Enablers or Inhibitors? Unpacking the Emotional Power Behind In-Vehicle AI Anthropomorphic Interaction: A Dual-Factor Approach by Text Mining

Shizhen Bai, Dingyao Yu, Chunjia Han, Mu Yang, Nazrul Islam, Zaoli Yang, Rui Tang, and Jiayuan Zhao

***Abstract*—The intelligent strategy of the new energy vehicle (NEV) industry has triggered the rapid prevalence of in-vehicle anthropomorphic artificial intelligence (AI) assistants. There is still a lack of clarity regarding NEV users' attitudes toward this cutting-edge technology and whether they receive a satisfactory intelligent service experience. To circumvent potential emerging technology resistance, in this article, we utilize text analysis techniques for the identification of AI interaction emotions, love and disgust (enablers and inhibitors) with significant influence on user satisfaction, and validates the improving role of multimodality on the effectiveness of anthropomorphic interaction. In addition, this study innovatively constructs a multidimensional corpus of modality × emotion, using structural topic modeling to uncover the constituent elements and real-time changes of love and disgust emotions in different modalities, from which development opportunities and improvement directions for AI anthropomorphic interaction technologies are identified. The findings provide new insights into the application of emotion analysis methods to improve users' intelligent service experience and provide a realistic reference for mitigating emerging technology resistance in the NEV industry.**

***Index Terms*—Anthropomorphism, artificial intelligence (AI) interaction, discrete emotion analysis, dual-factor analysis, emerging technology resistance, multimodal, new energy vehicle (NEV), structural topic modeling (STM).**

Manuscript received 9 January 2023; revised 25 April 2023 and 12 September 2023; accepted 11 October 2023. This work was supported in part by the National Social Science Fund of China under Grant 23BJY151 and in part by the National Social Science Fund of China under Grant 23BJY191. Review of this manuscript was arranged by Department Editor T. Daim. *(Corresponding author: Chunjia Han.)*

Shizhen Bai and Dingyao Yu are with the School of Management, Harbin University of Commerce, Harbin 150000, China (e-mail: baishzh1962@126.com; ydy_tr@163.com).

Chunjia Han and Mu Yang are with the Department of Management, Birkbeck, University of London, WC1E 7HX London, U.K. (e-mail: chunjia.han@bbk.ac.uk; m.yang@bbk.ac.uk).

Nazrul Islam is with the Royal Docks School of Business and Law University of East London, University Square Stratford, E15 1NF London, U.K. (e-mail: nazrul.islam@uel.ac.uk).

Zaoli Yang is with the College of Economics and Management, Beijing University of Technology, Beijing 100871, China (e-mail: yangzaoli@bjut.edu.cn).

Rui Tang is with the School of Economics Teaching and Research, Party School of the Central Committee of C.P.C (Chinese Academy of Governance), Beijing 100871, China (e-mail: tr940419@163.com).

Jiayuan Zhao is with the School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150000, China (e-mail: hrbcuzjy@163.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TEM.2023.3327500.

Digital Object Identifier 10.1109/TEM.2023.3327500

I. Introduction

Enabled by digital technologies such as natural language processing (NLP), anthropomorphic artificial intelligence (AI) assistants are accelerating to cover application scenarios in the Industry 5.0 era and are already widely used to communicate and interact with customers in scenarios such as hotels, restaurants, airlines, banks, education, automobiles, and electronic devices [1]. Unlike early-stage voice bots that can only recognize specific commands, they are able to make more intelligent and natural conversational responses by analyzing the meaning and context of human language [2] and often integrate more anthropomorphic features and social cues [3], such as human-like names and appearance [4], in an attempt to attract and delight users. However, there is no systematic and scientific approach to reviewing the performance of this emerging AI product from the user's perspective.

Typical examples are the rapid popularity of anthropomorphic AI assistants in new energy vehicles (NEV). As an embedded virtual service robot for NEV users [5], which automates basic, repeatable, and standardized user interaction services, it is able to play an irreplaceable role in multiple application scenarios, such as autonomous driving and assisted driving [6]. As a representative industry of Industry 5.0, the popularity of in-vehicle anthropomorphic AI assistants caters to the intelligent strategy and trend of NEV [7], which accelerates the digital transformation of the NEV industry from the design, production, and marketing of products. Despite the significance of in-vehicle anthropomorphic AI assistants for the NEV industry, little existing research has focused on and understood the attitudes and experiences of NEV users toward this cutting-edge technology [8]. User experience and responses are considered important references to guide product design improvements and make innovation decisions [9], thus revealing the AI anthropomorphic interaction experience of NEV users is of particular interest to researchers and practitioners.

In addition, according to social cognitive theory, users' perceived experiences will eventually be transformed into behavioral outcomes [10], and users' satisfaction will significantly influence subsequent behavioral intentions [11]. This means that the AI anthropomorphic interaction experience of NEV users will influence the users' satisfaction with the AI assistant's intelligent services, and the users' satisfaction ratings will determine the willingness and intention to continue using the anthropomorphic AI assistant [11], [12]. To circumvent the potential emerging technology resistance [13] resulting from any form of intelligent service failure or crisis of confidence, e.g., reduced frequency of AI interaction due to erroneous voice navigation threatening driving safety, NEV users' anthropomorphic AI interaction experiences should further be separated into two categories of user satisfaction influencing factors, enablers and inhibitors [14], and empirically test their significant differential impact on users' satisfaction with intelligent services. This is valuable for NEV companies to maintain sustained user attractiveness and cope with emerging technology resistance by building on their strengths and improving their weaknesses. However, no research related to AI anthropomorphic interaction and user satisfaction with intelligent services has been conducted under the perspective of dual-factor analysis. This forms the first gap in this study.

After integrating existing research findings, we have found that studies addressing user experience and user feedback on AI interaction have mostly focused on two dimensions: perceived competence and social relationships. For example, usefulness, ease of use, entertainment, convenience, information search, and task execution functions have been shown to significantly influence the willingness to use and satisfaction with personal voice assistants [15], [16], [17], [18]. Task attractiveness can exert an influence on the continued willingness to use AI voice assistants by affecting user satisfaction [12]. Functional service failures due to incorrect or delayed recognition of voice reinforce users' resistance to using innovative technologies [13]. In addition to the perceived competence perspective, Choi and Drumwright [18] argue that anthropomorphic features of AI can exert an influence on trust and acceptance of AI interaction through perceived empathy and interaction quality. Pelau et al. [19] propose that the feelings of intimacy, understanding, enjoyment, and engagement generated by AI interaction influence users' continued willingness to use. Competence and sociality, as the two evaluation dimensions of social cognition [10], are able to cover the user experience of AI interaction but cannot be clearly categorized into enablers and inhibitors of user responses by empirical testing, ignoring the compounding effects of the intersection between perceived competence and social relationships on user responses, such as the repairing effect of anthropomorphic intimate social relationships on service failure [20]. Actually, the AI anthropomorphic interaction experience of NEV users is essentially an emotional response to a stimulus [21]. Considering the complexity of the user's emotional experience caused by the uncanny valley phenomenon, the granularity of emotions should go beyond the typical distinction between positive and negative, and discrete emotions can reflect the subtle distinctions within the user experience in a more nuanced way [21]. However, scholars have yet to focus on the rationale and necessity of measuring complex user experiences in terms of discrete emotions, which forms the second gap in this study. Therefore, integrating the above two research gaps, this study innovatively proposes the following research questions.

> RQ1: *Within a dual-factor theoretical perspective, how do the discrete emotions generated by AI anthropomorphic interaction affect NEV users' satisfaction with intelligent services?*

Unlike current research that uses quantified discrete emotions as variables to verify some causalities [22], [23], the researchers concluded that merely identifying the enabling and inhibiting effects of users' discrete emotions on the satisfaction of anthropomorphic AI interaction is still unable to provide NEV enterprises with more targeted product and technology improvement suggestions and that it is also necessary to use structural topic modeling (STM) to clarify the components of the most influential emotions and to understand the reasons for NEV users' discrete emotions during anthropomorphic AI interaction, as well as the real-time attention to the anthropomorphic AI interaction at different times. The biggest challenge in analyzing user-generated content (UGC) is how to extract valuable information from the vast amount of unstructured data. As a disruptive text analysis technology, STM is able to gain real-time insights and important findings from users with good speed and accuracy [24]. This forms the third gap in this study, and a second research question is posed accordingly.

> RQ2: *What are the components of the most dominant emotions that influence NEV user satisfaction? How do users' concerns about AI anthropomorphic interaction change over time?*

Notably, the researchers also observe that the modality of AI anthropomorphic interaction may influence the effectiveness of the interaction, and in turn, the evaluation of AI anthropomorphic interaction. However, there are currently voice assistants that have not only anthropomorphic human-like names but also anthropomorphic human-like figures. The fusion of multimodal information (voice and visual) will complement and validate the emotional cues generated by human–computer interaction [25], which will have an impact on the emotions of NEV users and consequently on the evaluation of AI anthropomorphic interaction. The influence of AI interaction modality is being formally explored for the first time in the NEV domain. To fill the fourth research gap, we propose the following research question to guide this study.

> RQ3: *How do the modalities of AI anthropomorphic interaction moderate the influence of discrete emotions on NEV user satisfaction?*

To answer the above questions, we attempted to obtain the textual data needed for the study from a large amount of UGC through data mining techniques, which break the quantitative limitations and overly subjective drawbacks of traditional questionnaire methods. Specifically, we collected user reviews of different NEV brands from the largest car websites in China. Recent studies have shown that the emotions contained in the text are more reflective of users' true feelings toward products and services than the textual content in online reviews [26]. The DUTIR Emotion Lexicon is introduced to identify and quantify discrete emotions generated by AI anthropomorphic interaction [27], and multiple linear regressions are used to empirically test the relationship between discrete emotions, users' satisfaction, and AI anthropomorphic interaction modalities. The main emotions that significantly affect user satisfaction are classified as enablers and inhibitors, integrated within a dual-factor framework, using an unsupervised machine learning approach, STM, to reveal the constituent elements of enablers and inhibitors and the real-time attention of NEV users to AI anthropomorphic interaction.

The contributions of this study are mainly reflected in the following aspects. First, adopting the text analysis method, we creatively propose to reveal the AI anthropomorphic interaction experience using discrete emotions based on UGC and introduce an emotion lexicon to quantify discrete emotions. Second, the main discrete emotions that significantly affect NEV users' satisfaction are classified as the final enablers and inhibitors, extending the application of dual-factor theory in the field of users' experience management. Third, the corpus integration model is innovated to perform STM for topic clustering, which clarifies the constituent elements and real-time changes of the main discrete emotions. Finally, the moderating role of the modality of AI anthropomorphic interaction is revealed.

II. Literature Review and Hypotheses Development

A. Users' Satisfaction and Acceptance of Anthropomorphism AI Assistants

AI assistants are voice-enabled integrated AI technologies that are seen as dynamic systems with the ability to learn customer preferences [28]. One of the most widely used consumer-oriented AI products uses users' inputs, such as voice, vision (images), and contextual information to help them by answering questions, making suggestions, and performing actions in natural language [29]. The development of NLP and visualization technologies has made AI assistants more anthropomorphic and enhanced their social attributes [30]. Anthropomorphism has been defined as the tendency to attribute human characteristics, behaviors, emotions, and intentions to nonhuman entities [31]. As conversational nonhuman agents, anthropomorphism AI assistants not only provide functional services to users but also establish a social relationship with them through human-like clues, allowing users to have an emotional experience similar to interacting with people [32].

Given the tremendous business value of anthropomorphism AI assistants for users and enterprises, Statista's industry analysis predicts that AI assistant usage is expected to grow from 3.25 billion in 2019 to 8.4 billion in 2024 [33]. However, Kinsella [34] reported that only 24% of U.S. adults use voice commands on a daily basis and only 30% of consumers find digital voice assistants easier to use to solve service problems [35]. This implies that user acceptance does not match the rate of popularity of anthropomorphism AI assistants. Potential emerging technology resistance is likely to exist as users have traditionally tended to resist interaction with automated and AI environments [36].

Studies on users' acceptance of anthropomorphism AI assistants mostly tested the roles of factors related to technology utilization originating from the technology acceptance model (TAM), unified theory of acceptance and use of technology, and the service robot acceptance model [2], [3], [37]. The conclusions suggest that users' acceptance of automated service technologies should depend not only on functional performance but also on their ability to meet social emotional and relational needs [38]. However, fewer studies have explored the users' acceptance of anthropomorphism AI assistants from the perspective of user satisfaction as a mediator [12], [39].

