

# More Tools, More Control

## Lessons from Young Users on Handling Unwanted Messages Online

Michal Luria, CDT  
Carol F. Scott, University of Michigan

November 2023



The **Center for Democracy & Technology (CDT)** is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.



# More Tools, More Control

## Lessons from Young Users on Handling Unwanted Messages Online

### Authors

**Michal Luria, Center for Democracy & Technology**  
**Carol F. Scott, University of Michigan**

### CONTENT ADVISORY

**This report depicts young people's experiences with unwanted and harmful messages online and contains some descriptions of harassment and sexual abuse online. Please read with care.**

### WITH CONTRIBUTIONS BY

Aliya Bhatia, Ari Goldberg, Tim Hoagland, Samir Jain, Mallory Knodel, Emma Llansó, Iverna McGowan, Gabriel Nicholas, and Dhanaraj Thakur. Illustrations by Alanah Sarginson.

### ACKNOWLEDGEMENTS

We would like to thank Jessica Pater, Sarita Schoenebeck, and Jessica Vitak for their feedback on an earlier draft of this report, and Cambridge Focus for assisting in participant recruitment. We would also like to thank participants in the study for sharing their experiences and ideas with us.

This work was made possible through a grant from the John S. and James L. Knight Foundation.

### SUGGESTED CITATION

Luria, M. and Scott, C. F. (2023). More Tools, More Control: Lessons from Young Users on Handling Unwanted Messages Online. Center for Democracy & Technology. <https://cdt.org/insights/more-tools-more-control-lessons-from-young-users-on-handling-unwanted-messages-online/>.

**References in this report include original links as well as links archived and shortened by the Perma.cc service.** The Perma.cc links also contain information on the date of retrieval and archive.



# Contents

<b>Executive Summary</b>	<b>6</b>
<b>Background</b>	<b>9</b>
<b>Method</b>	<b>11</b>
<b>Participants</b>	<b>14</b>
<b>Findings</b>	<b>15</b>
<b>Finding #1. Participants made a conscious decision to use messaging platforms, mostly to communicate with close circles, although they were aware of possible risks.</b>	<b>15</b>
<b>Finding #2. Encryption was not a consideration for platform choice, perhaps due to a lack of awareness of how it can be used to protect individuals in messaging communication.</b>	<b>17</b>
<b>Finding #3. The distribution of unwanted content was not even: some participants received high volumes of unwanted content, while others received none. Participants who received more unwanted messages tended to be African American men.</b>	<b>19</b>
<b>Finding #4. “Unwanted” messages were perceived as ones that are unsolicited, come from strangers, and frequently include sexual content.</b>	<b>19</b>
<b>Finding #5. Most unwanted messages were received on direct messaging on social media, as opposed to messaging platforms, to some extent due to their design for discoverability.</b>	<b>20</b>
<b>Finding #6. The most common “unwanted” messages included sexual language or material, followed by hateful language and unsolicited promotional links.</b>	<b>22</b>

# Contents

<b>Finding #7. Participants attempted to set up an environment that included people they wanted to interact with as best they could, through profile settings and stringent approval processes.</b>	<b>24</b>
<b>Finding #8. When receiving an unwanted message, young people make judgments and escalate as needed: Ignoring, blocking, reporting, or as a last resort, asking for help.</b>	<b>26</b>
<b>Recommendations</b>	<b>29</b>
<b>Conclusion</b>	<b>35</b>
<b>References</b>	<b>36</b>

# Executive Summary

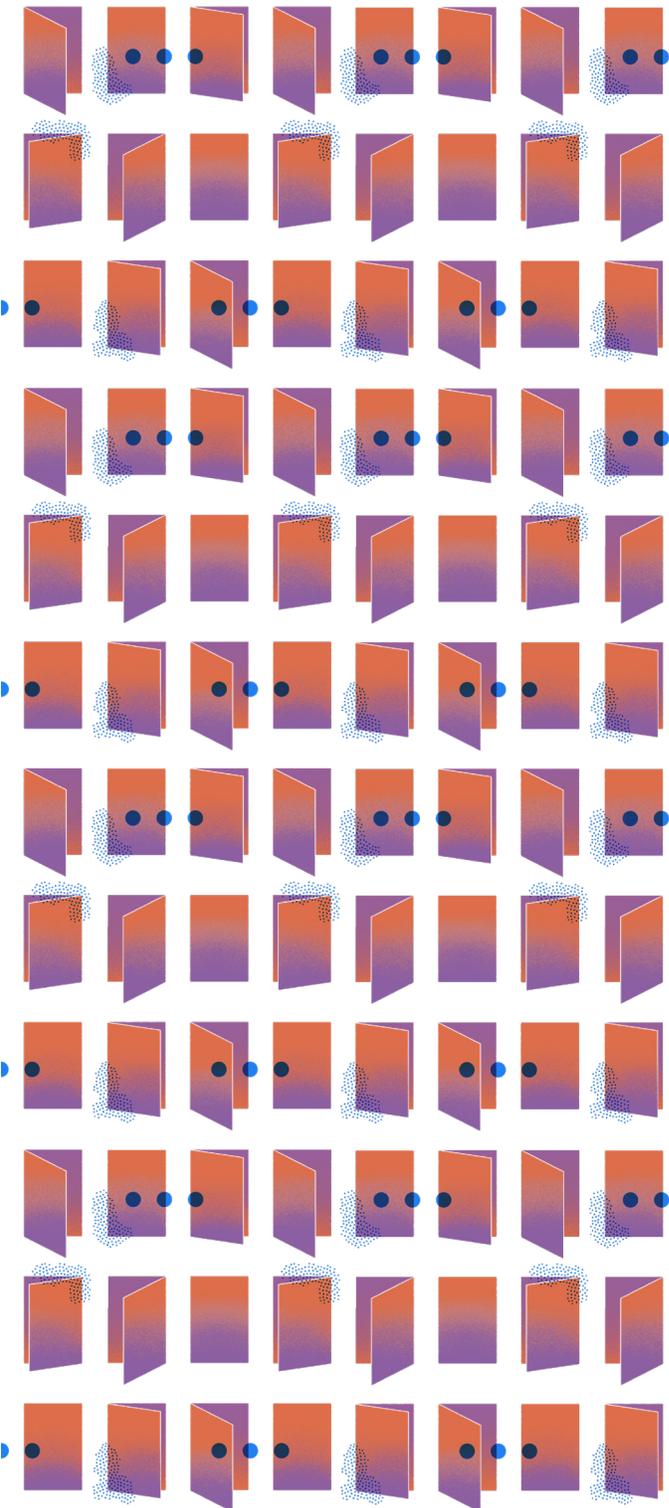
According to reports, about 95% of U.S. teens have access to smartphones and use the internet daily, which for most includes social media and messaging platforms ([Vogels et al., 2022](#); [Nesi et al., 2023](#)). This ubiquity has raised much concern about its contribution to a teen mental health crisis among parents, educators, and legislators ([Lima, 2023b](#)). Research paints a more complex picture of the impacts of social media and messaging, finding both benefits and risks to young people ([Weinstein, 2018](#); [Valkenburg, 2022](#)). As a result, it is difficult to determine the most effective protective measures for online services and legislators to take, even as the pressure to act grows.

Recent work highlighted the importance of considering both platform and user activity on platforms to better understand potential impacts, as well as to create safer online environments for young people ([Griffioen et al., 2020](#); [Valkenburg, 2022](#)). In this study, we aim to examine and understand young people’s experiences with direct messaging content across messaging apps (e.g., WhatsApp, iMessage) and private messaging on social media apps (e.g., Instagram DMs, Snapchat DMs). Messaging is one of the most popular activities for young people online, but these can also be a channel for concerning interactions, such as harassment ([Lwin et al., 2012](#); [Copp et al., 2021](#)) and sexual solicitation ([Jones et al., 2012](#); [Wolak et al., 2018](#)).

Working directly with 32 U.S. teenagers (ages 14-17, n=18) and young adults (ages 18-21, n=14) who use direct messaging regularly, we conducted a month-long, mixed-methods study. The study included an online survey, a 3-week diary study, and a semi-structured interview. We aimed to understand young people’s experiences when interacting through digital messages and to identify key opportunities to increase their sense of safety, agency, and control.

## Summary of Findings

1. Some American teenagers and young adults, especially young African American men or boys, regularly receive unwanted content through direct messaging.
2. Participants are aware of bad actors online and proactively try to assess risks and strategize how to handle privately received unwanted content.
3. Participants’ definition of “unwanted, unpleasant, or concerning” messages described mostly unsolicited messages that came from strangers and frequently included sexual



content. Such messages were not equally distributed; some participants in the study received as many as 7 unwanted interactions over the course of 3 weeks, while 6 participants received one message, and 7 received none at all.

4. The unequal distribution of unwanted messages might be partially explained by some participants' strategies to minimize their exposure, including setting their accounts to private, interacting primarily with close circles, and keeping their participation online to a minimum. Unfortunately, the latter may also contribute to a "chilling effect" among young people and may lead them to self-censor their activity and expression online.
5. Participants had strategies for dealing with unwanted messages after receiving them, including assessing and escalating them as needed. When deciding to escalate messages (rather than blocking or ignoring them), they primarily reported them and occasionally chose to share with friends and family.
6. Participants identified several tools they would like platforms to offer to better assess and address unwanted interactions; this surfaces opportunities for platforms to act to support users' needs.

The findings lead to several trauma-informed baseline guidelines for service providers, alongside areas for additional research that would improve available features for user safety on messaging platforms.

## Baseline Guidelines for Service Providers

- Platforms should provide users with basic response tools to unwanted messages, such as the ability to delete, block, and report.
- Platform accounts should be defaulted to private settings with limited discoverability, with the ability to select a public profile if desired.
- Platforms should introduce more friction or "speed bumps" in interactions with unknown profiles or potential strangers.
- Platforms should include more "just-in-time" notices that inform and educate users about potential risks on messaging platforms and tools for mitigating those risks.
- Platforms should be more transparent about reported cases by allowing users to track the outcome of a message or a person they had previously reported.

## Areas for Additional Consideration

- Platforms could introduce more user-side filtering by allowing users to define their own filtering and blocking criteria for private messages.
- Platforms might support interoperable blocklists that can be applied across different platforms, or allow third-party applications to apply a single blocklist.

- More efforts should be invested in educating young users and adults (e.g., parents and educators) about the platforms they select to communicate with others, and the possible benefits of selecting safer messaging spaces (e.g., the choice of phone number-based applications or end-to-end encrypted (E2EE) messaging platforms).

In line with the larger body of literature investigating social media use (e.g., [Whiting & Williams, 2013](#); [Scott et al., 2017](#); [Anderson & Jiang, 2018](#); [Valkenburg et al., 2022](#); [Vogels, 2022](#); [Scott et al., 2023](#)), our recommendations focus on equipping young people with tools and knowledge that could assist them in dealing with a range of risks of messaging, while still allowing them access to the many positive aspects and interactions that direct messaging platforms offer. Instead of attempting to fully control and shape users' exposure to messages, which is both impossible and not necessarily desired, we encourage platforms to hand over some of the control to users. In line with trauma-informed approaches, providing all users, not just young people, with tools and knowledge preserves their sense of control and agency, while helping them help themselves.



# Background

The United States is currently experiencing a youth mental health crisis, with more than half of American parents expressing concerns about their child's mental health (CDC, 2022; CDC, 2023). Many are concerned that social media and messaging platforms are significantly contributing to this crisis (Doucleff, 2023), and debates about how to best address this concern are widespread and ongoing.

