

Unraveling Entangled Feeds: Rethinking Social Media Design to Enhance User Well-being

Ashlee Milton

milto064@umn.edu

University of Minnesota - GroupLens
Minneapolis, Minnesota, USA

Dan Runningen

runni028@umn.edu

University of Minnesota - GroupLens
Minneapolis, Minnesota, USA

Loren Terveen

terveen@umn.edu

University of Minnesota - GroupLens
Minneapolis, Minnesota, USA

Harmanpreet Kaur

harmank@umn.edu

University of Minnesota - GroupLens
Minneapolis, Minnesota, USA

Stevie Chancellor

steviec@umn.edu

University of Minnesota - GroupLens
Minneapolis, Minnesota, USA

Abstract

Social media platforms have rapidly adopted algorithmic curation with little consideration for the potential harm to users' mental well-being. We present findings from design workshops with 21 participants diagnosed with mental illness about their interactions with social media platforms. We find that users develop cause-and-effect explanations, or folk theories, to understand their experiences with algorithmic curation. These folk theories highlight a breakdown in algorithmic design that we explain using the framework of *entanglement*, a phenomenon where there is a disconnect between users' actions and platform outcomes on an emotional level. Participants' designs to address entanglement and mitigate harms centered on contextualizing their engagement and restoring explicit user control on social media. The conceptualization of entanglement and the resulting design recommendations have implications for social computing and recommender systems research, particularly in evaluating and designing social media platforms that support users' mental well-being.

CCS Concepts

• **Human-centered computing** → **Social media**; *Participatory design*; *HCI theory, concepts and models*.

Keywords

Social Media, Recommender Systems, Algorithmic Curation, Mental Well-being, Mental Health

1 Introduction

Algorithmic curation has transformed social media platforms, like Instagram and TikTok, to deliver personalized content to billions of users [91]. Powered by recommender systems, algorithmic curation provides users with content from anywhere on the platform [81] and serves as the primary distribution mechanism for much of the content on current social media platforms. Prior work demonstrates that algorithmic curation has positive effects on users, including information discovery [61], identity and self-conceptualization [47, 55], and social connection [23, 59]. Furthermore, platforms can directly improve mental well-being for users with mental illness

by providing coping strategies and validation [7, 92], as well as information, community, and support [25, 61, 75].

However, concerns are growing that social media platforms and algorithms can harm people and their mental health [41, 46, 62]. TikTok and Instagram have received widespread public attention for their potential impact on teens and people with mental illness [29, 73, 93]. Prior academic work has also found that social media platforms propagate harmful information about mental health, including dangerous advice about eating disorders [13, 14], propagating challenges that promote self-harm [60], and even triggering or exacerbating dangerous behaviors [30, 36]. Disengaging from social media is not viable for many users with mental illness, as social media is their best source of mental health support in the face of societal stigma and limited resources [2, 62].

A growing body of work suggests that social media design features, particularly algorithmic content feeds, may drive harmful experiences for users with mental illnesses. Moreover, several recent HCI studies show that users identified the lack of control over algorithmically curated content as a primary source of harm to their well-being [59, 61, 82]. Participants used metaphors like “dopamine slot machine,” [61] “runaway train,” [59] and “domesticating” the feed [82] to describe their fraught relationship with algorithmic curation. Recent work has also identified that interface design decisions in social media may lead to cognitive issues, such as attention problems and decreased well-being [58]. As users with mental illness disproportionately experience harms from current social media curation systems, centering their experiences can reveal mechanisms of harm that affect all users and identify platform design opportunities that benefit everyone [21].

In this work, we focus on how users with mental illness interpret their interactions with social media platforms and algorithmic curation, and how system design affects their mental well-being. To achieve this, we conducted seven design workshops focused on two tasks designed to explore our interest in social media platforms: first, an activity to discover their current experiences with platforms, and second, a prototyping activity to design their ideal platform to support their mental well-being. These workshops were conducted with 21 participants, all of whom used social media platforms with algorithmic curation and identified as having anxiety, depression, or PTSD. We draw inspiration from design justice [20] and designing from the margins [31] to inform our approach. Designing with people with mental illness is vital, as conventional design processes



often fail to account for their distinct needs [21]. Design justice emphasizes how technological systems can perpetuate systemic inequities and underscores the importance of inclusive design processes that fairly distribute benefits and harms, ultimately leading to better designs for everyone [20, 21].

Our results highlight the cause-and-effect explanations, or folk theories, that our participants created to explain their experiences interacting with algorithmic curation. Combining these folk theories from our participants with users' experiences from prior work, we document a pervasive breakdown in the design of social media algorithmic curation that we conceptualize as *entanglement*. We propose *entanglement* as a framework in which participants cannot make clear connections between the actions they take when interacting with social media, the changes they observe in their curated feeds, and affective outcomes on their well-being. Entanglement extends Norman's action cycle [66] by introducing six types of errors: three related to the gulfs of execution ("Hiding", "Guessing", "Flattening") and three to the gulf of evaluation ("Dangling", "Overloading", "Disempowering"). Using the lens of entanglement, we identified novel interaction pathways that contributed to many of these breakdowns and led to negative affective consequences for our users. Our participants actively tried to combat entanglement through their design prototypes with three strategies: 1) contextualizing engagement to preserve users' emotional intentions, 2) consumption control for healthier interactions, and 3) reclaiming control through explicit user input.

We introduce a novel framework, entanglement, for explaining the problems users with mental illness encounter when interacting with current social media platforms and resulting harms to mental well-being. Entanglement extends Norman's action cycle through the idea of "joined worlds", which explains how merging recommender systems and social networking systems in social media platforms creates conflicts in users' mental models and interaction experiences. Finally, we propose design implications as starting points for combating entanglement through: 1) creating avenues for users to exert control through communicating their intentions and contextual information relevant to their interactions, and 2) shifting algorithmic optimization from accuracy to focus on contextual and emotional aspects of content and users.

2 Related Work

2.1 Social Media and Mental Well-being

Research consistently shows the importance of social media for people with mental illness: it provides a source of accessible information [59, 76] and community [26, 64, 83] for many. Social media makes it easy to connect with others about mental health through communities [25], hashtags [14, 35], and social media feeds [59]. Previous work has found that people with mental illness will look for social support from others with similar diagnoses [7, 63].

However, this same research recognizes a trade-off in the risks versus benefits of social media for people with mental illness [46, 62]. Social media can propagate unreliable information and malicious misinformation [7, 53], as well as dangerous content that encourages harmful mental health behaviors and outcomes [14]. More recently, people have begun to use social media for self-diagnosis, creating mixed consequences to users and healthcare providers

through pathologizing of normal behavioral and normalizing mental health disorders to combat stigma [19, 42]. Given the potential influence of social media on people's mental well-being, platforms have created interventions and policies to try to mitigate potential harms, such as screen timers, content moderation of dangerous content, and nudges to get off the platform. However, these policies can harm users by restricting access to content and communities that provide positive social support [36, 37] and silence users from marginalized identities speaking on their lived experiences [57].

Past research has established that social media is essential to users with mental illness, yet can also have complicated effects on users' mental well-being. In this work, we use the context of social media and mental health as a case study of algorithmic curation to examine the consequences to users and ways to design these systems to support them.

2.2 Algorithmic Curation in Social Media

Next, we provide an overview of the research on the effects of recommender systems/algorithmic curation on social media and users. Recommender systems and their impacts on users are a popular area of research [49, 50]. In recent years, recommender systems have been used on social media feeds to provide content matching users' tastes [8], which is often called algorithmic curation [40]. Algorithmic curation of social media can improve connectivity [8] and the delight of discovering new content [59]. However, other research finds that algorithmic curation polarization and filter bubbles [8, 12, 16], as well as echo chambers and rabbit holes [10, 54, 71].

The study of algorithmic curation in social media has evolved from initial explorations of user perceptions to more nuanced examinations of user experiences as algorithmic curation has become more pervasive. Early work in this area found that users were not always aware that algorithmic curation was occurring [33]. They reasoned about algorithmic curation in nuanced ways [32] and developed elaborate and nuanced folk theories of algorithms [28, 32]. Follow-up work highlighted a critical gap between user perceptions and the actual functioning of social media feeds driven by algorithmic curation [11, 38, 88]. More recently, research on algorithmic curation has examined how people engage with algorithms, finding that users struggled with sense-making and domestication of the algorithm that caused issues with their digital selves [82, 88]. The effects of algorithms on people were bidirectional, with studies showing both increases to mental health symptoms and coping mechanisms and social support from social media use [39, 48, 59, 61, 72].

2.3 Effects of Algorithmic Curation on Mental Well-being

Finally, we address prior work at the intersection of our interests: the impacts of algorithmic curation on mental well-being. This is in light of increased media attention on current social media and mental health [29, 73, 78, 93], and the lay hypothesis that feeds harm well-being [41].

Studies have demonstrated that algorithmic curation can be supportive of mental health. These benefits are wide-ranging, from supporting individuals with information-seeking [59, 61] to delivering help from professionals [5, 61, 76] to providing therapeutic value [5, 45]. Moreover, studies suggest that algorithmic curation

can help in mental health discourse [35, 57] and exposure to people’s lived experiences of mental illness [35, 59, 61]. This work indicates that algorithmic curation can serve as a tool for literacy and education, social support, and good advice for mental health.

