Aalto University
School of Science
Degree Programme in Industrial Engineering and Management

Tung Son Nguyen

Potential effects of chatbot technology on customer support: A case study

Master's Thesis
Espoo, April 30th, 2019

Supervisor: Professor Paul Lillrank
Advisor: Doctor Timo Smura

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Tung Son Nguyen

Title Potential effects of chatbot technology on customer support: A case study

Master's programme Industrial Engineering and Management

Major/Code Operations and Service Management / SCI 3049

Supervisor Professor Paul Lillrank

Advisor Doctor Timo Smura

Date 30.04.2019 **Number of pages** 69 **Language** English

Abstract

This study analyzed an experiment with using a chatbot for the customer support department at the case company. A modified version of the updated DeLone and McLean information system success model was utilized to evaluate potential effects of the chatbot on the operation of the customer support. Five dimensions of the model were measured before and after the chatbot implementation and then compared to determine if the chatbot can help improve the customer experience with the customer support of the case company.

Responses from 60 customers who had used the chatbot were obtained through a web-based survey. Results indicated that the addition of a chatbot to a traditional customer support model can improve customer experience, mainly on responsiveness measure, while maintain a similar level on information quality, system quality and user satisfaction dimensions. Results also suggested that unsuccessful chatbot attempts which require further human involvement may not necessarily worsen customer experience as many expect.

Based on the experiment, the study also provided three suggestions for firms when planning to adopt chatbot technology. First, the potential of chatbot should not be overestimated, it cannot replace human agents completely in customer support. Second, chatbot should handle only simple enough tasks and leave the more complex and trickier ones to human. And third, building a chatbot is a continuous process that requires careful resource planning not only for the initial development but also for the later stage of analyzing and turning conversations of the chatbot.

Keywords chatbot, customer support, information system success model

Acknowledgement

First and foremost, I would like to thank professor Paul for his quick guidance and patience throughout the whole development of this thesis.

Also I would like to personally thank my advisor Timo Smura for his constant support, without his help to push through some obstacles at the case company, this thesis might never have been completed.

Finally, a big thank you goes to my family and my friends Chocopie, Tram, Esther, Shaun, Allison who gave me unconditional supports during my hard times when doing this thesis.

Table of contents

List of abbreviations

AI	Artificial Intelligence
AIML	Artificial Intelligence Markup Language
FAQ	Frequently Asked Questions
KPI	Key Performance Indicator
IS	Information System
IT	Information Technology
WWW	World Wide Web

Introduction

Chatbot technology is emerging as one of the hot topics in recent years. The technology has been developed since 1966 when Joseph Weizenbaum presented a chatbot named Eliza. Today, it is progressively becoming popular on social media and messaging applications. In April 2018, Facebook reported that 100,000 bots had been created on their Messenger platform, within only one year after its introduction (Johnson, 2017). Meanwhile, research firm Canalys predicted that 56.3 million smart speakers, a special type of chatbot using voice, will be sold in 2018, up from an estimated 33 million units shipped in 2017 and 6 million units shipped in 2016 (Canalys.com, 2018). Besides the rapid growth of messaging platforms such as Facebook Messenger or Slack and voice services like Amazon Alexa or Apple Siri, the recent advances in Artificial Intelligence (AI) with new techniques such as machine learning or deep learning also helped dramatically improve the quality of chatbot on parsing human language, understanding contexts, composing replies or making decisions. Currently, chatbot is mainly used for two purposes, answering inquiries from users and executing more complex transactions like ticket booking.

The rise of chatbot is catching attention from many companies, especially in the customer service context (e.g. Cui et al., 2017; McGrath, 2018; Peterson, 2017; Chung et al., 2018; Flaiz, 2018). According to Gartner (2018), more than half of companies have already invested in chatbot and by 2020, chatbot will power 25% of all customer service operation. In a survey conducted by Oracle of 800 senior marketers and sales professionals across Europe, the Middle East and Africa, 80 percent of brands said they already used chatbot or planned to use it to serve customers by 2020 (Brynjolfsson and Mcafee, 2017).

The idea of using chatbot to replace or boost human workers in a customer contact center is under consideration in many organizations. This is understandable from the business perspective, since a chatbot can work anytime, twenty-four hours a day, seven days a week, and the initial cost for development can be quickly covered by much lower operating cost. But chatbot is not just about replacing human agents, it can even make human agents more productive by handling time-consuming repetitive tasks like querying data, scheduling meetings, or triggering transactions. For example, a chatbot integrated into an airline's website can answer queries about fees, rebook flights, and suggest additional services such as hotel and car reservations. Even if the chatbot cannot finish these exchanges, it is still able to collect initial information like customer's name or reservation number and transfer to a customer service representative to continue handling the case, saving considerable time for the company's call center. The potential improvement in terms of cost, efficiency and consistency is really driving organizations to implement chatbot to cover various tasks previously reserved for humans.

However, on the other side, there are also skepticisms about the use of chatbot, mainly related to the high failure rate in interactions with customers and the lack of personalized customer experience. In 2017, it was reported that Facebook Messenger bots failed to complete 70 percent of user requests, meaning only 30 percent of all conversations were ended successfully without the additional help of a human (Bozorgzadeh, 2017). During the same year, a travel search engine organized a survey among British people to learn what they know about chatbot and what they expect from it. The result was not positive with only 5 percent would consider bots more reliable than a human, and 75 percent had at least one concern, including data security, receiving incorrect answers, being misunderstood and the possibility of the bot's responses being somehow manipulated (Kayak.co.uk, 2017). These concerns surely may make companies hesitating on using chatbot for their operation.

When planning to use a new technology with high upfront cost like chatbot, it is important to understand what values it can bring to the table. A way to evaluate the quality of a technology is by looking at the customer experience with it. With that purpose in mind, the study aims to analyse potential impacts of chatbot on customer experience. The result of this study can give an idea if chatbot can be an effective channel for customer support operation compared to more traditional channels like Frequently Asked Questions (FAQ) or email. Hopefully, this study can be used as a reference for companies who are examining the possibility of using chatbot technology for their customer support operation.

Research gap

Until now, studies conducted on chatbot have mostly focused on using different techniques to improve capabilities and effectiveness of chatbot, or exploring user experience with chatbot in general experiment. Still, when it comes to understanding the effect of chatbot as a component of customer support in real business context, the literature that addresses this topic is still limited. For example, Xu et al. (2017) evaluated the quality of responses from chatbots measured by human judgment. However, the focus is comparing between different chatbot developing techniques, rather than between chatbot and humans. Meanwhile, Brandtzaeg and Følstad (2017) identified some key motivational factors of individuals for using chatbot, but the participants were self-selected and regarded as early adopters, therefore they may not present opinions of general users.

The aim of this study is to fill that gap by evaluating potential effects of chatbot on customer experience with a case company's customer support. Some important dimensions of customer experience would be measured before and after the implementation of a chatbot and then compared to determine if a chatbot can help improve that critical key performance indicator (KPI) of a customer support function.

Research question and hypothesis

The purpose of this study is to analyse potential effects of a chatbot in the customer support context, in order to help a company decide on adopting this technology for its customer support department. Furthermore, the notion of customer acceptance of chatbot is evaluated under multiple criteria to improve visibility over how a chatbot should be developed. In order to reach these objectives, two main research questions were created:

Research question 1: How does adding a chatbot to a traditional customer support (with FAQ and email channel) affect customer experience? The hypothesis was that with the addition of a chatbot, customer experience can be improved.

Research question 2: How may unsuccessful interactions of a chatbot affect user experience? This research question was studied with the assumption that the unsolved cases would be escalated and handled properly by human agents afterwards. The hypothesis was that unsuccessful chatbot attempts which require further human involvement worsen customer experience.

Research methodology

This study based on the updated DeLone and McLean information system (IS) success (2003) to define dimensions and measures capable of influencing customer experience with customer support. The data were collected by sending a questionnaire to customers who had used sorely the traditional support system (with FAQ and email) and then later to users who had used the chatbot. After that, the collected data between the two groups would be compared to find any significant differences in selected measures.

Structure of the thesis

This study is organized into five chapters. The first chapter gives a brief background of the study, identifies current research gap, defines research questions and research methodology. The second chapter reviews literature around chatbot technology, the development of the DeLone and McLean IS success model and base on that build a suitable model for the study. The third chapter presents the research methodology by describing the case study, how the questionnaire was designed, and the data were collected. The fourth chapter shows the research findings by applying statistical methods to compare data of groups of users. The final chapter discusses the findings to answer the research questions, as well as gives recommendations for the case company.

Literature review

Customer support

Loomba (1998) defined customer support as a set of activities to assure customers can use a product without trouble over its life cycle and is critical for maintaining customer satisfaction and loyalty (Armistead and Clark, 1992; Goffin and New, 2001). Particularly in high-tech industries with strong competition, companies realized that customer support can be the key to differentiate themselves from competitors (Loomba, 1998; Negash et al., 2003).

Goffin and New (2001) proposed seven key components of customer support: installation, user training, documentation, maintenance and repair, on-line support, warranty, and upgrades. The authors also pointed a major shift in the way each component influenced customer service in the past years. Nowadays, new technologies have managed to deliver more reliable but also more complex products. As a result, user training and on-line support have become more important elements of customer support to cope with more increased product complexity. Meanwhile in the past, since the failure rate of products was higher, reliable maintenance and repair were more crucial for companies (Goffin and New, 2001).

Anton (2000) summarized the four fundamental information needs of a customer when contacting customer support of a company:

1. The customer has a query and look for an answer to continue.
2. The customer needs the company to do a certain request.
3. The customer needs help from the company to solve an issue they have with the product.
4. The customer is not satisfied, or even angry with the product and demands an action from the company to settle the situation immediately.

In the past, customer support was mainly offered via face-to-face contact or over the phone. Gradually with new emerging technologies, other communication channels have been introduced and widely used including company webpage, email, forum, social network, and chat. Anton (2000) observed two phenomena after this evolvement. First, companies tend to quickly add more channels to contact their customer support, but then mostly cannot fill enough personnel in time to handle overwhelming influx of customer requests. And second, the appearance of new communication channels does not mean that old channels can be disbanded. Instead, customers want to have all of them available so they can choose the more convenient ones for them.

It has turned out to be progressively costly for organizations to keep up the quality of customer support across all channels. Hardalov et al. (2018) specified two main challenges here: (1) in order to be able to handle a target channel, each agent needs to acquire some certain specific trainings, which takes time and increase the cost; (2) customers have become more demanding with the availability of customer support, so it is challenging for firms to have agents available for 24×7 customer support operation. In this situation, chatbot stands out as a new potential channel which is accessible and efficient while maintaining cost at an acceptable level (Følstad et al., 2018).

Chatbot

Definition

The terminology “chatbot” emerged from the system CHATTERBOT, which was used by Mauldin (1994) to describe the system with the goal to pass the Turing test back in 1950 and therefore, could be considered as having the ability “to think”. Since then, many terms have been introduced and used interchangeably, such as conversational agent, intellectual agent, virtual agent, artificial conversational entity, virtual assistant or intelligent virtual assistant. All of them can be considered as synonyms of chatbot - the most commonly used term nowadays.

Since there have been several terms in use, their definitions are various also. For example, in the early days, Hutchens and Alder (1998) described conversation simulators as “computer programs which give the appearance of conversing with a user in natural language”. Schumaker et al. (2007) characterized a chatbot as “a system that seeks to mimic conversation rather than understand it”. Griol et al. (2013) explained a conversational agent as “a software that accepts natural language as input and generates natural language as output, engaging in a conversation with the user”. More recently, both Dale (2016) and Ciechanowski et al. (2018) took the term chatbot to refer to any software application that uses natural language to make a conversation with a human. In this study, the term chatbot is used to cover all other terms, and follow the definition of Date (2016) and Ciechanowski et al. (2018), as it clearly presents the underlying principle of every chatbot: the ability to interact with a human via natural language (either by auditory or textual means), understand the conversation and give responses properly.

Historical review

Based on the above definition, one of the earliest software applications that can be considered as a chatbot was actually invented a long time ago. In 1966, Joseph Weizenbaum introduced a chatbot named Eliza, it was designed to emulate a psychiatrist and was able to establish a conversation with human beings using basic natural language processing techniques (Weizenbaum, 1966). Eliza used the keyword mapping technique so that some keywords will trigger particular responses. Even being quite simple, it surprisingly succeeded in giving the illusion of understanding the user's issues and made many people believed they were talking with another human (this effect is currently called the "Eliza effect"). Interestingly, Weizenbaum (1976) even revealed that his own secretary asked him to leave the room so that she could have a private conversation with Eliza.

Human: Men are all alike.

ELIZA: In what way?

Human: They're always bugging us about something.

ELIZA: Can you think of a special example?

Human: Well, my boyfriend made me come here.

ELIZA: Your boyfriend made you come here?

Human: He says I'm depressed much of the time.

ELIZA: I'm sorry to hear that you are depressed much of the time.

Human: It's true. I'm unhappy.

ELIZA: Do you think coming here will help you not to be unhappy?

Human: I need some help, that much seems certain.

ELIZA: What would it mean to you if you got some help?

Human: Perhaps I could learn to get along with my mother.

ELIZA: Tell me more about your family.

An ELIZA sample conversation

Since then, many new types of chatbot have been developed and this topic has kept gaining attention from researchers, marked with the introduction of Loebner Prize in 1991, an annual competition between computer programs to be the most human-like using the unrestricted Turing test. In the history of this contest, one of the most famous chatbots is the open-source chatbot ALICE, the winner in 2000, 2001, 2004 (Shawar and Atwell, 2007; Hettige and Karunananda, 2015; Reshmi and Balakrishnan, 2016). Even if ALICE was unable to pass the Turing test, it built the foundation for the development of Artificial Intelligence Markup Language (AIML), which is used to define pattern-matching rules that connect user inputs with appropriate outputs (Radziwill and Benton, 2017). As of 2007, 50,000 matching patterns were built up by community for ALICE "brain" (Shawar and Atwell, 2007). Another notable chatbot from this competition is Mitsuku, the winner in 2013, 2016 and 2017 (Brandtzaeg and Følstad, 2017).