Yet addressing concerns surrounding young people's use of social media and messaging platforms is a complex endeavor. On the one hand, the potential disadvantages of their usage are many: They have been linked to poor mental health (Nesi et al., 2023; Seabrook et al., 2016), lack of sleep (Alonzo et al., 2019), and low self-esteem and distorted body image (Holland & Tiggemann, 2016), especially among young girls (Nesi et al., 2023). They can expose teenagers to sexual content (Brown & L'Engle, 2009) and content with excessive violence (Wisniewski et al., 2016), as well as harmful interactions including cyberbullying (Hamm et al., 2015; Vogels, 2022), harassment (Lwin et al., 2012; Copp et al., 2021), or sexual solicitation (Jones et al., 2012; Wolak et al., 2018). A recent global survey released by Microsoft suggested that 69% of teens were exposed to harmful content online in the past year, predominantly misinformation, violence, hate speech, and harassment (Microsoft, 2023).

On the other hand, social media and messaging platforms have been shown to be vital for many young people, allowing them to maintain social relationships (Seabrook et al., 2016), to express themselves (Uhls et al., 2017; Luria, 2022), and to explore and browse for entertainment and inspiration (Weinstein, 2018). Research shows that social media and messaging platforms can also increase well-being by providing youth with supportive communities that share their identities and interests (Anderson et al., 2022), particularly marginalized youth, such as adolescent girls of color (Nesi et al., 2023), teens who experience mental health challenges (Ridout & Campbell, 2018), and LGBTQ+ youth (Lucero, 2017; Berger et al., 2022; Miller, 2023).

Recent years have seen a significant push for U.S. legislation – at both federal and state levels – that attempts to better protect young people from possible negative impacts of social media and messaging. Proposals

have introduced many approaches, including limiting the time that young people can spend on a platform ([Moshiri, 2023](#)), mandating parental consent and control ([Singer, 2023](#)), creating age-appropriate design codes that limit exposure to some types of content ([Huddleston, 2023](#)), and even outright banning teenagers' access to platforms ([Lima, 2023a](#)). While these conversations are dominated by adults, youth advocacy groups are calling for young people's voices to be heard, arguing that they should be able to impact policy that will directly affect them ([Lima, 2023c](#)).

The focus of proposals and discussions has usually been broadly on social media platforms; some have addressed the internet as a whole. Yet the experiences of young people can differ across channels and digital platforms and require more nuanced consideration of different types and contexts of communication online ([Pater et al., 2015](#); [Griffioen et al., 2020](#)). For instance, 50% of adolescent girls reported that messaging platforms (e.g., WhatsApp, iMessage) positively affect them, and 10% said they have a "mostly negative effect." In contrast, 32% of adolescent girls said that Snapchat has a "mostly positive effect", and 26% thought it was mostly negative ([Nesi et al., 2023](#)). This variance highlights the need for research to understand how these differences play out.

Direct messaging (DM or DMing), which is the focus of this work, has received far less attention than public posts on social media ([Ali et al., 2022](#)) but is just as important to research. That is because many of the interactions that parents and policymakers are particularly worried about, like cyberbullying ([Hamm et al., 2015](#); [Vogels, 2022](#)), harassment ([Lwin et al., 2012](#); [Copp et al., 2021](#)), and sexual solicitation ([Jones et al., 2012](#); [Wolak et al., 2018](#)), occur on direct messaging platforms and channels that are hidden from public view.

We set out to help address this gap by identifying young people's challenges and concerns when using messaging services offered by both messaging apps (e.g., WhatsApp, iMessage) and social media apps (e.g., Facebook Messenger, Instagram DMs, Snapchat DMs). We also continue the effort of including young people in the U.S. by directly collaborating and engaging with youth in this study. Our work thus spotlights harmful online interactions experienced by young people on messaging platforms and offers youth-centered recommendations for tools and interventions that can increase their safety, but also give them choice, voice, and autonomy – ones that should be considered by both policymakers and messaging service providers.



# Method

This study took a human-centered, qualitative approach. We collaborated with young American people to explore their lived experiences of what we named “*unpleasant, concerning, or otherwise unwanted*” messages on direct messaging apps and features. The definition of “*unpleasant, concerning, or otherwise unwanted*” was left as open-ended as possible, as we assumed that words like “harm” or “risk” would trigger discussions about specific types of interactions that are often referred to in the public discourse, but do not necessarily get at the breadth and depth of young people’s experiences. Using an open-ended definition, therefore, allowed participating youth to shape what it did or did not include.

Our research questions, listed below, centered on direct messaging apps and features. Much less research has focused solely on private interpersonal interactions online, despite the influential role they play and the possible risks they impose on young people in the U.S. ([Razi et al., 2023](#)):

- (1) Which platforms are young people (ages 14-21) using for messaging, and how (if at all) do their experiences differ from one platform to another?
- (2) What messages do young people perceive as “unpleasant, concerning, or otherwise unwanted” (hereinafter referred to as unwanted messages)?
- (3) How are unwanted messages perceived and experienced by young people, and how do they impact them?
  - (3b) On which platforms are unwanted messages received?
  - (3c) Who do unwanted messages come from? Family, friends, peers, strangers, etc?
- (4) How do young people deal with or protect themselves from unwanted messages?
- (5) What tools would young people like to have to be able to better handle unwanted messages on messaging platforms?

We implemented a U.S.-based, 3-part online study to address these research questions from a youth-centric perspective. The study included: (1) an online survey about participants’ use and perceptions of messaging platforms; (2) a 3-week diary study in which participants shared their experiences on messaging platforms ‘in the moment’; and (3) a semi-structured interview about participants’ experiences on messaging platforms during the study and beyond. This combination

of methods allowed us to begin to synthesize an understanding of how U.S. youth experience conversations on messaging platforms and how they perceive and deal with unwanted content in these private channels.

In the online survey, participants answered questions about their use patterns of digital messaging platforms (e.g., who they interact with, what platforms they use, what kinds of data they share, etc.), as well as demographic questions intended to identify potential differences in experiences between groups, as well as to ensure diversity in our sample (age, gender, race, sexual orientation, location in the U.S., and education accommodations).

The diary study was the main part of the research. Diary studies are a human-centered qualitative research method that puts participants' reported experiences at the forefront and attempts to learn about them in real time ([Redmiles et al., 2019](#)). In diary studies, participants are asked to log their experiences over a particular time span, allowing researchers to collect longitudinal data that capture participant attitudes and behaviors. The data are also captured in context and shortly after they occur, instead of attempting to recall past events retrospectively ([Zimmerman & Wieder, 1977](#)). This in-the-moment recording provides more accurate and detailed responses in regard to people's experiences, as the memory of previous events and past experiences may shift over time ([Ohly et al., 2010](#)).

In our 3-week diary study, conducted on the research platform MetricWire, we asked participants to record a brief video-based "entry" in which they describe their experience every time they encountered an unwanted message on a messaging platform, whether in one-on-one interactions or in a group. If they hadn't experienced any unwanted messages, they were asked to submit an entry about a "week free of unwanted messages" at the end of each week. This approach provided participants with a self-driven, asynchronous way to share their everyday experiences. Further, the use of video was part of our youth-centric approach, given the rise in preference for video communication among this age group ([Pater et al., 2015](#)).

Lastly, we conducted 30-minute semi-structured interviews with participants via Zoom. In the interviews, we asked about their perceptions of interactions on messaging platforms and experiences of unwanted content more broadly, their strategies and tools for coping with unwanted messages, and the tools they wish they had. With permission, interviews were recorded (video and audio) for transcription.

Our final dataset included 126 diary entries created by 32 participants (mean duration of an entry = 1:38 minutes), 32 survey responses, and 22 qualitative 30-minute semi-

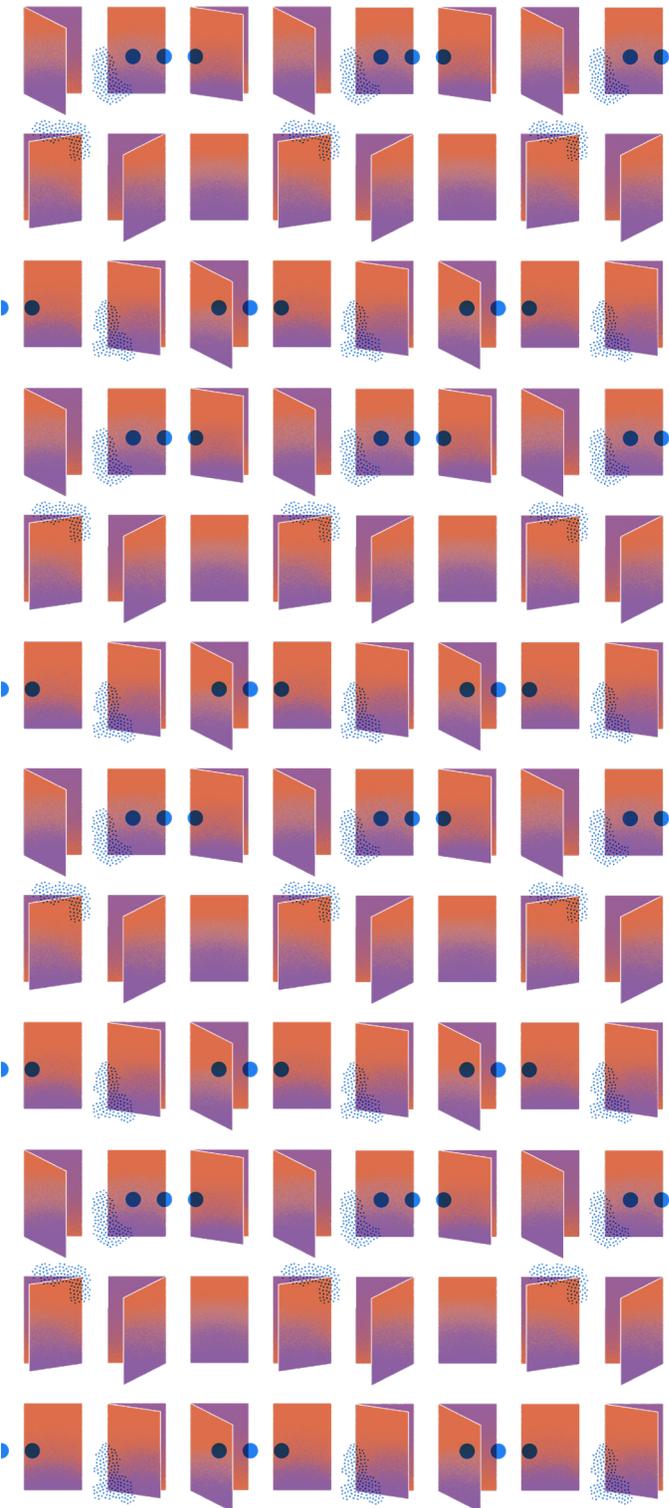
structured interviews. Recordings of diary entries and interviews were transcribed and analyzed to find common themes and patterns ([Braun & Clarke, 2006](#)). These themes are presented in our findings using direct participant quotes alongside descriptive statistics analyses from the survey. All study procedures were approved by the University of Michigan's Institutional Review Board (IRB).



# Participants

To be included, participants needed to be based in the U.S., be proficient in English and use messaging platforms at least once a week. A total of 18 teenagers (ages 14-17) and 14 young adults (ages 18-21) (n=32; mean age = 17.4 years) participated in the study. We considered participation for any individual who completed the onboarding survey and at least one diary entry. Of these, 22 participants completed all three parts of the study.

Of the 32 participants, 16 were men/boys, 10 women/girls, and 6 non-binary. Two were transgender. Ten (10) participants were African-American, 9 were white, 8 were multi-racial, 4 were Asian, and one participant was Latina. A total of 22 participants were heterosexual and 10 were LGBTQ+. Seven (7) participants receive different-ability learning accommodations at school (such as an individualized education plan or a 504 plan). Participants were compensated for both partial and full participation.



