République Tunisienne

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Gabès
Faculté des Sciences de Gabès

A. U. : 2021-2022

N° Ordre : LGLSI3<N°6>

Projet de Fin d'Études

Présenté à :

La Faculté des Sciences de Gabès

Département Informatique

En vue de l'obtention du diplôme de

LICENCE GENIE LOGICIELS ET SYSTEMES D'INFORMATION

TITRE DU PROJET :

CHATBOT POUR L'ORIENTATION DES CLIENT « GCT »

Réalisé Par :

Nom et prénom de l'étudiant

Soutenue le 04/07/2022
Devant le jury composé de :

M.	**Tahani Bouchrika**	***Président***
M.	**Soumaya Zoghbani**	***Examinateur***
M.	**Nouha Arfaoui**	***Encadrant***
M.		***Invité***

Acknowledgment:

I would like to express my sincere thanks and my deep gratitude to Mrs. Najet soudi for having welcomed me in this way for her support, her advice and her trust during my further training.
My thanks to the teaching team of the school, for having brought me the theoretical and practical help necessary for the smooth running of the course.
And more particularly Ms. Nouha Arfaoui for having been my academic supervisor during my end of study internship.

Table des matières

Projet de Fin d'Études 1
La Faculté des Sciences de Gabès 1
General Introduction 7
Chapter I : Preliminary study and Specification of Needs 8
1 - Presentation of the host organization 8
2 - Preliminary study 8
2.2 Objectives to be achieved 9
3 - Analyses of needs 10
3.1 Functional needs 10
3.2 Non-functional needs 10
4 - Use cases 10
4.1 Definition of UML 10
4.2 Identification of actors: 10
4.3 Initial use case diagram 10
Chapter II : Conceptual study 12
1- Refinement of use cases: 12
1.1 Refinement of use cases "ask a question": 12
1.2 Refinement of use cases "Provide feedback": 13
1.3 Refinement of use cases "Add question/response": 14
1.4 Refinement of "update responses" use cases: 15
2- Sequence diagrams: 16
2-1 Sequence Diagram: "Ask a Question" Scenario 16
2-2 Sequence Diagram: "Add Questions/Answers" Scenario 17
2-3 Sequence Diagram: "Response Update" Scenario 17
Chapter III : Realization 18
1 Working environment 18
1.1 Hardware environment 18
1.2 Software environment 18
2 Choix Techniques 20
2.1- What algorithm is used? 20
2.2 Which "Deep Learning Framework" is used? 21
2.3 Stemming vs Lemmatizing 21
3 Application Overview 22
3.1 General architecture 22
3.2 Front-End 22
3.3 Backend 26

Chapter IV Conclusion 29
Chapter V Bibliography 30

List of Figures

Figure I-1 Logo de GCT 8
Figure I-2 Poncho chatbot fail 9
Figure I-3 Chatbot fail 3 9
Figure I-4: Initial use case diagram 11
Figure II-1: Use case diagram "asks a question" 12
Figure II-2: Use case diagram "Provide feedback" 13
Figure II-3: Use case diagram "Add question/response" 14
Figure II-4: update responses Use Case Diagram 15
*Figure II-5: Sequence diagram "asks a question"" * 16
Figure II-6 : "add new question/answer" sequence diagram 17
Figure II-7 : Sequence diagram "response update" 17
Figure III-1 : Vs code 18
Figure III-2 Logo google Collab 18
Figure III-3 HTML logo 19
Figure III-4 JavaScript Logo 19
Figure III-5 CSS Logo 19
Figure III-6 : Django Logo 20
Figure III-7 flask Logo 20
Figure III-8 Python Logo 20
Figure III-9 Stemming vs Lemmatization [14] 21
Figure III-10 General Architecture 22
Figure III-11 chatbot UI closed 22
Figure III-12 Chatbot UI open 23
Figure III-13 Example of a chat 23
Figure III-14 Example of a chat 2 24
Figure III-15 Example of a message Unidentified 24
Figure III-16 Example of "bad word censor" 25
Figure III-17 le chatbot understands a mistype 25
Figure III-18 a question not in database 26
Figure III-19 intents dans le bot 26
Figure III-20 add intent 1 26
Figure III-21 add intent 2 27
Figure III-22 if post retrain model 27
Figure III-23 Unidentified message is added 27
Figure III-24 Api database tables 28

List of Tables

Tableau 1 : Descriptions of the "ask a question" use case _____ *12*
Tableau 2: Use case descriptions "Provide feedback" _____ *13*
Tableau 3: Use case descriptions 'Add question/response" _____ *14*
Tableau 4: Descriptions of the "update responses" use case _____ *15*
Tableau 5 keras vs tf vs pytorch [12] _____ *21*

General Introduction

After the rise of the web and mobile apps, virtual chatbot applications are the latest inventions of digital design. These applications are well known for automatic conversational agents that run on computer programming or a kind of artificial intelligence (AI) interaction between the users and machines with the intervention of natural language processing (NLP)

This presents in a report structured in 3 chapters

In the first chapter entitled "Preliminary study and Specification of Needs", we will present the framework of my project, namely the host organization, where I had carried out my internship. This chapter will present the analysis and criticism of the existing situation in order to propose adequate solutions, and the needs analysis to fully understand the context of the system and determine the main use cases.

In the second chapter called "Conceptual study", I had proposed a detailed design by presenting the refinements of the use cases, the class diagram as well as some sequence diagrams.

In the last chapter entitled "Realization", I presented the structure of the database, as well as the hardware and software environment and some application components produced.

Finally, I end this report with a general conclusion, also a set of annexes that contain some documents on which we based ourselves for the realization of this project.

Chapter I : Preliminary study and Specification of Needs

Introduction

In this chapter, we will present the GCT, it is the company in which I carried out my internship End of study project, Then I will study the shortcomings of the system used by the company or the absence of possibility of palliating them by proposing solutions, At the end of this chapter I will present the diagram of initial use case by specifying the actors.

1 - Presentation of the host organization

The Tunisian Chemical Group (GCT) is a Tunisian public company whose purpose is to produce and transform the phosphate extracted in Tunisia into chemicals such as phosphoric acid or fertilizers.

Its a result from the absorption or merger of companies operating in the fields of phosphate extraction (Compagnie des phosphates de Gafsa) and transformation — Industrial company of phosphoric acid and fertilizers in Sfax, Chemical industries in Gabes, Arab Society of Phosphate and Nitrogen Fertilizers in Gabes, Fertilizers in Gabes and Chemical Industry in Gafsa — between 1992 and 1994. [1]

Figure I-1 Logo de GCT

2 - Preliminary study

2.1 Criticism of the existing

A Chatbot is a computer program designed to initiate conversations with humans. For this purpose, it utilizes tools of Natural Language Processing. Some of the benefits of a chatbot are:

● Chatbot saves user and admin time.
● It provides instant service and improves user experience.
● It saves the cost of hiring a large number of customer service representatives.
● It automatically does most of the work for you (ie saves user-supplied data).
● It automates chat for you, which is the future of chat.
● This is a developing field and is expected to grow considerably over the next few years.

