COMMENTS ABOUT THE SIRI CHATBOT: A SENTIMENT ANALYSIS OF THE POSTINGS AT A MICROBLOGGING SITE

HIO NAM IO
UNIVERSITY OF MACAU
yb57021@um.edu.mo

CHANG BOON PATRICK LEE
UNIVERSITY OF MACAU
cblee@um.edu.mo

ZHAOTONG LIAN
UNIVERSITY OF MACAU
lianzt@um.edu.mo

ABSTRACT

This study used sentiment analysis to analyze 232,496 *Siri*-related comments posted on the Weibo website. The results indicate that most of the posts have positive sentiments. Further analyses conducted using keyword extraction indicate that 'come on to' is among the most frequently used words in the posts that have highly positive or highly negative sentiments. This finding about 'come on to' is unusual and unexpected as there has been no discussion about this aspect of user experience in the literature. The results provide motivation for future research to explore and validate the significance of 'come on to' when users interact with chatbots. The results also indicate that developers should build chatbots that can handle challenging inputs that involve 'come on to' from users as this would elicit positive sentiments about their user experience.

Keywords: chatbot, opinion analysis, sentiment analysis, *Siri*, social media, user experience

INTRODUCTION

Artificial intelligence (AI) involves the attempt to build computer systems that think and act like humans [9]. A renowned computer scientist has referred to AI as the 'New Electricity' [18]. This 'New Electricity' is transforming the current technology landscape. Computers that traditionally do simple tasks such as performing calculations, bookkeeping, and repetitive activities can now extend their capabilities to do more complex work such as

recognizing speech patterns, analyzing big data, and making smart recommendations.

An example of a widely-used AI application is the chatbot. Chatbots are also known as conversation agents. They conduct conversations with their users based on the human language [1]. Chatbots are different from computers that use command-line or graphical interface. When using command-line interface, syntax is required to communicate with the computer. In graphical interface, users need to click the appropriate icons to perform specific actions. In both interfaces, users are confined to a set

of predefined commands. On the other hand, chatbots respond to what users speak or write. Using a chatbot is similar to communicating with another person and therefore chatbots are very appealing to the users. Some chatbots have specific domain knowledge and they can provide answers to queries like train schedules and weather forecasts. Many organizations realize that chatbots are very useful and they have used them to provide 24/7 customer services. Organizations such as those in the e-commerce, healthcare, and banking industries have used chatbots to support their online business operations [20; 26].

For chatbots to work successfully, they must first understand users' inputs correctly. When chatbots misinterpret the users' inputs, they will not be able to give appropriate answers. Nowadays, system developers use AI to train chatbots to comprehend users' inputs much better than before. Besides understanding users' inputs, chatbots must also give satisfactory responses. Some chatbots failed because they provided inappropriate responses. A famous example is the case of Tay, an experimental chatbot launched by Microsoft. Tay was trained based on its interactions with Twitter's users. There was no control over the contents of Tay's outputs. Tay was shut down one day after it was deployed because of its obscene and inflammatory remarks [17]. Even though Tay attracted a lot of attention, the chatbot failed because its users were unhappy when using it.

As chatbots are a recent innovation, it is useful to determine people's sentiments about them. Apple's *Siri* was selected as the chatbot for this study due to three reasons: (1) *Siri* is one of the earliest chatbot available on mobile devices and many users have experienced interacting with *Siri* by using Apple's iOS devices [14]; (2) Compared to other famous chatbots in the market, *Siri* has a relatively unique name that is used in different languages, and this is helpful when collecting relevant data; (3) *Siri* has generated lots of discussions on social media. The purpose of this research is to address the following research question: 'What are people's sentiments about *Siri*?' Answering this research question can shed light on what people like or dislike about *Siri* and the results can help researchers gain insights into the users' experience with *Siri*. They can also help practitioners develop better chatbots in the future. To answer the research question, this study extracted *Siri*-related comments from a popular social media website called Weibo [29], and analyzed the data to understand people's sentiments about *Siri*.

The next section of this paper presents related works on chatbots, social media mining, and sentiment analysis. The subsequent sections describe the research method used in this study and the results and analysis. This is followed by a discussion of the results and their implications. The final section summarizes the findings and concludes with the limitations of the study.

RELATED WORKS

Chatbots

The performance of chatbots has improved over the years. In the early years of chatbot development, developers use the Artificial Intelligence Markup Language (AIML) to enable chatbots to 'understand' users' inputs. The AIML scripts contain pattern tags and possible answers. When chatbot detects the patterns from the users' inputs, it will select a possible answer from its knowledge base. There are some disadvantages of using AIML, however. First, it cannot afford the large knowledge base as developers need to record different patterns in the scripts to enable chatbots to understand users' inputs [25]. Second, when knowledge is dynamic, the response may not be up to date. Third, the human language involves many shades of meanings. It is not easy for chatbots to have a complete understanding just by recognizing some patterns in the sentences.

To overcome the disadvantages of AIML, modern chatbots apply Deep Learning (DL) and Natural Language Processing technologies to enhance their functionality and performance [13]. DL has attracted lots of attention since 2015 [11]. It is a sub-branch of machine learning and makes use of huge amount of data to train chatbots. Machine-learning refers to algorithms that detect patterns. It is able to learn how to make predictions and recommendations by processing data and experiences, rather than receiving explicit programming instructions. DL can often produce better results than traditional machine learning approaches because it makes use of what it has learned to make inferences about new data.

Chatbots are developed to serve different purposes. One way to classify chatbots is based on whether they are task-oriented or chat-oriented. Alime, which is a chatbot developed by the Alibaba Group, is a task-oriented chatbot. Alime serves online buyers by answering questions related to its e-commerce functions [21]. On the other hand, Apple's *Siri* is an example of a chat-oriented or general-purpose chatbot. *Siri* responds to various types of questions. For example, users can ask *Siri* to perform some tasks on their mobile devices, such as turning on an alarm, setting a schedule, or making a phone call. Users can also chat with *Siri* on many topics, such as having a broken heart, losing a job or when they are feeling down. There are other examples of chat-oriented chatbots in the marketplace. They include Amazon Alexa and Google Assistant. Microsoft China has also created a

chat-oriented chatbot called XiaoIce. XiaoIce is available on the WeChat messaging apps [24].

