Privacy as Social Norm: Systematically Reducing Dysfunctional Privacy Concerns on Social Media

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 123-A56-BU3

Privacy is essential to fully enjoying the benefits of social media. While fear around privacy risks can sometimes motivate privacy management, the negative impact of such fear, particularly when it is perceived as unaddressable (i.e., *dysfunctional fear*), can significantly harm teen well-being. In a co-design study with 136 participants aged 13-18, we explored how teens can protect their privacy without experiencing heightened fear. We identified seven different sources of dysfunctional fear, such as 'fear of a hostile environment' and 'fear of overstepping privacy norms.' We also evaluated ten designs, co-created with teen participants, that address these fears. Our findings suggest that social media platforms can mitigate dysfunctional fear without compromising privacy by creating a culture where privacy protection is the norm through default privacy-protective features. However, we also found that even the most effective privacy features are not likely to be adopted unless they balance the multifaceted and diverse needs of teens. Individual teens have different needs—for example, public and private account users have different needs—and teens often want to enjoy the benefits they get from slightly reducing privacy and widening their social reach. Given these considerations, augmenting default privacy features by allowing them to be toggled on and off will allow individual users to choose their own balance while still maintaining a privacy-focused norm.

CCS Concepts: • **Human-centered computing → Collaborative and social computing**.

ACM Reference Format:
Anonymous Author(s). 2024. Privacy as Social Norm: Systematically Reducing Dysfunctional Privacy Concerns on Social Media. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 37 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Due to the pervasive influence of social media, the landscape of teenage social interaction has largely migrated to the online realm [5]. This shift has led to an escalation in privacy concerns, highlighting the delicate balance between online engagement and personal data protection for teens [10, 75]. While debates have persisted about teenagers' attitudes towards privacy, recent studies indicate that they actively engage in various strategies to safeguard their personal information [78]. Nevertheless, significant concerns around teen privacy management remain underexplored.

A revelation from a 2019 Pew Research Center study [14] underscores an important concern around privacy management: 82% of Americans feel they have little to no control over their personal data, including location tracking. This sentiment is further pronounced among teenagers, who, as a vulnerable group, often feel overwhelmed and powerless in the face of privacy challenges. Research has illuminated several factors contributing to this vulnerability: teens frequently encounter a lack of self-efficacy when navigating privacy issues [30], face conflicting privacy norms among peers [69], and struggle to find viable solutions to their privacy concerns [15]. This confluence of factors can

Manuscript submitted to ACM

engender a state of ***dysfunctional fear*** [46] in teens, which is characterized by heightened anxiety over privacy that paradoxically diminishes their well-being and discourages proactive measures.

In this paper, given the critical need for teen privacy protection in social media while preventing amplified dysfunctional fear, we investigate the following questions:

(1) **RQ1.** What, if anything, do teens fear in their social media experiences? How do these fears affect their social media experiences?
(2) **RQ2.** How might the design of social media mitigate dysfunctional fear for teens? What are the implications of these designs?

To answer these questions, we conducted a mixed-methods study with a total of 137 13-to-18-year-old participants residing in the U.S. We first conducted co-design interviews with nineteen adolescents to understand privacy concerns they have difficulty addressing and how those affect their social media experience. The participants also shared design suggestions to address their privacy concerns. We then took the design suggestions and created high-fidelity prototypes for ten of the larger design idea themes that we identified from the co-design interview data. We also conducted a design evaluation survey with 72 private account owners and 64 public account owners.

Our results reveal that teens often experience a vague and persistent fear that remains unaffected by their privacy management strategies on current social media platforms. The fear concerned three overarching themes and seven subthemes, including 'fear of time collapse,' 'fear of a hostile environment,' and 'fear of overstepping privacy norms.' Such fears impact their quality of life, leading them to either withdraw from desired social media activities or become hypervigilant after posting. They also struggled to find sufficient features to address the fears, particularly when they have a public account. Moreover, the overall hostility of social media environments often inhibited their sense of safety. Even when protective options are available, they found that their peers' privacy expectations differ from their own, creating pressure to conform rather than prioritize their preferences. We find that features such as clarifying privacy norms during onboarding and providing default but optional features for privacy protection, such as screenshot blocking, can alleviate their fears while reducing the likelihood of privacy issues.

In this paper, we contribute empirical evidence of dysfunctional fear in the context of teen privacy on social media. We also provide an overview of features and affordances that could effectively reduce both privacy concerns and risks. Furthermore, we identify moments where teens' privacy needs collide with their peers' privacy expectations and suggest methods for platforms to foster a culture where privacy is considered the norm. We hope this work will prompt designers and researchers to consider the adverse effects of excessive or unaddressed privacy concerns that teens may experience online and systematically empower adolescent users to protect themselves while fully reaping the benefits of social media.

2 RELATED WORK

2.1 Teen Privacy Concerns on Social Media

Social media platforms offer a multitude of benefits, particularly for teens, including the enhancement of social capital and the facilitation of meaningful discourse. Central to these advantages, such as relationship formation, is self-disclosure [37, 39, 67, 68]. Often, disclosing personal information is a prerequisite to fully leveraging technological tools' potential, such as the different types of social capital benefits [38, 39]. The more users share about themselves, the greater the benefits they can derive from these systems. However, this increased disclosure also escalates the risk of what users perceive as privacy violations [31, 67]. It is also unsurprising that safety threats on social media platforms

can inhibit users' ability to gain these advantages. Both real and perceived threats can diminish the quality and benefits of interactions on these online social spaces [23, 34, 48, 51, 65]. This trade-off is particularly unfortunate considering the heavy reliance of modern teens on social media for socialization and identity development [29, 42].

Teens generally face many privacy concerns on social media, although their focus might be somewhat different from that of adults. Teens often focus their privacy efforts on evading direct supervision, such as from family members, avoiding interactions with strangers, and circumventing in-person drama with peers [78]. Moreover, due to their developmental stage, adolescents are highly self-conscious, often leading to the belief that they are constantly being watched or judged (i.e., *"imaginary audience"* [36]). This belief, coupled with the low audience transparency inherent in social media, heightens teens' concerns about audience scrutiny. In addition, teens' perceptions of privacy evolve as they encounter changing societal norms and expectations, leading to concerns about their digital footprint due to the persistent nature of social media content. Another privacy issue is oversharing, which involves posting excessive personal information or posting too frequently, a concern often developed by observing their peers [18, 78]. In addition, privacy concerns frequently arise due to the challenges in managing networked privacy [56] on social media. Although teens are quite adept at navigating and configuring settings in networked environments [78], networked privacy by its nature places individuals in social contexts where their privacy can be, and often is, violated, whether intentionally or unintentionally [43].

2.2 Designs and Strategies for Teen Privacy on Social Media

Contrary to early beliefs that teen privacy revolves around the *"privacy paradox,"* suggesting teens are less concerned with privacy than older individuals [70], it is now understood that valuing privacy is the norm among young people [15, 73, 78]. Teens experience a conflict between the desire to share information with their friends and the concern of unintended audiences viewing their content [12]. Thus, they actively engage in protecting their privacy through various strategies. These include, but are not limited to, posting in a strategically vague manner to obscure the meaning from those outside their network [78], disengaging from unsafe conversations [13], leveraging ephemeral modes of posting [24, 77], gradually decreasing the visibility of posts to mitigate risks associated with digital footprints [59, 63], adjusting privacy settings, and practicing self-censorship to prevent *"context collapse"* [32] or *"time collapse"* [20]. They also control their digital boundaries by filtering follow requests, blocking or removing followers [11], limiting audience reach [78], and creating trusted spaces [76]. Additionally, teens manage their online presence by archiving or deleting content [11], adhering to and respecting implicit networked privacy norms [62, 78], and employing other nuanced measures to maintain their digital privacy. Teens typically employ sophisticated privacy management techniques for their privacy protection needs, even when they believe they "have nothing to hide" [9].

The Computer-Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI) domains have made significant progress in finding ways to support teens' privacy protection better. The overall direction of the findings is to move away from designing based on paternalistic, protectionistic, and *"concern-centric"* approaches [11, 17, 47, 75] and toward a *"risk-centric"* [75] framework. Teens are turning to social media for connection and relaxation due to decreased public freedom and increased adult supervision in their physical environments. This shift to online spaces allows them to maintain social interactions in a setting that offers more privacy and autonomy, highlighting the need for a balance between guidance and independence [17]. Thus, the risk-centric framework suggests that teens can adopt more adequate privacy protection measures as a result of exposing themselves to increased online disclosures, thereby facing higher susceptibility to risky online interactions [47, 74, 75]. Additionally, measures that engage teens more directly and empower them, such as discussions and nudging them toward self-regulation, have been emphasized [11, 49, 75].

Specific designs that support teen privacy include nudging [7, 11, 57], multiple close friends list [78], and preventive measures such as restrictions for perpetrator and sensitivity filter [11]. Designs to address adult privacy concerns, such as improving the findability of privacy interfaces, may also be relevant for teens [27, 45].

2.3 Toward a Community-level Privacy Support for Teens on Social Media

Despite advancements in supporting teen privacy, teens continue to face significant challenges in privacy protection [64]. A key factor is a reliance on individual teens' conscious efforts for privacy management [71, 72], which often falls short when teens perceive their privacy concerns as being beyond their control [50]. This perception can arise from a *lack of self-efficacy* [30], *social norms conflicting with individual privacy needs* [69], or *limited options for adequately addressing privacy issues* [15]. Specifically, navigating privacy in networked public spaces requires users to manage technical affordances, interpersonal relationships, and social norms [78], and peer norms significantly influence teen privacy practices [69, 78]. Teens rely on trust that their friends will respect the boundaries of co-owned privacy [78], but social media platforms lack explicit rules and norms around this co-ownership. Consequently, when friendships deteriorate, the fragile agreements about what can and cannot be shared often break down, leading to breaches in privacy [52]. These challenges align with *Contextual Integrity theory* [60], which posits that privacy is maintained when information flows respect context-specific norms. For teens, the difficulty in managing these norms across various social media contexts can contribute to privacy concerns and feelings of reduced control.

Teens these days also face significant pressures from real-life social dynamics that complicate privacy management [73]. For example, constant digital documentation by peers and the fear of social repercussions like *"cancel culture"* (i.e., the practice of publicly shaming and boycotting individuals for perceived wrongdoings) make it difficult for teens to maintain control over their digital footprints [73]. The concept of *"receipts"* (i.e., proof screenshots, screen recordings, or messages saved for evidence of an indiscretion) and the culture of surveillance only exacerbate these issues, making teens feel more vulnerable and less empowered about their privacy [17, 28, 66, 73]. As such, the complexity and perceived lack of control often lead to *"network defeatism"* [30] among teens, a sense of fatalism regarding privacy decisions on social media [30, 43]. This reduced sense of self-efficacy in privacy control hampers teens' ability to address privacy issues and enhance their social media experiences [48, 53]. Furthermore, teens are more likely to overlook their vulnerability when they feel less efficacious, leading to disengagement from privacy protection [26]. Adults' well-meaning warnings about the permanence of online actions often add to this disempowerment, instilling a sense of fear and cynicism [73].

Echoing themes from fear of crime research, these phenomena underscore the need to understand worries and fears around privacy-related issues on social media. For example, fear of crime research differentiates between *dysfunctional worry*, which degrades the quality of life, and *functional worry*, which fosters precaution and awareness. Fear of crime research has found that only a quarter of worried individuals used their fear constructively, without impacting their life quality [46]. Fear of cybercrime studies reveal a similar pattern: while fear can be a motivator for protective behaviors, it often restricts activities and thus the quality of life, particularly when the fear response is disproportionate to the actual risk [19][1]. The influence of social environment on fear, as highlighted by the fear of crime research, may also be relevant to social media privacy concerns, suggesting that factors like (perceived) community disorder and (perceived) lack of social support can amplify fear [40, 54, 55]. Therefore, in this paper, we argue that addressing social media privacy concerns requires a balance between vigilance and system-wide reassurance [65], as well as measures to increase

[1]It is important to note, however, that our research does not aim to quantify the actual risks of privacy issues for teens on social media or determine what constitutes dysfunctional fear in this context. Rather, we seek to understand the emotional and psychological mechanisms of (dysfunctional) fear of crime and how they might apply to social media privacy concerns.

self-efficacy. Following the *"Privacy by Design"* principle [22], proactive approaches that integrate privacy into the design are essential. This is particularly important for teens, who are highly susceptible to peer pressure and social norms. However, it is crucial that community-wide privacy features do not merely create a *"safety theater,"* [65] which focuses solely on reducing fear without real safety, as a false sense of security can be dangerous [8, 25, 41, 61, 71]. The aim should be to empower users to move away from dysfunctional fear by transforming their concerns into constructive actions [46].

3 METHOD

3.1 Co-design Interviews

Materials and Procedures. We conducted co-design interviews in two stages: 30-minute entry and 60-minute exit interviews. We explored participants' privacy concerns on social media platforms during entry interviews. We asked generally where they felt most comfortable sharing, with whom they shared, and what (if any) issues or features influenced their sharing habits.

During exit interviews, we asked participants about the barriers (if any) that prevent them from comfortably sharing on social media. To do so, we used the Miro platform [4] to create a virtual whiteboard with nine sticky notes, each representing the improvements they seek on social media (and its related subcategories of concerns) identified in our survey. Examples of the larger themes on changes sought were "prevent posts from being shared out of context," while sub-themes of privacy concerns included "Content being shared with strangers" and "Screenshots taken / content saved." (The full list of the sticky notes is available in the supplementary materials.) We used these to prompt a conversation where participants highlighted the barriers most relevant to them.

Afterward, we conducted a co-design sketching exercise with the participants to dive deeper into these concerns' significance and relevance and to brainstorm potential design solutions. To prompt ideas for designs, we asked questions such as "What are your perceived social norms for privacy? Can those be flexible or broken? Are these different across different platforms?" Then, for each design idea the participants came up with, we asked them to sketch and explain their concept. We inquired about the possible benefits and downsides of incorporating the feature and whether such effects would differ across platforms or people.

The first author conducted the semi-structured entry and exit interviews via Zoom, lasting approximately 30 to 40 minutes and 60 to 90 minutes, respectively. Participants were compensated with $10 and $20 Amazon gift cards for their participation in the entry and exit interviews.

Participants and Recruitment. We reached out to individuals from our previously established participant pool, which was initially formed through purposive and convenience sampling. This sampling involved advertisements on social media platforms such as Instagram and Facebook, specifically targeting individuals aged 13 to 18 in the United States. Additionally, we invited individuals previously involved in or interested in our studies and provided consent to be contacted for future research opportunities. In total, we invited 80 participants to complete a general screening survey for our co-design study. Of the 47 respondents who completed the screener, 22 were chosen to participate in the co-design study. Participants were selected to reflect a broad range of demographic characteristics. Of these 22 invitees, 20 initially participated in the study. However, one participant withdrew after the entry interview. Consequently, **19 participants** (demographic details in Appendix A) completed all stages of the co-design study.