# Findings

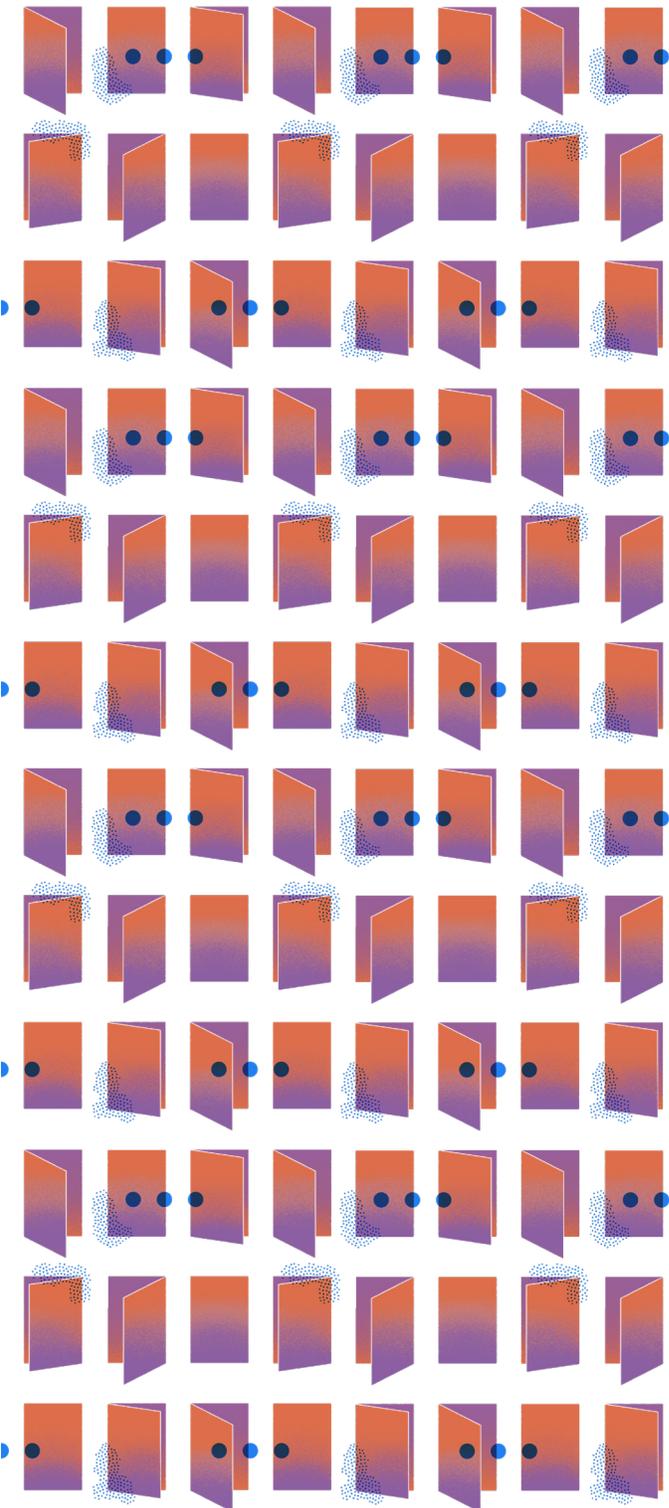
## A. Choice to Use Messaging Platforms

**Finding #1. Participants made a conscious decision to use messaging platforms, mostly to communicate with close circles, although they were aware of possible risks.**

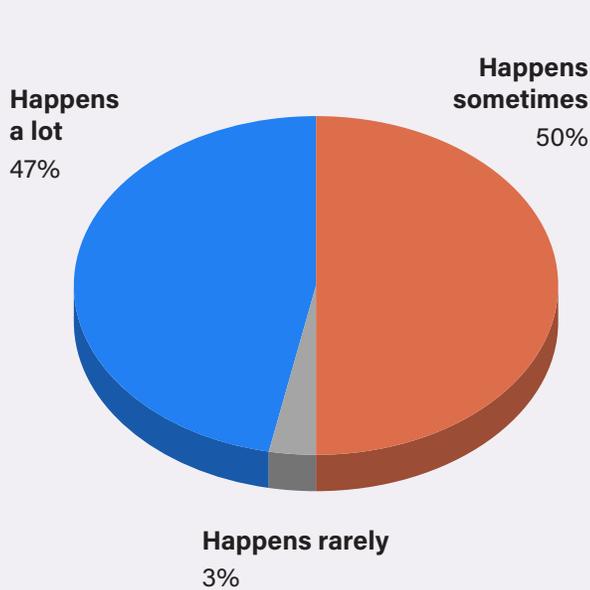
Participants are proactively choosing to use messaging platforms as a medium for their interpersonal communication. They are not passive consumers of content on messaging. Rather, messaging platforms provide a way for them to stay connected and in touch with friends and family:

*I don't really get to see my friends that much during the summer, so [messaging] gives me an opportunity to really talk to them, and just hang out again, like we used to, during school time. (Boy, 16)*

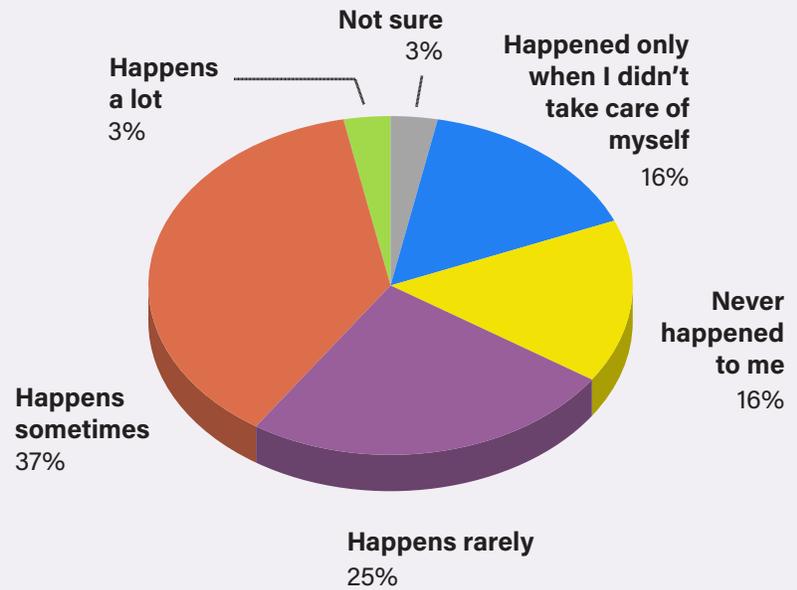
They were aware of the risks such platforms may pose; the vast majority of study participants (97%) said that people their age experience unwanted interactions on messaging at least sometimes (see Figure 1), and most participants (81%) have experienced unwanted messages themselves (see Figure 2).



How common do you think it is for people your age to experience concerning or uncomfortable situations on DM platforms?



How common has it been for you to experience concerning or uncomfortable situations on direct messaging platforms?



▲ **Figure 1. Participants' perception of how common unwanted messages on messaging platforms are among their age group.** None of the participants thought that "it happens only to those that do not take care of themselves," and none said "I haven't heard of any instances of concerning situations."

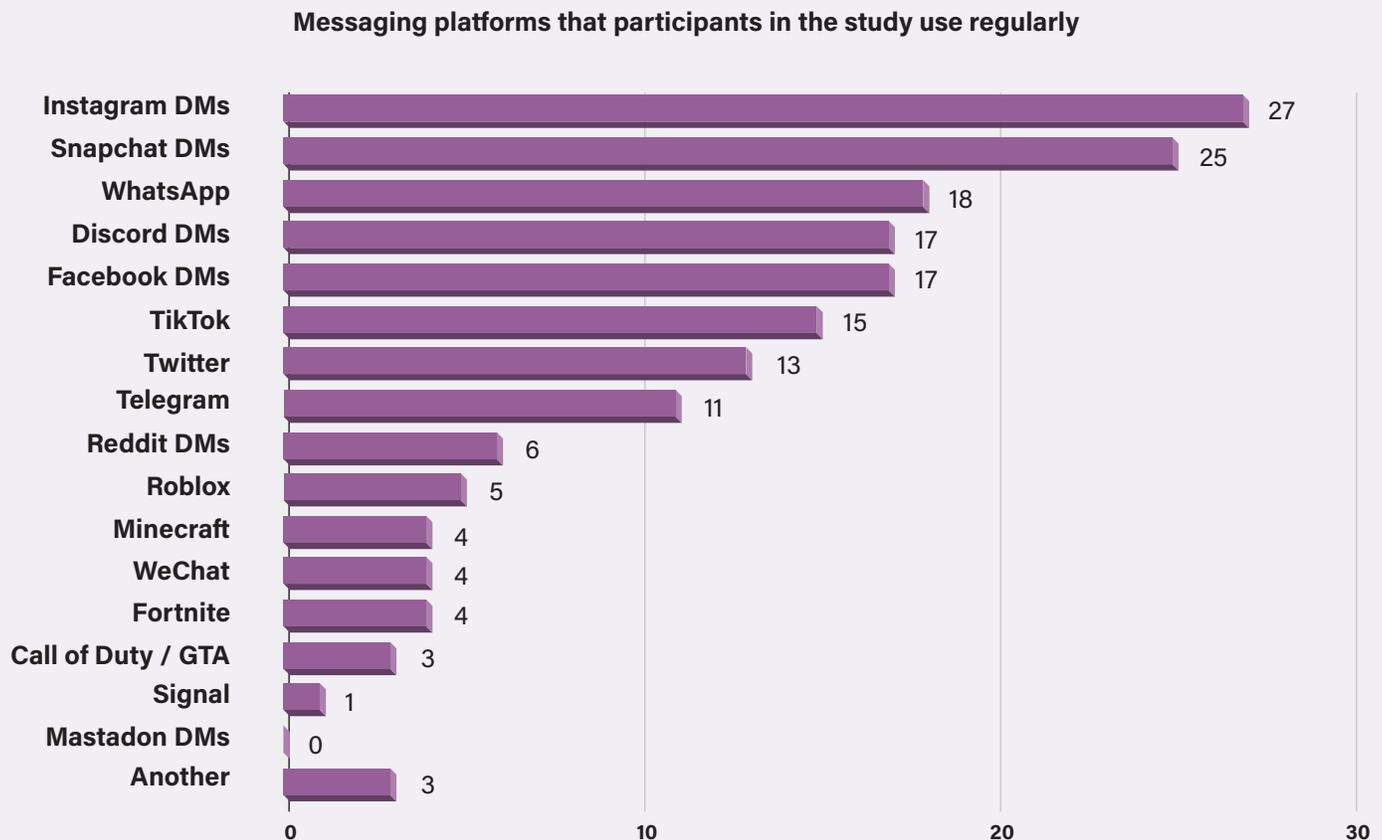
**Figure 2. Participants' perception of how common it has been for them to experience unwanted messages on messaging platforms.**

Still, several noted that they believed these risks are overstated:

*I feel like people vastly overestimate how many unwanted messages we get on platforms [...] The risks and dangers of social media are so emphatically spread by teachers and parents, etc. It's an active risk that I take, and I feel like I'm just comfortable with that. (Non-binary, 18)*

*It's not that big of a deal. If it's something really serious that could cause harm to me or other people, I will obviously tell an adult. Most of the time, I've never got anything that extreme. (Girl, 14)*

Participating young people used an average of 5 different messaging platforms on a regular basis (see Figure 3), most commonly Instagram (n=27) and Snapchat (n=25), followed by WhatsApp (n=18), Facebook Messenger (n=17), and Discord (n=17). The primary reason for their choice of platform was to be where others are. Direct messaging apps (such as WhatsApp) were more commonly used with family and friends, while DM features on social media (like Facebook Messenger) were used for group communication and what they perceived as weaker social ties.