Another way to classify chatbots is based on their inputs. Some chatbots can only understand text inputs, while others can also respond to voice inputs. Google Assistant can respond to both types of inputs. A recent study has compared some popular speech-based chatbots and found that these chatbots provide many services [14].

There are a few recent studies that focused on some specific aspects of chatbots [2; 10; 22]. These studies are mostly conducted with small groups of participants. The current study is different from prior studies because the data used for this study is based on a large collection of secondary data extracted from a social media website. According to the authors' knowledge, there is no prior study that used social media data to analyze users' sentiments about chatbots.

Social media mining

Social media websites facilitate the creation and sharing of information, ideas, and opinions through the Internet. Twitter and Weibo are two popular social media websites where users can express their moods and opinions about specific products or events in their posts. The rapid growth of social media has created huge amount of data related to user opinions and comments. It is useful to mine these data to extract valuable knowledge, patterns, and insights [7]. Indeed, many recent studies have used data from social media for their research [5]. Results obtained from social data mining can help to support inductive work [15]. They may also present researchers the opportunity to generate new theories [3].

There are many different types of social networking websites. For example, Instagram and Pinterest focus on photo sharing, Youtube and Youku on videos, while Facebook, Twitter, and Weibo provide services, such as allowing users to share text, photos, videos, and links. Depending on the website policies, researchers may have limited access to historical data on these websites due to security reasons. Access to data may also be limited because some social media websites do not provide an application programming interface to enable researchers to have access to the data.

Studies based on social data mining is evolving across many disciplinary areas. These studies used various data sources and text mining techniques to analyze social media data. For example, studies related to tourism have used web crawlers to collect customer review data from tourism websites such as Expedia.com and Tripadvisor.com to analyze guest experience [7; 28]. Studies on information systems have used Twitter data to conduct

research on information security issues using techniques such as topic modeling and sentiment analysis [12]. In finance, researchers have applied sentiment analysis on social media data to predict stock movements [19].

Sentiment analysis

Sentiment analysis has been widely applied to analyze social media data. The purpose of sentiment analysis is to mine textual comments to detect favorable or unfavorable opinions about specific subjects. Researchers have developed many techniques to judge the sentiment polarity of comments. These techniques ranged from the simple approach such as counting the sentiment scores for each word in a corpus, to the more advanced machine learning approaches such as the kNN, Naive Bayes, and support vector machine [6]. The details for various techniques used in sentiment analysis can be found in [16].

Sentiment analysis is an automated process and it can be conducted very quickly. It is unlike manual coding of data which takes a much longer time. There are still challenges in performing sentiment analysis, however. First, the sentiment analysis software may not be able to understand the human language completely as some words or sentences have multiple meanings. Second, there may be loss of information as the results of sentiment analysis provide only a summary measure. It will be useful, therefore, to supplement sentiment analysis with other research techniques such as keyword extraction and content analysis. Using these techniques can help to facilitate better understanding of the data [8].

RESEARCH METHOD

Weibo is a microblogging website. Microblogging refers to a combination of blogging and instant messaging that allows users to create short messages to be posted and shared with an audience online. The research method for this study involved using the word *Siri* to search for related posts in Weibo. The *Siri*-related posts were used as the corpus for sentiment analysis.

Figure 1 shows the workflow for the research. Several tools were used to collect and analyze the data. First, GooSeeker (http://www.gooseeker.com/) was used to crawl and search for the related posts in Weibo. GooSeeker is a web crawler software that crawls the text data from different webpages. It crawled the posts from Weibo based on the search word and the specified time period. The time window for this study was from October 2011, which was the time Apple introduced the iPhone 4S with a beta version of *Siri*, to June 2017. Altogether, there were 241,898 *Siri*-related posts during the study period. The posts were written in Simplified Chinese and only

textual data were used in this study. After data pre-processing, such as removing duplicate posts from the same user, 232,496 posts were available for analysis.

This research used the BosonNLP software to conduct sentiment analysis. The sentiment analysis engine in BosonNLP (http://bosonnlp.com/) was trained using data from social networks and news corpora [31]. The software is therefore considered 'Weibo-friendly'. Several prior studies [27; 30] had used BosonNLP to conduct sentiment analysis.

