# CHAPTER ONE

# INTRODUCTION

# 1.1 **Introduction**

The Oxford Advanced Learner’s Dictionary of Current English (2004) defines communication as the activity or process of expressing ideas and feelings or of giving people information. Giffin & Patten (1976) also state that communication is the process of creating meaning as well as ascribing it. It is the exchange of ideas and interaction among group members. Earlier, there was only verbal and non-verbal communication, but now there is another more effective form of communication, which is Online communication. Many people have started to communicate with the help of online communication. In fact, during the recent pandemic of COVID-19, online communication has picked up like anything.

More and more information is being churned out online ever than before. There is a lot of information for the reader to read online. People have started doing everything online, including but not limited to banking, reserving tickets, booking travel, planning travel, purchasing any and every kind of things, teaching, conducting meetings and seminars, one on one or group discussions, dating, sending information.

Every other activity which is possible is being done online. While basic tech literacy is essential for online activities, many can perform activities even with very little knowledge. The growth of online communication is fast and rapidly replacing traditional communication methods.

Over the years’ online communication has been affected with different problems which include:

**Privacy** Although the internet has made communication easier, quicker and convenient, privacy problems have emerged. From email hacking to phone signal interceptions, more and more people are now troubled about their private information.

**Lack of job security** Since technology keeps on changing, job security has become a problem. This means that IT experts need to be continuously learning to keep up with the changes if they want to retain their jobs.

**Overruling cultures** While the internet has made the universe a global village, some cultures have consumed others. For instance, teenagers in western countries have influenced most teenagers in other parts of the world on how they dress, act and behave.

**Toxic Words** According to research on computer-mediated communication by Lincoln Dahlberg and others. A term now heavily used by reporters, “trolling” is frequently used interchangeably to refer to bullying and hate speech, muddying the waters around the word’s definition and descriptive power. As a catchall media label, “trolling” invokes a kind of nebulous Internet folk devil rather than an actual person or persons behind the computer screen. It obscures the underlying hate speech. If observers were to shift away from such uses of “troll” and “trolling,” they could actually name specific toxic behaviors the sexism, racism, homophobia, transphobia, that they actually represent. Toxic behavior is pervasive in every online environment.

Toxicity can be manifested in different ways. It spans a wide spectrum that includes subtle and indirect signals; that can often be no less toxic than overly offensive language (Jurgens et al., 2019). While the research community has made enormous progress in detecting overly offensive language and hate speech (Schmidt and Wiegand, 2017; Waseem et al., 2018; Fortuna and Nunes, 2018; Qian et al., 2019), there has been less focus on computationally evaluating other subtle expressions of toxicity. The Ever increasing toxicity in online communication has given rise to the need of writing assistant tools that help prevent the use of toxic words in online communication in real time. Writing assistant tools are tools that help check for a specified error in writing and help in correcting or preventing them. AI writing assistants have come a long way since a Stanford student created the first spell-check software in the early 1970s. Since then, spell check, grammar checking, and autocorrect functionality have become commonplace, and people expect artificial intelligence to help them in their daily lives.

Having consider the problems attached to online communication this work will be focusing on the use of Toxic words. Which will lead to the development of a tool called Toxic Checker.

The researcher will be developing a writing assistant tool that check for the presence of toxic words in a text and prevent it from been sent.

# 1.2 Statement of the problem

The problem was actually motivated by the urge to totally eliminate the use of toxic language in online communication, which has led to the withdrawal of some people from the online environment with the aim of avoiding hate speech and online bully.

The most commonly use form of toxic language include, hate speech, online bullying, cyberbullying, trolling, xenophobia, racism, sexist

An attempt to deal with this issue is the introduction of crowdsourcing voting schemes that give the possibility to denounce inappropriate comments in online environments to the users. Among many others, Facebook, for example, allows its users to report a post in terms of violence or hate speech. This scheme allows Facebook to identify fake accounts, offensive comments, etc. However, these methodologies are often inefficient, as they fail to detect toxic comments in real time, becoming a requirement within social network communities. A toxic post might have been published online much earlier than the time it is reported and, during the time it is online, it might cause problems and offenses to several users which might have undesired behaviors (e.g., leaving the underlying social platform).

This has increase the need for a writing assistant agent for the prevention of the use of toxic language in our day to day activities in the online community

# 1.3 Aim and Objectives

The aim of the research is to develop a writing assistant that help prevent the use of toxic language in online forum and the specific objectives are as follows:

Develop a dictionary of toxic words and phrases and synonyms.

Develop an agent to sense toxic constant once keyed and prevent it.

Implement the agent in an online forum.

# 1.4 Research Methodology:

To achieve the aim of this work the waterfall development model will be adopted and the following steps will be taken:

Source for the frequently use toxic words in online communication across various social medias (Facebook, Twitter, Whatsapp)

Designing a tool using python together with javascript that check for the presence of toxic words in user input and helps prevent it and also check for the overall toxicity of a sentence and determine if it will be allowed or not.

# 1.5 Contribution to Knowledge

On completion of this work we should now have a tool that check user input for toxic words before allowing user to send the message.

# 1.6 Research questions

What is Toxicity in Online communication?

How can toxicity be prevent in online communication in real time?

# 1.7 Scope of the study

The work is focus on the prevention of individual toxic words that is commonly use across popular social media platforms which include facebook, blog posts, twitter and whatsapp but will be implement on a small blog page.

# 1.8 Definition of Terms

Toxic: extremely harsh, malicious, or harmful

Online Communication: is how people communicate, connect, transact to send, retrieve, or receive information of any kind via the internet using digital media

Research: A careful study of a subject to discover facts, establish a theory or develop a plan of action based on the facts discovered.

# CHAPTER TWO

# LITERATURE REVIEW

# 2.0 Introduction

This chapter gives an overview of the general knowledge on communication, online communication, toxicity in online communication, writing assistant tools and also related works to the research aim.

First part of the literature review talks about communication in great detail ranging from the definition of definition to the types to the forms and also leads us to the emergence of online communication, how much it has been of advantages to us in the society and the disadvantages of online communication.

Toxicity has been one of the major issue in online communication, more about this was given during the course of this chapter and how much techniques has been deployed in tackling toxicity and also the propose pattern which include the writing assistant tools.

Software agent in a broader view was also mentioned in the cause of the literature review, the various definition of software agents

Writing assistant agents has a history that runs into ages and also its emergence and growth over the days were discuss in this chapter and also the approach use during the course of the development of the tool (Dictionary based approach was discuss).

The chapter concluded with the brief overview of the related works to the research work.

# 2.1 Communication

The general view of communication is that it is an interaction within a social context. Communication usually involves a sender (source) and a receiver. It involves the interlocutors exchanging signals. These signals could be verbal or graphic, it could be gestural or visual (photographic). In essence, communication involves using codes that are done with the eyes, body movement or sounds made with the voice. Whichever way it is done, there is always a process in which someone initiates a meaning intent that is passed to the interlocutor (receiver). Daniel (2016) asserts that it is when feedback, which involves the receiver responding to the signal by initiating another circle of meaning exchange, has been sent to the sender (source) that the communication process has gone full circle and become complete.

Giffin & Patten (1976) also state that communication is the process of creating meaning as well as ascribing it. It is the exchange of ideas and interaction among group members. The Oxford Advanced Learner’s Dictionary of Current English (2004) defines communication as the activity or process of expressing ideas and feelings or of giving people information. One can safely say that communication is the act of transferring information and messages from one place to another and from one person to another. In a related manner, the Online Business Dictionary describes communication as a two-way process. It involves participants reaching a mutual understanding beyond merely encoding and decoding information, news, ideas and feelings. It is important that they also create and share the meaning content in the messages passed (see Daniel, 2013). In addition to this, communication is also seen as a means of connecting people or places. It is also regarded as an important key function of management because an organization cannot operate without communication between levels, departments and employees (*Online Business Dictionary*, 2010; Okenimpkpe, 2010). Again, communication can be defined as a field of study concerned with the transmission of information and broadcasting. It can involve any of the various professions, which have to do with the transmission of information such as advertising, public relations, broadcasting and journalism. The foregoing shows that communication is something human beings do every day in different ways and through different means. That is, the modern man communicates through different methods like speaking, using telephones, blogging, television, art, hand and body gestures and facial expressions. This can happen in closed intimate settings or over long distances. An example is the internet. The acts of communication draw on a number of inter and intra-personal skills like observing, speaking, questioning, analyzing and assimilating. It enables collaboration and cooperation.

Above all, language is the basic level of communication between one human being and another. It is the means by which we pass on our ideas, feelings, knowledge and requests. Awoniyi (1982) affirms that without communication there would be chaos. He adds that human existence and civilization as we know today would disappear without communication. He further claims that there are at least five important elements of the communicative process. This is shown in Table below.

**Table: Elements of communication**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sender (Encoder) | Message | Medium of Transmission | Medium of Receipt | Receiver (Decoder) |

Daniel (2016), following NOUN (2004), elaborates on how these elements of communication interact to make communication happen. The process requires the sender to encode message, through a medium of transmission, which the receiver receives and decodes through the medium of receipt. In essence, the sender must encode the message (i.e. the information being conveyed) into a form that is appropriate to the communication channel and the receiver(s) decode the message in order to understand the meaning content of the message. The goal of communication is for the receiver to understand the message sent. This is explained as the ability to categorize sender’s intent, understand the message sent and act upon it (Merriam-Webster Dictionary).

# 2.1.1 Human and Non- Human Communication

At this juncture, it is necessary to state that there is human and non-human communication. Human communication may be classified as verbal, non-verbal, written and other forms of electronic communication. Non-human communication includes animal communication, communication among plants and fungi and bacteria quorum sensing.

