Typing Less, Saying More? – The Effects of Using Generative AI in Online Consumer Review Writing

Research Paper

Maximilian Habla[1]

[1] Ulm University, Institute of Business Analytics, Ulm, Germany

Abstract. Generative AI (GenAI) is reshaping various everyday tasks, but little is known about its effect on online consumer review writing. This study investigates how using GenAI affects the informativeness of reviews. Through a scenario-based online experiment, we find that reviewers using GenAI perceive a reduced cognitive load while writing more informative reviews. Contrary to our expectation, cognitive load does not mediate the effect of using GenAI on the informativeness of reviews. Further analyses suggest that reviews written using GenAI exhibit higher linguistic complexity and more positive sentiment. Our findings contribute to literature on cognitive load theory, GenAI, and review writing, while offering valuable insights for practitioners. We propose avenues for digital platforms to draw from the increased informativeness while mitigating potential drawbacks associated with using GenAI. In summary, our results ensure that reviews remain effective in informing consumer purchasing decisions.

Keywords: Online Reviews, Informativeness, GenAI, Cognitive Load Theory.

1 Introduction

Online consumer reviews (hereafter referred to as “reviews”) have become a fundamental component of digital platforms, with approximately 90% of consumers consulting them to inform their purchasing decisions (Turner & Rainie, 2020). However, the mere presence of a large number of reviews is insufficient, as reviews must also be informative to effectively support consumer purchasing decisions (Mudambi & Schuff, 2010). Informativeness in this context refers to the extent to which a review covers a range of distinct aspects (e.g., “Pizza”, “Salad”) and broader topics (such as “Food” or “Service”) relevant to the product (Habla et al., 2024). When reviews provide broad and diverse product aspects, they enable consumers to make well-founded comparisons and, ultimately, informed purchasing decisions (Scholz & Dorner, 2013; Son et al., 2019). Writing informative reviews is a cognitively demanding process that requires reviewers to recall their product experiences, select relevant aspects, and structure their thoughts (Poniatowski et al. 2019). To facilitate this, digital platforms often employ structured review templates. These templates use strategic design features to support reviewers in their review writing (Gutt et al. 2019). For instance, Yelp prompts reviewers to provide aspects on topics such as food and service in restaurant reviews, while

TripAdvisor provides structured questions that guide reviewers step by step through their review writing. Although such templates support reviewers in writing informative reviews, they share one common limitation: they still require reviewers to recall and manually type their product experiences, leaving a substantial cognitive load. Given this challenge, reviewers may increasingly turn to Generative AI (GenAI) as a tool to support their review writing. GenAI has demonstrated remarkable capabilities across everyday tasks (Zemke et al. 2025) and now reaches hundreds of millions of users weekly via tools like ChatGPT (Rooney, 2025) and built-in assistants such as Windows Copilot and Apple Intelligence. Particularly in writing contexts, GenAI has proven to enhance structure, clarity, and fluency while simultaneously reducing cognitive load (Noy & Zhang, 2023). While the benefits of using GenAI are clear to reviewers, its effect on the informativeness of their reviews remains uncertain. Thus, it is crucial to understand whether reviewers invest the freed cognitive resources by using GenAI to include a broader and more diverse set of product aspects in their reviews. Consequently, we state the following research question:

How does using GenAI affect the informativeness of reviews?

To address our research question, we develop hypotheses based on cognitive load theory and conduct a scenario-based online experiment that ensures realism (Fink, 2022). Depending on the experimental group, participants write a restaurant review using one of two review templates designed after Yelp. In the treatment group, the review template embeds GenAI, while participants in the control group write a review without GenAI support. This design allows us to systematically compare the effects of using GenAI to traditional review writing on the informativeness of reviews. The results reveal that reviews written using GenAI are more informative, while their reviewers perceive a lower cognitive load. However, contrary to our expectations, we find that cognitive load does not mediate the relationship between using GenAI and the informativeness of reviews. These findings contribute to theory and practice, providing a basis for ensuring that reviews remain effective in informing consumer purchasing decisions.

2 Related Literature

The majority of existing research on reviews emphasizes their economic implications, such as their effect on purchase intentions and sales (see Babić Rosario et al. 2020), or their role in informing consumers' purchasing decisions (Rietsche et al. 2019, for an overview). Central to these studies is the recognition that informativeness affects a review's effectiveness in supporting these decisions (Mudambi & Schuff, 2010; Son et al. 2019). To enhance the informativeness of reviews, researchers examined various design features. Financial incentives, for instance, increase the number but often result in less informative reviews (Burtch et al. 2018; Khern-am-nuai et al. 2018). Conversely, non-financial incentives like gamification first encourage informative review writing, but the effect diminishes over time and can introduce sentiment bias (Moro et al. 2019). Contextual factors also play a role, for example showing the number of existing reviews negatively correlates with informativeness (Rohde et al., 2022), while multidimensional rating scales encourage discussion of more product aspects (Chen et al., 2018).

Gutt et al. (2019) propose that review templates, a set of strategically selected design features, positively affect the number and content of reviews. Consistent with this, further research finds that review templates improve the informativeness of reviews but do not reduce reviewers' cognitive load (Habla et al. 2024; Poniatowski et al. 2019).

