The Effect of Gender-Aware Curation Algorithms on User Engagement in User-Generated Content Platforms

Xinyi Zhang[1], Khim-Yong Goh[2], Chenshuo Sun[3], Renyu Zhang[4]

[1] National University of Singapore, Singapore, xinyizhang@u.nus.edu

[2] National University of Singapore, Singapore, gohky@comp.nus.edu.sg

[3] New York University, New York, USA, csun@stern.nyu.edu

[4] The Chinese University of Hong Kong, Hong Kong, China, philipzhang@cuhk.edu.hk

Although gender-aware curation algorithms (i.e., incorporating users' gender data for content curation or recommendation) have been increasingly adopted, their effects on user engagement in user-generated content (UGC) platforms remain unclear. In this work, we collaborated with a leading short-video platform in Asia to conduct a large-scale randomized experiment by assigning about 400,000 users to either a gender-aware or gender-neutral algorithm for ranking comments. We examined engagement outcomes for comment consumption (i.e., browsing duration) and comment production (i.e., posting incidence). Our findings show that gender-aware curation (relative to gender-neutral approach) increased the average browsing duration of comments by 1.0%. Intriguingly, it had asymmetric effects on comment production where female users (who were 25% less likely to comment than males) increased their commenting likelihood by 2.6%, whereas male users were 3.5% less likely to do so, underscoring the increased gender diversity in online user engagement due to gender-aware curation. Mechanism analyses suggest that the heterogeneous effects originate from differing engagement motives: male users aim to assert individual perspectives for self-presentation purposes, whereas female users are motivated by a desire for community belonging. In sum, our results show that gender-aware curation algorithms can boost user engagement and foster gender diversity of online discussions on UGC platforms. This study underscores the importance of incorporating gender data into curation algorithms, contributes theoretically to research on curation algorithms, user engagement, and gender-based behavioral differences, and offers practical insights for platform managers to design more engaging UGC platforms.

Key words: Curation Algorithm, Gender Data, User-Generated Content Platforms, Content Consumption, Content Production

1. Introduction

User engagement, encompassing both *content consumption* and *content production*,[1] is critical for user-generated content (UGC) platforms such as TikTok and Twitter (Dou et al. 2013, Zhao et al.

[1] The operationalization of *content consumption* (*production*) depends on the form of the focal UGC. In prior studies on short video platforms, videos were the focal UGC and comments were secondary (Zeng et al. 2023), so *content consumption* (*production*) refers to *video consumption* (*production*). In our study, comments are the focal UGC, so *content consumption* (*production*) refers to *comment consumption* (*production*).

2024).[2] These platforms rely on high engagement to retain users, foster community growth, and drive monetization. Of all the platform features that foster user engagement (Agarwal et al. 2025), user comments play a pivotal role (Cheng et al. 2024). Positioned beneath the original posts, these comments offer dedicated spaces for users to share opinions, engage in discussions, seek clarification, and connect with others. In particular, 77% of TikTok users read comments on videos[3], and Reddit users spend about 52% of their total time on comment sections[4]. To manage the vast volume of user comments, UGC platforms use content curation algorithms to filter, highlight, and rank comments, and to personalize displays based on users' historical engagement data (Berman and Katona 2020).

Rather than relying solely on engagement data, many platforms are now exploring *gender-aware curation algorithms*, i.e., explicitly incorporating users' gender data, in addition to user engagement data, into curation algorithms (Kelley et al. 2022).[5] Examples include YouTube, Qloo, and Spotify.[6] This trend is motivated by growing industry consensus and academic findings that men and women differ systematically in engagement patterns and content preferences (Varma et al. 2023), and such differences may not be fully reflected in engagement data alone, but may be better captured through gender information (Sun et al. 2024, Xu and Zhang 2022). In contrast, some platforms like Netflix explicitly state that they do not use gender data in their recommendation algorithms, relying only on engagement data such as viewing history.[7] This divergence in industry practices reflects ongoing debate about the effects of incorporating gender data into curation algorithms on user engagement (Pinney et al. 2023).

Against this backdrop, our study builds on research gaps from three streams of literature to motivate our research questions: (1) curation algorithms, (2) user engagement on UGC platforms, and (3) gender-based behavioral differences. Notably, the effects of gender-aware curation on user engagement are ambiguous from a theoretical perspective. On one hand, incorporating gender information may improve personalization by aligning comments more closely with users' preferences, encouraging users to spend more time reading and posting comments (Sun et al. 2024, Adomavicius et al. 2008). On the other hand, the benefits of gender-aware curation may be limited if users' gender information can already be inferred from user engagement data (Kosinski et al. 2013). Moreover, even when personalization improves, user engagement may not necessarily

[2] These platforms operate as two-sided markets where users act as both content users and producers. Thus, fostering both consumption and production is essential for sustained growth (Zeng et al. 2023).

[3] See `https://blog.brandbastion.com/tiktok-marketing-strategy`.

[4] See `https://backlinko.com/reddit-users`.

[5] Using gender information for personalization is generally acceptable on entertainment-focused UGC platforms (e.g., TikTok), where the goal is to enhance user experience rather than make high-stakes decisions (e.g., loan approvals) (Binns et al. 2018).

[6] See explanations from YouTube, Qloo, and Spotify.

[7] See `https://help.netflix.com/en/node/100639`.

increase. Users may find like-minded comments repetitive or feel that their own perspectives have already been expressed, reducing both their interest in reading and motivation to comment (Butler 2001, Zhou et al. 2025, Warlop et al. 2018).

These mixed predictions call for rigorous causal evidence to understand the exact impact of gender-aware curation algorithms. From a practical standpoint, such evidence can guide curation algorithm design for platform managers and data regulation policies for policymakers. However, on the consumption side, existing research on curation algorithms has not empirically tested the economic impact of explicitly incorporating gender information (Sun et al. 2024). On the production side, prior studies on curation algorithms have largely operationalized user engagement as consumption outcomes (e.g., viewing or browsing) with curated content, leaving production behaviors underexplored (Sun et al. 2024). Moreover, studies on user engagement on UGC platforms have primarily focused on other information technology (IT) features (e.g., reputation systems and identity disclosure) (Ma and Agarwal 2007, Wasko and Faraj 2005) but not on gender-aware curation algorithms. In short, neither the curation algorithm literature nor the user engagement literature has established causal evidence on how gender-aware curation algorithms shape user engagement. Motivated by both theoretical ambiguity and practical significance, we propose our first research question: *What is the impact of gender-aware curation algorithms on user engagement (consumption and production behaviors) for an online UGC platform?*

Beyond the effectiveness of gender-aware curation algorithms, it is also intriguing to investigate how the effects differ across genders. Prior sociolinguistic research (Wardhaugh and Fuller 2021, Coates 2015) finds that men and women engage in conversation for different motivations—men value distinctiveness, while women seek consensus. Therefore, if gender-aware curation amplifies like-minded viewpoints, male users may engage less due to reduced uniqueness-driven motivations, whereas female users may engage more due to a stronger sense of belonging. These possibilities underscore the need for empirical research to dig into the heterogeneous effects of gender-aware curation on user engagement across genders. This is practically important as UGC platforms strive to ensure content delivery does not disproportionately affect any groups, thereby maintaining diverse user bases, fostering opinion diversification, and supporting long-term growth (Lambrecht and Tucker 2019). However, prior studies on user engagement on UGC platforms have primarily focused on other IT features rather than gender-aware curation algorithms, and thus overlooked gender-based heterogeneity in the effects of gender-aware curation on user engagement (Ma and Agarwal 2007, Wasko and Faraj 2005). Likewise, although research on gender-based behavioral differences has established that men and women differ in engagement motivations, it has largely neglected algorithmic influences (Sun et al. 2024, Huang et al. 2017, Bucher-Koenen et al. 2024). In short, both streams of literature have overlooked how gender-aware curation algorithms interact with gender-specific motivations to shape user engagement in both consumption and production behaviors. To address this research gap, we propose our second research question: *How does the user engagement effect of gender-aware curation algorithms differ between female and male users?*

To answer the research questions, we collaborated with a leading short-video platform in Asia (hereafter referred to as "Platform A" due to a non-disclosure agreement) to conduct a large-scale randomized field experiment. Like TikTok, comments are a key driver of user engagement on Platform A[8]. Before October 2023, the platform ranked first-level comments based on comment attributes and user historical engagement, without incorporating demographic data like gender as input (detailed in Section 3.1). Recognizing that male and female users have distinct preferences over comments, Platform A started to introduce a gender-aware curation algorithm that leverages the user's gender information to rank comments in October 2023.

In our experiment, users were randomly assigned to either the treatment group, where first-level comments were ranked using the gender-aware curation algorithm, or the control group, where the algorithm remained gender-neutral. The experiment ran from October 30th to November 19th, 2023. Our analyses of experimental data yield several important findings. First, we find that the gender-aware curation algorithm boosted users' comment consumption in terms of browsing duration in the comment section by 1.0%, with no heterogeneous effect across genders. Second, it influenced comment production in a heterogeneous manner: female users were 2.6% more likely to comment, whereas male users were 3.5% less likely to do so. Given that female users are 25% less likely to post comments on platforms, the results imply that a gender-aware curation algorithm leads to higher gender diversity in online user engagement. Third, mechanism analyses suggest the heterogeneous effects stem from different engagement motivations by users of different genders: male users seek to assert distinctiveness for self-presentation purposes, while female users are motivated by consensus and a desire for community belonging.

In summary, our study shows that gender-aware comment curation algorithms, compared to gender-neutral ones, can improve user engagement by increasing comment consumption overall and comment production among female users. Our work delivers several theoretical and practical contributions. First, we advance the literature on curation algorithms (Zhou et al. 2025, Sun et al. 2024) by providing the first causal evidence that incorporating gender data into curation algorithms indeed enhances user engagement. This also contributes to data anonymization research (Kosinski et al. 2013, Xu and Zhang 2022) by demonstrating the added value of explicit gender data beyond inferences based on user engagement data. We further contribute by enriching the operationalization of user engagement, moving beyond consumption to incorporate production

[8] See `https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you`.

behaviors as a complementary dimension. Second, we contribute to the literature on user engagement on UGC platforms (Phang et al. 2015, Huang et al. 2019) by highlighting the overlooked role of gender-aware curation algorithms in shaping user engagement and revealing how the effects of curation algorithms vary by gender. Third, we extend research on gender-based behavioral differences (Varma et al. 2023, Venkatesh and Morris 2000) by showing how curation algorithms interact with gender-specific motivations would result in asymmetric engagement responses. Lastly, we provide actionable insights for platform managers, emphasizing the potential of gender-aware curation to increase user engagement and promote gender diversity in UGC platform participation.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 details our research context and empirical strategy. Section 4 presents the effect of gender-aware curation algorithms on user engagement in terms of comment consumption and production. Section 5 explores the mechanisms underlying these effects. Section 6 presents additional analyses and robustness tests. Lastly, Section 7 summarizes the findings and discusses the theoretical and practical contributions, as well as directions for future research.

2. Literature Review and Theoretical Background

2.1. Curation Algorithms

Our work is most closely related to research on the economic impact of curation algorithms. One stream examines the effects on user engagement and satisfaction (Zhou et al. 2025, Sun et al. 2024). While many report positive effects, some studies caution that overly personalized or homogeneous content may reduce engagement by limiting content diversity (Zhou et al. 2025, Berman and Katona 2020), inducing boredom (Warlop et al. 2018), or raising privacy concerns (Karwatzki et al. 2017). Another stream investigates content diversity, with mixed findings on whether curation algorithms reinforce filter bubbles or promote exploration beyond users' initial preferences (Bakshy et al. 2015, Hosanagar et al. 2014). A third stream examines how algorithm designs, such as incorporating variety-seeking preferences, affect engagement (Li and Tuzhilin 2024, Zeng et al. 2024).

Despite these advances, two key gaps remain in this stream of literature. First, research has largely overlooked the impact of incorporating personal data like gender into curation algorithms. While Sun et al. (2024) has examined the role of personal data in e-commerce recommender systems, the study has not isolated gender as a distinct factor. Understanding its independent effect is important, as gender is both behaviorally meaningful (Varma et al. 2023) and central to ongoing debates on fairness, discrimination, and responsible algorithmic design (Kelley et al. 2022). In addition, engagement on e-commerce platforms is goal-oriented and transactional, whereas engagement on UGC platforms is driven by identity expression and social interaction. Therefore, gender-aware curation may boost engagement in e-commerce by matching users to preferred products, but may reduce engagement on UGC platforms by limiting content diversity (Zhou et al. 2025) and self-presentation opportunities (Oh et al. 2023).

Moreover, the effects of explicitly integrating gender data into curation algorithms on user engagement are still uncertain. Incorporating gender data may help curation algorithms better capture user preferences and enhance engagement (Sun et al. 2024), as men and women exhibit distinct content preferences and engagement patterns (Venkatesh and Morris 2000, Huang et al. 2019) that may not be fully captured by engagement data alone (Lian et al. 2018, Zhou et al. 2025). However, studies on data anonymization suggest that algorithms can infer gender-related preferences from engagement data (Kosinski et al. 2013), and even if gender-aware curation improves personalization, it may not increase engagement due to privacy concerns or more homogeneous content delivery (Zhou et al. 2025, Karwatzki et al. 2017). Our paper reconciles these views with causal evidence comparing gender-aware and gender-neutral algorithms on user engagement and further examines whether the effects vary by gender.

Second, prior studies have primarily operationalized user engagement by consumption behaviors with curated content (i.e., in our context, comment reading), leaving production behaviors (e.g., comment posting) understudied. However, both behaviors are essential to UGC platforms (Zeng et al. 2023), and driven by different motivations (Phang et al. 2015). Consumption is driven by interest, relevance, or entertainment value, whereas production is often motivated by social recognition, community belonging, or self-expression (Phang et al. 2015, Wasko and Faraj 2005). A feature that boosts consumption may not similarly affect production. For example, Kane and Ransbotham (2016) shows that more developed content increases user consumption but discourages production. A small but growing literature examines how curation algorithm designs influence engagement in terms of production. Qian and Jain (2024) analytically shows that favoring quality over personalization can encourage high-quality UGC production. Zeng et al. (2024) empirically shows that prioritizing high-quality videos increases consumption but discourages production among less visible users. Zou et al. (2024) analytically shows that emphasizing quality over variety can reduce production from niche producers. Wang et al. (2025) propose a curation algorithm that jointly considers consumption and production objectives. However, these studies have yet to examine how incorporating personal data, such as gender, shapes engagement in production.