Based on the foregoing, possible answers to the question- “What is communication?” may be: the transmission, exchange of thoughts, messages or information through the medium of speech, signals, signs, writing or behaviour. It can also be said to be the art and technique of using words effectively to impart information or ideas. In simple terms, communication means speaking to or with some other person near or far away, either face-to-face through speech, telephone calls or text messages, body language, signs and so on**.** It can be done by both human and non-human beings. However, it is important to note that non-humans do not speakbut are likely to communicate through signals or symbolic behaviour. The essential point to observehere is that non-humans communicate too.

Above all, language is the basic level of communication between one human being and

another. It is the means by which we pass on our ideas, feelings, knowledge and requests.

Awoniyi (1982) affirms that without communication there would be chaos. He adds that

human existence and civilization as we know today would disappear without

communication.

# 2.1.2 Categories of Communication

Related to the above are various categories of communication. Of these categories, more than one may occur at a given time. For example, if you are with a group of close friends having a general conversation, you can at the same time be carrying on another tacit, non-verbal or sign communication with someone of your choice, which others in the group may not be aware of. Whatever the category or channel of communication, understanding is the desired outcome or goal of any communication act. Senders and receivers are very important in the communication process. There are many ways by which communication can take place as well as many skills one can employ in the process of communication. Another example is that a parent can communicate displeasure to a child over their behaviour in the midst of people by sign language without breaking the flow of conversation and without others being aware. These signs include frowning, clearing of the throat or stepping on the child’s toes.

# 2.1.3 Types and Levels of Communication

There are many types and varieties of communication, depending on the medium used or the way in which information is exchanged. For example, communication can be done through the internet, telephone mobile phones), speaking, singing, dancing, sign language, touch and eye contact, body language and even the way one dresses. All these inform the types of communication that exist. Teaching is also a type of communication because a teacher passes information to students through channels such as face-to-face in the class, by means of speech, visual aids and experiments, through written notes, on the internet which could be through

Skype, video conferencing or blogging. In addition, there is human (verbal, non-verbal, business or written) and non-human (occurring among animals, plants and fungi, and bacteria) communication. Generally, the following are the types of communication.

**Verbal communication:** This, as the name implies, is communicating using words. This includes sounds, words, language and speaking. Speaking is an effective way of communicating: it is classified into interpersonal communication and public speaking (Aarti, 2011). Interpersonal communication occurs when one person speaks directly to another. Here, communication would be informal: one can say what one really feels, though this is bound by the social norms guiding the sender and the receiver. Public speaking occurs when one person speaks to a large group. In this case, communication is formal; rule bound and centred more on the speaker getting some kind of result. In all of these, speakers may want to entertain, inform, persuade or argue. Verbal communication can also be referred to as oral or spoken. It can make use of visual aids and non-verbal elements to facilitate meaning and enhance rapport and achieve high level of understanding by removing ambiguity and garnering immediate feedback.

**Non-verbal communication**: Non-verbal communication is the process of conveying meaning in the form of non-word messages. It covers all information, messages and ideas we convey without using words; making use of physical communication such as tone of voice, touch, smell and body motion. Non-verbal communication includes music, dance, painting, drama and sculpture. Symbols and sign language are also included (cf. Daniel, 2016). This is because body language, facial expressions, clearing of the throat, physical contact and dress convey a lot of information. A good example of non-verbal communication is sign language, which can be used by anyone at anytime. A non-verbal sign such as bell ringing is a common sign you and I understand. By itself, ringing of a bell means nothing. However, in recognized contexts, it may mean “time for school”, “change of lesson”, “is somebody at home”, “end of lesson”, or “second hand goods for sale.” According to Wilson (1972), not only can almost anything be used as a sign, but almost every sign can be used to communicate several different things. He explains

further that everything depends on agreement about and understanding of the ways in which we use signs. To him, the significance of any sign depends on the context in which it is used.

**Written communication:** This means communicating with other people through written words. Good written communication is practised in many different languages. Email and text messages, reports, articles and memos are some of the ways of using written communication, both for business and personal purposes. An advantage of written communication is that it can be edited and amended many times before it is finally sent to the person intended. Writing is a human invention. There are as many legends and stories on the invention of writing as there are on the origin of language. Legend has it that Cadmus, the prince of Phoenicia, invented the alphabet and brought it to Greece. Stories also have it that the early drawings by ancient humans formed the seeds of present day picture writing. Many such drawings are clearly picture writings in pictograms. Historically, written communication first emerged through the use of pictograms, which were made on stone (Yule,2010). Later, writing began to appear on paper, papyrus, clay and wax. Now, communication is carried on by the transfer of information through controlled waves and electronic signals. Note that the four language skills (listening, speaking, reading and writing) are used in communication but writing is the most complex. Each of these skills is discussed extensively in the subsequent chapters.

**Visual communication:** This is a visual display of information such as topography, photography, signs, traffic codes, symbols and designs. Television and video clips are the electronic form of visual communication (Aarti, 2011).

Intra-personal communication is a level of communication that is general to all types of communication. It is described by Answers.com as language use or thought that is internal to the communicator. Daniel (2016) graphically presented this as “communicating with yourself” (p.32). In essence, talking to yourself is not always about being ‘crazy’ or ‘loony’, she avows. She notes that we do it all the time. To her, it is about tha kind of information that involves the speaker also being the receiver of the message but which most of us are not conscious of even doing. As such, when you write a shopping list, you are engaging in intra-personal

**Non-human communication**

Non-human communication essentially has to do with the communication that does not involve human beings. This is usually described as extra communication. It does suggest that plant and animal communication can be put in this category. In the same vein, cellular communication among micro-organisms could be described in this manner. One could safely describe the transfer of chemical reactions among bacteria and fungi as non-human communication. Even the transfer of viruses as well as the multiplication of cancer cells could be categorised as this kind of communication. I am certain that it is such communication that made the Ebola virus travelled so fast and wide and killed many people along the way during the period it was raging in West Africa recently.

Animal communication can be defined as any behaviour of one animal that affects the behaviour of other animals within the same environment, either now or at a later time. Zoo semiotics is the name that has been used to describe scholarship in this category. It is considered important to the development of the study of animal cognition. Nonetheless, Martinelli (2010) asserts that there is still “a certain curiosity” which surrounds it. This attitude is said to manifest in scepticism and ‘exoticism’. This does not remove the fact that it exists as its fifty years of practice (Martinelli, 2010) appears to affirm.

As noted above, communication is observed within the plant organism. Communication between plant cells, between plants of the same or related species, and between plants and non-plant organism also occurs. These are possible because of the apparent decentralised ‘nervous system’ of plants (Witzany, 2006). This shows that plant also communicate with other plants, especially, when exposed to attack behaviour from plants within their vicinity. In this way, they are actually warning neighbouring plants of possible danger around them. It could therefore be said that communication is not an instrument exclusive for use by only humans, plants and animals, but also available to microorganisms like bacteria. One may be right to relate this to what happens with viral spread such as when an infection is shared within the body of a victim of viral of viral attack. You will agree that such communication may result into the victim’s ailment or, sometimes, even death.

**Mass Communication**

Daniel (2016) describes mass communication as massive. This is because it is usually from a single source to a very large group or audience. The message sent out is not limited to one or two people as with the interpersonal kind of communication. As the name suggests, its audience is usually a very large group of people. For this reason, such media as the newspapers, radio, television or the internet are used to pass the message. Simply then, we can say that mass communication is another means of communicating information, ideas, and messages through various print or electronic media to a large number of people.

Nonetheless, we need to also note that mass communication, also refers to a course of study. For example, in the Adekunle Ajasin University, there is a department of Mass Communication. In the same vein, many institutions of higher learning offer it’s a course. This is the course that is studied and helps to train many media practitioners around us.

**Codes of communication**

Lawal (2004,18) affirms that language is comprehensively a communication code, a signalling or semiotic system. The *Oxford Advanced Learner’s Dictionary of Current English* defines code as a system of words, letters, numbers or symbols that represent a message or records information secretly. Craig (2000) is of the opinion that communication can be classified according to code, medium or channels through which it occurs and that language is a primary code (or, in different perspectives, a medium or channel) of human communication. He asserts further that code refers to the way a sign system is structured to constitute a particular correlation of signs to meaning. Language is therefore a primary code as well as a medium and channel of communication.

Human spoken and pictorial languages can be described as a system of symbols. These are usually set in patterns of communication, which are called grammars. Many languages of the world make use of patterns of the sound or graphs for symbols, which enable communication with others around them (cf. Yule,2010; Daniel,2011). Yule (2010) actually describes sign language in details. And it is obvious that some broadcasting houses in Nigeria and elsewhere also see that using sign language as a sub-sect of their news broadcast is an important part of connecting with their audience, no matter their physical needs. One cannot but agree with them.

Other codes of communication include signs such as traffic codes, kinesics (gesture, facial expression and body movement), haptics (touch), and paralanguage (vocal pitch, intonation).

# 2.2 Online Communication

The term "online communication" refers to reading, writing, and communication via networked computers. It encompasses synchronous computer-mediated communication (whereby people communicate in real time via chat or discussion software, with all participants at their computers at the same time); asynchronous computer-mediated communication (whereby people communicate in a delayed fashion by computer, using programs such as e-mail); and the reading and writing of online documents via the World Wide Web. Second language researchers are interested in two overlapping issues related to online communication: (1) how do the processes which occur in online communication assist language learning in a general sense (i.e., online communication for language learning); and (2) what kinds of language learning need to occur so that people can communicate effectively in the online realm (i.e., language learning for online communication).

Online social media platforms are arguably among the most culturally significant technological innovations of the 21st century. The numerous benefits include the wide distribution of content crossing geographic boundaries, and enabling interaction and exchanges that are nearly free of physical constraints except for infrastructure. Communities have emerged around every conceivable special interest from science to travel, from politics to child-rearing. The easy spread of data, information, and knowledge was expected to foster informed decision-making, cultural exchanges, and the coordination of activities online and in the physical world.

