

# Crafting Tomorrow: The Influence of Design Choices on Fresh Content in Social Media Recommendation

Srijan Saket\*  
srijanskt@gmail.com  
ShareChat  
USA

Mohit Agarwal\*  
mohitagarwal0212@gmail.com  
ShareChat  
India

Rishabh Mehrotra  
erishabh@gmail.com  
Sourcegraph  
UK

## Abstract

The rise in popularity of social media platforms, has resulted in millions of new, content pieces being created every day. This surge in content creation underscores the need to pay attention to our design choices as they can greatly impact how long content remains relevant. In today's landscape where regularly recommending new content is crucial, particularly in the absence of detailed information, a variety of factors such as UI features, algorithms and system settings contribute to shaping the journey of content across the platform. While previous research has focused on how new content affects users' experiences, this study takes a different approach by analyzing these decisions considering the content itself.

Through a series of carefully crafted experiments we explore how seemingly small decisions can influence the longevity of content, measured by metrics like Content Progression (CVP) and Content Survival (CSR). We also emphasize the importance of recognizing the stages that content goes through underscoring the need to tailor strategies for each stage as a one size fits all approach may not be effective. Additionally we argue for a departure from traditional experimental setups in the study of content lifecycles, to avoid potential misunderstandings while proposing advanced techniques, to achieve greater precision and accuracy in the evaluation process.

## CCS Concepts

• **Computing methodologies** → *Online learning settings*.

## Keywords

Recommendations, Early Stage, Design Choices, Fairness

## ACM Reference Format:

Srijan Saket\*, Mohit Agarwal\*, and Rishabh Mehrotra. 2024. Crafting Tomorrow: The Influence of Design Choices on Fresh Content in Social Media Recommendation. In *Proceedings of the 3rd International Workshop on Industrial Recommendation Systems (IRS 2024), Held in conjunction with CIKM-2024, October 25, 2024, Boise, ID, USA* ACM CIKM 2024 3rd International Workshop on Industrial Recommendation Systems, Boise, Idaho, USA, 8 pages. <https://doi.org/10.1145/3627673.3679083>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
CIKM IRS 2024, October 25, 2024, Boise, ID, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 979-8-4007-0436-9/24/10  
<https://doi.org/10.1145/3627673.3679083>

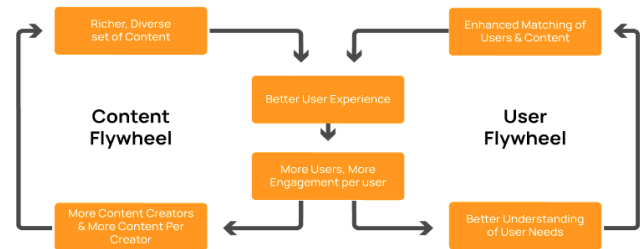


Figure 1: Content and User Flywheel

## 1 Introduction

In the past few years, video platforms have become extremely popular, changing how people watch and interact with digital content. Platforms like Instagram Reels, TikTok, and YouTube Shorts offer a wide variety of videos that differ in length, genre, and type, making it tricky to recommend the right content effectively. ShareChat alone receives approximately 2 million new pieces of content daily, complicating efforts to determine the right UI to surface content, select the correct algorithms for accurate recommendations, and make system choices, especially when considering budget constraints.

There is a rich body of literature that discusses fairness, primarily from user-centric perspectives (either individual or group-level) [9, 16, 22]. While we acknowledge the importance of this aspect in recommendation systems, we argue that it is not sufficient on its own. Extensive work has also been done on provider-side fairness and its impacts, such as Biega et al.'s work on Amortized Attention [2], Diaz et al.'s work on Expected Exposure [8], and Zehlike et al.'s work on fair ranking [26, 27], among others. Additionally, there is further research on balancing multi-sided concerns, such as Mehrotra et al. [17], and mixed-perspective work like that of Epps-Darling et al. [10]. This paper, however, asserts that understanding the different phases content goes through and the decisions that determine content fairness also play a crucial role in the operation of these video platforms.

The content flywheel in Fig 1 illustrates that a rich and diverse content set leads to a better user experience, which in turn drives higher user engagement, motivating creators to produce more content. The user flywheel is tightly coupled with the content flywheel in the sense that an improved user experience (driven through content and others) leads to increased engagement, providing more feedback data to better understand user interests, which ultimately results in richer training data for machine learning models. To the best of our knowledge, there is limited work on understanding the collective impact of UI, algorithm and system choices for early stage

\*Equal Contribution

recommendation on different stages of content lifecycle [Section 2].

We posit that the content lifecycle in short video applications (and other content marketplaces) can be characterized by multiple distinct stages: the initial phase, the intermediate phase, and the mature phase till it finally expires [23, 25, 30]. Each of these stages presents unique attributes which are covered in detail in section 2.

These varying stages necessitate practitioners to adopt distinct models and make specific design choices tailored to the specific phase of content development [13, 20, 28]. The decisions made regarding design at each stage of content development exert a substantial influence on the success of content within the platform. Consequently, this has a ripple effect on both users and creators, ultimately playing a crucial role in the patterns of content consumption throughout the entire platform.