BosonNLP will compute a sentiment score for each post. The score ranges from 0 to 1. When the score is close to 0, it means the post has a negative sentiment. Otherwise, the post has a positive sentiment when it is close to 1. The study used the Picdata tool (http://www.picdata.cn/) to perform keyword extraction. The extraction was done based on the count of high-frequency words in the positive and negative posts. After extracting the keywords, the study conducted a content analysis of the posts associated with the keywords.
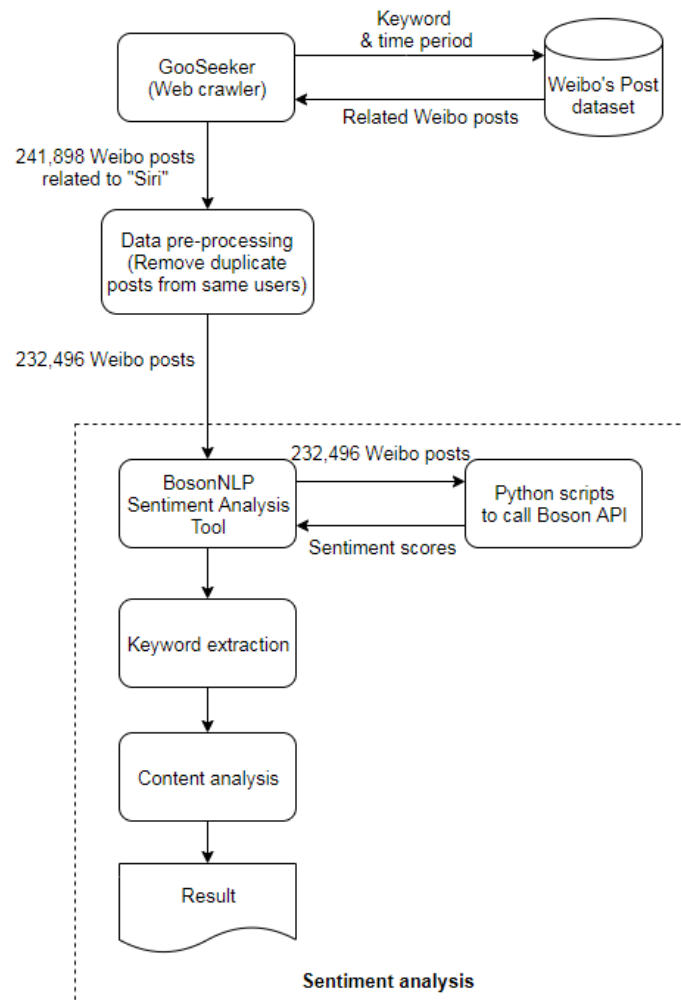

Figure 1: The workflow of the sentiment analysis

RESULTS AND ANALYSIS

The overall trend of *Siri*-related posts

Figure 2 shows the number of *Siri*-related posts in Weibo across time. The figure shows that discussions about *Siri* experienced several peaks and troughs. In October 2011, when Apple first embedded *Siri* in its iOS operating system during the release of iPhone 4S, *Siri* generated a lot of interest as the concept of a chatbot was relatively new. After *Siri* was launched, however, the initial interest dropped and there were fewer postings between 2013 and 2014. From the year 2015, DL became available and using DL in chatbots enabled them to perform much better than before. It is possible this could have contributed to the sporadic interests in *Siri* during 2015 and 2016.

Results of sentiment analysis

The sentiment score of 0.5 is the threshold value to distinguish between positive and negative sentiments

[23]. Table 1 presents the results of sentiment scores. The results show that 76 percent of the posts have scores that are above 0.5. The rest of the posts have scores that are 0.5 or less. The results show that most of the posts have positive sentiments.

Results of keyword extraction

The study selected the posts with the most positive sentiments (score more than 0.9) and the most negative sentiments (score less than 0.1) and extracted keywords based on their frequencies. Keywords such as 'Apple', 'Weibo' and 'iPhone' were excluded as they were used mainly for identification purposes. Table 2 presents the top 10 keywords from the posts with sentiment scores more than 0.9 while Table 3 presents the top 10 keywords from the posts with sentiment scores less than 0.1. The tables show that the keywords with positive scores appear in higher frequency than those with negative scores. Hence, these results also show that the sentiments of the comments are mostly positive.
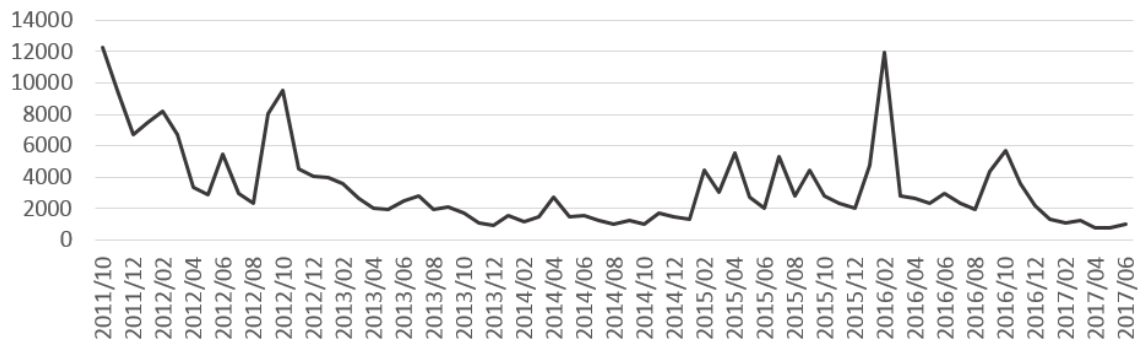

Figure 2: Number of Weibo posts commenting about "*Siri*"

Table 1: Results of sentiment scores

Sentiment type / Score	No. of posts (% in all posts)	Mean of sentiment scores (μ)	Std. Dev of sentiment scores (σ)
Positive / (0.5, 1]	176,923 (76.10%)	0.83	0.13
Negative / [0, 0.5]	55,573 (23.90%)	0.27	0.14

Table 2: Top 10 keywords from posts with sentiment scores > 0.9

Terms	Frequency	Percentage
'Come on to' (调戏)	6,156	24%
Chinese (中文)	4,785	19%
Intelligent (智能)	2,772	11%
Assistant (助手)	2,307	9%
Lovely (可爱)	2,232	9%
Funny (好玩)	1,684	7%
Powerful (强大)	1,497	6%
Bored (无聊)	1,428	6%
Cantonese (粤语)	1,235	5%
Singing (唱歌)	1,216	5%
Total	25,312	

Table 3: Top 10 keywords from posts with sentiment scores < 0.1

Terms	Frequency	Percentage
English (英语)	839	19%
'Come on to' (调戏)	707	16%
Chinese (中文)	669	15%
Bored (无聊)	428	10%
Reply (回答)	413	10%
Chat (聊天)	272	6%
Angry (气死)	337	8%
Sound (声音)	296	7%
English-Speaking (英语口语)	199	5%
Master (主人)	183	4%
Total	4,343	

