FinerMe: Examining App-level and Feature-level Interventions to Regulate Mobile Social Media Use

ADIBA ORZIKULOVA, KAIST, Republic of Korea
HYUNSUNG CHO, Carnegie Mellon University, USA
HYE-YOUNG CHUNG, Hanyang University, Republic of Korea
HWAJUNG HONG, KAIST, Republic of Korea
UICHIN LEE, KAIST, Republic of Korea
SUNG-JU LEE*, KAIST, Republic of Korea

Many digital wellbeing tools help users monitor and control social media use on their smartphones by tracking and setting limits on their usage time. Tracking is typically done at the granularity of phone- or app-level; however, recent social media apps provide various *features* such as direct messaging, comment reading/posting, and content uploading/viewing. While it is possible to track and analyze *within-app* feature usage, little is known about the effect of granularity on smartphone interventions. We designed and developed *FinerMe* to explore how the granularity of interventions (app-level vs. feature-level) affects the usage of popular social media such as Instagram and YouTube on smartphones. We conducted a field study with 56 participants over 16 days that consisted of three phases: baseline collection, self-reflection, and self-reflection with restrictive interventions. The results showed that while both app-level and feature-level interventions similarly reduced social media use, feature-level interventions enabled users to spend less time on passive app features related to content consumption (e.g., following feed on Instagram, and viewing comments on YouTube) than app-level interventions. Moreover, when self-reflection is combined with restrictive interventions at the feature-level, users were more reflective on their usage behavior than when done at the app-level.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Interaction paradigms**; **User studies**.

Additional Key Words and Phrases: digital wellbeing, smartphone intervention design, application feature, intervention granularity

ACM Reference Format:
Adiba Orzikulova, Hyunsung Cho, Hye-Young Chung, Hwajung Hong, Uichin Lee, and Sung-Ju Lee. 2023. FinerMe: Examining App-level and Feature-level Interventions to Regulate Mobile Social Media Use. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 274 (October 2023), 30 pages. https://doi.org/10.1145/3610065

1 INTRODUCTION

Mobile Social Media (SM) allows us to stay connected with our friends, build and maintain profiles, join groups with like-minded people, get informed about recent news and events, and

*corresponding author

Authors' addresses: Adiba Orzikulova, KAIST, Republic of Korea, adiorz@kaist.ac.kr; Hyunsung Cho, Carnegie Mellon University, USA, hyunsung@cs.cmu.edu; Hye-Young Chung, Hanyang University, Republic of Korea, hy2315@hanyang.ac.kr; Hwajung Hong, KAIST, Republic of Korea, hwajung@kaist.ac.kr; Uichin Lee, KAIST, Republic of Korea, uclee@kaist.ac.kr; Sung-Ju Lee, KAIST, Republic of Korea, profsj@kaist.ac.kr.

2573-0142/2023/10-ART274 $15.00
https://doi.org/10.1145/3610065

create and exchange diverse content instantly. However, SM can also be a great source of distraction and may have negative effects such as negative emotion [7, 60], weaker self-control [39, 46], poor academic performance [18], addiction [2, 61], anxiety, depression [5], eating concerns, and sleep disturbance [40]. Despite being aware of such negative consequences, many people still struggle to self-regulate and limit their SM use [47].

A considerable amount of research has been devoted to self-regulating mobile SM and smartphone usage. Prior studies examined various digital self-regulation and intervention techniques based on self-reflection [50, 59], restriction [29, 31, 41], social support [22, 26, 30, 32], and nudging [34–36]. These mechanisms help users to regulate their smartphone or specific application usage *as a whole* by informing about the accumulated smartphone or app usage statistics, limiting device or app usage during a pre-defined period or context, discouraging usage, leveraging social support, and applying different rules to trigger interventions.

However, prior works paid little attention to the unique characteristics of emerging social media apps that require frequent context switching between various features within an app. For instance, people can create and watch videos (YouTube videos, Facebook watch, Instagram stories), form and join groups (Facebook groups), follow people and contents of interest (YouTube channels, Instagram following feed, and Facebook news feed), browse system-recommended contents (Instagram suggested posts and videos), and even shop (Instagram shopping, Facebook shopping). These various app features can be associated with *active* and *passive* SM use. Active SM use involves direct interaction with others (e.g., direct messaging), whereas passive SM use involves consuming provided content (e.g., following feeds and watching videos).

To maintain user engagement, social media apps provide a mixture of features that users deem essential or addictive [7] and place them close to each other to facilitate frictionless user flow. Several recent studies explored feature-level user behavior analysis and intervention design [7, 12, 33, 46]. HabitLab [34] implemented feature-level interventions for social media web apps, including removing feed/comments from Facebook and hiding sidebars/comments from YouTube. Similarly, another work [47] investigated the effectiveness of feature-level interventions such as removing news feeds, goal reminders, and applying white background to help users reduce time spent on a Facebook web app. Although these studies examined feature-level interventions, only limited app features such as feeds, comments, and reactions were considered in web settings. In mobile settings with more app features involved, Cho et al. [7] segmented social media app usage into a sequence of app feature uses to find out which app feature use patterns resulted in more regretful experiences. Lukoff et al. [46] studied which internal mechanisms (e.g., recommendations, search, playlist, and autoplay) of YouTube were associated with a lack of self-control.

However, there is a lack of empirical research investigating the actual impact of feature-level interventions for regulating today's social media use. We explore the differences in social media use behavior change and user preferences when they are intervened at app- and feature-levels. We aim to learn what dimensions we need to consider when designing digital self-control tools incorporating app features. We designed and implemented app-level and feature-level versions of a self-regulatory intervention app, *FinerMe*, for Instagram and YouTube, two of the most popular social media apps [45]. We conducted an in-the-wild, between-group experiment (n=56), followed by in-depth interviews (n=36). Our field experiment consisted of three phases with five days each: Phase 1 (Baseline data collection), Phase 2 (Self-reflection intervention), and Phase 3 (Self-reflection and usage restriction intervention). There was also a transition step (one day) between Phase 2 and Phase 3 for setting time limit goals. We applied the same type of interventions at two granularities: (1) interventions based on self-reflection provide information about usage (e.g., **app-level**: use time and the number of visits to the Instagram app, **feature-level**: usage time and the number of visits to Instagram's *following feed* feature (see Appendix A)), (2) goal setting allows setting limits on usage (e.g., **app-level**: use YouTube app maximum one hour per day, **feature-level**: browse YouTube *home* feature maximum 10 minutes per day), and (3) restrictive interventions limit usage after exceeding the daily limit (e.g., **app-level**: limit access to the Instagram app, **feature-level**: limit access to Instagram's *watching reels* feature).

The main research question we want to answer is: *How does applying interventions at the app-level and feature-level impact mobile social media use behavior?*

Our study discovered that self-reflection combined with restrictive interventions (Phase 3) reduced overall time spent on social media apps for both intervention groups. The app feature use analysis revealed that participants in the feature-level intervention group reduced usage of the passive features (e.g., viewing comments on YouTube, following feed on Instagram) more compared with the participants in the app-level intervention group. Moreover, feature-level restrictive interventions showed that self-awareness of regretful usage was positively correlated with daily mean use time. Based on our findings, we discuss feature-level intervention design space for self-regulatory digital tools.

The key contributions of this paper are as follows. To our best knowledge, this is the first work investigating the impact of intervention granularity (app-level vs. feature-level) on smartphone app usage. Through the deployment of our app-level and feature-level self-regulatory system, *FinerMe*, we discovered that feature-level interventions reduce time spent on passive features more than app-level counterparts. Furthermore, our study revealed that a combination of self-reflection and restrictive interventions is more likely to induce regret among feature-level participants. As social media applications provide various features, we discuss the importance of exploring novel design spaces for integrating app features into smartphone interventions. The findings from our study suggest that digital wellbeing experts should take users' app usage patterns, intervention goals, preferences, and app feature characteristics into account when designing self-regulatory tools harnessing app features.

2 RELATED WORK

2.1 Digital Self-Control and Supporting Tools

Due to the negative consequences of technology overuse, tech giants such as Apple and Google introduced tools (ScreenTime on iOS [67] and Digital Wellbeing on Android [69]) embedded in mobile operating systems to promote digital wellbeing. These tools allow users to monitor their screen time, set time limits, and restrict themselves during pre-defined times and periods. With rising concerns of people struggling to limit their social media use, even social media companies introduced in-app tools to promote digital wellbeing. For instance, Meta integrated self-tracking and regulative functionalities into Facebook and Instagram apps [51]. However, these tools are disabled by default [20], hard to find [55], and do not offer fine-grained control of limiting diverse app features. It is debated that such reluctance to fully promote digital wellbeing stems from the fact that these tools would harm the business model of social media companies [20, 55].

A great amount of research effort has been made to designing self-regulatory tools and interventions to enhance digital wellbeing. They can broadly be categorized into interventions based on self-tracking and monitoring [15, 50, 59] and restrictive self-regulatory mechanisms [29, 31, 41, 72].

Self-tracking and monitoring are one of the most popular types of self-regulative digital interventions [21, 48, 50, 52, 58, 59]. Many productivity tools such as RescueTime [58], ManicTime [48], and Forest [56] track screen time and app usage to help users regulate technology overuse. MyTime [21] notifies users when they open monitored apps or exceed the daily time limits set by users. Similarly, GoodVibrations [52] provides vibration feedback, which helped participants to reduce daily usage by 20% during the intervention period and made users more aware of app usage patterns. Although these tracking and monitoring mechanisms were found to be effective in raising users' awareness about overall smartphone usage, there has been little work on the analysis of such self-regulation interventions when applied at the feature-level.

Restrictive self-regulatory mechanisms such as phone lockouts [41], lockout tasks serving as interaction friction [29, 72], synchronous lockouts leveraging social support [31, 32], and lockout in multi-device settings [27] are also well studied. Although such restrictive approaches are known to be effective in regulating social media use compared with self-reflection approaches, they also resulted in annoyance and discomfort when smartphone usage is necessary. Usage purposes could be different depending on user contexts, and user needs may be related to the usage of specific apps or features. This insight from prior studies [29] suggests that app-level and feature-level use patterns of a user should be considered when designing restrictive interventions.

2.2 Granularity of Smartphone Interventions

Digital self-control tools target the use of a specific device, application, or app feature. However, little work has been done on how the granularity of such interventions impacts users' behavior [34, 50]. For example, Monge Roffarello et al. [50] implemented an app called Socialize, where users could set various (tracking, restricting) interventions at the device level and app level. The results revealed that participants preferred app-level interventions to device-level interventions: they set more app-level interventions and reduced time significantly for specific apps (Facebook and Instagram) rather than the total reduction of use time on their mobile devices. Although their work compares the differences in intervention granularity, these comparisons were done only at the device- and app levels, but not at the feature level.

Social media apps such as YouTube and Instagram offer a wide range of app features (e.g., chatting, posting, video streaming, video calling, shopping, news viewing, etc.). Prior studies revealed that some of these app features can be addictive, regretful [7], or contain dark patterns [33]. HabitLab [34] is one of the earlier works that considers different app features when designing digital interventions for websites that people have less control over. On Facebook, HabitLab browser extension enables removing the news feed, comments, and clickbait from the news feed. For the YouTube website, users can remove sidebar links or comments, and be prompted before watching a video. Another work by Lyngs et al. [47] removed the news feed from the Facebook website as a feature-level intervention that resulted in a significant decrease in visit duration. Although these works applied interventions at a feature-level, they work only in web settings and offer interventions only for a limited set of features (feed, comments, and links).

For mobile platforms, several recent studies examined app usage with a wide variety of app features. Lukoff et al. [46] conducted co-design sessions with YouTube users and analyzed which internal mechanisms or YouTube features (e.g., recommendations, ads, playlists, autoplay, notifications, watch history, and stats) were associated with self-control. The results suggested that users prefer a YouTube interface with mechanisms that can help users feel higher self-control. Similarly, Zhang et al. [76] re-designed a mobile Twitter client app called Chirp to explore the differences in internal and external mechanisms affecting the users' self-agency and found that the internal mechanisms such as filtering the tweets were more effective than traditional external mechanisms (viewing Twitter usage in a dashboard). Cho et al. [7] analyzed feature-level usage behaviors (e.g., viewing, uploading, chatting) and the nature of contents (e.g., posts, videos, notifications). The app features that involved a more passive form of interaction including browsing feeds were more correlated with high regret levels.

