From Guest to Family: An Innovative Framework for Enhancing Memorable Experiences in the Hotel Industry

Abdulaziz Alhamadani*, Khadija Althubiti‡, Shailik Sarkar*, Jianfeng He*, Lulwah Alkulaib* §,
Srishti Behal*, Mahmood Khan‡, and Chang-Tien Lu*
* Department of Computer Science, Virginia Tech, Falls Church, VA 22043 USA
‡ Department of Hospitality and Tourism Management, Virginia Tech, Falls Church, VA 22043 USA
§ Department of Computer Science, Kuwait University, Kuwait
{hamdani, kalthubiti, shailik, jianfenghe, lalkulaib, behalsrishti08, Mahmood, ctlu}@vt.edu

***Abstract*—This paper presents an innovative framework developed to identify, analyze, and generate memorable experiences in the hotel industry. People prefer memorable experiences over traditional services or products in today's ever-changing consumer world. As a result, the hospitality industry has shifted its focus toward creating unique and unforgettable experiences rather than just providing essential services. Despite the inherent subjectivity and difficulties in quantifying experiences, the quest to capture and understand these critical elements in the hospitality context has persisted. However, traditional methods have proven inadequate due to their reliance on objective surveys or limited social media data, resulting in a lack of diversity and potential bias. Our framework addresses these issues, offering a holistic solution that effectively identifies and extracts memorable experiences from online customer reviews, discerns trends on a monthly or yearly basis, and utilizes a local LLM to generate potential, unexplored experiences. As the first successfully deployed, fast, and accurate product of its kind in the industry, This framework significantly contributes to the hotel industry's efforts to enhance services and create compelling, personalized experiences for its customers.**

***Index Terms*—Hotel industry, Memorable Experience, Keyword Extraction, Text Generation, Social media data mining**

I. Introduction

In 1998, Pine and Gilmore introduced the "experience economy" concept to describe the emerging consumer demand for experiences over products and services [45]. They observed that customers were no longer satisfied with simply buying services; they sought to buy unique, memorable experiences. Hence, companies in today's competitive market should go beyond the function and upgrade their offerings to provide an experience.

ASONAM '23, November 6–9, 2023, Kusadasi, Turkiye

ACM ISBN 979-8-4007-0409-3/23/11…$15.00
https://doi.org/10.1145/3625007.3632331

Csikszentmihalyi emphasized the significance of experiences in generating a profound sense of pleasure that leads to favorable memories [14]. Creating a memorable experience (ME) is taking root in the hospitality industry. The industry is known as a customer-centric industry; therefore, its products are highly experience-oriented [35]. The shift from service delivery to the experience creation industry began as researchers realized that strategies focusing solely on service, quality, and price are no longer the primary drivers of competitive advantage, and staging an experience to be a ME has become increasingly important within the core capabilities of hotels, for instance, and plays a crucial role in determining whether guests will choose to revisit [24], [39], [40]. Today, destination managers and tourism businesses must strive to provide memorable experiences as a new standard to meet since travelers now desire authentic and meaningful experiences that cater to their leisure and spiritual desires [32].

Experiences, as Pine and Gilmore [45] stated, are "inherently personal, existing only in the mind of an individual who has been engaged on an emotional, physical, intellectual, or even spiritual level." The mix of distinctively subjective factors makes experience hard to be quantified or measured accurately [12], [16]. However, researchers in hospitality and tourism have attempted to define the experience construct in the context of their field (tourist's experience, traveler experience, guest experience) and explore the link between tourism experiences and memory by applying different methods to measure it. For example, Kim and Chen [31]characterize tourist experiences as intangible, unique, ongoing, and highly subjective occurrences, and these experiences can be understood from two perspectives: the immediate, moment-to-moment encounters and the overall assessment of the experience. Vada et al. [64] define a memorable tourism experience (MTE) as a positive encounter that is retained and remembered by individuals even after the actual event has taken place. Seyfi et al. [51]suggest that the quality of the experience is a stronger predictor of creating MEs than the quality of service. It is a challenging task to identify MEs, but those defined criteria can help in capturing customers' memorable experiences.

The traditional methods used to measure tourists' MEs in hospitality and tourism include questionnaire surveys, self-reported diaries, interviews, observant participation, and the employment of the experiential sampling method [28]. However, in more recent times, memorable experience research has expanded to incorporate innovative techniques such as social media analytics [54]. Current efforts have been limited to objective surveys or a small portion of social media data, making it hard to generalize its results due to bias and lack of diversity.

To address the previously discussed challenges and, most importantly, to efficiently deploy a framework that can identify memorable experiences within the hotel industry on a real-world, large-scale platform, we propose **From Guest To Family (G2F)**. In this paper, we detail the development of our platform and deploying it as an asset product for hotel industry management to enhance their hotel services and stay ahead of the industry. The development of G2F consists of three main steps: 1) efficiently identifying the reviews that include positive or negative MEs, which is based on the K-Means algorithm to cluster the reviews based on their sentiment scores and user reviews' rating; (2) extracting the representative keywords that are trending for MEs in a monthly or yearly pattern based on an advanced keywords extraction algorithm; (3) novel and unexplored text generation of reviews that include MEs based on local Large Language Model (LLM) and the extracted keywords. To the best of our knowledge, the proposed G2F is the first successfully deployed fast and accurate product in capturing, analyzing, and generating MEs in the hotel industry. Our contributions are summarized as follows:

- We have created a comprehensive platform capable of distinguishing and extracting both positive and negative MEs from online customer reviews within the hotel industry.
- Our platform also provides an analytical tool that uncovers trends in MEs on a monthly or yearly basis, thereby enabling hotel management to identify unique or recurrent key terms to improve their services.
- Lastly, we've built a platform that leverages a local LLM and the extracted key terms to generate potential yet unexplored MEs, assisting hotel management in anticipating and preparing for future scenarios.

II. Related Work

A. Memorable experiences in Hospitality

Tourism, often seen as a journey of experiences, involves various elements like accommodation, local interaction, transportation, attractions, and culinary experiences [23]. Studies by [63] and [32] focused on the link between tourism experiences and memory, the elements that make experiences memorable, and the conceptualization of the term "memorable tourism experience." They identified key dimensions and developed a scale encompassing hedonism, refreshment, local culture, meaningfulness, knowledge, involvement, and novelty. While these focused primarily on positive experiences, more recent studies have started considering negative experiences as potentially memorable components [56].

