UNBREAKABLE:

Designing for Trustworthiness in Private Messaging

October 2022

Dalberg

SUPPORTED BY

OMIDYAR NETWORK™

Executive Summary

INTRODUCTION

Private messaging platforms like WhatsApp, Signal and Telegram have become an integral part of our day-to-day lives and yet much of what is shared remains private when compared with open forums on the internet.

We know that these private messaging platforms have a profound impact on our digital behavior and emotional well-being, yet it is hard to step back and see the forest for the trees given their ubiquitous nature. While these platforms play an essential role in securing our privacy, they also expose users to a range of risks that undermine their sense of security and trust. This undermining of trust can affect their perceptions of peer platform users, corporations and even governments. We each have our own personal and evolving opinions about how private messaging platforms can be made more trustworthy based on our lived experience, whether through better design choices, more comprehensible policies or more transparent governance models.

"I no longer go by my old name, just because the internet is a place. You can't search me by my documented name, it's a decision I made long ago. Also weary about sharing photos and geotagging, I no longer post often as I used to. I try to keep my face hidden to strangers and mostly identifiable to family."

"I follow up with a phone conversation and see how true it is or go online to verify that story before I choose to re-broadcast it. But in some cases, many people don't verify, they just repost and repost and it causes panic and in a few hours they find out it's fake."

"Sometimes, especially when counseling, the information shared (with me) is very sensitive. For instance, if you are doing counseling and you (message someone that they) should separate from their husband, this (message) is sensitive and you need security."

It's not often we get to hear directly from a diverse cross-section of users spanning geographies about their lived experiences of private messaging platforms. That is the purpose of this research: to both share a cross-cutting view of the experiences and concerns of a diverse range of users – and to connect the dots back to specific design decisions that platform providers should make to improve trustworthiness.

Over the course of ten weeks, our team engaged a total of 185 diverse participants from Colombia, Nigeria and the US. While we have drawn our own conclusions in this report, **we hope that this research can serve as a resource to many different stakeholders as they consider ways that the design of these platforms can be improved, including:**

1 // Platform owners and providers: To negotiate competing product priorities and adjust product planning to address user concerns and diminishing perceptions of trust within messaging experiences.

2 // Policymakers: To better assess the risks that matter to residents and citizens related to security, democracy, and information integrity, understand and prioritize the harms that occur on private messaging platforms, and inform meaningful policy solutions.

3 // Advocacy, Civil society: To buffer advocacy efforts with data points and anecdotal evidence of the harms a diverse set of global users experience on private messaging platforms and examples of concrete changes that could improve trustworthiness.

4 // Researchers: To equip the trust and safety research field with actionable user-centered data, and offer a blueprint for mixed methods methodologies focused on user experiences of private messaging platforms. Researchers have the opportunity to replicate this approach in other markets and with other communities to further quantify these harms.

5 // UX designers: To augment their own user research and data analytics, and influence product priorities in line with user trustworthiness.

6 // Platform value chain players & governments offering services on private messaging platforms: To assess the potential impact of user concerns on the trustworthiness of services they offer on private messaging platforms. User perceptions of private messaging platform trustworthiness will shape their trust in services offered by governments and other value chain players on private messaging platforms, as we have seen during the pandemic. A lack of trust will likely lead to less engagement with both the messaging platforms and corresponding services offered on top.

FINDINGS

Since these markets are distinct and individual journeys within private messaging platforms are personal, there is always a risk of generalization when attempting to summarize this sort of user experience research. Nonetheless, there are some common patterns that seem to transcend these differences:

A// We found that people across very different markets have become incredibly sophisticated in how they understand and navigate the intricacies of these platforms.

Across geographies, most users have built up fairly complex ways of engaging and adapting to risks and concerns as they perceive them (for ex: switching into airplane mode so that other users won't be able to tell if they have read their messages). This finding in particular calls into question the assumption that people are not likely to adjust their preferences even if these options were made more easily accessible. Even with insufficient features, people are finding a myriad of workarounds to address gaps and minimize shortcomings.

B// Heightened perception of risk generally arises in response to specific situations, not all of which can be attributed entirely to the platform providers themselves.

The risks that are most top of mind vary by market (e.g., in Nigeria, it was fraud. In the US, it was corporate surveillance). While private messaging platforms are responsible for some of the vulnerabilities and design gaps which make the risks more likely to materialize into harm, some factors leading to risks – for instance, cultural norms or existence of bad actors – are not fully preventable by messaging service providers. Still, because platform design and governance can enable and exacerbate these harms, platform providers have a responsibility to both understand them and take steps to mitigate them. Given these complexities, users generally do not have a full understanding of where to direct or who to attribute their concerns to. Often, they take on a sense of responsibility for themselves ("I should have known better") so their response choices bear little connection to the risk itself, and tend to fade over time. Regardless, perceptions of trust in messaging platforms change rapidly and irreversibly in response to these acute situations.

C// Users also face a huge gap in terms of recourse and redress, which is a critical element of trustworthiness.

The platforms themselves do not offer many clear affordances for seeking redress, particularly affordances that do not come with some reciprocal social costs (flagging another person's bad behavior or misinformation often leaves users more vulnerable to harassment).

D// Most users do not feel that they have real choice and can "venue-shop" based on personal preferences.

Even those with heightened awareness (human rights activists, for example) or high levels of technical knowledge find it practically challenging to avoid defaulting to the most common and pervasive platforms (WhatsApp in most cases). Because of this, choice alone cannot be held up as the silver bullet for ensuring better practices in the messaging platform market. While it's critical that new entrants prioritize trustworthy and safe platform design, existing platforms also need to take user concerns seriously and commit to enhancing trustworthiness with, inter alia, their design choices.

APPROACH

The user experience of platforms like WhatsApp have become second nature to users in Colombia, Nigeria and the US. The design choices of platform providers are something users work around every day, sometimes unaware of how they shape both their personal behavior and that of others, as well as their very expectations of what private messaging platforms can and should be. Human-Centered Design (HCD) approaches help us to make apparent dynamics and behaviors that are latent or under the surface.

For this reason, it was critical that we take a participatory, Human-Centered Design (HCD) approach to pierce this veil and bring forward the voices and cross-cutting concerns of private messaging platform users. What risks are they most aware of when using messaging platforms? Where and how do these risks show up in their day-to-day behavior? Who do they hold responsible, and do they feel that they have any opportunity for recourse or redress? What choices and tradeoffs are they comfortable making to safeguard their data privacy and security and where do they feel powerless?

To gain insight into these questions, our team engaged a total of 185 participants over the course of 10 weeks. We met with ecosystem experts from several countries in the context of co-creation workshops, and community leaders and platform users in 1-on-1 and small group discussions in Colombia, Nigeria and the US.

All sessions were conducted remotely due to COVID-19 except for the community-led sessions. A breakdown of our research is as follows:

Colombia

50 total participants,

- 10 in depth 1:1 remote sessions,
- 9 remote small group discussions,
- 31 in person interviews led by community leaders

Nigeria

64 total participants,

- 10 in depth 1:1 remote sessions,
- 4 remote small group discussions,
- 42 in person interviews led by community leaders

USA

54 total participants,

- 10 in depth 1:1 remote sessions,
- 4 remote small group discussions,
- 32 in person interviews led by community leaders

REPORT CONTENTS

In such opaque and highly personal environments, how might we better understand opportunities to intervene to address a set of common concerns? What would a better experience look like? To fill in that picture, this report breaks down what we heard into the following areas of analysis:

→ EXPERIENCES: It is critical to first contextualize these findings within a holistic view of people's everyday experiences and patterns of behavior on private messaging platforms. This report shares three sets of representative experiences from each market we looked at as a way of highlighting commonalities and differences from user perspectives.

→ HARMS: We identified the key risks leading to various harms that are most important to users across the three markets and are likely to have the biggest impact on their sense of trustworthiness. Any future design improvements should start by prioritizing the risks that are most important to the users themselves.

→ GAPS: The lack of mental models (other than text messaging) for how private messaging platforms work creates many gaps for users as they navigate risks and experiences of harm. Users lack supporting resources to evaluate and attribute their growing sense of concern. Who should they trust (their group admin? WhatsApp customer support?) when they encounter these gaps? In most cases the platforms provide few paths to recourse in the moment and little to no feedback to understand how their concerns might be resolved.

→ DESIGN OPPORTUNITIES: What can design really accomplish to minimize these risks, fill in these gaps and build trust once it is lost? Our research identified many pressing concerns regarding trustworthiness related to common elements of private messaging platform design, such as: group dynamics, misinformation and generalized anxiety relating to mental health. In each case, it is not hard to begin to see a path to provide users with better tools to manage risk and make informed choices – a number of which we illustrate with sample designs that were prototyped and tested with users to further inspire change. These designs are not prescriptive: they are meant to be representative of how a private messaging platform provider MIGHT address a specific gap or design opportunity. We recognize that any design changes are likely to come with tradeoffs and potentially impact business goals related to customer growth and engagement. **Some key areas where users responded most positively to potential design improvements include:**

- Securing and/or modifying account information
- Providing accessible & tailored security & privacy controls
- Providing support mechanisms & emergency controls
- Improving verification & permission mechanisms
- Improving administrative & management tools

CONCLUSION

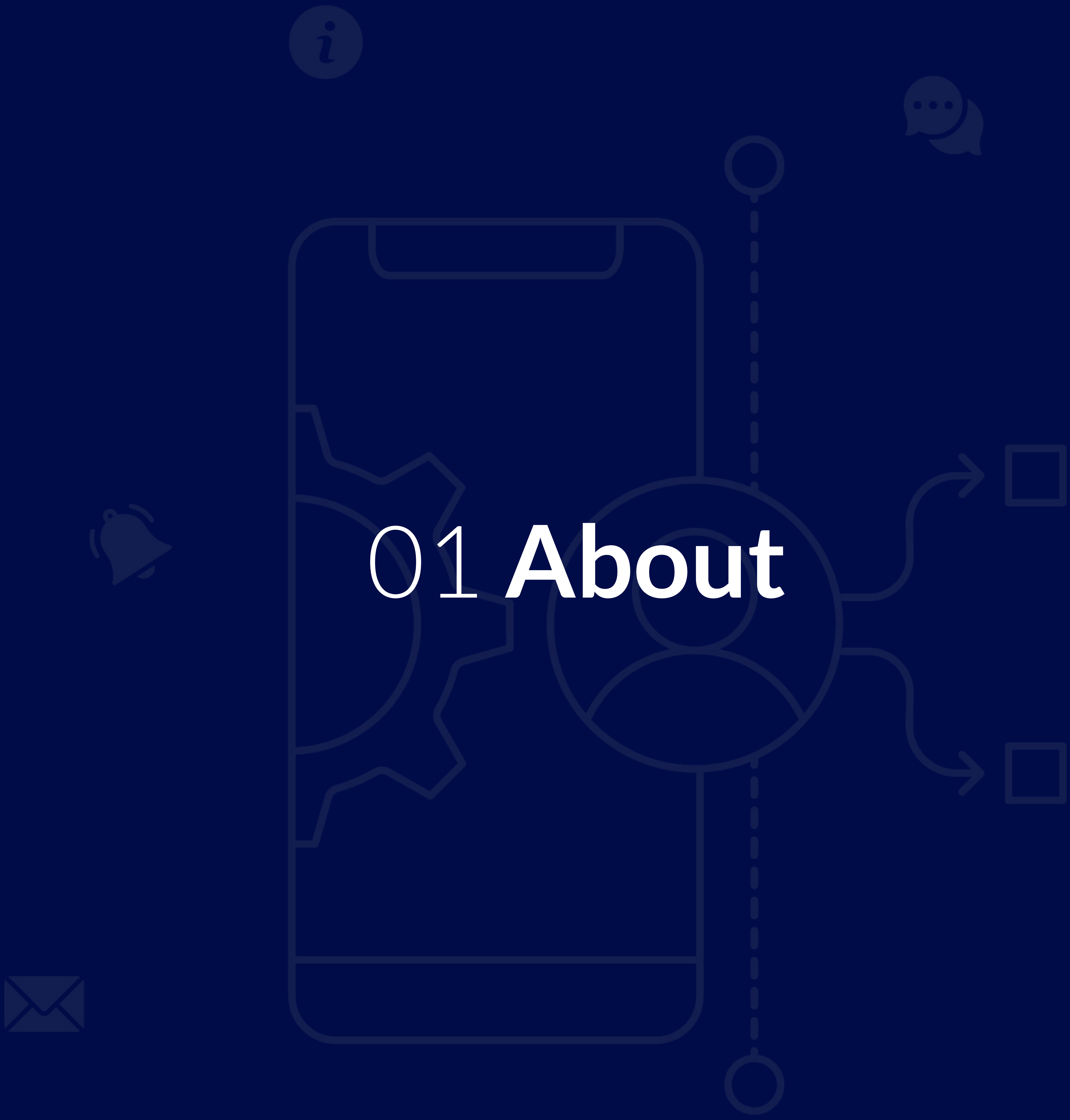
There is much that private messaging platform providers can do differently if they choose to prioritize trustworthiness in platform design. User choice is not a sufficient excuse to justify the current shortcomings. Our research suggested that few users feel that they have real choice in the market despite the availability of multiple private messaging platforms.

Pointing to the retention and engagement of users as a sign that they are satisfied with current interaction models and tradeoffs does not ring true. We heard consistently that the tradeoffs of leaving a dominant environment, – WhatsApp in most cases, – are incredibly daunting for all users, even the most security-conscious like human rights activists. Platform providers have a long way to go in bettering the design of their services, (though we are seeing discrete instances of intentional trustworthy design with recent changes by WhatsApp that allow users to leave group chats without alerting others, for example). We would also encourage private messaging service providers to be transparent in how they engage users in regular cycles of feedback using the sort of methods we employed for our research study – not just analyze user data behind closed walls.

The dialogue around trustworthiness has remained at a theoretical level for too long. We hope these findings will help those advocating for change (whether policymakers, researchers or activists) point to real and concrete design choices that can increase trustworthiness on private messaging platforms. We also hope this research offers stakeholders a provocation to consider more fundamental changes to the environments in which these platforms operate, whether it be business models or interoperability standards. In that sense, these recommendations are complementary to a number of related initiatives for fighting disinformation and dangerous speech on private messaging platforms – including research, technical partnerships, dialogue and convening with policymakers and technology leaders, and public advocacy — and should be seen as an integrated part of this broader effort.