While prior studies revealed the importance of feature-level self-control support in social media usage, there is a lack of experimental studies on how the intervention granularity affects social media usage behaviors. We attempt to fill this gap by implementing the intervention mechanisms at different granularity levels (app-level and feature-level) and conducting a field experiment to learn about the changes in social media use behavior and user preferences.

3 FINERME

Our objective is to observe how self-regulatory smartphone interventions of different granularity (i.e., app-level and feature-level) impact participants' social media use behavior. To accomplish this goal, we designed and implemented two different versions of *FinerMe*, a self-regulatory intervention system that supports app-level and feature-level interventions. *FinerMe* works with YouTube and Instagram mobile apps, and can be easily extended to other social media platforms.

3.1 Theoretical Groundings and Design Concepts

Our intervention system consists of three core design components: (1) self-reflection, (2) goal setting, and (3) restriction leveraging interaction friction. Regulatory mechanisms based on self-reflection are one of the most common design interventions promoting digital wellbeing [21, 59, 67, 69]. Although interventions based on self-reflection are widely used, previous literature [8] demonstrated that self-reflection alone is not sufficient to change user behavior. On the other hand, earlier studies [37, 38] highlighted the importance of leveraging self-reflection as an intervention. These works showed that users might struggle with reducing smartphone screen time due to their lack of mindfulness and awareness of their usage time and patterns. In addition, a recent study [7] introduced that feature-level self-reflection helped participants create actionable plans on which features of the app they wanted to spend less time. To quantitatively evaluate whether feature-level self-reflection impacts usage behavior in comparison with app-level self-reflection, we integrated self-reflection into our system components. We devised a self-reflection component of our intervention system based on the social cognitive theory (SCT) of self-regulation proposed by psychologist Albert Bandura [4]. Bandura suggests that self-regulation can further be divided into three fundamental sub-functions: self-observation (tracking and monitoring one's own behavior, causes and effects of such behavior), judgment (comparing one's behavior to personal standards), and self-reaction (changing one's behavior considering the outcomes of their behavior). Our interventions support self-reflection by assisting in sub-functions of self-regulation, which we detail in Section 3.2.

Goal setting is a prevalent approach in behavioral change systems [57] and is widely used as an intervention strategy in various areas ranging from physical [54, 63, 65] and mental health [73] to digital wellbeing [1, 28, 62]. According to the goal setting theory (GST) by Locke and Latham [42–44], for the goals to have an effect, (1) goals should be accepted by users, and (2) feedback on the progress of the goal should be provided. GST also states that the efficacy of the goal is determined by two key aspects: (1) difficulty and (2) specificity. More specific and harder goals are known to be more effective than less specific and easy ones. Hence, as one of the intervention components, we incorporate the goal-setting theory in the context of daily time limit setting.

The last design component of our intervention system is restriction, leveraging interaction friction. Many existing self-regulation mechanisms [28, 29, 31, 32] limit smartphone/app use to enhance digital wellbeing. Our restrictive techniques are based on the concept of interaction design frictions [11]. Design friction is typically known as a boundary or hindrance that a user may encounter while interacting with the technology. This means that design frictions should be minimized for successful interaction design [74] that facilitates higher user engagement. On the contrary, such design frictions can be utilized to reduce accidental human error [24] or promote positive behavioral changes [11, 64]. For instance, lockout tasks, one type of interaction design friction, facilitated the checking behaviors in data-entry tasks [19] and discouraged smartphone app use attempts [29, 72]. We thus incorporate such design interaction friction into the restriction

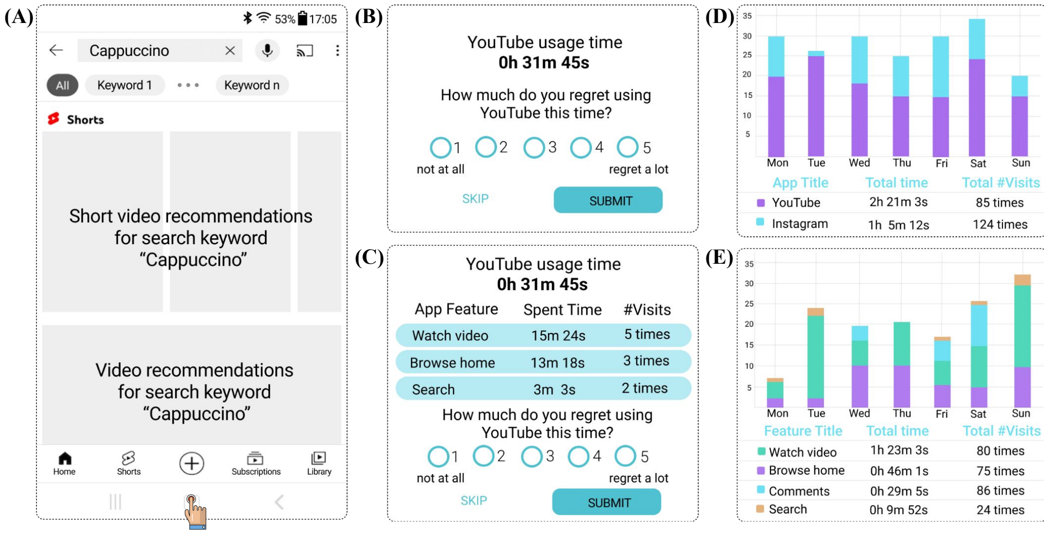


Fig. 1. Self-reflection. A user ends the current app session (e.g., by clicking the Home button) (A). Guided instant self-reflection samples a user's in-the-moment experience in terms of regret at app-level (B) and feature-level (C). Self-reflection on daily use: accumulated social media usage (time spent and visit frequency) visualizations at app-level (D) and feature-level (E).

component of our intervention system when a user exceeds the daily time limit set for a specific app or app feature.

3.2 System Design Components

3.2.1 Guided instant self-reflection. As part of self-reflection, guided instant self-reflection allows users to view their usage statistics and evaluate their experience/behavior. This happens after finishing an app session by rating their usage regret level on a Likert scale from 1 (not at all) to 5 (regret a lot) as shown in Figure 1(B-C). Similar to Finesse [7], we use 'regret' as a proxy to evaluate users' own behaviors. The authors provided theoretical backgrounds from psychology and behavioral economics about the role of 'regret' on user behavior. The ending of an episode of mobile interaction, in our case the end of an app session, was found to be effective in delivering notifications [17]. Moreover, this way of *instant* self-reflection allows us to sample a more reliable user experience as users' app usage experience is still fresh. Guided instant self-reflection mainly reflects the two sub-processes of the self-regulation theory of SCT: self-observation and judgement [4]. By viewing information on app usage (app-level) and app feature usage (feature-level), users can monitor and observe how they spend their time. By assessing their regret level, users can evaluate their behavior (e.g., whether they spent more time on unintended apps or app features).

To lower a user's burden, we apply the probability-based experience sampling method (ESM) [7] instead of prompting a user to reflect on their usage on every target app usage instance. This sampling policy takes the app session duration and time of day into account. The app session duration can be one of three categories: (1) less than 30 seconds, (2) over 30 seconds and less than five minutes, and (3) over five minutes. Thirty-second and five-minute durations are known as short [53] and average app session [25] duration for social media app uses correspondingly. In addition, the use time of the day could be one of the eight three-hour bins in a day, such as from 0 AM to 3 AM, 3 AM to 6 AM, and up to 9 PM to 12 AM. Considering this, the probability-based

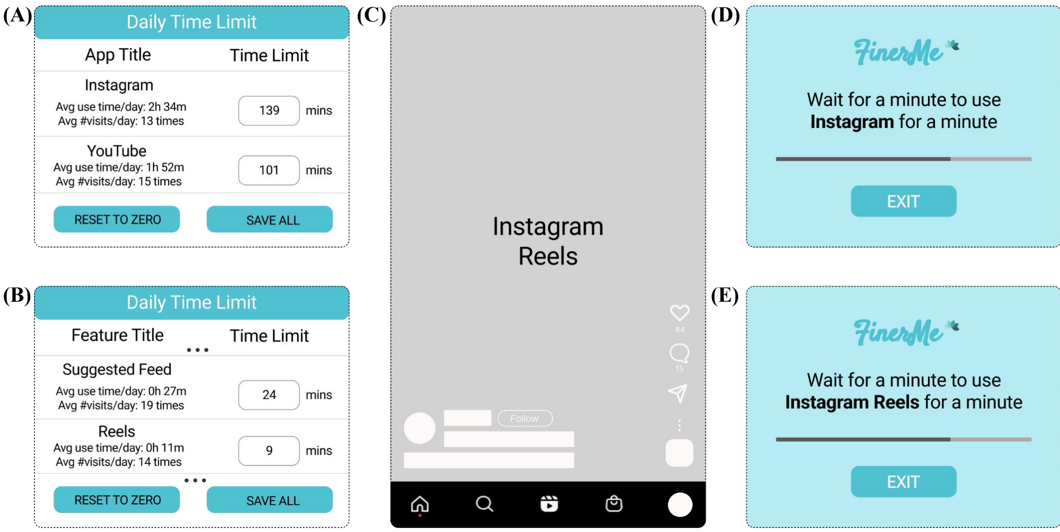


Fig. 2. Daily time limit goal setting at app-level (A) and feature-level (B). A user is watching Instagram *Reels* during Phase 3 (C). Restricting a user in the interaction friction window at app-level (D) and feature-level (E).

ESM prompts users to reflect on their usage at most once per session duration per app, and per time of the day with 50% probability. For instance, a user who has already been asked to reflect on an Instagram session lasting more than 5 minutes from 6 PM to 9 PM will not be prompted again for Instagram sessions lasting more than 5 minutes during the same time frame.

Both app-level and feature-level guided instant self-reflection components display the total time spent for the recently executed app session as shown in Figure 1. The feature-level *FinerMe* additionally shows app-feature statistics (spent time, visit frequency) in a vertically scrollable view (Figure 1(C)). App features are sorted in decreasing order according to the time spent. By displaying additional information on app feature usage, we want to observe whether such feature-level information impacts users' social media usage patterns and regret levels.

3.2.2 Self-reflection on daily use. As the second type of self-reflection, *FinerMe* provides visualizations of aggregated social media use statistics over different days as shown in Figure 1(D-E). This daily use-based self-reflection strategy is grounded in self-observation, judgment (although we do not log the judgment data), and self-reaction from the self-regulation theory of SCT [4]. Self-reflection on daily use is different from instant self-reflection per usage session, which only shows information about the previous app session and might contain limited information for a user to compare their use patterns in previous days and/or weeks. Therefore, to complement such limit of the instant self-reflection, the self-reflection on daily usage statistics portrays one's app usage patterns over a period (i.e., daily, weekly, or monthly) so that users can track usage patterns over time and potentially adjust their behavior after evaluating their progress (self-reaction). App-level cumulative use presents the total time spent and visit frequency for target YouTube and Instagram applications on a daily, weekly, and monthly basis in stacked bar charts with distinctive colors for two apps. Feature-level cumulative use shows similar statistics but at feature level with colors indicating different features. By providing such statistics on the app and app feature usage, we examine (1) if users will detect any new usage patterns, (2) if such information would impact their social media usage behavior, and whether (1) and (2) are different based on the granularity of the usage statistics provided.

3.2.3 Time limit setting. Daily time limit goal setting is a prerequisite step before applying restrictive interventions. In this step, users set daily time limit goals for target apps (i.e., *app-level* intervention group) and app features (i.e., *feature-level* intervention group), as shown in Figure 2 (A-B). As we discussed in Section 3.1, one of the ways to make goals effective is to be accepted by a user. If the system sets a very tight goal, it can cause annoyance and frustration due to the inability to achieve a pre-defined goal. On the other hand, setting a loose goal is considered less effective [57]. To help a user set an achievable goal, we set the default time limit goal to be equal to 90% of the average daily app and feature usage collected during the baseline and self-reflection phases for app-level and feature-level intervention groups, respectively. A 10–20% reduction in use was recommended in the literature [28] as a starting goal, and we thus established an initial daily usage limit goal equal to 90% of the usage collected over the previous ten days. In addition to the average daily usage, we show the daily visit frequency of an app or app feature when a user wants to reduce usage based on frequency. We designed a goal-setting component at the app level and feature level to observe the differences in users' goal-setting behaviors and preferences.

