Legal Advisor Chatbot for Cyber Crime in India

Darshana Borse, Prof. Khalid Alfatmi, Neha Saindane, kalyani Deore, Rohan Baviskar
Department of Computer Engineering, Shri Vile Parle Kelavani Mandal's Institute Of Technology, Dhule, India
darshanaborse687@gmail.com, alfatmi.khalid@gmail.com, nehasaindane.16@gmail.com, kalyanivd08@gmail.com, rohanbaviskar545@gmail.com

Abstract—**Individuals and corporations are becoming increasingly reliant on technology. Nowadays, using technology for our everyday activities is extremely frequent. For example, we use a personal computer or a smart phone to pay our life insurance premiums, power bills, flight or train booking, bus tickets, order books, or any Cybercrime encompasses a wide range of attacks, including cyber extortion, cyber warfare, identity theft, social media crimes, computer virus or malware distribution, Internet fraud, phishing, carding, spamming, child pornography, and intellectual property rights violations, among others. Because of the rising number of cyber-attacks these days, internet users must be aware of these threats and exercise caution when engaging in online activities. The common people don't know about the cybercrime legal process, applicable law articles, and prison terms, as well as many people, don't have access to legal services.Also, To obtain information about cybercrime, lawyers must go through a significant number of text documents. As a general person, official vocabulary is difficult for a common person who does not have the legal knowledge to understand the problem or rules. To overcome all these problems and the legal process should be easy to use for common people, law students or lawyers we have proposed a chatbot that is a Legal Adviser for cybercrime. The goal is to create an intelligent chatbot that can assist law students and efficiently handle client cases. The chatbot should make client engagement simple and quick, ensuring that clients are satisfied and that they receive legal advice quickly. It will suggest legal procedures, applicable legal articles, and prison conditions and lawyers if necessary.**

Keywords— Cyber Crime Laws, Chatbot, Natural Language Processing(NLP), Artificial Intelligence(AI), text data processing, Recurrent Neural Networks(RNN), Machine Learning(ML), Legal Advisor

I. INTRODUCTION

A. Overview of Cyber Crimes

India saw a major consequence in reported cybercrime in 2020 compared to the previous year. More than 50,035 cybercrime incidents were recorded. Nowadays, Cybercrime has increased day by day as new technologies and digitization take place. Research showed that only 16% of victims asked criminals to stop the attack. Since the pandemic, many people's daily activities take place at home. entertainment. At the same time, cybercrime has increased worldwide.People hook up with the web for work, school, shopping,etc.

cybercrime is a crime committed using a digital computer for purpose of violating privacy, stealing identities, intellectual property committing frauds, as computers and the internet are earlier introduced to America, Americans were the people got targeted by this kind of crime but in the 21st century, there are hardly any places where computer and internet haven't reached and so nor cybercrime. As new technologies are being introduced they are not just providing opportunities but also it is causing new types of crimes. This crime is not a traditional crime because for carrying out cybercrime you don't have to be physically present at the place of crime, attackers can commit cybercrime from anywhere in the world by using the virtual identity.

People are unaware of cybercrime and IPCs alike, their negligence in security or small steps of the attacker can bring you big trouble, you are opening the aisle for cybercriminals, therefore, it is important to keep track of what is happening to you.

There are following main types of cybercrimes:

1. Phishing scams- These are the most common types of scams where attackers trick unsuspected users into giving their personal information by creating profiles and websites of popular brands which look very identical to original ones.

2. Internet fraud- In this type of fraud attackers communicate with the person who needs money by offering a substantial sum of money must be returned or they asked for a small amount of money promising to return a substantial sum of money and when a person falls for this trap they will simply say process had some complication and will require more money and this will continue till attacker think that there is nothing left to gain from the victim and will go to another victim. In a similar scam, they will offer a job but for that, some amount will be asked to pay for the paperwork

3. Identity theft: In this crime, there are two types of indent theft one is impersonation .in which no card details and payment methods were stolen but personal details are leaked this is less dangerous than when the card details are stolen with the personal details and attackers can use it for which you will have to pay the charges or they can transfer money from your account

4. Online harassment and cyberstalking: cyber-harassment or cyberbullying is committed through social media in the forms of posts, comments, and direct messages in some cases I can also be through emails. Containing threats or defamatory messages to individual or group

In Cyberstalking, stalkers try to gain more personal information about victims from the internet and leak the personal information on online toxic communities with malicious intent. Cyber Crimes affect victims in various ways mainly it affects the mental health of the victim. There are cases where victims had a mental illness that further led to suicide.

