Deakin University

SIG788- OnTrack Submission

Task 5.2 P

Submitted by

Neethu Sidhardhan
s223494027
Attempt # 1
23/4/2023

Target Grade: P

Task Details – Natural language processing on Azure

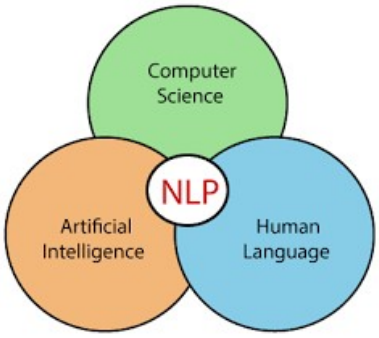
- Summarise the main points that is covered in this week.
- Reflect on the knowledge that you have gained by reading the contents and attending the sessions of this week.

Natural Language Processing (NLP)

Natural Language Process refers to a branch of computer science and more specifically to artificial intelligence. Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc.

The field of NLP involves making computers to perform useful tasks with the natural language's humans use. The input and output of an NLP system can be −

- Speech
- Written Text



Impact of NLP in our daily life:

1. **Language Translation**: - If a person who doesn't understand Spanish and he need to understand what the contents are written in Spanish. For that purpose, we use the translator. Google Translator is one feature that will help in translating from English too Spanish and too other specific languages.

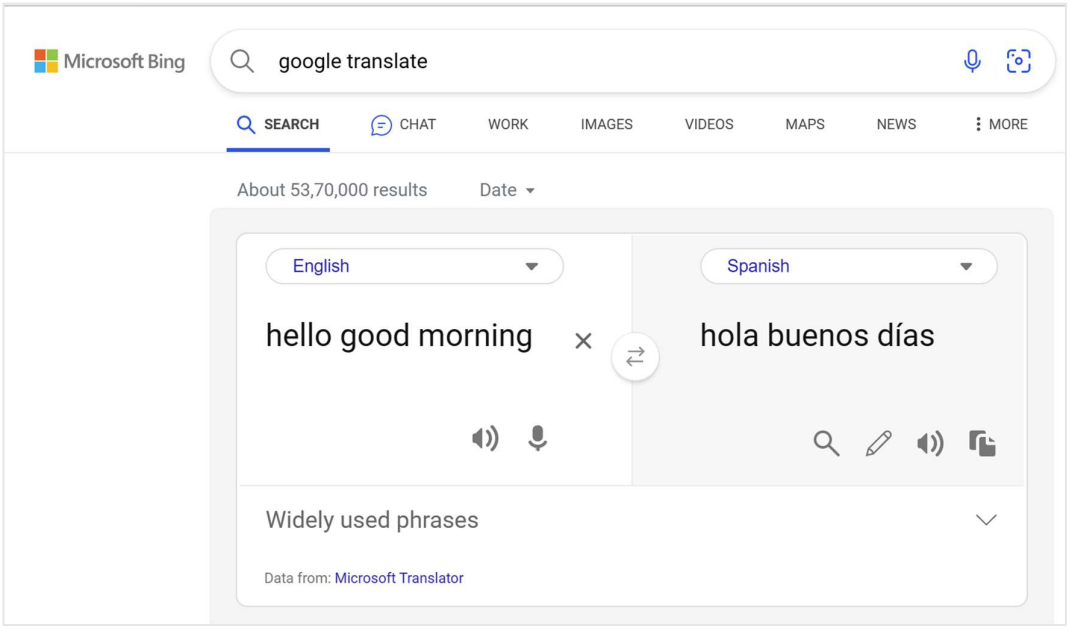


Fig 1 Google Translator

2. **Voice Assistance**: - It's a software that carries out everyday task via voice command. For example, Amazon Alexa, Siri, Google Maps, Google Assistant, Data Bot, etc. In Google Assistant can help in making a call, writing an email as per your narration, even sending the mail, taking a picture, like that so many things are worked out in voice assistance.

Fig 2: Voice Assistance

Applications involved in NLP:

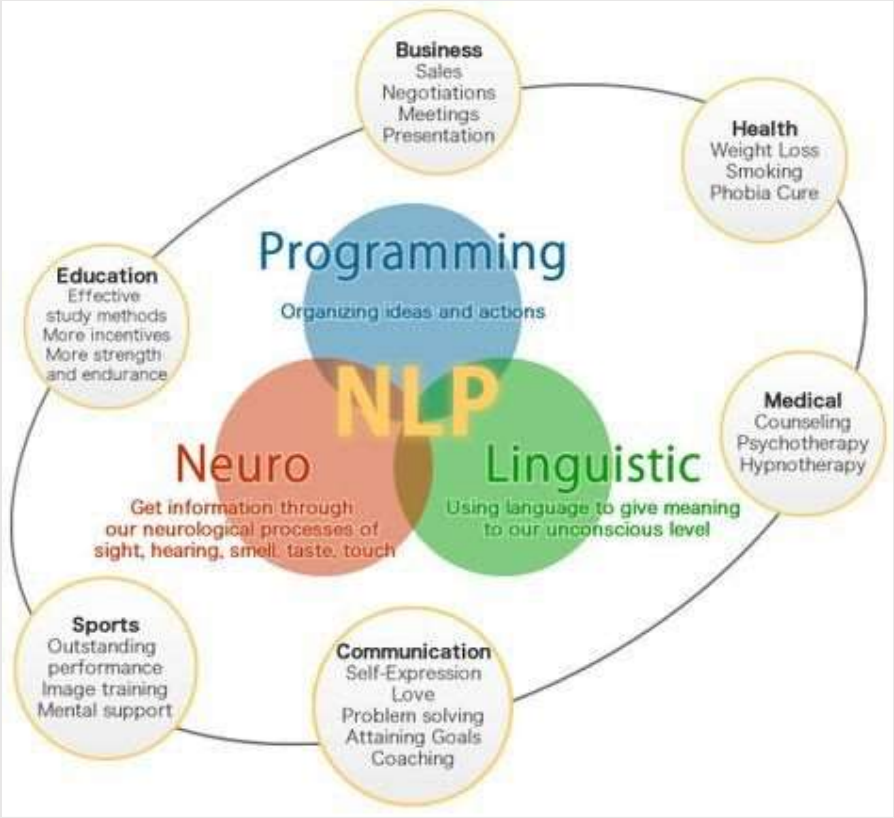


Fig 3 NLP Applications

Now, NLP has become an integral part of the industry. Most of the industries have implemented artificial intelligence in their organisation for better business growth.

- Health care: NLP is one of the ways used to collect data for diagnosing the patients.
- Sports: NLP is used for understanding the performance as well as mental support of the players
- Education: Various study reference materials can be generated by using NLP.

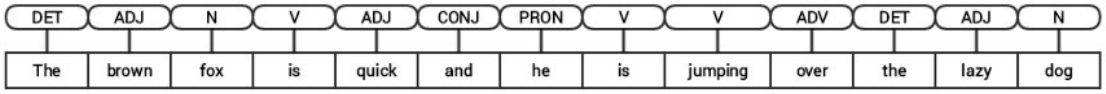
Basic Concepts of NLP:

Natural Language Process is one of the methods used for data mining. We get different types of reviews in Twitter, Google, Yahoo, LinkedIn, etc. For human being, its very easy to understand and recognise the concepts. When this comes to business, we need machine language that helps the business in interpreting and analysing the data, in the same way human beings communicate.

The below mentioned are the various concepts of NLP.