Despite these developments, there is still a lack of consensus on theories and measurement of the concept. The scales used are often seen as inadequate in capturing the true essence of what makes a tourism experience memorable. Most studies have utilized close-ended surveys, interviews, open-ended questionnaires, and travel blog narrative analysis, with few drawing on content analysis [55]. The need for more comprehensive and updated research on the topic is palpable, as several scholars urge for further studies to deepen our understanding of memorable tourism experiences [23], [32], [55].

B. Keywords Extraction

Automatic Keyword Extraction (AKE) is designed to quickly and efficiently identify a small yet representative set of words that accurately reflect the key topics in a text document without requiring time-consuming manual annotation by experts [66]. Different terminologies are used to describe the most significant information extracted from a text, such as key phrases, key segments, key terms, or keywords extraction. However, they all serve the same purpose [5]. Past efforts in KE techniques are mainly supervised or unsupervised. Supervised methods for keyword extraction typically require a substantial labeled training dataset to achieve high performance, making them often limited to specific domains. Consequently, unsupervised methods such as TextRank [41], Yake [9], EmbedRank [6], SIFRank [59], AttentionRank [15], and MDERank [71] have emerged as widely adopted and robust alternatives. Starting with TextRank [41], a graph-based approach, KE algorithms have continuously evolved to address the limitations of previous methods, resulting in improved accuracy, efficiency, and relevance of extracted keywords. Based on the best of our knowledge of the current state-of-the-art (SOTA) works, both SIFRank [59] and MDERank [71] demonstrate notable strengths in terms of robustness, efficiency, and keyword relevance.

KE is widely used across multiple domains and industries for various purposes such as tracking research trends [52], analyzing pandemic trends [65], or improving educational methods [17]. In the hospitality and tourism domain, Le Huy et al. [36] suggested a KE approach based on BilSTM-CRF combined with BERT for effectively extracting key phrases related to information and search methods in the field of tourism. A different study [37] developed a tool called VisTravel that used the TextRank [41] to identify and extract essential words from travel reviews, enabling the tourism management team to gain insights into customers' opinions. Additionally, this study [68] introduced an online hotel review analysis using KE based on TF-IDF algorithm to extract the top 20 keywords that reflect the most concerning factors of hotel consumers on hotel services. Chang et al. [10] also developed a visual analytics framework for exploring insights from hotel ratings and reviews. They incorporated their keyword extraction method

by integrating a sentiment-based model learned through SVM. While keyword extraction techniques have been applied in the tourism and hotel industries, previous studies have ignored high-accuracy or SOTA keyword extraction methods, which offer more relevant, accurate keywords and have efficient performance and robustness in various lengths of keyphrases [71]. Consequently, those results are affected by the limitations of low-accuracy KE methods.

C. Large Language Models for Text Generation

Text generation and *text summarization* share the common goal of producing coherent and comprehensible texts tailored to individual users' needs. Text summarization creates concise summaries of longer documents or texts. There are two types of summaries: *Extractive*, which assembles summaries from the source text [29], [48], [72], and *Abstractive* which generates summaries that contain novel words to simulate human summaries [20], [44], [50]. Various methods have been proposed to help travelers choose hotels. Hu et al. [25] used extractive summarization to identify informative sentences based on author reliability, review time, usefulness, and conflicting opinions. Tsai et al. [62] created high-quality summaries by identifying helpful reviews and categorizing sentences into location, sleep quality, room, service, value, and cleanliness. Nathania et al. [22] developed a tool to generate paragraph and phrase-based summaries and analyze annual sentiment trends. However, these methods may lack coherence and novelty and be limited to the original text.

To overcome those limitations, text generation (Abstractive summarization is only a specific form) can create new text from scratch or based on a given prompt or input. It can be achieved by using many techniques, but our focus here is on Large Language Models (LLMs) such as GPT-3 [8] and LLaMA [60]. There are many applications of those LLMs in many fields such as education [19], [38], [49], [70], healthcare [18], [34], [46], finance [1], [3], [69], Law [7], [42], [61], software development [53], [58], and scientific research [11], [33], [57]. To the best of our knowledge, text generation in LLMs has not yet been implemented in generating MEs in the hotels industry.

III. Methodology

The development of G2F consists of three main parts, as illustrated in Figure 1. The first step is identifying the most positive and negative MEs from the reviews through sentiment analysis and K-Means clustering methods. The second step is extracting the yearly and monthly most representative keywords of the positive and negative MEs based on implementing the advanced keywords extraction algorithm. This step enables G2F to conduct a yearly or monthly analysis of the positive and negative MEs' keywords by facilitating the implementation of keywords distribution of the extracted keywords. In the third step, given the yearly/monthly extracted words, we customize four prompts to be inputted into a local open-source text generation LLM (Vicuna) to obtain different and unexplored positive and negative MEs in the hotel industry. Each prompt contains a concatenation of the top 20 extracted keywords, 500 random extracted keywords, and a positive/negative customized prompt. G2F details are discussed in the following subsections.

Algorithm 1 Proposed Method of G2F Framework

Input: HotelRec Data HTA_N^{attr}
Output: Hotels-Memorable-Experiences Data HME_O^M
Initialize: K-clusters Optimal value k = 0, i = 0

while $i < N$ **do**
  $ApplySentimentAnalysis(HTA_i^{Text})$
  $AppendSentimentScores(HTA_i^{pos,neg,neu,com})$
end while
procedure K-Means($HTA_i^{pos,neg,neu,com,rating}$)
  Find and assign optimal value k for clusters
  Assign and append cluster labels to each row HTA_N^{clu}
  for each Top and Lowest clusters in HTA_N^{attr} **do**
    Preprocess Text HTA_N^{text}
    Apply KE Algorithm to Preprocessed Text
    Group cluster rows by <year,month>
    Calculate Keywords Distribution by <year,month>
    $LocalVicuna \leftarrow Prompt(CustomizedText + keywods)$
    Store Keywords for each Cluster and Generated Memorable Experiences Texts to HME_O^M
  end for
end procedure
Return HME_O^M

A. Memorable Experiences Identification from Reviews

The main goal of G2F is to extract MEs from customer reviews. Most hotel research (see sec II) considers only positive MEs, such as wedding days at hotels or great local food on a fantastic view. However, MEs can also be highly negative, like a toilet clog that causes a pungent odor to the room, and the hotel management does nothing about it. Each positive or negative experience has a long-term effect, such as revisiting or never booking there again. G2F considers both positive and negative experiences by considering the most positive and negative hotel reviews to extract those experiences.