Theoretically, the impact of gender-aware comment curation on comment production can be ambiguous. On one hand, social identity theory posits that a stronger sense of community fosters engagement (Wasko and Faraj 2005). By amplifying comments tailored to each gender's interests (Bakshy et al. 2015, Hosanagar et al. 2014), gender-aware curation algorithms may reinforce homophily and belonging, encouraging more postings. On the other hand, social exchange theory argues that users post to gain social rewards (e.g., status and recognition) or maintain distinctiveness (Wasko and Faraj 2005, Cao et al. 2024). Prioritizing like-minded content may inadvertently reduce these incentives if users perceive their perspectives as already well-represented (Butler 2001). Consistent with this view, Kane and Ransbotham (2016) shows that greater visibility on Wikipedia reduces editing as content becomes more refined, and Beaudoin (2002) finds that users in online learning post less once they feel their views are already voiced. This aligns with research on the bystander effect (Wong et al. 2021), group size and user engagement (Butler 2001), and social loafing and free riding (Lount Jr and Wilk 2014), all of which highlight how perceived redundancy, diffused responsibility, and reduced accountability can discourage posting. We extend this literature by empirically examining how gender-aware curation algorithms influence user engagement, with a particular focus on the often-overlooked production dimension and its heterogeneous effects across genders.

2.2. User Engagement on UGC Platforms

Our work is closely related to research on the antecedents of user engagement on UGC platforms. Existing studies commonly operationalize user engagement as consumption or production of UGC (Kane and Ransbotham 2016, Cao et al. 2024). Consumption involves low-effort behaviors such as reading, liking, or sharing, whereas production refers to high-effort behaviors such as posting and replying (Phang et al. 2015). The classification depends on the focal UGC context. For example, on short video platforms, when videos are the focal UGC, viewing videos is consumption, whereas posting videos is production (Zeng et al. 2024). In our context, where comments are the focal UGC, reading comments is consumption, whereas posting comments is production.

Research on user engagement in terms of consumption has examined factors such as content attributes (Cheng et al. 2024, Oh et al. 2023), social cues (e.g., peer endorsements (Agarwal et al. 2025)), and platform design features (e.g., content length limits (Gu and Zhao 2024), feed integration (Cao et al. 2024), and curation algorithms (Zeng et al. 2024)). This research stream is less developed, partly because consumption behaviors are less visible and often require platform-provided clickstream data (Oh et al. 2023). In contrast, research on user engagement in the form of production (also called user contribution) is more extensive and has examined (1) why users produce and (2) why lurkers do not. Motivations for production generally fall into two categories: community engagement and reputation building. Shared identity and social bonds encourage replies and discussions (Phang et al. 2015, Huang et al. 2017), as active users often engage to strengthen social bonds or fulfill social obligations (Ma and Agarwal 2007). Social reputation is another key driver, with users producing UGC to showcase expertise, gain visibility, or differentiate themselves from others (Wasko and Faraj 2005, Huang et al. 2019, Ma and Agarwal 2007). This aligns with social exchange theory, where contributions function as a form of currency for status and recognition (Cao et al. 2024). Other motivators include altruism (Qiao et al. 2020), monetary incentives (Wang et al. 2022), and reciprocity (Ma and Agarwal 2007).

In contrast, lurkers, users who consume without producing, face distinct psychological and social barriers. Many report low confidence due to shyness, feelings of inadequacy, or uncertainty about their expertise (Brescoll 2011, Coffman 2014, Phang et al. 2015). Social pressure and fear of negative evaluation further discourage participation, leading to self-censorship (Brescoll 2011, Peng et al. 2025). Beyond individual confidence, lack of community belonging also inhibits engagement, as users who feel socially disconnected see little value in engagement (Huang et al. 2017, Phang et al. 2015). Additionally, some perceive their input as redundant, particularly in large or highly active communities where similar perspectives are already expressed (Beaudoin 2002). This phenomenon goes beyond free-riding behaviors where individuals withhold contributions assuming others will act, to a crowding-out effect where existing contributions create a sense of saturation that suppresses further input (Wong et al. 2021).

Despite this extensive research, two key gaps remain. First, we lack evidence on whether and how gender-aware curation algorithms affect user engagement. In terms of consumption, no study has empirically tested the effects of incorporating gender data into UGC curation. In terms of production, prior research has focused on IT artifacts such as reputation systems, identity disclosure, and virtual co-presence (Ma and Agarwal 2007, Wasko and Faraj 2005), but has overlooked the role of curation algorithms. It is unclear whether these algorithms foster participation by strengthening community bonds or suppress it by reducing the reputational rewards of differentiation. Second, potential heterogeneous effects by gender remain underexplored. Prior research suggests that men value distinctiveness, while women prioritize consensus and community bonds (Coates 2015, Cross et al. 2011). If gender-aware curation algorithms amplify like-minded opinions, these algorithms may encourage engagement among female users while discouraging it among male users, i.e., effects that have yet to be examined. Huang et al. (2019) also documents gender heterogeneity in user engagement on UGC platforms but focuses on performance feedback interventions rather than curation algorithms. Our study fills in these gaps by empirically evaluating the causal impact of gender-aware comment curation on user engagement in terms of comment consumption and comment production, and its heterogeneous effects by gender.

2.3. Gender-Based Behavioral Differences

Scholars across fields such as economics, sociolinguistics, and information systems (IS) have long examined gender-based behavioral differences across various domains such as self-promotion (Peng et al. 2025), performance feedback and user engagement (Huang et al. 2019), conformity behaviors (Griskevicius et al. 2006), conversational engagement communicative style (Coates 2015, Brescoll 2011, Leaper and Ayres 2007), prosocial behaviors (Soutschek et al. 2017), technology adoption(Venkatesh and Morris 2000), financial investment (Bucher-Koenen et al. 2024), workplace networking (Jeong et al. 2022), and risk-taking behaviors (Varma et al. 2023).

To explain these differences, prior studies often draw on agency-communion theory and self-construal theory (Varma et al. 2023). Agency-communion theory posits that human behavior is shaped by agency, which emphasizes independence and self-assertion, and communion, which emphasizes social connection and harmony, with men more agentic and women more communal (Helgeson 1994). Similarly, self-construal theory argues that men are more likely to develop an independent self-construal (e.g., autonomy, assertiveness, competitiveness), whereas women tend to develop an interdependent self-construal (e.g., relationships, connection) (Cross et al. 2011). Related perspectives like gender self-schema theory (Huang et al. 2019) offer consistent arguments.

Despite the extant work, little is known about how curation algorithms incorporate these gender-specific differences to produce heterogeneous effects on user engagement. Our study fills in this gap by offering empirical evidence that examines the nuanced effect of gender-aware curation on user engagement by gender. By increasing personalization, the gender-aware curation algorithm exposes users to more like-minded comments (Bakshy et al. 2015, Adomavicius et al. 2008), which may shape male and female engagement behaviors differently. For female users, this exposure may increase engagement for several reasons. First, exposure to like-minded comments can reduce female users' hesitation to engage by addressing concerns about competence and fear of negative judgment (Varma et al. 2023, Coffman 2014), as well as social pressure to self-censor due to gender-based social norms (Brescoll 2011, Peng et al. 2025), especially when discussing nonconforming topics such as the military (Bordalo et al. 2019). Second, like-minded comments strengthen female users' perceived belongingness and social identity, thereby encouraging engagement (Ma and Agarwal 2007, Huang et al. 2017), particularly in male-dominated settings where female users might otherwise feel underrepresented. Third, by fostering a less confrontational environment, like-minded comments cater to women's preference for harmonious discussions and encourage engagement (Wardhaugh and Fuller 2021, Griskevicius et al. 2006). Fourth, based on muted group theory, exposure to aligned perspectives offers expressive cues that help female users articulate their views, thereby lowering their engagement barriers (Coates 2015, Wardhaugh and Fuller 2021).

For male users, gender-aware curation algorithms may reduce the incentive to engage for several reasons. First, men are generally motivated by the desire for visibility, recognition, and opportunities for social comparison in group discussions (Peng et al. 2025, Coates 2015). Like-minded comments reduce the perceived uniqueness and social rewards of engaging, weakening their drive to stand out or assert distinct perspectives (Wasko and Faraj 2005, Butler 2001). Second, men often adopt an information-oriented communication style, engaging when they believe their input adds new or valuable insights (Venkatesh and Morris 2000). By reinforcing existing viewpoints, the algorithm may reduce the perceived informational value, leading to lower engagement. Third, as dominant participants in online discussions (Peng et al. 2025), male users are more prone to social loafing. When similar comments are visible, male users may feel less obliged to engage, assuming others have expressed shared perspectives.

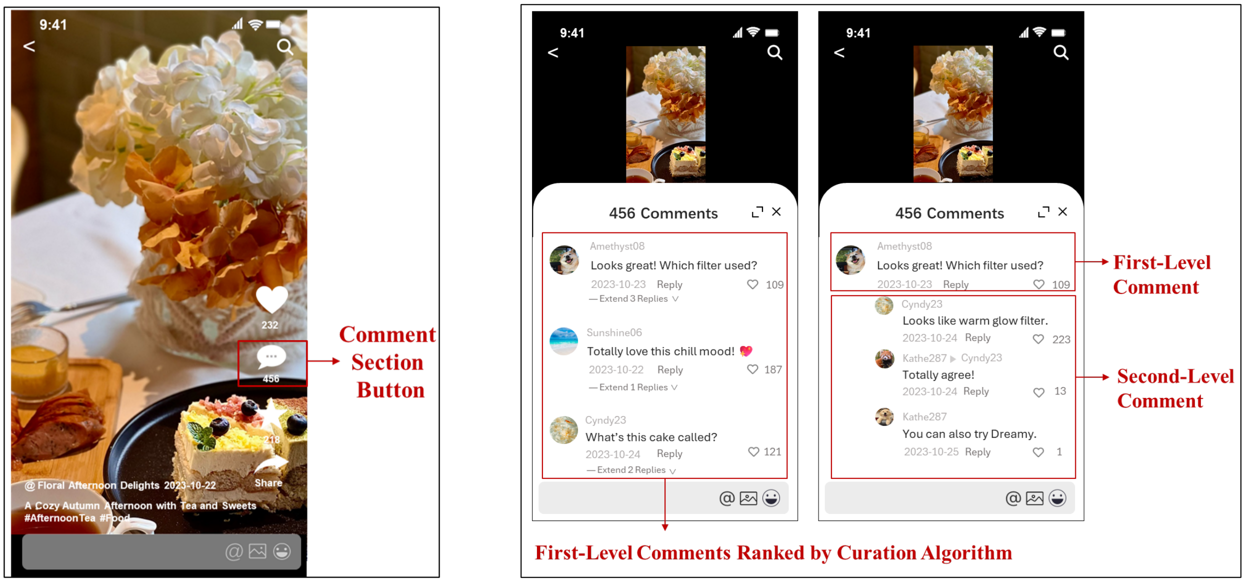
3. Research Context and Empirical Strategy

3.1. Research Context

We collaborated with one of the largest short-video platforms in Asia (Platform A), which has over 300 million daily active users. Like TikTok, Platform A allows users to watch videos and interact for free, and generates revenue primarily through online advertising. Because advertising effectiveness depends on users' content consumption and interaction, user engagement is central to Platform A's business model. To encourage active user engagement, Platform A prioritizes interactive features, with user comments playing a central role. Within the comment section, users can browse, reply to, and discuss about videos. After watching a video, users can access the comment section by tapping a button on the video page, which minimizes the video thumbnail while expanding the comment section below (see Figure 1 (a)). Users primarily engage in the comment section either by reading existing comments (comment consumption) or by posting and replying to comments as commenters (comment production). The comment section follows a two-level hierarchy. First-level comments are direct comments on the video, appearing immediately below the video as primary responses and displaying the commenter's avatar, username, content, and likes count. Second-level comments, including replies to both first-level and other second-level comments, are accessible through the "extend replies" button beneath first-level comments and are displayed in chronological order (see Figure 1(b)). Due to the large volume of comments, Platform A employs a curation algorithm to rank first-level comments to ensure an engaging experience.

The comment curation algorithm was trained based on data from all platform users. Prior to October 2023, the training inputs consisted of two types of features: (1) comment attributes such as the total number of likes and replies a comment had received, and semantic features such as comment text embeddings, and (2) users' historical engagement with comments, such as browsing duration, likes, clicks on the "extend replies" button, and replies. Importantly, the algorithm did not incorporate personal user data, such as gender, partly due to uncertainty about its effects on engagement beyond historical engagement data.

The algorithm employed content-based, collaborative filtering, and context-aware methods to rank comments based on predicted engagement. For example, content-based methods leverage semantic similarity in comment text, so that users are more likely to see comments with similar



(a) (b)

Figure 1 **User Interface for Accessing Comment Sections on Platform A**

features to those they have previously engaged with. Collaborative filtering draws on patterns of comparable users, so that users are more likely to see comments that these similar users have engaged with. In practice, for a given user and the set of first-level comments under a video, the model predicted the probability that the user would like, reply to, or click the "extend replies" button for each comment, based on both the comment features and patterns learned during training[9]. These predicted probabilities were combined into a composite score through a multi-objective function, with weights manually set by the operation team on Platform A to balance engagement outcomes. First-level comments were then ranked by this score, whereas second-level comments were displayed chronologically.

3.2. Experiment Design

We observed that male and female users exhibited significantly different content preferences, consistent with longstanding evidence of gender-based behavioral differences documented across diverse domains (Huang et al. 2019, Venkatesh and Morris 2000). Building on this observation, we collaborated with Platform A to design and implement a gender-aware curation algorithm in October 2023. This algorithm incorporates users' gender data into the ranking process, but does not include the gender of video or comment producers. Gender data provides complementary information beyond engagement data: it supports cold-start scenarios where engagement data are sparse, offers a stable demographic signal less volatile than engagement data, and allows the

[9] Although browsing duration is an important engagement metric, the objective function at the experimental period only included the prior-mentioned three explicit binary behaviors.

