Li Z., Chan F.Y., Gao C. & Ye Q. (2024). Will sentiment extraction based on ChatGPT yield better predictive outcomes? Evidence from an online travel agency. In Li, E.Y. *et al.* (Eds.) *Proceedings of The International Conference on Electronic Business, Volume 24* (pp. 196-208). ICEB'24, Zhuhai, China, October 24-28, 2024



# Will Sentiment Extraction based on ChatGPT yield Better Predictive Outcomes? Evidence from an Online Travel Agency

Zhihao Li [1, *]
Fun Yi Chan [2, *]
Chaoyue Gao [3]
Qiang Ye [4]

*Corresponding author
[1] PhD candidate, Harbin Institute of Technology, Harbin, China, leezh@hit.edu.cn
[2] PhD candidate, Harbin Institute of Technology, Harbin, China, funyichan@stu.hit.edu.cn
[3] Associate Professor, University of Science and Technology of China, Hefei, China, gaochaoyue@ustc.edu.cn
[4] Professor, University of Science and Technology of China, Hefei, Country, yeqiang@ustc.edu.cn

**ABSTRACT**

The rise of Large Language Models (LLMs) has significantly advanced natural language processing, outperforming many traditional tools such as rule-based systems and machine learning models. ChatGPT, a leading example, has exhibited exceptional capabilities in textual analysis. This study examines whether ChatGPT can outperform traditional sentiment analysis methods in the context of sales prediction leveraging online review data from online travel agencies, Booking and Expedia. We employ review ratings, and sentiment analysis tools, including VADER, RoBERTa, and ChatGPT, to predict revenue metrics. We find that both VADER and RoBERTa exhibit comparable predictive power to review ratings, whereas ChatGPT's sentiment scores demonstrate a weaker correlation with revenue metrics. Grounded in Heuristic-Systematic Models (HSM) from dual process theory, we posit that customers rely predominantly on heuristic cues (review ratings and keywords of extreme words) for decision making, which are better captured by traditional sentiment analysis tools. In contrast, ChatGPT's evaluation, which emphasizes systematic review content processing, aligns less with consumer behavior in this context. This study contributes theoretically to extending HSM to illustrate how AIGC moderates systematic information processing in sales prediction. It also offers empirical insights into the comparative effectiveness of sentiment analysis tools, providing a practical implication for e-commerce platforms and managers regarding the adoption of AIGC in strategic decision-making. Caution is advised when integrating AIGC into sales and operational strategies.

*Keywords*: ChatGPT, LLM, sentiment analysis, rating bias, hospitality industry.

## INTRODUCTION

Online reviews serve as an important source of information for potential buyers to reduce uncertainty, enhance product fit, and influence purchasing decisions (Pavlou & Dimoka, 2006). The relationship between online reviews and product sales performance has attracted widespread attention in recent years, highlighting the crucial role of customer reviews in sales forecasting (Chevalier & Mayzlin, 2006). this dynamic is particularly evident in the hotel industry, where reviews substantially influence hotel reservations (Vermeulen & Seegers, 2009). Positive reviews can enhance a hotel's reputation and attractiveness, thereby increasing reservation and occupancy rates, while negative reviews diminish evaluations and reduce reservation likelihood (Sparks & Browning, 2011). In contemporary business practice, traditional sentiment analysis methods, predominantly rely on lexicon-based techniques or machine learning models, are widely utilized to analyze various review dimensions, including ratings, review quality and review sentiment, to estimate sales revenue and inform operation strategies (Mehta & Pandya, 2020). However, the emergence of advanced large language models (LLMs), such as ChatGPT, introduces a novel approach to sentiment extraction (Lopez-Lira & Tang, 2023), potentially reshape the landscape of review analysis and sales prediction.