Recent GenAI literature adds another dimension to this discussion (see Zemke et al. (2025)). A growing number of studies demonstrate GenAI's potential to support writing tasks across educational and professional settings, thereby improving structure, clarity, and fluency while reducing the perceived cognitive load (Li et al. 2024; Noy & Zhang, 2023). Yet, only a few studies have investigated using GenAI in the context of reviews. To date, Erlebach (2024) finds that GenAI effectively supports consumers in processing a large number of reviews. Moreover, Knight et al. (2024) and Amos & Zhang (2024) suggest that reviews tagged as "AI-written" are generally perceived as less effective. In addition, Kovács (2024) asked ChatGPT to rewrite Yelp reviews and let consumers compare them, suggesting that they often could not distinguish between them.

In summary, strategically selecting design features has already been demonstrated to support reviewers in writing more informative reviews. To the best of our knowledge, however, no prior study investigates how review writing using GenAI affects the informativeness of reviews. This leaves a gap that this research aims to fill.

3 Theoretical Background and Hypotheses Development

We develop our hypotheses based on cognitive load theory (Sweller et al. 1998; Sweller et al. 2011; Sweller et al. 2019), which explains how the limitations of working memory constrain human cognitive architecture (Paas et al. 2010). When tasks require cognitive resources that exceed working memory capacities, negative outcomes, such as reduced information processing efficiency, are likely to occur. Cognitive load theory distinguishes between two primary components: intrinsic cognitive load and extraneous cognitive load (Skulmowski & Xu, 2022). Intrinsic cognitive load arises from the inherent complexity of the task itself and is heavily affected by an individual's prior knowledge. Consequently, intrinsic cognitive load can only be effectively reduced through changes in either the task itself or in the individual's existing knowledge related to the task (Erlebach, 2024). Conversely, extraneous cognitive load emerges from cognitive activities that are not associated with prior knowledge. Taken together, these two components jointly constitute an individual's perceived cognitive load for a given task.

Translating this to the context of review writing, reviewers experience intrinsic cognitive load primarily when recalling aspects of their product experience and making evaluative judgments. Extraneous cognitive load arises to a certain extent from the cognitive processes associated with structuring and organizing their written reviews. Recent studies indicate that extraneous cognitive load in writing tasks is affected by structural complexity, suggesting that strategies such as writing in chunks can reduce extraneous cognitive load (Putri et al. 2022; Rahmat, 2023). Building upon this, we posit that using GenAI can reduce reviewers' perceived cognitive load by supporting them in structuring their experiences, refining linguistic expression, and writing a coherent review. Accordingly, we state our first hypothesis as follows:

***H1**: Using GenAI reduces reviewers' perceived cognitive load.*

Furthermore, cognitive load theory posits that reducing cognitive load facilitates the redistribution of some of the freed cognitive resources toward other cognitive processes, thus enhancing overall task performance (Sweller et al. 1998; Sweller et al. 2011). In the context of review writing, the primary task is to write reviews that effectively support consumers in informing their purchasing decisions by describing and evaluating product experiences (Son et al. 2019). Using GenAI, by reducing cognitive load associated with structuring and writing, may enable reviewers to allocate a certain extent of the freed cognitive resources toward recalling and evaluating more aspects of their product experience. This may enhance the informativeness of their reviews. In other words, the overall reduction of the cognitive load using GenAI might lead to more informative reviews. Therefore, we suggest a mediating role of perceived cognitive load between using GenAI and the informativeness of reviews. Thus, we state our second hypothesis as follows:

***H2**: The positive effect of using GenAI on the informativeness of a review is mediated by reviewers' perceived cognitive load.*

It is important to note that our theoretical foundation in cognitive load theory does not suggest a direct effect of using GenAI on the informativeness of reviews. The theory posits a clear causal chain where performance improvements on a primary task are the result of freed cognitive resources that were previously occupied by cognitive processes. In context of our study, this implies that GenAI enhances informativeness because it first reduces cognitive load associated with structuring and writing the review. The freed cognitive resources can then be reallocated to primary cognitive tasks of recalling and evaluating more diverse product aspects. Thus, the relationship between using GenAI and informativeness of reviews is expected to be fully mediated by cognitive load. Our research model, summarizing our hypotheses, is outlined in Figure 1.

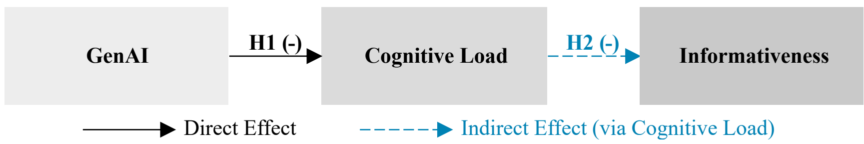

Figure 1. Research Model

4 Research Methodology and Experimental Design

We employ a scenario-based online experiment to investigate how using GenAI affects the informativeness of reviews. This allows us to isolate causal effects while controlling for confounding factors and balancing realism with controlled conditions (Fink, 2022).

4.1 Scenario Description

Among various industries, the restaurant sector is particularly impacted by reviews, as consumers frequently rely on them to inform themselves about food quality, service, and overall experience before deciding on a particular restaurant (Jeong & Jang, 2011).