The most distinctive outputs of this study– concrete, user-informed design recommendations – are just a starting point. To some, our design recommendations might seem incremental in the face of the scale and severity of user risks and concerns experienced on private messaging platforms. These recommendations do not point to a comprehensive end state which, if implemented, would satisfy all user needs and address all experiences of harm. Instead, the design recommendations in this report can provide a path towards beginning to address these harms if they are implemented within a user-centered and iterative process. They can help pave the way for a more trustworthy messaging future.

Table of contents



01 About

Project overview

> ***"I no longer go by my old name, just because the internet is a s— place. you can't search me by my document name, it's a decision I made long ago. Also weary about sharing photos and geotagging, I no longer post often as I used to. I try to keep my face hidden to strangers and mostly identifiable to family."***

In the digital realm, end-to-end private messaging plays an important role in upholding individual rights to privacy and free speech. Platforms like WhatsApp and Signal allow residents to communicate with each other without the fear of governments, advertisers, or even snooping family members listening in or moderating the content of their communication. But these digital environments are not without many harms that undermine end user trustworthiness. Given their widespread adoption, it is critical that platform providers prioritize design choices that strengthen, not undermine, trust. That sounds great in principle, but where should they turn for guidance?

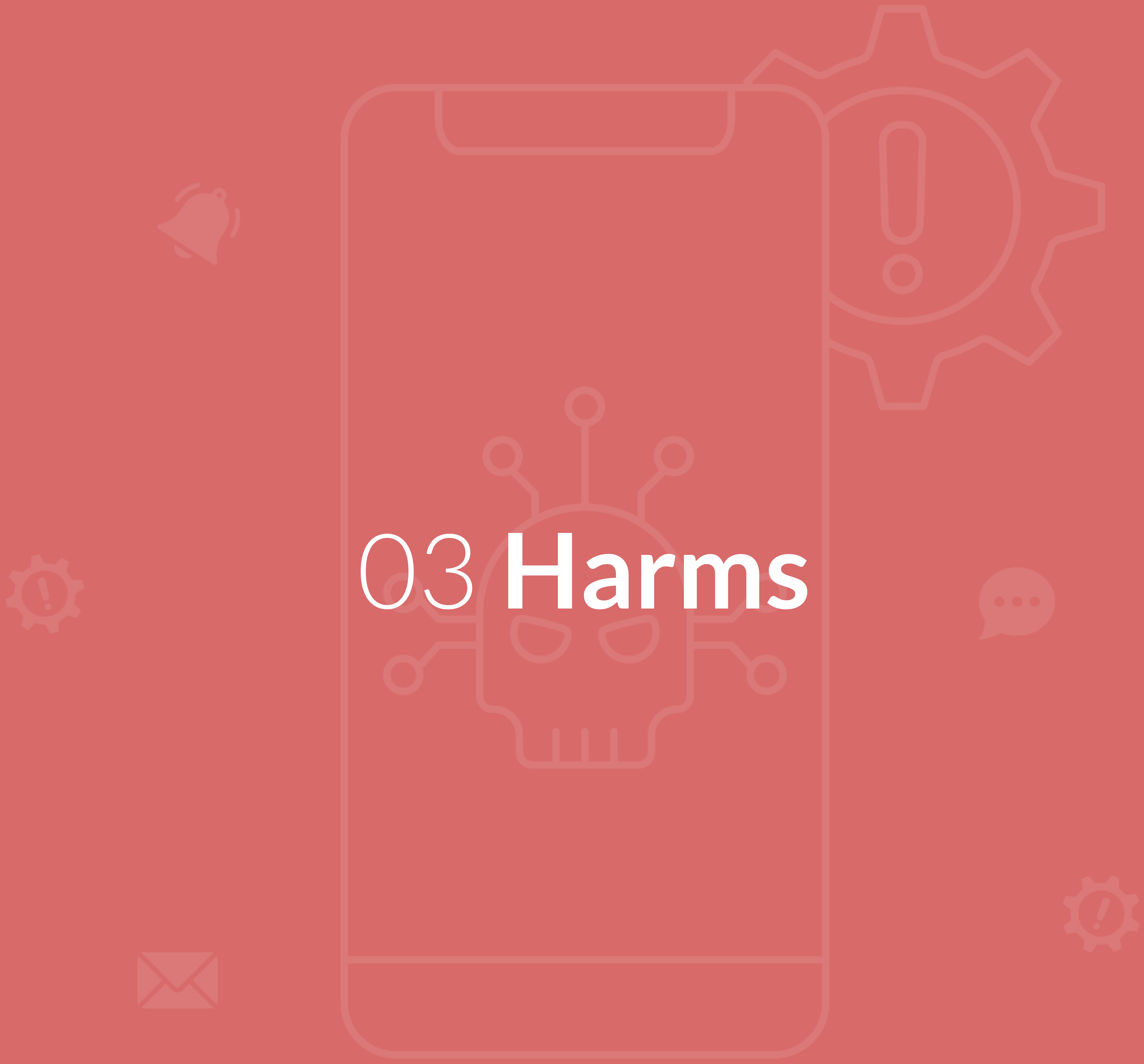
The goal of this report is to share design opportunities that address harms that exist on private messaging apps and matter the most to a globally diverse selection of individuals. These design opportunities aim to enhance individual experience to provide a safer and secure messaging environment.

What is at stake? For participants, private messaging can deliver offensive and inappropriate content, it can channel disinformation and "fake news", and it can be used by nefarious actors to defraud unsophisticated or unsuspecting individuals. For example, our research reveals the rampant cases of hacking and scamming in both Nigeria and Colombia leading participants to look for alternative options (e.g., 3rd party apps) to protecting their accounts and verifying unknown contacts, even though these 3rd party apps compromise their privacy and security.

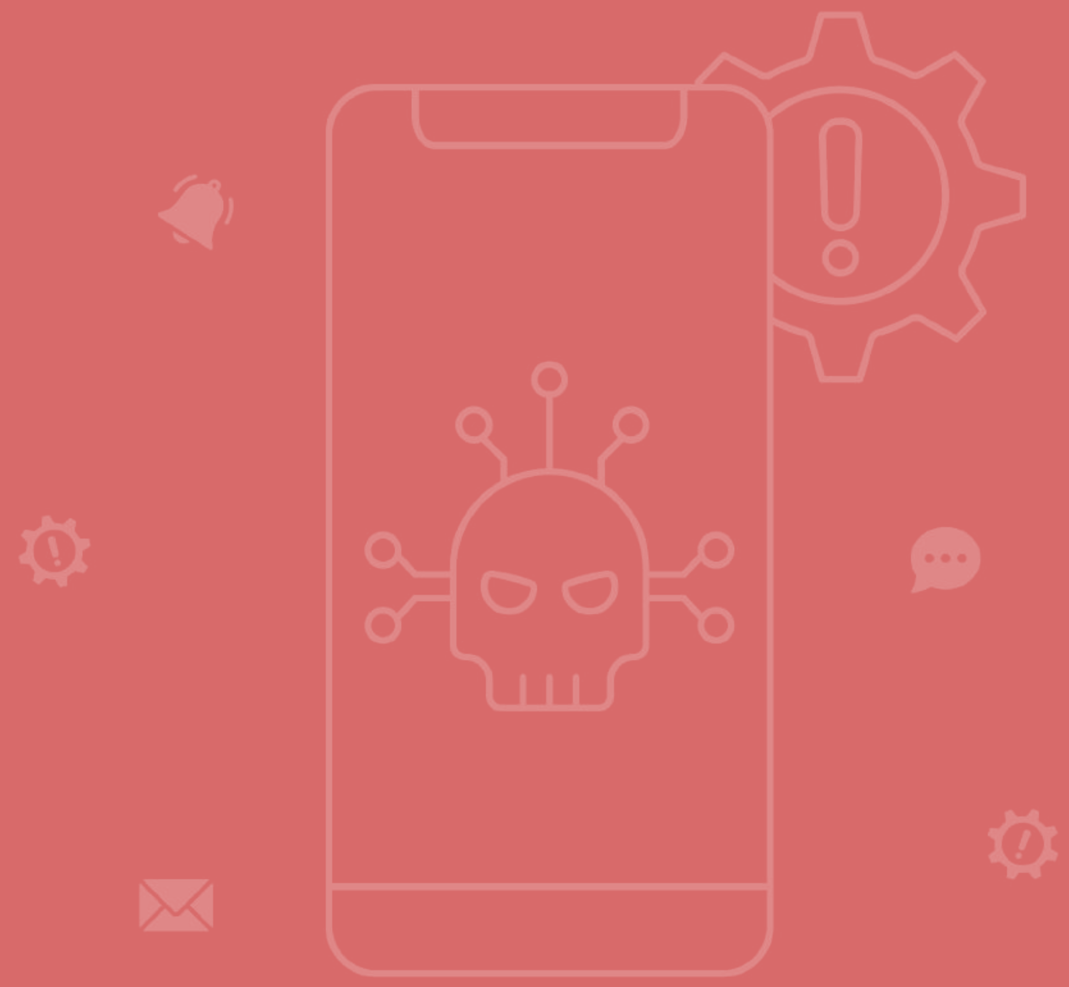
Platform providers may be tempted to view widespread adoption and high levels of engagement by individuals and groups as a reason to feel confident in current design choices. But our research participants are deeply concerned about their level of dependence on messaging services and their lack of control over the experiences within these messaging environments. Encryption alone does not confer a sense of safety and security, as it is poorly understood by almost everyone we spoke with. Participants are unsure of whom to trust – even scrutinizing the statements and reported behavior of senior executives like Mark Zuckerberg (Meta Platforms) or Pavel Durov (Telegram Messenger) as proxies for the relative integrity of Whatsapp or Telegram. It is only by investing in more effective and better-informed design choices that providers can help individuals and groups manage the risks inherent in these platforms; and work together to create chat environments that are safe, supportive and responsive to our changing needs.

This research looked to surface and test a preliminary set of design solutions that are likely to reduce the deleterious potential of private messaging platforms. As civil society organizations continue to push for more responsible technology, we hope our findings can be used by private messaging providers and other third-party players to build on the emerging ideas and test and implement potential solutions. While we do not expect our work to be the end-point in designing the right answer, we do hope it is an important step in that direction.

03 Harms



Harms



The use of private messaging platforms can expose individuals to a range of harms that impact trustworthiness.

In this section, we cover six harms that were most frequently mentioned by the participants we spoke to in Nigeria, Colombia and the United States. For each one, we explore the different ways that these impact the user experience as well as the platform design gaps that can amplify these harms.

1. Vulnerability to **adverse mental health impacts**

The negative psychological impacts that arise from using private messaging platforms.

2. Vulnerability to **targeted harassment for youth and young adults**

The use of private messaging platforms to exploit the vulnerability of youth and minors.

3. Vulnerability to **manipulation or exposure to offensive content**

The use of private messaging platforms to knowingly or unknowingly circulate content that can be perceived as being hateful, offensive and/or misleading.

4. Vulnerability to **hacking, scamming, blackmailing, extortion, fraud, and harassment**

The direct misuse of private messaging platforms by adults as distinct from those affecting youth and children.

5. Vulnerability to **encryption and data breaches via modified and third-party supporting platforms**

The different ways that private messaging platforms' security, privacy and encryption features are bypassed by individuals through the use of modified and third-party supporting private messaging platform apps.

6. Vulnerability to **digital surveillance and monitoring**

The potential use of private messaging platforms by governments and corporations to survey or monitor individuals.

Design gaps that enable these harms

Sitting across all six harms we found a total of seven product design gaps that seem to have the biggest impact on platform trustworthiness for the participants we spoke with.

A Easy access to personal identifying data

Personal information on most private messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

B Limited verification and consent focused features for contacts and groups

There is a lack of mechanisms for verifying contacts or groups, while permissions for new contacts and groups are either not set by default or are non-existent.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within complex menu structures.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app.

E Limited support and lack of adequate reporting mechanisms

From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

F Limited content management tools

There are very few features that help individuals manage and organize the content they receive. This can cause some individuals, mainly those who participate in large groups and/ or receive large amounts of content, to feel overwhelmed.

G Lack of transparency regarding access to personal data

There are gaps around who can access, use, and potentially misuse personal data *(e.g., how and if companies and governments can access personal data)*. And private messaging platforms don't communicate transparently and in a user-friendly way how they manage and protect individuals' data.

Not all design gaps are relevant to all harms. But our research indicates that each of these design gaps has an adverse effect on more than one harm leading to a cumulative effect that raised deep concerns among every user we spoke with. The chart below presents a summary view of the specific design gaps and their negative impacts across the primary harms that emerged from our discussions across the three countries.

HARMS IMPACTING USER TRUSTWORTHINESS PRODUCT DESIGN GAPS	1. Vulnerability to adverse mental health impacts	2. Vulnerability to targeted harassment for youth and young adults	3. Vulnerability to manipulation (misleading content, mis/disinformation) or exposure to offensive content	4. Vulnerability to hacking, scamming, blackmailing, extortion, fraud, and harassment	5. Vulnerability to encryption and data breaches via modified and third-party supporting platforms	6. Vulnerability to digital surveillance and monitoring
A. Easy access to personal identifying data	✓	✓	✓	✓		✓
B. Limited verification and consent focused features for contacts and groups	✓	✓	✓	✓		
C. Generalized and hidden privacy & security controls for contacts and groups	✓	✓	✓	✓	✓	
D. Existence of modified (MOD) and third-party supporting apps ecosystem	✓	✓	✓	✓	✓	
E. Limited support and lack of adequate reporting mechanisms	✓	✓	✓	✓		
F. Limited content management tools	✓	✓	✓		✓	
G. Lack of transparency regarding access to personal data	✓				✓	✓

1.

Vulnerability to **adverse mental health impacts**

This harm covers the negative psychological impacts that arise from using private messaging platforms. This includes psychological difficulties such as screen-time fatigue, overdependence on private messaging platforms, impact of large volumes of notifications and content, unclear or uncertain exposure to digital surveillance, impact of exposing digital behaviors to other individuals (e.g., user is online, read receipts), and social pressures from other contacts, whether real or perceived.

