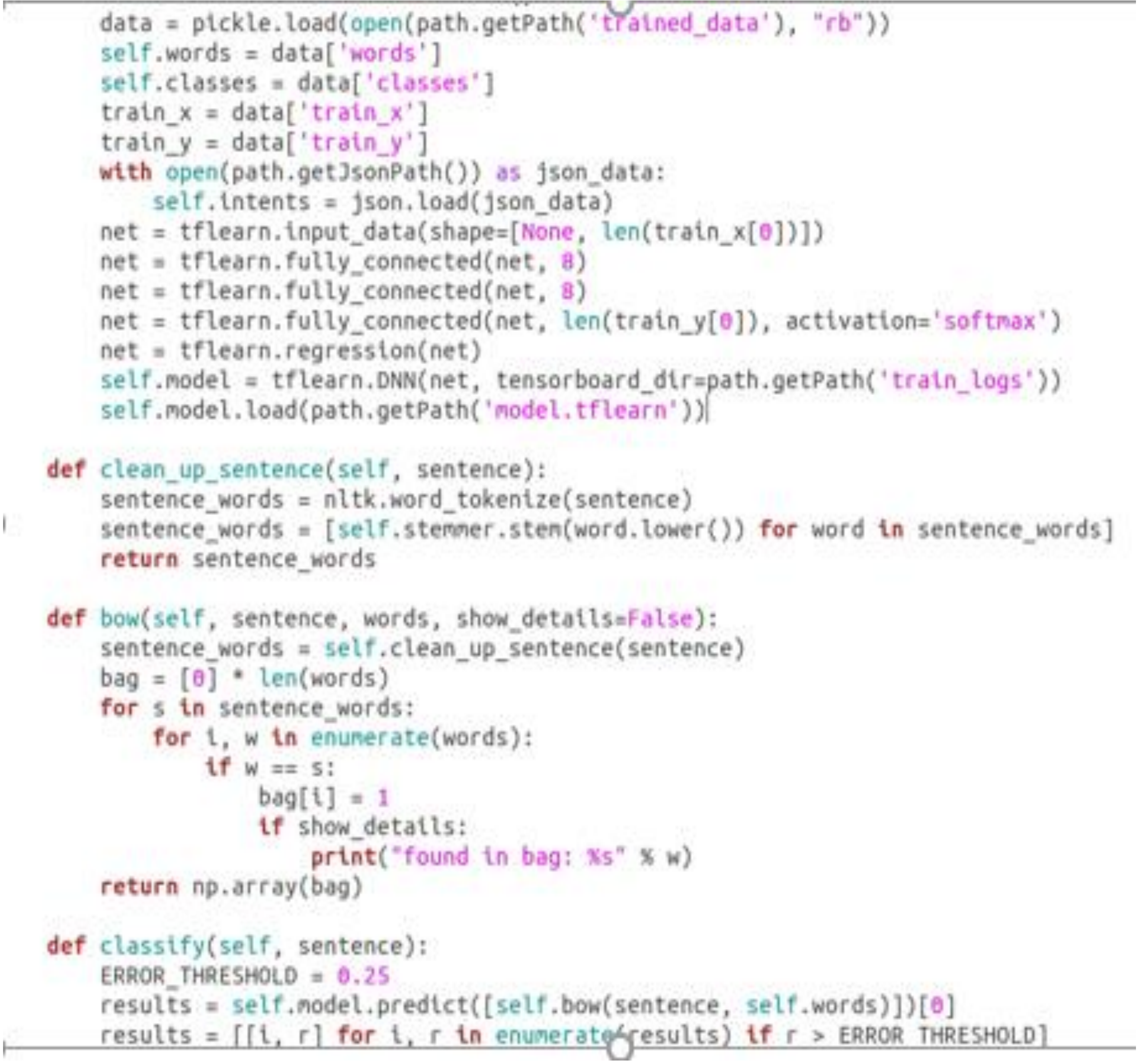
```
import numpy as np
import tensorflow as tf
import tflearn
import random
import pickle

import path
import json

stemmer = LancasterStemmer()
with open(path.getJsonPath()) as json_data:
    intents = json.load(json_data)

words = []
classes = []
documents = []
ignore_words = ['?']