3.2.4 Restrictive interventions. As discussed in Section 3.1, we consider interaction design friction as a restrictive intervention when a user exceeds the daily time limit for an app or app feature. One option would be to entirely block an app or app feature for the rest of the day. However, prior studies warned that strong restrictive interventions, despite being more effective than non-restrictive or weakly restrictive mechanisms, resulted in considerable annoyance and frustration [28]. To avoid such consequences, we introduce restrictive intervention with a relaxed condition. To be precise, instead of imposing a complete usage restriction, we design a *'friction window'* that restricts interaction with a target app or an app feature during a one-minute period once the user exceeds the daily time limit (Figure 2 (D-E)). The one-minute interval was chosen empirically based on the feasibility study we conducted prior to a formal study. By posing friction while interacting with target apps, we investigate the differences at the app level and feature level in terms of goal (time) limit-setting and usage behavior when users encounter restrictions at different granularities.

3.3 Implementation and Design Variations

We implemented *FinerMe* to work with two of the most popular social media applications (YouTube and Instagram) where the study was performed [45]. To implement *FinerMe*, we used the Android Accessibility API [13]. Similar to Finesse [7], we implement an accessibility service that is triggered on scroll, click, focus, window changed, and window state changed UI events taken by a user. To detect app features, we leverage the uiautomatorviewer [14] tool by Android to extract the resource ID, content description, bounds, and text information on target applications defined by the app developers. We do not use any private information related to the context of the information the user consumes, such as video content, video content description, or the content of the message the user is sending or receiving. *FinerMe* detects and tracks the usage of 14 features for Instagram and 11 features for YouTube (see Appendix A).

3.3.1 App-level FinerMe. An app-level version of *FinerMe* intervenes with the participants at a coarse-grained level to help them regulate social media usage. Feedback related only to whole app usage and visit frequency is provided for interventions based on self-reflection. When users set daily time limits, the app-level *FinerMe* guides them by setting the default time limit for each app as 90% of their average daily use. Additional information on daily average visit frequency is provided if users want to reduce the usage of frequently visited apps. After setting a daily time limit,

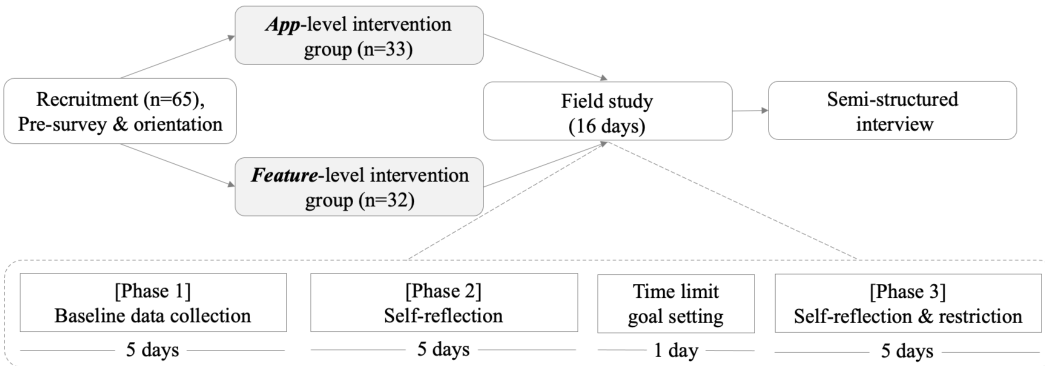


Fig. 3. Flowchart of the study procedure.

restrictive interventions in the form of a friction window are applied. When participants exceed the daily time limit for a specific target app, they are restricted by an overlay window covering the whole screen. When users feel the need to use the app more (e.g., for utilitarian purposes), they can use it for one more minute after they wait on the friction window for one minute.

3.3.2 Feature-level FinerMe. A feature-level version of *FinerMe* implements interventions at a feature-level granularity. Social media usage statistics (time spent, visit frequency) are shown for each app feature as opposed to the app level, where only app usage statistics are shown. For instance, for the guided instant self-reflection, together with the time spent for the previous app session, users are provided duration and frequency information about each app feature they visited as in Figure 1(B-C). Similarly, for self-reflection on daily use, feature-level *FinerMe* provides statistics based on app feature usage instead of app usage. There are separate tabs for Instagram and YouTube so that the users can switch to corresponding tabs to obtain more in-depth information on feature usage. During the time limit setting period, users set a time limit on specific features based on the statistics instead of setting a time limit on the applications. Similarly to app-level *FinerMe*, we set 90% of the average usage for each app feature as the initial default value. Participants could reduce the default value if they wanted to use less. Once users exceed the daily limit for a specific app feature, the friction window intervenes them (Figure 2(D-E)). They can wait on that overlay page to continue using the app feature for one minute (as in app-level intervention). We also allow transition time among features to prevent accidental blocking on unintended feature use.

4 USER STUDY METHODS

We conducted an in-the-wild deployment study with 56 Android smartphone users over 16 days followed by semi-structured in-depth interviews to examine how regulatory smartphone interventions of different granularity (*i.e., app-level* and *feature-level*) affect participants' social media use behavior. Our study was approved by the university's Institutional Review Board (IRB).

4.1 Study Procedure

Our study consisted of three parts: (1) a pre-survey and orientation, (2) a 16-day *FinerMe* field study, and (3) an in-depth exit interview, as shown in Figure 3. After the orientation, we divided participants into two intervention groups (*app-level* and *feature-level*) considering gender balance and the number of SM apps used (e.g., YouTube only, Instagram only, both YouTube and Instagram).

We compensated participants with approximately USD 42 for the 16-day user study and an additional USD 8 if they participated in the semi-structured post-interviews.

4.1.1 Pre-survey and Orientation. Along with the recruitment announcement, we asked participants to answer a short pre-survey to (1) exclude participants who did not meet our inclusion criteria (detailed in Section 4.2) and (2) understand the motivation for participation in the study. We organized three online Zoom [77] orientation sessions on different days to explain the experiment process, the type of data we collect, how to install and use *FinerMe*, and compensation criteria.

4.1.2 Field Study and Design Rationale. We designed our 16-day field study that consists of three main phases each lasting for five days: (1) Phase 1: Baseline data collection, (2) Phase 2: intervention based on self-reflection, (3) Phase 3: intervention based on self-reflection and restriction; and one intermediate step, a one-day period for time-limit setting between Phase 2 and Phase 3 (Figure 3). Note that we allowed the participants to set the time limit goal only once, during the one-day interval between Phase 2 and Phase 3. Participants were not allowed to modify their daily time limits during Phase 3. As described in Section 3.3, the feature tracker of *FinerMe* leverages the UI layout and accessibility node information (e.g., resource ID, class name, combination of components) on the target social media apps to detect app features. This implies that even minor updates in target apps (YouTube and/or Instagram) could result in inaccurately detected features. Therefore, we designed a short-term (16-day) experiment to minimize potential log errors from the app feature detection algorithm. We allocated the same number of days (five days) for each experiment phase to see the difference in usage patterns. We collected usage logs of target applications during the first five days without any intervention (Phase 1). The majority of the participants started the study on Saturday to balance the number of weekdays and weekends in the five-day period in each phase.

We collected usage time and visit frequency for each application and feature for both intervention groups. We then applied non-restrictive interventions based on self-reflection for the next five days. We proactively prompted users to rate their regret level after finishing a target application session to sample participants' instant self-reflection. Similar to Finesse [7], to minimize users' effort and prevent careless responses, we showed an instant self-reflection prompt based on sampling policy rather than showing it on every app usage session. For self-reflection on daily use, we provided statistics on total usage time and visit frequency by visualizing them. This information was accessible throughout the rest of the experiment period. We then gave users one-day intervals to set the time limit. A one-day interval was chosen to provide enough time for participants to observe their usage patterns collected for the previous ten days and set the daily time limit goals during the available time of that day. Participants could change the limit during this one-day period, but only to lower than their initial default value. After the one-day period of time limit setting, we applied restrictive interventions on top of the self-reflection, and this part of our experiment also lasted for five days.

4.1.3 Semi-Structured Interview. After the field study, we conducted one-on-one semi-structured interviews with participants. The interview participation was optional. Most of the participants who lived on campus attended the interview onsite, while we used Zoom [77] to interview others. Each interview lasted 30 minutes to one hour, during which we asked participants about their overall experience with *FinerMe* in each phase of the field study and in what aspects of our intervention system they liked or disliked in terms of regulating social media use. For the users in the feature-level intervention group, we asked how and why intervening at the feature level was or was not helpful.

4.2 Participants

We posted a call for participation on large university community forums. We recruited 65 participants (26 females and 39 males, age: M=23.01, SD=3.12) through the screening process based on our criteria: individuals who (1) use an Android smartphone as their primary phone, (2) use at least one of the target social media applications (YouTube, Instagram) on a daily basis, and (3) are motivated to reduce social media app usage time. The group consisted of 62 university students, one professor, and two unemployed people.

We divided participants into app-level and feature-level intervention groups, considering balance in gender, age, and application usage. From the initial participants, usage logs from four participants were discarded due to unstable data collection. Another two participants were excluded from the study as they were using both Android and iOS smartphones simultaneously, and we could not verify if the Android phone was their primary device. Lastly, we removed the data of three participants, as they dropped out during Phase 3 of the experiment due to their schedule. As a result, we used the social media usage logs from the remaining 56 participants (31 in app-level and 25 in feature-level intervention groups). After the field study, we conducted in-depth semi-structured interviews with 36 participants to corroborate our quantitative findings.

4.3 Method of Data Analysis

We analyzed the participants' social media usage logs, ESM responses from guided instant self-reflection prompts to rate participants' regret levels immediately after finishing an app session, and transcriptions from the semi-structured exit interviews.

4.3.1 SM Use Time Analysis by Intervention Groups. We inspected how the mean *daily* social media usage time changed over three different phases of our experiment, as time spent is one of the objective measurements to evaluate the intervention system. First, we verified that the two intervention groups did not statistically differ from each other using Mann-Whitney U Test [49], an alternative non-parametric test to an independent t-test. Next, to investigate the differences in usage time between groups and the type of interventions, we conducted an analysis of variance using a Generalized Linear Mixed Model (GLMM) [70] with two factors: intervention granularity (two levels: app-level, feature-level) (between subject, fixed effect) and intervention type (three levels: no intervention, self-reflection, restrictive interventions) (within-subject, random effect). The interaction effect between intervention granularity and intervention type was also considered when modeling GLMM. We modeled GLMM in R [66] using Gamma regression with a Log link function. Social media use time had skewed distributions, and thus, log-scale transformation was used to fit them into the gamma distributions.

4.3.2 SM Use Time Analysis by Intervention Type. To further understand the within-subject factor differences, we conducted the Friedman test, a non-parametric equivalent of one-way repeated measures ANOVA [70]. A significant Friedman test result is further analyzed by conducting post-hoc Conover pairwise tests [10] with Bonferroni correction to compensate for the multiple comparisons and avoid false positives. We conducted Friedman tests and post-hoc analyses for both overall social media (Instagram + YouTube) usage and app-specific (YouTube only, Instagram only) usage. After performing statistical tests on overall usage, we then focus on app-specific usage analyses to consider differences in usage behaviors.

One of the inclusion criteria to participate in our study was being an active user of at least one of the target social media applications. Therefore, if a participant was not an active user of one of the two applications, we excluded the usage of a non-active app from the data analysis. For instance, if

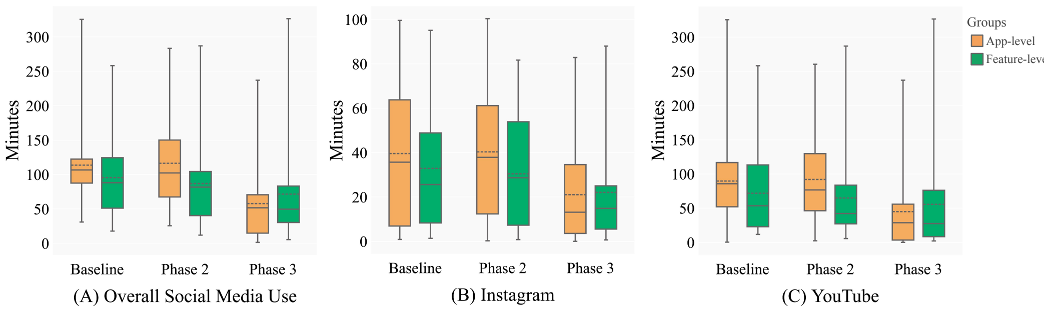


Fig. 4. Daily average social media use time for app-level (orange), and feature-level (green) intervention groups over three phases. (A) Overall social media use (Instagram + YouTube), (B) Instagram only, and (C) YouTube only. Phase 1 (Baseline): no intervention, Phase 2: interventions based on self-reflection, and Phase 3: interventions based on self-reflection and restriction.

a participant is an active user of YouTube, but only visited Instagram two or three times during the entire experiment period, we only considered the YouTube app use for our data analysis.