Individual's experience of this harm

The adverse impact of platforms on mental health was expressed as being of medium to high concern for many of the participants we spoke to in all three countries, particularly for participants in Colombia and the US. Most participants described their concern as being connected to a distrust of technology which in turn stems from worries related to increases in:

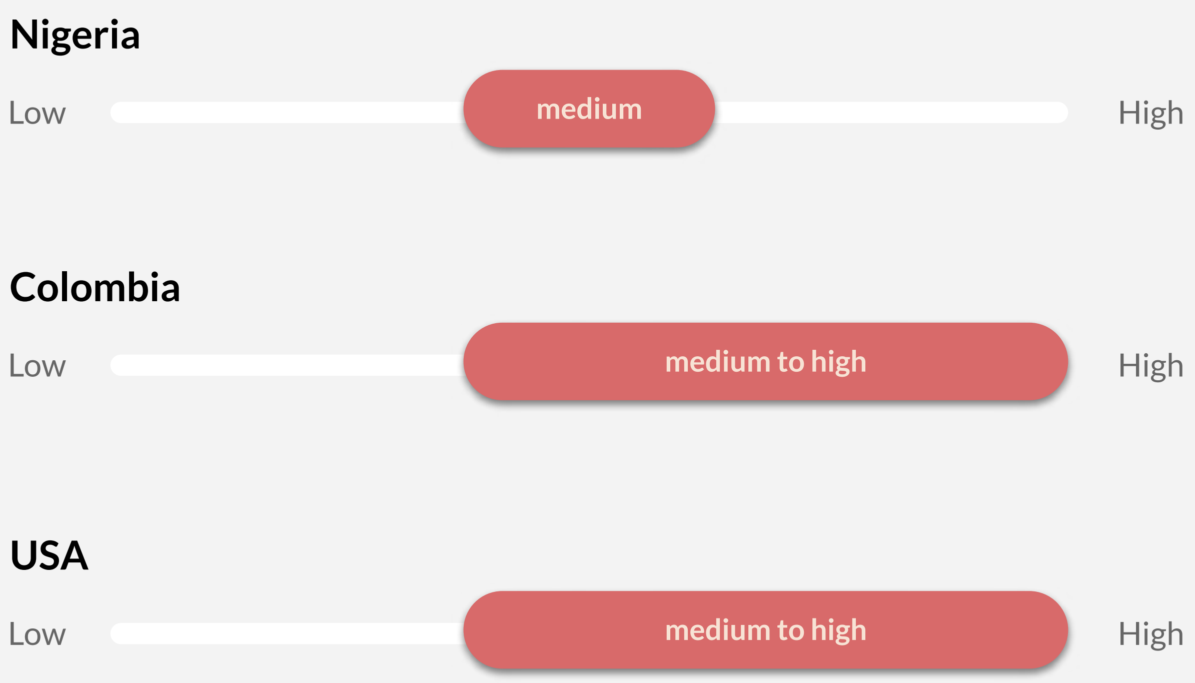
- Digital surveillance and monitoring by both corporations and governments
- Frequent security breaches and the difficulty of guaranteeing privacy
- Overdependence on using private messaging platforms in general
- Anxiety related to the expanding use of private messaging platforms in their day-to-day lives

Participants in all three countries shared some common mental health concerns, often tied to features that are meant to increase platforms' stickiness. For example, in all three countries (particularly in Colombia) features that communicate user behavior to others (e.g., user is online, read receipts, user is writing) were frequently pointed to as the cause of unwelcome social pressures. In addition, individuals vulnerable to privacy and security concerns, such as activists and dissenters, shared the cumulative mental stress that resulted from concerns that they might be targeted at any time for surveillance by the government and police. Other less-vulnerable individuals similarly expressed feeling a heightened sense of surveillance concerns in relation to the misuse of their data by corporations, particularly in the US. Due to this, some participants turned to modified (MOD) or third-party supporting private messaging platform apps to access workarounds that could relieve these pressures.

The map on the next page presents a cross-country perspective on the impact of platforms on mental health from participants in Colombia, Nigeria and the United States. Particular attention is paid to the design gaps that contribute to the widespread experience of this harm by the participants we spoke with.

Level of concern expressed by participants

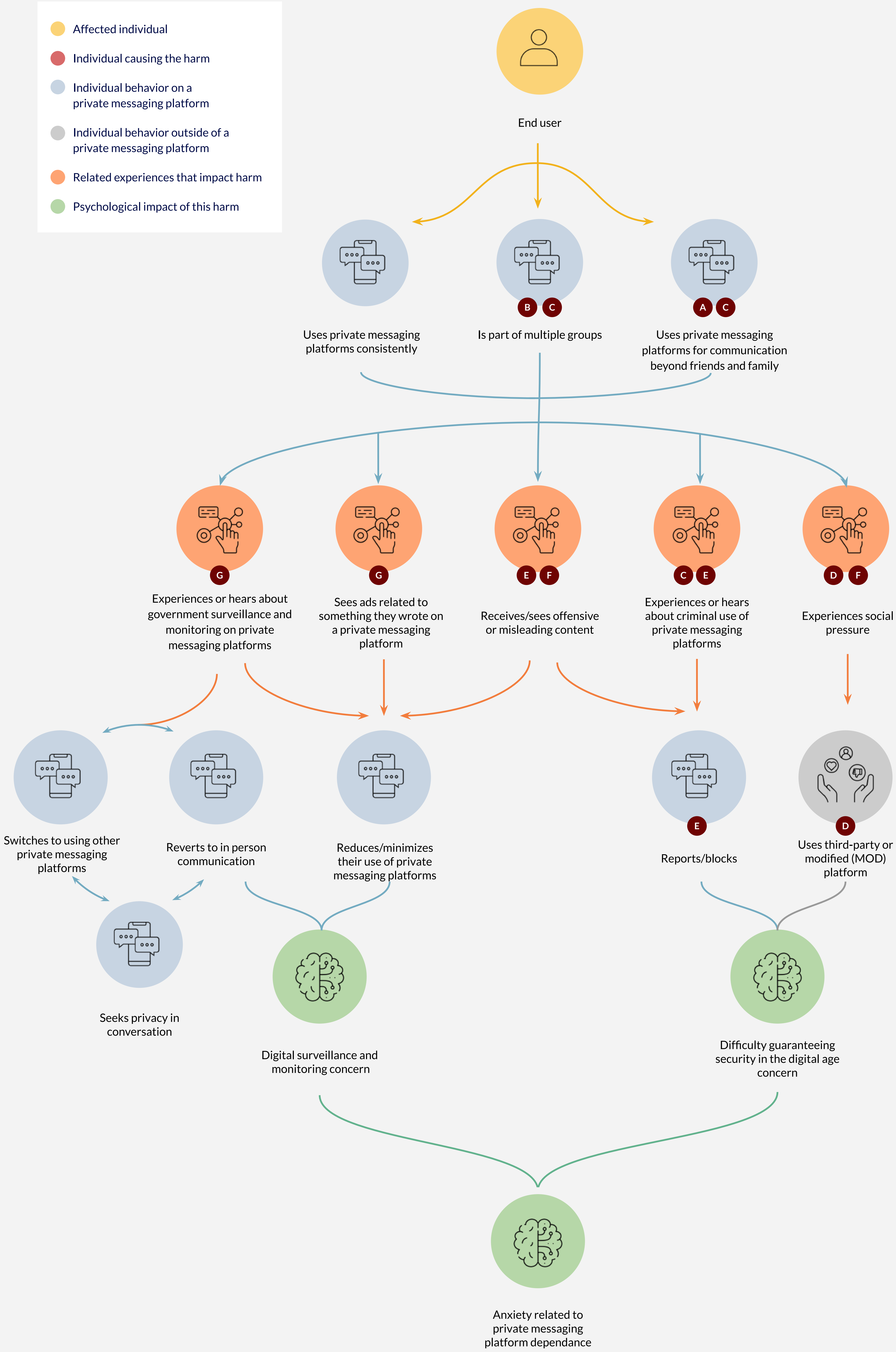
The scales below reflect the relative range of concern that participants we spoke with expressed for this harm across different countries.



> *It happens to me with my clients that sometimes I leave my computer on and they say I wrote to you at 2am and I saw you were online and you didn't answer me.... Sometimes I have to do everything so that they don't see me online, like putting it in airplane mode.*
>
> 38-year-old Colombian man, using WhatsApp, Telegram and FB Messenger

> *My phone notification gets clogged up, you have so many apps telling you to look at me. I don't even have time to eat, let alone look at all of this. Most of the notifications are with those apps that have many members. So I need to look at the settings of these.*
>
> 33-year-old Nigerian man

Vulnerability to **adverse mental health impacts** - illustrative process



- ● Affected individual
- ● Individual causing the harm
- ● Individual behavior on a private messaging platform
- ● Individual behavior outside of a private messaging platform
- ● Related experiences that impact harm
- ● Psychological impact of this harm

End user

Uses private messaging platforms consistently

Is part of multiple groups (B, C)

Uses private messaging platforms for communication beyond friends and family (A, C)

Experiences or hears about government surveillance and monitoring on private messaging platforms (G)

Sees ads related to something they wrote on a private messaging platform (G)

Receives/sees offensive or misleading content (E, F)

Experiences or hears about criminal use of private messaging platforms (C, E)

Experiences social pressure (D, F)

Switches to using other private messaging platforms

Reverts to in person communication

Reduces/minimizes their use of private messaging platforms

Reports/blocks (E)

Uses third-party or modified (MOD) platform (D)

Seeks privacy in conversation

Digital surveillance and monitoring concern

Difficulty guaranteeing security in the digital age concern

Anxiety related to private messaging platform dependance

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

A Easy access to personal identifying data

Personal information on most messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

B Limited verification and consent-focused features for contacts and groups

There is a lack of mechanisms for verifying contacts or groups, while permissions for new contacts and groups are either not set by default or are non-existent.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within little-used menu structures.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app without others knowing.

E Limited support and lack of adequate reporting mechanisms

From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

F Limited content management tools

There are very few features that help individuals better manage and organize the content they receive.

G Lack of transparency regarding access to personal data

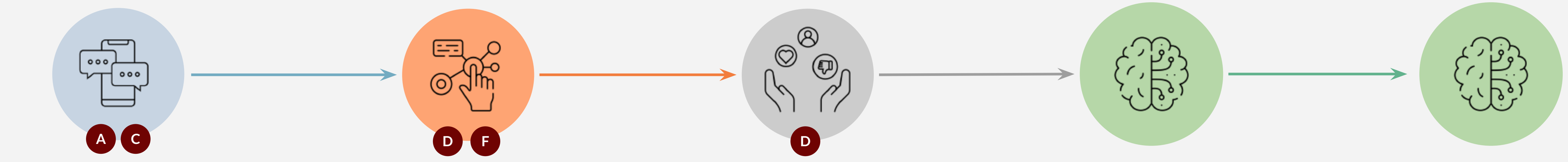
There are gaps around who can access, use, and potentially misuse personal data *(e.g., how and if companies and governments can access personal data)*. At the same time, messaging platforms don't communicate transparently and in a user-friendly way how they manage and protect personal data.

Vulnerability to **adverse mental health impacts** - illustrative journey

Jennifer - *Globe Trotter*

Jennifer is an interior designer and loves to travel the world. She has friends across continents whom she keeps in touch with often. She also manages many groups on private messaging apps that stretch across different time zones and continents. She uses several platforms in order to stay in touch with all of her contacts. She often uses WhatsApp because many of her contacts are on it, but she prefers to use Signal because it's simple and has fewer distractions.

> ***When I was in Bali I connected with some people on WhatsApp there, and when I got back to Sweden, Facebook tried to connect me to those people. Then I realized okay this is how it is connected together. This was not a coincidence.***



Uses private messaging platforms for communication beyond friends and family

Since Jennifer has many contacts across the globe, she uses multiple messaging platforms to communicate with family and friends.

She is open to using any private messaging platform that her contacts prefer, hence ends up using multiple platforms.

Experiences social pressure

Jennifer often feels overwhelmed when communicating. She feels people are always demanding her to respond immediately just because they see she's online or has read a message that was sent to her. These kinds of expectations, coupled with the use of multiple platforms, put a lot of pressure on her when using these platforms.

Uses 3rd party or clone app

Friends have shared tricks of measures Jennifer can use to block others from seeing that she's online, like responding while being on airplane mode. Still, Jennifer has decided to download a modified (MOD) messaging app to get more control over her privacy.

Difficulty guaranteeing security and privacy in the digital age

While the modified messaging app has helped Jennifer get more control over her privacy, she has noticed that the app also offers features that violate the privacy of others, like the ability to read messages that have been deleted. This has made her suspicious of how the app is getting access to this data.

Anxiety related to private messaging platform dependance

Jennifer has not been able to find a workaround that offers her control over her privacy when communicating.

Constant notifications and growing expectations when communicating has become a constant feature. She is yet to find a suitable solution that does not compromise her privacy or that of her contacts.

🟡 Affected individual

🔴 Individual causing the harm

🔵 Individual behavior on a private messaging platform

⚪ Individual behavior outside of a private messaging platform

🟠 Related experiences that impact harm

🟢 Psychological impact of this harm

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

A Easy access to personal identifying data

Personal information on most messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within little-used menu structures.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app without others knowing.

F Limited content management tools

There are very few features that help individuals better manage and organize the content they receive.

Product opportunities

We identified seven primary product opportunities to address the design gaps that contribute to the proliferation of this harm.

Design Opportunity 1

Securing and/or modifying account information

Product design gaps addressed

A Easy access to personal identifying data

Design Opportunity 2

Improving verification & permission mechanisms

Product design gaps addressed

B Limited verification and consent-focused features for contacts and groups

Design Opportunity 3

Providing accessible & tailored security & privacy controls

Product design gaps addressed

C Generalized and hidden privacy & security controls for contacts & groups

Design Opportunity 4

Managing access to modified & third-party supporting platforms

Product design gaps addressed

D Infringement by modified (MOD) & third-party supporting apps ecosystem

Design Opportunity 5

Providing support mechanisms & emergency controls

Product design gaps addressed

E Limited support and lack of adequate reporting mechanisms

Design Opportunity 6

Improving administrative & management tools

Product design gaps addressed

F Limited content management tools

Design Opportunity 7

Providing data use transparency & the ability to manage data

Product design gaps addressed

G Lack of transparency regarding access to personal data

2.

Vulnerability to **targeted harassment for youth and young adults**

This harm covers the use of private messaging platforms to exploit the vulnerability of youth and minors which our users saw as a distinct concern from more general misuses. This includes physical, sexual, and psychological abuses that are conducted or aided via messaging platforms such as, phishing, circulation of child sexual abuse material, hacking, exposure to online predators, and cyberbullying.

