Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 8 No. 3, November 2023, pp. 331 – 340
DOI:

Implementation Chatbot on Discord for Information Assistance and Conflict Prevention

Zudha Pratama*[1], **Ery Mintorini**[2]
[1,2]University of Dian Nuswantoro, Penanggungan 41 A Street, Bandar Lor, Mojoroto, Kediri, Indonesia
E-mail : zudhapratama@dsn.dinus.ac.id[1], ery.mintorini@dsn.dinus.ac.id*[2]
**Corresponding author*

Karmila[3], **Didik Hermanto**[4]
[3,4]University of Dian Nuswantoro, Penanggungan 41 A Street, Bandar Lor, Mojoroto, Kediri, Indonesia
E-mail : karmila@dsn.dinus.ac.id[3], didik.hermanto@dsn.dinus.ac.id*[4]

Abstract - Discord, which was originally created for the gamer community, can now be found used by hobby groups and communities that are used for shared learning purposes. But the downside is the gamer culture that comes with it. Rude and toxic words that are synonymous with the gamer community should be avoided in study group communities. Meanwhile, the facilities for minimizing harsh and toxic words are still limited to word filters that can be tricked so that they can still be sent to the chat room. This can trigger conflict and interfere with learning activities together. This paper proposed an information assistance chatbot that is able to answer question, and conflict prevention with detection toxic sentences using pre-processing from NLP (Natural Language Processing) and text classification so that the chatbot is able to limit toxic sentences a little more accurately than the word filter feature alone. Also, Chatbots are given the ability to determine the value / level of toxic conversations so that they are had been able to determine the punishment action to be carried out by warning, suspending, or even being issued for the most severe cases. In addition, by looking at the frequency of sending messages from several senders, which indicates toxic, it was able to determine when the conflict occurs. The result shows that chatbot can work fine to answer question and detecting toxic include do punishment to toxic sender. With 10% error on detecting conflict and 30% error on answer question. That 30% error false positive on make an answer that should not be answered.

Keywords – chatbot, string matching, classification

1. INTRODUCTION

Discord was created as an alternative to chat applications that did not provide a satisfying experience for gamers in 2015. The use of the Discord platform has reached 150 million in 2021. Discord is very popular because of chat channels that provide user role control, more varied stickers, leveling experience and several other excellent features that support the activities of game fans and other hobbies. However, behind a series of superiors, there are unfavorable facts in Discord written in a study, stating that Discord poses a risk of bringing the community towards de-contextualization, inter-textual racism and a shift in community norms towards toxicity [1]. This risk will lead to conflict if not handled properly. One solution to

overcome this conflict problem is to use a chatbot. Chatbot is applied as a computer mediation to replace humans because of its high efficiency and cost effectiveness [2]. The intended mediation is to mediate in the process of communication between forum members. For example in Discord, every message sent by forum members is checked and if there is any indication of being dangerous it will be warned and deleted. However, the currently available chatbots are still only limited to filtering bad words. Not yet able to distinguish between harsh but non-toxic words that do not cause conflict, or words that are not harsh but contain toxic meanings that do not cause conflict.

Rude words are words that are considered impolite or inappropriate to use in certain situations, such as in formal conversation or in front of children. Examples of harsh words include dirty words or words that contain swear words. Meanwhile, the word toxic refers to words that demean, insult, or hurt the feelings of others. Typically, the word toxic is used in an online context, such as in comments or messages on social media. Toxic words can spark arguments, online bullying, or exacerbate an already sensitive situation. A scientific article that discusses sensitive words, toxic words and politically charged words, explained that there are theoretical indicators to understand how netizens interact with filter systems or Internet censorship systems. Censorship from artificial intelligence and human intelligence influences online chat on Chinese social media [3]. It was also found that users are also creative to circumvent the filter system by recoding words to pass through sensitive word sensor filters which can be seen as a growing digital culture practice [4].