Satisfaction is a positive affective state associated with the feeling of fulfillment and is the result of a comparison between actual and expected behavioral outcomes, reflecting the user's emotional state and level of pleasure [39]. Satisfaction is a crucial concept as it predicts positive behavior [40] and can exert influence on users' behavioral intentions after usage [11]. Various studies have shown that the main determinant of users' intention to continue using is satisfaction [41], [42]. Therefore, it is crucial to emphasize the importance of user satisfaction in the acceptance of AI anthropomorphic interaction.

B. Dual-Factor Theory

The dual-factor theory states that humans have different sets of needs, leading to a positive or negative outcome, resulting in satisfaction or dissatisfaction [43]. Combining Han and Marikyan's studies [12], [39], we argue that users' acceptance of anthropomorphism AI assistants is the outcome of satisfaction, and satisfaction mediates the impact of users' AI anthropomorphism interaction experience on users' acceptance. In order to targetedly enhance users' acceptance of anthropomorphic AI assistants, we have to precisely improve users' satisfaction. Therefore, it is essential to identify the enablers and inhibitors of users' AI interaction satisfaction from the perspectives of positive and negative user experience under the guidance of dual-factor theory. Satisfaction and acceptance can be increased by promoting the enablers and improving the inhibitors.

Based on a satisfaction perspective, the dual-factor theory has been widely used to understand technology acceptance behavior [44], which also includes robot acceptance. Sharma and Mishra [45] developed a dual-factor research model that considered users' perceptions of enablers and inhibitors of mobile payments and verified that enablers positively influenced user satisfaction and drove continued use, while inhibitors exerted a negative effect on satisfaction. Based on the dual-factor theory, Talwar et al. [46] validated the intention of users' positive word-of-mouth for continued use of mobile wallets by considering perceived information quality, perceived competence, and perceived benefit as enablers of word-of-mouth effectiveness, and perceived cost, perceived risk, and perceived uncertainty as inhibitors of word-of-mouth effectiveness. Balakrishnan et al. [47] constructed a dual-factor analysis framework that positioned status quo bias factors as inhibitors (sunk cost, regret avoidance, inertia, perceived value, switching costs, and perceived threat) and positioned the TAM as enablers (perceived ease of use and perceived usefulness), identifying attitudes and resistance to adoption of AI voice assistants. In the context of promoting technology acceptance, researchers have identified different enablers and inhibitors.

We refer to previous studies and try to measure experiences in terms of emotions [48], considering positive AI interaction emotions as enablers and negative AI interaction emotions as inhibitors, and integrate them into a dual-factor framework to formulate research hypotheses for the enabling and inhibiting effects of complex AI interaction experiences on satisfaction respectively.

C. Users' Experience, Emotions, and Satisfaction

Within the scope of consumer experience management research, there are multiple ways of defining experience for users, for example: "interactions between organizations and customers [48]," "the collection of points at which companies and consumers exchange sensory stimuli, information, and emotions [49]," or "nondeliberate, spontaneous responses and reactions to particular stimulation [50]," etc. In this study, the AI anthropomorphic interaction experience of NEV users should be considered as the user's specific emotional responses when confronted with the anthropomorphic service provided by the in-vehicle AI assistant. As an interface connecting enterprises and users, the anthropomorphic interaction between in-vehicle AI assistants and NEV users is essentially one of the important channels for enterprises to maintain user stickiness and gain continuous insight into the user experience.

Current scholars commonly identify users' AI anthropomorphic interaction experiences in terms of effectiveness and sociality dimensions. For example, Marikyan et al. [39] explored the impact of experience elements such as perceived anthropomorphism, perceived entertainment, perceived intelligence, perceived social presence, and performance expectations on AI assistants' satisfaction. Fernandes and Oliveira [3] classified the factors influencing the acceptance of AI assistants into functional elements (perceived usefulness, perceived ease-of-use, and subjective social norms), social elements (perceived humanness, perceived social interactivity, and perceived social presence), and relational elements (trust and rapport). Han and Yang [12] verified that task attraction, social attraction, physical attraction, and security/privacy risk can exert a positive influence on continued usage intentions by improving user satisfaction.

However, we believe it is more convincing to measure NEV users' AI anthropomorphic interaction experiences in terms of discrete emotions. This is because users' discrete emotional responses to specific stimuli are more subtle [51] and can accommodate the complexity of anthropomorphic interaction experiences relative to binary emotions (positive and negative). Evidence is provided by past research. Orea-Giner et al. [52] used text-mining techniques to explore the relationship between emotions and sentiments generated by customers during hotel–robot interactions and the potential impact on hotel ratings. From the micro-, meso-, and macrolevels, Bagozzi et al. [53] classified emotions resulting from AI service interaction into basic emotions (e.g., joy, sadness, and fear), self-conscious emotions (e.g., pride, guilt, embarrassment), and moral emotions (e.g., contempt, righteous anger, social disgust). Taking the use of AI Chatbots by employees in the digital workplace as a specific type of AI system, Gkinko and Elbanna revealed the emotional experience of this system: hope, tolerance, and empathy [54].

Furthermore, users' AI anthropomorphic interaction experiences identified from the competence and social dimensions cannot be clearly classified as enablers and inhibitors. However, there is an interactive influence between perceived competence and perceived sociality as two evaluation dimensions of social cognition. For example, a robot service failure might be forgiven by the user because of its cute appearance [55] or humorous language [20], thus masking the true negative impact of the service failure on user satisfaction. This does not help managers to identify and avoid potential pitfalls that undermine user satisfaction timely. This study, therefore, uses discrete emotions that can convey explicit attitudes to cover the functional and social AI anthropomorphic interaction experience of NEV users.

Current mainstream methods for quantifying discrete emotions mainly include machine learning-based methods, lexicon-based methods, and hybrid methods [56], [57]. However, due to the lack of mature annotated datasets in the Chinese and NEV AI technology environments, which makes machine learning methods and hybrid methods difficult and the accuracy of emotion recognition is not as good as using emotion lexicons, we choose an emotion lexicon-based approach to quantify discrete emotions for AI anthropomorphic interaction. The specific quantification process is given in Section III.

Regarding the identified discrete emotion categories, we referred to the six basic emotions proposed by Ekman [58] and optimized them according to the chosen emotion lexicon to form the final emotion categories: anger, disgust, fear, sadness, surprise, love, and joy. As the antecedents of satisfaction, the impact of discrete emotions on user evaluation has formed a common conclusion [59], [60]; thus, we agree that negative emotions such as anger, disgust, fear, and sadness negatively affect user evaluation [52], while positive emotions, such as surprise, love, and joy positively affect user service outcomes [61]. Therefore, we argue that negative discrete emotions generated by AI anthropomorphic interaction negatively affect NEV users' satisfaction and positive discrete emotions generated positively affect NEV users' satisfaction. Satisfaction with the AI anthropomorphic interaction is measured using intelligent ratings provided by NEV users, which are determined based on the metadata of the reviews we obtained. Accordingly, we formulated the following research hypothesis.

> ***H1:*** *The negative emotions generated by AI anthropomorphic interaction, anger (H1a), disgust (H1b), fear (H1c), and sadness (H1d), have a significant negative impact on the intelligent rating.*

> ***H2:*** *The positive emotions generated by AI anthropomorphic interaction, surprise (H2a), love (H2b), and joy (H2c), have a significant positive impact on the intelligent rating.*

D. Moderating Effect of Modalities of AI Anthropomorphic Interaction

The development of AI interaction technology has led users to warm up to AI agents that add anthropomorphic visual elements besides anthropomorphic voice [62] for seeking a higher degree of anthropomorphism interactive experience. Anthropomorphic voice interaction is mainly reflected in the fact that AI voice assistants have human-like voices that can distinguish between

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

genders, human-like names as waking words [63], and conscious conversations like a human. However, pieces of evidence show that the popularity of AI interaction may also depend on its anthropomorphic appearance. Compared to disembodied AI agents, robots with an embodied human-like figure are more likely to prompt social interaction [64] and have a more positive impact on users' acceptance.

However, this perception of anthropomorphic visual interaction is not rigorous as the visual cues are considered to be static information but do not interact with users visually. A truly anthropomorphic visual interaction should have a dynamic embodied human-like figure. For example, when opening a car door, a voice assistant with a human-like figure turns its head toward you and says "Hi" with a smile.

No attention has been paid to the modalities in AI anthropomorphic interaction. Compared to single-modal, multimodal AI anthropomorphic interaction is capable of visual interaction alongside voice interaction, and the information of voice and visual can be consistent and complementary to each other, thus achieving better interaction results. The dual coding theory shows [65] that the human brain can enhance information recall and recognition by processing information visually and verbally simultaneously, as information between different modalities is corroborated and complemented by each other [66]. Similarly, emotional cues are also mutually corroborated and complemented, eliminating emotional differences and enabling more accurate emotional transmission during interaction [67].

Considering the impact of modality fusion on discrete emotions, we sought to explore how the fused interaction of voice and visual would affect the NEV user's AI anthropomorphic interaction experience. Specifically, how will modalities of AI anthropomorphic interaction affect the relationship between discrete emotions and intelligence rating? Accordingly, we propose the following research hypotheses.

***H3:** Compared with single-modal AI anthropomorphic interaction, multimodal AI anthropomorphic interaction significantly mitigates the negative influence of negative emotions anger (H3a), disgust (H3b), fear (H3c), and sadness (H3d) on intelligent rating.*

***H4:** Compared with single-modal AI anthropomorphic interaction, multimodal AI anthropomorphic interaction significantly enhances the positive impact of positive emotions surprise (H4a), love (H4b), and joy (H4c) on intelligent rating.*

E. User-Generated Content and Text Analysis

In recent years, UGC has become an increasingly important source of information for enterprises and consumers, with millions of users posting reviews on various platforms, such as Amazon, TripAdvisor, and Yelp. These reviews not only provide textual content but also valuable insights to identify users' emotions and experiences. After experiencing products or services, users write and post various reviews related to feelings and feedback, which are authentic and timely, reflecting their true opinions and reactions [68]. The popularity of social media and online platforms has led to an explosion of online reviews, which has led to an increased focus on obtaining users' perceptions and experiences of product usage from their reviews [21]. Unlike traditional methods that rest on questionnaires, interviews, and focus groups to capture users' real needs and determine product development and marketing strategies, UGC offers managers new and abundant data opportunities to measure consumer perceptions, attitudes, and intentions as a high-quality alternative source of information [69].

Several recent studies on human–computer interaction have used text mining and analysis techniques to not only reveal users' concerns about text contents but also to identify and quantify users' emotional experiences. Park et al. [70] used a text-mining approach to examine the impact of multiple dimensions of hotel service robot attractiveness on customer emotions from online reviews. Filieri et al. [57] verified that consumer emotions shared in online reviews are critical in predicting consumer willingness to adopt service robots. Huang et al. [71] developed a framework for representing the customer experience of service robots through the analysis of reviews from the hospitality and tourism industry, providing insight into customer–robot interaction in terms of sensory, conative, cognitive, and emotional experiences.

The empirical examination of the relationship between NEV users' AI interaction emotions and intelligence rating is able to classify the main emotions that significantly affect user satisfaction into enablers and inhibitors, but it is not able to trace the components of the main emotions and reveal the deeper reasons for the creation of enablers and inhibitors. The user's emotional experience is not static; as the time of usage and the frequency of interaction increase, the users may encounter an emotional turnaround due to the lack of novelty or the failure of the service [72]. This study attempts to identify the generation causes and real-time changes in the emotional experience of NEV users based on text clustering results with the help of a subversive text analysis method STM.

III. Methodology and Implementation

In Fig. 1, we build the overall framework of the research design and reveal the specific process of data acquisition and analysis.

A. Data Acquisition and Preparation

The NEV online reviews used in this study come from the CarFriend forum of the DCar website. As one of the largest online platforms for car information and transactions in China, it only allows real car owners to share their car experiences and evaluation results. However, this source of high-quality textual data has not received much attention from scholars in UGC-related studies.

In-vehicle AI anthropomorphic assistants can be separated into two categories: single-modal (voice) and multimodal (voice and vision). For example, BYD gives a human-like name (Xiaodi) to its AI assistant but only supports voice interaction. NIO's AI assistant has an embodied human-like figure in addition to a human-like name (Nomi) and supports both voice and visual interaction.

