

ChatReview: A ChatGPT-enabled natural language processing framework to study domain-specific user reviews

Brittany Ho, Ta'Rhonda Mayberry, Khanh Linh Nguyen, Manohar Dhulipala, Vivek Krishnamani Pallipuram^{*}

School of Engineering and Computer Science, University of the Pacific, 3601 Pacific Avenue, Stockton CA, USA

ARTICLE INFO

Keywords:

Natural Language Processing
ChatGPT
Sentiment analysis
Prompt engineering
Intelligent search engines
Recommender system

ABSTRACT

Intelligent search engines including pre-trained generative transformers (GPT) have revolutionized the user search experience. Several fields including e-commerce, education, and hospitality are increasingly exploring GPT tools to study user reviews and gain critical insights to improve their service quality. However, massive user-review data and imprecise prompt engineering lead to biased, irrelevant, and impersonal search results. In addition, exposing user data to these search engines may pose privacy issues. Motivated by these factors, we present ChatReview, a ChatGPT-enabled natural language processing (NLP) framework that effectively studies domain-specific user reviews to offer relevant and personalized search results at multiple levels of granularity. The framework accomplishes this task using four phases including data collection, tokenization, query construction, and response generation. The data collection phase involves gathering domain-specific user reviews from public and private repositories. In the tokenization phase, ChatReview applies sentiment analysis to extract keywords and categorize them into various sentiment classes. This process creates a token repository that best describes the user sentiments for a given user-review data. In the query construction phase, the framework uses the token repository and domain knowledge to construct three types of ChatGPT prompts including explicit, implicit, and creative. In the response generation phase, ChatReview pipelines these prompts into ChatGPT to generate search results at varying levels of granularity. We analyze our framework using three real-world domains including education, local restaurants, and hospitality. We assert that our framework simplifies prompt engineering for general users to produce effective results while minimizing the exposure of sensitive user data to search engines. We also present a one-of-a-kind Large Language Model (LLM) peer assessment of the ChatReview framework. Specifically, we employ Google's Bard to objectively and qualitatively analyze the various ChatReview outputs. Our Bard-based analyses yield over 90% satisfaction, establishing ChatReview as a viable survey analysis tool.

1. Introduction

While the theoretical foundations of natural language processing (NLP) were laid in the 1950s (Turing, 1950), the pre-trained generative transformers (GPT) introduced by OpenAI (Radford, Narasimhan, Salimans, & Sutskever, 2018) have transformed the user search experience. These large language models (LLMs) have permeated every conceivable area, ranging from software engineering (Ozkaya, 2023) to fake news detection (Caramancion, 2023) to intensive care medicine (Lu, Wu, Qi, & Cheng, 2023). Several domains including e-commerce, local businesses, education, hospitality, and government employ text-based surveys to identify nuanced areas for improvement. The LLMs are lucrative instruments to conduct survey research in these fields. Specifically, the stakeholders in these areas may query the LLM models

with excerpts of their survey data and obtain suggestions to improve the quality of service. Although LLMs are promising tools, they are yet to prove themselves as reliable survey research tools (Jansen, gyo Jung, & Salminen, 2023). For a vast majority of general users, several challenges arise in the pursuit of using the LLM models for large survey analysis. First, feeding the survey data directly into these models compromises data privacy. One popular example is the Samsung data breach where the employees shared their source code and meeting notes with ChatGPT (Gurman, 2023). Second, it is highly challenging to effectively extract user sentiments from an exceedingly large survey data and choose an appropriate set of domain attributes for analysis. Third, A simple feed of large survey data into the LLM models may yield

^{*} Corresponding author.

E-mail addresses: t_ho20@u.pacific.edu (B. Ho), t_mayberry@u.pacific.edu (T. Mayberry), k_nguyen173@u.pacific.edu (K.L. Nguyen), m_dhulipala@u.pacific.edu (M. Dhulipala), vpallipuramkrishnamani@u.pacific.edu (V.K. Pallipuram).

<https://doi.org/10.1016/j.mlwa.2023.100522>

Received 31 August 2023; Received in revised form 23 November 2023; Accepted 17 December 2023

Available online 28 December 2023

2666-8270/© 2023 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

incoherent results. Fourth, the LLM models require effective prompts to generate satisfactory responses. An imprecise prompt may yield biased, incorrect, and irrelevant results.

Using the above as motivation, we present ChatReview, a user-friendly framework that analyzes large survey data and provides domain-specific suggestions to improve the quality of service. By keeping a general user in mind, the ChatReview framework addresses the aforementioned challenges to provide relevant and personalized recommendations at varying levels of granularity. The framework accomplishes this task using four phases namely, data collection, tokenization, query construction, and response generation.

In the first phase, data collection, the ChatReview framework inputs the user survey data and removes any instances that may identify the stakeholders. This process offers a degree of privacy and reduces the probability of biased results. In the second phase, tokenization, the framework applies sentiment analysis to survey data to extract keywords (tokens) and categorize them into various sentiment classes. This process results in a token repository that characterizes the survey data. In the third phase, query construction, the framework generates domain-specific prompts for the generative artificial intelligence (AI) using a two-step process. First, the framework constructs the domain attributes either by employing a user's or a generative AI's assistance. Second, the framework maps the tokens to the domain attributes in order to identify the essential ones that best explain the data. For these essential attributes, the ChatReview framework designs three types of prompts including explicit, implicit, and creative prompts. In the fourth phase, response generation, the framework sends the essential attribute prompts to the generative AI and obtains recommendations. The explicit prompts aim to obtain specific answers, whereas the implicit prompts allow the generative AI some flexibility in response generation. The creative prompts enable generative AI to devise innovative recommendations by providing the highest level of flexibility. This prompting method employed by the framework yields responses at varying levels of granularity, offering a holistic insight into the survey data.

We test the ChatReview framework using three real-world domains including education, local business, and hospitality. Using conventional machine learning metrics including accuracy, precision, recall, F1-score, and mean absolute error (MAE), our results show that ChatReview can match human analysis, making it a viable human assistant for analyzing large survey data originating from any field. Unlike traditional studies that perform one-on-one comparisons between LLM models, we also present one of the earliest efforts to carry out an LLM peer evaluation. Specifically, we employ Google's Bard¹ as the peer reviewer for the ChatReview framework. To evaluate the framework, Bard objectively and qualitatively analyzes various ChatReview outputs including the domain attribute selection and final recommendations. Our Bard analysis yields over 90% satisfaction rating, thereby establishing ChatReview as a valuable survey analysis tool. The ChatReview framework presented in this article makes the following contributions:

1. The framework securely analyzes large survey data to automatically generate domain-specific suggestions to improve the quality of service.
2. It allows users to employ their domain knowledge for analysis. Optionally, the framework can also generate the domain attributes for analysis in case the users' domain knowledge is limited.
3. Using state-of-the-art NLP models, it maps the sentiments embedded in the survey to domain attributes, thereby identifying the essential ones that characterize the survey data.
4. It implements prompt engineering to generate responses at varying levels of granularity, which is conducive to a holistic analysis.

5. To our knowledge, it is one of the first attempts to introduce the concept of peer assessment for LLM models. Specifically, the framework evaluates itself using a peer LLM model, Google's Bard.
6. The framework makes a strong case that it is not a replacement for human analysis, but a valuable aid for a thorough study.

The rest of the article is organized as follows. Section 2 explains the ChatReview framework. Section 3 illustrates the framework using three real-world domain case studies. Section 4 evaluates the ChatReview framework's ability to match the human analysis. This section also presents a peer assessment strategy to evaluate the ChatReview framework using Google's Bard. Section 5 discusses the related work and showcases the limitations of the state-of-the-art in the area of automatic textual survey study. Section 6 provides conclusions, discusses ChatReview's limitations, and explores avenues for future work.

2. ChatReview framework

Fig. 1 shows the ChatReview framework comprising four phases including data collection, tokenization, query construction, and response generation. The framework inputs user reviews and outputs recommendations at different granularity levels to improve the quality of service. Sections 2.1–2.4 explain these phases.

2.1. Phase 1: Data collection

In this phase, ChatReview inputs user reviews in the JavaScript Object Notation (JSON) format with one free-form review per line. The data sources include open-source repositories such as Google reviews, Rate My Professor reviews, and Yelp, among others. Optionally, users may also employ data originating from their organization. To promote user privacy and reduce bias, the framework deletes any instances that identify the names, gender, and locations of the stakeholders (the organization and their users). The framework ignores any reviews in crass language because they add little to no value in a holistic analysis. The framework requires users to collect sufficient reviews for a compelling study; our framework recommends more than 20 reviews as optimal.

2.2. Phase 2: Tokenization

In the tokenization phase, the framework inputs the JSON-formatted user reviews and extracts keywords or tokens that effectively describe the user sentiments. First, as a preprocessing step, the framework eliminates stop words such as 'and', 'the', 'but', 'to', etc. Second, the framework inputs the remaining tokens into the popular sentiment analysis model, National Research Council Lexicon (NRCLex) (Mohammad & Turney, 2013), to categorize them into one of the ten sentiment categories. These include fear, anger, anticipation, trust, surprise, positive, negative, sadness, disgust, and joy. The NRCLex model comprises a library of ≈ 27000 words that are distributed across the ten sentiment categories. The NRCLex model computes the similarity between the tokens and the words in the library to identify the most similar sentiment category. We chose the NRCLex model because it is more accessible to general users versus other models such as Word2Vec (Church, 2017), Bidirectional Encoder Representations from Transformers (BERT) (Devlin, Chang, Lee, & Toutanova, 2019), and topic modeling (Albalawi, Yeap, & Benyoucef, 2020), which require some NLP expertise.

Given the nature of the domain under study, the ChatReview framework analyzes only a subset of the sentiment categories and creates a token repository. Consider an example of a restaurant hoping to improve its service quality. They may choose to study anger, negative, and disgust sentiments to gauge their users' experience. In this case, the framework considers the categorized tokens from these sentiment classes to construct a token repository that best describes user sentiments regarding the restaurant. The framework also allows users to select user sentiments that are pertinent to their analysis. The framework forwards the token repository to phase-3 query construction.