for intent in intents['intents']:
    for pattern in intent['patterns']:
        w = nltk.word_tokenize(pattern)
        words.extend(w)
        documents.append((w, intent['tag']))
        if intent['tag'] not in classes:
            classes.append(intent['tag'])
```

Appendix D Generate a model from the training data



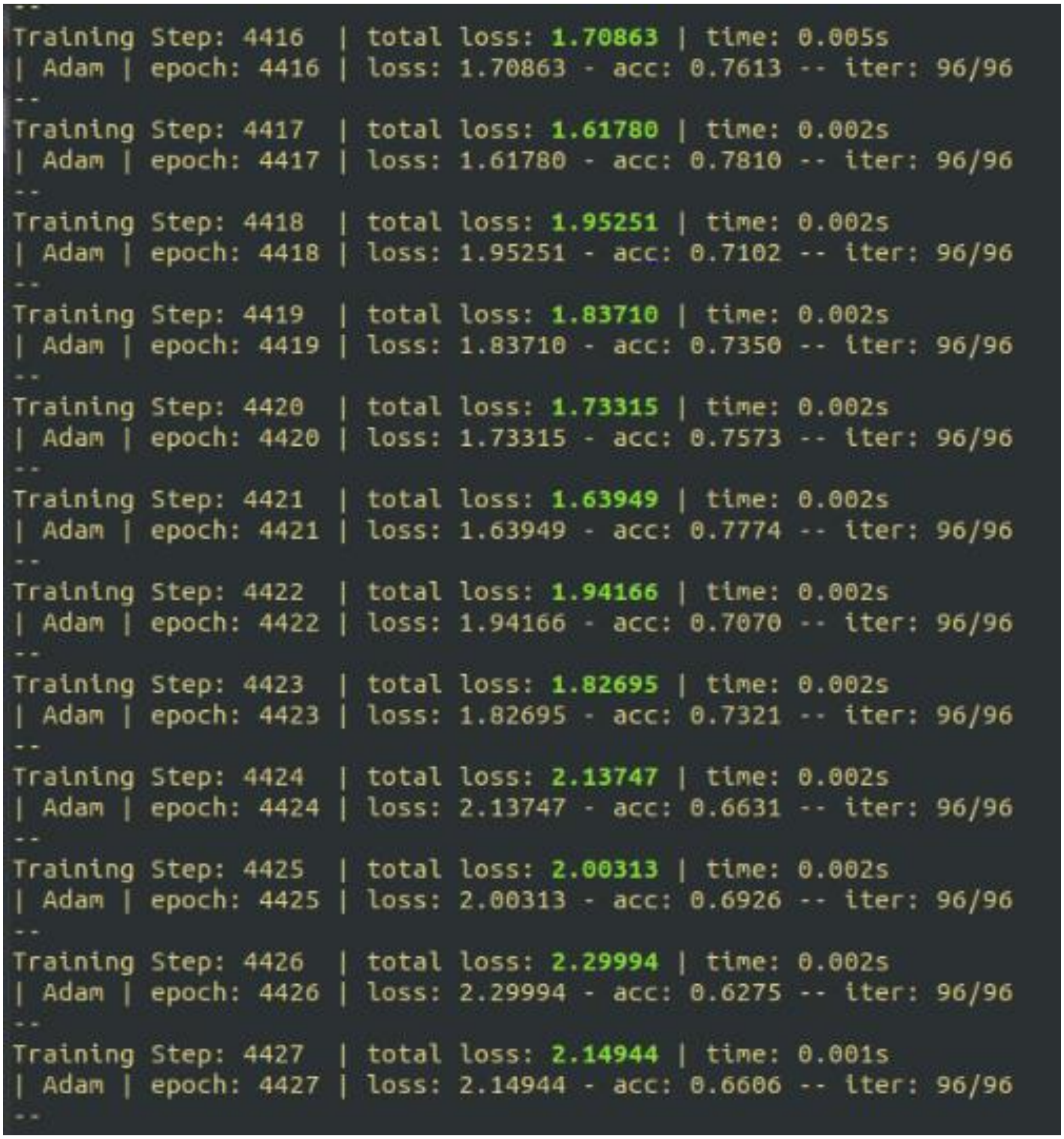
```
for doc in documents:
    bag = []
    pattern_words = doc[0]
    pattern_words = [stemmer.stem(word.lower()) for word in pattern_words]
    for w in words:
        bag.append(1) if w in pattern_words else bag.append(0)

    output_row = list(output_empty)
    output_row[classes.index(doc[1])] = 1

    training.append([bag, output_row])

random.shuffle(training)
training = np.array(training)

train_x = list(training[:, 0])
train_y = list(training[:, 1])

tf.reset_default_graph()
net = tflearn.input_data(shape=[None, len(train_x[0])])
net = tflearn.fully_connected(net, 8)
net = tflearn.fully_connected(net, 8)
net = tflearn.fully_connected(net, len(train_y[0]), activation='softmax')
net = tflearn.regression(net)

model = tflearn.DNN(net, tensorboard_dir=path.getPath('train_logs'))
model.fit(train_x, train_y, n_epoch=20000, batch_size=500, show_metric=True)