Reflexive Thematic Analysis. We undertook a reflexive thematic analysis [21] of the 38 entry and exit interview transcripts. Our approach was inductive, yet we concentrated on privacy-related aspects—especially privacy fears that participants felt undermined their social media experience—since the interviews encompassed a broader range of topics than could be addressed in a single paper. The analysis was structured into three phases over 2.5 weeks.

In the initial phase, the first author, two co-authors, and the last author engaged in line-by-line coding of two identical transcripts. The initial codes were descriptive, closely reflecting the data. Following this phase, the first author compiled these notes to draft the preliminary codebook. Subsequently, in the second phase, the same team independently coded two additional transcripts using Atlas.ti software [6] using this preliminary codebook and adding new codes as necessary. The first author then reviewed these codes and newly found themes to refine the codebook further.

In the final phase, the primary author coded all transcripts to refine the codebook. The two co-authors then coded 20 of the 38 transcripts, validating the clarity and validity of the codebook and the assigned codes. Throughout each phase, the team convened to discuss higher-level themes, disagreements, or ambiguities related to the codebook. Higher-level themes included "Fear that teens experience over privacy," "Existing strategies for mitigating fear," "Designs that affect teens' fear," and "Implications of the designs." Sub-codes included "Ripple audience," "Perspective taking," "Prompts for self-regulation," and "Privacy-utility trade-off." Each iteration of the codebook, complete with definitions and example quotes for each code, is available in the supplementary materials.

3.2 Design Evaluation Survey

Materials and Procedures. For each of the ten design idea themes developed during the inductive coding process of the co-design interviews, the first and fourth authors—both advanced-degree students in technology design programs—created high-fidelity prototypes of ten sample features (See Figure 1 for sample prototypes and the supplementary materials for the entire set of prototypes) that were provided to the survey respondents. These authors selected specific implementations for each design based on their potential to probe the defining characteristics of each design idea.

We incorporated the ten prototypes into a survey, inviting participants to evaluate each prototype in terms of general reactions (e.g., "What (if anything) do you dislike about this feature?"); usefulness in addressing privacy concerns (e.g., "With this feature, I would feel less worried about privacy-related problems on my [public/private] account"); likelihood of backfiring (e.g., "If I use this feature on my public account, I'm concerned it might lead to awkward or uncomfortable situations with my friends or people I know in real life."); and to answer general questions on demographics; social media use (e.g., "On which social media platforms do you have (or have you had) a public account? (choose all that apply)"); and privacy concerns related to social media (e.g., "Thinking about your [public/private] account on [platform], please rate your concern that someone might expose your private conversations with others, either on purpose or by accident").

Through these questions, we aimed to understand the extent of teens' dysfunctional fear of privacy risks on social media and assess each feature's potential strengths and weaknesses. The survey's estimated completion time was 38 minutes, with participants receiving $12 Amazon gift cards for their time. The median completion time was 48.2 minutes, with an interquartile range of 26.9-106.3 minutes.

We hypothesized that the effectiveness of, and thus the reactions to, the features would significantly differ between public and private account owners. We also anticipated they would have different types and extents of privacy concerns. Consequently, we created two versions of the same survey, one for public and another for private account owners. Each version asked participants to consider the features and privacy concerns in the context of the public or private account

they most frequently use. We permitted participants to respond to both surveys if they owned both types of accounts. The complete wording of the two versions of surveys can be found in the supplementary materials.

Participants and Recruitment. We invited 443 selected individuals who met one of two criteria: 1) those who have participated in our past studies and demonstrated sincerity in their participation, and 2) new recruits via an Instagram advertisement who answered the free-response question in the screening survey in good faith. At no point in the recruitment process did we specify a need for teens with specific privacy concerns. Instead, we targeted teens that would "tell us about [their] social media experience." We received a total of 201 responses: 96 from private account owners and 105 from public account owners.

Each response was reviewed, and we excluded those that appeared to be 1) submitted multiple times by an individual, 2) generated by generative AI, or 3) not genuine, such as responses not addressing the question or only offering vague positive comments such as *"I would love this"* without any evidence of having reviewed the prototypes throughout the entire survey. The detailed process and criteria we employed for this process are available in the supplementary materials. After this filtering process, data from 72 private account owners and 64 public account owners remained. Among these, 18 participants responded to both surveys. The demographics of the **118 participants** (who provided 136 responses) are detailed in Appendix A.

Evaluating Design Features (One-Sample t-Tests). We conducted one-sample t-tests on their reactions to each prototype and an overview of all prototypes. These tests assessed whether participants' responses to each survey question significantly deviated from a hypothesized mean of 3, representing "neither agree nor disagree," the midpoint of our survey scale.

Identifying Dysfunctional Privacy Concerns. Following the framework of Jackson and Gray [46], we classified privacy concerns on social media into three categories: unworried, functional worry, and dysfunctional worry. Our goal was not to determine what constitutes an "appropriate" level of worry. Rather, we sought to identify instances of intense worry that diminish the quality of life and are not or cannot be accompanied by effective, constructive actions. Details about our adoption of this categorization algorithm can be found in Appendix B.

Clustering Open-Ended Feedback. We analyzed participant feedback about design features using the MPNet-Base-V2 language model [2] from the Sentence Transformers library [3], converting responses into 320-dimensional vectors. This process resulted in 136 embeddings for both 'likes' and 'dislikes.' We then applied k-means clustering [44] to these embeddings, selecting an optimal number of clusters (between 3 and 7) based on the silhouette coefficient [35] for meaningful categorization. Each cluster was labeled by its main theme, with representative quotes provided in Appendix C. These clusters helped us understand the design feedback in a time-efficient and targeted manner; While the clusters are not definitive, we used the clustering results as a guide for stratified and purposive sampling when evaluating the open-ended responses.

3.3 Ethical Considerations

Both procedures were approved by our Institutional Review Board (IRB) prior to data collection. We sent consent forms (including parental consent) to participants before the interviews. During the interviews, we summarized the consent form and procedures, allowing participants to ask questions. We obtained both written and verbal consent to inquire about personal social media experiences and to record responses.

Given the sensitivity of the topic, we took additional precautions. We explicitly informed participants that they could decline to answer any questions that made them uncomfortable. Participants were given the option to turn off their cameras. We emphasized prior to starting the interviews that our goal was to understand teens' perspectives, not to judge their experiences or behaviors.

3.4 Use of Generative AI

We utilized ChatGPT-4 [1] to generate initial ideas for section headings, prototype names, and theme titles. Typical prompts were: "[summary of content in the section and long description of the key points we aimed to deliver through the section, written by the authors] What might be some headings for a section that suggests this? Give me a long list." We then refined these initial ideas to finalize versions that most accurately delivered what we were looking for. Additionally, the AI helped refine grammar in parts of our paper, draft accessibility texts for figures, and improve the conciseness of our writing. Typical prompts were "Just fix grammar without summarizing or anything" and "Is any part grammatically wrong or awkward?"

4 RESULTS

4.1 How Does Fear Affect Teens' Social Media Experiences?

As shown in prior literature, teens actively took measures to protect their privacy [78]. These measures include withholding information, considering their audience's perspective, reviewing and reversing content, assessing boundaries, and developing mental models. Often, they develop these models by observing their peers to understand what constitutes appropriate privacy practices.

However, as evidenced in prior literature, teenagers also perceive privacy concerns or fears as being beyond their control due to various reasons such as a lack of self-efficacy [30], social norms that conflict with individual privacy needs [69], or limited options for adequately addressing privacy issues [15]. Our survey results also confirmed the presence of dysfunctional fear in teens; the survey revealed that among our survey respondents with a private account, 15.3% experience dysfunctional fear. Among public account owners, the figure is 28.1%, almost twice as much.

From our co-design interviews, we identified ten sources of such fears, grouped under three broader themes: 'fear of content being shared out of context,' 'fear of digital vulnerability,' and 'fear of online missteps.' These fears prevent teens from fully utilizing the benefits of social media. We also outlined design approaches that teens suggested to mitigate each type of fear. Although we discussed specific design ideas with specific types of fear, this does not imply that these ideas are exclusively effective for the particular type of fear they are associated with. We associated them based on where they were most frequently mentioned.

4.1.1 Fear of Uncontrolled Audience Reach. The teen participants frequently expressed concerns about their digital content being shared or interpreted outside its original context, driven by the persistent, replicable, and scalable nature of digital media [16] and uncertainties surrounding networked boundaries. To address these fears, participants suggested enhancing visibility and boundary control options, increasing interaction transparency, implementing screenshot controls, and providing flexible content management features such as ephemerality and easy post-deletion.

Imaginary Audience. Eighteen participants (P02-P19) shared anxieties about potentially having large, unknown audiences on social media. This aligns with previous research findings, highlighting the opaque social media visibility mechanisms [33]. These teen participants experienced low audience transparency and felt they had little control over

whether their social media content might unintentionally reach unwelcome or unintended audiences. This fear was especially prevalent among those with public accounts. On platforms like YouTube, algorithms can cause content to *"go viral,"* exposing it to *"anyone who uses the platform"* (P04). Even those with private accounts were not immune to this fear, feeling inhibited from sharing due to their *"large following"* (P17). Additionally, even when utilizing features like Instagram's Close Friends Story, some teens felt these lists were too *"general"* (P17) and thus not sufficiently exclusive for certain types of content. In some instances, teens experienced anxiety over not knowing how the audience was engaging with their content, fearing excessive scrutiny or negative judgment from unseen viewers, a phenomenon similar to 'stage blindness'—the anxiety of being on stage and unable to see the audience due to bright lights. As P19 succinctly put it, *"We are afraid of the unknown; when you post, you don't really know who has seen it and who hasn't, or who disliked it."*

One design idea proposed by all nineteen teen participants to alleviate the fear of an imaginary audience involved enhancing visibility and *boundary control* options for public accounts while strengthening these controls for private accounts. They were already actively using various visibility and boundary control mechanisms and strategies on mainstream social media platforms, including rejecting friend requests, removing followers or friends, selective sharing, intimate reconfiguration, and privatizing accounts, as found in prior research. However, such boundary and visibility controls limited the benefits teens sought from social media. For instance, private accounts made it *"harder to meet new people"* (P14), conflicting with their desire to *"constantly networking and expanding [their] audience"*, especially for those pursuing activities like music. As P15 explained, with a public account, you *"you sort of give up the boundary . . . to other people,"* and must *"accept"* the *"opportunity costs."* They, therefore, suggested that there is a need for *"a way for people that want to be public to still be able to have some sort of ... knowing what's going on with their content."* For private accounts, participants desired more nuanced selective sharing options, with various *"groups,"* rather than the limited options, such as the *"only two groups"* that Instagram currently offers. Ideally, these options would cater to different interests, as suggested by P19.

Fifteen participants (P01, P02, P04-P09, P11-P15, P17, P18) suggested increasing *interaction transparency* as another design direction. They particularly desired clarity about *"how people interact with your posts,"* such as identifying viewers and their interactions (P07). For instance, P09 proposed a feature like *"profile views"* that would allow users to see who has viewed their regular posts, similar to Instagram Stories, as *"knowing who sees you"* contributes to a sense of security. Understanding *"how people interact with your posts"* (P07) would not only *"set off warning bells,"* but also provide them with the opportunity to block someone if necessary. This knowledge would offer them *"peace of mind,"* as they would *"prefer to know... rather than be anxious about who is sharing [their] content."* It would also allow them to recognize and address uncomfortable interactions (P07). Although most participants valued having insight into their content's interactions, some believed that *"sometimes ignorance is better,"* particularly when expected responses to their posts were not forthcoming. They preferred to assume the post was unseen rather than acknowledge it was ignored (P14). Nevertheless, they still acknowledged the importance of interaction transparency and recommended making this feature optional, with the ability to *"toggle it on and off"* (P14).

Boundary Violations. Among our 19 participants, 17 (P01-P09, P11, P12, P14-P19) expressed concerns that their content, intended for a specific audience, might be intentionally or unintentionally exposed to individuals outside a mutually understood privacy boundary. This trust is sometimes broken with malicious intent, as in cases where *"people take screenshots and send them to individuals whom you didn't intend to see them"* (P07). At other times, the implicit expectations of the shared boundary are breached unintentionally. An example is, *"We'll screenshot the messages we sent*

to each other and then send them to someone else, which we started doing because we thought it was funny... but then it gets out of hand" (P06). This is concerning because such breaches of the shared boundary can lead to misunderstandings or misrepresentations when content is viewed outside its original context by those unfamiliar with its background or if a trusted party betrays trust by purposely sharing *"one little snippet"* (P15) of the original content.

Twelve (P01, P02, P06, P07, P09, P11-P13, P15-P18) participants proposed the idea of *screenshot control*, either through notifications or blocking mechanisms. Many felt that screenshot notification would provide an *"extra layer of security"* and a means to *"confront [other users]"* (P14) if necessary. However, some noted drawbacks, such as feeling uncomfortable with Snapchat's feature of notifying others about accidental screenshots (P13). P09 mentioned that knowing someone took a screenshot of their post could be unsettling, leading to thoughts that *"mess up [their] day."* Some participants preferred that screenshots be entirely blocked, similar to how services like Disney Plus or Netflix prevent screenshots by blacking out the screen (P07). Others suggested that screenshots should *"disappear"* when shared, preventing others from saving them (P06). Interestingly, some teens acknowledged scenarios where they did not mind screenshots being taken, as these could serve as a form of validation, for instance, if something was *"funny"* (P11) or captured special *"moments"* (P18) with friends. Therefore, they proposed having more control over the screenshot function, such as a default block with the option to *"allow screenshots"* when desired (P16). This approach would enable users to decide when to permit screenshots, balancing privacy with the desire to share certain content more freely.

Four participants (P01, P08, P11, P14) suggested the concept of *conditional ephemerality* to address concerns about boundary violations in messaging. They appreciated Snapchat's feature of ephemeral messages with the option to save them and proposed similar functionalities for other messaging platforms. According to them, ephemerality provides a sense of comfort for *"just in case... something happens"* (P14), yet they acknowledged situations where permanent messages are preferable. For instance, receiving a message *"at 6 am"* and wanting to revisit it *"at 9 am"* requires the messages to be available for reference and memory (P11). Additionally, they noted that in scenarios involving *"issues or drama,"* including *"legal issues,"* the ephemerality of messages complicates the process of tracing back conversations and can lead to challenges in proving claims (P08). Therefore, they proposed a more flexible messaging system that allows users to choose between ephemeral and permanent messages based on the context and necessity.