Relying only on the reviewers' rating scores from 1 (the lowest) to 5 (the highest) is not sufficient or reliable to distinguish the reviews from the most positive or negative reviews. For example, a customer may be having a bad day and gives a rating of 1 to a hotel that does not offer discounts or cash. On the other hand, a rating of 5 can be given to a biased customer who compliments how the hotel is clean because he knows someone there. Therefore, we corroborated the rating score of the reviews with the sentiment analysis scores to have more reliable scores for the most positive and negative reviews. We employ VADER [27] for sentiment

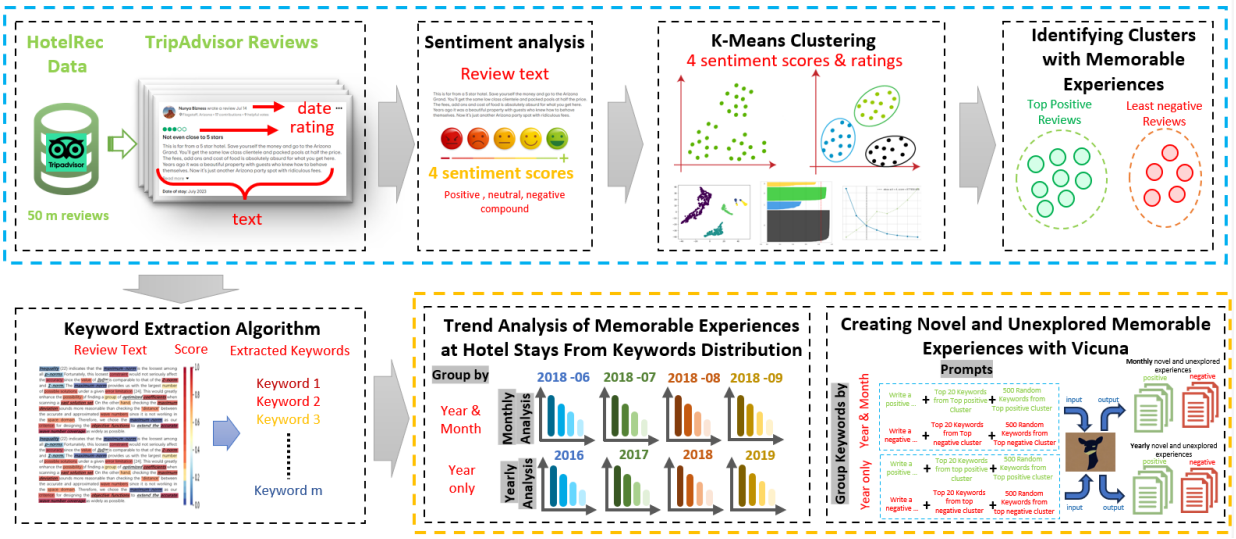

Fig. 1: The illustrative architecture of the proposed G2F framework.

analysis, a lexicon comprised of human-annotated phrase-emotion pairs. VADER returns three emotion scores (positive, negative, neutral) ranging from 0 to 1 and a compound ranging from -1 to 1 score for a given text input.

G2F utilizes the reviews' sentiment analysis scores and ratings to implement the K-Means clustering algorithm to group the reviews with the most positive and negative sentiment scores and ratings. In this work, we employed K-means over other clustering methods, such as Hierarchical clustering and DBSCAN or GMM, because K-means is computationally more efficient on a large-scale dataset, and other methods, such as DBSCAN, operates on the concept of density-based clustering where a point that does not belong to the density neighborhood of any clusters can be regarded as noise which is not ideal for us as we are interested in including the extreme instances in our cluster-based analysis. Consequently, we can focus on the top positive and negative clusters to extract MEs. The K-Means clustering algorithm divides N reviews in $attr$ dimensions into K clusters, then minimizes the sum of squares of the distances between each $r \in N$ review within each cluster. Given a the set of N reviews represented as $R = \{R_1, R_2, \cdots, R_n\}$ where each review in R is a $attr-$dimensional vector, the algorithm aims to divide the R reviews into $(k \leq n)$ clusters. $C = \{C_1, C_2, \cdots, C_k\}$, then optimize the sum of squares from each review to the centroid of its cluster. The problem is formulated as follows:

$$arg_C min \sum_{i=1}^{k} \sum_{r \in C_i} \|r - \mu_i\|^2 \quad (1)$$

Where C is the clusters whose points are the reviews represented by vectors where each of its elements is the attributes (rating, neutral, positive, negative, compound). The size of the cluster C_i is $|C_i|$, L^2 norm is represented by $\|.\|$, and μ_i is the centroid of those points in C_i such that

$$\mu_i = \frac{1}{|C_i|} \sum_{r \in C_i} r \quad (2)$$

We then find the optimized K clusters and set the labels to each review to its grouped cluster. The top positive cluster and the most negative cluster are identified as the reviews that contain memorable experiences. Finally, The labeled reviews will be used to find the most representative keywords in the next step.

B. Trend Analysis of Memorable Experiences Using Keyword Extraction

Up to this point, we have all the reviews from HotelRec accompanied by their cluster label based on the previously discussed technique. We implement a state-of-the-art KE algorithm on the top positive and most negative cluster reviews to extract the most representative keywords from each cluster. Therefore, MDERank [71] is considered a good candidate for the task because it outperforms all previous KE algorithms in terms of F1 score and has available code for implementation. However, MDERank performance is slower than SIFRank and only outperforms it by an average of 1.8 F1. Consequently, we implement SIFRank to extract the keywords.

In SIFRank [59], a hotel review from HotelRec undergoes tokenization and part-of-speech tagging. Subsequently, noun phrases (NPs) are extracted using a pattern-based NP-chunker, utilizing the part-of-speech tags, and these NPs are considered candidate keywords. The document's tokens are then fed into a pre-trained language model to acquire token representations, which may include multi-layer word embeddings. Using a sentence embedding model, the NPs and the entire document are transformed into NP embeddings and document embeddings, respectively, ensuring they share the same number of layers and dimensions. The similarity between candidate keyphrases and the document's topic is assessed using the cosine distance

between the NP embeddings and document embeddings. Finally, the top-N most similar candidate keywords are chosen as the final keywords for the hotel review.

The first step to implementing SIFRank into G2F involves preprocessing the reviews in the HotelRec dataset. This text preprocessing includes tokenization, stop word removal and special character removal. Then, given a preprocessed review from $HTA_N^{text} = \{r_1, r_1, \cdots, r_n\}$, $r \in HTA$ where HTA_N^{attr} denotes HotelRec dataset, and a set of selected candidate keywords $W = \{w_1, \cdots, w_i, \cdots, w_m\}$, where a candidate w_i consists of one or multiple tokens, as $w_i = \{w_i^1, \cdots, w_i^l\}$, and $m \leq n$, SIFRank's task is to select C candidates from W, where $(C \leq m)$. The candidates in C are scored and ranked from most important to least. The values range from 0 to 1, with higher values indicating greater relevance of the candidate keyword to the review's topic. Conversely, lower values indicate the keyword's increasing irrelevance to the topic.

