

# The Value of AI-Generated Metadata for UGC Platforms: Evidence from a Large-scale Field Experiment

Xinyi Zhang<sup>1</sup>, Chenshuo Sun<sup>2</sup>, Renyu Zhang<sup>3</sup>, Khim-Yong Goh<sup>4</sup>

<sup>1</sup> National University of Singapore, Singapore, xinyizhang@u.nus.edu

<sup>2</sup> Peking University, Beijing, China, csun@stern.nyu.edu

<sup>3</sup> The Chinese University of Hong Kong, Hong Kong, China, philipzhang@cuhk.edu.hk

<sup>4</sup> National University of Singapore, Singapore, gohky@comp.nus.edu.sg

AI-generated content (AIGC), such as advertisement copy, product descriptions, and social media posts, is becoming ubiquitous in business practices. However, the value of AI-generated metadata, such as title, remains unclear on user-generated content (UGC) platforms. To address this gap, we conducted a large-scale field experiment on a leading short-video platform in Asia to provide about 1 million users access to AI-generated titles for their uploaded videos. Our findings show that the provision of AI-generated titles significantly boosted content consumption, increasing valid watches by 1.6% and watch duration by 0.9%. When producers adopted these titles, these increases jumped to 7.1% and 4.1% respectively. This viewership-boost effect was largely attributed to the use of this generative AI (GAI) tool increasing the likelihood of videos having a title by 41.4%. The effect was more pronounced for the groups more affected by metadata sparsity. Mechanism analysis revealed that AI-generated metadata improved user-video matching accuracy in the platform’s recommender system. Interestingly, for a video for which the producer would anyway have posted a title, adopting the AI-generated title will decrease its viewership on average, implying that AI-generated titles may be of lower quality than human-generated ones. However, when producers chose to co-create with GAI and significantly revised the AI-generated titles, the videos outperformed their counterparts with either fully AI-generated or human-generated titles, showcasing the benefits of human-AI co-creation. This study highlights the value of AI-generated metadata and human-AI metadata co-creation in enhancing user-content matching and content consumption for UGC platforms.

*Key words:* Generative AI, Video metadata, User-video matching, Short-video platforms, Human-AI co-creation

---

## 1. Introduction

Generative AI (GAI) and AI-generated content (AIGC) have demonstrated significant values and potentials across industries by efficiently producing high-quality content, such as generating advertising copy, improving customer service, and enhancing media production.<sup>1</sup> These GAI tools help

<sup>1</sup> <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/how-generative-ai-can-boost-consumer-marketing>.

streamline content creation, reduce manual effort, and improve output quality. Market research<sup>2</sup> shows that after adopting GAI, 75% of marketers reported higher content production, while 79% observed improvements in content quality. Such an efficiency boost has also fueled the growth of the GAI market, which was valued at USD 11.6 billion in 2023 and is projected to grow to USD 175.3 billion by 2033, with a compound annual growth rate of 31.2%.

On user-generated content (UGC) platforms such as short-video platforms like TikTok, content metadata, such as titles and tags, often include descriptive information about the content. This metadata provides structured details such as themes and subjects that may allow recommender systems to better categorize and understand content (Wei et al. 2024). Research on YouTube (Hoiles et al. 2017) has shown that improved video metadata correlated with a 25% increase in video search visibility. A Nielsen report indicates that books with complete metadata related to 2.2 times higher sales than those with incomplete metadata.<sup>3</sup>

Despite its potential value, metadata is largely sparse for online platforms. Malik and Tian (2017) observe that only 42,000 (i.e., 0.26%) of 1.6 million YouTube videos had complete metadata. Similarly, a dataset of over 50,000 IMDb movies showed that less than 25% had complete metadata.<sup>4</sup> Other platforms, such as Spotify<sup>5</sup> and TikTok,<sup>6</sup> have reported similar challenges. The metadata sparsity issue is primarily driven by the time-consuming nature of metadata generation for content producers. Additionally, because metadata is less visible by users and, hence, does not directly engage them, producers often undervalue its importance (Peng et al. 2023). Platforms, however, could not mandate metadata generation as it may deter content uploads and eventually harm the platform’s ecosystem.<sup>7</sup>

The recent emergence of GAI technologies has provided UGC platforms with new solutions to such metadata sparsity challenges. Specifically, leveraging AI-generated content (AIGC) to automatically generate metadata such as titles and hashtags could substantially reduce a content creator’s burden to come up with such metadata. This raises an important question of both academic and practical interest: *Does AI-generated metadata deliver value for a UGC platform?* While AI holds promise in generating detailed and relevant metadata (Agrawal et al. 2023, Wei et al. 2024), the answer to this question is not yet conclusive. As discussed above, metadata is instrumental for user-content matching for UGC platforms, but no causal evidence has been reported

<sup>2</sup> <https://market.us/report/generative-ai-in-content-creation-market/>.

<sup>3</sup> [https://pnmais.com/wp-content/uploads/2023/11/Report\\_NielsenBookData\\_MVB\\_Metadata\\_Frankfurt\\_2022.pdf](https://pnmais.com/wp-content/uploads/2023/11/Report_NielsenBookData_MVB_Metadata_Frankfurt_2022.pdf).

<sup>4</sup> <https://www.kaggle.com/datasets/rajugc/imdb-movies-dataset-based-on-genre?resource=download>.

<sup>5</sup> <https://soundcharts.com/blog/music-metadata>.

<sup>6</sup> A dataset of 1,729 TikTok videos indicated that 45% of short videos have missing hashtags. See [https://www.kaggle.com/datasets/vbradculbertson/tiktok-trending-metadata?select=sug\\_users\\_vids1.csv](https://www.kaggle.com/datasets/vbradculbertson/tiktok-trending-metadata?select=sug_users_vids1.csv).

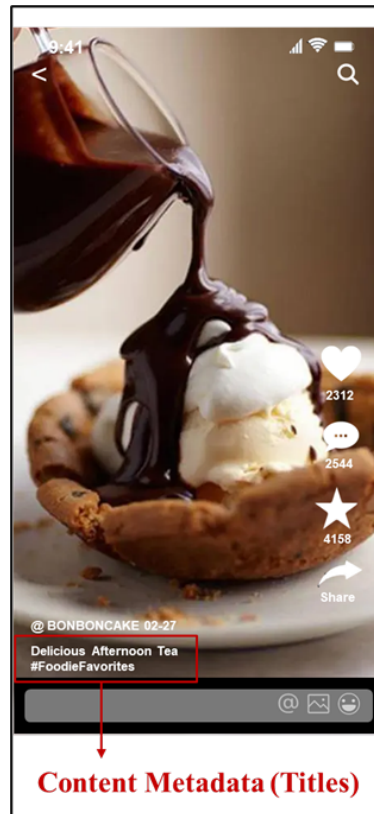
<sup>7</sup> <https://www.multicollab.com/blog/user-generated-content/>.

in the literature to quantify its downstream effects on content consumption in real-world settings. Furthermore, AI-generated metadata could introduce new challenges, such as hallucination, misinformation, stereotypes, and biased responses (Liang et al. 2021, Susarla et al. 2023). Hence, AI may generate misleading metadata that misrepresents the content, resulting in poor user-content matching and disengaging the matched users. Therefore, rigorous empirical research is needed to quantify the actual value of AI-generated metadata for UGC platforms. Existing GAI studies (Chen and Chan 2023, Su et al. 2024, Reisenbichler et al. 2022) have largely focused on the impact of AIGC that directly engages users (e.g., online advertising creatives), with the influence of AI-generated metadata mostly overlooked. Similarly, data augmentation studies (Wei et al. 2024, Ellis et al. 2018) have largely overlooked the potential of GAI to generate video metadata.

Another critical question faced by a UGC platform is: *How should platforms best leverage AI-generated metadata to boost content consumption?* Current GAI research (Chen and Chan 2023, Zhou and Lee 2024) has not provided conclusive evidence on whether AI-generated metadata can fully replace human-generated ones. On one hand, AI may outperform humans by learning from (very) large datasets and identifying patterns overlooked by humans (Chen and Chan 2023), particularly when humans lack the necessary skills or knowledge for metadata generation. On the other hand, GAI may struggle with unique context-specific cases (Longoni et al. 2019, Granulo et al. 2021), where human insights and private information about the content goal and intended audience are crucial for creating context-rich metadata (Sun et al. 2022). Hence, it remains unclear whether AI-generated metadata, human-generated metadata, or a co-creation of both would best work on UGC platforms. The answer to this question is crucial for a platform to make informed operational decisions on deploying AI-generated metadata in its product.

To answer these questions, we collaborated with a leading short-video platform in Asia (hereafter “Platform A”) to conduct a large-scale randomized field experiment. Users on Platform A primarily consume videos recommended by the platform. Video titles are the metadata we focus on and rarely noticed by viewers (see Figure 1 where video titles were small and positioned at the bottom left corner of the viewers’ interface<sup>8</sup>). This metadata is a crucial part of the video profile data used in the recommender system that matches users with relevant videos on Platform A. However, like many UGC platforms, Platform A faces the significant challenge of metadata sparsity. In our dataset (see Section 3.3 for details), only 60.7% of videos had titles during the pre-treatment period. In response to this challenge, Platform A developed a GAI tool to automatically generate video titles. Specifically, the platform sampled videos with well-matched titles to build a training set,

<sup>8</sup> A similar example can be seen on social media platforms like Instagram, where titles and tags are often hidden and require users to click “expand” to view them. Similarly, on e-commerce platforms like Amazon or Taobao, product specifications are often tucked away in dropdown menus, making them less visible unless being actively searched for.



**Figure 1 Viewer Interface for Watching Videos on Platform A**

which was then used to fine-tune a multi-modal large-language model (LLM) similar to GPT-4. Once a user uploaded a new video, the GAI tool would capture several frames from the video, extract any text in these frames, and process these combined visual and textual elements as inputs to generate a title.

In our experiment, users were randomly assigned to either the treatment or the control condition. Content producers in the treatment group received access to AI-generated titles on the video posting page, and could adjust the titles as needed, whereas producers in the control group could only write titles by themselves. Our study covered the treatment period from August 8<sup>th</sup> to 21<sup>st</sup>, 2023 and included 2,048,033 producers, each of whom posted at least one video.

The analyses of our experiment results yield several important insights. We present three findings about **the effects of AI-generated metadata on content consumption**. First, we find that access to AI-generated video titles significantly boosted video consumption, increasing valid watches<sup>9</sup> by 1.6%, and watch duration by 0.9%. When producers adopted these titles, the increase jumped to 7.1% and 4.1% respectively. Second, this viewership-boosting effect was likely

<sup>9</sup>For videos between 3 and less than 7 seconds, a viewer watch is valid if the watch duration matches the video duration. For videos of 7 seconds or longer, this count is recorded if the watch duration is at least 7 seconds. This metric is designed by the platform to measure the consumption of a video.

due to reduced title sparsity on Platform A. We find that access to AI-generated titles increased the likelihood of a video having a title by 41.4% and tags by 72.4%. Third, AI-generated titles disproportionately benefited hedonic-content videos (e.g., personal vlogs) and low-skilled content producers more due to their originally sparse video titles. Specifically, utilitarian-content videos (e.g., news and reviews) in the treatment group saw a relative decrease of 3.1% in valid watches and 3.0% in watch duration compared to hedonic-content videos. In contrast, low-skilled producers experienced an additional increase of 1.6% in valid watches, and 1.3% in watch duration.

Next, we examined the **mechanism for AI-generated metadata enhancing content consumption**. Prior GAI research (e.g., [Chen and Chan 2023](#)) mostly focuses on prominently displayed AI-generated content that directly engages users. Hence, the mechanisms in these settings cannot be directly applied to explain our results on AI-generated metadata, which are rarely noticed by viewers. Building on studies of data augmentation in recommender systems ([Wei et al. 2024](#)), we propose that the enriched metadata helped the platform’s recommender system better understand video content and more accurately match videos to fitted users. To validate this hypothesis, we analyzed an additional dataset of 93,618,096 recommendation sessions of the videos produced during our experiment. With this new dataset, we show that the Areas Under the ROC Curve (AUCs) to predict a user’s engagement behaviors such as liking, sharing, and following<sup>10</sup> are significantly higher ( $p < 0.01$ ) for treatment videos than for control videos. These findings confirm that AI-generated metadata indeed addressed sparse metadata issues, improved user-video matching accuracy, and ultimately drove higher video consumption and engagement.

Finally, we examined **how content producers should most effectively co-create with AI-generated metadata**. Interestingly, we find that when videos already had human-generated titles, access to AI-generated titles decreased content consumption, with declines of 37.9% in valid watches and 32.6% in watch duration. This suggests that AI-generated titles were generally of lower quality than existing human-generated ones. However, when producers chose to co-create with AI to significantly revise AI-generated titles, content consumption would improve. Specifically, each 10% decrease in textual similarity between AI-generated titles and actual titles adopted by the producers increased the video valid watch by 9.8% in valid watches and watch duration by 9.2%. Moreover, lower similarity scores were also associated with richer linguistic attributes, where a 10% decrease in similarity correlated with a 4.8% increase in lexical density, 3.7% in lexical variation, and 1.1% in entropy. These findings highlight the value of producers and AI co-creating metadata on a UGC platform. Additionally, we surveyed 1,925 treatment group users with open-ended questions on the

<sup>10</sup> While our main analysis focuses on viewership outcomes (e.g., valid watch and watch duration), this additional dataset lacks predicted probabilities for these measures. Instead, it includes predictions for downstream engagement behaviors such as liking, sharing, and following. As these behaviors occur at later stages of the user journey, their accurate predictions imply that viewership outcomes, which occur earlier, are also likely to be predicted accurately.

usage of AI-generated titles. The qualitative feedback pointed to an “inspiration effect,” where AI-generated titles inspired content producers to create better titles, highlighting the potential for human-AI metadata co-creation to boost content consumption.