Given this critical role, our study focuses on the restaurant sector to analyze how using GenAI affects the informativeness of reviews. To mirror real-world behavior, participants in our online experiment are instructed to write a review based on their most recent visit to a Mexican restaurant. We opt for this to ensure comparability across reviews by focusing on a specific cuisine. Furthermore, asking participants to draw on personal experiences, rather than hypothetical ones, increases the ecological validity and authenticity of their written reviews.

4.2 Treatment Variation

We employ a between-subjects experimental design, where participants are randomly assigned to one of two experimental groups: a control group or a treatment group. In the control group, they are presented with a review template designed after Yelp. The template includes three design features: a star rating, proposing topics to review (i.e., food, service, and ambiance), and a text input field (see Figure 2, Panel A). This represents a traditional review template approach and establishes a baseline for comparison.

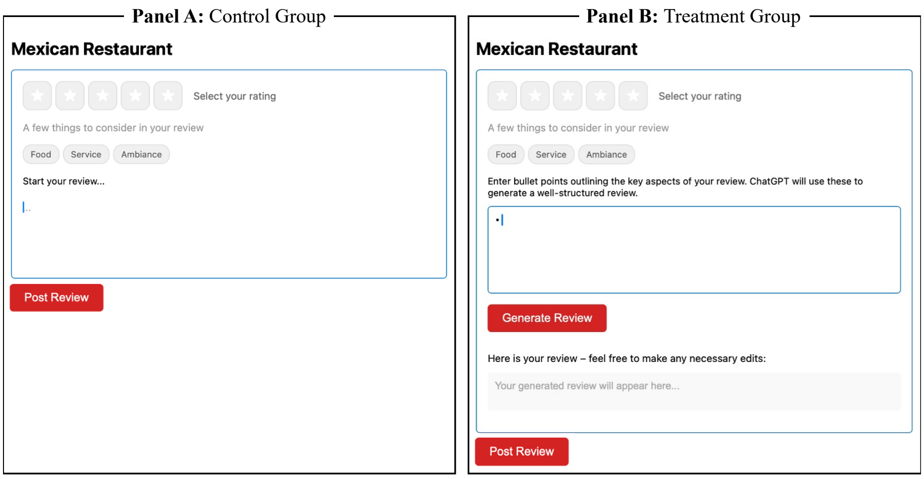

Figure 2. Review Templates

In contrast, participants in the treatment group are presented with a slightly modified version of Yelp's review template. While this version retains the same fundamental structure, it embeds GenAI, allowing participants to provide aspects of their experience in bullet point form. Upon clicking on "Generate Review", the bullet points are processed by ChatGPT-4o[1], which generates a coherent text that appears directly below the input field (see Figure 2, Panel B). To ensure the embedding represents interaction with GenAI as realistically as possible, we do not employ any form of prompt engineering to refine the generated text, such as instructing ChatGPT to write a certain length. More specifically, ChatGPT is simply instructed to "write a review for a

[1] We opt for ChatGPT, as it is one of the most popular GenAI tools (Zemke et al. 2025).

Mexican restaurant based on the following input:". However, participants can refine the generated text by manually editing it or modifying their input to have ChatGPT regenerate it. Embedding ChatGPT into the review template facilitates systematic tracking of participants' interactions, enabling detailed analysis of how using GenAI affects the informativeness of reviews. To this end, we disabled the copy-and-paste function, preventing participants in the control group from using external GenAI tools, while allowing precise tracking of GenAI interactions within the treatment group.

4.3 Variables[2]

To evaluate our hypotheses, we draw on both the participants' responses to the post-experimental questionnaire and their written reviews. Below, we detail how we operationalize each variable, including relevant control measures.

We measure participants' perceived *Cognitive Load*[3] using one construct from the NASA Task Load Index (Hart & Staveland, 1988) and a 7-point Likert scale, ranging from 1 (low) to 7 (high).

To comprehensively assess the informativeness of the written reviews, we adopt the procedure proposed by Habla et al. (2024), designed to account for the breadth and diversity of product information in reviews. Informativeness is thus operationalized through two distinct yet complementary variables. Firstly, the *Number of Aspects* mentioned in each review is assessed to capture the breadth of product-specific information provided. Secondly, we measure the diversity of information provided through the *Number of Topics* discussed in each review. Consistent with Habla et al. (2024), we establish nine topics that are commonly relevant to restaurant reviews: "Food", "Service", "Drinks", "Cleanliness", "Atmosphere/Ambience", "Value-For-Money", "Location", "Kid-friendliness", and "Wheelchair accessibility." However, to extract the aspects from reviews, we employ the PyABSA framework (Yang et al. 2023). The aspects extracted by this method are subsequently lemmatized and deduplicated, thereby ensuring each aspect is counted only once per review, resulting in the measure referred to as the *Number of Aspects*. Next, those aspects are mapped onto the predefined topics employing a topic modeling approach based on the Word2Vec algorithm (Mikolov et al. 2013). In instances where multiple aspects within a single review pertain to the same topic, the topic is counted only once, thereby determining the *Number of Topics*.

To account for participants' prior experiences and potential external influences, several control variables are included to ensure the robustness of our results. Specifically, participants' familiarity with writing and reading reviews is assessed (Habla et al. 2024; Lee & Jin Ma, 2012), along with the recency of their last visit to a Mexican restaurant. Additionally, prior usage of GenAI (Dhillon et al. 2024) and attitudes toward writing (Graham et al. 2017) are examined. Finally, sociodemographic information, including age and sex, is collected to provide further context for the analysis.

