

Online Appendices for “Navigating the Sea of Reviews: Unveiling the Effects of Introducing AI-Generated Summaries in E-Commerce”

Online Appendix A. Summary Statistics of the US-UK Matched-Product Dataset

Table A.1. Summary Statistics of US-UK Matched-Product Dataset

Variable	Definition	Mean	SD	Min	Max
RevVolume	the number of reviews for a product in a week	2.05	4.64	0	72
AISummary	whether a product has AIGS (1: Yes; 0: No) in a week	0.23	0.42	0	1
Rank	the sales rank of a product at the end of a week	23,917.60	47,973.89	2	800,706
Price	the price of a product, averaged through a week	29.60	32.32	1.28	256.00
CumReviews	the cumulative number of reviews for a product up to a week	1,402.12	2,605.07	17	23,656
CumAvgRating	the cumulative average rating for a product up to a week	4.33	0.34	1.50	5.00
ProdTenure	the number of months since a product’s release on Amazon up to a week	71.95	46.10	0	264
AvgNumTopic	the average number of distinct semantic topics discussed in reviews for a product in a week	3.74	2.00	0	18
AvgAIGSAttr	the average number of AIGS attribute keywords or their cognitive synonyms mentioned in reviews for a product in a week	0.59	0.66	0.00	5.00
AvgNonAIGSAttr	the average number of non-AIGS product attributes mentioned in consumer reviews for a product in a week	1.70	1.34	0	15
AvgAIGSAttrLen	the average character count of sentences mentioning AIGS product attributes in consumer reviews for a product in a week	78.12	48.13	10	465
AvgAIGSSim	the average semantic similarity between review sentences mentioning AIGS product attributes and the corresponding AIGS text discussing the same product attributes for a product in a week	0.42	0.13	0.04	0.83
AIGSMentRev	the number of reviews that mention AIGS attribute keywords for a product in a week	0.88	2.38	0	46

NoAIGSMentRev	the number of reviews that do not mention any AIGS attribute keywords for a product in a week	1.17	2.65	0	29
NumSameSen	the number of same-sentiment occurrences in reviews of a product in a week	0.70	2.16	0	38
NumDiffSen	the number of different-sentiment occurrences in reviews of a product in a week	0.34	1.13	0	19

Notes: SD is for standard deviation; Min is for minimum; Max is for maximum.

Online Appendix B. Unveiling the Black Box of AIGS

Impact of Review-level Characteristics

To examine whether AIGS disproportionately reflects reviews with certain characteristics, we begin by identifying which reviews are reflected by the AIGS. We utilize the product attribute keywords listed below the AIGS text, which correspond to the product attributes mentioned within the summary text. We define a review as “included” in the AIGS if any of these keywords or their semantic synonyms appear in it. To identify synonyms, we employ the Natural Language Toolkit (NLTK) in Python (Bird et al. 2009), retrieving the top 5 synonyms of each keyword from NLTK WordNet. (Results are robust when we use top 10 and top 20 synonyms.)

Impact of Vine reviews. We first examine whether AIGS disproportionately reflects Vine reviews. For review r of product i , we construct a dummy variable, $Included_{ri}$, which equals 1 if any of the AIGS keywords appear in the review, and 0 otherwise. We restrict our analysis to pretreatment reviews in the treatment group. We use a fixed effects generalized linear model (GLM) with a logit link function, given that the dependent variable is a dummy variable. Specifically, we perform the following equation:

$$\Pr(Included_{ri} = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 Vine_r + \beta_2 Negative_r + \beta_3 DayToTreat_r + X_r + \mu_i + \varepsilon_{ri}) \quad (\text{B.1})$$

In this model, $Included_{ri}$ represents the dependent variable (whether review r is included in the AIGS of product i), and $Vine_r$ is the independent variable (a dummy variable indicating whether review r is an Amazon Vine review). logit^{-1} is the inverse of a logit function. We add a series of review-level control variables. Specifically, we control for review valence, denoted by $Negative_r$, which is operationalized as an indicator equal to 1 if it has a one-star or two-star rating, and 0 otherwise (Mudambi and Schuff 2010). We also control for review recency, denoted by $DayToTreat_r$, which represents the number of days between the review date and the treatment implementation date. We further add X_r to control for the rating of the review ($Rating_r$), the length (i.e., character count) of the review ($Length_r$), and the number of helpful votes the review received ($NumHelpful_r$). In addition, we incorporate a series of linguistic variables derived from the Linguistic Inquiry and Word Count (LIWC) tool (Pennebaker et al. 2022), which have been documented in prior literature as critical characteristics of reviews. To control for

emotionality, we include the percentage of positive emotion words (emo_pos_r) and negative emotion words (emo_neg_r ; Oh et al. 2023), as well as more specific emotions such as anger (emo_anger_r) and anxiety (emo_anx_r ; Yin et al. 2014). We also account for pronoun usage associated with self- and other-focus (Lei et al. 2021), including the percentage of words that are first-person singular pronouns (i_r), first-person plural pronouns (we_r), and second-person pronouns (you_r). We further control for cognitive and affective processing by including the percentage of cognitive words ($Cognition_r$) and affective words ($Affect_r$; Huang et al. 2017). We further include product fixed effects, μ_i , to control for any inherent differences in product attributes. The results, presented in Table B.1, show that the coefficient of $Vine_r$ is statistically insignificant, suggesting that AIGS does not systematically favor Vine reviews in its summary generation.

Table B.1. Impact of Review Characteristics on Inclusion in AIGS

Variables	Included
Vine	0.044 (0.112)
Negative	0.031 (0.040)
DayToTreat	-0.0006 (0.0007)
Length	0.004*** (0.0002)
NumHelpful	0.0004 (0.004)
Rating	0.070 (0.047)
emo_pos	-0.005 (0.003)
emo_neg	0.010 (0.007)
emo_anger	-0.015 (0.023)
emo_anx	0.005 (0.039)
i	-0.008 (0.005)
we	0.005 (0.025)
you	-0.010 (0.008)
Cognition	0.001 (0.002)
Affect	-0.010*** (0.002)
Product FE	Yes
Observations	29,178

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effects.

Moreover, to rule out the possibility that our main findings are driven by Vine reviews, we re-estimate our main regression model after excluding all Vine reviews. The estimation results from this analysis are displayed in Table B.2. These results are substantially consistent with our main results, suggesting that our main findings are robust when the potential impact from Vine reviews is removed.

Table B.2. Impact of AIGS on Consumer Review Volume Excluding Vine Reviews

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)
AISummary	0.060*** (0.008)	0.050*** (0.008)
Control variables	No	Yes
Product FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	69,322	69,322

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effects.

Impact of Review Valence. We also pay particular attention to whether the AIGS disproportionately reflects reviews of a particular valence. As shown in Table B.1, the coefficient of $Negative_r$ is statistically insignificant. The results suggest that AIGS does not systematically prioritize negative reviews over positive ones, which supports Amazon’s claim that the AI-generated summaries reflect frequently mentioned content rather than sentiment bias.

Impact of Review Recency. We are also interested in whether AIGS is generated based on all previous reviews or primarily on recent ones. As presented in Table B.1, the coefficient of $DayToTreat_r$ is insignificant, suggesting that review recency does not systematically affect inclusion in AIGS. In other words, AIGS is likely generated based on all previous reviews rather than being biased toward more recent ones.

Discussion. As shown in Table B.1, with the exception of review length ($Length_r$) and the percentage of affective words ($Affect_r$), other coefficients are not significantly associated with the likelihood of a review being included in the AIGS. These results provide some insight into the types of review content that are more likely to be reflected in the AIGS. Specifically, the coefficient of $Length_r$ is significantly positive, indicating that longer reviews are more likely to be incorporated into the AIGS. This is consistent with the possibility that longer reviews contain a larger amount of product-relevant information, which may make them more likely to be extracted and aggregated in the summarization process. The coefficients of $Affect_r$ are significantly negative, indicating that reviews containing a higher percentage of affective words are less likely to be incorporated into the AIGS. This might be due to the possibility that reviews with a higher proportion of affective language may place greater emphasis on subjective or emotion-focused expressions and convey relatively less product-relevant substantive information, which may reduce the likelihood that they are incorporated into the AIGS.

