ENHANCING CONVERSATIONAL AI MODEL PERFORMANCE AND EXPLAINABILITY FOR SINHALA-ENGLISH BILINGUAL SPEAKERS

2022-056

Project Proposal Report

Dissanayake D.M.I.M.
(Hameed M.S., Jayasinghe D.T., Sakalasooriya S.A.H.A.)

B.Sc. (Hons) Degree in Information Technology Specialising in Data Science

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

January 2022

ENHANCING CONVERSATIONAL AI MODEL PERFORMANCE AND EXPLAINABILITY FOR SINHALA-ENGLISH BILINGUAL SPEAKERS

2022-056

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology Specialising in Data Science

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

January 2022

DECLARATION, COPYRIGHT STATEMENT AND THE STATEMENT OF THE SUPERVISORS

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Dissanayake D.M.I.M.	IT19069432	
Hameed M.S.	IT19064932	
Jayasinghe D.T.	IT19075754	
Sakalasooriya S.A.H.A.	IT19051208	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Name of the supervisor: Dr. Lakmini Abeywardhana
Signature of the supervisor: Date: 11/02/2022

Name of the co-supervisor: Ms. Dinuka Wijendra
Signature of the co-supervisor: Date: 11/02/2022

ABSTRACT

Machine learning model explainability and interpretability is an emerging research interest among Machine Learning and Artificial Intelligence researchers, and this field is known as Explainable Artificial Intelligence (XAI). It refers to the ability to explain how machine learning models make decisions. With the advancement of machine learning, it is often ambiguous to humans how state-of-the-art machine learning models give predictions. For instance, most machine learning models are black-box models, and how models like transformers generate predictions is often puzzling and even confusing. Explainability of the model predictions is considered crucial for various reasons, including identifying if the models have any biases, increasing the trustworthiness of the models, and finetuning or debugging the models. Explainable artificial intelligence has made its way towards the domain of natural language processing with the recent research conducted such as “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. This research component focuses on developing an explainable artificial intelligence algorithm called “DIME: dual interpretable model-agnostic explanations” by enhancing existing explainable AI techniques and utilizing the developed method to explain the predictions given by the “DIET Classifier”: the transformer-based deep learning text classification model used in Rasa conversational AI assistants. DIME algorithm focuses on using global explanations of the model to derive local model explanations, hence the word “dual” in DIME.

Keywords: Natural Language Processing, Explainable AI, DIET Classifier, DIME, Conversational AI

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
AI	Artificial Intelligence
API	Application Programming Interface
CLI	Command Line Interface
CPU	Central Processing Unit
DIET	Dual Intent Entity Transformer
DIME	Dual Interpretable Model-Agnostic Explanations
GDPR	General Data Protection Regulation
GFI	Global Feature Importance
GUI	Graphical User Interface
GPU	Graphics Processing Unit
IO	Input/Output
IT	Information Technology
LIME	Local Interpretable Model-Agnostic Explanations
LFC	Local Feature Contribution
NLP	Natural Language Processing
NLU	Natural Language Understanding
NLG	Natural Language Generation
PDP	Partial Dependency Plot
RAM	Random Access Memory
SLIIT	Sri Lanka Institute of Information Technology
SHAP	Shapley Additive Explanations
TPU	Tensor Processing Unit
UI	User Interface
XAI	Explainable Artificial Intelligence
CDD	Conversation Driven Development
SaaS	Software-as-a-service
CaaS	Conversational AI-as-a-Service

1. INTRODUCTION

1.1 Background & Literature survey

Many top-tier applications and devices utilize Natural Language Processing (NLP) in this era of Artificial Intelligence. Almost all domains have adopted NLP in ample ways to perform Natural Language Understanding (NLU), Natural Language Generation (NLG), or both. One such popular application of NLP is chatbots, and its name has gradually become Conversational Artificial Intelligence (Conversational AI) after chatbots started to utilize artificial intelligence to generate highly natural human-like dialogs and conversations instead of executing hard-coded rules and pattern matching. Artificial Intelligence chatbots or conversational AIs use advanced machine learning models that train on human conversation data, and hence they are capable of handling complex queries presented to them and possess the ability to have a more natural conversation [3]. Traditional NLP depended on inherently explainable techniques, also known as white-box techniques, such as rules-based approaches and decision tree classifiers. However, with the rapid advancement of machine learning and deep learning models, their human interpretability has decreased over time due to their complicated nature. It is known as the trade-off between model explainability and accuracy [4]. These less-interpretable models and techniques are known as black-box models/black-box methods, such as word embeddings, encoder-decoder architecture, and other deep learning-based approaches [5].

Interpretability and explainability are two terms in Explainable AI that are considered interchangeable but can often be confusing since they have two distinct definitions [6]. [7]. An interpretation is the process of mapping abstract concepts to a domain that humans can comprehend, while an explanation is the set of features that have contributed towards a specific decision made [4]. [8]. Explainability is the term mainly considered in Explainable AI since it focuses on generating explanations for the predictions made by non-transparent excessively complicated models to identify the set of features that contributed the most to the given prediction. Attempting to interpret such models without XAI might not always be practical. Although explanations are a part of interpretations, explanations alone do not always provide consistent model interpretations. As mentioned in [6]. and [8]. transparency of white-box models is a property of interpretability and not explainability so, the two terms must not be confused with one another. In general, less complicated models are well interpretable, while complex models require explanations generated to provide an acceptable interpretation of how the model gives predictions.

There are a few ways of implementing XAI. The main two approaches are known as intrinsic/self-explaining and post hoc. Intrinsic/self-explaining models are white-box models that are interpretable by design, while post hoc refers to generating explanations for existing black-box models [4]. Moreover, XAI approaches can be model-specific, model-agnostic, local interpretable, or global-interpretable [9]. Model-specific XAI techniques apply to a specific family of machine learning models, while model-agnostic XAI approaches are suitable for various types of models available. Recent research conducted in the XAI domain based on NLP has developed model-agnostic tools such as LIME [1]. and SHAP [10]. that have proven their value in contrast to the model-specific approaches. Different machine learning model types use these tools to generate model explanations, although their internal structure is diverse. Local interpretable explanation generation focuses on explaining a single data instance. Especially in NLP, for a data instance such as a sentence, paragraph, or question from a text corpus used to train models such as text classifiers can get explanations only considering the data instance itself but discarding how the model behaves for other data. This approach generates more data instances close to the original data instance based on distance matrices such as cosine similarity and trains an interpretable surrogate model that acts as a white-box model using those newly obtained data and text classifier model predictions for those data instances [1]. [11]. . Shapley values [15]. are significant when calculating the contribution of each feature while deriving explanations. The research paper [10]. explains how Shapley values assist in calculating the individual contribution of features. In the global interpretable explanation generation approach, all data points the model trained on are considered. In other words, in global interpretable explanations generation, the behavior of the whole model is considered [9]. . Feature importance is one of the properties that can interpret a model globally. It uncovers how significant a given feature is to a fully trained model considering all predictions done. There are numerous ways to calculate the global feature importance including, permutation feature importance and partial dependency plot-based (PDP) feature importance [16].

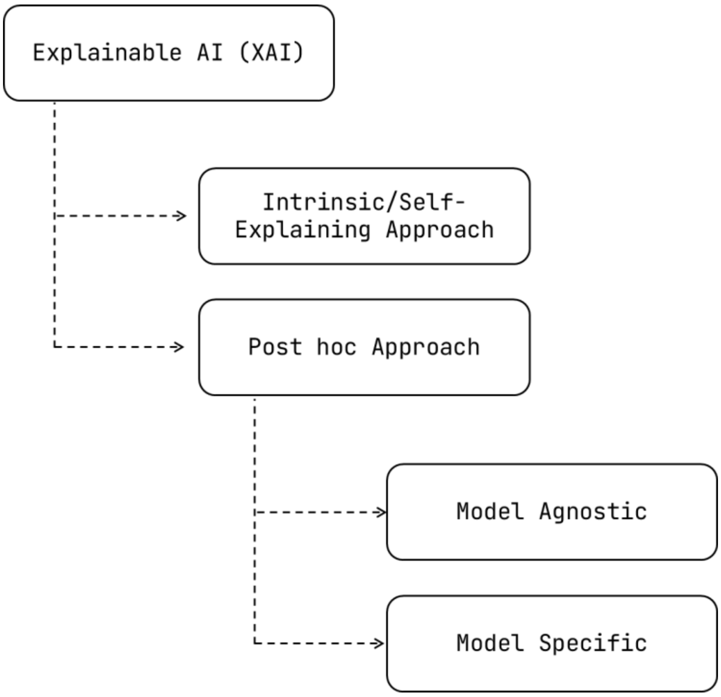


Figure 1.1: Summarized Explainable AI approaches

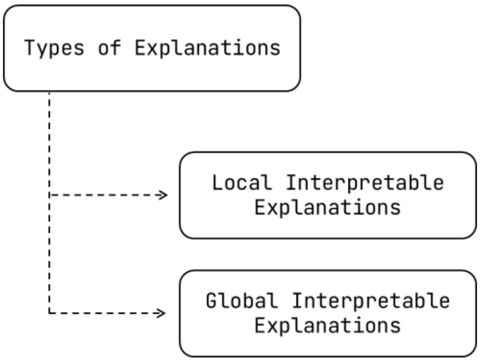

Types of Explanations
Local Interpretable Explanations
Global Interpretable Explanations

Figure 1.2: Summarized types of model explanations

There are several reasons why Explainable AI should be a requirement rather than an optional tool in the AI domain. One of the reasons is to get a clear idea of how a machine learning model makes predictions. The generated explanations can aid in constructing feature importance visualizations and model interpretations, and it is crucial to enhance the human understandability of why the model predicted what it predicted. In [4]. , the research concludes how trusting black-box machine learning models has led to a pressing issue, why human interpretability for machine learning models is essential, and how XAI can help bridge the gap between interpretability and accuracy. New regulations like GDPR have introduced a right to explanation due to the rise of opaque and complex algorithms to enforce trust through acceptable and clear explanations. Another reason to generate model explanations is to validate complex machine learning models. Since tools such as LIME [1]. and SHAP [10]. visualize the contribution of the features towards a prediction made by the model, model explanations make it possible to validate if the model has considered acceptable features to make predictions. Moreover, it can identify any biases the model contains, and this process makes it possible to debug machine learning models. While there can be many other reasons why complex models should be made explainable, the reasons mentioned above should be more than enough to notice why model explainability is a must.

1.2 Research Gap

Researchers in the XAI domain usually select one or many XAI approaches (Figure 1.1) and one or many out of types of explanations (Figure 1.2). Numerous research papers seem to have focused on explaining black-box models. Every XAI research that has introduced a tool has also comprised visualization techniques to display the contributions of the features. Recently published literature survey papers such as [3]. -[5]. , [8]. , and [9]. are immensely helpful for a researcher to get an idea about the existing research gap.

Recent research carried out in the XAI domain that covers the area of NLP, including LIME [1]. [11]. and SHAP[10]. , has incorporated various explainability techniques to calculate and visualize local and global explanations. Similar research papers to the above-specified research papers have been listed in Table 1.1 in contrast to the methods incorporated in this research component to define the research gap [12]. -[14].

Table 1.1: Comparison of XAI research done that applies to the domain of NLP and methodology used in contrast to this research.

Research Name	**Can be utilized for NLP**	**Intrinsic/ Post hoc**	**XAI Scope (Local/ Global)**	**Model-specificity**	**Visualization Technique**	**Feature contribution calculation**
LIME	Yes	Post hoc	Local	Agnostic	Highlighted text + Raw contribution value plots	Cosine Distance + Local linear surrogate model + Ridge Regression
SHAP	Yes	Post hoc	Local or Global	Agnostic	Highlighted text + Raw SHAP value plots	LIME, DeepLIFT, and other approaches with SHAP values.
Deep LIFT	Yes (Deep SHAP)	Post hoc	Local	Specific	Refer DeepSHAP [10].	Back-propagating contributions
Self-Explain	Yes	Self-explaining	Local	Specific	Highlighted features	Regularization with explanation specific losses
DIME *(This Research)*	Yes	Post hoc	**Local** with aid of **Global** feature importance	Agnostic	Highlighted features + Raw contribution score plots	**Global feature importance** + Shapley Values based on **model confidence**

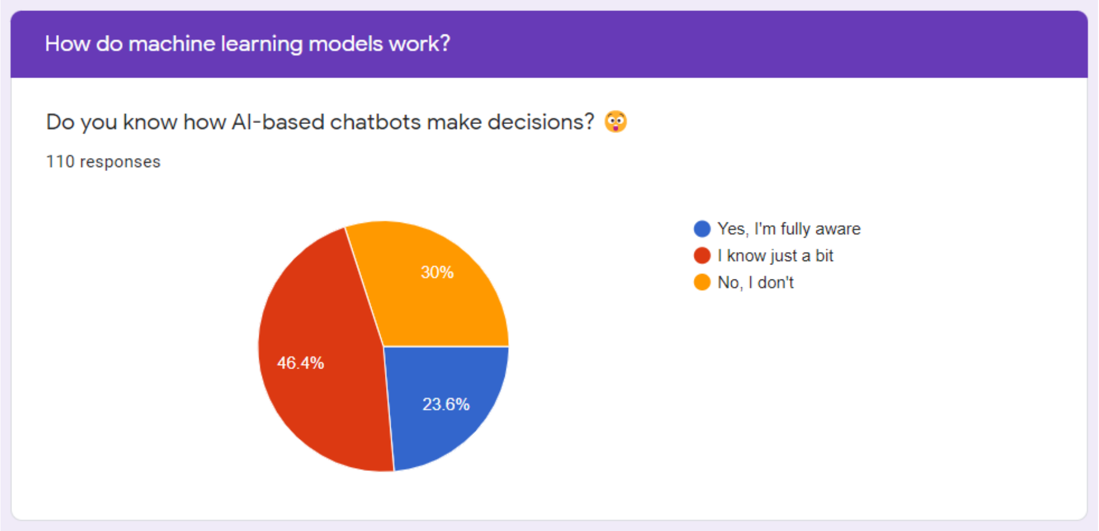
As demonstrated in the Table 1.1, there are similar research papers in contrast to this research in terms of applicability for NLP, Intrinsic/Post hoc, Model Specificity, and Visualization techniques. However, this research focuses on a scope of explanations that is notably different from the methods described in other research papers. Here, the local model explanations will consider the impact of global feature importance, both as a weighting factor and a feature selection criterion, when generating model explanations. The other specialty is that the method discussed in this research will consider model confidence given by the DIET classifier when calculating the feature contributions towards the predictions made by the models.

1.3 Research Problem

Algorithms and machine learning models are getting complex every second. Even widely used applications such as chatbots have employed machine learning and NLP due to their high performance and accuracy compared to traditional approaches. As previously explained, these complex algorithms and advanced machine learning techniques, such as deep learning, are opaque by design, and humans are having a hard time interpreting how these work under the hood and making predictions [4]. . Although many XAI-based explainers exist, they cannot generate explanations for all machine learning model types. It is sometimes impossible to uncover model explanations for application-specific machine learning models, even with model-agnostic explainers such as LIME [1]. and SHAP [10]. The machine learning model inside the chatbot framework Rasa is an example of such a model. The models trained and stored within Rasa contain a cluster of machine learning models, NLP pipeline components, and metadata. Rasa NLU is responsible for extracting the model, loading the pipeline components, and processing text data. Generating model explanations for text classifiers such as the DIET classifier is tedious since Rasa does not provide a way to examine it. Extracting prediction probabilities from DIET is also not directly supported, which is required by many model-agnostic text explainers [2]. [17]. However, it is better if there is a way to interpret these application-specific machine learning models by its users to trust the predictions given by the models used within these widely used frameworks.