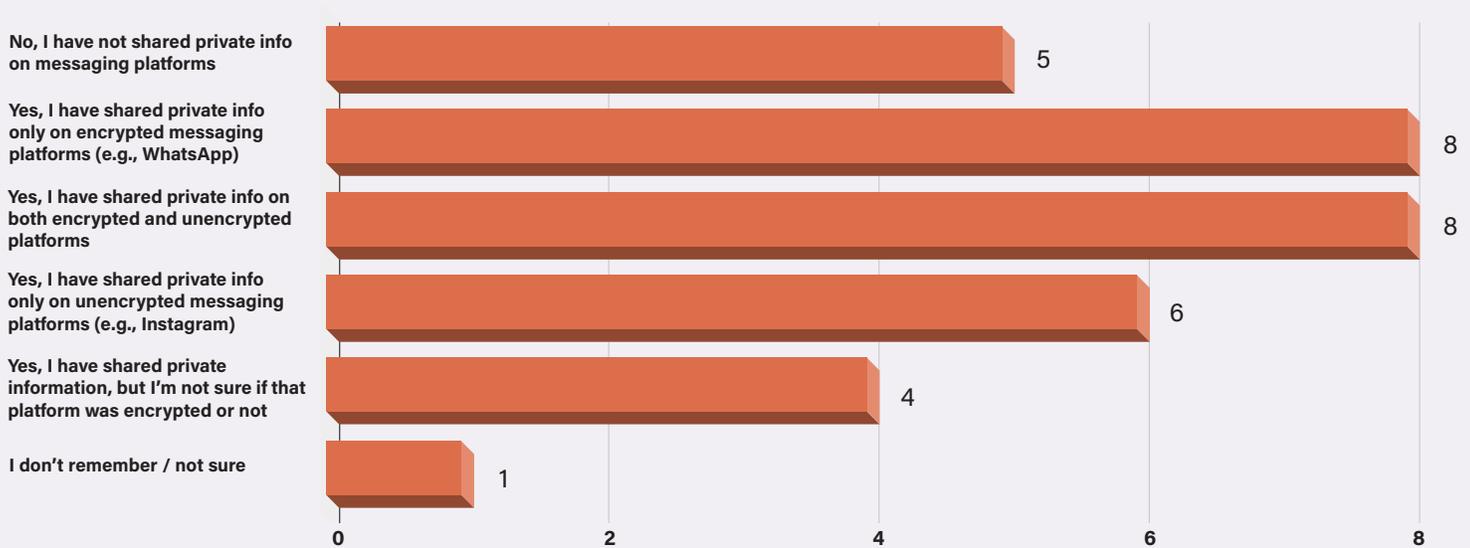
▲ **Figure 3. Messaging platforms that participants in the study use regularly.** Instagram DMs were the most common platform used among the study participants (27 participants used it on a regular basis), followed by SnapChat DMs (25), WhatsApp (18), Discord DMs (17), Facebook DMs (17), TikTok (15), Twitter, now named X (13), and Telegram (11).

In some instances, platforms were selected by educators, such as coaches or teachers, for group communication (e.g., Snapchat, Facebook Messenger). This highlights that youth can be in regular communication on messaging platforms with trusted adults who are not their parents. It also places some responsibility on these adults to select a platform with proper practices and safety tools before asking young people to join (more in the recommendations section).

**Finding #2. Encryption was not a consideration for platform choice, perhaps due to a lack of awareness of how it can be used to protect individuals in messaging communication.**

One of our survey aims was to understand how young people use end-to-end encrypted (E2EE) messaging platforms like WhatsApp in contrast to ones that are not end-to-end encrypted. E2EE is a key consideration for some groups like journalists and activists when choosing which platform to communicate through ([Belair-Gagnon et al., 2018](#)). However, this is not the case for the general public. According to Bai and colleagues, users frequently have poor or incorrect mental models about E2EE, including its benefits for protecting oneself online ([Bai et al., 2020](#)).

### Have you shared private personal information on direct messaging platforms in the last 6 months?



▲ Figure 4. Participants' responses to whether they shared private personal information (e.g., name, photos of you, contact information, address, etc) on messaging platforms in the last 6 months.

When we asked participating youth about how they decide which messaging platforms to regularly use, none of them mentioned encryption as a point of consideration. When asked in the survey about information sharing practices on encrypted vs. unencrypted platforms in the last 6 months, more than half of participants reported sharing personal information (defined as “things like your name, information about you, photos of you, your contact information, your address, etc”) on unencrypted platforms or on a platform they did not know whether it was encrypted or not. Some participants answered that they have only shared private information on encrypted platforms, and a handful reported never sharing private information on messaging platforms at all (see Figure 4).

Most of the unwanted messages that were reported in the diary study happened on platforms that were not E2EE (83%). That's not to say that these are more common on unencrypted platforms—participants tended to use more unencrypted platforms than encrypted ones. Rather, it shows that unwanted interactions can happen in both encrypted and unencrypted channels.

## Findings

### B. Unwanted Messages on Messaging Platforms

**Finding #3. The distribution of unwanted content was not even: some participants received high volumes of unwanted content, while others received none. Participants who received more unwanted messages tended to be African American men.**

Over the course of three weeks, 32 participants entered a total of 126 diary entries (mean per participant = 3.9 entries). Of those, 72 entries were of “unwanted” messages participants received, and 54 were entries of “a week free of unwanted messages.” The average quantity of unwanted messages received per participant was 2.38 messages in three weeks.

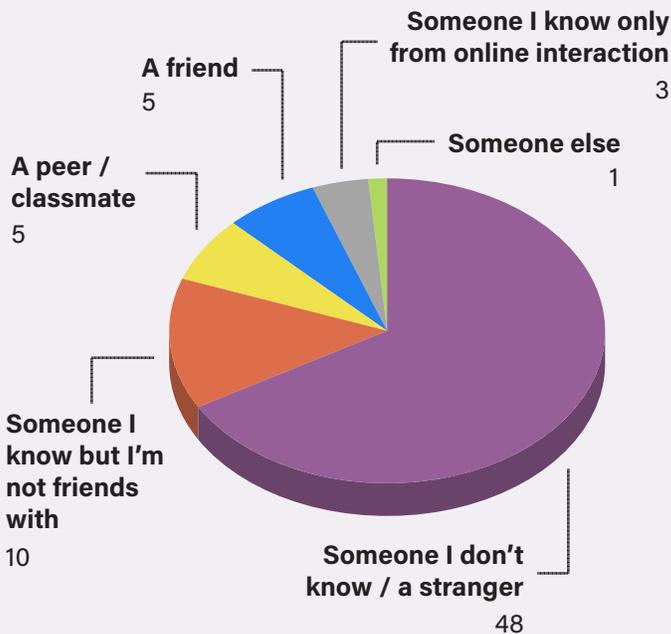
We found that the distribution of unwanted content was not even across participants. Several participants were more exposed than their peers, with up to 7 unwanted interactions over the course of 3 weeks. In contrast, 7 participants received none, and 6 participants received only one unwanted message during their three-week participation. The individuals receiving the most unwanted content tended to be African American men or boys, in line with prior work that shows that men and African American people can be subjected to some forms of online abuse more than other groups ([Thakur et al., 2022](#); [Vogels, 2022](#); [Moody, 2023](#)). This pattern extended to sexual content too, which was the most common form of unwanted messages in the study. Future work requires attending to these individuals to identify some of the circumstances that are contributing to this imbalance and to collaborate with them to design the tools they need to stay safe online.

**Finding #4. “Unwanted” messages were perceived as ones that are unsolicited, come from strangers, and frequently include sexual content.**

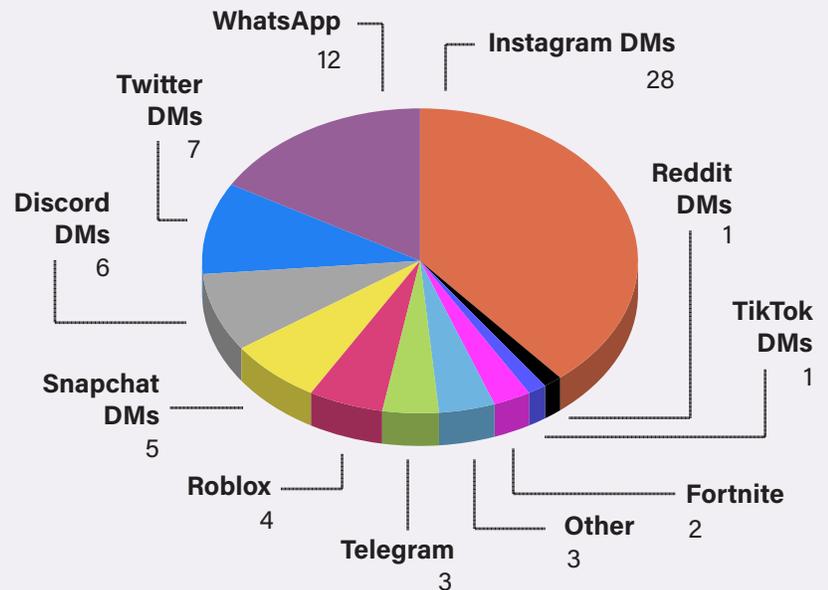
We intentionally left the definition of “unwanted” messages open-ended, so that participants in the study could define their own inclusion boundaries. When asked in the interview, most participants described “unwanted” messages as primarily those that are sent from people they do not know in person. Many participants also thought that “unwanted” is usually unsolicited, and frequently includes sexual content.

Their perceptions were confirmed in the logged diary entries (see Figure 5)—most reported unwanted content came from strangers (67%), or people they have interacted with before, but only online (4%).

**Who was the unwanted message from, or the interaction with?**



**Which messaging platform was the unwanted message or interaction on?**



▲ Figure 5. Participants reported who unwanted messages on messaging platforms came from.

Figure 6. Participants reported on which platform an unwanted message or interaction was received.

**Finding #5. Most unwanted messages were received on direct messaging on social media, as opposed to messaging platforms, to some extent due to their design for discoverability.**

In our study, 42 out of the total 72 “unwanted” messages (58%) were on social media platforms (see Figure 6). Instagram was the most common, with close to 40% of all unwanted messages coming through Instagram DMs. A total of 21 were on direct messaging platforms, such as WhatsApp (29%), and 6 were on gaming platforms like Roblox (8%).

This finding was supported in the interview portion of the study—participants flagged Instagram as the platform on which they thought they receive the most unwanted content. While these findings are likely impacted by the popularity of Instagram (84% of participants used Instagram on a regular basis), some participants pointed out that the problem might be in some parts of the design, specifically that anyone is able to follow or interact with them on Instagram:

*I think most platforms, other than Instagram, are safe because you have to add your friends before you can talk to them [...] I'm pretty sure Instagram just lets you talk*

*to anyone you want. I mean, I guess some influencers think it's better [...] but for normal people like me, it just holds a lot of unwanted interaction. (Girl, 14)*

*[Unwanted messages have] been really common on Instagram because I feel the culture of followers on Instagram is [that] you don't necessarily need to know [people you chat with]. (Boy, 16)*

Several participants flagged other platforms on which they commonly receive unwanted messages, including Twitter (now called X)<sup>1</sup>, TikTok, and Discord, due to their platform policy and design decisions:

*I think that [Twitter] don't flag things fast enough or at all. There's a lot of nudity, a lot of violence [...] I think Twitter is the most lenient when it comes to anyone bullying anyone, anyone wanting to post anything they're not supposed to, or anything in general. (Woman, 21)*

*The [direct messages] I remember were really violent ones on TikTok [...] A lot of rape threats, which was awful. I think that TikTok was different because the proximity to strangers was a lot closer. Anyone was just looking at my content [...] I would just post happy little videos of myself, and every single day I get a bunch of DMs from guys with the most horrific things. There was a countdown to my 18th birthday. (Non-binary, 18)*

*The thing with Discord is when you're on the same server with somebody on a particular platform, they are able to send you a message [...] Most of the time, those messages are unwanted because they're disturbing and they keep coming. I learned to switch off direct message[s] from people who I'm on the same server with. (Man, 18)*