Combining the sales and market share of NEV in China, we select BYD as a sample for single-modal AI anthropomorphic

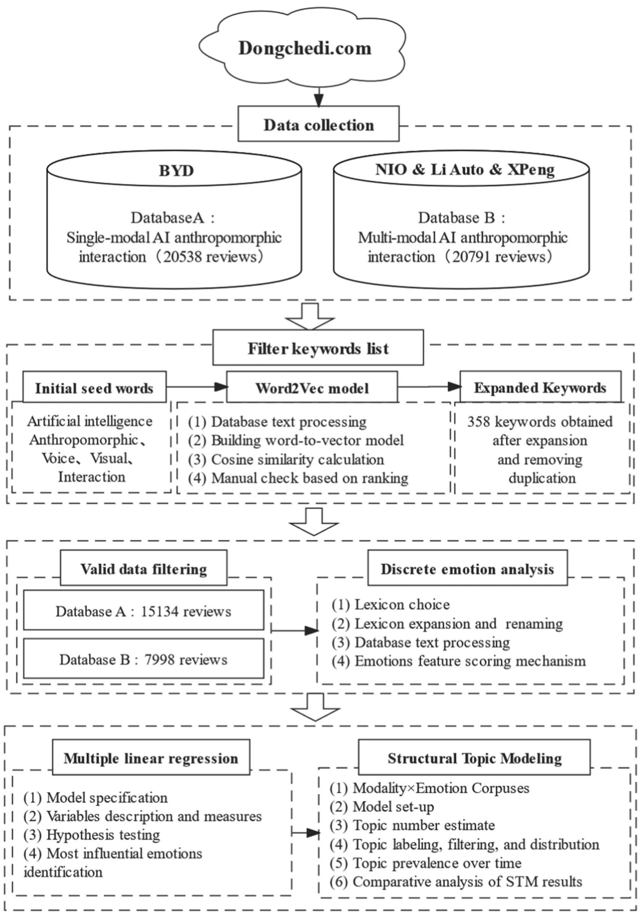


Fig. 1. Overall framework for research design.

interaction and select Li Auto, NIO, and Xpeng as samples for multimodal AI anthropomorphic interaction. The underlying technology for the above NEV brands' AI voice assistants is provided by the same intelligent voice company iFLYTEK, which occupies the largest market share in China's automotive intelligent voice interaction field, so they are comparable in terms of technology and quality. We developed an automated data crawler using Python to collect a total of 41 329 reviews of all the vehicle products released from June 1, 2018 to October 31, 2022. Among them, 20 538 reviews are from BYD, constituting database A for single-modal AI anthropomorphic interaction, and 20 791 reviews are from Li Auto, NIO, and Xpeng, which constitute database B for multimodal AI anthropomorphic interaction.

To reduce the interference of noisy data, we need to restrict our attention to the paragraphs related to "AI anthropomorphic interaction." Therefore, we must set keywords to filter the valid comments. To make the filtering of reviews more scientific and to ensure the maximum amounts of valid reviews, we have not set all the filtering keywords manually but first set five seed words "Artificial Intelligence," "Voice," "Visual," "Anthropomorphism," and "Interaction," where "Visual" is only for Database B, the other four seed words are both for Database A and Database B. Then a machine learning algorithm with a word embedding model is employed to obtain extended keywords with similar meanings to the seed words [73]. Finally, the valid reviews containing the seed words and the extended keywords are filtered from the database A and B. The specific steps are as follows.

First, building the word-to-vector model. Word2vec, a general-purpose word embedding model, maps uncomputable, unstructured words to computable, structured numeric vectors [74] and is used to quantitatively mine word-to-word relationships to better understand the semantic similarity between any given word and with other words in a database [75]. To create the Word2vec model, we perform text cleaning, including Chinese Jieba splitting, removing punctuation, numbers, stop words, and meaningless symbols, etc. The training of the Word2vec model is done by the Gensim library in Python. After training, all words in database A and database B, including the five seed words, are transformed into a 300-dimensional vector representing the meaning of the word.

The cosine similarities between the seed word vector and the other word vectors are then calculated and ranked. The top 1000 words that are most closely associated with the seed word vector (highest cosine similarity) are listed and manually checked to eliminate any ambiguous, inappropriate words that are not relevant to the topic. In the end, only 358 extended keywords are retained, which together with the 5 seed words constitute the final list of filtered keywords. After filtering, 14 483 valid reviews remained in database A and 7998 valid reviews remained in database B.

B. Lexicon-Based Discrete Emotions Quantification

Emotion lexicons consist the words representing specific emotions. The emotional tendency and intensity contained in each review are determined by matching the emotion words in the emotion lexicon and text database [76]. However, mainstream emotion lexicons such as NRC [77] are better at emotion analysis for English texts than for Chinese texts. This is due to the semantic bias of the emotion words in the emotion lexicons when translated into Chinese, resulting in less accurate recognition and classification of emotions. Therefore, to correctly identify the emotion categories of Chinese reviews, we chose to use the Emotion Lexicon Ontology Database of the Dalian University of Technology Institute of Information Retrieval (DUTIR) as a Chinese emotion lexical resource.

The DUTIR Emotion Lexicon is based on Ekman's six major categories of emotions (anger, disgust, fear, joy, sadness, and surprise), with the addition of the emotion category "good" to provide a more detailed classification of positive emotions [78]. Considering the appropriateness of the Chinese expression and the results of existing research [61], we rename the emotion category "good" to "love." Furthermore, while generalizability is an essential strength of the DUTIR Emotion Lexicon, it also means that the lexicon lacks specific emotion words for NEV and AI interaction scenarios. To compensate for this absence of data sources, we count and rank the word frequencies of text databases A and B. Three social science researchers are recruited to manually select 157 high-frequency emotion words related to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
EMOTION CATEGORIES IN DUTIR EMOTION LEXICON ONTOLOGY DATABASE

No.	Categories	Subcategories	Example words
1	Joy	joy, reliable	happy, glad, smiling, relieved, reassured
2	Love	respect, praise, trust, love, wish	reasonable, handsome, excellent, blessing
3	Anger	anger	anger, rage, furious, indignant, mad
4	Sadness	sadness, disappointment, guilt, miss	grief, regret, despair, guilt, pity, missing
5	Fear	panic, fear, shame	panic, flustered, timid, alarmed, nervous
6	Disgust	boredom, abhorrence, denounce, jealousy, doubt	depressed, fidgety, upset, shameful, stupid, hateful, jealous, suspicious, vain
7	Surprise	surprise	strange, miraculous, shocked, marvel

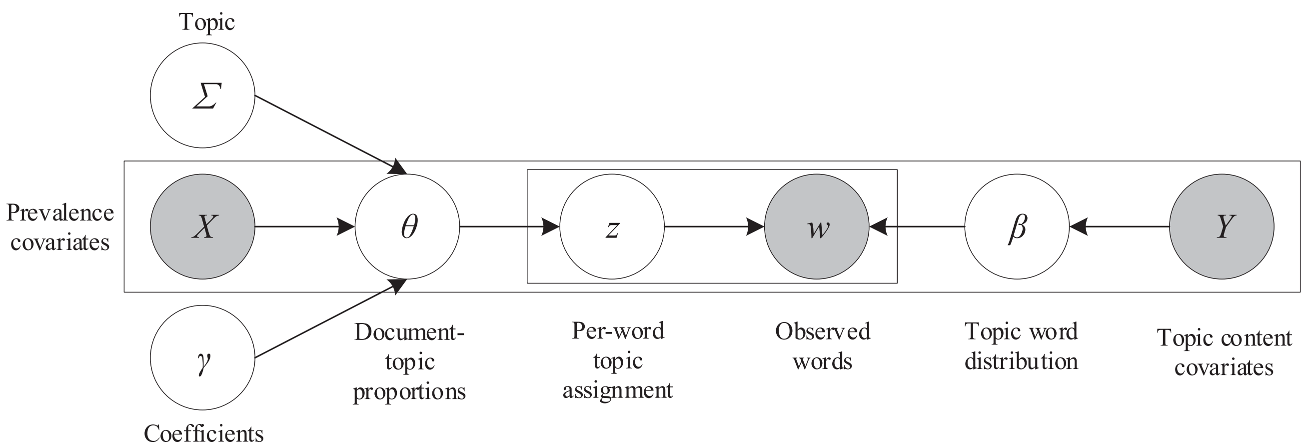


Fig. 2. Graphical illustration of the STM approach.

NEV and AI interaction to be added to the DUTIR Emotion Lexicon. The final emotions are divided into 7 major categories and 21 subcategories, with a total of 27 623 emotion words. Table I presents the final emotion classification of the DUTIR Emotion Lexicon.

For an operational approach to quantify discrete emotions in valid reviews using the DUTIR Emotion Lexicon, we refer to the research of the authors in [79] and [80]. The discrete emotions featuring scoring mechanism is used to construct a discrete emotion feature space containing seven emotion dimensions. The structure of the emotion dimensions is described as: DE < Joy, Love, Anger, Sadness, Fear, Disgust, Surprise >. The feature scores of discrete emotions in each dimension are measured by the ratio of the number of specific emotion words contained in a review to the total number of words in the review. For example, the process of calculating the emotion feature score for the LOVE dimension contained in review x can be represented by the following equation:

$$\begin{aligned} &DE_x < \text{love} > \\ &\quad = (\text{love} - \text{relatedwords}/\text{words in a review}) * 100. \end{aligned}$$

C. Structural Topic Modeling

Topic modeling is an unsupervised machine learning method, which is suitable for analyzing UGC, such as online reviews [81]. This series of methods checks the cooccurrence relationship between words in a text document and outputs a set of words with high cooccurrence probability, namely, the topic. We choose STM [82] because of the advantage that it allows merging the metadata of data samples to explain local popularity [81], [82]. Fig. 2 shows the illustration of STM and visualizes the text analysis process. The key processes of STM can be summarized as follows.

STM is a hierarchical model, in which the prevalence of each topic of document d (denoted as θ_d) is derived from the logical-normal distribution, and its average value is a function of the document covariate X_d. In this article, the document d represents a comment provided by visitors

$$\theta_d \sim \text{LogisticNormal}\left(X_{d\gamma}, \sum\right).$$

Then, given the topic-popularity vector, a specific topic $z_{d,n}$ is associated with the position that needs to be filled through the following process, where n is the index of each word in document d:

$$z_{d,n} \sim \text{Multinomial}\left(\theta_d\right).$$

Next, assign the words $w_{d,n}$ of each document to the topics

$$w_{d,n} \sim \text{Multinomial}\left(\beta_{d,z}\right)$$

where $\beta_{d,z}$ is the probability of choosing a vocabulary word w to fill a certain position in document d given the topic assignment variable z.

STM is considered to be superior in mining valuable users' insights and real-time focus from unstructured data. For example, Yang and Han [83] used STM to conduct a real-time survey of the UGC on Twitter to reveal the impact of COVID-19 on the hospitality industry, the challenges it faces, and the industry's response. Bai [84] explored the impact of visitor experience on visitor satisfaction in theme parks over time using dynamic visitor-generated review data.

IV. Analysis

A. Model Specification, Variables, and Measures

One of the critical objectives of this study is to investigate the influence of discrete emotions generated by in-vehicle AI anthropomorphic interaction on NEV users' intelligent service satisfaction and to examine the moderating effect of the modalities on such influence. In conjunction with the research hypothesis presented in this article, we employ multiple linear regressions to construct a hypothesis-testing Model. The model constructed is defined as follows:

Intelligent rating $= \beta_0 + \beta_1(\text{Anger}_i) + \beta_2\ (\text{Disgust}_i) + \beta_3(\text{Fear}_i) + \beta_4\ (\text{Sadness}_i) + \beta_5(\text{Surprise}_i) + \beta_6(\text{Love}_i) + \beta_7(\text{Joy}_i) + \beta_8(\text{modality}_i) + \beta_9(\text{Anger}_i \times \text{modality}_i) + \beta_{10}(\text{Disgust}_i \times \text{modality}_i) + \beta_{11}(\text{Fear}_i \times \text{modality}_i) + \beta_{12}(\text{Sadness}_i \times \text{modality}_i) + \beta_{13}(\text{Surprise}_i \times \text{modality}_i) + \beta_{14}(\text{Love}_i \times \text{modality}_i) + \beta_{15}(\text{Joy}_i \times \text{modality}_i) + \beta_{16}(\text{reviewer expertise}_j) + \beta_{17}(\text{review length}_i) + \varepsilon_{i,j}$ where β_0 is the intercept term, β_i ($i = 1,2\ldots17$) is the regression coefficient, *i* represents the review, *j* represents the reviewer, and $\varepsilon_{i,j}$ is the residual term.

The intelligence rating of the NEV in each review is used as the dependent variable. The DCar website allows for online ratings on the intelligence experience of the NEV users, reflecting users' satisfaction with intelligent service, such as AI voice interaction. A 5-point scale is used, with 1 representing very unintelligent and 5 meaning very intelligent.

