

Positivity Bias in AI-Generated Summaries of User-Generated Content: Exploring Its Sources and Impact on Public Sentiment

Anna Yan Liu

Institute of Communications Research, College of Media, University of Illinois Urbana-Champaign, Urbana IL, USA

Maggie Mengqing Zhang

Institute of Communications Research, College of Media, University of Illinois Urbana-Champaign, Urbana IL, USA

Abstract

Recently, platforms have been increasingly deploying generative AI (GenAI) to summarize user-generated content (UGC) into AI-generated summaries (AIGS). However, the potential bias in AIGS and its impact on the public remain inadequately examined. We used Weibo, a leading social media in China, as a case to investigate these important questions, focusing on public sentiments. Specifically, we explored whether AIGS are biased in representing emotions in UGC and whether such representation influences subsequent public sentiment. We empirically identify two sources of bias in the algorithmic processes underlying the production of AIGS from UGC: *the sampling process*, in which GenAI selects a subset of UGC, and *the summarizing process*, in which the summary is generated from the sampled content. Comparing emotions in AIGS, sampled UGC, and all UGC, we found evidence of bias in both processes. In our case, GenAI tends to favor positive UGC during the sampling process and produces summaries that further amplify this positivity, leading to an overrepresentation of positive sentiments in AIGS. Additionally, we utilized a Difference-in-Differences (DiD) design to explore AIGS in public sentiment dynamics. Findings suggest that AIGS alone are insufficient to influence public sentiment significantly. Overall, this study provides important implications for deploying GenAI in public online discussions.

Keywords: Generative AI (GenAI), AI-generated summaries (AIGS), Bias, Source of bias, Social media, Public sentiment

Introduction

With the rise of generative AI (GenAI) technology, digital platforms are increasingly developing AI-generated summaries (AIGS) features that can automatically summarize user-generated content (UGC). For instance, review platforms like Google Maps and E-commerce platforms like Amazon have launched AIGS features to help users quickly grasp the essence of customer reviews (Alavi & Nozari, 2025; Su et al., 2024). Some video-based social media platforms have introduced AIGS features to generate textual summaries of user-generated video content (Wei et al., 2025). Essentially, AIGS marks a fundamental shift in the application of GenAI: integrating GenAI into online communities, GenAI emerges as a public-facing communicative information source that synthesizes prior UGC and “retells” it to audiences, shaping public discussion and users’ perceptions of issues on a broader scope. This potentially large-scale influence of GenAI in the public discourses motivates our study.

This study builds on three bodies of research. The first one focuses on AI biases. Previous studies have theoretically discussed the three main categories of AI bias sources, namely data, model design, and implementation (Ferrara, 2023b; Mehrabi et al., 2022). Research also empirically analyzed biased representation of GenAI outputs, such as biased demographic representations in AI-generated visuals (Currie et al., 2024). However, the deployment of AI to summarize large volumes of user-generated content on public platforms may introduce new forms of bias. A second body of research has empirically examined the impact of AIGS on users, using laboratory experiments (Ouyang & Xu, 2024) and natural experiments (Alavi & Nozari, 2025) to explore how the introduction of AIGS shapes user behaviors. Some studies focused on video-based social media (Wei et al., 2025), where AIGS converts user-generated video into a textual summary, offering audiences an alternative format of the same content. Other studies concentrated on E-commerce platforms, on which AIGS aggregate multiple user-generated reviews into a single summary, influencing consumers’ subsequent engagement with individual reviews, review posting behaviors, and purchase decisions (Su et al., 2024). However, few studies have examined the implementation of AIGS on public discussion-oriented social media, where users stay informed and express personal opinions on trending issues (Dong & Lian, 2021). This context is worth exploring, as AIGS on social media may shape public opinion dynamics on societal issues. The third one focuses on the role of emotions in influencing users’ online engagement and public discussions (e.g., Naskar et al., 2020). Previous studies on emotional diffusion have primarily exam-

ined the direct spread of emotions between users on social media (Sasaki et al., 2021) or between humans and machines (Prinz, 2022). We focus on the introduction of AIGS, which summarizes and “retells” public discussions, as it may introduce new dynamics in how emotions spread online and have an impact on users’ online behaviors.

Using the case of AIGS for trending hashtags on the leading Chinese social media platform Weibo, this study examines whether AIGS are biased in representing public sentiment in UGC. We specifically investigated whether bias arose during two stages of Weibo AIGS production: (1) GenAI’s selection of content to represent all UGC (the sampling process), and (2) GenAI’s production of AIGS based on selected content (the summarizing process). We conducted sentiment analysis on AIGS content, the sampled content, and all the corresponding UGC. Through comparing sampled content versus all UGC, as well as AIGS content versus sampled content, we found that the Weibo GenAI tends to favor positive content during the sampling process and produce summaries that further amplify this positivity during the summarizing process, leading to an overrepresentation of positive public sentiment in AIGS.

We explored the role of AIGS in influencing the emotional dynamics in UGC. During the introduction phase of the AIGS feature, Weibo applied it to some hashtags but not others, enabling a natural experiment setting and allowing us to employ a Difference-in-Differences (DiD) design to examine whether emotions in AIGS impact subsequent public sentiment. Since AIGS exhibits varying sentiments across different hashtags, we distinguished between hashtags with positive and negative AIGS to better isolate its effect on public sentiment. We conducted two sets of analyses: one using hashtags with positive AIGS as the treatment group and hashtags without AIGS as the control group, and the other replicating this design with issues featuring negative AIGS as the treatment group. To evaluate the influence of AIGS on subsequent public sentiment, we compared changes in sentiment in UGC over time between the treatment and control groups in each analysis. The results revealed no significant difference in the changes of sentiment between the treatment and control groups, indicating that AIGS alone are insufficient to shape public sentiment. Further interpretation of findings is provided in the discussion section.

Focusing on the emerging role of GenAI as a public-facing information source on social media, this study empirically examined AIGS bias and its effects on public sentiment, making four key contributions to GenAI research. First, most studies have categorized sources of GenAI bias into data,

model design, and implementation (AlMakinah et al., 2025; Mehrabi et al., 2022). Focusing on the deployment of GenAI to summarize large volumes of UGC on public platforms, this study examines new forms of bias emerging from this process, extending the theoretical framework of AI bias sources to include bias introduced in the common algorithmic processes of AIGS production: the sampling and summarizing processes. Second, most studies have focused on theoretical discussions of the sources, consequences, and solutions of AI bias (e.g., Ferrara, 2023b). This study provides empirical evidence of new sources of GenAI bias by showing that both the sampling and summarizing processes involved in Weibo AIGS production tend to favor the representation of positive sentiments. Third, prior empirical studies mostly focused on the biased demographic representations in GenAI outputs, such as distorted gender and race presentation in AI-generated occupational visuals (Currie et al., 2024; Kekez et al., 2025; Sun et al., 2023). This study focuses on bias in the representation of public sentiment, which is a form of bias related to personal opinion and expression that has the potential to influence public discourse on societal issues. Lastly, prior explorations on the effect of AIGS on user behavior have focused on video-based social media or E-commerce platforms (Alavi & Nozari, 2025; Kim et al., 2024). Our study extends this investigation to public discussion on social media, where AIGS are used to summarize public opinions on trending issues, potentially shaping people's perception of and actions towards social issues. Although our findings suggest that AIGS alone are insufficient to significantly influence public sentiment, this study serves as an initial step toward our further understanding of the bias and effect of GenAI when it functions as a public-facing information source in the public information sphere on social media.

