Exploring Design Approaches for Reducing Viewers' Discomfort with Distressing Short-form Videos

Miran Park*
wsjun1214@snu.ac.kr
Information Science and Culture
Studies
Seoul National University
Seoul, South Korea

Kyuri Park*
kyuri117@snu.ac.kr
Information Science and Culture
Studies
Seoul National University
Seoul, South Korea

Hyewon Cho*
cocohw@snu.ac.kr
Information Science and Culture
Studies
Seoul National University
Seoul, South Korea

Hwan Choi*
chlghks123@snu.ac.kr
Information Science and Culture
Studies
Seoul National University
Seoul, South Korea

Hajin Lim
hajin@snu.ac.kr
Information Science and Culture
Studies
Seoul National University
Seoul, South Korea

ABSTRACT

The growing popularity of short-form video platforms and their reliance on algorithmic recommendations highlights the risk of viewers unintentionally encountering distressing content. Therefore, we investigated viewers' experiences with distressing content and developed design approaches to alleviate their discomfort. Through in-depth interviews, we discovered that participants perceived and reacted differently to "socially inappropriate content," which violated societal norms, and to "personally discomforting content," which triggered negative reactions on a personal level. Further, participants expressed frustration with the lack of transparency in content reporting processes, the challenges in tailoring recommendation algorithms to avoid distressing content, and the limitations of post-exposure feedback mechanisms. To address these challenges, we conceptualized three design approaches focused on enhancing reporting process transparency, providing users with granular control over content recommendations, and allowing for preemptive adjustments to their content feeds. Our findings and proposed design approaches may provide valuable directions for improving viewer well-being on short-form video platforms.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

distress, discomfort, algorithmic experience, content curation algorithm, short-form videos, video platform, content filtering, algorithmic recommendation, digital well-being

*These authors contributed equally to this research.

1 INTRODUCTION

In recent years, short-form videos, defined as short videos lasting between 5 and 90 seconds, have surged in popularity on platforms such as TikTok, Reels (of Meta, Instagram), and Shorts (of Youtube). Alongside this growth, there has been a noticeable increase in titillating, provocative, and stimulating short-form videos, often crafted to capture viewers' attention [21]. The situation is further intensified by the content curation algorithms of these platforms, which may inadvertently heighten users' exposure to such distressing content. Designed to offer an immersive viewing experience, these algorithms automatically select and play videos, thus raising the possibility of users unwittingly encountering distressing content without any prior warning.

In response to these challenges, constant efforts have been made to develop algorithms capable of filtering out such content. For example, several multimodal moderation techniques, incorporating motion, visual, and audio data, have been employed to classify harmful content, including hate speech and aggression, across different content sharing platforms [1, 19]. Further, short-form video platforms have outlined what constitutes "harmful" content within their Terms and Conditions (T&C), specifying categories[1] such as violence, sexual abuse. These platforms have also stated that they reserve the right to restrict or terminate the usage rights of individuals who post content falling within these harmful categories.

Yet, such technical approaches and platform policies aimed at censoring and mitigating distressing content often encounter limitations in accurately categorizing content due to the subjective and

---

[1]Violence, dangerous organizations, illegal content, self-harm, child sexual abuse, sexual abuse, human trafficking, bullying and harassment, hate speech, graphic content, sexual solicitation, spam, impersonation, misinformation, political propaganda [3]

dynamic nature of what is considered "distress." The challenge lies in the ambiguity of what constitutes distressing content and its subjective perception by different users [22, 26]. Previous studies have explored users' subjective experiences with distressing content, aiming to capture the broad spectrum of perceptions [14, 24, 25]. They highlighted the importance of a nuanced approach that takes into account the individual experiences of users rather than relying solely on broad, predefined categories. Accordingly, this paper aims to investigate how short-form viewers identify, perceive, and react to distressing content by conducting in-depth interviews.

The inherent ambiguity and subjectivity in defining distressing content possibly lead platforms to adopt passive regulatory measures, largely dependent on user reports based on the categories specified in the platforms' T&C. This approach, however, may fall short in capturing the fuller spectrum of user discomfort. Although platforms have introduced personalization features, such as the "not interested" button, to tailor the viewing experience, the opacity surrounding the impact of such feedback on the curation algorithms may leave users unprotected from distressing content. This situation underscores the pressing need for the development of design approaches that could effectively protect users from distressing experiences on short-form video platforms. In response, our study seeks to propose design approaches that could mitigate the adverse effects of encountering distressing short-form videos.

Specifically, we aim to address two research questions (RQs):

- **RQ1**: How do users define, perceive, and respond to distressing short-form videos?
- **RQ2**: What would be the design approaches to reducing users' discomfort with distressing short-form videos?