Given limited processing capacity, customers often reduce the effort required to evaluate different reviews attributes when making purchase decisions (Hu et al., 2014). They tend to rely on heuristic information processing, focusing on easily accessible cues, such as numerical ratings and emotionally charged keywords, rather than systematically evaluating the content of each review. Systematic information processing, which involves deeply analyzing every review in detail, is often impractical due to the sheer volume of available reviews. Traditional sentiment analysis tools typically capture keywords reflecting extreme emotions, aligning with the heuristic process strategies customers usually employ when browsing reviews. These tools, therefore, may more effectively predict purchase behaviors. While ChatGPT shows superior natural language processing capabilities, and is better suited to individuals' systematic information processing, which probably extract more nuanced and comprehensive sentiment cues but may not effectively capture the heuristic processes customers use to make decision. Consequently, its predictive power in sales performance may not surpass that of numerical ratings or traditional sentiment analysis methods on e-commerce platforms. Therefore, this study seeks to explore whether ChatGPT can outperform

traditional methodologies and numerical rating in predicting sales by extracting sentiment from hotel reviews, potentially reshaping how businesses leverage reviews for decision-making.

To address this research question, we conducted an empirical analysis using numerical review ratings alongside three sentiment analysis tools: VADER, RoBERTa, and ChatGPT. We collected hotel review from Texas on two leading online travel agencies, Booking and Expedia. These tools were applied to generate sentiment scores from the review text, which were then used to predict the hotel's revenue indicators for the next month. This approach allowed us to evaluate the effectiveness of both traditional sentiment analysis tools and ChatGPT in predicting financial performance, thereby assessing ChatGPT's potential effectiveness in understand individuals' information processing within the context of e-commerce platforms.

Our finding indicate that the sentiment scores generated by VADER and RoBERTa exhibit similar predictive power to numerical ratings. In contrast, ChatGPT's sentiment score demonstrated weaker correlations with hotel revenue metrics.

Although ChatGPT has shown strong performance in sentiment and textual extraction in prior studies (Bond et al., 2023), we speculate that its sentiment extraction process are less directly aligned with predicting revenue prediction in the context of customer purchase behaviors. Drawing on heuristic-systematic model from dual process theory, we suggest that customers are more inclined to reply on heuristic cues, such as ratings and keywords reflecting extreme sentiment, to reduce cognitive effort rather than engaging in systematic processing of reading review content, which ChatGPT excels at. This preference for heuristic processing may explain why traditional sentiment analysis methods and numerical ratings yield better predictive accuracy for sales performance. This study makes significant theoretical and practical contributions. Theoretically, first, we extend the application of dual process models by considering AI agents, like ChatGPT, as facilitators of systematic information processing. Additionally, we integrate LLMs, particularly ChatGPT, into the research domain of sentiment analysis and sales prediction, which has traditionally relied on keyword-based methods or non-multimodal natural language processing models (Mehta & Pandya, 2020). Practically, the study provides empirical evidence on the comparative effectiveness of different sentiment analysis tools in predicting sales. We advise managers and e-commerce platforms to exercise caution when incorporating AIGC, such as ChatGPT, into their sales and operational strategies, as traditional tools may offer more reliable predictions in certain contexts.

## LITERATURE REVIEW

The burgeoning field of predicting hotel sales through online reviews has garnered significant attention from researchers and practitioners alike. With the advent of big data and sophisticated analytical techniques, the hospitality industry is increasingly capable of utilizing customer feedback to predict demand and improve marketing strategies. Online reviews have become a crucial factor in predicting product sales across various industries. Several studies have explored this relationship, highlighting the importance of review volume, valence, and other characteristics.

### Online Reviews and Product Sales Prediction

The volume and ratings of online reviews are consistently shown to impact sales. Cui et al. (2012) investigated the relationship between user-generated content (such as review quantity) and product sales. Their analysis of data from the Chinese e-commerce market revealed that the number of reviews is a significant driver of sales, especially in highly competitive markets.