Results of keyword analysis

The keyword analysis found that some keywords appear in both Tables 2 and 3, indicating that these keywords have both positive and negative sentiments. Some keywords are related and they can be grouped into themes. For example, keywords such as Chinese, Cantonese, English, and English-speaking are related to language. Other keywords are related to using *Siri* for different usage intention – utilitarian or hedonic purposes. Utilitarian purpose means users use *Siri* to accomplish a functional task, while hedonic purpose means users use *Siri* for entertainment and fun. Some of the keywords are also related to the users experiencing anger. To consolidate the results of the keyword extraction, the keywords were grouped into five themes. The themes and the related keywords (shown in brackets) are as follows:

Theme#1: Sentiments related to 'come on to' ('Come on to', Bored),

Theme#2: Sentiments related to language (Chinese, Cantonese, English, English-speaking),

Theme#3: Positive sentiments arising from utilitarian purposes (Intelligent, Assistant, Powerful),

Theme#4: Positive sentiments arising from hedonic purposes (Lovely, Funny, Singing), and

Theme#5: Negative sentiments that anger users (Reply, Angry, Sound, Master, Chat).

Figure 3 shows the grouping of keywords based on the themes that have been identified. The themes are located on the chart using two dimensions: sentiment polarity (positive and negative) and usage intention (hedonic-based versus utilitarian-based). The first and second themes include both positive and negative sentiments. The third and fourth themes have positive sentiments, while the last theme has negative sentiments. With regard to the purpose of using *Siri*, the first, fourth, and fifth themes are related to the hedonic aspects of *Siri*, while the second, third, and fifth themes are related to the utilitarian aspects of *Siri*. The themes are described in the following sections.

Figure 3: Grouping of keywords

Theme#1: Sentiments related to 'come on to'

     'Come on to' is the most frequently used words in the positive posts and the second most frequently used words in the negative posts. 'Come on to' is translated from the Chinese phrase '调戏'. Given the high frequency of occurrence of this phrase, this study infers that it has special significance to *Siri*. The meaning of '调戏', as provided by Oxford Chinese, Yabla, Lingq, Google, Baidu, and Yandex, includes 'to take liberties with', 'making sexual advances', 'to flirt, 'to play around with', 'to assail a woman with obscenities', 'to dally', 'to 'molest', and 'to trick'. As '调戏' has a broad meaning, the study translated it to 'come on to' to reflect Weibo's context, which is informal and colloquial. Two language experts were consulted regarding the translation and they agreed that 'come on to' is a suitable translation.

     Table 4 presents some examples of positive posts pertaining to 'come on to'. One user shared screenshots of the conversation with *Siri* that showed how to 'come on to' *Siri*. This user added that *Siri* was very intelligent. The users loved to ask *Siri* to do some tasks related to 'come on to' or they would say something to *Siri* that was not easy when they talked to a human being. One user asked *Siri* to sing and tell a joke. Another said 'I love you' to *Siri*. The users wanted to see how *Siri* reacted to their comments. They were happy when they managed to 'come on to' *Siri*. One post commented, 'Haha, I succeeded to come on to *Siri* and asked her to sing'. Another said '… *Siri* asked me to go to bed'. Other posts in this theme described how they experienced joy when they 'come on to' *Siri*: 'Come on to *Siri* is very funny', 'I continue to come on to *Siri*. She can sing and tell a joke. *Siri* is so funny!', and 'A good day starts from coming on to *Siri*'.

     There are also negative sentiments related to 'come on to'. The negative comments happened when users were 'come on to' by *Siri*. For example, users felt ridiculed by *Siri* or when *Siri* could not accede to their

request to tell a joke. Table 5 illustrates a few situations when users were 'come on to' by *Siri*.

'Bored' is a term that appears frequently together with 'come on to'. Feeling bored is one reason why users 'come on to' *Siri*. Like 'come on to', 'bored' is related to both positive and negative sentiments. The positive sentiments occurred when *Siri* made users happy, comfortable, or when *Siri* showed them some powerful capabilities that would capture their attention. On the other hand, when users were not amused by *Siri*'s responses, their sentiments were low. Table 6 shows examples of posts for these two situations.

In summary, the theme for 'come on to' is unusual and unexpected. The literature has neither reported nor discussed this aspect of users' experience with chatbots. The comments about 'come on to' indicate that users 'come on to' *Siri* for hedonic purposes. The way *Siri* responded to 'come on to' had an impact on the users' sentiments. Users had positive sentiments when they managed to 'come on to' *Siri,* such as making *Siri* tell a joke or sing a song. They had low sentiments when they felt they were 'come on to' by *Siri* or they were ridiculed or denied their requests by *Siri*.

Table 4: Selected positive posts related to 'come on to'

Post content	Sentiment score
Come on to Siri with positive feelings	
This is a collection of how to **come on to** *Siri*, the bot is so intelligent. (User showed several screenshots of the conversation interacting with *Siri*)	0.91
Tonight I **come on to** *Siri* and say "I love you", then *Siri* asks me to go to bed. Hahaha…	0.94
Haha, I succeeded to **come on to** *Siri* and asked her to sing!	0.98
Come on to *Siri* is very funny.	0.97
I continue to **come on to** *Siri*. She can sing and tell a joke. *Siri* is so funny!	0.93
A good day starts from **coming on to** *Siri*.	0.96

Table 5: Selected negative posts related to 'come on to'

Post content	Sentiment score
When users are 'come on to' by Siri	
At the beginning, I feel bored and want to **come on to** *Siri*. However, I am come on to by her finally.	0.07
I am very bored and **come on to** *Siri*… but I'm furious as I am **come on to** by *Siri* all the time.	0.08
I ask *Siri* to tell a joke, but she didn't, and finally terminate our conversation… I think I have been **come on to** by *Siri*.	0.07
I feel bored, and want to **come on to** *Siri*. Then I am upset by her merciless ridicule.	0.01

Table 6: Selected posts related to 'come on to' and 'bored'