1. **Stop-words**: There are some commonly used words that humans used while interacting. But those words do not make any sense or add any value. Those words are deleted from the data considering as stop words.

2. **Lower Case**: There should be a uniformity in data. Hence, all the words are converted to lower case.
3. **Lemmatization**: This helps to reduce words into single form.
4. **Stemming**: This helps in reduce the words into their root form. There are 2 more types of stemming apart from *Porter Stemmer*. Those are Lancaster Stemming and Snowball. Snowball is an improvement over Porter stemming.
5. **Parts of Speech (POS)**: Parts of speech (POS) are specific lexical categories to which words are assigned, based on their role and context in a given sentence.



For example, in the above sentence, “The brown fox is quick, and he is jumping over the lazy dog,” the abbreviations denote the following parts of speech; DET: Dependency tag, ADJ: Adjective, N: Noun, V: Verb CONJ: Conjunction (coordinating), PRON: Pronoun, ADV: Adverb.

6. **Tokenization**: This helps to split a phrase, sentence, or paragraph into small units like words or terms. Each unit is called a token. There are different types of tokenization
7. **Vectorization/Word Embedding**: Vectorization helps to map the words to a vector of real numbers, which further helps into predictions. This helps to extract the important features.

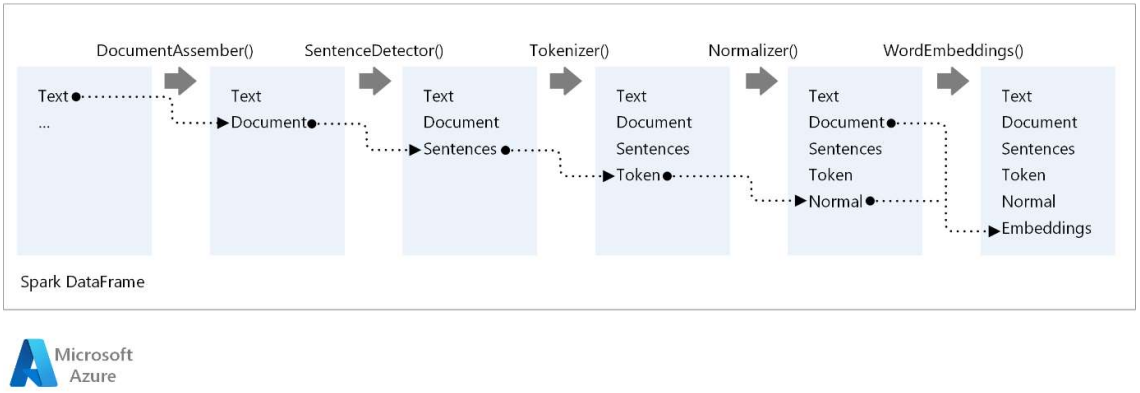


Fig 4: NLP Pipeline

Natural Language Process on Azure

Apache Spark is a parallel processing framework that supports in-memory processing to boost the performance of big-data analytic applications. Azure Synapse Analytics, Azure HDInsight, and Azure Databricks offer access to Spark and take advantage of its processing power.

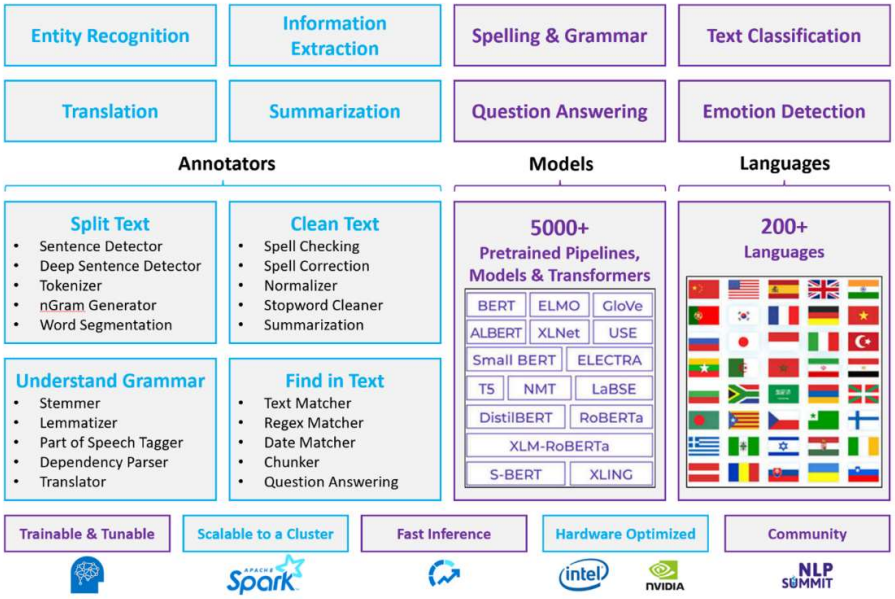


Fig 5 NLP process in Azure (Pic courtesy: Microsoft)

Azure Cognitive Service Language

Azure Cognitive Service for Language is a cloud-based service that provides Natural Language Processing (NLP) features for understanding and analysing text. Use this service to help build intelligent applications using the web-based Language Studio, REST APIs, and client libraries.

Features:

- **Named Entity Recognition (NER):** Named entity recognition is a preconfigured feature that categorizes entities (words or phrases) in unstructured text across several predefined category groups. For example: people, events, places, dates, and more.

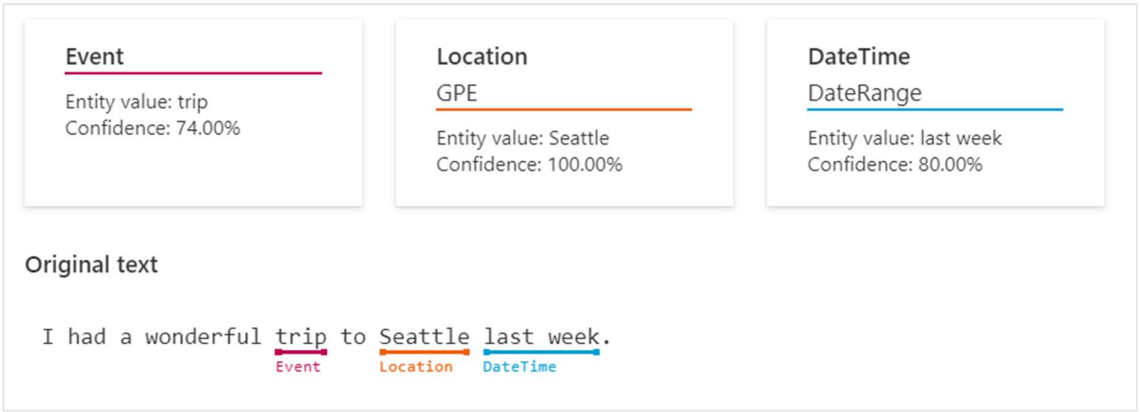


Fig 6: NER (Pic courtesy: Microsoft learning)

- **Language Detection**: Language detection is a preconfigured feature that can detect the language a document is written in, and returns a language code for a wide range of languages, variants, dialects, and some regional/cultural languages.

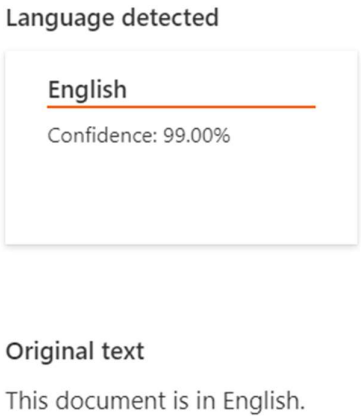


Fig 7: Language Detection

- **Conversational language understanding:** Conversational language understanding (CLU) enables users to build custom natural language understanding models to predict the overall intention of an incoming utterance and extract important information from it.