Many text explainers in the XAI domain focuses on generating either local or global model explanations. Almost none of the research has considered that these two kinds of model explanations can be blended to derive accurate local model explanations [5]. There is a possibility that using global feature importance might improve the quality of local model explanations, though it remains undiscovered. This research concentrates on finding the impact of global feature importance on local model explanations.

A recent survey that studied the current trends in machine learning model interpretability and explainability confirms that the majority prefer having explanations for the machine learning model predictions. The population of interest of the survey was 110 Undergraduates of the Faculty of Computing of Sri Lanka Institute of Information Technology, and it was decided based on the technical nature of the survey questions. Refer to **Error! Reference source not found.** to observe the complete list of survey questions and responses.



How do machine learning models work?

Do you know how AI-based chatbots make decisions? 😳

110 responses

Figure 1.3: Summary of survey responses received for the question "Do you know how AI-based chatbots make decisions?”

Responses in the Figure 1.3: Summary of survey responses received for the question "Do you know how AI-based chatbots make decisions?” clearly shows that more than three-fourths of the entire population have little to no understanding of the process of natural language understanding techniques used within chatbot frameworks. Figure 1.4: Summary of the responses received for the survey question "Are you interested to know how machine learning/ deep learning models make decisions?" indicates that 91.8% of the whole population prefers having explanations for the predictions made by machine learning models. These results confirm that there is indeed a demand for model prediction explanations and humans prefer interpretable model explanations.

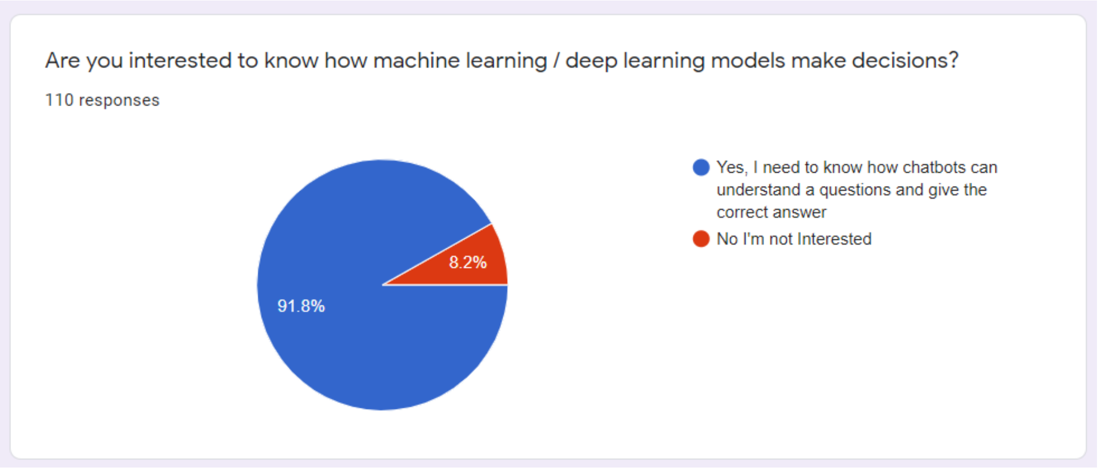


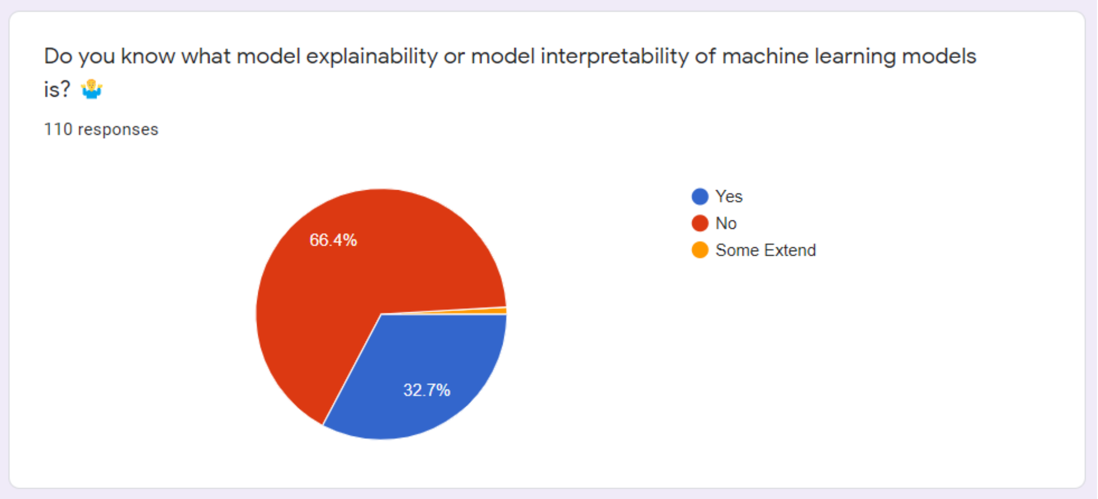
Figure 1.4: Summary of the responses received for the survey question "Are you interested to know how machine learning/ deep learning models make decisions?"

Although 32.7% of the population has claimed (**Error! Reference source not found.**) that they are aware of the terms model interpretability and explainability, **Error! Reference source not found.** clearly illustrates that only 26.4% have used text explainer tools such as LIME and SHAP. 66.4% of the total population are unaware of model explainability and interpretability, and this is due to the following reasons.

1. Deriving explanations and achieving interpretability requires additional work. XAI tools are not a part of the machine learning toolkits of machine learning-based applications by design. For instance, machine learning-enabled chatbots do not contain any text explainers. Users must generate explanations by themselves manually, which requires machine learning expertise.

2. Machine learning models created by some modern AI-enabled applications are different from a usual machine learning model. For instance, in the conversational AI framework, Rasa, the machine learning model is a collection of machine learning models and components compressed into a single ".bin" file rather than a single model. Especially in Rasa, a user should know how to find the classification model files in the zipped model files and the classification model architecture to generate model explanations. Thus, it is not easy to use the existing model-agnostic text explainers. It is preferable if there

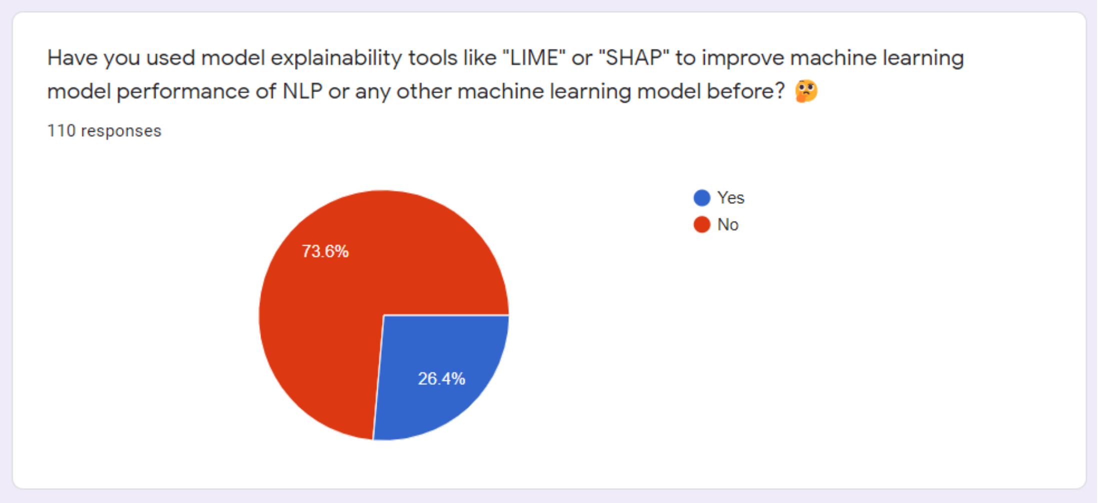
is a universal way to get model explanations without manually inspecting the zipped model of machine learning-based frameworks such as Rasa.



Do you know what model explainability or model interpretability of machine learning models is? 🤷‍♂️

110 responses

Figure 1.5: Summary of the responses received for the survey question "Do you know what model explainability or model interpretability of machine learning models is?"



Have you used model explainability tools like "LIME" or "SHAP" to improve machine learning model performance of NLP or any other machine learning model before? 🤔

110 responses

Figure 1.6: Summary of the responses received for the survey question "Have you used model explainability tools”

2. OBJECTIVES

2.1 Main Objectives

The main objective of the overall research project is to develop NLP tools for text preprocessing, feature engineering, training data and machine learning model improvements, model performance evaluation, and model explanation generation on DIET intent classifiers; that can be attached to a domain-specific Rasa conversational AI assistant designed for Sinhala-English code-switched text corpus.

The main objective of this individual research component is to develop an XAI algorithm, DIME (dual interpretable model-agnostic explanations), to deliver local model explanations with the help of global feature importance. As previously explained, currently, there is little to no strategy exists that tries to aid "global feature importance" to derive local model explanations. This implementation concentrates on deriving global feature importance for all tokens in the training data to investigate how important a word is to the fully trained model and utilizing these scores to derive local model-specific explanations.

2.2 Specific Objectives

Specific objectives of this research component can be listed down as follows.

1. Develop a strategy to calculate the global feature importance for individual features and normalize the scores in a logically explainable manner.

2. Develop a strategy to incorporate global feature importance to derive local model-agnostic explanations for the DIET classifier in Rasa framework.

3. Integrate the above strategies as a modular python package that can be easily installed.

4. Develop a visualization technique to illustrate the local text explanations with the contribution scores towards the prediction in a human-interpretable way.

5. Build a Server with API endpoints that can be consumed by the frontend applications and integrate model-agnostic explanation visualizations with the conversational AI maintenance frontend.

6. Integrate the XAI approach developed with the Rasa framework seamlessly and allow Rasa users to get model explanations for the DIET classifier.

The Methodology chapter (chapter **Error! Reference source not found.**) of this report explains the main objective and specific objectives mentioned above in detail.

3. METHODOLOGY

3.1 Requirements Gathering and Analysis

The requirements gathering phase mainly focused on studying existing research on XAI concepts for NLP and the necessity of an XAI component for machine learning-based products, especially chatbot frameworks such as Rasa. After identifying the current requirement for an XAI component to interpret machine learning models used in modern conversational AI frameworks, the survey mentioned in section 1.3 was conducted to study the following regarding this research component.

1. The percentage of users who understand how modern artificial intelligence-based chatbots work.
2. The percentage of users who would like to know how machine learning models make predictions when responding to user queries

3.1.1 Functional requirements

The functional requirements of the XAI approach, DIME, are as follows.

1. The proposed XAI approach DIME should logically calculate the global feature importance for any given fully trained DIET classifier.
2. DIME should normalize global feature importance scores between an acceptable range.
3. DIME should generate local model explanations using global feature importance scores as a part of its algorithm.
4. DIME should consider global feature importance to perform feature selection when generating local model explanations.
5. DIME should be applicable to any machine learning text classification model that outputs confidence scores for predictions.
6. DIME should ask the users number of features to generate explanations to reduce the number of calculations drastically.
7. DIME should visualize locally generated model explanations in a human-interpretable manner.
8. DIME should be seamlessly combinable with Rasa conversational AI assistants to generate explanations for non-technical users.

3.1.2 Non-functional requirements

The non-functional requirements of DIME are as follows.

1. DIME should perform calculations efficiently.
2. DIME should provide reliable local model explanations.
3. DIME should provide simple and easy-to-interpret model explanation visualizations.
4. DIME algorithm should be modular.

3.2 Feasibility Study

A feasibility study was carried out to investigate the technical, financial, legal, operational, and scheduling feasibility of the research component.

3.2.1 Technical feasibility

The research component requires having in-depth knowledge in Rasa framework, Rasa machine learning models, data science, and software development. There should be enough computing resources to train Rasa machine learning models. Training machine learning models in Rasa does not require advanced computing resources such as GPUs, TPUs, or High RAM/CPU instances, although it is better if such resources are available. Hard drive space should be a foremost concern as the Rasa framework may need a fair amount of disk space for caching model files. There should be a properly configured cloud infrastructure for the overall deployment of the final product.

3.2.2 Financial feasibility

The cloud infrastructure for training machine learning models or deployment should not be costly. There can be minor charges for domain name purchasing and cloud infrastructure resource usage, and they are acceptable since the final product contains a market value that can cover the costs involved.

3.2.3 Legal feasibility

Datasets used in the research components should be publicly available datasets. Any scraped data from websites that are not public should have the approval of the original owners and should have written consent. The research component should not extract content as it is from previous work done without the written permission of the original authors. Any python packages or other software used must be open source or legally purchased, and the original authors must be credited if requested.

3.2.4 Operational feasibility

This research component must not conflict with other research components of the same research project or other research projects. The research component should address the gap in the explainability of the DIET classifier and should preserve novelty in contrast to the previous work done. The research component should be viable to fulfill the scope stated by the research project requirements.

3.2.5 Scheduling feasibility

The tasks of the research component should adhere to the Gantt chart mentioned in section 5 of this proposal report and the pre-defined research project milestones.

3.3 Preparation of Datasets

All research components require two datasets in total. Both datasets are domain-specific datasets with Sinhala-English code-switched text data. However, there are notable differences between the datasets and subsections 3.3.1 and 3.3.2 explain these differences clearly. Both datasets utilize data augmentation techniques mainly to overcome the low-resource nature of Sinhala text data gathered, capture as many as code-switched phrases and collect as many distinct writing patterns as possible. All research group members will generate four versions of the samples in both datasets according to different code-switching styles. Duplicate data removal will be employed to ensure the quality of the collected data.

This research component only utilizes the domain-specific dataset mentioned in subsection 3.3.2. to train a Rasa conversational AI model since developing DIME only requires a fully trained Rasa model.

3.3.1 General dataset for machine learning model training

The first domain-specific dataset for machine learning model training and NLP text pre-processing tool development contains scraped Sinhala-English code-switched textual data from websites related to SLIIT and from news articles. (https://support.sliit.lk/, https://sliitinternational.lk/, and https://www.sliit.lk/ are few such websites). In addition to that, publicly available documents, such as PDFs and Docx files taken from the above SLIIT-based websites since they are both public and official. Dataset will utilize data augmentation techniques to overcome the low-resource issue.

3.3.2 Domain specific dataset for conversational AI training

The second domain-specific dataset for training the Rasa-based conversational AI contains handcrafted Sinhala-English code-switched textual data. This dataset should be designed carefully as a training dataset for the intent classification task of conversational AI according to the guidelines provided by Rasa. As in the previous case, the second dataset also will utilize data augmentation techniques to overcome the low-resource issue. There will be around 78 intents (classes), and each class initially contains a minimum of 10 examples as a standard. Note that the number of training examples may increase in the future to increase the overall performance of conversational AI.

3.4 Individual Component Architecture

The implementation of DIME, the XAI approach developed in this research component, will contain several sub-components, as illustrated in Figure 3.1.