In addressing these RQs, we first conducted in-depth interviews with 17 regular short-form video viewers (**Study 1**) to understand the specific challenges and needs these viewers faced when encountering distressing short-form videos (**RQ1**). The insights gained from these interviews informed the development of three design approaches aimed at mitigating the discomfort triggered by such videos (**RQ2**). Then, we gathered feedback on these design approaches to evaluate their potential to alleviate viewer discomfort (**Study 2**). Our research contributes to providing design implications to enhance the safety and well-being of users, ensuring a more secure and enjoyable short-form video viewing experience.

2 BACKGROUND & RELATED WORKS

2.1 Motivation of Short-Form Viewers

As primary motivations for consuming short-form videos, Khan and colleagues [17] identified three primary motives, including relaxation and entertainment, social interaction, and information sharing. Of these, relaxation and entertainment emerged as the primary drivers of content preference, often leading to a more passive consumption pattern. Similarly, in the context of TikTok, Dong and Xie [7] identified primary motivations for TikTok usage, including novelty seeking, stress relief, time passing, and escapism.

Given content consumption on short-form video platforms is largely passive and strongly driven by relaxation and entertainment motives, encountering distressing content may cause users severe discomfort, which could challenge their primary reasons for using

it. To gain deeper insights into this, we investigated the experiences of short-form video viewers with distressing content in Study 1.

2.2 Algorithmic Experience

To understand the user experience with distressing short-form videos, it would be crucial to explore the dynamics of user interactions with and perceptions of the algorithms that curate content on these platforms [15]. The concept of the algorithmic experience (AX), the user's experience and understanding of algorithms [2], provides a valuable framework. In enhancing AX, Alvarado and Waern [2] suggested increasing algorithmic transparency, enabling users to manage and control their algorithmic profiles, and raising overall algorithmic awareness. Further, Hamilton and colleagues [13] explored the optimal level of algorithm visibility for users. Additionally, Lukoff and colleagues [20] examined the balances between user and machine control for content curation algorithms. This body of work emphasizes the importance of ensuring algorithm visibility and transparency while expanding user agency, with a particular focus on users' understanding and interaction with curation algorithms.

Despite these insights, current short-form video platforms often fall short of facilitating meaningful user engagement with their content curation algorithms. This gap highlights the need for developing design approaches that could enhance user agency and control over the algorithmic curation process. Particularly, exploring design approaches to empower users to effectively manage or avoid the adverse effects of encountering distressing content would contribute to creating a safer and more personalized viewing environment. Therefore, we proposed three design approaches aimed at leveraging user control over short-form video curation algorithms and evaluated them in Study 2.

3 STUDY 1: USERS' PERCEPTIONS AND CHALLENGES WITH DISTRESSING SHORT-FORM VIDEOS

3.1 METHOD

To investigate how users perceive and respond to distressing short-form videos (RQ1), we conducted semi-structured interviews. We posted our recruitment post on popular online communities (e.g., Everytime) and recruited 17 participants (11 females and 6 males, with an average age of 25; see Table 1) who had been actively using short-form platforms for more than two months and had encountered distressing short-form videos. Interviews were conducted in person, during which participants were asked about their typical short-form video usage patterns and their experiences with distressing content. In order to emulate real-life usage and elicit genuine responses, we used the think-aloud approach [27], prompting participants to share their thoughts while navigating through their actual short-form video feeds during the interviews. The interviews were conducted in Korean and took approximately 45 minutes each. Participants were compensated with a gift voucher worth KRW 8,000 (approximately $6 USD) for their participation.

With the participants' consent, all interview sessions were audio-recorded, and the recordings were transcribed and analyzed using the thematic analysis method [5]. The thematic analysis process

Table 1: Participant Profile

	ID	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17
	Gender	F	F	F	F	F	F	F	F	F	M	F	M	M	M	F	M	M
Primary platform	Youtube Shorts	O	O	O	O	O	O	O	O	O	O	O		O	O	O	O	O
	TikTok	O																
	Instagram Reels	O		O	O	O	O	O			O	O	O	O	O	O	O	O
	X(Twitter)	O			O						O							
Daily usage duration (hours)		1.5	1.5	1	2	2	1.45	2.45	5	0.2	2	2	1.5	2	1	2	1.3	2
Participation in Study 2		O	X	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O

involved all researchers coding the transcripts individually and collaboratively identifying and developing common themes. Through several rounds of discussion, we characterized the types of content that participants found distressing, along with the challenges they faced in dealing with distressing short-form videos. All research materials and procedures were approved by the Institutional Review Board (IRB) of the university where the study was conducted.

3.2 FINDINGS

The majority of the participants (n=15) were heavy short-form video platform users, spending over an hour daily watching short-form videos. Participants predominantly used short-form platforms to "kill time" during idle moments. They particularly appreciated the accessibility to diverse content in an easy and fast manner: "*I like that I can watch many interesting videos in a short time*" (P09). The convenience of algorithm-driven recommendations was also a key factor in their viewing experience:"*I found the automatic recommendations very convenient*" (P15). Particularly, some participants highlighted the joy of encountering unexpected, entertaining content through these algorithmic recommendations: "*It's so fun and enjoyable when the algorithm recommends unexpected yet interesting content*" (P12).

