

Customer Emotions in Service Robot Encounters: A Hybrid Machine-Human Intelligence Approach

Raffaele Filieri

Department of Marketing, Audencia Business School, Nantes, France. Email:
affaele.filieri@audencia.com

Zhibin Lin

Durham University Business School, Durham, UK.
Email: zhibin.lin@durham.ac.uk

Yulei Li

Durham University Business School, Durham, UK.
Email: yulei.li@durham.ac.uk

Xiaoqian Lu

School of Business Administration, Jimei University, Ximen, China.
Email: xiaoqianlu@jmu.edu.cn

Xingwei Yang

Smith School of Business, Queen's University, Canada.
Email: xingwei.yang@queensu.ca

Abstract

Understanding consumer emotions arising from robot-customers encounters and shared through online reviews is critical for forecasting consumers' intention to adopt service robots. Qualitative analysis has the advantage of generating rich insights from data, but it requires intensive manual work. Scholars have emphasized the benefits of using algorithms for recognizing and differentiating among emotions. This study critically addresses the advantages and disadvantages of qualitative analysis and machine learning methods by adopting a hybrid machine-human intelligence approach. We extracted a sample of 9,707 customers reviews from two major social media platforms (Ctrip and TripAdvisor), encompassing 412 hotels in 8 countries. The results show that the customer experience with service robots is overwhelmingly positive, revealing that interacting with robots triggers emotions of joy, love, surprise, interest, and excitement. Discontent is mainly expressed when customers cannot use service robots due to malfunctioning. Service robots trigger more emotions when they move. The findings further reveal the potential moderation effect of culture on customer emotional reactions to service robots. The study highlights that the hybrid approach can take advantage of the scalability and efficiency of machine learning algorithms while overcoming its shortcomings, such as poor interpretative capacity and limited emotion categories.

Keywords: emotion; emotion detection; online customer reviews; machine learning; thematic analysis; service robot.

INTRODUCTION

Consumers increasingly express their emotional responses to service encounters through online consumer reviews (Filieri, 2015; Ye et al., 2020; Yin, Bond, & Zhang, 2014). Online reviews posted by customers on social media platforms represent valuable information about customer experiences, reflecting customers' actual experiences and not hypothesized intentions (e.g., D'Acunto, Volo, & Filieri, 2021; J. Li et al., 2018; Xiang et al., 2015). Traditionally, service researchers have adopted netnography and qualitative analysis to analyze online data (Heinonen & Medberg, 2018). These techniques can generate an in-depth understanding of customer experiences. Still, they are time-consuming and are unable to handle the enormous amount of social media data that is generated daily and is constantly growing (Dhaoui, Webster, & Tan, 2017).

In contrast, machine learning-based social media analytics methods, such as topic modeling, sentiment analysis, and emotion detection, are scalable and efficient (Vermeer et al., 2019) and can improve the generalizability of empirical research through computer algorithms and extensive use of data that reduce the limitations of small sample sizes (J. Li et al., 2018). Analytical technologies dealing with gigantic datasets can help identify hidden patterns in consumers' data and generate insights into the current and future consumer behaviors (Erevelles, Fukawa, & Swayne, 2016). However, the challenges for machine learning methods are their "black boxes" and the accuracy of results (Hayes et al., 2021).

The first aim of this study is to address the weaknesses of both manual and machine learning-based analytics by suggesting a hybrid machine-human intelligence approach that combines the strengths of the latest machine learning tools and those of interpretative research methods. Specifically, we first apply a state-of-the-art topic modeling algorithm, BERTopic

(Grootendorst, 2020), to identify the various robot services, and a deep learning model, the XLNet, to detect customer emotions from online reviews (Yang et al., 2019). This is supplemented with a critical analysis of the findings generated by machine learning, followed by a thematic analysis to explore customers' diverse emotional experiences with service robots (Fereday & Muir-Cochrane, 2006; Spiggle, 1994).

The study's second aim is to contribute to the literature on customers' emotions in service robot encounters (e.g., Belanche et al., 2020; Chuah & Yu, 2021; Mende et al., 2019; Tung & Au, 2018; Van Doorn et al., 2017; Wirtz et al., 2018). Service robots have been increasingly adopted in hospitality service settings in recent years. Large hotel chains have gradually adopted service robots (e.g., Hilton, Aloft, Crowne Plaza) for housekeeping and butler services, interacting with customers, and fulfilling concierge and front-desk tasks. The uncanny valley theory (Mori, MacDorman, & Kageki, 2012) and recent studies (e.g. Mende et al., 2019) emphasize that anthropomorphized robots may arouse negative feelings of uneasiness and discomfort among humans. Hence, understanding consumer emotional experiences with service robots is important for both services managers and scholars. Emotions are particularly important in consumption experiences and can help managers decide, for example, the most appropriate design for a service robot. Emotions can impact consumers' intentions to adopt new technology (Venkatesh, 2000), as well as consumers' evaluation and behavior (Bagozzi, Gopinath, & Nyer, 1999; Berger & Milkman, 2012; Richins, 1997; Yin et al., 2014). Researching consumer emotions is important in marketing (Bagozzi et al., 1999) and technology-related fields (Venkatesh, 2000).

Existing studies on service robots are mainly conceptual (Ivanov & Webster, 2020; Murphy, Hofacker, & Gretzel, 2017; Robinson et al., 2020; Tung & Au, 2018; Van Doorn et al., 2017; Wirtz, Kunz, & Paluch, 2021), while the limited empirical studies have primarily focused on the

technology, anthropomorphism enablers of robots, and intention to use (Blut et al., 2021; Tussyadiah & Park, 2018). Other scholars have instead used experimental methods to understand how customers would hypothetically interact and respond to service encounters with robots (e.g. Jörling, Böhm, & Paluch, 2019; McLeay et al., 2021; Mende et al., 2019; Tussyadiah, Zach, & Wang, 2020; Van Doorn et al., 2017). Very little research has focused on the specific emotions arising from customers' experiences with service robots (Chuah & Yu, 2021). Hence, this study attempts to: *a*) identify the various robot service encounters that trigger customer emotional responses and *b*) explore the ways customers feel about their encounters with hotel service robots. To achieve these goals, we extracted a sample of 9,707 customers reviews from two leading social media platforms (Ctrip and TripAdvisor), encompassing 412 hotels in 8 countries, to explore the ways customers feel about their encounters with service robots.

LITERATURE REVIEW

Generating Insights from Online Reviews

Online social media platforms have become a key channel through which customers express their feelings and emotions about their experiences with services or products (Gaind, Syal, & Padgalwar, 2019; Ullah et al., 2016). Traditionally, scholars have adopted netnography and other qualitative methods such as content, thematic, and discourse analysis to conduct online research (Heinonen & Medberg, 2018). However, the amount of digital data produced daily has grown exponentially passing from 4.4 zettabytes in 2013 to 64 zettabytes in 2020 and it is expected to reach 180 zettabytes in 2025 (i.e., 180 trillion gigabytes) (Statista, 2022).

To identify and analysing consumer data, various research tools have been developed, including machine learning-based text mining (Ordenes et al., 2014), topic modeling (Antons &

Breidbach, 2018; Berger et al., 2020; Ye et al., 2020), sentiment analysis (Dhaoui et al., 2017), trend analysis, and social network analysis (Mirzaalian & Halpenny, 2019). Practitioners have increasingly use social media analytics tools powered by machine learning algorithms (Hayes et al., 2021).

Machine Learning Approaches

Scholars who studied emotions in customer reviews have primarily focused on the positive or negative valence of reviews (e.g., Purnawirawan et al., 2015; Sparks & Browning, 2011). Recently, research has focused on uncovering specific emotions from consumer reviews and analyzing their impact on consumer evaluations and behaviors. For example, Yin et al. (2014) focused on the effects of two emotions - anger and anxiety - on the helpfulness of sellers' reviews from Yahoo shopping websites with the aid of a machine learning application, LIWC. Ullah et al. (2016) used Neural Language Processing (NLP) to analyze the emotional content in 15,849 reviews, revealing that reviews with extreme ratings contain more emotional expressions. Gaing et al. (2019) analyzed emotions mined from tweets based on NLP and emotion classification algorithms. Filieri, Yen, and Yu (2021b) reveal a variety of emotions expressed by Instagram users through pictures and hashtags using a mixed-method approach combining visual analysis and text analytics.

From the computer science perspective, detecting emotions from customer reviews is a text classification problem. Various machine learning models, such as logistic regression, Naïve Bayes classifiers, support vector machines Random Forest, and BERT, have been applied to identify emotional features in large textual datasets (D. Lee, Hosanagar, & Nair, 2018). A basic classification technique used by these machine learning methods is logistic regression. Logistic regression is limited to binary classifications and requires linear relationships between log odds

and the input variables to make predictive outcomes (LaValley, 2008). It lacks flexibility in describing non-linear relationships and multiclass classification problems (LaValley, 2008). Naïve Bayes classifiers are one of the simplest probabilistic classification algorithms. They do not perform well when modeling features with interaction effects and complicated relationships. More advanced models such as Support Vector Machines (SVM) are flexible in solving both linear and non-linear classification problems and often yield better performance (Kübler, Colicev, & Pauwels, 2020). Random Forest builds multiple uncorrelated decision trees, where each tree represents a vote toward the final class predictions (Breiman, 2001). It can deal with complex interactions between attributes, perform feature selection automatically, and non-linear model data (Hartmann et al., 2019).