B. Overview of Chabot

The chatbot is an Artificial Intelligence algorithm that processes natural language which analysis in order to provide an intelligent response. It is a computer program that

resembles human natural language through voice or text chat or both. Users talk or chat with chatbots like a real person and chatbot answers instantly like a real person via chat or text message. Rule-based chatbots, Rule-based chatbots, intellectually independent chatbots, and AI-powered chatbots are the three basic types of chatbots.

The main motive of chatbot is to make a structure of communication take after a natural conversation between humans. The first chatbot invented by MIT professor Joseph Weizenbaum in the 1960s is named ELIZA. In 2009, the WeChat company made an advanced chatbot in China.

Artificial Intelligence is a computer field where that focus on intelligent machines which work and think like a human. Artificial Intelligence has resembled human Intelligence

The following are the various sorts of chatbots:

1. Chatbots with a menu or a button: Button-based chatbots are the most prevalent sort of chatbot now in use. In the majority of circumstances, these chatbots are canonical call tree hierarchies supplied upon the user within the appearance of buttons. Similar to the machine-controlled phone menus we all use on a regular basis, these chatbots require the user to make multiple options in order to go deeper into the final word answer.[9]

2. Rule-Based Chatbots: If you can predict the sorts of inquiries your clients could have, you'll be ahead of the game, a linguistic chatbot could be the answer. If/then logic is mistreated by rules-based chatbots, resulting in informal flows. You must first define the language conditions for your chatbots. Conditions are frequently constructed to analyze the wording, their sequence, synonyms, and other characteristics. Your consumers will obtain the appropriate assistance in no time if the incoming query meets the criteria that your chatbot has specified.[9]

3. Chatbots that use keyword recognition: Unlike menu-based chatbots, keyword recognition-based chatbots will listen and respond properly to what users sort. These chatbots use configurable keywords in conjunction with an AI application - tongue method to figure out how to provide a relevant response to the user. When these chatbots are asked to respond to a large number of similar questions, they fall short. When there are keyword redundancies amongst several connected requests, human language technology chatbots can start to stutter. It's fashionable to find chatbots that combine keyword recognition and menu/button-based interaction. These chatbots allow users the choice of raising their inquiries directly or utilizing the chatbot's menu buttons if the keyword recognition technology isn't working or if the user wishes to utilize the chatbot's menu buttons.[9]

4. Machine Learning chatbots: Have you ever wondered what a dialogue chatbot is? A conversation chatbot is far more advanced than the three bots discussed previously. These chatbots employ machine learning (ML) and artificial intelligence (AI) to remember discussions with specific users so that they may be told and grow over time. Chatbots with discourse awareness, unlike keyword recognition-based bots, are smart enough to improve themselves based on what users want and how they want.[9]

5. The hybrid model: Businesses adore the sophistication of AI-chatbots, but they don't always have the skills or vast amounts of data to support them. That's why they choose hybrid models. The hybrid paradigm combines the best of both worlds with the simplicity of a rule-based chatbot and the accuracy of an AI-bot.[9]

6. Voice bots: To make casual interfaces even more conversational, businesses are increasingly resorting to speech-based chatbots or voice bots. Voice bots have become increasingly popular in recent years, with virtual assistants ranging from Apple's Siri to Amazon's Alexa, and the reason for this is the convenience they provide. It is significantly more practical. for a client to converse rather than sort. The tip customer receives resistance experiences from a voice-activated chatbot.[9]

C. Why Chabot is needed

Youngsters these days' use messengers for communication with friends, relatives and for connecting with brands, chat features are very popular among youngsters and brands for advertisement. chatbbt leads users to specific repetitive tasks and users do not have to read everything and search what to do, whereas the chatbot will give you exact suggestions or answers. The chatbot does not get bored or tired. chatbots are very time saving and efficient. People hesitate to talk about the legal issue with a lawyer, or to ask for a solution, or even simply to ask the lawyer for some information about the legal issue they are charging too much.[10]

Facebook and Hub spot released the below research:

- Every month, 2 billion messages will be sent between people and businesses.
- Messages are preferred by 56 percent of respondents over phone calls to customer support.
- 53 percent of respondents prefer to shop with companies that they can communicate with via text message.
- For customer assistance or help, 71% of respondents prefer to use messaging apps.[10]

II. LITERATURE SURVEY

A literature study was conducted to acquire an understanding of the topic, as well as related studies that are relevant to this research, are included in this chapter. This section covers topics such as human-machine interaction, chat interfaces, chatbots and chatbots, the legal profession, the evolution of chat interfaces, and the current condition

of chatbots in the legal area, among others.This aims to show how conversational interfaces can be used, as well as a few systems built around them, as well as apps in the commercial and legal realms.

