Asia Pacific Journal of Information Systems
Vol. 30 No. 4 (December 2020), 832-846

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2020.30.4.832

# An Experimental Comparison of the Usability of Rule-based and Natural Language Processing-based Chatbots

Yeji Lim[a], Jeonghun Lim[b], Namjae Cho[c,*]

[a] *Ph.D. Student, School of Business, Hanyang University, Seoul, Korea*
[b] *Ph.D. Student, School of Business, Hanyang University, Seoul, Korea*
[c] *Professor, School of Business, Hanyang University, Seoul, Korea*

**ABSTRACT**

Service organizations increasingly adopt data-based intelligent engines called chatbots in support of the interaction between customers and the companies. Two different types of chatbots have been suggested and introduced by companies leading the adoption of this emerging technology: rule-based chatbots and natural language processing-based chatbots. While the differences between these two types of technologies look relatively clear, the organizational and practical impacts of the differences have not been systematically explored. This study performed an experiment to compare the use of the two different types of chatbots used in practice by two comparable organizations. These two types of actual chatbots were used by Korean on-line shopping malls with similar business models (mobile shopping), length of history, size and reputation. The comparison was made based on such dimensions as usability, searchability, reliability and attractiveness. Contraty to conventional expectation that the superiority in technology will produce superior usability, the results show mixed superiority. The discussion on the reasons is presented.

*Keywords:* Chatbot, Rule-based, Natural Language Processing-based, Mobile Shopping, Machine Learning, Usability, User Experience, User Interface

## Ⅰ. Introduction

### The Background and Purpose of the Study

The use of intelligence technologies for organizational processes is rapidly increasing in various industries in Korea. A chatbot, representing chatting software robot, is one type of such fast spreading intelligence technology. Chatbots are used to help interaction between customers and a company. When

*Corresponding Author. E-mail: njcho@hanyang.ac.kr

※ This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2017S1A5A2A03068426).

customers have questions, comments, complaints, chatbots can interact with customers effectively and efficiently by understanding queries and questions from customers and providing answers and relevant rationales in response. Chatbots have wide range of applications and considered to enhance customer experiences, manage large number of customers, and be cost effective (Megha, 2019).

Chatbots generate answers based on either pre-designed rules or on the learned patterns found from previous interaction. The former is called rule-based chatbot as it is based on a set of rules designed and stored by the developer team of the company. The rules represent the summary of the experience and knowledge of the company. The later one, Natural language processing(NLP)-based chatbot - so called as by user companies in industry in practiceas this one provides more expressive interactions - is based on the machine learning ability and the patterns identified from the analysis of the large amount of data on evolving interactions with customers. This one is adaptable to the user's individualized language usage, searches, and preferences learned from continuing usage history (Bharti et al., 2020).

As the interaction between customers and a company is of prime importance in retail industry, the use of chatbot for customer interaction is considered to have strategic value. Many simple repetitive activities being performed by human workers, such as routine interaction with customers, can be replaced by chatbots equipped with interaction rules, intelligence algorithms, and large amount of data. For such reason, the adoption of chatbots by on-line shopping companies is expected to increase the quality of interaction and also save costs. Currently in Korea chatbots are being used by such leading on-line shopping companies as Lotte.com and Interpark.

User experience is considered as the consequence of interaction between users and system devices, contents, and the interaction environments (Moon et al., 2008). The quality of interaction experience via omnibus channels is considered critical as it is one major part of the holistic customer experience. In using shopping support systems users play dual roles; both as system users and as customers, making dual dimensions of interaction satisfaction of high importance (Mamani et al., 2012).

Morville (2004) defines customer experience as all accumulated behavioral, sentimental and knowledgeable memories created while using any services or products. In this context, user-customer experience can be considered as an overall experience that users think and feel while using a system, product, and service, both directly and indirectly.

Despite the rapid diffusion of intelligence technology and the prospective increase in the extensive use of chatbots by on-line shopping service providers, research on the practical impacts of the use of chatbots is still limited in diversity and depth. This research specifically focuses on the differences between two major types of chatbots: rule-based and natural language-based. Two major companies which declared the use of chatbots of each type are chosen for experimental comparison. The two chosen companies, Lotte.com and Interpark, are the leading on-line shopping companies in Korea. Both the companies launched its business at the same time (September, 1997) for the first time in Korea. Interpark's chatbot, called Talk Jibsa, is a rule-based chatbot as announced by the company. On the other hand, Lotte.com announced that their chatbot called Samantha is a natural language processing-based chatbot.