In summary, our study leverages AI-generated metadata to improve content-user matching and content consumption on UGC platforms. Our work provides several theoretical and practical contributions. First, we contribute to the research on the economic impact of GAI (Su et al. 2024, Zhou and Lee 2024) by exploring the value of a new type of AI-generated content, namely AI-generated metadata. We uncover a new mechanism where AI-generated content enhances engagement by addressing metadata sparsity and improving user-content matching. We also provide new insights on how human-AI metadata co-creation can further improve the accuracy of this matching process. Second, our findings contribute to the growing literature on platform operations (Filippas et al. 2023, Zeng et al. 2023), shedding light on how to leverage AI-generated metadata to improve user engagement and platform efficiency. Third, we complement the data augmentation literature (Peng et al. 2023, Wei et al. 2024) by empirically documenting GAI’s ability to address metadata sparsity and quantifying its economic impacts. Fourth, our study offers unique large-scale experimental evidence to causally examine the impact of GAI tools on a real-world platform, whereas earlier works are mostly based on observational data or lab-based methods. Lastly, we offer actionable insights for platform managers, emphasizing the importance of using GAI to augment metadata, enhance user-content matching, and ultimately increase content consumption.

The rest of the paper proceeds as follows. Section 2 reviews the relevant literature. Section 3 details our field setting, experimental design, and data. Section 4 presents the effect of AI-generated metadata on content consumption and the underlying mechanism. Section 5 explores how content producers collaborate with AI. Section 6 presents additional analyses and robustness tests. Last, Section 7 discusses the practical implications of our research and directions for future research.

## 2. Literature Review

Our paper speaks to three streams of literature: (1) GAI and its collaborations with humans; (2) platform operations; and (3) data augmentation and recommender system.

**GAI and its Collaborations with Humans.** Our work is most closely connected to research on the economic impact of GAI. Emerging literature has examined its economic impact across various fields including labor market (Liu et al. 2024), firm innovation (Cheng et al. 2022), marketing (Chen and Chan 2023, Su et al. 2024, Reisenbichler et al. 2022), artwork (Zhou and Lee 2024), and knowledge sharing (Burtch et al. 2023).

Our contribution to this literature is threefold. First and most importantly, the nascent literature that examines the effect of GAI tools on user engagement has mostly focused on the context

where AI-generated content prominently interacts with consumers. For example, the AI-generated summaries (AIGS) (Su et al. 2024) are displayed to users on product web pages, reducing information search costs and directly affecting consumer behaviors. Similar cases are observed in Chen and Chan (2023) and Reisenbichler et al. (2022), which studied advertisement copy that directly impacts consumers' purchasing decisions. These studies largely attribute the enhanced responses to the enhanced quality of AI-generated content. However, they have overlooked the potential of GAI to enhance user responses by improving user-content matching through augmented metadata within recommender systems. Our study fills in this research gap by using an innovative experimental design to examine AI-generated metadata, a type of content that is rarely visible to users and does not engage them directly. This design allows us to attribute changes in content consumption primarily to user-content matching rather than user engagement. We thus make a critical contribution by extending the mechanism from enhanced content quality to improved matching accuracy via mitigated metadata sparsity.

Second and method-wise, most prior GAI studies conduct laboratory experiments (Chen and Chan 2023) or treat the launch of ChatGPT or other large language model tools as an event and leverage its timing for econometric identification (usually with Difference-in-Differences method) (e.g., Zhou and Lee 2024, Burtch et al. 2023). Our study enhances these studies by using a randomized field experiment to causally assess the impact of GAI tool availability. This experiment provides the exogenous variation in the input of AI-generated metadata in the platform's recommender system and thus allows us to identify its causal effect on consumption outcomes based on a real-world setting.

Third, our research adds to the growing literature on human-AI collaboration, which has primarily shown that combining human input with AI tools outperforms both full automation and human-only approaches. This has been explored in studies on non-GAI tools (Anthony et al. 2023, Boyacı et al. 2024) and more recent work on GAI tools (Chen and Chan 2023, Zhou and Lee 2024, Wang et al. 2023a). Research on GAI tools has identified two primary modes of content co-creation: human-revised AI-generated content and AI-revised human-generated content (Chen and Chan 2023). These studies often compare the linguistic or visual attributes of the content generated by AI, humans, and their co-creation, to understand how these different approaches impact content consumption (Chen and Chan 2023, Zhou and Lee 2024, Reisenbichler et al. 2022). However, because the AI-generated content in these studies is directly shown to users, their findings mainly explain which features enhance user engagement, leaving the open question of what features and how human input can improve user-content matching, which is critical in recommender systems. For example, while emojis can increase user arousal to boost engagement, their complex symbolic nature can limit recommender systems to effectively match content with targeted users. We speak

to this open question by designing an experiment to explore AI-generated metadata, a type of AI-generated content that does not directly engage users, to offer a clear understanding of how humans can co-create with AI to enhance user-content matching. Our research advances the understanding of content co-creation process, focusing on human-revised AI-generated content by providing empirical evidence of how human modifications to AI-generated metadata can enhance user-content matching.

**Platform Operations.** Our research extends the growing literature that address operation problems on online platforms. This literature has examined how to build effective systems for pricing (Cui et al. 2022, Zhang et al. 2022), reviews systems (Cui et al. 2020a), logistic systems (Bai et al. 2022a), social fairness (Wang et al. 2023b, Clyde et al. 2024), content production (Zeng et al. 2023), advertisement delivery (Ye et al. 2023), and content consumption (Fang et al. 2023). It has also studied how to ensure service quality (Cui et al. 2020b) and participants’ responses to platform interventions (Lysyakov and Viswanathan 2023, Bai et al. 2022b).

We contribute to this stream of research in three ways. First, our work enriches the research that seeks to improve consumers’ content consumption on UGC platforms. Prior studies have explored consumer-side interventions such as content prioritization (Dukes and Liu 2024) and personalized content distribution (Wei et al. 2024), as well as producer-side interventions such as financial incentives (Kuang et al. 2019), performance feedback (Huang et al. 2019), social norms (Burtch et al. 2018), and more recently, AI or GAI tools (He et al. 2021). The producer-side interventions largely focus on improving content quality to boost content consumption. Our research takes this literature a step further by examining the impact of a new operational intervention (i.e., AI-generated metadata) on content consumption outcomes.

Second, our work speaks to the emergent literature that empirically tests the effectiveness of information-based interventions in solving operational problems for online platforms. Examples of prior interventions include providing producers with more information about customers (Buell et al. 2017), services or products (Kesavan and Kushwaha 2020, Sun et al. 2022), and competitors (Cui et al. 2020a, Zeng et al. 2023). These interventions primarily enhance consumer engagement by improving content production in terms of speed, capacity, and quality. Our study contributes to this literature by introducing a new type of information-based intervention—AI-generated metadata that leverages metadata generation to drive content consumption by enhancing user-content matching.

**Data Augmentation and Recommender System.** Recommender system studies have documented that challenges in user-generated video metadata, such as noise, sparsity, and incompleteness, inhibit accurate user-content matching (Wei et al. 2024). To address these issues, data



augmentation techniques are developed to enhance data quality. Solutions for handling data sparsity or missing data include, e.g., fuzziness methods (Choi et al. 2018), imputation via supervised learning (Ellis et al. 2018), active feature-value acquisition (Saar-Tsechansky et al. 2009), and Monte Carlo likelihood estimation (Peng et al. 2023).

We extend this body of literature in two ways. First, prior research has largely overlooked the potential of GAI to enhance metadata in recommender systems, and even when such attempts exist, they primarily focus on algorithmic aspects without validation in real-world settings (e.g., Agrawal et al. 2023, Wei et al. 2024). Our study contribute by providing a large-scale experimental evidence to quantify the economic value of AI-generated metadata in improving content consumption. Second, unlike past studies that retroactively impute missing data through algorithmic estimation, we propose a proactive approach that enables producers to access AI-generated metadata during metadata generation process. Our findings demonstrate that human-AI co-creation further enhances the value of data augmentation through GAI.

### 3. Field Setting, Experiment Design, and Data

#### 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, users on the platform can be either content producers or viewers. Producers post short videos on Platform A to enhance viewership/or engagement and attract new followers, aiming to increase advertising opportunities and revenue. Users visit Platform A either to be entertained by videos that catch their interest (organic browse) or to search for specific videos related to a topic (search-oriented browse).

Viewers can consume videos and engage with others for free on Platform A. They engage with producers mainly through viewership, but they can also like videos, leave comments, forward content to others both on and beyond Platform A, and follow producers for long-term video consumption and engagement. The platform generates revenue primarily through online advertising, i.e., disseminating advertising videos to viewers. Therefore, accurately matching the content with viewers to improve video consumption and engagement is crucial to Platform A’s business model.

Content distribution on Platform A is through two primary channels: organic recommendations and search-query-oriented recommendations. Organic recommendations generate a personalized video feed based on a user’s viewing history and preference, catering to those browsing without active searches. Search-query recommendations, on the other hand, respond directly to users’ text-based searches, tailoring content to match users’ specific queries. Here, the video title, which can include hashtags and descriptions, is the only content metadata used by the recommender system to improve content relevance.

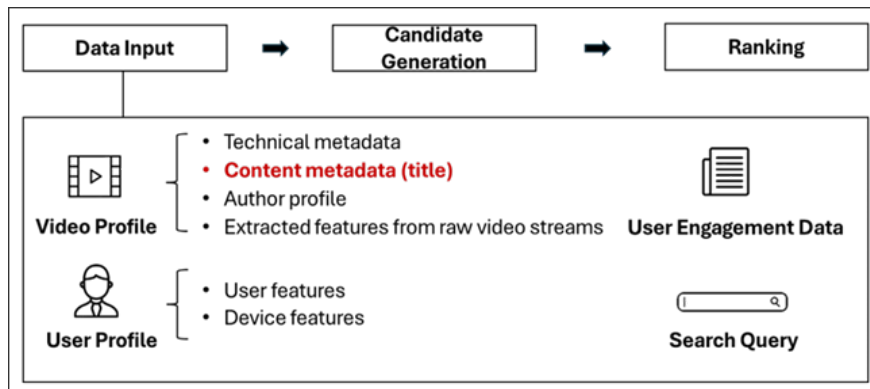


Figure 2 Recommender System Workflow on Platform A

Platform A’s recommender system functions in two stages (see Figure 2): candidate generation and ranking (Davidson et al. 2010). Both stages rely on four types of data inputs: (1) video profile, (2) user profile, (3) user engagement data (e.g., watch count, duration, likes), and (4) search queries (if any). Video profile data includes both technical metadata (e.g., file format, upload date, and video duration) and content metadata (i.e., titles), along with the producer profile and extracted features from processing raw video streams (e.g., video category). User profile data include user features (e.g., age and gender) and device features (e.g., device model). In the candidate generation stage, algorithms such as content-based, collaborative filtering, and context-aware methods are used to select relevant video candidates based on user input. A common approach is to pick videos that are closely related to the ones previously watched by the user, utilizing techniques like co-visitation counts and matrix factorization. The ranking stage then takes these candidates and prioritizes them based on the likelihood of user engagement. When posting videos, producers are encouraged to add titles with hashtags or descriptions in the title-setting box (see Figure 3(a)). There is no word limit for titles and a producer can also leave it blank, posting a video without a title. Allowing this flexibility helps maintain high upload rates, as mandatory title requirements could complicate the posting process and discourage participation.<sup>11</sup>

Video content metadata (titles) is a crucial part of video profile information in Platform A’s recommender system. Unlike technical metadata (e.g., upload times), which is automatically extracted by Platform A, titles require user input. Although the platform can still recommend videos using other inputs, titles provide more specific context than the broad insights generated by extracted features, such as visual patterns, audio cues, and category classifications.<sup>12</sup> These features, while useful, often lack the precise contextual information required for accurate recommendations.

<sup>11</sup> See <https://www.multicollab.com/blog/user-generated-content/>.

<sup>12</sup> Rather than directly processing raw video streams, which is computationally intensive, Platform A uses extracted features from these streams in its recommender system. These features are generated by a separate algorithm within the platform. This design can provide a more efficient video recommendation process.