An Approach to Obtaining Legal Assistance is the title of the paper. It provides a Virtual Legal Assistant (VLA) concept based on Artificial Intelligence, which is a four-component based architecture, with the four components being Knowledge Base, Question Generation Engine, Bot backed by Dialect Convertor, and Text Analytics, which can enable situations with an interactive and Artificial Intelligence (AI) based virtual assistant. Examining a set of facts and advising a client on a specific course of action based on existing legislation is typical of legal counsel. The proposed strategy considers the possibility of creating a similar digital assistant for the legal field, which could be used to train on millions of court cases and provide legal support to those who require it.[1]

They present the "Robot Lawyer" intellectual system in this work. This paper's major objective is to aid lawyers and the public in giving important information on legal processes. "Robot Lawyer" has an expert system that uses a set of rules to provide benchmark information and a neural network model to meet more complex needs.. The methods for processing textual data are also described in the article. The system was created with the Russian language in mind. [2]

This material has been thoroughly examined. Although the majority of current research focuses on a particular aspect of the LJP job, a true legal judgement prediction (LJP) system should give a judge with plausible judgement suggestions, including charges, applicable law articles, and jail term. They propose MANN, a multichannel attentive neural network model that uses a single framework to learn from the past, analyse documents, and complete the integrated LJP job, which is inspired by deep neural networks' remarkable success in a variety of applications. Using a textual account of a criminal case as input, MANN creates representations of latent entities appropriate to case facts, the defendant's persona, and relevant law documents using attention-based neural networks.Furthermore, they employ a two-tier framework that allows attentive sequence encoders of the case description to characterise semantic interactions from diverse sections in a hierarchical way at both the word and sentence levels. Four real-world datasets of criminal cases from mainland China were used in the study. The experimental data reveal that MANN achieves state-of-the-art LJP performance on all assessment measures.[3]

In this study, LAW-U is an artificial intelligence (AI) chatbot that gives legal assistance to sexual abuse survivors by proposing Supreme Court decisions that are most appropriate for their circumstance. The topic is chatbot LAWU Legal Guidance for Sexual Violence Victims and Survivors using Artificial Intelligence. [4]

A. Limitations in Existing systems:

- The existing system only provides the static information they don’t suggest lawyers to the user.
- One of the systems only available in Russian language.
- They directly predict the outcome of the case.

For India's digital landscape, the "Digital India" vision and the 2016 demonetization were game-changers. Increased cybercrime and incidences of online cheating have stemmed from the increase in internet usage over the last decade, as well as the pandemic. Within a fortnight, tens of thousands of Indians had established an online presence. Additionally, with the availability of low-cost mobile phones and high-speed data plans, an increasing number of Indians have embraced the digital revolution. The Information Technology Act of 2000, which was later updated to the IT (Amendment) Act of 2008, is Indian legislation that deals with penalties related to such offences.

Cybercrimes, according to the Information Technology Act, are subject to global jurisdiction, which implies that a complaint can be filed in any cyber cell in India, regardless of where the crime occurred or where the victim is now located. Many cyber cells have been established across India by criminal investigation teams with the primary objective of dealing with reports and investigations of offences. Currently, the bulk of Indian cities have their own cybercrime unit. The victim can file a complaint with the cyber police or the crime investigation department at any moment, either online or offline. A victim must submit a written complaint to any of the cybercrime cells to make a complaint. The victim's name, contact information, and address must be included in that written record. The Head of the cybercrime cell of the city has to be addressed to in the written complaint.