# Ⅱ. Conceptual Background

## 2.1. Chatbot

Chatbot refers to an interactive interface system which uses agent-based artificial intelligence system. Chatbot can interact like a human. Chatbots are pieces of software that are usually used as an interface between a company and its customers. The tasks performed by these bots can range from simple technical support for products to ordering products offered by the company (Keszocze et al., 2019). In recent years, the market for chatbots has grown rapidly. Many companies offer frameworks with varying features and pricing models (Amir, 2017).

Chatbots have been introduced in various industries including messenger services (Kim et al., 2017). Some case studies have been reported, for example, in news media industry (Roto, 2006) and marketing channel management (Han, 2017).

Chatbots are typically used in dialogue systems for various practical purposes including customer service or information acquisition. Some chatbots use sophisticated natural language processing systems, but many simpler ones scan for keywords within the input, then pull a reply with the most matching keywords, or the most similar wording pattern, from a database (Sidenko et al., 2019).

### 2.1.1. Rule-based Chatbot

A rule-based system consists of five parts: the user interface, inference engine, knowledge base, conflict resolution strategies, and working memory. Inference engine deduces rules stored on the knowledge base that users must apply based on the input entered by a user through the interface.

There are numerous online Chatbots that have been built with a Rule based technology. Knowledge in the Rule based Chatbots is organized and presented with conversational patterns (Arsovski et al., 2019). The knowledge base among all is considered the key element to this system. The knowledge base contains professional knowledge in a particular field extracted from human experts. Individual knowledge element is organized in the form of a set of rules. Each rule expresses instructions, recommendations, strategies, experiences, or relationships. A rule takes the form of a conditional declaration with 'IF (conditions) THEN (action)' structure. If a certain rule condition is met, the corresponding activity is executed or the target value is determined by selecting and applying appropriate rules.

As a system grows in the size of the knowledge base, complex combination of factor values is used to determine the target value. By incorporating the nature of a particular situation into the inference process, the system can execute optimal sequence of rules to solve complex problems in various areas. Sometimes explanatory natural language descriptions are used to present the inference path of rules fired until the final results are obtained. Rule-based systems have been applied in solving problems in many areas but where the rules were drawn.

### 2.1.2. Natural Language Processing Based Chatbot

As an important sub-theme of artificial intelligence, the natural language processing aims to handle the understanding, analysis, and creation of human-like languages. Natural Language refers to a spoken language that people use in everyday life for a communicative purpose. Natural language is differentiated from the formal written language artificially compiled up based on grammatical rules, which has been used by traditional computer systems. Natural language is semantically closer to the original intent of the assertion in the mind of the human designer, so natural language descriptions are easier and faster to formulate, and are less likely to contain errors than manually-generated formal assertion descriptions (Keszocze et al., 2019).

Examples of the uses of natural language processing include automatic translation, query response, information search and user-friendly interpretation of computer-processed results. When applied to chatbot, the natural language processing capability of the chatbot can convert voice, graphs, and texts into a comfortable interactive dialogue, vice versa. Human users would feel comfortable while exchanging information with machine. When applied to public service, scalable and flexible chatbot interaction can be provided for improved satisfaction of customers (Park, 2017).

The use of natural language processing-based chatbots is increasing recently due to the fast improvements in computer performance and developments in machine learning capability of computers. Despite of such high prospect, research on consumer response in practice, especially in the context of mobile shopping is scarce making heightened motivation of this research.

## 2.2. User Experience

### 2.2.1. The Concept of User Experience

ISO (International Organization for Standardization) standards defines user experience (UX) as "a comprehensive concept that includes users' preferences, beliefs, emotions, physical and mental behaviours and reactions, and perception that occurs before, during and after a product is used." Hassenzahl and Tractinsky (2006) define the user experience as 'an experience that can be conclusively gained by the user's internal state and system characteristics in any situation'. The user experience, thus, can be summarized as a continuous and holistic composition that combines the various aspects and circumstances that users encounter using a specific product and leads to a specific judgment and action (Hassenzahl, 2018).

Roto (2006) notes that users create a comprehensive experience by changing their feelings and attitudes toward a product, influenced not just by the momentary gain from using the product, but also by expectations or prior knowledge. In this vein, the user experience is related to all aspects of users' collective interaction with products, services, and companies. In summary, the user experience(UX) is the experience of a user from all memory, knowledge, behaviour, and emotions that accumulate when he or she use a particular product or service.