# 2.2.1 History of Online Communication

Online communication dates back to late 1960s, when U.S. researchers first developed protocols that allowed the sending and receiving of messages via computer (Hafner & Lyon, 1996). The ARPANET, launched in 1969 by a handful of research scientists, eventually evolved into the Internet, bringing together some 200 million people around the world at the turn of the millennium. Online communication first became possible in educational realms in the 1980s, following the development and spread of personal computers. The background on online communication in language teaching and research can be divided into two distinct periods, marked by the introduction of computer-mediated communication in education in the mid-1980s and the emergence of the World Wide Web in the mid-1990s.

In the first period, dating from the mid-1980s, language educators began to discover the potential of computer mediated communication for language teaching (Cummins, 1986). The integration of computer-mediated communication in the classroom itself divided into two paths: on the one hand, some educators began to use e-mail to set up long-distance exchanges, and, on the other hand, other educators began to use synchronous software programs (in particular, Daedalus Interchange (Daedalus Inc., 1989) to allow computer-assisted conversation in a single classroom. Long-distance exchanges and computer-assisted conversation had overlapping, but distinctive, justifications. Both types of activities were seen to shift the focus from language form to language use in meaningful context (e.g., Kelm, 1992; Meskill & Krassimira, 2000) and thereby increase student motivation (e.g., Meunier, 1998; Warschauer, 1996b). In addition, long-distance exchanges were viewed as brining about increased cultural knowledge from communication with native-speaking informants (e.g., Kern, 1995a; Soh & Soon, 1991), and making reading and writing more authentic and collaborative (e.g., Tella, 1992b). Those implementing computer-assisted conversation emphasized the linguistic benefits which could be achieved from rapid written interaction, such as better opportunities to process and try out new lexical or syntactic patterns as compared to oral interaction (e.g., Ortega, 1997; Warschauer, 1999).

Online communication is a new phenomenon, having first come into existence toward the end of the 20th century. It is growing at one of the fastest rates of any new form of communication in human history, and its long-term impact is expected to be substantial. A not uncommon, and, in my eyes, justifiable, view is that online communication represents the most important development in human communication and cognition since the development of the printing press (Harnad, 1991).

# 2.2.2 Types of Online Communication

**Video Conferencing**: Video conferencing have progressed to become one of the best types of online communication. Organization benefit from online conferencing through meetings. The reason for this is meetings let for the discussion of business improvement and the coordination of ideas. Video conferencing software like ezTalks Cloud Meeting let for organizing meetings. It does not matter about the physical locations of a participant therefore best for emergency times. This shows that this promises a quick redress of organization problems and raised productivity. It is affordable and facilitates nonverbal communication also.

**Whiteboards:** Interactive whiteboards are the most new types of communications tools on the internet. They are famous tools for online communication in education. They are useful because they let users to draw,write and communicate with the assistance of an interface that activates a real physical whiteboard. These have become the best online education tools due to the reason their features help in learning. Whiteboards need users to get a mouse to include some content.

**Forums:** There are many methods by that online forums are meant to as,these are discussion boards,discussion groups,and bulletin boards. They can be explained like areas where the users are permitted to post the questions or comments. These forums online are keyboard accessible. These tools use inaccessible for registration. These are insightful tools in education because they has plenty of valuable information.

**Voice over IP:**VoIP is a word which is availed in reference to a voice changing that is made over the internet link instead of over the normal phone line. Several big companies avail VoIP in place of the traditional phone line. It must be considered that it is probable to do VoIP calls over a system however with handsets or a USB headset. The calls are audio oriented and a good example of this kinds of communication tools is Skype. Web 2.0 has developed a new approach and vision from improving technology in to highly interactive and communicational. Magazine such as websites were replaced with interactive,powerful online spaces where authors and visitors can give and assist creating the site and shape the information.

**Email:** Email is also called as electronic mail is among the old tools in types of online communications. It let users to change photos,get news,and send files from different part of the globe. This tool is highly used for public communication,especially in mailing lists. This is not exhaustive because there are other fundamental online tool for communication like newsletters,websites and social networks. The internet has used a better amount of new key stream communication tools which has revolutionized things. The reason for this is different kinds of tools for online communication have greatly decreased the time spend in making and providing messages.

**Social Networks** Social networks has established from knowing that active users online might need to connected and get in touch with each other providing news,life updates,interests etc. Profiles pages which assist you to study about each other were placed in several websites and connected collectively with messaging,chats and extra features like music,video and photo sharing,schooling,group by interests and work. Few social networks are particularly made by interests,people who like music.

**Blogs**: Often referred to as social media, blogs exist on every topic imaginable, and in every industry. Blogs are web logs that are updated on a regular basis by their author. They can contain information related to a specific topic. Blog's give an opportunity to write content that is unique to you and your “practice”. While some people are uncomfortable with self-promotion, blog gives you an occasion to interact with your visitors while promoting who you are and what you do. Today blog’s are being used for all sorts of purposes. From companies that use blogs to communicate and interact with customers and other stakeholders to newspapers that incorporate blogs to their main website to offer a new channel for their writers. Individuals also created blogs to share with the world their expertise on specific topics

# 2.2.3 Advantages of Online Communication

Flexibility: accessible 24×7, any place as long as you have an internet connection

**Levelling**: reserved people who usually don’t speak up can say as much as they like while “loud” people are just another voice and can’t interrupt

**Documented**: unlike verbal conversation, online discussion is lasting and can be revisited

**Encourages reflection**: participants don’t have to contribute until they’ve thought about the issue and feel ready

**Relevance**: provides a place for real life examples and experience to be exchanged

**Choice**: a quick question or comment, or a long reflective account are equally possible

**Community**: over time can develop into a supportive, stimulating community which participants come to regard as the high point of their course

**Limitless**: you can never predict where the discussion will go; the unexpected often results in increased incidental learning

# 2.2.4 Disadvantages of Online Communication

**Text-based**: Predominantly relies on inputting text which can be challenging for those who don’t like to write or have poor keyboard skills, but with the advance of broadband connectivity and voice and video conference technology – this will be less of an issue.

**No physical cues**: without facial expressions and gestures or the ability to retract immediately there’s a big risk of misunderstanding

**Information overload**: a large volume of messages can be overwhelming and hard to follow, even stress-inducing

**Threads**: logical sequence of discussion is often broken by users not sticking to the topic (thread)

**Time lag**: even if you log on daily, 24 hours can seem like a long time if you’re waiting for a reply; and then the discussion could have moved on and left you behind