Chevalier and Mayzlin (2006) found that both the quantity and quality of online reviews significantly affect book sales on platforms like Amazon, and evidence from review-length data suggests that customers read review text rather than relying only on summary statistics. Dellarocas et al. (2007) emphasized that the volume, valence, and dispersion of online movie reviews all have a positive and statistically significant relationship with future box office sales. Duan et al. (2008) pointed out that user ratings do not affect movie sales after controlling for endogeneity of user reviews and product heterogeneity, while the number of postings is significantly correlated with movie sales after considering of the causality issue. And online reviews play an important role in consumer decision-making in the hospitality industry. Research by Ye et al. (2009) established a direct link between online review ratings and hotel room sales, indicating that positive reviews significantly enhance a hotel's attractiveness. Furthermore, Luca (2016) demonstrated that a one-star increase in Yelp ratings can lead to a 5-9% increase in revenue for independent restaurants, a finding likely extendable to hotels. In addition to review volume and numerical ratings, the textual content of reviews provides rich, qualitative data that reflects customer satisfaction and areas needing improvement.

Studies have shown that detailed, authentic reviews can be more influential than marketer information (Gretzel & Yoo, 2008). For instance, a review elaborating on excellent customer service or clean facilities can significantly boost potential guests' trust and willingness to book a hotel. Conversely, negative reviews detailing specific issues can deter prospective customers, emphasizing the need for hotels to address and manage online feedback proactively. Sentiment analysis, a subset of natural language processing (NLP), plays a pivotal role in extracting sentiments from textual reviews. Early approaches by Hu and Liu (2004) utilized lexicon-based methods to classify sentiments, laying the groundwork for more sophisticated techniques.

Moressss recent studies have employed machine learning models such as Support Vector Machines (SVM) (Medhat et al., 2014), Naive Bayes, and deep learning architectures like Recurrent Neural Networks (RNN) and Long Short-Term Memory networks (LSTM) to enhance sentiment classification accuracy. Studies employing BERT, LLMs have demonstrated their superior performance in sentiment analysis tasks compared to traditional models. For instance, Li et al. (2023) analyze

consumer sentiment using BERT-based technology and make predictions about Restaurant Survival based on this data. With the advent of the ChatGPT era, its unique and exceptional capabilities in sentiment analysis extraction have garnered widespread attention. Moreover, as an increasing number of platforms are adopting GPT for product recommendations, investigating the predictive validity of product sales based on GPT is particularly crucial. Therefore, this study aims to explore the predictive performance of sentiment indicators extracted by GPT in comparison to traditional forecasting metrics for product sales.

**Dual Process Theory**

The dual process theory has been widely applied in the field of Information System to explain individuals' information processing routes and cognition biases. The Heuristic–Systematic Model (HSM), a prevalent dual process model, posits that information processing occurs through two distinct but concurrent routes: heuristic and systematic, which influence each other in complex ways rather than functioning as mutually exclusive processes (Zhang et al., 2014). Heuristic processing is characterized by automatic, intuitive decision-making, requiring minimal cognitive effort and relying on a single or few salient cues, whereas systematic processing involves deliberate, effortful consideration of all available information to form a judgment (Arnott & Gao, 2019; Todorov et al., 2002). The model has been extensively used to explore the underlying mechanism of the influence of quality (systematic processing) and quantity (heuristic processing) of online reviews on customer purchase decisions (Zhang et al., 2014). Heuristic factors such as source trustworthiness and expertise, and systematic factors such as argument quality such as informativeness and persuasiveness have been identified as key elements in this process (Liu et al., 2012; Zhang et al., 2014). In e-commerce platforms, customers are often confronted with vast amounts of information but are constrained by their cognitive limitations. As a result, they tend to adopt simplifying strategies and heuristics to arrive at a decision (Hu et al., 2014; Tversky & Kahneman, 1974). Customers usually relying on various information cues in online reviews, such as numeric rating, volume of reviews, valence, and ratings variance. Information that is standardized and easier to process, such as numerical ratings (heuristic processing), is more likely to be utilized by customers when making purchase decisions, due to its low effort required and alignment across products. However, review contents, which requires more cognitive effort to evaluate, is associated with systematic processing. Therefore, we posit that customers are more likely to rely on heuristic cues over detailed review content when making decisions, leading to higher predictive power of sales from heuristic information processing compared to systematic processing.