model.save(path.getPath('model.tflearn'))
```

Appendix E Using the generated model during conversation

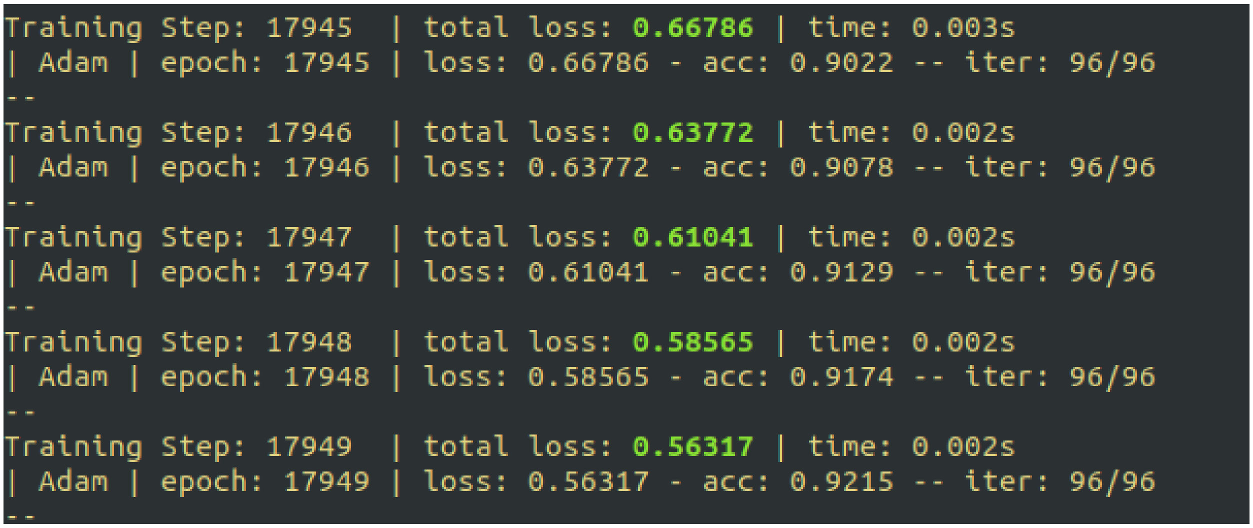


```
        data = pickle.load(open(path.getPath('trained_data'), "rb"))
        self.words = data['words']
        self.classes = data['classes']
        train_x = data['train_x']
        train_y = data['train_y']
        with open(path.getJsonPath()) as json_data:
            self.intents = json.load(json_data)
        net = tflearn.input_data(shape=[None, len(train_x[0])])
        net = tflearn.fully_connected(net, 8)
        net = tflearn.fully_connected(net, 8)
        net = tflearn.fully_connected(net, len(train_y[0]), activation='softmax')
        net = tflearn.regression(net)
        self.model = tflearn.DNN(net, tensorboard_dir=path.getPath('train_logs'))
        self.model.load(path.getPath('model.tflearn'))

    def clean_up_sentence(self, sentence):
        sentence_words = nltk.word_tokenize(sentence)
        sentence_words = [self.stemmer.stem(word.lower()) for word in sentence_words]
        return sentence_words

    def bow(self, sentence, words, show_details=False):
        sentence_words = self.clean_up_sentence(sentence)
        bag = [0] * len(words)
        for s in sentence_words:
            for i, w in enumerate(words):
                if w == s:
                    bag[i] = 1
                    if show_details:
                        print("found in bag: %s" % w)
        return np.array(bag)

    def classify(self, sentence):
        ERROR_THRESHOLD = 0.25
        results = self.model.predict([self.bow(sentence, self.words)])[0]
        results = [[i, r] for i, r in enumerate(results) if r > ERROR_THRESHOLD]
```

Appendix F Training the conversational AI

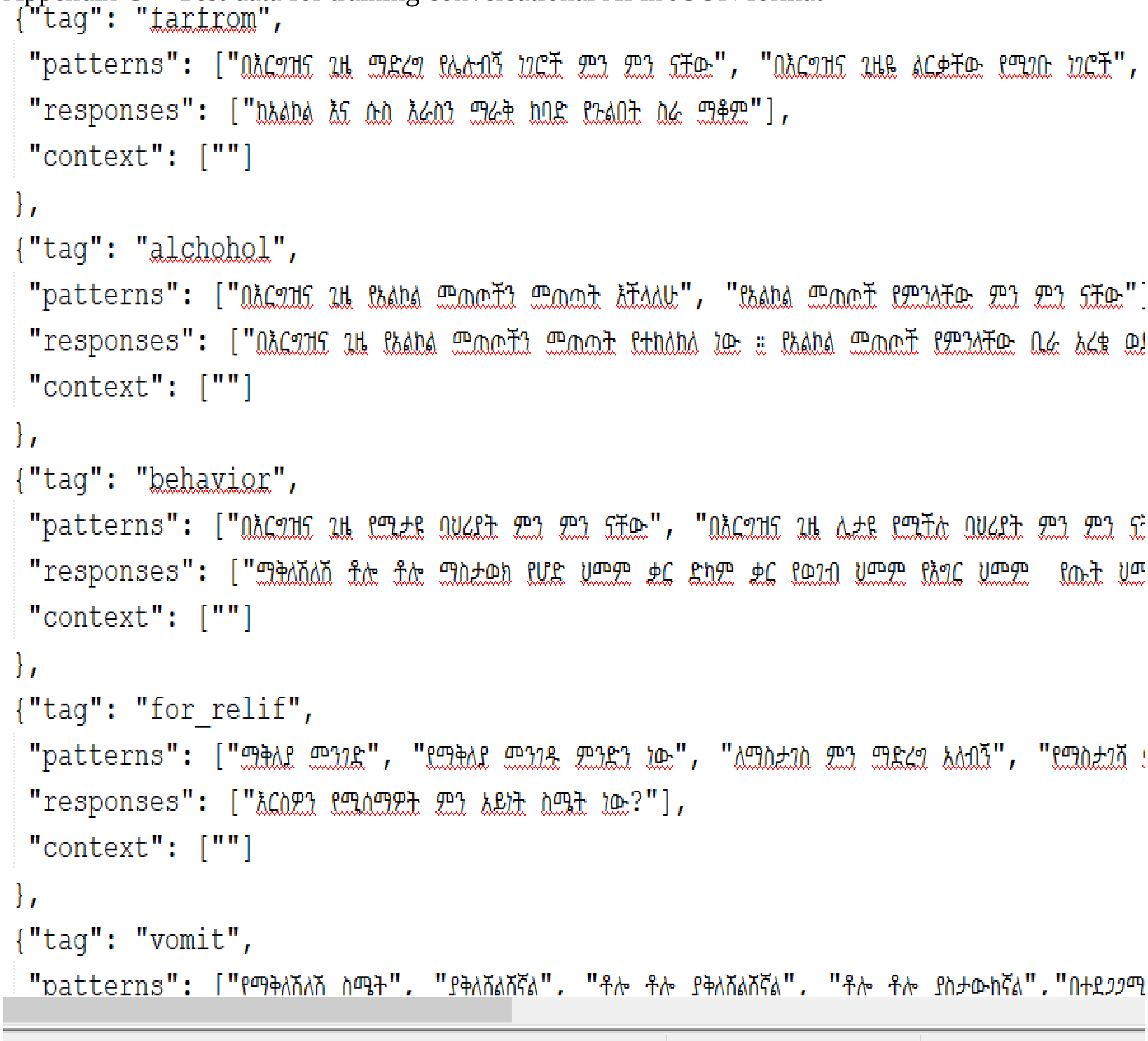


```
--
Training Step: 4416  | total loss: 1.70863 | time: 0.005s
| Adam | epoch: 4416 | loss: 1.70863 - acc: 0.7613 -- iter: 96/96
--
Training Step: 4417  | total loss: 1.61780 | time: 0.002s
| Adam | epoch: 4417 | loss: 1.61780 - acc: 0.7810 -- iter: 96/96
--
Training Step: 4418  | total loss: 1.95251 | time: 0.002s
| Adam | epoch: 4418 | loss: 1.95251 - acc: 0.7102 -- iter: 96/96
--
Training Step: 4419  | total loss: 1.83710 | time: 0.002s
| Adam | epoch: 4419 | loss: 1.83710 - acc: 0.7350 -- iter: 96/96
--
Training Step: 4420  | total loss: 1.73315 | time: 0.002s
| Adam | epoch: 4420 | loss: 1.73315 - acc: 0.7573 -- iter: 96/96
--
Training Step: 4421  | total loss: 1.63949 | time: 0.002s