|  |
| --- |
| Inefficient: it takes longer than verbal conversation and so it’s hard to reply to all the points in a message, easily leaving questions unanswered  **Isolation**: some learners prefer to learn on their own and don’t participate in the discussions  **Directionless**: participants used to having a teacher or instructor telling them what to do can find it a leaderless environment (and that’s where tutors come in)  **Toxic Words:** The use of harmful words has been a great concern for online communication 2.3 Toxic Language In these years, short text information is continuously being created due to the explosion of online communication, social networks, and e-commerce platforms. Through these systems, people can interact with each other’s, express opinions, engage in discussions, and receive feedback about any topic. However, a paramount inconvenience within online environments is that text spread by digital platforms can hide hazards, such as fake news, insults, harassment, and, more in general, comments that may hurt someone’s feeling. These comments can be considered to be the digital version of personal attacks (e.g., bullying behaviors) that can cause social problems (e.g., racism), and they are felt to be dangerous and critical by people who are struggling to prevent and avoid them. The risk of such a phenomenon has increased with the event of social networks and, more in general, within online communication platforms (https://medium.com/analytics-vidhya/twittertoxicity-detector-using-tensorflow-js-1140e5ab57ee). An attempt to deal with this issue is the introduction of crowdsourcing voting schemes that give the possibility to denounce inappropriate comments in online environments to the users. Among many others, Facebook, for example, allows its users to report a post in terms of violence or hate speech. This scheme allows Facebook to identify fake accounts, offensive comments, etc. However, these methodologies are often inefficient, as they fail to detect toxic comments in real time, becoming a requirement within social network communities. A toxic post might have been published online much earlier than the time it is reported and, during the time it is online, it might cause problems and offenses to several users which might have undesired behaviors (e.g., leaving the underlying social platform).  Social media can be a powerful tool that enables virtual human interactions, connecting people and enhancing businesses’ presence. On the other hand, since users feel somehow protected under their virtual identities, social media has also become a platform for hate speech and use of toxic language. Although hate speech is a crime in most countries, identifying cases in social media is not an easy task, given the massive amounts of data posted every day. Therefore, automatic approaches for detecting online hate speech have received significant attention in recent years (Waseem and Hovy, 2016; Davidson et al., 2017; Zampieri et al., 2019b)  Toxicity can be manifested in different ways. It spans a wide spectrum that includes subtle and indirect signals; that can often be no less toxic than overly offensive language (Jurgens et al., 2019). While the research community has made enormous progress in detecting overly offensive language and hate speech (Schmidt and Wiegand, 2017; Waseem et al., 2018; Fortuna and Nunes, 2018; Qian et al., 2019), there has been less focus on computationally evaluating other subtle expressions of toxicity  Online toxicity, defined as hateful communication that is likely to cause an individual user leave a discussion, can manifest itself in various ways, including cyberbullying, trolling, and the creation of online firestorms, defined as “rapid discharges of large quantities of negative, often highly emotional posts in the social media environment” ,where participants attack other groups or organizations. According to Patton et al., online toxicity may result in violent actions also in the physical world and should, therefore, be treated as a matter with serious social gravity. In online environments, toxic behavior is often seen enhanced by the fact that participants can typically comment anonymously and are not held accountable for their behavior in the same way as in offline interactions.  Online hate speech is can be seen as old as the Internet itself. Anti-Semitic and racist hate groups were active on Bulletin Board Systems as early as 1984. In the present time, some communities are specifically geared towards promoting hate speech and providing avenues for expressing politically incorrect values that may not comfortably be expressed in face-to-face interactions. Toxic commenting has also been found prevalent in general online discussion forums, news websites, and social media platforms. The existing research deals with multiple aspects, such as detection and classification of toxicity, assessing its impact on online communities, types of toxicity such as cyberbullying and trolling, and means of defusing online toxicity. To approach toxicity, researchers have investigated multiple social media platforms, such as Twitter, YouTube, Facebook, and Reddit , as well as comments in online discussion forums and news websites. Due to its high prevalence, toxicity has been identified as a key concern for the health of online communities.  defining what we mean by toxic comments. They largely fall into two categories: hate speech and online harassment (aka trolling).  There’s vigorous debate over what constitutes hate speech, but the United Nations defines it as “any kind of communication that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor”.  The UN Human Rights Office of the High Commissioner reported in March 2021 that hate  speech is on the rise worldwide, and that three-quarters of it is directed at minority groups.  Online harassment is another umbrella term that can be applied to many different kinds of comment. Pew Research Center categorizes it into six types: purposeful embarrassment, offensive name-calling, physical threats, stalking, sustained harassment, and sexual harassment. It says that 41% of Americans experienced at least one of these online in 2020. 2.3.1 Some Toxic Words anal, anus, ass, bastard, bitch, boob, cock, coon, cripple, cum, cunt, dick, dildo, dyke, fag, faggot, fuck, fudgepacker, fuk, greaseball, gypo, handjob, homo, jihadi, jizz, kike, knacker, kunt, muff, muzzie, nigga, nigger, niggur, peckerwood, penis, piss, poop, porch monkey, pussy, queer, raghead, rape, retard, sand nigger, semen, shit, shyster, slut, titties, twat, uncle tom, vagina, vulva, wank, yellow bone. |
| 2.4 Software Agents Software agent technology is a rapidly developing area of research and probably the fastest growing area of information technology (IT) (Nwana & Ndumu, 1996; Jennings & Wooldridge, 1996). Application domains in which agent solutions are being applied or researched into include workflow management, telecommunications network management, air-traffic control, business process reengineering, data mining, information retrieval/management, electronic commerce, education, personal digital assistants (PDAs), e-mail filtering, digital libraries, command and control, smart databases, and scheduling/diary management (Nwana & Ndumu, 1996).  Jennings & Wooldridge (1996) offer a relatively loose notion of an agent as a self-contained program capable of controlling its own decision making and acting, based on its perception of its environment, in pursuit of one or more objectives will be used here.  Wooldridge (1998) defines an intelligent agent as a system that enjoys the following four properties: autonomy (agents operate without the direct intervention of humans or others, and have control over their actions and internal state), social ability (agents are able to cooperate with humans or other agents in order to achieve their tasks), reactivity (agents perceive their environment, and respond in a timely fashion to changes that occur in it), and pro-activeness (agents do not simply act in response to their environment, they are able to exhibit goal-directed behavior by taking the initiative).  According to Hayes (1999), an agent is an entity (either computer, or human) that is capable of carrying out goals, and is part of a larger community of agents that have mutual influence on each other. Agents may co-exist on a single processor, or they may be constructed from physically, but intercommunicating processors (such as a community of robots) (Hayes, 1999). The key concepts in this definition are that agents can act autonomously to some degree, and they are part of a community in which mutual influence occurs (Hayes, 1999). |

# 2.4 Writing Assistant System

Writing assistants like Writer are online tools designed to catch grammatical errors and writing mistakes to help to make your writing better. There are human writing assistants, but they tend to either help authors write manuscripts or work in the writers’ room to help screenwriters craft movies and plays. AI writing assistants have come a long way since a Stanford student created the first spell-check software in the early 1970s.

Since then, spell check, grammar checking, and autocorrect functionality have become commonplace, and people expect artificial intelligence to help them in their daily lives — whether they realize it’s AI or not. New AI writing technologies have emerged beyond Grammarly and a cohort of copycats.

Now, AI writing assistants specialize in resume writing (Skillroads), English as a second language (Ginger, Grammarly, Linguix, LanguageTool), customer service and product UX (Writer), advertising (Persado, Phrasee), and many more niches, professional disciplines, and use cases for individuals and businesses.

They are pretty easy to work with, too. You type something into the program. Then the writing assistant checks and fixes your mistakes, in real-time, to improve the quality of your writing.

# 2.4.1 History of Writing Assistant Tools

**1971 The first spell check** is open-sourced, by a graduate student at Stanford’s Artificial Intelligence Laboratory.

**1981 Grammatik** is the earliest grammar checker to hit the market; more of a style and diction checker for PCs, it was years before it could check actual grammatical mistakes. Grammatik was eventually acquired by WordPerfect Corporation and integrated with the WordPerfect word processor.

**1993** Dean Hachamovitch at Microsoft invents **autocorrect for Word**, creating a UX pattern that becomes familiar to billions of computer users. An early grammar checker is released as well.

**2007** Autocorrect comes to **Apple’s mobile OS**.

**2010-2020** Years of progress in machine learning, natural language processing (NLP), and neural networks that allow for the next-generation of paid and free writing assistants to be AI-based.

The advances in deep learning (i.e., neural networks) meant that massive corpuses of well-edited content could train a grammar model on what good grammar looks like — no complex rules necessary. Further advances in machine learning made setting up and deploying these models into products that were easy for users became easier, too.

**Feb 2019** Google’s AI-based grammar checker comes to **Google Docs**.

**Apr 2019 Writer’s AI-based grammar and proofreading check** is released, allowing AI-based writing assistance across the web through its Chrome browser extension.

**Sep 2019 Grammarly** releases its tone detector.

**Nov 2019 OpenAI** releases full version of GPT-2, an AI-based text generator that can create human-sounding long-form content (including fake news).

Some other types of writing assistant agents include:

**WriteBetter**, a corpus-based writing assistant designed to be integrated into Microsoft Word, Google Docs, and Overleaf. This integration makes its use straightforward and easy as users can see corpus-based examples (1) in real-time while writing in the word processor or (2) just selecting a piece of text in their document.

**Jasper** is an AI copywriting tool with superb expertise. Jasper uses natural language processing (NLP) to instantly create original content that sounds human and appeals to search engine

**Zoho Writing Assistant** is a cloud-based word processor akin to Google Docs. With Zoho, you can create written content on a minimalist interface, review your work with teammates, and then export it to a platform of your choice.

**Writing Assistant** is an AI-powered software that works to improve written content by enhancing grammar and clarity. It examines copy, flags errors, and suggests changes for improvement.

**ShortlyAI**, recently acquired by Conversion AI, is an AI writing assistant that continues writing for you. This tool requires that you write at least one sentence and then it keeps trucking along the rest of the way for you.

**ProWritingAid** is dedicated to helping everyone become a better writer. It spots spelling, grammar, and general writing errors like repetitiveness, ensuring clarity and readability.   
2.5 Dictionary Based Approach

The methods for generating opinion lexicon falls into two main categories, dictionary and corpus-based approaches. The former involves a static dictionary of semantically relevant words tagged with both a polarity label and semantic orientation score or reliability label Dictionary based approach: it is a method that translates a word by word as a dictionary without correlating the meaning of words between them

A text based on the use of a dictionary containing words and their classified categories. For example, the word “agree” belongs to the word categories: assent, affective, positive emotion, positive feeling, and cognitive process

The lexicon based approach uses sentiment dictionary with opinion words and match them with the data for determining polarity.

in the dictionary based techniques, the idea is to first collect a small set of opinion words manually with known orientations, and then to grow this set by searching in the WordNet dictionary for their synonyms and antonyms.

# 2.6 Related works

**Defining and Detecting Toxicity on Social Media: Context and Knowledge are Key Amit Sheth1 , Valerie L. Shalin2 , Ugur Kursuncu** aim at detecting toxicity in social media communication and also provide an approach that identifies multiple dimensions of toxicity and incorporates explicit knowledge in a statistical learning algorithm to resolve ambiguity across such dimensions which require more sophisticated Natural Language Processing (NLP) and Machine Learning (ML) methods to both detect and use the features indicative of toxicity.   
Offensive Language Detection: Perspective API is a popular toxicity detector for detecting offensive conversations.

Waseem et al. (2018) devised a taxonomy and created a dataset to detect hate speech and discrimination. In this paper, they identified the multiple influences on the detection of toxic exchange beyond conventional content analysis. For toxicity detection they provided a framework founded on behavioral and social theory.

**Chatzakou et al. (2017)** The work is motivated by the need to test and provide a system that categorizes emotion in online activities.  The main contributions of this work are to: (a) detect primary emotions, social ones, and those that characterize general affective states from online text sources, (b) compare and validate different emotional analysis processes to highlight those that are most efficient, and (c) provide a [proof of concept](https://www.sciencedirect.com/topics/engineering/proof-of-concept) case study to monitor and validate online activity, both explicitly and implicitly. They used soft classification techniques (Naïve Bayes classifier) to produce the class membership of probabilities of each tweet. Chatzakou, Vakali, and Kafetsios used handcrafted features with different statistical machine learning methods to classify news and social media data into different emotion classes. They performed their experiments on English and Greek. Study results demonstrate that the methodologies at hand succeed to detect a wider spectrum of emotions out of text sources.

Zampieri et al. (2019a) SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval) The task was based on a new dataset, the Offensive Language Identification Dataset (OLID), which contains over 14,000 English tweets. It featured three sub-tasks. In sub-task A, the goal was to discriminate between offensive and non-offensive posts. In sub-task B, the focus was on the type of offensive content in the post. Finally, in sub-task C, systems had to detect the target of the offensive posts. The evaluation results have shown that the best systems used ensembles and state-of-the-art deep learning models such as BERT. Overall, both deep learning and traditional machine learning classifiers were widely used. They released a corpus for offensive posts named OffensEval which has been encouraging researchers to study offensive contents.