The next milestone for conversational bots was created by IBM through the Watson AI project. In 2007, a team at IBM started developing a computer system with the target of winning the American TV show Jeopardy!. There were three challenges of Jeopardy! that IBM faced: (1) the broadness of the questions with rich and varied natural language expressions; (2) the requirement of highly accurate and confident answers; (3) the time pressure to find the answer and buzz quickly to beat competitors (Thompson, 2010; Ferrucci et al., 2013). After 4 years of development, Watson finally conquered its grand challenge in 2011 by beating the two highest ranked players in a two-game Jeopardy! match. This event marked a big step towards a vision in which computer programs can understand, process and respond to humans properly (Markoff, 2011).

Later, the early 2010s marked the rise of virtual assistants including Apple Siri, Microsoft Cortana, Google Assistant, Amazon Alexa and others (Dale, 2016). Following up, a number of platforms were introduced by tech companies to support the creation of conversational agents, such as IBM Watson, Microsoft Bot Framework or DialogFlow (Følstad et al., 2018). Besides that, another noteworthy event is the opening of Facebook's Messenger Platform in 2016 (Yeung, 2016). This allowed developers to design and develop bots for Facebook users to interact with. Considering Facebook Messenger had more than 1.2 billion monthly active users in that year (Cohen, 2017), chatbot technology has gained significant popularity among digital users, as well as attracted the attention from other tech companies.

As shown in this brief summary, a lot of progress has been made since the early days of chatbot technology. This does not mean however those current solutions are without limitations which will be highlighted in the latter section.

Service quality

Service quality has always been a main interest in marketing research. Early on, researchers had mostly focused on service quality and its impact in non-Internet-based context. The period was heavily marked by the introduction of SERVQUAL instrument developed by Parasuraman, Zeithaml, and Berry (1988). This instrument was constructed to evaluate customer perceptions of service quality.

SERVQUAL instrument includes five measures: tangibles, reliability, responsiveness, assurance, and empathy (Table 1). At first, it was designed to measure the quality of services with face-to-face interactions. Multiple studies have been conducted to verify, adapt, develop, and complete that scale in various domain (e.g. healthcare, banking, retailing, telecommunications, information systems, public services) (Carman, 1990; Lee and Ulgado, 1997; Van der Wal et al., 2002; Jiang et al., 2000; Cook and Thompson, 2000).

Table 1: Five measures of SERVQUAL (Parasuraman, Zeithaml, and Berry, 1988)

Tangibles	Physical facilities, equipment, and appearance of personnel
Reliability	Ability to perform the promised service dependably and accurately
Responsiveness	Willingness to help customers and provide prompt service
Assurance	Knowledge and courtesy of employees and their ability to inspire trust and confidence
Empathy	Caring, individualized attention the firm provides its customer

Later, with the growing importance of technology in customers' interactions with firms, many researchers shifted their focus to a new form of service, the electric service. Since the introduction of the World Wide Web (WWW), an ever-increasing number of organizations have been utilizing it to give customers direct services and related information. The process of creating and delivering services has advanced from conventional communication channels to Web-based information systems (Tan et al., 2003). The interest in measuring the service quality in this context promoted further SERVQUAL research and led to the introduction of new conceptual frameworks such as WebQual (Loiacono et al., 2002) (Table 2) or E-Servqual (Parasuraman et al., 2005) (Table 3). These frameworks cover multiple attributes capable of affecting the electronic service quality including information quality, website aesthetics, purchase process, website convenience, product selection, merchandise availability, price offerings, website personalization, system availability, timeless of delivery, order accuracy, delivery condition, service level, return handling/policies, security, and privacy (Zeithaml et al., 2000; Loiacono et al., 2002; Parasuraman et al., 2005; Loiacono et al., 2007; Holloway and Beatty 2008).

Table 2: WebQual framework constructed by Loiacono et al. (2002)

Dimension	Measure	Description
Usefulness	Informational fit-to-task	The information provided fits what users need to complete their tasks
	Tailored communications	Users receive tailored information that help users complete their tasks
	Trust	Users trust that their personal information is safe when interacting with the website
	Response time	Users feel that the website responds promptly without long waiting time
Ease of use	Ease of understanding	Contents of the website is easy to understand
	Intuitive operations	The website is easy to use and navigate
Entertainment	Visual appeal	Users feel that the website is appealing visually
	Innovativeness	Users feel that the website is creative and unique
	Emotional appeal	Users feel pleasing when using the website
Complementary relationship	Consistent image	The website reflects the image of the company consistently
	On-line completeness	The website allows users to complete all or most of important transactions with the company

	Relative advantage	Compared to other channels, users prefer to use the website to interact with the company compared channel

Table 3: E-Servqual framework constructed by Parasuraman et al. (2005)

Measure	Description
Efficiency	The ease and speed of accessing and using the site.
Fulfillment	The extent to which the site's promises about order delivery and item availability are fulfilled.
System availability	The correct technical functioning of the site.
Privacy	The degree to which the site is safe and protects customer information.
Responsiveness	Effective handling of problems and returns through the site.
Compensation	The degree to which the site compensates customers for problems.
Contact	The availability of assistance through telephone or online representatives.

However, the above approaches were heavily built upon the concept of service quality from marketing research. This study focuses on using a new IT technology as a platform to pursue the organization's goals of reducing operation cost and at the same time, keeping or improving the service quality of the customer support. Therefore, some of the variables mentioned previously related to price, merchandise or delivery become less relevant in this circumstance. Another approach that utilizes IS theory was proposed to determine and measure the effectiveness of the new technology in this context.

DeLone and McLean IS success model

Keen (1987) explained the mission of IS as: "the effective design, delivery, use and impact of information technologies in organizations and society". Based on Keen's view of IS, DeLone and McLean (1992) analyzed 180 articles and classified over 100 measures into six major dimensions of IS success:

1. System quality – the desired engineering-oriented attributes of an IS, which focus on the performance and usability of the system
2. Information quality – the desirable attributes of an IS's outputs, which focus on the quality and usefulness of the information produced by the system
3. Use – the consumption and utilization of the output of an IS
4. User satisfaction – the level of satisfaction when utilizing an IS
5. Individual impact – the effect of an IS on each individual user
6. Organisational impact – the effect of an IS on the performance of its organization

As presented in Figure 1, system quality and information quality simultaneously affect both use and user satisfaction. Meanwhile, use and user satisfaction influence each other and both have effect on individual impact, which may eventually create some organizational impact.

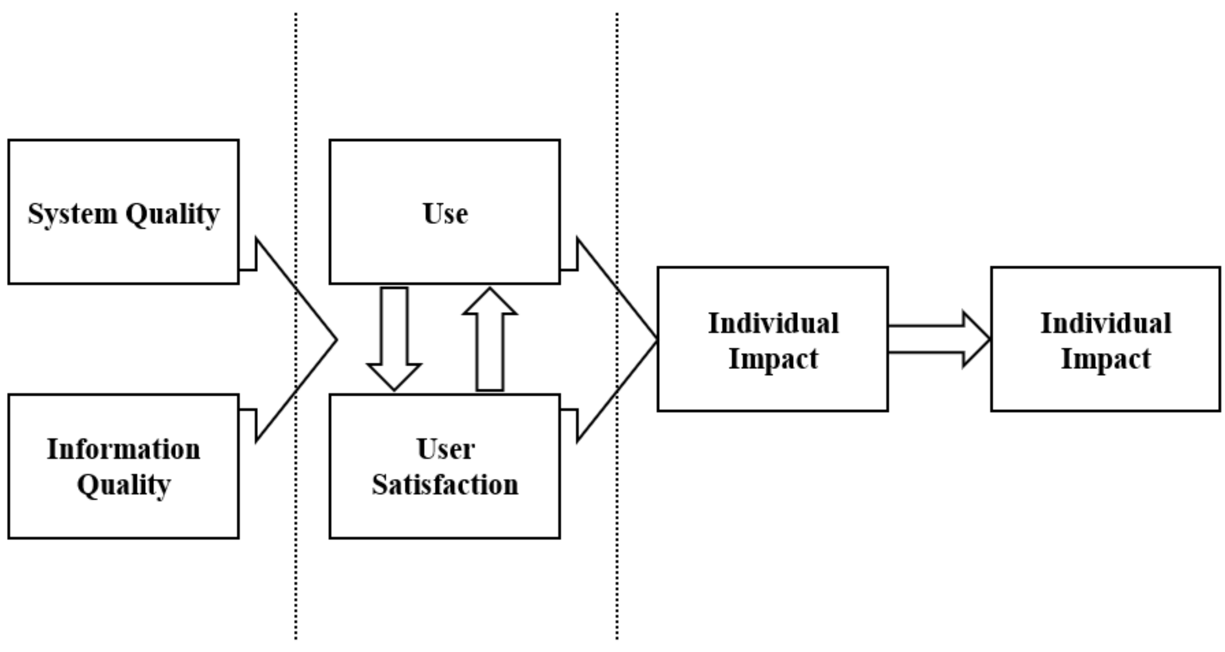

Figure 1: DeLone and McLean IS success model (1992)

Based on the feedbacks from other researchers, in 2003 DeLone and McLean updated their model of IS success (Figure 2) by adding a new dimension – service quality. This dimension was introduced to address the influence of service as a contributor to IS success, as many IS organizations transformed into a hybrid model of being both information provider (delivering an information product) and service provider (delivering support to end users) (DeLone and McLean, 2003). While system quality and information quality focus on the quality of the outputs of an IS, service quality helps measure the perceived quality of the service provided by IS department and IT personnel. The IS success model now builds upon three quality dimensions: system quality, information quality and service quality**.** Use dimension was further classified into two, intention to use and actual use. Meanwhile, individual impact and organizational impact were combined into net benefits, which have a positive impact on intention to use / use and user satisfaction.

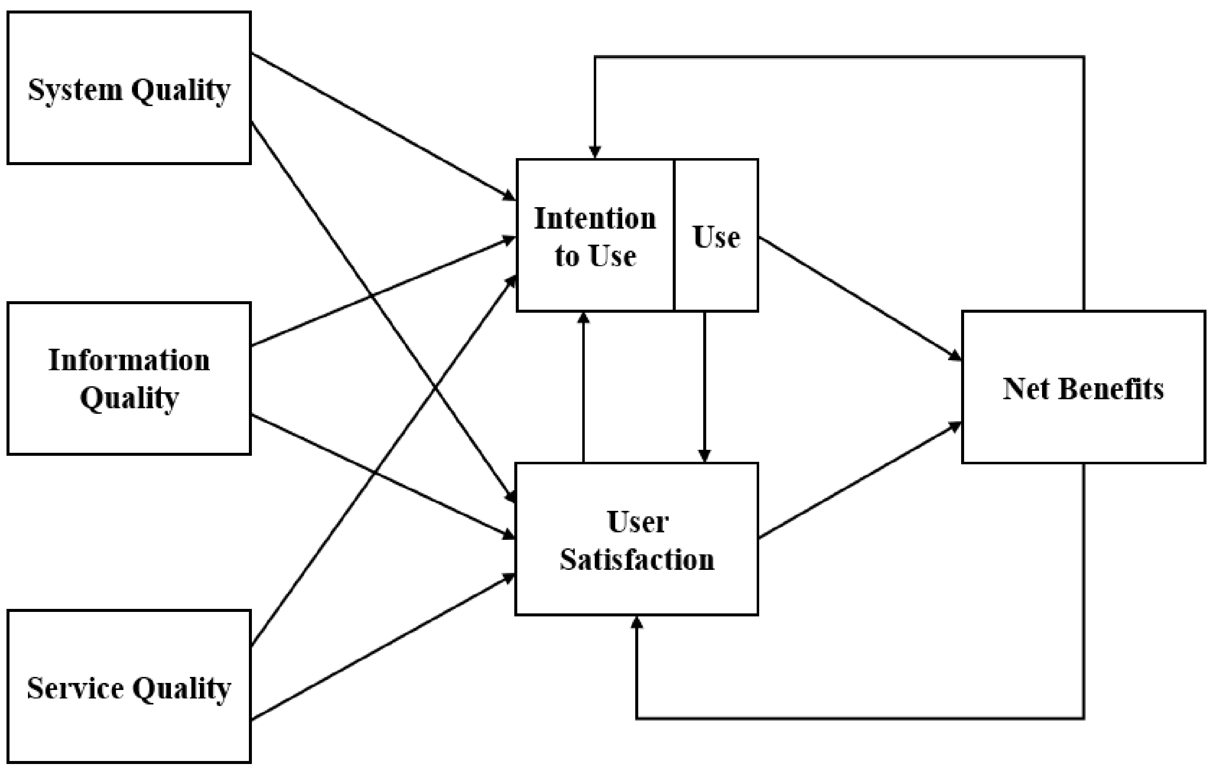

Figure 2: The updated DeLone and McLean IS success model (2003)

According to DeLone and McLean (2003), this model can be used to evaluate the success of a whole IS or even a single component of IS. They suggested that if the goal is to measure the success of a whole IS, service quality may become the most important variable. Meanwhile, information quality or system quality may be more important if the scope is narrowed down to a single component. The authors did not enforce any sets of measures for these dimensions, but instead they advised that depending on the objectives and context of the study, researchers should select appropriate dimensions and measures to reflect important characteristics of the study object. This can be seen as one of the main advantages of the model since it can be used in various contexts (Petter et al., 2008).

Information quality

Information quality represents the quality of the information delivered by an IS as perceived by its users (DeLone and McLean, 1992). There have been different approaches proposed to define information. While some researchers used information and data interchangeably, others made a clear distinction between them (Lillrank, 2002). From a hierarchical view, Lillrank (2002) explained data as a factual content of information. Without a context, a number delivers no meaning and thus it is just data. Meanwhile if that number is clarified as a net profit of a company, it can be understood and used by relevant audiences for practical purposes. The author further formalized this concept and developed the $M = f(D, C)$ equation, in which D is a piece of data, C is the context of the data, f is how relevant knowledge is used to analyze the data in the context, and then M is the result of the process of transforming data into information.

Based on that idea, Lillrank (2002) postulated two approaches to evaluate information quality. On one hand, the quality of information can be subject to technical quality, it is defined by comparing the intention of sender and the result of the process to transmit information to receiver. In this approach, quality of data and context need to be agreed upon in advance, while receiver should have a certain level of knowledge and competence. The expectation is that the receiver will capture correctly the meaning as intended by the sender. On the other hand, when data, context and ability and expertise of receiver are unstable, the quality of information must be negotiated and agreed between sender and receiver. Because of the difference between what producer can offer and what receiver may require, both parties need to work together during the transmission to find a common understanding, which is called negotiated quality by the author.