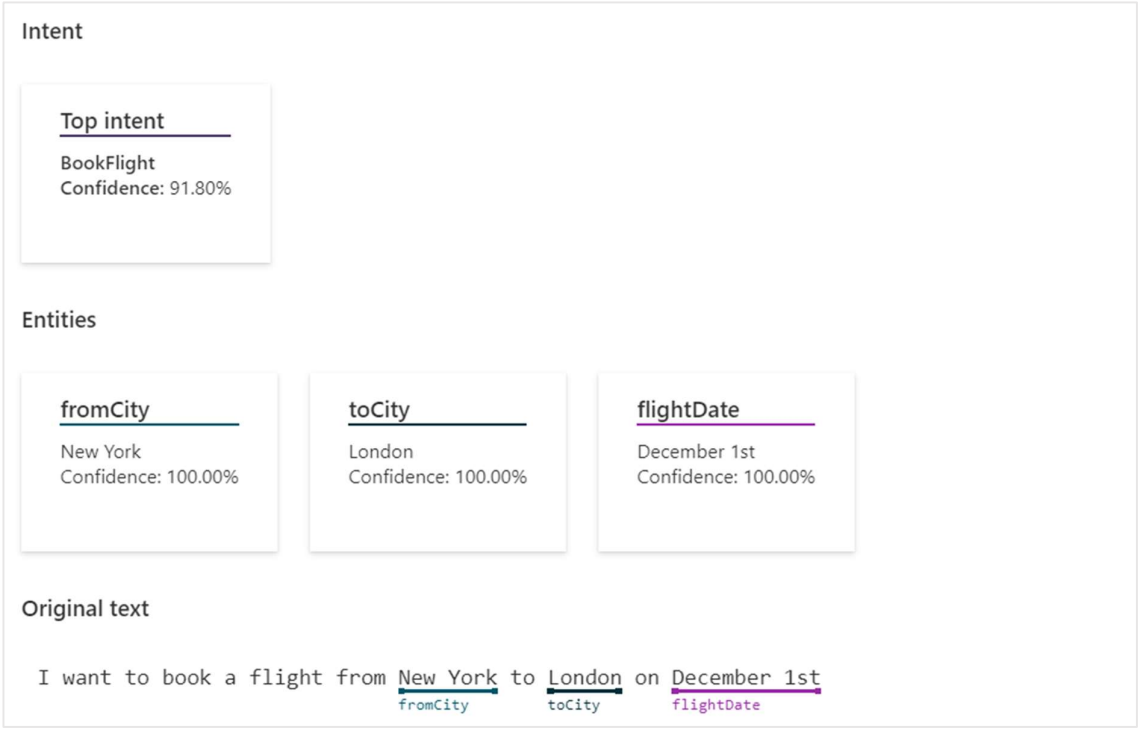


Fig 8: CLU

Intent: An intent represents an action that the user wants to perform and the entity represents a keyword that you want to be extracted from the user utterance. It refers to the goal the user has in mind while typing question or comment.

Entities: An entity represents values which is collected from the user in a conversation. Depending on the context of the conversation, the required response can either be a single value or group of specific values.

Utterance: Utterances can be referred to as phrases or trigger words used to activate an artificially intelligent model

- **Question Answering**: Question answering is a custom feature that finds the most appropriate answer for inputs from your users, and is commonly used to build conversational client applications, such as social media applications, chat bots, and speech-enabled desktop applications.

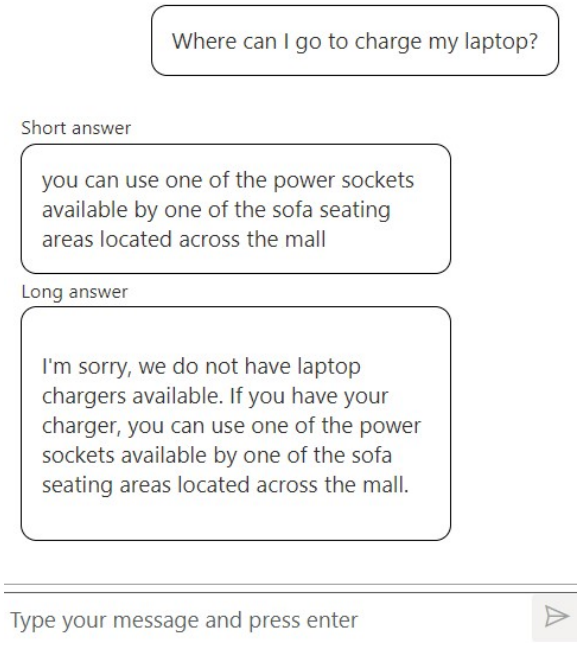


Fig 9: Q&A

Conversational Language Understanding (CLU):

Conversational language understanding is one of the custom features offered by Azure Cognitive Service for Language. It is a cloud-based API service that applies machine-learning intelligence to enable you to build natural language understanding component to be used in an end-to-end conversational application.

Conversational language understanding (CLU) enables users to build custom natural language understanding models to predict the overall intention of an incoming utterance and extract important information from it.

To simplify building and customizing your model, the service offers a custom web portal that can be accessed through the Language studio

CLU can be used in multiple scenarios across a variety of industries. Some examples are

- End-to-end conversational bot
- Human assistant bots
- Command and control application
- Enterprise chat bot

Project development lifecycle

Creating a CLU project typically involves several different steps.

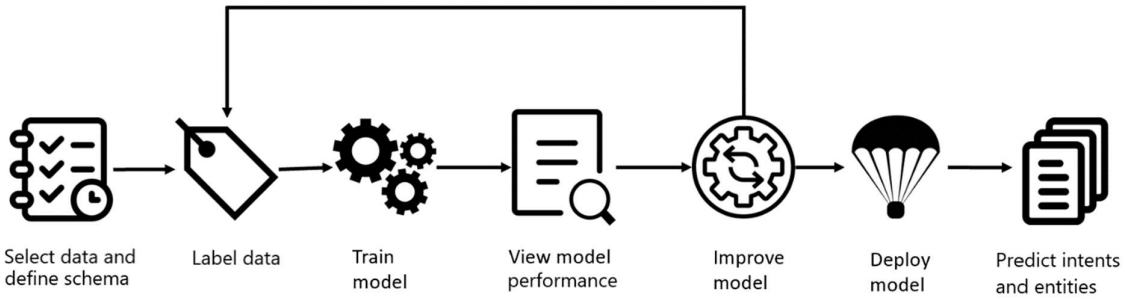


Fig 10: Project Development lifecycle

Hands on in CLU

1. **Login to Language Studio:** https://language.cognitive.azure.com/clu/projects