However, all participants reported continuous exposure to distressing videos while watching short-form videos through platforms' curation algorithms. They noted encountering distressing videos quite frequently: "*Maybe 3 out of 10 videos?*" (P10), "*Usually 3-4 out of 10 videos*" (P02).

As a result, participants expressed substantial frustration due to encountering distressing videos. They mentioned that such experiences not only disrupted their viewing flow:"*I want to watch enjoyable videos, but these kinds of videos distract me*" (P15). Also, they reported being exposed to distressing content sometimes affected their everyday life: "*I'm afraid it might come to mind when I try to sleep*" (P08). Some participants reported experiencing prolonged anxiety and lingering discomfort: "*It doesn't get resolved immediately as an uneasy feeling remains*" (P02). In severe cases, P3 mentioned experiencing symptoms similar to post-traumatic stress disorder: "*I happened to see videos of actual incidents, and for the next two days, I felt really bad and dizzy. For a while, even just looking at something similar to that video made me feel short of breath*" (P03).

Interestingly, most participants made a clear distinction in their experiences with distressing content, categorizing them into two

distinct types: "**socially inappropriate videos**" and "**personally discomforting videos**." When identifying **socially inappropriate content**, participants applied social norms, legal standards, and ethical considerations. The most frequently cited examples of such content included real accidents and crime (64%) and inappropriate or false information (64%), followed by sexually explicit (57%) and violent content (57%).

Conversely, the category of **personally discomforting videos** was defined by a broader and more subjective range of criteria that varied significantly among participants. These videos included content that, while not necessarily socially inappropriate, was personally distressing for various reasons. Examples of such content ranged widely, encompassing specific themes, such as "*political issues*" (P07), "*dance challenges*" (P10), and "*eating shows*" (P16), to specific visual triggers like "*birds*" (P05), "*insects*" (P08), and "*pores*" (P09), as well as certain auditory elements such as "*crying sounds*" (P04), "*synthesized voices*" (P10), and "*loud noises*" (P15).

Participants reported adopting distinct approaches for managing socially inappropriate videos and personally distressing videos. For socially inappropriate content, most participants reported having utilized the platform's "report" function, aiming for the removal of such content to safeguard the wider community. Conversely, when encountering videos that were personally discomforting but not necessarily socially inappropriate, they sought to influence their curation algorithms by signaling their disinterest, typically through options like "Not interested." This dual approach was encapsulated by P2's account: "*I report the video when it makes not only me but others uncomfortable. For videos that are only distressing to me, I press 'Do not recommend.'*" As such, participants made conscientious efforts when dealing with content that impacted the broader user base versus content that affected them personally.

Despite these strategies, many participants expressed frustration with the ongoing presence of distressing videos. Further, they highlighted the specific challenges they faced in effectively minimizing the exposure to distressing videos in their feeds.

*3.2.1 **Obstacles in 'Reporting' Socially Inappropriate content***. Most participants responded that they had used the 'report' function when they found the content socially inappropriate: "*I report things that are morally questionable*" (P06). Reporting was further perceived as a social action that engaged the community to address concerns. For instance, participant P01 shared her experience of seeking help in removing socially inappropriate short-form videos by involving a wider community: "*When deep fake videos*
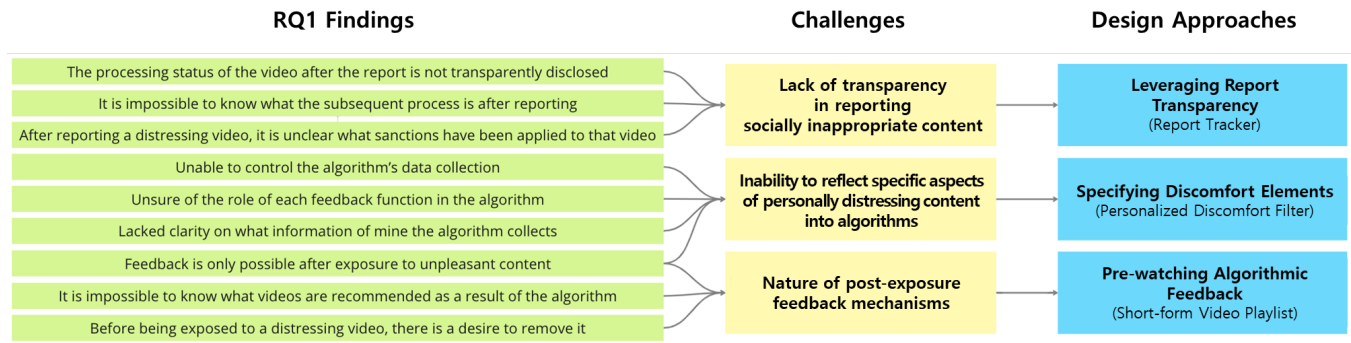
RQ1 Findings	Challenges	Design Approaches
The processing status of the video after the report is not transparently disclosed	**Lack of transparency in reporting socially inappropriate content**	**Leveraging Report Transparency** (Report Tracker)
It is impossible to know what the subsequent process is after reporting		
After reporting a distressing video, it is unclear what sanctions have been applied to that video		
Unable to control the algorithm's data collection	**Inability to reflect specific aspects of personally distressing content into algorithms**	**Specifying Discomfort Elements** (Personalized Discomfort Filter)
Unsure of the role of each feedback function in the algorithm		
Lacked clarity on what information of mine the algorithm collects		
Feedback is only possible after exposure to unpleasant content	**Nature of post-exposure feedback mechanisms**	**Pre-watching Algorithmic Feedback** (Short-form Video Playlist)
It is impossible to know what videos are recommended as a result of the algorithm		
Before being exposed to a distressing video, there is a desire to remove it		

