

# Utilizing Virtual Humans as Campus Virtual Receptionists

Moh. Zikky<sup>1\*</sup>, Marvel Natanael Suhardiman<sup>2</sup>, Kholid Fathoni<sup>3</sup>

<sup>1,2,3</sup>Department of Creative Multimedia Technology, Politeknik Elektronika Negeri Surabaya, Indonesia

<sup>1</sup>zikky@pens.ac.id\*; <sup>2</sup>marvelns.gt@student.pens.ac.id; <sup>3</sup>kholid@pens.ac.id

\*Corresponding author

## ABSTRACT

To imitate human-like behavior is one of the greatest feats a computer software could achieve. Computers can produce close-to-realism avatars with similar looks and behaviors in this modern era. One of the works that computer software could achieve now is conveying information in a place that a receptionist usually does. Therefore a computer software capable of that is called a Virtual Receptionist. This paper aims to explore the use of virtual humans as virtual receptionists and compare it to human receptionists to find both advantages and disadvantages. This research utilizes a virtual human model that imitates the behavior of a human receptionist. Its movements are based on real-life movements recorded with motion capture. It could also communicate with users by processing the voice input using speech-to-text technology recorded by a microphone. The recorded input will then be analyzed to determine whether it contains information stored in a database. The virtual human will then show the user the answer to their question accordingly. Utilizing virtual humans can make the process more interactive and exciting because of its futuristic feel. This way, campuses can have appealing introductory media and support campuses to be more open to the public in the future. However, the agent can only respond with prepared answers and not generate its own when necessary. Transcribed text will be analyzed for words that indicate the user's required information. In this case, the information would be the information of research laboratories in the post-graduate building of the *EEPIS* campus.

Keywords: Virtual Human, Virtual Receptionist, Open Campus.

This is an open-access article under the [CC-BY-SA](#) license.



## Article History

Received : Apr, 05<sup>th</sup> 2023

Accepted : May, 22<sup>nd</sup> 2023

Published : May, 31<sup>st</sup> 2023

## I. INTRODUCTION

In this modern era, the development of technology, especially computer software, is growing exponentially. With this rapid growth, we can see new feats a computer program achieves almost every day. From generating images, songs, and even websites, computer software is now even more capable of doing human activities. The use of technology in everyday life has a different feel, an atmosphere of futuristic and modernity in its implementations. Besides assisting, communicating with humans is one of technology's greatest goals. There are many approaches to this goal, such as using sensors and motion detection [1]. The use of virtual humans as virtual receptionists is one of the prime examples of when humans could communicate with a machine and another human being [2].

In addition to its futuristic feel, the virtual receptionist has advantages such as being more cost-effective and broader worktime flexibility. Unfortunately, there aren't many implementations of virtual humans in the field, as most virtual receptionists are still in chatbots and video calls. This is a missed opportunity because virtual humans can be considered a solution in this field. Also, besides all the advantages virtual humans have, the COVID-19 pandemic has created the need to distance physical interactions with other humans.

This paper will discuss using a virtual receptionist in the *EEPIS* campus post-graduate building. This paper will also reference similar works that could provide insight into the project. In the research conducted by Heru Ardiyanto [3], a virtual agent is added to an app that navigates rooms of the *EEPIS* undergraduate building. Both of these research use speech recognition technology to transcribe incoming audio and the output process to determine what information would be queried from the database. The difference is that in this research, the virtual human will act as a receptionist and give information to users based on information stored in the database. It will also give information on laboratories in the *EEPIS* campus post-graduate building. It consisted of the name, code, head of the laboratory, and a description of the research and practice done in the corresponding laboratory.

The other project used as a reference in this project is the BabyX [4] project. A virtual human could respond to its user's emotions and give an appropriate response. For example, if a user gives a happy gesture such as smiling, the model will display a happy animation such as smiling and waving. The model will display a scared animation if the user gives a negative gesture, such as anger or condemnation.

The next process of developing this application is integrating Google Cloud Platform (GCP) services such as speech-to-text, translation, and Firebase real-time database. Users will be prompted to ask questions regarding the lab, such as “Where is the interactive multimedia lab?”. The system would then translate the input from Indonesian to English. The transcribed audio will then be processed to find keywords used as query parameters to the database, which is “interactive multimedia,” in this case. After that, the system will query information about the interactive multimedia lab from the database. If the user gives invalid input or the information isn’t provided in the database, the system will display an error message and prompt the user to try using the recommended input again. Integrating the virtual human and the processing service is done using a game engine [5] in this case, Unity Game Engine.