Identifying the needs of a user is a prerequisite to an asymptotic user experience. Then, it goes beyond simply providing what users say they want. In order to achieve a high level of user experience, companies are recommended to carefully provide services in marketing, engineering, and interface design. Further, the cultural and social factors as well as the usage context should also be scrutinized for high quality of experiences of interaction with users (Arhippainen et al., 2003).

Extending this concept, users' experience in the interaction with computer systems can be considered to related to all processes of interaction between a user and the device, contents, usage contexts and system interfaces (Kim, 2012). Konstan and Reidl (2012) suggest to go beyond the accuracy of a prediction system during the design of both algorithms and systems and to take into account user experiences. In suggesting a new type of cognitive search system, the difference from traditional search systems is highlighted in the sense that the new approach improves the user experience (UX) through optimization factors such as response time, ease of use, friendly interface, and user interaction (Mamani et al., 2012).

In this vein, we may define the user experience with chatbot as all perception, cognitive consequences, and emotions, expectations that users develop before, during and after the use of chatbot systems.

#### 2.2.2. Assessing User Experience

The Honeycomb Model is a diagram introduced by Peter Moville in his book "Search 2.0: Evolution of Discovery (Morville, 2006)". This model serves as a good tool for the evaluation of usability and other dimensions and goals related to experiences. Morville's Honeycomb model is a survey instrument to assess the usability of a system based on seven components: Usefulness, usability, attractiveness, searchability, accessibility, reliability and value (Morville, 2004). It helps us to measure the user experience from a holistic perspective using seven different aspects of the user experience. These user experience assessments help us to understand different aspects for diverse situations.

In this study, we reconstructed the model into five factors to fit into the chatbot situation. The purpose is to compare rule-based vs. natural language processing-based chatbots. The definitions of the five dimensions is summarized in <Table 1>.

## Ⅲ. Research Method

### 3.1. Experimental Treatment: Chatbots

#### Treatment 1: Interpark's Talkjipsa, a Rule Based Chatbot

The 'Talkjipsa' from Interpark is the most well-known rule-based chatbot in South Korea. The rules of Talkjipsa recognizes different needs of users and respond. Talkjipsa draws answers based on pre-defined Q&A. It is designed to search for products that users want while gradually making choices in pre-determined categories. For example, in order to search for black T-shirts for men, the chatbot searches clothes, then go through men's clothing, and gradually reduce the range to the next colour requirements category. (Equipment-> Male Clothing-> T-shirt category)

#### Treatment 2: Lotte.com's Samantha, a Natural Language Processing-based Chatbot

'Samantha' by Lotte.com is the representative natu-

<Table 1> User Experience Assessment Dimensions

| Variables | Definition |
|---|---|
| Usefulness | Whether the products, services, systems are truly useful to users. |
| Usability | In using products and/or services, whether the user can have fewer concerns or difficulties in usability. |
| Searchability | If a user is able to find relevant information or services they need easily and flexibly. |
| Reliability | Measures if users can trust in the service for both direct and contextual purposes. |
| Attractiveness | Whether the experiences satisfy human five senses in terms of mental and emotional aspects. |


<Figure 1> Interpark's Talkjipsa


<Figure 2> Lotte.com's Samantha

ral language processing-based chatbot in South Korea. Samantha is much younger than Interpark's Talkjipsa. It understands conversations with users in a natural language and answers based on cumulative learning. Unlike Talkjipsa. A user of Samantha types "Find me a black T-shirt" to make the chatbot search for the product they want. Talkjipsa and Samantha represent two different approaches of chatbot technologies. This study aims to use these real contexts for a semi-field experiment to study practical implications of the two technologies experimentally.

### 3.2. Measurement of Dependent Variables

Five dependent variables were included in the research. They are usefulness, usability, searchability, reliability and attractiveness. Following the practice of existing literature we reviewed, the dependent vari-