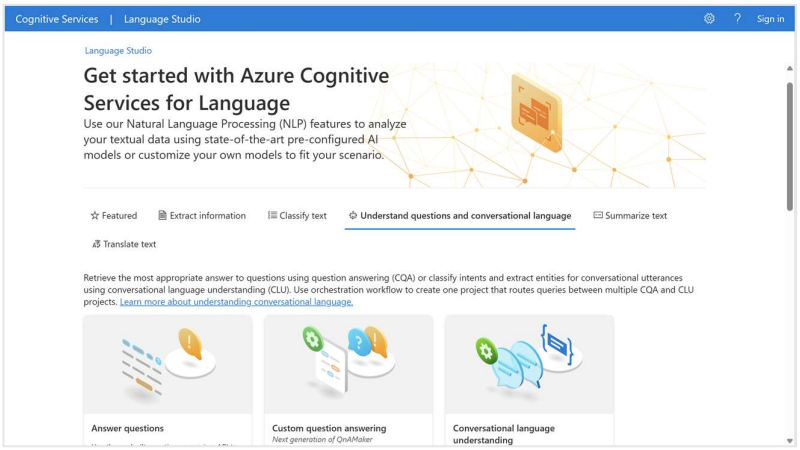


Fig 11: Language Studio

2. **Create Language resources and deploy the language group**

3. **Create a Conversational Project**

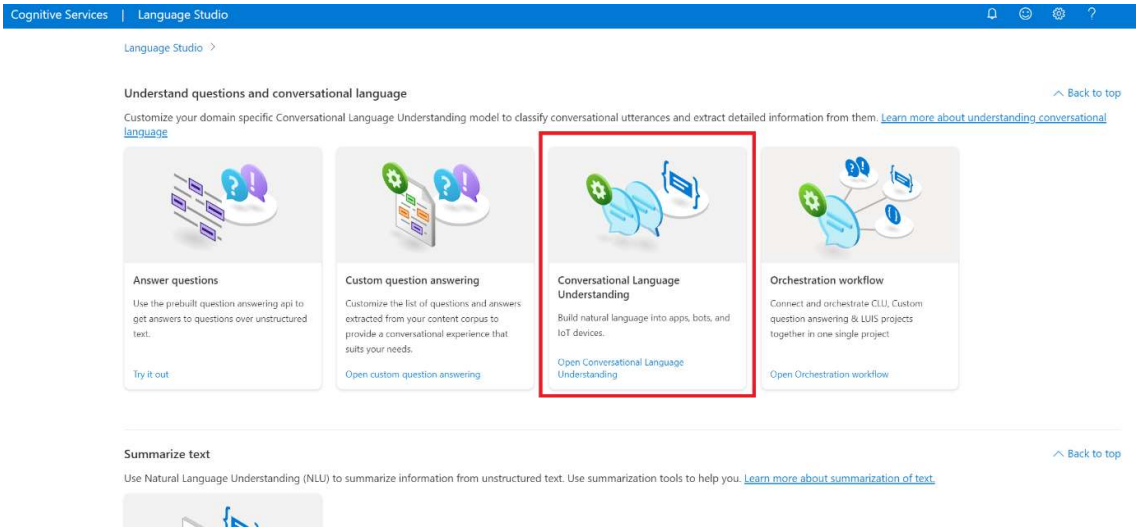


Fig 12: Create Conversational Project

3.1 Define your schema: Know your data and define the actions and relevant information that needs to be recognized from user's input utterances. In this step you create the intents that you want to assign to user's utterances, and the relevant entities you want extracted.