**Safi Samghabadi et al. (2020)** data instance is annotated into one of the three aggression classes - Not Aggressive, Covertly Aggressive, Overtly Aggressive, as well as one of the two misogyny classes - Gendered and Non-Gendered. We propose an end-to-end neural model using attention on top of BERT that incorporates a multi-task learning paradigm to address both the sub-tasks simultaneously. Our team, “na14”, scored 0.8579 weighted F1-measure on the English sub-task B and secured 3rd rank out of 15 teams for the task. The code and the model weights are publicly available at https://github.com/NiloofarSafi/TRAC-2. Keywords: Aggression, Misogyny, Abusive Language, Hate-Speech Detection, BERT, NLP, Neural Networks, Social Media

**Xu et al. (2012)** studied bullying and propose a fast training procedure to recognize these emotions without explicitly producing a conventional labeled training dataset. And also use sentiment analysis approach.

**Rajamanickam et al. (2020)** Joint Modelling of Emotion and Abusive Language Detection present the first joint model of emotion and abusive language detection, experimenting in a multi-task learning framework that allows one task to inform the other. Their results demonstrate that incorporating affective features leads to significant improvements in abuse detection performance across datasets.

# CHAPTER THREE

# RESEARCH METHODOLOGY

# 3.1 METHODOLOGY

The methodology describes the broad deep-thinking behind choosing a research method, including whether it is using qualitative or quantitative methods, or a mixture of both, and why. The methods section describes actions to be taken to investigate a research problem and the rationale for the application of specific procedures or techniques used to identify, select, process, and analyse information applied to understanding the problem, thereby, allowing the reader to critically evaluate a study’s overall validity and reliability.

According to Clifford Woody, research comprises defining and redefining problems, formulating hypothesis or suggesting solutions; collecting, organizing and evaluating data; making deductions and reaching conclusions; and finally, carefully testing the conclusions to determine whether they fit the formulated hypothesis. Also, research encompasses "creative and systematic work undertaken to increase the stock of knowledge, including knowledge of humans, culture and society, and the use of this stock of knowledge to devise new applications". Methodology on the other hand is the general research strategy that outlines the way in which research is to be undertaken and, among other things, identifies the methods to be used in it. These methods, described in the methodology, define the means or modes of data collection or, sometimes, how a specific result is to be calculated.

Therefore, research methodology can be defined as a way to systematically solve the research problem (which includes procedures for collecting and analysing data necessary to define or solve the problem for which the research is based on). It may be understood as a science of studying how research is done scientifically.

In Howell (2013), methodology is defined as the general research strategy that outlines the ways in which research is to be undertaken and, among other things, identifies the methods to be used in it. These methods, described in the methodology, define the means or modes of data collection or, sometimes, how a specific result is to be calculated.

# 3.1.2 Methodology Details

The aim of this research is to develop a blog web app and prevent users from using toxic words or phrase in the community by creating a tool using javascript responsiveness on web application to prevent users from using some predefined toxic words that we’ve gathered from reliable sources and developing a toxic classifier using the Logistic Regression algorithm and also the google perspective API to prevent toxic contents that were copied and paste from getting access to the page.

Part A: Designing the Blog Webpage

1. **App Set Up**: Since the web app will be developed using flask web app, we have to go through the main steps of setting up the flask app
2. **APP Structure and Design:** Set up a proper structure for the app using html for the structure and including CSS and Boostrap for the design and Javascript for the web page responsiveness
3. **Storing Data:** We will use SQLITE for the database system of this app since it’s easily integrated with python
4. **Toxic Checker:** After building the web page the toxic checker will be developed using JAVASCRIPT because of its large community and its strength on web applications.

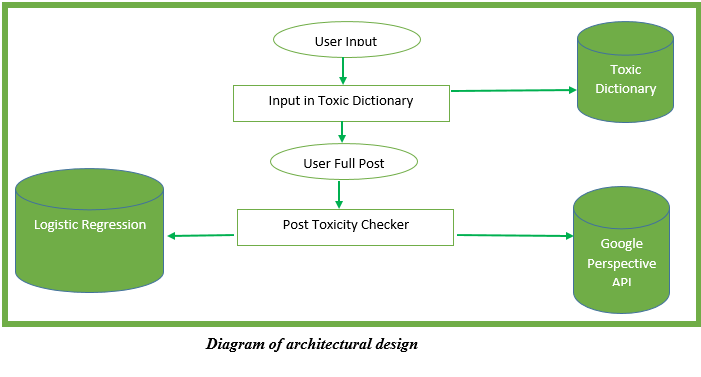
Part B: Building The Classifier

1. **Data Collection:** Both the predefined toxic words and the toxic comments that were used to train the models were gotten from the same source, the dataset provided by kaagle on kaagle for the toxic comment classification challenge, and also this same dataset was used in training the google perspective API for toxic comment classifier.
2. **Data Extraction:** Quite a length of python code was used to extract the most frequent words from the toxic comment classification dataset and a manual cleaning was carried out to remove words that are frequent but not toxic
3. **Data Preparation:** Data was prepared by performing few preprocessing and cleaning on the data which includes:

* Stop words removal
* Removing Special Characters
* Performing Stemmatiztion
* Lemmatization
* Transforming words to vector using TFIDFVECTORIZER

1. **Data Analysis:** The project didn’t require much of analysis but a few exploratory data analysis was carried out on data mainly focusing on the count of the comment by severity of comment.

**3.2 Architectural Design**



**3.2.1 Details of Architectural Design**

**I. User Input:** The User input is a page on the web app where the user will be allowed to input the content of their post it’s a form which was designed with the combination of html, css and flask framework in python for web development.

**II. Input in Toxic Dictionary Checker:** After a user typed a word and click spacebar on the user input page, this section checks if the word inputted is present in the toxic dictionary this functionality was developed using JavaScript.

**III. Toxic Dictionary:** This is a repository of toxic words and phrases that were generated by filtering out the frequently used words in our dataset and further subjected to manual cleaning by deleting words that appear frequently that weren’t toxic. This is a txt file that will be loaded into the page onload thorough javascript and store it into an array of toxic words for easy validation of toxicity.

**IV. Post Toxicity Checker:** If the user was successful in using words that are toxic but weren’t present in our toxic dictionary this phase of the architecture subjects it to further check by checking for the degree of toxicity in the text. This uses various techniques

**V. Logistic Regression:** This is the section of the model building for the toxic comment classifier and the following steps were carried out in building the models.

1. Data Collection: The same dataset that was used in generating the toxic words i.e the toxic comment classification competition dataset was used in the model building.
2. Data Cleaning: The data was clean firstly by removing stop words and by doing replacement of text
3. Data Preparation: the data was lemmatize and stemmatize and also base on the dataset the problem tends to be an unbalanced classification dataset due to large range of clean comments to toxic comment, to fix this the data was under sampled to prepare it for the model building and lastly vectorization was performed using the TFIDFVECTORIZER .  
   **stemmatization**: It is defined as the process which produces variants of a root/base word. In simple words, it reduces a base word to its stem word. We use stemming to shorten the look-up and normalize the sentences for a better understanding. example: 1.for the root word “like “ Stemming will include : -”likes” -”liked” -”likely” -”liking”  
   **lemmatization:** It is the process of assembling the inflected parts of a word such that they can be recognized as a single element, called the word‟s lemma or it‟s vocabulary form. This process is the same as stemming but it adds meaning to particular words. In simple words, it connects text with alike meanings to a single word. It is defined as an algorithm technique of finding the lemma of a word which is a root word rather than a root stem. It is based on the intended meaning the word is trying to convey. Example: 1. rocks: rock 2. Corpora: corpus 3. Better: good  
   **vectorization:** Vectorization is jargon for a classic approach of converting input data from its raw format (i.e. text ) into vectors of real numbers which is the format that ML models support. This approach has been there ever since computers were first built, it has worked wonderfully across various domains, and it’s now used in NLP. In Machine Learning, vectorization is a step in feature extraction. The idea is to get some distinct features out of the text for the model to train on, by converting text to numerical vectors.
4. Data Splitting: The data was split in to train and test for the purpose of training and evaluation of the model. 80% for training the model and 20 percent for evaluation
5. Model Building: The Logistic Regression Model was built using the python built-in library scikit-learn (sklearn) and were trained using the train section of our data.
6. Model Evaluation. Since the model is a classification model the following were used to evaluate the models this includes: precision, recall and f1score   
   **accuracy**: In general, the accuracy metric measures the ratio of correct predictions over the total number of instances evaluated.   
   **precision:** Precision is used to measure the positive patterns that are correctly predicted from the total predicted patterns in a positive class. **recall:** Recall is used to measure the fraction of positive patterns that are correctly classified **fscore**: In statistical analysis of binary classification, the F-score or F-measure is a measure of a test's accuracy.

Note: TP-> True Positive FP-> False Positive TN->True negative FN-> False Negative

1. Model Storing: The model was stored as a pickle file using the joblib library in python..

**VI. Perspective API:** This is a toxic comment classifier that was developed by google team it was also used to return a score ranging between 0-1 which signify the degree of toxicity of the text.

# 3.3 Software Methodology

The waterfall model is a breakdown of project activities into linear sequential phases, meaning they are passed down onto each other, where each phase depends on the deliverables of the previous one corresponds to a specialization of tasks. The approach is typical for certain areas of engineering design.