However, much of this same work also points to potential risks of algorithmic curation for mental health. This concern is underscored by ongoing public dialogue about the consequences of social media on mental health, especially in adolescents [41, 64, 89]. Studies show that social media can expose users to harmful content, including dangerous advice about eating disorders [14] and methods for suicide [60]. Problems with algorithmic curation and frustrations about control can lead to strong reactions by users [36, 85] and the perpetuation of social inequalities [88]. A notable concern is that users report that algorithmic curation can trigger or exacerbate ill-being [59, 61] or risky behaviors [30, 36]. Close to our work, Milton et al. [61] identifies frustrating experiences with user interfaces and algorithmic curation, where their participants likened feeds to “dopamine slot machines”. In this case, participants blamed specific interface characteristics for their ill-being and believed they had no control over the feed.

In sum, prior work emphasizes through folk theories [28, 32] and user experiences [59, 61, 82] that algorithmic feeds may be to blame for people’s distressing experiences. We build on the call from work close to ours [59] by incorporating and consulting people with mental illness on the design of such systems [6, 51, 77]. Taking such a standpoint is crucial because individuals with mental illness may disproportionately be affected by the harms of algorithmic curation [4, 36, 37].

3 Methods

We conducted seven design workshops with 21 participants who identified as having a mental illness and used social media. During these workshops, participants completed two tasks: one aimed at examining their experiences with algorithmic curation and its impact on their well-being, and a design task where they imagined ways to improve social media platforms to support their mental well-being. The study was approved by the Institutional Review Board at the University of Minnesota (STUDY00019753). Here, we outline information about our participants, workshops, analysis methods, and positionality statement.

3.1 Participants

We posted recruiting messages on social platforms that used algorithmic curation, primarily X/Twitter, Instagram, and Reddit. We also sent recruitment emails via mailing lists at the author’s institution and contacted past research participants who consented to be invited to future research studies. Recruiting messages included a link to a Qualtrics form that verified people’s eligibility, recorded their consent to participate, and collected demographic and social media usage information. Respondents were eligible for participation if they were 18 years or older, had been diagnosed (or self-diagnosed) with depression, anxiety, or PTSD, and engaged with social media platforms that provided recommended content (e.g., TikTok, Instagram, YouTube). Following design justice principles, we centered people with mental illness in our study, as those

Demographic Variables	N	Demographic Variables	N
Age		Ethnicity	
18-20	5	African American	2
21-25	8	Arab American	1
26-35	6	Asian	1
36-45	1	Indian	1
46-55	1	Latino	1
Gender		Middle Eastern	1
Female	13	Turkish	1
Male	5	White	17
Non-binary	1	Education	
Queer	1	Bachelor’s degree	8
Transfemme	1	Doctorate degree	2
Sexuality		Master’s degree	3
Asexual	1	Some college, no degree	8
Bisexual	7	Income (USD)	
Heterosexual	5	1 - 9,999	10
Hetero-flexible	1	10,000 - 24,999	4
Lesbian	2	25,000 - 49,999	4
Pansexual	1	50,000 - 74,999	1
Queer	3	75,000 - 99,999	1
Questioning	1	150,000 and greater	1

Table 1: Aggregated Demographic Information of Participants. Participants self-identified on intake; some minor standardization was applied (e.g., “women” and “females” collapsed). Demographic counts are not mutually exclusive and may not sum to the total sample.

most affected by algorithmic curation’s impact on mental well-being, providing them with essential knowledge for redesigning these systems [20, 43]. We chose depression, anxiety, and PTSD as our focus conditions based on their high prevalence in the United States [68], enabling our findings to reflect the experiences of a significant portion of social media users with mental illness. The survey also included a modified University of California, San Diego Brief Assessment of Capacity to Consent (UBACC) [44] to assess participants’ ability to consent to our study, per our IRB’s requirements for including people with mental illness.

The first two authors conducted workshops from March to July 2024 until theoretical saturation was reached [15]. 21 respondents met our eligibility requirements, scored above 15 on the UBACC (which demonstrated capacity to consent), consented to our study, and ultimately signed up for a workshop. Participants received their choice of a \$40 gift card to Target or Amazon for their participation.

Participants were aged 18 to 55 ($M=25.5$), and were primarily white ($N=17$) and female ($N=13$). Our participants also, on average, held a post-secondary degree. Table 1 overviews the participants’ demographics. We did not collect information about participants’ diagnoses of depression, anxiety, or PTSD for eligibility. Instead, we requested self-identification of these conditions, rather than formal assessment, due to well-known socioeconomic and stigma barriers to accessing diagnostic mental health care [18] and to align with the intentions of standpoint theory [6]. Participants reported on which platforms they used, with Instagram, YouTube, and TikTok each mentioned by over half of the pool. All these platforms feature videos as a significant form of content. Most participants reported

using one of these platforms at least daily, with only two indicating a frequency of weekly or less.

3.2 Workshops

We used design workshop methodologies [70] to capture collective perspectives and foster intervention in designs through participants' shared experiences. Given the sensitive nature of mental well-being, we carefully designed our workshops to be supportive environments. Prior research shows that workshops with people with mental illness can increase social inclusion and positively impact well-being [79, 84] and can be productive for creative brainstorming and generative engagement [90]. To mitigate potential risks and make participants feel more comfortable, we kept workshop sizes small (2 to 4 participants) and only took notes on participants' discussions in paraphrases and captured pictures of the resulting artifacts. Seven workshops were held: four in person and three remotely via Zoom to improve recruitment.

Participants who met the eligibility requirements were invited to one design workshop based on their availability. Each workshop took about two hours. Workshops began with introductions, setting ground rules, and icebreakers to help participants feel more at ease with the researchers and with one another. Most of the workshop was dedicated to the two study activities: one focused on exploration and discovery, with the other on prototyping [86]. Researchers provided prompts for each activity and asked follow-up questions to facilitate discussion. Throughout the workshop, participants were reminded to reflect on their mental well-being experiences while using their preferred algorithmically curated social media platforms.

3.2.1 Activity 1: Exploration and Discovery. For the first activity, participants were asked to discuss their interactions with algorithmically curated social media platforms and to identify the impacts on their mental health and well-being. This activity was modeled after the critical thinking exercise referred to as "always, sometimes, never". We provided a shared board for participants to visually orient their experiences along two axes: how often interactions occurred (from "always" to "never") and how participants felt about the effect those experiences had on their mental well-being (from "like" to "dislike"). Figure 1a shows an example of this activity. We specifically prompted participants to "think about the interactions that you have had (or wish you had) with your preferred algorithmically curated social media feed," then asked them to write down the interactions and place them on the board. Participants were encouraged to discuss their experiences with one another if they felt comfortable, enabling them to identify common trends and insights. This task allowed participants to explore and recall their interactions with platforms.

During the activity, the facilitating researchers would ask questions to clarify understanding and connection to algorithmic curation when necessary to keep the discussion focused. Researchers further asked follow-up questions at the end of the activity about the interactions provided and where they were placed on the board to promote discussion and gain further insights into participants' experiences and their effects on mental well-being.

3.2.2 Activity 2: Prototyping. In the second activity, participants were asked to design an ideal algorithmically-curated social media platform, focusing on how they envisioned the system supporting their mental well-being. The prompt given to participants stated, "Building on the experiences you talked about in the last activity, design your ideal algorithmically curated social media platform, focusing on how you want it to support your mental well-being." The goal was to have participants generate a prototype based on the individual and collective experiences they had identified in the previous task, as well as their ideas for improving the platforms. Toward the end of the sessions, participants shared their prototypes with the group and discussed in a design show-and-tell [3]. Throughout the activity, the researchers would ask questions, when relevant, about participants' discussions to understand their motivations for design ideas, connections to interactions from the previous activity, and how they connected to algorithmic curation and promoting mental well-being. Given our focus on design based on participants' needs, researchers refrained from providing any input on the designs. Researchers did ask about design aspects and connections to existing platform features at the end of the activity.

In-person sessions provided office and crafting supplies (see Figure 1b) to facilitate the creation of participant prototypes. Remote session participants had a dedicated space on a Google Jamboard slide where they could draw, type, and paste content to help them convey their designs. All participants were encouraged to design in a way that they felt most capable of communicating their ideas (e.g., drawing, writing). Participants were also free to continue discussing their experiences while they worked, and many did.

3.3 Thematic Analysis

We used inductive thematic analysis for our qualitative results [17, 65]. Themes emerged from analyzing the researchers' notes and the artifacts collected from the two tasks. Between workshop sessions, researchers conducted initial open coding and code collating. The first two authors conducted a secondary analysis phase after all workshops to identify underlying conceptual threads connecting emerging themes. During this secondary analysis, the authors noted a prevalent trend in users' interactions with social media, which led to the development of the entanglement framework (see Section 4.2) to reinterpret the emergent theme. Additional discussions between all authors confirmed the resulting findings. All quotes are either from field notes or the artifacts produced by participants.

3.4 Positionality Statement

Most authors of this study identify as active or peripheral members of the mental health research community [1] and are active participants on social media platforms. We believe that including community members in the research team leads to more thoughtful research practices and more open dialogue among participants in at-risk populations. However, including member researchers does not inherently protect us from introducing bias.

The drive to create thoughtful research practices led us to adopt a design justice approach [6, 20] in our work by centering on participants' needs (e.g., small workshops, acknowledging self-diagnosis), which we believe promoted better inclusion for people with mental illness. All authors were involved in the study's design, refinement,

4.1.2 Theory of Algorithmic Leakage. Participants saw their algorithmic feeds as unintentional disclosure agents, “leaking” personal information through recommended content. In this theory, feeds function as mirrors [23, 55], reflecting users’ identities through content. As feed content may be visible in shared or unsafe places, participants felt their private information, especially related to mental well-being and emotions, was exposed, prompting hypervigilance.