Literature Review

AI Bias

AI bias generally refers to a systematic tendency or error in a model that leads to unfair outcomes (Ferrara, 2023a). Being consistent and predictable, bias differs from random differences, which are unpredictable variations (Tejani et al., 2024). Most definitions of AI bias focus on its tendency to favor certain groups of people, such as those defined by gender or race (Fletcher et al., 2021; Ntoutsi et al., 2020). In this study, we focus on AI bias in the inaccurate representation of public discussion. Thus, following Ferrara (2023b), we define AI bias more broadly as “the presence of systematic misrepresentations, attribution errors, or factual distortions that result in favoring certain

groups or ideas, perpetuating stereotypes, or making incorrect assumptions based on learned patterns” (p.2).

AI bias results from a complex set of factors that are commonly divided into three categories (Mehrabi et al., 2022). The first one is **data bias**, which originates from the training data used to build AI models, as exemplified by the non-representative samples that fail to reflect the broader population, omitted or missing variables and data points, inconsistencies during data cleaning, errors in annotation, etc. (Ferrara, 2023b; Mayuravaani et al., 2024; Reyero Lobo et al., 2023). The second is **AI model bias**, which arises from algorithmic design decisions such as simplifying linear relationships, mistaking correlation for causation during model development, determining the selection and weighting of variables in models, and defining optimization objectives, among others (Kadiresan et al., 2022; Mienye et al., 2024). The third is **implementation bias**, which includes the deployment of AI in certain real-world settings and the users’ interaction behaviors with AI (Gray et al., 2024; Mienye et al., 2024). For instance, bias can result from discrepancies between the context in which a model is trained and the one in which it is applied, as well as from users’ interactions with AI based on their pre-existing beliefs and preferences (AlMakinah et al., 2025; Mehrabi et al., 2022). Furthermore, these three categories are deeply shaped by **broader structural forces**, as exemplified by cultural traditions and dynamics, social structures, political reasons, historical legacies, etc. (Afreen et al., 2025; Mayuravaani et al., 2024).

These biases lead to unfair AI outputs, such as inaccurate representations of demographic groups in AI-generated visuals (Sun et al., 2023) and discriminatory decision-making in job candidate selection (Deshpande et al., 2020). These unfair outcomes have profound social consequences, including shaping public perceptions and behaviors, thereby reinforcing existing inequalities (Gorska & Jemielniak, 2023). Driven by the social impact of AI biases, studies have explored solutions, as exemplified by developing technological tools to reduce bias and produce fairer AI outputs (Deshpande et al., 2020), designing social processes to limit the influence of bias in human decision-making (Reed et al., 2025), creating frameworks to evaluate and audit AI systems (Landers & Behrend, 2023), and establishing laws and regulations to ensure accountability in AI models (Nadeem et al., 2022).

AI-generated Summaries and Potential Bias

In the context of content generation, most theoretical frameworks and empirical studies on AI bias examine the implementation of AI models to pro-

duce complex outputs from abstract prompts, such as generating visuals from textual prompts (Ferrara, 2023a, 2023b). For instance, researchers have analyzed visual outputs generated by ChatGPT and Midjourney, revealing biased gender and racial representations in AI-generated images of specific occupations (Currie et al., 2024; Kirk et al., 2021; Sun et al., 2023; Zhou et al., 2024). In recent years, generative AI has been increasingly deployed on public platforms to summarize large volumes of UGC, such as long videos and online reviews, into concise AIGS (Kim et al., 2024; Yuan et al., 2025). These AIGS provide users with concise overviews of the original UGC, serving as alternative sources to the substantial content created by users (Alavi & Nozari, 2025; Su et al., 2024).

The deployment of AI in summarizing UGC introduces potential biases due to the common workflows of algorithmic summarization systems. Research on extractive summarization algorithms, although not based on generative AI, has shown that bias can arise from the sampling processes underlying these models (Dash et al., 2019). Typically, such algorithms assign an importance score to each input text based on its estimated informational value and then select high-scoring segments for summary generation (Dash et al., 2019). Building on this line of prior research, this study explores the potential bias in AIGS in the representation of emotions in public discourse. More specifically, using Weibo's AIGS feature as a case, we focus on two key algorithmic processes involved in summary generation, namely the sampling process and the summarizing process, to assess how accurately Weibo's AIGS reflects the sentiment of the user-generated content they summarize:

RQ1: Whether and how Weibo AIGS are biased in representing public sentiment due to the algorithmic (a) sampling process, and (b) summarizing process?

The Effect of AIGS on Users

As digital platforms increasingly adopt AIGS to summarize UGC, recent studies have begun to empirically examine their influence on users through laboratory or natural experiments, mostly focusing on the contexts of video-based and E-commerce platforms (Ouyang & Xu, 2024; Su et al., 2024).

Studies focusing on video-based social media platforms consider AIGS as a feature to convert the format of information from video to text. These studies explored how the introduction of AIGS to summarize each user-generated video content influences audiences' behaviors on video-based social media, such as the Chinese platform Bilibili (Kim et al., 2024; Wei et al., 2025). While findings suggest that AIGS of video content can reduce

audiences' cognitive load (Ouyang & Xu, 2024), studies have reported mixed outcomes regarding how AIGS influences audience engagement. For example, Kim et al. (2024) found that AIGS of videos, especially long or utilitarian videos, increased audience engagement, while Wei et al. (2025) found that the introduction of AIGS reduced audiences' reviews, upvotes, and comments of videos. Overall, the introduction of AIGS feature on video-based social media shaped how the video content is consumed by audiences.

Other studies in the context of E-commerce platforms have examined how AIGS generate a single summary from multiple pieces of textual UGC, such as Amazon reviews (Alavi & Nozari, 2025). Using natural experiments, studies explored how the introduction of AIGS influenced users' review reading and posting behaviors, reporting mixed effects (Alavi & Nozari, 2025; Su et al., 2024). For example, Alavi and Nozari (2025) found that AIGS reduced the topical diversity of subsequent user-generated reviews, attributing this to an anchoring effect; while Su et al. (2024) found that AIGS increased content richness of reviews, explained by users' heightened perception of the social impact of their reviews as reflected in AIGS. In general, the introduction of AIGS on E-commerce platforms influences users' engagement with reviews, potentially shaping subsequent purchase decisions.