Individual's experience of this harm

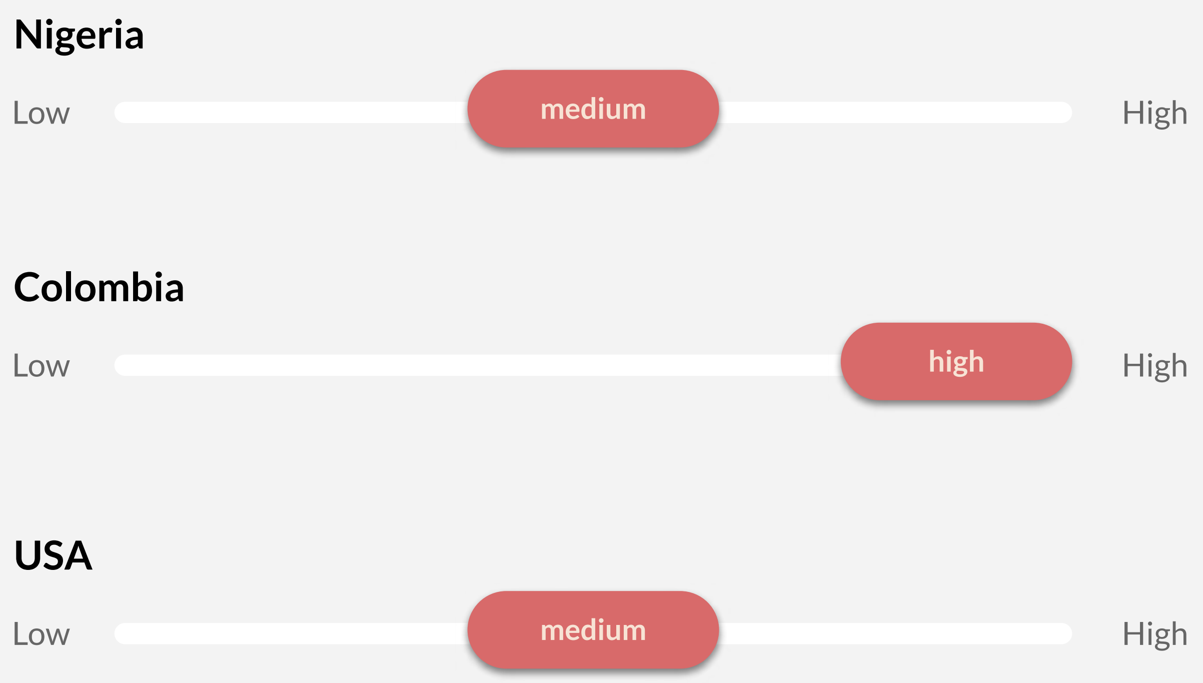
The use of private messaging platforms to exploit the vulnerability of youth and minors was expressed as being of medium concern for participants in Nigeria and the US and of extremely high concern for participants in Colombia. While all Colombian participants worried about this harm, Colombian parents expressed the most severe anguish over this risk. As such, in an effort to gain greater visibility of what could be affecting their children, they often turned to the use of third-party apps to clone the account of their children on their phones. This, however created additional stress as they tried to manage the right level of parental monitoring as their children grow up across multiple app environments.

While our research points to the need for protecting the vulnerability of minors and youth against criminal and abusive usage of private messaging platforms, we did not gather perspectives from minors (18 and under) directly as part of this study. Additional research would be needed to further understand the specific needs and use cases from the perspective of minors. The youth we spoke to (ages 18-20) expressed unique uses for private messaging platforms, including the use of WhatsApp and Telegram as dating apps, study groups, and places to meet new friends outside their circle.

The map on the next page presents a cross-country perspective on the use of private messaging platforms to exploit the vulnerability of youth and minors from the participants in Colombia, Nigeria, and the United States. Particular attention is paid to the design gaps that contribute to the widespread experience of this harm by the participants we spoke with.

Level of concern expressed by participants

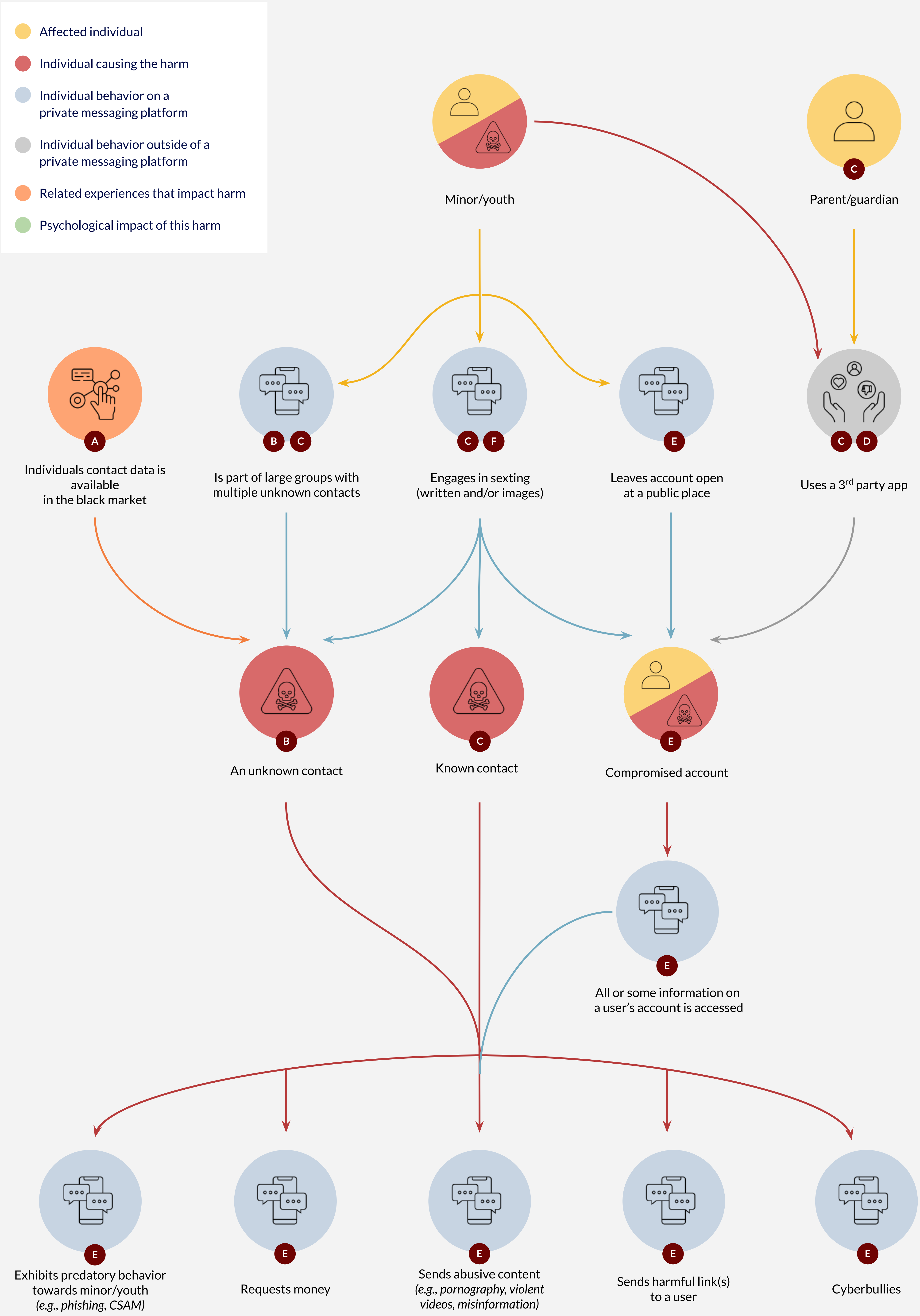
The scales below reflect the relative range of concern that participants we spoke with expressed for this harm across different countries.



> *"At the beginning of the pandemic we started to use WhatsApp like a dating app. Posting on each others WhatsApp stories the link to our account so other people could see if they liked it or not."*
>
> 18-year-old Colombian boy, using WhatsApp, Telegram and FB Messenger

> *"When you download an app you never check what instructions or terms and conditions agreement it has, you just install it and that's it because you need it."*
>
> 18-year-old Colombian boy

Vulnerability to **targeted harassment for youth and young adults** - illustrative process



● Affected individual
● Individual causing the harm
● Individual behavior on a private messaging platform
● Individual behavior outside of a private messaging platform
● Related experiences that impact harm
● Psychological impact of this harm

Minor/youth

C Parent/guardian

A Individuals contact data is available in the black market

B C Is part of large groups with multiple unknown contacts

C F Engages in sexting (written and/or images)

E Leaves account open at a public place

C D Uses a 3rd party app

B An unknown contact

C Known contact

E Compromised account

E All or some information on a user's account is accessed

E Exhibits predatory behavior towards minor/youth *(e.g., phishing, CSAM)*

E Requests money

E Sends abusive content *(e.g., pornography, violent videos, misinformation)*

E Sends harmful link(s) to a user

E Cyberbullies

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

A Easy access to personal identifying data

Personal information on most messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

B Limited verification and consent focused features for contacts and groups

There is a lack of mechanisms for verifying contacts or groups, while permissions for new contacts and groups are either not set by default or are non-existent.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within the menu.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app.

E Limited support and lack of adequate reporting mechanisms

From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

F Limited content management tools

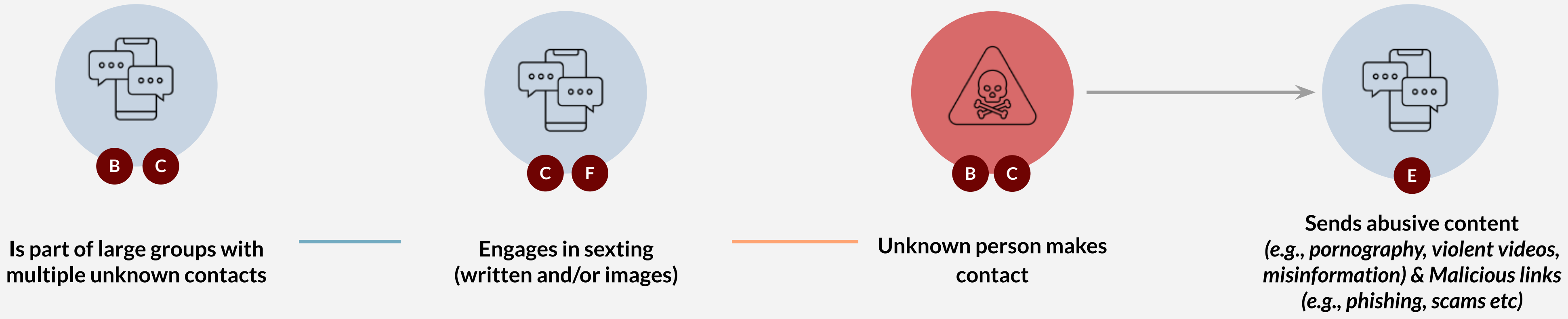
There are very few features that help individuals better manage and organize the content they receive.

Vulnerability to **targeted harassment for youth and young adults** - illustrative journey

Carlos - *At Risk Adolescent*

Carlos, who just turned 18, has been actively using WhatsApp since he was a 14-year-old. He was introduced to WhatsApp by his parents so that they could communicate with one another. He remembers that it was fairly easy to set up an account and start using WhatsApp. Over time, it has become his primary means of communicating not just with his parents but also friends, school administrators and others. Because of his love for technology, Carlos currently uses several platforms, including WhatsApp, Telegram and Discord, alongside other social media platforms for communication and socializing/learning.

> ***"When you download an app you never check what instructions or terms and conditions agreement it has, you just install it and that's it because you need it."***



Is part of large groups with multiple unknown contacts

Engages in sexting (written and/or images)

Unknown person makes contact

Sends abusive content ***(e.g., pornography, violent videos, misinformation) & Malicious links (e.g., phishing, scams etc)***

Carlos belongs to many groups on different private messaging platforms (e.g., Whatsapp, Telegram).

These groups are usually meant for entertainment/football, training/education, and other social activities.

Carlos being a sexually active young adult, actively experiments with new forms of erotic communications.

He often learns about these new forms of communications through online browsing and groups on private messaging apps.

Since Carlos is a member of many groups that do not have strict member and content restrictions, he often receives messages and contact requests from unknown contacts who want to befriend him or offer "opportunities".

Carlos often comes across adult/abusive content on private messaging apps. This tends to happen in unregulated groups as well as through personal contacts.

- ● Affected individual
- ● Individual causing the harm
- ● Individual behavior on a private messaging platform
- ● Individual behavior outside of a private messaging platform
- ● Related experiences that impact harm
- ● Psychological impact of this harm

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

B Limited verification and consent focused features for contacts and groups
There is a lack of mechanisms for verifying contacts or groups, while permissions for new contacts and groups are either not set by default or are non-existent.

C Generalized and hidden privacy and security controls for contacts and groups
Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within little-used menu structures.

E Limited support and lack of adequate reporting mechanisms
From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

F Limited content management tools
There are very few features that help individuals better manage and organize the content they receive.

Design opportunities

We identified six opportunities to address the design gaps that contribute to the proliferation of this harm.

Design Opportunity 1

Securing and/or modifying account information

Product design gaps addressed

A Easy access to personal identifying data

Design Opportunity 2

Improving verification & permission mechanisms

Product design gaps addressed

B Limited verification and consent-focused features for contacts and groups

Design Opportunity 3

Providing accessible & tailored security & privacy controls

Product design gaps addressed

C Generalized and hidden privacy & security controls for contacts & groups

Design Opportunity 4

Managing access to modified & third-party supporting platforms

Product design gaps addressed

D Infringement by modified (MOD) & third-party supporting apps ecosystem

Design Opportunity 5

Providing support mechanisms & emergency controls

Product design gaps addressed

E Limited support and lack of adequate reporting mechanisms

Design Opportunity 6

Improving administrative & management tools

Product design gaps addressed

F Limited content management tools

3.

Vulnerability to **manipulation** (misleading content, mis/disinformation) **or exposure to offensive content**

This harm covers the use of private messaging platforms to knowingly and unknowingly spread content that can be perceived as being hateful, offensive and/or misleading such as pornography, violent images and videos, and misinformation.