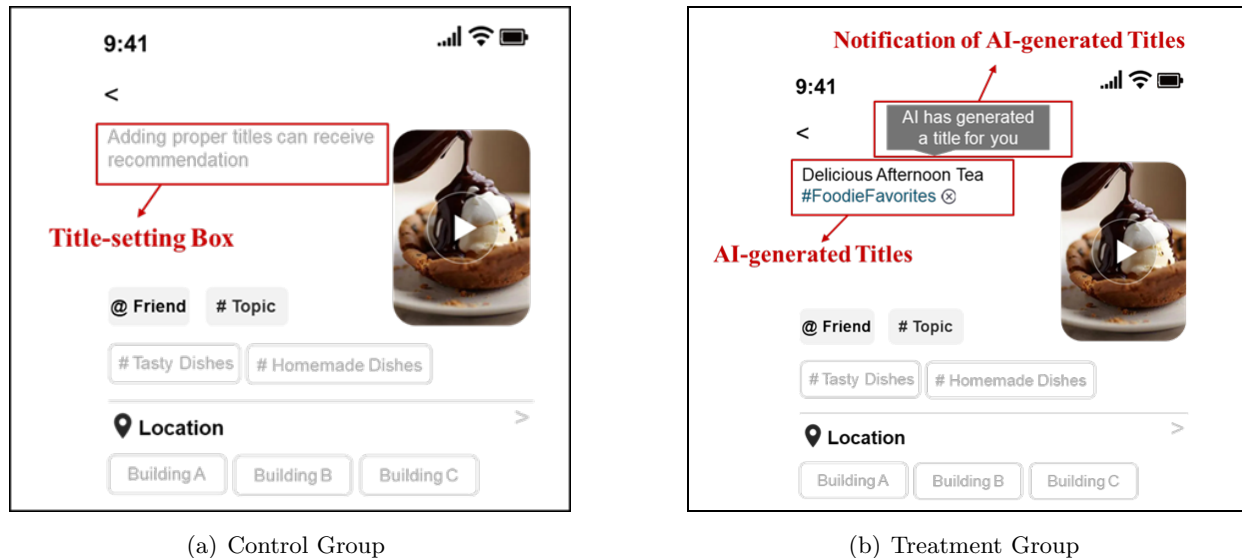


Figure 3 Using AI-generated titles on producers' video posting page

In contrast, video titles offer structured and concise details about a video's themes and objects, helping the system better understand content and enhance user-content matching (Panniello et al. 2016). In organic recommendations, for instance, content-based algorithms utilize video titles to recommend videos similar to what users have already watched (Adomavicius et al. 2008). For users who frequently watch videos with titles with keywords “cooking” or “recipes,” the system can recommend other videos with related keywords. In search-query recommendations, titles provide direct text matches to user searches, addressing challenges in cross-modality matching by offering clear textual references that improve accuracy. In our study, organic recommendations drove the vast majority of video discovery, and search-oriented recommendations only accounted for less than 1% of total viewership. Titles are also vital in addressing the cold-start problem, where new videos without engagement data rely heavily on descriptive titles to be categorized and recommended (Wei et al. 2024). Additionally, titles help mitigate the effects of noisy engagement data, ensuring stable and relevant recommendations even when user engagement is inconsistent due to short video life cycles (Davidson et al. 2010).

However, like many UGC platforms, Platform A faces the significant challenge of metadata sparsity. In our dataset (details are presented in Section 3.3), only 60.7% of videos had titles during the pre-treatment period. To address this, Platform A introduced GAI tools in July 2023, leveraging GPT-4, a multimodal model capable of processing both text and images. Leveraging transformers with self-attention mechanisms, GPT-4 produces coherent and relevant text. Platform A fine-tuned this model by curating a manually-selected dataset consisting of videos with well-matched titles as the training set. The fine-tuning process allows the model to learn specific patterns in the relationship between video content (both visual and textual) and their corresponding titles on

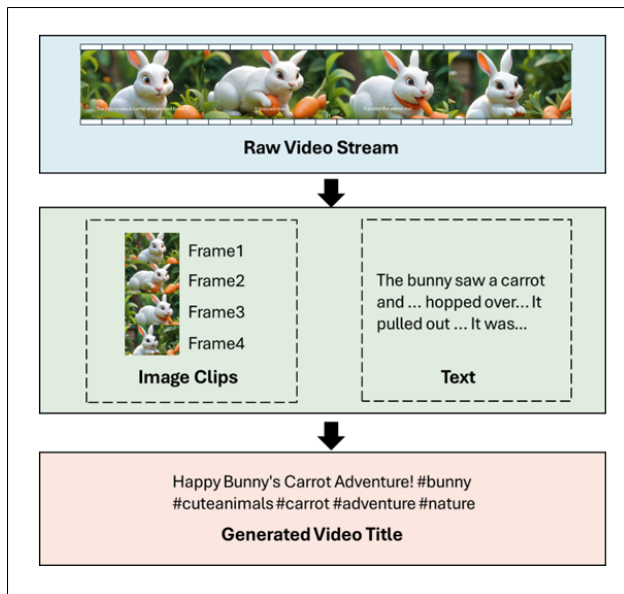


Figure 4 Process of Generating Titles

Platform A. To generate metadata for new videos, as shown in Figure 4, the platform’s GAI tool captured multiple frames from the video stream, extracting visual elements (e.g., key objects or scenes) and any text present in the frames (e.g., subtitles, on-screen text). This combination of visual and textual data was then fed into the fine-tuned model, which produced a title that best reflects the input content. When new videos were uploaded, the GAI tool captured frames, extracted visual elements and any on-screen text, and fed this data into the model to generate relevant titles automatically. This process allows the platform to generate coherent metadata without requiring human input for every video.

### 3.2. Experiment Design

To causally examine the values of AI-generated metadata, we conducted a field experiment on the video posting page to simulate the metadata input change that feeds into the video recommender system. This experiment lasted from July 20<sup>th</sup> to August 21<sup>st</sup>, 2023. Producers involved in our experiment were randomly assigned to the control and treatment groups. Treatment group producers could access an AI-generated title in the title-setting box on the video posting page after they uploaded the video (see Figure 3(b)), and there is a notification next to the AI-generated title indicating that the provided title is generated by AI. In contrast, control group producers could not access such a tool to generate titles via AI (see Figure 3(a)). In addition, in 2023, external GAI tools for generating video titles were unlikely to be widely used since major language models did not support video processing, and few video platforms offered such features. Thus, the risk of contamination, where the control group was unintentionally influenced by the experimental intervention, was limited. Treatment producers had the flexibility to either delete, amend, or fully

adopt the AI-generated title, and regardless of their choices, viewers could not see any indication in their interface of whether the titles were generated by AI. In addition, as shown in Figure 1, video titles were small and positioned at the bottom left corner of the viewers’ interface, so we assumed that any observed changes in video consumption and engagement between the treatment and control groups were unlikely to be driven by changes in the visibility of the titles to viewers.

### 3.3. Data and Variables

Due to some technical issues, Platform A only stored the AI-generated titles between August 8<sup>th</sup> and August 21<sup>st</sup> during the experiment. Thus, our dataset was segmented into two periods: (1) pre-treatment period: July 10<sup>th</sup> to July 19<sup>th</sup> and (2) treatment period: August 8<sup>th</sup> to August 21<sup>st</sup>. Our study included 2,048,033 producers who posted at least one video during our treatment period, with 1,024,940 in the treatment group and 1,023,093 in the control group. During the treatment period, producers in the treatment group uploaded 5,377,560 videos, while those in the control group posted 5,361,424 videos. During the pre-treatment period, only 60.7% of videos had titles, which indicates that metadata sparsity is prevalent on Platform A. For each producer, we obtained data for video viewership outcomes, producer characteristics, and video characteristics. To accommodate variations in video posting times during the treatment period, we calculated the cumulative viewership outcomes for each video over the first two weeks after its posting (Zeng et al. 2023).

Table 1 presents the summary of the variables used in our analysis. Our independent variables are the treatment group dummy ( $Treat_i$ ), coded as 1 if producer  $i$  was assigned to the treatment group, and a binary indicator ( $Adopt_{i,j}$ ) for whether producer  $i$  adopted an AI-generated title for video  $j$ , coded as 1 if the posted video title exactly matched the AI-generated title. Dependent variables, which capture viewers’ video consumption, include viewership metrics such as the number of valid watches ( $ValidWatch_{i,j}$ ), and viewers’ total watch duration in minutes ( $WatchDuration_{i,j}$ ). To analyze the heterogeneous effects of AI-generated titles, we used two moderator variables:  $Utilitarian_{i,j}$  and  $LowSkill_i$ .  $Utilitarian_{i,j}$  indicate whether video  $j$  of producer  $i$  was a utilitarian-content video, coded as 1 for know-how and news categories<sup>13</sup>, and 0 otherwise.  $LowSkill_{i,j}$  indicate low-skilled producers, coded as 1 if producer  $i$ ’s number of followers was below the median (420 followers)<sup>14</sup>, and 0 otherwise. .

<sup>13</sup> Know-how categories generally include educational or instructional videos aimed at elucidating practical skills and knowledge. News categories generally include political news and current affairs. The category classifications are developed by Platform A.

<sup>14</sup> We varied this threshold by applying the 30<sup>th</sup>, 40<sup>th</sup>, 60<sup>th</sup>, and 70<sup>th</sup> percentiles as alternative cutoffs, and the results remained qualitatively consistent. We also employed alternative measurements, such as whether the cumulative number of videos uploaded by producer  $i$  exceeds the median (more details are provided in Section 6).

To account for various confounding factors, we include a set of control variables ( $Controls_{ij}$ ) reflecting producer, video, and day attributes as follows. First, we include producers’ follower counts and a dummy indicating whether they were key opinion leaders to control producer popularity. Second, we include producers’ tenure on the platform (in years), the number of users they followed, and a dummy for multi-homing presence on the focal and rival platforms to control for producers’ experience and/or expertise level. Third, we include gender and provincial location dummies to control for producers’ demographic and geographic variations. Fourth, we include dummy variables indicating whether the video was publicly visible and whether the video was composed of clips or images to account for video type. Fifth, we include the video’s duration and a binary indicator of whether the producer manually sets a video cover to control for video quality. Sixth, we include dummies for video posting dates and categories to control for temporal and categorical variations. Table 2 presents the summary statistics of our focal variables. A correlation matrix of these variables is shown in Table 16 of Online Appendix A. To protect Platform A’s sensitive information,<sup>15</sup> the mean values of four key variables ( $ValidWatch$ ,  $WatchDuration$ ,  $Follower$ , and  $Following$ ) presented in the tables have been scaled by multiplying some positive numbers.

**Table 1 Variable Definitions**

Variable	Description
$Treat_{ij}$	Coded as 1 if producer $i$ of video $j$ was assigned to the treatment group, or else as 0.
$Adopt_{ij}$	Coded as 1 if the title of video $j$ posted by producer $i$ exactly matches the AI-generated title.
$ValidWatch_{ij}$	Number of valid watches for video $j$ of producer $i$ .
$WatchDuration_{ij}$	Viewers’ total watch duration (in minutes) for video $j$ of producer $i$ .
$Utilitarian_{ij}$	Coded as 1 if video $j$ of producer $i$ is a utilitarian-content video, or else as 0.
$LowSkill_i$	Coded as 1 if producer $i$ is a low-skilled producer, or else as 0.
$Follower_i$	Number of users that follow producer $i$ .
$KOL_i$	Coded as 1 if producer $i$ is a key opinion leader, or else as 0.
$Experience_i$	Tenure of producer $i$ (in years) on the platform.
$Following_i$	Number of users that producer $i$ follows.
$Multihome_i$	Coded as 1 if producer $i$ multihomes on other short video platforms, or else as 0.
$Female_i$	Coded as 1 if producer $i$ is female, or else as 0.
$Province_i$	Province location dummies for producer $i$ .
$PublicVisible_{ij}$	Coded as 1 if video $j$ is publicly visible.
$VideoDuration_{ij}$	Duration of video $j$ (in minutes).
$Cover_{ij}$	Coded as 1 if producer $i$ manually sets a video cover for video $j$ .
$ContentType_{ij}$	Coded as 1 if video $j$ is composed of videos (i.e., not images), or else as 0.
$PostDate_{ij}$	Video posting date dummies.
$Category_{ij}$	First-level category dummies for video $j$ of producer $i$ .

*Notes:* All variables are coded as described in the table. Video metadata variables are collected from the platform’s system logs. The variables  $WatchDuration_{ij}$  and  $ValidWatch_{ij}$  measure engagement metrics.

<sup>15</sup> The authors have a Non-Disclosure Agreement with Platform A.

**Table 2** Summary Statistics of Focal Variables

Variable	Mean	SD	Min	Max
<i>Treat</i>	0.501	0.500	0	1
<i>Adopt</i>	0.117	0.322	0	1
<i>ValidWatch</i>	236.543	8,718.901	0	10,504,166
<i>WatchDuration</i>	249.240	10,995.043	0	16,021,141.428
<i>Utilitarian</i>	0.067	0.251	0	1
<i>LowSkill</i>	0.500	0.500	0	1
<i>Follower</i>	477.066	649.083	0	5,133
<i>KOL</i>	0.005	0.073	0	1
<i>Experience</i>	2.209	1.763	0.003	5.631
<i>Following</i>	2,940.926	50,989.248	0	37,260,268
<i>Multihome</i>	0.732	0.443	0	1
<i>Female</i>	0.649	0.477	0	1
<i>PublicVisible</i>	0.921	0.270	0	1
<i>VideoDuration</i>	0.466	1.020	0	43.532
<i>Cover</i>	0.137	0.344	0	1
<i>ContentType</i>	0.744	0.437	0	1

*Notes:* All variables are calculated based on video-level and producer-level data. SD stands for standard deviation, and Min and Max represent the minimum and maximum values observed for each variable. Values for *ValidWatch*, *WatchDuration*, *Follower*, and *Following* have been scaled.