This study focuses on an under-examined context: public discussion on trending issues facilitated by social media platforms. Social media facilitates public discussion on diverse issues, enabling users to stay informed and express personal opinions (Dong & Lian, 2021). The launch of AIGS on social media is worth exploring, because of its potential to influence how people perceive societal issues and express personal opinions, thereby shaping public discussion.

We focus on the emotional aspect of AIGS on social media, as emotions are found to spread among individuals on social media, influencing users' online expression and public discussions (Naskar et al., 2020). Prior research has shown that emotions spread through online social networks (Goldenberg & Gross, 2020; Sasaki et al., 2021). For instance, sentiment in Facebook posts has been found to influence the emotional expressions of peers (Kramer et al., 2014). Similarly, using a natural experiment on Twitter, study showed that the emotional valence of tweets can affect the emotions expressed by other users (Ferrara & Yang, 2015).

As GenAI increasingly functions as a communicative agent, recent studies have explored the spread of emotions in human-machine communication (Prinz, 2022). Studies have explored how emotional transmissions between humans and GenAI are influenced by factors like the type of com-

munication counterpart (machine vs. human) and design features, such as bot appearance and AI’s emotional mimicry (Liu et al., 2024; W. Yang & Xie, 2024). However, these studies have primarily focused on GenAI in one-on-one human-machine interactions. When generating AIGS from UGC, GenAI summarizes previous UGC and “retells” it to broader audiences, with the potential to shape public emotions at scale. The emerging role of GenAI as a public-facing information source motivates this study to examine whether and how AIGS may influence public sentiment. More specifically:

RQ2: Whether and how the emotions of AIGS influence sentiment in subsequent posts associated with the hashtag?

Data

Background

To explore these questions, we used Weibo as the case. In 2024, Weibo, one of the leading social media platforms in China, launched AIGS for trending hashtags (Zhang, 2024). Weibo AIGS aggregates user-generated posts, allowing users to stay informed about public discussions alongside reading individual users’ posts.

While Weibo stated that AIGS are produced from user-generated posts, as is shown in Figure 1, a closer examination found that Weibo GenAI produces AIGS for a trending hashtag from a sample of posts (referred to as “reference posts” hereafter), rather than from the full set of associated user-generated posts (referred to as “all posts”). This provided an opportunity to examine whether biases exist in the two underlying processes through which Weibo AIGS are generated to represent public sentiment: the sampling process through which reference posts are selected from all posts, and the summarizing process through which AIGS are produced based on reference posts.

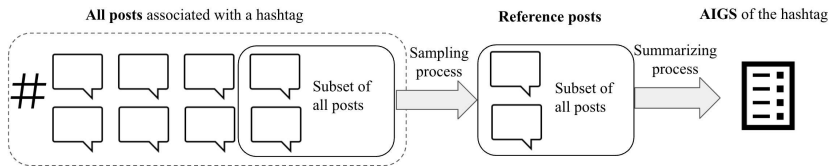


Figure 1: Weibo AIGS Generation process

At the time of our data collection (December 2024), Weibo introduced the AIGS feature for some hashtags but not others. As shown in Figure 2,

for hashtags with the AIGS feature, users accessed AIGS content by clicking on the hashtag, scrolling down to the folded AIGS, and then clicking it to view the full summary on a new page. Because AIGS was applied only to some hashtags, this created an opportunity to examine whether and how the emotions conveyed in AIGS influence subsequent public sentiment through a natural experiment.

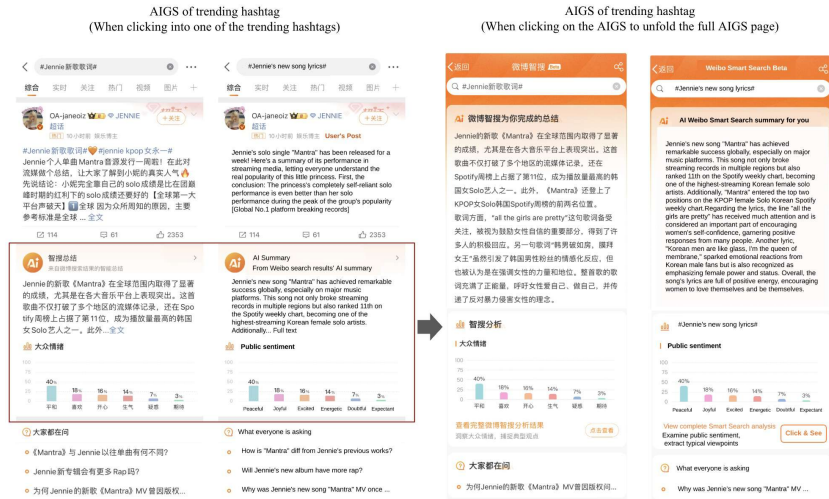


Figure 2: Weibo AIGS feature at the time of data collection (December 2024)

Data Collection

Using Weibo AIGS as the case, we used web scraping tools to collect two sets of data to explore our research questions. We first compiled a list of 11,628 hashtags that appeared on the Weibo hot search list throughout November 2024 and collected two datasets based on this list.

First, to examine RQ1a and RQ1b, which explore potential biases in the underlying algorithmic processes through which Weibo GenAI produces AIGS, we collected 2,421 hashtags with the AIGS feature implemented and their corresponding AIGS content, and 22,172 reference posts that were used to generate AIGS. Additionally, we gathered 644,485 all posts associated with these hashtags. For each hashtag, all posts associated with it cover a 7-day window, spanning 3 days before, the day of, and 3 days after its last appearance on the hot search list.

Second, RQ2 investigates the influence of AIGS emotions on subsequent

public sentiment. Based on the list of hashtags that appeared on the hot search list in November 2024, for each hashtag, we collected all posts associated with it within 7 days, including 3 days before, the day of, and 3 days after the last appearance of this hashtag on the hot search list. We then divided all posts associated with each hashtag into two subsets based on the publication time: those posted on or before the hashtag’s last appearance on the hot search list (referred to as “before hot search posts” hereafter), and those posted afterward (“after hot search posts”). The data for further analysis included (1) 2,209 hashtags with AIGS and 597,700 all posts associated with these hashtags; and (2) 2,728 hashtags without AIGS and 399,598 corresponding all posts¹. As introduced, at the time of data collection, Weibo implemented AIGS for only some hashtags. We included all hashtags with AIGS and randomly sampled hashtags without AIGS. Each hashtag was classified into either the treatment group (with AIGS) or the control group (without AIGS) for further analysis, based on whether valid AIGS content was retrieved through the web scraping tools.

Method

Measurement: Sentiment Score

To clean the raw data, for reference posts and all posts, we conducted text preprocessing by using the *tweet-preprocessor* Python library (Özcan, 2020) to remove URLs and username mentions in the texts. We retained emojis, since they are important elements that reflect sentiment. To measure the sentiment, we used *multilingual-e5-small-aligned-sentiment* (Tseng, 2024), a pre-trained multilingual transformer-based model. This model is fine-tuned specifically for multilingual text, including Chinese, which is the main language used in our data, and has shown satisfactory performance in sentiment-related tasks. The model outputs a sentiment score for each text on a continuous scale, typically ranging from -2 to 2, where negative scores indicate negative sentiment and positive scores indicate positive sentiment. We applied this model to texts of: (1) AIGS content, (2) reference posts that were used to generate AIGS, and (3) all posts.