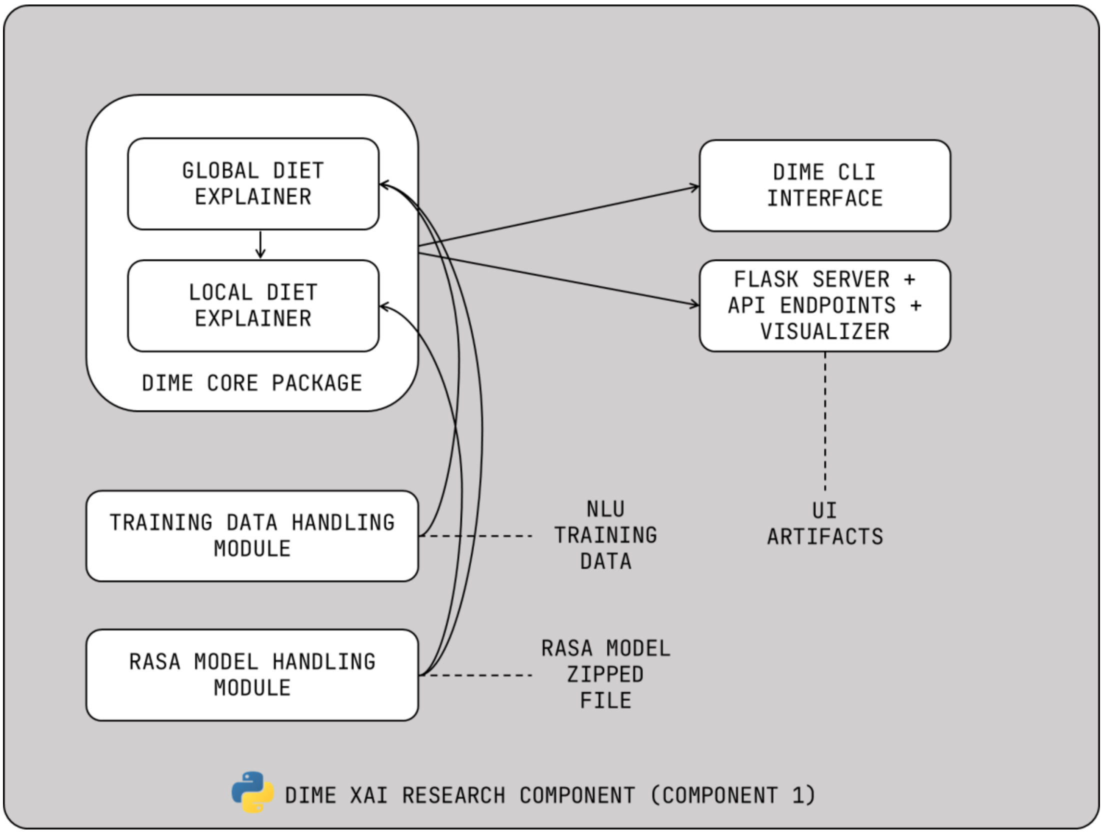

GLOBAL DIET EXPLAINER
LOCAL DIET EXPLAINER
DIME CORE PACKAGE
DIME CLI INTERFACE
FLASK SERVER + API ENDPOINTS + VISUALIZER
TRAINING DATA HANDLING MODULE
NLU TRAINING DATA
UI ARTIFACTS
RASA MODEL HANDLING MODULE
RASA MODEL ZIPPED FILE
DIME XAI RESEARCH COMPONENT (COMPONENT 1)

Figure 3.1: High-level architectural diagram of DIME

The core python package of DIME will contain two sub-packages to calculate the global feature importance and generate local model explanations. The "local explainer" module will use the results generated by the "global explainer" module. The global explainer will utilize the data and model handling modules to perform the required IO operations. The model handler will execute model loading, and the data handler module will handle training data-related operations. DIME will consist of a CLI interface and a local server built with Flask to reveal visualizations for the explanations generated. CLI tool and Notebooks will only get raw values instead of visualizations, leaving room for further improvements to the DIME module.

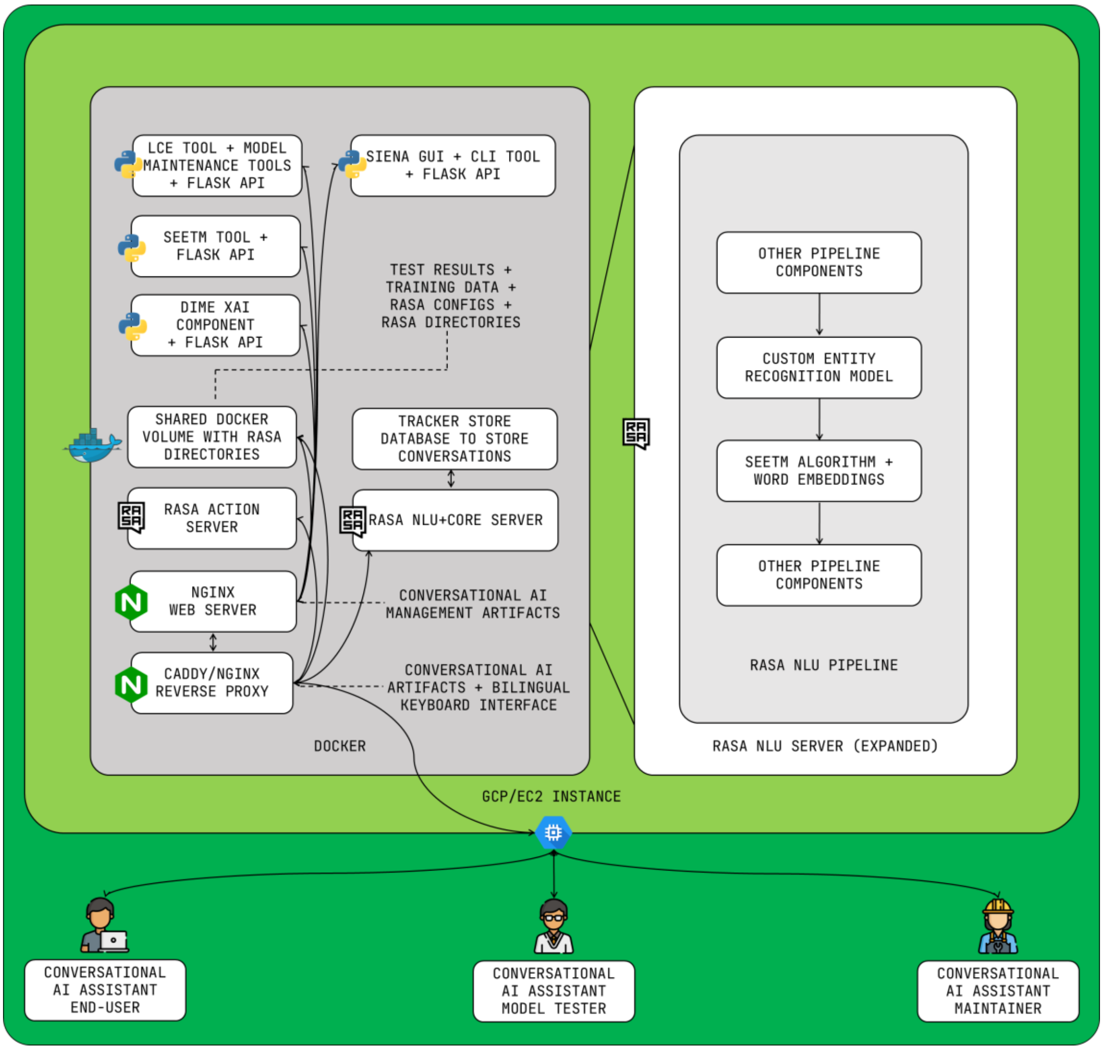
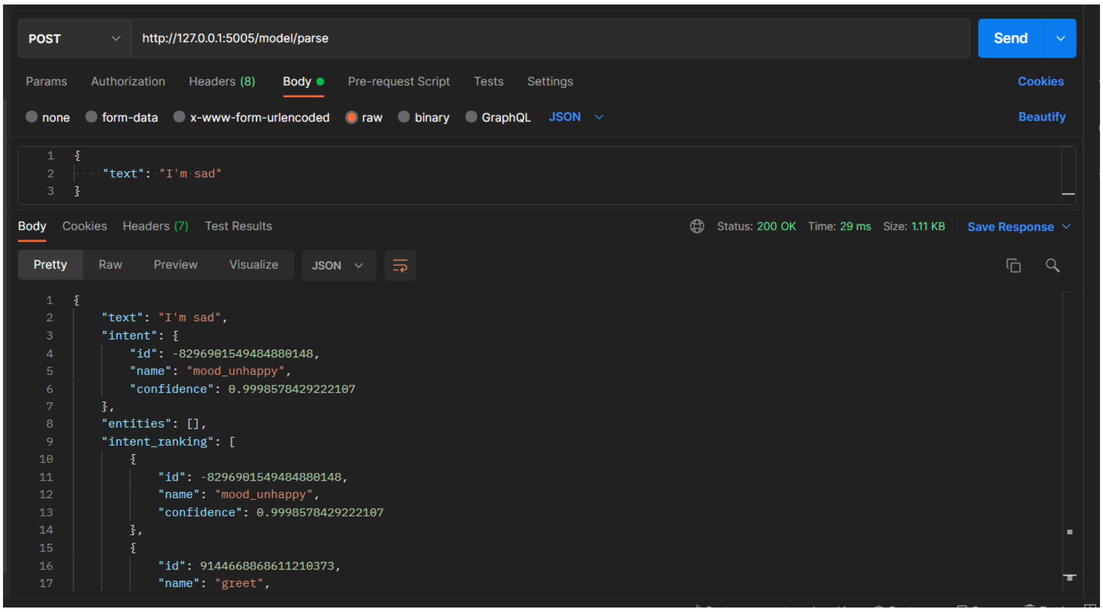

LCE TOOL + MODEL MAINTENANCE TOOLS + FLASK API
SIENA GUI + CLI TOOL + FLASK API
SEETM TOOL + FLASK API
TEST RESULTS + TRAINING DATA + RASA CONFIGS + RASA DIRECTORIES
DIME XAI COMPONENT + FLASK API
SHARED DOCKER VOLUME WITH RASA DIRECTORIES
TRACKER STORE DATABASE TO STORE CONVERSATIONS
RASA ACTION SERVER
RASA NLU+CORE SERVER
NGINX WEB SERVER
CONVERSATIONAL AI MANAGEMENT ARTIFACTS
CADDY/NGINX REVERSE PROXY
CONVERSATIONAL AI ARTIFACTS + BILINGUAL KEYBOARD INTERFACE
DOCKER
OTHER PIPELINE COMPONENTS
CUSTOM ENTITY RECOGNITION MODEL
SEETM ALGORITHM + WORD EMBEDDINGS
OTHER PIPELINE COMPONENTS
RASA NLU PIPELINE
RASA NLU SERVER (EXPANDED)
GCP/EC2 INSTANCE
CONVERSATIONAL AI ASSISTANT END-USER
CONVERSATIONAL AI ASSISTANT MODEL TESTER
CONVERSATIONAL AI ASSISTANT MAINTAINER

Figure 3.2: High-level architectural diagram of the purposed solution with all research components integrated

Figure 3.2 illustrates how DIME fits into the overall system architecture after integrating all research components. The "DIME XAI component + Flask API" is where DIME fits in. DIME component will use a docker container for the deployment, and it will take advantage of the reverse proxy to avoid directly exposing the server contents and the port for security concerns.

3.5 Machine Learning Model Training and Testing

The proposed XAI component requires a fully trained Rasa model with a DIET classifier included as an NLP pipeline component. The Rasa NLU component of the Rasa package allows to load a trained model and parse data to get predictions. The model predictions include intent predicted by the DIET classifier and the entities extracted (Figure 3.3). DIME will use the confidence scores given with the model predictions of the DIET classifier to generate global and local model explanations. The Rasa conversational AI will be trained using the dataset mentioned in subsection 3.3.2.


POST http://127.0.0.1:5005/model/parse

```
{
    "text": "I'm sad"
}
```

Status: 200 OK Time: 29 ms Size: 1.11 KB

```
{
    "text": "I'm sad",
    "intent": {
        "id": -8296901549484880148,
        "name": "mood_unhappy",
        "confidence": 0.9998578429222107
    },
    "entities": [],
    "intent_ranking": [
        {
            "id": -8296901549484880148,
            "name": "mood_unhappy",
            "confidence": 0.9998578429222107
        },
        {
            "id": 9144668868611210373,
            "name": "greet",
```

Figure 3.3: A sample of predictions given by a fully trained Rasa model, including intent, intent ranking, and entities predicted using DIET classifier explored with Postman.

3.6 Generating Model Explanations

3.6.1 Calculating global feature importance

The global feature importance calculation will take the intent predictions for all training data instances given by a fully trained Rasa model with a DIET classifier attached. The DIME algorithm will take the training dataset instead of a separate testing dataset to capture feature importance for all tokens that the model has seen during training. When calculating the global importance of a specific word, the DIME algorithm will get the predictions from the original model for the training data instances by removing the token of interest from the training data. It will perform this process repeatedly and find the probability of the correct model predictions. The algorithm will then consider the change in probability of accurate model predictions as the global feature importance of the token of interest.

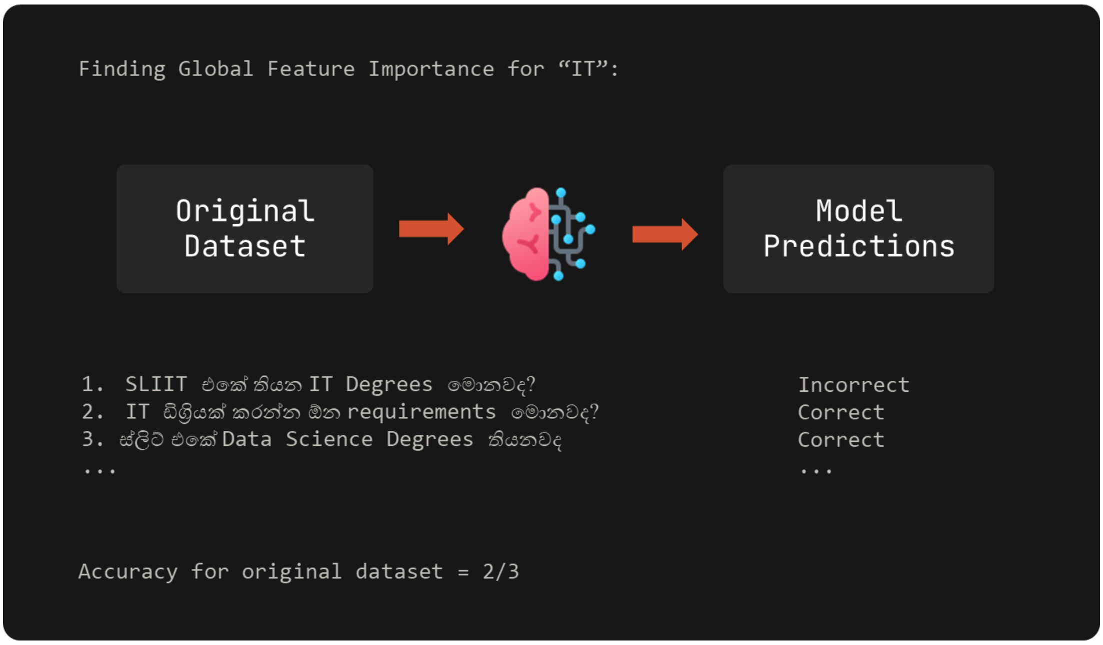
Consider the sample training data instances and the corresponding predictions given in Table 3.1. First, the algorithm finds the probability of correct predictions without alternating training data. In the following case, the percentage of accurate predictions given by the model is about 0.6667. If the token of interest is "IT", it will be removed from training, and the probability of correct predictions given by the model is calculated. The percentage of accurate predictions is around 0.3333 after the alternations are done (Table Table 3.2). Then the change in probability of accurate model prediction is calculated and assigned as the global feature importance of the token "IT". Thus, the feature importance of IT roughly equals to 0.3334. The above process is repeated for all terms, and the results will be cached to minimize the number of repeated calculations.