In the research that was carried out this time, a chatbot was built that was able to answer some of the problems above. Related to how to deal with indications of conflict and harsh words that are not necessarily toxic or not harsh but toxic. Regarding the filtering system which is not easily fooled by rewriting it in other words but the meaning is still dangerous. As well as chatbots for mediation and information facilities in online communities on Discord. The chatbot that will be developed uses the Chatbot API Discord which will be written in Python. In making the chatbot, the researcher will use a technique to parse text messages sent to the chat room and then classify them. Is it a question related to information or is it a dangerous toxic statement. In addition, it also calculates the activity frequency of the message sender in the chat room to be an additional parameter for indications of the start of a conflict.

The sub-discussion of previous research studies is used to review and study research results from researchers who have topics of discussion that have similarities with current research, namely chatbots. Lalwani in his article use an algorithm to check sentence similarity (NLP) is applied to the modified input to check its similarity with the questions of a predefined question-set, whose answers are available to make chatbot. [5] Next, Lee [6] in a journal entitled Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional, designs a Chatbot as a mediation for one's self-disclosure about personal experiences, thoughts and feelings. Because it is very important for mental health professionals to understand a person's mental status. Next Zhou [2] compares chatbots vs. human agent, on the quality of the anticipated communication and the mechanisms underlying it. He wrote in a journal entitled Talking to a bot or a wall? How chatbots vs. human agents affect anticipated communication quality. Nguyen [7] in his journal entitled NEU-chatbot: Chatbot for admission of National Economics University introduces an artificial intelligence-based admissions chatbot for campus admission activities that handle students so they can immediately get daily curriculum updates, new student admissions, school fees, etc. Pawlik [8] through a journal he wrote entitled A method for improving bot effectiveness by recognizing implicit customer intent in contact center conversations, developed a chatbot method to increase bot effectiveness by identifying customer intentions implicit in the conversation history of the contact center.

The previous research above has discussions related to the use of chatbots in disputes or conflicts, mediation facilities, the quality of anticipated communication or communication that has hidden statements, then chatbots as an information center, as well as identifying intentions based on conversation history, which is related to the process of recognizing toxic statements based on conversation history which was reported. The proposed chatbot is a chatbot that is also able to answer questions and act as a mediator to prevent conflict through detecting toxic expressions.

2. RESEARCH METHOD

2.1 Definition of Chatbots

Asbjørn Følstad and Petter Brandtzaeg say that virtual assistants or chatbots are software agents that interact with users through natural language conversations [9]. With natural language conversations that involve Natural Language Processing, Chatbots are able to provide a more personalized and efficient human interaction experience, as well as increase effectiveness and efficiency in various fields, such as customer service, e-commerce, and digital marketing. Chatbot is a technology that is increasingly important and growing rapidly in the future.

2.2 Definition of Mediation

In language, the meaning of mediation in KBBI is defined as the participation of the mediating side in resolving disputes as a consultant. In Mediation there are three elements, the first element is the process of resolving disputes between two or more parties to the litigation. The second element, in the settlement of disputes, parties involved from outside the dispute. And the third element, the party involved in settling the case acts as an advisor, he does not have any authority over the decision maker. Mediation is "to meditate" or "be in the middle" [10].

2.3 Bad word (rough words)

Badwords or harsh and impolite words are often associated with the use of inappropriate or demeaning words. The use of harsh and impolite words not only harms the person being targeted, but can also reflect a lack of ethics and courtesy on the part of the person speaking them. The use of harsh and impolite words can also have an impact on communication and interaction between individuals. The use of harsh words can trigger conflict, hinder the problem-solving process, and damage relationships between groups. Badwords are words or phrases that are generally considered blasphemous, vulgar or offensive. It is also called swearing, swearing. Sometimes these words don't just come out when people are angry or upset about something or someone. However, swearing is often displayed in several mass media such as magazines, newspapers, advertisements and films; this is quite common, although sometimes it is still considered taboo and deviates from social norms. [11]

2.4 Toxic Behavior

Toxic is a term often used to describe behavior or attitudes that harm others, especially in interpersonal relationships. Over-criticism: Having constructive and constructive criticism is a good thing, but over-criticism and often belittling or hurting others can be toxic behavior. Here are some types of toxic behavior:

1. Manipulation: Manipulating others by controlling information or situations for personal gain can cause discomfort and harm to others.