¹We excluded hashtags that appeared in only one period to ensure each had both pre- and post-treatment observations, resulting in a smaller size of hashtags and associated all posts than the previous dataset.

Analysis

RQ1a. Whether and how Weibo AIGS are biased in representing public sentiment due to the sampling process?

To address RQ1a, which examines potential bias in the sampling process, we conducted two analyses comparing the sentiment scores of reference posts and all posts for each hashtag with AIGS. The first analysis serves as the main test, for which we employed the Wilcoxon Signed-Rank test to compare the mean sentiment scores of reference posts and all posts associated with the same hashtag, examining whether the two sets differ significantly. Additionally, as a robustness check, we calculated the Jensen-Shannon Divergence (JSD) score for each pair of reference posts and all posts, which measures the difference between their probability distributions. We then conducted a one-sample t-test across all hashtags to assess whether the JSD scores significantly exceeded the threshold of 0.5, indicating a systematic difference between these two distributions.

Main analysis: Wilcoxon Signed-Rank test. We conducted a Wilcoxon Signed-Rank test for each hashtag, comparing the sentiment scores of all posts to the median sentiment score of the corresponding reference posts, producing a p-value for each of the 2,421 hashtags.

To assess whether sentiments in reference posts systematically differ from all posts, we counted how many of the 2,421 Wilcoxon Signed-Rank tests showed significant differences. In addition, to assess whether significant differences occurred more frequently than expected by chance, we conducted a binomial test on these t-test results. A significant binomial result would indicate systematic differences between the sentiment scores of reference posts and those of all posts, suggesting that the algorithmic selection process introduces bias into the AIGS.

In addition to examining *whether* the sentiment scores of reference posts and all posts differ significantly through the Wilcoxon Signed-Rank test, we assessed *how* they differ by comparing their mean sentiment scores. We calculated the mean sentiment scores of reference posts and all posts for each hashtag. We then averaged these per-hashtag means across all hashtags to examine overall differences in sentiment scores.

Robustness check: Jensen-Shannon Divergence. As a robustness check, we calculated a JSD score for each pair of sentiment score distributions of reference posts and all posts associated with the same hashtag. JSD quantifies the similarity between two distributions, producing a score typically ranging from 0 to 1, with larger values indicating greater differences. To

examine whether these differences were systematically large across the 2,421 pairs of comparisons, we conducted a one-sample t-test, comparing these JSD scores against a threshold of 0.5. A significant result would indicate that reference posts and all posts differ systematically in sentiment scores, suggesting bias in the sampling process.

RQ1b. Whether and how Weibo AIGS are biased in representing public sentiment due to the summarizing process?

Similar to the analysis for RQ1a, we also conducted a main analysis and a robustness check to examine RQ1b. We explored biases in the summarizing process by comparing the sentiment scores of the AIGS and their corresponding reference posts. The main analysis was also a pair-by-pair Wilcoxon Signed-Rank test to examine, for each pair, whether the sentiment scores of AIGS and those of reference posts are significantly different. The robustness check was a cross-pair one-sample t-test analysis. We compared the average gaps between the sentiment scores of AIGS and those of reference posts with 0, which indicates no difference, to examine whether the gaps between the sentiment scores of AIGS and those of reference posts are significantly large.

Main analysis: Wilcoxon Signed-Rank test. We also conducted a Wilcoxon Signed-Rank test for each pair, comparing the sentiment scores of AIGS content and the corresponding reference posts for each of the 2,421 pairs. We also counted the significant results across all the test results and conducted a binomial test, a significant result of which would indicate that the summarizing process systematically introduces biases in the AIGS.

Moreover, to examine *how* the sentiment scores of AIGS and reference posts are different, we compared their mean sentiment scores. Similar to RQ1a, for each hashtag, we computed the mean sentiment score of its reference posts, then compared the overall average of these scores to the overall average sentiment scores of the AIGS across all hashtags.

Robustness check: Sentiment score differences. Similar to the robustness check for RQ1a, in this analysis, we examined whether the gaps between sentiment scores of AIGS and those of corresponding reference posts are significantly large through a one-sample t-test across all 2,421 pairs. We calculated the gap between the AIGS sentiment score and the average sentiment score of corresponding reference posts for each pair. Then we compared these gaps in sentiment scores across all pairs with 0, which indicates no difference between the sentiment scores of AIGS and reference posts. A significant result would suggest bias in the summarizing process.

RQ2. Whether and how do the emotions of AIGS influence sentiment in subsequent posts associated with the hashtag

To explore the effect of emotions in AIGS on the subsequent public sentiment, we used a Difference-in-Difference (DiD) design in natural experiments on Weibo. Before the DiD analysis, we conducted an event study plot on sentiment scores to examine the parallel assumption. As Figure 3 shows², the parallel trends assumption was met, indicating that there is no evidence of systematic platform-level selection bias between hashtags with and without the AIGS feature.

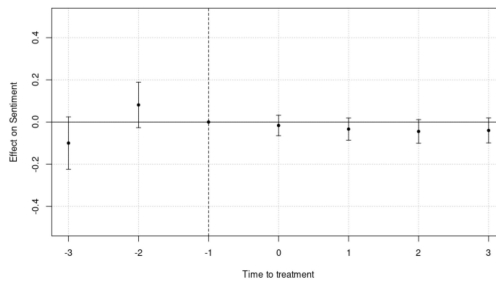


Figure 3: Event Study Plot on Sentiment Scores (Treatment versus Control)

Considering that the direction of AIGS sentiment may have different impacts on public sentiment, we first divided hashtags by the sentiment of their AIGS into two treatment groups. The first treatment group included hashtags with positive AIGS as the treatment, while the second included hashtags with negative AIGS. In both settings, the control group consisted of hashtags that did not receive AIGS.

As mentioned in data collection, for each hashtag, we divided all posts associated with it into two groups (before and after hot search) because the AIGS feature is typically implemented on the day the hashtag appears on the hot search list, according to Weibo's design. For each hashtag, we calculated

²The X-axis shows time to treatment: 0 is the time when treatment was applied, negative values represent days before the treatment, and positive values represent days after the treatment. The Y-axis, effect of sentiment, reports the estimated coefficients, which capture the difference in sentiment between treated hashtags (with AIGS) and control hashtags (without AIGS). Vertical error bars indicate confidence intervals; if a confidence interval crosses 0, the estimated effect is not statistically significant, meaning no significant difference between the treated and control groups. The vertical dashed line indicates the reference time point before the treatment. Overall, Figure 3 shows that, in the pre-treatment period, there is no significant difference between hashtags with and without the AIGS, therefore suggesting no systematic platform-level selection bias in which types of hashtags receive AIGS. In the post-treatment period, the coefficients also remain statistically insignificant, indicating that AIGS do not meaningfully influence public sentiment after treatment.

mean sentiment scores for before and after hot search posts, respectively. We then conducted two DiD analyses, respectively, to compare differences in mean sentiment scores over time between treatment and control groups.