After extracting the keywords, they play a pivotal role in analyzing the trends of MEs over the years and months. The analysis begins by focusing on the top positive and most negative clusters. These clusters are instrumental in understanding the essence of MEs. The reviews within each cluster are then organized based on the year and further categorized into monthly sub-groups. Within these groups and sub-groups, we calculate the frequency distribution of words, shedding light on their significance in both the yearly and monthly contexts. This meticulous process allows us to draw meaningful insights and patterns from the data, enabling a comprehensive understanding of the evolving trends in MEs.

C. Creating Novel Memorable Experiences with Local LLM Text Generation

Given the representative MEs extracted keywords from all the reviews that were grouped by year and months according to their cluster of either the highest positive reviews or the most negative reviews, we utilize those keywords to create novel and unexplored memorable customer experiences for future hotel stays. To generate the texts for this step, we implement a local text generation LLM Vicuna [13]. Vicuna-13B [13] represents an open-source chatbot trained using the sophisticated fine-tuning techniques of LLaMA. It utilizes dialogues from ShareGPT, a platform where users share their conversations, as its foundational training data. We did not use any ChatGPT [8] to avoid any privacy issues. According to ChatGPT's privacy policy[1], one of the resources that ChatGPT gathers its information is the conversation or prompts that are typed into the chatbot itself. Furthermore, An initial assessment deploying GPT-4 as a benchmark indicates Vicuna-13B surpasses 90% of the performance quality exhibited by established models like OpenAI's ChatGPT and Google Bard. Furthermore, it excels beyond other models such as LLaMA and Stanford Alpaca in over 90% of instances [13].

To generate a future and unexplored customer ME for a hotel stay, a prompt is customized and then inputted into Vicuna as illustrated in Algorithm. 2. The prompts to generate a positive ME differ from the negative ones. Each prompt is generated by $Prompt(CustomizedText + keywords)$ function (see line 13 Algorithm. 1). An example for the unchanged part of the prompt ($CustomizedText$) for a positive one is ***"Write a positive memorable experience hotel review from the following keywords:"***. We aimed to make the unchanged part of the prompt as short as possible to simplify it for Vicuna. Then, we concatenate the top 20 extracted keywords and 500 random ones from the same group to $CustomizedText$. The top 20 extracted keywords guarantee the review will be about positive fundamental hotel concepts. The random 500 keywords make the text generation unique because they will differ each time. Here is an explained instance from the pseudo-code structure in Algorithm. 2: to create a positive prompt for July 2018, we concatenate the positive $CustomizedText$, the top 20 keywords denoted as $TopHTA_c^d$ where d is the date and c is the cluster, and another 500 random extracted keywords denoted as $RandHTA_c^d$ from the grouped reviews of the exact date and cluster. We then input the concatenated prompt into Vicuna to get the final output.

[1]https://openai.com/policies/privacy-policy

Algorithm 2 Pesudo code for used prompt in Vicuna

Input: Top-20-KW $TopHTA_c^d$, Rand-500-KW $RandHTA_c^d$
Output: New-Hotels-Memorable-Experiences-Review NR_c^d

1: $LocalVicuna \leftarrow Prompt(CustomizedText + keywords)$

```
PROMPT_OBJECT{
 "PosPromptText": (
  "CustomizedText": "Write a positive ..",
  "Top-20-Keywords": "w1, w2, ..., w20",
  "Rand-500-Keywords": "w1, w2, ..., w500"
  ),
 "NegPromptText": (
  "CustomizedText": "Write a negative ..",
  "Top-20-Keywords": "w1, w2, ..., w20",
  "Rand-500-Keywords": "w1, w2, ..., w500"
  )
}
Output{
  "Pos_Rev" : GenText(PosPromptText),
  "Neg_Rev" : GenText(NegPromptText)
}
```

Return $NR_c^d \leftarrow Output$

IV. Experiment and Results

In this section, we first introduce the dataset and then perform machine-based and human-based evaluation metrics on keywords extractions for memorable hotel experiences and text generation. We then demonstrate the experimental results in a series of evaluations. In addition, a case study is provided to showcase an actual demonstration of the objectives of G2F framework.

A. Dataset

We conduct our study on a dataset collected only for hotels called HotelRec [2]. HotelRec is a comprehensive repository

of hotel reviews collected from TripAdvisor[2]. According to the third quarterly report in November 2019, available on the U.S. Securities and Exchange Commission website, TripAdvisor is established as the world's preeminent online travel platform, featuring roughly 1.4 million hotels [3]. The dataset contains a period of nineteen years, from February 1, 2001, to May 14, 2019, and stores 50,264,531 worldwide hotel reviews. These reviews are provided by a user base totaling 21,891,294 individuals. HoteRec analysis of user contribution reveals a diverse distribution, with 67.55% of users writing a single review and 90.73% contributing less than five. The average review count per user is 2.24, with the median at one review. User evaluations, represented through obligatory overall ratings, signify the collective hotel experience. Each review in the dataset includes a user profile, the user profile, the hotel URL, the overall rating, the summary, the user-written text, the date, and detailed sub-ratings on different hotel aspects when applicable. As it stands, HotelRec is unrivaled in size as a public dataset in the hotel industry and is the largest text-based recommendation dataset in any single domain.

B. Experiment Settings and Evaluation Metrics

The K-Means clustering algorithm is simple and efficient. However, a major challenge in this technique is determining the K number of clusters that should be chosen to group the data. To determine K we first use a statistical technique called the elbow method. The method calculates the sum of squared distances from each data point to its assigned center point, or centroid, during each iteration of the K-Means algorithm. Each iteration is carried out with a differing number of clusters. The result is displayed in the lower right chart in Figure. 2. The chart shows that $K = 4$ is the optimal number of clusters and also the most efficient one. In addition, $K = 5$ is also seen as the optimal number of clusters, but it is not timewise efficient when $k = 5$. However, some could argue that the elbow method is highly ambiguous because it does not contain a definite elbow [30] and is also considered unreliable in some cases. Therefore, we use the Silhouette method to find the optimal K number of clusters. The silhouette coefficients for each point signify how well a point aligns with other data in its cluster and how poorly it aligns with data from the nearest cluster, specifically, the cluster whose average distance from the data point is the smallest [47]. The value of the silhouette ranges between [-1, 1], and the closer the value to 1, the better K clusters we have. The four charts in Figure. 2 show the average silhouette scores when $K = \{2, 3, 4, 5\}$. We see $k = 4$ and $k = 5$ are the closest to 1, with scores of 0.83 and 0.85, respectively. Therefore, looking at the elbow method and silhouette scores together, 5 is the optimal K.