Figure 1: Thematic map of the findings from the data analysis

of celebrities were circulating, I asked for help in reporting them on online communities. Many people responded, and the videos were eventually blocked."

Despite their intention to use the report function, some participants expressed frustration with the lack of feedback and transparency in such reporting processes: "*I got notified that my reporting was received, but then I don't know what happens next. That's a bit frustrating*" (P03). This uncertainty was shared by several participants, with concerns centering on the lack of information about the follow-up procedures and the outcome of the reported content. This lack of transparency had led to diminished trust in platforms' handling of such issues and raised skepticism about reporting: "*It doesn't seem to have any effect. Reporting doesn't mean the video gets deleted*" (P06). Consequently, this deterred some participants from reporting socially inappropriate content: "*Since I don't know how the reports are handled, I tend not to report*" (P15).

*3.2.2 **Obstacles in 'Controlling' Curation Algorithms**.* When confronted with personally discomforting videos, a majority of participants actively attempted to manage the algorithm to avoid similar content in the future, while a few simply chose to swipe away. They mentioned that they frequently used feedback functions like 'Do not recommend,' 'Dislike,' and 'Not interested.' However, despite these efforts, most participants felt the platform did not adequately prevent re-exposure to discomforting content. Therefore, to further "*purify*" (P05) their feed, some tried to influence their algorithms by actively watching or liking a large amount of preferred content or even completely clearing their viewing history. P09 went a step further by avoiding the curated feed altogether, relying solely on the search function to navigate short-form videos and thus exerting more control over her viewing experience.

Overall, participants highlighted significant challenges in accurately reflecting the specific aspects of videos that caused them distress to content curation algorithms. A common observation among participants was the algorithm's effectiveness in recognizing their preferred topics but its inability to adequately identify and filter out disliked topics, leading some participants to stop providing feedback about disinterest: "*I used to press 'Dislike' often, but now I don't. I don't know if it makes a difference*" (P10). As a whole, they felt there was a notable lack of clarity regarding the specific impact of feedback on their curation algorithms. An interesting case was shared by P08, who noticed that attempting to block

unwanted content inadvertently resulted in the loss of preferred content: "*When I blocked a streamer's short-form that I did not like, my favorite streamers' short form stopped appearing, too*" (P08).

More fundamentally, the post-exposure feedback mechanisms, which require users to be exposed to distressing content before being able to provide feedback, were identified as a significant design flaw: "*Isn't it a bit ridiculous? You can only react after you've already felt the discomfort*" (P01).

3.3 Design Requirements

Study 1 findings revealed that participants employed different approaches and encountered specific challenges when dealing with 'socially inappropriate' versus 'personally discomforting' short-form videos. With socially inappropriate content, a significant issue was the lack of transparency in the reporting process, leaving participants uncertain about the outcome of their reports. For personally discomforting videos, participants struggled with effectively influencing the curation algorithm. Also, they expressed frustration with the nature of the current feedback mechanisms, which only allowed for feedback after exposure to distressing content.

To address these identified challenges, we brainstormed and developed three design approaches (see Figure 1):

(1) **Leveraging Report Transparency**: A design approach to inform users about the status and outcomes of their reports, providing clear feedback on the actions taken in response to reported content.

(2) **Specifying Discomfort Elements**: A design approach to allow users to pinpoint and select specific elements within a video that they find discomforting, offering more granular control over content curation algorithms.

(3) **Pre-watch Algorithmic Feedback**: A design approach to allow users to review recommendations made by algorithms before engaging with the content, granting users preemptive control over their viewing experience.