Passive mental health disclosures were a primary concern, showcasing ongoing stigma and fear of judgment around mental well-being. Participants noted that recommendations reveal prior interactions, with P04 saying, “everyone knows that you get recommendations based on what you have already seen”, and P17 adding “I feel like I will be judged for having mental health content on my feed.” While prior work has shown that feeds reflect users’ identities [23, 55] and can even influence their self-concept [22, 47], our participants emphasized the risks of passive mental health disclosure in shared spaces. Concerns about unintentional disclosure increased anxiety and reinforced the sense that feeds exposed more than users intended.

Participants’ concerns extended beyond feeds, as they perceived that platforms share data with other platforms, increasing disclosure risks. Participants gave examples of “searching for some [mental health] symptoms, and it seems to have crossed over into my feed” (P09) or “cookies on one site can travel to other sites and share the information” (P04), creating a need for them to be cognizant of how their interactions off-platform bleed into their algorithmic feeds. Together, these beliefs show that even when users are hypervigilant with their feeds, their interactions, particularly with mental health content, may be used in unanticipated ways, increasing the potential for harmful disclosure across platforms.

4.1.3 Theories of Algorithmic Exploitation. Our participants described two folk theories about how algorithmic feeds exploit their interactions. They believed these malicious intentions were deliberately done to affect their well-being. Both theories illustrate how participants attempted to make sense of their feed behaviors, which left them feeling distrustful and frustrated.

Theory of Corporate Intentions. Participants believed companies made deliberate choices that affected their mental well-being, evidenced by engagement-driven algorithmic design. Participants felt that “[companies] know their algorithm is addictive” (P10). They reasoned that corporations implemented algorithms to exploit people’s interactions and engagement to make money by keeping them on the platform (P05). Our participants attributed strong intentionality to an obscured process of algorithmic curation to understand their experiences, similar to prior issues with TikTok “techlash” [82] and amplifies the “Platform Directed” operationalization seen in folk theories about Twitter’s algorithm [28].

The belief that the algorithm was intentionally pushing engagement influenced how participants viewed platform-led mental well-being interventions (P10), like “stop scrolling” messages or “notifications to take a break.” As participants believed companies acted with exploitative intent, they viewed these interventions as “very patronizing because it was from the platform” (P19), provoking anger rather than self-reflection. However, when similar messages

came from creators, participants found them helpful (P18). Participants’ intuitions that companies’ intentions were exploitative made them distrust platforms’ efforts to support users’ mental well-being.

Theory of Engagement Abuse. While the previous folk theory focused on the *why*, participants also created beliefs about *how* the algorithm accomplished corporate goals. Participants theorized that the underlying algorithm used every interaction to promote engagement at the expense of mental well-being.

This belief came from participants’ experiences of “any engagement causes the recommendations to shift” (P14). Participants felt feeds purposely “prioritize attention and engagement over users’ well-being” (P04), reinforced by the sense that “[companies are] filling your feed with stuff *they* want you to see, not stuff *you* want to see” (P19). Participants’ belief that platforms deliberately chose outcomes led to increased feelings of distrust, as these engagement-driven outcomes were perceived as manipulative. This distrust, combined with participants’ coping with uncertainty in their interactions, contributed to participants’ overall experiences of anxiety and frustration when interacting with the platforms.

4.2 The Entangled Action Cycle

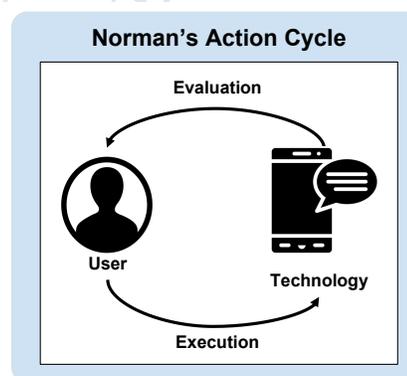


Figure 2: Visual representation of Norman’s Action Cycle

From participants’ folk theories, we identified patterns in their interactions with social media and the outcomes they saw in their feeds. These patterns reveal a disconnect between participants’ perceptions of their actions and changes in their feeds. This disconnect is clearly visible through users with mental illness experiences, who are more attuned to emotional changes and interact with greater emotional intentionality. To understand this disconnect, we developed the framework of *entanglement*, which describes participants’ inability to make clear connections between the actions they take on a platform and the resulting system responses they observe. This framework was developed during a secondary analysis phase to reinterpret emergent themes in participants’ experiences arising from interactions with social media algorithmic curation. In this section, we describe our framework of entanglement and the emotional outcomes it had on participants.

Our framework builds on Norman’s action cycle [66], popularized in the *The Design of Everyday Things*. The action cycle is a seminal psychological model that describes how people interact

with objects in the world. Norman posits that people “continually act upon and monitor the world, evaluating its state...[and] executing actions” [66]. This approach aligns with our participants’ folk theories that describe a cause-and-effect model of their interactions. By situating participants’ accounts of their interactions within the action cycle, we can more clearly examine their experiences to identify where these disconnects, or entanglements, occur.

In HCI, this action cycle has four components: user, technology/artifact, execution, and evaluation [66]. As illustrated in Figure 2, the cycle starts with a user having a goal they wish to accomplish, which they execute through iterations with the relevant technology, and then they evaluate that technology’s response to determine if they have achieved their goal. For Norman, good design makes this cycle clear and easy. In contrast, poorly designed objects, such as “Norman doors” [66], lack clear indications of whether they should be pushed or pulled. Thus, bad design causes users problems when using the underlying object to achieve their goals. Problems that emerge in this cycle are classically described as *gulfs of execution* and *gulfs of evaluation* (or less commonly, slips [67]).

Currently, gulfs in the action cycle neglect a critical dimension of participants’ experiences: emotions. These gulfs take a mechanical view of users’ interactions or possibly contextualize this as user error [67]. However, our participants noted the affective/emotional aspects to their intentions and experiences with algorithmically curated feeds, which were caused by the interface. Entanglement addresses this gap by centering on emotions in user actions and system responses, characterized through six issues that encompass emotional aspects in both the execution and evaluation phases of the action cycle. Thus, entanglement extends the current action cycle to include important emotional aspects relevant to users when interacting with technological systems.

In the rest of the section, we define and show examples of these six categories:

- Issues relating to execution: **Hiding, Guessing, and Flattening.**
- Issues relating to evaluation: **Dangling, Overloading, and Disempowering.**

4.2.1 Entanglement Issues in the Gulf of Execution. We first consider issues on the execution side of the action cycle, as shown in Figure 3. The issues we outline in this section relate to Norman’s *gulf of execution*, the difference between a user’s goals and what a system allows them to accomplish [66]. They build on this traditional understanding to capture our participants’ experiences and emotional responses.

Hiding occurs when users cannot find a feature to accomplish their goal because the feature is missing or unavailable. Hiding is close to a classic gulf of execution, arising when a user has a single goal but perceives that there are no features to accomplish it. The feature may be hidden, non-existent, or unclear in its functionality. This issue occurred when participants wanted to “reset” their feeds after getting “stuck in a rabbit hole, and then my feed is trashed” (P12), leading to uncomfortable content with no way to correct it. Participants would look for ways to reset, but no options were available. One participant noted their feeds “somehow went from scientific and gardening videos to bare-feet anti-vaccine content” (P12). The inability to “reset” a feed is particularly an issue for

mental well-being and emotions, as participants reported feeling distressed by the stream of content they were viewing. Without recourse to correct their feed, participants experienced intensified emotional impacts. An example of such content was that which romanticizes disabilities and “belittles experiences” which they felt “creates insecurity” (P14) for them.

Another goal participants could not accomplish was removing content from their feeds. Some content had adverse effects on participants, such as the content mentioned earlier that romanticizes disabilities or content that “trivializes/glamorizes inpatient treatment” (P17). Participants wanted to remove such content, not only for themselves but also for others, to mitigate the harm it could cause to people’s mental well-being. However, many did not see a way to accomplish this, and participants could not remove it from their feeds. Being unable to remove content left our participants frustrated and continually exposed to content they did not wish to see, which could make them feel uncomfortable or relive distressing experiences. Hiding also aligns with the *theory of corporate intentions*, as hiding or failing to provide mechanisms for users to take actions to influence their feeds can be seen as a deliberate act of forcing engagement through the inability to act at the expense of users’ mental well-being.

Guessing happens when a user faces multiple, unclear features for accomplishing a goal, forcing them to experiment to determine which feature will achieve their goal. The most common Guessing scenario related to algorithmic curation was when participants did “not want to see stuff [specific types of content]” (P19). Participants identified several features they guessed would help, including “not interested”, or “do not recommend”. However, no social media platforms clearly indicated which feature would actually permanently remove content. P01 noted it was “hard to convey that you’re not interested in something”, while P02 felt they had to “jump through hoops to stop seeing some content”. Guessing is an entanglement focused on the need to engage with multiple features that drive engagement, which aligns with the *theory of engagement abuse* when a purposively ambiguous design forces multiple interactions.

Guessing creates confusion in users’ minds, leading them to continue using features with uncertainty to regain control of their feeds. In cases of conflicting features, some participants, such as P12, began blocking creators to avoid viewing specific types of content. Blocking, however, is a creator-focused feature, not a content one, and would exclude an entire creator’s content from their timeline. This created tension: participants wanted to see some of a creator’s content, but not all of it, and, since they were using blocking as a workaround, they got stuck. One example was a creator whose content spanned “from ghost video to horror... [it has] similar undertones but the content is extremely different” (P15). P15 wanted more fine-grained control but couldn’t find it. The fact that participants went so far as to use a creator-focused feature like “blocking” for a content-focused goal shows how much guessing they had to do to accomplish their goal. Participants noted increased mental burden and emotional exhaustion while trying to remove content already harming their mental well-being.