This research aims to create a functional virtual receptionist that utilizes virtual humans. There are a lot of virtual receptionists that have a similar function but are mainly text-based applications [6] and conference calls. The usage of a virtual human was implemented so that engaging conversations with users may take place. Users are also prompted to use verbal interaction with the agent making it more interactive than communicating via text or calls [7]. Aside from interactivity, this research also intends to help solve problems in employing human receptionists in campus areas with unflexible work hours and memorizing limitations. Human receptionists can only be contacted within office hours, while virtual receptionists don’t have time limitations as they can work continuously. Virtual receptionists could also store more information correctly without worrying about forgetting or being mistaken.

## II. METHOD

The flow of Fig. 1 is the application flow grouped into several states. The first is the Start phase. During this phase, the model will display the idle Animation and stand by to receive input. If the input is transcribed as “Hello!” the application will change its state to Listening. In this phase, the model will play a listening animation, and then the system will record and process the upcoming input.

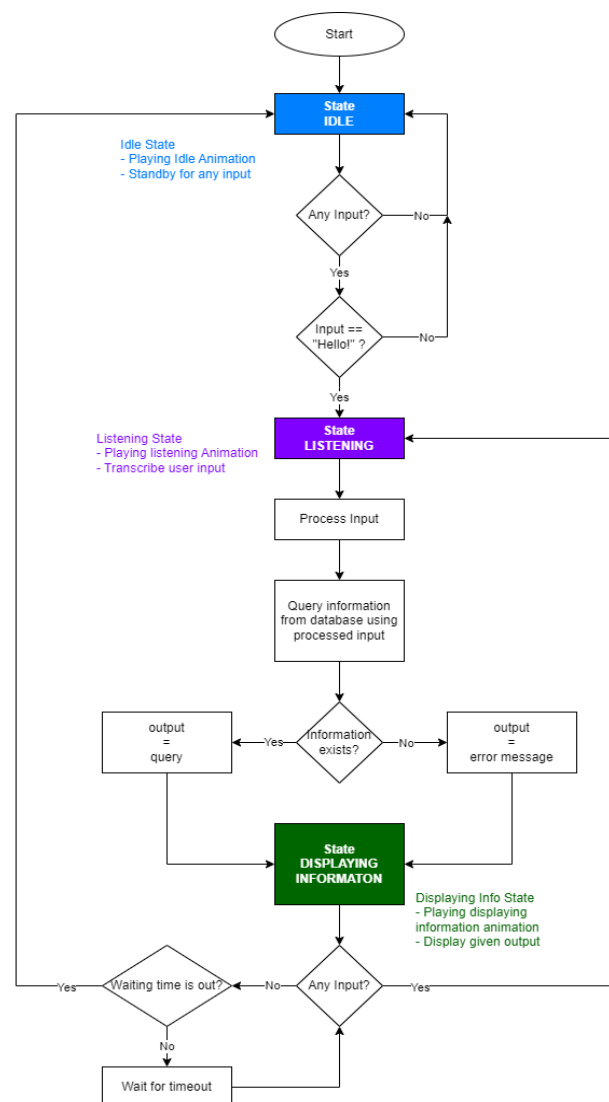


Fig. 1. Application Flow

The next phase would be Displaying Information state. During this phase, the system would display information based on the query result. If there is any result, the system will display the contained information. Otherwise, it will display an error message that prompts the user to try again. The model will also play the displaying information animation. To end this phase, the user needs to say, "Thank You!". After this state ends, the system will reset to its Start phase after a certain timeout. The following are the system designs used in this study, including (1) virtual human creation, (2) database creation, and (3) text processing systems. Each section will be described in sections A, B, and C.

#### A. Virtual human creation

This section will explain the process of creating the virtual human that will be used as the virtual receptionist. The first step is to create the 3D model. The 3D model was created in VRoid Studio with the image of a young lady wearing the *EEPIS* campus alma mater jacket. This process is done using Vroid Studio and Photoshop. Vroid Studio is used to create the base mesh for the model. Photoshop is used to edit the model's mesh texture to modify its appearance. After the model is created, the next step would be creating the model's animations using motion capture. Fig. 2 is one of the prime examples of a close-to-realism virtual human whom humans could communicate with.