The independent variables are the discrete emotion scores contained in each review, which measure the intensity of emotions generated by the AI anthropomorphic interaction. These emotions express the complexity of the AI anthropomorphic interaction experience. Among them, Joy, Love, and Surprise are regarded as potential enablers, and Anger, Sadness, Fear, and Disgust are regarded as potential inhibitors.

The moderating variable is the modality of AI anthropomorphic interaction and is a dummy variable. It takes the value of 0 if only single-modal AI anthropomorphic interaction is supported, and 1 if multimodal AI anthropomorphic interaction is supported.

To avoid potential confounding effects and to make the test more convincing, we select review length [85] as control variables in the review content dimension and choose reviewer expertise [86] as control variables in the reviewer characteristics dimension. However, the regression results show that the effects of all of the above control variables are negligible (with coefficients close to 0) and are, therefore, not specifically analyzed in the subsequent section.

B. Empirical Analysis

Prior to conducting the multiple linear regressions, this study examined the multicollinearity in the regression analysis. The result shows that none of the correlation coefficients between the variables exceed 0.286 and are much less than 0.5, indicating a low correlation between the variables. In addition, the variance inflation factors (VIF) of the variables are all distributed between 1 and 1.1, well below the commonly determined threshold of 10, and the tolerances (1/VIF) are all greater than 0.1. Therefore, the results of the regression analysis are not significantly affected by the problem of multicollinearity. We perform a multiple linear regression on the constructed Model through Stata software and Table II illustrates the results, where M1–M9 stand for Models 1–9, respectively.

Model 1 contains only independent variables and reveals the direct effects. Using Model 1 as the baseline model, moderating variables and interaction terms are added to Models 2–8, respectively, and Model 9 contains the full range of variables.

Model 1 indicates that among the negative emotions, anger and disgust have a significant negative effect on the intelligence rating. Of which, the weakening effect of anger (-1.764) is stronger than that of disgust (-0.952). This may be because the anger generated by the failure of AI anthropomorphic interaction brings stronger dissatisfaction and resistance to intelligence. This coincides with current research findings that emotional intensity affects the rating of online reviews [87]. In contrast, fear and sadness do not have significant effects on intelligence rating; we speculate this is due to the low frequency and probability of the two emotions being triggered by AI anthropomorphic interaction in driving scenarios. Ultimately, H1a and H1b are supported but H1c and H1d are rejected.

Within the positive emotions in Model 1, surprise, love, and joy all exert significant positive effects on intelligence rating. Among them, surprise (1.464) has a stronger contribution than love (0.264) and joy (0.119). This indicates the unexpected surprises brought about by AI anthropomorphic interaction can make users feel more satisfied. Therefore, H2a, H2b, and H2c pass the test.

Models 2 to 5 demonstrate that the modality of AI anthropomorphic interaction can significantly and positively moderate the relationship between negative emotions (except for anger) and intelligent rating. It is worth noting that multimodal AI anthropomorphic interaction fails to alleviate the negative effect of anger on intelligence rating, perhaps because the emotional polarity of anger is too extreme, leading to a strong distrust of intelligence among users [88]. Although sadness does not demonstrate a significant role in the main effect of Model 1, the addition of the interaction modality alters the significance of sadness and mitigates its negative effect on intelligent rating in Model 5. We think this might have something to do with the two-sided nature of the effectiveness of multimodal interaction, where a bad multimodal interaction causes sad emotions, such as disappointment and significantly reduces user satisfaction, while a wonderful multimodal interaction amplifies positive emotions to counteract user dissatisfaction. But before and after the addition of the interaction modality, fear consistently remained

TABLE II
MULTILINEAR REGRESSION RESULTS

Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9
Direct Effects									
Anger	-1.764**	-1.280*	-1.214*	-1.272*	-1.270*	-1.267*	-1.246*	-1.243*	-1.001*
Disgust	-0.952***	-1.053***	-1.186***	-1.051***	-1.051***	-1.052***	-1.040***	-1.047***	-1.091***
Fear	0.12	-0.011	-0.006	-0.338	-0.016	-0.008	-0.013	-0.033	0.095
Sadness	-0.043	-0.454	-0.47	-0.46	-0.875**	-0.455	-0.458	-0.461	-0.461
Surprise	1.464**	0.966	0.93	0.963	0.959	0.23	0.895	0.946	0.784
Love	0.264***	0.315***	0.324***	0.313***	0.316***	0.317***	0.334***	0.198**	0.307***
Joy	0.119***	0.083**	0.105***	0.085**	0.087**	0.085**	0.067	0.103**	0. **
Moderating Effects									
Modality		0.150***	0.129***	0.147***	0.146***	0.148***	0.114***	0.131***	0.113***
Anger×Modality		0.002							-0.036
Disgust×Modality			0.011***						0.004*
Fear×Modality				0.021**					-0.007
Sadness×Modality					0.024***				-0.000
Surprise×Modality						0.041**			0.007
Love×Modality							0.006***		0.004***
Joy×Modality								0.011***	0.002
Observations	22481	22481	22481	22481	22481	22481	22481	22481	22481
R^2	0.010	0.037	0.039	0.037	0.038	0.037	0.040	0.039	0.040
Adj. R^2	0.010	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040

Notes: ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

nonsignificant. So, H3a and H3c fail to pass the test, and H3b and H3d are supported.

Similarly, From Models 6 to 8, the modality of AI anthropomorphic interaction also significantly and positively moderates the relationship between positive emotions (surprise, love, joy) and intelligent rating. This can be attributed to the fact that the fusion of visual and voice facilitates the association, complementarity, and corroboration of emotional cues, eliminating emotional gaps between different modal information [25] and improving the emotional experience of anthropomorphic interaction. Thus, H4a, H4b, and H4c are all supported. And in Model 9 all-variable regression, our conclusions are further verified.

Based on the change in significance of the variables and their contribution to the dependent variable from Models 1 to 9, among the seven discrete emotions, we have identified only disgust and love as being able to maintain a consistently significant influence on intelligent rating. Love and disgust are also considered to be the ultimate enablers and inhibitors. We will explore the components of the two most influential emotions by means of topic clustering in Section IV-C.

C. Topic Clustering Analysis

Regression analyses can only reveal the effect of discrete emotions of NEV users on intelligent rating but they cannot indicate the reasons why this effect occurs. To this end, we decided to apply the STM method to cluster topics for the two most influential emotions, Love (enabler) and Disgust (inhibitor), in order to clarify their constituents and real-time changes. First, we need to construct a specific corpus; then, construct the model and estimate the optimal number of topics; next, filter the topics and assign topic labels; finally, visualize the prevalence of each topic over time.

In order to conduct topic clustering on the two most influential emotions, we first need to construct the corpus of specific emotions. In this study, Mehraliyev et al.'s [89] sensory dimension classifier and Liu et al.'s [61] emotion dimension classifier are referenced for text classification. First, reviews containing disgust and love emotion words are identified in the database A and B, respectively. Then, the sentences containing the disgust and love emotion words in each review are further tracked. Finally, the modalities and emotions are distinguished and the sentences tracked are extracted and linked to form the corpuses. Python executed text classification and tracking. The final set of Modality × Emotion corpuses is constructed, including Single-modal × Love corpus, Multimodal × Love corpus, Multimodal × Disgust corpus, and Single-modal × Disgust corpus. We will perform topic clustering on these corpuses to uncover the constituent elements of the two emotions in the different modalities. The process of text classification and tracking is listed in Fig. 3.

In contrast to other methods in the category of topic modeling methods, STM is unique in allowing the interpretation of topic popularity in conjunction with metadata from data samples [90]. For example, in our study case, metadata refers to the information related to online reviews, such as review created time, vehicle purchased time, review votes, and additional reviews. Therefore, we define two covariates, namely "review votes" and "vehicle purchased time." We apply STM to Modality × Emotion Corpuses. In other words, we consider each corpus as a document. In STM, a document is defined as a mixture of topics, which means that a document consists of multiple topics [91].

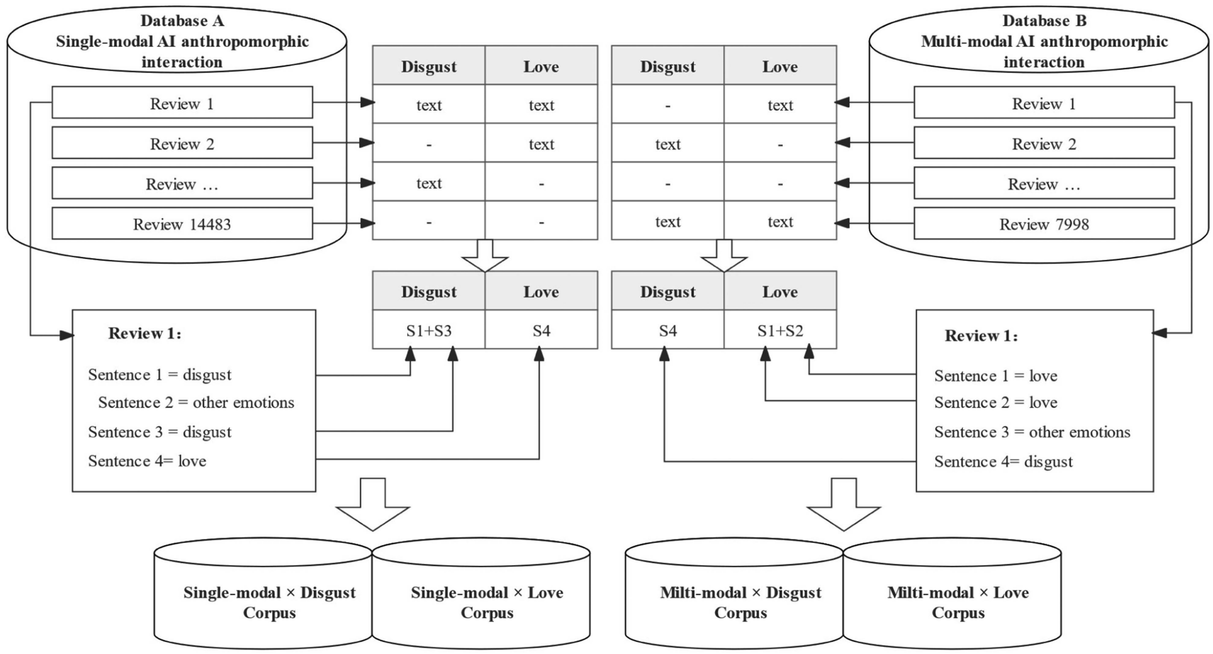


Fig. 3. Visualization process of creating the Modality × Emotion Corpuses.

Thus, a corpus is a mixture of topics, and a topic is a mixture of words, where each word has a probability of belonging to a particular topic. The *stm* package [92] in R was used to build the model. The prevalence function is shown as

$$\text{prevalence} \sim \text{reviewvotes} + s\,(\text{vehicle purchased time})$$

where s is a smoothing function of vehicle purchase time, and review votes is a topic popularity covariate indicating how many useful votes a review received.

The number of topics K is one of the most crucial user-specified parameters in STM, which helps in the substantive interpretation of the modeling results [93]. Using the function SearchK from the stm and furr packages in R, we evaluate models trained in parallel on sparse matrices with a range of different K values, i.e., from 3 to 15. The K value is not determined to maximize the fit [94] but should depend on the intrinsic nature of the corpus [95]. The textual data contained in the four corpuses are relatively homogeneous, so the K value should not be large. By comparing the variation of the fit metrics in the y-axis (Held-out Likelihood, Semantic Coherence, Residuals, and Lower Bound) corresponding to the different K values in the x-axis shown in Fig. 4, we identify the optimal number of topics (K values) for each corpus refer to Han et al.'s [24] justification, in which $K = 7$ is selected for Single-modal × Love Corpus; $K = 9$ is selected for Multimodal × Love Corpus; $K = 9$ is selected for Multimodal × Disgust Corpus; and $K = 5$ is selected for Single-modal × Disgust Corpus.

Based on the number of topics determined for corpuses, we execute the STM four times separately. The STM outputs a list of top words and a large number of representative review examples for each of the topics in the four corpuses. These words occur most frequently in the topic and least frequently in other topics, which can distinguish one topic from the others. Two social science researchers in the field of digital marketing and anthropomorphic brand management are recruited to determine the allocation of topic labels by analyzing the top words and representative review examples of the topics.