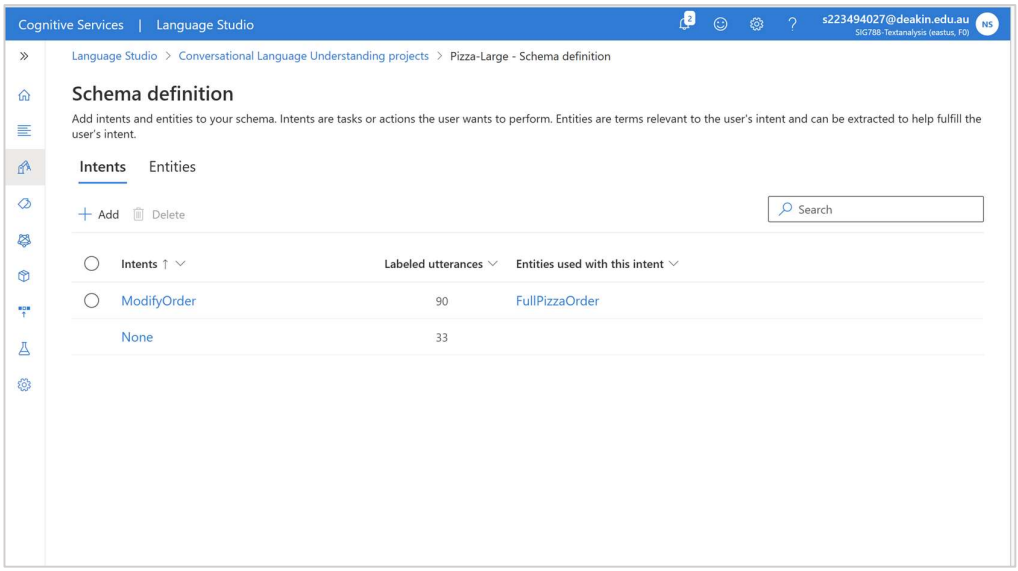


Fig 13: Schema Definition

3.2 Label your data

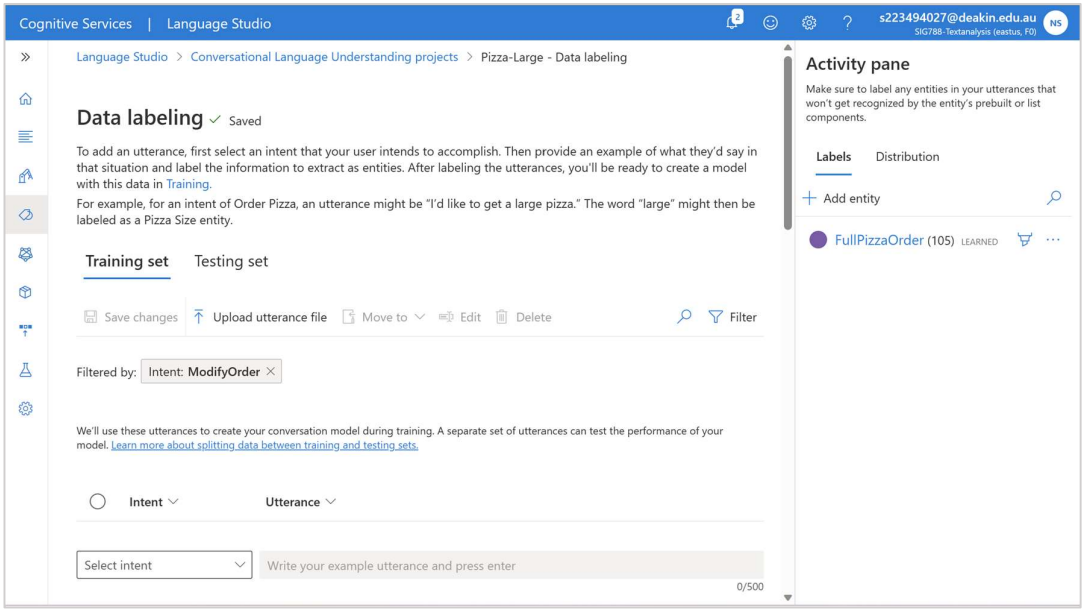


Fig 14: Data Labelling

3.3 Train the model

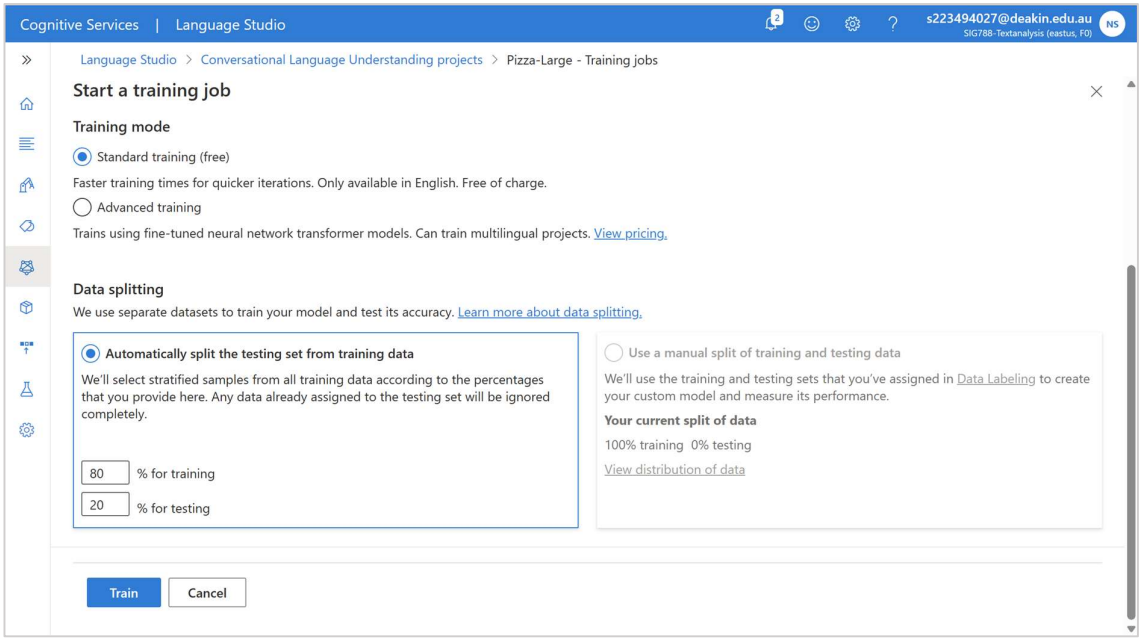


Fig 15: Train Model

3.4 View the model performance:

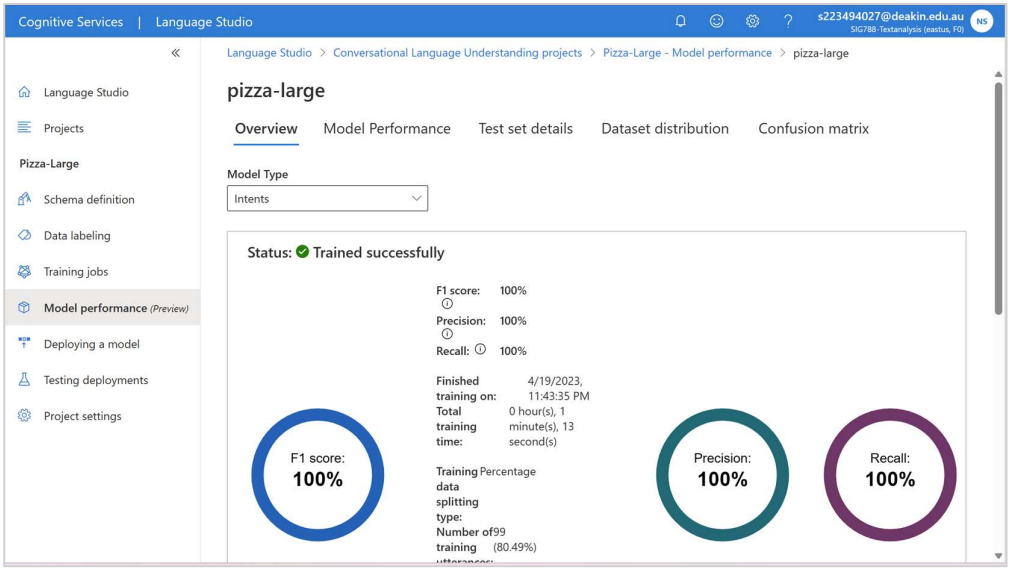


Fig 16:Performance

3.5 Improve the model: After reviewing the model's performance, you can then learn how you can improve the model.

3.6 Deploy the model:

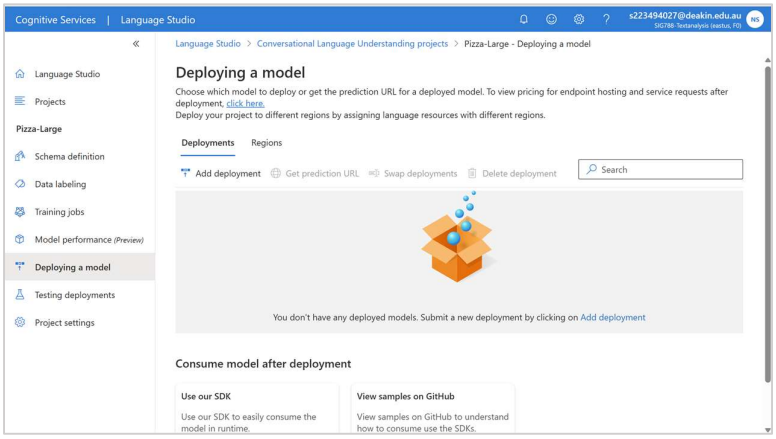


Fig 17: Deploy the model

3.7 Predict intents and entities: Use your custom model to predict intents and entities from user's utterances.

Evaluation metrics for conversational language understanding models

Data set is split into two parts. A set of training and a set of testing. The training data is used to train the data while the testing set is used as a test for model after training to calculate the model performance and evaluation.

Model evaluation is triggered automatically after training is completed successfully. The evaluation process starts by using the trained model to predict user defined intents and entities for utterances in the test set, and compares them with the provided tags (which establishes a baseline of truth). The results are returned so you can review the model's performance. For evaluation, conversational language understanding uses the following metrics:

- Precision (Accuracy of the model)
- Recall (model's ability to predict positive classes)
- F1 Score (function of precision and recall)

Confusion matrix

A Confusion matrix is an N x N matrix used for model performance evaluation, where N is the number of entities or intents. The matrix compares the expected labels with the ones predicted by the model.

- The values in the diagonal are the True Positive values of each intent or entity.
- The sum of the values in the intent or entities rows (excluding the diagonal) is the false positive of the model.
- The sum of the values in the intent or entities columns (excluding the diagonal) is the false Negative of the model