On this study, information quality is the focus and it should not be confused with data quality. The study is conducted to analyze various dimensions of quality as perceived by customers in the customer support context, rather than assessed sorely by the organization itself. Therefore, data quality is not an appropriate measurement here as it does not cover how data are interpreted and transformed into information, the process depends greatly on customers rather than can be fully under control of the organization. Also, the quality of information should be approached as negotiated quality, since while data and context can be somehow calculated and controlled by the organization, the knowledge and competence of customers are great unknown with significant impact on the quality of information delivered. With that in mind, various factors have been explored and analyzed as the source of information quality perception (Table 4). Some of the most recognized measures of are accuracy, completeness, relevance, timeliness and understandability.

For this study, accuracy, completeness, timeliness and understandability were selected to measure the quality of information produced by the chatbot. Accuracy and completeness were combined under accuracy since the latter normally comes side by side with the former (Miller, 1996). This measure reflects the correctness and the degree of up-to-date of the information that the chatbot delivered to users and how content users were with these attributes. Similarly, information completeness indicates the degree to which the delivered information was sufficient and comprehensive for the users to fulfill their needs. This quality is hard to be assessed since it is a relative concept, a piece of information may be complete under one user's perspective but may not be enough in the view of another (Nelson et al., 2005). Lastly, understandability ensures the information is clear, straightforward and easy to understand (Wang and Strong (1996). The last two measures are quite essential in this research since the amount of information the chatbot can deliver at a time is quite limited. Therefore, the content to be delivered needs to be short but concise to make sure users can receive and understand all necessary information to solve their problems.

Table 4: Common measures of information quality

Measure	References
Accuracy	Bailey and Pearson (1983), Baroudi and Orlikowski (1988), Huh et al. (1990), Miller (1996), Strong et al. (1997), Rai et al. (2002), Nelson et al. (2005), Gable et al. (2008)
Availability	Sedera et al. (2004), Gable et al. (2008)

Completeness	Bailey and Pearson (1983), Baroudi and Orlikowski (1988), Huh et al. (1990), Miller (1996), Wang and Strong (1996), DeLone and McLean (2002), Rai et al. (2002), Iivari (2005), Nelson et al. (2005), Gable et al. (2008)
Coherence	Miller (1996)
Conciseness	Sedera et al. (2004), Gable et al. (2008)
Consistency	Iivari (2005)
Currency	Bailey and Pearson (1983), Huh et al. (1990), Iivari (2005), Nelson et al. (2005)
Format	Doll and Torkzadeh (1988), Miller (1996), Rai et al. (2002), Sedera et al. (2004), Iivari (2005), Nelson et al. (2005), Gable et al. (2008)
Language	Bailey and Pearson (1983)
Security	Miller (1996), Wang and Strong (1996), Strong et al. (1997), DeLone and McLean (2002)
Precision	Bailey and Pearson (1983), Baroudi and Orlikowski (1988), Rai et al. (2002), Iivari (2005)
Relevance	Bailey and Pearson (1983), Miller (1996), McKinney et al. (2002), DeLone and McLean (2002), Sedera et al. (2004), Gable et al. (2008)
Reliability	Bailey and Pearson (1983), Baroudi and Orlikowski (1988), McKinney et al. (2002)
Scope	McKinney et al. (2002)
Timeliness	Bailey and Pearson (1983), Doll and Torkzadeh (1988), Miller (1996), Strong et al. (1997), Nelson et al. (2005), Gable et al. (2008), Wang and Strong (1996), Strong et al. (1997), Gable et al. (2008)

Understandability	Wang and Strong (1996), McKinney et al. (2002), Sedera et al. (2004), Gable et al. (2008)
Usability	Rai et al. (2002), Sedera et al. (2004), Gable et al. (2008)
Volume of output	Bailey and Pearson (1983), Iivari (2005)

System quality

An overview of most common measures for system quality is shown in Table 5. Among them, ease of use is the most common one due to a large number of studies driven by the technology acceptance model developed by Davis in 1989 (Petter et al., 2008). Perceived ease of use refers to "the degree to which a person believes that using a particular system would be free of effort" (Davis, 1989). Users are more likely to accept a new system or technology which they feel it is easy to use. Therefore, ease of use is arguably one of the most crucial factors that decide the success of a new system.

Meanwhile, other measures are either not significant enough, or not measurable from users' point of view (accessibility, flexibility, integration, recoverability customization), or somewhat overlap with other measures of service quality dimension (which was added later to the updated DeLone and McLean model) (reliability, response time), or too similar with ease of use measure (usability). Because of that, ease of use would be the only measure selected for system quality dimension. This is reasonable for this study since system quality dimension was originally developed to measure the technical performance of computer systems rather than how end users feel about these systems (DeLone and McLean, 1992).

Table 5: Common measures for system quality

Measure	References
Accessibility	Bailey and Pearson (1983), Srinivasan (1985), McKinney et al. (2002), Nelson et al. (2005), Gable et al. (2008)
Adaptability	DeLone and McLean (2003)
Availability	DeLone and McLean (2003)
Convenience	Iivari (2005)
Customization	Sedera and Gable (2004), Gable et al. (2008)
Data accuracy	Gable et al. (2008)
Data currency	Gable et al. (2008)
Ease of use	Belardo et al. (1982), Bailey and Pearson (1983), Doll and Torkzadeh (1988), Davis (1989), Seddon and Kiew (1996), Rai et al. (2002), Sedera and Gable (2004), Gable et al. (2008)
Entertainment	McKinney et al. (2002)
Error-proneness	Srinivasan (1985)
Flexibility	Bailey and Pearson (1983), Sedera and Gable (2004), Iivari (2005), Nelson et al. (2005), Gable et al. (2008)
Integration	Bailey and Pearson (1983), Sedera and Gable (2004), Iivari (2005), Nelson et al. (2005), Gable et al. (2008)
Interactivity	McKinney et al. (2002)
Navigation	McKinney et al. (2002)

Reliability	Balardo et al. (1982), Srinivasan (1985), DeLone and McLean (2003), Nelson et al. (2005), Gable et al. (2008)
Response time	Balardo et al. (1982), Bailey and Pearson (1983), Srinivasan (1985), DeLone and McLean (2003), Iivari (2005), Nelson et al. (2005)
Recoverability	Iivari (2005)
Sophistication	Sedera and Gable (2004), Gable et al. (2008)
Usability	Bailey and Pearson (1983), McKinney et al. (2002), Rai et al. (2002), DeLone and McLean (2003)
User friendliness	Seddon and Kiew (1996)

Service quality

As briefed shortly before, service quality is the dimension added later to the updated DeLone and McLean IS success model in 2003. This change followed many researchers' appeals in realization of the growing service role of the IS function at that time. Petter et al. (2008) described service quality in the updated DeLone and McLean model as "the quality of the system that support users receive from the IS department and IT support personnel".

To measuring IS service quality dimension, SERVQUAL, a popular measurement instrument from marketing literature, was proposed by several researchers (Jiang et al., 2002). SERVQUAL can be used to evaluate the service quality of IT departments by measuring the differences between users' perceptions and expectation of the IT department (Petter et al., 2008). This capability of SERVQUAL matches with the purpose of this study, as it can take and combine the whole customer support with both IS components and support personnel as a single unit instead of individuals and measure the degree to which this unit can meet or exceed what users expect from the service. While the original instrument contains five measures, this study considered only four measures: reliability, responsiveness, assurance, and empath. The remaining measure tangibles was omitted since the interactions between the customer support and the customers happened only in the Internet environment without any tangible channels.

Intention to Use/Use

In the updated model, DeLone and McLean (2003) further clarified use dimension, which is a broad context that should be considered from several perspectives. They suggested that depending on the context, researchers should choose either “intention to use”, which is an attitude, or “use”, which is a behavior, to represent the dimension. This modification is a response to an argument of Seddon and Kiew (1996) that in case of mandatory use, usefulness may be a better success measure than use. Seddon and Kiew’s approach to address usefulness is similar to the concept of perceived usefulness in technical acceptance model by Davis (1989). This model developed two variables, perceived ease of use and perceived usefulness, which were proved to be determinants of attitude toward use, intention to use, and actual use (Urbach and Müller, 2012). Since perceived ease of use had been used for system quality dimension, perceived usefulness was chosen to represent use dimension.

User Satisfaction

User satisfaction dimension measures the users’ level of satisfaction with an IS. This dimension was traditionally used as a measure of IS success (Bailey and Person, 1983) and it has been measured indirectly via information quality, system quality, and other variables (Rai et al., 2003). User satisfaction can be a very useful measurement when the use of an IS is mandatory and therefore the volume of use is not an applicable measure of IS success (Urbach and Müller, 2012).

Various measurement instruments have been developed to measure user satisfaction, some of them are presented in Table 6. However, some of them also includes measures of information quality, system quality, and service quality, rather than only focusing on user satisfaction. Others like repeat purchases or repeat visits are easy to be measured from technical standpoint, however they are not appropriate for this study since using the customer support is mandatory for customers to solve their problems. Since user satisfaction is a subjective feeling of users, it is reasonable to use self-reporting measures to understand how satisfied users are after using an IS to fulfill their needs.

Baroudi and Orlikowski (1988) proposed an idea that a single measure can be used to measure user satisfaction if the purpose of a study is to access overall satisfaction as a whole rather than in particular areas (Rai et al., 2002). Based on that idea and their worry about survey length and respondent convenience, Rai et al. (2002) used a single measure to measure how users rated their overall satisfaction with the IS. Since then, this approach has been adopted by many researchers with different sets of questions. This study would utilize the same approach since its goal in this dimension was to analyse the overall satisfaction of users with the customer support as a whole.

Table 6: Common measures for user satisfaction

Measure	**References**
A list of 39 factors affecting user satisfaction	Ives et al. (1983)
Accuracy	Doll and Torkzadeh (1988)
Adequacy	Seddon and Yip (1992), Seddon and Kiew (1994), Almutairi and Subramanian (2005)
Content	Doll and Torkzadeh (1988)
Ease of use	Doll and Torkzadeh (1988)
Effectiveness	Seddon and Yip (1992), Seddon and Kiew (1994), Almutairi and Subramanian (2005)
Efficiency	Seddon and Yip (1992), Seddon and Kiew (1994), Almutairi and Subramanian (2005)
Enjoyment	Gable et al. (2008)
Format	Doll and Torkzadeh (1988)
Information product	Baroudi and Orlikowski (1988)
Information satisfaction	Gable et al. (2008)
Overall satisfaction	Baroudi and Orlikowski (1988), Seddon and Yip (1992), Seddon and Kiew (1994), Rai et al. (2002), Luarn and Lin (2003),

	Almutairi and Subramanian (2005), Roca et al. (2006), Gable et al. (2008), Negash and Igbaria (2003)
Repeat purchases	DeLone and McLean (2003)
Repeat visits	DeLone and McLean (2003)
Staff and services	Baroudi and Orlikowski (1988)
System satisfaction	Gable et al. (2008)
Timeliness	Doll and Torkzadeh (1988)
User knowledge and involvement	Baroudi and Orlikowski (1988)
User surveys	DeLone and McLean (2003)

Net benefits

Net benefits were added to the updated model to cover two former distinct dimensions individual impact and organizational impact of the original model. As explained by the authors, the change was introduced since the original phrase "impacts" can be positive or negative, which may cause a possible confusion whether the results are good or bad. Also, the term "net" in "net benefits" rightly reflects the idea of capturing the difference between positive and negative effects of an IS, as no result is totally positive without any negative consequences. The choices of what benefits should be measured and who gain benefits depend on the system being assessed, the goal of the study, and the level of analysis (Urbach and Müller, 2012).

Since the purpose of this study is to learn how adding a chatbot to the customer support can affect customer experience without taking into account other important factors of an IS like cost saving or additional sale, this dimension was excluded since it could not be measured accurately with the collected data. Instead, the focus would be put on the other five dimensions.

Final research model

With all the dimensions and measures chosen for the context of this study, the final research model in Figure 3 was constructed to evaluate the success of the customer support under IS point of view. Perceived usefulness and user satisfaction would be used to measure the effectiveness of the customer support, influenced by the information quality, system quality, and service quality of the function. Between use and user satisfaction dimensions, the latter one was selected to be the final goal of the success model as it stands for the basis of a customer support success.

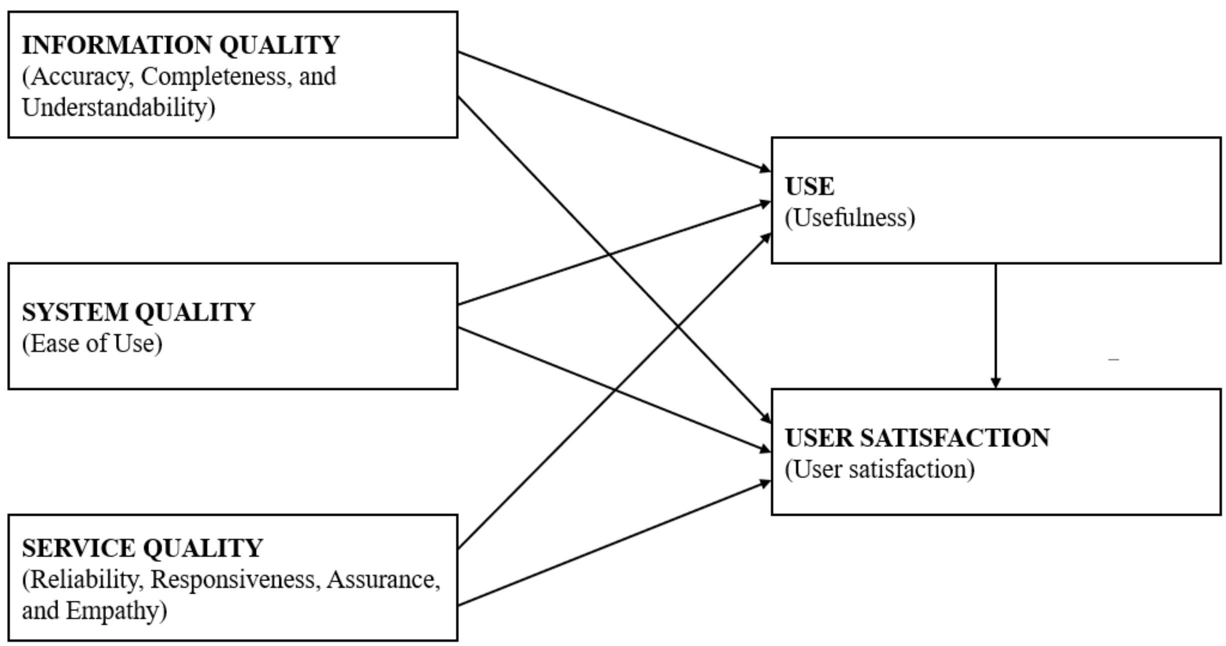

Figure 3: The research model to evaluate the success of the customer support under IS point of view

Research methodology

Case study

The case company is a company in software industry, it delivers an application to customers in multiple platforms like Android, iOS or Windows. The software installation process on each platform is different with multiple steps to be taken. Also, there are various promotions for customers after finishing the software installation. Because of the above complications, customers usually have questions during and right after the installation process. They can either try to find answer on the company's website or contact the support team to address their questions.