¹ www.google.com

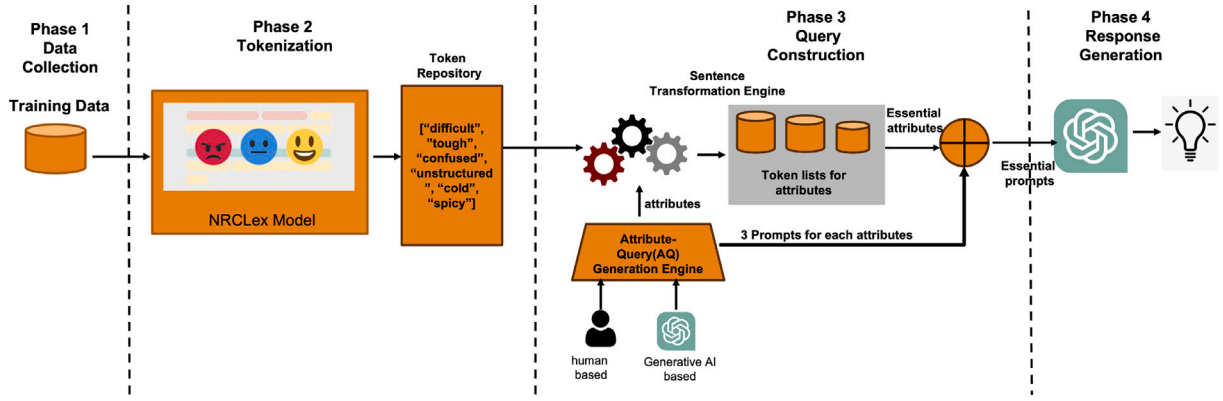


Fig. 1. The ChatReview framework comprises four phases: data collection, tokenization, query construction, and response generation.

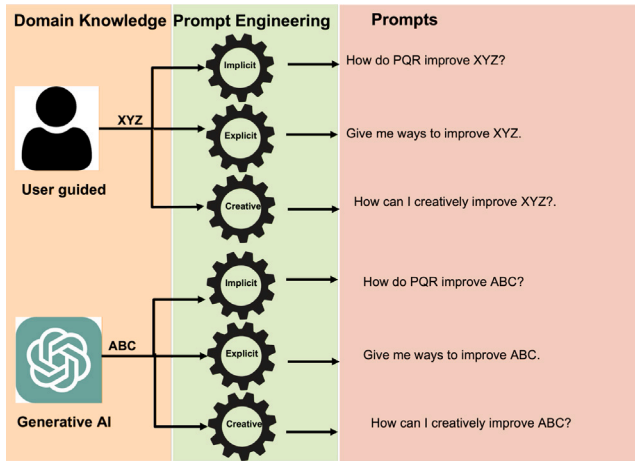


Fig. 2. The attribute-query (AQ) generation engine includes two components: domain knowledge and prompt engineering. The domain knowledge is either user-guided or enabled by a generative AI (ChatGPT in this example). The prompt types include implicit, explicit, and creative to obtain insights at different levels of granularity.

2.3. Phase 3: Query construction

In the query construction phase, the ChatReview framework inputs the token repository and domain-specific information (either provided by the user or by a generative artificial intelligence (AI)) to construct insightful questions (prompts) for the response generation phase. The framework accomplishes this task using two engines: Attribute-Query (AQ) Generation and Sentence Transformation (ST).

Attribute-Query (AQ) Generation Engine — Fig. 2 shows the AQ generation engine comprising two components namely, domain knowledge and prompt engineering. The domain knowledge component generates attributes that best describe the domain under study. For instance, the hospitality domain may be best described by attributes including, but not limited to, room quality, cleanliness, complimentary meals, and safety features, among others. The prompt engineering component employs these attributes to frame prompts at three levels of granularity: implicit, explicit, and creative.

The domain knowledge component generates the domain-specific attributes using one of the two options: user-guided and generative AI. The user-guided option allows users to select attributes using their domain knowledge. For instance, an educational institution aiming to improve its instruction quality may select attributes including class time, course organization, instructor's speaking ability, etc. The generative AI option allows users with limited domain knowledge to create the domain attributes automatically. For example, if the educational institution outsources the analysis to a third party, they may exercise

the generative AI option to identify the attributes that best describe the education domain.

The prompt engineering component creates three types of prompts including explicit, implicit, and creative prompts (Horizons, 2023). They offer varying degrees of flexibility in response generation, ranging from least (explicit) to most flexible (creative). This flexibility allows ChatReview to generate responses at different levels of granularity, promoting a comprehensive analysis of the survey data. The explicit prompts are focused questions that aim to get precise answers. For example, a restaurant owner seeking a specific answer regarding ambiance may ask, "How can I create an excellent restaurant ambiance at my XYZ location?" On the contrary, implicit prompts are intentionally indirect, offering generative AI more flexibility in providing answers. For instance, the previous question on ambiance can be implicitly stated as, "How do local restaurants offer excellent ambiance?" The creative prompts offer the generative AI the most flexibility by including words such as creative, innovate, produce, etc. These words make the generative AI think outside the box. The restaurant ambiance question can be creatively stated as follows: "Give me creative (or innovative) ways in which I can improve my restaurant's ambiance".

In summary, the AQ generation engine provides two significant outputs: (1) the domain attributes pertaining to the survey data and (2) the three types of prompts for each of the domain attributes.

Sentence Transformation Engine — The sentence transformation (ST) engine uses the all-MiniLM-L6-v2 sentence transformer model (Reimers & Gurevych, 2019) to map the tokens from the repository to the attributes identified by the domain knowledge component of the AQ generation engine (see Fig. 1). Specifically, the transformer maps the tokens to attributes based on their semantic similarity. The transformer computes the cosine similarity between a given token's vector and the vectors of each domain attribute. The cosine similarity score varies from -1 and 1 . If the vectors are highly similar, the score is close to 1 . If they are dissimilar, then the score is close to -1 . If they are orthogonal, the score is 0 . The sentence transformation engine checks whether the cosine similarity between a token and a domain attribute is greater than a predetermined threshold. By conducting an empirical analysis, we identify 0.25 as an effective threshold value. For a given token-attribute pair, if the cosine similarity is greater than 0.25 , then the pair is deemed related and the token is added to the attribute's list. The sentence transformation engine performs this task on all possible attribute-token pairs, resulting in a list of tokens for each attribute. The top three attributes with the longest token lists are regarded as the essential attributes that explain the survey data. The ChatReview framework allows users to choose the number of essential attributes that fit their analysis.

After the identification of the essential attributes, the framework passes their explicit, implicit, and creative prompts, henceforth called essential prompts, to the response generation phase.

Table 2

The user-selected education domain attributes and their respective three prompts generated by the attribute-query (AQ) generation engine of phase-3 query construction. The table lists ten selected attributes and their prompts.

| Domain attributes | Domains | | |
|---------------------------------------|--|---|---|
| | Explicit | Implicit | Creative |
| Course organization | How can I improve my course organization? | How do professors organize their courses? | How can I creatively organize my courses? |
| Homework effectiveness | How can I create effective homework? | How do professors create effective homework? | Give me creative ways to assign effective homework. |
| Grading | How do I improve my grading efficiency? | How do good professors grade? | How can I creatively improve my grading |
| Course relevance | How do I convey my course's relevance? | How do effective educators convey course relevance? | How can I be creative in conveying my course's relevance? |
| Enthusiasm for the subject | How do I exude enthusiasm for the subject matter? | How do seasoned professors display enthusiasm for the subject matter? | What are some creative ways in which I can convey my enthusiasm for the subject matter? |
| Preparedness | How should I prepare for the classes? | How do educators effectively prepare for their classes? | Give me creative ways to effectively prepare for my classes. |
| Ability to explain difficult concepts | How can I effectively explain difficult concepts? | How do seasoned professors explain difficult concepts in class? | Give me creative ways in which I can explain a difficult concept in class. |
| Presentation ability | How can I create a clean presentation for class? | What are the attributes of a good oral presentation? | Give me some creative ways in which I can present in class. |
| Responsiveness to students' needs | How do I respond to my student's needs? | How do good educators respond to their student's needs? | Give me some unique ways in which I can respond to my students' needs. |
| Effectiveness of examples | How should I create examples to explain concepts in class? | How do good educators provide good examples of concepts? | Give me creative ways in which I can create examples to demonstrate a concept. |

with only minor differences. For instance, the explicit prompt suggests inviting guest speakers and experts to boost enthusiasm for the subject matter. On the other hand, the implicit prompt responses are more broad and less direct. The creative prompt generates more innovative suggestions than the other two prompts. These suggestions include gamification with unique tools (e.g. use of leaderboards), teaching experimentation, mentorship, and real-world examples.

In summary, we observe that the implicit and explicit prompt responses are similar, whereas the creative prompt responses are more unique and innovative.

3.2.2. Local business domain

To execute phase-1 data collection, we compile openly available Google text reviews of selected restaurants and format them in the JSON format with one review per line. We create a hypothetical case where this review data corresponds to an imaginary restaurant. To promote the privacy of the entities, we format the data such that the entity names are unidentifiable and irretrievable. The framework passes this review data to the tokenization phase, which generates the token repository using the anger, negative, and disgust sentiments (see Table 1). Fig. 4 shows the resulting token repository. As seen in this figure, the large tokens that underscore the customers' emotions are 'disappointed', 'bad', 'horrible', 'stale', 'awful', and 'wrong'. These tokens suggest that the food quality and/or the service suffers at this restaurant.

In the query construction phase, the framework uses the token repository to perform token-to-domain attribute mapping. The AQ generation engine inputs the user-guided domain attributes and formulates their prompts given in Table 3. Next, the ST engine maps the tokens from the repository to the domain attributes. It identifies 'food flavor', 'cleanliness', and 'atmosphere' as the essential attributes that characterize the reviews. This identification is consistent with large tokens in Fig. 4 including 'disappointed', 'bad', 'horrible', 'awful', and 'stale'.