Individual's experience of this harm

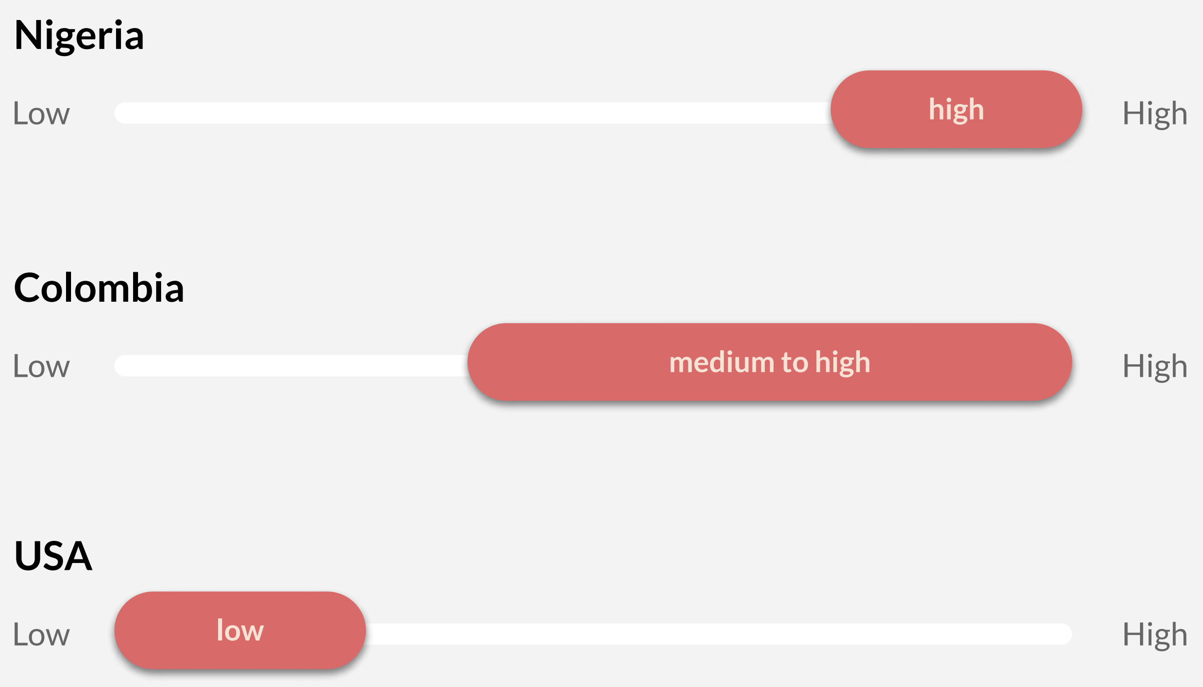
The circulation of offensive or misleading content on platforms was of high concern for the participants we spoke to in Nigeria and of medium to high concern for the participants we spoke to in Colombia. However, participants in the US largely opposed interventions geared at addressing this harm. This exposed two contradicting perspectives on ways to address this harm:

- **The Colombian & Nigerian perspective** - For most participants in Colombia and Nigeria, tackling this harm was important, particularly regarding the circulation of pornography. Tackling misinformation was also perceived as important, and participants in both countries placed great trust in the existence of a potential third-party organization that could be tasked with addressing the harm. Still, participants in both countries expressed some skepticism related to the specific ways that an intervention could effectively address this harm.
- **The US perspective** - Most participants we spoke to in the US were very vocal in expressing their concern that any efforts to address this harm by private messaging platform providers would likely have a negative impact on freedom of speech. Instead, American participants believed that the onus was on individuals themselves, not only to discern if something was misleading or correct but also to know how they could protect themselves, if they wanted to, from different forms of offensive content. This belief was particularly heightened for any intervention geared at the circulation of misinformation. Here, participants expressed overwhelming distrust in the involvement of any third-party organization that might be given the authority to address this harm. Political division, among other factors, would make the role of such an organization extremely polarizing.

The map on the next page presents a cross-country perspective on the circulation of offensive or misleading content on platforms from the participants in Colombia, Nigeria, and the United States. Particular attention is paid to the design gaps that contribute to the widespread experience of this harm by the participants we spoke with.

Level of concern expressed by participants

The scales below reflect the relative range of concern that participants we spoke with expressed for this harm across different countries.

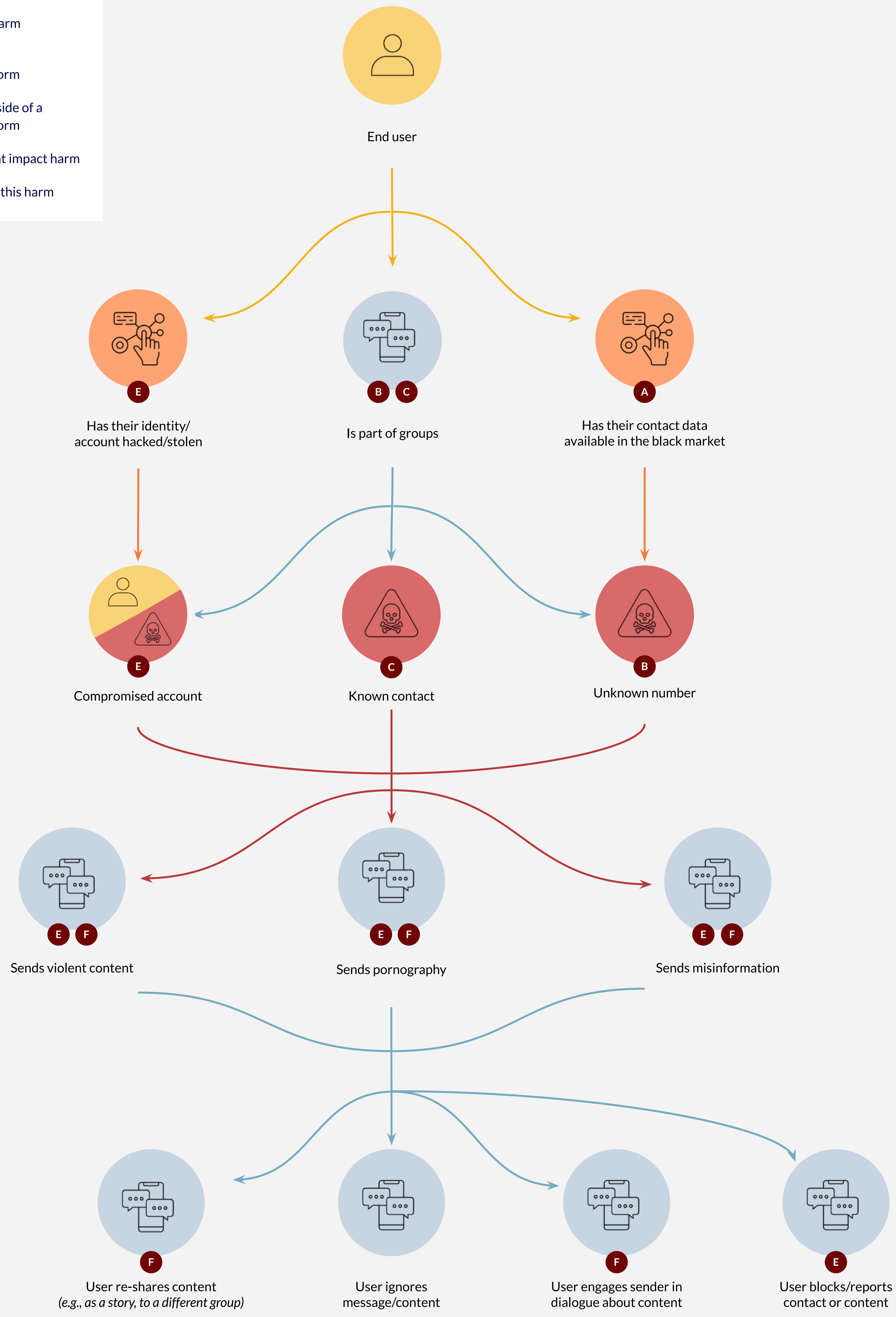


> "Misinformation is just a part of the internet and it is people's responsibility to know that and look out for themselves. Tech band-aids feel like they are not enough."
>
> 24-year-old man in the US, using Signal, and Telegram

> "I follow up with a phone conversation and see how true it is or go online to verify that story before I choose to re-broadcast it. But in some cases, many people don't verify, they just repost and repost, and it causes panic, and in less than 1 to 3 hours, they find out it's fake."
>
> 38-year-old Nigerian man

Understanding the **vulnerability to manipulation (misleading content, mis/disinformation) or exposure to offensive content** - illustrative process

- ● Affected individual
- ● Individual causing the harm
- ● Individual behavior on a private messaging platform
- ● Individual behavior outside of a private messaging platform
- ● Related experiences that impact harm
- ● Psychological impact of this harm


End user
E
Has their identity/ account hacked/stolen
B C
Is part of groups
A
Has their contact data available in the black market
E
Compromised account
C
Known contact
B
Unknown number
E F
Sends violent content
E F
Sends pornography
E F
Sends misinformation
F
User re-shares content (e.g., as a story, to a different group)
User ignores message/content
F
User engages sender in dialogue about content
E
User blocks/reports contact or content

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

A Easy access to personal identifying data

Personal information on most messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

B Limited verification and consent-focused features for contacts and groups

There is a lack of mechanisms for verifying contacts or groups, while permissions for new contacts and groups are either not set by default or are non-existent.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within the menu.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app.

E Limited support and lack of adequate reporting mechanisms

From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

F Limited content management tools

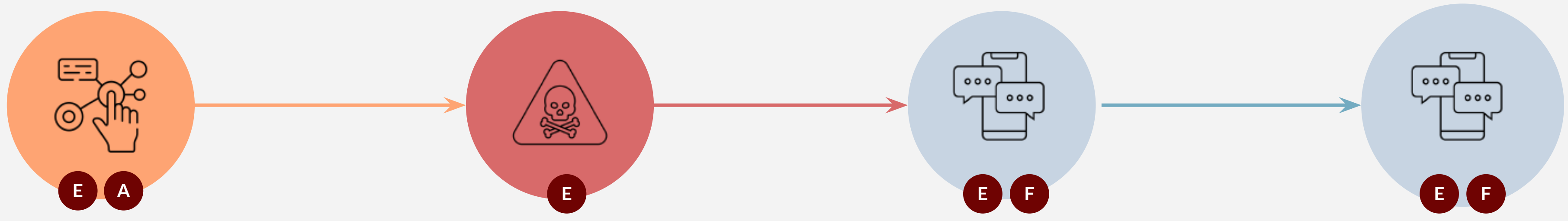
There are very few features that help individuals better manage and organize the content they receive.

Understanding the **vulnerability to manipulation (misleading content, mis/disinformation) or exposure to offensive content** - illustrative journey

Victor - *Citizen Journalist*

Victor is a 38-year-old development manager and community leader. He resides in northern Nigeria where he works with community organizations in assisting locals in empowerment initiatives and emergency support interventions. He primarily uses WhatsApp alongside Facebook Messenger and sometimes Telegram to communicate and organize community activities.

> ***"I follow up with a phone conversation and see how true it is or go online to verify that story before I choose to re-broadcast it. But in some cases many people don't verify, they just repost and repost and it causes panic and in less than 1 to 3 hours they find out it's fake."***



Has their identity/ account hacked/ stolen or Has their contact data available in the black market

Since Victor is a community leader, many residents come to him to request to use his phone to contact their loved ones in and outside of the country which often exposes his account to risks (e.g., malicious access to personal information/data, hacking etc)

Compromised account

Cases of account hacking have increased in the region where Victor resides. Often victims have no formal recourse to resolve such critical issues and they tend to give up on hacked account and create a new one losing important information in the process as well as having personal information/data in the hands of hackers

Hacker sends violent, disinformation or emergency scam content

Hackers often tend to use victims accounts to contact unsuspecting contacts through spread of disinformation and emergency scams.

Victims don't usually have an option to notify their contacts to be weary of messages sent to them from hacked accounts.

Other unsuspecting contact re-shares content (e.g., as a story, to a different group/contact) or unsuspecting contact engages sender in dialogue about content

Unsuspecting contacts who receive messages from compromised account of known contact, often tend to re-share public content received and sometimes send money to help their friend or family unaware that it was a scam

- 🟡 Affected individual
- 🔴 Individual causing the harm
- 🔵 Individual behavior on a private messaging platform
- ⚪ Individual behavior outside of a private messaging platform
- 🟠 Related experiences that impact harm
- 🟢 Psychological impact of this harm

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

A Easy access to personal identifying data

Personal information on most messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

E Limited support and lack of adequate reporting mechanisms

From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

F Limited content management tools

There are very few features that help individuals better manage and organize the content they receive.

Design opportunities

We identified six opportunities to address the design gaps that contribute to the proliferation of this harm.

Design Opportunity 1

Securing and/or modifying account information

Product design gaps addressed

A Easy access to personal identifying data

Design Opportunity 2

Improving verification & permission mechanisms

Product design gaps addressed

B Limited verification and consent-focused features for contacts and groups

Design Opportunity 3

Providing accessible & tailored security & privacy controls

Product design gaps addressed

C Generalized and hidden privacy & security controls for contacts & groups

Design Opportunity 4

Managing access to modified & third-party supporting platforms

Product design gaps addressed

D Infringement by modified (MOD) & third-party supporting apps ecosystem

Design Opportunity 5

Providing support mechanisms & emergency controls

Product design gaps addressed

E Limited support and lack of adequate reporting mechanisms

Design Opportunity 6

Improving administrative & management tools

Product design gaps addressed

F Limited content management tools

4. Vulnerability to **hacking, scamming, blackmailing, extortion, fraud, and harassment**

This harm covers the direct misuse of private messaging platforms by adults as distinct from those affecting youth and children. This includes common digital misuses such as the use of the platforms for scamming, conducting fraud schemes, hacking, or causing deliberate harassment.

Individual's experience of this harm

While the misuse of private messaging platforms was of extremely high concern to the participants we spoke to in Nigeria and Colombia, participants in the US did not mention commonly experiencing or being as concerned about this harm. Colombian and Nigerian participants shared similar concerns regarding platform misuse, yet Colombian participants experienced these misuse more frequently in connection with physical assault or other associated crimes. These crossover digital-to-physical crimes were often linked by individuals to the illegal sale of leaked contact databases that contain phone numbers and other personal information. For Nigerian participants, potential exposure to fraud or other scams was a primary risk associated with WhatsApp and other private messaging apps. The American participants we spoke with were instead more worried about the potential use of the platforms for aiding in organized crime, such as terrorism or media manipulation. That said, it is possible that this concern would have shown up much more prominently if our research had targeted specific groups with different vulnerabilities, such as the elderly.