Time collapse. Ten participants (P02, P03, P06, P07, P13, P15-P19) raised concerns about digital spaces creating a *"time collapse,"* where the past, present, and future merge [20]. Participants noticed a shift in their privacy boundaries over time, making them uncomfortable with their earlier posts, such as *"random posts"* (P06) and *"every single thought"* (P13), which they no longer wish to be shared on their social media. Additionally, there was an apprehension about the long-term impact of their current online activities. For example, P13 expressed, *"Digital footprints... I was taught that they will come back to haunt you and you see it nonstop. ... I don't want the things I'm doing now as a teenager to haunt me in 10 years."* P19 added, *"anyone, including potential employers or universities, might see it in the future."* These concerns were particularly challenging to address, as participants felt the need to anticipate judgment by a collapsed time of the future with potentially vastly different cultural or social norms. As P07 shared, *"I might not like the way I look in two years or something like that. Or I might feel that that was kind of like a weird thing to post in a couple years."*

Six participants (P01, P11, P14, P15-P17) expressed the importance of being able to easily *revisit and reverse* content, such as through deleting or archiving posts. This preference stems from the fact that non-ephemeral posts *"last longer and are more visible,"* leading to concerns that followers who join later can *"go back and see what [they were] doing like two years ago"* (P17). While many participants actively engaged in revisiting and reversing their content, they also felt that the process could be more user-friendly. For instance, P17 pointed out the inconvenience of deleting comments on

Instagram, describing the process as *"really annoying."* An additional suggestion from P16 was to receive a notification a few days after posting, asking *"Do you want to delete this post? It's been two days. Do you want to delete this post, or are you comfortable with it staying here for an extended period of time?"* This feature would provide users with a timely reminder to reassess their comfort with the content's continued visibility.

Eleven participants (P02, P03, P05-P07, P09, P13, P14, P16-P18) also proposed the idea of content *ephemerality*, viewing it as a form of *"assurance that nobody's really scrutinizing it too much,"* given that *"it's not something we can go back and look at later"* (P05). However, since ephemerality is already a feature in many platforms, and the participants generally expressed satisfaction with their experiences, we decided not to develop a prototype for this particular design idea in our design evaluation survey.

4.1.2 Fear of Online Hostility. Another frequently mentioned source of anxiety among the participants was the general unpredictability and perceived risks within social media environments, leaving them vulnerable and anxious as they navigated these platforms. To address these concerns, participants suggested enforcing robust community standards, enhancing security measures, and implementing features like pseudonymity to mitigate online hostility and cybersecurity breaches.

Hostile environment. Another source of fear that fourteen (P01-P08, P10, P12, P13, P15, P17, P18) of the participants mentioned was concerns about the chaotic or negative vibe of a specific platform, which often created a sense of discomfort or fear of privacy issues due to the overwhelming and often hostile environment. P18, for example, voiced concerns about encountering *"not very good influencers"* on Snapchat, contributing to a generally *"not so positive"* vibe. Although the level of disorder on specific platforms may not be directly related to privacy concerns, it still prevented participants from feeling *"safe"* (P18), thereby indirectly increasing their sense of vulnerability. This contrasts with P08's experience on Pinterest, where they felt *"a lot safer"* due to the absence of *"malicious"* users and an environment focused on *"look[ing] at pretty aesthetic things."* The teens' concerns were exacerbated by their perception that certain platforms do not prioritize user safety and fail to adequately regulate such disorders. P15's comment highlights this sentiment: *"Twitter's a hellscape and it's awful and especially with the new leadership people get away with harassment really really easily."*

Twelve participants (P01-P03, P05, P07, P09, P10, P13, P15, P17-P19) expressed a desire for more robust enforcement of *community standards* on social media platforms. They observed considerable *"risks"* (P18) and *"negativity"* (P13) on major platforms like Instagram and Twitter and believed that mitigating these issues was crucial. One method they proposed was to regulate negativity by enabling users to report others and restricting toxic users. For instance, P02 suggested that *"a third party could review and decide to remove a rude comment or prevent the user from posting for 30 days."* P10 offered another approach, noting that negative comments often get posted on platforms like Instagram or Twitter without any hindrance. They advocated for a more preventative approach where such comments would be blocked from posting. P17 raised concerns about the complexity of privacy policies and the necessity for clearer enforcements, noting that *"teenagers don't really read those because they're long and complicated."* They suggested a simpler, more frequent reminder system, such as notifications when posting or commenting, to remind users to *"make sure this is appropriate"* and to contribute to *"keeping a good community."*

Cybersecurity breach. Nine participants (P01, P02, P04-P06, P09, P13, P14, P17) expressed concerns about online threats that stem from unauthorized access or misuse of personal digital information. Much of this fear arose from indirect victimization experiences. P14 shared a specific instance: *"A couple years ago... one of those online therapy*

websites that you pay for and the people got access to all the records and were able to like blackmail people. They were like if you don't give us money, we'll tell your relative you don't like them that kind of thing. And while I don't think it would be that extreme for a group of teenagers, sometimes there's just that chance of somebody that just for some reason didn't like you decided to get into your phone or your account and just like went and did bad things." Many participants mentioned various problems such as *"data leak"* and *"impersonation"* (P09), *"doxxing"* and *"bots"* (P06), and *"child trafficking and... child pornography"* (P13). While they were concerned, the perception that these issues *"happen to everyone"* made certain participants, like P05, feel *"just not really comfortable"* while others felt these threats were beyond their control, resigning to the belief that *"it just happens"* (P09).

Five participants (P02, P03, P09, P11, P15) suggested that *enhanced security*, including options like *pseudonymity*, could alleviate their cybersecurity breach concerns. They appreciated existing security measures, such as mandatory password changes when unusual login activities are detected and the use of two-factor authentication (P11). The ability to remain pseudonymous was particularly valued. P02 liked not having their account connected to a phone number, reducing stalker concerns. P09 preferred changing their username on different platforms for anonymity, and P15 appreciated the anonymity on Twitter, which allows interactions with a familiar circle while maintaining privacy from others.

4.1.3 Fear of Personal Privacy Missteps. Participants also expressed concerns about unknowingly engaging in actions on digital platforms that could lead to personal risks or unintended negative consequences. To address these concerns, participants suggested implementing prompts for self-regulation to prevent overexposure and clarifying privacy norms on platforms to help navigate and enforce proper privacy boundaries.

Overexposure. Six participants (P05, P08-P10, P13, P14) expressed worries about intentionally or unintentionally revealing too much personal information or aspects of their lives on digital platforms, leading to a loss of privacy and control over their digital footprint. The concern often revolved around 'oversharing,' with some, like P08, unsure about *"how much of [their] life is like appropriate to share."* At times, this overexposure was facilitated by the affordances of social media platforms that make sharing easier. For instance, P05 pointed out that on BeReal, it is *"concerning"* how *"really easy to accidentally make your BeReal like public"* and inadvertently share their location. P14 shared a similar experience with Snapchat, where they *"just send like random stuff on accident like even just like of [their] pocket or something it'll get bumped."* These accidental shares sometimes resulted in disclosing information intended to be more private.

Six participants (P02, P05, P07, P08, P12, P17) suggested that *prompts for self-regulation* could help them avoid overexposing or oversharing online. For instance, P12 noticed they were more cautious on Instagram than on Snapchat, as Instagram requires *"clicking at least two or three buttons"* to post, encouraging more thoughtfulness. P07 proposed a feature that asks *"Are you sure you want to post this?"* and displays the user's followers as a reminder. P08 agreed, saying such prompts provide a moment to reconsider: *"Do I really want to share this? Will I regret this in the long run?"* This offers a chance to reassess one's confidence in what they're about to share. P12 also mentioned the idea of *"daily posting limits"* to prevent *"oversharing"*.

Overstepping privacy norms. Fifteen participants (P02, P03, P05-P09, P11-P14, P16-P19) expressed apprehension regarding the inadvertent overstepping of subtly established privacy expectations on social media platforms, which could potentially result in discomfort among other users or friends. Although these concerns are not privacy issues in themselves, they led to difficulties in navigating proper privacy management. For example, many participants felt

pressured to add people against their will, as P03 mentioned, *"pressured to add people [are] not that close to your account."* They also found it challenging to remove friends they were uncomfortable with because, as P09 noted, *"then you know you're like the aggressor in that situation."* Furthermore, even when desiring privacy boundaries, participants found it awkward to enforce them, especially with friends. P02 shared that they found it *"awkward to say that you don't want things to be screenshotted,"* and P16 found it *"awkward to set boundaries."* These privacy norms were often implicit and varied across participants and their individual friends. However, most were highly vigilant about not overstepping any privacy norms, and the possibility of overstepping these norms and the resulting inability to enforce privacy due to fear of doing so were seen as difficult to avoid.

Thirteen participants (P01-P05, P07, P10, P11, P13-P17) emphasized the need for platforms to *clarify privacy norms,* functioning as an *"air cover."* P02 underscored the importance of platforms stating explicitly that it is okay to reject friend requests, providing *"reassurance"* in complex social situations. P13 recommended adding disclaimers such as, *"Always make sure you know who you're following. You're always in control. You don't have to add someone if you don't want to,"* empowering users to make informed choices. P15 highlighted the need for platforms to clearly explain their privacy policies, thereby fostering a *"culture"* that prioritizes user privacy and safety and encourages users *"to set boundaries,"* as P03 stated. The participants observed that current platforms fail to communicate these norms effectively. P05 noted that people often inadvertently breach privacy because it is *"assumed to be common knowledge."* However, they remarked, *"it's kind of been proven that that's not necessarily true,"* highlighting a discrepancy between assumed understanding and actual adherence to privacy norms. In essence, clarifying privacy norms and empowering users to use these guidelines as a rationale to prioritize their privacy, especially in situations where it might seem inappropriate or awkward, can be helpful.

4.1.4 Vague and Persistent Fear. Many participants (P02-P04, P07-P10, P13-P16, P18, P19) often described experiencing non-specific, underlying feelings of threat or unease related to their general interactions on social media without being able to pinpoint a clear source. For instance, P03 mentioned, *"I don't post my face right now, but I have an inner monologue that keeps telling me what I'm doing could get me into trouble. It's really just paranoia."* Some expressed general concerns about social media usage, influenced by their parents' cautiousness. P04 noted that their parents are *"very cautious"* about social media and advised them that they should *"[not] get social media at all"* (P04). Media portrayals also fueled such fears. Despite perceiving that in the movies they are *"overdramatized"* and *"very exaggerated"* (P04), they still harbored significant concerns, particularly on public accounts where *"there's a lot of weird stuff going on, like predators"* (P14). Since the sources of these fears were elusive, the teens experienced persistent anxiety. Even after implementing all known mitigation strategies and precautions, a vague and lingering anxiety remained, leaving them uncertain about how to alleviate this fear, suggesting the possibility of dysfunctional fear.

While the source of vague fear cannot be specified or directly mitigated, five participants (P07, P09, P13, P17, P18) shared design ideas that would provide them with *reassurance* about their privacy fear in general. P13 proposed making it simpler for users to verify their privacy settings, suggesting, *"So after I post, if I'm unsure whether I included someone in the privacy settings, I should be able to check it quickly, in just two or three clicks, rather than navigating through settings. It should be more straightforward."* P17 emphasized the importance of control over one's profile, *"Especially with your own profile, having a lot of control over what people post and how you express yourself"* contributes to this reassurance. P18 agreed, noting the benefits of options like *"closing my comments or likes"* to make things *"more private for yourself,"* which they found to be a *"huge reassurance."* Additionally, P13 suggested that platforms could enhance feelings of safety by sharing positive affirmations, such as *"you are safe, you are loved, and we care about you."*

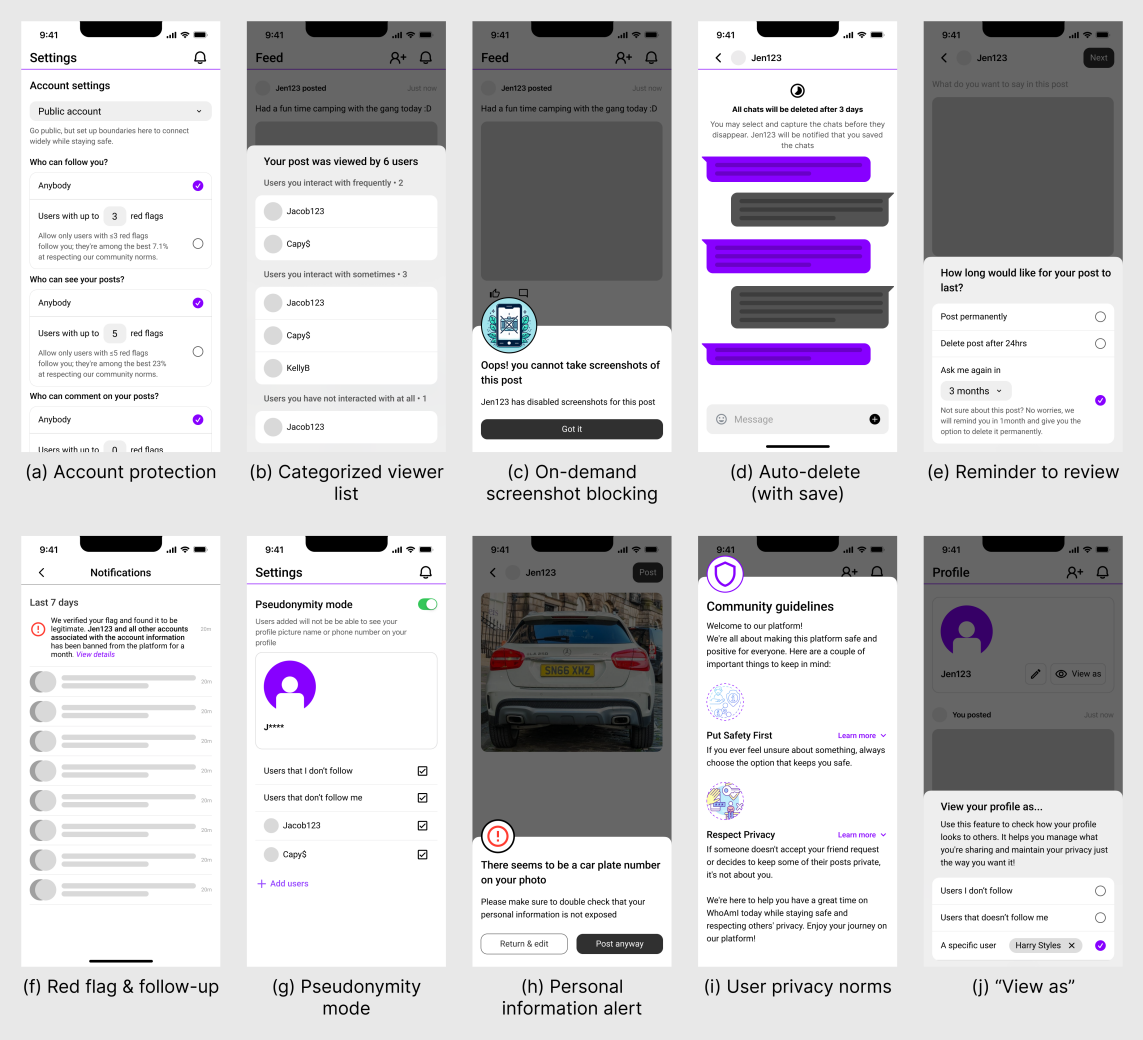


Fig. 1. Sample screens of prototypes of the ten design approaches derived from the co-design study with teen participants. Survey respondents were provided additional screens for clarity. A complete taxonomy of the designs is available in Table 1.