Fig. 2. Recording Animation Using Motion Capture

The motion capture process intends to record live movement and translate it into 3D Animation [8]. This process involves multiple cameras surrounding the actor. It will then emit infrared lights to be reflected by the markers attached to a suit. Markers are probes that will reflect incoming light to be detected by the camera. The position of the marker will be used to translate the bones' position of a skeleton model in the application. Before recording, a specific set of skeleton bones would need to be selected. Then the actor has to put on the suit with markers placed on points where the camera would track.

The virtual human was developed using motion capture technology to imitate human-like behavior. Using this technology, the 3D model can have fluid movements close to human behavior ranging from simple movements to complex such as dancing [9]. The 3D model in Fig. 3 was created using software such as VRoid Studio, Blender, and Photoshop to create an acceptable model for use in the campus area. The model itself took the form of a young lady wearing an *EEPIS* alma mater suit. The animations recorded are idle, listening, giving information, and expressing gratitude.

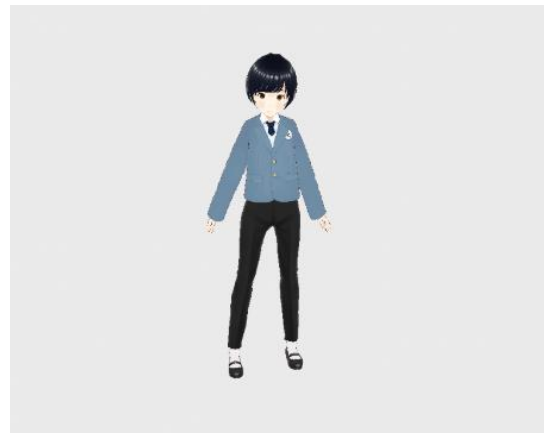
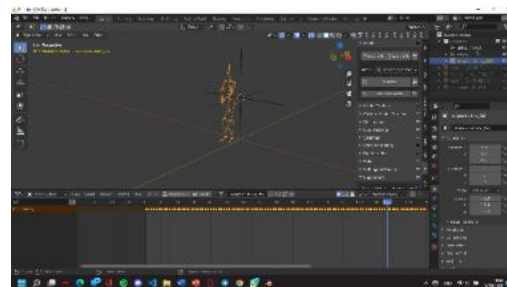


Fig. 3. Finished 3D model

After the suit is set, the first step is calibrating the cameras. This process uses a calibrator to ensure the cameras track the markers. The recording process in Fig. 4 (a) can start if the calibration is done. All motions in Fig. 4 (b) should start with doing a T-Pose at the center of the recording area. The T-Pose is used as a camera reference point [10]. If the movement is detected clearly, the skeleton's bones in the software in Fig. 4 (c) will move according to the position and rotation of the markers in the suit. The recorded animations for the virtual human in this research include idle, listening, conveying information, bowing to express gratitude, and head shaking to express dismissal.



(a) Motion capture recording process



(b) Inserting model animation in Blender



(c) Integrating the model into Unity Game Engine

Fig.4. The Virtual Human Creation

The recorded animations and the model will then be imported into Blender software to be inserted into the 3D model. This process was done by parenting the Animation to the 3D model mesh. The exported output would be a .fbx object that includes all the recorded animations. The final step would be integrating the model using Unity Game Engine to customize the behaviors based on the application state. An animator controller would control the states based on the designated parameters.

### B. Database creation

The database creation process for this application will be explained in Fig. 5. The database is hosted in Firebase real-time Database. The advantage of using an online hosted database is that the data can be modified without needing any changes to the application.

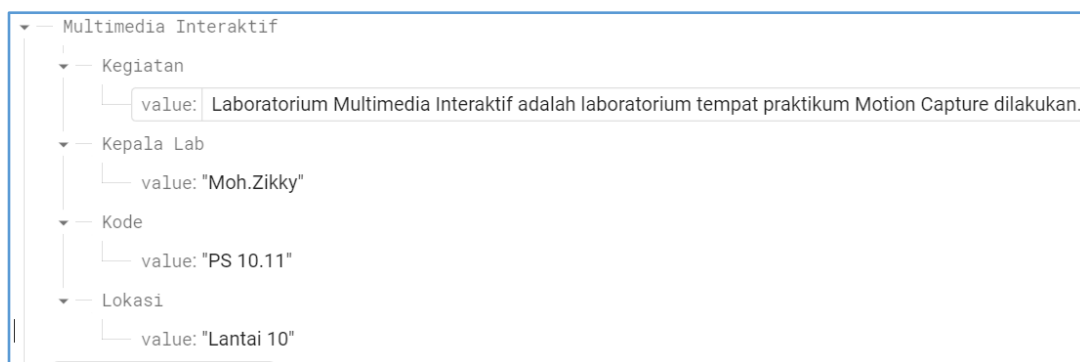


Fig. 5. Database Structure

The query parameter would be the laboratory name that acts as the parent directory. The stored information in the database is

stored in the child directory, such as the location, room number, name of the person in charge, and a short description of the lab activities.