Post content	Sentiment score
Positive sentiment	
Come on to *Siri* is very funny, I can chat with *Siri* when I feel **bored** in the future.	0.97
Siri is so powerful, I can **come on to** my mobile phone when I feel **bored**.	0.97
I feel **bored** during the class and **come on to** Siri, what an interesting robot!	0.93
Negative sentiment	
I am really **bored**, couldn't fall asleep, then I **come on to** *Siri*, but I am still **bored**…	0.07
It is raining and I cannot go anywhere, I am very **bored** and **come on to** *Siri*. However, *Siri* cannot understand me.	0.05
I couldn't fall asleep then I **come on to** *Siri*. She is very **boring** and stupid. I think if she is a human being, I will be very angry.	0.08

Theme#2: Sentiments related to language

The main language among Weibo users is Chinese. During its initial launch, *Siri* was available only in English. *Siri* started to support Chinese and Cantonese only after 2012. During the initial launch, many users were disappointed because *Siri* could not understand their English accents. This led many users to lament their English-speaking ability. The users felt positive when *Siri* started supporting the Chinese language. Hence, in Figure 3, the theme related to language overlays both positive and negative sentiments.

Table 7 presents the posts related to languages. The posts are categorized into two headings: (1) *Siri* in Chinese/Cantonese and (2) *Siri* in English/Chinese. The posts pertaining to Chinese/Cantonese are positive, while those pertaining to English/Chinese are negative. Many of the positive posts associate language with how smart *Siri* was and how easy it was to communicate with *Siri*.

In summary, the language theme shows that users were happy when *Siri* supported languages such as Cantonese and Chinese. They were happy because *Siri* could understand what they said and they perceived *Siri* as smart and easy to use.

Table 7: Selected posts related to languages

Post content	Sentiment score
Siri in Chinese/Cantonese	
This is the official introduction video about iPhone4S, *Siri* sounds good, despite it doesn't support **Chinese**!	0.91
Except not able to speak **Chinese**, *Siri* is quite smart.	0.93
The **Cantonese** version of *Siri*, whether you want to purchase or not, you should look at how funny it is.	0.94
The **Cantonese** version of *Siri* is awesome.	0.96
Siri in English/Chinese	
Learning **English** is more and more important, otherwise, you cannot even use *Siri*, and will be eliminated gradually…	0.07
Limited in hand, regretted in mind. Poor **English** makes you not able to come on to *Siri*.	0.03
Siri cannot understand my **English**; I feel regret about my poor **English-speaking**.	0.03
Siri fights against my **English-speaking**! I just say "12345", and she cannot understand, and she interprets as "one two three phone five"	0.03
When will the **Chinese** version of *Siri* be launched……the **English** version is difficult to use……	0.03

Theme#3: Positive sentiments arising from utilitarian purposes

Weibo users showed positive sentiments when they used *Siri* for utilitarian purposes. They used keywords like 'intelligent', 'assistant' and 'powerful' to describe *Siri*. They said *Siri* can understand the user's commands and it can perform different tasks, such as adding an event to the calendar, calling a friend, and switching on an alert function. Analysis of the posts containing the three keywords ('intelligent', 'assistant' and 'powerful') shows that *Siri* brings users positive sentiments because of its usefulness. On the whole, this theme shows that users felt happy when *Siri* provided them with utilitarian services. Table 8 illustrates some posts on *Siri*'s utility.

Theme#4: Positive sentiments arising from hedonic purposes

The users had positive comments when they used *Siri*'s for hedonic purposes. *Siri* can entertain by 'singing' and the users found *Siri* 'lovely' and 'funny'. The chatbot helped users to avoid boredom. Users treated *Siri* as a "friend" or a "toy" for entertainment to satisfy their intrinsic needs. This theme shows that users enjoyed using *Siri*. Table 9 illustrates some examples where the users provided positive sentiments when they were using *Siri* for hedonic purposes.

Table 8: Selected posts containing 'intelligent', 'assistant' and 'powerful'

Post content	*Sentiment score*
The uniqueness of *Siri* is the integration of speech recognition and artificial **intelligence**, and connected with calendar, contacts, GPS and other mobile apps, which makes *Siri* not only able to recognize the voice commands, but also react based on different user locations.	0.92
Siri can make iPhone 4S become an **intelligent** robot, users can make use of *Siri* to read messages, restaurants introduction, weather inquiry, set alarm with voice, etc.	0.93
Siri can understand natural languages according to context and human history. *Siri* has artificial **intelligence** and will understand what you mean. For example, you want to take a nap and prepare to set an alarm, just say 'wake me up after 20 minutes'.	0.95
The iPhone 4S built-in **intelligent** voice **assistant** *Siri* not only can complete the daily weather inquiries, sending text messages, searching and other functions, it can also accompany you to chat.	0.93
Siri is the most **powerful** function in iPhone 4S, like a chatbot. One more powerful function is that *Siri* can solve math problems.	0.97
Now, the functions of *Siri* have become more and more **powerful**.	0.95

Table 9: Selected posts containing the terms 'lovely', 'funny' and 'singing'

Post content	*Sentiment score*
Too many people ask *Siri* to **sing** PPAP. Let me try. She is very **lovely**.	0.96
Siri is very **lovely**. I ask to unlock my mobile phone password. It gives me various answers.	0.91
Notice the conversations between my son and *Siri*, AI is very **lovely**.	0.99
Siri is very **funny**, I will have something to do when I am bored!	0.96
I found something funny when I was **bored**, you can try to ask *Siri* to tell a story or a joke repeatedly, she is very **funny**.	0.94
Hahaha, go to say Happy New Year to *Siri*! It is very **funny** hahaha.	0.99
I test *Siri* to **sing**, it will reject at the first time, if you request again, it will start to "act cute". It is too **funny**.	0.96

Theme#5: Negative sentiments that angered users

When *Siri* cannot satisfy the users, the users' emotions may become negative. The terms 'reply', 'angry' and 'sound' are associated with the users' negative sentiments. Sometimes, *Siri* cannot understand the requirements from the users and *Siri* provided unsatisfactory replies. At other times, the users had high expectations about *Siri* and thought it could do everything, and so they asked *Siri* to solve very difficult problems. The users also liked to control *Siri* and treated *Siri* like a servant. In one post, for example, a user wrote: '*Siri*, how can you say such things to your master?'