4.2 How Does Design Affect Teens' Privacy-related Fears?

We developed ten prototypes corresponding to each of the ten design ideas suggested by teens during our co-design interviews. Although there are various ways to implement the specific features of each design idea, we chose to use prototypes as design probes instead of conceptual design ideas. To elicit more concrete feedback from teens, we presented specific implementations of our design ideas. While these examples represent just a few possible approaches, they facilitated clearer understanding and discussion. The core rationale behind our designs was to empower users with enhanced privacy control, offering them more choices and greater autonomy than existing features on mainstream platforms or those explored in prior research. Our aim was to ensure users feel empowered and in control of their privacy,

Table 1. A taxonomy of design ideas generated in co-design study. The example feature name and description are shown to the teens during the design evaluation survey.

Design Approach	Example Feature	Feature Description
Boundary control	Account protection (Fig. 1(a))	"Provides extra security layers for your public profile, keeping your info safe while you connect with more people."
Interaction transparency	Categorized viewer list (Fig. 1(b))	"Sorts your audience based on how often you interact, so you can decide if less frequent contacts should stay on your follower list."
(Conditional) Ephemerality	Auto-delete (with save) (Fig. 1(c))	"Automatically erases your content after a set time, but gives you a heads-up to save them if you want to keep them."
Context control	On-demand screenshot blocking (Fig. 1(d))	"Stops others from taking screenshots of your posts if you want to, keeping your shared moments private."
Revisit/reverse	Reminder to review (Fig. 1(e))	"Reminds you to look back at old posts and reassess if they're still appropriate or reflect who you are today."
Community standards	Red flag & follow-up (Fig. 1(f))	"Reports breaches of community rules and get notified when action is taken, ensuring your concerns are addressed."
Enhanced security	Pseudonymity mode (Fig. 1(g))	"Completely disconnects your real identity from your account, making it untraceable to certain people you choose, for ultimate privacy."
Prompts for self-regulation	Personal information alert (Fig. 1(h))	"Alerts you if you're about to share personal info, helping you think twice about your privacy."
Clarification of privacy norms	User privacy norms (Fig. 1(i))	"Guides for safe online sharing, reminding you that being privacy-conscious is smart, not odd."
Safety reassurance	"View as" (Fig. 1(j))	"Lets you see your profile through someone else's eyes, so you can be sure you're sharing only what you intend to."

thereby addressing dysfunctional worries. Importantly, we focused on implementing actual protective measures rather than creating privacy theater, thus providing substantive safeguards while alleviating vague and/or lingering anxiety. A comprehensive list of design strategies, sample prototypes, and detailed descriptions provided to participants during the follow-up design evaluation survey are available in Table 1. Additionally, example screenshots of all prototypes are displayed in Figure 1[2].

Overall, teens showed high interest in the prototypes we shared throughout the design evaluation survey. As illustrated in Figure 2, follow-up survey respondents expressed interest in and positive reactions to all features. While some prototypes were preferred more than others, one-sample t-test results revealed that the mean responses to the questions 'My reaction to this feature is,' 'I would be interested in trying this feature on my [public/private] account,' 'With this feature, I would feel less worried about privacy-related problems on my [public/private] account,' and 'With this feature, the likelihood of privacy-related problems on my [public/private] account will decrease,' were significantly higher than 3 ('Neutral' and 'Neither agree nor disagree,' respectively) for all ten features. Conversely, one-sample t-test

[2]The license plate number shown in (h) is a fictional placeholder, not an actual vehicle registration.

results showed that the mean responses to the questions 'This feature might make my overall experience using my [public/private] account worse,' 'I find this feature to be annoying,' and 'If I use this feature on my [public/private] account, I'm concerned it might lead to awkward or uncomfortable situations with my friends or people I know in real life' were all significantly lower than 3 for all features. [3]

Furthermore, as seen in Figure 3, with the features applied together, it would decrease their fear in all the different types of fear we identified above. This revealed that the prototypes were generally perceived as possibly effective in addressing dysfunctional fear, as they decreased the perceived risk and the actual risk while not revealing many new tradeoffs, such as causing discomfort or potential problems with other users. Still, through free response questions, they shared some of their concerns and areas for improvement, which we explore below. Given the overwhelmingly positive reactions to the Likert-scale questions, *we deliberately focused on examining the potential pitfalls of these features* to provide a balanced perspective and highlight possible trade-offs or concerns associated with implementing these privacy measures.

Account protection. As an example of *boundary/visibility control*, we designed a feature (Fig. 1(a)) to address issues with hostile interactions from strangers while balancing users' concerns about limiting their audience against the opportunity to reach a larger network. This feature allows users to experiment with and choose their own balance in this trade-off by filtering other users based on the number of warnings they've received on the platform. This empowers privacy-conscious users to implement stricter audience limitations, while those prioritizing network expansion can minimize restrictions.

Participants responded very positively to this feature, as evidenced by the following results:

- Overall positive reaction: $(mean = 4.27, sd = 0.890, t(135) = 16.7, p < .0001, d = 1.43)$
- Interest in trying the feature: $(mean = 4.09, sd = 0.954, t(135) = 13.3, p < .0001, d = 1.14)$
- Belief that the feature would reduce privacy worries: $(mean = 4.09, sd = 1.00, t(135) = 12.7, p < .0001, d = 1.09)$
- Expectation of decreased privacy concerns: $(mean = 3.91, sd = 1.04, t(135) = 10.2, p < .0001, d = 0.874)$

Importantly, participants also:

- Strongly disagreed that the feature would worsen their overall social media experience: $(mean = 1.92, sd = 1.01, t(135) = -12.5, p < .0001, d = -1.07)$
- Did not find the feature annoying: $(mean = 1.84, sd = 0.929, t(135) = -14.6, p < .0001, d = -1.25)$
- Moderately disagreed that the feature would cause awkward situations with friends: $(mean = 2.22, sd = 1.18, t(135) = -7.72, p < .0001, d = -0.662)$

Specifically, participants acknowledged its potential as *"a good step in between private and fully public"* (R056) and its empowerment of users by granting them enhanced *"control"* (R124) and *"security"* (R118). However, feedback also included concerns about the accuracy and transparency of the red flag system. Suggestions were raised for clearer definitions and criteria for red flagging, as well as a more nuanced system where flags could expire or become *"gray flags"* (R005) over time. Questions were raised about the practicality of quantifying red flags, accompanied by requests for manual override options to maintain connections despite red flags. Private account users noted potential redundancy with their existing selective audience management. Additionally, concerns were expressed regarding the possibility of users circumventing the system by creating new accounts.

[3]Despite the hypothetical nature of several questions, participants consistently rated their confidence in their responses to the question 'How confident are you about the responses you gave above?' as nearly 5 ('Very confident'), which is significantly higher the neutral rating ('Moderately confident').

Categorized viewer list. Our *interaction transparency* feature (Fig. 1(b)) builds upon existing functionalities like Instagram's Story viewer list but addresses a key limitation identified by participants: lengthy user lists often fail to provide complete audience transparency [33] or awareness of potentially unwanted profile interactions. To enhance this, we categorized viewers based on the closeness of their interactions with the user. This approach helps users identify unwanted interactions without compromising the positive feedback they receive from seeing that close contacts view their profiles.

Participants responded very positively to this feature and expressed a moderately high level of confidence in its effectiveness for enhancing privacy:

- Overall positive reaction: ($mean = 4.10, sd = 1.03, t(135) = 12.4, p < .0001, d = 1.07$)
- Interest in trying the feature: ($mean = 3.92, sd = 1.22, t(135) = 8.76, p < .0001, d = 0.751$)
- Belief that the feature would reduce privacy worries: ($mean = 3.68, sd = 1.06, t(135) = 7.53, p < .0001, d = 0.646$)
- Expectation of decreased privacy concerns: ($mean = 3.54, sd = 1.07, t(135) = 5.83, p < .0001, d = 0.500$)

Moreover, participants also:

- Strongly disagreed that the feature would worsen their overall social media experience: ($mean = 1.92, sd = 1.01, t(135) = -12.5, p < .0001, d = -1.07$)
- Did not find the feature annoying: ($mean = 1.84, sd = 0.929, t(135) = -14.6, p < .0001, d = -1.25$)
- Moderately disagreed that the feature would cause awkward situations with friends: ($mean = 2.22, sd = 1.18, t(135) = -7.72, p < .0001, d = -0.662$)

The feature was well-received, primarily because it enables users to more easily detect an *"unwanted person"* (R048) viewing their account. By organizing viewers into categories, users can quickly scan for anomalies or unexpected interactions, enhancing their sense of control and awareness. Nevertheless, concerns were expressed that it might become *"almost too intricate of a feature to enjoy the fun in posting even if the intention is to protect privacy"* (R013) and that it would make it *"too easy to get caught up in looking at who viewed your posts"* (R020). Suggestions to enhance the feature included sharing additional information, such as the *"day and time last interacted"* (R046). Participants also proposed an option to toggle the feature on and off, acknowledging that this level of transparency might be uncomfortable for some users.

Auto-delete with save. This feature (Fig. 1(c)) was designed for *conditional ephemerality*, providing automatic deletion while giving users the option to save posts. It also alerts other users when conversations are saved, addressing potential privacy concerns when the default ephemerality is overridden. This feature elicited the most diverse responses among participants. Many appreciated its ability to *prevent people from bringing up old things"* (R067), particularly in protecting privacy from *strict parents"* (R069). The notification system for saved conversations without *"asking for consent"* (R016) was particularly valued.

However, concerns were raised about losing the benefits of permanent content, as it can be *fun looking back at old convos"* (R006), or it might make retrieving information from the chat difficult (R018). Feedback highlighted issues such as potential *bullying"* (R125) and the lack of *evidence from the chat"* when needed (R115). Suggestions included options to *turn it on or off"* (R077), or the ability to *"never delete chats with friends [they] talk to often"* (R124).

Quantitative data revealed generally positive responses:

- Moderately positive reaction: ($mean = 3.43, sd = 1.34, t(135) = 3.77, p < .001, d = 0.323$)
- Moderate interest in trying the feature: ($mean = 3.40, sd = 1.40, t(135) = 3.37, p < .001, d = 0.289$)

- Moderate belief that the feature might reduce privacy worries: ($mean = 3.43, sd = 1.25, t(130) = 3.92, p < .001, d = 0.343$)
- Moderate expectation of decreased privacy concerns: ($mean = 3.32, sd = 1.22, t(135) = 3.03, p = .003, d = 0.260$)

Moreover, participants also:

- Moderately disagreed that the feature would worsen their overall social media experience: ($mean = 2.49, sd = 1.27, t(135) = -4.72, p < .0001, d = -0.405$)
- Felt the feature is not too annoying: ($mean = 2.75, sd = 1.39, t(135) = -2.10, p = .037, d = -0.180$)
- Did not strongly agree that the feature would cause awkward situations with friends: ($mean = 2.76, sd = 1.36, t(135) = -2.08, p = .040, d = -0.178$)

On-demand screenshot blocking. This feature (Fig. 1(d)) addresses teens' complex relationship with *screenshot control*. Our participants expressed nuanced views on screenshots, informed by their experiences with existing apps like Snapchat that notify users when screenshots are taken. They acknowledged that screenshots can threaten privacy by enabling out-of-context sharing, but also recognized legitimate reasons for taking them, such as preserving positive content. Participants noted that accidental screenshots can lead to awkward situations due to automatic notifications, and some preferred not to know when screenshots are taken to avoid overthinking. Given this range of preferences, we designed a feature allowing users to disable screenshots for individual posts. This approach aims to provide flexible control over content sharing, shape platform norms by conveying that screenshots are not always acceptable, and respect the original poster's intentions.

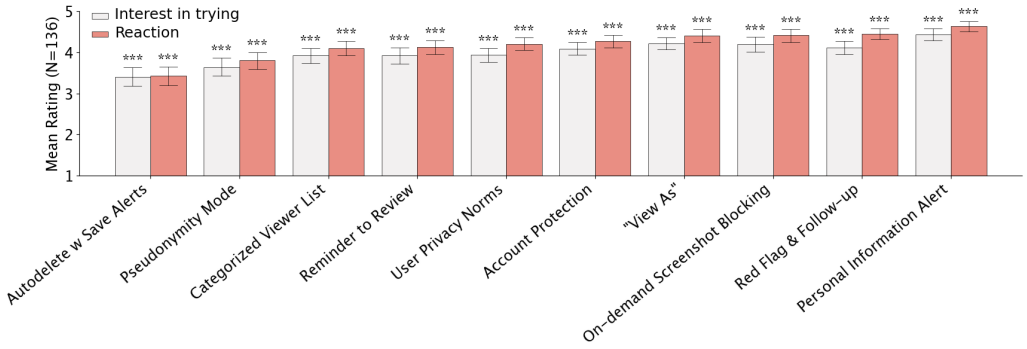
As expected, this feature was appreciated for its ability to block screenshots and for offering the option to apply this *"not to your whole profile, but just per post"* (R069). Concerns were raised about its effectiveness, noting potential circumvention by users who might use *"a second device or find a way to screenshot/screen record anyway, as many do with Snapchat"* (R078). Feedback included worries about potential exploitation for malicious purposes, with the possibility for *"some people to use this feature to post cruel or harmful things without risk of them being saved and/or shared"* (R029). While addressing such behind-the-scenes betrayal directly through social media design is challenging, this feedback highlights the importance of clarifying privacy norms so that users feel empowered to defend themselves when necessary.