**ChatGPT in Text Analysis**

ChatGPT shows excellent performance in classification, summarization and text generation, showcasing its versatility and effectiveness (Dong et al., 2023). Wei et al. (2022) conducted a comparative study focusing on the cognitive abilities of ChatGPT in relation to human auditors. They posed identical questions to both ChatGPT and human auditors, revealing that the responses generated by ChatGPT exhibited a striking resemblance to those of human auditors, particularly in sentiment, diction, and linguistic context. Wang (2023) fine-tuned ChatGPT by training with curated datasets and enhanced its generalization ability. In terms of sentiment analysis. In their study, a customized ChatGPT can help investment funds become more self-informed and more aligned with shareholder interest preferences. Zhang et al. (2023) used the GPT-3.5 model and the traditional BERT model to generate sentiment scores for the headlines and subheadings of Financial Times news, and the results showed that the GPT-3.5 score was more advantageous in predicting stock trends. Hu et al. (2023) compared GPT-3, FinBERT, and word lists from Loughran and McDonald (2011) by using these methods on Chinese MD&A disclosure for sentiment analysis. They found that both GPT-3 and FinBERT outperformed the word list approach. Although previous studies have demonstrated their superior capabilities, the effectiveness of these methods in extracting sentiment from reviews is still questioned.

## METHODOLOGIES

**Data**

Expedia and Booking, two leading online travel agencies, are employed as our research data sources. In the Top Charts of the Travel category in the US iOS store, these two apps are ranked top ten, demonstrating their comparable popularity and influence. In addition, the review policies of both agencies are committed to ensuring that the reviews are authentic and reliable[1][2]. They only allow users who have completed a subscription through their services to write reviews. This ensures that the reviews can reflect the real experience of users.

We used Python crawlers to collect 1,018,707 and 1,460,569 reviews from 2,244 and 2,231 hotels in Texas, from Booking and Expedia respectively. These review data contain tourists' ratings of hotels in addition to text. We also obtained the hotel's monthly total revenue and Revenue Per Available Room (RevPAR) from Search Texas Tax. All data on the website is obtained directly from the Texas Comptroller. After intersecting the hotels and the review time of the two platforms, we had 366,885 and 283,634 reviews from 688 and 668 hotels on Booking and Expedia, respectively, from August 2020 to August 2023. Among them, the non-empty data are 342,677 and 171,077. According to our research question, we formed monthly

[1] https://www.expedia.com/lp/b/content-guidelines

[2] https://www.booking.com/reviews_guidelines.html

review data based on hotels, with 23,175 and 22,141 samples from Booking and Expedia, respectively. We then used a bag-of-words technique and two deep learning models including VADER, RoBERTa, and ChatGPT to analyze the sentiment of the comments.

VADER is a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments expressed in social media contexts. Due to its simplicity and effectiveness, VADER is widely used in research to analyze large volumes of online short data quickly (Elbagir & Yang, 2019). As a rule-based and lexicon-based approach, it has similar shortcomings to similar technologies: it relies on a predefined lexicon, which may not perform well for new words or specific domain terminology not included in the lexicon. And it lacks the ability to understand context, VADER may struggle with complex semantics and polysemous words (Ribeiro et al., 2016).

RoBERTa (Robustly optimized BERT approach) is an advanced variant of BERT (Bidirectional Encoder Representations from Transformers), it refines BERT by using larger mini-batches, removing the next sentence prediction objective, and training on a larger dataset. This results in a more robust understanding of context and nuances in text, making it highly effective for sentiment analysis (Liu, 2019). RoBERTa's capability to capture subtle semantic and syntactic nuances allows for more accurate sentiment classification (Tenney, 2019). Its drawback is that, as a deep learning model, it requires a certain amount of computing resources.