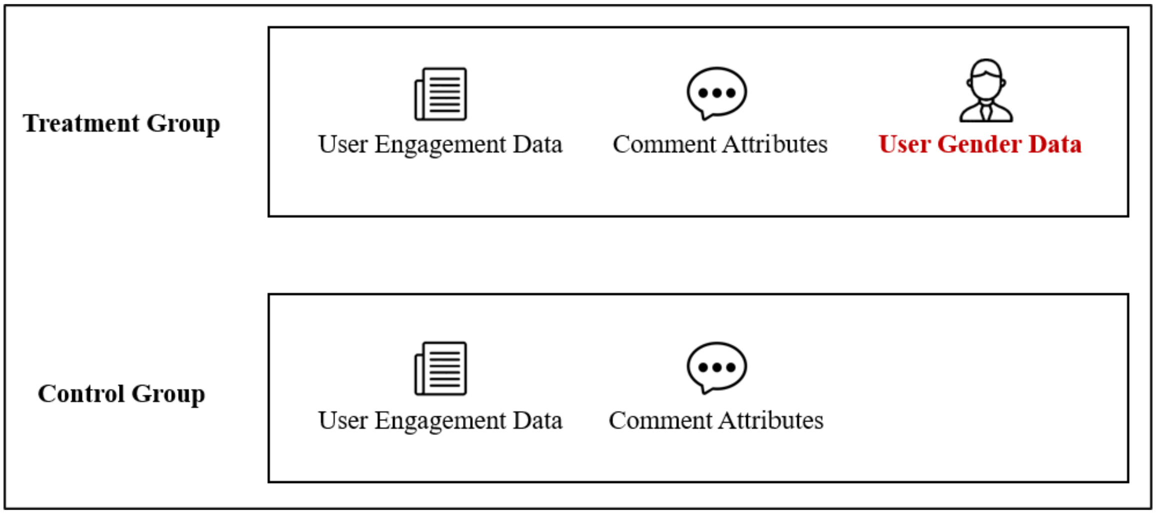


Figure 2 **Data Inputs for Curation Algorithms in Treatment and Control Groups**

model to differentiate heterogeneous preferences across users of different genders. This allows the algorithm to better capture how gender interacts with comment features and engagement behaviors. Notably, this experimental manipulation did not alter video recommendations.[10]

To causally examine the impact of incorporating gender data into the curation algorithm, we conducted a field experiment in comment sections of video pages on Platform A. The experiment ran from October 30th to November 19th, 2023, during which users were randomly assigned to either the treatment or control group. Users in the treatment group saw first-level comments ranked using a gender-aware curation algorithm that incorporated the users' gender. Those in the control group consumed first-level comments ranked by a gender-neutral algorithm that excluded gender data (see Figure 2). Importantly, users were unaware of the experiment, i.e., they were neither notified about the ranking criteria nor informed that the order of comments shown might differ from that of other users. This design ensured that any observed differences in user engagement outcomes should be attributed to the treatment rather than user expectations about experimental manipulation or algorithmic changes.

3.3. Data and Variables

We defined the first day after the experiment began with the treatment stably implemented as the treatment period and the day immediately before the experiment as the pre-treatment period, following Zeng et al. (2023)[11]. This one-day time unit aligns with the granularity of our data

[10] We compared the average share of recommended videos across first-level categories between treatment and control groups using two-sample t-tests. The results, reported in Online Appendix A.1, show no significant differences (p-values > 0.1), confirming that the gender-aware comment curation algorithm did not affect the platform's video recommendation system.

[11] The experiment launched on October 30th, but the platform was still adjusting the algorithm's parameters that day. Thus, we defined October 29th as the pre-treatment period and October 31st as the treatment period for our main analyses.

provided by Platform A. Platform A aggregates daily metrics (e.g., daily active commenters, daily browsing duration of comments), which decompose into user-level daily behaviors (e.g., whether a user browsed comments on a given day, how much time the user spent browsing comments, and how many comments the user posted).[12]

Our data sample included 383,135 users (i.e., who had browsed comments) on October 31st, with 191,420 in the treatment group and 191,715 in the control group. Among these users, the treatment group comprised 101,992 female and 89,428 male users, while the control group included 102,340 female and 89,385 male users. To ensure that our analyses captured the engagement with recent video content, we restricted the sample to videos published within the past month (September 30th to October 31st, 2023). To isolate focal users' engagement with others' content, we excluded videos uploaded by focal users, as focal users may exhibit distinct engagement patterns. Additionally, we restricted the sample to videos with at least one comment view from both groups to ensure comparability. In total, users in the treatment group consumed comments on 684,684 videos, while those in the control group did so on 684,310 videos. Our final dataset included 12,356,563 user-video-level observations in the treatment period, restricted to videos where users viewed at least one comment and unconditional on whether they posted any comment. The dataset included comment engagement outcomes, user characteristics, and video attributes. Table 1 presents the summary of the variables used in our analyses. Our independent variable is the treatment group dummy ($Treat_i$), coded as 1 if user i was assigned to the treatment group.

Our dependent variables are two user engagement metrics: $CommentDuration_{ij}$, which captures comment consumption measured by user i's browsing duration (in minutes) on the comments for video j, and $PostComment_{ij}$, which indicates comment production incidence as a binary indicator of whether user i posted a comment on video j.[13] $Male_i$ is our moderator variable, coded as 1 if comment user i is male.[14]

To account for various potential confounding factors, we include a set of control variables (denoted as $Controls_{ij}$ in our econometrics models) capturing user, video producer, and video attributes as follows. First, we control for the timing of comment consumption by including the number of days between the posting date of video j and the date the user i browsed its comments. Second, we account for user popularity, experience, and activity by including the user's followers

[12] Our main analyses focus on users' comment consumption behavior and are thus conditional on users having browsed the comment section.

[13] In our data sample, users who commented posted an average of 1.2 comments per video. Using a dummy variable ($PostComment_{ij}$) to indicate whether a user commented on a video should yield similar results to using the total number of comments. In our robustness tests (detailed in Section 6), we used the number of comments posted ($CommentNum_{ij}$) as an alternative measure, and the results are qualitatively consistent.

[14] Platform A determines the gender of a user by combining the self-reported gender with viewing behavior. This approach helps reduce misclassification when a user's self-reported gender does not align with the observed behaviors.

Table 1 Variable Definition

Variables	Description
$Treat_i$	Coded as 1 if user i was assigned to the treatment group, or else as 0.
$CommentDuration_{ij}$	Browsing duration (in minutes) by user i on the comments for video j.
$PostComment_{ij}$	Coded as 1 if user i has posted any comment for video j, or else as 0.
$Male_i$	Coded as 1 if user i is male, or else as 0.
$PostDate_{ij}$	Number of days between the video posting date and the date on which the user viewed the comments for video j.
$Follower_i$	Number of users that follow user i.
$Experience_i$	Tenure of user i (in years) on the platform.
$VideoUpload_i$	Number of videos user i uploaded during the last 7 days.
$Region_i$	Location dummies for user i, including the northern and southern regions of the focal country, and the overseas region.
$AgeSegment_i$	Dummy variables indicating user i's age segments, which include 0-12, 12-17, 18-23, 24-30, 31-40, 41-59, and 60+ years old.
$ProducerMale_j$	Coded as 1 if the producer of video j is male, or else as 0.
$ProducerFollower_j$	Number of users that follow the producer of video j.
$ProducerExperience_j$	Tenure of producer of video j (in years) on the platform.
$VideoDuration_j$	Duration of video j (in minutes).
$ContentType_j$	Coded as 1 if video j is composed of videos (i.e., not images), or else as 0.
$Category_j$	First-level category dummies for video j (e.g., Beauty, Games).

Notes: All variables are coded as described in the table. Video metadata variables are collected from the platform's system logs.

count, platform tenure (in years), and the number of videos produced in the past seven days. We also control for demographic and geographic variations using age-segment and location dummies. Third, we include video producer attributes, such as gender, tenure, and follower count, to control for differences in video producer popularity and experience. Fourth, we include video characteristics, including duration (in minutes) and a binary indicator of whether the producer manually sets a video cover, to account for producer effort and video quality. Fifth, we include video category dummies to capture differences in video categories. Table 2 presents the summary statistics of our focal variables. A correlation matrix of the variables is shown in Table B.1 of Online Appendix B.

3.4. Randomization Check

To verify the validity of randomization for treatment assignment, we compared treatment users (N = 191,420) and control users (N = 191,725) on their pre-treatment comment engagement outcomes, user characteristics, and attributes of the videos these users engaged with.[15] Pairwise t-tests results presented in Table 3 show no significant differences between treatment and control groups on these observable attributes. These results confirm that the treatment and control users in our sample were comparable, suggesting that any difference between conditions after the experiment started should be attributed to our experimental manipulation, i.e., whether the first-level comments were ranked by a gender-aware curation algorithm.

[15] We also conducted a randomization check using data from one week prior to the experiment. The results, as reported in Table C.1 of Online Appendix C, are qualitatively consistent.

Table 2 Summary Statistics of Focal Variables

Variables	Mean	SD	Min	Max
Treat	0.501	0.500	0	1
PostComment	0.024	0.153	0	1
Male	0.454	0.498	0	1
PostDate	3.899	6.216	0	30
Experience	4.343	2.467	0	12.220
VideoUpload	1.276	4.276	0	1,398
ProducerMale	0.532	0.499	0	1
ProducerExperience	3.581	2.605	0	12.260
VideoDuration	1.172	1.890	0.018	300.500
ContentType	0.892	0.311	0	1

Notes: All variables are calculated based on video-level and user-level data. SD stands for standard deviation, and Min and Max represent the minimum and maximum values observed for each variable. To protect Platform A's sensitive information, the summary statistics for *CommentDuration*, *Follower*, and *ProducerFollower* are not displayed.

Table 3 Randomization Check Results

Variables	Full Sample (1)	Male Users (2)	Female Users (3)
CommentDuration	0.796	0.745	0.112
PostComment	0.853	0.209	0.348
Male	0.458	-	-
PostDate	0.694	0.516	0.924
Follower	0.247	0.164	0.905
Experience	0.568	0.678	0.698
VideoUpload	0.271	0.846	0.211
ProducerMale	0.806	0.270	0.251
ProducerFollower	0.555	0.408	0.963
ProducerExperience	0.434	0.926	0.319
VideoDuration	0.704	0.227	0.135
ContentType	0.823	0.320	0.209

Notes: Values represent p-values from t-tests comparing the treatment and control groups. The first column reports results for the full sample, while the second and third columns report p-values for male and female users separately. A dash (-) indicates the variable is not applicable for the subgroup.

3.5. Econometric Models

Our unit of analysis was at the user-video level to capture changes in comment engagement outcomes for each video where a user had clicked on the comment section. We used an ordinary least squares (OLS) regression with robust standard errors as follows:

$$DV1_{ij} = \beta_0 + \beta_1 Treat_i + \beta_2 Controls_{ij} + \gamma_j + e_{ij} \tag{1}$$

where $DV1_{ij}$ refers to our two user engagement metrics, $CommentDuration_{ij}$ and $PostComment_{ij}$. $Treat_i$ is a binary indicator equal to 1 if the user i was in the treatment group. $Controls_{ij}$ include all prior-mentioned control variables[16]. We also include $Male_i$, a dummy indicating whether user i is of the male gender, to control for users' gender. γ_j captures the video fixed effects, and e_{ij} is the error term. To adhere to data confidentiality requirements, $CommentDuration_{ij}$ was standardized. Highly-skewed control variables were log-transformed with a one-unit increment to account for zero values, following the semi-log approach in Cole and Sokolyk (2018). β_1 is the focal model coefficient of interest. To examine whether gender-aware curation yielded heterogeneous outcomes for male and female users, we also specify a moderating effect model where an interaction term between $Treat_i$ and $Male_i$ is introduced into Equation (1) as:

$$DV1_{ij} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 Male_i + \alpha_3 Treat_i \times Male_i + \alpha_4 Controls_{ij} + \gamma_j + \varepsilon_{ij} \tag{2}$$

where ε_{ij} is the error term. α_3 is the other focal model coefficient of interest.

4. Impact of Gender-Aware Curation Algorithm

We first estimated the effect of gender-aware curation algorithms on comment consumption and examined whether these effects differed between female and male users. Next, we analyzed the impact on comment production outcomes and the heterogeneous effects across different genders.

4.1. Comment Consumption

We first looked at whether adding gender data to the curation algorithms helped increase the time users spent reading comments. We estimated Equation (1), where $DV1_{ij}$ denotes $CommentDuration_{ij}$ which tracks how long a user spends in the comment section of a video. The results, shown in Column (1) of Table 4 ($\beta_1 = 0.005$, p-value < 0.01), suggest a clear improvement. Users who were shown comments ranked by the gender-aware algorithm spent about 1.0% more time in the comment section compared to those shown comments ranked by a gender-neutral algorithm.[17] Given that our sample covers nearly 1% of total platform users who collectively engage with user comments for around 2 million minutes in the comment section, these findings translate into an increase of roughly 3 million more minutes spent reading comments. This finding echoes prior research showing that algorithmic interventions (Chen and Chan 2024) and platform-developed artifacts can significantly shape user behavior and engagement on digital platforms (Yang et al. 2020, Jeong et al. 2022).

To further understand the increase in comment consumption, we employed the number of comments a user viewed ($CommentView_{ij}$) as an alternative outcome variable. Using the same

[16] $PostDate_{ij}$ and video attributes are absorbed by video fixed effects and their coefficients are not separately identified.

[17] A 1% relative effect size is often considered meaningful in large-scale online field experiments, where even small improvements can translate into practical impact (Kohavi et al. 2020).