Impact of Number of Prior Reviews

To examine how the number of prior reviews included in the AIGS influences the impact of AIGS on subsequent consumer contributions, we introduce a moderating variable: the number of reviews written prior to the treatment that are included in the AIGS. Specifically, we perform the following regression model:

$$\log(\text{RevVolume}_{it} + 1) = \beta_0 + \beta_1 \text{AISummary}_{it} + \beta_2 \text{AISummary}_{it} \times \text{NumPriorRev}_i + \text{Controls} + \mu_i + \delta_t + \varepsilon_{it} \quad (\text{B.2})$$

where NumPriorRev_i denotes the number of reviews of product i written prior to the treatment that are included in the AIGS. Columns (1) and (2) in Table B.3 present the results of the regression analysis. The coefficients of the interaction between AISummary_{it} and NumPriorRev_i are consistently positive and statistically significant. This suggests that AIGS has a stronger effect when it reflects a greater number of prior reviews. One possible explanation is that a larger number of included reviews allows AIGS to more comprehensively capture existing consumer opinions. This breadth signals to potential contributors that AIGS effectively synthesizes diverse perspectives. As a result, contributors may feel more motivated to add their own perspectives, perceiving their reviews as meaningful additions to an informative and visible body of collective knowledge.

Table B.3. Moderating Roles of Numbers of Prior Reviews Included and Attribute Keywords

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)	(3) log(RevVolume + 1)	(4) log(RevVolume + 1)
AISummary	0.072*** (0.009)	0.061*** (0.009)	0.071*** (0.009)	0.061*** (0.009)
AISummary × NumPriorRev	0.006*** (0.001)	0.015*** (0.002)		
AISummary × NumAIGSAttr			0.017*** (0.003)	0.015*** (0.003)
Control variables	No	Yes	No	Yes
Product FE	Yes	Yes	Yes	Yes
Weekly FE	Yes	Yes	Yes	Yes
Observations	69,322	69,322	69,322	69,322

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effects.

Impact of Number of Attribute Keywords

To examine how the number of attribute keywords tagged by AIGS affects its impact on subsequent consumer review contributions, we introduce this number as a moderating variable in the following regression model:

$$\log(\text{RevVolume}_{it} + 1) = \beta_0 + \beta_1 \text{AISummary}_{it} + \beta_2 \text{AISummary}_{it} \times \text{NumAIGSAttr}_i + \text{Controls} + \mu_i + \delta_t + \varepsilon_{it} \quad (\text{B.3})$$

where $NumAIGSAtr_i$ denotes the number of attribute keywords tagged by AIGS for product i . As shown in Columns (3) and (4) in Table B.3, the coefficients of the interaction between $AISummary_{it}$ and $NumAIGSAtr_i$ are consistently positive and statistically significant. This suggests that the positive impact of AIGS on subsequent review volume is more pronounced when it tags a greater number of product attributes. One possible explanation is that a greater number of attribute keywords indicates that AIGS more effectively captures diverse aspects of existing consumer reviews. As a result, contributors may view the existing review content as collectively more informative and valuable, thereby being more motivated to add their own perspectives to this collective knowledge base.

AIGS Updating Progress

Clarifications regarding the updating process of AIGS:

- Although Amazon does not officially disclose its AIGS update policy, our empirical data indicate that AIGS typically updates within a week if there are new reviews. This suggests that AIGS updates in a timely manner to incorporate recent review content.
- While consumers are not explicitly notified of AIGS updates, the promptness of AIGS updates increases the chance that returning consumers notice the changes in the AIGS.
- We consistently use the updated AIGS content in our analyses to accurately capture its impact on subsequent reviews.

Online Appendix C. Callaway and Sant’Anna Staggered DID

Since the AIGS feature has been gradually assigned to products (some units received the treatment earlier, while others received it later), there may be concerns about the negative weights in DID estimates arising from varying treatment times, as highlighted in recent econometrics literature (Goodman-Bacon 2021). This is because the estimated coefficient reflects a composite average from three different pairs of comparisons: (1) units treated earlier as the treatment while units treated later as the control; (2) units treated later as the treatment while units treated earlier as the control; and (3) all treated units as the treatment while units never treated as the control. As demonstrated in previous literature (Callaway and Sant’Anna 2021; De Chaisemartin and d’Haultfoeuille 2020; Goodman-Bacon 2021), the second pair of comparisons can introduce bias into the estimates because the units treated earlier are actually treated during both the earlier and later periods.

The staggered DID model by Callaway and Sant’Anna (2021) calculates the average treatment effects for each product while accounting for variations in treatment timing. By grouping products according to when they first have AIGS, the model estimates treatment effects that explicitly account for the timing of AIGS introduction. Critically, by comparing treated products to those not yet exposed to AIGS at each time period, this approach avoids biased comparisons where earlier-treated units are incorrectly

used as the control group for later-treated units. Following prior literature (Li and Wang 2025; Liu et al. 2024), we obtain the results using the staggered DID model by Callaway and Sant’Anna (2021). As shown in Table C.1, the coefficients of $AI\text{Summary}_{it}$ remain significantly positive. This aligns with our main analysis, confirming the robustness of our findings.

Table C.1. Impact of AIGS on Consumer Review Volume using CS method

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)
AISummary	0.039** (0.015)	0.052** (0.019)
Control variables	No	Yes
Observations	69,322	69,322

Notes: We perform the estimation using the bootstrap method ($n = 5,000$) in R with the did package; Robust standard errors clustered at the product level are in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix D. Additional Robustness Checks

Random (Shuffled) Implementation Test

We conduct a random (shuffled) implementation test (Cheng et al. 2020). We begin by randomly shuffling the original independent variable (e.g., AISummary) to generate a “pseudo” treatment. We then run the main DID model using the “pseudo” treatment as the independent variable and record the resulting coefficient. This process is repeated 1,000 times. As expected, the distributions of the pseudo treatment effects, shown in Figure D.1, are centered around zero. Table D.1 shows that the observed effect is significantly different from the pseudo effect, indicating that the likelihood of obtaining a similar effect purely by chance is extremely low ($p < 0.001$).

Table D.1. Random (Shuffled) Implementation Test

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)
μ of Random β	-0.00018	-0.00017
σ of Random β	0.00011	0.00011
Replications	1000	1000
Estimated β	0.071	0.060
Control variables	No	Yes
Product FE	Yes	Yes
Weekly FE	Yes	Yes
Z-Score	648.100	603.873
P-Value	$p < 0.001$	$p < 0.001$

Notes: μ of Random β represents the mean of the pseudo effects, and σ of Random β denotes their standard error. The test calculates the probability of obtaining an effect of a similar magnitude to the observed effect purely by chance, which is extremely low ($p < 0.001$).

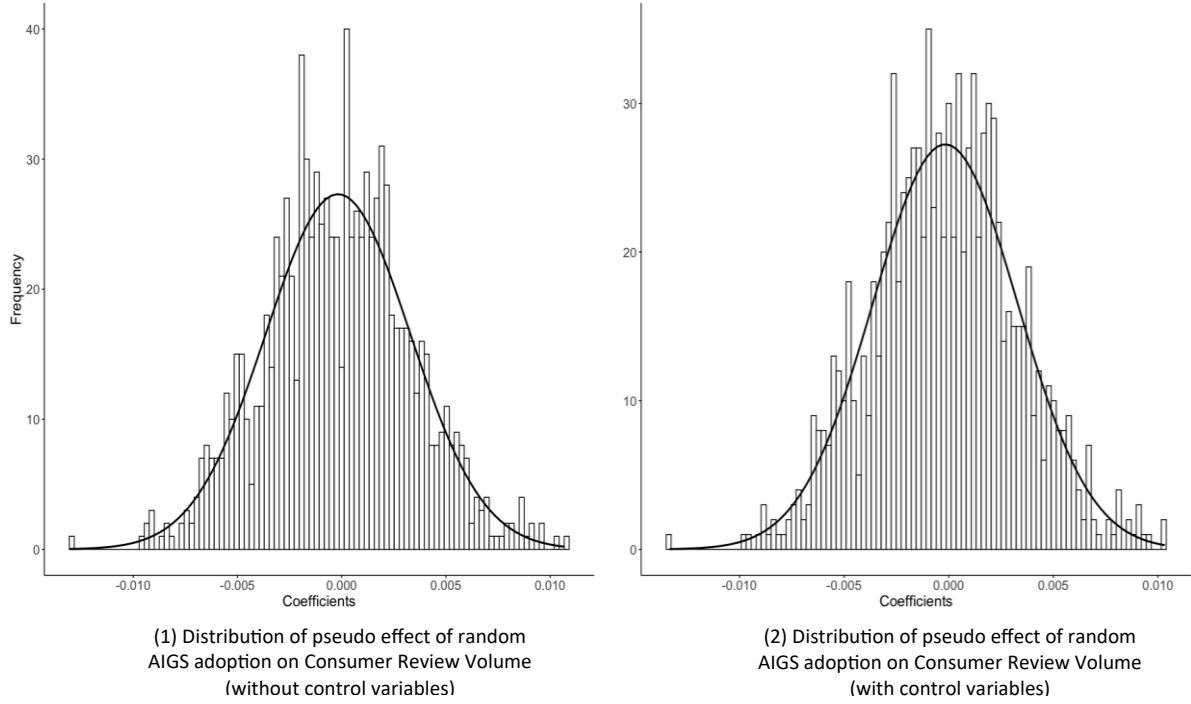


Figure D.1: Distribution of Coefficients based on Random Treatments

Pre-treatment Placebo Test

We also conduct pre-treatment placebo test by employing the FECT package in R. As shown in Figure D.2, the placebo equivalence p-value is less than 0.01, indicating that our results remain robust to the pre-treatment placebo test (Liu et al. 2024). This provides supporting evidence for the parallel trends assumption and helps rule out the possibility that our results are driven by pre-treatment effects.