<Table 2> Survey Contents by Assessment Element

| Variables | Definition |
|---|---|
| Usefulness | Did the chatbot help spending less time searching for products compared to typical online shopping?<br>Did you get help form chatbot when choosing a product?<br>Was the chatbot more convenient compared to traditional online shopping methods?<br>Did the chatbot service provide useful information? |
| Usability | Was the entire process from product discovery to selection seamlessly connected?<br>Was the overall screen configuration easy to browse?<br>Could you sort the products recommended by Chatbot in different ways? |
| Searchability | Was it easy to find the product you want?<br>Could you search for similar products?<br>If there was no product you want, did it give you a further search advice?<br>Was the search provided sufficiently detailed product information? |
| Reliability | Could you trust the product recommended by Chatbot Service?<br>Were chatbot services reliable?<br>Could you trust the price of the product that the Chatbot found?<br>Could you trust the information that the Chatbot recommended? |
| Attractiveness | Were you satisfied with the experience of the chatbot service?<br>How satisfied were you with the use of the chatbot service?<br>Are you willing to use it again?<br>Are you willing to recommend the chatbot service to others? |

ables were measured using a questionnaire with 20 questions in the form of Likert-type 5-poimt scale. <Table 2> summarizes the questionnaire items of the research instrument used to measure the five dimensions of the user experience, the dependent variables.

### 3.3. Experimental Procedure

Eight seven students from the Business School of Hanyang University served as research subjects. Hanyang University is one of the typical major universities in Korea. The students are considered to represent major group of on-line shopping consumers in terms of age strata and the level of education. Most of the participants were in their 20s and accustomed to use smartphones.

Repeated measurement approach was used in the experiment The subjects were told to use both Talkjipsa and Samantha and fill up the questionnaire administered by the research team. The subjects were randomly divided into 2 groups for alternating sequence. One group was told to use Talkjipsa first, and the other group was told to use Samantha first. The major charicteristics of the experimental procedure is summarized in <Table 3>.

## Ⅳ. Study Results

### 4.1. Data Collection and Data Analysis

The experiment produced 79 usable sample data after excluding 8 incomplete responses. The number of male subjects was slightly larger than the number of female subjects. 76% of the subjects were under age 24 (and over 18), and all the subjects were under age 30. 47% of the subjects had previous experiences in using chatbots. And the major purpose of their previous use of chatbots was the consultation with regard to the services they used. Diverse locations were used by the subjects in using computers and shopping sites.

In each experimental session, all participants were told to take a note of the task completion time immediately after they chose the product they wanted to

<Table 3> Experimental Procedure

| | |
|---|---|
| Participants | Students of Business School, Hanyang University, Seoul, Korea<br>(second, third and fourth year) |
| Attendance benefits | Souvenir, class participation score |
| Nature of experiments<br>Test kit | Repeated Measures Approach with mixed sequence<br>Treatment 1: Talkjipsa (Rule-based chatbot)<br>Treatment 2: Samantha (NLP- based chatbot) |
| Experimental duration | Time limit to complete the task<br>10minutes for session 1 + 10minutes for session 2<br>Total 20minute were allowed to finish. |
| Experimental task | Selecting a desired product using the given Chatbot |
| Experimental processes | Group 1 and Group 2 were randomly divided.<br>Session 1: Group 1 used Talkjipsa first and Group 2 used Samantha first. After completing the task using the first chatbot, the subjects were asked to fill up the study questionnaire.<br>Session 2: Group 1 used Samantha and Group 2 used Talkjipsa first and filled up the study questionnaire after completing the task. |

<Table 4> Basic Statistics of the Nature of the Experimental Subjects

| Gender | | Age | |
|---|---|---|---|
| male<br>female | 53%<br>47% | under 24<br>25~29 | 76%<br>24% |
| Experience with chatbot use | | Type of use | |
| have<br>have not | 47%<br>53% | service consultation<br>delivery consultation<br>Shopping<br>Financial consultation<br>other | 30%<br>14%<br>12%<br>11%<br>33% |
| Typical location of the use | | | |
| Home<br>Public transport<br>Company, School<br>Others or any location | | 22%<br>12%<br>10%<br>56% | |

purchase. The average time to use Talkjipsa was 2.01 minutes and the average time to use Samantha was 2.32 minutes.

The summary of the demographic nature and basic statistics of the 79 subjects is presented in <Table 4>.

## 4.2. Correlation Analysis

The correlation coefficients and mean standards deviation between the variables are shown in <Table 5> and <Table 6>.