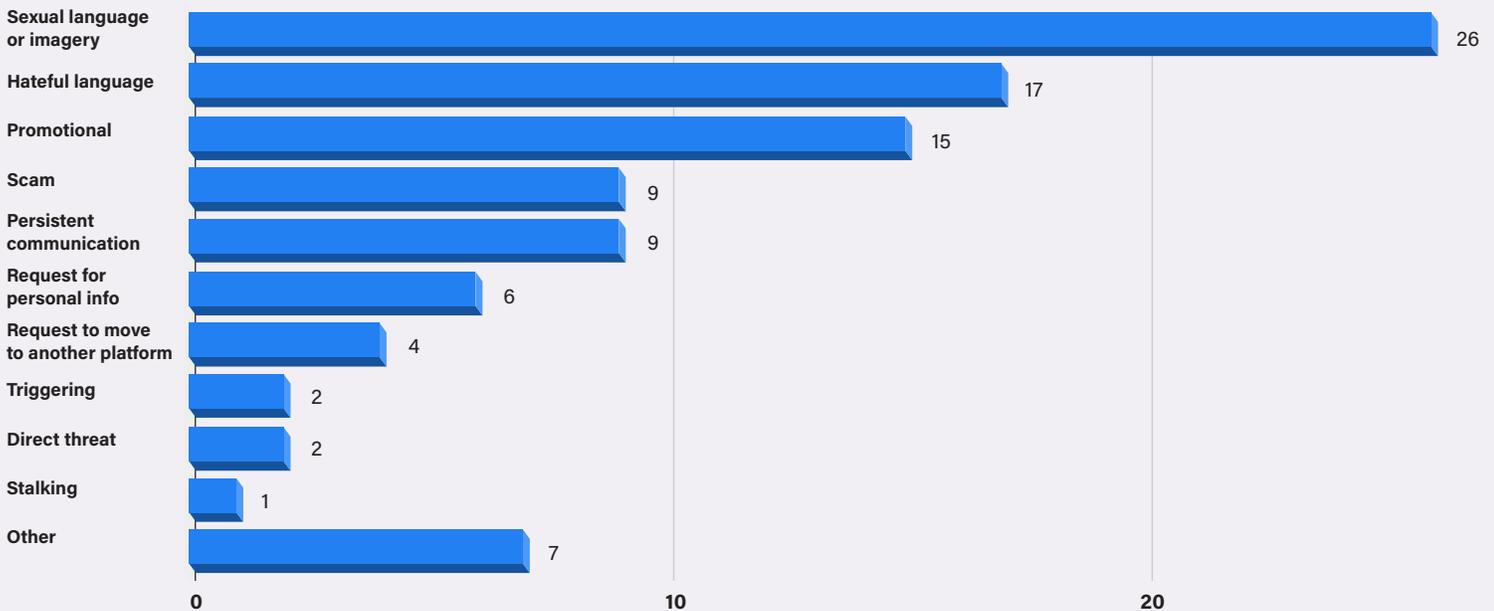
Gaming platforms that allow messaging between players were also flagged as ones that offer users few protections against unwanted messages:

*Then he started saying some very, very offensive stuff, which I won't go into, because it was really bad. It was really offensive. I don't know how he bypassed the Roblox moderation chat filter system [...] I do feel like I didn't have as much tools to really deal with the situation, due to the fact that Roblox Chat is just really underdeveloped. It's like as soon as you friend somebody, you just got to chat with them. (Boy, 16)*

---

1 Note that we use "Twitter" in this report because at the time of the research with participants that was the name of the company and the name participants used.

### Categorization of the types of interaction unwanted messages exhibited



▲ Figure 7. Types of interactions that unwanted messages in the study exhibited.

*Most of it [unwanted messages] actually happens on Roblox [...] I think that people tend to be more unattended on there. Like, for some of my peers, a lot of my peers, their parents check their phones and stuff like that [but parents don't check Roblox]. (Non-binary, 14)*

The sender differed between platforms too; most unwanted messages on direct messaging platforms were from people that participants knew, such as their peers or friends, as opposed to direct messages on social media, which were primarily from strangers, or people participants only knew online.

**Finding #6. The most common “unwanted” messages included sexual language or material, followed by hateful language and unsolicited promotional links (see Figure 7).**

From a total of 72 entries of “unwanted” messages, 26 included sexual content of some kind (e.g., sexual language or explicit imagery). The majority of sexual content was sent to participants through social media DMs:

## Findings

*I did see something that was uncomfortable and unwanted. I opened one request and was exposed to a male sending a picture of their genitals. Unfortunate[ly], it happens often, and it's sick that people do things like that. (Woman, 21)*

*She started sending me unsolicited nude photos, just started sending images without me asking for them. (Boy, 16)*

A significant portion of reported sexual messages also included sextortion or sexual solicitation, an attempt to coerce participants into participating in sexual behavior:

*I got [a message] from a guy who was persistently asking for my nudes. He kept asking about my sex life and said he wanted us to be partners, like, gay partners [...] I'm way [too] young to be interacting with that person in such a way. (Man, 15)*

*Started out with some generalized chatting, where are you from? What do you do? Stuff like that. Then, it transform[ed] into going to another direct media app, Snapchat, and they asked to play an inappropriate game in exchange for pictures and coerce you. (Man, 21)*

Hateful language, such as foul language, racism, and sexism was also common:

*He responded to my message and was like, that I should get my Black ass out of the server, out of the group. (Man, 16)*

*Someone just came to me in my DMs and had to body-shame me because I have actually gained weight. (Woman, 16)*

Other noted unwanted communications were requests to click a promotional link (“I got multiple marketing messages to subscribe to their channels.” [Man, 20]), unpleasantly persistent interactions (“She was spamming me nonstop, and just like would not stop even when I told her to stop.” [Non-binary, 14]), and scam attempts (“a stranger with a fake account was trying to get me to buy something, and click on a suspicious link” [Man, 19]).

## C. Current Protection Measures

**Finding #7. Participants attempted to set up an environment that included people they wanted to interact with as best they could, through profile settings and stringent approval processes.**

Findings indicate that young people are not only making a choice to use social media and messaging platforms, but that they are active participants, not passive consumers, when it comes to messaging. Below, we list some of the specific strategies the participants shared when operating online, with the goal of minimizing their own and others' exposure to unwanted and harmful messages.

**Setting their account to private.** Many participants purposely chose to set their account to private, noting it as key to keeping them safe. They also only allowed friends or people they knew offline to follow them:

*I have a private account, so I don't normally get DMs that I don't want. (Woman, 14)*

*On Instagram, it's not like I'm posting things for everybody to just see, it's a private account. I'm only interacting with my close friends. I think that's the main thing. I don't have a public account. If I had a public account it would be very depressing [...] One of my friends has a public account and she gets all random DMs and I'm like, "How is this happening?" (Woman, 16)*

**Being selective about chat partners.** When having a private account, participants were also able to be more intentional about who they follow, befriend, and talk to, generally avoiding strangers:

*I've just been talking to, majority are just school friends. I feel like social media is as dangerous as who you know. (Boy, 15)*

*I think it's a lot to do with how I operate on social media. I try and keep a really low profile overall and I just stick to friends and family [...] having a smaller profile makes less people who may be bad actors want to come after me. (Man, 19)*

*I make sure if someone requests to follow me, I don't do it right away. Normally, I'll see the person [offline], so I'll be like, "Hey, did you request to follow me on Instagram?" Then they'll be like, "Yes." If they say no, then it's not their account, and I won't accept it. (Girl, 14)*

**Lowering profile for less risk.** For our participating youth, protecting themselves from unwanted messages was not only about limiting who they talk to, but also about

## Findings

their activity and choice of exposure. Several participants mentioned that when they keep a low profile, unwanted messages are rare:

*Usually depending on how high the likes I get is how many more uncomfortable things I get [...] Whenever I post, it's going to happen once or twice. (Woman, 21)*

*Every time you post, I guess more people see it. They can talk to you, but if you don't post... like that week I didn't post, so I didn't get any unwanted interactions. (Girl, 14)*

This finding follows prior literature that states that high-profile accounts have faced a lot of hate and abuse online, implying that higher profile accounts have higher stakes ([Dahlen, 2021](#)). This suggests that unwanted messages can contribute to a chilling effect, as prior work has pointed out with other population groups ([Thakur et al., 2022](#)):

*[The direct message] was so offensive and I felt so bad. I had to take [my post] down. (Girl, 16)*

**Being selective about platforms.** Several participants noted that different platforms offered different levels of safety. Generally, participants felt safer on texting apps (like WhatsApp), as these are phone number-based and were therefore perceived as easier to control access to—participants use them to interact with close circles who have their number, making it difficult for bad actors to find and approach them:

*I feel safe on WhatsApp. It's just one-on-one conversation or a group conversation that you just talk within yourself. (Boy, 17)*

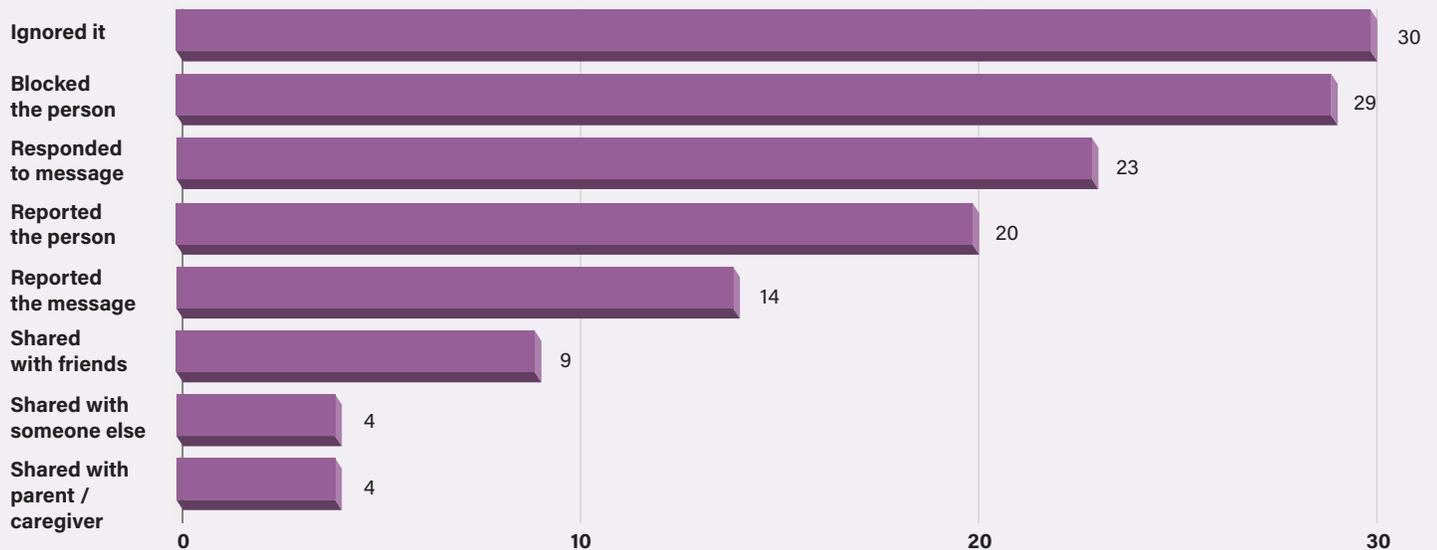
*I think with WhatsApp, it's rare to get DMs like that, unless, of course, the person has your number. I feel like you control who has access to your number. (Man, 21)*

While most participants kept using social media DMs and turned to protection measures as described above, in a couple of cases participants decided to remove platforms they thought involved too much risk:

*Instagram, there are lots of strangers there, so chatting on Instagram is actually a no for me [...] It's a decision. There are lots of unknown people on Instagram, and I'm not wanting unknown messages from strangers. (Boy, 17)*

*I used to have TikTok. TikTok was the worst [...] Anyone was just looking at my content in a way that doesn't happen on other platforms and so I deleted it because it just got way too much. (Non-binary, 18)*

### How participants reported responding to unwanted messages received during the study



▲ Figure 8. Participants' reported responses to unwanted messages received during the study.

## D. Current Response Measures

**Finding #8. When receiving an unwanted message, young people make judgments and escalate as needed: Ignoring, blocking, reporting, or as a last resort, asking for help.**

For the 72 unwanted messages that were reported in the diary study, the most common responses were ignoring the unwanted message (30 instances) and blocking the sender (29 instances), followed by responding to (23 instances) and reporting the sender (20 instances). Sharing with someone was last on the list (17 instances), with only 4 instances of sharing with a caregiver or parent (see Figure 8).