This work centers on examining how design decisions made in the initial phases of content development have a cascading effect on the success of content, user experiences, and the overall consumption patterns on a large-scale platform [3, 4, 14]. This includes decisions including number of views received by a fresh content, algorithm used for recommendation, as well as user interface and surface choices to provide visibility to the content.

The paper delves into various crucial aspects, showcasing how a combination of system, algorithm and design choices can have a catastrophic impact on content success within the platform. Major contributions of the paper include:

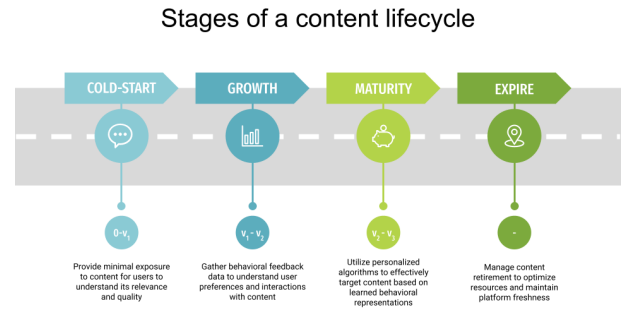
- (1) Identifying the impact of varied minimum exposure to fresh content on content progression outcomes.
- (2) Recognising the importance of differential treatment to time sensitive categories in early phase of the recommendation funnel [20].
- (3) Establishing that both the initial impression count and its rate are pivotal in shaping content longevity on the platform.
- (4) Analyzing how advanced algorithms for personalisation affect content development and persistence.
- (5) Examining the influence of presenting new content across various user interface options using metrics related to content advancement and user contentment.

## 2 Content Lifecycle

### 2.1 Overview of Content Lifecycle

The content lifecycle on a social media platform represents the sequential process through which content progresses, aiming to maximize its reach and engagement [11, 18, 21]. This content propagation creates a flow of information where each stage informs the next, allowing the platform to adapt to user behavior and preferences over time, as shown in Fig. 2. This ongoing process helps maintain a dynamic and engaging content ecosystem. Here’s how it works in four stages:

- (1) **Early stage recommendation:** In the ‘early’ phase, behavioral data is notably absent. The main objective at this stage is to introduce content to a limited audience to gather initial feedback and engagement, which helps gauge its relevance and quality. This process aids in assessing the content’s potential for wider appeal.



**Figure 2: Proposed stages a content goes through during its existence on a short-video platform**

- (2) **Growth:** As content progresses to the ‘growth’ phase, there is a marginal increase in available behavioral data, although it is still accompanied by a limited understanding of the content. The key objective at this stage is to generate behavioral feedback data to learn behavioral representations [19, 29]. We analyze user interactions (likes, comments, shares, etc.) to identify patterns and preferences. This data is then used to refine the content’s targeting and presentation.
- (3) **Maturity:** In the ‘mature’ phase, an abundance of behavioral data becomes available for analysis and the implementation of supervised machine learning models. The key objective at this stage is to utilize personalized algorithms for effective targeting. These personalized algorithms use the learned behavioral representations to fine-tune content recommendations, maximizing the likelihood of engagement by presenting content that aligns with user preferences.
- (4) **Expiration:** Content is allowed to naturally phase out after it has run its course. Additionally, a defined threshold ensures that storage and serving resources are optimised, maintaining platform freshness.

### 2.2 Data Context

**2.2.1 App Description.** We consider production traffic from one of the largest multi-lingual short video applications, ShareChat<sup>\*</sup>, that delivers content in over 14 languages, with a user base exceeding 180 million monthly active users. All the content are user-generated belonging to a variety of genres.

**2.2.2 Details of Data.** We compiled a dataset of video posts created in the Hindi language for an in-depth analysis. We expect other languages to behave similarly because of the randomness in the data selection process. This dataset includes *randomly* sampled user interaction data collected over two months from more than 4 million users, covering over 14 million video posts. This allows us to track the evolution of content across different viewership levels and its effect on user satisfaction metrics. We assessed this impact through both implicit indicators, such as successful video plays (1 if  $watch\_time > 0.98 \times video\_duration$ , 0 otherwise) and skips ( $watch\_time < 3s$ ), as well as explicit signals like clicks, likes, and

<sup>\*</sup><https://sharechat.com/>

shares. Additionally, the dataset includes information about the feed that led to an impression [Section 3.3]. The dataset is structured as time-series data, with each post associated with a view counter at different points in time. In the experiments, users are randomly assigned to different treatment groups to assess the influence on satisfaction metrics.