**Requirement Analysis**

**Testing**

**Implementation**

**System Design**

**Maintenance**

**Deployment**

Figure 3.1: Waterfall Model adopted for the Research

# 3.4 Development Requirement

This section consists of the tools and the technology to be use in developing the proposed system

# 3.4.1 System Design Technology

A web application of a Blog page will be design in which people will be able to post on different topics but the system is smart enough to detect and prevent Toxicity, the web app will be develop using the following tool and technology.

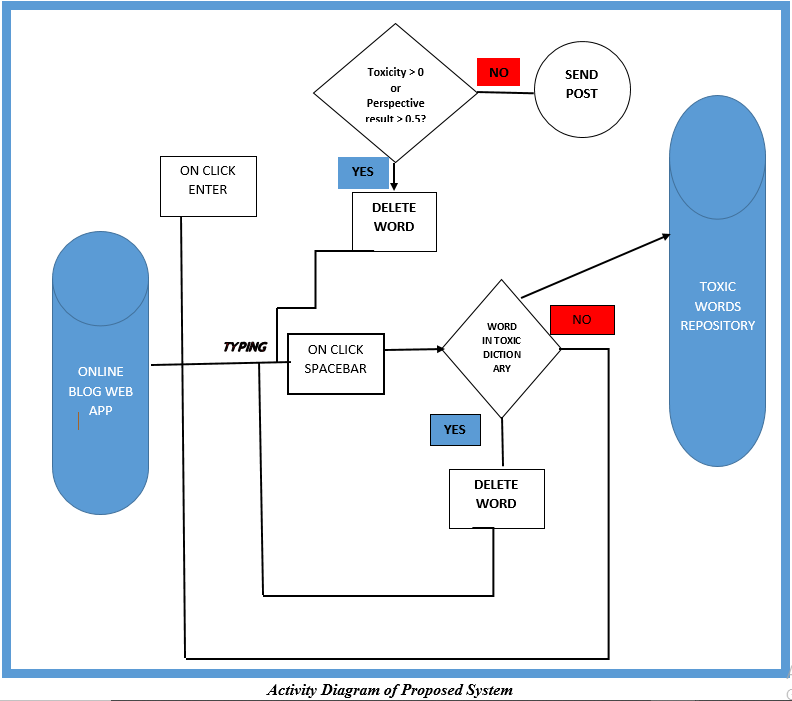
1. Python: is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation.Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.  
   Guido van Rossum began working on Python in the late 1980s as a successor to the ABC programming language and first released it in 1991 as Python 0.9.0.Python 2.0 was released in 2000. Python 3.0, released in 2008, was a major revision not completely backward-compatible with earlier versions. Python 2.7.18, released in 2020, was the last release of Python 2.  
   Python consistently ranks as one of the most popular programming languages.
2. 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. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.  
   Applications that use the Flask framework include Pinterest and LinkedIn.
3. Sqlite: is a database engine written in the C programming language. It is not a standalone app; rather, it is a library that software developers embed in their apps. As such, it belongs to the family of embedded databases. It is the most widely deployed database engine, as it is used by several of the top web browsers, operating systems, mobile phones, and other embedded systems.  
   Many programming languages have bindings to the SQLite library. It generally follows PostgreSQL syntax, but does not enforce type checking by default
4. 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.
5. JavaScript:  often abbreviated as **JS**, is a programming language that is one of the core technologies of the World Wide Web, alongside HTML and CSS. As of 2022, 98% of websites use JavaScript on the client side for webpage behavior, often incorporating third-party libraries. All major web browsers have a dedicated JavaScript engine to execute the code on users' devices.  
   JavaScript is a high-level, often just-in-time compiled language that conforms to the ECMAScript standard.[10] It has dynamic typing, prototype-based object-orientation, and first-class functions. It is multi-paradigm, supporting event-driven, functional, and imperative programming styles. It has application programming interfaces (APIs) for working with text, dates, regular expressions, standard data structures, and the Document Object Model (DOM).
6. HyperText Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser. It is frequently assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript.  
   Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.  
   HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page.

# 3.5 Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actionswith support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e., workflows), as well as the data flows intersecting with the related activities. Although activity diagrams primarily show the overall flow of control, they can also include elements showing the flow of data between activities through one or more data stores

Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system.

Below is the activity diagram of the system:



# CHAPTER FOUR

# SYSTEM DESIGN AND IMPLEMENTATION

# 4.0 INTRODUCTION

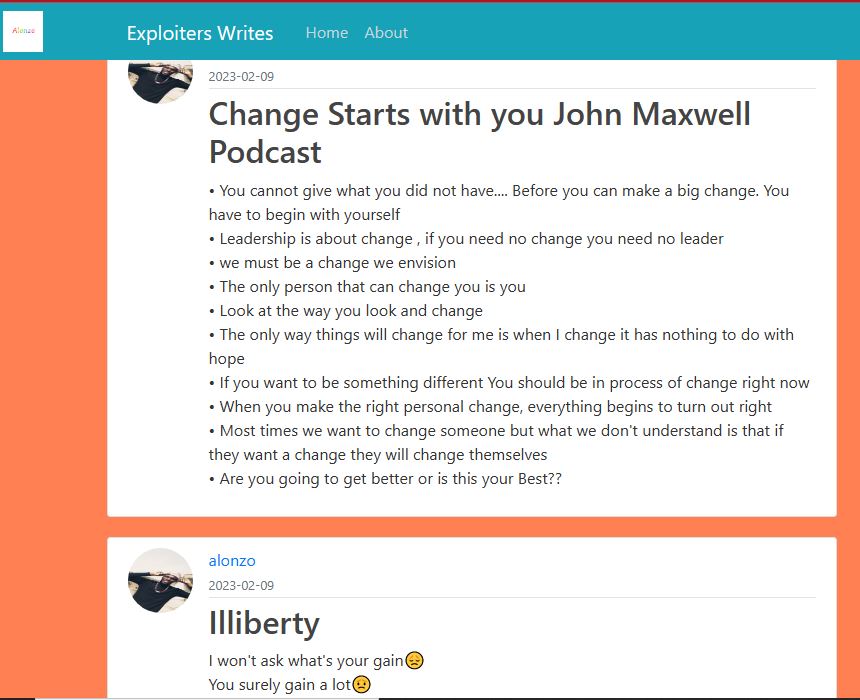
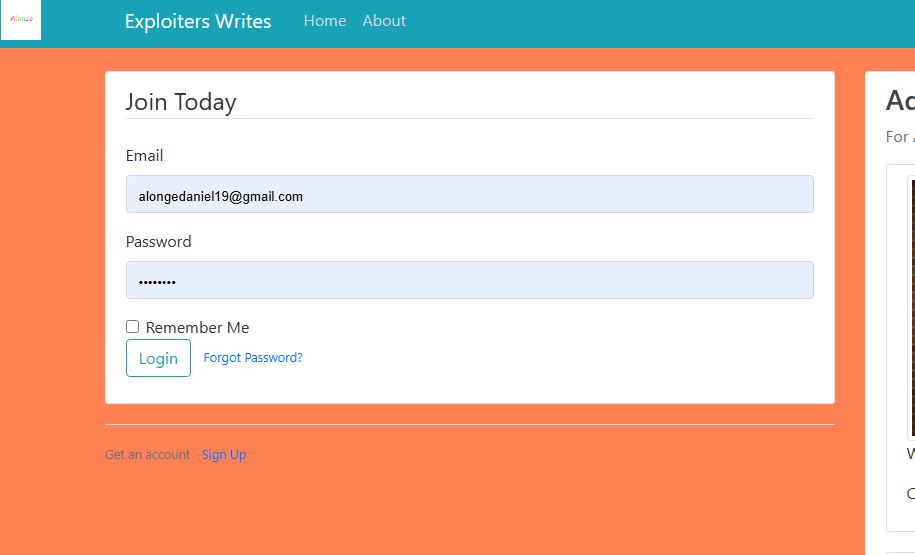
This is the software development stage and it is based on the user requirements and the detailed analysis of a system that was analysed on from the system analysis. Basically this system design is converting the requirements into a tangible reality. Hence, the system designed must meet user’s requirements. In this chapter, we will present the results and analysis of our research on developing a blog web app to prevent the use of toxic language in the community. The research aimed to create a tool using JavaScript responsiveness to prevent users from using predefined toxic words that were gathered from reliable sources. Additionally, we developed a toxic classifier using the combination Logistic Regression algorithm and the Google Perspective API to prevent toxic content that was copied and pasted from getting access to the page.

# 4.1 SYSTEM DESCRIPTION

After thorough research on how a blog page works we then after develop a blog page called the exploiter page, although it’s not as sophisticated as the blog websites in the real world scenario but it’s suitable and large enough for the purpose of this research.

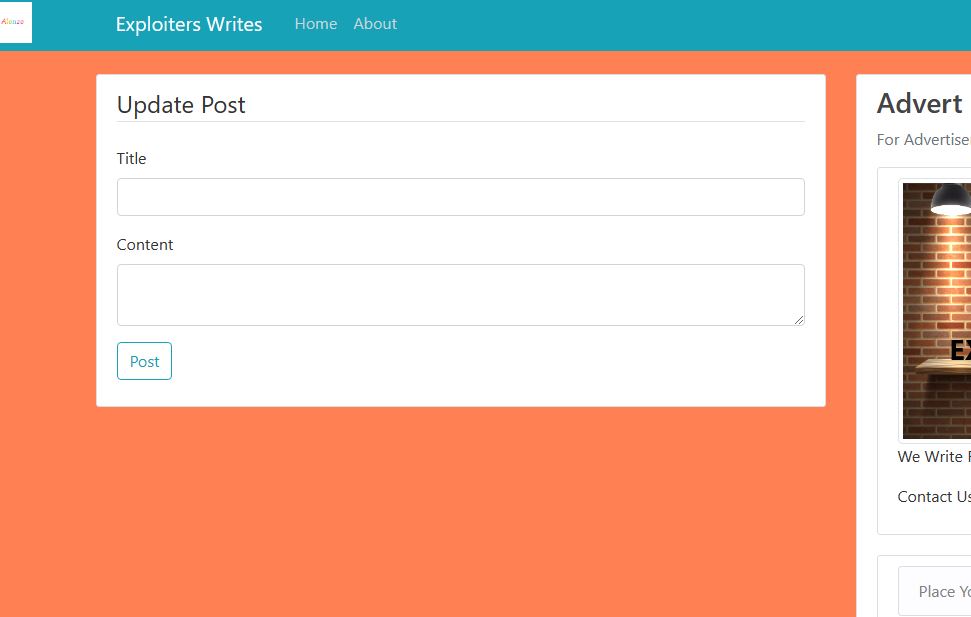
The blog post as mentioned in chapter 3 was developed using flask a python micro web development framework together with the traditional web languages which includes HTML, CSS and JAVASCRIPT.

