

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."

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.

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.

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

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

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**

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.

Thank you

Dalberg

SUPPORTED BY