To explore the components of emotions generated by AI anthropomorphic interaction in a more focused manner, the topics generated in each corpus that are irrelevant to AI anthropomorphic interaction are eliminated in the process of topic label assignment. After filtering, Single-modal × Love Corpus retains four of the seven topics; Multimodal × Love Corpus keeps four of the nine topics; Multimodal × Disgust Corpus maintains four of the nine topics; and Single-modal × Disgust Corpus holds two of the five topics. In Table III, each corpus is considered a quadrant, and we report the labels, ratios, and top ten words of topics retained in each quadrant.

To display the changes of the prevalence of each topic in the Modality × Emotion Quadrants over time, we plot the prevalence of each topic in Fig. 5 as a smoothed function of the time of car purchase, where the x-axis is the time series and the y-axis is the proportion of topics, measuring prevalence. It facilitates us to explore the dynamic trends of the constituent elements of disgust and love emotions over time, so as to explore the opportunities of technological innovation breakthroughs in the trends.

V. Results and Discussion

A. Topics in the Single-Modal × Love Quadrant

The four topics in the Single-Modal × Love Quadrant reveal the components of the love emotion, providing crucial clues to explore the popularity of voice interaction technology.

Based on the top words of Topic 7 (The sense of high-tech enhanced by design) and the representative review examples listed,

TABLE III
TOPIC LABELING IN THE MODALITY × EMOTION QUADRANTS

Rank	Topic label	Ratio	Top words
Single-modal×Love Quadrant			
Rank	**Topic label**	**Ratio**	**Top words**
Topic7	The sense of high-tech enhanced by design	21.54%	内饰(interior),语音(voice), 设计(design), 科技(technology), 旋转(rotation), 中控(central control), 屏幕(screen),车内(in-car), 大屏(large screen), 控屏(control screen)
Topic2	The sense of security from artificial intelligence	17.01%	功能(function), 智能(intelligence), 操作(operation), 方便(convenience), 控制 (control), 操控 (dominate), 识别(recognition), 开车(driving), 交互(interaction),小迪(Xiaodi)
Topic3	Better convenient and comfortable experience	15.25%	语音(voice), 空调(air conditioning), 控制(control), 座椅(seats), 舒适(comfortable), 方便(convenient), 打开(open),不错(nice), 调节(adjust), 车内(in-car)
Topic1	Better user-assist and entertainment experience	7.40%	导航(navigation),手机(mobile phone),智能化(intelligence),系统(system),音乐(music), app, 联网(networking),听歌(listening to music), 车内(in-car), 喜欢(like)
Multi-modal×Love Quadrant			
Rank	**Topic label**	**Ratio**	**Top words**
Topic7	Perceived warmth in AI anthropomorphic interaction	13.35%	喜欢(love),内饰(interio),nomi,特别(special),氛围(atmosphere), 孩子(children), 感觉(feeling), 颜色(color), 互动(interaction), 小朋友(children)
Topic5	Perceived intelligence in AI anthropomorphic interaction	13.32%	语音(voice),智能(intelligence),系统(system),交互(interaction), 驾驶(driving),智能化(intelligence),小鹏(Xiaopeng),人机交互(human-computer interaction),不错(good),车机(car machine)
Topic3	Better convenient and safe driving experience	12.75%	功能 (function), 操作 (operation), 方便 (convenience), 控制(control),识别(identification),导航(navigation), 小 p(Xiaopi), 完成(completion), 助手(assistant), 操控(control)
Topic1	Virtual companion and customized entertainment experience	10.61%	nomi,可爱(cute), 机器人(robot), 开车(driving),感觉(feeling), 灵魂(soul), 表情(expression), 选配(apolegamy), 选装(option), 女王(queen)
Multi-modal×Disgust Quadrant			
Rank	**Topic label**	**Ratio**	**Top words**
Topic2	Frequent network interruption and untimely system upgrades	15.7%	车机(vehicle machine), 系统(system), 升级(upgrade), 缺点(disadvantag), 问题 (proble), 手机 (mobile phone), 解决(solution),启动(startup), 软件(software), 不是(not)
Topic1	Slow response and awkward error recognition	12.42%	nomi, 反应迟钝(unresponsive), 蔚来(NIO), 开车(driving), 机器人(robot), 缺点(shortcomings), 互动(interaction), 上车(getting on),不用(not using), 路上(on the road)
Topic6	Low rate of AI voice recognition due to noise pollution	10.58%	空调(air conditioner), 识别率(recognition rate), 乘坐(ride), 性价比(cost performance), 噪音(noise), 语音(voice), 空气(air), 储物(storage), 控制(control), 提速(speed increase)
Topic5	Low level of voice customization and dialect recognition	7.13%	语音 (voice), 识别 (recognition), 智能 (intelligence), 反应(reaction), 人机 (human-computer), 不到 (less than), 交互(interaction),目前(now),智能化(intelligence),不是(not)
Single-modal×Disgust Quadrant			
Rank	**Topic label**	**Ratio**	**Top words**
Topic2	Vehicle-engine interaction system with delayed response	26.40%	车机(vehicle machine), 系统(system), 缺点(disadvantage), 不满意 (dissatisfied), 比亚迪 (BYD), 现在 (present), 问题(problem), 使用(use), 升级(upgrade), 时间(time)
Topic3	High costs of AI interaction	7.42%	不够用(not enough), 流量(traffic), 手机(mobile phone), 导航(navigation), 远程 (remote), app, 启动 (start-up), 功能(function), 车机(vehicle machine), 联网(networking)

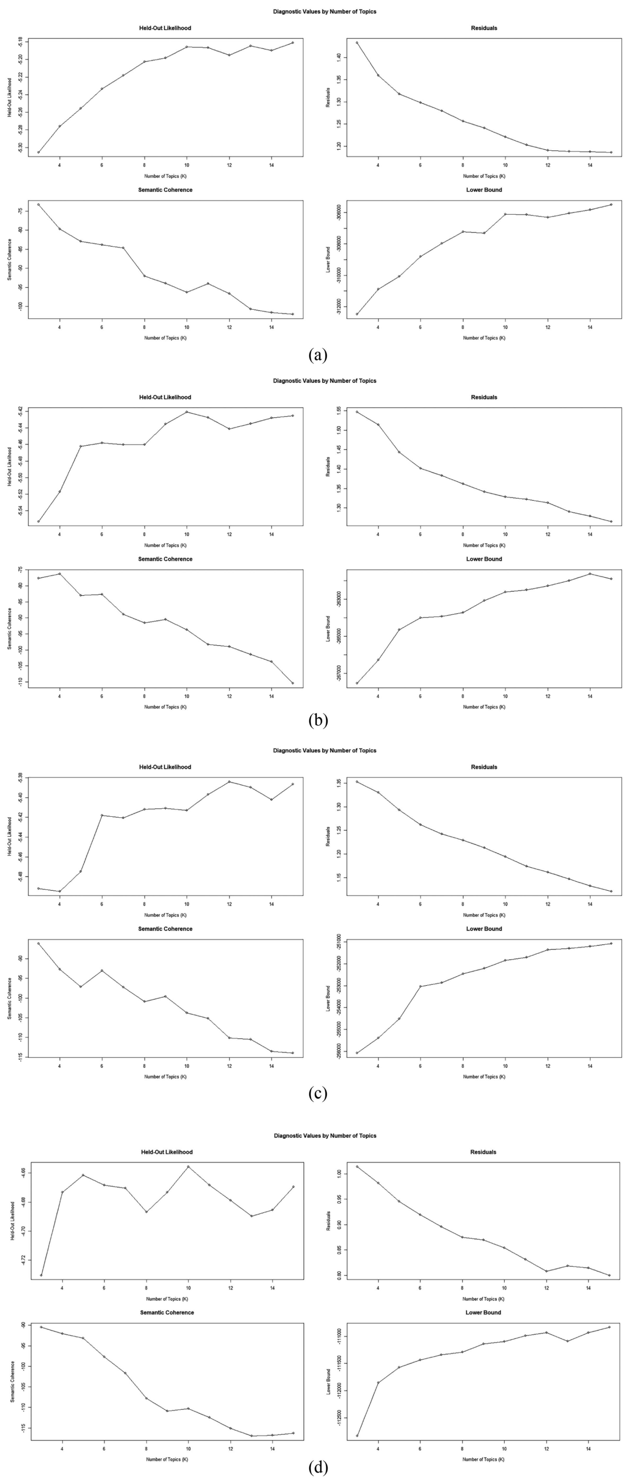


Fig. 4. (a) *K* of Single-modal × Love Corpus. (b) *K* of Multimodal × Love Corpus. (c) *K* of Multimodal × Disgust Corpus. (d) *K* of Single-modal × Disgust Corpus.

it is easy to see that as a carrier of AI voice interaction in the vehicle, the technological design elements of the center control screen (larger screen size, rotatable screen, etc.) can enhance the users' high-tech perception of AI voice interaction, resulting in a more enjoyable AI voice interaction experience. Elshan et al. [96] verified multiple relationships between design elements and user acceptance of AI interaction, which supports our findings. Although Topic 7 currently has the highest proportion (21.54%) in this quadrant, it is obvious from Fig. 5(a) that the attention to a high-tech design is rapidly declining, perhaps because with the current surge in domestic NEV brands, technology-enabled center control screen designs have become more common and users are no longer attracted by the exterior design, with the focus on intelligence gradually returning to the functionality and effectiveness of the AI voice interaction in vehicles.

> [Topic 7 The sense of high-tech enhanced by design] "Interior: The design of the interior felt very high-tech when I first saw it, with a 15.6-inch rotatable center screen that is bursting with technology, smooth to use and strong in voice recognition. I think this is a great design."

Based on representative examples of reviews, Topic 2 (The sense of security from artificial intelligence) expresses the users' recognition of the sense of security brought by AI voice interaction. Existing research has revealed that in-vehicle AI voice interaction not only reduces safety risks by reducing driving distractions [97] but also reduces driving hazards by alleviating users' negative emotions through emotive voice communication [98]. In Fig. 5(a), the sense of security brought about by the NEV's powerful AI voice interaction function is generally on an upward trend but has been fluctuating since 2021 perhaps due to the high breakdown frequency of vehicle-machine systems over the past two years, causing users to be concerned about the intelligence level of NEV.

> [Topic 2: The sense of security from artificial intelligence] "Voice control: Voice control is very sensitive, a "Hello Di" takes care of everything, completely freeing up the user's hands and eliminating the need for distractions to affect safety."

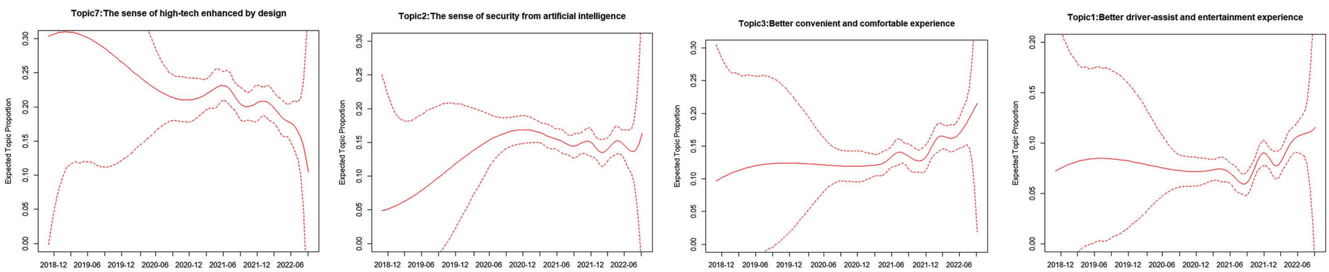
Topic 3 (Better convenient and comfortable experience) and Topic 1 (Better driver-assist and entertainment experience) focus on expressing drivers' enjoyment of the better convenience, comfort, assisted driving, and entertainment experience brought about by AI voice interaction. As Fig. 5(a) shows, concerns about Topic 3 and Topic 1 will continue to trend upward, even if the safety problems of the last two years have made some negative impact on the assisted driving experience.

B. Topics in the Multimodal × Love Quadrant

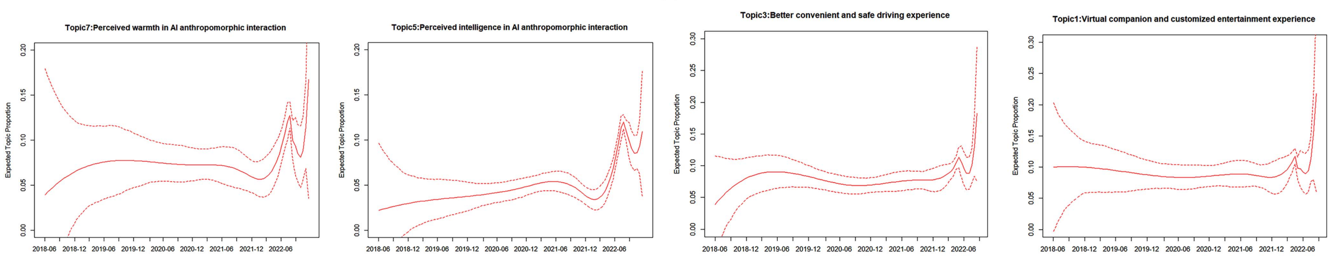
The four topics in the Multimodal × Love Quadrant uncover the constituent elements of love emotion. We compare the topics of Multimodal × Love Quadrant and Single-Modal × Love Quadrant to analyze the influence that the incorporation of anthropomorphic visual interaction, based on voice, exerts on the love emotion. An attempt is made to uncover the changes in the components of the love emotion due to modality differences of the AI anthropomorphic interaction.