Now that you have learned the benefits of using chatbots, one can put their power to work to improve user happiness. Whenever it comes to adopting chatbots within society, unfortunately, there are major challenges:

•They are sensitive to data security breaches.
•They may misunderstand the user's point of view.
•Chatbot security.

The GCT doesn't have an implanted chatbot or an older model they used before so all I can do is see what's causing other sites' chatbots to fail and see what I can do to fix them.

The top 3 chatbot fails

Failed chatbot #1: uxchat.me

Reason: engaging in an inappropriate conversation.

Fix: Don't teach your bot to answer a question it doesn't understand.

Failed chatbot #2: poncho

Reason: Not programming the Chatbot correctly

Figure I-2 Poncho chatbot fail

Fix: This is where a chatbot's natural language processing (NLP) capabilities really matter. Implementing a bot that has NLP will mitigate this situation.

Failed chatbot #3

Reason: No margin of error

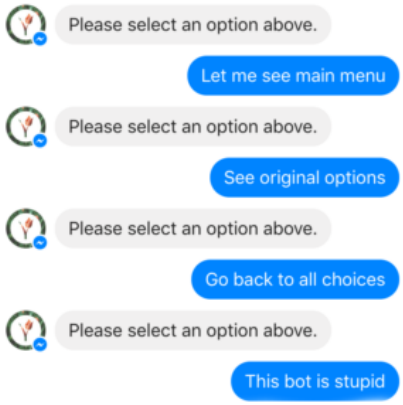

Figure I-3 Chatbot fail 3

Fix: make a prediction function with a margin of error variable. [2]

2.2 Objectives to be achieved

Goals:

- It offers 24/7 availability.
- Understands a language and responds in the same language
- Quality of conversation.

3 - Analyses of needs

3.1 Functional needs

Chatbots help improve the customer experience by helping or guiding customers through critical choices. We cite the most important needs of a chatbot:
• Complex dialogues: It can identify the intention of a question to provide a precise answer and suggest options to confirm or solve the problem.
• Training our bot, we have to train the bot and it gives a response after understanding the requests.
• Bad word filter
• Automatic saving of incomprehensible / unidentified questions
• Make the chatbot understand the mistype of user message

3.2 Non-functional needs

Non-functional requirements include the following:
• Security: Unauthorized users must not have access to the system.
• Scalability: The system must always function correctly, regardless of updates.
• Compatibility: The proposed system must be compatible with all web browsers.

4 - Use cases

4.1 Definition of UML

The **Unified Modeling Language (UML)** is a general-purpose, developmental, modeling language in the field of software engineering that is intended to provide a standard way to visualize the design of a system. [3]

4.2 Identification of actors:

Our application has two actors
The administrator: takes care of the most important roles in the system

- Add new questions/response
- Ensure the good quality of responses based on feedback

The user:

- Provides feedback
- Ask a question

4.3 Initial use case diagram

Use cases structure the need for uses and the goals of a system. They identify the users of the system (actor) and their interactions with the system, in addition the use cases make it possible to classify the actors and structure the objectives of the system

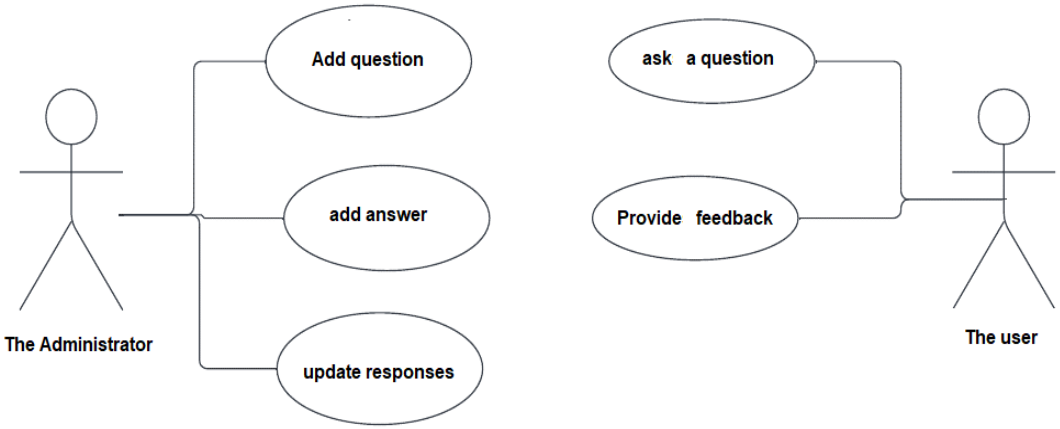

Figure I-4: Initial use case diagram

Conclusion: This chapter allowed us to detail the general framework of the system, the host company and the application that we will identify the different functionalities for.

Chapter II : Conceptual study

Introduction

To ensure the success of my project, a design step is essential to understand how our system works and to give a better result at the end. This chapter is for presenting the refinement of the use cases, then defining the class diagram and ends by presenting the sequence diagrams

1- Refinement of use cases:

1.1 Refinement of use cases “ask a question”:

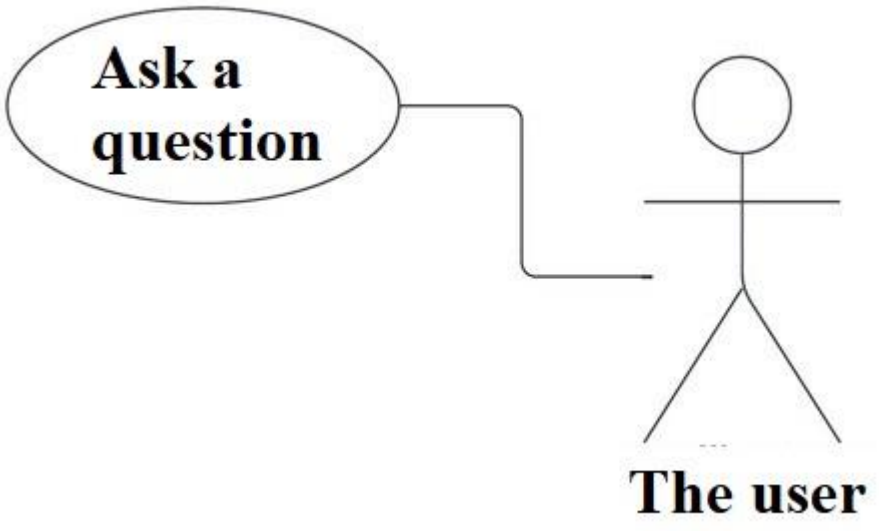

Figure II-1: Use case diagram “asks a question”