The framework uses the three prompt types pertaining to the essential attributes in phase 4 for response generation.

Table 7 provides the three prompts and their responses for the food-flavor attribute. As seen in the table, the implicit and explicit prompts generate five similar suggestions including the use of fresh ingredients, experimentation with herbs and spices, incorporation of acidity, experimentation with cooking techniques, and feedback from customers. The explicit prompt also generates some unique insights including balanced flavors, the use of umami-rich ingredients, and the fusion of ethnic flavors. Some unique recommendations by the implicit prompt include the pairing of ingredients and flavor consistency. As expected, the creative prompt yields innovative suggestions with specific examples. For instance, infusing flavors using oil and vinegar, playing with texture using toasted nuts, and using fermented ingredients such as kimchi. In summary, we observe that explicit, implicit, and creative prompt responses have sufficient diversity, providing the user with a comprehensive analysis.

3.3. Generative AI guided domain studies

This section discusses the framework using the domain knowledge created by a generative AI, specifically ChatGPT 3.5. Section 3.3.1 discusses the hospitality domain with a focus on hotels. Section 3.3.2 studies the local business domain using restaurants as a case study.

3.3.1. Hospitality domain

To execute the data collection phase, we gathered online hotel reviews to simulate customers' feedback on a fictitious hotel. In the tokenization phase, ChatReview inputs the reviews to generate a token repository using negative, anger, fear, and disgust emotions (see Table 1). Fig. 5 shows the resulting token repository. As shown in the figure, the large tokens include 'bad', 'smell', 'terrible', 'homeless', and 'sketchy'. These tokens indicate that the customers are unhappy

Table 3

The user-selected local business (restaurant) domain attributes and their respective three prompts generated by the attribute-query (AQ) generation engine of phase-3 query construction. The table lists ten selected attributes and their prompts.

| Domain attributes | Domains | | |
|-------------------------------|--|--|--|
| | Explicit | Implicit | Creative |
| Quality of food and beverages | How can I improve the quality of food and beverages provided for my restaurant? Give me examples | How do restaurants improve the quality of food and beverages provided? | Give me creative ways to improve the quality of food and beverages provided for my restaurant. |
| Order accuracy | How can I improve the accuracy of order for my restaurant? Give me examples | How do restaurants improve their accuracy of orders? | Give me creative ways to improve the accuracy of order for my restaurant. |
| Service accompaniments | How can I improve the service accompaniments provided (plates, glassware, cutlery, etc.) for my restaurant? Give me examples | How do restaurants improve their service accompaniments provided (plates, glassware, cutlery, etc.)? | Give me creative ways to improve the service accompaniments provided (plates, glassware, cutlery, etc.) for my restaurant. |
| Reservation services | How can I improve the quality of service provided by Special Functions staff (making reservations, etc.) for my restaurant? Give me examples | How do restaurants improve the quality of service provided by Special Functions staff (making reservations, etc.)? | Give me creative ways to improve the quality of service provided by Special Functions staff (making reservations, etc.) for my restaurant. |
| Portion size | How can I improve the portion size (possible adjustments) for my restaurant? Give me examples | How do restaurants improve their portion size? | Give me creative ways to improve the portion size (possible adjustments) for my restaurant. |
| Flavor | How can I improve the flavor of dishes for my restaurant? Give me examples | How do restaurants improve the flavor of dishes? | Give me creative ways to improve the flavor of dishes for my restaurant. |
| Presentation | How can I improve the presentation of dishes for my restaurant? Give me examples | How do restaurants improve their presentation of dishes? | Give me creative ways to improve the presentation of dishes for my restaurant. |
| Cleanliness | How can I improve my restaurant's cleanliness? Give me examples | How do restaurants maintain a clean space? | Give me creative ways to promote my restaurant's cleanliness. |
| Wait time | How can I improve the wait time for food service for my restaurant? Give me examples | How do restaurants improve their wait time for food service? | Give me creative ways to improve the wait time for food service for my restaurant. |
| Atmosphere | How can I improve the atmosphere of my restaurant? Give me examples | How do restaurants improve their atmosphere? | Give me creative ways to improve the atmosphere for my restaurant. |



Fig. 4. The token repository for the local business domain using a sample online review data. This repository captures the anger, negative, and disgust sentiments of the customers.

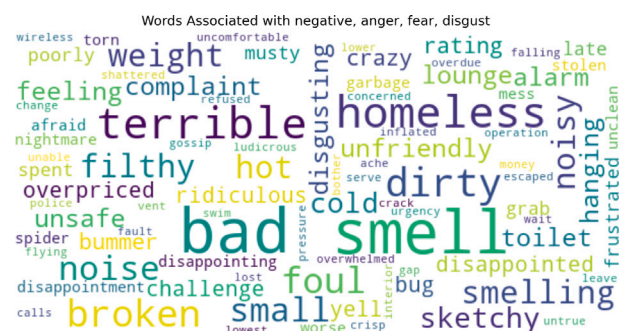


Fig. 5. The token repository for the hospitality domain using online data. This repository captures the anger, negative, fear, and disgust sentiments of the customers.

Table 4

The ChatGPT generated hospitality domain attributes and their respective three prompts generated by the attribute-query generation (AQ) engine of phase-3 query construction. The table lists ten selected attributes and their prompts.

| Domain attributes | Domains | | |
|-----------------------------|--|---|---|
| | Explicit | Implicit | Creative |
| Cleanliness and hygiene | How can I improve cleanliness and hygiene quality of my hotel? | How do good hotels work on their cleanliness and hygiene? | Give me creative ways to improve cleanliness and hygiene quality of my hotel? |
| Comfort | How can I offer the most comfortable stay to my customers? | How do good hotels promote comfortable stay of their customers? | Give me creative ways to increase customer's comfort level at my hotel. |
| Efficient staff | How do good hotels maintain staff efficiency? | How do good hotels maintain staff efficiency? | Give me creative ways to maintain staff efficiency, |
| Convenient location | How can I improve my hotel's profits despite its location? | How do good hotels maintain profitability despite their location? | How can I be creative to work around my hotels peculiar location. |
| Basic facilities | What basic facilities can my hotel offer? | What basic facilities do good hotels offer? | How can I be creative with basic facilities at my hotel? |
| Good food | How can I improve my customers' dining satisfaction? | How do good hotels address their customers' dining satisfaction? | How can I offer creative dining experience to my customers? |
| Efficient checkin | How can my hotel promote checkin and checkout efficiency? | How do good hotels promote checkin and checkout efficiency? | How can I creatively improve customer checkin and checkout experience? |
| Ambiance | How can my hotel offer excellent ambiance? | How do good hotels offer excellent ambiance? | Give me creative ways to improve my hotel's ambiance. |
| Pricing | How can I create a lucrative pricing model for my hotel? | How do good hotels work on their pricing? | How can I be creative with my hotel pricing? |
| Responsiveness to customers | How can I best analyze customer feedback? | How do good hotels respond to customer feedback? | How can I creatively address customer's constructive criticism? |

with the hotel's hygiene and location. The framework passes the token repository to the query construction phase.

The AQ generator engine obtains the domain attributes using ChatGPT 3.5 generative AI. Specifically, we asked ChatGPT, "What are the top ten attributes of an excellent hotel?" We consider the top ten attributes generated by ChatGPT as the domain attributes for the study. The attributes and their three prompts appear in Table 4. Next, the ST engine maps the tokens from Fig. 5 to the domain attributes. The engine identifies the following three essential attributes: 'comfort', 'pricing', and 'good food'. The tokens including 'terrible', 'bad', and 'smell' qualitatively map to the comfort attribute. These tokens along with the additional ones such as 'hot' and 'cold' map to the 'good food' attribute. The framework uses the 'overpriced' token to identify 'pricing' as an essential attribute. The framework misses considering the 'convenient location' attribute since 'homeless', 'sketchy', and 'afraid' tokens map to it. Nonetheless, the identified essential attributes reasonably capture the token repository. The ChatReview framework uses the three prompts of the essential attributes for response generation in phase 4. Table 8 provides the prompts and responses for the 'comfort' attribute.

Given the responses in Table 8, we observe that explicit and implicit prompts generate similar results. However, the implicit response captures additional insights including comfortable seating, workplace, and recreational facilities such as swimming pools, fitness centers, etc. The creative responses are more focused and innovative where each creative idea is accompanied by an example. Some creative suggestions not captured by the explicit and implicit prompts include a pillow menu, technology conveniences such as voice-controlled assistants, and personalized recommendations for local attractions, restaurants, and activities.

3.3.2. Local business domain

For this study, the data collection and tokenization phases are identical to Section 3.2.2. The framework passes the token repository

(see Fig. 4) to phase-3 query construction. The AQ generation engine employs ChatGPT to generate ten domain attributes, which appear in Table 5 along with their explicit, implicit, and creative prompts.

The ST engine maps the tokens from the token repository to domain attributes. The engine identifies the following three essential attributes: 'marketing', 'service', and 'good food'. These attributes correlate with the large tokens including 'disappointed', 'bad', 'horrible', 'small', 'avoid', and 'inconvenient'. Table 9 provides the prompts and their responses for the 'marketing' domain attribute. In this case, both the explicit and implicit prompts generate similar responses. The creative responses offer more outside-the-box suggestions versus the other two prompts. For example, ideas including cooking classes and unique promotions such as a "secret menu" are distinctive recommendations.

In summary, this section discusses the framework's capability to analyze survey data and recommend strategies to improve the quality of service. We use three real-world domains including education, local business, and hospitality. The section studies the framework by constructing the domain knowledge using two methods: user-guided and generative AI. The user-guided domain attributes are input by the user whereas the generative AI-guided domain attributes are obtained by providing simple prompts to the generative AI. We observe that explicit and implicit prompts generate similar responses, however, the implicit prompt responses generally cover broader areas than their explicit counterparts. The creative prompts generate unique recommendations versus the other two prompt types. Overall, the three prompt types yield responses at varying levels of granularity; together they offer a well-rounded set of recommendations. The next section evaluates the framework's ability to effectively identify the essential domain attributes that best describe the survey data.