---

2 We provide details on all variables used and the extended descriptive statistics in an online appendix available via the following link: https://github.com/ReviewGenAI/OnlineAppendix

3 In the original version, the construct is called mental demand, but it corresponds to cognitive load (Gieshoff & Heeb, 2023).

4.4 Experimental Procedure

The experimental procedure begins by introducing participants to the scenario. At this stage, participants are explicitly informed about present attention checks, clarifying that failing these checks will result in their exclusion without payment. Following their consent, participants proceed to a fictitious review writing platform. Depending on their experimental group, participants write their review either with a traditional review template (control group) or a review template including GenAI (treatment group). Upon completing the review-writing task, participants are required to submit their review by clicking "Post Review", an action enabled only after selecting a star rating and typing at least one character. After submission, participants are redirected to a post-experimental questionnaire. To confirm that participants correctly observe the treatment variation, we first ask whether GenAI was used to write the review as a manipulation check. Subsequently, participants respond to questions regarding their perceived cognitive load during review writing and the control variables. Notably, our attention checks are integrated into the questionnaire[4]. This ensures that only attentive participants contribute data, increasing the validity of our results.

4.5 Implementation and Data Collection

The experiment was implemented using the web-based survey platform SoSci Survey. To ensure clarity and eliminate potential ambiguities, we conducted an extensive pretest following established guidelines (Reynolds & Diamantopoulos, 1998). We first tested our implementation with students from a German university and collected feedback. We then pre-tested with participants from Prolific, a widely used crowd-working platform known for high-quality academic research (Peer et al. 2021). Collecting this feedback ensured that the experimental procedure and task instructions were clear. As no issues were reported, we proceeded with the main experiment. Participants in the pre-test and the main experiment were required to be at least 18 years old, current U.S. residents, and native English speakers. After completing the experiment, participants received $0.70 in compensation, in line with Prolific's standards.

5 Analyses and Results

In total, 87 participants correctly answered the manipulation and attention check questions and completed the experiment. Prior to the analyses, we excluded speeders who responded at a pace outside the 1st percentile range (2 observations), as recommended by Ford (2017). Thus, our final dataset comprises 85 participants randomly assigned to two experimental groups. The average age of participants was 39 years, with a distribution of 59 female and 26 male participants[5]. The experiment took approximately four

---

[4] Importantly, these questions are asked after the evaluation of the perceived cognitive load to avoid any biases in participants' answers.

[5] Additional demographic information and descriptive statistics can be found in our online appendix: https://github.com/ReviewGenAI/OnlineAppendix

minutes to complete, with the review-writing task itself averaging one minute and 45 seconds. To ensure the internal consistency of our scales, we assess Cronbach's alpha for our multi-item variable (Cortina, 1993). As Cronbach's alpha exceeds the recommended threshold of 0.7 (Hair et al. 2009), all proposed items are retained and aggregated for the respective variable. We verified the appropriateness of our sample randomization by comparing the control variables between the experimental groups. Specifically, we conducted Wilcoxon rank-sum tests for ordinal, t-tests for ratio-scaled, and Chi-square tests for nominal variables. As no significant differences emerged between the experimental groups, we conclude that our experiment is well-randomized. To get a first impression of our data, Table 1 provides the summary statistics of our dependent variables.

Table 1. Summary Statistics of the Dependent Variables

	Treatment (n = 43)		Control (n = 42)		Differences
Variable	Mean	SD	Mean	SD	in Means
Cognitive Load	2.558	1.452	3.500	1.452	-0.942^{***}
Number of Aspects	4.372	1.448	2.595	1.170	1.777^{***}
Number of Topics	2.558	0.765	1.786	0.717	0.772^{***}

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. *Statistical significance for differences is based on Wilcoxon rank-sum test (row 1), t-tests (row 2 & row 3). SD denotes standard deviation.*

5.1 Main Results

To analyze the effect of using GenAI on reviewers' perception of *Cognitive Load* (i.e., H1), we estimate an ordinary least squares (OLS) model:

$$\textit{Cognitive Load} = \alpha + \beta\ \textit{GenAI} + \gamma\ \textit{Controls} + \epsilon \tag{1}$$

where *Cognitive Load* represents the dependent variable of interest. *GenAI* is the binary independent variable, it denotes either using *GenAI* (encoded as one) or not using *GenAI* (encoded as zero). The model further includes all our control variables (i.e., *Controls*) as outlined in Section 4.3, for example, the recency of the participant's last restaurant visit, along with the remaining error term (i.e., ϵ). Robust standard errors are used. Panel A of Table 2 presents the estimation results. We observe a significantly negative coefficient for *GenAI* ($\beta = -0.985$, $p < 0.01$), which indicates that reviewers using *GenAI* experience less *Cognitive Load* compared to those not using *GenAI*. This result suggests that reviewers using *GenAI* perceive substantially lower *Cognitive Load* during review writing. Thus, we find support for H1.