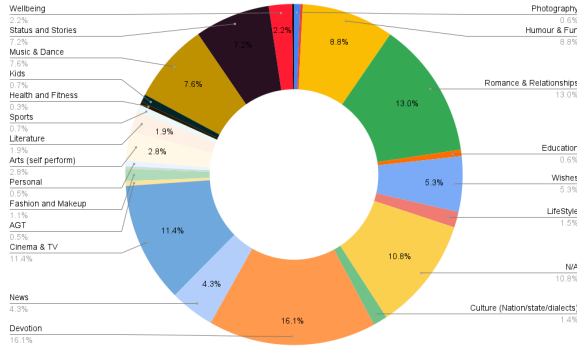


Figure 3: Genre Distribution

2.2.3 *Genres.* The dataset includes content that falls into about 25 broad categories, ranging from topics like *Devotion* and *News* to *Romance & Relationships*, among others. Some categories contain a higher density of content compared to others, as shown in Fig. 3. Our analysis explores how the initial exposure of content influences changes in its distribution across these categories on the platform. Additionally, we examine whether all categories behave uniformly or if there are notable differences worth highlighting.

### 2.3 Content Lifecycle

The success of content on short-video apps largely depends on how it's managed in the early stages. Initial exposure acts as a test to see if the content is relevant and appealing to users. Poor targeting during this time can harm the content's long-term success. But this is just the beginning. Later stages should focus on gathering useful feedback from user interactions to differentiate between great and average content. The goal of a strong *content lifecycle* is to develop a clear understanding of the content by combining user behavior with content semantics. This final phase is known as the 'Mature' stage. The more content that reaches this mature stage, the larger the pool of high-quality content available for recommendations (referred to as the candidate pool [6, 7]), ensuring users are satisfied and the content ecosystem remains healthy.

2.3.1 *Definitions.* The content lifecycle unfolds in distinct stages:

- **Reaching 'minViews':** At this initial stage, a fixed budget is allocated to all platform content. We represent it in this paper as  $views_{min}$  for the platform. Specific slots on a user's feed are reserved for exposure. A user's feed serves as the virtual space where content consumption takes place.
- **Reaching Maturity:** Achieving maturity involves learning a high-quality, low-dimensional representation of content (32

dimensions in our case), primarily through behavioral feedback gathered during initial exposure. Such systems leverage models like Field-aware Factorization Machines (FFMs), wide-and-deep neural networks, and two-tower models to learn vector representations for users and content, often referred to as *embeddings* [5, 12, 15]. Notably, content performance metrics show improvement with refined representations [19]. The paper uses the average across all posts as a threshold for this stage.

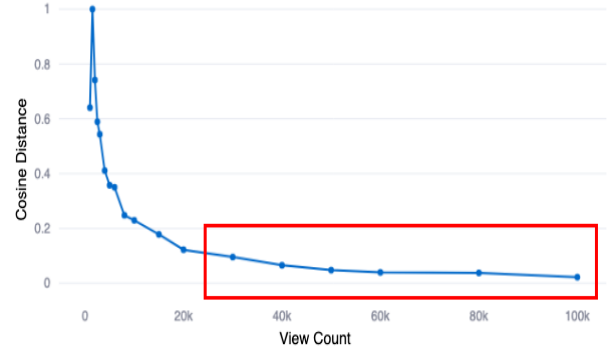


Figure 4: Embedding Maturity Curve

- **Content Progression:** Content shelf-life depends on the interaction of various lifecycle stages [2]. This section explores how specific design, algorithmic, and system choices affect the likelihood of content surpassing certain view thresholds. These thresholds are carefully selected to reflect peaks in the content maturity curve. Fig 4 illustrates the saturating cosine distance curve of content embedding at a given time  $t$  and the embedding at the converged state, plotted against the number of views  $x$  the content has received on the platform. We will use **Conditional View Probability (CVP)** as one of the metrics to measure content progression in early-stage recommendation problems. CVP measures the probability of content reaching a certain number of views ( $x$ ) given that it has already received at least a minimum number of views ( $y$ ). We propose the following formula for CVP:

$$CVP(x | y) = \frac{|\{c \in C | v(c) \geq x, v(c) \geq y\}|}{|\{c \in C | v(c) \geq y\}|} \quad (1)$$

where,  $C$  is the set of all content;  $v(c)$  is the number of views for content  $c$ ;  $x$  and  $y$  are view thresholds.

- **Content Success Metrics:** Evaluating content success depends on both user satisfaction and content progression metrics. Additionally, the study examines how category reach changes based on different algorithmic and system choices. We measure this using the **Content Survival Rate (CSR)**, which is the probability that a piece of content remains active and relevant over a specified period, given design choices  $\tilde{X}$ . By 'active,' we mean that the content continues to receive a significant number of views organically. We propose the following formula for CSR:

$$CSR(t' | y, t, \tilde{X}) = \frac{|\{c \in C | v(c, t + t') - v(c, t) \geq x\}|}{|\{c \in C | v(c, t) \geq y\}|} \quad (2)$$

where,  $x, t > 0$ ; the numerator is the set of all items that receive a minimum of  $x$  views in between time  $t$  and  $t'$ , suggesting content activeness; and the denominator is set of all items with at least  $y$  views within time  $t$ .

Each stage of the content lifecycle plays a crucial role in determining the overall success and longevity of content on the platform.