As *Siri* used voice to communicate with users, it did not allow users to type. There are pros and cons in using a voice chatbot. One advantage is that users can send a command or message to *Siri* in a shorter time without typing. However, forcing users to use their voice to interact with a chatbot also has problems. For instance, in some noisy situations, it is difficult to speak clearly to *Siri*. Also, *Siri* may speak unexpectedly, causing embarrassment to the users. Table 10 shows examples where *Siri* made the users angry.

In summary, this theme shows how *Siri* created negative sentiments among the users. The comments show that *Siri* still needs improvement to provide better service in the future.

Table 10: Selected posts where *Siri* made the users angry

Post content	Sentiment score
When I am bored and chatted with *Siri*, I asked *Siri* to tell me a joke, but she keeps rejecting me. So **angry**!	0.01
My little girl was bored and chatted with *Siri*. She had several requests for *Siri*, but *Siri* cannot give a satisfying **reply**, and my little girl got angry: What can you do for me?! Pick up your baggage and go away, don't stay here to waste my RAM!	0.05
I asked *Siri* what I can do if I did not want to get up. It **replied** it cannot answer my question.	0.09
It rained today, but I didn't have an umbrella. Then, I took out my mobile phone, and asked *Siri*, "What can I do when I don't have an umbrella when it is raining?" –It said, "I have found this information for you" –I said, "You should find someone to pick me up" –It said, "I have tried my best" –I said, "Rubbish! I am **angry** with *Siri*."	0.03
Siri, you are so naughty, how can you say such things to your	0.05
During the forum today, when my boss was having a speech next to me, I played the game with my mobile phone. Suddenly my boss turned to me, I was scared. When I was going to turn off the game, I turned on *Siri* unexpectedly, then everybody can hear a loud **sound** from *Siri*, "I'm sorry, I don't understand what you just said" …	0.07
What an embarrassing moment. When my teacher was talking and I was doing my assignment, I didn't know how my elbow touched the home button of my mobile phone, and this was the moment my teacher finished the class, then my *Siri* suddenly **sounded** out "I'm sorry, I don't understand what you just said."	0.05

DISCUSSION

This section discusses the results related to using *Siri* for utilitarian and hedonic purposes. It then goes on to discuss the implications of the results for research and practice.

Siri for utilitarian purposes

Referring to Figure 3, the terms 'intelligent', 'assistant' and 'powerful' are keywords that have positive sentiments arising from using *Siri* for utilitarian purposes. They show that users were happy with the utilitarian functions. For example, people asked *Siri* to assist them to perform some daily tasks, such as setting an alarm or enquiring about the weather. The more intelligent and powerful the assistant, the more the users perceived *Siri* to be useful. *Siri* could enhance the users' convenience and therefore improve their experience with the chatbot.

The ease of use is also an important factor influencing the user experience. Even if a system provides powerful functions and capability to enhance its performance, there may be barriers for users to use the system if they need to spend much effort to use it. In the case of *Siri*, language is critical for the users to interact with the chatbot. During the initial launch of *Siri* when the chatbot did not support the 'Chinese' language, people reflected their negative sentiments complaining about the difficulty of using the chatbot. For example, one user said *Siri* could not understand his/her English accent. When *Siri* supported 'Cantonese', however, people showed their surprise and happiness with their positive posts.

Siri is a voice chatbot. Many chatbot developers enable users to initiate conversation using specific phrases. In the case of Apple devices, users can start by saying "Hey *Siri*". However, voice chatbots may create embarrassing situations when they speak unexpectedly. Even though it is convenient to use voice chatbots, balancing convenience and controllability may be critical to the user experience.

Siri for hedonic purposes

The content analysis of the posts indicates that users often treat *Siri* as a person. In traditional IS that are either utilitarian or hedonic systems, the users recognize the system as a machine, and will not treat it as a person. However, in the context of a chatbot such as *Siri*, users seem to treat the system as a person and not merely a traditional IS. For example, no one will treat an automated teller machine as a person and 'come on to' it. Even in a hedonic system, an online game, for example, the users can understand their roles in the game and they will not

speak to the gaming system the way they do to a human. In the case of *Siri*, this study found many cases where users treated *Siri* as a friend or another human being. For instance, users used *Siri* when they felt bored, and they asked *Siri* some difficult questions to explore how *Siri* would respond. The users often used "she" instead of "it" when they talked about *Siri*. The use of words such as 'lovely', 'funny' and 'singing' reflected the users' pleasure and positive emotions after chatting with *Siri*. People were attracted to use *Siri* when it responded in a lovely or funny way, as this made them happy and they enjoyed using it. These results suggest that if the chatbot can personify as a human, then people may be attracted to use it.

Personification is a figure of speech in which inanimate objects are characterized in terms of human attributes, thus representing the object as a living and feeling person [4]. As an aside, personification can be applied in the context of chatbots in different forms, such as (1) using the human face or human features as a chatbot icon, (2) incorporating the properties that human beings own, such as gender, age or birth date, into the chatbots, (3) enabling different ways for the chatbot to interact with the users, for example, using voice or texts, and (4) characterizing the chatbots based on their conversation styles, like embedding human cultures or human characters into the chatbots. Technically speaking, the corpus used to train the chatbots can contribute to personification. For example, the conversation corpus of mature people should be used if the chatbot is to personify an old man.

Implications for research and practice

The results of this study show that people chat with *Siri* when they are bored, and they like to explore how *Siri* will react or perform some difficult tasks, such as singing or telling a joke. 'Come on to' appears very frequently among the comments and using *Siri* to 'come on to' it is quite an unexpected finding. There is no mention about 'come on to' in the literature on the users' experience with chatbots. Future research can explore and validate the significance of 'come on to' when users interact with chatbots.