To investigate whether perceived *Cognitive Load* mediates the relationship between using *GenAI* and the informativeness of reviews (i.e., H2), we conduct a mediation analysis employing the PROCESS macro Model 4 for R (Hayes, 2022). Following Hayes (2022), we assess the statistical significance of the effects via a bootstrapping procedure with 5000 bias-corrected resamples. In addition, we base standard errors and 95% confidence intervals. We analyze informativeness by means of two separate measures: the *Number of Aspects* provided, and the *Number of Topics* addressed in reviews. The results are summarized in Panel B of Table 2. For the *Number of Aspects*, using *GenAI* exhibits a significant positive direct effect (β = 1.915, $p < 0.01$). This indicates reviewers using *GenAI* provided approximately two additional aspects compared to those not using *GenAI*. The mediation analysis revealed no significant indirect effect, suggesting that the perceived *Cognitive Load* does not mediate the effect of using *GenAI* on the *Number of Aspects*. Similarly, for the *Number of Topics*, using *GenAI* also has a significant positive direct effect (β = 0.801, $p < 0.01$), indicating that reviewers using *GenAI* addressed experiences to a greater *Number of Topics* in their reviews. Again, we find no mediation through *Cognitive Load*. Therefore, we conclude that the perceived *Cognitive Load* does not serve as a mediator between using *GenAI* and the informativeness of reviews, leading us to reject H2.

To ensure the robustness of our findings, we analyzed whether manually editing the AI-generated text influenced the outcomes. A separate regression among treatment group participants, which included a binary indicator for manual edits, showed no significant effect on either informativeness (i.e., *Number of Aspects, Number of Topics)* or *Cognitive Load*. This confirms that the observed effects stem from using GenAI itself.

Table 2. Main Results

Panel A. Effects on Cognitive Load			
	Effect	SE	
GenAI → Cognitive Load (H1)	-0.985^{***}	0.328	
Panel B. Effects on Informativeness			
	Effect	SE	95% CI
GenAI → Number of Aspects	1.915^{***}	0.343	[1.232;2.598]
GenAI → Number of Topics	0.801^{***}	0.201	[0.401;1.202]
GenAI → Cognitive Load → Number of Aspects (H2)	-0.028	0.126	[-0.324;0.200]
GenAI → Cognitive Load → Number of Topics (H2)	-0.019	0.064	[-0.165;0.101]

In Panel A: Control variables and robust standard errors (SE) are used.
In Panel B: Standard errors (SE) and 95% confidence intervals (CI) for the indirect effect are based on 5000 bootstrapping resamples. Mediation analyses are based on the PROCESS macro Model 4 (Hayes, 2022).
Further Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. *Notably, none of the control variables showed a statistically significant effect on the dependent variables in any of the analyses.*

5.2 Additional Insights

Our results suggest that reviews written using GenAI are more informative, as measured by both a higher *Number of Aspects* and *Number of Topics*. To confirm that this increase comes from reviewers providing more aspects to GenAI, rather than GenAI artificially amplifying the content, we conduct a further analysis as a robustness check. Thus, we compare the *Number of Aspects* provided by reviewers in the input text field to ChatGPT with those present in their posted reviews. A t-test reveals no significant difference between the *Number of Aspects* provided in the input text field and the posted review ($p > 0.1$). Specifically, reviewers using GenAI provided about four aspects in their input fields (see Table 3, row 1), like the *Number of Aspects* in their posted review (see Table 1, row 1). This implies that the increased informativeness of reviews written using GenAI stems from reviewers providing more aspects of their experience.

Table 3. Further Measures

Variable	Treatment (n = 43) Mean	Treatment (n = 43) SD	Control (n = 42) Mean	Control (n = 42) SD	Differences in Means
Number of Aspects GenAI Input	4.419	1.867	-	-	-
Linguistic Complexity	15.882	1.368	10.024	3.956	5.857^{***}
Sentiment	0.917	0.277	0.607	0.516	0.310^{***}
Star Rating	4.279	0.734	4.190	0.994	0.089

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. *Statistical significance for differences is based on t-tests. SD denotes standard deviation. Significance is preserved applying a regression analysis including all controls.*

Beyond informativeness, GenAI may further affect the written reviews, potentially impacting the effectiveness in informing consumer purchasing decisions. Prior research finds that text written using GenAI is more grammatically refined and complex than human writing (Herbold et al. 2023). As linguistic complexity impairs comprehension, this could reduce the effectiveness of the review (Ghose & Ipeirotis, 2011; Korfiatis et al. 2012). Thus, we assess linguistic complexity of reviews by calculating the Gunning Fog Index. An index incorporating sentence length and advanced vocabulary proportion to estimate the required years of education for comprehension. Results suggest that reviews written using GenAI require more than 15 years of education to be easily comprehensible, compared to about 10 years for reviews written not using GenAI. A t-test reveals a significant increase in linguistic complexity ($p < 0.01$; see Table 3, row 2).

Sentiment constitutes another crucial dimension of reviews informing consumer purchasing decisions (Yin et al. 2014). Importantly, texts written using GenAI often adopt a more positive sentiment (e.g., Hohenstein et al. 2023). To analyze whether this extends to reviews, we employ the VADER sentiment analysis framework (Hutto & Gilbert, 2014), which rates textual sentiment from -1 (negative) to +1 (positive). The results reveal that reviews written using GenAI tend to be very positive, with a sentiment score of 0.917 compared to 0.607 for reviews written without GenAI. A t-test supports a significant increase in sentiment ($p < 0.01$; see Table 3, row 3). To determine whether this shift in sentiment stems from more positive product experiences among reviewers using GenAI, we analyze the star ratings associated with their reviews. Reviewers using GenAI and not using GenAI both rate their experience with around four stars (see Table 3, row 4). According to a t-test, there is no significant difference in the star ratings ($p > 0.1$), suggesting that the shift toward more positive sentiment is predominantly textual.