### C. Text processing system

In order to acquire the query parameter, the transcribed text needs to be processed. This process will utilize Google Cloud Platform's Natural Language Processing (NLP) and translation service. After the audio in Bahasa is transcribed, the system will translate it into English. This ensures that the NLP process returns the best results, as it works best in English.

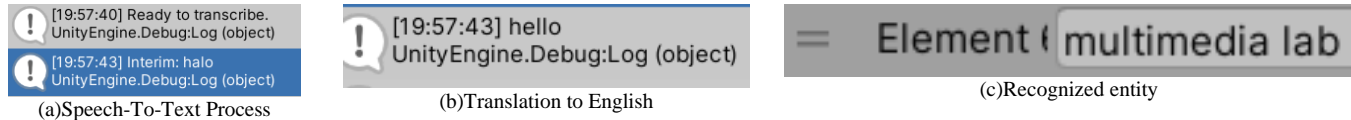


Fig. 6. Text Processing System Process

The NLP service, in particular, is entity recognition, which would categorize words into certain categories such as location, person, etc. [11]. The recognized word will then be matched with the database key values. If the two match, the system will return the selected word as the query parameter. If the result is below a designated threshold, then it will return failure. For example, the sentence "Please show me the creative multimedia laboratory" or "Could you show me the creative multimedia laboratory?" will return a neutral to a positive score. Still, sentences such as "I don't want to go to the multimedia laboratory" would return a negative score.

On the other hand, entity recognition would categorize words into certain categories such as location, person, etc. It will be used to detect a keyword in the transcribed text. For example, the sentence "Show me the creative multimedia laboratory" would result in "multimedia laboratory" as an entity.



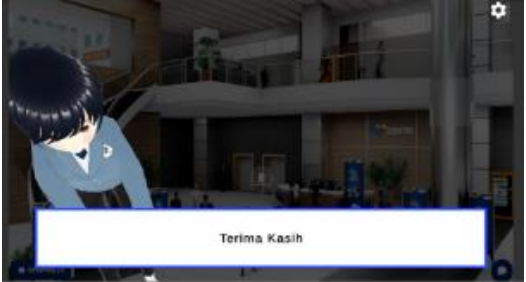
Suppose the sentiment analysis score is neutral to positive, and the entity recognition returns a result. In that case, the recognized word will then be matched with the database key values, which are the names of the laboratories. If the two match, the system will return the selected word as the query parameter. The query parameter would then select the information stored in the database, such as the laboratory name, room code, location, person in charge, and a description of the research and practice conducted there. Failure in the sentiment analysis or entity recognition would display an error message to the user that their questions weren't worded properly and prompt them to try again.

## III. RESULT AND DISCUSSION

The experiments and analyses in this research will explore the implementation of virtual humans as *EEPIS* Campus virtual receptionists. From a virtual human perspective, the five recorded animations were used in each application state, shown in the results in Table I.

TABLE I  
VIRTUAL HUMANS RECORDED ANIMATION

Virtual Human Animation	State
<p>(a)Idle Animation</p>	The starting phase of the application. In this state, the application will welcome the player
<p>(b)Listening to Animation</p>	If the application receives any input, the virtual human will play the listening animation and prompt the user to ask a question

Virtual Human Animation	State
 <p>(c)Dismissal Animation</p>	If the application cannot answer the question, it will play the head-shaking Animation, and it will apologize through the dialog box
 <p>(e)Displaying Animation</p>	If an answer exists, the application will play the displaying Animation and show the queried data from the database
 <p>(f)Thanking Animation</p>	After the session ends, the model will play a bowing animation and say Thank You! to the user

As the machine needs to understand the meaning of a question, transcribing the audio into text is not enough. The application needs to understand the question to determine what action to take next. Therefore, a process needs to be developed to reach the best outcome. In this research, the approach taken was to use entity recognition as part of the NLP framework. The application is expected to answer the user's question accordingly. The disadvantage of this method is that it can only understand common questions, and using complex sentences is prone to error.

Analysis [12] is used to identify negation in the statement [13] to dampen negative feelings [14]. It is considered unacceptable if the score is below the threshold of 0.0. The system also has all the laboratory names listed to be matched with the process results. It will also remove signs such as exclamation and question marks. Table II shows the sentiment analysis score of common questions, while Table III describes the detected entities in the sentences.