This study also provides implications for the personification of chatbots. As the chatbot is trained to respond based on its training corpus, the training corpus will have an impact on the users' chatbot experience. Based on different groups of target users, different conversation corpus could be used to train the chatbots. It is imaginable that young users would not understand sentences that are used by the older generation. Also, when chatbots reply with interesting responses, they can help to generate discussions on social media. This implies that

organizations can leverage on the training corpus to promote their chatbots.

The results have valuable implications for chatbot developers. First, even though chatbots may have specific functions and are developed to serve in specific domains, such as finance or e-commerce, they should be trained not only to answer questions in specific areas. They should also be able to handle challenging questions and inputs such as those involving 'come on to'. This can help to enrich the users' experience. Second, the chatbot responses are important to determine whether the users like or dislike the chatbot. Developers, therefore, need to have quality control over the corpus used to train the chatbot or they should perform more comprehensive pilot tests before production. Third, voice chatbot like *Siri* is convenient as users can interact with the chatbot easily without having to type. It can also create problems, however, when the chatbot speaks unexpectedly. To avoid such problems, developers could include multiple communication options for the chatbots.

CONCLUSION

This research conducted a sentiment analysis based on *Siri*-related comments from the Weibo website. The results showed that most comments were positive. Keyword extraction and content analysis were used to further analyze the results. Several frequently used words that accompanied the highly positive or highly negative comments were identified. These keywords were grouped into five themes. The first theme was about 'come on to'. Recall, 'come on to' was the most frequently used words in the positive posts and the second most frequently used words in the negative posts. These results highlighted the significance of 'come on to'. It was possible that people like to 'come on to' *Siri* because in the real world, people cannot say something rude or impolite to other people, and they wondered what would happen when they were to do it to *Siri*. Future research on the users' experience with 'come on to' can help to increase our knowledge about users' engagement with chatbots.

There were positive and negative sentiments related to 'come on to'. These sentiments were related to how *Siri* responded to 'come on to'. The implication of this result is that if the chatbot is not able to respond like a human being, the users may be disappointed. If the chatbot can respond and behave like a human being, the users are likely to be more satisfied when using it. Chatbot developers should therefore consider how they would handle 'come on to' inputs from the users. Also, e-commerce business owners who make use of chatbots in their operational tasks need to develop chatbots that fulfill the users' hedonic needs. Even though these chatbots are developed

for specific purposes, they also need to respond like a human when the users pose comments related to 'come on to'.

The second theme among the keywords was related to language. This theme showed that language is an important communication interface between users and *Siri*. In the context of a voice chatbot, if the chatbot cannot understand what the users say, it will greatly upset the users. It is therefore important for chatbot developers to use a suitable language. The third and fourth themes showed that users had positive sentiments when they used *Siri* for utilitarian and hedonic purposes. They were happy because they perceived *Siri* as 'intelligent', 'powerful', 'lovely', and 'funny'. Chatbot developers should continue to develop chatbots that reinforce these attributes that users like. The fifth theme showed that users had negative sentiments because they were angry with *Siri* for a number of reasons. Some had high expectations of *Siri* while others complained that *Siri* spoke unexpectedly. Developers should use these comments to improve the services provided by their chatbots.

There are several limitations related to this research. First, the data used were based on posts collected from a Chinese microblogging website. Researchers need to be careful when generalizing the results across countries and cultures. Future research can validate the results of this study by using data from other sources as well as from people residing in different geographical locations. Second, with regards to the sentiment analysis tool that is used in this study, the accuracy is around 80 to 85 percent [31]. It may be difficult to achieve absolute accuracy as sometimes the comments may be ironic or ambiguous. Bearing in mind there may be some margin of error, the results of sentiment analysis are nevertheless useful as they offer a big picture of the sentiments embedded in a massive amount of data. Also, as illustrated in this study, researchers can supplement sentiment analysis with keyword extraction and content analysis to provide details and interpretations of the sentiments. The third limitation of this study is that some comments in Weibo could have been written in Traditional Chinese and these comments might not be analyzed accurately. It should be noted, however, the BosonNLP software is trained using data from social media and news articles that could also include some Traditional Chinese. The impact of this limitation may therefore be limited.

To conclude, the results of this study have contributed to our understanding about people's sentiments about *Siri*. The results related to 'come on to' are noteworthy as they are new and interesting, and they have not been reported in the literature. It is important to understand how people interact with chatbots as this knowledge can help to guide successful development of chatbots in

the future. More research on people's experience with chatbots will therefore be very useful.

ACKNOWLEDGEMENTS

REFERENCES

[1] AbuShawar, B. and Atwell, E. 'ALICE chatbot: trials and outputs,' *Computación y Sistemas*, Volume 19, Number 4, 2015, pp. 625-632.

[2] Araujo, T. 'Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions,' *Computers in Human Behavior*, Volume 85, 2018, pp. 183-189.

[3] Berente, N., Seidel, S. and Safadi, H. 'Research commentary—data-driven computationally intensive theory development,' *Information Systems Research*, Volume 30, Number 1, 2019, pp. 50-64.

[4] Delbaere, M., McQuarrie, E. and Phillips, B. 'Personification in advertising,' *Journal of Advertising*, Volume 40, Number 1, 2011, pp. 121-130.

[5] Dindar, M. and Yaman, N.D. '#IUseTwitterBecause: content analytic study of a trending topic in Twitter,' *Information Technology & People*, Volume 31, Number 1, 2018, pp. 256-277.

[6] Feldman, R. 'Techniques and applications for sentiment analysis,' *Communications of the ACM*, Volume 56, Number 4, 2013, pp. 82-89.