Participants perceived this feature very positively and highly effective in addressing privacy concerns:

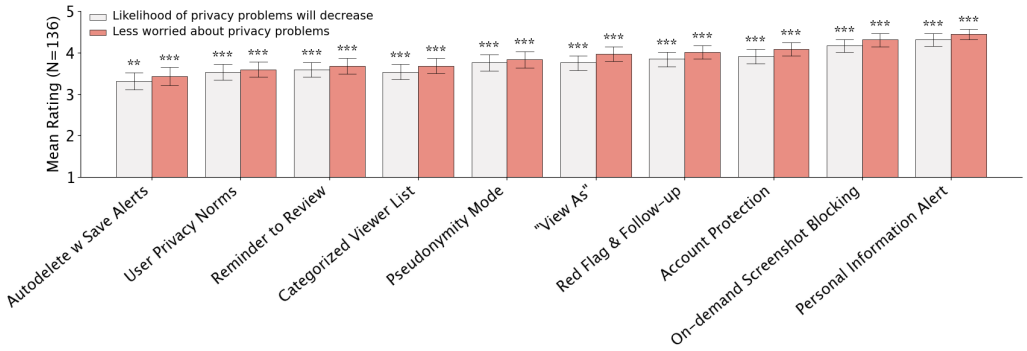
- Overall positive reaction: ($mean = 4.41, sd = 0.946, t(135) = 17.4, p < .0001, d = 1.49$)
- Strong interest in trying the feature: ($mean = 4.21, sd = 1.08, t(135) = 13.1, p < .0001, d = 1.12$)
- Strong belief that the feature might reduce privacy worries: ($mean = 4.32, sd = 0.948, t(135) = 16.2, p < .0001, d = 1.39$)
- High expectation of decreased privacy concerns: ($mean = 4.17, sd = 0.931, t(135) = 14.6, p < .0001, d = 1.26$)

Participants also indicated minimal drawbacks or trade-offs:

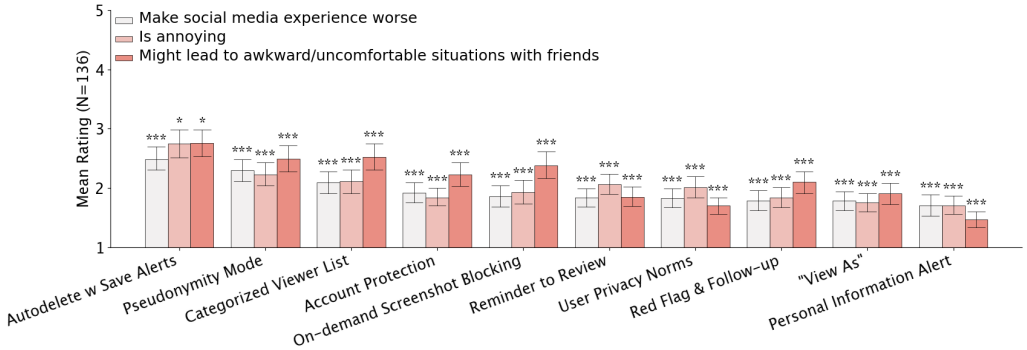
- They strongly disagreed that the feature would worsen their overall social media experience: ($mean = 1.85, sd = 1.08, t(135) = -12.4, p < .0001, d = -1.06$)
- They did not find the feature annoying: ($mean = 1.93, sd = 1.20, t(135) = -10.5, p < .0001, d = -0.897$)
- They disagreed that the feature would cause awkward situations with friends: ($mean = 2.38, sd = 1.29, t(135) = -5.59, p < .0001, d = -0.479$)



(a) Interest and reaction toward features



(b) Privacy-related effectiveness of the features



(c) Level of potential trouble caused by features

Fig. 2. Bar graphs illustrating the mean rating (N=136; 72 responses from private account owners and 64 from public account owners; a total of 118 participants responded) of the ten design prototypes in the design evaluation survey. Each bar illustrates the mean rating of participants' responses to the questions, with standard error bars. The significance levels of the one-sample t-tests against the hypothesis that the mean rating is a neutral score are denoted using asterisks over the bar, where one asterisk denotes $p < .05$, two asterisks denote $p < .01$, and three asterisks denote $p < .001$.

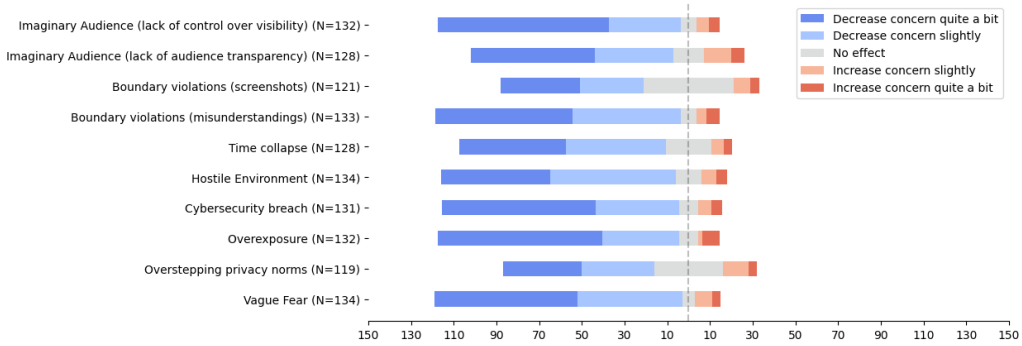


Fig. 3. A diverging bar graph illustrates the ratings (N=136) for responses to the question, "How would the features affect your [specific concern]?" Although we initially identified seven categories of concern, we subdivided the fear related to "imaginary audiences" and "boundary violations" into two separate categories and added a category for vague fears, thereby totaling ten distinct types of concern.

Reminder to review. This feature (Fig. 1(e)) facilitates *revisit and reverse*, addressing the dynamic nature of teens' identities and evolving social norms. It prompts users to reassess past content at different time intervals, recognizing that teens' perspectives on their posts may change over time. By sending notifications, the feature eliminates the need for manual searching and reduces the likelihood of overlooking old content. This approach empowers users to make informed decisions about keeping or modifying past posts, encouraging regular reflection on their digital footprint while balancing the preservation of meaningful memories with the curation of their current online presence.

The feature was well-received by participants for helping users to *"stay current with [their] beliefs and values"* (R039) by allowing them to *"reconsider or delete posts that reflect 'the old me'"* (R025). It was praised for providing an opportunity for *"self-reflection"* and *"reassurance"* (R132). However, there were concerns about the potential for notifications to become *"annoying"* (R018). Suggestions for improvement included providing *"tips on what should stay up and what should not"* (R002) and adding an *"archive option alongside a delete option"* (R026). A participant who identified themself as an *"infrequent poster"* (R030) felt the feature was unnecessary, as they are likely aware of their social media content. Additionally, there was worry that the feature might inadvertently remind users of *"things made in the past that you may not be comfortable with"* (R099). Again, a desire for the ability to *"toggle this on and off"* (R030) was highlighted.

The reactions were confirmed in our quantitative data:

- Overall positive reaction: $(mean = 4.13, sd = 1.01, t(135) = 13.0, p < .0001, d = 1.12)$
- Strong interest in trying the feature: $(mean = 3.92, sd = 1.19, t(135) = 9.03, p < .0001, d = 0.775)$
- Moderate belief that the feature might reduce privacy worries: $(mean = 3.68, sd = 1.15, t(135) = 6.84, p < .0001, d = 0.586)$
- Moderate expectation of decreased privacy concerns: $(mean = 3.59, sd = 1.08, t(135) = 6.36, p < .0001, d = 0.546)$

Participants also reported minimal drawbacks:

- Disagreement that the feature would worsen their social media experience: $(mean = 1.83, sd = 0.931, t(135) = -14.6, p < .0001, d = -1.26)$
- Low perception of the feature as annoying: $(mean = 2.06, sd = 1.12, t(135) = -9.79, p < .0001, d = -0.840)$

- Disagreement that the feature would cause awkward situations with friends: ($mean = 1.85, sd = 0.942, t(135) = -14.3, p < .0001, d = -1.23$)

Red flag and follow-up. This feature (Fig. 1(f)) was designed to assist with the enforcement of *community standards*, addressing participants' skepticism about mainstream social media apps' commitment to user privacy. The feature was appreciated for providing follow-up, which gives users *"peace of mind"* (R026) as intended. However, concerns were raised about the validity and feasibility of the "red flag" feature. As R045 pointed out, *"If this [hypothetical] platform becomes successful, you need to realize the number of people on the platform. The report system would get overwhelmed and not do its job. Even with an AI moderator, so many people would get wrongfully flagged. Additionally, it's so easy nowadays to get something like a fake email address or a VOIP phone number, making it appear as if different people are reporting. Even if you based it on IP address, VPNs and alternate devices exist."* Other worries included the potential for the feature to be *"abused"* (R102), with people flagging *"just to be annoying"* (R038). While detailing the operation of the red flag system is beyond the scope of this study, the feedback highlights the importance of the safety reassurance that the system provides users.

Despite the concerns, the feature was very well received, as corroborated in our quantitative data:

- Overall positive reaction: ($mean = 4.45, sd = 0.806, t(135) = 21.0, p < .0001, d = 1.80$)
- High interest in trying the feature: ($mean = 4.11, sd = 0.979, t(135) = 13.2, p < .0001, d = 1.13$)
- Strong belief that the feature might reduce privacy worries: ($mean = 4.01, sd = 0.947, t(135) = 12.4, p < .0001, d = 1.06$)
- Expectation of decreased privacy concerns: ($mean = 3.85, sd = 0.995, t(135) = 9.91, p < .0001, d = 0.850$)

Participants also reported minimal drawbacks:

- Strong disagreement that the feature would worsen their social media experience: ($mean = 1.79, sd = 0.977, t(135) = -14.5, p < .0001, d = -1.24$)
- Very low perception of the feature as annoying: ($mean = 1.84, sd = 1.01, t(135) = -13.4, p < .0001, d = -1.15$)
- Disagreement that the feature would cause awkward situations with friends: ($mean = 2.10, sd = 1.12, t(135) = -9.31, p < .0001, d = -0.798$)

Pseudonymity mode. This feature (Fig. 1(g)) aims to *enhance security* by allowing users to detach their real-world identity from shared content while maintaining social connectivity. The pseudonymity mode can be selectively applied to non-followers, users not followed, or specific individuals. This granular control enables teens to balance privacy protection with social connections.

This feature received mixed reactions. It was praised for enabling users to *"manage your social media presence completely without having to block other users"* (R101), offering an additional *"layer of protection,"* (R039), especially for public posts. However, there were concerns that it might encourage people to *"express harmful opinions more comfortably"* (R003), and some felt that *"just blocking the user"* might be a simpler solution (R108). Additionally, participants brought to attention that the feature could be *"used maliciously, such as by anonymous users viewing others' posts without being identified"* (R014). There were also worries that this mode might make it *"harder to gain more followers"* (R002) as it is *"harder for someone to find you"* (R050). Concerns were also expressed about whether the term *"pseudonymity"* would be familiar to teens. Moreover, reluctance towards pseudonymous communication was noted, with sentiments like *"you shouldn't be on social media if you are trying to hide your identity"* (R044). This feedback highlights that privacy features

can sometimes exacerbate privacy concerns and that users' preferences and views vary greatly, emphasizing the need for having the features be optional.

The mixed reactions to this feature were reflected in our survey data:

- Overall positive reaction: ($mean = 3.80, sd = 1.23, t(135) = 7.61, p < .0001, d = 0.653$)
- Interest in trying the feature: ($mean = 3.64, sd = 1.30, t(135) = 5.72, p < .0001, d = 0.491$)
- Belief that the feature might reduce privacy worries: ($mean = 3.83, sd = 1.16, t(135) = 8.37, p < .0001, d = 0.717$)
- Expectation of decreased privacy concerns: ($mean = 3.76, sd = 1.19, t(135) = 7.43, p < .0001, d = 0.637$)

Despite concerns about potential drawbacks, participants generally did not anticipate significant negative impacts:

- Disagreement that the feature would worsen their social media experience: ($mean = 2.29, sd = 1.16, t(135) = -7.09, p < .0001, d = -0.608$)
- Low perception of the feature as annoying: ($mean = 2.23, sd = 1.21, t(135) = -7.44, p < .0001, d = -0.638$)
- Mild disagreement that the feature would cause awkward situations with friends: ($mean = 2.49, sd = 1.27, t(135) = -4.66, p < .0001, d = -0.399$)

Personal information alert. This feature (Fig. 1(h)), designed to nudge for self-regulation, was recognized for helping users avoid *"overlook[ing] details they may have missed"* (R032), particularly beneficial for those who *"post more impulsively"* (R070). Concerns were raised about its accuracy and potential intrusiveness. Some users were apprehensive that it could *"sense things wrong and be annoying"* (R052), and there were privacy concerns regarding the perception of it *"checking what [they] post"* (R014). Suggestions for implementation included having the option to turn it *"on and off, but with the default setting as on"* (R011). Additionally, users thought the feature could be enhanced by allowing in-app content blurring. This feedback suggests that sometimes, a privacy issue associated with a privacy feature is overlooked because the benefit of the feature is significantly larger and more widely accepted.

This feature received the most positive reactions from teens:

- Overwhelmingly positive reaction: ($mean = 4.64, sd = 0.685, t(135) = 27.9, p < .0001, d = 2.39$)
- Strong interest in trying the feature: ($mean = 4.43, sd = 0.814, t(135) = 20.5, p < .0001, d = 1.76$)
- Strong belief that the feature might reduce privacy worries: ($mean = 4.44, sd = 0.796, t(135) = 21.1, p < .0001, d = 1.81$)
- Strong expectation of decreased privacy concerns: ($mean = 4.32, sd = 0.900, t(135) = 17.1, p < .0001, d = 1.46$)

Participants also strongly disagreed with potential drawbacks:

- Strong disagreement that the feature would worsen their social media experience: ($mean = 1.70, sd = 1.03, t(135) = -14.8, p < .0001, d = -1.27$)
- Very low perception of the feature as annoying: ($mean = 1.71, sd = 0.928, t(135) = -16.3, p < .0001, d = -1.39$)
- Strong disagreement that the feature would cause awkward situations with friends: ($mean = 1.46, sd = 0.816, t(135) = -22.0, p < .0001, d = -1.88$)

User privacy norms. Intended to *clarify privacy norms*, this feature (Fig. 1(i)) was positively received in general. Participants appreciated its guidance in online safety management, emphasizing that *"You aren't left to figure out how to protect your safety on your own"* (R041) and noting its role in *"help[ing] establish norms for the platform"* (R060). It was commended for making *"being cautious"*, not something to be *"ashamed or embarrassed of"* (R026) and for ensuring *"being respectful [is] normal"* (R047). Additionally, participants noted that with this feature, *"users won't be able to deny*

knowing the rules and policies" (R033). The feature's effectiveness in setting clear expectations was highlighted, described as *"short"* (R023) and able to *"very clearly but simply explain things"* (R032). However, some participants desired more specificity, wanting guidelines that are *"more specific to potential scenarios"* and offer *"more details of what constitutes 'safe"'* (R050). Concerns were raised that some people might *"just click off and not read or care about it"*; including *"a quiz of some sorts"* was suggested to confirm users' understanding and ensure they *"actually read over the guidelines"* (R035).