The dataset we are using is not highly dimensional because it has five features, but it may challenge the algorithm to group the clusters optimally. In Figure. 2, we plot the t-SNE to visualize the clusters and corroborate that the optimal K is 5. The t-SNE chart in the lower left of Figure. 2 shows how well a sample of 7 million points is separated, and there is a slight overlap. The top positive cluster is well separated with no overlap in the purple color. The mosst negative cluster in yellow overlaps with the second most negative cluster. Although the separation among the most negative clusters is not well-established, it is evident from the chart that our proposed method successfully extracts the top positive and most negative clusters so we can utilize the clusters for KE of positive and negative MEs.

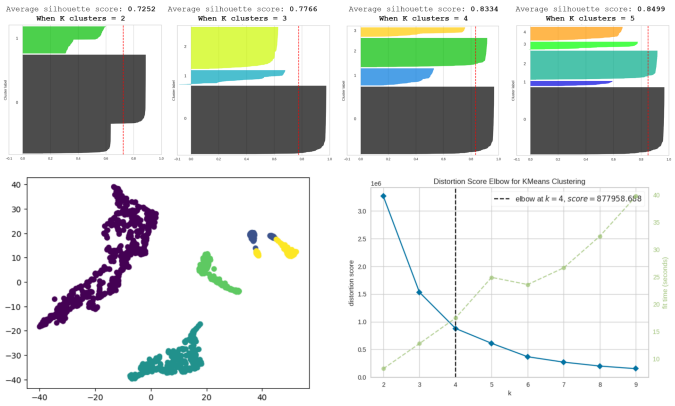

Fig. 2: Choosing optimal K for K-Means. The top four figures show the Silhouette scores for 2,3,4, and 5 clusters. The bottom left is the t-SNE figure. The right lower figures show the Elbow Test for K

Our framework contains two key aspects for evaluation in the hotel industry: **a) memorable experience-driven keyword extraction**, and **2) Creating Novel Memorable Experiences with Local LLM Text Generation.**

A. Machine-based evaluation of memorable experience-driven keyword extraction: We evaluate the utilization of SIFRank to extract memorable experience-related keywords based on Precision, Recall, and F1 value. All the SOTA keyword extraction methods are evaluated on several datasets of different domains, such as Inspec [26], SemEval2017 [4], or DUC2001 [67]. Needless to say, it is necessary to evaluate the accuracy of memorable experiences-related keyword extraction from HotelRec Dataset. To accomplish this, an industry expert meticulously annotated 100 reviews from HotelRec containing positive or negative memorable experiences. Extracting representative keywords from each review, we conduct the evaluation by comparing the annotated keywords to those extracted by SIFRank.

B.1. Machine-based evaluation of creating novel memorable experiences with Local LLM text generation. We evaluate the generated texts (positive/negative ME reviews) by a local Vicuna in two measures: ROUGE (ROUGE-1, ROUGE-2, ROUGE-L) score and BLEU (BLEU-1, BLEU-2, BLEU-3, BLEU-4). At first, we generated 30 prompts with the same standards in Algorithm. 2. The same prompts are given to an industry expert to generate memorable experiences based on those prompts. Similarly, those prompts were inputted into

[2]https://www.tripadvisor.com/
[3]https://www.sec.gov/ix?doc=/Archives/

TABLE I: Machine-based evaluation results of SIFRank performance on HotelRec compared to extracted keywords by a hotel industry expert

KE Method	Percision	Recall	F1-Score
SIFRank	0.86	0.54	0.63

Vicuna. The similarity between the two output texts of each prompt (human vs. machine) is calculated by ROUGE and BLEU based on the overlap of unigrams, bigrams, and the longest common sequences.

B.2.Human-based evaluation of creating novel MEs with Local LLM text generation. Relying only on ROUGE and BLEU scores is not reliable enough and favors the scoring against generated texts that delivers the same content but rephrases used words. Therefore, we perform a human evaluation based on Likert scale scoring [43]. The prevalent technique involves assigning ratings to a generated text (the review generated by Vicuna) based on a source document (the review generated by the expert). This often takes the form of an independent assessment, where each generated text is evaluated independently rather than compared directly with others. The evaluative criteria generally include consistency, fluency, informativeness, and relevance. Each generated text is scored on a scale from 1, being the poorest, to 5, considered the best. Two anonymous annotators were given the same task to conduct the evaluation.

C. Results and Analysis

The machine-based evaluation results of the SIFRank method to extract memorable experience-related keywords reveal an interesting trade-off between precision and recall. As seen in Table I, the precision of 0.86 indicates that when the method identifies keywords, it is highly accurate, with only 14% of the extracted keywords being false positives. This suggests that the method is proficient at selecting relevant and appropriate keywords, making it a valuable identifying representative keywords of ME. However, the recall score of 0.54 indicates that the method misses 46% of the actual keywords present in the text. One reason for the low recall score is the different lengths of the extracted keywords between the human-annotated keywords and the extracted ones by SIFrank. For example, SIFRank extracts from one to three words as one keyphrase, while the keyphrase length extracted by the annotator can consist of 4 words. Nonetheless, the overall F1 score of 0.63 demonstrates a reasonably balanced performance, indicating that the method strikes a fair compromise between precision and recall. Better annotation of human-generated keywords from HotelRec could potentially improve recall without sacrificing precision, leading to a more effective keyword extraction approach.

