

Understanding Differences Between Human Language Processing and Natural Language Processing by the Synchronized Model

Yixin Wang^{1, †}, Shisen Yue^{2, *, †}, Yanyi Zhong^{3, †}

¹ School of International Education, Nanchang Hangkong University, Nanchang, Jiangxi, China

² School of Foreign Language, Shanghai Jiaotong University, Shanghai, China

³ School of English Studies, South China Business College University, Guangzhou, Guangdong, China

*Corresponding author. Email: ²lyw520@sjtu.edu.cn

[†]Those authors contributed equally.

ABSTRACT

Chat applications using Artificial Intelligence (AI) based on Natural Language Processing (NLP) platforms have been reported to be gradually accepted by people. This research aims to investigate differences between human language processing and Natural Language Processing (NLP) system, which is the core technology of most chat applications, using the synchronized language model. To achieve this objective, this research first distribute and collect questionnaires with questions such as the frequency and motivation of using AI chatbots among university students. The study then evaluate the selected chatbot with linguistic method and knowledge through semantics and pragmatics. Practically, this study proposes valid approaches to perfect existing chatbots. This study suggests that AI chatbots based on NLP can be applied to complete tasks but differ apparently from the human language processing system. The conclusion drawn from this study is that if the AI chatbot is developed to recognize misspelled words and their vocabulary is expanded, it will enhance the applicability of AI chatbots and fit them into people's lives.

Keywords: *Natural Language Processing, Human Language Processing, Artificial Intelligence, Linguistics.*

1. INTRODUCTION

Artificial intelligence chatbots based on Natural Language Processing have been receiving increasing popularity around the world. According to statistics conducted by Retale, near 60% of the Millennial generation have used chatbots, among which 70% have a positive user experience. However, some users of AI chatbots also provide feedback that difference between chatting with a chatbot and a person is obvious. Little research has investigated this kind of difference through linguistic perspective, essentially, the difference between Natural Language Processing that underlying chatbot and human language processing system.

Natural Language Processing (NLP) is strongly related to artificial intelligence. It is a branch of linguistics, computer science, and artificial intelligence, mainly focusing on how to program computers to process and analyze large amounts of natural language data ("Natural Language Processing", 2021) [1]. The

quality of language model is positively correlate to the similarity of their output texts to words people naturally form. The most explanatory method for presenting what happens within a Natural Language Processing system is utilizing the 'levels of language' approach. This is also referred to as the synchronized model of language. (Liddy, E.D., 2001) [2]. This model divides language into phonology, morphology, lexical, syntactic, semantics, pragmatics and discourse levels, among which we choose two metrics to evaluate the language processing of an selected chatbot named Replika, namely the semantics and pragmatics levels, so as to compare the difference between human language processing and Natural Language Processing.

Previous literature indicates that humans treat humanoid identity differently from humans. Such results coincide with our research data, in which we find that most people reflect that they easily identify the chatbot identity through chatting. But this contradicts a study on the media equation, in which the author argues

people treat nonhuman technological objects and mediated representations socially and naturally (Lee-Won, R.J., Joo, Y.K., & Park, S.G., 2020) [3].

The first approach of this study is to distribute and collect a questionnaire among university students who have used AI chatbots, including questions such as what aspects of the communication process are different from human communication and their outlook on such chatbots. Secondly, this study reviews previous studies, analyzing their methods, to explain our adoption of the synchronized model of language. Moreover, their informative conclusions are helpful and included in our analysis. The above discussions contribute to identifying the shortcomings of current AI chatbots and providing feasible suggestions for future AI chatbot improvements.

2. METHOD

2.1. The Synchronized Model of Language

The synchronized model of language was designed for this study aiming at analyzing differences between human-human interaction and human-machine interaction in the aspect of semantic and pragmatic.