4.3.3 App Feature Use Time Data Analysis. We hypothesized that there will be usage time changes not only for overall social media or app-specific use but also for app feature use time. Therefore, we conducted statistical tests for app feature use time dynamics. We selected the top five app features with the highest use times during the Baseline collection period for both apps and tested whether there were any changes in daily use time as a result of our interventions.[1] Similar to app use time analyses, we conducted Friedman tests to measure the changes in app feature use time and post-hoc Conover pairwise tests with Bonferroni correction to understand the relative difference between phases.

4.3.4 Regret ESM Data Testing. We analyzed the regret ESM responses of users to examine whether their regret was related to the change in use time. We calculated the mean regret score per user in each condition and measured the correlation between use time and the mean regret score. We conducted a Spearman's rank correlation test [75] as it does not make any assumptions on data distribution.

4.3.5 Interview Data Analysis. Before analyzing the interview data, we transcribed all audio recordings of 36 interviews into text. We then conducted a thematic analysis [6] on the transcribed data. We analyzed the interview transcripts to answer (1) how users experienced our intervention system, *FinerMe*, (2) what users liked or disliked about it, and (3) whether app feature information was helpful to regulate social media use for the feature-level intervention group.

5 RESULTS

We present the results of social media app usage dynamics analysis to answer our main research question: *How do app-level and feature-level interventions affect regulated social media use?* We also describe the analysis results of qualitative data extracted from follow-up in-depth interviews.

5.1 Use Time by Intervention Granularity and Intervention Type

We verified that there was no statistically significant difference in the daily SM use time between the two intervention groups during the Baseline period (Phase 1) (Mann-Whitney U=453, n_1=31,

[1]The following are the top five app features for YouTube (i.e., *comments*, *browse home*, *search*, *watch shorts*, and *watch videos*), and for Instagram (*direct messaging*, *following feed*, *watch reels*, *suggested feed*, and *view story*).

	App-level				Feature-level			
		Post-hoc Conover				Post-hoc Conover		
App	Friedman	Ph1 & Ph2	Ph2 & Ph3	Ph1 & Ph3	Friedman	Ph1 & Ph2	Ph2 & Ph3	Ph1 & Ph3
Overall	χ^2=32.71, p<.001***	p=1.00	p<.001***	p<.001***	χ^2=12.48, p=.002**	p=1.00	p=.042*	p=.004**
Instagram	χ^2= 9.24, p=.010**	p=1.00	p=.017*	p=.110	χ^2= 4.11, p=.128	p=1.00	p=.310	p=.230
YouTube	χ^2=26.60, p<.001***	p=1.00	p<.001***	p<.001***	χ^2=10.64, p=.005**	p=1.00	p=.060	p=.009**

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table 1. Summary of statistical findings for daily social media app usage. Friedman test was followed by post-hoc Conover with Bonferroni adjustment. Ph1, Ph2, and Ph3 stand for Phase 1 (Baseline), Phase 2 (Self-reflection), and Phase 3 (Self-reflection and restrictive interventions) respectively.

n_2=25, p=0.3, two-tailed). We performed a GLMM test to analyze the changes in daily SM use time with respect to the intervention granularity (app-level and feature-level) and intervention type (no intervention, self-reflection, self-reflection combined with restriction). We found that the effect of intervention granularity was not significant (Wald $\chi^2_{(1,N=56)}$=0.36, p=.55), whereas the effect of intervention type (Wald $\chi^2_{(2,N=56)}$=58.23, p<.001) and the interaction effect between intervention granularity and intervention type (Wald $\chi^2_{(2,N=56)}$=10.61, p=.005) were significant. In Table 1, we summarize the main statistical findings regarding app use time.

5.2 Overall Use Time Reduction by Intervention Type

We analyzed the SM use time based on the intervention type, by examining (1) the reduction of both YouTube and Instagram use time, and (2) the contribution of each intervention phase to the use time reduction. Daily mean use time, although not significant, increased slightly for app-level participants during Phase 2 (self-reflection only) compared with the Baseline (Phase 1). It then decreased dramatically in Phase 3 (self-reflection and restriction) as shown in Figure 4A; (Baseline: M=113.6, SD=61.1; Phase 2: M=116.3, SD=68.7; Phase 3: M=57.5, SD=58.1). Participants in the feature-level group showed a consistent downward trend over the experiment period: (Baseline: M=95.5, SD=59.8; Phase 2: M=86.8, SD=58.2; Phase 3: M=71.5, SD=70.3).

5.2.1 Self-reflection with restriction (Phase 3) reduced use time. The results of a Friedman test shown in Table 1 revealed that the mean daily time use on target apps was significantly reduced in both groups (app-level: $\chi^2_{(2,N=31)}$=32.71, p <.001; feature-level: $\chi^2_{(2,N=25)}$=12.48, p=.002). Post-hoc comparisons using the Conover test with Bonferroni adjustment revealed a significant difference between Baseline and Phase 3 (app-level: p<.001; feature-level: p=.004), and between Phase 2 and Phase 3 (app-level: p<.001; feature-level p=.042), but non-significant differences between Baseline and Phase 2 for both groups. These results demonstrate that the use time reduction was most prominent when we use both self-reflection and restriction.

Participants' responses in the exit interviews agreed with the findings. A common theme was that the *friction window* introduced in Phase 3 helped the participants to self-regulate app usage as they used target apps only when they had something important to do. P45 noted, *"I waited one minute [of a friction window] only if I had to use [a target app]; in other cases, I quit."* P2 also stated, *"It gives you time to wait and see and then think about whether you really need it"*. Interaction friction was also helpful in avoiding habitual checking (P3), and the act of waiting to continue using a target app made participants feel ashamed due to a lack of self-control (P66) and assisted them to promptly quit a target app, *"I almost did not wait for an interaction friction window. It makes me ashamed... So when I saw the blue [friction] window, I thought to myself 'I have to quit'."*

5.2.2 Self-reflection (Phase 2) increased awareness, but failed to reduce use time. During the self-reflection period (Phase 2), none of the intervention groups significantly reduced usage time. Our qualitative analysis results showed that self-reflection helped participants to become more aware of their usage patterns. P1 said, *"Yeah, I think I was more conscious of my usage of the two [apps]."* Compared with app-level users, participants in the feature-level group obtained more information about their use patterns. For instance, P38 stated, *"I found out that I watch YouTube 'shorts' more than I thought, and I also [view] 'comments' when watching 'video'."* P64 added, *"I was surprised [to find out] that I 'browse home' so much on YouTube."*

Although self-reflection increased awareness of usage patterns, it alone was not enough to reduce use time. P41 mentioned, *"I don't think it can have a great impact on the behavior of me watching YouTube or Instagram unless it restricts me in some way. It just makes me aware, but I don't think I am going to do something different based on those statistics."* Another interesting note from the interview was that such non-restrictive interventions based on self-reflection might not be strong enough to progress into an action or behavior change because of their already-formed habits. P44 explained, *"To be honest, it is hard to change my behavior by just seeing how much I spent on which parts, ... because of my past habit."*

5.3 Use Time Reduction by Target Applications

Instagram and YouTube provide distinct functionalities. Users visit them with different motivations and spend time differently. We inspected if the effect of intervention methods varied by target apps.

5.3.1 Instagram use time reduction was different by intervention groups. As shown in Figure 4B, time spent on Instagram was generally shorter than that on YouTube: daily mean use time logged during Baseline data collection was 39.59 minutes (SD=32.29) for app-level and 32.89 minutes (SD=28.55) for feature-level intervention groups. In terms of usage time reduction, we observed an interesting pattern (see Table 1): app-level group reduced Instagram's daily mean use time significantly [$\chi^2(2)$=9.24, p=.01], while feature-level did not [$\chi^2(2)$=4.11, p=0.128] despite a downward trend in the daily mean use time. Post-hoc Conover test results for the app-level group showed a significant difference in mean Instagram usage time between Phase 2 and Phase 3 (p=.017), and a non-significant difference between Baseline and Phase 2, and Baseline and Phase 3.

Although both Instagram and YouTube have a mixture of utilitarian and hedonistic features [7], Instagram has more utilitarian features (e.g., *direct messaging, upload story, in-app web view, search*) than YouTube (e.g., *search*). The interview results revealed that the feature-level intervention users encountered the interaction friction window more frequently because a restriction was applied for each app feature individually. This means that when they encountered the interaction friction window, they waited more frequently to use essential/utilitarian app features such as *upload story* (P51), *direct messaging* (P39), or answer the survey questions uploaded to Instagram (*in app web view*) (P50). As a result, feature-level users tended to spend more time on Instagram as a whole compared with app-level users. P36 described this situation as follows: *"When I was just looking at the feed and when [the friction window] came up, I got annoyed to wait a minute and just quit [the app]. But [in a situation] when I had to send an Instagram DM, the [friction] window popped up, and I had to wait to send a message. I sent a message [after waiting]."* On the other hand, as we applied restrictions for the Instagram app as a whole (not for each individual feature) for app-level users, the usage of essential features was not restricted by using other features, thus leading to a significant reduction in the overall Instagram use time.

5.3.2 YouTube use time reduction follows the same pattern as overall use time reduction. Daily mean use time spent on YouTube app during the Baseline period was approximately twice that of Instagram use time for both groups: 89.68 minutes (SD = 68.42) for app-level and 71.80 (SD = 63.68) for feature-level groups (Figure 4C). Unlike Instagram, the results of the Friedman test

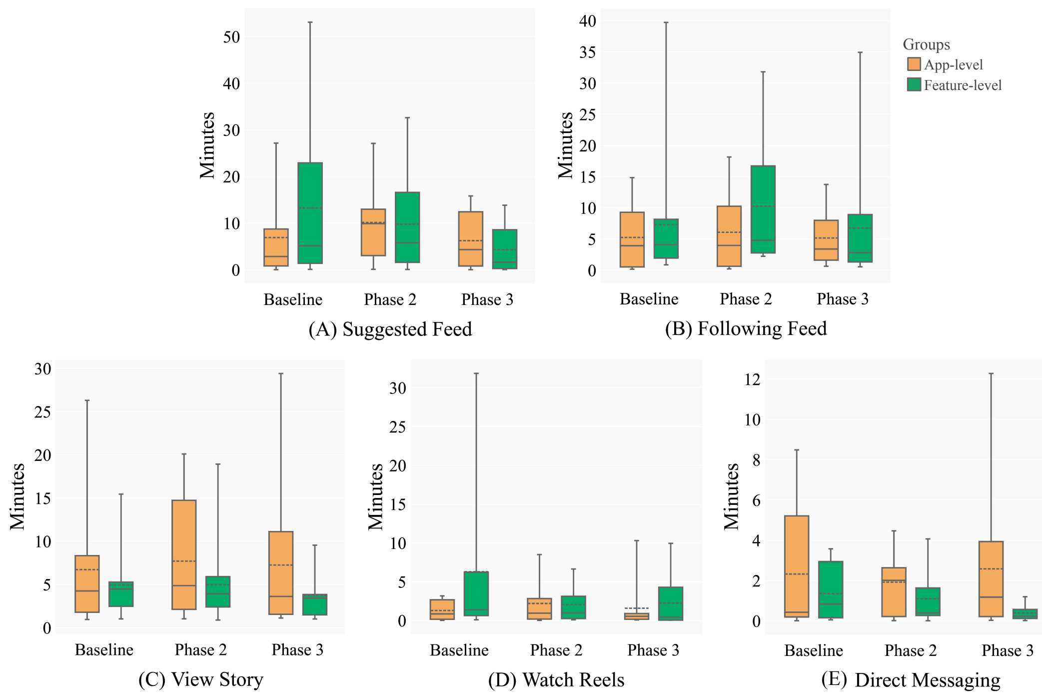


Fig. 5. Top five Instagram feature use time for app-level (orange), and feature-level (green) intervention groups. (1) Four passive features: (A) Suggested feed, (B) Following feed, (C) View story, (D) Watch reels; (2) One active feature: (E) Direct Messaging.

indicated a statistically significant reduction in YouTube use time for both groups as shown in Table 1 (app-level: $\chi^2(2) = 26.6$, $p < .001$; feature-level: $\chi^2(2) = 10.64$, $p = .005$). Post-hoc Conover test results showed that the main contribution comes from Phase 3, as we found significant differences between Baseline and Phase 3 (app-level: $p < .001$, feature-level: $p = .009$), and between Phase 2 and Phase 3 (app-level: $p < .001$, feature-level: $p = .06$).