| Adam | epoch: 4421 | loss: 1.63949 - acc: 0.7774 -- iter: 96/96
--
Training Step: 4422  | total loss: 1.94166 | time: 0.002s
| Adam | epoch: 4422 | loss: 1.94166 - acc: 0.7070 -- iter: 96/96
--
Training Step: 4423  | total loss: 1.82695 | time: 0.002s
| Adam | epoch: 4423 | loss: 1.82695 - acc: 0.7321 -- iter: 96/96
--
Training Step: 4424  | total loss: 2.13747 | time: 0.002s
| Adam | epoch: 4424 | loss: 2.13747 - acc: 0.6631 -- iter: 96/96
--
Training Step: 4425  | total loss: 2.00313 | time: 0.002s
| Adam | epoch: 4425 | loss: 2.00313 - acc: 0.6926 -- iter: 96/96
--
Training Step: 4426  | total loss: 2.29994 | time: 0.002s
| Adam | epoch: 4426 | loss: 2.29994 - acc: 0.6275 -- iter: 96/96
--
Training Step: 4427  | total loss: 2.14944 | time: 0.001s
| Adam | epoch: 4427 | loss: 2.14944 - acc: 0.6606 -- iter: 96/96
--
```



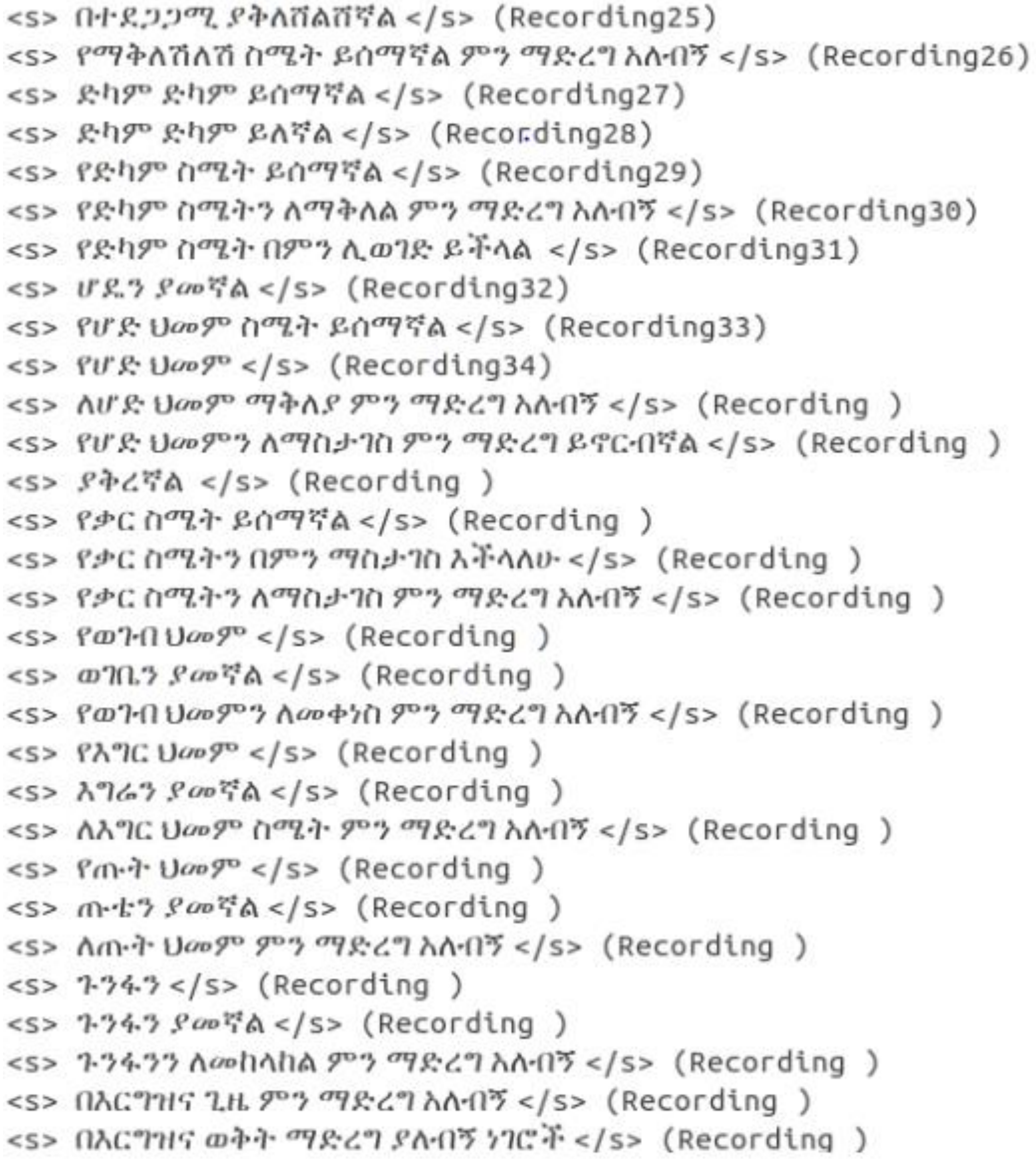
```
Training Step: 17945  | total loss: 0.66786 | time: 0.003s
| Adam | epoch: 17945 | loss: 0.66786 - acc: 0.9022 -- iter: 96/96
--
Training Step: 17946  | total loss: 0.63772 | time: 0.002s
| Adam | epoch: 17946 | loss: 0.63772 - acc: 0.9078 -- iter: 96/96
--
Training Step: 17947  | total loss: 0.61041 | time: 0.002s
| Adam | epoch: 17947 | loss: 0.61041 - acc: 0.9129 -- iter: 96/96