Table 4 Impact of Gender-Aware Curation Algorithm on Comment Consumption

Variables	*CommentDuration* (1)	*CommentView* (2)	*Level1CommentView* (3)	*Level2CommentView* (4)
Treat	**0.005*** **	**0.005*** **	**0.0003**	**0.010*** **
	(0.001)	(0.001)	(0.001)	(0.001)
Male	-0.005***	-0.004***	-0.011***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Relative effect size	1.0%	0.7%	0.1%	2.3%
Controls	Yes	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes	Yes
Observations	12,356,563	12,356,563	12,356,563	12,356,563
R-squared	0.139	0.128	0.127	0.156

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors. *CommentDuration*, *CommentView*, *Level1CommentView*, and *Level2CommentView* are standardized for data confidentiality.

specification Equation (1), we find that the number of comments viewed increased by 0.7% in the treatment group, as shown in Column (2) of Table 4. These results are qualitatively consistent with the effect observed for comment duration. Given the two-level comment structure on Platform A, we next decomposed the total number of comment views ($CommentView_{ij}$) into the first- ($Level1CommentView_{ij}$) and second-level comment views ($Level2CommentView_{ij}$). By re-estimating Equation (1) for each level, we find that the gender-aware curation algorithm increased first-level comment views by 0.1% (not statistically significant) and second-level comment views by 2.3%, as shown in Columns (3) and (4) of Table 4. These results suggest that although the gender-aware curation algorithm ranked first-level comments, it indirectly boosted second-level comment consumption. Users, after browsing first-level comments they find interesting, become more inclined to browse the second-level replies. Collectively, these findings demonstrate that compared to a gender-neutral algorithm, a gender-aware curation algorithm leads to more active engagement with both levels of comments. These results help address an ongoing debate in data anonymization research (Kosinski et al. 2013) by showing that gender data can offer incremental value beyond behavioral data alone. The findings also provide empirical support for the economic relevance of demographic data (Sun et al. 2024).

Next, we examined whether the effect of the gender-aware algorithm differed between male and female users. Using the regression specification in Equation (2), we find that the interaction term was not significant in Table 5 ($\alpha_3 = 0.0005$, p-value > 0.1). This suggests that incorporating gender data into the curation algorithm equally boosted comment consumption across genders.

4.2. Comment Production

Beyond comment consumption, we also asked how gender-aware curation algorithms influenced comment production. To answer this question, we estimated Equation (1) where $DV1_{ij}$ denotes $PostComment_{ij}$, a binary indicator of whether user i commented on video j. The insignificant

Table 5 Heterogeneous Impact of Gender-Aware Curation Algorithm on Comment Consumption

Variables	*CommentDuration*
	(1)
Treat	0.005***
	(0.001)
Male	-0.006***
	(0.001)
Treat × *Male*	**0.0005**
	(0.001)
Controls	Yes
Video fixed effects	Yes
Observations	12,356,563
R-squared	0.139

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors. *CommentDuration* is standardized for data confidentiality.

Table 6 Impact of Gender-Aware Curation Algorithm on Comment Production

Variables	*PostComment*
	(1)
Treat	**-0.0001**
	(0.0001)
Controls	Yes
Video fixed effects	Yes
Observations	12,356,563
R-squared	0.201

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors.

coefficient of *Treat* (β_1 = -0.0001, p-value > 0.1) in Table 6 suggests that the gender-aware curation algorithm did not significantly affect the overall production of comments.

We next investigated the heterogeneous effects of gender-aware curation algorithms on comment production across genders. To this end, we estimated Equation (2) with $PostComment_{ij}$ as the dependent variable. As shown in Column (1) of Table 7, the coefficient for *Treat* is positive and significant (α_1 = 0.0006, p-value < 0.01), indicating that female users in the treatment group were 2.6% more likely to comment than those exposed to the gender-neutral algorithm.[18] In contrast, the interaction term between *Treat* and *Male* is negative and significant (α_3 = -0.0015, p-value < 0.01), resulting in a negative net effect for male users (0.0006 - 0.0015 = -0.0009). Specifically, treated male users were 3.5% less likely to produce comments compared to the control group[19]. These asymmetric effects across genders likely explain the nonsignificant overall treatment effect observed in Table 6. By contrast, recall from Table 5 that no heterogeneous treatment effects emerged for

[18] This is calculated as: 0.0006 / 0.023 = 2.6%.

[19] This is calculated as: (0.0006 - 0.0015) / 0.026 = -3.5%.

comment consumption. This is plausible because the reduced uniqueness-driven motivation of male users operates mainly on the production side, where self-presentation is a key driver (Wasko and Faraj 2005, Oh et al. 2023). In consumption, however, both male and female users generally enjoy viewing content that matches their preferences, so the treatment effects are similar for different genders (Bakshy et al. 2015). These distinctions underscore the importance of examining user engagement separately across its consumption and production dimensions.

Table 7 Heterogeneous Impact of Gender-Aware Curation Algorithm on Comment Production

Variables	*PostComment* (1)
Treat	**0.0006*****
	(0.0001)
Male	0.006***
	(0.0001)
Treat × *Male*	**-0.0015*****
	(0.0002)
Control baseline (mean)	0.024
Control baseline (male)	0.026
Control baseline (female)	0.023
Relative effect size (male)	-3.5%
Relative effect size (female)	2.6%
Controls	Yes
Video fixed effects	Yes
Observations	12,356,563
R-squared	0.201

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors.

These findings are practically important. Previous research (Peng et al. 2025) has documented that women are underrepresented in online contributions, often remaining silent due to lower confidence or a weaker sense of belonging (Wardhaugh and Fuller 2021).[20] The positive coefficient of *Male* in Column (1) of Table 7 confirms the presence of a gender gap, with male users being 0.006 units or about 25%[21] more likely to comment than female users.

To better understand this heterogeneous effect, we decomposed $PostComment_{ij}$ into two binary indicators: one for whether the user posted a first-level comment ($Level1PostComment_{ij}$), and one for whether the user posted a second-level comment ($Level2PostComment_{ij}$). We then re-estimated Equation (2) using each as the dependent variable. As shown in Table 8, the results align well with those in Table 7. Among female users, the gender-aware algorithm increased the likelihood

[20] For example, women contribute less to Wikipedia editing. See `https://en.wikipedia.org/wiki/Gender_bias_on_Wikipedia`.

[21] This is calculated as: 0.006 / 0.024 = 25%.

of posting first-level comments by 2.4%[22] and second-level comments by 4.3%[23]. In contrast, male users in the treatment group showed a 3.0%[24] decline in the first-level comments and a 1.4%[25] decline in the second-level comments.

Table 8 Impact of Gender-Aware Curation Algorithm on Comment Production Across Comment Types

Variables	*Level1PostComment* (1)	*Level2PostComment* (2)
Treat	**0.0004*****	**0.0003*****
	(0.0001)	(0.0001)
Male	0.004***	0.002***
	(0.0001)	(0.0001)
Treat × *Male*	**-0.001*****	**-0.0004*****
	(0.0001)	(0.0001)
Control baseline (male)	0.020	0.007
Control baseline (female)	0.017	0.007
Relative effect size (male)	-3.0%	-1.4%
Relative effect size (female)	2.4%	4.3%
Controls	Yes	Yes
Video fixed effects	Yes	Yes
Observations	12,356,563	12,356,563
R-squared	0.203	0.099

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors.

5. Mechanism Analysis

5.1. Comment Consumption

So far, we have shown that gender-aware curation algorithms, compared to gender-neutral ones, can boost users' comment consumption. We next investigated the underlying mechanism and proposed that this effect was driven by improved personalization of comments shown. Prior research on curation algorithms (Adomavicius et al. 2008, Sun et al. 2024) suggests that personalization directs users toward more niche and personally relevant content, which can increase user engagement and satisfaction. If gender-aware curation enhanced personalization, we should observe that treated users were recommended fewer popular comments. To test this, we examined the average popularity of first-level comments browsed by users.[26] Popularity is measured by the log-transformed average number of views ($PopularityView_{ij}$), likes ($PopularityLike_{ij}$), and replies ($PopularityReply_{ij}$) received by all the first-level comments browsed by user i on video j. We then

[22] This is calculated as: 0.0004 / 0.017 = 2.4%.

[23] This is calculated as: 0.0003 / 0.007 = 4.3%.

[24] This is calculated as: (-0.001 + 0.0004) / 0.020 = -3.0%.

[25] This is calculated as: (-0.0004 + 0.0003) / 0.007 = -1.4%.

[26] This is because the curation algorithm on Platform A ranks first-level comments.

re-estimated Equation (1) using these three metrics as alternative dependent variables. As shown in Table 9, treated users were exposed to less popular (i.e., more niche) comments. Specifically, the average views, likes, and replies of browsed first-level comments decreased by 4.0%, 3.7%, and 2.0%, respectively. These findings support our proposition that gender-aware curation improved personalization, thereby increasing comment consumption.

Table 9 Effect of Gender-Aware Curation on Recommended First-Level Comment Popularity

Variables	*PopularityView* (1)	*PopularityLike* (2)	*PopularityReply* (3)
Treat	**-0.041***** (0.0002)	**-0.038***** (0.0003)	**-0.020***** (0.0003)
Relative effect size	-4.0%	-3.7%	-2.0%
Controls	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes
Observations	12,356,563	12,356,563	12,356,563
R-squared	0.953	0.952	0.929

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Values in parentheses are robust standard errors. *PopularityView*, *PopularityLike*, and *PopularityReply* are log-transformed.

5.2. Comment Production

We next examined why the gender-aware curation algorithm produced asymmetric effects on comment production, increasing the likelihood of commenting among female viewers while decreasing it among male viewers. Prior research has shown that men value self-presentation and are more motivated to engage when they can assert unique perspectives (Cross et al. 2011, Coates 2015), whereas women are motivated by a desire for social connection and shared understanding (Wardhaugh and Fuller 2021, Griskevicius et al. 2006, Venkatesh and Morris 2000). We therefore posit that gender-aware curation discourages male commenting by reducing opportunities for self-presentation, while it encourages female commenting through enhanced perceived belonging.

5.2.1. The Self-Presentation Effect for Male Users

We conducted two analyses to examine whether the observed decline in male comment production under gender-aware curation can be attributed to self-presentation motives. First, prior research suggests that audience size shapes individuals' willingness to share self-presenting content (Oh et al. 2023, Barasch and Berger 2014). If self-presentation motives underlie the observed decline, male users with a larger follower base in the treatment group should exhibit stronger reductions in comment production. To test this proposition, we extend Equation (2) by adding interaction terms of $ProducerFollower_{ij}$ with both $Treat_i$ and $Male_i$, as well as a three-way interaction between $Treat_i$, $Male_i$, and $ProducerFollower_{ij}$ specified as:

$$\begin{aligned} DV1_{ij} = {} & \lambda_0 + \lambda_1\, Treat_i + \lambda_2\, Male_i + \lambda_3\, ProducerFollower_{ij} + \lambda_4\, Treat_i \times Male_i \\ & + \lambda_5\, Treat_i \times ProducerFollower_{ij} + \lambda_6\, Male_i \times ProducerFollower_{ij} \\ & + \lambda_7\, Treat_i \times Male_i \times ProducerFollower_{ij} + \lambda_8\, Controls_{ij} + \gamma_j + \varepsilon_{ij}, \end{aligned} \tag{3}$$

where λ_7 is the focal model coefficient of interest. Columns (1) and (2) of Table 10 report the estimation results when $DV1_{ij}$ refers to *CommentDuration* and *PostComment*, respectively. The coefficient of the three-way interaction term is not significant in Column (1). In contrast, the coefficient is significantly negative (p-value < 0.01) in Column (2), indicating that male users with larger follower bases reduce their comment production more strongly under gender-aware curation. This finding provides empirical support for the proposed self-presentation motives for male users, and further demonstrates that the motivations underlying consumption and production indeed differ (Phang et al. 2015, Oh et al. 2023).

Table 10 Heterogeneous Impact of Gender-Aware Curation Algorithm by Number of Followers

Variables	*CommentDuration* (1)	*PostComment* (2)
Treat	0.004*	-0.001***
	(0.002)	(0.0002)
Male	-0.012***	0.002***
	(0.002)	(0.0003)
ProducerFollower	-0.003***	0.003***
	(0.0003)	(0.00005)
Treat × *Male*	0.003	0.003***
	(0.003)	(0.004)
Treat × *ProducerFollower*	0.0002	0.0004***
	(0.0004)	(0.0001)
Male × *ProducerFollower*	0.001***	0.001***
	(0.0004)	(0.0001)
Treat × *Male* × *ProducerFollower*	**-0.001**	**-0.001***
	(0.001)	(0.0001)
Controls	Yes	Yes
Video fixed effects	Yes	Yes
Observations	12,356,563	12,356,563
R-squared	0.139	0.201

Notes: ***p<0.01; **p<0.05; *p<0.1. Values in parentheses are robust standard errors. *CommentDuration* is standardized for data confidentiality.

Second, the desire for self-presentation is often manifested in higher information or content quality that users produce (Barasch and Berger 2014). To test whether male users exhibit such tendencies more strongly than female users, we examined the information quality of comments posted by control group users.[27] This analysis includes only observations where user i posted at

[27] Analyses of control group data allow us to identify baseline gender differences in comment quality, as the control group was not affected by the experimental manipulation.

least one comment for video j. If male users are indeed driven by self-presentation incentives, comments of male users should display higher information quality than those of female users. We operationalized information quality through multiple measures as follows.

First, we calculated the proportion of simple interaction comments posted by user i on video j ($SimpleInteraction_{ij}$). These comments are usually characterized by low-effort and generic content (e.g., "agree" or "reasonable"), offering limited informational value. Comments were classified as simple or complicated using Platform A's proprietary classification algorithm[28]. Second, we used language richness, a widely-used concept in linguistic studies, as a proxy for information quality (Qiao et al. 2020). Specifically, we considered three key linguistic metrics: (1) lexical density ($LexicalDensity_{ij}$), which captures the proportion of content words (such as nouns, verbs, and adjectives) relative to the total number of words, with higher values reflecting stronger informational content; (2) lexical variation ($LexicalVariation_{ij}$), which measures the ratio of unique words to total words, representing vocabulary diversity and linguistic complexity; and (3) entropy ($Entropy_{ij}$), which quantifies text unpredictability and is computed as:

$$Entropy = -\sum_{k=1}^{n} P_k \log P_k \tag{4}$$

where P_k represents the probability of each unique word k in the comments. Entropy reflects linguistic variability, with higher values indicating greater unpredictability in word usage. Third, we used the average comment length ($Length_{ij}$) as a proxy for information quality, as longer sentences often reflect more complex reasoning and higher cognitive effort.