Following chart shows the rate of the cyber crime in India : [9]

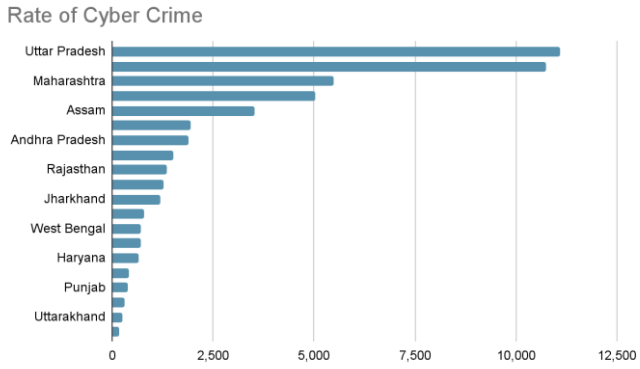

Fig 1. Rate of Cyber Crime in India

Many cybercrime cells have been established by the criminal investigation team in various places across India, to handle reports and investigations of cybercrime. Currently, most Indian cities have a dedicated cybercrime unit. You can file a complaint with the cyber police or the criminal investigation department at any time, either in person or online. The first and most important step in punishing cybercrime is to file a complaint against the offence. You must register a written complaint with any jurisdiction's cybercrime cell. You must include your name, contact information, and mailing address in the written complaint. The written complaint should be addressed to the Head of the Cybercrime Cell in the city where the cybercrime complaint

is being filed. A cybercrime falls under the purview of worldwide jurisdiction under the IT Act, which implies that a cybercrime complaint can be filed with any of India's cyber cells, regardless of where the crime was perpetrated or where the victim is now residing/staying.

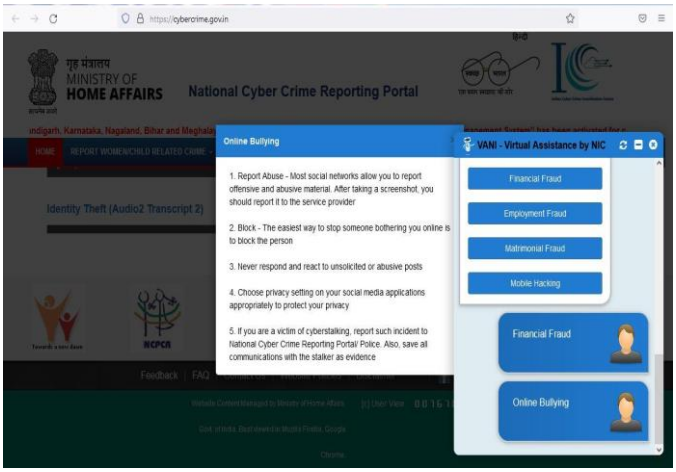

Fig 2. Chatbot on Government Website

The Government of India has the official website for Cybercrime. On that site we can report the cybercrime. The site contains one chatbot which is CyberDost. As we can see in the fig 1 the image of the CyberDost. CyberDost has only given the safety tips for the people and registered the complaint. CyberDost does not give the proper information about the legal procedures, applicable legal articles, and prison conditions and lawyers.[8]

III. METHODOLOGY

Here we propose a system using Machine Learning (ML) and Natural Language Processing (NLP) which is to inform and guide people about cyber laws and legal procedures. In this system, user input their problem and our chatbot extracts the keywords from the system and calculates the suitable response for the user like either the information about the laws or the legal procedure.

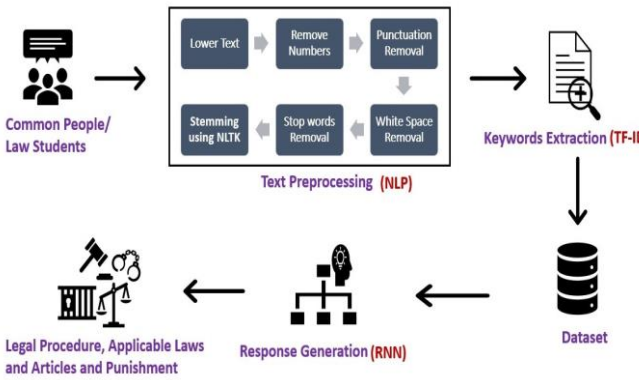

Fig 3. Architecture

Firstly, the user input their problem and the following are the steps to process that input:

A. Text Cleaning/processing:

Text cleaning or text preprocessing is required when working with text in Natural Language Processing (NLP). Because real-life human written text data comprises a variety of words with a wrong spelling, short words, special symbols, and so on, we must clean this type of noisy text data before feeding it to the machine learning model.