**Ignoring unwanted messages.** Aligned with how people reported responding to specific messages, in the interviews, most participants said they usually felt unaffected by unwanted messages, frequently choosing to ignore them. In part, their indifference may have been due to the volume of unwanted messages they receive, suggesting more needs to be done to address it:

*I just opened it up, saw what it was, and just moved on. (Man, 21)*

## Findings

*I just kind of ignore it, and just don't look at the chat really. Of course you can report and all, but at this rate it becomes so common, there are so many messages like that that you don't pay attention. (Non-binary, 21)*

*I usually [report], but if it's too many, I get tired, exhausted. It's time-consuming. (Woman, 21)*

**Blocking and reporting.** While ignoring was the automatic response, some messages triggered an assessment of whether to do more, like block or report. Participating youth decided whether to do so based on their perception of the severity of the interaction. For instance, if it seemed threatening to them, or if the sender may pose a risk of harming others:

*It's not that serious to me [...] But when it gets too threatening or any of those things, I report it as well. (Woman, 21)*

*I ignore, and then I'll choose to either block the number that it came from or report it if it is something suspicious. (Man, 19)*

Overall, participants perceived ignoring and blocking as a measure to protect themselves, while reporting was a contribution to the community, a way to help others receive less unwanted content:

*I love to report their account as well, just to make the whole thing, the whole platform even safer for the people, so I just report them, then I block them. (Boy, 16)*

*[I report] so that the app can figure out and do something about it before it happens to other people. (Girl, 14)*

But similar to prior work ([Thorn, 2020](#); [Vilk & Lo, 2023](#)), we found that reporting is perceived as time-consuming, especially when platforms provide very little feedback or evidence that users' efforts are effective. Instead, participants turn to blocking:

*Even if I report it, there is no tracking mechanism or complaint mechanism or redressal mechanism to have a redressal from the company point of view that they would block a user or filter such language or filter such messages. I'm left with no option but to ignore unless it's a very serious offense. (Man, 20)*

*[It] doesn't seem like they're even looking or even trying to investigate [...] What's the point in reporting it if it's not even working. I just choose to block most of the time because it seems reporting is not really doing anything [...] They'll tell me, "Thank you for reporting, we'll get back to you." They don't ever. (Woman, 21)*

**Sharing with friends, caregivers, or others.** When sharing, participants primarily shared with friends. Sharing with caregivers was the lowest on the action list—only 4 instances of 72 unwanted messages in the diary study were reported to parents or caregivers (6%). In the interviews, participants explained that they would turn to parents or another trusted adult only in a more severe case, that they feel would require adult intervention or support. Otherwise, they believed they could handle unwanted content themselves, using tools such as blocking and reporting:

*If it's just something random and stupid, I can handle that myself. It's not that big of a deal. If it's something really serious that could cause harm to me or other people, I will obviously tell an adult. Most of the time, I never got anything that extreme. (Girl, 14)*

*[The messages] never get to the point where I personally think that I need to get my parents involved because I think that I know how to deal with them myself, like by blocking them. (Girl, 14)*

Some of the reasons for avoiding sharing with parents included concerns that parents might overreach, their inability to do anything about it and the potential hassle of constantly sharing such interactions given their regular frequency:

*I told my mom one time and she's like, "This is the danger of social media. You should delete all your accounts and stuff." I think that adults tend to take it a lot more seriously than kids do, and so I feel a little bit alienated telling them. (Non-binary, 18)*

*Telling my parents, I don't think they will do so much, or maybe telling a friend wouldn't do so much. They'll just see the message, they just see the tweets. They'll just acknowledge that people are just wayward out here. (Girl, 16)*

*Some kids will eventually think that telling their parents everything they see as unwanted is going to become a task. (Boy, 16)*

Overall, participants were aware of available tools and their different strengths and weaknesses. Both protection and response measures that participants reported taking suggest that American youth are not passive users on messaging platforms; they are active participants in both creating the interactions they desire and avoiding the ones they don't. In the study, the vast majority of participants reported taking intentional and calculated steps in an attempt to prevent unwanted messages from coming into their accounts, and in their response to ones that do.



# Recommendations

We present five recommendations that messaging service providers should adopt to set a baseline for how to design messaging services with youth safety in mind. While some platforms have already implemented several baseline guidelines, smaller and newer platforms may not have these safeguards in place for users to prevent and handle unwanted communication.

Beyond these baseline recommendations, we present additional features and designs that providers should consider adopting. Although these may require further exploration and research, based on our findings, they have strong potential to increase young people's safety online and decrease the large amount of unwanted content that some youth have to deal with on a regular basis.

Collectively, these recommendations lean into and expand the strategies young people already reported using to keep themselves safe.

## Guideline #1: Block, report, delete

*[On Roblox] you can't really do anything [about unwanted messages], because even if you unfriend them, you still see the chat. I still saw the messages. I couldn't delete them. I couldn't get rid of them. (Boy, 16)*

While most platforms allow users to block and report others, or delete messages they've received, participants in the study noted that that is not the case for all platforms. Users should have the ability to delete an unwanted direct message they've received, or to permanently stop communication with a particular user so that they don't have to deal with this unwanted content any longer. This should include all types of communication: text messages, visual media (photos and videos), and audio streams (as common on gaming platforms).

## Guideline #2: Private account by default

*I don't really have to deal with anyone saying anything rude to me because I have a private account and I only use it to connect with my friends. (Woman, 19)*

Setting their account to private is a key strategy that youth use to reduce their exposure to strangers and to unwanted messages. We recommend that all platforms have a "private" account setting, which is enabled by default for all users when they first sign up to use a service.

Our findings show that participants commonly begin with a (default) public account, experience a negative interaction, and then decide to change their account to private. This is not always easy to do (e.g., not easy to find the setting or involves a lot of steps), and risks exposing young people to unnecessary risk until the change is made. Instead, defaulting to private accounts would allow youth (and adult users) to learn the platform with minimal exposure to a broader user base, and then decide for themselves whether to change their account to “public” or not. Given that in some cases, the reason for joining a social media platform is to meet new people, both youth and adult users should have the ability to switch to “public” as soon as they want to.

### **Guideline #3: Friction in interaction with strangers**

*Usually on Twitter, for example, it says to accept or decline the message, but I think it should do that before it shows any pictures [...] like “someone wants to message you” before they can even message you. (Woman, 21)*

Most participants consider “unwanted messages” as those that come unsolicited from strangers. To reduce this specific type of interaction and reduce bad actors’ ability to harm youth, we recommend that platforms design several friction points, or “speed bumps,” for interacting with unknown accounts for all users. These friction points should be trauma-informed ([Scott et al., 2023](#)), and designed to be strategic in reducing harm ([Single, 1996](#); [Marlatt & Witkiewitz, 2002](#); [Leslie et al., 2008](#)). Some examples include:

1. *Message requests*: A user receives a request from someone who would like to get in contact with them. Only after the user accepts this request can the person send a message.
2. *Hide visual media from strangers*: Initially hide all images and video that come from accounts the user has not yet connected with, asking them if they would like to view the content. This can include a warning about why it is not recommended to open media from unknown sources, as is common in email.
3. *‘No mutual friends’ warning*: When users are at the decision point for adding a friend or following someone new, ask the user if they are sure that they would like to add this contact, as they have no mutual friends.
4. *Suspicious profile warning*: In the same scenario, inform users if the requesting user is likely a fake account, has no followers, was created in the last few weeks, has not been verified, or any other potential yellow or red flag.

## Recommendations

### Guideline #4: Just-in-time notices about best private messaging practices

*I personally don't think that there are a lot of messaging apps that do a good job of teaching people how to conduct themselves online. All the ads I see for Instagram, Facebook, Snapchat say: 'have as many friends as you want.'* (Man, 19)

Platforms should provide users with “just-in-time” notices about how to safely use their messaging features in the form of short immediate prompts as part of basic interactions (as opposed to focusing primarily on long-form guides and safety information that rarely gets read). Just as platforms have increased efforts to inform users about what they see on social media feeds in context (e.g., content warnings on specific posts) and introduced well-being interventions (e.g., a self-help message pops up when someone types “depression” in search), just-in-time safety measures for messaging can help mitigate unwanted messages.

Platforms can draw from unique data they have about the kinds of messages that users report to make the most accurate and useful recommendations that are also trauma-informed, such as explaining tools they provide and how to use them, or what should be considered a red flag (e.g., links from unknown sources).

### Guideline #5: Feedback on reporting

*I think it would give me more peace of mind personally to know [platforms] were able to block that account or ban it or whatever else.* (Man, 19)

Participants perceived reporting as high-effort but not very effective or rewarding. One of the explanations participants gave for this impression was the lack of information about the reporting process and outcome, which aligns with findings from prior research about why users are unmotivated to report content (Vilk & Lo, 2023). One participant offered YouTube’s Reporting History dashboard (YouTube, 2018) as a successful model for informing users about content moderation decisions made about content they have reported.

Having simple, research-based reporting processes with few steps and usable designs (as introduced by Twitter [Twitter, 2021]) is not enough—we recommend that platforms allow users to track previously reported content, and view the platform’s handling and response to it. This feedback would be essential in encouraging users to continue reporting bad actors and reduce their presence on platforms.

**In addition to these baseline recommendations that we recommend all platforms implement, our findings point to several additional features that can be impactful in increasing youth safety on messaging platforms.**

## Recommendations

### User-controlled filtering

*“I think [I received few unwanted messages this week] because I just have a new filter on Discord, where you can rule out that type of stuff, which I think is pretty interesting. It really saves me the trouble.” (Boy, 16)*

Currently, some platforms automatically filter out messages that are suspected to be unwanted by users. However, several participants mentioned going into their filtered messages to check if anything had been unnecessarily blocked, ultimately causing them to view messages they sometimes would have preferred not to see.

Instead of automatic filtering that users may not always find reliable, we suggest that platforms allow users to create their own filters, as has been previously suggested for social media more broadly ([Scott et al., 2023](#)). One approach could be for platforms to allow users to select specific keywords to block (as Twitch allows its streamers to do). The downside is that this approach does not consider context, and therefore will only be effective to some extent. Another approach would be to allow users to turn a set of service-provided classifiers on and off, and dynamically evaluate what content is being filtered out as a result. Limitations to this approach may be that the tool’s efficacy and risk of overblocking would depend on how well trained the classifier is. A third approach would be to open services up to third-party developed tools that allow filtering and blocking, and would give users more choice and control over what gets blocked (similar to previously developed tools, such as Block Party for Twitter ([Grauer, 2021](#))). Although more research is needed to understand the exact tradeoffs between approaches to user-side filtering, our findings suggest that putting the power to filter back into users’ hands could minimize unwanted content while easing the concern of overfiltering.

### Bulk blocking

*I blocked one of their accounts. They ended up sending me another message, from another account, and the thing is that they didn’t change their name, or nothing. (Woman, 21)*

Other than simply ignoring, blocking was the most common strategy used by participating youth to address unwanted content. Based on the usefulness of blocking, we recommend that platforms consider extending blocking capabilities. One way of doing so would be to give users an option to block multiple accounts from the same origin. For instance, when blocking a user on Instagram, the interface asks whether you would like to block the user, or to “block [user] and other accounts they may have or create.”

*I wish there were more cross-platform communication. I get that they're different companies and it's not really interoperable but just if I had a way to block his number, block his IP address, block his email across platforms. (Non-binary, 18)*

A couple of participants shared experiences of having the same individual contact them on other messaging platforms after being blocked on one, which made them feel intimidated and unsafe. These experiences raise an opportunity for cross-platform blocking tools. For example, with interconnected platforms (e.g., what you post on Facebook can instantly be posted to Instagram, too), control features could also be interconnected, if requested by the user.