The list below presents the most common forms of misuse concerns we heard from the participants we spoke with:

- Hacking *(e.g., cloning of private messaging app account)*
- Scamming
- Blackmailing and extortion
- Fraud
- Harassment and/or stalking
- Robbery

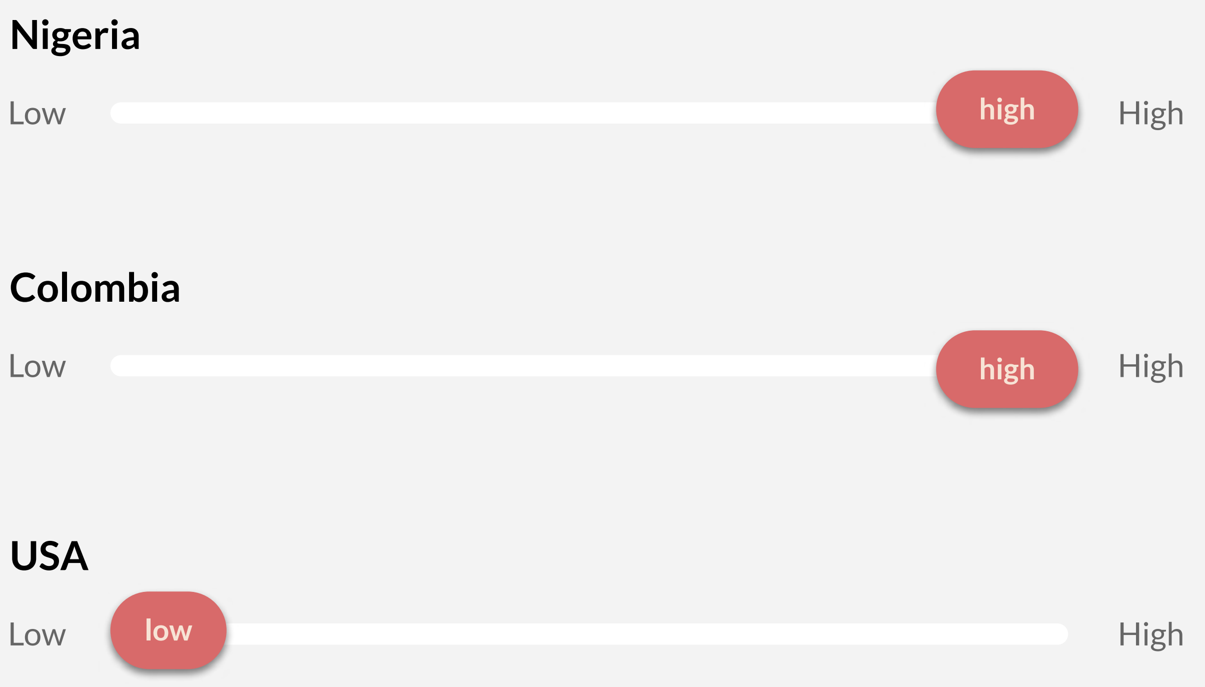
In the efforts to protect themselves against the harms listed above, participants in Nigeria and Colombia used distinct mechanisms. Below is a snapshot of differing behaviors across the two countries:

- In Nigeria the primary mechanism used by most participants was to turn on two-step verification. Individuals who are more technologically comfortable also relied on the use of third-party apps to verify links or for caller identification. Only a select few of those who were the most technologically comfortable knew that they could change their privacy settings so only contacts could add them to groups.
- In Colombia two-step verification was not mentioned by any of the participants we spoke with, instead participants spoke of having a general degree of 'vigilance' when communicating on the platforms. Like Nigeria, some technologically comfortable individuals also turned to third-party apps to verify links or for caller identification.

The map on the next page presents a cross-country perspective on the misuse of private messaging platforms from participants in Colombia, Nigeria and the United States. Particular attention is paid to the design gaps that contribute to the widespread experience of this harm by the participants we spoke with.

Level of concern expressed by participants

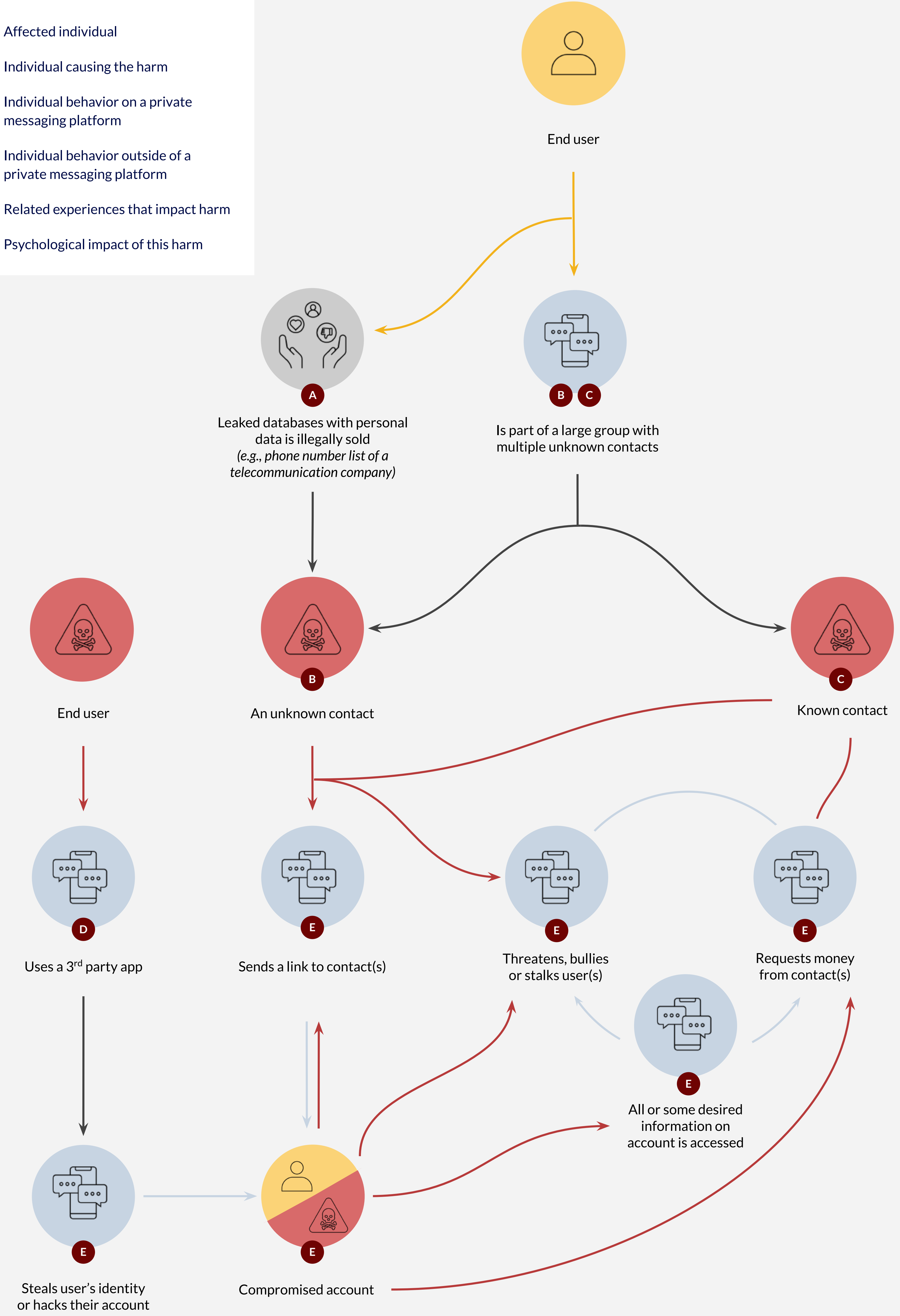
The scales below reflect the relative range of concern that participants we spoke with expressed for this harm across different countries.



> *"I set up two-step verification because anyone can takeover a WhatsApp account and mess with people. People get access through by dropping links on groups."*
>
> 49-year-old Nigerian man

Vulnerability **to hacking, scamming, blackmailing, extortion, fraud, and harassment** - illustrative process

- 🟡 Affected individual
- 🔴 Individual causing the harm
- 🔵 Individual behavior on a private messaging platform
- ⚪ Individual behavior outside of a private messaging platform
- 🟠 Related experiences that impact harm
- 🟢 Psychological impact of this harm



End user

A Leaked databases with personal data is illegally sold *(e.g., phone number list of a telecommunication company)*

B C Is part of a large group with multiple unknown contacts

End user

B An unknown contact

C Known contact

D Uses a 3rd party app

E Sends a link to contact(s)

E Threatens, bullies or stalks user(s)

E Requests money from contact(s)

E All or some desired information on account is accessed

E Steals user's identity or hacks their account

E Compromised account

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

A Easy access to personal identifying data

Personal information on most messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

B Limited verification and consent-focused features for contacts and groups

There is a lack of mechanisms for verifying contacts or groups, while permissions for new contacts and groups are either not set by default or are non-existent.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within the menu.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app.

E Limited support and lack of adequate reporting mechanisms

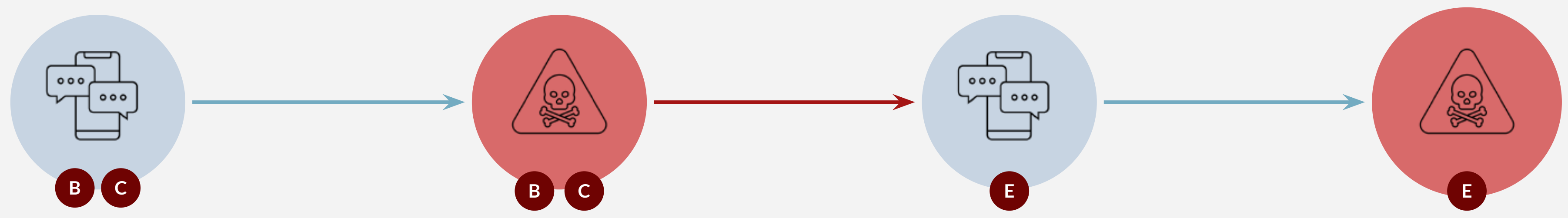
From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

Vulnerability **to hacking, scamming, blackmailing, extortion, fraud, and harassment** - illustrative journey

Diego - *Low Tech Entrepreneur*

Diego is 44 years old and runs a small transport company in Villeta. Due to the nature of his job, he is constantly multi-tasking, chasing payments to coordinating and resolving transportation issues. He heavily relies on WhatsApp and Telegram for the day-to-day operations of his business. He belongs to many groups on WhatsApp, and he finds these groups to be very valuable to him in marketing his business. He also uses private messaging apps to receive and share important traffic/emergency information with other drivers in his region.

> ***"I use an app called Troller, I don't have a good memory.... and with this can identify if a call that comes to my cell phone is reliable. It only works with calls, and there I can see if it is reliable or a scam or robbery."***



Is part of a large group with multiple unknown contacts

Receives content from unknown/known contact

Unknown contact sends a malicious a link to the user or known contact sends a request for money from contacts

Compromised account (e.g., individuals information is stolen or account hacked)

Diego belongs to many business groups (e.g., buyers and sellers) as well as other groups information groups (e.g., truckers and drivers, traffic updates etc).

Through these groups he interacts with many unknown contact and in the process his personal information is exposed to unknown contacts

Diego often receives information about new business opportunities or introductions to potential clients through known and unknown contacts.

Although he lacks tools to verify authenticity of unknown contacts before engaging

Through his varied communications with known and unknown contacts, he often receives suspicious messages (e.g., known contact claiming to be in an emergency and requiring money, or fraudulent links).

He sometimes calls the sender to confirm the emergency before sending money although this can be time consuming and costly.

Diego is worried about possibility of his account being hacked because he uses it for his business and would lose important information and contacts.

- ● Affected individual
- ● Individual causing the harm
- ● Individual behavior on a private messaging platform
- ● Individual behavior outside of a private messaging platform
- ● Related experiences that impact harm
- ● Psychological impact of this harm

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

B Limited verification and consent focused features for contacts and groups

There is a lack of mechanisms for verifying contacts or groups, while permissions for new contacts and groups are either not set by default or are non-existent.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within little-used menu structures.

E Limited support and lack of adequate reporting mechanisms

From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

Design opportunities

We identified five opportunities to address the design gaps that contribute to the proliferation of this harm.

Design Opportunity 1

Securing and/or modifying account information

Product design gaps addressed

A Easy access to personal identifying data

Design Opportunity 2

Improving verification & permission mechanisms

Product design gaps addressed

B Limited verification and consent-focused features for contacts and groups

Design Opportunity 3

Providing accessible & tailored security & privacy controls

Product design gaps addressed

C Generalized and hidden privacy & security controls for contacts & groups

Design Opportunity 4

Managing access to modified & third-party supporting platforms

Product design gaps addressed

D Infringement by modified (MOD) & third-party supporting apps ecosystem

Design Opportunity 5

Providing support mechanisms & emergency controls

Product design gaps addressed

E Limited support and lack of adequate reporting mechanisms

5.

Vulnerability to **encryption and data breaches via modified and third-party supporting platforms**

This harm covers the different ways that private messaging platforms security, privacy and encryption features are bypassed by users through the use of modified and third-party supporting private messaging platform apps; and the detrimental impact their adoption has on other individuals.

Individual's experience of this harm

This harm was expressed as being of medium to high concern for some of the participants we spoke to in Nigeria and Colombia but was not mentioned by any user we spoke to in the US. In both Nigeria and Colombia, participants rely heavily on WhatsApp as a standard and ubiquitous form of communication. participants in both countries reported expansive use of the app which goes beyond communication between friends and family *(e.g., work, news, networking, dating, emergency reporting, communication with businesses and services)*. In their pursuit of additional features to give them more control, some participants have taken to using modified (MOD) versions of private messaging platforms or third-party supporting apps. Some of these participants seem to not be aware that they are using a MOD private messaging platform app. Those that are, or participants who know someone that uses them, are concerned about the privacy and security risks they pose to others. But these concerns are not always sufficient to get them to stop using MODs entirely.