4. Discussion

This section assesses the various functionalities of the ChatReview framework. Section 4.1 evaluates the framework's ability to automatically identify the essential attributes that best describe the survey

Table 5

The ChatGPT generated local business (restaurant) domain attributes and their respective three prompts generated by the attribute-query (AQ) engine of phase-3 query construction. The table lists ten selected attributes and their prompts.

| Domain attributes | Domains | | |
|-----------------------------|--|--|---|
| | Explicit | Implicit | Creative |
| Hygiene | How can my restaurant work on its cleanliness and hygiene? | How do good restaurants work on their cleanliness and hygiene? | Give me creative ways to improve my restaurant's cleanliness and hygiene? |
| Menu | How can my restaurant craft an excellent menu? | How do good restaurants craft an excellent menu? | How can I creatively craft an excellent menu? |
| Service | How can my restaurant maintain staff efficiency? | How do good restaurants maintain staff efficiency? | How my restaurant creatively maintain staff efficiency? |
| Sourcing and sustainability | How can my restaurant source the ingredients in a sustainable way? | How do good restaurants source their ingredients in a sustainable way? | give me creative ways to source my ingredients in a sustainable way. |
| Marketing | How can I market my restaurant? | How do good restaurants market themselves? | Give me creative ways to market my local restaurant. |
| Good food | How can my restaurants address the customers' dining satisfaction? | How do good restaurants address their customers' dining satisfaction? | How can I be creative in offering an excellent dining experience? |
| Efficient operation | How can my restaurant promote dining efficiency? | How do good restaurant promote dining efficiency? | How can I be creative with my customer's dining efficiency? |
| Ambiance | How can my restaurant offer excellent ambiance? | How do good restaurants offer excellent ambiance? | Give me creative ways to promote excellent ambiance at my restaurant. |
| Community connection | How can my restaurant connect with the community? | How do good restaurants connect with the community? | How can my restaurant creatively engage with the local community? |
| Responsiveness | How can my restaurant respond to customer feedback? | How do good restaurants respond to customer feedback? | Give me creative ways to address customers' feedback. |

data. This study is critical because, without identifying the essential attributes to frame the correct questions, even the precise responses are of little value. Section 4.5 presents one of the first LLM peer assessment efforts. Specifically, we ask Bard to rate the quality of the essential attributes and their responses given by the framework.

4.1. Attribute evaluation

This section discusses the ChatReview framework's ability to identify the essential attributes. We consider three studies: user-guided education domain, generative AI-guided local business domain, and generative AI-guided hospitality domain. For each of these studies, we select co-authors as users of the framework. Prior to this study, they built their domain knowledge using online research and speaking with the domain experts. Therefore, we label them as users with 'average' domain knowledge. We asked the users to read the survey comments and map them to one or more domain attributes. After this mapping, we rank the domains based on the number of mapped comments. The more comments mapped to a given domain, the higher its ranking. To keep the study unbiased, we do not disclose the framework's results to the users. The ChatReview framework also performs this ranking process in phase 3, giving us the ChatReview ranking of the attributes. We compare the user ranking and ChatReview ranking to evaluate the framework's ability to identify the essential domain attributes.

It is worth noting that our goal with the single-user analysis is to evaluate the framework's capabilities with a single generic user's perspective with average domain knowledge. Future work includes testing the framework with multiple users with varying expertise.

To evaluate the divergence in user's and ChatReview's ranking, we compute the mean absolute error (MAE) given in Eq. (1) where N is to total number of domain attributes (10 in our study). The ChatReview framework identifies the top three attributes as the essential attributes.

We assert that the framework performs well in identifying these attributes if they appear among the top five identified by the user. A satisfactory identification of the essential attributes is paramount because the framework uses them to formulate prompts for the generative AI, which then offers the final recommendations.

$$MAE = \frac{\sum_{i=1}^{i=N} |user_{rank_{domain_i}} - ChatReview_{rank_{domain_i}}|}{N} \quad (1)$$

In addition to MAE, we also obtain the accuracy, precision, recall, and F1-score metrics using Eqs. (2)–(5) where TP , TN , FP , and FN are true positives, true negatives, false positives, and false negatives, respectively. In this analysis, a true positive occurs when both the user and ChatReview rank an attribute among the top five. If the user ranks an attribute in the bottom five and the ChatReview ranks it in the top five, then it is a false positive. Similarly, a true negative occurs when both the user and ChatReview rank an attribute among the bottom five. A false negative occurs when the user ranks an attribute in the top five and ChatReview ranks it in the bottom five. Sections 4.2–4.4 provide the analysis for the three studies.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$precision = \frac{TP}{TP + FP} \quad (3)$$

$$recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

4.2. User-guided education domain attribute ranking

Fig. 6 shows the user's and ChatReview's domain ranking. As seen in the figure, the ChatReview's ranking fairly tracks the user ranking,

Table 6

The three prompts and their respective responses for the education domain attribute: ‘enthusiasm for the subject’.

| | Explicit | Implicit | Creative |
|----------|---|--|--|
| Question | How do I exude enthusiasm for the subject matter? | How do seasoned professors display enthusiasm for the subject matter? | What are some creative ways in which I can convey my enthusiasm for the subject matter? |
| Response | <p>Fostering a higher level of enthusiasm among instructors for a subject can greatly impact the overall learning experience for students. Here are some examples of how you can achieve this:</p> <ol style="list-style-type: none"> 1. Professional development opportunities: Offer workshops, seminars, or conferences .. 2. Guest speakers and experts: Invite guest speakers or subject matter experts to share their experiences and insights with instructors... 3. Collaborative projects: Encourage instructors to collaborate on projects related to the subject. This could involve joint research, curriculum development, or organizing events... 4. Recognition and appreciation: Recognize and appreciate instructors' efforts and achievements in the subject area. This can be done through awards, certificates... 5. Peer learning and sharing: Facilitate opportunities for instructors to share their knowledge and experiences with each other... 6. Engaging teaching methods: Encourage instructors to incorporate innovative and engaging teaching methods into their classes. This could include using multimedia, interactive activities, or real-world examples.. 7. Continuous feedback and support: Provide regular feedback and support to instructors... 8. Encourage professional networks: Encourage instructors to join professional networks or associations related to the subject... | <p>There are several ways educational institutions can foster a higher level of enthusiasm for a subject among instructors:</p> <ol style="list-style-type: none"> 1. Professional Development Opportunities: Institutions can provide regular professional development opportunities for instructors to enhance their knowledge and skills in their subject area... 2. Collaboration and Networking: Encouraging collaboration and networking among instructors can help create a supportive community where they can share ideas, resources, and best practices... 3. Recognition and Appreciation: Recognizing and appreciating instructors' efforts and achievements can significantly boost their enthusiasm. Institutions can acknowledge exceptional instructors through awards, certificates, or public recognition... 4. Autonomy and Flexibility: Providing instructors with autonomy and flexibility in their teaching can foster enthusiasm. Allowing them to design their curriculum, choose teaching methods, and incorporate their own interests and expertise can create a sense of ownership and excitement... 5. Resources and Support: Institutions should ensure that instructors have access to adequate resources and support to effectively teach their subject. This includes providing up-to-date textbooks, technology tools, research materials, and teaching aids... 6. Student Engagement and Success: When instructors witness their students' enthusiasm and success in the subject, it can greatly enhance their own enthusiasm. Institutions can create an environment that promotes student engagement, active learning, and achievement... 7. Continuous Feedback and Evaluation: Regular feedback and evaluation processes can help instructors understand their strengths and areas for improvement... | <ol style="list-style-type: none"> 1. Gamify the learning experience: Introduce elements of gamification into the subject by creating challenges, leaderboards, and rewards for instructors who excel... 2. Encourage innovation and experimentation: Provide instructors with the freedom to explore new teaching methods, technologies, or approaches to the subject... 3. Create a community of practice: Establish a platform or regular meetings where instructors can share their experiences, resources, and best practices related to the subject... 4. Offer professional development opportunities: Organize workshops, conferences, or training sessions specifically tailored to the subject... 5. Recognize and celebrate achievements: Acknowledge and appreciate the efforts and achievements of instructors in the subject. This can be done through awards, certificates, or public recognition, boosting their enthusiasm and sense of accomplishment. 6. Provide ongoing support and mentorship: Assign experienced instructors as mentors to new or struggling instructors... 7. Incorporate real-world applications: Highlight the practical relevance and real-world applications of the subject in various fields. Invite guest speakers from relevant industries... 8. Encourage interdisciplinary connections: Showcase how the subject can be integrated with other disciplines or subjects... 9. Provide resources and tools: Ensure instructors have access to comprehensive and up-to-date resources, textbooks, online platforms, or software related to the subject... 10. Foster a positive and supportive work environment: Create a positive and inclusive work environment where instructors feel valued, respected, and supported... |

implying the tool's ability to replicate the human review process. Using Eq. (1), we obtain MAE value equal to 2, implying that the divergence between the user's and ChatReview's ranking is not more than ± 2 ranks. ChatReview identifies the three essential attributes: ‘enthusiasm for the subject matter’, ‘preparedness’, and ‘ability to explain difficult concept’. These three attributes are ranked among the top five by the user (1, 4, and 2, respectively), exhibiting congruity between the user and ChatReview in prioritizing the attributes.

The accuracy, precision, recall, and F1-scores (using Eqs. (2)–(5)) are 0.7, 0.75, 0.6, and 0.67, respectively. These metric values are expected given that our data and analysis pertain to the real world. A precision value greater than 0.5 highlights the framework's ability to identify the attributes that are truly important. A recall value greater than 0.5 implies the framework's ability to not drop the truly important attributes. In summary, ChatReview suitably replicates a human's analysis of the survey data pertaining to the education domain.