### 3.4. Randomization Check

To verify the randomization effectiveness, we compared treatment producers ( $N=1,024,940$ ) and control producers ( $N=1,023,093$ ) on their pre-treatment video engagement outcomes, producer characteristics, and video attributes. The results of pairwise  $t$ -tests in Table 3 show no significant differences between treatment and control groups on these observable attributes. These results confirm that the treatment and control producers in our sample were comparable, suggesting that any difference between conditions after the experiment started should be attributed to our experimental manipulation—that is, whether producers had access to and/or adopted AI-generated titles.

## 4. Effects of AI-generated Metadata on Content Consumption

Our investigation began by examining the effects of AI-generated metadata (i.e., titles) on the producers’ content consumption outcomes of posted videos. Motivated by past studies (Huang et al. 2021; Sun et al. 2019), we aimed to study two types of causal effects: (1) the effect of treatment (i.e., access to AI-generated titles) on video viewership (intention-to-treat effect, ITT); and (2) the effect of treatment-induced adoptions (i.e., adoption of AI-generated titles) on video viewership (local average treatment effect, LATE). Our unit of analysis was at the producer-video level to capture changes in viewership outcomes for each video uploaded by producers.

**Table 3 Randomization Check Results**

Variable	Treatment Producers	Control Producers	<i>P</i> -value of <i>t</i> -test
<i>ValidWatch</i>	163.055	176.388	0.248
<i>WatchDuration</i>	137.386	146.698	0.453
<i>Utilitarian</i>	0.065	0.066	0.238
<i>LowSkill</i>	0.627	0.627	0.652
<i>Follower</i>	1286.531	1261.090	0.601
<i>KOL</i>	0.003	0.003	0.830
<i>Experience</i>	2.343	2.341	0.553
<i>Following</i>	378.537	378.720	0.817
<i>Multihome</i>	0.856	0.855	0.103
<i>Female</i>	0.610	0.610	0.597
<i>PublicVisible</i>	0.940	0.941	0.259
<i>VideoDuration</i>	0.465	0.465	0.897
<i>Cover</i>	0.155	0.156	0.135
<i>ContentType</i>	0.757	0.757	0.725

Notes: The *p*-value column represents the significance level from a *T*-test comparing the treatment and control groups. Values for *ValidWatch*, *WatchDuration*, *Follower*, and *Following* have been scaled.

#### 4.1. Effects of Having Access to AI-generated Metadata on Content Consumption

We used the ordinary linear squares (OLS) regression specification with robust standard errors to causally estimate the effects of having access to AI-generated titles on viewership outcomes:

$$Outcome_{ij} = \beta_0 + \beta_1 Treat_i + \beta_2 Controls_{ij} + e_{ij} \quad (1)$$

where  $Treat_i$  is a binary indicator equal to 1 if the producer  $i$  was in the treatment group,  $Controls_{ij}$  include all prior-mentioned producer-, video-, and day-level attributes, and  $e_{ij}$  is the error term.  $Outcome_{ij}$  represented our two viewership metrics, include  $ValidWatch_{ij}$  (the number of valid watches) and  $WatchDuration_{ij}$  (viewers' total watch duration). All continuous variables in  $Outcome_{ij}$  were log-transformed, incremented by 1 to account for zero viewership outcomes, following the semi-log approach in Cole and Sokolyk (2018). Highly-skewed control variables were also log-transformed.

The model estimation results in Table 4 show that AI-generated titles boosted content consumption. Specifically, column (1) indicates an increase of 1.6%<sup>16</sup> in valid watches in the treatment group compared to the control group ( $\beta_1 = 0.016$ ,  $p$ -value < 0.01). Results in column (2) indicate that treatment group videos enhanced watch duration by 0.9% from the control group ( $\beta_1 = 0.009$ ,  $p$ -value < 0.01). Given our sample covers 2% of total platform users, who posted over 10 million videos during our experiment, this result translates to billions of additional valid watches and billions of extra minutes in watch duration across the platform, demonstrating significant economic benefits.

<sup>16</sup> The marginal effect size is calculated as:  $\text{Exp}(0.016)-1=1.6\%$ . The same calculation method is applied throughout this paper.



**Table 4 Results of Having Access to AI-generated Titles on Content Consumption**

Dependent Variable	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Treat</i>	0.016*** (0.001)	0.009*** (0.001)
Relative Effect Size	1.6%	0.9%
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.297	0.322

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

To investigate the underlying forces of the boosted viewership outcomes, we examined the textual characteristics of video titles using alternative dependent variables in Equation (1). Specifically, we utilized two binary indicators:  $Is\_title_{ij}$ , which indicates whether video  $j$  had a title (either GAI or human-generated), and  $Is\_tag_{ij}$ , which denotes whether the video title included tags. After re-estimating the OLS Equation (1) with these variables, the results in Table 5 demonstrate that having access to AI-generated titles increased the likelihood of a video having a title by 41.4% and enhances the probability of having tags by 72.4%<sup>17</sup>. These results suggest that having access to AI-generated titles effectively reduced metadata sparsity by increasing title and tag completeness, which in turn boosted viewership.

**Table 5 Results of Video Title Characteristics Analysis**

Dependent Variable	<i>Is_title</i> (1)	<i>Is_tag</i> (2)
<i>Treat</i>	0.244*** (0.0003)	0.247*** (0.0003)
Control Baseline (Mean)	0.590	0.341
Relative Effect Size	41.4%	72.4%
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.165	0.126

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

Building on these findings, we next examined whether this viewership-boosting effect was stronger for groups more affected by metadata sparsity. We introduced two moderator variables based on content type and producer skill level, as informed by prior research. According to social exchange theory, digital content creators are driven by motives like personal fulfillment or follower growth for financial gain (Wasko and Faraj 2005). Utilitarian-content videos (e.g., news and reviews) tend to have more detailed metadata to attract followers, while hedonic-content videos (e.g.,

<sup>17</sup> The relative effect size is calculated as  $0.244/0.590=0.414$  and  $0.247/0.341=0.724$ .

personal vlogs) typically lack such detail due to being more focused on self-expression. Additionally, digital divide studies (Nattamai Kannan et al. 2024) imply that low-skilled producers may under-value metadata due to a limited understanding of how recommender systems work. Accordingly, we constructed two moderator variables:  $Utilitarian_{ij}$ , identifying whether a video was utilitarian, and  $LowSkill_i$ , indicating low-skilled producers.

During the pre-treatment period, 61.3% of utilitarian videos had titles, compared to 57.3% of hedonic videos. Similarly, 65.3% of videos from low-skilled producers had titles, compared to 70.5% of those from high-skilled producers. These statistics align with the theoretical argument above, suggesting that AI-generated metadata may have a more pronounced effect on hedonic content and videos from low-skilled producers. To test this hypothesis, we incorporated these moderator variables and their interaction terms with  $Treat_i$  in Equation (1) to assess their influence.

Results in Table 6 show that utilitarian-content videos in the treatment group experienced a relative decrease in valid watches by 3.1% ( $p$ -value $<0.01$ ), and watch duration by 3.0% ( $p$ -value $<0.01$ ) than hedonic-content videos. Overall, utilitarian-content videos in the treatment group showed a 1.4%<sup>18</sup> decrease in valid watches, and a 1.9% decrease in watch duration, which is likely due to the more detailed human-generated metadata already associated with utilitarian-content videos. In contrast, Table 7 shows that low-skilled producers, compared to high-skilled ones, experienced an increase of 1.61% in valid watches ( $p$ -value $<0.01$ ), and 1.31% in watch duration ( $p$ -value $<0.01$ ) due to the access to AI-generated titles. These findings align with prior research showing that low-skilled workers disproportionately benefit from GAI tools (Chen and Chan 2023). Altogether, we find that the viewership-boosting effects of AI-generated metadata are stronger for hedonic-content videos and videos produced by low-skilled producers, due to their originally more sparse metadata.

**Table 6 Heterogeneous Effects of AI-generated Metadata Access on Content Consumption Across Video Types**

Dependent Variable	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Treat</i>	0.018*** (0.001)	0.011*** (0.001)
<i>Treat * Utilitarian</i>	-0.032*** (0.004)	-0.030*** (0.004)
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.297	0.322

Notes: \*\*\* $p$  $<0.01$ , \*\* $p$  $<0.05$ , \* $p$  $<0.1$ . Values in parentheses are robust standard errors. *Utilitarian* is absorbed by the video category dummies in *Controls* and therefore not reported in the table.

<sup>18</sup> This is calculated as:  $\text{Exp}(-0.032+0.018)-1 = -1.4\%$ .

**Table 7 Heterogeneous Effects of AI-generated Metadata Access on Content Consumption Across Producer Types**

Dependent Variable	Types	
	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Treat</i>	0.004*** (0.002)	-0.001 (0.001)
<i>LowSkill</i>	-0.783*** (0.002)	-0.755*** (0.001)
<i>Treat * LowSkill</i>	0.016*** (0.002)	0.013*** (0.002)
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.235	0.256

Notes: \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . Values in parentheses are robust standard errors.

#### 4.2. Results of Adopting AI-generated Metadata on Viewership Outcomes

To identify the effect of adopting AI-generated metadata, we cannot simply compare producers who adopted AI-generated titles with those who did not, because omitted variables (e.g., producers' inherent capability to generate titles) may drive both producers' decision to adopt AI-generated titles and their subsequent video viewership outcomes. Instead, we used the random assignment of producers to the treatment group ( $Treat_i$ ) as an instrumental variable (IV) for the adoption decision of AI-generated titles (Huang et al. 2021, Sun et al. 2019). We employed the following two-stage least squares (2SLS) regression specification:

$$Adopt_{ij} = \gamma_0 + \gamma_1 Treat_i + \gamma_2 Controls_{ij} + \epsilon_{ij} \quad (2)$$

$$Outcome_{ij} = l_0 + l_1 \hat{Adopt}_{ij} + l_3 Controls_{ij} + \eta_{ij} \quad (3)$$

where  $\epsilon_{ij}$  and  $\eta_{ij}$  are error terms. In Equation (2),  $Adopt_{ij}$  is a binary indicator for whether producer  $i$  adopts an AI-generated title for video  $j$ , and is instrumented with  $Treat_i$ . In Equation (3),  $\hat{Adopt}_{ij}$  refers to the instrumented  $Adopt_{ij}$ , i.e., the fitted value of  $Adopt_{ij}$  from Equation (2).  $\beta_1$  is the coefficient of interest indicating LATE.  $Treat_i$  is a valid IV for two reasons. First, it satisfies the relevance assumption as only the treatment group can access AI-generated titles, significantly influencing adoption. This is evidenced by a high first-stage  $F$ -statistic of 1,700,000. Second, it satisfies the exclusion restriction because the treatment assignment is random and should not correlate with other observed or unobserved covariates. Moreover, title generation occurs after video upload and just before posting, removing direct influence on video production.

The main effect results are presented in Table 8. The positive coefficients of  $Adopt_{ij}$  indicate that adopting AI-generated titles increased valid watches by 7.1% ( $p$ -value $<0.01$ ), and watch duration by 4.1% ( $p$ -value $<0.01$ ). These findings align with our ITT results but reflect a greater magnitude of

effect, demonstrating that adopting AI-generated titles significantly enhances content consumption. Similarly, the coefficients of the interaction term in Table 9 and Table 10 show this effect was more pronounced for hedonic-content videos and low-skilled producers, consistent with our ITT results in Table 6 and Table 7. Specifically, utilitarian-content videos showed a relative decrease of 14.0% in valid watches ( $p$ -value $<0.01$ ), and 13.2% in watch duration ( $p$ -value $<0.01$ ) compared with hedonic-content videos. In contrast, low-skilled producers, relative to high-skilled ones, received an incremental increase of 6.7% in valid watches ( $p$ -value $<0.01$ ), and 5.5% in watch duration ( $p$ -value $<0.01$ ) due to the adoption of AI-generated titles. However, the net effect was negative for utilitarian-content videos, with a decrease of 1.2% in valid watches, and 8.8% in watch duration. This implies that AI-generated titles, while helpful in addressing video title sparsity issues, may not surpass the quality of some existing human-generated titles.

**Table 8 Results of Adopting AI-generated Titles on Content Consumption**

Dependent Variable	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Adopt</i>	0.069*** (0.004)	0.040*** (0.004)
Relative Effect Size	7.1%	4.1%
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.297	0.322

Notes: \*\*\* $p<0.01$ , \*\* $p<0.05$ , \* $p<0.1$ . Values in parentheses are robust standard errors.

**Table 9 Heterogeneous Effects of Adopting AI-generated Titles on Content Consumption Across Video Types**

Dependent Variable	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Adopt</i>	0.078*** (0.004)	0.049*** (0.004)
<i>Adopt * Utilitarian</i>	-0.151*** (0.019)	-0.141*** (0.018)
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.297	0.322

Notes: \*\*\* $p<0.01$ ; \*\* $p<0.05$ ; \* $p<0.1$ . Values in parentheses are robust standard errors. *Utilitarian* is absorbed by the video category dummies in *Controls* and therefore not reported in the table.

**Table 10 Heterogeneous Effects of Adopting AI-generated Titles on Content Consumption Across Producer Types**

Dependent Variable	Types	
	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Adopt</i>	0.019*** (0.007)	-0.004 (0.006)
<i>Lowskill</i>	-0.785*** (0.002)	-0.756*** (0.001)
<i>Adopt * LowSkill</i>	0.065*** (0.008)	0.054*** (0.007)
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.235	0.256

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

### 4.3. Mechanism: AI-Generated Metadata Facilitates User-Content Matching

So far, we have shown that AI-generated titles significantly increased viewership outcomes. Next, we explore the mechanism behind this effect. Given the importance of metadata in user-content matching of recommender systems, we hypothesized that AI-generated titles improved video viewership by enhancing user-video matching accuracy. To illustrate, AI-generated titles probably help recommender systems better interpret video content. This enhanced interpretation should enable the system to more accurately predict which users are more likely to engage (e.g., view/like/share videos or follow the producer) and recommend/match the video to these specific users. This improved user-video matching accuracy translates into the higher consumption outcomes observed in Sections 4.1 and 4.2.