Table 3.1: Sample training data and predictions

Original training data	**Predicted Class**	**Is correct prediction**
SLIIT එකේ තියන IT Degrees මොනවද?	Ask_Hotline	Incorrect
IT ඩිග්‍රියක් කරන්න ඕන requirements මොනවද?	Ask_Requirements	Correct
ස්ලිට් එකේ Data Science Degrees තියනවද	Ask_Degrees	Correct

Table 3.2: Sample altered training data and respective predictions

Altered training data after removing the token “IT”	**Predicted Class**	**Is correct prediction**
SLIIT එකේ තියන Degrees මොනවද?	Ask_Hotline	Incorrect
ඩිග්‍රියක් කරන්න ඕන requirements මොනවද?	Ask_Requirements	Incorrect
ස්ලිට් එකේ Data Science Degrees තියනවද	Ask_Degrees	Correct



Finding Global Feature Importance for “IT”:

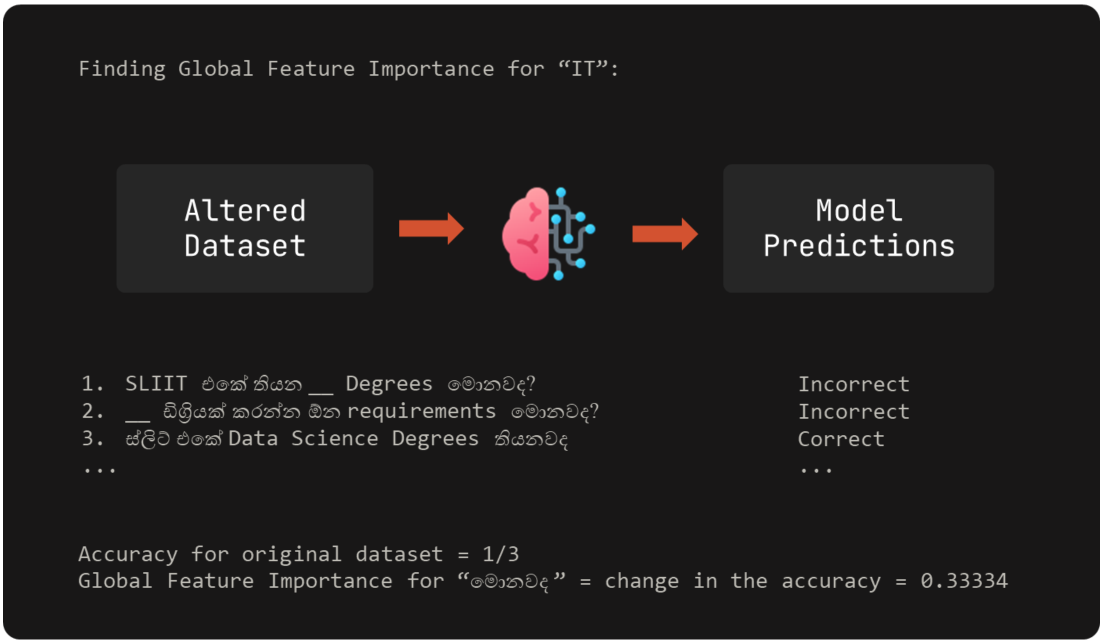
Original Dataset

Model Predictions

```
1.  SLIIT එකේ තියන IT Degrees මොනවද?                   Incorrect
2.  IT ඩිග්‍රියක් කරන්න ඕන requirements මොනවද?            Correct
3.  ස්ලිට් එකේ Data Science Degrees තියනවද               Correct
...                                                  ...
```

Accuracy for original dataset = 2/3

Figure 3.4: Providing the original dataset and recording the model accuracy



Finding Global Feature Importance for “IT”:

Altered Dataset

Model Predictions

```
1.  SLIIT එකේ තියන __ Degrees මොනවද?                   Incorrect
2.  __ ඩිග්‍රියක් කරන්න ඕන requirements මොනවද?            Incorrect
3.  ස්ලිට් එකේ Data Science Degrees තියනවද               Correct
...                                                  ...
```

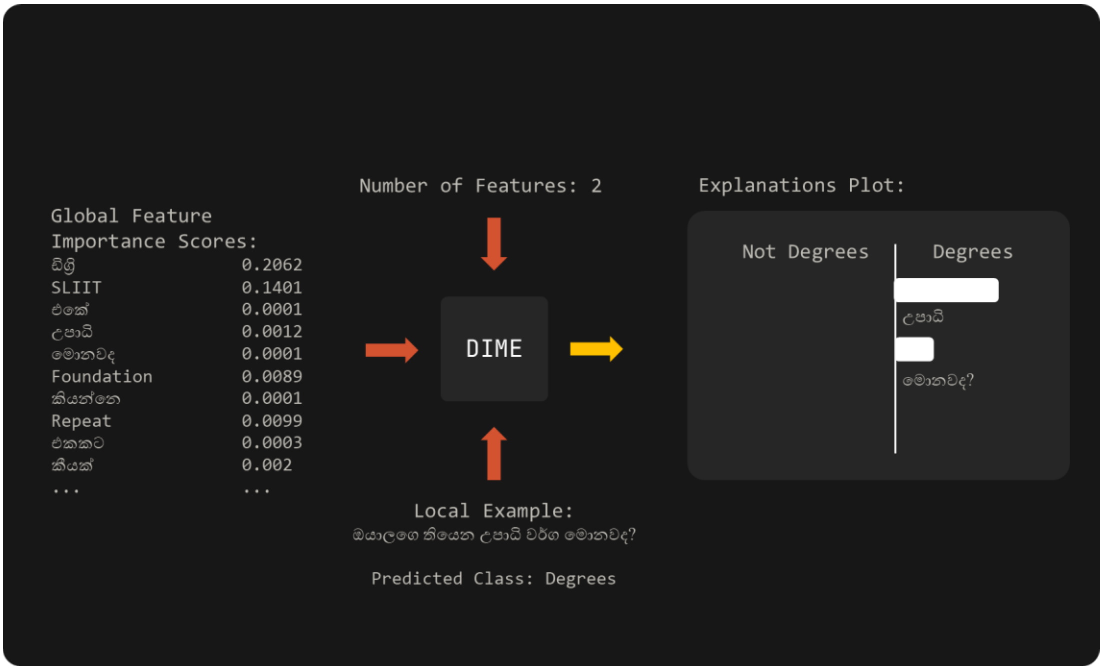
Accuracy for original dataset = 1/3
Global Feature Importance for “මොනවද” = change in the accuracy = 0.33334

Figure 3.5: Providing the altered dataset by removing “IT” and recording the model accuracy

The implementation of the research component will attempt another approach called permutation feature importance as an auxiliary approach to calculate the global feature importance. Instead of removing features from training data, permutation feature importance randomly shuffles the value of a token to calculate the mean decrease in the probability of accurate model predictions.

3.6.2 Generating local model explanations

The local model explanations consider a single data instance to generate model prediction explanations. The implementation incorporates Shapley value calculation to find the contribution of each word in each data instance towards the prediction given by the model. Shapley values are calculated by generating new data instances by constructing subsets of features of original instance tokens and finding the corresponding change in the model confidence score. Calculation of change in model confidence for each subset of tokens of an instance can be computationally expensive. Thus, local model explanations will utilize global feature importance found as explained previously as a technique to perform feature selection. However, a user must state the number of tokens and the feature selection technique to be used, considering them as hyperparameters. An alternative method will be provided to pass the minimum global feature importance as a percentage to only select features with an importance score equal to or higher than the specified value. For this, the global feature importance scores for all tokens will be sent through a Softmax function. A Softmax function is a mathematical function that can fit all global feature importance scores to a probability distribution that sums up to 1. The local model explanation scores also will be sent through a Softmax function to get the contribution of each token towards the prediction as a percentage/probability.



Global Feature Importance Scores:

ඩිග්‍රි	0.2062
SLIIT	0.1401
එකේ	0.0001
උපාධි	0.0012
මොනවද	0.0001
Foundation	0.0089
කියන්නෙ	0.0001
Repeat	0.0099
එකකට	0.0003
කීයක්	0.002
...	...

Number of Features: 2

Local Example:
ඔයාලගෙ තියෙන උපාධි වර්ග මොනවද?

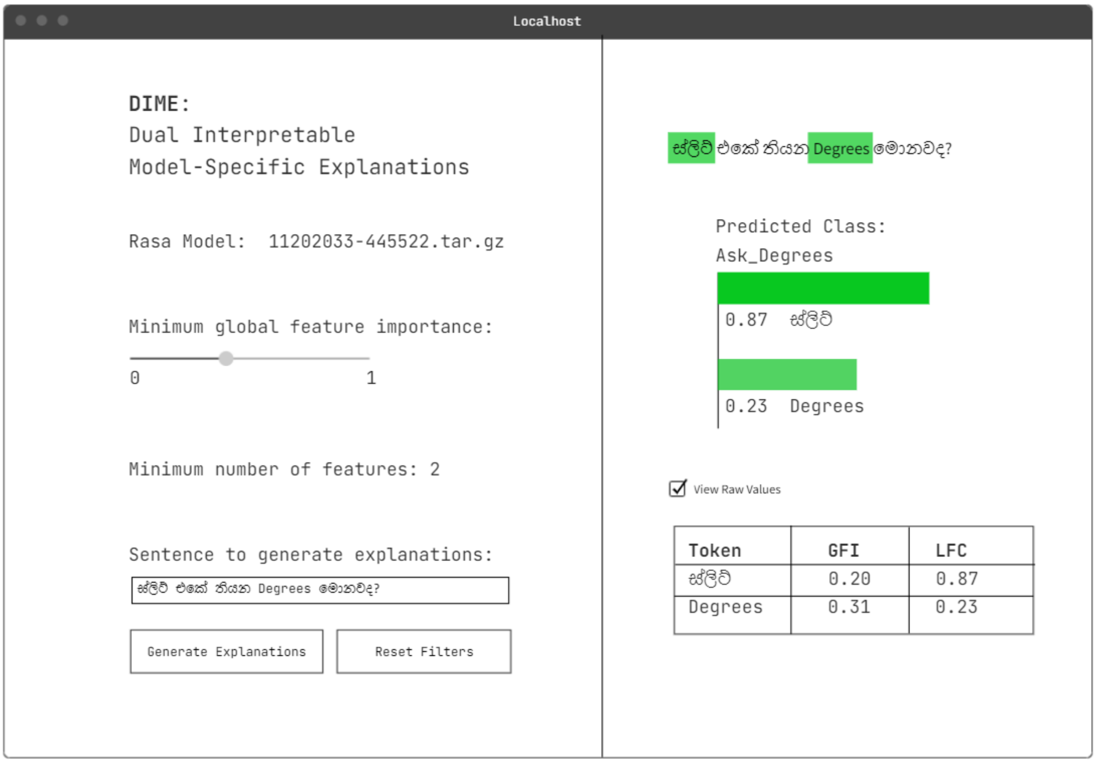
Predicted Class: Degrees

Explanations Plot:

Figure 3.6: Generating local explanations using global feature importance

3.7 User Interfaces and Text Visualizations

DIME will mainly have two interfaces, including a web interface and a CLI. The CLI will allow generating explanations and viewing the raw scores in the terminal. The CLI is useful when DIME is executed within a server with no access to a GUI or if it is impossible to run the webserver. However, the CLI interface will be less interpretable. The web interface will comprise interpretable explanations with word highlighting and feature contribution scores towards a specific prediction.



DIME:
Dual Interpretable
Model-Specific Explanations

Rasa Model: 11202033-445522.tar.gz

Minimum global feature importance: 0 1

Minimum number of features: 2

Sentence to generate explanations: ස්ලිට් එකේ තියන Degrees මොනවද?

ස්ලිට් එකේ තියන Degrees මොනවද?

Predicted Class:
Ask_Degrees

☑ View Raw Values

Token	GFI	LFC
ස්ලිට්	0.20	0.87
Degrees	0.31	0.23

Figure 3.7: Proposed web interface wireframe of the DIME local server for model explanation visualizations.

3.8 Tools and Technologies

Python 3.8 will be used for the implementation of the DIME XAI algorithm. Python packages such as NumPy, Pandas, and Sklearn will be utilized for handling the dataset. Rasa 2.8.12 will be used to handle Rasa machine learning models and train the conversational AI for both testing and final implementation. PyCharm, Visual Studio Code, and Google CoLab will be used as Integrated Development Environments. Anaconda Python distribution will be used for convenient python environment management.

A local server will be developed using Flask to build the required API endpoints and the interactive web application for the visualizations. DIME package and required Rasa source code scripts will be included in a docker container for efficient deployment. The DIME docker container will be integrated with the docker containers of packages from the other research components using the docker-compose tool, and the overall system will be deployed on either a GCP or an EC2 instance according to the final system requirements. A reverse proxy will be set up using Caddy Server 2.0 and NGINX webservers to secure the production server by minimizing the exposed ports to the public. The summary of the tools to be used is as follows.

Table 3.3: Summary of Tools and Technologies to be used according to the tasks

Task	Tools to be used
Algorithm Implementation and DIME package development	Python 3.8, NumPy, Pandas, sklearn, Rasa 2.8.12, PyCharm, Visual Studio Code, Google CoLab
Server development as a standalone frontend for the DIME visualizations	Flask, JavaScript, Bootstrap, React JS (optional), Chart JS, Streamlit
NoSQL Database development	MongoDB Atlas, MongoDB Compass
Cloud Infrastructure management	One out of Google Cloud Platform or Amazon Web Services
Conversational AI development by Integrating all research components (Backend)	Rasa 2.8.12, MongoDB Atlas, Gensim, spaCy, Docker
Conversational AI frontend development	React JS, Bootstrap, socket.io, JavaScript
Overall system deployment	Caddy 2, NGINX, Git, Docker, docker-compose

4. DESCRIPTION OF PERSONAL AND FACILITIES

Figure 4.1 breaks down all tasks of this individual research component under 5 main tasks. Deployment is done at the end by integrating all four research components.