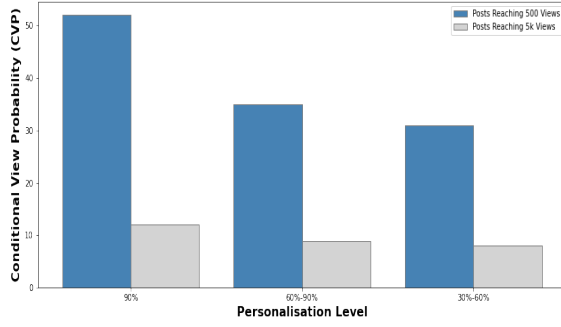


Figure 5: Effect of varying Personalisation on CVP

### 3 Design Choices Impact Success Outcomes

#### 3.1 System Design Choice

**3.1.1 Number of minimum impressions.** : The value of  $views_{min}$  itself is a fundamental aspect of system design choice. To investigate this, we experimented with varying  $views_{min}$  while keeping the other parameters consistent.

**3.1.2 Latency to provide minimal exposure.** : Latency, defined as the time taken to ensure the post receives the minimum guaranteed views,  $views_{min}$ , is another significant design choice. We segmented content into different categories to understand the impact of latency on time-sensitive vs non time-sensitive categories.

Our hypothesis is that by adjusting the  $views_{min}$  based on system design choices, the platform can better improve content progression and success metrics. Additionally, optimizing latency, especially for time-sensitive content, can greatly increase the reach and effectiveness of such posts on the platform.

#### 3.2 Algorithmic Personalisation Choices

We discuss three distinct approaches regarding how personalisation in the early-stage can affect the content journey.

**3.2.1 Random allocation.** Each new content is assigned a random 32-dim embedding at the start, devoid of any personalized information or guidance from the platform. This approach provides no prior knowledge about user preferences or content genres. This method is simplest to implement but lacks personalisation, potentially resulting in suboptimal recommendations. Users may not find

content that aligns with their interests, leading to reduced engagement and satisfaction. For each new post  $p_i$ , a random embedding  $e_{random,i}$  is assigned:

$$e_{random,i} \sim \mathcal{N}(0, \sigma^2)$$

**3.2.2 Genre-based Average.** In this method, we use a semi-personalized strategy by leveraging the average embeddings of high-view posts ( $v(c) \geq 10k$ ) based on content genres. When new content is created within a specific genre, its embedding is initialized as the average embedding of high-view posts in that genre. This method offers a moderate level of personalization by considering the performance of previously successful content within the same genre, which enhances the likelihood of higher views for the new post. Let  $G_j$  represent the set of posts in genre  $j$ , and  $e_{genre-avg,j}$  be the average embedding of high-view posts in that genre. For a new post  $p_i$  in genre  $j$ , its embedding  $e_{genre-based,i}$  is initialized as the average genre embedding:

$$e_{genre-based,i} = e_{genre-avg,j} \quad (3)$$

where

$$e_{genre-avg,j} = \frac{1}{|G_j|} \sum_{p \in G_j} e_p$$

**3.2.3 Model-based.** This involves a fully personalized model-based embedding initialization. Here, we develop a sophisticated model based on MEMER [1] that learns to initialize embeddings for new posts using both post embeddings of high-view content and multimodal content features (e.g., visual, text, audio). Details of the model are covered in the paper linked. Utilizing a model-based approach achieves a high degree of personalisation as the embeddings are tailored to individual posts, considering not only the genre but also the content's unique characteristics. This personalized initialization aims to enhance recommendations, leading to increased CVP, CSR and user satisfaction. The MEMER model is trained using interactions, high-view post embeddings ( $e_p$ ), user embeddings ( $u_p$ ), and multimodal content features ( $f_p$ ). However, during inference for a new post  $p_i$ , only the content features ( $f_i$ ) can be utilized to infer a new embedding:

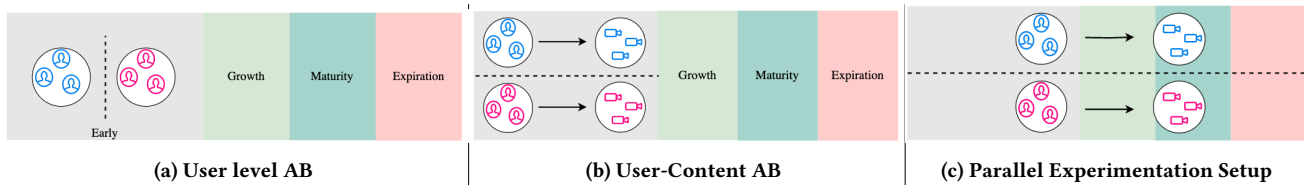
$$e_{model-based,i} = MEMER\_Inference(f_i, MEMER(e_p, u_p, f_p, interactions))$$

Summarising, selecting the appropriate embedding initialization has significant impact on how the  $views_{min}$  are provided and how the content continues to progress beyond that. We look at offline metrics AUC, F1, Relatmpr [24] for comparison of algorithms and online metrics to track the content progression (CVP) and user satisfaction (engagement).