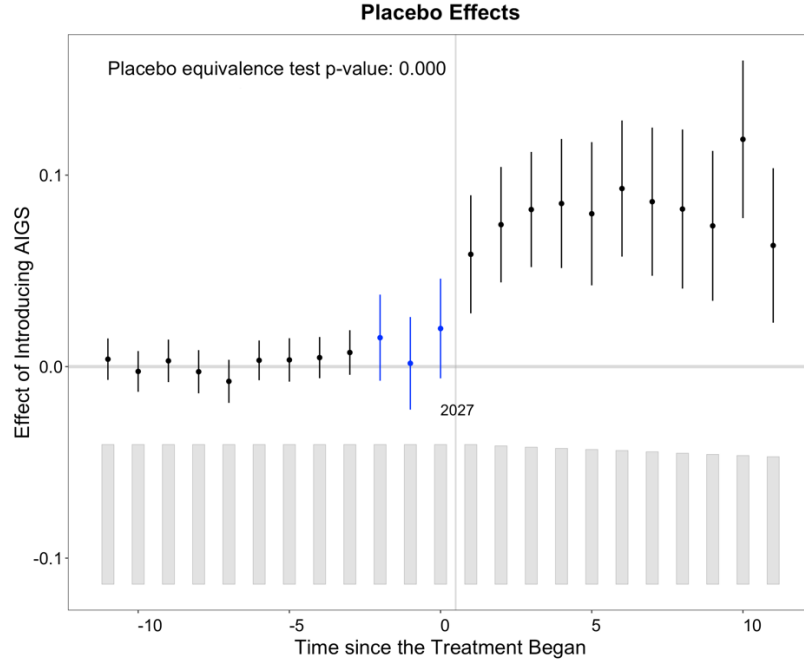


Figure D.2: Pre-treatment Placebo Test

Hazard Rate Test

We perform a hazard rate test to examine whether products with higher prior review activity are more likely to receive AIGS earlier. The hazard rate test predicts the weekly hazard of AIGS adoption as a function of products' recent review volume and other product characteristics. The survival time is calculated as the number of weeks since the introduction of AIGS on Amazon, and the event is defined as the first week in which a product receives AIGS. We include the log-transformed review volume in each of the five preceding weeks, denoted by $\log(\text{RevVolume}_{i,t-k} + 1)$, where $k = 1, \dots, 5$. If products with higher recent review activity are systematically more likely to receive AIGS earlier, one or more of these measures of lagged review volume should significantly predict the hazard of AIGS adoption.

The results are reported in Table D.2. The coefficients of the lagged measures of log-transformed review volume capture the possibility of reverse causality. As shown, these coefficients are statistically insignificant, providing no evidence that products with higher recent review activity are more likely to receive AIGS earlier. The results remain consistent when we include the log-transformed review volume in each of the three or seven preceding weeks. Taken together, these findings effectively rule out the possibility of reverse causality.

Table D.2. Cox Proportional Hazard Model

Variables	Efron (1)	Breslow (2)	Exact (3)
$\log(\text{RevVolume}(t - 1) + 1)$	0.053 (0.042)	0.044 (0.051)	0.055 (0.053)

log(RevVolume($t - 2$) + 1)	0.021 (0.047)	0.017 (0.052)	0.020 (0.051)
log(RevVolume($t - 3$) + 1)	0.066 (0.050)	0.042 (0.049)	0.065 (0.052)
log(RevVolume($t - 4$) + 1)	-0.043 (0.049)	-0.032 (0.055)	-0.044 (0.049)
log(RevVolume($t - 5$) + 1)	0.017 (0.052)	-0.011 (0.052)	0.021 (0.059)
log(Price)	0.159*** (0.024)	0.111*** (0.024)	0.122*** (0.023)
log(Rank)	-0.012 (0.020)	-0.010 (0.019)	-0.012 (0.019)
log(CumReviews + 1)	0.019 (0.057)	0.020 (0.059)	0.027 (0.053)
CumAvgRating	-0.055 (0.059)	-0.042 (0.044)	-0.036 (0.050)
ProdTenure	0.004*** (0.001)	0.003*** (0.001)	0.004*** (0.001)
Observations	13,867	13,867	13,867
No. of events	2,027	2,027	2,027

Notes. This table reports the estimated coefficients and robust standard errors (in parentheses) in the Cox proportional hazards model. Columns (1), (2), and (3) use the Efron method, Breslow method, and the exact partial-likelihood method to handle tied failures, respectively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Pre-trend Test

To further address the concern about the presence of a pretreatment trend, we follow Hartman and Hidalgo (2018) and conduct an equivalence test using two one-sided t -tests (TOST). The test examines whether the average prediction errors between the observed outcomes of treated products and their estimated counterfactual outcomes during the pretreatment periods are sufficiently close to zero. The test is considered passed if the average pretreatment prediction errors and their 90% confidence intervals fall within the prespecified equivalence bounds (Liu et al. 2024).

As shown in Figure D.3, the average pretreatment prediction errors and their 90% confidence intervals fall within these bounds. Accordingly, we reject the null hypothesis of inequivalence ($p < .001$). These results indicate that the pretreatment differences between the observed and counterfactual outcomes are sufficiently small to be considered equivalent to zero.

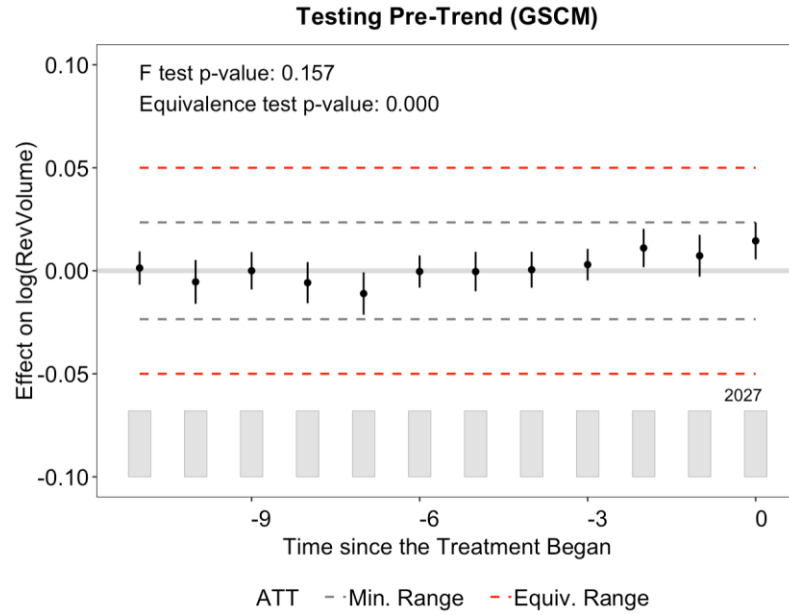


Figure D.3: Testing for no Pretreatment Trend

Using Untransformed Review Volume as the Dependent Variable

As highlighted by Chen and Roth (2024) in their work, the use of log transformations presents misspecifications when the variable of interest, Y , includes zero values. Since traditional $\log(Y)$ is undefined at $Y=0$, researchers commonly adopt “log-like” transformations such as $\log(Y+1)$. The critical issue of these log-like transformations is their “unit-dependence”: the Average Treatment Effect (ATE) estimates derived from the transformations are dependent on the units of the outcome (Y) and thus do not reliably approximate percentage effects, which is particularly pronounced when the treatment changes the outcome (Y) from zero to a non-zero value.