Use case: “ask a question”:

Use cases	Ask a question
Actor	The user
precondition	The chatbot is put online on the site
Post condition	The user receives the correct answer
Main scenario	The user enters the question

Tableau 1 : Descriptions of the “ask a question” use case

1.2 Refinement of use cases “Provide feedback”:

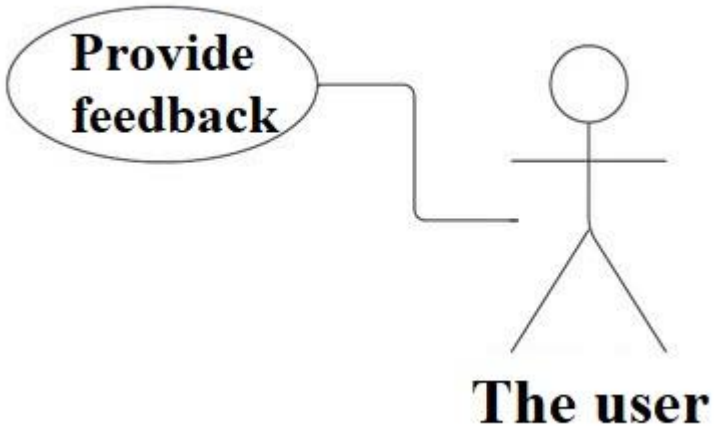

Figure II-2: Use case diagram “Provide feedback”

Use case: “provide feedback”

Use cases	Provide feedback
Actor	The user
precondition	The user asked a question and the answer was wrong/incorrect
Post condition	Answer was not satisfactory
Main scenario	The user sends an email to the administrator

Tableau 2: Use case descriptions “Provide feedback”

1.3 Refinement of use cases “Add question/response”:

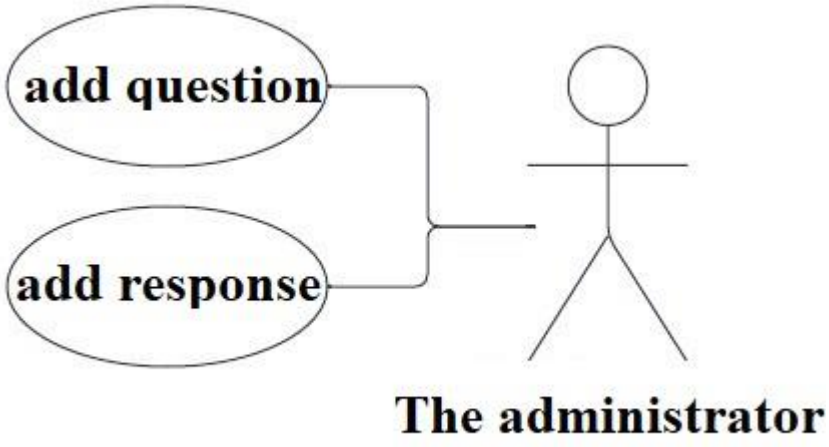

Figure II-3: Use case diagram “Add question/response”

Use cases: “Add question/answer”

Use cases	Add question/response
Actor	The Administrator
precondition	The Administrator has access to the database
Post condition	Added Question/Answer
Main scenario	The administrator adds a question/answer The API recycles the model

Tableau 3: Use case descriptions ’Add question/response”

1.4 Refinement of “update responses” use cases:

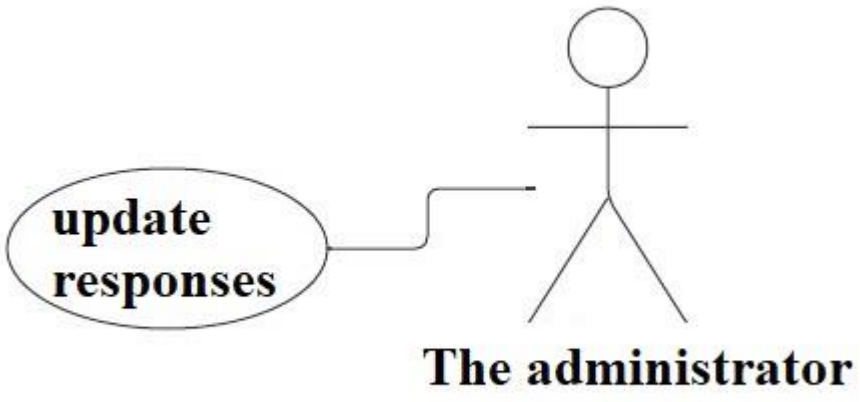

Figure II-4: update responses Use Case Diagram

Use cases: “update responses”

Use cases	response update
Actor	The Administrator
precondition	Admin receives feedback from users
Post condition	The administrator is logged in
Main scenario	The administrator checks his email/api He extracts questions/answers from feedback He updated the data The API recycles the model

Tableau 4: Descriptions of the “update responses” use case

2- Sequence diagrams:

The sequence diagram documents the interactions to be implemented between the classes to achieve a result, such as a use case. The sequence diagram lists objects horizontally and time vertically. It models the execution of the different messages as a function of time.

2-1 Sequence Diagram: “Ask a Question” Scenario

Figure II-5 describes the scenario:

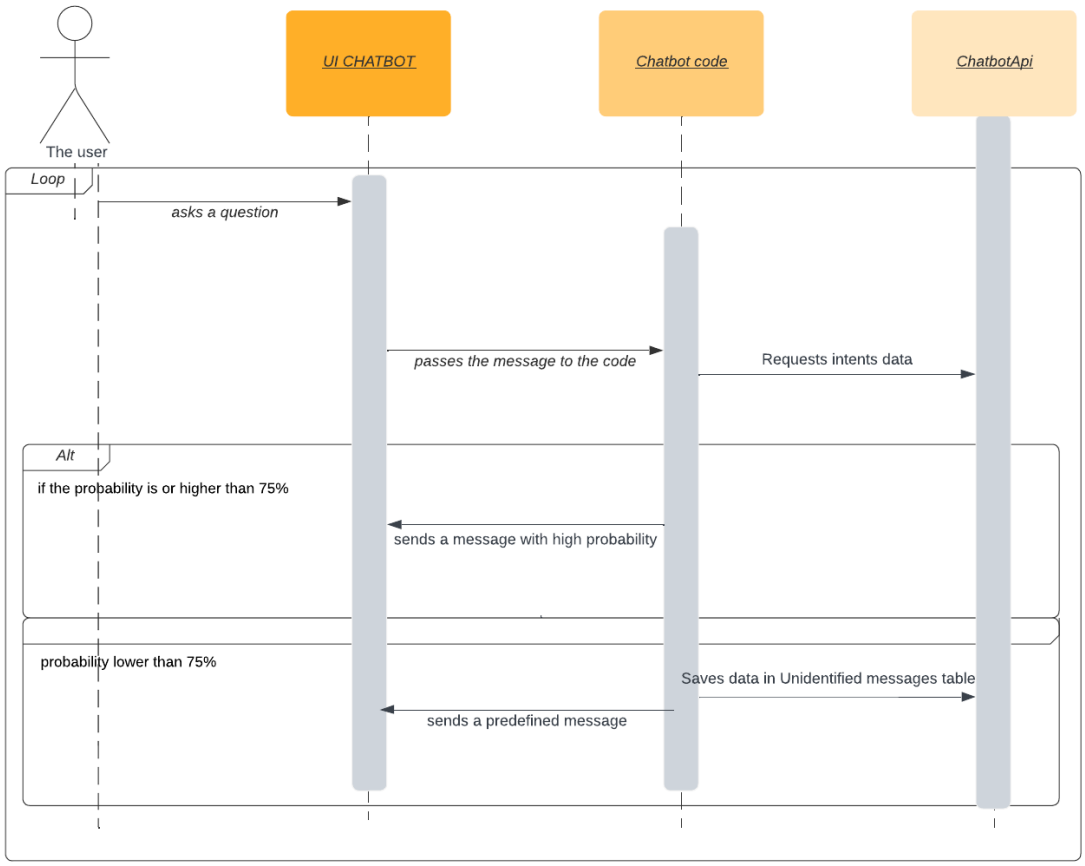

Figure II-5: Sequence diagram "asks a question""

When the user sends a message in the chatbot UI it launches the chatbot code , the chatbot requests the intents data from the API , the chatbot then proceeds to check the highest probability that it got from the prediction if the probability is higher than 75% which is the threshold I used in my model it will send a response from the corresponding tag with the >=75% probability else it will send a predefined message asking the user to contact support threw email and saving the unidentified message in the database for the administrator to check.

2-2 Sequence Diagram: “Add Questions/Answers” Scenario

Figure II-6 describes the scenario

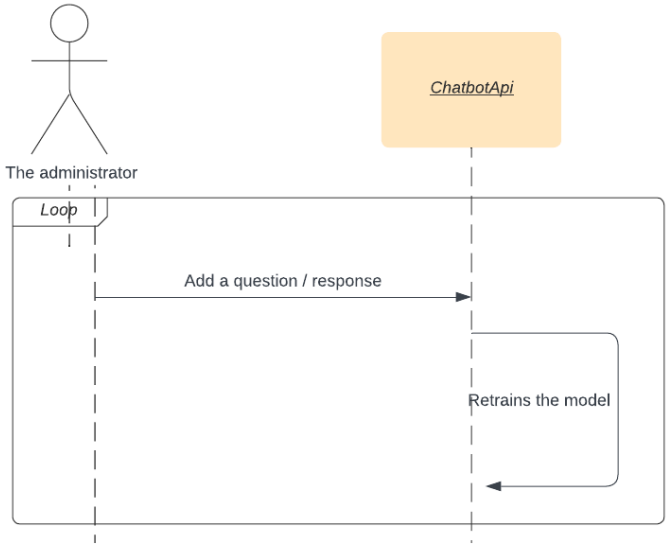

Figure II-6 : “add new question/answer” sequence diagram

When the administrator adds a question / response in the database the chatbot retrains itself.

2-3 Sequence Diagram: “Response Update” Scenario

Figure II-7 describes the scenario

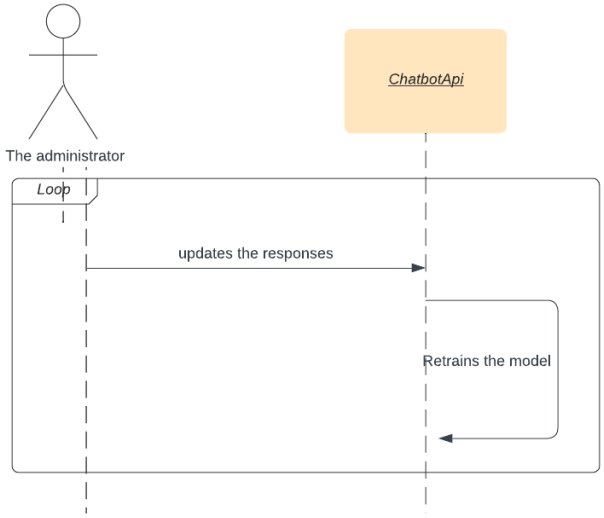

Figure II-7 : Sequence diagram “response update”

When the administrator updates the responses in the database the chatbot retrains itself.

Chapter III : Realization

Introduction
After having finished the design part, I just have to decide in which environments we will work in order to expose the technical choices used, the languages adopted and I illustrate at the end the realization based on some screenshots.

1 Working environment

1.1 Hardware environment

During the different phases of my project (specification of needs, design and development), I had used a laptop with the following characteristics:
Brand: Dell
Processor: i7 77hq
RAM: 16GB
Hard drive: 1TB

1.2 Software environment

Vs code:
Visual Studio Code is an extensible code editor developed by Microsoft for Windows, Linux and macOS. [4]

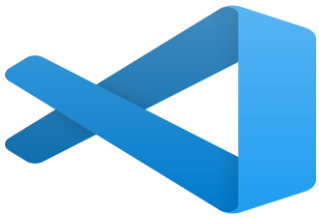

Figure III-1 : Vs code

Google Collab
Collaboratory, often shortened to "Collab", is a product of Google Research. Collab allows anyone to write and run Python code of their choice through the browser. It is an environment particularly suited to machine learning, data analysis and education.

Figure III-2 Logo google Collab

1.2.1 Front end technology

Html:

The **Hypertext Markup Language or HTML** is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript. [5]

Figure III-3 HTML logo

JavaScript:

often abbreviated JS, is a programming language that is one of the core technologies of the World Wide Web, alongside HTML and CSS. [6]

Figure III-4 JavaScript Logo

CSS:

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML or XML (including XML dialects such as SVG, MathML or XHTML). CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript. [7]

Figure III-5 CSS Logo

1.2.2 **Backend Technologies**

Django is a free and open-source, Python-based web framework that follows the model–template–views (MTV) architectural pattern. [8]

Figure III-6 : Django Logo

Flask

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions [9]

Figure III-7 flask Logo

1.2.3 Language used

Python

Python is a high-level, interpreted, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. [10]

Figure III-8 Python Logo

2 Choix Techniques

2.1- What algorithm is used?