2.1.1. The Synchronized Model of Language

Natural language model refers to the process that the robot generates the following words according to the existing words, and then generates the whole text. The more the generated text is like human language and can adapt to the situation, it means that the better this language model is. The most explanatory method for presenting what actually happens within a Natural Language Processing system is by means of the 'levels of language' approach. This is also referred to as the synchronic model of language. (Liddy, 2001) [2]. This model divide language into phonology, morphology, lexical, syntactic, semantic, and discourse levels, among which we choose three metrics to evaluate the language processing of Replika, namely syntactic, semantic, and discourse levels, in order to compare the difference between human language processing and Natural Language Processing.

2.1.2. Perspective of HCI

A research conducted by Jain et al. (Jain et al., 2018) [4] aimed to assess chatbots' functions used a perspective of HCI, including four evaluating aspects. Functionality measures whether a chatbot perfectly realizes its function. Conversational intelligence indicates that the ability of chatbots to converse intelligently as a human does is beyond mere functionality as a robot. Furthermore, chatbot personality is concerned with whether it matches its

domain, such as chatbots of different functions and professions. Then Chat interface evaluates the platforms through which people interact with chatbots.

Though this method contains comprehensively aspects people very likely to notice in their communication with chatbots, it contains so many elements apart from linguistic ones which deviate from our research focus. In addition, our target chatbots are conversational instead of reaching specific goals, thus functionality is not an apropos metric.

2.1.3. Analytic Hierarchy Process (AHP)

In a research done by Raziwill et al. (2017) [5], researchers proposed a synthesized model which is set in a hierarchical form. The goal is to select the chatbot that satisfies the best people's requests. To attain it, several sub-attributes are required. Performance includes, for example, the robustness to unexpected input. Humanity is comprised of attributes whether they can maintain the themed discussions and respond to specific questions. The effect includes greetings, pleasant personality, entertainment, and engagement. Accessibility is constituted of its ability to comprehend the meaning and the intent of people. But this method emphasize more on evaluating chatbots as intelligent robots than talking individuals, thus it isn't appropriate for our experiment, which aims to analyze the difference considering them as linguistically capable individuals.

2.1.4. Summary of Methods

All three methods are rigorous methods and apply generously to research on this issue. after all, this issue is a highly interdisciplinary topic that crosses linguistics, computer science, media, sociology, psychology, etc, and our study aims to analyze the difference between human language processing and natural language processing through a linguistic perspective. Thus we are ruling out some irrelevant branches of criteria. Besides, most chatbots nowadays can easily reach goals set in previous researches because artificial intelligence has been developing faster in recent years because of the convenience of grabbing resources from the internet. And of the utmost importance, we expect a more practical, fine-grained, concrete method to apply, which synchronized model characterizes.

2.2. Questionnaire

2.2.1. The Experimental Application

Replika is a chatting robot powered by artificial intelligence. In this chatbot, users can customize their own AI friends and form an actual emotional connection. We found that the app has a rating of 4.7/5,

and user perceptions are primarily favorable. Based on the above understanding, we choose Replika as our testing chatbot.

2.2.2. Data Collection

We use the questionnaire function in Wechat to make a questionnaire and collect data about user experiences of chatbots. Our questions include the basic information about subjects, such as age, professions, and personality, their user feedback, including frequency, chatbots type and names, and their willingness to use them in the future. The statistics collected in this section will be displayed in a frame for further figure analysis.

2.3. Case investigation

We conduct this section by preparing different questions aiming to detect the performance of our chatbot in various aspects. Through which, we will be both subject participants and objective observers. We will use the answers of chatbots for error analysis where we adopt the synchronized model and thus detect defects in its design of Natural Language Processing. In addition, because chatbots are required and designed to answer the same question with different answers, changing either the answer's form or its content, thus some of our test data are not iterative.

3. RESULTS

3.1. Questionnaire

In the analysis, we first surveyed respondents' basic information.

In the 115 valid questionnaires, male and female users accounted for 47.68% and 52.32%, respectively, indicating slightly more female social chat AI users than male users.