Logistic Regression, SVM, Naïve Bayes, and Random Forest are supervised learning methods where each data point needs to have a label or target. Pre-trained models are used for NLP tasks, such as classification and question answering. For instance, Google developed a pre-trained model called BERT, to produce unsupervised language representations for textual data. BERT achieves high performance compared to traditional machine learning NLP-based approaches (González-Carvajal & Garrido-Merchán, 2020). For emotion classification, the latest model available is XLNet, which outperforms BERT on several tasks, including emotion detection (Yang et al., 2019).

The current state-of-the-art model to extract latent topics from textual data is BERTopic (Grootendorst, 2020). The topics generated by BERTopic are more interpretable than traditional topic modeling methods, such as Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003). Moreover, unlike LDA, BERTopic does not require pre-setting the number of topics, which can increase the objectivity of its results.

Despite the rapid advances in machine learning, today, even the most advanced models still have a “black-box” problem, i.e. they cannot provide an explanation that is understandable by humans of why a specific emotion label is assigned to a piece of text (Bolukbasi et al., 2021). Research has shown that the accuracy of the results generated by machine learning algorithms is questionable (Hayes et al., 2021). This is problematic for scholars who wish to generate theoretical insights from online textual data (i.e., online customer reviews). To overcome these limitations, we advocate integrating traditional qualitative analysis with machine learning models.

Qualitative Analysis

Service researchers have traditionally used qualitative analysis, and specifically netnography, to analyze online data (Heinonen & Medberg, 2018). The main advantage of qualitative analysis is that it can generate a rich, accurate, and in-depth understanding of customer experiences from online data (Kozinets, 2002). At the same time, one of its weaknesses is the requirement for manual inputs from the researchers. The value of qualitative methods has been proven in various research fields and can be combined with quantitative methods (Heinonen & Medberg, 2018). For example, a researcher may triangulate qualitative findings generated from the analysis of online reviews with those uncovered from interviews and use them for developing hypotheses to be assessed using quantitative methods.

This study applies thematic analysis (Fereday & Muir-Cochrane, 2006; Spiggle, 1994) to a relatively large sample of customer reviews to supplement machine learning algorithms to better understand the range of emotions triggered by customer encounters with service robots. Thematic analysis is a type of qualitative analysis for researchers to sense the core meanings within a dataset (Fereday & Muir-Cochrane, 2006; Spiggle, 1994). The main objective of thematic analysis is to identify the themes that are relevant and important for understanding the

research phenomenon (Fereday & Muir-Cochrane, 2006). Researchers rely on personal interpretation and inference to categorize the raw data into themes. The categorization process enables the researchers to compare and contrast the narratives and subsequently, grasp the nuances of the phenomenon (Spiggle, 1994).

Customer Emotions

Emotion refers to “a mental state of readiness that arises from cognitive appraisals of events or thoughts; has a phenomenological tone; is accompanied by physiological processes; is often expressed physically (e.g., facial expressions, tone of voice); and may result in specific actions to affirm or cope with the emotion, depending on its nature and meaning for the person having it” (Bagozzi et al., 1999, p. 184).

Russell and Mehrabian (1977) developed the three-factor theory of emotions showing that emotions can be consistently categorized based on *valence* (i.e., pleasure versus displeasure), *arousal* (i.e., activation, activity), and *power* (i.e., dominance versus submissiveness). Emotions can be conceptualized in terms of *structure, dimensions, and content* (Bagozzi et al., 1999; Laros & Steenkamp, 2005; Watson & Spence, 2007). By *structure*, scholars refer to the presence of a hierarchy of emotions where specific emotions are manifestations of broader emotional states (Bagozzi et al., 1999). The *dimension* of emotions refers to the different affective dimensions of valence and level of arousal between emotions and the effects they have on consumer behavior (Watson & Spence, 2007). With regards to *content*, scholars refer to the emotions as general affective states such as positive *versus* negative, while appraisal theorists (e.g. Bagozzi et al., 1999; Lazarus & Lazarus, 1991; Smith & Lazarus, 1993) recommend that specific emotions should be studied separately and not combined, as each emotion has a distinct set of appraisals. Accordingly, Wong (2004) reveals that different emotions lead to different outcomes; for

example, enjoyment predicts loyalty, whereas happiness is a better predictor of relationship quality.

Consumer psychology scholars have long studied emotions. Izard (1977) suggests that emotions are revealed by specific patterns of facial expressions and identifies ten fundamental emotions, namely interest, enjoyment, surprise, distress (sadness), anger, disgust, contempt, fear, shame/shyness, and guilt. Plutchik's (1980) evolutionary approach recognizes eight primary emotions (fear, anger, joy, sadness, acceptance, disgust, expectancy, and surprise) paired in opposite directions, such as joy *versus* sadness. Ortony, Clore, and Collins (1990) developed a model consisting of 22 emotions. Ekman (1992) categorizes emotions into happiness, surprise, fear, anger, sadness, and disgust. Richins (1997) enlists twelve emotions: anger, discontent, worry, sadness, fear, shame, envy, loneliness, romantic love, love, peacefulness, contentment, optimism, joy, excitement, and surprise.

Service Robots

The term robot comes from the Czech word *robota*, which indicates 'forced labor' or 'slavery'. Robots can be described as mechanical devices programmed to perform specific physical tasks (Belanche et al., 2020). Service robots are used for front-desk and concierge purposes (i.e., robot 'Connie' at Hilton), for luggage transportation and room services (robot 'Botlr' at Aloft Hotel), and for cleaning and sanitizing rooms (e.g., robot 'Vi-YO-Let' at Yotel).

Scholars have found that robot anthropomorphic lifelike design provides social presence and warmth (Van Doorn et al. (2017) and can facilitate consumer intention to adopt robots (Tussyadiah & Park, 2018). The physical human appearance of a robot may affect customers' attitudes towards robots and service expectations (Belanche et al., 2020). Tung and Au (2018)

adopt a sample of 329 online reviews posted by customers of four hotels that use service robots to identify the dimensions of user experience (i.e., embodiment, emotion, human-oriented perceptions, the feeling of security, and co-experience). Yu (2020) uses the comments of YouTube users on two videos about Hennna Hotel in Japan and analyzes their reactions regarding the dimensions of anthropomorphism, perceived safety, animacy, perceived intelligence, and likeability.

Several scholars argue that empathetic intelligence makes humans irreplaceable from robots (Čaić, Odekerken-Schröder, & Mahr, 2018; Huang & Rust, 2018). For example, substituting human care givers with robots may dehumanize care, cause emotional concerns, and lead to social isolation, particularly for the elderly (Čaić et al., 2018). An apathetic, emotionless, innately cold robot does not seem like the ideal caregiver (Stahl & Coeckelbergh, 2016). Accordingly, Wirtz et al. (2018) argue that a robot's emotional expressions may not be perceived as genuine, especially in long and high involvement encounters. Drawing upon the Uncanny Valley Theory (Mori et al., 2012), Mende et al. (2019) argue that after a first initial impression of surprise, hotel guests may feel humanoid robots eerie, unsafe, and even a threat to human identity.

As shown in the literature review, there is little empirical research on customer emotions in service robot encounters using online reviews data. The only exception is the work of Chuah and Yu (2021), who used machine learning methods to measure customers' positive, negative and neutral emotions expressed through Instagram pictures towards Sophia, which is described as the first anthropomorphic robot capable of expressing various human feelings/emotions. However, Chuah and Yu's (2021) research fails to link to consumer emotions literature, focusing on a

specific type of robot in one country, impeding the generalization of customer emotions with service robots.

METHODOLOGY

The current study adopts a machine-human intelligence approach to understanding customer emotions from online reviews. We first adopted two machine learning approaches, BERTopic (Grootendorst, 2020) and the XLNet (Yang et al., 2019), to detect customer emotions in service robot encounters. This is a topic-based-emotion-detection process to explore people's emotions towards robots from online reviews. This process enables researchers to explore the general emotions expressed from unstructured textual data, such as online consumer reviews, and to explore the specific emotions towards various robot services. The machine learning models are then supplemented with qualitative analysis to verify the results and explore customer emotional experience in greater depth. To perform qualitative data analysis, we adopted the thematic analysis approach (Fereday & Muir-Cochrane, 2006; Spiggle, 1994).

The hybrid machine-human intelligence approach involves three main steps:

- a) use BERTopic to extract various customer encounters with robot services from the data.
- b) use XLNet to detect emotions from the data.
- c) conduct a critical manual analysis of the output of machine learning.
- d) conduct a thematic analysis from a sub-sample of customer reviews.