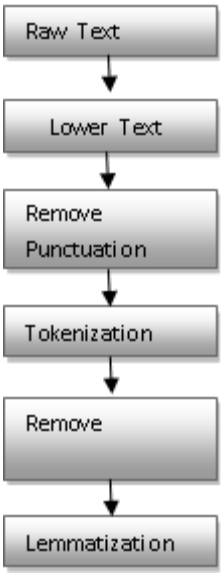

Fig 4. Data Pre-processing

- *Lower text*: The main purpose for changing the text to all lowercase is to prevent keywords like "Hello" and "hello" from being confused. It limits the number of words that can be stored in a dictionary at any given time.[5]
- *Removal of punctuation*: Punctuation and non-alphanumeric terms were removed from the original text. Steps in-text pre-processing that will assist the model in improving its performance. [5]

- *Tokenization*: Word tokenization is the process of splitting the text into words is called the token. Tokenization is an important part of the field of Natural Language Processing. [5]

NLTK provides two sub-modules for tokenization:

- word tokenizer
- sentence tokenizer

- *Stop word removal:* Stop words are terms like a, an, the, you, your, and others that are frequently used in written writings. Stop Words are frequently found in text documents. However, they are often ineffective for text analysis, thus it is preferable to eliminate them from the text. If stop words are removed, we may concentrate on the vital words. A list of stop words is provided by the NLTK package. Stop words are terms like a, an, the, you, your, and others that are frequently used in written writings. Stop Words are frequently found in text documents. However, they are often ineffective for text analysis, thus it is preferable to eliminate them from the text.

If stop words are removed, we may concentrate on the vital words. A list of stop words is provided by the NLTK package. [5]

• Lemmatise: Because Lemmatization is closely similar to Stemming, we'll utilize it here. The primary distinction is that Lemmatization considers the morphological analysis of words and converts them in a meaningful way, like a Legal Adviser, and can be employed in cases of cybercrime.[5]

B. Keywords Extraction and text summarization

It's a text-analysis method. Within a short period of time, we can gain valuable insights into the subject. It aids in the condensing of the content and the identification of significant keywords. It saves time by not having to read the entire document. We'll use the TF-IDF method in this case. [7]

Term Frequency - The number of times a term appears in a text. The total number of words in the document is divided by the number of times the term t appears in the text to arrive at this figure. [7]

Inverse Document Frequency - What exactly does a word in the document mean? (total number of sentences/total number of sentences with term t) log (total number of sentences/total number of sentences with term t) (total number of sentences divided by the total amount of words) TF-IDF — The TF-IDF Foundation is a non-profit organization This score is used to determine whether or not words are relevant. It's calculated using the formula TF * IDF. [7]

C. Response Generation:

RNN is generally used in time series analysis and other scenarios where a sequence of data must be processed. The network learns from what it has just observed in this type of activity, which is known as short-term memory. It can more accurately forecast future data using this memory. The length of time that information on previous data is retained is not set in stone, rather it is determined by the weights assigned to it. [6]

So, let's see the 4 layers of the RNN model:

1. *Embedding Layer*: The number of unique words in the dataset is the initial argument to the embedding layer. This argument is defined as large enough to allow for the unique encoding of every word in the corpus. The word size has been set to 20000 in this project. The number of embedding vectors is displayed in the second argument. The size of the embedding will be used to display each word in the corpus.
2. *LSTM Layer*: A Recurrent Neural Network's most significant notion is this. The LSTM (Long Short Term Memory) layer overcomes the Vanishing Gradient problem, giving the model the ability to predict the next word based on recent past memory.
3. *Activation Function* : The activation function "Hyperbolic tangent(tanh(x))" is utilised in the RNN model because it retains the value between -1 and 1. Backpropagation adjusts the weights at each node by multiplying them by gradients. If the gradient value is higher, the weight value for that node will grow dramatically. As a result, the weights of other nodes will be reduced to a minimum and will not be included in the learning process. As a result, the model will have a large bias. The tanh(z) hyperbolic function is employed to avoid this. It maintains a consistent distribution among the network weights by bringing the values between -1 and 1.
4. *Output Layer* :The "Sigmoid" activation function is employed in the output layer.It decreases the value in the same way that "Hyperbolic Tangent" does, but it does so between 0 and 1. The idea behind this is that if you multiply a value by 0, it becomes zero and can be discarded. When a value is multiplied by one, it remains zero and is just present here. As a result, when utilising the sigmoid function, only the most relevant and important values are considered in forecasts.