Flattening is when a single feature accomplishes multiple distinct goals, erasing the nuance of users’ contextual needs and what the feature actually claims to do. This issue was prominent in participants’ discussions of “clicks”, “comments”, and “watch time” for

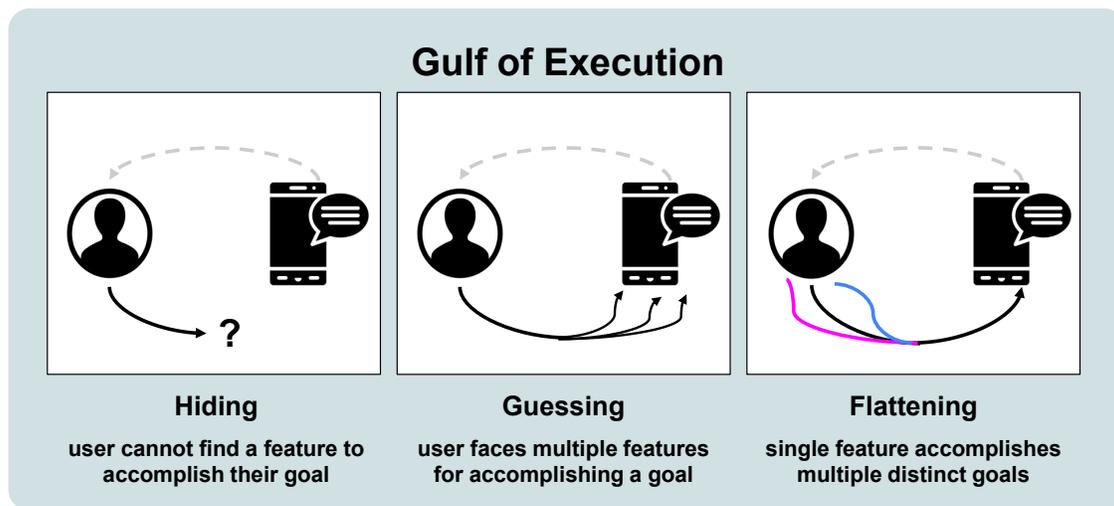


Figure 3: Visual representation of the entanglements in the Gulf of Execution

short-form videos on platforms like TikTok and Instagram. Participants perceived that the platform inferred these features to mean more than users intended, leading their feeds to become out of control. Our participants confirmed that “there is no distinction between clicking on something because you like it or hate it. The platforms don’t care either way” (P06). Comments were another area where Flattening came up. Comments accomplish different goals, such as showing support or interest, correcting content when erroneous information is present in a post, or arguing with others. All of these contextual uses of comments are collapsed into a single category: simple positive engagement. Participants noted having to weigh the pros and cons of interacting using such features, particularly in cases where they wished to stop the propagation of stigmatizing mental health content but also did not want their feeds to become inundated with such content, as the sheer volume could cause distress. The *theory of emotional mismatch* is directly connected to flattening, as it reveals the removal of emotional dimensions from actions before algorithm interpretation.

Participants reported that the flattening of the “watch time” feature was particularly detrimental to their mental well-being. We note that watch time or basic engagement is a *passive* feature on platforms, where users have little control to “stop” watching content. Thus, “watch time” suffered from the same issue as commenting, as participants noted that platforms “can’t tell the difference between watching something and being satisfied or enraged by it” (P01). Participants described getting “sucked into watching the content” and either “reveling in self-hatred” (P01) or finding “the dopamine from watching short videos can keep you stuck” (P10). Content such as “Rage bait” exemplifies the lack of nuance in watch time and showcases its effects on users as it “drives engagement instead of with a positive end goal in mind” (P13). Participants’ inability to look away from the content was taken as a measure of their engagement, regardless of whether they liked it or how it affected their mental well-being. These examples show that participants felt

a single watch-time metric cannot distinguish between genuine engagement and compulsive viewing, which at times led them to internalize the emotions presented to them. Similarly, the *theory of algorithmic leakage* and the *theory of engagement abuse* are relevant to flattening, as the collapsing of goals to drive engagement can also unintentionally lead to unwanted disclosure.

4.2.2 Entanglement Issues in the Gulf of Evaluation. Next, we move to examining entanglement related to the gulf of evaluation of the action cycle, as seen in Figure 4. Recall that gulfs of evaluation are the degree to which the system’s state can be interpreted in terms of a user’s goals [66]. We build on the concept of the *gulf of evaluation* in outlining nuanced issues in current social media.

Dangling occurs when users cannot perceive any changes in the system from their interactions with features. Almost every participant said that features for showing disinterest in content did not work. Some were “confused about whether clicking on dislike would stop the recommendations” (P14). Participants described confusion stemming from a perceived disconnect between their emotional intentions, the system’s interpretation of their actions, and the subsequent responses. While participants could sometimes find a feature to alter their feed content, the feature seemed to have no effect. Participants initially felt hopeful after discovering these features, but became frustrated and worried when they did not notice an effect. The *theory of emotional mismatch* and the *theory of corporate intentions* support this entanglement, with engagement features designed in ways that are ambiguous and lack emotional consideration, driving emotional misalignment with feed content.

Dangling is exemplified when contrasting “disliking” and “liking”. P18 noted the difference between how “seriously” feeds responded to their positive engagement (“likes”) versus negative engagement (“dislikes”), if it was available at all. Similarly, participants found “blocking” users and channels unhelpful in removing unwanted content. P12 had to “block them like seven times” yet “it still shows me [that content]”. Some felt that the “algorithm is testing” them,

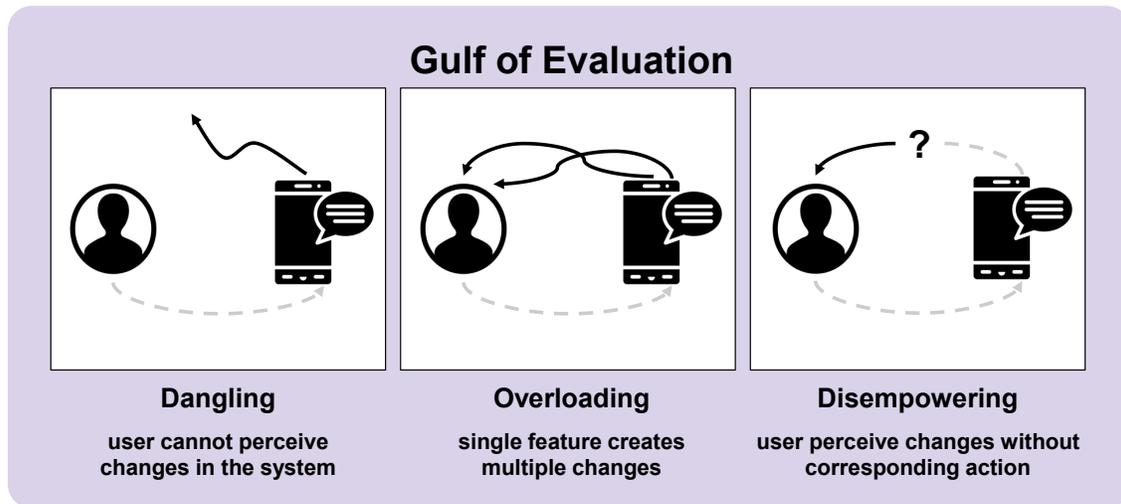


Figure 4: Visual representation of the entanglements in the Gulf of Evaluation

as content “will come back” (P18). Knowing that content could reappear, even if participants could successfully remove it, was noted to increase feelings of anxiety. Participant P07 described having to “filter your content for yourself, even though the algorithm is supposed to send you content that relates to you”. However, it was unclear “how an action you’re about to take would impact your experience” (P04), as participants could not interpret any changes in their feeds.

Overloading happens when a single feature creates multiple and often conflicting changes in the system that confuse users about the feature’s outcome. The most prominent examples of overloading were “sharing” and “liking” content, which both serve multiple purposes for users. “Sharing” allows users to send content to others and build social connections, but participants believe it is also used for positive engagement, affecting both the sender’s and recipient’s feeds. Participants echoed P17’s statement of wishing they “could share posts but also not send other personal metadata with it” as “people seem to get content suggested to them when I share things ... related to topics I engage with” (P17). The example of sharing illustrates how participants connected their experiences of emotional distress to the unpredictability of overloaded system responses. Participants felt anxious about what would be revealed about them and their mental health status through the metadata sent with the content via sharing. We can observe the *theory of engagement abuse* in overloading, where platforms leverage a single interaction in multiple ways to drive engagement. Interactions that participants did not perceive as connected to engagement clearly support this theory’s connections to overloading.

“Liking” is also used for both social connections and algorithmic engagement. Participants noticed that “if I like one psychology quote, all of my timeline fills up with inspirational quotes” (P18), meaning that “liking” also results in more similar content in feeds, regardless of the intention of the user’s action. At a creator level, “the algorithm assumes that if you like one video from a creator,

you’ll like all their others, regardless of differences in exact content” (P01). Participants struggled to understand changes to their feeds, as they perceived a mismatch between their intentions and the multiple ways the system used their interactions. Overloading has clear connections to both the *theory of emotional mismatch* and the *theory of algorithmic leakage* through the compression and exchange of emotionally laden information without regard for the emotional or social implications it creates.

Disempowering is when users perceive system changes happening without any corresponding features that give them control over initiating or stopping such changes. Participants experienced this issue through features such as infinite scroll and autoplay, which led to compulsive content viewing. When these features were combined to create a continuous content stream, users felt compelled to engage, regardless of their intentions or the harm it caused to their well-being. Yet the “watch time” registered as positive engagement due to “any engagement is positive and will get you more content” (P20). Participants described an emotionally distressing cycle in which they had no agency to escape content that harmed their mental well-being, leaving them frustrated, anxious, and disempowered.