Data Collection

We collected online review data from two major customer-generated media platforms (Filieri, Algezau, McLeay, 2015), Ctrip and TripAdvisor. Ctrip attracts the largest amount of

travel customers and reviews in China. We used the keyword “robot” in Chinese to extract customer reviews from November 16th, 2019, to December 11th, 2019. A total of 7,177 reviews related to “robot” about 341 hotels in 35 cities in China. We conducted a preliminary analysis of this dataset using the labeled training data provided by Harbin Institute of Technology. Each review was labelled with one of the emotions: *neutral*, *happy*, *angry*, *sad*, *fear* and *surprise* based on 34,768 reviews extracted from Ctrip. We used the Bert-Base-Chinese Model for the training process. The overall accuracy of the training model reached 76%. The results of emotion detection show that the vast majority of customer emotions in their online reviews contained the emotion “joy” (82.86%), followed by “neutral” (9.45%), while very few emotions are “angry” (3.61%) and “surprise” (3.57%), whereas sad and fear are rare, accounting for only 0.28% and 0.24% of the total reviews, respectively.

TripAdvisor is the world's largest traveler community, and it has been used in many studies on customer experience with services (Fileri, Galati, & Raguseo, 2021a). To extract TripAdvisor data, we first set the keyword "robot" to search for hotels with reviews related to robot services. We used a Python scraper to automatically collect the URLs of hotels with at least ten reviews that mentioned the word “robot”. Next, we removed the online customer reviews that used the word “robot”, but they were not referring to service robots; for example, “staff does their thing in routine, like robots” and “Receptionist just saying we are not robots”. We then scraped the reviews for the selected hotels, and we collected 2,530 service robot-related customer reviews from TripAdvisor between April 2017 to December 2020 of 71 international hotels from 7 countries.

We conducted a preliminary analysis of the TripAdvisor dataset using BERT. The Bert-Base-Uncased Model was used to convert texts to numbers. We then trained the emotion

detection model using the dataset developed by Saravia et al. (2018), who labeled social media posts using six emotions: *anger*, *fear*, *joy*, *love*, *sadness* and *surprise*, based on 20,000 English posts extracted from Twitter. Similar to findings obtained from Ctrip reviews, the results of emotion detection from TripAdvisor reviews show that most emotions expressed were joy (85.85%), with very few reviews discussing negative emotions such as anger (4.90%), fear (3.16%), and sadness (1.74%). Given the similarity in the results, we decided to combine the two datasets for further analysis. Considering the language differences in the two datasets, which would lead to the use of different training sets and emotion categories, we decided to translate the Ctrip reviews into English with the aid of machine translation (Baidu Translate). The translation was proofread by another author to ensure accuracy.

Data Pre-processing

We first extracted sentences that are relevant to robots. It is important to note that each review often contains multiple service touchpoints and ambivalent evaluations. For example, the following review: “I love the robot but the price is too high. Friendly staff. Excellent atmosphere” includes information about robots, price, staff, and atmosphere. However, the goal of this study is to detect customer emotions referring to service robots’ encounters. Hence, we used the Python library, spaCy, and one of its pipelines, “clausie”, to extract the clauses within each review. SpaCy can accomplish multiple natural language processing tasks, such as tokenization, tagger, parser, entity recognition, lemmatization. Going back to the previous example (i.e., “I love the robot, but the price is too high...”), spaCy would remove all irrelevant comments, leaving only what is relevant for our analysis (i.e., “I love the robot”).

Extracting Latent Topics

We used BERTopic for identifying customers' emotional responses triggered by encounters with service robots. To achieve the optimal results, we first fine-tuned the hyperparameters of the model. We removed topics with less than 1% of the total dataset, that is topics containing less than 74 reviews would be ignored by the algorithm (Ryoo, Wang, & Lu, 2021). We also used the CountVector module under the Scikit-Learn library to convert the reviews into a matrix of token counts with a maximum of 2-grams and remove common English stopwords (Pedregosa et al., 2011). For the embedding model, we used the "all_mpnet_base_v2" sentence-embedding model for both accuracy and efficiency (Grootendorst, 2020).

The outputs of the BERTopic algorithm were the topics with the five most representative keywords. The meaning of each topic requires human interpretation; therefore, two of the co-authors interpreted and named each topic based on their keywords and exemplar reviews.

Detecting Emotions

We used XLNet to detect customer emotions from online reviews. We adopted a pre-trained English XLNet, called "xlnet-base-cased", to train the multiclass emotion detection model. The architecture of this model has 12 layers, 768 hidden neurons, 12 heads, and 110 million parameters in total. The training and test dataset was developed by Garbas (2019), containing texts with emotion classifications into a balanced dataset, including five emotions, namely *neutral, joy, sadness, anger, and fear*. The whole dataset was split into training (7,934 reviews) and test (3,393 reviews) datasets. We trained the XLNet model using the training dataset and evaluated its performance on the test dataset. The evaluation of the model on the test dataset reports that our model performs well. As shown in Table 1, the overall accuracy of the model is as high as 82% and the accuracy of detecting each of the five emotions is from 78% to 86%.

[Table 1 about here]

Benchmark Models for Emotion Detection

We then further validate the proposed emotion classification model by comparing its performance with other benchmark models using the Mathews Correlation Coefficient (MCC) metric (Boughorbel et al., 2017; Matthews, 1975), which is calculated as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP, TN, FP and FN refer to the number of true positive, true negative, false positive, and false negative predictions. The higher the MCC metric, the better the model performs. The MCC can achieve a high score only if the model performs well in all the four confusion matrix categories (TP, TN, FP, and FN). Therefore, MCC is more reliable than other metrics in multiclass classification, such as accuracy and precision.

We selected five benchmark models: (1) BERT; (2) Support Vector Machine (Linear); (3) Logistic Regression; (4) Naïve Bayes (Multinomial); and (5) Random Forest. Table 2 presents the MCC performance for our proposed model and the benchmark models. The results show that XLNet outperforms other benchmark models, including the state-of-the-art BERT model. Therefore, we are confident to use this trained model to classify the reviews data.

[Table 2 about here]

MACHINE LEARNING RESULTS

Customer Encounters with Robot Services

We applied the BERTopic algorithm to identify customer encounters with various service robots discussed in customer reviews. The BERTopic initially identified 19 topics and showed

the top-5 words within each topic, as shown in Figure 1. One of the researchers of this study provided a label for each topic, and another author reviewed and confirmed it, following the approach suggested by Guo, Barnes, and Jia (2017) and Filieri et al. (2021b)

The naming of the topic was based not only on the logical connection between the keywords but also on the comprehension of three example reviews for each topic, followed by the integration of similar topics. For example, topic number 15 was named “cooking service”, based on the top-five keywords ‘egg’, ‘breakfast’, ‘eggs’, ‘omelet’ and ‘making’. Furthermore, the three example reviews were “but don't like the egg robot”, “Good location and the robot that makes eggs in the morning is fun to watch”, and “Breakfast had a good spread including a robot egg chef”. Finally, 13 unique topics were identified (see Figure 2). The majority of customer reviews discussed the robot service in general without mentioning any specific service. For example, “this hotel has a robot”. “The robot service amazed me”. In addition, reviewers discuss how their children react or feel toward service robots.

Most of the other topics are customer reactions to their encounters with the specific services delivered by robots. A few examples are listed next. Delivery service: “I was able to telework from my room and service was delivered by a robot, awesome!!!”, and “There's a robot to send things. It's very cute!”. Room service: “the room service robot is the coolest, you definitely have to order something to see him in action”, and “The robots that do room service (water and extra towels) are so cute and entertaining”. Way-leading service: “it is very interesting for robots to lead the way!”, and “Robot leads the way, stupid robot!”. Lobby Service: “There are robots standing in the lobby, which are very cute”, “There is a robot in the lobby, named Xiao Lang, who is warm and hospitable. My child had a great fun with the robot”. Cooking service: “They had this Super fun robot omelet machine maker, which was pretty cool”, and “impressed by the

egg chef robot to cook omelets and sunny side up.” Question answering: “I especially like the small robot. It answers all questions”, and “There is a very cute intelligent robot as an advisor, which can answer intelligently with the guests, which is very popular with children”.

[Figures 1 & 2 about here]

We then manually classified these topics into three main categories about robot service, namely *general discussion*, *specific services*, and *feelings*. Figures 3 and 4 show the three categories with specific topics and their distribution within the dataset. The results show that reviews mostly discuss the specific services delivered by robots, such as delivery service and cooking service. The rest involves general discussion and feelings. General service involves a general discussion on the robot services in a particular hotel. For example, “The hotel has robot service”. Another important topic is the expression of emotions. This category, as shown in Figure 3, mainly includes children's reactions to service robots, such as excitement and love. For example, “I liked it a lot, and my daughter went crazy about the robots”.

We compared the topics generated from the BERTopic with those derived from authors' interpretation (see Table 3) to validate the results. The table shows that the same ten topics were identified by both the machine and human analysis, while two topics (water delivery and general discussion) were missed in human analysis and another two topics (towel delivery and toiletries delivery) were missed by BERTopic. We measured the overlap of the topics extracted through these two approaches using the Jaccard coefficient (Guo et al., 2017; Tirunillai & Tellis, 2014). If $N(Dim_{bertopic})$ represents the set of topics derived from Ctrip and TripAdvisor reviews using the BERTopic algorithm and $N(Dim_{human})$ represents the set of topics summarized by human researchers, we calculate the Jaccard coefficient as:

$$JC = \frac{|N(Dim_{bertopic} \cap Dim_{human})|}{|N(Dim_{bertopic} \cup Dim_{human})|}$$

A higher Jaccard coefficient means the degree of overlap between the two sets of topics is higher. Jaccard coefficient for BERTopic is 0.71, implying that the BERTopic can generate similar topics by human interpretation while reducing the large amount of time that is needed for manual analysis. Therefore, we argue that the topics generated by the BERTopic algorithm are reliable and valid (Guo et al., 2017).