### Rule-based Chatbot (Talkjipsa) Use Sessions

With regard to the Talkjipsa session, the factor analysis using the Verimax method showed that factor loadings ranged between 0.654 and 0.799. The internal consistency of the scales as measured by Cronbach Alpha value were consistently higher than

<Table 5> Correlations among Variables in Rule-based Chatbot (Talkjipsa) Use

| | | Usefulness | Usability | Searchability | Reliability | Attractiveness |
|---|---|---|---|---|---|---|
| Usefulness | Pearson Correlation | 1 | .424** | .373** | .321** | .430** |
| | Sig. | | .000 | .001 | .001 | .000 |
| Usability | Pearson Correlation | .424** | 1 | .493** | .406** | .394** |
| | Sig. | .000 | | .000 | .000 | .000 |
| Searchability | Pearson Correlation | .373** | .493** | 1 | .591** | .501** |
| | Sig. | .001 | .000 | | .000 | .000 |
| Reliability | Pearson Correlation | .321** | .406** | .591** | 1 | .600** |
| | Sig. | .001 | .000 | .000 | | .000 |
| Attractiveness | Pearson Correlation | .430** | .394** | .501** | .600** | 1 |
| | Sig. | .000 | .000 | .000 | .000 | |

Note: **Correlation is significant at the 0.01 level

<Table 6> Correlations among Variables in NLP -based Chatbot (Samantha) Use

| | | Usefulness | Usability | Searchability | Reliability | Attractiveness |
|---|---|---|---|---|---|---|
| Usefulness | Pearson Correlation | 1 | .488** | .297** | .342** | .302** |
| | Sig. | | .000 | .001 | .001 | .001 |
| Usability | Pearson Correlation | .488** | 1 | .415** | .466** | .559** |
| | Sig. | .000 | | .000 | .000 | .000 |
| Searchability | Pearson Correlation | .297** | .415** | 1 | .360** | .472** |
| | Sig. | .001 | .000 | | .001 | .000 |
| Reliability | Pearson Correlation | .342** | .466** | .360** | 1 | .319** |
| | Sig. | .001 | .000 | .001 | | .001 |
| Attractiveness | Pearson Correlation | .302** | .559** | .472** | .319** | 1 |
| | Sig. | .001 | .000 | .000 | .001 | |

Note: **Correlation is significant at the 0.01 level

<Table 7> Factor Analysis Results and Reliability of Variables

| | N of items | Factor Loading | Cronbach's alpha | Total explained Variance (%) |
|---|---|---|---|---|
| Usefulness | 4 | .654 | .795 | 72.513 |
| Usability | 4 | .714 | .779 | |
| Searchability | 4 | .798 | .748 | |
| Reliability | 4 | .792 | .746 | |
| Attractiveness | 4 | .799 | .753 | |
| Kaiser-Mayer-Olkin : .796,<br>Bartlett's chi-squared : 118.379, The degree of freedom : 10, Sig. : .000 | | | | |

0.7 making all the variables judged to be reliable to use for further analysis. The results of KMO and Bartlett's statistics were 0.796 and significant at 0.000. <Table 7> shows the result of the factor analysis.

#### Natural language Processing-based (Samantha) Chatbot Use Sessions

With regard to the Samantha session, factor analysis using the Verimax method showed that factor loadings ranged between 0.711 and 0.828. The internal consistency of the scales as measured by Cronbach Alpha were higher than 0.7 making all the variables judged to be reliable to use. The results of KMO and Bartlett's statistics were 0.772 and significant at the level of 0.000. <Table 8> summarizes the results of factor analysis.

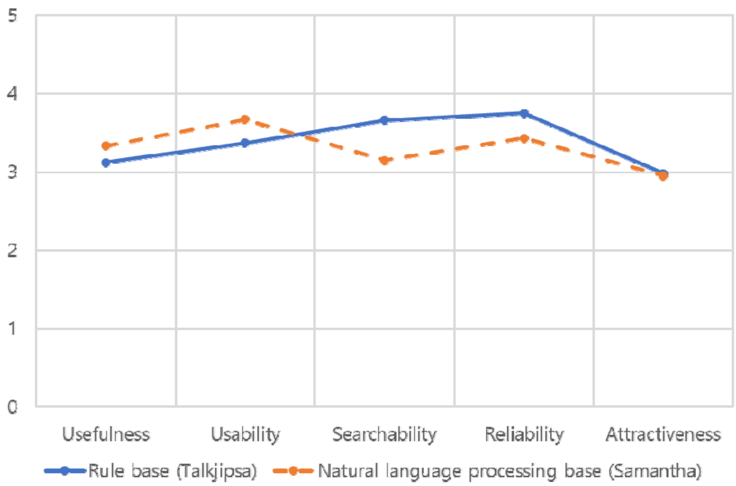
### 4.3. Patterns of Experimental Outcome Scores and Overall Comparison

The average scores of the five outcome variables are depicted in <Figure 3>.