2. Bullying or bullying: Constantly hurting, threatening, or making fun of others can cause significant emotional and psychological distress and can have long-term effects on mental health.
3. Gaslighting: Manipulating a person's belief in reality to make him doubt himself or feel crazy, often with the aim of gaining control or gain.
4. Excessive negativity: Constantly and for no reason spreading gossip or speaking ill of others can create an unhealthy environment and damage interpersonal relationships.
Mitchell Kusy and Elizabeth Holloway in their book [12] entitled Toxic Workplace!: Managing Toxic Personalities and Their Systems of Power, also have their own views on the meaning of toxic. In short, according to them, toxic is a counterproductive pattern that weakens a person, both individually and in a group, and can even occur in the long term.

2.5 Discords

Discord is an online communication platform that was originally designed for gamers, but is now being used by communities of all kinds. This platform provides features such as text chat, voice chat, and video chat, as well as the ability to share files and screen sharing. Discord allows users to create and join servers, which are basically virtual spaces where people can talk and collaborate. On the server, users can create channels for specific topics and invite others to join. Discord also offers a wide variety of customization options, including the ability to add bots that can perform functions such as playing music, moderating servers, and providing information.

Discord has features that are simple, practical, attractive, and easy to use. This platform can be accessed through various devices. Also, users do not need to have an account to join the main chat channel, depending on the channel owner's policies. To join the channel, users only need to open the invitation link (Instant Invite) generated by the channel owner and shared with them. Users who are already in the channel can also share the invite link. In addition, users can set how long the link will be valid, how many times it can be opened, and whether the link is temporary members only or not. There are also options to make the links easier or more difficult to access [13].

We can also create a chatbot for Discord using the Discord Bot API. Here are the general steps for creating a chatbot using the Discord Bot API. First, Create a Discord developer account and create a new Discord app at https://discord.com/developers/applications/. Then create a new bot on the app page and set bot permissions according to your needs. Next create a bot token and store it safely. Download and install the Python library discord.py or another library that supports building Discord bots. Write your bot code following the Discord Bot API documentation. Finally run your bot code and test your bot by joining the Discord server and calling bot commands.

2.6 NLP (Natural Language Processing) and Pre-processing

NLP (Natural Language Processing) is a field of science that focuses on understanding and processing human natural language by computers. NLP is a branch of Artificial Intelligence and Linguistics that focuses on enabling computers to comprehend written or spoken human language expressions [14]. One part of the NLP techniques involves an initial preprocessing step, which is a vital phase. Data preprocessing encompasses various operations, including the elimination of stop words (common words in a language), punctuation removal, part-of-speech tagging, and the conversion of abbreviations into full words or phrases to facilitate comprehension. Preprocessing is for data cleaning for the modeling stage. its to identify and filter out (1) sentences in languages other than the target language, (2) linguistically unconnected and/or corrupted sentences, and (3), duplicate sentences [15].

3. RESULTS AND DISCUSSION

The researchers get the problems that arise in the discord group through literature studies and experience when joining the gaming community on discord. The following are the problems the researchers found and the formulation of the functional requirements for chatbot capabilities in Table 1.

Table 1. Found Problems and Functional Requirements

Found Problems	Functional Requirements
Questions for new members of the community.	The bot functionality of answering about the rules of the game in the community, community history, community structure and other information related to community introduction
Posts that contain indications of the beginning of a conflict	The bot functionality of determining the status of the chat room based on message text and also the frequency of messages.
Posts that contain indications of toxic statements	The bot functionality of classifying messages whether toxic or not based on the history of reported toxic sentences
Problem member.	The bot functionality takes punishment action by reminding, temporarily stopping and expelling members who are indicated to be problematic.

From the functional requirements above, the researcher compose the algorithm by utilizing several string-matching processes, checking the frequency of messages from log messages, using pre-processing of Natural Language Processing that combined with classification, weighting values from classification process to decide the punishment. The following are the steps to compose the algorithm in creating a chatbot that has the capabilities according to the functional requirements above.