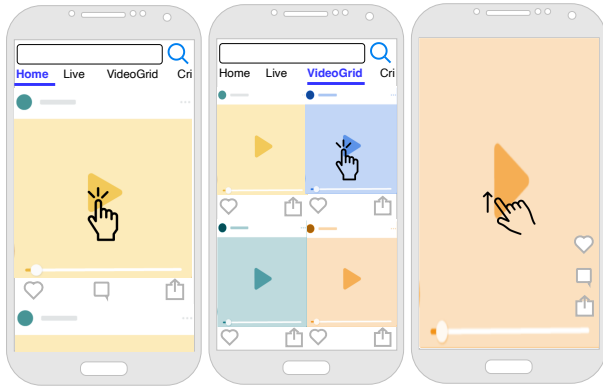
#### 3.3 UI Surface Choices

The user interface (UI) in a recommender system profoundly influences how users engage with the provided content. In this paper, we examine the effects of three UI surface designs on content progression and user interaction as illustrated in Figure 7.

- (1) *HomeFeed - Curated Choice:* It presents a vertically scrollable interface with videos that require a click to play. At any given time, a maximum of two videos are visible on the screen, similar to Instagram's home feed. In a single feed fetch, around 12 videos are loaded, and the content refreshes



**Figure 6: The figure shows evolution of experimentation framework to gauge the impact of design choices across user satisfaction and platform content distribution**



**Figure 7: Representation of different UI Surface Choices - Home, VideoGrid, VideoScroll Feed resp.**

as you scroll down. Users actively engage with the content, explicitly selecting videos based on their preferences. The intent is to encourage active user engagement and promote user-driven content selection.

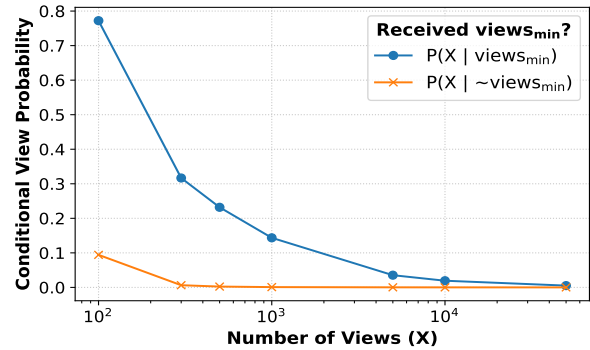
- (2) *VideoGridFeed - Variety Grid*: It displays video content as multiple tiles on a single screen, with each video requiring a click to play. The videos are arranged in a 4x4 grid, allowing users to view a variety of options at the same time. This setup aims to balance user choice and exposure, potentially leading to a more diverse content experience.
- (3) *VideoScrollFeed - Seamless Stream*: This UI offers autoplayed videos, and users can scroll to move to the next video, similar to popular platforms like Reels or TikTok. At any given time, only one video is in full-screen mode. On this surface, users consume content automatically without needing to make an explicit selection, so content progression is continuous and smooth. The goal is to make content consumption effortless and maximize consumption efficiency.

We investigate the impact of providing  $views_{min}$  across different UI surfaces by measuring the CVP and engagement across them.

## 4 Experimentation & Results

### 4.1 Experimentation Setup

**4.1.1 User level AB.** : In the conventional AB test setup, the user base is divided into two equal groups: test and control as shown in Fig 6a. The test group is exposed to treatment  $\tilde{X}_1$ , while the control group is exposed to  $\tilde{X}_2$ . Comparative metrics are used to measure



**Figure 8: CVP if content receives  $views_{min}$  vs. it doesn't**

the relative difference between the two groups, and the results are reported.

However, a major challenge arises due to the variable budget. For instance, if the test variant starts favoring highly-engaging categories like "Romance & Relationships", the distribution of views across test and control groups becomes uneven. When this is scaled to 100%, the relative numbers no longer hold true.

**4.1.2 User-Content AB.** : This modification extends the previous setup by not only splitting the user base but also the content corpus; shown in Fig 6b. Each user set is paired 1:1 with a specific content set. This helps control spillage, as the effect of content set  $\tilde{C}_1$  is confined to user set  $\tilde{U}_1$  and does not impact users in the alternate set  $\tilde{U}_2$ , and vice versa.

However, a limitation arises when experiments are confined to early-stage recommendation analysis. As the content progresses through its lifecycle, both groups may begin to influence each other, potentially leading to discrepancies in the scaled setup.

**4.1.3 Parallel Experimentation Setup.** : This represents an advanced iteration of the user-content level AB testing. Here, the split is implemented across various stages of the content lifecycle, allowing for a more comprehensive evaluation. The framework is described in Fig 6c.