[Table 3 about here]

Emotions Detected

The emotions are detected from the three main categories: *specific services, general discussion, and feelings expressed*. The results show that the distribution of emotion across all three categories is similar. "Joy" and "Fear" are the most frequently expressed emotions. "Fear", however, appears more often in reviews discussing specific services than the other two topics. For example, "The robot sent water and scared me".

Table 4 shows the distribution of each emotion detected from the overall data. The most frequently expressed emotion, "Joy", accounts for over 60% of the total reviews. "Fear" is the second most expressed. Nearly 30% of people did not expect some of the services are provided by robots instead of human staff. All the other three emotions altogether, including "Anger", "Sadness" and "Neutral", account for around 10% of the total reviews. The XLNet model generates the category of neutral, which involves 4.38% of the reviews in our dataset. However, "Neutral" indicates lacks emotions and therefore it will not be discussed in the following section, where we only illustrate the results of the emotion categories generated by the machine learning.

[Table 4 about here]

Joy

Based on emotions detected by XLNet of the online reviews, customers mainly expressed a feeling of joy when interacting with service robots (61.02%). Joy indicates a state of happiness and felicity, a great pleasure and delight. Customers, and particularly their kids, enjoyed interacting with robots: *and the star of the show were the robots Cleo and Leo. Our kids just loved this experience and requested more than they needed from the bots. How fun. Hotels usually aren't fun (T2292).*

However, our manual analysis of the reviews grouped under the dimension of Joy by XLnet enabled us to identify some inconsistencies. For example, in many reviews, customers made considerations about the robot's service attributes or appearance, using very often adjectives like cute and cool: *When we got to our room we had the most adorable little robot (Mr. Wes,) come to our room and bring us water, snacks, and a welcome letter. It is by far the cutest and coolest thing ever (T878).* In other reviews, customers discussed that the robot was the *highlight* of their stay, without revealing any emotional reaction: *the highlight of our stay was the robotic services where the robots are at your room service, saving bottles of water and also making eggs at the breakfast buffet (T42).*

The 'remarkable' nature of the experience with service robots is relevant but conceptually different from joy, which is a feeling of great happiness. However, not all the customer reviews grouped under this topic by XLNet indicated joy. For instance, some customers disclose a feeling of surprise (amazement) in the encounter with service robots. Amazement is a feeling of great surprise or wonder. Although the feelings of amazement and joy are both positive, and

amazement can lead to joyful emotions, they have different meanings: *Also they have the most amazing robotic omelet-making machine!! It makes the best omelets...*(T428). Some customer reviews express feelings of love, which, although potentially leading to or being caused by joy, indicate a different type of feeling. In this case, the feeling of love towards robots makes customers forgive the robot's service failure: *We loved the room robot, although it did bring us shaving cream instead of water!!* (T1779).

Some customers stated that the hotel they have chosen was their favorite because of the adoption of service robots. Favorite indicate consumers' preference for a person or a thing over all others of the same kind. It is not necessarily an emotional category and does not indicate joy. *My favorite thing about this hotel is the robot that delivers small items for guests, unique and lovely* (T1469). Customers labeled service robots by using terms like novel, modern, futuristic, creative, high-tech, and unique. However, these adjectives refer to the robot concept and not to customer emotions in robots' encounters: *It is unusual for us to ask for something from the front desk, but we couldn't resist the novelty of having it delivered by a robot. So we requested extra body wash, which smelled great!* (T2446).

In many reviews, customers appreciate the helpfulness of service robots. However, the concept of helpfulness refers to the utilitarian nature of robots that contrasts with the hedonic nature of joy: *Staff, including the amazing robot room service, are very helpful and provide good service* (T599).

In summary, the manual analysis critically shed light on some of the inconsistencies between the concept of joy and the content of online reviews attached to this concept by XLNet.

Fear

Table 4 shows 28.52% of reviewers expressed feelings of fear toward robots, which was the second-largest emotion category discovered by XLNet. Fear feelings focus on the interaction with robots, such as *Frightened by the robot a little (T2249). The staff were lovely and could not be more helpful - beware of the robot who frightened the life out of me when it got into the lift with me!!! (T636)*. Customers also found service robots creepy *though I found the whole robot room service thing to be kind of creepy (T2290)*.

The manual analysis, however, reveals that most of the customer reviews grouped under this topic did not focus on emotions of fear, but rather on discontent. For instance, customers voiced their disappointment because some robots were out of order or because the claims of the robot hotel did not meet their experience: *We were very excited about the robot hotel but found it underwhelming. There are two robots at the check-in and that is it. When you first get there, you are like 'woah check them out that's freaky' but you can't actually interact with them until 3pm when check in is, you can't get near them before that...(T1192). It was advertised as a dinosaur robot hotel.... There was none of that. There were two dinosaurs for check in and nowhere else.... We chose this hotel for what was advertised and were incredibly disappointed (T1115)*.

Anger

Some customers (5.16%) express anger about service robots. However, the manual analysis reveals that consumers are disappointed or frustrated rather than angry, and this feeling refers to the service delivered by robots, which they sometimes find annoying and slow. Marketing research provides evidence that feelings of discontent resulting from service failures are distinct from more specific negative emotions, such as anger or sadness (Bougie, Pieters, & Zeelenberg, 2003; Romani, Grappi, & Dalli, 2012). *The robots are a little pesky, but there is no need to engage with them, except when they create a traffic jam at the elevators (T768)*.

Some customers were not happy about the ‘unreasonable’ charges for the use of robots. *How can you with a straight face say that I need to pay for the services of this R2D2-looking thing, whether or not I use him to get me a granola bar that I can't walk to the lobby for?* (T819).

Sadness

A small number of reviewers express sad feelings (i.e., 0.91%), as shown in Table 4. The malfunctioning or the lack of efficiency of service robots created feelings of sadness, particularly among kids. *Disappointed that the robot-powered bag storage was full when I wanted to use it* (T178). Parents were particularly disappointed, rather than sad, because their kids could not play with the service robots. *Sadly, the Robot dog, which my kids were looking forward to interacting with, was not charged for our stay* (T549).

The fact that robots were used to replace human workers generated some sad feelings, too. Hotel customers were concerned about humans being replaced by robots, feeling sad about the employment prospect of the hotel staff. *...and a robot that delivers non-food room-service requests, for example, a toothbrush. This did not feel uncomfortable, but it made me sad to think about how computers can substitute for people* (T1332).

RESULTS FROM THEMATIC ANALYSIS

The preceding manual analysis of machine learning results revealed the limitations of the available training dataset for emotion detection, which could only identify five emotions (i.e., joy, fear, anger, sadness, and neutral) and forced the categorization of some conceptually different emotions into the five types of emotions mentioned above. Therefore, we decided to integrate a thematic analysis to provide an in-depth understanding of the specific types of emotions arising from consumer reviews and to clarify the inconsistencies that emerged in the previous analysis.

Thematic analysis is a flexible qualitative data analysis method that aims to search, analyze and report repeated patterns of meaning (themes) across a dataset (Fereday & Muir-Cochrane, 2006; Spiggle, 1994). We used systematic random sampling to select the reviews from our dataset for manual analysis, aiming for 50% of the sample (i.e., 4,853 reviews). The researchers read the reviews to familiarize themselves with the data and generate initial codes. All robot-related review excerpts were underlined and later copied into a word document consisting of 451 pages and 265.790 words. Subsequently, the excerpts from the reviews related to robot experiences were transferred into the coding document and organized predominantly using a theory-driven approach (Fereday & Muir-Cochrane, 2006), which was based on the emotional categories available from the literature (e.g., Richins, 1997). This approach is appropriate for specific research questions (Fereday & Muir-Cochrane, 2006; Spiggle, 1994), like in this study that aims to shed light on the properties and dimensions of consumer experience with service robots. The semantic or explicit level of themes identification was preferred (over the latent level) because our goal was not to look for anything beyond what the hotel customers have written in their reviews (Boyatzis, 1998).

Data analysis proceeded with the identification, review, and definition of themes. The theoretical saturation was achieved after reading the first 1,000 reviews, however, we proceeded until we reached the target sample. The analysis enabled us to identify four macro-dimensions of customer experience with service robots, which summarized the topics discussed in the online customer reviews, such as *emotions with robots' encounters*, *robot attributes*, *services delivered by robots*, and *robot service evaluation*. The *first dimension* included the themes related to the *emotions* triggered by robots' encounters (sorted by frequency of mentions): *joy*, *love*, *surprise*, *interest*, *excitement*, and *discontent*. In the thematic analysis the 'keyness' or prevalence of a

theme does not depend on how many times it appears in a data corpus (Braun & Clarke, 2006), however considering we are dealing with a large dataset, and we want to compare these results with the results of the machine learning analysis, we have decided to quantify the themes pertaining to emotions. Hence, Table 5 presents the ranking of emotions based on the number of times they appear in the dataset compared to the machine learning output. The *second dimension* included adjectives related to *robot attributes* such as modern, cool, intelligent, cute, lovely, friendly, quirky, weird, popular, and playful. The *third dimension* includes the *service delivered by robots* such as room service, omelet maker, and reception. Finally, reviewers commented on the *service quality of robots*, which was evaluated as convenient, amazing, unique, fast, distinctive, and slow.