3.1 Chatbot basic ability to identifying *question for asking information*

Basic chatbot capabilities, which involve identifying questions for asking information. This capability is achieved through of string-matching techniques combined with the Levenshtein Distance algorithm. Before that message will be pre-processed first. This process was adopted from Natural Language Processing (NLP) preprocessing that contain sub-process tokenizing, lowercasing, removing punctuation and stop words. Then, text matching allows the chatbot to compare user input to pre-saved questions along with their answer pairs. The Levenshtein Distance algorithm measures the similarity between two strings, enabling the chatbot to judge the proximity of the user's input to the existing question text. By leveraging this technique, chatbots can differentiate statements in sent messages and increase their capacity to provide appropriate responses.

3.2 Determine when conflict indicate occur by frequency of messages

Detecting the occurrence of conflicts involves analyzing and monitoring message frequency within the message log. Concurrently, the frequency of messages in the log is tracked over time. When an unusual spike in message frequency aligns with a notable increase in detected toxic language, it can indicate the onset of a conflict. This integrated approach leverages both linguistic analysis and communication patterns, providing a more comprehensive understanding of conflict.

3.3 Detect toxic with classification

Detecting toxic language in text using classification algorithm like Naive Bayes is trained on labeled toxic and non-toxic text samples. Imbalanced data can be addressed, and the Naive Bayes model can be deployed for predictions after preprocessing new input text. Continuous monitoring and model updates are necessary due to the evolving nature of toxic language.

3.4 Decide the punishment

Evaluating the extent of toxicity within content is crucial in determining suitable penalties. By gauging the harmful nature of the content, appropriate actions can be taken. Factors such as the language used, intent, and potential impact are taken into account. These assessments can lead to a tiered system where content is categorized into different levels of toxicity. Each level corresponds to a predefined punishment, with more severe penalties applied to higher levels of toxicity. This method ensures that consequences are proportionate to the harm caused by the content, creating a balanced and effective system for maintaining positive online interactions.

3.5 Design the Algorithm

To understanding what will be applied in the chatbot that will be made, consider Figure 1 below.