6 Discussion and Conclusion

Our study contributes to the literature by integrating cognitive load theory (Sweller et al. 1998; Sweller et al. 2011; Sweller et al. 2019) into the context of review writing using GenAI. In doing so, we extend prior research and provide a theoretical basis for understanding how using GenAI affects reviewers' perceived cognitive load. In this sense, our findings suggest that using GenAI reduces reviewers' perceived cognitive load. However, contrary to cognitive load theory, which posits that decreasing cognitive load should free cognitive resources to enhance performance in other demanding parts of a task, such as recalling diverse product experiences, we do not find evidence supporting cognitive load as a mediator for the informativeness of reviews. One possible explanation is that not all reviewers allocate the newly freed cognitive resources from using GenAI in the same way. Some reviewers may reinvest these resources to provide more aspects of their product experience, while others, already satisfied with their input, do not. Alternatively, reviewers might allocate these cognitive resources towards evaluating and adapting the GenAI-generated text to better fit their requirements (Lee et al. 2025), rather than using them to recall and analyze additional product aspects. Both explanations disrupt the anticipated mediation pathway that follows cognitive load theory. Nevertheless, GenAI could still enhance the informativeness of reviews through other means. Thus, our results indicate a direct effect of GenAI on the informativeness of reviews, despite the absence of an indirect effect mediated by perceived cognitive load. In conclusion, we contribute to literature by finding that using GenAI can increase the informativeness of reviews. Further, we lay an important theoretical foundation for understanding the cognitive processes associated with using GenAI in review writing.

The findings of this study also contribute valuable insights for practitioners, highlighting the potential implications of using GenAI to support review writing. Using GenAI can increase the informativeness of reviews, by addressing a broader range of product aspects and diverse topic coverage. Consequently, consumers stand to benefit from a broader informational base when informing purchasing decisions. However, alongside these benefits, our results uncover critical concerns that practitioners must address. Reviews written using GenAI exhibit a significantly higher linguistic complexity, corresponding to approximately 15.9 years of education. In contrast, reviews written without GenAI exhibit a complexity level equivalent to around 10 years of education, aligning more closely with the average reading level of the general population (OECD, 2024). This gap raises accessibility concerns, as many consumers may find it difficult to fully comprehend the more complex language of reviews written using GenAI. As a result, even with their increased informativeness, the practical value of these reviews may be diminished. Digital platform providers should therefore consider implementing strategies to ensure that the added informativeness does not come at the cost of accessibility, such as simplifying GenAI outputs, offering user-specific readability options, or integrating summaries tailored to a broader audience. Moreover, our findings reveal that reviews written using GenAI tend to express a more positive sentiment, even though there is no corresponding change in star ratings. This observation suggests that negative aspects of the product experience are understated if the textual sentiment of the additional aspects added to the review appears more favorable than the actual product experience. Such a shift presents risks for consumers, who may be misled by the diverse aspects evaluated in the review. To mitigate these concerns, digital platform providers must implement strategies such as GenAI detection mechanisms that flag review writing using GenAI or develop guidelines to account for sentiment shifts. Another approach is to integrate review summaries into their digital platforms, which can benefit from the increased informativeness of reviews while mitigating sentiment shifts and linguistic complexity.

Our study has some limitations, which serve as valuable starting points for future research. First, our study focused on the review writer's perspective, although using GenAI leads to more informative reviews, future research should analyze whether the reviews also effectively support consumers in informing their purchasing decisions. Second, we evaluated using GenAI in a controlled online experiment within the context of restaurant reviews. This approach allowed for the careful isolation of effects, but it did not fully replicate real-world conditions. While we controlled for the recency of the last visit, reviews are often written immediately after an experience, and our experimental design may not fully capture the nuances associated with in-the-moment review writing. Further, we evaluated the effects of using GenAI for only one product category, but the effects may vary across different products (e.g., electronics). Thus, future research should examine using GenAI in real-world settings and investigate whether our results generalize to other products. Third, we focused on a single GenAI model. This choice enabled an exploration of review writing using GenAI, but different GenAI models can vary in text generation, originating from their training data (Zemke et al. 2025). Therefore, future research should replicate this study with other GenAI models (e.g., Google Gemini) to underscore our findings. Fourth, although we find that using GenAI reduces the perceived cognitive load and increases the informativeness of reviews, our results reveal no mediation effect. This finding contradicts assumptions from cognitive load theory and suggests that additional factors must be considered. Therefore, future research should replicate the experiment and focus on other influencing factors to further understand how using GenAI affects the informativeness of reviews. Fifth, we aimed to understand how using GenAI affects the informativeness of reviews in general. Hence, we embedded ChatGPT into Yelp's review template to ensure that all reviewers' inputs and GenAI outputs were collected. As this approach shows promise in supporting reviewers in writing more informative reviews, future research should follow the design science research paradigm and develop a refined GenAI review template. This can include tailored prompt designs and strategic integration of design features. The resulting artifact can combine the advantages of using GenAI, such as increased informativeness, with the reduction of potential drawbacks.