To analyze these textual attributes, we employed an OLS regression with robust standard errors, specified as:

$$DV2_{ij} = \theta_0 + \theta_1 Male_i + \theta_2 Controls_{ij} + \gamma_j + \eta_{ij} \tag{5}$$

where $DV2_{ij}$ represents the textual attributes of comments posted by commenter i of video j, γ_j denotes video fixed effects, and η_{ij} is the error term. Other variables are consistent with those used in Equation (1). The focal coefficient of interest is θ_1. Regression results are presented in Columns (1)-(5) of Table 11.

The results support our proposition. Column (1) shows that male users were 1.2% less likely to post simple interaction comments than female users. In Columns (2)-(4), the positive coefficients of *Male* indicate that comments posted by male users exhibited richer informational content, with lexical density, lexical variation, and lexical entropy values exceeding those of female users by 0.017,

[28] The classification algorithm was trained on a large dataset of comments from Platform A, with part of the data labeled by human annotators and the rest automatically labeled by GPT-4. The model distinguishes multiple comment categories and achieves around 90% accuracy overall.

0.022, and 0.089 units (5.0%,[29] 5.4%, and 9.1%, respectively). Additionally, results in Column (5) show that male users wrote longer comments, with an average length of 5.2%[30] greater than that of female users. Taken together, these findings consistently suggest that male users tend to produce comments with higher informational value. This supports the argument that male users are more likely to comment when they can contribute novel or distinctive content (Coates 2015), consistent with male users' stronger self-presentation motivation (Oh et al. 2023).

Table 11 Gender Differences in Comment Information Quality (Control Group)

Variables	*SimpleInteraction* (1)	*LexicalDensity* (2)	*LexicalVariation* (3)	*Entropy* (4)	*Length* (5)
Male	**-0.012*****	**0.017*****	**0.022*****	**0.089*****	**0.051*****
	(0.003)	(0.003)	(0.003)	(0.010)	(0.007)
Control baseline (female)	0.210	0.340	0.409	0.974	4.494
Relative effect size	5.7%	5.0%	5.4%	9.1%	5.2%
Controls	Yes	Yes	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	91,148	91,148	91,148	91,148	91,148
R-squared	0.370	0.627	0.686	0.652	0.700

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors. *Length* is log-transformed. This table is based on observations in control group where user i posted at least one comment on video j (147,834 observations), as comment attributes are undefined otherwise. Videos with only a single commenting user provide no within-video variation and are absorbed by video fixed effects, leaving 91,148 observations for estimation.

5.2.2. The Consensus-Seeking Effect for Female Users

Next, we test the proposition that female users place a greater value on consensus and community-building (Cross et al. 2011, Coates 2015). As an alternative dependent variable in Equation (5), we first examined the proportion of comments that explicitly tag friends, a behavior often linked to seeking agreement, reinforcing social bonds, and fostering connection (Huang et al. 2017). Column (1) of Table 12 shows that male users were 28.1%[31] less likely to tag friends in their comments, supporting the proposition that women engage more actively in community-building behaviors. This aligns with prior research suggesting that women tend to share ideas within close networks rather than through public discussions (Brescoll 2011, Coates 2015).

We then turn to analyze the emotional tone of the comments, with the expectation that consensus-oriented communication should correlate with affirmative or praiseworthy language (Leaper and Ayres 2007). To test this, we first analyzed the share of comments classified as positive[32]. As shown in Column (2) of Table 12, male users posted a 5.7% lower share of

[29] This is calculated as: 0.017 / 0.340 = 5.0%. The same calculation method is applied subsequently.

[30] This is calculated as: exp(0.051) - 1 = 5.2%.

[31] This is calculated as: (-0.047) / 0.167 = -28.1%.

[32] This classification is provided by Platform A.

positive comments than female users. We further examined neutral and negative comments. Columns (3) and (4) show that male users posted 1.0% more neutral comments and 12.1% more negative comments than female users. These results consistently suggest that female users tend to communicate in a more positive and affirming tone, while male users are relatively more likely to adopt neutral or critical styles. This aligns with prior research showing that women are more inclined toward cooperative, consensus-building dialogue and tend to avoid adversarial exchanges (Griskevicius et al. 2006, Coates 2015, Leaper and Ayres 2007). Combined with our earlier finding that gender-aware curation algorithms increased commenting incidence among female users but decreased it among male users, the gender differences in comment sentiment suggest that such curation enhances comment diversity by adding more emotionally affirming and community-oriented perspectives.

Furthermore, we examined gender differences in comment-liking behavior, using a binary indicator of whether a user liked any comments on a video ($LikeComment_{ij}$). Based on control group data, Column (5) shows that male users were 25.9% less likely to like others' comments than female users. Interestingly, this finding contrasts with our earlier result that male users were 25% more likely to post comments. This may be because commenting often reflects a desire to express personal opinions, but liking others' comments is often seen as a form of communal affirmation, which signals agreement and a willingness to engage in consensus-oriented interaction. This is consistent with prior research suggesting that men prioritize distinctiveness and status-seeking over community building (Cross et al. 2011), and with evidence that women are more likely to engage in prosocial behaviors than men (Soutschek et al. 2017).

Table 12 Gender Differences in Consensus-Seeking Behaviors (Control Group)

Variables	*TagFriend* (1)	*Positive* (2)	*Neutral* (3)	*Negative* (4)	*LikeComment* (5)
Male	**-0.047*****	**-0.012*****	**0.008****	**0.004****	**-0.007*****
	(0.003)	(0.003)	(0.003)	(0.002)	(0.0002)
Control baseline (female)	0.167	0.209	0.758	0.033	0.027
Relative effect size	-28.1%	-5.7%	1.0%	12.1%	-25.9%
Controls	Yes	Yes	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	91,148	91,148	91,148	91,148	5,920,955
R-squared	0.547	0.366	0.406	0.337	0.086

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors. Columns (1) to (4) are based on observations in control group where user i posted at least one comment on video j (147,834 observations). Videos with only a single commenting user provide no within-video variation and are absorbed by video fixed effects, resulting in 91,148 observations for estimation. Column (5) uses the full control group sample unconditional on posting any comments. Videos with only one observation are likewise absorbed by video fixed effects, leaving 5,920,955 observations.

As an alternative test of the consensus-seeking effect, we examined whether the positive impact of gender-aware curation on females' comment production was weaker in female-dominated categories than in male-dominated or gender-neutral ones. The rationale is that in female-dominated categories where community bonds and like-minded perspectives are already strong, gender-aware curation should provide less incremental value. Consistent with this expectation, we find that the positive effect of gender-aware curation on females' comment production is smaller in female-dominated categories, while no such heterogeneity is observed for male users (see Table D.2 in Online Appendix D).

6. Additional Analyses and Robustness Tests

This section is devoted to further discussions and analyses to supplement our main results. The detailed regression results are relegated to Online Appendix E.

6.1. Effects of Gender-Aware Curation Algorithm on User Engagement Over Time

As our main analyses rely on data from the first treatment day, one potential concern is that the observed outcomes may be driven by short-term novelty effects. To address this, we extended the analyses to a three-week period (October 31st to November 19th, 2023) to assess the treatment effects on user engagement in terms of comment consumption and comment production over time. We aggregated variables at the user-day level and replicated the main analyses in Section 4. Table E.1 and Table E.2 show qualitatively consistent results. These findings indicate that gender-aware curation fosters long-term engagement rather than short-term novelty-driven responses, reinforcing its potential in shaping sustained user engagement via comments.

6.2. Analyses of Comment Exposure

To assess how the gender-aware curation algorithm altered the comments shown to users, we analyzed two dimensions of comment exposure: (a) gender homophily (whether users were shown more comments posted by users of the same gender), and (b) controversialness (whether users were exposed to more debatable or contentious comments). We find no evidence that the treatment increased gender homophily, as both male and female users were less exposed to female-authored comments. However, male users were more likely to be shown controversial comments, consistent with their greater preference for debatable content (Table E.3).

6.3. Effects of Gender-Aware Curation on Textual Features of User Comments

In Section 4.2, we show that the gender-aware curation algorithm affected users' likelihood of commenting, with effects varying by gender. A natural follow-up question is whether gender-aware curation also changed how users commented, i.e., the characteristics of the comments users posted. To investigate this, we used the comment features introduced in Section 5.1 (e.g., lexical richness, lexical density, and length) as outcome variables in Equation (2). As shown in Table E.4, the coefficients of *Treat* and its interaction term with *Male* are statistically insignificant across all measures, suggesting that the treatment shaped the likelihood of posting comments but did not change the style or informational quality of comments.

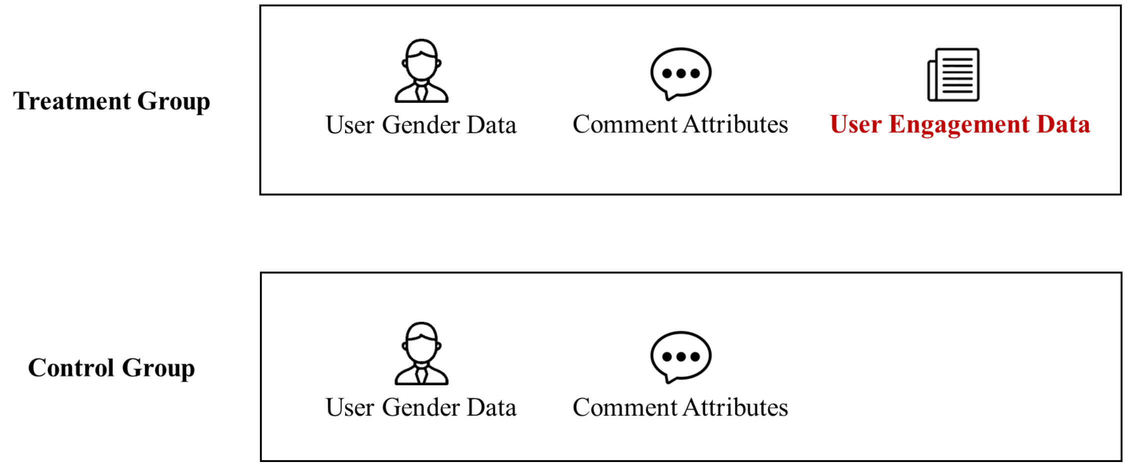


Figure 3 Data Inputs for Treatment and Control Algorithms in the Follow-Up Experiment

6.4. Gender-Based Heterogeneity with Alternative Curation Data Inputs

To examine whether the gender-based heterogeneity in comment production observed in Section 4.2 is specific to gender-aware curation or can also be observed when other types of user data are incorporated, we conducted a follow-up field experiment in June 2024. In this additional experiment, the comment curation algorithm for the control group incorporated two types of features: (1) users' gender data and (2) comment attributes. For the treatment group, the algorithm used the same features as in the control group and additionally incorporated users' historical engagement with comments (e.g., browsing duration, likes, and replies)(see Figure 3). We re-estimated the econometric models in Section 3.5 based on the follow-up experimental data. As shown in Table E.8, we find no significant treatment effect on comment production or heterogeneous effects across genders. This is plausibly because the engagement data in our context primarily capture consumption rather than production behaviors and therefore did not reproduce the heterogeneous production effects observed under gender-aware curation in Section 4.2.

6.5. Alternative Operationalization of Variables

We employed alternative operationalizations of comment production to ensure the robustness of our results. Specifically, we replaced the binary indicator of whether a user commented on a video with the total number of comments user i posted for video j ($CommentNum_{ij}$), a count measure that captures both the occurrence and intensity of commenting. Results in Table E.9 are consistent with our main findings.

7. Conclusions and Discussion

7.1. Discussion of Findings

Our study, which examines the causal impact of a gender-aware curation algorithm on user engagement in the comment sections of a UGC platform for short videos, reveals several key findings. First, we find that incorporating gender data into curation algorithms significantly enhances such engagement by improving personalization. Specifically, our gender-aware comment algorithm increased users' browsing time in the comment section by 1.0%, compared to a gender-neutral approach. Moreover, we did not observe heterogeneous treatment effects on comment consumption across genders. This is plausible because both men and women generally enjoy viewing content that matches their preferences (Bakshy et al. 2015), so gender-aware curation boosted browsing behavior similarly across groups. This finding is consistent with recent IS research showing that adding personal attribute data to curation systems can improve engagement (e.g., Sun et al. 2024). However, unlike prior studies that considered personal attribute data in aggregate, our study isolates gender, a distinct and behaviorally salient personal attribute, and shows that incorporating it into the curation algorithm can boost user engagement. This addresses an open question in the data anonymization literature (Kosinski et al. 2013, Xu and Zhang 2022) about whether explicit gender data adds value beyond gender-related preferences inferred from engagement data, and complements research on gender-based behavioral differences (Varma et al. 2023, Venkatesh and Morris 2000) by confirming systematic differences between men and women in engagement patterns and content preferences.

Second, beyond comment consumption, we find that the gender-aware algorithm influences comment production, but in a heterogeneous way: male users were 3.5% less likely to comment, whereas female users were 2.6% more likely to do so. Mechanism analyses suggest that this asymmetry effect was driven by differing commenting motives across genders: male users aim to assert individual perspectives for self-presentation purposes, whereas female users are prompted by a desire for community belonging. These findings extend prior research on curation algorithms (Zhou et al. 2025, Bakshy et al. 2015), which has primarily measured engagement through consumption, by showing that curation can shape engagement along the production dimension. Because UGC platforms rely on both consumption and production to sustain success, recognizing production as a key dimension of engagement is essential. Our results also advance the research on user engagement on UGC platforms (Huang et al. 2019, Baek and Shore 2020) by highlighting the overlooked role of curation algorithms in influencing the production of engagement, and contribute to research on gender-based behavioral differences (Varma et al. 2023) by demonstrating how gender-specific motivations interact with algorithmic design to produce asymmetric engagement responses by gender.