Large Language Models (LLMs), such as GPT-3 and its successors, represent a significant leap in NLP capabilities. These models, trained on diverse and extensive datasets, possess a deep understanding of language, context, and sentiment. Their ability to generate human-like text and understand complex language constructs makes them particularly suited for sentiment analysis in hotel reviews. LLMs can handle the intricacies of informal language, context shifts, and implicit sentiments often found in online reviews. In terms of disadvantages, LLMs can hardly be explained (Rudin, 2019) and they may occasionally misinterpret sarcasm, irony, or context-specific slang. They performance can be computationally intensive, requiring significant resources for real-time analysis (Strubell et al., 2020).

**Prompt Engineering in LLM**

Prompt engineering is a critical subfield within the broader domain of artificial intelligence, particularly leveraging LLMs. It involves the design, optimization, and fine-tuning of prompts to elicit desired responses from language models like ChatGPT. We added some features to Prompt related to giving correct sentiment scores to ensure the reliability of responses. These are:

1. Role prompt: It is crucial in AI language models as it helps define the context and perspective from which the model should respond. Since tourists are the main consumers, we assume that the role of AI is that of a tourist, which can more accurately give the perception from the perspective of tourists. (Line 4)
2. Goal prompt: ChatGPT requires a specific question to provide tailored answers, and in our research, we need it to provide sentiment ratings for comments separately. (Line 4-8)
3. Using prompt: It is necessary to provide ChatGPT with a specific output format to obtain structured data and reduce the number of tokens used. This output result will be beneficial for subsequent data processing and comparison. (Line 8-12)
4. Zero shot prompt: Zero shot can directly utilize the capabilities of LLM to obtain answers without providing examples. This will help reduce usage costs, but the quality of the response may be lower than that of the few shots prompt.

Here is our prompt:

```
1.  Messages = [
2.      {
3.          "role": "system",
4.          "content": As a tourist looking for hotels online, your task is to analyze the sentiment of
5.  the following texts. Texts will be separated by \" | \" and given as text delimited by triple quotes.
6.  Please consider the overall tone of the discussion, the emotion conveyed by the language used, and
7.  the context in which words and phrases are used. Give the sentiment score towards the text separately,
8.  measured between -1 and 1. Your response should be in the following JSON format: \
9.          JSON = {\
10.                         \"sentiment\": [n1, n2, ..., nx] \
11.                 } \
12.                 The reply should not contain information other than the json.
13.     }
14.     {
15.                 "role": "user",
16.                 "content":
17.                         "Here are the comments: \"\"\"" + comment + "\"\"\"\n"
18.     }
19. ]
```

## EMPIRICAL RESULT

We give the variable definitions in Table 1 and the descriptive statistics in Table 2. All IVs are min-max normalized to facilitate comparison of prediction results. Based on these variables, we regress the hotel's performance for the next month.

Table 1: Measurements and source of variables in analysis

| | Variables | Measurements |
|---|---|---|
| DV | Total Revenue | Hotel monthly total revenue for month *t*, from Search Texas Tax |
| | RevPAR | Hotel monthly revenue per available room for month *t*, from Search Texas Tax |
| IV | Rating | Hotel monthly average rating of hotel reviews for the month *t-1*, from Expedia and Booking |
| | VADER Sentiment | Average sentiment score of hotel reviews text given by VADER for the month *t-1* |
| | RoBERTa Sentiment | Average sentiment score of hotel reviews text given by RoBERTa for the month *t-1* |
| | ChatGPT Sentiment | Average sentiment score of hotel reviews text given by ChatGPT for the month *t-1* |
| CV | Review Num | Hotel monthly number of reviews for month *t-1* |