To evaluate user-video matching accuracy, we used the Area Under the ROC Curve (AUC), a widely applied metric in recommender system studies (Chen et al. 2024, Bi et al. 2024). AUC measures how well the model predicts viewer engagement behaviors by comparing the predicted and actual viewer engagement outcomes. A higher AUC indicates more accurate user-video matching. To calculate AUC, we collected a proprietary dataset from the platform’s recommender system, documenting video recommendations from November 1<sup>st</sup> to November 30<sup>th</sup>, 2023 for videos posted during our experiment.<sup>19</sup> It included 93,618,096 records with details on predicted engagement probabilities (i.e., like videos, share videos, and follow producers) and actual viewer behaviors for each user-video pair.

While our main analysis focuses on viewership outcomes (e.g., valid watch and watch duration), this additional dataset does not include predictions for these measures. Instead, it focuses on

<sup>19</sup> Videos produced during our experiment period were still recommended to users with their titles unchanged after the experiment. Because the platform did not archive the predicted engagement data during our experiment period, we used the data collected in November, 2023.

downstream engagement behaviors, which occur at later stages of the user journey. These predicted behaviors serve as essential inputs for the recommender system to match users with content that aligns with their preferences. Higher AUC values for like, share, and follow indicate that the recommender system effectively predicts user interactions. As these behaviors occur at later stages of the user journey, their accurate predictions imply that viewership outcomes, which occur earlier, are also likely to be predicted accurately. Collectively, a higher AUC for these engagement behaviors reflects improved user-video matching accuracy and supports our hypothesis that AI-generated titles enhance video viewership and engagement through better user-video matching.

To compare whether AUC values differ significantly across treatment and control group videos, we performed 1,000 bootstrap resampling iterations to calculate AUC and  $p$ -values for both treatment and control videos. The results in Table 11 showed significant improvements. The AUC for shares increased from 0.823 in the control group to 0.848 in the treatment group, an improvement of 0.026 ( $p$ -value<0.01). For likes, the AUC rose from 0.892 to 0.921, an increase of 0.029 ( $p$ -value<0.01), and for follows, it increased from 0.867 to 0.887, a rise of 0.019 ( $p$ -value<0.01). These results confirmed that AI-generated titles significantly improved user-video matching accuracy. These findings support our hypothesis that AI-generated titles enhance video viewership and engagement through better user-video matching.

**Table 11 AUC Comparison**

Variable	Treatment Group	Control Group	Difference	$P$ -value
<i>Share</i>	0.848	0.823	0.026	<0.01
<i>Like</i>	0.921	0.892	0.029	<0.01
<i>Follow</i>	0.887	0.867	0.019	<0.01

## 5. AI vs. Human-Generated Titles

Beyond the impact of AI-generated metadata on content consumption, we next explored whether and how human content producers can co-create with AI-generated metadata to further enhance consumption outcomes, to address whether content producers should be offered the option to modify AI-generated metadata.

### 5.1. Effect of Human-AI Co-creation on Content Consumption

We began our exploration by comparing the effectiveness of AI-generated titles to human-generated titles. While Section 4 demonstrated that AI-generated titles boosted video viewership by addressing title sparsity, their impact on videos that already have human-generated titles remained unclear. To investigate this, a direct comparison between titled videos in the treatment and control groups would be misleading. This is because some videos in the treatment group may only

be titled because of the access to AI-generated titles, potentially indicating lower producer effort and video quality. Thus, such a direct comparison could underestimate the true effect of accessing AI-generated titles on viewership for titled videos. To address this, we used the propensity score matching (PSM) method with the radius matching algorithm<sup>20</sup>, employing one-to-many matching to pair each titled video in the treatment group with several of the most “similar” titled videos in the control group based on pre-treatment covariates<sup>21</sup> (more details are available in Appendix B). Using the matched sample, we re-estimated Equation (1), applying weights to account for the multiple matches per treated unit. Interestingly, the results in Table 12 show that videos in the treatment group experienced a decrease of 37.9%<sup>22</sup> in valid watches and 32.6% in watch duration compared to the control group. These results suggest that AI-generated titles may not outperform existing human-generated titles in terms of quality.

**Table 12 Model Estimation Results for Titled Videos (Matched Sample)**

Dependent Variable	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Treat</i>	-0.476*** (0.002)	-0.394*** (0.001)
Relative Effect Size	-37.9%	-32.6%
Controls	YES	YES
Observations	3,885,089	3,885,089
R-square	0.329	0.351

*Notes:* \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . Values in parentheses are robust standard errors. The control baseline used here is calculated based on the matched sample.

To further explore this phenomenon, we next analyzed the heterogeneous effect of textual similarity on viewership. We followed the text-mining literature (Burtch et al. 2022) and employed a well-established approach to construct textual similarity. Specifically, we employed the cosine distance between numeric representations of textual content in vector space. This measure is constructed using term-frequency inverse-document frequency (TF-IDF<sup>23</sup>) within an embedding space derived from a broader corpus of video titles in both the treatment and control groups in our sample. Using the TF-IDF algorithm, we computed cosine similarities between AI-generated titles

<sup>20</sup> We also applied other matching algorithms (e.g., kernel matching) and found robust results.

<sup>21</sup> We removed 2,226,922 titled videos (51.56% of the total titled videos) in our treatment group that did not receive AI-generated titles due to algorithmic issues.

<sup>22</sup> The marginal effect is calculated as:  $\text{Exp}(-0.476) - 1 = -37.9\%$ . The same calculation method is applied subsequently.

<sup>23</sup> TF-IDF is a scaled matrix that captures how frequently each term appears in a document relative to its frequency across all documents in the dataset. This scaling down-weights common words that are less informative for distinguishing between documents, thus emphasizing words that are unique to specific documents. If a word is highly unique and only appears in a single document, its impact is preserved, while words common across many documents are given less weight. Calculating cosine distances between document vectors in this TF-IDF derived space effectively measures similarity in terms of distinctive word usage.

and actual video titles adopted by producers ( $Similarity_{ij}$ ). We then incorporated this similarity measure ( $Similarity_{ij}$ ) into Equation (1) by adding an interaction term between  $Treat_i$  and  $Similarity_{ij}$ .<sup>24</sup>

$$Outcome_{ij} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 (Treat_i \times Similarity_{ij}) + \alpha_3 Controls_{ij} + \mu_{ij} \quad (4)$$

The negative coefficient of the interaction term in Table 13 indicates that a 10% increase in similarity between AI- and human-generated titles led to an 9.8% decrease in valid watches<sup>25</sup>, and a 8.2% decrease in watch duration. However, interestingly, when considering the positive coefficient for  $Treat_i$ , we find that treatment group videos with less than 20.8%<sup>26</sup> and 20.9% similarity (i.e., low overlap with AI-generated titles) outperformed the control group in both valid watches and watch duration. These findings suggest that AI-generated titles may be of lower quality than human-generated titles, and thus higher similarity to these titles tended to reduce viewership. However, when content producers revised AI-generated titles—resulting in lower similarity—the negative effect diminished, and treatment group videos ultimately performed better than those in the control group. These findings suggest that while AI-generated titles reduced production costs and addressed title sparsity, producers should revise these titles rather than adopt them without changes. This aligns with prior research showing that human-AI collaboration outperformed both full automation and human-only approaches (Boyacı et al. 2024), emphasizing the benefits of combining human judgment with AI efficiency (Chen and Chan 2023, Zhou and Lee 2024).

**Table 13** Heterogeneous Estimation Results for Titled Videos (Matched Sample)

Dependent Variable	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Treat</i>	0.213*** (0.002)	0.178*** (0.002)
<i>Treat * Similarity</i>	-1.026*** (0.003)	-0.852*** (0.002)
Controls	YES	YES
Observations	3,885,089	3,885,089
R-square	0.382	0.393

Notes: \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . Values in parentheses are robust standard errors.

There is an endogeneity concern that the videos with significantly revised AI-generated titles, resulting in lower similarity between the AI-generated titles and actual titles adopted by producer,

<sup>24</sup> The main effect of  $Similarity_{ij}$  is not included in this specification because it would be absorbed by the interaction term ( $Treat_i \times Similarity_{ij}$ ) due to collinearity.

<sup>25</sup> The marginal effect size is calculated as:  $\text{Exp}((-1.026) * 0.1) - 1 = -9.8\%$ .

<sup>26</sup> This is calculated as  $0.213 / 1.026 = 20.8\%$ . The same calculation method is applied subsequently.



may inherently indicate higher producer effort and video quality. In other words, the observed outcomes could stem from these underlying differences rather than the benefits of human-AI co-creation. To address this issue, we employed the propensity score matching (PSM) method with a radius matching algorithm, utilizing one-to-many matching. Specifically, for each titled video in the treatment group with cosine similarity to AI-generated titles below 20%, we matched it with several control group videos that were most similar based on pre-treatment covariates (details are presented in Online Appendix C). Using the matched sample, we re-estimated Equation (1) and applied weights to account for multiple matches per treated unit. The results in Table 14 show that videos in the treatment group outperformed those in the control group in terms of valid watches and watch duration. These findings suggest that the observed higher viewership outcomes are likely driven by human-AI co-creation.

**Table 14 Model Estimation Results for Titled Videos with Lower Textual Similarity (Matched Sample)**

Dependent Variable	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)
<i>Treat</i>	0.153*** (0.002)	0.121*** (0.002)
Relative Effect Size	16.5%	12.9%
Controls	YES	YES
Observations	3,533,720	3,533,720
R-square	0.242	0.269

*Notes:* \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . Values in parentheses are robust standard errors. The control baseline used here is calculated based on the matched sample.

## 5.2. Effect of Human-AI Co-creation on Lexical Richness

To better understand how increased human input boosted video viewership, we used lexical richness, a key linguistic concept in language studies, signaling information quality, as our alternative dependent variable. We followed prior research (Qiao et al. 2020) and measured it through multiple dimensions, including lexical density ( $LexicalDensity_{ij}$ ), lexical variation ( $LexicalVariation_{ij}$ ), and entropy ( $Entropy_{ij}$ ). Lexical richness is an important proxy for cognitive effort in text crafting and a signal of information quality. For example, Goes et al. (2014) used lexical density to measure the informational value of reviews. Lexical density is the proportion of content words (such as nouns, verbs, and adjectives) to the total number of words in a text. Lexical variation is the ratio of unique words to the total number of words. Entropy quantifies text unpredictability, computed as:

$$Entropy = - \sum_{k=1}^n P_k \log P_k \quad (5)$$

where  $P_k$  is the probability of each unique word. We also include sentence length as an additional confounder in this analysis. The results in Table 15 show that each 10% increase in similarity score related to a 4.8%,<sup>27</sup> 3.7%, and 1.1% decrease in lexical density, lexical variation, and entropy, respectively. These results show that titles with lower similarity tend to be more descriptive and information-rich. Such titles may provide clearer context and relevant keywords that align more effectively with the video content, improving the recommender system’s ability to match videos with the targeted users.<sup>28</sup> For instance, in a video featuring peaceful natural scenery—trees and flowing water—the AI-generated title, “Enjoy the Beauty of Nature #ScenicNature,” captures the general theme but lacks specificity. In contrast, a human-revised AI-generated title, “Lush Mountains and Flowing Streams: Embrace Nature’s Serenity,” offers greater lexical richness by adding more descriptors. Nouns like “Mountains” and “Streams” highlight the visual elements of the video content, while descriptive terms such as “Lush” and “Flowing” convey the element states, enhancing lexical density and variation. These context-specific enhancements allow the title to better describe and align with the video content, making it easier for the recommender system to interpret the video for more accurate content-user matching (Panniello et al. 2016).

**Table 15 Results of Lexical Richness Analysis (Matched Sample)**

Dependent Variable	<i>LexicalDensity</i> (1)	<i>LexicalVariation</i> (2)	<i>Entropy</i> (3)
<i>Treat</i>	0.016*** (0.0002)	0.017*** (0.0002)	0.011*** (0.0007)
<i>Treat * Similarity</i>	-0.245*** (0.0002)	-0.301*** (0.0002)	-0.442*** (0.0007)
Control Baseline (Mean)	0.509	0.812	3.889
Controls	YES	YES	YES
Observations	3,885,089	3,885,089	3,885,089
R-square	0.621	0.674	0.803

Notes: \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . Values in parentheses are robust standard errors.

To investigate why AI-generated titles led to high-quality but dissimilar titles, we surveyed 1,925 treatment group users in August. This survey featured open questions on the usage of AI-generated titles. The qualitative feedback highlighted an *inspiration effect*, where users perceived AI-generated titles as a creative catalyst. For example, one user noted, “It’s already great; it may not always be precise, but it provides some inspiration for our posts!” Another mentioned its help with “writer’s block,” saying, “When I can’t think of anything, it helps a bit.” This evidence highlights the role of AI-generated titles as catalyst in content creation,<sup>29</sup> motivating producers to re-write or come

<sup>27</sup> The marginal effect size is calculated as:  $(-0.245 \times 0.1) / 0.509 = -4.8\%$ .