In Section 4.1, we follow prior research (Chen and Roth 2024) and use Poisson quasi-maximum likelihood estimation (QMLE) as an alternative approach to address the issue of log-like transformations. In addition, we directly use *RevVolume* without log-transformation as the dependent variable and re-estimate our main regression model. As shown in Table D.3, the coefficients of $AISummary_{it}$ are significantly positive. This analysis reinforces the robustness of our main findings by demonstrating results consistent with our main analysis, mitigating concerns related to log transformations.

Table D.3. Impact of AIGS on Consumer Review Volume without Log Transformations

Variables	(1) RevVolume	(2) RevVolume
AISummary	0.390*** (0.053)	0.334*** (0.053)
Control variables	No	Yes
Product FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	69,322	69,322

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effect.

Excluding Top 1% of Products Ranked by Pre-policy Review Volume

To ensure our findings are not driven by a small number of highly active products, we conduct an additional robustness check that excludes the top 1% of products ranked by their pre-policy review volume. Specifically, we calculate the 99th percentile of pre-policy review volume and re-estimate our main regression model after excluding products whose pre-policy review volume exceeds this threshold. The findings of the main analysis are presented in Table D.4. Columns (1) and (2) respectively display the results of the regressions without and with the inclusion of control variables. The coefficients of $AI\text{Summary}_{it}$ are consistently positive across the analysis, consistent with the main results. These findings indicate that our results are not driven by a small number of products with exceptionally high pre-policy review activity.

Table D.4. Impact of AIGS on Consumer Review Volume Excluding Top 1% of Products Ranked by Pre-policy Review Volume

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)
AISummary	0.068*** (0.009)	0.057*** (0.009)
Control variables	No	Yes
Product FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	68,494	68,494

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effects.

Excluding Products Unavailable in Any Week During the Observation Window

We further consider product availability during the observation window and find that only 39 products (approximately 1.3% of the sample) were unavailable on Amazon in one or more weeks, while all remaining products were available throughout the entire observation window. We re-estimate our main regression model after excluding these products. The findings of the analysis are presented in Table D.5. Columns (1) and (2) respectively display the results of the regressions without and with the inclusion of control variables. The coefficients of $AI\text{Summary}_{it}$ are consistently positive across the analysis, consistent with the main results. The results suggest that the estimated treatment effects are not driven by differential product availability following treatment.

Table D.5. Impact of AIGS on Consumer Review Volume Excluding Products Unavailable in Any Week During the Observation Window

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)
AISummary	0.072*** (0.009)	0.061*** (0.009)
Control variables	No	Yes

Product FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	68,425	68,425

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effects.

Online Appendix E. Alternative Measures of the Moderators

We conduct additional analyses using two product-level moderators constructed from pre-policy information: the cumulative number of reviews and cumulative variance of ratings. First, we define the moderating factor *WellReviewed*. Based on their cumulative number of reviews prior to the policy, we categorize a product as either “well-reviewed” if the number is above the median, or “under-reviewed” otherwise. We then incorporate *WellReviewed_i*, a dummy variable denoting whether product *i* is well-reviewed or under-reviewed (1 = well-reviewed, 0 = under-reviewed), as the moderating factor and re-estimate Equation (4). Columns (1) and (2) in Table E.1 present the main results of the regression analysis. The coefficients of the interaction between *AISummary_{it}* and *WellReviewed_i* are significant. The results indicate that introducing AIGS has a stronger positive effect on products with higher cumulative review volumes compared to those with lower volumes.

Table E.1. Moderating Roles of Product’s Cumulative Review Volume and Rating Variance

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)	(3) log(RevVolume + 1)	(4) log(RevVolume + 1)
AISummary	0.020** (0.008)	0.017** (0.008)	0.048*** (0.010)	0.036*** (0.010)
AISummary × WellReviewed	0.085*** (0.011)	0.083*** (0.018)		
AISummary × Polarizing			0.044*** (0.012)	0.046*** (0.012)
Control variables	No	Yes	No	Yes
Product FE	Yes	Yes	Yes	Yes
Weekly FE	Yes	Yes	Yes	Yes
Observations	69,322	69,322	69,322	69,322

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effect.

We then examine whether the impact of introducing AIGS varies with different levels of cumulative rating dispersion for products. Following previous research (Li 2018), we define products as either “polarizing” when their variance of the ratings exceeds the median value, or “non-polarizing” otherwise. We then incorporate the rating dispersion dummy, *Polarizing_i*, which denotes whether a product is polarizing or non-polarizing (1 = polarizing, 0 = non-polarizing), as the moderating factor and re-estimate Equation (5). Columns (3) and (4) in Table E.1 display the main results of the regression analysis. The coefficients of the interaction between *AISummary_{it}* and *Polarizing_i* are consistently positive, indicating a more pronounced positive effect of introducing AIGS for products with higher rating dispersion

compared to those with lower rating dispersion. These results align with the analyses using the dummy moderators updated weekly, further supporting the robustness of our findings.

Moderation Analysis Using Continuous Specifications

To complement the analyses based on discretized measures of the moderating factors, we conduct additional robustness checks using their continuous specifications directly. First, we incorporate $CumReviews_i$, the cumulative number of reviews for product i before the policy, as the moderating factor. We mean-center the log-transformed variable before incorporating it into the regression to facilitate the interpretation of coefficients. We re-estimate Equation (4) using this continuous measure of the moderating factor. Columns (1) and (2) in Table E.2 present the results of the regression analysis. The coefficients of the interaction between $AI Summary_{it}$ and $CumReviews_i$ are significantly positive. Consistent with the results using the discretized measure, the results indicate that introducing AIGS has a stronger positive effect on products with higher cumulative review volumes compared to those with lower volumes.

Table E.2. Moderating Roles of Product's Cumulative Review Volume and Rating Variance (Continuous Measures)

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)	(3) log(RevVolume + 1)	(4) log(RevVolume + 1)
AI Summary	0.050*** (0.008)	0.043*** (0.008)	0.071*** (0.009)	0.060*** (0.009)
AI Summary × CumReviews	0.046*** (0.005)	0.044*** (0.005)		
AI Summary × RatingVariance			0.019** (0.010)	0.020** (0.010)
Control variables	No	Yes	No	Yes
Product FE	Yes	Yes	Yes	Yes
Weekly FE	Yes	Yes	Yes	Yes
Observations	69,322	69,322	69,322	69,322

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effect.

We then incorporate $RatingVariance_i$, the rating variance of reviews for product i before the policy, as the moderating factor and re-estimate Equation (5). We mean-center the variable before incorporating it into the regression. Columns (3) and (4) in Table E.2 present the results of the regression analysis. The coefficients of the interaction between $AI Summary_{it}$ and $RatingVariance_i$ are consistently positive, indicating a more pronounced positive effect of introducing AIGS for products with higher rating dispersion compared to those with lower rating dispersion. These results align with the analyses using the discretized measures, further supporting the robustness of our findings.

Online Appendix F. Additional Analyses on Alternative Explanations

Additional Analyses on Complementary Contributions

In Section 4.4.3, we document that contributors make complementary contributions after the introduction of AIGS. We further examine whether the increase in review volume occurs among reviews that mention AIGS attributes, reviews that do not, or both. Using the US-UK matched-product dataset, we construct two dependent variables that count the number of reviews that mention at least one AIGS attribute keyword (AIGSMentRev) and the number that do not (NoAIGSMentRev) for product i in week t at platform j , and re-estimate Equation (3) for each. As shown in Table F.1, the coefficients of AISummary are significantly positive for both, indicating that the numbers of reviews that mention and that do not mention AIGS attributes both increase after the introduction of AIGS. This shows that the rise in review volume is not limited to reviews that introduce new information, a finding consistent with our framework in which enhanced collective efficacy, our primary mechanism, raises overall contribution while complementary contribution shapes the content that contributors add.

Table F.1. Impact of AIGS on Reviews That Mention and Do Not Mention AIGS Attributes

Variables	(1) log(AIGSMentRev + 1)	(2) log(AIGSMentRev + 1)	(3) log(NoAIGSMentRev + 1)	(4) log(NoAIGSMentRev + 1)
AISummary	0.085*** (0.027)	0.078*** (0.027)	0.177*** (0.039)	0.169*** (0.039)
Control variables	No	Yes	No	Yes
Product–platform FE	Yes	Yes	Yes	Yes
Weekly FE	Yes	Yes	Yes	Yes
Observations	2,432	2,432	2,432	2,432

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product-platform level are in parentheses; FE stands for fixed effect.