Table 2: Descriptive statistics

| Variables | Obs | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|---|
| Total Revenue | 24,534 | 399603.2 | 804089.6 | 1 | 5.66E+07 |
| RevPAR | 24,573 | 103.502 | 379.352 | 0.0002 | 22082.23 |
| Booking Review Num | 22,164 | 15.936 | 18.627 | 1 | 345 |
| Expedia Review Num | 20,197 | 8.177 | 8.949 | 1 | 126 |
| Booking Rating | 22,164 | 0.750 | 0.171 | 0 | 1 |
| Expedia Rating | 20,197 | 0.704 | 0.240 | 0 | 1 |
| Booking VADER Rating | 22,164 | 0.641 | 0.122 | 0 | 1 |
| Expedia VADER Rating | 20,197 | 0.670 | 0.180 | 0 | 1 |
| Booking RoBERTa Rating | 22,164 | 0.667 | 0.195 | 0 | 1 |
| Expedia RoBERTa Rating | 20,145 | 0.539 | 0.281 | 0 | 1 |

As shown in the following 4 tables, we regressed the revenue data of hotels on Booking and Expedia, fixing the time effect. Each group uses the average rating of hotel reviews, and the average sentiment score extracted by VADER, RoBERTa, and ChatGPT as the dependent variable.

Table 3: Booking total revenue prediction

| Variables | Rating | VADER | RoBERTa | ChatGPT |
|---|---|---|---|---|
| Sentiment | 13.610*** | 15.667*** | 11.878*** | 12.783*** |
| | 9.683 | 8.176 | 9.731 | 8.634 |
| Review Num | 0.001 | -0.003 | 0.001 | -0.009 |
| | 0.042 | -0.265 | 0.093 | -0.685 |
| Total Revenue | 0.948*** | 0.952*** | 0.949*** | 0.957*** |
| | 402.015 | 413.356 | 406.303 | 416.138 |
| Observations | 21783 | 21783 | 21783 | 20,999 |
| Number of Hotel | 692 | 692 | 692 | 692 |
| TIME FE | YES | YES | YES | YES |
| R-squared | 0.511 | 0.511 | 0.511 | 0. 534 |

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Booking RevPAR prediction

| Variables | Rating | VADER | RoBERTa | ChatGPT |
|---|---|---|---|---|
| Sentiment | 20.525*** | 23.426*** | 16.820*** | 18.476*** |
| | 14.624 | 12.183 | 13.819 | 12.476 |
| Review Num | -0.046*** | -0.047*** | -0.044*** | -0.046*** |
| | -3.708 | -3.783 | -3.552 | -3.781 |
| RevPAR | 0.880*** | 0.886*** | 0.883*** | 0.895*** |
| | 272.411 | 279.666 | 276.314 | 282.778 |
| Observations | 21820 | 21820 | 21820 | 21035 |
| Number of Hotel | 693 | 693 | 693 | 693 |
| TIME FE | YES | YES | YES | YES |
| R-squared | 0.535 | 0.535 | 0.535 | 0.557 |

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Expedia total revenue prediction

| Variables | Rating | VADER | RoBERTa | ChatGPT |
|---|---|---|---|---|
| Sentiment | 10.129*** | 11.062*** | 5.179*** | 8.992*** |
| | 9.809 | 8.213 | 6.021 | 8.88 |
| Review Num | 0.132*** | 0.131*** | 0.127*** | 0.124*** |
| | 4.58 | 4.513 | 4.371 | 4.25 |
| Total Revenue | 0.945*** | 0.947*** | 0.949*** | 0.946*** |
| | 362.684 | 368.165 | 371.169 | 363.685 |

| | | | | |
|---|---|---|---|---|
| Observations | 19869 | 19869 | 19818 | 19782 |
| Number of Hotel | 671 | 671 | 670 | 671 |
| TIME FE | YES | YES | YES | YES |
| R-squared | 0.514 | 0.514 | 0.512 | 0.513 |