Finally, because female users in our setting were 25% less likely to comment than males, the observed increase in female commenting under gender-aware curation carries important diversity implications. By encouraging engagement from an underrepresented group, gender-aware curation can enhance the gender diversity in online engagement. Moreover, because attributes of comments (e.g., emotional tone, friend tagging) systematically differ between male and female users, greater gender diversity also implies improved content diversity in online discussions. This complements emerging evidence in the IS field that IT can improve engagement among underrepresented populations (e.g., Jeong et al. (2022)) and suggests practical pathways for platforms seeking more inclusive and balanced engagement.

7.2. Theoretical Contributions

Our work offers several theoretical contributions as follows. First, we contribute to the curation algorithm research stream (Zhou et al. 2025, Sun et al. 2024) by presenting the first causal evidence that integrating gender data into these algorithms enhances user engagement on UGC platforms. This finding also enriches data anonymization research (Kosinski et al. 2013, Xu and Zhang 2022) by demonstrating that explicit gender information adds value beyond what engagement data captures. Moreover, we extend this literature by enriching the operationalization of user engagement, moving beyond consumption to incorporate production behaviors as a complementary dimension.

Second, we contribute to research on user engagement on UGC platforms (Xu et al. 2021, Huang et al. 2019) by highlighting the previously underexplored role of gender-aware curation algorithms in shaping user engagement on UGC platforms. More importantly, we take this literature one step further by revealing significant gender-based heterogeneity in the effects of gender-aware curation on user engagement. Specifically, we find no heterogeneous treatment effects on comment consumption. In contrast, for comment production, gender-aware curation increased participation among female users while reducing it among male users. These novel findings enrich curation algorithm research (Sun et al. 2024, Li and Tuzhilin 2024) by highlighting the need to incorporate user heterogeneity and motivation differences when developing and evaluating curation algorithms.

Third, we extend broad research on gender-based behavioral differences (Varma et al. 2023, Venkatesh and Morris 2000) by investigating how gender-aware curation algorithms shape user engagement. Although prior studies have established that men and women differ in engagement motivations, these studies have largely overlooked algorithmic influences. We take this literature one step further by demonstrating how gender-aware curation algorithms interact with gender-specific motivations to produce asymmetric comment production responses, increasing comment production among female users while reducing it among male users.

7.3. Practical Implications

Our findings offer several insights for platform operators of UGC platforms. First, we suggest platforms integrate users' gender data into curation algorithms to boost UGC consumption, as our findings show that the gender-aware approach extends users' browsing duration in comment sections, relative to gender-neutral approaches. This operational strategy can extend to broader

UGC contexts like video feeds and article recommendations, particularly where female and male users exhibit distinct content preferences. However, we also caution that this approach is more suited to enhancing personalization and may be less appropriate when platforms seek to foster exposure to a more diverse range of content. Moreover, we caution that prolonged user engagement can be associated with increased screen time, which in turn may lead to well-being issues and addiction risks (Fassi et al. 2025). This issue may be especially relevant when the platform's main content consumers are young children.

Second, our results indicate that gender-aware curation asymmetrically impacts comment production, with female users increasing comment production and male users decreasing theirs. To address the decline, platforms could implement gender-specific strategies, such as presenting male users with fewer like-minded opinions (e.g., comments that challenge typical male perspectives) to stimulate unique contributions while offering female users more ideologically consistent content (e.g., comments aligning with prevailing female views) to reinforce community bonds. Additionally, platforms could incorporate content production metrics or assign them greater weight in algorithm objectives to alleviate the decline in male users' content production (Wang et al. 2025).

Third, our mechanism analyses indicate that the consensus-seeking effect for female users is stronger in gender-neutral and male-dominated content categories. Consequently, in resource-constrained settings, platforms could prioritize applying gender-aware curation algorithms in these areas for female users. Moreover, because gender-aware curation may discourage content production for male users, platforms could adopt targeted curation strategies that consider both the user's gender and the nature of the content (e.g., category or topic). For example, gender-aware algorithms can be selectively applied for female users in male-dominated categories, while gender-neutral or more diverse recommendation strategies may be used for male users to avoid over-personalization and support continued contribution.

7.4. Limitations and Future Research

The limitations of our work open up interesting avenues for future research. First, our analyses focus on the comment sections of short-video platforms, which are mainly entertainment-oriented. Future research could extend this investigation to knowledge-oriented platforms, such as Q&A forums like Quora, Stack Overflow, or Zhihu, where the nature of user engagements may differ. In entertainment-driven environments, commenting behavior is often influenced by social engagement and low-cost participation, whereas in knowledge-sharing platforms, contributions require higher cognitive effort, domain expertise, and reputation. Understanding how gender-aware curation operates in these distinct settings could offer deeper insights into its broader applicability across different digital ecosystems. Second, although our study focuses on the impact of incorporating the gender of users, future work could explore whether incorporating the gender of both video producers and comment producers affects curation outcomes. Moreover, exploring the interplay between gender and other demographic or behavioral factors (such as age or content preferences) may further refine our understanding of content curation and inform industry practices.

References

Adomavicius, Gediminas, Zan Huang, Alexander Tuzhilin. 2008. Personalization and recommender systems. *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*. INFORMS, 55–107.

Agarwal, Ashish, Shun-Yang Lee, Andrew B. Whinston. 2025. The effect of popularity cues and peer endorsements on assertive social media ads. *Information Systems Research* **36**(2) 1167–1182. doi:10.1287/isre.2021.0606.

Baek, Jiye, Jesse Shore. 2020. Forum size and content contribution per person: A field experiment. *Management Science* **66**(12) 5906–5924.

Bakshy, Eytan, Solomon Messing, Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on facebook. *Science* **348**(6239) 1130–1132.

Barasch, Alixandra, Jonah Berger. 2014. Broadcasting and narrowcasting: How audience size affects what people share. *Journal of Marketing Research* **51**(3) 286–299.

Beaudoin, Michael F. 2002. Learning or lurking?: Tracking the "invisible" online student. *The Internet and Higher Education* **5**(2) 147–155.

Berman, Ron, Zsolt Katona. 2020. Curation algorithms and filter bubbles in social networks. *Marketing Science* **39**(2) 296–316.

Binns, Reuben, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, Nigel Shadbolt. 2018. 'it's reducing a human being to a percentage' perceptions of justice in algorithmic decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, Andrei Shleifer. 2019. Beliefs about gender. *American Economic Review* **109**(3) 739–773.

Brescoll, Victoria L. 2011. Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly* **56**(4) 622–641.

Bucher-Koenen, Tabea, Rob Alessie, Annamaria Lusardi, Maarten van Rooij. 2024. Fearless woman: Financial literacy, confidence, and stock market participation. *Management Science* .

Butler, Brian S. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research* **12**(4) 346–362.

Cao, Zike, Yingpeng Zhu, Gen Li, Liangfei Qiu. 2024. Consequences of information feed integration on user engagement and contribution: A natural experiment in an online knowledge-sharing community. *Information Systems Research* **35**(3) 1114–1136.

Chen, Zenan, Jason Chan. 2024. Large language model in creative work: The role of collaboration modality and user expertise. *Management Science* **70**(12) 9101–9117.

Cheng, Xiaoye, Hillol Bala, Mochen Yang. 2024. Engaging users on social media business pages: The roles of user comments and firm responses. *MIS Quarterly* **48**(2).

Coates, Jennifer. 2015. *Women, men and language: A sociolinguistic account of gender differences in language*. Routledge.

Coffman, Katherine Baldiga. 2014. Evidence on self-stereotyping and the contribution of ideas. *Quarterly Journal of Economics* **129**(4) 1625–1660.

Cole, Rebel A, Tatyana Sokolyk. 2018. Debt financing, survival, and growth of start-up firms. *Journal of Corporate Finance* **50** 609–625.

Cross, Susan E, Erin E Hardin, Berna Gercek-Swing. 2011. The what, how, why, and where of self-construal. *Personality and Social Psychology Review* **15**(2) 142–179.

Dou, Yifan, Marius F Niculescu, DJ Wu. 2013. Engineering optimal network effects via social media features and seeding in markets for digital goods and services. *Information Systems Research* **24**(1) 164–185.

Fassi, Luisa, Amanda M Ferguson, Andrew K Przybylski, Tamsin J Ford, Amy Orben. 2025. Social media use in adolescents with and without mental health conditions. *Nature human behaviour* 1–17.

Griskevicius, Vladas, Noah J Goldstein, Chad R Mortensen, Robert B Cialdini, Douglas T Kenrick. 2006. Going along versus going alone: when fundamental motives facilitate strategic (non) conformity. *Journal of Personality and Social Psychology* **91**(2) 281.

Gu, Zheyin, Xuying Zhao. 2024. Content length limit: how does it matter for a consumer-to-consumer media platform? *Information Systems Research* **35**(4) 1785–1801.

Helgeson, Vicki S. 1994. Relation of agency and communion to well-being: Evidence and potential explanations. *Psychological Bulletin* **116**(3) 412.

Hosanagar, Kartik, Daniel Fleder, Dokyun Lee, Andreas Buja. 2014. Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation. *Management Science* **60**(4) 805–823.

Huang, Ni, Gordon Burtch, Bin Gu, Yili Hong, Chen Liang, Kanliang Wang, Dongpu Fu, Bo Yang. 2019. Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science* **65**(1) 327–345.

Huang, Ni, Yili Hong, Gordon Burtch. 2017. Social network integration and user content generation. *MIS Quarterly* **41**(4) 1035–1058.

Jeong, Christina Y., Jason Chan, Yue Guo. 2022. The role of online socialization at the workplace: Impact on reducing gender disparity. *Academy of Management Annual Meeting Proceedings* .

Kane, Gerald C, Sam Ransbotham. 2016. Content as community regulator: The recursive relationship between consumption and contribution in open collaboration communities. *Organization Science* **27**(5) 1258–1274.

Karwatzki, Sabrina, Olga Dytynko, Manuel Trenz, Daniel Veit. 2017. Beyond the personalization–privacy paradox: Privacy valuation, transparency features, and service personalization. *Journal of Management Information Systems* **34**(2) 369–400.

Kelley, Stephanie, Anton Ovchinnikov, David R Hardoon, Adrienne Heinrich. 2022. Antidiscrimination laws, artificial intelligence, and gender bias: A case study in nonmortgage fintech lending. *Manufacturing Service Operations Management* **24**(6) 3039–3059.

Kohavi, Ron, Diane Tang, Ya Xu. 2020. *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press.

Kosinski, Michal, David Stillwell, Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* **110**(15) 5802–5805.

Lambrecht, Anja, Catherine Tucker. 2019. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science* **65**(7) 2966–2981.

Leaper, Campbell, Melanie M Ayres. 2007. A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review* **11**(4) 328–363.

Li, Pan, Alexander Tuzhilin. 2024. When variety seeking meets unexpectedness: Incorporating variety-seeking behaviors into design of unexpected recommender systems. *Information Systems Research* **35**(3) 1257–1273.

Lian, Jianxun, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, Guangzhong Sun. 2018. xdeepfm: Combining explicit and implicit feature interactions for recommender systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1754–1763.

Lount Jr, Robert B, Steffanie L Wilk. 2014. Working harder or hardly working? posting performance eliminates social loafing and promotes social laboring in workgroups. *Management Science* **60**(5) 1098–1106.

Ma, Meng, Ritu Agarwal. 2007. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research* **18**(1) 42–67.

Oh, Hyelim, Khim-Yong Goh, Tuan Q Phan. 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.

Peng, Hao, Misha Teplitskiy, Daniel M Romero, Emőke-Ágnes Horvát. 2025. The gender gap in scholarly self-promotion on social media. *Nature Communications* **16**(1) 5552.

Phang, Chee Wei, Atreyi Kankanhalli, Bernard CY Tan. 2015. What motivates contributors vs. lurkers? an investigation of online feedback forums. *Information Systems Research* **26**(4) 773–792.

Pinney, Christine, Amifa Raj, Alex Hanna, Michael D Ekstrand. 2023. Much ado about gender: Current practices and future recommendations for appropriate gender-aware information access. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 269–279.

Qian, Kun, Sanjay Jain. 2024. Digital content creation: An analysis of the impact of recommendation systems. *Management Science* **70**(12) 8668–8684.

Qiao, Dandan, Shun-Yang Lee, Andrew B Whinston, Qiang Wei. 2020. Financial incentives dampen altruism in online prosocial contributions: A study of online reviews. *Information Systems Research* **31**(4) 1361–1375.

Soutschek, Alexander, Christopher J Burke, Anjali Raja Beharelle, Robert Schreiber, Susanna C Weber, Iliana I Karipidis, Jolien Ten Velden, Bernd Weber, Helene Haker, Tobias Kalenscher, et al. 2017. The dopaminergic reward system underpins gender differences in social preferences. *Nature Human Behaviour* **1**(11) 819–827.

Sun, Tianshu, Zhe Yuan, Chunxiao Li, Kaifu Zhang, Jun Xu. 2024. The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy. *Management Science* **70**(4) 2645–2660.

Varma, Rajita, Raghu Bommaraju, Siddharth S Singh. 2023. Female chief marketing officers: When and why do their marketing decisions differ from their male counterparts'? *Journal of Marketing Research* **60**(6) 1154–1176.

Venkatesh, Viswanath, Michael G Morris. 2000. Why don't men ever stop to ask for directions? gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly* 115–139.

Wang, Jing, Gen Li, Kai-Lung Hui. 2022. Monetary incentives and knowledge spillover: Evidence from a natural experiment. *Management Science* **68**(5) 3549–3572.

Wang, Yuyan, Long Tao, Xian Xing Zhang. 2025. Recommending for a multi-sided marketplace: A multi-objective hierarchical approach. *Marketing Science* **44**(1) 1–29.

Wardhaugh, Ronald, Janet M Fuller. 2021. *An introduction to sociolinguistics*. John Wiley & Sons.

Warlop, Romain, Alessandro Lazaric, Jérémie Mary. 2018. Fighting boredom in recommender systems with linear reinforcement learning. *Advances in Neural Information Processing Systems* **31**.

Wasko, Molly McLure, Samer Faraj. 2005. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* 35–57.

Wong, Randy Yee Man, Christy MK Cheung, Bo Xiao, Jason Bennett Thatcher. 2021. Standing up or standing by: Understanding bystanders' proactive reporting responses to social media harassment. *Information Systems Research* **32**(2) 561–581.