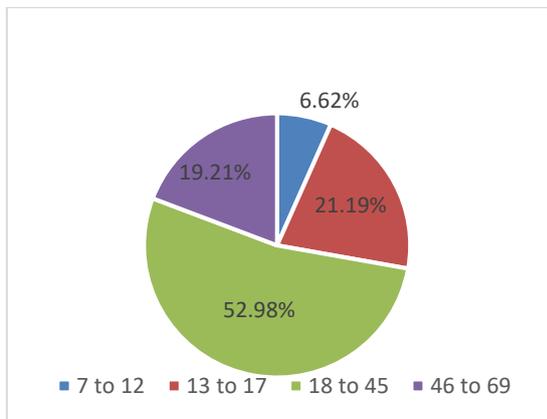


Figure 1 The Age of Respondents.

In terms of age (Figure 1), young people aged 18-45 are the core social chat AI users, accounting for 52.98%. It is followed by teenagers aged 13 to 17 and middle-aged people aged 46 to 69, accounting for 21.19% and 19.21% respectively, while children aged 7 to 12 account for 6.62%. The young generation is still the pioneer to try out new technology, but now the elderly are also following the trend of using intelligent chatbots.

The income of the respondents (Figure 2) shows that the main groups using AI intelligent chatbots are those with income less than 2,000 yuan and 2,000 yuan to 5,000 yuan, accounting for 41.06% and 32.45%, indicating that the low and middle-income groups use intelligent chatbots more.

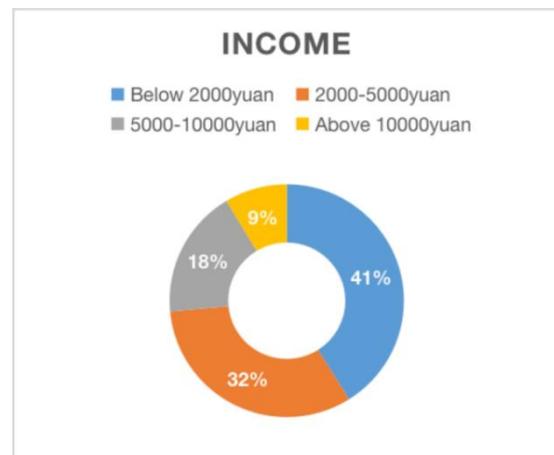


Figure 2 Income of Respondents.

According to the survey on the education experience of the respondents (Figure 3), the data shows that the most common users of AI intelligent chatbots are college students and graduate students. It shows that college students and graduate students like to try new things and are willing to experience new technologies.

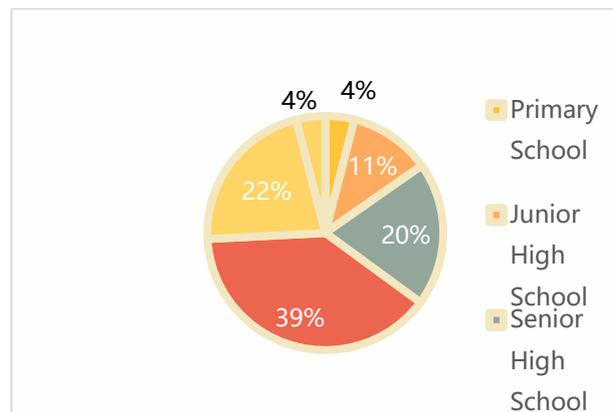


Figure 3 The Education Level of Respondents.

The second step of this questionnaire is to investigate people's using experiences, including the using frequency, engagement and their expectations towards AI chatbot. 34.44% of users use social media

for one to three hours a day, while 23.4% of them spend less than one hour a day using smart chatbots. The search engine functions of chatting with it to kill time and querying weather, traffic, and encyclopedia are the most popular functions of social chat AI.