The customer support function of the case company is organized in a very traditional model. The team maintains an FAQ section on the company's website and answers questions received via email. Another channel under monitoring is social media but this is not a priority and the numbers of requests received via this channel is negligible. The personnel work eight hours a day for 5 days a week and the promised response time to customers are three working days.

The case company was interested in experimenting with a chatbot for its customer support for a main reason. There are some periods during a month in which the number of new users increases significantly, thus leading to a long backlog of questions to be answered. In this situation, the response time is unavoidably prolonged, which definitely affect customer experience negatively. The case company wanted to do a quick experiment with chatbot during this period to reduce the load on customer support personnel and therefore improve the response time. There were skeptical opinions but in general they were eager to see the result before making further investment on this technology.

Questionnaire design

Based on the model presented in Figure 3, a questionnaire was prepared by adapting a number of standard instruments to the context. There are five groups of questions that measure ten measures, details of these questions are listed in Table 7. First, the three questions on three measures of information quality are all from Miller (1996). Second, a question on ease of use four is based on Davis's technical acceptance model (1989). Third, the nine questions on service quality are all from Parasuraman et al. (1988). Fourth, the two questions on usefulness are from Rai et al.

(2002). Lastly, the three questions on user satisfaction are from Negash and Igbaria (2003). The questionnaire uses a 5-point Likert scale with values covering from 1 – Strongly disagree – to 5 – Strongly agree.

Table 7: Questions to be asked to assess measures

Dimension	Measure	Adapted from	Items
Information Quality	Completeness	Miller (1996)	I can find all the detailed information I need to solve my problem
	Accuracy		I receive accurate and up-to-date information from the customer support
	Understandability		I can easily understand the information provided by the customer support
System Quality	Ease of Use	Davis (1989)	I find the customer support easy to use
Service Quality	Reliability	Parasuraman et al. (1988)	I receive the right solution for my problem
			I can easily contact a person if I cannot solve the problem by myself
	Responsiveness	Parasuraman et al. (1988)	I receive response for my problem in a timely manner
			I feel that the customer support is always available to help me
	Assurance	Parasuraman et al. (1988)	I feel confident that my problem would be solved with the help from customer support

			I trust to use the customer support in the future
	Empathy	Parasuraman et al. (1988)	I feel that the customer support is friendly and personal
			I feel that the customer support understands my problem well
			I have enough channels (e.g. FAQ, email, social media...) to raise my problem and receive necessary support
Use	Usefulness	Rai et al. (2002)	I can solve my problem quickly by using the customer support
			I feel that the customer support is useful
User satisfaction	User satisfaction	Negash and Igbaria (2003)	My expectations on the customer support are met
			I am satisfied with the amount of time it takes to solve my problem
			I am satisfied with the quality of the support I receive

Data collection

To collect the data for this study, the online questionnaire was sent to customers who contacted the customer support or used the chatbot during or right after installing the application of the company. This approach ensured that the comparisons between comparing groups were valid as all ratings came from new customers who contacted the support team for the first time and thus they were not impacted by good or bad experiences before. The process happened as follows. First, before the chatbot was deployed, the customers who contacted the customer support via email were invited to take the questionnaire. Then after the chatbot deployment, customers who used the chatbot were contacted to collect their opinions. The latter group consisted of both who used the chatbot only or chatbot and then interacted with human agents to solve their problems.

Demographic statistics for the respondents are presented in Table 8. The gender ratio was 70% female and 30% male, matching with the gender ratio of the customer base. The majority of the respondents (70%) are young adults and more than half of the sample have at least a college degree or higher education levels.

Table 8: Demographic data for respondents

		Count	Percentage
Number of respondents		60	100
Gender	Male	18	30
	Female	42	70
Age	Under 18	1	1.7
	18 to 44	42	70
	Over 45	17	28.3
Education level	Some high school	7	11.7
	Graduated high school	13	21.7
	Some college	19	31.7
	Associate’s degree	5	8.3
	Bachelor’s degree	10	16.6
	Post graduate degree	6	10

Findings

At the beginning, the customer experiences with two customer support systems, one with the chatbot enabled and one without it, were examined. The data of ten quality measures including information completeness, information accuracy, information understandability, usefulness, ease of use, reliability, responsiveness, assurance, empathy, and satisfaction were analyzed descriptively and then compared between two systems ($n_1 = n_2 = 30$) to produce insights for research question 1. After that, the data of the customer system with the chatbot enabled were split into two subsets, one received from the users who used the chatbot only and were able to solve their problems, and the other one from the users who already used the chatbot but also required further assistance from human agents afterwards to settle their cases completely. Then descriptive analysis and nonparametric statistical test ($n_1 = n_2 = 15$) were performed again to answer research question 2.

Research question 1: Potential effects of the chatbot on the customer support

Information quality

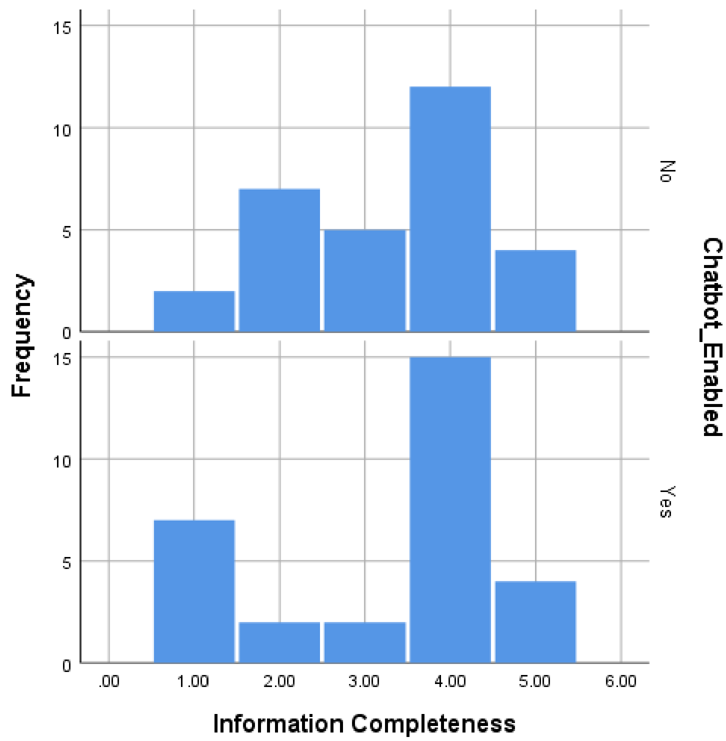

Figure 4: Distribution of ratings in information completeness measure for two customer support systems, one with the chatbot enabled and one without it

Table 9: Descriptive statistics for ratings in in information completeness measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Information Completeness	No	3.30	1.18	4.00	2.00	-0.363
	Yes	3.23	1.43	4.00	2.25	-0.668

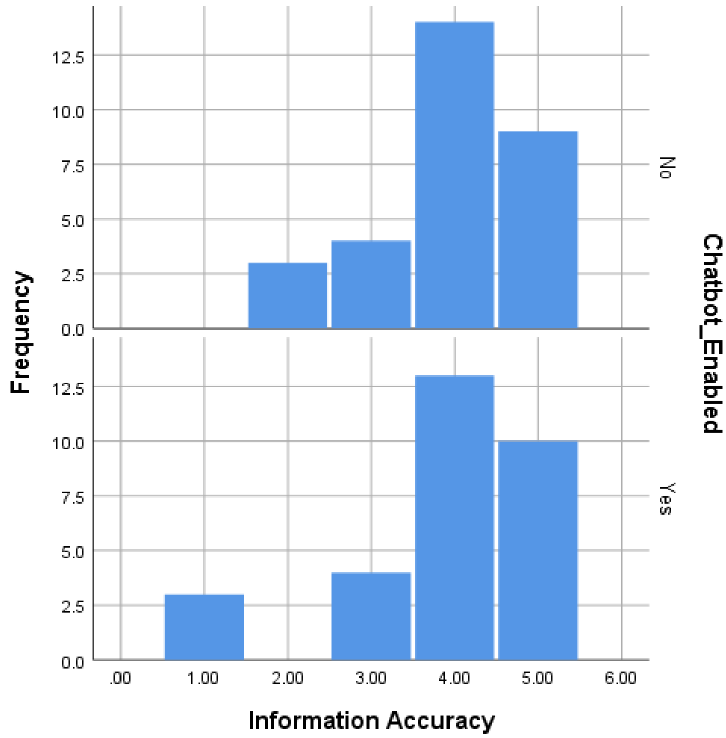

Figure 5: Distribution of ratings in information accuracy measure for two customer support systems, one with the chatbot enabled and one without it

Table 10: Descriptive statistics for ratings in information accuracy measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Information Accuracy	No	3.97	0.93	4.00	1.25	-0.763
	Yes	3.90	1.18	4.00	1.25	-1.395

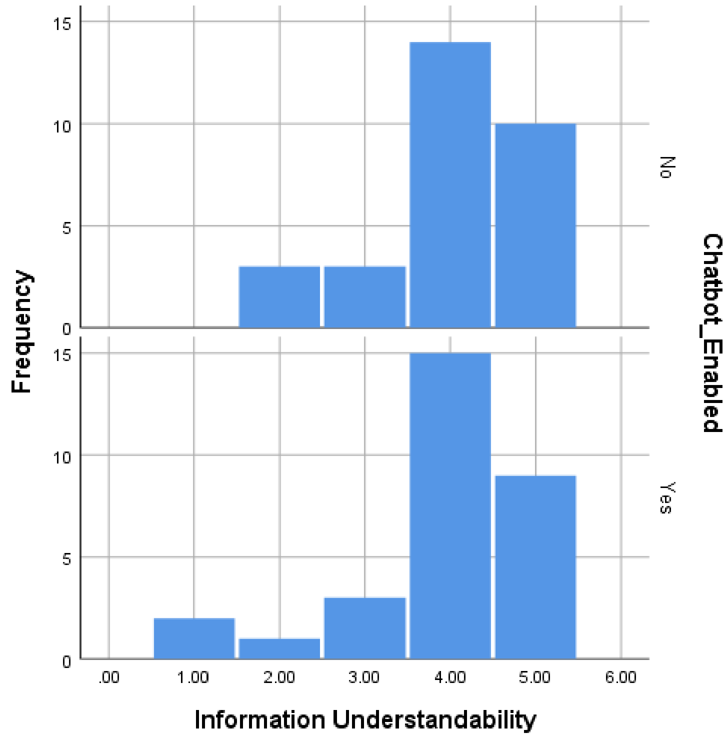

Figure 6: Distribution of ratings in information understandability measure for two customer support systems, one with the chatbot enabled and one without it

Table 11: Descriptive statistics for ratings in information understandability measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Information Understandability	No	4.03	0.93	4.00	1.00	-0.902
	Yes	3.93	1.08	4.00	1.00	-1.441

As presented, all three information quality measures were negatively skewed. When comparing between groups, the medians and IQRs were equal while the means of the customer support system without chatbot were slightly higher across all measures.

The result of Mann-Whitney U test showed that there is no significant difference between the two customer support systems in all three measures (Mann-Whitney U, $p = 0.963$ for information completeness, p = 0.912 for information accuracy, and p = 0.810 for information understandability).

System quality

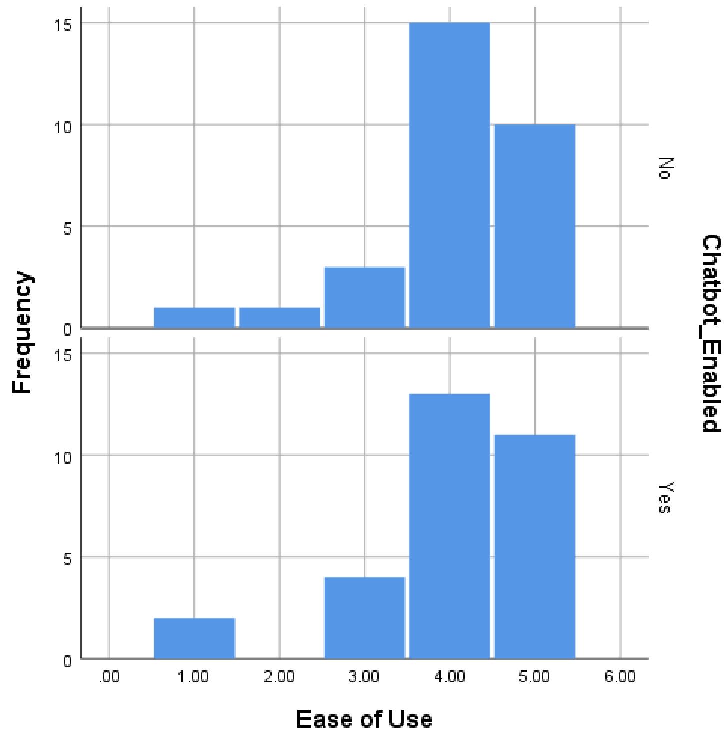

Figure 7: Distribution of ratings in ease of use measure for two customer support systems, one with the chatbot enabled and one without it

Table 12: Descriptive statistics for ratings in ease of use measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Ease of use	No	4.07	0.94	4.00	1.00	-1.455
	Yes	4.03	1.07	4.00	1.00	-1.533

Similar to information quality measures, the distributions of ratings in ease of use measure were negatively skewed. Both the means, medians and IQRs between comparing groups were very similar to each other.