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Expedia RevPAR prediction

| Variables | Rating | VADER | RoBERTa | ChatGPT |
|---|---|---|---|---|
| Sentiment | 14.380*** | 16.471*** | 6.683*** | 12.859*** |
| | 13.864 | 12.151 | 7.763 | 12.631 |
| Review Num | 0.024 | 0.029 | 0.031 | 0.016 |
| | 0.859 | 1.049 | 1.109 | 0.579 |
| RevPAR | 0.879*** | 0.883*** | 0.889*** | 0.881*** |
| | 247.467 | 251.174 | 255.808 | 248.091 |
| Observations | 19906 | 19906 | 19855 | 19819 |
| Number of Hotel | 672 | 672 | 671 | 672 |
| TIME FE | YES | YES | YES | YES |
| R-squared | 0.542 | 0.541 | 0.54 | 0.541 |

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Firstly, our study shows that the regression coefficients for all measures are positive and statistically significant, indicating that more positive reviews are associated with an increase in the hotel's total revenue and RevPAR for the next month. It is consistent with previous research, which demonstrates that consumers often rely on reviews when choosing a hotel, with positive reviews increasing their likelihood of booking a particular hotel (Ye et al., 2009). Secondly, we observe a divergence in the prediction results of the sentiment scores generated by ratings, traditional sentiment analysis tools, and ChatGPT. The ratings and sentiment scores derived from traditional tools exhibited stronger correlations with revenue metrics compared to those produced by ChatGPT, suggesting that customers reply on the heuristic cues rather than systematic cues in decision making process.

## CONTRIBUTION AND DISCUSSION

This study investigates the predictive power of sentiment score extracted by ChatGPT on sales performance, in comparison to compared to numerical ratings and traditional sentiment analysis tools. Our findings reveal that reviews with more positive sentiment are significantly associated with higher hotel revenue the following month, which is consistent with previous research. However, despite the superior capabilities of ChatGPT in processing textual analysis, it does not outperform numerical ratings or traditional sentiment analysis tools in predicting sales outcomes. We suggest that customers do not actively engage in systematic cues processing, as ChatGPT does, instead, they reply on heuristic cues, such as ratings and key emotional terms. Theoretically, this study contributes to extending HSM by illustrating how AIGC moderates heuristic and systematic information processing in sales prediction. Empirically, it provides insights into the comparative effectiveness of sentiment analysis tools, offering practical implications for e-commerce platforms to the adoption of AIGC in sales and operation strategies.

For future research direction, we suggest exploring the performance of ChatGPT across different contexts to further understand the applications and limitations of LLMs. Furthermore, previous studies have noted that LLMs are prone to absorbing and amplifying social biases presented in training data, such as gender, racial and cultural biases (Bender et al., 2021). LLMs may also struggle to accurately interpret emotions in reviews that contain sarcasm or irony, potentially leading to misjudgments (Bommasani et al., 2021). Future research could delve deeper into these challenges. This study also has certain limitations. LLMs technology is evolving rapidly, and as these models improve, the predictive power of sales forecasting based on review sentiment may change. Moreover, while ChatGPT excels in natural language understanding, it may not fully capture the complexity, nuances, or cultural differences embedded in user comments, particularly those involving sarcasm or irony.

## REFERENCES

Arnott, D., & Gao, S. (2019). Behavioral economics for decision support systems researchers. *Decision Support Systems*, *122*, 113063. https://doi.org/10.1016/j.dss.2019.05.003

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (FAccT 2021), Online, March 3-10, 610–623. https://doi.org/10.1145/3442188.3445922

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., & Brunskill, E. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. https://doi.org/10.48550/arXiv.2108.07258

Bond, S. A., Klok, H., & Zhu, M. (2023). Large language models and financial market sentiment. *Available at SSRN 4584928*. https://doi.org/10.2139/ssrn.4584928

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, *43*(3), 345-354. https://doi.org/10.1509/jmkr.43.3.345

Cui, G., Lui, H.-K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, *17*(1), 39-58. https://doi.org/10.2753/JEC1086-4415170102

Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, *21*(4), 23-45. https://doi.org/10.1002/dir.20087