Based on the output top words and representative review examples, Topic 7 is labeled as Perceived warmth in AI anthropomorphic interaction. It is the most obvious variation in the components of the love emotion following the addition of anthropomorphic visual interaction. Warmth is considered to be a feeling of closeness and friendliness [10], a core component of

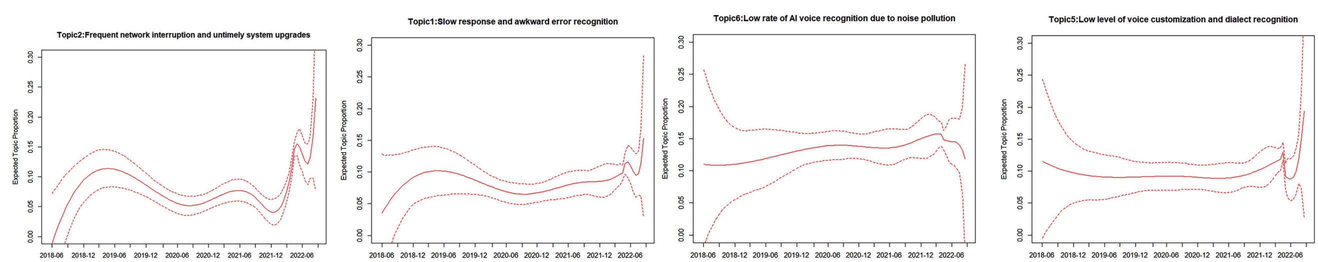
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



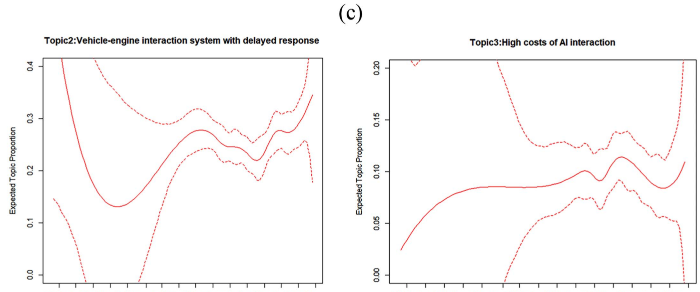
(a)



(b)



(c)



(d)

Fig. 5. (a) Topic prevalence over time in the Single-modal × Love Quadrant. (b) Topic prevalence over time in the Multimodal × Love Quadrant. (c) Topic prevalence over time in the Multimodal × Disgust Quadrant. (d) Topic prevalence over time in the Single-modal × Disgust Quadrant.

the human essence [99]. Visual avatar such as "Nomi" gives the disembodied AI voice agent in the vehicle an embodied human-like figure, enabling the drivers and passengers (especially kids) to feel the warmth as friends during the AI interaction, which helps to close the psychological distance between human and AI agents [1]. During AI anthropomorphic interaction, Nomi conveys a fusion of visual and voice, which amplifies the emotional power of warmth and reinforces the love emotion. The focus on perceived warmth in Topic 7 has grown exponentially since the end of 2021, when multimodal AI anthropomorphic interaction began to be used as an influential emotional marketing strategy in several NEV brands.

> [Topic 7: Perceived warmth in AI anthropomorphic interaction] "In terms of human–car interaction, I feel like I have an extra friend, especially "Nomi." The first thing I feel when I get in the car every day is the "Good morning" from Nomi, which makes me feel very warm. With Nomi it's not just a car, it's a friend with a warm heart."

Similarly, the integration of voice and visual modalities enhances the level of intelligence of the AI anthropomorphic interaction. Due to the addition of anthropomorphic visual interaction, the AI interaction is improved and the drivers receive more visual feedback, especially when active interaction occurs, for example, when Nomi turns into a greeting emoji and says "Hello" when the driver gets into the car. Therefore, Topic 5 is labeled as Perceived intelligence in AI anthropomorphic interaction, and its trend over time and the reasons for it are similar to Topic 7 due to the widespread use of multimodal AI anthropomorphic interaction techniques. Also, the topic focused on the Single-Modal × Love Quadrant receives the same attention in the Multimodal × Love Quadrant. This suggests that the modality differences do not have a significant influence on the perceived convenience, perceived usefulness, and safe driving focus in Topic 3 (Better convenient and safe driving experience).

It is notable that we have discovered interesting phenomena again in Topic 1 (Virtual companion and customized entertainment experience). The visual AI voice assistant such as Nomi plays the role of a virtual companion that functions as a safety reminder and mood regulator when driving alone [98], alleviating drivers' isolation [100]. Moreover, richer customized entertainment experiences are obtained by changing the visual anthropomorphic figures of the AI voice assistant. According to Fig. 5(b), Topic 1 has shown a surge since 2022, which may be related to the recent customization marketing strategies adopted by the NEV companies, such as the Nomi option and exterior products.

> [Topic 1: Virtual companion and customized entertainment experience] "Nomi I particularly like, it enhances the intelligence of the whole car, I sometimes talk to Nomi when I am driving alone, Nomi's expressions change to give a feeling of companionship, naughty Nomi makes me feel like I have a friend with me all the time, no more boredom when I am alone."

C. Topics in the Multimodal × Disgust Quadrant

The Multimodal × Disgust Quadrant also contains four definitively labeled topics that reveal the components of disgust emotion. We compare the topics of the Multimodal × Disgust Quadrant and Multimodal × Love Quadrant in an attempt to uncover the reasons for the emotional differences of the users confronted with multimodal AI anthropomorphic interaction.

We first analyze the four topics in this quadrant overall and find that all of them are related to technical support. Topic 2 suggests that the reasons for drivers' disgust are frequent network interruptions and untimely system upgrades. The former is perhaps owing to the inability to achieve network coverage in remote areas or interference with signal reception. Without a network, multimodal AI anthropomorphic interaction will not be of any use. The latter is due to cost considerations, the vehicle-machine systems will not be as up-to-date as it was in the past; this means that drivers will have to endure a system with bugs for a certain period of time. Topic 1 indicates that drivers suffer from slow response and awkward error recognition. This may be attributed to the lack of accuracy and sensitivity in recognition, or it may have been suffered from the network and system breakdown of Topic 1. Topic 6 argues that drivers are dissatisfied with the low rate of voice recognition due to noise pollution, which implies that poor sound insulation in NEV will be a hidden hazard affecting the AI interaction experience. Topic 5 points out the issue of the low level of voice customization and dialect recognition, which denotes that there is still space for improvement in pervasiveness and individualization.

This explains why the same multimodal AI anthropomorphic interaction produces differences in emotions. NEV users enjoy the benefits and experience of multimodal AI anthropomorphic interaction, but the lack of technical support causes the interaction to fail, resulting in a poor experience and disgust emotion. According to Fig. 5(c), only Topic 6 has recently shown a decreasing trend, which means that NEV companies have improved in isolating noise interference from AI voice interaction. However, the remaining topics all show an overall upward trend, which implies an intense conflict between the strong demand for AI voice interaction and the unstable AI technology support, resulting in a sharp increase in the attention to the corresponding topics.

D. Topics in the Single-Modal × Disgust Quadrant

Finally, Topic 2 (Vehicle–engine interaction system with delayed response) in the Single-Modal × Disgust Quadrant is still concerned with the technical issues of AI interaction, showing that the modality differences do not shake the focus on technical issues in the disgust emotion. In addition, Topic 3 (High costs of AI interaction) also raises the issue of the cost of AI interaction, mainly the cost of network traffic. This is probably because NEV companies implemented the policy of giving away Internet traffic during the initial launch of new vehicles, but as time spent, users have to cover the extra-budgetary costs of Internet traffic in order to maintain their usage habits.

VI. Conclusion and Implications

A. Conclusion

Based on a dual-factor perspective, applying the DUTIR Emotion Lexicon and STM, this study aims to find the enablers and inhibitors of NEV users' satisfaction from the discrete emotions triggered by in-vehicle AI anthropomorphic interaction and to address the potential emerging technological resistances in NEV domain. New insights are provided into the improvement directions and development opportunities for anthropomorphic AI assistants based on the constituent elements of love and disgust emotion and their real-time variations. Our findings cater to the intelligent strategy of NEV in Industry 5.0 and have significant implications for the NEV industry and academics, as well as for other industries facing resistance to emerging technologies.

B. Theoretical Implications

The theoretical implications of this study are mainly reflected in the following aspects. First, this study selects the DCar website, a text data source that has not yet received much attention, and trains a Word2vec machine learning model to filter valid text data with high accuracy, guaranteeing the high quality of the data used in this study. Discrete emotions related to AI anthropomorphic interaction in NEV reviews are extracted and quantified through the DUTIR Emotion Lexicon, which provides feedback on the complex emotional experiences of NEV users. Our study differs from previous binary sentiment analysis (positive and negative) in filling the gap of discrete emotions research in the domain of NEV and AI interaction, expanding the research scenario of the dual-factor theory, and is a useful attempt to apply emotion analysis techniques to user experience management.

Second, the feasibility and rationality of utilizing the user's emotion feedback to improve the satisfaction of AI interaction have been verified through multiple linear regression and STM. The findings innovatively illustrate the underlying mechanism, whereby anthropomorphic features and multimodal can improve the effectiveness of AI interaction, providing a theoretical basis to improve AI techniques and cope with the emerging technology resistance based on users' discrete emotions.

Finally, we demonstrate the superiority and effectiveness of STM. We innovatively construct a set of multidimensional corpuses of Modality $\times$ Emotion and further explore the components of love and disgust emotion. The effects of modality differences on the components of emotion are analyzed under the same emotion, and the causes of emotion differences are identified under the same modality. Notably, we fill a research gap on how to unlock the value of unstructured "real-time" data in the NEV industry.

C. Managerial Implications

This study also provides some managerial implications. First, we provide a technical approach based on UGC identification and real-time attention to users' emotions, which is an inspiration for NEV companies to identify the users' AI interaction experience and maintain continuous product improvement. Second, in terms of multimodal and anthropomorphic perspectives, it reveals the reasons for user dissatisfaction from the negative emotions and clarifies the reasons for user satisfaction from the positive emotions, which provides directional guidance for product acceptance enhancement and application promotion and is able to effectively avoid potential technological resistance due to unsatisfactory experience. Finally, operational improvement guidelines are provided for third-party websites. Advanced emotion analysis algorithms should be used to recognize and label the types of emotions in NEV reviews, guiding NEV companies to access the platforms and leverage the advantages of third-party websites to pay attention to users' real-time emotional variations and respond agilely to their intelligent concerns.

D. Limitations

This study still has certain limitations requiring improvement. First, only Chinese reviews from a single car website are selected, and no reviews in other languages, such as English are collected, resulting in a too-homogeneous sample source and sample data, which may bias the results of the emotion analysis. In future studies, we consider collecting review data from multiple third-party websites in multiple languages. Second, only a single emotion lexicon is selected and the accuracy of the emotion analysis results has not been critically evaluated. Future research plans to use multiple emotion analysis techniques, including machine learning and hybrid techniques. Third, the small number of NEV brands supporting multimodal AI anthropomorphic interaction results in a too short time series of data, which negatively affects the topic prevalence of STM.

REFERENCES

[1] X. Li and Y. Sung, "Anthropomorphism brings us closer: The mediating role of psychological distance in User-AI assistant interactions," *Comput. Hum. Behav.*, vol. 118, May 2021, Art. no. 106680, doi: 10.1016/j.chb.2021.106680.

[2] J. Balakrishnan and Y. K. Dwivedi, "Conversational commerce: Entering the next stage of AI-powered digital assistants," *Ann. Oper. Res.*, 2021, doi: 10.1007/s10479-021-04049-5.

[3] T. Fernandes and E. Oliveira, "Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption," *J. Bus. Res.*, vol. 122, pp. 180–191, 2021, doi: 10.1016/j.jbusres.2020.08.058.