This feature was very well-received, with participants viewing it as highly necessary. However, they were less confident in its ability to address privacy concerns directly, highlighting the importance of shifting overall privacy norms:

- Overall positive reaction: ($mean = 4.20, sd = 0.987, t(135) = 14.2, p < .0001, d = 1.21$)
- Strong interest in trying the feature: ($mean = 3.93, sd = 1.03, t(135) = 10.6, p < .0001, d = 0.909$)
- Moderate belief that the feature might reduce privacy worries: ($mean = 3.59, sd = 1.09, t(135) = 6.28, p < .0001, d = 0.539$)
- Moderate expectation of decreased privacy concerns: ($mean = 3.54, sd = 1.11, t(135) = 5.61, p < .0001, d = 0.481$)

Participants also strongly felt the feature would have minimal negative impacts:

- Strong disagreement that the feature would worsen their social media experience: ($mean = 1.82, sd = 0.918, t(135) = -14.9, p < .0001, d = -1.28$)
- Low perception of the feature as annoying: ($mean = 2.01, sd = 1.13, t(135) = -10.2, p < .0001, d = -0.877$)
- Strong disagreement that the feature would cause awkward situations with friends: ($mean = 1.70, sd = 0.855, t(135) = -17.8, p < .0001, d = -1.52$)

"View as". Participants generally appreciated the feature (Fig. 1(j)) designed to enhance *safety reassurance*. Despite its similarity to Facebook's "View As" function, we included it in our survey as most teens reported being unfamiliar with Facebook during interviews. They found it particularly useful *"if you have certain things blocked from certain users"* (R012). The feature was perceived to be useful in allowing users to *"step' into the shoes of others so you can make sure that what you're changing about your account is what you want others to see"* (R062), eliminating *"the hassle of having to use different accounts to see what my main account looks like to others,"* which can be *"very tiring"* (R039). However, there were views that this feature might be redundant for private accounts as *"there aren't THAT many people too concerned with their private content"* (R044), and some felt that *"on Instagram, your profile already looks pretty much how others will see it"* (R087). Participant reactions to the feature underscored the importance of psychological reassurance regarding privacy and safety for some users, even if the feature does not directly provide additional control.

Respondents showed a strong positive reaction to this feature:

- Highly positive reaction: ($mean = 4.40, sd = 0.898, t(135) = 18.2, p < .0001, d = 1.56$)
- Strong interest in trying the feature: ($mean = 4.22, sd = 0.956, t(135) = 14.9, p < .0001, d = 1.28$)
- Belief that the feature might reduce privacy worries: ($mean = 3.97, sd = 1.03, t(135) = 11.0, p < .0001, d = 0.947$)
- Expectation of decreased privacy concerns: ($mean = 3.76, sd = 1.06, t(135) = 8.30, p < .0001, d = 0.711$)

Participants also strongly disagreed with potential negative impacts:

- Strong disagreement that the feature would worsen their social media experience: ($mean = 1.78, sd = 0.932, t(135) = -15.3, p < .0001, d = -1.31$)
- Very low perception of the feature as annoying: ($mean = 1.75, sd = 0.964, t(135) = -15.1, p < .0001, d = -1.30$)

Table 2. Assessment of potential drawbacks and/or trade-offs for each prototype.

Feature Name	Impact on Other Users	Utility Restriction	User Burden	Community-Dependent
Auto-delete (with save)	Yes	Yes	Yes	Yes
Reminder to review	No	No	Yes	No
Account protection	No	Yes	No	No
On-demand screenshot blocking	Yes	Yes	No	Yes
Personal information alert	No	No	No	No
User privacy norms	No	No	No	Yes
Categorized viewer list	No	No	Yes	No
"View as"	No	No	Yes	No
Red flag & follow-up	No	No	No	No
Pseudonymity mode	No	Yes	No	Yes

- Disagreement that the feature would cause awkward situations with friends: ($mean = 1.90, sd = 1.11, t(135) = -11.5, p < .0001, d = -0.989$)

5 DISCUSSION

5.1 Dysfunctional Fear and Privacy Norms

Social media is central to teen communication, with privacy crucial for their safety and well-being. However, the focus on reducing privacy risks has often led to increased fear in teens' social media use. This fear is exacerbated by adult narratives that, while intended to promote vigilance, often resemble frightening folklore lacking specific details. Such stories contribute to a persistent perception of uncontrollable risks among teens, potentially leading to withdrawal from social media or dysfunctional anxiety that diminishes the quality of life without enhancing safety.

Current social media designs for mitigating teen privacy risks emphasize self-regulation, placing significant responsibility on individuals. However, teens must navigate multiple, often conflicting privacy norms across various contexts. Contextual Integrity theory [60] suggests that privacy is maintained when information flows appropriately according to contextual norms. Yet, for teens, adhering to these norms is often compromised by conflicting social pressures and platform designs that don't account for nuanced contexts.

Our study provides empirical evidence of dysfunctional fear in teens' privacy management. Teens expressed vague but persistent fears about audience control and feeling watched, coupled with concerns about potential negative social consequences of protective actions. They face dilemmas between privacy and social reach, experiencing significant stress while struggling to find effective solutions.

We found that social media platform design can significantly influence whether teens feel empowered to take constructive privacy-related actions. Our findings highlight a critical limitation in current approaches: the overemphasis on individual choice fails to account for the powerful influence of social contexts and norms that can override.

5.2 A Design Agenda for Privacy Assurance

Participants perceived tremendous potential and proposed numerous ideas for improving the status quo. Their primary agenda for change involved enhancing safety reassurance through system-level approaches, offering more options

for users to tailor their privacy settings, and prioritizing privacy across the platform, thereby establishing privacy protection as the norm among users.

In our prototypes, based on interview data, we aimed to incorporate extensive user choice, allowing users to determine their own balance between utility and privacy. We enhanced existing features with nuanced controls, such as improved screenshot control and viewer lists supporting audience transparency. We also introduced new features like reminders to review past posts and reintroduced designs that teen participants were unfamiliar with, such as Facebook's "View As" function.

Based on our design evaluation results, we outline the following three key design recommendations:

Cultivate a Culture of Privacy. Teens often experience dysfunctional fear of privacy issues on social media, leading to a perceived lack of self-efficacy. Paradoxically, this can result in them surrendering their privacy, believing issues to be inevitable. This suggests that the current model of individual user control is insufficient, highlighting the need for platforms to provide both protection and reassurance to effectively address teens' privacy concerns.

To enhance teenagers' confidence in managing privacy, social media platforms should promote privacy as a social norm by visibly prioritizing it in design and features. This includes explicitly communicating privacy expectations and providing varied protection options. Implementing features that offer tangible safety reassurance, such as follow-up notifications on reported incidents, can further bolster user confidence. Platforms should emphasize the importance of privacy during onboarding and throughout the user experience, empowering users to prioritize their privacy even when it conflicts with peer norms. For instance, explicitly stating that users should not feel obliged to accept all friend requests can help teens navigate complex social dynamics.

This approach shifts the narrative, making privacy protection a standard practice rather than an option for only the most cautious users. By providing a supportive environment, platforms can help teens navigate complex social interactions while maintaining privacy, acknowledging the significant influence of peer perceptions during adolescence. This strategy allows teens to learn safe social media navigation through managed risk-taking, as suggested by previous research, ultimately fostering a more privacy-conscious and empowered user base.

Recognize Multifaceted Needs. Based on survey respondents' concerns, we identified four main areas of potential pitfalls of the privacy features: negative impact on other users, restriction of social media utility, user burden (e.g., features triggered without active engagement), and community-dependency (where individual engagement alone is insufficient to reduce privacy risks). We analyzed each prototype against these criteria, summarizing the results in Table 2. As anticipated, features without these drawbacks—Personal Information Alert and Red Flag & Follow-up—were most positively received. Conversely, Auto-delete (with save), which exhibited all four issues, was the least favored.

These findings underscore that effective designs must balance privacy enhancement with utility while considering impacts on user experience, social dynamics, and perceived effectiveness. First, privacy is secondary to engagement for teens on social media. Features that significantly impact utility or impose a high user burden are less likely to be embraced. Second, teens are highly attuned to peer dynamics and social norms, preferring features that do not burden their friends. Third, community-dependent features may be perceived as less effective, potentially due to their limited impact on individual self-efficacy or agency. These highlight the delicate balance required in designing privacy features for teens: they must enhance privacy without compromising the core social and functional benefits that draw teens to these platforms.

Support Diverse Needs. Our study revealed that teens have varied privacy requirements, particularly evident in the contrast between private and public account users. While private account holders prioritize strict privacy controls, public account users, such as aspiring artists, seek a balance between visibility and protection. To address this spectrum of needs, participants proposed design improvements that offer nuanced privacy options. For public accounts, suggestions included disconnecting personal information from public visibility, systematically filtering out users who fail to meet community standards, and enhancing interaction transparency through better-organized viewer lists. These features aim to enable public account holders to engage with a larger audience without fully compromising their privacy, demonstrating the need for flexible privacy solutions that cater to diverse user goals.

Participants also advocated for a range of privacy-enhancing features applicable to all users, reflecting the individual variation in privacy concerns. These included alerts for potential privacy compromises, systematic visibility controls like auto-delete or screenshot blocking, and the ability to view one's profile from others' perspectives. Recognizing that the perceived necessity and potential drawbacks of these features vary among individuals, participants emphasized the importance of user choice. They preferred the option to toggle features on and off, with privacy-focused settings as the default. This approach not only accommodates individual preferences but also reinforces privacy as the norm while allowing users to customize their experience.

5.3 Limitations and Future Work

We did not conduct a thematic analysis of the survey's free-response questions, so it is unclear which reactions were most prevalent among respondents. Additionally, our survey did not measure the varying intensities of different fears (e.g., whether some types of fear are more likely to lead to dysfunctional behavior than others), nor did it explore how different design ideas might affect each type of fear. We also did not investigate the distinct responses of public and private account holders to various fears and design ideas in our study. Further, our co-design interview participants were not limited to teens with a dysfunctional fear of privacy issues, nor were our design evaluation survey respondents. While this means that we did not explore features specifically targeting dysfunctional fear, the findings from our study are transferable to designs for dysfunctional fear. This is because the teen participants indicated that our features reduce fear and risks, critical elements in decreasing dysfunctional fear.

While we received numerous responses outlining the potential drawbacks of each feature, we utilized prototypes, meaning that in real-world and long-term scenarios, these features might encounter unforeseen issues. Using specific features as examples for each design idea made it easier for teens to understand and provide detailed feedback. However, there are multiple ways to implement these design ideas, and our findings may not be generalizable to all implementations. Although our design directions are intended to be broader concepts that can be adapted to different implementations, this is a limitation of our study. Another assumption in our prototypes was the feasibility and accuracy of user reporting systems, which is not currently a given. This assumption was necessary as solving this problem was beyond the scope of our research.

A significant challenge, consistent with prior research, is addressing fear in the context of real-life peer dynamics on social media disclosure. Future research should delve into how the content shared by teens might impact their social interactions (e.g., fear of judgment by peers based on a single post) and explore more effective solutions for this issue.

6 CONCLUSION

In this paper, we explored the privacy-related fears experienced by teens on social media and investigated how design can mitigate these fears without compromising privacy protection. Our study involved co-design interviews with 19 teens

and a design evaluation survey with 136 U.S. teen participants. Our findings indicate that teens frequently experience vague and persistent fears about privacy that they feel unable to manage effectively, leading to dysfunctional fear that inhibits their use of social media and negatively impacts their well-being. These fears fall into three primary categories: fear of uncontrolled audience reach, fear of online hostility, and fear of personal privacy missteps. Teens suggested various design improvements to address these fears, such as prompts for self-regulation and clearer communication of privacy norms. We developed ten concrete prototypes for these design ideas, some of them based on existing features, with an emphasis on increasing user agency over the status quo. Survey participants found these design prototypes helpful in reducing their privacy concerns and enhancing their sense of control on social media. Our study provides empirical evidence of dysfunctional fear in teens' privacy management on social media and offers guidelines to support user empowerment. By incorporating system-wide guidance, enforcement, and options to make privacy both the norm and a priority, social media platforms can help teens navigate complex social interactions while maintaining privacy.

REFERENCES

[1] [n. d.]. ChatGPT. https://openai.com/chatgpt/. Accessed: 2024-7-13.

[2] [n. d.]. sentence-transformers/all-mpnet-base-v2 · Hugging Face. https://huggingface.co/sentence-transformers/all-mpnet-base-v2. Accessed: 2024-7-15.

[3] [n. d.]. SentenceTransformers Documentation — Sentence Transformers documentation. https://sbert.net/. Accessed: 2024-7-15.

[4] [n. d.]. The Visual Collaboration Platform for Every Team. https://miro.com/. Accessed: 2023-7-18.

[5] 2023. *Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory*. Technical Report.

[6] 2024. ATLAS.ti. https://atlasti.com/. Accessed: 2024-1-15.

[7] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Comput. Surv.* 50, 3 (Aug. 2017), 1–41.

[8] Alessandro Acquisti and Ralph Gross. 2006. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *International workshop on privacy enhancing technologies*. Springer, 36–58.

[9] Michael Adorjan and Rosemary Ricciardelli. 2019. Student perspectives towards school responses to cyber-risk and safety: the presumption of the prudent digital citizen. *Learning, Media and Technology* 44, 4 (2019), 430–442.

[10] Zainab Agha. 2023. To Nudge or Not to Nudge: Co-Designing and Evaluating the Effectiveness of Adolescent Online Safety Nudges. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(Idc '23)*. Association for Computing Machinery, New York, NY, USA, 760–763.

[11] Zainab Agha, Karla Badillo-Urquiola, and Pamela J Wisniewski. 2023. "Strike at the Root": Co-designing Real-Time Social Media Interventions for Adolescent Online Risk Prevention. *Proc. ACM Hum.-Comput. Interact.* 7, Cscw1 (April 2023), 1–32.

[12] Denise E Agosto and June Abbas. 2017. "Don't be dumb–that's the rule I try to live by": A closer look at older teens' online privacy and safety attitudes. *New Media & Society* 19, 3 (2017), 347–365.

[13] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Joshua Gracie, Munmun De Choudhury, Pamela J Wisniewski, and Gianluca Stringhini. 2022. Understanding the Digital Lives of Youth: Analyzing Media Shared within Safe Versus Unsafe Private Conversations on Instagram. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) *(Chi '22, Article 148)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[14] Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. 2019. Americans and privacy: Concerned, confused and feeling lack of control over their personal information. (2019).

[15] Grant Blank, Gillian Bolsover, and Elizabeth Dubois. 2014. A new privacy paradox: Young people and privacy on social network sites. In *Prepared for the Annual Meeting of the American Sociological Association*, Vol. 17.

[16] Danah Boyd. 2008. Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. In *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*. unknown, 1–26.

[17] danah boyd. 2014. *It's Complicated: The Social Lives of Networked Teens* (1 ed.). Yale University Press, New York.

[18] Sydney Elaine Brammer, Narissra Maria Punyanunt-Carter, and Robin S Duffee. 2022. Oversharing on social networking sites: A contemporary communication phenomenon. *Computers in Human Behavior Reports* 8 (2022), 100236.