Results

RQ1a. Whether and how Weibo AIGS are biased in representing public sentiment due to the sampling process?

The results of both the main analysis and the robustness check show a significant difference in emotions between all posts associated with hashtags and the reference posts for AIGS, suggesting that the sampling process introduces bias to AIGS. We found that the Weibo algorithm tends to select posts with positive emotions as reference posts for AIGS.

For the main analysis, we conducted a Wilcoxon Signed-Rank test comparing the sentiment scores of all posts associated with a hashtag and the mean of the sentiment scores of its reference posts. Of the 2,421 pairs of comparisons, 1,939 show a statistically significant difference, while 482 are not significant. Then, a binomial test was conducted to assess whether the number of significant results ($p < .05$) across 2,421 comparisons exceeded what would be expected by chance. As shown in Figure 4, the number of observed significant results ($n = 1,939$) is significantly higher than the expected count under the null hypothesis ($p < .001$), showing a systematic difference in sentiment scores between all posts associated with hashtags and the reference posts used for generating AIGS.

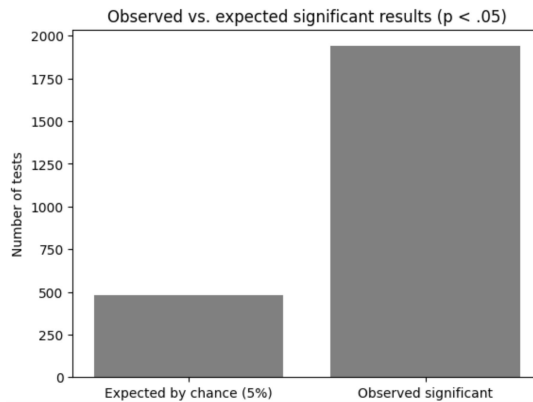


Figure 4: Observed vs. Expected Number of Significant Results in Binomial Test, Comparing Reference Posts and All Posts

We also conducted a two-sided one-sample t -test across all the pairs as a robustness check, comparing their JSD scores with the threshold of 0.5. The

results show a statistically significant difference, $t(2,420) = 64.30, p < .001$, with a mean JSD score ($M = 0.66, 95\% \text{ CI } [0.65, 0.66]$) significantly higher than the threshold. This suggests divergence between the distributions of sentiment scores of reference posts and all posts across hashtags.

To explore how the emotions among reference posts and all posts differ, we further compared the average of the mean sentiment scores of reference posts across the 2,421 hashtags ($M = 0.45$) with that of all posts ($M = 0.39$). This difference suggests that the Weibo algorithm tends to select posts with more positive emotions as reference posts to produce AIGS.

RQ1b. Whether and how Weibo AIGS are biased in representing public sentiment due to the summarizing process?

Both the main analysis and robustness check show a significant difference between the emotions of AIGS and their corresponding reference posts, suggesting that the summarizing process also introduces bias into the AIGS. The emotions of AIGS were found to be more positive than those of the reference posts.

Our main analysis involved the Wilcoxon Signed-Rank test comparing the sentiment scores of reference posts with those of AIGS for each hashtag. Among the 2,421 comparisons, 1,561 show significant differences. We then conducted a binomial test to examine whether the number of observed significant results ($p < .05$) exceeded what would be expected by chance. Figure 5 shows that the observed number of significant results ($n = 1,561$) is significantly higher than expected ($p < .001$), indicating that the sentiment scores of the AIGS and their corresponding reference posts differ systematically.

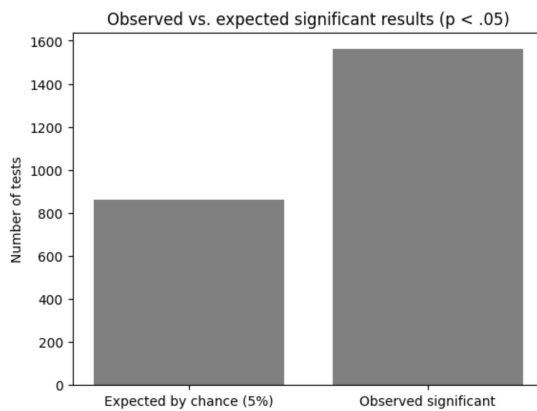


Figure 5: Observed vs. Expected Number of Significant Results in Binomial Test, Comparing AIGS and Reference Posts

For RQ1b, we also conducted a one-sample *t*-test as a robustness check across all the pairs, comparing the differences between AIGS and reference posts with 0. The result is statistically significant, $t(2,420) = 41.09, p < .001$, with a 95% CI [0.56, 0.62]. We computed the mean of sentiment scores of the 2,421 AIGS ($M = 1.04$) and the average of the 2,421 mean sentiment scores of reference posts ($M = 0.45$), with a gap of 0.59. These results suggest that the Weibo algorithm tends to produce AIGS with more positive sentiment than the reference posts on which they are based.

RQ2. Whether and how do the emotions of AIGS influence sentiment in subsequent posts associated with the hashtag

As shown in Table 1, the two DiD analyses compared the treatment groups (positive AIGS or negative AIGS) with the control group, respectively. The results suggest that emotions conveyed in AIGS have no significant influence on the subsequent public sentiment.

Table 1: Effect of AIGS on Public Sentiment

	Treatment group 1: Positive AIGS	Treatment group 2: Negative AIGS
Treatment	0.459 (0.029)***	-1.121 (0.042)***
Period	0.003 (0.025)	0.003 (0.026)
Treatment × Period	-0.041 (0.041)	0.013 (0.059)
N	8,590	6,740

The left side of Table 1 presents the comparison between treatment group 1 (positive AIGS) and the control group. To examine whether positive emotions in AIGS positively affected public sentiment, a DiD linear regression was conducted. The model included main effects for the treatment (positive AIGS vs. no AIGS), period (before vs. after), and their interaction. The interaction term, which represents the DiD effect, is not statistically significant, $b = -0.041, SE = 0.041, p = .255$. These findings show no significant difference in sentiment change over time between posts associated with hashtags featuring positive AIGS and those without AIGS.

The right side of Table 1 presents the results of comparing treatment group 2 (negative AIGS) and the control group. We also conducted a DiD linear regression to examine whether negative AIGS lead to an increase in negative emotions in public discussion. The regression included main effects for the treatment (negative AIGS vs. no AIGS), period (before vs. after), and their interaction. The interaction term is not statistically significant, b

= 0.013, $SE = 0.059$, $p = .858$, indicating no significant difference in sentiment change between the treatment and control groups over time.