The evaluation results in Table II of the LLM (Large Language Model) for generating text reviews indicate its effectiveness in capturing the essence of the original reviews. The mean ROUGE scores, which measure the similarity between the generated text and the reference (original) text,

TABLE II: Machine-based evaluation results comparing the generated text by Vicuna to generated text by a hotel expert

Metric Name	Accuracy
ROUGE-1	0.5559
ROUGE-2	0.3374
ROUGE-L	0.5201
BLEU-1	0.5741
BLEU-2	0.4226
BLEU-3	0.3230
BLEU-4	0.2471

TABLE III: Human-based evaluation results comparing the generated text by Vicuna to generated text by a hotel expert. C = Consistency, F = fluency, I = informativeness, R = relevance

Method	**C**	**F**	**I**	**R**
Annotator 1	4.625	4.5	4.757	4.5
Annotator 2	4.263	4.25	4.375	4.125

are quite promising. For ROUGE-1, the score of 0.5559 suggests that more than half of the unigrams (individual words) in the generated reviews match those in the reference reviews. Similarly, for ROUGE-2, with a score of 0.3374, the model demonstrates reasonable success in reproducing meaningful word sequences of two words in length from the reference text. Moreover, the ROUGE-l score of 0.5201 reveals that the LLM performs well in preserving the reviews' overall linguistic structure and continuity. The ROUGE-L metric considers the longest common subsequence between the generated and reference texts, indicating that the LLM can produce reviews that capture the essence and context of the original reviews reasonably well. For BLEU-1, the score of 0.5741 indicates that over 57% of the unigrams (individual words) in the generated reviews match those in the reference reviews. This suggests that the LLM is reasonably successful in producing words that align with the original reviews. For BLEU-2, the score of 0.4226 represents the similarity in bigrams (sequences of two words) between the generated and reference texts. The model's ability to reproduce meaningful two-word sequences is evident, though there is still room for improvement. BLEU-3 and BLEU-4 scores, 0.3230 and 0.2471, respectively, account for trigrams (sequences of three words) and 4-grams (sequences of four words). These scores demonstrate that the LLM's performance in generating longer sequences of words is relatively lower compared to unigrams and bigrams. To summarize, Vicuna showed its effectiveness in generating text reviews. It did not only generate reviews from the exact keywords but also showed its ability to write novel and unexplored reviews that describe positive and negative MEs.

Table. III shows the results of the human-based evaluation of the generated customer ME reviews according to the Likert scale. The generated reviews received high scores across all categories from the first and second annotators. It received average scores of 4.625 and 4.2625 for consistency,

indicating that our method of creating the prompt assessed Vicuna to produce text with high coherence and did not contradict itself. The fluency scores of 4.5 and 4.25 show that the generated texts were very high in terms of grammar, sentence structure, and readability. The flow of the writing was very smooth and appeared human-like text. Regarding informativeness, impressive scores of 4.757 and 4.375 indicate that the generated reviews were highly insightful and valuable. Lastly, the relevance scores of 4.5 and 4.125 suggest that the framework's outputs were highly pertinent to the given prompt or task. According to the annotators' evaluation, these scores show that employing Vicuna to receive the created prompt from our proposed framework demonstrated a high level of proficiency in generating positive and negative ME reviews.

D. Case Study

In this section, we perform case studies to demonstrate the capabilities of G2F in the hotel industry. The first case study demonstrates how G2F can identify trending keywords related to MEs for hotel stays, providing valuable insights for hotel owners and managers looking to improve the quality of their service. The second case study showcases the ability of G2F to combine with local LLMs to generate new and unique experiences that can help the hotel industry stay ahead of the curve in terms of customer satisfaction and service innovation.

To showcase the usefulness of our proposed G2F, we present a qualitative analysis of a real-world case study from July 2018. According to the Gensler Hospitality Index report, several fundamental factors play a crucial role in creating a good hotel experience: cleanliness, safety, quality/value, and having friendly and hospitable staff are statistically significant drivers [21]. Figure 3 shows that the framework captured the fundamental representative keywords for making an experience good at hotel stays. For example, in the cleanliness domain, some of the extracted keywords were "spotlessly clean, "clean bed," and "clean room." In the staff domain, example keywords such as "friendly," "helpful," and "welcoming" were extracted and highly mentioned. All the primary domain keywords have been highly mentioned not only for one month but for every month. This shows that G2F successfully identified the fundamental factors.

Moreover, G2F captured representative keywords from the defining terms of a ME mentioned by Kim [32] such as local culture and novelty. G2F identified local events associated with hotel stays, such as the Soccer World Cup in Russia in the summer of 2018. The unique keywords extracted from customer reviews were "Russian channels," "world cup themed," "repainted walls," and "Russian tour." Hotel management can benefit from such analysis by preparing for such events by providing such services in their hotels, like providing TV channels of the event or creating a themed atmosphere for such an event by repainting the walls. The last capability of G2F is creating novel and unexplored experiences. Figure 3 shows an example review text created by G2F that covers the fundamental aspects of a ME. Creating reviews like that can help hotel management polarize ideal future MEs and raise their level of service by preparing for unexplored scenarios.

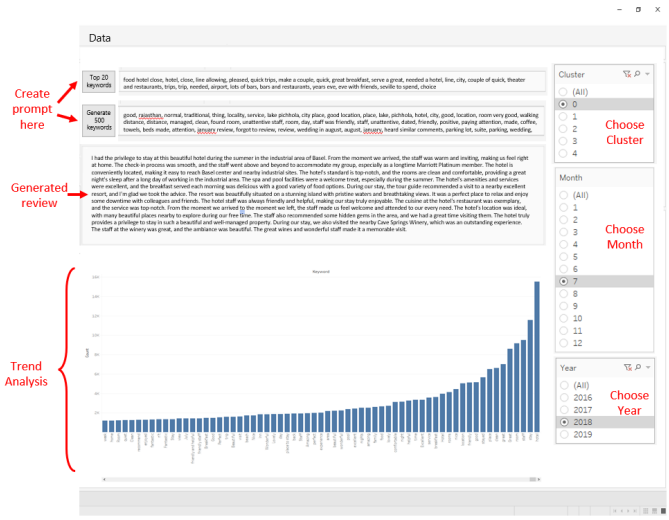

Fig. 3: Example of the case study from G2F for July 2018

V. Conclusion

In conclusion, this paper underlines the growing significance of MEs in the hotel industry and their impact. G2F is a solution for hotels to efficiently capture, analyze, and generate MEs for the hotel industry and make guests feel like family. The platform's evaluation results demonstrate its effectiveness in identifying relevant keywords associated with MEs and generating novel, consistent, and informative customer reviews. G2F sets a new standard in leveraging technology to enhance hotel services, ultimately leading to improved customer satisfaction and a competitive advantage in the market.

References

[1] Julio Cesar Salinas Alvarado, Karin Verspoor, and Timothy Baldwin. Domain adaption of named entity recognition to support credit risk assessment. In Ben Hachey and Kellie Webster, editors, *Proceedings of the Australasian Language Technology Association Workshop, ALTA 2015, Parramatta, Australia, December 8 - 9, 2015*, pages 84–90. ACL, 2015.

[2] Diego Antognini and Boi Faltings. Hotelrec: a novel very large-scale hotel recommendation dataset. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 4917–4923, Marseille, France, May 2020. European Language Resources Association.