[4] D. H. Kim, H. S. Lee, and M. J. Chung, "Biologically inspired models and hardware for emotive facial expressions," in *Proc. IEEE 14th Workshop Robot Hum. Interactive Commun.*, 2005, pp. 679–685.

[5] S. Schanke, G. Burtch, and G. Ray, "Estimating the impact of 'humanizing' customer service chatbots," *Inf. Syst. Res.*, vol. 32, no. 3, pp. 736–751, Sep. 2021, doi: 10.1287/isre.2021.1015.

[6] Q. Xin, M. Alazab, R. G. Crespo, and C. E. Montenegro-Marin, "AI-based quality of service optimization for multimedia transmission on Internet of Vehicles (IoV) systems," *Sustain. Energy Technol. Assessments*, vol. 52, Aug. 2022, Art. no. 102055, doi: 10.1016/j.seta.2022.102055.

[7] C. Llopis-Albert, F. Rubio, and F. Valero, "Impact of digital transformation on the automotive industry," *Technol. Forecasting Social Change*, vol. 162, Jan. 2021, Art. no. 120343, doi: 10.1016/j.techfore.2020.120343.

[8] P. Grover, A. K. Kar, and Y. K. Dwivedi, "Understanding artificial intelligence adoption in operations management: Insights from the review of academic literature and social media discussions," *Ann. Oper. Res.*, vol. 308, no. 1/2, pp. 177–213, Jan. 2022, doi: 10.1007/s10479-020-03683-9.

[9] A. K. Rathore and P. V. Ilavarasan, "Pre- and post-launch emotions in new product development: Insights from Twitter analytics of three products," *Int. J. Inf. Manage.*, vol. 50, pp. 111–127, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.05.015.

[10] S. T. Fiske, A. J. C. Cuddy, and P. Glick, "Universal dimensions of social cognition: Warmth and competence," *Trends Cogn. Sci.*, vol. 11, no. 2, pp. 77–83, Feb. 2007, doi: 10.1016/j.tics.2006.11.005.

[11] C. M. Chiu, C. S. Chiu, and H. C. Chang, "Examining the integrated influence of fairness and quality on learners' satisfaction and web-based learning continuance intention," *Inf. Syst. J.*, vol. 17, no. 3, pp. 271–287, Jul. 2007, doi: 10.1111/j.1365-2575.2007.00238.x.

[12] S. Han and H. Yang, "Understanding adoption of intelligent personal assistants: A parasocial relationship perspective," *Ind. Manage. Data Syst.*, vol. 118, no. 3, pp. 618–636, 2018, doi: 10.1108/imds-05-2017-0214.

[13] D. Hong and C. H. Cho, "Factors affecting innovation resistance of smartphone AI voice assistants," *Int. J. Hum.-Comput. Interact.*, vol. 39, pp. 2557–2572, 2023, doi: 10.1080/10447318.2022.2080899.

[14] R. T. Cenfetelli, "Inhibitors and enablers as dual factor concepts in technology usage," *J. Assoc. Inf. Syst.*, vol. 5, no. 11, 2004, Art. no. 16, doi: 10.17705/1jais.00059.

[15] S. Malodia, N. Islam, P. Kaur, and A. Dhir, "Why do people use artificial intelligence (AI)-enabled voice assistants," *IEEE Trans. Eng. Manage.*, to be published, doi: 10.1109/tem.2021.3117884.

[16] P. Kowalczuk, "Consumer acceptance of smart speakers: A mixed methods approach," *J. Res. Interactive Marketing*, vol. 12, no. 4, pp. 418–431, 2018, doi: 10.1108/jrim-01-2018-0022.

[17] D. Sledgianowski and S. Kulviwat, "Using social network sites: The effects of playfulness, critical mass and trust in a hedonic context," *J. Comput. Inf. Syst.*, vol. 49, no. 4, pp. 74–83, 2009.

[18] T. R. Choi and M. E. Drumwright, "'OK, Google, why do I use you?' Motivations, post-consumption evaluations, and perceptions of voice AI assistants," *Telematics Inform.*, vol. 62, Sep. 2021, Art. no. 101628, doi: 10.1016/j.tele.2021.101628.

[19] C. Pelau, D.-C. Dabija, and I. Ene, "What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry," *Comput. Hum. Behav.*, vol. 122, Sep. 2021, Art. no. 106855, doi: 10.1016/j.chb.2021.106855.

[20] H. Y. Yang, H. Xu, Y. Zhang, Y. Liang, and T. Lyu, "Exploring the effect of humor in robot failure," *Ann. Tourism Res.*, vol. 95, Jul. 2022, Art. no. 103425, doi: 10.1016/j.annals.2022.103425.

[21] M. Sykora, S. Elayan, I. R. Hodgkinson, T. W. Jackson, and A. West, "The power of emotions: Leveraging user generated content for customer experience management," *J. Bus. Res.*, vol. 144, pp. 997–1006, May 2022, doi: 10.1016/j.jbusres.2022.02.048.

[22] A. Orea-Giner, L. Fuentes-Moraleda, and T. Villace-Molinero, "Does the implementation of robots in hotels influence the overall TripAdvisor rating? A text mining analysis from the Industry 5.0 approach," *Tourism Manage.*, vol. 93, Dec. 2022, Art. no. 104586, doi: 10.1016/j.tourman.2022.104586.

[23] M. J. Chen and C. K. Farn, "Examining the influence of emotional expressions in online consumer reviews on perceived helpfulness," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102266, doi: 10.1016/j.ipm.2020.102266.

[24] C. Han, M. Yang, and A. Piterou, "Do news media and citizens have the same agenda on COVID-19? An empirical comparison of Twitter posts," *Technol. Forecasting Social Change*, vol. 169, 2021, Art. no. 120849, doi: 10.1016/j.techfore.2021.120849.

[25] H. Xie, J. Mao, and G. Li, "Sentiment classification of image-text information with multi-layer semantic fusion," *Data Anal. Knowl. Discov.*, vol. 5, no. 6, pp. 103–114, 2021, doi: 10.11925/infotech.2096-3467.2020.1159.

[26] M. D. Rocklage, D. D. Rucker, and L. F. Nordgren, "Mass-scale emotionality reveals human behaviour and marketplace success," *Nature Hum. Behav.*, vol. 5, no. 10, pp. 1323–1U65, Oct. 2021, doi: 10.1038/s41562-021-01098-5.

[27] J. Martinez-Miranda and A. Aldea, "Emotions in human and artificial intelligence," *Comput. Hum. Behav.*, vol. 21, no. 2, pp. 323–341, Mar. 2005, doi: 10.1016/j.chb.2004.02.010.

[28] V. Kumar, A. Dixit, R. G. Javalgi, and M. Dass, "Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing," *J. Acad. Marketing Sci.*, vol. 44, no. 1, pp. 24–45, Jan. 2016, doi: 10.1007/s11747-015-0426-9.

[29] T. M. Brill, L. Munoz, and R. J. Miller, "Siri, Alexa, and other digital assistants: A study of customer satisfaction with artificial intelligence applications," *J. Marketing Manage.*, vol. 35, no. 15/16, pp. 1401–1436, Oct. 2019, doi: 10.1080/0267257x.2019.1687571.

[30] G. McLean and K. Osei-Frimpong, "Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants," *Comput. Hum. Behav.*, vol. 99, pp. 28–37, Oct. 2019, doi: 10.1016/j.chb.2019.05.009.

[31] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: A three-factor theory of anthropomorphism," *Psychol. Rev.*, vol. 114, no. 4, pp. 864–886, Oct. 2007, doi: 10.1037/0033-295x.114.4.864.

[32] Y. Ye et al., "Speech interface reformulations and voice assistant personification preferences of children and parents," *Int. J. Child-Comput. Interaction*, vol. 21, pp. 77–88, Sep. 2019, doi: 10.1016/j.ijcci.2019.04.005.

[33] H. Tankovska, "Number of digital voice assistants in use worldwide 2019–2024," 2020. [Online]. Available: https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/

[34] B. Kinsella, "54% of US adults have used voice commands and 24% use them daily, but smart speaker adoption slows – Edison and NPR report," 2020. [Online]. Available: https://voicebot.ai/2020/01/08/54-of-us-adults-have-used-voice-commands-and-24-use-them-daily-but-smartspeaker-adoption-slows-edison-and-npr-report/

[35] CGS, "2019 CGS customer service chatbots & channels survey," 2019. [Online]. Available: https://www.cgsinc.com/en/resources/2019-CGS-Customer-Service-Chatbots-Channels-Survey

[36] Y. K. Dwivedi et al., "Setting the future of digital and social media marketing research: Perspectives and research propositions," *Int. J. Inf. Manage.*, vol. 59, Aug. 2021, Art. no. 102168, doi: 10.1016/j.ijinfomgt.2020.102168.

[37] J. Wirtz et al., "Brave new world: Service robots in the frontline," *J. Service Manage.*, vol. 29, no. 5, pp. 907–931, 2018, doi: 10.1108/josm-04-2018-0119.

[38] J. van Doorn et al., "Domo Arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' Service experiences," *J. Service Res.*, vol. 20, no. 1, pp. 43–58, Feb. 2017, doi: 10.1177/1094670516679272.

[39] D. Marikyan, S. Papagiannidis, O. F. Rana, R. Ranjan, and G. Morgan, ""Alexa, let's talk about my productivity": The impact of digital assistants on work productivity," *J. Bus. Res.*, vol. 142, pp. 572–584, Mar. 2022, doi: 10.1016/j.jbusres.2022.01.015.

[40] B. L. Rich, J. A. Lepine, and E. R. Crawford, "Job engagement: Antecedents and effects on job performance," *Acad. Manage. J.*, vol. 53, no. 3, pp. 617–635, Jun. 2010, doi: 10.5465/amj.2010.51468988.

[41] A. Bhattacherjee, "Understanding information systems continuance: An expectation-confirmation model," *MIS Quart.*, vol. 25, no. 3, pp. 351–370, Sep. 2001, doi: 10.2307/3250921.

[42] Y. G. Xie, S. P. Zhao, P. Y. Zhou, and C. Y. Liang, "Understanding continued use intention of AI assistants," *J. Comput. Inf. Syst.*, 2023, vol. 63, no. 6, pp. 1424–1437, 2023, doi: 10.1080/08874417.2023.2167134.

[43] B. Mausner and B. B. Snyderman, *The Motivation to Work*. Piscataway, NJ, USA: Transaction, 1993.

[44] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quart.*, vol. 13, no. 3, pp. 319–340, Sep. 1989, doi: 10.2307/249008.

[45] S. K. Sharma and A. Mishra, "Enablers and inhibitors of mobile payments in rural India: A dual-factor theory perspective," *Inf. Syst. Front.*, 2022, Art. no. 17, doi: 10.1007/s10796-022-10355-5.

[46] M. Talwar, S. Talwar, P. Kaur, A. Islam, and A. Dhir, "Positive and negative word of mouth (WOM) are not necessarily opposites: A reappraisal using the dual factor theory," *J. Retailing Consum. Serv.*, vol. 63, Nov. 2021, Art. no. 102396, doi: 10.1016/j.jretconser.2020.102396.

[47] J. Balakrishnan, Y. K. Dwivedi, L. Hughes, and F. Boy, "Enablers and inhibitors of AI-powered voice assistants: A dual-factor approach by integrating the status quo bias and technology acceptance model," *Inf. Syst. Front.*, 2021, Art. no. 22, doi: 10.1007/s10796-021-10203-y.

[48] C. Shaw and J. Ivens, *Building Great Customer Experiences*. Berlin, Germany: Springer, 2002.

[49] J. K. Fawcett, "Emotion marketing: The Hallmark way of winning customers for life," *J. Consum. Marketing*, vol. 19, no. 1, pp. 74–82, 2002, doi: 10.1108/jcm.2002.19.1.74.4.

[50] L. Becker and E. Jaakkola, "Customer experience: Fundamental premises and implications for research," *J. Acad. Marketing Sci.*, vol. 48, no. 4, pp. 630–648, Jul. 2020, doi: doi: 10.1007/s11747-019-00718-x.

[51] R. P. Bagozzi, M. Gopinath, and P. U. Nyer, "The role of emotions in marketing," *J. Acad. Marketing Sci.*, vol. 27, no. 2, pp. 184–206, 1999, doi: 10.1177/0092070399272005.

[52] A. Orea-Giner, L. Fuentes-Moraleda, and T. Villace-Molinero, "Does the implementation of robots in hotels influence the overall TripAdvisor rating? A text mining analysis from the Industry 5.0 approach," *Tourism Manage.*, vol. 93, Dec. 2022, Art. no. 104586, doi: 10.1016/j.tourman.2022.104586.