TABLE II  
SENTIMENT ANALYSIS OF USER INPUT COMMANDS

User input commands	Score	Compatibility
Please show me the creative multimedia lab	0.1	compatible
Please show me the creative multimedia lab	0.0	compatible
Show me the creative multimedia lab and the networking lab	0.2	compatible
I don't want to go to the creative multimedia lab, but the networking lab	-0.2	incompatible

Table II shows that punctuation marks may impact the sentiment analysis score. Hence the sentence must be clear to ensure a more accurate result. Negation in sentences also proved to impact the score negatively. Therefore the user must be prompted to avoid using it as a higher sentiment analysis score is preferred.

TABLE III  
DETECTED ENTITIES OF USER INPUT COMMANDS



User input commands	Entity	Compatibility
Please show me the creative multimedia lab!	"multimedia lab" (Other)	compatible
I don't want to go to the creative multimedia lab, but the networking lab	"networking lab" (Other)	compatible
Show me the creative multimedia lab and the networking lab	"networking lab" (Other)	compatible

As for the recognized entities described in Table III, the best case is to have the laboratory name marked. In both questions, entity recognition has successfully detected the desired entities, the laboratory names. But the problem is when there is more than one recognized entity, as the system can only process one request at a time. Determining the words that would be used as parameters was to test the names individually using entity analysis. Using the optimal condition where the transcribed text is "Show me the {laboratory name} lab", the result of the entity analysis process is described in Table IV.

TABLE IV  
LABORATORY NAME ENTITY ANALYSIS RESULT

Laboratory Name	Entity Analysis result
Aquaculture Laboratory	aquaculture
Biomedical Laboratory	biomedical
Electric Machine & Controls Laboratory	controls
Knowledge Engineering Laboratory	knowledge
Manufacturing Precision Technology Laboratory	manufacturing
Mobile & Wireless Communication Laboratory	mobile
Interactive Multimedia Laboratory	interactive
Computer Network & Web Engineering Laboratory	network
Renewable Energy Laboratory	renewable
Computer Architecture dan RTOS Laboratory	RTOS
Device & Sensor Technology Laboratory	device
Telecommunication Signal Processing Laboratory	telecommunication
Signal, Vision, & Graphics Laboratory	signal
Studio Broadcasting Laboratory	broadcasting
Digital Imaging Laboratory	imaging
Computer-Aided Learning Laboratory	learning
Workshop Laboratory	workshop
Human-Centric Multimedia Laboratory	human

Aside from the technical perspective, this research also compares whether the application has a more cost-effective production and development fee than a conventional receptionist. A comparison between the workload needs to be conducted. Table V describes the cost used to run the application, taken from the pricing page of the Google Cloud Platform [14]. The study by Mehl [15] shows that, on average, women and men spoke about 16,000 words daily. For a five day a week and four weeks month period, it can be estimated that an average person speaks about 300.000-350.000 words.

TABLE V  
APPLICATION MONTHLY RUNNING COST

Feature	Amount	Price
Entity Analysis	5K – 1M requests monthly	\$1.00
Sentiment Analysis	5K – 1M requests monthly	\$1.00
Translation	Per million characters monthly	\$20
Total		\$22

The salary range of a receptionist varies from IDR 1.800.000 to 4.000.000 or USD 150 – 275 a month. Compared to the monthly wage, the application offers a lower running cost of less than USD 25 a month, or approximately IDR 375.000 to IDR 400.000. This shows that utilizing a virtual receptionist has a lower starting cost than a conventional receptionist. Note that this comparison is approximate and may vary for each use case.

#### IV. CONCLUSION

To utilize virtual humans as virtual receptionists in campus areas, it would need to imitate human behavior to respond based on input. The first step is to transcribe the incoming input, which in this case, is implemented using speech-to-text. After receiving the input, the next step would be to respond appropriately. This research process was done through NLP processing and database query. Failure in one of these processes would result in a less desirable outcome. Therefore, the system must understand the input correctly for the desired outcome.