Joy

The feeling of joy is the most frequently mentioned feeling, confirming the results of the machine learning analysis. Joy is often expressed through emotional words like ‘like’, ‘happy’, ‘enjoy’, ‘amusing’, ‘fun’, which very often refer to the experience of children with robots.

A great touch that my boyfriend and I enjoyed a lot was the robot who brought up your room service! [ID1563]

Love

Love is a strong feeling of affection and emotional attachment which can be felt towards people, activities, objects, and places (Carroll & Ahuvia, 2006). The emotion of love is often triggered while the robot is performing his tasks. Once again, this emotion is often described to be associated with the reaction of the children of the reviewer.

The robot was an absolute surprise for us, as we'd never seen anything like that! My girls loved it! [ID42]

Surprise

In general, surprise can have a positive, neutral and negative valence. In our analysis, surprise has mainly a positive connotation, being associated with a feeling of awe. Surprise is often expressed through exclamation marks, which convey the feeling of great surprise.

Awesome and amazing are the emotional words more often used to express surprise. *I had a room on 8th floor facing pool/back side of hotel yes they have a robot butler how awesome!!!!*

[ID 484] *I tried the Robot room service! WAW! Amazing to be able to experience such technology. [ID1587] the visit by the robot Elina was a fun surprise for the kids [1712]*

Interest

Feeling interested is “the feeling of being engaged, caught up, fascinated, or curious. There is a feeling of wanting to investigate, become involved, or expand the self by incorporating new information and having new experiences...” (Izard, 1991, p. 100). Interest is expressed with words like interesting, fascinated, and curious. The feeling of interest refers to individuals’ willingness to know more about how the robots deliver their specific service.

There are robots leading customers to the room. It's very interesting and the children are very happy [ID 2982] Our kids were fascinated by the robot [ID 17]

Excitement

Excitement is a temporary euphoric state involving feelings of thrill, arousal and enthusiasm. The findings highlight that customers’ children are particularly excited to see and play with robots.

*There is also a high-tech Xiaoman robot. The children are very excited to see it [ID 4191]
The reception by the GM (Julian) of the hotel was a lovely touch and the fun robot that delivers
things to your room had the kids intrigued and excited. [ID 627] My 6yo was super excited to see
Jeno the robot wondering about. [ID 667]*

Discontent

Discontent represents the feeling of dissatisfaction with one's possessions, status, or situation. Although discontent is a negatively valenced emotion, in our study, it did not have a negative connotation. Interestingly, this feeling was mostly expressed when service robots were out of service, or to indicate the missed opportunity to see them in action or play with them (often associated to customers' children).

*My 6yo was super excited to see Jeno the robot wondering about. He was terribly
disappointed when our room service was delivered by a human being instead of Jeno [ID 667]
We also called down a couple times to have extra washcloths brought to our room and
specifically asked that the robot bring it up. I guess he wasnt working because he never showed
up. :([ID1522]*

Fear, Anger, and Sadness

We came across very few reviews expressing feelings of fear, anger, and sadness. These negative emotions were expressed in less than 1% of the consumer reviews analyzed. Furthermore, consumers weight more the benefits of robots (i.e., privacy, quiet) compared to the negative emotions of fear that they can provoke, suggesting their willingness to accept them. *I think the robots at the concierge will make me feel a little bit creepy but I think they are so useful when I want to stay quiet because exhausted from long day trip [ID1467]*

Finally, some customers observed that the service delivered by humans could be, at times, more awkward compared to the one offered by robots. *Room service always seems a little awkward, but with the robot delivering it makes it less awkward for sure* [ID1461]

Movement versus not movement

We noticed that the robots' movement modality (*versus* non-movement) was an important factor in triggering (or not) emotional experiences. Specifically, we noticed that consumers were more likely to express (positive) emotions when the robot was in movement compared to when it was standing still.

Cultural Differences

Our thematic analysis reveals several interesting differences between the Chinese customers (Ctrip dataset) and the English-speaking, mainly Western customers (TripAdvisor dataset). First, for Chinese customers, the emotion of fear is more oriented toward robots' appearance and behaviors, while western customers focus on the eye contact. For example, one TripAdvisor customer wrote, *Also there's a robot concierge. Truthfully, Pepper kinda freaked me out so I didn't look her in the eyes* (ID539). For Western customers, the appearance of service robots reminded them of the movie character of Dalek, a merciless alien in Doctor Who: *Stepping into the hotel, we were greeted by smiling staffs (and that cute robot who unfortunately reminds me of a Dalek)* (ID1048).

Second, Western customers tend to be concerned about the idea that robots could replace human employees, for example, *it made me sad to think how computers can substitute people.* (ID1332). We did not find similar comments in the Chinese dataset.

Third, some Chinese customers expressed discontent when the service robot was not available; in contrast, some Western customers disliked being served by machines, due to their impersonal nature and lack of flexibility.

I was told a robot could deliver for me. (Seriously? I like people. Real people.) (ID718)

Technology was cool e.g. robots but felt a little impersonal/functional and frustrating at times. (ID1052)

DISCUSSION AND CONCLUSION

This study utilizes a hybrid machine-human intelligence approach that integrates machine learning, critical manual analysis, and qualitative analysis (thematic analysis) to understand customer emotional experience with hotel service robots. The study has important methodological, theoretical, and managerial implications.

Methodological Implications

Our hybrid machine-human intelligence approach offers both the strengths of machine tools in performing emotion categorization of a large dataset; and of thematic analysis in offering a human interpretation of machine learning outputs and in providing an in-depth understanding of the emotions and feelings expressed by customers who have experienced service robot encounters. The study adopts customer reviews published on two large social media platforms, providing an accurate analysis of the actual customer experience with service robots. Previous studies on consumer emotions adopted single methods using experiments, literature review, and factor analysis focusing on consumer's intentions or perceptions (e.g., Akdim, Belanche, & Flavián, 2021; Jörling et al., 2019; McLeay et al., 2021; Mende et al., 2019; Van Doorn et al.,

2017), and not on actual experiences. Our study suggests the suitability of a hybrid method to analyze customers' emotional experiences with service robots.

This study demonstrates that the proposed hybrid machine-human intelligence approach overcomes the inadequacies of the most advanced machine learning methods, which categorize various emotions under a limited number of macro-dimensions that do not necessarily represent the same content of the emotion/feeling or emotional states in the customer review (i.e., helpfulness is not an emotion, discontent has a positive connotation). Although some of these emotions are positively associated (i.e., love can lead to joy), they represent different emotional states with different arousal/intensity (i.e., love is a stronger feeling than joy). Our results show that although machine learning techniques become more advanced, they cannot yet match the human capacity to interpret consumer emotional experiences. This study sheds light on how researchers can integrate machine learning algorithms with human-based qualitative analysis. The study shows that the hybrid method is particularly useful to handle what machine learning cannot do in dealing with complex topics that require a more nuanced, in-depth analysis, such as consumer emotions. Our results indicate that the integration of traditional thematic analysis based on a smaller but statistically representative sub-sample of reviews, although more time-consuming, can obtain a more accurate and thorough picture of the phenomenon investigated.

The thematic analysis in our hybrid approach highlights a more nuanced view of emotions, confirming the prominence of enjoyment as the primary emotion in customer-robot encounters, but revealing a richer variety of emotions such as *love*, *surprise*, *interest*, *excitement*, and *discontent*. The findings reveal that customer reviews discussing robots mainly refer to *emotions in customer-robot encounters*, *services delivered by robots*, *robot attributes* (e.g., cool, cute), and *robot service evaluation* (e.g., highlight, convenient). Thematic analysis corrects the inaccuracy

of the findings of machine learning outputs, overcoming the weakness of these techniques in providing an accurate analysis of customers' emotions and the limitation of the available training dataset that only identifies limited types of emotions (i.e., joy, fear, anger, sadness and neutral). The results of this study also confirm the poor accuracy of NLP for negative sentiments analysis (Dhaoui et al., 2017).

Theoretical Implications

Consumers' emotional reactions to service offerings have important effects on consumer behavior (Bagozzi et al., 1999). Our study focuses on consumer emotions triggered by the interactions with service robots implemented in the hospitality sector and shared through online textual reviews. Emotional wording is often used in digital conversations (Berger & Milkman, 2012) and when expressed through online consumer reviews can have an impact on consumers' attitudes, evaluation, and behavior (Berger & Milkman, 2012; Yin et al., 2014).

This study responds to a call for additional research on emotional experience with AI-enabled technologies (Mende et al., 2019). This is one of the first studies that adopt a relatively large dataset of customer's actual experiences with different types of service robots, thus hotels in our sample employed humanoid robots with anthropomorphic features (e.g., room delivery robots) and non-humanoid robots (i.e., mechanical arms).