Xu, Heng, Nan Zhang. 2022. Implications of data anonymization on the statistical evidence of disparity. *Management Science* **68**(4) 2600–2618.

Xu, Yuqian, Mor Armony, Anindya Ghose. 2021. The interplay between online reviews and physician demand: An empirical investigation. *Management Science* **67**(12) 7344–7361.

Yang, Lusi, Zhiyi Wang, Jungpil Hahn. 2020. Scarcity strategy in crowdfunding: An empirical exploration of reward limits. *Information Systems Research* **31**(4) 1107–1131.

Zeng, Zhiyu, Hengchen Dai, Dennis J Zhang, Heng Zhang, Renyu Zhang, Zhiwei Xu, Zuo-Jun Max Shen. 2023. The impact of social nudges on user-generated content for social network platforms. *Management Science* **69**(9) 5189–5208.

Zeng, Zhiyu, Zhiqi Zhang, Dennis Zhang, Tat Chan. 2024. The impact of recommender systems on content consumption and production: Evidence from field experiments and structural modeling. *Available at SSRN 4915562* .

Zhao, Weihong, Wen Wang, Siva Viswanathan. 2024. Spillover effects of generative ai on human-generated content creation: Evidence from a crowd-sourcing design platform. *Available at SSRN 4693181* .

Zhou, Wei, Mingfeng Lin, Mo Xiao, Lu Fang. 2025. Higher precision is not always better: Search algorithm and consumer engagement. *Management Science* **71**(7) 6204–6226.

Zou, Tianxin, Yue Wu, Miklos Sarvary. 2024. Quality-variety tradeoffs in recommendation systems on content platforms. *Columbia Business School Research Paper No. 4966901* .

Online Appendices

A. Randomization Check of Recommended Videos

Table A.1 Randomization Check of Recommended Videos (First-Level Categories)

First-Level Category	p-value
Military	0.750
Travel	0.997
Education	0.755
Lifestyle	0.612
Entertainment	0.121
Relationships	0.421
Oddities	0.119
Fashion	0.129
Appearance	0.247
Agriculture	0.641
Books	0.647
Comedy	0.460
Real Estate/Home	0.953
Games	0.787
Film/TV/Show	0.511
Selfie	0.598
Pets	0.321
Sports	0.480
Humanities	0.872
Current Affairs	0.969
Astrology/Mysticism	0.584
Other	0.719
Art	0.910
Parenting	0.578
Fitness	0.765
Anime/Manga	0.277
Food	0.524
Photography	0.245
Finance	0.656
Dance	0.157
Beauty	0.905
Health	0.743
Automotive	0.416
Technology	0.471
Music	0.972
Science	0.614

Notes: Values represent p-values from t-tests comparing treatment and control groups, based on each user's average share of recommended videos by first-level categories, without conditioning on opening the comment section.

B. Pearson Correlation Matrix of Focal Variables

Table B.1 Pearson Correlation Matrix of Focal Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) *Treat*	1												
(2) *CommentDuration*	0.00	1											
(3) *PostComment*	0.00	0.13	1										
(4) *Male*	0.00	0.02	0.01	1									
(5) *PostDate*	0.00	-0.01	-0.01	0.01	1								
(6) *Follower*	0.00	0.00	0.00	0.00	0.00	1							
(7) *Experience*	0.00	0.01	-0.05	0.00	-0.04	0.01	1						
(8) *VideoUpload*	0.00	-0.01	0.08	-0.08	0.00	0.09	-0.08	1					
(9) *ProducerMale*	0.00	0.02	0.00	0.30	0.00	0.00	0.01	-0.03	1				
(10) *ProducerFollower*	0.00	0.00	-0.01	-0.01	-0.04	0.00	0.05	-0.01	0.03	1			
(11) *ProducerExperience*	0.00	0.00	-0.02	-0.03	-0.05	0.01	0.12	-0.01	0.02	0.17	1		
(12) *VideoDuration*	0.00	0.05	-0.02	0.01	0.03	0.00	0.05	-0.04	0.04	0.01	0.00	1	
(13) *ContentType*	0.00	0.01	-0.05	0.04	0.00	0.00	0.06	-0.08	0.10	0.05	0.02	0.20	1

C. Randomization Check: One Week Before Treatment

Table C.1 Randomization Check Results (One Week Before the Experiment)

Variables	Full Sample (1)	Male Users (2)	Female Users (3)
CommentDuration	0.315	0.498	0.409
PostComment	0.981	0.439	0.444
Male	0.458	–	–
PostDate	0.280	0.478	0.422
Follower	0.247	0.164	0.905
Experience	0.568	0.678	0.698
VideoUpload	0.271	0.846	0.211
ProducerMale	0.612	0.105	0.176
ProducerFollower	0.979	0.296	0.247
ProducerExperience	0.756	0.777	0.813
VideoDuration	0.683	0.787	0.407
ContentType	0.744	0.467	0.247

Notes: Values represent p-values from t-tests comparing the treatment and control groups. The first column reports results for the full sample, while the second and third columns report p-values for male and female users separately. A dash (-) indicates the variable is not applicable for the subgroup.

D. Heterogeneous Effects in Female-Dominated Video Categories

In Section 5.2.2, we examined comment production behaviors across genders to uncover the social motivations driving female users' engagement. Building on these findings, we employed an alternative approach to test whether gender-aware curation leads to a greater sense of confidence, consensus, and belonging among female users, which may explain their increased engagement in comment production. We introduced $FemaleCategory_j$, a dummy variable equal to 1 if video j belongs to a female-dominated category. A category was classified as female-dominated if both the proportion of female users and the proportion of female commenters exceeded 60%.[1]. These metrics were calculated based on control group data covering 35 first- and 210 second-level video categories.[2] To ensure accuracy, we manually reviewed the classifications and validated them using large language models. The details of the classified female-dominated categories are presented in Table D.1. Most identified female-dominated categories (e.g., Beauty-Cosmetics and Fashion-Wedding) align with widely recognized gender-specific interests. To examine heterogeneous treatment effects, we estimated the following OLS regression with robust standard errors:

$$PostComment_{ij} = \beta_0 + \beta_1 Treat_i + \beta_2 (Treat_i \times FemaleCategory_j) + \beta_3 Controls_i + \gamma_j + \delta_{ij} \qquad (6)$$

where δ_{ij} is the error term. β_2 is the focal model coefficient of interest.[3] We estimated Equation (6) separately for male and female users. Column (1) of Table D.2 shows a negative interaction term between $Treat$ and $FemaleCategory$ ($\beta_2 = -0.0007$, p-value < 0.01), indicating that the positive effect of the gender-aware curation algorithm on comment production for female users was 3.0% weaker in female-dominated categories compared to gender-neutral or male-dominated categories. Specifically, the treatment increased female commenting likelihood by 3.5%[4] in gender-neutral or male-dominated categories, but only by 0.4%[5] in female-dominated categories. These results show that female users experienced a weaker comment-production-boosting effect in contexts where like-minded content and community bonds are already strong, supporting our proposition that gender-aware curation increases female participation by fostering confidence and community belonging. For male users, Column (2) shows that the interaction term is not statistically significant ($\beta_2 = 0.0003$, p-value > 0.1), indicating no heterogeneous effects across different video categories.

[1] This 60% threshold is widely used in practice. For instance, government guidelines (Ontario Pay Equity Office) and research reports (see page 6 of the Australian Fair Work Commission report) conventionally classify an occupation as female-dominated if women comprise more than 60% of the workforce.

[2] These video categories are provided by Platform A.

[3] $FemaleCategory_j$ is absorbed by the video fixed effects and its coefficient is not separately identified.

[4] This is calculated as: 0.0008 / 0.023 = 3.5%.

[5] This is calculated as: (0.0008 - 0.0007) / 0.023 = 0.4%.

Table D.1 Female-Dominated Video Categories

First-Level Category (1)	Second-Level Category (2)	Share of Female Commenters (3)	Share of Female Users (4)
Beauty	Cosmetics	0.85	0.92
Beauty	Makeup Imitation	1.00	0.92
Games	Dress-up Games	1.00	0.91
Parenting	Childrearing	0.74	0.90
Parenting	Pregnancy & Childbirth	0.81	0.89
Games	Dating Simulation Games	0.80	0.88
Food	Mukbang	0.72	0.86
Art	Others	0.75	0.85
Games	Management Simulation Games	1.00	0.83
Fashion	Marketing & Sales	0.74	0.82
Fashion	Wedding	0.60	0.82
Beauty	Skincare	0.75	0.82
Beauty	Others	0.73	0.81
Parenting	Kids Sharing	0.67	0.81
Film/TV/Show	Variety Shows	0.68	0.78
Games	Nurturing Game	1.00	0.78
Education	Learning Skills	1.00	0.77
Education	Educational Peripherals	0.77	0.75
Oddities	Niche Subcultures	0.86	0.75
Parenting	Others	0.69	0.75
Entertainment	Gossip	0.66	0.75
Film/TV/Show	Campus Short Drama	0.72	0.74
Art	Handicrafts	0.80	0.73
Games	Adventure Games	0.75	0.72
Beauty	Cosmetic Procedures	0.66	0.71
Fashion	Outfits	0.62	0.71
Photography	Gear Photography	0.71	0.71
Books	Novels	0.69	0.70
Astrology/Mysticism	Divination & Fortune Telling	0.69	0.70
Beauty	Beauty & Wellness	0.69	0.70
Film/TV/Show	Historical Short Drama	1.00	0.69
Fashion	Others	0.63	0.69
Film/TV/Show	Suspense Short Drama	1.00	0.68
Appearance	Others	0.61	0.68
Art	Art & Design	0.84	0.68
Art	Painting	0.74	0.68
Games	Casual Puzzle Games	0.75	0.67
Anime/Manga	Cool Anime Art	0.64	0.66
Education	Academic Education	0.67	0.65
Art	Traditional Crafts	0.74	0.63
Appearance	Costume Change	0.67	0.63
Food	Cooking Tutorials	0.66	0.61

Notes: The share is calculated from the control group data. Values represent proportions of female users.

To ensure robustness, we replicated the analysis using alternative thresholds of 65% and 70% to define female-dominated categories. As shown in Columns (1) and (3) of Table D.3, the coefficients of the interaction term between *Treat* and *FemaleCategory* remain significantly negative, confirming the robustness of our results.

Table D.2 Heterogeneous Effect of Gender-Aware Curation Algorithm on Comment Production Across Video Types, DV = PostComment

Sample	Female Users (1)	Male Users (2)
Treat	0.0008***	-0.001***
	(0.0001)	(0.0001)
Treat × *FemaleCategory*	**-0.0007*****	**0.0003**
	(0.0002)	(0.0004)
Control baseline (mean)	0.023	0.026
Relative effect size	-3.0%	-
Controls	Yes	Yes
Video fixed effects	Yes	Yes
Observations	6,628,193	5,473,616
R-squared	0.223	0.192

Notes: ***p<0.01; **p<0.05; *p<0.1. Values in parentheses are robust standard errors. A dash (-) indicates the variable is not applicable for the subgroup.

Table D.3 Results Using Alternative Thresholds for Female-Dominated Category, DV = PostComment

Threshold	> 65%		> 70%	
	Female Users (1)	Male Users (2)	Female Users (3)	Male Users (4)
Treat	0.0008***	-0.001***	0.0007***	-0.0008***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Treat × *FemaleCategory*	**-0.0007*****	-0.00003	**-0.001*****	-0.00004
	(0.0002)	(0.0005)	(0.0004)	(0.001)
Control baseline (mean)	0.023	0.026	0.023	0.026
Relative effect size	-3.0%	-	-4.3%	-
Controls	Yes	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes	Yes
Observations	6,628,193	5,473,616	6,628,193	5,473,616
R-squared	0.223	0.192	0.223	0.192

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Values in parentheses are robust standard errors. A dash (-) indicates the variable is not applicable for the subgroup.

E. Additional Analyses and Robustness Tests

E.1. Effects of Gender-Aware Curation Algorithm on User Engagement Over Time

As our main analyses rely on data from the first treatment day, one concern is that the observed outcomes may be partially driven by short-term novelty effects. To address this, we extended our analyses to a three-week period, from October 31st to November 19th, 2023. Due to the large volume of data over this extended period, estimating effects at the user-video level became computationally impractical. We therefore aggregated the data at the user-day level and estimated the treatment effects separately for each of the three weeks to assess how the effects evolved over time. Specifically, we estimated the following specifications with robust standard errors:

$$DV1_{it} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 Controls_{it} + \nu_t + \psi_{it} \tag{7}$$

$$DV1_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Male_i + \beta_3 (Treat_i \times Male_i) + \beta_4 Controls_{it} + \nu_t + \zeta_{it} \tag{8}$$

where $DV1_{it}$ represents the two user engagement metrics, i.e., $CommentDuration_{it}$ (the browsing duration in minutes that user i spends on comments on day t) and $PostComment_{it}$ (a binary indicator for whether user i commented on day t). $Controls$ includes all previously mentioned control variables, aggregated at the user level. ν_t denotes the day fixed effects. ψ_{it} and ζ_{it} are error terms. $CommentDuration_{it}$ is standardized for confidentiality, and highly-skewed variables are log-transformed.

Table E.1 shows the impact of the gender-aware curation algorithm on comment consumption, with $CommentDuration_{it}$ as the dependent variable. Columns (1)-(3) indicate that the coefficients of $Treat$ remain consistently positive across all three weeks, with the magnitude increasing over time. In contrast, Columns (4)-(6) show no significant heterogeneous effect across genders. Table E.2 presents the estimation results with $PostComment_{it}$ as the dependent variable. The coefficients of $Treat$ in Columns (1)-(3) are not statistically significant. The negative coefficients of the interaction term between $Treat$ and $Male$ in Columns (4)-(6) reveal consistent heterogeneous effects across genders, with male users becoming less likely to comment and female users more likely to do so. The results from both tables are qualitatively consistent with our main analyses in Section 4, specifically those reported in Table 4, Table 5, Table 6, and Table 7, which are based on the dataset from the first treatment day. Taken together, these findings suggest that the observed effects in Section 4 are unlikely to be short-term novelty responses. Rather, the findings reflect longer-term engagement patterns, which aligns with our experimental design, where users were not explicitly informed about the algorithmic changes.