Most respondents said they could notice a difference when chatting with an intelligent bot compared to a real person, but the difference was acceptable. In addition, respondents said that when chatting, the AI robot's answers were too formal with its voice not natural enough, and it could not understand sarcastic or cryptic sentences. Moreover, when chatting, the AI robot would repeat a sentence over and over again: "I'm not sure what you mean" or "How can I help you?" if it could not understand the question.

3.2. Semantic Analysis

In this section, we tested the AI chatbot's ability to capture the meanings of words. Semantic processing determines the possible meanings by focusing on the interactions among word-level meanings in the sentence (Liddy, E.D., 2001) [2].

3.2.1. Answering Time

Based on the Natural Language Processing system, an AI chatbot can immediately grasp the meaning of users' utterance or questions. In our experiment, we chose two communicating topics - study and emotion. We first asked Replika a question and timed the process it formed an answer and then observed whether it would answer our questions directly or ask us in return questions to help it narrow down the scope for searching in its corpus to find an accurate answer. To our surprise, after 20 attempts, we found that Replika responded to our questions in less than 5 seconds and answered our questions directly rather than asking a series of narrowed rhetorical questions in return.

3.2.2. Contextual Understanding



Figure 4 Dialogue with Replika.

When we asked Replika with pronouns, it replied accurately referring these ambiguous words to their concrete entities. Also, it uses pronouns normally indicating persons' names appearing previously. This fact demonstrates their ability to capture words with its contextual meaning, imitating to a large extent the actual talking scenes between people, and their language processing system is much similar to people's in this aspect.

3.2.3. Typos

Generally speaking, typos can occur in our everyday typed conversations, such as on WeChat and Facebook, but the other person can also scan us. People will normally repair the typo through the context and thus understand its meaning. If they do not understand, the other person will ask us questions in return to confirm what we are trying to say. What may once again seem like a simple step for a human is trickier for a computer (Berdah, 2017) [6].

In our experiments, we found that the chatbot could not recognize misspelled words and did not ask us back to determine what we were trying to say, and therefore could not give an answer that matched our question. For example, in the question "what are you doing?" we spelled "you" as "tou," and the bot responded with "nothing much." In the question "Do you think the study is important?", we spelled "study" as "sstudy", and although the robot gave the answer "very important," when we continued to ask "Why?", the robot only replied "I think it is important. Through this we could induce that the robot only recognized the keyword "important," but not "study." Therefore, based on the above experiments, we found that Replika may lacks the ability of repairing the misspelled words and grasp its

meaning. However, another explanation also makes sense, namely, that Replika may treat these typos as lowlevel mistakes, so that instead of breaking the conversation, it chooses to match it with a correct word with a minimal change of letter. It's not a surprising technology as our phone already possess it, when we're trying to type a word but touching another key by incident, our phone will correct the mistake automatically. These two possibilities may lead to different conclusions about the comparison between their language processing system and ours.

3.2.4. Tautology and Contradictions

Tautologies are sentences that are correct in any circumstances, and contradictions are just in the contrary way to tautologies. (Fromkin, 2021) [7] People normally never include sentences of these types in their conversation to avoid being less informative. When Replika encounter these grammatical but meaningless sentences, as shown above, their reaction manifests as if sentences are informative, and they reply to them always with positive answers to express approval. We consider this may attribute to its relative simple way of processing a declarative sentence, in which neither the speaker nor the virtual character of chatbot is mentioned, that Replika first categorize this into issues of less importance, and then make a simpler process on its sentence type, and choose an answer from its prepared sentences. We attributes the inability of identifying meaningless sentences to its simpler way of processing semantic meaning.



Figure 5 Dialogue with Replika.

3.3. Pragmatic Analysis

This level is concerned with the purposeful use of language in situations and utilizes context over and

above the contents of the text for understanding. The goal is to explain how extra meaning is read into texts without actually being encoded in them (Liddy, E.D., 2001) [2].