--
Training Step: 17948  | total loss: 0.58565 | time: 0.002s
| Adam | epoch: 17948 | loss: 0.58565 - acc: 0.9174 -- iter: 96/96
--
Training Step: 17949  | total loss: 0.56317 | time: 0.002s
| Adam | epoch: 17949 | loss: 0.56317 - acc: 0.9215 -- iter: 96/96
--
```

Appendix G Test data for training conversational AI in JSON format



```
{"tag": "tartrom",
 "patterns": ["በእርግዝና ጊዜ ማድረግ የሌለብኝ ነገሮች ምን ምን ናቸው", "በእርግዝና ጊዜዬ ልርቃቸው የሚገቡ ነገሮች",
 "responses": ["ከአልኮል እና ሱስ እራስን ማራቅ ከባድ የጉልበት ስራ ማቆም"],
 "context": [""]
},
{"tag": "alchohol",
 "patterns": ["በእርግዝና ጊዜ የአልኮል መጠጦችን መጠጣት እችላለሁ", "የአልኮል መጠጦች የምንላቸው ምን ምን ናቸው"]
 "responses": ["በእርግዝና ጊዜ የአልኮል መጠጦችን መጠጣት የተከለከለ ነው ። የአልኮል መጠጦች የምንላቸው ቢራ አረቄ ወ
 "context": [""]
},
{"tag": "behavior",
 "patterns": ["በእርግዝና ጊዜ የሚታዩ ባህሪያት ምን ምን ናቸው", "በእርግዝና ጊዜ ሊታዩ የሚችሉ ባህሪያት ምን ምን ና
 "responses": ["ማቅለሽለሽ ቶሎ ቶሎ ማስታወክ የሆድ ህመም ቃር ድካም ቃር የወገብ ህመም የእግር ህመም  የጡት ህ
 "context": [""]
},
{"tag": "for_relif",
 "patterns": ["ማቅለያ መንገድ", "የማቅለያ መንገዱ ምንድን ነው", "ለማስታገስ ምን ማድረግ አለብኝ", "የማስታገሻ
 "responses": ["እርስዎን የሚሰማዎት ምን አይነት ስሜት ነው?"],
 "context": [""]
},
{"tag": "vomit",
 "patterns": ["የማቅለሽለሽ ስሜት", "ያቅለሽልሸኛል", "ቶሎ ቶሎ ያቅለሽልሸኛል", "ቶሎ ቶሎ ያስታውከኛል","በተደጋጋሚ
```