Table E.1 Effects of Gender-Aware Curation Algorithm on Comment Consumption Over Time

Variables	*CommentDuration*			*CommentDuration*		
	Week 1	Week 2	Week 3	Week 1	Week 2	Week 3
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	**0.012*****	**0.015*****	**0.018*****	0.011***	0.015***	0.014***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Male	-0.086***	-0.088***	-0.116***	-0.087***	-0.088***	-0.120***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Treat × *Male*				**0.002**	**0.001**	**0.001**
				(0.002)	(0.003)	(0.003)
Relative effect size	1.8%	2.1%	2.3%	-	-	-
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,595,567	2,021,120	1,568,256	2,595,567	2,021,120	1,568,256
R-squared	0.027	0.034	0.013	0.027	0.034	0.013

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Values in parentheses are robust standard errors. All variables are aggregated at the user level. $CommentDuration_{it}$ is standardized for confidentiality, and highly-skewed variables are log-transformed. A dash (-) indicates the variable is not applicable for the subgroup.

Table E.2 Effects of Gender-Aware Curation Algorithm on Comment Production Over Time

Variables	*PostComment*			*PostComment*		
	Week 1	Week 2	Week 3	Week 1	Week 2	Week 3
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	**0.0001**	**0.0004**	**0.0005**	**0.001****	**0.003*****	**0.002****
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.003***	0.001	0.003**	0.001	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Treat × *Male*				**-0.003*****	**-0.006*****	**-0.004*****
				(0.001)	(0.001)	(0.001)
Control baseline (male)	0.267	0.281	0.298	0.267	0.281	0.298
Control baseline (female)	0.280	0.290	0.306	0.280	0.290	0.306
Relative effect size (male)	-	-	-	-0.7%	-1.1%	-0.7%
Relative effect size (female)	-	-	-	0.4%	1.07%	0.6%
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,595,567	2,021,120	1,568,256	2,595,567	2,021,120	1,568,256
R-squared	0.091	0.098	0.095	0.091	0.098	0.095

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Values in parentheses are robust standard errors. All variables are aggregated at the user level, and highly skewed variables are log-transformed. A dash (-) indicates the variable is not applicable for the subgroup.

E.2. Analyses of Comment Exposure

To examine how the gender-aware curation algorithm influenced the comments shown to users, we tested whether the treatment increased gender homophily, i.e., whether users were more likely to see comments produced by users of the same gender. We introduced $FemaleCommenter_{ij}$, the proportion of first-level comments shown to user i on video j that female users posted. We re-estimated Equation (2), using $FemaleCommenter_{ij}$ as the dependent variable. As shown in the Column (1) of Table E.3, both male and female users were exposed to more comments from male commenters under the treatment (α_1 = -0.003, p-value < 0.01), although this shift was slightly smaller for male users ($\alpha_3 = 0.001$, p-value < 0.01). This suggests that the observed gender-based differences in exposure were driven primarily by the content of the comments, rather than by the gender of the commenters.

We then investigated whether the treatment increased exposure to more controversial comments. We introduced $Controversialness_{ij}$, defined as the average controversy score of the first-level comments shown to user i on video j. Following Cheng et al. (2024), the controversy score for each first-level comment was calculated as the variation in sentiment among its second-level comments (i.e., the dispersion in the ratio of positive to negative second-level comments), and was set to missing if the second-level comments were not available. As shown in the Column (2) of Table E.3, the treatment significantly increased exposure to more controversial comments (α_1 = 0.0001, p-value < 0.01), with a stronger effect for male users (α_3 = 0.0001, p-value < 0.01). This finding is consistent with prior studies (Wardhaugh and Fuller 2021) suggesting that men have a stronger preference for debatable or contentious content. Together, manipulation checks via three approaches confirm that the gender-aware curation algorithm meaningfully altered which comments were displayed.

Table E.3 Additional Dimensions of Comment Exposure for Control Group Viewers

Variables	*FemaleCommenter* (1)	*Controversialness* (2)
Treat	**-0.003*****	**0.0001*****
	(0.0001)	(0.00001)
Male	-0.005***	-0.00004**
	(0.0002)	(0.00002)
Treat × *Male*	**0.001*****	**0.0001*****
	(0.0002)	(0.00002)
Controls	Yes	Yes
Video fixed effects	Yes	Yes
Observations	12,356,563	11,301,075
R-squared	0.761	0.723

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Values in parentheses are robust standard errors. Observations where user i did not browse any second-level comments under video j are excluded from Column (2), as *Controversialness* cannot be defined in such cases.

E.3. Effects of Gender-Aware Curation on Textual Features of User Comments

Comment production can be viewed as a two-stage process (as shown in Figure E.1): users first decide whether to post a comment, and then decide what kind of comment to post (i.e., textual attributes). In the main text (Section 4.2), we showed that the gender-aware curation algorithm influenced users' comment production behavior, with effects that varied by gender. A natural follow-up question is whether the treatment also changed the characteristics of the comments users produced. To investigate this, we used the comment features discussed in Section 5.1 as outcome variables in Equation (2). As shown in Table E.4 and Table E.5, the coefficients of *Treat* and its interaction term with *Male* are statistically insignificant across all measures. These findings suggest that gender-aware curation (compared to the gender-neutral approach) shaped the likelihood of posting comments but not the style or informational quality of comments.

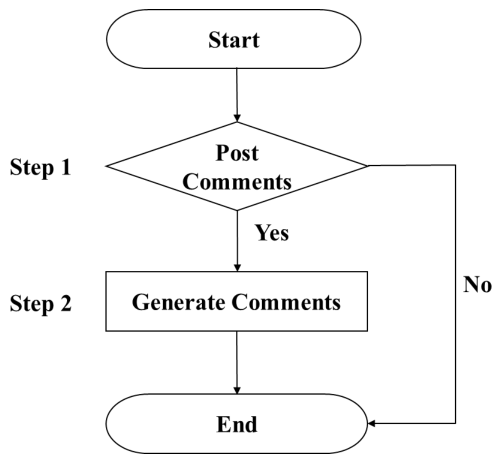


Figure E.1 Two-Stage Process of User Comment Production

Table E.4 Effects of Gender-Aware Curation on Textual Features of User Comments

Variables	*SimpleInteraction* (1)	*LexicalDensity* (2)	*LexicalVariation* (3)	*Entropy* (4)	*Length* (5)
Treat	**-0.001**	**0.001**	**0.003**	**0.009**	**0.008**
	(0.002)	(0.002)	(0.002)	(0.007)	(0.005)
Male	-0.010***	0.019***	0.024***	0.090***	0.052***
	(0.003)	(0.002)	(0.002)	(0.008)	(0.006)
Treat × *Male*	**-0.002**	**-0.004**	**-0.003**	**-0.015**	**-0.012**
	(0.003)	(0.003)	(0.003)	(0.010)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	215,038	215,038	215,038	215,038	215,038
R-squared	0.376	0.635	0.692	0.641	0.696

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Values in parentheses are robust standard errors. *Length* is log-transformed. Columns (1) to (5) are based on observations from both treatment and control groups where user i posted at least one comment on video j (295,706 observations), since comment attributes are otherwise undefined. Videos with only a single commenting user provide no within-video variation and are absorbed by video fixed effects, leaving 215,038 observations for estimation.

Table E.5 (Continued) Gender Differences in Comment Type Exposure for Control Group users

Variables	*TagFriend* (6)	*Positive* (7)	*Neutral* (8)	*Negative* (9)
Treat	**0.0004*****	**0.001**	**-0.0004**	**-0.001**
	(0.0001)	(0.002)	(0.002)	(0.001)
Male	-0.007***	-0.009***	0.006**	0.004**
	(0.0002)	(0.003)	(0.003)	(0.001)
Treat × *Male*	**-0.0005*****	**-0.005**	**0.004**	**0.001**
	(0.0002)	(0.003)	(0.004)	(0.002)
Controls	Yes	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes	Yes
Observations	12,356,563	215,038	215,038	215,038
R-squared	0.073	0.375	0.410	0.308

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Values in parentheses are robust standard errors. Columns (7) to (9) follow the same sample definition as Columns (1) to (5). Column (6) uses the full sample unconditional on posting any comments. Since each video was viewed by at least one user in both treatment and control groups, there are always at least two observations per video, so none are absorbed by video fixed effects.

E.4. Gender-Based Heterogeneity with Alternative Curation Data Inputs

In the main text (Section 4.2), we showed that incorporating gender data into the curation algorithm heterogeneously affected comment production between male and female users. To examine whether this heterogeneity is specific to gender-aware curation or also occurs when other types of user data are incorporated, we conducted a follow-up field experiment in June 2024. In this additional experiment, the curation algorithm for the control group incorporated two types of features: (1) users' gender data and (2) comment attributes. For the treatment group, the algorithm used the same features as in the control group and additionally incorporated users' historical engagement data with comments (e.g., browsing duration, likes, and replies).

The experiment was conducted from June 7^{th} to June 12^{nd}, 2024. As the system parameters were still being adjusted on June 7^{th}, we designated June 8^{th} (the first stable day) as the treatment period and June 6^{th} as the pre-treatment period. Our final sample consisted of 369,266 users on June 8^{th}, with 184,725 users in the treatment group and 184,541 users in the control group. We collected data on user engagement and attributes at the user, producer, and video levels. Summary statistics for the treatment period are reported in Table E.6. To validate the experimental design, we conducted randomization checks using pre-treatment data. Insignificant differences of key variables in Table E.7 confirm that the treatment and control groups were comparable prior to the intervention, suggesting that any differences observed during the treatment period can be attributed to the inclusion of user engagement data in the treatment group.

Table E.6 Summary Statistics of Focal Variables

Variables	Mean	SD	Min	Max
Treat	0.500	0.500	0	1
PostComment	0.024	0.152	0	1
Male	0.410	0.492	0	1
PostDate	6.140	8.444	0	38
Experience	3.765	2.547	0	12.790
VideoUpload	1.322	4.093	0	874
ProducerMale	0.502	0.500	0	1
ProducerExperience	4.565	2.649	0.533	13.560
VideoDuration	0.987	1.767	0	110.800
ContentType	0.893	0.310	0	1

Notes: To protect Platform A's sensitive information, the summary statistics for *CommentDuration*, *Follower*, and *ProducerFollower* are not displayed.

We then re-estimated Equations (1) and (2) in Section 3.5 using this new data sample, where $DV1_{ij}$ denotes $CommentDuration_{ij}$ and $PostComment_{ij}$. As shown in Column (1) of Table E.8, the positive coefficient of *Treat* suggests that incorporating engagement data increased comment consumption. In Column (2), the interaction term with *Male* is statistically insignificant, indicating

Table E.7 Randomization Check Results

Variables	Full Sample (1)	Male Users (2)	Female Users (3)
CommentDuration	0.782	0.790	0.936
PostComment	0.769	0.615	0.389
Male	0.681	-	-
PostDate	0.437	0.060	0.389
Follower	0.774	0.316	0.549
Experience	0.874	0.325	0.404
VideoUpload	0.153	0.643	0.138
ProducerMale	0.492	0.873	0.873
ProducerFollower	0.265	0.446	0.446
ProducerExperience	0.424	0.713	0.713
VideoDuration	0.107	0.267	0.267
ContentType	0.837	0.452	0.452

Notes: Values represent p-values from t-tests comparing the treatment and control groups. The first column reports results for the full sample, while the second and third columns report p-values for male and female users separately. A dash (-) indicates the variable is not applicable for the subgroup.

no evidence of gender-based heterogeneity. These findings are qualitatively consistent with our main results. However, the insignificant coefficient of *Treat* in Column (3) indicates no significant treatment effect on overall comment production, consistent with our main findings in Section 4.2. In Column (4), the coefficients of both *Treat* and its interaction term with *Male* are statistically insignificant. This suggests that we find no significant treatment effect on comment production or heterogeneous effects across genders. This is plausibly because the engagement data in our context primarily capture consumption rather than production behaviors and therefore did not reproduce the heterogeneous production effects observed under gender-aware curation in Section 4.2.

Table E.8 Impact of Gender-Aware Curation Algorithm on User Engagement

Variables	*CommentDuration* (1)	*CommentDuration* (2)	*PostComment* (3)	*PostComment* (4)
Treat	**0.003*****	**0.002****	**0.0001**	**0.0002**
	(0.001)	(0.001)	(0.0001)	(0.0001)
Male	-0.002*	-0.003**	0.002***	0.002***
	(0.001)	(0.001)	(0.0001)	(0.0002)
Treat × *Male*		**0.002**		**-0.0002**
		(0.002)		(0.0002)
Controls	Yes	Yes	Yes	Yes
Video fixed effects	Yes	Yes	Yes	Yes
Observations	6,861,498	6,861,498	6,861,498	6,861,498
R-squared	0.145	0.145	0.235	0.235

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Values in parentheses are robust standard errors. *CommentDuration* is standardized for data confidentiality.

E.5. Alternative Operationalization of Variables

To ensure the robustness of our main findings, we employed alternative operationalizations of comment production. Since the effects of gender-aware curation on comment production may depend on how commenting behavior is measured, testing the alternatives helps verify the consistency and generalizability of our results. We replaced the binary indicator of whether a user commented on a video with the number of comments user i posted for video j ($CommentNum_{ij}$). This count measure captures both the occurrence and intensity of commenting. As shown in Table E.9, the results remain qualitatively consistent with our main findings.

Table E.9 Results Using Alternative Dependent Variables for Comment Production, DV = CommentNum

Variables	(1)	(2)
Treat	**-0.0002**	**0.003*****
	(0.001)	(0.001)
Male	0.024***	0.027***
	(0.001)	(0.001)
Treat × *Male*		**-0.007*****
		(0.001)
Relative effect size (male)	-	-3.5%
Relative effect size (female)	-	2.6%
Controls	Yes	Yes
Video fixed effects	Yes	Yes
Observations	12,356,563	12,356,563
R-squared	0.149	0.149

Notes: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$. Values in parentheses are robust standard errors. *CommentNum* is standardized for data confidentiality.