[3] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *CoRR*, abs/1908.10063, 2019.

[4] Isabelle Augenstein, Mrinal Das, Sebastian Riedel, Lakshmi Vikraman, and Andrew McCallum. Semeval 2017 task 10: Scienceie - extracting keyphrases and relations from scientific publications. In Steven Bethard, Marine Carpuat, Marianna Apidianaki, Saif M. Mohammad, Daniel M. Cer, and David Jurgens, editors, *Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada, August 3-4, 2017*, pages 546–555. Association for Computational Linguistics, 2017.

[5] Slobodan Beliga. Keyword extraction: a review of methods and approaches. *University of Rijeka, Department of Informatics, Rijeka*, 1(9), 2014.

[6] Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl, and Martin Jaggi. Simple unsupervised keyphrase extraction using sentence embeddings. In Anna Korhonen and Ivan Titov, editors, *Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018, Brussels, Belgium, October 31 - November 1, 2018*, pages 221–229. Association for Computational Linguistics, 2018.

[7] Andrew Blair-Stanek, Nils Holzenberger, and Benjamin Van Durme. Can GPT-3 perform statutory reasoning? *CoRR*, abs/2302.06100, 2023.

[8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

[9] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Mário Jorge, Célia Nunes, and Adam Jatowt. Yake! collection-independent automatic keyword extractor. In *Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings 40*, pages 806–810. Springer, 2018.

[10] Yung-Chun Chang, Chih-Hao Ku, and Chun-Hung Chen. Social media analytics: Extracting and visualizing hilton hotel ratings and reviews from tripadvisor. *International Journal of Information Management*, 48:263–279, 2019.

[11] Tzeng-Ji Chen. Chatgpt and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association*, 86(4):351–353, 2023.

[12] Xin Chen, Zhen-feng Cheng, and Gyu-Bae Kim. Make it memorable: Tourism experience, fun, recommendation and revisit intentions of chinese outbound tourists. *Sustainability*, 12(5):1904, 2020.

[13] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality, March 2023.

[14] M Csikszentmihalyi. Flow: The psycology of optimal experience, steps toward enhancing the quality of life. 1990.

[15] Haoran Ding and Xiao Luo. Attentionrank: Unsupervised keyphrase extraction using self and cross attentions. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 1919–1928. Association for Computational Linguistics, 2021.

[16] Abdallah M Elshaer and Asmaa M Marzouk. Memorable tourist experiences: the role of smart tourism technologies and hotel innovations. *Tourism Recreation Research*, pages 1–13, 2022.

[17] Jordán Pascual Espada, Jaime Solís Martínez, Irene Cid Rico, and Luis Emilio Velasco Sánchez. Extracting keywords of educational texts using a novel mechanism based on linguistic approaches and evolutive graphs. *Expert Systems with Applications*, 213:118842, 2023.

[18] Nino Fijačko, Lucija Gosak, Gregor Štiglic, Christopher T Picard, and Matthew John Douma. Can chatgpt pass the life support exams without entering the american heart association course? *Resuscitation*, 185, 2023.

[19] Simon Frieder, Luca Pinchetti, Ryan-Rhys Griffiths, Tommaso Salvatori, Thomas Lukasiewicz, Philipp Christian Petersen, Alexis Chevalier, and Julius Berner. Mathematical capabilities of chatgpt. *CoRR*, abs/2301.13867, 2023.

[20] Sebastian Gehrmann, Yuntian Deng, and Alexander Rush. Bottom-up abstractive summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4098–4109, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[21] Institute Gensler. Hospitality experience index. Technical report, Gensler research institute, 2018.

[22] Gabriela Nathania H., Ryan Siautama, Amadea Claire I. A., and Derwin Suhartono. Extractive hotel review summarization based on tf/idf and adjective-noun pairing by considering annual sentiment trends. *Procedia Computer Science*, 179:558–565, 2021. 5th International Conference on Computer Science and Computational Intelligence 2020.

[23] Sameer Hosany, Erose Sthapit, and Peter Björk. Memorable tourism experience: A review and research agenda. *Psychology & Marketing*, 39(8):1467–1486, 2022.

[24] Seyedasaad Hosseini, Rafael Cortes Macias, and Fernando Almeida Garcia. Memorable tourism experience research: a systematic review of the literature. *Tourism Recreation Research*, 48(3):465–479, 2023.

[25] Ya-Han Hu, Yen-Liang Chen, and Hui-Ling Chou. Opinion mining from online hotel reviews – a text summarization approach. *Information Processing Management*, 53(2):436–449, 2017.

[26] Anette Hulth. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2003, Sapporo, Japan, July 11-12, 2003*, 2003.

[27] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

[28] Miyoung Jeong and Hyejo Hailey Shin. Tourists' experiences with smart tourism technology at smart destinations and their behavior intentions. *Journal of Travel Research*, 59(8):1464–1477, 2020.

[29] Baoyu Jing, Zeyu You, Tao Yang, Wei Fan, and Hanghang Tong. Multiplex graph neural network for extractive text summarization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 133–139, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

[30] David J Ketchen and Christopher L Shook. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, 17(6):441–458, 1996.

[31] Hyangmi Kim and Joseph S Chen. Memorable travel experiences: recollection vs belief. *Tourism Recreation Research*, 46(1):124–131, 2021.

[32] Jong-Hyeong Kim, JR Brent Ritchie, and Bryan McCormick. Development of a scale to measure memorable tourism experiences. *Journal of Travel research*, 51(1):12–25, 2012.

[33] Anton Korinek. Language models and cognitive automation for economic research. Technical report, National Bureau of Economic Research, 2023.

[34] Tiffany H Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models. *PLoS digital health*, 2(2):e0000198, 2023.

[35] Béchir Ben Lahouel, Nathalie Montargot, et al. Children as customers in luxury hotels: what are parisian hotel managers doing to create a memorable experience for children? *International Journal of Contemporary Hospitality Management*, 32(5):1813–1835, 2020.

[36] Hien Ngo Le Huy, Hoang Ho Minh, Tien Nguyen Van, and Hieu Nguyen Van. Keyphrase extraction model: a new design and application on tourism information. *Informatica*, 45(4), 2021.

[37] Qiusheng Li, Yadong Wu, Song Wang, Maosong Lin, Xinmiao Feng, and Haiyang Wang. Vistravel: visualizing tourism network opinion from the user generated content. *Journal of Visualization*, 19:489–502, 2016.