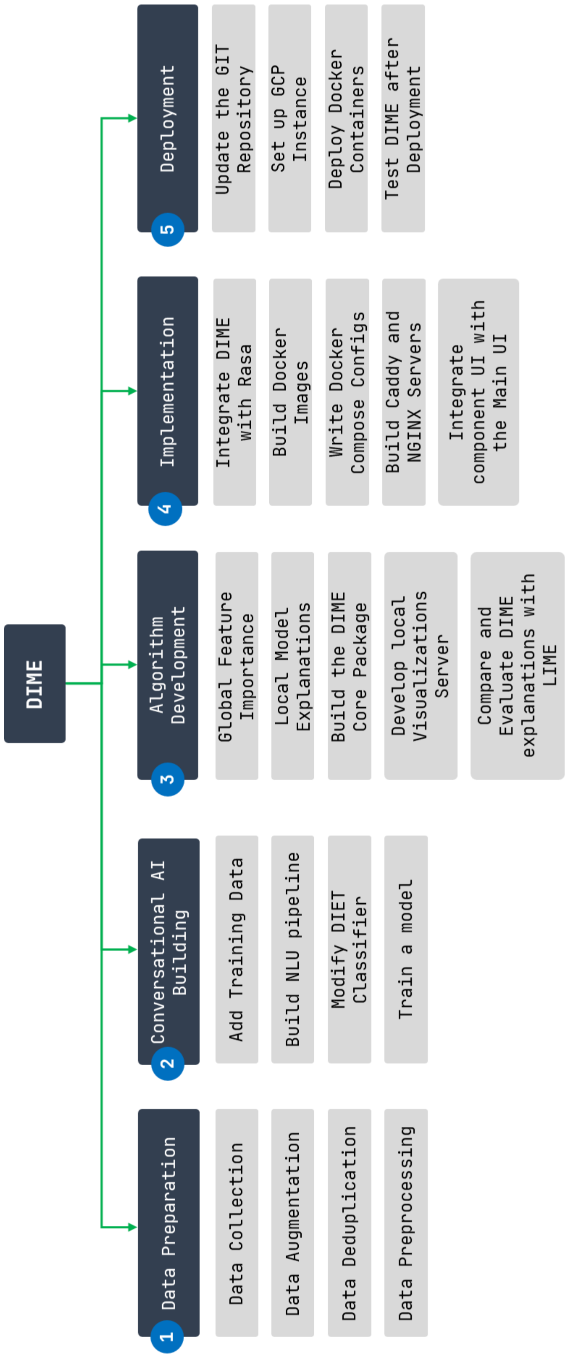
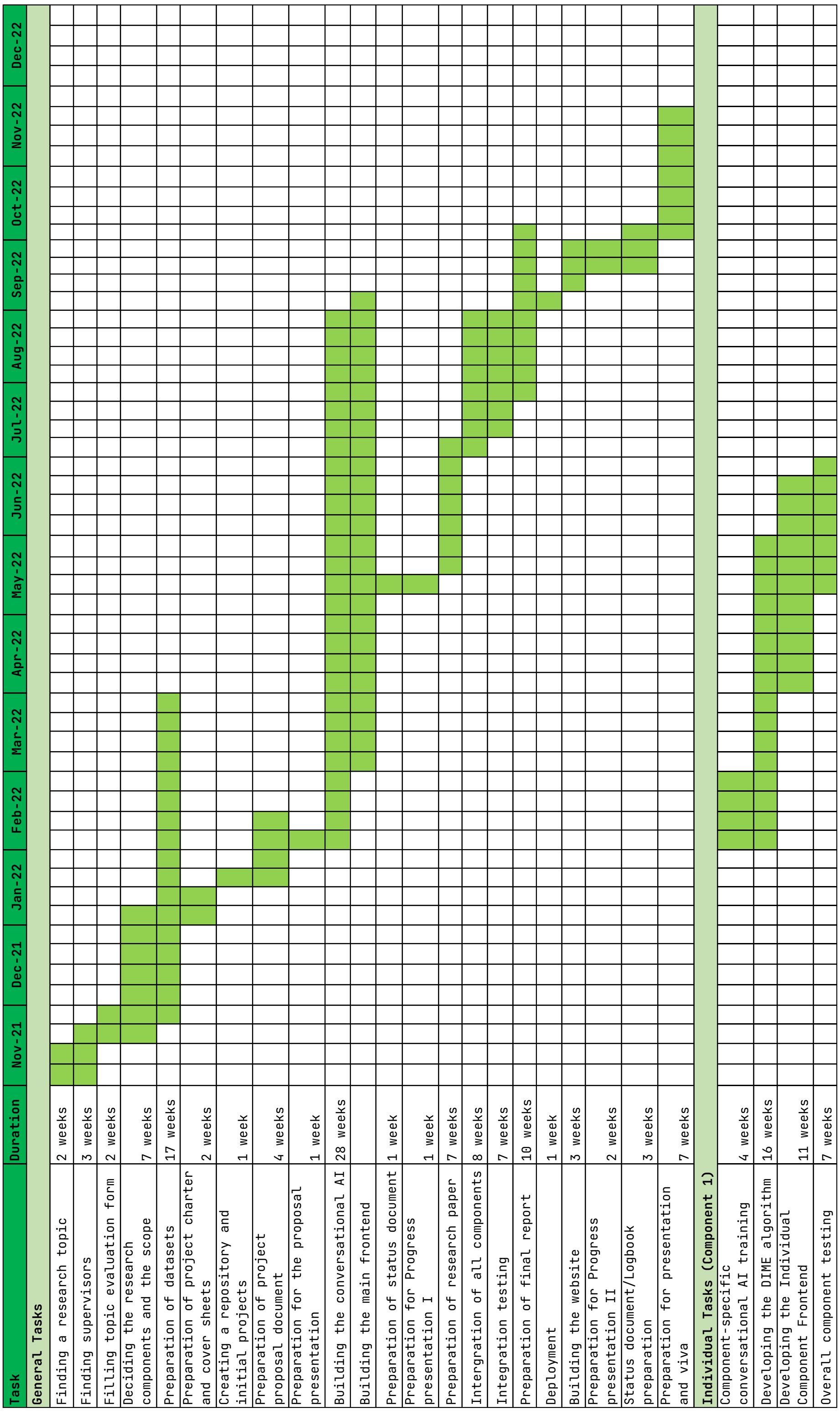


Figure 4.1: Work breakdown structure of the individual research component

5. GANTT CHART



Task	Duration	Nov-21	Dec-21	Jan-22	Feb-22	Mar-22	Apr-22	May-22	Jun-22	Jul-22	Aug-22	Sep-22	Oct-22	Nov-22	Dec-22
General Tasks															
Finding a research topic	2 weeks	🟩													
Finding supervisors	3 weeks	🟩													
Filling topic evaluation form	2 weeks	🟩													
Deciding the research components and the scope	7 weeks	🟩	🟩	🟩											
Preparation of datasets	17 weeks	🟩	🟩	🟩	🟩	🟩									
Preparation of project charter and cover sheets	2 weeks			🟩											
Creating a repository and initial projects	1 week			🟩											
Preparation of project proposal document	4 weeks			🟩	🟩										
Preparation for the proposal presentation	1 week				🟩										
Building the conversational AI	28 weeks				🟩	🟩	🟩	🟩	🟩	🟩	🟩				
Building the main frontend						🟩	🟩	🟩	🟩	🟩	🟩	🟩			
Preparation of status document	1 week							🟩							
Preparation for Progress presentation I	1 week							🟩							
Preparation of research paper	7 weeks							🟩	🟩	🟩					
Intergration of all components	8 weeks									🟩	🟩				
Integration testing	7 weeks									🟩	🟩				
Preparation of final report	10 weeks									🟩	🟩	🟩	🟩		
Deployment	1 week											🟩			
Building the website	3 weeks											🟩			
Preparation for Progress presentation II	2 weeks											🟩			
Status document/Logbook preparation	3 weeks											🟩	🟩		
Preparation for presentation and viva	7 weeks												🟩	🟩	
Individual Tasks (Component 1)															
Component-specific conversational AI training	4 weeks				🟩										
Developing the DIME algorithm	16 weeks				🟩	🟩	🟩	🟩							
Developing the Individual Component Frontend	11 weeks						🟩	🟩	🟩						
Overall component testing	7 weeks							🟩	🟩						

Figure 5. 1: Gantt chart for overall project and the individual component

6. COMMERCIALISATION PLAN

Each research component of the overall research project provides applications for various Natural Language Processing and Machine Learning model evaluation-related tasks. Although the outputs of each research component can help design many products and services, they were all integrated to build a conversational AI that fully supports Sinhala-English code-switching. The main reason for developing a conversational AI are as follows.

1. Conversational AIs are a trending topic, and there is an increasing demand for Sinhala-based conversational AIs. Many businesses are looking to increase their customer reach by using conversational AIs for a vast range of business tasks, including providing technical assistance, automating manual tasks such as booking tickets, and providing the information requested by the customers. Although that is the case, it is hard to find Sinhala-based conversational AIs, especially in Sri Lanka. Thus, developing Sinhala-based conversational AIs was identified as a potential business opportunity.

2. Although there are many conversational AI development frameworks, most of them lack evaluation tools to debug machine learning models. In frameworks like Rasa, model evaluations are highly technical, and the average developers without machine learning knowledge fail to identify problems and debug the machine learning models. Solving this issue by allowing non-technical users to maintain and evaluate machine learning models and conversational AIs is another business opportunity where evaluation tools can be built and released as add-on features.

3. Businesses are moving towards cloud-based solutions, especially SaaS products from traditional standalone applications, web applications, and on-premises solutions, where the maintenance cost is high. A cloud-based highly configurable conversational AI would be an appropriate solution for many businesses where the effort for maintenance is considerably low.

The end product of the overall research concentrates on designing a solution for the above potential business ideas and opportunities. The conversational AI developed as the research end-product is mainly a SaaS or simply a CaaS (conversational AI-as-a-service) product that a business can purchase with a set of add-on features, as illustrated in Figure 6.1. In addition to the CaaS packages, on-premises and demo cloud-based conversational AI packages are available for affordability and convenience. The end product of the overall research has the following archivable user benefits.

1. Businesses can purchase a CaaS package and eliminate conversational AI maintenance efforts.

2. Developers can use code-less maintenance tools to maintain conversational AIs they purchase without having in-depth knowledge about the backend deployment and the conversational AI development framework.

3. Businesses can purchase evaluation tools as add-ons and generate model explanations and evaluate machine learning models themselves to avoid extra maintenance costs. Here, the generated evaluations are easy to understand by non-technical users, which is not a feature of any existing chatbot framework.

4. Users of conversational AI can easily use the Sinhala-English code-switchable keyboard interface, and businesses can attract more users from having this feature as it eliminates the need to use third-party Sinhala typing services.

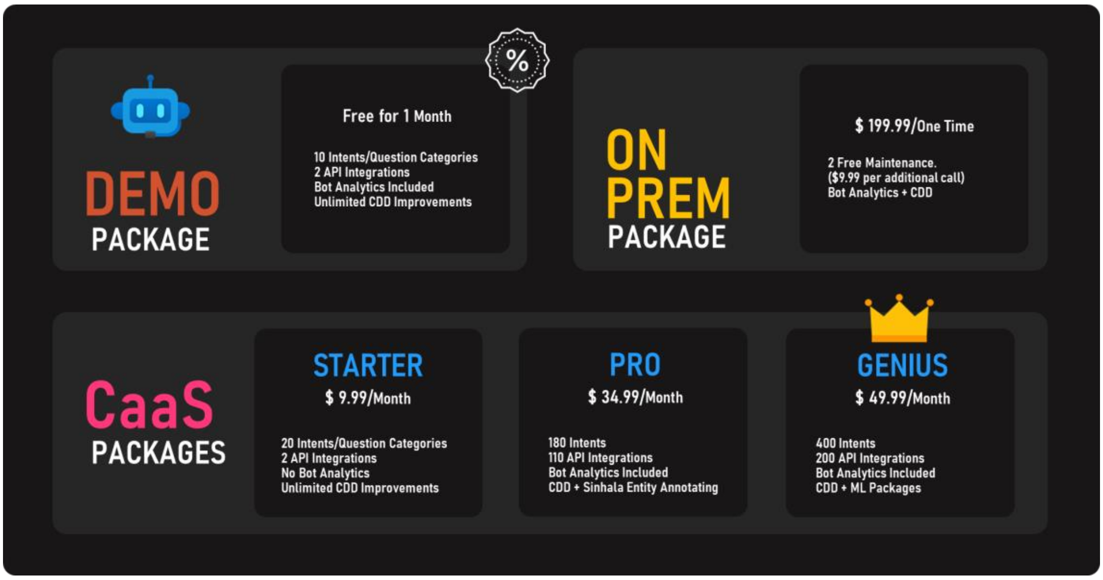
5. Businesses have the freedom to purchase either the CaaS packages or on-premises packages as per their need, while anyone can test the demo conversational AI for a period of 1 month before deciding to purchase any of the other packages that have a cost assigned.

6. Businesses can reach a vast customer range through Sinhala-English code-switching-based conversational AIs, especially within the Sri Lankan market, and it is possible to eliminate the need for having a dedicated customer care staff. It will dramatically lower the expenses of the business.

The proposed commercialisation plan of the end product of the overall research component contains a set of convenient packages and the offered features of each package differ based on the cost attached to it. The distribution of the features and the package cost were carefully planned and designed by analyzing the existing purchasable packages of similar SaaS products and conversational AIs. Table 6.1 clearly illustrates the feature variation and the cost difference of the packages offered by the proposed commercialisation plan.

Table 6.1: Feature variations and cost difference of packages proposed by the commercialisation plan

Feature	**Packages**				
	Demo	**On-Prem**	**CaaS**		
			Starter	**Pro**	**Genius**
Intents	10	Unlimited	20	180	400
API Integrations	2	Unlimited	2	110	200
Bot Analytics	✓	✓	-	✓	✓
CDD	✓	✓	✓	✓	✓
Sinhala Entity Annotating	✓	-	-	✓	✓
ML Evaluation Tools	-	-	-	-	✓
Maintenance Fee	-	2 Free + $9.99 per additional call	-	-	-
Trial Duration	1 Month	-	-	-	-
Package Price	Free	$199.99	$9.99	$34.99	$49.99



DEMO PACKAGE

Free for 1 Month

10 Intents/Question Categories
2 API Integrations
Bot Analytics Included
Unlimited CDD Improvements

ON PREM PACKAGE

$ 199.99/One Time

2 Free Maintenance.
($9.99 per additional call)
Bot Analytics + CDD

CaaS PACKAGES

STARTER

$ 9.99/Month

20 Intents/Question Categories
2 API Integrations
No Bot Analytics
Unlimited CDD Improvements

PRO

$ 34.99/Month

180 Intents
110 API Integrations
Bot Analytics Included
CDD + Sinhala Entity Annotating

GENIUS

$ 49.99/Month

400 Intents
200 API Integrations
Bot Analytics Included
CDD + ML Packages

Figure 6.1: proposed commercialisation plan

7. BUDGET AND BUDGET JUSTIFICATION

The budget justification for all four research components is mentioned in **Error! Reference source not found.**.

Table 7.1: Budget justification for the overall research project

Component Name	Individual Item Price (LKR)	Number of Items	Duration	Total Item Price (LKR)
Domain Name	2148.43/year	1	1 year	2148.43
GCP Instance	10683.84/month	1	6 months	64103.06
Reference Book: Basaka Mahima by J.B. Dissanayake	1250.00	1	-	1250.00
Research Paper Publication	25000.00	1	-	25000.00
Grand Total	-	-	-	92,501.49

REFERENCES

[1]. M. T. Ribeiro, S. Singh, C. Guestrin, "'why should I trust you?': Explaining the predictions of any classifier," 16 Feb 2016. [Online]. Available: https://arxiv.org/abs/1602.04938

[2]. T. Bunk, D. Varshneya, V. Vlasov, A. Nichol, "DIET: Lightweight Language Understanding for Dialogue Systems," 2020. [Online]. Available: https://arxiv.org/abs/2004.09936

[3]. G. Caldarini, S. Jaf, K. McGarry, "A Literature Survey of Recent Advances in Chatbots," 2022. [Online]. Available: https://arxiv.org/abs/2201.06657

[4]. F. K. Došilović, M. Brčić and N. Hlupić, "Explainable artificial intelligence: A survey," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 0210-0215, doi: 10.23919/MIPRO.2018.8400040.

[5]. M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, P. Sen, "A Survey of the State of Explainable AI for Natural Language Processing," 2020. [Online]. Available: https://arxiv.org/abs/2010.00711

[6]. Y. Jin and B. Sendhoff, "Pareto-Based Multiobjective Machine Learning: An Overview and Case Studies," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., vol. 38, no. 3, pp. 397-415, May 2008.