4.3. AI guided local business domain attribute ranking

Fig. 7 provides the user's and ChatReview's ranking of the local business (restaurant) domain attributes guided by the generative AI. While the ChatReview ranking tracks the user ranking for most of the attributes, the ChatReview ranking overestimates the marketing attribute (2 vs. 6 by the user) and underestimates the efficient operation (9 vs. 4 by the user) and responsiveness to customers (6 vs. 3 by the user) attributes. The MAE in this case is 2.5, meaning that the average ranking divergence is ± 2.5 ranks. The ChatReview framework identifies two essential attributes that are within the top five selected by the user. These include ‘service’ (ranked 1 vs. 2 by the user) and ‘good food’ (ranked 3 vs. 1 by the user). This observation implies that ChatReview adequately follows a generic user's analysis, showing the framework's capability to identify essential attributes. For this study, the values of accuracy, precision, recall, and F1-score are each equal to 0.4. We identify that the framework performance is lower than expected. A

Table 7
The three prompts and their respective responses for the local business domain attribute: ‘food flavor’.

| | Explicit | Implicit | Creative |
|----------|--|--|---|
| Question | How can I improve the flavor of dishes for my restaurant? Give me examples. | How do restaurants improve the flavor of dishes? | Give me creative ways to improve the flavor of dishes served at my restaurant. |
| Response | <p>There are several ways to improve the flavor of dishes in your restaurant. Here are some examples:</p> <ol style="list-style-type: none">1. Use fresh and high-quality ingredients: Opt for fresh produce, meats, and seafood to enhance the natural flavors of your dishes. Avoid using processed or low-quality ingredients.2. Experiment with herbs and spices: Incorporate a variety of herbs and spices to add depth and complexity to your dishes. For example, try using fresh basil, thyme, rosemary, or spices like cumin, paprika, or turmeric to enhance the flavors.3. Balance flavors: Aim for a balance of sweet, salty, sour, and umami flavors in your dishes...4. Incorporate acidity: Adding a touch of acidity can brighten up flavors and make dishes more vibrant. Consider using citrus juices, vinegars, or pickled ingredients...5. Experiment with cooking techniques: Explore various cooking techniques like grilling, roasting, braising, or sautéing...6. Create homemade sauces and dressings: Prepare your own sauces, dressings, and marinades to control the flavor profile and quality...7. Incorporate umami-rich ingredients: Umami is the fifth taste, known for its savory and rich flavor. Include ingredients like mushrooms, soy sauce, miso paste, Parmesan cheese, or tomatoes to add umami...8. Consider ethnic flavors and fusion: Experiment with different cuisines and flavors to create unique and exciting dishes...9. Seek feedback from customers: Encourage your customers to provide feedback on the flavors of your dishes...10. Train your chefs and staff: Invest in training your chefs and staff on flavor profiles, cooking techniques, and ingredient combinations... | <p>Restaurants can improve the flavor of dishes in several ways:</p> <ol style="list-style-type: none">1. Quality ingredients: Using fresh, high-quality ingredients is crucial for enhancing the flavor of dishes. Restaurants often source ingredients from local suppliers or farms to ensure freshness and superior taste.2. Seasoning and spices: Skilled chefs know how to use a variety of seasonings and spices to enhance the flavor profile of a dish. They carefully balance flavors, such as salt, pepper, herbs, and spices, to create a harmonious taste.3. Cooking techniques: Different cooking techniques can bring out unique flavors in ingredients. Restaurants may use methods like grilling, roasting, braising, or sautéing...4. Sauces and dressings: Restaurants often create their own signature sauces and dressings to elevate the flavor of a dish...5. Marination and brining: Marinating meats or vegetables before cooking can infuse them with flavors and tenderize them...6. Presentation and garnishes: Restaurants pay attention to the visual appeal of a dish, as it can influence the perception of taste. Garnishes like fresh herbs, edible flowers, or citrus zest can add brightness and freshness to a plate.7. Pairing ingredients: Combining ingredients that complement each other can enhance the overall flavor of a dish...8. Consistency and attention to detail: Restaurants strive to maintain consistency in flavor by following standardized recipes...9. Feedback and continuous improvement: Restaurants value customer feedback and use it to improve their dishes... | <ol style="list-style-type: none">1. Experiment with unique spice blends: Create your own signature spice blends by combining different herbs, spices, and seasonings...2. Infuse flavors: Infuse oils, vinegars, or even water with herbs, fruits, or vegetables to add a subtle yet distinctive flavor to your dishes. For example, infuse olive oil with garlic and rosemary for a flavorful drizzle over roasted vegetables.3. Play with textures: Incorporate contrasting textures in your dishes to enhance the overall flavor experience. For instance, add a crunchy element like toasted nuts or crispy fried onions to a creamy soup or salad.4. Incorporate umami-rich ingredients: Umami is known as the fifth taste, adding a savory and satisfying flavor to dishes. Include umami-rich ingredients like mushrooms, miso paste, soy sauce, or Parmesan cheese to enhance the overall taste.5. Utilize fermentation: Fermented ingredients like kimchi, sauerkraut, or pickles can add a tangy and complex flavor to various dishes...6. Smoke it up: Introduce smoky flavors to your dishes by using a smoker or adding smoked ingredients like smoked paprika, chipotle peppers, or smoked salt...7. Incorporate fresh herbs and edible flowers: Fresh herbs like basil, cilantro, or mint can elevate the flavor profile of your dishes. Additionally, edible flowers like lavender, rose petals, or marigold can add a delicate and aromatic touch to desserts or salads.8. Create unexpected flavor combinations: Experiment with unconventional flavor pairings to surprise and delight your customers. For example, combine sweet and savory elements like adding a drizzle of honey to a savory dish or incorporating a hint of spice into a dessert.9. Use seasonal and locally sourced ingredients: Fresh, seasonal produce tends to have more vibrant flavors. Incorporating locally sourced ingredients...10. Offer customizable options: Allow customers to personalize their dishes by offering a variety of sauces, toppings, or condiments... |

plausible reason could be the framework’s limitation in identifying conflicting reviews.

4.4. AI guided hospitality domain attribute ranking

Fig. 8 shows the user’s and ChatReview’s ranking of generative AI-guided attributes of the hospitality domain. We observe some difference between the user and ChatReview rankings for this domain study. The MAE for this case is 3, implying a divergence of ± 3 ranks. Nonetheless, the two essential attributes identified by the framework are ranked among the top five by the user. These include ‘comfort’ and ‘responsiveness’. This observation indicates the framework’s reasonable performance in identifying the essential attributes. The accuracy,

precision, recall, and F1-score for this study are 0.5, 0.5, 0.6, and 0.54, respectively. The precision and recall values (≥ 0.5) indicate the framework’s ability to account for truly important attributes and not discard them.

In summary, this section discusses the framework’s ability to identify the essential attributes for a given survey/review data by comparing them with the user-identified essential attributes. This comparison is critical because a satisfactory identification of the essential attributes influences the quality of recommendations given by phase 4. The discussion also implicitly evaluates the performance of the NRClex and the all-MiniLM-L6-v2 sentence transformation models. They appear in phases 2 and 3, respectively, which collectively identify the essential

Table 8
The three prompts and their respective responses for the hospitality domain attribute: ‘comfort’.

| | Explicit | Implicit | Creative |
|----------|---|---|--|
| Question | How do I offer the most comfortable stay to my customers? | How do good hotels promote comfortable stay of their customers? | Give me creative ways to increase customer's comfort level at my hotel. |
| Response | <ol style="list-style-type: none">1. Provide a warm welcome: Greet your customers with a friendly smile and make them feel valued from the moment they arrive.2. Clean and well-maintained accommodations: Ensure that your rooms are spotlessly clean, well-maintained, and free from any unpleasant odors...3. Comfortable bedding: Invest in high-quality mattresses, pillows, and linens to ensure a good night's sleep for your guests. Offer a variety of pillow options to cater to different preferences.4. Temperature control: Maintain a comfortable temperature in the rooms, allowing guests to adjust the thermostat according to their preferences. Provide fans or heaters as needed.5. Soundproofing: Minimize noise disturbances by soundproofing rooms or providing earplugs for light sleepers. Ensure that common areas are also quiet and peaceful.6. Adequate lighting: Ensure that rooms are well-lit with a combination of natural and artificial lighting. Provide bedside lamps and reading lights for convenience.7. Privacy: Respect your guests' privacy by providing adequate window coverings, soundproof walls, and secure locks. Avoid unnecessary interruptions or intrusions.8. Amenities and facilities: Offer a range of amenities such as toiletries, towels, hairdryers, and ironing facilities...9. High-speed internet: Provide reliable and fast internet access throughout your property, as many guests consider it essential for both work and leisure. | <ol style="list-style-type: none">1. Quality bedding and amenities: Good hotels invest in high-quality mattresses, pillows, and linens to ensure a comfortable sleep experience...2. Efficient temperature control: Hotels maintain a comfortable temperature throughout the property, allowing guests to adjust the room temperature according to their preferences...3. Soundproofing: To minimize noise disturbances, hotels use soundproofing techniques such as double-glazed windows, thick walls, and carpeting...4. Excellent customer service: Good hotels prioritize exceptional customer service. They train their staff to be attentive, friendly, and responsive to guest needs...5. Cleanliness and maintenance: Hotels maintain high cleanliness standards by regularly cleaning rooms, common areas, and facilities...6. Comfortable seating and workspace: Hotels provide comfortable seating options in rooms and common areas, including ergonomic chairs and sofas...7. In-room amenities: These may include flat-screen TVs, mini-fridges, coffee makers, safes, and complimentary Wi-Fi...8. Quality dining options: Hotels often have on-site restaurants or room service options that offer high-quality food and beverages...9. Recreational facilities: Many hotels offer recreational facilities such as swimming pools, fitness centers, spas, or wellness areas...10. Attention to detail: This includes providing sufficient lighting, offering blackout curtains for better sleep, ensuring good water pressure in showers... | <ol style="list-style-type: none">1. Personalized welcome: Send a personalized welcome message or email to guests before their arrival, addressing them by name and expressing your excitement to have them stay at your hotel...2. Upgrade amenities: Offer complimentary amenities such as bathrobes, slippers, and high-quality toiletries to enhance the comfort level of your guests...3. Pillow menu: Provide a pillow menu with a variety of options, including memory foam, feather, or hypoallergenic pillows...4. Room customization: Allow guests to customize their room settings, such as adjusting the temperature, lighting, or even choosing the artwork displayed in their room...5. Technology convenience: Install smart devices in rooms, such as voice-controlled assistants or mobile apps, that allow guests to control various aspects of their room, such as temperature, lighting, or entertainment systems...6. Cozy common areas: Create cozy and inviting common areas within the hotel, such as a lounge with comfortable seating, a fireplace, or a reading nook...7. Wellness amenities: Offer wellness amenities such as a fitness center, spa, or yoga classes...8. Personalized recommendations: Provide personalized recommendations for local attractions, restaurants, or activities based on guests' preferences...9. Thoughtful gestures: Surprise guests with small, thoughtful gestures such as a welcome gift, a handwritten note, or a complimentary snack waiting in their room...10. Prompt and attentive service: Train your staff to provide prompt and attentive service, anticipating guests' needs and going the extra mile... |

attributes. Our discussion highlights the framework’s satisfactory performance in identifying the essential attributes. The analysis of final recommendations requires more exhaustive tools such as a review conducted by domain users with varying levels of expertise. We aim to perform this task in the future.