There are 3 algorithms used in chat bots nlu - nlp - nlg l but although natural language processing (NLP), natural language understanding (NLU) and natural language generation (NLG) are all related topics, they are distinct.
NLP is an AI that plays with the language we speak, to get something well defined out of it. It can be as simple as identifying names from a sentence or as complex as finding out people's emotions towards a movie, by processing movie reviews. Simply put, a machine uses NLP models to read and understand the language spoken by a human.
In real life, NLP can be used to:

-Chatbot
-Summary of text
- Text categorization
-Marking up parts of speech
-Rooting
-Text mining. [11]

2.2 Which “Deep Learning Framework” is used?

For deep learning frameworks, I found that there are 3 that are most used are:
-**PyTorch** is a relatively new deep learning framework based on Torch. Developed by Facebook's AI Research Group and open sourced on GitHub in 2017.
-**TensorFlow** is an end-to-end open-source deep learning framework developed by Google and released in 2015.
-**Keras** is an efficient, high-level neural network application programming interface (API) written in Python.

	Keras	PyTorch	TensorFlow
Architecture	Simple, concise, readable	Complex, less readable	Not easy to use
Datasets	Smaller datasets	Large datasets, high performance	Large datasets, high performance
Popularity	Most popular	Third most popular	Second most popular
Speed	Slow, low performance	Fast, high-performance	Fast, high-performance

Tableau 5 keras vs tf vs pytorch [12]

2.3 Stemming vs Lemmatizing

Stemming vs lemmatization in Python is all about reducing the texts to their root forms. These techniques are used by chatbots and search engines to analyze the meaning behind the search queries. [13]

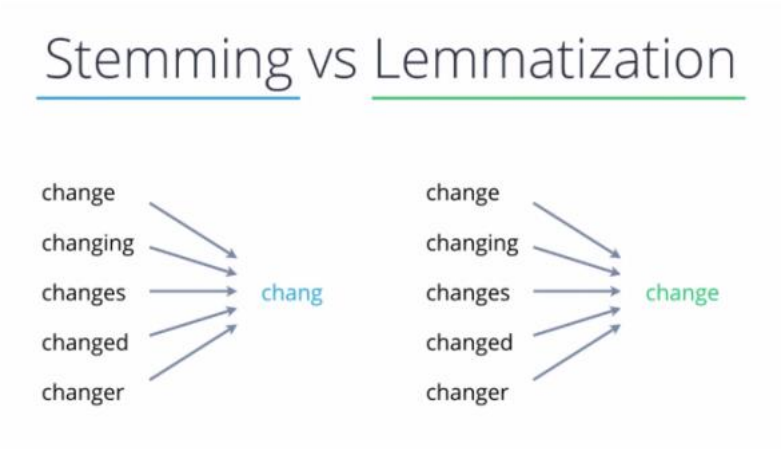

Figure III-9 Stemming vs Lemmatization [14]

Both stemming and lemmatization but stem can return a stem that isn’t an actual word whereas lemma is an actual language word but Stemming follows an algorithm with steps to perform on the words which makes it faster. It’s a difference in quality vs speed and I opted to use lemmatization as I had a bunch of fails from stemming.

3.1 General architecture

Before moving to interfaces created in this project it's crucial to understand how the whole system work, we are going to look at the interaction between Web interface and chatbot application in order to better understand the functioning of the system as a whole.
Let's start from the perspective of the user who have access to the application as seen in the Figure III-9 below.

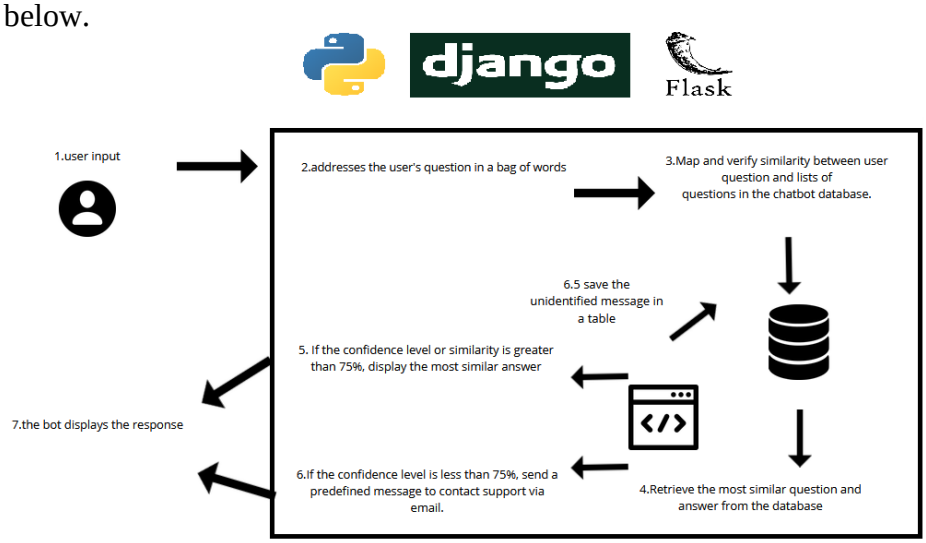

Figure III-10 General Architecture

First the user will open the chatbot-UI and send a message that message will be sent to the chatbot that will turn the message into a list of words and stem them then create a bag-of-word (a table of 1 and 0 for each word in the bag that exists in the sentence) it will load the neural network model saved from the training after that it will generate probabilities from the model using the bag of words created filter the prediction by a predefined threshold sort by probability if there is a match it will give a answer from that match else it will give a predefined response to contact support through email and save the unidentified message in the Unidentified table in database.