Samantha, a natural language processing-based chatbot was superior than Talkjipsa in terms of usefulness and usability. On the other hand, the response scores with regard to the rule-based chatbot, Talkjipsa were higher than Samantha in 2 dependent variables:

<Table 8> Factor Analysis Results and Reliability of Variables

| | N of items | Factor Loading | Cronbach's Alpha | Total explained Variance (%) |
|---|---|---|---|---|
| Usefulness | 4 | .659 | .754 | 78.455 |
| Usability | 4 | .828 | .783 | |
| Searchability | 4 | .711 | .739 | |
| Reliability | 4 | .778 | .747 | |
| Attractiveness | 4 | .744 | .722 | |
| Kaiser-Mayer-Olkin : .772<br>Bartlett's chi-squared : 125.267, The degree of freedom : 10, Sig. : .000 | | | | |




<Figure 3> Average Assessment Scores of the Dependent Variables

<Table 9> T-test Results of Outcome Variable Comparison of Means

| | Talkjipsa (rule base) | Samantha (NLP base) | *t* | *p* |
|---|---|---|---|---|
| Usefulness | 3.12 | 3.33 | 2.848 | .001 |
| Usability | 3.37 | 3.67 | 2.877 | .000 |
| Searchability | 3.65 | 3.16 | 4.508 | .000 |
| Reliability | 3.71 | 3.34 | 2.995 | .000 |
| Attractiveness | 2.98 | 2.95 | 2.229 | .003 |

searchability and reliability. The difference in the scores of attractiveness were not as big asa the other variables. The score of Talkjipsa was slightly higher than the one of Samantha.

<Table 9> shows the result of t-test comparing the average values of the dependent variables. All the differences between the two type of technology were signigicant. One notable result is that the minor superiority of Talkjipsa in attractiveness was also found significant.

The result is interesting as Samantha's machine learng technology is said superior in every sense of technology. It is also closer to natural language than Talkjipsa. However, practical response to the two technology by users were not consistent to the technological expectation, which may reflect the view of developers.

### 4.4. Itemized Comparison of the Experimental Outcome Scores

As shown in <Table 9> the practical responses to the superiority of technology is inconsistent despite the technical superiority of the natural languare-based chatbot, which use machine learning algorithm. To clarify the comparison by examining differences item by item, we performed a further itemized comparison between the two technology using all the 20 items of dependant variables.

<Table 10> summarizes the itemized scores of dependent variables of the two experimental sessions. In terms of Usefulness, Interpark's Talkjipsa score was slightly higher than Samantha in help and usefulness of information, but Samantha was way high in search time and shopping convenience than Talkjipsa.

<Table 10> Itemized Assessment Score Comparison

| Evaluation item | | Talkjipsa (rule base) | Samantha (natural language processing base) | $t$ | $p$ |
|---|---|---|---|---|---|
| Usefulness | Search Time | 2.85 | **3.63** | 3.232 | .000 |
| | Chatbot'S Help | **3.37** | 2.96 | 2.313 | .000 |
| | Shopping Convenience | 2.78 | **3.65** | 3.609 | .000 |
| | Useful Information | **3.48** | 3.06 | 2.241 | .002 |
| Usability | Execution Process | 3.34 | **3.81** | 2.674 | .000 |
| | Search Process | 3.30 | **3.78** | 2.988 | .000 |
| | Screen Configuration | 3.29 | **3.89** | 3.356 | .000 |
| | Alignment Method | **3.53** | 3.20 | 2.491 | .001 |
| Searchability | An Easy Search | **3.63** | 3.52 | 2.554 | .000 |
| | A Similar Product | **3.84** | 2.82 | 6.345 | .000 |
| | Additional Search | **3.82** | 2.61 | 6.679 | .000 |
| | Detailed Information | 3.48 | **3.54** | 2.455 | .001 |
| Reliability | Recommendation Reliability | 3.51 | **3.59** | 2.248 | .001 |
| | Chatbot Reliability | **3.68** | 3.47 | 2.298 | .000 |
| | Price Reliability | **3.86** | 3.08 | 3.831 | .000 |
| | Information Reliability | **3.78** | 3.23 | 3.604 | .000 |
| Attractiveness | Usage | 3.05 | **3.20** | 2.254 | .001 |
| | Product Satisfaction | **3.15** | 3.06 | 2.225 | .003 |
| | Reuse Purpose | **2.85** | 2.76 | 2.237 | .003 |
| | Recommendation | **2.86** | 2.78 | 2.201 | .004 |

Among the 4 items of Usability, Samantha scores were higher than Talkjipsa in 3 tiems (execution process, search process, and screen configuration). The score of Talkjipsa was higher in alignment method (supporting flexible sorting) than Samantha. But the difference was not as big as the other items.