YouTube usage reduction was more significant than Instagram use time reduction. One main reason behind this was that the number of features that users deem essential is fewer on YouTube than that on Instagram. For instance, participants reported waiting on the interaction friction window when they had to use *direct message*, *upload post/story*, *search*, *in app web* features on Instagram. However, participants rarely mentioned the features they consider essential except for *watch videos* of important content (e.g., lecture, tutorial, and workout). For those important use cases, several participants admitted that they instantly accessed YouTube from other devices (e.g., desktops, laptops, and iPads). For instance, P44 used YouTube from other devices, saying *"I had times when I accessed YouTube via the browser when my professor uploaded lectures, or when I really needed to look up some information. But I couldn't do so for Instagram."*

5.4 Impact on App Feature Use Behaviors

As discussed in Section 2, social media apps typically contain a mixture of features that users deem essential or addictive. Moreover, social media use can also be categorized into *active* and *passive use*. In fact, Verduyn and colleagues [68] found that passive social media use (e.g., passive consumption of media contents) is negatively correlated with digital wellbeing, while active social

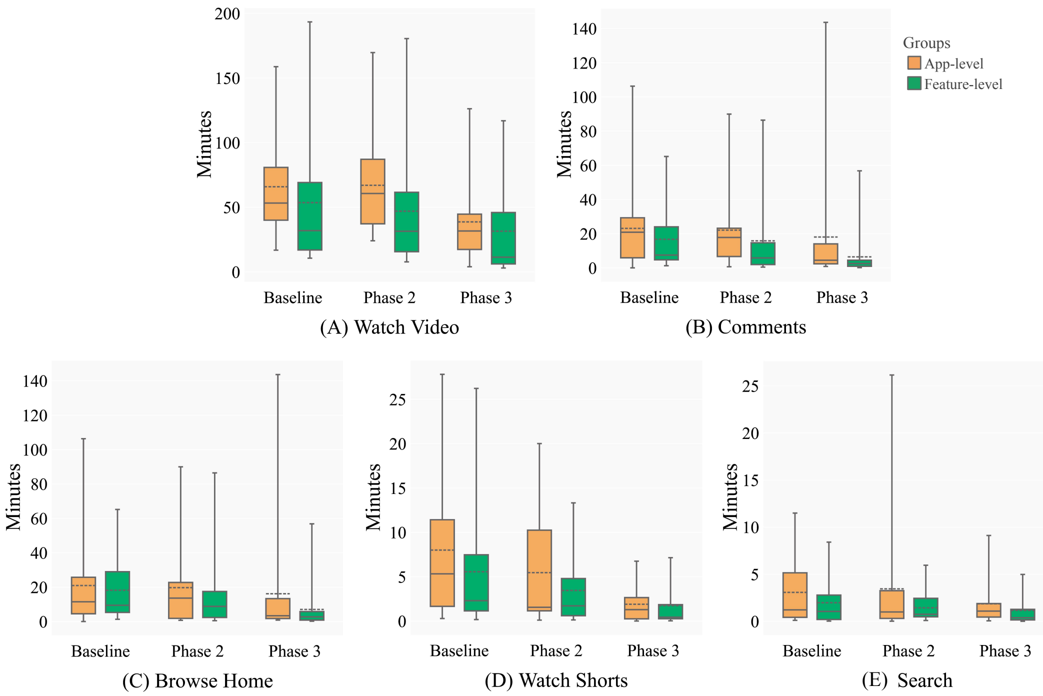


Fig. 6. Top five YouTube feature use time for app-level (orange), and feature-level (green) intervention groups. (1) Four passive features: (A) Watch Video, (B) Comments, (C) Browse Home, (D) Watch Shorts; (2) One active feature: (E) Search.

media use (e.g., direct messaging) was found to have a positive impact on one's digital wellbeing. App features, in turn, can also be associated with active or passive social media use. For instance, Instagram features such as *direct messaging*, *upload story*, *upload post*, and *search* can be considered active features as they involve active decision makings and user actions, while *following feed*, *suggested feed*, *view story*, *watch reels*, *other's feed*, *view by hashtag*, *in app web view*, *notifications*, and *shopping* can be regarded as passive features as these features require only passively viewing the action from a user. Similarly, YouTube's *search* feature is an active feature, while all remaining features can be viewed as passive features: *watch video*, *comments*, *browse home*, *watch shorts*, *channel*, *subscriptions*, *playlists*, *explore*, *my library*.

By logging app feature use patterns for both intervention groups, we examined how app-level and feature-level interventions affected app feature use behaviors. We analyzed app feature use behaviors for the top five features of both YouTube and Instagram apps based on the daily mean use time spent during the Baseline period. Interestingly, the top four features with the highest time spent for both apps (*suggested feed*, *following feed*, *view story*, *watch reels* on Instagram, and *watch video*, *comments*, *browse home*, *watch shorts* on YouTube) were passive app features, and the fifth most used app features (*direct messaging* on Instagram and *search* on YouTube) were the only active features. We summarize the important statistical findings related to feature use in Table 2.

5.4.1 Feature-level intervention users reduced the usage of passive app features for both target applications compared with app-level intervention users. For **Instagram**, the top four features with the highest daily mean use time were passive features: *suggested feed* (M=8.45, SD=10.20), *following* *feed* (M=6.86, SD=8.03), *view story* (M=5.76, SD=5.94), and *watch reels* (M=2.60, SD=5.17) as seen in Figure 5A-D. In fact, the first three of these four app features were reported to be the top three most 'regretful' Instagram features [7]. The results of the Friedman test showed a significant reduction for all these three features for participants in the feature-level intervention group, while app-level users reduced their use time significantly for only the *suggested feed* feature as shown in Table 2. A phase-level analysis revealed that the contribution was primarily from Phase 3 (when we applied interventions based on self-reflection and restriction together). Both app-level and feature-level users showed a marginal difference for *watch reels*.

In **YouTube**, as expected, the most used app feature with the daily mean use time during the Baseline collection period was *watch video* (M=49.79, SD=43.36). It was followed by *comments* (M=16.57, SD=24.25), *browse home* (M=6.47, SD=6.67), and *watch shorts* (M=4.31, SD=5.86) as illustrated in Figure 6. Here again, the first three features were considered the top three most regretful YouTube features. Commenting is considered a type of active social media use according to the literature [68]. However, we found out that in the case of YouTube, participants were mostly "viewing" comments instead of actually "typing" comments. Moreover, *comments* had the highest individual regret ratio among all YouTube features in a previous study [7]. Therefore, we refer to the *comments* on YouTube as a passive feature.

Both app-level and feature-level intervention groups reduced the amount of time spent *watching videos* as the Friedman test indicated a significant difference as seen in Table 2.

Only feature-level users reduced the time spent on *comments* significantly, according to the outcome of the Friedman test. Feature-level interventions also tended to be effective in reducing the time spent *browsing home*, another passive type of YouTube feature that merely displays the list of videos based on the YouTube algorithm. For *watch shorts*, app-level users reduced the use time significantly, and feature-level users showed a marginally significant reduction. Similar to Instagram, post-hoc analysis showed that self-reflection and restriction (Phase 3) primarily contributed to a significant reduction in passive feature usage.

Overall, these results collectively revealed that the app features associated with passive social media use tended to be better regulated in the feature-level group as opposed to the app-level group. We analyzed the qualitative data to understand why and how feature-level users reduced features associated with passive SM use. The thematic analysis revealed that participants intentionally tried to reduce their time spent on these features after becoming aware of their usage patterns. P62 said, *"I did not know that I would spend so much time on 'viewing comments'. But [*FinerMe*] lets me know how much I spend on '[viewing] comments' or '[watching] shorts'. [After recognizing my pattern], I intentionally exited the comments window when I was watching playlists, reminding myself, 'Oh, I viewed comments a lot.'"* P50 also noted the impact of feature-level information on her behavior: *"I became more conscious of using certain features that I use often. I think I noticed such changes in my use behavior."* Another theme we discovered is related to goal setting. Although this theme is not applicable to all feature-level participants, some participants mentioned setting shorter time goal limits for features with the highest use time. P44 stated, *"My default upper limit for 'comments' was around 60 minutes. But I thought 50 minutes would be enough, so I changed my goal to 50 minutes."*.

5.4.2 Use of active features remains similar in both intervention groups. The fifth most used app feature for both applications was active use: *direct messaging* (M=1.55, SD=2.24) on Instagram (Figure 5E), and *search* (M=2.12, SD=3.49) on YouTube (Figure 6E). Interestingly, neither app- nor feature-level users showed any statistically significant difference in daily mean use time of active features for both target apps (Table 2). This again verifies that by providing fine-grained control over the app components, we can support users to reduce the usage of unwanted passive features, while maintaining the same use amount for the active/essential app features.

		App-level				Feature-level			
			Post-hoc Conover				Post-hoc Conover		
App	Features	Friedman	Ph1 & Ph2	Ph2 & Ph3	Ph1 & Ph3	Friedman	Ph1 & Ph2	Ph2 & Ph3	Ph1 & Ph3
Instagram	Suggested Feed	χ^2= 6.73, p=.035*	p=.137	p=.088	p=1.00	χ^2=10.17, p=.006**	p=1.00	p=.044*	p=.027*
	Following Feed	χ^2= 0.54, p=.761	p=1.00	p=1.00	p=1.00	χ^2=11.17, p=.004**	p=.105	p=.011*	p=.956
	View Story	χ^2= 0.60, p=.741	p=1.00	p=1.00	p=1.00	χ^2= 6.17, p=.046*	p=1.00	p=.110	p=.160
	Watch Reels	χ^2= 4.67, p=.097	p=.530	p=.150	p=1.00	χ^2= 4.67, p=.097	p=1.00	p=.530	p=.150
	Direct Messaging	χ^2= 0.20, p=.901	p=1.00	p=1.00	p=1.00	χ^2= 2.17, p=.338	p=1.00	p=.500	p=.960
YouTube	Watch Video	χ^2=10.13, p=.006**	p=1.00	p=.020*	p=.049*	χ^2=20.32, p<.001***	p=1.00	p=.002**	p<.001***
	Comments	χ^2= 2.13, p=.344	p=1.00	p=1.00	p=.470	χ^2=24.00, p<.001***	p=.178	p=.018*	p<.001***
	Browse Home	χ^2= 4.59, p=.101	p=.400	p=1.00	p=.140	χ^2= 5.47, p=.065	p=1.00	p=.340	p=.088
	Watch Shorts	χ^2= 5.37, p=.068	p=.503	p=1.00	p=.086	χ^2=13.00, p=.001***	p=.143	p=.428	p=.004**
	Search	χ^2= 0.13, p=.935	p=1.00	p=1.00	p=1.00	χ^2= 4.35, p=.113	p=1.00	p=.290	p=.200

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table 2. Summary of statistical findings for daily feature usage. Friedman test was followed by post-hoc Conover with Bonferroni adjustment. Ph1, Ph2, and Ph3 stand for Phase 1 (Baseline), Phase 2 (Self-reflection), and Phase 3 (Self-reflection and restrictive interventions) respectively.

5.5 Intervening at Feature-level: User Preferences and Impact on Behavior

5.5.1 Feature-level restriction increases awareness about usage. As discussed in Section 3.2.1, we used the guided instant self-reflection with probabilistic ESM that allows us to sample the users' experience on how much they felt regretful about the app session they just finished. We compared if there were any differences in regret ESM data over a period for two intervention groups. Using Spearman's rank correlation test, we analyzed the relationship between use time and mean regret score. The results revealed that during Phase 3 (self-reflection and restrictive interventions), the mean regret score was correlated with use time only for participants in the feature-level intervention group: $r(25)$=0.435, p=0.03. However, we did not find any significant correlation for app-level users: $r(31)$=0.051, p=0.78. Accordingly, these results indicate that feature-level self-reflection and restrictive interventions (Phase 3) raised awareness about usage. This finding is consistent with the Decision Justification Theory (DJT) [9] in which regret is linked to awareness of having made a bad or unsatisfactory decision.