3.2 Front-End

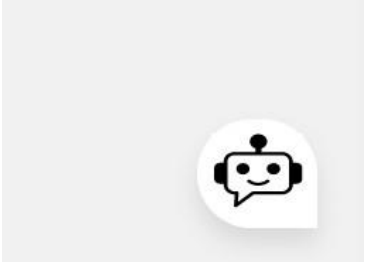

Figure III-11 chatbot UI closed

Figure III-10 represents the chatbot UI when the site is first opened the ui is closed

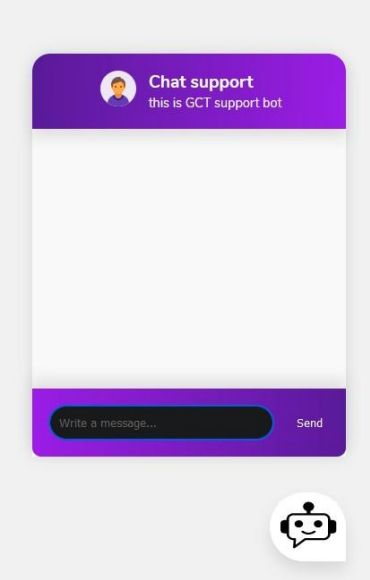

Figure III-12 Chatbot UI open

Figure III-11 represents the chatbot UI when its clicked open

3.2.1 Chatbot

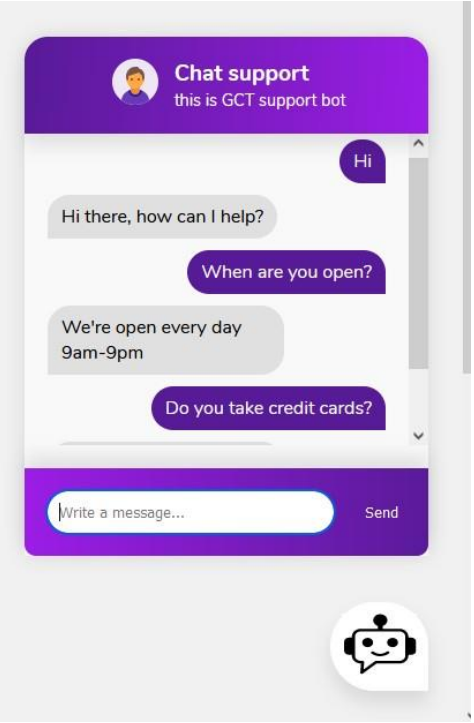

Figure III-13 Example of a chat

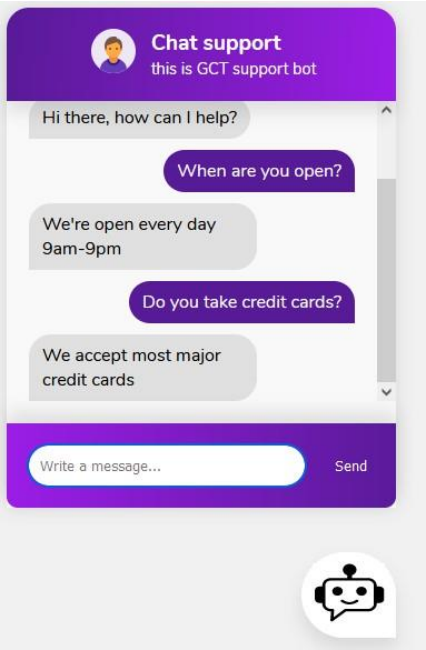

Figure III-14 Example of a chat 2

Le figure III-15 et figure III-16 represents an example of chat with the bot.

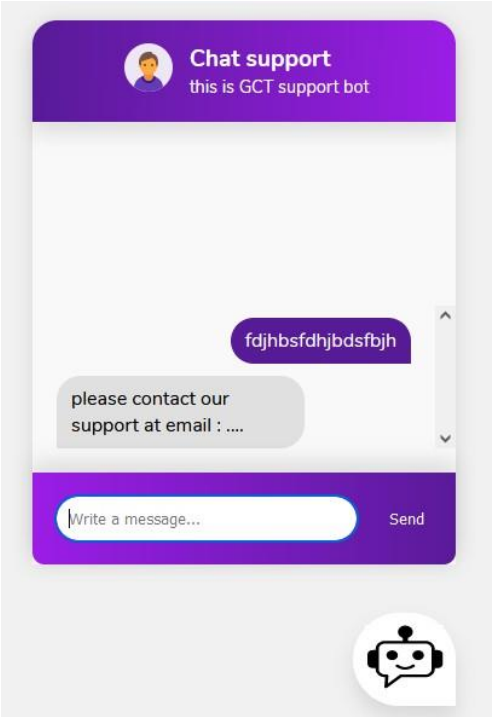

Figure III-15 Example of a message Unidentified

Le figure III-14 represents the case where the user enters a malformed message.

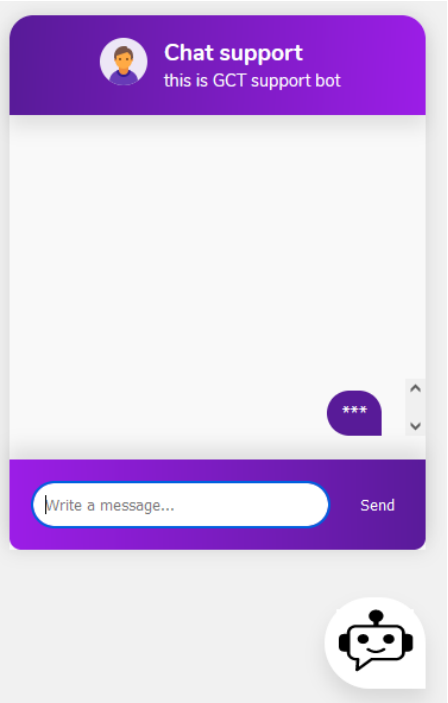

Figure III-16 Example of "bad word censor"

Le figure III-15 represents the case where the user enters a bad word and it gets censored

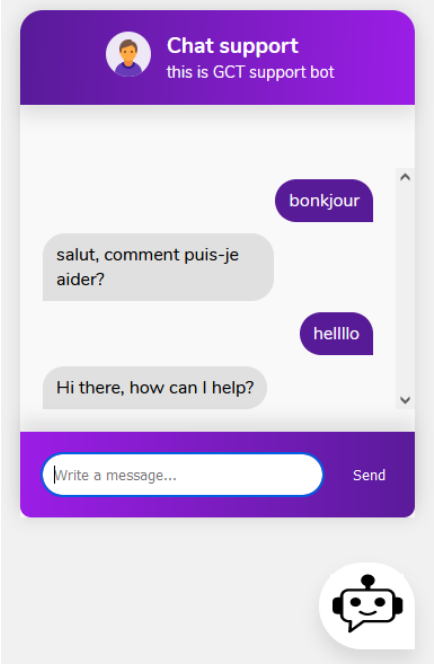

Figure III-17 le chatbot understands a mistype

Figure III-15 represents the case where the user enters a message with a typing error and the bot does a good job of detecting it.