4.5. Peer assessment using bard

For this study, we use generative AI-guided domain studies pertaining to the hospitality and local business domains. We ask Google’s Bard to rate (on a scale of 1 to 5) the domain attributes and final recommendations generated by the ChatReview framework. We also present Bard’s qualitative analysis of ChatReview’s domain attribute selection and final recommendations. Section 4.5.1 motivates LLM-based peer assessment. Sections 4.5.2 and 4.5.3 present the Bard analyses for the hospitality and local business domains, respectively.

4.5.1. LLM peer assessment motivation

Because the large language models (LLMs) train using the existing body of knowledge and seek continuous improvement, we assert that techniques for effective human education can be easily adapted to

improve the LLMs. One such technique is student peer assessment wherein students provide an unbiased opinion of their peer’s work (Yin, Chen, & Chang, 2022). We adapt this technique to evaluate ChatReview using Google Bard as the peer reviewer. It is worth noting that this assessment is unlike conventional studies that perform one-on-one comparisons between similar models and frameworks. We assert that this peer review has four major benefits: (1) it highlights any ‘knowledge gap’ within the ChatReview framework. Google Bard may have significantly more data access versus ChatGPT whose training data is several months behind and never current, (2) it enables users to develop a level of confidence in the LLM’s output, (3) Google Bard is justifiably less biased than a human reviewer. Its reviews may be more comprehensive, accounting for a wider perspective versus a human reviewer, and (4) Its review process is significantly faster than a human reviewer.

4.5.2. Hospitality domain

To qualify the domain attributes identified by ChatReview’s AQ generation engine, we input the domain attributes listed in Table 4 into Bard and use the prompt: ‘On a scale of 1 to 5, rate the importance of these attributes for a highly successful hotel.’ Fig. 9 provides the

Table 9
The three prompts and their respective responses for the local business domain attribute: ‘marketing’.

| | Explicit | Implicit | Creative |
|----------|--|---|--|
| Question | How can I market my restaurant? | How do good restaurants market themselves? | Give me creative ways to market my local restaurant. |
| Response | <ol style="list-style-type: none">1. Develop a strong online presence: Create a professional website for your restaurant, optimize it for search engines, and ensure it is mobile-friendly...2. Online food delivery platforms: Partner with popular food delivery services like Uber Eats, Grubhub, or DoorDash to expand your reach and attract customers...3. Local SEO: Optimize your website and online listings with relevant keywords, location-specific information, and positive customer reviews...4. Offer promotions and discounts: Attract new customers and encourage repeat visits by offering special promotions, discounts, or loyalty programs...5. Collaborate with influencers: Partner with local influencers or food bloggers who have a significant following in your area...6. Host events or themed nights: Organize special events, such as live music nights, trivia nights, or themed dinners, ...7. Engage with the local community: Participate in local events, sponsor local sports teams or charities, and collaborate with neighboring businesses...8. Utilize email marketing: Collect customer email addresses and send regular newsletters or promotional emails to keep them informed about new menu items, upcoming events, or exclusive offers.9. Encourage customer reviews: Encourage satisfied customers to leave positive reviews on platforms like Google, Yelp, or TripAdvisor...10. Monitor and respond to feedback: Regularly monitor online reviews and social media mentions of your restaurant. Respond promptly and professionally... | <ol style="list-style-type: none">1. Online presence: Establishing a strong online presence is crucial in today's digital age. Restaurants create professional websites with menus, photos, and contact information...2. Online reviews and ratings: Positive reviews and high ratings on platforms like Yelp, Google, and TripAdvisor can significantly impact a restaurant's reputation...3. Local advertising: Restaurants often advertise in local newspapers, magazines, and radio stations to reach a wider audience...4. Loyalty programs: Many restaurants offer loyalty programs to reward repeat customers. These programs often include discounts, freebies, or exclusive offers, encouraging customers to return and refer others.5. Special events and promotions: Hosting special events like wine tastings, live music nights, or themed dinners can attract new customers and create buzz...6. Partnerships and collaborations: Collaborating with other local businesses, such as wineries, breweries, or food suppliers, can help restaurants expand their reach...7. Email marketing: Collecting customer email addresses and sending regular newsletters or promotional offers can keep customers engaged and informed about upcoming events or new menu items.8. Influencer marketing: Partnering with influencers or food bloggers who have a significant following can help restaurants gain exposure...9. Community involvement: Participating in local events, sponsoring sports teams, or supporting charitable causes can enhance a restaurant's reputation and build a positive image within the community.10. Excellent customer service: Word-of-mouth is a powerful marketing tool... | <ol style="list-style-type: none">1. Host a themed event night: Organize a special themed night at your restaurant, such as a trivia night, karaoke night, or live music night. Promote it through social media...2. Collaborate with local influencers: Partner with local bloggers, food critics, or social media influencers who have a significant following in your area...3. Offer unique promotions: Create exclusive promotions that are only available at your restaurant. For example, you could offer a “secret menu” item that customers can only order if they mention a specific phrase or code word...4. Organize cooking classes or workshops: Share your culinary expertise by hosting cooking classes or workshops at your restaurant...5. Participate in local events: Get involved in community events, festivals, or charity fundraisers...6. Create a loyalty program: Implement a loyalty program that rewards customers for their repeat visits. Offer incentives such as discounts, free appetizers, or exclusive access to special events...7. Utilize social media creatively: Use platforms like Instagram, Facebook, and Twitter to showcase your restaurant's ambiance, dishes, and behind-the-scenes moments...8. Collaborate with other local businesses: Partner with nearby businesses, such as local breweries, wineries, or dessert shops, to create special promotions or cross-promote each other...9. Offer delivery or takeout options: In today's fast-paced world, many people prefer the convenience of ordering food to their doorstep...10. Create a visually appealing outdoor seating area: If your restaurant has outdoor seating, invest in creating an inviting and visually appealing space... |

ratings of the ten attributes selected by the ChatReview framework. As seen in the figure, the domain attributes including cleanliness and hygiene, comfort, efficient staff, and responsiveness to customers obtain the highest rating equal to 5. The four attributes including convenient location, basic facilities, good food, and efficient check-in obtain a rating of 4 out of 5. Bard gives the lowest rating of 3 out of 5 to ambiance and pricing attributes. Bard provides the following justification for its rating: ‘Cleanliness and hygiene, comfort, efficient staff, and responsiveness to customers are all essential for a highly successful hotel. Convenient location, basic facilities, good food, and efficient check-in are also important, but they are not as essential as the first four attributes. Ambiance and pricing are the least important attributes, but they can still play a role in attracting and retaining guests.’ Bard collectively rates the attributes as 4.4 out of 5 (88% overall rating), highlighting an effective performance of the ChatReview framework.

We use Bard to assess the quality of the creative responses generated by ChatReview for the ‘comfort’ attribute. Specifically, we input the ten creative responses in Table 8 into Bard and use the prompt: ‘Rate these suggestions to improve the customer’s comfort level in a stay hotel, keeping their creativity in mind.’ Bard provides a 5 out of 5 to all of

the ten suggestions (100% overall rating), highlighting the framework’s effectiveness in generating recommendations. Bard provides the following justification for its rating: ‘All of your suggestions are creative and effective ways to improve the customer’s comfort level in a stay hotel. They are all well-thought-out and show a deep understanding of the needs and wants of hotel guests.’

4.5.3. Local business domain

To evaluate the quality of domain attributes generated by ChatReview’s AQ generation engine, we input the generated domain attributes to Bard and use the prompt: ‘On a scale of 1 to 5, rate the importance of these attributes for a highly successful local restaurant.’ Fig. 10 provides Bard’s rating to the local business domain attributes given in Table 5. As seen in the figure, Bard rates the four attributes (hygiene, menu, service, good food, and responsiveness to customers) as 5/5; one attribute (source and sustainability) as 4/5, three attributes (marketing, efficient operation, and ambiance) as 3/5, and community connection as 1/5. Bard collectively rates them as 4.5 out of 5 (90% overall rating), implying a good LLM performance. Bard provides the following

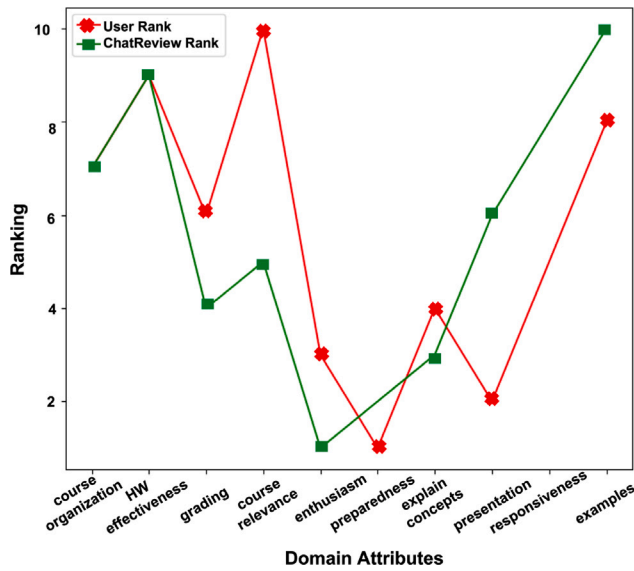


Fig. 6. The attribute ranking comparison between a user and ChatReview for the user-guided education domain study.