<sup>28</sup> See <https://hivo.co/blog/creating-descriptive-titles-for-content-with-ai-a-how-to-guide>.

<sup>29</sup> Similar arguments can be found in <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/how-generative-ai-can-boost-consumer-marketing>.

up with their own titles. These findings align with recent studies on GAI and creativity (Zhou and Lee 2024).

## 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 Appendix D.

**Viewership Diversity.** In the main text, we focus on the economic impact of AI-generated titles on content consumption, particularly consumption quantity (e.g., number of valid watch). However, their implications can be multifold, encompassing both quantity and diversity. This section analyzed how the access to AI-generated titles affects video viewership diversity using the Herfindahl-Hirschman Index (HHI), a widely used measure of market concentration in economic and antitrust analyses (Narayanan et al. 2009). HHI is calculated by squaring the market share of each entity and summing the results, with values ranging from close to 0 to 1. A lower HHI indicates a more competitive environment, while a higher HHI signals dominance by one or a few large entities. In our analysis, we used valid watches to calculate HHI. The index ranged from  $1/N$  to 1, where  $N$  was the total number of videos in our context. As shown in Table 19, the treatment group with access to AI-generated titles had a significantly lower HHI of 0.0002 compared to 0.0003 in the control group, representing a 50% reduction in platform-level HHI. This indicates a substantial increase in viewership diversity, aligning with our earlier finding that AI-generated titles disproportionately benefited low-skilled producers. These results are consistent with recent GAI studies (Zhou and Lee 2024) and contribute to the growing body of research on GAI’s impact on socioeconomic inequality (Capraro et al. 2024).

**Channel Analysis.** In Section 3.1, we have discussed that organic and search-oriented recommendations are two main video recommendation channels on Platform A. Building on this, we next analyzed viewership outcomes for each channel separately and replicated the main analysis in Equation (1). As shown in Table 20 and Table 21, the positive coefficients of *Treat* are qualitatively aligned with the results in Table 4. These results indicate that AI-generated titles improve content consumption across both channels, reinforcing the effectiveness of AI-generated titles in enhancing user-video matching.

**Impact on Content Production.** One potential explanation for the boosted viewership is that access to AI-generated titles changes users’ video production behavior. For example, producers with access to AI-generated titles might spend more time refining each video’s content and producing fewer, but higher-quality, videos. To examine this possibility, we conducted a producer-level *t*-test comparing both the total number of videos and the average time gaps (in hours) between videos

produced by each producer in the treatment and control groups during the treatment period. The results, shown in Table 22, indicate no statistically significant difference in the number of videos produced and time gaps between videos between the two groups. These results suggest that the increase in viewership in the treatment group is unlikely to be driven by a change in producers’ video production behavior.

**Alternative Operationalization of Variables.** We employed several alternative operationalizations of variables to ensure the robustness of our results. First, we used the number of watches ( $Watch_{ij}$ ), complete watches<sup>30</sup> ( $CompleteWatch_{ij}$ ), and the number of likes ( $Like_{ij}$ ) for video  $j$  of producer  $i$  as alternative dependent variables for viewership outcomes. The ITT results presented in Table 23 and 24, and LATE results shown in Table 25 and 26 of Online Appendix D.4, are qualitatively consistent with our main findings.

Second, in Section 3.3,  $LowSkill_i$  is coded as 1 if producer  $i$ ’s number of followers is below the median. To test robustness, we alternatively coded  $LowSkill_i$  as 1 if the cumulative number of videos posted by producer  $i$  exceeds the median (157). The ITT and LATE results, presented in Table 27, remain qualitatively consistent with our main findings.

Third, in Section 3.3,  $Adopt_{ij}$  is coded as 1 for exact matches between AI-generated titles and posted video titles. As a robustness test, we coded  $Adopt_{ij}$  as 1 if the cosine similarity between the AI-generated title and the video’s posted title met or exceeded specified thresholds (i.e., 95%, 90%, 85%, and 80%). The LATE results in Table 28, using  $ValidWatch_{ij}$  as the dependent variable, are qualitatively aligned with our findings.

Fourth, we employed the Levenshtein algorithm as an alternative method to calculate the textual similarity ( $Similarity_{ij}$ ) between video titles and AI-generated titles in Section 5. The Levenshtein algorithm, also known as the edit-distance metric, measures the minimum number of insertions, deletions, and substitutions required to transform one string into another, with each operation having a cost of one. Higher edit counts indicate lower similarity between AI-generated titles and actual video titles. The regression results for viewership outcomes and title lexical richness, presented in Tables 29 and 30, are qualitatively consistent with our findings.

## 7. Conclusion and Discussion

Previous research has shown that AI-generated content, such as advertisements, effectively engages users by enhancing content quality. However, the value of AI-generated content that does not directly interact with users, such as metadata, remains less understood. To address this gap, we conducted a randomized field experiment on a short-video platform where AI-generated titles,

<sup>30</sup> A complete watch is coded as 1 when the watch duration exactly matches the full video duration.

displayed in the bottom left corner, were rarely noticed by users. This setup allows us to isolate the effect of AI-generated metadata without direct user interaction. Our results show that AI-generated titles significantly boosted video consumption. Specifically, access to AI-generated video titles increased valid watches by 1.6% and watch duration by 0.9%. Further analysis suggests that this impact was primarily driven by addressing metadata scarcity, as evidenced by a notable increase in title availability. Moreover, we find that this effect was amplified for utilitarian videos and those produced by low-skilled creators, i.e., groups more affected by metadata sparsity. Specifically, utilitarian-content videos in the treatment group saw a relative decrease of 3.1% in valid watches and 3.0% in watch duration compared to hedonic-content videos. In contrast, low-skilled producers experienced an additional increase of 1.6% in valid watches, and 1.3% in watch duration. Mechanism analysis further indicates that AI-generated titles enhanced user-video matching accuracy. However, AI-generated titles were often of lower quality than human-generated ones, and only human-revised AI titles improved engagement further, highlighting the potential of combining AI with human input for better outcomes.

### 7.1. Practical Implications

Our results shed light on several important managerial implications. First, our study demonstrates that AI-generated metadata can significantly boost content discovery by improving user-content matching through mitigating metadata sparsity. Therefore, we encourage platform owners to invest in GAI tools that generate metadata, which can address operational challenges related to sparse metadata. While much of the focus has been on GAI’s ability to create user-facing content (e.g., advertisements and articles), our results emphasize the equally crucial role of AI-generated metadata in improving platform operations and boosting content discovery. This is relevant for platforms where content consumption is primarily driven by recommendations, such as UGC platforms and e-commerce sites.

Second, by demonstrating that AI-generated metadata disproportionately benefits low-skilled producers and hedonic-content videos, our work reveals the importance of tailoring platform strategies to support those most affected by metadata sparsity. Platforms should consider focusing their efforts on these segments when deciding whether to scale up the implementation of AI-generated metadata tools and how to maximize their effectiveness. For example, platforms can prioritize rolling out these tools to groups more affected by metadata sparsity, such as novice producers, to generate the most immediate and noticeable impact on content consumption.

Third, while GAI tools streamline video title generation, our results show that the quality of these AI-generated titles often falls short of human-generated ones. Therefore, rather than automatically integrating AI-generated titles into their recommender systems, platforms are encouraged to display

these titles to content producers and provide the option to modify or enhance the metadata. Building on this, platforms are also recommended to incentivize content producers to revise or enhance AI-generated metadata rather than adopting them automatically. For example, platforms could place prominent reminders or offer traffic awards, to encourage producers to revise AI-generated titles. This can allow producers to inject their creativity and domain expertise to improve metadata. Additionally, our results show that improved titles with greater linguistic richness were associated with better viewership outcomes. To capitalize on this, platforms should consider offering workshops or tutorials to equip content producers with the skills needed to effectively use AI tools. By training producers to create more contextually rich and detailed metadata, platforms can enhance content discoverability and drive higher engagement.

## 7.2. Limitations and Future Research

The limitations of our work open up interesting avenues for future research. First, while we focus on the value of AI-generated metadata in improving user-content matching, we only explore content-based metadata. Future research could explore the role of user-related metadata, such as AI-generated user profiles. GAI can generate synthetic user profiles based on minimal inputs or demographic similarities, which may help the system to better predict user preferences and improve matching accuracy. Additionally, examining how different types of AI-generated metadata (e.g., content and user metadata) interact or complement each other also deserves exploration. Second, in our study, the GAI algorithm generated titles solely based on video content, without incorporating producer attributes or their historical content. Future research could explore how to design and improve AI-generated metadata to induce stronger effects. One potential direction is to incorporate producer-specific data, such as frequently used keywords from past videos or audience engagement patterns, to generate personalized AI-generated titles.

## 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.
- Agrawal, Saurabh, John Trenkle, Jaya Kawale. 2023. Beyond labels: Leveraging deep learning and llms for content metadata. *Proceedings of the 17th ACM Conference on Recommender Systems*. 1–1.
- Anthony, Callen, Beth A. Bechky, Anne-Laure Fayard. 2023. “collaborating” with ai: Taking a system view to explore the future of work. *Organization Science* **34**(5) 1672–1694.
- Bai, Bing, Tat Chan, Dennis Zhang, Fuqiang Zhang, Yujie Chen, Haoyuan Hu. 2022a. The value of logistic flexibility in e-commerce. *Available at SSRN 4206229* .
- Bai, Bing, Hengchen Dai, Dennis J Zhang, Fuqiang Zhang, Haoyuan Hu. 2022b. The impacts of algorithmic work assignment on fairness perceptions and productivity: Evidence from field experiments. *Manufacturing & Service Operations Management* **24**(6) 3060–3078.

- Bi, Xuan, Mochen Yang, Gediminas Adomavicius. 2024. Consumer acquisition for recommender systems: A theoretical framework and empirical evaluations. *Information Systems Research* **35**(1) 339–362.
- Boyacı, Tamer, Caner Canyakmaz, Francis de Véricourt. 2024. Human and machine: The impact of machine input on decision making under cognitive limitations. *Management Science* **70**(2) 1258–1275.
- Buell, Ryan W, Tami Kim, Chia-Jung Tsay. 2017. Creating reciprocal value through operational transparency. *Management Science* **63**(6) 1673–1695.
- Burtch, Gordon, Qinglai He, Yili Hong, Dokyun Lee. 2022. How do peer awards motivate creative content? experimental evidence from reddit. *Management Science* **68**(5) 3488–3506.
- Burtch, Gordon, Yili Hong, Ravi Bapna, Vladas Griskevicius. 2018. Stimulating online reviews by combining financial incentives and social norms. *Management Science* **64**(5) 2065–2082.
- Burtch, Gordon, Dokyun Lee, Zhichen Chen. 2023. Generative ai for ugc and online community engagement. Available at SSRN 4521754 .
- Capraro, Valerio, Austin Lentsch, Daron Acemoglu, Selin Akgun, Aisel Akhmedova, Ennio Bilancini, Jean-François Bonnefon, Pablo Brañas-Garza, Luigi Butera, Karen M Douglas, et al. 2024. The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus* **3**(6).
- Chen, Jiawei, Luo He, Hongyan Liu, Yinghui Yang, Xuan Bi. 2024. Background music recommendation on short video sharing platforms. *Information Systems Research* .
- Chen, Zenan, Jason Chan. 2023. Large language model in creative work: The role of collaboration modality and user expertise. *Management Science* .
- Cheng, Zhaoqi, Dokyun Lee, Prasanna Tambe. 2022. Innovae: Generative ai for mapping patents and firm innovation. Available at SSRN 3868599 .
- Choi, Tsan-Ming, Stein W Wallace, Yulan Wang. 2018. Big data analytics in operations management. *Production and Operations Management* **27**(10) 1868–1883.
- Clyde, Nicholas, Dennis Zhang, Bing Bai. 2024. The impact of ridesharing platforms on healthcare access. Available at SSRN 4968892 .
- Cole, Rebel A, Tatyana Sokolyk. 2018. Debt financing, survival, and growth of start-up firms. *Journal of Corporate Finance* **50** 609–625.
- Cui, Ruomeng, Jun Li, Dennis J. Zhang. 2020a. Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science* **66**(3) 1071–1094.
- Cui, Ruomeng, Meng Li, Qiang Li. 2020b. Value of high-quality logistics: Evidence from a clash between sf express and alibaba. *Management Science* **66**(9) 3879–3902.
- Cui, Ruomeng, Meng Li, Shichen Zhang. 2022. Ai and procurement. *Manufacturing & Service Operations Management* **24**(2) 691–706.