[53] R. P. Bagozzi, M. K. Brady, and M. H. Huang, "AI service and emotion," *J. Service Res.*, vol. 25, no. 4, pp. 499–504, Nov. 2022, doi: 10.1177/10946705221118579.

[54] L. Gkinko and A. Elbanna, "Hope, tolerance and empathy: Employees' emotions when using an AI-enabled chatbot in a digitalised workplace," *Inf. Technol. People*, vol. 35, no. 6, pp. 1714–1743, Dec. 2022, doi: 10.1108/itp-04-2021-0328.

[55] X. Y. Lv, Y. Liu, J. J. Luo, Y. Q. Liu, and C. X. Li, "Does a cute artificial intelligence assistant soften the blow? The impact of cuteness on customer tolerance of assistant service failure," *Ann. Tourism Res.*, vol. 87, Mar. 2021, Art. no. 103114, doi: 10.1016/j.annals.2020.103114.

[56] R. V. Kubler, A. Colicev, and K. H. Pauwels, "Social Media's impact on the consumer mindset: When to use which sentiment extraction tool?," *J. Interactive Marketing*, vol. 50, pp. 136–155, May 2020, doi: 10.1016/j.intmar.2019.08.001.

[57] R. Filieri, Z. B. Lin, Y. L. Li, X. Q. Lu, and X. W. Yang, "Customer emotions in service robot encounters: A hybrid machine-Human intelligence approach," *J. Service Res.*, vol. 25, no. 4, Nov. 2022, Art. no. 10946705221103937, doi: 10.1177/10946705221103937.

[58] P. Ekman, "An argument for basic emotions," *Cogn. Emotion*, vol. 6, no. 3/4, pp. 169–200, May–Jul. 1992, doi: 10.1080/02699939208411068.

[59] L. Zhu, Y. Lin, and M. M. Cheng, "Sentiment and guest satisfaction with peer-to-peer accommodation: When are online ratings more trustworthy?," *Int. J. Hospitality Manage.*, vol. 86, Apr. 2020, Art. no. 102369, doi: 10.1016/j.ijhm.2019.102369.

[60] S. M. Mohammad, *Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States From Text (Emotion Measurement)*. Cambridge, U.K.: Woodhead, 2016, pp. 201–237.

[61] J. Liu, Y. Yu, F. Mehraliyev, S. Hu, and J. Chen, "What affects the online ratings of restaurant consumers: A research perspective on text-mining big data analysis," *Int. J. Contemporary Hospitality Manage.*, vol. 34, pp. 3607–3633, 2022, doi: 10.1109/ijchm-06-2021-0749.

[62] K. Letheren, J. Jetten, J. Roberts, and J. Donovan, "Robots should be seen and not heard horizontal ellipsis sometimes: Anthropomorphism and AI service robot interactions," *Psychol. Marketing*, vol. 38, no. 12, pp. 2393–2406, Dec. 2021, doi: 10.1002/mar.21575.

[63] S. M. Branham, A. R. M. Roy, and M. A. Comp, "Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users," in *Proc. 21st ACM SIGACCESS Int. Conf. Comput. Accessibility*, 2019, pp. 446–458.

[64] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robot. Auton. Syst.*, vol. 42, no. 3/4, pp. 143–166, Mar. 2003, doi: 10.1016/s0921-8890(02)00372-x.

[65] A. Paivio, "Dual coding theory - Retrospect and current status," *Can. J. Psychol.-Revue Canadienne De Psychologie*, vol. 45, no. 3, pp. 255–287, Sep. 1991, doi: 10.1037/h0084295.

[66] P. Razzaghi, K. Abbasi, M. Shirazi, and N. Shabani, "Modality adaptation in multimodal data," *Expert Syst. With Appl.*, vol. 179, Oct. 2021, Art. no. 115126, doi: 10.1016/j.eswa.2021.115126.

[67] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, "Image-text sentiment analysis via deep multimodal attentive fusion," *Knowl.-Based Syst.*, vol. 167, pp. 26–37, Mar. 2019, doi: 10.1016/j.knosys.2019.01.019.

[68] A. Mohsenpour, L. Biddle, K. Krug, and K. Bozorgmehr, "Measuring deterioration of small-area housing environment: Construction of a multi-dimensional assessment index and validation in shared refugee accommodation," *SSM-Popul. Health*, vol. 13, Mar. 2021, Art. no. 100725, doi: 10.1016/j.ssmph.2020.100725.

[69] P. Marshall, "A latent allocation model for brand awareness and mindset metrics," *Int. J. Market Res.*, vol. 64, no. 4, Jul. 2022, Art. no. 14707853211040052, doi: 10.1177/14707853211040052.

[70] H. Park, S. Jiang, O. K. D. Lee, and Y. Chang, "Exploring the attractiveness of service robots in the hospitality industry: Analysis of online reviews," *Inf. Syst. Front.*, 2021, Art. no. 21, doi: 10.1007/s10796-021-10207-8.

[71] D. Huang, Q. R. Chen, J. H. Huang, S. J. Kong, and Z. Y. Li, "Customer-robot interactions: Understanding customer experience with service robots," *Int. J. Hospitality Manage.*, vol. 99, Oct. 2021, Art. no. 103078, doi: 10.1016/j.ijhm.2021.103078.

[72] P. Hu, Y. B. Lu, and Y. M. Gong, "Dual humanness and trust in conversational AI: A person-centered approach," *Comput. Hum. Behav.*, vol. 119, Jun. 2021, Art. no. 106727, doi: 10.1016/j.chb.2021.106727.

[73] K. Li, X. Liu, F. Mai, and T. Zhang, "The role of corporate culture in bad times: Evidence from the COVID-19 Pandemic," *J. Financial Quant. Anal.*, vol. 56, no. 7, Nov. 2021, Art. no. Pii s0022109021000326, doi: 10.1017/s0022109021000326.

[74] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning word vectors for 157 languages," in *Proc. 11th Int. Conf. Lang. Resour. Eval.*, 2018, pp. 3483–3487.

[75] Q. Le and T. Mikolov, "Distributed representation of sentences and documents," in *Proc. Int. Conf. Mach. Learn.*, 2014, vol. 32, pp. 1188–1196.

[76] A. Jurek, M. D. Mulvenna, and Y. Bi, "Improved lexicon-based sentiment analysis for social media analytics," *Secur. Inform.*, vol. 4, 2015, Art. no. 9.

[77] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, Aug. 2013, doi: 10.1111/j.1467-8640.2012.00460.x.

[78] L. Xu, H. Lin, Y. Pan, H. Ren, and J. Chen, "Constructing the affective lexicon ontology," *J. China Soc. Sci. Tech. Inf.*, vol. 27, no. 2, pp. 180–185, Apr. 2008.

[79] M. S. I. Malik and A. Hussain, "Helpfulness of product reviews as a function of discrete positive and negative emotions," *Comput. Hum. Behav.*, vol. 73, pp. 290–302, Aug. 2017, doi: 10.1016/j.chb.2017.03.053.

[80] R. Gang and H. Taeho, "Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews," *Inf. Process. Manage.*, vol. 56, no. 4, pp. 1425–1438, Jul. 2019, doi: 10.1016/j.ipm.2018.04.003.

[81] J. Buschken and G. M. Allenby, "Sentence-based text analysis for customer reviews," *Marketing Sci.*, vol. 35, no. 6, pp. 953–975, Nov./Dec. 2016, doi: 10.1287/mksc.2016.0993.

[82] M. E. Roberts et al., "Structural topic models for open-ended survey responses," *Amer. J. Political Sci.*, vol. 58, no. 4, pp. 1064–1082, 2014.

[83] M. Yang and C. J. Han, "Revealing industry challenge and business response to COVID-19: A text mining approach," *Int. J. Contemporary Hospitality Manage.*, vol. 33, no. 4, pp. 1230–1248, May 2021, doi: 10.1108/ijchm-08-2020-0920.

[84] S. Z. Bai et al., "Exploring thematic influences on theme park visitors' satisfaction: An empirical study on Disneyland China," *J. Consum. Behav.*, 2023, Art. no. 17, doi: 10.1002/cb.2157.

[85] D. Weathers, S. D. Swain, and V. Grover, "Can online product reviews be more helpful? Examining characteristics of information content by product type," *Decis. Support Syst.*, vol. 79, pp. 12–23, Nov. 2015, doi: 10.1016/j.dss.2015.07.009.

[86] M. X. Han, "Examining the effect of reviewer expertise and personality on reviewer satisfaction: An empirical study of TripAdvisor," *Comput. Hum. Behav.*, vol. 114, Jan. 2021, Art. no. 106567, doi: 10.1016/j.chb.2020.106567.

[87] M. Ghasemaghaei, S. P. Eslami, K. Deal, and K. Hassanein, "Reviews' length and sentiment as correlates of online reviews' ratings," *Internet Res.*, vol. 28, no. 3, pp. 544–563, 2018, doi: 10.1108/IntR-12-2016-0394.

[88] S. Zhou, X. Sun, B. Liu, and G. Burnett, "Factors affecting pedestrians' trust in automated vehicles: Literature review and theoretical model," *IEEE Trans. Hum.-Mach. Syst.*, vol. 52, no. 3, pp. 490–500, Jun. 2022, doi: 10.1109/thms.2021.3112956.

[89] F. Mehraliyev, A. P. Kirilenko, and Y. Choi, "From measurement scale to sentiment scale: Examining the effect of sensory experiences on online review rating behavior," *Tourism Manage.*, vol. 79, Aug. 2020, Art. no. 104096, doi: 10.1016/j.tourman.2020.104096.

[90] M. E. Roberts, B. M. Stewart, and E. M. Airoldi, "A model of text for experimentation in the social sciences," *J. Amer. Statist. Assoc.*, vol. 111, no. 515, pp. 988–1003, Sep. 2016, doi: 10.1080/01621459.2016.1141684.

[91] M. E. Roberts et al., "Structural topic models for open-ended survey responses," *Amer. J. Political Sci.*, vol. 58, no. 4, pp. 1064–1082, Oct. 2014, doi: 10.1111/ajps.12103.

[92] M. E. Roberts, B. M. Stewart, and D. Tingley, "STM: An R package for structural topic models," *J. Statist. Softw.*, vol. 91, no. 2, pp. 1–40, Oct. 2019, doi: 10.18637/jss.v091.i02.

[93] K. Li, Q. Zhang, S. Kwong, M. Li, and R. Wang, "Stable matching-based selection in evolutionary multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 18, no. 6, pp. 909–923, Dec. 2014, doi: 10.1109/tevc.2013.2293776.

[94] D. Berliner, B. E. Bagozzi, and B. Palmer-Rubin, "What information do citizens want? Evidence from one million information requests in Mexico," *World Develop.*, vol. 109, pp. 222–235, Sep. 2018, doi: 10.1016/j.worlddev.2018.04.016.

[95] N. Hu, T. Zhang, B. Gao, and I. Bose, "What do hotel customers complain about? Text analysis using structural topic model," *Tourism Manage.*, vol. 72, pp. 417–426, Jun. 2019, doi: 10.1016/j.tourman.2019.01.002.

[96] E. Elshan, N. Zierau, C. Engel, A. Janson, and J. M. Leimeister, "Understanding the design elements affecting user acceptance of intelligent agents: Past, present and future," *Inf. Syst. Front.*, vol. 24, no. 3, pp. 699–730, Jun. 2022, doi: 10.1007/s10796-021-10230-9.

[97] H. Alnizami, I. Alvarez, and J. E. Gilbert, "Socializing under the influence of distracted driving: A study of the effects of in-vehicle and outside-of-the-vehicle communication while driving," in *Proc. Int. Conf. Hum. Factors Transp.*, 2016, vol. 484, pp. 243–255.

[98] M. Braun, J. Schubert, B. Pfleging, and F. Alt, "Improving driver emotions with affective strategies," *Multimodal Technol. Interaction*, vol. 3, no. 1, Mar. 2019, Art. no. 21, doi: 10.3390/mti3010021.

[99] N. Haslam, P. Bain, L. Douge, M. Lee, and B. Bastian, "More human than you: Attributing humanness to self and others," *J. Pers. Social Psychol.*, vol. 89, no. 6, pp. 937–950, Dec. 2005, doi: 10.1037/0022-3514.89.6.937.

[100] K. Merrill, J. Kim, and C. Collins, "AI companions for lonely individuals and the role of social presence," *Commun. Res. Rep.*, vol. 39, no. 2, pp. 93–103, Mar. 2022, doi: 10.1080/08824096.2022.2045929.