These proposed design approaches were conceptualized into specific features and mockups to visually represent them. Detailed explanations and feedback on these were presented in Study 2.

4 STUDY 2: EVALUATION OF DESIGN APPROACHES TO MITIGATE VIEWERS' DISCOMFORT

4.1 METHOD

To evaluate our design approaches to reducing users' discomfort with distressing short-form videos, we developed them as potential features for popular short-form platforms like YouTube and visualized them through mockup images using Figma.

Among the 17 participants who initially participated in Study 1, 15 of them expressed interest and participated in Study 2 (see Table 1). The study was conducted either in-person or virtually, with each session lasting approximately 30 minutes. After obtaining participants' consent, we presented each of the three mockup images, each representing a different design approach. Participants were then invited to share their perspectives on each feature, focusing on its potential utility, limitations, and suggestions for improvement. Subsequently, they rated each feature's capability to alleviate discomfort caused by distressing short-form videos and their willingness to utilize the feature using a 5-point scale (see Table 2). After completing evaluations for all three proposed design approaches, participants were compensated with a KRW 8,000 (approximately $6 USD) gift card. Employing the same thematic analysis method used in Study 1, we examined the interview transcripts from Study 2. Also, we averaged the ratings on each feature.

4.2 Design and Evaluation of "Report Tracker"

As a way to address the need for greater **transparency in the reporting process for socially inappropriate videos**, we proposed the "**Report Tracker**" feature (see Figure 2). This feature aims to keep users informed about the status and outcomes of their reports. Through this, users can navigate to their report history page (Figure 2: 1st screen), where they can see a list of the short-form videos they have reported. By selecting a specific report from the list, users can review the current status of their report (Figure 2: 2nd screen). Furthermore, when a final decision is made regarding the reported content, users will receive a popup notification detailing the action taken by the platform, with the content guidelines applied (Figure 2: 3rd screen).
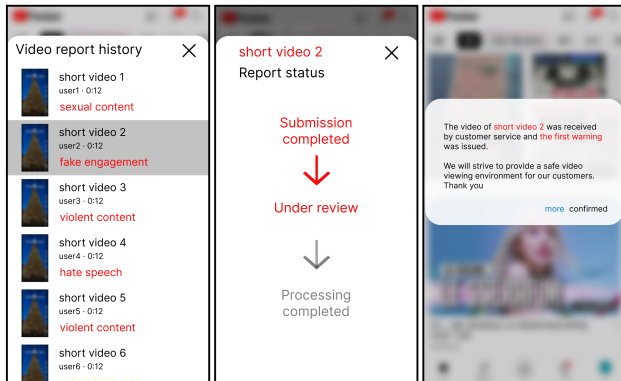
Overall, the Report Tracker was well-received by participants, who highly rated its effectiveness in addressing socially inappropriate content (average = 4.2, SD = 0.94). Their willingness to use it in the future, however, received a slightly lower but moderate rating (average = 3.8, SD = 1.01). Participants valued the feature's potential to boost both the transparency and the effectiveness of the reporting process: "*It seems like an important feature that will enhance the transparency and effectiveness of reporting*" (P11). P01 further highlighted that this design approach could also validate users' contributions to platform safety: "*The popup would make me feel that my efforts to report socially harmful content are finally appreciated*" (P01). Additionally, some found its potential to build trust in the platform: "*If I see my reports have been received and processed, my trust in the platform will certainly increase*" (P09).

However, some participants expressed concern about re-exposure to distressing feelings associated with reported short-form videos: "*When I review my report history, I would be reminded of those unpleasant feelings*" (P09).

4.3 Design and Evaluation of "Personalized Discomfort Filter"

To provide users with **granular control over the content curation algorithm**, we proposed the "**Personalized Discomfort Filter**" (see Figure 3), which allows them to identify specific elements that cause discomfort. As part of the platform's onboarding process, users are invited to identify what they find distressing, with options spanning visual, auditory, and thematic categories (Figure 3: 1st screen). This initial selection process aims to tailor the algorithm to reduce the presence of such content in the user's feed. Furthermore, when users encounter a video that they find distressing, they have the option to click the "dislike" icon, which opens a popup window (Figure 3: 2nd screen). This window lists the discomfort elements identified in the current video, leveraging advanced multimodal feature extraction techniques (e.g., [4, 18]), and allows users to add distressing elements. Users can also view and manage all discomfort elements that they have selected via a management page (Figure 3: 3rd screen), allowing them to continuously refine their preferences as their tastes or sensitivities change over time.
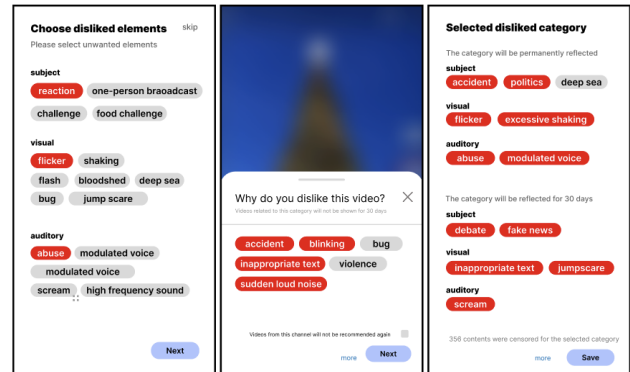
Figure 2: Mockup images of the Report Tracker