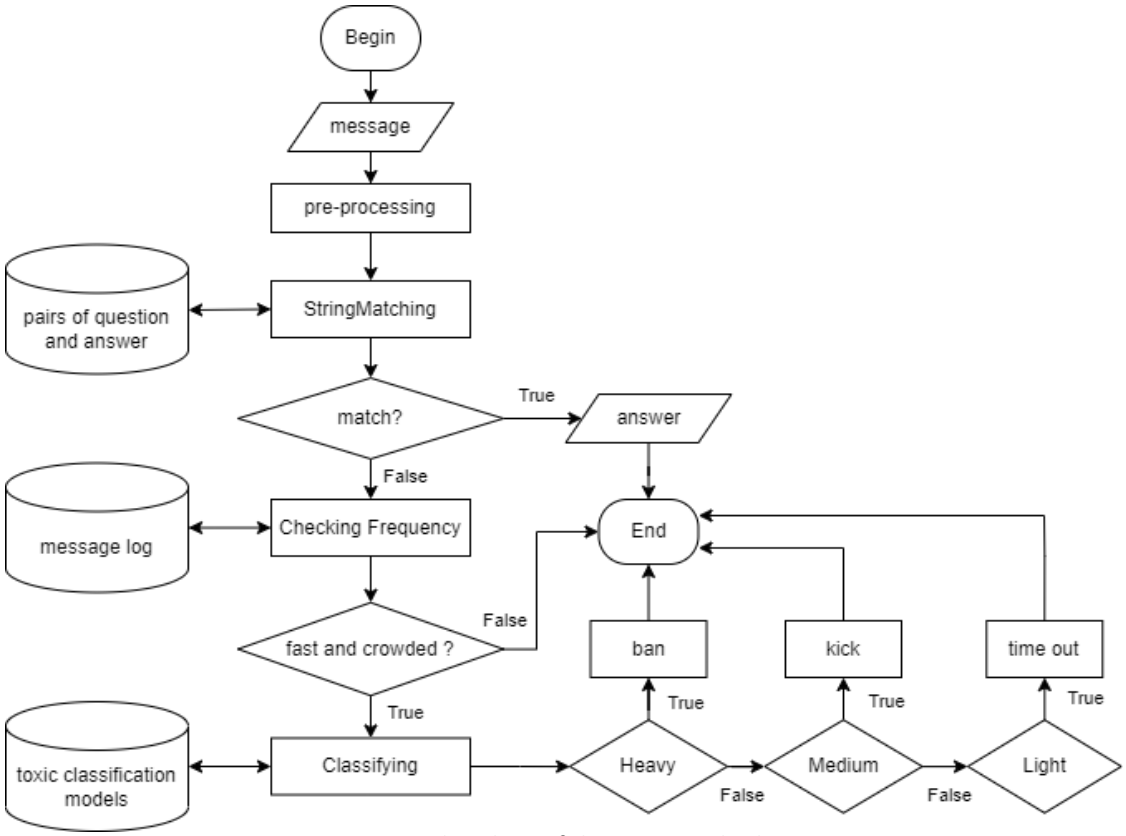

Figure 1. Flowchart of the propose chatbot.

The process starts from an input of a message that is sent by user to the Discord. Then the chatbot will do the pre-processing process which consists of tokenizing, lowercasing, removing punctuation. Then, the next process is string matching using the Levenshtein distance between the message and the question in datasource that contain pair of question and answer. If it is found or match the questions, the chatbot will provide a response output an answer to these questions based on the datasource. If it doesn't match, process will be continued to frequency checking process. it will do based on the message log entered in the chat room. If message log is still within a reasonable frequency, then it will be finished without to any actions. But, if it within a certain time the message is sent too quickly by several users, then it will continue to the toxic classification processing. The Toxic Classification process will compare the message with the data model in the datasource. This will generate a classification value that is

heavy, medium, or light toxicity. If the classification result is heavy toxic, it will be banned, if it is indicated to be medium toxic, it will be punished by a kicked and if it is lightly toxic, it will be punished by a timeout. Ban is a punishment for being kicked out and not allowed to re-enter the group or blacklisted, kick is a punishment for being kicked out but you can still rejoin, and timeout is an action to temporarily stop access to sending messages to the chat room.

3.6 Implementation

This phase will begin with write all process to python code on replit. Then registering an account to discord developer page for get the API Token. Include adjust the bot permission configuration for bot access rights to the API bot feature. API Token will be used to connect between the python code on replit and chatbot application on the developer's discord page. Figure 2 is the process of registering API Token for chatbot that we name it CoMeBo (Conflict Mediator Bot).

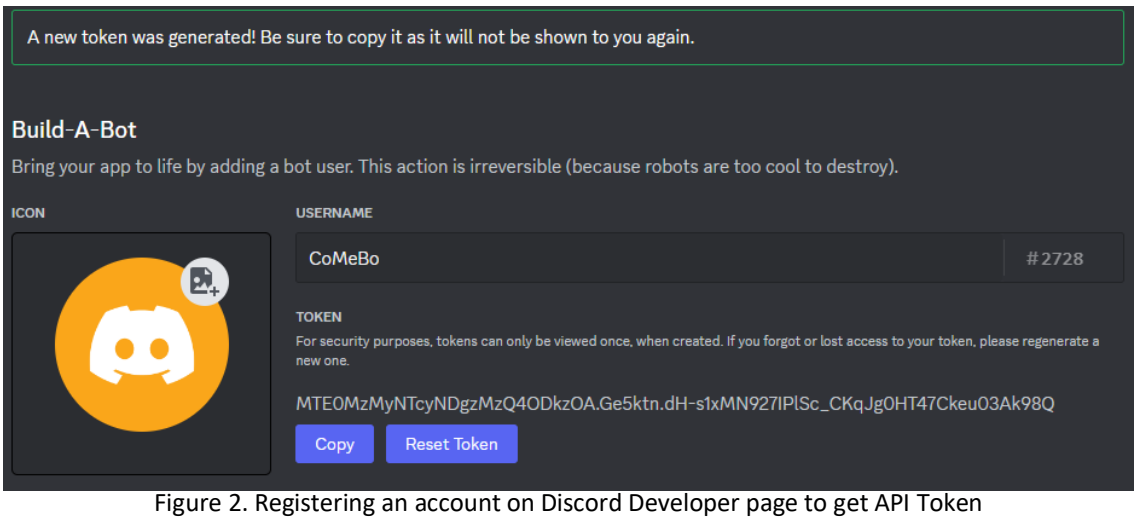

Figure 2. Registering an account on Discord Developer page to get API Token

Next prepare the json file as datasource to hold the question-and-answer pair data and hold the message log. The message log contains the sending time and the sender's name which is used to determine the frequency with which messages are sent. Training examples of toxic sentences and labeling it to classify whether heavy, medium, or light toxic.