Under a broader conceptualization of complementarity, reviews mentioning AIGS attributes may also be deemed to complement the AIGS content by either deepening discussions or offering alternative perspectives. We therefore conduct an additional analysis to further distinguish between reviews that express the same sentiment as the AIGS on the mentioned attributes and those that express a different sentiment.

Specifically, we leverage the sentiment tags assigned to AIGS attribute keywords. Each attribute keyword in the AIGS is associated with a sentiment tag that reflects the prevailing sentiment expressed toward that attribute in the set of existing reviews (Schermerhorn 2023). When the prevailing sentiment toward an attribute is positive, the keyword is tagged as “positive”; when the prevailing sentiment is negative, it is tagged as “negative.” For each attribute mentioned in a review, we apply a RoBERTa-based sentiment classifier fine-tuned on Amazon product reviews to determine whether the sentence(s) discussing the attribute express a positive or negative sentiment (AnkitAI 2025; Liu et al. 2019). As such, for each attribute, we obtain (i) the sentiment tag assigned to the attribute in the AIGS, and (ii) the sentiment expressed toward the attribute in the review. We then compare the two sentiments.

We define a same-sentiment occurrence when the sentiment expressed in the review is the same as the sentiment tag for that attribute in the AIGS. We define a different-sentiment occurrence when the sentiment expressed in the review is different from the sentiment tag in the AIGS. We then count the number of same-sentiment and different-sentiment occurrences. At the product–week–platform level, we aggregate these counts and construct two dependent variables: $NumSameSen_{ijt}$ and $NumDiffSen_{ijt}$, which capture the number of same-sentiment and different-sentiment occurrences in reviews of product i in week t at platform j , respectively. Accordingly, we perform the following regressions:

$$\left(\frac{\log(NumSameSen_{ijt} + 1)}{\log(NumDiffSen_{ijt} + 1)} \right) = \beta_0 + \beta_1 AISummary_{ijt} + Controls + \mu_{ij} + \delta_t + \varepsilon_{ijt} \quad (F.1)$$

This analysis enables us to examine more precisely whether contributors primarily reinforce or complement AIGS content. If contributors only complement the AIGS content, we would expect a significant increase in $NumDiffSen$ but not in $NumSameSen$. However, as shown in Table F.2, the coefficients of $AISummary_{ijt}$ are positive for both dependent variables. The results indicate that, following the introduction of AIGS, expressions with a different sentiment from the AIGS and those with the same sentiment as the AIGS both increase. The increase in different-sentiment occurrences shows that contributors complement the AIGS content, while the increase in same-sentiment occurrences shows that contributors also express sentiments consistent with the AIGS content. This is consistent with our framework, in which enhanced collective efficacy, our primary mechanism, raises overall contribution while complementary contribution shapes the content that contributors add (Section 4.4.3). In other words, by enhancing contributors’ beliefs that their individual reviews can make a meaningful contribution, AIGS encourages contributions that either reinforce or supplement the AIGS content.

Table F.2. Impact of AIGS on Same-Sentiment and Different-Sentiment Occurrences

Variables	(1) log(NumSameSen + 1)	(2) log(NumSameSen + 1)	(3) log(NumDiffSen + 1)	(4) log(NumDiffSen + 1)
AISummary	0.074*** (0.029)	0.074*** (0.029)	0.039*** (0.013)	0.038*** (0.013)
Control variables	No	Yes	No	Yes
Product–platform FE	Yes	Yes	Yes	Yes
Weekly FE	Yes	Yes	Yes	Yes
Observations	4,646	4,646	4,646	4,646

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product-platform level are in parentheses; FE stands for fixed effect.

Moreover, we conduct an additional analysis using the experimental data to further examine complementary contributions. Specifically, we utilize the AIGS attribute keywords and the textual comments collected from participants in our laboratory experiment. We manually code reviews based on whether they mention AIGS attribute keywords. We then conduct binary logistic regression analyses with experimental condition as the independent variable to examine (1) the likelihood of writing a review that

mentions AIGS keywords and (2) the likelihood of writing a review that does not mention AIGS keywords. If contributors only complement the AIGS content, we would expect an increase in reviews that do not mention AIGS keywords but not in those that mention AIGS keywords.

The results show that participants in the treatment condition (56.25%, 45/80) are significantly more likely than those in the control condition (40.00%, 32/80) to write a review that mentions AIGS keywords ($\chi^2(1, 158) = 4.23$, odds ratio = 1.93, $p = .040$). Meanwhile, participants in the treatment condition (7.50%, 6/80) are also more likely than those in the control condition (1.25%, 1/80) to write a review that does not mention AIGS keywords ($\chi^2(1, 158) = 3.74$, odds ratio = 6.41, $p = .053$). These findings indicate that, with AIGS, the likelihood of writing reviews that both mention and do not mention AIGS keywords increases. The increase in reviews that do not mention AIGS keywords reflects complementary contributions, and the increase in reviews that mention AIGS keywords shows that contributors also continue to discuss AIGS product attributes. Consistent with our framework, collective efficacy, our primary mechanism, drives the overall rise in contribution, while complementary contribution shapes the content that contributors add. The additional experimental analysis, which adopts the broader conceptualization of complementarity described above, further supports the robustness of our findings.

Reduced Redundancy Concerns

Applying similar reasoning, we further examine whether reduced concerns about information redundancy underlie the effect of AIGS. If AIGS increases review volume primarily by reducing concerns about information redundancy, we would expect an increase in reviews that supplement the AIGS content but not in those that reinforce the AIGS content. However, our results consistently indicate that AIGS encourages contributions that either reinforce or supplement the AIGS content. This evidence suggests that reduced concerns about information redundancy alone cannot fully explain the observed effect.

Online Appendix G. Effects on Rating and Sales

The Effect on Rating

While our primary analysis examines how AIGS influences subsequent review contributions, changes in review ratings provide additional insights into the effects of AIGS. We conduct an additional analysis to examine how the introduction of AIGS affects the rating of subsequent consumer reviews. Specifically, we use $AvgRating_{it}$, the average rating of reviews for product i in week t , as the dependent variable and re-estimate our main regression model. As shown in Table G.1, the significantly positive coefficients of $AISummary_{it}$ suggest that the average rating of subsequent reviews increases following the introduction of AIGS. Given that higher review ratings are generally associated with greater consumer satisfaction (Yin et al. 2017), these results suggest that consumer experiences improve following the introduction of AIGS.

Table G.1. Impact of AIGS on Consumer Ratings

Variables	(1) AvgRating	(2) AvgRating
AISummary	0.124*** (0.038)	0.111*** (0.038)
Control variables	No	Yes
Product FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	28,789	28,789

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effect.

This additional analysis complements our main findings. It suggests that AIGS enhances the helpfulness of reviews for future buyers by aggregating existing consumer opinions into concise and structured summaries. This, in turn, contributes to better consumer purchase experiences. While our proposed mechanism focuses on how AIGS fosters collective efficacy and motivates contribution, the observed increase in subsequent review ratings supports the outcome that contributors hoped for. Specifically, it suggests that their reviews indeed help future buyers make more informed and satisfying purchase decisions. This reinforces the idea that AIGS not only encourages contribution through enhanced perceptions of collective efficacy but also helps realize the shared goal of improving consumer experiences.

The Effect on Sales: Matthew Effect or Long-tail Effect?

We also examine how the introduction of AIGS interacts with pretreatment baseline sales rank to affect the subsequent sales rank. Products are defined either as “top-ranked” if they rank in the top 20% prior to the policy announcement or as “lower-ranked” otherwise. The 20% cutoff is derived from the classic Pareto principle, commonly known as the 80/20 rule of sales concentration, i.e., the top 20% of products often generate 80% of sales (Brynjolfsson et al. 2010). Accordingly, we specify our regression as follows:

$$\log(Rank_{it}) = \beta_0 + \beta_1 AISummary_{it} + \beta_2 AISummary_{it} \times TopRanked_i + Controls + \mu_i + \delta_t + \varepsilon_{it} \quad (G.1)$$

The dependent variable, $Rank_{it}$, indicates product i 's sales rank in Beauty & Personal Care department in week t (a higher value of sales rank indicates worse sales performance). Since Amazon's sales data are not publicly accessible, we use sales rank as a proxy for sales, a well-established approach in prior studies (e.g., Park et al. 2023; Sun 2012). As shown in Table G.2, the coefficients of the interaction between $AISummary_{it}$ and $TopRanked_i$ are consistently negative. Because we do not invert the sales rank variable, the negative coefficients of $AISummary_{it}$ signify increased sales performance. Therefore, our finding indicates a greater positive effect of introducing AIGS on sales of top-ranked products compared to lower-ranked products.