[38] Brady D Lund and Ting Wang. Chatting about chatgpt: how may ai and gpt impact academia and libraries? *Library Hi Tech News*, 40(3):26–29, 2023.

[39] Asmaa Marzouk, Azza Maher, and Toka Mahrous. The influence of augmented reality and virtual reality combinations on tourist experience. *Journal of the Faculty of Tourism and Hotels-University of Sadat City*, 3(2):1–19, 2019.

[40] Asmaa M Marzouk. Egypt's image as a tourist destination: an exploratory analysis of dmo's social media platforms. *Leisure/loisir*, 46(2):255–291, 2022.

[41] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411, 2004.

[42] John J. Nay. Law informs code: A legal informatics approach to aligning artificial intelligence with humans. *CoRR*, abs/2209.13020, 2022.

[43] Ani Nenkova and Rebecca J. Passonneau. Evaluating content selection in summarization: The pyramid method. In Julia Hirschberg, Susan T. Dumais, Daniel Marcu, and Salim Roukos, editors, *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, HLT-NAACL 2004, Boston,*

Massachusetts, USA, May 2-7, 2004, pages 145–152. The Association for Computational Linguistics, 2004.

[44] Jonathan Pilault, Raymond Li, Sandeep Subramanian, and Chris Pal. On extractive and abstractive neural document summarization with transformer language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9308–9319, Online, November 2020. Association for Computational Linguistics.

[45] B Joseph Pine, James H Gilmore, et al. *Welcome to the experience economy*, volume 76. Harvard Business Review Press Cambridge, MA, USA, 1998.

[46] Arya Rao, John Kim, Meghana Kamineni, Michael Pang, Winston Lie, and Marc D Succi. Evaluating chatgpt as an adjunct for radiologic decision-making. *medRxiv*, pages 2023–02, 2023.

[47] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[48] Qian Ruan, Malte Ostendorff, and Georg Rehm. HiStruct+: Improving extractive text summarization with hierarchical structure information. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1292–1308, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[49] Jürgen Rudolph, Samson Tan, and Shannon Tan. Chatgpt: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 2023.

[50] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[51] Siamak Seyfi, C Michael Hall, and S Mostafa Rasoolimanesh. Exploring memorable cultural tourism experiences. *Journal of Heritage Tourism*, 15(3):341–357, 2020.

[52] Abheesht Sharma, Gunjan Chhablani, Harshit Pandey, and Rajaswa Patil. DRIFT: A toolkit for diachronic analysis of scientific literature. In Heike Adel and Shuming Shi, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2021, Online and Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 361–371. Association for Computational Linguistics, 2021.

[53] Giriprasad Sridhara, Sourav Mazumdar, et al. Chatgpt: A study on its utility for ubiquitous software engineering tasks. *arXiv preprint arXiv:2305.16837*, 2023.

[54] Erose Sthapit. Exploring tourists' memorable food experiences: A study of visitors to santa's official hometown. *Anatolia*, 28(3):404–421, 2017.

[55] Erose Sthapit, Peter Björk, and Dafnis N Coudounaris. Emotions elicited by local food consumption, memories, place attachment and behavioural intentions. *Anatolia*, 28(3):363–380, 2017.

[56] Erose Sthapit, Peter Björk, and Jano Jiménez Barreto. Negative memorable experience: North american and british airbnb guests' perspectives. *Tourism Review*, 76(3):639–653, 2021.

[57] Chris Stokel-Walker. Chatgpt listed as author on research papers: many scientists disapprove. *Nature*, 613(7945):620–621, 2023.

[58] Weisong Sun, Chunrong Fang, Yudu You, Yun Miao, Yi Liu, Yuekang Li, Gelei Deng, Shenghan Huang, Yuchen Chen, Quanjun Zhang, Hanwei Qian, Yang Liu, and Zhenyu Chen. Automatic code summarization via chatgpt: How far are we? *CoRR*, abs/2305.12865, 2023.

[59] Yi Sun, Hangping Qiu, Yu Zheng, Zhongwei Wang, and Chaoran Zhang. Sifrank: A new baseline for unsupervised keyphrase extraction based on pre-trained language model. *IEEE Access*, 8:10896–10906, 2020.

[60] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models. *CoRR*, abs/2302.13971, 2023.

[61] Dietrich Trautmann, Alina Petrova, and Frank Schilder. Legal prompt engineering for multilingual legal judgement prediction. *CoRR*, abs/2212.02199, 2022.

[62] Chih-Fong Tsai, Kuanchin Chen, Ya-Han Hu, and Wei-Kai Chen. Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management*, 80:104122, 2020.

[63] Vincent Wing Sun Tung and JR Brent Ritchie. Exploring the essence of memorable tourism experiences. *Annals of tourism research*, 38(4):1367–1386, 2011.

[64] Sera Vada, Catherine Prentice, Noel Scott, and Aaron Hsiao. Positive psychology and tourist well-being: A systematic literature review. *Tourism Management Perspectives*, 33:100631, 2020.

[65] Sakshi Vatsa, Surbhi Mathur, Mansi Garg, and Rajni Jindal. Covid-19 tweet analysis using hybrid keyword extraction approach. In *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, pages 136–140. IEEE, 2021.

[66] Didier A Vega-Oliveros, Pedro Spoljaric Gomes, Evangelos E Milios, and Lilian Berton. A multi-centrality index for graph-based keyword extraction. *Information Processing & Management*, 56(6):102063, 2019.

[67] Xiaojun Wan and Jianguo Xiao. Single document keyphrase extraction using neighborhood knowledge. In Dieter Fox and Carla P. Gomes, editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008*, pages 855–860. AAAI Press, 2008.

[68] Jing Wang and Jianjun Zhu. Research on hotel customer preferences and satisfaction based on text mining: Taking ctrip hotel reviews as an example. In *INFORMS International Conference on Service Science*, pages 227–237. Springer, 2021.

[69] Hongyang Yang, Xiao-Yang Liu, and Chris Wang. Fingpt: Open-source financial large language models. *ArXiv*, abs/2306.06031, 2023.

[70] Xianjun Yang, Yan Li, Xinlu Zhang, Haifeng Chen, and Wei Cheng. Exploring the limits of chatgpt for query or aspect-based text summarization. *CoRR*, abs/2302.08081, 2023.

[71] Linhan Zhang, Qian Chen, Wen Wang, Chong Deng, ShiLiang Zhang, Bing Li, Wei Wang, and Xin Cao. MDERank: A masked document embedding rank approach for unsupervised keyphrase extraction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 396–409, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[72] Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. Extractive summarization as text matching. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6197–6208, Online, July 2020. Association for Computational Linguistics.