5.5.2 Different strokes for different folks: For whom, why do/do not feature-level interventions work? In the follow-up interviews, we asked how and why intervening at the feature level was/was not helpful. The major theme we extracted from the interview results suggested that participants appreciated feature-level information and preferred feature-level intervention as it helped them to realize where exactly they waste time and reduce time spent on unnecessary parts of an app. P39 stated, *"I liked that I could set the time limit by features. There are some useless features... and I can reduce using [those useless] parts"* Similarly, P50 pointed out, *"When I am restricted at app-level, I don't know where I waste my time. I think restricting by features, for this reason, was helpful for me".* Furthermore, participants discovered usage patterns that were different from their expectations. P47 explained, *"I thought I spent most of my time only watching videos on YouTube. But then I realized that I spent time navigating the app a lot. I also spent a lot of time browsing home."*

In contrast, digesting such fine-granular information and micro-controlling target apps felt too burdensome for participants who wanted to simply reduce the overall time spent. P51 mentioned, *"The goal of reducing SNS is to reduce the usage of the app itself, rather than how much I watch 'shorts', or 'video'... That's why I think [intervening at] app-level is enough."*. Moreover, some participants felt burdened by setting time limits for each feature individually. A notable finding is that only 29% of interviewed feature-level participants modified limits on daily feature use compared with 67% of the interviewed app-level participants. P52 explained this situation as follows: *"If I want to set a time limit goal for each app feature individually, I have to think how much I typically spend time on* *this feature, and why I would spend too much time on this feature. It might be difficult to establish good criteria for this limit, which makes me think that setting goals at app-level might be better.".*

6 DISCUSSION

We explored the differences between applying coarse-grained *app-level* and fine-grained *feature-level* smartphone interventions to regulate mobile social media use. The key findings pertaining to the feature-level interventions are: (1) feature-level interventions are more effective in reducing the amount of time spent on app features associated with passive social media use, and (2) feature-level restrictive interventions showed a positive correlation between awareness of regretful usage and the use time reduction. We now discuss four major design implications extracted from our study, limitations, and future work.

6.1 Regulating Social Media Use at Different Granularity

Our work demonstrated that participants in both intervention groups significantly reduced the time spent on social media during the intervention period compared with the baseline. However, for both groups, the mechanisms based merely on self-reflection were insufficient to reduce time, as a significant reduction in use time occurred during Phase 3 when we applied restrictive interventions (interaction friction window) together with self-reflection. Although mechanisms based on self-reflection are widely adopted in existing digital wellbeing tools [67, 69] and behavioral change design [21, 59], our results showed that these types of easily *bypassable* interventions did not have a consequential effect on use time reduction.

Nevertheless, participants who used various features of target applications in the feature-level group appreciated the provided information on app feature usage, and it made them become better aware of their fine-grained usage patterns. Participants liked being informed about their app feature use patterns, as they *knew exactly where they wasted time*, and it made them *use those features more consciously* when they visited a target app. This finding is consistent with prior work [7] that reported that feature-level information on instant self-reflection was beneficial for people who use various features of apps.

We also explored if there were any app-specific differences when we intervened at different granularities. Our results confirmed that people use two popular social media platforms differently, and their use time reduction based on the intervention granularity differed. In general, our participants spent more time on YouTube than Instagram and reduced usage time more in Phase 3 than that in the Baseline. For Instagram, daily mean use time reduction was only significant for app-level users, but there was a non-significant downward trend for feature-level participants. This was because the feature-level restriction is applied for each individual feature, and participants in this group encountered interaction friction windows more frequently than app-level users.

None of the prior studies compared the app feature use behaviors before and during the intervention period when users are intervened at the app- and feature-level. We tracked app feature use behaviors throughout the experiment for both intervention groups. The results showed that feature-level participants tended to use the top four passive features less than app-level participants. Specifically, the use time reduction for *comments* on YouTube and *following feed* on Instagram (top two features with the highest individual regret ratio [7]) were significant for only feature-level participants but not for app-level participants.

Another interesting finding from our study is the relationship between use time reduction and self-evaluation of usage regret. We found a significant correlation between use time reduction and individual regret ratio of only feature-level users. Participants labeled the app sessions as highly regretful if their usage was increased from the previous phase and marked the app sessions as less regretful when their use time was less than in the previous phase. This tendency became prominent during Phase 3 when we applied interventions based on self-reflection and restriction at the same time. This can be explained by the targeted exposure of interaction friction at the feature level. Interaction friction brought an opportunity for self-judgment on the gap between their goals and current achievements on the target features. Frequent encounters helped participants to realize that they were overusing the target features, resulting in highly regretful app sessions. Furthermore, this led them to successfully reduce the regretful usage time of the target features.

6.2 Design Implications

Modern social media apps are equipped with various features and keep users engaged by placing active (e.g., direct messaging) and passive (e.g., suggested feed) features in one place. However, most existing digital wellbeing tools typically intervene at the phone- or app-level, rarely considering the characteristics of features within an app. To bridge this gap, we designed app-level and feature-level interventions and analyzed the impact based on the granularity and type of intervention. We share and discuss key design implications derived from our study.

6.2.1 Reducing user burden of feature management. Our results suggest that the number of features per app should be carefully considered. In our prototype, YouTube and Instagram usage is divided into 11 and 14 unique features, respectively. Therefore, a user who uses both YouTube and Instagram had to manage (set a daily time limit and understand usage in visualizations) 25 features. Some participants felt burdened to go over each feature and set limits one by one. Also, when viewing the usage statistics, some users looked only at the top features sorted by usage time or overall usage in general. There were also features that most users barely used. Focusing on the main features of a target app would ease the burden and still allow users to control usage within the app. Combining similar features or excluding rarely accessed features could be another choice.

To lower such control overload, it is possible to help users to detect the undesirable use of features and then give suggestions to set more personalized interventions such as tighter limits on the features. Our results showed the potential of developing a method that can detect regretful use at a feature level. An automated algorithm can leverage semantic information about each feature as well as the actual content consumed during sessions. Traditional item-based collaborative filtering or user feedback-based learning methods (e.g., multi-armed bandits) can be applied to developing personalized feature recommenders.

6.2.2 Compiling a list of irregular yet essential features. When designing feature-level restrictions, we offered users to set a daily time limit for each pre-defined feature on YouTube and Instagram. To prevent higher limits, we set an upper bound for each feature to be equal to 90% of the average usage of that feature based on data collected during Baseline collection and Phase 2. However, we found that feature usage is irregular for some participants, and the same set of features is not used regularly: e.g., one might upload a story on Instagram once in a few days, but not every day. For such cases, the default time limit was set low, and when the participants wanted to use those essential yet not frequently used features, they were intervened in the friction window and had to wait for a minute to use that feature. To avoid such situations, one could create a list of irregular but essential features and apply different rules to determine the upper bound to set a daily time limit. Thus, it is important to profile feature-level usage patterns and their regret levels to personalize feature selection and goal setting.

6.2.3 Providing flexibility by offering a mixture of app- and feature-level interventions. While we only considered a dichotomous intervention in our experiment, it is possible to design a mixture of app-level and feature-level interventions, which can also be personalized as mentioned earlier. The interview results showed that people have varying preferences (feature-level vs. app-level) concerning how to be intervened and what information they like to see to understand their usage better. Thus, it is important to allow users to select the part of the intervention they want to be fine-grained or coarse-grained. Some users suggested they would like to see their usage by features but restrict it at the app level. This means that some users prefer a mixture of app-level and feature-level interventions based on their needs. For instance, interventions based on self-reflection can be applied at the feature level and restrictions at the app level. Moreover, participants might set app-level restriction goals (e.g., using Instagram for 1 hour per day) as their primary goals, and sub-goals for specific features as a feature-level intervention (e.g., limiting viewing suggested feed to 10 minutes per day within a 1-hour daily limit).

6.2.4 Employing alternative user study designs. Our study design falls under the category of an "active control trial" [3] where the first group receives the treatment (app-level interventions) and the second group also receives similar treatment (feature-level interventions). As our study's primary goal is to compare the impact of feature-level interventions with that of app-level interventions, we decided to leave the control group out of the experiment. Nonetheless, we acknowledge that incorporating a control group into the study design would enhance the depth of our analysis of the impact of the intervention type. Another viable design option would be employing a randomized cross-over trial [23]. This study design would involve participants receiving both app-level and feature-level interventions (with the option of including a control period) in random order. A washout period would follow each intervention method to avoid any potential carryover effects.

6.3 Limitations and Future Work

There are several limitations that could be addressed in future work. First, similar to Cho et al. [7], our feature detection algorithm is rule-based and this approach is not robust against the application updates. We see an opportunity to extend our work to facilitate the detection of features agnostic to apps or app updates. A recent research [71] showed that app screenshots could be used to infer the relationships between UI elements. A follow-up study [16] leveraged deep learning to detect screen similarities and transitions between screens and a user's UI event actions (e.g., scroll) from app screenshots. We believe that these developments could be utilized to overcome the limitations of rule-based feature detection and expand our approach to cover other applications. Second, this work only considered interventions based on self-reflection and restriction. Future work could explore other intervention types and nudges [21] such as punishment, rewards, and social support. Considering a variety of intervention types at different granularities could yield a deeper understanding of the differences in the impact of feature- and app-level interventions. Lastly, most participants in our study were university students who used Android phones. For generalizability, participants from various demographics in terms of age, occupation, and primary device type would be beneficial.

7 CONCLUSION

We designed app-level and feature-level smartphone interventions for mobile social media self-regulation. We implemented and evaluated two versions of our system, app-level *FinerMe*, and feature-level *FinerMe*, to investigate the differences in the impact of interventions based on self-regulation and restriction at two different levels of granularity. Our experiment results showed that both intervention groups effectively reduced usage during the restrictive intervention period. The app-feature use analysis demonstrated that the feature-level intervention group reduced the use of passive app features more than the app-level participants. Moreover, feature-level restrictive interventions led to higher awareness about use time reduction. Nevertheless, our qualitative findings suggested that feature-level interventions are not a one-size-fits-all solution and require careful design decisions. Our study sheds light on how app-level and feature-level regulatory mechanisms impact social media use and provides important design implications for self-regulatory mechanisms incorporating app features.

8 ACKNOWLEDGEMENTS

This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2022-0-00064, Development of Human Digital Twin Technologies for Prediction and Management of Emotion Workers' Mental Health Risks).

REFERENCES

[1] Mohamed Basel Almourad, Amen Alrobai, Tiffany Skinner, Mohammed Hussain, and Raian Ali. 2021. Digital wellbeing tools through users lens. *Technology in Society* 67 (2021), 101778. https://doi.org/10.1016/j.techsoc.2021.101778

[2] Cecilie Schou Andreassen. 2015. Online social network site addiction: A comprehensive review. *Current Addiction Reports* 2, 2 (2015), 175–184. https://doi.org/10.1007/s40429-015-0056-9

[3] American Psychological Association. 2022. Active Control Trial. https://dictionary.apa.org/active-control-trial/.

[4] Albert Bandura. 1991. Social cognitive theory of self-regulation. *Organizational behavior and human decision processes* 50, 2 (1991), 248–287. https://doi.org/10.1016/0749-5978(91)90022-L

[5] Jennifer L Bevan, Jeanette Pfyl, and Brett Barclay. 2012. Negative emotional and cognitive responses to being unfriended on Facebook: An exploratory study. *Computers in Human Behavior* 28, 4 (2012), 1458–1464. https://doi.org/10.1016/j.chb.2012.03.008

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[7] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. 2021. Reflect, Not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 456 (oct 2021), 36 pages. https://doi.org/10.1145/3479600

[8] Emily I. M. Collins, Anna L. Cox, Jon Bird, and Cassie Cornish-Tresstail. 2014. Barriers to Engagement with a Personal Informatics Productivity Tool. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (Sydney, New South Wales, Australia) *(OzCHI '14)*. Association for Computing Machinery, New York, NY, USA, 370–379. https://doi.org/10.1145/2686612.2686668

[9] Terry Connolly and Marcel Zeelenberg. 2002. Regret in decision making. *Current directions in psychological science* 11, 6 (2002), 212–216.

[10] William Jay Conover. 1999. *Practical nonparametric statistics*. Vol. 350. john wiley & sons.

[11] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1389–1397. https://doi.org/10.1145/2851581.2892410

[12] Siddhartha Datta, Konrad Kollnig, and Nigel Shadbolt. 2022. Mind-Proofing Your Phone: Navigating the Digital Minefield with GreaseTerminator. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 523–536. https://doi.org/10.1145/3490099.3511152

[13] Android Developers. 2021. Android Accessibility Service. https://developer.android.com/guide/topics/ui/accessibility.

[14] Android Developers. 2021. UIAutomatorViewer. https://developer.android.com/training/testing/other-components/ui-automator.