In summary, the objective of this study is to investigate how using GenAI affects the informativeness of reviews. To this end, we conducted an online experiment in which reviewers either wrote a review using GenAI or not. Our findings reveal that using GenAI reduces reviewers' perceived cognitive load (H1). Additionally, our results confirm that using GenAI increases the informativeness of reviews. However, contrary to our expectations, we find that cognitive load does not mediate the relationship between using GenAI and the informativeness of reviews (H2). Finally, additional analyses of the reviews suggest that using GenAI affects reviews beyond informativeness. Specifically, reviews written using GenAI exhibit an increased linguistic complexity and tend to have a more positive sentiment. These findings contribute to both theory and practice and call for further investigation by future research.

References

Amos, C. & Zhang, L. (2024), 'Consumer reactions to perceived undisclosed ChatGPT usage in an online review context', *Telematics and Informatics* **93**, 102163.

Babić Rosario, A., De Valck, K. & Sotgiu, F. (2020), 'Conceptualizing the electronic word-of mouth process: What we know and need to know about eWOM creation, exposure, and evaluation', *Journal of the Academy of Marketing Science* **48**(3), 422–448.

Burtch, G., Hong, Y., Bapna, R. & Griskevicius, V. (2018), 'Stimulating Online Reviews by Combining Financial Incentives and Social Norms', *Management Science* **64**(5), 2065–2082.

Chen, P.-Y., Hong, Y. & Liu, Y. (2018), 'The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments', *Management Science* **64**(10), 4629–4647.

Cortina, J. M. (1993), 'What Is Coefficient Alpha? An Examination of Theory and Applications', *Journal of Applied Psychology* **78**(1), 98–104.

Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S. & Robert, L. P. (2024), Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models, *in* 'Proceedings of 2024 Conference on Human Factors in Computing Systems', ACM, pp. 1–18.

Erlebach, S. (2024), (When) Is a Summary Really Worth a Thousand Words? – How Textual and Visual Customer Review Summaries Affect Cognitive Load During Online Purchase Decisions, *in* 'Proceedings of the 19th International Conference on Wirtschaftsinformatik'.

Fink, L. (2022), 'Why and How Online Experiments Can Benefit Information Systems Research', *Journal of the Association for Information Systems* **23**(6), 1333–1346.

Ford, J. B. (2017), 'Amazon's Mechanical Turk: A Comment', *Journal of Advertising* **46**(1), 156–158.

Ghose, A. & Ipeirotis, P. G. (2011), 'Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics', *IEEE Transactions on Knowledge and Data Engineering* **23**(10), 1498–1512.

Gieshoff, A. C. & Heeb, A. H. (2023), 'Cognitive load and cognitive effort', *Translation, Cognition and Behaviour* **6**(1), 3–28.

Graham, S., Kiuhara, S. A., Harris, K. R. & Fishman, E. J. (2017), 'The Relationship among Strategic Writing Behavior, Writing Motivation, and Writing Performance with Young, Developing Writers', *The Elementary School Journal* **118**(1), 82–104.

Gutt, D., Neumann, J., Zimmermann, S., Kundisch, D. & Chen, J. (2019), 'Design of review systems – A strategic instrument to shape online reviewing behavior and economic outcomes', *The Journal of Strategic Information Systems* **28**(2), 104–117.

Habla, M., Napirata, S., Wrabel, A., Kupfer, A. & Zimmermann, S. (2024), Never Again "The Pizza was Great!" – Developing Design Principles for Dynamic Review Templates, *in* 'Proceedings of the 19th International Conference on Wirtschaftsinformatik'.

Hart, S. G. & Staveland, L. E. (1988), 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research', *Advances in Psychology* **52**,139–183.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2009), *Multivariate Data Analysis*. Prentice Hall.

Hayes A. F. (2022), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Press.

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z. & Trautsch, A. (2023), 'A large-scale comparison of human-written versus ChatGPT-generated essays', *Scientific Reports* **13**(1).

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J. & Jung, M. F. (2023), 'Artificial intelligence in communication impacts language and social relationships', *Scientific Reports* **13**(1), 5487.

Hutto, C. & Gilbert, E. (2014), ‘VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text’, *Proceedings of the International AAAI Conference on Web and Social Media* **8**(1), 216–225.

Jeong, E. & Jang, S. (2011), ‘Restaurant experiences triggering positive electronic word-of-mouth (eWOM) motivations’, *International Journal of Hospitality Management* **30**(2), 356–366.

Khern-am-nuai, W., Kannan, K. & Ghasemkhani, H. (2018), ‘Extrinsic versus Intrinsic Rewards for Contributing Reviews in an Online Platform’, *Information Systems Research* **29**(4), 871–892.

Knight, S., Bart, Y. & Yang, M. (2024), ‘Generative AI and User-Generated Content: Evidence from Online Reviews’, *SSRN Electronic Journal*.

Korfiatis, N., García-Bariocanal, E. & Sánchez-Alonso, S. (2012), ‘Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content’, *Electronic Commerce Research and Applications* **11**(3), 205–217.

Kovács, B. (2024), ‘The Turing test of online reviews: Can we tell the difference between human-written and GPT-4-written online reviews?’, *Marketing Letters* **35**(4), 651–666.