Demo of the Legal Advisor ChatBot :

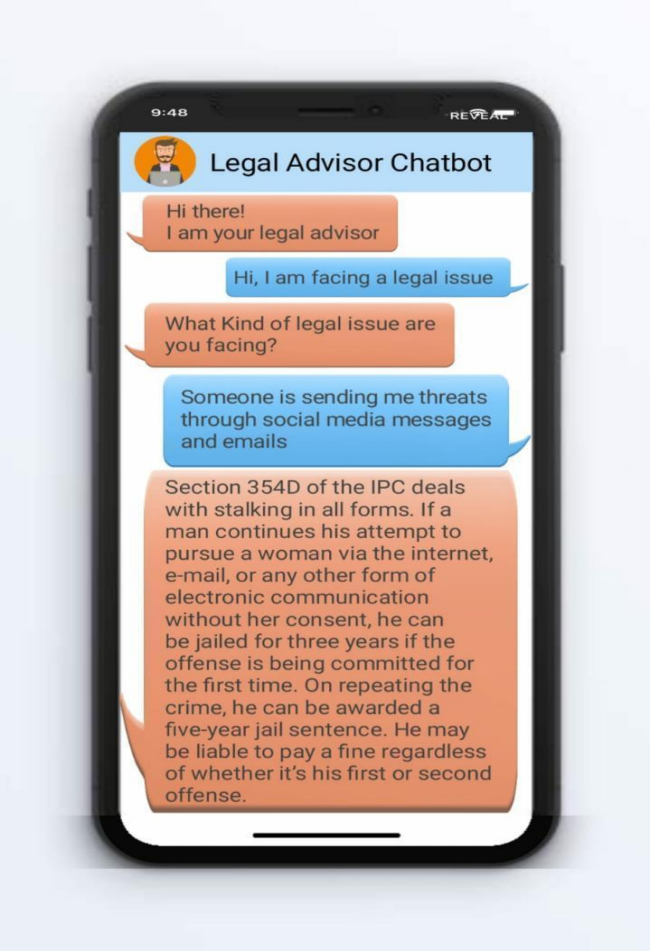

Fig .5 Legal Advisor ChatBot

IV. CONCLUSION

The goal is to create a chatbot that can assist individuals and solve problems quickly. To guarantee that users are satisfied, the chatbot should make user interaction simple and quick, and they should be able to acquire any information they need quickly from the chatbot's legal knowledge. Many people do not have access to legal services and are unaware of the cyber crime legal process, applicable legislation articles, and prison sentences. To

comprehend the issue or rules. To address all of these issues and to make the legal process more accessible to ordinary people, law students, and lawyers, we have developed a chatbot that can serve as a legal advisor and can be employed in the event of cybercrime.

V. ACKNOWLEDGMENT

We would like to convey our gratitude to Prof. Khalid Alfatmi for sharing their pearls of wisdom with us during this work

References

[1] Vorada Socatiyanurak ,Nittayapa Klangpornkun, Adirek Munthuli Phongphan Phienphanich Lalin Kovudhikulrungsri ,Nantawat Saksakulkunakorn, Phonkanok Chairaungsri and Charturong Tantibundhit , "LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors" IEEE Access 2021.

[2] Nishant Jain and Gaurav Goyal, "An Approach to Get Legal Assistance Using Artificial Intelligence" 8th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) 2020.

[3] Shang li,Hongli zhang,Lin ye, Xiaoding guo, and Binxing fang, "MANN: A Multichannel Attentive Neural Network for Legal Judgment Prediction"IEEE Access 2019.

[4] A. Khasianov, I. Alimova, A. Marchenko, G. Nurhambetova, E. Tutubalina and D. Zuev, "Lawyer's Intellectual Tool for Analysis of Legal Documents in Russian" International Conference on Artificial Intelligence Applications and Innovations (IC-AIAI) 2018.

[5] Saurav Pradha,Malka N. Halgamuge,Nguyen Tran Quoc Vinh, "Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data"

[6] Panitan Muangkammuen, Narong Intiruk, Kanda Runapongsa Saikaew, "Automated Thai-FAQ Chatbot using RNN-LSTM" IEEE Access 2018.

[7] Steven Bird, Ewan Klein, Edward Loper, "Natural Language Processing with Python"

[8]https://cybercrime.gov.in/

[9]https://www.engati.com/blog/types-of-chatbots-and-their-applications?utm_content=blog

[10]https://www.hubspot.com/stories/chatbot-marketing-future