- Davidson, James, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. 2010. The youtube video recommendation system. *Proceedings of the fourth ACM conference on Recommender systems*. 293–296.
- Dukes, Anthony, Qihong Liu. 2024. The consumption of advertising in the digital age: Attention and ad content. *Management Science* **70**(4) 2086–2106.
- Ellis, Scott C., Shashank Rao, Dheeraj Raju, Thomas J. Goldsby. 2018. Rfid tag performance: Linking the laboratory to the field through unsupervised learning. *Production and Operations Management* **27**(10) 1834–1848.
- Fang, Zhen, Ming Fan, Apurva Jain. 2023. Content proliferation and narrowcasting in the age of streaming media. *Production and Operations Management* **32**(10) 3295–3310.
- Filippas, Apostolos, Srikanth Jagabathula, Arun Sundararajan. 2023. The limits of centralized pricing in online marketplaces and the value of user control. *Management Science* **69**(12) 7202–7216.
- Goes, Paulo B, Mingfeng Lin, Ching-man Au Yeung. 2014. “popularity effect” in user-generated content: Evidence from online product reviews. *Information Systems Research* **25**(2) 222–238.
- Granulo, Armin, Christoph Fuchs, Stefano Puntoni. 2021. Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts. *Journal of Consumer Psychology* **31**(1) 72–80.
- He, Qinglai, Yili Hong, TS Raghu. 2021. The effects of machine-powered platform governance: An empirical study of content moderation. *Available at SSRN 3767680* .
- Hoiles, William, Anup Aprem, Vikram Krishnamurthy. 2017. Engagement and popularity dynamics of youtube videos and sensitivity to meta-data. *IEEE Transactions on Knowledge and Data Engineering* **29**(7) 1426–1437.
- 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, Probal Mojumder, Tianshu Sun, Jinchi Lv, Joseph M Golden. 2021. Not registered? please sign up first: A randomized field experiment on the ex ante registration request. *Information Systems Research* **32**(3) 914–931.
- Kesavan, Saravanan, Tarun Kushwaha. 2020. Field experiment on the profit implications of merchants’ discretionary power to override data-driven decision-making tools. *Management Science* **66**(11) 5182–5190.
- Kuang, Lini, Ni Huang, Yili Hong, Zhijun Yan. 2019. Spillover effects of financial incentives on non-incentivized user engagement: Evidence from an online knowledge exchange platform. *Journal of Management Information Systems* **36**(1) 289–320.
- Liang, Paul Pu, Chiyu Wu, Louis-Philippe Morency, Ruslan Salakhutdinov. 2021. Towards understanding and mitigating social biases in language models. *International Conference on Machine Learning*. PMLR, 6565–6576.



- Liu, Jin, Xingchen Xu, Xi Nan, Yongjun Li, Yong Tan. 2024. "generate" the future of work through ai: Empirical evidence from online labor markets. URL <https://arxiv.org/abs/2308.05201>.
- Longoni, Chiara, Andrea Bonezzi, Carey K Morewedge. 2019. Resistance to medical artificial intelligence. *Journal of Consumer Research* **46**(4) 629–650.
- Lysyakov, Mikhail, Siva Viswanathan. 2023. Threatened by ai: Analyzing users' responses to the introduction of ai in a crowd-sourcing platform. *Information Systems Research* **34**(3) 1191–1210.
- Malik, Haroon, Zifeng Tian. 2017. A framework for collecting youtube meta-data. *Procedia Computer Science* **113** 194–201.
- Narayanan, Sriram, Sridhar Balasubramanian, Jayashankar M Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management science* **55**(11) 1861–1876.
- Nattamai Kannan, Karthik Babu, Eric Overby, Sridhar Narasimhan. 2024. Can improvements to mobile internet service help reduce digital inequality? an empirical analysis of education and overall data consumption. *Management Science* .
- Panniello, Umberto, Michele Gorgoglione, Alexander Tuzhilin. 2016. Research note—in cars we trust: How context-aware recommendations affect customers' trust and other business performance measures of recommender systems. *Information Systems Research* **27**(1) 182–196.
- Peng, Jiayu, Jungpil Hahn, Ke-Wei Huang. 2023. Handling missing values in information systems research: A review of methods and assumptions. *Information Systems Research* **34**(1) 5–26.
- 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.
- Reisenbichler, Martin, Thomas Reutterer, David A Schweidel, Daniel Dan. 2022. Frontiers: Supporting content marketing with natural language generation. *Marketing Science* **41**(3) 441–452.
- Saar-Tsechansky, Maytal, Prem Melville, Foster Provost. 2009. Active feature-value acquisition. *Management Science* **55**(4) 664–684.
- Su, Yi, Qili Wang, Liangfei Qiu, Runyu Chen. 2024. Unveiling the effects of introducing ai-generated summaries in e-commerce. *Available at SSRN 4872205* .
- Sun, Jiankun, Dennis J. Zhang, Haoyuan Hu, Jan A. Van Mieghem. 2022. Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science* **68**(2) 846–865. doi:10.1287/mnsc.2021.3990.
- Sun, Tianshu, Lanfei Shi, Siva Viswanathan, Elena Zheleva. 2019. Motivating effective mobile app adoptions: Evidence from a large-scale randomized field experiment. *Information Systems Research* **30**(2) 523–539.

- Susarla, Anjana, Ram Gopal, Jason Bennett Thatcher, Suprateek Sarker. 2023. The janus effect of generative ai: Charting the path for responsible conduct of scholarly activities in information systems. *Information Systems Research* **34**(2) 399–408.
- Wang, Wen, Siqi Pei, Tianshu Sun. 2023a. Unraveling generative ai from a human intelligence perspective: A battery of experiments. *Available at SSRN 4543351* .
- Wang, Zhihan (Helen), Jun Li, Di (Andrew) Wu. 2023b. Mind the gap: Gender disparity in online learning platform interactions. *Manufacturing & Service Operations Management* **25**(6) 2122–2141.
- 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.
- Wei, Wei, Xubin Ren, Jiabin Tang, Qinyong Wang, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, Chao Huang. 2024. Llmrec: Large language models with graph augmentation for recommendation. *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 806–815.
- Ye, Zikun, Dennis J Zhang, Heng Zhang, Renyu Zhang, Xin Chen, Zhiwei Xu. 2023. Cold start to improve market thickness on online advertising platforms: Data-driven algorithms and field experiments. *Management Science* **69**(7) 3838–3860.
- 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.
- Zhang, Xingyue, James A Dearden, Yuliang Yao. 2022. Let them stay or let them go? online retailer pricing strategy for managing stockouts. *Production and Operations Management* **31**(11) 4173–4190.
- Zhou, Eric, Dokyun Lee. 2024. Generative artificial intelligence, human creativity, and art. *PNAS Nexus* **3**(3) 052.

# Online Appendices

## A. Pearson Correlation Matrix of Focal Variables

**Table 16** Pearson Correlation Matrix of Focal Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) <i>Treat</i>	1													
(2) <i>Adopt</i>	0.36	1												
(3) <i>ValidWatch</i>	0.00	-0.01	1											
(4) <i>WatchDuration</i>	0.00	-0.01	0.80	1										
(5) <i>Utilitarian</i>	0.01	-0.01	0.01	0.01	1									
(6) <i>LowSkill</i>	0.00	0.02	-0.02	-0.02	0.00	1								
(7) <i>Follower</i>	0.00	-0.02	-0.17	0.13	0.01	-0.06	1							
(8) <i>KOL</i>	0.00	-0.02	0.02	0.01	0.01	-0.07	0.12	1						
(9) <i>Experience</i>	0.00	0.09	0.00	0.00	-0.05	-0.29	0.04	0.06	1					
(10) <i>Following</i>	0.00	0.08	0.01	-0.01	0.02	-0.42	-0.01	0.00	0.15	1				
(11) <i>Multihome</i>	0.00	-0.03	0.01	0.01	-0.01	-0.02	0.02	0.03	0.11	-0.05	1			
(12) <i>Female</i>	0.00	0.01	-0.01	-0.01	-0.03	-0.04	-0.01	-0.03	0.04	-0.03	0.06	1		
(13) <i>PublicVisible</i>	0.00	-0.04	0.01	0.01	0.03	-0.06	0.01	0.02	-0.03	0.06	-0.01	-0.09	1	
(14) <i>VideoDuration</i>	0.00	0.02	0.03	0.04	0.05	-0.01	0.03	0.01	0.06	0.05	0.00	-0.06	0.02	1

## B. Propensity Score Matching (PSM) Results for Titled Videos

From the total of 10,738,984 videos, we first excluded 3,444,235 videos which do not have titles from both the treatment and control groups. Additionally, we removed 2,226,922 titled videos (51.56%) in our treatment group that do not receive AI-generated titles due to algorithmic limitations. The filtered sample consisted of 2,092,219 videos in the treatment group and 2,975,608 in the control group. Next, we used the Propensity Score Matching (PSM) method to identify a sample that was similar in observed characteristics during the pre-treatment period. The matching followed a two-step procedure: first, we ran a logit regression using the pre-treatment variables (i.e., all the moderating and control variables mentioned in Section 3.3) and obtained predicted propensity scores for each unit. Second, we employed a one-to-many radius matching algorithm where all control units for which the propensity scores fall within a pre-defined radius (also known as caliper) from the propensity scores of the treatment units are matched. This ensures multiple matches for each treated unit. We then applied weights in the subsequent analysis to account for this one-to-many structure. Next, we obtained a new sample after discarding unmatched units (3,885,089 videos are matched). To evaluate the matching quality, we performed  $t$ -tests of equality of means before and after matching to verify whether our matching has successfully balanced the attributes between the treatment and control group videos. The results in Table 17 show that the mean differences between the groups were no longer statistically significant, indicating that the matching process successfully reduced bias associated with observable attributes.

**Table 17** Differences in Mean Before and After Matching (PSM, Radius Matching)

Variable	Sample	Mean Treated	Mean Control	%Bias	T-statistics	P-value
<i>Utilitarian</i>	Before-matched	0.065	0.069	-1.900	-20.650	0.000
	After-matched	0.065	0.073	-3.400	-0.050	0.960
<i>LowSkill</i>	Before-matched	0.523	0.471	10.200	113.610	0.000
	After-matched	0.523	0.528	-1.000	-0.010	0.989
Ln( <i>Follower</i> )	Before-matched	2.975	2.710	28.700	315.750	0.000
	After-matched	2.975	2.905	3.800	0.110	0.911
<i>KOL</i>	Before-matched	0.004	0.008	-4.500	-48.470	0.000
	After-matched	0.004	0.005	-1.100	-0.020	0.986
<i>Experience</i>	Before-matched	2.642	2.072	31.900	354.270	0.000
	After-matched	2.642	2.505	7.700	0.110	0.915
Ln( <i>Following</i> )	Before-matched	2.907	2.989	-8.200	-89.620	0.000
	After-matched	2.907	2.881	2.600	0.040	0.968
<i>Multihome</i>	Before-matched	0.720	0.768	-10.900	-121.070	0.000
	After-matched	0.720	0.735	-3.400	-0.050	0.962
<i>Female</i>	Before-matched	0.658	0.658	-0.100	-1.500	0.000
	After-matched	0.658	0.649	1.900	0.030	0.979
<i>Public Visible</i>	Before-matched	0.921	0.949	-11.200	-126.320	0.000
	After-matched	0.921	0.930	-3.400	-0.040	0.965
Ln( <i>VideoDuration</i> )	Before-matched	0.327	0.269	16.000	177.980	0.000
	After-matched	0.327	0.337	-2.600	-0.040	0.971
<i>Cover</i>	Before-matched	0.135	0.196	-16.300	-178.890	0.000
	After-matched	0.135	0.160	-6.700	-0.100	0.918
<i>ContentType</i>	Before-matched	0.828	0.660	39.200	425.900	0.000
	After-matched	0.828	0.850	-5.100	-0.080	0.934

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . %Bias measures the standardized difference in covariate means, scaled by the standard deviation of the sample including both treatment and control groups. The closer it is to zero, the better the balance between the treatment and control groups. Values for *Follower* and *Following* have been scaled.

### C. PSM Results for Videos with Low Textual Similarity

From the 2,092,219 titled videos in the treatment group and 2,975,608 in the control group (as described in Online Appendix B), we first excluded 1,532,819 treatment videos with cosine similarity to AI-generated titles higher than 20%. This filtering resulted in 559,400 videos in the treatment group and 2,975,608 in the control group. Using the Propensity Score Matching (PSM) method with a radius matching algorithm, we identified a matched sample based on pre-treatment characteristics (i.e., all moderating and control variables in Section 3.3), following the detailed procedure in Online Appendix B. After discarding unmatched units, the final matched sample included 3,533,720 videos. Post-matching  $t$ -tests of equality of means, shown in Table 18, indicate that the mean differences between the treatment and control groups were no longer statistically significant, confirming that the matching process effectively reduced bias.

**Table 18** Differences in Mean Before and After Matching (PSM, Radius Matching)

Variable	Sample	Mean Treated	Mean Control	%Bias	T-statistics	P-value
<i>Utilitarian</i>	Before-matched	0.065	0.069	-1.900	-20.650	0.000
	After-matched	0.066	0.069	-1.100	-0.040	0.966
<i>LowSkill</i>	Before-matched	0.523	0.471	10.200	113.610	0.000
	After-matched	0.533	0.485	9.700	0.360	0.717
Ln( <i>Follower</i> )	Before-matched	2.975	2.710	28.700	315.750	0.000
	After-matched	2.716	2.700	1.700	0.070	0.948
<i>KOL</i>	Before-matched	0.004	0.008	-4.500	-48.470	0.000
	After-matched	0.006	0.007	-1.400	-0.060	0.955
<i>Experience</i>	Before-matched	2.642	2.072	31.900	354.270	0.000
	After-matched	2.541	2.075	25.600	0.930	0.353
Ln( <i>Following</i> )	Before-matched	2.907	2.989	-8.200	-89.620	0.000
	After-matched	2.911	2.947	-3.600	-0.140	0.886
<i>Multihome</i>	Before-matched	0.720	0.768	-10.900	-121.070	0.000
	After-matched	0.777	0.767	2.600	0.100	0.922
<i>Female</i>	Before-matched	0.658	0.658	-0.100	-1.500	0.000
	After-matched	0.660	0.657	0.700	0.030	0.980
<i>Public Visible</i>	Before-matched	0.921	0.949	-11.200	-126.320	0.000
	After-matched	0.959	0.952	3.200	0.130	0.898
Ln( <i>VideoDuration</i> )	Before-matched	0.327	0.269	16.000	177.980	0.000
	After-matched	0.277	0.263	4.000	0.150	0.881
<i>Cover</i>	Before-matched	0.135	0.196	-16.300	-178.890	0.000
	After-matched	0.226	0.194	7.800	0.280	0.777
<i>ContentType</i>	Before-matched	0.828	0.660	39.200	425.900	0.000
	After-matched	0.648	0.647	0.200	0.010	0.994

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . %Bias measures the standardized difference in covariate means, scaled by the standard deviation of the sample including both treatment and control groups. The closer it is to zero, the better the balance between the treatment and control groups. Values for *Follower* and *Following* have been scaled.