With regard to Searchability, Talkjipsa was superior in 3 out of 4 items (ease of search, reaviewing similar products, performing additional search). Samantha was superior in detail of information, but here again the difference was not big.

In terms of Reliability, Talkjipsa was higher than Samantha again in 3 out of 4 items (chatbot reliability, trustworthiness of price information, and information reliability). Samantha's score in recommendation reliability was slightly higher than that of Talkjipsa.

In terms of attractiveness, Talkjipsa was superior to Samantha in the level of satisfaction, reusability, and recommendation to others, and Samantha was superior to Talkjipsa in usage attractiveness. However, the differences in all the four items were not very big.

## Ⅴ. Conclusion

This experiment compared and analysed the usability of two different types of chatbots used in mobile shopping sites; rule-based (Talkjipsa) and natural language-based (Samantha). The natural language-based Samantha is based on machine learning algorithm technology, which is newer and considered as superior than traditional rule-based technology. So, the conventional expectation prefers Samantha to Talkjipsa in every sense. We can easily expect that users do not have to go through the hierarchies of categories presented by the rule-based chatbot when they use natural language processing query.

However, different to the conventional wisdom, the two types seem to had relative pros and cons in practice. Rule-based chatbot was superior in searchability and reliability. On the other hand, natural language processing-based chatbot was superior in usefulness and usability. Rule-based chatbot provided more complicated search, but was considered less intuitive. Interestingly, although natural language-based interaction was supposed to be easier to use and flexible, some users seem to prefer to use guided search of rule-based chatbots, especially when they are not familiar to use chatbots.

Several reasons can be considered. One possible reason is the familiarity of user interface to the existing base of users. Users are still feel more comfortable toward the structured interfaces than the less structured natural language interaction. In this regard, designers of chatbot systems should be advised to consider such social perception and havits in designing different aspect of a chatbot, rather than applying new technique in every aspects of the interface.

Another reason can be the amount of interaction data and the speed of growth of the amount of data to be used for machine learning. Samantha, after learning sufficiently from past customer interaction, may respond to users with more convenience and comfort. However the amount of such practical interaction data is not comparable to typical development settings such as Go game. So the learning speed must be limited. However, this expectation that Samantha will do better in the future is still based on technical assuptions without considering why users are in favour of the structured interfaces of Talkjipsa in such a large number of functional items.

In both chatbots, the level of reuse intention and recommendation for others were not as high as expected. The results mean that the level of maturation in the design of interfaces is still in the stage of developing in both technologies.

College students of Hanyang University served as research subjects and Peter Mobil's Honeycomb model was used as the frame of reference to evaluate the quality of experience. As typical experimental studies using students as research subjects, we need to excuse the limited external validity and generalizability in term of the type of users. However, due to the nature of the experimental task (mobile shopping task) the subjects we used in this study can be considered as largely similar to the major group of users are young people with high familiarity to on-line shopping. We believe college students taking MIS course tend to represent this group of re this group of consumers so that the sample selection is not very far from general practices.

In future research, two considerations can be incorporated: generalizability and maturation effect. As we used college students as research subjects, systematic extension of the user groups in older or younger ages will be a natural extension of this experiment. We expect that, the growth in terms of the rule set of rule-based chatbot and data set in natural language-based machine learning chatbot will make evolutionary differences. Study of dynamic evolution of the effects of the two types of chatbots will help companies to find the right policy for the establishment and transformation of the use of chatbot technologies. Further, a future research may test the effectiveness of the use of customized design of interaction to different types of users (Tanya et al., 2017).

## <References>


[1] Arhippainen, L., and Tähti, M. (2003). Empirical evaluation of user experience in two adaptive mobile application prototypes. *In MUM 2003. Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia, 11*, 27-34.

[2] Arsovski, S., Osipyan, H., Oladele, M. I., and Cheok, A. D. (2019). Automatic knowledge extraction of any Chatbot from conversation. *Expert Systems with Applications, 137*, 343-348.

[3] Bharti, U., Bajaj, D., Batra, H., Lalit, S., Lalit, S., and Gangwani, A. (2020). Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after COVID-19. *In 2020 5th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 870-875.

[4] Cho, N., and Park, S. (2001). Development of electronic commerce user consumer satisfaction index(ECUSI) for internet shopping. *Industrial Management & Data Systems, 101*(8), 400-405.