Figure 3: Personalized Discomfort Filter

Table 2: Survey Results

| Survey Question (1:Disagree to 5:agree)	Report Tracker	Personalized Discomfort Filter	Short-Form Video Playlist
(Potential Effectiveness) *This feature will mitigate issues caused by distressing content.*	Mean: 4.2 (SD: 0.94)	Mean: 4.53 (SD: 0.52)	Mean: 4.29 (SD: 0.91)
(Intention to Use) *I am generally inclined to use this feature in the future.*	Mean: 3.8 (SD: 1.01)	Mean: 4.4 (SD: 0.63)	Mean: 3.8 (SD: 1.21)

Participants responded very favorably to the Personalized Discomfort Filter feature, giving it high ratings for its potential effectiveness in alleviating personal discomfort (average = 4.53, SD = 0.52) and expressing strong interest in using it (average = 4.4, SD = 0.63). The ability of this feature to allow users to exclude specific elements was highly valued for creating a tailored and safe viewing experience: "*If I could filter out the short-form videos with insects, I definitely would use this feature*" (P08).

However, concerns were raised about the feature potentially becoming burdensome if users were prompted with a selection popup every time they clicked 'dislike.' Suggestions to mitigate this included allowing users to control the frequency of the popup or providing users with the option to activate or deactivate the feature when disliking a video. Additionally, the optimal granularity of discomfort elements was discussed among participants. Some participants expressed concern that too broad categories might diminish the diversity of content, while others were concerned that too many specific choices could make the process overly complex. Additionally, there was a divide in preferences regarding the strictness of content filtering. Some sought to completely avoid certain discomforting elements, whereas others advocated for a less restrictive filter, preferring to retain the possibility of encountering a wider array of content.

4.4 Design and Evaluation of "Short-Form Video Playlist"

To allow users to **pre-screen and modify algorithmic recommendations**, the "**Short-Form Video Playlist**" (see Figure 4) was proposed. By tapping the 'Playlist' icon on the viewing interface (Figure 4: 1st screen), users can access the current list of recommended short-form videos (Figure 4: 2nd screen). This list showcases key information for each recommended content, including their thumbnails, titles, the number of likes, and video duration, allowing users to make informed choices about what to watch. If a user identifies content they find potentially distressing or unappealing, they have the option to remove it from the playlist by swiping it away.

Participants highly rated the playlist for its potential to mitigate the harms caused by distressing content (average = 4.29, SD = 0.91) and were moderately inclined to use it in the future (average = 3.8, SD = 1.21). They highlighted the benefits of taking proactive measures to prevent exposure to distressing videos, thus providing a safer viewing experience: "*It'll be safer for people who are psychologically vulnerable since they can filter out things before being exposed to them*" (P04). Some participants pointed out that this feature would be particularly useful in certain situations where there was a sudden increase in distressing videos related to major accidents: "*If another major accident happens, I would use it to avoid*
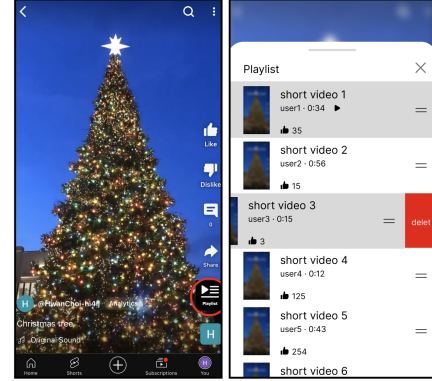
Figure 4: Short-Form Video Playlist

related short-form videos" (P01). Moreover, the playlist was valued for providing insights into the current recommendations of the algorithm, offering a clear view of the content curation process: "*The playlist makes it convenient to see my feed and check what my algorithm is recommending these days*" (P13).

However, some participants expressed concerns that this pre-screening capability might detract from the spontaneity and enjoyment derived from unexpected content discoveries: "*Knowing too much in advance about what will be showing up could reduce the enjoyment of watching short-form videos*" (P05). As such, they expressed reservations about this feature, specifically regarding its potential to alter the spontaneous nature of content discovery on short-form video platforms.

5 DISCUSSION

Our research aimed to uncover participants' perspectives and reactions to distressing short-form videos, with a particular emphasis on the nuanced and subjective nature of discomfort, which offered a unique perspective compared to the existing literature's focus on socially harmful content categorization and detection (e.g., [1, 3, 8, 9, 23]). This approach allowed us to identify specific user challenges and propose design approaches aimed at enhancing the viewing experience on short-form video platforms.