This setup is budget-conscious and enables the measurement of metrics beyond the early stage, such as assessing the impact of design choices during the cold-start phase on content progression up to expiration. Nevertheless, it is acknowledged that this setup is more intricate and costly to establish compared to the previous

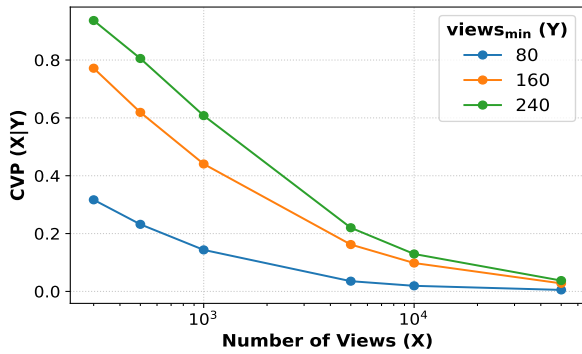
	AUC	F1	RelImpr (%)
Random	0.500	0.151	-0.15
Genre Average	0.605	0.190	66.7
MEMER	0.631	0.207	83.3

(a) Offline Metrics for Algorithmic Choices for Successful Video Watch

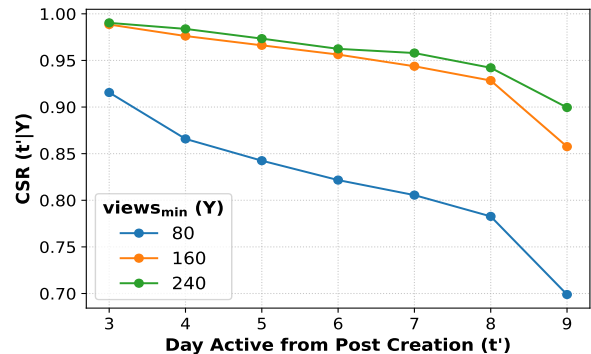
	Genre Average	MEMER
CVP(500  $views_{min}$ )	0.7235	0.7646
Engagement/Views	0.0099	0.0145
Successful Video Play	0.2633	0.2871

(b) Online metrics for Algorithmic Choices for User-Post AB Test

Table 1: Offline and Online Results from the Algorithmic Personalisation Choices



(a) CVP( $x|y$ )



(b) CSR( $t'|y$ )

Figure 9: Content Progression Metrics given Differential  $views_{min}$

two methods. Additionally, there may still be minor data leakages if there are multiple sources contributing to views.

## 4.2 Across System Design Choices

**4.2.1 Impact of Meeting Minimum View Thresholds.** To establish the initial motivation for providing  $views_{min}$ , we compare two scenarios: one where content does not receive the promised views within 48 hours of creation (this time frame can be adjusted), and the other where it does. As shown in Figure 8, we observe that content that achieves the minimum pledged views within this time frame exhibits a higher progression rate (CVP) compared to content that does not. Notably, this difference is particularly pronounced for lower view thresholds ( $< 1k$ ) and eventually levels off.

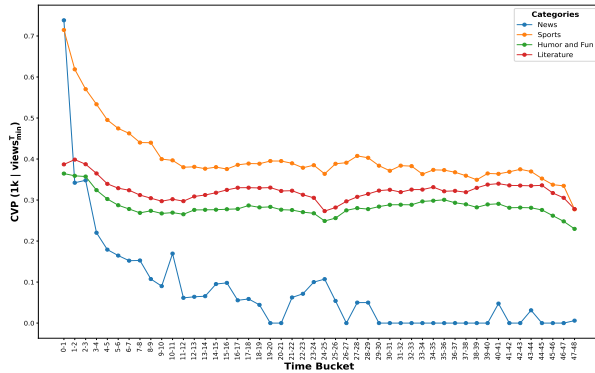


Figure 10: CVP( $1k|views_{min}^T$ ) across content categories

**4.2.2 Optimizing View Allocation for New Content.** Next, we explore varying the allocated view budget ( $views_{min}$ ) for new content on the platform while keeping the algorithm and UI consistent throughout the experiment. In Figure 9, we observe that generally, allocating a higher  $views_{min}$  results in more posts achieving a greater number of views (CVP) and a longer shelf life (CSR). However, the relative increase becomes less pronounced as we continue to raise the  $views_{min}$  threshold.

While this observation seems straightforward, its practical implications are more complex. It suggests that simply increasing  $views_{min}$  won't continue to positively affect content progression metrics indefinitely. This is because new content generally performs worse on the platform compared to *mature* content due to the lack of user feedback data. This introduces an optimization challenge: we aim to maximize user satisfaction and content progression while ensuring that every new piece of content receives  $views_{min}$ . Without setting a cap on the exposure of new content, user satisfaction could be compromised, negatively impacting overall business metrics.