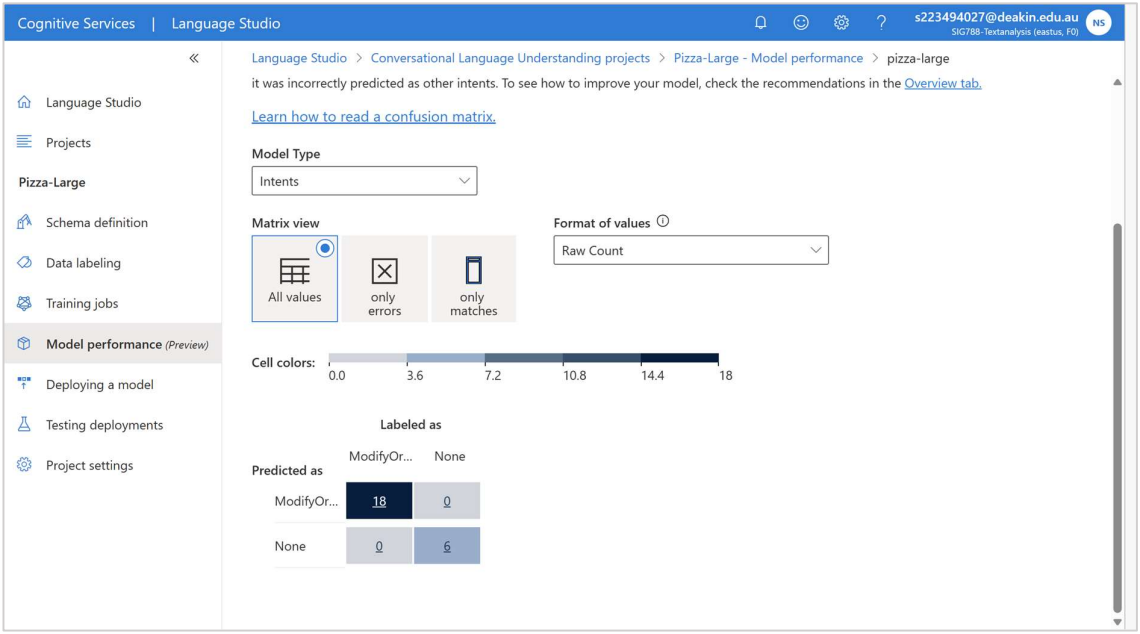


Fig 18: Confusion Matrix

Question Answering

Question answering provides cloud-based Natural Language Processing (NLP) that allows you to create a natural conversational layer over your data. It is used to find appropriate answers from customer input or from a project.

Question answering provides cloud-based Natural Language Processing (NLP) that allows you to create a natural conversational layer over your data. It is used to find appropriate answers from customer input or from a project.

Question answering comprises of two capabilities:

Custom question answering: Using this capability users can customize different aspects like edit question and answer pairs extracted from the content source, define synonyms and metadata, accept question suggestions etc.
Prebuilt question answering: This capability allows users to get a response by querying a text passage without having the need to manage knowledgebases.

Hand on with Question Answering

1. **Login to Language Studio:** https://language.cognitive.azure.com/clu/projects

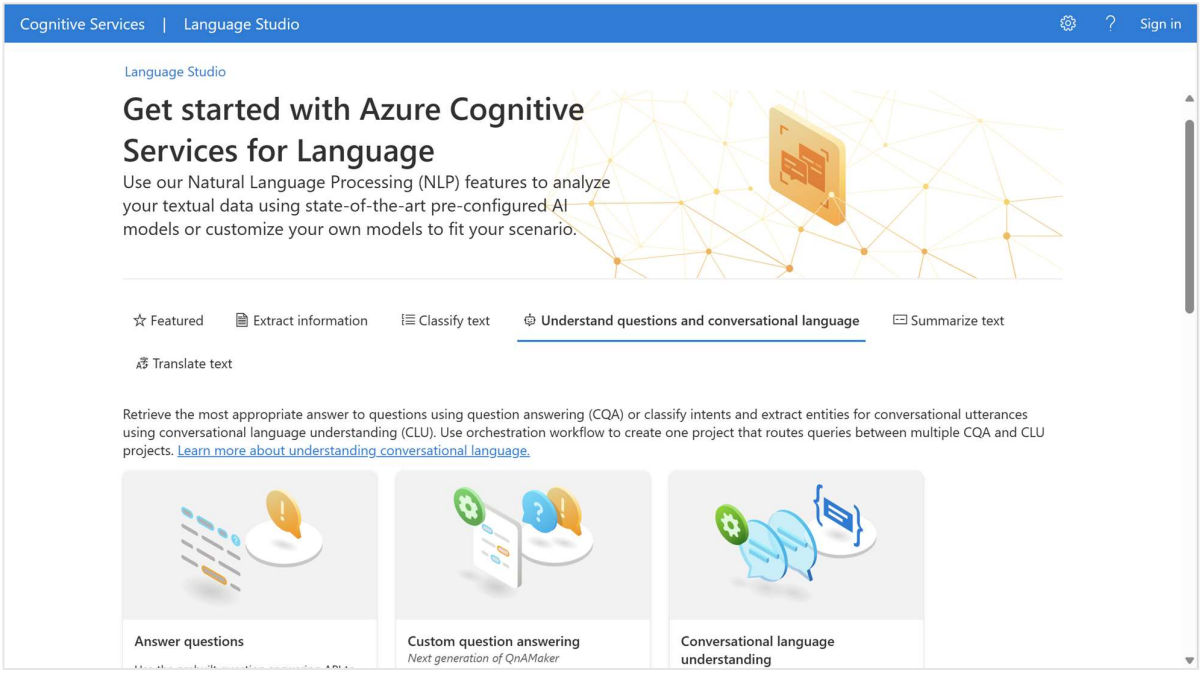


Fig 19: Language Studio

2. **Open Custom question answering**

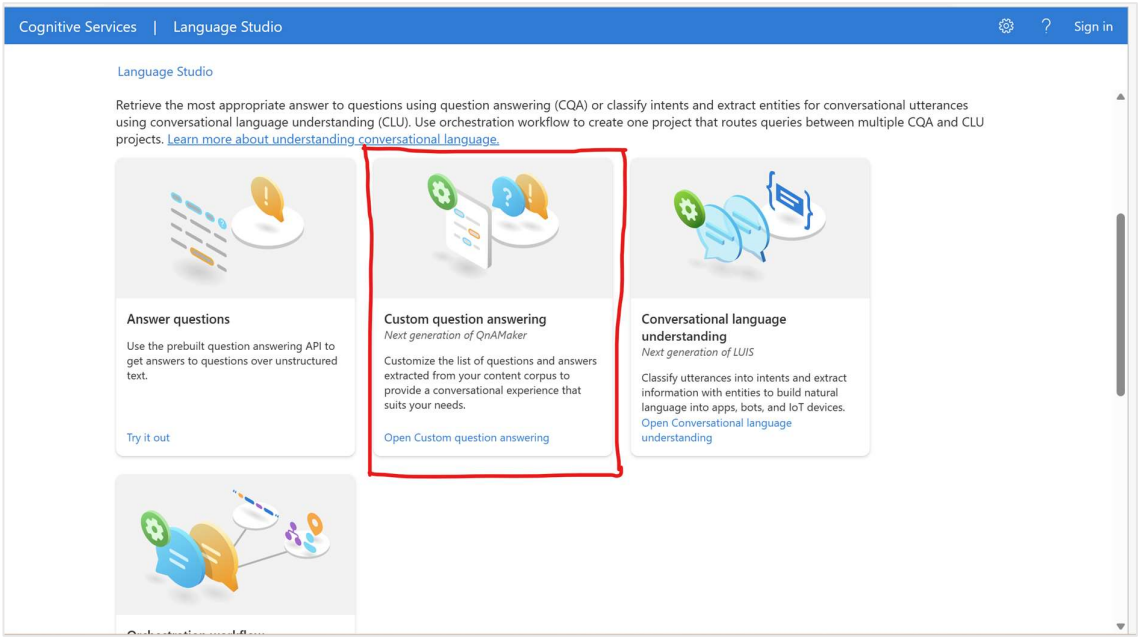


Fig 20: Custom Question Answering screen

3. If your resource is not yet connected to Azure Search select Connect to Azure Search. This will open a new browser tab to Features pane of your resource in the Azure portal.