Table G.2. The Impact of AIGS on Sales Rank: Moderating Role of Pretreatment Sales Rank

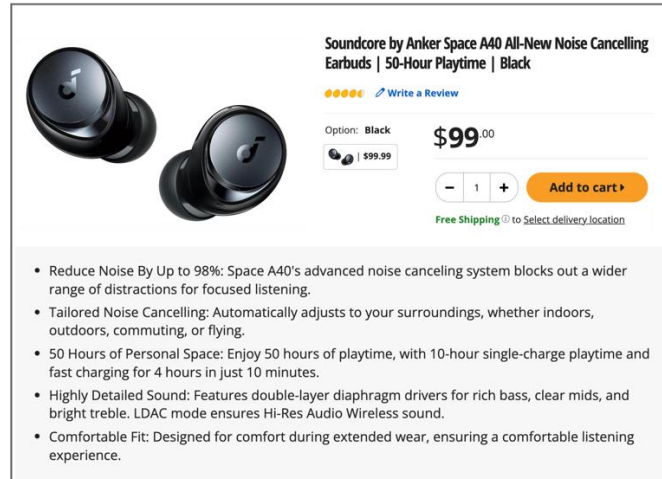
Variables	(1) log(Rank)	(2) log(Rank)
AISummary	-0.072*** (0.019)	-0.073*** (0.019)
AISummary $\times$ TopRanked	-0.070** (0.034)	-0.070** (0.034)
Control variables	No	Yes
Product FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	69,322	69,322

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effect.

This result establishes an important economic implication of AIGS: beyond increasing review volume, AIGS also shapes downstream sales outcomes and the distribution of sales across products. Specifically, the sales gains associated with AIGS are significantly stronger for products that were already top-ranked, whereas the positive effects for lower-ranked products are comparatively limited. This might be due to the possibility that lower-ranked products have fewer existing reviews, which limits the amount and richness of information that can be incorporated into the AIGS. Consequently, the AIGS for these products may be less informative to consumers, which weakens their influence on subsequent purchasing decisions. As a result, AIGS disproportionately enhances the sales performance of products that are already top-ranked, amplifying existing demand concentration within the marketplace.

This finding regarding product sales dynamics highlights a broader competitive implication for e-commerce platforms employing AIGS. By strengthening the sales performance of already high-performing products, AIGS may increase dispersion in seller outcomes and reinforce incumbent advantages. Given the essential role small businesses play in innovation and job creation, the observed “the rich get richer and the poor get poorer” Matthew effect of AIGS could threaten the e-commerce ecosystem by jeopardizing small businesses. This analysis therefore adds to the ongoing discussion about fairness issues within AI applications (Brynjolfsson et al. 2010; Elberse 2008; Gu et al. 2013). From a managerial and platform governance perspective, the adoption of AIGS entails a trade-off between aggregate efficiency gains and distributional balance. While AIGS improves overall marketplace performance, it may reinforce existing disparities. Platforms seeking to maintain a competitive and diverse ecosystem in the long term should take into account these distributional consequences when designing and implementing AIGS.

Online Appendix H. Product Information Page in Experiment



Online Appendix I. Measures in Laboratory Experiment

Collective efficacy

Please indicate the extent to which you agree with the following statements (1 = Strongly disagree, 7 = Strongly agree):

1. “My contributions, together with others’, can effectively influence future buyers’ decisions.”
2. “With the current review structure, contributors can create valuable insights for future buyers.”
3. “On this platform, the collective effort of review contributors efficiently improves the information available to future buyers.”
4. “If more contributors share reviews on this platform, future buyers can gain higher-quality information.”
5. “The collective contributions of review contributors can help future buyers make better decisions more easily.”

Past review experience

To what degree have you had experience with writing online product reviews?

1 = “I have no experience with writing online product reviews.”

7 = “I have a lot of experience with writing online product reviews.”

Product knowledge

To what degree do you have knowledge about wireless Bluetooth earbuds as a product category?

1 = “I have no knowledge about wireless Bluetooth earbuds.”

7 = “I have a lot of knowledge about wireless Bluetooth earbuds.”

Online Appendix J. Rank-to-Sales Transformation

In our main regression model, we use sales rank data as a proxy for sales. Following prior literature (Brynjolfsson et al. 2003; Chevalier and Goolsbee 2003), we conduct an additional rank–sales transformation to estimate sales using Amazon sales rank. Specifically, we assume a Pareto distribution of sales and estimate the relationship between rank and sales using the formulation:

$$\ln(Rank - 1) = c - \theta \ln(Sales)$$

To estimate θ , we follow the approach used in prior literature and conduct an experiment on Amazon. Specifically, we select a product in Amazon’s Beauty & Personal Care department with a sales rank of 1,799,673, for which daily sales are assumed to be close to zero. We purchase 20 units of the product and observe its rank improve to 286,884. We then purchase an additional 20 units, after which the rank further improves to 210,620. Using these observed changes in rank and the corresponding changes in sales, we solve for the implied Pareto shape parameter θ . The estimated θ is 0.446.

Notably, this estimate is lower than values reported in prior studies based on the book category, which typically range from 0.9 to 1.3. A possible explanation is that book sales tend to be more concentrated among top-ranked products, whereas sales of beauty and personal care products in our context may be more dispersed across products. Using the estimated parameters, we calculate estimated sales using Amazon sales rank and then re-estimate our main regression model. The results are consistent when estimated sales are used as a control variable in place of sales rank.

Using the estimated sales, we further construct an approximate measure of review propensity, *EstReviewRate*, calculated by dividing weekly review volume by the estimated weekly sales. Using this estimated review propensity measure as the dependent variable, we re-estimate our main regression model. As shown in Table J.1, the coefficients of *AI Summary* remain significantly positive. The results suggest that the positive effect of AIGS cannot be explained solely by changes in purchase volume and are consistent with the interpretation that AIGS increases consumers’ propensity to contribute reviews.

Table J.1. Impact of AIGS on Estimated Consumer Review Rate

Variables	(1) EstReviewRate	(2) EstReviewRate
AI Summary	0.002*** (0.001)	0.002*** (0.001)
Control variables	No	Yes
Product FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	69,306	69,306

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product level are in parentheses; FE stands for fixed effect.

It is worth noting that the estimated sales are derived from sales rank rather than transaction-level purchase data and therefore provide only an approximation of actual purchase volume. Accordingly, this

analysis should be viewed as additional supporting evidence for the main findings rather than a precise measure of review propensity.

Online Appendix K. Discussion on Alternative Theoretical Explanations

While alternative theories such as social learning theory and the theory of planned behavior also seem to account for the increase in review volume, they face challenges in explaining changes in review content. If AIGS primarily triggers a social learning mechanism (Bandura 1997; Qiu et al. 2021), we would expect subsequent reviews to increasingly mention the product attributes already highlighted in the AIGS. However, our findings show that subsequent reviews discuss attributes already covered by AIGS to a lesser extent, while exhibiting enhanced content richness. From the perspective of the theory of planned behavior, AIGS may increase review contributions by enhancing perceived behavioral control (Ajzen 1991; Pavlou and Fygenson 2006), as the information provided by AIGS may make the reviewing task appear more structured or manageable. Yet, this perspective does not inherently explain why contributors exert effort to elaborate on experiences beyond the existing AIGS content. Instead, our findings are more consistent with a collective efficacy mechanism, which suggests that AIGS highlights the collective value of consumer reviews.

Online Appendix L. Additional Discussion on Generalizability

Generalizability to other platforms. Similar to Amazon, platforms such as Newegg, Microsoft Store, Yelp, and Google Maps integrate AIGS prominently into their user interfaces, using similar formats to summarize consumer opinions (Roth 2024; Saldanha 2024; Sardo 2023). For example, Newegg introduces AIGS with ‘What reviewers are saying about the product,’ while Google Maps prefaces them with labels such as ‘People say.’ The consistent adoption and presentation of AIGS across platforms suggests that the effects observed in our study are not exclusive to Amazon’s implementation. To further strengthen the robustness of our findings, we select Newegg’s AIGS content, rather than Amazon’s, as the basis for our experimental materials. The alignment between our experimental results and our main findings supports the validity of our findings across different AIGS implementations.