[15] Sindhu Kiranmai Ernala, Moira Burke, Alex Leavitt, and Nicole B. Ellison. 2020. How Well Do People Report Time Spent on Facebook? An Evaluation of Established Survey Questions with Recommendations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376435

[16] Shirin Feiz, Jason Wu, Xiaoyi Zhang, Amanda Swearngin, Titus Barik, and Jeffrey Nichols. 2022. Understanding Screen Relationships from Screenshots of Smartphone Applications. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 447–458. https://doi.org/10.1145/3490099.3511109

[17] Joel E. Fischer, Chris Greenhalgh, and Steve Benford. 2011. Investigating Episodes of Mobile Phone Activity as Indicators of Opportune Moments to Deliver Notifications. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 181–190. https://doi.org/10.1145/2037373.2037402

[18] Fausto Giunchiglia, Mattia Zeni, Elisa Gobbi, Enrico Bignotti, and Ivano Bison. 2018. Mobile social media usage and academic performance. *Computers in Human Behavior* 82 (2018), 177–185. https://doi.org/10.1016/j.chb.2017.12.041

[19] Sandy J.J. Gould, Anna L. Cox, Duncan P. Brumby, and Alice Wickersham. 2016. Now Check Your Input: Brief Task Lockouts Encourage Checking, Longer Lockouts Encourage Task Switching. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3311–3323. https://doi.org/10.1145/2858036.2858067

[20] Simon Hill and Andy Boxall. 2018. Can Google really save us from wasting time, or are we on our own? https://www.digitaltrends.com/mobile/is-google-digital-wellbeing-a-salvation-or-a-sham/.

[21] Alexis Hiniker, Sungsoo (Ray) Hong, Tadayoshi Kohno, and Julie A. Kientz. 2016. MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4746–4757. https://doi.org/10.1145/2858036.2858403

[22] Alexis Hiniker, Bongshin Lee, Kiley Sobel, and Eun Kyoung Choe. 2017. Plan & Play: Supporting Intentional Media Use in Early Childhood. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 85–95. https://doi.org/10.1145/3078072.3079752

[23] John R Hughes, Dave C Casal, and Arthur S Leon. 1986. Psychological effects of exercise: A randomized cross-over trial. *Journal of Psychosomatic Research* 30, 3 (1986), 355–360. https://doi.org/10.1016/0022-3999(86)90013-9

[24] "ISO". 1996. *"ISO 14119:2013(en)" Safety of machinery: Interlocking devices associated with guards - principles for design and selection.* British Standards Institution.

[25] J.Clement. 2016. Most popular mobile social networking apps in the United States as of September 2019, by average session duration. https://www.statista.com/statistics/579411/top-us-social-networking-apps-ranked-by-session-length/.

[26] Inyeop Kim, Gyuwon Jung, Hayoung Jung, Minsam Ko, and Uichin Lee. 2017. Let's FOCUS: Mitigating Mobile Phone Use in College Classrooms. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 63 (sep 2017), 29 pages. https://doi.org/10.1145/3130928

[27] Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-Device Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 64 (sep 2017), 21 pages. https://doi.org/10.1145/3130932

[28] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 16 (mar 2019), 29 pages. https://doi.org/10.1145/3314403

[29] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType: Lockout Task Intervention for Discouraging Smartphone App Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300927

[30] Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamiLync: Facilitating Participatory Parental Mediation of Adolescents' Smartphone Use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 867–878. https://doi.org/10.1145/2750858.2804283

[31] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock n' LoL: Group-Based Limiting Assistance App to Mitigate Smartphone Distractions in Group Activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 998–1010. https://doi.org/10.1145/2858036.2858568

[32] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-Based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1235–1245. https://doi.org/10.1145/2675133.2675244

[33] Konrad Kollnig, Siddhartha Datta, and Max Van Kleek. 2021. I Want My App That Way: Reclaiming Sovereignty Over Personal Devices. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 393, 8 pages. https://doi.org/10.1145/3411763.3451632

[34] Geza Kovacs. 2019. *HabitLab: In-the-wild Behavior Change Experiments at Scale.* Stanford University.

[35] Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S. Bernstein. 2019. Conservation of Procrastination: Do Productivity Interventions Save Time Or Just Redistribute It?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300560

[36] Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2018. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 95 (nov 2018), 25 pages. https://doi.org/10.1145/3274364

[37] Robert LaRose. 2007. Uses and gratifications of Internet addiction. *Internet addiction: A handbook and guide to evaluation and treatment* (2007), 55–72. https://doi.org/10.1002/9781118013991.ch4

[38] Robert LaRose, Carolyn A Lin, and Matthew S Eastin. 2003. Unregulated Internet usage: Addiction, habit, or deficient self-regulation? *Media Psychology* 5, 3 (2003), 225–253. https://doi.org/10.1207/S1532785XMEP0503_01

[39] Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim, Subin Yang, Koji Yatani, Gahgene Gweon, Kyong-Mee Chung, and Junehwa Song. 2014. Hooked on Smartphones: An Exploratory Study on Smartphone Overuse among College Students. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2327–2336. https://doi.org/10.1145/2556288.2557366

[40] Liu Yi Lin, Jaime E Sidani, Ariel Shensa, Ana Radovic, Elizabeth Miller, Jason B Colditz, Beth L Hoffman, Leila M Giles, and Brian A Primack. 2016. Association between social media use and depression among US young adults. *Depression and anxiety* 33, 4 (2016), 323–331. https://doi.org/10.1002/da.22466

[41] Markus Löchtefeld, Matthias Böhmer, and Lyubomir Ganev. 2013. AppDetox: Helping Users with Mobile App Addiction. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) *(MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 43, 2 pages. https://doi.org/10.1145/2541831.2541870

[42] Edwin Locke and Gary Latham. 2015. Goal-setting theory. *Organizational behavior* 1 (2015), 159–183.

[43] Edwin A Locke and Gary P Latham. 1990. *A theory of goal setting & task performance.* Prentice-Hall, Inc.

[44] Edwin A Locke and Gary P Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist* 57, 9 (2002), 705. https://doi.org/10.1037/0003-066X.57.9.705

[45] Asia Pacific Net Ltd. 2021. Korea Digital Marketing 2021. https://www.asiapacdigital.com/digital-marketing-insight/korea-digital-marketing-2021.

[46] 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 368, 17 pages. https://doi.org/10.1145/3411764.3445467

[47] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376672

[48] ManicTime. 2016. Website. https://www.manictime.com.

[49] Patrick E McKnight and Julius Najab. 2010. Mann-Whitney U Test. *The Corsini encyclopedia of psychology* (2010), 1–1. https://doi.org/10.1002/9780470479216.corpsy0524

[50] Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300616

[51] Facebook Newsroom. 2018. New Tools to Manage Yout Time on Facebook and Instagram. https://newsroom.fb.com/news/2018/08/manage-your-time/.

[52] Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good Vibrations: Can a Digital Nudge Reduce Digital Overload?. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/3229434.3229463

[53] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous computing* 16, 1 (2012), 105–114. https://doi.org/10.1007/s00779-011-0412-2

[54] Erin S Pearson. 2012. Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components. *Patient education and counseling* 87, 1 (2012), 32–42. https://doi.org/10.1016/j.pec.2011.07.018

[55] Sarah Perez. 2018. YouTube rolls out new tools to help you stop watching. https://techcrunch.com/2018/05/11/youtube-rolls-out-new-tools-to-help-you-stop-watching/.

[56] Shaokan Pi. 2015. Forest. https://www.forestapp.cc/.

[57] Charlie Pinder, Jo Vermeulen, Benjamin R. Cowan, and Russell Beale. 2018. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Trans. Comput.-Hum. Interact.* 25, 3, Article 15 (jun 2018), 66 pages. https://doi.org/10.1145/3196830

[58] RescueTime. 2016. Website. https://www.rescuetime.com.

[59] John Rooksby, Parvin Asadzadeh, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2016. Personal Tracking of Screen Time on Digital Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 284–296. https://doi.org/10.1145/2858036.2858055

[60] Mintra Ruensuk, Taewan Kim, Hwajung Hong, and Ian Oakley. 2022. Sad or Just Jealous? Using Experience Sampling to Understand and Detect Negative Affective Experiences on Instagram. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 147, 18 pages. https://doi.org/10.1145/3491102.3517561

[61] Tracii Ryan, Andrea Chester, John Reece, and Sophia Xenos. 2014. The uses and abuses of Facebook: A review of Facebook addiction. *Journal of behavioral addictions* 3, 3 (2014), 133–148. https://doi.org/10.1556/jba.3.2014.016

[62] Dale H Schunk. 2001. Self-Regulation through Goal Setting. ERIC/CASS Digest. (2001).

[63] Mical Kay Shilts, Marcel Horowitz, and Marilyn S Townsend. 2004. Goal setting as a strategy for dietary and physical activity behavior change: a review of the literature. *American Journal of Health Promotion* 19, 2 (2004), 81–93. https://doi.org/10.4278/0890-1171-19.2.81

[64] Gideon Simons. 2015. UX is not Usability. https://medium.com/@papasimons/ux-is-not-usability-a9f82e4a60ba.

[65] Victor J Strecher, Gerard H Seijts, Gerjo J Kok, Gary P Latham, Russell Glasgow, Brenda DeVellis, Ree M Meertens, and David W Bulger. 1995. Goal setting as a strategy for health behavior change. *Health education quarterly* 22, 2 (1995), 190–200.

[66] R Core Team et al. 2013. R: A language and environment for statistical computing. (2013).

[67] Screen Time. 2020. Screen Time. https://support.apple.com/en-us/HT208982.

[68] Philippe Verduyn, Oscar Ybarra, Maxime Résibois, John Jonides, and Ethan Kross. 2017. Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review* 11, 1 (2017), 274–302. https://doi.org/10.1111/sipr.12033

[69] Digital Wellbeing. 2018. Digital Wellbeing. https://www.android.com/digital-wellbeing/.

[70] JO Wobbrock and M Kay. 2016. Modern Statistical Methods for HCI. (2016).

[71] Jason Wu, Xiaoyi Zhang, Jeff Nichols, and Jeffrey P Bigham. 2021. Screen Parsing: Towards Reverse Engineering of UI Models from Screenshots. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 470–483. https://doi.org/10.1145/3472749.3474763

[72] Xuhai Xu, Tianyuan Zou, Han Xiao, Yanzhang Li, Ruolin Wang, Tianyi Yuan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K Dey. 2022. TypeOut: Leveraging Just-in-Time Self-Affirmation for Smartphone Overuse Reduction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 442, 17 pages. https://doi.org/10.1145/3491102.3517476

[73] Philip T Yanos, Alicia Lucksted, Amy L Drapalski, David Roe, and Paul Lysaker. 2015. Interventions targeting mental health self-stigma: A review and comparison. *Psychiatric rehabilitation journal* 38, 2 (2015), 171. https://doi.org/10.1037/prj0000100

[74] Victoria Young. 2015. Strategic UX: The Art of Reducing Friction. http://www.dtelepathy.com/blog/business/strategic-ux-the-art-of-reducing-friction.

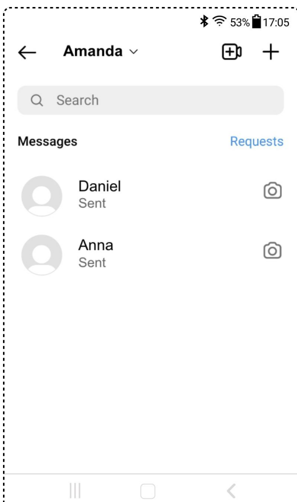
[75] Jerrold H Zar. 2014. Spearman Rank Correlation: Overview. *Wiley StatsRef: Statistics Reference Online* (2014). https://doi.org/10.1002/9781118445112.stat05964

[76] Mingrui Ray Zhang, Kai Lukoff, Raveena Rao, Amanda Baughan, and Alexis Hiniker. 2022. Monitoring Screen Time or Redesigning It? Two Approaches to Supporting Intentional Social Media Use. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 60, 19 pages. https://doi.org/10.1145/3491102.3517722

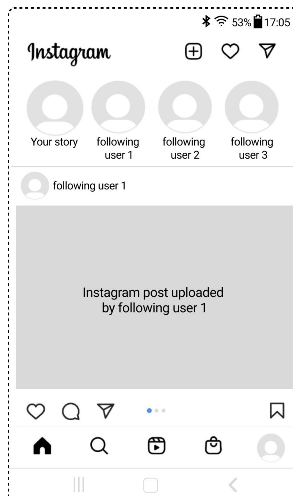
[77] Inc. Zoom Video Communications. 2021. Video Conferencing, Web conferencing, Webinars, Screen Sharing - Zoom. https://zoom.us/.