[19] Jelle Brands and Janne Van Doorn. 2022. The measurement, intensity and determinants of fear of cybercrime: A systematic review. *Computers in Human Behavior* 127 (2022), 107082.

[20] Petter Bae Brandtzaeg and Marika Lüders. 2018. Time collapse in social media: Extending the context collapse. *Social Media+ Society* 4, 1 (2018), 2056305118763349.

[21] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[22] Ann Cavoukian. 2009. Privacy by design. (2009).

[23] Hsuan-Ting Chen and Yonghwan Kim. 2013. Problematic use of social network sites: The interactive relationship between gratifications sought and privacy concerns. *Cyberpsychology, Behavior, and Social Networking* 16, 11 (2013), 806–812.

[24] Hsuen Chi Chiu and Chien Wen (tina) Yuan. 2021. To Last Long or to Fade Away: Investigating Users' Instagram Post and Story Practices. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. Acm, Virtual Event USA, 32–35.

[25] Hichang Cho, Jae-Shin Lee, and Siyoung Chung. 2010. Optimistic bias about online privacy risks: Testing the moderating effects of perceived controllability and prior experience. *Computers in Human Behavior* 26, 5 (2010), 987–995.

[26] Hui-Lien Chou and Chien Chou. 2023. How teens negotiate privacy on social media proactively and reactively. *New Media & Society* 25, 6 (2023), 1290–1312.

[27] Camille Cobb, Lucy Simko, Tadayoshi Kohno, and Alexis Hiniker. 2020. User Experiences with Online Status Indicators. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Acm, Honolulu HI USA, 1–12.

[28] Frances Corry. 2021. Screenshot, save, share, shame: Making sense of new media through screenshots and public shame. *First Monday* (2021).

[29] Katie Davis. 2012. Friendship 2.0: adolescents' experiences of belonging and self-disclosure online. *J. Adolesc.* 35, 6 (Dec. 2012), 1527–1536. https://doi.org/10.1016/j.adolescence.2012.02.013

[30] Ralf De Wolf. 2020. Contextualizing how teens manage personal and interpersonal privacy on social media. *New media & society* 22, 6 (2020), 1058–1075.

[31] Bernhard Debatin. 2011. Ethics, privacy, and self-restraint in social networking. In *Privacy online: Perspectives on privacy and self-disclosure in the social web*. Springer, 47–60.

[32] Vanessa P Dennen, Stacey A Rutledge, Lauren M Bagdy, Jerrica T Rowlett, Shannon Burnick, and Sarah Joyce. 2017. Context collapse and student social media networks: Where life and high school collide. In *Proceedings of the 8th international conference on social media & society*. 1–5.

[33] Michael A DeVito, Jeremy Birnholtz, and Jeffery T Hancock. 2017. Platforms, People, and Perception: Using Affordances to Understand Self-Presentation on Social Media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Acm, Portland Oregon USA, 740–754.

[34] Tobias Dienlin and Miriam J Metzger. 2016. An extended privacy calculus model for SNSs: Analyzing self-disclosure and self-withdrawal in a representative US sample. *Journal of Computer-Mediated Communication* 21, 5 (2016), 368–383.

[35] Duy-Tai Dinh, Tsutomu Fujinami, and Van-Nam Huynh. 2019. Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient. In *Knowledge and Systems Sciences: 20th International Symposium, KSS 2019, Da Nang, Vietnam, November 29–December 1, 2019, Proceedings 20*. Springer, 1–17.

[36] David Elkind and Robert Bowen. 1979. Imaginary audience behavior in children and adolescents. *Developmental psychology* 15, 1 (1979), 38.

[37] Nicole B Ellison, Cliff Lampe, and Charles Steinfield. 2010. With a little help from my friends: How social network sites affect social capital processes. *A networked self* (2010), 132–153.

[38] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of computer-mediated communication* 12, 4 (2007), 1143–1168.

[39] Nicole B Ellison, Jessica Vitak, Charles Steinfield, Rebecca Gray, and Cliff Lampe. 2011. Negotiating privacy concerns and social capital needs in a social media environment. *Privacy online: Perspectives on privacy and self-disclosure in the social web* (2011), 19–32.

[40] Stephanie W Greenberg. 1986. Fear and its relationship to crime, neighborhood deterioration, and informal social control. *The social ecology of crime* (1986), 47–62.

[41] Ralph Gross and Alessandro Acquisti. 2005. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. 71–80.

[42] Jessica L Hamilton, Jacqueline Nesi, and Sophia Choukas-Bradley. 2022. Reexamining Social Media and Socioemotional Well-Being Among Adolescents Through the Lens of the COVID-19 Pandemic: A Theoretical Review and Directions for Future Research. *Perspect. Psychol. Sci.* 17, 3 (May 2022), 662–679.

[43] Eszter Hargittai and Alice Marwick. 2016. "What can I really do?" Explaining the privacy paradox with online apathy. *International journal of communication* 10 (2016), 21.

[44] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)* 28, 1 (1979), 100–108.

[45] Jane Im, Ruiyi Wang, Weikun Lyu, Nick Cook, Hana Habib, Lorrie Faith Cranor, Nikola Banovic, and Florian Schaub. 2023. Less is Not More: Improving Findability and Actionability of Privacy Controls for Online Behavioral Advertising. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(Chi '23, Article 661)*. Association for Computing Machinery, New York, NY, USA, 1–33.

[46] Jonathan Jackson and Emily Gray. 2010. Functional fear and public insecurities about crime. *The British Journal of Criminology* 50, 1 (2010), 1–22.

[47] Haiyan Jia, Pamela J Wisniewski, Heng Xu, Mary Beth Rosson, and John M Carroll. 2015. Risk-taking as a Learning Process for Shaping Teen's Online Information Privacy Behaviors. 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, 583–599.

[48] Mohsen Jozani, Emmanuel Ayaburi, Myung Ko, and Kim-Kwang Raymond Choo. 2020. Privacy concerns and benefits of engagement with social media-enabled apps: A privacy calculus perspective. *Computers in Human Behavior* 107 (2020), 106260.

[49] Hyunjin Kang, Wonsun Shin, and Junru Huang. 2022. Teens' privacy management on video-sharing social media: the roles of perceived privacy risk and parental mediation. *Internet Research* 32, 1 (2022), 312–334.

[50] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. {"My$}$ Data Just Goes ${$Everywhere:"$}$ User Mental Models of the Internet and Implications for Privacy and Security. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*. 39–52.

[51] Hanna Krasnova, Oliver Günther, Sarah Spiekermann, and Ksenia Koroleva. 2009. Privacy concerns and identity in online social networks. *Identity in the Information Society* 2 (2009), 39–63.

[52] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in it together: interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3217–3226.

[53] Doohwang Lee, Robert Larose, and Nora Rifon. 2008. Keeping our network safe: a model of online protection behaviour. *Behaviour & Information Technology* 27, 5 (2008), 445–454.

[54] Seong-Sik Lee, Kyung-shick Choi, Sinyong Choi, and Elizabeth Englander. 2019. A test of structural model for fear of crime in social networking sites. *International Journal of Cybersecurity Intelligence & Cybercrime* 2, 2 (2019), 5–22.

[55] Theo Lorenc, Mark Petticrew, Margaret Whitehead, David Neary, Stephen Clayton, Kath Wright, Hilary Thomson, Steven Cummins, Amanda Sowden, and Adrian Renton. 2013. Fear of crime and the environment: systematic review of UK qualitative evidence. *BMC public health* 13 (2013), 1–8.

[56] Alice E Marwick and Danah Boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New media & society* 16, 7 (2014), 1051–1067.

[57] Hiroaki Masaki, Kengo Shibata, Shui Hoshino, Takahiro Ishihama, Nagayuki Saito, and Koji Yatani. 2020. Exploring Nudge Designs to Help Adolescent SNS Users Avoid Privacy and Safety Threats. 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, 1–11.

[58] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[59] Reham Mohamed, Paulina Chametka, and Sonia Chiasson. 2020. The Influence of Decaying the Representation of Older Social Media Content on Simulated Hiring Decisions. 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, 1–19. https://doi.org/10.1145/3313831.3376346

[60] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[61] Sina Ostendorf, Silke M Müller, and Matthias Brand. 2020. Neglecting long-term risks: Self-disclosure on social media and its relation to individual decision-making tendencies and problematic social-networks-use. *Frontiers in Psychology* 11 (2020), 2913.

[62] Jochen Peter and Patti M Valkenburg. 2011. Adolescents' online privacy: Toward a developmental perspective. *Privacy online: Perspectives on privacy and self-disclosure in the social web* (2011), 221–234.

[63] Sabid Bin Habib Pias, Imtiaz Ahmad, Taslima Akter, Apu Kapadia, and Adam J Lee. 2022. Decaying Photos for Enhanced Privacy: User Perceptions Towards Temporal Redactions and 'Trusted' Platforms. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 1–30. https://doi.org/10.1145/3555538

[64] Afsaneh Razi, Zainab Agha, Neeraj Chatlani, and Pamela Wisniewski. 2020. Privacy challenges for adolescents as a vulnerable population. In *Networked Privacy Workshop of the 2020 CHI Conference on Human Factors in Computing Systems*.

[65] Elissa M Redmiles, Jessica Bodford, and Lindsay Blackwell. 2019. "I just want to feel safe": A Diary Study of Safety Perceptions on Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 405–416.

[66] Alexis Shore and Kelsey Prena. 2023. Platform rules as privacy tools: The influence of screenshot accountability and trust on privacy management. *New Media & Society* (2023), 14614448231188929.

[67] Monika Taddicken and Cornelia Jers. 2011. The uses of privacy online: trading a loss of privacy for social web gratifications? In *Privacy online: Perspectives on privacy and self-disclosure in the social web*. Springer, 143–156.

[68] Lisa Collins Tidwell and Joseph B Walther. 2002. Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human communication research* 28, 3 (2002), 317–348.

[69] Sabine Trepte. 2020. The Social Media Privacy Model: Privacy and Communication in the Light of Social Media Affordances. *Commun. Theory* 31, 4 (May 2020), 549–570.

[70] Michel Walrave, Ini Vanwesenbeeck, and Wannes Heirman. 2012. Connecting and protecting? Comparing predictors of self-disclosure and privacy settings use between adolescents and adults. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 6, 1 (2012).

[71] Joseph B Walther. 2011. Introduction to privacy online. In *Privacy online: Perspectives on privacy and self-disclosure in the social web*. Springer, 3–8.

[72] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Franziska Roesner, and Kurt Thomas. 2023. "There's so much responsibility on users right now:" Expert Advice for Staying Safer From Hate and Harassment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(Chi '23, Article 190)*. Association for Computing Machinery, New York, NY, USA, 1–17.

[73] Emily Weinstein and Carrie James. 2022. *Behind Their Screens: What Teens Are Facing (and Adults Are Missing)*. MIT Press.

[74] Pamela Wisniewski. 2018. The Privacy Paradox of Adolescent Online Safety: A Matter of Risk Prevention or Risk Resilience? *IEEE Secur. Priv.* 16, 2 (2018), 86–90.

[75] Pamela J. Wisniewski, Jessica Vitak, and Heidi Hartikainen. 2022. Privacy in Adolescence. In *Modern Socio-Technical Perspectives on Privacy*. Springer International Publishing.

[76] Sijia Xiao, Danaë Metaxa, Joon Sung Park, Karrie Karahalios, and Niloufar Salehi. 2020. Random, Messy, Funny, Raw: Finstas as Intimate Reconfigurations of Social Media. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Acm, Honolulu HI USA, 1–13.

[77] Bin Xu, Pamara Chang, Christopher L. Welker, Natalya N. Bazarova, and Dan Cosley. 2016. Automatic Archiving versus Default Deletion: What Snapchat Tells Us About Ephemerality in Design. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. Acm, San Francisco California USA, 1662–1675. https://doi.org/10.1145/2818048.2819948

[78] Dorothy Zhao, Mikako Inaba, and Andrés Monroy-Hernández. 2022. Understanding Teenage Perceptions and Configurations of Privacy on Instagram. *Proceedings of the ACM on Human-Computer Interaction* 6, Cscw2 (2022), 1–28.

A DEMOGRAPHICS DATA

Table 3. Demographics of the Co-design Survey Participants (N=19)

Gender identity	Girls (63%), Boys (26%), Non-binary or third gender (5%), Boy and non-binary or third gender (5%)
Age	red13 (10.5%), 14 (10.5%), 15 (15.8%), 16 (21.1%), 17 (21.1%), 18 (21.1%)
Race	White (53%), Asian or Asian-American (32%), Black or African American (11%), White and Asian or Asian-American (5%)
Hispanic or Latin-American Origin	No (84%), Yes (16%)
Social Media Usage	Instagram (100%), BeReal (84%), Snapchat (74%), Twitter (68%), TikTok (63%), Discord (11%), Reddit (11%), Tumblr (5%)
Platforms Where Participants Reported Sharing Most Frequently	Instagram (37%), BeReal (32%), Snapchat (11%)), Twitter (5%), YouTube (5%), Discord (5%), Rinsta (5%)
Platforms Where Participants Reported Sharing Most Comfortably	Instagram (26%), BeReal (26%), Snapchat (16%), Discord (16%), Rinsta (5%), Pinterest (5%), Finsta (5%)

Table 4. Demographics of the Follow-up (Design Evaluation) Survey Participants (N=136)

Gender identity	Girls (59.6%), Boys (33.8%), Non-binary or third gender (2.9%), Girl and non-binary or third gender (1.5%), Boy, girl, and non-binary or third gender (1.5%), Prefer Not to Disclose (0.7%)
Age	13 (0%), 14 (5.1%), 15 (18.4%), 16 (18.4%), 17 (29.4%), 18 (28.7%)
Race	White (47.8%), Black or African American (25.7%), Asian or Asian-American (20.6%), White and Black or African American (1.5%), White and Asian or Asian-American (0.7%), White and American Indian or Alaska Native (1.5%), Black or African-American, American Indian or Alaska Native (1.5%), Asian or Asian-American, Black or African-American (0.7%), Other (0.7%)
Hispanic or Latin-American Origin	No (83.1%), Yes (16.9%)
Social Media Usage	Instagram (Real) (85.3%), Snapchat (55.9%), TikTok (65.4%), BeReal (38.2%), Instagram (Spam) (39.7%), Twitter (40.4%) Other (7.9%): Reddit (2.9%), Youtube (1.5%), Pinterest (0.7%), Locket (0.7%), Lemon8 (0.7%), Pintrest (0.7%), Whatsapp (0.7%), Zeeme (0.7%)

B IDENTIFYING DYSFUNCTIONAL WORRY

The following is the algorithm for identifying "Unworried" vs. "Functional Worry" vs. "Dysfunctional Worry" groups, adopted from Jackson and Gray (2010) [46]. We first ask the following five questions as part of the worry group survey:

I. Overall, how much do you worry about privacy-related issues on the [public/private] account on [platform name]? (1: Not at all worried; 2: Not very worried; 3: Fairly worried; 4: Very worried)
II. What steps do you take to protect yourself on your [public/private] account on [platform name]? (0: Never; 1: Occasionally; 2: Often; 3: Always, on five different ways of privacy protection measures)
III. To what extent do you feel these measures are effective in protecting your privacy on your [public/private] account on [platform name]? (0: Not at all; 1: A little; 2: Moderately; 3: Quite a bit; 4: Very much)
IV. How much do your concerns about privacy issues on your [public/private] account on [platform name] affect your overall quality of life? (0: Not at all; 1: A little; 2: Moderately; 3: Quite a bit; 4: Very much)
V. How much do the steps you take to protect your privacy on [public/private] account on [platform name] affect your quality of life? (0: Not at all; 1: A little; 2: Moderately; 3: Quite a bit; 4: Very much)

Table 5. Coding response to questions in the worry group survey.