Specifically, our investigation into users' distressing experiences (Study 1) revealed that participants differentiated between "socially inappropriate" and "personally distressing content", each category eliciting distinct challenges that are currently under-addressed by the platforms. Participants expressed frustration over the opaque nature of the reporting process for socially inappropriate videos, pointing out a need for greater transparency. Additionally, when

dealing with personally distressing content, participants noted challenges in influencing the platforms' curation algorithms to avoid specific types of distressing content. They also highlighted the inherent limitations of the current feedback mechanisms, which only allowed them to react to distressing content after having been exposed to it. Overall, these findings underscore the need for a more nuanced approach to content moderation on short-form video platforms, one that considers both the societal impact of harmful content and the individual experiences of discomfort.

Further, the insights from in-depth interviews pointed us toward the development of three key design approaches aimed at addressing the key needs of transparency, control, and preventative action for distressing content. Specifically, the "Report Tracker" feature aimed to enhance procedural transparency by allowing users to track the status and outcomes of their reports on socially inappropriate content. The "Personalized Discomfort Filter" was designed to enable users to specify distressing elements they wished to avoid, offering a tailored viewing experience. Finally, the "Short-Form Video Playlist" was intended to enable users to pre-screen recommended videos, providing a feedback mechanism for avoiding distressing content proactively.

We further evaluated these design approaches and corresponding features (Study 2). The overall evaluation from participants was largely positive, with each feature being recognized for its potential to alleviate user discomfort in various ways.

Specifically, the "Report Tracker" feature was particularly praised for enhancing transparency within the platform's moderation system. By keeping users informed about the progress and outcomes of their reports, this feature was seen as a critical step towards building greater trust between users and the platform. In addition, such transparency was seen as integral to ensuring users that their concerns are taken seriously and addressed, which could encourage more responsible and engaged community participation.

Next, the "Personalized Discomfort Filter" received positive feedback for empowering users to tailor their viewing experience by allowing users to identify and avoid specific elements that caused discomfort. Also, some believed that this feature could address the need for a more individualized approach to content moderation that respected individual sensitivities and preferences.

Furthermore, the "Short-Form Video Playlist" was recognized for its proactive approach to content filtration. Participants saw that offering users the ability to screen and modify their content feed before exposure to potentially distressing videos could enhance the overall safety and comfort of the viewing experience.

As a whole, the results of this study emphasize the importance of user agency in the design of short-form video platforms, particularly through the implementation of design approaches that allow users to exercise greater control over the content they are exposed to. This focus on increasing user control over algorithmic processes directly contributes to the broader discourse in HCI research about balancing algorithmic curation with user preferences and agency [2, 13, 16, 20]. Further, the increasing societal concerns around content filtering and platform moderation practices, highlighted by debates on how these practices might limit users' freedom of expression and content access [6, 10, 12], underscore the importance of our design approaches.

However, our findings also suggest a potential trade-off between maximizing user agency and maintaining the core appeal of short-form video platforms such as spontaneous and effortless content discovery. While all of the design approaches were highly rated for their potential to mitigate discomfort, the intention to use some features was relatively lower. This discrepancy may suggest that, although viewers appreciate having control over their content exposure, they also value the serendipitous nature of content discovery inherent in short-form video platforms. Moreover, the preference for passive and effortless consumption, identified as a key motivation for engaging with short-form videos [7, 11, 17, 20], suggests the need for balancing user control with the desire for effortless content discovery.

6   LIMITATIONS & FUTURE WORK

The limitations of this study present important directions for future research. Firstly, our participant pool was primarily college students in their 20s, which restricts the diversity of user experiences and perceptions considered. Therefore, future studies will need to include a broader and more diverse demographic to ensure findings reflect a broader spectrum of experiences on short-form video platforms. Additionally, the current research primarily relies on conceptualized design approaches without empirically testing these in real-world scenarios. Therefore, the development and real-world testing of interactive prototypes could potentially validate our proposed design approaches as effective means to mitigate distressing content encounters.

ACKNOWLEDGMENTS

REFERENCES

[1] Sharifa Alghowinem. 2019. A safer youtube kids: An extra layer of content filtering using automated multimodal analysis. In *Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 1.* Springer, 294–308.

[2] Oscar Alvarado and Annika Waern. 2018. Towards algorithmic experience: Initial efforts for social media contexts. In *Proceedings of the 2018 chi conference on human factors in computing systems.* 1–12.

[3] Arnav Arora, Preslav Nakov, Momchil Hardalov, Sheikh Muhammad Sarwar, Vibha Nayak, Yoan Dinkov, Dimitrina Zlatkova, Kyle Dent, Ameya Bhatawdekar, Guillaume Bouchard, et al. 2023. Detecting harmful content on online platforms: what platforms need vs. where research efforts go. *Comput. Surveys* 56, 3 (2023), 1–17.