Lee, H.-H. & Jin Ma, Y. (2012), ‘Consumer perceptions of online consumer product and service reviews’, *Journal of Research in Interactive Marketing* **6**(2), 110–132.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. & Wilson, N. (2025), The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers, *in* ‘Proceedings of the ACM Conference on Human Factors in Computing Systems’, ACM.

Li, H., Wang, Y., Luo, S. & Huang, C. (2024), ‘The influence of GenAI on the effectiveness of argumentative writing in higher education: evidence from a quasi-experimental study in China’, *Journal of Asian Public Policy*, 1–26.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013), ‘Efficient Estimation of Word Representations in Vector Space’, arXiv:1301.3781.

Moro, S., Ramos, P., Esmerado, J. & Jalali, S. M. J. (2019), ‘Can we trace back hotel online reviews’ characteristics using gamification features?’, *International Journal of Information Management* **44**, 88–95.

Mudambi, S. M. & Schuff, D. (2010), ‘Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com’, *MIS Quarterly* **34**(1), 185–200.

Noy, S. & Zhang, W. (2023), ‘Experimental evidence on the productivity effects of generative artificial intelligence’, *Science* **381**(6654), 187–192.

OECD. (2024), *Do Adults Have the Skills They Need to Thrive in a Changing World?: Survey of Adult Skills 202*3, OECD Publishing.

Paas, F., van Gog, T. & Sweller, J. (2010), ‘Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives’, *Educational Psychology Review* **22**(2), 115–121.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z. & Damer, E. (2021), ‘Data quality of platforms and panels for online behavioral research’, *Behavior Research Methods* **54**(4), 1643–1662.

Poniatowski, M., Neumann, J., Goerzen, T. & Kundisch, D. (2019), ORGANIZING THEIR THOUGHTS – HOW ONLINE REVIEW TEMPLATES AFFECT THE REVIEW TEXT, *in* ‘Proceedings of the 27th European Conference on Information Systems’.

Putri, M. D., Rahmat, A. & Sanjaya, Y. (2022), ‘Chunking Technique with Writing is Thinking to Control Students' Extraneous Cognitive Load’, *Bioedukasi: Jurnal Pendidikan Biologi* **15**(1), 12–19.

Rahmat, N. H. (2023), ’Reducing Cogntive Overload in Online Academic Writing: A Case Study’, *International Journal of Academic Research in Business and Social Sciences* **13**(2), 1377–1391.

Reynolds, N. & Diamantopoulos, A. (1998), ‘The effect of pretest method on error detection rates: Experimental evidence’, *European Journal of Marketing* **32**(5/6), 480–498.

Rietsche, R., Frei, D., Stoeckli, E. & Söllner, M. (2019), NOT ALL REVIEWS ARE EQUAL- A LITERATURE REVIEW ON ONLINE REVIEW HELPFULNESS, *in* ‘Proceedings of the 27th European Conference on Information Systems’.

Rohde, C., Kupfer, A. & Zimmermann, S. (2022), ‘Explaining reviewing effort: Existing reviews as potential driver’, *Electronic Markets* **32**(5), 1169–1185.

Rooney, K. (2025), https://www.cnbc.com/2025/02/20/openai-tops-400-million-users-despite-deepseeks-emergence.html. Accessed: 01.03.2025.

Scholz, M. & Dorner, V. (2013), ‘The Recipe for the Perfect Review?’, *Business & Information Systems Engineering* **5**(3), 141–151.

Skulmowski, A. & Xu, K. M. (2022), ‘Understanding Cognitive Load in Digital and Online Learning: A New Perspective on Extraneous Cognitive Load’, *Educational Psychology Review* **34**(1), 171–196.

Son, J., Negahban, A. & Chiang, D. T. (2019), Topic Diversity of Online Consumer Reviews and Its Effect on Review Helpfulness, *in* ‘Proceeding of the 25th Americas Conference on Information Systems’.

Sweller, J., Ayres, P. & Kalyuga, S. (2011), *Cognitive Load Theory*. Springer.

Sweller, J., van Merrienboer, J. J. G. & Paas, F. G. W. C. (1998), ‘Cognitive Architecture and Instructional Design’, *Educational Psychology Review* **10**(3), 251–296.

Sweller, J., van Merrienboer, J. J. G. & Paas, F. G. W. C. (2019), ‘Cognitive Architecture and Instructional Design: 20 Years Later’, *Educational Psychology Review* **31**(2), 261–292.

Turner, E. & Rainie, L. (2020), https://www.pewresearch.org/short-reads/2020/03/05/most-americans-rely-on-their-own-research-to-make-big-decisions. Accessed: 01.02.2025.

Yang, H., Zhang, C. & Li, K. (2023), PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis, *in* ‘Proceedings of the 32nd ACM International Conference on Information and Knowledge Management’, pp. 5117–5122.

Yin, D., Bond, S. D. & Zhang, H. (2014), ‘Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews’, *MIS Quarterly* **38**(2), 539–560.

Zemke, H., Stahmann, P. & Janiesch, C. (2025), Leveraging Generative Artificial Intelligence for Design Thinking in Creative Processes: A Literature Review, *in* ‘Proceedings of the 58th Hawaii International Conference on System Sciences`, pp. 142–151.