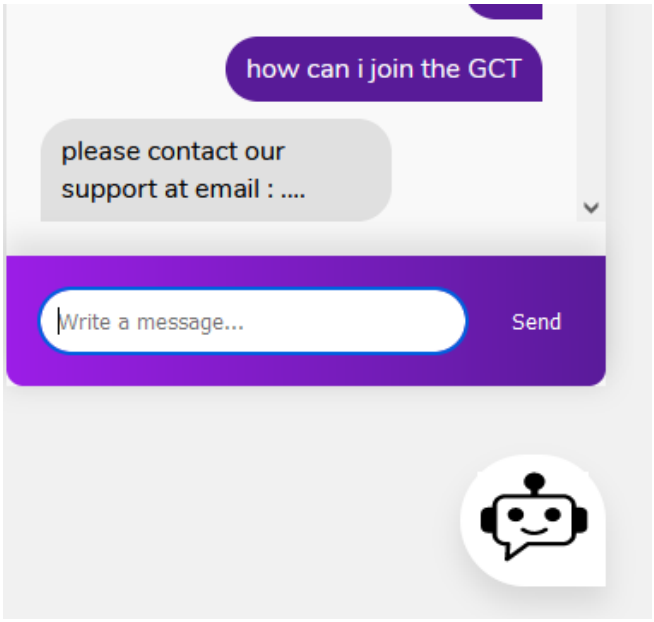

Figure III-18 a question not in database

Le figure III-17 represents the case where the user enters a message that is not in the bot intent base.

3.3 Backend

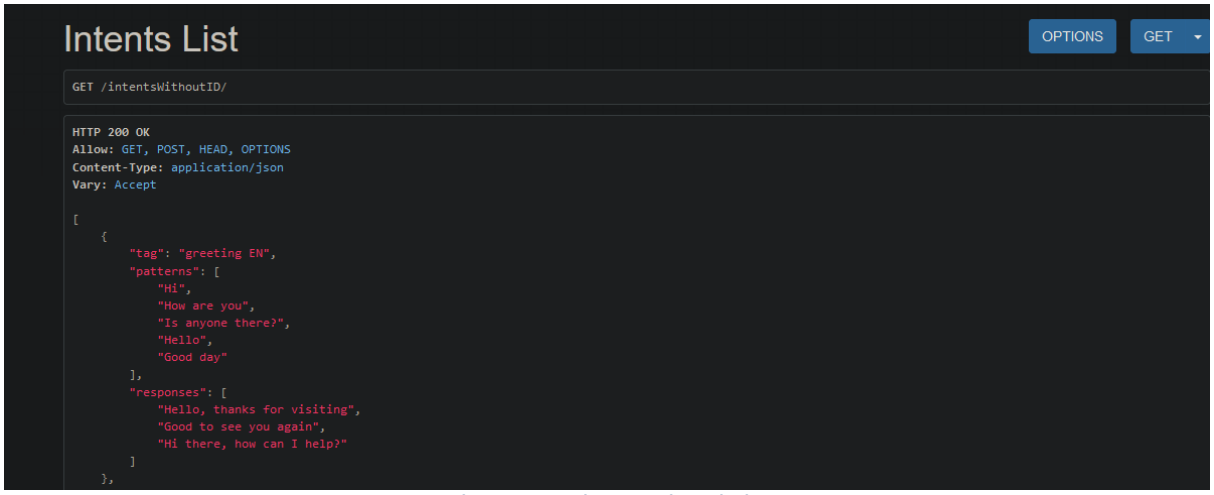

Figure III-19 intents dans le bot

Figure III-18 Represents the data of intents in the api currently

Figure III-20 add intent 1

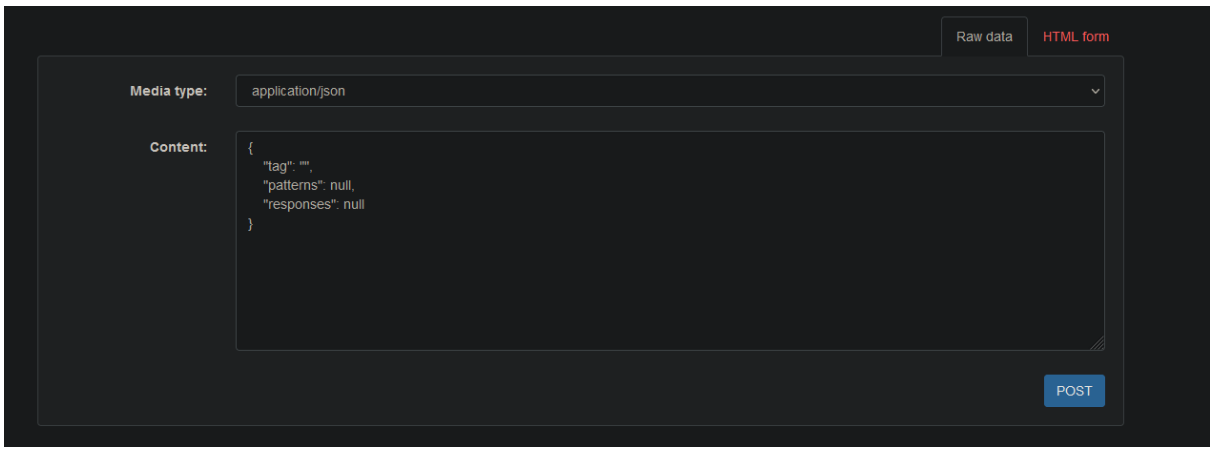

Figure III-21 add intent 2

Figure III-19 and Figure III-20 Represents how the admin can add data

```
    def create(self, request):
        serializer = self.get_serializer(data=request.data)
        serializer.is_valid(raise_exception=True)
        self.perform_create(serializer)
        headers = self.get_success_headers(serializer.data)
        chat()
        return Response(serializer.data, status=status.HTTP_201_CREATED, headers=headers)
```

Figure III-22 if post retrain model

Figure III-21 show that the retrain function is called when a post has happened

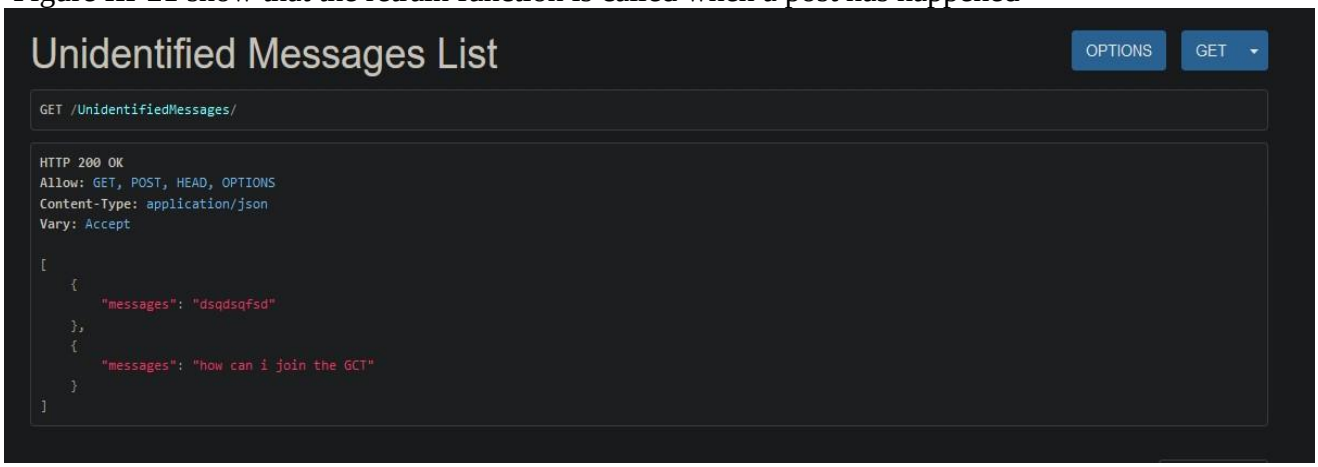

Figure III-23 Unidentified message is added

Figure III-22 show the Unidentified Message table where the unidentified messages are saved for the admin to choose from and maybe add in the intents database