[5] Hassenzahl, M., and Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & Information Technology, 25*(2), 91-97.

[6] Hassenzahl, M. (2018). The thing and I: Understanding the relationship between user and product. In *Funology 2*. Springer, Cham, pp. 301-313.

[7] Han, S. R. (2017). *Business trends and implications of Chatbot*. Digieco Report.

[8] Keszocze, O., and Harris, I. G. (2019). Chatbot-based assertion generation from natural language specifications. *In 2019 Forum for Specification and Design Languages(FDL)*, IEEE, 1-6.

[9] Kim, J. T., Jung, S. R., and Chung, H. (2017). Trends of messenger-based chatbot technology and military application. *Defense & Technology, 459*, 118-127.

[10] Kim, J. W. (2012). *Human computer interaction*. Ahn Graphics.

[11] Konstan, J. A., and Riedl, J. (2012). Recommender systems: From algorithms to user experience. *User Modeling and User-Adapted Interaction, 22*(1-2), 101-123.

[12] Mamani, J. R. C., Álamo, Y. J. R., Aguirre, J. A. A., and Toledo, E. E. G. (2019). Cognitive services

to improve user experience in searching for academic information based on chatbot. *In 2019 IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing(INTERCON)*, IEEE, 1-4.

[13] Moon, J. H., Lim, S. T., Park, C. L., Lee, I. S., and Kim, J. W. (2008). Conceptual study on user experience in HCI: Definition of UX and introduction of a new concept of CX (Co-Experience). *Journal of the HCI Society of Korea, 3*(1), 9-17.

[14] Megha, D. (2019). [Online] Available at https://medium.com/searce/paving-theway-to-a-chatbot-revolution-c7ea21fa2f2f

[15] Morville, P. (2006). *Search 2.0: The evolution of discovery(Ambient findability)*. Seoul: Hanvit Media.

[16] Morville, P. (2004). *User experience design*. Ann Arbor: Semantic Studios LLC.

[17] Oh, S. W. (2016). (ro) Bot as media An experimental discussion on news chatbot. *Korean Journal of Communication and Information, 79*, 70-103.

[18] Park, D. A. (2017). A study on conversational public administration service of the Chatbot based on artificial intelligence. *Journal of Korea Multimedia Society, 20*(8), 1347-1356.

[19] Roto, V. (2006). *Web browsing on mobile phones: Characteristics of user experience*. Helsinki University of Technology, pp. 834-837.

[20] Shevat, A. (2017). Designing bots: Creating conversational *experiences*. O'Reilly Media, Inc., p. 120.

[21] Sidenko, I., Kondratenko, G., Kushneryk, P., and Kondratenko, Y. (2019). Peculiarities of human machine interaction for synthesis of the intelligent dialogue Chatbot. *In 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications(IDAACS)*, IEEE, 1056-1061.

[22] Tanya, R., Tanlamai, U., and Cho, N. (2017). Design guidelines of a web portal's functions and layouts to fit individual user's personality traits. *International Journal of Innovation and Learning, 22*(3), 271-292.

# ◆ About the Authors ◆

**Yeji Lim**

Yeji Lim has a master's degree in MIS at Hanyang University and is a Ph.D candidate in Business Administration. Interesting research fields are business intelligence, Big data analytics, IoT and the Family business.

**Jeonghun Lim**

Jeonghun Lim graduated from the Y.E.S. MBA at Hanyang University and is a Ph.D candidate in Business Administration. Interesting research fields are AI, Big data, IoT and the Family business.

**Namjae Cho**

Namjae Cho is a professor of MIS at the School of Business of Hanyang University, Seoul, Korea, He received his doctoral degree in MIS from Boston University, U.S.A. He has published research papers in journals including Industrial Management and Data Systems, Computers and Industry, International Journal of Information Systems and Supply Chain, Journal of Data and Knowledge Engineering. He also published several books including "Supply Network Coordination in the Dynamic and Intelligent Environment(IGI Global)" and "Innovations in Organizational Coordination Using Smart Mobile Technology(2013, Springer)". He consulted government organizations and several multinational companies. His research interest includes technology planning and innovation, analysis of IT impacts, strategic alignment and IT governance, knowledge management and industrial ICT policy, design thinking, and the management of family business.

Submitted: August 8, 2019; 1st Revision: February 29, 2020; 2nd Revision: August 28, 2020; Accepted: November 4, 2020