**4.2.3 Analysis of View Rates Across Content Categories.** Finally, we conduct a more detailed analysis, exploring the impact of view rates across time-sensitive and timeless content categories. As shown in Figure 10, we find that for time-sensitive categories like *News*, the Conditional-View-Probability (CVP) rate is greatly influenced by how quickly we provide the  $views_{min}$ . In fact, the drop in numbers varies significantly between the initial time buckets and later ones. This trend differs in categories like *Humor & Fun* videos, which are not time-sensitive. Posts in these categories tend to consistently achieve the target view count regardless of when the  $views_{min}$  is

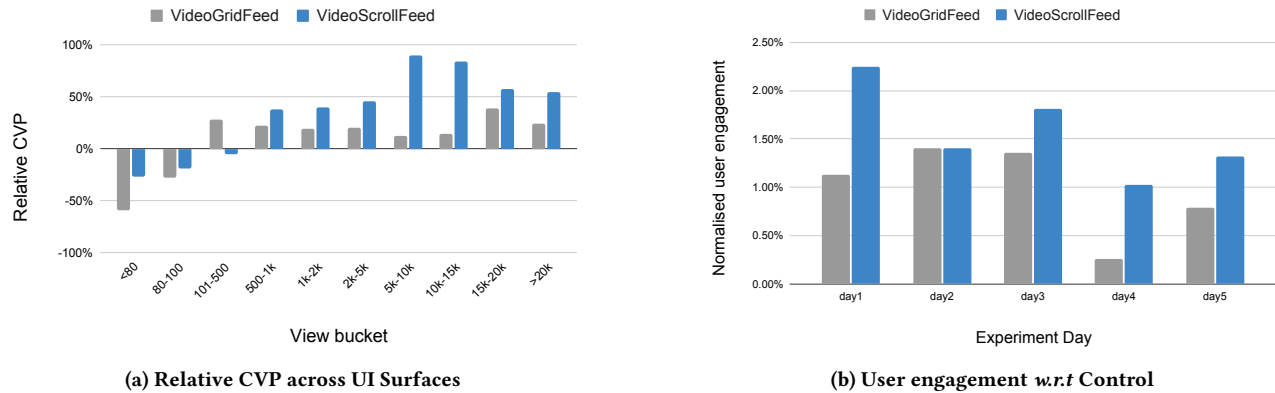


Figure 11: Content & User Engagement Metrics across different UI surfaces

provided. Please note that although there are 25 categories in the dataset, we have picked a few distinct ones to highlight the stark differences.

### 4.3 Across Algorithmic Personalisation Choices

Initially, for motivation, Figure 5 shows the impact of varying personalisation on CVP. Next, we examine the effect of changing the underlying algorithm while keeping the initial view budget and UI surface constant for the experiment, assessing the impact on content progression and user satisfaction. We introduce three approaches: random allocation (for benchmarking), genre average, and MEMER, which are explained in Section 3.2.

The offline and online results of different algorithmic choices are summarized in Table 1. We observe a significant improvement in offline AUC when using a sophisticated algorithm like MEMER compared to heuristic approaches such as random allocation or genre average (83.3% *RelatImpr* (RI) [24] for Successful Video Watch [1a]). These offline gains also translate into online metrics, as seen in higher CVP rates and increased user engagement in Table 1b. User engagement includes explicit signals such as likes, shares, and downloads, as well as implicit signals like successful video watch, which is a binary signal based on video watch time (explained in Sec 2.2.2).

### 4.4 Across UI Surfaces

We keep the algorithm constant and introduce initial exposure to new content across various UI interfaces, as shown in Figure 7. We evaluate content performance across different view buckets and find that video feeds show higher Conditional-View-Probability (CVP) and user engagement rates compared to the control group (*HomeFeed*), as depicted in Figures 11a and 11b. Additionally, among video feeds, *VideoScrollFeed* displays a higher CVP and user engagement rate, potentially because user intent is already captured at this stage of the funnel, making users more likely to explore fresh content. In contrast, *VideoGridFeed* may show lower performance due to the added complexity of requiring a click on the thumbnail for the video to start playing, which impacts content progression metrics. The hierarchy of valuable impressions follows the order: *VideoScrollFeed* > *VideoGridFeed* > *HomeFeed*.

## 5 Discussion and Conclusion

We believe that design decisions have a significant impact on a system’s functionality. Our research explores the effects of these choices on content, examining various aspects, including system configurations, algorithmic selections, and UI designs. We show the tangible effects on content progression and longevity, and further analyze the differences in minimal exposure rates across different content categories. Additionally, we highlight the limitations of conventional A/B testing in accurately capturing content lifecycle metrics and propose more effective evaluation alternatives. This emphasizes the critical importance of informed design decisions in shaping the performance and longevity of content within the system.

There are several promising directions for future research. First, considering a variable budget allocation for new content could be beneficial, as it may not be necessary to allocate the same initial budget for every piece of new content. This adjustment could potentially free up resources to be redirected towards more promising content. Additionally, the distinct behavior observed in time-sensitive categories compared to others warrants further investigation. Exploring strategies for pacing initial views to ensure content relevance presents an intriguing area of study. Moreover, identifying the time-sensitive nature of content through behavioral feedback could offer valuable insights. Finally, an interesting area for future work involves evaluating the impact of design, algorithm, and system choices on content creators. This could be achieved by implementing relevant A/B test frameworks, providing insights into how these choices influence creation patterns and creator incentives.