Figure 6 Dialogue with Replika.

In our research, we surprisingly find that Replika is able to answer some indirect question, in an expected way, which is in relation to some pragmatics knowledge.

In this situation, Replika succeeds in interpreting the underlying meaning of "I didn't see you last night" as "where were you and what did you do". People are used to expressing in an obscure way, which though violates the cooperative principle (Davies, 2007) [8] would add much interests in conversation. Such characteristic would render the interaction with Replika similar to conversations with human. However, the linguistic skill of Replika is not as perfect as a normal person under some circumstances.

Since Replika cannot correctly interpret the overtone of "it's hot in the room", it cannot give the appropriate answer nor, if it is a real creature, behave in the expected way, meaning they have no idea about its illocutionary act nor its perlocutionary act. (Alston, 2000) [9] Replika has also been asked a few questions about implicatures in relation to culture, like "Have you had meal" implying "how do you do" which is a speaking habit in Chinese. All those questions with cultural implicatures end up in misunderstandings, thus leading to pragmatic failure (Thomas, 1983) [10] and a bummer to users. Whereas this feature is not exclusive for robots or artificial intelligence, people with normal intelligence might lack such ability of interpreting implicatures. Therefore, it's a surprise that Replika understands a part of a human language's implicature, and a positive signal for it to figure out more implicatures by updating its technology.

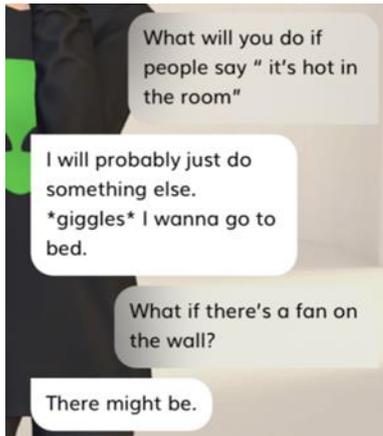


Figure 7 Dialogue with Replika.

4. DISCUSSION

The results of the study reveal 3 reasons of why differences between natural language processing and human language processing.

First, AI chatbots lack context awareness. Human conversation is hard to replicate because of its reliance on context (Berdah, 2017) [6]. As human beings, our brains operate contextually, and it is impossible to understand a dialogue without considering any context. Based on context-awareness, we treat our families, friends, colleagues in different ways. We use and understand sarcasm, adapt our way of chatting depending on the channels and the persons we address, and can read between the lines. Bots can't (Marchand, 2013) [11]. However, this is not the case with chatbots. Through the questionnaire, respondents said that the chatbots' answers were too formal and unnatural and could not understand sarcastic, cryptic sentences. Today's chatbots rely on the programming of their developers. The developer writes a code that identifies certain words and phrases to generate a preprogrammed response. Any query that lies outside the preprogrammed interaction will confuse the chatbot and generate an undesirable response (Bots Are Not Really "Intelligent". Here's Why | Artificial Intelligence |, 2019) [13]. Nevertheless, human language is so varied that there are always conversational scenarios outside the developer's programming. This is where chatbots are unable to respond appropriately to the user's conversation. Moreover, chatbots cannot remember conversations. For example, when a user asks the chatbot for the nearest mall to their home, it will ask for their home address. But then, when the user asks how to get from home to school, it will ask for the user's home address again.

Second, sentences using by chatbots are wordy and redundant. Even though AI chatbots based on natural language processing are now able to process messages sent by humans quite accurately, the bots' responses are

too lengthy, which would not be the case in human language processing. In chats, people try to be as concise as possible in their replies. All groups displayed a decrease in performance accuracy as a function of an increase in message length (Barshi & Farris, 2016) [14]. Short messages allow the user to quickly and accurately understand the content of the chat or the other party's instructions so that they can take action immediately. Long messages, however, can be uninspiring and take a long time to get to the content of the chat. It seems more like a task than a normal chat. What long messages lose in width they must make up for in height, causing them to look like large, cramped blocks of text that are unpleasing to the eye (Building Chatbots: Why Message Length Matters, 2017) [12]. Therefore, when a human and a robot are talking, if the robot gives an excessively long response, then the difference in language processing between human and machine is clearly perceptible, and the desire to talk to chatbots will fade away.