Mann-Whitney U test was performed again and the result showed that there was no significant difference between the two customer support systems in this dimension (Mann-Whitney U, p = 0.287 for ease of use).

Service quality

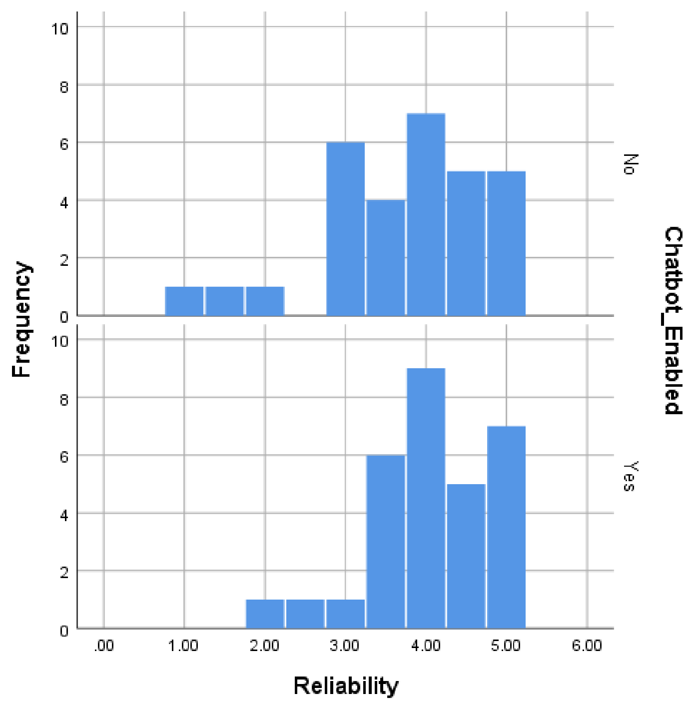

Figure 8: Distribution of ratings in reliability measure for two customer support systems, one with the chatbot enabled and one without it

Table 13: Descriptive statistics for ratings in reliability measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Reliability	No	3.73	1.02	4.00	1.50	-0.918
	Yes	4.07	0.76	4.00	1.13	-0.740

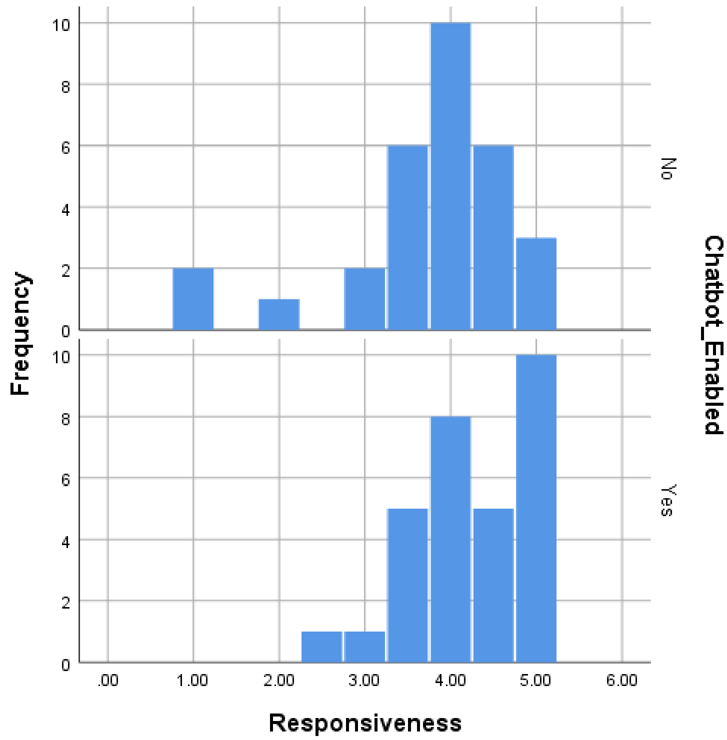

Figure 9: Distribution of ratings in responsiveness measure for two customer support systems, one with the chatbot enabled and one without it

Table 14: Descriptive statistics for ratings in responsiveness measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Responsiveness	No	3.77	0.99	4.00	1.00	-1.557
	Yes	4.25	0.69	4.25	1.13	-0.586

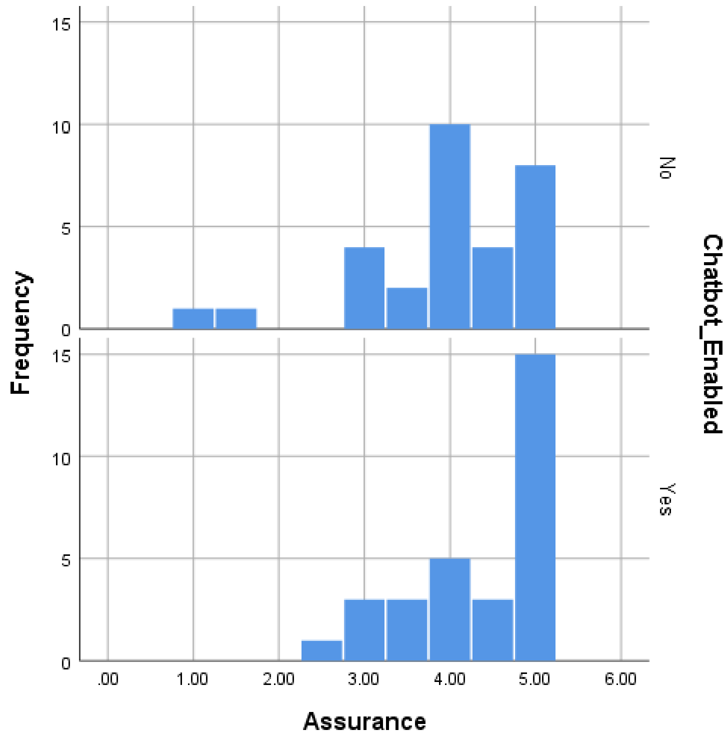

Figure 10: Distribution of ratings in assurance measure for two customer support systems, one with the chatbot enabled and one without it

Table 15: Descriptive statistics for ratings in assurance measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Assurance	No	3.98	1.00	4.00	1.50	-1.383
	Yes	4.35	0.79	4.75	1.13	-0.874

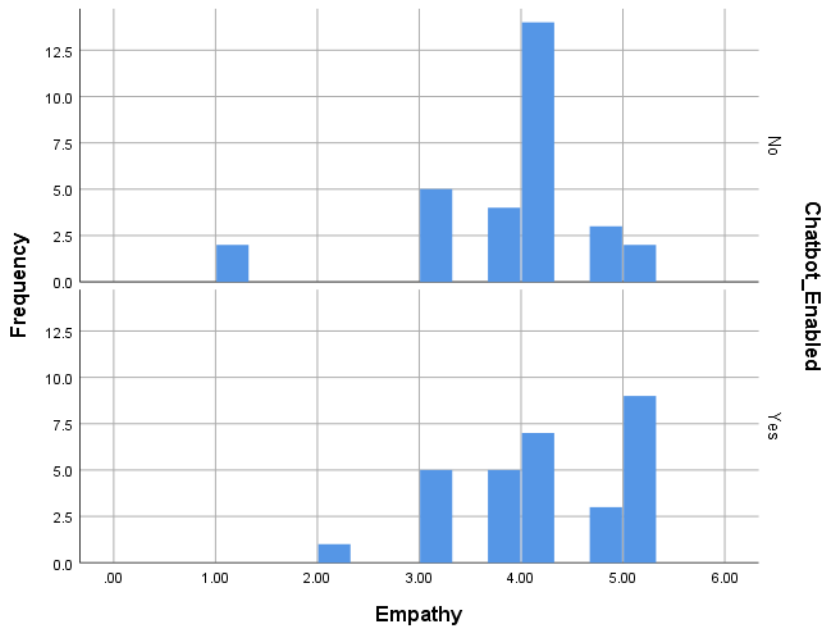

Figure 11: Distribution of ratings in empathy measure for two customer support systems, one with the chatbot enabled and one without it

Table 16: Descriptive statistics for ratings in empathy measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Empathy	No	3.87	0.89	4.00	0.75	-1.871
	Yes	4.14	0.76	4.00	1.33	-0.432

Service quality dimension revealed more differences between the two customer support systems. When compared to the initial customer support without chatbot, the new one with the chatbot enabled seemed to have higher means and medians with noticeable gap across all four service quality measures.

However, when performing Mann-Whitney U test to verify the differences, only the distributions of responsiveness ratings could be proved to have significant differences between two customer support systems (Mann-Whitney U, $p = 0.049$). The customer support system with the chatbot enabled had higher mean ranks (34.83) than the initial one without the chatbot (26.17).

Meanwhile, there was no significant difference found among reliability, assurance and empathy dimensions (Mann-Whitney U, $p = 0.222$, 0.124, and 0.134 respectively).

Use

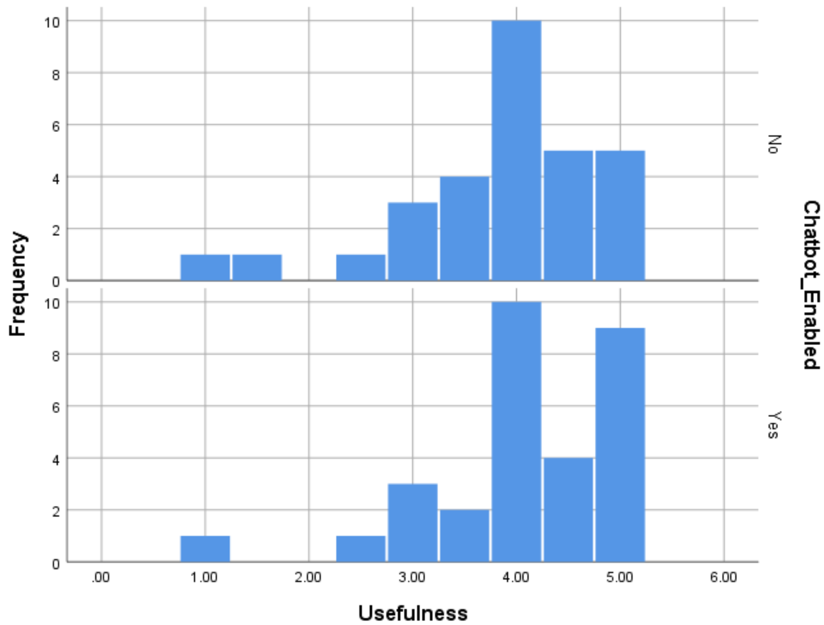

Figure 12: Descriptive statistics for ratings in usefulness measure for two customer support systems, one with the chatbot enabled and one without it

Table 17: Descriptive statistics for ratings in usefulness measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Usefulness	No	3.85	0.97	4.00	1.00	-1.294
	Yes	4.08	0.92	4.00	1.13	-1.441

The distributions of ratings in this measure were negatively skewed. When comparing statistics between two customer support systems, the medians and IQRs were equal while the means of the customer support system with the chatbot enabled were slightly higher.

Mann-Whitney U test did not indicate any significant differences between the two customer support systems in this measure (Mann-Whitney U, $p = 0.287$).

User Satisfaction

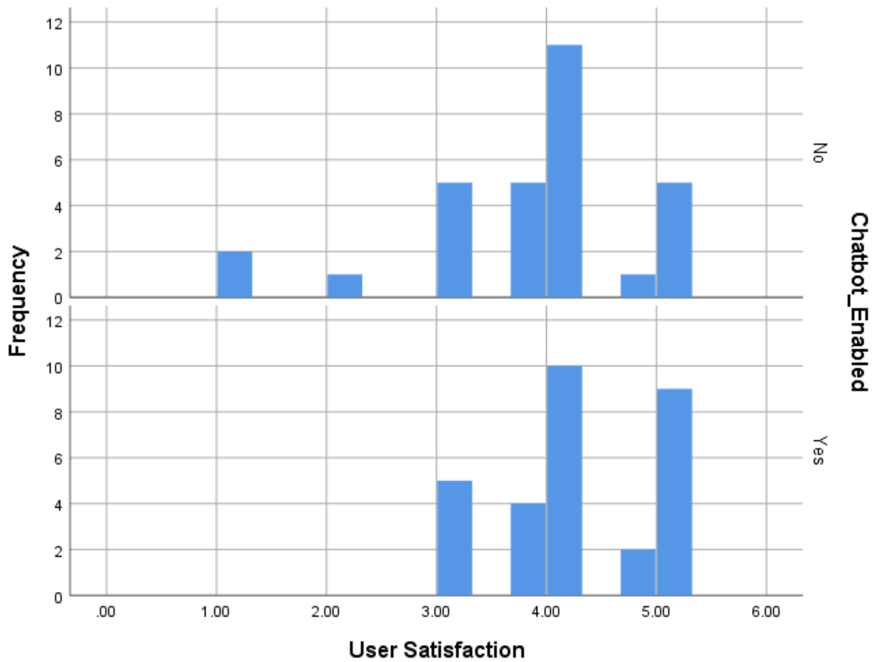

Figure 13: Distribution of ratings in user satisfaction measure for two customer support systems, one with the chatbot enabled and one without it

Table 18: Descriptive statistics for ratings in user satisfaction measure for two customer support systems, one with the chatbot enabled and one without it

Measure	Chatbot enabled	Mean	SD	Median	IQR	Skewness
Satisfaction	No	3.78	1.00	4.00	1.00	-1.357
	Yes	4.2	0.65	4.00	1.33	-0.027

In this dimension, one noticeable observation is that the customer support system with the chatbot enabled had much higher mean with no value at the bottom of the scale compared to the other system.

However again, Mann-Whitney U test did not support any significant difference between the two customer support systems in this dimension (Mann-Whitney U, $p = 0.134$).