We further conducted two robustness checks, as shown in Appendix Table A.1. The first one controlled the potential differences between hashtags in the treatment and control groups by incorporating Propensity Score Matching (PSM) into the DiD regression. The following variables were used for matching: (a) hashtag length, measured by the number of words in the hashtag; (b) hashtag type, categorized into Entertainment, News, or Lifestyle, aligning with Weibo's classification; (c) hashtag duration, measures the time a hashtag remained on the hot search list (in minutes); and (d) hashtag popularity, based on Weibo's index reflecting user engagement through searches, posts, forwards, and interactions. The rationale for including each variable is as follows: hashtag length indicates the textual richness of the information conveyed by the hashtag itself, which is often related to sentiment; hashtag type reflects the nature of the event being discussed, which can shape the sentiment of public discussion; hashtag duration and hashtag popularity capture the degree and persistence of public attention associated with each hashtag, which may also correlate with sentiment. Each hashtag in the treatment group was matched with one control hashtag, and only the matched pairs were included in the DiD regression. The results are consistent with our main analysis.

The second robustness check addresses a concern related to our treatment-time definition. We defined treatment time as a hashtag's last appearance on the hot search list to ensure the AIGS feature had been applied. Because some treatment-group hashtags appeared more than once, there may have been prior exposure before the last appearance. Thus, we re-estimated the model after excluding these hashtags, and the results are consistent with the main analysis.

Discussion

The results of RQ1a and RQ1b reveal that Weibo GenAI systematically overrepresents positive emotions when producing AIGS from UGC. We empirically found that both the sampling and summarizing processes are sources of AIGS bias. Weibo GenAI tends to filter in posts with positive emotions in the sampling process and further amplifies this positivity in the summarizing process, producing AIGS that overrepresent positive public sentiment. Our observations of collected data indicate that Weibo GenAI tends to select posts by influential users as reference posts, such as those with large follower bases or verified institutional accounts. While this pattern offers a direct

explanation for the overrepresentation of positivity, the AI bias observed in our case is also deeply rooted in the unique social, political, cultural, and institutional context (Afreen et al., 2025; Mehrabi et al., 2022), which profoundly shapes the training data, model design, and implementation of Weibo GenAI in the Chinese context. In China, the government has a long tradition of regulating online discussions, exemplified by initiatives that promote “positive energy” and a “clear and bright cyberspace” (Yang & Tang, 2018). Although the primary aim of this study is to examine the existence of bias introduced in the sampling and summarizing processes, these structural factors and sociopolitical tendencies provide a critical background for a deeper understanding of the overrepresented positivity in AIGS in the Chinese social media and public discussion context.

By demonstrating the existence and examining the sources of AIGS bias, this work makes several contributions to the growing body of research on AI bias. First, the sources of AI bias in prior studies are mostly categorized into data, model design, or implementation (AlMakinah et al., 2025; Mehrabi et al., 2022). This study extends the theoretical framework by focusing on new forms of bias that emerge from the sampling and summarizing algorithmic processes through which AIGS are produced from large-scale UGC. Second, most studies on AI bias to date have been theoretical discussions on its sources, consequences, and solutions (Ferrara, 2023b). This work provides empirical evidence of new sources by showing that both the sampling and summarizing processes introduce bias. Third, prior empirical studies mostly focused on the biased demographic representations (e.g., gender, race) in GenAI outputs (Currie et al., 2024; Kekez et al., 2025; Sun et al., 2023). We focused on a different type of bias in representing public opinions, which is related to personal expression that has the potential to influence public discussions. Lastly, prior explorations on the effect of AIGS on user behavior have focused on video-based social media or E-commerce platforms (Alavi & Nozari, 2025; Kim et al., 2024). Our study extends this investigation to social media, where AIGS are used to summarize public opinions, potentially shaping people’s perception of and actions towards social issues.

While results suggest that Weibo AIGS are biased in representing public sentiment, our exploration for RQ2 found no observable effect of AIGS sentiments on subsequent public sentiments. This finding may be attributed to users’ actual engagement with AIGS and their selective perception of AIGS content. To begin with, the introduction of the AIGS feature to the platform does not guarantee users’ actual engagement with it. Users may pay limited attention to AIGS or choose not to interact with them. Moreover, users’ mo-

tivation for using the platform may influence their engagement with AIGS. From a use and gratification (U&G) approach, E-commerce users are mostly motivated by functional needs like efficient purchase decision-making (Krasnikolakis, 2022). In contrast, people use social media mostly for information on ongoing issues, social interaction, entertainment, passing time, etc. (Dolan et al., 2016; Whiting & Williams, 2013). Thus, E-commerce users may engage with AIGS to reduce cognitive efforts and save time from reading individual UGC, while social media users may spend time engaging with individual UGC and other users. Thus, social media users' engagement with alternative information sources beyond AIGS (e.g., peers, influencers) could reduce the importance of Weibo AIGS as an information source in shaping users' opinions and expression.

Moreover, even among users who engage with AIGS, selective perception may occur when they process AIGS content. Selective perception is people's tendency to process new information based on what they are already mentally prepared to perceive, with their perception shaped more by internal expectations than by the information itself (Dearborn & Simon, 1958). Although AIGS overrepresent positive sentiment, they still aggregate and present some different content from public discussion. When engaging with AIGS, users may selectively process and perceive the content that aligns with their preexisting beliefs or attitudes.

While the results suggest that AIGS alone are insufficient to significantly influence public sentiment, this study contributes to research on AIGS and user behaviors by offering context-specific insights. Prior studies have primarily focused on video-based social media and e-commerce platforms (Alavi & Nozari, 2025; Kim et al., 2024), whereas our study extends the investigation to AIGS in public discussions on social media, where AIGS summarize public opinions and may shape people's perceptions of social issues.

Conclusion, Limitations, and Future Research

Focusing on AIGS of UGC in representing public sentiment, this study empirically found that both the sampling and summarizing processes introduce bias in AIGS. In our case, GenAI tends to favor positive content during the sampling process and further amplify this positivity during the summarizing process, leading to an overrepresentation of positive public sentiment in AIGS. Although the results indicate that AIGS alone are insufficient to significantly influence subsequent public sentiment, our empirical evidence shows that GenAI filters and amplifies positive emotions in AIGS, raising critical concerns about this new source of GenAI bias when implementing

AI to generate summaries from large volumes of UGC in the public online spaces. As the AIGS feature is increasingly introduced, developed, and established on various digital platforms, it is important to continuously audit GenAI biases and their outputs, as well as monitor the potential impact of biased AIGS on user behaviors.