Dong, M. M., Stratopoulos, T. C., & Wang, V. X. (2023). A Scoping Review of ChatGPT Research in Accounting and Finance. *Available at SSRN 4680203*. *https://doi.org/10.1016/j.accinf.2024.100715*

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007-1016. https://doi.org/10.1016/j.dss.2008.04.001

Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists*, (IMECS 2019), Hong Kong, China, March 13-15, *122*(16).. https://doi.org/10.1142/9789811215094_0005

Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. In *Information and communication technologies in tourism 2008* (pp. 35-46). Springer. http://dx.doi.org/10.1007/978-3-211-77280-5_4

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, (SIGKDD 2004), Seattle WA, USA, August 22 - 25, 168-177. https://doi.org/10.1145/1014052.1014073

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings Lead You to the Product, Reviews Help You Clinch it? The Mediating Role of Online Review Sentiments on Product Sales. *Decision Support Systems*, *57*, 42-53. https://doi.org/10.1016/j.dss.2013.07.009

Hu, N., Liang, P., & Yang, X. (2023). Whetting all your appetites for financial tasks with one meal from GPT? A comparison of GPT, FinBERT, and dictionaries in evaluating sentiment analysis. *Available at SSRN 4426455*.

Li, H., Bruce, X., Li, G., & Gao, H. (2023). Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews. *Tourism Management*, *96*, 104707. https://doi.org/10.1016/j.tourman.2022.104707

Liu, Y. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*. https://doi.org/10.48550/arXiv.1907.11692

Liu, Z. M., Liu, L., & Li, H. (2012). Determinants of Information Retweeting in Microblogging. *Internet Research*, *22*(4), 443-466. https://doi.org/10.1108/10662241211250980

Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*. https://doi.org/10.48550/arXiv.2304.07619

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The journal of finance*, *66*(1), 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Luca, M. (2016). Reviews, reputation, and revenue: The case of Yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*(12-016).http://doi.org/10.2139/ssrn.1928601

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices and applications. *International Journal of Scientific and Technology Research*, *9*(2), 601-609.

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information systems research*, *17*(4), 392-414. https://doi.org/10.1287/isre.1060.0106

Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*, 1-29. https://doi.org/10.1140/epjds/s13688-016-0085-1

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, *1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, *32*(6), 1310-1323. https://doi.org/10.1016/j.tourman.2010.12.011

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on artificial intelligence*, (AAAI 2020), New York, USA, February 7-12, 34(09), 13693-13696. https://doi.org/10.1609/aaai.v34i09.7123

Tenney, I. (2019). BERT rediscovers the classical NLP pipeline. *arXiv preprint arXiv:1905.05950*. https://doi.org/10.48550/arXiv.1905.05950

Todorov, A., Chaiken, S., & Henderson, M. D. (2002). The Persuasion Handbook: Developments in Theory and Practice. In. SAGE Publications, Inc. https://doi.org/10.4135/9781412976046

Tversky, A., & Kahneman, D. (1974). Judgment Under Uncertainty - Heuristics and Biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, *30*(1), 123-127. https://doi.org/10.1016/j.tourman.2008.04.008

Wang, C. (2023). Outsourcing Voting to AI: Chan ChatGPT Advise Index Funds on Proxy Voting Decisions? *Fordham J. Corp. & Fin. L.*, *29*, 113-189. http://doi.org/10.2139/ssrn.4413315

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837. https://doi.org/10.48550/arXiv.2201.11903

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180-182. https://doi.org/10.1016/j.ijhm.2008.06.011

Zhang, B., Yang, H., & Liu, X.-Y. (2023). Instruct-fingpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. *arXiv preprint arXiv:2306.12659*. https://doi.org/10.48550/arXiv.2306.12659

Zhang, K. Z. K., Zhao, S. J., Cheung, C. M. K., & Lee, M. K. O. (2014). Examining the Influence of Online Reviews on Consumers' Decision-Making: A Heuristic-Systematic Model. *Decision Support Systems*, *67*, 78-89. https://doi.org/10.1016/j.dss.2014.08.005