Other areas of future exploration could be allowing users to bring in their own blocklists from different platforms, or alternatively, allowing third-party developers to create tools that support a single blocklist across multiple platforms.

## Protection features for specific groups

A finding in this study that requires further examination and research is the potentially disproportionate experience of unwanted messages on messaging platforms among African American men and boys. Prior research has exemplified similar disproportionate experiences of harm online, such as sexting scams targeted at teen boys ([Moody, 2023](#)) and disinformation targeted at women of color politicians ([Thakur et al., 2022](#)).

The choice of features and settings can affect the chance of receiving unwanted messages, and in our study, participants demonstrated awareness of that. Some participants reported making the tradeoff of more exposure and engagement with people unknown to them, even though that increased their risk of unwanted messages. A goal for research and industry work should be to ensure that users who make particular choices, e.g. to have a public profile, still have other tools and settings they can use to reduce the risk of unwanted messages.

Further, because we observed certain demographic trends in which participants made decisions about services to use and settings to employ, and concurrently received different levels of unwanted messages, research into how to ensure every user has some set of tools available to them should also focus on the needs of specific communities, who may have different risk profiles and tolerances, as well as different goals for messaging interactions. In the spirit of user-centered research, platforms can work with these user groups to better understand specific needs and to create effective tools and features accordingly.

## Recommendations

### Education about safer messaging platforms

The findings point out that most teenagers and young adults use messaging platforms as a means of communication with people they know in person: family, friends, and classmates. Nevertheless, they don't always consider how the design of the platform can impact which interactions they will encounter. Not only young people themselves make decisions about which platforms to use—several participants noted signing up to platforms they haven't used before in order to participate in a range of school-related chat groups that were set up by educators and other trusted adults (e.g., sporting teams). Thus, both adults and young people must be mindful of the tradeoffs of platform selection.

Further, we did not find evidence to support that participants were aware of, or considered using end-to-end encrypted (E2EE) platforms to increase their safety online, in line with prior research that has highlighted the need for better mental models to understand how E2EE can increase online safety ([Bai et al., 2020](#)). Thus, the benefits of E2EE should be part of any digital literacy efforts aimed at young people, as well as adults, who also have a stake in which platforms young people use.

E2EE messaging services, for their part, should make sure they have blocking, reporting, and other functions that youth indicate they use and want to use in messaging services, as indicated in the recommendations above. In other words, if E2EE services want to attract youth, they should explain the privacy and safety benefits that they can offer from the perspective of what youth care about most.

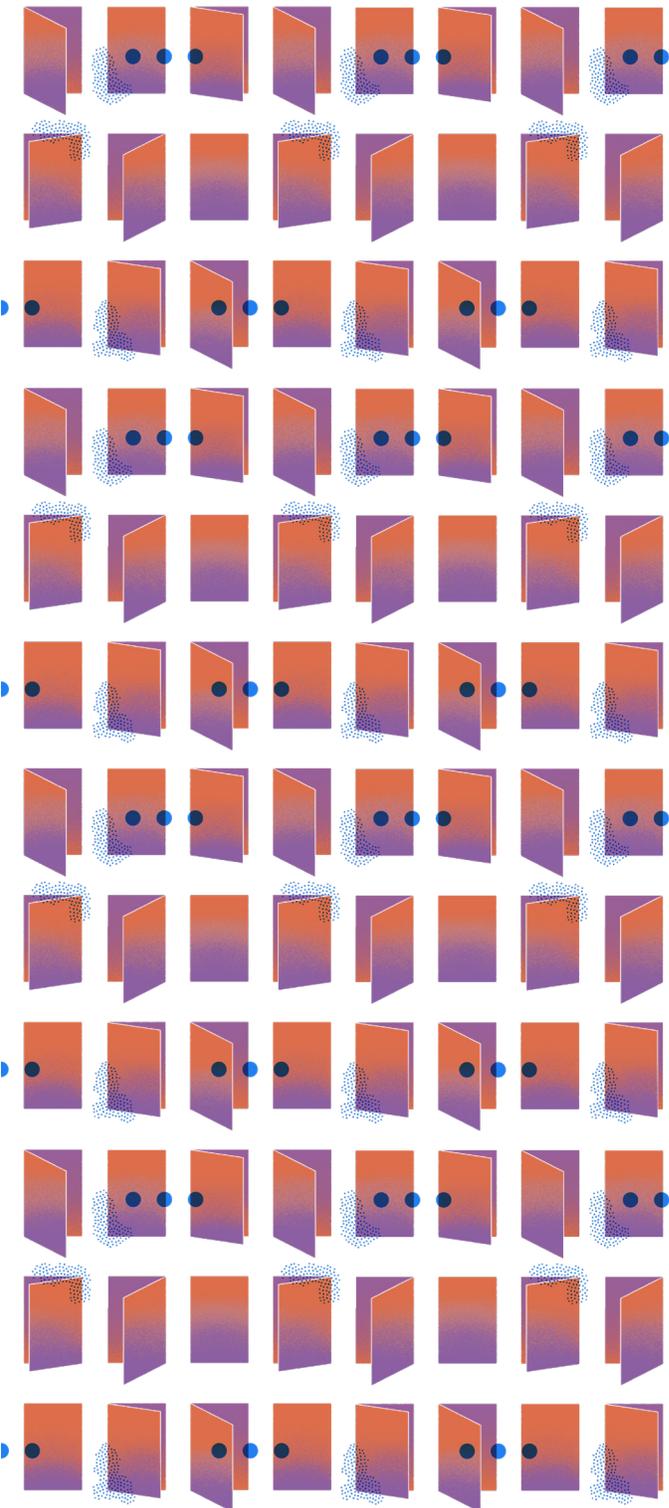


# Conclusion

This study surfaces teenagers' and young adults' personal experiences of using and interacting on messaging platforms, including texting, messaging apps, and social media direct messaging. Using a combination of a survey, a 3-week diary study, and semi-structured interviews, we probed at what some young people in the U.S. deem "unwanted" messages, who these messages come from, and at what frequency, and what current tools and strategies youth use to handle them.

The study finds that many American youth regularly experience unwanted content in messaging. To deal with it, they become active participants and exercise their right to shape their messaging experiences. Youth seem to build strategies, make use of existing tools, and actively resist unwanted messages to keep their interpersonal communication spaces positive.

The right to private, digital communications is a cornerstone of free expression. Thus, as we outline above and as indicated by participants in the study, it is important to provide youth with even more tools and strategies that enable them to use messaging platforms in a way that supports their needs, while keeping their digital environment safe.



# References

- Ali, S., Razi, A., Kim, S., Alsoubai, A., Gracie, J., De Choudhury, M., Wisniewski, P. J., & Stringhini, G. (2022). Understanding the Digital Lives of Youth: Analyzing Media Shared within Safe Versus Unsafe Private Conversations on Instagram. *CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3491102.3501969> [<https://perma.cc/FAS6-BN5X>]
- Alonzo, R. T., Hussain, J., Anderson, K., & Stranges, S. (2019). Interplay between social media use, sleep quality and mental health outcomes in youth: A systematic review. *Sleep Medicine*, 64(64), S365–S365. <https://pubmed.ncbi.nlm.nih.gov/33385767/> [<https://perma.cc/4FDM-4FCK>]
- Anderson, M., & Jiang, J. (2018, November 28). Teens and their experiences on social media. *Pew Research Center: Internet, Science & Tech*. <https://www.pewresearch.org/internet/2018/11/28/teens-and-their-experiences-on-social-media/> [<https://perma.cc/9R8C-G2L5>]
- Anderson, M., Vogels, E. V., Perrin, A., & Rainie, L. (2022, November 16). Connection, Creativity and Drama: Teen Life on Social Media in 2022. *Pew Research Center: Internet, Science & Tech*. <https://www.pewresearch.org/internet/2022/11/16/connection-creativity-and-drama-teen-life-on-social-media-in-2022/> [<https://perma.cc/UM6U-M4A>]
- Bai, W., Pearson, M., Kelley, P. G., & Mazurek, M. L. (2020). Improving Non-Experts' Understanding of End-to-End Encryption: An Exploratory Study. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, 210–219. <https://doi.org/10.1109/EuroSPW51379.2020.00036> [<https://perma.cc/ZN4M-SXZ7>]
- Belair-Gagnon, V., Agur, C., & Frisch, N. (2018). Mobile sourcing: A case study of journalistic norms and usage of chat apps. *Mobile Media & Communication*, 6(1), 53–70. <https://doi.org/10.1177/2050157917725549> [<https://perma.cc/RQ96-TWHC>]
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa> [<https://perma.cc/Y8BG-P56R>]
- Berger, M. N., Taba, M., Marino, J. L., Lim, M. S. C., & Skinner, S. R. (2022). Social Media Use and Health and Well-being of Lesbian, Gay, Bisexual, Transgender, and Queer Youth: Systematic Review. *Journal of Medical Internet Research*, 24(9), e38449. <https://doi.org/10.2196/38449> [<https://perma.cc/MC2V-2KH6>]
- Brown, J. D. & L'Engle, K. L. (2009). *X-Rated: Sexual Attitudes and Behaviors Associated With U.S. Early Adolescents' Exposure to Sexually Explicit Media*. <https://journals.sagepub.com/doi/abs/10.1177/0093650208326465> [<https://perma.cc/T423-JN5W>]
- CDC. (2022). *Youth Risk Behavior Survey Data Summary & Trends Report: 2011-2021*. [https://www.cdc.gov/healthyyouth/data/yrbs/yrbs\\_data\\_summary\\_and\\_trends.htm](https://www.cdc.gov/healthyyouth/data/yrbs/yrbs_data_summary_and_trends.htm) [<https://perma.cc/8TKH-4USD>]
- CDC. (2023, February 13). *U.S. Teen Girls Experiencing Increased Sadness and Violence*. Centers for Disease Control and Prevention. <https://www.cdc.gov/media/releases/2023/p0213-yrbs.html> [<https://perma.cc/EC7N-YBGJ>]

- Copp, J. E., Mumford, E. A., & Taylor, B. G. (2021). Online sexual harassment and cyberbullying in a nationally representative sample of teens: Prevalence, predictors, and consequences. *Journal of Adolescence*, 93(1), 202–211. <https://doi.org/10.1016/j.adolescence.2021.10.003> [<https://perma.cc/4XC9-3AA5>]
- Dæhlen, M. (2021, September 9). *Why are young climate activists being harassed online?* <https://sciencenorway.no/environment-harassment-internet/why-are-young-climate-activists-being-harassed-online/1907887> [<https://perma.cc/X8ZU-Z6AT>]
- Doucleff, M. (2023, April 25). The truth about teens, social media and the mental health crisis. *NPR*. <https://www.npr.org/sections/health-shots/2023/04/25/1171773181/social-media-teens-mental-health> [<https://perma.cc/ND4E-26C6>]
- Grauer, Y. (2021, December 23). *How to Filter Out Twitter Trolls by Using Block Party*. Consumer Reports. <https://www.consumerreports.org/electronics/social-media/filter-out-twitter-trolls-by-using-block-party-a1074825198/> [<https://perma.cc/99XG-99M3>]
- Griffioen, N., Van Rooij, M., Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward improved methods in social media research. *Technology, Mind, and Behavior*, 1(1). <https://doi.org/10.1037/tmb0000005> [<https://perma.cc/GZ5T-A885>]
- Hamm, M. P., Newton, A. S., Chisholm, A., Shulhan, J., Milne, A., Sundar, P., Ennis, H., Scott, S. D., & Hartling, L. (2015). Prevalence and effect of cyberbullying on children and young people: A scoping review of social media studies. *JAMA Pediatrics*, 169(8), 770–777. <https://doi.org/10.1001/jamapediatrics.2015.0944> [<https://perma.cc/8RFQ-EB7W>]
- Holland, G., & Tiggemann, M. (2016). A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. *Body Image*, 17, 100–110. <https://doi.org/10.1016/j.bodyim.2016.02.008> [<https://perma.cc/Uln7-C9HG>]
- Huddleston, J. (2023, April 6). *Would New Legislation Actually Make Kids Safer Online?* | Cato Institute. <https://www.cato.org/briefing-paper/would-new-legislation-actually-make-kids-safer-online-analyzing-consequences-recent> [<https://perma.cc/B25U-WWSD>]
- Jones, L. M., Mitchell, K. J., & Finkelhor, D. (2012). Trends in youth internet victimization: Findings from three youth internet safety surveys 2000–2010. *Journal of Adolescent Health*, 50(2), 179–186. [https://www.jahonline.org/article/S1054-139X\(11\)00338-7/fulltext](https://www.jahonline.org/article/S1054-139X(11)00338-7/fulltext) [<https://perma.cc/9AKB-ZZ88>]
- Leslie, K., Canadian Paediatric Society, & Adolescent Health Committee. (2008). Harm reduction: An approach to reducing risky health behaviours in adolescents. *Paediatrics & Child Health*, 13(1), 53–56. <https://doi.org/10.1093/pch/13.1.53> [<https://perma.cc/PS35-TK32>]
- Lima, C. (2023a, February 2). Analysis | A new bill would ban anyone under 16 from using social media. *Washington Post*. <https://www.washingtonpost.com/politics/2023/02/02/new-bill-would-ban-anyone-under-16-using-social-media/> [<https://perma.cc/6TVQ-X3RA>]
- Lima, C. (2023b, May 10). Officials say social media is hurting teens. Scientists say it's complicated. *Washington Post*. <https://www.washingtonpost.com/politics/2023/05/10/officials-say-social-media-is-hurting-teens-scientists-say-its-complicated/> [<https://perma.cc/2ELQ-L9JU>]