While reasons for using MOD and third-party supporting apps vary, below are some of the most common reasons we heard from participants:

Modified private messaging platforms

- **Individuals seeking greater privacy while communicating.** These individuals seek additional privacy controls that help them avoid social pressures. MODs allow them the ability to toggle on and off the communication of their own actions on WhatsApp *(e.g., user is online, read receipts, user is writing, last seen, private mode for chatting)*
- **Individuals seeking greater security while communicating** *(e.g., app lock, password lock for particular chats, ability to hide chats)*
- **Individuals seeking better content management** *(e.g., separate folders for DMs and group chats, configuration of automatic reply messages, ability to share large media files, recall and scheduling of WhatsApp messages, disabling of app notifications)*
- **Individuals seeking greater customization of the apps user interface design** *(e.g., changing fonts, features, and colors)*
- **Individuals seeking to preserve access to deleted messages** *(e.g., anti-delete features)*

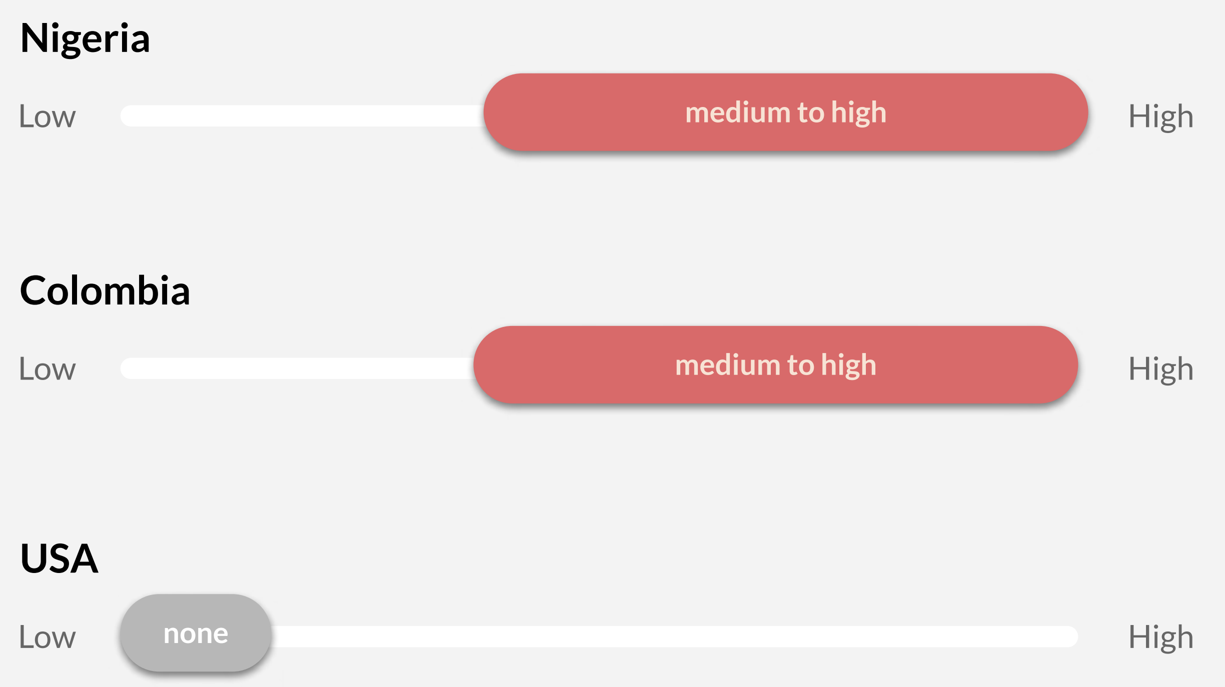
Third-party supporting apps

- Parents seeking to monitor their children's account
- Partners seeking to monitor their corresponding partners
- Individuals seeking private message reading
- Individuals seeking to preserve access to deleted messages
- Individuals seeking to download/ copy the stories of other contacts
- Individuals seeking to get easier contact adding features *(e.g., tapping a phone to get the WhatsApp account of another user)*

The map on the next page presents a cross-country perspective on the use of modified (MOD) & third-party supporting platforms from participants in Colombia, Nigeria, and the United States. Particular attention is paid to the design gaps that contribute to the widespread experience of this harm by the participants we spoke with.

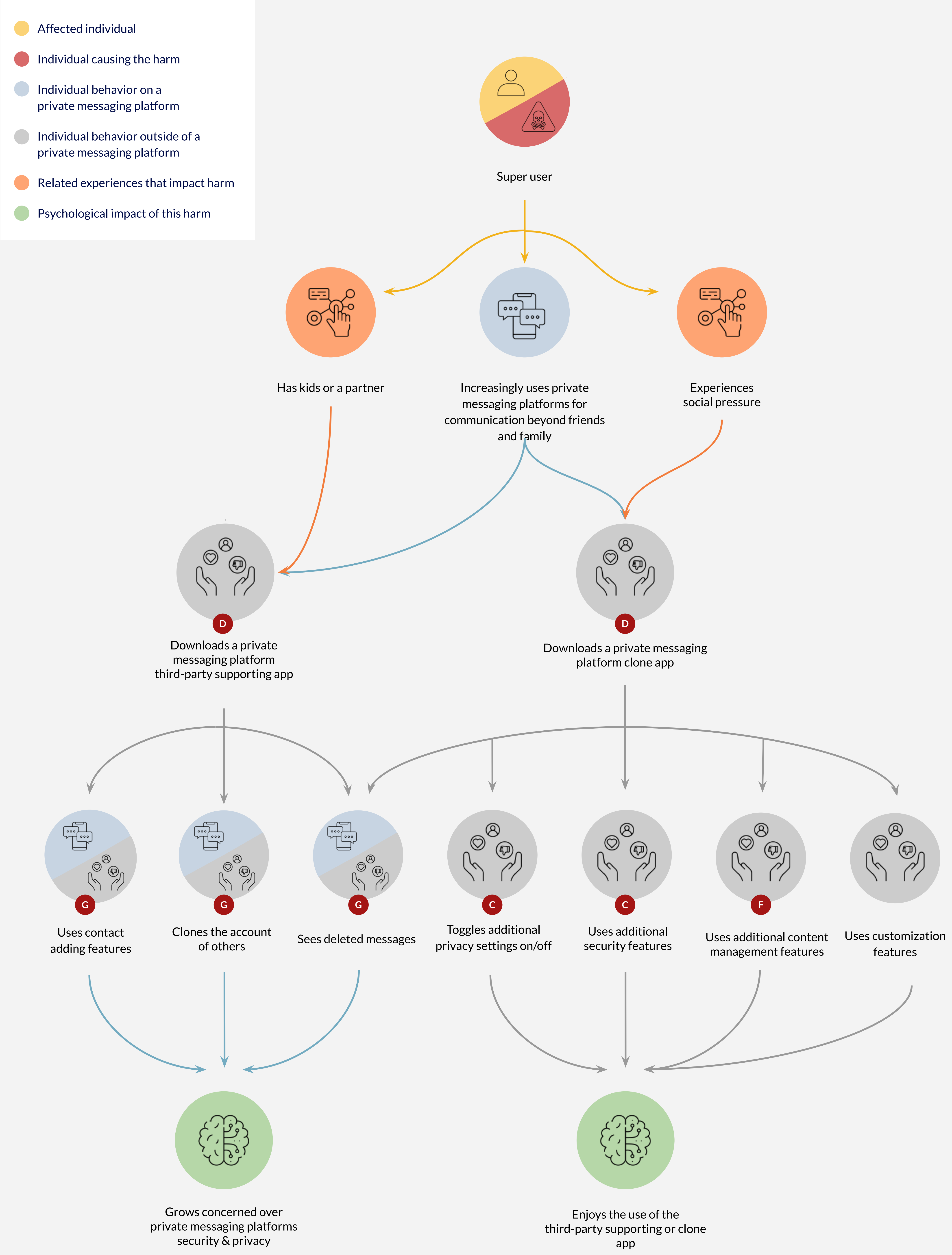
Level of concern expressed by participants

The scales below reflect the relative range of concern that participants we spoke with expressed for this harm across different countries.



> *"There is an app that duplicates WhatsApp accounts. I used to use it with my daughter so I could have more control over who she talks to, what she posts, the conversations she has, and until what time she uses the cell phone. After some time we stopped using it because she understood and educated herself about the use and management of these platforms."*
>
> 28 -year-old Colombian woman

Vulnerability to **encryption and data breaches via modified and third-party supporting platforms** - illustrative process



🟡 Affected individual

🔴 Individual causing the harm

🔵 Individual behavior on a private messaging platform

⚪ Individual behavior outside of a private messaging platform

🟠 Related experiences that impact harm

🟢 Psychological impact of this harm

Super user

Has kids or a partner

Increasingly uses private messaging platforms for communication beyond friends and family

Experiences social pressure

D Downloads a private messaging platform third-party supporting app

D Downloads a private messaging platform clone app

G Uses contact adding features

G Clones the account of others

G Sees deleted messages

C Toggles additional privacy settings on/off

C Uses additional security features

F Uses additional content management features

Uses customization features

Grows concerned over private messaging platforms security & privacy

Enjoys the use of the third-party supporting or clone app

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within little-used menu structures.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app without others knowing.

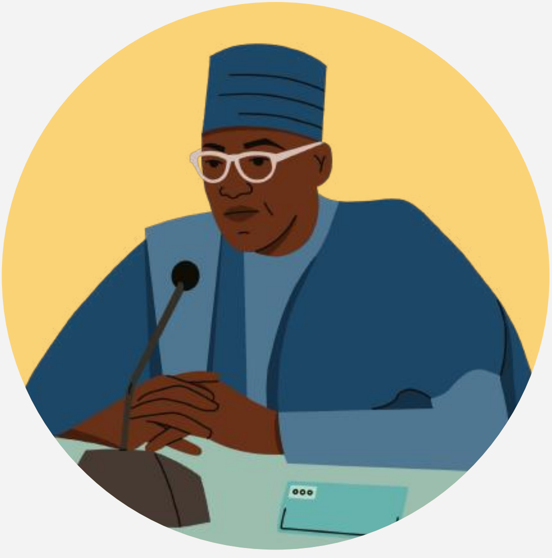
F Limited content management tools

There are very few features that help individuals better manage and organize the content they receive.

G Lack of transparency regarding access to personal data

There are gaps around who can access, use, and potentially misuse user data *(e.g., how and if companies and governments can access user data).* At the same time, messaging platforms don't communicate transparently and in an user-friendly way how they manage and protect personal data.

Vulnerability to **encryption and data breaches via modified and third-party supporting platforms** - illustrative journey



Emmanuel -

Low Tech Influencer

Emmanuel is a 48- year-old church leader who lives in Lagos, Nigeria. He runs a local church and is also an inspirational speaker and life coach to his followers. He often meets his followers through scheduled public gatherings, however after he started using a private messaging app, he quickly realized the convenience it offered through individual and group communications, so he now heavily uses private messaging apps to engage his followers. He primarily uses WhatsApp to communicate, but he also connects on other platforms like Telegram in order to reach as many followers as possible, wherever they are most comfortable.

> ***“Sometimes, especially when counseling, the information shared is very sensitive. For instance, if you are doing counseling and you should separate from your husband, this is sensitive and you need security.”***

Increasingly uses private messaging platforms for communication beyond friends and family

Emmanuel has become a frequent user of private messaging apps, he especially finds them very helpful to communicate with larger groups of people as well as the convenience of talking to contacts wherever they are.

He perceives private messaging platforms as an important channel to engage his growing followers.

Downloads a private messaging platform third-party supporting app

Due to the sensitive nature of Emmanuel’s communications (e.g., marriage counseling via private message apps), he explores ways to secure his account and the information in it and sometimes ends up downloading third-party apps that offer such security and privacy options to use alongside related private message apps.

Uses contact adding features or finds out some individuals can see deleted messages

Emmanuel frequently adds new contacts to his growing contact list and, sometimes some of his contacts use unverified messaging apps which are capable of bypassing user actions/settings (e.g., access to deleted messages or restricted posts)

Grows concerned over private messaging platforms security & privacy

Emmanuel is often worried about the privacy and security of his account and its information, especially recognizing his struggles with technology which makes him unable to fully maximize the privacy and security settings of preferred apps or being able to discern malicious links and other supplementary apps.

- ● Affected individual
- ● Individual causing the harm
- ● Individual behavior on a private messaging platform
- ● Individual behavior outside of a private messaging platform
- ● Related experiences that impact harm
- ● Psychological impact of this harm

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app without others knowing.

G Lack of transparency regarding access to personal data

There are gaps around who can access, use, and potentially misuse user data *(e.g., how and if companies and governments can access user data)*. At the same time, messaging platforms don’t communicate transparently and in an user-friendly way, how they manage and protect personal data.

Design opportunities

We identified four opportunities to address the design gaps that contribute to the proliferation of this harm.

Design Opportunity 3

Providing accessible & tailored security & privacy controls

Product design gaps addressed

C Generalized and hidden privacy & security controls for contacts & groups

Design Opportunity 4

Managing access to modified & third-party supporting platforms

Product design gaps addressed

D Infringement by modified (MOD) & third-party supporting apps ecosystem

Design Opportunity 6

Improving administrative & management tools

Product design gaps addressed

F Limited content management tools

Design Opportunity 7

Providing data use transparency & the ability to manage data

Product design gaps addressed

G Lack of transparency regarding access to personal data

6.

Vulnerability to **digital surveillance and monitoring**

This harm covers the different ways users on private messaging platforms experience and perceive infringement of their privacy. It speaks to concerns of surveillance and monitoring of conversations by those in power, be it for business reasons (targeted advertising) or political reasons (surveillance of critics and activists).

Individual's experience of this harm

This harm was of low to medium concern for participants we spoke to in Nigeria and Colombia. For participants in the US, this was of much higher concern. Apart from activists and journalists, Colombian and Nigerian participants reported that they are not worried about government surveillance since they don't believe that they have anything to hide. Some participants even advocated for government surveillance to curb criminal activities on digital platforms. These individuals were more concerned about access and/or monitoring of private conversations by other entities/contacts.