By drawing upon the literature on consumer emotions (Bagozzi et al., 1999; Laros & Steenkamp, 2005), we developed an innovative mixed-method approach using advanced machine learning techniques (BERTopic and XLNET) and thematic analysis to advance our understanding of customer emotions in service robot encounters. This study complements previous studies on robots dealing with the utilitarian and cognitive determinants of employees'

and consumers' intention to adopt robots (e.g. Blut et al., 2021; de Kervenoael et al., 2020; Turja et al., 2020). This study affirms the prominence of positive emotions in the customer experiences with service robots as shared in online reviews by extending the work of scholars investigating consumer emotions in technology consumption contexts (Kulviwat et al., 2007; S. Lee, Ha, & Widdows, 2011), or focusing on emotions as important enablers of technology acceptance and use (Kulviwat et al., 2007).

The findings of our machine tools highlight that human-robot interactions provoke emotions like joy, anger, fear, and sadness. This result aligns with previous findings using machine learning and measuring the feelings towards a female humanoid (Sophia) (Chuah & Yu, 2021).

The findings of our thematic analysis reveal a more nuanced picture of the customer emotional experience phenomenon, revealing that customer reviews discuss the following themes: *emotions in robots encounters*, *services delivered by robots* (e.g., omelet making robot), *robot attributes* (e.g., cool, cute), and *robot service evaluation* (e.g., highlight, convenient).

Overall, the feeling of joy consistently is the most frequent emotion in customer encounters with service robots. This result aligns with the findings of a study on customer experience with robots in the hospitality sector (Tung & Au, 2018), and with a recent survey study on healthcare professionals, which show that dispositional attitude and perceived enjoyment were the most significant predictors of humanoid robots' use intention (Turja et al., 2020).

The qualitative work of our hybrid approach reveals a richer variety of positive emotions such as *love*, *surprise*, *interest*, *excitement*, and *discontent*. Interestingly, the analysis identified a negligible number of negative emotions (less than 1%), and a negative emotion like discontent was often associated with the unavailability of service robots due to malfunctioning or being

under repair. Thus, our analysis reveals an *overwhelmingly positive* evaluation of customer experience with service robots from an emotional perspective, which contrasts with Yu's (2020)'s analysis of YouTube video comments about robots showing that people have anticipated negative emotions towards robots. We also could not find the expression of feelings of insecurity when interacting with service robots in our analysis, which does enable us to confirm previous findings based on the analysis of a smaller number of reviews (Tung & Au, 2018). In general, this result contrasts with most of previous experimental studies that reveal that consumers might have a negative emotional reaction to service robots (e.g., Akdim et al., 2021; Mende et al., 2019).

Furthermore, the analysis shows that consumer show interest and curiosity towards service robots, supporting previous studies (Akdim et al., 2021). Some scholars excluded interest from their categorization of emotions because is a non-valenced descriptor of cognition (Ortony et al., 1990). However, in this study, interest was the fourth most mentioned emotions in service robots' encounters.

The thematic analysis also indicates that the sociodemographic characteristics of customers matter in the emotional experiences with service robots. For example, the civil status of the customers (i.e., family with kids) and age could be considered as a moderator on customer emotional reactions to service robots; that is young consumers (children) predominantly experience positive emotions such as joy, love, and excitement in service robot encounters. Scholars have revealed that robots can be an effective therapy for children with autism because they are less intimidating, less complex, and emotional than humans (Cabibihan et al., 2013). The result is partly in line with Kanda et al.'s (2004) experimental findings showing that children are willing to interact with robots, although they rapidly lose interest (after one week). However,

children may also like robots more than adults because of their playful and fun appearance that reassembles to robots' anime they watch on television programs.

Our study further reveals the potential moderation effect of culture on customer emotional reactions to service robots. Specifically, we found that compared to Chinese customers, Western customers tended to *a*) focus more on the facial expression of robots, especially the robot's eyes, *b*) feel sad about human jobs being replaced by robots, and *c*) prefer to be served by humans, because machines are perceived as impersonal and inflexible. Culture has received little attention in the recent service robot literature. Previous studies suggest that culture can influence attitudes toward robots (Bartneck et al., 2007), with initial evidence suggesting Asian customers have a higher level of robot acceptance (MacDorman, Vasudevan, & Ho, 2009; Rau, Li, & Li, 2009) and people from low-context cultures (e.g., Germany) are less engaged with robots when the sociability of a task is low (D. Li, Rau, & Li, 2010). Our findings tend to corroborate these findings while providing additional new insights into the cross-cultural implications of service robots deployment.

Furthermore, we reveal that robots trigger emotions when they are in movement, that is, while delivering their service and that emotions are mainly expressed towards humanoid robots, while we observed a neutral or cognitive reaction to non-humanoid types (i.e., mechanical arms).

Managerial Implications

Our results show that service robots can be a source of differentiation for the hotels that introduce them. Customers and robots co-create novel and unique experiences when interacting together (Tung & Au, 2018). Robots create positive emotional reactions such as astonishment, curiosity, amazement, and surprise by creating unforgettable customer experiences, especially

among family with children. Hence, the results of this study provide strong support for the adoption of robot in hotels that target families with kids.

However, negative experiences with robot services are also detected. It is also important to learn from the negative customer experience that the hotels should focus on improving service quality, increasing functionality, and human-robot interactions of robots as future development plans for the hospitality discipline. Moreover, hotels should have cultural sensitivity when designing and deploying service robots. For example, avoid designing a robot that resembles an object that arouses unpleasant feelings for a certain cultural group. Hotels in Western countries should take measures to reassure that there is no job loss due to the adoption of robots.

We recommend hotels that are offering only a very limited number of services through robots to not promote themselves as robot hotels because this can create high expectations in the customers. Customers expect that robot hotels are solely or mainly equipped with robots. However, if they find a gap between their expectation and their experience, they will voice negative feelings of disappointment, frustration, and discontent in their reviews. Service robots can be a double-edged sword. Hotels that advertise themselves as service robots create high expectations about the robotization of services delivered in that hotel. Therefore, one recommendation is that a robot hotel can create a real differentiation if the service is entirely delivered by robots. For instance, a hotel such as Yotel, which bases its competitive advantage on its brand image as a robot hotel, should be more careful about the quality of the service experience with robots compared to a hotel like the Ritz Carlton hotel, which relies on different points of differentiation.

The findings show that service robots sometimes malfunction, or they are not capable of responding to unexpected customer requests, which can create frustration in the most exigent

customers, such as business customers, who are less keen to tolerate inefficient services.

Therefore, we believe that hotels with a clientele composed of business customers should reduce the use of entertaining or funny robots.

Limitations and Future Research Directions

This study has several limitations, and further research is recommended. First, our machine learning models use the pre-labeled training dataset (Garbas, 2019). As a result, we could not detect emotions beyond the five labels set in the training dataset. Using researchers' own training data can ensure the emotions detected can match the scholars' research attention. Therefore, before adopting our hybrid approach, future researchers may start with manual analysis based on a proportion of the datasets to categorize various emotions and label the training data accordingly for the machine learning models, i.e. the manual-machine learning-manual combination.

Second, future studies may combine machine learning with traditional statistical methods to predict customer ratings or hotel sales performance based on customer emotional experiences detected from online reviews, and test whether national culture performs a significant moderation role.

Third, the empirical context of service robots in our study is still at the early stage of application. When the novelty of service robots gradually wears off, customers' knowledge and expectations of robots might change. Future research could investigate whether customer emotions change as service robots become commonplace. Studies may also compare the differences in customer emotions for new and repeated customers of robot services.

Fourth, the study focused on physical service robots adopted in hotels. Future research could assess whether the emotions found in this study can be found for various robot types.

Researchers could focus on different service robot types (i.e., virtual AI-robots) adopted in different service contexts (e.g., healthcare, e-commerce). Finally, future studies may need to explore the impact of service failure and recovery on consumer experience (Belanche et al., 2020). Robot services, like human services, are subject to service failures that may lead to experiencing co-destruction (Čaić et al., 2018).

REFERENCES

- Akdim, K., Belanche, D., & Flavián, M. (2021). Attitudes toward service robots: analyses of explicit and implicit attitudes based on anthropomorphism and construal level theory. *International Journal of Contemporary Hospitality Management* (In Press), <https://doi.org/10.1108/IJCHM-1112-2020-1406>.
- Antons, D., & Breidbach, C. F. (2018). Big data, big insights? Advancing service innovation and design with machine learning. *Journal of Service Research*, 21(1), 17-39.
- Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, 27(2), 184-206.
- Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *Ai & Society*, 21(1), 217-230.
- Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2020). Service robot implementation: a theoretical framework and research agenda. *The Service Industries Journal*, 40(3-4), 203-225.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of marketing*, 84(1), 1-25.
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192-205.