This study has limitations. First, our data were collected in November 2024, which was a relatively early phase of Weibo AIGS development, during which this feature was only available on the mobile app and activated for a given hashtag after achieving a number of users' requests. Additionally, the visual presentation of AIGS on Weibo is less prominent than on other platforms. For instance, on Amazon, AIGS appear at the top of the review section, whereas on Weibo, they are embedded within the content feed when users search for a hashtag (Figure 2). This placement likely reduces users' attention to and engagement with AIGS, thereby limiting their potential influence on subsequent user behavior. As of May 2025, AIGS has been launched on both the mobile and web versions of Weibo, expanding its presence on the platform. On the web version, AIGS now appears as one of the clickable tabs displayed when users search for a hashtag. The stage of Weibo AIGS development may influence its adoption and perceived importance among users at the platform level. Consequently, the influence of AIGS on public opinion is likely to evolve as the feature becomes more established and widely adopted on Weibo. Future research could replicate this study on Weibo when its AIGS feature is more established. Second, the data collection occurred several weeks after November 2024, when some users may have deleted their posts, resulting in a small number of invalid posts. However, the proportion of invalid posts was minimal compared to the overall dataset and is unlikely to have affected data quality. Lastly, we used hashtags without AIGS as the control group, collected after the AIGS feature was introduced to Weibo. This design may be affected by confounding factors. While using a control group from before the system-wide launch would provide a cleaner comparison, we chose the current approach to minimize the impact of external time-related confounders.

This study serves as a starting point for future studies to empirically explore bias in AIGS and its impact on public discussion. Future research could extend the examination of the sampling and summarizing processes of producing AIGS to other platforms beyond Weibo to generalize the insights derived from this study. Additionally, this study focused on the biased representation of public sentiments and its impact. Future research could investigate other aspects of AIGS beyond emotion, such as topic richness

and diversity in public discussion, and examine how these biases influence the dynamics of public opinion expression. Moreover, this study relies on observational data to reveal behavioral patterns. However, we did not explore the underlying mechanism of how users process AIGS and respond to them. Nor did we investigate how social, political, and cultural factors shape the design and implementation of GenAI at the institutional level. Thus, in addition to analyzing behavioral traces, future research could explore the psychological processes underlying user behaviors through laboratory experiments or investigate how structural factors shape the design and deployment of Weibo GenAI through interviews with platform developers and workers. Overall, this exploratory study is an initial step for future research to further investigate the adoption of GenAI as a communicative, public-facing information source in the public information sphere.

References

- Afreen, J., Mohaghegh, M., & Dobarjeh, M. (2025). Systematic literature review on bias mitigation in generative AI. *AI and Ethics*, 5(5), 4789–4841. <https://doi.org/10.1007/s43681-025-00721-9>
- Alavi, S., & Nozari, S. (2025). Algorithmic summaries and content homogenization: The changing landscape of ugc. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5106785>
- AlMakinah, R., Goodarzi, M., Tok, B., & Canbaz, M. A. (2025). Mapping artificial intelligence bias: A network-based framework for analysis and mitigation. *AI and Ethics*, 5(3), 1995–2014. <https://doi.org/10.1007/s43681-024-00609-0>
- Currie, G., John, G., & Hewis, J. (2024). Gender and ethnicity bias in generative artificial intelligence text-to-image depiction of pharmacists. *International Journal of Pharmacy Practice*, 32(6), 524–531. <https://doi.org/10.1093/ijpp/riae049>
- Dash, A., Shandilya, A., Biswas, A., Ghosh, K., Ghosh, S., & Chakraborty, A. (2019). Summarizing user-generated textual content: Motivation and methods for fairness in algorithmic summaries. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–28. <https://doi.org/10.1145/3359274>
- Dearborn, D. C., & Simon, H. A. (1958). Selective perception: A note on the departmental identifications of executives. *Sociometry*, 21(2), 140. <https://doi.org/10.2307/2785898>
- Deshpande, K. V., Pan, S., & Foulds, J. R. (2020). Mitigating demographic bias in ai-based resume filtering. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 268–275. <https://doi.org/10.1145/3386392.3399569>
- Dolan, R., Conduit, J., Fahy, J., & Goodman, S. (2016). Social media engagement behaviour: A uses and gratifications perspective. *Journal of Strategic Marketing*, 24(3-4), 261–277. <https://doi.org/10.1080/0965254X.2015.1095222>

- Dong, X., & Lian, Y. (2021). A review of social media-based public opinion analyses: Challenges and recommendations. *Technology in Society*, 67, 101724. <https://doi.org/10.1016/j.techsoc.2021.101724>
- Ferrara, E. (2023a). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1), 3. <https://doi.org/10.3390/sci6010003>
- Ferrara, E. (2023b). Should chatgpt be biased? Challenges and risks of bias in large language models [arXiv:2304.03738]. <https://doi.org/10.48550/arXiv.2304.03738>
- Ferrara, E., & Yang, Z. (2015). Measuring emotional contagion in social media (C. T. Bauch, Ed.). *PLOS ONE*, 10(11), e0142390. <https://doi.org/10.1371/journal.pone.0142390>
- Fletcher, R. R., Nakeshimana, A., & Olubeko, O. (2021). Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. *Frontiers in Artificial Intelligence*, 3, 561802. <https://doi.org/10.3389/frai.2020.561802>
- Goldenberg, A., & Gross, J. J. (2020). Digital emotion contagion. *Trends in Cognitive Sciences*, 24(4), 316–328. <https://doi.org/10.1016/j.tics.2020.01.009>
- Gorska, A. M., & Jemielniak, D. (2023). The invisible women: Uncovering gender bias in AI-generated images of professionals. *Feminist Media Studies*, 23(8), 4370–4375. <https://doi.org/10.1080/14680777.2023.2263659>
- Gray, M., Samala, R., Liu, Q., Skiles, D., Xu, J., Tong, W., & Wu, L. (2024). Measurement and mitigation of bias in artificial intelligence: A narrative literature review for regulatory science. *Clinical Pharmacology & Therapeutics*, 115(4), 687–697. <https://doi.org/10.1002/cpt.3117>
- Kadiresan, A., Baweja, Y., & Ogbanufe, O. (2022). Bias in ai-based decision-making. In M. V. Albert, L. Lin, M. J. Spector, & L. S. Dunn (Eds.), *Bridging Human Intelligence and Artificial Intelligence* (pp. 275–285). Springer International Publishing. https://doi.org/10.1007/978-3-030-84729-6_19
- Kekez, I., Lauwaert, L., & Begičević Redep, N. (2025). Is artificial intelligence (Ai) research biased and conceptually vague? A systematic review of research on bias and discrimination in the context of using AI in human resource management. *Technology in Society*, 81, 102818. <https://doi.org/10.1016/j.techsoc.2025.102818>
- Kim, A., Lu, Y., Ma, T., & Tan, Y. (2024). Less is more? Impact of ai-generated summaries on user engagement of video-sharing platforms. <https://doi.org/10.2139/ssrn.5020305>
- Kirk, H. R., Jun, Y., Volpin, F., Iqbal, H., Benussi, E., Dreyer, F., Shtedritski, A., & Asano, Y. (2021). Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models. *Advances in neural information processing systems*, 34, 2611–2624.
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790. <https://doi.org/10.1073/pnas.1320040111>