By contrast, many users in the US are highly concerned about the infringement of their privacy on private messaging platforms. The US is comprised of a variety of users who all agreed on the need for stronger measures to protect users' privacy and security online. Activists in the US reported the use of sophisticated tools for surveillance by government agencies and other tech organizations. Many privacy-sensitive users have resorted to using other forms of communications outside of private messaging apps (e.g., emails, in-person) to pass sensitive information.

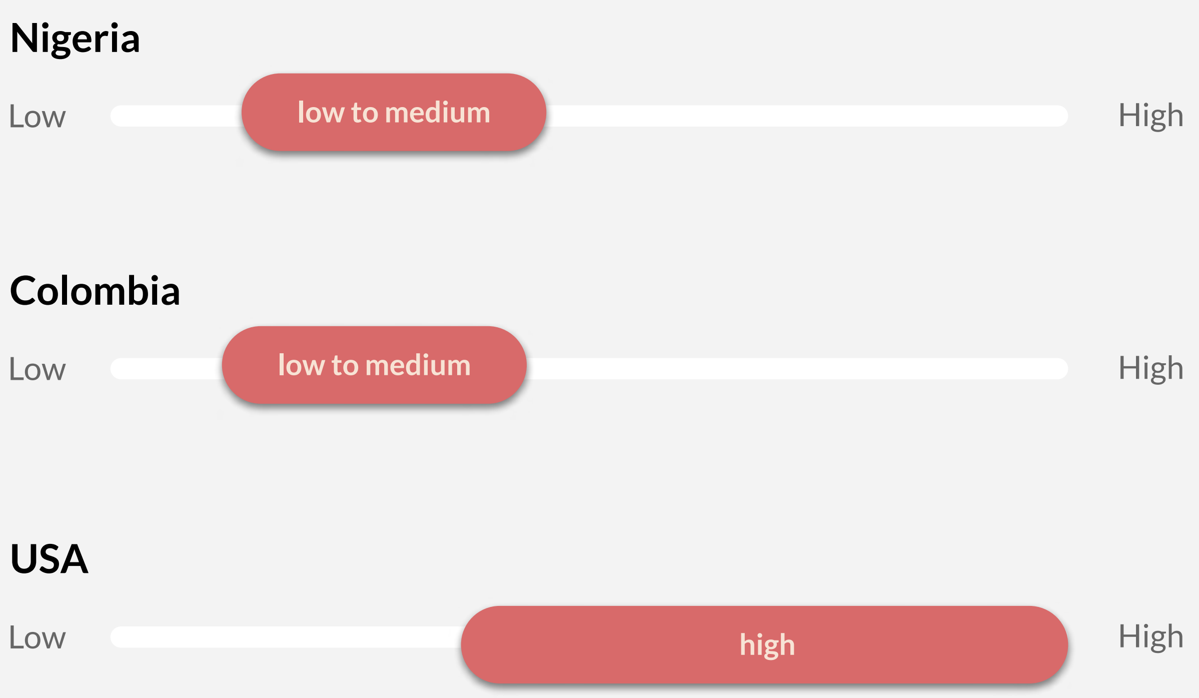
Tracking of movements using geolocation was reported as the main concern among activists/journalists in both Colombia and the US. These location services are relied on by activists to rally support or help during demonstrations, especially in Colombia. But they are also used by authorities to track the gathering of demonstrators which can lead to violent confrontations. Activists in the US are even warier of location data and often leave their phones behind for sensitive meetings. Methods of surveillance/monitoring of concerns to users in Nigeria, Colombia and the US include:

- Monitoring of conversations for targeted ads (e.g., through digital assistants Siri/Alexa)
- Monitoring and tracking of movements
- Using devices to eavesdrop
- Spying (e.g., in private messaging groups by law enforcement)
- Remote screen and audio recording

The map on the next page presents a cross-country perspective on surveillance and monitoring based on feedback from users in Colombia, Nigeria and the United States. Particular attention is paid to the design gaps that contribute to the widespread concerns about this harm among some of the users we spoke with.

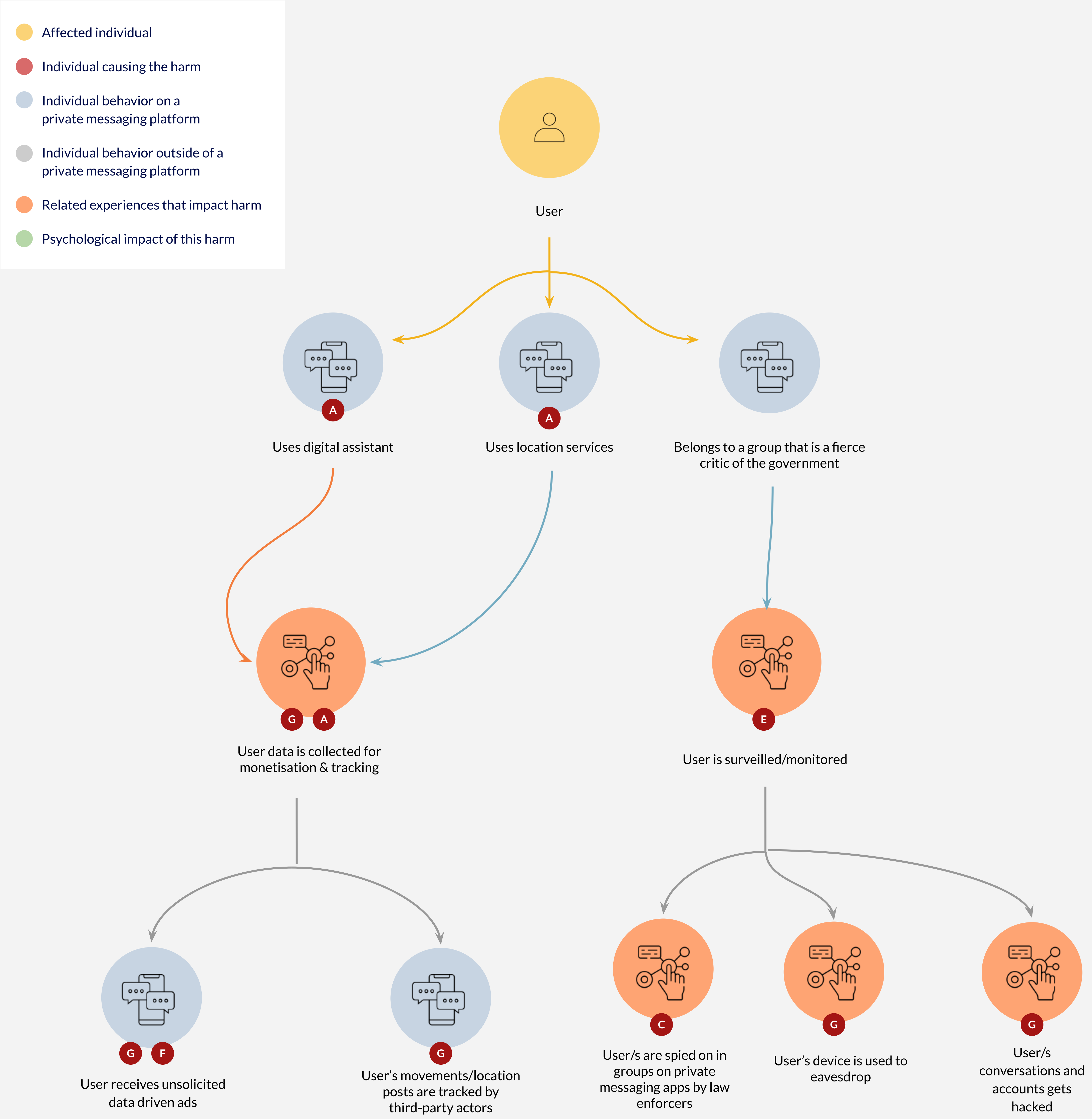
Level of concern expressed by participants

The scales below reflect the relative range of concern that participants we spoke with expressed for this harm across different countries.



> *"During the protests WhatsApp helped a lot because people who were lost in the streets could easily locate us because Facebook would get blocked and the networks were blocked as well."*
>
> 24 -year-old Colombian activist

Vulnerability to **digital surveillance and monitoring** - illustrative process



● Affected individual

● Individual causing the harm

● Individual behavior on a private messaging platform

● Individual behavior outside of a private messaging platform

● Related experiences that impact harm

● Psychological impact of this harm

User

A Uses digital assistant

A Uses location services

Belongs to a group that is a fierce critic of the government

G A User data is collected for monetisation & tracking

E User is surveilled/monitored

G F User receives unsolicited data driven ads

G User's movements/location posts are tracked by third-party actors

C User/s are spied on in groups on private messaging apps by law enforcers

G User's device is used to eavesdrop

G User/s conversations and accounts gets hacked

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

A Easy access to personal identifying data

Personal information on most messaging platforms is easily accessible, while the use of phone numbers as account identifiers makes it easy to connect with any contact.

C Generalized and hidden privacy and security controls for contacts and groups

Security and privacy controls are presented as generic settings applied equally to all contacts and groups while also remaining hidden behind multiple steps within complex menu structures.

E Limited support and lack of adequate reporting mechanisms

From tech literacy and customer support to emergency and reporting tools, there are limited to no support mechanisms available. Those that exist are not perceived as being useful or adequately functional.

G Lack of transparency regarding access to personal data

There are gaps around who can access, use, and potentially misuse user data (*e.g., how and if companies and governments can access user data)*. And messaging platforms don't communicate transparently and in an user-friendly way how they manage and protect individual's data.

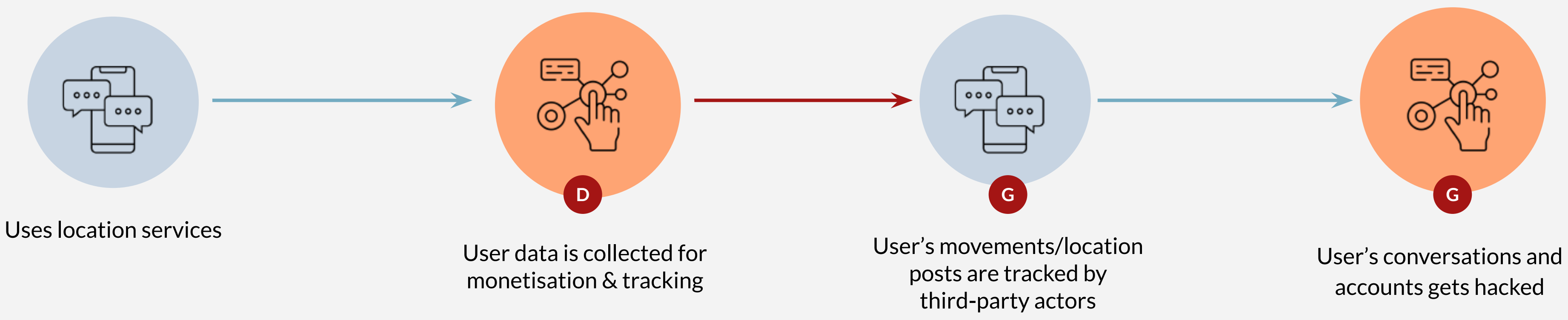
Vulnerability to **digital surveillance and monitoring** - illustrative journey

Barbara -

Concerned Activist

Barbara is a 24-year-old Colombian who is self-employed and lives in Bogota. She is passionate about human rights and actively involved in protests and championing the rights of others. She is very concerned and protective of her privacy when using private messaging which deeply influences her adoption and usage of different platforms. Also, due to her activism, Barbara has become a target of government surveillance from time-to-time, so her safety is also becoming a cause for concern. She used to use WhatsApp as her primary private messaging app but switched to Telegram when WhatsApp updated its privacy policies, which made her distrustful of the platform.

> ***"A password in this day and age feels rudimentary, because if someone really wants to, they can still access your account."***



Uses location services

D User data is collected for monetisation & tracking

G User's movements/location posts are tracked by third-party actors

G User's conversations and accounts gets hacked

Barbara uses location services frequently to share her location with friends and loved ones, especially during trips and geo-tagging posts on social media.

Besides social use, Barbara and her activist friends often use location services to find each other and rally support and help during street demonstrations.

Government agencies, including law enforcement, frequently collect/monitor location data of targeted groups or individuals for monitoring of movements and coordinated crackdowns against gatherings and demonstrations.

Other unauthorized contacts (e.g., law enforcement agents or stalkers) can track Barbara's location information to monitor her movements and activities due to the lack of better location privacy tools.

Barbara often receives strange links and contact invites from unknown contacts which leads her to believe that either someone is actively trying to get access to her account or she is being monitored. But due to a lack of sufficient verification and safety tools, she often struggles to discern the authenticity of a new contact or ways to better secure her account.

- 🟡 Affected individual
- 🔴 Individual causing the harm
- 🔵 Individual behavior on a private messaging platform
- ⚪ Individual behavior outside of a private messaging platform
- 🟠 Related experiences that impact harm
- 🟢 Psychological impact of this harm

Platform design gaps that enable this harm

Below is a list of the relevant platform design gaps that enable behaviors that exacerbates this harm.

D Infringement by modified (MOD) and third-party supporting apps ecosystem

There are multiple modified (MOD) and third-party supporting private messaging platform apps that offer individuals additional features that they can use in combination with or in replacement of their private messaging platform app without others knowing.

G Lack of transparency regarding access to personal data

There are gaps around who can access, use, and potentially misuse user data (*e.g., how and if companies and governments can access user data*). At the same time, messaging platforms don't communicate transparently and in an user-friendly way how they manage and protect personal data.

Design opportunities

We identified seven opportunities to address the design gaps that contribute to the proliferation of this harm.

Design Opportunity 1

Securing and/or modifying account information

Product design gaps addressed

A Easy access to personal identifying data

Design Opportunity 2

Improving verification & permission mechanisms

Product design gaps addressed

B Limited verification and consent-focused features for contacts and groups

Design Opportunity 3

Providing accessible & tailored security & privacy controls

Product design gaps addressed

C Generalized and hidden privacy & security controls for contacts & groups

Design Opportunity 4

Managing access to modified & third-party supporting platforms

Product design gaps addressed

D Infringement by modified (MOD) & third-party supporting apps ecosystem

Design Opportunity 5

Providing support mechanisms & emergency controls

Product design gaps addressed

E Limited support and lack of adequate reporting mechanisms

Design Opportunity 7

Providing data use transparency & the ability to manage data

Product design gaps addressed

G Lack of transparency regarding access to personal data

Thank you

Dalberg

SUPPORTED BY

OMIDYAR NETWORK™