## References

- [1] Mohit Agarwal, Srijan Saket, and Rishabh Mehrotra. 2023. MEMER - Multimodal Encoder for Multi-Signal Early-Stage Recommendations. In *Companion Proceedings of the ACM Web Conference 2023* (Austin, TX, USA) (*WWW '23 Companion*). Association for Computing Machinery, New York, NY, USA, 773–777. <https://doi.org/10.1145/3543873.3587679>
- [2] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*. 405–414.
- [3] Allison JB Chaney, Brandon M Stewart, and Barbara E Engelhardt. 2018. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM conference on recommender*

- systems. 224–232.
- [4] Minmin Chen. 2021. Exploration in recommender systems. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 551–553.
  - [5] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishu Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 7–10.
  - [6] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*. 191–198.
  - [7] James Davidson, 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. In *Proceedings of the fourth ACM conference on Recommender systems*. 293–296.
  - [8] Fernando Diaz, Bhaskar Mitra, Michael D Ekstrand, Asia J Biega, and Ben Carterette. 2020. Evaluating stochastic rankings with expected exposure. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 275–284.
  - [9] Michael D Ekstrand, Anubrata Das, Robin Burke, Fernando Diaz, et al. 2022. Fairness in information access systems. *Foundations and Trends® in Information Retrieval* 16, 1-2 (2022), 1–177.
  - [10] Avriel Epps-Darling, Henriette Cramer, and Romain Takeo Bouyer. 2020. Artist gender representation in music streaming.. In *ISMIR*. 248–254.
  - [11] Kim Falk. 2019. *Practical recommender systems*. Simon and Schuster.
  - [12] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).
  - [13] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2020), 3549–3568.
  - [14] Kevin Hu, Michiel A Bakker, Stephen Li, Tim Kraska, and César Hidalgo. 2019. Vizml: A machine learning approach to visualization recommendation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
  - [15] Yuchin Juan, Yong Zhuang, Wei-Sheng Chin, and Chih-Jen Lin. 2016. Field-aware factorization machines for CTR prediction. In *Proceedings of the 10th ACM conference on recommender systems*. 43–50.
  - [16] Will Ma, Pan Xu, and Yifan Xu. 2020. Group-level fairness maximization in online bipartite matching. *arXiv preprint arXiv:2011.13908* (2020).
  - [17] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th acm international conference on information and knowledge management*. 2243–2251.
  - [18] Shaina Raza and Chen Ding. 2022. News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review* (2022), 1–52.
  - [19] Srijan Saket, Olivier Jeunen, and Md Danish Kalim. 2024. Monitoring the Evolution of Behavioural Embeddings in Social Media Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2935–2939.
  - [20] Jianling Wang, Haokai Lu, Sai Zhang, Bart Locanthi, Haoting Wang, Dylan Greaves, Benjamin Lipshitz, Sriraj Badam, Ed H Chi, Cristos J Goodrow, et al. 2023. Fresh Content Needs More Attention: Multi-funnel Fresh Content Recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5082–5091.
  - [21] Shoujin Wang, Longbing Cao, Yan Wang, Quan Z Sheng, Mehmet A Orgun, and Defu Lian. 2021. A survey on session-based recommender systems. *ACM Computing Surveys (CSUR)* 54, 7 (2021), 1–38.
  - [22] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems* 41, 3 (2023), 1–43.
  - [23] Siqi Wu, Marian-Andrei Rizoio, and Lexing Xie. 2018. Beyond views: Measuring and predicting engagement in online videos. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.
  - [24] Ling Yan, Wu-Jun Li, Gui-Rong Xue, and Dingyi Han. 2014. Coupled group lasso for web-scale ctr prediction in display advertising. In *International Conference on Machine Learning*. PMLR, 802–810.
  - [25] Honglin Yu, Lexing Xie, and Scott Sanner. 2015. The lifecycle of a youtube video: Phases, content and popularity. In *Proceedings of the international AAAI conference on web and social media*, Vol. 9. 533–542.
  - [26] Meike Zehlike, Ke Yang, and Julia Stoyanovich. 2022. Fairness in ranking, part i: Score-based ranking. *Comput. Surveys* 55, 6 (2022), 1–36.
  - [27] Meike Zehlike, Ke Yang, and Julia Stoyanovich. 2022. Fairness in ranking, part ii: Learning-to-rank and recommender systems. *Comput. Surveys* 55, 6 (2022), 1–41.
  - [28] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)* 52, 1 (2019), 1–38.
  - [29] Wayne Xin Zhao, Sui Li, Yulan He, Edward Y Chang, Ji-Rong Wen, and Xiaoming Li. 2015. Connecting social media to e-commerce: Cold-start product recommendation using microblogging information. *IEEE Transactions on Knowledge and Data Engineering* 28, 5 (2015), 1147–1159.
  - [30] Fan Zhou, Xovee Xu, Goce Trajcevski, and Kungpeng Zhang. 2021. A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Computing Surveys (CSUR)* 54, 2 (2021), 1–36.

Received 15 August 2024