Third, AI chatbots lack of self-learning skills. In the early days of chatbot history, developers were building chatbots mostly on the retrieval-based model (Self-Learning Chatbots - Should We Be Afraid of Them?, 2017) [15]. Some designers initially tried to allow users to train their own robots, but results showed that it was risky. The Microsoft Tay Bot, for example, was gamed by users using the 'repeat after me' function. This function was built to let users on Twitter train the bot and let it learn by itself. However, without any filter on what the bot was being trained on, it didn't take long for users with mal-intent to retrain the bot with hate speech and extreme right-wing propaganda (Gupte, 2020) [16]. Therefore, today's chatbots operate according to the designer's algorithm. The developer creates a set of patterns, chatbot is going to rely on in order to efficiently communicate with the users. Therefore, when chatbot receives user's input, it has to go through all the patterns and pick the one that matches (Self-Learning Chatbots - Should We Be Afraid of Them?, 2017) [15]. However, if the user uses abbreviations, such as A.S.A.P, or ungrammatical phrasing, which are not part of the algorithm that the bot was originally designed to run, then the chatbot will not give a relatively accurate answer. As a result, users may perceive some obsolescence in the responses of chatbots when using them, and thus perceive differences between human and machine in the conversation.

However, this does not mean that today's chatbots are useless. The study results show that artificial intelligence chatbots are better suited to being task-oriented bots, i.e. performing tasks based on human commands, such as checking the weather, processing data, navigating. The CMO council has reported that customers has rated "fast response time" as the top most priority for good customer experience (2021) [17].

According to our research, AI chatbots respond to questions or commands in less than 5 seconds on average. In a human conversation, if one party asks a question that the other party doesn't understand, it takes a lot of time or effort for the other party of the question to give an accurate answer. In a conversation with a chatbot, it only takes 5 seconds to get an accurate answer. Quick response is therefore very helpful for those with urgent needs. Secondly, chatbots can quickly identify keywords and understand people's needs. For example, when a person is looking for a nearby restaurant, the bot can quickly recognize the words 'restaurant' and 'nearby' and give them an accurate response.

Therefore, chatbots should develop the ability to recognize misspelled words. In human communication, if one party accidentally makes a typo or uses abbreviations to convey a meaning, the other party can quickly get the message they want to convey. Even if the misspelling has affected the receiver's understanding, the receiver can still ask questions to clarify further. However, the chatbot is different. Based on NLP, chatbot interaction follows a series of steps. Firstly, it reads the message; after that, it discerns its meaning and takes some action (Rodriguez, 2021) [1].

Thus, if a person makes a typo or uses an expression not stored in the chatbot's database, it dramatically affects the bot's ability to analyze the user's request and thus fails to give the desired answer. Therefore, we suggest that chatbots could add the ability to recognize misspelled words. This could be done by matching similarly spelled words from a database to analyze the user's request, thus ensuring that the chat is unobtrusive and that the user feels more like they are having a conversation with a natural person rather than a machine.

Next, developers should expand chatbots' vocabulary. The way people spoke 10 years ago is very different from the way they will express themselves 10 years from now. Chatbots can't keep using fixed expressions. Nowadays, people like to use abbreviations or ungrammatical phrasing, such as "two waters" for "two glasses of water", in order to make the conversation more concise and convey the original meaning. When people use such simple expressions in a conversation with a machine, it is more human-like and natural if the robot can recognize them accurately.