[4] Yuki Asano, Mandela Patrick, Christian Rupprecht, and Andrea Vedaldi. 2020. Labelling unlabelled videos from scratch with multi-modal self-supervision. *Advances in Neural Information Processing Systems* 33 (2020), 4660–4671.

[5] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[6] Thiago Dias Oliva. 2020. Content moderation technologies: Applying human rights standards to protect freedom of expression. *Human Rights Law Review* 20, 4 (2020), 607–640.

[7] Zhiwen DONG and Tian Xie. 2022. Why People Love Short-Form Videos? The Motivations for Using Tiktok and Implications for Well-Being. *The Motivations for Using Tiktok and Implications for Well-Being* (2022).

[8] Hazım Kemal Ekenel and Tomas Semela. 2013. Multimodal genre classification of TV programs and YouTube videos. *Multimedia tools and applications* 63 (2013), 547–567.

[9] Vaishali U Gongane, Mousami V Munot, and Alwin D Anuse. 2022. Detection and moderation of detrimental content on social media platforms: Current status and future directions. *Social Network Analysis and Mining* 12, 1 (2022), 129.

[10] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (2020), 2053951719897945.

[11] Li Gu and Xun Gao. 2022. What drives me to use TikTok: A latent profile analysis of users' motives. *Frontiers in psychology* 13 (2022), 992824.

[12] Andrew M Guess. 2021. (Almost) Everything in moderation: new evidence on Americans' online media diets. *American Journal of Political Science* 65, 4 (2021), 1007–1022.

[13] Kevin Hamilton, Karrie Karahalios, Christian Sandvig, and Motahhare Eslami. 2014. A path to understanding the effects of algorithm awareness. In *CHI'14 extended abstracts on human factors in computing systems*. 631–642.

[14] Richard Jackson Harris and Lindsay Cook. 2011. How content and co-viewers elicit emotional discomfort in moviegoing experiences: Where does the discomfort come from and how is it handled? *Applied Cognitive Psychology* 25, 6 (2011), 850–861.

[15] John Herrman. 2019. How TikTok is rewriting the world. *The New York Times* 10 (2019), 412586765–1586369711.

[16] Hyunjin Kang and Chen Lou. 2022. AI agency vs. human agency: understanding human–AI interactions on TikTok and their implications for user engagement. *Journal of Computer-Mediated Communication* 27, 5 (2022), zmac014.

[17] M Laeeq Khan. 2017. Social media engagement: What motivates user participation and consumption on YouTube? *Computers in human behavior* 66 (2017), 236–247.

[18] Zekang Li, Zongjia Li, Jinchao Zhang, Yang Feng, and Jie Zhou. 2021. Bridging text and video: A universal multimodal transformer for audio-visual scene-aware dialog. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), 2476–2483.

[19] Kang Liu, Feng Xue, Dan Guo, Le Wu, Shujie Li, and Richang Hong. 2023. Megcf: Multimodal entity graph collaborative filtering for personalized recommendation. *ACM Transactions on Information Systems* 41, 2 (2023), 1–27.

[20] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the design of youtube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[21] Kimeko McCoy. 2024. *How Nationwide is navigating the short-form video boom*. Retrieved January 10, 2024 from https://digiday.com/marketing/how-nationwide-is-navigating-the-short-form-video-boom/

[22] James Pierce, Sarah Fox, Nick Merrill, and Richmond Wong. 2018. Differential vulnerabilities and a diversity of tactics: What toolkits teach us about cybersecurity. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[23] N Rea, G Lacey, R Dahyot, and C Lambe. 2006. Multimodal periodicity analysis for illicit content detection in videos. (2006).

[24] Adelais Reichmann, Ines Bauda, Bettina Pfeffer, Andreas Goreis, Mercedes Bock, Paul Plener, Oswald D Kothgassner, et al. 2023. Post-Traumatic Stress after Corona Virus Disease 19 (COVID-19): The Role of Gender and Distressing Social Media Exposure as Risk Factors. *Digital Psychology* 4, 1 (2023), 14–26.

[25] Gautam Kishore Shahi and William Kana Tsoplefack. 2022. Mitigating harmful content on social media using an interactive user interface. In *International Conference on Social Informatics*. Springer, 490–505.

[26] Spandana Singh. 2019. Everything in moderation: An analysis of how Internet platforms are using artificial intelligence to moderate user-generated content. *New America* 22 (2019), 1–42.

[27] Maarten W Van Someren, Yvonne F Barnard, Jacobijn AC Sandberg, et al. 1994. The think aloud method: a practical approach to modelling cognitive processes. *London: AcademicPress* 11 (1994), 29–41.