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine learning research*, 3(Jan), 993-1022.
- Blut, M., Wang, C., Wunderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, 49, 632–658.
- Bolukbasi, T., Pearce, A., Yuan, A., Coenen, A., Reif, E., Viégas, F., & Wattenberg, M. (2021). An Interpretability Illusion for BERT. *arXiv preprint arXiv:2104.07143*.
- Bougie, R., Pieters, R., & Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science*, 31(4), 377-393.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*: sage.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Cabibihan, J.-J., Javed, H., Ang, M., & Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics*, 5(4), 593-618.
- Čaić, M., Odekerken-Schröder, G., & Mahr, D. (2018). Service robots: value co-creation and co-destruction in elderly care networks. *Journal of Service Management*, 29(2), 178-205.
- Carroll, B. A., & Ahuvia, A. C. (2006). Some antecedents and outcomes of brand love. *Marketing letters*, 17(2), 79-89.

- Chuah, S. H.-W., & Yu, J. (2021). The future of service: The power of emotion in human-robot interaction. *Journal of Retailing and Consumer Services*, *61*, 102551.
- D'Acunto, D., Volo, S., & Filieri, R. (2021). "Most Americans like their privacy." Exploring privacy concerns through US guests' reviews. *International Journal of Contemporary Hospitality Management*, *33*(8), 2773-2798.
- de Kervenoael, R., Hasan, R., Schwob, A., & Goh, E. (2020). Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors' intentions to use social robots. *Tourism management*, *78*, 104042.
- Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904.
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, *5*(1), 80-92.
- Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research*, *68*(6), 1261-1270.
- Filieri, R., Galati, F., & Raguseo, E. (2021a). The impact of service attributes and category on eWOM helpfulness: an investigation of extremely negative and positive ratings using latent semantic analytics and regression analysis. *Computers in Human Behavior*, *114*, 106527.

- Filieri, R., Yen, D. A., & Yu, Q. (2021b). # ILoveLondon: An exploration of the declaration of love towards a destination on Instagram. *Tourism Management*, 85, 104291.
- Gaind, B., Syal, V., & Padgalwar, S. (2019). Emotion detection and analysis on social media. *arXiv preprint arXiv:1901.08458*.
- Garbas, L. (2019). *Emotion Classification in Short Messages*. Retrieved from: <https://github.com/lukasgarbas/nlp-text-emotion>
- González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT against traditional machine learning text classification. *arXiv preprint arXiv:2005.13012*.
- Grootendorst, M. (2020). BERTopic: leveraging BERT and c-TF-IDF to create easily interpretable topics (Version v0.9.2): Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.4381785>
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467-483.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20-38.
- Hayes, J. L., Britt, B. C., Evans, W., Rush, S. W., Towery, N. A., & Adamson, A. C. (2021). Can social media listening platforms' artificial intelligence be trusted? Examining the accuracy of Crimson Hexagon's (now Brandwatch Consumer Research's) AI-Driven analyses. *Journal of Advertising*, 50(1), 81-91.
- Heinonen, K., & Medberg, G. (2018). Netnography as a tool for understanding customers: implications for service research and practice. *Journal of Services Marketing*, 32(6), 657-679.

- Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155-172.
- Ivanov, S., & Webster, C. (2020). Robots in tourism: A research agenda for tourism economics. *Tourism Economics, 26*(7), 1065-1085.
- Izard, C. E. (1977). Differential emotions theory. In *Human emotions* (pp. 43-66): Springer.
- Jörling, M., Böhm, R., & Paluch, S. (2019). Service robots: Drivers of perceived responsibility for service outcomes. *Journal of Service Research, 22*(4), 404-420.
- Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction, 19*(1-2), 61-84.
- Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research, 39*(1), 61-72.
- Kübler, R. V., Colicev, A., & Pauwels, K. H. (2020). Social media's impact on the consumer mindset: When to use which sentiment extraction tool? *Journal of Interactive Marketing, 50*, 136-155.
- Kulviwat, S., Bruner II, G. C., Kumar, A., Nasco, S. A., & Clark, T. (2007). Toward a unified theory of consumer acceptance technology. *Psychology & Marketing, 24*(12), 1059-1084.
- Laros, F. J., & Steenkamp, J.-B. E. (2005). Emotions in consumer behavior: a hierarchical approach. *Journal of Business Research, 58*(10), 1437-1445.
- LaValley, M. P. (2008). Logistic regression. *Circulation, 117*(18), 2395-2399.
- Lazarus, R. S., & Lazarus, R. S. (1991). *Emotion and adaptation*: Oxford University Press on Demand.
- Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science, 64*(11), 5105-5131.

- Lee, S., Ha, S., & Widdows, R. (2011). Consumer responses to high-technology products: Product attributes, cognition, and emotions. *Journal of Business Research*, 64(11), 1195-1200.
- Li, D., Rau, P.-L., & Li, Y. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2(2), 175-186.
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301-323.
- MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *Ai & Society*, 23(4), 485-510.
- McLeay, F., Osburg, V. S., Yoganathan, V., & Patterson, A. (2021). Replaced by a Robot: Service Implications in the Age of the Machine. *Journal of Service Research*, 24(1), 104-121.
- Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, 56(4), 535-556.
- Mirzaalian, F., & Halpenny, E. (2019). Social media analytics in hospitality and tourism: A systematic literature review and future trends. *Journal of Hospitality and Tourism Technology*.
- Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley: The original essay by Masahiro Mori. *IEEE Robotics & Automation*, 19(2), 98-100.
- Murphy, J., Hofacker, C., & Gretzel, U. (2017). Dawning of the age of robots in hospitality and tourism: Challenges for teaching and research. *European Journal of Tourism Research*, 15, 104-111.
- Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., & Zaki, M. (2014). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*, 17(3), 278-295.

- Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions*: Cambridge university press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.
- Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3-33): Elsevier.
- Purnawirawan, N., Eisend, M., De Pelsmacker, P., & Dens, N. (2015). A meta-analytic investigation of the role of valence in online reviews. *Journal of Interactive Marketing*, *31*, 17-27.
- Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587-595.
- Richins, M. L. (1997). Measuring emotions in the consumption experience. *Journal of Consumer Research*, *24*(2), 127-146.
- Robinson, S., Orsingher, C., Alkire, L., De Keyser, A., Giebelhausen, M., Papamichail, K. N., . . . Temerak, M. S. (2020). Frontline encounters of the AI kind: An evolved service encounter framework. *Journal of Business Research*, *116*, 366-376.
- Romani, S., Grappi, S., & Dalli, D. (2012). Emotions that drive consumers away from brands: Measuring negative emotions toward brands and their behavioral effects. *International Journal of Research in Marketing*, *29*(1), 55-67.
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality*, *11*(3), 273-294.

- Ryoo, J. H., Wang, X., & Lu, S. (2021). Do Spoilers Really Spoil? Using Topic Modeling to Measure the Effect of Spoiler Reviews on Box Office Revenue. *Journal of Marketing*, 85(2), 70-88.
- Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., & Chen, Y.-S. (2018). *Carer: Contextualized affect representations for emotion recognition*. Paper presented at the Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.
- Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition & emotion*, 7(3-4), 233-269.
- Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management*, 32(6), 1310-1323.
- Spiggle, S. (1994). Analysis and interpretation of qualitative data in consumer research. *Journal of Consumer Research*, 21(3), 491-503.
- Stahl, B. C., & Coeckelbergh, M. (2016). Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems*, 86, 152-161.
- Statista. (2022). Amount of data created, consumed, and stored 2010-2025. Retrieved from <https://www-statista-com.audenciagroup.idm.oclc.org/statistics/871513/worldwide-data-created/>. Retrieved Mar 18, 2022. <https://www-statista-com.audenciagroup.idm.oclc.org/statistics/871513/worldwide-data-created/>
- Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4), 463-479.
- Tung, V. W. S., & Au, N. (2018). Exploring customer experiences with robotics in hospitality. *International Journal of Contemporary Hospitality Management*, 30(7), 2680-2697.

- Turja, T., Aaltonen, I., Taipale, S., & Oksanen, A. (2020). Robot acceptance model for care (RAM-care): a principled approach to the intention to use care robots. *Information & Management, 57*(5), 103220.
- Tussyadiah, I. P., & Park, S. (2018). Consumer evaluation of hotel service robots. In B. Stangl & J. Pesonen (Eds.), *Information and Communication Technologies in Tourism 2018* (pp. 308-320). Cham: Springer.
- Tussyadiah, I. P., Zach, F. J., & Wang, J. (2020). Do travelers trust intelligent service robots? *Annals of Tourism Research, 81*, 102886.
- Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems, 81*, 41-53.
- Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research, 20*(1), 43-58.
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research, 11*(4), 342-365.
- Vermeer, S. A., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing, 36*(3), 492-508.
- Watson, L., & Spence, M. T. (2007). Causes and consequences of emotions on consumer behaviour: A review and integrative cognitive appraisal theory. *European Journal of Marketing, 41*(5/6), 487-511.

- Wirtz, J., Kunz, W., & Paluch, S. (2021). The service revolution, intelligent automation and service robots. *The European Business Review*, 38-44.
- Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. *Journal of Service Management*, 29(5), 907-931.
- Wong, A. (2004). The role of emotional satisfaction in service encounters. *Managing Service Quality: An International Journal*, 14(5), 365-376.
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International journal of hospitality management*, 44, 120-130.
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *ArXiv, abs/1906.08237*.
- Ye, F., Xia, Q., Zhang, M., Zhan, Y., & Li, Y. (2020). Harvesting online reviews to identify the competitor set in a service business: evidence from the hotel industry. *Journal of Service Research*, 1094670520975143.
- Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS quarterly*, 38(2), 539-560.
- Yu, C.-E. (2020). Humanlike robots as employees in the hotel industry: Thematic content analysis of online reviews. *Journal of Hospitality Marketing & Management*, 29(1), 22-38.