## D. Additional Analyses and Robustness Tests

### D.1. Results for Viewership Diversity Analysis

For analyses reported in the main text (as explained in Section 4), we focused on the influence of AI-generated titles on video viewership. Here, we extend this analysis to explore their impact on viewership diversity. We first divided the total videos in our sample into two groups: videos produced by treatment group users and videos by control group users. We then calculated HHI based on valid watches to measure viewership concentration for each group, defined as:

$$\text{HHI} = \sum_{j=1}^N s_j^2 \quad (6)$$

where  $s_j$  represents the market share of video  $j$  within its respective group, calculated as:

$$s_i = \frac{V_j}{\sum_{j=1}^N V_j} \quad (7)$$

where  $V_j$  is the number of valid watches for video  $j$ , and the denominator is the total number of valid watches across all  $N$  videos within the group.

We used 1000 bootstrap iterations for each group to calculate HHI. In each iteration, we randomly sampled videos with replacements from each group to create a bootstrap sample and calculated the HHI for this sample using equations (6) and (7). This process yielded 1000 HHI values for both the treatment and control groups, enabling a bootstrap-based significance test to assess whether the difference in viewership concentration between the two groups was statistically significant. The results in Table 19 show that the treatment group with access to AI-generated titles had a significantly lower HHI of 0.0002 compared to 0.0003 in the control group ( $p < 0.05$ ), representing a 50% reduction in platform-level HHI. A lower HHI indicates a more evenly distributed viewership across videos, reflecting greater diversity. This substantial increase in viewership diversity is consistent with our earlier finding in Section 4 that AI-generated titles disproportionately benefited lower-skilled producers.

**Table 19 Results of Platform-level HHI Analysis**

Treatment Group	Control Group	Difference	P-value
(1)	(2)	(3)	(4)
0.0003	0.0002	-0.0001	< 0.05

*Notes:* We perform 1000 bootstrap iterations to compute the HHI and p-value.

## D.2. Results for Channel Analysis

In Section 4.3, we suggested that the positive effect of AI-generated titles on video viewership is attributed to enhanced recommendation accuracy. Theoretically, if true, this effect should benefit both recommendation channels on Platform A (organic and search-oriented recommendations). To test this, we used detailed video consumption data to identify the recommendation source for each watch. Next, we aggregated the valid watch counts for each video by channel and replicated the analysis in Equation (1) for each channel. The results, presented in Table 20 and 21, show a 0.3% increase in both valid watches and watch duration for treatment group videos in the search-oriented channel, and a stronger 1.5% increase in valid watches with a 0.9% increase in watch duration in the organic channel. These findings validate the effectiveness of AI-generated titles in enhancing viewership metrics across both channels.

**Table 20 Results for Search-Oriented Recommendations**

Dependent Variable	<i>ValidWatch</i>	<i>WatchDuration</i>
	(1)	(2)
<i>Treat</i>	0.003*** (0.0003)	0.003*** (0.0002)
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.166	0.165

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

**Table 21 Results for Organic Recommendations**

Dependent Variable	<i>ValidWatch</i>	<i>WatchDuration</i>
	(1)	(2)
<i>Treat</i>	0.015*** (0.001)	0.009*** (0.001)
Controls	YES	YES
Observations	10,738,984	10,738,984
R-square	0.295	0.321

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

### D.3. Results for Video Production Comparison

In Section 4.3, we suggested that the positive effect of AI-generated titles on video viewership is attributed to enhanced recommendation accuracy. An alternative explanation for the increased viewership could be a supply-side shift in production behavior. Specifically, treatment producers with access to AI-generated titles might have altered their production behavior by focusing on producing fewer but potentially higher-quality videos. This shift could theoretically enhance perceived video quality, thereby increasing viewers' content consumption. To examine this possibility, we conducted a producer-level  $t$ -test comparing the total number of videos and average time gaps (in hours) between videos produced by each producer in the treatment and control groups during the treatment period. As shown in Table 22, there is no statistically significant difference in the number of videos produced ( $p$ -value  $> 0.1$ ) and the time gaps between videos ( $p$ -value  $> 0.1$ ) across the two groups. These results suggest that the increase in viewership in the treatment group is unlikely to be driven by a change in producers' video production behavior.

**Table 22 Comparison of Video Production Between Treatment and Control Producers**

Variable	Treatment Group	Control Group	Difference	$P$ -value
<i>NumVideo</i>	5.246	5.240	0.006	0.660
<i>TimeGap</i>	51.241	51.249	-0.008	0.928

### D.4. Results for Alternative Operationalization of Variables

To ensure the robustness of our main findings, we employed alternative operationalizations for key variables in our analysis. Given that the observed positive effect of AI-generated titles may vary depending on how viewership and adoption are defined, testing these alternative definitions allows us to verify the consistency and generalizability of our results. First, for viewership, we used the number of watches ( $Watch_{ij}$ ), complete watches ( $CompleteWatch_{ij}$ ), and the number of likes ( $Like_{ij}$ ) per video as alternative dependent variables. These measures capture different aspects of viewer engagement and may offer additional insights into how AI-generated titles impact a range of interactive forms of engagement. The ITT and LATE results, shown in Tables 23-26, are

qualitatively consistent with our findings in Section 4.1 and 4.2, confirming that AI-generated titles positively impact users' content consumption and engagement.

**Table 23 ITT Results Using Alternative Dependent Variables**

Dependent Variable	<i>Watch</i>	<i>CompleteWatch</i>	<i>Like</i>
	(1)	(2)	(3)
<i>Treat</i>	0.015*** (0.001)	0.007*** (0.001)	0.021*** (0.001)
Controls	YES	YES	YES
Observations	10,738,984	10,738,984	10,738,984
R-square	0.313	0.252	0.297

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

**Table 24 Heterogeneous ITT Results Using Alternative Dependent Variables**

Dependent Variable	<i>Watch</i>		<i>CompleteWatch</i>		<i>Like</i>	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Treat</i>	0.017*** (0.001)	-0.0003 (0.002)	0.008*** (0.001)	-0.001 (0.001)	0.025*** (0.001)	0.011*** (0.001)
<i>LowSkill</i>		-0.911*** (0.002)		-0.623*** (0.001)		-0.849*** (0.001)
<i>Treat * Utilitarian</i>	-0.035*** (0.004)		-0.013*** (0.004)		-0.056*** (0.003)	
<i>Treat * LowSkill</i>		0.022*** (0.002)		0.010*** (0.002)		0.014*** (0.002)
Controls	YES	YES	YES	YES	YES	YES
Observations	10,738,984	10,738,984	10,738,984	10,738,984	10,738,984	10,738,984
R-square	0.313	0.252	0.252	0.190	0.297	0.258

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . Values in parentheses are robust standard errors. *Utilitarian* is absorbed by the video category dummies in *Controls* and therefore not reported in the table.

**Table 25 LATE Results for Main Effects Using Alternative Dependent Variables**

Dependent Variable	<i>Watch</i>	<i>CompleteWatch</i>	<i>Like</i>
	(1)	(2)	(3)
<i>Adopt</i>	0.063*** (0.004)	0.029*** (0.004)	0.091*** (0.004)
Controls	YES	YES	YES
Observations	10,738,984	10,738,984	10,738,984
R-square	0.313	0.252	0.297

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

Second, in Section 3.3,  $LowSkill_i$  is coded as 1 if producer  $i$ 's number of followers is below the median. For robustness checks,  $LowSkill_i$  is coded as 1 if the cumulative number of videos posted



**Table 26 Heterogeneous LATE Results Using Alternative Dependent Variables**

Dependent Variable	<i>Watch</i>		<i>CompleteWatch</i>		<i>Like</i>	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Adopt</i>	0.073*** (0.005)	-0.0008 (0.007)	0.033*** (0.004)	-0.006 (0.006)	0.107*** (0.004)	0.051*** (0.006)
<i>LowSkill</i>		-0.913*** (0.002)		-0.624*** (0.001)		-0.852*** (0.001)
<i>Adopt * Utilitarian</i>	-0.162*** (0.020)		-0.250*** (0.016)		-0.058*** (0.018)	
<i>Adopt * LowSkill</i>		0.089*** (0.008)		0.052*** (0.007)		0.042*** (0.008)
Controls	YES	YES	YES	YES	YES	YES
Observations	10,738,984	10,738,984	10,738,984	10,738,984	10,738,984	10,738,984
R-square	0.313	0.252	0.252	0.190	0.297	0.258

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . Values in parentheses are robust standard errors. *Utilitarian* is absorbed by the video category dummies in *Controls* and therefore not reported in the table.

by producer  $i$  exceeded the median (313). The ITT and LATE results in Table 27 are qualitatively consistent with our main findings.

**Table 27 Model Estimation Results Using Alternative Measures for Low-skilled Producers**

Dependent Variable	ITT		LATE	
	<i>ValidWatch</i> (1)	<i>WatchDuration</i> (2)	<i>ValidWatch</i> (3)	<i>WatchDuration</i> (4)
<i>Treat</i>	0.006*** (0.001)	-0.0002 (0.001)	0.025*** (0.006)	-0.0005 (0.005)
<i>LowSkill</i>	-0.162*** (0.002)	-0.188*** (0.001)	-0.164*** (0.002)	-0.189*** (0.001)
<i>Treat * LowSkill</i>	0.009*** (0.002)	0.009*** (0.002)	0.050*** (0.000)	0.043*** (0.008)
Controls	YES	YES	YES	YES
Observations	10,738,984	10,738,984	10,738,984	10,738,984
R-square	0.204	0.223	0.203	0.222

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

Third, in Section 4.2, *Adopt* was defined as an exact match between the AI-generated title and the posted video title. However, exact matching may overlook cases where titles are similar but not identical. To address this, we coded *Adopt* as 1 if the cosine similarity between the AI-generated title and the posted title met or exceeded thresholds of 95%, 90%, 85%, and 80%. This approach allows us to capture adoption behaviors more flexibly and examine whether even partial adoption drives viewership outcomes. Next, we re-estimated Equation (2) and (3) in our LATE analysis, using *ValidWatch* as the dependent variable. The positive coefficients of *Adopt* in Table 28 validate the effectiveness of AI-generated titles in enhancing content consumption outcomes.

Lastly, we used the Levenshtein distance, also known as the edit distance, to compute the similarity between AI-generated and posted titles as an alternative to cosine similarity used in Section

**Table 28** Alternative Adoption Thresholds of AI-generated Titles and Effect on Content Consumption, DV = ValidWatch

Similarity Threshold	> 95%	> 90%	> 85%	> 80%
	(1)	(2)	(3)	(4)
<i>Adopt</i>	0.068*** (0.004)	0.068*** (0.004)	0.067*** (0.004)	0.066*** (0.004)
Controls	YES	YES	YES	YES
Observations	10,738,984	10,738,984	10,738,984	10,738,984
R-square	0.295	0.295	0.295	0.295

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

5. This metric calculates the minimum edits (e.g., insertions, deletions, substitutions) needed to transform one title into another, with higher distances indicating lower similarity. Using this alternative measure, we then replicated the analysis in Section 5. The positive coefficients of *Treat* and negative coefficients of the interaction term between *Treat* and *Similarity*, as shown in Tables 29 and 30, are qualitatively aligned with our main analysis. Together, these alternative variable operationalizations indicate that our analyses are robust across different measures of consumption, adoption, and title similarity.

**Table 29** Model Estimation Results Using Levenshtein Algorithm (Matched Sample)

Dependent Variable	<i>ValidWatch</i>	<i>WatchDuration</i>
	(1)	(2)
<i>Treat</i>	0.251*** (0.002)	0.208*** (0.002)
<i>Treat</i> * <i>Similarity</i>	-1.062*** (0.003)	-0.880*** (0.002)
Controls	YES	YES
Observations	3,885,089	3,885,089
R-square	0.385	0.396

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

**Table 30** Results of Lexical Richness Analysis Using Levenshtein Algorithm (Matched Sample)

Dependent Variable	<i>LexicalDensity</i>	<i>LexicalVariation</i>	<i>Entropy</i>
	(1)	(2)	(3)
<i>Treat</i>	0.031*** (0.0002)	0.034*** (0.0002)	0.030*** (0.001)
<i>Treat</i> * <i>Similarity</i>	-0.263*** (0.0003)	-0.322*** (0.0002)	-0.464*** (0.001)
Controls	YES	YES	YES
Observations	3,885,089	3,885,089	3,885,089
R-square	0.634	0.689	0.802

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