Research Question 2: The impact to the user experience when the chatbot cannot solve problems by itself

Information quality

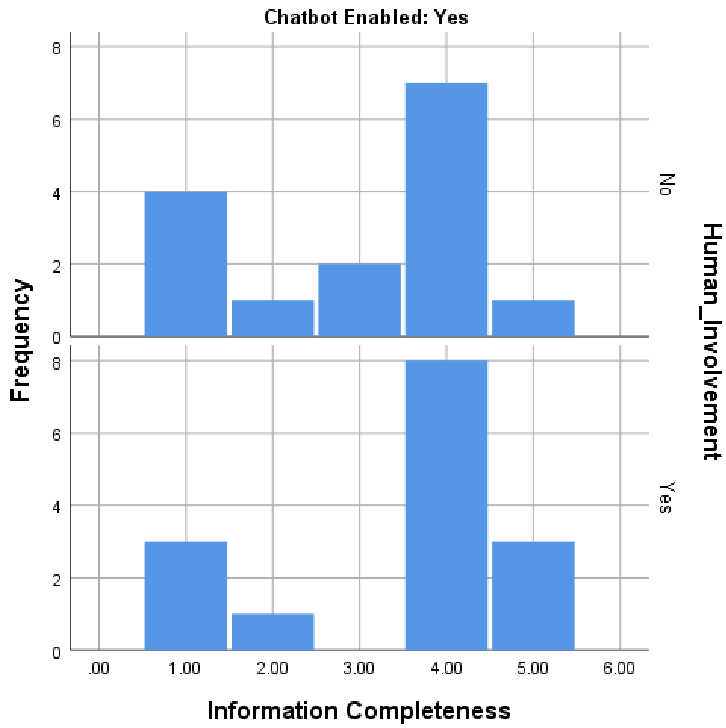

Figure 14: Distribution of ratings in information completeness measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 19: Descriptive statistics for ratings in information completeness dimension in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Information completeness	No	3	1.00	4.00	3.00	-1.352
	Yes	3.47	1.46	4.00	2.00	-0.963

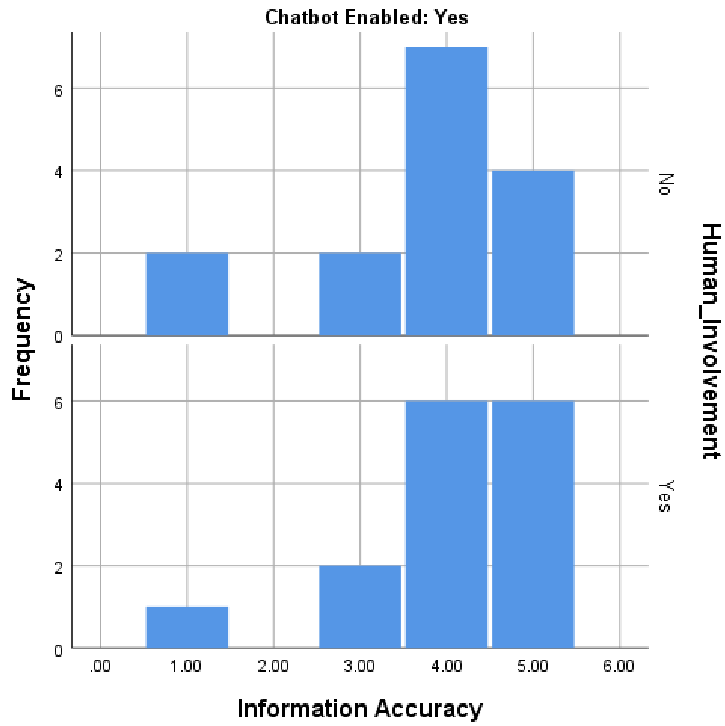

Figure 15: Distribution of ratings in information accuracy measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 20: Descriptive statistics for ratings in information accuracy measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	**Human involved**	**Mean**	**SD**	**Median**	**IQR**	**Skewness**
Information accuracy	No	3.73	1.28	4.00	2.00	-1.312
	Yes	4.07	1.10	4.00	1.00	-1.635

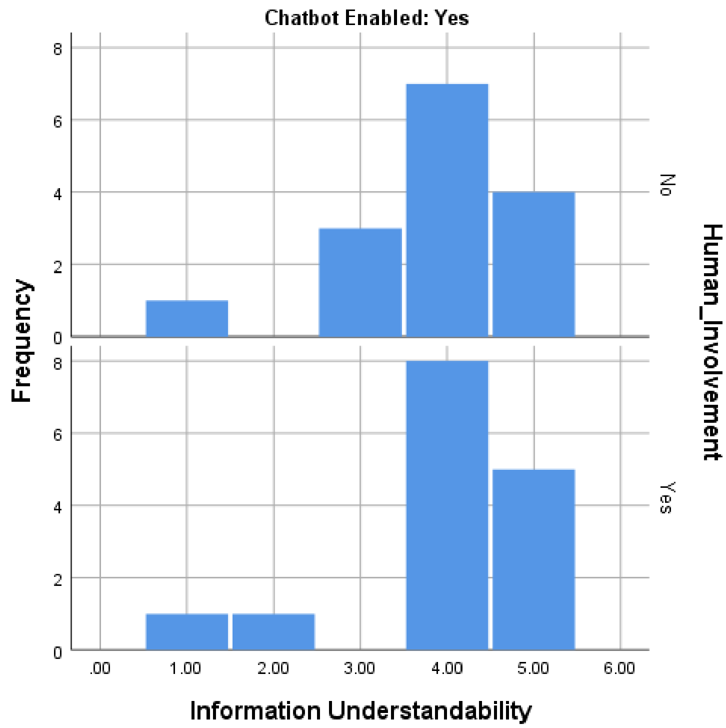

Figure 16: Distribution of ratings on information understandability measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 21: Descriptive statistics for ratings in information understandability measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Information understandability	No	3.87	1.06	4.00	2.00	-1.361
	Yes	4.00	1.13	4.00	1.00	-1.696

In both scenarios, the distributions of all three information quality measures were negatively skewed. While the medians of these distributions were equal between comparing groups, the IQRs in the scenario with human involvement were lower with quite big margins compared to the other scenario without human involvement across all dimensions. This indicated that the ratings on information quality were more consistent around the medians in the scenario with human involvement.

Meanwhile, Mann-Whitney U test did not detect any no significant difference between the two scenarios in all three dimensions (Mann-Whitney U, p = 0.254 for information completeness, p = 0.439 for information accuracy, and p = 0.528 for information understandability).

System quality

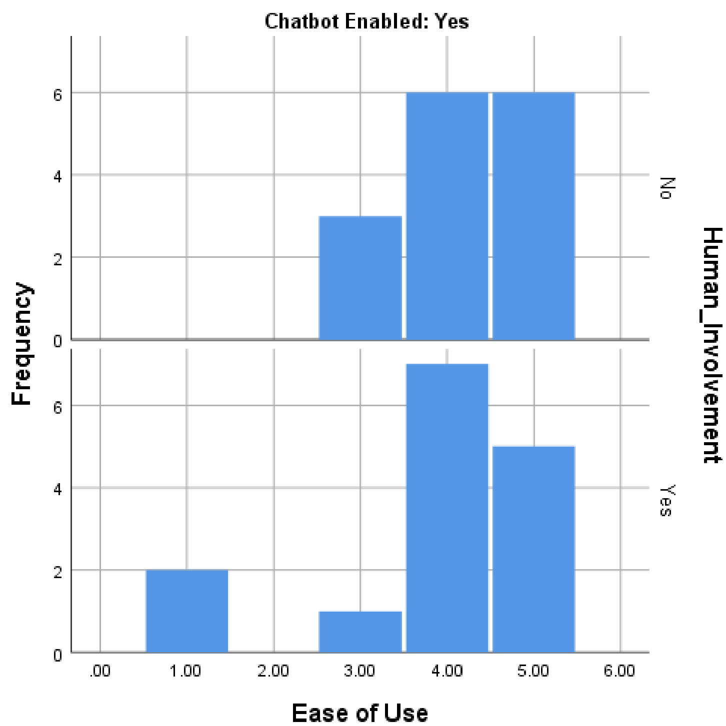

Figure 17: Distribution of ratings in ease of use measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 22: Descriptive statistics for ratings in ease of use measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Ease of use	No	4.20	0.77	4.00	1.00	-0.383
	Yes	3.87	1.30	4.00	1.00	-1.511

The distributions of ratings on ease of use measure were negatively skewed for both scenarios. While the medians and IQRs are equal, the mean in the scenario without human involvement was notably higher compared to the one with human involvement.

The result of Mann-Whitney U test showed that there was no significant difference in both dimensions between the two scenarios (Mann-Whitney U, $p = 0.367$ for Usefulness, and $p = 0.689$ for Ease of Use).

Service quality

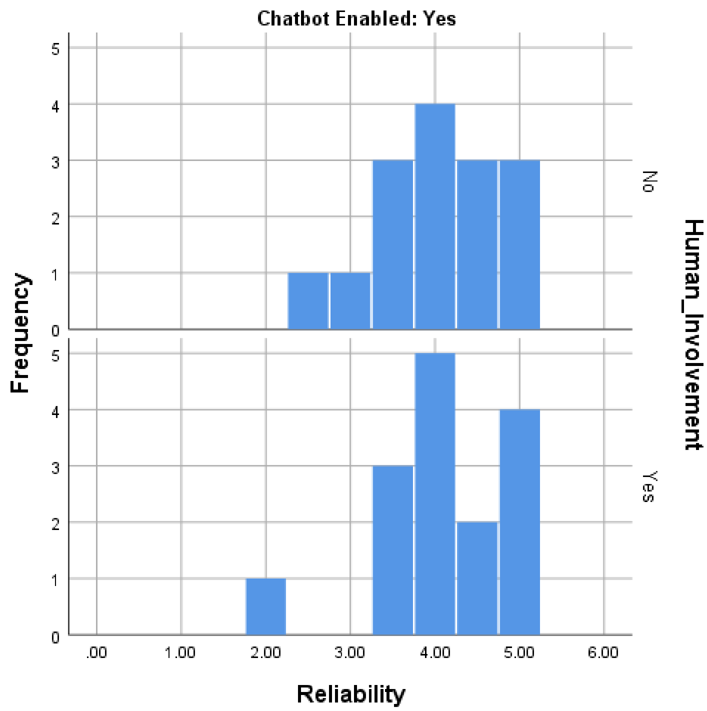

Figure 18: Distribution of ratings in reliability measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 23: Descriptive statistics for ratings in reliability measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Reliability	No	4.03	0.74	4.00	1.00	-0.431
	Yes	4.10	0.80	4.00	1.50	-1.080

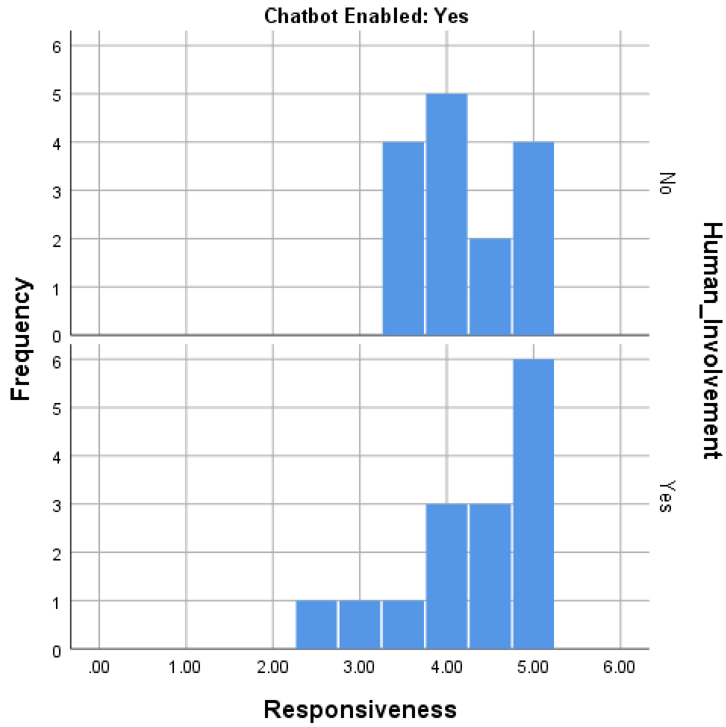

Figure 19: Distribution of ratings in responsiveness measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 24: Descriptive statistics for ratings in reliability measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Responsiveness	No	4.20	0.59	4.00	1.50	0.275
	Yes	4.20	0.84	4.50	1.50	-1.063

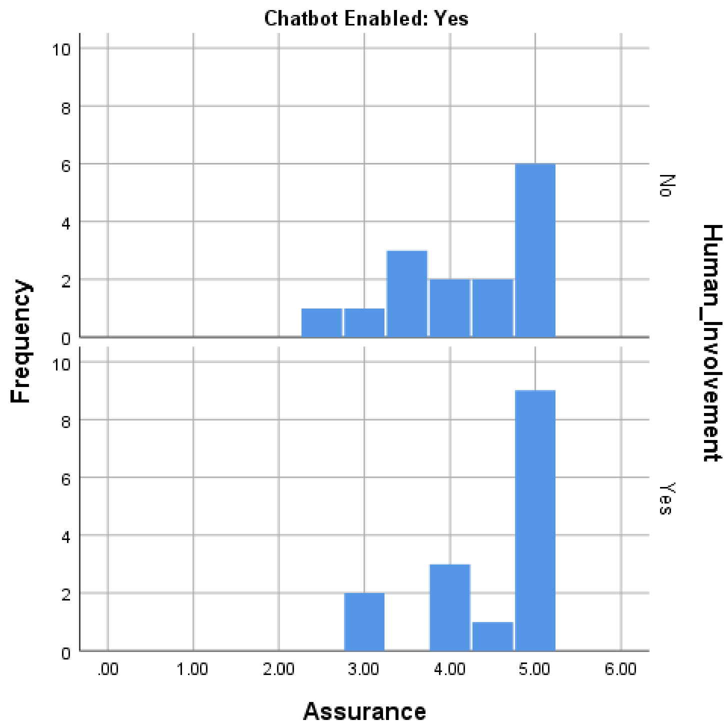

Figure 20: Distribution of ratings in assurance measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 25: Descriptive statistics for ratings in assurance measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Assurance	No	4.20	0.84	4.50	1.50	-0.632
	Yes	4.50	0.73	5.00	1.00	-1.261

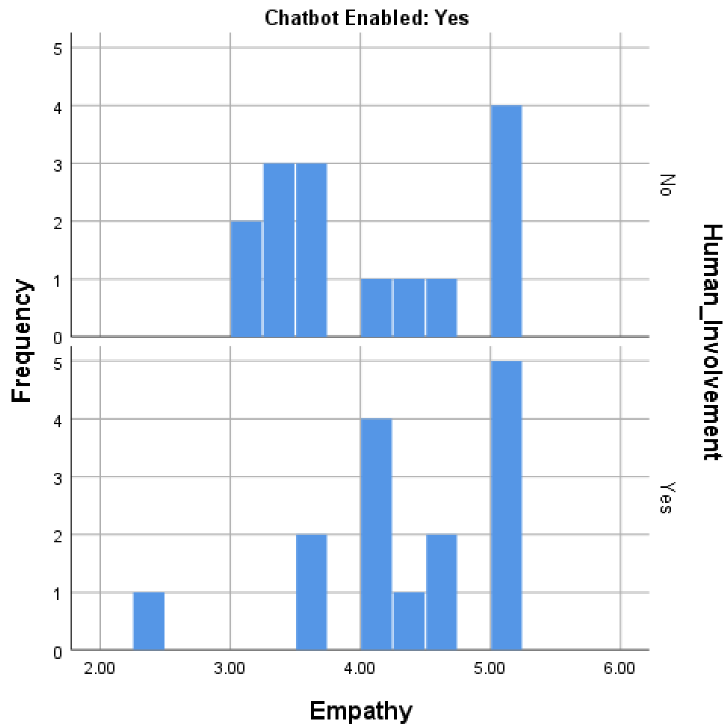

Figure 21: Distribution of ratings in empathy measure for in scenarios, the chatbot solved the problem without and with human involvement

Table 26: Descriptive statistics for ratings in empathy measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Empathy	No	4.00	0.77	3.67	1.67	0.243
	Yes	4.29	0.74	4.33	1.00	-1.237

Among four measures of service quality dimension, while reliability and responsive were quite similar between two scenarios, assurance and empathy revealed some interesting observations. In the scenario with human involvement, the median of assurance dimension hit the highest possible value (5.00) with a low IQR (1.00). Meanwhile in empathy dimension, the scenario without human involvement received the lowest median score (3.67) across all dimensions in the whole study.