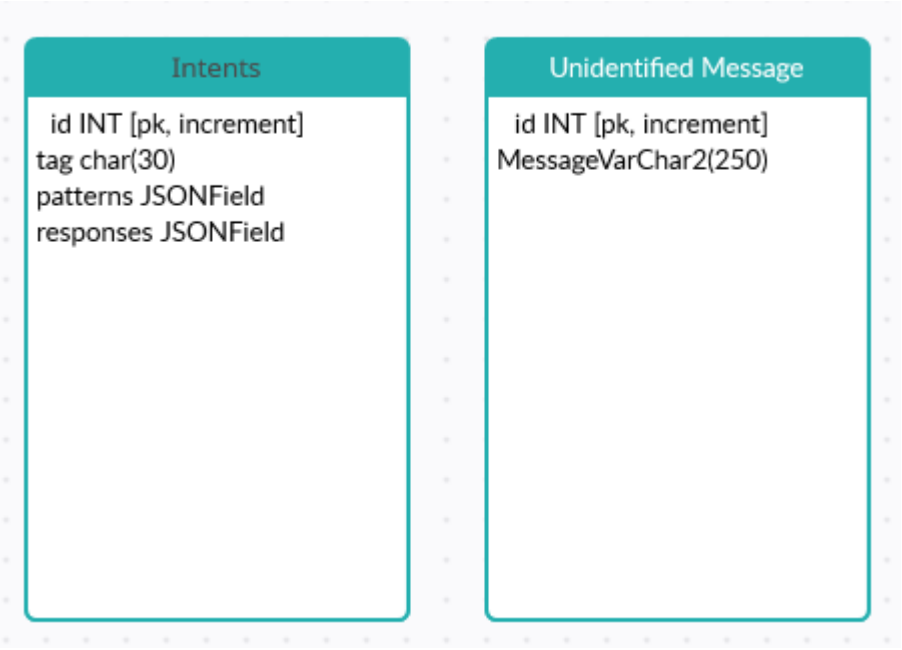

Figure III-24 Api database tables

Figure III-23 represents my Api database tables

Chapter IV Conclusion

Throughout this report, we have set out the work we have done in as part of our end-of-study project within the “GCT” the objective of this project is to is to create a reliable and engaging chatbot to ease communication between the “GCT” and public.

To do this, we began with a study of the general context of the project. Then, we carried out a study of the existing solution analyzing their limitation and comparing them to our solution. Subsequently, we implemented the needs of our project. Thus, this analysis served us in the design phase of our solution. Finally, we crowned our efforts with the implementation of our application.

The contribution of this project was considerable in terms of working methodology and technical skills. Indeed, this project allowed me on the one hand, to apply the theoretical knowledge acquired during the training of the LGLCI cycle and on the other hand to enrich my knowledge in terms of design and development thanks to the challenges that I faced, especially since this project is uses cutting edge technologies like NLP, Deep learning. It taught me project management and the use of different working tools.

From a project perspective, this project could get a little better work on the admin side UI and adding functionalities for user side such as voice to text system and making the bot response auditable.

Chapter V Bibliography

[1] "Groupe chimique tunisien wiki," [Online]. Available: https://fr.wikipedia.org/wiki/Groupe_chimique_tunisien.

[2] "Chatbot Fails," [Online]. Available: https://netomi.com/the-worst-chatbot-fails-and-how-to-avoid-them.

[3] "Unified Modeling Language," [Online]. Available: https://en.wikipedia.org/wiki/Unified_Modeling_Language.

[4] "Visual Studio Code," [Online]. Available: https://fr.wikipedia.org/wiki/Visual_Studio_Code.

[5] "HTML," [Online]. Available: https://en.wikipedia.org/wiki/HTML.

[6] "JavaScript," [Online]. Available: https://en.wikipedia.org/wiki/JavaScript.

[7] "CSS," [Online]. Available: https://en.wikipedia.org/wiki/CSS.

[8] "Django (web framework)," [Online]. Available: https://en.wikipedia.org/wiki/Django_(web_framework).

[9] "Flask (web framework)," [Online]. Available: https://en.wikipedia.org/wiki/Flask_(web_framework).

[10] "Python (programming language)," [Online]. Available: https://en.wikipedia.org/wiki/Python_(programming_language).

[11] "What are the Differences Between NLP, NLU, and NLG?," [Online]. Available: https://www.xenonstack.com/blog/difference-between-nlp-nlu-nlg.

[12] "keras vs tf vs pytorch," [Online]. Available: https://www.simplilearn.com/keras-vs-tensorflow-vs-pytorch-article.

[13] " Stemming and lemmatization," [Online]. Available: https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html.

[14] "stemming vs lemmatization in nlp," [Online]. Available: https://analyticsindiamag.com/explained-stemming-vs-lemmatization-in-nlp/.

[15] "UML (informatique)," [Online].

chatbot pour l'orientation des client du GCT.

Mouafak Abdel Majid el Zehri

Résumé: Dans ce projet nous nous intéressons à la conception et à la mise en place d'un chatbot d'intelligence artificielle. Le chatbot sera déployé sur le site GCT. Notre chatbot s'entraînera sur les entrées des utilisateurs pour fournir des réponses appropriées.

Mots clés: Chatbot, intelligence artificielle (AI), Traitement automatique des langues (NLP), Réseau de neurones (NN), Apprentissage automatique (ML), Apprentissage profond (DL).

Abstract: In this project we are interested in the conception and implementation of a artificial intelligence chatbot. The chatbot will be deployed on GCT site. Our chatbot will train itself on user inputs to provide appropriate responses.

Key-words: Chatbot, Artificial Intelligence (AI), Natural Language Processing (NLP), Neural Network (NN), Machine Learning (ML), Deep Learning (DL).