Appendix H transcribed text and its corresponding audio record



```
<s> በተደጋጋሚ ያቅለሸልሸኛል </s> (Recording25)
<s> የማቅለሽለሽ ስሜት ይሰማኛል ምን ማድረግ አለብኝ </s> (Recording26)
<s> ድካም ድካም ይሰማኛል </s> (Recording27)
<s> ድካም ድካም ይለኛል </s> (Recording28)
<s> የድካም ስሜት ይሰማኛል </s> (Recording29)
<s> የድካም ስሜትን ለማቅለል ምን ማድረግ አለብኝ </s> (Recording30)
<s> የድካም ስሜት በምን ሊወገድ ይችላል  </s> (Recording31)
<s> ሆዴን ያመኛል </s> (Recording32)
<s> የሆድ ህመም ስሜት ይሰማኛል </s> (Recording33)
<s> የሆድ ህመም </s> (Recording34)
<s> ለሆድ ህመም ማቅለያ ምን ማድረግ አለብኝ </s> (Recording )
<s> የሆድ ህመምን ለማስታገስ ምን ማድረግ ይኖርብኛል </s> (Recording )
<s> ያቅረኛል  </s> (Recording )
<s> የቃር ስሜት ይሰማኛል </s> (Recording )
<s> የቃር ስሜትን በምን ማስታገስ እችላለሁ </s> (Recording )
<s> የቃር ስሜትን ለማስታገስ ምን ማድረግ አለብኝ </s> (Recording )
<s> የወገብ ህመም </s> (Recording )
<s> ወገቤን ያመኛል </s> (Recording )
<s> የወገብ ህመምን ለመቀነስ ምን ማድረግ አለብኝ </s> (Recording )
<s> የእግር ህመም </s> (Recording )
<s> እግሬን ያመኛል </s> (Recording )
<s> ለእግር ህመም ስሜት ምን ማድረግ አለብኝ </s> (Recording )
<s> የጡት ህመም </s> (Recording )
<s> ጡቴን ያመኛል </s> (Recording )
<s> ለጡት ህመም ምን ማድረግ አለብኝ </s> (Recording )
<s> ጉንፋን </s> (Recording )
<s> ጉንፋን ያመኛል </s> (Recording )
<s> ጉንፋንን ለመከላከል ምን ማድረግ አለብኝ </s> (Recording )
<s> በእርግዝና ጊዜ ምን ማድረግ አለብኝ </s> (Recording )
<s> በእርግዝና ወቅት ማድረግ ያለብኝ ነገሮች </s> (Recording )
```