Code Label	Code Description	Coded responses to survey questions
(1a)	Not worried about crime	"1: Not at all worried" or "2: Not very worried" to Question I.
(1b)	Worried about crime	"3: Fairly worried" or "Very worried" on Question I.
(2a)	Worry has no impact on quality of life	"0: Not at all" or "1: A little" on Question IV.
(2b)	Worry has impact on quality of life	"2: Moderately" or "3: Quite a bit" or "4: Very much" on Question IV.
(3a)	Don't take precautions	Average of five sub-scores to Question II. rounds to 0.
(3b)	Take precautions	Average of five sub-scores to Question II. rounds to 1, 2, or 3.
(4a)	Precautions have no impact on safety	"0: Not at all" or "1: A little" on Question III.
(4b)	Precautions have impact on safety	"2: Moderately" or "3: Quite a bit" or "4: Very much" on Question III.
(5a)	Precautions have no impact on quality of life	"0: Not at all" or "1: A little" on Question V.
(5b)	Precautions have impact on quality of life	"2: Moderately" or "3: Quite a bit" or "4: Very much" on Question V.

Based on the coded responses above, we categorize respondents into three groups—"Unworried," "Functional Worry," and "Dysfunctional Worry" based on the conditions below:

- Unworried Group: (1a)
- Functional Worry Group: {(1b) && (2b) && (3b) && (4b) && (5a)} || {(1b) && (2a)}
- Dysfunctional Worry Group: {(1b) && (2b) && (3a)} || {(1b) && (2b) && (3b) && (4a)} || {(1b) && (2b) && (3b) && (4b) && (5b)}

C UMAP [58] CLUSTERS OF OPEN-ENDED RESPONSES IN FOLLOW-UP SURVEY

Table 6. UMAP cluster labels and example quotes for each of the 10 prototypes used in the follow-up design evaluation survey. The numbers in parenthesis in the "Cluster Label" column indicate the number of responses that were classified into the cluster.

Feature	Cluster Label	Example Quotes
Account Protection (Likes)	0: User Restriction (33)	"*I like that you can restrict certain users that have a certain amount of red flags*" (R6); "*I like that this feature could prevent suspicious users or bots from interacting with your content.*" (R51)
	1: Extra Security (34)	"*It lets you have extra security measures*" (R104); "*i like the security idea. it gives me a sense of comfort as i believe the extra security layers is an advantage*" (R123)
	2: Customizability (66)	"*It gives the user more options on how to customize their accounts privacy*" (R108); "*I love the specificity and variety in the options for who can view your profile.*" (R119)
Account Protection (Dislikes)	1: Loopholes (32)	"*How do you filter for people who are very new and suspiciously active? What if someone creates a new account to follow you?*" (R32); "*I think it's good, but people can still create new accounts to stalk.*" (R58)
	2: Complexity (16)	"*It looks complicated*" (R87); "*Too many sub features*" (R90)
	3: Uncertain Criteria (28)	"*i dont know exactly how meaningful the red flags are- how are they given?*" (R79); "*Im just concerned about how red flags will be handed out and how accurate it will be. Also countermeasures and ways to remove red flags that were placed unjustifiably.*" (R116)
	0, 4, 5: None (45)	"*No complaints! :)*" (R21); "*I do not dislike anything.*" (R47)
Categorized Viewer List (Likes)	0, 2: Reviewing Followers (114)	"*The fact I can see who I've been talking to and who I haven't been talking to also it might help filter out fake accounts*" (R93); "*It sorts your friends based on how much you chat with them. So, you can keep your real buddies at the top and decide who stays on your follower list.*" (R107)
	1: Contentment (18)	"*I love this!!*" (R35); "*It's interesting*" (R102)
Categorized Viewer List (Dislikes)	2: Too much Transparency (54)	"*I think it sacrifices the privacy of the viewers a little bit: some people might not want others to know how often they view/interact with their content.*" (R1); "*I dont like the idea of having a list of people who view your stuff, it can become an obsession issue where people are constantly checking who viewed their posts and may use it to cause drama or rumors. It's unnecessary and causes drama and mental health issues.*" (R36)
	3: Excessive (15)	"*seems a little extra*" (R37); "*Too complex*" (R89)
	0, 1: None (51)	"*Great feature. I dislike nothing.*" (R27); "*Nothing really*" (R109)
Auto-delete with Save (Likes)	0, 5: Record Control (58)	"*i like that i can control how long my chats are present, minimizing my digital footprint*" (R19); "*I like that there isnt a record of everything you have said in the past.*" (R62)
	2: Mutual Transparency (35)	"*I like the second feature where it notifies both sides if a screenshot is taken.*" (R124); "*It tells me in advance that saving will notify the others. And can save as an image.*" (R23)

Continued on next page

Table 6 – continued from previous page

Feature	Cluster Label	Example Quotes
	1, 3, 4: None (37)	"*nothing*" (R30); "*I don't like this feature.*" (R111)
Auto-delete with Save (Dislikes)	1: Time Period (25)	"*3 days is long enough to show it to other people IRL and have you conversation spread without you knowing.*" (R77); "*Yes it seems like 3 days is too early for the message to be deleted*" (R78)
	2: Lost Record (65)	"*Being able to permanently erase chats can clear evidence of crimes committed on the app and so on.*" (R42); "*I forget conversations from a couple days ago easily, so this increases the likelyhood of me forgetting and never remembering.*" (R109)
	0: None (33)	"*I honestly have no complaints about this feature!*" (R20); "*I don't dislike anything about this feature.*" (R43)
On-demand Screenshot Blocking (Likes)	0, 1, 4: Control over Sharing (108)	"*Good way to prevent unwanted sharing of a post meant specifically for a certain audience. Nice!*" (R24); "*I really like this feature because it makes sure the content I share will be less likely to be shared with others; it makes it harder for the content to go places.*" (R56)
	2, 3: Contentment (26)	"*I really like this!*" (R35); "*Looks amazing*" (R99)
On-demand Screenshot Blocking (Dislikes)	0: No Notification (25)	"*I would make it to where (like snapchat) it lets the owner of the account know who, and when they screenshotted, i think that could help prevent a lot of mental health issues and bullying.*" (R33); "*Maybe a sort of notification can be sent to me after someone tries to take a screenshot, probably with the person's name asking me if I would like to let them take the shot*" (R87)
	3: Loopholes (25)	"*It prevents some unconsensual sharing, however it isn't stopping people from showing other in person or finding another method.*" (R17); "*Unfortunately, people will find their way around it, like using a different phone to take a picture of it.*" (R105)
	1, 2: None (69)	"*I love it as it is :)*" (R37); "*I do not dislike anything.*" (R113)
Reminder to Review (Likes)	0: Contentment (11)	"*This is amazing*" (R73); "*It is cool.*" (R109)
	1: Self-Reflection (115)	"*WOW! This is a phenomenal feature. I tend to go back to my social media and reconsider or delete posts that reflect "the old me" or are just too old for my liking.*" (R23); "*It reminds you to check your old posts, so you make sure they still fit who you are now. Keeps your online self in sync with the real you.*" (R107)
	2: None (6)	"*nothing really*" (R52); "*Nothing*" (R108)
Reminder to Review (Dislikes)	0: Unnecessary (6)	"*I think it's unnecessary*" (R38); "*I just don't think it's necessary for me.*" (R74)
	3: Annoying (28)	"*it seems annoying and kind of pointless*" (R3); "*I don't like the part where it gives a reminder, and I would hope that this feature doesn't show every single time you post because it would be annoying.*" (R18)

Continued on next page

Table 6 – continued from previous page

Feature	Cluster Label	Example Quotes
	5: Inflexibility (40)	"*I don't like that it only gives the option to delete after 24 hours and not a chosen time period, although that might be just for the example.*" (R17); "*I dislike how the opposing option to posting is to delete permanently. I feel a better option to deleting would be an option to private/archive the post so it's not available to the public but still to the user.*" (R25)
	1, 2, 4: None (44)	"*I dont dislike anything about it.*" (R5); "*none*" (R69)
Red Flag and Follow-up (Likes)	0, 2: Reassurance (37)	"*I like that you are able to receive updates about reports, along with the details of the punishment.*" (R102); "*I love how it allows you to know if your concern has been heard instead of wondering if it has or hasn't been taken action for.*" (R128)
	1: Contentment (20)	"*Everything about the features*" (R7); "*Love it*" (R36)
	3, 4: Safer Environment (77)	"*I like this because it allows users of the app to hold each other accountable and create a safer environment.*" (R57); "*I like that it is included because it should be standard with any sharing platform. It allows users to take care of themselves by reporting people/content that make them feel unsafe.*" (R115)
Red Flag and Follow-up (Dislikes)	0: Lack of Variation (26)	"*As it shows here it says banned for a month from the platform. I'd hope that there would be varying degrees of punishments.*" (R24); "*I think it seems pretty good. I would make sure that the ban period/consequence for the flag varies depending on how bad the flag was.*" (R68)
	2: Misuse (34)	"*This feature can be abused and people can start reporting things just because they don't like the person that posted it, etc. Also, a lot of times the app ignores reports that are serious.*" (R101); "*The only thing I have to say about this feature is that I would hope that others wouldn't misuse the report button just to get others banned unreasonably.*" (R114)
	1, 3: None (56)	"*I dont dislike it! Its quite nice*" (R4); "*Nothing I dislike about this feature*" (R71)
Pseudonymity Mode (Likes)	0: Audience Control (16)	"*I like how I can choose who can see my profile and who can't*" (R27); "*I like the control over who can interact/view my account*" (R50)
	1: Contentment (22)	"*It's definitely helpful*" (R9); "*It's good feature*" (R101)
	2, 5, 6: Advanced Anonymity (70)	"*I like how you can customize what you appear as to other people; it adds a layer of anonymity and acts as a preventative measure against potential harassers.*" (R59); "*How it truly makes it you private.*" (R63)
	3, 4: None (23)	"*I don't like it.*" (R55); "*I don't think I really like this*" (R116)
Pseudonymity Mode (Dislikes)	1: Misuse and Challenges in Connecting (77)	"*I don't like this feature because it could be used in the wrong way. For example, someone could use this feature to bully someone by hiding their profile from the victim.*" (R47); "*It can be hard to figure out who someone is. It also can keep people from finding potential friends online.*" (R20)
	0, 2: None (48)	"*I don't have any specific issues with it.*" (R27); "*There's nothing I dislike*" (R61)

Continued on next page

Table 6 – continued from previous page

Feature	Cluster Label	Example Quotes
Personal Information Alert (Likes)	0, 1, 2: Error Prevention and Awareness (134)	"*Ooh I like this, because sometimes people can miss things, so if the app's able to catch things like that before the post is even sent out, that'd be super helpful.*" (R130); "*I like that it can help people realize that their post may contain sensitive information that really shouldn't be available to see. Especially for people who post more impulsively.*" (R68)
Personal Information Alert (Dislikes)	0: Inaccuracy and Inflexibility (53)	"*I do wonder if the feature would always work properly every time(picking up the information or failing to pick up the information).*" (R57); "*Maybe they should have a studio feature where you can blur/edit things in the app so you don't have to redo the whole post, instead you can go into the studio for one specific photo in the post (if multiple)* (R31)
	1, 2: None (60)	"*Nothing to dislike*" (R66); "*I can't think of anything.*" (R99)
User Privacy Norms (Likes)	0, 4: Explicitness (51)	"*I like that it spells out the guidelines for you and makes it easy to understand how to be safe on the internet.*" (R19); "*I like that it has pretty concise bullet points that are easily understandable for everyone.*" (R67)
	1: Contentment (7)	"*The features is very useful*" (R77); "*It's a great feature*" (R97)
	2, 3: Establishing Norm (50)	"*I love that it makes being respectful normal instead of making out people who want privacy and respect online to be weird.*" (R45); "*I like that it's straight out there: being privacy-conscious isn't odd, but rather smart.*" (R129)
	5: Safety Reminders (25)	"*It is a reminder that the internet can be a dangerous place, and privacy should be prioritized*" (R10); "*I like that it gives people a reminder of how to keep themselves safe online.*" (R64)
User Privacy Norms (Dislikes)	2: Vagueness and Potential Ignorance (23)	"*I think some of the guidelines are conceptual/not really concrete so some users may not really read thoroughly through them.*" (R50); "*Most people do not really take the time to read such guidelines or terms when they use an app, so this might go ignored by many people.*" (R108)
	5: Annoying (17)	"*seems a bit annoying, i wouldnt even read all those words to be honest id just click skip*" (R30); "*It depends on how often the reminder pops up but if it's often it might become annoying.*" (R11)
	6: Lengthy (23)	"*It's very wordy and I have a feeling lots of teenagers would just skip right past it and not read it*" (R32); "*some might find it too long to read, maybe summarizing a bit would help*" (R91)
	0, 1, 3, 4: None (55)	"*I have no issues with the feature*" (R33); "*Nothing at all*" (R102)
"View As" (Likes)	0, 3: Reassurance (88)	"*This would be extremely useful to make sure one does not have anything exposed that they would not want to be.*" (R20); "*You can check how your profile looks to others, making sure you share only what you want. It's like a double-check for your online image.*" (R109)
	1: Contentment (17)	"*I like it for the most part.*" (R66); "*It's interesting*" (R75)
	2: Different POV (29)	"*It allows me to see how friends, peers, and maybe strangers would see my account from their point of view/perspective.*" (R44); "*It gives you better insight on how others see your account, without you having to assume or ask another person to check for you.*" (R126)

Continued on next page

Table 6 – continued from previous page

Feature	Cluster Label	Example Quotes
"View As" (Dislikes)	0, 3: Excessive and Unnecessary (32)	"*I think this feature can easily become quite complicated, and especially if multiple users in the following list are being shown different parts.*" (R25); "*It's kind of unneeded because what others see on your account is pretty much what you can see on your end*" (R51)
	1, 2, 4, 5: None (79)	"*don't dislike anything*" (R5); "*I have nothing negative to say.*" (R13)