Participants also found that their feeds would drastically change with “all the progress [personalization] gets reset” (P10) and showing generic content. Participants did not understand why these changes happened or what caused them. They were left with the heavy mental burden of curating their feeds again and possibly being exposed to content that evoked feelings of distress, rage, or anxiety, which they had previously successfully removed from their feeds. These examples showcase how participants struggled to connect their interactions with system responses, leaving them unable to control or manage their feeds and, in some cases, feeling demoralized and confused. The *theories of algorithmic exploitation* are the foundation of the entanglement of disempowering with its direct

connection to concepts of forced and misrepresented engagement to drive compulsive consumption, and thus profits.

4.3 Mitigating Entanglement Through Design

Participants worked to resolve entanglements caused by algorithmic curation to protect their mental well-being. Our analysis reveals three major design themes our participants proposed to address entanglement: contextualizing engagement to preserve users' emotional intentions, consumption control for healthier interactions, and reclaiming control through explicit user input. Each high-level design theme is attempting to mitigate entanglement in the gulfs of execution and evaluation that our participants experienced.

4.3.1 Contextualizing Engagement. Participants' designs highlight a key tension: their contextual, intentional, and nuanced engagement was interpreted by platforms in simplistic, binary ways. Their designs aimed to create separation in three dimensions of intention: interest, amplification, and protection.

Interest. Many participants made design suggestions to refine the "like" button and comments to better capture users' intentions and interests. By separating engagement to reflect users' intentions around interest, participants aimed to combat the entanglement issues of "Flattening" and "Overloading" where they felt features controlled too many facets of the feed. For example, P03 suggested adding "dislike" options, but P11 noted that it could be "not good for creators' mental health" and that the algorithm tends to promote negative content. For comments, P01 proposed adding sentiment, which would prevent all comments from being interpreted as positive engagement.

Amplification. Participants wanted more nuanced ways to promote content to others. They distinguished between "liking" content and wanting to amplify, or see more of it in their feeds. Adding avenues for amplification divorced from interest aims to address "Flattening" and "Overloading" by disentangling user intentions and system actions. For instance, P01 added "show more or less" options to their design, explaining that it added dimensionality beyond "like or dislike" by indicating a desire to amplify content rather than indicating interest. Separating interest and amplification aims to address participants' complaints about mental health content, where it was of interest. Still, they do not always want it amplified, as it tends to overwhelm their feeds.

Protection. In adapting "blocking" features as seen in Figure 5b, participants created ways to remove harmful content regardless of community guidelines to protect their mental well-being. They wanted the ability to block channels, keywords, or topics, with P08 wanting to "throw out" content and "never have it come back". These designs aim to mitigate the entanglement issues participants experienced with "Guessing", "Hiding", and "Dangling", by creating clear removal avenues for harmful content. Rather than "reporting" which aligns with community guidelines, participants imagined softer protections for removing personally harmful content by "flagging [content] that [is] personally upsetting or unwanted" (P03) to "protect against triggers" (P10). Additional features included "descriptions of why something might have been flagged" (P10), "a disclaimer if multiple people have flagged the content" (P03), and "auto-blur" for triggering posts (P21).

4.3.2 Consumption Control. The second area of participant solutions revealed a need for interventions to support healthy content consumption and mental well-being; participants found current social media inventions missing or ineffective. Their designs created interventions that promoted intentional consumption through controlling engagement effects on feeds, intentional feed layouts, and personalized intervention strategies.

Controlling Engagement Effects. Participants wanted real-time control over their engagement to adjust feeds when they started having mood swings, which affected users' mental well-being. They envisioned sliders for real-time feed changes, allowing "prioritized or deprioritized engagement" (P05). Sliders to manage diversity, or "how far" (P05) recommendations deviate, were also suggested to control the appearance of content "on the fringe" (P09) of their interests. At the extremes, these sliders could allow engagement to be "incognito" (P04) and not affect their feeds. This direct control of engagement aims to address "Hiding" and "Disempowering" entanglement issues by exposing paths for user control.

Intentional Layouts. Participants criticized current social media layouts for being optimized for mindless consumption rather than mental well-being. Block layouts with "aesthetic similar to Pinterest" (P16) (see figure 5c) and eliminating "infinite scroll" (P10) and auto-play, which participants noted encouraged "doom scroll" (P03), were proposed to create more intentional interactions by removing forced watch time. These designs aim to address the entanglement issues of "Flattening", "Disempowering", and "Overloading" by introducing intentionality and reducing problematic engagement metrics.

Personalized Interventions. While participants found platform interventions aimed at content consumption, like "stop scrolling" message, potentially helpful, the lack of personalization made them ineffective. Participants wanted "the ability to voluntarily turn on the warnings to take a break from social media" (P09) with customizable conditions and messages, including condition logic for specific behaviors and message contents and presentation. Examples of some conditions included "screen time" that asked "if the user wants to continue" (P03) or limiting "the number of posts that you see in a day" (P06). For messaging, participants wanted "more comprehensive mental health resources outside of the suicide hotline" (P10). Personalized intervention designs aim to mitigate "Hiding" and "Disempowering" entanglements by making intervention features visible and controllable for users.

4.3.3 Reclaiming Feed Control through Explicit Input. Frustration with unpredictable feed behaviors permeated participant designs, which emphasized user-centered, explicit controls. The designs introduced mechanisms to create contextual feeds by organizing and categorizing content to support their mental well-being.

Organizing for Contextual Feeds. Participants repeatedly requested separate feeds that allow for explicit organizational control over their feeds. These types of feeds were called "containers" (P04), "groupings" (P07), or "multiple FY pages" (P17), but shared similar functionality to create multiple feeds with fewer "mood swings". For example, P17 wanted a mental health feed containing "memes, personal experience, resources, etc" in one dedicated place. Participants wanted to move content between feeds manually and "toggle them on and off at will" (P07) when the algorithm learned

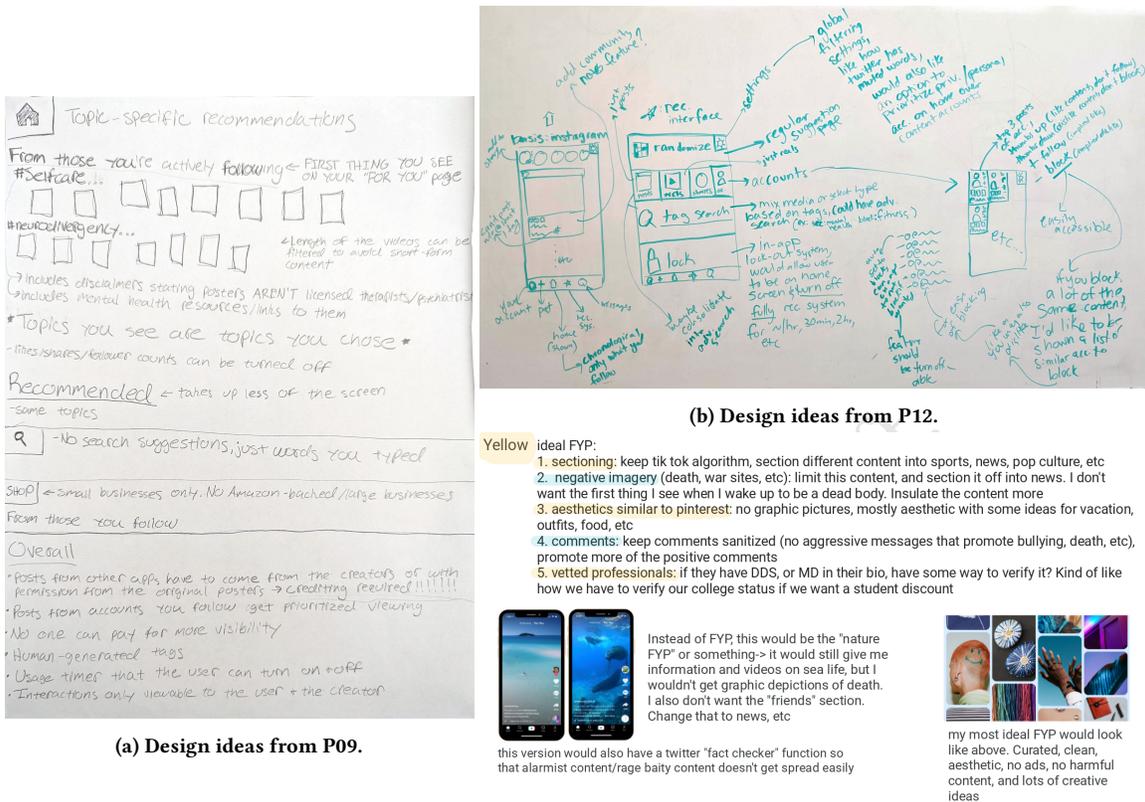


Figure 5: Examples of participant prototypes.

unwanted behaviors or included harmful content. Participants believed this design enabled them to create contextual interactions in their feeds and exercise greater control over content recommendations, thereby limiting issues such as disclosure leaks. Enabling explicit contextual control over interactions and feeds aims to address the entanglement issues of 'Flattening' and 'Disempowering'.

Categorizing for Content Control. Building on the previous design, participants wanted to provide “more explicit input into what the algorithms show them” (P12) through categorizing content with specific themes, or “tags”. Tags, as seen in Figures 5a and 5b, would enable clear feedback and explicit control over harmful content, allowing users to add, remove, mute, or block content based on its theme. Several participants wanted filtering capabilities like Tumblr or AO3, with some requesting “require tags on posts”. Most participants favored “users curating their own tags” (P07) as platform-generated tags could make “communities might feel boxed-in” (P07). Some participants thought about using tags as a starting point for content in the aforementioned feeds. Tags aim to address the entanglement issues of “Guessing” and “Hiding” by providing dedicated, transparent mechanisms for content control.