Generalizability to other product categories. To further address concerns regarding the generalizability of our findings, we conduct an additional cross-platform analysis using real-world proprietary data from a U.S.-based kitchenware manufacturer. This partner company provides us with weekly data for **kitchenware products** sold on Amazon.com and Walmart.com between May and October 2023. During this period, Amazon implemented AIGS, whereas Walmart did not, allowing us to use Walmart as a control group. Based on this kitchenware dataset, we implement a DID analysis with the following regression model:

$$\log(\text{RevVolume}_{ikt} + 1) = \beta_0 + \beta_1 \text{AISummary}_{ikt} + \text{Controls} + \mu_{ik} + \delta_t + \varepsilon_{ikt} \quad (\text{L. 1})$$

where the dependent variable, $RevVolume_{ikt}$, represents the review volume for product i on platform $k \in \{Amazon, Walmart\}$ during week t . $AI\text{Summary}_{ikt}$ is a dummy variable that indicates whether AIGS is displayed on the product page for product i on platform k in week t . $Controls$ represents control variables. μ_{ik} represents product–platform specific fixed effects.

As shown in Table L.1, the coefficients of $AI\text{Summary}_{ikt}$ are consistently positive across regressions without and with the inclusion of control variables, i.e., Columns (1) and (2), respectively. The results demonstrate that products on Amazon receive a significant increase in consumer reviews after the introduction of AIGS compared to those on Walmart, where AIGS is not implemented. The findings from the kitchenware dataset align with our main results in Amazon’s Beauty & Personal Care category, further supporting the generalizability of our findings to other product categories.

Table L.1. Impact of AIGS on Consumer Review Volume: Evidence from Kitchenware Products

Variables	(1) log(RevVolume + 1)	(2) log(RevVolume + 1)
AI\text{Summary}	0.094^{***} (0.019)	0.092^{***} (0.018)
Control variables	No	Yes
Product–platform FE	Yes	Yes
Weekly FE	Yes	Yes
Observations	6,992	6,992

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors clustered at the product-platform level are in parentheses; FE stands for fixed effects.

Online Appendix M. Additional Clarifications on Experimental Design

Our experimental design follows established paradigms in prior research and aligns with the broader experimental literature on consumer review contribution (e.g., Burtch et al. 2018; Chen et al. 2010; Hou and Ma 2022), in which participants experience the product before being exposed to the focal treatment and deciding whether to contribute a review. These studies examine whether a focal factor affects individuals’ decisions to contribute a review given product experience, which is the same logic that we adopt in our experiment. Importantly, introducing the treatment at this stage enables clean randomization and helps ensure that differences observed in review contribution can be attributed to the treatment rather than to variation in product experience.

In line with this approach, rather than replicating the full marketplace journey, we focus specifically on the contribution stage under controlled conditions in our laboratory experiment. Our goal is to explore whether exposure to AIGS influences individuals’ decisions to contribute a review, conditional on product experience. Importantly, although the experimental procedure abstracts from the full marketplace journey observed in real settings, the psychological mechanism we theorize operates at the contribution stage rather than the purchase stage. Collective efficacy concerns contributors’ beliefs that their review meaningfully

adds to a visible and structured body of collective knowledge. This belief is formed when individuals evaluate whether and how to contribute, regardless of whether AIGS exposure occurs before or after product use. While AIGS may influence expectation formation and satisfaction in real-world settings, our experiment is designed to isolate the contribution-stage mechanism. The observed effect demonstrates that AIGS affects review contribution decisions even when expectation-based satisfaction pathways are held constant. Therefore, although the broader consumption pathway may vary across contexts, the core mechanism of enhanced perceived collective efficacy remains conceptually consistent across laboratory and marketplace settings.

Online Appendix N. Screenshots of Product Review Pages in the Experiment

Treatment condition	Control condition
Customer Reviews of the Soundcore Space A40 WHAT REVIEWERS ARE SAYING ABOUT THE PRODUCT Summary AI AI-generated from the contributions of customer reviews The Soundcore by Anker Space A40 earbuds have received mixed reviews. Most users appreciate their comfortable fit, easy setup, good noise cancellation, and long battery life. The sound quality is generally considered great, and the app is useful for setup and customization. However, there are complaints about the earbuds not staying secure in the ear, and the noise cancellation not performing up to expectations. Additional issues include a non-protective case, occasional high-pitched screeching sounds, and lack of Linux support. Overall, users should weigh these factors according to their personal priorities when considering this product. Pros Comfortable fit Easy to setup Good noise cancellation Cons Case not protective Lost! Great noise cancellation and sound! Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black Mostly listen to classical music and baseball games. Noise cancellation is great and I like that there are three full, partial, and transparent. The sound is also good; better than my Google pixel earbuds. The preset sound profiles are great, plus a customizable option. I would definitely buy again. jfehrenb Great sound. Fits well. Small too. Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black Product fits multiple sizes of ears. Great sound. Small profile. Easy setup. bml Soundcore by Anker Space A40 Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black This is my first pair of noise canceling earbuds. I used these while mowing using a ZTR riding mower. I am very impressed with them. The mower sounded as if it were in the far distance with only a slight humming sound. I did not need to have the volume of the music very loud. I would recommend these earbuds.	Customer Reviews of the Soundcore Space A40 Lost! Great noise cancellation and sound! Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black Mostly listen to classical music and baseball games. Noise cancellation is great and I like that there are three full, partial, and transparent. The sound is also good; better than my Google pixel earbuds. The preset sound profiles are great, plus a customizable option. I would definitely buy again. jfehrenb Great sound. Fits well. Small too. Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black Product fits multiple sizes of ears. Great sound. Small profile. Easy setup. bml Soundcore by Anker Space A40 Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black This is my first pair of noise canceling earbuds. I used these while mowing using a ZTR riding mower. I am very impressed with them. The mower sounded as if it were in the far distance with only a slight humming sound. I did not need to have the volume of the music very loud. I would recommend these earbuds. Cameron Good ANC and great audio clarity. Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black I really enjoy using this set of earbuds. Battery life on them is great. I am not a heavy user, only use them for 5-6 hours a week, and the battery last about 4 months for me. Great audio clarity. ANC is pretty good, but there is no "Off" selection. Either they are in "Through" mode or in ANC. I like to be able to turn ANC off while outside to eliminate wind noise. bobby God tier for the price. Verified Owner Related Product: COMPATIBILITY: Wireless COLOR: Black I bought these after receiving a \$99 pair of AirPods that were literally E-waste quality. These were \$20 cheaper, and like 500x better. I've bought a few Anker products recently and all of them have been incredibly high-quality for the price.

Online Appendix O. Attribute Extraction Procedure

This appendix details the procedures underlying the attribute-level analyses reported in Section 4.4.3. We describe the large language model (LLM)-assisted procedure used for (i) extracting product attributes from consumer reviews and (ii) normalizing semantically equivalent expressions into a common attribute label, as well as (iii) the human validation of the extraction.

1. LLM-Assisted Attribute Extraction

For each consumer review, we use an LLM to identify the set of product attributes discussed in the review text. Recent research shows that LLMs can perform text annotation and information extraction tasks at a level comparable to or exceeding that of trained human coders (Gilardi et al. 2023), which motivates our use of an LLM for this task. We employ GPT-4.1 with the temperature set to zero to minimize randomness and enhance reproducibility of the output. For each review, the model is instructed to return the product attributes that the review explicitly discusses, expressed as single attribute words or short attribute phrases.

We process reviews individually to avoid cross-review contamination, and we retain the model output without manual editing prior to normalization. The extraction prompt is reproduced below.

Prompt:

You are an expert annotator analyzing e-commerce product reviews in the Beauty & Personal Care department. Given the text of a single product review, identify all distinct product attributes discussed in the review.

A product attribute is an identifiable dimension along which a product can be evaluated. In our context, product attributes include physical and compositional characteristics, components, sensory properties, usability, functionality, performance, and direct effects of the product that are explicitly discussed in the review.

Instructions:

1. Identify only attributes that are explicitly discussed in the review. An attribute may be expressed indirectly through a clear description, even when the attribute label itself does not appear in the text. For example, “it lasts all day” may indicate “Battery life.”