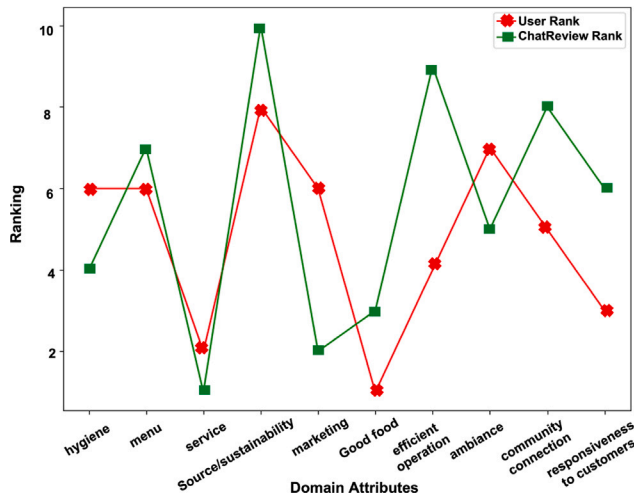


Fig. 7. The attribute ranking comparison between a user and ChatReview for the generative AI-guided local business domain study.

justification for its rating: ‘Hygiene, menu, service, good food, and responsiveness to customers are all essential for a highly successful local restaurant. Sourcing and sustainability, efficient operation, ambiance, and community connection are also important, but they are not as essential. Marketing is the least important attribute, but it can still help attract new customers.’

We evaluate the ten creative prompt responses generated by the ChatReview framework for the marketing attribute. These responses appear in Table 9. Specifically, we input the ten creative prompt responses to Bard and use the prompt: ‘Rate these suggestions to improve the marketing of a local restaurant, keeping their creativity in mind’. Bard’s response to this prompt is as follows: ‘Overall, all of the marketing suggestions you have provided are creative and effective ways to promote a local restaurant. I would rate them all 4/5 or 5/5, except hosting a themed event night and participating in local events, which I would rate 4/5.’ The above Bard’s response yields an overall Bard rating equal to 4.8 (96%), implying ChatReview’s excellent performance in offering creative suggestions.

In summary, this section introduces a unique approach to evaluating the effectiveness of large language models (LLMs) using a peer

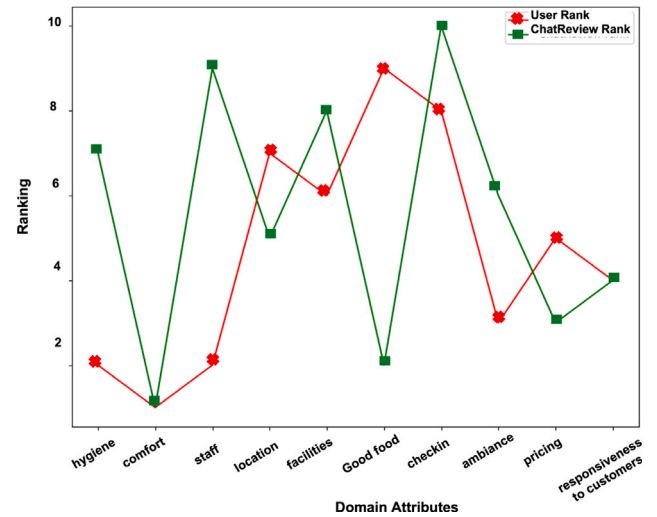


Fig. 8. The attribute ranking comparison between a user and ChatReview for the generative AI-guided hospitality domain study.

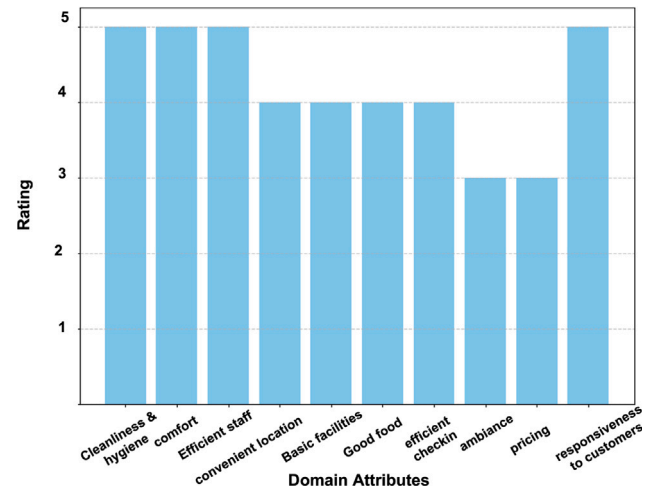


Fig. 9. The Bard rating of domain attributes identified by ChatReview for the generative AI-guided hospitality domain study.

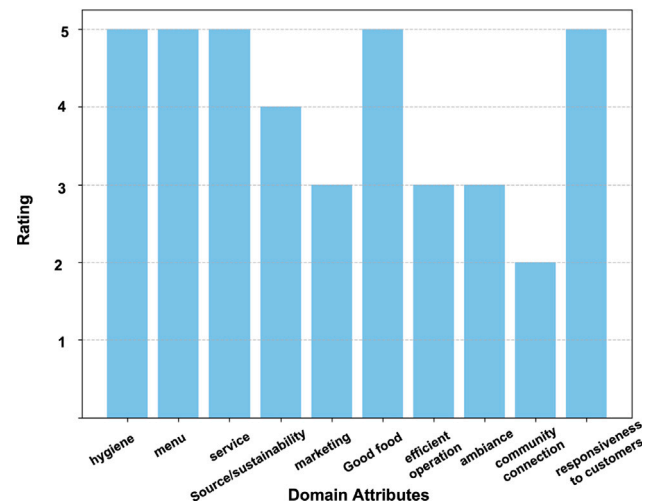


Fig. 10. The Bard rating of domain attributes identified by ChatReview for the generative AI-guided local business domain study.

assessment approach. We employ Bard as the peer evaluator of the ChatReview framework. ChatReview receives Bard rating of over 90% for its ability to identify relevant domain attributes and provide effective recommendations.

5. Related work

There is a widespread interest in the application of natural language processing (NLP) and large language models (LLMs) to domain-specific studies. In what follows, Section 5.1 expounds on the recent advances in the application of NLP for sentiment analysis on surveys originating from the education and business domains. Section 5.2 discusses the most recent efforts that apply ChatGPT for survey analysis.

5.1. NLP-based survey analysis in education and business

Rani and Kumar (2017) present a sentiment analysis system that employs NRC Emotion Lexicon (EmoLex) to identify sentiments and emotions embedded in student surveys. Using the EmoLex sentiments, they compute students' satisfaction and dissatisfaction with the instructor. They evaluate their system by correlating the identified emotions with the overall course rating. The sentiment analysis presented by Rani and Kumar (2017) is similar to our framework's tokenization phase. However, their system is limited to sentiment identification only, leaving the rest of the analysis to users. In contrast, the ChatReview framework identifies the embedded sentiments, connects them with the domain attributes (classroom teaching, presentation, concept explanation, etc.), and researches ideas for improvement.

Jiménez, Finamore, Casanova, and Simões (2021) perform a student survey research using the BERT model to assess the COVID-19 impact on higher education. They only focus on the polarity classification task and classify the reviews into positive, negative, and neutral categories. They test a variety of tuning strategies to improve the modeling performance. Similar to Rani and Kumar (2017), their work is limited to sentiment analysis only.

Shaik et al. (2023) present a survey analysis of multiple machine learning (ML) and NLP techniques to mine opinions from educational data. They explore emotion analysis, sentiment annotation, machine learning techniques, and the impact of sentiment analysis on procedures to improve pedagogy. Their analysis involves the use of complex NLP and ML models that are inaccessible to general users. The literature reports some other works that use unique techniques for analysis. Abdi et al. (2023) apply deep learning techniques to perform student feedback analysis. In contrast to our work, the above two efforts also focus on sentiment extraction and do not address methods to generate domain-specific solutions. Mouronte-López, Ceres, and Columbrans (2023) also perform sentiment analysis in the education domain. They analyze Twitter feeds to study the student perception of the education system. Unlike our study which aims to provide domain-specific solutions for users, their work focuses more on the sociological aspects.

The NLP studies are gaining significant attention in restaurant and hospitality domains. Asani, Vahdat-Nejad, and Sadri (2021) use the SentiWordNet dictionary (Denecke, 2008) to extract the sentiments embedded in the user's preference. They develop a recommender system that matches the user's preferences with the previous users to recommend good restaurants. Their primary stakeholders are users, whereas our ChatReview framework's target users are the business owners. van Leeuwen and Koole (2022) present an unsupervised machine learning method to cluster users based on their attributes and develop marketing strategies. They employ hierarchical clustering for this task. Similar to our framework's phase 3, their approach identifies the important attributes for marketing. However, they use predefined strategies versus using LLMs that can retrieve significant and potentially unique approaches.