5 Discussion

5.1 Theory Implication: Entanglement, Joined Systems, and Norman Design Interactions

Our findings show that users experience entanglement, where they are unable to connect their actions to their consequences on social media platforms. Attempts to curate content for their mental well-being through interface interactions left participants entangled, exposing them to emotionally erratic feed behaviors that harmed their well-being. Participants heavily critiqued social media platforms and their emergent issues with entanglement, aspects of which have been explored in prior work on self-control and infinite scrolling feeds for platforms like YouTube [58].

Our work reveals a fundamental design problem that we hypothesize has led to the current entanglement seen on social media platforms. Two distinct technological systems, each with its own function, have been merged into a single interface and platform (e.g., social networking and algorithmic curation systems in platforms like TikTok and Instagram). Classic examples of technological systems, such as Norman’s doors and teapots [66], provide clear, distinct functions that users can easily troubleshoot when problems arise; in contrast, modern algorithmic systems and their functions are often opaque. In many prior systems, such as adding search

to email, integrating the system into the interface creates a new function clearly distinct from the original system.

However, by integrating the distinct functions of social networking and algorithmic curation into a single, intertwined interface, current social media platforms obscure how the system works. Our results on “Guessing” and “Dangling” showed that traditional social media features (likes, comments, shares) now play entangled roles, serving both as platform interactions and as data for algorithmic personalization. Prior research shows that both social networking [30, 36] and recommender systems [60] negatively affect users’ mental well-being, further complicating the issue. The intertwined interfaces are evident in the features platforms offer, which have remained essentially unchanged after the integration of algorithmic curation, with only minimal additions, such as the “not interested” feature. Said plainly, the interface of social networking has been forced to control algorithmic curation on social media, leading to the entanglements we found in our study.

Our work on entanglement builds on the foundational work of Norman [66] to investigate what happens when technological platforms, particularly those that affect users’ behaviors and mental well-being, attempt to combine two systems. Successful navigation of Norman’s action cycle requires a strong mental model of how a technology works. However, users cannot construct these mental models to reason about the actions to take to influence algorithmic behaviors. Overlaying recommender system affordances (e.g., implicit user preferences) onto traditional social media features (e.g., liking and commenting) obscures crucial information for building mental models, thereby creating entanglement.

The concept of entanglement can further help us understand and anticipate platform challenges arising from integrating different systems in the future. A timely example of where entanglement could be applied is generative AI in existing search engines. Search engines and generative AI differ in their processes and outcomes: one *retrieves* information, while the other *generates* it. Yet generative AI is currently integrated into existing search engines. We can already see that entanglement may be occurring in this case, with reports of issues with trust and cognitive damage [56, 95]. Entanglement could explain why distrust is occurring, as users struggle to distinguish between credible information from websites and that generated by generative AI. By using the proposed framework to analyze users’ experiences interacting with the system, researchers can identify the specific issues users face, thereby disentangling the system to support users’ needs and well-being.

5.2 Design Implications: Combating Entanglement

Entanglement disrupts users’ ability to form mental models of the feeds by obscuring connections between actions and system response, directly affecting their mental well-being. Addressing entanglement requires rethinking the relationship between social media, recommender systems, users, and content. A straightforward solution to entanglement could be to map all desired features to buttons. However, that would over-encumber the interface, creating an equally stressful experience for users. We must find a balanced solution to support users without burdening them when interacting with these systems.

Addressing entanglement requires a fundamental reconceptualization of how algorithmic curation supports users rather than “one-size-fits-all” interface changes or algorithmic updates. While companies have financial incentives to keep users engaged on their platforms, users grow dissatisfied with platform design practices and are likely to migrate to alternative options. Thus, supporting users’ needs ultimately should align with the platform’s long-term interests. Some platforms have even begun implementing wellness features, such as watch notifications and activity centers, to promote well-being. In this section, we propose two design implications based on participants’ designs (see Section 5): restoring intentional control through contextually aware interfaces and reimagining algorithmic curation through a contextually aware lens.

5.2.1 Restoring Intentional Control: “Choose Your Own Adventure” Interventions. Our first design implication addresses entanglement issues by changing interfaces to establish context and communicate user intentions, thereby rebuilding the connections between users’ actions and system responses. We frame these as “choose your own adventure” interventions, embodying users’ control over their feeds to support their mental well-being.

Personalized Safeguards for Mental Well-being. Existing platform interventions for mental well-being, such as screen-time timers, “stop scrolling” messages, and crisis pop-ups, were ineffective for our participants. Participants perceived these interventions as generic and disingenuous, given their distrust of the platform’s intentions. Moreover, in Section 4.1, participants believed that these simplistic interventions removed their agency and choice in safeguarding their mental well-being.

We recommend a suite of user-tailored interventions that mitigate entanglement through restoring user agency and trust. For example, in Section 4.3.2, participants envisioned customizable feed layouts and direct manipulation of feed behaviors to alter distressing patterns of behavior directly. Volume- or topic-based threshold or filter approaches were also suggested to allow mapping from users’ plain language to interface behaviors, e.g., after 10 videos on a heavy topic, do not show any more for the rest of the day. Personalized interventions support users’ agency in protecting their mental well-being.

These intervention approaches push against the “one size fits all” responses to extreme behaviors, instead prioritizing users’ agency and early interventions to mediate harmful behavior patterns. However, implementing these interventions requires platforms to shift from viewing mental well-being as legal risk mitigation to genuine user support. Developing and evaluating diverse, well-grounded interventions requires joint research efforts between HCI and psychology to ensure they support mental health and well-being needs.

Disentangling Context Through Separate Feeds. Current social media platforms are cross-contaminating user intentions and mental states by compressing them into single feeds. Participants designed modular feeds that respect users’ diverse contexts, enabling them to maintain isolated feeds for distinct intentions, interests, and mental states.

We therefore recommend separate feeds that promote content organization. For example, platforms could provide multiple feeds, similar to browser “tabs,” allowing users to easily create, modify, and remove feeds on demand. These “tabs” could understand users’

intentions through various methods, such as content “tags” (Section 4.3.3), aesthetic preferences, or explicit questions about emotional and mental states to inform feed creation. Platforms like BlueSky have adopted user-curated feeds that users can “follow” or manually curate. This approach effectively expands a user’s ego-network beyond direct following to include users who contribute to those feeds. What our participants wanted differed from existing feeds: they wanted to retain the serendipitous discovery enabled by algorithmic curation while maintaining control over feed content.

This approach builds on participants’ content organization designs from Section 4.3.3, which research shows supports autonomy and self-esteem [69]. Implementing separate feeds requires maintaining separate user models and respecting contextual boundaries, which builds on prior calls for algorithmic feed control [59, 82]. These designs must balance user control with the cognitive burden imposed by such systems. Evaluating these trade-offs through usability studies would need insights from psychology to assess mental health and well-being impacts.

5.2.2 Reimagining Algorithmic Curation: From Accurate to Emotionally Dynamic. Our second approach targets entanglement by shifting the curation algorithm’s optimization from accuracy to contextual and emotional relevance. Current algorithms are seen as accurate but fail to account for context or user intentions.

Content-Aware Recommendations. Current algorithmic curation can not distinguish between different topics and tones of content that affect users’ interactions and intentions in engagement. This lack of contextual awareness stems from joined systems (defined in Section 5.1) as social media is inherently emotional [87], while recommender systems focus more on binary metrics.

We propose an ensemble approach using multiple specialized models, each tailored to specific content types and contexts. Instead of a single monolithic model, multiple smaller models could be trained for specific contexts and user intentions. For example, separate models could be used for traditional recommendation contexts like “BookTok” content, and for more complex content like “MentalHealthTok” content that focuses on social support and mental well-being. An ensemble approach would support the multi-feed “tabs” suggested in Section 5.2.1, with each tab powered by the appropriate models.

The ensemble approach extends prior work on the contextual ambiguity in feeds [59, 94] and media richness theory [24]. Fully addressing contextual awareness requires collaboration between HCI, social computing, and recommender systems researchers to integrate users’ insights about context into algorithmic design.

User-Aware Recommendations. Beyond content context, algorithmic curation must consider user context, particularly emotional and mental states, which determine users’ capacity to engage with specific topics or tones. While recommender systems research has explored context [52], including location [34] and mood [27, 74], mental state remains unexplored despite participants identifying it as distinct from mood or emotions.

We propose integrating mental state through explicit user specification combined with system inference. For example, users could specify they prefer comforting content over emotionally heavy topics in the morning, allowing the system to adjust recommendations based on time, emotional capacity, and content weight. This

approach explicitly incorporates the user’s mental state into system behavior. However, modeling the complex interplay among mental state dimensions requires significant advances in user modeling technology and contextual inference. Conceptualizing and operationalizing mental state as a context is also a rich area for future research for psychology and HCI. Additionally, we need an understanding of how contextual aspects appear in interaction data and are digitized for algorithmic interpretation. User modeling and interaction analysis provide starting points for such work.

6 Limitations and Future Work

The limitations of our study primarily revolve around the participant pool. Our initial recruitment focused on university campuses despite attempts to reach a broader online audience. Thus, participants lean heavily toward college-aged adults, all of whom have at least some college education. Our participants were also skewed white and female, which continues the trend of struggling to recruit men and racial/ethnic minorities into studies for mental health [9, 80]. Lastly, while we tracked whether participants reported experiencing mental health conditions (specifically depression, anxiety, and PTSD), we did not record which conditions or co-morbid diagnoses they specifically had out of respect for the participants’ privacy. Participants were free to self-disclose during the workshops, but these conditions and individuals do not represent the entirety of the mental health community.