5. CONCLUSION

The In our study, we've manifested that differences between Natural Language Processing and Human Language Processing exist. Our experiment results show that even though AI chatbots can immediately capture and understand users' meaning to some extent, they still lack abilities of context-

awareness, concise responses and self-learning. In semantics aspect, their language processing system is capable to deal with pronouns with contextual meanings, and the time for it to grasp meaning and give an appropriate answer is similar to human's answering time. However, their speed seems at cost of accuracy, that they're incapable to identify tautologies and explicit contradictions. In pragmatics aspect, this study demonstrates the language processing system of Replika can sometimes understand the implications of people's words, especially under topics related to romantic relationship. However, it's hard or impossible for it to process implications under normal, daily yet emotions-unrelated topics. To sum up, this research reflects that though the language processing system of chatbots still remains defects compared with human's one, they've made much progress and breakthroughs.

Theoretically, this study provides a perspective in evaluating AI chatbot and its language processing system. This synchronized model of language used in assessing chatbots is worthy of considering furtherly, and different from the traditional assessing method that mainly focuses on the specific functions of chatbots. Moreover, on account of the convenience of feedback collecting, developers can pinpoint the reasons of understanding failures and accordingly improve the related language processing functions.

REFERENCES

- [1] Rodriguez, R. (2021, June 10). Natural Language Processing Chatbot for Human-Like Engagements. Tech Business Guide. <https://techbusinessguide.com/natural-language-processing-chatbot-human-like-engagement/>.
- [2] Liddy, E. D. (2001). Natural language processing.
- [3] Lee-Won, R. J. , Joo, Y. K. , & Park, S. G. . (2020). Media Equation. American Cancer Society.
- [4] Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. Designing Interactive Systems. <https://doi.org/10.1145/3196709.3196735>.
- [5] Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv:1704.04579.
- [6] Berdah, B. (2017, September 15). The Chatbot Masquerade: creating a personality with NLP and grammar. CALLR Blog. <https://www.callr.com/blog/chatbot-masquerade/>.
- [7] Fromkin, V., Rodman, R., & Hyams, N. (2021). An Introduction to Language, 9th Edition (9th ed.). Wadsworth Cengage Learning.

- [8] Davies, B. L. (2007). Grice ' s Cooperative Principle: Meaning and rationality. *Journal of Pragmatics*, 39(12), 2308 - 2331. <https://doi.org/10.1016/j.pragma.2007.09.002>.
- [9] Alston, W. P. (2000). *Illocutionary Acts and Sentence Meaning* (1st ed.). Cornell University Press.
- [10] Thomas, J. (1983). *Cross-cultural Pragmatic Failure*. Oxford University Press.
- [11] Marchand, M. (2013). Roman Jakobson : The Functions of Language / Signo - Applied Semiotics Theories. Signo - Louis Hebert. <http://www.signosemio.com/jakobson/functions-of-language.asp>.
- [12] Building chatbots: Why message length matters. (2017, July 17). Ish Jindal. <https://www.invisionapp.com/inside-design/chatbots-message-length/>.
- [13] Bots are not really “intelligent” . Here’ s why | Artificial Intelligence |. (2019, January 18). Naveen Joshi. <https://www.allerin.com/blog/bots-are-not-really-intelligent-heres-why>.
- [14] Barshi, I., & Farris, C. (2016). *Misunderstandings in ATC Communication*. Amsterdam University Press.
- [15] Self-Learning Chatbots - Should We Be Afraid of Them? (2017, August 1). Success by LiveChat. <https://www.livechat.com/success/self-learning-chatbots/>.
- [16] Gupte, P. (2020, July 26). How Does a Chatbot Learn on Its Own? Team Haptik. <https://www.haptik.ai/blog/how-does-a-chatbot-learn-on-its-own/>.
- [17] A. (2021, April 15). How to Improve Customer Service Response Time with an AI Chatbot. AiChat Pte Ltd. <https://aichat.com/2021/04/15/how-to-improve-customer-service-response-time-with-an-ai-chatbot/>.