[7] Gundecha, P. and Liu, H. 'Mining social media: a brief introduction,' *New Directions in Informatics, Optimization, Logistics, and Production*, September 2012, pp. 1-17.

[8] Hsieh, H. and Shannon, S. 'Three approaches to qualitative content analysis,' *Qualitative Health Research*, Volume 15, Number 9, 2005, pp. 1277-1288.

[9] Laudon, K. and Laudon, J. *Management Information System*, Pearson Education, 16th Global Edition, Malaysia, 2020, pp. 459-460.

[10] Laumer, S., Maier, C. and Gubler, F.T. 'Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis,' *Proceeding of the 27th European Conference on Information System (ECIS)*, 2019.

[11] LeCun, Y., Bengio, Y. and Hinton, G. 'Deep learning,' *Nature*, Volume 521, Number 7553, 2015, pp. 436-444.

[12] Lee, T., Sung, W. and Kim, H. 'A text mining approach to the analysis of information security awareness: Korea, United States, and China', *Proceedings of the 20th Pacific Asia Conference on Information Systems*, Taiwan, June 26 to July 1, 2016.

[13] Li, J., Monroe, W., Ritter, A., Jurafsky, D., Galley, M. and Gao, J. 'Deep Reinforcement Learning for Dialogue Generation,' *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, November 1-5, 2016, pp. 1192–1202.

[14] López, G., Quesada, L. and Guerrero, L.A. 'Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces, '*Advances in Human Factors and Systems Interaction*, Volume 592, Springer, Cham, 2018.

[15] McAbee, S., Landis, R. and Burke, M. 'Inductive reasoning: The promise of big data,' *Human Resource Management Review*, Volume 27, Number 2, 2017, pp. 277-290.

[16] Medhat, W., Hassan, A. and Korashy, H. 'Sentiment analysis algorithms and applications: A survey,' *Ain Shams Engineering Journal*, Volume 5, Number 4, 2014, pp. 1093-1113.

[17] Neff, G. and Nagy, P. 'Automation, algorithms, and politics| talking to Bots: Symbiotic agency and the case of Tay,' *International Journal of Communication*, Volume 10, Number 17, 2016, pp. 4915-4931.

[18] Ng, A. 'Artificial Intelligence is the New Electricity', *presentation at the Stanford MSx Future Forum*, January 2017.

[19] Nguyen, T.H., Shirai, K. and Velcin, J. 'Sentiment analysis on social media for stock movement prediction,' *Expert Systems with Applications*, Volume 42, Number 24, 2015, pp. 9603-9611.

[20] Palanica, A., Flaschner, P., Thommandram, A., Li, M. and Fossat, Y. 'Physicians' perceptions of chatbots in health care: cross-sectional web-based survey,' *Journal of Medical Internet Research*, Volume 21, Number 4, 2019, e12887.

[21] Qiu, M., Li, F.L., Wang, S., Gao, X., Chen, Y., Zhao, W., Chen, H., Huang, J. and Chu, W. 'Alime Chat: A Sequence to Sequence and Rerank based Chatbot Engine,' *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Volume 2, July, 2017, pp. 498-503.

[22] Shum, H.Y., He, X.D. and Li, D. 'From Eliza to XiaoIce: challenges and opportunities with social chatbots,' *Frontiers of Information Technology & Electronic Engineering*, Volume 19, Number 1, 2018, pp. 10-26.

[23] Tao, Y., Zhang, F., Shi, C. and Che, Y. 'Social media data-based sentiment analysis of tourists' air

quality perception' *Sustainability,* Vol. 11, 2019, pp.1-23.

[24] Thies, I.M., Menon, N., Magapu, S., Subramony, M. and O' Neill, J. 'How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians,' *IFIP Conference on Human-Computer Interaction*, Springer, Cham, 2017. pp. 441-459.

[25] Thomas, N.T. 'An e-business chatbot using AIML and LSA,' *Proceedings of 2016 International Conference on Advances in Computing, Communications and Informatics*, India, September 21-24, 2016, pp. 2740-2742.

[26] Trivedi, J. 'Examining the customer experience of using banking chatbots and its impact on brand love: the moderating role of perceived risk,' *Journal of Internet Commerce*, Volume 18, Number 1, 2019, pp. 91-111.

[27] Wang, N., Li, Q., Liang, H., Ye, T. and Ge, S. 'Understanding the importance of interaction between creators and backers in crowdfunding success,' *Electronic Commerce Research and Applications*, Volume 27, 2018, pp. 106-117.

[28] Xiang, Z., Schwartz, Z., Gerdes, J. and Uysal, M. 'What can big data and text analytics tell us about hotel guest experience and satisfaction?' *International Journal of Hospitality Management*, Volume 44, 2015, pp. 120-130.

[29] Yan, R. 'Know your Chinese social media,' *The New York Times,* November 19, 2018.

[30] Yang, C. and Zhang, C. 'Dynamic effectiveness of electronic word-of-mouth and online display advertising on offline sales,' *Proceedings of the 20th Pacific Asia Conference on Information Systems*, Taiwan, June 26 to July 1, 2016.

[31] Ying, K., Jingchang, P. and Minglei, W. 'Research on Sentiment Analysis of Micro-blog's Topic Based on TextRank's Abstract,' *Proceedings of the 2017 International Conference on Information Technology*, Singapore, December 27-29, 2019, pp. 86-90.

AUTHOR BIOGRAPHIES

Chang Boon Patrick Lee is Associate Professor of Business Information Systems at the Faculty of Business Administration, University of Macau. He has a wide range of research interests that include e-commerce management, management of computer professionals, and applications of technology in the gaming and hospitality industry.

Hio Nam Io is a PhD candidate at the Faculty of Business Administration, University of Macau. His research interests include human-robot interactions, social media mining, and e-commerce technology.

Zhaotong Lian is Professor and Associate Dean (Research) of the Faculty of Business Administration, University of Macau. His research interests are Business Analytics, Supply Chain and Service Management.