We still believe our findings open several possible research directions. Entanglement and the consequences on joined systems are a rich area of future research, allowing us to understand integrated platforms and identify issues that affect users’ experiences and well-being. Considering the impacts of joined systems, this set a foundation for work evaluating current platforms through a new lens, updating existing knowledge and theories in social computing and HCI. Regarding design, our participants had concrete ideas for changing algorithmic curation to better support their mental well-being through features that enable explicit user control and mechanisms to organize recommended content. These ideas can be further studied in the areas of interface design and recommender systems when considering mental well-being. Expanding the research to more users, including those with other mental illnesses or marginalized identities, is vital to ensure designs are not over-generalized.

7 Conclusion

In this work, we examine how users experience algorithmic curation in social media and how they would like to see these systems evolve. Through design workshops, we discovered that users develop folk theories to make sense of their experiences with algorithmic curation. These folk theories showcase core issues with algorithmic design that we explain through entanglement, a phenomenon in which users cannot connect their actions to platform outcomes. We contribute to theory by building on Norman’s action cycle to develop the entanglement framework, adding new dimensions to the gulfs of execution and evaluation, and reconceptualizing integrated systems. Our findings provide design implications that rethink current social media algorithmic curation to support users’ mental well-being.

Acknowledgments

We would like to thank Leah Ajmani and Mo Houtti for their invaluable feedback on this work. Funding for this research was partially provided by UL Research Institutes through the Center for Advancing the Safety of Machine Intelligence.

References

- [1] Patricia A Adler and Peter Adler. 1987. *Membership roles in field research*. Vol. 6. Sage, USA.
- [2] Antonia Aguirre Velasco, Ignacio Silva Santa Cruz, Jo Billings, Magdalena Jimenez, and Sarah Rowe. 2020. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. *BMC psychiatry* 20, 1 (2020), 1–22.
- [3] Yoko Akama, Roslyn Cooper, Laurene Vaughan, Stephen Viller, Matthew Simpson, and Jeremy Yuille. 2007. Show and tell: Accessing and communicating implicit knowledge through artefacts. *Artifact: Journal of Design Practice* 1, 3 (2007), 172–181.
- [4] Nazanin Andalibi and Justin Buss. 2020. The human in emotion recognition on social media: Attitudes, outcomes, risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [5] Holly Avella. 2023. "TikTok≠ therapy": Mediating mental health and algorithmic mood disorders. *New Media & Society* (2023), 14614448221147284.
- [6] Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1301–1310.
- [7] Corey H Basch, Lorie Donelle, Joseph Fera, and Christie Jaime. 2022. Deconstructing TikTok videos on mental health: cross-sectional, descriptive content analysis. *JMIR formative research* 6, 5 (2022), e38340.
- [8] Ron Berman and Zsolt Katona. 2020. Curation algorithms and filter bubbles in social networks. *Marketing Science* 39, 2 (2020), 296–316.
- [9] Gillian Brown, Max Marshall, Peter Bower, Adrine Woodham, and Waqas Waheed. 2014. Barriers to recruiting ethnic minorities to mental health research: a systematic review. *International journal of methods in psychiatric research* 23, 1 (2014), 36–48.
- [10] Megan A Brown, James Bisbee, Angela Lai, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2022. Echo chambers, rabbit holes, and algorithmic bias: How YouTube recommends content to real users. Available at SSRN 4114905 (2022).
- [11] Taina Bucher. 2019. The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. In *The social power of algorithms*. Routledge, 30–44.
- [12] L Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. 2019. Controlling polarization in personalization: An algorithmic framework. In *Proceedings of the conference on fairness, accountability, and transparency*. 160–169.
- [13] Stevie Chancellor, Yannis Kalantidis, Jessica A Pater, Munmun De Choudhury, and David A Shamma. 2017. Multimodal classification of moderated online pro-eating disorder content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3213–3226.
- [14] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #thyghapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. ACM, 1201–1213.
- [15] Kathy Charmaz, Liska Belgrave, et al. 2012. Qualitative interviewing and grounded theory analysis. *The SAGE handbook of interview research: The complexity of the craft* 2 (2012), 347–365.
- [16] Jaeho Cho, Saifuddin Ahmed, Martin Hilbert, Billy Liu, and Jonathan Luu. 2020. Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. *Journal of Broadcasting & Electronic Media* 64, 2 (2020), 150–172.
- [17] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* 3 (2015), 222–248.
- [18] Nicholas C Coombs, Wyatt E Meriwether, James Caringi, and Sophia R Newcomer. 2021. Barriers to healthcare access among US adults with mental health challenges: A population-based study. *SSM-population health* 15 (2021), 100874.
- [19] Anjuli Corzine and Ananya Roy. 2024. Inside the black mirror: current perspectives on the role of social media in mental illness self-diagnosis. *Discover Psychology* 4, 1 (2024), 40.
- [20] Sasha Costanza-Chock. 2018. Design justice: Towards an intersectional feminist framework for design theory and practice. (2018).
- [21] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
- [22] Kelley Cotter, Julia R DeCook, Shaheen Kanthawala, and Kali Foyle. 2022. In FYP we trust: The divine force of algorithmic spirituality. *International Journal of Communication* 16, 2022 (2022), 1–23.
- [23] Zoë Natalia Cullen, Angela Y Lee, Brenna Davidson, Jeffrey T Hancock, and Nicole B Ellison. 2025. Not Just For You': How the Algorithmic Crystal Mediates Communication and Identity Work on TikTok's FYP. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–27.
- [24] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.
- [25] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 71–80.
- [26] Ivan De Martino, Rocco D'Apolito, Alexander S McLawhorn, Keith A Fehring, Peter K Sculco, and Giorgio Gasparini. 2017. Social media for patients: benefits and drawbacks. *Current reviews in musculoskeletal medicine* 10, 1 (2017), 141–145.
- [27] Yashar Deldjoo, Markus Schedl, Paolo Cremonesi, and Gabriella Pasi. 2020. Recommender systems leveraging multimedia content. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–38.
- [28] Michael A DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms ruin everything" # RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3163–3174.
- [29] Meagan Drilling. 2022. What to Know About Using TikTok as a Mental Health Resource. Retrieved September 8, 2022 from <https://www.healthline.com/health-news/what-to-know-about-using-tiktok-as-a-mental-health-resource>
- [30] Michele P Dyson, Lisa Hartling, Jocelyn Shulhan, Annabritt Chisholm, Andrea Milne, Purnima Sundar, Shannon D Scott, and Amanda S Newton. 2016. A systematic review of social media use to discuss and view deliberate self-harm acts. *PLoS one* 11, 5 (2016), e0155813.
- [31] Sheena Erete, Aarti Israni, and Tawanna Dillahunt. 2018. An intersectional approach to designing in the margins. *interactions* 25, 3 (2018), 66–69.
- [32] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I like it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2371–2382.
- [33] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 153–162.
- [34] Zeshan Fayyaz, Mahsa Ebrahimian, Dina Nawara, Ahmed Ibrahim, and Rasha Kashef. 2020. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *applied sciences* 10, 21 (2020), 7748.
- [35] Jessica L Feuston and Anne Marie Piper. 2018. Beyond the Coded Gaze: Analyzing Expression of Mental Health and Illness on Instagram. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 51:1–51:21. doi:10.1145/3274320
- [36] Jessica L Feuston and Anne Marie Piper. 2019. Everyday experiences: small stories and mental illness on Instagram. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.
- [37] Jessica L Feuston, Alex S Taylor, and Anne Marie Piper. 2020. Conformity of eating disorders through content moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–28.
- [38] Richard Fletcher and Rasmus Kleis Nielsen. 2019. Generalised scepticism: How people navigate news on social media. *Information, communication & society* 22, 12 (2019), 1751–1769.
- [39] Junling Gao, Pinpin Zheng, Yingnan Jia, Hao Chen, Yimeng Mao, Suhong Chen, Yi Wang, Hua Fu, and Junming Dai. 2020. Mental health problems and social media exposure during COVID-19 outbreak. *PLoS one* 15, 4 (2020), e0231924.
- [40] Tarleton Gillespie. 2014. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society* 167, 2014 (2014), 167.
- [41] Jonathan Haidt. 2024. *The anxious generation: How the great rewiring of childhood is causing an epidemic of mental illness*. Random House.
- [42] Farah Hasan, Melissa M Foster, and Hyunyi Cho. 2023. Normalizing anxiety on social media increases self-diagnosis of anxiety: The mediating effect of identification (but not stigma). *Journal of health communication* 28, 9 (2023), 563–572.
- [43] Susan Hekman. 1997. Truth and method: Feminist standpoint theory revisited. *Signs: Journal of women in culture and society* 22, 2 (1997), 341–365.
- [44] Dilip V Jeste, Barton W Palmer, Paul S Appelbaum, Shahrokh Golshan, Danielle Glorioso, Laura B Dunn, Kathleen Kim, Thomas Meeke, and Helena C Kraemer. 2007. A new brief instrument for assessing decisional capacity for clinical research. *Archives of general psychiatry* 64, 8 (2007), 966–974.
- [45] Stephanie Joseph, Stephanie Miodus, and Frank Farley. 2024. Apps and Care: Clinical and Societal Applications of Media. In *Handbook of Media Psychology: The Science and The Practice*. Springer, 169–186.
- [46] Fazida Karim, Azeizat A Oyewande, Lamis F Abdalla, Reem Chaudhry Ehsanullah, and Safeera Khan. 2020. Social media use and its connection to mental health: a systematic review. *Cureus* 12, 6 (2020).
- [47] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic folk theories and identity: How TikTok users co-produce Knowledge