4. Create a new project

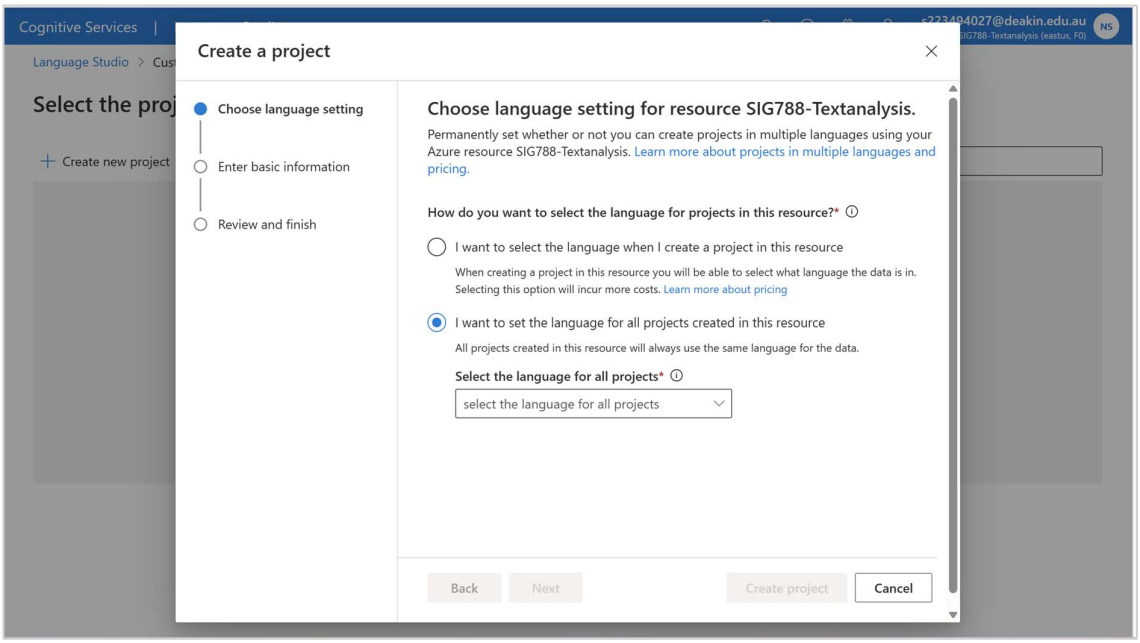


Fig 21: Create a new project

We can create projects by using URL, Files or Chit Chat.