Again, no significant difference could be found between the two scenarios in all measures (Mann-Whitney U, p = 0.733 for reliability, p = 0.466 for responsiveness, p = 0.275 for assurance, and p = 0.205 for empathy).

Use

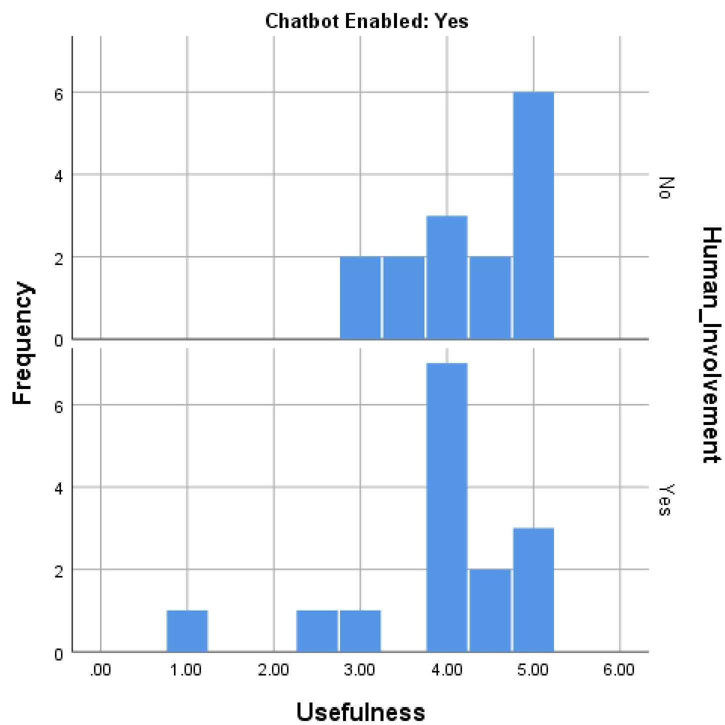

Figure 22: Distribution of ratings in usefulness measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 27: Descriptive statistics for ratings in usefulness measure in two scenarios, the chatbot solved the problem without and with human involvement

Variable	Human involved	Mean	SD	Median	IQR	Skewness
Usefulness	No	4.27	0.75	4.50	1.50	-0.508
	Yes	3.90	1.06	4.00	0.50	-1.642

In this measure, the median and mean in the scenario without human involvement were higher compared to the one with human involvement.

The result of Mann-Whitney U test showed that there was no significant difference in this measure between the two scenarios (Mann-Whitney U, $p = 0.367$).

User Satisfaction

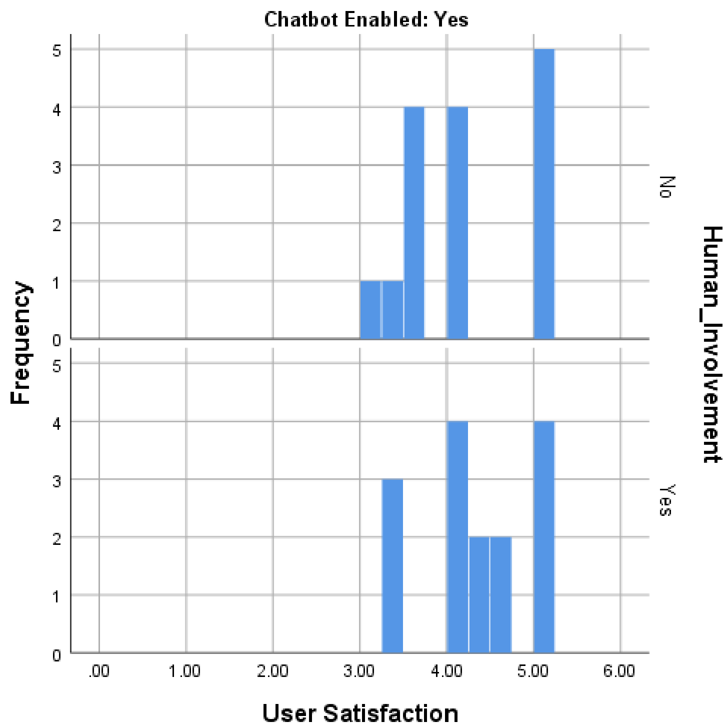

Figure 23: Distribution of ratings in user satisfaction measure in two scenarios, the chatbot solved the problem without and with human involvement

Table 28: Descriptive statistics for ratings in user satisfaction measure in two scenarios, the chatbot solved the problem without and with human involvement

Measure	Human involved	Mean	SD	Median	IQR	Skewness
User satisfaction	No	4.13	0.69	4.00	1.33	0.218
	Yes	4.27	0.62	4.33	1.00	-0.282

In this dimension, the scenario without human involvement had higher mean, median with lower SD and IQR. Still, Mann-Whitney U test did not support the hypothesis of existing significant difference between the two scenarios (Mann-Whitney U, $p = 0.551$).

Discussion and conclusion

In this chapter, the findings are reviewed further to answer two research questions. Limitations and further research are discussed after that.

Answering research questions

Research question 1: How does adding a chatbot to a traditional customer support (with FAQ and email channel) affect customer experience?

One of the main responsibilities, if not the biggest one, of customer support is delivering quality information to customers, either to answer their enquiries, or to instruct step-by-step how to solve the issues they encountered. Generally, each organization always tries to create and maintain a knowledge base which contains consistent and effective known practices or methods to handle cases escalated by its customers. A knowledge base is a collection of items that exist under multiple forms, such as FAQ articles, email templates used to reply to customers, or even personal notes prepared by agents. This is indeed the first source of information to prepare chatbot scenarios for a customer support.

Also, it is important to understand one characteristic of chatbot related to information quality, the constraint of information throughput that can be delivered at a time. In many cases, the amount of information required to handle a known issue cannot be handed over in a chat window. A common solution for this problem is rerouting customers to FAQ articles, which allows more freedom to display required information. This phenomenon further creates a tightly coupling between a chatbot and a knowledge base of a company. The fact that all information quality measures had a very similar scores between two customer support systems before and after the chatbot implementation in the study strongly indicated that chatbot cannot individually improve information quality of customer support. Besides that, the $M = f(D, C)$ equation mentioned in literature review chapter can provide another reasonable explanation here: while D (data) and C (context) can somehow be under control of the organization, f (the knowledge and competence of users) varies greatly among individuals and is hard to be fixed to a certain standard. There were cases when an explanation was detailed enough for some users but hard to understand for others. Therefore, , it is important to recognize that information quality must be enhanced as a whole across all channels of a customer support, with the goal of delivering meaningful information to groups of users who have the lowest level of knowledge and understanding on products or services provided.

Looking into how the customers used the chatbot during the experiment, the data showed that out of 521 conversions, there were 142 cases in which the chatbot could not solve by itself so then it suggested the customers to contact the support team. This reality happened because either the chatbot could not understand the contexts and find the appropriate solutions, or these cases were too complicated and required human agents to make the right decisions. That reflected a very clear picture of the state of chatbot technology at the moment, it is not perfect and ready yet to completely replace human agents, but it definitely can help improve the operation of customer support.

It is essential to clarify the role of chatbot in customer support as a tool to deliver first level support, to both organizations and customers. On one hand, it can be used to handle simple questions while for more complicated cases, it should always redirect the flow to human agents. By letting chatbot take care of repetitive tasks quickly, support teams can have more time for other complex ones. With this idea in mind at the beginning of the process to develop a chatbot for its customer support system, an organization can focus on filtering and defining those simple enough scenarios to let the chatbot handle instead of making it too complex for customers to use. On the other hand, its customers should be communicated (or even educated) at the beginning of the conversions that the chatbot cannot solve all questions and when they want, they can always easily switch to other channels or find human agents to handle their cases. By setting the right mentality for customers when using chatbot, organizations can increase the possibility of getting customer acceptance and therefore improve perceived ease of use and usefulness measures.

Although chatbot offers an opportunity to reduce operational cost of customer support and serve more customers, many organizations still hesitate to implement this technology. There are a few main reasons behind this hesitation, the fear of losing personal touch with customers, the lack of empathy to handle frustrated customers, or the uncertainty of chatbot learning capability. Those concerns are all valid and indeed the best customer experience is normally delivered by humans. No matter how advanced chatbot technology can be developed in the future, there are situations and subjects that a human can handle better than a bot. However, there are certain factors that make chatbot become a strong candidate to be added to customer support, not because it is the best one, but the best possible one.

Throughout the study, ten dimensions were compared between two customer support systems with the chatbot enabled and without it. Still, the result showed that only one dimension recorded a significant difference between how customers rated these two systems, responsiveness. Behind the cost factor, this is definitely one of the greatest strengths of chatbot, it can reduce customer support response time and provide 24/7 service coverage. Customer support response time is the period between when a inquiry is sent by a customer and when customer support provides the first response for that inquiry. This is one of the most important metrics in customer support which indicates how long an organization take to respond to a customer.

In the case of the company in which the study was conducted, there is a big time zone difference between the customer base and the customer support team, also the customer support team only works eight hours on workdays. Therefore, the promised response time is within three working days. That means in the worst scenario, a customer sends a question on Friday may receive a response next Wednesday. Technologies keep advancing and have influenced customers to raise their expectation on customer support all the time, three working days does not seem to be a good standard anymore and it may decrease customer satisfaction. Also, another element that makes reducing customer support response time even more important for the company is the nature of its business. In a short window when a customer decides to give the company's application a try, any concerns, questions or issues they have should be addressed as soon as possible. The longer the customer has to wait to get a response, the least likely the installation process may happen. In this situation, neither building a new customer support team near the customer base or making the current one work 24/7 are viable options. Developing a chatbot to fill the gap, eliminate repetitive questions so the support team can have more time to reach other complex cases earlier, is definitely a sensible option that the company should consider.

Besides that, adding chatbot as a new channel to get support can help deliver a message to customers that firms are working constantly to improve their customer support. From customer perspective, they may not know about the benefit of reducing operational cost. What they really care is that they can receive necessary support to solve their problems quickly, accurately and appropriately, whether via a chatbot, an FAQ article or a human agent. That may somehow explain why even the chatbot still had some flaws during the experiment, it still received quite positive ratings on other service quality measures besides responsiveness and user satisfaction dimension.

The result of the questionnaire supported the hypothesis of research question 1 that the addition of a chatbot to a traditional customer support model can improve customer experience, mainly on responsiveness measure, while maintain a similar level on information quality, system quality and user satisfaction dimensions.

Research question 2: *How may unsuccessful interactions of a chatbot affect user experience?* This research question was studied with the assumption that unsolved cases would be escalated and handled properly by human agents afterwards.

As pointed out in Findings chapter, while comparing three information quality dimensions between two scenarios with and without human involvement, Mann-Whitney U test did not verify any significant differences. Still, the ratings on information quality seem more consistent around the medians in the scenario with human involvement. There are probably two explanations behind this result. First, the deployed chatbot was built mainly using two mechanisms: (1) decision tree with predefined questions that a user can choose to ask the bot, and (2) Natural Language Processing and algorithmic probability to analyse a user input and to determine an appropriate response. In some occasions, the latter mechanism was not able to analyse and understand what users inputted so it returned an apology to them. That certainly affected the ratings of some users on measures like information completeness and information accuracy.

Second, the chatbot itself was developed so that in a number of certain situations, it reroutes the conversions to human agents so that more detailed information can be delivered to users, or more complex decisions can be made. This is due to some limitations on capability of the chatbot platform which was used. Also, since a chat window cannot deliver such long paragraphs like a FAQ article or an email, all predefined answers are normally the shorter versions of the ones prepared for other channels. Therefore, some users may feel that the responses they received were not sufficient or harder to understand.

Similarly, examining through system quality and service quality dimensions revealed the superior of the ratings on the scenario with human involvement in usefulness and assurance dimensions. At first glance the observation here may sound skeptical since in this scenario, the chatbot could not solve users' cases within initial conversions so they had to spend more time later with human agents to settle their cases completely. However, it is important to remind that these customers were instructed to rate the customer support as a whole, not rate the chatbot individually. So when the chatbot could not solve, or had been programmed to not solve, some certain types of issues, the process of escalating to human may not be as harmful as the initial prediction. The similar ratings in responsiveness measure between two scenarios with and without human involvement in some manner support this hypothesis. Within a certain degree of expectation on how long it may take the customer support to solve their problems, especially more complicated ones, the customers might not mind that much if they need to go through another step after the chatbot. Moreover, those users who experienced the process of escalation from the chatbot to human agents may feel more assured that their problems would be taken cared of completely via any channel.