2. Use a concise and meaningful attribute label that captures the product aspect being discussed.

3. Use a level of granularity similar to the following types of attributes:

- * physical characteristics, such as Size, Weight, Thickness, Texture, or Color
- * components, materials, and ingredients, such as Clip, Brass, Alcohol, or Vitamin c
- * functionality and performance, such as Charging, Sharpness, Durability, or Speed
- * usability dimensions, such as Ease of use, Ease of installation, or Ease of removal
- * sensory and experiential attributes, such as Comfort, Scent, Softness, or Stickiness
- * product effects and outcomes, such as Moisturizing, Pain relief, Effect on skin, or Effect on hair
- * product problems or resistance to problems, such as Leakage, Shedding, Smudge, or Water resistance
- * evaluative product dimensions, such as Quality, Value, Appearance, or Authenticity

4. Preserve established attribute concepts rather than unnecessary raw details. For example:

- * “lasts all day” → “Battery life”
- * “easy to clean” → “Ease of cleaning”
- * “12 inches long” → “Length”
- * “red” → “Color”
- * “contains vitamin C” → “Vitamin c”
- * “the brass part” → “Brass”
- * “does not smudge” → “Smudge resistance”
- * “smudges easily” → “Smudge”

5. Do not include:

- * sentiment or opinions without an identifiable attribute, such as “I love it” or “It is amazing”
 - * descriptive modifiers such as “good,” “bad,” “beautiful,” or “strong” when they merely express an evaluation of another attribute
 - * exact numerical values or measurement expressions when they can be represented by an attribute such as Length, Size, Weight, or Thickness
 - * brands, product names, retailers, URLs, or unrelated delivery experiences
 - * user characteristics unless the review explicitly discusses the product’s suitability or effect for that characteristic
 - * activities or usage occasions that do not describe an attribute of the product
6. Do not mechanically split a meaningful attribute phrase into separate words. Retain phrases such as “Battery life,” “Sensitive skin,” “Ease of application,” “Smudge resistance,” and “Heat resistance” when the full phrase represents one attribute concept.
7. Separate concepts only when the review clearly discusses distinct attributes. For example, “the brass component is sturdy” may yield:
- Brass
- Sturdiness
8. Use a single noun or a short noun phrase. Use singular form unless the concept is conventionally plural.
9. Remove duplicate attributes within the review. Merge expressions that refer to the same attribute concept.

2. LLM-Assisted Attribute Normalization

Consumers often refer to the same underlying attribute using different surface expressions (for example, “battery,” “battery life,” and “how long the charge lasts”). To avoid treating these as distinct attributes, we further use an LLM to normalize semantically equivalent expressions into a single canonical attribute label. We provide the model with the list of extracted attribute expressions and instruct it to group expressions that can be used interchangeably to refer to the same underlying product attribute and to assign each group a single canonical attribute label, again setting the temperature to zero. The normalization prompt is reproduced below.

Prompt:

You are an expert annotator standardizing product attributes extracted from e-commerce reviews in the Beauty & Personal Care department. Given a list of attribute expressions, group expressions that can be used interchangeably to refer to the same underlying product attribute.

Instructions:

Assign each group a single canonical attribute label. For each attribute expression in the list, return one row with:

- attribute: the original input attribute
- canonical_attribute: the canonical attribute label of its group

Because the number of extracted attribute expressions exceeded the LLM’s token limit, we divided them into batches of 300. This batch-wise procedure could assign different canonical labels to conceptually equivalent attributes across batches. To ensure consistency, we conducted an additional consolidation round in which the initial canonical labels were further grouped and standardized by the LLM, allowing semantically equivalent labels to be merged into a common canonical attribute label.

After normalization, each extracted attribute expression corresponds to a canonical attribute label. Following the procedure described in Section 4.4.2, an attribute is treated as covered by AIGS if its corresponding canonical attribute label matches one of the AIGS attribute keywords or their cognitive synonyms retrieved from NLTK WordNet (Bird et al. 2009). Attributes whose canonical attribute label does not match are treated as non-AIGS attributes.

3. Human Validation

To assess the reliability of the LLM-assisted extraction, two research assistants with experience in e-commerce and online reviews independently annotated the product attributes discussed in a random sample of 200 consumer reviews, without access to the LLM output. The two annotators agreed on 99.75% of the binary review-by-attribute judgments, corresponding to a Cohen’s kappa of .90. All disagreements were resolved through discussion to arrive at a consensus coding. We then compared the human-coded attributes with the LLM-extracted attributes for the same reviews. The two sources agreed on 99.70% of the binary review-by-attribute judgments, corresponding to a Cohen’s kappa of .86 (Cohen 1960), indicating a high level of agreement. The remaining discrepancies were examined and did not reveal systematic extraction errors, supporting the validity of the LLM-assisted procedure.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- AnkitAI. (2025). *Reviews-roberta-base-sentiment-analysis*. HuggingFace model. Available at: <https://huggingface.co/AnkitAI/reviews-roberta-base-sentiment-analysis>.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman & Co.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Sebastopol, CA.
- Brynjolfsson, E., Hu, Y., & Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49(11), 1580–1596.
- Brynjolfsson, E., Hu, Y., & Smith, M. D. (2010). Research commentary—long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns. *Information Systems Research*, 21(4), 736–747.
- 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.
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Chen, J., & Roth, J. (2024). Logs with zeros? Some problems and solutions. *The Quarterly Journal of Economics*, 139(2), 891–936.
- Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review*, 100(4), 1358–1398.
- Cheng, Z., Pang, M. S., & Pavlou, P. A. (2020). Mitigating traffic congestion: The role of intelligent transportation systems. *Information Systems Research*, 31(3), 653–674.
- Chevalier, J., & Goolsbee, A. (2003). Measuring prices and price competition online: Amazon. com and BarnesandNoble. com. *Quantitative Marketing and Economics*, 1(2), 203–222.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996.
- Elberse, A. (2008). Should you invest in the long tail?. *Harvard Business Review*, 86(7/8), 88.
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), e2305016120.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Gu, B., Tang, Q., & Whinston, A. B. (2013). The influence of online word-of-mouth on long tail formation. *Decision Support Systems*, 56, 474–481.
- Hartman, E., & Hidalgo, F. D. (2018). An equivalence approach to balance and placebo tests. *American Journal of Political Science*, 62(4), 1000–1013.
- Hou, J., & Ma, X. (2022). Space norms for constructing quality reviews on online consumer review sites. *Information Systems Research*, 33(3), 1093–1112.
- Huang, N., Hong, Y., & Burtch, G. (2017). Social network integration and user content generation. *MIS Quarterly*, 41(4), 1035–1058.
- Lei, Z., Yin, D., & Zhang, H. (2021). Focus within or on others: The impact of reviewers' attentional focus on review helpfulness. *Information Systems Research*, 32(3), 801–819.
- Li, X. (2018). Impact of average rating on social media endorsement: The moderating role of rating dispersion and discount threshold. *Information Systems Research*, 29(3), 739–754.
- Li, Z., & Wang, G. (2025). Regulating powerful platforms: Evidence from commission fee caps. *Information Systems Research*, 36(1), 126–140.
- Liu, L., Wang, Y., & Xu, Y. (2024). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 68(1), 160–176.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

- 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.
- Oh, H., Goh, K. Y., & Phan, T. Q. (2023). Are you what you tweet? The impact of sentiment on digital news consumption and social media sharing. *Information Systems Research*, 34(1), 111–136.
- Park, S., Shin, W., & Xie, J. (2023). Disclosure in incentivized reviews: Does it protect consumers?. *Management Science*, 69(11), 7009–7021.
- Pavlou, P. A., & Fygenon, M. (2006). Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS Quarterly*, 30(1), 115–143.
- Pennebaker, J. W., Boyd, R. L., Booth, R. J., Ashokkumar, A., & Francis, M. E. (2022). Linguistic Inquiry and Word Count: LIWC-22. Pennebaker Conglomerates.
- Qiu, L., Chhikara, A., & Vakharia, A. (2021). Multidimensional observational learning in social networks: Theory and experimental evidence. *Information Systems Research*, 32(3), 876–894.
- Roth, E. (2024). Google Maps will use AI to answer questions about the new restaurant you want to try. Accessed December 17, 2024, <https://www.theverge.com/2024/10/31/24283970/google-maps-gemini-ai-answer-questions/>.
- Saldanha, C. (2024). Yelp releases a series of new discovery, contribution, services and AI-powered features. Accessed May 26, 2024, <https://blog.yelp.com/news/winter-product-release-2024/>.
- Sardo, G. (2023). Welcoming AI to the Microsoft Store on Windows. Accessed May 26, 2024, <https://blogs.windows.com/windowsdeveloper/2023/05/23/welcoming-ai-to-the-microsoft-store-on-windows/>.
- Schermerhorn, V. (2023). How Amazon continues to improve the customer reviews experience with generative AI. Accessed May 26, 2024, <https://www.aboutamazon.com/news/amazon-ai/amazon-improves-customer-reviews-with-generative-ai>.
- Sun, M. (2012). How does the variance of product ratings matter?. *Management Science*, 58(4), 696–707.
- 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.
- Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, 54(3), 447–463.