5. Edit the knowledge base: Here you can edit the questions as well as answers before proceeding with test.

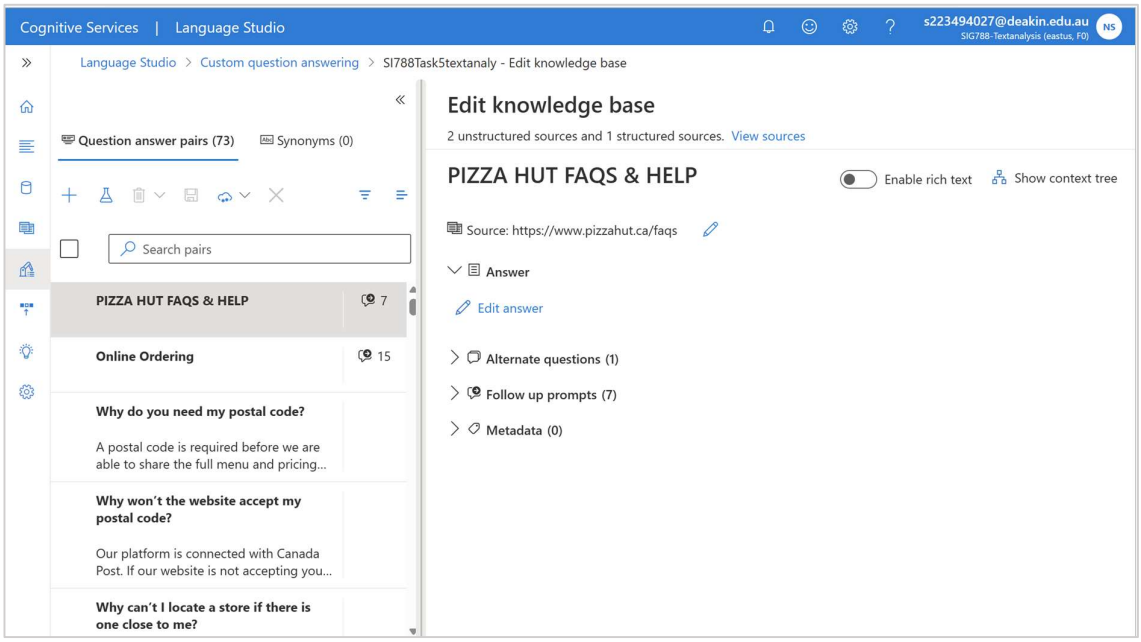


Fig 22: Edit knowledge base

6. Test the Question and Answering

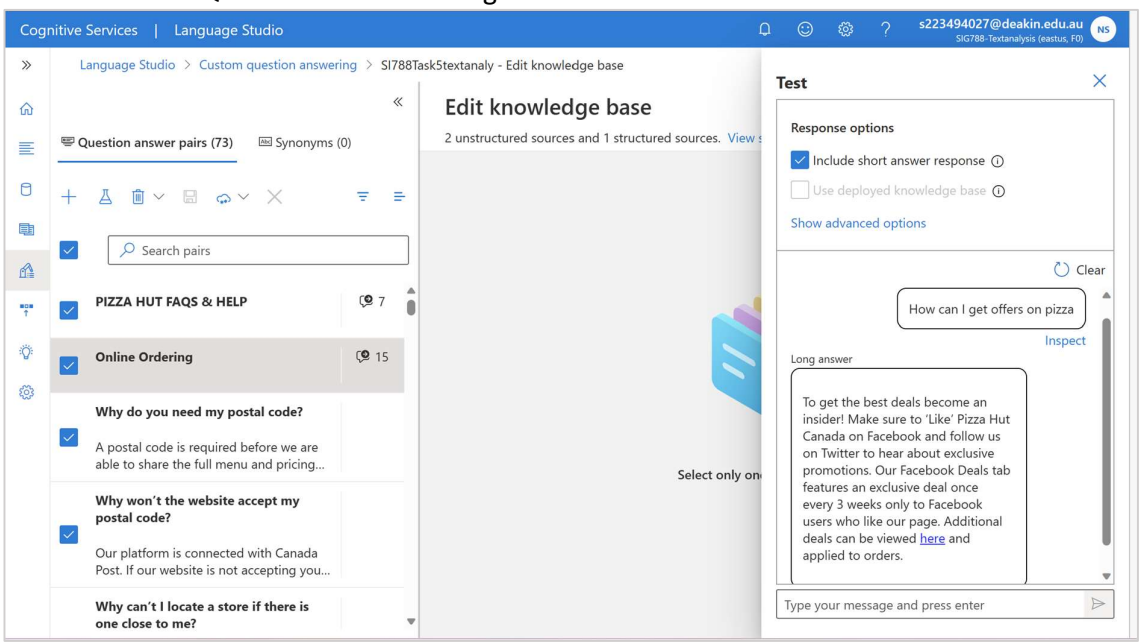


Fig 23: Test the Q&A

7. Deploy the knowledge base.

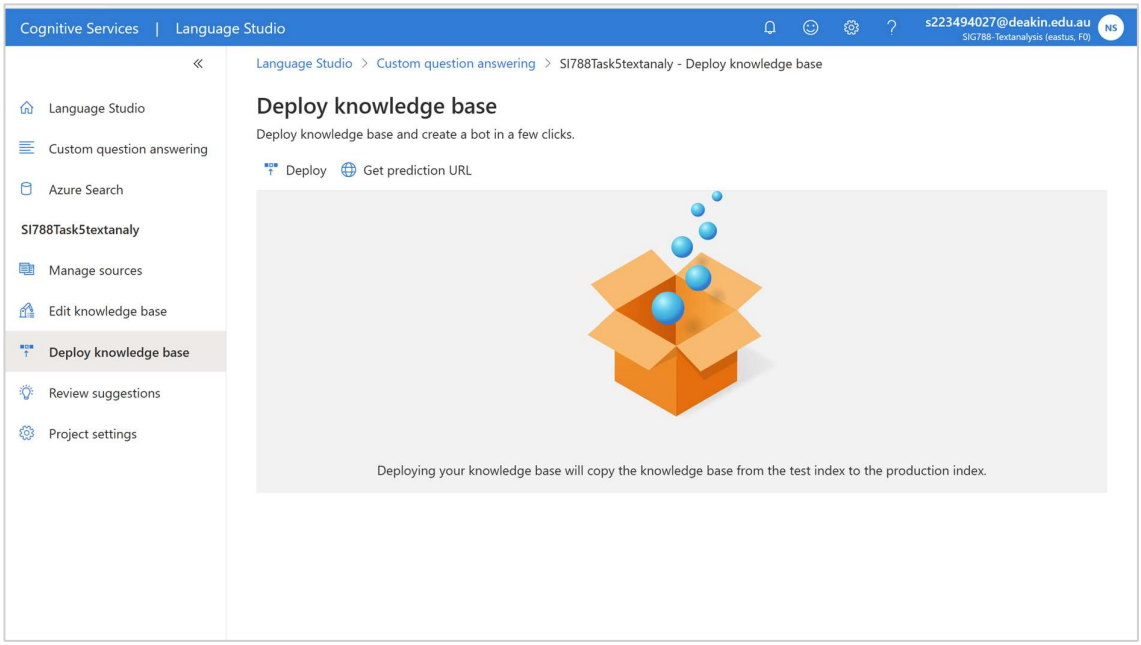


Fig 24: Deploy knowledge base

Once the knowledge base is successfully deployed, we get the option to create a bot.

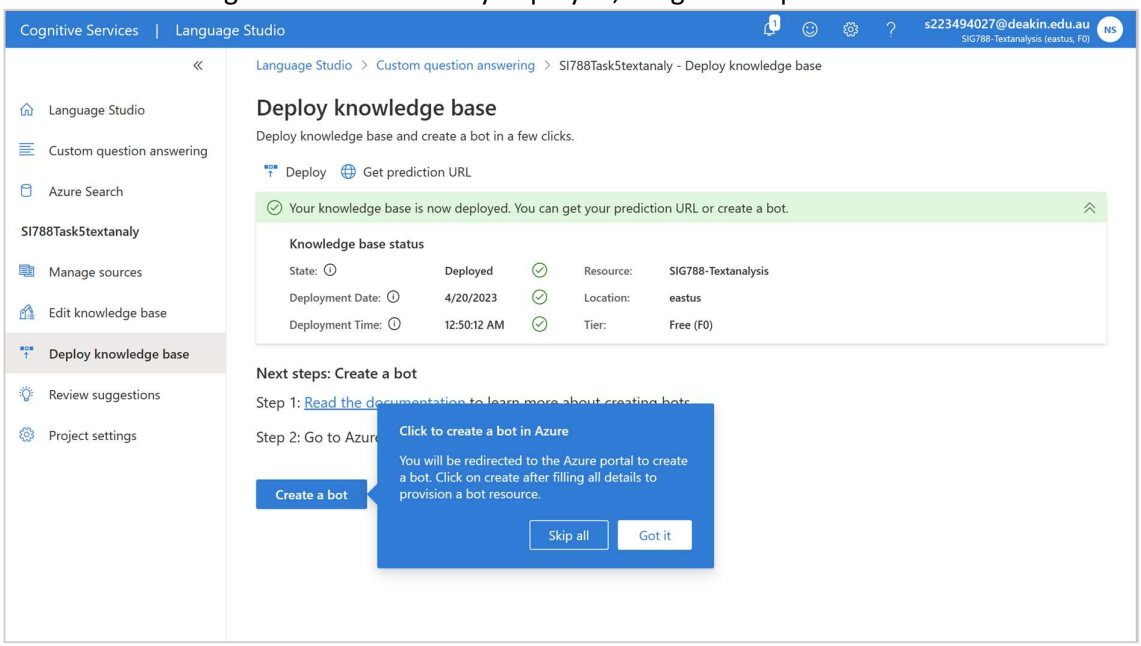


Fig 25: Create a bot

The page redirects to Azure portal where the model need to be deployed and tested in Web chat. The chat bot can be attached to different channels like Team, Twitter, LinkedIn, Facebook, Instagram, Skype, etc.

References

1. Anon., 2023. *Tutorialspoint.* [Online]
   Available at:
   https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.htm
   [Accessed 23 April 2023].
2. Anon., n.d. *Haptika.* [Online]
   Available at: https://docs.haptik.ai/nlu/what-is-the-difference-between-intent-and-entity#:~:text=An%20intent%20represents%20an%20action%20that%20the%20user,issue%2C%20the%20intent%20is%20what%20they%20really%20mean.
   [Accessed 23 April 2023].
3. Anon., n.d. *towardsai.* [Online]
   Available at: https://towardsai.net/p/nlp/natural-language-processing-concepts-and-workflow-48083d2e3ce7
   [Accessed 23 April 2023].
4. Biswas, P., 2021. *Towards Data Science.* [Online]
   Available at: https://towardsdatascience.com/basic-concepts-of-natural-language-processing-nlp-models-and-python-implementation-88a589ce1fc0
   [Accessed 23 April 2023].
5. IBM, n.d. *IBM.* [Online]
   Available at: https://www.ibm.com/in-en/topics/natural-language-processing#:~:text=Natural%20language%20processing%20(NLP)%20refers,same%20way%20human%20beings%20can.
   [Accessed 23 April 2023].
6. Microsoft, n.d. *https://learn.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/natural-language-processing.* [Online]
   Available at: https://learn.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/natural-language-processing
   [Accessed 23 April 2023].
7. Microsoft, n.d. *https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/.* [Online]
   Available at: https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/
   [Accessed 23 April 2023].
8. Microsoft, n.d. *https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/conversational-language-understanding/how-to/create-project?tabs=language-studio%2CLanguage-Studio.* [Online]
   Available at: https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/conversational-language-understanding/how-to/create-project?tabs=language-studio%2CLanguage-Studio
   [Accessed 23 April 2023].
9. Microsoft, n.d. *https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/question-answering/overview.* [Online]
   Available at: https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/question-answering/overview
   [Accessed 23 April 2023].
10. Microsoft, n.d. *learn.microsoft.* [Online]
    Available at: https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/conversational-language-understanding/overview
    [Accessed 23 April 2023].
11. Nakisa, D. B., n.d. *Great Learning.* [Online]
    Available at: https://olympus.mygreatlearning.com/courses/91262/pages/5-dot-1-natural-

language-processing-2?module_item_id=3833109
[Accessed April 2023].