So, for research question 2, the result of the study did not support the hypothesis that unsuccessful chatbot attempts which require further human involvement worsen customer experience.

Lastly to summarize the experiment with chatbot in the case company, there are some valuable lessons that organizations may want to consider before implementing this technology. First, do not overestimate the potential of chatbot, it cannot replace human agents completely in customer support. The technology is advancing progressively but there are always matters that a human can handle better than a bot. Finding the right balance between chatbot and human resources in customer support is the key to be successful with this technology. Second, do not overcomplicate the chatbot you want to build. It should handle only simple enough tasks and leave the more complex and trickier ones to human. A bad conversion handled completely by a chatbot is worse than a case handled by a human after a proper escalation from a chatbot. And third, building a chatbot is a continuous process that keeps going on after the first bot release. Using chatbot in customer support requires organizations to stay updated with all analytics of the conversations between their chatbot and users. Unfortunately, while this is a time-consuming process, it is the only way to tune the chatbot to keep up with all new issues and how customers communicate about them to customer support.

Limitations and further research

A limitation of the study is the size of the sampling which occurred due to the time and cost restrictions. For future work, it would be valuable to reproduce the study with a larger sample size in different industries or other companies to see whether the findings are still valid in a bigger context.

Another limitation is that the users who received the questionnaire were the ones who just joined the company's program and contacted the customer support for the first time. It would be interesting to measure potential effects of chatbot on customers who already used the customer support before and now would experience the addition of a chatbot. This experiment may reveal further observations on how chatbot truly affects current customers of a company, who got used to a traditional customer support system of the company already.

Furthermore, as pointed out the limitations of chatbot in handling too complex requests or conversations with the involvement of emotions, it would be beneficial to analyze other case studies to build a certain framework that can be used to categorize the right channel for each scenario. That would help reduce the cost of developing a new chatbot for organizations by not overcomplicating the chatbot unnecessarily, as well as effectively streamline the experience of customers with their services.

References

Almutairi, H. and Subramanian, G.H. (2005). An empirical application of the DeLone and McLean model in the Kuwaiti private sector. *Journal of Computer Information Systems*, 45(3), pp.113-122.

Anton, J. (2000). The past, present and future of customer access centers. *International Journal of Service Industry Management*, 11(2), pp.120-130.

Armistead, C.G. and Clark, G. (1992). *Customer service and support: Implementing effective strategies*. Financial Times-Pitman.

Bailey, J.E. and Pearson, S.W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management science*, 29(5), pp.530-545.

Baroudi, J.J. and Orlikowski, W.J. (1988). A short-form measure of user information satisfaction: a psychometric evaluation and notes on use. *Journal of Management Information Systems*, 4(4), pp.44-59.

Belardo, S., Karwan, K.R. and Wallace, W.A. (1982). DSS component design through field experimentation: an application to emergency management. In *Proceedings of the Third International Conference on Information Systems* (Vol. 93).

Bozorgzadeh, A. (2017). *Facebook's botched Messenger rollout isn't hurting chatbots. [online] VentureBeat*. Available at: https://venturebeat.com/2017/08/01/facebooks-botched-messenger-rollout-isnt-hurting-chatbots/ [Accessed 14 Jul. 2018].

Brynjolfsson, E. and Mcafee, A., (2017). The business of artificial intelligence: What it can and cannot do for your organization. *Harvard Business Review Digital Articles*, pp.3-11.

Brandtzaeg, P.B. and Følstad, A. (2017). Why people use chatbots. In *International Conference on Internet Science* (pp. 377-392). Springer, Cham.

Canalys.com. (2018). *Smart speakers are the fastest-growing consumer tech; shipments to surpass 50 million in 2018*. [online] Available at: https://www.canalys.com/newsroom/smart-speakers-are-fastest-growing-consumer-tech-shipments-surpass-50-million-2018 [Accessed 14 Jul. 2018].

Carman, J.M. (1990). Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions. *Journal of Retailing*, 66(1), pp. 33-55.

Chung, M., Ko, E., Joung, H. and Kim, S.J., 2018. Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*.

Ciechanowski, L., Przegalinska, A., Magnuski, M. and Gloor, P. (2018). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*.

Cohen, D. (2019). Facebook Messenger Now Has More Than 1.2 Billion Monthly Active Users. [online] Adweek.com. Available at: https://www.adweek.com/digital/facebook-messenger-1-2-million-monthly-active-users/ [Accessed 24 Feb. 2019].

Cook, C. and Thompson, B. (2000). Reliability and validity of SERVQUAL scores used to evaluate perceptions of library service quality. *The Journal of Academic Librarianship*, 26(4), pp.248-258.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C. and Zhou, M., 2017. Superagent: A customer service chatbot for e-commerce websites. *Proceedings of ACL 2017, System Demonstrations*, pp.97-102.

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), pp.811-817.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pp.319-340.

DeLone, W.H. and McLean, E.R. (1992). Information systems success: The quest for the dependent variable. *Information systems research*, 3(1), pp.60-95.

Delone, W.H. and McLean, E.R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), pp.9-30.

Doll, W.J. and Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. *MIS quarterly*, pp.259-274.

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D. and Mueller, E.T. (2013). Watson: beyond jeopardy!. *Artificial Intelligence*, 199, pp.93-105.

Flaiz, W. (2018). *Using Chatbots As A Customer Service Virtual Agent*. [online] Forbes.com. Available at: https://www.forbes.com/sites/forbescommunicationscouncil/2018/02/21/using-chatbots-as-a-customer-service-virtual-agent/#38baab2f2a90 [Accessed 5 Mar. 2019].

Følstad, A., Nordheim, C.B. and Bjørkli, C.A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *International Conference on Internet Science* (pp. 194-208). Springer, Cham.

Gable, G.G., Sedera, D. and Chan, T. (2008). Re-conceptualizing information system success: The IS-impact measurement model. *Journal of the association for information systems*, 9(7), p.18.

Gartner (2018). *Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020*. [online] Available at: https://www.gartner.com/en/newsroom/press-releases/2018-02-19-gartner-says-25-percent-of-customer-service-operations-will-use-virtual-customer-assistants-by-2020 [Accessed 16 Jun. 2018].

Goffin, K. and New, C. (2001). Customer support and new product development-An exploratory study. *International Journal of Operations & Production Management*, 21(3), pp.275-301.

Griol, D., Carbó, J. and Molina, J.M. (2013). An automatic dialog simulation technique to develop and evaluate interactive conversational agents. *Applied Artificial Intelligence*, 27(9), pp.759-780.

Hardalov, M., Koychev, I. and Nakov, P. (2018). Towards Automated Customer Support. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications* (pp. 48-59). Springer, Cham.

Hettige, B., & Karunananda, A. S. (2015). Octopus: A Multi Agent Chatbot. In *Proceedings of 8th International Research Conference KDU*.

Holloway, B.B. and Beatty, S.E. (2008). Satisfiers and dissatisfiers in the online environment: A critical incident assessment. *Journal of service research*, 10(4), pp.347-364.

Huh, Y.U., Keller, F.R., Redman, T.C. and Watkins, A.R. (1990). Data quality. *Information and software technology*, 32(8), pp.559-565.

Hutchens, J.L. and Alder, M.D. (1998). Introducing megahal. In *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning* (pp. 271-274). Association for Computational Linguistics.

Iivari, J. (2005). An empirical test of the DeLone-McLean model of information system success. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 36(2), pp.8-27.

Ives, B., Olson, M. and Baroudi, J.J. (1983). The measurement of user information satisfaction.

Jiang, J.J., Klein, G. and Crampton, S.M. (2000). A note on SERVQUAL reliability and validity in information system service quality measurement. *Decision Sciences*, 31(3), pp.725-744.

Jiang, J.J., Klein, G. and Carr, C.L. (2002). Measuring information system service quality: SERVQUAL from the other side. *MIS quarterly*, pp.145-166.

Johnson, K. (2017). Facebook Messenger hits 100,000 bots. [online] VentureBeat. Available at: https://venturebeat.com/2017/04/18/facebook-messenger-hits-100000-bots/ [Accessed 16 Jun. 2018].

Kayak.co.uk. (2017). *KAYAK Mobile Travel Report: Chatbots in the UK*. [online] Available at: https://www.kayak.co.uk/news/mobile-travel-report-2017 [Accessed 14 Jul. 2018].

Keen, P.G. (1987). MIS research: Current status, trends and needs. *Information systems education: Recommendations and implementation*, 1, p.13.

Lee, M. and Ulgado, F.M. (1997). Consumer evaluations of fast-food services: a cross-national comparison. *Journal of Services Marketing*, 11(1), pp.39-52.

Lillrank, P., 2003. The quality of information. *International Journal of Quality & Reliability Management*, *20*(6), pp.691-703.

Loiacono, E.T., Watson, R.T. and Goodhue, D.L. (2002). WebQual: A measure of website quality. *Marketing theory and applications*, 13(3), pp.432-438.

Loiacono, E.T., Watson, R.T. and Goodhue, D.L. (2007). WebQual: An instrument for consumer evaluation of web sites. *International Journal of Electronic Commerce*, 11(3), pp.51-87.

Loomba, A.P. (1998). Product distribution and service support strategy linkages: an empirical validation. *International Journal of Physical Distribution & Logistics Management*, 28(2), pp.143-161.

Luarn, P. and Lin, H.H. (2003). A customer loyalty model for e-service context. *J. Electron. Commerce Res.*, 4(4), pp.156-167.

McKinney, V., Yoon, K. and Zahedi, F.M. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Information systems research*, 13(3), pp.296-315.

Markoff, J., 2011. Computer wins on 'jeopardy!': trivial, it's not. *New York Times*, 16.

Mauldin, M.L. (1994). Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In *AAAI* (Vol. 94, pp. 16-21).

McGrath, R. (2018). *How To Improve Customer Service With Chatbots. [online] Chatbots Magazine*. Available at: https://chatbotsmagazine.com/ill-never-buy-from-them-again-using-chatbots-to-avoid-bad-customer-service-e6a967360244 [Accessed 5 Mar. 2019].

Miller, H. (1996). The multiple dimensions of information quality. *Information Systems Management*, 13(2), pp.79-82.

Negash, S., Ryan, T. and Igbaria, M. (2003). Quality and effectiveness in web-based customer support systems. *Information & management*, 40(8), pp.757-768.

Nelson, R.R., Todd, P.A. and Wixom, B.H. (2005). Antecedents of information and system quality: an empirical examination within the context of data warehousing. *Journal of management information systems*, 21(4), pp.199-235.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1988). Servqual: A multiple-item scale for measuring consumer perc. *Journal of retailing*, 64(1), p.12.

Parasuraman, A., Zeithaml, V.A. and Malhotra, A. (2005). ES-QUAL: a multiple-item scale for assessing electronic service quality. *Journal of service research*, 7(3), pp.213-233.

Peterson, B. (2017). *How chatbots could change customer service over the next 5 years*. [online] Nordic.businessinsider.com. Available at: https://nordic.businessinsider.com/how-chatbots-could-change-customer-service-over-the-next-5-years-2017-9?r=US&IR=T [Accessed 14 Jul. 2018].

Petter, S., DeLone, W. and McLean, E. (2008). Measuring information systems success: models, dimensions, measures, and interrelationships. *European journal of information systems*, 17(3), pp.236-263.

Radziwill, N.M. and Benton, M.C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv*:1704.04579.

Rai, A., Lang, S.S. and Welker, R.B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. *Information systems research*, 13(1), pp.50-69.

Reshmi, S., & Balakrishnan, K. (2016). Implementation of an inquisitive chatbot for database supported knowledge bases. *Sādhanā*, 41(10), 1173-1178.

Roca, J.C., Chiu, C.M. and Martínez, F.J. (2006). Understanding e-learning continuance intention: An extension of the Technology Acceptance Model. *International Journal of human-computer studies*, 64(8), pp.683-696.

Seddon, P. and Yip, S.K. (1992). An empirical evaluation of user information satisfaction (UIS) measures for use with general... *Journal of Information Systems*, 6(1), pp.75-92.

Seddon, P. and Kiew, M.Y. (1996). A partial test and development of DeLone and McLean's model of IS success. *Australasian Journal of Information Systems*, 4(1).

Sedera, D. and Gable, G.G. (2004). A factor and structural equation analysis of the enterprise systems success measurement model. Association for Information Systems.

Schumaker, R.P., Ginsburg, M., Chen, H. and Liu, Y. (2007). An evaluation of the chat and knowledge delivery components of a low-level dialog system: The AZ-ALICE experiment. *Decision Support Systems*, 42(4), pp.2236-2246.

Shawar, B.A. and Atwell, E. (2003). Using dialogue corpora to train a chatbot. In *Proceedings of the Corpus Linguistics 2003 conference* (pp. 681-690).

Shawar, B.A. and Atwell, E. (2007). Chatbots: are they really useful?. In *Ldv Forum* (Vol. 22, No. 1, pp. 29-49).

Srinivasan, A. (1985). Alternative measures of system effectiveness: associations and implications. *MIS quarterly*, pp.243-253.

Strong, D., Lee, Y.W. and Wang, R.Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), pp.103-111.

Tan, K.C., Xie, M. and Li, Y.N. (2003). A service quality framework for web-based information systems. *The TQM Magazine*, 15(3), pp.164-172.

Thompson, C., 2010. What is IBM's Watson. *New York Times Magazine (June 2011)*. https://www.nytimes.com/2010/06/20/magazine/20Computer-t.html.

Urbach, N. and Müller, B. (2012). The updated DeLone and McLean model of information systems success. In *Information systems theory* (pp. 1-18). Springer, New York, NY.

Van der Wal, R.W.E., Pampallis, A. and Bond, C. (2002). Service quality in a cellular telecommunications company: a South African experience. *Managing Service Quality: An International Journal*, 12(5), pp.323-335.

Wang, R.Y. and Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), pp.5-33.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), pp.36-45.

Weizenbaum, J. (1976). Computer power and human reason: From judgment to calculation.

Xu, A., Liu, Z., Guo, Y., Sinha, V. and Akkiraju, R., (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3506-3510). ACM.

Yeung, K. (2016). Facebook opens its Messenger platform to chatbots. [online] VentureBeat. Available at: https://venturebeat.com/2016/04/12/facebook-opens-its-messenger-platform-to-chatbots/ [Accessed 22 Apr. 2019].