- Krasnikolakis, I. (2022). Uses and gratifications theory in e-commerce: Habit and social presence. *EURASIAN JOURNAL OF BUSINESS AND MANAGEMENT*, 10(1), 1–18. <https://doi.org/10.15604/ejbm.2022.10.01.001>
- Landers, R. N., & Behrend, T. S. (2023). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, 78(1), 36–49. <https://doi.org/10.1037/amp0000972>
- Liu, W., Zhang, S., Zhang, T., Gu, Q., Han, W., & Zhu, Y. (2024). The AI empathy effect: A mechanism of emotional contagion. *Journal of Hospitality Marketing & Management*, 33(6), 703–734. <https://doi.org/10.1080/19368623.2024.2315954>
- Mayuravaani, M., Ramanan, A., Perera, M., Senanayake, D. A., & Vidanaarachchi, R. (2024). Insights into artificial intelligence bias: Implications for agriculture. *Digital Society*, 3(3), 48. <https://doi.org/10.1007/s44206-024-00142-x>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35. <https://doi.org/10.1145/3457607>
- Mienye, I. D., Obaido, G., Emmanuel, I. D., & Ajani, A. A. (2024). A survey of bias and fairness in healthcare ai. *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, 642–650. <https://doi.org/10.1109/ICHI61247.2024.00103>
- Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Gender bias in AI-based decision-making systems: A systematic literature review. *Australasian Journal of Information Systems*, 26. <https://doi.org/10.3127/ajis.v26i0.3835>
- Naskar, D., Singh, S. R., Kumar, D., Nandi, S., & Rivaherrera, E. O. D. L. (2020). Emotion dynamics of public opinions on twitter. *ACM Transactions on Information Systems*, 38(2), 1–24. <https://doi.org/10.1145/3379340>
- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdil, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulous, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., ... Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1356. <https://doi.org/10.1002/widm.1356>
- Ouyang, C., & Xu, D. (2024). Does AI video summary help? The effects of AI-generated video summary. *ACIS 2024 Proceedings*. <https://aisel.aisnet.org/acis2024/120>
- Özcan, S. (2020). Tweet-preprocessor 0.6.0. Tweet-Preprocessor 0.6.0. Retrieved November 22, 2025, from <https://github.com/s/preprocessor>
- Prinz, K. (2022). *The smiling chatbot: Investigating emotional contagion in human-to-chatbot service interactions* (1st ed). Springer Fachmedien Wiesbaden GmbH.
- Reed, C., Wynn, M., & Bown, R. (2025). Artificial intelligence in digital marketing: Towards an analytical framework for revealing and mitigating bias. *Big Data and Cognitive Computing*, 9(2), 40. <https://doi.org/10.3390/bdcc9020040>
- Reyero Lobo, P., Daga, E., Alani, H., & Fernandez, M. (2023). Semantic Web technologies and bias in artificial intelligence: A systematic literature review (D. Gromann, Ed.). *Semantic Web*, 14(4), 745–770. <https://doi.org/10.3233/SW-223041>

- Sasaki, W., Nishiyama, Y., Okoshi, T., & Nakazawa, J. (2021). Investigating the occurrence of selfie-based emotional contagion over social network. *Social Network Analysis and Mining*, *11*(1), 8. <https://doi.org/10.1007/s13278-020-00712-0>
- Su, Y., Wang, Q., Qiu, L., & Chen, R. (2024). Navigating the sea of reviews: Unveiling the effects of introducing ai-generated summaries in e-commerce. <https://doi.org/10.2139/ssrn.4872205>
- Sun, L., Wei, M., Sun, Y., Suh, Y. J., Shen, L., & Yang, S. (2023). Smiling women pitching down: Auditing representational and presentational gender biases in image-generative AI (S. González-Bailón & E.-Á. Horvát, Eds.). *Journal of Computer-Mediated Communication*, *29*(1), zmad045. <https://doi.org/10.1093/jcmc/zmad045>
- Tejani, A. S., Ng, Y. S., Xi, Y., & Rayan, J. C. (2024). Understanding and mitigating bias in imaging artificial intelligence. *RadioGraphics*, *44*(5), e230067. <https://doi.org/10.1148/rg.230067>
- Tseng, A. (2024). Agentlans/multilingual-e5-small-aligned-sentiment · hugging face. Retrieved November 22, 2025, from <https://huggingface.co/agentlans/multilingual-e5-small-aligned-sentiment>
- Wei, O. R., Zhou, J., Hong, W., & Hui, K.-L. (2025). The impact of ai summaries on creators and viewers of video-sharing platforms. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5213073>
- Whiting, A., & Williams, D. (2013). Why people use social media: A uses and gratifications approach. *Qualitative Market Research: An International Journal*, *16*(4), 362–369. <https://doi.org/10.1108/QMR-06-2013-0041>
- Yang, P., & Tang, L. (2018). “Positive energy”: Hegemonic intervention and online media discourse in china’s xi jinpings era. *China: An International Journal*, *16*(1), 1–22. <https://doi.org/10.1353/chn.2018.0000>
- Yang, W., & Xie, Y. (2024). Can robots elicit empathy? The effects of social robots’ appearance on emotional contagion. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100049. <https://doi.org/10.1016/j.chbah.2024.100049>
- Yuan, L., Chan, F. Y., Gao, C., Leung, A., Gu, B., & Ye, Q. (2025). Ai-generated summaries as differentiating reference: Impact on user content generation in online communities. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5180947>
- Zhang, J. (2024). Weibo’s Wang Wei: The “Zhiwei” large model delivers a Weibo-specific application experience for users. Retrieved November 22, 2025, from <https://tech.chinadaily.com.cn/a/202409/23/WS66f11b5ba310a792b3abd8fb.html>
- Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). Bias in generative ai [arXiv:2403.02726]. <https://doi.org/10.48550/arXiv.2403.02726>

Appendix

Table A.1: *Robustness checks on the effect of AIGS on Public Sentiments*

	Robustness Check 1		Robustness Check 2	
	Treatment group 1: Positive AIGS	Treatment group 2: Negative AIGS	Treatment group 1: Positive AIGS	Treatment group 2: Negative AIGS
Treatment	0.264 (0.029)***	-0.918 (0.047)***	0.484 (0.030)***	-1.125 (0.046)***
Period	0.002 (0.029)	0.017 (0.047)	0.003 (0.025)	0.003 (0.026)
Treatment × Period	-0.040 (0.041)	0.000 (0.067)	-0.045 (0.043)	0.007 (0.065)
N	6,156	2,484	8,160	6,492

Table A.1 demonstrates two robustness checks on the effect of AIGS on public sentiment. Robustness check 1 controlled the potential differences between hashtags in the treatment and control groups by incorporating Propensity Score Matching into the DiD regression. Robustness check 2 excluded the hashtags that appeared on the hot search list more than once to re-estimate the model. The results of both robustness checks are consistent with the main analysis.