[7]. G. Montavon, W. Samek, and K.-R. Muller, "Methods for interpreting and understanding deep neural networks," Digit. Signal Process., vol. 73, pp. 1-15, Feb. 2018

[8]. S. R. Islam, W. Eberle, S. K. Ghafoor, en M. Ahmed, “Explainable Artificial Intelligence approaches: A survey,” 2021. [Online]. Available: https://arxiv.org/abs/2101.09429

[9]. P. Gohel, P. Singh, en M. Mohanty, “Explainable AI: current status and future directions”, 2021. [Online]. Available: https://arxiv.org/abs/2107.07045

[10]. S. Lundberg, S. Lee, “A Unified Approach to Interpreting Model Predictions,” 2017. [Online]. Available: https://arxiv.org/abs/1705.07874

[11]. H2O.ai, *Interpretable Machine Learning Using LIME Framework - Kasia Kulma (PhD), Data Scientist, Aviva*. (Dec. 19, 2017). Accessed: Dec. 25, 2021. [Online Video]. Available: https://www.youtube.com/watch?v=CY3t11vuuOM&list=TLPQMjYxMjIwMjHjkkneYrcmyg&index=2

[12]. Xainlp2020.github.io, “*XAI for Natural Language Processing*”, [online] Available at: https://xainlp2020.github.io/xainlp [Accessed 24 January 2022].

[13]. D. Rajagopal, V. Balachandran, E. Hovy, en Y. Tsvetkov, “SelfExplain: A self-explaining architecture for neural text classifiers”, 2021. [Online]. Available: https://arxiv.org/abs/2103.12279

[14]. A. Shrikumar, P. Greenside, en A. Kundaje, “Learning important features through propagating activation differences”, 2017. [Online]. Available: https://arxiv.org/abs/1704.02685

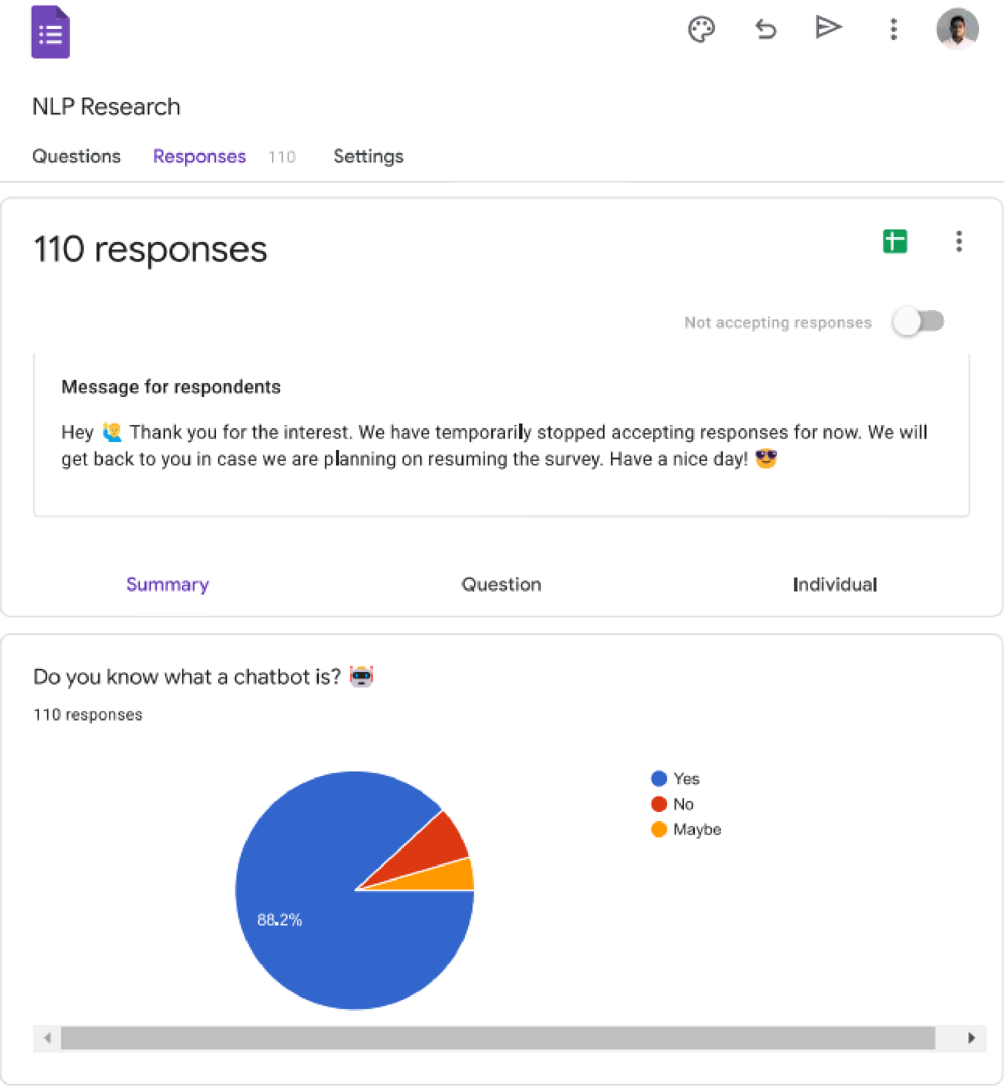
[15]. Lloyd S Shapley. “A value for n-person games”. In: Contributions to the Theory of Games 2.28 (1953), pp. 307–317.

[16]. C. Molnar, 2022. “*Interpretable Machine Learning*”, 2022. [online] Christophm.github.io. Available at: https://christophm.github.io/interpretable-ml-book [Accessed 24 January 2022].

[17]. T. Bocklisch, J. Faulkner, N. Pawlowski, A. Nichol, “Rasa: Open Source Language Understanding and Dialogue Management,” 2017. [Online]. Available: https://arxiv.org/abs/1712.05181

APPENDICES

Appendix A: Survey Form



NLP Research

Questions Responses 110 Settings

110 responses

Not accepting responses

Message for respondents

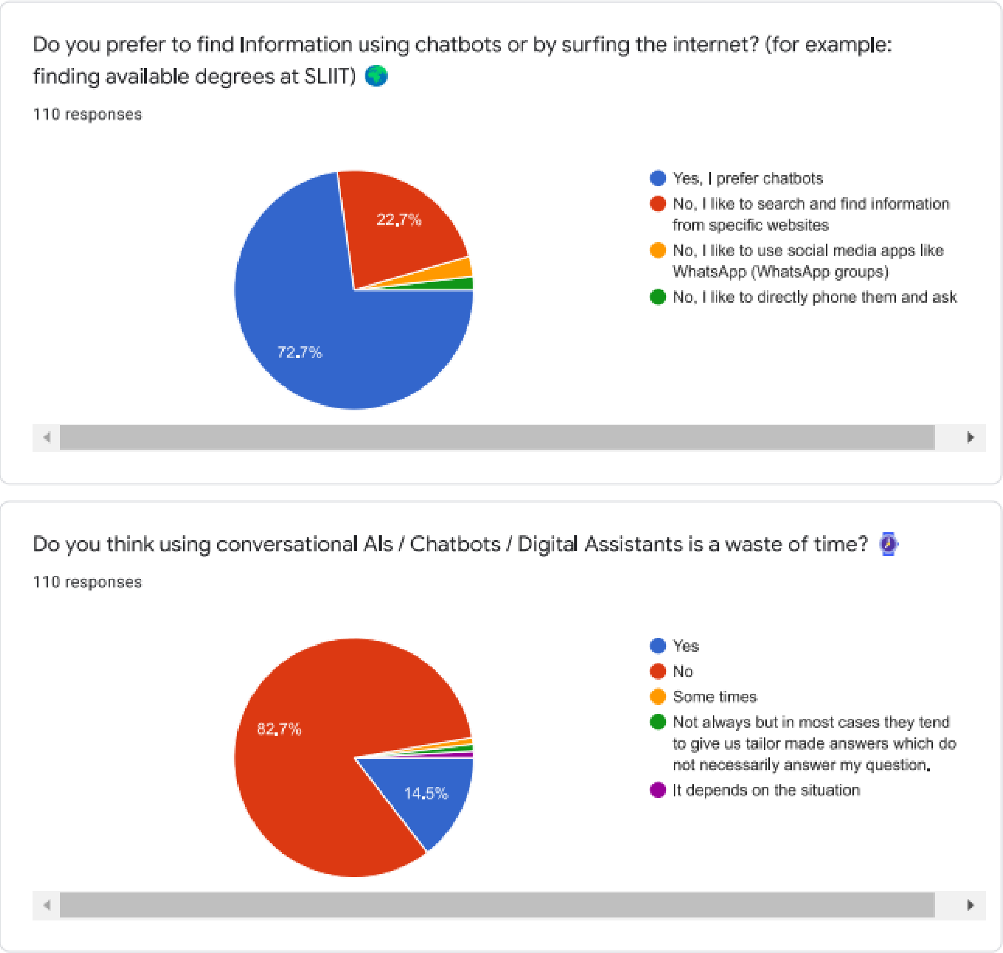
Hey 🙋 Thank you for the interest. We have temporarily stopped accepting responses for now. We will get back to you in case we are planning on resuming the survey. Have a nice day! 😎

Summary Question Individual

Do you know what a chatbot is? 🤖

110 responses

Figure A.1: Complete survey form questions and responses – part 1



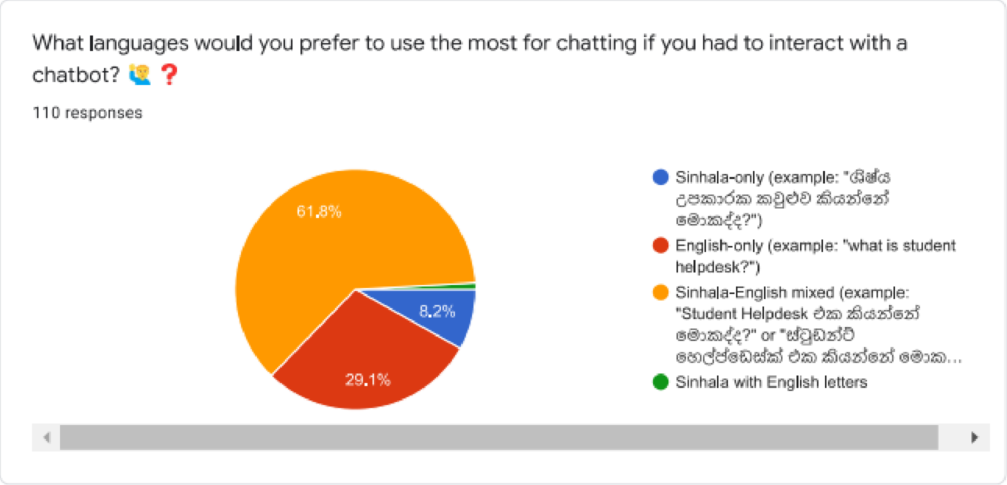
Do you prefer to find Information using chatbots or by surfing the internet? (for example: finding available degrees at SLIIT) 🌍

110 responses

Do you think using conversational AIs / Chatbots / Digital Assistants is a waste of time? ⏱

110 responses

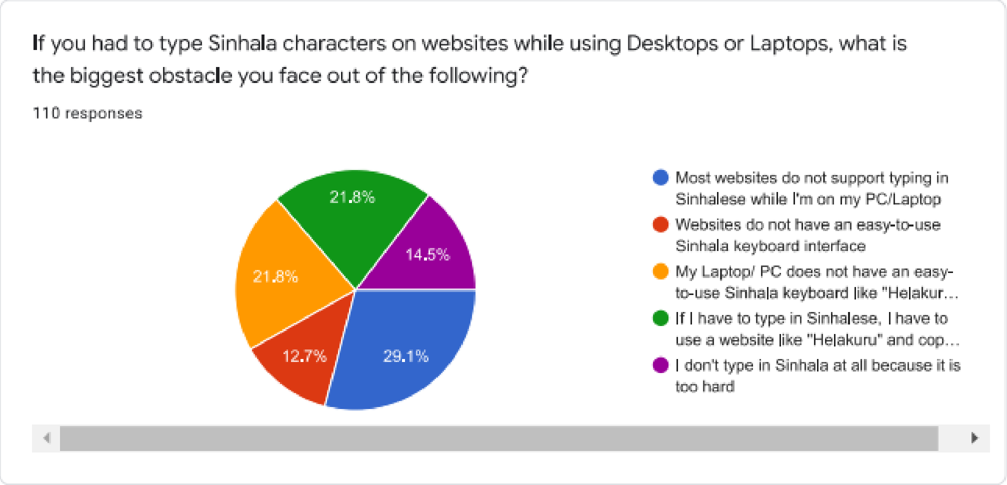
Figure A.2: Complete survey form questions and responses – part 2



What languages would you prefer to use the most for chatting if you had to interact with a chatbot? 🙋 ❓

110 responses

- ● Sinhala-only (example: "ශිෂ්‍ය උපකාරක කවුළුව කියන්නේ මොකද්ද?")
- ● English-only (example: "what is student helpdesk?")
- ● Sinhala-English mixed (example: "Student Helpdesk එක කියන්නේ මොකද්ද?" or "ස්ටුඩන්ට් හෙල්ප්ඩෙස්ක් එක කියන්නේ මොක...
- ● Sinhala with English letters

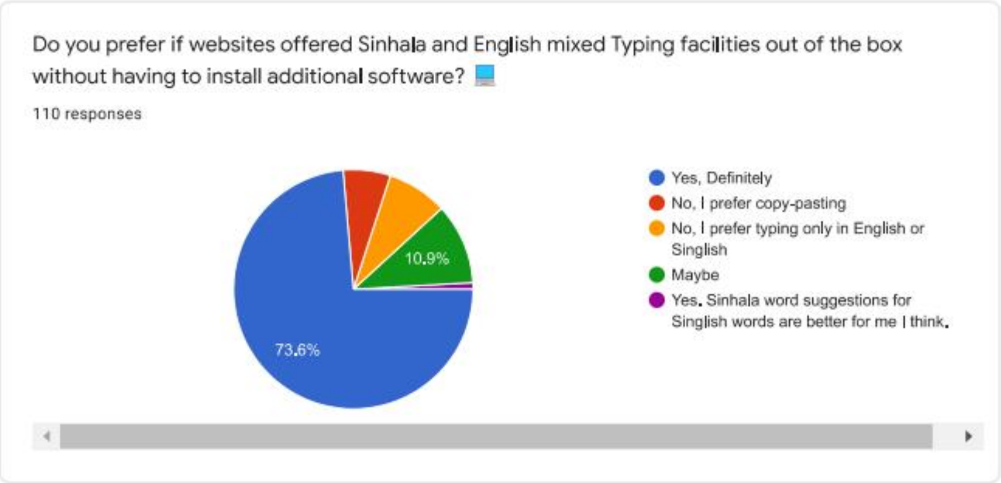


If you had to type Sinhala characters on websites while using Desktops or Laptops, what is the biggest obstacle you face out of the following?