3.7 Testing and Analysist

Scenario testing is simple, it is carried out to test each functional requirement running as it should. But we toke two scenarios because some functional requirements were checked together at the same time. First, Testing Identifying Question, Test the string-matching function with make some question-answer on datasource then make message which similar with question testing string 10 times. Second Test, Conflict Detection, Classify Toxic and Deside Punishment Test were do in one time. Researcher sending messages from 3 accounts, 10 messages per account with different sending intervals to test counter frequency function. Then testing the result classifying when conflict detected. And the last is checking result of deciding punishment.

Table 2. Testing String-Matching Function

Testing Code	Test Case	Similarity (%)	Chatbot Response	Conclusion
A1	Question is in datasource	92	Answer	True

A2	Question is in datasource	93	Answer	True
A3	Similar object question but different question word	74	Answer	True
A4	Similar object question but different question word	79	Answer	True
A5	Different object question but similar question word	47	Not Answer	True
A6	Different object question but similar question word	79	Answer	False Positive
A7	Not a question, but some words similar	78	Answer	False Positive
A8	Not a question, but some words similar	73	Answer	False Positive
A9	Not a question and no words are similar	30	Not Answer	True
A10	Not a question and no words are similar	46	Not Answer	True

Table 2 above shows that question detection system using string matching that is able to run and generate answers for every message that should be answered. However, there are several messages that are not questions that should not be answered, but are answered by chatbots because of their high percentage of similarities. There are 30% false positives occurs.

Table 3. Testing Frequency Counter, Classifying, and Punishment Function

Testing Code	Test Case	Function Result			Conclusion
		Counter	Classifying	Punishment	
B1	2 user message each other with normal time intervals	Normal	-	-	True
B2	3 user message each other with normal time intervals	Normal	-	-	True
B3	2 user message each other with fast time intervals	Normal	-	-	True
B4	3 user message each other with fast time intervals	Crowded	-	-	True
B5	2 user messages with normal time intervals, one user fast time	Crowded	-	-	True
B6	3 user send messages to each other at normal time intervals, and one user uses word toxic	Normal	-	-	True
B7	3 user send messages each other with fast time intervals, and one user uses word toxic	Crowded	Conflict	-	False
B8	3 user send messages to each other with fast time intervals, and some user use light toxic words	Crowded	Conflict	Timeout	True
B9	3 user send messages to each other with fast time intervals, and some user use medium toxic words	Crowded	Conflict	Kick	True
B10	3 user send messages to each other with fast time intervals, and some user use heavy toxic words	Crowded	Conflict	Ban	True

Testing above shows 90% success. The failure occurs in the code B7 scenario when a message containing a toxic sentence is sent by one user while several other users are sending messages quickly. In design it should be if only one person is toxic then it should be left alone, because the system aims to prevent conflict, not just toxic filters. for example, in the code B6 scenario, the system lets toxic sentences enter when no one responds, even though the toxic system allows it because it does not cause conflict, the chat room can still be considered safe.

4. CONCLUSION

From the several experiments above, it can be concluded that the system created and applied to the chatbot is able to provide quite good results. The system for checking questions to Informastion Assistance who need chatbot information was able to answer properly and with

70% success. An error of 30% is obtained from the similarity value obtained from inappropriate questions, this can be corrected if you increase the question-and-answer database, because if there are more comparative data it will provide a closer similarity value so that other answers will not appear whose questions are indeed not recognized by the system.

Conflict Prevention System that consisting of crowded detection of the frequency of message delivery and classification of message content including to determine punishment. The system was able to run well with 90% success. 10% of failures stem from system failures to detect a combination of conflict conditions and the presence of toxic sentences. This needs to be reviewed regarding the conflict detection method using the frequency of sending messages whether it is relevant or not. For further research, it is necessary to formulate this problem and add test cases that may have never been tested or increase the number of test texts, for example each test case is repeated several times. Apart from adding test cases, future research needs to develop a chatbot mediator that is equipped with the ability to private message with people who are detected as toxic as if to provide an understanding regarding their personal mistakes. Apart from providing punishment for mistakes, we need to provide understanding regarding the mistakes.