A APPENDIX

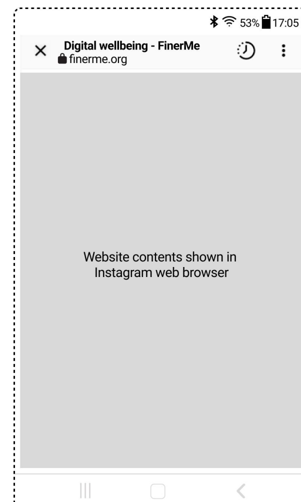
Instagram app features



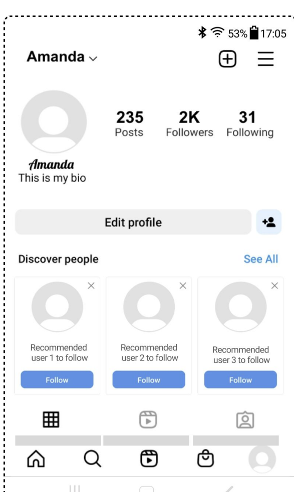
Direct Messaging
Chat with other users



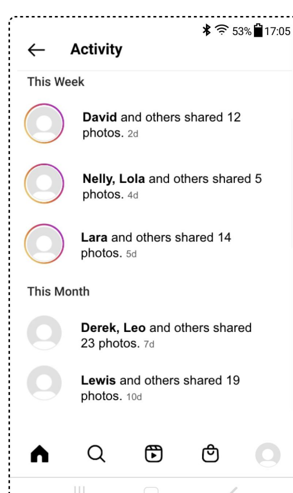
Following Feed
View following users' feeds



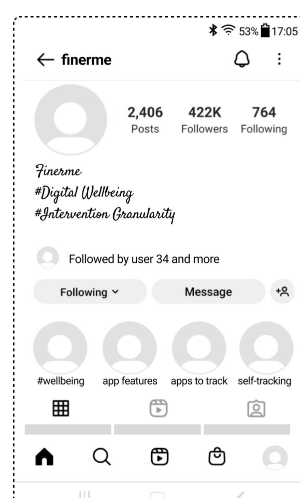
In App Web View
Use Instagram's web browser



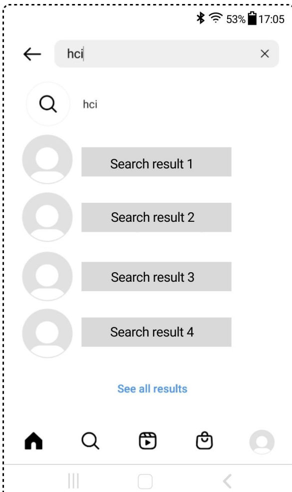
My Profile
View the user's own profile



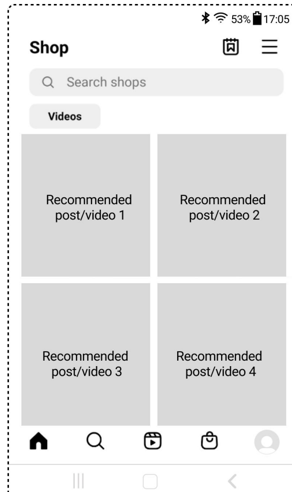
Notifications
Check notifications



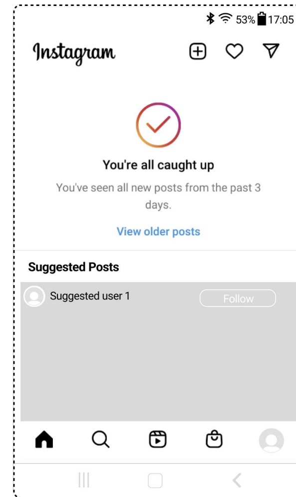
Other's Profile
View other users' profile



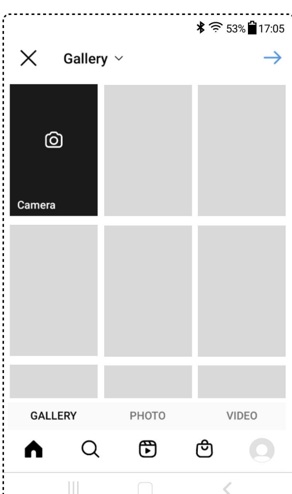
Search
Use keywords to search



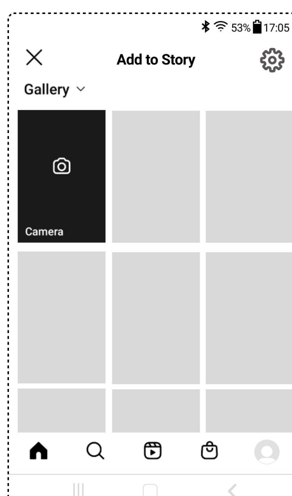
Shopping
Browse for shopping



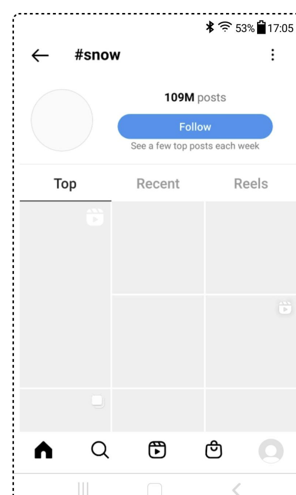
Suggested Feed
View system-recommended feed



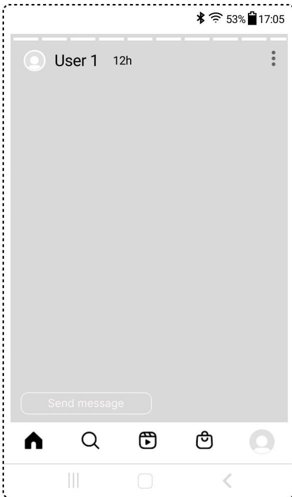
Upload Post
Publish an Instagram post



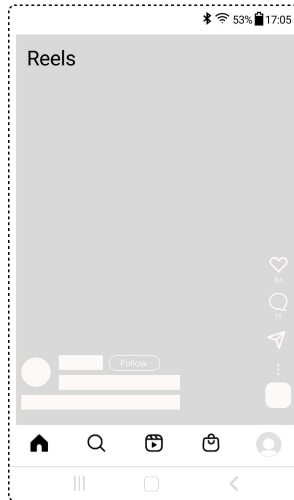
Upload Story
Publish an Instagram story



View by Hashtag
View hashtag-filtered posts

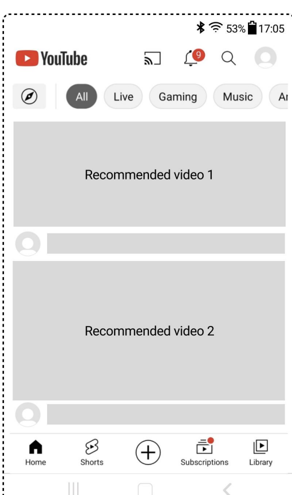


View Story
View an Instagram story

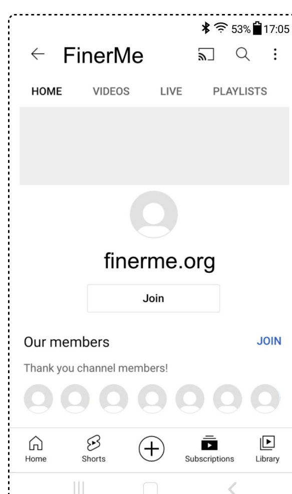


Watch Reels
Watch short (up to 60s) clips

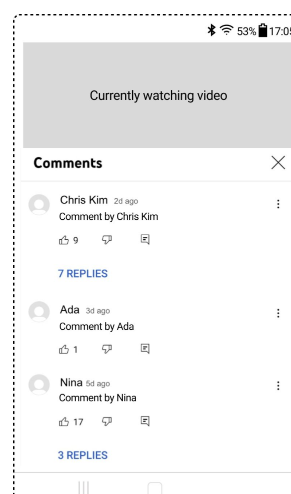
YouTube app features



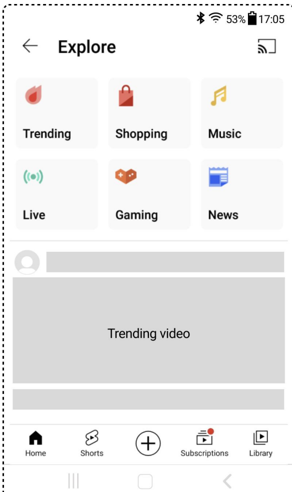
Browse Home
Browse suggested videos



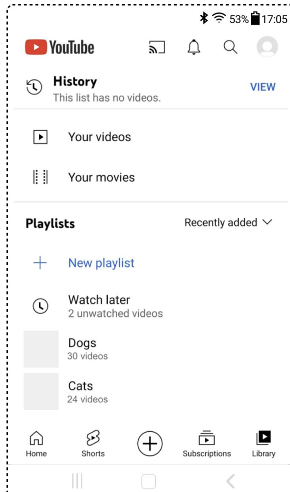
Channel
View channel contents



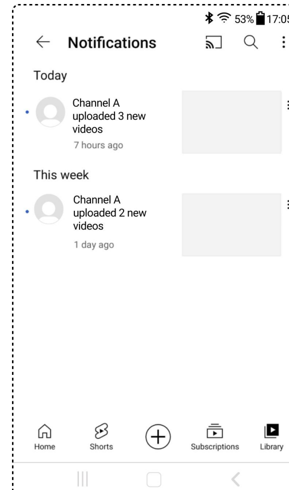
Comment
Read comments



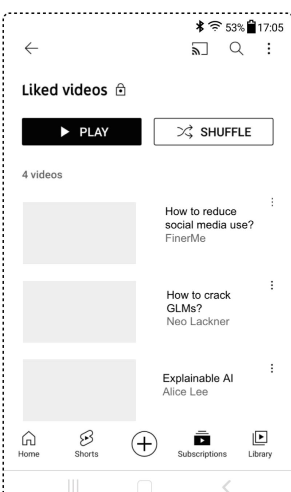
Explore
Explore trending videos



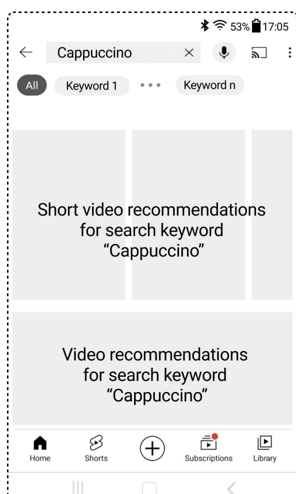
My Library
View history, saved videos



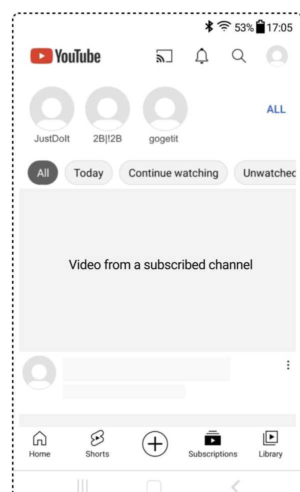
Notifications
Check notifications



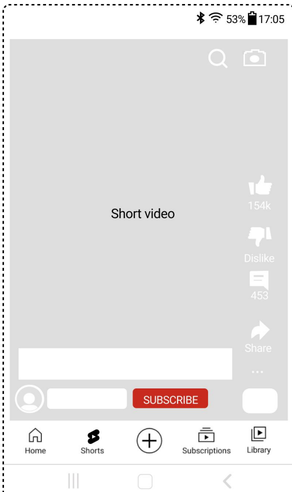
Playlists
Browse saved videos or playlists



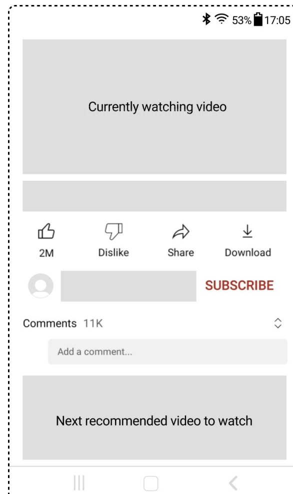
Search
Search videos



Subscriptions
View following channel’s videos



Watch Reels
Watch short (up to 60s) clips



Watch Video
Watch videos of unlimited size

Received July 2022; revised January 2023; accepted March 2023