110 responses

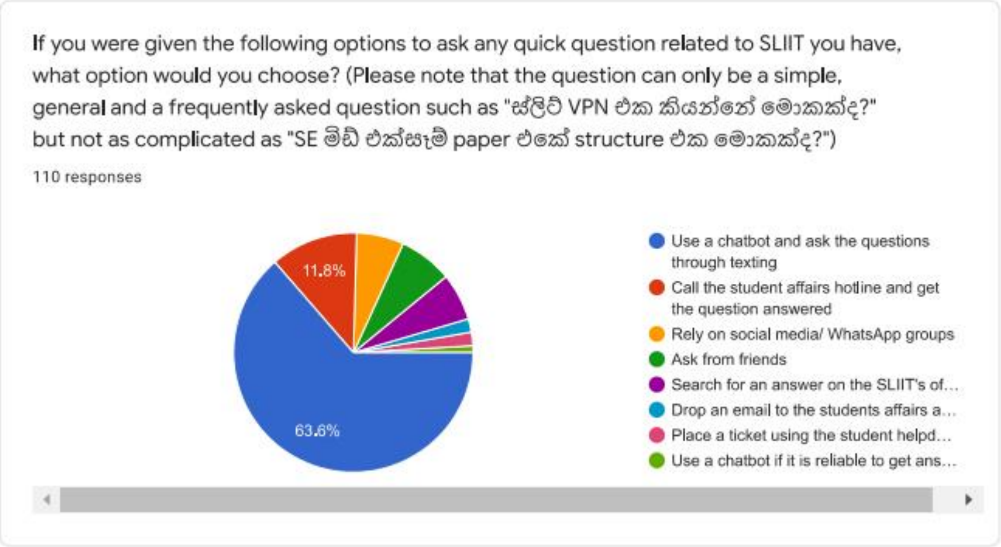
- ● Most websites do not support typing in Sinhalese while I'm on my PC/Laptop
- ● Websites do not have an easy-to-use Sinhala keyboard interface
- ● My Laptop/ PC does not have an easy-to-use Sinhala keyboard like "Helakur...
- ● If I have to type in Sinhalese, I have to use a website like "Helakuru" and cop...
- ● I don't type in Sinhala at all because it is too hard

Figure A.3: Complete survey form questions and responses – part 3



Do you prefer if websites offered Sinhala and English mixed Typing facilities out of the box without having to install additional software? 💻

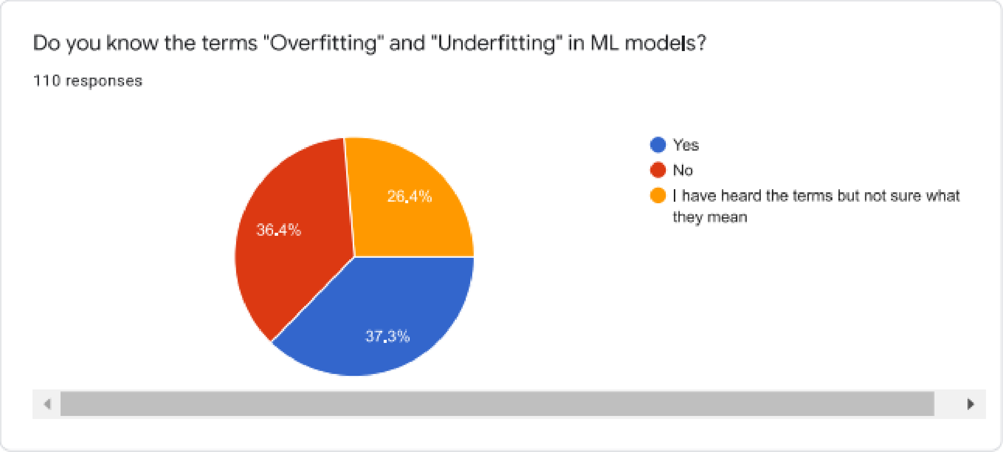
110 responses



If you were given the following options to ask any quick question related to SLIIT you have, what option would you choose? (Please note that the question can only be a simple, general and a frequently asked question such as "ස්ලිට් VPN එක කියන්නේ මොකක්ද?" but not as complicated as "SE මිඩ් එක්සෑම් paper එකේ structure එක මොකක්ද?")

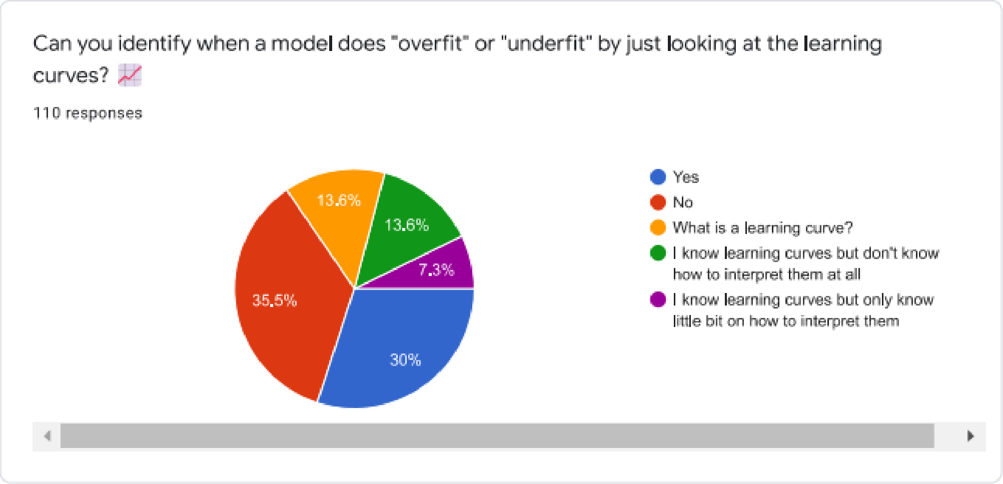
110 responses

Figure A.4: Complete survey form questions and responses – part 4



Do you know the terms "Overfitting" and "Underfitting" in ML models?

110 responses

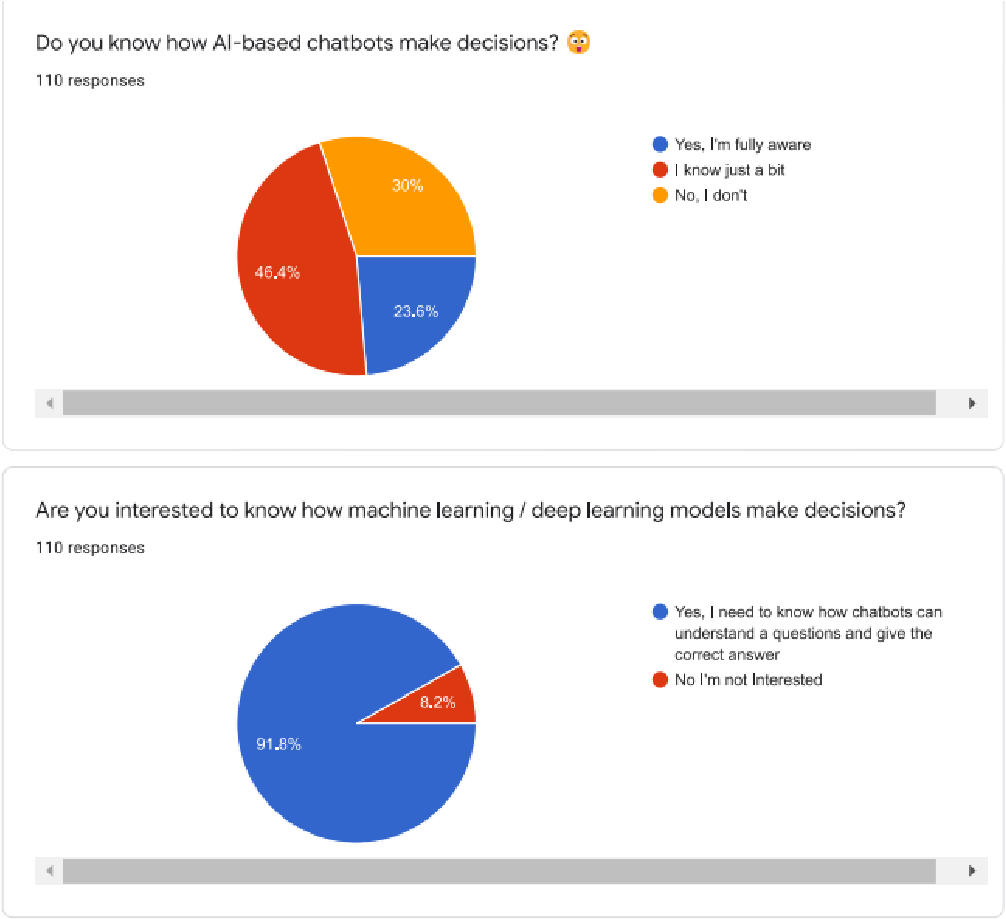


Can you identify when a model does "overfit" or "underfit" by just looking at the learning curves? 📈

110 responses

● Yes
● No
● What is a learning curve?
● I know learning curves but don't know how to interpret them at all
● I know learning curves but only know little bit on how to interpret them
13.6%
13.6%
7.3%
35.5%
30%

Figure A. 5: Complete survey form questions and responses – part 5



Do you know how AI-based chatbots make decisions? 😳

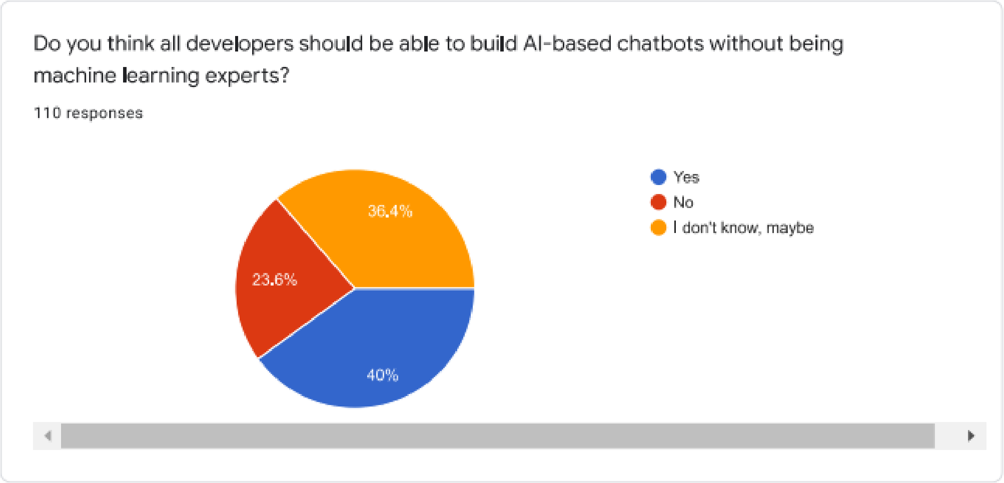
110 responses

Are you interested to know how machine learning / deep learning models make decisions?

110 responses

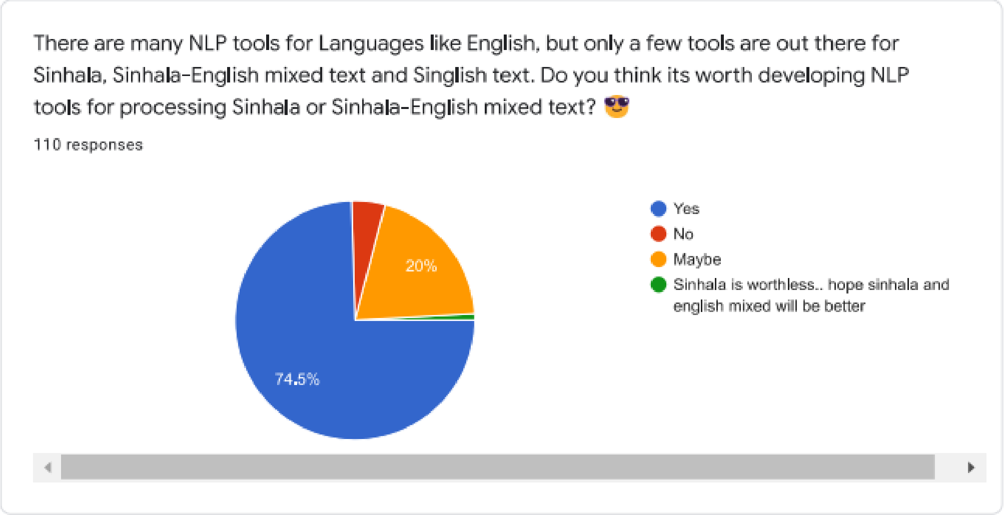
8.2%
91.8%
● Yes, I need to know how chatbots can understand a questions and give the correct answer
● No I'm not Interested

Figure A.6: Complete survey form questions and responses – part 6



Do you think all developers should be able to build AI-based chatbots without being machine learning experts?

110 responses

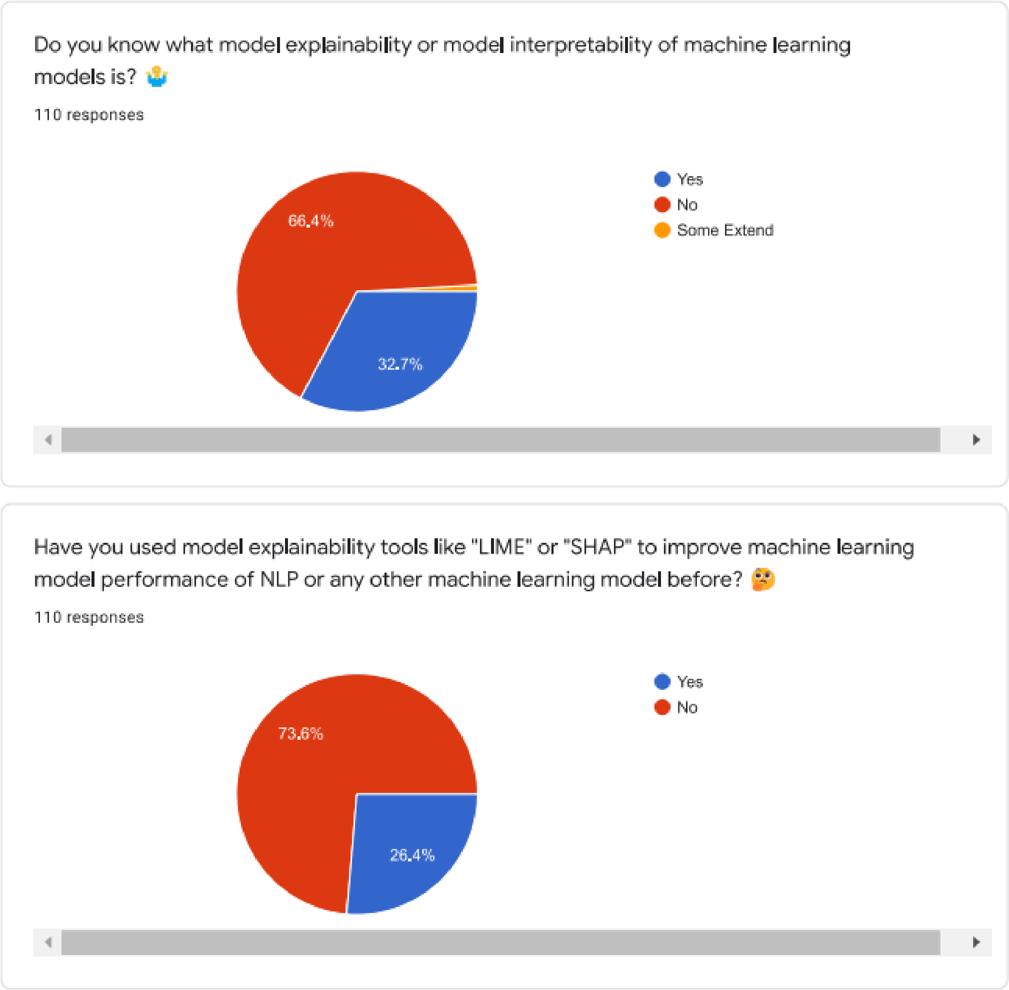


There are many NLP tools for Languages like English, but only a few tools are out there for Sinhala, Sinhala-English mixed text and Singlish text. Do you think its worth developing NLP tools for processing Sinhala or Sinhala-English mixed text? 😎

110 responses

20%
74.5%
● Yes
● No
● Maybe
● Sinhala is worthless.. hope sinhala and english mixed will be better