REFERENCES

[1] E. K. Johnson and A. Salter, "Embracing discord? The rhetorical consequences of gaming platforms as classrooms," Comput. Compos., vol. 65, p. 102729, 2022, doi: https://doi.org/10.1016/j.compcom.2022.102729.

[2] Q. Zhou, B. Li, L. Han, and M. Jou, "Talking to a bot or a wall? How chatbots vs. human agents affect anticipated communication quality," Comput. Human Behav., vol. 143, p. 107674, 2023, doi: https://doi.org/10.1016/j.chb.2023.107674.

[3] X. Wang, K. Juffermans, and C. Du, "Harmony as language policy in China: an Internet perspective," Lang. Policy, vol. 15, no. 3, pp. 299–321, 2016, doi: 10.1007/s10993-015-9374-y.

[4] W. Ye and L. Zhao, "'I know it's sensitive': Internet censorship, recoding, and the sensitive word culture in China," Discourse, Context Media, vol. 51, p. 100666, 2023, doi: https://doi.org/10.1016/j.dcm.2022.100666.

[5] Lalwani, Tarun and Bhalotia, Shashank and Pal, Ashish and Rathod, Vasundhara and Bisen, Shreya, "Implementation of a Chatbot System using AI and NLP". International Journal of Innovative Research in Computer Science & Technology (IJIRCST), vol. 6, no. 3, 2018, doi: http://dx.doi.org/10.2139/ssrn.3531782

[6] Y.-C. Lee, N. Yamashita, and Y. Huang, "Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional," Proc. ACM Human-Computer Interact., vol. 4, pp. 1–27, 2020.

[7] T. T. Nguyen, A. D. Le, H. T. Hoang, and T. Nguyen, "NEU-chatbot: Chatbot for admission of National Economics University," Comput. Educ. Artif. Intell., vol. 2, p. 100036, 2021, doi: https://doi.org/10.1016/j.caeai.2021.100036.

[8] Ł. Pawlik, M. Płaza, S. Deniziak, and E. Boksa, "A method for improving bot effectiveness by recognising implicit customer intent in contact centre conversations," Speech Commun., vol. 143, pp. 33–45, 2022, doi: https://doi.org/10.1016/j.specom.2022.07.003.

[9] A. Følstad and P. Brandtzaeg, "Chatbots and the new world of HCI," interactions, vol. 24, pp. 38–42, 2017, doi: 10.1145/3085558.

[10] McDowell, P. Media and Mediation in the Eighteenth Century. Oxford Handbooks Online, 2017, 1–22.

[11] R. A. Maulidiatsani, “The Portrait Of Swearwords And The Social Background Of The Characters In The Breakfast Club Movie,” Lang. Horiz. J. Lang. Stud., vol. 3, no. 1, 2015.
[12] M. Kusy, Toxic workplace! managing toxic personalities and their systems of power, 1st ed. San Francisco: Jossey-Bass, 2009.
[13] Ayob, M. A., Hadi, N. A., Pahroraji, M. E. H. M., Ismail, B., & Saaid, M. N. F. , “Promoting ‘Discord’ as a Platform for Learning Engagement during Covid-19 Pandemic,” Asian Journal of University Education, vol. 18, no. 3. UiTM Press, Universiti Teknologi MARA, Jul. 31, 2022. doi: 10.24191/ajue.v18i3.18953.
[14] Khurana, D., Koli, A., Khatter, K. et al. “Natural language processing: state of the art, current trends and challenges”. Multimed Tools Appl, vol. 82, pp. 3713–3744, 2023. https://doi.org/10.1007/s11042-022-13428-4.
[15] A. Petukhova and N. Fachada, “TextCL: A Python package for NLP preprocessing tasks,” SoftwareX, vol. 19. Elsevier BV, p. 101122, Jul. 2022. doi: 10.1016/j.softx.2022.101122.