The Main pages of the blog page includes:

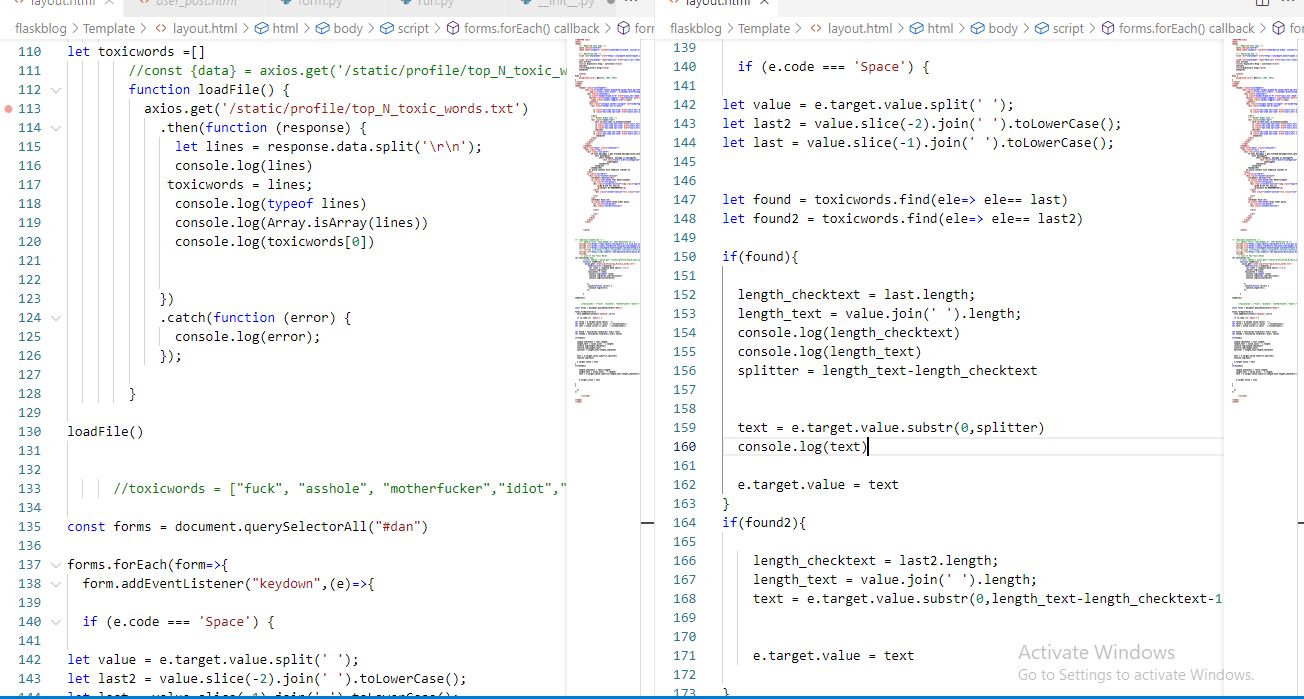
1. **Homepage:** This is the first page of the blog post where the post of all users are displayed which can be access by a register and unregistered users 
2. **The Login/Sign Up Page:** Where the user signup and login  
   
3. **Post Page:** This is the page where the user write and post contents and the blog page and can only be access by registered users alone.
4. **The Profile Page:** This is the page where the users make edits on their profile e.g. profile pictures, name and password

# 4.1.1 INPUT DESCRIPTION

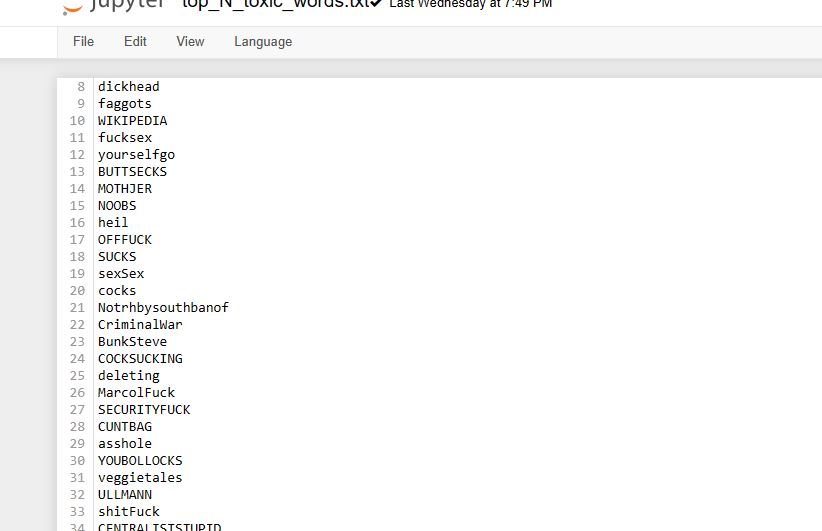
The system takes input from the post page section of the blog which contains a form with two field the title field and the input field. In this section of the system user input the title of their content and input the body which will be check by the system using the reactiveness of JAVASCRIPT to see if the user input is present in the predefined toxic words dictionary and remove it if it is.



*A Snapshot of the post page*



*Javascript code that check enhances the functionality of checking for user input in toxic dictionary*

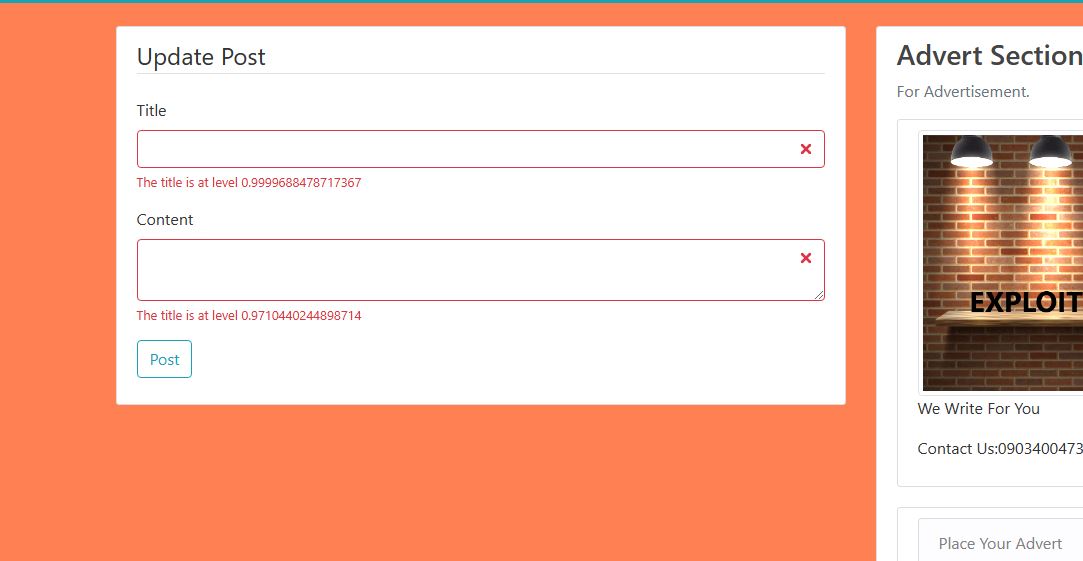


*The Predefined Toxic words*

# 4.1.2 OUTPUT DESCRIPTION

Once The user is able to submit the post or perhaps the user was able to manipulate their way in getting pass the first check for toxic word and the user click on post button a second check is performed on the user input by checking for the toxicity of the text using the logistic regression model built for text classification alongside with the google perspective API used for toxicity measurement.

If the sum of the probability score return by the logistic regression is greater than 0.5 and less than 0.7 the system will check if the perspective API score is above 0.6 and prevent the post but if the probability score of the logistic regression is greater than 0.7 the system prevent the post from getting cross to the website without checking the Perspective API score.



*Result of models*

# 4.2 Implementation Requirement

The model development requirements are into two main parts, software and hardware requirements.