TABLES AND FIGURES

Table 1. Performance evaluation for each label:

Emotions	Precision	Recall	f1-score	Support
Anger	0.74	0.82	0.78	693
Sadness	0.90	0.82	0.86	679
Joy	0.81	0.87	0.84	707

Fear	0.83	0.78	0.80	638
Neutral	0.81	0.79	0.80	676
Accuracy			0.82	3393
Macro average	0.82	0.82	0.82	3393
Weighted average	0.82	0.82	0.82	3393

Table 2. Performance comparison with benchmark models on the test dataset

Model	MCC
XLNet	0.8053
BERT	0.7984
Support Vector Machine (Linear)	0.5901
Logistic Regression	0.5890
Naïve Bayes (Multinomial)	0.4962
Random Forest	0.4330

Table 3. A comparison **topic** classification between BERTopic and human interpretation

Topics	BERTopic	Human
Feelings	✓	✓
Towel delivery	✗	✓
Water delivery	✓	✗
Cooking	✓	✓
Room service	✓	✓
General delivery	✓	✓
Lobby service	✓	✓
Children’s feeling	✓	✓
Food delivery	✓	✓
Toiletries delivery	✗	✓
Coffee delivery	✓	✓
Take-away delivery	✓	✓
Question Answering	✓	✓
General discussion on robot	✓	✗

Notes: ✓ = included; ✗ = not included. Jaccard coefficient: 0.71.

Table 4. The distribution of emotion

Emotion	Counts	Percentage
Joy	4349	61.02%
Fear	2033	28.52%
Anger	368	5.16%
Neutral	312	4.38%
Sadness	65	0.91%

Table 5. A comparison of dimensions between BERTopic and human interpretation

Emotion	Bert Topic Counts/%	Qualitative analysis
Joy	4349 (61.02%)	1008 (21%)
Fear	2033 (28.52%)	1 (0.02%)
Anger	368 (5.16%)	3 (0.06%)
Neutral	312 (4.38%)	620 (13%)
Sadness	65 (0.91%)	22 (0,45%)
Love	-	963 (20%)
Surprise	-	569 (12%)
Interest	-	266 (5%)
Excitement	-	64 (1,3%)
Discontent	-	63 (1,3%)

Topic Word Scores

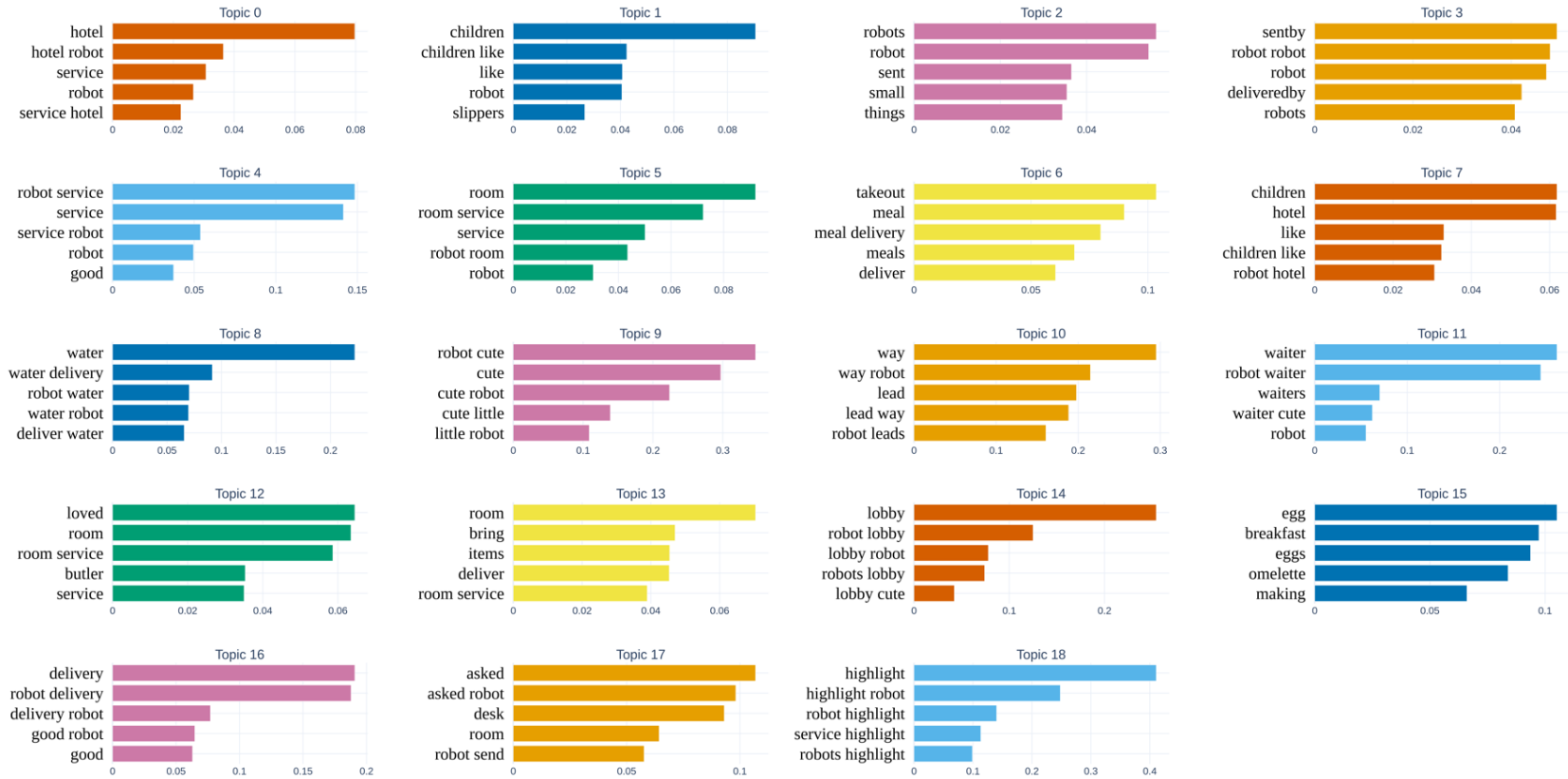


Figure 1. Keyword scores within each topic

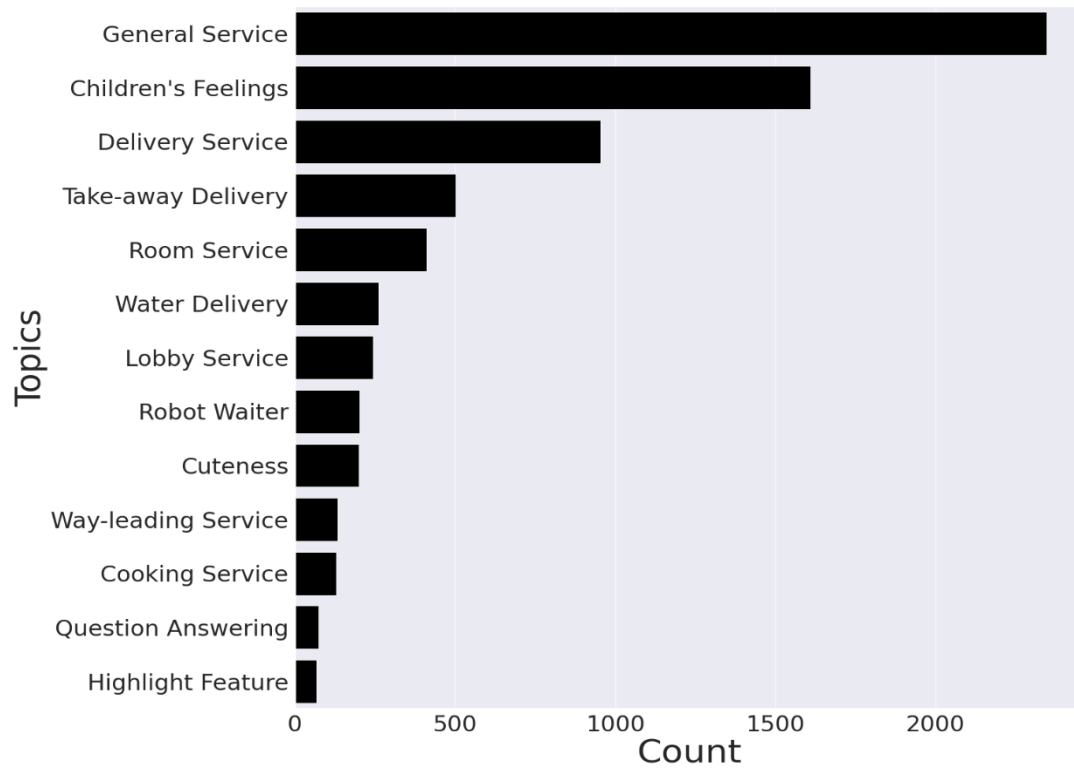


Figure 2. The distribution of keywords within each topic