- of identity and engage in algorithmic resistance. *Proceedings of the ACM on human-computer interaction* 5, CSCW2 (2021), 1–44.
- [48] Betül Keles, Niall McCrae, and Anmarie Grealish. 2020. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International journal of adolescence and youth* 25, 1 (2020), 79–93.
- [49] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22 (2012), 441–504.
- [50] Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction* 22 (2012), 101–123.
- [51] Rachel Kornfield, Jonah Meyerhoff, Hannah Studd, Ananya Bhattacharjee, Joseph Jay Williams, Madhu Reddy, and David C Mohr. 2022. Meeting users where they are: user-centered design of an automated text messaging tool to support the mental health of young adults. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [52] Saurabh Kulkarni and Sunil F Rodd. 2020. Context Aware Recommendation Systems: A review of the state of the art techniques. *Computer Science Review* 37 (2020), 100255.
- [53] Annie YS Lau, Elia Gabarron, Luis Fernandez-Luque, and Manuel Armayones. 2012. Social media in health—what are the safety concerns for health consumers? *Health Information Management Journal* 41, 2 (2012), 30–35.
- [54] Mark Ledwich and Anna Zaitsev. 2019. Algorithmic extremism: Examining YouTube’s rabbit hole of radicalization. *arXiv preprint arXiv:1912.11211* (2019).
- [55] Angela Y Lee, Hannah Mieczkowski, Nicole B Ellison, and Jeffrey T Hancock. 2022. The algorithmic crystal: Conceptualizing the self through algorithmic personalization on TikTok. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–22.
- [56] Hao-Ping Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems*. 1–22.
- [57] Valerie Lookingbill and Kimanh Le. 2024. “There’s Always a Way to Get Around the Guidelines”: Nonsuicidal Self-Injury and Content Moderation on TikTok. *Social Media+ Society* 10, 2 (2024), 20563051241254371.
- [58] 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.
- [59] Ashlee Milton, Leah Ajmani, Michael Ann DeVito, and Stevie Chancellor. 2023. “I See Me Here”: Mental Health Content, Community, and Algorithmic Curation on TikTok. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [60] Ashlee Milton and Stevie Chancellor. 2022. The Users Aren’t Alright: Dangerous Mental Illness Behaviors and Recommendations. *The 5th FAccTRec Workshop: Responsible Recommendation* (2022).
- [61] Ashlee Milton, Juan F Maestre, Abhishek Roy, Rebecca Umbach, and Stevie Chancellor. 2024. Seeking in Cycles: How Users Leverage Personal Information Ecosystems to Find Mental Health Information. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
- [62] John A Naslund and Kelly A Aschbrenner. 2019. Risks to privacy with use of social media: understanding the views of social media users with serious mental illness. *Psychiatric services* 70, 7 (2019), 561–568.
- [63] J. A. Naslund, K. A. Aschbrenner, L. A. Marsch, and S. J. Bartels. 2016. The future of mental health care: peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences* 25, 2 (2016), 113–122. doi:10.1017/S2045796015001067
- [64] John A Naslund, Ameya Bondre, John Torous, and Kelly A Aschbrenner. 2020. Social media and mental health: benefits, risks, and opportunities for research and practice. *Journal of technology in behavioral science* 5 (2020), 245–257.
- [65] Kimberly A Neuendorf. 2018. Content analysis and thematic analysis. In *Advanced research methods for applied psychology*. Routledge, 211–223.
- [66] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
- [67] Donald A Norman. 1981. Categorization of action slips. *Psychological review* 88, 1 (1981), 1.
- [68] National Institute of Mental Health. 2023. Mental illness. <https://www.nimh.nih.gov/health/statistics/mental-illness>
- [69] World Health Organization et al. 2022. World mental health report: transforming mental health for all. (2022).
- [70] Rikke Ørningreen and Karin Levinsen. 2017. Workshops as a research methodology. *The Electronic Journal of e-learning* 15, 1 (2017), 70–81.
- [71] Derek O’Callaghan, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. 2015. Down the (white) rabbit hole: The extreme right and online recommender systems. *Social Science Computer Review* 33, 4 (2015), 459–478.
- [72] Michelle O’reilly, Nisha Dogra, Natasha Whiteman, Jason Hughes, Seyda Eruyar, and Paul Reilly. 2018. Is social media bad for mental health and wellbeing? Exploring the perspectives of adolescents. *Clinical child psychology and psychiatry* 23, 4 (2018), 601–613.
- [73] Kari Paul. 2022. What TikTok does to your mental health: ‘It’s embarrassing we know so little’. Retrieved June 6, 2024 from <https://www.theguardian.com/technology/2022/oct/30/tiktok-mental-health-social-media>
- [74] Marco Polignano, Fedelucio Narducci, Marco de Gemmis, and Giovanni Semeraro. 2021. Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors. *Expert Systems with Applications* 170 (2021), 114382.
- [75] Claudette Pretorius, Derek Chambers, David Coyle, et al. 2019. Young people’s online help-seeking and mental health difficulties: Systematic narrative review. *Journal of medical Internet research* 21, 11 (2019), e13873.
- [76] Claudette Pretorius, Darragh McCashin, and David Coyle. 2022. Mental health professionals as influencers on TikTok and Instagram: What role do they play in mental health literacy and help-seeking? *Internet Interventions* 30 (2022), 100591.
- [77] Caroline Ramazanoglu and Janet Holland. 2002. From truth/reality to knowledge/power, taking a feminist standpoint. *Feminist Methodology Challenges and Choices* (2002), 68–69.
- [78] Kitty Rusking. 2022. For Some Women With ADHD, TikTok Is the First Place They Felt Heard. Retrieved December 1, 2022 from <https://time.com/6211695/adhd-tiktok-women/>
- [79] Javier Saavedra, Elvira Pérez, Paul Crawford, and Samuel Arias. 2018. Recovery and creative practices in people with severe mental illness: evaluating well-being and social inclusion. *Disability and rehabilitation* 40, 8 (2018), 905–911.
- [80] Catherine Sanchez, Adrienne Grzenda, Andrea Varias, Alik S Widge, Linda L Carpenter, William M McDonald, Charles B Nemeroff, Ned H Kalin, Glenn Martin, Mauricio Tohen, et al. 2020. Social media recruitment for mental health research: a systematic review. *Comprehensive psychiatry* 103 (2020), 152197.
- [81] Donghee Shin, Kerk F Kee, and Emily Y Shin. 2022. Algorithm awareness: Why user awareness is critical for personal privacy in the adoption of algorithmic platforms? *International Journal of Information Management* 65 (2022), 102494.
- [82] Ellen Simpson, Andrew Hamann, and Bryan Semaan. 2022. How to tame’ your’ algorithm: LGBTQ+ users’ domestication of TikTok. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–27.
- [83] Meghan Sit, Sarah A Elliott, Kelsey S Wright, Shannon D Scott, and Lisa Hartling. 2022. Youth Mental Health Help-Seeking Information Needs and Experiences: A Thematic Analysis of Reddit Posts. *Youth & Society* (2022), 0044118X221129642.
- [84] Maddy Slattery, Hayley Attard, Victoria Stewart, Helena Roennfeldt, and Amanda J Wheeler. 2020. Participation in creative workshops supports mental health consumers to share their stories of recovery: A one-year qualitative follow-up study. *PLoS one* 15, 12 (2020), e0243284.
- [85] Jessie J Smith, Lucia Jayne, and Robin Burke. 2022. Recommender systems and algorithmic hate. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 592–597.
- [86] Clay Spinuzzi. 2005. The methodology of participatory design. *Technical communication* 52, 2 (2005), 163–174.
- [87] Steffen Steinert and Matthew James Dennis. 2022. Emotions and digital well-being: On social media’s emotional affordances. *Philosophy & technology* 35, 2 (2022), 36.
- [88] Joëlle Swart. 2021. Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social media+ society* 7, 2 (2021), 20563051211008828.
- [89] Patti M Valkenburg. 2022. Social media use and well-being: What we know and what we need to know. *Current opinion in psychology* 45 (2022), 101294.
- [90] Brooke White, Katie Quehl, Jingjing Wang Costello, and Jingjing Wang Costello. 2022. A Workshop on Brainstorming Workshops Successful brainstorming with diverse creators in a hybrid work environment. In *Proceedings of the 14th Conference on Creativity and Cognition* (Venice, Italy) (C&C ’22). Association for Computing Machinery, New York, NY, USA, 494–497. doi:10.1145/3527927.3531204
- [91] Belle Wong. 2023. Top Social Media Statistics And Trends Of 2024. Retrieved September 10, 2024 from <https://www.forbes.com/advisor/business/social-media-statistics/>
- [92] Lu Yan and Yong Tan. 2014. Feeling blue? Go online: an empirical study of social support among patients. *Information Systems Research* 25, 4 (2014), 690–709.
- [93] Venessa Yurkevich. 2023. Why experts worry TikTok could add to mental health crisis among US teens. Retrieved June 6, 2024 from <https://www.cnn.com/2023/01/11/tech/tiktok-teen-mental-health/index.html>
- [94] Alice Qian Zhang, Ashlee Milton, and Stevie Chancellor. 2023. # Pragmatic or# Clinical: Analyzing TikTok Mental Health Videos. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 149–153.
- [95] Yue Zhang, Jiao Tian, and Fengyi Deng. 2025. In generative AI we trust: an exploratory study on dimensionality and structure of user trust in ChatGPT. *Interacting with Computers* (2025), iwaf029.