5.2. Survey analysis using ChatGPT

Given that ChatGPT's popularity has risen in the last few years, the literature reports a few studies that employ this tool to promote quality of service in different fields. Wang, Xie, Ding, Feng, and Xia (2023) evaluate ChatGPT 3.5's ability to analyze sentiments in four areas including standard evaluation, polarity shift, open-domain evaluation, and sentiment inference evaluation. They conclude that ChatGPT has significant inference capabilities. Koc, Hatipoglu, Kivrak, Celik, and Koc (2023) conduct a similar study where they feed the customers survey into ChatGPT 4. They observe quality management responses given by ChatGPT 4 within seconds. The above studies involve direct data feed into the GPT tool, which may compromise data privacy for highly sensitive domains (e.g. government, hospitals, etc.). The research has shown that trust is a significant factor in the widespread adoption of ChatGPT (Choudhury & Shamszare, 2023). To protect user privacy as much as possible, the ChatReview framework does not feed the survey results into the GPT tool. Our framework's tokenization and query construction phases create a layer of abstraction, thereby obscuring sensitive data including the names of entities, gender, and racial identification from the GPT tool.

6. Conclusion

6.1. Summary

We present ChatReview, an NLP framework to analyze user survey data emanating from multiple domains and offer insights to improve the quality of service. The framework comprises multiple phases where each phase provides insights into the survey data while promoting user privacy.

In the first phase, data collection, the ChatReview framework removes any instances that identify the stakeholders, promoting data privacy and reducing bias. In the second phase, tokenization, the framework extracts words (tokens) from the survey that relate to the ten most common sentiments, creating a token repository. This token collection offers users a preliminary insight into user sentiments embedded in the large survey data. In the third phase, query construction, the framework performs three vital tasks. First, it generates important attributes that are crucial to the domain. The framework performs this task either via the user's guidance or by using generative artificial intelligence (AI). The user-guided attribute generation allows for human autonomy, offering users flexibility in customizing their study. The generative AI-guided attribute generation allows users to fill gaps in their domain knowledge while spending minimal time in conducting research. Second, the framework maps the tokens from the repository to the domain attributes, thereby identifying areas for improvement. Third, the framework formulates three types of prompts including explicit, implicit, and direct prompts. In the response generation phase, the framework employs ChatGPT to generate responses in varying levels of granularity, providing users with a holistic and deep insight into their application domain.

We test our framework using three real-world domains including education, local business, and hospitality. We assert that ChatReview is effective in analyzing surveys and offering insights in all four phases. Our studies also find that ChatReview is adequate in replicating human analyses. While this tool does not replace a human, we advocate that it can serve as a valuable assistant for humans in analyzing large survey data.

6.2. Limitations and future work

We also identify several limitations of the ChatReview framework that beget significant future research activities, all of which are critical to our social fabric. The current ChatReview framework is gender and race-neutral. However, gender and racial biases embedded in the

reviews greatly skew the analysis (Chavez & Mitchell, 2020; Davis, Worsnop, & Hand, 2022; Kamalahmadi & Yu, 2023). For instance, research (Kreitzer & Sweet-Cushman, 2022) has shown that female educators are typically reviewed critically versus their male counterparts. Given its fluidity, the gender-based analysis is highly crucial. The racial biases in the reviews pose a similar problem. A burning question to be addressed in the future is, “How can AI identify such biases, mitigate their effects, and offer insights effectively?”

ChatReview explores three specific prompting techniques including implicit, explicit, and creative prompting. Future work includes delving deeper into prompt engineering and testing other methods such as zero-shot prompting explored by Fatouros, Soldatos, Kouroumalis, Makridis, and Kyriazis (2023). We assert that this study is pivotal in shaping the future of generative AI and its widespread adoption.

Future work includes using domain-guided deep learning techniques with NLP (for instance, Zhang et al., 2023 use physics-informed deep learning). We also aim to explore other evaluation instruments such as a thorough user survey after seeking appropriate institutional review board (IRB) approval. We aim to apply these methods to further highlight ChatReview's strengths and mitigate its weaknesses. We uploaded the source code and the data for the general public use at: <https://github.com/vkpallipuram/ChatREVIEW>.

CRediT authorship contribution statement

Brittany Ho: conceptualization, Software, Testing, Reviewing. **Ta'Rhonda Mayberry:** Framework evaluation, Reviewing. **Khanh Linh Nguyen:** Framework evaluation, Reviewing. **Manohar Dhulipala:** Data collection, Testing. **Vivek Krishnamani Pallipuram:** Supervision, Conceptualization, Verification, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank University of the Pacific's Undergraduate Research Office, USA for supporting this research via the Summer Undergraduate Research Fellowship (SURF) 2023 award.

References

- Abdi, A., Sedrakyan, G., Veldkamp, B., et al. (2023). Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems. *Soft Computing*, 27, 14073–14094. <http://dx.doi.org/10.1007/s00500-023-07926-2>.
- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3, <http://dx.doi.org/10.3389/frai.2020.00042>.
- Asani, E., Vahdat-Nejad, H., & Sadri, J. (2021). Restaurant recommender system based on sentiment analysis. *Machine Learning with Applications*, 6, Article 100114. <http://dx.doi.org/10.1016/j.mlwa.2021.100114>.
- Caramancion, K. M. (2023). Harnessing the power of ChatGPT to decimate mis/disinformation: Using chatgpt for fake news detection. In *2023 IEEE world AI IoT congress* (pp. 0042–0046). <http://dx.doi.org/10.1109/AIoT58121.2023.10174450>.
- Chavez, K., & Mitchell, K. M. (2020). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science*, 53(2), 270–274. <http://dx.doi.org/10.1017/S1049096519001744>.
- Choudhury, A., & Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis. *Journal of Medical Internet Research*, 25, Article e47184. <http://dx.doi.org/10.2196/47184>.
- Church, K. W. (2017). Word2Vec. *Natural Language Engineering*, 23(1), 155–162. <http://dx.doi.org/10.1017/S1351324916000334>.
- Davis, S. R., Worsnop, C. J., & Hand, E. M. (2022). Gender bias recognition in political news articles. *Machine Learning with Applications*, 8, Article 100304. <http://dx.doi.org/10.1016/j.mlwa.2022.100304>.
- Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. In *2008 IEEE 24th international conference on data engineering workshop* (pp. 507–512). <http://dx.doi.org/10.1109/ICDEW.2008.4498370>.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv abs/1810.04805*.
- Fatouros, G., Soldatos, J., Kouroumalis, K., Makridis, G., & Kyriazis, D. (2023). Learning sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, 14, Article 100508. <http://dx.doi.org/10.1016/j.mlwa.2023.100508>.
- Gurman, M. (2023). *Samsung Bans Staff's AI Use after spotting ChatGPT data leak*. Bloomberg, URL: <https://www.bloomberg.com>.
- Horizons, D. S. (2023). *Mastering generative AI and prompt engineering: A practical guide for data scientists: Technical report*.
- Jansen, B. J., gyo Jung, S., & Salminen, J. (2023). Employing large language models in survey research. *Natural Language Processing Journal*, 4, Article 100020. <http://dx.doi.org/10.1016/j.nlp.2023.100020>.
- Jiménez, H. G., Finamore, A. C., Casanova, M. A., & Simões, G. (2021). Sentiment analysis of student surveys - A case study on assessing the impact of the COVID-19 pandemic on higher education teaching. *International Educational Data Mining Society*.
- Kamalahmadi, M., & Yu, Y. (2023). Racial and gender biases in customer satisfaction surveys: Evidence from a restaurant chain. 53(2), <http://dx.doi.org/10.2139/ssrn.4418420>.
- Koc, E., Hatipoglu, S., Kivrak, O., Celik, C., & Koc, K. (2023). Houston, we have a problem! The use of ChatGPT in responding to customer complaints. *Technology in Society*, 74, Article 102333. <http://dx.doi.org/10.1016/j.techsoc.2023.102333>.
- Kreitzer, R., & Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: a review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, 20, 73–84. <http://dx.doi.org/10.1007/s10805-021-09400-w>.
- Lu, Y., Wu, H., Qi, S., & Cheng, K. (2023). Artificial intelligence in intensive care medicine: Toward a chatgpt/GPT-4 way. *Annals of Biomedical Engineering*, 51, <http://dx.doi.org/10.1007/s10439-023-03234-w>.
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436–465. <http://dx.doi.org/10.1111/j.1467-8640.2012.00460.x>.
- Mouronte-López, M., Ceres, J., & Columbrans, A. (2023). Analysing the sentiments about the education system through Twitter. *Educational Information Technology*, 28, 10965–10994. <http://dx.doi.org/10.1007/s10639-022-11493-8>.
- Oesper, L., Merico, D., Isserlin, R., & Bader, G. D. (2011). WordCloud: a cytoscape plugin to create a visual semantic summary of networks. *Source Code for Biology and Medicine*, 6(1), 7.
- Ozkaya, I. (2023). Application of large language models to software engineering tasks: Opportunities, risks, and implications. *IEEE Software*, 40(3), 4–8. <http://dx.doi.org/10.1109/MS.2023.3248401>.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pretraining. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
- Rani, S., & Kumar, P. (2017). A sentiment analysis system to improve teaching and learning. *Computer*, 50(5), 36–43. <http://dx.doi.org/10.1109/MC.2017.133>.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing*. Association for Computational Linguistics.
- Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2, Article 100003. <http://dx.doi.org/10.1016/j.nlp.2022.100003>.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.
- van Leeuwen, R., & Koole, G. (2022). Data-driven market segmentation in hospitality using unsupervised machine learning. *Machine Learning with Applications*, 10, Article 100414. <http://dx.doi.org/10.1016/j.mlwa.2022.100414>.
- Wang, Z., Xie, Q., Ding, Z., Feng, Y., & Xia, R. (2023). Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv:2304.04339*.
- Yin, S., Chen, F., & Chang, H. (2022). Assessment as learning: How does peer assessment function in students' learning? *Frontiers in Psychology*, 13, <http://dx.doi.org/10.3389/fpsyg.2022.912568>.
- Zhang, J., Zhao, Y., Shone, F., Li, Z., Frangi, A. F., Xie, S. Q., et al. (2023). Physics-informed deep learning for musculoskeletal modeling: Predicting muscle forces and joint kinematics from surface EMG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 484–493. <http://dx.doi.org/10.1109/TNSRE.2022.3226860>.