- Lima, C. (2023c, September 5). The young activists shaking up the kids' online safety debate. *Washington Post*. <https://www.washingtonpost.com/technology/2023/09/05/kosa-bill-2023-youth-activists-social-media/> [<https://perma.cc/ZD9A-V87D>]
- Lucero, L. (2017). Safe spaces in online places: Social media and LGBTQ youth. *Multicultural Education Review*, 9, 117–128. <https://doi.org/10.1080/2005615X.2017.1313482> [<https://perma.cc/PE5D-8HCC>]
- Luria, M. & Foulds, N. (2022). Hashtag-Forget: Using Social Media Ephemerality to Support Evolving Identities | *N Factors in Computing Systems*. <https://dl.acm.org/doi/abs/10.1145/3411763.3451734> [<https://perma.cc/C7PY-PN36>]
- Lwin, M. O., Li, B., & Ang, R. P. (2012). Stop bugging me: An examination of adolescents' protection behavior against online harassment. *Journal of Adolescence*, 35(1), 31–41. <https://doi.org/10.1016/j.adolescence.2011.06.007> [<https://perma.cc/238N-5X5Y>]
- Marlatt, G. A., & Witkiewitz, K. (2002). Harm reduction approaches to alcohol use: Health promotion, prevention, and treatment. *Addictive Behaviors*. [https://doi.org/10.1016/S0306-4603\(02\)00294-0](https://doi.org/10.1016/S0306-4603(02)00294-0) [<https://perma.cc/U4EL-NBPR>]
- Microsoft. (2023). *Global Online Safety Survey 2023: Parents' and kids' perceptions of online safety*. <https://www.microsoft.com/en-us/online-safety/digital-civility> [<https://perma.cc/3227-KSSR>]
- Miller, C. C. (2023, May 24). For One Group of Teenagers, Social Media Seems a Clear Net Benefit. *The New York Times*. <https://www.nytimes.com/2023/05/24/upshot/social-media-lgbtq-benefits.html> [<https://perma.cc/2ZKZ-39TZ>]
- Moody, C. (2023, October 2). 'IDK what to do': Thousands of teen boys are being extorted in sexting scams. *Washington Post*. <https://www.washingtonpost.com/parenting/2023/10/02/teen-boys-sextortion/> [<https://perma.cc/5E7F-2DXZ>]
- Moshiri, A. (2023, March 24). Utah is first US state to limit teen social media access. *BBC News*. <https://www.bbc.com/news/world-us-canada-65060733> [<https://perma.cc/R8X2-F59W>]
- Nesi, J., Mann, Supreet, & Robb, Michael B., W. (2023). *Teens and Mental Health: How girls really feel about social media*. <https://www.common sense media.org/research/teens-and-mental-health-how-girls-really-feel-about-social-media> [<https://perma.cc/MT8K-34EV>]
- Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary Studies in Organizational Research: An Introduction and Some Practical Recommendations. *Journal of Personnel Psychology*, 9(2), 79–93. <https://doi.org/10.1027/1866-5888/a000009> [<https://perma.cc/LG6N-LNQY>]
- Pater, J. A., Miller, A. D., & Mynatt, E. D. (2015). This Digital Life: A Neighborhood-Based Study of Adolescents' Lives Online. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2305–2314. <https://doi.org/10.1145/2702123.2702534> [<https://perma.cc/NB8P-VJPB>]
- Razi, A., Alsoubai, A., Kim, S., Ali, S., Stringhini, G., De Choudhury, M., & Wisniewski, P. J. (2023). Sliding into My DMs: Detecting Uncomfortable or Unsafe Sexual Risk Experiences within Instagram Direct Messages Grounded in the Perspective of Youth. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1–29. <https://doi.org/10.1145/3579522> [<https://perma.cc/UB2Y-GY94>]

- Redmiles, E. M., Bodford, J., & Blackwell, L. (2019). "I just want to feel safe": A Diary Study of Safety Perceptions on Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, 13, 405–416. <https://doi.org/10.1609/icwsm.v13i01.3356> [<https://perma.cc/AW8K-WD5A>]
- Ridout, B., & Campbell, A. (2018). The Use of Social Networking Sites in Mental Health Interventions for Young People: Systematic Review. *Journal of Medical Internet Research*, 20(12), e12244. <https://doi.org/10.2196/12244> [<https://perma.cc/7LWN-4NJV>]
- Scott, C. F., Bay-Cheng, L. Y., Prince, M. A., Nochajski, T. H., & Collins, R. L. (2017). Time spent online: Latent profile analyses of emerging adults' social media use. *Computers in Human Behavior*, 75, 311–319. <https://doi.org/10.1016/j.chb.2017.05.026> [<https://perma.cc/6DRF-QSM4>]
- Scott, C. F., Marcu, G., Anderson, R. E., Newman, M. W., & Schoenebeck, S. (2023). Trauma-Informed Social Media: Towards Solutions for Reducing and Healing Online Harm. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–20. <https://doi.org/10.1145/3544548.3581512> [<https://perma.cc/H7ZU-LA5Y>]
- Seabrook, E. M., Kern, M. L., & Rickard, N. S. (2016). Social Networking Sites, Depression, and Anxiety: A Systematic Review. *JMIR Mental Health*, 3(4), e5842. <https://doi.org/10.2196/mental.5842> [<https://perma.cc/3V2J-CMTU>]
- Singer, N. (2023, June 8). Louisiana Passes Bill That Would Require Parental Consent for Kids' Online Accounts. *The New York Times*. <https://www.nytimes.com/2023/06/08/business/louisiana-consent-minors-online.html> [<https://perma.cc/8WCK-5XFD>]
- Single, E. (1996). Harm Reduction as an Alcohol-Prevention Strategy. *Alcohol Health and Research World*, 20(4), 239–243. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6876518/> [<https://perma.cc/5VC9-TGYH>]
- Thakur, D., Hankerson, D. L., Luria, M., Savage, S., Rodriguez, M., & Valdovinos, M. G. (2022). *An Unrepresentative Democracy: How Disinformation and Online Abuse Hinder Women of Color Political Candidates in the United States*. Center for Democracy & Tecnology. <https://cdt.org/insights/an-unrepresentative-democracy-how-disinformation-and-online-abuse-hinder-women-of-color-political-candidates-in-the-united-states/> [<https://perma.cc/38C9-CHCE>]
- Thorn. (2020). *Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking*. <https://www.thorn.org/blog/new-research-from-thorn-22-of-minors-report-having-online-sexual-interactions-with-adults/> [<https://perma.cc/J2KD-574E>]
- Twitter. (2021). *Twitter's new reporting process centers on a human-first design*. <https://blog.twitter.com/content/blog-twitter/common-thread/en/topics/stories/2021/twitters-new-reporting-process-centers-on-a-human-first-design.html> [<https://perma.cc/DHX5-96ER>]
- Uhls, Y. T., Ellison, N. B., & Subrahmanyam, K. (2017). Benefits and Costs of Social Media in Adolescence. *Pediatrics*, 140(Supplement\_2), S67–S70. <https://doi.org/10.1542/peds.2016-1758E> [<https://perma.cc/DM6K-M28M>]
- Valkenburg, P. M. (2022). Social media use and well-being: What we know and what we need to know. *Current Opinion in Psychology*, 45, 101294. <https://doi.org/10.1016/j.copsyc.2021.12.006> [<https://perma.cc/4F5V-8S9Q>]

- Valkenburg, P. M., van Driel, I. I., & Beyens, I. (2022). The associations of active and passive social media use with well-being: A critical scoping review. *New Media & Society*, 24(2), 530–549. <https://doi.org/10.1177/14614448211065425> [<https://perma.cc/DS5S-8XU8>]
- Vilk, V., & Lo, K. (2023). *Shouting into the Void: Why Reporting Abuse to Social Media Platforms Is So Hard and How to Fix It—PEN America*. <https://pen.org/report/shouting-into-the-void/> [<https://perma.cc/7R83-KEFD>]
- Vogels, E. A. (2022, December 15). Teens and Cyberbullying 2022. *Pew Research Center: Internet, Science & Tech*. <https://www.pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/> [<https://perma.cc/7PMU-KNLH>]
- Vogels, E. A., Gelles-Watnick, R., & Massarat, N. (2022, August 10). Teens, Social Media and Technology 2022. *Pew Research Center: Internet, Science & Tech*. <https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/> [<https://perma.cc/7PXV-EACN>]
- Weinstein, E. (2018). The social media see-saw: Positive and negative influences on adolescents' affective well-being. *New Media & Society*, 20(10), 3597–3623. <https://doi.org/10.1177/1461444818755634> [<https://perma.cc/HR8E-GUBK>]
- Whiting, A., & Williams, D. (2013). Why people use social media: A uses and gratifications approach. *Qualitative Market Research: An International Journal*, 16(4), 362–369. <https://doi.org/10.1108/QMR-06-2013-0041> [<https://perma.cc/NL2N-BPRK>]
- Wisniewski, P., Xu, H., Rosson, M. B., Perkins, D. F., & Carroll, J. M. (2016). Dear Diary: Teens Reflect on Their Weekly Online Risk Experiences. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3919–3930. <https://doi.org/10.1145/2858036.2858317> [<https://perma.cc/XAP2-RQRD>]
- Wolak, J., Finkelhor, D., Walsh, W., & Treitman, L. (2018). Sextortion of Minors: Characteristics and Dynamics. *Journal of Adolescent Health*, 62(1), 72–79. <https://doi.org/10.1016/j.jadohealth.2017.08.014> [<https://perma.cc/67TB-QFNS>]
- YouTube. (2018). *More information, faster removals, more people—An update on what we're doing to enforce YouTube's Community Guidelines*. YouTube. <https://blog.youtube/news-and-events/more-information-faster-removals-more/> [<https://perma.cc/MH5E-AMKF>]
- Zimmerman, D. H., & Wieder, D. L. (1977). *The Diary: Diary-Interview Method*. <https://journals.sagepub.com/doi/abs/10.1177/089124167700500406> [<https://perma.cc/QSK5-3V3D>]

 [cdt.org](https://cdt.org)

 [cdt.org/contact](mailto:cdt.org/contact)

 **Center for Democracy & Technology**  
1401 K Street NW, Suite 200  
Washington, D.C. 20005

 202-637-9800

@CenDemTech