# 4.2.1 Software Requirement

Software is a collection of programs or instructions written in any computer language, which enables flexibility to do whatever the user wants. This package can only be run and developed with the following minimum software

Minimum OS: windows 7

Database: SQLITE

1. Software Requirements
   1. Visual Studio
   2. Anaconda
   3. Flask APP

# 4.2.2 Hardware Requirement

The major hardware requirements for the smooth execution of the system are as follows;

Minimum Ram: 4 GB

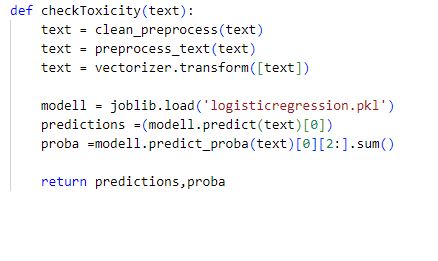
Minimum Hard disk: 30GB

A mouse or mouse sensitive used on laptops

Keyboard

Uninterrupted power supply (UPS)

Processor (Preferably Core i3 above)

****

***Check toxicity function***

****

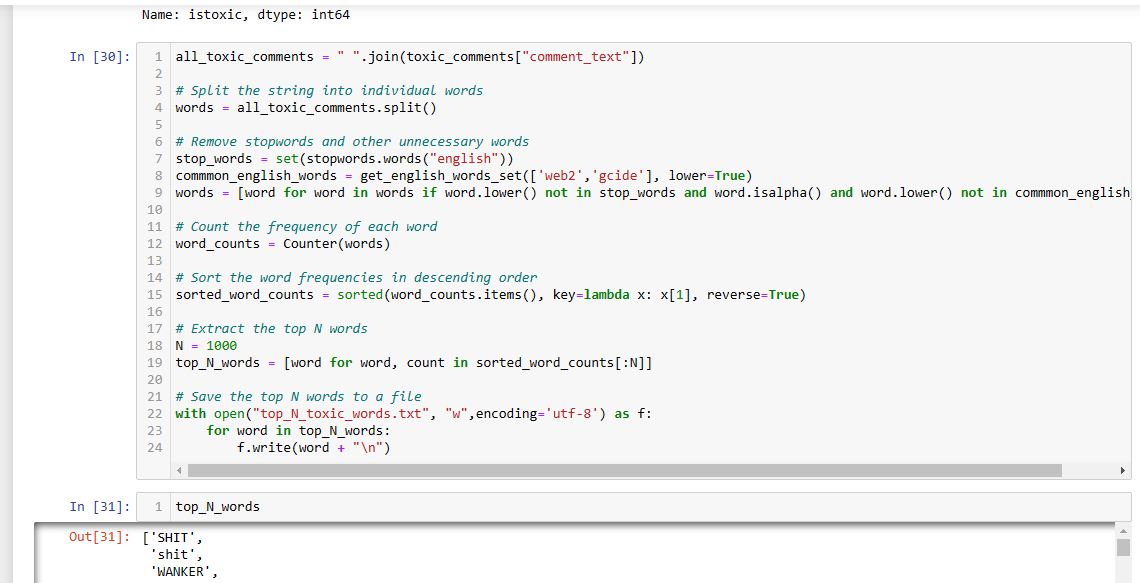
***Model Building***

****

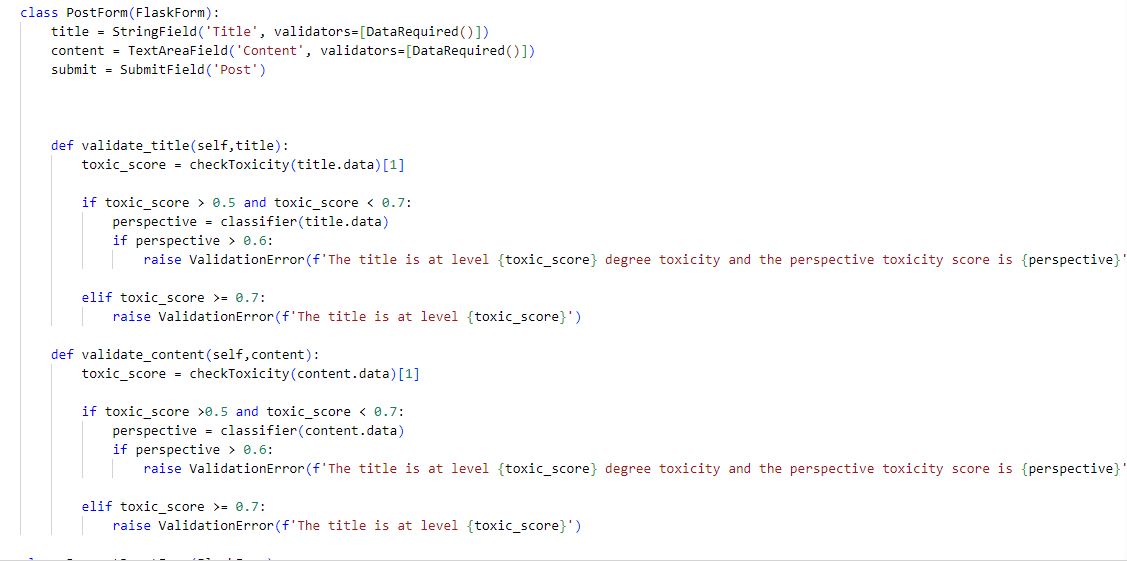
***Storing Model***

****

***Perspective API***

****

***Code to filter out the top toxic words***

****

***Validating User Input code***

# CHAPTER FIVE

# SUMMARY, CONCLUSION, RECOMMENDATION

# 5.0 INTRODUCTION

This chapter presents the conclusion of the research work carried out on developing a writing assistant tool to address the issue of toxicity in online communication. The objectives of the research were achieved, and the outcomes were presented in chapter 4. This chapter discusses the key findings, contributions, limitations, and future works of the research.

# 5.1 SUMMARY

The Ever increasing toxicity in online communication has given rise to the need of writing assistant tools that help prevent the use of toxic words in online communication in real time. Writing assistant tools are tools that help check for a specified error in writing and help in correcting or preventing them. AI writing assistants have come a long way since a Stanford student created the first spell-check software in the early 1970s. Since then, spell check, grammar checking, and autocorrect functionality have become commonplace, and people expect artificial intelligence to help them in their daily lives.

This study aimed to develop a writing assistant that help prevent the use of toxic language in online by developing a dictionary of toxic words and phrases, and then developing an agent to sense toxic constant once keyed and prevent it. To achieve this aim, the study employed a dictionary-based approach in developing the writing assistant tool.

The study found that toxicity is a major issue in online communication, and there is a need for tools that can help reduce its impact. The writing assistant tool developed in this study is a step towards addressing this issue. The tool was designed to prevent the use of toxic words and phrase when a user tries to. The study also found that software agents can be useful in developing writing assistant tools.

In developing the writing assistant, the model for text classification and online forum for the exhibition the following tools was used, Visual Studio code, Flask(Python web frame work), Javascript, html, css, bootstrap e.t.c.

**Limitation:** The research has some limitations that should be noted. First, the tool developed in this research is limited to detecting toxic language in English only. Second, the tool may not be effective in detecting certain types of toxic language, such as sarcasm or irony. Third, the tool relies on a dictionary-based approach, which may not be effective in detecting new or evolving forms of toxic language. Fourth, the tool may generate false positives or false negatives, which may affect its effectiveness.

**Contribution:** This research makes several contributions to the field of online communication and writing assistant tools. First, it provides a solution to the problem of toxicity in online communication, which is a significant issue in today's society. Second, it demonstrates the effectiveness of a dictionary-based approach in detecting toxic language in real-time. Third, it contributes to the growing body of knowledge on writing assistant tools and their applications in different domains.

# 5.2 CONCLUSION

In conclusion, this study developed a writing assistant tool that can assist in reducing toxicity in online communication. The tool was designed using a dictionary-based approach and prevent the successful entry of words when a user inputs a toxic word or phrase. The study found that software agents can be useful in developing writing assistant tools and that further research is needed to develop more sophisticated tools that can detect and suggest alternative phrases for toxic language.

# 5.3 RECOMMENDATION

Based on the findings of this study, the following recommendations are made:

* Further research should be conducted to develop more sophisticated writing assistant tools that can detect and suggest alternative phrases for toxic language.
* More research is needed to investigate the effectiveness of writing assistant tools in reducing toxicity in online communication.
* Future studies should explore other approaches to developing writing assistant tools in real time, such as machine learning and natural language processing.
* Online platforms should encourage the use of writing assistant tools by integrating them into their platforms to help users communicate more effectively and reduce the use of toxic language.

# REFERENCES

Ben Miller. (2018). Notoriously Toxic: Understanding the Language and Costs of Hate and Harassment in Online Games [Working Paper].

Daniel, I.O.A. (2013). Communication as socio-cultural meaning exchange. The example of Richard Wright’s Black Boy. International Journal of Applied Linguistics and English Literature, 2(5), 173-177.

Dang, Y., Zhang, Y., & Chen, H. (2010). A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. IEEE Intelligent Systems, 25(4), 46-53.

Hafner, K., & Lyon, M. (1996). Where wizards stay up late: The origins of the Internet. New York: Simon & Schuster.

Harnad, S. (1991). Post-Gutenberg galaxy: The fourth revolution in the means of production and knowledge. Public-Access Computer Systems Review, 2(1), 39-53.

Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). Deceiving Google’s Perspective API built for detecting toxic comments. arXiv preprint arXiv:1702.08138.

Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu, & Amy Bellmore. (2012). Learning from bullying traces in social media. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 656-666).

Kelm, O. (1992). The use of synchronous computer networks in second language instruction: A preliminary report. Foreign Language Annals, 25(5), 441-454.

Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, & Ritesh Kumar. (2019). Predicting the type and target of offensive posts in social media. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 1415-1420).

Meskill, C., & Krassimira, R. (2000). Curriculum innovation in TEFL: Technologies supporting socio-collaborative language learning in Bulgaria. In M. Warschauer & R. Kern (Eds.), Network-based language teaching: Concepts and practice (pp. ??-??). New York: Cambridge University Press.

Merriam-Webster Dictionary. Retrieved from <https://www.merriam-webster.com/dictionary/>

Nwana, H., & Ndumu, D. (1996). A brief introduction to software agent technology. Unicom seminar on "Real-world applications of intelligent agent technology," London UK, 6/11-13/96, 278-292.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Vol. 10 (pp. 79-86).

Patton, D. U., Eschmann, R. D., Elsaesser, C., & Bocanegra, E. (2016). Sticks, stones and Facebook accounts: What violence outreach workers know about social media and urban-based gang violence in Chicago. Computers in Human Behavior, 65, 591-600.

Phillips, W. M. (2012). The impact of new media technologies on social interaction in the household