After conducting this research, the result is that virtual receptionists could be a consideration to be operated in campus areas in the future. They have the advantage of being more interactive, cost-effective, and have flexible work hours. Things that should be acknowledged when utilizing a virtual human as a virtual receptionist is that it should imitate real human behavior, such as movement and response to an action. Implementing the two elements correctly will make the virtual human closer to imitating an actual human. This research does not intend to replace conventional receptionist jobs entirely. The technology still needs much improvement and workaround to reach its full potential. In the future, 4.0 technologies, such as Artificial Intelligence (AI) and Machine Learning (ML), can make the application generate better responses. As for the virtual human, facial motion capture and high-definition rendering of the model could also be implemented to make the virtual human closer to reality.

#### REFERENCES

- [1] Zikky, M., Yuniar Hakkun, R., & Rafsanjani, B. (2019). Indonesian Sign Language API (OpenSIBI API) as The Gateway Services for Myo Armband. *International Journal of Artificial Intelligence & Robotics (IJAIR)*, 1(1), 16-25. doi: 10.25139/ijair.v1i1.2026.
- [2] Garrido, Piedad & Martinez, Francisco & Guetl, Christian. (2010). Adding Semantic Web Knowledge to Intelligent Personal Assistant Agents. *CEUR Workshop Proceedings*. 687.
- [3] Zikky, Mohammad & Basuki, Achmad & Ardiyanto, Mohamad Heru, INTERACTIVE AGENT TOUR GUIDE IN EEPIS VIRTUAL CAMPUS TOUR WITH VOICE COMMAND (2018)
- [4] Sagar, Mark & Seymour, Mike & Henderson, Annette. (2016). Creating connection with autonomous facial Animation. *Communications of the ACM*. 59. 82-91. 10.1145/2950041
- [5] Gadia, Davide & Celata, Tommaso & Notarangelo, Antonio & Ripamonti, Laura & Maggiorini, Dario. (2018). G.E.M.I.X.: Game Engine Movie Interaction eXperience.
- [6] Srinivasa, Arul, Madheswari, A. Neela, The Role of Smart Personal Assistant for improving personal Healthcare, *International Journal of Advanced Engineering, Management and Science (IJAEMS)* [Vol-4, Issue-11, Nov-2018], <https://dx.doi.org/10.22161/ijaems.4.11.5>
- [7] Unismuh, Junaid. (2020). The Verbal Interaction Between Lecturers and Students in Classroom. 10.13140/RG.2.2.22728.34567.
- [8] Nogueira, P. (2011). Motion Capture Fundamentals: A Critical and Comparative Analysis on Real-World Applications. Faculdade de Engenharia da Universidade do Porto. Programa Doutoral em Engenharia Informática. Instituto de Telecomunicações. Retrieved from [https://paginas.fe.up.pt/~prodei/ds12/papers/paper\\_7.pdf](https://paginas.fe.up.pt/~prodei/ds12/papers/paper_7.pdf)
- [9] Nurindiyani, Artiarini & Pramadihanto, Dadet & Afifah, Rosyidina. (2019). Motion Modeling of Traditional Javanese Dance: Introduction of Javanese Dancer Gesture with 3D Models. 195-201. 10.1109/ELECSYM.2019.8901523.
- [10] Wu, Qingqiang & Xu, Guanghua & Li, Min & Longting, Chen & Zhang, Xin & Xie, Jun. (2018). Human Pose Estimation Method based on Single Depth Image. *IET Computer Vision*. 12. 10.1049/iet-cvi.2017.0536.
- [11] Naseer, Salman & Ghafoor, Muhammad & Sohaib, & Khalid Alvi, Sohaib & Kiran, Anam & Rehman, Shafique Ur & Murtaza, Ghulam & Campus, Jehlum & Jehlum, Pakistan. (2022). Named Entity Recognition (NER) in NLP Techniques, Tools Accuracy and Performance.
- [12] Aqlan, Ameen & Bairam, Dr. Manjula & Naik, R Lakshman. (2019). A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. 10.1007/978-981-13-6459-4
- [13] Mahany, Ahmed, Heba Khaled, Nouh Sabri Elmitwally, Naif Aljohani, and Said Ghoniemy. 2022. "Negation and Speculation in NLP: A Survey, Corpora, Methods, and Applications" *Applied Sciences* 12, no. 10: 5209. <https://doi.org/10.3390/app12105209>
- [14] [https://cloud.google.com/natural-language/pricing#pricing\\_units](https://cloud.google.com/natural-language/pricing#pricing_units) (Accessed March 2023)
- [15] Mehl, Matthias & Vazire, Simine & Ramirez-Esparza, Nairán & Slatcher, Richard & Pennebaker, James. (2007). Are Women Really More Talkative Than Men?. *Science* (New York, N.Y.). 317. 82. 10.1126/science.1139940.