Figure A.7: Complete survey form questions and responses – part 7



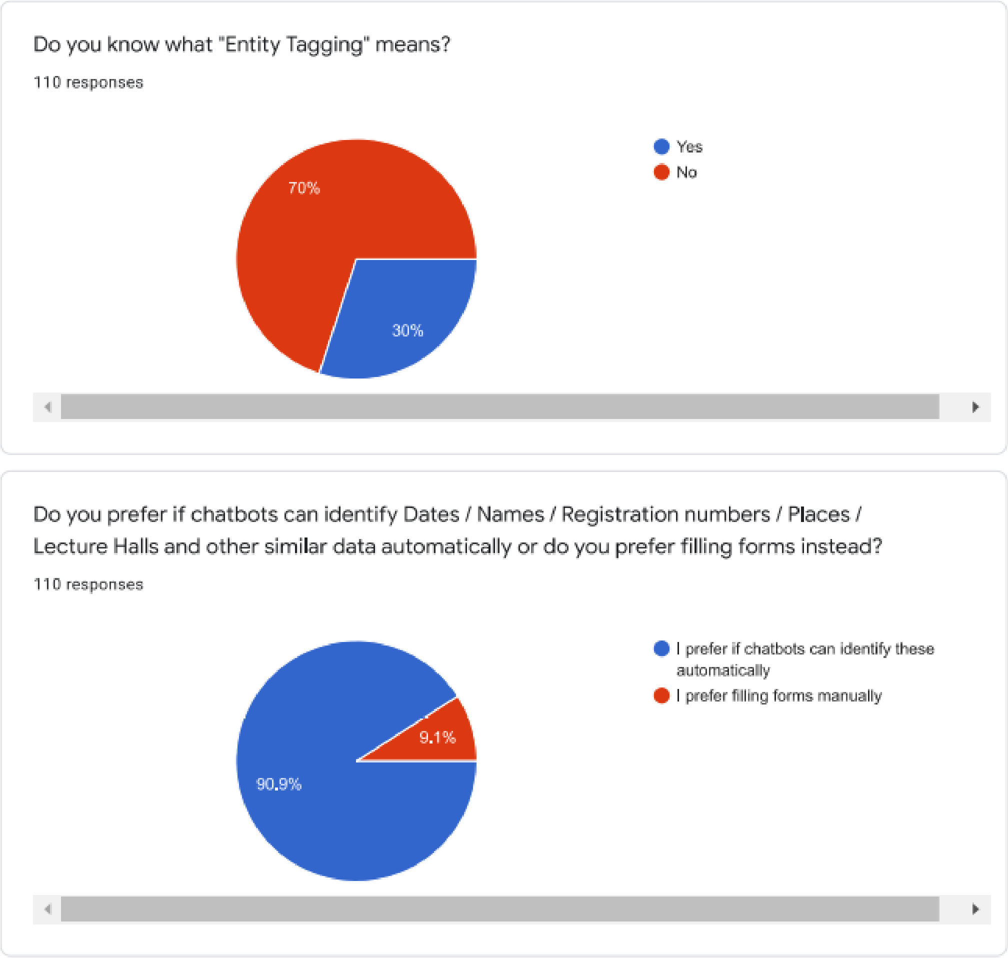
Do you know what model explainability or model interpretability of machine learning models is? 🤷

110 responses

Have you used model explainability tools like "LIME" or "SHAP" to improve machine learning model performance of NLP or any other machine learning model before? 🤔

110 responses

Figure A.8: Complete survey form questions and responses – part 8



Do you know what "Entity Tagging" means?

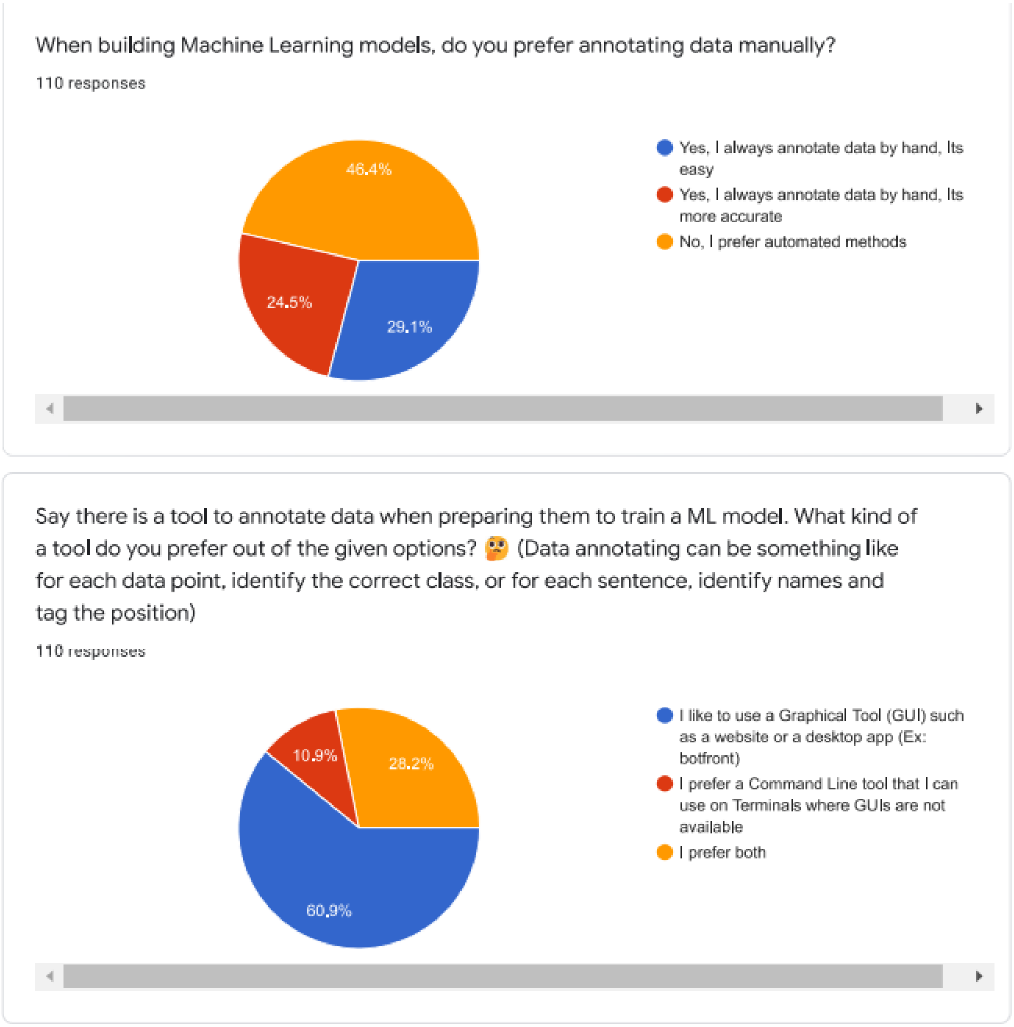
110 responses

Do you prefer if chatbots can identify Dates / Names / Registration numbers / Places / Lecture Halls and other similar data automatically or do you prefer filling forms instead?

110 responses

9.1%
90.9%
● I prefer if chatbots can identify these automatically
● I prefer filling forms manually

Figure A.9: Complete survey form questions and responses – part 9



When building Machine Learning models, do you prefer annotating data manually?

110 responses

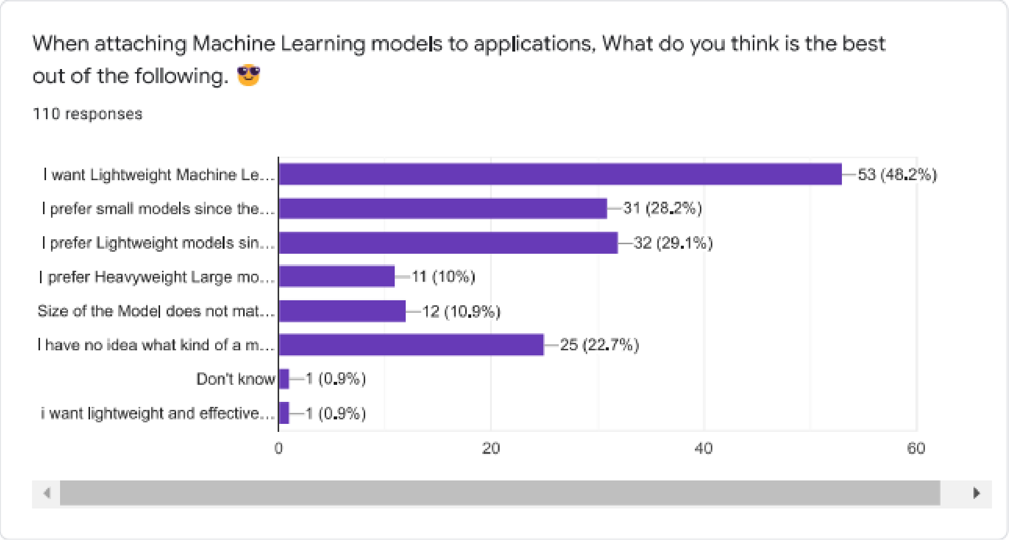
46.4%
24.5%
29.1%
● Yes, I always annotate data by hand, Its easy
● Yes, I always annotate data by hand, Its more accurate
● No, I prefer automated methods

Say there is a tool to annotate data when preparing them to train a ML model. What kind of a tool do you prefer out of the given options? 🤔 (Data annotating can be something like for each data point, identify the correct class, or for each sentence, identify names and tag the position)

110 responses

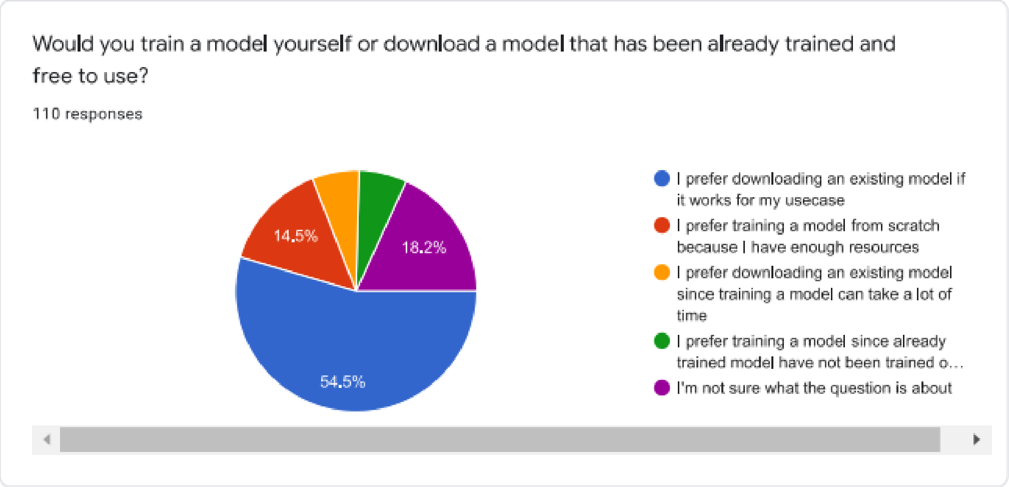
10.9%
28.2%
60.9%
● I like to use a Graphical Tool (GUI) such as a website or a desktop app (Ex: botfront)
● I prefer a Command Line tool that I can use on Terminals where GUIs are not available
● I prefer both

Figure A.10: Complete survey form questions and responses – part 10



When attaching Machine Learning models to applications, What do you think is the best out of the following. 😎

110 responses



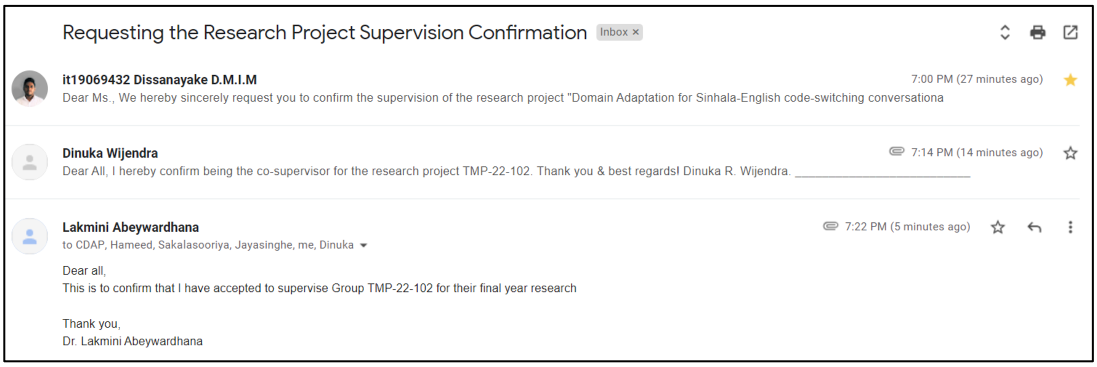
Would you train a model yourself or download a model that has been already trained and free to use?

110 responses

Thank you! 🙇

Figure A.11: Complete survey form questions and responses – part 11

Appendix B: Supervision Confirmation Emails



Requesting the Research Project Supervision Confirmation

it19069432 Dissanayake D.M.I.M 7:00 PM (27 minutes ago)

Dear Ms., We hereby sincerely request you to confirm the supervision of the research project "Domain Adaptation for Sinhala-English code-switching conversationa

Dinuka Wijendra 7:14 PM (14 minutes ago)

Dear All, I hereby confirm being the co-supervisor for the research project TMP-22-102. Thank you & best regards! Dinuka R. Wijendra. ________________________

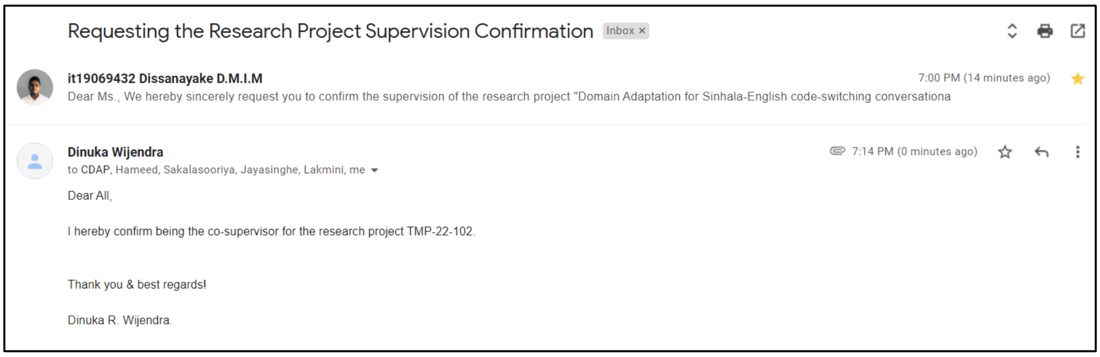
Lakmini Abeywardhana 7:22 PM (5 minutes ago)

to CDAP, Hameed, Sakalasooriya, Jayasinghe, me, Dinuka

Dear all,
This is to confirm that I have accepted to supervise Group TMP-22-102 for their final year research

Thank you,
Dr. Lakmini Abeywardhana

Figure B.1: Research project supervision confirmation email



Requesting the Research Project Supervision Confirmation

it19069432 Dissanayake D.M.I.M 7:00 PM (14 minutes ago)

Dear Ms., We hereby sincerely request you to confirm the supervision of the research project "Domain Adaptation for Sinhala-English code-switching conversationa

Dinuka Wijendra 7:14 PM (0 minutes ago)

to CDAP, Hameed, Sakalasooriya, Jayasinghe, Lakmini, me

Dear All,

I hereby confirm being the co-supervisor for the research project TMP-22-102.

Thank you & best regards!

Dinuka R. Wijendra.

Figure B. 2: Research project co-supervision confirmation email